Data characteristics and their relation to closed patterns discovery algorithms

Engelbert Mephu Nguifo

LIMOS, Clermont University, Blaise Pascal University and CNRS
Clermont-Ferrand, France
mephu@isima.fr

Abstract. Closed frequent patterns discovery remains a challenge in data mining. During the last decade, different works on data mining algorithms have based their performance evaluation on one dataset characteristic: its density (or on the contrary its sparseness). The incoming of massive datasets in different applications, points out the important goal to design efficient algorithms. The density measurement have shown to be a direction to reach such goal, especially when dealing with formal context of concept lattices. This talk will discuss this notion and describe some metrics defined to characterize dataset density for patterns discovery purpose.

References

1. Yahia, S.B., Hamrouni, T., Nguifo, E.M.: Frequent closed itemset based algorithms: a thorough structural and analytical survey. SIGKDD Explor. Newsl. **8**(1) (June 2006) 93–104
2. Boley, M., Grosskreutz, H.: Approximating the number of frequent sets in dense data. Knowl. Inf. Syst. **21**(1) (October 2009) 65–89
3. Emilion, R., Lévy, G.: Size of random galois lattices and number of closed frequent itemsets. Discrete Appl. Math. **157**(13) (July 2009) 2945–2957
4. Flouvat, F., Marchi, F., Petit, J.M.: A new classification of datasets for frequent itemsets. J. Intell. Inf. Syst. **34**(1) (February 2010) 1–19
5. Kuznetsov, S.O., Obiedkov, S.A.: Comparing performance of algorithms for generating concept lattices. Journal of Experimental & Theoretical Artificial Intelligence **14**(2-3) (2002) 189–216
6. Hamrouni, T., Yahia, S.B., Nguifo, E.M.: Looking for a structural characterization of the sparseness measure of (frequent closed) itemset contexts. Information Sciences (0) (2012) –
7. Négrevergne, B., Termier, A., Méhaut, J.F., Uno, T.: Discovering closed frequent itemsets on multicore: Parallelizing computations and optimizing memory accesses. In: HPCS. (2010) 521–528
8. Uno, T., Asai, T., Uchida, Y., Arimura, H.: An efficient algorithm for enumerating closed patterns in transaction databases. In: In Proc. DS '04, LNAI 3245. (2004) 16–31