Expert Systems with Applications 36 (2009) 8503–8508

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A new method for ranking discovered rules from data mining by DEA

Mehdi Toloo [a,*], Babak Sohrabi [b], Soroosh Nalchigar [b]

[a] Department of Mathematics, Islamic Azad University of Central Tehran Branch, Tehran, Iran
[b] Department of Information Technology Management, Faculty of Management, University of Tehran, Tehran, Iran

ARTICLE INFO

Keywords:
Data mining
Data envelopment analysis
Association rule
Interestingness

ABSTRACT

Data mining techniques, extracting patterns from large databases have become widespread in business. 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. Evaluating and ranking the interestingness or usefulness of association rules is important in data mining. This paper proposes a new integrated data envelopment analysis (DEA) model which is able to find most efficient association rule by solving only one mixed integer linear programming (MILP). Then, utilizing this model, a new method for prioritizing association rules by considering multiple criteria is proposed. As an advantage, the proposed method is computationally more efficient than previous works. Using an example of market basket analysis, applicability of our DEA based method for measuring the efficiency of association rules with multiple criteria is illustrated.

1. Introduction

With the rapid growth of databases in many modern enterprises data mining has become an increasingly important approach for data analysis. In recent years, the field of data mining has seen an explosion of interest from both academia and industry (Olafson, Li, & Wu, 2008). Increasing volume of data, increasing awareness of inadequacy of human brain to process data and increasing affordability of machine learning are reasons of growing popularity of data mining (Marakas, 2004).

One of the main objectives of data mining is to produce interesting rules with respect to some user's point of view. This user is not assumed to be a data mining expert, but rather an expert in the field being mined (Lenca, Meyer, Vaillant, & Lallich, 2008). The problem of discovering association rules has received considerable research attention and several fast algorithms for mining association rules have been developed (Srikant, Vu, & Agrawal, 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, Hsu, Chen, and Ma (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 & Kumar, 2000) and limited business resources (Choi, Ahn, & Kim, 2005). According to Chen (2007), selecting the more valuable rules for implementation increases the possibility of success in data mining. For example, in market basket analysis, understanding which products are usually bought together by customers and how the cross-selling promotions are beneficial to sellers both attract marketing analysts. The former makes sellers to provide appropriate products by considering the customers' preferences, and the later allows sellers to gain increased profits by considering the sellers' profits. Customers' preferences can be measured based on support and confidence in association rules. On the other hand, seller profits can be assessed using domain related measures such as sale profit and cross-selling profit associated with the association rules (Chen, 2007).

In previous studies dealing with the discovery of subjectively interesting association rules, most approaches require manual input or interaction by asking users to explicitly distinguish between interesting and uninteresting rules (Chen, 2007). Srikant et al. (1997) presented three integrated algorithms for mining association rules with item constraint. Moreover, Lakshmanan et al. (1998) extended the approach presented by Srikant et al. to consider much more complicated constraints, including domain, class, and SQL-style aggregate constraints. Liu et al. (2000) presents an Interestingness Analysis System (IAS) to help the user identify interesting association rules. In their proposed method, they consider two main subjective interestingness

* Corresponding author. Tel.: +98 9122390003.
E-mail addresses: m_toloo@yahoo.com (M. Toloo), Bsohrabi@ut.ac.ir (B. Sohrabi), Nalchigar@ut.ac.ir (S. Nalchigar).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.10.038

measures, unexpectedness and actionability. Choi et al. (2005), using analytic hierarchy process (AHP) presented a method for association rules prioritization which considers the business values which are comprised of objective metric or managers' subjective judgments. They believed that proposed method makes synergy with decision analysis techniques for solving problems in the domain of data mining. Nevertheless this method requires large number of human interaction to obtain weights of criteria by aggregating the opinions of various managers. Chen (2007) developed their work and proposed a data envelopment analysis (DEA) based methodology for ranking association rules while considering multiple criteria. During his ranking procedure, he uses a DEA model, proposed by Cook and Kress (1990), to identify efficient association rules. Then, he applies another DEA model, developed by Obata and Ishii (2003), to discriminate efficient association rules. It should be noted that his proposed method requires the first model to be solved for all DMUs and the second model to be solved for efficient DMUs. As a drawback, this approach requires considerable number of linear programming (LP) models to be solved. Moreover, this approach includes some redundant computations and considerations. Therefore there is a need for a method which is able to rank association rules more efficiently. This paper tries to fill the gap by developing a new integrated DEA model which is able to identify most efficient association rule by solving only one mixed integer linear programming (MILP) and proposing a new method for ranking association rules with multiple criteria. The proposed method is computationally efficient and helps user to get fast results.

DEA is a non-parametric linear programming based technique for measuring the relative efficiency of a set of similar units, usually referred to as decision making units (DMUs). Because of its successful application and case studies, DEA has gained too much attention and widespread use by business and academy researchers. Evaluation of data warehouse operations (Mannino, Hong, & Choi, 2008), selection of flexible manufacturing system (Liu, 2008), assessment of bank branch performance (Camanho & Dyson, 2005), examining bank efficiency (Chen, Skully, & Brown, 2005), analyzing firm's financial statements (Edirisinghe & Zhang, 2007), measuring the efficiency of higher education institutions (Johnes, 2006), solving facility layout design (FLD) problem (Ertay, Ruan, & Tuzkaya, 2006) and measuring the efficiency of organizational investments in information technology (Shafer & Byrd, 2000) are examples of using DEA in various areas. Similar to Chen (2007), this paper uses DEA as a post-processing approach. After the rules have been discovered from the association rule mining algorithms, DEA is used to rank those discovered rules based on the specified criteria. The main contribution of this paper is to develop a new integrated DEA model for finding most efficient association rule (by solving only one LP) and to propose a new method for ranking discovered association rules of data mining.

The rest of this paper is organized as follows. In section 2, briefly, association rule is described. Section 3, presents DEA models and section 4 discuss a previous method for ranking association rules. Our proposed method is introduced in section 5. Then, applicability of our method is illustrated in section 6. The paper closes with some concluding remarks in section 7.

2. Association rule

Association rule mining, introduced by Agrawal, Imielinski, and Swami (1993), has been widely used from traditional business applications such as cross-marketing, attached mailing, catalog design, loss-leader analysis, store layout, and customer segmentation to e-business applications such as the renewal of web pages and web personalization (Choi et al., 2005).

Given a set of transactions, where each transaction is a set of literals (called items), an association rule is an expression of the form $X \Rightarrow Y$, where X and Y are sets of items. The intuitive meaning of such a rule is that transactions of the database which contains X to contain Y. An example of an association rule is: "40% of transactions that contain bread also contain milk; 3% of all transactions contain both these items". Here 40% is called the confidence of the rule, and 3% the support of the rule. It should be noted that associations may include any number of items on either side of the rule. An efficient algorithm is required that restricts the search space and checks only a subset of all association rules, yet does not miss important rules (Chen, 2007). Many algorithms can be used to discover association rules from data to extract useful patterns. Apriori algorithm is one of the most widely used and famous techniques for finding association rules (Agrawal & Srikant, 1994; Agrawal et al., 1993). Apriori operates in two phases. In the first phase, all itemsets with minimum support (*frequent* itemsets) are generated. This phase utilizes the downward closure property of support. In other words, if an itemset of size k is a frequent itemset, then all the itemsets below $(k-1)$ size must also be frequent itemsets. Using this property, candidate itemsets of size k are generated from the set of frequent itemsets of size $(k-1)$ by imposing the constraint that all subsets of size $(k-1)$ of any candidate itemset must be present in the set of frequent itemsets of size $(k-1)$. The second phase of the algorithm generates rules from the set of all frequent itemsets.

Association rule mining is a popular technique for market basket analysis, which typically aims at finding buying patterns for supermarket, mail-order and other customers. By mining association rules, marketing analysts try to find sets of products that are frequently bought together, so that certain other items can be inferred from a shopping cart containing particular items. Association rules can often be used to design marketing promotions, for example, by appropriately arranging products on a supermarket shelf and by directly suggesting to customers items that may be of interest (Chen, 2007).

3. DEA models

DEA is a data-oriented approach for relatively evaluating the performance of a group of entities referred to DMUs. It was introduced by Charnes, Cooper, and Rhodes (1978) based on Farrell's pioneering work. They generalized the single-output to single-input ratio definition of efficiency to multiple inputs and outputs. In their original DEA model, Charnes, Cooper and Rhodes (CCR model) proposed that the efficiency of a DMU can be obtained as the maximum of a ratio of weighted outputs to weighted inputs, subject to the condition that the same ratio for all DMUs must be less than or equal to one. The DEA model must be run n times, once for each unit, to get the relative efficiency of all DMUs. The envelopment in CCR is constant returns to scale meaning that a proportional increase in inputs results in a proportionate increase in outputs. Banker, Charnes, and Cooper (1984) developed the BCC model to estimate the pure technical efficiency of decision making units with reference to the efficient frontier. It also identifies whether a DMU is operating in increasing, decreasing or constant returns to scale. So CCR models are a specific type of BCC models.

Assume that there are n DMUs, $(\mathrm{DMU_j} : j = 1, 2, \ldots, n)$ which consume m inputs $(\mathbf{x}_i : i = 1, 2, \ldots, m)$ to produce s outputs $(\mathbf{y}_r : r = 1, 2, \ldots, s)$. The BCC input oriented (BCC-I) model evaluates the efficiency of $\mathrm{DMU_o}$, DMU under consideration, by solving the following linear program:

$$
\begin{aligned}
&\max \sum_{r=1}^{s} u_r y_{rj} - u_0 \\
&s.t. \\
&\sum_{i=1}^{m} w_i x_{io} = 1 \\
&\sum_{r=1}^{s} u_r y_{rj} - u_0 - \sum_{i=1}^{m} w_i x_{ij} \le 0 \quad j = 1,2,\ldots,n \\
&u_0 \textit{ free} \\
&w_i \geqslant \varepsilon \quad i = 1,2,\ldots,m \\
&u_r \geqslant \varepsilon \quad r = 1,2,\ldots,s
\end{aligned}
\tag{1}
$$

where x_{ij} and y_{rj} (all nonnegative) are the inputs and outputs of the DMU_j, w_i and u_r are the input and output weights (also referred to as multipliers). x_{io} and y_{ro} are the inputs and outputs of DMU_o. Also, ε is non-Archimedean infinitesimal value for forestalling weights to be equal to zero.

New applications with more variables and more complicated models are being introduced (Emrouznejad, Tavares, & Parker, 2007). In many applications of DEA, finding the most efficient DMU is desirable. Amin and Toloo (2007) proposed an integrated model for finding most efficient DMU, as follows:

$$
\begin{aligned}
&M^* = \min M \\
&s.t. \\
&M - d_j \geqslant 0 \quad j = 1,2,\ldots,n \\
&\sum_{i=1}^{m} w_i x_{ij} \leqslant 1 \quad j = 1,2,\ldots,n \\
&\sum_{r=1}^{s} u_r y_{rj} - \textstyle\sum_{i=1}^{m} w_i x_{ij} + d_j - \beta_j = 0 \quad j = 1,2,\ldots,n \\
&\sum_{j=1}^{n} d_j = n - 1 \\
&0 \leqslant \beta_j \leqslant 1, \quad d_j \in \{0,1\} \quad j = 1,2,\ldots,n \\
&w_i \geqslant \varepsilon \quad i = 1,2,\ldots,m \\
&u_r \geqslant \varepsilon \quad r = 1,2,\ldots,s
\end{aligned}
\tag{2}
$$

where d_j as a binary variable represents the deviation variable of DMU_j. DMU_j is most efficient if and only if $d_j = 0$. The constraint $\sum_{j=1}^{n} d_j = n - 1$ forces among all the DMUs for only single most efficient unit. It should be noted that Model (2) is based on CCR model and identify most CCR-efficient DMU. Indeed, Model (2) is not applicable for situations in which DMUs operating in variable return to scale. To overcome this drawback, Toloo and Nalchigar (2009) proposed an integrated model which is able to find most BCC-efficient DMU.

4. Chen's proposed method

In this section, Chen (2007) proposed method for ranking association rules is discussed. In fact, his proposed method uses a DEA model, proposed by Cook and Kress (1990), for identifying efficient association rules. This model is as follows:

$$
\begin{aligned}
&\max \textstyle\sum_{j=1}^{k} \mathrm{w_j} v_{oj} \\
&s.t. \\
&\sum_{j=1}^{k} \mathrm{w_j} v_{ij} \leqslant 1, \quad i = 1,2,\ldots,m; \\
&w_j - w_{j+1} \geqslant d(j,\varepsilon) \quad j = 1,2,\ldots,k-1 \\
&w_k \geqslant d(k,\varepsilon)
\end{aligned}
\tag{3}
$$

where w_j denotes the weight of the jth place; v_{ij} represents the number of jth place votes of candidate $i (i = 1,2,\ldots,m,$ $j = 1,2,\ldots,k)$ and $d(\bullet,\varepsilon)$, known as the discrimination intensify function, is nonnegative and nondecreasing in ε and satisfies $d(\bullet,\varepsilon) = 0$.

Model (3) should be resolved for each candidate $o, o = 1,2,\ldots,m$. The resulting objective value is the preference score of candidate o. Because of the fact that DEA frequently generates several efficient candidates (Obata & Ishii, 2003), Chen's proposed method uses another DEA model, proposed by Obata and Ishii (2003), for discriminating efficient association rules. It should be noted that this model does not employ any information about inefficient candidates and should be solved only for efficient association rules.

The above-mentioned Chen's method, however, has the following properties:

- Chen's method requires computing v_{ij} from y_{ij} (jth outputs of ith association rule). Although, the algorithm of computing v_{ij} from y_{ij} is polynomial, it is time consuming. Identifying efficient association rules can be done through a more simple and efficient way.
- Result of Chen's method is immensely dependent on discrimination intensify function.
- Suppose that there are e efficient association rules which are obtained from Model (3). To rank e efficient units, Chen's method includes solving $(n + e)$ LPs.

In the next a new method for ranking association rules of data mining is proposed which ranks association rules more efficiently and does not include redundant computations.

5. Proposed model

In evaluation of association rules of data mining, the criteria such as support, confidence, itemset value and cross-selling profit are to be maximized and can be considered as outputs (Chen, 2007). To rank efficient association rules, Chen's used voting models, because these models consider only output data of units. Continuing previous works of Amin and Toloo (2007) and Toloo and Nalchigar (2009), in this section we propose a new integrated DEA model which is able to identify most CCR-efficient DMU by considering only outputs data of DMUs. The model proposes as

$$
\begin{aligned}
&M^* = \min M \\
&s.t. \\
&M - d_j \geqslant 0 \quad j = 1,2,\ldots,n \\
&\sum_{r=1}^{s} u_r y_{rj} + d_j - \beta_j = 1 \quad j = 1,2,\ldots,n \\
&\sum_{j=1}^{n} d_j = n - 1 \\
&0 \leqslant \beta_j \leqslant 1, d_j \in \{0,1\} \quad j = 1,2,\ldots,n \\
&w_i \geqslant \varepsilon \quad i = 1,2,\ldots,m \\
&u_r \geqslant \varepsilon \quad r = 1,2,\ldots,s
\end{aligned}
\tag{4}
$$

where d_j as a binary variable represents the deviation variable of DMU_j. DMU_j is most efficient if and only if $d_j = 0$. Similar to Amin and Toloo (2007) and Toloo and Nalchigar (2009), the main idea of Model (4) is trying to find only one most efficient DMU, but in situations in which DMUs are evaluated based on their outputs data. Indeed, Model (4) is a customized version of previous models. Hence, the following LP, which is a customized version of Amin and Toloo (2007) epsilon model, is proposed to determine the non-Archimedean epsilon:

$$\begin{aligned}&\varepsilon^* = \max \varepsilon\\&s.t.\\&\textstyle\sum_{r=1}^{s} u_r y_{rj} \leqslant 0 \quad j = 1,2,\ldots,n\\&u_r - \varepsilon \geqslant 0 \quad r = 1,2,\ldots,s\end{aligned} \tag{5}$$

In this section, using Model (4), we propose a new method for ranking DMUs. The proposed method, which is based on a simple idea, is described as follows:

Step 0: Let $T = \phi$ and e = number of DMUs to be ranked.
Step 1: Solve following model:

$$\begin{aligned}&M^* = \min M\\&s.t.\\&M - d_j \geqslant 0 \quad j = 1,2,\ldots,n\\&\sum_{r=1}^{s} u_r y_{rj} + d_j - \beta_j = 1 \quad j = 1,2,\ldots,n\\&\sum_{j=1}^{n} d_j = n - 1\\&d_j = 1 \quad \forall j \in \mathrm{T}\\&0 \leqslant \beta_j \leqslant 1, \quad d_j \in \{0,1\} \quad j = 1,2,\ldots,n\\&w_i \geqslant \varepsilon \quad i = 1,2,\ldots,m\\&u_r \geqslant \varepsilon \quad r = 1,2,\ldots,s\end{aligned} \tag{6}$$

Suppose in optimal solution $d_p^* = 0$.

Step 2: Let $\mathrm{T} = \mathrm{T} \cup \{p\}$.
Step 3: If $|\mathrm{T}| = e$, then stop; otherwise go to Step 1.

Indeed in Step 1 of proposed algorithm, a DMU is identified as most CCR-efficient unit. After entering this DMU to T in Step 2, in Step 3 if all DMUs are ranked, the algorithm finishes, otherwise it goes to next iteration. By continuing the iterations to e times, decision maker is able to rank DMUs by considering only outputs data of them.

6. Illustrative example

To show applicability of proposed method, an example of market basket data is adopted from Chen (2007). Association rules first are discovered by the Apriori algorithm, in which minimum support and minimum confidence are set to 1.0% and 10.0%, respectively. Forty-six rules then are identified and presented in Table 1. We are to rank association rules.

By solving Model (6) for data presented in Table 1 (with considering suitable value for epsilon) DMU_{18} is easily identified as most CCR-efficient association rule ($d_{18}^* = 0, d_{j\neq 18}^* = 1$). In second iteration of proposed method, a constraint $d_{18} = 1$ is added to model. This added constraint ensure that in second iteration of algorithm, DMU_{18} will not again identified as most efficient unit. By solving Model (6) in second iteration, optimal solution is ($d_{23}^* = 0, d_{j\neq 23}^* = 1$) which implies that DMU_{23} is second CCR-efficient association rule. By continuing this process user can rank all association rules. Table 2 presents results of ranking efficient rules in comparison to Chen's method.

It is notable that our proposed method rank efficient units by solving 11(= e) MILPs; however, Chen's proposed method solve 57($n + e$ = 46 + 11) LP models to rank efficient association rules. As another advantage, the proposed method is able to rank all association rules (by solving 45 MILPs); in contrast, Chen's method ranks only efficient rules.

In addition, instead of using Model (3) and computing v_{ij}, one can identify efficient association rules by adding a virtual input (whose value is equal to for all DMUs) to problem using basic

Table 1
Data of association rules.

Association rule number (DMU)	Support (%)	Confidence (%)	Itemset value	Cross-selling profit
1	3.87	40.09	337.00	25.66
2	1.42	18.17	501.00	11.63
3	2.83	17.64	345.00	11.29
4	2.34	30.83	163.00	19.73
5	2.63	23.90	325.00	15.30
6	1.19	55.65	436.00	35.61
7	1.19	47.42	598.00	30.35
8	1.19	15.70	436.00	52.91
9	1.19	10.82	598.00	36.45
10	1.19	12.32	436.00	20.08
11	1.19	12.32	598.00	40.04
12	3.87	38.08	337.00	103.97
13	1.18	15.09	710.00	41.19
14	2.44	15.22	554.00	41.56
15	2.14	28.21	372.00	77.02
16	2.51	22.81	534.00	62.26
17	1.19	50.92	436.00	139.02
18	1.19	45.25	598.00	123.52
19	1.19	11.70	436.00	43.54
20	1.19	11.70	598.00	62.50
21	1.42	13.99	501.00	61.16
22	1.18	12.23	710.00	53.45
23	1.50	13.64	698.00	59.59
24	2.83	27.82	345.00	78.17
25	2.44	25.27	554.00	71.00
26	1.25	15.97	718.00	44.87
27	1.22	34.89	339.00	98.04
28	1.30	35.12	435.00	98.68
29	1.42	33.81	534.00	95.01
30	1.91	25.26	380.00	70.97
31	1.43	37.14	618.00	104.35
32	2.38	21.63	542.00	60.78
33	1.18	30.24	366.00	84.98
34	1.23	29.36	626.00	82.51
35	1.58	22.65	354.00	63.64
36	2.34	22.99	163.00	22.76
37	2.14	22.14	372.00	21.92
38	1.91	11.94	380.00	11.82
39	2.03	18.42	360.00	18.23
40	1.19	30.73	436.00	30.43
41	2.63	25.87	325.00	67.52
42	2.51	25.98	534.00	67.81
43	1.50	19.16	698.00	50.02
44	2.38	14.85	542.00	38.75
45	2.03	26.73	360.00	69.78
46	1.19	30.73	598.00	80.22

Table 2
Ranking of proposed method in comparison to Chen's method.

Ranking	Association rule number (DMU)	
	Chen's method	Proposed method
1	26	18
2	22	23
3	18	26
4	17	12
5	7	31
6	23	43
7	6	22
8	43	6
9	31	17
10	12	1
11	1	7

CCR model. Appendix A presents the results of CCR model by considering a single input (whose value is equal to 1 for all DMUs) and four outputs. To provide further insight, by applying Amin & Toloo's model to data of Appendix A, most efficient association rule is identified as DMU_{18}, similar to our proposed model. It is while, Chen's method finds DMU_{26} as most efficient DMU.

7. Conclusion

Data mining popularity is growing at a lightning-fast pace. 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. In this paper, we developed a new integrated DEA model which is able to identify most CCR-efficient DMU by considering only outputs data of them, without any input. This model is applicable for finding most efficient association rule. Consequently, by utilizing proposed model, we introduced a new method for ranking association rules with multiple criteria. In comparison to previous works, our method is computationally efficient and also ranks all association rules.

Appendix A

Efficient association rules identified by CCR.

Association rule number (DMU)	Outputs				Input	Efficiency score of CCR
	Support (%)	Confidence (%)	Itemset value	Cross-selling profit		
1	3.87	40.09	337.00	25.66	1	1
2	1.42	18.17	501.00	11.63	1	0.779447
3	2.83	17.64	345.00	11.29	1	0.837651
4	2.34	30.83	163.00	19.73	1	0.709984
5	2.63	23.90	325.00	15.30	1	0.783182
6	1.19	55.65	436.00	35.61	1	1
7	1.19	47.42	598.00	30.35	1	1
8	1.19	15.70	436.00	52.91	1	0.685432
9	1.19	10.82	598.00	36.45	1	0.847433
10	1.19	12.32	436.00	20.08	1	0.666243
11	1.19	12.32	598.00	40.04	1	0.847433
12	3.87	38.08	337.00	103.97	1	1
13	1.18	15.09	710.00	41.19	1	0.988858
14	2.44	15.22	554.00	41.56	1	0.999116
15	2.14	28.21	372.00	77.02	1	0.782319
16	2.51	22.81	534.00	62.26	1	0.989038
17	1.19	50.92	436.00	139.02	1	1
18	1.19	45.25	598.00	123.52	1	1
19	1.19	11.70	436.00	43.54	1	0.66774
20	1.19	11.70	598.00	62.50	1	0.878682
21	1.42	13.99	501.00	61.16	1	0.794483
22	1.18	12.23	710.00	53.45	1	1
23	1.50	13.64	698.00	59.59	1	1
24	2.83	27.82	345.00	78.17	1	0.837651
25	2.44	25.27	554.00	71.00	1	0.999116
26	1.25	15.97	718.00	44.87	1	1
27	1.22	34.89	339.00	98.04	1	0.751815
28	1.30	35.12	435.00	98.68	1	0.807445
29	1.42	33.81	534.00	95.01	1	0.896064
30	1.91	25.26	380.00	70.97	1	0.750386
31	1.43	37.14	618.00	104.35	1	1
32	2.38	21.63	542.00	60.78	1	0.9763
33	1.18	30.24	366.00	84.98	1	0.696249
34	1.23	29.36	626.00	82.51	1	0.957845
35	1.58	22.65	354.00	63.64	1	0.670822
36	2.34	22.99	163.00	22.76	1	0.604651
37	2.14	22.14	372.00	21.92	1	0.753352
38	1.91	11.94	380.00	11.82	1	0.724173
39	2.03	18.42	360.00	18.23	1	0.722315
40	1.19	30.73	436.00	30.43	1	0.747445
41	2.63	25.87	325.00	67.52	1	0.783182
42	2.51	25.98	534.00	67.81	1	0.989038
43	1.50	19.16	698.00	50.02	1	1
44	2.38	14.85	542.00	38.75	1	0.9763
45	2.03	26.73	360.00	69.78	1	0.748978
46	1.19	30.73	598.00	80.22	1	0.925912

References

Agrawal, R., Imielinski, T., & 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., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of the international conference on very large data bases*, 407–419.

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

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

Camanho, A. S., & Dyson, R. G. (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, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research, 2*, 429–444.

Chen, M. C. (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., & 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, D. H., Ahn, B. S., & Kim, S. H. (2005). Prioritization of association rules in data mining: Multiple criteria decision approach. *Expert Systems with Applications, 29*, 867–878.

Cook, W. D., & Kress, M. (1990). A data envelopment model for aggregating preference rankings. *Management Science, 36*, 1302–1310.

Edirisinghe, N. C. P., & 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., & Parker, B. (2007). A bibliography of data envelopment analysis (1978–2003). *Socio-Economic Planning Sciences, 38*, 159–229.

Ertay, T., Ruan, D., & Tuzkaya, U. R. (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., & 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, S. T. (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., & Ma, Y. (2000). Analyzing the subjective interestingness of association rules. *IEEE Intelligent Systems, 15*, 47–55.

Mannino, M., Hong, S. N., & Choi, I. J. (2008). Efficiency evaluation of data warehouse operations. *Decision Support Systems, 44*, 883–898.

Marakas, G. M. (2004). *Decision support systems in the 21st Century*. Second Edition: Prentice-Hall of India.

Ng. R. T., Lakshmanan, L. V. S., Han, J., & 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* (pp. 13–24).

Obata, T., & 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., & Wu, S. (2008). Operations research and data mining. *European Journal of Operational Research, 187*, 1429–1448.

Shafer, S. M., & Byrd, T. A. (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., & Agrawal, R. (1997). Mining association rules with item constraints. In *Proceedings of the third international conference on knowledge discovery and data mining, KDD-97* (pp. 67–73).

Tan, P. N., & Kumar, V. (2000). Interestingness measures for association patterns: A perspective. In *KDD 2000 workshop on postprocessing in machine learning and data mining*, Boston, MA, August.

Toloo, M., & Nalchigar, S. (2009). A new integrated DEA model for finding most BCC–efficient DMU. *Applied Mathematical Modelling, 33*, 597–604.