Introduction : knowledge quality measures in a data-mining process

Vipin Kumar*, **

* Director, Army High Performance Computing Research Center
** Professor, Department of Computer Science & Eng, University of Minnesota
kumar@cs.umn.edu
http://www.cs.umn.edu/~kumar

The author. Vipin Kumar is currently the Director of Army High Performance Computing Research Center and Professor of Computer Science and Engineering at the University of Minnesota. His research interests include High Performance computing, data mining, and their applications to information assurance. He has authored over 200 research articles, and co-edited or co-authored 9 books including the widely used text book ``Introduction to Parallel Computing", and an upcoming edited collection, "Managing Cyber Threats: Issues, Approaches and Challenges" to be published by Kluwer Academic Publishers.
Kumar has served as chair/co-chair for many conferences/workshops in the area of data mining and parallel computing, including the IEEE International Conference on Data Mining (2002) and the 15th International Parallel and Distributed Processing Symposium (2001). Kumar serves as the chair of the steering committee of the SIAM International Conference on Data Mining, and serves on the editorial boards of Knowledge and Information Systems, IEEE Computational Intelligence Bulletin, Annual Review of Intelligent Informatics, Parallel Computing, the Journal of Parallel and Distributed Computing, and has served on the editorial boards of IEEE Transactions of Data and Knowledge Engineering (93-97), IEEE Concurrency (1997-2000), and IEEE Parallel and Distributed Technology (1995-1997). He is a Fellow of IEEE, a member of SIAM, and ACM, and a Fellow of the Minnesota Supercomputer Institute.

Advances in information technology and data collection methods have led to the availability of large data sets in commercial enterprises and in a wide variety of scientific and engineering disciplines. Examples of large data sets are genomic data, climate data, and market basket data collected at commercial outlets. There is an unprecedented opportunity to analyze such data and extract intelligent and useful information. The potential for return to the society from this analysis is huge. For example, the understanding of genome expression data can lead to the development of better and cheaper drugs that have fewer side effects. The information collected from the analysis of market-basket data can improve the profitability of a corporation. Indeed, researchers in the fast growing discipline of knowledge