# APPLICATION OF BOOSTING CLASSIFICATION ALGORITHM IN TOURISM SYSTEM

**Xiang Nan[1*], Assoc. Prof. Ts. Dr. Rajamohan Parthasarathy[2]**

**Abstract—**
Tourism industry is one of the most powerful and largest industries in the current global economy, which has created huge economic benefits. With the rapid development of China's tourism industry, tourism information has become an important strategic resource, and the development of e-commerce has provided important support for tourism informatization, which makes its positive role in the value chain of the tourism industry increasingly prominent. In order to improve the service level of tourism e-commerce, a tourism system based on Boosting classification algorithm is constructed, and the random gradient descent algorithm, adaptive gradient optimization algorithm and density method are used to optimize and improve the Boosting classification algorithm. The experimental results show that compared with the traditional Kmeans classification, the improved Boosting classification algorithm can increase the browsing volume by 28.21% and the purchase volume by 15.37%. Through the application of Boosting classification algorithm, it can provide users with personalized tourism goods, improve their purchase volume, and thus improve the competitiveness of the tourism market.

**Keywords—** Classification algorithm; Boosting classification algorithm; Tourism system; Tourism products; Traditional classifier

[1*]PhD Research Scholar, Faculty of Engineering, Built Environment and Information Technology SEGi University Kota Damansara, Malaysia
[2]PhD Research Supervisor, Faculty of Engineering Built Environment and Information Technology Segi University Kota Damansara, Malaysia prajamohan@segi.edu.my 446717589@qq.com.

***Correspondence Author:-** Xiang Nan
[*]PhD Research Scholar, Faculty of Engineering, Built Environment and Information Technology SEGi University Kota Damansara, Malaysia

**DOI: -** 10.48047/ecb/2023.12.si5a.0246

## I. INTRODUCTION

In recent years, with the rapid development of our economy and the improvement of people's quality of life and living standards, travel consumption has become an important part of our people's daily life. With the emergence of e-commerce and the development of classification algorithm, a new consumption mode has emerged in tourism services. However, the search process of most tourism systems is tedious and the product recommendation interface is single, which is difficult to satisfy users [1].

At present, there are many researches on the combination of classification algorithm and e-commerce at home and abroad, mainly using classification algorithm, classification algorithm and correlation analysis. However, the current classification algorithm still has some defects, its results are interfered by the data set, and the segmentation accuracy of the multi-index group is not high. In order to improve the service quality of its users, it is necessary to further strengthen the data mining. In this paper, Boosting classification algorithm is optimized and improved to improve the analysis ability of Boosting classification algorithm, and it is used in tourism system services. It is hoped that the method of data analysis and integration can be used to provide personalized tourism product recommendation interface for tourists, so as to improve the service quality of tourism system
[2].

## II.SYSTEM DESIGN

A. Boosting tourism e-commerce platform based on Boosting classification algorithm.

Classification algorithm is to use the database for automatic data retrieval, the existing data induction, sorting, reasoning, and mining valuable knowledge from the data, so as to analyze the overall trend; Only in this way can we reasonably predict and make future changes. On this basis, a tourism system platform is established by using the classification algorithm, which can predict and judge the tourists according to their relevant information, and provide them with personalized services. It can not only help consumers quickly find the tourism commodities they need, but also improve their sales performance and enhance their competitiveness in the industry. Through the analysis of the visit records of the users on the tourism e-commerce platform, the characteristics of their browsing preferences are dug out and classified as the browsing attributes based on products. In this way, consumers can be provided with personalized travel products, and users' browsing experience can be optimized on the platform to help them quickly find the right travel products for them.

Classification algorithm is one of the most common methods in data mining. It divides data objects into multiple classes or clusters according to attributes and other relations. Data objects belonging to the same class or cluster have some similarity. The similarity function is used to divide the similarity of data objects.

Data set: includes n data samples of p dimensions, and the data matrix of the data set is expressed as follows:

$$\mathbf{S}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

Firstly, the classification center of the data set is selected, and on this basis, k classification points are randomly selected, and the Euclidean distance function d between them and the classification center is expressed as follows.

$$d(x_i,x_j)=\left(\left|x_{i1}-x_{j1}\right|^2+\left|x_{i2}-x_{j2}\right|^2+\cdots+\left|x_{ip}-x_{jp}\right|^2\right)^{\frac{1}{2}} \quad (2)$$

The difference matrix C of data objects is expressed as follows:

$$\mathbf{C}=\begin{bmatrix} 0 & & & & \\ d(x_2,x_1) & 0 & & & \\ d(x_3,x_1) & d(x_3,x_2) & 0 & & \\ \vdots & \vdots & \vdots & 0 & \\ d(x_n,x_1) & d(x_n,x_2) & d(x_{2n},x_2) & \cdots & 0 \end{bmatrix} \quad (3)$$

Using the average method, the classification points of each category are re divided to obtain a classification group. The algorithm iterates repeatedly until k classification groups do not change.

B. Optimization of Loss Function of Boosting Classification Algorithm

The selection of learning rate will have a certain impact on the convergence rate, and an appropriate learning rate can improve the convergence of the algorithm. In the later stage of training, the minimum value is ignored in order to prevent the parameters from converging too fast. Since the traditional Boosting classification method uses a fixed learning rate, it would be easy to boost the algorithm jitter, which would affect the performance of the algorithm [3]. On this basis, the exponential decay mean of historical gradient is

combined with the square exponential decay mean of historical gradient, and the attenuation learning rate of this method is obtained by using the adaptive method. This method uses the gradient mean and the deviation method to estimate the moving average, and uses the deviation correction method to reduce the initialization error and improve the operability of the algorithm. The formula for the average attenuation of the first and second order gradients is as follows:

$$\begin{cases} m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \\ v_t = \beta_2 m_{t-1} + (1-\beta_2) g_t^2 \end{cases} \quad (4)$$

In the formula, m and v represent first-order momentum and second-order momentum respectively; mt is mean estimation; vt means there is a bias error estimation; β1 represents the first-order momentum attenuation coefficient; β2 represents the second-order momentum attenuation coefficient; t represents the initialization time step; gt stands for partial derivative vector [4].

In the initialization process, m and v are taken as 0 vectors respectively, so mt and vt will deviate from 0, thus affecting the performance of the algorithm. Therefore, the offset correction mechanism is added into the adaptive gradient optimization algorithm, so that the learning rate of each iteration can maintain in the given interval. The correction function for the average and biased estimates looks like this [5]:

$$\begin{cases} \hat{m}_t \leftarrow m_t / (1-\beta_1^t) \\ \hat{v}_t \leftarrow v_t / (1-\beta_2^t) \end{cases} \quad (5)$$

In the formula, mˆ t represents the corrected mean value estimation; vˆt represents the biased error estimation after correction; $\beta_1^t$ represents the first-order momentum attenuation coefficient after correction; $\beta_2^t$ represents the corrected second-order momentum attenuation coefficient [6].

In Boosting classification algorithm, an adaptive gradient optimization algorithm based on exponential decay learning rate is introduced in order to improve the convergence of Boosting classification algorithm [7].

C. Boosting classification algorithm initial center point optimization

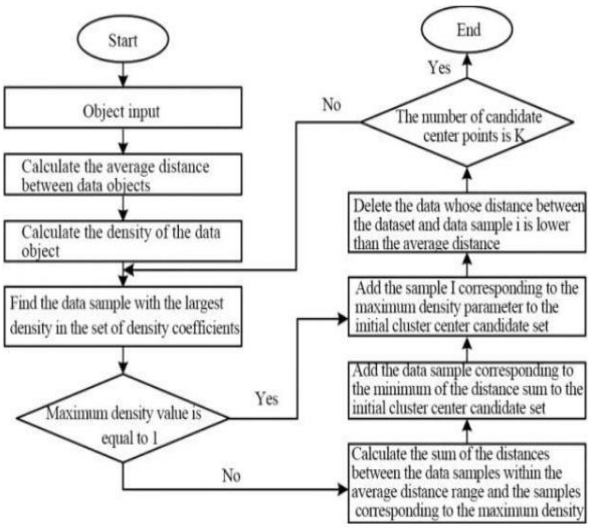


**Fig.1.** Optimization process of initial clustering center based on density method.

In the traditional Boosting classification method, the initial value of the method will directly affect the final result of the method because they are highly dependent on the location of the initial classification center point [8].

On this basis, density method was used to optimize the initial positioning of k classification points, and combined with density parameters to determine k initial classification points. Boosting classification algorithm optimization process based on density method is shown in Figure 1.

In data sample set: $\mathbf{S} = \{x_1, x_2, \cdots, x_n\}$ the average distance calculation function between data objects is shown as follows:

$$M(S) = \frac{2}{n(n-1)} \times \sum_{i \neq j, i, j=1}^{n} d(x_i, x_j) \quad (6)$$

In the formula, M(S) represents the average distance; d(xi, xj) represents the distance between data object xi and data object xj [9].

III. Experiment and Discussion

Through the improvement of traditional Boosting classification and Boosting classification algorithm, the response speed of the two methods is compared in 20 different clustering experiments, and the comparison results of the two methods are given, as shown in Figure 2.

Asshownin Figure 2, theconventional Boosting classification algorithm has an average response time of 0.724 seconds, and the maximum response time is 0.861 seconds, while the traditional Boosting classification algorithm has a lower operation speed and a longer response time. The improved Boosting classifier is between 0.498 s and 0.647 s.

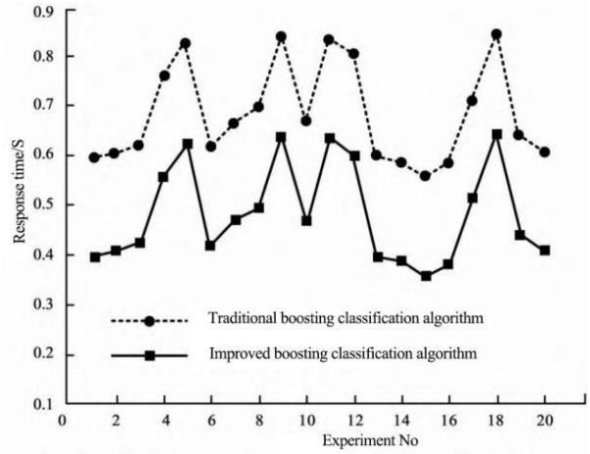


**Fig.2.** Performance comparison of Boosting classification algorithm before and after improvement.

Through the optimization of Boosting classification, a better Boosting classification method is obtained. The average response time is reduced by 0.226 s, and the system response speed is improved by 31.2%. On this basis, the random gradient descent method is adopted, the regularization term is introduced, and the adaptive gradient optimization method is used to determine the learning rate, so as to improve the operation efficiency of Boosting classification and accelerate the convergence of the algorithm; Therefore, the response speed of Boosting classification method is improved, and its performance is also greatly improved [10].

Through online experiments, the improved Boosting classification algorithm was personalized recommended and statistically analyzed. The data statistics of the tourism e-commerce platform based on the improved Boosting classification algorithm are shown in Table 1.

**Table I** Data Of Tourism E-Commerce Platform Based On Improved Boosting Classification Algorithm

| **Weeks** | Week 1 | Week 2 | Week 3 | Week 4 | Aver age value |
|---|---|---|---|---|---|
| **Platform views** | 1026 8 | 1254 7 | 1296 4 | 1287 3 | 1216 3 |
| **Recommen ded product views** | 1518 | 1895 | 2 048 | 2035 | 1816 |
| **Tourism product purchases** | 159 | 201 | 211 | 207 | 195 |
| **Recommen ded product purchases** | 40 | 57 | 63 | 61 | 55 |
| **Recommen ded click rate of browsing behavior** | 14.78% | 15.10% | 15.80% | 15.81% | 15.37% |
| **Recommen ded transaction rate of** | 25.16% | 28.36% | 29.86% | 29.47% | 28.21% |
| **Purchase behavior** | | | | | |
| **Recommen ded traffic conversion rate** | 2.64 % | 3.01 % | 3.08 % | 3.00 % | 2.93 % |

As shown in Table 1, in four weeks, the average weekly platform visits of the travel e-commerce platform based on the Boosting classification algorithm reached 12163, while the weekly visits of the tourism products personalized recommended through the Boosting classification algorithm reached 1874. On the platform, 15.41% of users were recommended in their browsing behavior. On the tourism e-commerce platform, consumers bought an average of 194.5 commodities every week, of which 55.25 were recommended for tourism. Among consumers, 28.41% chose tourism commodities.

This method combines the improved Boosting classification algorithm with clustering analysis to make personalized recommendations. The average weekly recommendation traffic conversion rate is 29.48%; The personalized recommendation can effectively promote the order transaction on the e-commerce platform. Using the improved Boosting classification algorithm, Apriori association rule mining algorithm, user collaboration screening algorithm and other methods, the online production of tourism e-commerce platform is studied [11]

.

**Table II** Sales Of Recommended Tourism Products By Three Algorithms

| **Algorithms** | **Browsing behavior Recomme ndation rate** | **Purchase behavior Recomme ndation rate** | **Recomm ended conversi on rate** |
|---|---|---|---|
| User—based CF | 12.95% | 20.47% | 2.36% |
| Apriori | 13.44% | 23.63% | 2.58% |
| Improved Boosting classification | 15.37% | 28.21% | 2.93% |

As shown in Table 2, on the basis of Boosting classification, the recommended conversion rate of this method is 2.93%, which is higher than 2.58% of Apriori and 2.36% of User based CF, while the recommended conversion rates are 0.35% and 0.57% higher respectively. The improvement of Boosting classification algorithm can provide consumers with differentiated tourism commodities, thus effectively increasing the sales volume of e-commerce platform and improving the economic benefits of tourism enterprises.

## IV. CONCLUSION

With the emergence of e-commerce, the marketing model of tourism products has also changed, and online booking has become a trend. In order to improve the service quality of tourism system, a new tourism system platform is established by using Boosting classification algorithm. The experiment shows that the improved Boosting classifier has a 31.2% increase in response time of 0.498 s compared with the traditional classifier, and has good performance. Through the improvement of
Boosting classification, the recommended traffic conversion rate of the tourism platform has reached 2.93%, which has been greatly improved compared with other algorithms.

## REFERENCES


1. Shao Mengliang, Qi Deyu. An improved random forest Boost multi label text classification algorithm. Computer Application and Software, 2022,39 (11): 215-221+303
2. Zhang Xilong, Han Meng, Chen Zhiqiang, Wu Hongxin, Li Muhang. Boosting Classification Algorithm for Unbalanced Drift Data Streams Based on
3. Hellinger Distance. Computer Engineering and Science,
4. 2022,44 (05): 788-799
5. Wang Zhiqiang, Wang Xianchuan. Design of a depth learning image classification algorithm based on Boosting. System Simulation Technology, 2021,17 (02):
6. 109-113
7. Du Shiyu, Han Meng, Shen Mingyao, Zhang Chunyan, Sun Rui. An Iterative Weighted Integrated
8. Classification Algorithm Based on Boosting. Computer Application Research, 2021,38 (04): 1038-1043
9. Qiao Yuan, Liao Xiaoping, Shao Kaixia. Cancer classification based on jump pattern mining algorithm. Computer and Modernization, 2018 (05): 100-105
10. Wang Lai, Fan Chongjun, Yang Yunpeng, Yuan
11. Guanghui. KFDA Boosting Algorithm for Unbalanced Data Classification. Computer Application Research, 2019, 36 (03): 807-811
12. Wang Shixun, Pan Peng, Chen Deng, Lu Yansheng. An Adaptive Multi class Boosting Classification Algorithm.
13. Computer Science, 2017,44 (07): 185-190
14. Huang Bin. Research on Chinese Emotion Classification Based on Boosting Algorithm. Electronic Technology and Software Engineering, 2017 (12): 190-191
15. Xiong Wei. Research on News Text Classification Based on Boosting Algorithm. Electronic Technology and Software Engineering, 2017 (08): 174-175
16. Li Yijing, Guo Haixiang, Li Yanan, Liu Xiao. Classification of an Integrated Learning Algorithm Based on Boosting in Unbalanced Data. System Engineering Theory and Practice, 2016,36 (01): 189-199
17. Chen Wei, Ren Daowen, Pang Meiyan. Design of Online Education Information Management System Based on Data Mining Algorithm[J]. Wireless Communications and Mobile Computing, 2022,2022.