



Université Claude Bernard



DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **21 décembre 2017**

Nom de famille et prénom de l'auteur : **GHARROUDI Ouadie**

Titre de la thèse : « *Apprentissage multi-label ensembliste dans le context supervisé et semi-supervisé* »

Résumé



L'apprentissage multi-label est un problème d'apprentissage supervisé où chaque instance peut être associée à plusieurs labels cibles simultanément. Il est omniprésent dans l'apprentissage automatique et apparaît naturellement dans de nombreuses applications du monde réel telles que la classification de documents, l'étiquetage automatique de musique et l'annotation d'images.

Nous discutons d'abord pourquoi les algorithmes multi-label de l'état-de-l'art utilisant un comité de modèle souffrent de certains inconvénients pratiques. Nous proposons ensuite une nouvelle stratégie pour construire et agréger les modèles ensemblistes multi-label basés sur k-labels. Nous analysons ensuite en profondeur l'effet de l'étape d'agrégation au sein des approches ensemblistes multi-label et étudions comment cette agrégation influence les performances de prédictive du modèle en fonction de la nature de fonction cout à optimiser.

Nous abordons ensuite le problème spécifique de la sélection de variables dans le contexte multi-label en se basant sur le paradigme ensembliste. Trois méthodes de sélection de caractéristiques multi-label basées sur le paradigme des forêts aléatoires sont proposées. Ces méthodes diffèrent dans la façon dont elles considèrent la dépendance entre les labels dans le processus de sélection des variables.

Enfin, nous étendons les problèmes de classification et de sélection de variables au cadre d'apprentissage semi-supervisé. Nous proposons une nouvelle approche de sélection de variables multi-label semi-supervisée basée sur le paradigme de l'ensemble. Le modèle proposé associe des principes issus de la co-training en conjonction avec une métrique interne d'évaluation d'importance des variables basée sur les out-of-bag.

Testés de manière satisfaisante sur plusieurs données de référence, les approches développées dans cette thèse sont prometteuses pour une variété d'applications dans l'apprentissage multi-label supervisé et semi-supervisé.

Testés de manière satisfaisante sur plusieurs jeux de données de référence, les approches développées dans cette thèse affichent des résultats prometteurs pour une variété domaine d'applications de l'apprentissage multi-label supervisé et semi-supervisé.