PhD Forum

Preface

For the second year, ICDM hosts a PhD Forum on Data Mining. The aim of the Forum is to provide an international environment in which students can meet and exchange their ideas and experiences both with peers and with senior researchers from the Data Mining Community. The Forum is particularly aimed at PhD students in the early stages of their career and Master’s students planning to pursue their research in a PhD programme.

The PhD Forum spans 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.

The PhD Forum has attracted 28 submissions from 16 countries all over the world (Algeria, Belgium, China, France, Germany, Greece, Honk-Kong, India, Iran, Ireland, Malaysia, New Zealand, Nigeria, Saoudi Arabia, UK, USA). Each paper has been reviewed by at least 2 program committee members among people of internationally-renowned value. Finally, 8 submissions have been selected as regular papers for an oral presentation at the PhD Forum, and 10 submissions have been accepted as poster presentations. These submissions cover a broad spectrum of topics and reflect recent advances in Data Mining research in theoretical as well as practical contexts.

Below we give a brief overview of the contributions, which will appear in the workshop. For this we classify the contributions into groups based on their theme.

This classification is shown below.

- **Core data mining and machine learning**: These contributions focus on core data-mining and machine-learning themes including classification through regression, rule selection and text mining. The contributions of this category are: “Selecting accurate and comprehensible regression algorithms through meta learning”, “Active learning based rule extraction for regression”, “Effective text classification by a supervised feature selection approach”, and “Using conceptual clustering to improve the accuracy of classification”.

- **Cross-computer science** papers: These are contributions that apply techniques developed in one computer-science area to solve a problem in another area—again within computer science. The contributions of this type are: “Multi-slice modularity optimization in community detection and image segmentation”, “Towards a particle swarm optimization-based regression rule miner”, and “Kernel design for Internet traffic data aggregation and classification”.

- **Data mining for (social) networks and social media**: The papers focusing on this theme relate to social networks and social media, and are entitled: “Modeling of collective synchronous behavior on social media” and “Adaptive recommendations based on data mining of user behavioral patterns”. These papers are primarily concerned with modeling (social-network) users so that they provide them with better recommendations.
• **Privacy and data management**: Two contributions of the workshop are concerned with privacy preserving data analysis and secure data analysis on the cloud. The titles of these papers are: "Sorted neighborhoods for multidimensional privacy-preserving blocking" and "Modeling confidentiality archetype and data mining in cloud computing".

• **Data mining for bibliometrics and job-market analysis**: These contributions are entitled: "How third-party funding institutions thematically influence job advertisements for professorial appointments at German universities" and "Towards a more transparent research evaluation using citation Classification" and use classification techniques and natural language processing.

• **Data mining for new application domains**: Finally, the last category concerns contributions where data-mining techniques are applied to new, domain-specific application domains. The corresponding papers are: "Imputation of HLA genes from SNP data", "A Committee machine with different adaptive neuro fuzzy inference system for water saturation prediction", "Biomechanics: an unrealised opportunity for the data mining community - Limitations and possibilities of continuous data analysis in biomechanics", "Generic data mining system in semiconductor manufacturing", and "An adaptive multi level spatial clustering algorithm for medical databases". These papers cover a wide range of new applications domains, namely, bioinformatics, water-saturation systems, biomechanics, semiconductor manufacturing and medical databases. The set of data-mining and machine learning techniques used in these papers is also very diverse, including both supervised and unsupervised techniques. This last group of contributions illustrates, by its variety, that exciting and innovative applications of data mining await us in the coming years.

Our first acknowledgements go to the program committee members, whose reviewing efforts ensured the selection of a competitive and scientifically sound technical program. We also wish to thank the authors of all submitted papers for their innovation and creativity. Finally, we gratefully acknowledge the Artificial Intelligence Journal Division for providing the Forum with financial support.

15 October 2012
Brussels, Belgium
Boston, USA

Yann-Aël Le Borgne
Evimaria Terzi
**Co-Chairs of PhD Forum 2012**
Program Committee

Aris Anagnostopoulos, Sapienza University of Rome, Italy

Mauro Birattari, University of Brussels, Belgium

Gianluca Bontempi, Machine Learning Group, Universite Libre de Bruxelles, ULB, Belgium

Gemma Garriga, INRIA, France

Pierre Geurts, University of Liège, Belgium

Niina Haiminen, IBM Research, USA

Theodoros Lappas, Boston University, USA

Florent Masseglia, INRIA, France

Pauli Miettinen, MPI-INF, Germany

Spiros Papadimitriou, Google, United States

Panagiotis Papapetrou, Aalto University, Finland

Dino Pedreschi, University of Pisa, Italy

Marco Saerens, UCL, Belgium

Hanghang Tong, IBM T.J. Watson Research Center, USA

Panayiotis Tsaparas, University of Ioannina, Greece