3rd IPDPS Workshop on High Performance Data Mining

Preface

The explosive growth in data collection in business and scientific fields has literally forced upon us the need to analyze and mine useful knowledge from it. Data mining refers to the entire process of extracting useful and novel patterns/models from large datasets. Due to the huge size of data and amount of computation involved in data mining, high-performance computing is an essential component for any successful large-scale data mining application.

This workshop provided a forum for presenting recent results in high performance computing for data mining including applications, algorithms, software, and systems. High-performance was broadly interpreted to include scalable sequential as well as parallel and distributed algorithms and systems. Relevant topics for the workshop included:

1. Scalable and/or parallel/distributed algorithms for various mining tasks like classification, clustering, sequences, associations, trend and deviation detection, etc.
2. Methods for pre/post-processing like feature extraction and selection, discretization, rule pruning, model scoring, etc.
3. Frameworks for KDD systems, and parallel or distributed mining.
4. Integration issues with databases and data-warehouses.

These proceedings contain 9 papers that were accepted for presentation at the workshop. Each paper was reviewed by two members of the program committee. In keeping with the spirit of the workshop some of these papers also represent work-in-progress. In all cases, however, the workshop program highlights avenues of active research in high performance data mining.

We would like to thank all the authors and attendees for contributing to the success of the workshop. Special thanks are due to the program committee and external reviewers for help in reviewing the submissions.

February 2000

Mohammed J. Zaki
Vipin Kumar
David B. Skillicorn
Editors

Workshop Co-Chairs

Mohammed J. Zaki (Rensselaer Polytechnic Institute, USA)
Vipin Kumar (University of Minnesota, USA)
David B. Skillicorn (Queens University, Canada)

Program Committee

Philip K. Chan (Florida Institute of Technology, USA)
Alok Choudhary (Northwestern University, USA)
Umeshwar Dayal (Hewlett-Packard Labs., USA)
Alex A. Freitas (Pontifical Catholic University of Parana, Brazil)
Ananth Grama (Purdue University, USA)
Robert Grossman (University of Illinois-Chicago, USA)
Yike Guo (Imperial College, UK)
Jiawei Han (Simon Fraser University, Canada)
Howard Ho (IBM Almaden Research Center, USA)
Chandrika Kamath (Lawrence Livermore National Labs., USA)
Masaru Kitsuregawa (University of Tokyo, Japan)
Sanjay Ranka (University of Florida, USA)
Vineet Singh (Hewlett-Packard Labs., USA)
Domenico Talia (ISI-CNR: Institute of Systems Analysis and Information Technology, Italy)
Kathryn Burn-Thornton (Durham University, UK)

External Reviewers

Eui-Hong (Sam) Han (University of Minnesota, USA)
Wen Jin (Simon Fraser University, Canada)
Harsha S. Nagesh (Northwestern University, USA)
Srinivasan Parthasarathy (University of Rochester, USA)