

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 (5ième rang de la compétition internationale).

Summary

In this paper, we propose a study on the use of topological weighting learning to transform the representation space of a dataset in order to increase the quality of learning and adapting it to the case of Transfer Learning. For the transformation of the feature space, we use weighted topological models (lwo-SOM) and PCA or SVD to find a new space of representation. New data representation is based on their projections on the topological weighted model. Each example in the dataset is described by a new representation consisting of the distances of this example to all components of the topological model (prototype). For the Transfer Learning, we propose a new method where the representation of data is done in the same way as in the first phase, but using a pruned topological model. The experiments are presented as part of an International Challenge where we obtained good results (5th place).