Guest Editorial: Special issue on Discovery Science

Takuya Kida[1] · Tetsuji Kuboyama[2] · Takeaki Uno[3] · Akihiro Yamamoto[4]

This special issue focuses on Discovery Science, the research field that develops and analyzes methods for discovering scientific knowledge. According to the growth of massive and complex data stored in databases and distributed via networks, knowledge discovery methods from such data have been developed actively. Discovery Science occurs between developing methods and research in various scientific domains. The methods are technically based on machine learning, data mining, intelligent data analysis, and big data analysis, and aimed at discovering new scientific knowledge which could not be found with previous methods.

This special issue follows the 20th International Conference on Discovery Science, held in Kyoto, Japan (October 15–17, 2017). We opened a call for papers to solicit submissions on Discovery Science, received 12 submissions, and selected four for inclusion in this issue. The articles have been rigorously peer-reviewed according to the journal's high standards. Every accepted contribution is concerned with methods for knowledge discovery based on mathematical definitions, and has results obtained by applying the methods to data in a specific scientific area. The accepted articles are briefly summarized below. The proposed methods have mathematical foundations, and therefore, would be applicable to data in areas other than those addressed in the articles. The latter two articles aim to remove the limitation of methods popular in a specific research area.

In the article "Effective Approximation of Parametrized Closure Systems over Transactional Data Streams," Daniel Trabold, Tamás Horváth, and Stefan Wrobel propose an algorithm for mining strongly closed itemsets from transactional data streams. The class of strongly closed itemsets is very compact but rich enough in extracting knowledge, and

✉ Akihiro Yamamoto
  akihiro@i.kyoto-u.ac.jp

  Takuya Kida
  kida@ist.hokudai.ac.jp

  Tetsuji Kuboyama
  kuboyama@tk.cc.gakushuin.ac.jp

  Takeaki Uno
  uno@nii.ac.jp

[1] Hokkaido University, Sapporo, Japan

[2] Gakushuin University, Tokyo, Japan

[3] National Institute of Informatics, Tokyo, Japan

[4] Kyoto University, Kyoto, Japan

stable against the databases. The authors assume that the transactions are provided as a stream, and they design the mining algorithm with the algebraic and algorithmic properties of strongly closed itemsets. They demonstrate the strongly closed itemsets work as indicators of concept drifts in transactional data streams. Through application to computer-aided product configuration, they realize concept identification by proposing an algorithm of "learning with subset queries."

The article "Feature Ranking for Multi-Target Regression" by Matej Petković, Dragi Kocev, and Sašo Džeroski examines multi-target regression (MTR, for short), a model used when there are multiple dependent variables. MTR is now attracting attention in solving real life problems in various fields. For ranking descriptive features in MTR, the authors propose scores of the features in two approaches, algorithm adaptation and data transformation, by extending a score previously proposed.

In the article "Ranking by Inspiration: A Network Science Approach" by Livio Bioglio, Valentina Rho, and Ruggero Gaetano Pensa, a new definition of influence, "the inspiration rank," is introduced in order to capture the diffusion of inspiration within a network, particularly in the citation network of bibliography. Although the models of contagion processes in epidemiology and life science are applied to investigation of the diffusion of information, they are not complex enough because scientific and cultural aspects affect such diffusion. The authors show that the inspiration rank is effective in detecting important information in a citation network built upon large bibliographic datasets, and they assert that it can be used as an alternative or complementary bibliographic index.

In the article "Exploiting Causality in Gene Network Reconstruction Based on Graph Embedding," Gianvito Pio, Michelangelo Ceci, Francesca Prisciandaro, and Donato Malerba propose a new inductive method for reconstruction of gene networks, which is an important task in bioinformatics. The method is developed based on graph embedding to identify indirect relationships of arbitrary length to remove links, and take into account possible community structures and possible similar roles.

In summary, this special issue includes papers demonstrating strong research results in Discovery Science, which treats the development of discovery methods itself as science, and it also applies the developed methods to practical data. We hope the readers enjoy them. We would like to express our gratitude to all of the authors, reviewers, and the editorial staff of the *Machine Learning* journal.