

Clustering visuel semi-interactif

Lydia Boudjeloud-Assala*, Philippe Pinheiro**,
Alexandre Blansché*, Thomas Tamisier** et Benoît Otjacques**

* Laboratoire d'Informatique Théorique et Appliquée, LITA-EA 3097,
Université de Lorraine, Metz, F-57045, France
prenom.nom@univ-Lorraine.fr

**LIST- Luxembourg Institute of Science and Technology,
Esch-Sur-Alzette, Luxembourg
prenom.nom@list.lu

Résumé. Nous proposons dans cet article une approche de clustering visuel semi-interactif. L'approche proposée utilise la perception visuelle pour guider l'utilisateur dans le processus interactif. Les clusters sont extraits de manière successive et itérative, puis évalués selon leur ordre d'extraction. Pour l'utilisateur, l'approche semi-interactive permet non seulement d'évaluer les classes en fonction d'un critère déterminé mais aussi d'évaluer l'influence de l'extraction d'un cluster sur ceux précédemment extraits. Un protocole de test est présenté afin de comparer cette approche avec les approches purement automatiques et purement interactives. Cet article est un résumé d'un papier accepté¹ pour un journal international.

1 Introduction

Dans le processus d'extraction de connaissances à partir de données, il y a au moins deux moyens de faire collaborer les méthodes automatiques avec des méthodes visuelles interactives. Il est possible d'utiliser les méthodes de visualisation en prétraitement de l'algorithme automatique ou en post-traitement de ce même algorithme. En prétraitement de données, on s'aperçoit que, bien souvent, une intuition initiale des concepts cachés peut être acquise de façon visuelle dans les très grandes quantités d'information. Cette étape peut également guider l'utilisateur dans le choix des algorithmes de fouille les plus pertinents ou de leurs paramètres. En post-traitement des connaissances, les méthodes de visualisation sont plutôt utilisées pour interpréter et évaluer des résultats en se basant sur des représentations graphiques plus accessibles que des colonnes de chiffres ou un ensemble de règles. Ces différentes interactions illustrent l'intérêt de faire coopérer des méthodes automatiques et des méthodes visuelles interactives. La compréhension des résultats est ainsi accrue et la précision des algorithmes automatiques peut être facilement améliorée. Une des possibilités pour augmenter la part de la visualisation dans les algorithmes de fouille de données est de faire coopérer l'algorithme automatique de fouille de données avec un algorithme visuel interactif. On parle alors de fouille

1. Interactive and iterative visual clustering, Information Visualization Journal, pp 1-17, 2015 (DOI: 10.1177/1473871615571951)