

Fouille de données interactive

Apprentissage de préférences

Rendu : Non obligatoire mais peut permettre de rattraper un TP précédent.

L'objectif de ce TP est d'étudier les différentes étapes d'un processus de fouille de données interactives. Notamment, nous étudierons dans cette séance comment utiliser les retours de l'utilisateur afin d'apprendre un modèle de préférences, puis, comment utiliser ce modèle dans les futures itérations de la boucle d'interaction.

Pour rappel, un processus de fouille de données interactive est composée de 3 étapes principales, chacune comportant différents verrous scientifique. Dans cette séance de TP, nous aborderons plus particulièrement l'étape « Learn » qui vise à construire un modèle de préférences à partir des retours de l'utilisateur. Nous nous interrogerons également sur l'ensemble de motifs à présenter à l'utilisateur à chaque étape en assurant une diversité locale ou globale (i.e., à travers les différentes itérations).

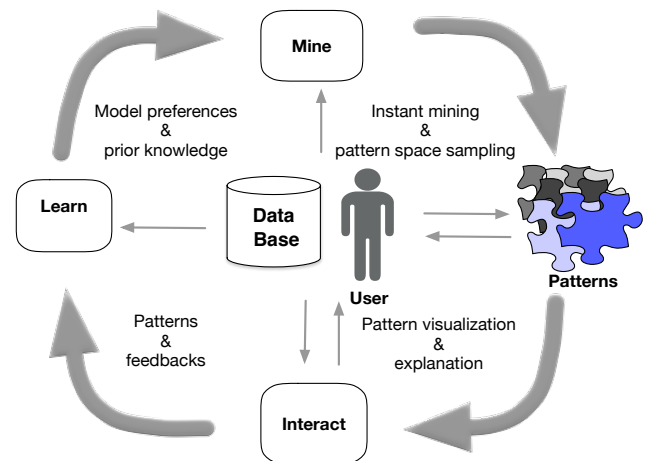
Afin de nous focaliser sur l'étape « Learn » dans ce TP, nous faisons des choix sur les deux autres étapes :

- «**Mine**» : on utilisera l'algorithme d'échantillonnage des motifs fréquents de la première séance.
- «**Interact**» : pas d'interface graphique pour le moment. On s'appuiera sur des ensembles de likes, dislikes et « neutres » (pas d'avis émis).

1. Quels descripteurs pour caractériser les motifs ? Construire un ensemble de descripteurs décrivant un motif (support, aire, contraste par rapport à un attribut cible, attributs présents dans le motifs, etc.).
2. Comment utiliser les évaluations individuelles des motifs pour apprendre des préférences ? Pour cela, il est nécessaire de faire une « pairwise transformation ». Plus les comparaisons paires à paires, plusieurs choix sont possibles (un motif aimé est préféré à un motif qui a reçu un feedback négatif, un motif aimé est supérieur à un motif neutre, un motif neutre supérieur à un motif négatif).
3. Apprendre les préférences : à partir de la transformation réalisée à l'étape précédente, apprendre un modèle de préférence à partir de l'algorithme SVM-Rank.
4. Exploiter le modèle : post-traitement d'une ensemble de motifs (les ordonner). Quels sont les bons motifs à retourner à l'utilisateur ? (voir slide 76/97 du cours sur la fouille de données interactive).

Références :

- "Optimizing Search Engines Using Clickthrough Data", T. Joachims. Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM, 2002.
- Dzyuba, V, van Leeuwen, M, Nijssen, S & De Raedt, L [Interactive Learning of Pattern Rankings](#). International Journal on Artificial Intelligence Tools vol.23(6), World Scientific, 2014
- Rüping, Stefan: [Ranking Interesting Subgroups](#). In: Proceedings of the 26th International Conference on Machine Learning (ICML 2009), Bottou, Leon and Littman, Michael (Eds.), Montreal, Omnipress, 913-920, 2009.



A new exploratory analysis: easy to use (interactive, no skill required), that produces more **insightful patterns** based on user **preferences** and **background knowledge** and well-adapted for **big data** (instant-mining, pattern-space sampling)

