Preface to the IEEE ICDM 2011 PhD Forum

Alfredo Cuzzocrea
ICAR-CNR and University of Calabria
Rende, Italy
cuzzocrea@si.deis.unical.it

Rosa Meo
University of Torino
Torino, Italy
meo@di.unito.it

I. RATIONALE

For the first year, ICDM hosts a Forum dedicated to PhD students. The aim of the ICDM PhD Forum is to provide an international environment in which PhD students can meet, exchange their ideas and experiences with peers and senior researchers from the Data Mining Community, in an international scope. Here, PhD students have a unique opportunity to present their ideas and discuss on the work-in-progress in preparation of the PhD dissertation and on the major interests in the Data Mining field.

The PhD Forum spans over all the topics of Data Mining and other research fields in which Data Mining benefits from cross-fertilization, such as: machine learning, statistics, databases, natural language processing, information retrieval, WWW, data visualization, multimedia, bioinformatics, knowledge-based systems, pattern matching and high performance computing.

II. DESCRIPTION

The PhD Forum has attracted 14 submissions from 11 countries all over the world (Canada, China, France, Hong Kong, India, Italy, Japan, Slovenia, Taiwan, UK, USA). Each paper has been reviewed by at least 3 program committee members among people of internationally-renowned value. Finally, 6 papers have been selected for the final presentation at the PhD Forum. These papers cover a broad spectrum of topics and reflect recent advances in Data Mining research in theoretical as well as practical contexts.

III. CONTRIBUTIONS

Next, we provide a brief overview of the 6 contributions finally selected.

In On Efficient Distance-based Similarity Search J. Liu, H. Chen, K. Furuse, H. Kitagawa and J. Xu Yu address two sub-problems within the broad topic of similarity search, focusing on the enhancement of search efficiency based on their common clue “distance”. One sub-problem is solving efficiently the fundamental query types, k-nearest neighbor and range queries that regard distance comparison in term of nearness. The other sub-problem is a relatively special query type, i.e., the reverse furthest neighbor query (RFN), oppositely considering the distance in term of farness. For the former sub-problem, the authors propose an original index scheme, called “function index”; whose purpose is to index expensive distance functions for efficient query processing in the multi-dimensional (even high-dimensional) space. For the latter sub-problem the authors concentrate on the theoretical analysis and on the algorithm design to enhancing the query efficiency.

In Active Learning of Transfer Relationships for Multiple Related Bayesian Network Structures, D. Oyen presents a novel algorithm that learns multiple Bayesian networks for a collection of unsupervised machine learning tasks when limited data is available and a metric of the relatedness of the tasks can be obtained from metadata. The novelty of this proposal replies in a first multitask network learning algorithm that incorporates domain knowledge about the relatedness of the tasks, contrary to traditional approaches that are purely data-driven.

In Understanding and Exploiting the Connections between NMF and SVM, V. Potluru explores the connections between Support Vector Machines (SVM) and Nonnegative Matrix Factorization (NMF), as standard tools for data analysis, to importing algorithms from the SVM world to solve the NMF problem and vice-versa. In particular, one such algorithm from the SVM world is adapted to solve NMF. The preliminary results clearly show that the performance of this algorithm is competitive with one solver of the state-of-the-art in NMF.

In Community Evolution in Dynamic Social Networks - Challenges and Problems, M. Takaffoli focuses the attention on the dynamic aspects of social networks. As the author recognizes, studying the evolution of these networks over time could provide insight into the changes that occurred in the patterns of the users' interaction and also into the future trends of the networks. Furthermore, in a dynamic scenario, the analysis of communities and their evolution can help to determine the characteristics and the structural properties of the network. With these ideas in mind, an overview of the existing research and challenges in the area of dynamic social network analysis is provided.

In Nearest Neighbor Voting in High-Dimensional Data: Learning from Past Occurrences, N. Tomasev and D. Mladenic propose a new approach for exploiting the so-called hubness phenomenon, a recently described aspect of the curse of dimensionality inherent to the nearest-neighbor methods, in the context of the high-dimensional k-nearest neighbor classification. The authors argue that some of the neighbor occurrences carry more information than others, by the virtue of being less frequent events. Based on this main intuition, a new algorithm, Hubness Information k-Nearest Neighbor (HIKNN), is proposed and experimentally assessed.

Finally, in Towards a Framework for Detecting and Managing Opinion Contradictions, M. Tsytseraru focuses the attention on Sentiment Analysis, which has recently gained
in interest due to the large amount of potential applications and the increasing number of opinions expressed, in particular in the Web. The proposal is represented by a framework on top of sentiment analysis for detecting contradictions. The framework introduces a statistical model of contradictions based on the mean value and variance of sentiments among different posts. The author also demonstrates the effectiveness and the scalability of the proposed method in capturing contradictions on both the noisy synthetic data and the real-world one.

IV. ACKNOWLEDGMENTS

First, we would like to express our sincere gratitude to Prof. Osmar Zaiane for proposing, for the first time, the PhD Forum at ICDM. We are confident that this will be the first edition of a very fortunate series. We also would like to thank the authors of all submitted papers. Their innovation and creativity has resulted in a pretty strong technical program. Finally, we are highly indebted to the program committee members, whose reviewing efforts ensured in selecting a competitive and scientifically valid technical program.