

Transformation de l'espace de description pour l'apprentissage par transfert

Nistor Grozavu *, Younès Bennani*, Lazhar Labiod**

*LIPN-UMR 7030, Université Paris 13,
99, av. J-B Clément, 93430 Villetaneuse, France
email: {firstname.secondname}@lipn.univ-paris13.fr

**LIPADE, Paris Descartes University
45, rue des Saints Pères, 75006 Paris, France
email: {lazhar.labiod}@parisdescartes.fr

Résumé. Dans ce papier, nous proposons une étude sur l'utilisation de l'apprentissage topologique pondéré et les méthodes de factorisation matricielle pour transformer l'espace de représentation d'un jeu de données "sparse" afin d'augmenter la qualité de l'apprentissage, et de l'adapter au cas de l'apprentissage par transfert. La factorisation matricielle nous permet de trouver des variables latentes et l'apprentissage topologique pondéré est utilisé pour détecter les plus pertinentes parmi celles-ci. La représentation de nouvelles données est basée sur leurs projections sur le modèle topologique pondéré.

Pour l'apprentissage par transfert, nous proposons une nouvelle méthode où la représentation des données est faite de la même manière que dans la première phase, mais en utilisant un modèle topologique élagué.

Les expérimentations sont présentées dans le cadre d'un Challenge International où nous avons obtenu des résultats prometteurs (5ème rang de la compétition internationale).

1 Introduction

L'exploration des données, un domaine en pleine évolution et interdisciplinaire, a reçu beaucoup d'intérêt dans de nombreux domaines scientifiques. L'objectif de l'exploration de données est d'extraire des connaissances à partir d'ensembles de données volumineux en combinant les méthodes de statistique et d'intelligence artificielle avec les méthodes de la gestion de bases de données.

La taille des données peut être mesurée selon deux dimensions, le nombre de variables et le nombre d'observations. Ces deux dimensions peuvent prendre des valeurs très élevées, ce qui peut poser un problème lors de l'exploration et l'analyse de ces données. Pour cela, il est fondamental de mettre en place des outils de traitement de données permettant une meilleure compréhension des données.

Dans cette étude, nous nous intéressons à la réduction de dimension de l'espace de description dans le cadre de l'apprentissage non-supervisé à travers la factorisation matricielle et la