

La carte GHSOM comme alternative à la SOM pour l'analyse exploratoire de données

Françoise Fessant*, Fabrice Clérot*
Pascal Gouzien*

* Orange Labs, 2 av. Pierre Marzin, 22307 Lannion, France
francoise.fessant@orange-ftgroup.com

Résumé. L'objectif de cet article est de faire de la carte auto-organisatrice hiérarchique (GHSOM) un outil utilisable dans le cadre d'une démarche d'analyse exploratoire de données. La visualisation globale est un outil indispensable pour rendre les résultats d'une segmentation intelligibles pour un utilisateur. Nous proposons donc différents outils de visualisation pour la GHSOM équivalents à ceux de la SOM.

1 Introduction

Le modèle des cartes auto-organisatrices hiérarchiques (ou GHSOM pour *Growing Hierarchical Self Organizing Map*) est un arbre de cartes SOM qui s'adapte aux données d'apprentissage par expansion ou agrandissement des feuilles SOM. La taille des branches et la configuration des feuilles varient en fonction des données. Ce modèle a été proposé initialement par Rauber et al. (2002) comme une alternative à la carte SOM traditionnelle. La carte SOM suppose de fixer a priori l'architecture initiale (le nombre de prototypes et la topologie du réseau). La GHSOM se construit sans que l'utilisateur ait à définir la granularité du modèle ni sa profondeur. Seule la forme des feuilles est fixée a priori : les feuilles sont des grilles bidimensionnelles carrées. Le processus d'apprentissage est géré par différents paramètres qui contrôlent l'expansion et l'élargissement des feuilles. Moins contraint que la SOM, il offre de meilleures performances de quantification car ses prototypes se positionnent mieux dans l'espace des données. Dans cet article nous nous intéressons à l'adaptation des outils de visualisation et d'interprétation des classifications de la SOM à la GHSOM. L'objectif est d'en faire un outil utilisable dans le cadre d'une démarche d'analyse exploratoire de données pour laquelle il est nécessaire de disposer de représentations graphiques et de visualisations très parlantes des données aussi bien quantitatives que qualitatives.

2 La carte GHSOM

Le processus d'apprentissage combine une phase d'élargissement et une phase d'expansion qui sont contrôlées par deux paramètres α et β . Les cartes d'un niveau sont indépendantes les unes des autres. Le modèle est initialisé par la création de deux cartes SOM :