How well go Lattice Algorithms on currently used Machine Learning TestBeds ?

Huaiguo Fu, Engelbert Mephu Nguifo

CRIL-CNRS FRE2499, Université d'Artois
Rue de l'université SP 16, 62307 Lens cedex. France
{fu,mephu}@cril.univ-artois.fr

Abstract. Many research papers in classification or association rules increase the interest of Concept lattices structures for data mining (DM) and machine learning (ML). To increase the efficiency of concept lattice-based algorithms in ML, it is necessary to make use of an efficient algorithm to build concept lattices. In fact, more than ten algorithms for generating concept lattices were published. As real data sets for data mining are very large, concept lattice structure suffers from its complexity issues on such data. The efficiency and performance of concept lattices algorithms are very different from one to another. So we need to compare the existing lattice algorithms with large data. We implemented the four first algorithms in Java environment and compared these algorithms on about 30 datasets of the UCI repository that are well established to be used to compare ML algorithms. Preliminary results give preference to Ganter's algorithm, and then to Bordat's algorithm, which do not fit well with the recommendations of Kuznetsov and Obiedkov. Furthermore, we analyzed the duality of the lattice-based algorithms.

1 Introduction

Concept is an important and basic means of knowledge representation, since it represents abstraction and generalization of objects. A concept defines a subset of objects which shares some common attributes or properties. Concept lattice structure [Ganter and Wille, 1999] has shown to be an effective tool for data analysis, knowledge discovery, and information retrieval, etc [Mephu Nguifo *et al.*, 2002]. It shows how objects can be hierarchically grouped together according to their common attributes. Researchers of different domains study it in theory and application of data analysis and formal knowledge representation etc.

Several algorithms are proposed to build concepts or concept lattices of a context : Bordat [Bordat, 1986], Ganter (NextClosure) [Ganter, 1984], Chein [Chein, 1969], Norris [Norris, 1978], Godin [Godin *et al.*, 1995], Nourine [Nourine and Raynaud, 1999], Carpineto [Carpineto and Romano, 1996], and Valtchev[Valtchev and Missaoui, 2001], etc. Some algorithms can generate also diagram graphs of concept lattices. The performance of the lattice algorithm is very important for its application to data mining (DM). In fact real data sets for DM are very large, e.g. the customer data of a company. In the worst case, the generation of lattice nodes increases exponentially. The efficiency of concept lattice algorithms are different from one to another. So we need