Preface

Data Mining on Linked Data Workshop (DMoLD'13) with Linked Data Mining Challenge was held on 23 September 2013 in Prague, Czech Republic. It was part of ECML/PKDD2013, the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. It was the first workshop in a series foreseen for researchers and practitioners interested in ‘KDD first’.

Linked data (LD), published on the web in RDF format, has been so far nearly untouched by advanced data mining methods. Due to the nature of data it poses specific challenges. There is a larger variety of link types in the graph. Presence of semantic links between datasets makes data more complex. It also allows inclusion of new linked datasets into the mining dataset nearly on the fly, making the feature selection problem extremely hard. Moreover, there are heterogeneous data models, with varying degrees of completeness and credibility. The potential of interlinking to external datasets is tempting. The series should provide forum for discussions on adapting techniques from KDD rather than inventing LD-tailored approaches from scratch.

The workshop consisted of a keynote speech, an open track, a challenge track, and a discussion panel. The Program Committee consisting of 13 members accepted two papers for the open track and two for the challenge track. As the workshop was foreseen mostly for KDD community it featured the challenge event addressing a real business KDD problem rather than problems specific to Semantic Web data handling. In 2013 the core of the Linked Data Mining Challenge (LDMC) held as part of the workshop was a dataset from the highly topical public procurement domain.

Heiko Paulheim gave a keynote talk on “Exploiting Linked Open Data as Background Knowledge in Data Mining.”

The workshop was supported by the grant from European Union’s 7th Framework Programme provided for the project LOD2 Creating Knowledge out of Interlinked Data (GA no. 288176).

August 2012

Claudia d’Amato
Petřík Berka
Vojtěch Svátek
Krzysztof Węcel
Conference Organization

Programme Chairs

Claudia d’Amato, University of Bari, Italy
Peter Berka, University of Economics, Prague, Czech Republic
Vojtěch Svátek, University of Economics, Prague, Czech Republic
Krzysztof Węcel, Poznan University of Economics, Poland

Program Committee

Jose María Alvarez-Rodríguez, South East European Research Center, Greece
Sören Auer, Universität Leipzig, Germany
Bettina Berendt, KU Leuven, Belgium
Marko Grobelnik, JSI, Ljubljana, Slovenia
Fazel Famili, NRC Institute for Information Technology Ottawa, Canada
Nicola Fanizzi, University of Bari, Italy
Agnieszka Ławrynowicz, Poznan University of Technology, Poland
Pablo Mendes, Free University Berlin, Germany
Martin Nečaský, Charles University Prague, Czech Republic
Zbigniew Raś, University of North Carolina, Charlotte, USA
Jerzy Stefanowski, Poznan University of Technology, Poland
Volker Tresp, University of Munich, Germany
Djamel Zighed, University Lumière, Lyon, France

Additional Reviewers

Domain experts in public contract data analysis:

- Jana Chvalkovská, Charles University, Prague
- Jiří Skuhrovec, Charles University, Prague
Table of Contents

Invited Talk

Exploiting Linked Open Data as Background Knowledge in Data Mining .
Heiko Paulheim

Open Track

Lattice Based Data Access (LBDA): An Approach for Relating Data and Linked Open Data in Biology ..........................
Mehwish Alam, Melisachew Wudage Chekol, Adrien Coulet, Amedeo Napoli and Malika Smail-Tabbone

A Fast and Simple Graph Kernel for RDF ......................
Gerbens Dirk de Vries and Steven de Rooij

Challenge Track

Linked Data Mining Challenge (LDMC) 2013 Summary ........
Vojtech Svatok, Jindrich Mynarz, Petr Berka

Graph Kernels for Task 1 and 2 of the Linked Data Data Mining Challenge 2013 ..............................................
Gerbens Dirk de Vries

A Machine Learning approach to the Linked Data Mining Challenge 2013
Eneldo Loza Mencia, Simon Holthausen, Axel Schulz and Frederik Janssen