Improve Data Mining and Knowledge Discovery through the use of MatLab

Gholam Ali Shaykhian
Information Technology (IT)
Directorate
Technical Integration Office (IT-G)
ali.shaykhian@nasa.gov

Dawn (Elliott) Martin
Safety and Mission Assurance
Directorate
Technical Management Branch (SA-G1)
dawn.elliott@nasa.gov

Robert Beil
NASA Engineering & Safety Centers
(NESC)
Systems Engineering Office(C1-20)
robert.j.beil@nasa.gov

National Aeronautics and Space Administration (NASA)
Kennedy Space Center, FL 32899, United States

Abstract

Data mining is widely used to mine business, engineering, and scientific data. Data mining uses pattern based queries, searches, or other analyses of one or more electronic databases/datasets in order to discover or locate a predictive pattern or anomaly indicative of system failure, criminal or terrorist activity, etc. There are various algorithms, techniques and methods used to mine data; including neural networks, genetic algorithms, decision trees, nearest neighbor method, rule induction association analysis, slice and dice, segmentation, and clustering. These algorithms, techniques and methods used to detect patterns in a dataset, have been used in the development of numerous open source and commercially available products and technology for data mining.

Data mining is best realized when latent information in a large quantity of data stored is discovered. No one technique solves all data mining problems; challenges are to select algorithms or methods appropriate to strengthen data/text mining and trending within given datasets. In recent years, throughout industry, academia and government agencies, thousands of data systems have been designed and tailored to serve specific engineering and business needs. Many of these systems use databases with relational algebra and structured query language to categorize and retrieve data. In these systems, data analyses are limited and require prior explicit knowledge of metadata and database relations; lacking exploratory data mining and discoveries of latent information.

This presentation introduces MatLab® (MATrix LABoratory), an engineering and scientific data analyses tool to perform data mining. MatLab was originally intended to perform purely numerical calculations (a glorified calculator). Now, in addition to having hundreds of mathematical functions, it is a programming language with hundreds built in standard functions and numerous available toolboxes. MatLab's ease of data processing, visualization and its enormous availability of built in functionalities and toolboxes make it suitable to perform numerical computations and simulations as well as a data mining tool. Engineers and scientists can take advantage of the readily available functions/toolboxes to gain wider insight in their perspective data mining experiments.