

# Visualisation et classification avec les cartes topologiques catégorielles

Mustapha Lebbah<sup>\*,\*\*</sup>, Fouad Badran<sup>\*\*</sup>,  
Sylvie Thiria<sup>\*</sup>

<sup>\*</sup> Laboratoire LODYC, Université Paris 6, Tour 45-5<sup>e</sup> étage, boîte 100  
4 place Jussieu 75252 Paris cedex 05 France

lebbah,thiria@lodyc.jussieu.fr,

<sup>\*\*</sup> CEDRIC, Conservatoire National des Arts et Métiers,  
292 rue Saint Martin, 75003 Paris, France  
badran@cnam.fr

**Résumé.** Ce papier introduit les cartes topologiques dédiée à la visualisation et à la classification de données composées de variables catégorielles. Pour visualiser ou classer ces données par des cartes topologiques, les méthodes classiques utilisent une phase de codage de prétraitement de ces données en données numériques et appliquent l'algorithme classique des cartes topologiques. Dans ce papier nous proposons un modèle de cartes topologiques dédiées aux données catégorielles. Ce modèle est basé sur un formalisme probabiliste où chaque cellule est représentée par une table de probabilités. Deux exemples réels permettent de valider ce modèle. Les résultats obtenus montrent l'apport de ce modèle dans la visualisation et la classification de données catégorielles.

## 1 Introduction

La visualisation des données est une étape importante dans la phase exploratoire d'une analyse de données. Cette étape est plus difficile quand il s'agit de données qualitatives pour lesquelles il existe moins de méthodes standard. Les cartes topologiques sont de plus en plus utilisées comme outils de visualisation, puisqu'elles permettent de projeter sur des espaces discrets qui sont généralement de dimensions deux. Le modèle de base, proposé par Kohonen [9], est dédié uniquement aux données numériques, il a cependant été appliqué avec succès au traitement de données textuelles [8]. Cet algorithme a été appliqué aussi sur des données binaires (données catégorielles codées en binaires) précédé d'un prétraitement ou d'un changement de représentation spécifique des données [7, 13]. Des extensions et reformulations du modèle de Kohonen ont été proposé dans la littérature : Cartes topologiques probabilistes [1], Generative Topographic Mapping [6, 2]. En se basant sur le formalisme classique des cartes topologiques, nous avons déjà proposé un modèle de cartes topologiques dédiés aux données binaires [10]. Dans ce papier, nous présentons le modèle CTM de cartes topologiques dédié aux données catégorielles. Ce modèle est basé sur le formalisme des cartes topologiques probabilistes, l'apprentissage consiste alors à estimer les paramètres du modèle en maximisant la fonction de vraisemblance des observations de la base d'apprentissage. L'algorithme d'apprentissage que nous proposons est une application de l'algorithme EM. Au paragraphe 2 nous présentons le modèle CTM et l'algorithme d'apprentissage associé,