Prediction Model for Club Recruitment: Time Series Algorithm in Recruitment Data

Chen,Meng[1,2]

1School of Arts and Social Sciences, Hong Kong Metropolitan University, Hong Kong, 999077, China
2Xi'an University, Xi'an, Shaanxi, 710065, China

Abstract: Club recruitment activities are an important way for universities to carry out campus cultural construction and student organizations,and are an important component of campus culture.However, due to the randomness of club members, there are problems with the accuracy of predicting the number of applicants and the effectiveness of resource allocation in predicting the number of university club members during the recruitment process. This article conducts a questionnaire survey analysis on students' acceptance of club activities and their participation in club activities at different time points.It utilized the Autoregressive Integrated Moving Average Model (ARIMA) in time series analysis algorithms to design a data prediction model for club recruitment, and analyzed the university club recruitment model. Research has shown that the club recruitment prediction model designed based on ARIMA in this article can achieve a maximum accuracy of 96.7% and optimize resource allocation.

Keywords: Club recruitment; Autoregressive moving average model; Residual analysis; Resource allocation optimization

DOI: 10.62639/sspjinss03.20240104

1. Introduction

Club recruitment refers to activities carried out among college students based on interests and aimed at exercising abilities. It is one of the important forms of campus cultural construction and student organization work.The prediction model for club recruitment can help club managers more accurately understand and predict the trend of new recruitment,thereby formulating relevant countermeasures.Accurate prediction is beneficial for optimizing resource allocation and accurately planning event planning and promotional strategies. However, in the process of recruiting new members for clubs, club activities are often influenced by many factors, and the instability of these factors often leads to low accuracy in predicting data. How to achieve accurate prediction from historical data in club recruitment activities is a meaningful and valuable research direction.In addition, the problem of poor allocation of prediction model sources also urgently needs to be solved.

Through the analysis and comparison of time series analysis algorithms, this article selects the ARIMA model and applies it to the prediction of recruitment data to achieve the construction of the prediction model. In order to test the reliability of the model, residual analysis was conducted in this paper, and the influencing factors of the community recruitment prediction model were analyzed by designing variables. In actual data collection, due to the different participation of students in club activities in different universities, there may also be some deviations in the collected club information. This article collected data from a total of 1000 samples of students in universities from 2020 to 2022. It mainly includes samples that have participated in club activities and samples that have not participated in club activities but have participated in other student organization activities. In the process of data

(Manuscript NO.: JINSS-24-4-80001)

About the Author

Chen,Meng (1993-), Master, School of Arts and Social Sciences, Hong Kong Metropolitan University, Xi'an University, Reseach Area: Drama Education, Student Affairs Management Research.

collection, this article uses the China University Student Association Data Platform to obtain data and preprocesses it through sample classification to ensure the accuracy and consistency of the data. Afterwards, the data can be applied to the prediction model, and the traditional moving average algorithm can be compared with the prediction model based on ARIMA model. It includes data analysis on the accuracy, efficiency, and resource allocation of prediction results to evaluate the predictive performance of the model.

2. Related Works

With the development of big data, data-driven decision-making is increasingly being applied in daily life. In many fields, time series analysis algorithms are widely used for predicting and analyzing data with temporal dependencies due to their advantages such as high prediction accuracy, strong adaptability, and flexibility. In the planning and implementation process of club recruitment activities, it is necessary to predict the number of new recruits to achieve efficient allocation of resources and promote smooth recruitment activities.

In recent years, there are some scholars who have studied and analyzed this issue. Through the research on the recruitment problems of innovation and entrepreneurship associations in higher vocational colleges, Zhang Qiang analyzed the problems of innovation and entrepreneurship associations in college students, and gave the corresponding recruitment countermeasures [1]. In order to improve the administration and information communication of the association, Zhang Wenjie discussed the organization platform of university campus community information release platform based on the model-view-controller architecture mode design. The research shows that it is of great significance to use scientific and technological means to optimize the association's activities and improve the operational efficiency [2]. On the basis of the analysis of the characteristics and limitations of university associations, Zhang Heming put forward the strategy to promote the healthy development of associations, and emphasized that the development of associations needs to comprehensively consider the internal and external environment [3]. Kumar R integrated ARIMA and evolutionary algorithm. A multi-stage time series data analysis and prediction method based on ARIMA is established. It is found that combining evolutionary algorithm can optimize ARIMA model parameters and improve the prediction accuracy [4]. Li S introduced the timing analysis applied to remote sensing images, and proposes some countermeasures for their applicability in processing and algorithm [5]. Ren Weijie reviewed causal relationships in multivariate time series and expanded their application from individual predictions to the study of system dynamics and causal relationships [6]. Wang X proposed a new method for processing ultra long time series data using a distributed ARIMA model, demonstrating the potential application of time series analysis in big data environments. Research has shown that distributed ARIMA models can effectively process and analyze time series data in large datasets [7]. Schaffer A L provided practical guidance for evaluating large-scale health interventions by using the ARIMA model for interruption time series analysis. Research has shown that the ARIMA model can promote evaluation efficiency [8]. Deif M A emphasized the importance of combining traditional statistical models with modern optimization algorithms [9]. Xu D explored the application of ARIMA-LSTM hybrid models in environmental science. Especially in drought prediction, research has shown that by integrating traditional time series analysis models with advanced deep learning techniques, the accuracy and efficiency of predictions can be significantly improved [10]. The research of these scholars indicates that ARIMA models have applications in multiple fields, and the use of ARIMA models in drought prediction can improve prediction accuracy. In addition, combining ARIMA models with evolutionary algorithms can optimize model parameters and improve prediction accuracy. There is relatively little research on the ARIMA model in predicting club recruitment. This article applies the ARIMA model to predict club recruitment activities, thereby utilizing the characteristics of the ARIMA model to optimize and improve the prediction accuracy of club recruitment activities.

3. Methods

(1) Time series analysis algorithm

Time series analysis is the study of historical data to predict future trends and behaviors. In the traditional prediction analysis, there is moving average method and trend prediction method. The former is a series of data with the average of a period of time, which is suitable for short-term data analysis. However, the moving average method is not suitable for long-term data. Moreover, the moving average method needs to calculate the parameters, which is easy to produce large errors in the calculation process. Trend prediction is the extraction of one or more trend types from the historical data of one or more time periods, in order to make predictions about future changes. On the basis of statistics and modeling of the data, through the trend analysis of the data sequence in a certain period, the future data can be predicted, so as to provide the basis for decision making.

In this paper, ARIMA is selected as the algorithm applied to predict the new model of community recruitment. ARIMA consists of an autoregressive sliding average model and an autoregressive moving average model, which can be used to predict time series. The algorithm can take into account the trend, seasonality and residuals of time series data, and can play a good role in the new prediction model of community recruitment. However, ARIMA needs to calculate multiple parameters, which can increase its complexity and computational difficulty. In order to solve this problem, this paper adopts the linear regression method in the long-term prediction to improve the computational complex problem [11].

(2) Design of a prediction model for club recruitment

The steps of the prediction model designed by ARIMA in this article are shown in Figure 1, including model selection, acquisition of sample time series data, and preprocessing of input data for model prediction.

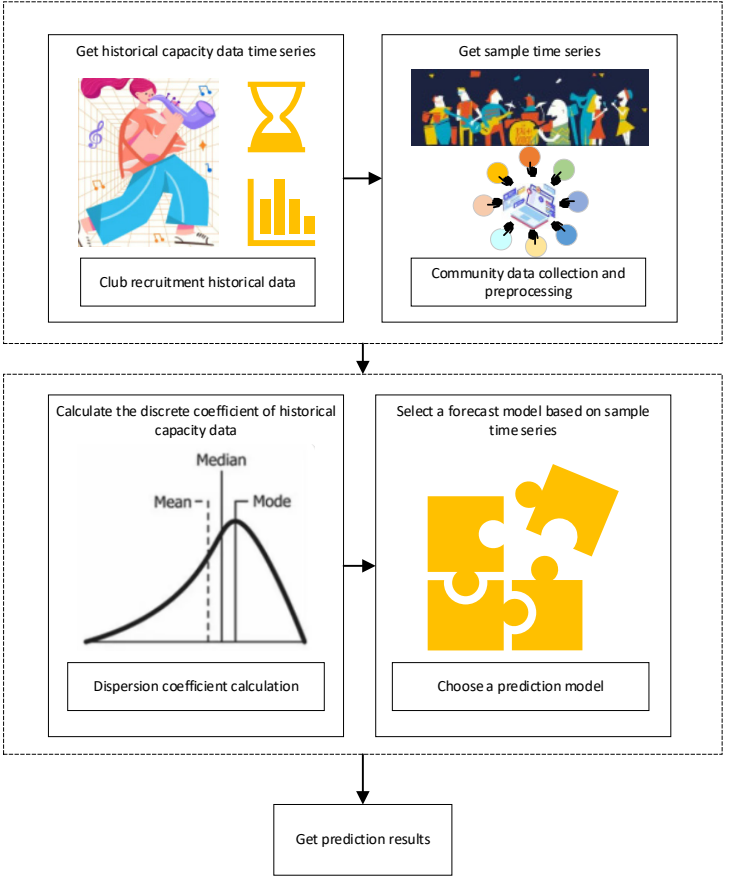

Figure 1. Steps for designing a predictive model

As shown in Figure 1, the design of the community recruitment prediction model in this article first obtains the historical capacity data time series, and then preprocesses the historical capacity data to obtain the sample time series. Afterwards, the historical capacity data is subjected to discrete coefficient calculation, and a prediction model is selected based on the discrete coefficient. The formula for calculating the dispersion coefficient is as follows:

$$CV = \frac{\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ...+ (x_n - \bar{x})^2}{n}}}{\bar{x}} \quad (1)$$

Among them, $\bar{x}$ is the average, n is the amount of data, and $x_1, x_2, ..., x_n$ are the specific values of each data. Afterwards, the selected prediction model is established based on the sample time series and tested through residual analysis to determine whether the prediction model is qualified. In residual analysis, a established model is used to predict historical data and obtain predicted values. The actual observed values are then subtracted from the predicted values to obtain a residual sequence. In subsequent analysis, autocorrelation function graphs can be used to test whether the residual sequence has autocorrelation. The stationarity of residual sequences can be verified through unit root tests, and relevant adjustments and transformations can be made based on the results. Afterwards, data collection and preprocessing are carried out, and the processed data is input into the prediction model to obtain the prediction results. ARIMA is a composite sequence composed of autoregressive terms and moving average terms, used to describe trends over a period of time. In ARIMA, the moving average term represents the difference between the current and past values of the original data. The moving average term cannot represent the trend itself, but is a statistical measure that reflects changes in the trend.

(3) Data collection and preprocessing

The data studied in this article was collected from the recruitment activities of college student clubs in 2020, spanning from November 2020 to November 2022. Considering the availability and researchability of data, this article used the China University Student Association Data Platform to obtain the data. This platform contains registration information, club names, establishment time, current president and vice president, responsible persons, and other information of student organizations from various universities across the country. Meanwhile, considering the diverse forms of activities in university student clubs, including collective activities, academic activities, and hobbies, this article selects a portion of them as samples from various university student organizations. This article divides the samples into two categories: one is organizations that are not on the club member list, that is, organizations that are not on the club member list. The other type is organizations on the list of club members, which belong to the list of club members. This article mainly analyzes organizations that are not on the list of club members. After registering on the platform, these organizations can have a dedicated system responsible for entering and maintaining student organization information. In order to prevent duplicate entries, this article distinguishes organizations that do not belong to the club member list from organizations of the same type on the platform.

(4) Variable design

This article aims to investigate the impact of three factors: student acceptance of club activities, internal structure and trends of club changes, and student participation in club activities at different time points on the prediction model of the number of university club members. By simplifying the analysis, the number of students participating in club activities can be used as the dependent variable, and the attitudes of students towards club activities, internal structure and trends of club changes, and student participation in club activities at different time points can be used as independent variables. ARIMA can be used to establish a prediction model for the number of university club members.

4. Results and Discussion

In order to provide a more intuitive evaluation of the club recruitment prediction model established based on ARIMA in this article, the traditional moving average method was selected as a comparison. They respectively input data into the moving average method and the prediction model established by ARIMA for prediction. It can compare and evaluate the differences between the two methods in practical applications, and propose corresponding improvement measures for the prediction results.

(1) Residual analysis of club recruitment prediction model

In order to verify the effectiveness of the club recruitment prediction model designed in this article, residual analysis was conducted, and the analysis results are shown in Table 1 and Figure 2:

Table 1. Residual Analysis Data

Observation	Predicted value	Residual	Standardized residual	Squared residual
1	132	14	0.9	0.81
2	145	9	-0.2	0.04
3	198	22	0.4	0.16
4	166	16	0.6	0.36
5	213	19	0.7	0.49
6	149	6	-0.3	0.09
7	169	4	0.5	0.25
8	178	13	0.4	0.16

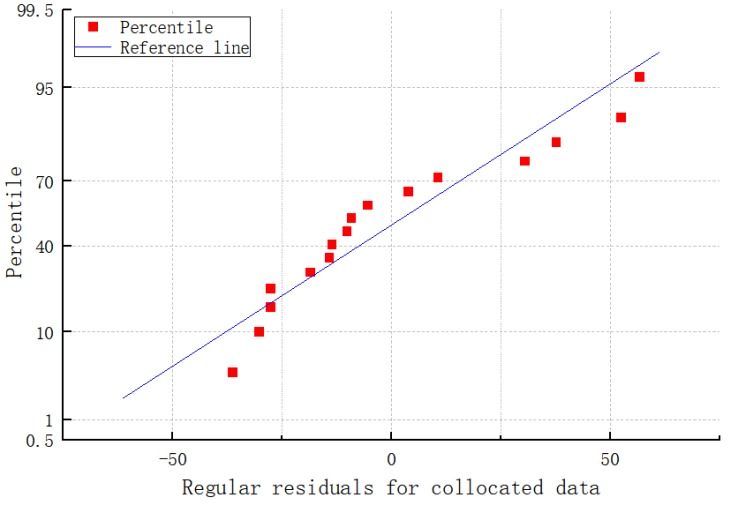

Figure 2. Residual analysis of the prediction model for club recruitment

From Table 1 and Figure 2, it can be seen that the standardized residuals of the model designed in this article are between -0.3 and 0.9. This indicates that most residuals are concentrated near the mean, indicating a small gap between the model's predictions and actual values. The squared residuals range from 0.04 to 0.81. A low squared residuals mean smaller prediction errors, which in turn increase the stability of the model.

(2) Influencing factors of the prediction model for the number of university club members

This article selects student acceptance of clubs, internal factors of clubs, and student club participation at different times as the main reasons for influencing factor analysis. This article obtained the rating results of relevant influencing factors by adjusting the attitudes of students towards club activities, as shown in Table 2. The total score for acceptance is 100, and the total score for stability is 10:

Table 2. Analysis of influencing factors

Time point	Club members (number)	Acceptance level	Structural stability	Participation in club activities (%)
2020.11	150	84	9.4	91

2021.5	163	92	9.1	88
2021.11	165	87	8.4	76
2022.5	172	93	8.8	86
2022.11	184	95	9.6	94

From Table 2, it can be seen that the higher the structural stability of university clubs, the higher the corresponding participation rate of student club activities. At different times, students have varying levels of participation in clubs, with the highest number of club members being 184 and an acceptance score of 95. From November 2020 to November 2022, the overall membership of the club has been increasing, indicating an increase in the club's operations, activity quality, and influence. In terms of structural stability, the score range is 8.4-9.6 points, indicating good internal management of the club, stable relationships among members, and no frequent leadership changes or organizational restructuring events. In terms of participation in club activities, the participation rate is between 76% -94%, and the participation rate of student club activities fluctuates greatly, indicating that the participation rate of club activities is relatively unstable.

In addition, to verify whether the above factors have a significant impact on the prediction model of the number of university club members, this paper conducted a significance analysis through the design of independent and dependent variables, and the results are shown in Table 3:

Table 3. Significance analysis

Variable	Willingness		
	Coefficient	Significance	Exp (B)
Manner	0.081	0.031	0.048
Structure	1.174	0.021	0.162
Character	1.469	0.769	0.811
Time	1.269	0.014	2.869

From Table 3, it can be seen that the significance values of attitude, structure, and time are all below 0.05, indicating that these three factors have a significant impact on the number of club members. The coefficients are all positive, indicating a positive correlation between students' positive attitude towards club activities, club structure and time, and the number of club members. The number of club members is influenced by semester cycles, holidays, and other seasonal events.

(3) Prediction accuracy of club recruitment prediction model

This article compares the accuracy data of the prediction models output by ARIMA method and moving average algorithm, and the obtained data is shown in Figure 3:

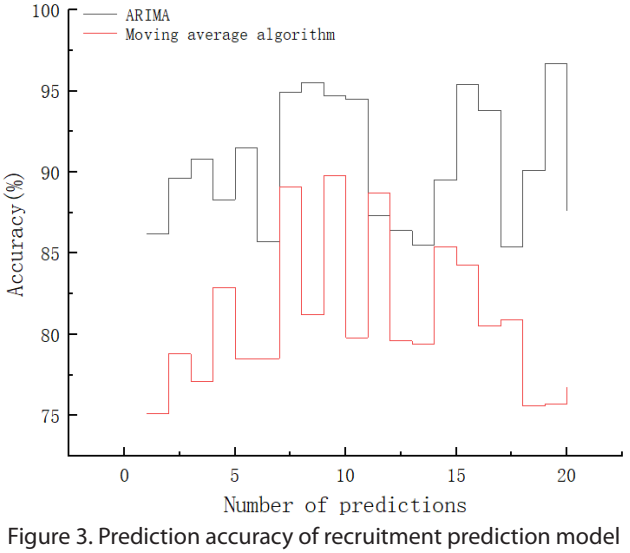

Figure 3. Prediction accuracy of recruitment prediction model

From Figure 3, it can be seen that in most cases, the accuracy of the ARIMA based community recruitment prediction model is between 85% -97%, while the prediction accuracy of the moving average algorithm based prediction model is between 75% -90%. This indicates that the overall prediction accuracy based on ARIMA can improve the prediction accuracy of traditional prediction models during the prediction process. It can also be observed that the prediction accuracy of some ARIMA models is lower than that of the moving average algorithm. This indicates that although ARIMA has higher accuracy than the moving average method in most cases, there are also specific conditions or datasets under which the ARIMA model performs worse than the moving average method.

(4) Prediction efficiency of club recruitment prediction model

In addition to analyzing the prediction accuracy of the model, this article also collected the time spent on training and predicting models using different methods, in order to analyze the prediction efficiency of the model. The training time data of the model is shown in Figure 4:

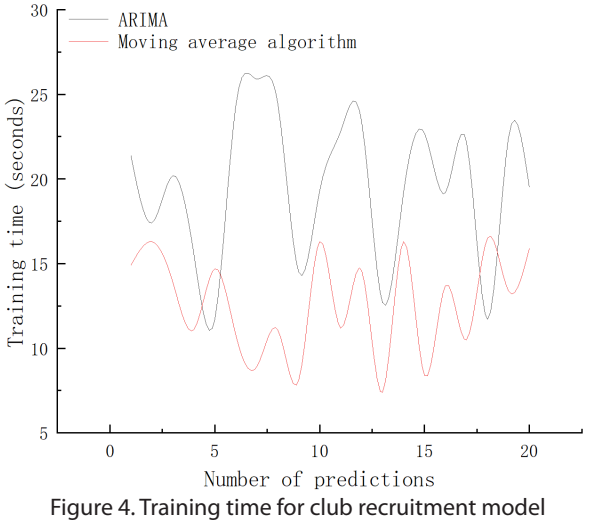

Figure 4. Training time for club recruitment model

From Figure 4, it can be seen that the training time of the ARIMA based recruitment prediction model for club recruitment is generally higher than that of the moving average algorithm. The maximum training time of the ARIMA based recruitment prediction model exceeded 25 seconds, while the maximum training time of the moving average algorithm was 16.5 seconds. This indicates that the training efficiency of traditional prediction models is higher than that of the method used in this paper. This may be due to ARIMA requiring more time to analyze and statistically analyze data characteristics, and increasing training time during the optimization process. The predicted time data of the club recruitment model is shown in Figure 5:

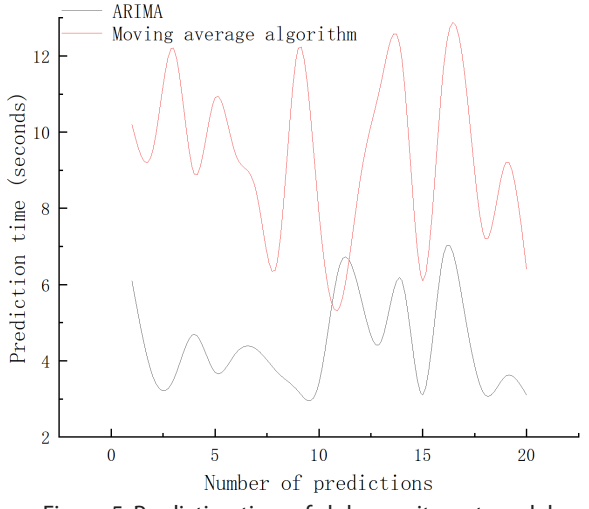

Figure 5. Prediction time of club recruitment model

From Figure 5, it can be seen that overall, the ARIMA based community recruitment prediction model has less prediction time, indicating higher efficiency. Compared to traditional moving average algorithm prediction models, the prediction model designed in this paper has a significant improvement in prediction efficiency. The maximum prediction time difference between the two methods is about 9 seconds, indicating that ARIMA can save up to 9 seconds of prediction time. This indicates that after implementing model training, ARIMA may be able to process input data faster and generate prediction results.

(5) Prediction model for club recruitment and resource allocation

Due to the problem of poor resource allocation performance in traditional prediction models, a comparative analysis of the resource allocation performance of the prediction model designed in this paper is conducted to evaluate the resource allocation performance of the ARIMA based community recruitment prediction model. The specific data is shown in Figure 6:

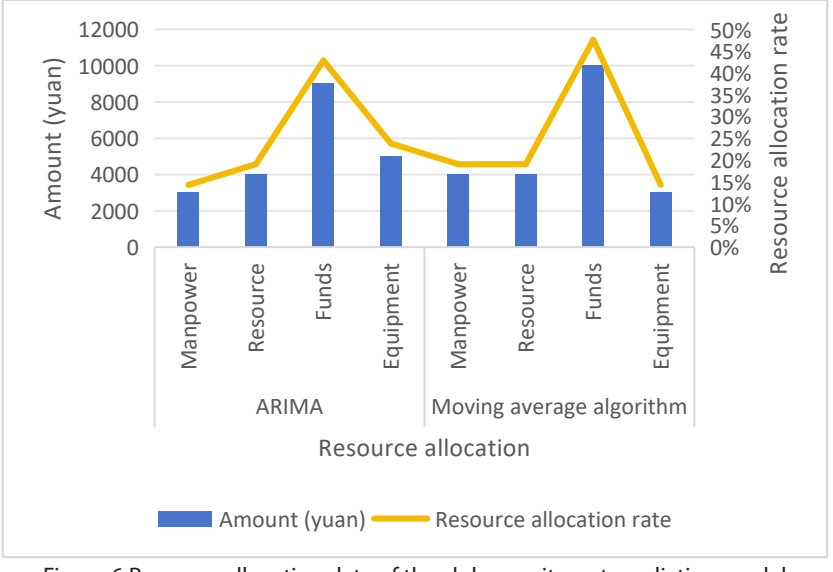

Figure 6 Resource allocation data of the club recruitment prediction model

From Figure 6, it can be seen that in terms of resource allocation, the prediction models based on ARIMA and moving average algorithms mainly allocate resources to funds.However, unlike traditional moving average algorithms,ARIMA based prediction models also tend to focus on equipment resources, while traditional methods tend to focus on manpower. More resources can be invested in equipment, and the club can improve the quality and efficiency of acquisition, increase the cost of equipment investment, and bring about equipment updates and additions, thereby providing better technical support for club members. In contrast, although human resources are important, the efficiency improvement brought about by the increase of manpower is not linear. With the advancement of technology, the activities of clubs can increasingly rely on high-quality equipment. In the long-term recruitment of clubs, labor costs can continue to increase, and equipment investment can be utilized in the long term.Therefore,the ARIMA model based prediction of club recruitment has more advantages in resource allocation.

5. Conclusions

This article applies ARIMA to a club recruitment prediction model, predicting the number of new recruits by analyzing historical recruitment data. The designed model solves the problems of low prediction accuracy and unreasonable resource allocation in traditional prediction models, and also improves the prediction efficiency of the model.The training process of the model takes a relatively long time, and future research can continuously optimize the model and algorithm to improve the efficiency of model training.This article provides more effective

planning and decision-making support for club recruitment activities, which has important theoretical and practical value. In order to further enhance the practicality and impact of research, future research can consider introducing more dimensional data and exploring more types of predictive models to provide data support and strategic recommendations for the development of university clubs.

References

[1] ZHANG Qiang,JIANG Meixia,ZOU Mi.Research on the Difficulties, Strategies and Practice of Recruiting Innovation and Entrepreneurship Associations in Higher Vocational Colleges[J].The Theory and Practice of Innovation and Enntrepreneurship, 2022, 5(2): 133-136.

[2] Zhang Wenjie, Ji Qingnan,Xie Haojie, Hong Xuyang,Liu Huanyu, Wu Jiatong. Design of university campus community information release organization platform based on MVC architecture model [J]. Design, 2023, 8(4): 3456-3465.

[3] Zhang Heming, Yan Siyu.A brief analysis of the characteristics and restrictive factors of university societies [J].Innovative Education Research,2022,10(10): 2453-2458.

[4] Kumar R, Kumar P, Kumar Y. Multi-step time series analysis and forecasting strategy using ARIMA and evolutionary algorithms[J].International Journal of Information Technology, 2022, 14(1): 359-373.

[5] Li S, Xu W,Li Z. Review of the SBAS InSAR Time-series algorithms, applications, and challenges[J]. Geodesy and Geodynamics, 2022, 13(2): 114-126.

[6] Ren Weijie,Han Min. Review of research on multivariate time series causality analysis[J]. Acta Automata, 2021, 47(1): 64-78.

[7] Wang X, Kang Y, Hyndman R J. Distributed ARIMA models for ultra-long time series[J]. International Journal of Forecasting, 2023, 39(3): 1163-1184.

[8] Schaffer A L, Dobbins T A,Pearson S A.Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions[J].BMC medical research methodology, 2021, 21(1): 1-12.

[9] Deif M A, Solyman A A A, Hammam R E. ARIMA model estimation based on genetic algorithm for COVID-19 mortality rates[J]. International Journal of Information Technology & Decision Making, 2021, 20(06): 1775-1798.

[10]Xu D, Zhang Q,Ding Y.Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting[J].Environmental Science and Pollution Research, 2022, 29(3): 4128-4144.

[11]Su Yaqian,Cui Chaoran,Qu Hao.Time series forecasting based on self-attention moving average[J].Journal of Nanjing University (Natural Science Edition), 2022, 58(4): 649-657.