

Fouille de données par programmation visuelle structurée avec KD-Ariane

Régis Clouard, François Rioult

GREYC CNRS UMR6072 - ENSICAen - Université de Caen Basse-Normandie
regis.clouard@ensicaen.fr, francois.rioult@unicaen.fr
<https://forge.greyc.fr/projects/kdariane>

Résumé. Nous présentons ici la plate-forme KD-Ariane, un déploiement d'outils pour la fouille de données dans l'environnement de programmation visuelle Ariane. Ce déploiement facilite la conception de chaînes structurées de traitements pour l'extraction de connaissance dans les données.

1 Introduction

La programmation visuelle de composants de haut niveau permet d'améliorer le développement des chaînes de traitement. C'est pourquoi les plate-formes libres pour l'analyse de données (Weka, RapidMiner, Knime¹) proposent des environnements de développement visuel. Pour autant, il y est difficile d'ajouter ses propres réalisations, par exemple des binaires compilés, car ces plate-formes requièrent l'utilisation d'une API dédiée, imposant par exemple un format spécifique pour la manipulation des données. Les possibilités de programmation structurée y sont également limitées : ces plate-formes sont conçues pour enrober leurs propres composants de base pour l'apprentissage automatique.

Nous présentons ici la plate-forme KD-Ariane, un déploiement d'outils pour la fouille de données dans l'environnement de programmation visuelle Ariane. Ce déploiement facilite la conception de chaînes structurées de traitements pour l'extraction de connaissance dans les données (Clouard (2009)). KD-Ariane est simple à prendre en main et possède un fort potentiel pédagogique. Les composants graphiques exécutent simplement des commandes système, rendant aisé l'interfaçage avec les bibliothèques disponibles. La puissance réside dans les possibilités de programmation structurée offertes par les boucles, les macro-composants et leur partage avec d'autres utilisateurs sous forme de routines. Enfin, les chaînes développées avec KD-Ariane sont exportables en scripts `shell` ou `perl`.

Ariane : c'est une plate-forme de programmation visuelle, initialement conçue pour valoriser les opérateurs de traitement d'image de la bibliothèque Pandore (Pandore (2013)). Nous illustrons ici son potentiel dans la réalisation de chaînes de traitement pour l'extraction de connaissances dans les bases de données.

Le principe est celui de la programmation visuelle : des composants sont assemblés pour réaliser la chaîne de traitements. Ces traitements sont réalisés sur des fichiers de données, afin de produire d'autres données, ou des résultats sous forme de chaîne de caractère ou de valeur numérique.

1. <http://www.cs.waikato.ac.nz/ml/weka/index.html>, <http://rapidminer.com/>, <http://www.knime.org/>