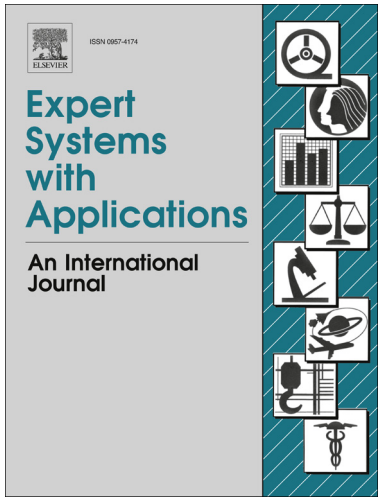

EXPEDITE: EXPress ClosED ITemset Enumeration

Giulio Aliberti, Alessandro Colantonio, Roberto Di Pietro, Riccardo Mariani

PII: S0957-4174(14)00810-0
DOI: http://dx.doi.org/10.1016/j.eswa.2014.12.031
Reference: ESWA 9756

To appear in: *Expert Systems with Applications*

Please cite this article as: Aliberti, G., Colantonio, A., Pietro, R.D., Mariani, R., EXPEDITE: EXPress ClosED ITemset Enumeration, *Expert Systems with Applications* (2014), doi: http://dx.doi.org/10.1016/j.eswa.2014.12.031

EXPEDITE: EXPress ClosED ITemset Enumeration

Giulio Aliberti[a], Alessandro Colantonio[b], Roberto Di Pietro[a,*], Riccardo Mariani[b]

[a] *Department of Mathematics and Physics, Roma Tre University*
[b] *Bay31 AG, Switzerland*

Abstract

In this paper, we introduce *EXPress closED ITemset Enumeration* (EXPEDITE), a new *Frequent Closed Itemset* (FCI) miner designed to speed up the process of FCIs extraction from a dataset of transactions. Compared to the state of the art, EXPEDITE provides a CPU time saving of up to two orders of magnitude without compromising other dimensions of performance (e.g. memory). The reason why it is so fast is that EXPEDITE wastes less time in mining intermediate item sets that are discarded in later phases of the algorithm. More specifically, it cuts down the number of both duplicate FCIs—those generated multiple times by the algorithm—and infrequent itemsets—those with low support or no supporting transactions. This feature, enjoyable by both sparse and dense datasets, is analytically motivated first, and then experimentally supported by extensive tests on real datasets. As a further contribution, we propose two alternative implementations of EXPEDITE that perform even better than the basic version, although they rely on particular features of the input dataset.

Keywords: data mining, knowledge discovery, closed itemsets, frequent itemsets, algorithms

*Corresponding author. L.go S. Leonardo Murialdo, 1. ZIP code: 00146 Roma, Italy. Phone: +39 06 90285694

Email addresses: aliberti@mat.uniroma3.it (Giulio Aliberti), alessandro@bay31.com (Alessandro Colantonio), dipietro@mat.uniroma3.it (Roberto Di Pietro), riccardo@bay31.com (Riccardo Mariani)

URL: http://www.bay31.com/ (Alessandro Colantonio), http://ricerca.mat.uniroma3.it/users/dipietro/ (Roberto Di Pietro)