

Subspace Clustering et Visualisation des Flux de Données

Ibrahim Louhi^{*,**} Lydia Boudjeloud-Assala^{*}
Thomas Tamisier^{**}

^{*}Université de Lorraine, Laboratoire d'Informatique Théorique et Appliquée.
{ibrahim.louhi, lydia.boudjeloud-assala}@univ-lorraine.fr

^{**}Luxembourg Institute of Science and Technology.
{ibrahim.louhi, thomas.tamisier}@list.lu

Résumé. Dans ce papier nous proposons une nouvelle approche de subspace clustering pour les flux de données, permettant à l'utilisateur de suivre visuellement le changement dans le comportement du flux. Cette approche détecte l'impact des variables sur l'évolution du flux, Tout en visualisant les étapes du subspace clustering en temps réel. En premier lieu nous appliquons un clustering sur l'ensemble de variables afin d'identifier les sous-espaces. Ensuite un clustering est appliqué sur les individus dans chaque sous-espace.

1 Introduction

Le clustering est une des techniques utilisées pour la fouille de données, qui essaye de regrouper les individus similaires selon certains critères dans le même groupe appelé cluster. Cependant, parfois les données comportent des informations cachées qui ne sont pas visibles sur l'espace original de variables. Parmi les techniques utilisées pour découvrir ces informations, le subspace clustering cherche à trouver des clusters sur tous les sous-espaces de données.

La tâche du subspace clustering se complique encore plus quand il s'agit de flux de données. Comment identifier des clusters dans ses sous-espaces pertinents tout en respectant les contraintes du traitement du flux de données. Plusieurs approches ont été proposées pour effectuer du subspace clustering sur les flux de données. Cependant aucune à notre connaissance ne permet de visualiser en temps réel l'évolution du flux et des sous-espaces. La visualisation permet dans le contexte du subspace clustering de mieux comprendre les résultats obtenus, et le plus important, d'explorer les données au niveau des différents sous-espaces.

Dans ce papier nous présentons dans un premier temps un bref état de l'art de quelques techniques utilisées pour le subspace clustering sur des données statiques, la visualisation des sous-espaces, et le subspace clustering appliqué aux flux de données. Ensuite, nous présentons notre approche pour effectuer et visualiser un subspace clustering de flux. Nous discuterons également les résultats obtenus, et nous illustrons l'utilité de notre approche et des perspectives pouvant l'améliorer.