

# Motifs séquentiels flous et attributs quantitatifs

Céline Fiot

Les motifs séquentiels ont été étudiés depuis maintenant plusieurs années. Ils permettent de traiter de gros volumes de données et d'en extraire des règles incluant la dimension temporelle de la forme :  $\langle (\text{lecteur de DVD DVD})(\text{écran}) \rangle$  signifiant que les clients qui achètent simultanément un lecteur de DVD et un ou plusieurs DVD achètent plus tard un écran. Cependant, les algorithmes proposés ne travaillent que sur des données binaires qui indiquent l'absence ou la présence d'éléments. Or la plupart des données réelles intéressantes dans le contexte de données séquentielles sont numériques (par exemple pour décrire des données issues de capteurs prélevées au cours du temps).

Une première proposition [SA96] traite ce type de bases et d'attributs grâce à un découpage des attributs en intervalles. Plus récemment, l'utilisation de la théorie des sous-ensembles flous a permis des coupures moins brutales entre les intervalles, ce qui conduit à l'obtention de règles plus pertinentes. Ainsi, [KFW98] et [HCTS03, CH02, CTCH01, HLW01] ont proposé des extensions des techniques d'extraction de motifs séquentiels et de règles d'association.

Concernant les règles d'association tout d'abord, [KFW98] propose une nouvelle définition du support et de la confiance pour des règles d'association sur des attributs quantitatifs.

Ces nouvelles définitions sont basées sur l'introduction d'ensembles flous qui remplacent les items classiques. Ainsi, on ne cherchera plus des règles du type "75% des gens qui achètent des couches achètent aussi des bières" mais "60% des gens qui achètent beaucoup de chocolat achètent peu de chips". Les items ne sont plus des attributs discrets (couches, bières, chocolat), mais des attributs quantitatifs (beaucoup de chocolat...).

Pour l'extraction de motifs séquentiels flous, le principe est le même que pour les règles d'association : un motif séquentiel flou consiste en une séquence fréquente de plusieurs attributs quantitatifs. [HCTS03, CH02, CTCH01, HLW01] propose la recherche de motifs séquentiels flous et plus généralement de connaissances floues relatives au comportement d'achats des consommateurs.

Cependant ces propositions ne sont pas satisfaisantes car [HLW01] présente une solution partielle et les travaux dans [HCTS03, CH02, CTCH01] ne permettent pas de différencier clairement les co-occurrences d'événements (chocolat et pain) par rapport aux occurrences successives (chocolat) puis (pain). Or ce point est fondamental dans la définition même d'un motif qui intègre le concept de séquence. Notre proposition a pour objectif la définition d'une méthode adaptée et complète de traitement des données quantitatives par les motifs séquentiels flous.

Dans ce contexte, nous proposons une extension des méthodes existantes pour les règles d'association floues qui permettra l'extraction de motifs séquentiels flous. Tout d'abord nous définissons formellement le support d'une séquence floue à partir du support d'un itemset flou. Plusieurs algorithmes sont proposés pour le calcul du support des séquences de la base et la recherche des motifs séquentiels.

## References

- [CH02] R.-S. CHEN et Y.-C. HU : A novel method for discovering fuzzy sequential patterns using the simple fuzzy partition method. *Journal of the American Society for Information Science*, 54(7):660–670, 2002.
- [CTCH01] R.-S. CHEN, G.-H. TZENG, C.-C. CHEN et Y.-C. HU : Discovery of fuzzy sequential patterns for fuzzy partitions in quantitative attributes. In *ACS / IEEE International Conference on Computer Systems and Applications (AICCSA)*, pages 144–150, 2001.
- [HCTS03] Y.-C. HU, R.-S. CHEN, G.-H. TZENG et J.-H. SHIEH : A fuzzy data mining algorithm for finding sequential patterns. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 11(2):173–193, 2003.
- [HLW01] T.P. HONG, K.Y. LIN et S.L. WANG : Mining fuzzy sequential patterns from multiple-items transactions. In *Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, pages 1317–1321, 2001.
- [KFW98] C. M. KUOK, A. W.-C. FU et M. H. WONG : Mining fuzzy association rules in databases. *SIGMOD Record*, 27(1):41–46, 1998.
- [SA96] R. SRIKANT et R. AGRAWAL : Mining quantitative association rules in large relational tables. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pages 1–12, Montreal, Quebec, Canada, 4–6 1996.