

Towards generic and efficient constraint-based mining, a constraint programming approach

Tias Guns*

*KU Leuven, Celestijnenlaan 200A, Leuven, Belgium
tias.guns@cs.kuleuven.be
<http://people.cs.kuleuven.be/tias.guns/>

1 Introduction

The fields of data mining and constraint programming are amongst the most successful subfields of artificial intelligence. Yet, their methodologies are quite different. Constraint programming advocates a declarative modeling and solving approach to constraint satisfaction and optimisation problems. Data mining on the other hand has focussed on handling large and complex datasets that arise in particular applications. Pattern mining more specifically aims to extract interesting patterns from a dataset, where *interestingness* is often defined by the application at hand. Current ad-hoc methods often focus on special-purpose algorithms to specific problems and interestingness criteria. This typically yields complex code that is very efficient, but hard to modify or reuse in other applications. Hence, less attention has been devoted to the issue of general and generic solution strategies.

Nevertheless, there is a need for generic techniques that can handle variations of known tasks, as well as application-driven constraints (Dzeroski et al., 2010; De Raedt et al., 2011). The typical iterative nature of the knowledge-discovery cycle (Han et Kamber, 2000), in which the data and problem definition are iteratively defined based on prototyping and small scale evaluations. In this case, the problem specification typically changes between iterations, which may in turn require changes to the algorithms.

This is acknowledged in the field of *constraint-based mining*, which adopts the methodology of formulating a problem in terms of constraints (Nijssen, 2010; Boulicaut et Jeudy, 2005). For example, for itemset mining (Agrawal et al., 1993), a wide variety of other constraints and a range of algorithms for solving these constraint-based itemset mining problems (Mannila et Toivonen, 1997; Jr. et al., 2000; Pei et Han, 2000; Pei et al., 2001; Bucila et al., 2003; Han et al., 2007; Soulet et Crémilleux, 2005; Bonchi et Lucchese, 2007) has enabled the application of itemset mining to numerous other problems, ranging from web mining to bioinformatics (Han et al., 2007). Generic frameworks in the constraint-based mining literature have focussed on the (anti-)monotonicity of constraints (Mannila et Toivonen, 1997; Pei et Han, 2000; Bucila et al., 2003) leading to systems such as ConQueSt (Bonchi et Lucchese, 2007), MusicDFS (Soulet et Crémilleux, 2005) and Molfea (De Raedt et Kramer, 2001). While many typical data mining tasks consist of (anti-)monotonic constraints, many other constraints do not fit in this framework, such as finding *closed* patterns in dense data (Pasquier et al., 1999;