

*Université Hassan 1^{er}
Faculté des Sciences Juridiques
Economiques et Sociales de Settat
Laboratoire de Modélisation Mathématique
et Calcul Economique*

*Université Mohammed V-Agdal
Faculté des Sciences de Rabat
Le Laboratoire de Recherche
en Informatique et Télécommunications*

**Habilitation à Diriger des Recherches
Présenté par Pr. Abdeljalil ELOUARDIGHI
Spécialité Informatique
Doctorat National de la Faculté des Sciences de Rabat
Professeur Assistant à la FSJES de Settat**

**Sélection de variables pertinentes dans un processus d'extraction de
connaissance à partir de grande base de données**

Soutenue le 17 Septembre 2011 devant la commission d'examen :

Mr. Ahmed Hammouch : PES à l'ENSET de Rabat (Président)

Mr. Rachid Oulad Haj Thami PES à l'ENSIAS de Rabat (Rapporteur externe)

Mr. Boujemâa Achchab PES à la FSJES de Settat (Rapporteur interne)

Résumé :

Dans la cadre de la présentation du rapport d'activité en vue de l'obtention de l'habilitation à diriger des recherches en Informatique. Nous avons présenté un document réparti en deux volets. Dans le premier volet, nous avons présenté l'ensemble des travaux pédagogiques (Cours, TD et TP) que nous assurons depuis notre recrutement à la Faculté des Sciences Juridiques, Economiques et Sociales de Settat (FSJES). Sont exposés dans cette partie les manuels et photocopiés rédigés individuellement, les enseignements dispensés dans l'établissement d'accueil et dans d'autres établissements publics et Ecole d'Ingénieur, et dans différents parcours (Deug, Licence, Master et Cycle Ingénieur).

Le deuxième volet concerne les travaux de recherche, dans lequel on retrouve l'ensemble des travaux réalisés et leurs contextes.

Le travail de recherche présenté constitue une contribution dans le domaine du Datamining ou extraction des connaissances à partir de grande base de données. En particulier, ces travaux traitent le problème de la sélection des variables qui est une étape cruciale dans un processus de Datamining.

Tous Ces travaux ont été réalisés dans le Laboratoire de Recherche en Informatique et Télécommunication (LRIT) de la Faculté des Sciences de Rabat et au sein du Laboratoire de Modélisation Mathématique et Calcul Economique de la FSJES de Settat.

Ces travaux de recherches peuvent être synthétisés en quatre volets. Dans un premier volet nous avons testé et analysé le critère statistique Lambda de Wilks, développé en analyse discriminante, pour sélectionner les variables les plus pertinentes. Nous avons montré que ce critère permet d'améliorer les performances d'un algorithme de classification.

Dans un deuxième volet, nous nous sommes intéressés au problème de sélection de variables en classification supervisée. Nous avons présenté un nouveau critère de sélection basé sur le gain d'interaction qui prend en considération l'interaction entre les variables. Les tests que nous avons menés ont montré que le critère proposé donne, en général, des résultats encourageant, comparés avec d'autres critères de sélections de variables.

Dans un troisième volet, nous nous sommes intéressés à la sélection de gènes les plus pertinents, utilisant les données ADN. Notre objectif était de trouver un ensemble de gènes qui discrimine le mieux ces individus. Nous avons testé et évalué une nouvelle méthode de sélection de gènes pour les données ADN combinant la méthode MRMR (Minimum de Redondance -Maximum de Pertinence) et les Algorithmes Génétiques (GA). Dans une première étape, MRMR est utilisé pour filtrer les gènes redondants. Dans la deuxième étape, l'AG utilise la précision d'une fonction fitness (fonction d'évaluation) pour sélectionner les gènes les plus discriminants. La comparaison de la méthode MRMR-GA avec d'autres méthodes utilisées pour la sélection des gènes, a montré que notre méthode est capable de trouver un plus petit sous-ensemble de gènes et qui donne, dans la plus part des cas, la meilleure précision de classification.

Dans un quatrième volet nous avons utilisé la sélection de variable dans le cas d'un problème de reconnaissance de visage. L'objectif est de trouver le plus petit sous-ensemble de variables (caractéristiques) qui maximise la capacité de reconnaissance de visage. Dans cette étude, nous avons proposé une nouvelle théorie de l'information pour la sélection de caractéristiques, nommé Ponderated Mutual Information (PMI), qui prend en compte l'interaction entre les variables. Des performances significatives ont été obtenues en appliquant cette méthode pour la reconnaissance de visages.

Ces travaux de recherche ont donné lieu à de nombreuses publications dans des journaux internationaux et dans des congrès et colloques nationaux et internationaux.