Mining Implications from Lattices of Closed Trees

José L. Balcázar*, Albert Bifet*, Antoni Lozano *

* Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
{balqui,abifet,antoni}@lsi.upc.edu

Abstract. We propose a way of extracting high-confidence association rules from datasets consisting of unlabeled trees. The antecedents are obtained through a computation akin to a hypergraph transversal, whereas the consequents follow from an application of the closure operators on unlabeled trees developed in previous recent works of the authors. We discuss in more detail the case of rules that always hold, independently of the dataset, since these are more complex than in itemsets due to the fact that we are no longer working on a lattice.

1 Introduction

In the field of data mining, one of the major notions contributing to the success of the area has been that of association rules. Many studies of various types have provided a great advance of the human knowledge about these concepts. One particular family of studies is rooted on the previous notions of formal concepts, Galois lattices, and implications, which correspond to association rules of maximum confidence.

These notions have allowed for more efficient works and algorithmics by reducing the computation of frequent sets, a major usual step towards association rules, to the computation of so-called closed frequent sets, a faster computation of much more manageable output size, yet losing no information at all with respect to frequent sets.

It was realized some time ago that the plain single-relational model for the data, as employed by the computation of either closed sets or association rules, whereas useful to a certain extent, was a bit limited in its applicability by the fact that, often, real-life data have some sort of internal structure that is lost in the transactional framework. Thus, studies of data mining in combinatorial structures were undertaken, and considerable progress has been made in recent years. Our work here is framed in that endeavor.

In previous work, we have proposed a mathematical clarification of the closure operator underlying the notion of closed trees in datasets of trees; the closure operator no longer works on single trees but on sets of them. In a sense, made precise there, closed trees do not constitute a lattice. A mathematically precise replacement lattice can be defined, though, as demonstrated in (Balcázar et al., 2006), consisting not anymore of trees but of sets of trees, and with the peculiar property that, in all experiments with real-life data we have undertaken, they turn out to be actually lattices of trees, in the sense that every closed set of trees was, in all practical cases, a singleton.