Editorial

Ali K. Kamrani*

Department of Industrial Engineering,
University of Houston,
E206 Engineering Bldg 2,
Houston, TX 77204-4008, USA
Fax: +1-713-743-4190
E-mail: akamrani@uh.edu
*Corresponding author

Hamid R. Parsaei

Department of Industrial Engineering,
University of Houston,
E206 Engineering Bldg 2,
Houston, TX 77204-4008, USA
Fax: +1-713-743-4190
E-mail: parsaei@uh.edu

Mahammad Aloudat

Baker Hughes,
Baker Oil Tools, Navigation facility
6023 Navigation Blvd,
Houston, TX 77011, USA
Fax: +1-713-923-9654
E-mail: mahammad.aloudat@bakerhughes.com

Biographical notes: Ali K. Kamrani received his PhD in Industrial Engineering from the University of Louisville in 1991. Currently, he is an Associate Professor and PhD Program Coordinator at the University of Houston, Department of Industrial Engineering, Houston, TX, USA. His current research interests include cellular manufacturing, systems engineering, collaborative engineering, product design life cycle and rapid prototyping.

Hamid R. Parsaei is a Professor and Chairman of the Department of Industrial Engineering at the University of Houston. He received his MS and PhD in Industrial Engineering from Western Michigan University and the University of Texas at Arlington, respectively. His research interests are in design of manufacturing systems and economic and financial decision analysis. He has published 250 papers in peer reviewed journals and conference proceedings. He is a Registered Professional Engineer in Texas and a Fellow of the Institute of Industrial Engineers.

Mahammad Aloudat received his master degree from Sam Houston State University in Huntsville, Texas (1999); a Doctoral degree in Industrial Engineering from the University of Houston (2006). He has been working with Baker Oil Tools an oil company located in Houston since 2001. Currently, he is a Manufacturing Manager at the Baker Oil Tools company. He is a Lecturer at the Technology Department at University of Houston. His areas of research are data mining, quality improvement, lean manufacturing and cellular manufacturing.

Data mining has been referred to as knowledge management or knowledge engineering. Until recently, it has been an obscure technology, discussed mainly by theoreticians in the artificial intelligence fields. Data mining is the process of applying computational techniques to identify patterns in order to produce models. It has been defined as a process of searching through details of data in order to identify patterns or trends. Many times, large databases are searched for relationships, trends and patterns, which prior to the search, are neither known to exist nor visible. These relationships or trends are usually assumed to be there by engineers and marketers, but need to be proven by the data itself. Data mining process will identify these patterns. The new knowledge generated from the data mining process allows the user community to be better at what it does. Often, a problem that arises is that large databases are searched for very few facts that will give the desired information. Moreover, the algorithm and search criteria used in a single database may change when a new trend or pattern is to be studied. Each database may need a different search criterion as well as new algorithms that can adapt to the conditions and problems of the new data. It is a known fact that the rate of growth of data sets completely exceeds the rates that traditional 'manual' analysis techniques that can cope with.

This means that if a company uses a regular technique for extracting knowledge from a database, vast amounts of data will be left unsearched, as the data growth surpasses the traditional mining procedures. These factors call for a need of a technology that will enable humans to tackle a problem using large amounts of data without disregarding or losing valuable information that may help solve any kind of problem involving large data sets. During the 1980s, many organisations built infrastructural databases, containing data about their products, clients and competitors. These databases were a potential gold mine, containing terabytes of data with much 'hidden' information that was difficult to understand. With the great strides shown by artificial intelligence researchers, machine-learning techniques have grown rapidly. Neural networks, genetic algorithms and other applicable learning techniques are making the extraction of knowledge form large databases easier and more productive than ever. Data mining is being used widely.

Large organisations are utilising data discovery to analyse their client files. In the UK, the BBC has applied data mining techniques to analyse viewing figures. However, it has been seen that the use of data discovery brings a lot of problems. As much as 80% of this process is preparing data, and the remaining 20% is about mining. It is very difficult to introduce data mining into a whole organisation.

A lot of data mining projects are disregarded as options because of these additional problems:

- Lack of clear vision
- Quality and integrity of the data
- Privacy and legal restrictions
- Technical reasons
- Analysis and interpretation problems

In addition to these problems, the companies also need to have a minimum level of expertise on the data mining processes. Scenarios, in which the area expert does not have a specific question and asks the analyst to come up with some interesting results, are sentenced to fail. The same holds true for situations where the expert provides the data analyst with a set of data and a question, expecting the analyst to return the exact answer to that question. Knowing what to ask and what to expect from the information in a database is not enough. Knowledge of the available data from within is also required. This will enable the expert and the data analyst to know where the data is and have it available depending on the problem being studied. Data mining expertise is required in order to select the appropriate algorithm for the data mining problem and the questions being raised. Data mining can bring significant gains to organisations, for example through better-targeted marketing and enhanced internal performance. The long-term goal of data mining is to create a self-learning organisation that makes optimal use of the information it generates.

The goal of this Special Issue is to cover a variety of topics and issues related to data mining with specific application to knowledge discovery. This Special Issue consists of seven selected articles. *Aloudat, Kamrani and Nasr* will examine cell quality performance improvement through the integration of data mining, Artificial Neural Network (ANN) techniques and cellular manufacturing. The aim is to study and predict the factors that impact quality product in cellular manufacturing, such as material complexity, operation type, machinist and quantity in order to improve cell performance. *Liu, Loh, Youcef-Toumi and Tor* describe their recent research on a Knowledge Management and Retrieval system based on Hierarchical Text Classification scheme. The proposed system organises the large volume of manufacturing related electronic documents according to the manufacturing knowledge taxonomy, then classifies and further routes the searching queries based on manufacturing concepts to the corresponding categories for documents retrieval. Automatic text summarization has long been studied. The growth in the information size on the web results in more demands for automatic methods for text summarization. *Qazvinian, Hassanabadi and Halavati* have developed a Genetic Algorithm-based sentence selection system to make summary. The fitness function is based on readability, cohesion and topic-relation factors. *Youssef and Rebai* proposed a new approach to solve the classification problems via Fuzzy Linear Programming Models (FLPM) to address the major disadvantage of the parametric procedures and their requirement of certain assumptions such as normality, equality of variance-covariances matrix and the absence of outliers. In their paper Songram and Boonjing propose a new method, called closed multidimensional sequential pattern mining, for mining multidimensional sequential patterns. The new method is an

integration of closed sequential pattern mining and closed item set pattern mining. *Pintér, Madeira, Vieira, Majzik and Pataricza* presented another application of data intelligence from the point of view of computer dependability experts. In the last paper *Kamrani and Gonzalez* propose a new methodology for solving complex combinatorial problem using genetic algorithms. A sample TSP is solved using the developed GA model.

The editors would like to thank the reviewers of this Special Issue. Without their assistance this project would not have been possible. We would also like to thank Dr. Dorgham and Mr. Jim Corlett from Inderscience Publishers for their support and assistance throughout the development of this special issue. Finally, we would like to thank our contributors by allowing us to share the results of their research with the rest of the engineering community.