

Contributions aux communications inter-vues pour l'apprentissage collaboratif

Denis Maurel

10 septembre 2018

Mots clés : Apprentissage non-supervisé, clustering collaboratif, reconstruction collaborative

Résumé : Cette thèse présente plusieurs méthodes d'optimisation et d'amélioration des communications inter-vues dans un contexte d'apprentissage collaboratif. Les idées présentées sont réparties suivant deux axes principaux :

Le premier axe concerne l'amélioration des communications dans le cadre du clustering collaboratif, un paradigme dans lequel plusieurs jeux de données, appelées vues, sont utilisés pour effectuer un premier clustering local avant de s'échanger des informations afin de parvenir à un consensus sur leurs résultats. Contrairement à l'ensemble learning, le clustering collaboratif modifie chaque résultat local plutôt que de fusionner l'ensemble des résultats locaux en un unique modèle. La première méthode d'amélioration des communications inter-vues consiste en une méthode d'apprentissage de coefficients permettant à une vue locale de pondérer l'information en provenance des vues externes. Cette méthode se base sur la résolution d'un problème sous contraintes constitué du critère usuel du clustering collaboratif auquel à été ajouté deux contraintes sur les coefficients de pondérations.

Une seconde contribution consiste en la définition d'une méthode permettant d'effectuer un apprentissage au cours du temps de cartes auto-adaptatrices (aussi appelées cartes de Kohonen), suivie de son adaptation au clustering collaboratif. Cette méthode permet entre autre la mise à jour des résultats obtenues via le clustering collaboratif étant donnés les éventuels changements dans la distribution des données qui pourraient survenir au cours du temps.

Le second axe consiste en la définition d'un nouveau paradigme collaboratif, appelé reconstruction collaborative. Dans ce paradigme, plusieurs vues collaborent pour compléter les données localement manquantes de certaines vues. Cette méthode se base sur l'apprentissage de réseaux de neurones permettant de faire le lien entre les données externes et les données locales. La combinaison des informations ainsi récupérées est assurée par une méthode d'apprentissage de coefficients de pondération permettant de privilégier les caractéristiques les mieux reconstruites par chaque vue externe.