Mining Conjunctive Sequential Patterns*

Chedy Raïssi[1,3], Toon Calders[2], and Pascal Poncelet[3]

[1] LIRMM, University of Montpellier, France
raissi@lirmm.fr
[2] Eindhoven University of Technology, Netherlands
t.calders@tue.nl
[3] LGI2P, Ecole des Mines d'Alès, France
pascal.poncelet@ema.fr

In this paper we study the discovery of frequent sequences and we aim at extending the non-derivable condensed representation in frequent itemset mining to sequential pattern mining. We start by showing a negative example: in the context of frequent sequences, the notion of non-derivability is *meaningless*.

This negative result motivated us to look at a slightly different problem: the mining of *conjunctions* of sequential patterns. This extended class of patterns turns out to have much nicer mathematical properties. For example, for this class of patterns we are able to extend the notion of non-derivable itemsets in a non-trivial way, based on a new unexploited theoretical definition of equivalence classes for sequential patterns. As a side-effect of considering conjunctions of sequences as the pattern type, we can easily form association rules between sequences. We believe that building a theoretical framework and an efficient approach for sequence association rules extraction problem is the first step toward the generalization of association rules to all complex and ordered patterns.

We also present a new depth-first approach to mine non-derivable conjunctive sequential patterns and show its use in mining association rules for sequences. This approach is based on a well known combinatorial theorem: the Möbius inversion. A performance study using both synthetic and real datasets illustrates the efficiency of our mining algorithm. These new introduced patterns have a high-potential for real-life applications, especially for network monitoring and biomedical fields with the ability to get non-redundant sequential association rules with all the classical statistical metrics such as confidence, conviction, lift etc.

References

1. Raïssi, C., Calders, T., Poncelet, P.: Mining conjunctive sequential patterns. Data Mining and Knowledge Discovery 17(1), 77–93 (August 2008)

* This is an extended abstract of an article published in the Data Mining and Knowledge Discovery journal [1]

W. Daelemans et al. (Eds.): ECML PKDD 2008, Part I, LNAI 5211, p. 19, 2008.