

Preface

We are living in an age of great opportunity. There is an ever-rising flood of digital data from many sources, from tweets, photographs and social network communications to high-resolution sensor data across nearly all sciences. The great opportunity, and also the challenge, lies in making order out of chaos and deliver more usable information.

Not that there is a shortage of great data analysis tools. The fields of Knowledge Discovery in Databases (KDD) and Machine Learning (ML) have reached a mature stage, offering plenty of techniques to solve complex data analysis tasks on all types of data. The problem is that, with the increased sophistication of data analysis techniques, there is a greater than ever need for clear guidance on how to use these techniques.

Indeed, data needs to be analysed in often non-trivial refinement processes, which require technical expertise about methods and algorithms, skill in successful workflow design, experience with how a precise analysis should proceed, and knowledge about an exploding number of analytic approaches.

The underlying idea of this workshop is that constructing such workflows could be greatly facilitated by leveraging machine learning techniques. In other words, can't we *learn*, based on examples, how to select algorithms and build workflows? Research over the last decades has shown that this is in fact an inherently multidisciplinary problem, requiring an multidisciplinary solution. We can distinguish the following themes that have emerged in the literature:

Planning, from the field of Artificial Intelligence, provides a principled approach to generate and build workflows. Indeed, building a knowledge discovery workflow can be cast as a planning problem in which (sub)tasks such as data preprocessing and modelling must be achieved by devising an optimal sequence of operators.

Ontologies, from the field of Information Science, allow us to express knowledge about data, algorithms and model in such a way that systems can interpret and reason about these concepts. It yields precise descriptions of data format requirements so that workflows don't break, and express the inner working of algorithms so we can reason about how to tune them.

Meta-learning, from the field of Machine Learning, allows us to build models about algorithm behavior under various conditions, and thus to predict which (combinations of) techniques are most useful. Meta-learning approaches have grown ever more sophisticated, building on ontological information about algorithms, large databases of machine learning experiments, and domain-specific data features.

It is interesting to see how these fields interact: ontologies lead to better meta-learning techniques, which in turn allow for more accurate workflow planning, in turn providing more examples from which to learn. This workshop, therefore, is particularly aimed at exploring the possibilities of integrating these fields. It offers a forum for exchanging ideas and experience concerning the state-of-the-art from these different areas and outline new directions for research.

These proceedings include 7 contributions on the latest advances in this area, many of them integrating planning, ontologies and meta-learning as discussed above. Moreover, they include several demonstrations of working systems in this area, which will be of great use to end users. We thank everybody for their sincere interest and their contributions, and especially thank our invited speakers:

Ross King (Manchester University): *The potential of automated workflow planning in the Robot Scientist*.
Filip Zelezny (Czech Technical University): *Planning to learn: Recent developments and future directions*.

We hope you will find it an interesting and inspiring workshop, leading to great new collaborations.

Leiden, August 2012

Joaquin Vanschoren
Pavel Brazdil
Jörg-Uwe Kietz

Main areas covered by the workshop

Of particular interest are methods and proposals that address the following issues:

- Planning to construct workflows
- Exploitation of ontologies of tasks and methods
- Representation of learning goals and states in learning
- Control and coordination of learning processes
- Experimentation and evaluation of learning processes
- Recovering / adapting sequences of DM operations
- Meta-learning and exploitation of meta-knowledge
- Layered learning
- Multi-task learning
- Transfer learning
- Active learning
- Multi-predicate learning (and other relevant ILP methods)
- Learning to learn
- Learning to plan
- Intelligent design and learning

A Brief History

PlanLearn-2012 is the 5th workshop in the series, which has been collocated with a range of conferences:

- PlanLearn-2007 was associated with ECML/PKDD-07 in Warsaw, Poland. The invited speaker was Larry Hunter (Univ. of Colorado at Denver and Health Sciences Center) who presented a talk entitled *Historical Overview of the area Planning to Learn*.
- PlanLearn-2008 was associated with ICML/COLT/UAI in Helsinki, Finland. The invited speaker was Raymond J. Mooney (University of Texas at Austin) who presented a talk on *Transfer Learning by Mapping and Revising Relational Knowledge*.
- PlanLearn-2010 was associated with ECAI in Helsinki, Finland. Invited speakers were Michele Sebag (LRI, U. Paris-Sud) on *Monte-Carlo Tree Search: From Playing Go to Feature Selection*, and Luc de Raedt (Katholieke Universiteit Leuven) on *Constraint Programming for Data Mining*.
- PlanSoKD-2011 was held in conjunction with the Service-Oriented Knowledge Discovery workshop at ECML/PKDD in Athens, Greece.

More information on the workshop and previous editions can be found at:

<http://datamining.liacs.nl/planlearn.html>

Acknowledgements

Many thanks to Rui Leite for typesetting these proceedings.

Workshop Organization

Workshop Chairs

Joaquin Vanschoren (LIACS, Leiden University)

Pavel Brazdil (LIAAD-INESC Porto L.A. / FEP, University of Porto)

Jörg-Uwe Kietz (Department of Informatics, University of Zürich)

ECAI Workshop Chair

Jérôme Lang and Michèle Sebag

Workshop Program Committee

Michael Berthold, U. Konstanz, Germany, and KNIME

Pavel Brazdil, LIAAD-INESC Porto L.A. / FEP, University of Porto, Portugal

André Carlos Ponce de Carvalho, USP, Brasil

Claudia Diamantini, Università Politecnica delle Marche, Italy

Johannes Fuernkranz, TU Darmstadt, Germany

Christophe Giraud-Carrier, Brigham Young University, USA

Krzysztof Grabczewski, Nicolaus Copernicus University, Poland

Melanie Hilario, U Geneva, Zwitterland

Norbert Jankowski, Nicolaus Copernicus University, Poland

Alexandros Kalousis, U Geneva, Zwitterland

Pance Panov, Jozef Stefan Institute, Slovenia

Bernhard Pfahringer, U Waikato, New Zealand

Vid Podpecan, Jozef Stefan Institute, Slovenia

Ricardo Prudêncio, Universidade Federal de Pernambuco Recife (PE), Brasil

Stefan Rüping, FhG-IAIS, Germany

Floarea Serban, University of Zurich, Switzerland

Carlos Soares, FEP, University of Porto, Portugal

Maarten van Someren, University of Amsterdam, The Netherlands

Vojtech Svatek, Univ. of Economics, Prague, Czech Republic

Ljupco Todorovski, Jozef Stefan Institute, Slovenia

Joaquin Vanschoren, U. Leiden / KU Leuven

Michel Verleysen, UC Louvain, Belgium

Ricardo Vilalta, University of Houston, USA

Filip Zelezny, Czech Technical University, Czech Republic

Technical support

Management of EasyChair: Joaquin Vanschoren, Pavel Brazdil and Jörg-Uwe Kietz