

Table des matières

Introduction générale	i
I Amélioration du <i>Boosting</i> face aux données bruitées	1
1. Introduction	4
2. Contexte des données bruitées	6
2.1. Notations	6
2.2. Le concept de bruit	7
3. Problématique et motivation	9
3.1. Evaluation de l'apprentissage	9
3.2. La décomposition biais-variance	11
3.3. Les raisons de l'efficacité de l'agrégation de classifieurs	12
4. Les méthodes d'agrégation de classifieurs	14
4.1. Agrégation des classifieurs existants	14
4.2. Stratégies globales utilisant des mécanismes aléatoires	14
4.3. Stratégies globales adaptatives	16
4.4. Comparaison des méthodes d'agrégation	18
5. Le <i>Boosting</i> : faiblesses et améliorations	18
5.1. Le sur-apprentissage	19
5.2. La vitesse de convergence	24
6. Contributions : Amélioration du <i>Boosting</i> face aux données bruitées (Glouton-Boost, AdaBoost Hybride)	26
6.1. GloutonBoost	26
6.2. Explication de GloutonBoost	27
6.3. AdaBoost Hybride	29
6.4. Explication d'AdaBoost Hybride	30
7. Performances de GloutonBoost et AdaBoost Hybride	31
7.1. Comparaison avant bruitage	32
7.2. Comparaison sur des données bruitées	36
7.3. Comparaison de la vitesse de convergence	37

8.	Conclusion et perspectives	39
II	Classification associative adaptative et déséquilibre	41
1.	Introduction	44
2.	Contexte : le déséquilibre des classes	45
2.1.	Présentation des données déséquilibrées	45
2.2.	Méthodes pour les données déséquilibrées	46
2.3.	Le <i>Boosting</i> face aux données déséquilibrées	50
3.	Apprentissage de règles	51
3.1.	Présentation de l'apprentissage par induction de règles	51
3.2.	Les règles d'association prédictives	54
3.3.	Cadre général de l'apprentissage à base de règles	56
4.	Proposition d'un modèle à base de règles adapté au <i>Boosting</i> : phase d'extraction de règles	62
4.1.	Les algorithmes d'extraction des itemsets fréquents	62
4.2.	FP-Growth	63
4.3.	Contribution 1 : adaptation de FP-Growth à la classification supervisée (FCP-Growth)	65
4.4.	Contribution 2 : amélioration de la qualité des règles extraites (FCP-Growth-P)	69
4.5.	Etude de la complexité de FP-Growth et de FCP-Growth-P	71
4.6.	Performances de FCP-Growth et FCP-Growth-P	72
5.	Proposition d'un modèle à base de règles adapté au <i>Boosting</i> : phase de prédiction	77
5.1.	Contribution 3 : prédiction à partir d'une base de règles significatives pondérées (W-CARP)	77
5.2.	Performances de W-CARP	79
6.	Contribution 4 : CARBoost, un algorithme de classification associative adaptative	84
6.1.	Intérêt des méthodes ensemblistes	84
6.2.	Pour une classification associative adaptative : CARBoost	85
6.3.	Les performances de CARBoost face aux méthodes usuelles	88
6.4.	Les performances de CARBoost face à Boost HPWR	93
7.	Analyse théorique et expérimentale de CARBoost	95
7.1.	Motivations	95
7.2.	Comportement théorique de CARBoost	96
7.3.	Caractéristiques expérimentales	99
8.	Conclusion et perspectives	110

III Application : Détection d'intrusions	112
1. Introduction	114
2. Contexte : Les systèmes de détection d'intrusions : IDS	114
2.1. Architecture d'un système de détection d'intrusions DIS	115
2.2. Approche comportementale	116
2.3. Approche par scénarios	117
2.4. L'apprentissage automatique et les IDS	117
2.5. Les méthodes ensemblistes et les IDS	119
3. Présentation des données KDD-Cup 1999 et motivations	120
3.1. Présentation des données	120
3.2. Motivations	123
4. Expériences et résultats	124
4.1. Expériences	124
4.2. Analyse des matrices de confusion	124
4.3. Comparaison des performances des différentes méthodes	127
4.4. Coûts comparés	128
5. Conclusion et Perspectives	129
Conclusion et perspectives	131
1. Conclusion	131
2. Perspectives	132
2.1. Perspectives théoriques	132
2.2. Perspectives applicatives	133