

Sujet : Approche déclarative pour l'énumération des motifs graduels

Encadrant : Badran Raddaoui, INF, Equipe ACMES/SAMOVAR

Les données numériques représentent des informations quantitatives sur quelque chose qui est mesurable. Elles sont toujours collectées à partir de sources et de domaines d'application variés, notamment l'environnement et l'écologie, les sciences humaines et sociales, le commerce et la finance, les sciences de la vie et l'astronomie. De nos jours, les données numériques sont abondantes en raison de la prolifération des dispositifs de mesure et de collecte de données et des capteurs. Ce type de données est généralement analysé du point de vue de l'apprentissage machine, de l'analyse statistique et logique des données. D'autre part, relativement peu d'approches de fouille de modèles, fonctionnant généralement sur des données catégorielles, sont conçues pour traiter des données numériques. Parmi elles, on peut citer l'extraction quantitative d'ensembles d'éléments/règles d'association, les modèles d'intervalle à partir de données numériques en utilisant l'analyse conceptuelle formelle, les ensembles d'attributs numériques corrélés au rang d'extraction et l'extraction progressive de modèles. L'extraction graduelle d'ensembles d'items permet de découvrir des co-variations fréquentes entre les attributs, comme "plus l'âge est élevé, plus le salaire est élevé et plus le temps libre est faible". Ce dernier type de modèles est très expressif car il représente la variabilité des valeurs numériques, souvent recherchée dans de nombreux domaines d'application, y compris par exemple la biologie où la plupart des progrès sont réalisés par l'analyse des données du génome, la médecine où les chercheurs en psychologie se concentrent sur les corrélations entre la mémoire et les points de ressenti du "Manuel de diagnostic des troubles mentaux" et les marchés financiers où l'on aimerait découvrir des covariations entre divers indicateurs économiques et financiers, comme la hausse des taux d'intérêt et la baisse des prix des obligations et vice versa. Plus généralement, la découverte de modèles graduels est pertinente dans toutes les données numériques où l'on a besoin de récupérer la relation entre les attributs en termes de variabilité. Les approches de l'état de l'art [2, 3, 4, 5] ont été conçues pour traiter de la découverte progressive d'un modèle ou de ses variantes. Un modèle progressif extrait par la plupart de ces approches est fourni avec un support unique ou une sous-séquence de transactions, appelée extension, qui le soutient. Ces informations,

lorsqu'elles sont fournies, nous permettent d'expliquer à l'utilisateur pourquoi un tel ensemble d'éléments est progressif. Elle augmente la robustesse du processus d'extraction, souvent utile dans plusieurs domaines d'application. L'un des principaux problèmes des approches d'extraction de modèles progressifs est l'espace de combinaison exponentiel à explorer et le problème de la gestion de la quantité de modèles extraits qui peuvent également être de taille exponentielle. Récemment, une approche à base de contraintes pour l'extraction des motifs graduels à partir d'une base de données numérique a été proposée [1].

L'objectif de ce stage est d'étudier et d'implémenter une approche de fouille de motifs graduels en se basant sur un encodage vers la logique propositionnelle (Propositional Satisfiability ou SAT en anglais). Ce modèle SAT déclaratif inclut différents types de contraintes qu'il faut encoder pour extraire les motifs graduels fréquents. L'approche sera implémenter en C/C++ en étendant le solveur de l'état de l'art MiniSAT (<http://minisat.se/>).

Bibliographie

- [1] Jerry Lonlac, Said Jabbour, Badran Raddaoui, Lakhdar Sais: Extracting Frequent Gradual Patterns using Constraints Modeling. CoRR/abs/1903.08452, 2019
- [2] Said Jabbour, Jerry Lonlac, Lakhdar Sais: Mining Gradual Itemsets Using Sequential Pattern Mining. Fuzz-IEEE, pages 1–6, 2019
- [3] Jerry Lonlac, Yannick Miras, Aude Beauger, Vincent Mazenod, Jean-Luc Peiry, and Engelbert Mephu. An approach for extracting frequent (closed) gradual patterns under temporal constraint. In FUZZ-IEEE, pages 878–885, 2018.
- [4] Anne Laurent, Benjamin Négrevergne, Nicolas Sicard, and Alexandre Termier. Pgp-mc: Towards a multicore parallel approach for mining gradual patterns. In DASFAA, Part I, pages 78–84, 2010.
- [5] Anne Laurent, Marie-Jeanne Lesot, and Maria Rifqi. GRAANK: exploiting rank correlations for extracting gradual itemsets. In Flexible Query Answering Systems, 8th International Conference, FQAS, pages 382–393, 2009.