

# Data-driven integration evaluation from the perspective of Adaboost and its application in WeChat public number ranking

Yicheng Gong<sup>1</sup>, Juan Zhao<sup>1</sup>, and Dongyang Zhang<sup>1,\*</sup>

<sup>1</sup>Wuhan University of Science and Technology School of Science, Hubei, Wuhan, China

**Keywords:** Integrated evaluation, AdaBoost, Data driven, WeChat subscription.

**Abstract.** The traditional comprehensive evaluation is difficult to model when dealing with large data with large parameters and complex structure, and it cannot adapt to the update of data. In order to improve this situation, this paper draws on the Adaptive Learning Adaboost perspective in statistical learning to develop a data-driven integrated evaluation model that updates the weight of sample weights and weak evaluation models with data. Three specific weak evaluation models were selected: data-driven Topsis method, principal component analysis method and factor analysis method. Taking the ranking of WeChat public account as an example, the results show that the accuracy of the integrated evaluation model is 88.57%, which is 17.14%, 31.43% and 28.57% higher than the data-driven Topsis method, principal component method and factor analysis method.

## 1 Preface

In order to effectively use data to service production and life, scientific evaluation methods are needed to help people make correct decisions from a large number of choices. Traditional comprehensive evaluation has two shortcomings when dealing with big data with many parameters and complicated structure: first, it is difficult to model when it encounters data with too many parameters and too complicated structure; second, with the update of the data, the model is accurate The degree will also get lower and lower. Therefore, the effective use of the big data environment to improve evaluation methods is a need to keep pace with the times.

Scholars at home and abroad have conducted many studies on integrated evaluation. For example: grey statistical method and entropy technology combined with fuzzy method (2007, Pang Qinghua)<sup>[1]</sup>, data envelopment analysis method (2009, Wu DD)<sup>[2]</sup>, analytic hierarchy process combined with approaching ideal point method (Topsis) (2010, Yao Shengbao)<sup>[3]</sup>, fuzzy comprehensive evaluation (2013, Cai Zexiang, Liu Ye)<sup>[4]</sup>. Although these studies used the integration method, they were not integrated from a data-driven perspective. In this paper, the Boosting method is used to construct an integrated evaluation model by combining the

---

\* Corresponding author: [1044600738@qq.com](mailto:1044600738@qq.com)

Topsis method commonly used in multi-attribute decision-making evaluation methods with the principal component analysis method and factor analysis method to adapt to multiple indicators, in order to improve the above integration methods, which are easily limited by data types and quantities. Deficiencies.

With the popularity of big data today, the data-driven thinking has been applied to various industries for statistical evaluation. In the industrial field, the genetic algorithm (GA) and factor analysis method are used to optimize the evaluation indicators, and a data-driven construction safety evaluation model is finally combined with the Bayesian network (BN) to give a detailed quantitative evaluation procedure of the model (2017, Bai Xiaoping and Pu Tao) [5]; In the business field, a data-driven model and neural network method were used to establish a comprehensive carrying capacity evaluation model to evaluate the comprehensive carrying capacity of the Tianjin coastal area (2012, Liu Aizhen, Si Qi and Li Mingchang) [6] and data-driven Topsis method (2018, Zhao Juan, Gong Yicheng) [7]; in the medical field, a data-driven method is used to evaluate and improve the strict blood glucose control scheme in the intensive care unit (2016, Kuiper BI With Number SI) [8].

At present, there are not many evaluations using data-driven ideas in the social field. Therefore, this article selects 70 public accounts in WeChat from March 5, 2018 to March 11, 2018 as the research object. A total of seven important indicators were selected-the number of posts (articles), the total reads (thousands), the headline reads (thousands), the average reads (thousands), the highest reads (thousands), the number of likes (thousands) And the heat index to measure the market influence of the major WeChat public accounts (all data in this article were collected by the octopus collector).

## 2 Integrated evaluation model from AdaBoost perspective

The integrated evaluation model based on AdaBoost idea is to use the data-driven Topsis method, principal component analysis method and factor analysis method to train the three comprehensive evaluation methods as weak evaluation models, and then calculate their respective weights according to their error rates, and finally get A strong evaluation model composed of these three weak evaluation models.

The experiment was performed in the following six steps:

Step 1. Prepare the target dataset.

Step 2. Standardize the data.

Step 3. Take the three evaluation methods of data-driven Topsis method, principal component analysis method and factor analysis method as three weak learners (denoted as  $f_1(x), f_2(x), f_3(x)$ ). The data are evaluated to obtain scores and ranking results.

Step 4. Based on the results of step 3 compared with the actual list, calculate the evaluation error rate  $e_k$  (where  $k = 1, 2, 3$  weak learners), and obtain the weight  $\alpha$  of each comprehensive evaluation method by the following formula.

$$\alpha_k = \frac{1}{2} \left| \log \frac{1-e_k}{e_k} \right| \dots \dots \dots (1)$$

Step 5. Get the final learner-linear integrated evaluation model  $G(x)$ :

$$G(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) \dots \dots \dots (2)$$

Step 6. Test and compare the final integrated evaluation model. First, it compares with each single comprehensive evaluation method to verify the validity of the model; then compares with the traditional combination method to verify the superiority of the combination method.

### 3 Evaluation of the influence of the integrated model on WeChat public account from the perspective of AdaBoost

Use SPSS software to standardize the scores of the three comprehensive evaluation methods.  $f_1(x), f_2(x), f_3(x)$  referred to as  $f_1^*(x), f_2^*(x), f_3^*(x)$ .

Based on the AdaBoost idea, the three weak evaluation models are integrated. First, the weights of  $f_1^*(x), f_2^*(x)$  and  $f_3^*(x)$  are obtained according to the formula:  $\alpha_1, \alpha_2, \alpha_3$ , and then linear Combination to get the final integrated evaluation model  $G(x)$  as

$$G(x) = 0.199f_1^*(x) + 0.062f_2^*(x) + 0.088f_3^*(x) \dots \dots \dots (3)$$

Finally, calculate the value of  $G(x)$ , and finally get the strong evaluation model. The final results and rankings are shown in Table 1.

**Table 1.** Evaluation results and comparison of WeChat public account by integrated model.

	Public Number Ranking	$G(x)$	Actual Ranking
1	People's Daily	1.79	People's Daily
2	Xinhua News Agency	1.32	Xinhua News Agency
.....	.....	.....	.....
69	Daily Music	-0.28	Daily Music
70	Fashion COSMO	-0.41	Fashion COSMO

It can be seen from Table 1 that in the evaluation of the influence of the integrated evaluation model on WeChat public accounts, there are 62 correctly ranked public accounts with an error rate of 11.43%.

### 4 Comparison between integrated evaluation and three weak evaluators on WeChat public account

#### 4.1 Evaluation of WeChat public account by data-driven topsis method

The specific ranking results are shown in Table 2.

**Table 2.** Evaluation of WeChat public account by data-driven Topsis method.

	Public Number Ranking	Actual Ranking
1	People's Daily	People's Daily
2	Xinhua News Agency	Xinhua News Agency
.....	.....	.....
69	Daily Music	Daily Music
70	Fashion COSMO	Fashion COSMO

The score and ranking of the data-driven Topsis method, among which there are 50 correctly ranked apps, with an error rate of 28.57%.

#### 4.2 Evaluation of WeChat public account by principal component analysis

The specific results are shown in Table 3.

**Table 3.** Evaluation of WeChat public account by principal component analysis.

	Public Number Ranking	Actual Ranking
1	People's Daily	People's Daily
2	Xinhua News Agency	Xinhua News Agency
... ..	... ..	... ..
69	World of Poetry	Daily Music
70	Ten Point Digest	Fashion COSMO

Using SPSS software for comprehensive evaluation of principal components, there were 40 correct rankings with an error rate of 42.86%.

### 4.3 Evaluation of WeChat public account by factor analysis

The specific results are shown in Table 4.

**Table 4.** Evaluation of WeChat public account by factor analysis.

	Public Number Ranking	Actual Ranking
1	People's Daily	People's Daily
2	Xinhua News Agency	Xinhua News Agency
... ..	... ..	... ..
69	China Search	She published
70	Reader	Reader

Using SPSS software for comprehensive evaluation using factor analysis, there were 42 correct ones with an error rate of 40%.

### 4.4 Comparative analysis and conclusion of integrated evaluation

This paper compares the ranking results of the data-driven integrated evaluation model from the perspective of Adaboost with the data-driven Topsis method, principal component analysis method, and factor analysis method. The specific data are shown in Table 5.

**Table 5.** Comparison of the results of the four evaluation methods.

Evaluation method	Data-driven Topsis	Principal Component Analysis	Factor Analysis	Integrated evaluation method
Number of correct evaluations	50	40	42	62
Correct rate	71 43%	57 14%	60%	88 57%

It can be seen that the accuracy rate of the evaluation method integrated in this paper is the highest, reaching 88.57%, which is greatly improved compared with the other three weak evaluator evaluation methods. The comparison results show the feasibility of the integrated method and the effectiveness of the model improvement.

## 5 Conclusion

This article builds an integrated evaluation model from the perspective of AdaBoost based on the data-driven concept and the idea of improving integration. The data-driven Topsis method, principal component method, and factor analysis method are given weights in the integrated model based on the error rates of their respective evaluation results. Apply the

integrated evaluation model to the impact evaluation of WeChat public account. The results show that the model accuracy is 88.57%, which is 17.14%, 31.43%, and 28.57% higher than the data-driven Topsis method, principal component analysis method, and factor analysis method, respectively, indicating the feasibility and effectiveness of the integrated evaluation method.

However, there are still two shortcomings in this article: First, in this paper, only three weak evaluation methods are selected in the evaluation model integration, but more types of comprehensive evaluation methods are selected to integrate the evaluation of complex structure data, which is expected to be more comprehensive. Accurate integrated evaluation model; the second is to consider applying the integrated evaluation model in different fields, which is expected to enhance the application value of the model.

## References

1. Pang Qinghua. An integrated evaluation method of enterprise informationization level [J]. *Statistics and Decision*, 2007 (2): 154-155.
2. Wu D D. Performance evaluation: An integrated method using data envelopment analysis and fuzzy preference relations [J]. *European Journal of Operational Research*, 2009,194 (1): 227-235.
3. Yao Shengbao. Decision-making of Enterprise Technology Innovation Project Based on Integrated Evaluation Method [J]. *Science Research Management*, 2010, 31 (5): 102-107
4. Cai Zexiang, Liu Yan. Integrated Evaluation of the Operational Performance of New Rural Financial Institutions: An Empirical Analysis Based on Jiangsu Province [J]. *Audit and Economic Research*, 2013 (2): 89-96.
5. Bai Xiaoping, Pu Tao. Research and application of data-driven construction safety evaluation model [J]. *Chinese Journal of Safety Science*, 2017, 27 (9): 134-139.
6. Liu Aizhen, Si Qi, Li Mingchang. Research on Evaluation Method of Comprehensive Carrying Capacity of Tianjin Coastal Area Based on Data-Driven Neural Network Model [J]. *Journal of Water Resources and Water Engineering*, 2012, 23 (4): 83-85.
7. Juan ZHAO, Yi-Cheng GONG. Data Driven Topsis Method and Its Application in E-commerce APP.ICBDC 2018 [C]. 2018 International Conference o Big Data and Computing.
8. Kuiper BI, Number SI. A data driven approach to evaluate and improve the protocol for strict glucose control in the Intensive Care Unit. [D]. Eindhoven: Eindhove University, 2016.