Global Journal of Computer Sciences: Theory and Research

Volume 8, Issue 3, (2018) 120-125

www.gjcs.eu

Associative classification of the Jordanian hospitals efficiency based on DEA

Ahmad Alaiad, Department of Computer Information System, Jordan University of Science and Technology, Irbid 22110, Jordan

Hassan Najadat*, Department of Computer Information System, Jordan University of Science and Technology, Irbid 22110, Jordan

Nusaiba Al-Mnayyis, Department of Computer Science, Jordan University of Science and Technology, Irbid 22110, Jordan

Ashwaq Khalil, Department of Computer Science, Jordan University of Science and Technology, Irbid 22110, Jordan

Suggested Citation:

Alaiad, A., Najadat, H., Al-Mnayyis, N. & Khalil, A. (2018). Associative classification of the Jordanian hospitals efficiency based on DEA. *Global Journal of Computer Sciences: Theory and Research*. *8*(3), 120–125.

Received from March 15, 2018; revised from July 11, 2018; accepted from November 10, 2018.
Selection and peer review under responsibility of Prof. Dr. Dogan İbrahim, Near East University, Cyprus.

Abstract

Data envelopment analysis (DEA) has been widely used in many fields. Recently, it has been adopted by the healthcare sector to improve efficiency and performance of the healthcare organisations, and thus, reducing overall costs and increasing productivity. In this paper, we demonstrate the results of applying the DEA model in Jordanian hospitals. The dataset consists of 28 hospitals and is classified into two groups: efficient and non-efficient hospitals. We applied different association classification data mining techniques (JCBA, WeightedClassifier and J48) to generate strong rules using the Waikato Environment for Knowledge Analysis. We also applied the open source DEA software and MaxDEA software to manipulate the DEA model. The results showed that JCBA has the highest accuracy. However, WeightedClassifier method achieves the highest number of generated rules, while the JCBA method has the minimum number of generated rules. The results have several implications for practice in the healthcare sector and decision makers.

Keywords: Component, DEA, DMU, output-oriented model, health care system.

* ADDRESS FOR CORRESPONDENCE: Hassan Najadat, Department of Computer Information System, Jordan University of Science and Technology, Irbid 22110, Jordan.
E-mail address: najadat@just.edu.jo/ Tel.: +962-27211111-2499

1. Introduction

Nowadays, healthcare sector becomes a high priority for all countries over the world. Researchers pay more attention to healthcare services, as well as healthcare providers to improve patient's health in these countries and therefore improve individual productivity. Managing the healthcare operation has a substantial impact on the expenditures on healthcare in these countries. Hence, it is very important to evaluate the efficiency of healthcare services, institutions and providers.

Generally speaking, the lack of utilisation of resources or waste of the available resources would reduce the efficiency of healthcare services. Since healthcare may be an impediment to countries' economies, discussing the efficiency of the healthcare services becomes a persistent need to adjust the consumed resources and the desired outputs. In addition, it enables us to know the level of healthcare services provided in these countries.

There are many proposed methods to evaluate the efficiency of any implicit methodologies, viz, borders analysis and Data Envelopment Analysis (DEA). The DEA has become a common benchmarking tool for evaluating the efficiency of decision-making units (DMU) where any system or organisation can be decomposed into many components that can be considered as DMUs (Cooper, 2013; Ji & Lee, 2010). In this paper, we introduced the DEA in the healthcare sector to measure and evaluate the relative performance of Jordanian hospitals and to facilitate the task of controlling efficiency. Open Source DEA (OSDEA), MaxDEA and Waikato Environment for Knowledge Analysis (WEKA) tools have been used to apply the DEA model and association classification methods, respectively. The association classification algorithms used are JCBA and WeightedClassifier. These algorithms have been used to generate strong rules in order to represent the association between the attributes and class attribute in the dataset; 48J decision tree classifier is used also to compare the result of the algorithms.

This paper unfolds as follows. Section II presents background about the DEA and association classification technique in data mining. Section III demonstrates the related work. Section IV represents Jordanian hospitals dataset. Finally, Section V demonstrates the methodology and result.

2. Background

In this Section, we discuss the Associative Classifications (ACs) techniques and the DEA approach to estimate the performances of DMUs. Furthermore, we represent the WEKA tool; an open source data mining tool.

2.1. Associative classifications technique

AC is a branch of data mining field. It merges two known data mining tasks, association rule mining and classification to build a classifier (model) for prediction purpose (Thabtah, 2007).

Classification is a process of building a model that best represents a set of elements that may have common features. It has been used to distinguish data elements and to predict a clear class for objects that unknown class label. Association Rule Mining includes two steps, the first is to find the frequent item sets and the second is to produce association rules from that frequent item sets. Based on a minimum support threshold, the item sets are determined as frequent or not. The produced rules must satisfy both of the minimum support thresholds and the minimum confidence thresholds (Han, Kamber & Pei, 2011).

Association Classification is an individual case of association rule mining in which only the class attribute is considered in the rule's consequent (right-hand side) as shown in Eq. (1), (Thabtah, 2007). By applying the AC process, you will be able to foretell the efficiency of any new case given only its inputs and outputs. Also, you will know the most important inputs affect the efficiency.

$$P1 \wedge P2 \wedge \ldots \wedge PN \rightarrow \text{'Class} = C\text{'} \quad (1)$$

2.2. Data envelopment analysis

DEA is a nonparametric linear programming approach for estimating and evaluating the efficiency, productivity and the performance of a set of peer entities called Decision Making Units (DMUs). DMUs provide diverse outputs from diverse inputs. In 1978, Charnes, Cooper and Rhodes (CCR) introduced DEA as a tool for managerial and performance measurement (Charnes, Cooper & Rhodes, 1978). DMU generally refers to any entity that is evaluated in terms of its ability and efficiency in converting inputs into outputs. DEA approach has different properties and becomes vastly used in evaluating many kinds of entities in many different areas. Since the DEA requires some assumptions, that made it a suitable for evaluating the performance in many states, especially when the relations are complex or unknown among inputs and outputs involved in DMUs (Cooper, 2013; Ji & Lee, 2010).

DEA can deal with numerous inputs and numerous outputs at the same time without any distribution to the data. In terms of proportional change in inputs or outputs, the DEA is divided into two main models. The first model is the input-oriented model that focuses on minimising inputs while at least preserving the output levels. The second one is the output-oriented model, which is focused on maximising outputs without the need for maximising the given inputs (Han et al., 2011; Thabtah, 2007). In terms of returns to scale, DEA models can be split into two main efficiency measurement models by adding weight constraints: the first model is a constant return to scale and the second model is a variable return to scale (Cooper, 2013). DEA is described in general as in Eq. (2), (Charnes et al., 1978).

$$\text{Efficiency} = \sum \text{Weighted outputs} / \sum \text{Weighted inputs} \quad (2)$$

2.3. WEKA

WEKA is stood for Waikato Environment for Knowledge Analysis, Java-based open source software, developed at the University of Waikato in New Zealand and released under the General Public License. It contains a lot of methods used in data pre-processing, classification, regression, clustering, association rules and visualisation (Hall et al., 2009).

3. Related work

DEA can be applied in different fields and has been widely used by researchers. Some researchers used two-stage models of DEA in order to make the second one estimates whether the first unit works well. The two-stage models applied on 6000 Stock Exchange Tehran companies to analyse the financial statement. The result represented the efficiency of these companies; some of their efficiency was reduced in the second model, while the other has the same efficiency in both models. This makes the second model not the best one (Abolfathi, Zadeha & Abolfathia, 2013).

Another case of using DEA with two phases applied in the Brazilian National Agency of Electric Energy in order to understand the electricity distributors' performance. The first phase of DEA model was used to gather the environmental variables of the privilege regions and the second one was used to estimate the efficiency for these environmental variables (de Andrade et al., 2014).

Once the input and output data are not accurate, researchers worked to evaluate the DEA efficiency using slacks-based measures and the SBM models that are used to specify the efficiency intervals for each DMU (Azizi, Kordrostami & Amirteimoori, 2015). The DEA approved that it has the best performance (Azizi et al., 2015). Other research is based on a combination of the DEA with other method called Factor analysis in order to evaluate the DMUs efficiency without wasting the information. There was a various high correlation between the combination methods and the DEA method without combination (Nadimi & Jolai, 2008). Some researchers used a random data and

stochastic DEA approach to estimate the efficiency of operating entities (DMU). The result of using a constraint with stochastic DEA leads to increase the lower bound of an efficiency score of the entity with some pre-selected probability (Wei, Chen & Wang, 2014).

Some studies are using DEA to compare the healthcare systems among countries. The study of Serbia healthcare systems was compared with European countries systems using three inputs and three outputs. Two sets of countries were created with an identical level of development (Mitrovic, Vujosevic & Savic, 2015). Another research used Japanese municipal hospitals dataset in order to apply two different models of DEA based on the input and output factors. The first one based on the chronicles dataset assesses the difference of data that imposing chronological order weights, while the second one estimates the score and confidence ranges of future efficiency for all DMUs (Tone & Ouenniche, 2016).

4. Dataset

Our data are collected from the annual statistical report to the Ministry of Health in the Hashemite Kingdom of Jordan (2017). The dataset consists of 28 instances (hospitals) and seven attributes. For input and output variables of DEA approach, we chose as input, equipment and personnel (Najadat & Alsmadi, 2011; O'Neill et al., 2008), the number of Doctors, number of nurses, the number of beds and the number of the lab tests in each hospital. Because output composed the attributes that express quality and productivity (Najadat & Alsmadi, 2011; O'Neill et al., 2008), we chose outpatients number, number of surgical operations and number of emergencies as output. Table 1 displays these attributes and its statistics, which were collected using WEKA for input variables and output variables, respectively.

Table 1. Dataset information about attributes

Attribute	Type	Min	Max	Mean	Std Dev
Number of surgical operations	Output	33	25819	3180.25	4996.188
Number of Emergencies	Output	0	558858	106284. 143	104233. d893
Out-Patients	Output	751	86354	12867.071	15929.978
Number of Doctors	Input	20	981	109.869	181.231
Number of nurses	Input	43	1819	333.607	265.479
Number of Beds	Input	21	1101	165.143	211.172
Lab tests	Input	0	2703044	754342. 536	583625.484

5. Methodology and results

We used the OSDEA software and MaxDEA tool to evaluate the Jordanian hospital's efficiency by using the CCR output-oriented model. Our choice came after multi experiments that showed there are no huge differences in applying the different models on our dataset. Because the CCR model is the most vastly known and used in DEA and it supposes constant returns to scale, it is chosen to be the model for our methodology (O'Neill et al., 2008). Our methodology also combines the results of DEA with the AC data mining techniques. After that, it compares the results of AC algorithms decision tree classifier J48 to estimate the accuracy of these algorithms. Then, we enlarge our data using 'SMOTE' filter which is implemented in WEKA and repeat the previous step to check if the size of data impact or not in accuracy.

The association classification task finds the strong association between the attributes and class label. According to the DEA, the class label is represented by the efficiency score, where any value less than 0.9 represents inefficient class and higher than 0.9 represents the efficient class. To be capable of using the outputs of DEA with the association classification algorithms, the efficiency score is determined as 'E' and other values are determined as 'N'.

By applying the AC process, we will be able to foretell the efficiency of any new hospitals given only its inputs and outputs. Also, we will know that the most important inputs affect the efficiency. For the given data set, we have 15 inefficient hospitals and 13 efficient hospitals.

The association classification algorithm that has been used in WEKA is JCBA, which is A CBA classifier implemented using Java. It works with class association rules as a decision list classifier where one class is allowed in the consequence; it has two steps of pruning: optimal and obliges.

Another classification algorithm used was WeightedClassifier, which is also Java implemented Classifier. It weights a set of class association rules. In WeightedClassifier, no pruning is performed and, for classification, there are three options you can choose: the default one makes all rules with equal weight, the second one is to weight the rules linearly and the third option is to weight all rules with the inverse function (Hall et al., 2009).

After we applied the JCBA, WeightedClassifier and J48 algorithms on our dataset, we got the following results in Table 2.

Table 2. Classification algorithms results before enlarge data

Algorithm	Instances classified correct of. No	Accuracy	Error square mean root	Rules of. No	Precision. Avg	Recall. Avg	F. Avg
JCBA	17	7143.60	6268.0	2	657.0	607.0	589.0
Classifier weighted	17	7143.60	4949.0	16	657.0	607.0	589.0
J48	17	7143.60	5892.0	5	606.0	607.0	599.0

Results in Table 2 shows that the three classifiers have the same and slightly good accuracy with a different number of rules. So, we enlarged the data three times using 'SMOTE' filter, and then again applied the three classifiers. Table 3 shows the results after enlarging data.

Table 3. Classification algorithms result after enlarge data

Algorithm	Instances classified correct of. No	Accuracy	Error square mean root	Rules of. No	Precision. Avg	Recall. Avg	F. Avg
JCBA	64	0488.78	468.0	5	777.0	78.0	777.0
Classifier weighted	58	7317.70	49.0	22	703.0	707.0	680.0
J48	63	8293.76	46.0	7	780.0	768.0	771.0

As we see in results, the accuracy gets better after enlarging the data for the three classifiers and the number of rules also increase for WeightedClassifier and J48, but considering accuracy, JCBA classifier has the highest accuracy, so by these results, it can be approved that the used AC algorithms are good and we can use in case of knowing the most important inputs affect the efficiency of hospitals.

6. Conclusion

This paper shows the result of applying the DEA model using OSDEA software and MaxDEA tool in Jordanian health care systems to determine the efficient and inefficient hospitals. Then, the class label has been introduced to the original dataset as a class label to discover the relationship between the dataset attributes and class label. The rules have been discovered by using some association classification techniques in WEKA. These algorithms are JCBA and WeightedClassifier.

In general, all the algorithms have a good accuracy before and after enlarge data; however, the main measurement to compare these algorithms is the number of rules. WeightedClassifier method achieves the highest number of generated rules, while the JCBA method has the minimum number of generated rules.

References

Abolfathi, E., Zadeha, M. R. H. & Abolfathia, M. (2013). *Analyzing financial statements of listed companies in Tehran stock exchange with a hybrid model of data envelopment analysis (DEA) and artificial neural network (ANN)* (Doctoral dissertation, Master Thesis in industrial management).

Azizi, H., Kordrostami, S. & Amirteimoori, A. (2015). Slacks-based measures of efficiency in imprecise data envelopment analysis: an approach based on data envelopment analysis with double frontiers. *Computers & Industrial Engineering, 79*, 42–51.

Charnes, A., Cooper, W. W. & Rhodes, E. (1978), Measuring the efficiency of decisionmaking units, *European Journal of Operational Research, 2*, 429–444.

Cooper, W. W. (2013). Data envelopment analysis. In *Encyclopedia of operations research and management science* (pp. 349–358). USA: Springer.

de Andrade, G. N., Alves, L. A., da Silva, C. E. R. F. & de Mello, J. C. C. S. (2014). Evaluating electricity distributors efficiency using self-organizing map and data envelopment analysis. *IEEE Latin America Transactions, 12*(8), 1464–1472.

Han, J., Kamber, M. & Pei, J. (2011). *Data mining: concepts and techniques*. Elsevier.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter, 11*(1), 10–18.

Ji, Y. B. & Lee, C. (2010). Data envelopment analysis. *The Stata Journal, 10*(2), 267–280.

Martic, M., Novakovic, M. & Baggia, A. (2009). Data envelopment analysis-basic models and their utilization. *Organizacija, 42*(2), 37–43.

Mitrovic, Z., Vujosevic, M. & Savic, G. (2015). Data envelopment analysis for evaluating Serbia's health care system. *Management, 75*, 39–46.

Nadimi, R. & Jolai, F. (2008). Joint use of factor analysis (FA) and data envelopment analysis (DEA) for ranking of data envelopment analysis. *Language, 770*, 10908.

Najadat, H. & Alsmadi, I. (2011). An interactive DEA based management system for hospitals. *In Proceedings of the International MultiConference of Engineers and Computer Scientists, 2*.

O'Neill, L., Rauner, M., Heidenberger, K. & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences, 42*(3), 158–189.

Thabtah, F. (2007). A review of associative classification mining. *The Knowledge Engineering Review, 22*(01), 37–65.

Tone, K. & Ouenniche, J. (2016). DEA scores' confidence intervals with past-present and past-present-future based resampling. *American Journal of Operations Research, 6*(02), 121.

Wei, G., Chen, J. & Wang, J. (2014). Stochastic efficiency analysis with a reliability consideration. *Omega, 48*, 1–9.

(2017). Retrieved May 8 2017, from http://www.moh.gov.jo/Echobusv3.0/SystemAssets/58a5a982-55a1-463a-868e-45c77d9df09b.pdf