

Apprentissage d’espaces prétopologiques dans un cadre multi-instance pour la structuration de données

Gaëtan Caillaut, Guillaume Cleuziou

LIFO, Université d’Orléans, Rue Léonard de Vinci, 45067 Orléans Cedex 2
prenom.nom@univ-orleans.fr

Résumé. Nous présentons dans cet article une méthode supervisée de structuration (en DAG) d’un ensemble d’éléments. Étant donné une structure cible et un ensemble de relations sur ces éléments, il s’agit d’apprendre un modèle de structuration par combinaison des relations initiales. Nous formalisons ce problème dans le cadre de la théorie de la prétopologie qui permet d’atteindre des modèles de structuration complexes.

Nous montrons que la non-idempotence de la fonction d’adhérence rentre dans le cadre du formalisme de l’apprentissage (supervisé) multi-instance et nous proposons un algorithme d’apprentissage reposant sur le dénombrement des « sacs » positifs et négatifs plutôt que sur un ensemble d’apprentissage standard.

Une première expérimentation de cette méthode est présentée dans un cadre applicatif de fouille de textes, consistant à apprendre un modèle de structuration taxonomique d’un ensemble de termes.

1 Introduction

La structuration de données et de connaissances est le sujet d’intenses recherches depuis plusieurs décennies et concerne un très large champ d’applications. Au-delà même d’une utilité applicative directe, le besoin de structurer des données, des variables ou des concepts est devenu incontournable dans de nombreux processus décisionnels.

Différentes formes de structures peuvent être envisagées parmi lesquelles on peut citer les graphes, les arbres ou encore les DAGs (graphes orientés sans cycles). Selon le contexte, deux problématiques d’apprentissage peuvent être identifiées : l’apprentissage de la structure elle-même (ex. modèle Bayésien) ou l’apprentissage d’un modèle de structuration. Nous nous intéressons dans cette étude à la seconde problématique consistant à inférer un modèle de structuration en DAG à partir d’une structure cible et d’une description sur les données.

Nous proposons dans cet article, un cadre générique d’apprentissage supervisé d’un modèle de structuration complexe d’un ensemble d’éléments E en un DAG $\mathcal{G}$ dans lequel chaque nœud représente un élément ou un sous-ensemble d’éléments. La complexité des structures ciblées par notre approche repose sur l’usage du formalisme prétopologique permettant une modélisation fine du processus de propagation de la domination entre éléments ou ensembles d’éléments. Le cadre est présenté comme *générique* dans le sens où il considère en entrée une collection de relations $\{\mathcal{R}_1, \dots, \mathcal{R}_K\}$ binaires sur $E \times E$, plutôt qu’une description de type