

Interprétation visuelle et interactive d'épisodes séquentiels

Pierre-Loup Barazzutti * Amélie Cordier ** Béatrice Fuchs ***

* LIRIS CNRS, UMR 5205, Université Lyon 1

** amelie.cordier@liris.cnrs.fr

*** LIRIS CNRS, UMR 5205, Université Lyon 3

beatrice.fuchs@liris.cnrs.fr

Durant le processus d'extraction de connaissances à partir de données (ECD), de nombreuses interactions ont lieu avec un utilisateur humain expert du domaine ou analyste qui y joue un rôle moteur. Pour aider l'analyste dans sa tâche, de nombreux outils permettent toutes sortes de visualisations réalistes des résultats, mais leurs capacités d'interaction restent souvent limitées à des manipulations graphiques. Nous proposons une approche interactive et visuelle de l'ECD supportée par un système de gestion de traces. Nous nous focalisons dans un premier temps sur l'étape d'interprétation.

1 Traces et système à base de traces

Nous nous plaçons dans le cadre de l'étude des *traces modélisées* (Champin et al., 2013) où une trace est issue de l'observation d'un phénomène et est constituée d'une séquence d'éléments observés temporellement situés que nous appelons des *obsels*. Une trace est modélisée c'est-à-dire qu'elle se conforme à un *modèle de trace* décrivant les types d'obsels, leurs attributs et leurs relations et permettant d'interpréter la trace afin de faciliter son exploitation ultérieure. Les traces sont manipulées par des opérations élémentaires appelées *transformations* qui sont de différents types : filtrage d'obsels, fusion de traces, etc. Parmi ces transformations, la *réécriture* est une opération qui crée une nouvelle trace appelée *trace transformée* qui vise à augmenter progressivement le niveau de compréhension et d'abstraction de la trace initiale. La réécriture consiste à construire une nouvelle trace t_2 à partir d'une trace primaire t_1 en remplaçant dans t_2 des motifs, c'est-à-dire des séquences d'obsels, de t_1 par de nouveaux types d'obsels résumant chaque motif et en constituant une interprétation. Un système à base de traces est un système permettant de collecter, de traiter et de visualiser des traces. La réécriture de traces se situe au cœur du dispositif mis en place dans l'interprétation interactive.

Ce travail a été réalisé dans le cadre du projet OFS - Open Food System, programme investissements d'avenir.

2 Extraction de connaissances à partir de traces

Les traces sont étudiées dans le cadre d'un processus d'ECD. Durant le post-traitement, les motifs issus de la fouille sont proposés à l'analyste qui y choisit ceux qu'il estime les plus pertinents compte tenu de ses connaissances du domaine. Le principe de l'interprétation est de produire une trace transformée par réécriture de la trace initiale où les motifs choisis sont remplacés par de nouveaux types d'observés issus de l'interprétation par l'analyste. Pour appréhender méthodiquement l'étape d'interprétation, nous proposons une démarche itérative pour assister l'analyste dans sa tâche, car la fouille produit souvent un nombre important de motifs caractérisés par une forte redondance combinatoire.

Tout d'abord, la liste des motifs est complétée par des indicateurs de qualité permettant de les trier et ainsi mettre en avant les motifs potentiellement les plus intéressants. Les motifs sont alors affichés et l'analyste peut les sélectionner pour visualiser leurs occurrences dans la trace. Lorsqu'un motif est sélectionné, l'analyste peut décider de créer un nouveau type d'obsel pour mémoriser son interprétation dans la trace transformée. À chaque sélection de motif, la liste des motifs est mise à jour dynamiquement en éliminant toutes les occurrences de motif ayant au moins un obsel en commun avec les occurrences du motif sélectionné. Les indicateurs des motifs restants sont alors recalculés et la liste est triée en conséquence. Les motifs ne satisfaisant plus les contraintes sont éliminés, ce qui diminue graduellement le nombre de résultats restants. Ce filtrage automatique facilite ainsi le travail de l'analyste en lui permettant de se focaliser plus rapidement sur d'autres motifs. Lorsque l'analyste a sélectionné tous les motifs et défini les types d'obsels correspondants, il peut alors réitérer le processus d'ECD avec d'autres contraintes sur la même trace ou sur la trace transformée.

Ce scénario d'interprétation interactive a été mis en oeuvre dans le prototype TRANSMUTE qui est présenté de façon plus détaillée dans l'appel à démonstration (Barazzutti et al., 2016).

Références

- Barazzutti, P.-L., A. Cordier, et B. Fuchs (2016). Transmute : un outil interactif pour assister l'extraction de connaissances à partir de traces. In *EGC-2016*, pp. à paraître.
- Champin, P.-A., A. Mille, et Y. Prié (2013). Vers des traces numériques comme objets informatiques de premier niveau : une approche par les traces modélisées. *Intellectica* (59), 171–204.