

# Séminaire TEST

Ruocong Zhang

19 septembre, 2014

## Résumé

Les travaux présentés dans cette thèse portent sur la prédiction des rendements d'actifs financiers par des méthodes d'apprentissage statistique, motivée par le problème de sélection de titres en gestion de portefeuille. En particulier, nous nous intéressons à la prédiction du signe du rendement à un horizon donné. Nous développons une suite de méthodes traitant le problème depuis la construction d'une base de données d'entraînement jusqu'aux tests de performance. Les contributions principales de ces travaux sont d'une part la mise en oeuvre d'un processus complet pourvoyant à toutes les étapes de la prédiction, d'autre part la proposition d'une méthode d'apprentissage multi-tâche pour résoudre simultanément les tâches de prédiction et les relations de dépendance entre elles. Après avoir introduit le cadre de la gestion de portefeuille et les approches usuelles, nous exposons les enjeux de l'apprentissage statistique en gestion de portefeuille, ainsi que les objectifs opérationnels qui, selon nous, définissent une méthode pertinente pour l'application. Les travaux s'organisent autour de quatre thèmes. Le traitement de la prédiction commence par la représentation des données de séries temporelles en descripteurs des actifs financiers. Nous présentons des méthodes de représentation susceptibles de produire des descripteurs robustes et interprétables afin de servir d'observations en entrée de la prédiction. La méthode retenue est l'approximation linéaire par morceaux, où la segmentation est construite par un arbre issu des algorithmes de type Coifman-Wickerhauser ou CART. Une fois obtenue la description des actifs, nous traitons la prédiction du signe des rendements en classification binaire. Nous utilisons des arbres de classification construits par CART et agrégés par Random Forests. Afin de réduire le risque, nous proposons de conditionner la prédiction à des scores de confiance spécifiquement conçus à partir des arbres. Nous nous intéressons ensuite à l'apprentissage multi-tâche. Nous proposons de considérer les tâches de prédiction comme les sommets d'un graphe, et de pénaliser le risque empirique par le laplacien du graphe. Cette formulation permet de résoudre conjointement les tâches et le graphe de dépendance entre elles. Un protocole de test par fenêtres glissantes est enfin proposé pour évaluer la performance de la méthode dans les conditions d'utilisation des prédictions. Nous soulignons dans ce cadre l'importance du choix des paramètres d'apprentissage. Ces méthodes ont abouti à l'émission de recommandations financières publiées en temps réel. Nous concluons par une discussion sur les performances en sélection de titres.