ISSN: 1813-4890

Evaluation of Port Efficiency based on DEA and Machine Learning Model

Weiqi Liu

School of Economics and Management, Chongqing University of Posts and Telecommunications, Chongqing 400000, China

Abstract

Evaluate the efficiency of 18 major ports in the Yangtze River Basin. First, the port efficiency index system is constructed by analyzing 80 port efficiency literatures through text mining. Second, the SBM-DEA-ML model is constructed by optimizing the traditional DEA model using machine learning algorithms. The smoothness of the new model is absolutely absolute. The efficient frontier has two advantages: one is that the smooth curve efficient frontier is more tolerant of outliers than the traditional linear efficient frontier, and makes up for the deficiencies that the efficient frontier of the DEA model is easily offset by statistical noise; The efficient frontier is more convenient for the horizontal and vertical comparison of port efficiency, and solves the efficiency evaluation conflict caused by the change of the relative effective frontier of the DEA model when a new evaluation unit is added.

Keywords

Ports in the Yangtze River Basin; Efficiency Evaluation; DEA Model; Machine Learning Model.

1. Introduction

Ports are a valuable strategic resource for promoting national and regional economic development, and an important support for promoting the development of water transport and improving the comprehensive transportation system. In recent years, the infrastructure construction of ports in the Yangtze River Basin has been further improved with the rapid development of the Yangtze River Economic Belt, but with it, the available port coastline resources in the Yangtze River Basin have also declined. Considering that the Yangtze River Economic Belt spans the three major economic zones of the east, the middle and the west, and involves the three major urban agglomerations of the Yangtze River Delta, the middle reaches of the Yangtze River, and Chengdu-Chongqing, although the number of ports is huge and the water, land and air are connected in all directions, due to the wide span, plus Due to socio-economic and technological factors, port efficiency is not ideal. In this context, it is particularly important to introduce a scientific and reasonable port efficiency measurement model to accurately measure the development level of ports in the Yangtze River Basin.

2. Research Model Building

2.1. Index System Construction

This paper uses text mining technology to count the word frequency data related to the index system in a large number of port efficiency literature, and builds a port efficiency index system based on the word frequency data. The steps are as follows: First, obtain the relevant literature on port efficiency required for word segmentation from CNKI, including 50 core journals of Peking University and above, and 30 master's thesis; secondly, use Python and Jieba libraries to set up custom dictionaries, stop word lists, and then carry out Jieba word segmentation, and

ISSN: 1813-4890

finally count the word frequency after word segmentation, and obtain the input-output indicators commonly used by scholars at present, so as to construct the port efficiency index system in this paper, so as to enhance the objectivity and scientificity of index selection. The final input index selects the length of production wharf, the number of berths for production, and the number of 10,000-ton berths, and the output index selects container throughput and cargo throughput.

2.2. Evaluation Model Construction

The disadvantage of CCR, BCC and SBM models is that the linear efficient frontier is easily affected by extreme values [1]. If there are some extreme values in the evaluation data, it will greatly affect the movement of the efficient frontier, which also shows that the relative efficiency is evaluated based on DEA [2].

Efficiency evaluation using the traditional DEA model is based on the relative efficiency of all evaluation units, and the relative efficient frontier is constructed [3]. When a new evaluation unit needs to be added, the efficient frontier needs to be rebuilt, and the overall efficiency value may change [4]. The evaluation unit will have different efficiency values under the calculation of the old and new efficient frontier, which makes it difficult for decision makers to compare; moreover, the efficient frontier of the traditional DEA model is linear or piecewise linear, and the efficient frontier is formed entirely according to the input and output of the evaluation unit, and is highly sensitive to outliers and statistical noise. If the data is affected by statistical noise, the DEA effective frontier is easily distorted, making it difficult to accurately evaluate the efficiency of the evaluation unit. In view of the above two problems, in terms of the relative efficient frontier, the machine learning algorithm can be used to construct an absolutely effective frontier, which can be effectively solved, avoiding the trouble of incomparability between the new and the old effective frontier of the same evaluation unit caused by the constant change of the effective frontier; In the aspect that the linear boundary is easily affected by outliers, the efficient frontier constructed by the machine learning algorithm is nonlinear[5]. Compared with the linear or piecewise linear efficient frontier of the traditional DEA model, the efficient frontier constructed by machine learning presents a smooth curve, which is more It is easy to tolerate outliers, which makes up for the deficiency of DEA that is easily affected by outliers.

The idea of constructing a smooth absolute efficient frontier in this paper: firstly, a relatively efficient frontier is constructed through the SBM-DEA model; secondly, all evaluation units are adjusted to the efficient frontier based on the relaxation measured by the SBM-DEA model; finally, the machine learning algorithm is used Constructing the regression model establishes a smooth absolute efficient frontier.

3. Empirical Analysis

3.1. Data Sources

The training data set used to build the SBM-DEA-ML model is the input-output data of 18 major ports in the Yangtze River basin from 2005 to 2020. The input index data is mainly from the "China Statistical Yearbook", and the output index data mainly comes from " China Port Yearbook and Statistical Bulletin of National Economic and Social Development of each port city.

3.2. Model Training

This paper uses the input-output data adjusted by redundant variables as the dataset to train the SBM-DEA-ML model to construct a new absolute efficient frontier. When modeling, in order to prevent the model from overfitting, the data set was randomly divided by 4:1 as a training set and a test set, and then six machine learning algorithms were used for training and testing,

ISSN: 1813-4890

and finally the one with the best performance was selected. The algorithm builds the SBM-DEA-ML model.

In order to comprehensively compare the performance of each model, this paper uses three metrics to evaluate different machine learning (ML) algorithms, namely mean square error (MSE), mean absolute error (MAE), and R2 (coefficient of determination, which reflects the full variation of the dependent variable) The proportion that can be explained by the independent variable through the regression relationship, the closer to 1, the better the regression effect).

The results of random training of six machine learning models are shown in Table 1.

Table 1. Machine Learning Model Accuracy Comparison

Numble	Model	MSE	MAE	R2
1	Extra Trees Regressor	116951.6631	151.1074	0.9916
2	Decision Tree Regressor	159361.1864	210.3146	0.9788
3	Random Forest Regressor	416290.8330	294.9262	0.9828
4	Ridge	950158.8627	526.6058	0.9251
5	LinearRegression	2665328.9445	715.9417	0.7239
6	BP Neural Networks	14910119.4093	2281.4387	0.2983

According to the results in the table, the top 4 model fitting effects are extreme random forest (ETR), decision tree (DTR), random forest (RFR), and ridge regression (R). Among them, the extreme random forest (ETR) mean square error (MSE) is 116951.6631, the mean absolute error (MAE) is 151.1074 and the R2 is 0.9916, indicating that the ETR model is superior to other algorithms in terms of result prediction accuracy and model revealing ability.

The performance of extreme random forest (ETR) on the test set is shown in Figure 1.

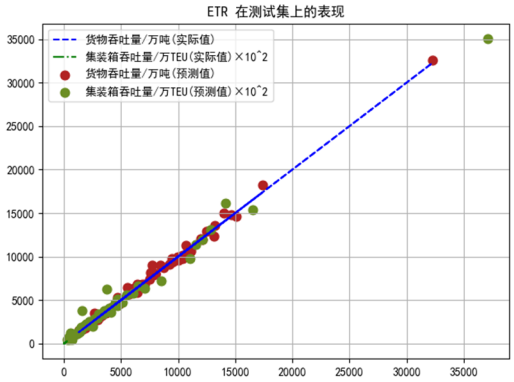

Fig 1. The performance of the ETR model on the test set

The residual distribution of the ETR model on the test set is shown in Figure 2.

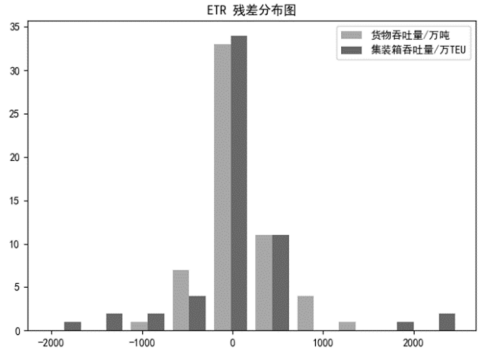

Fig 2. Residual distribution of the ETR model in the test set

ISSN: 1813-4890

4. Summary

The text mining technology is used to analyze a large number of port efficiency literatures to construct the port efficiency index system. Secondly, the SBM-DEA-ML efficiency evaluation model is constructed by combining the machine learning model and the DEA model to overcome the shortcomings of the traditional DEA model. The research conclusions are as follows.

The traditional DEA model builds a relative efficient frontier. When a new evaluation unit is added, the same evaluation unit may cause efficiency conflicts, and the linear efficient frontier of the traditional DEA model is easily affected by outliers and causes distortion, which makes the efficiency evaluation biased. In this paper, the new SBM-DEA-ML model formed by the combination of the extreme random forest model and the SBM-DEA model is constructed as an absolute efficient frontier, which can solve the efficiency value conflict problem caused by the continuous change of the effective frontier of the traditional DEA model. The efficient frontier is a smooth curve, which makes up for the shortcoming that the linear boundary is easily affected by outliers. The new SBM-DEA-ML model improved by combining the machine learning model obtains a lower absolute efficiency value than the traditional model, overcoming the shortcomings of the linear effective frontier and relative efficiency, which not only truly reflects the efficiency of the ports in the Yangtze River Basin, but also facilitates the ports. Horizontal and vertical comparison of efficiency.

References

[1] Charnes A, Cooper W W, Rhodes E. Measuring the efficiency of decision-making units[J]. European journal of operational research, 1978, 2(6): 429-444.

[2] Tone K. A slacks-based measure of efficiency in data envelopment analysis[J]. European journal of operational research, 2001, 130(3): 498-509.

[3] Roll Y, Hayuth Y. Port performance comparison applying data envelopment analysis (DEA)[J]. Maritime policy and Management, 1993, 20(2): 153-161.

[4] Elsayed A, Khalil N S. Evaluate and Analysis Efficiency of Safaga Port Using DEA-CCR, BCC and SBM Models–Comparison with DP World Sokhna[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2017, 245(4): 042033.

[5] Liu C C. Evaluating the operational efficiency of major ports in the Asia-Pacific region using data envelopment analysis[J]. Applied economics, 2008, 40(13): 1737-1743.