Journal of the Operational Research Society (2009) 60, S16–S23

www.palgrave-journals.com/jors/

50 years of data mining and OR: upcoming trends and challenges

B Baesens[1,2]*, C Mues[2], D Martens[1,3] and J Vanthienen[1]

[1]K.U. Leuven, Leuven, Belgium; [2]University of Southampton, Southampton, UK; and [3]University College Ghent, Ghent, Belgium

Data mining involves extracting interesting patterns from data and can be found at the heart of operational research (OR), as its aim is to create and enhance decision support systems. Even in the early days, some data mining approaches relied on traditional OR methods such as linear programming and forecasting, and modern data mining methods are based on a wide variety of OR methods including linear and quadratic optimization, genetic algorithms and concepts based on artificial ant colonies. The use of data mining has rapidly become widespread, with applications in domains ranging from credit risk, marketing, and fraud detection to counter-terrorism. In all of these, data mining is increasingly playing a key role in decision making. Nonetheless, many challenges still need to be tackled, ranging from data quality issues to the problem of how to include domain experts' knowledge, or how to monitor model performance. In this paper, we outline a series of upcoming trends and challenges for data mining and its role within OR.

Journal of the Operational Research Society (2009) **60,** S16–S23. doi:10.1057/jors.2008.171
Published online 11 February 2009

Keywords: data mining; learning algorithms; decision support systems; applications; prediction

Introduction: data mining

The research area of data mining refers to the process of extracting previously unknown patterns or models from (often very large) data sets. Data mining techniques and applications can be categorized as being either predictive or descriptive (Witten and Frank, 2000). Predictive data mining entails predicting the value for a certain target variable, based on historical data. When this target is discrete, we refer to the task at hand as classification. There are many applications, including predicting the repayment behaviour of loan applications (known as credit scoring), predicting customer 'churn', and classifying an insurance claim as fraudulent or not. The credit scoring example problem is illustrated in Figure 1. Regression, on the other hand, is the task of predicting the value of a continuous target variable. Typical examples are stock price, credit loss, and sales amount prediction.

Descriptive data mining aims at finding patterns that describe underlying relationships in the data, for example association rules and clustering. Association rule mining looks for frequently occurring patterns in the data and is often used for market basket analysis. A well-known example is the rule: if someone buys nappies then this person is also likely to buy beer. Although its actual truthfulness has been questioned, its use as a marketing vehicle for data mining has surely been effective.

Since data mining is often used to support decision making, it can to a large extent be considered part of operational research (OR), a substantial part of which also focuses on developing various kinds of decision models from data. Hence, clearly there is substantial potential for cross-fertilization between both disciplines. In this paper, we begin by giving a brief overview of the two-way interaction between data mining and OR. We then discuss some recent upcoming trends and challenges facing data mining approaches, which may among other things involve the application of OR techniques. In particular, we address the issue of data quality, the need for interpretable white-box data mining models, the role of domain knowledge, the need for backtesting and stress testing data mining models, networked-based learning, and some exciting new application areas.

Data mining and its role within OR

The origins of data mining predate the birth of OR. Durand (1941) used quadratic discriminant analysis for analysing a credit scoring data set of more than 7000 observations. Credit scoring, where the aim is to distinguish good payers from defaulters based on a set of customer characteristics, can undoubtedly be considered as one of the most popular application fields for both data mining and OR techniques. Many of the specific features of this problem statement have led to the development of new data mining/OR techniques, which will be illustrated later in this paper.

Mangasarian (1965) applied linear programming to perform linear and nonlinear separation of patterns. This

*Correspondence: *B Baesens, Department of Decision Sciences and Management, K.U. Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.*
E-mail: Bart.Baesens@econ.kuleuven.be

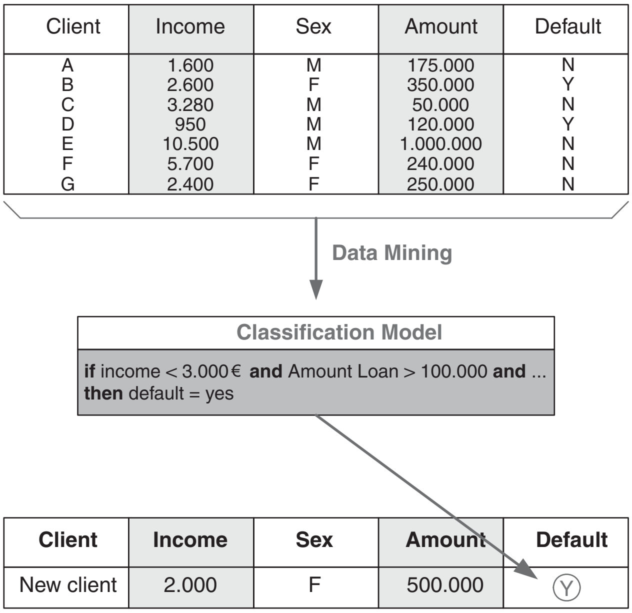


Figure 1 Predicting the creditworthiness of loan applications based on historical data.

OR-based solution to a data mining problem has been validated in several empirical application settings, for example credit scoring (Thomas *et al*, 2002), customer relationship management (CRM) (Padmanabhan and Tuzhilin, 2003), and breast cancer detection (Mangasarian *et al*, 1995). Furthermore, the modern data mining algorithm support vector machines (SVM) essentially uses ideas from mathematical programming to develop powerful nonlinear prediction models (Burges, 1998). The principles of mathematical programming have also been successfully used in other data mining settings, for example for data clustering (Bradley *et al*, 1997), data visualization (Abbiw-Jackson, 2006), decision tree construction (Bennett, 1992), attribute selection (Bredensteiner and Bennett, 1998; Yang and Olafsson, 2005), and classification cut-off setting (Thomas *et al*, 2002). Bradley *et al* (1999) give an overview of mathematical programming formulations and challenges for data mining.

Earlier OR-based forecasting methods have also been introduced for data mining. Popular examples include linear regression and time series analysis. Both techniques have been successfully applied in data mining settings, and have laid the foundation for developing new techniques, for example projection pursuit regression, multivariate adaptive regression splines, neural networks, and regression trees. See Fildes *et al* (2008) for an overview of the historical relationship between forecasting and OR.

The OR concepts of Markov processes and Monte Carlo sampling have also been widely used in data mining, for example model selection and Bayesian network learning (Baesens *et al*, 2002; Giudici and Castelo, 2003). Social network analysis also often uses the Markov property for learning structures from networked data (see below).

Recently, ant colony optimisation (ACO) algorithms have been introduced as a new promising OR technique (Dorigo and Stützle, 2004). Initial successful applications include the travelling salesman problem, manufacturing control, and routing of packages through the Internet (Di Caro and Dorigo, 1998). However, the same optimization method can also be applied for data mining. For example, Martens *et al* (2007b) used ACO to develop AntMiner+, a system capable of extracting if–then classification rules from data. Genetic/evolutionary algorithms are another example of optimization algorithms that can serve both traditional OR and data mining applications. As such, the latter have been successfully used to perform feature selection, train neural networks, infer classification rules, or optimize the weights of classifiers in a data mining ensemble (Freitas, 2007).

From the above examples, it becomes clear that both OR and data mining are growing more closely intertwined, with no clear-cut boundary in between (Olafsson, 2006). Both have contributed significantly to the analysis of decision problems, and the discussion of whether a technique should be considered OR or data mining seems increasingly irrelevant.

In this paper, we will outline a series of upcoming trends and challenges for data mining and its role within OR. It is obvious that these challenges will require new developments, for example by devising new (optimization) algorithms, or re-using earlier introduced OR/data mining methods in new settings.

Data quality

Both OR and data mining rely on the use of data. Bad data yields bad models. This is often referred to as the GIGO principle: garbage data in, garbage results out. Hence, it is of crucial importance that data is of good quality in order to obtain good OR and/or data mining models. Furthermore, it has been shown in many data mining studies that simple models generally perform well on most data sets. For example, the benchmarking study of Baesens *et al* (2003b) showed that logistic regression typically performs quite well for classification, whereas Holte (1993) obtained a similar finding using simple classification rules. Hence, in several areas, the best way to augment the performance of a data mining/OR model often is to improve data quality. Measuring data quality is however not an easy endeavour and many challenges lie ahead in trying to improve data quality. In the next section, we discuss five aspects of data quality: data accuracy, data completeness, data bias and sampling, data transformations, and data definition.

Data accuracy determines the extent to which the data accurately and consistently measures what it is intended to measure. Poor data accuracy can be caused by data entry errors and/or measurement errors. Outliers may also be a sign of poor data accuracy, although it should be noted that an outlier does not necessarily represent an invalid observation. For example, when considering ratio variables in a credit scoring data mining system, it is often observed that these variables have highly dispersed distributions, because the division of two Gaussian distribution is a Cauchy distribution with fat tails. Hence suitable methods must be adopted to deal with these valid, outlying observations, so that the data mining system is best able to detect genuine patterns in the data. One way is using robust winsorizing schemes (Van Gestel *et al*, 2005). However, other interesting outlier handling schemes could be developed, taking into account the type of variable, the reason why it is an outlier, and of course the data mining technique adopted.

Data completeness refers to the proportion of missing values and/or observations in the data. Missing information should clearly be minimized, especially for variables that are retained in the data mining model; however, if present, appropriate procedures should be used to deal with it. Although many approaches have been suggested to deal with missing values, for example (multiple) imputation routines or introducing a separate category/variable, it is still not always clear which is the best approach given the problem domain and the data mining technique adopted.

The next data quality criterion relates to data bias and sampling. When building a data mining model, one typically starts from a sample of data. A first important issue relates to the sampling strategy to be adopted. It is very important to select the observations carefully, using both active and adaptive learning strategies. Also, selecting a (near-)optimal set of variables is far from straightforward, particularly in high-dimensional data sets, which often occur in bio-informatics applications. It is important to note that, in many cases, the chosen sample is not representative of the population for which the data mining model is intended. For example, when building data mining models for credit scoring, one only has the target variable (good/bad) available for the sample of past accepts, whereas the future population may not be exactly the same, because of the past credit policy. This leads to the problem of reject inference, which has been the subject of intensive study using a variety of techniques, but has to date not been solved unambiguously (Thomas *et al*, 2002). Closely related to this is the problem of policy inference, where data mining models are being used to apply different policies to different customers, making it more difficult to decide whether the actual outcome was due to the policy applied or to customer characteristics. Similar problems also often occur in other data mining settings. Hence, studying the impact of data bias and its impact on data mining models is clearly an important issue of future research.

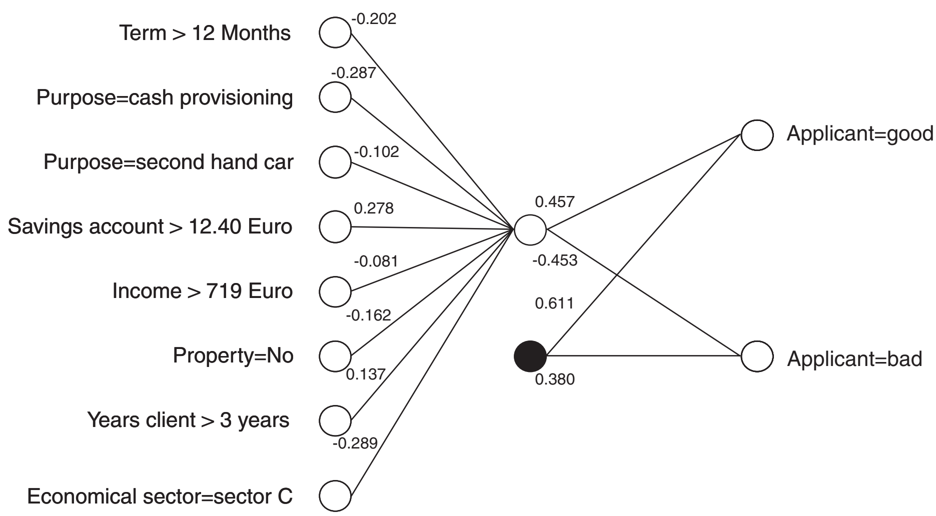
Data transformations are often applied to make attributes more informative, so that the data mining technique can more easily detect patterns. A popular example is coarse classification, which categorizes categorical and/or continuous variables for more robust analysis or for introducing nonlinear effects into linear models. Another example are Box–Cox transformations, which are a set of simple logarithmic transformations that may significantly improve the performance of classification models (Van Gestel *et al*, 2005). Principal component analysis is also often used to transform the variables to a set of uncorrelated principal components, typically at the cost of decreased interpretability. A crucial issue here is the choice of the most appropriate transformations for the data mining task at hand (ie classification, regression, clustering, and so on) given model priorities, such as accuracy, interpretability, or re-training efforts, for example.

Data definition relates to the way data attributes have been defined. Consider for example a corporate credit risk rating system in which the ratio debt/earnings is used as a variable in a linear/logistic regression based data mining system. Because earnings can be both negative and positive, the impact of that variable on the default risk is not uniquely defined. By defining the variable in that way, one may already force the model to make a suboptimal decision, both in terms of performance and interpretability. Hence, it is of crucial importance to think carefully about how variables have been defined and what might be their impact on the target variable of interest.

In summary, in the preceding paragraphs we have outlined various issues related to data quality. We clearly stressed that these issues should be dealt with in the best way possible, in order to obtain the most powerful data mining system. However, further research is needed on how to improve data quality by either redesigning data entry processes (using, eg, validation constraints or business rules), or by developing new procedures to optimally encode, transform or redefine data, thereby allowing the data mining/OR system to better detect patterns in the data. It is clear that traditional OR techniques such as linear programming and genetic algorithms can definitely also provide added value in this context.

Interpretable data mining models

In recent years, many data mining algorithms for a variety of different problems have been developed, for example neural networks, SVM, Bayesian networks, genetic algorithms, fuzzy techniques, and swarm intelligence. An important focus of these developments was the optimal detection of patterns in data, given a predetermined performance metric, such as classification accuracy, R-squared, or mean squared error. Although it is undoubtedly important, technical performance is however not the only relevant criterion for the successful deployment of data mining models in business. Increasingly, interpretability and model readability are considered key critical success factors. A popular illustration of this is the recently proposed Basel II Capital Accord, which encourages financial institutions to develop data mining models estimating default risk, loss risk and exposure risk. Given the strategic impact of these models on measures such as capital levels maintained, financial regulators are reluctant to approve the use of complex, black-box models. Thus there is a need for white-box data mining models that give a clear and transparent view of the patterns in the data. The well-known principle of Ockam's razor embodies this idea



if Term >12 months **and** Purpose=cash provisioning **and** Savings Account<=12.40 Euro **and** Years client<=3 **then** *Applicant=bad*

if Term >12 months **and** Purpose=cash provisioning **and** Owns Property=No **and** Savings Account <=12.40 Euro **then** *Applicant=bad*

if Purpose=cash provisioning **and** Income>719 Euro **and** Owns Property=No **and** Savings Account <= 12.40 Euro **and** Years client<=3 **then** *Applicant=bad*

if Purpose=second hand car **and** Income>719 Euro **and** Owns Property=No **and** Savings Account <= 12.40 Euro **and** Years client<=3 **then** *Applicant=bad*

if Savings Account <=12.40 Euro **and** Economical sector=Sector C **then** *Applicant=bad*

Default class: *Applicant=good*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Savings Account	<= 12.40 Euro													> 12.40 Euro
2. Economical sector	Sector C	other												-
3. Purpose	-	cash provisioning							second-hand car				other	-
4. Term	-	<= 12 months				> 12 months			-				-	-
5. Years client	-	<= 3			> 3	<= 3	> 3		<= 3			> 3	-	-
6. Owns Property	-	Yes	No		-	-	Yes	No	Yes	No		-	-	-
7. Income	-	-	<= 719 Euro	> 719 Euro	-	-	-	-	-	<= 719 Euro	> 719 Euro	-	-	-
1. applicant = good	-	x	x	-	x	-	x	-	x	x	-	x	x	x
2. applicant = bad	x	-	-	x	-	x	-	x	-	-	x	-	-	-

Figure 2 Developing interpretable data mining models using neural network rule extraction and decision tables (Baesens *et al*, 2003a). The network above was used to extract propositional if–then rules presented in the middle, which were then transformed to a decision table given at the bottom of the figure.

of model simplicity. However, aiming for model understandability often comes at the price of decreased performance, so trade-offs between model readability and model performance need to be taken into account. In recent years, attempts have been made to balance this trade-off. Baesens *et al* (2003a) used neural network rule extraction as a way of opening up

Table 1 Taxonomy of possible domain constraints to incorporate into the data mining model

	Univariate				*Multivariate*
		Ordinal			
	Nominal	*Linear*	*Nonlinear*		
			Piecewise linear	*Non-piecewise linear*	
Soft					
Hard					

the neural network black box and shed light on its decision logic, by extracting propositional if–then rules mimicking its behaviour. Furthermore, in order to represent the rules in the most user-friendly way, decision tables were adopted. This entire process is illustrated in Figure 2. The translation of this work from neural networks to SVM is currently also being addressed (eg Martens *et al*, 2007a).

Similarly, Martens *et al* (2007b) used ACO algorithms to extract propositional if–then rules from data. It should be noted that other types of rules, such as M-of-N rules, oblique rules, or fuzzy rules, can be considered as viable alternatives. Further research is needed into the most suitable rule representation in terms of simplicity and ease of use, in order to derive interpretable and user-friendly data mining models.

Van Gestel *et al* (2005) introduced an alternative way of obtaining interpretable models by using an additive combination of linear (logistic regression) and nonlinear (SVM) modelling techniques to find an optimal balance between model interpretability, obtained from the linear part, and model performance, obtained from the nonlinear part. Note that also Bayesian networks are considered white-box probabilistic models with a high degree of model transparency and readability.

Despite recent advances, many challenges still lie ahead in obtaining interpretable data mining models. Examples of such challenges are:

- What is the preferred model representation in terms of simplicity, for example rule-based models, linear models, or graphical models?
- How can we choose the optimal balance between model interpretability and model performance?
- How can we suitably quantify not just performance but both criteria?

Incorporating domain knowledge

Although data mining techniques were originally introduced to analyse large-scale data sets, the same techniques are increasingly being used to mine small data sets. Examples of these are in medical settings where only a limited number of observations may be available, or financial institutions with low default portfolios (eg exposures to sovereigns, or banks). Detecting patterns from small volume data sets requires the use of special techniques. In this context, one often tries to include domain-specific expert knowledge during the learning process. In this way, the purpose is to derive a model using an optimal combination of limited data with any available business expert knowledge. The problem of how to consolidate automatically generated data mining knowledge with knowledge reflecting experts' domain expertise constitutes the so-called knowledge fusion problem (Martens *et al*, 2006).

There are many types of constraints that a domain expert might wish to incorporate. The most popular constraint is the monotonicity constraint, stating that an increase in a certain input variable cannot lead to a decrease in the output variable. A credit scoring example is that an increasing income, keeping all other variables equal, should yield a decreasing probability of loan default. Although almost all research in this field is focused on this constraint, the taxonomy of possible domain constraints, shown in Table 1, indicates that this is arguably too limited a view. It is, however, the most common type of constraint to be incorporated. Each constraint can be either mandatory, which we term a hard constraint, or simply preferred, which we term a soft constraint. An example of multivariate constraints is requiring that variable X is more important in the model than variable Y. A detailed overview of examples and existing research can be found in Martens and Baesens (2009).

In the early OR approaches to data mining, such as linear programming, it was very easy to include domain knowledge. For example, an additional constraint specifying that the weight of one variable should be greater than a constant or the weight of some other variable could often do the job. However, in newer techniques, such as rule extraction techniques, neural networks, SVM, ACO systems, or Bayesian networks, it is harder to identify the best way to take account of domain knowledge. Hence, in this area, there are many issues and challenges for further study. Examples are:

- What is the best way of formalizing domain knowledge?
- How can we best elicit domain knowledge from business experts?
- What is the best way to include domain knowledge when learning patterns from data?
- How can we resolve conflicts between patterns learnt from data and domain knowledge?

Backtesting and stress testing data mining models

Data mining models are typically constructed using a snapshot of historical data. Once implemented in practice, the performance of these models may degrade over time. Three obvious reasons for this are sampling variability, macro-economic influences, and endogenous effects. Sampling variability is the uncertainty due to the fact that the sample only gives a limited view of the population. Macro-economic influences reflect the impact economic up- or downturns may have on the model. Endogenous effects represent changes induced by the firm itself, for example strategy changes or exploration of new market segments. These three effects can have a strong impact on the performance of a data mining model, and they typically occur in a mixed way. Clearly, this is a complicating factor when backtesting the models, that is, tracking their performance over time. Hence, backtesting frameworks should constantly monitor the behaviour of data mining models, and perform diagnostic checks on its performance taking into account the influences mentioned above. In doing so, it is important to be aware of the philosophy that was adopted when designing the data mining models. A distinction can sometimes be made between point-in-time (PIT) data mining models, which take into account both cyclical and non-cyclical information, and through-the-cycle (TTC) data mining models which only focus on non-cyclical inputs. For example, a customer score can measure churn behaviour or default risk during the upcoming year (PIT), or over a longer term horizon (TTC). Knowing this helps to judge the significance of temporary fluctuations due (say) to macro-economic volatility. Clearly, both PIT and TTC represent the extremes of a continuum, and it is vital to know towards which extreme the data mining system is situated before any backtesting exercise can start.

When working out a backtesting framework, a first critical question is which test statistics, metrics and/or measurements should be used for monitoring data mining model performance. For example in a clustering setting, one might be interested in monitoring cluster stability, whereas in a classification setting one might be interested in monitoring data stability, discriminatory power and/or probability calibration. When developing these backtesting diagnostics, one should clearly take into account the three influences measured above, combined with the fact that customer behaviour might be correlated. Measuring and taking these correlation effects into account is a key challenge for a successful backtesting exercise. Furthermore, clear advice should be given regarding the significance levels to be adopted. Next, all the diagnostic checks need to be combined in a user-friendly intelligent dashboard environment allowing constant, real-time monitoring of the data mining model. Finally, action plans should be available that specify what to do in case the data mining model starts degrading in performance, ranging from simple parameter adjustment and incremental learning to complete model re-estimation. Although preliminary work has been done in this area (see, eg Castermans *et al* (2007) for an overview of backtesting credit risk models), more research needs to be conducted to appropriately answer the above questions and come up with powerful backtesting frameworks that can be implemented in a user-friendly way.

Since data mining models are increasingly being used as key inputs for strategic organization processes (eg allocating marketing budgets or capital calculation in a Basel II setting), it is crucial to know how they behave under adverse economic circumstances. The aim of stress testing is to determine the impact of stress events such as macro-economic downturns on data mining models. This can range from simple sensitivity checks to advanced scenario-based analysis, based on historical or hypothetical scenarios. It is clear that also in this area, much work lies ahead to come up with reliable methods for stress testing data mining models and/or making them less stress sensitive.

Network-based learning

Typically, data mining is performed on a set of independent and identically distributed (iid) observations. However, in practice most observations relate to agents acting in a network-based structure. Examples of networked data include web pages connected by hyperlinks, research papers linked by citations, and customers linked by social networks (eg telephone calls, or social network sites such as LinkedIn or Facebook). It is believed that quantifying these social network structures and including them as part of the data mining learning process can have a significant impact on the type and power of patterns and/or models being learnt. For example, graph-based data mining algorithms can be used for community discovery and/or web structure mining, or information about an individual's social network may improve the performance of response or churn models in the marketing domain.

However, many challenges lie ahead. Since the class membership of one entity may influence the class membership of another, relational data mining algorithms are being developed that take the network dependencies into account. Furthermore, methods of collective inference are being devised to determine how the unknown nodes in a social network can be jointly estimated. Also, since a network functions as a complete structure, it is hard to separate it into a training and independent test set for performance calculation. Although preliminary work has been done in this area (see eg Lu and Getoor, 2003; Macskassy and Provost, 2007), there is currently a great need for more empirical applications of network-based learning, in order to assess its added value over existing approaches, and potentially, for new algorithms to be developed. Again, this creates new opportunities for using existing OR techniques in exciting novel settings.

New application areas

Having been originally introduced in well-known problem domains such as CRM, or specific applications such as churn prediction, fraud detection, or credit scoring, data mining is currently being applied in many new exciting application areas. One interesting novel application is web analytics, where data mining is being used for analysing all kinds of data collected from the World Wide Web (Liu, 2006). This involves web usage mining, which studies web logs for detecting web usage patterns, web content mining, which focuses on analysing the content of web pages, and web structure mining, which aims at mining communities from a set of hyperlinks.

Another popular application domain is bio-informatics, where data mining is being studied for analysing biological data, such as protein sequence databases (De Moor *et al*, 2003). Text mining applications are also becoming increasingly popular. Typical challenges in both are the relatively small number of observations compared to the high number of dimensions, which calls for special types of algorithms that are able to cope with such sparse data. Data mining is also becoming increasingly popular for classical medical analysis problems, for example cancer detection or drug effect analysis. A key challenge in this field is the previously discussed problem of merging patterns derived from data with knowledge extracted from domain experts such as medical practitioners.

Data mining techniques are currently also being applied to terrorism prevention. The application of social network mining techniques seems very promising in this area. However, there are major issues here relating to data confidentiality, civil liberties and the protection of privacy. Consumer privacy is also a major concern for mining RFID (Radio Frequency Identification) data, particularly in a retail context. RFID is an increasingly popular supply-chain technology that overcomes several of the limitations of traditional barcodes and which can provide huge amounts of data on product identity, location, and movement. RFID mining would be of particular interest for OR as it would allow the possibility to track products as they move through the supply chain or (say) are transported across large warehouse spaces, which can aid in the further optimization of these processes.

Another exciting application is in software engineering, where data mining has been introduced to predict faults in software, or software development efforts (Lessmann *et al*, 2008). More recently, work has also begun on mining data in software repositories for tasks ranging from identifying code usage patterns to the analysis of change patterns to assist in future development or, for example, assigning programmers to particular tasks. Research in this area is aided by the public availability of large amounts of data on open-source software development projects.

Closely related to this is the area of process mining, where data mining techniques are being used to mine process models such as Petri nets from event logs of an information system (Van der Aalst and Weijters, 2004). This allows organizations to better understand their key operational processes and workflows by showing them in what order and by whom various activities are being performed in practice, and how this may differ from the processes designed ‘on paper’.

Another emergent area is real-time data mining applications, in which interesting patterns or anomalies are being discovered on a continuous basis. In the real world, data and customer behaviour are changing on a continuous basis and thus the idea of building a ‘static’ data mining model that is subsequently used for a fixed period of time may no longer be appropriate. The data mining model has to be amenable to very quick, if not automatic updating to allow for real-time decision making. Applications are typically found in product configuration and pricing (Seow and Thomas, 2007), fraud prevention, and also in network intrusion detection systems (Lee *et al*, 2001).

The use of data mining techniques for analysing videos (video mining) also offers many exciting applications, for example, image/video retrieval, video summarization, visual surveillance, and real-time decision making during sports matches. Closely related to this is the application of data mining to astronomy data, where the aim is to analyse data obtained by telescopes to find new phenomena, relationships, and useful knowledge about the universe.

Conclusions

The importance of data mining has steadily grown in recent decades, with an ever-increasing range of techniques being developed. Whereas traditionally many applications have been in the finance and marketing domains, its use has now spread to several other domains. However, many challenges and opportunities for data mining still lie ahead. In this paper we identified a number of outstanding issues and current trends relating to data quality, the interpretability of data mining models, the incorporation of domain knowledge into the data mining process, the need to backtest and stress test models, network-based learning, and a number of new application fields. For many of these, it was argued that data mining may continue to benefit from the introduction of techniques from the OR area.

Acknowledgements—This research was supported by the Odysseus program (Flemish Government, FWO) under grant G.0915.09.

References

Abbiw-Jackson R, Golden BL, Raghavan S and Wasil EA (2006). A divide-and-conquer local search heuristic for data visualization. *Computers & OR* **33**: 3070–3087.

Baesens B, Egmont-Petersen M, Castelo R and Vanthienen J (2002). Learning bayesian network classifiers for credit scoring using Markov Chain Monte Carlo search. In: Sanfeliu A, Villanueva JJ, Vanrell M, Alquézar R, Jain AK and Kittler J (eds). *Proceedings of the Sixteenth International Conference on Pattern Recognition (ICPR'2002)*. IEEE Computer Society: Québec, Canada, August, pp 49–52.

Baesens B, Setiono R, Mues C and Vanthienen J (2003a). Using neural network rule extraction and decision tables for credit-risk evaluation. *Mngt Sci* **49**(3): 312–329.

Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J and Vanthienen J (2003b). Benchmarking state-of-the-art classification algorithms for credit scoring. *J Opl Res Soc* **54**(6): 627–635.

Bennett KP (1992). Decision tree construction via linear programming. In: Evants M (ed). *Proceedings of the Fourth Midwest Artificial Intelligence and Cognitive Science Conference*. Utica, Illinois, Southern Illinois University: Illinois, pp 97–101.

Bradley PS, Mangasarian OL and Street WN (1997). Clustering via concave minimization. In: Mozer MC, Jordan MI and Petsche T (eds). *Advances in Neural Information Processing Systems*. MIT Press: Cambridge, MA.

Bradley PS, Fayyad UM and Mangasarian OL (1999). Mathematical programming for data mining: Formulations and challenges. *Informs J Comput* **11**(3): 217–238.

Bredensteiner EJ and Bennett KP (1998). Feature minimization within decision trees. *Comput Optim Appl* **10**: 111–126.

Burges JC (1998). A Tutorial on support vector machines for pattern recognition. *Data Mining Knowl Discov* **2**: 121–167.

Castermans G, Martens D, Van Gestel T, Hamers B, Baesens B (2007). An overview and framework for PD backtesting and benchmarking. In: Crook J, Thomas L and Hand DJ (eds.) *Conference on Credit Scoring and Credit Control X*, Edinburgh, UK, July. Edinburgh University Press: Edinburgh.

De Moor B, Marchal K, Mathys J and Moreau Y (2003). Bioinformatics: Organisms from Venus, technology from Jupiter, algorithms from Mars. *Eur J Control* **9**: 237–278.

Di Caro G and Dorigo M (1998). Antnet: Distributed stigmergetic control for communications networks. *J Artif Intel Res* **9**: 317–365.

Dorigo M and Stützle T (2004). *Ant Colony Optimization*. MIT Press: Cambridge, MA.

Durand D (1941). *Risk Elements in Consumer Instalment Financing*. National Bureau of Economic Research: New York.

Fildes R, Nikolopoulos K, Crone SF and Syntetos AA (2008). Forecasting and operational research: A review. *J Opl Res Soc* **59**: 1150–1172.

Freitas AA (2007). A review of evolutionary algorithms for data mining. In: Maimon O and Rokach L (eds). *Soft Computing for Knowledge Discovery and Data Mining*. Springer: Berlin, pp 61–93.

Giudici P and Castelo R (2003). Improving Markov Chain Monte Carlo model search for data mining. *Mach Learn* **50**(1–2):127–158.

Holte RC (1993). Very simple classification rules perform well on most commonly used datasets. *Mach Learn* **11**(1): 63–90.

Lee W, Stolfo SJ, Chan PK, Eskin E, Wei F, Miller M, Hershkop S, Junxin Z (2001). Real time data mining-based intrusion detection. In: *DARPA Information Survivability Conference & Exposition II*, DISCEX '01: Anaheim, CA, pp 89–100.

Lessmann S, Baesens B, Mues C and Pietsch S (2008). Benchmarking classification models for software defect prediction: a proposed framework and novel findings. *IEEE Trans Software Eng* **34**: 485–496.

Liu B (2006). *Web Data Mining: Exploring Hyperlinks, Content and Usage Data*. Springer: Berlin.

Lu Q and Getoor L (2003). Link-based classification. In: Fawcett T and Mishra N (eds.) *Proceeding of the Twentieth Conference on Machine Learning (ICML-2003)*. AAAI Press: Washington, DC.

Macskassy SA and Provost F (2007). Classification in networked data: a toolkit and a univariate case study. *J Mach Learn Res* **8**: 935–983.

Mangasarian OL (1965). Linear and nonlinear separation of patterns by linear programming. *Opns Res* **13**: 455–461.

Mangasarian OL, Street WN and Wolberg WH (1995). Breast cancer diagnosis and prognosis via linear programming. *Opns Res* **43**(4): 570–577.

Martens D and Baesens B (2009). Building acceptable classification models. *Ann Inform Syst* (forthcoming).

Martens D, De Backer M, Haesen R, Baesens B, Mues C, Vanthienen J (2006). Ant-based approach to the knowledge fusion problem. In: Dorigo M, Gambardella L, Birattari M, Martinoli A, Poli R and Stützle T (eds.) *Lecture Notes in Computer Science*, ANTS Workshop 2006, Brussels, Belgium, pp 84–95, September. Springer: Berlin.

Martens D, Baesens B, Van Gestel T and Vanthienen J (2007a). Comprehensible credit scoring models using rule extraction from support vector machines. *Eur J Opl Res* **183**(3): 1466–1476.

Martens D, De Backer M, Haesen R, Vanthienen J, Snoeck M and Baesens B (2007b). Classification with Ant Colony Optimization. *IEEE Trans Evolut Comput* **11**(5): 651–665.

Olafsson S (2006). Introduction to operations research and data mining. *Comput Opns Res* **33**(11): 3067–3069.

Padmanabhan B and Tuzhilin A (2003). On the use of optimization for data mining theoretical interactions and eCRM opportunities. *Mngt Sci* **49**(10): 1327–1343.

Seow HY and Thomas LC (2007). To ask or not to ask: that is the question. *Eur J Opl Res* **183**: 1513–1520.

Thomas LC, Edelman DB and Crook JN (2002). *Credit Scoring and Its Applications*. Society for Industrial Mathematics: Philadelphia, PA.

Van der Aalst WMP and Weijters AJMM (2004). Process mining: a research agenda. *Comput Ind* **53**(3): 231–244.

Van Gestel T, Baesens B, Van Dijcke P, Suykens J, Garcia J and Alderweireld T (2005). Linear and nonlinear credit scoring by combining logistic regression and support vector machines. *J Credit Risk* **1**(4).

Witten IH and Frank E (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers Inc.: San Francisco, CA.

Yang J and Olafsson S (2005). Optimization based feature selection with adaptive instance sampling. *Comput Opns Res* **33**(11): 3088–3106.

Received June 2008;
accepted August 2008 after one revision