

Apprentissage et classification pour des systèmes complexes

Applications aux systèmes automatiques et à la gestion du risque médical

Soutenue le 04 Juillet 2017 à l'Université de Lille 1
Pour l'obtention de

l'Habilitation à Diriger des Recherches

Spécialité : Génie informatique, Automatique, Traitement du signal et des images
par

Salah ZIDI

Maitre-assistant – Université Qassim (Arabie Saoudite)

Composition du jury

<i>Rapporteurs :</i>	Prof. Ismail BISKRI	Université du Québec à Trois-Rivières, Canada
	Prof. Hassani MESSAOUD	ENIM, Tunisie
	Prof. Stéphane CANU	INSA Rouen, France
<i>Examineurs :</i>	Prof. Salah MAOUCHE	Université de Lille1, France
	Prof. Gilles GONCALVES	Université d'Artois, France
	Prof. Slim HAMMADI	Ecole Centrale de Lille, France
<i>Invité :</i>	DR. Christophe DI POMPEO	Université de Lille 2, France

A mes parents

A ma femme et mes enfants

A toute ma famille

Remerciements

J'ai été constamment soutenu par l'amitié et la confiance de mon directeur de thèse et garant scientifique, Salah MAOUCHE. Il m'a beaucoup appris. Je lui adresse toute ma reconnaissance pour sa disponibilité, ses encouragements et ses conseils.

Je remercie Messieurs les professeurs Ismail BISKRI, Stéphane CANU et Hassani MESSAOUDI qui ont accepté les tâches ingrates d'évaluer ce travail et de rédiger un rapport sur ce manuscrit.

Je remercie également Messieurs les professeurs Slim HAMMADI, Gilles GONCALVES et Christophe DI POMPEO d'avoir accepté de participer à mon jury de HDR.

J'adresse ma gratitude aux sociétés BIOMEDIQA, ARCHIMED et ARIZONA au sein desquelles j'ai pu réaliser une bonne partie de ces travaux de recherche. C'est aussi l'occasion de remercier tous mes anciens collègues qui ont été témoins de mon parcours professionnel.

Monsieur Fouad MAALLOUL, PDG de la société BIOMEDIQA, et Monsieur Mongi ZIDI, PDG du groupe ARCHIMED, m'ont accueilli et m'ont ouvert les portes de la recherche et de l'innovation dans un monde industriel. Je les en remercie.

Pendant ces trois dernières années au sein de la CBE à l'université AL Qassim, j'ai eu l'occasion de travailler avec des enseignants-chercheurs du département MIS qui m'ont aidé à lancer des projets de recherche et à travailler dans de bonnes conditions. Ils m'ont donné d'indispensables coups de main au cours de ces années. Je remercie messieurs les docteurs Mohamed ALODIB, Tarek MOULAHI, Elamri LAOUAMER, Bachir ALAYA et Sami TWIL. Ils sont devenus plus que des collègues de travail.

J'ai eu le plaisir de coopérer avec des collègues de différents laboratoires et équipes de recherche. Je tiens à remercier les professeurs Moufida KSOURI, Kaouther LAABIDI, Kamel ABDERREHIM et Jean-Pierre BOUREY pour leur confiance.

Tous les travaux présentés dans ce manuscrit n'auraient pas vu le jour sans le travail des doctorants que j'ai eu le plaisir d'encadrer. Je remercie Sabra ELFERCHICHI, Foued THELJANI, Mounira TARHOUNI, Lamaa SALLAMI et Thibault JULIEN pour leur confiance et le temps consacré à notre collaboration.

J'aimerais finir par remercier particulièrement mes parents et ma femme pour m'avoir encouragé et pour m'avoir toujours soutenu.

Très nombreux sont ceux que je n'ai pas pu nommer, qu'ils trouvent ici l'expression de ma gratitude.

Avant-propos

Ce document retrace mes dix années de carrière après l'obtention d'un doctorat en automatique et informatique industrielle. Pendant ces années je suis passé d'un enseignant-chercheur à l'université Lille 1 à un responsable recherche et innovation au sein du groupe Archimed et d'autres entreprises notamment Biomedica et Arizona-CT. Puis de nouveau en poste d'enseignant-chercheur (maitre-assistant) au sein de l'université d'AlQassim en Arabie Saoudite.

Ingénieur ou théoricien ?

A 18 ans et après avoir eu mon Bac, j'ai hésité avant de m'inscrire dans un parcours d'étude universitaire pour devenir ingénieur. D'un côté, ma passion les mathématiques et la magie de la recherche fondamentale et de l'autre côté, la réalité du marché de travail et mes rêves de créateur innovant dans des domaines de l'ingénierie ont rendu le choix difficile. "Il y a beaucoup de math en ingénierie" est une phrase que j'ai beaucoup entendue et qui a fait pencher la balance du côté de l'ingénierie tout en espérant apprendre et surtout appliquer les théories mathématiques dans des domaines industriels. C'est peut être aussi le chemin le plus court vers l'innovation. Enfin, j'ai choisi de m'inscrire dans une école préparatoire pour entrer dans une école d'ingénieur. Sans le prouver, j'étais convaincu par ce choix. Ce n'est pas « scientifique » mais je n'étais pas encore chercheur. La réussite m'a encore poussé vers ce chemin et j'ai terminé par avoir la gratification du meilleur projet universitaire en Tunisie en 2003.

Néanmoins, cette récompense ne m'a pas fait oublier mes anciens envies pour la recherche fondamentale et je suis revenu à la question de départ. Cette fois avec un peu d'expérience (diplôme d'ingénieur, des stages et un mastère recherche) j'ai pu choisir tout en me basant sur des arguments et une approche relativement fiable. J'ai donc continué dans un parcours de recherche en m'inscrivant en thèse à l'université Lille1, Sciences et Technologies.

Ma thèse de doctorat est-elle un point de départ ?

Pendant ma thèse, j'ai réussi à proposer une méthode d'aide à la prise de décision pour la régulation des réseaux de transport multimodal. Basée sur l'apprentissage statistique, cette approche m'a permis de manipuler des fondements théoriques capables de modéliser des systèmes complexes. Étant donné que les systèmes actuels sont de plus en plus complexes, ces approches issues de l'Intelligence Artificielle (IA) sont très utiles. En effet, Cette complexité rend les problèmes liés à ces systèmes de plus en plus difficiles. Qu'ils soient liés à la modélisation, la spécification ou l'implémentation et la commande allant jusqu'au diagnostic (fonctionnement), ces problèmes peuvent être souvent traités en tant que problèmes de prise de décision. Et les méthodes de l'IA ont toujours montré une grande fiabilité avec des résultats attractifs.

Ainsi ma première réussite avec une technique d'apprentissage statistique m'a motivé d'aller plus loin sur l'application de ces nouvelles approches pour la prise de décision dans des systèmes plus complexes. Bien qu'elles soient basées sur des fondements théoriques très robustes, ces méthodes ont des limites qui méritent plus d'exploration par les chercheurs. Par exemple, pendant ma thèse j'étais confronté à un premier problème de sélection d'attributs pertinents.

À partir de ce constat, j'ai commencé à monter un axe de recherche dans ce contexte.

Pour ces raisons, je suppose que la thèse était mon point de départ pour la carrière d'un chercheur en analyse de données. Convaincu par les résultats de ces nouvelles solutions et par l'adéquation des problèmes avec la prise de décision, Je les ai proposés pour résoudre des problèmes industriels réels en collaboration avec des industriels notamment Biomedica et Archimed.

Une validation industrielle de la recherche...

En poste d'ATER à l'Université Lille 1, j'ai commencé à collaborer avec des entreprises de la région lilloise pour le montage de projets de recherche dans le contexte d'analyse de données et des systèmes d'aide à la décision. Encore une fois, je suis revenu sur la question de départ : vaut-il mieux d'aller vers l'ingénierie et le monde industriel pour réaliser des projets d'innovation ? Les propositions que j'ai eu, de la part de ces sociétés, m'ont garanti de rester sur des projets de recherche en lien avec l'axe que j'essaye de développer. J'ai alors intégré l'équipe R&D de la société Biomedica, avec comme premier projet, le développement de la plate-forme RMP qui sera présentée dans ce mémoire. Puis, j'ai accompagné l'entreprise Arizona et le groupe Archimed sur des projets de recherche et d'innovation.

Bien que la validation et la réalisation industrielle aient donné plus de valeur à mes contributions, j'ai toujours préféré rester sur le volet recherche et innovation. En 2014, j'ai décidé de revenir à l'enseignement mais dans un environnement différent. J'ai eu le poste de maître assistant à l'université d'ALQassim en Arabie Saoudite...

Table des matières

Avant-propos.....	3
Première partie : Rapport de Synthèse.....	15
1. Notice individuelle	16
1.1. État Civil	16
1.2. Formation Universitaire	17
1.3. Emplois Occupés.....	18
2. Activités pédagogiques	19
2.1. Bilan général	19
2.2. Matières enseignées par établissement	20
2.2.1. Université d'Al Qassim (Arabie Saoudite)	20
2.2.2. Institut Universitaire de Technologie (IUT A) Lille (France)	23
2.2.3. Université Lille1 (France)	24
2.2.4. Telecom Lille (France)	27
2.2.5. Ecole Nationale d'Ingénieur de Sfax-ENIS (Tunisie).....	28
2.2.6. Université Libre de Sfax (Tunisie)	28
2.3. Productions Pédagogiques	29
2.4. Activités d'intérêt collectif.....	30
2.5. Encadrements de Projets et suivi de stages :.....	31
2.6. Formations pédagogique Complémentaires.....	31
3. Activités et responsabilités administratives	31
4. Activités de Recherche	32
4.1. Mastère recherche	32
4.2. Doctorat.....	32
4.3. Activités de recherche post doctorales.....	34
4.3.1. Contexte Général.....	34
4.3.2. Sélection de variables et extraction de caractéristiques	34
4.3.3. Classification et régression pour les systèmes automatiques complexes	35
4.3.4. Gestion du risque médical.....	37
4.3.5. Reconnaissance de caractères.....	37
5. Encadrement.....	37
5.1. Co-encadrement de thèses	37
5.1.1. Thèses soutenues.....	37

5.1.2.	Thèses en cours	39
5.2.	Encadrement de Mastères Recherche.....	39
6.	Participation et Coordination de Projets.....	40
6.1.	Projet KACST N°34-440.....	40
6.2.	Projet NOOR N° NRC1-126	40
6.3.	Projet interne « Qassim University N° 2984 »	40
6.4.	Projet interne « Qassim University N° 1250 »	41
6.5.	Projet SYRACUSE.....	41
6.6.	Projet Stade Ubiquitaire	41
6.7.	Projet ATRAQ.....	42
7.	Rayonnement scientifique.....	42
7.1.	Comités d'Organisation de Sessions	42
7.2.	Participation à des Jurys de Thèses	42
7.3.	Évaluation pour des Revues	42
7.4.	Évaluation pour des Conférences.....	42
7.5.	Participation aux Groupes de Recherche et Comités scientifiques.....	42
8.	Publications.....	43
8.1.	Revue Internationale avec Comité de Lecture.....	43
8.2.	Chapitres d'Ouvrage Collectif	44
8.3.	Congrès Internationaux avec Actes et Comité de Lecture.....	44
9.	Activités industrielles	46
9.1.	Principales missions par entreprise	46
9.1.1.	ARCHIMED : Docteur-Ingénieur Recherche et Innovation (2010 – 2013)	46
9.1.2.	BIOMEDIQA : Responsable R&D (2009).....	47
9.1.3.	ARIZONA-CT : Consultant Recherche et Innovation (2009).....	48
9.2.	Développement d'outils logiciels	49
9.2.1.	CRYSTAL FRAMEWORK - socle documentaire et applicatif commun	49
9.2.2.	DocFactory	49
9.2.3.	RMP Risk Management Platform.....	49
9.2.4.	DOSITRACE	50
9.2.5.	Simplicit-e.....	50
Deuxième partie : Mémoire de recherche.....		51
Chapitre 1 : Sélection et Extraction d'attributs pour les problèmes de classification		52
1.	Introduction	53

2.	Classification automatique de données.....	53
2.1	Apprentissage supervisé.....	54
2.2	Apprentissage non-supervisé.....	54
2.3	Apprentissage semi-supervisé.....	55
2.4	Apprentissage actif	55
3.	Défis de la classification de données de grande dimension	55
3.1	Fléau de dimensionnalité	55
3.2	Phénomène de l'espace vide.....	56
3.3	Phénomène de Sur-apprentissage	56
4.	Réduction de dimension	57
5.	Sélection de variables (attributs).....	58
6.	Extraction d'attributs	60
6.1	Approches linéaires	60
6.2	Approches non linéaires.....	61
7.	Nouvelle approche de sélection de variables	62
7.1	Description de la solution proposée	62
7.2	Algorithme d'optimisation	63
7.3	Application à la régulation des réseaux de transport.....	65
7.4	Résultats	65
8.	Approche par classification non supervisée pour l'extraction d'attributs	67
8.1	Principe de la méthode.....	67
8.2	Mesure de similarité	68
8.3	Classification des attributs.....	70
9.	Information mutuelle pour l'extraction d'attributs	71
9.1	Principe de l'Information Mutuelle	71
9.2	Estimation par fenêtre de Parzen.....	72
9.3	Information mutuelle quadratique	73
9.4	Information Mutuelle pour la sélection et l'extraction d'attributs	74
9.5	Approche proposée pour l'extraction	74
9.5.1	Formulation basée sur IM	74
9.5.2	Minimisation de l'IM par descente de gradient	76
9.5.3	Algorithme d'extraction avec IM.....	78
10.	Expérimentations.....	79
10.1	Expérimentation sur les base UCI	79

10.2	Extraction pour la reconnaissance de visage	81
10.3	Extraction pour le diagnostic d'un procédé industriel.....	82
Chapitre 2 : Apprentissage et classification pour l'identification et le diagnostic des systèmes automatiques complexes		84
1.	Introduction	85
2.	SVM pour l'identification des systèmes non linéaire	85
2.1	De la modélisation mathématique..... aux nouvelles méthodes.....	85
2.1.1	Modèles non linéaires	86
2.1.2	Modèles dynamiques non linéaires	88
2.1.3	...Aux nouvelles approches de l'identification	89
2.2	Théorie de l'apprentissage statistique.....	90
2.2.1	Pour la classification.....	90
2.2.2	Pour la régression	93
2.3	Least Squares Support Kernel Machines (LS-SKM) pour l'identification des systèmes	96
2.3.1	Least Squares Support Vector Machines (LS-SVM)	96
2.3.2	Approche multi noyaux (Multi-kernel LS-SVM)	97
2.3.3	Least Squares Support Kernel Machines (LS-SKM).....	99
2.3.4	Validation	101
2.4	Support Kernel Regression (SKR) pour l'identification des systèmes	103
3.	Apprentissage statistique pour l'identification des systèmes dynamiques hybrides ...	105
3.1	Modélisation des systèmes dynamiques hybrides	105
3.1.1	Approche de modélisation discrète	105
3.1.2	Approche modélisation mixte	106
3.1.3	Approche modélisation continue.....	106
3.1.4	Problème d'identification pour les systèmes dynamiques hybrides.....	106
3.2	Classification pour l'estimation de l'état discret	107
3.2.1	Estimation hors ligne.....	107
3.2.2	Estimation en ligne	108
3.3	Régression pour l'identification des systèmes linéaire commutant	110
4.	Classification dynamique pour le diagnostic des systèmes automatiques	112
4.1	Le diagnostic des systèmes	112
4.2	Classification par enveloppe convexe pour le diagnostic	113
4.2.1	Algorithme DCH : Calcule de l'enveloppe convexe	114
4.2.2	Algorithme DCH : Classification en ligne	115

4.3	Nouvelle approche de classification dynamique par SVDD	116
4.3.1	Classification par SVDD	116
4.3.2	Nouvelle version pour la classification dynamique	117
4.4	Validation	118
5.	Conclusion	119
Chapitre 3 : Classification pour la gestion du risque médical		121
1.	Introduction	122
2.	Gestion des évènements précurseurs ou indésirables	122
2.1	Evènement indésirable	122
2.2	Etude contextuelle	123
2.2.1	Bilan et constat en France	123
2.2.2	Etat de l'art sur la gestion des risques	124
2.3	Vers une solution informatique de traçabilité	124
2.4	Gestion automatique des Evènements Indésirables	125
2.4.1	Le retour d'expertise	125
2.4.2	Une nouvelle approche pour la gestion des EI	126
2.4.3	N-grams pour l'extraction textuelle	128
2.4.4	Construction du vecteur	128
2.4.5	Classification des EI	129
2.4.6	Mesure de similarité	129
2.4.7	Validation expérimentale	129
3.	Gestion automatique des épidémies	130
3.1	Nouvelle approche de gestion automatique des épidémies	131
3.1.1	Un système informatique de partage d'expertise	131
3.1.2	Classification semi supervisée pour la détection des épidémies	132
3.1.3	Nouvelle technique de réduction de dimension	134
3.1.3.1	Analyse Formelle de Concepts (AFC)	134
3.1.3.2	AFC pour la réduction de dimension	134
3.1.4	Classification multi-classe avec dimension dynamique	136
3.2	Validation expérimentale	137
3.2.1	Cas de 5 épidémies	137
3.2.2	Résultats	138
3.2.2.1	Cas d'épidémie connue : Classification supervisée	138
3.2.2.2	Cas d'épidémie inconnue : Classification semi-supervisée	139

4. Conclusion	140
Conclusion et perspectives	141
Bilan	142
Synthèse des travaux présentés dans ce mémoire	142
Projets en cours	143
Perspective de recherche	144
Bibliographie	146

Liste des tableaux

Tableau 1-1: resultats de l'algorithme svm avant la selection.....	66
Tableau 1-2: précision de classification sur la base sonar.....	80
Tableau 1-3: Précision de classification sur la base Pima	80
Tableau 1-4: Précision de classification sur la base Breast Cancer	80
Tableau 1-5: Précision de classification sur la base Ionosphère	81
Tableau 1-6: Précision de classification sur la base Wine	81
Tableau 1-7:Précision de classification sur la base Iris.....	81
Tableau 1-8: Précision de classification sur la base Yale	82
Tableau 1-9: Précision de classification sur la base ORL	82
Tableau 1-10: Taux d'erreur de classification sans extraction	83
Tableau 1-11: Taux d'erreur de classification avec extraction.....	83
Tableau 2-12: Les fonctions noyaux.....	93
Tableau 2-13: Valeurs de R_{tot}^2 et temps de calcul pour les différentes méthodes.....	103
Tableau 2-14: Évolution de R_{tot}^2 en fonction des paramètres m_1, n_1	105
Tableau 2-15: Taux de classification de NS-SVDD comparée aux autres méthodes.....	119
Tableau 3-16: Résultats de classification des évènements indésirables.....	130
Tableau 3-17: Contexte de valeurs.....	135
Tableau 3-18: Contexte binaire.....	135
Tableau 3-19: Résultats de la classification MSVM.....	138
Tableau 3-20: la distance euclidienne entre les nouveaux vecteurs et les différentes classes.....	139
Tableau 3-21: Distance minimale entre les classes.....	139

Liste des figures

Figure 0-1: proportions par spécialité, et par cours, td, tp.....	20
Figure 1-1: Phénomène de sur-apprentissage	57
Figure 1-2: Principe de la sélection de variables	58
Figure 1-3: Principe de l'extraction de caractéristiques	58
Figure 1-4: L'approche filtre pour la sélection des variables	59
Figure 1-5: L'approche Symbioses pour la sélection des variables.....	59
Figure 1-6: L'approche Intégrée pour la sélection des variables	60
Figure 1-7: Principe de la méthode de sélection.....	63
Figure 1-8: Recherche Tabou pour la sélection de variables.	63
Figure 1-9: Organigramme de l'algorithme proposé utilisant la recherche Tabou.....	64
Figure 1-10: Organigramme de l'algorithme proposé utilisant l'algorithme génétique.....	65
Figure 1-11: Cas de convergence vers l'optimum global (1.5%).	67
Figure 1-12: Analyse de tendance et mesure de similarité.....	69
Figure 1-13: Procédure de maximisation de l'IM entre les attributs et leur centre.	75
Figure 2-14: Séparation linéaire.....	91
Figure 2-15: la régression.....	94
Figure 2-16: Division du vecteur observation pour une régression multi-noyaux.....	99
Figure 2-17: Principe de l'identification des systèmes par LS-SKM.....	101
Figure 2-18: Réglage des paramètres de la méthode LS-SKM en se basant sur L'OEP.	102
Figure 2-19: Évolutions de la sortie du système et de la sortie du modèle basées sur la méthode LS-SKM(lin-lin).	102
Figure 2-20: Tracé de dessus : l'évolution du mode réel. Tracé de dessous : l'évolution du mode identifié.	108
Figure 2-21: Structure d'identification en ligne.	109
Figure 2-22: L'évolution du mode réel et de son estimé.	109
Figure 2-23: L'évolution du signal de sortie du système $y(k)$ et la sortie estimée $\hat{y}(k)$	110
Figure 2-24: Classification des données dans l'espace de régression.....	110
Figure 2-25: Procédure d'identification des systèmes linéaires commutant.....	111
Figure 2-26: Evolution de la sortie du modèle proposé $\hat{y}(k)$ et de la sortie du système $y(k)$	112
Figure 2-27: Maintien dynamique de l'enveloppe convexe.....	115
Figure 2-28: Les différentes zones de fonctionnements estimés par le classifieur SVDD.	119
Figure 3-29: Une nouvelle approche pour la gestion des événements précurseurs	127
Figure 3-30: Une collaboration Médicale électronique pour la gestion des épidémies.	132
Figure 3-31: Nouvelle approche de classification semi-supervisée pour la détection automatique des épidémies	133
Figure 3-32: le réseau conceptuel de 5 classes avec 33 attributs.....	136
Figure 3-33: arbre binaire de classification pour 5 epidemies.....	137
Figure 3-34: De l'intelligence à intégrer dans les nouveaux systèmes multi-agents	145

Introduction Générale

Pendant cette dernière décennie les approches d'apprentissage et de classification ainsi que d'autres techniques de l'intelligence artificielle ont donné des résultats pertinents dans différents domaines d'applications. Dans le cadre de mes travaux de thèse de doctorat, j'ai fait appel à une de ces techniques de classification pour la régulation des réseaux de transport multimodal. Une première application à travers laquelle j'ai constaté, la réussite de ces approches avec des taux d'apprentissage relativement intéressants. Néanmoins, il est difficile de confirmer la même réussite pour des systèmes plus complexes ou avec des exigences différentes.

Indépendamment de leurs types, domaines d'applications ou complexité les systèmes actuels sont souvent très liés à la bonne prise de décision. En effet, la régulation et le diagnostic de systèmes automatiques, la reconnaissance de visage, des images ou de caractères est aussi une prise de décision, la gestion des problèmes administratifs et techniques est aussi basée sur la prise de décision. Ainsi, les contributions de chercheurs dans ce contexte sont, pour la plupart, orientées dans ce sens d'aide à la décision. Les approches d'apprentissage et classification automatique semblent être adéquates pour ce type de problème. Il s'agit d'un processus intelligent (faisant usage d'algorithmes) capable d'attribuer les données décrivant un problème, à une ou plusieurs classes faisant référence à une ou plusieurs solutions (décisions).

Pour le premier problème traité, j'ai été confronté à des difficultés liées, entre autres, à la détermination de données pertinentes pour la classification automatique. Ainsi j'ai proposé un premier sujet de recherche dans ce contexte. Il s'agit d'un premier axe dans lequel j'ai dirigé des travaux sur la sélection de variables et l'extraction de caractéristiques. Cet axe de recherche fera l'objet du premier chapitre de la partie mémoire de ce présent document. Mes contributions dans cet axe nous ont été utiles pour les autres problématiques d'aide à la décision dans les domaines médicaux et de surveillance industrielle.

Ce document est organisé en deux parties principales. La première partie représente une synthèse de mes activités de recherche d'enseignement et mes activités industrielles. Elle est constituée d'un bref curriculum vitae retraçant mon état civil mes diplômes ma carrière et les différentes responsabilités que j'ai assumées. Dans cette partie, je présente aussi mes activités d'encadrement, mes productions pédagogiques ainsi que les projets de recherche auxquelles j'ai participé.

La deuxième partie de ce document est un mémoire de recherche. Elle présente la plus grande partie de mes travaux de recherche. Les contributions qui seront présentées portent sur trois principaux axes :

1. Le premier axe concerne l'extraction de caractéristiques et la sélection de variables. Dans ce contexte, je présenterai des solutions de détermination d'attributs pertinents pour un problème de classification.

i- Une première méthode de sélection de variables de type Wrapper sera présentée dans la première partie de ce chapitre. Elle se base sur une exploration de l'espace de combinaisons d'attributs à l'aide de l'approche d'optimisation Tabou et dont le but à atteindre est un taux d'erreur d'apprentissage minimal.

ii- La deuxième approche portera sur l'extraction de caractéristiques. Elle sera présentée dans la deuxième partie de ce chapitre et elle se base sur une classification des attributs selon une analyse des tendances puis ces derniers seront remplacés par le centre de la classe calculé par la mesure Information Mutuelle.

Ces travaux de recherche ont été élaborés, essentiellement, dans le cadre de la thèse de Sabra El Ferchichi [1].

2. Le deuxième axe de recherche porte sur l'application des approches d'apprentissage et classification pour l'identification et le diagnostic des systèmes automatiques complexes. Il fera l'objet du deuxième chapitre de la partie mémoire de recherche. Ce chapitre sera composé de trois parties :

i- La première partie sera consacrée à l'identification des systèmes automatiques non-linéaire. Pour cette problématique, nous avons proposé trois approches basées sur des méthodes multi-noyaux de l'apprentissage statistique. La première approche appelée Multi-Kernel LS-SVM est une variante de la méthode LS-SVM. Elle consiste à appliquer une somme pondérée de plusieurs noyaux. Cette approche a été améliorée par la deuxième méthode appelée LS-SKM qui consiste à une division du vecteur observation en des sous-vecteur, en associant à chacun une fonction noyau. La troisième méthode utilise la même division du vecteur observation avec l'application du principe Multi-noyau mais pour une régression. Ces trois approches seront présentées dans la première partie du chapitre 2 et elles ont fait l'objet du travail de thèse de Mounira Tarhouni [2]

ii- La deuxième partie portera sur l'identification des systèmes commutant. Deux solutions seront présentées dans cette partie, dont la première est pour l'estimation de l'état discret, en ligne et hors-ligne, et la deuxième est une approche de régression pour l'identification des systèmes commutant, basés sur le principe de l'apprentissage dynamique. Ces solutions ont fait l'objet des travaux de thèse de Lamaa Sellami [3].

iii - La troisième partie concernera une autre problématique des systèmes automatiques complexes. Il s'agit du diagnostic de ses systèmes. Un premier algorithme de classification séquentielle sera décrit dans cette partie. Il se base sur le calcul dynamique de l'enveloppe convexe. Appelé DCH (Dynamic Convex Hall), il permet de détecter les classes de fonctionnement normal et celles de la panne. La deuxième approche, proposée dans ce contexte de diagnostic, est une méthode de classification dynamique par SVDD. Adaptée aux données stationnaires, cette méthode sera présentée en deux versions dont la première est mono-classe et la deuxième est Multi-classe. Cette partie a fait l'objet des travaux de thèse de Foued Theljani [4].

3. Le troisième axe de recherche porte sur la gestion des risques médicaux. En se basant sur le retour d'expertise, deux problématiques seront traitées dans le troisième chapitre de la partie mémoire, dans deux parties différentes.

i- La première partie présentera une nouvelle approche de gestion des risques liés aux événements indésirables ou précurseurs dans les établissements de soins. Cette méthode utilise l'approche n-grams pour l'extraction textuelle et une méthode de classification pour automatiser le partage d'expertise et résoudre les problèmes liés à ces événements.

ii- La deuxième partie portera sur la gestion automatique des risques liés aux épidémies. Dans les mêmes contextes médicaux et d'apprentissage et classification, je présenterai une nouvelle approche de classification semi-supervisée pour la détection et la prédiction des épidémies à partir de données partagées entre plusieurs centres et traitées par l'approche d'Analyse Formelle de Concepts (AFC).

Les trois chapitres présenteront aussi la validation expérimentale de ces approches.

On finira par une conclusion sur ces contributions et les résultats atteints. Des perspectives de ces contributions seront aussi présentées et ouvriront les portes pour de futurs projets de recherche et innovation.

Première partie :

Rapport de Synthèse

1. Notice individuelle

1.1. État Civil

Nom : **ZIDI**

Prénom : **Salah**

Date et lieu de naissance : Le 3 juin 1980 à Gafsa (en Tunisie)

Situation Familiale : Marié, 2 Enfants

Adresse Professionnelle : Université AlQassim, College of Business and Economics (CBE), Buraydah 52571, Saudi Arabia

Téléphone professionnelle : +(966) (6) 3800050 (poste : 3017)

E-Mail : s.zidi@qu.edu.sa

Page Google Scholar <https://scholar.google.fr/citations?user=Ee-9F1UAAAAJ&hl=fr>

1.2. Formation Universitaire

2007 **Doctorat**, spécialité Automatique et Informatique industrielle, Laboratoire d'Automatique, Génie Informatique et Signal (LAGIS), Université des Sciences et Technologies de Lille.

Titre : SARR : Système d'Aide à la Régulation et la Reconfiguration des réseaux de transport multimodal.

Cadre : Effectuée à l'Université des Sciences et Technologies de Lille, la thèse se situe dans le projet : « Système d'aide à la régulation du réseau de transport valenciennois et de ses pôles d'échanges ». Convention de Recherche GRRT

Date de la soutenance : le 04 juillet 2007

Membres du jury : Alexandre DOLGUI et Mohamed EL BAGDOURI (Rapporteurs), Mireille BAYART, Aziz MOKRIM, Abdelhakim ARTIBA et Christian TAHON (Examineurs), Salah MAOUCHE (Directeur) et Slim HAMMADI (Co-Directeur).

Mention : Très Honorable.

Qualification : - en mars 2008 aux fonctions de maître de conférences en 61^{ème} section du CNU (Génie informatique, automatique et traitement du signal) et 27^{ème} section du CNU (Informatique).

2004 **Mastère de recherche en Nouvelles Technologies des Systèmes Informatiques dédiés**, Ecole Nationale d'Ingénieur de Sfax (ENIS), Tunisie.

Stage effectué au G.M.S (Groupe en micro-technologie et système sur puce), Ecole Nationale d'Ingénieur de Sfax

sous la direction de Mohamed MASMOUDI.

Projet de recherche : Étude des systèmes d'exploitation dédiées aux cartes à puce : Étude de faisabilité : développement d'une partie logicielle dédiée aux cartes à puce à base du système ouvert NESOS et implantation sur une plateforme matérielle LEON.

2003 **Ingénieur en génie électrique**, Ecole Nationale d'Ingénieur de Sfax (ENIS), Tunisie

PFE : Réalisation d'une borne de réapprovisionnement des porte-monnaie virtuels intégrant 3 modes : lecteur des pièces, lecteur des billets et lecteur des cartes à puces.

Prime ministériel en Tunisie du meilleur Projet universitaire

Stage effectué au sein de la société ARCHIMED, Lille, France.

2000 Cycle préparatoire aux études d'ingénieurs, Institut Préparatoire aux études d'ingénieurs de Nabeul (IPEIN), spécialité Mathématiques-Physiques (MP).

1998 Baccalauréat, spécialité mathématiques, Lycée Houcine Bouzaiene à Gafsa (Tunisie).

1.3. Emplois Occupés

Depuis Aout 2014 Assistant professor, Qassim University, Saudi Arabia, College of Business and Economics CBE, Management of Information Systems (MIS) Department.

Depuis juillet 2014 Auto-entrepreneur, Consultant Recherche et innovation en informatique, Développement logiciel GED et Archivage à valeur probante, Accompagnement pour les dossiers CIR CII.

Janv.2010 – Dec.2013 Docteur Ingénieur Recherche et Innovation, Groupe Archimed, Lille, France

Sept.2009 – Dec.2010 Consultant Recherche et Innovation, Société ARIZONA Consulting, Lille, France

Juin.2009 – Sept.2009 Responsable R&D, Société BIOMEDIQA, Lille, France

Sep.2007 – Aout.2009 ATER à plein temps Université des Sciences et Technologies de Lille, UFR d'IEEA (Informatique, Électronique, Électrotechnique, Automatique).

2. Activités pédagogiques

J'ai assuré des activités d'enseignement depuis l'année universitaire 2003/2004, durant mon stage de mastère de recherche. Pendant une année, j'ai eu l'occasion d'apprendre et développer ma pédagogie d'enseignement en tant que vacataire. J'ai continué à assurer aussi des activités d'enseignement pendant ma thèse en tant que vacataire, deux ans après ma soutenance en tant qu'ATER mais aussi trois ans en plus en tant que vacataire pendant mes activités et responsabilité industrielles. Pendant cette période allant de 2010 et 2013 j'ai eu l'occasion d'enrichir mon expérience en termes de compétences pédagogiques à travers des activités d'enseignement en tant que vacataire et dans le cadre d'un contrat de collaboration entre le groupe Archimed et l'IUT A de Lille.

Depuis Août 2014, recruté en tant qu' "assistant professor" (équivalent Maître de conférences ou encore maître assistant) à l'université Al Qassim en Arabie Saoudite, j'ai eu à assurer des modules en informatique et systèmes d'information.

J'ai acquis une expérience de 10 ans d'enseignement. Dans cette partie de ce présent document, je détaille les heures académiques de cours, travaux dirigés et travaux pratiques, ainsi que l'activité d'encadrement de projets déconnectée de mon activité de recherche. Je présente aussi les établissements et les formations qui ont fait l'objet de cette expérience.

2.1. Bilan général

MATIERE	VOLUME HORAIRE	ETABLISSEMENT	TYPE
programmation orientée objet : java	180	QASSIM UNIVERSITY	COURS
DSS	180	QASSIM UNIVERSITY	COURS
Systèmes d'information	90	QASSIM UNIVERSITY	COURS
Nouvelles technologies	90	QASSIM UNIVERSITY	COURS
Sécurité informatique	45	QASSIM UNIVERSITY	COURS
Structure de données et Gestion de fichiers	45	QASSIM UNIVERSITY	COURS
XML/XSLT	30C + 8TD/TP	IUT-A LILLE	COURS + TD/TP
VBA	8C+ 9TD+ 32TP	UST LILLE	COURS + TD+TP
Régulation industrielle	18TD + 16 TP	UST LILLE	TD+TP
Outil de simulation	16	UST LILLE	TP
Supervision	16	UST LILLE	TP
Logique	16	UST LILLE	TP
Processeur de traitement	22TD + 16 TP	UST LILLE	TD + TP
Informatique industrielle	22TD	UST LILLE	TD
Outils pour l'Informatique Industrielle	22	UST LILLE	TD
Outil de modélisation (programmation C)	12	UST LILLE	TP
BD	32	UST LILLE	TP
HTML	9TD+ 32TP	UST LILLE	TD + TP
Architecture des ordinateurs	15TD + 18TP	TELECOM LILLE	TD + TP

Electronique	60	ENIS	TP
Electronique Numérique	15	ENIS	TP
Circuit Logique	12	FLS	TP
Automate programmable	12	FLS	TP
Technologies	12	FLS	TP
Electronique Numérique	15	FLS	COURS

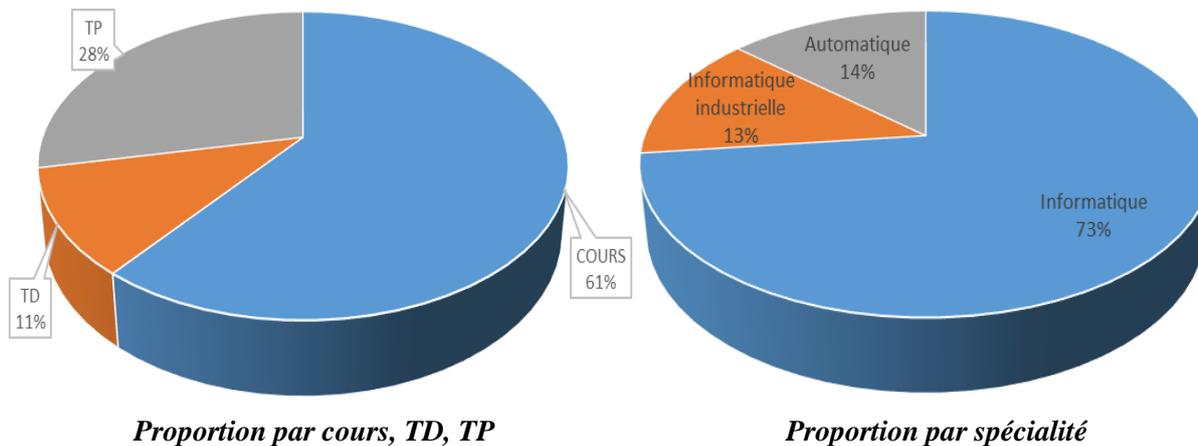


FIGURE 0-1: PROPORTIONS PAR SPECIALITE, ET PAR COURS, TD, TP.

2.2. Matières enseignées par établissement

Au cours de ces 10 dernières années, j'ai eu l'opportunité d'enseigner dans 6 établissements différents dans trois pays (la Tunisie, la France et l'Arabie saoudite). Pour des niveaux allant de Bac + 2 à Bac + 5, j'ai enseigné en 3 langues différentes (français pour la plupart des matières, arabe et anglais).

Depuis l'année universitaire 2003/2004, mes activités d'enseignement (Cours, Travaux Dirigés et Travaux Pratiques) se situent essentiellement dans les domaines de l'informatique générale, l'informatique industrielle et l'automatique.

2.2.1. Université d'Al Qassim (Arabie Saoudite)

Assistant professor ; depuis 2014 ; 270h de cours/an

Depuis Aout 2014, j'ai signé un contrat en poste "Assistant professor" (équivalent maître de conférences) à l'université Al Qassim au centre nord de l'Arabie saoudite (à 350 Km de la capitale Ryadh). J'ai intégré le département « Gestion des Systèmes d'Information » (MIS : Management of Information Systems) de la faculté d'économie et gestion (CBE : College of Business and Economics). Au sein de cet établissement, le défi était important puisque l'établissement a eu l'accréditation internationale AACSB (audit en 2016). J'ai eu l'opportunité de participer à cette réussite en tant que **président d'un comité AOL (Assurance Of Learning)**. Très centré sur la pédagogie pour la partie Assurance of Learning (AoL), le processus AACSB est intéressant car il vise à s'assurer de la qualité des apprentissages et c'est un bon outil pour mesurer les sorties du processus universitaire. Dans cette même logique d'amélioration et d'évaluation de nos efforts pédagogiques, j'ai monté plusieurs modules. J'ai enseigné essentiellement des matières en relation avec **l'informatique générale et les systèmes d'information**. Avec un volume horaire de **270 heures par an (135 heures**

par semestre), j'ai dispensé des cours pour différents niveaux License (Bac+2, Bac+3 et Bac+4) et j'ai participé au montage des TD et TP en tant que responsable des modules.

➤ ***Programmation Orientée objet : Langage JAVA avancé***

Ce cours de programmation Java permet d'initier les étudiants à la programmation orientée objet. Il leur permet de comprendre et d'appliquer les principales notions de ce concept, à savoir l'abstraction, l'encapsulation, l'héritage, le polymorphisme ...

Bien que ce cours ne constitue pas une introduction à la programmation java, il commence par un bref rappel sur les bases de ce langage avant de présenter des détails plus évolués.

Selon le processus AOL, les objectifs pédagogiques (learning goals) de ce module sont :

- Connaissances théoriques : assurés par le cours et mesuré par les examens théoriques
- Connaissances pratiques et techniques : assurés par les TD/TP et mesurés par les examens pratiques
- Réflexion critique : assurés par le cours et les projets et mesurés par les examens et les projets de développement

ÉTABLISSEMENT	CBE, Qassim University KSA
DEPARTEMENT	MIS
TYPE DE CONTRAT	Assistant Professor
CURSUS	3 ^{ème} année License (semestre 5)
REPARTITION	45 heures cours/semestre (1,5h x 2 séances/semaine) pendant 4 semestres
DESRIPTIF	Programmation JAVA niveau 2

➤ ***Systèmes d'aide à la décision et systèmes experts***

Ce cours présente les concepts de base des systèmes d'aide à la décision (DSS) et leurs applications dans les organisations pour appuyer la prise de décision. Il décrit aussi comment les organisations peuvent exploiter des systèmes informatisés dans le processus décisionnel. Ce cours porte sur les caractéristiques, les composants, les logiciels et les modèles utilisés pour développer ces systèmes. En outre, le cours aborde les entrepôts de données, l'exploration de données, le business intelligence et systèmes experts. Quelques algorithmes de l'intelligence artificielle sont expliqués dans ce cours à savoir l'Algorithme génétique, les réseaux de neurone et la logique flou. Ils sont présentés dans le contexte de leur usage pour la prise de décision.

Selon le processus AOL, les objectifs pédagogiques (learning goals) de ce module sont :

- Connaissances théoriques : assurés par le cours et mesurés par les examens théoriques
- Connaissances pratiques et techniques : assurés par les TP et mesurés par les examens pratiques
- Réflexion critique : assurés par le cours et les projets et mesurés par les examens et les projets

ÉTABLISSEMENT	CBE, Qassim University, KSA
DEPARTEMENT	MIS
TYPE DE CONTRAT	Assistant Professor
CURSUS	4 ^{ème} année License (semestre 8)
REPARTITION	45 heures cours/semestre (1,5h x 2 séances/semaine) pendant 4 semestres
DESRIPTIF	Les systèmes d'aide à la décision et les systèmes expert.

➤ *Systemes d'information*

Sans entrer beaucoup dans les détails techniques, ce cours explique le triangle stratégique et l'importance des nouveaux systèmes d'information. Il présente aussi l'architecture de ces systèmes et les différents composants. Les derniers chapitres portent sur les sources externes et internes, la gestion des projets de systèmes d'information et la gestion des connaissances.

Selon le processus AOL, les objectifs pédagogiques (learning goals) de ce module sont :

- Connaissances théoriques : assurés par le cours et mesurés par les examens théoriques
- Connaissances pratiques et techniques : assurés par les TP et mesurés par les examens pratiques
- Communication orale et écrite : assurés et mesuré par les présentations et les rapports
- Responsabilité et travail en équipe : assurés et mesurés par les projets collectifs.

ÉTABLISSEMENT	CBE, Qassim University, KSA
DEPARTEMENT	MIS
TYPE DE CONTRAT	Assistant Professor
CURSUS	3 ^{ème} année License (semestre 6)
REPARTITION	45 heures cours/semestre (1,5h x 2 séances/semaine) pendant 2 semestres 2015
DESRIPTIF	Gestion des systèmes d'information.

➤ *Séminaire en gestion des systèmes d'information : sécurité informatique*

Ce cours porte sur des sujets spécifiques liés à la spécialité. Il traite le problème de la sécurité informatique de point de vue juridique, éthique et professionnelle. Il présente aussi les dernières technologies liées à ce sujet.

- Introduction à la sécurité de l'information
- Le besoin de sécurité
- Questions juridiques, éthiques et professionnels en sécurité de l'information
- Gestion des risques
- Planification de la sécurité
- Technologie de sécurité : détection d'intrusion et les systèmes de prévention, et autres outils de sécurité
- Cryptographie
- Mise en œuvre de la sécurité de l'information

Selon le processus AOL, les objectifs pédagogiques (learning goals) de ce module sont :

- Connaissances théoriques : assurés par le cours et mesurés par les examens théoriques
- Enjeux Éthique, juridique et professionnels : assurés et mesurés par les projets réalisés par les étudiants et proposés et encadrés par l'enseignant.
- Communication orale et écrite : assurés et mesurés par les présentations et les rapports individuels.
- Responsabilité et travail en équipe : assurés et mesurés par les projets collectifs.
- Responsabilité et travail en équipe : assurés et mesurés par les projets collectifs.

ÉTABLISSEMENT	CBE, Qassim University, KSA
DEPARTEMENT	MIS
TYPE DE CONTRAT	Assistant Professor
CURSUS	4 ^{ème} année License (semestre 7)
REPARTITION	45 heures cours/semestre (1,5h x 2 séances/semaine) pendant 1 semestre 2015
DESRIPTIF	Sécurité d'information.

➤ *Structure de données et Gestion de fichiers*

Ce cours aide les étudiants à comprendre les structures des données et des fichiers. Il leur permet d'apprendre la partie programmation qui concerne la gestion de mémoire et la gestion de fichiers en JAVA (création, lecture, écriture, FIFO, LIFO, LinkedList, ArrayList, arbre,...). Il explique aussi le stockage secondaire et l'indexation multi-niveau.

Selon le processus AOL, les objectifs pédagogiques (learning goals) de ce module sont :

- Connaissances théoriques : assurés par le cours et mesurés par les examens théoriques
- Connaissances pratiques et techniques : assurés par les TP et mesurés par les examens pratiques

ÉTABLISSEMENT	CBE, Qassim University, KSA
DEPARTEMENT	MIS
TYPE DE CONTRAT	Assistant Professor
CURSUS	4 ^{ème} année License (semestre 8)
REPARTITION	45 heures cours/semestre (1,5h x 2 séances/semaine) pendant 1 semestre 2016
DESRIPTIF	Structures de données et gestion de fichiers.

➤ *Sujets spéciaux en système d'information : Nouvelles technologies*

Avec des collègues du département, Nous avons monté ce module en 2015. J'ai participé à 50% (4 chapitres = 7 semaines d'enseignement). Ce module porte sur des nouvelles technologies dans le domaine TIC. En 2015 et 2016, nous avons enseigné les sujets suivants :

- Data Mining
- Big Data
- Cloud Computing
- Internet of thinks
- E-services
- E-government

Selon le processus AOL, les objectifs pédagogiques (learning goals) de ce module sont :

- Connaissances théoriques : assurés par le cours et mesurés par les examens théoriques
- Connaissances pratiques et techniques : assurés par les TD/TP et mesurés par les examens pratiques
- Réflexion critique : assurés par le cours et les projets et mesurés par les examens et les projets de développement

ÉTABLISSEMENT	CBE, Qassim University, KSA
DEPARTEMENT	MIS
TYPE DE CONTRAT	Assistant Professor
CURSUS	4 ^{ème} année License (semestre 8)
REPARTITION	45 heures cours/semestre (1,5h x 2 séances/semaine) pendant 2 semestres
DESRIPTIF	Cloud computing, Internet of thinks, e-Gov, e-services.

2.2.2. Institut Universitaire de Technologie (IUT A) Lille (France)

Vacataire 2010 – 2012 ; 15h/an cours+4hTD/TP

➤ *XML/XSLT*

Dans le cadre d'un contrat entre la société Archimed et l'IUT A de Lille1, j'ai monté un module pour la formation Licence Professionnelle Systèmes Informatiques et Logiciels option DA2I (*Développement et Administration de sites Internet et Intranet*). Dans cette formation, une large part des interventions est réservée aux entreprises. J'ai participé par un module sur le métalangage de balisage **XML** et ses transformations **XSLT** avec un volume horaire de **15heures/an de cours avec 4 heures de TP**.

ÉTABLISSEMENT	IUT A, Lille, France
DEPARTEMENT	MIS
TYPE DE CONTRAT	Contrat entre société Archimed Group et l'IUT
CURSUS	3 ^{ème} année License (DA2I)
REPARTITION	15 heures cours/an (4h x 3 séances+3h) +4h TP pendant 2 ans
DESRIPTIF	XML/XSLT.

2.2.3. Université Lille1 (France)

2 ans ATER (100% poste complet) 2007 – 2009 : 192h/an équivalent TD

J'ai passé deux ans en tant qu'Attaché temporaire d'enseignement et de recherche (**ATER**) à l'UFR d'IEEA (Informatique, Électronique, Électrotechnique, Automatique) de l'Université des Sciences et Technologies de Lille (USTL). En étant en poste à 100% pendant deux ans, j'ai effectué une charge totale de **192 heures équivalentes TD** en automatique, informatique générale et informatique industrielle pour des étudiants en licence Sciences et Technologies, Mention EEA option Ingénierie Électrique (IE) et des étudiants en master Automatique et Systèmes Electriques (ASE) et master Mesures Instrumentation Procédés (MIP).

Automatique

Ma formation d'ingénieur en génie électrique et mes travaux de recherche en identification, régulation et diagnostique des systèmes automatiques m'ont apporté les connaissances nécessaires pour enseigner différentes thématiques relevant de l'automatique.

➤ *Régulation industrielle*

J'ai enseigné 9 séances de 2 heures de TD et 4 séances de 4 heures de TP pour des étudiants de licence I.E. Dans ce module les thèmes suivants sont traités :

- Asservissement de vitesse ;
- Asservissement de position ;
- Régulation thermique.

ÉTABLISSEMENT	USTL
DEPARTEMENT	UFR IEEA
TYPE DE CONTRAT	ATER
CURSUS	3 ^{ème} Licence ST Mention EEA Semestre 6 Parcours IE
REPARTITION	18h TD+16 heures TP 4 x (4h /semaine)
DESRIPTIF	Asservissement de vitesse ; asservissement de position ; régulation thermique.

➤ *Outil de simulation*

L'objectif de ces manipulations est de maîtriser les principales fonctionnalités du plus représentatif des logiciels de calcul scientifique : le logiciel de calcul Matlab et sa Toolbox graphique Simulink.

ÉTABLISSEMENT	USTL
DEPARTEMENT	UFR IEEA
TYPE DE CONTRAT	ATER
CURSUS	Master ASE M1 (Bac +4)
REPARTITION	16 heures TP 4x(4h /semaine) X 2 groupes
DESRIPTIF	Etude et simulation de systèmes automatiques avec MATLAB.

Informatique industrielle

➤ *Supervision*

Le module traite, entre autre, de la mise en œuvre de diverses solutions structurées pour la réalisation de Grafcet et la commande d'un processus logique par ordinateur. Dans ce module j'ai utilisé le noyau temps réel MTR86w32, pour écrire des applications système temps réel multitâche. J'ai proposé aussi des réalisations sur Intouch dans ces manipulations.

ÉTABLISSEMENT	USTL
DEPARTEMENT	UFR IEEA
TYPE DE CONTRAT	ATER
CURSUS	Master ASE M1 + Master MIP M1 (Bac +4)
REPARTITION	16 heures TP 4x(4h /semaine) X 2 groupes
DESRIPTIF	Etude et simulation de systèmes automatiques avec MATLAB.

➤ *Logique*

Des TP réalisés sur la maquette logique M13. Dans ce module les thèmes suivants sont traités:

- Introduction à la logique booléenne
- Circuits intégrés de la logique combinatoire
- Circuits arithmétiques simples : addition et soustraction de nombres entiers positifs
- Eléments de logique séquentielle
- Circuits synchrones

ÉTABLISSEMENT	USTL
DEPARTEMENT	UFR IEEA
TYPE DE CONTRAT	ATER (poste complet : 192 heures) cour, TD et TP
CURSUS	1 ^{ière} Licence ST Mention EEA Semestre 1
REPARTITION	16 heures TP 4 X (4h /semaine)
DESRIPTIF	Simulation sur la maquette logique M13

➤ *Processeur de traitement*

Dans ce module j'ai assuré des TD (22h) et des TP (16h). Les thèmes suivants sont traités, dans ce module :

- Organisation interne : UAL, séquenceur, registres
- Organisation externe : bus, chronogrammes

- Langage machine et assembleur
- Les interruptions

ÉTABLISSEMENT USTL
 DEPARTEMENT UFR IEEA
 TYPE DE CONTRAT ATER
 CURSUS Master ASE M1 (Bac +4)
 REPARTITION 22 heures TD (2h /semaine) x 11 semaines + 16h TP
 DESCRIPTIF Asservissement de vitesse ; asservissement de position ; régulation thermique.

➤ ***Informatique industrielle***

J'ai enseigné le module informatique industrielle pour deux niveaux différents (master M1 et licence). Pour les formations master MIP et ASE le module porte sur les **Automates programmables industriels** avec une simulation et une implémentation utilisant (**TSX+Automgen**). Pour le niveau licence, le module porte sur la logique séquentielle asynchrone et sur l'étude de systèmes séquentiels et combinatoires.

ÉTABLISSEMENT USTL
 DEPARTEMENT UFR IEEA
 TYPE DE CONTRAT ATER
 CURSUS Master ASE et MIP M1 (Bac +4) + 3^{ème} Licence Mention EEA option I.E Semestre 5
 REPARTITION 22 heures TD (2h /semaine) x 11 semaines
 DESCRIPTIF Asservissement de vitesse ; asservissement de position ; régulation thermique.

➤ ***Outils pour l'Informatique Industrielle***

Ce module est composé de deux parties. Une première sur le grafcet et la seconde partie de la programmation C. J'ai enseigné dans ce module des TD et des TP.

ÉTABLISSEMENT USTL
 DEPARTEMENT UFR IEEA
 TYPE DE CONTRAT ATER
 CURSUS Master ASE M1 (Bac +4)
 REPARTITION 22 heures TD (2h /semaine) x 11 semaines
 DESCRIPTIF Grafcet + Programmation C.

➤ ***Outil de modélisation pour l'E.E.A ; TD/TP***

Ce module porte sur la programmation C.

ÉTABLISSEMENT USTL
 DEPARTEMENT UFR IEEA
 TYPE DE CONTRAT ATER
 CURSUS 2^{ème} Licence ST Mention EEA Semestre 4 option IE
 REPARTITION 12 heures TP 3 x (4h /semaine) x 3 groupes
 DESCRIPTIF Programmation C

Informatique Générale

➤ **VBA ; Cours + TD et TP**

Ce module porte sur la programmation Visual Basic pour Applications. J'ai enseigné le module complet (Cours + Td et TP) et j'ai participé au montage des TP.

ÉTABLISSEMENT USTL
DEPARTEMENT UFR IEEA
TYPE DE CONTRAT ATER
COURSUS Master MIP M1 (Bac +4)
REPARTITION 8h cours + 9h TD+ 32 heures TP 4 x (4h /semaine) x 2 groupes
DESRIPTIF Visual Basic pour Applications

➤ **Base de données**

4 TP dont les manipulations de la base de données se font sur Acces.

ÉTABLISSEMENT USTL
DEPARTEMENT UFR IEEA
TYPE DE CONTRAT ATER
COURSUS Master MIP M1 (Bac +4)
REPARTITION 32 heures TP 4 x (4h /semaine) x 2 groupes
DESRIPTIF Base de données Acces

➤ **HTML**

Une initiation à la programmation web de base. De la programmation HTML

ÉTABLISSEMENT USTL
DEPARTEMENT UFR IEEA
TYPE DE CONTRAT ATER
COURSUS Master MIP M1 (Bac +4)
REPARTITION 9h TD+ 32hTP 4 x (4h /semaine) x 2 groupes
DESRIPTIF HTML

2.2.4. Telecom Lille (France)

2005 – 2006 : vacataire : 15hTD +18h TP

Durant ma thèse, j'ai assuré des heures d'enseignements à l'ENIC - Télécom Lille 1 en tant que Vacataire. Ce module porte sur l'architecture des ordinateurs avec un volume horaire total de 33 heures sur les deux ans (7.5h TD/an + 9h TP/an) pour des étudiants en 3^{ème} année école d'ingénieur.

➤ **Architecture des ordinateurs**

Les TD et TP de ce module porte sur la programmation Assembleur sur le microprocesseur INTEL 8086.

ÉTABLISSEMENT PENIC -TELECOM LILLE 1
DEPARTEMENT Informatique et réseaux
TYPE DE CONTRAT Vacations

CURSUS	3 ^{ème} année école d'ingénieur
REPARTITION	7,5 heures TD (1,5h x 5 séances) +9 heures TP (1,5h x 6 séances)
DESCRIPTIF	microprocesseur INTEL 8086

2.2.5. Ecole Nationale d'Ingénieur de Sfax-ENIS (Tunisie)

Vacataire 2003-2004 ; 75h TP

Pendant mon stage de recherche en mastère, j'ai eu l'opportunité d'enseigner 2 TP en électronique à l'ENIS en tant que vacataire. Ma formation d'ingénieur en génie électrique dans la même école et au sein du même département m'a beaucoup aidé pour réussir cette première expérience pédagogique.

➤ *Electronique 1*

ÉTABLISSEMENT	Ecole Nationale d'Ingénieur de Sfax (E. N. I. S) - Tunisie
DEPARTEMENT	Génie Électrique
TYPE DE CONTRAT	Vacations
CURSUS	2 ^{ème} année école d'ingénieurs
REPARTITION	60 heures TP ((4h x 3 séances+3h x1 séance) x 4 groupes)
DESCRIPTIF	Réalisation de mini-projets électroniques.

➤ *Electronique Numérique*

ÉTABLISSEMENT	Ecole Nationale d'Ingénieur de Sfax (E. N. I. S) - Tunisie
DEPARTEMENT	Génie Électrique
TYPE DE CONTRAT	Vacations
CURSUS	1 ^{ère} année école d'ingénieur
REPARTITION	15 heures TP ((4h x 3 séances+3h x1 séance) x 1 groupe)
DESCRIPTIF	Montage sur maquette, tracés de caractéristiques, détermination de seuil, Gain...

2.2.6. Université Libre de Sfax (Tunisie)

Vacataire 2003-2004 : 15h cours + 7.5h TD + 36h TP

Pendant l'année universitaire 2003/2004 et en même période que l'expérience de l'ENIS, j'ai eu l'occasion d'enseigner un module de Cours et un autre en TD et 2 TP à l'ULS. J'ai beaucoup appris de cette expérience d'un volume horaire total de 60 heures.

➤ *Circuit Logique*

ÉTABLISSEMENT	Université Libre de Sfax (U. L. S) - Tunisie
DEPARTEMENT	Génie Électrique
TYPE DE CONTRAT	Vacations
CURSUS	1 ^{ère} année génie électrique
REPARTITION	12 heures TP (3h x 4 séances x 1 groupe)
DESCRIPTIF	Calcul théorique et vérification pratique sur simulateur logique

➤ *Automate programmable*

ÉTABLISSEMENT	Université Libre de Sfax (U. L .S) - Tunisie
DEPARTEMENT	Génie Électrique
TYPE DE CONTRAT	Vacations
CURSUS	1 ^{ière} année génie électrique
REPARTITION	12 heures TP ((3h x 4 séances) x 1 groupe)
DESCRIPTIF	Utilisation de Step7

➤ *Technologie des circuits imprimés*

ÉTABLISSEMENT	Université Libre de Sfax (U. L .S) - Tunisie
DEPARTEMENT	Génie Électrique
TYPE DE CONTRAT	Vacations
CURSUS	1 ^{ière} année génie électrique
REPARTITION	12 heures TP (3h x 4 séances x 1 groupe)
DESCRIPTIF	Technologie des circuits imprimés (les étapes de réalisation), réalisation d'un circuit

➤ *Electronique Numérique (Cours)*

ÉTABLISSEMENT	Université Libre de Sfax (U. L .S) - Tunisie
DEPARTEMENT	Génie Électrique
TYPE DE CONTRAT	Vacations
CURSUS	1 ^{ère} année génie électrique
REPARTITION	15 heures cours ((3h x 5 séances) x 1 groupe)
DESCRIPTIF	Historique, les niveaux de tensions logiques, les portes logiques (caractéristiques électriques), les bascules, ...

2.3. Productions Pédagogiques

Depuis ma première expérience pédagogique en 2004 et jusqu'à ces dernières années universitaires à l'université Al Qassim, j'ai participé avec mes collègues à la rédaction de nombreux documents pédagogiques destinés aux étudiants de différents niveaux. Dans ce cadre universitaire, j'ai eu l'opportunité de participer à la préparation des polycopiés de cours, des exercices de TD, des sujets de travaux pratiques et des examens et j'ai aussi contribué à la mise à jour des différents supports pédagogiques.

Durant mes expériences professionnelles et surtout au sein du groupe Archimed, j'ai rédigé des documents et supports pédagogiques pour la formation des ingénieurs et des responsables commerciaux.

La liste suivante, non exhaustive, comporte la grande partie des supports que j'ai produits ou participé à leur élaboration :

Productions Universitaires

- Support de cours sous forme de diaporama pour le module programmation orienté objet (langage JAVA niveau avancé) destiné aux étudiants en 3^{ième} année licence (semestre 5) au département MIS College of Business and Economics de l'université Al Qassim.

- Polycopié d'exercices de programmation JAVA, destiné aux étudiants en 3^{ième} année licence (semestre 5) au département MIS College of Business and Economics de l'université Al Qassim.
- Support de cours sous forme de diaporama pour le module systèmes d'aide à la décision et systèmes experts destiné aux étudiants en 4^{ième} année licence (semestre 8) au département MIS College of Business and Economics de l'université Al Qassim.
- Support de cours sous forme de diaporama pour le module systèmes d'information, destiné aux étudiants en 3^{ième} année licence (semestre 6) au département MIS College of Business and Economics de l'université Al Qassim. Pour ce cours j'ai participé par 2 chapitres sur 8.
- Support de cours sous forme de diaporama pour le module Sujets spéciaux en système d'information, destiné aux étudiants en 3^{ième} année licence (semestre 6) au département MIS College of Business and Economics de l'université Al Qassim. Pour ce cours j'ai participé par 4 chapitres sur 6 dont les sujets sont Cloud computing ; internet of things, e-services et e-gov.
- Support de cours et de travaux pratiques pour le module XML/XSLT, destiné aux étudiants en 3^{ième} année licence Professionnelle Systèmes Informatiques et Logiciels option DA2I à l'IUT A de Lille
- Mise à jour du support de cours pour le module programmation VBA, destiné aux étudiants en Master MIP M1 à l'UFR IEEA à l'USTLille.
- Support de cours pour le module technologies au département génie électrique à l'ENIS, destiné aux étudiants en 1^{ière} année école d'ingénieur.
- Sujets de travaux pratiques et de travaux dirigés dans différentes disciplines : programmation, supervision, nouvelles technologies, XML/XSLT ...,

Productions pédagogique dans un cadre industriel

- Support pédagogique sur la Signature électronique : document d'étude + Support de cours sous forme de diaporama, destiné aux ingénieurs et aux commerciaux d'Archimed Group.
- Support pédagogique sur le Protocole Oauth : document + Support de cours sous forme de diaporama + démonstration pratique, destiné aux ingénieurs et aux commerciaux d'Archimed Group.
- Support pédagogique sur la gestion électronique de document et l'archivage à valeur probante pour le Centre National d'Informatique CNI, le résultat d'un séminaire professionnel que j'ai organisé en 2010.

2.4. Activités d'intérêt collectif

- Responsable pédagogique de plusieurs modules à l'université Al Qassim College of Business and Economics:
 - o Module de "Programmation Orienté objet"
 - o Module de "systèmes d'aide à la décision et systèmes experts"
 - o Module de "Gestion de fichiers"
- Responsable de comité AOL de l'objectif pédagogique "Communication oral" (Oral communication) à l'université Al Qassim College of Business and Economics
- Membre de comité AOL de l'objectif pédagogique "Connaissance" (knowledge) à l'université Al Qassim College of Business and Economics

2.5. Encadrements de Projets et suivi de stages

Depuis 2007, j'ai encadré des projets et des stages dans le cadre de mes activités d'ATER à l'UFR IEEA à l'université de Lille1. Ces projets sont destinés aux étudiants de master M1 et M2 pour les deux formations MIP et ASE SII.

Par ailleurs, j'ai encadré un projet de fin d'étude pour un étudiant en *cycle d'ingénieur en génie électrique de l'ENIS*. Ce projet est dans le cadre d'une collaboration entre le LAGIS de l'UST Lille et l'ENIS en Tunisie.

1. Fahmi AYADI, *PFE cycle d'ingénieur en génie électrique de l'ENIS*, titre : Outil d'apprentissage pour un système d'aide à la régulation des réseaux de transport, février 2007 - juillet 2007, Enc. 100%
2. Ali tej eddine Ben Hammouda, Olivier Raguette, Omar Skalante, *Projet de master MIP à l'UFR IEEA de L'USTL*, Titre : Réalisation d'un simulateur d'exploitation d'un réseau de transport, janvier 2008 – mai 2008, Enc. 100%.
3. Yohan Quertainmont et Jérémy Beuscart, *Projet de master ASE professionnel option Système d'Information pour l'Industrie (SII) à l'UFR IEEA de L'USTL*, Titre : Développement d'un outil de simulation de convergence d'un algorithme d'optimisation, janvier 2009 – juin 2009, Enc. 100%.
4. Jonathan Degroote, suivie d'un Stage professionnel master MIP, stage réalisé à SNCF, Titre : développement d'un outil informatique pour assurer le suivi du flux des pièces peinture, dans le cadre du projet de rénovation des rames Thalys, Co-encadrement entre *l'UFR IEEA de L'USTL et SNCF*, Tuteur Enc. 50%.

2.6. Formations pédagogique Complémentaires

JANVIER 2016	Formation avec certificat sur la plateforme de e-learning Blackboard organisée par l'université d'Al Qassim
NOV 2015	Formation avec certificat sur la préparation des fiches descriptives et des rapports de modules pédagogiques.
DEC 2003	Cycle de formation pédagogique, organisée dans le cadre des formations de master à l'E.N.I.S.
AVRIL 2003	Université de printemps de pédagogie, organisée dans le cadre des formations de master à l'E.N.I.S.

3. Activités et responsabilités administratives

- ✓ Membre de la Comité d'excellence en enseignement (Teaching Excellence committee), au sein de la CBE (College of Business and Economics), 2017.
- ✓ Responsable de l'équipe AOL (Assurance Of Learning) de l'objectif pédagogique « Communication orale » au sein de la CBE, 2014 -2016
- ✓ Membre de l'équipe AOL de l'objectif « connaissance pratique » 2014-2017
- ✓ Membre de jury de promotion au sein de la CBE, 2015
- ✓ Membre de l'équipe administrative et pédagogique du Master MBA au sein de la CBE, 2016-2017.

- ✓ Responsable Recherche et Innovation au sein du groupe Archimed 2010-2013
- ✓ Responsable R&D au sein de la société Biomediqua, 2009

4. Activités de Recherche

Suite à une formation de mastère recherche à l'ENIS complétée par un stage de recherche qui porte sur les systèmes d'exploitation des cartes à puces, j'ai commencé une thèse de doctorat à l'Université des Sciences et Technologies de Lille en octobre 2004. Effectuée au sein du Laboratoire d'Automatique Génie Informatique & Signal (LAGIS), ce doctorat porte sur la reconfiguration et la régulation des réseaux de transport multimodal.

Recruté sur un poste ATER, j'ai continué mes activités de recherche au sein du LAGIS et en collaborations industrielles et internationales pendant 2 ans et j'ai participé à l'encadrement de sujets de mastère et thèse.

Depuis 2009, j'ai quitté l'université pour travailler dans des sociétés de la région en TIC pour des missions de recherche et innovation. J'ai aussi continué à exercer mes activités de recherche et d'encadrement. Cette expérience industrielle m'a aussi permis d'appliquer mes travaux de recherche dans les domaines d'apprentissage, classification et analyse de données.

4.1. Mastère recherche

SPECIALITE	Nouvelles Technologies Des Systèmes Informatique Dédiés (N.T.S.I.D), E.N.I.S. (Ecole Nationale D'ingénieurs De Sfax (Tunisie))
FORMATION	Systèmes embarqués ; spécification des systèmes distribués ; conception logicielle/matérielle ; classification et reconnaissance des formes ; recherche d'information & data mining ; processeurs dédiés et intelligents ; traitement d'image.
PROJET	"Étude des systèmes d'exploitation dédiées aux cartes à puce" développement d'une partie logicielle dédiée aux cartes à puce à base du système ouvert NESOS et implantation sur une plate-forme matérielle LEON.
STAGE	G.M.S (Groupe en Microtechnologie et Système Sur Puce) E.N.I.S, de septembre 2003 au juin 2004
SOUTENANCE	LE 21 JUIN 2004
COMMISSION D'EXAMEN	Pr. Mohamed ABID, Pr. Adel Mahfoudi et Pr. Mohamed MASMOUDI

4.2. Doctorat

TITRE	"SARR : Système d'Aide à la Régulation et la Reconfiguration des réseaux de transport multimodal"
DOMAINE	Automatique & Informatique Industrielle
ETABLISSEMENT	Université des Sciences et Technologies de Lille (U.S.T.L), Salle

	308, Bâtiment P2.
DIRECTEUR	Salah MAOUCHE, Professeur, Université Lille1 (USTL)
CO-DIRECTEUR	Slim HAMMADI, Professeur, Ecole Centrale de Lille
LABORATOIRE DE RATTACHEMENT	Laboratoire d'Automatique, Génie Informatique & Signal, LAGIS (UMR CNRS 8146), U.S.T.L.
CADRE	La thèse se situe dans le projet : « système d'aide à la régulation du réseau de transport valenciennois et de ses pôles d'échanges », Convention de Recherche GRRT
EQUIPE	ID (Ingénierie de la Décision)
FINANCEMENT	Groupement Régional pour la Recherche dans les Transports (GRRT). Gestion : EGIDE
PERIODE	Octobre 2004 – juillet 2007
SOUTENANCE	04 juillet 2007
JURY D'EXAMEN	Président : Mireille BAYART (USTLille) Rapporteurs : Alexandre DOLGUI (Ecole Nationale Supérieure des Mines de Saint-Etienne) ; Mohamed EL BAGDOURI (UTBM) Examineurs : Aziz MOKRIM (UTC) ; Abdelhakim ARTIBA (Supméca); Christian TAHON (UVHC)
MENTION	Très honorable
QUALIFICATION	en mars 2008 aux fonctions de maître de conférences en 61^{ème} section du CNU (Génie informatique, automatique et traitement du signal) et 27^{ème} section du CNU (Informatique).

SUJET DE LA THESE

RESUME DU MEMOIRE Ce travail de recherche s'inscrit dans le projet « système d'aide à la régulation du réseau de transport valenciennois et de ses pôles d'échanges » ; l'objectif étant la minimisation des temps d'attente des usagers des systèmes de transports urbains en mode normal et perturbé. En fait, La planification des réseaux de transport réalisée en temps anticipé, se base sur les prévisions des conditions de circulation, des durées des trajets et de la demande de transport. Ainsi, il est bien difficile en temps réel de suivre les horaires issus de ce processus de planification à cause de l'apparition des phénomènes complexes et aléatoires affectant le trafic. Afin d'éviter la détérioration de la qualité de service, les perturbations provoquées par ces phénomènes doivent être gérées rapidement à travers un processus de régulation. Dans ce contexte, nous proposons un système d'Aide à la décision nommé SARR (Système d'Aide à la Régulation et la Reconfiguration), pour assister les exploitants dans ces tâches de gestion d'exploitation. Les premières contributions portent sur l'élaboration d'un module d'apprentissage et de classification utilisant la technique S.V.M (Séparateurs à Vaste Marge) pour le diagnostic et la régulation des cas familiers. Mais dans les cas de perturbations plus complexes, simultanées

et difficiles à gérer, le régulateur doit refaire une planification totale ou partielle du réseau en temps réel. Nous proposons alors une première approche à colonie de fourmis pour la reconfiguration spatiale nommée ACFRS (Algorithme à Colonie de Fourmis pour la Reconfiguration Spatiale). Un autre algorithme est développé pour la régulation horaire appelé ACFRH (Algorithme à Colonie de Fourmis pour la Régulation Horaire). Ces deux approches présentent de bonnes performances en termes de convergence, exploitation et exploration des espaces de recherche qui ont été prouvés à travers des scénarii de simulation.

MOTS Transport multimodal, système d'aide à la décision, analyse multicritère,
CLES régulation, reconfiguration, methaheuristique, optimisation, classification
S.V.M, colonie de fourmis.

4.3. Activités de recherche post doctorales

Depuis ma soutenance de thèse en juillet 2007, j'exerce mes activités de recherche portant sur la classification et l'analyse de données. Depuis cette date, j'ai encadré 4 masters, 3 thèses de doctorat soutenues et 3 thèses en cours. Depuis 2009, j'ai eu l'opportunité de proposer des projets industriels qui m'ont permis d'appliquer les approches et techniques proposés dans mes travaux de recherche.

4.3.1. Contexte Général

Comme je l'ai expliqué précédemment dans ce présent mémoire, mes travaux de recherche portent sur l'apprentissage et la classification automatique de données. Sans négliger les autres approches, je me suis basé sur les techniques d'apprentissage statistique notamment le SVM et j'ai proposé des solutions pour des problèmes liés à des systèmes complexes. Outre la reconnaissance de caractères, ces travaux ont traité essentiellement deux axes différents à savoir l'étude des systèmes automatiques et la gestion du risque médical. Ces deux axes ainsi que le sujet sur la reconnaissance de caractères seront détaillés dans la section mémoire de recherche.

4.3.2. Sélection de variables et extraction de caractéristiques

Les techniques d'apprentissage et classification automatique offrent des résultats attractifs. Cependant, leurs performances restent limitées si les variables d'entrées sont non-pertinents vis-à-vis du problème. Un des plus importants sujets de recherche est l'amélioration de ces approches à travers la résolution du problème de sur-apprentissage. Pour remédier à ce problème, plusieurs approches de réduction de dimensionnalité et d'extraction de caractéristiques ont été proposées. L'objectif de ces méthodes est de construire une base de représentation en choisissant parmi toutes les variables disponibles les variables les plus informatives relativement au problème à traiter. Lorsque cette base est constituée d'un sous ensemble de variables d'origine, on parle de méthode de sélection de variables. Dans le cas où cette base est issue de combinaisons linéaires ou non-linéaires de variable d'origines, on parle de méthode d'extraction de caractéristiques. Dans ce même contexte, j'ai travaillé sur une première idée de sélection de variable et qui a fait l'objet du sujet de mastère de Sabra El Ferchichi en 2007 [5] [6]. Puis j'ai proposé un sujet de thèse de doctorat sur l'extraction. Cette thèse a été soutenue en 2013 par la même étudiante [7] [1].

4.3.3. Classification et régression pour les systèmes automatiques complexes

Depuis 2008, j'ai eu l'opportunité de collaborer avec des chercheurs du laboratoire d'Analyse et Commande des Systèmes à l'Ecole Nationale d'Ingénieurs de Tunis sur l'identification, la surveillance, et le diagnostic des systèmes automatiques complexes. Dans ce contexte, j'ai encadré 2 mastères sur la surveillance des systèmes multimodaux et le diagnostic des systèmes complexes et j'ai participé à l'encadrement de 2 thèses de doctorats sur l'identification des systèmes non linéaires et le diagnostic des systèmes industriels. J'ai aussi travaillé sur l'étude des systèmes à commutation dans une collaboration avec l'unité de recherche Commande, Numérique des Procédés Industriels de l'Ecole Nationale d'Ingénieur de Gabès (Tunisie). Pour ces problèmes liés aux systèmes complexes, nous avons proposé des solutions basées sur des techniques d'apprentissage et classification automatique. Ces approches apportent ainsi des contributions complémentaires les unes aux autres tout en considérant des hypothèses et des contextes de travail différents. Aussi, la principale partie de ces travaux, et qui sera exposée dans la suite du manuscrit, peut être présentée selon trois axes:

✓ Surveillance des systèmes

Notre contribution, dans cet axe, consiste au développement d'une nouvelle approche basée sur la technique SVM pour la surveillance des systèmes complexes. Ce sujet a fait l'objet du travail de Master de Mounira TARHOUNI soutenue en Avril 2009 [8]. L'algorithme proposé [9] a été appliqué pour la surveillance du procédé pilote "niveau d'eau" disponible au sein d'unité de recherche ACS de l'ENIT et les résultats trouvés par cette méthode de classification ont été comparés à d'autres approches telle que celle basée sur le calcul de validité. Cette contribution sera détaillée dans le chapitre 2 de la partie mémoire de recherche. A la fin de ce travail, j'ai conclu qu'il faut travailler sur l'identification de ces systèmes pour avoir un modèle solide capable de décrire leur comportement.

✓ Apprentissage statistique pour l'identification des systèmes non linéaires

Les premières contributions, dans cet axe, ont été dans le cadre de la thèse de doctorat de Mounira TARHOUNI [2] soutenues en Mai 2014. En se basant sur la théorie d'apprentissage statistique, nous avons proposé une modélisation mathématique des systèmes non linéaires. En effet, l'étude de ce type de systèmes dynamiques exige une représentation, dite modèle, capable de reproduire leurs comportements. Dans ce travail d'identification, nous nous sommes basé sur les méthodes à noyau avec uniquement une collection d'observations entrée/sortie du système.

Dans ce contexte, nous avons proposé une première approche d'apprentissage statistique à noyau multiple [10] dont l'idée de base est de trainer l'algorithme Least Squares Support Vector Machines (LS-SVM) avec plusieurs fonctions noyaux combinées. Avec des pondérations, nous avons choisi un noyau linéaire et un noyau Gaussien et la méthode a assuré des performances meilleures que la méthode LS-SVM classique qui fait recours à une fonction noyau unique.

Dans une autre contribution, nous avons amélioré cette méthode et nous l'avons appliqué pour l'identification des systèmes à plusieurs entrées/plusieurs sorties (MIMO). Cette amélioration consiste en l'ajout d'un algorithme d'optimisation pour le choix de paramètres [11]. Il s'agit de la méthode d'optimisation par Essaim Particulaire (OEP). Couplée avec l'approche LS-SVM cette approche nous a permis de fixer les paramètres de l'algorithme d'identification et les résultats de cette approche ont été améliorés.

Nous avons passé, par la suite, de la méthode LS-SVM à noyau multiple à une nouvelle approche appelé Least Squares Support Kernel Machines (LS-SKM) [12]. Cette dernière se base sur la division du régresseur en plusieurs vecteurs et nous avons adopté un noyau pour chacun de ces vecteurs.

Nous avons proposé une nouvelle formulation de la méthode LS-SKM pour pouvoir identifier les systèmes de type NARMA [13]. Cette nouvelle formulation est capable d'améliorer la flexibilité de ce type de modèle (NARMA) en se basant sur les couples croisées (Entrée/Entrée, Entrée/Sortie et Sortie/Sortie). Pour chacun de ces trois couples nous avons adopté une fonction noyau.

Nous avons aussi proposé une nouvelle version de la méthode SVM pour la régression dite Support Vector Regression (SVR). Cette contribution consiste à reprendre le même principe de division du vecteur régresseur en plusieurs blocs avec l'adoption d'un noyau par bloque. Il s'agit de la méthode Support Kernel Regression (SKR) [14] [15].

Ces contributions ont été proposées dans le cadre de la thèse de doctorat de Mounira TARHOUNI. Elles seront détaillées dans le chapitre 2 de la partie mémoire de recherche.

Depuis 2015, J'ai commencé un travail sur l'identification des systèmes commutant. Dans cet axe, je travaille actuellement sur deux volets. Le premier consiste à adapter la régression par apprentissage statistique pour l'identification de ce type de systèmes. Le deuxième volet concerne l'application d'une approche de classification pour l'estimation de l'état discret. Ces travaux étaient dans le cadre de la thèse de doctorat de Lamaa SELLAMI soutenue en octobre 2016.

✓ **Apprentissage et classification pour le diagnostic des systèmes industriels**

Outre la surveillance et l'identification, La sûreté de fonctionnement assuré par le diagnostic des systèmes est un besoin principal dans le milieu industriel. En effet, la détection précoce d'un dysfonctionnement contribue à éviter toute dégradation pouvant toucher le système à surveiller et les éléments du milieu voisin (utilisateurs et matériels). Dans nos contributions pour ce problème, la tâche du diagnostic est perçue comme un problème de classification. Notre première contribution dans cet axe consiste à proposer une nouvelle approche de diagnostic des systèmes multimodes basée sur la classification automatique. Cette méthode nous a permis de proposer un processus de diagnostic admettant les capacités nécessaires pour distinguer le mode de fonctionnement dans lequel le système évolue et susceptible de s'adapter aux situations inconnues a priori (modes imprévisibles). Pour ces situations nous avons travaillé avec des approches de classification non supervisées notamment l'algorithme des *k-moyennes*. Ce sujet a fait l'objet du projet de master de Foued THELJANI soutenu en 2009 [16] [17]. Nous avons aussi proposé un algorithme basé sur la géométrie algorithmique (computational geometry), notamment le principe de l'enveloppe convexe [18]. Il exploite la mesure de la densité pour la classification.

Le deuxième volet de cet axe concerne la classification dynamique pour le diagnostic et le suivi des systèmes automatiques évolutifs. Dans ce contexte, nous avons proposé une amélioration de la technique SVDD (Support Vectors Domain Description) à noyau [19]. Il est adapté pour tenir compte de différentes situations liées à la non-stationnarité de données. L'évaluation de ces approches est effectuée sur des systèmes automatique simulés ainsi que des procédés industriels réels [20] [21]. Les travaux sur l'algorithme de densité et la classification dynamique ont fait l'objet de la thèse de doctorat de **Foued THELJANI** [4].

Les contributions de cet axe seront détaillées dans le chapitre 2 de la partie mémoire de recherche.

4.3.4. Gestion du risque médical

Pour le problème de la gestion du risque médical, mes travaux portent sur deux volets dont le premier concerne la gestion des événements précurseurs et des incidents dans les centres hospitaliers. Le second volet traite la gestion des risques liés aux épidémies. Les contributions sur ce deux axes feront l'objet du 3^{ème} chapitre dans la suite du manuscrit.

J'ai commencé les travaux sur la gestion des événements précurseurs, depuis 2009, en collaboration avec la société BIOMEDIQA. Nous avons développé la plateforme RMP (Risk Management Platform) qui sera présenté dans la partie activités industrielles. Il s'agit d'un système d'aide à la décision qui permet de valoriser les éléments importants pour assister et accompagner les opérateurs dans un centre médicale. Dans ce contexte d'aide à la décision, nous avons proposé des algorithmes pour la mise en œuvre des actions correctives. L'idée consiste à appliquer des nouvelles approches d'apprentissage et classification pour exploiter une base d'expertise. Les difficultés liées à cette problématique concernent principalement les données textuelles décrivant les événements indésirables ou explicitant les actions correctives. En effet, outre La redondance d'informations corrélées ou non pertinentes, ces données textuelles ne sont pas exploitables pour les approches de classification. Dans le cadre du master de Thibault JULIEN [22], nous avons travaillé sur l'extraction textuelle pour résoudre une part de cette difficulté. Puis nous avons proposé une approche par apprentissage automatique pour la classification des événements [23] [24]. En 2016, J'ai proposé un sujet de thèse de doctorat dans ce même contexte d'analyse de données médicales à l'Ecole Centrale de Lille.

Pour ce qui concerne la gestion du risque lié aux épidémies, mes travaux ont débuté en juin 2015 dans le cadre d'un projet financé par l'université d'Al Qassim (Qassim University). Dans le cadre de ce travail, nous nous plaçons aussi dans un contexte d'aide à la décision dont l'objectif est d'assister les opérateurs médicaux pour détecter la possibilité d'apparition d'une épidémie et limiter ses risques.

4.3.5. Reconnaissance de caractères

En 20012, j'ai lancé un projet de recherche et innovation dans le cadre du dispositif MOBIDOC financé par l'Union Européenne dans le cadre du programme PASRI et administré par l'ANPR. En collaboration entre le groupe ARCHIMED et le laboratoire Stratégie d'Optimisation et Informatique IntelligentE (SOIE) en Tunisie, ce projet consiste au développement d'un OCR multilingue. Nous avons débuté par un sujet de thèse de doctorat (en cours 3^{ème} année). Nous avons proposé des solutions basées sur des techniques d'apprentissage statistique pour la reconnaissance de caractères arabes notamment les caractères manuscrits [25] [26].

5. Encadrement

5.1. Co-encadrement de thèses

5.1.1. Thèses soutenues

Co-encadrement (60%) de la thèse de **Sabra El Ferchichi** avec Salah Maouche (20%) et Moufida Lahmari Ksouri (20%)

Titre : Sélection et extraction d'attributs pour les problèmes de classification

Date de soutenance : 01 Juillet 2013.

Etablissement : Cotutelle entre l'université de Lille1 et l'ENIT (Tunis).

Durée de la thèse : 40 mois.

Jurys : Stéphane Canu (Prof., Institut National des Sciences Appliquées Rouen) et Moncef Tagina (Prof., Ecole Nationale des Sciences de l'Informatique, Tunisie) (rapporteurs), Olivier Colot (Prof., Université des Sciences et Technologies Lille1), Aziz Moukrim (Pr., Université de Technologie de Compiègne), Kamel Abderrahim (MC., Ecole Nationale d'Ingénieurs de Gabes) et Kaouther Laabidi (MC., Ecole Sup des Sciences et Techniques de Tunis) (Examineurs),

Salah Maouche (Prof., Université des Sciences et Technologies Lille1) (Directeur), Moufida Ksouri (Prof., Ecole Nationale d'Ingénieur de Tunis) et Salah ZIDI (Consultant R&I chez ARCHIMED-Group) (Co-directeur).

Mention : Très Honorable.

Publications : 2 articles en revues internationales (International Journal of Computers Communications & Control et International Journal of Computer Applications), 1 chapitre d'ouvrage (Springer, International Journal of Computer Applications) et 3 communications dans des conférences internationales. Sabra El Ferchichi travaille actuellement en tant que Assistante à l'université de Bizerte (Tunisie).

Co-encadrement (40%) de la thèse de **Foued Theljani** avec Moufida Lahmari Ksouri (20%) et Kaouther LAABIDI (40%)

Titre : Classification dynamique de données pour le diagnostic des systèmes

Date de soutenance : 22 Juin 2013.

Etablissement : ENITunis

Durée de la thèse : 36 mois.

Jurys : Abdelkader Chaari (Maître de conférences, ESSTT) et Kamel Ben Saad (Maître de conférences, ENIT) (Rapporteurs), Faouzi Bouani (Professeur, ENIT) et Salah Maouche (Professeur, UST Lille) (Examineurs), Moufida Ksouri (Professeur, ENIT) (Directeur), Kaouther Laabidi (Maître de conférences, ESSTT) et Salah ZIDI (Consultant R&I, ARCHIMED-Group) (Co-directeurs).

Mention : Très Honorable.

Publications : 4 articles en revues internationales (International Journal of Artificial Intelligence, International Journal on Artificial Intelligence Tools, Journal of Dynamic Systems, Measurement, and Control et Arabian Journal for Science and Engineering) et 5 communications dans des conférences internationales.

Foued Theljani travaille actuellement en tant que Maître-Assistante à l'Institut Supérieur des Sciences Appliquées et de la Technologie de Mateur (université de Bizerte, Tunisie).

Co-encadrement (50%) de la thèse de **Mounira TARHOUNI** avec Moufida Lahmari Ksouri (20%) et Kaouther LAABIDI (30%)

Titre : Contribution à l'identification des systèmes non linéaire par les Séparateurs à Vaste Marge (SVM)

Date de soutenance : 22 Juin 2014.

Etablissement : Faculté des Sciences de Tunis FST

Durée de la thèse : 40 mois.

3 Jurys : Abdelkader CHAARI (Maître de conférences, ESSTT) et Salah SALHI (Maître de conférences, ISI El Manar) (Rapporteurs), MOUFIDA Ksouri (Professeur, ENIT) (Directeur), Kaouther LAABIDI (Maître de conférences, ESSTT) et Salah ZIDI (Consultant R&I, ARCHIMED-Group) (Co-encadrants).

Mention : Très Honorable.

Publications : 4 articles en revues internationales (Int. J. of Modelling, Identification and Control, International Journal of Innovative Computing, Information and Control, Int. J. of Modelling, Identification and Control) et 4 communications dans des conférences internationales.

Mounira TARHOUNI est actuellement en poste Maître-Assistante à l'Institut Supérieur d'Informatique et de Multimédia de Gabes, université de Gabes, Tunisie.

Co-encadrement de la thèse de **Lamaa SALLAMI** avec Kamel ABDERREHIM (Université de Gabes, Tunisie).

4 Titre : Identification de systèmes linéaires commutant

Date de soutenance : 12 octobre 2016.

Lamaa SALLAMI est actuellement en poste Maître-Assistante à l'université AlQassim en Arabie Saoudite.

5.1.2. Thèses en cours

Co-encadrement de la thèse de **Marwa AMARA** avec Khaled GHEDIRA et Kamel ZIDI.

Titre : Reconnaissance de caractères arabes

1 Début : 01 octobre 2012.

Cadre : dispositif MOBIDOC financé par l'Union Européenne dans le cadre du programme PASRI.

Date de soutenance prévue juillet 2017.

5.2. Encadrement de Mastères Recherche

1 **Thibault JULIEN**, Master recherche, Spécialité : Modélisation et Management des Organisations, École Centrale de Lille, titre : Extraction textuelle pour la gestion des évènements indésirables, Avr. 2009 - Sept. 2009, co-Enc. : 80% avec Jean-Pierre BOUREY (Prof., École Centrale de Lille / LAGIS).

2 Sabra El Ferchichi, Master recherche, Spécialité : Génie électrique, École Nationale d'Ingénieurs de Tunis, titre : Sélection de variables et extraction de caractéristiques pour une méthode d'apprentissage, Sept. 2007 - Juin. 2008, co-Enc. : 80% avec avec Moufida KSOURI (Prof., École Nationale d'Ingénieurs de Tunis).

3 Foued THELJANI, Master recherche, Spécialité : Génie électrique, École Nationale d'Ingénieurs de Tunis, titre : Classification non supervisée pour le diagnostic des systèmes industriels, Sept. 2008 - Juin. 2009, co-Enc. : 60% avec Kaouther LAABIDI (MCF., Ecole Sup des Sciences et Techniques de Tunis).

4 Mounira TARHOUNI, Master recherche, Spécialité : Génie électrique, Faculté des sciences de Tunis, titre : Surveillance des systèmes Par Séparateurs à Vaste Marge, Sept. 2008 - Avril. 2009, co-Enc. : 60% avec Moufida KSOURI (Prof., École Nationale d'Ingénieurs de Tunis).

6. Participation et Coordination de Projets

6.1. Projet KACST¹ N°34-440

Titre : A Secure e-Quran Certification System Integrating Audio, Image and Text Watermarking Technologies for Verification of Integrity and Content-Authenticity

Contexte : Projet financé par le KACST sur l'Authenticité et la standardisation des applications logicielles sur le Quran.

Durée : 24 mois (depuis Septembre 2015)

Financement : KACST

Etablissement : Le centre de recherche Noor Center², Medine, Arabie Saoudite

6.2. Projet NOOR N° NRC1-126

Titre : Authentication and Protection of Quran Resources using Text and Image Based Watermarking

Contexte : Authenticité et protection du Quran

Durée : 24 mois ma participation pendant 6 mois (depuis juin 2015)

Financement : Noor Center

Etablissement : Le centre de recherche Noor Center, Medine, Arabie Saoudite

6.3. Projet interne « Qassim University N° 2984 »

¹ KACST : King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, www.k3cst.edu.sa

² Noor Center, www.Noorcenter.edu.sa

Titre : From evaluation to reducing the risk in the medical field.

Contexte : Gestion des risques médicales

Durée : 12 mois (2015-2016)

Financement : Qassim University

Etablissement : Qassim University, Arabie Saoudite

6.4. Projet interne « Qassim University N° 1250 »

Titre : Preventing failures and faults in Wireless Sensor Networks

Contexte : Diagnostique et tolérance aux fautes des réseaux de capteurs

Durée : 12 mois (2016-2017)

Financement : Qassim University

Etablissement : Qassim University, Arabie Saoudite

6.5. Projet SYRACUSE

Titre : SYRACCUSE

Contexte : Appel à Projets lancé par le Fonds national pour la Société Numérique (FSN) avec une Aide à l'innovation par OSEO sur les technologies des contenus numériques, Un produit logiciel d'exploitation et de diffusion de contenus numériques.

Durée : 24 mois

Financement : OSEO + Fonds national pour la Société Numérique (FSN)

Etablissement : ARCHIEMD Group, Lille, France

Partenaires : ARCHIEMD Group, INRIA (Lille, France)

6.6. Projet Stade Ubiquitaire

Titre : Stade ubiquitaire

Contexte : Une aide à l'innovation par OSEO sur une Etude de faisabilité d'un espace ubiquitaire

Durée : 12 mois

Financement : OSEO

Etablissement : ARIZONA-CT (Lille, France)

Partenaires : ARIZONA-CT, Ecole Centrale de Lille, OSEO

6.7. Projet ATRAQ

Titre : Automation-Traceability-Risk Management-Accuracy-Quality assurance

Contexte : Projet mené par BIOMEDIQA et financé par ANR dans le cadre du JEI sur la Gestion du risque médical

Durée : Février 2007 - Juin 2010

Financement : BIOEMDIQA + ANR

Etablissement : BIOEMDIQA (Lille, France)

7. Rayonnement scientifique

7.1. Comités d'Organisation de Sessions

- Session spécial pour le journal Journal of Innovation in Digital EcoSystems (JIDES), titre : "Data Mining techniques for Digital Ecosystems"

7.2. Participation à des Jurys de Thèses

- **Sabra El Ferchichi**, thèse de doctorat, Université des Sciences et Technologies Lille1, titre : "Sélection et extraction d'attributs pour les problèmes de classification", Juillet 2013.
- **Foued Theljani**, thèse de doctorat, Ecole Nationale d'Ingénieurs de Tunis, titre : "Classification dynamique de données pour le diagnostic des systèmes", Juin 2013
- **Mounira Tarhouni**, thèse de doctorat, Faculté des Sciences de Tunis, titre : "Contribution à l'identification des systèmes non linéaire par les Séparateurs à Vaste Marge (SVM)", Juin 2014.

7.3. Évaluation pour des Revues

- International Journal of Production Research (*IJPR*)
- International Journal of Computers and Applications (*IJCA*)
- Journal of Innovation in Digital EcoSystems (*JIDES*)

7.4. Évaluation pour des Conférences

- FLAIRS: The 29 & 30th International FLAIRS Conference
- TELS: Telecommunication Systems
- The International Conference on Artificial Intelligence and Computer Engineering
- The International Conference on Information, Intelligence, Systems and Applications (IISA)
- journées JNMACD et JDMACS.
- (IASTED) International Association of Science and Technology for Development

7.5. Participation aux Groupes de Recherche et Comités scientifiques

- Participation aux groupes thématiques du GdR Modélisation, Analyse et Conduite des Systèmes dynamiques (GDR MACS) : GT ORT (Groupe Thématique Ordonnancement et Réseaux de Transport).
- Membre de la communauté scientifique du centre de recherche sur la technologie de l'information pour le saint Coran et ses sciences (le centre NOOR)(<http://www.nooritc.org/>)
- Membre du comité scientifique du Journal « Journal of Innovation in Digital EcoSystems (JIDES) » <https://www.journals.elsevier.com/journal-of-innovation-in-digital-ecosystems/>
- Membre du comité scientifique de la conférence « The International Conference on Artificial Intelligence and Computer Engineering » <http://www.aice2017.org/>
- Membre du comité scientifique de conférence « The International Conference on Information, Intelligence, Systems and Applications (IISA) » <http://iisa2017.unipi.gr/>

8. Publications

8.1. Revues Internationales avec Comité de Lecture

1. Foued THELJANI, Kaouther LAABIDI, **Salah ZIDI**, et Moufida Lahmari Ksouri, An Efficient Density-Based Algorithm for Data Clustering. International Journal on Artificial Intelligence Tools, 2017.
2. **Salah Zidi**, Bechir Alaya, Tarek Moulahi, Lamri Laouamer, Formal Concept Analysis and Statistical learning theory for aiding detection and classification of epidemics, Wulfenia journal, Vol 24, Issue 1, pp 63-78, 2017.
3. Tibault Julien, **Salah Zidi**, Amir Zidi, Foued Maaloul, A decision support system for adverse events in radiotherapy, Physica Medica: European Journal of Medical Physics, Vol.31, e39-e40, 2015.
4. Bechir Alaya, **Salah Zidi**, Lamri Laouamer, et al. Effect of selective replication strategy and dynamic admission control to QoS management in video on demand systems. Journal of Innovation in Digital Ecosystems, Vol 2, Issues 1–2, pp 45–54, 2015.
5. Foued Theljani, Kaouther Laabidi, **Salah Zidi**, and Moufida Ksouri, A New Kernel-Based Classification Algorithm for Systems Monitoring: Comparison with Statistical Process Control Methods, Arabian Journal for Science and Engineering, Volume 40, Issue 2, pp 645-658, 2015
6. Foued THELJANI, Kaouther LAABIDI, **Salah ZIDI**, et Moufida Lahmari Ksouri, Tennessee Eastman Process diagnosis based on dynamic classification with SVDD, Journal of Dynamic Systems, Measurement, and Control, 2015, vol. 137, no 9, pp 091006.
7. Mounira Tarhouni, **Salah Zidi**, Kaouther Laabidi and Moufida Lahmari Ksouri, Support Kernels Regression For NARMA System Identification (SKRNARMA), International Journal of Modelling, Identification and Control (IJMIC), 2014 Vol.22, No.2. pp.139 – 149, 2014.
8. Foued Theljani, Kaouther Laabidi, **Salah Zidi** et Moufida Lahmari Ksouri, Convex hull based clustering algorithm, International Journal of Artificial Intelligence IJAI, Vol. 10, No. S13, 2013.
9. Sabra El Ferchichi, **Salah Zidi**, Salah Maouche, Kaouther Laabidi and Moufida Lahmari Ksouri "Feature Clustering based MIM for a New Feature Extraction

Method." *International Journal of Computers Communications & Control* 8.5 (2013): 699-707.

10. Mounira Tarhouni, **Salah Zidi**, Kaouther Laabidi and Moufida Lahmari Ksouri, Least squares support kernel machines (LS-SKM) for identification, *International Journal of Modelling, Identification and Control*, vol. 17, no. 1, pp 68-77, 2012.
11. Sabra El Ferchichi, **Salah Zidi** and Kaouther Laabidi, A New Unsupervised Clustering-based Feature Extraction Method, *International Journal of Computer Applications*, 57(6):43-49, 2012
12. Sabra EL FERCHICHI, **Salah ZIDI**, and Kaouther Laabidi, Genetic Algorithm and Tabu search for Feature Selection, *Studies in Informatics and Control (SIC)*, vol.18, No. 2, pp 181-187, 2009.
13. **Salah Zidi**, Salah Maouche, Système d'aide à la régulation et à la reconfiguration des réseaux de transports : SVM et Algorithmes à Colonie de Fourmis, *JESA (Journal Européen des Systèmes Automatisés)*, Vol 43, No10, pp.1121-1148, 2009.
14. **Salah Zidi**, Salah Maouche and Slim Hammadi, Ant Colony with Dynamic Local Search for the Time Scheduling of Transport Networks, *International Journal of Computers, Communications & Control, IJCCC*, Vol. I, No. 4, pp. 110-125, 2006.

8.2. Chapitres d'Ouvrage Collectif

1. Sabra El Ferchichi, **Salah Zidi**, Kaouther Laabidi, et Moufida Lahmari Ksouri Salah Maouch, (2012), A New Feature, Extraction Method Based on Clustering for Face Recognition, Book Chapter, *Engineering Applications of Neural Network IFIP, Advances in Information and Communication Technology*, 26/01/2012, Volume 363/2011, 247-253, DOI: 10.1007/978-3-642-239571_28, Print ISBN: 978-3-642-23956-4, Online ISBN: 978-3-642-23957-1
2. Marwa Amara, Kamel Zidi, **Salah Zidi**, Khaled Ghedira (2014), Arabic Character Recognition Based M-SVM: Review, *Advanced Machine Learning Technologies and Applications, Communications in Computer and Information Science* Volume 488, November 2014, pp 18-25, DOI: 10.1007/978-3-319-13461-1_3, Print ISBN: 978-3-319-13460-4, Online ISBN: 978-3-319-13461-1
3. Marwa Amara, Kamel Zidi, Khaled Ghedira, & **Salah Zidi**, New Rules to Enhance the Performances of Histogram Projection for Segmenting Small-Sized Arabic Words. In *Hybrid Intelligent Systems* (pp. 167-176). Springer International Publishing, DOI: 10.1007/978-3-319-27221-4_14, Print ISBN: 978-3-319-27220-7, Online ISBN: 978-3-319-27221-4, 2016

8.3. Congrès Internationaux avec Actes et Comité de Lecture

1. Sellami, Lamaa, **Salah Zidi**, and Kamel Abderrahim. "Active mode estimation via clustering algorithm for switched linear systems." *System Theory, Control and Computing (ICSTCC)*, 2016 20th International Conference on. IEEE, 2016.
2. Sellami, Lamaa, **Salah Zidi**, and Kamel Abderrahim. "Identification of switched linear systems using self-adaptive SVR algorithm." *Control and Automation (MED)*, 2016 24th Mediterranean Conference on. IEEE, 2016.
3. **Salah Zidi**, Tibault Julien, Anis Mjirda, Foued Maaloul, Textual extraction and classification for medical risk management: A new Risk Management Platform to manage

- undesired medical events, 4th International Conference on Advanced Logistics and Transport (ICALT), Valenciennes, 2015.
4. Foued Theljani, Kaouther Laabidi, **Salah Zidi** et Moufida Lahmari Ksouri, Multi-model approach for systems diagnosis based on SVDD classification, International Conference on Control, Engineering & Information Technology CEIT'13, Sousse - Tunisia , 2013,
 5. Mounira Tarhouni, **Salah Zidi**, Kaouther Laabidi and Moufida Lahmari-Ksouri, System identification based on Support Kernels Regression, International Conference on Control, Engineering and Information Technology CEIT'13, Sousse - Tunisia, 2013.
 6. Foued Theljani, Kaouther Laabidi, **Salah Zidi** et Moufida Lahmari Ksouri, Systems monitoring based on dynamic classification with SVDD , IEEE International Multi-Conference on Systems, Signals and Devices IEEE SSD'13, Hammamet – Tunisia, pp 1-6, 18-21, 2013.
 7. Foued Theljani, Kaouther Laabidi, **Salah Zidi** et Moufida Lahmari Ksouri, Support Vector Domain Description for non-stationary data, IEEE 6th International Conference on Application of Information & Communication Technologies IEEE AICT'12, pp 376-380, Tbilisi – Georgia, 2012.
 8. Sabra EL FERCHICHI, **Salah ZIDI**, Kaouther Laabidi, Moufida KSOURI and Salah MAOUCHE, A new feature extraction method based on clustering for face recognition, 12th conference on Engineering Applications of Neural Networks, EANN, Corfou, Grèce, 2011.
 9. Sabra EL FERCHICHI, **Salah ZIDI**, Kaouther Laabidi, Moufida KSOURI and Salah MAOUCHE, Feature extraction for atmospheric pollution detection, international Conference on Communications, Computing and Control applications, CCCA'11, Hammamet, Tunisie, 2011.
 10. Foued Theljani, Kaouther Laabidi, **Salah Zidi** et Moufida Lahmari Ksouri, 2011, New Approach for Systems Monitoring Based on semi-Supervised Classification, IEEE International Conference on Communications, Computing and Control Applications IEEE CCCA'11, Hammamet - Tunisia, pp 159-164, 2011.
 11. Mounira Tarhouni, **Salah Zidi**, Kaouther Laabidi et Moufida Lahmari-Ksouri, 2011, A Nonlinear MIMO System Identification based on Improved Multi-Kernel Least Squares Support Vector Machines (Improved Multi-Kernel LS-SVM), 8th International Multi-Conference on Systems, Signals and Devices, Sousse, Tunisie.
 12. Foued Theljani, Kaouther Laabidi, **Salah Zidi** et Moufida Lahmari Ksouri, Classification non supervisée pour la surveillance d'un système automatique, 8ème Conférence Internationale de MOdélisation et SIMulation (MOSIM'10), Hammamet – Tunisia, 2010.
 13. Mounira Tarhouni, Kaouther Laabidi, **Salah Zidi** et Moufida Lahmari-Ksouri, Surveillance des systèmes complexes par Séparateurs à Vaste Marges (SVM), 8ème Conférence Internationale de MOdélisation et SIMulation (MOSIM'10), Hammamet-Tunisie, 2010.
 14. Mounira Tarhouni, Kaouther Laabidi, **Salah Zidi** et Moufida Lahmari-Ksouri, System Identification based on Multi-kernel Least Squares Support Vector Machines (Multi-kernel LS-SVM), International Conference on Neural Computation (ICNC 2010), Valencia, Spain, 2010.

15. Sabra EL FERCHICHI, Kaouther Laabidi, **Salah ZIDI** and Salah MAOUCHE, Feature Selection Using an SVM Learning Machines, In SCS'09, 2009 international conference on Signals, Circuits and Systems, Djerba, Tunisie, 2009.
16. **Salah Zidi** and Salah Maouche, Ant Colony Optimization for the rescheduling of multimodal transport networks, CESA'06 IMACS-IEEE Multiconference on Computational Engineering in Systems Applications, pp. 965-971, Beijing, China, 2006
17. **Salah Zidi**, Salah Maouche and Slim Hammadi, Real time route planning of the public transportation system, ITSC2006 The 9th International IEEE Conference on Intelligent Transportation Systems, IEEE Toronto, 2006.
18. **Salah Zidi**, Salah Maouche and Slim Hammadi, Ant colony optimization for spatial reconfiguration of multimodal transport network, 11th IFAC-Symposium on Control in Transportation Systems CTS2006, Delft, The Netherlands, 2006.
19. **Salah Zidi**, Kamel ZIDI and Salah Maouche, Approches métaheuristiques pour des problèmes de transport multimodal, LT' 2006 Workshop International : Logistique et Transport, pp. 237-242, Hammamet - Tunisie, 2006.
20. **Salah Zidi**, Salah Maouche and Slim Hammadi, Support vector machine for regulation of multimodal transport network, EANN2005, The 9th International Conference on Engineering Applications of Neural Networks, pp. 205-212, Lille, France, 2005.

9. Activités industrielles

9.1. Principales missions par entreprise

9.1.1. ARCHIMED : Docteur-Ingénieur Recherche et Innovation (2010 – 2013)

✓ Présentation du groupe ARCHIMED

ARCHIMED est un éditeur et intégrateur de logiciels de gestion documentaire et de partage des connaissances pour la culture, l'enseignement, l'administration et les entreprises.

Né en 1993, le groupe ARCHIMED propose des solutions innovantes qui répondent aux problématiques actuelles des entreprises et des administrations : comment identifier, collecter, puis correctement organiser et diffuser la connaissance, pour qu'elle contribue à la croissance de l'entreprise et au développement d'une région ?

Depuis 2010, ARCHIMED est devenu le premier éditeur français de logiciels pour les bibliothèques et centres de documentation. Le groupe est à ce jour présent dans plus de 600 bibliothèques municipales et universitaires, pour la réalisation d'espaces publics numériques, de systèmes de gestion et de portails documentaires.

✓ Missions

1. Accompagnement pour le CIR et le CII

Pendant 3 ans au sein du groupe ARCHIMED j'ai eu l'opportunité de monter les dossiers annuels de la demande contrat d'impôt à la recherche CIR et du contrat d'impôt à l'innovation CII. Pour chaque projet je prépare le rapport annuel justificatif des travaux éligibles CIR et CII.

2. Développement du Socle logiciel Crystal FrameWork

J'ai intégré l'équipe socle en 2012 et j'ai pu participer au développement du socle logiciel du groupe ARCHIMED. Il s'agit de la plateforme Crystal Framework. Elle sera présentée plus en détail dans la section suivante (Développement d'outils logiciels).

3. Montage et coordination de projets R&I

- *OCR multilingue*

Avec la société Archimed j'ai proposé un projet de recherche qui porte sur la reconnaissance de caractères arabes pour un OCR multilingue. Ce projet a fait l'objet de la thèse de doctorat de Marwa Amara.

- *Le projet SYRACUSE*

Le projet SYRACUSE - alliant les compétences complémentaires de la société ARCHIMED, leader sur le marché des Systèmes d'Information et de Gestion des Bibliothèques, et l'INRIA, reconnu pour ses travaux dans les domaines de moteurs de recherche et de recommandations - ambitionne de proposer sur le marché la première Market place à vocation culturelle, à destination des emprunteurs de livres (et non des acheteurs). SYRACUSE peut se définir comme la transposition des concepts de Market Place tels Amazon vers les réseaux de distribution de contenus numériques (bibliothèques).

- *Réalisation de veilles technologiques*

J'ai aussi réalisé des veilles technologiques et des études notamment sur le protocole Oauth, sur l'intégration de lecteurs de code à barre et QR code dans le logiciel DocFactory et sur d'autres technologies qui ont été intégrées ou développées pour les produits édités par Archimed.

4. Développement des processus de Signature électronique

La conservation à valeur probante des documents est l'un des défis de la société Archimed pour ses solutions de GED. Dans ce contexte, j'ai eu l'occasion de participer par le développement d'un premier processus pour la signature avancée XADES est un second équivalent pour les documents PDF PADES. J'ai intégré ces processus dans la solution logicielle DocFactory qui sera détaillée dans la section développement d'outils logiciels.

9.1.2. BIOMEDIQA : Responsable R&D (2009)

✓ Présentation du Groupe BIOMEDIQA

Le Groupe BIOMEDIQA, créé en 2004, a choisi de se diversifier afin d'assurer une meilleure répartition des risques. Il fait figure aujourd'hui de leader sur le marché de la physique médicale, l'assurance qualité et la radioprotection au sein des établissements de santé, tant publics que privés.

Il est aujourd'hui seul à être présent sur tout le segment des rayonnements ionisants à usage médical, regroupant à la fois les activités liées à l'imagerie médicale, telles que la radiologie médicale, dentaire et vétérinaire, ou encore la médecine nucléaire, mais également les activités liées au traitement des pathologies, telles que la radiothérapie et la curiethérapie.

A côté de ses métiers historiques, le Groupe gère un portefeuille de projets innovants lui permettant d'adapter son développement à l'évolution du marché. Une équipe de recherche

et développement en TIC a été créé en 2009 et a pu proposer deux gammes logicielles à savoir RMP et DOSITRACE.

✓ Missions

J'ai eu l'occasion de collaborer avec cette société pour la première fois en tant que chercheur de l'Université Lille1. Nous sommes partis d'un "brainstorming" pendant lequel j'ai proposé une collaboration dans le cadre d'un projet sur la gestion du risque médical, plus précisément la gestion des événements précurseurs. En septembre 2009, j'ai été recruté en qualité de responsable R&D. Outre le montage des dossiers et la gestion de l'équipe plusieurs missions m'ont été confiées pour la mise en place des premiers produits.

1. Conception et développement de la RMP et DOSITRACE

J'ai participé au montage d'une équipe de trois ingénieurs et un technicien (en physique médicale). Depuis, l'équipe a bien évolué (8 ingénieurs) avec deux produits logiciels sur le marché. J'ai eu l'opportunité de participer au développement du premier logiciel. Il s'agit de la plateforme RMP qui concerne la gestion des événements précurseurs.

Le second logiciel web est une application de contrôle de la dose patient en radiothérapie. Les deux logiciels seront présentés dans la partie produit logiciels de ce présent mémoire.

2. Encadrement

Dans ce même contexte de développement de deux plateformes, j'ai encadré **Thibault Julien** dans le cadre de son stage de master qui a porté sur l'extraction textuelle pour la gestion des événements précurseurs. J'ai aussi encadré les deux étudiants de licence **Abdelkader ELhaj** et **ANAS SAGHROUNI** pendant leur stage en entreprise.

9.1.3. ARIZONA-CT : Consultant Recherche et Innovation (2009)

✓ Présentation d'Arizona

Arizona Consulting & Technology s'est créée avec l'idée que les Collectivités et les Entreprises devraient irrémédiablement se faire conseiller et accompagner dans le domaine des TIC (Technologies de l'Information et de la Communication). Elle se propose d'accompagner ses clients dans plusieurs domaines, liés aux nouvelles technologies de la communication et de l'information en rassemblant les meilleures compétences autour d'un projet et en s'engageant dans la mise en œuvre de ce dernier.

Arizona C&T est une équipe composée de consultants, techniciens formateurs, et chefs de projets agissant sur la France entière et accompagnant ses Clients dans toutes les démarches (amont ou aval) de lancement de projets : assistance à maîtrise d'ouvrage ou maîtrise d'œuvre.

✓ Missions

1. Accompagnement pour la mise en place d'une solution GED à valeur probante

Dans une première mission de consulting au sein de la société ARIZONA j'ai eu l'opportunité d'accompagner le Centre National d'Informatique (CNI) en Tunisie pour la mise en place d'un logiciel de gestion électronique de documents et d'archivage électronique à valeur probante. J'ai dirigé une équipe de trois ingénieurs pour la rédaction des termes de référence (cahier de charge détaillé) puis en tant que consultant externe, j'ai piloté le reste des étapes notamment une assistance à maîtrise d'ouvrage et maîtrise d'œuvre, la mise en place des processus

automatiques workflow et la validation des différentes fonctionnalités de la création à l'archivage. Cette mission d'accompagnement a été précédée par un séminaire sur la GED et l'archivage à valeur probante. J'ai participé à l'organisation et à la préparation du contenu pédagogique et j'ai présenté pendant une journée la GED et ses fonctionnalités.

2. Étude de faisabilité d'un espace ubiquitaire

Dans le cadre d'un projet financé par OSEO, j'ai proposé une étude de faisabilité d'un espace ubiquitaire. Cette étude concerne la mise en place des nouvelles technologies dans un espace sportif ou de festivités (stade, salle de fête, salle de sport...). Ce projet a fait l'objet d'une collaboration avec l'école centrale de Lille. Suite à l'étude, Arizona avec le pôle régional numérique ont organisé une journée pour lancer un appel à projet chez des entreprises numériques de la région.

9.2. Développement d'outils logiciels

9.2.1. CRYSTAL FRAMEWORK - socle documentaire et applicatif commun

Crystal Framework représente la couche applicative commune aux solutions logicielles développées par ARCHIMED. J'ai eu l'opportunité de participer au développement des versions 4.4 et 4.5. Plusieurs modules ont été innovés dans ces deux versions. Entre autres, nous avons intégré SOLR pour l'indexation et nous avons développé un nouveau module DSI (Diffusion Sélective d'Information) pour une notification multicanal (mail, réseaux sociaux, mobiles (SMS)...), multiservice (notamment le service de suivi de recherche) et multifréquence, permettant de définir différentes fréquences de diffusion : tous les jours, toutes les semaines, tous les mois, etc...

Puisque l'indexation d'un gros volume de documents était problématique, nous avons opté pour un nouveau module d'ordonnancement dans ces versions. Ce module contient un API permettant la gestion des tâches, des planifications et de l'historique des exécutions. Nous avons aussi amélioré la gestion de montée en charge et de la tolérance aux pannes.

9.2.2. DocFactory



DocFactory est un logiciel de numérisation édité par ARCHIMED.

Il permet le traitement "qualité" automatique, la normalisation, la conversion et la conservation à valeur probante des documents. La solution Docfactory donne la possibilité de définir plusieurs chaînes de traitement (workflow) facilement paramétrables et avec un contrôle de performance des serveurs (monitoring). J'ai participé par le développement de deux processus de signature électronique. Le premier étant pour la signature PDF avancée PADES et le second est un processus de signature XADES.

9.2.3. RMP Risk Management Platform



RMP est une plateforme de gestion des risques et des événements indésirables en radiothérapie et en imagerie médicale grâce à différentes fonctionnalités conforme à la réglementation telles que par exemple la planification et l'animation des réunions CREX, l'analyse de la criticité de chaque événement indésirable afin de prioriser les traitements. Outre le montage des dossiers et la conception, j'ai participé au développement de ce produit full web dans sa première version. Il est développé en Java JEE.

9.2.4. DOSITRACE



DOSITRACE est une solution de traçabilité et d'optimisation de la dose délivrée aux patients en radiologie, scintigraphie et radiothérapie. Cette solution permet d'automatiser le recueil de la dose patient et son intégration dans le RIS (ou autre système d'information) à chaque examen radiologique.

DOSITRACE est un DACS (Dose Archiving and Communication System). Il permet également grâce à différentes fonctionnalités telles que l'historique dosimétrique pour chaque patient ainsi qu'un système d'alerte, de limiter les surexpositions des patients. Etant un puissant outil statistique, il permet aussi aux médecins d'optimiser leurs temps de travail en priorisant leurs activités.

Il s'agit d'une plate-forme web éditée en JAVA JEE. J'ai participé à la conception et au développement de la première version. En 2014 et en tant que consultant externe, j'ai audité le développement de l'équipe pour la dernière version.

9.2.5. Simplicite

Simplicite est une suite logicielle modulaire développée par Cap RH du Groupe ARCHIMED depuis 2006, capable de gérer efficacement le capital intellectuel, de mobiliser les talents et d'améliorer les performances individuelles et collectives au sein d'une entreprise.

Elle est composée de 3 briques pouvant s'intégrer parfaitement aux autres logiciels de gestion des ressources humaines de l'entreprise (paie, gestion des temps...). Elle permet d'optimiser la gestion de la formation des collaborateurs de l'entreprise, que cette formation se fasse en salle ou en ligne, par la mise en place de plans de formation adaptés, issus pour partie des besoins liés au développement des compétences.

Les trois principaux modules sont e-Compétences, e-Gestion et e-Cursus.



J'ai eu l'occasion de participer au développement d'une partie de ce logiciel en 2006.

Deuxième partie :

Mémoire de recherche

Chapitre 1 : Sélection et Extraction d'attributs pour les problèmes de classification

1. Introduction

La classification automatique est une catégorisation algorithmique d'objets. Elle s'agit d'une affectation d'un objet à une classe ou catégorie, en se basant sur des données statistiques et des observations. Elle fait couramment appel à l'apprentissage automatique. Les deux axes à savoir l'apprentissage et la classification automatique (Machine Learning) ont été conçus pour la résolution automatique des problèmes complexes par la prise de décisions sur la base d'observations de ces problèmes.

Plusieurs travaux de recherche réalisés ces dernières années ont fait naître un développement théorique important dans le domaine d'apprentissage et classification. Néanmoins, le processus est toujours confronté à quelques limitations. Outre ceux liés à l'algorithme lui-même, il y a des problèmes qui sont plutôt liés aux données. Le choix des attributs adéquats pour ces données à classer est une première problématique qui a fait l'objet de plusieurs travaux de recherche. En effet, la catégorisation d'une observation est toujours basée sur ses variables qui la distinguent des autres observations de classes différentes.

Plus on augmente le nombre de variables ou d'attributs décrivant chaque donnée, plus on dispose d'informations concernant ces données et plus on en facilite et on en améliore la classification. Cependant, la qualité de la classification ne dépend pas du nombre d'informations à disposition mais de la pertinence de ces informations. Il est ainsi important de déterminer avec la manière la plus utile ces variables pertinentes. Deux grandes familles de solutions interviennent à ce niveau : la sélection des variables et l'extraction des caractéristiques.

Un des plus importants sujets en matière d'apprentissage automatique est la résolution des problèmes de fléau de dimension et de sur-apprentissage. Les approches de réduction de dimensionnalité ont été proposées pour remédier à ces types de problèmes. Ces méthodes peuvent être regroupées en deux classes à savoir les méthodes de sélection de variables et les méthodes d'extraction de caractéristiques. En ce qui concerne la sélection de variables, plusieurs algorithmes sont disponibles dans la littérature et qui seront présentés dans ce présent chapitre. Quant à l'extraction de caractéristiques, elle constitue une nouvelle alternative qui suscite de plus en plus d'intérêt. Elle s'intéresse à améliorer les performances d'un classifieur donné sans se soucier de la dimensionnalité. Elle repose sur le principe de transformation linéaire ou bien non linéaire de l'ensemble original des variables pour construire un nouvel ensemble qui sera plus déterminant pour la tâche de classification.

2. Classification automatique de données

Pour la prise de décision ou l'assistance à la prise de décision, la classification automatique semble être l'une des meilleures méthodes appliquées et intégrées dans les nouveaux systèmes. Outre ces problèmes de prise de décision, l'analyse de données est un autre domaine très attrayant notamment avec le volume de données qui ne cessent d'accroître sur le web. Les approches de classification automatique ont aussi montré une efficacité pertinente pour ce domaine. Souvent liée à l'apprentissage, la classification automatique consiste en la catégorisation algorithmique de données représentatives d'un ensemble d'objets. La fouille de données connue sous l'expression de "Data Mining" ou encore extraction de connaissances fait appel à la classification comme l'une des techniques les plus utilisées.

- La classification est par définition l'organisation d'un ensemble d'objets en k classes différentes selon des critères prédéfinis. Les objets de la même classe (ou sous ensemble) sont donc similaires selon les mêmes critères prédéfinis. La classification

automatique consiste à générer des règles de décision pour associer à chaque objet (observation) un label qui donne son appartenance à l'une des classes.

Si on suppose \mathbf{X} un ensemble de n objets \mathbf{x}_i dont chacun est décrit par p variables descriptives, le but de la classification est d'attribuer chaque vecteur \mathbf{x}_i (de p attributs) à l'une des k classes. La fonction \mathbf{F} de décision de ce processus devient :

$$\begin{aligned} F: R^P &\rightarrow \{1, \dots, k\} & 1.0 \\ x_i &\rightarrow y_i \end{aligned}$$

$y_i = j$ si x_i appartient à la $j^{\text{ème}}$ classe.

L'apprentissage automatique connu sous le terme anglais "Machine Learning" est le secret de la réussite des techniques de classification. Champ d'étude de l'intelligence artificielle, il s'agit d'un domaine en plein essor. Il consiste à étudier la façon dont les machines (au sens large) peuvent apprendre à partir des données. Les recherches dans ce domaine s'orientent principalement à proposer des algorithmes permettant d'apprendre automatiquement à reconnaître des modèles complexes et prendre des décisions intelligentes sur une base de données [27]. Outre l'ordinateur, que nous considérons le plus concerné par l'apprentissage automatique, toute machine automatique peut être plus intelligente en utilisant ces algorithmes. Ça lui permet d'améliorer ses performances par un processus systématique, et ainsi d'accomplir des tâches difficiles à réaliser par des moyens algorithmiques plus classiques.

Les techniques d'apprentissage automatique et leurs applications pour l'analyse des données ne cessent d'évoluer surtout avec la quantité étendue et les grandes masses de données disponibles sur le web et l'apparition de nouveaux concepts notamment le "Cloud Computing" (informatique en nuage) et le "Big Data". Selon l'information sur les données, il existe quatre classes de techniques d'apprentissage à savoir l'apprentissage supervisé, l'apprentissage non-supervisé, l'apprentissage semi-supervisé et l'apprentissage actif.

2.1 Apprentissage supervisé

A la base, ce type d'apprentissage est le synonyme de la classification. Pour ce type d'apprentissage les classes sont prédéterminées et les exemples sont connus. On parle alors de données labélisées ou étiquetées. En se basant sur une base d'exemples appelée aussi base d'apprentissage, le but de la technique est de généraliser pour des données inconnues ce qu'il a pu « apprendre » grâce aux données déjà traitées par des experts.

Une base d'apprentissage est un ensemble de couples $\langle x_i, y_i \rangle$ dont $\{x_i; i: 1..n\}$ est l'ensemble des entrées du système d'apprentissage et il s'agit de l'ensemble des objets ou observations à classer et $\{y_i; i: 1..n\}$ est l'ensemble des sorties du même système et il s'agit de la décision d'attribution de l'observation à une classe donnée selon l'expert. Les données d'apprentissage sont donc labélisées et le but de la phase d'apprentissage est de déterminer une représentation compacte d'une fonction de prédiction qui permet d'associer à une nouvelle entrée x (non connue dans la base d'apprentissage) une sortie convenable y .

2.2 Apprentissage non-supervisé

Dénoté aussi clustering, cet apprentissage est non supervisé car il traite des données non labélisées à l'entrées. Il consiste à classer un ensemble d'objets sans avoir une expertise préalable ni sur les classes ni sur l'appartenance des données aux classes. En effet, ce processus de clustering permet la catégorisation d'un ensemble d'observations non étiquetées

en se basant sur leur distribution et la corrélation des attributs. Ce type d'apprentissage se différencie de l'apprentissage supervisé par le fait qu'il n'y a pas de sortie a priori. Cette technique attribue les données les plus similaires au même groupe homogène et les données seront considérées comme différentes à d'autres groupes distincts.

2.3 Apprentissage semi-supervisé

Il s'agit d'un type d'apprentissage qui utilise des données étiquetées et d'autres non étiquetées. En se plaçant entre les deux premiers types, cette classe de techniques d'apprentissage automatique a permis d'améliorer significativement la qualité d'apprentissage. Les observations labélisées sont utilisées pour l'apprentissage et les autres observations non labélisées sont utilisées pour affiner la séparation des classes de manière à trouver les meilleurs limites ou contours qui optimisent la catégorisation des données. Parfois avec un nombre important d'observations à classer, il est difficile d'avoir une intervention d'un expert sur la totalité du jeu de données. Dans ce cas, l'apprentissage semi supervisé devient intéressant puisque l'étiquetage ne sera que sur une partie des données.

2.4 Apprentissage actif

Ce type d'apprentissage est une extension (Pour certains chercheurs, il est considéré en tant que l'un des modèles) de l'apprentissage semi-supervisé. Il donne à l'utilisateur un rôle actif dans le processus d'apprentissage. En effet, l'expert peut être demandé à intervenir au cours du processus pour labéliser une observation. Le but est d'optimiser la qualité du model. Contrairement au processus des autres types d'apprentissage où les données sont connues et imposées, en apprentissage actif, les données ainsi que des informations précises sont demandés par l'algorithme.

3. Défis de la classification de données de grande dimension

3.1 Fléau de dimensionnalité

Grace à un progrès technologique vertigineux, à l'apparition de nouvelles notions dans la technologie numérique (tel que l'internet des objets et le "Big Data"), à l'augmentation exponentielle des performances dans le secteur informatique et à l'arrivée de nouvelles machines et systèmes intelligents la quantité des données utiles est de plus en plus importante. Elle double tous les deux ans pour certains secteurs. En effet, toutes les entreprises et les individus génèrent aujourd'hui une masse de données très grande (mégadonnées).

Utiliser efficacement ces grandes quantités d'information est un défi qui nécessite le maniement des méthodes d'analyse automatique pour pouvoir les traiter, trier et exploiter. Néanmoins, le défi est encore plus difficile en présence de données de grandes dimensions. En effet, les processus d'acquisition de données ont beaucoup progressé ce qui a favorisé cette augmentation des dimensions. Ainsi les observations sont de tailles grandes dans les différents domaines notamment les deux domaines les plus concernés par la classification à savoir l'analyse textuelle [28] et le traitement d'image. Par exemple, la qualité des appareils photo aujourd'hui a beaucoup évolué ce qui favorise l'obtention des images de grandes dimensions.

Les difficultés liées à la manipulation de ces données de grandes dimensions sont associées à un problème baptisé "Fléau de dimensionnalité". En effet, avec l'augmentation de dimension, les propriétés bien connues de l'espace de dimension ordinaire 2D et 3D ne seront plus valables. Ce phénomène de fléau de dimensionnalité pose un problème délicat pour les méthodes de classification supervisée ainsi que la classification non supervisée [29]. Le

problème liée à ce phénomène dont le terme est inventé par Richard Bellman [30] consiste à une augmentation du volume de l'espace ce qui rend les données isolées et éparpillées. Ce problème est un obstacle pour les méthodes de classification automatique. En effet, il faut avoir un nombre important d'observations pour pouvoir valider la même distribution de données que dans un espace moins volumineux. Ce qui n'est pas toujours le cas pour ces méthodes d'apprentissage automatique. Souvent, elles font usage à des ressemblances d'un nombre réduit d'exemples dans un espace de possibilités de dimension élevée.

Plusieurs chercheurs ont proposé des études théoriques de l'effet de la dimension en classification [31], [32], [33], [34]. Pour expliquer ce problème, Leo Breiman [33] présente un exemple illustratif de 100 observations dans un espace unidimensionnel $[0, 1]$. Ces 100 observations deviennent des points éparpillés dans l'espace correspondant à 10 dimensions $[0, 1]^{10}$. Pour effectuer une analyse statistique ou classification de ces observations avec une couverture équivalente à celle des 100 points dans l'espace initial $[0,1]$, il faut plus que 10^{20} observations.

Les problèmes modernes les plus concernés par ce phénomène sont les problèmes de traitement d'image notamment la reconnaissance de visage. Ils admettent des données en très grande dimension mais le nombre d'observations disponibles est généralement beaucoup plus faible.

3.2 Phénomène de l'espace vide

Outre le phénomène de fléau de dimensionnalité, d'autres chercheurs [35], [36] ont montré un autre effet de la grande dimension mis en évidence par le phénomène de l'espace vide. Présenté pour la première fois par Scott et Thompson [36], ce phénomène a été illustré par l'exemple du volume de la boule [1]. En observant l'évolution du volume de la boule unité en fonction de la dimension de l'espace dans lequel elle se trouve, il apparaît qu'il devient quasiment nul quand la dimension devient grande. Et pour mieux appréhender ce phénomène, certains chercheurs comme Huber [37] et Bouveyron [35] ont comparé le volume de la boule de rayon 0.9 à celui de la boule unité. Le rapport décroît vers 0 quand la dimension augmente. La probabilité qu'un point se trouve dans l'espace compris entre les boules de rayon 0.9 et 1 suit la formule suivante en fonction de la dimension p : $P(p) = 1 - 0.9^p$. Ces exemples ont montré que l'espace de dimension p est presque vide puisque la plus grande majorité des points se situe aux alentours d'un espace de dimension $p - 1$.

3.3 Phénomène de Sur-apprentissage

C'est un autre phénomène accentué par la grande dimension des données [35]. Aussi appelé phénomène de sur-ajustement, il touche l'apprentissage supervisé et semi-supervisé. Il concerne les classifieurs complexes. En effet, ces classifieurs, tenant compte d'un trop grand nombre de paramètres, sont très ajustés à la forme des données d'apprentissage. Ce problème les rend très dépendants de ces données. Dans ce cas, un classifieur ne pourra pas être efficace pour traiter de nouvelles données, différentes de celles d'apprentissage et surtout si ces données ne sont pas représentatives du processus qui les a générées. À l'inverse du jeu d'apprentissage, le taux d'erreur du jeu de validation a tendance à croître quand le degré de complexité du classifieur croît, comme le montre la Figure 1-1. On dit que le classifieur a sur-appris et il n'est plus assez général à la prédiction.

Il y a aussi le phénomène inverse appelé sous-apprentissage qui survient lorsque le classifieur est trop simple. Dans ce cas, il ne pourra ni être efficace sur les données d'apprentissage ni sur

les données de validation. Il est donc important de trouver le bon degré de complexité pour obtenir un classifieur efficace.

Puisque la complexité du classifieur est en général liée à la dimension de l'espace, le phénomène de sur-apprentissage peut survenir surtout dans les cas des grandes dimensions des données. Outre la validation croisées et la régularisation, la réduction de dimension de données est aussi une solution envisageable pour ce phénomène.

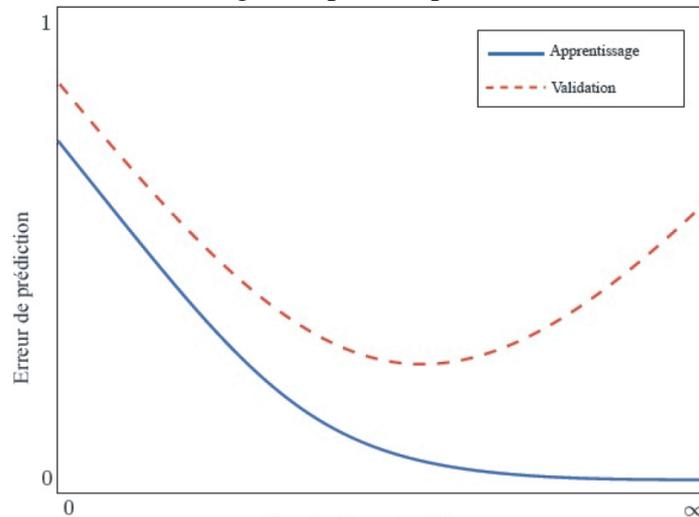


FIGURE 1-1: PHENOMENE DE SUR-APPRENTISSAGE

4. Réduction de dimension

Les nouvelles techniques et les outils d'acquisition de données, nous ont permis, ces deux dernières décennies, d'acquérir des masses de données de tailles importantes de point de vue du nombre d'exemples et du nombre d'attributs. Dans la plupart des situations, on dispose d'un nombre de variables qui dépasse largement le nombre des observations. Ceci pose un problème majeur pour les techniques d'apprentissage automatique supervisé et non supervisé. En effet, outre la complexité algorithmique, la grande dimension de données réduit les performances de ces techniques de classification.

La solution naturelle qui peut faire face à ce problème et aux phénomènes que nous avons présentés précédemment est la réduction de dimension. Elle prend le problème à la source et peut être mise en œuvre avant le traitement des données dans un processus de classification pour limiter les effets et améliorer les performances. Il s'agit de réduire la dimension de l'espace des exemples (le nombre d'attributs dans une observation) pour apporter des éléments de réponse à ces problèmes. Néanmoins, la réduction doit être optimale et sans perte d'information. L'objectif étant la détermination d'un sous-ensemble optimal de caractéristiques pertinentes. Plus petit que l'ensemble initial, ce sous-ensemble est plus représentatif du problème. Il s'agit d'éliminer les informations non-pertinentes et redondantes selon un critère utilisé.

Tel que décrits par Guérif [38], les objectifs de la réduction de dimension sont :

- Faciliter la visualisation et la compréhension des données,
- Réduire l'espace de stockage nécessaire,
- Réduire le temps d'apprentissage et d'utilisation,
- Identifier les facteurs pertinents.

Les techniques de réduction de dimension sont traditionnellement classées en deux grandes catégories à savoir les approches de sélection de variables (variable selection) et les approches d'extraction de caractéristiques (feature extraction). Pour la première catégorie (sélection de variables), comme son nom l'indique, ses techniques consistent à choisir des attributs jugés les plus pertinents dans l'espace de mesure (voir Figure 1-2). Tandis que les techniques de la deuxième catégorie (extraction de caractéristiques) visent à effectuer des transformations sur les variables de départ pour pouvoir extraire une observation plus petite en taille mais avec la même pertinence d'information (voir Figure 1-3). Les résultats de ces approches seront attribués par la suite à l'entrée du système de classification et apprentissage.

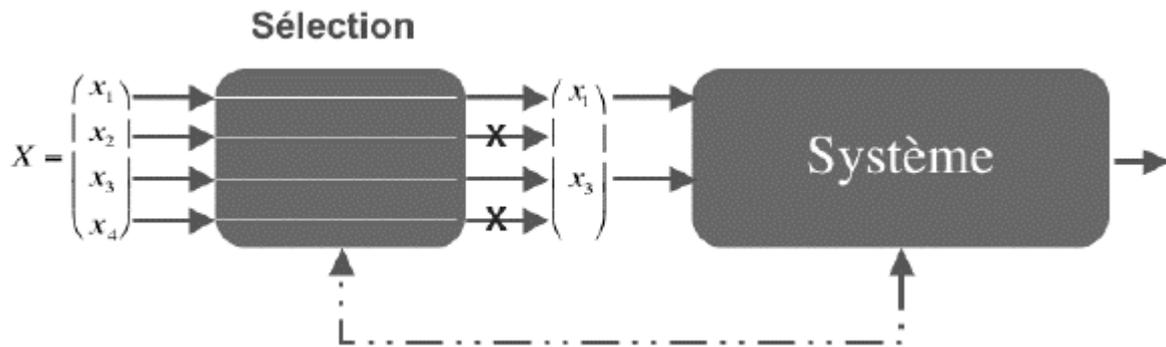


FIGURE 1-2: PRINCIPE DE LA SELECTION DE VARIABLES [38].

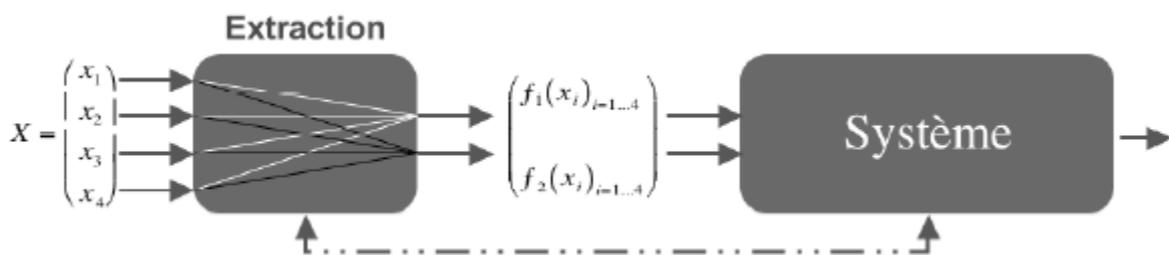


FIGURE 1-3: PRINCIPE DE L'EXTRACTION DE CARACTERISTIQUES [38].

5. Sélection de variables (attributs)

Tel que défini par Bennani [39], la sélection de variables est un procédé permettant de choisir le sous-ensemble optimal de variables pertinentes, à partir d'un ensemble de variables originales, selon un certain critère de performance. La sélection de variables a fait l'objet de plusieurs recherches en statistique et en informatique. Il s'agit d'une problématique de plus en plus importante et d'une complexité croissante.

Les difficultés de ce procédé viennent du fait qu'il faut définir la pertinence des variables et surtout déterminer la meilleure façon pour mesurer cette pertinence et de déterminer la façon de former le sous ensemble optimal. Le critère d'optimalité est un élément essentiel dans le procédé de sélection de variables. Il est aussi difficile de savoir le nombre optimal p de variables à sélectionner parmi n variables initiales. En classification, les variables sont corrélées et la performance d'un classifieur dépend de la combinaison des variables plus que des variables séparées. Cela veut dire qu'une variable A peut-être pertinente en présence d'une autre variable B et non pertinente dans le cas contraire. Il est donc difficile de trouver la meilleure combinaison de variables surtout en grande dimension. En effet, il existe $(2^n - 1)$ combinaisons possibles pour un ensemble de n variables puisque chaque variable admet deux

possibilités : sélectionnée ou non. Le nombre de combinaisons pour trouver un sous-ensemble de p variables parmi n , est égal à : $\binom{n}{p} = C_n^p = \frac{n!}{p!(n-p)!}$.

Dans le cas d'une grande dimension ce nombre de combinaisons devient très élevé. En plus, Le nombre optimal de variables n'est pas connu a priori. Ce qui rend l'espace de recherche très grand. Son exploration complète est alors rude, compliquée et gourmande en temps. Ainsi, la recherche exhaustive n'est pas envisageable dans ce cas. Il s'agit d'un problème NP-difficile. Plusieurs chercheurs ont traité ce problème de sélection de variables à l'aide de différentes approches d'optimisation exactes et heuristiques [40] et d'autres méthodes de recherche de type Branch & Bound [41].

En fonction de l'évaluation des performances et de la définition d'une mesure de pertinence, il existe trois classes d'approches de sélection de variables :

Les méthodes Filtres (filters) : Ces méthodes sélectionnent les variables indépendamment des performances de l'algorithme de classification/régression qui va les utiliser. Elles se basent sur les caractéristiques générales de l'ensemble des données sans tenir compte des relations entre les variables. Le but étant de sélectionner certaines variables et de supprimer d'autres (les moins intéressantes). Outre leur temps de calcul faible, ces méthode offre des résultats robustes au problème de sur-apprentissage. Mais sans prendre en considération la relation entre les variables, elles risquent de sélectionner des variables redondantes. Elles sont donc appliquées en pré-traitement. Comme le montre la Figure 1-4, ces approches ne prennent pas en compte le biais de l'algorithme de classification. Elle sélectionne les variables indépendamment de ce processus ce qui peut ne pas être très rigoureux [42] [43] [44].

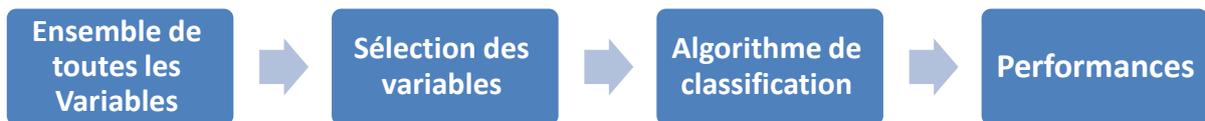


FIGURE 1-4: L'APPROCHE FILTRE POUR LA SELECTION DES VARIABLES

Les méthodes Symbioses (wrappers) : Ces approches prennent en compte l'influence des attributs sur les performances de l'algorithme de classification (Apprentissage). Dans ces approches l'algorithme d'apprentissage est utilisé pour l'évaluation [45] [46]. Tel que présenté dans la Figure 1-5, cette approche évalue plusieurs sous-ensembles de variables à travers l'algorithme de classification puis elle sélectionne les meilleurs. Pour un nombre de variables très grand au départ, ces approches risquent de tomber dans le problème de sur-apprentissage surtout si le nombre d'observations est insuffisant. Pour la même raison l'espace de recherche s'agrandi et le temps de calcul devient important.

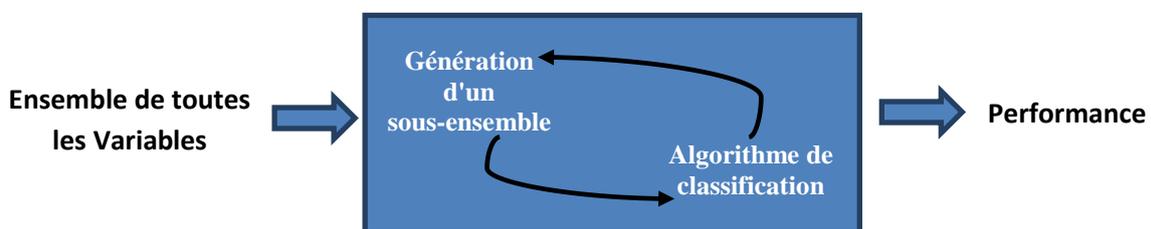


FIGURE 1-5: L'APPROCHE SYMBIOSES POUR LA SELECTION DES VARIABLES.

Les méthodes Intégrés (embedded) : Ces approches combinent les avantages des méthodes filtre et des méthodes symbioses. La sélection de variables s'effectue pendant le (parallèlement au) processus de classification. Les variables pertinentes sont alors sélectionnées de façon à optimiser le critère d'apprentissage utilisé [38] [40] [47].

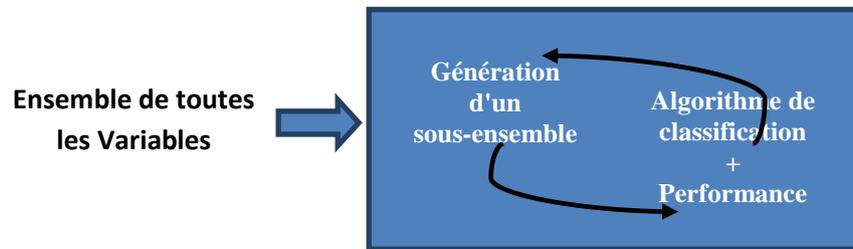


FIGURE 1-6: L'APPROCHE INTEGREE POUR LA SELECTION DES VARIABLES

6. Extraction d'attributs

Le processus d'extraction d'attributs ou des caractéristiques consiste à construire P caractéristiques à partir de N variables originales dont $P \leq N$. Ce processus permet de déterminer un ensemble d'attributs, plus petit que l'ensemble d'origine, sans perte d'information. A travers des transformations linéaires ou non linéaires effectuées sur les variables originales, ces méthodes donnent naissance à de nouveaux attributs porteurs d'information pertinente pour la classification. Contrairement à la sélection des variables, le processus d'extraction des caractéristiques construit un nouvel ensemble de variables, résultantes d'une transformation des variables d'origine, et qui contiennent la plus grande part possible de l'information initiale.

Plusieurs approches ont été proposées pour l'extraction des caractéristiques dont nous pouvons classer en deux catégories à savoir les approches linéaires et les approches non-linéaires. Dans la suite de ce paragraphe nous présenterons les plus connues.

6.1 Approches linéaires

Tel que leur nom l'indique, ces méthodes consistent à appliquer une transformation linéaire sur les attributs. Cette transformation repose sur une distance euclidienne qui utilise à son tour un produit scalaire ordinaire. Il existe principalement trois méthodes classiques. Il y a eu par la suite plusieurs variantes de ces méthodes pour donner naissance à des approches non-linéaires.

✓ Analyse en Composantes Principales (ACP)

L'analyse en composantes principales (ACP) (en anglais Principal Component Analysis (PCA)) est la méthode la plus ancienne. Elle est à l'origine d'une idée parue dans un article de Karl Pearson en 1901 [48]. Appliquée pour différents problèmes d'apprentissage et classification [49], cette approche consiste à déterminer des axes de projections orthogonaux qui maximisent la variance de la projection dans l'espace de caractéristiques. Ces axes sont utilisés pour former une base d'un sous-espace vectoriel de dimension plus réduite. Par projection des points originaux dans ce nouveau sous-espace, on obtient des points dont les coordonnées (nouvelles variables) sont non corrélées. L'ACP se calcule en diagonalisant la matrice de corrélations, le plus souvent en utilisant une décomposition en valeurs singulières (SVD). Linéaire et simple à mettre en œuvre, cette approche est la plus utilisée [38].

Cette approche ne prend pas en compte les dépendances non-linéaires entre les variables et ne peut donc pas fournir une projection fidèle pour une distribution non-linéaire de points. Plusieurs variantes de cet approche ont été alors proposées dont les plus utilisées sont les méthodes varimax et varimin [50] [51] [52] qui se base sur une rotation orthogonale des axes de la même façon que les méthodes quartimax [53] et equamax. D'autres variantes de l'ACP s'appuient sur des rotations obliques comme les approches oblmin et promax.

✓ **Positionnement Multidimensionnel (MDS)**

Le principe de cette approche consiste à construire une représentation en faible dimension d'un ensemble de points d'apprentissage tout en conservant les distances entre eux. Pour un positionnement de N points en m dimensions, l'algorithme MDS est capable de générer une approximation en dimension $d < m$ de ces N points qui respecte exactement les distances fournis [54]. C'est une approche plus flexible que l'ACP mais elle requiert plus de mémoire de stockage. Cette méthode a évolué de la version métrique en se basant sur le modèle euclidien et le produit scalaire pour la mesure des distances vers la version non métrique qui focalise sur l'information au lieu de la distance [55].

✓ **Analyse discriminante de Fisher**

Cette approche appelée analyse discriminante linéaire de Fisher ou l'Analyse Factorielle Discriminante - Fisher Discriminant Analysis (FDA), consiste à transformer les données vers une représentation dans un espace vectoriel de faible dimension qui maximise la variance inter-classe et minimise la variance intra-classe. Elle a été proposée par Fisher en 1936 [56] et elle est adaptée aux observations dont les classes sont connues. Pour k classes il y aura, au plus, $(k-1)$ axes discriminants. Il est nécessaire que les projections des k centres soient les plus séparées que possible contrairement aux projections des données qui doivent être groupées autour du centroïde.

✓ **Analyse en composantes indépendantes**

L'analyse en composantes indépendantes ACI est une approche linéaire connue historiquement en tant que méthode de séparation aveugle de source. Pour un vecteur aléatoire $X=(X_1, X_2, \dots, X_n)$, le principe de cette méthode consiste à identifier le modèle génératif linéaire instantané non bruité suivant :

$$X=A.S$$

Où A est une matrice carrée inversible et $S=(S_1, S_2, \dots, S_n)$ une variable aléatoire latente dont les composantes sont mutuellement indépendantes. On se propose d'estimer A , pour atteindre S_1, \dots, S_n à partir de la donnée d'un échantillon X^1, X^2, \dots, X^n .

Plusieurs problèmes de différents domaines et types ont fait l'objet d'une résolution à l'aide de la méthode d'analyse en composantes indépendantes [57] [58]. Aussi plusieurs approches ont été proposées pour résoudre le problème de l'ACI [57] [59] [60].

6.2 Approches non linéaires

Les méthodes linéaires ne sont intéressantes et ne peuvent résoudre les problèmes d'extraction que si les données ont une distribution linéaire. Elles ne peuvent traiter le cas des données d'une variété non linéaire. En effet, ces méthodes reposent sur l'usage d'une distance euclidienne à travers le produit scalaire ordinaire. D'autres approches non linéaires ont été proposées pour palier à ce problème. La théorie des espaces de Hilbert est l'une des solutions proposées pour permettre de définir d'autres produits scalaires différents du produit scalaire ordinaire et basés sur des fonctions noyaux $k(x,y)$. Ces fonctions noyaux sont aussi des mesures de similarité entre les points traités. Le fait de remplacer le produit scalaire habituel par la fonction noyau rend ces méthodes non linéaires. Cette notion de fonction noyau a été introduite par Vapnick [61] dans le cadre de la théorie de l'apprentissage statistique ou la méthode Séparateur à Vaste Marge appelée aussi support Vecteur de Machine (SVM).

✓ **Kernel PCA**

Cette approche est une version non-linéaire de l'ACP. Proposée par Scholkopf [62], cette méthode consiste à remplacer le produit scalaire ordinaire par une fonction noyau [63]. Dans

ce cas l'ACP sera applicable pour des données situées sur une variété non-linéaire. Si on considère la transformation suivante : $x \in R^d \rightarrow \phi(x) \in H$ avec H est un espace de Hilbert, il existe une fonction noyau (kernel) $K(x_i, x_j)$, tel que $\forall x_i, x_j \in R^d, K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. On n'a pas à calculer $\phi(x_i)$ explicitement puisqu'on aura besoin uniquement du produit scalaire remplacé par la fonction K . D'autres approches non-linéaires ont été proposées et qui font usage à des fonctions noyaux comme la méthode LDA [64].

✓ **ISOMAP**

Cette approche est très similaire à l'approche MDS dont la seule différence est que l'approche ISOMAP utilise la distance du graphe alors que la méthode MDS utilise la distance euclidienne. Un graphe qui relie chaque point à ses k plus proches voisins est ainsi construit. La longueur du plus court chemin entre deux points dans le graphe est utilisée pour estimer la distance géodésique. Une fois les distances calculées, on peut appliquer MDS pour déterminer un positionnement des points dans un espace de dimension réduite [65] [66].

✓ **Plongement localement linéaire LLE**

La méthode LLE (Local Linear Embedded) aborde le problème de la réduction de dimension par reconstruction des points. Chaque point est reconstruit à partir de ses plus proches voisins. Tout en préservant le voisinage. Cette méthode procède par une projection vers un espace linéaire de faible dimension [67].

✓ **Approche neuromimétique : Réseaux de neurones autorégressifs**

Il s'agit de la version ACP neuronale non linéaire. Son principe consiste à minimiser l'erreur moyenne de reconstruction des points à travers leur projection dans un espace de dimension réduite.

7. Nouvelle approche de sélection de variables

7.1 Description de la solution proposée

Cette approche a été proposée dans le cadre des travaux de Mastère de Sabra Elferchichi [5] [6]. Il s'agit d'une méthode de type "wrapper" pour la sélection des variables. En partant d'un ensemble d'observations chacune décrite par la liste complète de variables, on applique une méthode dite d'élimination en arrière pour sélectionner les attributs les plus pertinents. Cette sélection est réalisée à travers une recherche guidée par la méthode Tabou [68]. Cette approche d'optimisation a été utilisée plusieurs fois et par plusieurs chercheurs pour la sélection des variables [69] [70] [71] [72]. D'autres méthodes d'optimisation et des approches hybrides ont été aussi proposées pour le même problème de sélection de variables. La plupart de ces méthodes ont été présentées dans une revue publiée en 2011 [73]. La pertinence des variables est décidée par une phase d'apprentissage par SVM exécutée suite à chaque sélection. L'objectif de cet algorithme d'optimisation étant la recherche de la combinaison d'attribut qui minimise le taux d'erreur de classification. Ce taux d'erreur est calculé sur une base de données de test. Un sous ensemble optimal d'attributs enfin trouvé, sera retenu pour la classification. Le principe général de la solution est illustré par la Figure 1-7 suivante :



FIGURE 1-7: PRINCIPE DE LA METHODE DE SELECTION

Comme le montre la Figure 1-8, cette procédure de sélection des variables est composée de trois phases à savoir la recherche tabou pour la génération du sous ensemble d'attributs, une phase d'apprentissage pour la génération d'un hyperplan séparateur et une phase de validation par détermination du taux d'erreur. Ce taux d'erreur est la fonction coût à optimiser qui sera retourné vers le bloc de recherche Tabou. Ces trois phases seront itérées jusqu'à ce qu'une condition d'arrêt soit vraie. Cette condition d'arrêt concerne une satisfaction en termes de taux d'apprentissage c'est-à-dire un taux d'erreur minimale.

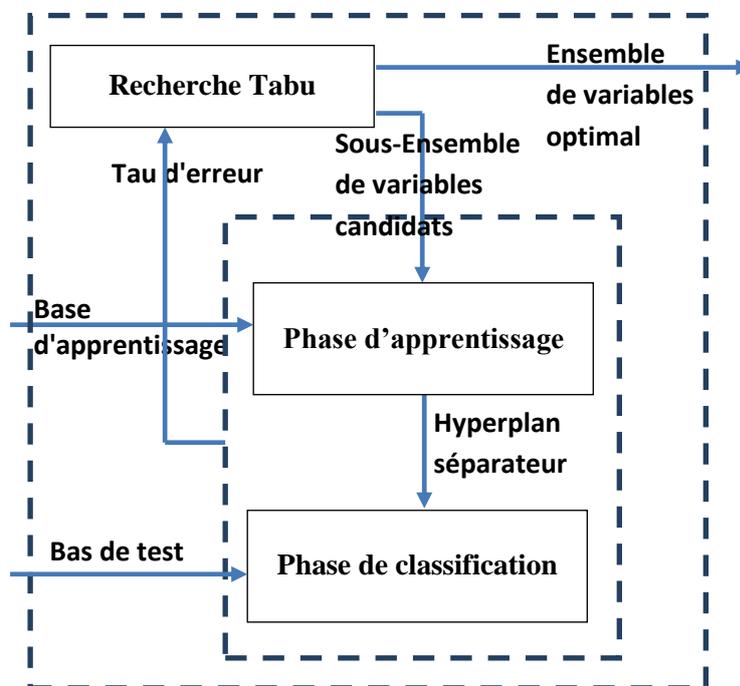


FIGURE 1-8: RECHERCHE TABOU POUR LA SELECTION DE VARIABLES.

7.2 Algorithme d'optimisation

Comme le montre la Figure 1-9, l'algorithme est basé sur la recherche Tabou [68] pour la sélection de p variables parmi n . Nous sommes devant un problème complexe dont le nombre de combinaison est égal à $\frac{n!}{p!(n-p)!}$. Outre l'initialisation des paramètres tels que la liste tabou et le nombre d'itération, dans la première phase, on divise l'ensemble des observations en deux bases. La première, composée de $1/3$ des vecteurs, est une base d'apprentissage alors que le reste ($2/3$) des observations sera utilisé pour les tests.

Partant d'une solution initiale qui comporte tous les attributs (n variables), on génère une population de voisins autour de ce point initial d'une façon aléatoire. Pour avoir un voisin il suffit de retirer quelques variables de l'observation et on aura des voisins avec $p < n$ variables.

L'évaluation des voisins se fait à travers le calcul de la fonction coût. Cette dernière est le taux d'erreur de la classification calculé sur la base de test. On utilise la méthode SVM pour la classification et la résolution du lagrangien est faite avec l'algorithme de points intérieurs. Cet algorithme s'inspire du principe du problème d'optimisation simplexe. L'entrée de ce module est la base d'apprentissage extraite de la base principale. Selon la combinaison des paramètres proposée par le module de génération de voisinage, la base d'apprentissage sera modifiée. La sortie sera un hyperplan séparateur optimal caractérisé par l'offset b et les multiplicateurs de Lagrange α_i correspondants.

Après avoir évalué toute la population de voisins, on sélectionne le meilleur voisin qui a le coût le moins faible c'est-à-dire le taux d'erreur le plus faible. Si ce taux d'erreur est inférieur à celui de la valeur optimale, on assigne donc la valeur du meilleur voisin à la valeur du vecteur optimal (combinaison optimal).

Suite à la sélection du meilleur voisin, on le charge dans la liste Tabou avec sa fonction coût correspondante. Et pendant la génération d'un voisin, on vérifie s'il existe déjà dans la liste. Il sera ainsi interdit de la rajouter encore une fois pour ne pas abuser de l'espace de stockage de la liste, étant donné que sa taille est fixée dès le début. Pour plus d'exploration de l'espace de recherche, nous avons utilisé le principe du saut aléatoire.

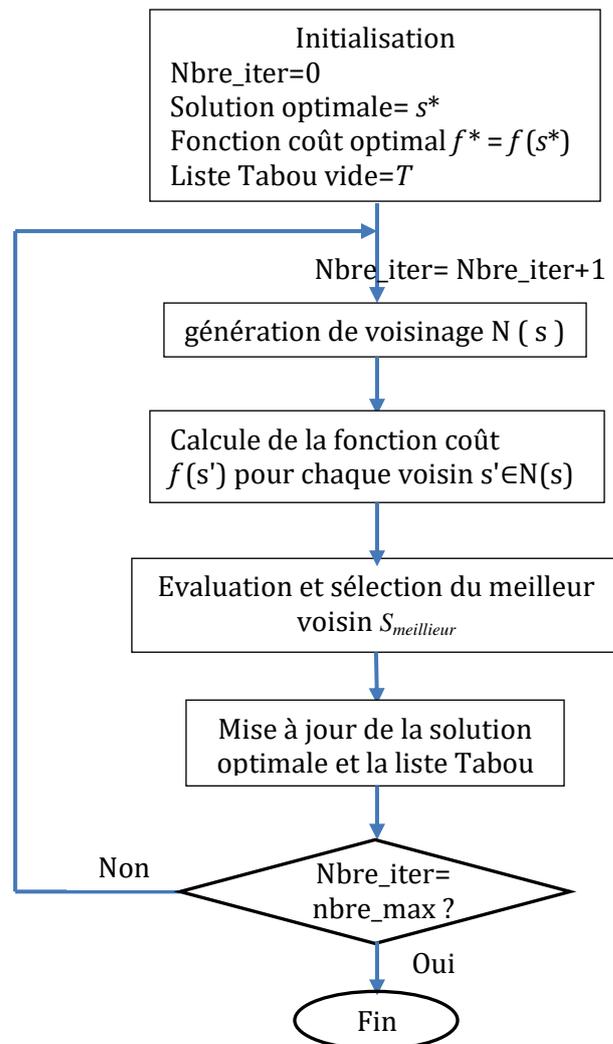


FIGURE 1-9: ORGANIGRAMME DE L'ALGORITHME PROPOSE UTILISANT LA RECHERCHE TABOU.

Nous avons remplacé la recherche Tabou par un algorithme génétique (Figure 1-10) et nous avons comparé les deux approches [5]. Les résultats de l'algorithme génétique sont légèrement plus intéressants. Pour uniquement 200 itérations, ce dernier arrive à avoir un optimum avec un taux d'erreur inférieur à celui de la recherche tabou.

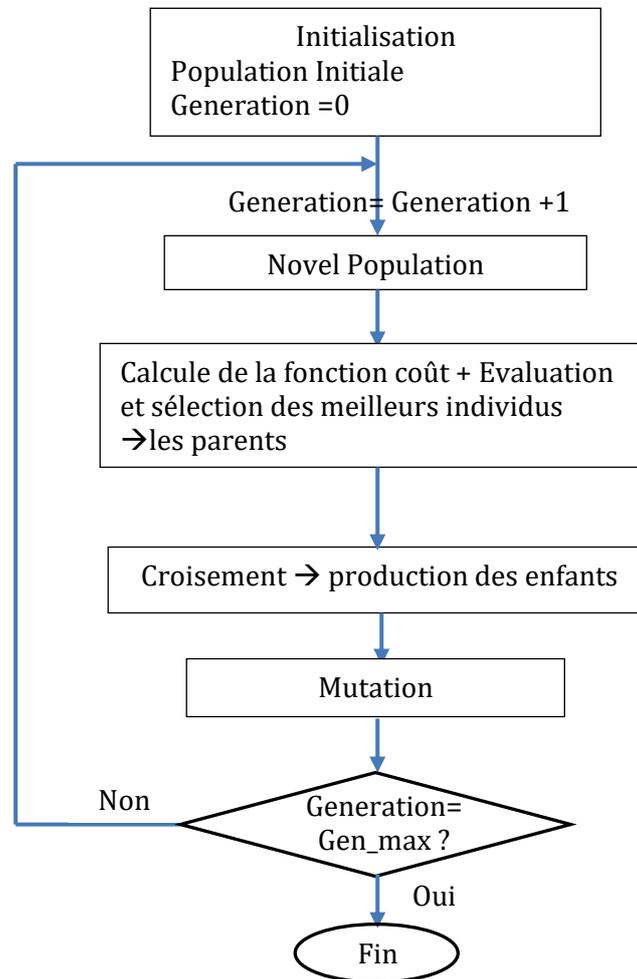


FIGURE 1-10: ORGANIGRAMME DE L'ALGORITHME PROPOSE UTILISANT L'ALGORITHME GENETIQUE.

7.3 Application à la régulation des réseaux de transport

L'approche a été testée sur un problème de classification des perturbations d'un réseau de transport multimodal. Il s'agit d'un problème qui a été traité dans le cadre de ma thèse [74]. Le contexte de ce problème est celui des systèmes d'aide à la décision pour la régulation et la reconfiguration d'un réseau de transport public en cas de perturbation. L'approche de classification utilisée pour ce problème est un algorithme SVM d'apprentissage statistique. Nous avons continué avec la même approche d'apprentissage statistique pendant la sélection. Cette classification des perturbations a donné un taux d'erreur égal à 5.3% partagé sur plusieurs niveaux puisqu'il s'agit d'une classification MSVM (Multiclass Support Vector Machines).

7.4 Résultats

En partant d'une base relativement petite de 198 vecteurs perturbation, nous l'avons divisée en deux bases dont le 1/3 pour la première base d'apprentissage et les 2/3 pour la deuxième base

des tests. La taille de la base nous a obligés d'appliquer une validation croisée. Au départ les observations sont de dimension 24 chacune. Ces vecteurs représentent différentes sortes de perturbations de l'exploitation d'un réseau de transport. En se basant sur la base d'apprentissage, nous avons effectué des tests de validation sur la base de test, avec les résultats illustrés dans le Tableau 1-1 suivant :

TABLEAU 1-1: RESULTATS DE L'ALGORITHME SVM AVANT LA SELECTION.

Vecteurs d'apprentissage		Vecteurs de test	
67		131	
Support vecteur	Non support vecteur	Vecteurs Bien classés	Vecteurs mal classés
25	42	127	4 =3%
Total = 198			

Les paramètres suivants ont été fixés à ces valeurs suite à plusieurs exécutions de réglage :

- Le nombre maximal d'itérations : 3000 itérations,
- La taille de la liste taboue : 44,
- Le nombre de voisins : 7 voisins,
- Le pas des sauts aléatoires : les itérations à la quelles on fait un saut sont {45, 145, 245, 345, 445, 545, 645, 745, 845, 945, 1045, 1245, 1345, 1545, 2145},
- Génération du voisinage : La représentation des solutions est sous la forme d'un vecteur de bits binaires (0 ou 1) pour indiquer la présence ou non de variable dans la même position dans le vecteur observation. Nous avons utilisé une fonction de voisinage qui change deux bits seulement, aléatoirement, dans le vecteur. Suite à plusieurs exécutions de l'algorithme, on a opté pour ces valeurs de réglage.

La fonction à optimiser est un taux d'erreur. La méthode que nous suivons est l'élimination en arrière, donc notre solution initiale est composée de la combinaison des 24 paramètres tous présents. Le taux d'erreur correspondant à cette combinaison est de 0.03 soit 3%. Notre but est alors de chercher un taux inférieur à cette valeur. Plusieurs essais ont été faits pour aboutir à ce réglage qui a donné un résultat optimale égal à 1,5% c'est à dire la moitié de la valeur initiale 3%. L'algorithme atteint son optimum global qui est le 1,5% dans un laps de temps égal à 4.78 secondes. Sinon l'algorithme converge vers un optimum local=2,29% de fréquence un sur dix cas. Parmi les combinaisons correspondantes à un taux d'erreur égal à 1,5%, on trouve les combinaisons suivantes, elles sont les plus fréquentes :

$x_{opt} = \{ [10111111101111101110011], [101110111110101111111011], [101111100101111110111001], [11111010110110111111101], [1011111111011111111011] \}$. En effet, on remarque que l'algorithme converge généralement vers 1500 itérations, 2000 ou bien 2900 itérations. Durant la recherche, l'algorithme rencontre plusieurs valeurs d'optimums locaux {3%, 3,8%, 0.053 et 0.0229}. La Figure 1-11 représente un cas de convergence où on remarque bien que l'algorithme débute à une valeur de 3,8%, il y reste 221 itérations, puis passe à la valeur 3% pendant 110 itérations et ensuite il reste 1555 itérations à la même valeur 2,29%, il y reste 1555 itérations et enfin arrive à 1,5% à l'itération 1885.

Comme on peut le remarquer sur la Figure 1-11, un des optimums locaux est très fréquent dans la recherche. Il s'agit de la solution qui a la valeur 0.0229 (2.29%). C'est un optimum local dans lequel l'algorithme trouve des difficultés à en sortir. Dans ce cas l'algorithme n'a pas réussi à bien explorer tout l'espace avant de retomber dans cet optimum local d'une façon

prématurée. Puis il a pu converger vers l'optimum global d'une valeur de 0.015 (1.5%). Relativement à la valeur initiale avant la sélection, les deux solutions trouvées par l'algorithme (l'optimum local et l'optimum global) peuvent être considérés comme acceptables et même comme de bons résultats. En effet, outre le taux d'erreur qui a baissé, le nombre de variables a été aussi diminué de 24 initialement à 19 après la sélection dont le vecteur représentatif de la solution correspondante est $x_{opt} = [1001111111101111101101]$.

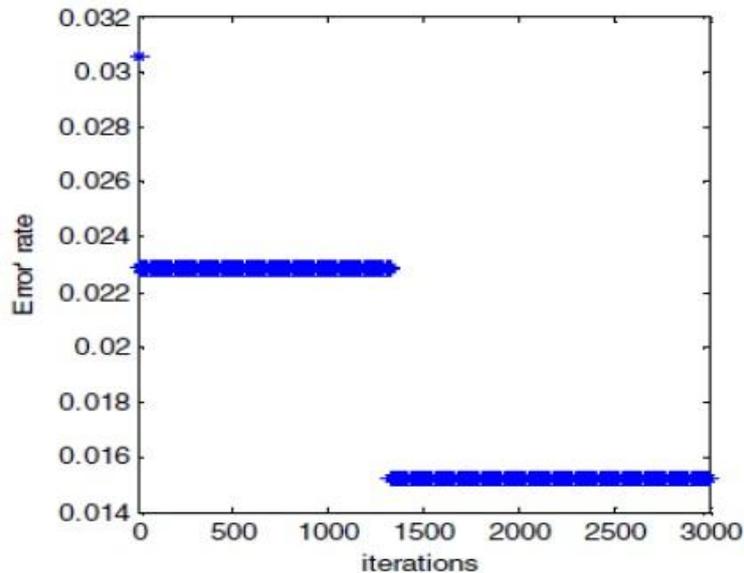


FIGURE 1-11: CAS DE CONVERGENCE VERS L'OPTIMUM GLOBAL (1.5%).

8. Approche par classification non supervisée pour l'extraction d'attributs

8.1 Principe de la méthode

Dans cette section, on présente le principe de l'approche d'extraction de caractéristiques² que nous avons proposée dans le cadre de la thèse de sabra El Ferchichi [1]. Cette méthode est basée sur une classification des attributs indépendamment de la structure de données et de leurs appartenances. Il s'agit d'identifier les clusters d'attributs qui ont les mêmes comportements sur l'ensemble des données de la base. La méthode fait usage d'une analyse de la tendance des attributs pour la détermination de la similarité entre ces derniers. Le but étant l'identification de toute sorte de redondance d'information dans la base de données. En effet, la redondance d'information est une caractéristique intrinsèque de données de grande dimension, qui complique la tâche d'apprentissage et dégrade les performances de la classification [75]. Néanmoins, ces attributs jugés non informatifs ou non pertinents peuvent interagir implicitement en combinaison avec d'autres attributs pour la discrimination des données. Groupés avec d'autres, ces variables donnent, parfois, une information pertinente pour la classification. Ainsi, leur élimination totale n'est pas toujours la bonne décision. L'approche d'extraction que nous avons proposée cherche à identifier les meilleures combinaisons des attributs qui ont des comportements similaires. Nous avons utilisé une mesure de similarité basée sur l'analyse de la tendance et capable d'identifier les interactions internes des attributs. Ensuite nous avons appliqué une méthode de Clustering pour déterminer les groupes (classes) d'attributs similaires. Les centres de ces classes peuvent regrouper les informations portés par ces attributs en une représentation unique. L'ensemble des attributs du même cluster ont été alors remplacés par leur centroïde.

La méthode clustering a été appliquée sur l'espace d'attributs original composé de D attributs V_1, \dots, V_D . V_i est un vecteur de dimension L dont L est le nombre d'observations dans la base de données. Une distance entre deux vecteurs d'attribut V_i et V_j a été bien définie pour identifier la partition optimale. Elle est basée sur une mesure de similarité bien étudiée pour répondre aux objectifs de l'extraction. Cette mesure de similarité sera présentée dans le paragraphe suivant.

Les $d < D$ groupes d'attributs ou Clusters $C_{k \in \{1, \dots, d\}}$ ont été, par la suite, représentés par leur centroïdes donnés par la formule suivante : $g_k = f(S_k) = \sum_{i=1}^L \sum_1^{n_k} w_i V_i$. S_k est l'ensemble des vecteurs attributs appartenant au cluster C_k et f est la transformation appliquée à l'ensemble des attributs dans C_k , définie par les poids élémentaires W_i attribuées à chaque attribut appartenant à C_k . L'ensemble de centroïdes obtenu représente alors l'ensemble des d nouveaux attributs qui vont représenter l'ensemble original des L observations de la base de données.

8.2 Mesure de similarité

La définition d'une mesure de similarité est une tâche importante voir essentielle dans la plupart des procédures de classification ou de Clustering [76]. Ces techniques cherchent les groupements d'individus similaires. Ainsi, nous avons besoin d'une métrique pour mesurer cette similarité en prenant en compte les objectifs voulus. Plusieurs distances ont été conçues par des chercheurs pour mesurer la similarité entre deux points. La distance euclidienne [77] est l'une des distances les plus utilisées pour l'analyse de données. Utilisée pour évaluer la proximité des objets, elle calcule la moyenne de la différence entre deux vecteurs. La distance VDM ou Value Difference Metric est une autre mesure de similarité basée sur des techniques probabilistes simples et qui a été proposée aussi bien pour la classification que pour la sélection des variables [78].

La plupart des distances proposées auparavant n'utilisent aucune information sur l'aptitude des individus mais elles utilisent souvent la moyenne de la différence entre deux vecteurs. Ce qui n'est pas toujours adéquat puisqu'on peut trouver deux vecteurs avec la même moyenne mais qui ont des tendances différentes. Nous avons proposé une nouvelle mesure de similarité basée sur l'analyse de la tendance. Cette méthode permet de mesurer la différence des comportements de deux vecteurs attributs donnés. Pour caractériser un vecteur nous avons besoin d'analyser sa tendance le long de la base d'observations en se basant sur l'étude de la monotonie. Ainsi, pour déterminer la tendance d'un vecteur attribut V , nous calculons sa dérivée première, à travers la différence entre deux valeurs V_i et V_{i-1} , consécutives prises par l'attribut V dans les deux observations X_i et X_{i-1} .

$$\frac{dv}{dx} = \frac{v_i - v_{i-1}}{x_i - x_{i-1}} \quad 1.1$$

Par la suite, la monotonie est déterminée par le signe de cette dérivée première $\alpha = \text{sign}\left(\frac{dv}{dx}\right)$

$$Si \begin{cases} \frac{dv}{dx} < 0 \\ \frac{dv}{dx} > 0 \\ \frac{dv}{dx} = 0 \end{cases} \text{ Alors } \alpha = \begin{cases} -1 & (\text{décroissant}) \\ 1 & (\text{croissant}) \\ 0 & (\text{stable}) \end{cases} \quad 1.2$$

Un vecteur attribut $v \in R^L$ est représenté par L variables tel que $\alpha \in \{0,1,-1\}$. Après avoir déterminé la monotonie de chaque vecteur attribut, une comparaison des tendances de deux vecteurs est possible. Nous avons choisi une calcul de la différence (distance) entre deux vecteurs basée sur le nombre d'occurrence des trois signes de la dérivée première.

Cette distance est donnée par l'équation 1.3. Il s'agit de la racine carrée de la somme des différences absolues donné par l'équation 1.4, entre l'occurrence d'une valeur spécifique de α pour deux vecteurs données. Ceci permet de garantir l'indépendance de la distance développée entre deux vecteurs attributs par rapport aux positions de chaque α (expression d'une tendance) grâce à la distribution conditionnelle de chaque tendance extraite.

$$d(v_i, v_j) = \sqrt{\delta_1(v_i, v_j) + \delta_{-1}(v_i, v_j) + \delta_0(v_i, v_j)} \quad 1.3$$

Avec

$$\delta_\alpha(v_i, v_j) = \left| p(v_i/\alpha) - p(v_j/\alpha) \right|, \alpha \in \{0,1,-1\} \quad 1.4$$

$$p(v_i/\alpha) = \frac{\text{occurrence de } \alpha}{L} \quad 1.5$$

$p(v_i/\alpha)$ est déterminé en calculant le nombre d'apparition de la valeur $\alpha = 0$, $\alpha = 1$ ou $\alpha = -1$ dans le vecteur attribut v_i (voir Figure 1-12).

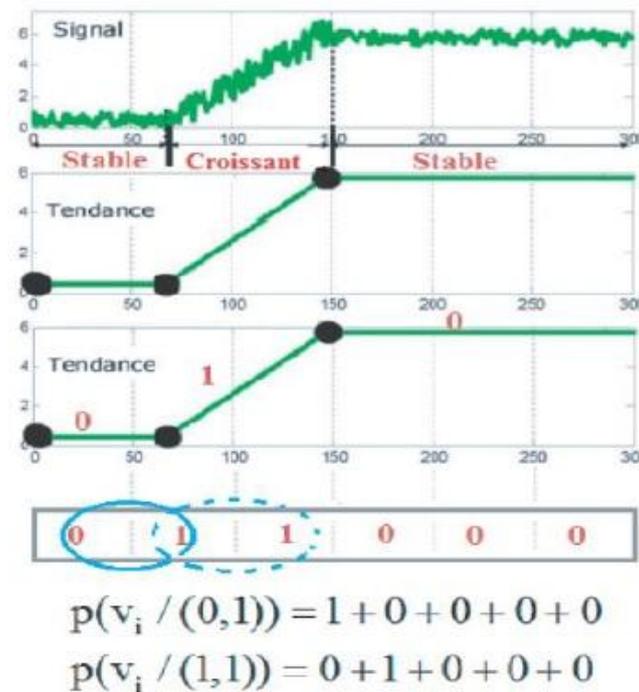


FIGURE 1-12: ANALYSE DE TENDANCE ET MESURE DE SIMILARITE.

Nous avons aussi proposé un calcul de similarité basé sur deux variables $\{\alpha, \beta\}$ au lieu d'une seule [79]. L'idée consiste en une comparaison de deux vecteur à travers le nombre d'occurrence du couple de valeur de α et β . 9 combinaisons sont possibles au lieu de 3 dans le premier cas.

$$(\alpha, \beta) \in \{(1,0), (1,1), (1,-1), (0,0), (0,1), (0,-1), (-1,0), (-1,1), (-1,-1)\}$$

Par la suite, Nous avons déterminé une matrice de similarité M contenant les distances entre les différents vecteurs attributs. $M = d(v_i v_j)$. Le comportement des attributs peut être comparé classe par classe et dans le cas de la version supervisée ou bien indépendamment des classes des observations dans le cas de la version non supervisé.

8.3 Classification des attributs

Après avoir analysé la tendance de chaque vecteur attribut v_i de la matrice des attributs V , la deuxième étape de notre algorithme consiste à construire une matrice de même dimension que V contenant la tendance de chaque attribut codée en 0, 1, et -1. Puis la probabilité conditionnelle $p(v_i/(\alpha, \beta))$ de trouver la combinaison (α, β) dans v_i est déterminée pour calculer les distances entre les attributs deux à deux. Ce qui permet de construire la matrice de similarité M . Puis une approche de "clustering" C-means est appliquée pour trouver les partitions dans l'espace d'attributs en se basant sur les valeurs de cette matrice M , ordonnées par ordre décroissant. Après avoir fixé un nombre initial de clusters, nous initialisons chaque cluster C_k par l'attribut classé en premier dans la matrice M . Ce même attribut est considéré premier centre du cluster. Par la suite, les premiers attributs classés dans M par rapport à ce centre sont sélectionnés. Le nouveau centre est alors mis à jour en calculant la moyenne des attributs de chaque cluster selon la formule suivante (1.6).

$$g_{j \in \{1 \dots d\}} = W_j V_j \quad 1.6$$

Avec $W_j = 1 \times \frac{1}{n_j}$ est le vecteur de transformation appliqué à l'ensemble de n_j attributs $V_j \in C_j$

Algorithme FECM de base

Require : $X \in R^{LD}$: matrice contenant les observations,

ε : Seuil,

C : nombre d'attributs initial

1 : Construire la matrice $V \in R^{LD}$ contenant les attributs originaux. Chaque ligne représente un vecteur attribut.

2 : extraire la tendance de chaque vecteur attribut v_i , de la matrice V

3 : calculer le signe de la tendance de chaque attribut : déterminer $sign(dv_i/dx)$

4 : calculer la probabilité conditionnelle $p(v_i/(\alpha, \beta))$ pour chaque couple (α, β)

5 : calculer la matrice de distance M , avec $M(i, j) = d(v_i, v_j)$ tel que $i, j \in 1 \dots D$.

6 : ordonner la matrice M dans l'ordre croissant

7 : initialiser le premier cluster $C_{k=1}, k = 1 \dots d$, avec le premier attribut v_1 dans la matrice V et mettre l'indice $index = 1$

8 : fixer le nombre final des clusters $d, d = D/\varepsilon$

9 : construire les clusters :

10 : sélectionner les ε attributs classés en premiers par rapport à l'attribut v_{index} dans la ligne numéro $index$ dans la matrice M .

11 : mettre les attributs sélectionnés dans le cluster C_k

12 : calculer le centroid g_k correspondant au cluster C_k .

13 : mettre à jour le nombre d'attributs $C \leftarrow C - \varepsilon$

14 : mettre à jour l'indice $index \leftarrow index + \varepsilon + 1$

15 : déterminer les attributs communs entre les clusters $\forall (k_1, k_2) \in \{1 \dots d\}$ intersection entre clusters

16 : attribuer chacun des attributs communs au cluster le plus proche selon la distance euclidienne

17 : mettre à jour les centres g_k des clusters

18 : Return les nouveaux attributs sont l'ensemble des centres g_k
19 : Return la transformation de la matrice de donnée $T(X) \in R^{dL}$

Nous avons fait face à un problème de chevauchement des clusters qui consiste à avoir un même attribut dans une partie commune entre plusieurs classes (clusters). Pour remédier à ce problème, nous avons appliqué une procédure d'intersection entre chaque pair de clusters. Dans cette procédure nous attribuons un attribut commun au centre le plus proche.

9. Information mutuelle pour l'extraction d'attributs

Après avoir achevé la phase de classification dans notre algorithme, chaque groupe d'attributs (cluster) sera remplacé par son centre déterminé par une transformation des vecteurs membres de la classe. Cette transformation appliquée sur les attributs constituant le cluster est réellement l'extraction de caractéristique voulue par ce travail. Les centres des clusters sont donc la partie la plus importante. Ils sont souvent déterminés par un calcul de la moyenne des coordonnées des observations. Ces méthodes classiques de calcul du centre ne prennent pas en compte l'interdépendance des variables (attributs). Nous avons alors travaillé sur l'application de l'Information Mutuelle (IM) en tant que critère de recherche du centre du cluster. Dans ce présent mémoire, je présenterais le principe de cette mesure utilisée dans la théorie de l'information. Puis, dans une deuxième section, je proposerai les détails de l'application de cette mesure pour l'extraction d'attributs.

9.1 Principe de l'Information Mutuelle

L'Information Mutuelle est une mesure quantitative utilisée dans la théorie de l'information. Introduite pour la première fois par Shannon [80] en 1948, cette notion est une mesure de réduction d'incertitude d'une variable au vu de la connaissance d'une autre. Pour deux variables aléatoires, l'information mutuelle est une quantité mesurant leur dépendance statistique. L'information Mutuelle de deux variables discrètes X et Y est donnée par l'équation suivante :

$$I(X, Y) = \sum_x \sum_y P(X, Y) \ln \frac{P(X, Y)}{P(X)P(Y)} \quad 1.7$$

Les sommes dans l'équation 1.7 sont remplacées par des intégrales dans le cas continu et on aura l'équation suivante (1.8) :

$$I(X, Y) = \int_x \int_y P(X, Y) \log \frac{P(X, Y)}{P(X)P(Y)} dx dy \quad 1.8$$

Deux variables sont liées si leur Information Mutuelle MI est grande et dans le cas contraire (deux variables non liées) leur MI est petite.

Nous pouvons citer quelques propriétés de l'IM :

- $I(X, Y) = 0$ si et seulement si X et Y sont des variables aléatoires indépendantes.
- L'Information Mutuelle est positive ou nulle : $I(X, Y) \geq 0$
- L'information Mutuelle est symétrique : $I(X, Y) = I(Y, X)$

Il est difficile de déterminer les densités de probabilité dans le cas continu. Ainsi, pour le calcul de l'information mutuelle on utilise généralement le cas discret en divisant l'espace d'attributs en plusieurs partitions discrètes. Néanmoins il faut prendre en considération l'erreur inhérente dans le cas de variables continues.

Plusieurs méthodes d'estimation non paramétrique de l'IM ont été proposées par des chercheurs notamment l'approche d'estimation par histogramme [81], l'estimation par noyau [82] et l'estimation par fenêtre de Parzen [83] [84].

La première méthode d'estimation par histogramme procède par partitionnement des axes en différentes barres (ou classes) de largeurs différentes w_i et la probabilité normalisée est obtenue par la formule suivante : $P_i = \frac{n_i}{NW_i}$ avec n_i le nombre d'observation dans la classe i .

La fonction de probabilité est ainsi constante le long de la largeur de chaque barre. D'autres dérivées de cette approche ont été proposées. Elles font usage d'un ensemble d'histogrammes avec différentes largeurs de barres ou bien un partitionnement adaptatif du plan XY [85]. Pour la première dérivée, L'IM est calculée à partir de la moyenne des valeurs estimées pour chaque barre. La largeur de barre de l'histogramme est variable en fonction des données. Pour la deuxième dérivée, le plan est partitionné en divisant chaque axe en plusieurs segments équiprobable. La deuxième approche (estimation par noyau) estime la fonction de densité de probabilité de X par la superposition de fonctions noyaux, centrés sur les points de données.

9.2 Estimation par fenêtre de Parzen

L'estimation de densité par fenêtre de Parzen est une troisième approche qui consiste à utiliser l'entropie de Renyi combinée avec la méthode de la fenêtre de Parzen [86].

L'entropie de Renyi est une famille de mesures paramétriques qui inclue l'entropie de Shannon comme un cas particulier. Dans le cas d'une variable discrète C et une variable continue Y l'entropie de Renyi d'ordre α est définie par :

$$H_{R_\alpha}(C) = \frac{1}{1-\alpha} \log \sum_c p(c)^\alpha \quad 1.9$$

$$H_{R_\alpha}(Y) = \frac{1}{1-\alpha} \int_y p(y)^\alpha dy \quad 1.10$$

Avec $\alpha > 0$, $\alpha \neq 1$, $\lim_{\alpha \rightarrow 1} H_{R_\alpha} = H$. Dans notre cas, nous considérons le cas de la mesure quadratique $\alpha=2$.

La méthode de fenêtre de Parzen est l'une des méthodes non paramétriques pour estimer la densité de probabilité des variables. Cette méthode consiste à attribuer une fonction noyau (kernel) à chaque échantillon et elle estime la densité par la somme de ces fonctions noyaux. Ainsi, il s'avère que la combinaison de la mesure quadratique de Renyi et l'estimation de densité par la méthode de Parzen avec des noyaux gaussiens assure une économie significative en termes de temps de calcul [87].

Soit la fonction noyau suivante :

$$G(y, Cov) = \frac{1}{(2\pi)^{\frac{d}{2}} |Cov|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} y^T Cov^{-1} y\right) \quad 1.11$$

Avec y un échantillon aléatoire et Cov la matrice de covariance.

Pour deux fonctions noyau, on a une propriété très importante pour l'estimation de la densité. En effet, la convolution de deux Gaussiennes centrées en a_i et a_j est une Gaussienne centrée en $a_i - a_j$ avec une covariance égale à la somme des covariances originales tel comme l'indique la formule 1.12 :

$$\int_y G(y - a_i, Cov_1) G(y - a_j, Cov_2) dy = G(a_i - a_j, Cov_1 + Cov_2) \quad 1.12$$

Supposons maintenant que la densité de Y est estimée comme étant une somme de Gaussiennes sphériques centrée chacune en un échantillon y_i (estimateur de Parzen) :

$$P(Y) = \frac{1}{N} \sum_{i=1}^N G(y - y_i, \sigma I) \quad 1.13$$

Avec I une matrice unité. L'entropie quadratique de Renyi est donc donnée par :

$$\begin{aligned} H_{R_2}(Y) &= -\log \int_y p(y)^2 dy \\ &= -\log \frac{1}{N^2} \int_y \sum_{k=1}^N \sum_{j=1}^N G(y - y_k, \sigma I) G(y - y_j, \sigma I) dy \\ &= -\log \frac{1}{N^2} \int_y \sum_{k=1}^N \sum_{j=1}^N G(y_k - y_j, 2\sigma I) dy \end{aligned} \quad 1.14$$

L'entropie quadratique de Renyi est ainsi estimée par la somme d'interactions locales, le long de toutes les paires d'exemples. En pratique, seulement la moitié sera évaluée grâce à la symétrie.

9.3 Information mutuelle quadratique

L'entropie d'une fonction multivariée de densité peut être calculée d'une façon non paramétrique à partir de l'ensemble des données. Elle est perçue comme étant une interaction entre les paires des échantillons. Ceci peut être étendu au cas de l'information mutuelle entre les variables. Ainsi pour appliquer l'équation 1.12, l'information mutuelle doit être exprimée comme une fonction de la densité des variables au carré, ou bien sous une forme de produit de convolution de deux fonctions kernels. Pour trouver une distribution qui maximise/minimise la divergence sans calculer la valeur absolue de la divergence, les axiomes utilisés pour dériver la mesure peuvent être relaxés et le résultat de l'optimisation sera toujours la même distribution [88]. Les auteurs de [88] présentent un large nombre de telles mesures pour deux distributions discrètes P et Q , parmi lesquelles il y a la mesure suivante :

$$D(P, Q) = \frac{1}{\alpha(\alpha - 1)} \sum_{i=1}^N [p_i^\alpha - \alpha p_i q_i^{\alpha-1} + (\alpha - 1) q_i^\alpha], \alpha \neq 0, \alpha \neq 1 \quad 1.15$$

Pour $\alpha=2$, on peut étendre la mesure de l'équation précédente (1.15) au cas de densité continue et on aura :

$$D(f, g) = \int_y (f(x) - g(x))^2 dx \quad 1.16$$

Cette équation (1.16) avec les équations 1.7 et 1.8 conduisent à une formulation de la mesure d'information mutuelle quadratique I_T entre deux variables continues Y_1 et Y_2 .

$$I_T = \int \int (p(y_1, y_2) - p(y_1)p(y_2))^2 dy_1 dy_2 \quad 1.17$$

9.4 Information Mutuelle pour la sélection et l'extraction d'attributs

L'information mutuelle est un concept très utilisé en théorie de l'information pour la détermination et la mesure de toute dépendance linéaire ou non linéaire entre deux variables aléatoires. Ce concept a été largement appliqué dans des travaux de recherche sur la sélection des variables. Il a été utilisé, par exemple, comme un critère pour filtrer des variables [89], [90], [91]. Ces travaux ont appliqué des méthodes d'estimation paramétrique de cette mesure (IM) basées sur les histogrammes.

L'information mutuelle n'a été utilisée que dans le cadre de la classification supervisée pour l'extraction des caractéristiques. Le but étant de chercher les directions qui maximisent la mesure IM entre l'ensemble des variables et les labels des classes [92], [93], [94], [95]. Ces travaux ont appliqué des estimateurs différents pour le calcul de cette mesure et la réduction de la complexité [94]. Ainsi certains chercheurs ont combiné l'entropie de Renyi avec un estimateur non paramétrique de Parzen [92]. D'autres ont utilisé l'estimateur de Parzen pour calculer directement l'IM [91].

9.5 Approche proposée pour l'extraction

9.5.1 Formulation basée sur IM

Après avoir déterminé un cluster d'attributs avec l'algorithme de classification FECM de la section 1.8, nous devons trouver le centre de chaque classe. L'objectif de l'approche proposée dans cette section est donc la détermination du centre. En général, le centre est calculé à travers une transformation linéaire sur les données pour dériver un barycentre

$$G_j = W_j V_j = \text{avec } W_j = 1 \frac{1}{n_j} \quad 1.18$$

Avec n_j le nombre d'attributs dans la classe C_j concernée et V_j est la matrice contenant les attributs appartenant à ce cluster.

Notre procédure d'extraction consiste à utiliser les barycentres des clusters obtenus comme étant les nouveaux attributs ou descripteurs. Chaque cluster sera représenté par son barycentre. Ainsi, il devient délicat de déterminer la transformation la plus adéquate qui permet de trouver les centres les plus représentatifs. L'information mutuelle semble être une mesure appropriée pour cet objectif. Elle préserve les dépendances linéaire et non linéaire entre les variables (les attributs).

Nous avons donc appliqué cette mesure dont le critère adopté est sa maximisation entre chaque attribut dans un cluster et son centroïde correspondant. Nous avons défini une transformation que nous avons appelé f et que nous avons appliquée pour chaque attribut $V_i \in C_j$ de l'ensemble d'attributs V_j de C_j pour obtenir le centre $g_{ij} = f(w, v_i)$ et qui maximise $I(g_j, V_j)$.

Comme l'indique la Figure 1-13, le principe consiste à appliquer une procédure de descente de gradient pour chercher les poids w tel que $g=wV$ qui maximise l'information mutuelle $I(g, V)$ entre le centre g et les attributs d'un cluster V .

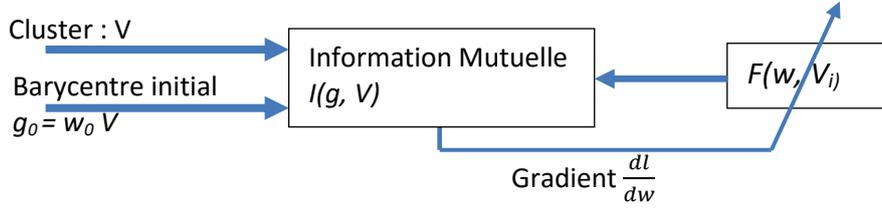


FIGURE 1-13: PROCEDURE DE MAXIMISATION DE L'IM ENTRE LES ATTRIBUTS ET LEUR CENTRE.

A partir de l'équation 1.17 on obtient :

$$I(g_j, V_j) = \int \int (p(g_j, v) - p(g_j)p(v))^2 dg_j dv \quad 1.19$$

Nous devons déterminer $p(g_j, v)$ pour pouvoir calculer l'information mutuelle. Puisque le centre recherché $g_j = f(w, v_j)$ appartient au même cluster C_j que les attributs, le cluster final $C_j^{final} = V_j \cup g_j$. La densité de probabilité jointe sera exprimée par l'équation 1.20 :

$$p(g_j, v) = p(g_j)p(g_j/v) = p(g_j)(p(v) - p(g_j)) \quad 1.20$$

Les deux dernières équations (1.19 et 1.20) nous donnent une équation plus simple de l'information mutuelle :

$$I(g_j, V_j) = \int \int (p(g_j))^4 dg_j dv \quad 1.21$$

Après application de l'estimation par la méthode de la fenêtre de Parzen, La densité de probabilité $p(g_j)$ de l'ensemble d'attributs du cluster final C_j^{final} comptant $n_j + 1$ attributs, est donnée par la formule suivante :

$$\frac{1}{n+1} \sum_{i=1}^{n+1} \phi(g_j - v_j, \sigma I) \quad 1.22$$

On note $I(g_j, V_j) = I$ et $g_j = y$ et $v = x$, pour simplifier l'écriture, et on sait que

$$\int Z_c N_x(\mu_c, \Sigma_s) dx = z_c$$

Ainsi on aura la IM donnée par :

$$\begin{aligned} I &= \int \int \frac{1}{(n+1)^4} \left(\sum_{i=1}^{n+1} \phi(y - y_i, \sigma I) \right)^4 dy dx \\ &= \frac{1}{(n+1)^4} \int \int \sum_{i=1}^{n+1} \phi(y - y_i, \sigma I) \sum_{j=1}^{n+1} \phi(y - y_j, \sigma I) \sum_{k=1}^{n+1} \phi(y - y_k, \sigma I) \sum_{l=1}^{n+1} \phi(y - y_l, \sigma I) dy dx \\ &= \frac{1}{(n+1)^4} \int \int \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} \sum_{l=1}^{n+1} \phi(y - y_i, \sigma I) \phi(y - y_j, \sigma I) \phi(y - y_k, \sigma I) \phi(y - y_l, \sigma I) dy dx \\ &= \frac{1}{(n+1)^4} \int \int \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} \sum_{l=1}^{n+1} \prod_{s=1}^4 \sigma_y(\mu_s, \Sigma_s) dy dx \end{aligned} \quad 1.23$$

$$= \frac{1}{(n+1)^4} \int \int \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} \sum_{l=1}^{n+1} \hat{z} \, dx$$

L'expression de l'information mutuelle devient alors :

$$\begin{aligned} I(g, V) &= \int_g \int_v \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} \sum_{l=1}^{n+1} \prod_{s=1}^{n+1} 4\phi_g(\mu_s, \Sigma_s) \, dg \, dv \\ &= \int_g \int_v \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} \sum_{l=1}^{n+1} \prod_{s=1}^{n+1} z \phi_g(\mu, \Sigma) \, dg \, dv \\ &= \int_v \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} \sum_{l=1}^{n+1} z \, dv \end{aligned} \quad 1.24$$

Avec

$$\begin{aligned} z &= \frac{|2\pi\Sigma_d|^{-\frac{1}{2}}}{\prod_{s=1}^4 |2\pi\Sigma_d|^{-\frac{1}{2}}} \prod_{a < b} \exp\left(-\frac{1}{2}(\mu_a - \mu_b)^T B_{ab}(\mu_a - \mu_b)\right), \\ \Sigma_d &= \left(\sum_{s=1}^{n+1} 4\Sigma_s^{-1}\right)^{-1}, \\ \Sigma_s &= \sigma^2 I, \\ B_{ab} &= \Sigma_s^{-1} \Sigma_d \Sigma_s^{-1} \end{aligned} \quad 1.25$$

Dans la Figure 1-13, nous maximisons $MI = I(g, V)$ pour obtenir une transformation optimale f^* et la matrice des poids optimale W^* et déterminer le centre g final. Nous avons utilisé un algorithme de descente de gradient pour trouver la solution. Selon la Figure 1-13, la descente de gradient a besoin de calculer le gradient $\frac{dI}{dw}$ ce qui induit la différentiation de $I(g, V)$. Rappelons que $g = WV$, on obtient alors la forme suivante :

$$\frac{dI}{dw} = \frac{dI}{dg} \frac{dg}{dw} = \frac{dI}{dg} V \quad 1.26$$

9.5.2 Minimisation de l'IM par descente de gradient

La procédure de descente de gradient procède par dériver I par rapport à w ce qui mène à calculer $\frac{dI}{dg}$ selon l'équation 1.24. Sachant que g est l'élément $(n+1)$ de l'ensemble V , on a donc : $\frac{dI}{dg} = \frac{dI}{dv_{n+1}}$.

On calcule alors la quantité :

$$\frac{dI}{dv_{n+1}} = \frac{dQ_1}{dv_{n+1}} + \frac{dQ_2}{dv_{n+1}} + \frac{dQ_3}{dv_{n+1}} + \frac{dQ_4}{dv_{n+1}} + \frac{dQ_5}{dv_{n+1}} \quad 1.27$$

Les quantités Q_1, Q_2, Q_3, Q_4 et Q_5 sont données par :

$$Q_1 = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n q(l, k, j, i) \quad 1.28$$

$$Q_2 = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n q(l = n + 1, k, j, i) \quad 1.29$$

$$Q_3 = \sum_{i=1}^n \sum_{j=1}^n q(l = n + 1, k = n + 1, j, i) \quad 1.30$$

$$Q_4 = \sum_{i=1}^n q(l = n + 1, k = n + 1, j = n + 1, i) \quad 1.31$$

$$Q_5 = q(l = n + 1, k = n + 1, j = n + 1, i = n + 1) \quad 1.32$$

Avec

$$\begin{aligned} q(l, k, j, i) = & \exp\left((v_i - v_j)^T B(v_i - v_j)\right) + (v_i - v_k)^T B(v_i - v_k) \\ & + (v_i - v_l)^T B(v_i - v_l) + (v_j - v_k)^T B(v_j - v_k) \\ & + (v_j - v_l)^T B(v_j - v_l) + (v_k - v_l)^T B(v_k - v_l) \end{aligned} \quad 1.33$$

Ainsi, on retrouve

$$\frac{dQ_1}{dv_{n+1}} = 0 \quad 1.34$$

$$\begin{aligned} \frac{dQ_2}{dv_{n+1}} &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \frac{dq(l, k, j, i)}{dv_{n+1}} = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{dq(l = n + 1, k, j, i)}{dv_{n+1}} \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \left(\frac{v_{n+1} - v_i}{\det(B)} + \frac{v_{n+1} - v_j}{\det(B)} + \frac{v_{n+1} - v_k}{\det(B)} \right) \exp(e_2) \end{aligned} \quad 1.35$$

$$\begin{aligned} \frac{dQ_3}{dv_{n+1}} &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \frac{dq(l, k, j, i)}{dv_{n+1}} \\ &= \sum_{i=1}^n \sum_{j=1}^n \left(\frac{2 * ((v_{n+1} - v_i) + (v_{n+1} - v_j))}{\det(B)} \right) \exp(e_3) \end{aligned} \quad 1.36$$

$$\begin{aligned} \frac{dQ_4}{dv_{n+1}} &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \frac{dq(l, k, j, i)}{dv_{n+1}} = \sum_{i=1}^n \frac{dq(l = n + 1, k = n + 1, j = n + 1, i)}{dv_{n+1}} \\ &= \sum_{i=1}^n 3 * (v_{n+1} - v_i) \exp(e_4) \end{aligned} \quad 1.37$$

$$\frac{dQ_5}{dv_{n+1}} = 0 \quad 1.38$$

Avec

$$\begin{aligned} e_2 = & (v_i - v_j)^T B(v_i - v_j) + (v_i - v_k)^T B(v_i - v_k) \\ & + (v_i - v_{n+1})^T B(v_i - v_{n+1}) + (v_j - v_k)^T B(v_j - v_k) \\ & + (v_j - v_{n+1})^T B(v_j - v_{n+1}) + (v_k - v_{n+1})^T B(v_k - v_{n+1}) \end{aligned} \quad 1.39$$

$$\begin{aligned} e_3 = & (v_i - v_j)^T B(v_i - v_j) + 2 * (v_i - v_{n+1})^T B(v_i - v_{n+1}) + 2 \\ & * (v_j - v_{n+1})^T B(v_j - v_{n+1}) \end{aligned} \quad 1.40$$

$$e_4 = 3 * (v_i - v_{n+1})^T B(v_i - v_{n+1}) \quad 1.41$$

$\frac{dl}{dv_{n+1}}$ est finalement donnée par l'expression suivante :

$$\begin{aligned} \frac{dl}{dv_{n+1}} = & 4 * \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{(v_{n+1} - v_i) + (v_{n+1} - v_j) + (v_{n+1} - v_k)}{\det(B)} \exp(e_2) \\ & + 6 * \sum_{i=1}^n \sum_{j=1}^n \frac{2 * ((v_{n+1} - v_i) + (v_{n+1} - v_j))}{\det(B)} \exp(e_3) \quad 1.42 \\ & + 4 * \sum_{i=1}^n (v_{n+1} - v_i) \exp(e_4) \end{aligned}$$

Pour simplifier les calculs au niveau de l'application de la descente du gradient, ce dernier ne sera calculé que sur un sous ensemble d'attributs dans chaque cluster et non pas sur tous les attributs de la classe.

9.5.3 Algorithme d'extraction avec IM

La première partie de ce deuxième algorithme d'extraction que nous avons proposé ressemble au premier algorithme présenté dans la section 1.8.3. Il s'agit d'une classification non supervisée des attributs. La deuxième partie consiste en une maximisation de l'IM pour la détermination de nouveaux centres optimaux. Ce module de recherche des centres pour chaque cluster applique la méthode de descente de gradient stochastique.

Pour commencer l'algorithme de maximisation de l'IM, un ensemble aléatoire d'attributs dans chaque cluster est choisi. Puis, On initialise le vecteur des poids w_o aléatoirement. La dérivée de l'IM est ensuite calculée pour déterminer la pente. Le poids suivant est ainsi calculé en fonction du poids précédent. Si la condition d'arrêt est vraie, l'algorithme s'arrête sinon il recherche un nouveau poids. Enfin, le poids optimal sera déterminé et le nouveau centroid est calculé.

Algorithme de FEMC avec IM

Require: $X \in R^{LD}$: matrice contenant les observations,

ε : Seuil,

C : nombre d'attributs initiaux

1: Construire la matrice $V \in R^{LD}$ qui contient les attributs originaux. Chaque ligne représente un vecteur attribut.

2: extraire la tendance de chaque vecteur attribut v_i de la matrice V

3: calculer le signe de la tendance de chaque attribut : déterminer $sign(dv_i/dx)$

4: calculer la probabilité conditionnelle $p(v_i/(\alpha, \beta))$ pour chaque possibilité $(\alpha, \beta) \in \{(1,0), (1,1), (1, -1), (-1,0), (-1,1), (-1, -1), (0,0), (0,1), (0, -1)\}$

5: calculer la matrice de distance M , avec $M(i, j) = d(v_i, v_j)$ tel que $i, j \in 1 \dots D$.

6: ordonner la matrice M dans l'ordre croissant

7: initialiser le premier cluster $C_{k=1}, k = 1 \dots d$, avec le premier attribut v_1 dans la matrice V et mettre l'indice $index=1$

8: fixer le nombre final des clusters $d, d = D/\varepsilon$

9: construire les clusters :

10: **While** $C \leq 1$ **do**

-
- 11** : sélectionner les ε attributs classés en premiers par rapport à l'attribut v_{index} dans la ligne numéro $index$ dans la matrice M .
 - 12** : mettre les attributs sélectionnés dans le cluster C_k
 - 13** : calculer le centroïde g_k correspondant au cluster C_k .
 - 14** : mettre à jour le nombre d'attributs $C \leftarrow C - \varepsilon$
 - 15** : mettre à jour l'indice $index \leftarrow index + \varepsilon + 1$
 - 16** : **End while**
 - 17** : déterminer les attributs communs entre les clusters $\forall (k_1, k_2) \in \{1 \dots d\}, C_{k_1} \cap C_{k_2}$: intersection entre clusters
 - 18** : attribuer chacun des attributs communs au cluster le plus proche selon la distance euclidienne
 - 19** : recherche des centres g_k des clusters par maximisation de MI : appeler algorithme de maximisation
 - 20** : **Return** les nouveaux attributs sont l'ensemble des centres g_k
 - 21** : **Return** la transformation de la matrice de donnée $T(X) \in R^{dL}$
-

Algorithme de maximisation de l'IM

- 1** : Choisir un sous ensemble aléatoire d'attributs dans chaque cluster
 - 2** : Initialiser la matrice des poids w_0
 - 3** : Calculer le dérivée de MI $\frac{dI}{dv_{n+1}}$ en fonction de $W(:,j)$ et, ci
 - 4** : Calculer le poids suivant $W(:,j+1) = W(:,j) + c * d'$
 - 5** : Tester si $norm(W(:,j+1) - W(:,j)) < delta$ donc break, sinon retour à l'étape **3**
 - 6** : déterminer le poids optimal $wopt = (W(:,end))$
 - 7** : **Return** le barycentre $bary = wopt * ci$
-

10. Expérimentations

10.1 Expérimentation sur les base UCI

Les algorithmes d'extraction précédemment présentés ont été validés sur différents problèmes de classification. Cette première expérimentation concerne des problèmes de classification issus des bases réelles prises au Centre d'apprentissage automatique et de systèmes intelligents (UCI machine learning repository³). Nous avons appliqué cette validation sur six bases différentes à savoir la base Sonar, la base Pima Indian Diabetes, la base Breast Cancer, la base Ionosphère, la base Wine et la base Iris.

Les tableaux suivants (Tableau 1-2, Tableau 1-3 et Tableau 1-4) présentent les résultats de nos algorithmes d'extraction pour les six bases. Comparés avec d'autres approches d'extraction tel que l'algorithme PCA et ICA, nos approches offrent des performances largement meilleurs pour certains cas. La validation a été appliquée via un classifieur SVM utilisant un noyau Gaussien avec les paramètres $\sigma=1$ et $C=1000$ pour la base Sonar. Avec une validation croisée, l'algorithme FECM arrive à des taux d'apprentissage meilleurs que les algorithmes PCA et ICA à part le cas de la dimension 1. Les résultats ont été améliorés avec l'usage de l'information Mutuelle. Même pour le cas de la dimension 1 les performances sont très proches de celles de l'ICA et mieux que le PCA. On note aussi que l'algorithme FECM arrive

³ <http://a.rchive.ics.uci.edu/ml/>.

au meilleur taux de bonne classification pour la base Sonar (85%) avec la dimension 12 alors que les deux autres algorithmes n'ont pas dépassé 75% pour la même dimension. De même, des résultats remarquables ont été achevés à travers notre approche pour la dimension 9. Ainsi, on constate que l'Information Mutuelle a pu améliorer la précision du classifieur surtout pour les faibles dimensions.

TABLEAU 1-2: PRECISION DE CLASSIFICATION SUR LA BASE SONAR

No. d'attributs	Précision de classification (SVM)			
	PCA	ICA	FEMC (sans IM)	FEMC (avec IM)
1	58.6	67.2	60.60	65.94
3	54.7	69.7	71.68	75.05
6	63.0	70.2	77.39	77.89
9	70.2	68.7	82.38	82.32
12	75.1	71.7	85.02	81.40
60	82.7			

Pour la base Pima, nous avons utilisé le même classifieur SVM avec $\sigma=10$ et $C=1000$. Cette base contient des données facilement séparables. Nous avons aussi appliqué une validation croisée. Tel que présentés dans le tableau suivant (Tableau 1-3), les taux d'apprentissage obtenus suite à une extraction avec FECM sont meilleurs pour la dimension 5 et ils arrivent au meilleur taux (82%). Pour les dimensions 1 et 2, l'algorithme ICA est meilleur. On note aussi que les résultats ont été améliorés avec l'application de l'IM.

TABLEAU 1-3: PRECISION DE CLASSIFICATION SUR LA BASE PIMA

No. d'attributs	Précision de classification (SVM)			
	PCA	ICA	FEMC (sans IM)	FEMC (avec IM)
1	66.3	73.2	68.08	67.80
2	75.1	76.7	69.72	69.85
3	75.5	76.8	75.72	75.72
5	75.5	77.2	76.80	82.32
8	78.0			

La base Breast Cancer comporte des données manquantes, qui ont été remplacées par la moyenne des attributs correspondant. Les valeurs prises par les attributs dans ces observations ne sont pas précises et donc les attributs sont légèrement bruités. Nous avons utilisé un classifieur SVM à noyau gaussien avec $\sigma=0.01$. Pour 1, 2 ou 3 attributs, les résultats de l'algorithme FECM dans ces deux versions sont meilleurs ou proches des résultats des autres algorithmes. Comme dans les autres bases, le meilleur résultat de classification est donné par notre approche avec IM.

TABLEAU 1-4: PRECISION DE CLASSIFICATION SUR LA BASE BREAST CANCER

No. d'attributs	Précision de classification (SVM)			
	PCA	ICA	FEMC (sans IM)	FEMC (avec IM)
1	85.8	85.1	96.72	96.86
2	94.7	90.3	96.57	96.71
3	95.9	91.3	94.71	94.28
6	96.6	94.3	85.11	85.11
9	96.6			

On peut constater à travers les tableaux suivants (Tableau 1-5, Tableau 1-6 et Tableau 1-7) les mêmes performances de l'algorithme FECM dans ces deux version (avec et sans IM) pour les différents bases UCI. Comme on peut le constater dans les tableaux suivants, l'algorithme

donne ses meilleures performances pour un nombre d'attributs moyen. Ils dépassent l'algorithme PCA et l'algorithme ICA dans la plupart des cas.

TABLEAU 1-5: PRECISION DE CLASSIFICATION SUR LA BASE IONOSPHERE

No. d'attributs	Précision de classification (SVM)			
	PCA	ICA	FEMC (sans IM)	FEMC (avec IM)
1	64.07	61.28	72.09	76.09
3	85.21	81.80	75.20	83.50
6	84.79	86.05	85.21	85.49
9	84.83	86.52	87.52	87.80
12	86.31	88.04	89.20	89.20
34	91.73			

TABLEAU 1-6: PRECISION DE CLASSIFICATION SUR LA BASE WINE

No. d'attributs	Précision de classification (KNN)			
	PCA	ICA	FEMC (sans IM)	FEMC (avec IM)
1	67.93	67.42	71.02	71.88
2	71.94	73.84	72.36	76.25
3	72.48	75.22	78.19	81.49
5	75.74	91.01	79.92	80.45
13	80.27			

TABLEAU 1-7: PRECISION DE CLASSIFICATION SUR LA BASE IRIS

No. d'attributs	Précision de classification (KNN)			
	PCA	ICA	FEMC (sans IM)	FEMC (avec IM)
1	80.0	88.67	87.33	89.33
2	94.67	95.33	97.33	96.67
3	96.0	96.0	98.0	98.0
4	96.0			

10.2 Extraction pour la reconnaissance de visage

La reconnaissance de visage est l'un des problèmes de reconnaissance de formes les plus étudiés. L'application d'une approche d'extraction pour ce problème est très importante puisque la dimension des images est relativement très grande par rapport aux autres données dans d'autres domaines. En plus de l'amélioration des performances, l'extraction des caractéristiques peut être aussi utile pour la compression de l'image. Plusieurs méthodes ont été appliquées pour ce problème.

Parmi les plus importants en terme de résultats on peut citer les méthodes Eigenfaces (utilisant l'ACP) et Fisherfaces (basée sur LDA) [96]. Ces deux approches permettent une caractérisation effective d'un espace de faible dimension tout en préservant la qualité perceptive d'une image de très grande dimension. La méthode ICA a aussi réussi à donner de bonnes performances surtout pour la reconnaissance de visage avec des changements en pose et pour la classification des actions faciales [97].

Pour ce problème de reconnaissance de visage, nous avons validé notre approche FECM sur deux bases différentes à savoir la base Yale et la base ORL. La première base (Yale) contient un ensemble d'images de 15 visages de personnes différentes. Pour chaque personne il y a 11 images selon des conditions différentes : de lumière, d'expressions du visage (yeux ouverts/yeux fermés, avec/sans sourire...) et de détails de visages (avec/sans lunettes). Cela fait un total de 165 images. La deuxième base (ORL : Olivetti Research Laboratory) contient un

ensemble d'images de 40 visages de personnes différentes avec 10 images pour chaque personne, cela fait un total de 400 images.

Les résultats dans les deux tableaux suivants (Tableau 1-8 et Tableau 1-9) montrent des résultats de notre algorithme FECM. Comparée aux autres approches à savoir KNN, PCA, ICA et LDA, la méthode a montré des performances intéressantes pour les deux bases. Pourtant dépassé en taux d'apprentissage par la méthode LDA (21% contre 8.48%), elle a donné le même nombre d'attributs 14 pour la première base tout en gardant un taux d'erreur moins que les autres méthodes. Pour la base ORL, elle a été moins bien en ce qui concerne le taux d'apprentissage avec 5% mais un nombre d'attributs beaucoup plus petit que les autres approches.

TABLEAU 1-8: PRECISION DE CLASSIFICATION SUR LA BASE YALE

Méthode	Taux d'erreur	Nombre d'attributs
KNN	21.82	783
PCA	24.85	30
ICA	23.03	30
LDA	8.48	14
FECM	21.00	14

TABLEAU 1-9: PRECISION DE CLASSIFICATION SUR LA BASE ORL

Méthode	Taux d'erreur	Nombre d'attributs
KNN	3.00	952
PCA	4.00	40
ICA	4.25	40
LDA	2.00	39
FECM	5.00	23

10.3 Extraction pour le diagnostic d'un procédé industriel

Le Tennessee Eastman Process (TEP) est un procédé qui a été conçu par la société Eastman Chemical Company, afin de fournir une simulation d'un procédé chimique industriel réel. Il sert à tester des méthodes d'asservissements et/ou de surveillance de procédé. Le TEP a été très utilisé pour comparer certaines méthodes de surveillance et le diagnostic des défauts [98], [99], [100]. Il est composé de cinq éléments principaux : un réacteur, un compresseur, un décapeur, un séparateur et un condenseur. Il comporte 53 variables en total dont 12 variables d'asservissement et 41 variables mesurables. Parmi les 41 variables mesurables, il y a 22 qui sont des variables mesurables en continu (ce sont les valeurs des capteurs du procédé). Les autres variables sont des mesures de compositions telles que des concentrations.

Tel que traité par plusieurs chercheurs, notamment ceux cités précédemment, le problème de diagnostic de ce procédé peut être vu en tant que problème de classification. Les 20 fautes à lesquelles il peut être soumis sont modélisées en classes d'observations. Ces fautes sont de diverses natures (saut en échelon de certaines variables internes, augmentation de la variabilité de certaines autres, ou bien fautes d'actionneurs tel qu'une vanne bloquée).

En se basant sur une base d'expertise comportant des observations de fonctionnement normal et d'autres pour chaque faute, notre approche FECM ainsi que d'autres méthodes ont été appliquées pour extraire les variables pertinentes à chaque situation. L'extraction permettra surtout de réduire la complexité. Ainsi, les approches de classifications seront aussi utiles pour le diagnostic en line. En comparant les résultats de la classification avant (Tableau 1-10) et après l'extraction (Tableau 1-11), on note que cette dernière a considérablement amélioré la classification. Indépendamment de l'approche de classification utilisée, les taux d'erreur ont

baissé sauf pour le KNN. Ces baisses ont atteint parfois 20% comme pour le LDA. L'extraction des attributs a permis aussi de réussir la classification avec uniquement 4 attributs au lieu de 52.

TABLEAU 1-10: TAUX D'ERREUR DE CLASSIFICATION SANS EXTRACTION

Taux d'erreur (%)			
SVM	KNN	LDA	QDA
52.00	35.38	68.83	47.5

TABLEAU 1-11: TAUX D'ERREUR DE CLASSIFICATION AVEC EXTRACTION

Nombre d'attributs	Taux d'erreur (%)			
	SVM	KNN	LDA	QDA
21	53.96	59.25	49.67	43.42
10	49.17	58.88	47.50	35.50
4	44.20	54.50	48.17	45.79

Conclusion

Identifier les attributs redondants, nous a permis de réduire la dimension des observations sans perte d'information importante ni de pouvoir discriminant. Nous avons proposé, dans un premier travail, une nouvelle approche de sélection de variables pour résoudre ce problème de redondance d'attributs. Cette méthode, testée sur un problème de classification des perturbations dans un réseau de transport, est de type Wrapper. Elle combine une méthode d'optimisation avec l'algorithme SVM. Deux algorithmes d'optimisation à savoir la méthode Tabou et l'algorithme génétique ont été, successivement, testée pour cette approche. Les résultats présentés dans ce chapitre sont intéressants surtout en les comparant aux autres approches de sélection de variables. Les travaux réalisés dans ce premier axe ont été publiés depuis 2009 [5] et ont fait l'objet d'un Master de recherche en collaboration avec l'ENIT [6].

La deuxième partie de ce chapitre concerne l'extraction des caractéristiques. Un sujet en lien avec le premier axe et qui a fait l'objet de la continuité de cette collaboration avec l'ENIT. Nous l'avons publié depuis 2011 [7] [79] [101] [102] et a fait l'objet des travaux de thèse de Sabra [1]. Nous avons proposé une approche basée sur une classification non supervisée des attributs. Nous avons fait appel à une mesure de similarité spécifique pour cette classification. Cette mesure nous a permis de regrouper les attributs qui ont les mêmes comportements d'une classe à l'autre. Les attributs seront remplacés par la suite par le centre de la classe. Dans une deuxième version de cet algorithme, nous avons appliqué l'Information Mutuelle pour le calcul du centre. Les résultats tels qu'ils sont présentés dans la section 10 sont aussi intéressants.

Je continue à travailler sur l'extraction et la sélection de variables avec plus d'exploration pour les domaines de type spécifiques tel que les données textuelles. Ce type de données non scalaire n'est pas exploitable à l'état brut dans le processus de classification. Ainsi, nous sommes amenés à appliquer des approches d'extraction de caractéristiques ou préalable du processus de classification ou d'apprentissage. L'un des axes de recherche dans mes perspectives est l'extraction textuelle. Une problématique que j'ai commencée à explorer pour l'analyse automatique de textes dans le domaine médical.

Chapitre 2 : Apprentissage et classification pour l'identification et le diagnostic des systèmes automatiques complexes

1. Introduction

Les nouveaux systèmes industriels sont de plus en plus complexes. En effet, outre leurs comportements conditionnés souvent par plusieurs paramètres dont une partie aléatoire, ces nouveaux systèmes sont composés de plusieurs procédés physiques et des processus numériques, que le progrès technologique a imposés. L'interaction de plus en plus importante entre ces procédés physiques (modèles d'évolution impliquant des signaux continus) et les processus numériques (ordinateurs, logiciels, composants logiques, etc.) a conduit, en automatique, à l'apparition et à la construction des systèmes dits hybrides avec des phénomènes à la fois continue et événementielle.

Ces systèmes dynamiques hybrides et d'autres système plus simples mais avec des comportements compliqués, doivent avoir un fonctionnement automatique pour simplifier leur commande et leur surveillance. Il est alors important pour le contrôle de ces systèmes d'établir les modèles les plus fidèles aux comportements. Néanmoins, on est souvent confronté à des systèmes non linéaires dont la mise en place d'une représentation exige des nouvelles approches compatibles avec les nouveaux processus automatiques intelligents.

Les nouvelles techniques issues de l'intelligence artificielle ne cessent de prouver leur capacité d'être auto-adaptatives en imitant la nature mais avec une capacité de calcul plus importante. Ainsi, comme d'autres chercheurs, j'ai pensé à ces techniques pour l'identification et le diagnostic de ces systèmes automatiques. A travers des propositions validées avec des jeunes chercheurs en thèse et en mastère, je présente dans ce chapitre l'importance de l'une des techniques de l'apprentissage et classification pour cet objectif de modélisation des systèmes automatiques. Il s'agit de la théorie d'apprentissage statistique. L'une des forces des techniques de classification et de régression issues de cette théorie est l'usage des espaces de Hilbert à noyau reproduisant surtout pour les distributions non-linéaire. Ils s'agissent des espaces qui généralisent les notions de l'espace euclidien classique (de dimension 2 ou 3) sur des espaces de dimensions plus grandes (finie ou infinie). Munis d'un produit scalaire qui peut être remplacés par une fonction noyau, ces espaces sont capables de résoudre certaines difficultés de la classification non-linéaire.

Dans ce même contexte, je propose deux approches de régression pour l'identification des systèmes automatiques non linéaires, dans une première partie. Dans la deuxième partie de ce chapitre, je présente une méthode pour l'identification des systèmes dynamiques hybrides en gardant les mêmes fondements théoriques. Et je termine par une troisième partie sur le diagnostic et la classification dynamique de données non stationnaires.

2. SVM pour l'identification des systèmes non linéaire

2.1 De la modélisation mathématique..... aux nouvelles méthodes

Un processus automatique est un système physique qui évolue au cours du temps sous l'effet de différentes actions internes et externes. Il est caractérisé par des entrées, des sorties et des variables d'état. Pour le control et le fonctionnement automatique, tout processus doit être régit par un modèle. Il s'agit d'une représentation mathématique décrivant les relations existantes entre ses entrées et ses sorties. Il a pour but de reproduire partiellement ou totalement le comportement du processus. Selon les informations disponibles et le processus de leur conception, il existe trois types de modèles à savoir :

✓ Les modèles de connaissance (boite blanche) qui consistent à établir les équations mathématiques du système reliant les entrées aux sorties. Ces modèles sont complètement déduits des principes fondamentaux (lois physiques, chimiques, économiques, ...) qui régissent

le fonctionnement du processus. Malgré la richesse et la précision, l'établissement et la mise en œuvre de ces modèles restent difficiles.

✓ Les modèles mathématiques (boite noire) qui sont établis à partir d'un ensemble d'expérimentations (mesures d'entrées/sorties du système). Ces modèles nécessitent peu d'informations à priori sur le procédé tels que l'ordre et la structure du système. Les paramètres n'ont généralement pas de signification physique. Souple et capable de donner une expression mathématique fidèle au comportement du système, ce type de modèles est le plus préféré par les automaticiens.

✓ Les modèles semi physique (boite grise) mis en œuvre lorsque les connaissances disponibles sont insuffisantes pour concevoir un modèle de connaissance. Cette modélisation prend en considération à la fois les connaissances disponibles et les données mesurées du système.

Dans la suite de cette section, nous allons parler du deuxième type (les modèles mathématiques). En effet, les automaticiens sont souvent confrontés à une méconnaissance des informations sur le processus. Ainsi, pour la plupart des systèmes on parle de boite noir représentant le processus qui interagit avec le milieu extérieur par le biais des signaux d'entrée ou de commande $u(t)$ et des signaux de sortie ou de mesure $y(t)$. L'établissement de ce type de modèle est une procédure itérative comportant cinq phases [103], à savoir l'extraction de données, le choix de la structure du modèle, le choix du critère d'estimation paramétrique, l'estimation paramétrique et la validation du modèle.

Le peu d'informations ou connaissances à priori disponibles au départ combinées à des données expérimentales permettent de passer de la boite noire à une boite grise. À partir de là, on peut sélectionner la structure convenable du modèle. En parlant des structures, il y a plusieurs types de modèles sur lesquels on peut se baser pour établir une représentation mathématique fidèle au fonctionnement du système :

2.1.1 Modèles non linéaires

➤ *Modèle de Volterra-Wiener* : Il s'agit d'un modèle polynomial non récursif dont sa relation Entrée/sortie est :

$$y(k) = \sum_{i_1=1}^{\infty} \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \dots \sum_{j_N=0}^{\infty} h_i(j_1, \dots, j_N) \prod_{l=1}^N u(k - j_l) \quad 2.1$$

u , y , et h_i sont respectivement l'entrée, la sortie et le noyau de Volterra d'ordre N . Les $j_l, l = 1, \dots, N$ sont des entiers. Un noyau d'ordre N peut être vu comme une généralisation à l'ordre N de la notion de réponse impulsionnelle bien connue dans le cas linéaire. En appelant M la mémoire du système, on peut considérer le modèle tronqué, à temps discret, à l'ordre N suivant :

$$y(k) = \sum_{i_1=1}^N \sum_{j_1=0}^{M-1} \sum_{j_2=0}^{M-1} \dots \sum_{j_N=0}^{M-1} h_i(j_1, \dots, j_N) \prod_{l=1}^N u(k - j_l) \quad 2.2$$

Ce modèle est le plus ancien et il est caractérisé par la linéarité vis-à-vis de ses paramètres et par sa stabilité garantie au sens entrée bornée, sortie bornée [104]. A cause du nombre très élevé de paramètres, ce modèle n'a pas connu beaucoup de succès en identification. Et pour résoudre ce problème, plusieurs chercheurs ont proposé des solutions de réduction de complexité et de décomposition [103] [104] [105] .

➤ *Modèle de Wiener* : Il consiste à mettre en cascade un modèle dynamique linéaire suivi d'un bloc non linéaire statique. Si la non-linéarité statique est de forme polynomiale d'ordre N

et le bloc linéaire dynamique représente une réponse impulsionnelle finie FIR (Finite Impulse Response), la relation d'entrée/sortie de ce modèle sera comme suit :

$$w(k) = \sum_{j_1=0}^{M-1} h(j_1)u(k - j_1) \quad 2.3$$

$$y(k) = g[w(k)] = \sum_{i=0}^N c_i w^i(k) \quad 2.4$$

Plusieurs approches ont été proposées pour l'identification de ce type de modèle tels que les splines cubiques [106], les modèles locaux linéaires [107], les fonctions affines par morceaux [108] et les réseaux de neurones à fonctions de base radiales [109].

➤ *Modèle de Hammerstein* : Il peut être vu comme étant le dual du modèle de Wiener composé des mêmes blocs mais placés dans un ordre inversé. Si la non-linéarité statique est de forme polynomiale d'ordre N et le bloc linéaire dynamique représente une réponse impulsionnelle finie FIR (Finite Impulse Response), la relation d'entrée/sortie de ce modèle sera comme suit :

$$w(k) = g[u(k)] = \sum_{i=0}^N c_i u^i(k) \quad 2.5$$

$$y(k) = \sum_{j_1=0}^{M-1} h(j_1)w(k - j_1) \quad 2.6$$

Plusieurs méthodes ont été proposées par des chercheurs pour l'identification de ce type de modèle. On peut citer la théorie de corrélation [110], les réseaux neuronaux [111], les fonctions orthogonales, les polynômes, le modèle linéaire par morceaux, les réseaux de neurones à fonctions de base radiales et l'algorithme génétique [112].

➤ *Modèle Wiener-Hammerstein* : Pour ce modèle, le premier bloc linéaire dynamique a été fondu à un sous-système qui laisse passer une bande de fréquence bien définie, le bloc non linéaire statique a été remplacé par un redresseur et le second bloc linéaire dynamique par un filtre passe bas [113].

➤ *Modèle de Hammerstein-Wiener* : Ce modèle est le plus complexe, il contient deux sous modèles non linéaires et un sous modèle linéaire dynamique. Il est le moins utilisé par rapport aux trois modèles précédents.

➤ *Modèles flous* : Développée par Lotfi Zadeh en 1965 [114], la logique floue a été très utilisée pour la modélisation des systèmes automatique non linéaire. Il existe deux types de modèles flous : les modèles flous de Mamdani [115] et les modèles flous de Takagi-Sugeno (TS) [116]. Cette modélisation floue permet une représentation des systèmes complexes en se basant sur les imprécisions du raisonnement humain. Il est constitué d'un ensemble de règles comportant des variables linguistiques

➤ *Multi-modèles* : Il est évident qu'un modèle n'est valable que dans un domaine de fonctionnement précis. D'où vient l'idée de représenter un système automatique par plusieurs modèles. Habituellement, ces systèmes sont traités par linéarisation autour d'un point de fonctionnement. En se basant sur une bibliothèque de modèles, cette approche attribue un modèle à un mode en particulier d'une partie de la plage de fonctionnement [117] [118] [119].

2.1.2 Modèles dynamiques non linéaires

Les modèles dynamiques présentent des particularités dont la plus importante est la possibilité d'avoir plusieurs représentations d'un même processus. Il existe des modèles sans perturbation aléatoire (bruit) et d'autres avec perturbations aléatoires. Parmi les modèles les plus connus on peut citer :

➤ *Modèle NDARMA* : Ce modèle, appelé NDARMA (Non Linéaire Déterministe Auto-Regressif à Moyenne Ajustée), est un modèle hypothèse déterministe (sans bruit additif). Il est de la forme suivante [120]

$$y(k) = f(y_p(k - d_y - 1), \dots, y_p(k - d_y - n), u(k - d_u - 1), \dots, u(k - d_u - m)) \quad 2.7$$

Où f est une fonction non linéaire que l'on désire estimer.

➤ *Modèle NARX* : Le modèle NARX est le modèle Non linéaire Auto-régressif avec entrée eXogène. Il est non linéaire et il consiste à considérer qu'un bruit agit sur le processus [121]. Il est de la forme suivante :

$$y_p(k) = f[y_p(k - d_y - 1), \dots, y_p(k - d_y - n), u(k - d_u - 1), \dots, u(k - d_u - m)] + e(k) \quad 2.8$$

➤ *Modèle NARMAX* : NARMAX (Non linéaire Auto-Regressif à Moyenne Ajustée avec entrée eXogène) est le modèle non linéaire le plus général car il modélise les deux types de bruits (d'état et de sortie). Ce modèle est une extension non linéaire du modèle linéaire ARMAX. Il est de la forme :

$$y(k) = f(y_p(k - d_y - 1), \dots, y_p(k - d_y - n), u(k - d_u - 1), \dots, u(k - d_u - m), e(k - d_e - 1), \dots, e(k - d_e - p)) + e(k) \quad 2.9$$

➤ *Modèle NARMA* : On considère que les systèmes non linéaires peuvent être décrits par la relation suivante :

$$y(k) = g(X(k)) + e(k) \quad 2.10$$

où g est une fonction non linéaire inconnue $X(k) = [y(k - 1), \dots, y(k - n), u(k - 1), \dots, u(k - m)]$, les deux réelles m et n représentent, respectivement, l'ordre de l'entrée $u(k)$ et l'ordre de la sortie $y(k)$. La structure dynamique d'un modèle de type NARMA est donnée par l'équation discrète mathématique suivante :

$$\begin{aligned}
\hat{y}(k) = & \bar{y} + \sum_{i=1}^m b_i u(k-i) \\
& + \sum_{i=1}^m \sum_{j=1}^m b_{ij} u(k-i) u(k-j) + \dots \\
& + \sum_{i=1}^m \dots \sum_{v=p}^m \sum_{l=v}^m b_{i\dots l} u(k-i) \dots u(k-l) \\
& + \sum_{i=1}^n a_i y(k-i) + \sum_{i=1}^n \sum_{j=1}^n a_{ij} y(k-i) y(k-j) + \dots \\
& + \sum_{i=1}^n \dots \sum_{v=p}^n \sum_{l=v}^n a_{i\dots l} y(k-i) \dots y(k-l) \\
& + \sum_{i=1}^m \sum_{j=1}^n c_{ij} y(k-i) y(k-j) + \dots \\
& + \sum_{i=1}^m \dots \sum_{v=1}^m \sum_{l=1}^n c_{i\dots l} u(k-i) \dots u(k-v) y(k-l) + \dots \\
& + \sum_{i=1}^m \sum_{j=1}^n \dots \sum_{l=1}^n c_{ij\dots l} u(k-i) y(k-j) \dots y(k-l)
\end{aligned} \tag{2.11}$$

où \bar{y} présente la valeur moyenne de y .

2.1.3 ...Aux nouvelles approches de l'identification

Depuis les années 80, plusieurs approches issues de l'intelligence artificielle ont permis une évolution considérable en ce qui concerne la modélisation des systèmes automatiques complexes. Souvent basées sur des fondements mathématiques, ces nouvelles méthodes offrent une représentation fidèle du fonctionnement d'un processus. J'ai cité dans la section 2.1.1 la logique floue. Il s'agit de l'une de ces nouvelles approches qui a été largement appliquée pour la modélisation des systèmes non linéaire. Un autre axe qui a bien réussi pour ces problèmes est celui de la classification et l'apprentissage automatique. Avec une grande capacité d'adaptation, de généralisation et de traitement des informations, les Réseaux de Neurones Artificiels "RNAs" sont des modèles mathématiques inspirés du fonctionnement du système nerveux. Cette approche a été aussi utilisée pour l'identification des systèmes automatiques de différents types notamment les systèmes dynamiques [122].

Dans ce même contexte d'apprentissage, on peut citer une nouvelle approche basée sur la minimisation du risque structurel et qui ne cesse d'offrir des résultats intéressants. Il s'agit de la théorie de l'apprentissage statistique. Proposée par Vapnik en 1995 [61], cette technique fait intervenir des espaces d'hypothèse de nature particulière pour résoudre les problèmes d'apprentissage. Ces espaces d'hypothèse sont des espaces de Hilbert à noyau reproduisant (Reproducing Kernel Hilbert Space : RKHS). Dans ces espaces, le problème de la modélisation est formulé comme un problème de minimisation d'une fonction similaire à celle habituellement utilisée par les automaticiens. La solution à ce problème de minimisation est obtenue grâce au théorème du représentant [123]. Cette solution, qui est une combinaison linéaire des fonctions noyaux de l'espace RKHS, est linéaire par rapport aux paramètres qui sont les coefficients de cette combinaison. Le nombre de ces paramètres est égal au nombre des observations utilisées lors de l'apprentissage. La théorie de l'apprentissage statistique fournit des algorithmes adéquats pour l'identification de ces paramètres. Parmi ces algorithmes, nous citons les Séparateurs à Vaste Marges (SVM) (en anglais Support Vector Machines) qui sont aptes à modéliser des formes complexes sans aucune connaissance a priori sur la distribution des données [124]. Dans tous les domaines de l'apprentissage artificiel, ces méthodes dites encore à noyaux ont connu ces dernières années une explosion importante

[125]. Ces méthodes ont des fondements théoriques solides et offrent des résultats expérimentaux captivants dans la plupart des problèmes de classification et de régression.

Ce présent chapitre présente mes contributions dans ce contexte d'application de la théorie d'apprentissage statistique pour l'identification des systèmes automatiques complexes.

2.2 Théorie de l'apprentissage statistique

2.2.1 Pour la classification

Comme je viens de la présenter dans la section précédente, la théorie de l'apprentissage statistique est basée sur le principe de minimisation du risque empirique pour l'étude d'un modèle conceptuel. Proposée par Vapnik [126] comme étant un problème d'inférence statistique basée sur un nombre limité d'observations, elle a été utilisée pour la classification supervisée binaire. Comme le montre la figure suivante (Figure 2-14), les observations (données) en question appartiennent à deux classes différentes : Classe 1 et Classe -1. Classifier ces données consiste à déterminer une règle de décision capable, à partir d'observations externes, d'assigner un objet à une classe parmi ces deux classes. Dans le cadre de l'apprentissage statistique supervisée, le problème de classification se résume de la façon suivante : à partir d'un ensemble d'exemples $\{x_i, y_i\}_{i=1 \dots l}$, on veut construire une fonction $f : X \rightarrow \{-1, 1\}$ qui permette de prédire si un nouveau point $x \in X$ appartient à la classe (-1) ou à la classe (1) en minimisant le risque structurel R défini de la manière suivante :

$$R = \int \frac{1}{2} |y - f(x)| dP(x, y) \quad 2.12$$

Avec :

$P(x, y)$: la loi de probabilité qui est inconnue.

(x, y) : couple aléatoire de l'échantillon d'apprentissage souvent noté par :

$$(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L) \subseteq (X \times Y)^L$$

Puisque la probabilité de distribution P est inconnue, R l'est aussi. L'idée de Vapnik consiste à estimer le risque structurel R par le risque empirique dit aussi erreur d'apprentissage. Ainsi, il a démontré que le risque structurel peut être borné de la manière suivante :

$$R \leq R_{emp} + \sqrt{\frac{h(\log(\frac{2N}{h}) + 1 - \log(\frac{h}{4}))}{N}} \quad 2.13$$

R_{emp} est le risque empirique donné par :

$$R_{emp} = \frac{1}{2N} \sum_{i=1}^N |y_i - f(x_i)| \quad 2.14$$

Et h représente la dimension de Vapnik-Chervonenkis (VC).

Pour une classe de fonctions F en classification binaire cette dimension h est le cardinal maximal d'un sous-ensemble $A \subset X$ tel qu'on puisse toujours trouver une fonction $f \in F$ qui classe parfaitement tous les éléments de A quelles que soit leurs étiquettes. D'après Vapnik [126] cette dimension est égale à $P+1$ pour l'ensemble $F = \{x \rightarrow \text{sign}(\langle w, x \rangle + b), w \in \mathbb{R}^p, b \in \mathbb{R}\}$ des hyperplans dans \mathbb{R}^p .

Vapnik a proposé un algorithme de séparation de région appelé Support Vector Machines (SVM) qui consiste à trouver l'hyperplan optimal qui sépare les données de deux classes différentes.

✓ *Cas de la classification linéaire*

On suppose que nous avons les données empiriques suivantes : $(x_1, y_1) \dots (x_i, y_i) \dots (x_m, y_m) \in R \times \{-1, 1\}$. Dans le cas d'une classification linéaire, l'algorithme SVM calcule un hyperplan qui sépare au mieux les échantillons de deux classes. Dans ce cas, la fonction f est linéaire en x_i et elle prend la forme générale suivante : $f(x_i) = \langle w, x_i \rangle + b$. Comme on peut le voir sur la Figure 2-14, Il existe une infinité d'hyperplans capables de séparer ces données. Mais il y a un seul hyperplan optimal. Il s'agit de l'hyperplan qui passe "au milieu" des points des deux classes. Il vérifie la condition suivante : $y_i (wx_i + b) \geq 1$ pour $i=1 \dots m$

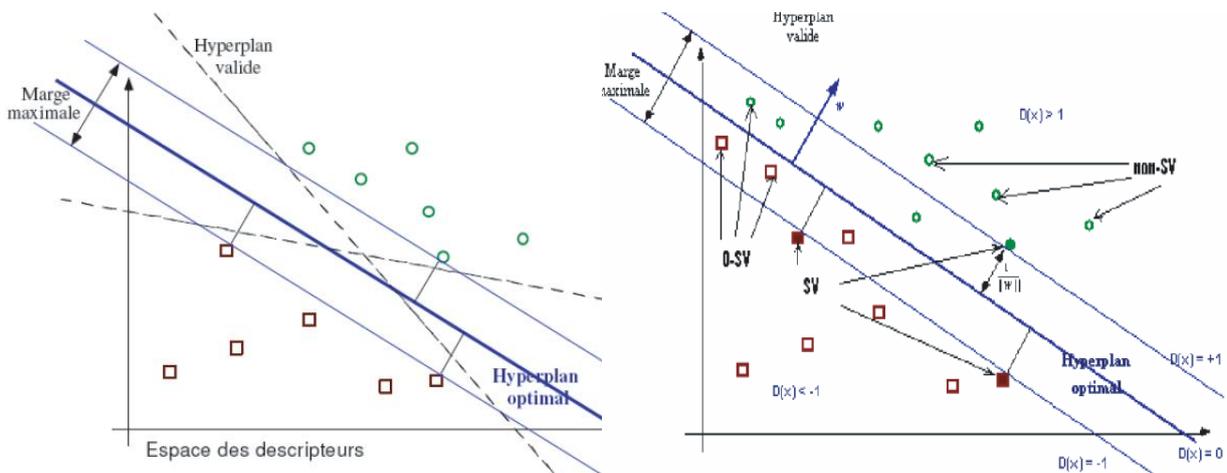


FIGURE 2-14: SEPARATION LINEAIRE.

La distance entre un point x_i et l'hyperplan est $D(x, H) = \frac{|\langle w, x \rangle + b|}{\|w\|^2}$. L'unique hyperplan optimal est celui qui maximise $M = \frac{2}{\|w\|}$ et donc maximiser $\|w\|^2$ ou bien $\frac{1}{2} \|w\|^2$. Cet hyperplan optimal est paramétré par w et b tel que :

$$\begin{cases} \min_{w,b} \frac{1}{2} \|w\|^2 \\ y_i (wx_i + b) \geq 1, \forall i \end{cases} \quad 2.15$$

Notre problème d'optimisation décrit par (2.15) est de la forme : $\min_{x \in \Omega} f(x)$ Sc $g_i(x) \leq 0$. A ce stade, on arrive à un problème d'optimisation quadratique convexe sous contraintes linéaires. Pour résoudre ce problème, on peut utiliser la méthode de lagrangien. Ainsi le lagrangien primaire devient :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (\langle w, x \rangle + b) - 1) \quad 2.16$$

Avec α_i appelés multiplicateurs de Lagrange associés aux contraintes.

Une condition nécessaire d'optimalité en x est la satisfaction des conditions de Karush-Kuhn-Tucker (KKT) :

$$\begin{cases} \frac{\partial L(x, \alpha, \beta)}{\partial x} = 0 \\ \frac{\partial L(x, \alpha, \beta)}{\partial \beta} = 0 \\ g_i(x) \leq 0 \\ \alpha_i \geq 0 \end{cases} \quad 2.17$$

En maximisant le lagrangien, notre problème primaire, décrit par le système (2.16), peut se traduire sous la forme :

$$\min_{w,b} \max_{\alpha_i} L(w, b, \alpha) = \max_{\alpha_i} \min_{w,b} L(w, b, \alpha) \quad 2.18$$

Notre problème étant convexe, alors la dualité est forte, d'où on peut écrire :

$$\max_{\alpha \geq 0} \min_{w,b} L(x, \alpha) \quad 2.19$$

En résolvant le système d'équations (2.17), on obtient les équations suivantes :

$$\sum_{i \in SV} \alpha_i y_i = 0 \text{ et } w = \sum_{i \in SV} \alpha_i x_i y_i \quad 2.20$$

SV : c'est l'ensemble des indices des supports vecteurs.

Le problème dual devient :

$$\begin{cases} \sum_i^L \alpha_i - \frac{1}{2} \sum_{i,j}^L \alpha_i \alpha_j y_i y_j < x_i, x_j > \\ \alpha_i \geq 0 \\ \sum_i^L \alpha_i y_i = 0 \end{cases} \quad 2.21$$

$$\begin{aligned} \alpha_i \neq 0 & \text{ pour } y_i (< w, x_i > + b) = 1 \\ \alpha_i = 0 & \text{ pour } y_i (< w, x_i > + b) > 1 \end{aligned}$$

On obtient alors le vecteur de poids de l'hyperplan $w = \sum_{i=1}^L \alpha_i y_i x_i$ et l'offset $b = 1 - < w, x_i >$ avec x_i est un vecteur support de classe connue (ici sa classe est 1).

Et on obtient la fonction de l'hyperplan

$$f(x) = < w, x > + b = \sum_{i \in SV} \alpha_i y_i < x_i, x > + b \quad 2.22$$

✓ Cas de la classification non linéaire

Dans le cas de la classification non linéaire, l'hyperplan séparateur de la section précédente n'est plus valable. On doit commencer par relâcher les contraintes et autoriser quelques observations de tomber dans la marge de sécurité (entre les SV et l'hyperplan). Puis, on applique le SVM non linéaire dont l'astuce consiste à trouver un espace de plus grande dimension dans lequel les projections des exemples soient linéairement séparables. Il s'agit d'un espace de Hilbert H muni d'un produit scalaire qui peut être remplacé par une fonction noyau de l'espace de départ (espace des observations). On suppose $\Phi: R^p \rightarrow H$

$$x_i \mapsto \Phi(x_i)$$

Notre problème d'optimisation s'écrit alors sous la forme suivante :

$$\begin{cases} \max_{\alpha_i} \sum_i^L \alpha_i - \frac{1}{2} \sum_{i,j}^L \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle \\ \sum_i^L \alpha_i y_i = 0 \\ C \geq \alpha_i \geq 0 \end{cases} \quad 2.23$$

Le produit scalaire $\langle \phi(x_i), \phi(x_j) \rangle$ peut être remplacé par une fonction noyau K (le Tableau 2-12 présente les fonctions noyau les plus connues et utilisées)

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad 2.24$$

Le système (2.23) se transforme alors sous la forme suivante :

$$\begin{cases} \max_{\alpha_i} \sum_i^L \alpha_i - \frac{1}{2} \sum_{i,j}^L \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \sum_i^L \alpha_i y_i = 0 \\ C \geq \alpha_i \geq 0 \end{cases} \quad 2.25$$

Et la fonction de décision sera :

$$f(x) = \langle w, x \rangle + b = \sum_{i \in SV} \alpha_i y_i K(\langle x_i, x \rangle) + b \quad 2.26$$

TABLEAU 2-12: LES FONCTIONS NOYAUX.

Type du noyau	Fonction noyau	Commentaires
Linéaire	$K(x, x') = xx' + c$	La constante c est fixée a priori par l'utilisateur.
Polynômial	$K(x, x') = (xx')^d$	La puissance d est déterminée a priori par l'utilisateur.
Gaussien	$K(x, x') = e^{-\frac{\ x-x'\ ^2}{2\sigma^2}}$	L'écart type est spécifié a priori par l'utilisateur.
Sigmoïde	$K(x, x') = \tanh(\alpha x^t x' + b)$	Le théorème de Mercer n'est vérifié que pour certaines valeurs de a et b .

2.2.2 Pour la régression

La régression consiste à prédire une valeur réelle. Contrairement à la tâche de classification, la base d'apprentissage dans ce cas est formée par $\{x_i, y_i\}_{i=1}^N$, avec $y_i \in \mathbb{R}$. Ce problème revient à chercher une fonction $f(x_i) = y$ qui soit le plus proche possible de y_i .

Basés sur le principe des SVM, les Support Vector Regression (SVR) permettent de traiter des problèmes de régression (linéaire ou non linéaire). Ils consistent à trouver la fonction $f(x)$ qui a au plus une déviation ε par rapport aux exemples d'apprentissage x_i, y_i , et qui est la plus plate possible. Cela revient à ne pas considérer les erreurs inférieures à ε et à interdire celles supérieures à ε [127]. Comme le montre la Figure 2-15 dans les deux cas, linéaire ou non linéaire, l'objectif de la régression est de trouver la meilleure fonction qui peut suivre la distribution des observations. Le principe est comparable avec celui de la classification par SVM. En effet, au lieu d'imposer une marge entre les points des deux classes, on impose une sorte de "chaussette" autour des points grâce au paramètre ε .

La régression est importante dans le travail d'identification que nous allons présenter dans ce présent chapitre. En effet notre contribution se base sur la technique SVR pour l'identification des systèmes non linéaire et les systèmes commutant. Récemment, plusieurs travaux ont été consacrés à l'utilisation des SVR pour l'approximation de fonctions et pour la prédiction. Cette méthode est largement utilisée pour l'identification des systèmes non linéaires [128] [129] [130] [131].

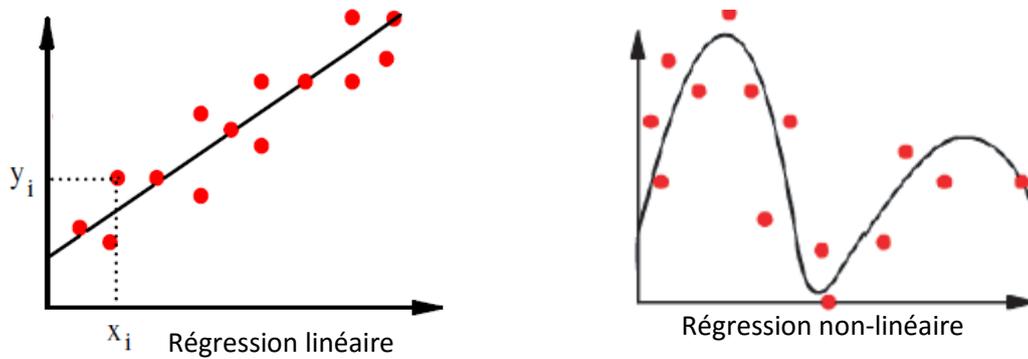


FIGURE 2-15: LA REGRESSION.

✓ Cas de la régression linéaire

Dans ce cas, la fonction f est décrite par :

$$f(x) = \langle w, x \rangle + b \quad 2.27$$

Avec $x \in R^q$ le vecteur d'entrée, $w \in R^q$ le vecteur des paramètres (ou poids), b une constante à déterminer et \langle, \rangle indique le produit scalaire dans R .

Assurer la platitude de la fonction f signifie que la norme des poids $\|w\|$ devra être minimisée. Le problème devient donc une minimisation de cette norme des erreurs inférieures à ε . Ainsi il peut être exprimé sous la forme suivante :

$$\begin{cases} \min \frac{1}{2} \|w\|^2, \\ \text{sc } |y_i - \langle w, x_i \rangle - b| \leq \varepsilon, i = 1, \dots, N \end{cases} \quad 2.28$$

Après introduction des variables de relâchement (slack variables) ξ_i, ξ_i^* pour rendre les contraintes du problème d'optimisation faisables, le problème peut être réécrit sous la forme suivante :

$$\begin{cases} \min_{w,b} \|w\|^2 + \sum_{i=1}^N (\xi_i + \xi_i^*), \\ |y_i - \langle w, x_i \rangle - b| \leq \varepsilon + \xi_i, \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i^* \geq 0, \xi_i \geq 0, i = 1, \dots, N. \end{cases} \quad 2.29$$

En utilisant une fonction de perte telle que la fonction ε -insensitive [2], on peut passer à une minimisation du lagrangien :

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) - \sum_{i=1}^N (\eta_i \xi_i + \eta_i \xi_i^*) - \sum_{i=1}^N \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) - \sum_{i=1}^N \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \quad 2.30$$

Et après calcul des dérivées par rapport au variables primaires (w, b, ξ_i, ξ_i^*) et remplacement dans l'équation on obtient:

$$\left\{ \begin{array}{l} \text{Max} -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^N (\alpha_i - \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \\ \text{sous les contraintes} \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0 \text{ et } \alpha_i^*, \alpha_i \in [0, C] \end{array} \right. \quad 2.31$$

$$w = \sum_{i=1}^N (\alpha_i^* - \alpha_i) x_i \quad 2.32$$

$$\begin{aligned} b &= y_i - \langle w, x_i \rangle - \varepsilon \text{ pour } \alpha_i \in [0, C], \\ b &= y_i - \langle w, x_i \rangle + \varepsilon \text{ pour } \alpha_i^* \in [0, C] \end{aligned} \quad 2.33$$

Au final, le modèle s'écrit sous la forme suivante :

$$f(x) = \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad 2.34$$

✓ Cas de la régression non linéaire

Dans le cas où il est impossible de trouver une fonction linéaire qui peut suivre les observations, la solution consiste à trouver un espace de plus grande dimension dans lequel on projette les exemples et dans un tel espace une régression linéaire est possible. Le principe est le même que dans le cas de la classification SVM non linéaire et il s'agit de l'idée qui fait la force de la théorie d'apprentissage statistique. Ainsi, en appliquant une fonction ϕ , on plonge les observations dans un espace de Hilbert de dimension plus élevé et encore une fois on va se baser sur les noyaux auto-reproduisant.

Le lagrangien correspondant est :

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) - \sum_{i=1}^N (\eta_i \xi_i + \eta_i \xi_i^*) - \sum_{i=1}^N \alpha_i (\varepsilon + \xi_i - y_i + \langle w, \phi(x_i) \rangle + b) - \sum_{i=1}^N \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, \phi(x_i) \rangle - b) \quad 2.35$$

Avec les conditions de KKT

$$\begin{aligned} \frac{\partial l}{\partial w} &= w - \sum_{i=1}^N (\alpha_i^* - \alpha_i) \phi(x_i) = 0, \\ \frac{\partial l}{\partial b} &= \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0, \end{aligned} \quad 2.36$$

$$\frac{\partial l}{\partial \xi} = C - \alpha_i - \eta_i = 0$$

$$\frac{\partial l}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0$$

On obtient

$$\begin{aligned} w &= \sum_{i=1}^N (\alpha_i^* - \alpha_i) \phi(x_i), \\ \eta_i &= C - \alpha_i, \\ \eta_i^* &= C - \alpha_i^*. \end{aligned} \quad 2.37$$

Ainsi, le problème peut être posé comme suit :

$$\left\{ \begin{aligned} \max L_D &= -\frac{1}{2} \sum_{i=1}^N (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_i, x_j) - \varepsilon \sum_{i=1}^N (\alpha_i^* - \alpha_i) + \sum_{i=1}^N y_i (\alpha_i^* - \alpha_i) \\ \sum_{i=1}^N (\alpha_i^* - \alpha_i) &= 0 \\ \alpha_i^*, \alpha_i &\in [0C] \end{aligned} \right. \quad 2.38$$

Et l'estimateur est donné par cette fonction :

$$f(x) = \sum_{i=1}^N (\alpha_i^* - \alpha_i) k(x_i, x) + b \quad 2.39$$

2.3 Least Squares Support Kernel Machines (LS-SKM) pour l'identification des systèmes

2.3.1 Least Squares Support Vector Machines (LS-SVM)

Comme une variante intéressante de SVM standard, Least Squares Support Vectors Machines (LS-SVM) a été proposée par Suykens Vandewalle et al [132] comme un outil intéressant pour l'identification des systèmes non linéaires complexes. Cette méthode est une simplification largement connu de l'approche SVM. Elle permet d'obtenir le modèle en résolvant un système d'équations linéaires au lieu d'un problème de programmation quadratique donné par les SVM standard [133]. Cette méthode a prouvé son efficacité dans des domaines d'applications multiples telles que la classification [134], la régression [135], l'identification des systèmes hybrides [136] et identification des systèmes de Hammerstein [137].

Le but de la technique LS-SVM consiste à construire une fonction f qui représente les dépendances de sortie y en entrée x et on suppose que cette fonction à la forme suivante :

$$y = \langle w, \varphi(x) \rangle + b \quad 2.40$$

Où $\varphi(\cdot)$ est la fonction non-linéaire qui fait correspondre l'espace d'entrée dans un espace caractéristique de dimension supérieure, voir infini. Et son estimation à travers ses paramètres w et b peut être reformulée comme selon la formule d'optimisation suivante :

$$\left\{ \begin{aligned} \min_{w_1, w_2, b, e} Jp(w_1, w_2, e) &= \frac{1}{2} \|w_1\|_2^2 + \|w_2\|_2^2 + C \frac{1}{2} \sum_{i=1}^N e_i^2 \\ \text{Sous les contraintes: } y_i &= \langle w, \varphi(x_i) \rangle + b + e_i, \\ i &= 1, 2, \dots, N \end{aligned} \right. \quad 2.41$$

Avec e_i sont des variables de relaxation et C est l'hyper-paramètre de régularisation. Et on arrive au lagrangien suivant :

$$L(w, b, e, \alpha) = J_p(w, e) - \sum_{i=1}^N \alpha_i (< w, \varphi(x_i) > + b + e_i - y_i) \quad 2.42$$

Les conditions d'optimalité sont :

$$\begin{cases} \frac{\partial l}{\partial w} = 0 \mapsto w = \sum_{i=1}^N \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \mapsto \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial l}{\partial e_i} = 0 \mapsto \alpha_i = C e_i, i = 1, \dots, N \end{cases} \quad 2.43$$

En remplaçant w et e_i dans le problème d'optimisation avec contraintes décrit précédemment, on obtient, alors :

$$y_i = \sum_{j=1}^N \alpha_j < \varphi(x_i), \varphi(x_j) > + b + \frac{1}{C} \alpha_i \quad 2.44$$

Les paramètres seront réécrits sous forme de solutions du système linéaire suivant :

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \Omega + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad 2.45$$

Avec $y = [y_1, \dots, y_N]^T$, $\vec{1} = [1, \dots, 1]^T$, $\alpha = [\alpha_1, \dots, \alpha_N]$ et Ω est une matrice carrée dans laquelle l'élément situé à la i^{me} colonne et la j^{me} ligne est égal à $\Omega_{i,j} = \varphi(x_i)^T \varphi(x_j) = K(x_i, x_j)$,

$i, j = 1, \dots, N$. K est une fonction noyau.

Nous pouvons obtenir directement α et b à partir du système d'équation précédent et obtenir l'estimateur LS-SVM :

$$y(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b \quad 2.46$$

2.3.2 Approche multi noyaux (Multi-kernel LS-SVM)

Dans cette section, je vais présenter la première approche que nous avons proposée dans ce contexte d'identification des systèmes non-linéaires. Elle consiste à appliquer deux fonction pondérées φ_1 et φ_2 au lieu d'une seule fonction φ donnée par l'équation suivante :

$$y(x) = a_1 < w_1, \varphi_1(x) > + a_2 < w_2, \varphi_2(x) > + b, \quad 2.47$$

Avec a_1 et a_2 étant deux facteurs de pondération. On arrive à la formulation suivante du problème d'optimisation :

$$\min_{w_1, w_2, b, e} J_p(w_1, w_2, e) = \frac{1}{2} \|w_1\|_2^2 + \|w_2\|_2^2 + C \frac{1}{2} \sum_{i=1}^N e_i^2 \quad 2.48$$

En appliquant les étapes indiquées dans la section précédente, on arrive au système linéaire suivant pour rétablir les paramètres :

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & a_1 \Omega_1 + a_2 \Omega_2 + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad 2.49$$

et la sortie estimée sera reformulée comme suit :

$$y(x) = \sum_{i=1}^N \alpha_i (a_1 K_1(x, x_i) + a_2 K_2(x, x_i)) + b \quad 2.50$$

Le réglage des paramètres semble être la tâche la plus complexe. En effet le nombre de ces paramètres (C, a_1, a_2, \dots) est important et l'espace est vaste ce qui nous oblige à effectuer plusieurs tests de validation avant de trouver les paramètres adéquats. Ainsi, nous avons proposé une méthode hybride qui combine un algorithme d'Optimisation par Essaim Particulaire avec contraintes (OEP) et notre méthode d'identification proposée basée sur LS-SVM à noyau multiple (Multi-kernel LS-SVM). L'OEP a été développée conjointement par R. Eberhart et J. Kennedy [138] en 1995. Cette méthode s'inspire du comportement collectif de groupes d'animaux, notamment, des bancs de poissons ou essaims d'oiseaux. Elle a été utilisée plusieurs fois conjointement avec des approches d'apprentissage [139].

Nous avons choisi de combiner un noyau linéaire par un noyau Gaussien. Dans ce cas, la méthode est notée LS-SVM (lin-rbf) et la fonction noyau résultante est donnée par la relation suivante :

$$K(x, x_i) = a_1 (xx_i^T + c) + a_2 \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad 2.51$$

L'idée principale est de prendre chaque paramètre à optimiser comme la position d'une particule qui peut se déplacer possédant ainsi une position et une vitesse. Dans notre cas, la position Z sera un vecteur de dimension 5 formé comme suit :

$Z(C, a_1, a_2, c, \sigma^2)$. Pour faciliter les notations, nous considérons que les deux paramètres p_1 et p_2 représentent les paramètres des deux noyaux utilisés (c-à-d $p_1 = c$ et $p_2 = \sigma^2$ dans ce cas LS-SVM (lin-rbf) et $p_1 = \sigma_1^2$ et $p_2 = \sigma_2^2$ dans le cas où on combine deux noyaux gaussien (LS-SVM (rbf-rbf)). On utilise une fonction cout F_{cout} pour juger la qualité du modèle obtenu s'il caractérise convenablement le comportement du système. Les étapes de l'algorithme que nous avons proposé sont décrites comme suit :

Algorithme d'optimisation par essaim particulaire, cas de la méthode LS-SVM à noyaux multiple.

- 1 :** Le vecteur position prends les 5 paramètres suivants (C, a_1, a_2, c, σ). Dans ce cas, 5 particules sont prises en compte.
 - 2 :** Initialisation : Génération des populations initiales aléatoirement : initialiser les positions
(C, a_1, a_2, p_1, p_2) de chaque particule ainsi que leurs vitesses correspondantes
 - 3 :** **pour** $i = 1$: Taille de swarm (on a pris taille de swarm=49) **faire**
 - 4 :** **tant que** le critère d'arrêt n'est pas atteint **faire**
 - 5 :** **pour** $i = 1$ à N (N c'est le nombre d'itérations) **faire**
 - 6 :** Déplacement de la particule
 - 7 :** Si les positions des particules ne violent pas les contraintes ($C > 0, a_1 \geq 0, a_2 \geq 0, p_1 > 0, p_2 > 0$)
 - 8 :** Calcul de la fonction objectif (fitness) notée F_{cout}
 - 9 :** Calcul du Pbest pour chaque particule et Gbest de la population
 - 10 :** Mise à jour des vitesses, positions, Pbest et Gbest des particules
 - 11 :** **fin pour**
 - 12 :** Condition d'arrêt : Atteinte du nombre maximum d'itérations ?
-

-
- 13 : Si oui
 14 : fin tant que
 15 : Si non-retour à l'étape 4
 16 : fin pour
-

2.3.3 Least Squares Support Kernel Machines (LS-SKM)

Nous avons remarqué que la formulation précédente d'une somme de deux fonctions noyaux pondérées est coûteuse en temps de calcul en appliquant les deux noyaux à toutes les observations. Aussi elle ne reflète pas une réalité des distributions de données. En effet, chaque donnée x_i est un vecteur formé de p attributs non homogènes et ayant des origines différentes. Nous avons alors pensé à découper ce vecteur et remplacer la recherche de la similarité entre deux vecteurs par une somme des similarités entre chaque deux sous vecteurs formant la base de données. Nous avons donc proposé une nouvelle approche dite Least Squares Support Kernel Machines (LS-SKM) [12].

nous supposons que chaque donnée x_i de la base d'apprentissage sera formée par les m sous vecteurs $x_{1,i}, x_{2,i}, \dots, x_{m,i}$ comme le montre la Figure 2-16 suivante :

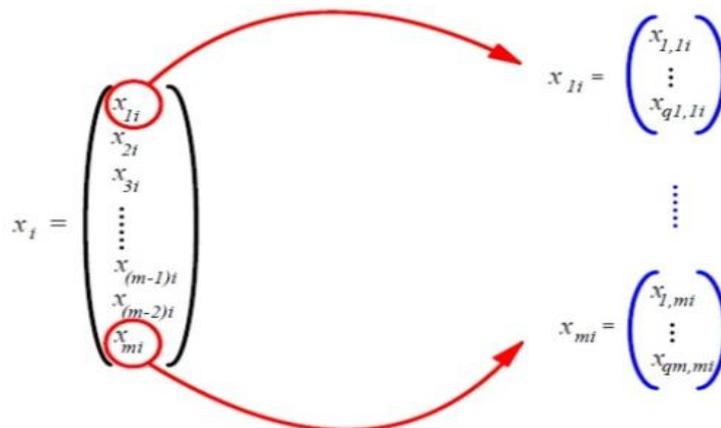


FIGURE 2-16: DIVISION DU VECTEUR OBSERVATION POUR UNE REGRESSION MULTI-NOYAUX.

Avec q_j est la dimension de chaque sous vecteur x_{qj} et $\sum_{j=1}^m q_j = q$ dont q est la dimension du vecteur régresseur complet x_i .

Suite à cette division, on applique plusieurs fonctions pondérées $\varphi_1, \dots, \varphi_m$, pour présenter la sortie du système, au lieu d'utiliser une seule fonction φ proposée par la méthode traditionnelle LS-SVM. La sortie devient comme suit :

$$y = \left\langle \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}, \begin{pmatrix} a_1 \varphi_1(x_1) \\ \vdots \\ a_m \varphi_m(x_m) \end{pmatrix} \right\rangle + b \quad 2.52$$

On considère que l'espace d'entrée χ est associée à m différents espaces caractéristiques F_1, \dots, F_m en utilisant des fonctions $\varphi_1, \dots, \varphi_m$ et $w = (w_1, \dots, w_m)$ et $x = (x_1, \dots, x_m) \in R^{l_1 + \dots + l_m}$. En se basant sur ces hypothèses, l'approche LS-SKM sera définie dans le problème d'optimisation suivant :

$$\left\{ \begin{array}{l} \min_{w_1, \dots, w_m, b, e} J_p(w_1, w_2, e) = \frac{1}{2} \|w_1\|_2^2 + \dots + \frac{1}{2} \|w_m\|_2^2 + C \frac{1}{2} \sum_{i=1}^N e_i^2 \\ \text{sous les contraintes: } y_i = \left\langle \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}, \begin{pmatrix} a_1 \varphi_1(x_{1i}) \\ \vdots \\ a_m \varphi_m(x_{mi}) \end{pmatrix} \right\rangle + b + e_i \\ i = 1, 2, \dots, N \end{array} \right. \quad 2.53$$

Avec C est le facteur de régularisation et e_i sont des variables des relaxations. Pour résoudre ce problème d'optimisation, nous avons défini la fonction de Lagrange suivante :

$$\begin{aligned} L(w_1, \dots, w_m, b, e, \alpha) &= J_p(w_1, \dots, w_m, e) \\ &- \sum_{i=1}^N \alpha_i (\langle w_1, a_1 \varphi_1(x_{1i}) \rangle + \dots + \langle w_m, a_m \varphi_m(x_{mi}) \rangle \\ &> + b + e_i - y_i) \end{aligned} \quad 2.54$$

avec α_i sont les multiplicateurs de Lagrange. Les conditions d'optimalités sont :

$$\left\{ \begin{array}{l} \frac{\partial l}{\partial w_1} = 0 \mapsto w_1 = \sum_{i=1}^N \alpha_i a_1 \varphi_1(x_{1i}), \\ \vdots \\ \frac{\partial l}{\partial w_m} = 0 \mapsto w_m = \sum_{i=1}^N \alpha_i a_m \varphi_m(x_{mi}), \\ \frac{\partial L}{\partial b} = 0 \mapsto \sum_{i=1}^N \alpha_i = 0 \\ s \frac{\partial l}{\partial e_i} = 0 \mapsto \alpha_i = C e_i, i = 1, \dots, N. \end{array} \right. \quad 2.55$$

En remplaçant l'expression de w et e_i dans le problème d'optimisation sous contraintes décrite par (2.53), nous obtenons :

$$\begin{aligned} y_i &= \sum_{j=1}^N \alpha_j (a_1 \langle \varphi_1(x_{1i}), \varphi_1(x_{1j}) \rangle + \dots + a_m \langle \varphi_m(x_{mi}), \varphi_m(x_{mj}) \rangle) + b \\ &+ \frac{1}{C} \alpha_i \end{aligned} \quad 2.56$$

Les paramètres seront rétablis sous forme de solution du système linéaire suivant :

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & a_1 \Omega_1 + a_m \Omega_m + C^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad 2.57$$

L'estimateur LS-SKM, $y(x)$, est obtenu et il est donné par l'expression suivante :

$$y(x) = \sum_{i=1}^N \alpha_i (a_1 K_1(x_1, x_{1i}) + \dots + a_m K_m(x_m, x_{mi})) + b \quad 2.58$$

Ce qui donne l'expression équivalente :

$$y(x) = \sum_{i=1}^N \alpha_i^1 K_1(x_1, x_{1i}) + \dots + \sum_{i=1}^N \alpha_i^m K_m(x_m, x_{mi}) + b \quad 2.59$$

Le principe de l'identification des systèmes non-linéaire à l'aide de l'approche LS-SKM est décrit par la Figure 2-17 suivante :

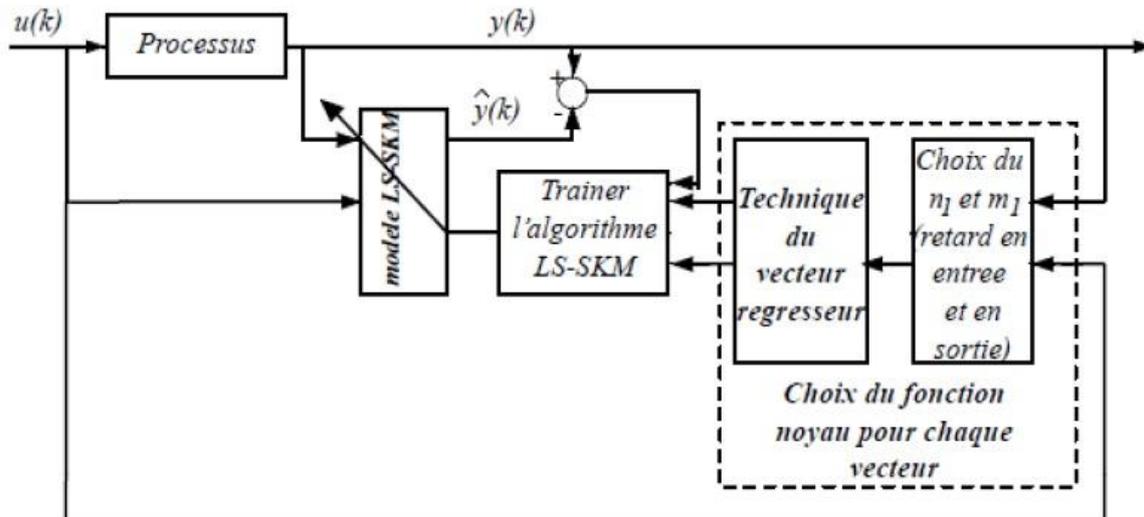


FIGURE 2-17: PRINCIPE DE L'IDENTIFICATION DES SYSTEMES PAR LS-SKM.

Le réglage des paramètres suit le même principe que le LS-SVM utilisant l'algorithme OEP. La Figure 2-18 décrit le principe.

Cette approche a été appliquée pour une identification en ligne. Avec une structure de prédictor à un pas, l'algorithme modifie récursivement le modèle en ajoutant une nouvelle mesure en temps réel.

L'approche LS-SKM a été aussi appliquée sur le modèle NARMA (Non Linear Auto Regressive Moving Average) [13]. Il a l'avantage d'être capable de modéliser le système avec un ordre relativement faible et de réduire le nombre de paramètres.

2.3.4 Validation

L'approche LS-SKM a été validée sur différents systèmes tel que le single-input/single-output (SISO) proposé dans [140] et dont le modèle est le suivant :

$$y(k) = \frac{y(k-1)y(k-2)(y(k-1) + 4.5)}{1 + y(k-1)^2 + y(k-2)^2} + u(k-1) \quad 2.60$$

Les résultats sont pertinents comme le montre la Figure 2-19 où on peut voir la sortie estimée correspondre parfaitement à la sortie réelle.

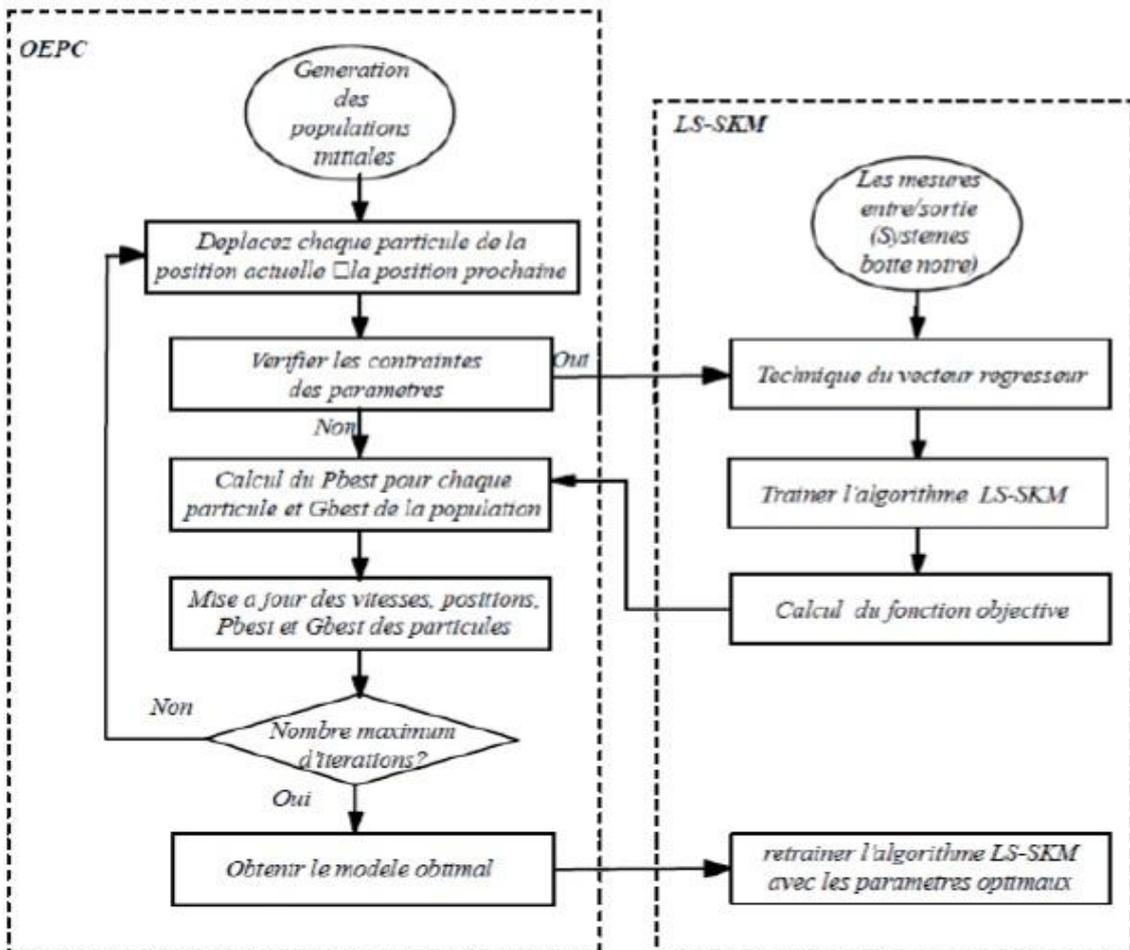


FIGURE 2-18: REGLAGE DES PARAMETRES DE LA METHODE LS-SKM EN SE BASANT SUR L'OEPC.

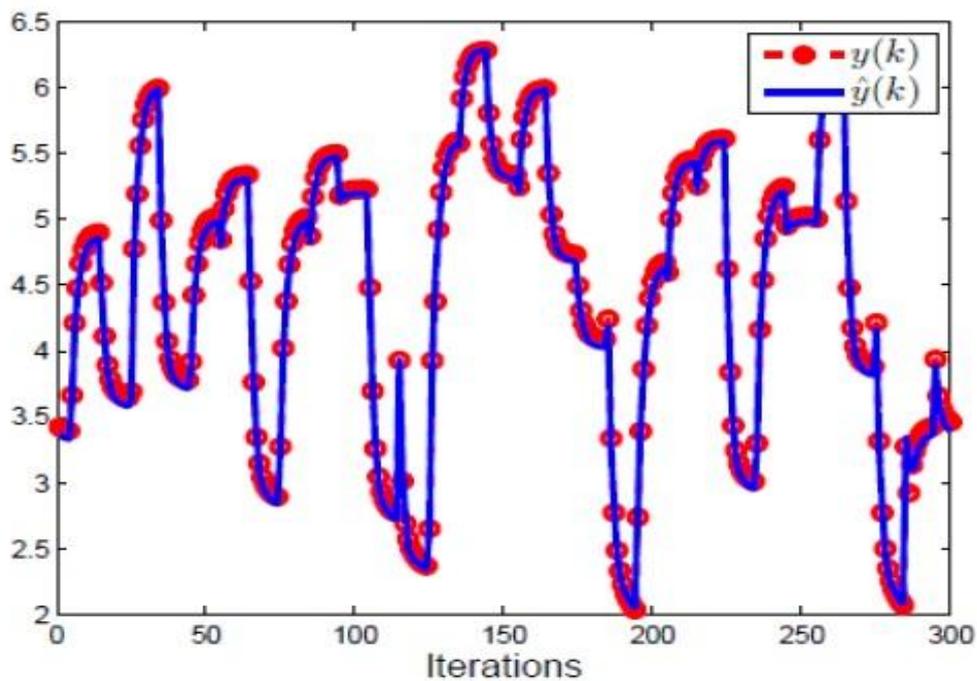


FIGURE 2-19: ÉVOLUTIONS DE LA SORTIE DU SYSTEME ET DE LA SORTIE DU MODELE BASEES SUR LA METHODE LS-SKM(LIN-LIN).

Etant donné que le coefficient de corrélation total (R_{tot}^2) mesure la similitude entre la sortie réelle du système et la sortie modèle, le Tableau 2-13 suivant montre l'efficacité de notre approche LS-SKM.

$$R_{tot}^2 = 1 - \frac{\sum_{k=1}^N \varepsilon^2(k)}{\sum_{k=1}^N y^2(k)} \quad 2.61$$

TABEAU 2-13: VALEURS DE R_{TOT}^2 ET TEMPS DE CALCUL POUR LES DIFFERENTES METHODES

	Linear LS-SVM	Gaussian Ls-SKM	LS-SKM (rbf-lin)	LS-SKM (lin-lin)
a_1	ϕ	ϕ	11.4117	10.747
a_2	ϕ	ϕ	2	1.2
C	16.1960	10.62	1	30
c	9.0596	ϕ	9.3708	10.2
$2\sigma^2$	ϕ	10.27	3	ϕ
$R_{tot}^2(\%)$	95.40	97.16	97.47	97.13
<i>Temps (sec)</i>	7.3632	7.8313	7.7376	7.1604

2.4 Support Kernel Regression (SKR) pour l'identification des systèmes

En adoptant la même division du vecteur observation (vecteur régresseur) que la section précédente, chaque donné x_i de la base d'apprentissage sera représenté par m sous vecteurs. Par la suite l'idée de l'approche Support Kernel Regression (SKR) consiste à appliquer le principe de la régression SVR mais Au lieu d'utiliser une seule fonction ϕ , on applique plusieurs fonctions pondérées ϕ_1, \dots, ϕ_m . Ces fonctions seront utilisées pour rapprocher la sortie du système comme suit :

$$y = \left\langle \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}, \begin{pmatrix} a_1 \phi_1(x_1) \\ \vdots \\ a_m \phi_m(x_m) \end{pmatrix} \right\rangle + b \quad 2.62$$

En se basant sur les mêmes hypothèses que dans l'approche LS-SKM en termes d'espace d'entrée, espaces caractéristiques, w et x , nous obtenons le problème d'optimisation suivant.

$$\begin{aligned} & \min \left\{ \frac{1}{2} \|w_{1i}\|^2 + \dots + \frac{1}{2} \|w_{mi}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \right. \\ & \left. \text{sc} \begin{cases} y_i - \langle w_{1i}, \phi_{1i}(X_{1i,i}) \rangle - \dots - \langle w_{mi}, \phi_{mi}(X_{mi,i}) \rangle - b \leq \varepsilon + \xi_i \\ \langle w_{1i}, \phi_{1i}(X_{1i,i}) \rangle + \dots + \langle w_{mi}, \phi_{mi}(X_{mi,i}) \rangle + b - y_i \leq \varepsilon + \xi_i^* \end{cases} \right. \end{aligned} \quad 2.63$$

Le lagrangien correspondant impliquant les variables duales α_i et η_i est le suivant :

$$\begin{aligned}
L := & \frac{1}{2} \|w_{1i}\|^2 + \dots \\
& + \frac{1}{2} \|w_{mi}\|^2 C \sum_{i=1}^N (\xi_i + \xi_i^*) \\
& - \sum_{i=1}^N (\eta_i \xi_i + \eta_i^* \xi_i^*) \\
& - \sum_{i=1}^N \alpha_i (\varepsilon + \xi_i - y_i + \langle w_{1i}, \phi_{1i}(X_{1i,i}) \rangle + \dots + \\
& \langle w_{mi}, \phi_{mi}(X_{mi,i}) \rangle + b) \\
& - \sum_{i=1}^N \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w_{1i}, \phi_{1i}(X_{1i,i}) \rangle + \dots + \\
& \langle w_{mi}, \phi_{mi}(X_{mi,i}) \rangle - b)
\end{aligned} \tag{2.64}$$

Avec η_i^* , η_i , α_i , α_i^* sont les multiplicateurs de Lagrange.

Les conditions d'optimalités sont :

$$\left\{ \begin{array}{l}
\frac{\partial L}{\partial b} = 0 \mapsto \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\
\frac{\partial L}{\partial w_{1i}} = 0 \mapsto w_{1i} - \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\
\vdots \\
\frac{\partial L}{\partial w_{mi}} = 0 \mapsto w_{mi} - \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\
\frac{\partial L}{\partial \xi_i^*} = 0 \mapsto C - \alpha_i^* - \eta_i^* = 0 \\
\frac{\partial L}{\partial \xi_i} = 0 \mapsto C - \alpha_i - \eta_i = 0 \\
i = 1, \dots, N.
\end{array} \right. \tag{2.65}$$

$$\begin{aligned}
\max & \left\{ \begin{array}{l}
\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) (k_u(X_{1i,i}, X_{1j,j}) + \dots + k_{mi}(X_{mi,i}, X_{mj,j})) \\
-\varepsilon \sum_{i,j=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*),
\end{array} \right. \tag{2.66} \\
\text{sc} & \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \text{ et } \alpha_i, \alpha_i^* \in [0, C].
\end{aligned}$$

La fonction de régression est alors donnée par la formulation suivante :

$$y(x) = \sum_{i=1}^N (\alpha_i^* - \alpha_i) [k_{1i,i}(X_{1i,i}, X_{1j,j}) + \dots + k_{mi}(X_{mi,i}, X_{mj,j})] + b \tag{2.67}$$

Puisque nous sommes confrontés au problème d'identification des systèmes non linéaires, la division du vecteur d'entrée $x_i = (u(i-1), \dots, u(i-m), y(i-1), \dots, y(i-n))$ se fait en s'appuyant sur deux critères : soit le temps soit le type d'entrée. La première séparation (division) se base sur l'hypothèse que l'entrée et la sortie du système à l'instant $(i-1)$ influe, fortement, sur la sortie modèle du système à l'instant i comparée aux autres instants. Pour cette raison, nous avons proposé d'utiliser un noyau pour le sous-groupe formé par $u(i-1)$ et $y(i-1)$ et un autre noyau pour le reste du vecteur.

La deuxième séparation est fondée sur la division du régresseur selon l'origine du signal. Puisque le problème en question est un problème d'identification, cette division prend les signaux du commande ($u(i-1), \dots, u(i-m1)$) dans un premier groupement et les sorties retardées ($y(i-1), \dots, y(i-n1)$) dans un autre groupement.

Nous avons appliqué cette approche SKR sur un système NARMA et nous avons eu aussi des résultats intéressants. A travers le Tableau 2-14 suivant, on peut comparer les deux approches SKR et LS-SKM pour le même système (modèle) de validation.

TABLEAU 2-14: ÉVOLUTION DE R_{tot}^2 EN FONCTION DES PARAMETRES m_1, n_1

	Quadratic SVR rbf	Quadratic SKR		ϵ -insensitive SVR rbf	ϵ -insensitive SVR rbf	
		(rbf-lin)	(rbf-rbf)		(rbf-lin)	(rbf-rbf)
$m_1 = 2, n_1 = 2$	92.7216	92.8239	89.8832	92.7037	92.7114	89.4828
$m_1 = 2, n_1 = 4$	79.4039	92.9119	87.8897	81.8926	92.6129	88.3924
$m_1 = 2, n_1 = 5$	68.8681	92.1809	85.8544	74.7954	92.2881	86.7706
$m_1 = 3, n_1 = 3$	86.5163	92.8242	88.5696	86.8895	92.6100	88.8937
$m_1 = 3, n_1 = 4$	77.6569	94.6908	88.3008	78.4300	94.7375	88.9023
$m_1 = 3, n_1 = 5$	69.2320	95.0638	88.3919	76.7768	94.7368	88.3002
$m_1 = 4, n_1 = 5$	67.7843	94.8251	92.1101	76.2927	94.4972	92.5302
$m_1 = 4, n_1 = 6$	51.7861	92.8111	85.7681	69.2924	91.2839	85.8577
$m_1 = 10, n_1 = 10$	8.6875	92.2183	73.8575	23.2393	91.8490	81.1670

3. Apprentissage statistique pour l'identification des systèmes dynamiques hybrides

3.1 Modélisation des systèmes dynamiques hybrides

Les systèmes dynamiques hybrides (SDH) font interagir simultanément des phénomènes à la fois continue et événementielle. Etablir des modèles précis pour ces systèmes complexes est une tâche relativement difficile pour les automaticiens. Ils s'agissent des systèmes commutant dont la difficulté de leur modélisation réside dans le fait qu'il faut estimer la dynamique discrète, modélisée par une variable nommée état discret, en même temps que l'identification des sous-modèles de la partie continue. La première partie concerne la détermination de la dynamique de commutation entre les sous-modèles de fonctionnement du système. Le nombre des sous-modèles est parfois inconnu. La commutation est souvent très rapide et gérée par des paramètres non contrôlables et aléatoires. La deuxième partie consiste à l'identification des sous-systèmes. Une tâche qui peut être classique comme elle peut envisager quelques difficultés. Toutes ces raisons et bien d'autres ont rendu le problème de l'identification de ces systèmes compliqué et suscite beaucoup de recherche. Ce problème a occupé l'attention de plusieurs chercheurs qui ont proposés des formalismes afin d'établir un modèle homogène permettant la conciliation entre les parties discrètes et continues.

Selon [141], il y a trois familles d'approches pour la modélisation des systèmes SDH à savoir l'approche discrète, l'approche mixte et l'approche continue. Ces approches se différencient par l'interaction entre la partie discrète et la partie continue.

3.1.1 Approche de modélisation discrète

Elle consiste à approximer la dynamique continue du système de façon à se ramener à un système à événement discret. Cette modélisation fait appel à la théorie classique de

supervision des systèmes à événements discrets (SED). La modélisation de ces systèmes est fondée sur la théorie de la supervision des SED qui applique le modèle automate et les langages formels. Parmi les modèles de l'approche discrète, on peut citer le modèle à base de la partition de l'espace d'état continu [142] et le modèle basé sur la notion d'invariants naturels [143].

3.1.2 Approche modélisation mixte

Elle considère dans une même structure les comportements continu et discret du système. Des outils de modélisation sont envisageables par cet approche mixte notamment : les automates hybrides [144], [145], [146] définissant un modèle de présentation fondamental de cette approche, les différents extensions des réseaux de Petri [147], les statecharts hybrides qui apportent des solutions au niveau des problèmes posés par des spécifications des modèles, particulièrement, par l'emploi d'une structure hiérarchisée [148].

3.1.3 Approche modélisation continue

Elle fait associer la partie événementielle dans le modèle de la dynamique continue du système afin d'utiliser les théories de l'automatique classique et avoir un modèle global et unifié. Plusieurs modèles peuvent être cités dans ce contexte notamment les systèmes linéaires à saut JLS (jump linear systems) [149], les systèmes linéaires à commutation SLS (switched linear systems) [150] et les systèmes affines par morceaux PWA (piecewise affines linear systems) [151]. Dans ce présent travail d'identification des SDH, nous avons focalisé nos études sur ce type de modélisation. Les travaux d'identification présentés dans les sections précédentes rentrent dans ce même contexte d'identification notamment les techniques LS-SVM, LS-SKM et SKR.

3.1.4 Problème d'identification pour les systèmes dynamiques hybrides

Le problème d'identification des systèmes dynamiques hybrides peut être posé différemment pour les modèles d'état et les modèles de régression. En effet, la première classe (les modèles d'état) est composée d'une collection finie de sous-systèmes linéaires indexés par un état discret modélisant le mode en cours d'évolution. Chaque sous-modèle est défini par l'équation suivante :

$$\begin{cases} x(k+1) = A_i x(k) + B_i u(k) \\ y(k) = C_i x(k) \end{cases} \quad 2.68$$

Avec i , l'état discret du système, indexe le mode en cours d'évolution, et il appartient à l'ensemble fini $S = \{1, \dots, s\}$ (avec s est le nombre de sous-modèles inconnu à priori). $x(k) \in R^n$, $y(k) \in R$ et $u(k) \in R$ sont respectivement le vecteur d'état continu, la sortie et la commande du système. A_i , B_i et C_i sont les matrices paramètres associées au sous-modèle i qui est actif à l'instant k .

L'identification pour cette classe consiste à estimer : le nombre des sous-modèles s , l'ordre du système n , un observateur $\hat{x}(k)$ de l'état continu du système, et l'état discret i du système. Et ceci à partir d'un ensemble de mesures entrée-sortie $\{y(k), u(k)\}_{k=1}^N$ définies sur un horizon de N observations,

Pour la deuxième classe (les modèles de régression), nous posons le problème d'identification des modèles à commutation linéaires sous la forme entrée-sortie suivante :

$$y(k) = \Phi_i(x(k)) \quad 2.69$$

Avec Φ_i est une fonction affine définissant le modèle qui gouverne le i^{me} sous-modèle, $i \in S = \{1, \dots, s\}$ est l'état discret du système, s est le nombre des sous-modèles et $x(k) = [y(k-1), \dots, y(k-na), u(k-1), \dots, u(k-nb)]$ est le vecteur de régression. Où $y(k)$ et $u(k)$ sont respectivement la sortie et l'entrée du système mesurées à l'instant k , n_a et n_b sont les ordres du système.

Pour cette classe, le problème d'identification consiste à : estimer le nombre des sous modèles s , les ordres du système n_a et n_b ; Identifier les fonctions affines Φ_i , $i = 1, \dots, s$; et estimer l'état discret i du système.

Les chercheurs ont proposé plusieurs approches pour l'identification de ces systèmes hybrides, on peut citer : l'approche bayésienne [152], l'approche basée sur la classification [153], l'approche basée sur la programmation mixte en nombres entiers [154], l'approche de l'adaptation du poids [155] et l'approche algébrique [156] [157], l'approche de l'erreur bornée [158]. La méthode bayésienne et la méthode de l'erreur bornée sont utilisées pour l'identification des systèmes commutant de type SARX dont la dynamique de commutation est arbitraire. Toutefois, Les autres approches se concentrent sur le problème d'identification du système affine par morceaux (PWA), où l'état discret dépend de l'état continu.

Ces approches ont été étendues dans plusieurs travaux. Notamment, l'approche bayésienne a été étendue aux modèles d'erreur de sortie par morceaux [159]. Tandis que, l'étape de classification à la base de la méthode décrite dans [160] a été améliorée dans [161]. Pour l'approche algébrique, l'algorithme récursive, proposé dans [162] [163], a été récemment étendu pour faire face aux systèmes commutant MIMO [164] [165] [166] et en tenant compte d'un temps de séjour minimum pour quelques 'une.

3.2 Classification pour l'estimation de l'état discret

Pour l'estimation de l'état discret d'une classe de système dynamique hybride, exprimés par des équations d'état en temps discret, Nous avons proposé deux alternatives, dans le cadre des travaux de thèse de Mme Sellami [3]. Ces deux alternatives consistent à une estimation hors ligne pour la première et une estimation en ligne pour la deuxième.

3.2.1 Estimation hors ligne

Nous avons proposé une nouvelle approche basée sur la technique de classification de Chiu [167], [168] et qui consiste à associer la sortie courante $y(k)$ à son sous-modèle approprié. L'état discret est estimé comme une indice de sous-modèle qui génère le vecteur de régression $(u(k), y(k))$, il est élaboré au sens de la minimisation d'un critère de décision $J_i(k)$, tel que :

$$\hat{i}(k) = \arg \min_{i=1, \dots, s} J_i(k) \quad 2.70$$

Le critère $J_i(k)$ peut être choisi par différentes méthodes, puisqu'on ne dispose pas d'un modèle spécifique qui modélise la dynamique de l'état discret du système linéaire commutant. Dans ce travail, nous avons considéré $J_i(k)$ comme une fonction de l'erreur quadratique entre la sortie du système $y(k)$ et la première composante du i^{me} centre y_i^* :

$$J_i(k) = |y(k) - y_i^*|^2 \quad 2.71$$

Il s'agit de la distance entre la donnée $y(k)$ et le centre y_i^* de la i^{eme} classe.

Pour valider l'approche, nous l'avons appliqué sur un modèle commutant linéaire à deux modes ($s = 2$) d'ordre $n = 3$ et nous avons appliqué une séquence d'entrée aléatoire. Le comportement de l'état discret estimé $\hat{i}(k)$ et de l'état discret réel $i(k)$, présentés sur la Figure 2-20, prouve la viabilité de cette méthode d'identification hors-ligne.

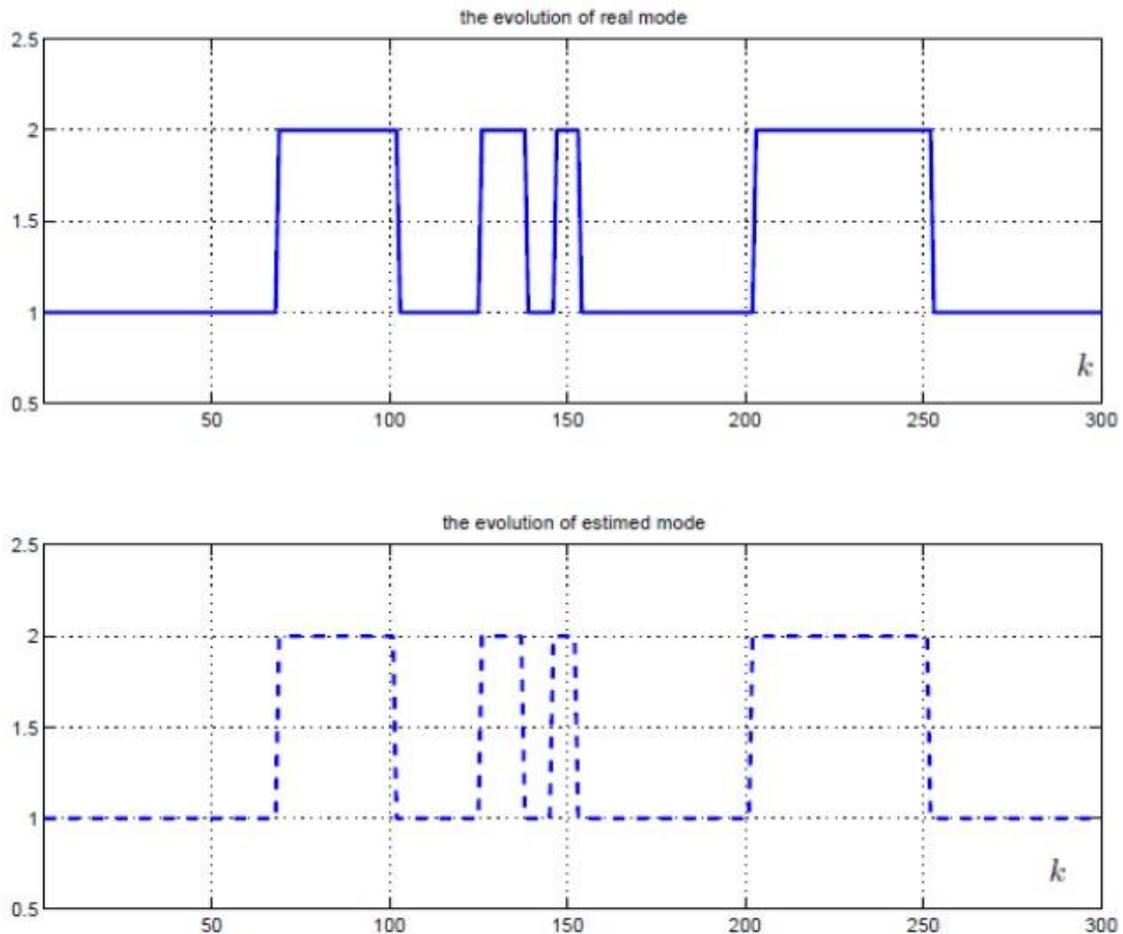


FIGURE 2-20: TRACE DE DESSUS : L'ÉVOLUTION DU MODE REEL. TRACE DE DESSOUS : L'ÉVOLUTION DU MODE IDENTIFIÉ.

3.2.2 Estimation en ligne

Nous avons proposé, dans le même cadre, un système d'identification en ligne basé sur la technique de détection des instants de commutation. Ce système comporte une phase d'apprentissage au préalable, basée sur l'algorithme de classification de Chiu. Cette phase permet de déterminer le nombre des sous-modèles s , et des groupements des données à leurs régions de validité appropriées. Ensuite, l'algorithme SVM multi-classes est appliqué afin de déterminer tous les hyperplans limites des différentes classes du système. Ces frontières peuvent servir à la détection de l'instant de commutation. En effet, les données de chaque classe sont localisées au voisinage de son centre lorsque le sous-modèle est en cours d'évolution. Mais ces données de régression s'éloignent vers ces frontières de la région d'activation du sous-modèle en se rapprochant de l'instant de commutation. L'instant $(k+1)$ sera un instant de commutation d'un premier sous-modèle i vers un deuxième sous-modèle j si la distance $d_{i,j}(k)$ entre la donnée disponible à l'instant k et l'hyperplan limite des sous-modèles i et j ($j \in \{1, \dots, s\}, j \neq i$) est inférieure à un seuil pratique δ donnée par l'opérateur. Dans le cas de concurrence entre plusieurs hyperplans on choisit le plus proche de la donnée. Finalement, selon la valeur de mode estimé le bloc de l'observateur est mis à jour. La Figure 2-21 présente la structure de l'identification en ligne.

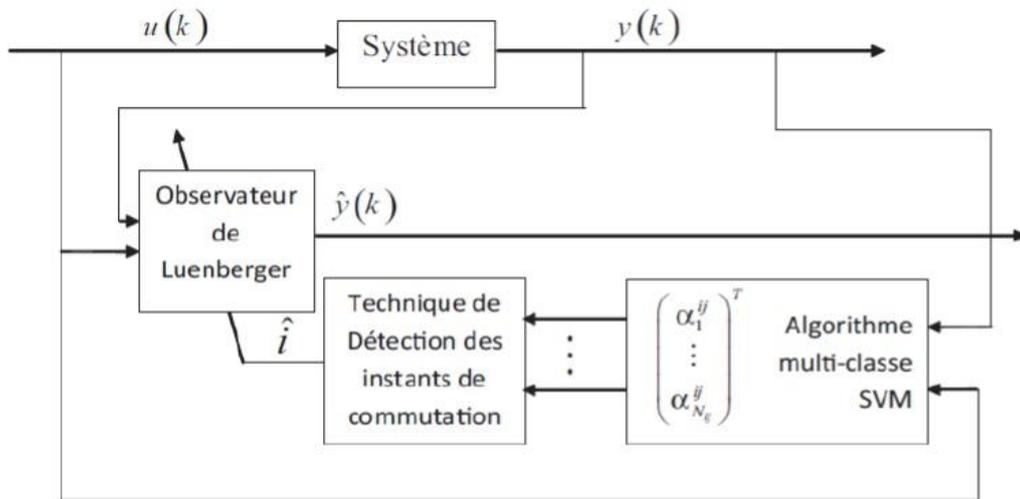


FIGURE 2-21: STRUCTURE D'IDENTIFICATION EN LIGNE.

Pour la validation de cette approche d'identification en ligne, nous l'avons appliquée sur un modèle SISO composé de deux sous-modèles ($s = 2$) d'ordre $n = 2$. Les trois figures Figure 2-22, Figure 2-23 et Figure 2-24 montrent l'efficacité de cette approche pour la détection de l'instant de commutation. Sur la première figure on voit bien que le décalage est négligeable voir nul entre le mode réel et le mode estimé. Ainsi que sur la Figure 2-23 les signaux des deux sorties réelle et estimée sont confondus. Sur la Figure 2-24, on peut remarquer le passage entre les deux modes à partir de la distribution des points dans les deux classes et leurs distances de l'hyperplan séparateur sachant que les croix en noirs sont les centres de chaque classe.

Nous avons aussi testé ces approches sur d'autres systèmes tels que le système chariot double avec accouplement élastique.

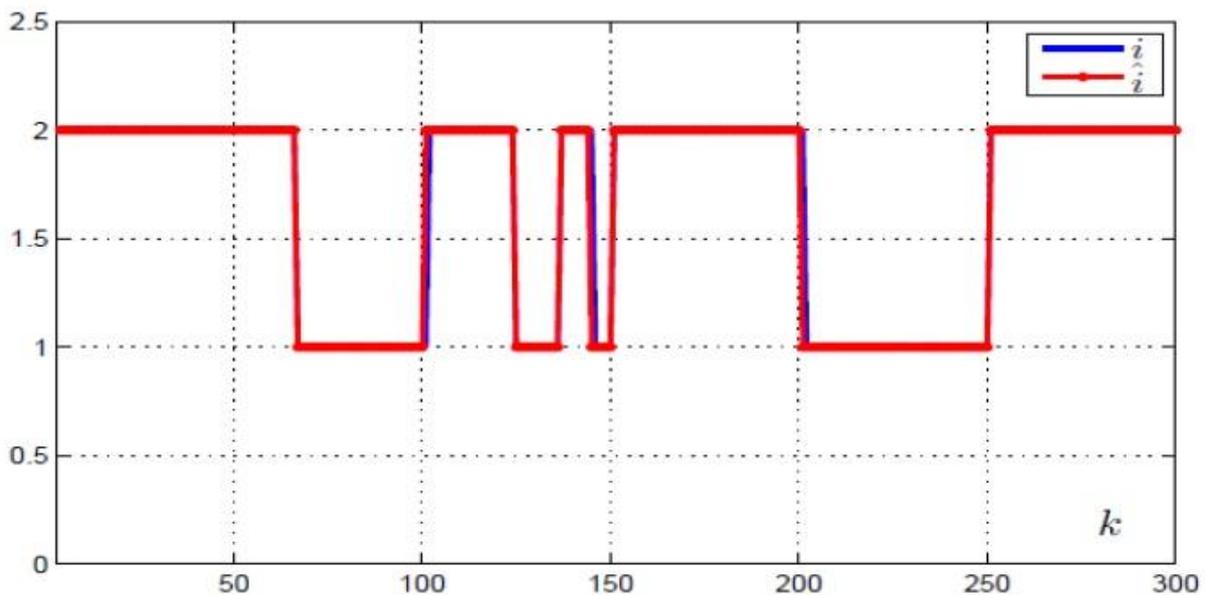


FIGURE 2-22: L'EVOLUTION DU MODE REEL ET DE SON ESTIME.

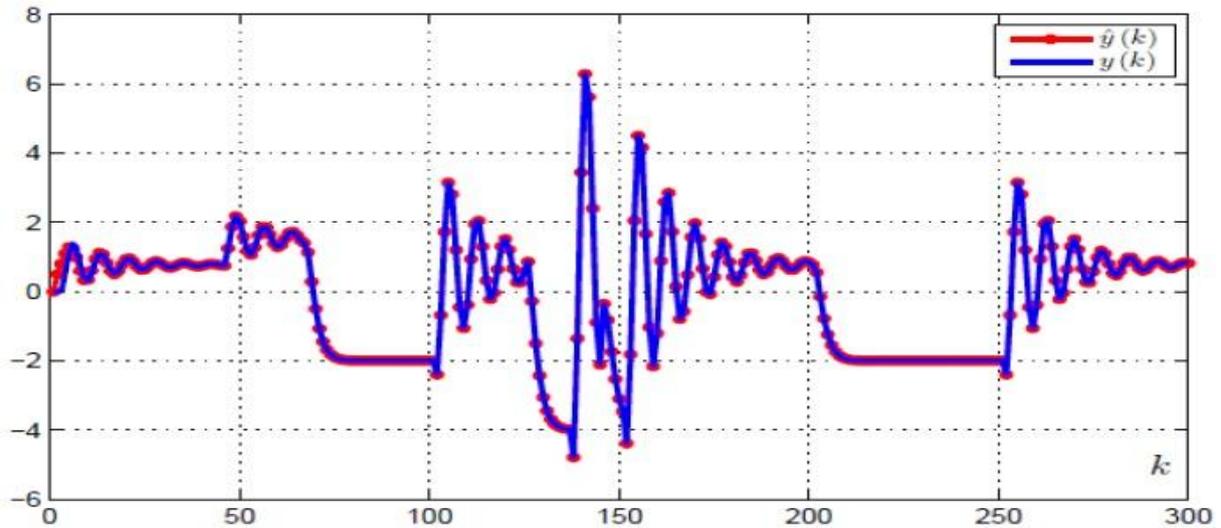


FIGURE 2-23: L'EVOLUTION DU SIGNAL DE SORTIE DU SYSTEME $y(k)$ ET LA SORTIE ESTIMEE $\hat{y}(k)$.

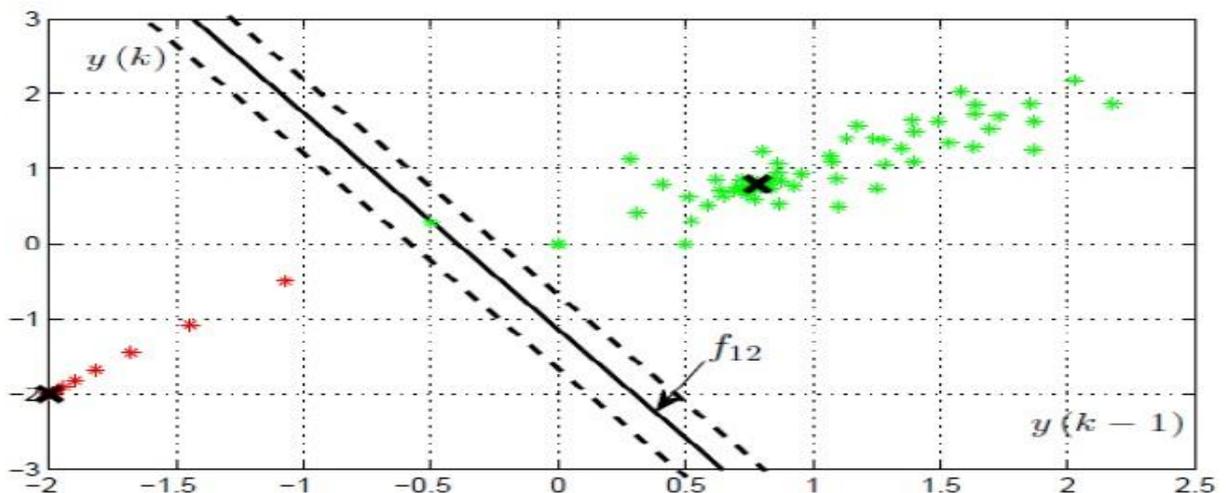


FIGURE 2-24: CLASSIFICATION DES DONNEES DANS L'ESPACE DE REGRESSION.

3.3 Régression pour l'identification des systèmes linéaire commutant

Plusieurs approches ont été développées et proposées par des chercheurs pour l'identification des systèmes hybrides font appel à des techniques d'apprentissage, de classification et de régression. On peut citer des travaux basés sur des techniques d'apprentissage statistique [169] [170] [171], ou d'autres techniques de classification [172] [173] [174]. Nous avons prouvé précédemment (la première partie de ce chapitre) l'efficacité des techniques de la régression pour l'estimation du comportement des systèmes automatiques. Dans le cadre de la thèse de Mme Sellami, nous avons travaillé sur l'adaptation d'une nouvelle approche d'identification des systèmes linéaires commutant. En se basant sur l'approche de classification dynamique proposée par Boubacar [124] et l'approche SKR que j'ai présenté dans ce présent document [15], nous avons développé un algorithme d'identification auto-adaptatif multi noyau que nous avons baptisé " AUSKR ". Cette approche consiste à classer en ligne les données de régression et à estimer simultanément la fonction de régression qui les gouverne. A l'apparition d'une nouvelle donnée, on commence par l'incorporer à la classe (mode) la plus appropriée via le calcul d'un critère de similarité. A partir de ce critère de

similarité on peut déterminer un ensemble Ω^{win} des classes gagnantes à chaque fois qu'une nouvelle donnée se présente. Ce critère est défini tel que :

$$\Omega^{win} = \{C_m^k / S(X_k, C_m^k) - \varepsilon_{th} \geq 0\} \quad 2.72$$

Par la suite, le modèle de connaissances est mis à jour en ligne suivant trois procédures dynamiques liées à l'environnement de données non-stationnaires. Le choix de l'une des procédures est décidé selon le cardinal de l'ensemble des classes gagnantes Ω^{win} .

✓ Si $\text{card}(\Omega^{win}) = 0$, on applique la procédure de création. C'est une phase de création de nouvelle classe (modèle) lors de l'apparition d'une nouvelle partition avec une étape d'initialisation de modèle. deux cas sont envisageables par rapport au nombre de classes déjà existants. S'il n'y a aucune classe déjà déterminée, on parle d'une étape d'initialisation. Dans le cas contraire (il y a déjà des classes) il s'agit de l'apparition d'une nouvelle partition (classe) contenant la nouvelle donnée.

✓ Si $\text{card}(\Omega^{win}) = 1$, on applique la procédure d'adaptation. Lors de cette phase les modèles des classes évoluant sont adaptés selon une règle de mise à jour récursive. L'algorithme récursif considéré est basé sur l'équation de descente de gradient dans l'espace RKHS

✓ Si $\text{card}(\Omega^{win}) \geq 2$, on applique la procédure de fusion : cette phase permet de fusionner les informations mutuelles par utilisation des règles de fusion et en tenant compte de problème de chevauchement de classes. Dans ce cas, deux ou plusieurs groupes partagent les données acquises. Ces groupes seront fusionnés pour créer un nouveau groupe.

Pour la validation de cette approche de régression dynamique, nous l'avons testé sur un système commutant entre trois sous-modèles linéaire avec des ordres invariants $n_a=2$ et $n_b=1$. La Figure 2-26 montre l'apport de cette approche AUSKR. En effet, les deux sorties estimées et réelles sont confondues.

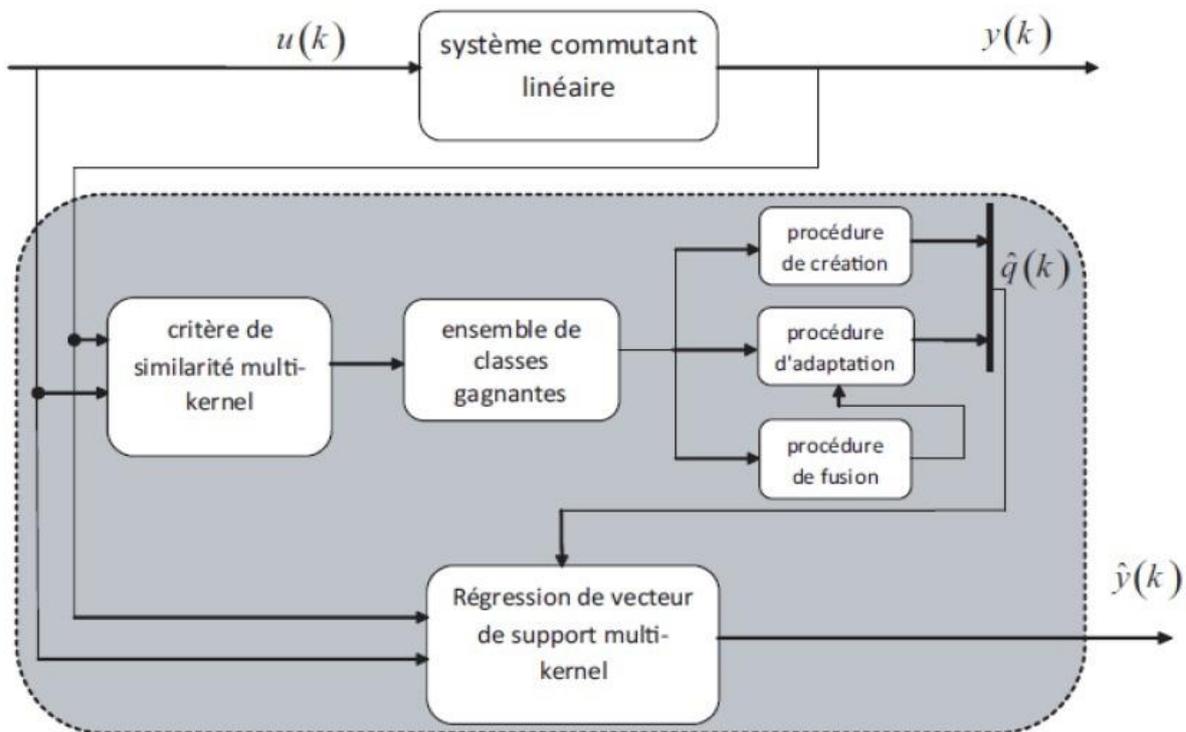


FIGURE 2-25: PROCEDURE D'IDENTIFICATION DES SYSTEMES LINEAIRES COMMUTANT.

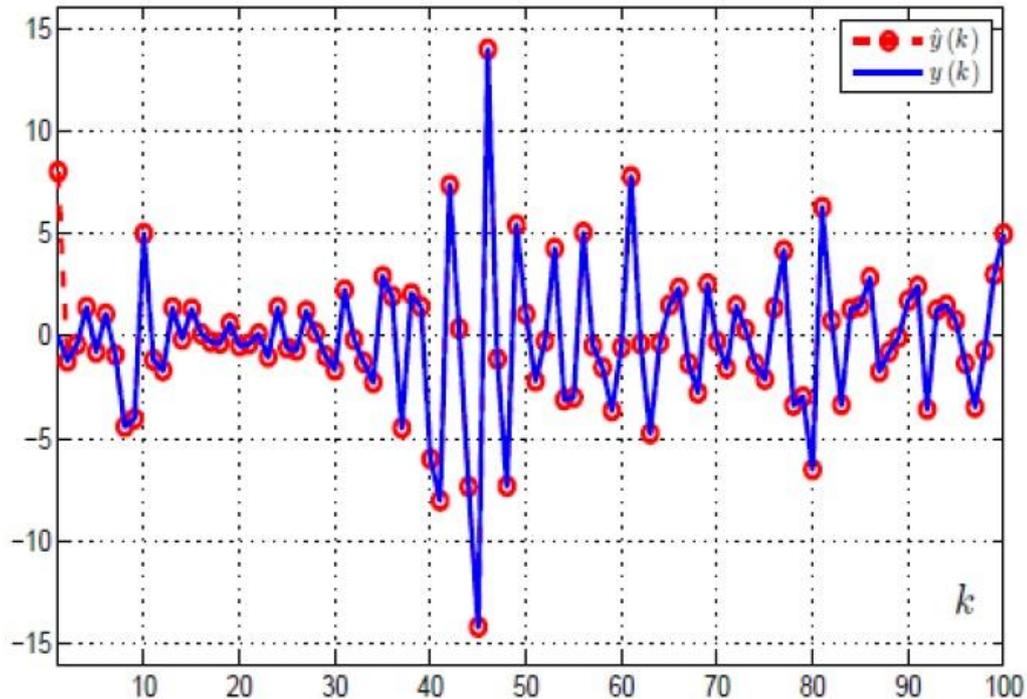


FIGURE 2-26: EVOLUTION DE LA SORTIE DU MODELE PROPOSE $\hat{y}(k)$ ET DE LA SORTIE DU SYSTEME $y(k)$.

4. Classification dynamique pour le diagnostic des systèmes automatiques

4.1 Le diagnostic des systèmes

Pour garantir une bonne sûreté de fonctionnement d'un système dans un milieu industriel, il convient de mettre en œuvre un processus de diagnostic convenable. Ce qui permet la détection d'un dysfonctionnement et contribue à éviter toute dégradation pouvant toucher le système à surveiller ou tout élément de son environnement (utilisateurs et matériels).

Les systèmes industriels actuels sont souvent multimodes (opèrent en plusieurs modes). Outre cette complexité, les modes et les zones de fonctionnement sont aussi susceptibles d'évoluer et subir des progressions. Par conséquent, tout modèle de connaissances adopté doit tenir compte de ces situations liées au non stationnarité des données.

Plusieurs chercheurs ont exploré le domaine pour proposer des solutions fiables à cette problématique. Dans sa mémoire de thèse [4] Foued a regroupé les approches existantes en trois familles :

✓ Les approches fondées sur les connaissances : appelées aussi approches relationnelles, elles sont basées sur les connaissances d'un expert sur le système, les défauts et leurs effets. Dans cette famille d'approches on trouve essentiellement l'approche AMDE (l'Analyse des Modes de Défaillances, de leurs Effets et de leur Criticité) [175] en anglais AMDEC (Failure Modes, Effects and Criticality Analysis). Cette méthode consiste à énumérer les défaillances et définir leurs causes, leurs conséquences et leurs criticités. Cette approche ne convient pas aux procédés complexes avec un nombre de pannes indéterministe, puisque elles nécessitent une énumération de toutes les combinaisons possibles des anomalies.

On peut citer aussi dans la même famille, l'approche Arbre de défaillances [176]. Elle consiste à déterminer, d'une manière hiérarchique, les enchaînements des causes pouvant conduire à un événement prédéfini

✓ Les approches à base de modèle : à partir d'une connaissance des lois physiques internes et externes qui régissent le système à surveiller, ces approches se basent sur des modèles mathématiques de comportement, tel que ceux présentés dans les sections précédentes. Les défauts de fonctionnement sont détectés et parfois localisés par le calcul de l'écart entre les observations et les prédictions faites par ces modèles.

On peut citer l'approche multi-modèle [177] [178] [179] [180]. Elle se base sur une décomposition du fonctionnement du système global en le modélisant par des sous modèles simples. Dans cette approche, le modèle unique est remplacé par une bibliothèque de modèles facilitant l'exploitation.

✓ Les approches de traitement de données et de Reconnaissance De Formes RDF : à partir d'un historique de données ou des signaux prélevés du système à surveiller, Ces approches, basées sur l'analyse qualitative, effectuent une segmentation temporelle de ces données en des unités appelées épisodes [181] en fonction de leur pente et de leur convexité. Ces épisodes seront classés selon un critère de similarité. Et delà vient l'idée d'appliquer les techniques de reconnaissance de formes tel que le k-means, les réseaux de neurones artificiels, l'algorithme de Kohonen,... [182] et [183]. Dans ce cadre, j'ai eu l'occasion d'encadrer deux masters [8] et [16] qui ont porté sur la surveillance des systèmes (entre autres multi-modèles) en se basant sur la technique SVM et d'autres techniques de classification.

Vers les nouvelles approches

Nous avons cité pour le diagnostic des systèmes des méthodes quantitatives à base de modèles, des méthodes relationnelles et des méthodes qualitatives basées sur le traitement de données et de RDF. On peut constater que ces approches ne sont pas adaptées à tout type de système et nous avons passé des approches fondées sur les connaissances et qui ne conviennent pas aux procédés complexes vers les approches à base de modèle. Cependant, à cause de la complexité de certains processus, une représentation mono-modèle de toute la plage de fonctionnement s'avère difficile et cette solution a été remplacée par l'approche multi-modèle. Néanmoins, Cette dernière n'est pas conçue pour détecter les modes imprévisibles et qui n'admettent pas des modèles associés dans la bibliothèque.

Le problème de diagnostic peut être perçu comme un problème de classification puisqu'il s'agit d'attribuer les observations aux modes de fonctionnement normaux ou défaillants. Beaucoup de travaux ont traité le diagnostic industriel de cet abord [184], [185]. Comme je l'ai présenté dans les sections précédentes, ces méthodes ont montré une capacité importante de s'adapter à la distribution des données disponibles. Elles imposent moins de contraintes dans le sens où les connaissances de maîtrise du système peuvent être minimales.

4.2 Classification par enveloppe convexe pour le diagnostic

Dans le contexte du diagnostic des systèmes, nous avons proposé, dans le cadre de la thèse de Foued [4], un algorithme de classification séquentielle basé sur le calcul dynamique de l'enveloppe convexe. Dans cette approche, une classe est représentée par les sommets de l'enveloppe convexe englobant les données. Cette enveloppe limite approximativement la région occupée par la classe. Elle s'agit d'une nouvelle modélisation de la classe différente de la modélisation par le centre de gravité tel que le K-means ou la modélisation par les axes principaux tel que le ACP ou encore la modélisation par la dispersion interne tel que l'algorithme de Fisher.

4.2.1 Algorithme DCH : Calcul de l'enveloppe convexe

Cet algorithme de classification automatique incrémentale, que nous avons appelé DCH (Dynamic Convex Hull), a été conçu pour le calcul de l'enveloppe convexe dynamique d'un nuage de points. Son principe est le suivant :

Soit $CH_{c,k}$ l'enveloppe convexe d'une classe $C_{c,k}$ à l'itération k . avec $c = 1, \dots, N_c$ et N_c désigne le nombre total des classes. Un nouveau point p_k arrivant en ligne à l'itération k est classé dans $C_{c,k}$ s'il vient de se situer à l'intérieur de l'enveloppe $CH_{c,k}$, sinon, il fait partie d'une classe distincte. Le point p_k se situe à l'extérieur de $CH_{c,k}$ s'il existe deux lignes de support à $CH_{c,k}$ qui passent par p_k .

L est une ligne de support d'un ensemble S de points dans le plan si elle passe par un point de S et ne fait aucune séparation entre deux autres points. Autrement dit, L passe au moins par un point de S tel que S reste entièrement dans l'une des deux demi-plans définis par L.

La localisation d'un point dans la pratique se fait avec un processus utilisant une modélisation mathématique plus claire. On suppose que $CH_{c,k}$ est formée par un nombre fini de n sommets $v_{c,i}$, tels que $CH_{c,k} = \{v_{c,1}, \dots, v_{c,i}, \dots, v_{c,n}\}$ ($i = 1, \dots, n$). Pour prouver que le point p_k est à l'extérieur de l'enveloppe, il suffit de vérifier l'existence d'un sommet $v_{c,i}$ de l'enveloppe convexe menant à ce que le segment $\overline{p_k v_{c,i}}$ soit une ligne de support. Le point $v_{c,i}$, dans ce cas, est dit un sommet de support. le segment $\overline{p_k v_{c,i}}$ est une ligne de support ($v_{c,i}$ est un sommet de support) si tous les sommets restants $v_{c,j}$ ($i, j = 1, \dots, n, i \neq j$) sont dans le même demi-plan délimité par ce segment. Dans ce cas la séquence $\overrightarrow{p_k v_{c,i} v_{c,j}}$ conserve la même orientation $\forall v_{c,j}$. Cette orientation peut être déterminée par le signe du déterminant de la matrice d'orientation O qui est définie par :

$$O = \begin{pmatrix} 1 & p_k \\ 1 & v_{c,i} \\ 1 & v_{c,j} \end{pmatrix} \quad 2.73$$

Avec $p_k, v_{c,i}, v_{c,j}$ sont des vecteurs lignes à deux dimensions.

$$\begin{cases} \det(o) < 0 \leftrightarrow \text{sens horlogique} \\ \det(o) > 0 \leftrightarrow \text{sens antihorlogique} \end{cases} \quad 2.74$$

Donc pour que le segment $\overline{p_k v_{c,i}}$ soit une ligne de support, il faut que le déterminant de la matrice d'orientation de la séquence $\overrightarrow{p_k v_{c,i} v_{c,j}}$ conserve le même signe quelque soit $v_{c,j}$.

Si le nouveau point arrivant p_k est situé à l'extérieur de l'enveloppe convexe, on effectue une adaptation dynamique de cette enveloppe en fonction de l'ajout du point (voir Figure 2-27). On recommence par déterminer un deuxième sommet de support $v_{c,j}$. Ensuite, on élimine les points situés à l'intérieur du triangle formé par les trois points $p_k, v_{c,i}$ et $v_{c,j}$, en appliquant la méthode des coordonnées barycentriques

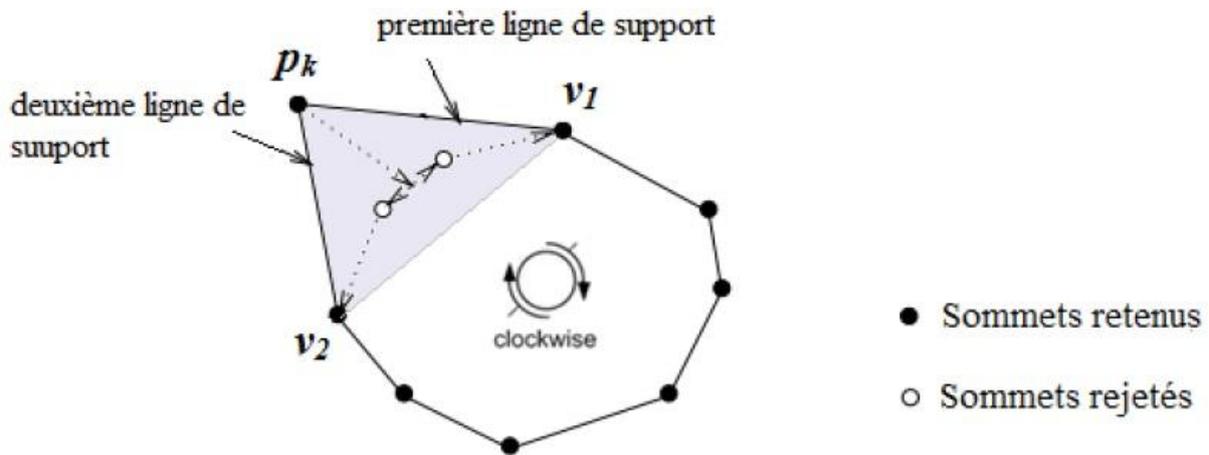


FIGURE 2-27: MAINTIEN DYNAMIQUE DE L'ENVELOPPE CONVEXE.

Le pseudo code suivant décrit les différentes étapes de la classification avec DCH

Algorithme : Dynamic Convex Hull

Entrées L'enveloppe convexe $CH_{c,k}$ d'un ensemble de données, un point p_k à classer.

Etape.1 Localisation

Chercher le premier sommet de support v_1 de $CH_{c,k}$

- 1 : si v_1 existe (p_k est à l'extérieur de $CH_{c,k}$), faire
 - Chercher le second sommet de support v_2
 - Retourner v_1 et v_2
 - Exécuter **Etape.2**
- 2 : si v_1 n'existe pas (p_k est à l'intérieur de $CH_{c,k}$), faire
 - Maintenir l'enveloppe convexe : $CH_{c,k+1} = CH_{c,k}$
 - Répéter **Etape.1** pour traiter le point suivant p_{k+1}

Etape.2 Restructuration

- 1 : Supprimer les sommets de $CH_{c,k}$ se situant dans le triangle formé par les points p_k, v_1, v_2
Retourner $S_{remain} = \{CH_{c,k}\} \setminus \{S_{reject}\}$
 - 2 : Maintenir l'enveloppe convexe : $CH_{c,k+1} = \{S_{remain}\} \cup \{p_k\}$
 - 3 : Répéter **Etape.1** pour traiter le point suivant p_{k+1}
-

4.2.2 Algorithme DCH : Classification en ligne

Le calcul de l'enveloppe convexe, en temps anticipé, est une phase de limitation géographique des points de la même classe. Elle remplace la phase d'apprentissage dans les autres approches de classification. Après avoir déterminé les frontières de chaque classe, on propose un autre processus de classification des nouveaux points. Il s'agit d'une classification en ligne ou on s'appuie sur la variation de la densité pour décider de l'appartenance des données aux classes. En effet, un point réside à l'intérieur d'une enveloppe convexe $CH_{c,k}$ si la densité au sein de la classe $C_{c,k}$ augmente par rapport à l'ancienne valeur, et vice-versa.

La densité est exprimée par le nombre de population par unité de surface, elle peut être déterminée selon l'expression :

$$D_{c,k}^m(C_{c,k}, \theta_m) = \frac{N_{c,k}}{A_{c,k}^m(C_{c,k}, \theta_m)} \quad 2.75$$

Nous avons utilisé la formule de Shoelace pour le calcul de l'aire des polygones simples [186], sachant que chaque classe on le remplace par une combinaison de plusieurs projection sur des sous-espaces différentes Θ_m . La formule est décrite comme suit :

$$A_{c,k}^m(C_{c,k}, \theta_m) = \frac{1}{2} \left| \sum_{i=1}^{n-1} v_{i,x} v_{i+1,y} - v_{i+1,x} v_{i,y} \right| \quad 2.76$$

Avec $A_{c,k}^m(C_{c,k}, \theta_m)$ est l'aire délimitée par l'enveloppe convexe planaire $CH_{c,k}^m$

La densité globale $Dg_{c,k}$ est par la suite estimée sur la somme de toutes les densités marginales, tel que :

$$Dg_{c,k} = \sum_{m=1}^{N_p} D_{c,k}^m(C_{c,k}, \theta_m) \quad 2.77$$

Pour une nouvelle observation à classer, on calcule les densités des classes et l'ensemble de classes gagnantes C_{win} sera :

$$C_{win} = \{C_{c,k} | Dg_{c,k} - Dg_{c,k-1} > T_{adapt}\} \quad 2.78$$

Avec

$$T_{adapt} = \frac{T_d}{\sigma_{c,k}} \quad 2.79$$

T_d est un seuil prédéfini et $\sigma_{c,k}$ est l'écart type calculé sur la distribution de données de la classe $C_{c,k}$.

4.3 Nouvelle approche de classification dynamique par SVDD

Pour le diagnostic des systèmes industriels, les classes illustrant explicitement les modes de fonctionnement peuvent être en perpétuelle évolution. Ceci fait appel à un autre type de classification, qui est la classification dynamique. Selon Boubacar, qui proposée les algorithmes AUDyC et SAKM [124], la problématique de la classification dynamique se résume en trois principaux aspects à savoir : la classification séquentielle, la classification non supervisée et la modélisation adaptative. On peut aussi résumer la distribution dynamique des données en des phénomènes qui concernent particulièrement l'apparition des nouvelles classes, l'évolution des classes et la fusion des classes.

Bien qu'elle effectue une classification incrémentale, l'approche DCH, proposée dans la section précédente, n'est pas parfaitement adaptée aux données stationnaires. Dans le cadre du même travail de thèse de Foued [4], nous avons proposé une nouvelle approche de classification dynamique basée sur la technique SVDD, proposée par [187] et [188].

4.3.1 Classification par SVDD

Il s'agit d'une méthode à noyau pour la classification mono-classes (one-class). Utilisée pour la détection de nouveauté (Novel detection problems), Cette méthode cherche la plus petite hyper-sphère englobant les instances d'une même classe. Le principe de l'apprentissage consiste alors à résoudre le problème d'optimisation suivante :

$$\min R^2 + C \sum_{i=1}^N \xi_i \quad 2.80$$

Sous contrainte

$$\begin{aligned} (x_i - a)^T (x_i - a) &\leq R^2 + \xi_i, \\ \xi_i &\geq 0 \quad \forall i = 1, \dots, N, \end{aligned} \quad 2.81$$

Avec R et a sont respectivement le rayon et le centre de l'hyper-sphère, C est le paramètre de régularisation et ξ_i est le variable de relaxation. La formule lagrangienne est alors :

$$L = R^2 + C \sum_i \xi_i - \sum_i \alpha_i (R^2 + \xi_i - (x_i - a)^T (x_i - a)) - \sum_i \gamma_i \xi_i \quad 2.82$$

Avec α_i et γ_i sont les multiplicateurs de Lagrange. En remplaçant les conditions de KKT dans l'expression de L , on obtient le problème dual suivant :

$$\max \frac{1}{2} \sum_{i=1}^N \alpha_i (x_i^T x_i) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j (x_i^T x_j), \quad 2.83$$

Sous contrainte

$$\sum_{i=1}^N \alpha_i = 1, \quad 0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, N. \quad 2.84$$

Selon les conditions de KKT

$$a = \sum_{i=1}^N \alpha_i x_i \quad 2.85$$

Une nouvelle instance z_t appartient à la classe en question si sa distance du centre a est inférieur ou égale au rayon R . sachant que la distance est calculée de la façon suivante :

$$D^2(z_t, a) = (z_t^T z_t) - 2 \sum_{i=1}^N \alpha_i (z_t^T x_i) + \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j (x_i^T x_j) \quad 2.86$$

Si $R^2 - D^2 \geq 0$ alors z_t appartient à la classe.

Pour les cas les plus complexes et non linéaires, on utilise le même principe que le SVM non-linéaire avec une projection dans un espace reproduisant RKHS et le produit scalaire sera remplacé par une fonction noyau.

4.3.2 Nouvelle version pour la classification dynamique

Le principe de la version proposée pour la classification dynamique consiste à incorporer chaque nouvelle instance, arrivant en ligne, dans l'ensemble de données. Il s'agit ensuite d'adapter les modèles de connaissances sans refaire le processus d'apprentissage à zéro. Deux cas se présentent pour la nouvelle instance.

1^{er} Cas : la distance du centre est inférieure ou égale au rayon : dans ce cas rien ne change et l'observation peut être considérée dans la classe.

2^{ème} Cas : la distance du centre est supérieure au rayon : dans ce cas la nouvelle observation est susceptible d'être un nouveau vecteur de support et l'ensemble d'apprentissage $S_{training}$ doit

être enrichi pour inclure le nouvel échantillon pour améliorer les résultats du problème d'optimisation. Chaque opération d'ajout appliquée sur l'ensemble d'apprentissage doit être suivie par une opération de suppression des données obsolètes (inutiles).

La suppression

Puisqu'un échantillon perd son importance tout en s'éloignant de la zone de haute densité, on supprime, à chaque insertion, le support vecteur qui vérifie la condition suivante :

$$SV_{del} = \arg \max_{x_i \in SV_s, i = 1, \dots, N_{SV_s}} (sign(\Delta d) * ||z_t - x_i||), \quad 2.87$$

Avec a_t et a_{t+1} respectivement le centre actuel de l'hyper-sphère décrivant le domaine de données et le futur centre quand z_t devient probablement un nouveau vecteur de support. Le centre a_{t+1} peut être simplement prédit par la relation récursive suivante :

$$a_{t+1} = a_t + \frac{1}{N_{SV_s} + 1} (z_t - a_t) \quad 2.88$$

Et Δd est la variation de la distance exprimée comme suit :

$$\Delta d = d_{t+1} - d_t, \quad 2.89$$

Avec d_t et d_{t+1} la distance entre un élément support de vecteur x_i et les centres a_t et a_{t+1} .

L'insertion

Un vecteur x_i de l'ensemble NSVs (Non-Support Vectors) peut être un futur vecteur de support s'il vérifie la condition suivante :

$$\sqrt{D^2(x_j, a_t)} \geq T, j = 1, \dots, N_{NSV} \quad 2.90$$

Avec T est un seuil de décision fixé à l'avance en fonction de la dispersion de données.

Cette approche a été adaptée à la classification multi-classe en se basant sur le degré de proximité (ou voisinage) entre une observation reçue en ligne et les classes de la partition. Puisque les classes sont susceptibles d'évoluer dans l'espace et de se chevaucher. Ainsi, une procédure de fusion est introduite.

4.4 Validation

Pour tester l'approche de classification dynamique sur des cas réels de diagnostic, nous l'avons appliqué sur deux systèmes différents pour différents cas de pannes. Pour les deux premières expérimentations, nous avons considéré un système hydraulique disponible au sein du laboratoire ACS et destiné à la validation des approches d'asservissement, Alors que pour d'autres cas, nous avons utilisé un procédé chimique nommé Tennessee Eastman Process (TEP) Nous avons pu évaluer les cas suivants :

✓ SVDD mono-classe : cas d'un système hydraulique

Nous avons commencé par une phase d'apprentissage en ne considérant qu'une classe. Il s'agit de la classe de fonctionnement normal. Nous avons essayé, par la suite (en temps réel de fonctionnement), d'assigner chaque observation arrivant à cette classe si non elle est considérée "outlier" et donc un point de fonctionnement défaillant. Nous avons utilisé le noyau gaussien et le processus de suivi parvient à identifier instantanément les différentes pannes au moment où elle est produite et la méthode prouve une considérable capacité de détection.

✓ SVDD multi-classe : une approche multi-modèle

Pour le système TEP, nous avons considéré deux modes de fonctionnement différents et pour chaque mode, nous avons associé une panne. Donc nous avons au total 4 classes. La transition d'un mode à un autre se fait par commutation, alors que la transition d'un modèle à un autre est une transition progressive lente. La Figure 2-28 suivante montre les 4 zones de fonctionnement et les transitions entre les modes.

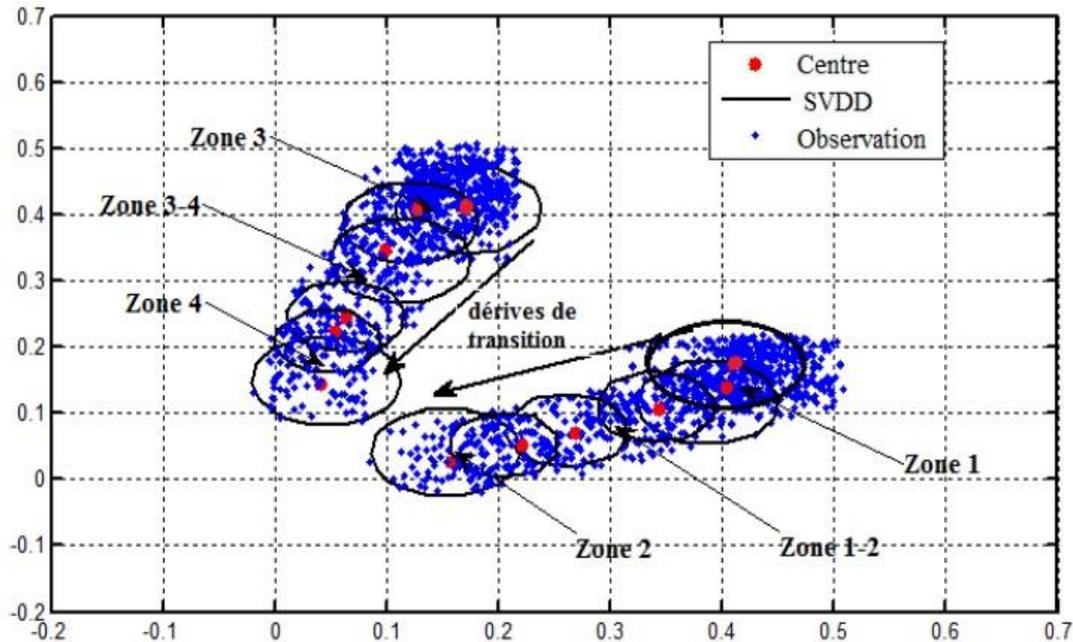


FIGURE 2-28: LES DIFFERENTES ZONES DE FONCTIONNEMENTS ESTIMES PAR LE CLASSIFIEUR SVDD.

✓ SVDD multi-classe : cas du TEP

Nous avons tenu compte uniquement de trois types de défauts parce qu'ils représentent un cas idéal des classes chevauchées et inséparables. Notre Approche, noté ici NS-SVDD, a été comparée au deux autres approches à savoir le LDA et le QDA sur trois différents espaces de projection ; un espace de dimension 52 contenant toutes les variables, un espace de trois dimensions contenant les variables {51,9,21} et un espace bidimensionnel contenant les variables {51,9}. Le Tableau 2-15 montre que l'approche proposée est plus efficace dans le cas de l'espace de haute dimension. Elle est aussi meilleur que le LDA dans tous les cas et pas loin de l'approche QDA.

TABLEAU 2-15: TAUX DE CLASSIFICATION DE NS-SVDD COMPAREE AUX AUTRES METHODES

Méthode	Erreur de classification (%)			Non-stationnaire
	All	{51, 9, 21}	{51,9}	
LDA	58.21	32.29	30.75	Non
QDA	60.75	7.33	7.38	Non
NS-SVDD	58	13.91	13.62	Oui

5. Conclusion

Dans ce chapitre, j'ai présenté des nouvelles approches basées sur la théorie d'apprentissage statistique pour l'identification et le diagnostic des systèmes automatiques complexes. Dans une première partie qui concerne les systèmes non linéaires, une première méthode a été

dévoilée qui applique une variante modifiée de la technique SVM standard. Il s'agit de l'approche LS-SVM multi-kernel à deux noyaux, faisant appel à l'algorithme d'optimisation OEP pour le réglage des paramètres. Puis en se basant sur une subdivision du vecteur observation (régression) cette approche a été modifiée et nous avons proposé l'approche LS-SKM multi-kernel. Validées sur des processus automatiques notamment le système de type NARMA, les deux méthodes ont donné des résultats attractifs.

Dans ce même volet d'identification des systèmes non-linéaires, le chapitre présente l'approche SKR multi-noyau qui découle de la méthode de régression SVR.

La deuxième partie a été consacrée à l'identification des systèmes dynamiques hybrides. Le même principe théorique a été adapté pour une classification dynamique pour déterminer le mode de fonctionnements de ses systèmes commutant. Pour suivre la commutation rapide entre les modes, les méthodes d'identification adoptées doivent être dotées des facultés d'apprentissage séquentiel (ou incrémental) avec une bonne capacité d'auto-adaptation. Ce travail d'identification a été précédé par une phase de détection de l'instant de commutation dans laquelle on applique aussi une classification SVM.

Dans la troisième partie de ce chapitre, j'ai présenté ma contribution dans le diagnostic des systèmes automatiques complexes. Encore une fois, les approches de classification ont montré leur capacité d'adaptation. Cette fois, j'ai pu améliorer l'orientation de l'équipe vers les données stationnaires tout en gardant des objectifs industriels pratiques pouvant motiver les jeunes chercheurs à aller plus loin dans ce volet.

Ces travaux ont été élaborés, en partie, dans le cadre de deux collaborations dont la première, qui concerne l'identification des systèmes non linéaires et le diagnostic, est avec le laboratoire d'Analyse, Conception et Commande des Systèmes (ACS) de l'Ecole National d'Ingénieurs de Tunis (ENIT). Et la seconde collaboration, portant sur l'identification des systèmes dynamiques hybride, était avec l'unité de recherche "COMmande Numérique des PRocédé Industriels" (CONPRI) l'Ecole national d'Ingénieurs de Gabes (ENIG).

Chapitre 3 : Classification pour la gestion du risque médical

1. Introduction

L'amélioration de la sécurité des soins est une préoccupation importante des systèmes de santé. Dans les établissements, il peut arriver des événements précurseurs ou indésirables qui sont parfois à l'origine de certains accidents. La survenue de quelques accidents graves dans différents centres hospitaliers, a montré que les procédures suivies ainsi que les moyens mis en œuvre doivent être complétés. En effet, les techniques de traitement devenant aujourd'hui de plus en plus complexes avec des machines assimilables de plus en plus à des "boîtes noires". Il devient donc nécessaire d'appliquer désormais les principes fondamentaux de sûreté, tels que ceux utilisés dans les installations nucléaires. L'un des plus importants principes est la déclaration et l'archive de tout événement susceptible d'entraîner une anomalie ou un accident. Les facteurs temps et précisions sont à prendre en compte dans ce type de problèmes. Il est donc important, voir obligatoire, de passer à des moyens informatiques qui permettent le traitement automatique de ces déclarations. En tant que responsable R&D à Biomedica, j'ai dirigé le développement d'une plateforme de gestion des événements précurseurs que nous avons appelé RMP (Risk Management Platform). J'ai intégré, par la suite, une partie intelligente de traitement automatique de données en se basant sur des approches : d'extraction, d'apprentissage et de classification. Dans ce même contexte, la première partie de ce chapitre présente la gestion intelligente des événements précurseurs dans les centres de soins.

La deuxième partie est consacrée à la gestion d'un autre risque aussi important que le premier. Il s'agit de la gestion du risque lié aux épidémies. Dans ce contexte, je propose ma contribution qui consiste à une nouvelle méthode de détection automatique des épidémies. Cette méthode permet de classer des données médicales dont le but est de prédire la survenue d'une épidémie. Dans une première phase, on prépare les données, en se basant sur une analyse formelle de concepts. Puis, le processus sera complété par une classification semi supervisée basée sur l'apprentissage statistique.

2. Gestion des événements précurseurs ou indésirables

2.1 Évènement indésirable

Un événement indésirable ou précurseur est un événement non prévu, dans un centre hospitalier, qui aura entraîné une anomalie dans le fonctionnement de la structure de soin, pouvant notamment nuire à la prise en charge des patients, causer des dommages aux personnes (patients ou personnels) et à leurs biens ou ceux de l'établissement, voir même engager un contentieux avec les patients ou leurs familles. L'évènement indésirable peut être relié à un ou plusieurs facteurs plus ou moins influents, à l'origine de la survenue de l'incident.

Par exemple, un couloir de l'hôpital fraîchement nettoyé avec une petite zone restée mouillée par oubli, peut causer des accidents de chute des personnes âgées. Ils peuvent être relativement graves. Ainsi, l'acte d'oubli est considéré comme un événement précurseur même dans le cas où il n'y a pas eu d'accident.

Les barrières de lit (garde-corps) des bébés doivent être en position haute (fermées) tant que le cadre médicale n'est pas présent à ses côtés. Oubliées en position ouverte est une situation qui peut causer des accidents graves. Elle doit être signalée comme événement indésirable.

Le secteur de la radiothérapie est le plus concerné par cette problématique. En effet, le risque peut se manifester rapidement à cause des rayons et la maîtrise des risques dans ce domaine implique une très grande rigueur en raison de :

- ✓ l'utilisation de très fortes doses (20 à 80 Gy) intentionnellement délivrées au "volume-cible clinique" ;
- ✓ l'utilisation d'un faisceau de rayonnement directement focalisé sur le patient ou bien des sources radioactives placées au contact des tissus pour lesquels toute erreur significative dans le positionnement peut avoir des conséquences graves ;
- ✓ nombreuses étapes et de nombreux intervenants pour réaliser le traitement depuis sa prescription jusqu'à la délivrance de la dose ; de la mise en œuvre de technologies complexes et sophistiquées (calculs par ordinateur et transferts de données) en parallèle d'activités manuelles (préparation de la protection de certains organes, de dispositifs d'immobilisation...).

Une erreur de saisie sur le dossier d'un patient peut provoquer un accident de sur-irradiation. Cette erreur doit être aussi déclarée en tant qu'évènement précurseur.

2.2 Etude contextuelle

2.2.1 Bilan et constat en France

Dans la plupart du pays du monde, les évènements précurseurs ou indésirables dans les établissements de soin sont nombreux. Mais, le nombre des accidents survenus à cause de ces évènements et les dommages causés par ces anomalies varient d'un pays à l'autre selon plusieurs facteurs notamment organisationnels.

En France, le bilan a été très lourd surtout dans le domaine de la radiothérapie. Il y a eu quatre accidents graves ayant affecté des patients depuis 2005. Ces accidents de radiothérapie externe étaient en 4 villes différentes (Grenoble, Lyon, Tours et Epinal). Ils ont été la cause de complications extrêmement sévères. En 2008 et dans une visite aux victimes de sur-irradiation du service de radiothérapie de l'hôpital Jean-Monet d'Epinal, la ministre de la santé de l'époque a annoncé un premier bilan de 5500 personnes traitées entre 1987 et 2006 qui ont été victimes de cet accident.

Le ministère a dressé une feuille de route en 6 étapes dont l'amélioration de la sécurité des installations, le renforcement de la qualité et de la sécurité des pratiques et le renforcement du programme d'inspection de l'ASN (Autorité de Sûreté Nucléaire). Cette dernière, après avoir assuré la transposition en droit français des dispositions communautaires relatives à la radioprotection des patients, a été amenée à renforcer ses actions de contrôle dans ce domaine et à mettre en place une procédure de déclaration des évènements significatifs ainsi qu'une échelle expérimentale de communication à destination du public sur ces évènements significatifs. Il s'agit de définir un processus de traçabilité.

Il convient à cet effet, de développer une culture de sûreté dans les services : l'erreur humaine ne doit pas conduire à des mises en causes individuelles, mais au contraire à la prise de conscience d'une responsabilité collective par rapport à celle-ci. Même si le mauvais geste est commis par une personne, l'analyse « facteur humain et organisationnel » conduit à présupposer que l'origine est plus complexe (environnement, manque de formation, manque d'ergonomie de tel ou tel appareil etc.). C'est l'analyse de ces « petits évènements », qui doit et de la sécurité des soins. Cette instance transmet le cas échéant la déclaration à l'autorité compétente et organise les actions d'amélioration au sein de l'établissement de santé. Restent à définir les règles d'anonymisation et de confidentialité, en fonction de la gravité des

conséquences sur le patient ; elles sont garanties par exemple dans le dispositif d'accréditation.

2.2.2 Etat de l'art sur la gestion des risques

En France, la gestion des risques médicaux, surtout en radiothérapie a été gérée par le renforcement de contrôle de l'ASN [189] [190] et par l'écriture d'un référentiel d'assurance qualité effectué par des organismes de l'état tel que l'ASN, l'Institut National du Cancer (INCA) et la Haute Autorité de santé (HAS). Ces organismes avec l'Etat ont ciblé des approches sur les facteurs organisationnels et humains [191] [192] [193]. L'accident d'Epinal est le plus important événement déclencheur d'une politique renforcée de prévention des risques, d'amélioration de la connaissance des incidents et accidents dans le domaine de la radiothérapie en France. Il a été un des éléments déclencheur de la révision d'une réglementation sécuritaire dans le domaine de la Radiothérapie et la mise en place de mesures correctives [194] [195] [196] [197]. Les praticiens ont proposé quelques solutions notamment organisationnelles telles que la création des cellules de retour d'expériences CREX.

La gestion des risques est un sujet qui a pris sa part dans la recherche scientifique et qui touche plusieurs secteurs. Le domaine de l'aéronautique est le plus concerné par ce sujet et aussi le plus proche à notre contexte de gestion du risque en médical [198]. Ils ont même mis en place un standard proposé par Flight Safety Foundation [199] et ils ont travaillé sur cette problématique depuis des années avec l'idée de retour d'expérience [200]. Les mêmes approches utilisées par les chercheurs en aéronautiques [201] ont été appliquées en médecine notamment en chirurgie, en médecine nucléaire et en hygiène hospitalière [202]. D'autres domaines ont fait aussi l'objet de quelques études et travaux de recherche comme la sécurité dans le transport ferroviaire [203].

2.3 Vers une solution informatique de traçabilité

La traçabilité (industriellement parlant) est régie par la norme ISO 9001:2000 et a vu son arsenal technologique considérablement étoffé depuis le début du 20ème siècle. [204] [205] [206]. Cette traçabilité commence à devenir une branche spécialisée de l'ingénierie, du fait de sa transversalité et de la complexité des moyens technologiques mis en œuvre. Dans un tel contexte, la législation de la traçabilité s'est renforcée dans tous les domaines. [207], [208]. [198]

Technologiquement, un système de traçabilité est (généralement) supporté par une infrastructure informatique et doit être alimenté soit par des opérateurs humains soit par des capteurs externes, voire même une combinaison des deux. [208].

Dans le cas qui nous préoccupe, il s'agit de tracer les actions des professionnels de santé dans différents hôpitaux. Ce traçage répond indirectement à un besoin législatif de signalement des incidents dans les plus brefs délais. Un traçage contextualisé ou mappabilisé, une fois analysé pourrait permettre l'anticipation, c'est-à-dire la prévention des risques en amont ou alors par l'analyse a posteriori d'incidents ou d'accidents, une amélioration des connaissances collectives et une gestion des risques par le retour d'expérience qui profiterait à tous les patients. On retient que :

- il existe au moins une modélisation d'analyse des risques à priori [206] [209], mais que cette modélisation requiert un panel d'experts pour l'évaluation de ces risques.
- la gestion des risques est d'autant plus compliquée que l'organisation l'est aussi ; auquel cas, il faut procéder à une analyse organisationnelle complexe [210] [211].

- la gestion des risques est fortement influencée par les facteurs humains [212] [213]; les facteurs humains sont à la fois la force et la faiblesse d'un système de gestion des risques
- la tendance est à la mise en place de cellules dites de retour d'expérience [214]. Ces cellules ont pour but d'analyser et de décider des mesures correctives à mettre en place face à une situation donnée. Ces cellules ont pour charte le principe de non-sanction qui permet le recueil d'informations critiques de manière volontaire.

Aujourd'hui, au vu de l'augmentation du nombre de plaintes des patients et suite aux nouvelles dispositions des organismes institutionnels, la pression se fait de plus en plus forte sur les établissements de soins, quant à la gestion des risques et des situations accidentelles. Ainsi, le recueil de ces événements devient donc indispensable, notamment dans un contexte d'amélioration continue de la qualité.

La société BIOMEDIQA, dans une démarche de recherche et de développement, a cherché à répondre aux attentes de chacun, en fournissant aux professionnels de santé les moyens de prendre pleinement conscience des risques présents dans leur établissement et de mettre en œuvre les dispositions qui s'avèrent, afin de réduire ces risques.

Ce recueil est fondé sur la participation de chaque acteur de l'établissement concerné, quelle que soit sa fonction, ayant un rôle auprès du patient. Il vise à répertorier les événements indésirables déclarés resitués dans leur contexte, et permettre de les qualifier, c'est-à-dire à identifier les événements indésirables graves (mortels ou engageant le pronostic vital ou conduisant à une invalidité ou une incapacité persistante et significative) et, parmi eux, ceux qui sont jugés évitables. Il est également nécessaire de quantifier l'événement indésirable, en terme de gravité, mais aussi de durée ; ressortir un lien de causalité entre l'événement lui-même et le contexte dans lequel il est survenu ; enfin de mentionner les mesures immédiatement prises après la survenu de l'événement indésirable.

A partir de là, les déclarations pourront être classées, regroupées et hiérarchisées en fonction de leur gravité, de leur fréquence, dans le but de définir et évaluer l'action corrective permettant de limiter le ou les risques concernés. Cet outil informatique a pour objectif d'amener les professionnels d'un établissement de santé à interagir et mener une réflexion sur les événements recensés, afin de les admettre, de les comprendre et de définir les actions correctives à mettre en œuvre pour éviter qu'il ne se reproduisent.

L'obligation de déclaration des incidents issue du code de la santé publique participe à la diffusion du retour d'expérience, en externe cette fois : les erreurs et les progrès des uns doivent pouvoir bénéficier aux autres, de façon à faire progresser la profession collectivement. Et ce partage d'expertise est clairement difficile voire impossible avec des moyen traductionnels et sans dématérialisation des informations en papiers. Par contre, le progrès technologique en TIC a beaucoup simplifié la diffusion et le partage de l'expertise dans différents domaines surtout avec une orientation mondiale vers l'économie de la connaissance. Néanmoins, la grande quantité d'information et la complexité des systèmes actuels exigent le développement des approches plus intelligentes. Dans ce contexte, j'ai dirigé le développement d'un système électronique de déclaration des évènements précurseurs qui fera l'objet de la section suivante.

2.4 Gestion automatique des Evènements Indésirables

2.4.1 Le retour d'expertise

Le mot clé de la réussite de la gestion des événements indésirables est le retour d'expertise. En effet, il s'agit de l'objectif principal du processus de déclaration de ces événements. C'est une expression de la communauté des affaires : Il exprime l'amélioration des connaissances des employés par l'analyse de leur expérience afin de faire bénéficier toutes les unités de l'organisation. En fonction des succès ou de l'échec de l'expérience, elle tend à éviter ou à renouveler une situation du même type. En ce qui concerne le domaine médical, ont défini le retour d'expertise comme un processus organisé et systématique pour :

- ✓ Analyser toutes les anomalies et tous les incidents.
- ✓ Trouver toutes les causes et séquences liées.
- ✓ Analyser et dégager les diverses leçons.
- ✓ Définir des mesures de corrections et d'amélioration.
- ✓ Diffuser les informations pertinentes aux parties intéressées.

Par exemple, en radiothérapie la sécurité des traitements constitue un défi majeur. En effet, un mauvais traitement peut causer un accident grave y compris le décès du patient. Toutefois, avec une bonne qualité de contrôle, les grands incidents peuvent être évités. C'est dans cet esprit qu'un programme de recherche a été créé à Angers, Lille et Villejuif, donnant naissance aux premières cellules de retour d'expérience "Cellule de Retour D'Expérience (CREx)", [215] dont leurs objectifs principaux sont :

- Recueillir des événements précurseurs prioritaires.
- Surveiller l'analyse systémique effectuée chaque mois.
- Décider et assurer les mesures correctives.

Pour ces raisons et d'autres, il devient important de préparer des plateformes matérielles et logicielles pour assister les professionnels dans ce processus de gestion du risque médical.

2.4.2 Une nouvelle approche pour la gestion des EI

Dans cette section, je présente ma contribution dans la gestion des risques liés aux événements indésirables dans les centres de soins. Cette contribution consiste au développement d'un système informatique de signalement des événements précurseurs avec leurs analyses automatisées. Ce système nous l'avons conçu en se basant sur des nouvelles approches issues de l'intelligence artificielle. Une plateforme Web baptisée RMP (Risk Management Platform) a été développée et proposée à des établissements de soins pour déclarer les événements indésirables. Ces événements seront traités et analysés avec des algorithmes permettant le partage intelligent de l'expertise de chaque établissement et notamment des cellules CREX. Nous avons fait usage à des approches de classification et apprentissage automatique. Nous appliquons la technique SVM et d'autres techniques pour classer les événements selon les actions correctives attribués souvent par des experts.

Les systèmes de signalement des événements indésirables en médecine se répartissent en trois grandes classes. La classe I regroupe les systèmes passifs reposant sur la déclaration des acteurs médicaux, la classe II les systèmes passifs reposant sur la déclaration des patients et la classe III rassemble les systèmes actifs fondés sur l'analyse de dossiers ou de traces électroniques. L'objectif de notre système est de rassembler toutes ses déclarations ou classe pour une analyse automatique et plus intelligente.

Tel qu'il est présenté dans la Figure 3-29 suivante, notre système de gestion des événements indésirables est basé sur un module d'apprentissage de l'expertise humaine. A travers la plateforme Web RMP proposée par Biomedica, les professionnels déclarent toute sorte d'événements indésirables. Les déclarations seront enregistrées dans des bases de données stockées de leur part dans des serveurs sécurisés de Biomedica. Le système effectue une

vectorisation automatique des données. Cette première étape consiste à construire un vecteur représentatif de chaque déclaration. Alors une indexation ou extraction d'attribut est faite pour les parties textuelles. Les autres éléments numériques sont transcrits directement dans le vecteur. Cette première opération de vectorisation comporte deux étapes à savoir l'extraction textuelle et la réduction de dimension. Nous avons commencé par tester une première approche d'extraction textuelle. Il s'agit de la méthode n-grammes mot p-grammes caractère. Néanmoins l'existence d'un lexique médical MeSH laisse les portes ouvertes pour une indexation de textes à partir d'un lexique dédié à la gestion des risques dans le domaine médicale.

La deuxième étape consiste à la réduction de dimension. Nous pouvons appliquer pour ce traitement l'une des approches présentées dans le premier chapitre.

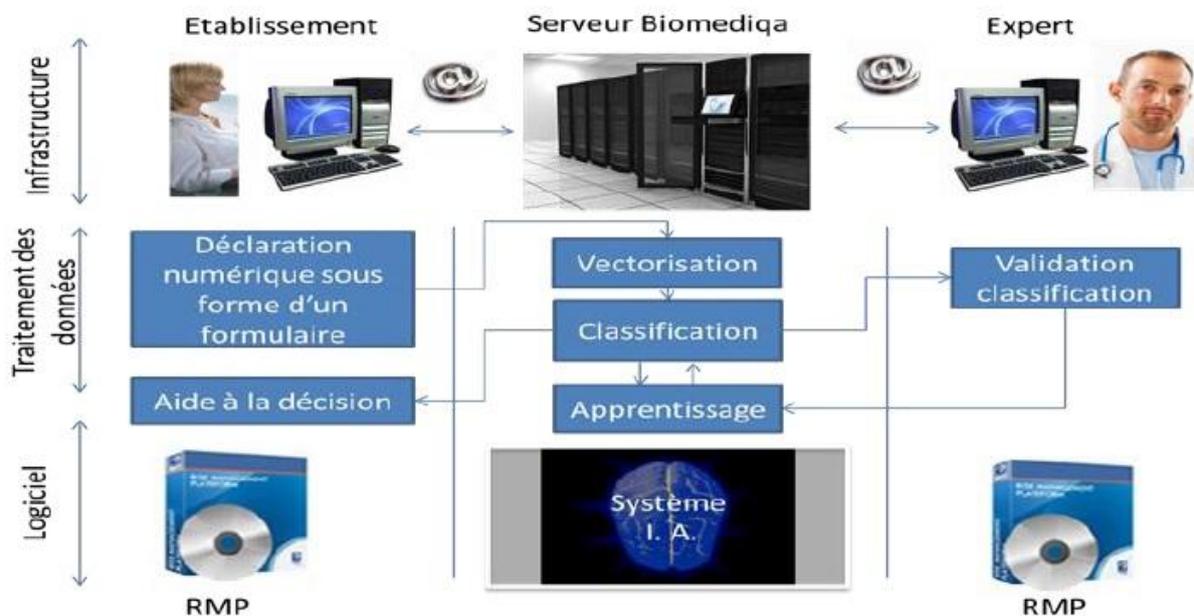


FIGURE 3-29: UNE NOUVELLE APPROCHE POUR LA GESTION DES EVENEMENTS PRECURSEURS

Pour toute déclaration effectuée sur la plateforme RMP le système commence par lui attribuer un vecteur numérique. Ce dernier sera classé par une approche de classification et d'apprentissage automatique. Une base d'apprentissage est alors préparée en temps anticipé. Elle s'agit d'une base d'exemples donnés par des experts médicaux. Ces professionnels valident aussi les résultats de classification de notre système. Les classes correspondent aux actions correctives. Au final le déclarant d'un événement indésirable sur la plateforme RMP sera assisté pour prendre des décisions correctives. Il s'agit bien d'un partage intelligent de l'expertise dans le domaine de la gestion de risque entre les établissements adhérents dans ce système. Sans échange de données en clair et en toute sécurité, les différents établissements peuvent bénéficier de ce partage d'expertise. Un événement indésirable vécu dans un premier établissement peut être traité rapidement en se basant sur l'expertise d'un autre établissement pour le même incident. Le premier établissement peut même éviter des conséquences de l'incident qui ont été perçus au sein du deuxième établissement. En effet, on peut trouver dans les notes correctives enregistrés dans le suivi de l'événement les actions à effectuer pour éviter des conséquences réelles.

2.4.3 N-grams pour l'extraction textuelle

Au niveau de la plateforme RMP, les utilisateurs, en remplissant les différents champs, doivent donner une description de l'événement et de l'action corrective mise en œuvre. Ainsi, les informations dans la base de données seront sous la forme textuelle en plus d'autres informations scalaires. Leur exploitation pour extraire des informations pertinentes n'est pas facile. Afin d'appliquer une classification numérique des événements précurseurs déclarés, nous avons cherché une représentation vectorielle pour chaque déclaration. Les parties les plus difficiles à exploiter sont les textes. Il s'agit donc d'une problématique d'extraction textuelle. En technologie d'information la première étape dans un processus de traitement d'un gros corpus avec des moyens statistiques est de subdiviser le texte en plusieurs unités lexicales (unités d'information) appelées *tokens*. Cette étape est appelée *Tokenisation*.

L'unité d'information élémentaire traditionnelle est le mot. Cependant, un mot utilisé comme unité lexicale ne donne pas une idée correcte sur le sens voulu par le texte. Aussi il n'est pas trivial au sens informatique de repérer un mot dans le texte. C'est le cas de la langue arabe dont les personnes sujets et compléments sont parfois attachés, ou la langue allemande dans le cas de termes composés [216]. Cette subdivision ne semble pas être très adéquate à des perspectives de classification numérique à des fins d'extraction de connaissances [217].

L'approche n-grams [216] est une autre méthode utilisée pour la représentation textuelle. Un n-grams est une séquence de n éléments (caractères ou mots). Pour un document donné, l'ensemble de n-grams peut être généré par impliquant des séquences de caractères (ou des mots) de longueur n sur le texte entier. L'idée, provenant des travaux de Claude Shannon [218] en théorie de l'information, consiste à déterminer les fréquences de toutes les séquences de n-grams de caractères ou de mots dans un texte pour un calcul de similarité. Elle permet aussi de déterminer les relations entre les mots [219]. Cette notion a été utilisée en traitement automatique du langage naturel et pour d'autres applications telles que le traitement de l'écrit [220] et l'indexation [221].

2.4.4 Construction du vecteur

Examinant la littérature, il semble qu'il existe trois approches pour construire un vecteur valable au traitement d'extraction de connaissances et qui contient les unités lexicales les plus pertinentes. La première approche consiste à construire un vecteur spécifique à partir des unités lexicales de chaque texte. Elle présente quelques inconvénients puisqu'elle renvoie un ensemble de vecteurs avec des dimensions variables qui les rendront difficiles à explorer et à comparer. La deuxième approche utilise un vecteur de référence, contenant un ensemble d'unités lexicales pertinentes. A partir de ce vecteur référence créé par un expert, chaque texte sera décrit avec la fréquence de la présence de ces unités lexicales. Cette approche est efficace, mais il faut impliquer un expert pour chaque domaine à étudier. Chose qui n'est pas toujours facile. La dernière approche considère tout le texte corpus et utilise un seul vecteur avec toutes les unités lexicales de tous les textes. Ce vecteur est décrit comme une matrice formée par "Unités lexicales X textes". Cette matrice ne contiendra pas les unités lexicales de fréquences élevées ni celles de fréquences basses. Sachant que la fréquence est définie par le nombre d'apparition d'une unité lexicale. L'inconvénient de cette méthode est que la matrice contiendra plusieurs cellules vides. Mais de nombreuses méthodes ont été proposées pour remédier à ce problème notamment les approches de sélection et d'extraction que j'ai présentée dans le premier chapitre.

Nous avons adopté la troisième méthode, pour construire notre vecteur représentatif. Pour conserver la signification des expressions dans le texte tout en limitant le nombre d'unités lexicales et les cellules vides, nous avons combiné les deux méthodes n-grams et n'-

grams dont la première est une n-grams de mots et la seconde est une n'-grams de caractères. La première nous permet de déterminer les séquences de mots. Puis, nous appliquons la décomposition des mots résultants avec la n'-grams de caractères. La valeur de n' doit être supérieure à 3 pour que les mots aient un sens. La dernière étape de ce processus de construction du vecteur consiste à réduire la dimension de la matrice en appliquant un nettoyage des cellules vides et une sélection de variables.

2.4.5 Classification des EI

Les étapes précédentes nous ont permis de préparer le vecteur représentatif d'une déclaration d'un évènement indésirable. À ce niveau, nous procédons à la classification des déclarations pour pouvoir proposer des actions correctives adéquates. Il s'agit d'une classification supervisée. Nous avons essayé deux approches à savoir la méthode SVM et la méthode k-NN (k-Nearest Neighbour : k-plus proche voisin). Bien que la classification soit la solution la plus adéquate pour un tel système de retour d'expertise, la distribution des observations (les évènements indésirables) est compliquée. Outre cette complexité de distribution de données, le vecteur observation est de dimension relativement grande. Pour cette raison nous avons choisi la méthode d'apprentissage statistique. Les classes sont préparées par un expert du groupe Biomedica selon leurs criticités et surtout les actions correctives attribuées et réussies.

Dans un premier temps, nous avons appliqué une classification supervisée et nous prévenons une extension de ce processus vers une classification semi-supervisée. En effet, plusieurs évènements précurseurs ou indésirables se répètent et ils sont connus par les professionnels ainsi que leurs actions correctives. Néanmoins, on peut se trouver devant certaines situations accidentelles jusque-là inconnues et qui non jamais été traitées auparavant. En tant que plateforme de retour d'expertise, la RMP doit ajouter une nouvelle classe pour chaque nouvelle situation.

2.4.6 Mesure de similarité

Il existe plusieurs méthodes dans la littérature pour le calcul de la similitude entre deux textes. Je peux citer la similarité cosinus (ou mesure cosinus) [222], l'indice et la distance de Jaccard [223] et l'information mutuelle que nous avons présenté dans le premier chapitre.

Nous avons adopté la distance χ^2 , prononcé "khi-deux" ou "khi carré" donnée par l'équation 3.1 pour mesurer la similarité entre deux documents (textes) [224] [225]. D_i et $D_{i'}$ sont les matrices représentatives des deux documents (deux déclarations de deux EI). On note par f_{ij} la fréquence de la j^{ieme} séquence de n' caractères du i^{ieme} mot.

$$d_{\chi^2}(D_i, D_{i'})^2 = \sum_j \frac{1}{f_j} \left(\frac{f_{ij}}{f_i} - \frac{f_{i'j}}{f_{i'}} \right)^2 \quad 3.1$$

$$\text{Où } f_i = \sum_j f_{ij} \text{ et } f_j = \sum_i f_{ij}$$

2.4.7 Validation expérimentale

Pour valider les approches proposées, nous avons utilisé des données réelles fournies par des déclarations de différents établissements médicaux où la plateforme RMP est installée et utilisée. Nous avons utilisé un corpus de 257 déclarations. Dans une première étape, ces déclarations ont été classées manuellement en 17 catégories par un expert du groupe Biomedica. Puisque nous n'avons que 257 déclarations, nous avons effectué une validation croisée avec un ensemble de 3 tests. Pour chaque test, nous avons utilisé 2/3 de données

comme base d'apprentissage et le 1/3 qui reste pour le test. Ci-dessous un exemple de déclaration :

- Titre : "Chute d'un patient"
- Description : "Après une séance de radiothérapie, le patient X se leva de la table et perdit sa stabilité qui provoque une chute du patient. Le patient a passé quelques secondes étourdi avant de se lever sans dommages "
- Indicateurs numériques avec une échelle de 1 à 4 qui définissent :
 - o Fréquence : 3
 - o Gravité : 1
 - o Non détectable : 1
- Criticité : égale au produit de la fréquence multipliée par la gravité et la non-détectabilité. Dans cet exemple la criticité est égale à 3.

Le vecteur représentatif de cette déclaration est alors :

Attribut	valeur
Chute	2
patient	3
Radiothérapie	1
Stabilité	1
étourdi	1
Gravité	1
fréquence	1
Non détectabilité	1

Les premières expériences concernent la classification SVM avec différentes valeurs de n et n' pour n -grams de caractères et n' -grams de mots. Le Tableau 3-16 montre les performances de notre approche de gestion des épidémies.

TABLEAU 3-16: RESULTATS DE CLASSIFICATION DES EVENEMENTS INDESIRABLES

Méthode	Précision obtenue	Résultat de la réduction
N-Gram words + N-Gram char / SVM	58%	
N-Gram words + N-Gram char / kNN	59%	
N-Gram words + N-Gram char / Réduction 1 / SVM	69%	80 paramètres
N-Gram words + N-Gram char / Réduction 2 / SVM	72%	65 paramètres
N-Gram words + N-Gram char / Réduction 1 / kNN	68%	40 paramètres
N-Gram words+ N-Gram char / Réduction 2 / kNN	73%	50 paramètres

3. Gestion automatique des épidémies

Ces dernières années ont été marquées par l'apparition de plusieurs épidémies qui ont frappés différentes régions du monde. Depuis 2002, le virus SRAS a provoqué une épidémie, en Chine et a fait 800 morts dans le monde. Depuis, il y a eu d'autres virus qui ont fait aussi des bilans lourds de plusieurs centaines de morts. Ces derniers jours, on ne parle que du virus Ebola en Afrique et il s'agit encore d'une épidémie inquiétante. L'Arabie Saoudite n'a pas fait l'exception. En effet, le coronavirus MERS a fait un bilan de mort inattendu en 2014 malgré les moyens financiers et organisationnels mis en place. Il a même touché des professionnels du corps médical.

Pour plusieurs raisons la gestion de ce type d'épidémie est très difficile voir compliquée. Trois facteurs sont très importants pour bien gérer les risques liés à ces virus.

Le temps est un premier facteur très important. Il faut être rapide dans ces situations pour soigner les premiers cas, limiter les risques de propagation de la maladie et surtout comprendre le virus, son évolution et ses conséquences.

Le deuxième facteur aussi important que le premier est celui de la récolte d'information. Pour ces situations, toute information peut être utile pour la détermination des causes et conséquences et encore une fois limiter les risques. Les chercheurs ont besoin de savoir tous les détails et puisque les virus de ce type sont souvent nouveaux, il est aussi important que ces chercheurs échangent ces informations entre eux.

Dans le même contexte d'échange d'information, le troisième point que nous considérons important à traiter est l'échange d'expertise ou le retour d'expérience des chercheurs et praticien. Les virus qui provoquent des épidémies sont souvent peu connus mais parfois présentent des symptômes et des signes qui se ressemblent avec d'autres parus auparavant et dans ce cas le retour d'expérience peut aider aussi au diagnostic. Il est aussi important de pouvoir anticiper et prévoir la possibilité d'une épidémie dès les premiers cas.

J'ai présenté quelques exemples de travaux de recherche dans ce contexte de gestion du risque dans la section précédente. Plusieurs domaines ont été concernés par cette problématique telle que l'aéronautique, le transport ferroviaire, la radiothérapie et la médecine nucléaire. Et j'ai noté l'importance du retour d'expérience pour tous ces domaines. Pour les risques liés aux épidémies, il y a eu aussi des travaux dont les plus pertinents ont été proposés par des mathématiciens statisticiens [226] [227] [228]. Ils ont travaillé surtout sur la modélisation des épidémies pour prédire l'impact d'une maladie dans une population et la capacité de la surmonter. Il y a eu d'autres propositions spécifiques de détection automatique de certaines anomalies médicales. Dans [229] par exemple, les chercheurs proposent un système de raisonnement basé sur des cas pour aider à la détection et la classification des infections nosocomiales

3.1 Nouvelle approche de gestion automatique des épidémies

3.1.1 Un système informatique de partage d'expertise

Pour le constat présenté précédemment, j'ai proposé le développement d'une plateforme de gestion du risque médical basée sur le partage d'une base de données entre plusieurs centres hospitaliers. Ainsi ma contribution consiste à développer une approche capable d'exploiter l'expertise qui sera dans cette base en se basant sur des nouvelles méthodes de classification et d'extraction de connaissances. L'objectif principal de ce processus est l'anticipation de la détection des épidémies. Le présent travail a été financé par l'université d'Al Qassim en Arabie Saoudite. Cette dernière (Arabie Saoudite) a été très touchée par le coronavirus MERS. La Figure 3-30 présente un exemple de partage de base de connaissance entre centre hospitalier dans ce royaume.

La solution envisagée sera donc composée de trois parties dont la première est le bloc de récolte de données. La deuxième partie est la plus importante et elle a fait l'objet de ce travail de recherche. Elle consiste à la classification automatique de ces données récoltées pour pouvoir détecter la possibilité d'une épidémie et effectuer une évaluation. La phase de classification sera précédée par une phase d'extraction de caractéristiques. Pour ce problème, nous utilisons la technique ACF (Analyse Formelle de Concepts) qui sera présentée dans la section 3.1.3. L'objectif de cette méthode est de détecter les attributs pertinents pour la classification.

La troisième partie concerne le retour d'expérience. En effet, après la récupération de données, il faut valoriser les éléments importants pour assister et accompagner les opérateurs dans un centre médical. Dans un premier temps, le système sera capable de proposer des statistiques et de calculer la criticité de la situation pour pouvoir prioriser le traitement et le raffinement de certains données. Les actions de prévention et de traitement prises par des opérateurs compétents et expérimentés seront par la suite enregistrées et partagées entre les centres adhérents à la base commune. L'idée consiste à appliquer des nouvelles approches d'apprentissage et classification pour exploiter une base d'expertise.



FIGURE 3-30: UNE COLLABORATION MEDICALE ELECTRONIQUE POUR LA GESTION DES EPIDEMIES.

Le MERS par exemple est un coronavirus de la même famille que le virus responsable du syndrome respiratoire aigu sévère (SRAS). Il provoque une infection des poumons, accompagnée de fièvre, de toux et de difficultés respiratoires. A la différence du SRAS, il génère une défaillance rénale. Dès les premiers cas déclarés de ce virus, une alerte sera déclenchée et une classe très proche de la classe du SRAS sera créé automatiquement et confirmée s'il y a eu plusieurs déclarations de cas venant de différents centres hospitaliers. Ainsi, le partage de données entre ces centres permettra la détection de toute situation sanitaire critique. Nous pourrons anticiper cette détection et gagner en temps.

3.1.2 Classification semi supervisée pour la détection des épidémies

Ma contribution dans ce contexte de gestion des épidémies est focalisée essentiellement sur la deuxième partie (ou bloc) du système informatique envisagé. Ce bloc fonctionnel est conçu pour classer les données des patients reçus de différentes sources (centre hospitaliers). Chaque cas (patient) est décrit par un vecteur numérique. Un attribut de ce vecteur porte l'information sur l'existence ou non d'un symptôme (fièvre, diarrhée, douleur abdominal...). Par exemple, un patient qui a de la fièvre et qui tousse mais il n'a pas ni douleur abdominal ni diarrhée, son vecteur représentatif sera (1, 1, 0, 0) si on suppose que les deux premiers attributs sont respectivement pour la fièvre et la toux et les deux derniers sont respectivement pour la douleur abdominal et la diarrhée. Ce vecteur patient est l'entrée de notre bloc de classification. Il s'agit alors de l'observation à classer dont les variables (attributs) à prendre en considération

constat pour notre approche de classification, par exemple, un vecteur représentatif d'un cas Ebola peut être très proche (à la différence de deux ou trois attributs) d'un autre vecteur d'un cas SARS. Les classes sont alors très proches, chevauchées et non linéairement séparables.

3.1.3 Nouvelle technique de réduction de dimension

Plusieurs travaux de recherche ont été proposés pour l'extraction de caractéristiques et la réduction de dimensions pour des problèmes de classification. Ces approches ont fait l'objet du premier chapitre. Pour la problématique de gestion des épidémies, j'ai appliqué une approche différente jusque-là non utilisée dans le contexte de réduction de dimension. Il s'agit de la méthode d'Analyse Formelle de Concepts (AFC). L'idée consiste à trouver des regroupements d'attributs communs entre différentes classes.

3.1.3.1 Analyse Formelle de Concepts (AFC)

L'Analyse Formelle de Concepts appelée aussi Analyse de Concepts Formels est un formalisme mathématique pour la théorisation des concepts et de la pensée conceptuelle. Ce formalisme relie la théorie des ensembles ordonnés, et les applications d'analyse de données. Il sert à déterminer des concepts structurés hiérarchiquement regroupant des objets possédant les mêmes attributs. Le résultat de l'AFC est une hiérarchie ou réseau de concepts appelé treillis de Galois

AFC est appliquée dans de nombreuses disciplines telles que la médecine, les réseaux de capteurs [230] et de l'ingénierie industrielle [231] et d'autres domaines d'analyse de données [232]. Cette technique utilise des contextes formels et des réseaux de concepts. Elle permet de rechercher automatiquement des ensembles d'objets et un ensemble de leurs attributs en commun.

Définitions :

- Contexte : Soient A et O respectivement un ensemble d'attributs et un ensemble d'objets. Un contexte C, décrit par A et O &, est une donnée matricielle dont les lignes représentent les objets et les colonnes représentent les attributs. Les données dans la matrice peuvent être représentées par des valeurs booléennes ou réelles.
- Concept formel : Soit $a \subset A$ et $o \subset O$. (a, o) un concept formel de C, si les attributs de l'ensemble admettent en commun les objets de l'ensemble o.

3.1.3.2 AFC pour la réduction de dimension

Pour le problème de gestion des risques liés aux épidémies, le contexte C est l'ensemble des données d'entrée (base de données d'apprentissage et de test). L'ensemble des objets (O) sont les vecteurs de la classe décrits par plusieurs attributs (A). Chaque attribut est représenté par le taux d'existence d'un symptôme (fièvre, diarrhée, douleur abdominale ou autres symptômes) dans les vecteurs de la même classe dans la base de données d'apprentissage. Le tableau suivant (Tableau 3-17) représente un contexte avec 5 classes (voir section 2 : classes a, b, c, d, e) et 5 attributs. Ensuite, ce contexte de valeurs est converti en un contexte binaire (Tableau 3-18) avec la transformation suivante :

$$C(i, j) = 0 \text{ if } cell(i, j) < 0.5$$

$$C(i, j) = 1 \text{ if } cell(i, j) \geq 0.5$$

Avec $cell(i, j)$ est la valeur de l'attribut (pourcentage d'existence du symptôme) i dans la classe (l'épidémie) j.

TABLEAU 3-17: CONTEXTE DE VALEURS.

	a	b	c	d	e
1	0.35	0.14	1	0.5	0.28
2	0.78	0.42	1	0.57	0.28
3	0.78	0.21	0	1	1
4	1	0.5	1	1	0.85

TABLEAU 3-18: CONTEXTE BINAIRE.

	a	b	c	d	e
1	0	0	1	1	0
2	1	0	1	1	0
3	1	0	0	1	1
4	1	1	1	1	1

Le réseau conceptuel décrivant le problème de la gestion des épidémies est généré à l'aide de l'outil "Lattice Miner" [233] (voir Figure 3-32). Lattice Miner est un outil de data mining pour créer, visualiser et explorer des réseaux de concepts (treillis de Galois). Il permet la génération de concepts formels et de règles d'association. Ce réseau de concepts est le résultat de l'application de la technique FCA pour extraire tous les concepts formels d'un contexte et trouver des groupes d'objets (classes ou épidémies) qui partagent des groupes d'attributs (symptômes). Pour notre problème de classification des vecteurs patient, l'approche AFC a assuré deux objectifs majeurs à savoir la réduction de dimension et la sélection des classes de chaque niveau de l'arbre binaire de la classification multi-classe MSVM.

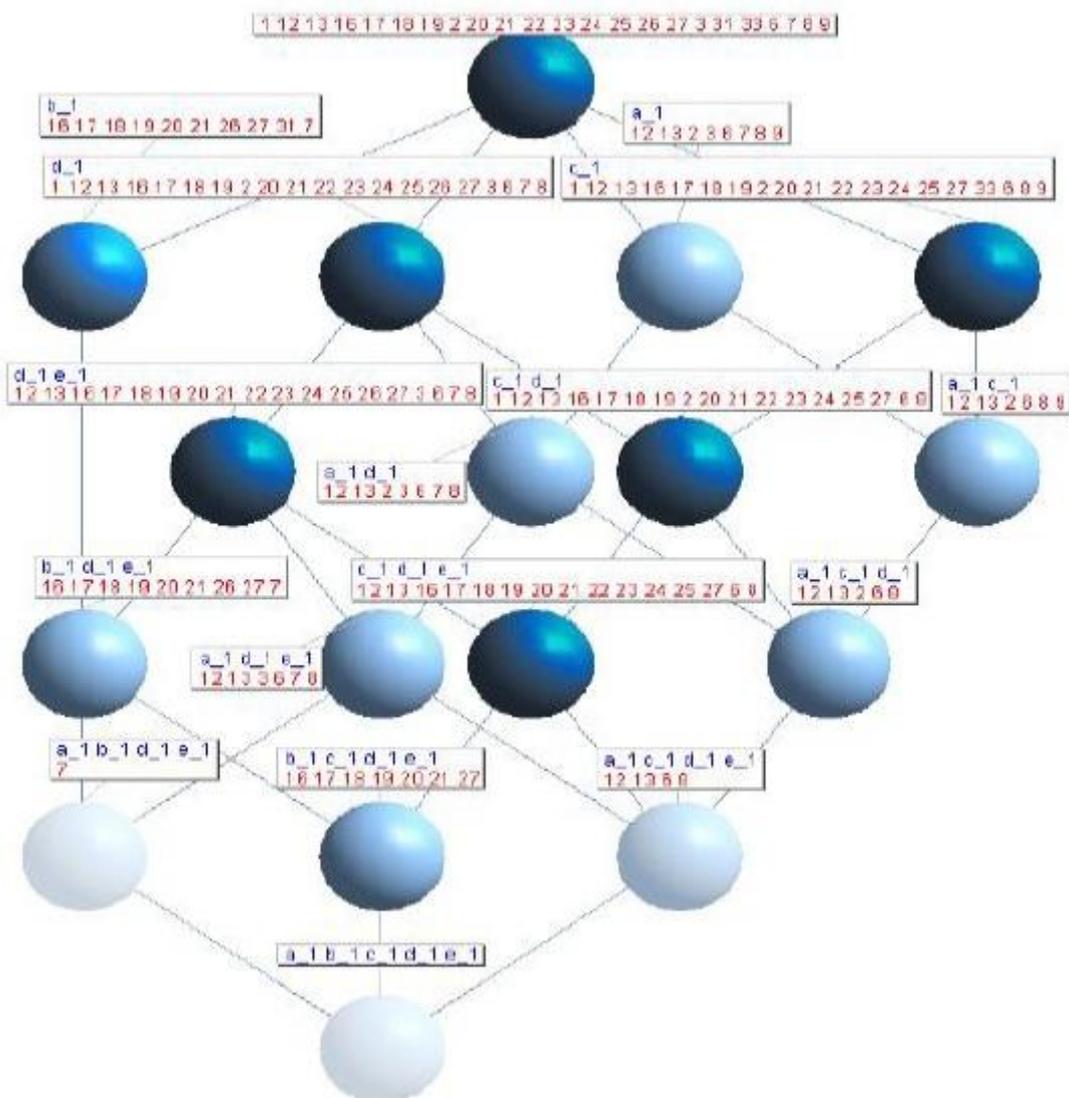


FIGURE 3-32: LE RESEAU CONCEPTUEL DE 5 CLASSES AVEC 33 ATTRIBUTS.

Le premier objectif assuré par la méthode AFC est le but principal voulu par l'usage de ce formalisme. Il s'agit de déterminer les attributs communs des différentes classes. Ces attributs ne seront pas essentiels pour la classification. Ainsi, la méthode AFC a permis de sélectionner les informations discriminantes. Le nombre de ces attributs diminue d'un niveau à l'autre dans l'arbre MSVM (voir section suivante).

Le deuxième objectif assuré par la méthode AFC, aussi important que le premier, est la détermination des combinaisons des classes de chaque niveau de l'arbre MSVM. Les classes qui ont plus d'attributs communs ont été regroupées ensemble pour pouvoir diminuer la dimension au maximum.

3.1.4 Classification multi-classe avec dimension dynamique

La méthode SVM est une approche de classification binaire. Mais il est possible d'appliquer cette approche pour une classification multi-classe en utilisant l'une des techniques dite MSVM (Multi-classe Support Vector Machines). On peut citer les techniques "un contre tous", "un contre un" et l'arbre binaire [234]. Ces techniques consistent à un séquençement de classification binaire avec une combinaison de groupement de classes différente pour chaque technique. Pour le problème de gestion des épidémies, nous avons choisi la technique de

l'arbre binaire (voir Figure 3-33). Elle est basée sur la division de l'ensemble des classes en deux groupes disjoints dans chaque nœud de l'arbre de décision. A chaque niveau de l'arbre on applique une classification SVM pur décider dans lequel des groupes l'échantillon inconnu doit être attribué.

Nous avons appliqué l'approche AFC, présentée dans la section précédente, à chaque niveau de l'arbre. Ce qui nous a permis de déterminer les combinaisons adéquates des groupements de classes à chaque niveau et la dimension optimale du vecteur observation aussi variable à chaque niveau. Ainsi, on parle d'une dimension dynamique optimale. Cette dimension peut aussi changer à chaque mise à jour de la base d'apprentissage en fonction des concepts formels possibles. C'est un aspect aussi innovant sur lequel je vais me baser pour des futurs projets qui seront détaillés dans mes perspectives de recherche.

3.2 Validation expérimentale

Dans cette section, nous présentons les résultats de la validation de notre solution proposée. Pour valider la détection des épidémies, l'approche a été appliquée sur différents cas et les tests ont été réalisés sur les deux phases du processus à savoir la phase temps anticipée et une simulation de la phase temps réel.

3.2.1 Cas de 5 épidémies

Pour valider le processus de classification proposé, nous avons choisis les cinq épidémies les plus connus dans la région et surtout les plus similaires. Cette similarité fait de notre problème de classification de plus en plus difficile. Les cinq classes sont alors :

- a- MERS (Coronavirus)
- b- SARS (Coronavirus)
- c- EBOLA
- d- Influenza A/H1N1
- e- Influenza A/H5N1 (Avian Influenza)

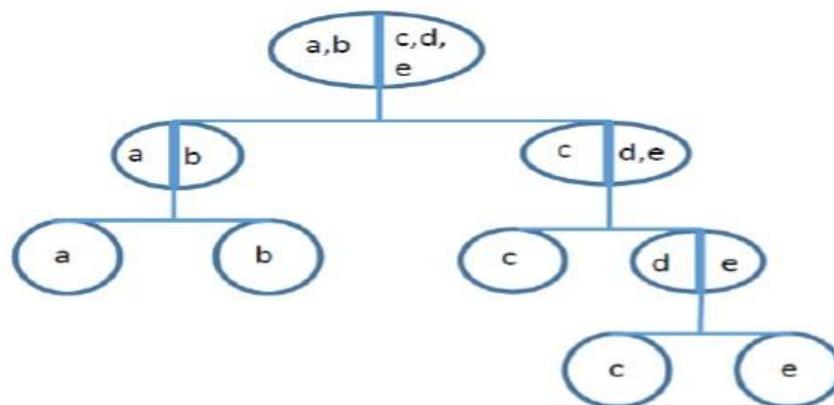


FIGURE 3-33: ARBRE BINAIRE DE CLASSIFICATION POUR 5 EPIDEMIES.

Nous avons appliqué notre approche MSVM sur des vecteurs patients, composé de 33 attributs. Ces observations ont été recueillies à partir de différentes sources fiables [235] [236] [237]. En appliquant la technique AFC, (Figure 3-32) on a trouvé des groupes (classes intermédiaires) à différents niveaux de l'arbre binaire de la Figure 3-33. Ces combinaisons de classes ont permis de réduire la dimension des données et avoir une dimension dynamique par rapport aux différents niveaux de l'arbre.

Pour appliquer une validation croisée, nous avons divisé la base en 3 parties et la validation a passé par 3 phases. À chacune, nous considérons une partie de données pour le test et les deux autres parties pour l'apprentissage.

3.2.2 Résultats

3.2.2.1 Cas d'épidémie connue : Classification supervisée

Le tableau suivant (Tableau 3-19) résume les résultats de la classification SVM pour différentes valeurs du paramètre de fonction noyau σ puisque nous utilisons un noyau gaussien (voir Tableau 2-12 chapitre 2). Le taux d'apprentissage est intéressant pour le premier niveau de classification. Ce qui n'est pas le cas pour les autres niveaux. En effet, la similarité est faible entre la première classe (a, b) et la seconde (c, d, e). Par exemple, dans le premier niveau, le taux d'erreur est égal à zéro pour les différentes valeurs de σ . Bien qu'il ne soit pas linéaire, la séparation entre les classes est facile en raison de la distance suffisante entre elles. Pour un système d'aide à la décision, il est intéressant d'avoir un taux d'apprentissage élevé dans le premier niveau. Nous pouvons garantir une attribution d'un vecteur au premier groupe de classes (2 ou 3 épidémies) sans erreur. Le tableau (Tableau 3-19) montre les mêmes résultats (taux d'apprentissage = 100%) pour $\sigma \geq 2$ dans le niveau 2 de l'arbre binaire. Par conséquent, nous pouvons trouver la bonne classe pour les patients de classe a, classe b ou classe c. Pour le niveau 3 (résultat final de la classification), les résultats sont compris entre 85% et 91,4% comme taux d'apprentissage (taux d'erreur compris entre 15% et 8,6%). Les meilleurs résultats (94,2%) sont pour $\sigma = 2,5$.

Pour $\sigma < 2$, on a eu de mauvais résultats avec un taux d'erreur compris entre 16% et 38%. Il est clair que les hyperplans séparateur (et les classes) sont suffisamment proches pour confirmer une classification non linéaire. Cette observation est également confirmée par les résultats de la classification en ligne qui sera présentée dans le paragraphe suivant.

TABLEAU 3-19: RESULTATS DE LA CLASSIFICATION MSVM.

	$\sigma=10$		$\sigma=5$		
	Errors	Learning rate	Errors	Learning rate	
Level-1	0	100%	0	100%	
Level-2	0	100%	0	100%	
Level-3	11	84.2%	10	85.7%	
Final	11	84.2%	10	85.7%	
		$\sigma=2.5$		$\sigma=2$	
		Errors	Learning rate	Errors	Learning rate
Level-1	0	100%	0	100%	
Level-2	0	100%	0	100%	
Level-3	6	91.4%	7	90%	
Final	6	91.4%	7	90%	
		$\sigma=1.6$		$\sigma=1$	
		Errors	Learning rate	Errors	Learning rate
Level-1	0	100%	8	88.5%	
Level-2	4(ab)	94.28%	5(cde)+ 9(ab)	80%	
Level-3	7	90%	12	82.85%	
Final	11	84.2%	26	62.8%	

3.2.2.2 Cas d'épidémie inconnue : Classification semi-supervisée

Pour la classification semi-supervisée, un ensemble de tests a été appliqué pour vérifier l'efficacité de l'approche. En se basant sur la distance euclidienne, Nous effectuons, en temps réel, une première vérification d'appartenance des nouvelles observations aux classes existants. Les résultats de sept nouveaux vecteurs patients sont présentés dans le Tableau 3-20. Il contient les distances entre ces vecteurs et différents centres des cinq classes présentées précédemment. Le Tableau 3-20 contient également la plus grande distance entre les centres et les autres vecteurs de la même classe. Tous les nouveaux vecteurs sont de la classe "choléra". Il s'agit d'une nouvelle épidémie (non listée dans les cinq classes existantes) et nous pouvons observer que les distances entre ces vecteurs et les centres sont élevées (distance > D_{\max}). Ainsi, il y a une faible similitude avec d'autres classes et le résultat est égal à 100% comme un taux d'apprentissage (erreur 0). D_{\max} est la distance entre le centre de classe et le point le plus éloigné de la même classe.

Les résultats de classification en ligne sont également vérifiés par les vecteurs utilisés dans le premier processus en considérant dans chaque cas une des classes comme une nouvelle épidémie non classée précédemment. Le Tableau 3-21 indique la distance minimale entre tous les points d'une classe et le centre des autres classes. Il y a un taux d'erreur égal à 25,8% (taux d'apprentissage égal à 74,2%). Ces résultats expliquent notre premier choix de classification supervisée. En effet, nous avons un taux de réussite de 94,2% contre 74,2% dans la deuxième phase.

Les résultats de classification en ligne sont également vérifiés par les vecteurs utilisés dans le premier processus en considérant dans chaque cas une des classes comme une nouvelle épidémie non classée précédemment. Le Tableau 3-21 indique la distance minimale entre tous les points d'une classe et le centre des autres classes. Il y a un taux d'erreur égal à 25,8% (taux d'apprentissage égal à 74,2%). Ces résultats expliquent notre premier choix de classification supervisée. En effet, nous avons un taux de réussite de 94,2% contre 74,2% dans la deuxième phase.

TABLEAU 3-20: LA DISTANCE EUCLIDIENNE ENTRE LES NOUVEAUX VECTEURS ET LES DIFFERENTES CLASSES.

	a	b	c	d	e
D_{\max}	3.9987	3.8881	5.1	5.1676	5.1130
1	4.1564	3.9064	5.3861	5.2770	5.1270
2	3.9084	3.8139	5.1971	5.0840	4.9281
3	3.9084	3.6415	5.1971	5.0840	4.9281
4	3.9084	3.6415	5.1971	5.0840	4.9281
5	3.9987	3.6415	5.1	5.1676	5.1130
6	3.9448	3.9064	5.1	5.0840	4.9713
7	3.9987	3.8881	5.1	5.1676	5.1130

TABLEAU 3-21: DISTANCE MINIMALE ENTRE LES CLASSES.

	a	b	c	d	e
a		3.5015	3.3396	3.4419	3.5857
b	3.6239		3.4449	2.9983	2.8785
c	3.3151	3.4189		1.9614	2.1712
d	3.4832	3.0431	1.9666		0.9258
e	3.3151	3.0195	2.2062	0.9949	

4. Conclusion

Ma contribution dans ce chapitre consiste au développement de nouvelles approches de gestion du risque médical. Ce problème est très lié à la prise de décision. Dans ce contexte, l'usage d'un outil informatique devient indispensable. Dans la plupart des domaines, notamment les plus critiques, les risques sont souvent gérés par le retour d'expérience. Le constat des problèmes dans le domaine médical en France et ailleurs montre qu'il est aussi critique que d'autres domaines et qu'il est confronté à plusieurs risques.

Dans ce chapitre, j'ai présenté deux solutions pour la gestion du risque médical. En me basant sur des techniques d'apprentissage et classification, les approches proposées assurent un retour d'expérience automatique très utile. En effet, les situations de risque ou accidentelles souvent se répètent avec des conditions similaires. L'expert, pour sa part, connaît les résultats des décisions prises auparavant et il est capable de mesurer les conséquences. Néanmoins, il est difficile de représenter et de stocker les connaissances de l'expert dans un système informatique avec les moyens de traitement de l'information traditionnels. Ainsi, j'ai eu recours aux nouvelles techniques d'extraction de connaissances et de fouille de données pour palier à ce problème.

Le chapitre est organisé en deux parties dont la première est consacrée à la gestion des événements précurseurs ou indésirables et la deuxième présente ma contribution pour la gestion des risques liés aux épidémies. Pour les deux problèmes, j'ai proposé des solutions basées, entre autres, sur l'apprentissage statistique à travers la technique SVM. L'approche de gestion des événements indésirables a été proposée dans un cadre industriel et validée via une plateforme que j'ai développée en collaboration avec la société Biomedica. Il s'agit de la plateforme web RMP actuellement sur le marché et proposée en mode SAAS.

Les déclarations des événements indésirables dans la plateforme RMP contiennent des parties de description textuelles. Ainsi, avec une équipe que j'ai dirigée au sein de la société, nous avons proposé une amélioration de l'approche n-grams pour l'extraction textuelle. Nous avons appliqué n-grams de caractères avec n-grams de mots pour pouvoir déterminer un vecteur numérique avec des attributs pertinents capables de représenter la déclaration de la façon la plus fidèle. Ce vecteur est alors une observation pour l'approche de classification. Ces contributions ont fait l'objet de la première partie.

Dans la deuxième partie de ce chapitre, j'ai présenté un travail dans le même contexte de gestion du risque médical qui concerne les épidémies. Mes premières contributions étaient dans le cadre d'un projet interne financé par l'université d'ALQassim en Arabie saoudite. Nous avons proposé une solution d'aide à la décision dans le cadre d'un partage d'expertise entre des centres de soins. Ma contribution théorique consiste à développer un processus de classification semi-supervisée avec une dimension dynamique des observations qui optimise le rapport dimension/efficacité d'apprentissage. Pour la réduction de dimension, j'ai appliqué la technique Analyse Formelle de Concepts (AFC).

Je peux résumer les perspectives de mes travaux dans cet axe en deux points importants. Le premier concerne l'extraction textuelle. En effet l'existence d'un lexique médical "MeSH" va beaucoup nous aider à effectuer une analyse automatique des textes médicaux tout en optimisant les algorithmes dédiés. Le deuxième volet concerne la partie apprentissage et classification. Nous devons travailler sur l'adéquation entre les actions correctives et les déclarations, pour la gestion des EI. Et pour la gestion des risques liés aux épidémies, nous devons définir un ou plusieurs modèles de prédiction pour anticiper les actions. Dans les deux cas, les classes ne seront plus statiques. Ainsi, nous allons explorer les moyens de classification dynamique.

Conclusion et perspectives

Bilan

Dans ce document, j'ai retracé mes dix années de carrières. Pendant cette période de 2007 à 2017, je suis passé d'une expérience d'enseignant-chercheur de l'université Lille 1 en France à l'université d'AlQassim en Arabie Saoudite. Entre les deux expériences, j'ai eu l'occasion d'occuper des postes dans des entreprises de la région lilloise. À travers ces expériences dans le domaine industriel, j'ai pu développer et valider des approches théoriques et j'ai eu l'occasion de mener des projets de recherche et innovation. Durant ces années d'expérience industrielle ainsi que mes expériences en tant que enseignant-chercheur, j'ai continué à travailler sur différents axes de recherche souvent dans le cadre de collaborations universitaires ou industrielles avec d'autres chercheurs et/ou professionnels.

J'ai participé à l'encadrement de cinq étudiants en thèse de doctorat sur des sujets que j'ai proposés entre 2008 et 2016. J'ai aussi encadré des étudiants en Master de recherche et d'autres en projets de fin d'étude et Master professionnel.

En terme de publications scientifiques, j'ai un bilan de :

- 14 Articles dans des revues internationales avec Comité de Lecture
- 3 Chapitres d'ouvrage collectif
- 19 Articles dans des congrès Internationaux avec Actes et Comité de Lecture

En ce qui concerne mon bilan pédagogique, J'ai eu l'occasion d'enseigner dans différents établissements dans trois pays différents avec un cumul de plus que 700 heures de cours et plus que 400 heures de TD/TP.

Pendant quatre ans d'expériences industrielles (professionnelles) entre responsable R&D, consultant et responsable recherche innovation, j'ai participé au développement de cinq logiciels, outre le montage des dossiers de recherche, la réalisation des veilles technologiques et la formation.

Synthèse des travaux présentés dans ce mémoire

Mes contributions scientifiques, présentées dans ce présent document, s'inscrivent dans le contexte de la classification automatique et l'analyse de données. Elles portent, essentiellement, sur trois axes de recherche.

Le premier axe de recherche, présenté dans le premier chapitre, concerne la réduction de dimension de données pour le problème de classification. Une première partie de cet axe est consacrée à la sélection de variables alors que la deuxième partie présente une nouvelle approche d'extraction de caractéristiques. Pour la première problématique, ma contribution a porté sur le développement d'une méthode de type "Wrapper". Basée sur l'approche d'optimisation Tabou, la solution proposée consiste à explorer l'espace de combinaisons d'attributs, pour sélectionner les plus pertinents de point de vue du taux d'apprentissage. Bien qu'elle ait le même principe que les autres approches existantes, cette méthode a donné des résultats pertinents. Mais, elle reste limitée à des espaces de dimensions relativement moyennes voire faibles et elle ne permet pas une bonne représentation des relations intra et inter classes entre les attributs. Pour ces raisons, j'ai travaillé sur une nouvelle approche d'extraction basée sur un algorithme de classification des attributs que j'ai présenté dans la deuxième partie du premier chapitre. Cette méthode remplace les attributs de la même classe par son centre. Ainsi, nous avons proposé la mesure "Information Mutuelle" pour la détermination du centre de chaque classe. Nous avons appliqué une analyse des tendances pour déterminer le vecteur attribut. Appliquée à des problématiques différentes, cette méthode

d'extraction a pu réduire considérablement la dimension tout en gardant un taux d'erreur faible.

Le deuxième axe de recherche que j'ai pu développer st aussi dans le même contexte de classification automatique. Mais cette fois, je l'ai appliqué à des problèmes de systèmes automatiques. Il a fait l'objet du deuxième chapitre de ce mémoire et concerne deux problématiques dont la première est l'identification des systèmes automatiques et la deuxième le diagnostic de ses systèmes. En ce qui concerne l'identification, j'ai proposé deux approches d'apprentissage statistique à noyaux multiples, Multi-kernel LS-SVM et LS-SKM. Puis, j'ai présenté une troisième approche multi-noyaux de régression utilisant le même principe de division de vecteur que la deuxième. Cette approche a été appelée SKR. Appliquées sur des systèmes non-linéaires, ces approches ont donné des résultats attractifs comparables aux autres méthodes. Dans un autre contexte d'identification des systèmes commutant, j'ai présenté, dans le deuxième chapitre, deux autres approches dont une pour déterminer l'état discret et l'autre pour l'identification. La deuxième partie de ce chapitre, je l'ai consacré aux travaux sur le diagnostic des systèmes automatiques complexes. Pour ce problème, j'ai présenté deux algorithmes de classification. Le premier, nommé DCH (Dynamic Convex Hall), est basé sur le calcul dynamique de l'enveloppe convexe. La deuxième approche consiste en une classification dynamique basée sur la méthode SVDD.

Le troisième axe de recherche présenté dans ce mémoire concerne la gestion des risques médicaux. Dans ce contexte, j'ai travaillé sur le retour et le partage d'expérience automatique. En me basant sur une analyse de données à travers une nouvelle approche d'apprentissage et extraction, j'ai proposé deux approches pour deux problématiques différentes. J'ai utilisé l'approche n-grams pour l'extraction de données textuelles suivi d'un apprentissage automatique pour assister les professionnels de la santé à prendre les bonnes décisions dans des situations particulières. Pour assurer un retour d'expérience automatique dans le but de gérer les événements précurseurs ou indésirables dans les établissements de soins, j'ai participé au développement de la plateforme RMP, en tant que responsable R&D au sein de la société Biomedica. Ce premier projet de gestion du risque médical a été initié en 2009, par une collaboration entre la société et l'école centrale de Lille. J'ai dirigé l'équipe dans ce projet de développement et j'ai eu l'occasion d'appliquer une nouvelle approche d'analyse de données appliquant la méthode n-grams et l'apprentissage statistique. Les résultats décrits dans la première partie du chapitre 3 ont montré clairement l'efficacité de ces méthodes. La deuxième problématique traitée dans ce contexte est la gestion des risques liés aux épidémies. La méthode que j'ai proposé pour ce problème consiste en une détection automatique, au préalable, des épidémies en se basant sur l'approche AFC et une classification semi-supervisée. La deuxième partie de ce chapitre 3 décrit cette approche basée sur une hybridation entre un algorithme de classification supervisée SVM et un autre classique non supervisé "k-means". Ce travail a été élaboré dans le cadre d'un projet interne financé par l'université d'ALQassim et publié en début 2017 [238].

Projets en cours

Pour la continuité de mes travaux de recherche scientifique, je suis actuellement (2017) sur quatre projets qui portent sur des sujets différents.

1. Diagnostic des réseaux de capteurs : Le progrès actuel et le développement prévu dans les quelques années dans le domaine de l'Internet des Objets (IoT) a rendu ce volet très exploré par les chercheurs. Pour ma part, je travaille actuellement sur le diagnostic des réseaux de capteurs, dans le cadre d'un projet interne et dont je suis le responsable dans une équipe de 4 chercheurs financé par l'université d'ALQassim. Bien que le réseau de capteurs puisse être

considéré comme un système automatique, il a certaines particularités, notamment l'énergie faible et les aléas de communication. Le but de ce projet est le développement d'une approche de prédiction automatique de panne dans un réseau de capteurs. Une partie qui concerne l'optimisation dans les réseaux VANET est aussi envisagée dans ce projet.

2. Surveillance d'un système électrique multi-sources : Face aux enjeux énergétiques auxquels fait face le monde aujourd'hui et également au tarissement des énergies fossiles, le développement des énergies renouvelables semble de plus en plus une alternative nécessaire. Cependant, ces sources d'énergie telles que l'énergie photovoltaïque et éolienne dépendent des conditions météorologiques souvent très variables et imprévisibles. Pour garantir une sécurité d'approvisionnement en énergie, il est nécessaire de lever le problème d'irrégularité des sources par l'optimisation de leur combinaison entre elles et/ou avec des unités de stockage (Batterie, pile à combustible) et des sources conventionnelles. La contribution principale, dans ce travail de recherche est l'élaboration d'une stratégie d'optimisation, de commande et de supervision d'un système multi-sources pour une meilleure gestion des flux de puissance et une meilleure rentabilité du système global.

Pour ce projet, j'envisage de proposer un sujet de thèse.

3. Une approche d'optimisation distribuée : En collaboration avec un chercheur mathématicien, nous travaillons actuellement sur le développement d'une approche distribuée d'optimisation dont la division de l'espace de recherche fait appel à une méthode de Clustering. Ce travail est aussi dans le cadre d'un projet interne financé par l'université AlQassim.

4. Dans le cadre du projet KACST4 N°34-440, j'ai proposé un nouveau standard sur la qualité des applications web et mobile qui portent sur le Coran et l'éducation religieuse. J'ai rédigé une première version que j'ai proposée au SASO (Saudi Standards, Metrology and Quality Organization).

En plus de ces projets, je continue à travailler sur l'analyse textuelle automatique des rapports médicaux pour les deux problèmes présentés dans le dernier chapitre.

Perspective de recherche

Sur le plan théorique, mes perspectives de recherche portent sur deux axes :

1. Analyse de données textuelles multi-langue: Le "big data" est l'une des technologies émergentes très utile pour les nouveaux systèmes. Un des axes majeurs de ce domaine est l'extraction textuelle. Bien que plusieurs chercheurs travaillent sur ce problème il y a moins de solutions d'extraction textuelle multi-langue. La difficulté de ce problème réside dans la quantité énorme de données et la particularité de chaque langue. J'envisage travailler sur le développement de nouvelles approches d'analyse de données textuelles faisant appel à des techniques de l'intelligence artificielle dédié aux données multi-langue.

2. Classification et analyse de données pour les systèmes Multi-Agent : Dans un autre contexte, les nouvelles approches de **classification** et **d'analyse de données** peuvent faire l'objet de la partie intelligente dans un **système multi-agent** ou toute autre système d'information ou d'aide à la décision. Tel qu'il est présenté dans la Figure 3-34 suivante, je pourrai intervenir à deux niveaux :

- 1- Au niveau de la composante décision en mettant en valeur le retour d'expérience des agents. Un algorithme d'apprentissage pourra améliorer les résultats à ce niveau.

⁴ KACST : King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, www.k3cst.edu.sa

- 2- Comme nous l'avons signalé auparavant, les nouveaux systèmes échangent et traitent une quantité de données importante et de différents types (numériques, textuelles, images, son, vidéo,...). Ces données existent en entrée et en sortie des différents modules. Ils doivent être analysés et préparés pour qu'ils soient exploitables. C'est un deuxième axe qui rejoint les domaines d'envergure tel que "Big Data" et "Cloud Computing".

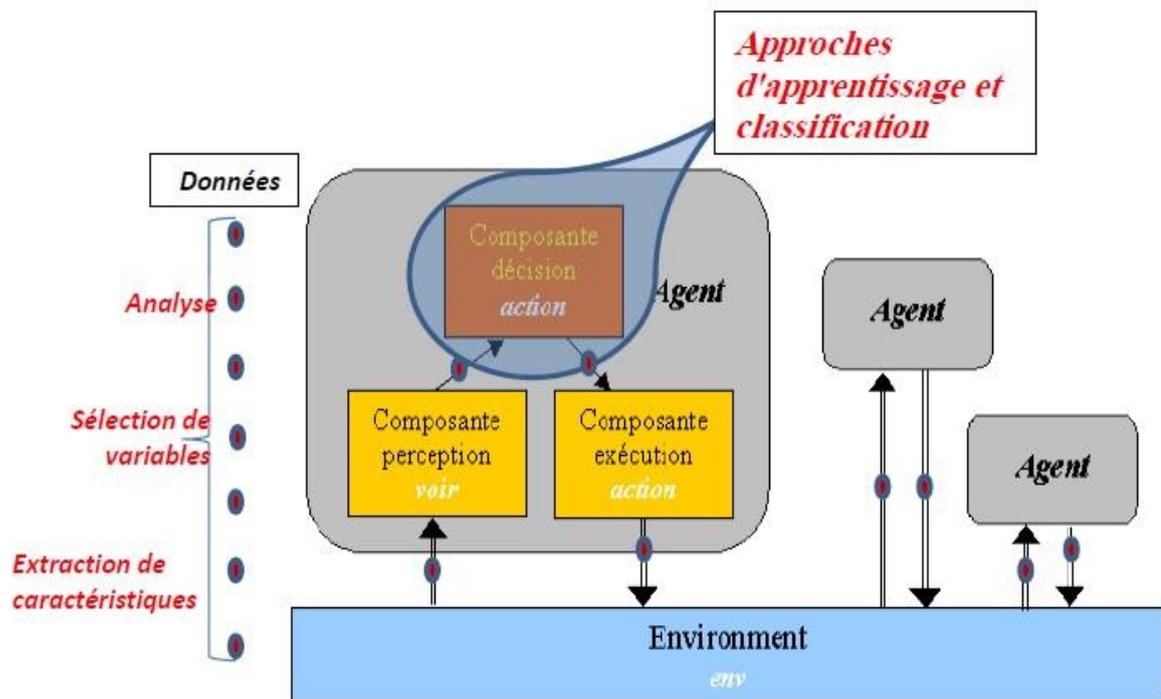


FIGURE 3-34: DE L'INTELLIGENCE A INTEGRER DANS LES NOUVEAUX SYSTEMES MULTI-AGENTS

3. Optimisation

L'un de mes projets en cours porte sur l'optimisation dans les **réseaux de capteurs** et les **VANET**. Il s'agit d'un projet financé par l'université d'AlQassim en Arabie Saoudite. Dans ce projet, je vais travailler sur l'étude et le développement de nouveaux modèles et approches d'optimisation dynamique afin de garantir un fonctionnement optimal de ces réseaux de capteurs. En plus d'une partie importante qui concerne la surveillance et le diagnostic de ces réseaux les modèles seront aussi appliqués dans un contexte de **l'internet des objets**. Une application prévue pour ce projet concerne le **domaine de la santé**.

Bibliographie

- [1] S. EL FERCHICHI, "Sélection et extraction d'attributs pour les problèmes de classification," PhD, Université Lille1, Lille, 2013.
- [2] M. TARHOUNI, "Contribution à l'identification des systèmes non linéaire par les Séparateurs à Vaste Marge (SVM)," PhD ENIT, Tunisie, 2014.
- [3] Lamaa SELLAMI, "Sur l'identification des systèmes dynamiques hybrides," PhD Université de Gabes, Tunisie, octobre 2017.
- [4] F. THELJANI, "Classification dynamique de données pour le diagnostic des systèmes.," PhD ENIT, Tunisie, 2013.
- [5] EL FERCHICHI, Sabra, LAABIDI, Kaouther, et ZIDI, Salah, "Genetic algorithm and tabu search for feature selection," *Studies in Informatics and control*, vol. 18, no. 2, pp. 181-187, 2009.
- [6] S. EL FERCHICHI, "Sélection de variables et extraction de caractéristiques pour une méthode d'apprentissage.," Tunis, 2007.
- [7] EL FERCHICHI, Sabra, ZIDI, Salah, LAABIDI, Kaouther, et al. , "Feature extraction for atmospheric pollution detection.," in *Communications, Computing and Control Applications (CCCA), 2011 International Conference on IEEE* , 2011.
- [8] M. TARHOUNI, "Surveillance des systèmes Par Séparateurs à Vaste Marge.," 2009.
- [9] Tarhouni, M., et al., "Surveillance des systèmes complexes par séparateurs à vaste marge (svm)," in *8e Conférence de MODélisation et SIMulation (MOSIM'10).*, 2010.
- [10] TARHOUNI, M., LAABIDI, K., ZIDI, S., et al., "Surveillance des systèmes complexes par séparateurs à vaste marge (svm).," in *8e Conférence de MODélisation et SIMulation (MOSIM'10).*, 2010.
- [11] M. L. K. Z. S. e. a. TARHOUNI, "A nonlinear MIMO system identification based on improved multi-kernel Least Squares Support Vector Machines (improved multi-kernel LS-SVM). In : Systems, Signals and Devices (SSD), 2011 8th Interna," in *2011 8th International multi-Conférence in Systems, Signals and Devices (SSD)* , 2011.
- [12] Tarhouni, M., Zidi, S., Laabidi, K., & Ksouri-Lahmari, M., "Least squares support kernel machines (LS-SKM) for identification.," *International Journal of Modelling, Identification and Control*, vol. 17, no. 1, pp. 68-77, 2012.
- [13] TARHOUNI, M., ZIDI, S., LAABIDI, K., & LAHMARI-KSOURI, M., "Least Squares Support Kernel Machines for identification of NARMA model (LS- SKM)," *NARMA. International Journal of*

Modeling and simulation for Engineering Research, vol. 1, no. 2, pp. 8-17, 2015.

- [14] Tarhouni, M., Zidi, S., Laabidi, K., & Ksouri-Lahmari, M., "Support kernels regression for NARMA system identification (SKRNARMA)," *International Journal of Modelling, Identification and Control*, vol. 22, no. 2, pp. 139-149., 2014.
- [15] TARHOUNI, M., ZIDI, S., LAABIDI, K., & KSOURI-LAHMARI, M., "System identification based on Support Kernels Regression.," in *Proceedings Engineering & Technology*, 2013.
- [16] F. THELJANI, "Classification non supervisée pour le diagnostic des systèmes industriels.," 2009.
- [17] THELJANI, Foued, LAABIDI, Kaouther, ZIDI, Salah et al., "Classification non Supervisée pour la Surveillance d'un Système Automatique.," in *8e Conférence Internationale de Modélisation et SIMulation*, Tunisie, 2010.
- [18] THELJANI, F., K. Laabidi, S. Zidi et M. Lahmari Ksouri., "Convex hull based clustering algorithm.," *International Journal of Artificial Intelligence IJAI*, vol. 10, no. S13, March March 2013.
- [19] THELJANI, Foued, Laabidi, K., Zidi, S., & Ksouri, M., "Multi-model approach for systems diagnosis based on SVDD classification.," *International Conference on Control, Engineering & Information Technology (CEIT'13) Proceedings Engineering & Technology*, vol. 3, pp. 100-105., 2013.
- [20] THELJANI, Foued, LAABIDI, Kaouther, ZIDI, Salah, et al., "Systems monitoring based on dynamic classification with SVDD.," in *Systems, Signals & Devices (SSD), 2013 10th International Multi-Conference on. IEEE*, 2013.
- [21] THELJANI, Foued, LAABIDI, Kaouther, ZIDI, Salah, et al., "Tennessee Eastman Process diagnosis based on dynamic classification with SVDD.," *Journal of Dynamic Systems, Measurement, and Control*, vol. 137, no. 9, p. 091006, 2015.
- [22] T. JULIEN, "Gestion des évènements indésirables.," 2009.
- [23] ZIDI, Salah, JULIEN, Thibaut, MJIRDA, Anis, et al., "Textual extraction and classification for medical risk management: A new Risk Management Platform to manage undesired medical events.," in *Advanced Logistics and Transport (ICALT), 2015 4th International Conference on IEEE*, 2015.
- [24] JULIEN, T., ZIDI, Salah, et al., "A decision support system for adverse events in radiotherapy.," *Physica Medica: European Journal of Medical Physics*, vol. 31, pp. e39-e40., 2015.
- [25] AMARA, Marwa, ZIDI, Kamel, ZIDI, Salah, et al., "Arabic character recognition based M-SVM," in *Advanced Machine Learning Technologies and Applications*, 2014.
- [26] Amara, M., Zidi, K., Ghedira, K., & Zidi, S., "New Rules to Enhance the Performances of Histogram Projection for Segmenting Small-Sized Arabic Words.," in *Hybrid Intelligent Systems*,

Springer International Publishing., 2016, pp. 167-176.

- [27] Jiawei, H. , Micheline, K., Jian, P., , Data Mining : Concepts and techniques, J. P. J. e. K. M. HAN, Ed., Elsevier, 2011.
- [28] Fallery, Bernard, and Florence Rodhain., "Quatre approches pour l'analyse de données textuelles: lexicale, linguistique, cognitive, thématique." , .," in *XVI ème Conférence de l'Association Internationale de Management Stratégique AIMS.*, 2007.
- [29] Bouveyron, Charles, and Stephane Girard., "Classification supervisée et non supervisée des données de grande dimension.,", *La revue de Modulad*, no. 40, pp. 81-102., 2009.
- [30] R. Bellman, Dynamic programming., Princeton University Press, 1957.
- [31] T. Pavlenko, "On feature selection, curse of dimensionality and error probability in discriminant analysis.," *Journal of Statistical Planning and Inference*, vol. 115 , p. 565–584, 2003.
- [32] T. Pavlenko and D. Von Rosen., "Effect of dimensionality on discrimination.," *Statistics*, vol. 35, p. 191–213, 2001.
- [33] L. Breiman., "Bagging predictors.," *Machine Learning*, vol. 24, no. 2, p. 123–140, 1996.
- [34] Evangelista, Paul F., Mark J. Embrechts, and Boleslaw K. Szymanski. , "Taming the curse of dimensionality in kernels and novelty detection.," in *Applied soft computing technologies: The challenge of complexity.*, Springer Berlin Heidelberg, 2006, pp. 425-438.
- [35] C. Bouveyron, " Modélisation et classification des données de grande dimension : application à l'analyse d'images.," 2006.
- [36] D. Scott and J. Thompson., "Probability density estimation in higher dimensions.," in *Computer science and statistics: proceedings of the Symposium on the Interface*, North-Holland, Amsterdam, 1983.
- [37] P. Huber., "Projection pursuit.," *The Annals of Statistics.*, vol. 13, no. 2, pp. 435–525,, 1985.
- [38] S. GUÉRIF, "Réduction de dimension en apprentissage numérique non supervisé .," 2006.
- [39] Y. Bennani., "Systèmes d'apprentissage connexionnistes : sélection de variables.," *Revue d'Intelligence Artificielle.*, vol. 15, no. 3-4, 2001.
- [40] J. Hamon, "Optimisation combinatoire pour la sélection de variables en régression en grande dimension: Application en génétique animale.," Lille, 2013.
- [41] Liu, Huan, and Hiroshi Motoda., Feature selection for knowledge discovery and data mining., vol. 454, Springer Science & Business Media, 2012.
- [42] Hall, M.A., & Smith, L.A. on Kohavi et Geaorge. H. John,, "Feature Selection for Machine

- Learning: Comparing a Correlation-Based Filter Approach to the Wrapper," in FLAIRS conference, , , pp.235-239, 1999.
- [43] Lorena, Luiz HN, André CPLF Carvalho, and Ana C. Lorena., "Filter feature selection for one-class classification.," *Journal of Intelligent & Robotic Systems* , vol. 80, pp. 227-243, 2015.
- [44] Li, Jiangeng, Xiaodan Li, and Wei Zhang., "A filter feature selection method based LLRFC and redundancy analysis for tumor classification using gene expression data.," in *12th World Congress In Intelligent Control and Automation (WCICA)*, 2016.
- [45] R. Kohavi et G. H. John., "Wrappers for feature subset selection. , , no. 1-2,," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324,, 1997.
- [46] Kamarudin, Muhammad Hilmi, Carsten Maple, and Tim Watson., "Hybrid feature selection technique for intrusion detection system.," *International Journal of High Performance Computing and Networking (IJHPCN)*, 2016.
- [47] Mares, Mihaela A., Shicai Wang, and Yike Guo., "Combining Multiple Feature Selection Methods and Deep Learning for High-dimensional Data.," *Transaction of Machine Learning and Data Mining*, vol. 9, no. 1, pp. 27-45., 2016.
- [48] K. Peason, " On lines and planes of closest fit to systems of point in space.," *Philosophical Magazine*, no. 2, pp. 559-572., 1901.
- [49] Batis, Carolina, et al., "Using both principal component analysis and reduced rank regression to study dietary patterns and diabetes in Chinese adults.," *Public health nutrition*, vol. 19, no. 02, pp. 195-203, 2016.
- [50] M. Bellmann, "Factor analysis of geometric figures with four attributes : A comparison of PCA, varimax and varimin.," *Personality and Individual Differences*, vol. 90, pp. 326-331, 2016.
- [51] Kaiser, Henry F., "The varimax criterion for analytic rotation in factor analysis.," *Psychometrika* , vol. 23, no. 3, pp. 187-200, 1958.
- [52] Jolliffe, Ian T., "Rotation of principal components: choice of normalization constraints.," *Journal of Applied Statistics* , vol. 22, no. 1, pp. 29-35., 1995.
- [53] Neuhaus, J. O., and C. Wrigley., "The quartimax method: an analytic approach to orthogonal simple structure. USAF Personnel & Training Research Center Research Bulletin.," *AFPTRC-TR-54-107*, vol. iv, no. 23, 1954.
- [54] Torgerson, Warren S. , "Multidimensional scaling: I. Theory and method.," *Psychometrika* , vol. 17, no. 4, pp. 401-419, 1952.
- [55] Verleysen, Michel, and Damien François., "The curse of dimensionality in data mining and time series prediction," in *International Work-Conference on Artificial Neural Networks.*, 2005.

- [56] Fisher, Ronald A., "The use of multiple measurements in taxonomic problems.," *Annals of eugenics* , vol. 7, no. 2, pp. 179-188, 1936.
- [57] P. Barbedor, "Analyse en composantes indépendantes par ondelettes.," 2006.
- [58] Fargeas, A., et al., "Nouvelle méthode basée sur l'analyse en composantes indépendantes hautement prédictive de toxicité en cas de radiothérapie prostatique.," *Cancer/Radiothérapie* , vol. 20, no. 6, p. 737, 2016.
- [59] C. Jutten, "Calcul neuromimétique et traitement du signal : analyse en composantes indépendantes.," 1987.
- [60] J.-F. Cardoso, "Analyse en composantes indépendantes.," in *Proc. of XXXIV Journées de Statistique*,, Bruxelles , 2002.
- [61] Vapnik, Vladimir N., "An overview of statistical learning theory.," *IEEE transactions on neural networks* , vol. 10, no. 5, pp. 988-999, 1999.
- [62] Schölkopf, Bernhard, Alexander Smola, and Klaus-Robert Müller., "Nonlinear component analysis as a kernel eigenvalue problem.," *Neural computation* , vol. 10, no. 5, pp. 1299-1319, 1998.
- [63] Liu, Bingli, Ke Guo, and Li Zhang., "Kernel Principal Component Analysis in the Application of Geochemical Comprehensive Feature Extraction.," in *Geostatistical and Geospatial Approaches for the Characterization of Natural Resources in the Environment.*, 2016.
- [64] Ye, Fei, Zhiping Shi, and Zhongzhi Shi., "A comparative study of PCA, LDA and Kernel LDA for image classification.," in *International Symposium Ubiquitous Virtual Reality ISUVR'09.* , 2009.
- [65] Patra, Dipa, et al., "Mathematical Analysis of PCA, MDS and ISOMAP Techniques in Dimension Reduction.," *International Journal* , vol. 3, no. 5, 2015.
- [66] Tenenbaum, Joshua B., Vin De Silva, and John C. Langford. , "A global geometric framework for nonlinear dimensionality reduction.," *Science* , vol. 290, no. 5500 , pp. 2319-2323, 2000.
- [67] Deng, Tingquan, Jinyan Liu, and Ning Wang., "Locally linear embedding preserving local neighborhood.," in *2th International Conference Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*,, 2016.
- [68] Lai, David SW, Ozgun Caliskan Demirag, and Janny MY Leung. , "A tabu search heuristic for the heterogeneous vehicle routing problem on a multigraph.," *Transportation Research Part E: Logistics and Transportation Review* , vol. 86 , pp. 32-52., 2016.
- [69] Zhang, Hongbin, and Guangyu Sun., "Feature selection using tabu search method.," *Pattern recognition*, vol. 35, no. 3, pp. 701-711, 2002.

- [70] L. e. a. Mousin, "Feature Selection using Tabu Search with Learning Memory: Learning Tabu Search.," in *Learning and Intelligent Optimization Conference LION*, 2016.
- [71] Wang, Yong, et al. , "Feature selection using tabu search with long-term memories and probabilistic neural networks.," *Pattern recognition letters* , vol. 30, no. 7, pp. 661-670., 2009.
- [72] Yusta, Silvia Casado., "Different metaheuristic strategies to solve the feature selection problem.," *Pattern Recognition Letters*, vol. 30, no. 5, pp. 525-534., 2009.
- [73] S. Kotsiantis, "Feature selection for machine learning classification problems: a recent overview.," *Artificial Intelligence Review*, pp. 1-20., 2011.
- [74] S. Zidi, "SARR: Système d'aide à la régulation et la reconfiguration des réseaux de transport multimodal.," 2007.
- [75] B. D. Ripley, "Pattern recognition and neural networks.," 2007.
- [76] Mao, Jianchang, and Anil K. Jain. , "A self-organizing network for hyperellipsoidal clustering (HEC).," *IEEE transactions on neural networks*, vol. 7, no. 1, pp. 16-29, 1996.
- [77] Lu, Binbin, et al., "The Minkowski approach for choosing the distance metric in geographically weighted regression.," *International Journal of Geographical Information Science* , vol. 30, no. 2, pp. 351-368, 2016.
- [78] Payne, Terry R., and Peter Edwards., "Implicit feature selection with the value difference metric.," in *European Conference on Artificial Intelligence*, 1998.
- [79] El Ferchichi, Sabra, et al., "A New Feature Extraction Method Based on Clustering for Face Recognition.," *Engineering Applications of Neural Networks.*, pp. 247-253., 2011.
- [80] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review.*, vol. 5, no. 1, pp. 3-55, 2001.
- [81] P. Deheuvels, "Estimation non paramétrique de la densité par histogrammes généralisés.," *Revue de Statistique Appliquée* , vol. 25, no. 3, pp. 5-42, 1977.
- [82] Moon, Young-Il, Balaji Rajagopalan, and Upmanu Lall., "Estimation of mutual information using kernel density estimators.," *Physical Review E*, vol. 52, no. 3, p. 2318, 1995.
- [83] Kwak, Nojun, and Chong-Ho Choi. , "Input feature selection by mutual information based on Parzen window.," *IEEE Transactions on Pattern Analysis and Machine Intelligence* ., vol. 24, no. 12, pp. 1667-1671, 2002.
- [84] Schaffernicht, Erik, et al., "On estimating mutual information for feature selection.," in *International Conference on Artificial Neural Networks.*, 2010.

- [85] Cellucci, Christopher J., Alfonso M. Albano, and Paul E. Rapp., "Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms.," *Physical Review E*, vol. 71, no. 6, p. 066208., 2005.
- [86] PRINCIPE, Jose C., XU, Dongxin, et FISHER, John., " Information theoretic learning.," in *Unsupervised adaptive filtering*, vol. 1, 2000, pp. 265-319.
- [87] K. Torkkola, "Feature extraction by non-parametric mutual information maximization.," *Journal of machine learning research*, vol. 3, pp. 1415-1438., 2003.
- [88] Kapur, Jagat Narain., *Measures of information and their applications.*, Wiley-Interscience, 1994.
- [89] R. Battiti, "Using mutual information for selecting features in supervised neural net learning.," *IEEE Transactions on neural networks* , vol. 5, no. 4, pp. 537-550, 1994.
- [90] Yang, H., and John Moody. , "Feature selection based on joint mutual information.," in *Proceedings of international ICSC symposium on advances in intelligent data analysis.* , 1999.
- [91] Kwak, Nojun, and Chong-Ho Choi. , "Input feature selection by mutual information based on Parzen window.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1667-1671, 2002.
- [92] K. Torkkola, "Feature extraction by non-parametric mutual information maximization.," *Journal of machine learning research*, vol. 3, pp. 1415-1438., 2003.
- [93] N. Kwak, "Feature extraction based on direct calculation of mutual information.," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 21, no. 07, pp. 1213-1231, 2007.
- [94] Hild, Kenneth E., et al. , "Feature extraction using information-theoretic learning.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1385-1392, 2006.
- [95] Fisher, John W., and Jose C. Principe., "A methodology for information theoretic feature extraction.," in *The 1998 IEEE International Joint Conference on Neural Networks Proceedings.IEEE World Congress on Computational Intelligence.*, 1998.
- [96] Xiang, C., and D. Huang., "Feature extraction using recursive cluster-based linear discriminant with application to face recognition.," *IEEE Transactions on Image Processing* , vol. 15, no. 12, pp. 3824-3832., 2006.
- [97] Kwak, Nojun, and Chong-Ho Choi., "Feature extraction based on ICA for binary classification problems.," *IEEE Transactions on Knowledge and Data Engineering* , vol. 15, no. 6, pp. 1374-1388., 2003.
- [98] Kulkarni, Abhijit, Vaidyanathan K. Jayaraman, and Bhaskar D. Kulkarni., "Knowledge incorporated support vector machines to detect faults in Tennessee Eastman Process.," *Computers & chemical engineering*, vol. 29, no. 10, pp. 2128-2133, 2005.

- [99] Chiang, Leo H., Mark E. Kotanchek, and Arthur K. Kordon., "Fault diagnosis based on Fisher discriminant analysis and support vector machines.," *Computers & chemical engineering* , pp. 1389-1401., 2004.
- [100] Verron, Sylvain, Teodor Tiplica, and Abdessamad Kobi., "Fault detection and identification with a new feature selection based on mutual information.," *Journal of Process Control*, pp. 479-490., 2008.
- [101] El Ferchichi, Sabra, et al., "Feature Selection using an SVM learning machine.," in *Signals, Circuits and Systems (SCS), 2009 3rd International Conference*, 2009.
- [102] El Ferchichi, Sabra, et al., "Feature Clustering based MIM for a New Feature Extraction Method.," *International Journal of Computers Communications & Control* , vol. 8, no. 5, pp. 699-707., 2013.
- [103] Khouaja, Anis, and Gerard Favier., "Identification of SVD-PARAFAC based third-order Volterra models using an ARLS algorithm.," *IFAC Proceedings Volumes*, vol. 38, no. 1, pp. 767-772., 2005.
- [104] Favier, Gérard, and Thomas Bouilloc., "Parametric complexity reduction of Volterra models using tensor decompositions.," in *17th European Signal Processing Conference*, 2009.
- [105] Kibangou, Alain Y., and Gérard Favier. , "Tensor analysis-based model structure determination and parameter estimation for block-oriented nonlinear systems.," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 514-525., 2010.
- [106] Y. Zhu, "Distillation column identification for control using Wiener model.," in *Proceedings of the 1999 American Control Conference*, 1999.
- [107] Kozek, Martin, and Sabina Sinanović., "Identification of Wiener models using optimal local linear models.," *Simulation Modelling Practice and Theory*, vol. 16, no. 8, pp. 1055-1066., 2008.
- [108] J. Vörös, "Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities.," *Systems & Control Letters*, vol. 56, no. 2, pp. 99-105..
- [109] Y. Zhu, "Distillation column identification for control using Wiener model.," in *Proceedings of the 1999 American Control Conference*, 1999.
- [110] Billings, S. A., and S. Y. Fakhouri., "Identification of systems containing linear dynamic and static nonlinear elements.," *Automatica* , vol. 18, no. 1, pp. 15-26., 1982.
- [111] Li-Duwaish, H., and M. Nazmul Karim. , "A new method for the identification of Hammerstein model.," *Automatica* , vol. 33, no. 10, pp. 1871-1875, 1997.
- [112] Hachino, Tomohiro, Katsuhisa Deguchi, and Hitoshi Takata. , "Identification of Hammerstein model using radial basis function networks and genetic algorithm.," 2004.

- [113] Hunter, Ian W., and M. J. Korenberg., "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models.," *Biological cybernetics*, vol. 55, no. 2, pp. 135-144., 1986.
- [114] L. A. Zadeh, "Fuzzy sets.," *Information and control* , vol. 8, no. 3, pp. 338-353., 1965.
- [115] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant.," *Proceedings of the Institution of Electrical Engineers.*, vol. 121, no. 12, 1974.
- [116] Takagi, Tomohiro, and Michio Sugeno., "Fuzzy identification of systems and its applications to modeling and control.," *IEEE transactions on systems, man, and cybernetics* , no. 1, pp. 116-132., 1985.
- [117] A. Akhenak, "Conception d'observateurs non linéaires par approche multi-modèle: application au diagnostic.," 2004.
- [118] M. Adam Medina, "Diagnostic de défauts des systèmes à représentation multi-modèle linéaire invariant dans le temps.," 2004.
- [119] M. Rodrigues, "Diagnostic et commande active tolérante aux défauts appliqués aux systèmes décrits par des multi-modèles linéaires.," 2005.
- [120] M. Lucea, "Modélisation dynamique par réseaux de neurones et machines à vecteurs supports : contribution à la maîtrise des émissions polluantes de véhicules automobiles.," 2006.
- [121] Loghmanian, Sayed Mohammad Reza, et al. , "Polynomial NARX model structure optimization using multi-objective genetic algorithm.," 2012.
- [122] Chen, S., and S. A. Billings, "Neural networks for nonlinear dynamic system modelling and identification.," *International journal of control*, vol. 56, no. 2, pp. 319-346., 1992.
- [123] G. Wahba, Spline models for observational data. Society for industrial and applied mathematics,, 1990.
- [124] H. A. Boubacar, "Classification Dynamique de données non-stationnaires : Apprentissage et Suivi de Classes évolutives.," 2006.
- [125] Smola, Alex J., and Bernhard Schölkopf., "Learning with kernels. GMD-Forschungszentrum Informationstechnik, .," 1998.
- [126] Vapnik, Vladimir Naumovich, and Vladimir Vapnik., *Statistical learning theory.*, vol. 1, New York: Wiley, 1998.
- [127] Smola, Alex J., and Bernhard Schölkopf., "A tutorial on support vector regression.," *Statistics and computing* , vol. 14, no. 3, pp. 199-222., 2004.

- [128] Gretton, A., Doucet, A., Herbrich, R., Rayner, P. J., & Scholkopf, B., "Support vector regression for black-box system identification.," in *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, 2001.
- [129] Thissen, U., Van Brakel, R., De Weijer, A. P., Melssen, W. J., & Buydens, L. M. C., "Using support vector machines for time series prediction.," *Chemometrics and intelligent laboratory systems*, vol. 69, no. 1, pp. 35-49, 2003.
- [130] I. E. Aissi., "Modélisation, Identification et Commande Prédictive des Systèmes Non Linéaires par Utilisation des Espaces de Hilbert à Noyau Reproduisant.," Tunisie, 2009.
- [131] T. Okba, E. Ilyes, G. Tarek et M. Hassani. , "A new approach for identification of MIMO non linear system with RKHS model," *WSEAS Transactions on Information Science and Applications*, vol. 7, no. 7, pp. 50,126, 2010.
- [132] SUYKENS, Johan AK et VANDEWALLE, Joos. , "Least squares support vector machine classifiers.," *Neural processing letters*, vol. 9, no. 3, pp. 293-300, 1999.
- [133] J. Valyon et G. Horvath., "A Robust LS-SVM Regression," *International journal of computational intelligence*, vol. 3, no. 3, pp. 243-248, 2006.
- [134] S. Tao, D. Chen et W. Zhao. , "Fast pruning algorithm for multi-output LS-SVM and its application in chemical pattern classification.," *Chemometrics and Intelligent Laboratory Systems*, , vol. 96, no. 1, pp. 63 - 69, 2009.
- [135] X. Liang, G. Yan, S. Li, Z. Long, X. Li et W. Long. , "Model Predictive Control for MIMO System Based on LSSVM and Mutative Scale Chaos Optimization.," in *Second international Conference on Intelligent Computation Technology and Automation*, ., 2009.
- [136] F. Lauer., "From Support Vector Machines to Hybrid System Identification.," 2008.
- [137] GOETHALS, Ivan, PELCKMANS, Kristiaan, SUYKENS, Johan AK, et al. , "Identification of MIMO Hammerstein models using least squares support vector machines.," *Automatica*,, vol. 41, no. 7, pp. 1263-1272., 2005.
- [138] Eberhart, R., & Kennedy, J. , "A new optimizer using particle swarm theory.," in *MHS'95., Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, , 1995.
- [139] Kalani, Hadi, Mohammad Sardarabadi, and Mohammad Passandideh-Fard., " Using artificial neural network models and particle swarm optimization for manner prediction of a photovoltaic thermal nanofluid based collector.," *Applied Thermal Engineering*, vol. 113 , pp. 1170-1177, 2017.
- [140] Verdult, Vincent, Lennart Ljung, and Michel Verhaegen. , "Identification of composite local linear state-space models using a projected gradient search," *International Journal of Control*,

vol. 75, no. 16-17, pp. 1385-1398., 2002.

- [141] A. Chombart., "Commande supervisée de systèmes hybrides.," 1997.
- [142] ANTSAKLIS, Panos J., STIVER, James A., et LEMMON, Michael., "Hybrid system modeling and autonomous control systems," in *Hybrid Systems*, vol. 736, Springer Berlin Heidelberg, 1993, pp. 366-392.
- [143] P. J. Antsaklis, J. A. Stiver, and M. D. Lemmon., "Interface and controller design for hybrid control systems.," 1995.
- [144] M. Kurovsky., " Étude des systemes dynamiques hybrides par representation d'état discrete et automate hybride.," 2002.
- [145] ALUR, Rajeev, COURCOUBETIS, Costas, HENZINGER, Thomas A., et al., "Hybrid automata: An algorithmic approach to the specification and verification of hybrid systems.," in *Hybrid systems.*, Springer Berlin Heidelberg,, 1993, pp. 209-229..
- [146] NICOLLIN, Xavier, OLIVERO, Alfredo, SIFAKIS, Joseph, et al. , "An approach to the description and analysis of hybrid systems.," in *Hybrid Systems.*, Springer Berlin Heidelberg,, 1993, pp. 149-178..
- [147] C. Petri., " Kommunikation mit automaten.," 1962.
- [148] KESTEN, Yonit et PNUELI, Amir., "Timed and hybrid statecharts and their textual representation.," in *Formal Techniques in Real-Time and Fault-Tolerant Systems.*, Springer Berlin Heidelberg, 1992, pp. 591-620..
- [149] DE SANTIS, Elena, DI BENEDETTO, Maria D., DI GENNARO, Stefano, et al., "Hybrid observer design methodology.," *Public deliverable D*, , vol. 7, 2003.
- [150] Zhang, Lixian, Songlin Zhuang, and Richard D. Braatz., "Switched model predictive control of switched linear systems: feasibility, stability and robustness.," *Automatica* , vol. 67, pp. 8-21., 2016.
- [151] DI BERNARDO, Mario, MONTANARO, Umberto, ORTEGA, Romeo, et al., "Extended hybrid model reference adaptive control of piecewise affine systems.," *Nonlinear Analysis: Hybrid Systems.*, vol. 21, pp. 11-21., 2016.
- [152] A. L. Juloski, S. Weiland, and W. Heemels. , "A bayesian approach to identification of hybrid systems.," *IEEE Transaction on Automatic Control.*, vol. 50, no. 10, pp. 1520–1533,, 2005.
- [153] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari. , "A clustering technique for the identification of piecewise affine systems.," *Automatica*, vol. 39, no. 2, pp. 205–217,, 2003.
- [154] J. Roll, A. Bemporad, and L. Ljung., "Identification of piecewise affine systems via mixed-integer

- programming.," *Automatica*, vol. 40, no. 1, p. 37–50, 2004.
- [155] RAGOT, José, MOUROT, Gilles, et MAQUIN, Didier., "Parameter estimation of switching piecewise linear system.," in *42nd IEEE Conference on Decision and Control*, Maui, Hawai, USA, 2003.
- [156] R. Vidal, S. Soatto, Y. Ma, and S. Sastry., "An algebraic geometric approach to the identification of a class of linear hybrid systems.," in *42nd IEEE Conference on Decision and Control*, Maui, Hawai, USA, , 2003.
- [157] MA, Yi et VIDAL, René., "Identification of deterministic switched ARX systems via identification of algebraic varieties.," in *Hybrid Systems: Computation and Control*, Springer Berlin Heidelberg, 2005, pp. 449-465.
- [158] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino., "A bounded-error approach to affine system identification.," *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1567– 1580,, 2005.
- [159] JULOSKI, A. Lj et WEILAND, S, "A Bayesian approach to the identification of piecewise linear output error models.," *IFAC Proceedings Volumes*, vol. 39, no. 1, pp. 374-379., 2006.
- [160] FERRARI-TRECCATE, Giancarlo, MUSELLI, Marco, LIBERATI, Diego, et al., " A clustering technique for the identification of piecewise affine systems.," *Automatica*, vol. 39, no. 2, pp. 205-217., 2003.
- [161] NAKADA, Hayato, TAKABA, Kiyotsugu, et KATAYAMA, Tohru., "Identification of piecewise affine systems based on statistical clustering technique.," *Automatica*, vol. 41, no. 5, pp. 905-913., 2005.
- [162] VIDAL, René et ANDERSON, Brian DO., "Recursive identification of switched ARX hybrid models: Exponential convergence and persistence of excitation.," in *43rd IEEE Conference on Decision and Control, CDC*, 2004.
- [163] R. VIDAL, "Recursive identification of switched ARX systems.," *Automatica*, vol. 44, no. 9, pp. 2274-2287., 2008.
- [164] L. Bako and R. Vidal., "Identification of switched mimo arx models.," in *Proc. of the 11th Int. Conf on Hybrid Systems : Computation and Control (HSCC)*, St. Louis, MO, USA, , 2008.
- [165] BAKO, Laurent, MERCÈRE, Guillaume, et LECÈUCHE, Stéphane., "On-line structured identification of switching systems with possibly varying orders.," in *European Control Conference (ECC)*, 2007.
- [166] V. Verdult and M. Verhaegen. , "Subspace identification of piecewise linear systems.," in *Proc. of the 43rd IEEE Conf. on Decision and Control (CDC)*, , Atlantis, Paradise Island, Bahamas, 2004.
- [167] S. Chiu., "Fuzzy model identification based on cluster estimation.," *Journal of intelligent and*

fuzzy systems, vol. 2, p. 267–278, 1994.

- [168] A. Shigeo., *Neural networks and fuzzy systems theory and applications.*, USA: Kluwer Academic Publishers,, 1997.
- [169] F. Lauer., "From support vector machines to hybrid system identification.," France, 2008.
- [170] LAUER, Fabien et BLOCH, Gérard. , "A new hybrid system identification algorithm with automatic tuning.," in *IFAC Proceedings Volumes.*, 2008.
- [171] LAUER, Fabien et BLOCH, Gérard. , "Switched and piecewise nonlinear hybrid system identification.," in *Hybrid Systems: Computation and Control in International Workshop on Hybrid Systems*, 2008.
- [172] K. Boukharouba., "Modélisation et classification de comportements dynamiques des systèmes hybrides.," 2011.
- [173] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari., "A clustering technique for the identification of piecewise affine systems.," *Automatica*, vol. 39, no. 2, p. 205–217, 2003.
- [174] H. Nakada, K. Takaba, and T. Katayama. , "Identification of piecewise affine systems based on statistical clustering technique.," *Automatica*, vol. 41, p. 905–913, 2005.
- [175] D. Stamatis., *Failure mode effect analysis : Fmea from theory to execution.*, Asq Press, 2003.
- [176] G. Zwingelstein., *Diagnostic des défaillances, théorie et pratique pour les systèmes industriels.*, Hermès Science Publications, , Juin 1995. .
- [177] H. Derbel., "Diagnostic à base de modèles des systèmes temporisés et d'une sous-classe de systèmes dynamiques hybrides.," Grenoble I, , 2009.
- [178] M. Chadli., "Stabilité et commande des systèmes décrits par les multi-modèles.," 2002.
- [179] T. Zouari, K. Laabidi et M. Ksouri., "Multimodel approach applied for failure diagnosis.," *International Journal on Sciences and Techniques of Automatic control and computer engineering IJ-STA*, ., vol. 2, no. 1, pp. 500–515,, 2008.
- [180] R. Orjuela., "Contribution à l'estimation d'état et au diagnostic des systèmes représentés par des multi modèles.," 2008.
- [181] COLOMER, J., MELENDEZ, J., et GAMERO, F. I., "Pattern recognition based on episodes and DTW. Application to diagnosis of a level control system.," in *16th International Workshop on Qualitative Reasoning*, 2002.
- [182] R. Casimir., "Diagnostic des défauts des machines asynchrones par reconnaissance des formes.," 2003.

- [183] O. Ondel., "Diagnostic par reconnaissance des formes : Application à un ensemble convertisseur-machine asynchrone.," 2007.
- [184] T. Kempowski., "Surveillance de procédés à base de méthodes de classification : d'un outil d'aide pour la détection et le diagnostic des défaillances.," Toulouse, 2004.
- [185] S. Verron., "Diagnostic et surveillance des processus complexes par réseaux de neurones bayésiennes.," 2007.
- [186] B. Braden., "The Surveyor's Area Formula.," *The College Mathematics Journal*, vol. 17, no. 4, p. 326–337, 1986.
- [187] TAX, David MJ et DUIN, Robert PW., " Support vector domain description.," *Pattern recognition letters*,, vol. 20, no. 11, pp. 1191-1199., 1999.
- [188] TAX, David MJ et DUIN, Robert PW., "Support vector data description.," *Machine learning*, vol. 54, no. 1, pp. 45-66, 2004.
- [189] ASN, *Bilan national des inspections réalisées dans les services de radiothérapie externe sur le thème de la radioprotection des patients par une approche sur les facteurs organisationnels et humains (Campagne 2007)*,, 2008.
- [190] J.-L. GODET, "Les actions de contrôle de l'Autorité de sûreté nucléaire (ASN) dans le domaine de la radiothérapie.," *Cancer/Radiothérapie*, vol. 11, no. 6, pp. 313-316., 2007.
- [191] *Lettre circulaire DGSNR/SD7/n°1027/2006 du 19 avril 2006 portant sur la prévention des incidents graves en radiothérapie par une approche sur les facteurs organisationnels et humains ;*
- [192] "Le facteur humain : un défi pour les autorités de sûreté nucléaire.," OCDE, 2004 .
- [193] M. Bourrier, "Facteurs organisationnels : du neuf avec du vieux.," 2003.
- [194] P.DUBIAU., "Gestion d'une urgence radiologique : Organisation et fonctionnement." *Editions Techniques de l'Ingénieur*..
- [195] J.-L. GODET, "Les actions de contrôle de l'Autorité de sûreté nucléaire (ASN) dans le domaine de la radiothérapie.," *Cancer/Radiothérapie*, vol. 11, no. 6, pp. 313-316., 2007.
- [196] P.DUBIAU., "Gestion d'une urgence radiologique : Retour d'expérience et gestion post-accidentelle." *Editions Techniques de l'Ingénieur*..
- [197] D. S. J.-M. e. E. F. PEIFFERT, "L'accident d'Épinal: passé, présent, avenir.," *Cancer/Radiothérapie*, , vol. 11, no. 6, pp. 309-312., 2007.
- [198] H. PETIT, "Analyse et gestion du risque dans l'aéronautique.," in *Congrès de la Société Française*

de Chirurgie Oncologique SFCO , 16 Septembre 2011.

- [199] "Flight safety foundation, Norme de base en matière de risque aéronautique, version 3,," juin 2010.
- [200] BOURDEAUX, Isabelle et GILBERT, Claude., "Procédures de REX, d'apprentissage et de vigilance organisationnelles: approches croisées.," *Programme Risques Collectifs et Situation de Crise,*, vol. 36, 1999.
- [201] "Direction générale de l'Aviation civile et direction de la sécurité de l'Aviation civile, du traitement des événements à la gestion des risques,," 2011.
- [202] J JOUNEAU, "Outils et techniques de gestion des risques adaptés à la prévention des IAS,," 2010.
- [203] C. TEA, "Retour d'expérience et données subjectives: quel système d'information pour la gestion des risques?: le cas de la sécurité dans le transport ferroviaire.," Paris, 2009.
- [204] F.COCHOY., "Sur les traces de la traçabilité : Histoire, diagnostic et propositions.," *Editions Techniques de l'Ingénieur,*, 10 nov 2006.
- [205] C.DOUCET., "Traçabilité et norme ISO 9001:2000 : Objectifs, moyens et méthodes." *Editions Techniques de l'Ingénieur.*
- [206] D.NAIRAUD., "Traçabilité des denrées alimentaires : Aspects généraux." *.s.l : Editions Techniques de l'Ingénieur.*
- [207] K.LENTZ., "Aspects juridiques de la traçabilité." *Editions Techniques de l'Ingénieur.*
- [208] J.-L.VIRUEGA., "La traçabilité : un enjeu stratégique." *Editions Techniques de l'Ingénieur.*
- [209] A.DASSENS, R.LAUNAY., "Étude systémique de l'analyse de risques : Présentation d'une approche globale." *.s.l : Editions Techniques de l'Ingénieur.*
- [210] J.BUCKI., "Analyse décisionnelle des systèmes complexes." *.s.l : Editions Techniques de l'Ingénieur.*
- [211] J.-M.ROUSSEAU, A.LARGIER., "Industries à risques : conduire un diagnostic organisationnel par la recherche de facteurs pathogènes." *.s.l : Editions Techniques de l'Ingénieur.*
- [212] L.SAMSON., "Approche systémique des facteurs humains dans l'entreprise." *.s.l : Editions Techniques de l'Ingénieur.*
- [213] G.DONNADIEU., "Motiver les hommes dans les entreprises : approche systémique." *.s.l : Editions Techniques de l'Ingénieur.*
- [214] Y.MORTUREUX., "Dimensions humaine et organisationnelle dans le retour d'expérience." *.s.l :*

- [215] LARTIGAU, E., COCHE-DEQUEANT, B., DUMORTIER, V., et al., "La cellule de retour d'expérience (CREx) en radiothérapie: un outil indispensable au management de la sécurité. Une évaluation après deux ans d'activité au centre Oscar-Lambret.," *Cancer/Radiothérapie*, vol. 12, no. 6, pp. 610-613., 2008.
- [216] Biskri, Ismaïl, and Sylvain Delisle., "Les n-grams de caractères pour l'aide à l'extraction de connaissances dans des bases de données textuelles multilingues.," in *Proceedings of TALN*, 2001.
- [217] F. Y. CHOI, "Advances in domain independent linear text segmentation.," in *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference. Association for Computational Linguistics*, 2000.
- [218] C. E. Shannon, "Prediction and entropy of printed English.," *Bell Labs Technical Journal*, vol. 30, no. 1, pp. 50-64., 1951.
- [219] C. E. SHANNON, "A mathematical theory of communication.," *ACM SIGMOBILE Mobile Computing and Communications Review*, , vol. 5, no. 1, pp. 3-55., 2001.
- [220] M. DAMASHEK, "Gauging similarity with n-grams: Language-independent categorization of text.," *Science*, vol. 267, no. 5199, p. 843., 1995.
- [221] Mayfield, James, and Paul McNamee. , "Indexing using both n-grams and words.," *TREC*, 1998.
- [222] P. D. TURNEY, "Mining the web for synonyms: PMI-IR versus LSA on TOEFL.," in *European Conference on Machine Learning*. , 2001.
- [223] HAMERS, Lieve, HEMERYCK, Yves, HERWEYERS, Guido, et al. , "Similarity measures in scientometric research: the Jaccard index versus Salton's cosine formula.," *Information Processing & Management*, vol. 25, no. 3, pp. 315-318., 1989.
- [224] R. e. a. Vázquez-Jiménez, "Applying the chi-square transformation and automatic secant thresholding to Landsat imagery as unsupervised change detection methods.," *Journal of Applied Remote Sensing*, vol. 11, no. 1, pp. 016016-016016., 2017.
- [225] Salah Zidi, Thibaut Julien, Anis Mjirda, Fouad Maaloul,, "Textual extraction and classification for medical risk management: A new Risk Management Platform to manage undesired medical events.," in *4th International Conference on Advanced Logistics and Transport (ICALT)*, 2015.
- [226] DURAD, MUHAMMAD HANIF, and MUHAMMAD NAVEED AKHTAR., "Analysis Of Sir Epidemic Models.," *VFAST Transactions on Software Engineering*, vol. 6, no. 1, pp. 1-6., 2015.
- [227] Kuznetsov, Yu A., and Carlo Piccardi., "Bifurcation analysis of periodic SEIR and SIR epidemic models.," *Journal of mathematical biology*, vol. 32, no. 2, pp. 109-121., 1994.

- [228] Andersson, Hakan, and Tom Britton., Stochastic epidemic models and their statistical analysis., vol. 151, Springer Science & Business Media, , 2012.
- [229] Gómez-Vallejo, H. J., et al. , "A case-based reasoning system for aiding detection and classification of nosocomial infections.," *Decision Support Systems*, vol. 84, pp. 104-116. , 2016.
- [230] MOULAH, Tarek; NASRI, Salem; GUYENNET, Herve., "Fault-tolerant flooding through formal concept analysis for wireless sensor networks.," in *21st International Conference in Telecommunications (ICT)*,, 2014.
- [231] K. E. Wolff, "A first course in Formal Concept Analysis," *StatSoft '93*, pp. 429-438,, 1994.
- [232] N. Messai, "Analyse de concepts formels guidée par des connaissances de domaine: Application à la découverte de ressources génomiques sur le Web.," 2009.
- [233] Boumedjout Lahcen and Leonard Kwuida. , "Lattice Miner: A Tool for Concept Lattice Construction and Exploration.," in *Suplementary Proceeding of International Conference on Formal concept analysis (ICFCA'10)*, , 2010.
- [234] MADZAROV, Gjorgji, GJORGJEVIKJ, Dejan, et CHORBEV, Ivan. , "A multi-class SVM classifier utilizing binary decision tree.," *Informatica*, vol. 33, no. 2, 2009.
- [235] "Communicable Disease Toolkit: Case management of epidemic-prone diseases.," world health organization, , iraq , march 2003.
- [236] " Infection prevention and control of epidemic-and pandemic-prone acute respiratory diseases in health care: WHO interim guidelines.," World Health Organization., 2007.
- [237] Lu, Lu, et al. , "Middle East respiratory syndrome coronavirus (MERS-CoV): challenges in identifying its source and controlling its spread.," *Microbes and Infection* , vol. 15, no. 8, pp. 625-629., 2013.
- [238] B. A. T. M. L. L. Salah Zidi, "Formal Concept Analysis and Statistical learning theory for aiding detection and classification of epidemics," *Wulfenia*, pp. 63-78, 2017.