

## Special volume on *Data Mining*

Victoria C.P. Chen

Published online: 11 September 2009  
© Springer Science+Business Media, LLC 2009

Data mining was first developed in the statistics and computer science communities. Its popularity and success rapidly grew, catching the attention of both academia and industry, including those in operations research. To address operations research related contributions to data mining, the Institute for Operations Research and Management Sciences (INFORMS) approved the founding of the Section on Data Mining in 2004. This is a highly interdisciplinary section, inviting participation from researchers and practitioners across the fields of statistics, computing, operations research, operations management, computer science, and information technology. To provide a forum bringing these groups together, the first *INFORMS Workshop on Artificial Intelligence and Data Mining*, a collaboration between the INFORMS Data Mining and Artificial Intelligence subdivisions, was held at the 2006 INFORMS Annual conference. This special volume of the *Annals of Operations Research* evolved out of this Workshop, and 10 of the 14 papers contain research originally presented at the Workshop.

The papers in this special volume offer a perspective into the potential for operations research methods to impact the field of data mining and vice-versa. New methods for both supervised and unsupervised learning are presented, with a variety of applications. Three papers are strongly linked to complex biomedical problems, three papers are motivated by business applications, and two papers study issues related to manufacturing.

Five papers address unsupervised learning problems, two for pre-processing complex data structures, two applied to business applications, and one applicable to multivariate statistical process control. Wang and Zhu develop a principal component analysis based method for cleaning noise from image data. Kim, Wang, and Hiremath present a Bayesian approach for aligning spectra from high-resolution nuclear magnetic resonance. Kumar and Patel extend Fisher's classical clustering method and then employ their method to forecast retail sales. Au, Duan, Hesar, and Jiang propose a LASSO-based framework to highlight irregularities in time series data and apply it to activity monitoring. Finally, Hu and Rungger convert

---

V.C.P. Chen (✉)  
University of Texas at Arlington, Arlington, USA  
e-mail: [vchen@uta.edu](mailto:vchen@uta.edu)

the problem of detecting changes in multivariate time series data into a supervised learning problem.

Two papers address issues in support vector methods. Wang, Jiang, and Tsui propose a smooth version of a hinge loss function for support vector machines. Wang and Zhu, in a second paper, develop a kernel learning method for support vector regression and apply it to financial market forecasting.

Four papers focus on the optimization engine to improve classification. For linear discrimination, Plastria, De Bruyne, and Carrizosa create a hybrid variable neighborhood search scheme, and Boylu, Aytug, and Koehler present a genetic algorithms approach and compare it to a mixed integer programming approach. Brooks and Lee utilize mixed integer programming for the challenging problem of multi-category classification. Fan and Chaovallitwongse develop a hybrid approach for binary classification in medical decision-making using nearest neighbor and integer programming.

In addition to that of Fan and Chaovallitwongse, hybrid methods were developed in three other papers. Agarwal, Colak, and Deane integrated neural networks with genetic algorithms to improve the search for good solutions in integer programming. Moon, Simpson, and Ku-mara combined product ontology, fuzzy c-means clustering, and association rule mining to discover product families for more efficiently structured manufacturing processes. Finally, Gray, Williams, Brown, Faulon, and Sale employed a fusion of support vector machines, neural networks, hidden Markov models, and a scoring method to conduct classification for the complex problem of protein phosphorylation prediction. Because of the nature of data mining, all of the methods in this special volume have hybrid characteristics that integrate ideas from different fields, and data mining will continue to evolve in this manner.

Much appreciation goes to the late Peter Hammer and to Sigurdur Ólafsson for establishing this special volume. Many thanks to the referees for their careful evaluations of the submitted papers. Finally, thanks to Endre Boros for maintaining the standards of the journal, and my gratitude to Katie D'Agosta for facilitating administrative matters.

August 2009