

Pondération locale des variables en apprentissage numérique non-supervisé ¹

Nistor Grozavu, Younès Bennani, Mustapha Lebbah

LIPN - CNRS UMR 7030 - Université Paris 13
99, avenue J-B. Clément, 93430 Villetaneuse

{Prenom.Nom}@lipn.univ-paris13.fr

Résumé. Dans cet article, nous proposons une nouvelle approche de pondérations des variables durant un processus d'apprentissage non supervisé. Cette méthode se base sur l'algorithme « batch » des cartes auto-organisatrices. L'estimation des coefficients de pondération se fait en parallèle avec la classification automatique. Ces pondérations sont locales et associées à chaque référent de la carte auto-organisatrice. Elles reflètent l'importance locale de chaque variable pour la classification. Les pondérations locales sont utilisées pour la segmentation de la carte topologique permettant ainsi un découpage plus riche tenant compte des pertinences des variables. Les résultats de l'évaluation montrent que l'approche proposée, comparée à d'autres méthodes de classification, offre une segmentation plus fine de la carte et de meilleure qualité.

1 Introduction

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. La réduction des dimensions est l'une des plus vieilles approches permettant d'apporter des éléments de réponse à ce problème. Les méthodes qui nous intéressent dans ce papier sont celles qui permettent de faire à la fois de la réduction de dimension et la classification non supervisée de données en utilisant les cartes auto-organisatrices (SOM : Self-organizing Map). Celles-ci sont souvent utilisées parce qu'elles sont considérées à la fois comme outils de visualisation et de partitionnement non supervisé de différents types de données. Elles permettent de projeter les données sur des espaces discrets qui sont généralement en deux dimensions. Plusieurs extensions des cartes auto-organisées ont été dérivées du premier modèle original proposé par Kohonen (Kohonen,

¹ Ce travail a été réalisé dans le cadre du projet Infom@gic du Pôle de Compétitivité Cap Digital (Image, Multimedia and Vie numérique).