Editorial

Hisham Al-Mubaid*

Department of Computer Science,
University of Houston – Clear Lake,
2700 Bay Area Blvd.,
Houston, TX, 77062, USA
E-mail: hisham@uhcl.edu
*Corresponding author

Emad S. Abouel Nasr

Department of Mechanical Engineering,
Helwan University,
Helwan, Cairo, Egypt
E-mail: emadsamir60@helwan.edu.eg

Biographical notes: Hisham Al-Mubaid received his PhD in Computer Science from the University of Texas at Dallas in 2000. He is currently an Associate Professor of Computer Science at the University of Houston-Clear Lake, and Chair of the Computer Information Systems program. His research interests include data mining, bioinformatics, machine learning and natural language processing.

Emad S. Abouel Nasr is an Assistant Professor in the Mechanical Engineering Department at Helwan University, Faculty of Engineering, Helwan, Cairo, Egypt. He received his PhD in Industrial Engineering from the University of Houston, TX, USA in 2005. His current research focuses on CAD, CAM, rapid prototyping, advanced manufacturing systems and collaborative engineering.

Manufacturing firms and industrial companies have to apply the latest technologies and utilise the most recent advances in manufacturing and industrial engineering to be able to compete and nourish in the nowadays competitive world markets.

Among these advances is how to utilise the volumes of manufacturing data generated during course of manufacturing including: design, production, marketing, sales, warranty, and more. Data mining and knowledge management in the industrial and manufacturing domains, the topic of this special issue, is interested in addressing the issue of how to benefit from manufacturing data to improve the manufacturing productivity and to enhance the product.

Data mining has been proved to be a success and important step in manufacturing and industrial engineering. Modern organisations produce and generate large volumes of data over time. This data include knowledge, patterns, and information that could lead to increased efficiencies, cost reduction and reliable operations. Knowledge generated from data mining process is an important factor in the success of these industries.

The aim of this special issue is to identify the research issues related to data mining and its application in engineering and business environment in light of global challenges.

The title of this issue 'Data mining and knowledge management' is to support, promote and publish high quality research results, advances and developments in the areas related to data mining within manufacturing and industrial engineering. This special issue will cover a variety of topics and issues related to data mining with specific application to knowledge extraction and management. The topics covered in this issue include: data mining concepts and applications, data mining for manufacturing system, fault detection and diagnostics, reliability and maintenance, process and quality control, CAD, CAM and engineering design, material requirement planning (MRP), decision support systems, data mining and knowledge acquisition, genetic algorithms (GAs), knowledge discovery from industrial data, supervised and unsupervised machine learning, artificial neural networks (ANN).

The first article, authored by Zanaty, Aljahdali and Cripps, addresses the data mining and knowledge management in the general sense and proposes a method of combing two kernel functions, polynomial and rbf, for support vector machine. The paper proves the effectiveness of the proposed approach through examples and experiments with data classification. The new kernel outperforms both poly and rbf.

Dehkordi, Badie and Zadeh presents, in the second article, a new method for hiding sensitive data from the original dataset so that when we do ARM, the sensitive transactions will not appear. The paper includes two data mining techniques: association rule mining and GA. Their approach is important for application that requires privacy preserving with the existence of huge amounts of data.

In the third article, Li, Nsofor, and Song present a comparative study of predictive data mining techniques. The paper studies five regression-based methods for data modelling and prediction, and provides a detailed comparison of the characteristics of these methods. Comparison experiments are conducted on four datasets of different properties, and detailed results are reported and discussed. The techniques studied in this paper include multiple linear regression (MLR), principal component regression (PCR), ridge regression (RR), partial least squares (PLS), non-linear partial least squares (NLPLS).

The fourth article by Geetha and Arock introduces a clustering algorithm based on particle swarm optimisation (PSO) and K-means for gene expression data. The proposed ideas represent an interesting way in data mining to combine these two techniques, i.e., PSO and d K-means.

Prakash Kumar and Tripathy present, in the fifth article, a modified algorithm called MMeR. This algorithm handles both numerical and categorical data simultaneously besides dealing with uncertainty. Moreover, this new algorithm provides better performance than most of the existing algorithms in the literature. Some well-known datasets are taken to test and illustrate the superiority of MMeR over most of the existing algorithms.

In the sixth article, Singh and Rupinder Singh introduce a methodology to study the feasibility of decreasing the shell wall thickness of mould cavity for economical and statistically controlled rapid casting solution of low brass alloy using three dimension printing technology. They begin their study from the identification of component benchmark, and then technological prototypes were produced at different shell wall thickness of mould cavity. They used a coordinate measuring machine to calculate the dimensional tolerances of the castings produced. Also, some important mechanical properties were compared to verify the suitability of the castings. The study suggested that for the shell thickness, having value less than the recommended one is more suitable

from dimensional accuracy and economic perspective. Moreover, production cost and production time has been improved by this methodology.

Krishnakumari and Vivekanandan develop, in the seventh article, an efficient scalable clustering algorithm designed for high-dimensional data which combines the ideas of linear discriminant analysis (LDA) based on PCA feature extraction along with K-means algorithm to select the most discriminative subspace. K-means clustering is used in the beginning to generate class labels and then LDA is used for subspace selection towards highest variance. The algorithm is designed to reduce the sum squared errors as much as possible for the partitions, while at the same time keep the partitions far apart as possible. Moreover, the clustering process is integrated with the subspace selection process based on LDA and the data are then simultaneously clustered while the feature subspaces are selected. Real datasets are used to show that the proposed method outperforms existing methods for clustering high-dimensional genomic data in terms of accuracy.

In the eighth article, Abouel Nasr, Al-Mubaid and Hussein present a data mining methodology for modelling material properties with machine learning. Material properties play a key role in manufacturing process throughout the product life cycle. The methodology signifies new ways to utilise data mining and machine learning in analysing and examining material properties with an experimental study. They propose and apply an effective feature reduction technique with data clustering and classification to extract the most important properties of materials which can benefit various manufacturing industries. Their evaluation results proved that the material databases and material properties can be easily and feasibly analysed and examined in data mining and the outcomes can be fairly useful for the manufacturing industries and other industrial engineering applications.