

Extraction des nombres de Betti avec un modèle génératif

Maxime Maillot *, Michaël Aupetit*
Gérard Govaert**

* CEA, LIST, F-91191 Gif-sur-Yvette
maxime.maillot@cea.fr, michael.aupetit@cea.fr,

**UTC, Heudiasyc, UMR CNRS 6599, BP 20529, F-60205 Compiègne Cedex
gerard.govaert@utc.fr

Résumé. L’analyse exploratoire de données multidimensionnelles est un problème complexe. Nous proposons d’extraire certains invariants topologiques appelés nombre de Betti, pour synthétiser la topologie de la structure sous-jacente aux données. Nous définissons un modèle génératif basé sur le complexe simplicial de Delaunay dont nous estimons les paramètres par l’optimisation du critère d’information Bayésien (BIC). Ce Complexe Simplicial Génératif nous permet d’extraire les nombres de Betti de données jouets et d’images d’objets en rotation. Comparé à la technique géométrique des Witness Complex, le CSG apparaît plus robuste aux données bruitées.

1 Introduction

Une approche récente de l’analyse exploratoire de données multidimensionnelles consiste en l’extraction d’invariants topologiques (Carlsson, 2009). Ces invariants permettent de caractériser la topologie de la population si l’on suppose qu’elle est définie par une collection de sous-variétés de l’espace \mathbb{R}^D dont sont issues les données. Les nombres de Betti sont de tels invariants : le premier encode le nombre de composantes connexes, le second le nombre de cycles indépendants, le troisième le nombre de cavités indépendantes, etc... Ces nombres expriment numériquement des caractéristiques topologiques que nous extrayons visuellement lors de l’analyse de nuages de points présentés dans un repère cartésien à deux dimensions. Ces invariants topologiques le sont par homotopie une classe très large de transformations non linéaires incluant les homéomorphismes, les similitudes et les isométries, on s’attend donc à ce que l’informaiton topologique soit plus robuste à la chaîne de mesure que l’information de nature géométrique.

Il existe des techniques d’extraction des nombres de Betti à partir d’un complexe simplicial (de Silva). Un complexe simplicial est un ensemble de simplexes tels que l’intersection de deux simplexes du complexe est soit vide soit fait aussi partie du complexe. Le plongement dans \mathbb{R}^D d’un k -simplexe est l’enveloppe convexe de $(k + 1)$ -points de cet espace. Un complexe simplicial plongé est donc une collection de variétés qui peut servir de modèle topologique et géométrique à la population génératrice des données. Nous devons lui associer un modèle de densité de probabilité afin de modéliser complètement le processus de génération des données.