

Ontologies floues et fouille de données

A. Laurent et S. Bringay
{laurent,bringay}@lirmm.fr

Les ontologies permettent une représentation formelle d'un domaine de connaissance sous la forme d'une terminologie hiérarchisée munie de relations sémantiques. Les représentations des ontologies de manière formelle reposent généralement sur une description en logique classique (e.g. la logique du premier ordre). Or celle-ci montre ses limites pour tous les faits qui ne s'expriment pas avec des « vrais » et des « faux ». Des approches floues ont alors été proposées pour rendre plus souples les possibilités de représentation en affectant des poids aux différents liens (e.g. « le patient a une fièvre moyennement forte » plutôt que « le patient a ou n'a pas de fièvre »).

Si des méthodes ont été proposées pour construire ces ontologies et les utiliser dans certains contextes, de nombreux problèmes restent ouverts (mise à jour automatique de ces ontologies, utilisation de ces ontologies floues dans le processus de fouille de données, mise en correspondance automatique d'ontologies, etc).

Dans le cadre de ce stage, il s'agira d'étudier comment tirer profit au mieux des ontologies floues dans le processus de fouille de données, et notamment pour l'extraction de motifs séquentiels les plus pertinents possibles.

Une des applications privilégiées sera le domaine de la santé où les ontologies sont très largement répandues et utilisées pour décrire les connaissances des différents sous-domaines.

Il s'agira donc notamment d'étudier comment optimiser l'extraction de motifs séquentiels s'appuyant sur les ontologies floues, pour l'étude de l'évolution des processus biologiques et cliniques chez les patients à partir des étapes suivantes :

1. étude bibliographique,
2. proposition de méthodes complètes de prise en compte d'ontologies floues pour l'extraction de motifs séquentiels, incluant le cadre formel (e.g. définitions, mesures de comparaison) et les algorithmes,
3. mise en oeuvre de ces solutions en les implémentant et en les validant sur des bases de données réelles et conséquentes.

Références bibliographiques

Gruber T. R. (1993). A translation approach to portable ontology specifications. In: Knowledge Acquisition, 5, 199-220.

Bodenreider, O., Smith,B., Kumar,A., and Burgun,A. (2007). Investigating subsumption in SNOMED CT: An exploration into large description logic-based biomedical terminologies. In: Artif.Int.Med., 39:183-195.

Tang, Z., Phan, S., Pan, Y., and Famili, F (2007). *Prediction of Co-Regulated Gene Groups through Gene Ontology*. In : Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CICBC2007, ISBN: 1-4244-0710-9). pp 178-184 1 au 5 avril, 2007

Parry, D.T. Evaluation of a fuzzy ontology based medical information system. International journal of Health Information Systems and Informatics, vol. 1, pp. 40-49, 2006.