2004 SIAM International Conference on Data Mining

Industry/government session: Thursday April 22, 2004, 3:00pm – 4:45pm

Session chair: Chandrika Kamath, Lawrence Livermore Laboratory

3:00pm: **Joseph Coughlan**, Intelligent Systems Project & NASA Ames Research Center

Today many different types of public data including NASA data are available on the Internet. At the same time one of the most common problems for scientific and engineering studies is having "access to data" for use in applications and analysis systems. This talk will provide background and then describe specific examples of knowledge discovery in the California wine industry, and in tropical ecosystems. These examples illustrate the ongoing need to develop more efficient discovery capabilities. To that end, NASA has been planning a project to support research and technology demonstration applications that enable more efficient and broader use of data generated by NASA science and exploration missions. An overview of that plan will be presented and commentary is welcome.

3:30pm: "Event Mining", **Sheng Ma**, IBM Research

Today, many different types of data have been collected related to the operations and performance of networks, systems, servers and applications. As the amount of data grows rapidly every day, autonomic computing requires efficient, effective tools to extract useful information out of data effectively and automatically in both real-time and off-line modes. In this talk, I will first discuss how machine learning and data mining approaches may help to achieve self-healing as described by autonomic computing, and the open research issues and related applications. Further, I will focus on our recent research on pattern discovery and prediction. I will present the developed algorithms based on machine learning and data mining principles and their results.

4:00pm: "Some lessons from data mining projects", **Frank Meyer**, France Telecom

In this talk, I will present, in a general way, some lessons from data mining projects carried out in a large company. Some projects were difficult and complex to complete, some were more usual. All of them led to some interesting lessons about:

- some currently critical steps in data mining processes,
- some currently possible differences between machine learning and industrial data mining approaches

4:30 - 4:45pm: Break

4:45pm: “Applications of Data Mining and InfoViz at Sandia National Labs”, **Kevin Boyack**, Sandia National Laboratories

Data mining and information visualization are being applied in many different fields of interest at Sandia, including research assessment, business intelligence, gap analysis, bioinformatics, and network analysis. The tools and techniques used in the different areas have much in common. Examples will be given.

5:15pm: “Data Mining in Aviation Safety and Security”, **Ashok Srivastava**, NASA Ames Research Center

Decision makers need fast, reliable, and interpretable analysis of multivariate, heterogeneous, and distributed data in order to constantly and consistently assess the performance of the aviation system to identify life threatening conditions and events. Data mining can play an important role in the discovery of potential problems in the aviation system by automatically analyzing text-based safety and security reports, flight plans, telemetry, and weather data.

This talk gives an overview of the Aviation Safety and Security program at NASA and describes areas where data mining plays a critical role. An example of a system known as the "Morning Report" will be given, which is a successful real-world application of data mining technology in this industry.