

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
École Nationale Polytechnique

Département : Génie Industriel



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Mémoire de Magister

en

Génie Industriel

Thème :

**Une approche multi-agents à base de SVM
pour la modélisation cognitive du processus
de prise de décision
Application au dépistage automatique de la maladie de
la Scoliose**

Proposé et Dirigé par :

Pr. BELMOKHTAR Oumhani

Etudié par :

Mr. MEGHOUFEL Mâamar

ENP 2012

REMERCIEMENTS

Ce projet a été effectué au département du Génie Industriel de l'École Nationale Polytechnique d'Alger. Tout au long de ma formation magistrale, j'y ai bénéficié du savoir-faire et d'une grande rigueur de recherche. Je tiens à exprimer ma plus profonde gratitude à tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce mémoire de magister.

Je suis profondément reconnaissant envers ma directrice de recherche le Professeur O.BELMOKHTAR qui m'a encadré pendant ces années de magister. Elle a mis en place les conditions propices au bon développement de mes travaux et a su prodiguer de inestimables réflexions. Je suis heureux de pouvoir lui exprimer ici mes sincères remerciements.

De même, je tiens à remercier tout particulièrement le docteur M. AIB, qui fut pour moi plus qu'un Co-directeur. Par son dynamisme sans bornes, son dévouement, son ouverture d'esprit et son exigence, il a su me transmettre l'énergie nécessaire à l'accomplissement d'un travail ambitieux. Je lui suis très reconnaissant, sur un plan professionnel et humain.

Last but not least, un grand merci à ma famille : mon père et ma mère, mes frères et sœurs, en particulier Ali & Brahim, ainsi qu'à tous les membres de ma famille. Sans leur soutien et leur estime, je n'aurais jamais eu le courage d'entreprendre un tel travail. Sans leur bienveillance et leur compréhension, je n'aurais jamais pu le finir.

DÉDICACES

A mes chers parents.

A mes frères et sœurs.

A ma fiancée

A toute ma famille.

A la mémoire de mon oncle Mohamed et de sa femme Kheira

A tous ceux que j'aime

ملخص :

غالبا ما ينظر إلى اتخاذ القرار نتيجة اختيار مقارن شامل بين مختلف البدائل. للأسف، أظهرت العديد من الدراسات التجريبية حدود هذا النموذج. صانع القرار يعمل في محيط من المشاكل معقد جدا ومنظم بشكل سيئ وعلاوة على ذلك، يتعرض صانع القرار لقائمة طويلة من القصور المعرفي.

الهدف من هذا العمل هو اقتراح نمذجة معرفية. في هذا المنظور من النمذجة، شرح القرار يقودنا أولا الى فهم جوهر الفعل بدلا من شرح اتخاذ القرار بشكل مباشر على الاطراد الملحوظ في اتخاذ القرارات من قبل نموذج رسمي.

لهذا اقترحنا الأنظمة متعددة الوكلاء مع هندسة تنظيمية قائمة على السبورة السوداء حيث قمنا بنمذجة الوكلاء بواسطة الفواصل الضخمة الهوامش. سنستعمل هذا النموذج للكشف الأوتوماتيكي عن مرض الجنف.

الكلمات المفتاحية : اتخاذ القرار، معرفية، محيط معقد جدا، النمذجة، الأنظمة متعددة الوكلاء، الفواصل الضخمة الهوامش.

Résumé :

La décision est souvent considérée comme le résultat d'un choix comparatif exhaustif entre les diverses solutions possibles. Malheureusement, de nombreuses études empiriques ont montré les limites de ce modèle. Le décideur opère dans un environnement de problèmes très complexe et mal structuré. De plus, il est exposé à une longue liste de biais cognitifs.

L'objectif de ce travail est de proposer une modélisation cognitive. Dans cette perspective de modélisation, expliquer une décision nous amène d'abord à comprendre l'acte de décision au lieu d'expliquer directement des régularités observées dans des décisions par un modèle formel.

Nous avons proposé un Système Multi-Agents cognitif avec une architecture organisationnelle à base de tableau noir et dont les agents sont modélisés avec les Séparateurs à Vaste Marge que nous avons appliqué au dépistage automatique de la maladie de la scoliose.

Mots clé : décision, cognitive, environnement complexe, modélisation, Système Multi-Agents, Séparateurs à Vaste Marge.

Abstract :

The decision has often been regarded as the result of an exhaustive and comparative choice between the various possible solutions. Unfortunately, many empirical studies showed the limits of this model. The decision maker operates in a very complex and badly structured environment. Furthermore, he is exposed to a long list of cognitive skews.

The objective of this work is to propose a cognitive modeling form. In this perspective, to explain a decision, we have to understand the act of decision instead of directly explain regularities observed in decisions by a formal model.

We proposed a cognitive Multi-Agents System with a blackboard organizational architecture and whose agents are modeled with the Support Vector Machines that we applied to the automatic detection of the disease of the scoliosis.

Key words: decision, cognitive, complex environment, modeling, Multi-Agents System, Support Vector Machines

TABLE DES MATIÈRES

<i>REMERCIEMENTS</i>	I
<i>DÉDICACES</i>	III
<i>RÉSUMÉ</i>	IV
<i>TABLE DES MATIÈRES</i>	V
<i>LISTE DES TABLEAUX</i>	X
<i>LISTE DES FIGURES</i>	XI
<i>LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES</i>	XIII
INTRODUCTION GÉNÉRALE	1
PARTIE 1 : ANALYSE COGNITIVE	6
CHAPITRE 1 : PRINCIPAUX CONCEPTS COGNITIFS DU DÉCIDEUR ET DE LA DÉCISION	7
1.1 Introduction.....	7
1.2 Cognition et décision	8
1.2.1 Processus de décision et le temps	11
1.2.2 Étapes de la décision.....	12
1.2.2.1 Définition du problème.....	13
1.2.2.2 Définition et utilisations de buts	14
1.2.2.3 Résolution du problème	15
1.2.2.4 Évaluation	15
1.2.2.5 Processus complet de décision.....	16
1.3 Analyse cognitive du décideur.....	16
1.3.1 Propriétés du décideur.....	18
1.3.2 Le décideur, une vue systémique	19
1.3.3 Le décideur dans un environnement complexe.....	22
1.4 Conclusion	23

PARTIE 2 : CONCEPTION DE L'OUTIL D'AIDE À LA DÉCISION	25
CHAPITRE 2 : MODÉLISATION D'UN DÉCIDEUR INDIVIDUEL PAR LES SÉPARATEURS À VASTE MARGE.....	26
2.1 Introduction.....	26
2.2 Prise de décision et intelligence artificielle	26
2.3 Intelligence artificielle et classification	27
2.3.1 Entre PMC et SVM.....	28
2.3.1.1 Perceptron multicouche	28
2.3.1.2 Séparateur à Vaste Marge	31
2.3.1.3 Choix final	33
2.4 Fondement théorique des SVM	34
2.4.1 Problème de la classification linéaire	35
2.4.1.1 Notion de marge.....	37
2.4.1.2 Marge Molle.....	40
2.4.2 Problème de la classification non linéaire	43
2.4.2.1 Noyaux.....	45
2.4.3 Performances en généralisation des SVM	47
2.4.3.1 Principe de régularisation	49
2.4.3.2 Empirical Risk Minimization (ERM)	51
2.4.4 Extension des SVM vers le cas multi-classes.....	52
2.4.4.1 One versus all.....	53
2.4.4.2 One versus one.....	53
2.5 Conclusion	53
CHAPITRE 3 : LES SYSTÈMES MULTI-AGENTS, VERS UNE INTELLIGENCE COLLECTIVE ORIGINELLE	54
3.1 Introduction.....	54
3.2 Intelligence collective originelle.....	55
3.2.1 Caractéristiques de l'intelligence collective originelle.....	57

3.2.1.1 Un tout émergent.....	57
3.2.1.2 Un espace holoptique.....	57
3.2.1.3 Un contrat social.....	57
3.2.1.4 Une architecture polymorphe.....	58
3.2.1.5 Un objet-lien en circulation.....	58
3.2.1.6 Une organisation apprenante.....	58
3.2.1.7 Une économie du don.....	58
3.3 Évolution vers une Intelligence Artificielle Distribuée.....	59
3.4 Systèmes Multi-Agents.....	60
3.4.1 Concept d'agent dans un SMA.....	61
3.4.1.1 Caractéristiques d'un agent cognitif.....	62
3.4.1.2 Fonctionnement d'un agent cognitif.....	63
3.4.2 Caractéristiques des systèmes multi-agents.....	66
3.4.2.1 Structures organisationnelles.....	67
3.4.2.2 Coopération.....	68
3.4.2.3 Coordination.....	70
3.4.2.4 Négociation.....	70
3.4.2.5 Communication.....	71
3.5 Modélisation <i>Blackboard</i> de la structure organisationnelle.....	71
3.5.1 La dualité du terme organisation.....	72
3.5.2 Le choix d'une organisation.....	72
3.5.3 L'architecture à base de tableau noire.....	73
3.6 Les cartes Auto-organisatrices comme dispositifs de contrôle.....	76
3.7 Conclusion.....	77
PARTIE 3: APPLICATION AU DÉPISTAGE AUTOMATIQUE DE LA MALADIE DE LA SCOLIOSE.....	78

CHAPITRE 4 : RECONSTRUCTIONS 3D DU SQUELETTE HUMAIN POUR UNE NOUVELLE APPROCHE DE DEPISTAGE DE LA MALADIE DE LA SCOLIOSE.....	79
4.1 Introduction.....	79
4.2 Pathologie de la scoliose.....	80
4.3 Dépistage conventionnel de la scoliose.....	82
4.3.1 Examen clinique de la scoliose.....	82
4.3.1.1 Examen dans le plan frontal.....	83
4.3.1.2 Examen dans le plan sagittal.....	83
4.3.2 Examen radiographique.....	84
4.4 Dépistage de la scoliose à l'aide de reconstruction 3D du rachis.....	86
4.4.1 L'imageur biplan à détecteurs gazeux.....	87
4.4.1.1 Principe physique.....	87
4.4.1.2 Spécifications techniques de l'imageur biplan.....	88
4.4.1.3 Avantages de l'imageur biplan.....	90
4.4.2 Reconstruction 3D du rachis.....	92
4.5 Problématique et objectifs de l'application.....	94
4.6 Conclusion.....	95
CHAPITRE 5 : MÉTHODOLOGIE : PRÉTRAITEMENT DES DONNÉES ET SIMULATION DE LA PRISE DE DÉCISION.....	96
5.1 Introduction.....	96
5.2 Base de données.....	96
5.3 Prétraitement de la base des données.....	99
5.3.1 Normalisation selon l'axe des Z.....	101
5.3.2 Création d'un centre de masse.....	102
5.3.3 Recalage.....	103
5.4 Extraction des caractéristiques principales.....	104
5.5 Simulation de la prise de décision individuelle.....	106
5.5.1 Choix de la fonction noyau.....	107

5.5.2	Paramètres à optimiser	108
5.5.3	Optimisation des paramètres	109
5.5.3.1	Initialisation des paires (C, σ)	110
5.5.3.2	Partition de la base des données.....	111
5.5.3.3	Validation croisée	113
5.5.4	Programmation numérique.....	115
5.5.5	Résultats	116
5.6	Simulation de la prise de décision collective	118
5.6.1	Les Agents	119
5.6.2	Le Tableau Noir	120
5.6.3	Le dispositif de contrôle.....	121
5.6.3.1	Topologie de la carte auto-organisatrice.....	121
5.6.3.2	Apprentissage de la carte auto-organisatrice	122
5.6.4	Programmation et résultats	123
5.7	Discussion des résultats	127
	CONCLUSION GÉNÉRALE.....	131
	ANNEXE I PERCEPTRON MULTI-COUCHE (MLP)	135
	ANNEXE II RÉSEAUX DE NEURONES DE KOHONEN ET KNN	142
	ANNEXE III VALEURS DES PARAMÈTRES C & σ DES SVM OPTIMAUX	145
	LISTE DES RÉFÉRENCES BIBLIOGRAPHIQUES	147

LISTE DES TABLEAUX

	Page
Tableau 5.1 : Tableau récapitulatif de la base de données.....	97
Tableau 5.2 : Représentation <i>dataset</i> de la base de données.	99
Tableau 5.3 : Partition en sous blocs de la base de données initiale.....	113

LISTE DES FIGURES

	Page
Figure 1.1 : Mécanisme de réduction de l'entropie.	21
Figure 2.1 : Hyperplan optimal et marge d'un SVM. Les ronds représentent des exemples de classe -1 et les carrés, des exemples de classe +1.....	38
Figure 2.2 : Illustration de la méthode des SVM pour les classes non séparables. Introduction des variables molles : à chaque point d'apprentissage non classé du bon côté par le séparateur est associée une variable d'erreur.	41
Figure 2.3 : Illustration de projection de données de l'espace e départ vers l'espace support	44
Figure 2.4 : Effet du paramètre σ . De gauche à droite σ^2 diminue	47
Figure 2.5 : La dimension VC pour l'hyperplan dans un espace de dimension 2 (donc pour la droite) est de 3.....	50
Figure 2.6 : Illustration du principe ERM appliquée par les SVM. La borne sur l'erreur de généralisation permet d'optimiser en parallèle la complexité de la fonction et de l'erreur empirique mesurée sur l'échantillon	52
Figure 3.1 : Schématisation du fonctionnement d'un agent cognitif	64
Figure 3.2 : Représentation d'un Système Multi-Agents.....	66
Figure 3.3 : Une architecture de système à base de tableau noire.	75
Figure 4.1 : Déformation scoliotique de la colonne vertébrale : l'axe X est orthogonal au plan frontal, l'axe Y est orthogonal au plan sagittal	80
Figure 4.2 : Mesure de l'angle de Cobb.....	85
Figure 4.3 : Imageur biplan à détecteurs gazeux : (a) Principe de l'imageur vue de dessus; (b) principe de l'imageur vue de face.	88
Figure 4.4 : Exemple d'une radiographie de face et de profil.	89
Figure 4.5 : Comparaison entre une image radiographique classique en (a) et une image radiographique obtenue avec l'imageur biplan en (b)	91
Figure 4.6 : Reconstruction 3D du squelette à partir de radiographies bipanes	93
Figure 4.7 : Mesure de l'angle de Cobb à partir d'une reconstruction 3D du rachis.....	94

Figure 5.1 : L'ensemble de vertèbres sur le rachis .	98
Figure 5.2 : Représentation de quatre colonnes vertébrales de la base de données dans l'espace 3D.	100
Figure 5.3 : Représentation de quatre colonnes vertébrales normalisées.	102
Figure 5.4 : Création d'un centre de masse pour les quatre colonnes vertébrales.	103
Figure 5.5 : Minimisation de l'erreur de rotation entre quatre colonnes vertébrales.	104
Figure 5.6 : Hétérogénéité des distances entre les centres de masse des vertèbres.	105
Figure 5.7 : La colonne vertébrale de la Figure 4.12 représentée par un ensemble de points équidistants échantillonnés par interpolation.	106
Figure 5.8 : Schématisation de la méthode <i>Grid-Search</i> pour l'optimisation des paramètres C et σ	110
Figure 5.9 : Partition de la base de données en sous blocs.	112
Figure 5.10 : Séquences de déroulement de la crose validation 5-fois.	114
Figure 5.11 : Programmation des SVM sous MATLAB	116
Figure 5.12 : Les meilleurs SVM choisis selon leur score de validation croisée (SVC) et le taux d'erreur de test correspondant.	117
Figure 5.13 : Système Multi-Agents décisionnel avec une architecture à base de Tableau Noir.	119
Figure 5.14 : Réseau de Kohonen avec carte rectangulaire de dimension $S=2 \times 3 = 6$ SVM.	122
Figure 5.15 : Algorithme d'apprentissage de Kohonen.	123
Figure 5.16 : Topologie finale de la structure organisationnelle des Agents.	124
Figure 5.17 : Décisions collective des six gents générées sur la base d'Apprentissage.	125
Figure 5.18 : Décisions collective des six gents générées sur la base de données Test.	127

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

PMC	Perceptron Multicouche (MLP en Anglais)
SVM	Séparateurs à Vaste Marge
SV	Vecteurs Supports
VC	Vapnik-Chervonenkis
ERM	Empirical Risk Minimization
KKT	Karush-Kuhn-Tucker
IAD	Intelligence Artificielle Distribuée
SMA	Systèmes Multi-Agents
RBF	Radial Basis Function
2D	Espace deux dimensionnels
3D	Espace trois dimensionnels
CT-Scan	Tomodensitométrie assistée par ordinateur
IRM	Imagerie par résonance magnétique
Rayons X	Imagerie par rayonnement électromagnétique à haute fréquence constitué de photons.
eV	Énergie mesurée en électronvolt ($1\text{eV} = 1,6 \cdot 10^{-19}$ Joule)
Ghz	Gigahertz est l'unité dérivée de fréquence du système international (si). Elle est équivalente à une oscillation par seconde (s^{-1} ou $1/\text{s}$)
Mo	Méga-octets unité de mesure en informatique mesurant la quantité de données

INTRODUCTION GÉNÉRALE

« Ce qu'il y a d'étonnant dans la décision, c'est qu'en théorie il n'y a pas de solution mais en pratique on choisit quand même »

Keen (1997)

La décision est au cœur d'une des grandes énigmes de la pensée et de l'action. Elle semble partout présente dans l'histoire. L'examen de la réalité des décisions montre que décider ne correspond pas à une phase précise, clairement identifiable où tombe le couperet. La langue française emploie l'expression « prendre une décision » un peu comme si la décision était un objet identifiable. Le point d'arrivée est confondu avec le processus. La langue anglaise, qui emploie l'expression de « decision making process » (la fabrication de la décision) rend mieux compte de ce fait. La décision est un cheminement : elle se construit, se négocie, suit des voies sinueuses au cours du temps. Elle pourrait sembler insaisissable.

Pourtant, la décision a longtemps été et encore souvent considérée comme le résultat d'un choix comparatif exhaustif entre les diverses solutions possibles. Le décideur doit mesurer avec soin les risques et les issues possibles de chaque formule, peser leurs avantages et inconvénients pour retenir finalement celle qui représente le meilleur rapport d'utilité. Cette analyse, en termes de calcul rationnel, postule l'existence d'un acteur unique qui agirait en vertu de préférences hiérarchisées en fonction d'une meilleure utilité fixée à priori.

Malheureusement, de nombreuses études empiriques ont montré les limites de ce séduisant modèle. Aujourd'hui, les processus de décision dans une organisation s'écartent sensiblement de cette représentation mythique et simpliste qui fait abstraction de plusieurs aspects nécessaires pour une représentation crédible du processus décisionnel.

Tout d'abord, il est devenu clair que le décideur moderne opère dans un environnement de problèmes très complexe et mal structuré où il doit intégrer beaucoup de facteurs intra- et

extra-organisationnelles ainsi que toutes les logiques qui leur sont attachées. La perception croissante de la complexité décisionnelle se traduit par un problème de contingence d'une trop grande liberté de choix où, le choix lui-même, est maintenant envisagé comme élément répétitif dans un long processus de traitement d'information. Le rôle du décideur implique donc d'essayer de trouver les moyens qui puissent réduire l'incertitude générée par cet environnement hostile qui décourage souvent toute anticipation. [Daniel, 1996]

De plus, le décideur lui-même ne ressemble guère à ce décideur aux capacités exceptionnelles. En effet, le décideur ne domine ni l'environnement auquel il est confronté ni l'organisation qu'il dirige. Il n'est pas non plus, ainsi que le mythe le dépeint, idéalement lucide et rationnel. Dans cette perspective, le terme de rationalité limitée, cher à H.A. Simon [Simon, 1981] prend tout son sens. La liste est longue des biais cognitifs auxquels le décideur est exposé. Le grand décideur mythique, *l'homo economicus*, y devient une aberration, aucun décideur ne peut tout calculer et tout prévoir dans une organisation un tant soit peu complexe aujourd'hui.

A partir de ces affirmations, nous dérivons les principaux objectifs de notre travail de mémoire de magister qui se veut exploratoire et prospectif, où nous allons proposer un nouveau modèle d'aide à la décision qui doit être en mesure d'orienter, fidèlement, le décideur qui se voit souvent incapable de prendre une décision sensée et cohérente.

L'aide à la décision est une activité qui s'appuie sur des modèles clairement explicites mais non nécessairement complètement formalisés, et qui aide à obtenir des éléments de réponses aux questions que se pose un acteur inclus dans le processus de décision. Le modèle tient une place centrale dans notre approche. En effet, Le plus grand déficit auquel nous devons faire face aujourd'hui se trouve au niveau des modèles. Ce n'est qu'avec des modèles que nous arriverons à réduire l'écart énorme qui existe entre les prétentions des théories et la réalité de l'observation empirique. Par contre, créer un modèle signifie non seulement décrire quelque chose, mais également l'expliquer en grande partie. La conviction que le modèle est en fait la pierre angulaire de la recherche devient de plus en plus populaire et en vient même

à substituer la respectable notion de théorie, notamment dans toutes les approches systémiques.

Au cours de ce travail, un des postulats sur lesquels nous nous sommes basé est que le modèle que nous proposons doit posséder une pertinence au niveau du sens. C'est-à-dire, nous visons une modélisation que l'on pourrait appelé essentialiste, opposée à une modélisation behavioriste. Dans cette perspective de modélisation, expliquer une décision nous amène d'abord à comprendre l'acte de décision au lieu d'expliquer directement des régularités observées dans des décisions par un modèle formel. En effet, la qualité d'un modèle peut se mesurer dans deux dimensions: celle de l'homomorphie avec le phénomène modélisé et de la maniabilité pour répondre à une question de recherche concrète.

Ainsi, nous avons découpé ce document en trois parties avec un total de cinq chapitres. Ce découpage reflète la démarche exploratoire et prospective que nous avons adoptée.

- La première partie est une analyse cognitive, cette partie est composée du premier chapitre qui traite les principaux concepts cognitifs du décideur et de la décision où nous allons exposer le paradigme des sciences cognitives qui sont un ensemble de disciplines concernées par la recherche sur la pensée humaine et le cerveau humain. Étant donné que le de décideur pense, ces nouvelles disciplines peuvent nous fournir un nombre important de théories, de connaissances plus spécifiques, et de techniques de modélisation qui seront la base de conception de notre modèle d'aide à la décision.
- Dans la deuxième partie du document, nous allons présenter notre approche de conception du modèle d'aide à la décision ainsi que les outils de simulation nécessaire pour sa mise en œuvre. Pour opérationnaliser notre approche de modélisation, nous proposons d'avoir recours aux méthodes de l'intelligence artificielle qui permettent de modéliser les structures de connaissances et leur manipulation comme ils permettent d'intégrer toutes sortes de logiques de l'action afin de les intégrer dans un système s'adaptant à la nécessité et la logique du type de la situation de décision. Cette partie est composée de deux chapitres :

- Dans le deuxième chapitre nous proposons de modéliser le décideur individuel à l'aide des Séparateurs à Vaste Marge qui s'apparentent aux réseaux de neurones et sont issues de la même théorie, ou du moins du même domaine : l'apprentissage statistique. Ces derniers permettent d'utiliser la souplesse des méta-modèles type réseaux de neurones tout en conservant une forte capacité de généralisation.
- Le troisième chapitre est une tentative de minimiser l'effet de la rationalité limitée des décideurs. Face à un environnement incertain, la décision est affaire de bonne organisation collective. Elle devient l'émanation de savoirs multiples dans les domaines techniques, sociaux, culturels et organisationnels. Résultantes complexes d'action et de stratégies individuelles et de construits collectifs. Nous nous sommes intéressés, en particulier, à l'intelligence collective originelle dont l'objet est l'étude et l'optimisation des propriétés émergentes intérieures-subjectives et extérieures-objectives des collectifs. Afin de conceptualiser cette image, nous proposons d'utiliser un Système Multi-Agents cognitifs avec une architecture à base de tableau noir.
- La dernière partie est une application dont le but est de tester la performance ainsi que la robustesse de notre outil d'aide à la décision. Nous nous sommes intéressés au dépistage automatique de la maladie de la scoliose. Le choix de cette application se justifie par la nature radicale et incertaine de la décision thérapeutique. En effet, Face à son patient, le médecin est souvent dans l'ignorance totale malgré le poids de son savoir et sa capacité d'expertise. Cette partie est composée de deux chapitres :
 - Le quatrième chapitre présente la pathologie de la maladie de la scoliose ainsi que la nouvelle technologie de l'imageur biplan à détecteurs gazeux nommé « EOS imaging ». Cette technologie à permet une nouvelle perception du problème de la maladie de la scoliose en utilisant les reconstructions 3D du rachis humain, ce qui a rendu possible la réalisation de cette application.
 - Dans le cinquième et dernier chapitre nous allons présenter en détaille la méthodologie que nous avons adoptée en commençant par la présentation de la

base de données, le prétraitement, l'extraction de l'information pertinente, la simulation de la prise de décision individuelle ainsi que la simulation de la prise de décision collective et nous terminons ce chapitre par une discussion des résultats obtenus.

Partie 1 :
Analyse cognitive

CHAPITRE 1

PRINCIPAUX CONCEPTS COGNITIFS DU DÉCIDEUR ET DE LA DÉCISION

1.1 Introduction

Max Weber, au début du XX^{ème} siècle, était le premier à formuler les principes d'une science sociale interprétative et explicative. Il postule que : « *La décision est un comportement intentionnel et social. Ainsi, un acteur qui agit avec intention (contrairement à celui qui se contente de réagir à des stimuli sans réfléchir) doit travailler à l'aide d'un modèle du monde au sein duquel il active des objets. Il doit créer des représentations, un monde symbolique qu'il doit manipuler consciemment ou inconsciemment pour arriver à prendre des décisions qui affecteront le "monde réel". En conséquence, les données fondamentales sur lesquelles se base une décision individuelle sont des constructions symboliques. Ceci signifie que la partie fondamentale de l'analyse du décideur (et des décisions) est le sens perçu par ce dernier* » [Daniel, 1996].

Cependant, la décision n'existe pas sans le problème. Pour cela, l'analyse de la décision ne doit pas se limiter à celle du choix entre certaines alternatives d'action. La décision est surtout engendrée par la définition de *l'espace du problème* et son traitement par le décideur. En d'autres termes, les décisions sont faites au moyen de processus symboliques et reposent sur des représentations. Techniquement, il est utile de décrire un décideur comme un système organisé symboliquement.

Le sens, en tant que donnée pour la décision, est symboliquement pré-structuré, produit par le savoir pratique d'une raison non-théorique. C'est-à-dire, ce ne sont pas nos théories ou nos modèles qui attribuent une signification première aux actions. Ainsi, la description scientifique de ces données ne dépend pas seulement d'un fondement théorique solide mais est également déterminée par le savoir de l'acteur social sur le monde, ce savoir souvent appelé *la pensée subjective*.

Pire encore, la recherche empirique montre que beaucoup de décisions ne sont prises que semi-consciemment et que d'autres ne sont pas explicites. D'une manière caricaturale, nous pouvons dire que le décideur ne réfléchit pas nécessairement lorsqu'il pense. Ainsi, même ses propres motivations peuvent lui échapper lors d'une décision [Hull, 1952]. Un chercheur en sciences sociales ne peut parler d'intentions et de sens sans comprendre, d'une certaine façon, ce que le sujet analysé pense et ce qu'il fait.

Dans cette perspective de recherche orientée vers la compréhension, il faut d'abord essayer de reconstruire le processus de décision et ses interactions avec le monde, ce qui implique d'imaginer certaines structures de connaissance internes du décideur avant de commencer la modélisation proprement dite.

1.2 Cognition et décision

La décision est au cœur d'une des grandes énigmes de la pensée et de l'action. Elle semble partout présente dans l'histoire. L'examen de la réalité des décisions montre que décider ne correspond pas à une phase précise, clairement identifiable où tombe le couperet. La langue française emploie l'expression « *prendre une décision* » un peu comme si la décision était un objet identifiable. Le point d'arrivée est confondu avec le processus. La langue anglaise, qui emploie l'expression de « *decision making process* » (la fabrication de la décision) rend mieux compte de ce fait. La décision est un cheminement : elle se construit, se négocie, suit des voies sinueuses au cours du temps. Elle pourrait sembler insaisissable [Plantie, 2006].

Pourtant, pour beaucoup de chercheurs, une décision n'est qu'un choix, chez d'autres, un peu moins extrémistes, nous retrouvons la notion de processus déjà rencontrée dans le concept de la résolution de problèmes en science cognitive. Nous introduisons donc le temps et le changement en plus des choix. Généralement, nous pouvons attribuer à tout processus de prise de décision les éléments linéaires suivants [Sfez, 1992] :

1. Un processus de recherche pour découvrir des buts ;

2. La formulation d'objectifs exacts;
3. La sélection d'alternatives stratégiques pour accomplir ces objectifs ;
4. L'évaluation des résultats.

Cependant, il est clair que cette liste d'éléments ne reflète que partiellement la réalité. Le processus de la prise de décision ne conduit pas linéairement de la sélection des buts à la conclusion, les buts d'une décision ne seront souvent découverts que durant le processus même de la mise en œuvre de la décision. Des schémas de ce type, même les plus élaborés, reflètent une vision biaisée par le plan normatif. En effet, les recherches empiriques sur les prises de décisions montrent qu'aucune de ces étapes ne peut être effectuée d'une façon optimale et à un moment précis de la décision. Toutefois, nous devons attribuer une certaine valeur heuristique à cette démarche simple, car il est vrai qu'un acteur (collectif ou individuel) est concerné par ces quatre éléments de décision. Dans certains cas, un tel schéma linéaire peut suffire à décrire une décision.

Les approches cognitives, quant à elles, traitent la décision comme le résultat d'un processus global de résolution de problèmes qui est souvent long et de nature presque continue. Elle se fait dans un réseau logique d'anticipation, de rétroaction et de correction. Il existe un processus de structuration progressive du champ du possible vers une fermeture des possibilités. En effet, la modélisation qui désire savoir comment un décideur (individuel ou collectif) prend une décision doit s'intéresser à la façon dont le décideur modélise le monde et au savoir-faire subjectif qui permet de traiter cette information.

En se basant sur ce point de vue, ces approches définissent la prise de décision comme étant un processus complexe dont les données sont instables et dépendent également de l'interaction du décideur avec son environnement [Meadows et Robinson, 1985]. La notion de décideur isolé est inutilisable dans la plupart des cas. Ainsi, un modèle de décideur devrait contenir également un modèle de l'environnement, même si ce dernier n'est pas sophistiqué.

Ainsi, les étapes de la décision, qui ont une signification plutôt logique que temporelle, et qui sont les plus souvent citées par les auteurs des approches cognitives sont :

1. la perception ;
2. la préparation ;
3. la décision ;
4. l'exécution.

Les opinions sur l'importance de ces étapes varient énormément. Cette variété de vues et de paradigmes s'explique en partie par le type de décisions auquel ils s'intéressent. Les éléments ontologiques à l'aide desquels nous pourrions définir des types de décision illustrent la richesse de ce concept. Ces derniers sont le(a) :

1. Objet de la décision: but, programme, opération, instrument,...
2. Organe de décision: organisation, groupe, individu,...
3. Type de prise de décision: routine, créatif, application d'un programme,...
4. Portée de la décision: stratégique, tactique, opérationnelle,...
5. Contrôle des éléments de la décision: bonne, moyenne,...

La différence structurelle entre certains types de décision peut être très grande. Cela explique en partie la divergence des paradigmes en ce qui concerne la définition de ce qu'est une décision et de ce qu'un décideur sait faire. Un grand nombre de ces théories de la décision ne concerne pas le phénomène global de la décision mais peut être utile pour un type de décision bien précis.

Aujourd'hui, il est devenu clair qu'un décideur moderne opère dans un environnement de problèmes très complexe et mal structuré où il doit intégrer beaucoup de facteurs intra- et extra-organisationnelles ainsi que toutes les logiques qui leur sont attachées. La perception

croissante de la complexité décisionnelle a focalisé la recherche sur la définition du problème, des processus de résolution de problème et de la mise en œuvre [Daniel, 1996]. Le choix lui-même est maintenant envisagé comme élément répétitif dans un processus de traitement d'information mettant en interaction des individus avec leur organisation.

1.2.1 Processus de décision et le temps

Le chercheur qui s'intéresse à la décision possède des problèmes communs à ceux de l'historien. Les analyses se font en règle générale ex post facto. Les processus qui nous intéressent sont rationalisés par le biais d'une interprétation. Malheureusement, le temps dans l'histoire est avant tout chronologique, les événements ont l'air unique. Ce constat est en contradiction avec le but des sciences cognitives dont le but est de trouver des généralisations, des lois, voire même un système.

En ce qui concerne la décision, nous désirons connaître les propriétés du problème, du décideur et de son environnement et trouver une logique qui conduit à une décision. Le temps dans un modèle de décision doit couler du stade de la possibilité vers le stade de l'accomplissement [Delley et al., 1982], sinon, l'observateur procède à une énumération inintéressante de faits. En d'autres mots, l'ordre intéressant des phases dans un processus de décision est plutôt logique que chronologique.

De nombreuses études ont mis en valeur la relativité des étapes de la décision. Ainsi, il arrive très rarement qu'un problème de décision soit bien défini au moment de la décision. Au contraire, décider veut souvent dire définir le problème. Chaque étape de la décision peut être redéfinie par l'étape suivante. Dans l'autre sens, un décideur peut anticiper des étapes de la décision sans avoir résolu les étapes précédentes. La notion de la décision est diluée dans le temps [Delley et al., 1982].

Finalement, la décision peut modifier le décideur lui-même et par là changer tout le processus de décision. Aussi, au lieu de parler des étapes ou des stades de décision, il faudrait voir la décision comme un processus qui met en place une sorte de système ayant une durée

de vie limitée qui a comme but de comprendre et de définir un problème, de trouver ou de créer des ressources qui fournissent des solutions et qui les mettent en œuvre.

1.2.2 Étapes de la décision

La décision est un processus de résolution de problème qui met en œuvre des connaissances de nature très variée. Dans cette section, nous discuterons brièvement de l'anatomie idéalisée d'un processus de décision globale isolée.

Pour mieux discuter cette anatomie, nous allons d'abord donner une définition de la décision simple. Elle correspond à la classe de problème relativement bien défini que nous retrouvons dans la littérature de la science cognitive [Doerner, 1987].

Un problème existe si:

1. Le décideur perçoit un état interne ou externe non désiré **A**.
2. L'état de départ **A** non désiré doit être transformé dans un état de but **B**.
3. Le décideur ne sait pas au départ, comment parcourir le chemin qui va du point de départ **A** vers le point de solution **B**.

Prendre ce type de décision simple équivaut en termes de psychologie cognitive à résoudre un problème bien structuré en appliquant des opérateurs cognitifs à un espace de problème. Etant donné, que même les décisions simples ne possèdent ni problème ni but bien défini, il convient d'élargir cette définition de problème en une recette d'action (décision complexe) [Daniel, 1996] :

1. Le décideur doit transformer la perception, peu précise, du problème en un problème plus précis et mieux structuré.
2. Les buts de la décision doivent être précisés.

3. Le problème doit être transformé en sous-problèmes abordables.
4. Chaque sous-problème peut être résumé par un but à atteindre. Pour chaque sous-but il s'agit de trouver des heuristiques (opérateurs) qui permettent de le résoudre.
5. La poursuite des sous-buts doit être coordonnée stratégiquement. Il s'agit aussi de résoudre des conflits entre sous-buts.
6. Une fois que tous les sous-buts ont été atteints, il faut évaluer si le but général a été atteint.

Ces points nous rapprochent du processus de décision le plus fréquent, la différence est que le processus de résolution d'un problème n'est que très rarement une affaire aussi bien structurée. Comme nous l'avons déjà fait remarquer, ces étapes de la décision forment une suite suggestive plutôt qu'une suite causale et temporelle. Il faut donc interpréter ces étapes de la décision comme un ensemble de processus cognitifs impliqués dans une décision. Elles peuvent exister en parallèle et sont en interaction. Seulement, leur importance relative varie beaucoup pendant le déroulement de la décision. Nous proposons donc d'étudier la décision plutôt comme une fonction de connaissances intervenues que comme le déroulement séquentiel de N étapes de décision. Examinons de plus près quelques-uns de ces processus qui impliquent des connaissances de résolution de problème :

1.2.2.1 Définition du problème

Au premier stade, le décideur perçoit un problème qui doit être exprimé dans des termes assez précis. Le décideur rassemble autant d'informations que nécessaires pour prendre sa décision. Ce qui est nécessaire est défini par plusieurs facteurs, entre autres par les exigences du problème et par les ressources que les agents peuvent investir. Une partie importante de cette information est fournie par l'environnement.

A l'aide de média variés, le décideur se renseigne sur le problème à résoudre. Très souvent, le décideur identifie au même moment les buts à atteindre à la fin du processus. Une autre source importante pour la définition du problème est la mémoire sémantique et

épistémique du décideur. Tout ce qu'il perçoit l'est en fonction de connaissances déjà existantes [Daniel, 1996]. Ainsi nous comprenons déjà pourquoi le stade suivant (définition de buts) est très lié à celui de la définition du problème.

1.2.2.2 Définition et utilisations de buts

Les buts ont des fonctions multiples à la fois au niveau du comportement général et au niveau instrumental (planification des actions). Par rapport aux activités du décideur, nous pouvons distinguer plusieurs classes de buts.

Il existe des buts de très longue durée, voire permanents. Ceux-ci sont souvent liés au bien-être du décideur et n'ont pas forcément un rapport très direct avec le problème qu'il doit résoudre. Toutefois, leur rôle est crucial. Quand, par exemple, un décideur doit prendre une décision délicate, il pense aussi aux implications qu'a sa décision sur lui et son organisation. Autrement dit, le but de préservation est activé. Un deuxième ensemble important de buts à long terme est lié à la fonction du décideur. Plus sa position est élevée, plus il a des buts généraux de maintien ou de poursuite de quelque chose qui concernent son environnement. Cognitivement parlant, ces buts restent souvent dans un état semi-actif. Ils sont scrutés périodiquement ou alors activés par un stimulus venant de l'environnement. Lorsqu'un problème apparaît, plusieurs de ces buts peuvent s'activer. Ils sont souvent liés entre eux, parfois de façon conflictuelle [Plantie, 2006].

Ces deux classes de buts (maintien du décideur, buts de l'environnement liés à sa fonction) peuvent être exceptionnellement créés lors d'un processus de décision, mais le plus souvent ils sont seulement activés, identifiés, poursuivis ou légèrement modifiés. Nous les appelons buts universaux ou centraux. Ces buts sont organisés dans un réseau relativement stable ayant des relations hiérarchiques (but et sous-but) et des relations de priorité (importance relative). Sans être logiquement cohérent, ce réseau joue un rôle important dans toute activité du décideur.

Les buts qui concernent plus spécifiquement la finalité d'une décision seront appelés buts généraux de la décision ou simplement objectifs de la décision. Dans les théories de choix traditionnels, on présuppose que ces buts sont connus par le décideur et qu'ils sont bien définis. Cependant, il s'est avéré que ce n'est pas le cas en règle générale. La définition de ces buts généraux de la décision n'est obtenue qu'après un long processus. Souvent, ils restent mal définis, voire contradictoires. Ces objectifs de décision sont toujours en interaction avec les buts universaux et parfois sont engendrés par ces derniers [Plantie, 2006].

1.2.2.3 Résolution du problème

Résoudre un problème identifié signifie plusieurs choses. Au problème à résoudre, nous associons au moins un but à atteindre. Généralement, atteindre ce but implique souvent atteindre plusieurs sous-buts en parallèle. Ces sous-buts peuvent être décomposés à leur tour. Ce processus est appelé factorisation car chaque sous-but reflète un problème partiel à résoudre.

Grâce à des connaissances stratégiques, le décideur essaie de diviser un problème en des unités plus réduites qui respectent un ordre hiérarchique, temporel et causal. Une fois cette factorisation accomplie, il s'agit de coordonner et de contrôler la poursuite des sous-buts et de régler d'éventuels conflits. Cette décomposition du problème en sous-buts (ou sous-problèmes selon un autre point de vue) ne représente que le niveau supérieur des opérations. La réalisation des sous-buts met en œuvre des connaissances multiples, telles que l'application d'instruments de pilotage ou l'allocation de ressources. Au plan cognitif, il s'agit d'élaborer, ou de retrouver et d'adapter, des plans que nous pouvons exécuter [Daniel, 1996].

1.2.2.4 Évaluation

Vers la fin du processus de décision, le décideur évalue l'effet des mesures (décisions) prises par rapport au but à atteindre. Il teste si les actions ont eu l'effet voulu. En fonction des résultats, certains processus de résolution de problème doivent être répétés et modifiés,

jusqu'à ce que le sous-but concerné ait été atteint de façon satisfaisante. Cette évaluation se fait souvent d'une façon informelle [Daniel, 1996].

1.2.2.5 Processus complet de décision

L'organisation globale d'un processus de décision est le résultat de l'interaction de plusieurs couches de processus. Il existe une ou plusieurs stratégies générales de la décision ou de résolution de problème. Ces stratégies ne sont pas des programmes fixes, mais des heuristiques qui organisent et qui contrôlent le processus. Certaines connaissances s'activent quand c'est nécessaire, mais d'autres sont mobilisées de façon « *top-down* » ou « *goal-driven* ». Ainsi, il ne faut pas confondre la description d'un processus de décision avec le savoir qui est impliqué dans sa génération [Daniel, 1996].

1.3 Analyse cognitive du décideur

Il existe une littérature importante, études cliniques à l'appui, qui démontre les limites du décideur. Cette littérature se concentre très souvent sur des problèmes de choix dans différents types de situation. Ces travaux comparent la stratégie adoptée par le sujet à une stratégie optimale basée sur la maximisation des résultats attendus. De façon générale, nous constatons que les décideurs ne maîtrisent pas les stratégies basées sur les valeurs, mais qu'ils utilisent toutes autres sortes de raisonnements. Voici quelques exemples [Daniel, 1996] :

- La perception biaisée du cadre globale : Les gens décident en fonction de la façon dont on leur présente un problème.
- La violation de simples règles de calcul : Les règles de transitivité (répercussion des préférences relatives sur l'ensemble) et de dominance (choix de la meilleure solution) sont aussi fréquemment ignorées.
- Le choix sans raisonnement : Souvent nous constatons que les gens prennent des décisions sans savoir pourquoi. Si nous interrogeons ces sujets sur les raisons de leur décision, ils inventent des raisons fausses.

- Les jugements affectifs : Les émotions ont un rôle très important. Cela se manifeste lors de l'évaluation d'alternatives affectivement chargées ou encore lors des choix dans des situations où le sujet se trouve sous des émotions fortes.

En règle générale, nous pouvons observer que la décision repose sur un raisonnement qui prend des formes très variées. Cependant, jusqu'à présent il existe un clivage très important entre les recherches sur le raisonnement et les recherches sur la prise de décision. Les premières sont plutôt descriptives tandis que les deuxièmes sont plutôt critiques ou normatives. En tout cas, aucune étude n'a montré la pertinence empirique des théories de choix rationnel.

Le décideur humain est certainement totalement incapable de générer des arbres de décision longs ou complexes, entre autre, à cause de ses capacités très limitées de calcul. Par contre, il est en mesure de faire des choix relativement rationnels pour des problèmes à dimension humaine. En somme, les décisions ne sont pas prises après avoir posé le problème et collecté toutes les informations, mais progressivement durant un long processus d'action et de planification. Le choix a une place limitée dans la décision et il faut évaluer un décideur en fonction de sa capacité à poursuivre (et réussir) un ensemble de buts à long terme.

Dans le passé, les recherches se sont centrées sur l'analyse des attitudes des décideurs pour prédire leur comportement [Doerner, 1987]. L'aspect cognitif du décideur était donc traité très négligemment dans ces modèles. Ces analyses par les attitudes ont une valeur très limitée pour l'analyse du décideur. Il s'est en effet avéré que la cohérence des attitudes à travers les situations est assez faible. La décision s'explique souvent mieux par les contraintes dans lesquelles elle est réalisée. Les attributs individuels (ainsi que d'autres variables psychologiques à structure statique) ne sont pas seulement instables à travers le temps et les situations, mais ils n'ont que très peu d'importance par rapport aux variables de l'environnement.

La plupart des auteurs intéressés par la modélisation de ce que nous appelons les *variables intervenantes* sont d'accord pour affirmer que les propensions de diagnostic et de décision

d'un individu sont en premier lieu le résultat d'ordres provenant de l'extérieur [Daniel, 1996]. Toutefois, les caractéristiques de l'individu ont aussi leur place, par exemple ses croyances et sa capacité à traiter une information complexe. Les approches cognitives se sont basées particulièrement sur ce postulat. Ces dernières affirment que la clé pour la modélisation du décideur est donc de s'intéresser plus particulièrement à la façon dont il représente les choses, autrement dit comment il modélise le monde. La question des mécanismes de décision proprement dite garde leur importance, bien sûr, mais ce n'est plus la question fondamentale. Ces observations faites en sciences cognitives coïncident d'une façon extraordinaire avec les travaux en intelligence artificielle qui ont découvert que c'est la représentation du savoir qui est la tâche la plus critique lors de la simulation [Plantie, 2006].

1.3.1 Propriétés du décideur

Le décideur humain est un organisme ancré dans un environnement: c'est un acteur social. Il doit être décrit et traité comme un système ouvert qui, activement, saisit de l'information, qui la traite et qui agit. Il possède une certaine indépendance cognitive: il est *équifinal*, c'est-à-dire, il peut atteindre le même but avec des moyens différents, et il est auto-organisateur. Décider veut dire agir. Une action, dans le cas idéal, possède une finalité, elle est consciente, planifiée et voulue [Clancy, 1992]. Un tel idéal-type doit être distingué du pur comportement qui n'est que réaction. L'action du décideur se trouve quelque part entre ces deux extrêmes, mais plus près de l'action idéale.

En faisant références à quelques travaux récents en sciences cognitives, le décideur est traité comme étant un système, c'est-à-dire : « *un objet, qui, dans un environnement, doté de finalités, exerce une activité et voit sa structure interne évoluer au fil du temps, sans qu'il perde pourtant son identité unique* » [Morin, 1990].

Cette définition nous permet de dire qu'un système c'est :

- quelque chose (n'importe quoi, présumé identifiable) ;
- qui dans quelque chose (environnement) ;

- pour quelque chose (finalité ou projet) ;
- fait quelque chose (activité = fonctionnement) ;
- par quelque chose (structure) ;
- et qui se transforme dans le temps (évolution).

La constitution d'un système ne peut pas se faire par simple énumération de ses éléments visibles, mais par une sorte de triangulation qui pondère une définition fonctionnelle (ce que l'objet fait), une définition ontologique (ce que l'objet est) et une définition génétique (ce que l'objet devient). Cette trilogie de l'Être, du Faire et du Devenir nous permettra seulement de construire les éléments et les interactions du système que nous désirons modéliser. Le modèle est ainsi un langage pour parler d'un système.

1.3.2 Le décideur, une vue systémique

Une façon simple de définir un système est de lui attribuer une fonction de décision par rapport à un environnement. Un environnement peut être conceptualisé de différentes façons. Certains ne s'intéressent à l'environnement que dans la mesure où il permet de fixer quelques paramètres externes du système. Pour d'autres auteurs, l'environnement devient un facteur clé pour définir un système, c'est-à-dire que tout ce qui est à l'intérieur d'un sous-ensemble de l'environnement et qui peut faire des transactions devient le système. La décision devient alors à la fois une fonction de « *l'input* » de l'environnement et une sortie venant d'une boîte qui opère quelques transformations sur cet « *input* » [Daniel, 1996].

Les positions divergent quant à la façon d'étudier cette boîte. La méthode empirique la plus sûre est de la traiter comme une boîte noire, c'est-à-dire d'inférer la nature de la boîte de décision en observant son comportement externe. Le désavantage de ces boîtes de décisions est leur potentiel décisionnel qui va au-delà de ce que l'on peut observer dans un laps de temps. Les observations n'auront pas de grande valeur sans théorie sur le décideur. La même remarque est vraie pour l'environnement du décideur. Dès que son niveau d'équilibre change, les données dont il doit tenir compte changent aussi.

Mais comment développer une théorie du décideur qui veut aller au-delà des banalités de la théorie du choix qui se désintéresse de l'ensemble de la décision? Les décisions ne peuvent pas être prises sans processus de décision. Et le processus de décision repose sur un traitement de l'information par un agent intelligent.

Parmi les premières théories du système, la cybernétique de Wiener [Wiener, 1961] était une des approches les plus intéressantes. Wiener a fourni à la science de nouvelles métaphores. Ces nouvelles métaphores sont la communication, le traitement de l'information, le contrôle, le « *feed-back* » et l'homéostasie. Ces termes modernes comme celui du « *feed-back* » n'étaient qu'une nouvelle façon d'analyser un comportement téléologique. Nous retrouvons même la notion de but qui n'était pas explicitement traité dans l'original de Wiener, toutefois à partir de la notion du maintien de l'équilibre, nous avons nécessairement introduit les buts d'équilibre et la survie et d'autres buts principaux que le système doit poursuivre pour se maintenir dans son environnement.

L'apport important de la cybernétique a été les notions de l'action continue sur l'environnement et de l'auto-adaptation du système. Cet apport a une influence directe sur la notion de but qui se réfère à la fois à la perception subjective du système et à un état objectif de l'environnement. Autrement dit, la théorie du système voit le décideur comme un « *goal-seeking feedback engine* » et comme un « *learning engine* » qui sont liées inextricablement à un environnement [Daniel, 1996].

En effet, la décision est un processus souvent long et de nature presque continue. Et elle se fait dans un réseau logique d'anticipation, de rétroaction et de correction. Une raison pour ne pas assimiler choix et décision réside dans la complexité [Daniel, 1996]. En terme systémiques, la complexité est la relation entre le système de décision et le monde. De manière plus formelle, il s'agit de la somme des événements possibles par rapport à la structuration du système.

La complexité est un problème de contingence d'une trop grande liberté de choix. Ainsi les systèmes de décision (organisations ou individus) inventent un grand éventail de

mécanismes pour ramener la contingence des possibilités d'actions à un seuil acceptable. Il existe aussi le mécanisme de double contingence, où un système sélectionne ses états en fonction des états possibles des autres systèmes. Cette contingence réciproque est le mécanisme qui assure la coordination et la convention humaines. Autrement dit, un acteur réduit les possibilités d'action car il tient compte des autres dans son environnement.

De plus, afin de réduire la complexité innée du monde, les décideurs possèdent des structures cognitives spécialisées pour représenter leurs connaissances sur leur environnement de décision. Ces derniers possèdent des procédures pour créer et transformer ces structures. Au niveau de l'organisation, il existe le même phénomène sous la forme de langages spécialisés et de procédures qui permettent de réduire la variance à des niveaux gérables. En d'autres termes, un système de décision tend à transformer l'entropie riche (complexité peu structurée) de l'environnement en entropie de moins en moins riche jusqu'à ce qu'il puisse interrompre ce processus par une décision [Sfez, 1992]. C'est-à-dire, il existe un processus de structuration progressive du champ du possible vers une fermeture des possibilités. Ce processus est représenté par la figure suivante.

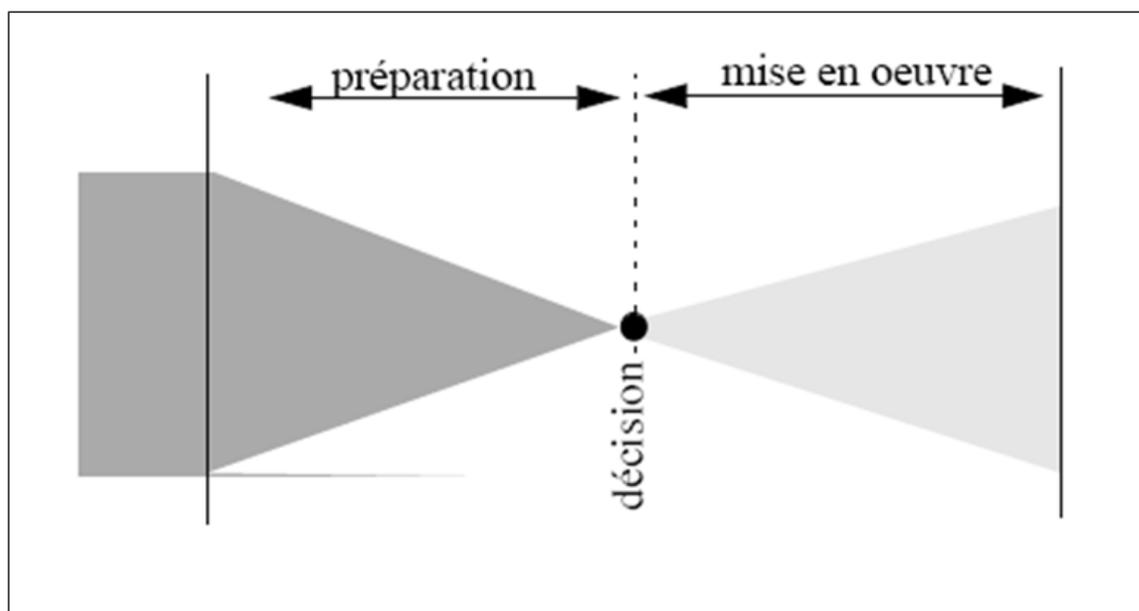


Figure 1.1 : Mécanisme de réduction de l'entropie [Daniel, 1996].

Ce mécanisme de structuration implique un code, une structure symbolique qui permet de traduire un état perçu dans un langage X vers un langage Y ou encore d'abstraire, de combiner et de nommer des phénomènes. Traduction, combinaison et abstraction sont les maîtres mots des transformations que subit une information au cours de la décision [Sfez, 1992]. La décision est donc toujours le résultat d'un ensemble de décisions subordonnées, d'un processus de délibération. Un modèle général de la décision ne peut pas éviter de distinguer entre différentes classes de décisions et décideurs. Les avantages intellectuels de la classification exigent une identification des étapes du processus de la décision, même si leur position chronologique est souvent douteuse. Il convient aussi de noter que beaucoup de décisions à l'intérieur d'un processus de décision étudié n'ont pas lieu pendant les phases de décision au sens simple du terme.

Un dernier facteur important en rapport avec la gestion de la complexité est lié à la capacité d'apprentissage d'un système de décision. L'histoire permet au décideur à un niveau plus immédiat un raisonnement à bases de cas précédents, et, à un niveau plus profond, la création de schémas d'actions [Daniel, 1996]. Il est vrai que la complexité d'un processus de décision peut être mesurée par le nombre d'étapes qu'il faut exécuter pour arriver à une solution. Toutefois, la complexité ne mesure pas la facilité avec laquelle un système peut trouver une solution à un problème.

1.3.3 Le décideur dans un environnement complexe

Une simulation de situations réelles de décision qui se veut crédible, doit tenir compte de la complexité de l'environnement. En effet, un problème est rarement résolu grâce à une décision unique. Il s'agit au contraire de piloter en permanence des variables critiques du système global et d'observer l'effet que des décisions (et actions qui en découlent) ont eu sur cet environnement.

Dans un tel environnement, il est souvent difficile de trouver un point de départ et une fin pour un processus de décision. Il faut procéder à des coupures subjectives, c'est-à-dire des périodes où le décideur se désintéresse d'un problème parce qu'il semble être résolu ou

autrement inintéressant [Daniel, 1996]. Généralement, une seule décision importante peut prendre entre quelques heures et plusieurs années. Dans le deuxième cas, il s'agit en règle générale de mettre en place tout un dispositif pour poursuivre un but. Ce genre d'activité se déroule en parallèle avec les activités courantes et un certain nombre de ressources du décideur y sont consacrées.

De plus, il faut noter que dans un environnement complexe nous remarquerons souvent que :

- Seuls des problèmes fermés (comme un problème arithmétique) peuvent être définis en termes d'état indésirable à transformer en état de but à l'aide d'un certain nombre d'opérateurs connus. Cependant, dans un environnement complexe la majorité des problèmes sont mal définis. L'état actuel du problème et le but à atteindre sont souvent mal connus et doivent être élaborés dans un dur labeur.
- Un décideur doit s'occuper de plusieurs problèmes à la fois. Ces problèmes ont une interdépendance qui peut être forte, par exemple, ils peuvent affecter des buts communs. Aussi, même si un problème est attaqué d'une façon isolée, il n'est jamais perçu totalement en dehors du contexte. Sa perception, la définition de buts, les moyens d'actions envisagés etc. sont affectés par cet effet "réseau".
- Dans de nombreux problèmes, les solutions ne sont pas trouvées par des moyens clairs et transparents, mais par des processus de raisonnement (individuels ou institutionnels) associatifs comme l'analogie.

1.4 Conclusion

L'idée-force de ce chapitre qui a déjà été bien souvent exposée dans ces pages est : le décideur est une machine de traitement de l'information qui interagit avec un environnement actif. Nous postulons en plus que le décideur peut être modélisé avantageusement comme processeur symbolique.

La qualité d'un modèle peut se mesurer dans deux dimensions: celle de l'homomorphie avec un modèle d'une validité théorique large et celle de l'utilité et de la maniabilité pour répondre à une question de recherche concrète. En sciences sociales, l'écart entre le modèle utilisé dans une recherche empirique et le modèle général de l'acteur humain est souvent important. Il est à la fois trop difficile de trouver les données nécessaires à une telle entreprise et il semble presque impossible de trouver un langage de modélisation qui permet la représentation et la maîtrise de la complexité de l'action humaine. Il est important de clairement identifier les éléments théoriques que nous perdons en descendant vers l'empirique. Seulement en adoptant cette stratégie d'ancrage global, nous serons en mesure de faire progresser de façon incrémentale ce type de modélisation qui veut réunir à la fois l'image humaniste de l'homme et celle plus formelle de la théorie des systèmes et des modèles en intelligence artificielle.

De plus, dans le domaine de l'intelligence artificielle, il est souvent probable de rencontrer des problèmes qui se manifestent dans d'autres disciplines systémiques, notamment, le problème de la reproduction de processus complexes et de leur gestion dans un modèle accessible et manipulable. L'accès aux données (la pensée, la communication ainsi que son support externe) reste incomplet. Les catégories centrales de l'action humaine comme celle de la pensée sont difficilement accessibles par des moyens scientifiques.

Toutefois, si nous postulons qu'un acteur réfléchit, il faut construire des modèles qui produisent des processus ayant un sens à ce niveau de description. En effet, il est important de viser une homomorphie réduite, adaptée à la fois aux exigences de l'économie de la recherche et à la nécessité de modéliser un acteur et non pas un système de paramètres dénué de tout sens. La frontière entre un modèle qui représente vraiment quelque chose et un modèle qui a une pure valeur prédictive n'est pas facile à identifier. Nous pouvons même dire que c'est le « *no man's land* » entre ces deux extrêmes qui sera notre terrain de recherche.

Partie 2 :

***Conception de l'outil
d'aide à la décision***

CHAPITRE 2

MODÉLISATION D'UN DÉCIDEUR INDIVIDUEL PAR LES SÉPARATEURS À VASTE MARGE

2.1 Introduction

L'intelligence artificielle est devenue un des terrains de bataille favoris de la philosophie de l'esprit. Comme peu de disciplines des sciences humaines, elle est explicite et revendicatrice. L'intelligence artificielle n'a pas d'objet de recherche académique bien défini à part l'intérêt porté au mental et aux représentations de connaissances. Elle s'est divisée en de nombreuses sous-disciplines focalisant sur des problèmes bien distincts tels que la vision, la résolution de problèmes, la compréhension du langage, l'apprentissage,... Cependant, il n'existe pas de paradigme unifié de recherche et certaines branches de l'intelligence artificielle sont devenues des terrains d'échanges multidisciplinaires où se côtoient philosophes, psychologues, informaticiens et autres qui s'intéressent aux diverses problématiques de l'intelligence [Daniel, 1996].

2.2 Prise de décision et intelligence artificielle

Comme nous l'avons précisé auparavant, la prise de décision est un processus global de résolution de problèmes souvent long et de nature presque continue. Elle se fait dans un réseau logique d'anticipation, de rétroaction et de correction. Une raison pour ne pas assimiler choix et décision réside dans la complexité. En effet, la complexité est un problème de contingence d'une trop grande liberté de choix. Ainsi les systèmes de décision (organisations ou individus) utilisent un grand éventail de mécanismes pour ramener la contingence des possibilités d'actions à un seuil acceptable.

Afin de réduire la complexité innée du monde, les décideurs possèdent des structures cognitives spécialisées pour représenter leurs connaissances sur leur environnement de décision. Ces derniers possèdent des procédures pour créer et transformer ces structures. En

d'autres termes, un système de décision tend à transformer l'entropie riche (complexité peu structurée) de l'environnement en entropie de moins en moins riche jusqu'à ce qu'il puisse interrompre ce processus par une décision. C'est-à-dire, d'associer le problème perçu à la classe de solutions ou, dans le cas idéal, à la solution correspondante.

En effet, nous pouvons dire que la prise de décision est essentiellement un problème de classification qui associe le problème d'entrée à une classe de solutions plus ou moins appropriées. Sauf que cette classification, pour ne pas tomber dans le piège du choix simple et directe, se fait sur la base de structures cognitives spécialisées peu explorées qui, après une série de transformations, restructure le problème en le rendant plus accessible ce qui permet une meilleure classification et par conséquent un meilleur choix de la (les) solution(s).

Par conséquent, pour modéliser un décideur individuel, nous allons utiliser les outils de l'intelligence artificielle qui se basent sur le concept de l'apprentissage statistique qui permet d'une part d'approcher le fonctionnement des structures cognitives. Et d'autre part, entre ces outils, nous allons choisir ceux qui permettent de réaliser des classifications avec une grande capacité de généralisation.

2.3 Intelligence artificielle et classification

En passant en revue la littérature de l'intelligence artificielle, il s'avère qu'il existe deux outils, à savoir le Perceptron Multicouche (PMC) et les Séparateurs à Vaste Marge (SVM), qui sont les mieux adaptés aux problèmes de classification avec une grande capacité de généralisation et qui ont pour argument leurs fondements mathématiques et des résultats pratiques très convaincant [Deheeger, 2008; Hurtado et Alvarez, 2001; Plantie, 2006; Taleb-Ahmed et Djadel, 2002].

Bien que le PMC et les SVM permettent de résoudre les mêmes types de problèmes, il existe néanmoins quelques différences entre ces derniers. Dans ce qui suit, nous allons essayer de présenter leurs logiques de fonctionnement, leurs avantages et inconvénients ainsi

que quelques études comparatives qui vont guider notre choix du réseau le mieux adapté à la modélisation d'un décideur individuel.

2.3.1 Entre PMC et SVM

2.3.1.1 Perceptron multicouche

Le perceptron multicouche (*cf. Annexe I*) fait partie de la famille des réseaux de neurones artificiels, qui sont à l'origine, une tentative de modélisation mathématique du cerveau humain [Ladjadj, 2003]. L'idée principale de ces derniers consiste à modéliser une grandeur mesurable (physique, chimique, économique, ...), dont on soupçonne qu'elle dépend d'un ou plusieurs facteurs qui constituent les composantes du vecteur d'entrée. L'objectif de la modélisation par les réseaux de neurones est de trouver, à partir des mesures disponibles, une relation, si elle existe, entre les variables d'entrée et la grandeur à modéliser.

Une propriété fondamentale qui distingue le perceptron multicouche est l'approximation universelle parcimonieuse. Cette expression traduit deux propriétés distinctes : d'une part les PMC sont des approximateurs universels, et d'autre part, une approximation à l'aide de PMC nécessite, en général, moins de paramètres ajustables [Dreyfus, 2004].

Les PMC appartient à la famille générale des réseaux non bouclé ou encore à propagation vers l'avant, c'est-à-dire, l'information se propage dans un sens unique, des entrées vers les sorties sans aucune rétroaction. Il est constitué d'unités élémentaires appelées neurones formels qui sont des intégrateurs qui effectuent la somme pondérée de leurs entrées. Le résultat de cette somme est ensuite transformé par une fonction de transfert de type sigmoïde qui produit la sortie du neurone [Chatelain, 2003a; , 2003b]. Les neurones constituent, par leurs tours, les couches du PMC à savoir la couche d'entrée, la ou les couches intermédiaires et la couche de sortie où seulement les couches voisines sont connectées par des connexions pondérées.

A. Avantages

- La propriété fondamentale des PMC, l'approximation universelle parcimonieuse, fait de ceux-ci une représentation mathématique très avantageuse pour la modélisation statique et dynamique non linéaire de processus [Oussar, 1998].
- Les PMC présentent l'avantage de pouvoir être utilisées dans de nombreuses applications sans avoir à faire d'hypothèses suivant le problème et permettent d'obtenir des frontières entre les classes aussi complexes que nécessaires. Les perceptrons multicouches peuvent ainsi approximer n'importe quelles fonctions mathématiques : ce sont des approximateurs universels [Huet, 2004].
- Un des avantages incontestables du perceptron est la possibilité de moduler la puissance du réseau en modifiant le nombre de neurones dans la couche cachée. Plus nous ajoutons de neurones, plus le modèle pourra apprendre des formes complexes. Mais ce n'est pas anodin, en augmentant le nombre de neurones, nous prenons également le risque de trop coller aux données de l'ensemble d'apprentissage [Leray, 2001].
- Une fois la phase d'apprentissage achevée, les PMC deviennent rapides lors de la décision, résultat d'une exécution de fonctions non linéaires de combinaisons linéaires. Cela autorise l'utilisation de vecteurs de caractéristiques assez importants comme il permet, dans quelques applications, la conception d'un réseau opérant en temps réel [Zaier, 2008].

B. Inconvénients

- Pour son apprentissage, le PMC utilise l'algorithme de la rétropropagation (*cf. Annexe I*) basé sur la règle de correction des erreurs. Cette technique est déterminante dans la recherche d'une fonction de substitution car elle va conditionner l'obtention des différents paramètres, essentiellement les poids des connexions. Il s'agit en fait d'un problème d'optimisation : trouver les poids qui permettent de minimiser une fonction coût. Cette fonction permet le contrôle de la précision de l'approximation obtenue.

Cependant, la règle de la rétropropagation des erreurs présente plusieurs inconvénients majeurs :

- La rétropropagation des erreurs repose sur une optimisation non-linéaire et peut aboutir à des minima locaux, Cela est caractéristique des méthodes à gradient [Deheeger, 2008].
- Cette technique est capable de déterminer des frontières de décision complexe. C'est aussi là un défaut : une frontière de décision complexe aura tendance à bien classer les exemples de la base d'apprentissage, au détriment de la capacité à généraliser [Ladjadj, 2003].
- Un autre problème est celui du nombre d'itérations lors de la phase d'apprentissage. Celui-ci est primordial puisque il apparait au bout d'un certain nombre d'itération le phénomène bien connu du sur-apprentissage ou « *over-fitting* », durant lequel le MLP commence à apprendre par cœur les exemples de la base d'apprentissage et perd sa capacité à généraliser. En utilisant une base de validation, nous pouvons calculer l'erreur de généralisation du réseau en fonction du nombre d'itération. Nous choisirons naturellement le nombre d'itération qui minimise l'erreur de généralisation. Ainsi une autre limitation du PMC est qu'il nécessite des bases d'apprentissages conséquentes [Chatelain, 2003a; , 2003b].
- L'initialisation des poids des connexions est également un problème délicat. En effet, l'initialisation se fait de manière aléatoire, c'est-à-dire, que nous regardons le comportement du réseau sur la base d'apprentissage pour chaque initialisation, et nous choisissons la meilleure configuration [Chatelain, 2003a; , 2003b].
- Pour un problème de classification, le PMC se contente de trouver un hyperplan classant bien les échantillons de la base d'apprentissage sans vérifier qu'il soit le plan optimal qui sépare le mieux les différentes classes ce qui influe négativement sur sa capacité de généralisation [Chatelain, 2003a; , 2003b].
- La technique de la rétropropagation des erreurs converge seulement pour des problèmes de classifications où les données sont séparables, comme elle perd beaucoup de son efficacité dès qu'il y a trop de descripteurs [Azé, 2008].

- Un autre problème des MLP est celui du dimensionnement, c'est-à-dire, que nous ne pouvons pas connaître *a priori* les dimensions du réseau pour un problème donné. L'expérience montre qu'il n'est pas nécessaire d'avoir plus d'une couche cachée [Dreyfus, 2004]. En revanche, nous ne pouvons pas déterminer *a priori* le nombre de neurones de la couche cachée nécessaire à un problème donné [Chatelain, 2003a; , 2003b].

2.3.1.2 Séparateur à Vaste Marge

Appelé également Support Vector Machine, les SVM s'apparentent aux réseaux de neurones et sont issues de la même théorie, ou du moins du même domaine : l'apprentissage statistique [Deheeger, 2008]. En effet, les SVM permettent d'utiliser la souplesse des méta-modèles type réseaux de neurones en conservant une forte capacité de généralisation.

Les SVM sont des classificateurs qui reposent sur deux idées clés, qui permettent de traiter des problèmes de discrimination non-linéaire, et de reformuler le problème de classification comme un problème d'optimisation quadratique [Deheeger, 2008].

La première idée clé est la notion de marge maximale. La marge est la distance entre la frontière de séparation et les échantillons les plus proches. Ces derniers sont appelés vecteurs supports. Dans les SVM, la frontière de séparation est choisie comme celle qui maximise la marge. Ce choix est justifié par la théorie de « *Vapnik-Chervonenkis* » (VC) ou théorie statistique de l'apprentissage, qui montre que la frontière de séparation de marge maximale garantit de bonnes propriétés de généralisation [Wisniewski, 2007]. Le problème est de trouver cette frontière séparatrice optimale, à partir d'un ensemble d'apprentissage. Ceci est fait en formulant le problème comme un problème d'optimisation quadratique, pour lequel il existe des algorithmes connus.

Afin de pouvoir traiter des cas où les données ne sont pas linéairement séparables, la deuxième idée clé des SVM est de transformer l'espace de représentation des données d'entrées en un espace de plus grande dimension (possiblement de dimension infinie), dans

lequel il est probable qu'il existe un séparateur linéaire. Ceci est réalisé grâce à une fonction noyau, qui doit respecter certaines conditions, et qui a l'avantage de ne pas nécessiter la connaissance explicite de la transformation à appliquer pour le changement d'espace. Les fonctions noyau permettent de transformer un produit scalaire dans un espace de grande dimension, ce qui est coûteux, en une simple évaluation ponctuelle d'une fonction. Cette technique est connue sous le nom de « *kernel trick* ».

A. Avantages

Les SVM ont bouleversé le monde de l'apprentissage artificiel en mettant un nouveau principe d'apprentissage, ERM « *Empirical Risk Minimization* » [Taleb-Ahmed et Djadel, 2002]. En effet, le principe ERM repose sur le concept de marge qui possède une interprétation mathématique simple, et permet d'implémenter un algorithme de classification de manière incrémentale. La zone de marge correspond en fait à une zone d'incertitude. Pour améliorer ou affiner cette marge, de nouveaux points d'apprentissage sont à choisir dans la base d'apprentissage. Pour le moment, mettons en évidence certains des points clés de cette technique d'apprentissage et qui donne aux SVM la suprématie dans le domaine de la classification.

- En premier lieu, l'entraînement des SVM dépend seulement de la valeur des produits scalaires des vecteurs de données. Ce qui signifie que si nous avons une façon d'avoir les produits scalaires, le calcul ne dépend pas de la dimensionnalité de l'espace des caractéristiques, c'est-à-dire, des attributs de description des données [Turenne, 2006].
- La recherche des paramètres et la *Vapnik-Chervonenkis* dimension qui contrôle la complexité du modèle et la généralisation de celui-ci, permettent de passer outre le problème de la dimension. Ainsi le nombre d'échantillons et la capacité de généralisation sont indépendants de la dimension du problème [Deheeger, 2008].
- Le fait d'avoir seulement besoin des produits scalaires signifie qu'il est possible de substituer des fonctions plus générales pour l'opération de produit scalaire traditionnelle, l'astuce du noyau, pour obtenir des classificateurs plus puissants sans

réellement changer quelque chose dans l'entraînement actuel et les procédures de classification [Turenne, 2006].

- Un autre point à se rappeler est que le classificateur résultant ne dépend pas de tous les points d'entraînement mais seulement de ceux proches de la marge, en d'autres termes, ceux qui aident à définir la frontière entre les deux classes [Turenne, 2006].
- Cette technique permet de travailler avec des populations restreintes. En effet, il est rare d'avoir un très grand nombre de données expérimentales et les données numériques sont parfois difficiles à obtenir (temps de calcul très longs) [Deheeger, 2008].
- La contrainte de marge maximale aide à réduire la variance des hypothèses SVM. Insister sur la norme minimale du vecteur de poids réduit la taille de la classe d'hypothèse et aide à s'écarter du problème de sur-apprentissage [Turenne, 2006].
- Finalement, la technique ERM garantit un optimum global unique. Cela élimine beaucoup de tracas des autres approches d'apprentissage et permet aux SVM d'avoir une meilleure capacité de généralisation [Turenne, 2006].

B. Inconvénients

L'inconvénient majeur lors de l'utilisation des SVM est le temps de calcul qu'ils exigent, ceci est d'autant plus vrai lorsque la taille de l'ensemble d'apprentissage est grande. D'un autre côté, plus l'ensemble d'apprentissage est grand, plus le classificateur, généré par l'apprentissage, voit sa capacité de généralisation augmenter [Taleb-Ahmed et Djadel, 2002].

2.3.1.3 Choix final

Après avoir énumérer les principaux avantages et inconvénients du PMC et des SVM, nous nous sommes référés, afin de conforter notre choix, à un nombre d'études comparatives des performances des deux réseaux [Hurtado et Alvarez, 2001; Kelemen, Liang et Franklin, 2002; Mirmehdi et al., 2003; Osowski, Siwek et Markiewicz, 2004; Tarantino, D'addabbo et Castellana, 2005; Tyagi, 2008], essentiellement pour des problèmes de classification et de

prise de décision. Ces dernières ont montré, la supériorité des SVM en termes de généralisation et de précision par rapport au PMC. En conséquence, nous adopterons ce type de réseaux pour notre application.

2.4 Fondement théorique des SVM

Les SVM introduites au début des années 90, constituent la mise en pratique du principe de minimisation du risque structurel ERM. Elles réalisent ainsi un grand succès de la théorie de l'apprentissage statistique. Aujourd'hui, nous pouvons dire sans exagérer que ces machines ont supplanté les réseaux de neurones et les autres techniques d'apprentissage. En effet, elles sont largement répandues en apprentissage statistique et ont eu beaucoup de succès dans quasiment tous les domaines où elles ont été appliquées.

Les SVM exploitent les concepts relatifs à la théorie de l'apprentissage statistique et à la théorie des bornes de *Vapnik-Chervonenkis* pour aborder d'une façon nouvelle le problème de la classification. Le compromis entre la capacité d'apprentissage et la capacité de généralisation pour ces machines est respectivement accompli en minimisant l'erreur empirique et dans le même temps, en essayant de maximiser une marge géométrique. La justification intuitive de cette méthode d'apprentissage est la suivante : si l'échantillon d'apprentissage est linéairement séparable, il semble naturel de séparer parfaitement les éléments des deux classes de telle sorte qu'ils soient le plus loin possible de la frontière choisie [Ben-Ishak, 2007].

Ces fameuses machines ont été inventées en 1992 par Boser et *al.* [1992], mais leur dénomination par SVM n'est apparue qu'en 1995 avec Cortes et *al.* [1974] [Ben-Ishak, 2007]. Depuis lors, de nombreux développements ont été réalisés pour proposer des variantes traitant le cas non-linéaire, pour adapter cette méthode d'apprentissage à la régression ou pour introduire d'autres formulations des SVM ayant des liens avec les méthodes de régularisation.

Dans cette section nous présenterons en détail les SVM dans le cadre de la classification binaire et nous décrirons leur extension pour traiter le cas non-linéairement séparable. Nous mettrons également en évidence l'approche ERM ainsi que le principe de régularisation.

2.4.1 Problème de la classification linéaire [Ben-Ishak, 2007; Essid, 2005]

Le problème de la discrimination linéaire remonte aux années 1930 quand Fisher [Ben-Ishak, 2007] a proposé la première procédure de classification binaire. Jusqu'à nos jours, plusieurs autres procédures ont été introduites dont certaines manquent de bases théoriques encourageantes. Entre les années 1960 et 1980 les statistiques ont connu une révolution menée par l'apparition de la théorie de l'apprentissage statistique. Et depuis 1992, les SVM en sont le plus grand succès.

En effet, le problème de classification rentre bien dans le cadre de l'apprentissage statistique supervisé. Le but est de prévoir la classe de sortie Y d'un vecteur d'entrée X en se basant sur les mesures des variables qui l'expliquent avec pour seule information celle contenue dans l'échantillon d'apprentissage D où :

$$D = \{(x_1, y_1), \dots, (x_l, y_l)\}; X \in \mathbb{R}^d, y_i \in \{-1, +1\} \quad 2.1$$

Les x_i sont ici des vecteurs d'attributs pouvant appartenir à deux classes possibles : une classe positive, notée +1, et une classe négative notée -1. Les y_i représentent donc les étiquettes ou les valeurs cibles associées à x_i . Ces exemples sont supposés tirés à partir d'une distribution de probabilité inconnue $P(X, Y)$. La tâche d'apprentissage à partir des exemples consiste à trouver une fonction de classification f , permettant de prédire l'étiquette y_i de x_i :

$$f : X \in \mathbb{R}^d \mapsto \mathbb{R} \quad 2.2$$

Toute observation x_i est affectée à la classe qui correspond au signe de $f(x_i)$: si $f(x_i) \geq 0$, x_i est affecté à la classe positive (+1) sinon elle sera dans la classe négative (-1). En classification linéaire la fonction f est linéaire en x_i et elle prend la forme générale suivante :

$$f(x_i) = \langle w \cdot x_i \rangle + b \quad 2.3$$

w , b et $\langle . \rangle$ représentent respectivement le vecteur de poids, le biais et le produit scalaire

La méthodologie d'apprentissage implique que les paramètres $(w, b) \in \mathbb{R}^d \times \mathbb{R}$ doivent être estimés à partir des données d'apprentissage.

Géométriquement, ce classificateur divise l'espace des variables explicatives X en deux demi-espaces correspondant chacun à une classe. Cette séparation est réalisée par l'hyperplan $\mathbf{H}_{(w,b)}$ défini par l'équation :

$$\langle w \cdot x \rangle + b = 0 \quad 2.4$$

où le vecteur de poids w est un vecteur normal à l'hyperplan de séparation, tandis que la variation du paramètre b , engendre une simple translation de l'hyperplan séparateur.

En résumé, l'objectif de la discrimination linéaire est de trouver la bonne fonction de décision f dans la classe de toutes les fonctions linéaires de X dans \mathbb{R} que nous noterons $\mathcal{F}_{\mathcal{L}}$. La classe de tous les hyperplans qui en découle sera noté \mathcal{H} . Cette forme simple de classificateur a été énormément utilisée dans la classification binaire. Cette procédure est appelée discrimination linéaire pour les statisticiens [Fisher 1936] [Daniel, 1996] et perceptron dans le domaine des réseaux de neurones [Rosenblatt 1956] [Daniel, 1996].

Plusieurs algorithmes itératifs simples motivés par l'optimisation de différentes fonctions de coûts ont été introduits dans les années 1960 pour séparer linéairement des observations issues de deux populations différentes. Les SVM cherchent plutôt à séparer parfaitement les deux classes de telle sorte que leurs éléments soient le plus loin possible de l'hyperplan trouvé. C'est pour définir cette notion de "plus loin" que l'on introduit la marge.

2.4.1.1 Notion de marge

Les SVM sont par essence des classificateurs bi-classes qui visent à séparer les exemples de chaque classe au moyen d'un hyperplan $H_{(w,b)}$ choisi de manière à garder un maximum de marge de séparation entre n'importe quels exemples d'apprentissage et $H_{(w,b)}$. De façon plus formelle, il s'agit de déterminer l'hyperplan optimal solution de :

$$\max_{w \in \mathbb{R}^d, b \in \mathbb{R}} \min_{X,i} \{ \|x - x_i\|; X \in \mathbb{R}^d, \langle w \cdot x \rangle + b = 0, i = 1, \dots, l \} \quad 2.5$$

En utilisant une mise à l'échelle appropriée de w et b il est possible de contraindre les exemples de chaque classe à satisfaire les conditions :

$$\langle w \cdot x_i \rangle + b \geq +1, \text{ pour } y_i = +1 \quad 2.6$$

$$\langle w \cdot x_i \rangle + b \leq -1, \text{ pour } y_i = -1 \quad 2.7$$

qui peuvent être combinées en une même inégalité :

$$y_i(\langle w \cdot x_i \rangle + b) - 1 \geq 0, \forall i = 1, \dots, l. \quad 2.8$$

Les deux hyperplans :

$$H_1 : \langle w \cdot x_i \rangle + b = +1 \quad 2.9$$

$$H_2 : \langle w \cdot x_i \rangle + b = -1 \quad 2.10$$

permettent de définir la marge. Remarquons que H_1 et H_2 sont parallèles (ils ont la même normale w) et qu'il n'existe aucun point entre les deux grâce aux équations 2.6 et 2.7. Par suite, la marge n'est autre que la distance entre H_1 et H_2 qui vaut $2 / \|w\|$. La Figure 2.1 donne une illustration détaillée de ce principe.

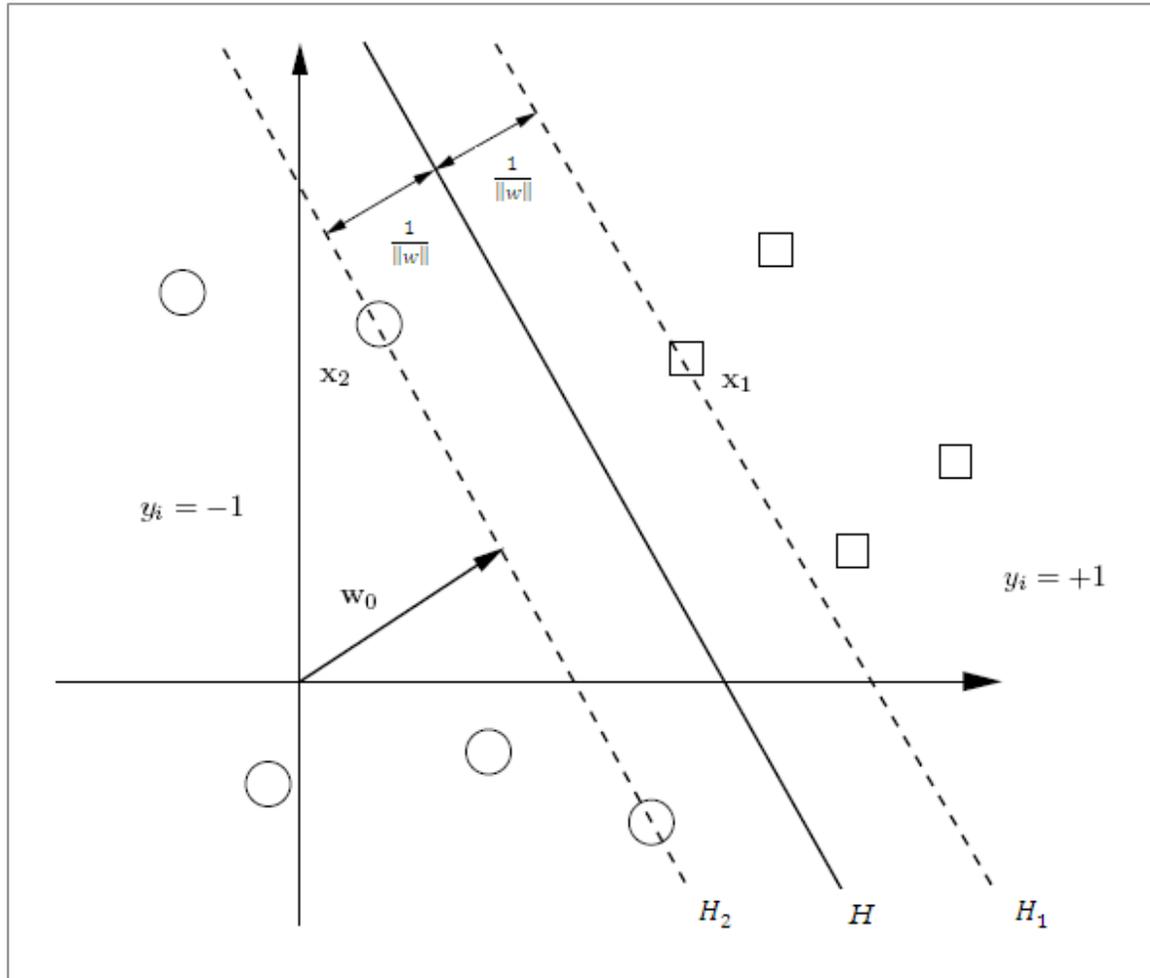


Figure 2.1 : Hyperplan optimal et marge d'un SVM. Les ronds représentent des exemples de classe -1 et les carrés, des exemples de classe +1 [Essid, 2005].

Les points qui se trouvent sur les hyperplans H_1 et H_2 sont appelés les vecteurs supports (SV : Support Vectors). Le problème posé ne dépend en fait que de ces points particuliers en ce sens que si tous les autres points sont éliminés, la solution du problème reste la même.

Ainsi, l'hyperplan optimal est solution du problème d'optimisation :

$$\begin{cases} \text{minimiser} & \tau(w) = \frac{1}{2} \|w\|^2, w \in \mathbb{R}^d, b \in \mathbb{R} \\ \text{Sous les contraintes} & c_i(x_i) = y_i(\langle w \cdot x_i \rangle + b) - 1 \geq 0, \forall i = 1, \dots, l. \end{cases} \quad 2.11$$

Le problème (équation 2.11) est un problème d'optimisation sous contraintes qui est résolu en introduisant des multiplicateurs de Lagrange $(\alpha_i)_{1 \leq i \leq l}$ et un Lagrangien [Thanh-Nghi et Fekete, 2007] :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i (\langle w, x_i \rangle + b) - 1] \quad 2.12$$

$$\text{Où } \alpha = [\alpha_1, \dots, \alpha_l]^T$$

Le Lagrangien L doit être minimisé par rapport aux variables dites primales w et b , et maximisé par rapport aux variables duales α_i : ce sont les conditions de Karush-Kuhn-Tucker (KKT) [Sholkopf et Smola, 2002].

Dans le cas où la fonction objectif, ici $\tau(w)$, et les contraintes, ici $c_i(x_i)$ sont convexes, les conditions KKT sont nécessaires et suffisantes, et la solution du problème est telle que :

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0 \quad 2.13$$

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \quad 2.14$$

$$y_i (\langle w, x_i \rangle + b) - 1 \geq 0 \quad 2.15$$

$$\alpha_i [y_i (\langle w, x_i \rangle + b) - 1] = 0 \quad 2.16$$

$$\alpha_i \geq 0 \quad 2.17$$

Les conditions (équations 2.13 et 2.14) donnent respectivement :

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad 2.18$$

$$\sum_{i=1}^l \alpha_i y_i x_i = w \quad 2.19$$

De plus, l'équation 2.16 implique que tous les points x_i qui ne sont pas vecteurs supports, *i.e.* ceux qui ne vérifient pas l'égalité $y_i(\langle w \cdot x_i \rangle + b) - 1 = 0$, sont associés à des α_i nuls. Ainsi, nous retrouvons que l'hyperplan optimal ne dépend que des n_s vecteurs supports du problème ($n_s \leq l$) :

$$\sum_{i=1}^{n_s} \alpha_i y_i x_i = w \quad 2.20$$

et la fonction de décision est définie par le signe de :

$$f(x) = \langle w \cdot x \rangle + b = \sum_{i=1}^{n_s} \alpha_i y_i \langle x \cdot x_i \rangle + b \quad 2.21$$

Le paramètre b peut être déterminé au travers de la condition (équation 2.16) en choisissant un indice i tel que $\alpha_i \neq 0$, ou encore en moyennant les valeurs obtenues en utilisant tous les points x_i associés à des α_i non nuls (pour une meilleure robustesse numérique).

En utilisant les équations 2.20 et 2.12, nous obtenons la formulation duale du problème :

$$\left\{ \begin{array}{l} \text{maximiser} \quad L_D(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle \\ \text{Sous les contraintes :} \quad \sum_{i=1}^l \alpha_i y_i = 0 \\ \quad \quad \quad \quad \quad \quad \quad \quad \alpha_i \geq 0 \end{array} \right. \quad 2.22$$

Il est intéressant de remarquer que w et b ont été éliminés et qu'il s'agit désormais de déterminer les α_i ce qui facilite énormément le calcul, surtout quand nous avons tous les points x_i qui ne sont pas vecteurs supports sont associés à des α_i nuls.

2.4.1.2 Marge Molle [Callut, 2003; Essid, 2005; Wisniewski, 2007]

La formulation du problème d'optimisation (équation 2.11) suppose que les données soient linéairement séparables, ce qui n'est généralement pas le cas pour des données réelles. En 1995, Corinna Cortes et Vladimir Vapnik [Vapnik, 1998] proposent une technique dite de

marge molle qui permet de relâcher les contraintes sur les vecteurs d'apprentissage en insérant des variables ressorts « *slack variables* ». À titre d'exemple (voir Figure 2.2), pour une donnée d'apprentissage (x_i, y_i) , il existe une variable ressort ξ_i qui traduit de combien la contrainte de la marge peut être violée.

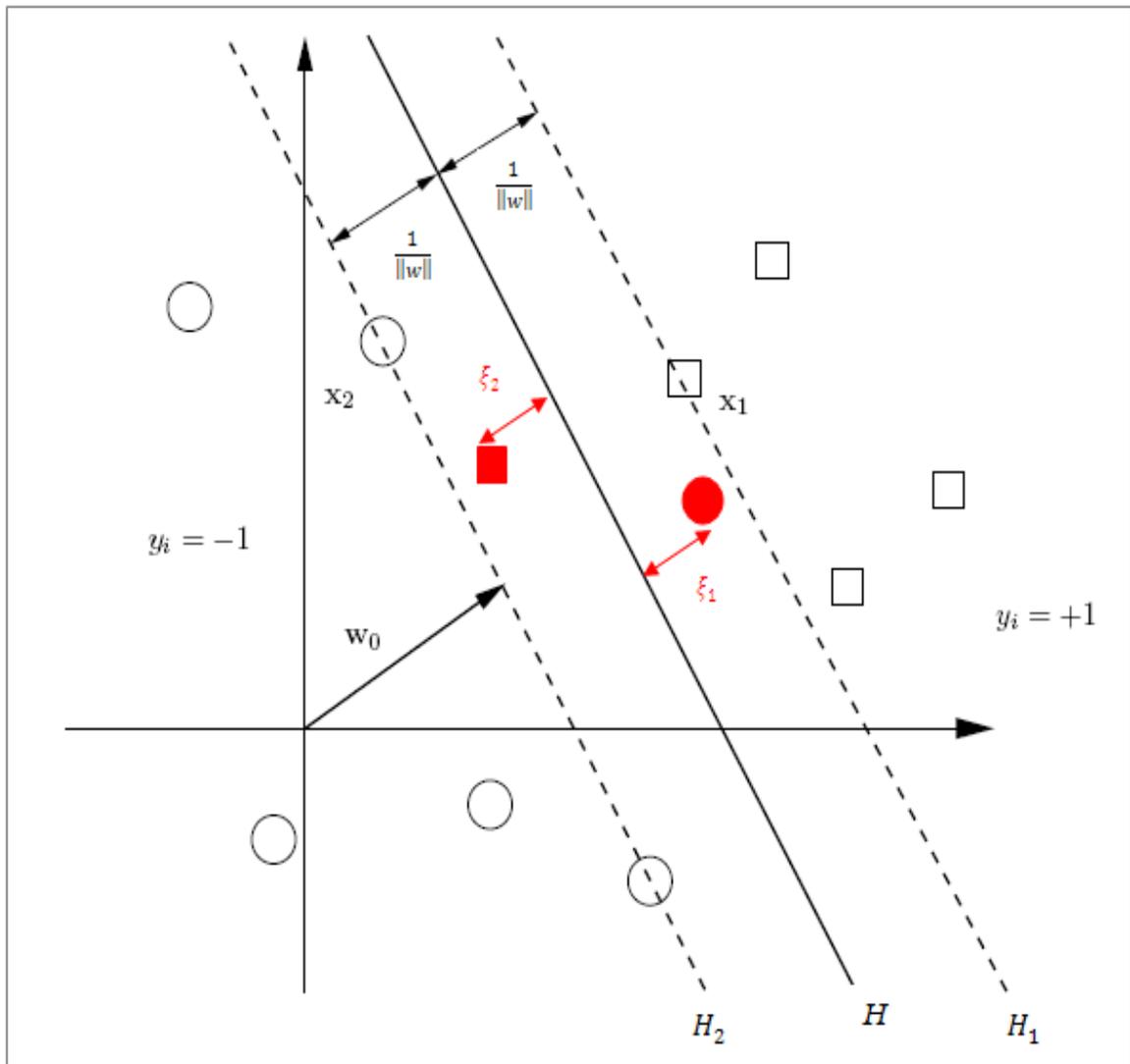


Figure 2.2 : Illustration de la méthode des SVM pour les classes non séparables. Introduction des variables molles : à chaque point d'apprentissage non classé du bon côté par le séparateur est associée une variable d'erreur [Deheeger, 2008].

Le problème d'optimisation correspondant est défini par :

$$\begin{cases} \langle w \cdot x_i \rangle + b \geq 1 - \xi_i & \text{Pour } y_i = +1 \\ \langle w \cdot x_i \rangle + b \leq -1 - \xi_i & \text{Pour } y_i = -1 \\ \xi_i \geq 0 & \forall i = 1, \dots, l \end{cases} \quad 2.23$$

Pour qu'un exemple d'apprentissage x_i soit mal classé, il faut que le ξ_i correspondant soit supérieur à 1. Par suite $\sum_i \xi_i$ est une borne supérieure sur le nombre d'erreurs de classification qui peuvent être pénalisées en modifiant la fonction objectif $\tau(w)$ par :

$$\tau(w, \xi) = \frac{\|w\|^2}{2} + C(\sum_i^l \xi_i) \quad 2.24$$

où $\xi = [\xi_1, \dots, \xi_l]^T$ et $C > 0$ est un paramètre permettant de contrôler le compromis entre le fait de maximiser la marge et minimiser les erreurs de classification commises sur l'ensemble d'apprentissage. Notons qu'il est souvent préférable de tolérer certaines erreurs, au bénéfice d'une marge plus grande car ces erreurs peuvent être dues à des « *outliers* », observations aberrantes, non-significatives de la classe qui leur est associée.

Pour ce cas de figure, le Lagrangien primal est :

$$L(w, b, \alpha, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i [y_i (\langle w \cdot x_i \rangle + b) - 1 + \xi_i] \quad 2.25$$

Le problème dual est plus simple, il prend la forme :

$$\begin{cases} \text{maximiser} & L_D(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle \\ \text{Sous les contraintes :} & \sum_{i=1}^l \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \end{cases} \quad 2.26$$

Par rapport au cas séparable, une contrainte supplémentaire sur les α_i a été introduite : ils admettent à présent la borne supérieure C . En réécrivant les conditions KKT nous retrouvons la même solution (équation 2.20) :

$$\sum_{i=1}^{ns} \alpha_i y_i x_i = w$$

Cette nouvelle condition ne dépend que des vecteurs supports (SV) à la différence que dans cette réalisation “molle” des SVM, w dépend, en plus des SV se trouvant à la marge (sur les hyperplans H_1 et H_2), de vecteurs supports se retrouvant à l’intérieur de la marge, appelés erreurs de marge, qui sont associés à des multiplicateurs $\alpha_i = C$; ils sont désignés par *Bounded Support Vectors* (BSV).

Les conditions KKT permettent en outre de déduire que les variables d’écart ξ_i sont nulles pour tous les vecteurs supports associés à des multiplicateurs α_i tels que $0 < \alpha_i < C$, ce qui permet de calculer b de la même façon que dans le cas séparable.

2.4.2 Problème de la classification non linéaire [Ben-Ishak, 2007; Deheeger, 2008; Essid, 2005]

Les SVM présentés ci-dessus ne concernent que les modèles linéaires. Il serait complètement illusoire de se dire que nous pourrions séparer efficacement n’importe quel jeu de données par un simple hyperplan.

Pour y remédier à ce problème, une idée simple consistée à doter les SVM d’un mécanisme permettant de produire des surfaces de décision non-planes. En effet, il s’agit de transformer les données de l’espace de départ \mathbb{R}^d dans un espace de Hilbert \mathbb{E} de dimension supérieure, possiblement infinie, dans lequel les données transformées deviennent linéairement séparables. Ainsi, en exploitant une application :

$$\Phi : \mathbb{R}^d \mapsto \mathbb{E} \tag{2.27}$$

L’algorithme SVM linéaire appliqué aux données $\Phi(x_i)$ dans l’espace \mathbb{E} , appelé également Espace Support, produit des surfaces de décision non-planes dans l’espace de départ \mathbb{R}^d , mieux appropriées aux données de départ pour un choix judicieux de Φ .

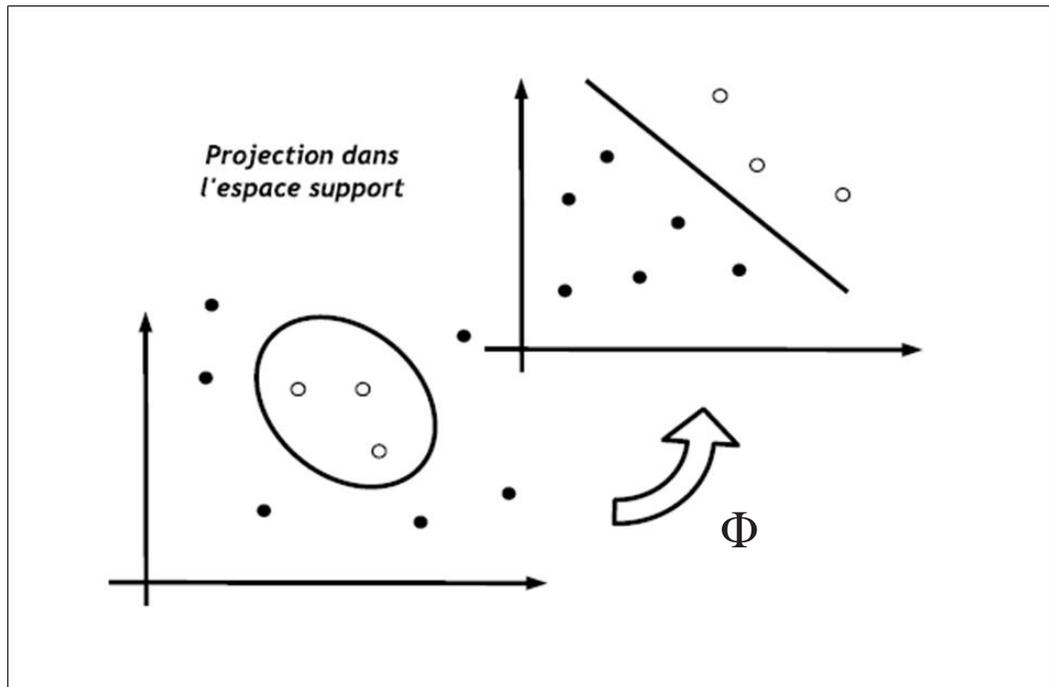


Figure 2.3 : Illustration de projection de données de l'espace de départ vers l'espace support [Deheeger, 2008].

Cette procédure peut être rendue très efficace en utilisant une astuce permettant d'effectuer les calculs nécessaires à l'algorithme dans l'espace de départ \mathbb{R}^d sans passer explicitement dans \mathbb{E} . Du fait que les données apparaissent dans tous les calculs uniquement sous forme de produits scalaires $\langle x_i, x_j \rangle$, il suffit de trouver une façon efficace de calculer $\Phi(x_i) \cdot \Phi(x_j)$. Cela est réalisé en faisant appel à une fonction *noyau* $k(x_i, x_j)$, définie par :

$$k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad 2.28$$

Tout le développement présenté dans le cas linéaire reste valable en remplaçant simplement les termes $\langle x_i, x_j \rangle$ par $k(x_i, x_j)$. La nouvelle fonction de décision est définie par le signe de :

$$f(x) = \sum_{i=1}^{ns} \alpha_i y_i k(s_i, x) + b \quad 2.29$$

où les s_i sont les vecteurs supports.

L'avantage d'une telle approche réside dans le fait qu'il n'est pas nécessaire de connaître Φ explicitement. Il suffit d'obtenir des noyaux convenables. C'est ce que nous discutons dans la section suivante.

2.4.2.1 Noyaux [Essid, 2005]

Sous quelles conditions une fonction $k(\mathbf{x}, \mathbf{y})$ symétrique est-elle associée à un espace \mathbb{E} et une transformation Φ vers cet espace ?

La réponse est donnée par les conditions de Mercer [Essid, 2005] qui stipulent qu'il existe une application Φ et un développement de $k(\mathbf{x}, \mathbf{y})$ de la forme :

$$k(x, y) = \sum_{i=1}^{+\infty} \Phi_i(x) \cdot \Phi_i(y) \quad 2.30$$

ce qui traduit le fait que $k(\mathbf{x}, \mathbf{y})$ décrit un produit interne dans un espace \mathbb{E} , si et seulement si pour toute fonction $g(\mathbf{x})$ sur \mathbb{R}^d , de norme L_2 finie (*i.e.* $\int g(\mathbf{x})^2 d\mathbf{x}$ est finie) la condition suivante est satisfaite :

$$\int k(x, y) g(x) g(y) dx dy \geq 0 \quad 2.31$$

Différentes formes de noyau (vérifiant les conditions de Mercer) ont été proposées. Nous présenterons dans ce qui suit les deux formes les plus utilisées à savoir :

a. Noyau Polynômial : Le noyau polynômial de degré δ correspond à une transformation Φ par laquelle les composantes des vecteurs transformés $\Phi(\mathbf{x})$ sont tous les monômes d'ordre δ formés à partir des composantes de \mathbf{x} . Par exemple, pour $d = \delta = 2$, le noyau.

$$k(x, y) = (x \cdot y)^2 \quad 2.32$$

Correspond à la transformation :

$$\Phi : \mathbb{R}^2 \mapsto \mathbb{E} = \mathbb{R}^3 \quad 2.33$$

$$x = [x_1, x_2]^T \mapsto [x_1^2, x_2^2, x_1 \cdot x_2]^T$$

Le noyau polynômial permet ainsi d'effectuer la classification sur des nouveaux attributs qui sont tous les produits d'ordre δ des attributs de départ.

Il est possible dans ce cas de calculer la dimension $d_{\mathbb{E}}$ de l'espace transformé \mathbb{E} correspondant à un noyau polynômial de degré δ en comptant le nombre de monômes d'ordre δ possibles. Il vient :

$$d_{\mathbb{E}} = C_{\delta+d+1}^{\delta} = \frac{(\delta+d-1)!}{\delta!(d-1)!} \quad 2.34$$

Signalons qu'il est également possible de recourir à des noyaux polynômiaux dits inhomogènes de la forme :

$$k(x, y) = (x \cdot y + 1)^{\delta} \quad 2.35$$

Ces modélisations permettent de prendre en compte tous les monômes d'ordre inférieur ou égal à δ .

b. Noyau radial exponentiel (RBF-Radial Basic Function) : Ce noyau a la forme générale suivante :

$$k(x, y) = \text{Exp} \left(-\frac{\|x-y\|^2}{2\sigma^2} \right) \quad 2.36$$

La Figure 2.4 montre les surfaces de décision correspondant à des valeurs croissantes de σ . nous pouvons constater que ce paramètre permet de contrôler la courbure des surfaces de décision. Les σ élevés correspondent des surfaces présentant des courbures plus importantes.

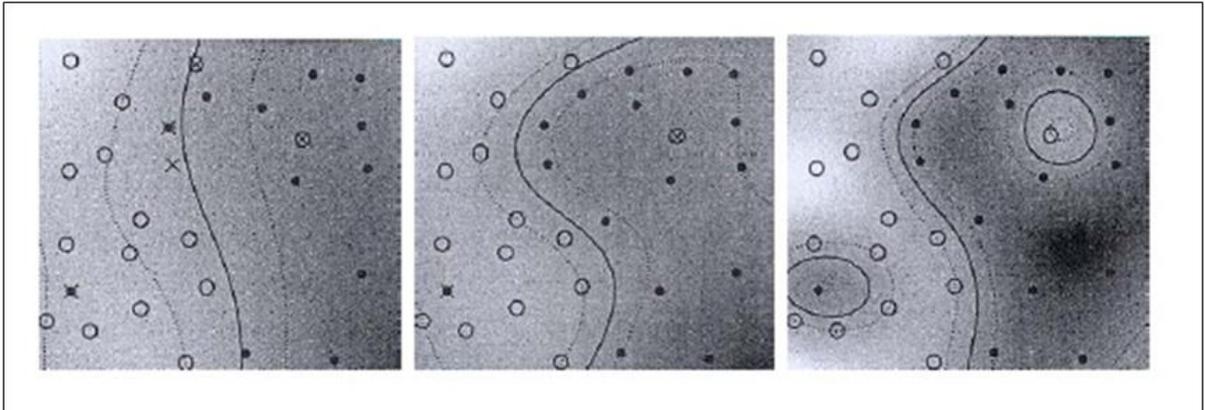


Figure 2.4 : Effet du paramètre σ . De gauche à droite σ^2 diminue [Sholkopf et Smola, 2002].

Sur la Figure 2.4 les lignes continues indiquent les surfaces de décision et les lignes interrompues les bords de la marge. Notons que pour les grandes valeurs de σ^2 , le classificateur est quasi linéaire et la surface de décision ne parvient pas à séparer les données correctement. A l'autre extrême, les valeurs trop faibles de σ^2 donnent lieu à des surfaces de décision qui suivent de trop près la structure des données d'apprentissage et il y a un risque de sur-apprentissage. Il est donc nécessaire de réaliser un compromis tel que celui réalisé dans l'image du milieu.

2.4.3 Performances en généralisation des SVM [Ben-Ishak, 2007; Deheeger, 2008; Essid, 2005]

Les SVM présentent en pratique de très bonnes performances en généralisation, c'est-à-dire, sur la classification de nouveaux exemples de test. Intuitivement, nous sentons que la marge joue en cela un rôle important. Il est en effet raisonnable de penser que si nous parvenons à séparer les exemples d'apprentissage avec une grande marge, il y a de fortes chances pour que de nouveaux exemples soient bien classés, ces derniers se situant dans les cas les plus défavorables à l'intérieur de la marge.

Une autre caractéristique frappante des SVM est qu'ils sont connus pour défier ce que nous appelle « *the curse of dimensionality* » puisqu'ils sont capables de fournir des bonnes performances de classification à partir d'un nombre réduit d'exemples d'apprentissage tout

en agissant dans des espaces de dimensions très élevées. Cela s'explique en partie par le fait que les SVM peuvent être considérés comme une réalisation du principe ERM.

Il est utile de définir au préalable quelques fonctions permettant de déterminer le degré d'apprentissage :

- tout d'abord la fonction de coût $L(y, h(x, w))$ de la machine d'apprentissage. y est la fonction à approximer et $h(x, w)$ la machine d'apprentissage, où w est un vecteur de paramètres définissant la machine d'apprentissage.
- Nous définissons également une fonction risque qui est la valeur espérée de la fonction coût :

$$R(w) = \int L(y, h(x, w)) dP(x, y) \quad 2.37$$

Si la distribution $P(x, y)$ est dérivable, l'expression de l'intégrale est réduite :

$$R(w) = \int L(y, h(x, w)) P(x, y) dx dy \quad 2.38$$

Comme la densité conjointe P est inconnue, le risque est souvent approximé par le risque empirique $R_e(w)$ évalué à partir d'un échantillon de n individus calculés jusque-là. Il s'agit alors de minimiser un tel risque. Cette approximation est le principe ERM.

Pour un problème de classification nous avons $y \in \{+1, -1\}$. La fonction $h(x, w) = \text{Sgn}(f(x, w))$ est recherchée. L'objectif est de trouver la valeur de w qui minimise l'erreur calculée sur l'échantillon. La fonction coût est la suivante :

$$L(y, h(x, w)) = \begin{cases} 0 & \text{si } y = h(x, w) \\ 1 & \text{si } y \neq h(x, w) \end{cases} \quad 2.39$$

Le risque empirique associé est de la forme :

$$R_e = \frac{1}{n} \sum_{i=1}^n L(y, h(x_i, w)) \quad 2.40$$

2.4.3.1 Principe de régularisation

Le problème de choix du méta-modèle est extrêmement important. En effet, ce choix doit être contrôlé puisqu'une fonction dont la complexité est excessive peut conduire à un sur-apprentissage.

De l'idée de contrôle de la complexité du modèle est née le principe de régularisation (ou de pénalisation). Il s'agit ici, non plus de minimiser uniquement le risque empirique, mais d'y ajouter un terme qui va dépendre de la complexité de la fonction. Le problème d'optimisation s'écrit alors :

$$\text{minimiser } R_r(w, n) = R_e(w, n) + \lambda \Psi(h(x, w)) \quad 2.41$$

où $\Psi(h(x, w))$ est une fonction de régularisation qui dépend de la complexité du méta-modèle. Le coefficient λ , coefficient de Lagrange positif et non nul, permet de contrôler le poids de la complexité de la fonction sur le risque global.

Dans le cas des SVM ce principe est assuré par la dimension VC. La VC dimension d'une fonction correspond à une mesure de sa complexité. La définition est la suivante : c'est le nombre maximum v d'échantillons qu'il est possible de séparer par le modèle selon les 2^v façons, où $v = d + 1$ dans le cas d'un modèle linéaire, et d est la dimension de l'espace.

La VC dimension est illustrée dans le plan où $v = 3$ pour la droite, comme le montre la Figure 2.5. Dans ce cas de figure, il y a $2^v = 8$ façons de classer les 3 points.

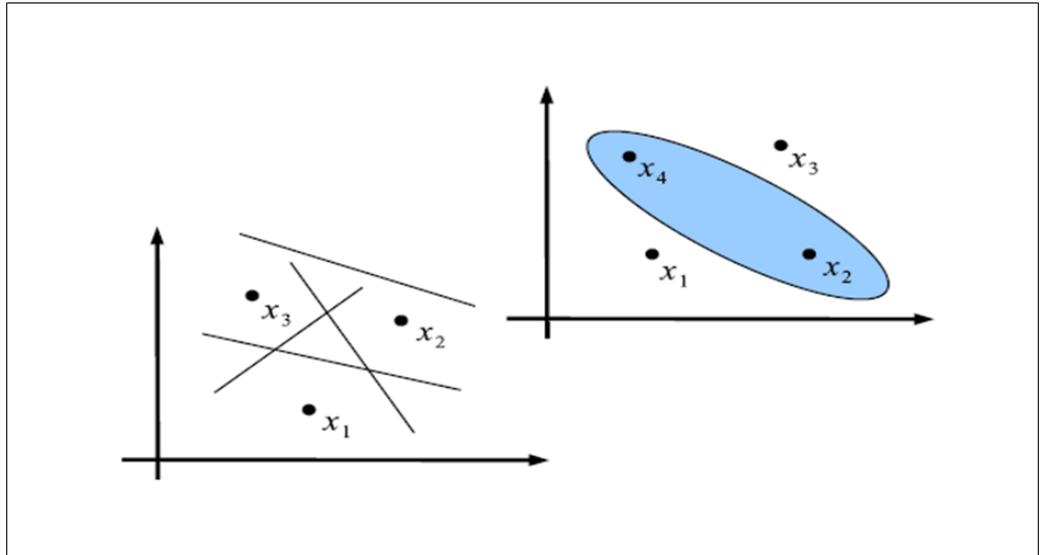


Figure 2.5 : La dimension VC pour l'hyperplan dans un espace de dimension 2 (donc pour la droite) est de 3 [Deheeger, 2008].

Sur Figure 2.5, à gauche, toutes les possibilités de séparer 3 points sont réalisables par une droite, en revanche, il est impossible de séparer de toutes les façons 4 points : à droite de la Figure 2.5, les points x_2 et x_4 ne peuvent être séparés des autres réalisations par une droite. Il est important de retenir que cette dimension mesure la complexité d'un méta-modèle, avant de passer à l'explication du principe ERM.

Vapnik [Vapnik, 1998] a démontré, dans une série de théorèmes, l'existence de bornes pour le risque de généralisation, bornes liées à la dimension VC du méta-modèle. Pour la classification binaire en particulier, il a montré qu'il y a une probabilité $1 - \eta$ que le risque de classification admette l'inégalité suivante :

$$R(w, n) \leq R_e(w, n) + \frac{q}{2} \left(1 + \sqrt{1 + \frac{4 + R_e(w, n)}{q}} \right) \quad 2.42$$

où

$$q = \frac{\alpha v (\ln(\beta^n / v) + 1) - \ln(\eta / 4)}{n} \quad 2.43$$

Les paramètres α et β sont compris dans $0 < \alpha \leq 4$ et $0 < \beta \leq 2$, v indique la dimension VC du méta-modèle. Pour une dimension VC égale à v , si le rapport n/v est grand (associé à un grand nombre de points), alors q est petit et il suffit simplement de minimiser le risque empirique, ce qui est fait habituellement. Dans le cas contraire, pour un petit nombre de points, il est nécessaire de minimiser les deux termes du risque pour minimiser la somme. Dans ce cas de figure, et lorsque les points d'apprentissage sont coûteux, il est utile de contrôler la variable v pour minimiser le risque, c'est-à-dire, contrôler la complexité du méta-modèle que nous utilisons.

2.4.3.2 Empirical Risk Minimization (ERM)

Les bornes de Vapnik permettent de minimiser le risque global d'un séparateur à partir d'un échantillon constitué de quelques points seulement, par le contrôle de la variable de la dimension VC. Le but de l'optimisation est alors de trouver la valeur de VC qui va minimiser la somme du risque empirique et de l'intervalle de confiance sur le risque de généralisation, c'est le principe de minimisation du risque empirique (ERM) (voir Figure 2.6).

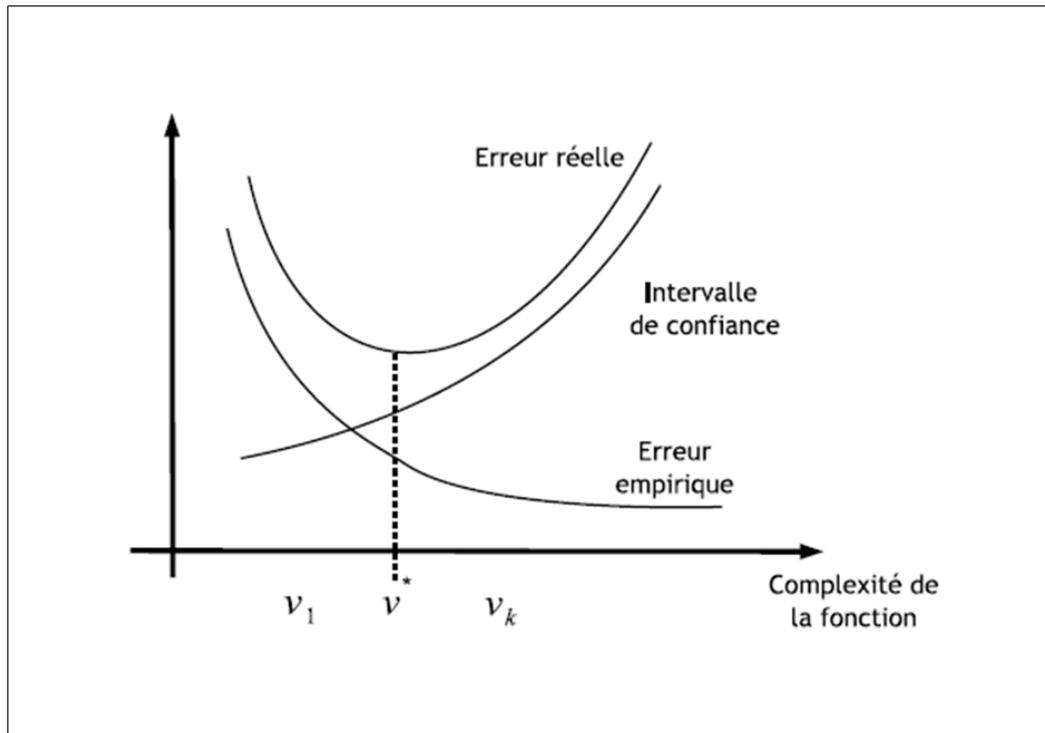


Figure 2.6 : Illustration du principe ERM appliquée par les SVM. La borne sur l'erreur de généralisation permet d'optimiser en parallèle la complexité de la fonction et de l'erreur empirique mesurée sur l'échantillon [Deheeger, 2008].

Il est possible d'obtenir, pour un minimum de points d'apprentissage, un séparateur avec une capacité maximale de généralisation (capacité maximale autorisée pour le nombre de points utilisé). Cette optimisation est réalisée de manière intrinsèque dans le concept de maximisation de la marge qui est à la base de la méthode des SVM.

2.4.4 Extension des SVM vers le cas multi-classes [Le-Meur, 2005]

Plusieurs méthodes ont été proposées pour étendre les SVM au cas où plus de deux classes sont à séparer. Ces méthodes sont applicables à tout classificateur binaire, et ne sont donc pas spécifiques aux SVM. Les deux plus connues sont appelées « *one versus all* » et « *one versus one* ». Formellement, les échantillons d'apprentissage et de test peuvent ici être classés dans M classes.

2.4.4.1 One versus all

Cette méthode consiste à construire M classificateurs binaires en attribuant le label 1 aux échantillons de l'une des classes et le label -1 à toutes les autres. En phase de test, le classificateur donnant la valeur de confiance ou la marge la plus élevée remporte le vote.

2.4.4.2 One versus one

Cette méthode consiste à construire $M(M - 1)/2$ classificateurs binaires pour tous les couples de classes possibles en confrontant chacune des classes. En phase de test, l'échantillon à classer est analysé par chaque classificateur dont la sortie fournit un vote partiel concernant uniquement un couple de classes et un vote majoritaire permet ensuite de déterminer la classe.

2.5 Conclusion

Ce chapitre a eu pour objectif d'exposer l'outil qui va être utilisé, dans le cadre de notre travail, pour modéliser notre décideur individuel et qui est les SVM. En effet, nous avons justifié le choix de cet outil ainsi que son élégante construction théorique. Le prochain chapitre représente la deuxième étape logique de notre travail car il va être consacré à la modélisation de la prise de décision par un groupe de décideurs où une approche multi-agent sera présentée.

CHAPITRE 3

LES SYSTÈMES MULTI-AGENTS, VERS UNE INTELLIGENCE COLLECTIVE ORIGINELLE

3.1 Introduction

Le dirigeant a bien souvent été représenté comme le grand stratège, à la tête d'un arsenal de prévisions, scénarios, calculs, méthodes, plans, programmes, projets, qu'une organisation disciplinée et efficace élabore sous sa direction, et dont il conduit fermement mais sereinement la réalisation. Malheureusement, nous pouvons affirmer qu'aujourd'hui l'action d'un dirigeant ne correspond pas à l'image mythique du grand patron, clairvoyant et omniprésent.

De nombreuses études empiriques ont montré que les processus de décision dans une organisation s'écartent sensiblement de cette représentation mythique. Tout d'abord, c'est d'un environnement incertain et hostile que naissent et disparaissent les menaces qui peuvent affecter l'organisation, le rôle du dirigeant implique donc de mettre cette dernière en adéquation avec cet environnement, qui décourage souvent toute anticipations, en essayant de trouver des moyens de réduire l'incertitude générée par la multiplicité des acteurs.

Cependant, le dirigeant lui-même ne ressemble guère à ce décideur aux capacités exceptionnelles. En effet, le décideur ne domine ni l'environnement auquel il est confronté ni l'organisation qu'il dirige. Il n'est pas non plus, ainsi que le mythe le dépeint, idéalement lucide et rationnel. La liste est longue des biais cognitifs auxquels les dirigeants sont exposés. De plus, le dirigeant est confronté à un dilemme : en privilégiant les approches méthodiques, traitant des données objectives et autant que possible quantifiées, il perd les informations qualitatives, impalpables mais essentielles, mais en faisant confiance à l'expérience, aux analogies, à sa perception, à l'intuition, il risque d'être victime de sa subjectivité. Le décideur n'aborde pas les situations d'un œil vierge. Il a notamment des représentations de ce qui est important ou non, de ce qu'il convient de faire ou non. Il s'appuie sur des règles et principes

qui sont pour lui des évidences, mais qui pour l'observateur apparaissent comme des croyances, c'est-à-dire des manières de penser dont la validité n'est pas démontrée.

Nous pouvons ainsi s'interroger sur la capacité des dirigeants à contrôler leurs décisions? Entre les incertitudes générées par l'environnement, les dérives organisationnelles, les croyances des dirigeants, y a-t-il donc une chance de voir se former des décisions sensées et cohérentes?

3.2 Intelligence collective originelle

L'intelligence collective n'est ni une nouveauté, ni une découverte, bien au contraire ! Elle est fondatrice des organisations sociales : groupe, tribu, entreprise, équipe, gouvernements, nations, associations..., lorsque les individus rassemblés pour échanger et collaborer, trouvent un avantage tant individuel que collectif supérieur à ce qui aurait été obtenu si chacun était resté isolé. L'intelligence collective constitue le fondement de ce que nous appelons les économies à somme positive.

C'est pourquoi, depuis quelques années, l'Intelligence Collective devient une discipline, avec son cadre théorique, son savoir-faire pratique, ses domaines de recherche, ses méthodologies et heuristiques. Elle vise à maximiser le potentiel d'action et de liberté des collectifs humains dans un contexte où la plupart des challenges que nous rencontrons semblent trop complexes à résoudre. L'Intelligence Collective développe les leviers pour la maîtrise du pilotage dans les météos capricieuses de la complexité.

Ainsi, la science de l'Intelligence Collective a pour objet l'étude et l'optimisation des propriétés émergentes *intérieures-subjectives* et *extérieures-objectives* des collectifs, et ce dans le but d'augmenter leur capacité d'existence, d'évolution et de plénitude. Ce faisant, elle invente les outils d'une gouvernance universelle (globale, locale, transversale, transculturelle...) tout en développant aussi des savoir-faire pratiques et immédiats pour les organisations d'aujourd'hui à travers une éthique de la collaboration [Noubel, 2007].

Chez l'humain les types de collectifs sont nombreux, cependant, il existe une constitution *optimale* qui permet de produire ce que nous appelons l'Intelligence Collective Originelle [Noubel, 2007]. Cette dernière est tout simplement l'intelligence en petit groupe où nous retrouvons un petit nombre de personnes en proximité sensorielle, donc spatiale, les unes vis-à-vis des autres.

Sur le plan de la psychologie sociale, un premier constat est que lors d'une prise de décision, cette constitution permet au groupe d'adopter des choix plus risqués que les individus, le groupe est alors supposé favoriser la dilution des responsabilités, la réduction de l'engagement des individus. Nous pouvons penser au contraire, que dans certaines circonstances, les individus s'engagent plus au sein d'un groupe : il a été en effet observé que la décision en groupe provoque un renforcement considérable des options initiales quelles qu'elles soient [Plantie, 2006].

Un autre point de la décision au sein d'un petit groupe concerne la nécessité de parvenir à un consensus [Plantie, 2006]. L'engagement des individus s'en trouve augmenté car, s'il n'y a pas nécessité d'aboutir à une décision commune, les membres d'un groupe débattent de leurs opinions personnelles, mais sans chercher à retravailler cognitivement les opinions exprimées par les autres. L'objectif d'un consensus ne doit pas être de supprimer les conflits, mais de les tolérer. Plus les options individuelles sont variées, plus la confrontation est importante et plus la reformulation du problème peut devenir explicite. La recherche d'un compromis par le groupe incite les individus à reconsidérer leurs positions sur le problème. Lorsque les membres du groupe se sont mis d'accord sur une conclusion, nous constatons qu'ils la maintiennent ensuite lorsque nous les interrogeons individuellement. La possibilité de laisser s'exprimer les opinions divergentes favorise donc la cohésion du groupe. Par ailleurs, il faut noter que l'innovation vient souvent des points de vue minoritaires dissidents. Ainsi, pour être véritablement efficaces, les mécanismes de décision doivent favoriser la contradiction.

3.2.1 Caractéristiques de l'intelligence collective originelle [Noubel, 2007]

L'intelligence collective originelle est le fruit de la réunion d'un nombre de caractéristiques majeures que nous retrouvons dans tous collectifs respectant la constitution citée dessus. Chaque caractéristique est à la fois cause et conséquence des autres, aucune ne saurait être prise isolément. Plus elles sont développées et coordonnées, plus le collectif se montre capable d'évoluer et de créer le futur en contexte complexe, inattendu et incertain.

3.2.1.1 Un tout émergent

Chaque groupe possède un *caractère*, un *style*, un *esprit* différents auxquels nous nous référons, comme s'il s'agissait d'une seule individualité... Notons que plus ce *Tout* est manifeste, plus il souligne implicitement la réussite du groupe en tant que tel.

3.2.1.2 Un espace holoptique

La proximité spatiale offre à chaque participant une perception complète et sans cesse réactualisée de ce Tout. Chacun, grâce à son expérience et expertise, s'y réfère pour anticiper ses actions, les ajuster et les coordonner avec celles des autres. Il existe donc un aller-retour incessant, qui fonctionne comme un miroir, entre les niveaux individuel et collectif. Nous nommerons *holoptisme* l'ensemble de ces propriétés, à savoir la transparence *horizontale* (perception des autres participants) à laquelle s'ajoute la communication *verticale* avec le Tout émergent du collectif. En d'autres termes, l'holoptisme est un espace qui permet à tout participant de percevoir en temps réel les manifestations des autres membres du groupe (axe horizontal) ainsi que celles provenant du niveau supérieur émergent (axe vertical).

3.2.1.3 Un contrat social

Le collectif est fondé autour d'un contrat social, tacite ou explicite, objectif ou subjectif, souvent les deux à la fois, accepté et mis en scène par chacun des participants. Le contrat

social porte non seulement les valeurs et les règles du groupe, mais également sa raison d'être, donc son inscription dans le futur.

3.2.1.4 Une architecture polymorphe

La cartographie des relations entre les participants se réactualise sans cesse en fonction des circonstances, des expertises, de la perception de chacun, des tâches à accomplir, des règles définies par le contrat social. Elle se magnétise autour des expertises, chaque expert, reconnu comme tel par le groupe, prenant tour à tour le « *lead* » au fil des besoins.

3.2.1.5 Un objet-lien en circulation

Nul doute que l'intelligence collective originelle se construit dans la convergence des individualités vers un objet collectivement poursuivi, que cet objet ne soit matériel ou symbolique. Quand ils appartiennent à l'espace symbolique, il est absolument nécessaire que ces objets soient clairement identifiés dans leur nombre et qualité par chaque participant du groupe, sinon cela mène à des situations floues et peuvent inciter la divergence dans le groupe.

3.2.1.6 Une organisation apprenante

L'apprentissage concerne non seulement le niveau individuel, mais il implique également l'existence d'un processus social qui prend en charge l'erreur, l'intègre et la transforme en objet de cognition partagé. Il engage le développement d'une intelligence relationnelle : ce que nous apprenons pour soi est valable pour les autres.

3.2.1.7 Une économie du don

Dans l'économie de compétition, nous prenons pour soi contre une compensation. Dans l'économie du don, nous donnons d'abord, nous recevons en retour une fois que le collectif a

gagné en richesse. En effet, l'économie du don constitue le socle absolu de la vie sociale, à tel point que nous n'en avons que rarement conscience. Quel collectif pourrait fonctionner sur le long terme s'il reposait sur une dynamique sacrificielle individuelle? Dans l'économie du don, chaque participant trouve un bénéfice individuel fort qui le motive à donner le meilleur de lui-même. L'économie du don organise la convergence entre les niveaux individuel et collectif.

3.3 Évolution vers une Intelligence Artificielle Distribuée

L'intelligence artificielle est reconnue comme étant une discipline qui a pour objectif de modéliser ou de simuler des comportements humains dits intelligents. Elle s'attache à la construction de programmes informatiques capables d'exécuter des tâches complexes, en s'appuyant sur une centralisation et une concentration de l'intelligence au sein d'un système unique [Kabachi, Vincent et Ouzrout, 1996]. L'évolution des applications de l'Intelligence Artificielle pour recouvrir des domaines complexes et hétérogènes tels que l'aide à la décision, la reconnaissance des formes et la conduite des processus industriels, etc., a montré les limites de cette approche, dues pour la plus part à la nécessité d'intégrer, au sein d'une même base de connaissances, l'expertise, les compétences et les connaissances d'individus différents qui, dans la réalité, communiquent et collaborent à la réalisation d'un but commun [Noomen, 1998].

C'est à la suite de ces considérations qu'au milieu des années soixante-dix, est apparu l'Intelligence Artificielle Distribuée (IAD), comme une branche à part entière [Kebair, 2009]. L'IAD a pour but de remédier aux insuffisances de l'approche classique de l'intelligence artificielle en proposant la distribution de l'expertise sur un groupe d'agents, non soumis à un contrôle centralisé, devant être capables de travailler et d'agir dans un environnement commun et de résoudre les conflits éventuels.

A la différence de l'Intelligence Artificielle classique qui modélise le comportement intelligent d'un seul agent, L'IAD s'intéresse à des comportements intelligents qui sont le produit de l'activité *coopérative* de plusieurs agents. Le passage du comportement individuel

aux comportements collectifs est considéré non seulement comme une extension mais aussi comme un enrichissement de l'approche classique [Labidi et Lejouad, 2004]. En effet, l'IAD met l'accent sur l'aspect social et considère que l'interaction est le fondement de la constitution d'organisations capables de mener à bien des tâches complexes, et qui de points de vues multiples et parfois contradictoires, il était possible de faire *émerger* de nouvelles solutions [Kebair, 2009]. Par conséquent, l'intelligence n'est pas le fruit d'un individu mais d'un ensemble d'individus en interaction, d'où la réalisation des systèmes dits multi-agents.

3.4 Systèmes Multi-Agents

L'évolution de l'Intelligence Artificielle a provoqué plus d'intérêts vers des problèmes plus complexes, plus réalistes et de plus grande échelle, qui dépassent les capacités d'un agent individuel dont la rationalité est limitée par ses connaissances, ses ressources de calcul et ses perspectives.

Si le domaine d'un problème est particulièrement complexe, large ou imprévisible, une première alternative serait de le décomposer en sous problèmes, mais cette hypothèse n'est pas toujours réaliste car beaucoup de problème ne peuvent être partitionnés de cette manière [Labidi et Lejouad, 2004]. Une autre alternative raisonnable consiste donc à développer un nombre de composants fonctionnels, spécifiques et modulaires, qui sont chacun spécialisés dans la résolution d'un aspect particulier du problème [Zidi, 2006]. En effet, les problèmes réels impliquent des systèmes ouverts distribués. Un système ouvert est celui dont la structure est capable de changer de façon dynamique. Les composants d'un tel système ne sont pas connus d'avance, peuvent changer au cours du temps et peuvent être constitués d'entités hétérogènes implémentés par des personnes différentes, à des instants différents et avec des techniques et outils différents. Quand des problèmes interdépendants surgissent, ces composants doivent être capables d'interagir et de se coordonner pour gérer cette interdépendance, en se munissant de techniques basées sur la négociation ou la coopération. Pour ce faire, ils doivent être dotés de capacités de perception et d'action sur leur

environnement, comme ils doivent posséder une certaine autonomie de comportement, on parle alors d'agents et par conséquent de Systèmes Multi-Agents (SMA).

Les SMA se caractérisent alors par l'autonomie et l'intelligence des composants impliqués. Toutefois, un agent isolé ne dispose pas d'une vision globale de son environnement.

3.4.1 Concept d'agent dans un SMA

De nos jours, il est encore difficile de donner une définition commune au terme agent, vu qu'il est utilisé de différentes manières et par un large public venant de disciplines diversifiées en particulier celui de l'Intelligence Artificielle. Ainsi, le terme agent peut vraiment être vu comme un terme général qui couvre un ensemble d'autres types d'agents spécifiques et hétérogènes. Cependant, pendant longtemps la plupart des définitions s'accordaient avec celle proposée par Ferber [Ferber, 1995] : « *Un agent est une entité autonome, réelle ou abstraite, qui est capable d'agir sur elle-même et sur son environnement, qui, dans un univers multi-agents, peut communiquer avec d'autres agents, et dont le comportement est la conséquence de ses observations, de ses connaissances et des interactions avec les autres agents.* »

Actuellement, il semble que la façon la plus générale avec laquelle le terme agent est utilisée a été définie par Wooldridge et Jennings [Wooldridge et Jennings, 1994], où l'agent dit *agent intelligent* est référé à un matériel ou plus souvent à un système informatique à base de logiciels, ce qui coïncide bien avec l'approche que nous avons proposé dans le chapitre précédent, où nous avons modélisé le décideur individuel, qui représentera notre fameux agent dans le SMA que nous allons développer par la suite, à l'aide d'un programme informatique basé sur les SVM.

La modélisation de notre agent à base de SVM fait de ce dernier un agent de type cognitif. En effet, dans la littérature des SMA nous retrouvons trois grandes écoles de pensées relatives à la notion d'agent [Zidi, 2006], où nous distinguons les agents cognitifs, réactifs et

hybrides. Les systèmes d'agents cognitifs sont fondés sur la coopération d'agents autonomes capables à eux seuls d'effectuer des tâches relativement complexes [Labidi et Lejouad, 2004]. Un système cognitif est dit de forte granularité « *Coarse Grain* », il regroupe un petit nombre d'agents en interaction ayant tous une capacité de raisonnement propre basée sur les principes de croyance et d'intention avec leurs propres base de connaissances et leurs propres aptitudes à traiter et à réagir aux informations qui leur parviennent, à travers des plans explicites leurs permettant d'atteindre leurs buts [Weber, 2007]. Les systèmes Multi-Agents cognitifs sont par excellence une des meilleures réalisations du concept de l'intelligence collective originelle discuté plus haut.

3.4.1.1 Caractéristiques d'un agent cognitif [Labidi et Lejouad, 2004]

L'avancement des travaux en IAD et SMA a conduit les chercheurs à définir non seulement la notion d'agent mais aussi quelques-unes de ses caractéristiques :

- ***Intentionnalité*** : Un agent intentionnel est un agent guidé par ses buts. Une intention est la déclaration explicite des buts et des moyens d'y parvenir. Elle exprime donc la volonté d'un agent d'atteindre un but ou d'exécuter une action.
- ***Rationalité*** : Un agent rationnel est un agent qui suit le principe dit principe de rationalité : « *Si un agent sait qu'une de ses actions lui permet d'atteindre un de ses buts, il la sélectionne* »

Les agents rationnels disposent de critères d'évaluation de leurs actions et sélectionnent, selon ces critères, les meilleures actions qui leur permettent d'atteindre le but. De tels agents sont capables de justifier leurs décisions. La notion de rationalité se rapporte au comportement cognitif de l'agent. Ce terme qualifie l'utilisation des ressources par L'agent.

- ***Autonomie*** : les agents fonctionnent sans interventions extérieures et possèdent un certain type de contrôle sur leurs actions et leurs états internes.

- **Engagement** : La notion d'engagement est l'une des qualités essentielles des agents coopératifs. Un agent coopératif planifie ses actions par coordination et négociation avec les autres agents. En construisant un plan pour atteindre un but, l'agent se donne les moyens d'y parvenir et donc s'engage à accomplir les actions qui satisfont ce but: L'agent croit qu'il est en mesure d'exécuter tout le plan qu'il a élaboré, ce qui le conduit (ainsi que les autres agents) à agir en conséquence.
- **Adaptabilité** : Un agent adaptatif est un agent capable de contrôler ses aptitudes (communicationnelles, comportementales, etc.) selon l'agent avec qui il interagit. Un agent adaptatif est un agent d'un haut niveau de flexibilité.

En conclusion, nous appelons agent intelligent un agent qui comporte une ou plusieurs des caractéristiques présentées ci-dessus.

3.4.1.2 Fonctionnement d'un agent cognitif [Labidi et Lejouad, 2004; Noomen, 1998]

L'architecture de fonctionnement d'un agent cognitif est illustrée par la Figure 3.1 [Labidi et Lejouad, 2004]. Sur cette Figure, les ellipses représentent les processus mis en œuvre lors du fonctionnement de l'agent.

Les agents sont immergés dans un environnement dans lequel et avec lequel ils interagissent. D'où leur structure autour de trois fonctions principales : *percevoir, décider et agir*. Parmi les sous-fonctions importantes d'un agent, nous pouvons citer : la détection de conflits, la révision des croyances, la coopération (négociation, coordination), l'apprentissage, etc. Cependant, toutes les fonctionnalités ne sont pas représentées sur la Figure en question (Figure 3.1).

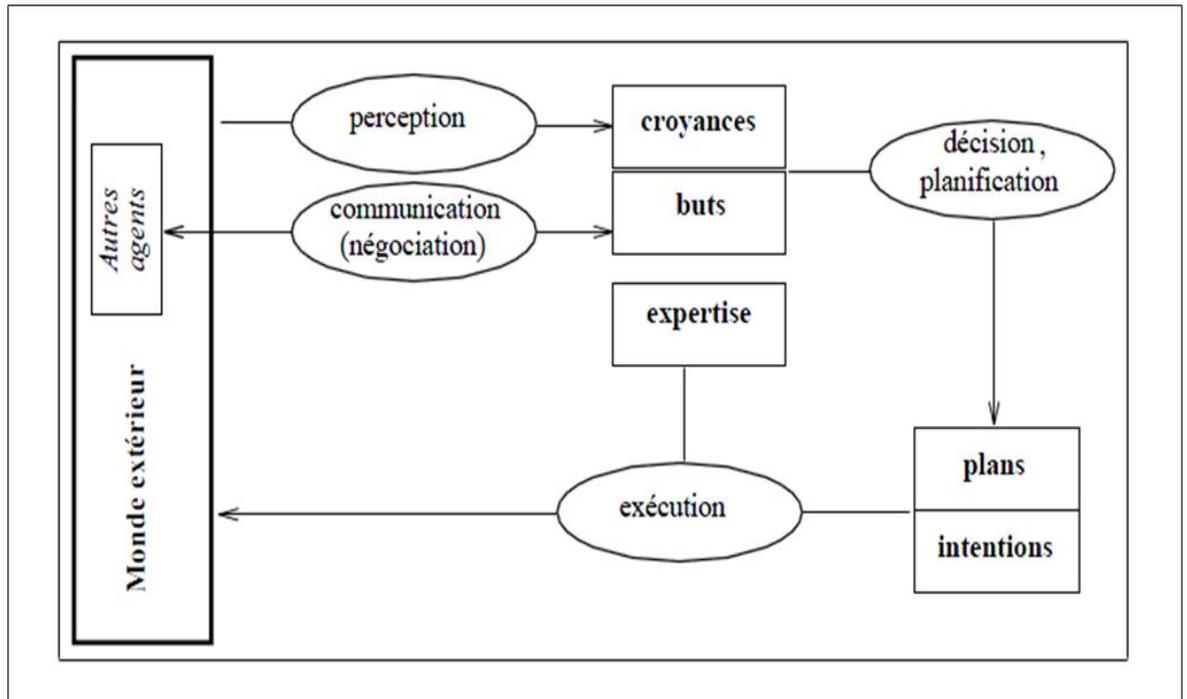


Figure 3.1 : Schématisation du fonctionnement d'un agent cognitif [Labidi et Lejouad, 2004].

Un agent a la possibilité d'acquérir des connaissances sur l'environnement externe (*perception*) comme il a aussi des capacités d'interaction avec les autres agents (*communication, négociation*). En fonction des connaissances et croyances dont il dispose et des buts qu'il se fixe suite à une perception ou à une interaction avec le monde extérieur, l'agent doit élaborer un plan d'action. Pour ce faire, il doit décider quel serait le but à retenir et à satisfaire en premier. Ensuite, il doit planifier en fonction de ce but et passer à l'exécution. Ces deux derniers processus doivent être alternés du fait du caractère dynamique des environnements multi-agents.

- **Perception** : Les connaissances ou les croyances d'un agent ont plusieurs origines :
 - Le savoir initial de l'agent ;
 - La perception de soi (perception proprioceptive) et du monde externe (perception extéroceptive) ;
 - La communication avec les autres agents.

Généralement, les informations issues de la perception et du savoir initial de l'agent sont considérées comme des connaissances certaines, puisqu'elles n'ont subi aucune mise à jour. Pour leur part, les connaissances provenant des autres agents sont considérées incertaines, puisqu'elles évoluent sans que l'agent en soit forcément informé. Nous pouvons, pour cela, associer à chaque connaissance son origine afin d'en évaluer la crédibilité et d'en permettre la vérification.

- **Communication** : La communication est essentielle pour un agent en ce qui concerne toute interaction avec d'autres. La communication permet à un agent d'enrichir ou de réviser ces connaissances, de remettre en cause ses plans d'action, de coopérer ou de négocier avec d'autres agents, etc.
- **Prise de décision** : Durant son exécution, un agent se fixe un certain nombre de buts, suite à ses observations et ses interactions avec le monde (perception, communication, négociation). Il se trouve donc confronté au problème de la sélection du but à satisfaire en premier et pour chaque but, de l'action qui permet de l'atteindre.

Face à de telles situations, l'agent analyse les différentes alternatives en termes d'utilité (quel avantage l'agent pourrait en retirer) et d'incertitude (quelle chance à l'action de fournir le résultat attendu). La prise de décision est une des caractéristiques des agents rationnels, l'agent tiendra compte de ses croyances pour faire son choix.

- **Planification** : La planification dans les systèmes d'Intelligence Artificielle classique repose sur l'hypothèse d'un univers statique (seules les actions du planificateur sont prises en compte), cet aspect est incompatible avec l'approche multi-agents. Les SMA, en revanche, offrent des possibilités de négociation autorisant une gestion locale des conflits et une planification dynamique.

Du fait de l'intervention d'autres agents, un plan peut être remis en cause. L'agent doit, pour cela, alterner planification et exécution et réviser des parties de son plan. La planification dans les SMA est une planification distribuée. En effet, il n'existe pas de plan global et chaque agent construit son propre plan en coordonnant avec les autres agents.

- **Exécution** : Finalement, l'exécution consiste à la mise en œuvre des plans générés par l'agent. L'exécution de ces plans est basée sur l'expertise de l'agent.

3.4.2 Caractéristiques des systèmes multi-agents

Les systèmes multi-agents sont des réseaux d'agents (solveurs) faiblement couplés, qui coopèrent pour résoudre des problèmes qui dépassent les capacités ou les connaissances individuelles de chaque agent. Les agents sont autonomes et peuvent être de nature hétérogène. En effet, comme l'illustre la Figure 3.2, un système multi-agent est composé d'un environnement d'objets qui sont manipulés par les agents, d'agents représentant les entités actives du système, d'interactions entre les agents et d'opérations qui permettent les perceptions et les différentes actions.

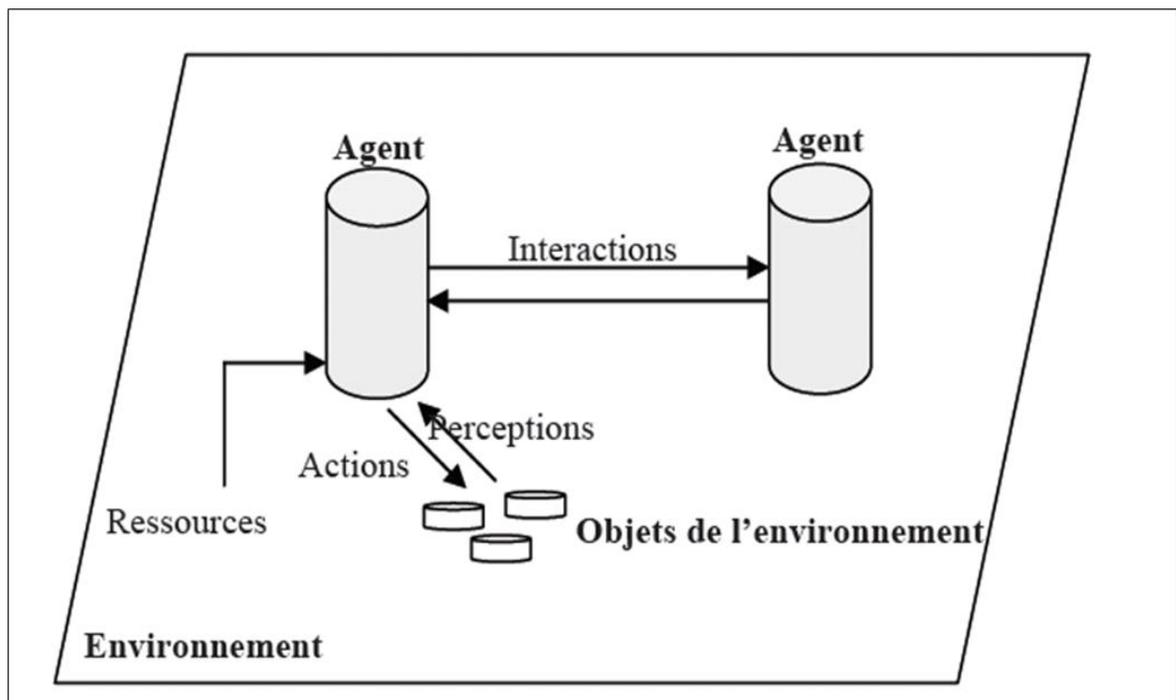


Figure 3.2 : Représentation d'un Système Multi-Agents [Zidi, 2006].

Ainsi, les agents d'un SMA, n'ayant qu'une vision locale de leur environnement, sont amenés à coopérer, et donc à communiquer, afin d'atteindre le but global du réseau. Aussi, il est nécessaire de définir la structure organisationnelle dans la société d'agents et d'y établir :

- la communication entre les agents,
- la coordination et la planification des tâches des agents,
- la négociation entre les agents pour la résolution des conflits.

En effet, dans tout type d'organisation, pour pouvoir résoudre avec cohérence les problèmes, les agents doivent communiquer entre eux, coordonner leurs activités et négocier dans les situations de conflits. Les conflits peuvent résulter d'une simple ressource limitée à des calculs plus complexes, où les agents tombent en désaccord, à cause des différences entre leurs domaines d'expertise. La coordination est nécessaire pour déterminer la structure organisationnelle d'un groupe d'agents et pour l'allocation des ressources et tâches. La négociation est nécessaire pour la détection et la résolution des conflits [Zidi, 2006].

3.4.2.1 Structures organisationnelles

Une structure organisationnelle est constituée d'un groupe d'agents, d'un ensemble d'activités effectuées par ces agents, d'un ensemble de connexions entre ces entités et d'un ensemble de buts ou de critères d'évaluation, avec lesquels les activités combinées des agents sont évaluées. Les structures organisationnelles imposent des contraintes sur la nature de la communication ou de la coordination.

Dans l'approche multi-agents, une organisation est une collection de rôles, de relations et de structures d'autorité qui gouvernent les comportements des agents impliqués [Kebair, 2009].

Généralement, Tous les SMA possèdent quelques-unes ou toutes ces caractéristiques et donc ont une certaine forme d'organisation, bien que ça peut être implicite et informel. En effet, beaucoup d'études récentes montrent que dans les entreprises performantes ainsi que la majorité des collectivités humaines coexistent deux types d'organisations [Kabachi, Vincent et Ouzrout, 1996] :

- celle qui est *formelle*, dont l'objectif est de guider ses membres à une aptitude de collaboration ;
- celle qui est *informelle*, où la collaboration entre individus est faite de façon spontanée.

De façon analogue, il existe dans la littérature de l'IAD deux approches pour la conception des organisations [Kabachi, Vincent et Ouzrout, 1996]:

- *l'approche statique* : selon cette approche, les liens d'autorités et de communication entre des agents, que l'on connaît tous a priori, sont complètement définis pendant la phase de conception du système. Ils servent à établir un moyen de contrôle global de la société, en la conduisant à un comportement censé résoudre un problème ;
- *l'approche dynamique* : selon cette approche, les liens d'autorités et de communication ne sont pas préétablis entre les agents. Ces liens sont créés de façon dynamique, que certains auteurs désigneraient par le terme émergence, lorsque les agents cherchent à résoudre un problème tout en essayant d'atteindre leurs propres buts.

3.4.2.2 Coopération

Une interaction est une mise en relation dynamique de deux ou plusieurs agents par le biais d'un ensemble d'actions réciproques. Pour un agent, interagir avec un autre constitue à la fois la source de sa puissance et l'origine de ses problèmes [Ferber, 1995].

La coopération est la forme générale de l'interaction. Plusieurs études en IAD se sont attachées à la définir, mais l'ensemble de ces études est plutôt hétéroclite et il règne une certaine confusion sur la signification et la mise en œuvre de la coopération [Kabachi, Vincent et Ouzrout, 1996]. La coopération doit-elle être définie en termes de buts, de stratégies ou de formes ? Désigne-t-elle un comportement ou une posture ?

Une des définitions les plus générales de la coopération est celle proposée par Galliers [Galliers, 1988] : « *La coopération existe entre deux agents par rapport à une proposition*

quand un agent reconnaît que l'autre agent a pour but la réalisation de cette proposition et qu'il s'est engagé à résoudre ce but comme un but commun, en relation avec les autres agents qui l'ont aussi pour but et ceci suite à un choix délibéré. Ainsi la coopération existe entre deux agents par rapport à une proposition, si au moins un des deux agents est engagé dans la réalisation de cette proposition en rapport avec les autres agents pour qui elle constitue un but et, ceci, suite au fait qu'ils croient avoir un but commun. »

Cette définition introduit trois conditions nécessaires à la coopération :

- reconnaître le but d'un autre agent ;
- s'impliquer dans sa résolution en l'adoptant comme son propre but de façon délibéré ;
- le considérer comme un but commun.

Dans le cadre d'une telle dynamique collective, un agent doit disposer en plus de la connaissance reflétant son degré d'implication dans cette dynamique (croyances, buts, intentions, engagements, modèle de soi et d'autrui) d'un certain nombre de compétences nécessaire pour la coopération, il doit pouvoir [Labidi et Lejouad, 2004]:

- mettre à jour le modèle du monde environnant ;
- intégrer des informations venant d'autres agents ;
- interrompre un plan pour aider d'autres agents.

Cependant, la coopération en général crée des conflits qui pourraient survenir entre agents lorsqu'il y a une relation de dépendance entre leurs actions. Pour éviter ces situations conflictuelles, il existe des outils pour la gestion des conflits entre agents à savoir : la coordination, la négociation et la communication.

3.4.2.3 Coordination [Zidi, 2006]

La coordination est un processus dans lequel les agents se sont engagés en vue d'assurer une communauté d'agents individuels agissant avec cohérence et harmonie. Les agents ont en fait besoin de la coordination pour empêcher les comportements chaotiques, pour être coordonnés de la même manière, puisqu'aucun agent ne possède une vue globale sur le système et parce qu'ils possèdent des capacités et expertises différentes, mais complémentaires.

La coordination comprend aussi l'allocation des tâches, qui consiste à affecter des responsabilités et des ressources nécessaires à la résolution de problèmes à un agent. Le créateur du système d'agents peut allouer toutes les tâches d'avance en engendrant ainsi une organisation de résolution des problèmes qui est non adaptable. Par contre, nous pouvons avoir une allocation dynamique et flexible des tâches.

La planification fait également partie de la coordination. Pour un agent, elle constitue un processus de construction d'une séquence d'actions en tenant compte seulement des objectifs, des capacités et des contraintes environnementales. La planification a pour rôle d'éviter les conflits. Elle peut être centralisée ou distribuée.

3.4.2.4 Négociation [Zidi, 2006]

La négociation est un processus de communication d'un groupe d'agents, qui a pour objectif d'atteindre un accord mutuellement accepté dans une affaire donnée. Elle est donc considérée comme une méthode de coordination et de résolution de conflits. Le processus de négociation peut induire des échanges d'informations, des relaxations des buts initiaux, des concessions mutuelles, des mensonges ou des menaces. Elle relève ainsi principalement de la recherche d'un consensus. Elle peut être compétitive ou coopérative.

La négociation compétitive est utilisée dans les situations où des agents d'intérêts différents tentent de faire un choix de groupe avec des alternatives bien définies. Les agents sont néanmoins compétitifs et non coopératifs.

Contrairement à la négociation compétitive, la négociation coopérative concerne les situations où les agents ont un but unique global envisagé pour le système. Il s'agit ainsi de systèmes distribués qui ont été conçus pour réaliser un seul objectif global. Ces agents sont aussi dits collaboratifs.

3.4.2.5 Communication [Zidi, 2006]

Les agents ont besoin de communiquer pour pouvoir interagir et échanger de l'information. Ils peuvent interagir soit en accomplissant des actions linguistiques (en communiquant entre eux), soit en accomplissant des actions non-linguistiques, qui modifient leur environnement.

En communiquant, les agents peuvent échanger des informations et coordonner leurs activités. Dans les SMA, deux stratégies principales ont été utilisées pour supporter la communication entre agents : les agents peuvent échanger des messages directement, ou ils peuvent accéder à une base de données partagée (appelée tableau noir ou "blackboard") dans laquelle les informations sont postées. Les communications sont à la base des interactions et de l'organisation sociale d'un SMA.

3.5 Modélisation *Blackboard* de la structure organisationnelle

Le concept d'organisation est un champ d'investigation très vaste. D'une manière générale, l'organisation est un modèle permettant aux agents de coordonner leurs actions au cours de la résolution d'une ou de plusieurs tâches. Elle définit d'une part, une structure comprenant un ensemble de rôles qui doivent être attribués aux agents et un ensemble de chemins de communication entre ces rôles. Elle définit d'autre part un régime de contrôle qui

dicte le comportement social des agents. Enfin, elle définit des processus de coordination entre agents afin d'assurer une réalisation cohérente des tâches [Guessoum, 2003].

3.5.1 La dualité du terme organisation

Jacques Ferber [Ferber, 1995] insiste sur la dualité entre l'aspect statique et l'aspect dynamique de l'organisation en soulignant que « les organisations constituent à la fois le support et la manière dont se passent les interrelations entre les agents, c'est-à-dire dont sont réparties les tâches, les informations, les ressources, et la coordination d'actions. De plus, et c'est ce qui rend ce terme si difficile à cerner, l'organisation est à la fois le processus d'élaboration d'une structure et le résultat même de ce processus. »

L'organisation est donc à la fois la manière dont un ensemble est constitué pour fonctionner et l'action d'organiser. Dans le premier sens, nous nous référons à un ordre, un régime, une structure ; dans le second sens, l'idée sous-jacente est celle d'un processus, d'une dynamique. L'organisation recouvre donc deux aspects : l'action et le résultat de l'action [Malville, 1999].

Cette dualité du terme montre qu'il existe toujours un aspect dynamique dans l'organisation: il n'existe, à proprement parler, pas d'organisations statiques. L'organisation est nécessairement dynamique et toujours en réorganisation de l'ensemble des entités et des liens qui unissent ces entités [Ferber, 1995].

3.5.2 Le choix d'une organisation

De ce qui précède, il s'avère que le choix d'une organisation est un problème majeur, spécialement pour les systèmes distribués puisqu'il conditionne les propriétés globales du système. En effet, Les travaux portant sur l'étude des organisations [Fox, 1981] montrent qu'il n'existe pas de « bonne » organisation dans un sens absolu. La pertinence d'une organisation est contextuelle. Une organisation appropriée pour un type de tâche peut être

inappropriée pour un autre. De la même façon, une organisation appropriée selon un critère peut ne pas l'être selon un autre.

D'une façon générale, le choix d'une organisation dépend des critères choisis pour évaluer la pertinence des organisations (les coûts, les temps de réponse, la robustesse) et des caractéristiques (l'incertitude et la complexité) des tâches et des informations dont disposent les agents. La complexité est définie par une demande excessive de rationalité (trop d'informations à gérer, tâches trop complexes ou trop nombreuses, trop de subalternes à coordonner). L'incertitude correspond à la différence entre l'information disponible et celle qui permettrait de prendre les meilleures décisions.

En soulignant la similitude entre les limites des capacités des agents et les limites de la rationalité humaine, M. Fox montre que le passage d'une organisation à une autre permet de répondre à un dépassement des limites des capacités des agents. La pertinence d'une organisation dépend de sa capacité à réduire la complexité et l'incertitude. En d'autres termes, une organisation pertinente est celle qui permet une adaptation dynamique aux changements en passant d'un modèle organisationnel à un autre de façon autonome afin de réduire la complexité et l'incertitude.

3.5.3 L'architecture à base de tableau noir [Ferber, 1995]

L'architecture à base de tableau noir est l'une des plus utilisée dans les systèmes multi-agents cognitifs symboliques et elle a donné lieu à une abondante littérature. Originellement développée dans le cadre de l'intelligence artificielle traditionnelle pour la reconnaissance de la parole, l'architecture de tableau noir s'est rapidement imposée en IAD comme une architecture suffisamment souple et puissante pour pouvoir implémenter les mécanismes de raisonnement et de calculs intervenant à l'intérieur des agents ainsi que les différentes structures organisationnelle qui unissent ces agents.

L'architecture de tableau noir présente de nombreux avantages dont, en tout premier lieu, une remarquable souplesse pour décrire des modules et articuler leur fonctionnement. Son

intérêt réside dans ce qu'elle est à la fois opportuniste et centralisée, les liaisons entre les modules (entre les agents) étant mutables. Elle est opportuniste au sens où les agents se déclenchent lorsque des configurations du tableau les intéressent, suite à des modifications provoquées par d'autres agents. Elle est centralisée par l'intermédiaire d'un module de contrôle qui ordonne l'ordre dans lequel les agents seront effectivement activés, en essayant de déterminer quelle est la meilleure action à effectuer compte tenu de l'état du système.

En tout état de cause, les avantages de ce type d'architecture sont indéniables : du fait de sa très grande plasticité, il est possible d'implémenter n'importe quelle structure d'agent en termes d'éléments de tableaux et d'agents. En particulier, toutes les autres architectures peuvent être modélisées en termes de tableaux noirs au prix parfois d'une certaine lenteur dans l'exécution. Le tableau noir se présente donc comme une sorte de "méta-architecture", c'est-à-dire une architecture pour implémenter des architectures.

Le modèle de tableau noir est fondé sur un découpage en modules (agents) indépendants qui ne communiquent directement aucunes informations, mais qui interagissent indirectement en partageant des informations. Ces modules travaillent sur un espace qui comprend tous les éléments nécessaires à la résolution d'un problème. L'architecture d'un système à base de tableau noir comprend trois sous-systèmes représentés sur la Figure 3.3 :

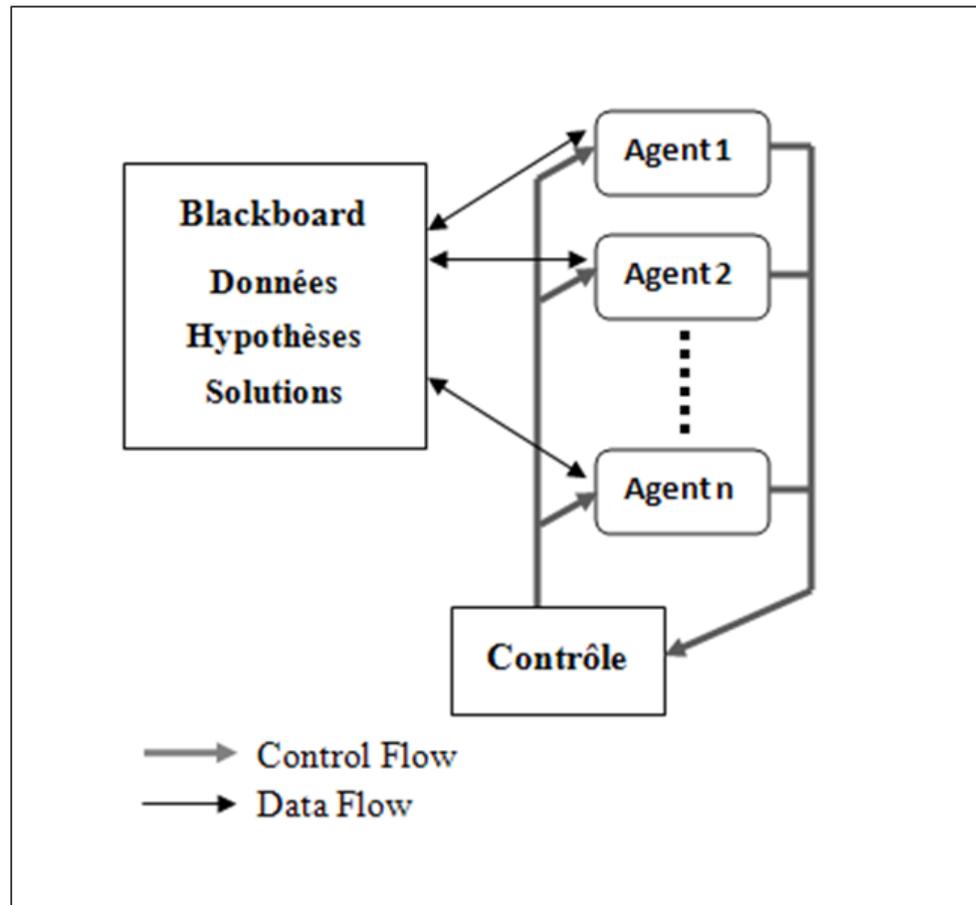


Figure 3.3 : Une architecture de système à base de tableau noire [Ferber, 1995].

- *Les agents.*
- *La base partagée* (le “tableau” proprement dit) qui comprend les états partiels d’un problème en cours de résolution, les hypothèses et les résultats intermédiaires et d’une manière générale toutes les informations que s’échangent les agents. Ces bases sont décomposées en hiérarchies qui structurent la modélisation du domaine d’application comme l’espace des hypothèses/solutions.
- *Le dispositif de contrôle* qui gère les conflits d’accès entre les agents, ces derniers intervenant de manière “opportuniste” c’est-à-dire sans être déclenchés effectivement par un système centralisé de contrôle. C’est cette partie qui a connu le plus de modifications au cours de l’évolution de ces architectures.

En effet, de nombreuses réalisations ont vu le jour, chacune présentant une vision différente du contrôle, la partie en fait la plus ardue d'un tableau noir. Le problème du contrôle dans un tableau noir revient à essayer de savoir ce qu'il convient de faire ensuite, c'est-à-dire de déterminer quel agent doit être déclenché.

Mais ce n'est qu'avec le système BB1 de B. Hayes-Roth que l'organisation du contrôle dans les tableaux noirs acquit ses lettres de noblesses, en considérant le contrôle comme un problème suffisamment important pour qu'il dispose de son propre tableau. L'architecture de BB1 comprend donc deux tableaux: le premier est destiné au traitement du problème du domaine et le second à la gestion du contrôle. De ce fait, il est possible de traiter l'activation des agents comme s'il s'agissait d'un problème indépendant du domaine d'application.

3.6 Les cartes Auto-organisatrices comme dispositifs de contrôle

En s'inspirant du système BB1 de B. Hayes-Roth, nous proposons, dans le cadre de notre travail, d'utiliser un réseau de neurones dit réseau de Kohonen (*cf. Annexe II*), ou encore carte auto-organisatrice de Kohonen. Le choix de ce type de réseau n'est pas fait au hasard, mais au contraire, il est justifié par deux propriétés majeures des cartes auto-organisatrices qui sont [Parizeau, 2004] :

- Les cartes auto-organisatrices sont des réseaux non supervisés avec un apprentissage compétitif qui permettent de modéliser l'espace des entrées avec des prototypes, où chaque neurone de ce réseau correspond aux prototypes. Dans notre cas, chaque neurone va correspondre à un agent.
- Elles permettent également de construire une carte à une ou deux dimensions qui définit les relations de voisinage entre les neurones ce qui permet de structurer l'espace d'entrées. Pour notre cas, les cartes auto-organisatrices offrent la possibilité de converger vers une structure d'organisation de ses neurones (agents) la plus appropriées via le processus d'apprentissage.

Comme nous l'avons précisé plus haut, ce modèle nécessite deux tableaux, le premier sera notre base de données du problème proprement dit qui servira à faire l'apprentissage des agents à base de SVM. Le second tableau sera destiné à la gestion du contrôle et à l'aboutissement de la structure organisationnelle des agents, ce dernier sera composé d'une part des résultats de simulation des différents agents, c'est-à-dire la décision de chaque agent après simulation, et de la décision réelle issue de la base de données initiale de l'autre part.

3.7 Conclusion

Le but de ce chapitre est de briser l'image mythique du décideur individuel, longtemps considéré comme étant le grand stratège clairvoyant et omniprésent. En effet, à cause des biais cognitifs auxquels il est exposé ainsi que l'hostilité et l'incertitude de l'environnement dans lequel il évolue, le décideur se voit souvent incapable de contrôler la crédibilité et la cohérence de ces décisions. En visant à maximiser le potentiel d'action et de liberté des décideurs dans un contexte où la plupart des challenges qu'ils rencontrent semblent trop complexes à résoudre, nous nous sommes intéressés à la décision collective, résultat d'une intelligence collective originelle dont le but est d'optimiser les propriétés émergentes *intérieures-subjectives* et *extérieures-objectives* des collectifs.

Par la suite, nous avons introduit les outils de simulation d'une telle architecture intelligente à savoir : les systèmes d'Intelligence Artificielle Distribuée et Multi-Agents qui représentent actuellement l'un des meilleurs moyens mis à la disposition des chercheurs pour modéliser et construire des systèmes ouverts et dont les entités (Agents) peuvent manipuler des concepts de haut niveau comme la rationalité et l'intentionnalité, où un modèle à base de « *blackboard* » avec un réseau de neurone de type cartes auto-organisatrices de Kohonen comme dispositif de contrôle est retenu.

La prochaine partie sera consacrée à la mise en œuvre de notre modèle avec comme premier objectif la démonstration de la souplesse et la robustesse de notre modèle de simulation de l'outil d'aide à la décision.

Partie 3:

*Application au dépistage
automatique de la
maladie de la Scoliose*

CHAPITRE 4

RECONSTRUCTIONS 3D DU SQUELETTE HUMAIN POUR UNE NOUVELLE APPROCHE DE DEPISTAGE DE LA MALADIE DE LA SCOLIOSE

4.1 Introduction

Afin de tester l'outil d'aide à la décision qu'on a préposé tout au long des chapitres précédents, on s'intéresse à un domaine très particulier qui est le domaine médical, plus exactement le domaine de l'orthopédie où le dépistage de la maladie de la scoliose fera l'objet de notre application.

Le choix de cette application se justifie par la nature radicale et incertaine de la décision thérapeutique. Celle-ci est fondamentalement radicale en raison d'une incomplétude des états de la nature. En effet, en matière médicale, les états de la nature ne sont jamais parfaitement connus, la liste des pathologies comme celle des thérapies est éphémère. Le praticien est dans l'impossibilité structurelle d'évaluer le risque maladie ou le risque guérison. L'environnement dans lequel le médecin prend sa décision est donc à information incomplète (incomplétude des sources d'information : diagnostic et études épidémiologiques) et à information imparfaite (asymétrie d'information). Face à son patient, le médecin est souvent dans l'ignorance totale malgré le poids de son savoir et sa capacité d'expertise. Pourtant, même dans ces situations, le médecin prend une décision. L'adage de la profession : *“guérir parfois, soulager souvent, consoler toujours”* prend ici tout son sens.

Pour faire face à cette radicalité inhérente du domaine, la science médicale a toujours été la première à adapter la technologie aux besoins vitaux des humains, en l'occurrence, la visualisation qui est devenue un élément clé pour comprendre et communiquer l'information médicale, ainsi que pour permettre de gérer des données de plus en plus complexes. En même temps, les avancées de la visualisation ont mené à une nouvelle génération d'outils qui permettent non seulement d'assister les praticiens dans le côté décisionnel mais aussi dans des interventions chirurgicales délicates et de très haute précision, autrement dit, très difficiles à réaliser.

4.2 Pathologie de la scoliose

La scoliose vient du mot grec *skolios* qui veut dire tortueux [Mary, 2004]. Le philosophe Hippocrate fut le premier à utiliser ce mot pour désigner toutes les déformations de la colonne vertébrale. Aujourd'hui, la définition est plus restreinte et se définit comme étant seulement une déformation de la colonne vertébrale telle que vue dans le plan frontal. Cependant, dans le milieu médical, la scoliose est reconnue comme une anomalie tridimensionnelle de l'axe du squelette (voir Figure 4.1). Elle est classifiée selon son étiologie en trois sous-groupes, soient la scoliose neuromusculaire, congénitale et idiopathique. Selon l'âge du patient au moment de l'apparition de la scoliose, elle est classifiée comme infantile (trois ans ou moins), juvénile (de quatre ans au début de la puberté), adolescente (après la puberté) ou adulte (après la maturité du squelette).

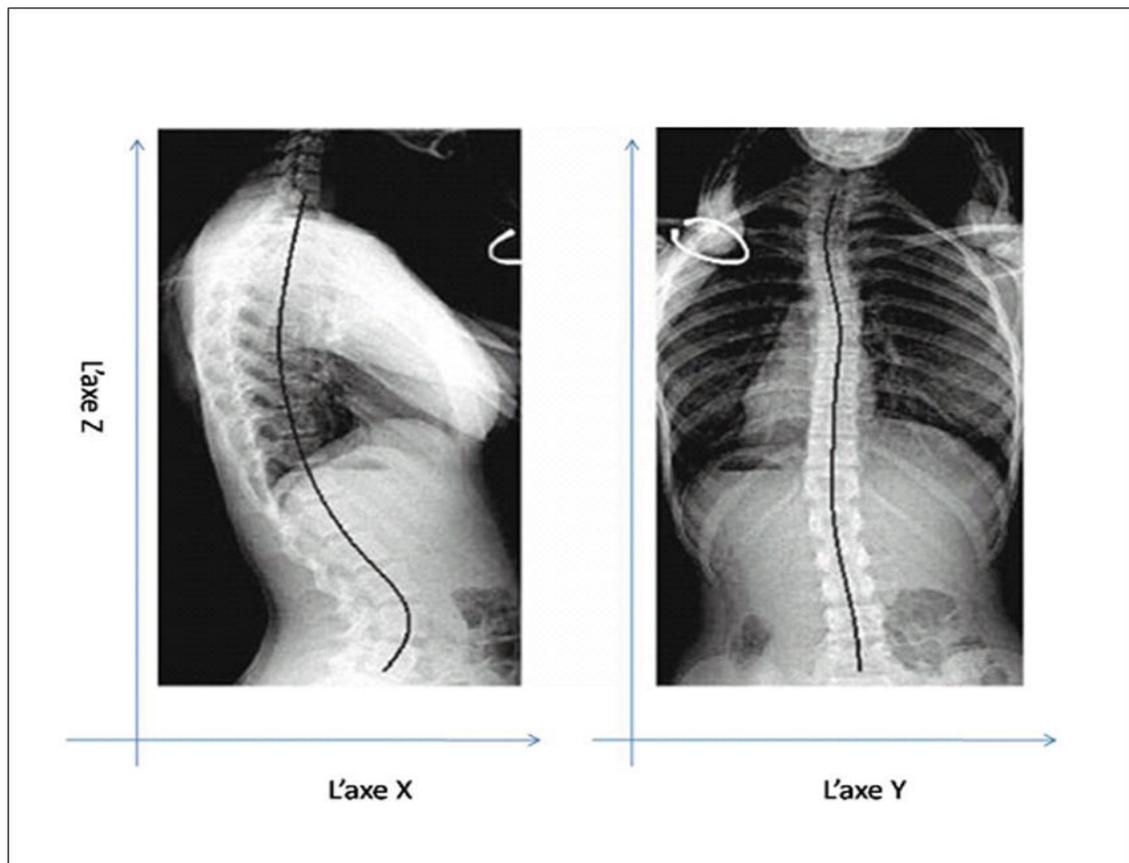


Figure 4.1 : Déformation scoliootique de la colonne vertébrale : l'axe X est orthogonal au plan frontal, l'axe Y est orthogonal au plan sagittal [Mary, 2004].

La scoliose dite idiopathique désigne les scolioses qui se produisent sans cause décelable et en dehors de tout contexte pathologique, elles représentent plus de 70 % des cas. Cette dernière est l'une des déformations orthopédiques les plus importantes chez les adolescents, particulièrement les jeunes filles qui sont 8 fois plus atteintes que les garçons. En effet, Selon les études rapportées en [Mary, 2004], la fréquence des scolioses idiopathiques enregistrées varient entre 1 à 5 individus pour une population de 1000 individus. Quelle que soit sa cause, la scoliose fait appel aux mêmes mesures thérapeutiques.

La scoliose idiopathique est une déformation potentiellement progressive qui affecte les adolescents durant leur phase de croissance et reste définitivement après qu'ils aient atteint l'âge adulte. Cette déformation de la colonne implique des asymétries de translation et d'angulation des vertèbres, de la cage thoracique et de la surface du dos qui imposent, sous l'effet de la croissance, des déformations plastiques progressives à l'ensemble des composants du rachis et du thorax [Mary, 2004]. Ces phénomènes impliquent donc la tridimensionnalité de la déformation scoliotique.

La scoliose présente plusieurs complications, en l'occurrence une déformation inesthétique qui a d'importantes répercussions psychologiques mais c'est surtout le retentissement fonctionnel qu'il convient de prévenir :

- Diminution de la capacité respiratoire avec insuffisance respiratoire et répercussion cardiaque (surtout pour des scolioses dorsales sévères) ;
- Douleurs rhumatismales.

Il est évident que toute scoliose sévère a un jour été modérée et c'est au stade de la scoliose modérée qu'il faut agir. En effet, tout le problème actuel de la scoliose est de dépister et de prévoir l'évolution d'une scoliose modérée afin de décider du traitement qui empêchera cette évolution naturelle. Généralement, la scoliose idiopathique se constitue avant l'âge de 6 ans, et même probablement lors de la première année. Un examen des radiographies pulmonaires de nourrissons devenus ensuite scoliotiques beaucoup plus tard, permet parfois de prévoir cette évolution sur les clichés.

Cependant, Il est impératif de distinguer deux affections totalement différentes [Mary, 2004] :

- La scoliose vraie ;
- Et l'attitude scoliotique beaucoup plus fréquente et beaucoup moins grave.

Il est important de ne pas se tromper car il est tout aussi grave d'étiqueter "attitude scoliotique" une véritable scoliose qui va continuer d'évoluer alors qu'un traitement bien conduit l'en aurait empêché que d'appeler "scoliose" une banale attitude scoliotique et d'imposer des traitements fastidieux inutiles.

4.3 Dépistage conventionnel de la scoliose

L'interrogatoire permet de recueillir les impressions du patient sur le trouble de la statique. Il se plaint de « se déformer, de se tasser, de perdre de la taille, d'avoir une hanche qui sort, d'être déséquilibré et de pencher en avant ou sur le côté, d'avoir les côtes qui entrent dans le bassin, d'avoir des difficultés d'habillement ». Chez l'enfant et l'adolescent, ce sont souvent les parents qui remarquent la déformation du tronc « il se tient mal » et le diagnostic est souvent fait tard car hélas, il n'existe plus de dépistage systématique et il y a un manque de formation des médecins [Marty, 2003].

4.3.1 Examen clinique de la scoliose

L'examen clinique doit être précis et suivre un protocole bien établi à la fois pour le diagnostic et pour le suivi de l'évolution. Ainsi, cet examen doit être chiffré et avoir une bonne reproductibilité. L'examen est fait dans les trois plans de l'espace, la scoliose étant une déformation tridimensionnelle.

La première chose à faire lorsqu'on examine le rachis d'un patient est de s'assurer qu'il s'agit bien d'un désordre de la statique rachidienne et non pas d'une anomalie sous-jacente.

C'est pour cette raison que l'examen clinique doit systématiquement commencer par l'étude de l'équilibre du bassin [Marty, 2003].

Une fois le bassin équilibré, le rachis lui-même peut être examiné. Plusieurs éléments doivent être systématiquement notés à chaque examen, car comme on le verra ultérieurement, l'évolutivité de la scoliose est un argument majeur pour proposer un traitement.

4.3.1.1 Examen dans le plan frontal

L'équilibre global du tronc est apprécié grâce au fil à plomb abaissé à partir de vertèbre C7, la 7^e vertèbre cervicale (la plus saillante). Le rachis est dit équilibré lorsque le fil à plomb passe dans le sillon fessier. Le déséquilibre éventuel s'apprécie par la distance entre ce pli et le fil à plomb. Cette mesure est essentielle mais ne signifie pas qu'il n'y a pas de scoliose ; une double courbure peut être parfaitement équilibrée [Mary, 2004].

La ligne des épaules est normalement horizontale ; un déséquilibre à ce niveau est également le signe d'une anomalie au niveau du tronc en dehors de toute pathologie scapulaire.

Pour pouvoir parler de scoliose, il faut qu'il y ait une rotation vertébrale, qui s'exprime par la ou les gibbosités selon le nombre de courbures. La présence d'une gibbosité (bosse) signe dans l'immense majorité des cas, l'existence d'une scoliose.

L'appréciation de l'asymétrie du pli de taille est un élément non mesurable, mais c'est un bon témoin de la présence d'une courbure au niveau lombaire ou thoraco-lombaire. Tous ces éléments étudient la déformation dans le plan frontal.

4.3.1.2 Examen dans le plan sagittal

L'aspect de la déformation dans le plan sagittal est essentiel. Un rachis équilibré dans ce plan fait que le fil à plomb abaissé du conduit auditif externe, passe en avant de T9, à

l'aplomb de L3, et enfin par le centre des têtes fémorales. Ceci est difficile à appréhender cliniquement. On se contente alors de mesurer les flèches de profil en C7, T9, L3, et au niveau du sacrum, Théoriquement sur un rachis normal, les flèches sont de 30 mm en C7, 0 mm en T7, 30 mm en L3 et 0 mm au niveau de sacrum. Ce sont des valeurs indicatives, car contrairement au plan frontal où le rachis normal est rectiligne, les courbes dans le plan sagittal sont beaucoup plus variables. L'appréciation subjective de l'aspect du dos (dos plat, dos creux, hypercyphose thoracique) est moins précise, mais très utile en pratique [Mary, 2004].

Tous les éléments que nous venons de voir donnent une vision globale de l'aspect de la déformation qui permet un suivi. Mais ces mesures sont statiques. Le caractère dynamique du rachis est aussi à appréhender.

4.3.2 Examen radiographique

Il complète l'examen clinique, et doit être réalisé de manière rigoureuse. Il comprend un cliché du rachis en entier de face, debout, bassin horizontal. Des clichés faits en décubitus sont également intéressants en cas d'inégalité de longueur, ou pour apprécier l'importance de la composante statique et dynamique de la déformation. L'essentiel pour pouvoir assurer un suivi correct de la scoliose, est de demander à chaque fois des clichés dans la même position pour pouvoir comparer [Marty, 2003]. Un cliché de profil est également réalisé, debout, les membres supérieurs à l'horizontale, reposant sur un appui. On y étudie les courbures de profil sur chaque secteur.

En cas de scoliose, la radiographie permet de mesurer l'angle de Cobb qui est le principal critère d'évaluation de la scoliose. Il est utilisé pour déterminer l'ampleur de la courbe d'une scoliose afin de classer la gravité de cette dernière [Mary, 2004].

L'angle de Cobb se mesure sur une radiographie de face sur laquelle on définit des vertèbres stratégiques : Les vertèbres limites pour une courbure sont celles qui sont les plus inclinées par rapport à l'horizontale. Il se mesure entre le plateau supérieur de la vertèbre

limite supérieure et le plateau inférieur de la vertèbre limite inférieure. Cet angle est mesuré sur chaque courbure (voir la Figure 4.2).

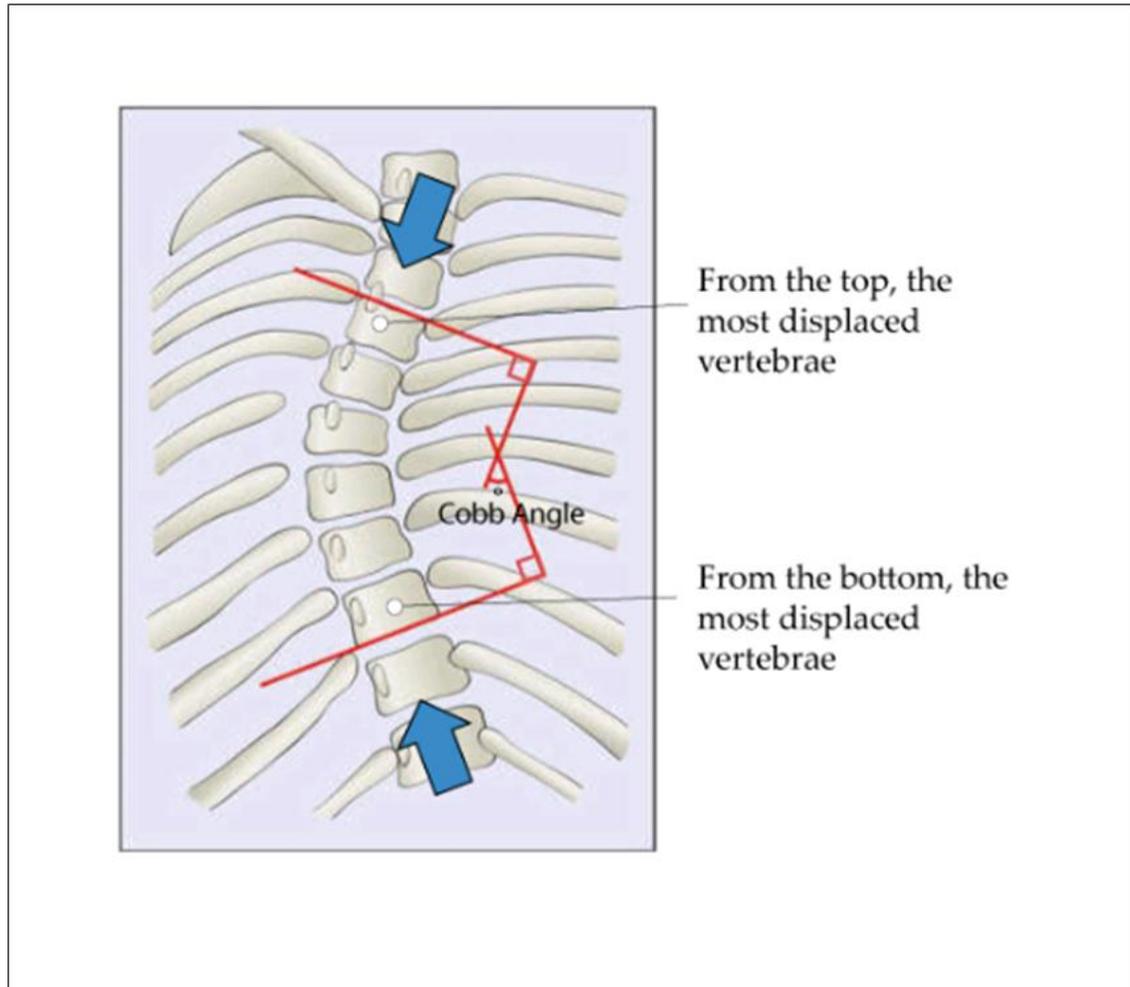


Figure 4.2 : Mesure de l'angle de Cobb [Mary, 2004].

L'angle de Cobb est une mesure de la courbure de la colonne vertébrale en degré, qui aide le médecin à déterminer quel type de traitement est nécessaire :

- Un angle de Cobb de 10 degrés est considéré comme l'angulation minimale pour définir une scoliose.

- Une courbe de 10 à 15 degrés ne nécessite généralement pas de traitement autre que des contrôles réguliers avec le chirurgien orthopédiste jusqu'à ce que le patient atteigne l'âge de puberté.
- Si la courbe de la scoliose est de 20 à 30 degrés, le médecin orthopédiste généralement prescrit un corset pour éviter le développement de la déformation de la colonne vertébrale.
- Si l'angle de Cobb est de 40 ou 50 degrés ou plus, la chirurgie peut être nécessaire pour corriger la courbe.

De ce qui précède, il s'avère que le problème actuel de la scoliose est de pouvoir la dépister avant qu'elle atteigne un angle de 30 degrés où généralement la situation deviendra très grave et évoluera vers des angles très importants.

4.4 Dépistage de la scoliose à l'aide de reconstruction 3D du rachis

L'angle de Cobb est mesuré à partir de la projection de la colonne vertébrale dans un plan 2D ce qui est insuffisant pour déterminer objectivement l'angle et qui ne nous donne pas l'information complète sur la forme de la colonne vertébrale dans l'espace.

En effet, la radiographie conventionnelle est la modalité d'imagerie couramment utilisée en milieu clinique pour le diagnostic des pathologies du rachis. Néanmoins, le cliché radiologique ne représente qu'une projection de la zone anatomique étudiée sur un plan 2D et le résultat de cette projection est sensible au positionnement du patient lors de la prise de la radiographie. Les angles sont donc calculés avec des biais de projection et leurs valeurs varient en fonction de la position du patient. De plus, cette modalité ne permet pas d'évaluer les rotations et torsions dans le plan transversal.

Dans de rares cas, un examen par tomodensitométrie (CT-Scan) ou par résonance magnétique (IRM) peut être indiqué afin d'évaluer les troubles de torsions et de rotations des structures osseuses. Cependant, la plupart de ces systèmes d'imagerie proposent une

acquisition en position couchée et non en charge, ce qui ne permet pas d'évaluer les phénomènes de compensation interarticulaire associés à la pathologie. De plus, les doses d'irradiation associées au CT-Scan sont encore élevées ce qui limite la réalisation des examens de manière fréquente. L'IRM peut être une alternative au CT-Scan, mais elle est très coûteuse et plus dédiée à l'étude des tissus mous.

Une nouvelle alternative est l'imageur biplan à détecteurs gazeux nommé « *EOS imaging* » [Feydy et al., 2008]. Ce système donne la possibilité d'obtenir des clichés biplans du corps entier, en position debout et avec de basses doses d'irradiation. Cette technologie permet, en utilisant des algorithmes de reconstruction 3D, d'avoir une reconstruction très précise et en taille réelle du squelette entier du patient.

4.4.1 L'imageur biplan à détecteurs gazeux

Cette technologie s'appuie sur les travaux du Pr. Charpak (Prix Nobel de Physique en 1992) qui inventa la chambre à fils : il s'agit d'une chambre remplie de Xénon (un gaz rare) et contenant de minuscules filaments de tungstène [Feydy et al., 2008].

4.4.1.1 Principe physique

Tandis que l'émetteur de rayons X descend en glissant sur un rail derrière le patient, la chambre à fils descend simultanément afin de recevoir les rayons X ayant été atténués par les différents organes et structures osseuses du patients. De ce fait, l'organe à radiographier est balayé verticalement par des rayons X qui "suivent" les filaments de tungstène à la sortie du patient. Au fur et à mesure de leur progression sur ces filaments, les rayons X arrachent des électrons (plus ou moins selon leur intensité) aux molécules de Xénon et les font migrer le long des filaments de tungstène. Les électrons stockés sur les différents fils de la chambre forment des tensions mesurées et amplifiées.

C'est cette amplification, aussi appelée "phénomène d'avalanche", qui va permettre, à partir d'une faible information (un seul photon X), d'obtenir un signal très important, ce qui

contribue à pouvoir réduire la dose d'entrée. Un ordinateur va ensuite analyser les différentes tensions mesurées en fonction de la position des différents filaments et du temps, afin de reconstituer les structures en présence et l'image de l'organe traversé. Il y a donc besoin d'une émission moins forte de rayons X au cours de l'examen.

Basé sur cette découverte ainsi que sur l'idée supplémentaire de positionner deux ensembles sources-détecteurs à 90° pour intégrer les techniques de reconstruction 3D, l'imageur EOS biplan à détecteurs gazeux est né (voir Figure 4.3).

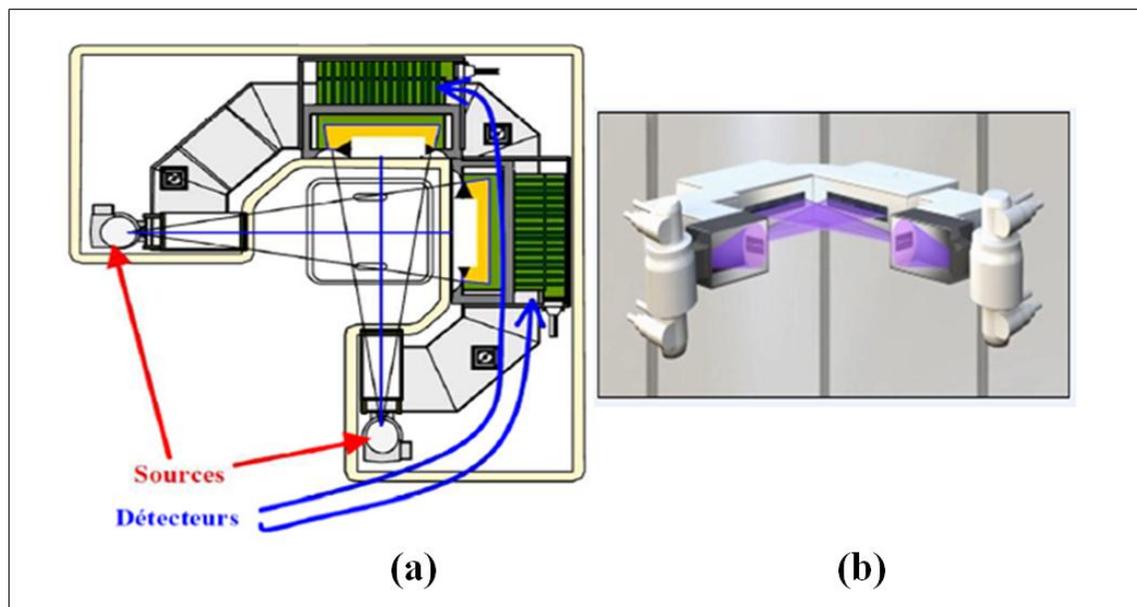


Figure 4.3 : Imageur biplan à détecteurs gazeux : (a) Principe de l'imageur vue de dessus; (b) principe de l'imageur vue de face [Feydy et al., 2008].

4.4.1.2 Spécifications techniques de l'imageur biplan

L'imageur biplan est un système de radiologie numérique, réalisant simultanément, par balayage vertical, deux vues orthogonales d'incidences 0° et 90° (Figure 4.4). Le mode de production des rayons X est classique et le générateur permet de faire des acquisitions avec une énergie émise de 40 à 140 keV (kilo électronvolts).

L'ensemble du système peut réaliser un balayage du corps entier en 20 secondes. Les données recueillies sont stockées sous forme numérique autorisant l'affichage sur écran et le traitement de l'image. La résolution finale (taille du pixel) est de $254 \mu\text{m}$. La grande sensibilité du détecteur, la capacité d'ajuster le gain, ainsi que l'absence de rayonnement diffusé, donnent au système de détection une dynamique réelle très importante qui assure une résolution en contraste élevée (30000 niveaux de gris), ce qui permet l'analyse simultanée des tissus osseux et de certains tissus mous [Feydy et al., 2008].



Figure 4.4 : Exemple d'une radiographie de face et de profil [Feydy et al., 2008].

4.4.1.3 Avantages de l'imageur biplan

Les avantages de cette nouvelle technologie sont nombreux :

- Acquisition d'images grand format, tête aux pieds, (jusqu'à 1.75 m) donnant une information globale du système ostéo-articulaire.
- Le patient est en position debout (en charge) dans la cabine - donc en position fonctionnelle.
- L'acquisition simultanée des deux clichés, face et profil, résout la problématique du mouvement du patient que l'on rencontre lorsque les deux radiographies sont réalisées successivement à partir d'une unique source conventionnelle.
- Les ensembles sources-détecteurs sont fixes, l'environnement 3D du système est donc parfaitement connu et ne nécessite pas de calibration à chaque nouvelle acquisition.
- Les paramètres d'acquisition proposés par défaut peuvent être ajustés selon le type de cliché et l'anatomie du patient.
- La collimation permet d'obtenir un cliché localisé ou un cliché grand format, en position de charge ou assise.
- L'ajustement automatique du gain permet de garantir un contraste d'image maximal, quels que soient les kV appliqués.
- Les radiographies obtenues sont numériques, ce qui permet l'accès à des traitements d'image performants tels que le renforcement des contrastes et de la luminosité, l'analyse de la texture, le zoom, etc.
- La possibilité de stockage et d'archivage numérique, permettant ainsi le partage et l'accès rapide aux examens.

La dose administrée au patient est 8 à 10 fois inférieure à celle d'une radiographie conventionnelle sur film (Figure 4.5), 800 à 1000 fois inférieure à celle du CT-Scan et de 2.9 à 9.2 fois inférieure à celle d'une radiographie numérique.

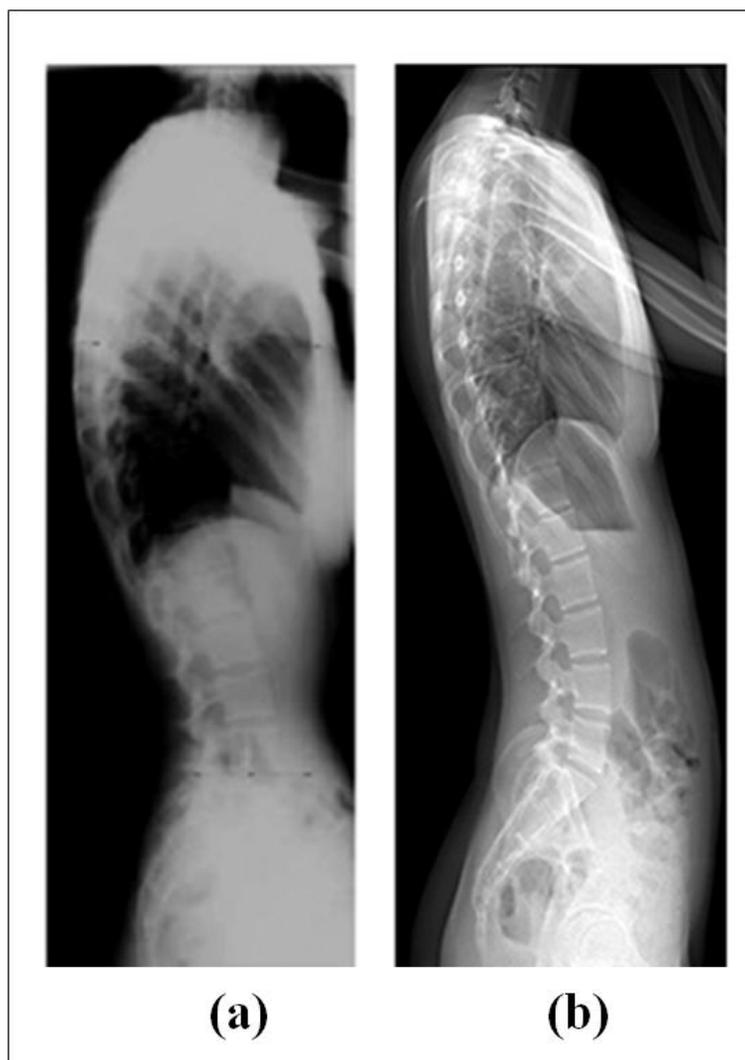


Figure 4.5 : Comparaison entre une image radiographique classique en (a) et une image radiographique obtenue avec l'imageur biplan en (b) [Feydy et al., 2008].

Ce dernier point est un avantage de taille. En effet, une étude à dernièrement ravivée les inquiétudes concernant les effets à long terme potentiels des rayonnements émis lors de l'acquisition d'images diagnostiques ; cette étude suggère que l'exposition à de multiples examens radiologiques dans l'enfance et l'adolescence pourrait majorer le risque de développer un cancer jusqu'à une fois sur mille [Feydy et al., 2008].

4.4.2 Reconstruction 3D du rachis

Les méthodes de modélisation 3D s'appuient sur la reconstruction 3D de points anatomiques 2D Stéréo-Correspondants, c'est-à-dire identifiés dans chacune des deux radiographies. Une fois les points repérés, leurs positions 3D dans l'espace sont calculées [Bertrand, 2005; Humbert et al., 2009; Meghoufel, 2004] (voir Figures 4.6 et 4.7). Le principe de base est l'utilisation de deux radiographies orthogonales, d'un environnement calibré afin que les données géométriques et positionnelles de l'environnement (source et plan de projection) soient connues, l'extraction manuelle ou automatique de l'information pertinente, recalage d'un modèle prédéfini (dit modèle générique ou morpho-réaliste) et, enfin, l'obtention d'un modèle tridimensionnel personnalisé. L'utilisation de l'imageur biplan qui donne la possibilité d'obtenir des clichés du corps entier en position fonctionnelle et avec de basses doses d'irradiation, a permis d'obtenir des modèles 3D personnalisés précis et de calculer les indices cliniques associés.

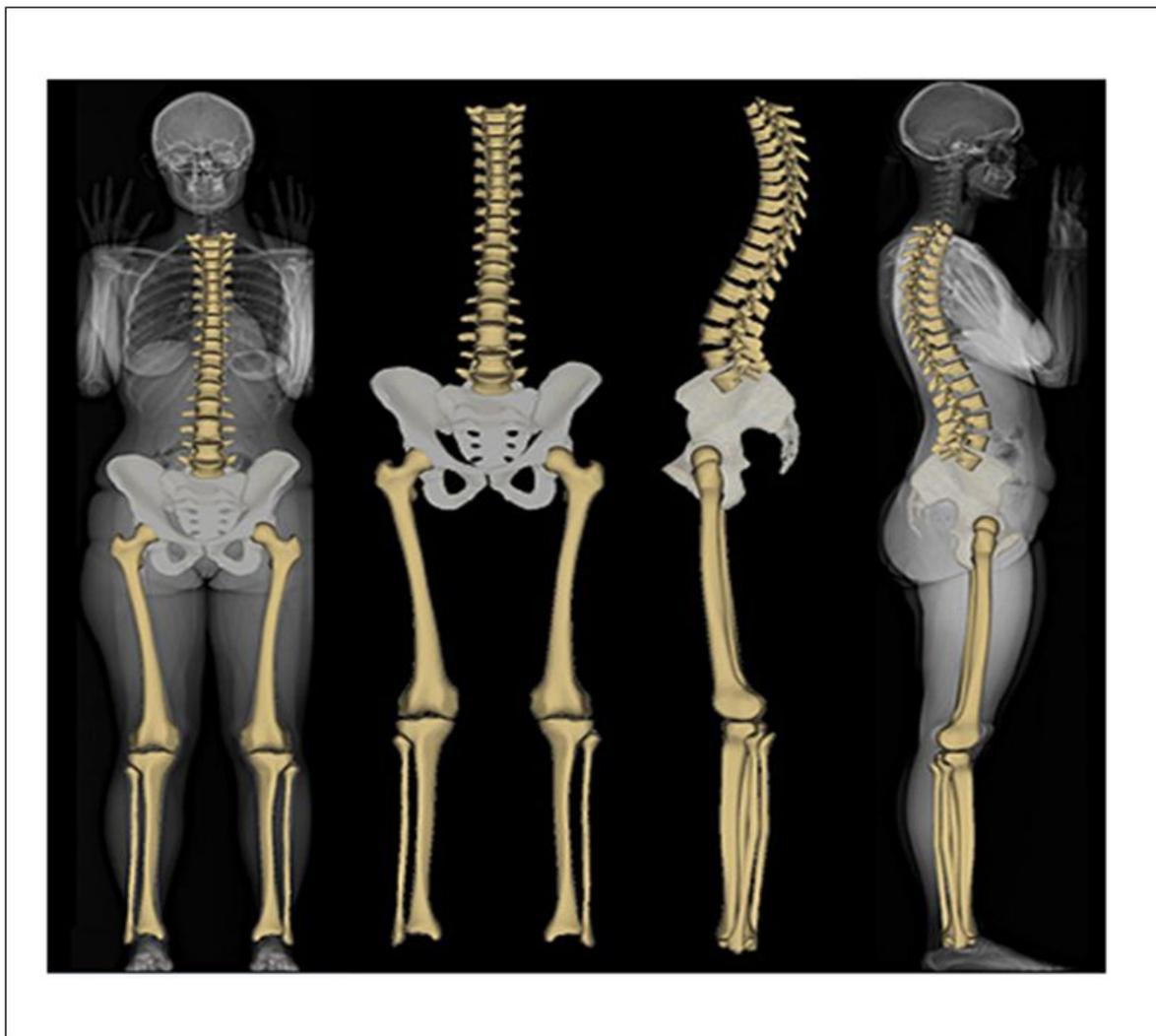


Figure 4.6 : Reconstruction 3D du squelette à partir de radiographies biplanes [Feydy et al., 2008].

En ce qui concerne la scoliose, l'indice clinique principal est l'angle de Cobb qui se mesure généralement à partir d'une radiographie de face 2D. Cependant, cette méthode de calcul présente l'inconvénient de ne pas permettre l'évaluation des rotations et torsions du rachis dans le plan transversal. Afin d'y remédier, beaucoup d'auteurs se sont intéressés aux techniques de reconstruction 3D du rachis à partir de radiographies biplanes, en l'occurrence, les travaux d'Humbert et *al.* [2009], qui sont d'ores et déjà utilisées en clinique routinière avec une précision proche du millimètre et avec un temps de reconstruction inférieur à 10

minutes. Ces techniques ont permis un calcul très précis de l'angle de Cobb et cela dans les trois dimensions de l'espace.

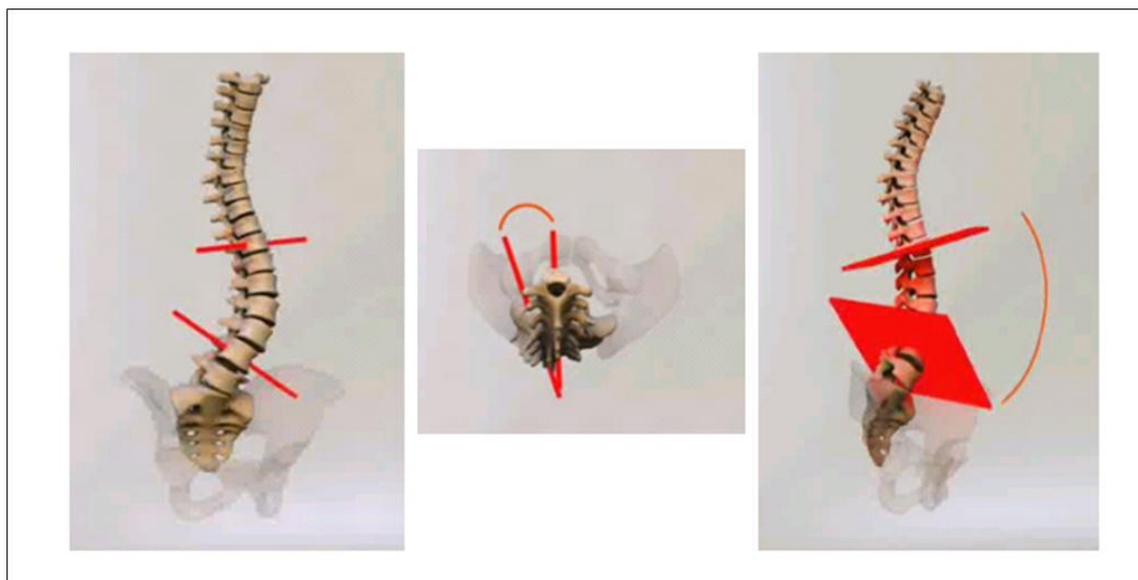


Figure 4.7 : Mesure de l'angle de Cobb à partir d'une reconstruction 3D du rachis [Feydy et al., 2008].

4.5 Problématique et objectifs de l'application

L'utilisation de l'imageur biplan donne la possibilité d'obtenir des clichés du corps entier en position fonctionnelle ce qui permet l'obtention de modèles 3D précis et le calcul des indices cliniques associés ainsi que le stockage et l'archivage numérique de ces derniers. Ces nouvelles possibilités ont permis la construction de bases de données consistantes, chose qui était auparavant trop coûteuse, voire impossible.

Le but de notre application est de proposer une nouvelle technique, moins lourde, de dépistage de la scoliose. Cette technique consiste à utiliser le système multi-agents à base de SVM comme outil d'aide à la décision qui va exploiter les bases de données générées par l'imageur biplan afin de discriminer automatiquement les colonnes vertébrales normales des scoliotiques.

4.6 Conclusion

Dans le prochain chapitre, nous allons présenter en détaille la méthodologie de travail que nous avons adopté afin d'aboutir à notre outil d'aide à la décision à base de Système Multi-Agents cognitifs que nous allons appliquer pour la discrimination automatique des personnes Scoliotiques des normales. Les résultats obtenus avec notre modèle vont être discuté afin de tirer les points forts et faibles de notre approche.

CHAPITRE 5

MÉTHODOLOGIE : PRÉTRAITEMENT DES DONNÉES ET SIMULATION DE LA PRISE DE DÉCISION

5.1 Introduction

Tout au long de ce chapitre, nous allons présenter la méthodologie adoptée afin d'aboutir à notre outil d'aide à la décision qui doit être en mesure de simuler la décision si un patient est scoliotique ou non.

Les principales étapes de notre méthodologie se résument comme suit :

- Présentation de la base des données;
- Prétraitement de la base des données ;
- Extraction des caractéristiques principales susceptibles à discriminer les colonnes normales des scoliotiques;
- Simulation de la prise de décision individuelle par les SVM ;
- Simulation de la prise de décision collective par le Système Multi-agents à base de SVM.
- Discussion des résultats

5.2 Base de données

Comme nous l'avons précisé plus haut, un des avantages de l'imageur biplan est que les radiographies obtenues sont numériques ce qui donne la possibilité de stockage et d'archivage numérique de ces dernières. Cet avantage a permis de construire des bases de données consistantes, chose qui était auparavant trop coûteuse, voire impossible.

La base de données utilisée dans notre étude a été collectionnée en quelques années dans des laboratoires spécialisés pour de divers projets de recherche [Champlain, 2004]. Elle est issue avec des radiographies numériques enregistrées par l'imageur biplan à détecteurs gazeux. L'obtention de cette base de données a été réalisée par internet à partir du site de l'École de Technologie Supérieure de Montréal, sous respect des propriétés intellectuelles des détenteurs des droits.

Cette base de données contient un total de 160 modèles de colonne vertébrale identifiées par des cliniciens; 91 colonnes vertébrales normales (angle de Cobb \leq 20 degrés) et 69 colonnes vertébrales scoliotiques (angle de Cobb $>$ 20 degrés).

Tableau 5.1 : Tableau récapitulatif de la base de données [Champlain, 2004].

Nombre Total des Cas	Normale	Scoliotique
160	91	69

Un cas de notre base de données est une colonne vertébrale (rachis) identifiée par 17 points anatomiques 3D précis représentés par leurs coordonnées dans l'espace. Ces points anatomiques sont les vertèbres Thoraciques de T1 à T12 plus les vertèbres lombaires de L1 à L5. Ces vertèbres sont celles susceptibles d'être atteintes des déformations scoliotiques (voir Figure 5.1).

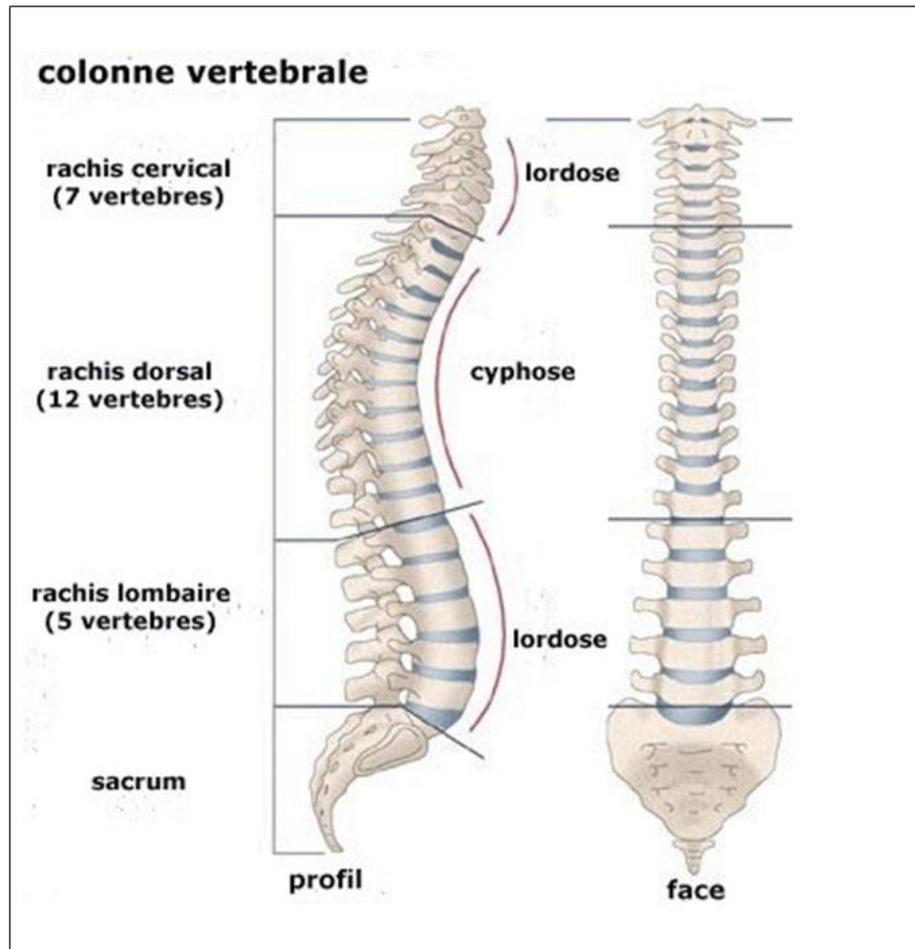


Figure 5.1 : L'ensemble de vertèbres sur le rachis [Wikipédia, 2011].

La base de données, comme nous l'avons présentée plus haut, est une matrice 3D à laquelle nous devons rajouter l'étiquette : N pour les cas normaux ou S pour les cas scoliotiques. Cependant, cette représentation matricielle est inexploitable pendant la phase de programmation. Pour y remédier, nous proposons de passer à une représentation matricielle, proposée par MATLAB, dite *dataset* [Duin et al.,2007], destinée essentiellement pour des problèmes de classification ce qui coïncide avec l'objet de notre application.

Tout simplement, *dataset* est une matrice de taille $m \times (k+1)$, le nombre de ligne m représente l'ensemble des objets à classer, dans notre cas les objets correspondent aux colonnes vertébrales. Chaque objet est décrit par un vecteur de caractéristiques composé de k descripteurs, dans notre cas les descripteurs sont les coordonnées cartésiennes des 17 points

anatomiques de chaque vertèbre ce qui nous donne un total de 51 descripteurs : 17 coordonnées sur l'axe des X , 17 coordonnées sur l'axe des Y et 17 coordonnées sur l'axe des Z , plus la colonne d'étiquette qui sert à préciser la classe d'appartenance de chaque colonne vertébrale.

Tableau 5.2 : Représentation *dataset* de la base de données.

	Descripteurs			
	Étiquette	17 Coordonnées selon les X	17 Coordonnées selon les Y	17 Coordonnées selon les Z
Cas de 1 à 91	Normale	✓	✓	✓
Cas de 92 à 160	Scoliotique	✓	✓	✓

5.3 Prétraitement de la base des données

Le prétraitement est une étape indispensable avant chaque prise de décision. En effet, le prétraitement fait partie des mécanismes qui permettent de réduire l'entropie riche de l'information d'entrée, souvent trop complexe et entachée de différents biais. Pendant cette phase, l'information est traitée, restructurée et filtrée afin d'assurer l'extraction de l'information pertinente qui sera par la suite la base de la prise de décision.

Dans le cas de notre application, les colonnes vertébrales de la base de données représentent des personnes de tailles différentes ce qui implique que leurs formes 3D (grandeurs géométriques) ne sont pas dans le même repère. Cette représentation initiale des colonnes vertébrales peut nuire à toute comparaison entre elles comme le montre la Figure 5.2.

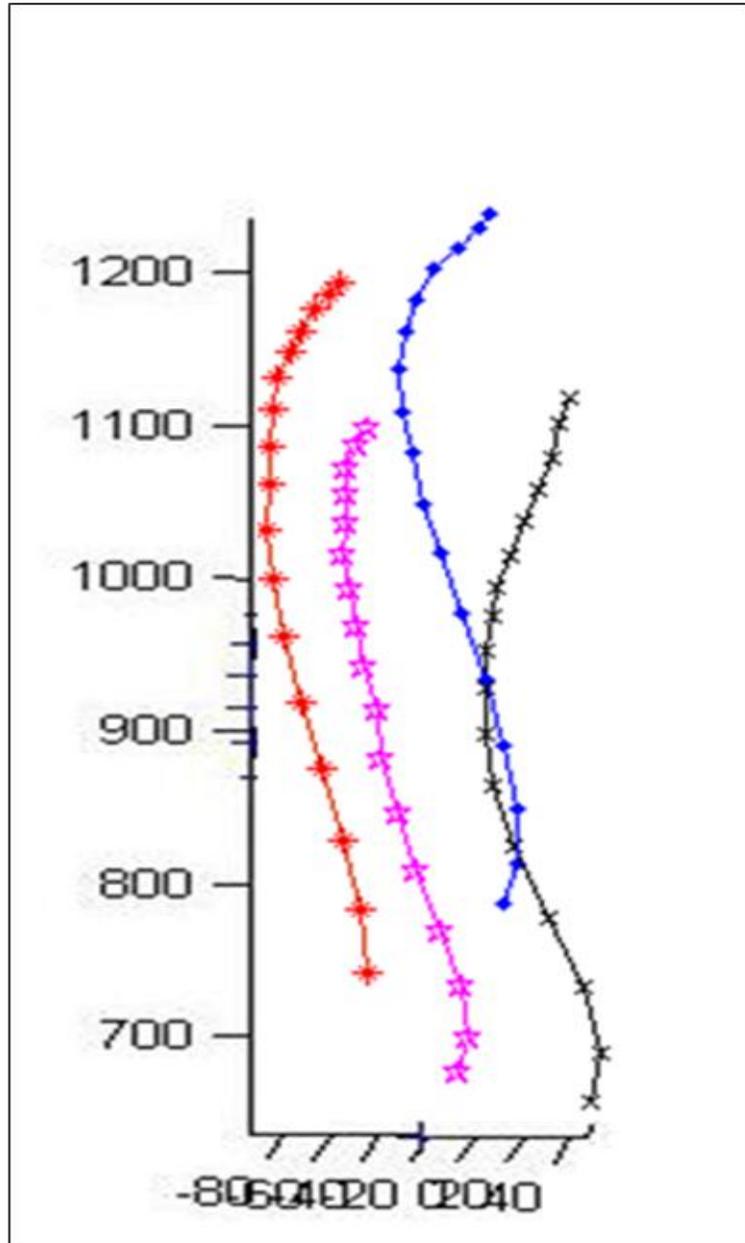


Figure 5.2 : Représentation de quatre colonnes vertébrales de la base de données dans l'espace 3D.

Une analyse au préalable s'impose dans cette étude. Étant donné que représenter une colonne vertébrale en 3D par des coordonnées cartésiennes (les 17 points) est une étape relativement simple, nous pouvons normaliser les données à pieds d'égalité pour pouvoir les traiter avec une analyse identique et pour pouvoir les comparer dans un espace unique.

Le prétraitement mathématique que nous proposons se résume en les étapes suivantes :

1. normalisation des longueurs de toutes les colonnes vertébrales selon l'axe des Z;
2. identification du centre de masse;
3. recalage des colonnes vertébrales.

Ces étapes du prétraitement nous permettent finalement d'extraire les caractéristiques principales utiles pour la classification. Une description détaillée de chaque étape est décrites dans les sections suivantes.

5.3.1 Normalisation selon l'axe des Z

Dans la vie réelle la scoliose peut toucher plusieurs personnes de déférent âge et de déférente taille cette déférence peut introduire des erreurs lors de la comparaison entre la colonne vertébrale d'entrée et les autre colonnes vertébrales de différentes classes. Pour cette raison une normalisation selon l'axe des Z est nécessaire comme première étape du prétraitement. Cette normalisation consiste à redéfinir un nouvel intervalle pour la variable z dont les bornes sont $[0,1]$, tout en conservant la forme globale de la colonne vertébrale. Pour cette fin nous utilisons un facteur de normalisation commun entre les trois axes X, Y, Z (Voir Figure 5.3).

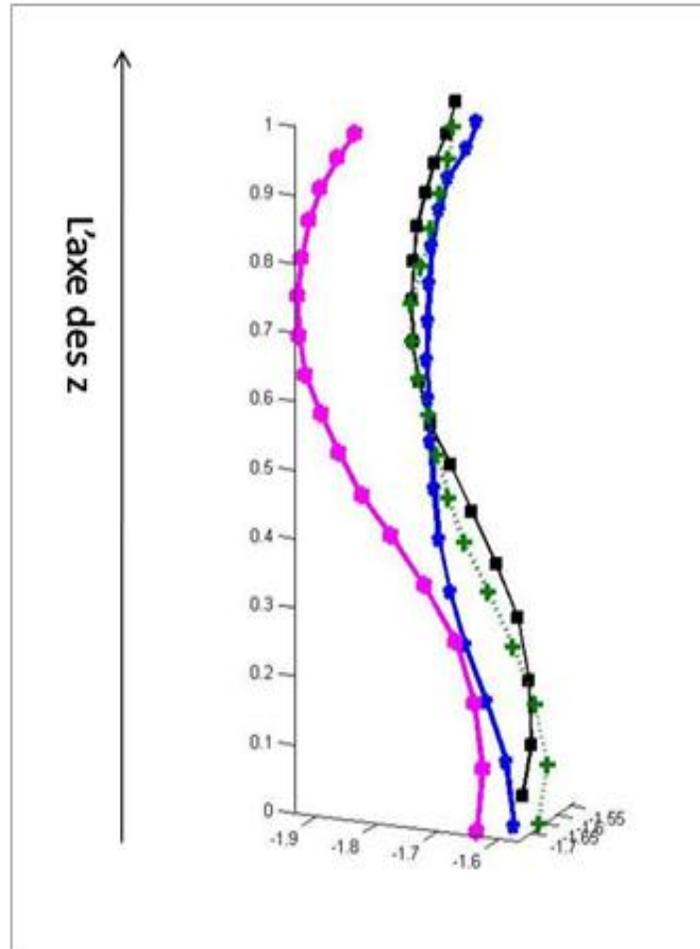


Figure 5.3 : Représentation de quatre colonnes vertébrales normalisées.

5.3.2 Création d'un centre de masse

Après avoir fait la normalisation selon l'axe des Z, les colonnes vertébrales doivent être regroupées toutes dans le même espace, pour cela nous avons créé un centre de masse en ramenant la moyenne de l'intervalle de la variable X et de la variable Y à 0 de telle sorte que chaque colonne doit passer par ce point comme montre la Figure 5.4.

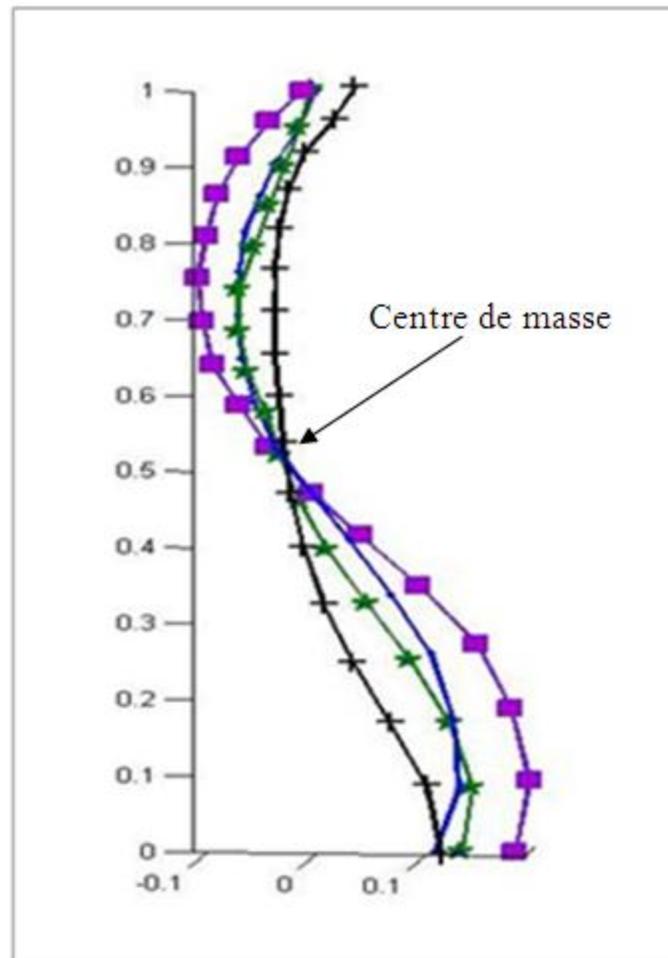


Figure 5.4 : Création d'un centre de masse pour les quatre colonnes vertébrales.

5.3.3 Recalage

Lors de l'acquisition de l'image radiographique une mauvaise position du patient peut introduire un biais sous forme de rotations, afin de minimiser ces angles de rotation nous avons appliqué un algorithme de recalage sur les colonnes vertébrales qui consiste à faire des rotations par rapport au centre de masse afin de les ramener; le plus possible, sur le même plan comme le montre la Figure 5.5.

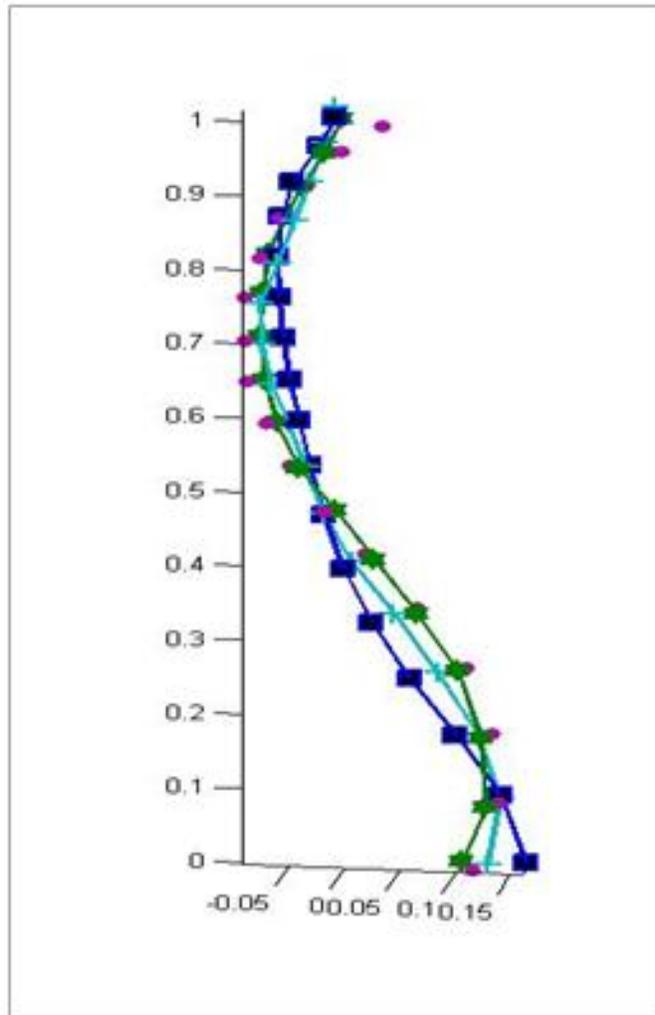


Figure 5.5 : Minimisation de l'erreur de rotation entre quatre colonnes vertébrales

5.4 Extraction des caractéristiques principales

Les colonnes vertébrales de notre base de données sont représentées par une forme 3D reconstruite à partir des images de la colonne vertébrale dans les deux plans frontale et sagittale; cette représentation est sous forme de points qui représentent les centres de masse des vertèbres. Cependant chaque colonne vertébrale se compose de deux parties; partie lombaire et partie thoracique; Les vertèbres de la partie lombaire sont plus grandes que celles de la partie thoracique ce qui fait que la distance entre les centres de masse n'est pas la

même; c'est-à-dire la distance entre les centres dépend des dimensions des vertèbres comme le montre la Figure 5.6.

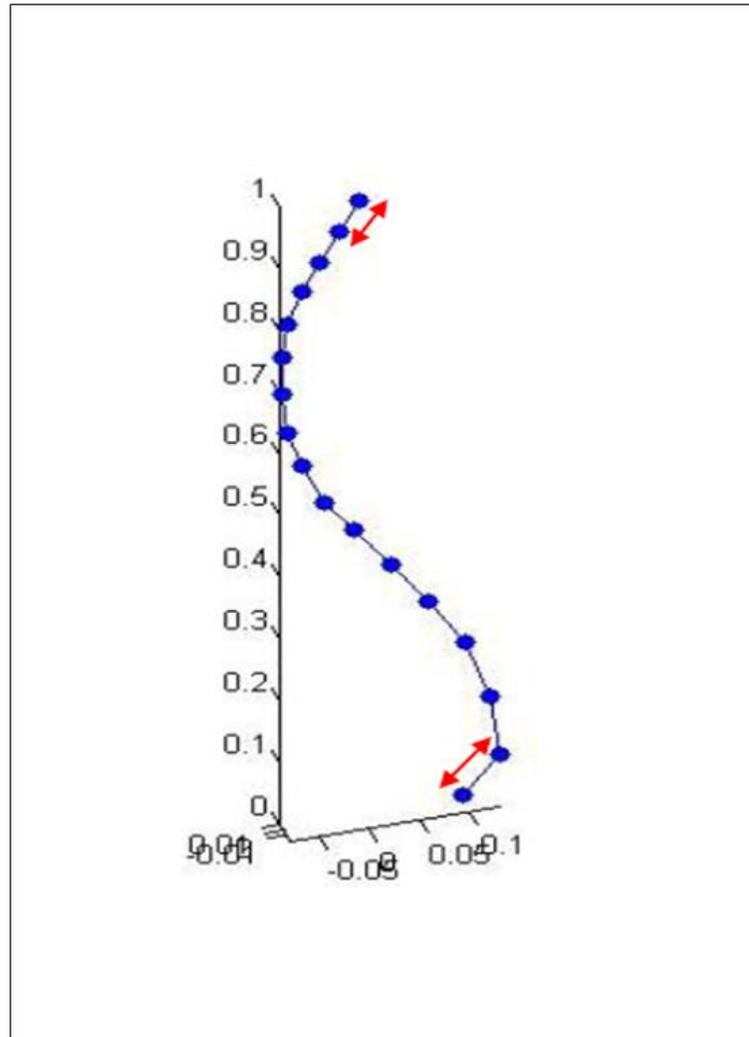


Figure 5.6 : Hétérogénéité des distances entre les centres de masse des vertèbres.

Afin de garantir une meilleure classification des colonnes vertébrales, nous avons restructuré la base de données tout en conservant l'information de base. Nous avons représenté chaque colonne vertébrale par un ensemble de points équidistants échantillonnés par interpolation, dans notre cas nous avons choisi un nombre de points égal à 17 (voir Figure 5.7).

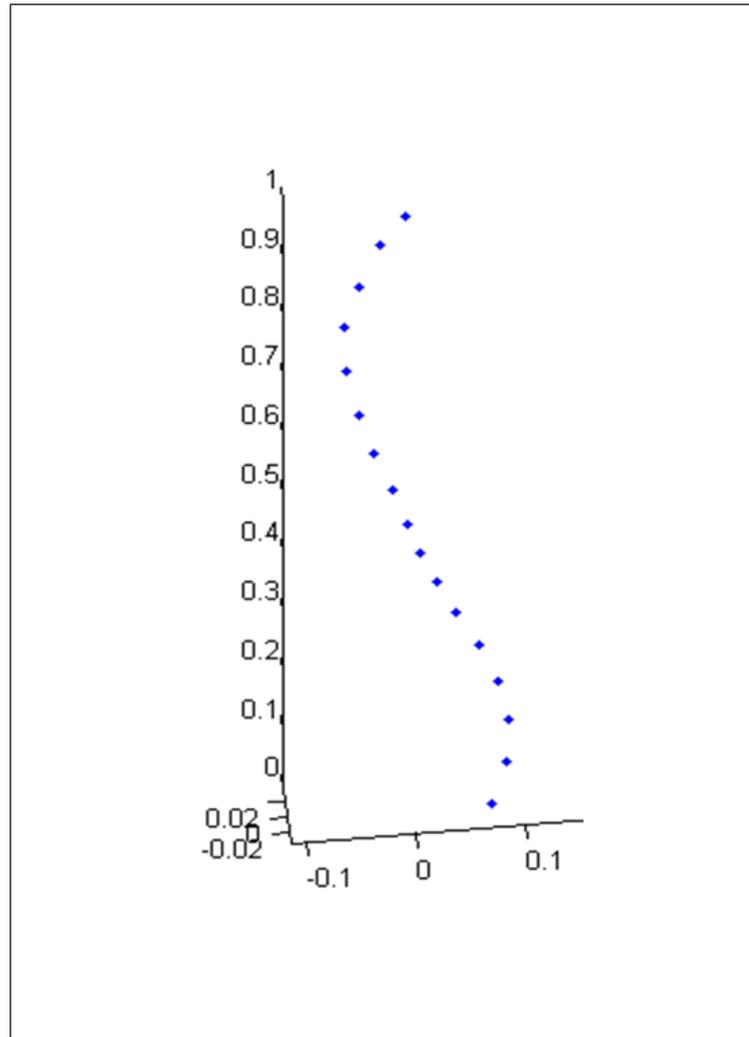


Figure 5.7 : La colonne vertébrale de la Figure 4.12 représentée par un ensemble de points équidistants échantillonnés par interpolation.

5.5 Simulation de la prise de décision individuelle

Après avoir fait le prétraitement ainsi que l'extraction des caractéristiques principales des colonnes vertébrales de notre base de données, nous allons maintenant exploiter ces données afin de simuler automatiquement la prise de décision.

Tout d'abord, nous allons programmer les SVM dont le but est de simuler la décision individuelle si un patient est scoliotique ou non. Mais puisque les SVM sont issus de la théorie de l'apprentissage statistique, c'est-à-dire, que pour garantir une meilleure capacité de

généralisation et, par conséquent, de classification, une phase apprentissage est indispensable afin d'optimiser leurs différents paramètres.

Les principales étapes de notre démarche se résument à :

- Le choix de la fonction noyau.
- L'optimisation des paramètres.
- L'évaluation de la performance en généralisation des réseaux obtenus.

5.5.1 Choix de la fonction noyau

Il est d'abord à noter que pour un problème de classification, le choix de la fonction noyau n'influence que peu le résultat. La souplesse des différentes fonctions conduit à l'obtention de résultats similaires quel que soit le choix de l'utilisateur [Germain, 2009]. Cependant, ceci est valable dans le cas où les données sont nombreuses. Dans notre cas, la règle est donc de fixer une fonction qui soit assez souple au regard du type de comportement à modéliser, sans oublier que le critère primordial reste la capacité de généralisation. Nous optons donc pour les fonctions « *Radial Basis Function* » (RBF), qui sont les fonctions les plus utilisées dans le contexte des méthodes de l'apprentissage statistique.

Ce choix se justifie par plusieurs raisons :

- les fonctions de type RBF permettent de modéliser des comportements fortement non linéaires. Rappelons que l'intérêt d'une méthode basée sur les SVM est justement d'apporter une réponse quant aux difficultés rencontrées par les surfaces de réponses quadratiques face à ces comportements non linéaires.
- l'utilisation du noyau RBF diminue la complexité du SVM qui se mesure par le nombre de paramètres nécessaire pour le représenter. En effet, par rapport au noyau polynômial, l'utilisation du noyau RBF facilite le choix du modèle final car il n'a que deux paramètres à fixer.

- **Le paramètre σ** : En second lieu, il est nécessaire de fixer le paramètre de la fonction noyau RBF qui a une grande influence sur la forme des surfaces de décisions générées par le SVM. Un mauvais choix de ce paramètre peut entraîner à des problèmes de sur-apprentissage.

5.5.3 Optimisation des paramètres

Un des problèmes de l'utilisation des SVM qui reste ouvert est lié au choix des paramètres des fonctions noyaux. Pour optimiser les résultats des classifications en SVM, par un choix des paramètres, différentes techniques sont possibles. Cependant, Il faut prendre garde au sur-apprentissage que peut entraîner une recherche optimale des paramètres par rapport à un échantillon de données.

Le but de cette optimisation est de trouver les paires (C, σ) qui garantissent que le classificateur soit en mesure de classer correctement des cas qui n'ont pas servi pendant la phase d'apprentissage. Pour ce faire nous proposons la démarche suivante :

1. Initialisation des paramètres (C, σ) par la méthode *Grid-Search*;
2. Partition de la base de données en 2 sous-ensembles : Test, Apprentissage (Entraînement & Validation) ;
3. Choix des meilleures paires (C, σ) par la méthode de la validation croisée :
 - a. Entraînement des SVM avec la base de données Entraînement ;
 - b. Evaluation de l'erreur des SVM avec la base de données Validation ;
4. Répétition des étapes 2 et 3 en utilisant les différentes paires (C, σ) ;
5. Choix des meilleurs SVM ;
6. Entraînement des SVM de l'étape 5 en utilisant la base Apprentissage ;
7. Evaluation et choix final des SVM optimaux en utilisant la base de données Test.

5.5.3.1 Initialisation des paires (C, σ)

Pour le noyau RBF, deux paramètres doivent être ajustés : C et σ . Le choix des paramètres qui correspondent le mieux au problème n'est pas intuitif. Pour ce faire, plusieurs méthodes de sélection de modèles (recherche de paramètres) peuvent être utilisées avec un seul but, qui est de trouver les paires (C, σ) de telle sorte que le classificateur puisse classer correctement de nouveaux cas.

Afin d'explorer le plus de paires possible d'un côté, et d'éviter les initialisations aléatoires d'un autre côté, nous avons opté pour une initialisation de type *Grid-Search*. Tout simplement, cette méthode permet de tester tous les points de la grille composé de toutes les valeurs possibles des deux paramètres C et σ .

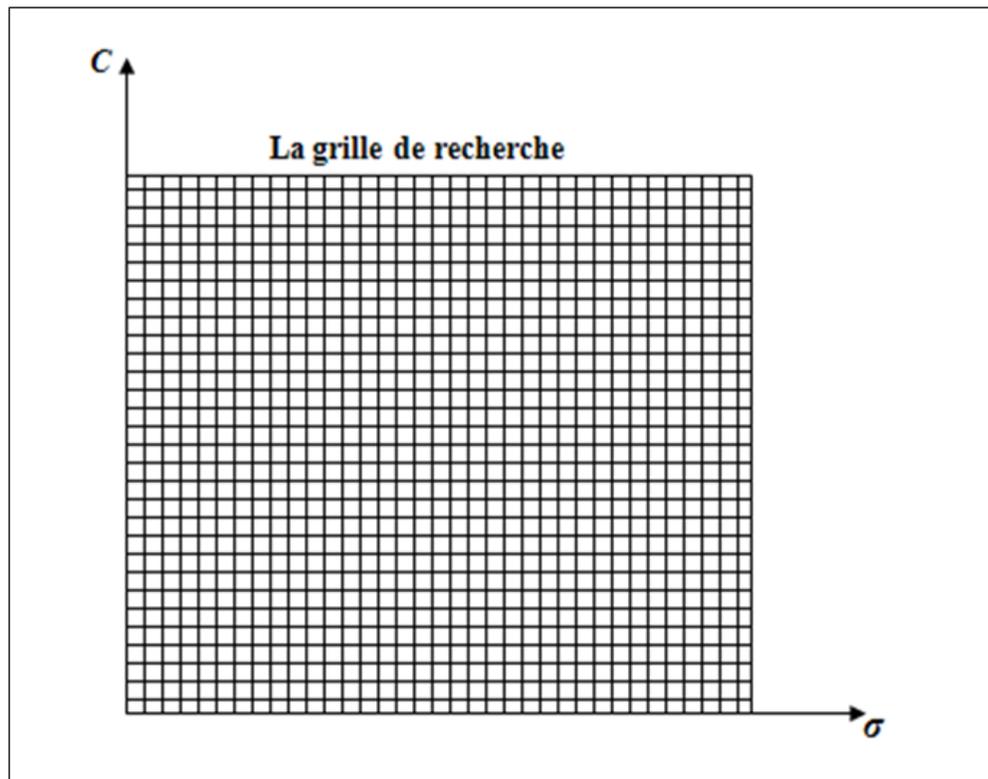


Figure 5.8 : Schématisation de la méthode *Grid-Search* pour l'optimisation des paramètres C et σ [Chih-Wei, Chih-Chung et Chih-Jen, 2010].

Nous avons créé pour chaque paramètre un intervalle qui comprend une séquence à croissance exponentielle comme suit :

$$\begin{cases} C \in [2^{-20}, & 2^{50}] \\ \sigma \in [2^{-50}, & 2^{20}] \end{cases} \quad 5.2$$

L'avantage de la méthode *Grid-Search* est d'offrir une protection contre les minima locaux. Cependant, dans le cas où le nombre de paramètres à ajuster est important, le nombre de déterminations devient excessif ce qui rend cette méthode très coûteuse en terme de temps de calcul.

5.5.3.2 Partition de la base des données

Pour sélectionner notre classificateur optimal, nous devons, d'un côté, déterminer ses paramètres optimaux à l'aide d'un processus d'apprentissage. De l'autre côté, il faut que le vrai risque de ce classificateur soit calculé en utilisant un ensemble de données qui n'a pas servi à la phase d'apprentissage. Pour cela, une partition de la base de données s'avère indispensable.

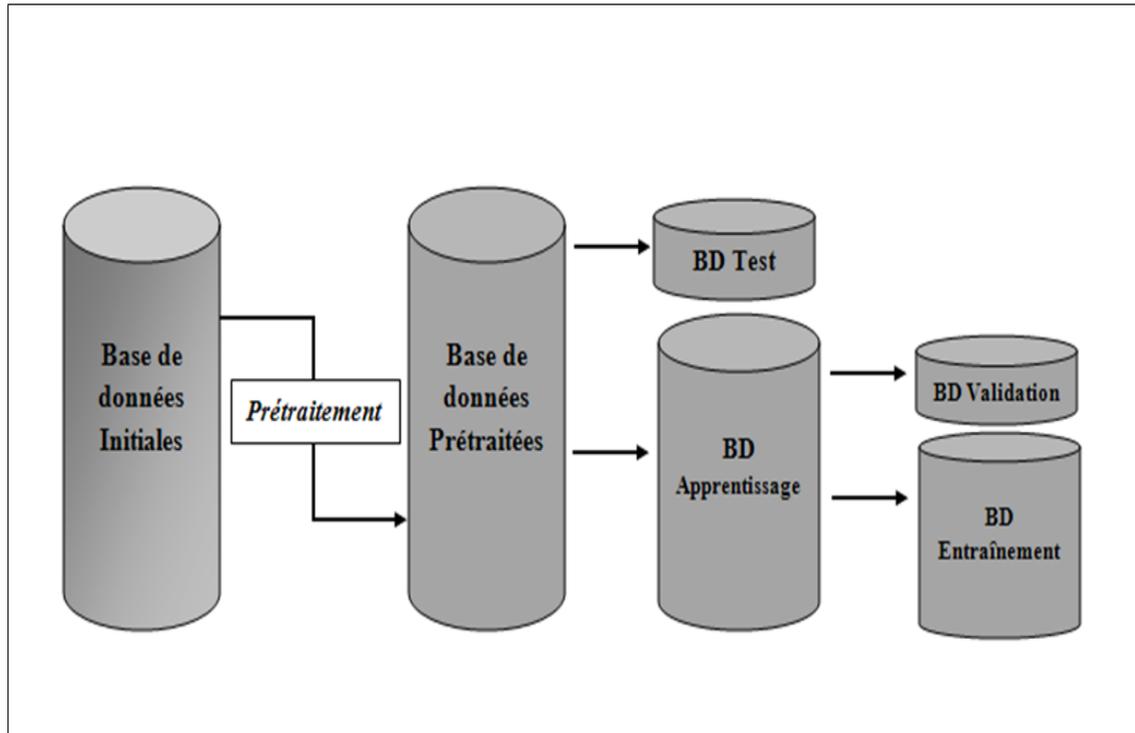


Figure 5.9 : Partition de la base de données en sous blocs.

- 1) **Base de données Test** : La méthode privilégiée pour estimer le vrai risque d'un classificateur consiste à le tester sur des exemples qui n'ont pas été utilisés pendant le processus d'apprentissage. Pour ce faire, nous réservons une portion des données étiquetées à notre disposition pour former un ensemble que nous appelons BD Test. L'erreur calculée du SVM sur l'ensemble BD Test est une estimation sans biais de la capacité de généralisation de ce dernier, et c'est sur la base de cette erreur que va se faire le choix final du meilleur SVM.

- 2) **Base de données Apprentissage** : Comme son nom l'indique, la base d'apprentissage représente l'ensemble des données qui vont servir à déterminer les meilleurs paires (C, σ) où une approche de validation croisée sera utilisée. Cette approche consiste à découper le jeu de données en k groupes tirés aléatoirement où, à chaque fois, $(k-1)$ groupes seront utilisés pour l'entraînement du SVM et le groupe restant sera l'ensemble de Validation.

- a) *Base de données Entraînement* : Ensemble de données utilisé pour effectuer l'apprentissage du réseau.
- b) *Base de données Validation* : Ensemble de données utilisé pour le calcul de l'erreur du réseau.

Le Tableau 1.3, ci-contre, représente la partition que nous avons apportée à notre base de données :

Tableau 5.3 : Partition en sous blocs de la base de données initiale.

Base de données (BD)	Nombre Total des Cas	Normale	Scoliotique
BD Initiale	160	91	69
BD Prétraitées	160	91	69
BD Test	30	16	14
BD Apprentissage	130 (5 x 26)	75 (5 x 15)	55 (5 x 11)
BD Entraînement	104 (4 x 26)	60 (4 x 15)	44 (4 x 11)
BD Validation	26 (1 x 26)	15 (1 x 15)	11 (1 x 11)

5.5.3.3 Validation croisée

La validation croisée k -fois (ou k -fold cross validation) est une méthode couramment utilisée pour sélectionner les paramètres d'un algorithme d'apprentissage. Comme nous l'avons mentionné plus haut, cette méthode consiste à partitionner l'ensemble d'apprentissage en k sous-ensembles disjoints, nous avons choisi $k = 5$ qui est une valeur typique utilisée souvent par les programmeurs [Kaastra et Boyd, 1995].

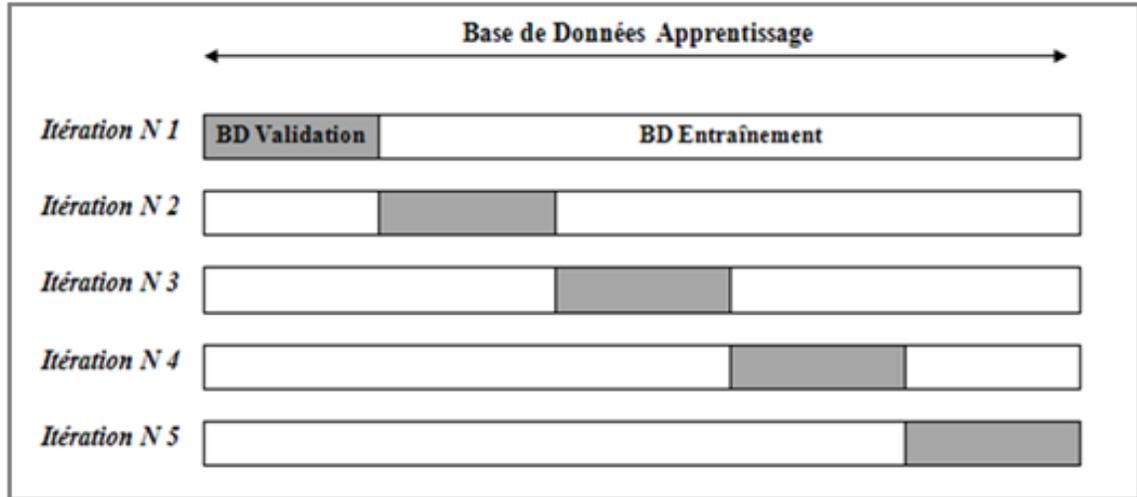


Figure 5.10 : Séquences de déroulement de la crosse validation 5-fois.

Nous effectuons ensuite les étapes de calcul suivantes pour chaque paire (C, σ) :

1. Itération à réaliser k fois :
 - construire une base d’Entraînement par l’union de $(k-1)$ sous-ensembles ;
 - effectuer un apprentissage du SVM avec les paramètres (C, σ) ;
 - calculer l’erreur quadratique moyenne sur l’ensemble de validation constituée par le sous-ensemble des observations restantes :

$$EQM = \sqrt{\frac{1}{N_v} \sum_{i=1}^{N_v} (r_i)^2} \quad 5.3$$

2. Calculer le score de validation croisée à partir des k quantités EQM :

$$SVC = \sqrt{\frac{1}{k} \sum_{i=1}^k EQM^2} \quad 5.4$$

Le chiffre à retenir doit être le plus petit.

Les meilleurs SVM qui vont être retenus à cette étape, sont ceux qui minimisent la valeur SVC.

5.5.4 Programmation numérique

Afin de programmer la démarche citée plus haut, nous avons utilisé le logiciel MATLAB, version 7.10.0 (R2010a) que nous avons enrichi avec la boîte à outil PRTools. En effet, PRTools (Pattern Recognition Tools) est le logiciel de base utilisé dans le livre « *Classification, parameter estimation and state estimation - an engineering approach using Matlab* » [Duin et al., 07], développé essentiellement pour des problèmes de classification et de reconnaissance de formes. Cette boîte à outil propose plus de 200 routines destinées essentiellement à la conception et l'évaluation de nouveaux algorithmes et à la construction de prototypes informatiques.

Le programme que nous avons élaboré, représenté par la Figure 5.11, permet de :

- sélectionner les meilleurs SVM avec la méthode de la validation croisée ;
- entraîner les SVM retenus avec la base de données Apprentissage ;
- tester la performance en généralisation des SVM retenus sur la base de données test;
- retenir les SVM les plus performants.

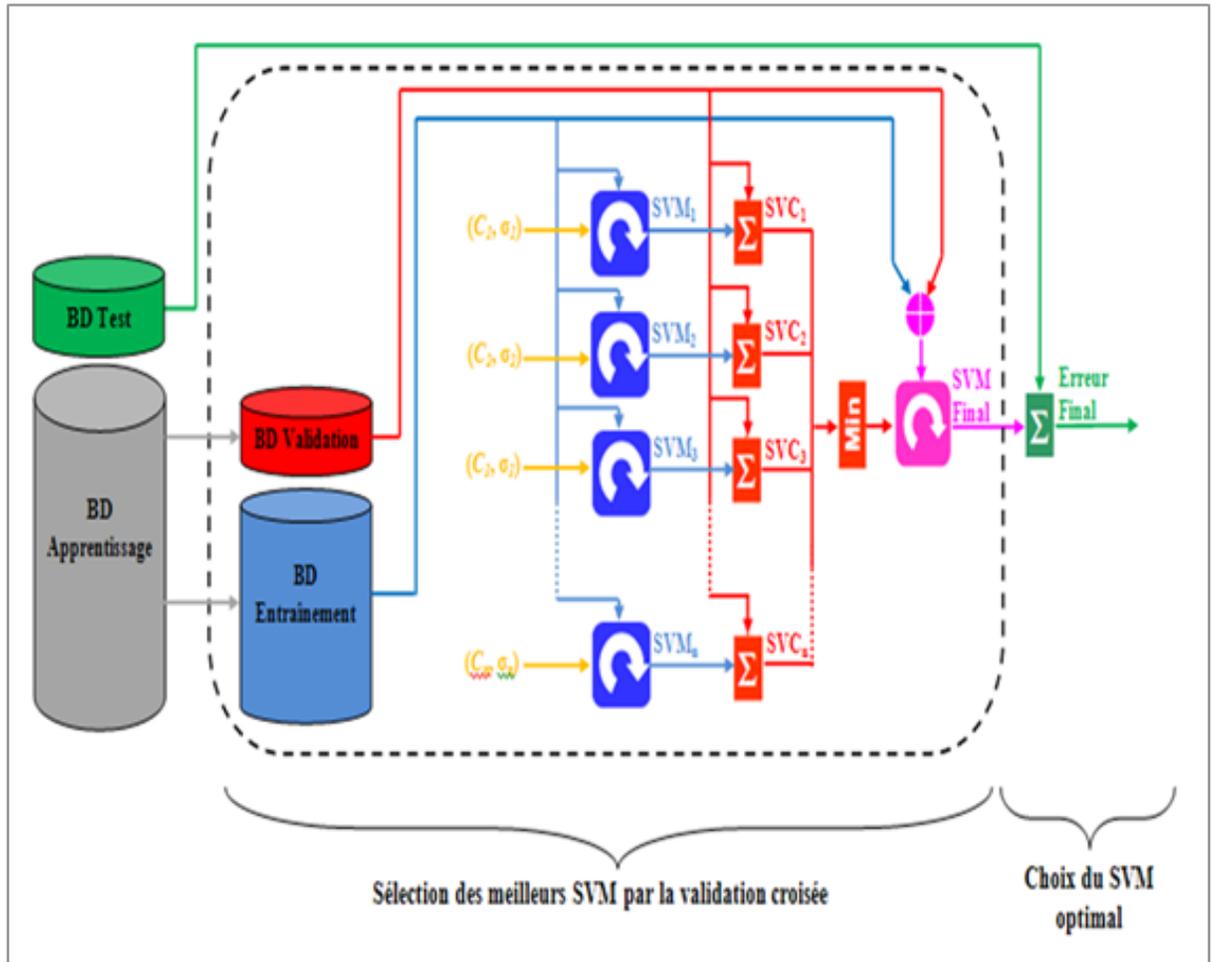


Figure 5.11 : Programmation des SVM sous MATLAB

5.5.5 Résultats

L'exécution du programme informatique a été faite par MATLAB, version 7.10.0 (R2010a), sur un ordinateur processeur Centrino Intel Pentium 1.73 Ghz avec une mémoire de 2 Go. Les résultats obtenus par notre démarche sont représentés sur la Figure 5.12.

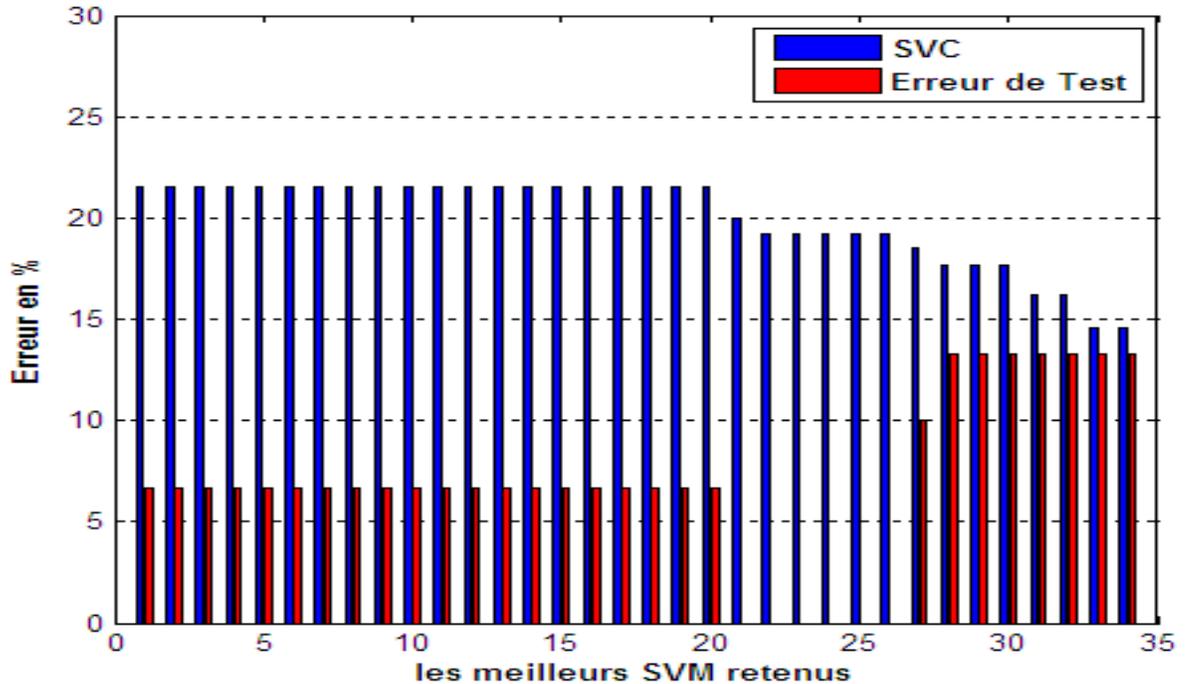


Figure 5.12 : Les meilleurs SVM choisis selon leur score de validation croisée (SVC) et le taux d’erreur de test correspondant.

L’optimisation nous a donné 34 SVM, accompagné par leur score de la validation croisée (SVC) et le taux d’erreur de test correspondant, les valeurs optimisées de C et de σ sont aussi fourni. (*cf. Annexe III*)

Malgré la complexité de la tâche de classification, due essentiellement au nombre important de descripteurs de chaque colonne vertébrale (51 descripteurs), les taux d’erreur des SVM retenus sont très satisfaisants. Cependant, nous remarquons sur la figure 5.12 qu’il existe trois sous-ensembles de SVM :

- de 1 - 20 : les vingt SVM de ce sous-ensemble classifient les 160 cas de la base de données avec une très bonne précision : une erreur de test inférieure à 7% et un score de validation croisée de l’ordre de 22%.
- de 21 - 26 : Les SVM de ce sous-ensemble sont les meilleurs classificateur obtenus par notre démarche d’optimisation, non seulement ils minimisent le score de

validation à 19%, ces derniers classifient parfaitement les cas de la base de données test et donne un taux d'erreur de test égal à 0%.

- de 27 - 34 : dans ce sous-ensemble, nous remarquons que les SVM ont les meilleurs scores de validation croisées, jusqu'à 14%, mais avec une faible capacité de généralisation où le taux d'erreur de test dépasse 13%. Cela se traduit par le phénomène de sur-apprentissage où les SVM commencent à apprendre par cœur les exemples de la base d'apprentissage au détriment de leur capacité de généralisation.

Par conséquent, les SVM qui vont être retenus pour la conception du système Multi-Agents dont le but est de simuler la prise de décision collective sont ceux du deuxième sous ensemble (de 21 à 26).

5.6 Simulation de la prise de décision collective

Cette section représente la deuxième et dernière étape de conception de l'outil d'aide à la décision que nous proposons. En effet, après avoir simulé la décision individuelle avec les Séparateurs à Vaste Marge, nous allons à présent regrouper les meilleurs SVM obtenus au sein d'un seul réseau afin d'aboutir à notre Système Multi-Agents dont le but est de simuler la prise de décision collective de l'ensemble des SVM.

Le Système Multi-Agents que nous proposons est de type cognitif avec une architecture à base de tableau noir. L'intérêt de cette architecture est d'être suffisamment souple et puissante pour pouvoir implémenter les différentes structures organisationnelles qui unissent les agents. La figure 5.13 illustre les principales composantes de notre SMA.

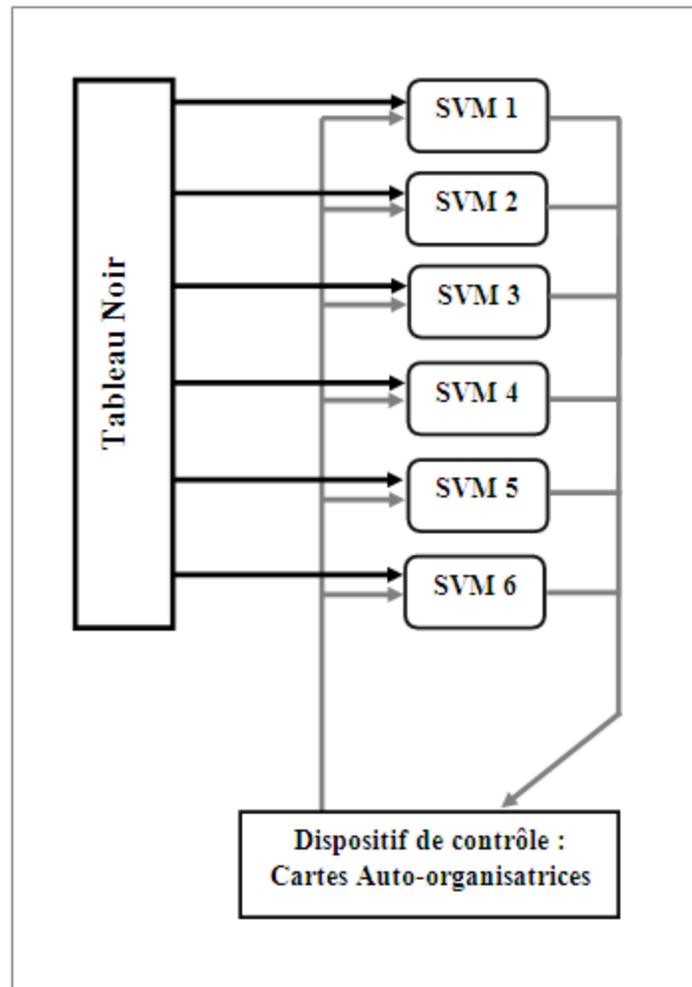


Figure 5.13 : Système Multi-Agents décisionnel avec une architecture à base de Tableau Noir.

5.6.1 Les Agents

Les agents de notre SMA sont les SVM optimaux retenus dans l'étape de simulation de la prise décision individuelle. La modélisation des agents avec des SVM fait de ces derniers des agents de type cognitif. Cette configuration donne à notre SMA deux caractéristiques :

- Le système d'agents cognitifs est fondé sur l'interaction et la coopération d'agents autonomes capables à eux seuls d'effectuer des tâches relativement complexes, à savoir, la prise de décision ;

- Le système cognitif que nous proposons est de forte granularité, car il regroupe un petit nombre d'agents en interaction, dans notre cas nous avons un total de six agents. Cette configuration fait de notre système Multi-Agents cognitifs une réalisation du concept de l'intelligence collective originelle discuté auparavant.

5.6.2 Le Tableau Noir

L'architecture à base de tableau noir est l'une des plus utilisée dans les systèmes multi-agents cognitifs. En tout état de cause, les avantages de ce type d'architecture sont indéniables : du fait de sa très grande plasticité, il est possible d'implémenter n'importe quelle structure organisationnelle des agents.

Le modèle de tableau noir est fondé sur un découpage en modules (agents) indépendants qui ne communiquent directement aucunes informations, mais qui interagissent indirectement en partageant des informations qui constitue le tableau noir.

Dans le cadre de notre application, le tableau noir est composé de deux tableaux qui sont :

- Le premier Tableau est notre base de données prétraitées qui a servi à faire l'apprentissage ainsi que la sélection des SVM optimaux et qui va servir, à cette étape, à obtenir les décisions individuelles générées par chaque SVM après simulation sur l'ensemble des cas de la base de données.
- Le second tableau sera utilisé au niveau du dispositif de contrôle dont le but est d'aboutir à la structure organisationnelle des agents. Ce tableau est composé d'une part des résultats de simulation des différents agents sur l'ensemble des cas de la base de données prétraitée, c'est-à-dire la décision de chaque SVM, en plus de la décision réelle issue de la base de données initiale de l'autre part.

5.6.3 Le dispositif de contrôle

Le choix d'une organisation est un problème majeur, spécialement pour les systèmes distribués puisqu'il conditionne les propriétés globales du système. Le rôle du dispositif de contrôle est de gérer les conflits d'accès entre les agents qui se déclenchent de manière opportuniste. A cette fin, nous avons utilisé les cartes auto-organisatrices de Kohonen qui sont des réseaux non supervisés avec un apprentissage compétitif et qui permettent de modéliser l'espace des entrées à l'aide d'une carte à deux dimensions ce qui offrent la possibilité de converger vers une structure organisationnelle des agents (SVM) la plus appropriées via le processus d'apprentissage.

5.6.3.1 Topologie de la carte auto-organisatrice

Une carte auto-organisatrice, comme son nom l'indique, s'organise par rapport aux stimuli présentés en respectant les contraintes topologiques de l'espace d'entrée. Il y a mise en correspondance de l'espace d'entrée avec l'espace du réseau. Les zones voisines de l'espace d'entrée sont voisines sur la carte auto-organisatrice. Il s'agit de réseaux non supervisés avec un apprentissage compétitif qui permettent de modéliser l'espace des entrées avec des prototypes (Agent) et de construire une carte à deux dimensions afin de structurer cet espace.

Dans notre cas, nous avons un nombre d'agent égal à six, pour cela nous allons utiliser une carte à deux dimensions 2 x 3 afin de modéliser notre espace d'entrée composé des résultats de simulation des SVM sur l'ensemble des cas de la base de données prétraitée, en plus de la décision réelle issue de la base de données initiale. La figure 5.14 illustre la topologie de la carte que nous proposons.

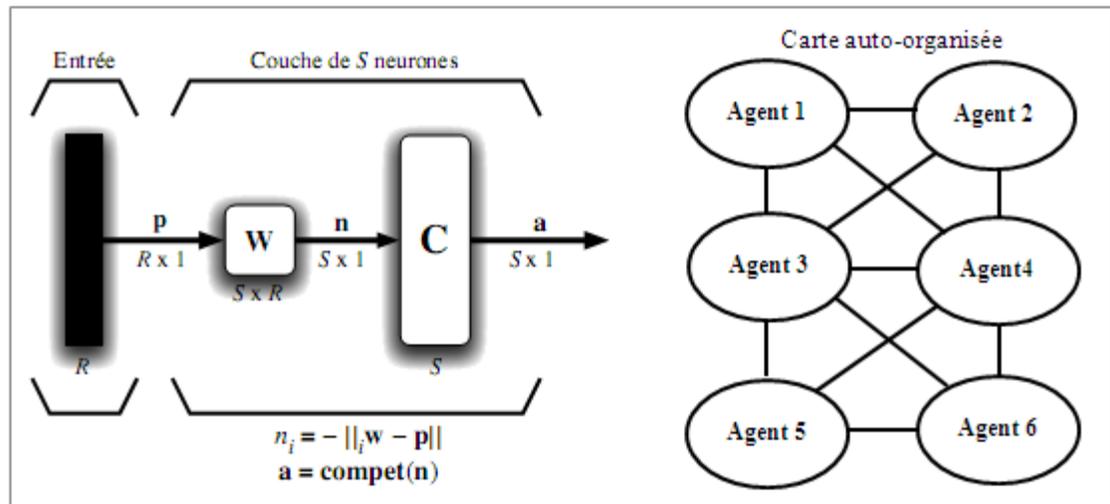


Figure 5.14 : Réseau de Kohonen avec carte rectangulaire de dimension $S=2 \times 3 = 6$ SVM.

5.6.3.2 Apprentissage de la carte auto-organisatrice

L'algorithme de base de Kohonen est résumé à la figure 5.15. Il consiste à échantillonner les stimuli d'apprentissage jusqu'à l'atteinte du critère d'arrêt spécifié par un nombre maximum d'itération. A chaque itération, on détermine l'agent gagnant (le plus proche) et on déplace celui-ci, ainsi que son voisinage actuel, dans la direction du stimulus courant en utilisant le taux d'apprentissage courant.

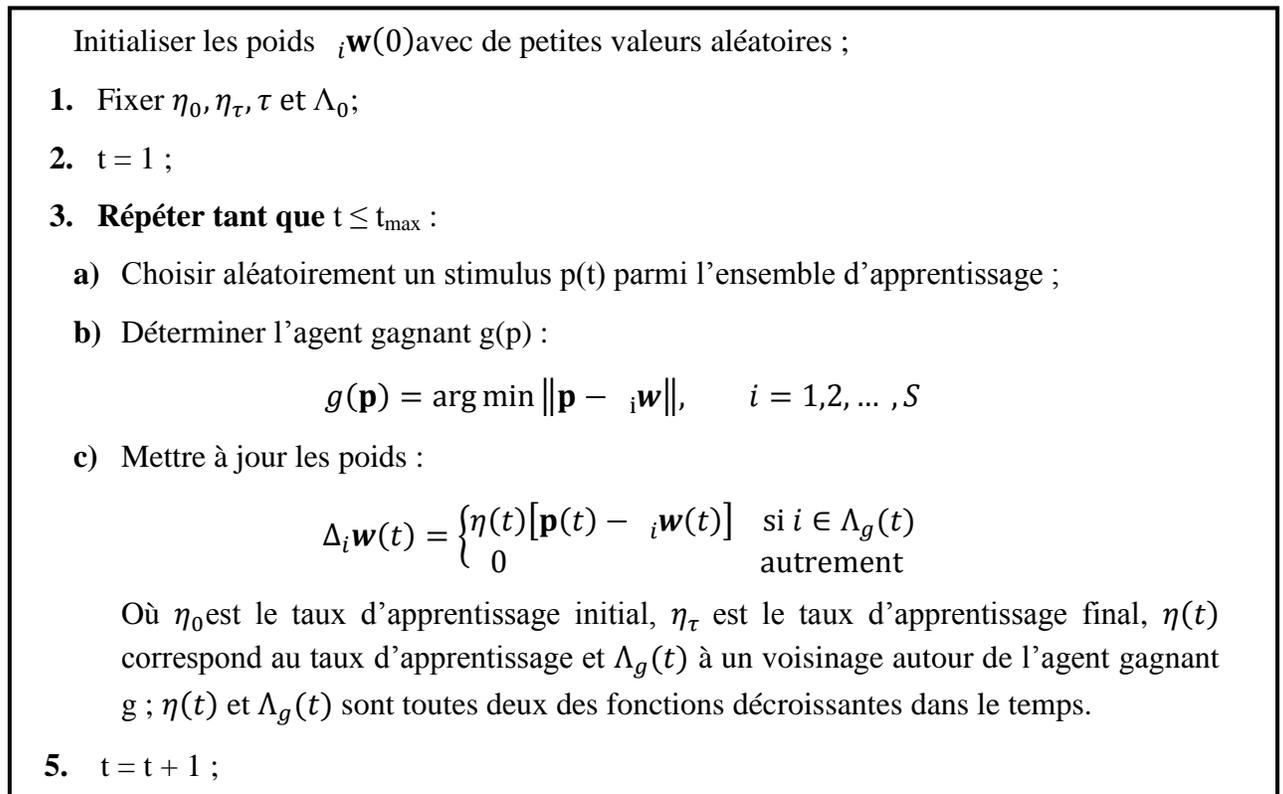


Figure 5.15 : Algorithme d'apprentissage de Kohonen. [Parizeau, 2004]

Cet algorithme est de type compétitif, il consiste à mettre à jour les poids W des arrêtes qui unissent les agents à chaque itération en déterminant l'agent gagnant g . Ce dernier est déterminé simplement en retenant l'indice du neurone pour lequel la distance avec le stimulus d'entrée est minimum.

5.6.4 Programmation et résultats

Afin de converger vers la structure finale des agents, une phase apprentissage de la carte auto-organisatrice est indispensable pour fixer les poids des connections entre les différents agents. Pour cela, nous avons partitionné notre deuxième tableau de la même façon que pour la base de données prétraitée, c'est-à-dire, une base Apprentissage composée de 130 cas (75 Normaux + 55 Scoliotiques) et une base Test composée de 30 cas (16 Normaux + 14 Scoliotiques)

De plus, nous avons fixé un nombre maximum d'itérations égale à 200 itérations, où nous avons remarqué que les poids des connections se stabilisent et convergent vers leur valeurs finales. La figure 5.16 représente la topologie finale de notre carte auto-organisatrice, après apprentissage, qui regroupe nos 6 agents.

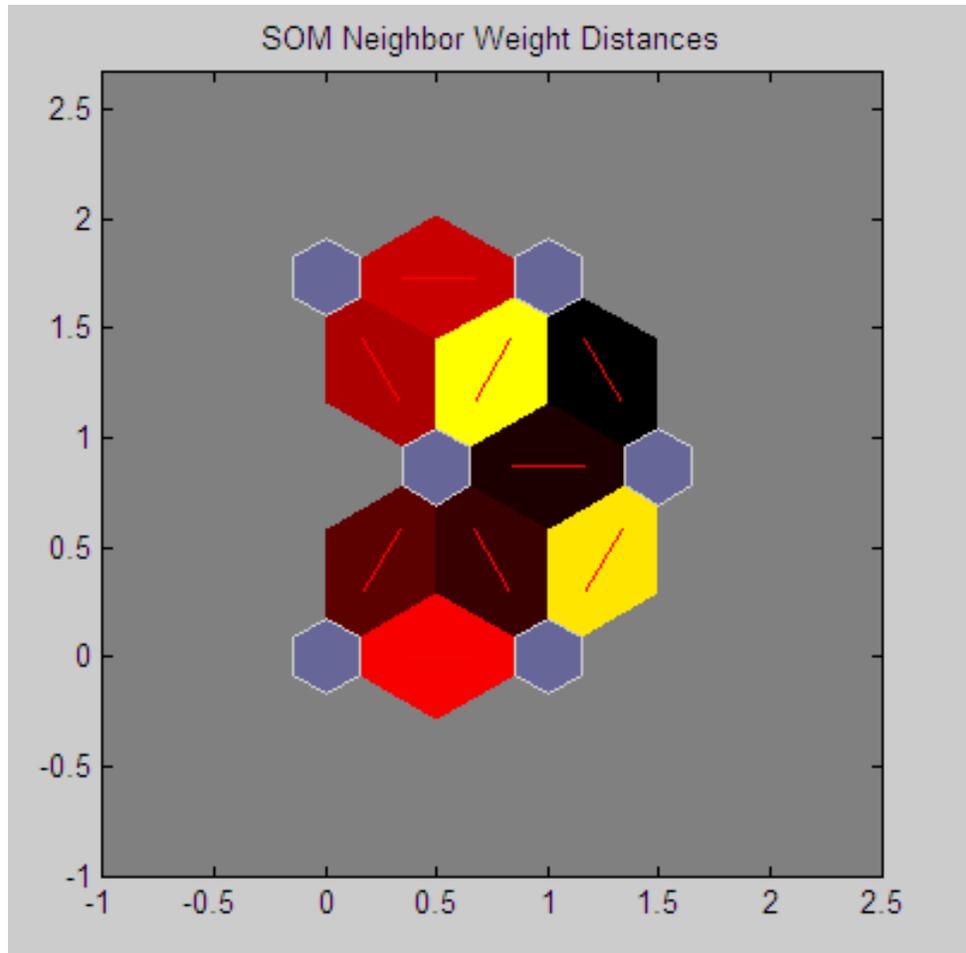


Figure 5.16 : Topologie finale de la structure organisationnelle des Agents.

Le code couleur de la figure 5.16 est le suivant :

- Les hexagones bleus représentent les agents.
- Les lignes rouges représentent les arrêtes qui relient les agents voisins.

- Les couleurs dans les régions contenant les arrêtes rouges indiquent les distances entre les agents :
 - Les couleurs plus foncées représentent de plus grandes distances.
 - Les couleurs plus légères représentent de plus petites distances.

Mais puisque la pertinence d'une organisation est contextuelle, ce n'est pas la structure organisationnelle des agents qui nous intéresse le plus mais c'est plutôt les résultats décisionnels obtenus avec cette dernière. Dans notre cas, les résultats de la phase d'apprentissage que nous avons obtenus avec cette topologie de la carte auto-organisatrice sont représentés à la figure 5.17.

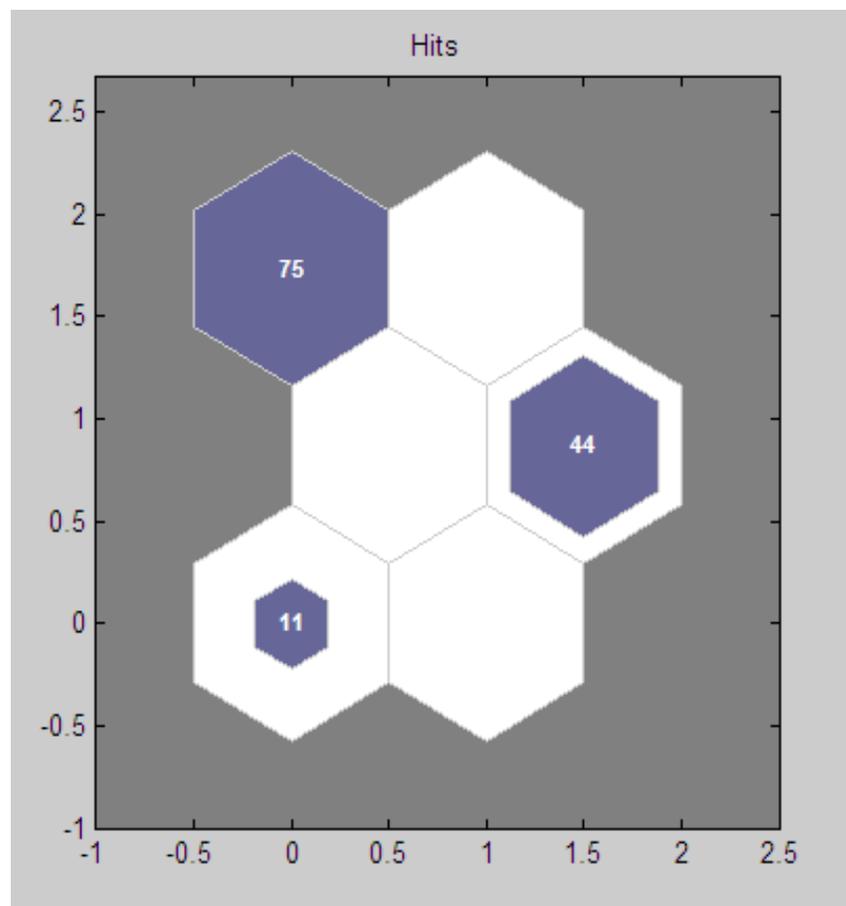


Figure 5.17 : Décisions collective des six agents générées sur la base d'Apprentissage.

Les résultats que donne l'ensemble des agents réunis au sein de notre carte auto-organisatrice sont excellents. En effet, à partir de la figure 5.17, nous remarquons que notre SMA classe parfaitement les 75 cas normaux de la base d'Apprentissage, alors que pour les 55 cas scoliotiques, nous avons 44 cas classés correctement, alors que pour les 11 cas restants, nous remarquons que le SMA a créé une nouvelle classe, que nous appelons classe de non consentement. Cette nouvelle classe comporte l'ensemble des cas sur lesquels les agents n'arrivent pas à aboutir à un consensus concernant la classe approprié de ces derniers. Ces résultats donnent un taux d'erreur d'apprentissage égale à 8.46 % qui est nettement meilleur que les taux d'erreur obtenus pendant la phase d'apprentissage lors de la simulation de la prise de décision individuelle.

En ce qui concerne la phase de test nous avons obtenus des résultats pareils à ceux obtenus lors de la simulation de la prise de décision individuelle, c'est-à-dire, notre SMA arrive à classer parfaitement tous les cas de la base de données Test et donne un taux d'erreur de test égale à 0 %. La figure 5.18 illustre ces résultats.

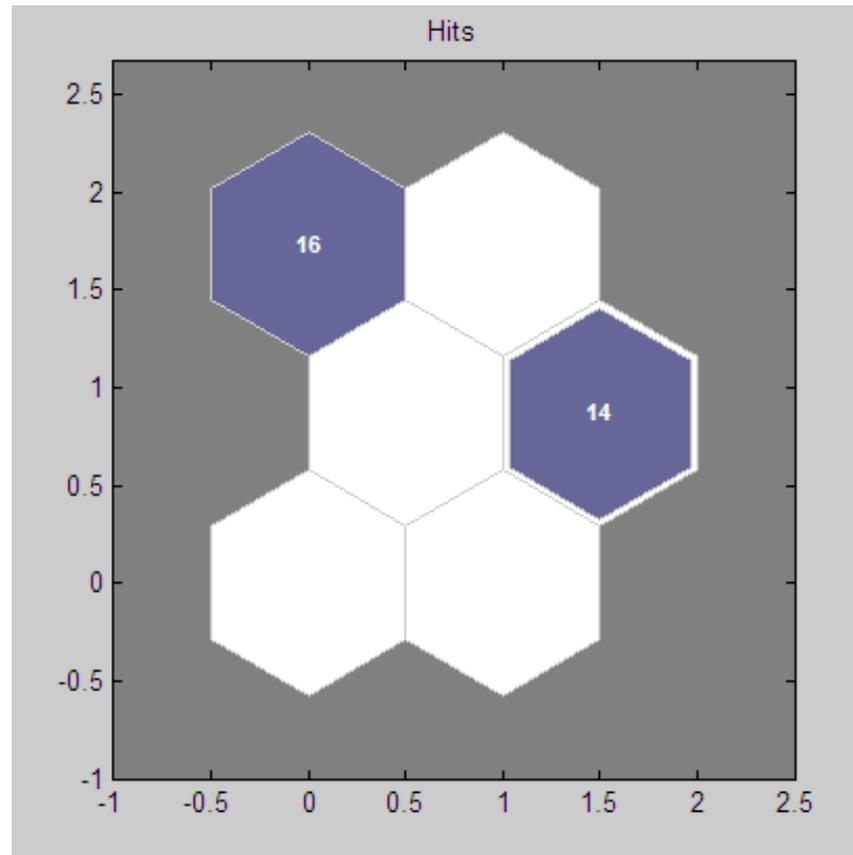


Figure 5.18 : Décisions collective des six gents générées sur la base de données Test.

5.7 Discussion des résultats

En tout état de cause, la performance ainsi que la robustesse du modèle que nous avons proposé sont indéniable. Cela s'explique par les résultats très satisfaisants que nous avons obtenus tout au long de notre application. La méthodologie que nous avons adoptée est organisée en quatre étapes principales qui sont :

- Présentation et restructuration de la base de données test sous forme *dataset* ;
- Prétraitement de la base de données et extraction des caractéristiques principales ;
- Simulation de la prise de décision individuelle par les SVM
- Simulation de la prise de décision collective par un Système Multi-Agents cognitif avec une architecture à base de tableau noir.

Le but du prétraitement est de ramener, le plus possible, les colonnes vertébrales représentées en 3D, à une représentation sur un plan 2D afin de faciliter la tâche de classification. Cependant, l'identification des 17 points anatomiques (les positions des vertèbres) ainsi que les transformations géométriques appliquées sur ces données, ont révélées que l'étape de l'identification du centre de masse des colonnes vertébrales pourrait biaiser l'information de base des colonnes vertébrales et influencer les résultats de la classification. En fait, si le degré de déviation de la colonne vertébrale (l'angle de Cobb) est grand entre les données, le centre de masse de chaque colonne vertébrale varie considérablement dans l'espace, ce qui peut engendrer des erreurs de translations dans la direction de l'axe des Z, et par conséquent, peut fausser la comparaison. Néanmoins, l'excellent résultat de classification montre l'effet minime de ce facteur.

Un deuxième constat que nous avons fait concerne l'étape de simulation de la décision individuelle, et plus exactement le temps d'exécution du programme informatique. En effet, nous avons constaté que le temps nécessaire pour l'exécution de la phase d'apprentissage dépasse largement les 26 heures, alors que pour la phase de test, l'exécution se fait instantanément. Le temps de calcul constitue donc une limitation des SVM comparativement à la taille de notre base de données (160 cas avec 51 descripteurs). Cette limitation deviendra majeure lorsqu'on applique les SVM sur des bases de données de grande tailles ou complexes et il faudra faire un compromis entre la pertinence de la classification (taux d'erreur de classification) et le temps d'exécution. Les facteurs qui ont influencés le temps d'exécution du programme informatique sont :

- La complexité de la tâche de classification due au nombre important de descripteurs pour chaque cas de la base de données ;
- L'initialisation des paramètres C et σ du noyau RBF par la méthode Grid-Search, où nous étions appelé à explorer 300^2 paires ce qui rend cette méthode très couteuse en terme de temps de calcul ;
- Le dernier facteur qui influence le temps d'exécution est le nombre de vecteurs supports que contient la base de données Apprentissage. En effet, le SVM est un

modèle dont la complexité est fonction du nombre de vecteurs de support, c'est-à-dire, ce classificateur contient autant de paramètres que de vecteurs supports qui servent à déterminer les hyper-plans qui délimitent la marge. Pour notre application nous avons un total de 103 vecteurs supports ce qui complique la détermination des hyper-plans.

Pour le cas traité dans ce mémoire, l'excellent taux de classification obtenu avec les SVM compense largement le temps d'exécution. De plus, nous nous sommes intéressé à comparer nos résultats avec ceux obtenus avec d'autres classificateurs, à savoir, le PMC et le K-NN (*cf. Annexe II*) qui ont été appliqué à la même base de donnée [Mezghanie et al., 2008]. Les taux d'erreur de test respectif des deux modèles sont 10% et 7%. Cette comparaison à démontrer, une fois de plus, la robustesse et la supériorité des SVM par rapport aux autres classificateurs. Mais contrairement au SVM, le MLP et le K-NN sont plus facile à implémenter comme ils sont largement moins couteux en termes de temps d'exécution.

La dernière étape de notre méthodologie est la simulation de la prise de décision collective. Comme prévu, dans cette étape nous avons pu améliorer considérablement le taux d'erreur de la phase d'apprentissage qui a passé de 19% à 8.46% tout en gardent une performance maximale lors de la phase de test où nous avons obtenu un taux d'erreur de test égale à 0% pour les deux étapes de simulation de la prise de décision.

Le deuxième constat que nous avons fait au niveau de cette dernière étape et que nous considérant comme étant un avantage majeur est la création d'une nouvelle classe que nous avons appelée classe de non consentement. Cette classe comporte l'ensemble des cas sur lesquels les agents n'arrivent pas à aboutir à un consensus concernant la classe approprié de ces derniers. Cet avantage nous permet de tirer deux considérations importantes, la première est d'ordre théorique alors que la deuxième est d'ordre pratique :

- La première considération concerne le taux d'erreur. Dans notre cas, nous n'allons pas parler de taux d'erreur mais plutôt de taux de non classification. En effet, pour les cas dont les agents n'arrivent pas à un consensus, le SMA crée une nouvelle

classe au lieu de les classer au hasard dans l'une des deux classes : Normal ou Scoliotique. C'est-à-dire que pour ces cas, notre outil d'aide à la décision affiche un résultat d'incapacité de classification ou de décision, ce que nous considérons meilleur qu'une fausse décision ;

- La deuxième considération concerne les cas de la classe de non consentement. Généralement, ces cas sont ceux qui se trouvent à la limite de la marge ou, même des fois, à l'intérieur de la marge (marge molle), ce qui explique la difficulté de leur classification. Du point de vue clinique, ces cas sont souvent des personnes qui ont une attitude scoliotique ou bien une scoliose modérée. Le résultat d'incapacité de classification généré par notre outil d'aide à la décision va pousser les cliniciens à prendre plus d'attention au sujet de ces cas, car il est important de noter qu'il est tout aussi grave d'étiqueter "attitude scoliotique" une véritable scoliose qui va continuer d'évoluer alors qu'un traitement bien conduit l'en aurait empêché que d'appeler "scoliose" une banale attitude scoliotique et d'imposer des traitements fastidieux inutiles.

CONCLUSION GÉNÉRALE

Le processus décisionnel est un processus cognitif, organisationnel (appartenant à un groupe d'individus de l'organisation, soumis aux rapports du groupe) et constructiviste (rien n'est défini à priori : ni le problème, ni les informations, ni les acteurs impliqués dans ce processus, ni les objectifs).

L'aide à la décision n'a pas pour ambition d'établir des vérités objectives ou de découvrir des vérités cachées mais de contribuer à construire des convictions individuelles, des décisions collectives, des compromis entre rationalités, des enjeux, des valeurs multiples et souvent conflictuelles. Elle vise à asseoir, sur des bases dites scientifiques faisant référence à des hypothèses de travail, des énoncés de propositions qu'elle soumet au discernement d'un décideur et/ou de divers acteurs engagés dans le processus de décision. Pour cela, elle prend appui sur des modèles.

Dans cette perspective, l'objectif de ce présent travail est de proposer une modélisation dite cognitive d'un outil d'aide à la décision. Mais puisque la qualité d'un modèle se mesure dans deux dimensions : celle de l'homomorphie avec le phénomène modélisé et de la maniabilité pour répondre à une question de recherche concrète. Nous avons adopté une approche exploratoire et prospective afin d'aboutir à une modélisation qui soit adaptée à la fois aux exigences de l'économie de la recherche et à la nécessité de modéliser un décideur dans son environnement et non pas un système de paramètres dénué de tout sens.

L'analyse cognitive nous a révélé que le décideur opère dans un environnement de problèmes très complexe et mal structuré dont les données sont instables et dépendent également de l'interaction du décideur avec cet environnement. Ainsi, la décision doit être perçue comme le résultat d'un processus global de résolution de problèmes qui est souvent long et de nature presque continue où le décideur, décrit comme un système ouvert qui, activement, saisit de l'information, qui la traite et qui agit, tend à transformer l'entropie riche (complexité peu structurée) de l'environnement en entropie de moins en moins riche jusqu'à ce qu'il puisse interrompre ce processus par une décision. C'est-à-dire, il existe un processus

de structuration progressive du champ du possible vers une fermeture des possibilités. La notion de décideur isolé est inutilisable dans la plupart des cas.

Pour opérationnaliser cette approche de modélisation, nous avons fait recours aux méthodes de l'intelligence artificielle. Nous avons proposé un Système Multi-Agents cognitif avec une architecture organisationnelle à base de tableau noir et dont les agents sont modélisés avec les Séparateurs à Vaste Marge qui sont considérés comme l'un des modèles à avoir le plus marqué le domaine de la classification en fournissant un cadre théorique statistique et non plus connexionniste à la théorie de l'apprentissage. Ce classificateur est à même de prédire des frontières de décision complexes grâce à l'introduction de noyaux non linéaires.

Ce modèle a finalement été appliqué pour le dépistage automatique de la maladie de la scoliose qui n'était jusqu'alors abordé que par des méthodes traditionnelles trop lourdes. Les résultats obtenus sont très encourageants dès lors que le taux d'erreur d'apprentissage est de 8.46% avec une parfaite habileté de classification des cas réservé pour la phase de test. Ce modèle constitue un outil d'aide à la décision intéressant et pourra être mis à la disposition de différentes organisations en vue de l'élaboration d'autres applications.

Néanmoins, tout au long de ce travail, nous avons rencontré des difficultés qui ouvrent des perspectives de plusieurs ordres :

- L'abondance de l'information dans une société très consommatrice a changé le paradigme de la décision : on pouvait autrefois énoncer « Le pouvoir, c'est le savoir ». Seul un petit nombre d'élus qui « gouvernaient » savaient et pouvaient donc décider. Aujourd'hui, tout le monde est censé avoir accès à l'information, le paradigme tombe. Le véritable problème du décideur en quête de l'information pertinente n'est plus l'accès à cette information qui tend à être uniformément distribuée, mais de savoir la trouver dans d'imposants flux informationnels extrêmement bruités. Cependant, l'incapacité de l'homme à traiter l'ensemble des flux d'informations imprécises, incertaines, incomplètes et contradictoires nécessaire à la décision semble montrer que la solution pour une aide à la

décision efficace relève des systèmes de traitement de l'information. L'étude du processus de la décision doit donc inclure un véritable processus de traitement de l'information en particulier lorsque la situation n'est pas complètement mathématisable, ce qui est généralement le cas lorsque la décision relève d'une organisation ou d'un quelconque collectif.

– Un des problèmes récurrent en apprentissage machine, que nous avons rencontré au cours de ce travail, est de trouver la bonne représentation des données. En effet, pour que l'outil de simulation apprenne facilement le phénomène étudié et converge vers une des solutions de ce problème, il est souvent nécessaire de construire une représentation adéquate, en utilisant des transformations complexes des données brutes. La construction automatique d'une bonne représentation est encore un problème aujourd'hui dans tous les types d'apprentissage machine. De plus, chaque représentation trouvée est souvent propre à un seul problème, et donc pour chaque problème, il faut recommencer. Il est très difficile de construire une représentation pouvant servir de base solide à plusieurs tâches.

– Les Séparateurs à Vaste Marge ont bouleversé le monde de l'apprentissage artificiel. Malheureusement, un des problèmes de l'utilisation des SVM qui reste ouvert est lié au choix des paramètres des fonctions noyaux. Le choix des paramètres qui correspondent le mieux au problème n'est pas intuitif. N'ayant pas d'information a priori sur le problème, le classificateur entraîné est loin des performances escomptées. En effet, la méthode Grid-Search combinée avec une validation croisée permettent d'approcher les valeurs optimales de ces paramètres au détriment d'un temps de calcul très coûteux. Cette limitation deviendra majeure lorsqu'on applique les SVM sur des bases de données de grande tailles ou complexes et il faudra faire un compromis entre la pertinence de la classification et le temps d'exécution.

– La dernière difficulté que nous avons rencontrée est liée au problème d'implémentation des systèmes multi- agents, notamment pour un non-spécialiste des systèmes multi-agents. La difficulté de réalisation d'un système multi-agent s'explique principalement de par la diversité, la richesse des paradigmes multi-agents, et la complexité des concepts sous-jacents (coordination, interaction, organisation, etc.). Cette complexité

rend l'utilisation de la majorité des outils existants toujours difficile et même, parfois, presque impossible.

Ces difficultés ne constituent pas pour autant un grand handicap, tous ces problèmes font aujourd'hui l'objet de recherches dont nous pouvons espérer qu'elles apportent des réponses satisfaisantes dans les années qui viennent. En effet, l'intelligence artificielle est devenue un des terrains de bataille favoris de la philosophie de l'esprit. Cependant, il n'existe pas de paradigme unifié de recherche à part l'intérêt porté au mental et aux représentations de connaissances et certaines branches de l'intelligence artificielle sont devenues des terrains d'échanges multidisciplinaires où se côtoient philosophes, psychologues, informaticiens et autres qui s'intéressent aux diverses problématiques de l'intelligence.

Enfin, il convient de dire que l'une des perspectives sera donc de continuer un travail de veille sur des outils qui font leurs preuves dans les domaines de l'intelligence. Aujourd'hui, toute discipline est dans l'obligation de redéfinir ses problèmes en évitant à la fois ce qui est simple et inintéressant et ce qui est trop difficile. L'espoir en ce qui concerne les recherches en Intelligence Artificielle est de trouver des modèles plus riches qui tiennent à la fois compte de la richesse de la réflexion humaine (les décideurs), des événements historiques (la mémoire des cas) et de l'environnement complexe (les contraintes et stimuli externes) qui n'ont pas encore atteint le stade de concrétisation.

«Everything should be made as simple as possible, but not simpler»,

Albert Einstein

ANNEXE I

PERCEPTRON MULTI-COUCHE (MLP)

I.1 Réseau de neurones de type perceptron

Le perceptron est un type de réseau de neurones; il appartient à la famille générale des réseaux (non bouclé) ou encore (à propagation vers l'avant), c'est-à-dire qu'en mode normal, l'information se propage dans un sens unique, des entrées vers les sorties sans aucune rétroaction. Son apprentissage est de type supervisé par la règle de correction des erreurs [Parlzeau, 2004].

A.I.1.1 Perceptron simple

Le perceptron simple est constitué d'une seule couche de neurones dont les fonctions d'activation sont de type seuils. Il s'agit d'une seule couche de S neurones totalement connectée sur un vecteur p de R entrées. La matrice W de dimension $S \times R$ représente l'ensemble des poids de la couche. Le vecteur b de dimension $S \times 1$ désigne l'ensemble des S biais de la couche. Les niveaux d'activation $n = Wp - b = [n_1 \ n_2 \ \dots \ n_S]^T$ des neurones de la couche servent d'argument à la fonction d'activation qui applique un seuil au niveau 0 pour produire le vecteur des sorties $a = [a_1 \ a_2 \ \dots \ a_S]^T$, où:

$$a_i = \begin{cases} +1 & \text{si } n_i > 0 \\ -1 & \text{autrement} \end{cases} \quad \text{A.I.1}$$

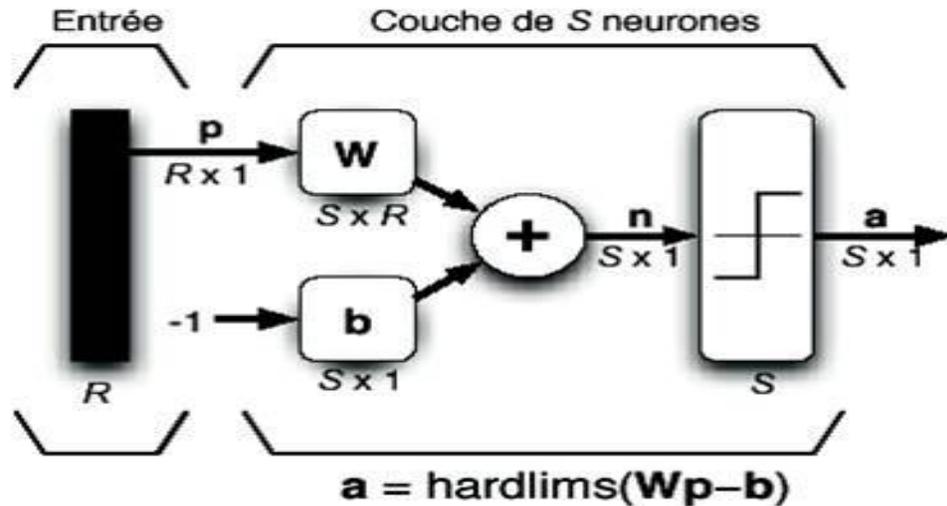


Figure A.I.1 : Perceptron à une couche avec fonction seuil [Parlzeau, 2004].

Pour $R = 2$ et $S = 1$, c'est-à-dire lorsque la couche n'est formée que d'un seul neurone relié à deux entrées: $\mathbf{p} = [p_1, p_2]^T$, $\mathbf{W} = [w_{11}, w_{12}]^T$, $\mathbf{b} = [b_1]$ et $\mathbf{a} = [a_1]$, où:

$$a_1 = \begin{cases} +1 & \text{si } w_{11} * p_1 + w_{12} * p_2 > b_1 \\ -1 & \text{autrement} \end{cases} \quad \text{A.I.2}$$

L'équation (A.I. 2) indique clairement que la sortie du réseau (neurone) peut prendre seulement deux valeurs distinctes selon le niveau d'activation du neurone, -1 lorsque ce dernier est strictement inférieur à 0 et +1 dans le cas contraire. Il existe donc dans l'espace des entrées une frontière délimitant deux régions correspondantes. Cette frontière est définie par la condition $w_{11} p_1 + w_{12} p_2 = b_1$ qui correspond à l'expression générale d'une droite, telle qu'illustrée à la Figure A.II.1 [Parlzeau, 2004].

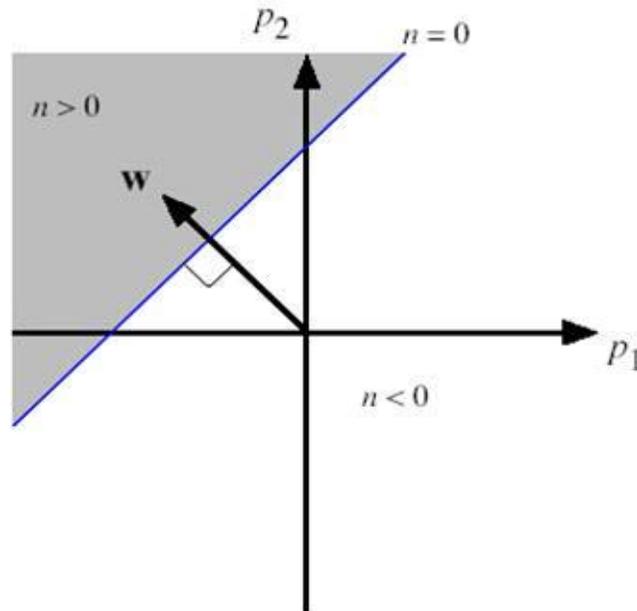


Figure A.I.2 : Frontière de décision pour un perceptron simple à 1 neurone et 2 entrées [Parlzeau, 2004; , 2004.]

En considérant le cas ($S > 1$), chaque neurone i possède son propre vecteur de poids et son propre biais, on obtient S frontières de décision distinctes. Toutes ces frontières de décision sont linéaires et permettent chacune de découper l'espace d'entrée en deux régions infinies, de part et d'autre d'une droite. Chaque neurone d'un perceptron simple permet donc de résoudre parfaitement un problème de classification à deux classes, à condition que celles-ci soient linéairement séparables. Pour cette raison le perceptron simple est souvent appelé (séparateur linéaire).

Comme nous l'avons dit précédemment, l'apprentissage du perceptron simple est de type supervisé par correction des erreurs. Soit les Q couples d'apprentissages:

$$\{(p_1, d_1), (p_2, d_2), \dots, (p_q, d_q), (p_Q, d_Q)\}$$

Pour pouvoir déterminer les poids et les biais du réseau permettant de classer au mieux les Q couples d'apprentissage, une règle d'apprentissage très simple a été développée. Cette règle consiste à définir un signal d'erreur e , tel que:

$$e = \frac{d - a}{2} \quad e \in \{-1, 0, +1\} \quad \text{A.I.3}$$

Pour chaque stimulus p_q les modifications synaptiques vont se faire de la manière suivante:

$$\Delta w = \begin{cases} p & \text{si } e = +1 \\ 0 & \text{si } e = 0 \\ -p & \text{si } e = -1 \end{cases} \quad \text{A.I.4}$$

Dans le cas où $b \neq 0$, on peut aussi mettre à jour le biais en observant simplement que celui-ci n'est rien d'autre qu'un poids comme les autres, mais dont l'entrée est fixée à -1 (entrée virtuelle). Ainsi: $\Delta b = -e$.

Dans le cas général où il y a S neurones, cette règle peut être réécrite sous forme matricielle de la façon suivante:

$$\begin{aligned} \Delta w &= ep^T ; \\ \Delta b &= -e; \end{aligned} \quad \text{A.I.5}$$

où $e = d - a = [e_1, e_2, \dots, e_S]^T$ est le vecteur des erreurs observé en sortie pour le stimulus p , Cette règle converge toujours en un nombre fini d'itérations et permet de trouver un hyperplan (plan dans un espace à n dimensions) qui sépare ou classe en deux classes les entrées du perceptron (si ces derniers sont linéairement séparables).

A.I.1.2 Perceptron multi couche «*multi layer perceptrone.*» PMC [Dreyfus, 2004]

En traitant les perceptrons à une seule couche de neurones, on a constaté que ces derniers ne pouvaient résoudre que des problèmes de classification linéairement séparables.

Les réseaux multicouches permettent de lever cette limitation. Il est même démontré qu'avec un réseau de trois couches de neurones (02 couches cachées), on peut construire des frontières de décision de complexité quelconque, ouvertes ou fermées, concaves ou convexes, à condition d'employer une fonction de transfert non linéaire et de disposer de suffisamment de neurones sur les couches cachées [Parlzeau, 2004].

Un réseau multicouche n'est rien d'autre que la mise en cascade de perceptrons simples (à une seule couche), en prenant les sorties d'une couche et en les injectant comme les entrées de la couche suivante.

La combinaison de plusieurs couches linéaires peut toujours se ramener à une seule couche linéaire équivalente. C'est pourquoi, pour être utile, un réseau multicouche doit toujours posséder des neurones avec fonctions de transfert non linéaires sur ses couches cachées. Sur sa couche de sortie, selon le type d'application, il pourra comporter des neurones linéaires ou non linéaires [Parlzeau, 2004].

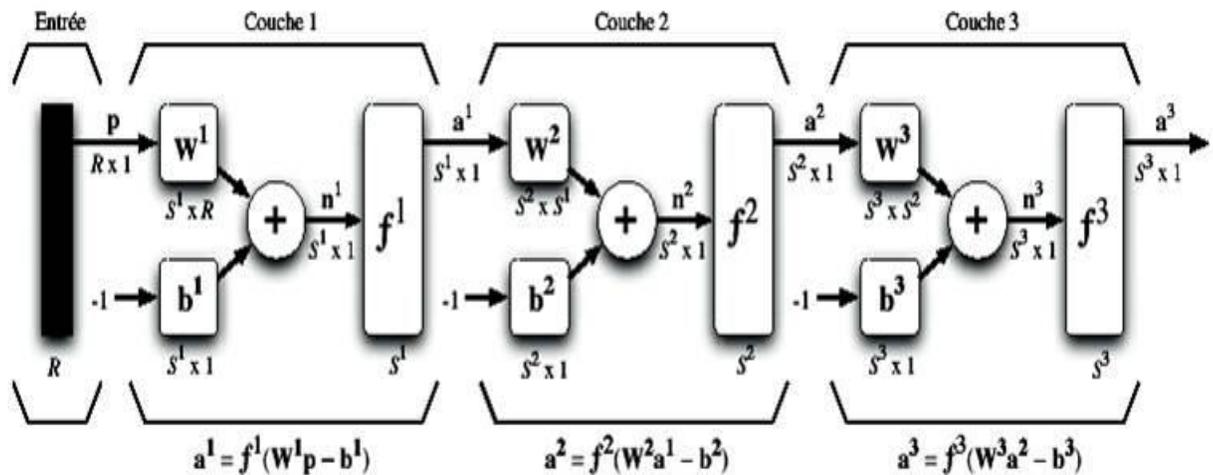


Figure A.I.3 : Représentation matricielle d'un réseau à trois couches [Howard et Beale, 2000].

A.I.2 Rétropropagation des erreurs:

L'algorithme de rétropropagation utilise comme indice de performance l'erreur quadratique moyenne, et permet un apprentissage de type supervisé avec un ensemble de Q association stimulus/cible. Mais avant de passer à l'algorithme de rétropropagation pour l'entraînement du *MLP*, il faut mentionner que ce n'est pas par hasard qu'on remplace la fonction de transfert seuil par la fonction sigmoïde, mais bien pour pouvoir procéder à un apprentissage automatique. Le problème avec la fonction seuil est que sa dérivée est toujours nulle sauf en un point où elle n'est même pas définie. Elle ne peut donc pas être utilisée avec la méthode de la descente du gradient qui va servir pour développer la méthode de la rétropropagation [Bertels, Neuberg et Vassiliadis, 2001].

La méthode de la rétropropagation des erreurs peut être résumée en un algorithme simple qui indique les démarches à suivre pour entraîner un MPL:

1. Initialiser tous les poids du réseau à des valeurs aléatoires (de préférence à de petites valeurs).
2. Pour chaque association (p_q, d_q) dans la base d'apprentissage:
 - a. Propager les entrées p_q vers l'avant à travers les couches du réseau:

$$a^0 = p_q$$

$$a^k = f^k(w^k a^{k-1} - b^k) \quad \text{pour } k = 1, \dots, M$$

- b. Rétropropager les sensibilités $(S_i^k = \frac{\partial F}{\partial n_i^k})$ vers l'arrière à travers les couches du réseau:

$$s^M = 2[f^M]'(n^M)(d - a^M)$$

$$s^k = [f^k]'(n^k)(w^{k+1})s^{k+1} \quad \text{pour } k = M - 1, \dots, 1$$

- c. Mettre à jour les poids et biais:

$$\Delta w^k(t) = -\eta s^k(t)(a^{k-1})^T(t) \quad \text{pour } k = 1, \dots, M$$

$$\Delta b^k(t) = \eta s^k(t) \quad \text{pour } k = 1, \dots, M$$

3. Si le critère d'arrêt est vérifié, alors STOP.
4. Sinon, permuter l'ordre de présentation des associations de la base d'apprentissage et aller à l'étape 2 [Parlzeau, 2004].

Plusieurs critères d'arrêts peuvent être utilisés avec l'algorithme de rétropropagation des erreurs. Le plus commun consiste à fixer un nombre maximum de périodes d'entraînement, ce qui fixe effectivement une limite supérieure sur la durée de l'apprentissage. Un deuxième critère commun consiste à fixer une borne inférieure sur l'erreur quadratique moyenne.

Lorsque l'indice de performance choisi diminue en dessous d'un objectif à atteindre, on considère simplement que le *MLP* a suffisamment bien appris ses données et on arrête l'apprentissage [Parlzeau, 2004].

À chaque période d'entraînement, il importe de permuter l'ordre de présentation des stimuli pour réduire la probabilité qu'une séquence de données garde le réseau dans d'un

minimum local. En effet, la performance de la méthode de la descente du gradient peut dépendre grandement de cet ordre de présentation qui engendre des trajectoires différentes dans l'espace des paramètres, et des trajectoires différentes peuvent conduire à des minimums locaux différents. Donc le fait de permuter les données à chaque période nous garantit que l'on ne tombera pas systématiquement sur les mêmes minima [Parlzeau, 2004].

ANNEXE II

RÉSEAUX DE NEURONES DE KOHONEN ET KNN

A.II.1 Réseau de neurones de Kohonen

Le réseau de neurones de *kohonen* appelé aussi la carte auto organisatrice (SOM ou la carte *kohonen*) a été introduit par *Teuvo, Kohonen*, ce type de réseau de neurone est largement utilisé dans l'analyse des données comme la reconnaissance de forme et la classification des signaux biomédicaux [Mezghani et al., 2008]; Il s'agit de réseaux non supervisés avec un apprentissage compétitif qui permettent de modéliser l'espace des entrées avec des prototypes et construire une carte à une ou deux dimensions afin de structurer cet espace [Parlzeau, 2004]. En effet, une carte auto-organisatrice, comme son nom l'indique, s'organise par rapport aux stimuli présentés en respectant les contraintes topologiques de l'espace d'entrée. Il y a mise en correspondance de l'espace d'entrée avec l'espace du réseau. Les zones voisines de l'espace d'entrée sont voisines sur la carte auto-organisatrice [Touzet, 1992].

Dans notre cas le vecteur d'entrée est un ensemble des points équidistants dans l'espace tridimensionnel chaque points est représenté par les cordonnées (x,y,z) . Soit le vecteur d'entrée $X=(X_1, X_2, \dots, X_I)$ et $W_j=(w_{j1}, w_{j2}, \dots, w_{jI})$ l'ensemble des vecteurs de poids de dimension I stockés aux nœuds j , avec $j=1, \dots, J$, comme montre la Figure A.II.1

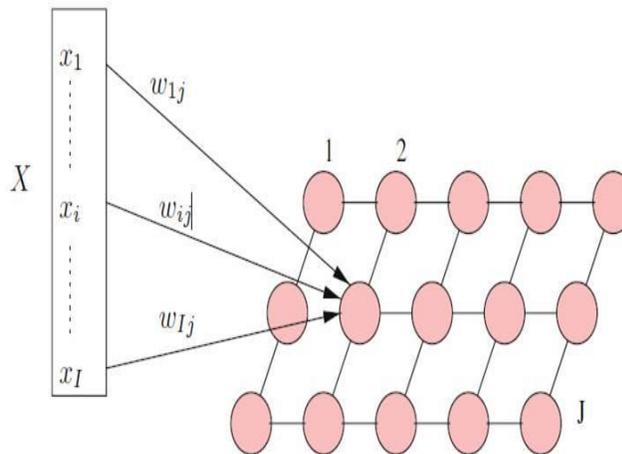


Figure A.II.1 : la mémoire de *kohonen* en deux dimensions avec J est le nombre de nœuds [Mezghani et al., 2008].

Le vecteur X d'entrée avec L est le nombre de points équidistants est comme suit $(x_1, \dots, x_L, y_1, \dots, y_L, z_1, \dots, z_L)$, L'algorithme de l'apprentissage de *kohonen* est basé sur l'apprentissage compétitif. Après l'initialisation des poids avec des petites valeurs, le processus d'apprentissage consiste à trouver le nœud j^* qui contient le vecteur de poids le plus proche de vecteur X présenté à l'entrée; Le carré de la distance entre un vecteur de poids W_j de la carte de *kohonen* et le vecteur d'entrée X est calculé comme suit:

$$d(X, W_j)^2 = \sum_{i=1}^L (x_i - w_{ji})^2 + (y_i - w_{j(L+i)})^2 + (z_i - w_{j(2L+i)})^2 \quad \text{A.II.1}$$

avec: $w_{ji}=1, \dots, L$, sont les poids en nœud j correspondant à la coordonnée x, $w_{j(L+i)}$, sont les poids en nœud j correspondant à la coordonnée y, $w_{j(2L+i)}$, sont les poids en nœud j correspondant à la coordonnée z.

À l'itération n, les vecteurs de poids à chaque nœud j de la mémoire sont ajustés en ajoutant un montant qui est une fonction de la distance de grille au nœud j^* le nœud qui contient le vecteur de poids le plus proche de vecteur d'entrée

$$w_{ji}^{(n+1)} = w_{ji}^n + \epsilon_n \cdot h_n^{j,j^*} \cdot (x_i^n - w_{ji}^n) \quad \text{A.II.2}$$

$$\epsilon_n = \epsilon_1 \cdot \left(\frac{\epsilon_2}{\epsilon_1}\right)^{\frac{n}{n_{max}}} \quad \text{A.II.3}$$

$$\delta_n = \delta_1 \cdot \left(\frac{\delta_2}{\delta_1}\right)^{\frac{n}{n_{max}}} \quad \text{A.II.4}$$

$$h_n^{j,j^*} = \text{EXP} - \frac{\|j - j^*\|^2}{2\delta_n^2} \quad \text{A.II.5}$$

Le facteur h^{j,j^*} appelé aussi la fonction de voisinage, agit comme un noyau de lissage et il définit l'influence du nœud j^* sur le nœud j au cours de l'ajustement des poids dans le nœud j, il diminue lorsque la distance de la grille entre le nœud j^* et les nœuds j augmente. Il

dépend aussi d'un paramètre σ qui diminue avec le nombre d'itérations entre les valeurs σ_1 et σ_2 , respectivement, la valeur initiale et la valeur finale (équation A.II.5).

Le paramètre ϵ varie avec le nombre d'itération avec ϵ_1 est la valeur initiale et ϵ_2 est la valeur finale (Eq A.II.9). Les paramètres δ_1 , δ_2 , ϵ_1 et ϵ_2 doivent être choisis de manière appropriée pour assurer la convergence de l'algorithme. Les valeurs des paramètres utilisés dans les expériences sont: $\delta_1=1$, $\delta_2=0,002$, $\epsilon_1=1$ et $\epsilon_2=0,0005$.

A.II.2 Classificateur (K plus proches voisins *K-NN*)

Plus connu en anglais sous le nom *K-nearest neighbor (K-NN)*, l'algorithme cherche les K plus proches voisins. la classification est effectuée par un vote majoritaire de son voisinage et l'objet est classé dans la classe qui est plus présente parmi les k plus proche voisins. la mesure la plus utilisée est la distance euclidienne, définie par [Gupta, Patil et Samanta., 2011].

$$d(p, q) = \sqrt{(p_i - q_i)^2} \quad \text{A.II.6}$$

ANNEXE III

VALEURS DES PARAMÈTRES C & σ DES SVM OPTIMAUX

SVM	C	Sigma	SVC	Erreur Test
1	2,16E-05	6,23E+05	2,15E-01	6,67E-02
2	2,57E-05	6,23E+05	2,15E-01	6,67E-02
3	4,32E-05	6,23E+05	2,15E-01	6,67E-02
4	5,13E-05	6,23E+05	2,15E-01	6,67E-02
5	6,10E-05	6,23E+05	2,15E-01	6,67E-02
6	7,26E-05	6,23E+05	2,15E-01	6,67E-02
7	8,63E-05	6,23E+05	2,15E-01	6,67E-02
8	1,03E-04	6,23E+05	2,15E-01	6,67E-02
9	1,22E-04	6,23E+05	2,15E-01	6,67E-02
10	1,45E-04	6,23E+05	2,15E-01	6,67E-02
11	1,73E-04	6,23E+05	2,15E-01	6,67E-02
12	2,05E-04	6,23E+05	2,15E-01	6,67E-02
13	2,44E-04	6,23E+05	2,15E-01	6,67E-02
14	2,90E-04	6,23E+05	2,15E-01	6,67E-02
15	3,45E-04	6,23E+05	2,15E-01	6,67E-02
16	4,11E-04	6,23E+05	2,15E-01	6,67E-02
17	4,88E-04	6,23E+05	2,15E-01	6,67E-02
18	5,81E-04	6,23E+05	2,15E-01	6,67E-02
19	6,91E-04	6,23E+05	2,15E-01	6,67E-02

20	1,64E-03	6,23E+05	2,15E-01	6,67E-02
21	1,25E-01	1,77E-01	2,00E-01	0,00E+00
22	1,49E-01	1,77E-01	1,92E-01	0,00E+00
23	1,77E-01	1,77E-01	1,92E-01	0,00E+00
24	2,10E-01	1,77E-01	1,92E-01	0,00E+00
25	2,50E-01	1,77E-01	1,92E-01	0,00E+00
26	2,97E-01	1,77E-01	1,92E-01	0,00E+00
27	3,54E-01	7,43E-02	1,85E-01	1,00E-01
28	4,20E-01	7,43E-02	1,77E-01	1,33E-01
29	5,00E-01	6,25E-02	1,77E-01	1,33E-01
30	2,83E+00	1,05E-01	1,77E-01	1,33E-01
31	1,81E+02	4,20E-01	1,62E-01	1,33E-01
32	2,15E+02	4,20E-01	1,62E-01	1,33E-01
33	3,62E+02	5,00E-01	1,46E-01	1,33E-01
34	4,31E+02	5,00E-01	1,46E-01	1,33E-01

LISTE DES RÉFÉRENCES BIBLIOGRAPHIQUES

- Azé, J. 2008. « Classification supervisée : Perceptron et SVM ». *Notes de cours, Laboratoire de Recherche en Informatique, Université Paris Sud, France.*
- Ben-Ishak, A. 2007. « Sélection de variables par les machines à vecteurs supports pour la discrimination binaire et multi classe en grande dimension ». *Thèse de doctorat, Faculté des Sciences de Luminy, Université d'Aix-Marseille II, France.*
- Bertels, K., L. Neuberg et S. Vassiliadis. 2001. « On Chaos and Neural Networks: The Backpropagation Paradigm. *Artificial Intelligence Review* ».
- Bertrand, S. 2005. « Modélisation géométrique 3D in vivo du tron humain à partir de l'imageur basse dose EOS™ ». *Thèse de doctorat, École National des Arts et Métiers, Paris, France.*
- Boser, A. , I. Guyon et V.N. Vapnik. 1992. « A Training Algorithm for Optimal Margin Classifiers ». *The Fifth Annual Workshop on Computational Learning Theory, ACM, Pittsburgh.*
- Callut, J. 2003. « Implémentation Efficace des Support Vector Machines pour la Classification ». *Thèse de doctorat, Département d'Informatique, Université Libre de Bruxelles, Belgique.*
- Champlain, N. 2004. « Recherche des facteurs biomécaniques dans l'aggravation des scolioses idiopathiques ». *Thèse de doctorat, École National des Arts et Métiers, Paris, France.*
- Chatelain, C. 2003a. « Les Multi-Layer Perceptron ». *1^{er} Rapport Interne.*
- Chatelain, C. 2003b. « Les Support Vector Machine (SVM) ». *2^{eme} Rapport Interne.*
- Chih-Wei, H. , C. Chih-Chung et L. Chih-Jen. 2010. « A Practical Guide to Support Vector Classification ». *Department of Computer Science National Taiwan University, Taipei 106, Taiwan.*
- Clancy, W.J. . 1992. « The Frame of Reference Problem in the Design of Intelligent Machines ». *Architectures for Intelligence, Proceedings of the 22nd Carnegie Symposium on Cognition, New Jersey, USA.*
- Cortes, F., A. Przeworski et J. Sprague. 1974. « Systems Analysis for Social Scientists. ». *Wiley, New York, USA.*

- Daniel, K.S. 1996. « Modélisation de la démarche du décideur politique dans la perspective de l'intelligence artificielle ». *Thèse de doctorat, Faculté des sciences économiques et sociales de l'Université de Genève, Suisse.*
- Deheeger, F. 2008. « Couplage mécano-fiabiliste : SMART – méthodologie d'apprentissage stochastique en fiabilité ». *Rapport Interne, Université Blaise Pascal et Institut Français de Mécanique Avancée, Clermont, France.*
- Delley, J.D., R. Derivaz, L. Mader, C.A. Morand et D. Schneider. 1982. « Le Droit en Action ». *Livre, St-Saphorin, Georgi.*
- Doerner, D. . 1987. « On the Logic of Failure: Thinking, Planning and Decision Making in Uncertainty and Complexity ». *Bamberg University, Lehrstuhl Psychologie II, Allemagne.*
- Dreyfus, G. 2004. « Réseaux de neurones : méthodologie et application ». *Ed. Eyrolles.*
- Duin, R.P.W., P. Juszczak, P. Paclik, E. Pekalska, D. De-Ridder, D.M.J. Tax et S. Verzakov. 2007. « PRTools4.1, A Matlab Toolbox for Pattern Recognition ». *Delft University of Technology.*
- Essid, S. 2005. « Classification automatique des signaux audio-fréquences : reconnaissance des instruments de musique ». *Université Pierre et Marie Curie, Paris, France.*
- Ferber, J. 1995. « Les systèmes multi-agents. vers une intelligence collective ». *Inter Editions, Paris, France.*
- Feydy, A. , S. Ferey, V. Merzoug, H. Guerini, A. Chevrot, J. Dubousset, Kalifa G. et Drape J.L. 2008. « Imagerie de la scoliose. Place du système EOS™ ». *Société d'imagerie musculo-squelettique. Congrès Thématique Paris, France*
- Fox, M. 1981. « An Organizational View of Distributed Systems ». *IEEE Transaction on Systems, Man, and Cybernetics, SMC-11, p. 70-80.*
- Galliers, J.R. 1988. « A Theoretical Framework for Computer Models of Cooperative Dialog. Acknowledging Multi-Agent Conflict ». *Open University, United Kingdom.*
- Germain, P. 2009. « Algorithmes d'apprentissage automatique inspirés de la théorie PAC-Bayes ». *Faculté des Sciences et de Génie, Université Laval, Québec, Canada.*
- Guessoum, Z. 2003. « Modèles et Architectures d'Agents et des Systemes Multi-Agents Adaptatifs ». *Université Pierre et Marie Curie, Paris, France*

- Gupta, S. P., V. M. Patil et S. Samanta. 2011. « A 3D-QSAR Study on a Series of Benzimidazole Derivatives Acting as Hepatitis C Virus Inhibitors: Application of kNN-Molecular Field Analysis ». *Med Chem*.
- Howard, D., et M. Beale. 2000. « Neural Network Toolbox For use with MATLAB Version 4.Ed. THE MATHWORKS. ».
- Huet, S. 2004. « Classification de Signaux Myo-électriques ». *Édition Simbad*.
- Hull, C. . 1952. « A behavior system ». *Yale University Press, New Haven, USA*.
- Humbert, L , J.A. Dde-Guise, B. Aubert, B. Godbout et W. Skalli. 2009. « 3D reconstruction of the spine from biplanar X-rays using parametric models based on transversal and longitudinal inferences ». *Med Eng Phys*.
- Hurtado, J.E. , et D.A. Alvarez. 2001. « Neural-network-based reliability analysis: a comparative study ». *Computer Methods in Applied Mechanics and Engineering*.
- Kaasra, I., et M. Boyd. 1995. « Designinig a neural network for forecasting financial and economic time series ». *Elsevier & Neurocomputing*.
- Kabachi, N., L. Vincent et Y. Ouzrout. 1996. « Une société d'agents pour la prise de décision dans les organisations productives ». *Actes des journées Francophones d'IA Distribuée et SMA*.
- Kebair, F. 2009. « Modélisation multi-agent de postes de commande coordonnateurs de prise de décisions stratégiques ». *Thèse de doctorat, Université du Havre & Laboratoire d'Informatique de Traitement de l'Information et des Systèmes, France*.
- Kelemen, A. , Y. Liang et S. Franklin. 2002. « A Comparative Study of Different Machine Learning Approaches for Decision Making ». *The Institute for Intelligent Systems, The University of Memphis, Memphis, USA*.
- Labidi, S. , et W. Lejouad. 2004. « De l'Intelligence Artificielle Distribuée aux Systèmes Multi-Agents ». *Institut National De Recherche En Informatique Et En Automatique, France*.
- Ladjadj, R. 2003. « Les réseaux de neurones ». *Ed. Ingénieur*
- Le-Meur, N. 2005. « De l'Acquisition des Données de Puces à ADN vers leur Interprétation : Importance du Traitement des Données Primaires ». *Thèse de doctorat, Université de Nantes, Ecole doctorale chimie biologie, France*
- Leray, P. 2001. « Quelques Types de Réseaux de Neurones : La Rétropropagation ». *INSA Rouen -Département ASI - Laboratoire PSI, France*.

- Malville, E. 1999. « L'auto-organisation de groupes pour l'allocation de tâches dans les Systèmes Multi-Agents, Application à CORBA ». *Thèse de doctorat, Université De Savoie, Ecole doctorale d'informatique, France.*
- Marty, C. 2003. « Comment examiner un trouble de la statique rachidienne ? ». *Revue du Rhumatisme 71*
- Mary, P. 2004. « La scoliose idiopathique de l'enfant et de l'adolescent ». *Revue : Les troubles de la statique, Médecine thérapeutique, Pédiatrie vol. 7.*
- Meadows, D.H., et J.M. Robinson. 1985. « The electronic oracle: computer models and social decisions ». *John Wiley, Chichester, NewYork, USA.*
- Meghoufel, Ali. 2004. « Extraction semi-automatique des points de référence sur une Paire de radiographies à rayons X ». *Master, Département du Génie de la Production Automatisée, École de Technologie Supérieure, Montréal, Canada.*
- Mezghani, N., R. Chav, L. Humbert, S. Parent, W. Skalli et J. De Guise. 2008. « A computer-based classifier of three-dimensional spinal scoliosis severity ». *International Journal of Computer Assisted Radiology and Surgery*, p. 55-60.
- Mirmehdi, M. , A. Osareh, T. Barry et R. Markham. 2003. « Comparative Exudate Classification using Support Vector Machines and Neural Networks ». *Department of Computer Science, University of Bristol, Bristol*
- Morin, E. 1990. « Introduction à la pensée complexe ». *ESF, Paris, France.*
- Noomen, H. 1998. « Outil de communication pour les systèmes multi-agents ». *Université De Sherbrooke, Département de Génie Electrique et de Génie Informatique, Québec, Canada.*
- Noubel, J.F. 2007. « Intelligence Collective, la révolution invisible ». *The Transistioner.*
- Osowski, S. , K. Siwek et T. Markiewicz. 2004. « MLP and SVM Networks – a Comparative Study ». *Proceedings of the 6th Nordic Signal Processing Symposium NORSIG, Finland*
- Oussar, Y. 1998. « Réseaux d'ondelettes et réseaux de neurones pour la modélisation statique et dynamique de processus ». *Thèse de doctorat, Université Pierre Et Marie Curie, Ecole doctorale Robotique, Paris, France.*
- Parizeau, M. 2004. « Réseaux de neurones ». *Rapport Interne, Université LAVAL, Ville de Québec, Canada.*

- Parlzeau, M. 2004. « Réseaux de neurones ». *Rapport Interne*.
- Parlzeau, M. 2004. « Réseaux de neurones. ».
- Plantie, M. 2006. « Extraction automatique de connaissances pour la décision multicritère ». *Thèse de doctorat, Ecole Nationale Supérieure des Mines de Saint-Etienne, France*
- Sfez, L. . 1992. « Critique de la décision ». *Presses de la Fondation des sciences politiques, 4^{ème} édition (édition originale:1973)*.
- Sholkopf, B. , et A.J. Smola. 2002. « Learning with kernels ». *The MIT Press, Cambridge, UK*.
- Simon, H. A. 1981. « Cognitive Science: The Newest Science of the Artificial ». *Cognitive Science*, 4:33–46.
- Taleb-Ahmed, A. , et H. Djadel. 2002. « SMOWR: Sequential Minimal Optimization with Weighted Ray ».
- Tarantino, C., A. D'addabbo et L. Castellana. 2005. « Neural Network Ensemble and Support Vector Machine Classifiers: an application to remote sensed data ». *ISSIA-CNR, Via Amendola, Bari (Italy)*.
- Thanh-Nghi, Do. , et J.D. Fekete. 2007. « Large Scale Classification with Support Vector Machine Algorithms ». *Sixth International Conference on Machine Learning and Applications, ICMLA*
- Touzet, C. 1992. « Les réseaux de neurones Artificiels Introduction au Connexionnisme ».
- Turenne, N. 2006. « Apprentissage automatique : SVM ». *INRA, France*.
- Tyagi, S. 2008. « A Comparative Study of SVM Classifiers and Artificial Neural Networks Application for Rolling Element Bearing Fault Diagnosis using Wavelet Transform Preprocessing ». *Proceedings of world academy of science, engineering and technology*, vol. 33.
- Vapnik, V. . 1998. « Statistical Learning Theory ». *Wiley-Interscience*.
- Weber, A. 2007. « Modélisation et Gestion de Flux par Systèmes Multi-agents, Application à un Système d'Aide à la Décision en Epidémiologie ». *Thèse de doctorat, Ecole Centrale De Lille, Ecole doctorale d'Automatique & Informatique Industrielle*.
- Wiener, N. . 1961. « Cybernetics or Control and Communication in the Animal and the Machine ». *MIT Press, Cambridge, USA*.

- Wikipédia. 2011. « L'ensemble de vertèbres sur le rachis ». *Consulté en décembre 2011*.
- Wisniewski, G. 2007. « Apprentissage dans les espaces structurés, Applications à l'étiquetage de séquences et à la transformation automatique de documents ». *Thèse de doctorat, Université Pierre et Marie Curie, Paris, France*.
- Wooldridge, M., et N.R. Jennings. 1994. « Intelligent Agents : Theory and Practice ». *Rapport Interne*, vol. 10.
- Xinghuo, Y. , M. Onder-Efe et K. Okyay. 2002. « A General Backpropagation Algorithm for Feedforward Neural Networks ». *IEEE transactions on neural networks*, vol. 13.
- Zaier, I. 2008. « Modélisation de l'épaisseur de glace des lacs par les Réseaux de Neurones Artificiels ». *Université d'avant-garde, INRS, Centre Eau Terre Environnement*.
- Zidi, K. 2006. « Système Interactif d'Aide au Déplacement Multimodal (SIADM) ». *Thèse de doctorat, Université des Sciences et Technologies de Lille, Ecole Doctorale d'Automatique et Informatique Industrielle, France*.