Mach Learn (2016) 105:1–2
DOI 10.1007/s10994-016-5589-1

EDITORIAL

Special issue on discovery science

Sašo Džeroski[1] · **Dragi Kocev**[1] · **Panče Panov**[1]

Received: 28 July 2016 / Accepted: 1 August 2016 / Published online: 8 September 2016

This special issue focuses on the topic of discovery science (DS). DS is a research discipline concerned with the development and analysis of methods for discovering scientific knowledge, coming from machine learning, data mining, and intelligent data analysis, as well as the applications of such methods in various scientific domains, including medicine, the natural sciences and the social sciences. Furthermore, DS considers the analysis of different types of complex data, such as structured, spatio-temporal and network data.

The special issue was preceded by the 17th International Conference on Discovery Science, held in Bled, Slovenia (October 8–10, 2014). In our open call for papers, we solicited submissions in all areas of discovery science. These included submissions focusing on the analysis of different types of complex data, such as structured, spatio-temporal and network data, and submissions addressing applications in scientific domains, such as environmental and life sciences. We received 17 diverse submissions showing the liveliness and the breadth of this field. Of the received submissions, we eventually selected four for inclusion in this special issue. The accepted articles have undergone two rounds of rigorous peer-review according to the journal's high standards.

The accepted contributions encompass a wide range of research topics, such as learning mixture models, mining subgraphs, learning rules for multi-label classification and anomaly detection, thereby appealing to both the experts in the respective fields and those who want a snapshot of the current breadth of topics covered by discovery science. Three papers are more on the theoretical/methodological side, while one is more on the application side. The accepted articles are briefly summarized below.

In "Explaining mixture models through semantic pattern mining and banded matrix visualization", Prem Raj Adhikari, Anže Vavpetič, Jan Kralj, Nada Lavrač, and Jaakko Hollmén present an approach to semi-automated data analysis using tools for pattern construction, exploration and explanation. The approach is tailored to multiresolution 0–1 data analysis and includes data clustering with mixture models, extraction of rules from clusters, and data/rule visualization using banded matrices. The use of an ontology describing the relation-

✉ Dragi Kocev
Dragi.Kocev@ijs.si

[1] Jožef Stefan Institute, Ljubljana, Slovenia

ships between the different resolutions facilitates the application to multiresolution data. The method is showcased on complex DNA copy number amplification data for which semantic pattern and cluster/pattern visualization was performed. Moreover, an evaluation on four public datasets shows the generality of the proposed methodology.

The article "Subjective interestingness of subgraph patterns", by Matthijs van Leeuwen, Tijl De Bie, Eirini Spyropoulou, and Cédric Mesnage addresses the issue of balancing between approximating actual interestingness and computational efficiency in subgraph mining. The authors propose to resolve the issue by formally specifying what makes a dense subgraph pattern interesting to a given user by considering users' prior beliefs. Two cases are considered here: (1) the user only has a belief about the overall density of the graph, and (2) the user has prior beliefs about the degrees of the vertices. Contrary to most of the existing approaches, the proposed method naturally allows for the subsequently found patterns to be overlapping. The empirical evaluation highlights the properties of the new interestingness measure given different prior belief sets, and the approach's ability to find interesting subgraphs that other methods failed to find.

In "Learning rules for multi-label classification: A stacking and a separate-and-conquer approach", Eneldo Loza Mencía and Frederik Janssen introduce two approaches for learning label-dependent rules for multi-label classification. The first solution is a bootstrapped stacking approach, built on top of a conventional rule learning algorithm by learning a separate ruleset for each label and including the remaining labels as additional attributes in the training instances. The second approach adapts the commonly used separate-and-conquer algorithm for learning multi-label rules by re-including the covered examples with the predicted labels so that this information can be used for learning subsequent rules. The experimental evaluation reveals (a) that the discovered dependencies contribute to the understanding and improved analysis of multi-label datasets, and (b) that the found multi-label rules are crucial for the predictive performance as the proposed approaches are better than conventional rules.

The article "Sequential anomalies: a study in the railway industry," by Rita P. Ribeiro, Pedro Pereira, and João Gama discusses an application of predictive maintenance for train doors. More specifically, a failure detection system is used to predict train door breakdowns before they happen using data from their logging system; i.e., sensor data from pneumatic valves that control the open and close cycles of a door. One of the major issues here is that a failure of a cycle does not necessarily indicates a breakdown, as a cycle might fail due to user interaction. The goal is to detect structural failures in the automatic train door system. Three methods for such structural failure detection are studied; i.e., outlier detection, anomaly detection, and novelty detection, using different windowing strategies. The proposed solution is a two-stage approach, where the output of a point-anomaly algorithm is post-processed by a low-pass filter to obtain a subsequence-anomaly detection. The proposed solution strongly reduces the false alarm rate.

In sum, this issue includes an exciting and diverse set of papers on discovery science: We hope that the readers will enjoy them. We want to thank all the authors and reviewers for the time invested in putting this special issue together. Special thanks are due to Michelangelo Ceci for serving as the action editor for the paper entitled "Explaining mixture models through semantic pattern mining and banded matrix visualization."

Sašo Džeroski, Dragi Kocev, Panče Panov, July 28th, 2016