

Report on the Fifth International Conference on Data Mining

Houston, Texas, November 27 – 30, 2005

ICDM 2005, the fifth conference in this series, has continued the tradition of providing an international forum for scientists, engineers, and practitioners to exchange and share their experience, new ideas, and latest research results on the design, analysis and implementation of data mining theory, systems and applications. The conference drew researchers and application developers from a wide range of data mining related areas such as statistics, machine learning, pattern recognition, databases and data warehousing, data visualization, knowledge-based systems and high performance computing.

This time around the conference committee received, a record high, 630 submissions from 1,702 authors and 52 countries all around the world. To accomplish the daunting task of reviewing these papers, we are fortunate to have an extremely strong technical program committee Co-chaired by Jiawei Han and Benjamin Wah and led by the following fourteen Vice-Chairs, who were in charge of different tracks of the conference:

Daniel Barbara, *George Mason University, USA*
Tamraparni Dasu, *AT&T Research Labs, USA*
Inderjit Dhillon, *University of Texas at Austin, USA*
Venkatesh Ganti, *Microsoft Research, USA*
Bart Goethals, *University of Antwerp, Belgium*
Dimitrios Gunopulos, *University of California, Riverside, USA*
Hillol Kargupta, *University of Maryland, Baltimore County & Agnik, LLC, USA*
George Karypis, *University of Minnesota, USA*
S. Muthu Muthukrishnan, *Rutgers University, USA*
Dino Pedreschi, *Univ. of Pisa, Italy*
Jian Pei, *State University of New York at Buffalo, USA*
Sunita Sarawagi, *Indian Institute of Technology, Bombay, India*
Arno Siebes, *Utrecht University, Netherlands*
Jeffrey Xu Yu, *Chinese University of Hong Kong, P. R. China*

Over a period of six weeks, the program committee, aided by over 140 highly qualified program committee members, worked tirelessly to provide in most cases three reviews for each paper. After significant deliberations of all papers with controversial reviews, the committee accepted 69 regular papers and 72 short papers in the following twenty tracks:

Program Track	Number Submitted	Regular Accepted	Short Accepted
Foundations	17	2	5
Data mining and machine learning algorithms	228	27	27
Mining semi-structured data, temporal, spatial, multimedia data	81	8	9
Mining data streams	22	4	2
Pattern recognition and trend analysis	9	1	2
Collaborative filtering/personalization	10	2	2
Data and knowledge representations	8	0	0
Query languages and user interfaces	1	0	0
Complexity, efficiency, and scalability issues	8	3	0
Data pre-processing/reduction, feature selection/transformation	34	4	3
Post-processing of data mining results	10	0	1
Statistics and probability in large-scale data mining	6	1	0
Soft computing and uncertainty management	29	1	2
Integration of data warehousing, OLAP, and data mining	10	0	0
Human-machine interaction and visual data mining	6	0	0
High performance and parallel/distributed data mining	11	1	2
Quality assessment and interestingness metrics	10	2	2
Security, privacy and social impact of data mining	23	3	4
Data mining applications	80	7	8
Others	27	3	3
Total	630	69	72

Those accepted papers were organized into 21 sessions and three tracks. A panel session, titled “Data Mining- Where to go?” and organized by Wen-Ran Zhang, was held on the afternoon of November 29, 2005. The panelists consisted of Jiawei Han, Vijay Raghavan, Bamshad Mobasher and Ramamohanarao Kotagiri. In addition there were 3 invited tutorials (the first 3, below), 2 contributed tutorials and 9 satellite workshops on November 27, 2005. Tutorials and Workshops committees were chaired, respectively, by Haesun Park and Pawan Lingras.

n **Tutorials Topics**

Models and Methods for Privacy-Preserving Data Mining and Data Publishing; Mining and Searching of Graph-Structured Databases; DMX, XML for Analysis and SQL Server Data Mining Platform; Clustering with Constraints; Bioinformatics and Bioimage Analysis.

n **Workshop Titles**

Multiagent Data Warehousing and Multiagent Data Mining; Optimization-based Data Mining Techniques with Applications; Data Mining Case Studies and Practice Prize; Temporal Data Mining: Algorithms, Theory and Applications; Knowledge Acquisition from Distributed, Autonomous, Semantically; Heterogeneous Data and Knowledge Sources; Computational Intelligence in Data Mining; Foundation of Semantic Oriented Data and Web Mining; Mining Complex Data; Privacy and Security Aspects of Data Mining.

Administrated by the Awards Chair, Katharina Morik, selecting the best paper of all submissions was carried out independently, and the best paper award was presented during the conference banquet.

It was indeed delightful to have four outstanding keynote speakers, Dr. John Elder of Elder Research Inc., Prof. Raj Reddy of Carnegie Mellon University, Prof. Sunita Sarawagi of Indian

Institute of Technology, and Dr. Arie Shoshani of Lawrence Berkeley Laboratory. Their thought-provoking speeches enlightened the participants with new developments and key research issues in the field.

ICDM 2005 introduced a couple of new events this past year: i) Qiang Yang and Xindong Wu organized a session entitled, '10 Challenging Problems in Data Mining.' The list of challenging problems were compiled by seeking inputs from a number of prominent researchers in the data mining field and is intended to provide guidance and direction to young researchers entering the field, and ii) One of the workshops, titled 'Data Mining Case Studies and Practice,' was organized to attract data mining practitioners from industry to share their data mining success stories. The workshop organizing committee also judged the various presentations and selected the two best presentations to receive first second place ICDM Data Mining Prizes, sponsored by Elder Research Inc.

The ICDM 2005 conference banquet was held on the evening of November 29 at the Houston Space Center. Following the introductory remarks by the steering committee chair, Xindong Wu, several awards were presented as follows:

1. The ICDM 2005 best paper, titled 'Mining Minimal Distinguishing Subsequence Patterns with Gap Constraints', was awarded to Xiaonan Ji, James Bailey and Guozhu Dong.
2. With unanimous support from the ICDM Steering Committee, Vijay Raghavan, the general chair of ICDM 2005, received the ICDM 2005 Outstanding Service Award.
3. Two ICDM 2005 Data Mining Practice Prizes, sponsored by the Elder Research Inc., were awarded: (a) 1st prize was awarded to Bharat Rao from Siemens Medical, and (b) 2nd prize was awarded to Dave DeBarr and Zach Elyer-Walker from MITRE Corporation.
4. Sponsored by IBM Research, five student travel awards were given to Arnab Bhattacharya, Huiping Cao, Varun Chandola, Xiaonan Ji and Raz Tamir.

The organizing committee wishes to note that ICDM 2005 was originally scheduled to be held in New Orleans, Louisiana. On August 29, Hurricane Katrina unleashed its fury, bringing death and destruction to New Orleans and very soon it became clear that Katrina had seriously crippled the infrastructure of New Orleans. With a strong resolve, determination and with the help of ICDM steering committee, IEEE Computer Society, the local arrangements committee members, including the faculty and 28 student volunteers from the University of Louisiana at Lafayette and the University of Houston, successfully re-organized and moved the conference in its entirety to Houston.

ICDM 2005 organizing committee proudly acknowledges corporate and other sponsors for their support: IBM Research, American Discount Ads, Inc., Elder Research Inc., the Center for Advanced Computer Studies at the University of Louisiana at Lafayette, University of Louisiana at Lafayette, the Department of Computer Science at the University of Houston, *Knowledge and Information Systems*, and Web Intelligence Consortium.

For additional information, please refer to the ICDM 2005 URL at: <http://www.cacs.louisiana.edu/~icdm05>