

Apprendre les contraintes topologiques dans les cartes auto-organisatrices

Guénaël Cabanes*, Younès Bennani*

*LIPN-CNRS, UMR 7030
99 Avenue J-B. Clément, 93430 Villetaneuse, France
cabanes@lipn.univ-paris13.fr

Résumé. La Carte Auto-Organisatrice (SOM : Self-Organizing Map) est une méthode populaire pour l'analyse de la structure d'un ensemble de données. Cependant, certaines contraintes topologiques de la SOM sont fixées avant l'apprentissage et peuvent ne pas être pertinentes pour la représentation de la structure des données. Dans cet article nous nous proposons d'améliorer les performances des SOM avec un nouvel algorithme qui apprend les contraintes topologiques de la carte à partir des données. Des expériences sur des bases de données artificielles et réelles montrent que l'algorithme proposé produit de meilleurs résultats que SOM classique. Ce n'est pas le cas avec une relaxation triviale des contraintes topologiques, qui résulte en une forte augmentation de l'erreur topologique de la carte.

1 Introduction

Une Carte Auto-Organisatrice ou Self-Organizing Map (SOM : Kohonen, 2001) se compose d'un ensemble de neurones artificiels, qui représentent la structure des données. Les neurones sont connectés avec des connexions topologiques pour former une grille à deux dimensions. Deux neurones connectés devraient représenter le même type de données, deux neurones distants (sur la carte) doivent représenter des données différentes. Ces propriétés sont assurées pendant le processus d'apprentissage grâce aux informations de voisinage qui imposent des contraintes topologiques.

Toutefois, dans l'algorithme SOM, l'information topologique est fixée avant le processus d'apprentissage et peut ne pas être pertinente par rapport à la structure des données. Pour résoudre ce problème, certains travaux ont été réalisés afin d'adapter le nombre de neurones au cours du processus d'apprentissage en fonction des données à analyser (Fritzke, 1995). Les résultats ont montré que la qualité du modèle est améliorée lorsque le nombre de neurones est appris à partir des données.

En dépit de ces résultats, il y a très peu de travaux qui abordent le problème de l'apprentissage des contraintes topologiques en fonction de la structure des données. Pourtant, à la fin du processus d'apprentissage, des neurones "voisins" peuvent ne pas représenter des données similaires (Cabanes et Bennani, 2007, 2008; Matsushita et Nishio, 2008). Dans l'algorithme False Neighbor-SOM (FN-SOM : Matsushita et Nishio, 2008), les auteurs proposent de conserver la topologie bidimensionnelle de la SOM, mais en associant à chaque "ligne" ou "colonne"