

Stage de master

Algorithme de clustering éactif pour l'exploration de données et l'anticipation de services

On cherche à développer un outil pour mieux comprendre les systèmes complexes (en particulier des systèmes sociaux-techniques, i.e incluant des acteurs humains, des systèmes techniques éventuellement en interaction avec des environnements naturels) afin de pouvoir agir sur ceux-ci en proposant les traitements/services les plus pertinents à un instant donné.

Pour cela nous faisons les hypothèses suivantes :

- Les systèmes sont décrits par des objets, des événements caractérisés par un vecteur de N dimensions (parmi lesquelles l'espace, le temps, les usages, ...)
- Il existe une mesure de la similitude entre ces objets/événements pris deux à deux
- La structuration en groupes, classes d'objets/événements doit permettre une meilleure compréhension du système considéré.

Etant donnée la complexité de ces systèmes et le nombre important d'objets le décrivant nous proposons de mettre en œuvre et d'étudier un algorithme de clustering capable de s'adapter aux intentions de l'utilisateur, même si ces intentions, floues initialement se précisent voir évoluent au cours du temps.

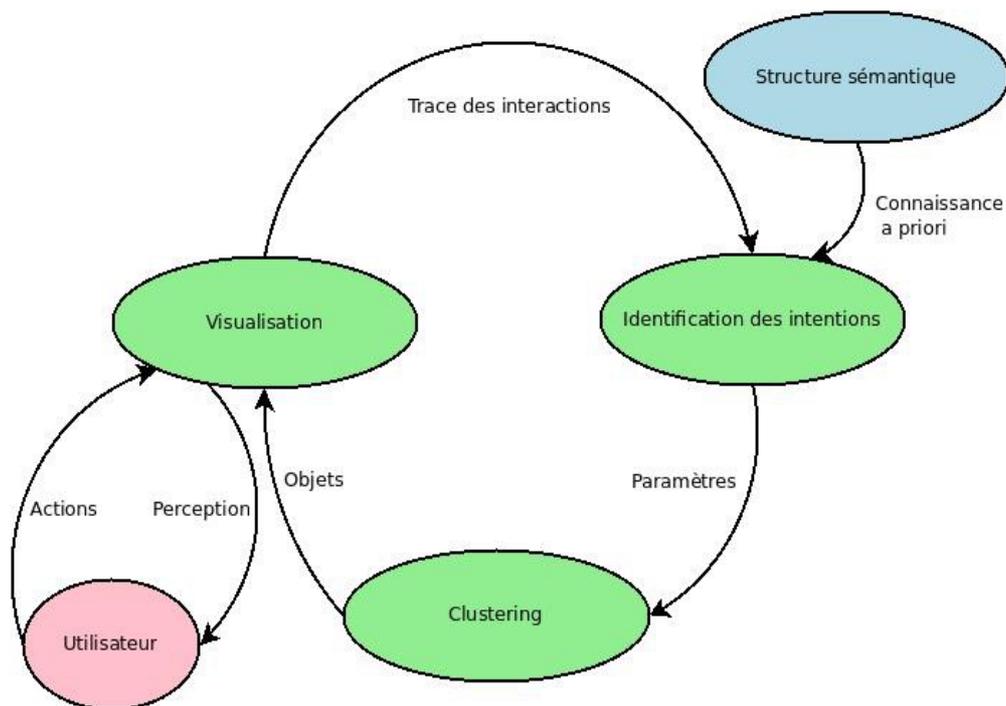
Pour atteindre cet objectif nous proposons d'étudier un algorithme reposant sur un paradigme général à savoir « *la marche pour chemin* » : une personne marchant dans un champ trace un chemin. Celui-ci devient de plus en plus attractif à chaque fois qu'il est emprunté. Au contraire, il tend à disparaître quand il est délaissé. Il y a donc interaction entre le marcheur et son environnement, on parle même de couplage fort c'est à dire qu'ils s'influencent mutuellement et en continu.

Concrètement il s'agit (1) de mettre en œuvre un couplage fort entre un utilisateur et un algorithme de clustering et (2) d'étudier les propriétés de ce couplage.

(1) une procédure de visualisation sert d'interface entre l'utilisateur et le système à développer : l'utilisateur y perçoit les résultats partiels d'un algorithme de clustering et y réagit en s'intéressant à certains objets présentés plus qu'à d'autres. Ces interactions sont analysées par une procédure d'identification des intentions de l'utilisateur qui modifie, en fonction de ces intentions inférées les paramètres de la procédure de clustering.

(2) nous chercherons à étudier certaines propriétés du couplage ainsi mis en œuvre :

- la capacité du système à s'adapter aux intentions de l'utilisateur
- la capacité du système à orienter les utilisateurs vers une classifications jugée pertinente a priori
- la dynamique de la prise d'initiative (qui de l'utilisateur ou du système propose de nouveaux centres d'intérêt).
- l'évolution de la variété des objets/événements présentés à l'utilisateur



Lieu : CERV (Brest)
Encadrant : Eric Maisel (ENIB)
Financement : Arkéa

Bibliographie :

- De Baker M, Haesen R., Martens D., Baesens B, « A Stigmergy Based-Approach to Data Mining », proc of AI 2005, LNAI 3809, pp 975-978, 2005
- Ong J, Abidi S., « Data Mining using Self-Organizing Kohonen Maps : a Technique for Effective Data Clustering and Visualisation », proc of International Conference on Artificial Intelligence (IC-AI'99)
- Proctor G, Winter C, « Information Flocking : Data Visualization in Virtual Worlds using Emergent Behaviours », proc of «Virtual Worlds : First International Conference, VW'98, Paris, 1998
- Rajaonarivo L, Courgeon M, Maisel E, De Loor P., « Inline Co-evolution between Users and Information Presentation for Data Exploration », proc. Of the 22nd International Conference on Intelligent User Interfaces, pp 215-219, IUI'17, 2017.
- Sklar E., Jansen C., Chan J., Byrd M., « Toward a methodology for Agent-Based Data Mining and Visualization », Proc. of Agents and Data Mining Interaction ADMI 2011, LNAI 7103 pp 4-15, 2012, Springer-Verlag.

