Iranian Journal of Management Studies (IJMS) *http://ijms.ut.ac.ir/*
Vol. 8, No. 4, October 2015 *Print ISSN: 2008-7055*
pp: 589-605 *Online ISSN: 2345-3745*

Evaluation of recommender systems: A multi-criteria decision making approach

Babak Sohrabi[1], Mehdi Toloo[2], Ali Moeini[3], Soroosh Nalchigar[1*]

1. Faculty of Management, University of Tehran,Iran
2. Department of Business Administration, Technical University of Ostrava, Czech Republic
3. School of Engineering, University of Tehran, Iran

(Received: 25 July 2014; Revised: 6 September 2015; Accepted: 8 September2015)

Abstract

The evaluation and selection of recommender systems is a difficult decision making process. This difficulty is partially due to the large diversity of published evaluation criteria in addition to lack of standardized methods of evaluation. As such, a systematic methodology is needed that explicitly considers multiple, possibly conflicting metrics and assists decision makers to evaluate and find the best recommender system among a given set of alternatives. This paper introduces Multi-Criteria Decision Making (MCDM) approach for evaluation of recommender systems. In particular, this paper proposes the use of Data Envelopment Analysis (DEA) approach, as a sub-category of MCDM, in order to solve this problem. Various DEA models are introduced and their applicability are illustrated. A real case of evaluation of recommender systems is used to demonstrate the approach.

Keywords

Data envelopment analysis, Evaluation, Metrics, Multi-criteria decision making, Recommender systems.

∗ **Corresponding Author: Email:** soroosh.n@gmail.com

Introduction

Recommender systems assist users to find potentially interesting items in a given domain (e.g., movies, books, applications, websites, and travel destinations) by suggesting the items that match their preferences, tastes, and needs. These systems utilize various sources of information about users to provide predictions and recommendations of items (Bobadilla *et al*., 2013). Collaborative filtering (Resnick et al., 1994), content-based filtering (Pazzani and Billsus, 2007), and hybrid filtering (Balabanović and Shoham, 1997) are main approaches used for designing recommender systems. Since the advent of recommender systems in early-to-mid 1990's, this field has gained great deal of attention from academia as well as industry and has advanced into the point where these systems play an important role in a wide variety of e-commerce applications (Konstan and Riedl, 2012; Adomavicius and Tuzhilin, 2005).

Evaluating recommender systems is a critical and challenging task for several reasons. One of the reasons is that there exist many quantitative metrics as well as additional qualitative evaluation techniques. While early evaluation works focused on “accuracy” metrics for evaluation, more recent researches (e.g., Wu *et al*., 2012; McNee *et al*., 2006; Fouss and Saerens, 2008) advocate that those metrics are not sufficient for evaluation of recommender systems so argue that the ratings are not necessarily representing the best recommender system. They mention that accuracy related metrics should be coupled with other criteria to gain a comprehensive evaluation of the recommender systems. As a result, many new metrics have been proposed to capture various aspects of the recommendation process. However, this brings up the question of how to combine or trade-off between multiple and sometimes conflicting criteria to find the best recommender system? How to evaluate a given set of alternative recommender systems with regarding to a set of metrics?

Existence of large number of metrics has been resulted in a need of a uniform and systematic way of evaluation of the recommender systems. Although the diverse set of metrics facilitates examining various aspects of recommender systems, there is still a lack of a common methodology to put together these metrics, compare, and rate the recommender systems. In other words, there is a need for a systematic method that considers multiple metrics together and evaluates the recommender systems. To address these problems, this paper introduces Multi-Criteria Decision Making (MCDM) approach for evaluation of recommender systems. MCDM is an important sub-discipline of operations research that deal with decision making problems in presence of multiple decision criteria (Zeleny, 1982). It is concerned with designing mathematical and computational models to assist the subjective evaluation of a finite number of decision alternatives under a finite number of performance criteria (Lootsma, 1999). There are a variety of existing techniques for solving MCDM problems, one of which is Data Envelopment Analysis (DEA).

DEA is a widely used optimization, based on non-parametric method for efficiency evaluation of a set of similar units, usually referred to as Decision Making Units (DMUs). The main idea of DEA is to assess the efficiency of a DMU based on its performance of generating outputs in means of input consumption. DEA models are widely recognized and applied to various industrial and non-industrial decision making contexts. Evaluation of information system projects (Sowlati *et al*., 2005), ranking of data mining association rules (Chen, 2007; Toloo *et al*., 2009), efficiency assessment of bank branches (Paradi and Zhu, 2013), designing of facility layouts in manufacturing systems (Ertay *et al*., 2006), and evaluation of data warehouse operations (Mannino *et al*., 2008) are examples of applications of DEA in various contexts. In this paper, we introduce and show applications of various DEA models for evaluation of recommender systems in presence of multiple evaluation metrics.

This paper shows how DEA models could assist organizational decision makers to evaluate a set of recommender systems and to find the best system among the given alternatives. The main advantage of the proposed approach over previously proposed ones is that it is systematic, non-parametric, and is able to handle multiple, and possibly conflicting criteria. The remainder of this article is structured as follows: Section 2 reviews related works on evaluation of recommender systems. Section 3 introduces basic DEA models. Section 4 introduces recent DEA models that are proposed to find the most efficient unit among a given set of alternatives. Section 5 illustrates application of various DEA models for evaluation of recommender systems. This paper ends at Section 6 with some concluding remarks and directions for future research.

Related literature

Evaluation of recommender systems has been addressed and discussed by various authors in the past. Breese *et al*. (1998) perhaps were the first ones who performed an evaluation of several recommendation algorithms over a number of data sets. They compared the algorithms in identical experiments using two classes of metrics: accuracy of predictions and utility of a ranked list of suggested items. Later, their approach was used by many other authors to evaluate recommender systems. Herlocker *et al*. (2004) performed a comprehensive survey of metrics for evaluation of recommender systems and conceptually analyzed their strengths and weaknesses. Moreover, they proposed a classification of recommendation systems from the user task perspective, that is the intentions of the users when using the recommender systems. They performed experiments, measured various accuracy measures on a dataset and found that while many of the measures are strongly correlated, there exist three classes of measures which are uncorrelated.

Aiming to provide a common ground round for evaluating recommender systems, Hernández del Olmo and Gaudioso (2008) proposed a framework along with a new metric for evaluation of recommender systems. They performed a comparison between their new metric and the traditional ones (e.g., accuracy). Gunawardana and Shani (2009) categorized the existing recommender system tasks into three major groups (the prediction task, the recommendation task, and the utility maximization task) and reviewed a set of well-known evaluation metrics for each task. Zaier *et al*. (2008) introduced and discussed the importance of the long tail theory (i.e., small number of items sale very high and a large number of items sale low) for recommender system applications. They performed a review of the existing datasets that are used to evaluate recommender systems algorithms and studied the distribution of data and its impact on recommendation quality.

McNee *et al*. (2006) discussed that the recommender systems that are most accurate according to existing accuracy criteria, does not necessarily recommend the most useful items to the users. They argued that to have a proper evaluation of these systems, we should move beyond the traditional accuracy metrics and their associated experimental methodologies. They provided three aspects of the recommendation process that are not captured by accuracy measures: the similarity of recommendation lists, recommendation serendipity, and the importance of user needs and expectations in a recommender. Similarly, Wu *et al*. (2012) explained that the accuracy criteria alone are not enough for a proper evaluation of recommender systems. They suggested to use a variety of criteria (as opposed to using only accuracy criteria), including coverage, diversity, serendipity and so on, for evaluating recommender systems. They performed experiments and compared five algorithms from different aspects, and concluded that use of various criteria in assessing recommender systems is meaningful. Also, Fouss and Saerens (2008) suggested that using a single criteria of accuracy in evaluating recommender systems is not sufficient and it should be coupled with other considerations. They proposed coverage, confidence metrics computing time and robustness as additional set of metrics for assessing the performance of recommender systems. They performed a comparison of six recommendation algorithms based on four criteria (accuracy, computing time, robustness, and novelty) and concluded that kernel-based algorithms provide the best results overall.

McLaughlin and Herlocker (2004) indicated that use of the Mean Absolute Error (MAE) as a metric for evaluation of recommender systems could be misleading and that ratings are not necessarily indicative of whether a user is likely to choose an item. They proposed a modified Precision metric for more accurate evaluation of the user experience. In order to cover all aspects that affect the effectiveness of recommender systems, Said *et al*. (2012) proposed a three-dimensional evaluation model for recommender systems. The three perspectives in this model were user requirements, business requirements and technological constraints. They briefly reviewed the existing evaluation metrics and mentioned their shortcoming. They described application of their proposed model on a specific use case. Table 1summarizes this literature.

Table 1. Summary of the literature

Authors	Summary of contribution
Breese *et al*. (1998)	Use of criteria such as Accuracy of predictions, Utility of a ranked list of suggested items
Herlocker *et al*. (2004)	Comprehensive review of existing criteria including: Mean Absolute Error, Precision, Recall, ROC Curves, Rank Accuracy, (and their related metrics), Prediction-Rating Correlation, Half-life Utility, Coverage, Learning Rate, Novelty and Serendipity, Confidence.
Fouss & Saerens (2008)	Use of criteria: Robustness, Recall, Novelty, and Computing time.
Hernández del Olmo & Gaudioso (2008)	Introduction of a new measure, namely final performance, and review of existing criteria such as: Recall, Precision, Accuracy.
Gunawardana & Shani (2009)	Review of accuracy evaluation metrics including: Root of the Mean Square Error, Mean Average Error, and Normalized Mean Average Error, Precision, Recall, ROC Curves, False Positive Rate
McLaughlin & Herlocker (2004)	Proposed modified Precision metric to be used beside Mean Absolute Error.
Wu *et al*. (2012)	Suggested use of other criteria beside accuracy, including: coverage, diversity, serendipity.
Said *et al*. (2012)	Suggested three evaluation perspectives: user requirements, business requirements and technological constraints.

Investigation of previous related works shows that the importance and the challenges of evaluation of recommender systems have been discussed. Moreover, many criteria have been proposed to cover various aspects of the recommendation process. However, there is a lack of systematic methodology that is able to evaluate and compare a set of alternative recommender systems based on several criteria. In other words, there is a need for a methodology that supports decision makers in finding the most appropriate recommender system while considering multiple evaluation metrics simultaneously. This paper fills this gap by proposing a methodology for evaluating a set of recommender systems and finding the most appropriate one.

DEA Models

DEA, initially proposed by Charnes, Cooper, and Rhodes (CCR model) in 1978 (Charnes *et al*., 1978), is a mathematical programming based approach used to evaluate the performance of a group of homogeneous DMUs. A DMU can be viewed as a system that converts a set of inputs into a set of outputs and its performance has to be evaluated. In this method, the efficiency performance of a given DMU is defined as the ratio of the sum of weighted outputs to the sum of weighted inputs. Since the advent of DEA, there has been a significant growth both in theoretical development as well as the range of applications (Cook and Seiford, 2009). Interested readers are referred to the paper by Liu *et al*. (2013) for a survey of DEA literature and to the paper by Gattoufi *et al*. (2004) for a taxonomy of DEA literature.

Suppose that there are n DMUs, $(DMU_j: j = 1,2,\dots,n)$ which utilize m input $(x_i: i = 1,2,\dots,m)$ to generate s outputs $(y_r: r = 1,2,\dots,s)$. The formulation of the classical DEA model, namely CCR model, is as follows:

$$
\begin{aligned}
&\text{Max} \sum_{r=1}^{s} u_r y_{ro} \\
&s.t. \\
&\sum_{i=1}^{m} w_i x_{io} = 1 \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} w_i x_{ij} \le 0 \quad j = 1,2,\dots,n \\
&w_i \ge \varepsilon \quad i = 1,2,\dots,m \\
&u_r \ge \varepsilon \quad r = 1,2,\dots,s
\end{aligned}
\tag{1}
$$

where x_{ij} and y_{ij} (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, DMU under consideration. Also, ε is non-Archimedean infinitesimal value for forcing the weights to be greater than zero. By solving Model 1, the efficiency of DMU_o is calculated. This model needs to be run once for each DMU (n times in total) in order to find the scores efficiency of alternative DMUs. Scores equal to one indicate efficient DMUs while scores less than one signify inefficient units. The CCR model assumes constant returns to scale. This means the model assumes that a proportional increase in inputs results in a proportionate increase in outputs. Later, Banker et al. (1984) introduced the assumption of variable returns to scale and extended the CCR model to BCC model. The BCC model evaluates the pure technical efficiency score of DMUs and identifies whether a DMU is operating in increasing, decreasing or constant returning to scale.

DEA Models for Finding Most Efficient Unit

Using the classic DEA models, the decision maker gets a categorization of all DMUs as either "efficient" or "inefficient" and these models fail to discriminate the efficient DMUs. In other words, the traditional DEA models does not assist decision makers to differentiate the efficient units and to find the best alternative. However, in many real world contexts, the decision makers need to find a single most efficient DMU among a given set of alternatives. To solve this problem, recently some new DEA models have been proposed in the literature. Ertay *et al*. (2006) has extended minimax DEA model to identify a single most efficient DMU and used to evaluate layout design of manufacturing systems. Amin and Toloo (2007) improved their work and proposed Model 2 for finding the most efficient DMU, given a set of units.

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

This model is an LP model in which the binary variable d_j represents the deviation of the DMU_j from efficiency frontier. Also, the ε^* is the maximum non-Archimedean for forestalling weights to be equal to zero. The objective function of this model is to minimize the maximum deviation from efficiency. After solving this model, DMU_j is the most efficient unit 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. Model 2 uses common set of optimal weights for all DMUs and hence it needs to be solved only once in order to find the most efficient unit. Model 2 assumes constant returns to scale and find most CCR-efficient unit. Therefore, it is not applicable for cases in which DMUs operate in variable returns to scale. Later, Toloo and Nalchigar (2009) extended this model and proposed a new DEA model for identifying the most BCC-efficient unit. Also, Amin (2009) showed that Model 2, in some situations, may result in more than one efficient DMU and proposed some improvements.

Toloo and Nalchigar (2011) proposed a new DEA model that is able to find the most efficient unit while the data of inputs and outputs of alternatives are imprecise (i.e. when the inputs and outputs of DMUs are given in terms of cardinal and ordinal data). Most recently, Toloo (2012) found some problems in their model and developed a new Mixed Integer Programing DEA (MIP-DEA) model for finding the most BCC-efficient DMU. His approach includes two steps. The first step recognizes a set of candidate DMUs for being most BCC-efficient unit. The second step finds the single most BCC-efficient unit among the candidates. He proposed following LP model to be used in the first step:

$$
\begin{aligned}
&\min\ d_{\max} \\
&s.t. \\
&d_{\max} - d_j \geq 0 \quad j = 1, 2, \ldots, n \\
&\sum_{i=1}^{m} w_i x_{ij} \leq 1 \quad j = 1, 2, \ldots, n \\
&\sum_{r=1}^{s} u_r y_{rj} + u_0 - \sum_{i=1}^{m} w_i x_{ij} + d_j = 0 \quad j = 1, 2, \ldots, n \\
&w_i \geq \varepsilon^* \quad i = 1, 2, \ldots, m \\
&u_r \geq \varepsilon^* \quad r = 1, 2, \ldots, s
\end{aligned} \tag{3}
$$

This model results in a common set of optimal positive weights for all DMUs, and DMU_j is a candidate for being most BCC-efficient unit if and only if $d_j^* = 0$. Having the set of candidate DMUs as outcome of Model 3, Toloo (2012) proposed following MIP-DEA integrated model to determine the most BCC-efficient unit:

$$
\begin{aligned}
&\min\ d_{\max} \\
&s.t. \\
&d_{\max} - d_j \geq 0 \quad j = 1, 2, \ldots, n \\
&\sum_{i=1}^{m} w_i x_{ij} \leq 1 \quad j = 1, 2, \ldots, n \\
&\sum_{r=1}^{s} u_r y_{rj} + u_0 - \sum_{i=1}^{m} w_i x_{ij} + d_j = 0 \quad j = 1, 2, \ldots, n \\
&\sum_{j=1}^{n} \theta_j = n - 1 \\
&d_j \leq M\theta_j \quad j = 1, 2, \ldots, n \\
&\theta_j \leq N d_j \quad j = 1, 2, \ldots, n \\
&d_j \geq 0 \quad j = 1, 2, \ldots, n \\
&\theta_j \in \{0, 1\} \quad j = 1, 2, \ldots, n \\
&w_i \geq \varepsilon^* \quad i = 1, 2, \ldots, m \\
&u_r \geq \varepsilon^* \quad r = 1, 2, \ldots, s
\end{aligned} \tag{4}
$$

where M and N are large numbers defined by the user and θ_j is a binary variable. This model is, indeed, an extended version of the Model 3 and includes some additional constraints to enforce finding a single most BCC-efficient unit. In this model, if $\theta_j = 0$, then the constraint $\theta_j \leq Nd_j$ is redundant and the constraint $d_j \leq M\theta_j$ forces that d_j is equal to zero. Otherwise, if $\theta_j = 1$, then $d_j \leq M\theta_j$ is a redundant constraint and the constraint $\theta_j \leq Nd_j$ insures d_j to be positive. These imply that in this model $d_j = 0$ if $\theta_j = 0$, and $d_j > 0$ if $\theta_j = 1$.

The main contribution of this paper is to introduce and to illustrate applications of DEA models for evaluation and selection of most efficient recommender systems. Towards this end, this paper first applies the traditional DEA models and shows how they could be used for differentiating efficient versus inefficient recommender systems. After that, we apply the models proposed by Toloo (2012) to further analyze the efficient recommender systems and to identify the most efficient one. A dataset of a real case of evaluation of recommender systems is used to demonstrate the approach.

Illustration

In this section, we show the applicability of DEA models on evaluation of collaborative recommendation methods. The data that we use in this section is obtained from the paper by Fouss and Saerens (2008), where they calculated four performance criteria for six recommender system algorithms. We opted to use their data due to availability of proper criteria and their values for a set of competing algorithm. The first three algorithms, namely Basic, Binary coefficient (Bin), and Cosine coefficient (Cos) are classical memory-based scoring algorithms while the last three algorithms, that is, regularized Laplacian kernel (K_{RL}), commute time kernel (K_{CT}), and markov diffusion kernel (KMD) are kernels on a graph. These algorithms are evaluated with regarding to robustness, recall, novelty, and computing time. The robustness criteria, robustness accuracy score, in Fouss and Saerens (2008) measures the ability of an algorithm to make good predictions in the presence of noisy data. The recall metric measures the capacity of the recommender algorithm in obtaining all the relevant items present in the pool. The novelty metric measures the ability of an algorithm to recommend items that are non-obvious and surprising for the user. Finally, the computing time metric is the time needed by each recommender system to provide its recommendations. The recall, novelty, and computing time metrics were calculated on the real MovieLens data set while the robustness of the algorithms were measured on- artificially generated datasets. Interested readers are referred to the papers by Fouss and Saerens (2008) and also Fouss *et al*. (2007) for details on calculation of the metrics.

Table 2 presents the data. Similar to previous applications of DEA (e.g., Doyle and Green, 1993; Sowlati *et al*., 2005; Ertay *et al*., 2006), the evaluation criteria that are to be minimized are viewed as DEA inputs, and the criteria to be maximized are considered as DEA outputs.Attention should be paid that, (Fouss and Saerens, 2008), a high novelty score shows that the recommendation algorithm tends to position obvious, frequently bought items among the top items in the ranked list. Therefore, the novelty score should be as low as possible for good performance and is considered as a DEA input in this illustration.

Table 2. Data of six recommender systems (Fouss and Saerens, 2008)

	DEA Input		DEA Outputs	
DMUs	**Computing Time (Sec)**	**Novelty (%)**	**Recall (%)**	**Robustness(%)**
Basic	0.0	247.24	25.60	27.47
Bin	65.6	244.95	31.13	28.01
Cos	79.7	240.46	31.29	27.93
K_{RL}	40.5	231.12	31.80	28.55
K_{CT}	24.9	234.51	32.20	28.28
K_{MD}	106.7	230.04	31.27	28.55

First, by using DEA-Solver[1], we applied the CCR and BCC models (input oriented versions) on the data of table 2. Results, which are presented in table 3 indicate that the output of these two models are similar and both implies that the second and third DMUs, namely Binary coefficient (Bin) and Cosine coefficient (Cos) are inefficient

1. http://www.saitech-inc.com/products/prod-dsp.asp

systems. Moreover, these results show that out of the six alternative recommender systems four of them are efficient systems, namely Basic, regularized Laplacian kernel (K_{RL}), commute time kernel (K_{CT}), and markov diffusion kernel (K_{MD}). Obviously, by using the basic DEA models, the decision makers are not able to have a proper evaluation and to find the most proper alternative based on the criteria at hand.

Table 3. Evaluation of recommender systems with basic DEA models (CCR and BCC)

DMUs	CCR efficiencies	BCC efficiencies
Basic	1.00	1.00
Bin	0.92	0.94
Cos	0.94	0.95
K_{RL}	1.00	1.00
K_{CT}	1.00	1.00
K_{MD}	1.00	1.00

To further analyze the performance of these systems, we use DEA for finding the most efficient recommender system among the alternatives. Among existing DEA models for finding most efficient unit (reviewed in previous section), we opt to apply the approach proposed by Toloo (2012) since it is the most recent in the domain and tackles the drawbacks of its previously proposed models. As mentioned in previous section, this approach includes two steps. The first step uses the Model 3 to find a set of candidate DMUs for being the most efficient unit. The second step uses Model 4 to find the single most efficient unit from those candidates.

Using GAMS operations research software[1], we solve Model 3 for the data presented in Table 2. The maximum value of non-Archimedean epsilon is $\varepsilon^* = 0.002970$. Using this value and solving model 3 for the dataset, following results are achieved:

$$d_j^* \begin{cases} = 0 & j = 5 \\ > 0 & j \neq 5 \end{cases} \quad \text{and} \quad d_{max}^* = 0.210937$$

which implies that DMU_5 is the only candidate for being most efficient recommender system. Having only one candidate means that,

1. http://www.gams.com/

indeed, Model 3 is able to discriminate the four efficient recommender systems (Table 2) and there is no need to solve the Model 4. To conclude, commute time kernel (K_{CT}) is the most efficient recommender systems among the given six alternatives.

Conclusion

Recommender systems are widespread in modern e-commerce. Evaluation and selection of recommender systems are a complex decision making task, mainly due to existence of numerous evaluation criteria in addition to lack of standardized methods of evaluation. This paper suggested the use of MCDM techniques as a solution that can better model the complexity of evaluating recommender systems. DEA models, as a sub-category of MCDM, were introduced and their applications were illustrated on comparisons of a set of recommendation algorithms.

Obviously, the goal of this paper was not to provide a comprehensive list of evaluation metrics, but rather to propose a MCDM approach and show how that would work for a given set of evaluation metrics. Future studycan perform a comprehensive review of evaluation criteria for recommender systems. Moreover, future works can extend and show application of the proposed approach with other existing evaluation metrics. We anticipate that in future, more qualitative and/or fuzzy metrics for evaluating users experience with recommender systems would be proposed. Future studies can extend the proposed method such that it handles qualitative and/or fuzzy metrics. These extensions are left to future researches.

References

Adomavicius, G. & Tuzhilin, A. (2005). "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions". *Knowledge and Data Engineering, IEEE Transactions on*, 17(6), 734-749.

Amin, G. R. (2009). "Comments on finding the most efficient DMUs in DEA: An improved integrated model". *Computers & Industrial Engineering*, 56(4), 1701-1702.

Amin, G. R. & Toloo, M. (2007). "Finding the most efficient DMUs in DEA: An improved integrated model". *Computers & Industrial Engineering*, 52(1), 71-77.

Balabanović, M. & Shoham, Y. (1997). "Fab: content-based, collaborative recommendation". *Communications of the ACM*, 40(3), 66-72.

Banker, R. D.; Charnes, A. & Cooper, W. W. (1984). "Some models for estimating technical and scale inefficiencies in data envelopment analysis". *Management science*, 30(9), 1078-1092.

Bobadilla, J.; Ortega, F.; Hernando, A. & Gutiérrez, A. (2013). "Recommender systems survey". *Knowledge-Based Systems*, 46, 109-132.

Breese, J. S.; Heckerman, D. & Kadie, C. (1998). "Empirical analysis of predictive algorithms for collaborative filtering". In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (pp. 43-52). Morgan Kaufmann Publishers Inc.

Charnes, A.; Cooper, W. W. & Rhodes, E. (1978). "Measuring the efficiency of decision making units". *European journal of operational research*, 2(6), 429-444.

Chen, M. C. (2007). "Ranking discovered rules from data mining with multiple criteria by data envelopment analysis". *Expert Systems with Applications*, 33(4), 1110-1116.

Cook, W. D. & Seiford, L. M. (2009). "Data envelopment analysis (DEA)–Thirty years on". *European Journal of Operational Research*, 192(1), 1-17.

Doyle, J. & Green, R. (1993). "Data envelopment analysis and multiple criteria decision making". *Omega*, 21(6), 713-715.

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(3), 237-262.

Fouss, F. & Saerens, M. (2008). "Evaluating performance of recommender systems: An experimental comparison". In *Web Intelligence and*

Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on (Vol. 1, pp. 735-738). IEEE.

Fouss, F.; Pirotte, A.; Renders, J. M. & Saerens, M. (2007). “Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation”. *Knowledge and Data Engineering, IEEE Transactions on*, 19(3), 355-369.

Gattoufi, S.; Oral, M. & Reisman, A. (2004). “A taxonomy for data envelopment analysis”. *Socio-Economic Planning Sciences*, 38(2), 141-158.

Gunawardana, A. & Shani, G. (2009). “A survey of accuracy evaluation metrics of recommendation tasks”. *The Journal of Machine Learning Research*, 10, 2935-2962.

Herlocker, J. L.; Konstan, J. A.; Terveen, L. G. & Riedl, J. T. (2004). “Evaluating collaborative filtering recommender systems”. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5-53.

Hernández del Olmo, F. & Gaudioso, E. (2008). “Evaluation of recommender systems: A new approach”. *Expert Systems with Applications*, 35(3), 790-804.

Konstan, J. A. & Riedl, J. (2012). “Recommender systems: from algorithms to user experience”. *User Modeling and User-Adapted Interaction*, 22(1-2), 101-123.

Liu, J. S.; Lu, L. Y.; Lu, W. M. & Lin, B. J. (2013). “Data envelopment analysis 1978–2010: A citation-based literature survey”. *Omega*, 41(1), 3-15.

Lootsma, F. A. (Ed.). (1999). *Multi-criteria decision analysis via ratio and difference judgment.* Kluwer Academic Publishers, Applied Optimization Series (Vol. 29). Springer.

Mannino, M.; Hong, S. N. & Choi, I. J. (2008). “Efficiency evaluation of data warehouse operations”. *Decision Support Systems*, 44(4), 883-898.

McLaughlin, M. R. & Herlocker, J. L. (2004). “A collaborative filtering algorithm and evaluation metric that accurately model the user experience”. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 329-336). ACM.

McNee, S. M.; Riedl, J. & Konstan, J. A. (2006). “Being accurate is not enough: how accuracy metrics have hurt recommender systems”. In *CHI'06 extended abstracts on Human factors in computing systems* (pp. 1097-1101). ACM.

Paradi, J. C. & Zhu, H. (2013). “A survey on bank branch efficiency and performance research with data envelopment analysis". *Omega*, 41(1), 61-79.

Pazzani, M. J. & Billsus, D. (2007). *Content-based recommendation* systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web Methods and Strategies of Web Personalization, vol. 4321, pp. 325–341. Springer, Berlin/Heidelberg.

Resnick, P.; Iacovou, N.; Suchak, M.; Bergstrom, P. & Riedl, J. (1994). "GroupLens: an open architecture for collaborative filtering of netnews". In *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (pp. 175-186). ACM.

Said, A.; Tikk, D.; Shi, Y.; Larson, M.; Stumpf, K. & Cremonesi, P. (2012). *Recommender systems evaluation: A 3D benchmark*. In ACM RecSys 2012 Workshop on Recommendation Utility Evaluation: Beyond RMSE, Dublin, Ireland (pp. 21-23).

Sowlati, T.; Paradi, J. C. & Suld, C. (2005). "Information systems project prioritization using data envelopment analysis". *Mathematical and Computer Modelling*, 41(11), 1279-1298.

Toloo, M. (2012). "On finding the most BCC-efficient DMU: A new integrated MIP–DEA model". *Applied Mathematical Modelling*, 36(11), 5515-5520.

Toloo, M. & Nalchigar, S. (2009). "A new integrated DEA model for finding most BCC-efficient DMU". *Applied Mathematical Modelling*, 33(1), 597-604.

Toloo, M. & Nalchigar, S. (2011). "A new DEA method for supplier selection in presence of both cardinal and ordinal data". *Expert Systems with Applications*, 38(12), 14726-14731.

Toloo, M.; Sohrabi, B. & Nalchigar, S. (2009). "A new method for ranking discovered rules from data mining by DEA". *Expert Systems with Applications*, 36(4), 8503-8508.

Wu, W.; He, L. & Yang, J. (2012). "Evaluating recommender systems". In *Digital Information Management (ICDIM), 2012 Seventh International Conference on* (pp. 56-61). IEEE.

Zaier, Z.; Godin, R. & Faucher, L. (2008). "Evaluating recommender systems". In *Automated solutions for Cross Media Content and Multi-channel Distribution, 2008. AXMEDIS'08. International Conference on* (pp. 211-217). IEEE.

Zeleny, M. (1982). *Multiple Criteria Decision Making* (Vol. 25). J. L. Cochrane (Ed.). New York: McGraw-Hill.