Artificial Intelligence 244 (2017) 1–5

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Editorial

Introduction to the special issue on Combining Constraint Solving with Mining and Learning

1. Introduction

Data mining, machine learning and constraint solving are major themes in artificial intelligence research. They have evolved quite independently, though in recent years, there is a growing interest in the potential of integrating these fields.

Data mining and machine learning are methods for extracting regularities out of data, for example which instances cluster together, what patterns appear frequently in the data or what function can discriminate positive from negative examples. On the other hand, constraint solving investigates generic methods for solving constraint satisfaction and optimization problems.

Until now, these fields have evolved quite independently. Nevertheless, complex mining and learning tasks can often be formalized in terms of constraints that need to be satisfied, ranging from logical constraints to (non-linear) numeric ones. This methodology is becoming increasingly compelling, as the community is progressively moving from addressing relatively simple tasks on tabular data to complex problems on structured data. The use of machine learning and data mining techniques to enhance constraint solving is equally promising, and only starting to be explored. Consequently, there is a mutual benefit in the study of these approaches, which can lead to advances in both fields.

For this reason, a number of workshops have been organized on the topic, as well as this special issue. A combination, or integration, of Machine Learning (ML) and Data Mining (DM) with Constraint Solving (CS) can work in two directions: CS techniques can be used to (partly) solve tasks in DM/ML; Alternatively, one can use DM/ML techniques to augment CS techniques. Both directions are increasingly being explored. Yet, a proper approach in either direction requires good knowledge of both research fields, and hence feedback from both communities. The aim of this special issue is to bundle recent advances on combinations of the two fields, and to foster interaction between researchers of both communities.

The integration of machine learning and constraint solving is rooted in recent trends in these research areas. Fields like statistical relational learning [1], probabilistic programming [2] and structured-output prediction [3] naturally integrate learning algorithms with reasoning engines for addressing the underlying sets of constraints. SAT solvers [4], constraint solvers [5], mixed integer programming [6] as well as high-level modeling languages such as OPL [7] and MiniZinc [8] are extending both the scale and diversity of problems that can be addressed, as well as making the technology more accessible to other researchers.

A number of developments in recent years have extended the reach and applicability of integrations between the two fields. In particular: an increasing number of discrete and symbolic data mining and learning problems, for which other mathematical programming approaches are less suited; the rise of interest in 'constraint-based' mining and learning and the applicability of generic constraint solvers for this purpose; the interest in learning to predict structured outputs, namely complex entities such as trees and graphs, rather than class labels. In general, the increased presence of hard constraints in machine learning and data mining creates a need for generic and flexible methods that deal with such hard constraints. This need, combined with the advancement of both the speed of computer hardware as well as the scalability and maturity of constraint solving systems, is creating new opportunities for interaction between these methods.

There is also an increasing interest in using machine learning to improve the solving of constraint problems, as well as improving the modeling and the embedding of learned constraints and objectives into the model. This is driven by advances in computer hardware and the availability of large storage capacities and multiple processing units, combined with the increased availability of data regarding all aspects of the solving process. These include internal data regarding the solving and the effect of solver parameters that can be used to learn and automatically adapt the solving behavior; external data regarding the problem domain that can help in learning the parameters of constraints or even entire parts of the constraint model; and external data useful to infer objective functions or other functions that could not otherwise be formalized in terms of simple constraints. These integrations are non-trivial, and many research questions remain regarding how to best approach them.

http://dx.doi.org/10.1016/j.artint.2017.01.002

In the following we briefly introduce the papers in this special issue, grouping them according to the aspects of integration they focus on.

2. Machine learning and data mining with constraints

In today's data-rich world, one wishes to guide the learning and mining process by information on the type of knowledge we are seeking. A powerful way to express this is through constraints.

In recent years, constraint solving has been shown to offer a generic methodology that fits many mining and learning settings. This special issue contains recent advancements in the following broad research domains: constraint solving in pattern mining, clustering and learning.

2.1. Pattern mining using constraint solving

Pattern mining and constraint-based mining seeks to extract patterns from a large set of observations that satisfy the provided constraints, for example that a pattern must appear in a given number of observations. Constraint solving techniques provide a general framework for addressing mining tasks subject to a wide range of constraints. The following papers investigate the possibilities for different classes of tasks.

MiningZinc: A declarative framework for constraint-based mining *by Tias Guns, Anton Dries, Siegfried Nijssen, Guido Tack and Luc De Raedt*

The authors aim at bridging the gap between problem-specific pattern mining procedures and declarative problem solving methods. They present MiningZinc, a modeling language allowing to specify data mining tasks as constraint solving problems. The language is coupled with an execution mechanism which automatically extracts different execution strategies combining generic and specialized solvers. Tasks ranging from closed itemset mining to discriminative pattern mining and mining pattern sets can be modeled and solved in this framework.

Mining Top-k motifs with a SAT-based framework *by Said Jabbour, Lakhdar Sais and Yakoub Salhi*

This work introduces a generalization of partial max-SAT and X-minimal model computation called Top-k SAT. Given a preference relation, it finds those solutions that have fewer than k other solutions preferred to it. This framework and the proposed algorithm is applied to constrained itemset mining and sequence mining. The authors show how the approach can be used to mine top-k closed itemsets under size constraints, as well as top-k sequences of items and sequences of itemsets.

Skypattern mining: From pattern condensed representations to dynamic constraint satisfaction problems *by Willy Ugarte, Patrice Boizumault, Bruno Cremilleux, Alban Lepailleur, Samir Loudni, Marc Plantevit, Chedy Raissi and Arnaud Soulet*

The use of a preference relation is studied in this work from a different angle: the paper studies the task of finding all Pareto-optimal solutions given a set of measures. It investigates under what conditions the number of measures can be safely reduced, through *skylineability*. It then compares and unifies the traditional data mining approach of post-processing the output of a dedicated algorithm with the novel use of a generic constraint solver, showing the strengths and weaknesses of each.

2.2. Clustering using constraint solving

In clustering the goal is to group instances based on similarity. The following works demonstrate that constraint solving is applicable to many different forms of clustering.

Constrained clustering by constraint programming *by Thi-Bich-Hanh Dao, Khanh-Chuong Duong and Christel Vrain*

A main body of work is on clustering points into a partition. This work presents a generic partition-based clustering framework using constraint programming, which finds exact, optimal clusters. Multiple clustering measures as well as clustering constraints can be expressed and combined in this framework. It is also shown how it can be used for multi-objective clustering.

A flexible ILP formulation for hierarchical clustering *by Sean Gilpin and Ian Davidson*

A widely used clustering algorithm in practice is hierarchical clustering, which returns a *dendogram* or hierarchy instead of a partitioning. As the authors show this problem can be expressed as an Integer Linear Programming problem. This also allows for expressing additional constraints on the instances and the structure, as well as relaxing some constraints to find overlapping groups.

Cost-optimal constrained correlation clustering via weighted partial Maximum Satisfiability *by Jeremias Berg and Matti Jarvisalo*

Correlation clustering assumes that each pair of instances can be classified as being similar or not similar, and the goal is to find a clustering that correlates well with this classification. The authors show how this problem can be transformed