Mach Learn (2011) 82: 91–93
DOI 10.1007/s10994-010-5209-4

EDITORIAL

Introduction to the special issue on mining and learning with graphs

S.V.N. Vishwanathan · Samuel Kaski · Jennifer Neville · Stefan Wrobel

Received: 1 June 2009 / Accepted: 1 May 2010 / Published online: 21 August 2010

Driven by application areas ranging from biology to the World Wide Web, research in Data Mining and Machine Learning is nowadays increasingly focusing on the analysis of structured data. Of particular interest are data that consist of interrelated parts or data that are characterized by collections of interrelated objects, linked together into complex graphs and structures. Dealing with such interrelated data is one of the major research challenges that we are facing. The aim of this special issue is to bring together papers from different sub-disciplines within Machine Learning and Data Mining that focus on the analysis of structured data.

This special issue was first conceived as a forum for publishing extended versions of papers presented at the annual workshop on Machine Learning and Graphs (MLG) 2008 co-located with the International Conference on Machine Learning (ICML) 2008 in Helsinki, Finland. Later, the scope was extended and we solicited relevant papers from researchers in all areas of machine learning and data mining, working on mining and learning with graphs

S.V.N. Vishwanathan (✉)
Departments of Statistics and Computer Science, Purdue University, 250 N University Street, West Lafayette, IN 47907-2066 USA
e-mail: vishy@stat.purdue.edu

S. Kaski
Department of Information and Computer Science, Aalto University School of Science and Technology, P.O. Box 15400, 00076 Aalto, Finland
e-mail: samuel.kaski@tkk.fi

J. Neville
Departments of Computer Science and Statistics, Purdue University, 305 North University Street, West Lafayette, 47907-2107, USA
e-mail: neville@cs.purdue.edu

S. Wrobel
Fraunhofer IAIS and Department of Computer Science, University of Bonn, Schloss Birlinghoven, 53754 Sankt Augustin, Germany
e-mail: stefan.wrobel@iais.fraunhofer.de

and relations. With help from 38 reviewers, we selected 6 papers for publication out of the 17 submissions.

The first paper *Efficiently Mining δ-tolerance Close Frequent Subgraphs*, by Takigawa and Mamitsuka focuses on an important aspect of current research on Pattern Mining in Graphs—the identification of problem abstractions that allow efficient discovery by reducing the number of relevant patterns in a reasonable way. The authors consider a parameterized problem setting where the parameter δ allows to continually blend the problem characteristics from closed frequent subgraphs to maximal frequent subgraphs. They propose a novel algorithm for this setting based on the idea of reverse search, which while maintaining the completeness of enumeration still allows for better pruning. As shown in the experiments, the parameter is effective at controlling the size of the output set, and the resulting algorithm allows discovery to be performed on real-world databases.

In their paper *Multi-way Set Enumeration with Weight Tensors*, Georgii et al. consider the problem of mining associations relating multiple types of instances, which are often represented by multi-way arrays or tensors. Examples of such data could include sales figures of a company based on the products, region, and time of the year. They generalize the notion of frequent item sets to n-sets and propose an enumerative mining algorithm for these patterns. These n-set patterns provide a higher-level view of the data, revealing associative relationships between groups of instances. They also show that their algorithm can tolerate missing observations to a certain degree and also take association weights into account.

Next, although many networks are naturally evolving over time, much of the past work in modeling graphs has focused on analyzing the structure of static graphs (i.e., with a fixed set of nodes and edges). Yang et al., in their paper *Detecting Communities and Their Evolutions in Dynamic Social Networks—A Bayesian Approach*, move beyond the assumption of static data and explicitly consider modeling a series of snapshot graphs which evolve over time. To reason about the evolution of communities in these snapshots, they develop a dynamic stochastic blockmodel with both online and offline methods for inference. The strength of the approach is that it explicitly models transitions in community memberships over time, within a Bayesian probabilistic model. The authors demonstrate the performance of the model on both synthetic and real world network datasets.

The learning of trust and distrust is a crucial aspect of social interaction among autonomous, mentally-opaque, networks of agents. The issue of trust learning, based on past observations and context information, is addressed in the paper *Statistical Relational Learning of Trust* by Rettinger et al. They show how to implement and learn context-sensitive trust using statistical relational learning in form of a Dirichlet process mixture model called Infinite Hidden Relational Trust Model (IHRTM). They demonstrate the effectiveness of their technique on user-ratings gathered from eBay.

Topic models of texts have been a major research direction in machine learning, and link-based ranking a major success story in information retrieval in the web. The paper *Topic Level Expertise Search over Heterogeneous Networks* by Tang et al. combines the two into a machine learning project that "works", in the words of one of the reviewers. State-of-the-art ideas are used to model jointly the contents and citations, and random walk at the topic level is used to do the ranking. The model is used in a working system for searching for experts and for citation influence studies.

Finally, a central issue in artificial intelligence is how to develop agents that learn and act in complex environments. Unfortunately, realistic environments are hard to model because they feature a variable number of objects, complex relations between them, and non-deterministic transition behavior. In this case, standard probabilistic sequence models may not be effective in capturing the relational complexity. On the other hand, statistical relational learning techniques might be too inefficient to cope with complex sequential data.

In the paper *Stochastic relational processes: Efficient inference and applications* Thon et al. introduce a simple model that occupies an intermediate position in this expressiveness/efficiency trade-off. It is based on CP-logic (Causal Probabilistic Logic), an expressive probabilistic logic for modeling causality. However, by specializing CP-logic to represent a probability distribution over sequences of relational state descriptions and employing a Markov assumption, inference and learning become more tractable and effective.

These six papers represent the breadth and diversity of work that focuses on mining and learning with graphs. The common themes include (1) a tradeoff between representation complexity and algorithmic efficiency, and (2) a focus on developing mining and learning algorithms for realistic real-world environments.

Acknowledgements We are much indebted to our authors and reviewers, who put a tremendous amount of effort to make this issue happen. We are also thankful to Foster Provost for constant guidance and encouragement. The CMT team at Microsoft were very responsive with technical requests and their help is gratefully acknowledged. Vishwanathan and Neville were supported by NSF grant IIS-0916686.