

IEEE Data Mining Forum 2008

Useful Activities to Promote Data Mining

Vipin Kumar
University of Minnesota

kumar@cs.umn.edu

www.cs.umn.edu/~kumar

SIAM International Conference on Data Mining

- Founded in 2001
 - Co-founders
J. Han, R. Grossman, V. Kumar
- Currently sponsored by American Statistical Association (ASA)
- Brings people from applied math & statistical community together with data mining
- Need to do more to focus on applications – not just on algorithms



SAM Journal: Aim and Scope

Announcing a new journal intended to encourage collaborations across disciplines, communicating novel data mining and statistical techniques to both novices and experts involved in the analysis of data from practical problems.

PREMIERING IN 2007!

STATISTICAL ANALYSIS **AND** DATA MINING

EDITORS-IN-CHIEF

Arnold Goodman
Center for Statistical Consulting, University of California, Irvine
agoodman@uci.edu

Chandrika Kamath
Center for Applied Scientific Computing, Lawrence Livermore National Laboratory
kamath2@llnl.gov

Vipin Kumar
Department of Computer Science and Engineering, University of Minnesota
kumar@cs.umn.edu



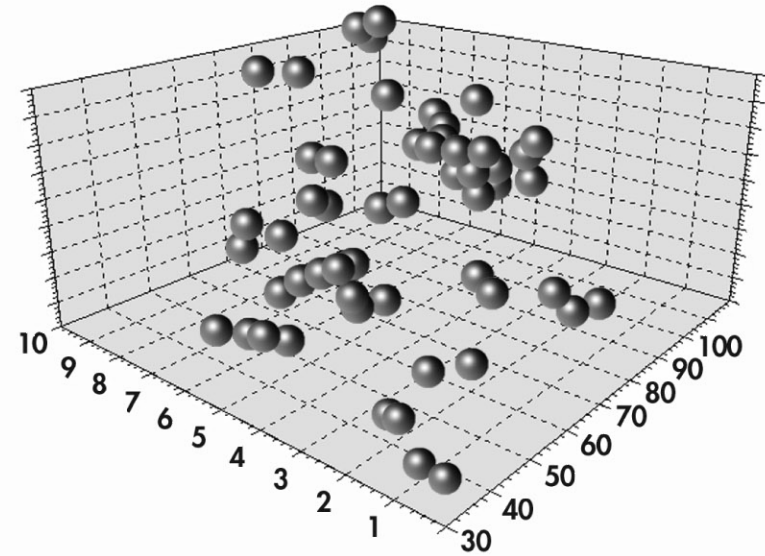
- Statistical Analysis and Data Mining addresses the broad area of data analysis, including data mining algorithms, statistical approaches, and practical applications.

- Forth coming special issues:
 - Business Analytics
 - Material Science
 - ChemInformatics

Data Mining and Knowledge Discovery Series Chapman & Hall/CRC

AIMS AND SCOPE:

- This series aims to capture new developments and applications in data mining and knowledge discovery, while summarizing the computational tools and techniques useful in data analysis.
- This series encourages the integration of mathematical, statistical, and computational methods and techniques through the publication of a broad range of textbooks, reference works, and handbooks.



- Series Editor: Vipin Kumar, University of Minnesota
- Series Acquisitions Editor: Randi Cohen, Chapman & Hall/CRC Press
- Series webpage: http://www-users.cs.umn.edu/~kumar/dmkd_series/

Data Mining and Knowledge Discovery Series Chapman & Hall/CRC (Cont.)

- List of current & forthcoming titles

PUBLISHED TITLES:

UNDERSTANDING COMPLEX DATASETS: DATA MINING WITH MATRIX DECOMPOSITIONS
David Skillicorn

COMPUTATIONAL METHODS OF FEATURE SELECTION
Huan Liu and Hiroshi Motoda

FORTHCOMING TITLES:

MULTIMEDIA DATA MINING: A Systematic Introduction to Concepts and Theory
Zhongfei Zhang and Ruofei Zhang

CONSTRAINED CLUSTERING: Advances in Algorithms, Theory, and Applications
Sugato Basu, Ian Davidson, and Kiri Wagstaff

TEXT MINING: Theory, Applications, and Visualization
Ashok Srivastava and Mehran Sahami

DATA MINING FOR DESIGN AND MARKETING
Yukio Ohsawa and Katsutoshi Yada

GEOGRAPHIC DATA MINING AND KNOWLEDGE DISCOVERY, Second Edition
Edited by
Harvey Miller and Jiawei Han

KNOWLEDGE DISCOVERY FOR COUNTERTERRORISM AND LAW ENFORCEMENT
David Skillicorn

DATA ANALYTICS: A Problem-Solving Approach
Sanjay Chawla

RELATIONAL DATA CLUSTERING: Models, Algorithms, and Applications
Bo Long, Zhongfei Zhang, and Philip Yu

SOCIAL COMPUTING: A DATA MINING PERSPECTIVE
Huan Liu, Jianping Zhang, and Arunabha Sen

NEXT GENERATION OF DATA MINING
Edited by Hillol Kargupta, Jiawei Han, Philip Yu, Rajeev Motwani, and Vipin Kumar

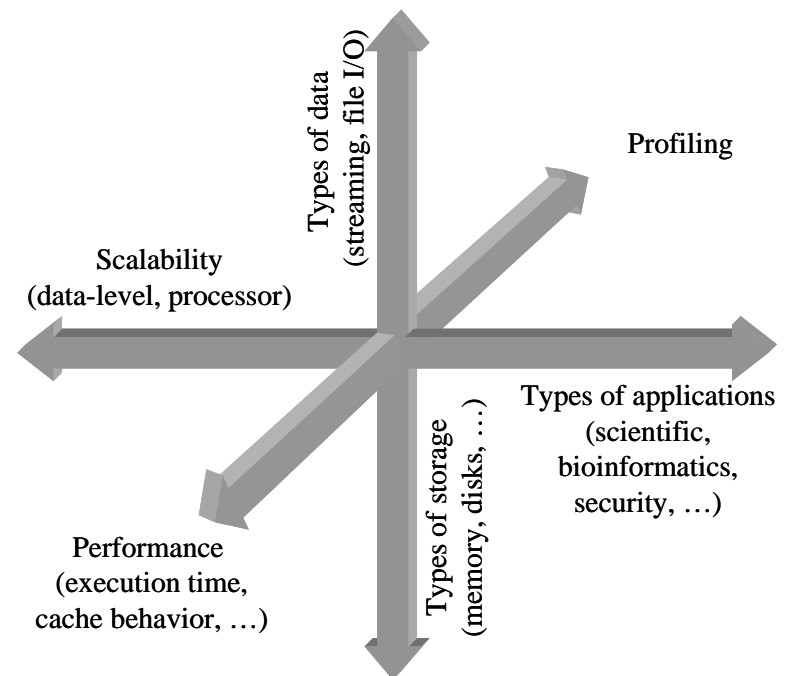
Scalable Benchmarks, Software and Data for Data Mining, Analytics and Scientific Discoveries

Projects funded by NSF

PIs: A. Choudhary and Gokhan Memik (NW) , V. Kumar and M. Steinbach (UM)

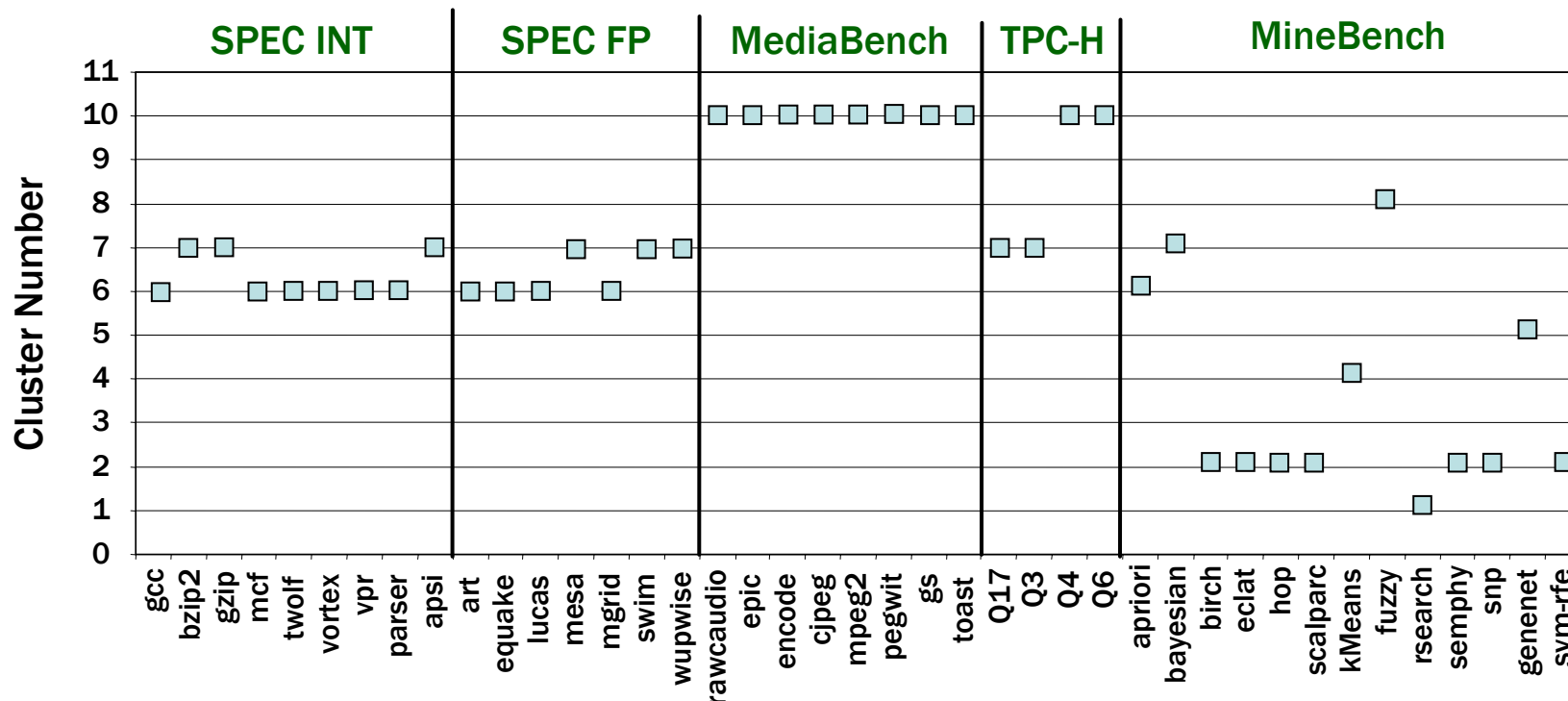
Goal: Establish a comprehensive benchmarking suite for data mining applications.

- Motivate the development of new processor architectures and system design for data mining
- Motivate the implementation of more sophisticated data mining algorithms that can work with the constraints imposed by current architecture designs
- Improvement the productivity of scientists and engineers using data mining application in a wide variety of domains



Do We Need Benchmarks Specific to Data Mining?

- Performance metrics of several benchmarks gathered from Vtune
 - Cache miss ratios, Bus usage, Page faults etc.
- Benchmark applications were grouped using Kohonen clustering to spot trends:



“Advances in the microprocessor industry would not have been possible without the SPEC benchmarks” - David Patterson)

Reference: [Pisharath J., Zambreno J., Ozisikyilmaz B., Choudhary A., 2006]



Thank you!