Preface

In industry, society, and science, advanced software is used for planning, scheduling, and allocating resources in order to improve the quality of service, reduce costs, or optimize resource consumption. Examples include power companies generating and distributing electricity, hospitals planning their surgeries, and public transportation companies scheduling their time-tables. This type of problem is often referred to as *constraint satisfaction and combinatorial optimization problems*.

Despite the availability of effective and scalable *solvers* that are applicable to a wide range of applications, current approaches to this problem are still unsatisfactory. The reason is that in all these applications it is very hard to acquire the constraints and criteria (that is, the *model*) needed to specify the problem, and, even if one has succeeded in capturing the model at one point, it is likely that it needs to be to changed over time to reflect changes in the environment. Therefore, there is an urgent need for optimizing and revising a model over time based on data that should be continuously gathered about the performance of the solutions and the environment they are used in.

Exploiting gathered data to modify the model is difficult and labour intensive with state-of-the-art solvers, as these solvers do not support data mining (DM) and machine learning (ML). However, existing frameworks for constraint satisfaction and combinatorial optimization problems do not support ML/DM techniques. In current ICT technology, DM and ML have almost always been studied independently from solving technology such as constraint programming (CP). On the other hand, a growing number of studies indicate that significant benefits can be obtained by connecting these two fields.

This led us to believe – almost five years ago – that it was the right time to develop the foundations of an integrated and cross-disciplinary approach to these two fields. A successful integration of CP and DM has the potential to lead to a new ICT paradigm with far-reaching implications that would change the face of DM/ML as well as CP technology. It would not only allow one to use DM techniques in CP to identify and update constraints and optimization criteria, but also to employ such constraints and criteria in DM and ML in order to discover models compatible with such prior knowledge. This book reports on the key results obtained on this research topic within the European FP7 FET Open project no. 284715 on "Inductive Constraint Programming" and a number of associated workshops and Dagstuhl seminars.

The book is structured in five parts. Part I contains an introduction to CP by Barry Hurley and Barry O'Sullivan and an introduction to DM by Valerio Grossi, Dino Pedreschi, and Franco Turini.

The next two parts address different challenges related to using ML and DM in a CP context. The first of these is the *model acquisition problem*, which aims at learning the different components of the CP model. This includes the identification of the domains to use, the constraints and possibly the preference or optimization function to be used. This is the topic of Part II. The first contribution, by Christian Bessiere, Abderrazak

Daoudi, Emmanuel Hebrard, George Katsirelos, Nadjib Lazaar, Younes Mechqrane, Nina Narodytska, Claude-Guy Quimper, and Toby Walsh, discusses an algorithm that acquires constraints by querying the user. The contribution by Nicolas Beldiceanu and Helmut Simonis describes a system for generating finite domain constraint models based on a global constraint catalog. The contribution by Luc De Raedt, Anton Dries, Tias Guns, and Christian Bessiere investigates the problem of learning constraint satisfaction problems from an inductive logic programming perspective. The contribution by Andrea Passerini discusses Learning Modulo Theories, a novel learning framework capable of dealing with hybrid domains.

The second challenge is that once the model is known, it needs to be solved. Reformulating models, optimizing the parameters of the solver, or considering alternative solvers is needed to solve the problem efficiently. Hints on how to improve a model and the best technique for solving it can be obtained by analyzing data collected during the run of solvers, or data collected from user studies. Part III reports on a number of techniques for model reformulation and solver optimization, that is: techniques for learning how to find solutions faster and more easily. In this part, a contribution by Lars Kotthoff provides a survey of algorithm selection techniques. Subsequently, Barry Hurley, Lars Kotthoff, Yuri Malitsky, Deepak Mehta, and Barry O'Sullivan present the Proteus portfolio solver and several improvements to portfolio techniques. Finally, Amine Balafrej, Christian Bessiere, Anastasia Paparrizou, and Gilles Trombettoni present techniques that adapt the level of consistency ensured by a solver during the search.

Part IV reports on the use of constraints and CP within a DM and ML context. This is motivated by the observation that many DM and ML tasks are essentially constraint satisfaction and optimization problems and that, therefore, they may benefit from CP principles and techniques. By specifying the constraints and optimisation criteria explicitly, DM and ML problem specifications become declarative and can potentially be solved by CP systems. Furthermore, several high-level modeling languages have been developed within CP that can potentially be applied or extended to ML and DM. The contribution by Anton Dries, Tias Guns, Siegfried Nijssen, Behrouz Babaki, Thanh Le Van, Benjamin Negrevergne, Sergey Paramonov, and Luc De Raedt introduces MiningZinc, a unifying framework and modeling language with associated solvers for DM and CP. Subsequently, Valerio Grossi, Tias Guns, Anna Monreale, Mirco Nanni, and Siegfried Nijssen show how many clustering problems can be formalized as constraint optimization problems.

Finally, Part V takes a more practical perspective. The first chapter by Christian Bessiere, Luc De Raedt, Tias Guns, Lars Kotthoff, Mirco Nanni, Siegfried Nijssen, Barry O'Sullivan, Anastasia Paparrizou, Dino Pedreschi, and Helmut Simonis reports on the iterative approach to inductive CP. The key idea is that the CP and ML components interact with each other and with the world in order to adapt the solutions to changes in the world. This is an essential need in problems that change under the effect of time, or problems that are influenced by the application of a previous solution. It is also very effective for problems that are only partially specified and where the ML component learns from observation of applying a partial solution, e.g., in the case of constraint acquisition. In addition, it reports on a number of applications of inductive CP in the areas of carpooling (with a contribution by Mirco Nanni, Lars Kotthoff, Riccardo Guidotti, Barry O'Sullivan, and Dino Pedreschi), health care (with a contribution by Barry Hurley, Lars Kotthoff, Barry O'Sullivan, and Helmut Simonis), and energy (with a contribution by Barry Hurley, Barry O'Sullivan, and Helmut Simonis).

The editors would like to thank the European Union for supporting the EU FET FP7 ICON project, the reviewers of the project, Alan Frisch, Bart Goethals, and Francesca Rossi, and the project officer Aymard de Touzalin for their constructive feedback and support, all participants of the ICON project for their contributions (Behrouz Babaki, Amine Balafrej, Remi Coletta, Abderrazak Daoudi, Anton Dries, Valerio Grossi, Riccardo Guidotti, Tias Guns, Barry Hurley, Nadjib Lazaar, Yuri Malitsky, Younes Mechqrane, Wannes Meert, Anna Monreale, Benjamin Negrevergne, Anastasia Paparrizou, Sergey Paramanov, Andrea Romei, Salvatore Ruggiero, Helmut Simonis, and Franco Turini), as well as the participants of the Dagstuhl seminars 11201 and 14411 and the Cocomile workshop series. Furthermore, they are grateful to all reviewers of chapters in this book.

September 2016

Christian Bessiere
Luc De Raedt
Lars Kotthoff
Siegfried Nijssen
Barry O'Sullivan
Dino Pedreschi

Springer
http://www.springer.com/978-3-319-50136-9

Data Mining and Constraint Programming
Foundations of a Cross-Disciplinary Approach
Bessiere, C.; De Raedt, L.; Kotthoff, L.; Nijssen, S.; O'Sullivan, B.; Pedreschi, D. (Eds.)
2016, XII, 349 p. 73 illus., Softcover
ISBN: 978-3-319-50136-9