Top 10 Data Mining Techniques in Business Applications: A Brief Survey

Abstract

Purpose: Data mining is widely considered necessary in many business applications for effective decision making. The importance of business data mining is reflected by the existence of numerous surveys in the literatures focusing on the investigation of related works using data mining techniques for solving specific business problems. However, there has been no recent study answering the following question: What are the widely used data mining techniques in business applications?

Design/methodology/approach: The aim of this paper is to examine related surveys in the literature and thus to identify the frequently applied data mining techniques. To ensure the recent relevance and quality of the conclusions, the criterion for selecting related studies are that the works be published in reputed journals within the past 10 years.

Findings: There are 33 different data mining techniques employed in eight different application areas. Most of them are supervised learning techniques and the application area where such techniques are most often been is bankruptcy prediction, followed by the areas of customer relationship management, fraud detection, intrusion detection, and recommender systems. Furthermore, the widely used 10 data mining techniques for business applications are the decision tree (including C4.5 and CART), genetic algorithm, k-nearest neighbor, multilayer perceptron neural network, naïve Bayes, and support vector machine as the supervised learning techniques and association rule, expectation maximization, and k-means as the unsupervised learning techniques.

Originality/value: The originality of this paper is to survey the recent ten years of related survey and review articles about data mining in business applications in order to identify the most popular techniques.

Article classification: General review

Keywords: data mining; business applications; machine learning; survey

1. Introduction

Data are key assets to any business enterprise, which means that techniques for data mining and knowledge discovery in databases (KDD) play an important role in many business applications and assist with the making of effective decisions. The aim is to identify useful and/or interesting knowledge from a given dataset, which can then be treated as a decision support system. In general, the process is comprised of four steps: data collection, data cleaning and preprocessing, data analysis, and interpretation and evaluation (Han and Kamber, 2001; Ye, 2003).

Of these steps, data analysis is essential to the performance of data mining tasks, based on employing relevant statistical and/or machine learning techniques. In early surveys of Wong et al. (1997) and Vellido et al. (1999), artificial neural networks were recognized as one of the most popularly used data mining algorithms for a diverse range of business applications in the 80s and 90s. Bose and Mahapatra (2001) reviewed a number of machine learning techniques in the context of mining business data. They found rule induction (i.e., decision trees) to be the most commonly used technique, followed by artificial neural networks and case-based reasoning. In addition, most applications were found in areas related to finance, where prediction of future eventualities is the dominant task category for business data mining.

Recently, there have been several studies surveying the data mining techniques used in specific business applications. These surveys also show that the artificial neural network is one of the widely employed techniques for various business problems, such as fraud detection (Ngai et al., 2011), bankruptcy prediction and credit scoring (Lin et al., 2012), recommender systems (Bobadilla et al., 2013), and software development effort estimation (Wen et al.,

2012).

Wu et al. (2008) has identified the top (or frequently used) 10 data mining algorithms, which covered classification (including C4.5, SVM, AdaBoost, k-NN, naïve Bayes, and CART), clustering (including k-means and EM), statistical learning (i.e., EM and Naïve Bayes), association analysis (i.e., Apriori), and link mining (i.e., PageRank). In particular, they focus on describing these algorithms, discussing the impact of these algorithms, and reviewing current and future research on these algorithms. However, the status of current data mining algorithms for business applications is not analyzed.

With the exception of the work of Bose and Mahapatra (2001) the purpose of these surveys has not been to answer the question of what are the top data mining techniques for popular business applications. More specifically, they only describe what kinds of techniques were used for which business domain problems. In addition, their survey of data mining techniques and business application areas is not new, where the works were published before 1999. Therefore, the objective of this study is to examine the recent relevant survey articles in order to identify the top (or frequently used) data mining algorithms for different business problems. Specifically, the surveys related to eight well-known business applications are analyzed, which are bankruptcy prediction, customer relationship management, fraud detection, intrusion detection, recommender systems, software development effort estimation, stock prediction/investment, and other financial time-series areas. Note that the methodology of surveying related survey/review articles is not new. For example, Shaughnessy and Slawson (2003) surveyed thirty five review articles on the treatment of type 2 diabetes.

The rest of this paper is organized as follows. Section 2 overviews the related survey literature about business data mining. Section 3 describes the research methodology and the results of the analysis are presented in Section 4. Finally, some conclusions are offered in Section 5.

2. Literature Review

2.1 Neural Networks for Business Applications

According to Wong et al. (1997), who conducted one of the earliest surveys for business data mining during the period between 1988 and 1995, neural networks (or artificial neural networks) were widely developed for a variety of business applications. They found the multilayer perceptron (MLP) neural network, based on the backpropagation learning algorithm, also called as backpropagation neural network (BPNN), to be the most popular network model for prediction or classification problems. On the other hand, the self-organizing map (SOM) was the most widely used neural network technique for the clustering problems. They identified seven application categories using neural networks: accounting/auditing, finance, human resources, information systems, marketing/distribution, production/operations, and others. Among them, production/operations and finance are the top two business applications.

Later, Vellido et al. (1999) conducted another survey for the period between 1992 and 1998. They presented six business application areas using neural networks, including accounting/auditing, finance, management, marketing, production, and others. Similar to Wong et al. (1997), MLP and SOM were the most commonly used classification and clustering neural network techniques, respectively. In addition, they found that most studies focused on using neural networks for finance problems, such as bankruptcy prediction and credit scoring.

In another study, Smith and Gupta (2000) overviewed different types of neural network models for a variety of business problems. They showed that neural networks were actually applied across a wide range of areas from business and commerce to industry. They reviewed seven kinds of business problems, namely marketing, retail, banking and finance, insurance, telecommunications, operations management, and other industries.

2.2 Data Mining in Specific Business Problems

Besides the surveys focusing on the use of neural networks for various business applications, recent surveys have discussed popular data mining techniques for specific business problems. For instance, Ngai et al. (2011) reviewed the literature on the data mining techniques used for fraud detection. In particular, four categories of financial fraud (bank fraud, insurance fraud, securities and commodities fraud, and other related financial fraud) and six classes of data mining techniques (classification, regression, clustering, prediciton, outlier detection, and visualization) were analyzed.

In Lin et al. (2012), data mining techniques for bankruptcy prediction and credit scoring were reviewed. Related techniques were classified into single, ensemble, and hybrid ones with ensemble and hybrid approaches being based on combining several single techniques in a parallel or linear manner.

Ngai et al. (2009) conducted a review of the related literature for the application of data mining techniques to customer relationship management (CRM). In their review, four CRM dimensions (customer identification, customer attraction, customer retention and customer development) and seven data mining techniques (association, classification, clustering, forecasting, regression, sequence discovery and visualization) were analyzed.

In Wen et al. (2012), machine learning techniques for software development effort estimation from four aspects were analyzed, which are type of learning techniques, estimation accuracy, model comparison, and estimation context.

Bobadilla et al. (2013) overviewed related techniques for the application of recommender systems for web applications. In their work, five machine learning related techniques were identified. Catania and Garino (2012) surveyed related works in the area of intrusion detection, where misuse and anomaly based approaches by machine learning techniques were analyzed.

3. Methodology

The first step to search related surveys in the literature is to identify some popular business application areas. In this paper, they are partially based on Bose and Mahapatra (2001), Smith and Gupta (2000), Vellido et al. (1999), and Wong et al. (1997). Business application areas that have been covered in at least two out of the four works mentioned above are accounting and auditing, banking and finance, marketing, production and operations, and telecommunications.

Since the above identified business application areas may not completely cover today's business problems, an additional criterion used for searching related survey articles is that the works have been published in reputed indexed journals within the past 10 years. By performing these two steps of collecting survey/review articles, we believe that most important business application areas can be identified. In total, 15 survey articles were analyzed, which are Ahmed et al. (2010), Atsalakis and Valavanis (2009), Bobadilla et al. (2013), Catania and Garino (2012), Kolias et al. (2011), Kumar and Ravi (2007), Lin et al. (2012), Ngai et al. (2011), Ngai et al. (2009), Park et al. (2012), Ponsich et al. (2013), Tsai et al. (2011), Verikas et al. (2010), Wen et al. (2012), and Wu et al. (2008).

Finally, the eight specific business areas that matched the publication criteria are shown in Table 1.

Table 1 The eight business application areas

Areas	No. of articles	Publications
Bankruptcy prediction	3	● European Journal of Operational Research ● IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews ● Soft Computing
Customer relationship management	1	● Expert Systems with Applications
Fraud detection	1	● Decision Support Systems

Intrusion detection	3	● Computers & Electrical Engineering ● Computers & Security ● Expert Systems with Applications
Recommender systems	2	● Expert Systems with Applications ● Knowledge-Based Systems
Software development effort estimation	1	● Information and Software Technology
Stock prediction/investment	2	● IEEE Transactions on Evolutionary Computation ● Expert Systems with Applications
Other financial time-series areas	1	● Econometric Reviews

In addition to the business application areas covered in the early surveys, we can see that intrusion detection and recommender systems are two important business applications that have received much more attention recently. This may be because the advancements of web and network technology make the popularity of e-commerce activities and increase the risk of network security. On the other hand, production and operations have been of less concern. In the areas of marketing and telecommunications, CRM may contain some of these problems. Similarly, fraud detection is an important problem in the area of accounting and auditing. Finally, for the problems of intrusion detection and software development, the former can be the practical problem of business security and the later for the software development outsourcing companies.

In total, 14 survey articles focusing on one of the eight specific application areas are examined and analyzed. In addition, the employed data mining techniques are presented for further comparison and discussion.

4. Results and Discussion

Table 2 summarizes related surveys in the literature related to specific business applications. In addition, related techniques employed for the eight specific application areas

are also presented.

Table 2 Data mining techniques in business applications

Application areas	Techniques
Bankruptcy Prediction	
Kumar and Ravi (2007)	● C4.5[1] ● CBR[2] ● DEA[3] ● FL[4] ● GA[5] ● MLP ● RS[6] ● SVM[7] ● Other soft computing methods, e.g. FL+MLP/GA
Lin et al. (2012)	● C4.5 ● CBR ● DEA ● FL ● GA ● LDA ● EM[8] ● K-means ● K-NN ● LR ● MLP ● NB[9] ● RS ● SOM ● SVM ● Other ensemble and hybrid combination

[1] C4.5: C4.5 decision tree
[2] CBR: case-based reasoning
[3] DEA: data envelopment analysis
[4] FL: fuzzy logic
[5] GA: genetic algorithm
[6] RF: rough sets
[7] SVM: support vector machine
[8] EM: expectation maximization
[9] NB: naïve Bayes

	methods
Verikas et al. (2010)	• C4.5 ensemble by boosting • FL+MLP+GA/MLP • GA+SVM/MLP/K-NN[10] • LR[11]+MLP • MLP ensemble by bagging • RBFNN[12] ensemble by bagging • RF+SVM/MLP/GA/K-NN • SOM+MLP • Stacking: combined MLP+K-NN+C4.5 with MLP • SVM ensemble by bagging • Voting: MLP+LDA[13]+CBR; MLP+LR+LDA+C4.5 • Other combinations of multiple techniques
Customer Relationship Management Ngai et al. (2009)	• AR[14] • BN • C4.5 • DEA • EM • GA • MCMC[15] • MLP • K-means • K-NN • LR • SA[16] • SVM
Fraud detection Ngai et al. (2011)	• BN • CART[17]

[10] K-NN: k-nearest neighbor
[11] LR: logistic regression
[12] RBFNN: radial basis function neural network
[13] LDA: linear discriminant analysis
[14] AR: association rules
[15] MCMC: Markov chain Mote Carlo
[16] SA: survival analysis
[17] CART: classification and regression tree

	• C4.5 • LDA • EA[18] • FL • K-NN • MLP • LR • NB • Probit • SOM • SVM
Intrusion detection	
Catania and Garino (2012)	• Chi-square • FL • GA • MCMC • MLP • NB • PCA[19] • SVM
Kolias et al. (2011)	• ACO[20] • ACO+SOM/SVM • PSO[21] • PSO+MLP/SVM/k-means
Tsai et al. (2009)	• C4.5 • FL • GA • K-NN • MLP • NB • SVM • Other ensemble and hybrid combination methods

[18] EA: evolutionary algorithms
[19] PCA: principal component analysis
[20] ACO: ant colony optimization
[21] PSO: particle swarm optimization

Recommender systems	
Bobadilla et al. (2013)	● BN ● FL ● GA ● K-NN ● MLP
Park et al. (2012)	● AR ● C4.5 ● K-means ● K-NN ● LA[22] ● LR ● MLP ● SOM ● Other heuristic methods
Software development effort estimation	
Wen et al. (2012)	● AR ● BN ● C4.5 ● CBR ● GA ● MLP ● SVR
Stock prediction/investment	
Atsalakis and Valavanis (2009)	● ARIMA[23] ● GA ● LR ● MLP
Ponsich et al. (2013)	● MOEA[24]
Other financial time series areas	
Ahmed et al. (2010)	● BN[25] ● CART

[22] LA: link analysis
[23] ARIMA: autoregressive integrated moving average model
[24] MOEA: multiobjective evolutionary algorithm
[25] BN: Bayesian network

	• GRNN[26] • K-NN • MLP • RBFNN • SVR[27]

From the machine learning viewpoint, these techniques can be classified into supervised and unsupervised learning techniques. Supervised learning can be thought of as learning from examples or learning with a teacher (Mitchell, 1997). The collected dataset is represented by a set of input-output examples where the input is composed of a number of variables and the output is the prediction (or output label) for the input variables. On the other hand, unsupervised learning focus on arranging a given collection of unlabelled input patterns into natural groupings or meaningful clusters based on a measure of similarity (Jain et al., 1999).

Using these criteria, the above techniques can be classified into the supervised and unsupervised learning categories and each category can also be divided into statistical and intelligent learning based techniques. Note that the definition of intelligent techniques is based on Kumar and Ravi (2007), which include most of the machine learning techniques, such as neural networks, support vector machines, fuzzy logic, evolutionary approaches, soft computing techniques, etc. Table 3 shows the classification of the data mining techniques employed in the eight business application areas. The techniques underlined are the ones that are identified by Wu et al. (2008) as the top 10 data mining algorithms.

Table 3 Classification of the employed data mining techniques

Supervised learning		**Unsupervised learning**	
Statistical tech.	*Intelligent tech.*	*Statistical tech.*	*Intelligent tech.*
• ARIMA	• ACO	• EM	• AR
• BN	• C4.5	• LA	• DEA
• Chi-square	• CART	• MCMC	• K-means
• LDA	• CBR	• PCA	• SOM
• LR	• EA	• SA	

[26] GRNN: generalized regression neural network
[27] SVR: support vector regression

• MCMC • NB • Probit	• FL • GA • GRNN • K-NN • MLP • MOEA • PSO • RBFNN • RS • SVM • SVR		

An examination of Table 3 shows that more supervised learning techniques have been employed in the literature than unsupervised learning ones. This directly relates to the identified eight business application areas (c.f. Table 1) that supervised learning techniques are the more suitable solution for most of the business problems, which are prediction oriented problems, than unsupervised learning techniques. In addition, more intelligent based techniques are used than statistical based ones. One of the main reasons is that most business applications are prediction oriented problems as has been pointed out in Bose and Mahapatra (2001), and supervised learning techniques are much more suitable for these problems.

According to Tables 2 and 3, Figure 1 shows the different types of techniques used in the eight application areas. Among these areas, popular business application areas can be identified. That is, 15 different techniques were employed for the bankruptcy prediction problem, 13 different techniques for the customer relationship management and fraud detection areas, 12 for intrusion detection, and 11 for the area of recommender systems. The results indicate that some relevant data mining techniques have not been studied in the areas of software development effort estimation, stock prediction/investment, and other financial time-series areas. However, we can observe that financial data mining is the most widely studied domain when bankruptcy prediction, stock prediction/investment, and other financial time-series areas are considered as the financial problem category.

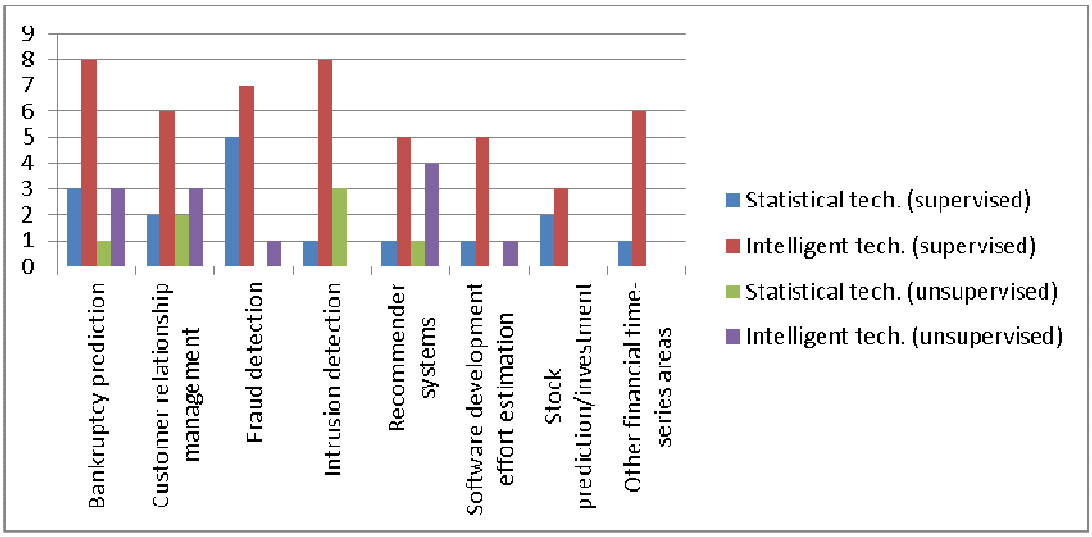

Figure 1 Techniques used in the eight application areas

Besides the techniques discussed above, ensemble learning (or soft learning) is another category of machine learning technique, usually based on combining multiple supervised and/or unsupervised learning techniques (Kittler et al., 1998). Specifically, linear and parallel combinations of multiple techniques can be implemented.

For the example, two techniques, *A* and *B,* used in the linear approach are combined, with a given a dataset *D*, where *D* is used as the input for the first component *A*, and then the output of *A* is used as the input for the second component *B*. The output of *B* is the final result for the dataset problem. One example of this is the popular neuro-fuzzy approach (Lin and Lee, 1996).

On the other hand, suppose that three techniques, *A*, *B*, and *C* are combined in a parallel manner and then dataset *D* is used as the input for *A*, *B*, and *C*, respectively. Then, the outputs obtained by *A*, *B*, and *C* are combined through some combination methods, such as simple voting or weighted voting. Other common combination methods for this type of ensemble technique are bagging (Breiman, 1996) and boosting (Freund, 1995).

According to Table 2, various ensemble techniques have only been employed in three application areas, which are bankruptcy prediction, intrusion detection, and recommender systems. This indicates that the likelihood of using ensemble or soft learning techniques is

problem dependent. In addition, the results also suggest that the relevant ensemble or soft techniques have not been fully examined in the other five application areas.

From an examination of the data mining techniques discussed in Wu et al. (2008) and this study, the top data mining techniques in business applications can be identified. Without a doubt, the top supervised learning techniques are the decision tree methods including C4.5 and CART, K-NN, and SVM. However, different from Wu et al. (2008), artificial neural networks (including MLP, GRNN, and RBFNN) and evolutionary algorithms (including ACO, GA, MOEA, and PSO) are two other types of widely employed techniques. More specifically, MLP from the category of artificial neural networks and GA from the category of evolutionary algorithms are the most frequently used techniques in the categories of artificial neural networks and evolutionary algorithms, respectively. NB is the most popular statistical technique.

Of the unsupervised learning techniques, the AR and K-means techniques are the top intelligent techniques and EM is the most commonly used statistical techniques for related business applications.

5. Conclusion

This study investigates the related survey literature in order to identify the frequently used data mining techniques in various business application areas. In total, eight business application areas where data mining techniques are used are surveyed. They are bankruptcy prediction, customer relationship management, fraud detection, intrusion detection, recommender systems, software development effort estimation, stock prediction/investment, and other financial time-series areas.

In addition, it is found that 33 different techniques have been employed in these eight business application areas. Among them, there are 24 and 9 supervised and unsupervised learning techniques, respectively. This demonstrates that most business problems are

prediction oriented. Furthermore, bankruptcy prediction is the most widely studied application area where 15 different techniques were used.

For the widely used data mining techniques, 10 different techniques are identified, which include 7 supervised and 3 unsupervised techniques, namely decision trees (including C4.5 and CART), genetic algorithms, k-nearest neighbors, multilayer perceptron neural networks, naïve Bayes, and support vector machines as the supervised learning techniques and association rules, expectation maximization, and k-means as the unsupervised learning techniques.

Therefore, the contribution of this paper is to provide the guideline for relevant academicians and practitioners to quickly decide more suitable techniques from the 10 techniques instead of the 33 techniques listed from Table 2. Moreover, the effort for the performance comparison can be reduced since only a few number of techniques is chosen.

Finally, there are four suggestions for future business data mining studies. First, although ensemble or soft learning techniques have been shown to outperform many single data mining techniques (Oza and Tumer, 2008; Rokach, 2010; Verikas et al., 2010), they have not received much attention in many business application areas, such as customer relationship management, fraud detection, software development effort estimation, stock prediction/investment, and other financial time-series areas. Second, for a specific domain problem, it is necessary to employ the frequently applied data mining techniques identified in this paper for comparison in order to reach a reliable conclusion and provide the most effective solution for that problem. However, there may be some exceptions for selecting these techniques based on certain situations and organizational parameters. Third, some other techniques, such context-analysis and text mining (Tsai and Hung, 2013), can be considered to further demonstrate the widely used data mining techniques. Fourth, the collected data (i.e. the survey/review articles) could be regularly updated in order to identify newer techniques and/or business application areas if any.

References

Ahmed, N.K., Atiya, A.F., Gayar, N.E., and EL-Shishiny, H. (2010) An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, vol. 29, no. 5-6, pp. 594-621.

Atsalakis, G. and Valavanis, K.P. (2009) Surveying stock market forecasting techniques – Part II: soft computing methods. *Expert Systems with Applications*, vol. 36, pp. 5932-5941.

Bobadilla, J., Ortega, F., Hernando, A., and Gutierrez, A. (2013) Recommender systems survey. *Knowledge-Based Systems*, vol. 46, pp. 109-132.

Bose, I. and Mahapatra, R.K. (2001) Business data mining – a machine learning perspective. *Information & Management*, vol. 39, pp. 211-225.

Breiman, L. (1996) Bagging predictors. *Machine Learning*, vol. 24, no. 2, pp. 123-140.

Catania, C.A. and Garino, C.G. (2012) Automatic network intrusion detection: current techniques and open issues. *Computers & Electrical Engineering*, vol. 38, pp. 1062-1072.

Freund, Y. (1995) Boosting a weak learning algorithm by majority. *Information and Computation*, vol. 121, no. 2, pp. 256-285.

Han, J. and Kamber, M. (2001) *Data mining: concepts and techniques*. Morgan Kaufmann.

Jain, A.K., Murty, M.N., and Flynn, P.J. (1999) Data clustering: a review. *ACM Computing Survey*, vol. 31, no. 3, pp. 264-323.

Kittler, J., Hatef, M., Duin, R.P.W., and Matas, J. (1998) On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239.

Kolias, C., Kambourakis, G., and Maragoudakis, M. (2011) Swarm intelligence in intrusion detection: a survey. *Computer & Security*, vol. 30, pp. 625-642.

Kumar, P.R. and Ravi, V. (2007) Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review. *European Journal of Operational Research*, vol. 180, pp. 1-28.

Lin, C.-T. and Lee, C.S.G. (1996) *Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems*. Prentice Hall.

Lin, W.-Y., Hun, Y.-H., and Tsai, C.-F. (2012) Machine learning in financial crisis prediction: a survey. *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, vol. 42, no. 4, pp. 421-436.

Mitchell, T. (1997) *Machine Learning*. McGraw Hill.

Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y., and Sun, X. (2011) The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. *Decision Support Systems*, vol. 50, pp. 559-569.

Ngai, E.W.T., Xiu, L., and Chau, D.C.K. (2009) Application of data mining techniques in customer relationship management: a literature review and classification. *Expert Systems with Applications*, vol. 36, p. 2592-2602.

Oza, N.C. and Tumer, K. (2008) Classifier ensembles: select real-world applications. *Information Fusion*, vol. 9, pp. 4-20.

Park, D.H., Kim, H.K., Choi, I.Y., and Kim, J.K. (2012) A literature review and classification of recommender systems research. *Expert Systems with Applications*, vol. 39, pp. 10059-10072.

Ponsich, A., Jaimes, A.L., and Coello, C.A.C. (2013) A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications. *IEEE Transactions Evolutionary Computation*, vol. 17, no. 3, pp. 321-344.

Rokach, L. (2010) Ensemble-based classifiers. *Artificial Intelligence Review*, vol. 33, pp. 1-39.

Smith, K. and Gupta, J.N.D. (2000) Neural networks in business: techniques and applications for the operations researchers. *Computers & Operations Research*, vol. 27, pp. 1023-1044.

Tsai, C.-F., Hsu, Y.-F., Lin, C.-Y., and Lin, W.-Y. (2011) Intrusion detection by machine learning: a review. *Expert Systems with Applications*, vol. 36, pp. 11994-12000.

Tsai, C.-F. and Hung, C. (2013) Popular research topics in multimedia. *Scientometrics*, vol. 95, no. 1, pp. 465-479.

Shaughnessy, A.F. and Slawson, D.C. (2003) What happened to the valid POEMs? A survey of review articles on the treatment of type 2 diabetes. *British Medical Journal*, vol. 327, pp. 1-7

Vellido, A., Lisboa, P.J.G., and Vaughan, J. (1999) Neural networks in business: a survey of applications (1992-1998). *Expert Systems with Applications*, vol. 17, pp. 51-70.

Verikas, A., Kalsyte, Z., Bacauskiene, M., and Gelzinis, A. (2010) Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. *Soft Computing*, vol. 14, pp. 995-1010.

Wen, J., Li, S., Lin, Z., Hu, Y., and Huang, C. (2012) Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology*, vol. 54, pp. 41-59.

Wong, B.K., Bodnovich, T.A., and Selvi, Y. (1997) Neural network applications in business: a review and analysis of the literature (1988-95). *Decision Support Systems*, vol. 19, pp. 301-320.

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J., and Steinberg, D. (2008) Top 10 algorithms in data mining. *Knowledge and Information Systems*, vol. 14, pp. 1-37.

Ye, N. (2003) *The handbook of data mining*. Lawrence Erlbaum.

Dr. Wei-Chao Lin is an associate professor at the Department of Computer Science and Information Engineering, Asia University, Taiwan. His research focuses on image retrieval and pattern recognition.

Dr. Shih-Wen Ke is an assistant professor at the Department of Information and Computer Engineering, Chung Yuan Christian University, Taiwan. His research interests cover information retrieval, data mining, and medical informatics.

Dr. Chih-Fong Tsai is a professor at the Department of Information Management, National Central University, Taiwan. He has published more than 60 technical publications in journals, book chapters, and international conference proceedings. His current research focuses on and data mining and its applications.