Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/02/jls.htm
2015 Vol. 5 (S2), pp. 874-882/Kheirinejad and Mohammadjafari
Research Article

PROVIDING A METHOD TO DETERMINE AND EVALUATE RULES HAVE LED TO PROJECTS' TIME DELAY USING INTEGRATED APPROACH OF DATA MINING AND DATA ENVELOPMENT ANALYSIS

***Mojtaba Kheirinejad and Marjan Mohammadjafari**
Department of Industrial Engineering, Kerman branch, Islamic Azad University, Kerman, Iran
**Author for Correspondence*

ABSTRACT
Implementing infrastructure plans in each area will cause enhancement in economic condition and commercial transactions in that area. Therefore, it needs heavy budget and long lifetime. So, delay in these kinds of projects will cause non-compensable harms to country's economy. On the other hand, by increasing data volume, new techniques are being used to analyze data. Anticipating time delay of civil projects and determining factors and features which cause time delay in these projects can have a considerable assistance to managers of these projects in order to have an appropriate view toward probable delays before setting up the project. In this paper, we have determined rules which will cause time delay more than one year in the civil projects of Tehran municipality using association rules algorithms of data mining. These rules have been determined using Apriori algorithm. Since, in evaluating rules, more than one criterion is important (support and confidence), Data Envelopment Analysis has been used to evaluate rules. Results obtained from ranking these rules can be used to predict time delays in many of civil projects especially in Iran municipalities.

Keywords: *Data Mining, Projects' Time Delay, Association Rules, Efficiency, Data Envelopment Analysis Model*

INTRODUCTION
One of the main objectives of data mining is to produce interesting rules with respect to user's point of view.
This user is not assumed to be a data mining expert (Lenca *et al.*, 2008). The problem of discovering association rules has received considerable research attention and several fast algorithms. Mining association rules have been developed (Srikant *et al.*, 1997). Using these techniques, various rules may be obtained and only a small number of these rules may be selected for implementation due, at least in part, to limitations of budget and resources (Chen, 2007).
According to Liu *et al.,* (2000) the interestingness issue has long been identified as an important problem in data mining. It refers to finding rules that are interesting/useful to the user, not just any possible rule. Indeed, there exist some situations that make necessary the prioritization of rules for selecting and concentrating on more valuable rules due to the number of qualified rules (Tan and Kumar, 2000) and limited business resources (Choi *et al.*, 2005).
According to Chen (2007), selecting the more valuable rules for implementation increases the possibility of success in data mining.
In this dissertation, we intend to determine and evaluate rules which have led to projects' time delay. To do this, civil projects' managers don't know which factors will cause a civil project to encounter delay and also don't know if a project is fitted in more than one rule, time delay which has been predicted by each of rules is more suitable.
Therefore, in this paper, by determining association rules based on the formed database and ranking them, we intend to anticipate probable time delay of projects. For this purpose, it is required to compare the new project with the rules extracted based on the data envelopment analysis models priorities after determining association rules. If it is fitted in an especial rule, it can be articulated that time delay relevant to that project is equal to delay of that rule.

MATERIALS AND METHODS

Association Rules

An association rule is a phrase like $X \rightarrow Y$ where X is a set of items and Y is just one item. Association rules techniques are an initial approach of data discovery which are often applied for vary massive data set. One example is relevant to the data related to one grocery's market basket (Chen *et al.*, 2005). Exploration of association rules provides valuable information for evaluating important correlations. These rules have been applied in the various fields including medical science (Exarchos *et al.*, 2006) and identifying fraud in the medical insurance. Commercial usage of market basket analysis also includes guarantee analysis (Agrawal *et al.*, 1993).

The most famous method is inductive algorithm. A number of methods have been initiated to identify market basket relations effectively. Agrawal and Srikant have proposed an inductive algorithm which have enhanced for many times. Initial methods have been described by Hip *et al.*, Vaitchev *et al.*, proposed a method for Frequent Item set Generation which has been followed by compact rules set and low storage requirement (Vitchev *et al.*, 2002). A great number of algorithms have been proposed to explore association rules in massive databases.

Most of them, such as Apriori algorithm identify correlations between transactions including explicit features which are using binary values. Some of data mining approaches are applying weighted association rules for binary values (Cai *et al.*, 1998) or are using time intervals (Hou *et al.*, 2001). Mild and Ratrer have recommended more calculative approaches.

In correlation, relation between especial elements in one data transaction with other elements of the same transaction is used to predict patterns. For example, a customer who purchases laptop (X), in 60 percent of cases, he is also buying mouse (Y). This pattern in 5/6 percent of laptop purchase ceases is occurring. Association rule in this case is expressed as the following: X implies Y where confidence factor is equal to 60 percent and support factor is equal to 5/6 percent. When support and confidence factors are expressed by verbal variables such as "up" and "down", association rule can be articulated in the form of fuzzy logic as the following: "where support factor is low, X is highly implying Y". In the most of qualitative variables, fuzzy association in data mining is an efficient and required technique.

By fast growth of data in databases of commercial enterprises, extracting important information is getting harder and harder. KDD process can be helpful in detecting efficient; equal and beneficial unknown patterns (Frawley *et al.*, 1991). Data mining plays an important role in KDD process and applies specific patterns for extracting required knowledge or favorite pattern from current databases for a specific goal. Most of the past studies have concentrated on ranking features. Data transfer in the real applications often includes quantitative features, so that data mining algorithms have also been recommended for the quantitative values.

These algorithms obtain association rules by classifying features' areas; combining adjacent parts and eventually transforming problem to the binary form (Srikant and Agrawal, 1996).

Extracting fuzzy association rules for the quantitative values have been addressed by a great deal of researchers. Most of proposed methods are based on Apriori algorithm (Kouk and Wong, 1998). These researchers have assumed all of features or linguistic (verbal) words as uniform. Of course, in the real usages, users are probably interested in the rules with more application. Reducing Minsup and Minconf for obtaining rules including high applied words is not desirable, because algorithm efficiency will be decreased and undesirable rules are simultaneously being generated (Lander *et al.*, 2003). Extracting weighted quantitative association rules based on fuzzy logic has been presented by Gyenesei using 2 different definitions from weighted support (with and without normalization) (Gyenesei, 2000).

Data Envelopment Analysis Technique

Charnes, Cooper, Rhodes model (CCR) has been formed based on an interesting program from mathematic linear programing which measures relative efficiency of Decision Making Units (DMU) with multiple inputs and outputs (Charnes *et al.*, 1978). Relative efficiency has been defined as the ratio of sum of weighted outputs on sum of weighted inputs.

Relative model of DEA for the m_{th} decision making unit has been displayed as the following:

$$\text{Max } E_m = \frac{\sum_{j=1}^{J} v_{jm}\, y_{jm}}{\sum_{i=1}^{I} u_{im}\, x_{im}}$$

Subject to:

$$0 \le \frac{\sum_{j=1}^{J} v_{jm}\, y_{jn}}{\sum_{i=1}^{I} u_{im}\, x_{in}} \le 1 \qquad n = 1,2,3,\ldots,N;$$

$$u_{im}, v_{jm} \ge 0 \qquad i=1,2,\ldots,I;\ j=1,2,\ldots,J;$$

Here, Y is output value of each decision making unit and X is input value of each decision making unit. For example, Yim is the Jth output rate of Mth DMU and Xim is the Ith input rate of Mth DMU (N is the number of decision making units).

To solve this model due to this issue that the purpose is maximizing efficiency value, problem can be solved by 2 methods:

1- According to output oriented nature: In this conception, to calculate efficiency, by holding inputs constant, it is tried to maximize outputs.

2- According to input oriented nature: In this method, to calculate efficiency, by holding outputs constant, it has been tried to minimize inputs.

For example, by setting the denominator equal to one and using model 1, we are obtaining an output oriented CCR:

$$\text{Max } Z = \sum_{j=1}^{J} v_{jm}\, y_{jm}$$

Subject to:

$$\sum_{i=1}^{I} u_{im}\, x_{im} = 1$$

$$\sum_{j=1}^{J} v_{jm}\, y_{jn} - \sum_{i=1}^{I} u_{im}\, x_{in} \le 0$$

$$n = 1,2,3,\ldots,N;\ u_{im}, v_{jm} \ge \varepsilon \qquad i = 1,2,\ldots,I; \qquad j = 1,2,\ldots,J;$$

Efficiency digit (Z) is changing from 0 to 1 and if efficiency of investigated unit is equal to 1, this unit is relatively efficient. Otherwise, this unit is inefficient. When DMU's efficiency will be 1 for more than one unit, to obtain DMUs' ranks, Anderson-Peterson (AP) model will be employed. In Anderson-Peterson (AP) model, by removing constraint related to that DMU, it is allowed to have efficiency more than one. In this model, each DMU which obtains the greater efficiency digit, among efficient units, has a higher performance (Edirisinghe and Zhang, 2007).

MATERIALS AND METHODS

Research Methodology

Experimental Design

Steps to do this paper have been detailed in 4 main phases:

Phase 1: Forming database

Step 1: Determining problem constraints

In this step, the main constraint of the problem is determined and in respect with problem objectives, respective information has to be extracted and general figure of database is determined.

Step 2: Determining records

In this stage, civil projects which we intend to examine and predict their delay time have to be determined. In respect with experts' opinion and information existed in civil sector of Tehran municipality and according to comments given by civil experts, projects which have been started up between 2007 to 2009 have been randomly chosen.

Step 3: Determining features

In this step, initial features about each of projects including quantitative and qualitative cases will be regulated which are eventually as the following:

Cost, number of predictable oppositions, passed budget No, grade of the main contractor, number of WBS activities, duration of project implementation, state or private main contractor, percent of FS processes to the total activities, kind of contract, percent of the first payment to the total payments and the year of project implementation.

Step 4: Determining class (attachment) in data base
In this step, with respect to experts' opinions, data bases have been classified in many groups such as projects which have been done in due time.7 classes will be considered in data base which are:
Until 3 months less than predicted time, predicted time, until 3 months delay, 3 months to 6 months delay, 6 months to one year delay, one year to 2 years delay and more than 2 years delay.
Phase 2: Validation of data base
For validation of data base in this paper, 10-fold Crusade validation is being used. If data base is suitable, accuracy of classification algorithms on data base has to be at least 80%.
Phase 3: Determining data base rules
In this step, by employing association algorithms and especially Apriori algorithm, it is approached to regulate database rules and support and confidence values of each of rules which indicate importance and correctness of rules.
Phase 4: Ranking and determining efficiency of obtained rules
In this step, DEA model for obtained rules has been detailed as the following:
Units: Obtained rules
Input: We consider constant digit such as 1.
Output: Support and confidence values of each of obtained rules (Tollu *et al.*, 2009).
By forming and solving model, database rules will be evaluated and efficiency and consequently association rules ranking will be obtained. Its' model is being formed as the following:

$$\text{Max } Z = \sum_{j=1}^{J} v_{jm}\, y_{jm}$$

Subject to:

$$\sum_{i=1}^{I} u_{im}\, x_{im} = 1$$

$$\sum_{j=1}^{J} v_{jm}\, y_{jn} - \sum_{i=1}^{I} u_{im}\, x_{in} \leq 0$$

$$n = 1,2,3,\ldots,N; \quad u_{im}, v_{jm} \geq \varepsilon \qquad i = 1,2,..,I; \qquad j = 1,2,..,J;$$

Efficiency digit (Z) is changing from 0 to 1 and if efficiency of studied unit is equal to 1, this unit is relatively efficient. Otherwise, this unit is accounted inefficient. When DMU's efficiency will be 1 for more than one unit, to obtain DMUs ranks, Anderson-Peterson (AP) model will be used.
Also, data collection in this paper has been done through forming databases of Tehran municipality civil projects and so, to determine association rules of database, Clementine software has been used. In the case of validation of data base, WEKA software and in the case of determining efficiency of association rules, DEAP software has been applied.

RESULTS AND DISCUSSION

1) Determining Records

In this step, projects which we want to examine and anticipate their validation have to be determined. In respect with experts' opinion and information existed in civil sector of Tehran municipality, 200 projects were randomly chosen and their information were used. Their number is also appropriate for data mining analysis.

2) Determining Features

In this step, initial features about each project which are including qualitative and quantitative cases are determined which have been detailed as the following:

1. Cost (more than 10 billion, between 5 to 10 billion, between 2 to 5 billion, less than 2 billions)
2. Number of predictable oppositions
3. Passed budget No (between 1818 to 1939)
4. Grade of the main contractor (first, second, third, fourth grade)
5. Number of WBS activities
6. Duration of project implementation
7. State or private main contractor
8. Percent of FS processes to the total activities (more than 50%, less than 50 %)

Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/02/jls.htm
2015 Vol. 5 (S2), pp. 874-882/Kheirinejad and Mohammadjafari
Research Article

9. Kind of contract (Auction, Price List, Lump sum)
10. Percent of the first payment to the total payments (less than 10 percent, between 10% to 20%, more than 20 percent)
11. The year of project implementation (2007,2008, 2009)

3) Determining Database Class

In this stage, the most important criterion (projects' delay rate) which municipality is trending to anticipate in the case of each project is determined. This time delay is categorized to 7 categories as the following:

1. Until 3 months less than predicted time (It has been coded by H letter in database)
2. Predicted time (It has been coded by D letter in database)
3. Until 3 months delay (It has been coded by B letter in database)
4. Until 6 months delay (It has been coded by C letter in database)
5. 6 months to one year delay (It has been coded by F letter in database)
6. One year to 2 years delay (It has been coded by E letter in database)
7. More than 2 years delay (It has been coded by G letter in database)

To validate database, we take an action to rank data using multiple algorithms. Table 1 indicates accuracy of implementing each of algorithms on database. This is done by WEKA software. In principle, this table indicates total accuracy of ranking algorithms in predicting projects' time delay.

Table 1: Determining validation of time delay database using 10-fold cross-validation

Algorithm	Database accuracy	Algorithm	Database accuracy
Simple Cart	%86	Meta Filtered Classifier	%78.5
J48	%84	END	%89.5
Decision Table	%69.5	Bagging	%80.5

Since, municipality trends to identify rules and thereby projects which have high time delay, in this step, data base rules are determined using Apriori algorithm. In table 2, rules and criteria of support and confidence for each of rules have been shown.

Table 2: Raw rules of database

	Consequent	Antecedent	Support %	Confidence %
1	C12 = E	C2 > 42	%33	%95
2	C12 = E	C10=I and C7=G and C1=A	%19	%100
3	C12 = E	C8=W and C6>3	%55	%90
4	C12 = E	C5>3000 and C10=I and C4=R	%23	%75
5	C12 = E	C2>29 and C7=G	%46	%100
6	C12 = E	C9=C and C10=I	%33	%85
7	C12 = E	C1=A and C7=G and C10=I	%17	%66
8	C12 = E	C7=N and C9=C	%15	%55
9	C12 = E	C5>2500 and C2>25	%39	%82
10	C12 = G	C1=A and C8=W	%40	%75
11	C12 = G	C8=W and C9=B	%50	%65
12	C12 = G	C4=R and C2>24	%45	%90
13	C12 = G	C10=I and C9=C	%22	%70
14	C12 = G	C9=C and C6>2	%18	%44
15	C12 = G	C2>20 and C10=I	%42	%76
16	C12 = G	C11=S and C7=N and C5>2300	%31	%49
17	C12 = G	C11=S and C2>20 C9=C	%19	%50
18	C12 = G	C11=S and C1=A and C8=W	%18	%50
19	C12 = G	C11=S and C9=B	%29	%65

Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/02/jls.htm
2015 Vol. 5 (S2), pp. 874-882/Kheirinejad and Mohammadjafari
Research Article

After rules codding, final results have been shown in table

Table 3: Codded rules of database

	Consequent	Antecedent	Support %	Confidence %
1	1 to 2 years delay	Number of oppositions more than 42	%33	%95
2	1 to 2 years delay	Percent of the first payment is less than 10% and private main contractor and high project cost	%19	%100
3	1 to 2 years delay	Percent of FS predicted processes is more than 50% and project predicted duration is more than 3 years	%55	%90
4	1 to 2 years delay	Number of WBS activities is more than 3000 and percent of the first payment is less than 10% and grade of the main contractor is 2	%23	%75
5	1 to 2 years delay	Number of oppositions is more than 29 and private main contractor	%46	%100
6	1 to 2 years delay	Kind of contract is price list and the first payment percent is 10%	%33	%85
7	1 to 2 years delay	High project cost and private main contractor and percent of the first payment is less than 10%	%17	%66
8	1 to 2 years delay	State main contractor and kind of contract is price list	%15	%55
9	1 to 2 years delay	Number of WBS activities is more than 2500 and number of oppositions is more than 25	%39	%82
10	Delay more than 2 years	High project cost and percent of FS predicted processes is more than 50%	%40	%75
11	Delay more than 2 years	Percent of FS predicted processes is more than 50% and kind of contract is lump sum	%50	%65
12	Delay more than 2 years	Grade of the main contractor is 2 and number of oppositions is more than 24	%45	%90
13	Delay more than 2 years	Percent of the first payment is less than 10% and kind of contract is price list	%22	%70
14	Delay more than 2 years	Kind of contract is price list and project duration is more than 2 years	%18	%44
15	Delay more than 2 years	Number of oppositions is more than 24 and percent of the first payment is less than 10%	%42	%76
16	Delay more than 2 years	Project start time is 2009 and state main contractor and number of WBS activities is more than 2300	%31	%49
17	Delay more than 2 years	Project start time is 2009 and number of oppositions is more than 20 and kind of contract is price list	%19	%50
18	Delay more than 2 years	Project start time is 2009 and project cost is high and private main contractor	%18	%50
19	Delay more than 2 years	Project start time is 2009 and kind of contract is lump sum	%29	%65

Eventually, in this step, by founding DEA model and solving it, efficiency of each of rules and thereby rules ranking are obtained. Forming DEA model is as the following:
Units: Units are rules extracted from delay data base
Inputs: We consider a constant digit such as 1

Output: Rate of support and confidence is considered for each of rules
Therefore, solving model which displays efficiency and rank of each of rules has been shown in table 4.

Table 4: Rules' efficiency and ranks

	Consequent	Antecedent	Efficiency	Rank
1	1 to 2 years delay	Number of oppositions more than 42	0.95	4
2	1 to 2 years delay	Percent of the first payment is less than 10% and private main contractor and high project cost	1	1
3	1 to 2 years delay	Percent of FS predicted processes is more than 50% and project predicted duration is more than 3 years	1	1
4	1 to 2 years delay	Number of WBS activities is more than 3000 and percent of the first payment is less than 10% and grade of the main contractor is 2	0.75	11
5	1 to 2 years delay	Number of oppositions is more than 29 and private main contractor	1	1
6	1 to 2 years delay	Kind of contract is price list and the first payment percent is 10%	0.85	7
7	1 to 2 years delay	High project cost and private main contractor and percent of the first payment is less than 10%	0.66	13
8	1 to 2 years delay	State main contractor and kind of contract is price list	0.55	16
9	1 to 2 years delay	Number of WBS activities is more than 2500 and number of oppositions is more than 25	0.83	8
10	Delay more than 2 years	High project cost and percent of FS predicted processes is more than 50%	0.79	10
11	Delay more than 2 years	Percent of FS predicted processes is more than 50% and kind of contract is lump sum	0.91	6
12	Delay more than 2 years	Grade of the main contractor is 2 and number of oppositions is more than 24	0.93	5
13	Delay more than 2 years	Percent of the first payment is less than 10% and kind of contract is price list	0.70	12
14	Delay more than 2 years	Kind of contract is price list and project duration is more than 2 years	0.44	19
15	Delay more than 2 years	Number of oppositions is more than 24 and percent of the first payment is less than 10%	0.81	9
16	Delay more than 2 years	Project start time is 2009 and state main contractor and number of WBS activities is more than 2300	0.56	15
17	Delay more than 2 years	Project start time is 2009 and number of oppositions is more than 20 and kind of contract is price list	0.50	17
18	Delay more than 2 years	Project start time is 2009 and project cost is high and private main contractor	0.50	17
19	Delay more than 2 years	Project start time is 2009 and kind of contract is lump sum	0.65	14

Conclusion

In this paper, we have determined rules and evaluate rules which have led to projects' time delay. Based on results obtained from components of Apriori algorithm and evaluating rules using DEA model, rules have been evaluated.

The most important rules (which have reached efficiency 1) are:

1. Percent of the first payment is less than 10%and private main contractor and high project costs which are led to delay between 1 year and 2 years.
2. Percent of FS predicted processes are more than 50% and project predicted duration is more than 3 years which are led to delay between 1 year and 2 years.

3. Number of oppositions is more than 29 and private main contractor which are led to delay between 1 year and 2 years.

In the case of aforementioned rules, approaches have to be taken in order to prevent from factors which caused time delay. By studying association rules, the first point existed in their ranking is that rules which are led to time delay more than 2 years have gained an important place in the ranking. That is, in time delay more than 2 years, there are features which are not in database. So, especial reasons will cause these delays. Also, by investigating rules which have reached efficiency 1, it can be observed that percent of the first payment has been mentioned less than 10%. Thus, the first payment is accounted an important factor in projects' delay. Also by studying rules, it can be articulated that number of oppositions is observed in the rules which have gained high efficiency. So, this factor also has a high importance. In regard with enhancement of results obtained from this research, it is recommended to other researchers to use more developed models of DEA in ranking rules. Also, they can use more criteria in evaluating rules. Also, they can define weight for criteria and involve those weights in calculations.

REFERENCES

Agrawal R, Imielinski T and Swami A (1993). Mining association between sets of items in massive database. *International Proceedings of the ACM-SIGMOD International Conference on Management of Data* 207–216.

Agrawal R and Srikant R (1994). Fast algorithms for mining association rules. *Proceedings of the International Conference on Very Large Data Bases* 407–419.

Amin Gholam R and Toloo M (2007). Finding the most efficient DMUs in DEA: An improved integrated model. *Computers & Industrial Engineering* **52** 71–77.

Banker RD, Charnes A and Cooper WW (1984). Some models for estimating technical and scale inefficiency in data envelopment analysis. *Management Science* **30** 1078–1092.

Camanho AS and Dyson RG (2005). Cost efficiency measurement with price uncertainty: A DEA application to bank branch assessments. *European Journal of Operational Research* **161** 432–446.

Charnes A, Cooper WW and Rhodes E (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research* **2** 429–444.

Chen MC (2007). Ranking discovered rules from data mining with multiple criteria by data envelopment analysis. *Expert Systems with Applications* **33** 1110–1116.

Chen X, Skully M and Brown K (2005). Banking efficiency in China: Application of DEA to pre- and post-deregulation eras: 1993–2000. *China Economic Review* **16** 229–245.

Choi DH, Ahn BS and Kim SH (2005). Prioritization of association rules in data mining: Multiple criteria decision approach. *Expert Systems with Applications* **29** 867–878.

Cook WD and Kress M (1990). A data envelopment model for aggregating preference rankings. *Management Science* **36** 1302–1310.

Edirisinghe NCP and Zhang X (2007). Generalized DEA model of fundamental analysis and its application to portfolio optimization. *Journal of Banking & Finance* **31** 3311–3335.

Emrouznejad A, Tavares G and Parker B (2007). A bibliography of data envelopment analysis (1978–2003). *Socio-Economic Planning Sciences* **38** 159–229.

Ertay T, Ruan D and Tuzkaya UR (2006). Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing systems. *Information Sciences* **176** 237–262.

Johnes J (2006). Measuring teaching efficiency in higher education: An application of data envelopment analysis to economics graduates from UK Universities 1993. *European Journal of Operational Research* **174** 443–456.

Lenca P, Meyer P, Vaillant B and Lallich S (2008). On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. *European Journal of Operational Research* **184** 610–626.

Liu ST (2008). A fuzzy DEA/AR approach to the selection of flexible manufacturing systems. *Computers & Industrial Engineering* **54** 66–76.

Liu B, Hsu W, Chen S and Ma Y (2000). Analyzing the subjective interestingness of association rules. *IEEE Intelligent Systems* **15** 47–55.
Mannino M, Hong SN and Choi IJ (2008). Efficiency evaluation of data warehouse operations. *Decision Support Systems* **44** 883–898.
Marakas GM (2004). *Decision Support Systems in the 21st Century*, second edition (Prentice-Hall of India).
Ng RT, Lakshmanan LVS, Han J and Pang A (1998). Exploratory mining and pruning optimizations of constrained association rules. *In Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD-98* 13–24.
Obata T and Ishii H (2003). A method for discriminating efficient candidates with ranked voting data. *European Journal of Operational Research* **151** 233–237.
Olafson S, Li X and Wu S (2008). Operations research and data mining. *European Journal of Operational Research* **187** 1429–1448.
Shafer SM and Byrd TA (2000). A framework for measuring the efficiency of organizational investments in information technology using data envelopment analysis. *Omega* **28** 125–141.
Srikant R, Vu Q and Agrawal R (1997). Mining association rules with item constraints. *In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, KDD-97 67–73.
Tan PN and Kumar V (2000). Interestingness measures for association patterns: A perspective. In KDD 2000 workshop on post processing in machine learning and data mining, Boston, MA, August.
Toloo M and Nalchigar S (2009). A new integrated DEA model for finding most BCC– efficient DMU. *Applied Mathematical Modeling* **33** 597–604.