J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 8813--8827

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Studying the Interactive Efficiency of Social Media in English Language Education Based on Time Series Prediction Algorithms

Jing Xia[1, ✉]

[1] *School of Marxism, Chongqing Vocational and Technical College of Industry and Trade, Chongqing, 408000, China*

ABSTRACT

Social media as a new type of media has become an important channel for people to obtain information and communicate, which brings new opportunities and challenges for English education. In this study, the Markov chain model is improved by using the weighted ward system clustering method and the fuzzy set theory to improve the prediction performance of the model. Then the ARIMA model and the improved Markov chain model are combined to construct an improved time series prediction model to realize the prediction of the interaction efficiency of social media in English education. The performance of the improved prediction model is superior compared to other comparative models, providing reliability for the subsequent prediction results. The prediction results show that the interactive efficiency of social media interaction data in English education shows an upward trend over time, and the number of readings and playbacks of English courseware resources as well as video resources increases from 18477 and 18147 to 88629 and 84571 in six months. The predicted results of this study indicate that social media has good interactive efficiency in English education, which can be utilized in the future to expand the dimension of education, build an English education platform, expand the teaching space and extend educational thinking, and play a percolating role in English education.

Keywords: Markov chain, ARIMA model, time series prediction model, social media, English education

1. Introduction

Time series forecasting algorithm is an important data analysis method used to predict the trend of numerical changes in the future period of time. This algorithm has a wide range of applications in various fields such as economics, finance, meteorology, and education [1-3]. By analyzing and modeling past observations, time series forecasting algorithms can help us to understand and explain the patterns of data changes and provide strong support for future decision making [4-5].

✉ Corresponding author.
E-mail address: gmjing1936@126.com (J. Xia).
Received 12 January 2024; Revised 05 March 2024; Accepted 25 December 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-489

Autocorrelation reflects the internal pattern of the time series data, which can be evaluated by calculating the autocorrelation function of the series. Trend indicates the overall change trend of the time series in the long term, and can be used to capture and predict future changes by fitting curves or models, which is of great significance when applied to the study of the interaction efficiency of social media in English education [6-9].

In recent years, with the continuous development and popularization of Internet technology, social media plays an increasingly important role in people's lives. As a powerful tool, social media are used by more and more people and organizations to communicate and transfer information, which also includes educational institutions and their educators [10-13]. In university English teaching, social media is not only conducive to improving students' English proficiency and learning efficiency, but also cultivating students' ability of active exploration and independent learning [14-16]. At the same time, the application of social media can also promote the exchange and communication between teachers and students, enhance the efficiency of interaction, and establish a closer teaching partnership [17-18].

Literature [19] explores the impact of social media on English language proficiency based on medical students' perspectives. Descriptive survey design, quantitative methods were used. The researchers noted that social media has shown many advantages in English language learning and the effects it brings are not affected by gender. Literature [20] explored the impact of social media on English language learning during CoviD-19. A survey was conducted on 166 undergraduate students and concluded that social media positively influenced students' writing style and reading skills. Literature [21] examined students' perceptions of social media in English vocabulary learning. A questionnaire survey of 193 undergraduate students was conducted using quantitative research methods and analyzed using descriptive statistical tools. The results of the study showed that many features of social media are beneficial to students' learning and development of English vocabulary. It also pointed out that social media can be used as an interactive and pedagogical platform between teachers and students. Literature [22] reveals the potential opportunities and challenges of social media in flipped classroom teaching. A literature review was conducted in terms of flipped learning activities, students' achievement and attitudes. The results of the analysis emphasized that the flipped classroom has positive effects in terms of learning and engagement, while students' lack of familiarity and proper adaptation became one of the most important challenges it faces. Literature [23] emphasizes the role of social media in English language learning. A questionnaire was launched to 100 English majors. The results showed that social media resources have a positive impact on students in higher education because socializing has a motivating effect on students and stimulates their interest in learning applied English language learning. Literature [24] examined students' attitudes towards social media. The study was conducted on 132 students using random sampling and SPSS version 23 and descriptive statistical tools were taken to tabulate and analyze the data. Social media has a positive impact on students' English writing proficiency. Literature [25] suggests that social media has great potential in teaching English, but there are drawbacks such as distraction and misinformation. The findings of the study reveal that it is necessary to develop effective strategies to deal with the shortcomings of social media in order to realize the utilization of social media media in English language education. Literature [26] applied a quasi-experimental research methodology to conduct a comparative experiment with 101 students. Compared to social media, mobile learning technology is more capable of improving students' behavior, socialization, and English learning outcomes. Whereas traditional teaching tools are more underperforming in these areas, future research may be more inclined towards the development of serious games in order to achieve this goal.

Based on the characteristics of online social media platforms in the Internet era, this paper applies them to English education, proposes a variety of interactive strategies for English education under social media integration, and proposes a time series prediction algorithm to analyze the interactive efficiency of social media in English education. The smoothness of social media interaction data in

English education is first judged, and the unqualified data are processed by using difference operation, and then the ARIMA model is constructed by using autocorrelation function and partial autocorrelation function and so on. Then the weighted ward system clustering method is utilized to establish the Markov chain state space of the samples, the data sequence is subjected to the Mahalanobis test, and the transfer state probability matrix is established. Subsequently, the model is improved and optimized using the affiliation formula and fuzzy set theory to obtain the EEMD-ARIMA-Markov time series forecasting model. Finally, this study analyzes the performance of this prediction model, and implements the interactive strategy of English education based on social media in School A. The prediction model is used to analyze the interactive efficiency of social media in English education, and to explore whether social media has a facilitating effect on the effect of English education.

2. Social media-based interactive strategies for English language education

1) Carrying out online teaching and building an efficient classroom. In the Internet era, students are increasingly active in social media, relying on online social media to share their daily learning life and using online channels to carry out multiple interactions. In this context, English teachers should keep abreast of the changes of the times, change their teaching concepts, and carry out English interactive online teaching to enhance the overall interactive teaching effect. In addition, the teaching and research team should develop interactive micro-courses for English education with the help of social media platforms, so that students can use fragmented time to watch English videos, audios and pictures on the mobile terminal, deeply understand the English content they have learned, and continue to improve their comprehensive English literacy.

2) Grasp the characteristics of social media platforms and continuously optimize English teaching content. English teachers should fully grasp the characteristics of social media platforms, formulate targeted teaching programs based on students' characteristics, and timely adjust the goal of English ability cultivation. Enhance their own technical application ability, integrate the English teaching resources on the media platform in an all-round way, regularly push the relevant teaching content, and strengthen the differentiation and relevance of the English teaching content.

3) Integration of educational methods to improve interactive effects. In the process of creating an interactive English education classroom based on social media, English teachers should flexibly integrate educational methods to realize the effective linkage between online and offline, so as to ensure the teaching effect of the interactive English classroom. For example, creating a flipped classroom, regularly uploading high-quality English teaching content, amplifying the convenience and effectiveness of social media teaching, so that students can complete the pre-course preparation, independently find out the content of the course points of doubt, and efficiently carry out the course content learning. At the same time, teachers can try to create class social media accounts, organize knowledge analysis, classroom questions, unit quizzes, unit theme discussions and other diversified English interactive activities to guide students to grasp the focus of the unit learning theme and make suggestions. Teachers can refine the overall teaching plan based on students' social media feedback, so that students can have a precise understanding of the logic and ideas of classroom teaching design. This not only helps to form an interactive atmosphere of free participation in English, but also maximizes the quality of interactive classroom teaching.

Therefore, this paper takes School A as a pilot to realize the interaction between English teachers and students based on social media (QQ, WeChat, Jieyin and Zhihu). This paper analyzes the interactive efficiency of social media in English education in this school from six aspects: the reading volume of classroom resources uploaded by teachers, the playing volume of video resources, the collection volume of learning resources, the forwarding volume of learning resources, the participation volume of English education activities, and the interaction volume between teachers and students. It lays the foundation for the application of social media in English education.

3. Improved construction of time series forecasting algorithms

3.1. Differential autoregressive moving average model construction

3.1.1. Data pre-processing. It is first necessary to determine whether the data is smooth or not, and if it does not satisfy smoothness, then the difference operation is required. The smoothness of the data refers to the time series generated by a certain random process, and each data in the sequence $\{X_t\}\{X_1, X_2, X_3, \ldots, X_t\}$ is taken from the same probability distribution.

1) The random time series is said to be smooth and the white noise process is smooth if the following three conditions are satisfied.

Data expectation μ is a constant independent of time t, i.e., $E(X_t) = \mu$. Data variance σ^2 is a constant independent of time t, i.e., $Var(X_t) = \sigma^2$. Data covariance γ_k is a constant related only to the time interval κ and independent of time t, i.e., $Cov(X_t, X_{t+k}) = \gamma_k$.

2) ADF test (unit root test). The object to be tested by the ADF test is the unit root. for an autoregressive process:

$$X_t = bX_{t-1} + a + \varepsilon_t \tag{1}$$

If the lag coefficient b in this autoregressive process is 1, it is said to be a unit root. If a unit root exists, it means that the residuals in this model are always on, and that the past data are deceptively related to the current data.

By the above two determination methods, if the data does not satisfy the smoothness condition, it is necessary to perform a difference operation on it. Then the ΔX_t obtained after the difference operation is re-examined for smoothness. If the difference after ΔX_f still fails to meet the stability conditions, then the difference operation is carried out several times until the ADF test is passed and the smooth data are obtained. The number of differences is $ARIMA(p,d,q)$ model parameter d. In general, the value of d is not greater than 5, if the number of differences is more than 5, the data will produce a large distortion, which means that the group of time series is not suitable for forecasting using the ARIMA model.

3.1.2. Model ordering. Model ordering is a key step in the algorithm, mainly to determine the autoregressive order p and the sliding average order q. The autocorrelation function ACF [27] mainly describes the degree of correlation between the current data of a time series and its past data, including direct and indirect correlation information between the two, i.e., it describes all the relationships between x_t and x_{t-k}. For a smooth time series $\{X_t\}$, its k th order correlation coefficient ρ is defined as:

$$\begin{aligned}\rho_k &= \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)} \times \sqrt{Var(X_{t+k})}} \\ &= \frac{Cov(X_0, X_k)}{\sqrt{Var(X_0)} \times \sqrt{Var(X_k)}} \\ &= \frac{r_k}{r_0}\end{aligned} \tag{2}$$

where the self-covariance c_k is defined as follows:

$$\begin{aligned} c_k &= Cov(X_t, X_{t+k}) \\ &= E((X_t - \mu)\times(X_{t+k} - \mu)) \\ &= \frac{1}{N}\sum_{t=k+1}^{N}(X_t - \mu)\times(X_{t+k} - \mu) \end{aligned} \tag{3}$$

Autocorrelation coefficient $r_k = \frac{c_k}{c_0}$. Then the correlation coefficient ρ_k can be introduced:

$$\rho_{\mathrm{k}} = \frac{N}{N-k}\times\frac{\sum_{t-k+1}^{N}(X_t - \mu)(X_{t+k} - \mu)}{\sum_{t-1}^{N}(X_t - \mu)(X_t - \mu)} \tag{4}$$

The partial autocorrelation function PACF [28] mainly describes the direct relationship between X_t and $X_{t\text{-}k}$, ignoring the effect of other shorter lag terms $(X_{t\text{-}1}, X_{t\text{-}2}, \ldots\ldots, X_{t\text{-}k\text{-}1})$. Its k th order correlation coefficient is defined as:

$$\begin{cases} \phi_{11} = \rho_1 \\ \phi_{22} = \dfrac{(\rho_2 - \rho_1^2)}{(1-\rho_1^2)} \\ \phi_{kk} = \dfrac{\rho_k - \sum_{j=1}^{k-1}\phi_{k-1,j}\rho_{k-j}}{k-1}, k = 3,4,5,\ldots \end{cases} \tag{5}$$

where $\phi_{ki} = \phi_{k-1,i} - \phi_{kk}\phi_{k-1,k-i}, j = 1,2,3,\ldots,k-1$.

In the ACF and PACF function methods, it is necessary to determine the value of p,q by determining whether the autocorrelation coefficient or partial autocorrelation coefficient of the model will significantly tend to 0 from a certain order. If there is such a tendency towards 0, the autocorrelation coefficient or partial autocorrelation coefficient of that order is called a truncated tail. If no such trend exists, it is called a trailing tail.

First calculate the coefficients for the $AR(p)$ model. First transform the $AR(p)$ model expression into the following form:

$$X_{t-k}X_t = \phi_1 X_{t-k}X_{t-1} + \phi_2 X_{t-k}X_{t-2} + \ldots + \phi_i X_{t-k}X_{t-i} + X_{t-k}\mu_t \tag{6}$$

Reduced to:

$$\gamma_k = \phi_1\lambda_{k-1} + \phi_2\lambda_{k-2} + \ldots + \phi_p\lambda_{k-i} \tag{7}$$

Then the autocorrelation coefficient of the moving average model can be found:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-i} \tag{8}$$

where $\rho_0 = 0, k > 0$.

After solving for ρ_k of each order, the system of Yule-Walker equations can be derived:

$$\begin{gathered} \rho_1 = \phi_{k1} + \phi_{k2}\rho_1 + \ldots + \phi_{kk}\rho_{k-1} \\ \rho_2 = \phi_{k1}\rho_1 + \phi_{k2} + \ldots + \phi_{kk}\rho_{k-2} \\ \vdots \\ \rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \ldots + \phi_{kk} \end{gathered} \tag{9}$$

For any $i<k$, φ_{ii} is the partial autocorrelation coefficient of order i of the sequence. The number of systems of equations is equal to the number of unknowns, so φ_{ii} can be found.

Similarly, for the $MA(q)$ model expression for a smooth sequence, set:

$$\gamma_k = E\left(x_t x_{t-k}\right) = \begin{cases} \sigma^2(1+\theta_1^2+\theta_2^2+\ldots+\theta_i^2), k=0 \\ \sigma^2(\theta_k+\theta_1\theta_{k+1}+\ldots+\theta_{k-i}\theta_i), 0<k\le i \\ 0, k>i \end{cases} \tag{10}$$

Then the autocorrelation coefficient ρ is introduced by Eq:

$$\rho_k = \frac{r_k}{r_0} = \begin{cases} 1, k=0 \\ (\theta_k+\theta_1\theta_{k+1}+\ldots+\theta_{k-i}\theta_i)/(1+\theta_1^2+\theta_2^2+\ldots+\theta_i^2), 0<k\le i \\ 0, k>i \end{cases} \tag{11}$$

Each $MA(q)$ has a corresponding $AR(\infty)$, and the partial autocorrelation coefficient of $MA(q)$ can be found indirectly through $AR(\infty)$.

3.2. Improved modeling of Markov chains

3.2.1. State space based on clustering of weighted ward systems. The basic idea of the traditional ward system clustering method [29] is to select the samples that minimize the increment of S to be merged, if when the samples have outliers, S will increase very fast because of these few outliers. When the number of clusters k is unknown, the final clustering effect is not significant if more similar samples cannot be clustered into one class just to minimize the increment of S. So in this paper, the weighted ward system clustering method is used, i.e., the weight is given to the sum of squares of the outliers within each class, and at this time the sum of squares of the outliers within the class is:

$$SS = \sum w_t S_t = \sum_{t=1}^{k}\sum_{i=1}^{n_t} w_t (x_i^{(t)} - \overline{x^{(t)}})^2 \tag{12}$$

where, $w_t = n_t/m$. n_t denotes the number of samples in category t. $x_i^{(t)}$ denotes the ith sample in category t. $\overline{x}^{(t)}$ is the mean of the samples in category t. k is the total number of categories. The weighted ward system clustering method only assigns weights to the sum of squares of the departures within each class, and the remaining steps are consistent with the ward system clustering method. This method provides more intraclass sums of squares of deviations for classes that contain a large number of samples and fewer intraclass sums of squares of deviations for samples that contain a small number of outliers. The data series are categorized into N states using the weighted ward system clustering method, i.e., the state space of the Markov chain is $C=\{1,2,\cdots,N\}$.

3.2.2. Mars test and transfer state probability matrices. In order to ensure that the data series are Markovian in nature [30], the data series are subjected to a Markov test before model prediction, which is usually performed using the χ^2 statistic. If the sum of the columns of the transfer frequency matrix is divided by the sum of the rows and columns, the resulting value becomes the “marginal probability”, i.e.:

$$p_j = \frac{\sum_{i=1}^{N} a_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}} \tag{13}$$

where a_{ij} is the frequency corresponding to state i to state j. χ^2 statistic is:

$$\chi^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}\left|\ln\frac{p_{ij}}{p_j}\right| \tag{14}$$

The statistic is following a χ^2 distribution with degree of freedom $(N-1)^2$. Given the significance level α, the quantile point $\chi_a^2(N-1)^2$ is known by querying the chi-square table. The value of the statistic χ^2 obtained through the calculation is compared with the quantile point $\chi_a^2(N-1)^2$ and if $\chi^2 > \chi_\alpha^2(N-1)^2$ it is considered that the sequence satisfies the Markov property, otherwise it is considered that the sequence cannot be predicted by a Markov chain. According to the state space $C=\{1,2,\cdots,N\}$ of the Markov chain, the k step transfer probability matrix can be obtained as:

$$p^{(k)} = \begin{bmatrix} p_{11}^k & p_{12}^k & \cdots & p_{1N}^k \\ p_{21}^k & p_{22}^k & \cdots & p_{2N}^k \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1}^k & p_{N2}^k & \cdots & p_{NN}^k \end{bmatrix} \tag{15}$$

where $p_{ij}^k = a_{ij}/a_i$ is the probability of state i transferring to state j after k steps, a_{ij} is the frequency corresponding to state i transferring to state j after k steps; and a_i is the total number of samples in state i.

3.2.3. Model improvements. In this paper, we use the affiliation formula to give the reference sample belongs to all the states of the affiliation, thus determining the initial state vector contains more information, then the nst data belongs to the state i of the affiliation is calculated as follows:

$$\mu_i(X_n) = \frac{|A_i - A_{i-1}|}{|A_i - X_n| + |X_n - A_{i-1}|}, i=1,2,\cdots,N \tag{16}$$

where $A_1, A_2, \cdots, A_{N-1}$ is the lower critical value of each level of criteria, $A_0 = \min\left\{X^t\right\} - \sigma_1; A_N = \max\left\{X^t\right\} + \sigma_2$, σ_1 and σ_2 are two non-negative real numbers. In order to make the initial state vector contain more information about the data, the affiliation formula is improved as follows:

$$\mu_i^*(X_n) = \frac{|A_i - A_{i-1}|}{|A_i - X_n| + |X_n - A_{i-1}|} \cdot \frac{|A_i - A_{i-1}|}{|A_N - A_0|}, i=1,2,\cdots,N \tag{17}$$

The initial state vector obtained using the improved affiliation formula not only contains information about the critical values of each rank, but also covers the ratio of the interval length of each rank to the total length, resulting in the initial state vector of the nst data as:

$$S(n) = (\mu_1^*(X_n), \cdots, \mu_N^*(X_n)) \tag{18}$$

The initial state vector of the reference sample is combined with its corresponding transfer probability matrix to obtain the state vectors of each order for the $n+1$st data. Normalize the krd order state vector of the $n+1$nd data as $S^k(n+1)$, and the predicted state of the sample to be tested is:

$$S(n+1) = (w_1, w_2, \cdots, w_N)\begin{bmatrix} S^1(n+1) \\ S^2(n+1) \\ \vdots \\ S^N(n+1) \end{bmatrix} = (s_1, s_2, \cdots, s_N) \tag{19}$$

where $s_1, s_2, \cdots, s_N$ is the probability that the precipitation of the sample to be measured belongs to each state.

In this paper, the level eigenvalue in fuzzy set theory [31] is used to predict the state corresponding to the $n+1$nd data, then the calculation method is:

$$d_i = \frac{s_i^\eta}{\sum_{i=1}^{N} s_i^\eta} \tag{20}$$

where η is the coefficient of maximum probability, generally 2 or 4, the larger its value represents the larger role of maximum probability. State eigenvalue $H = \sum_{i=1}^{N} id_i$, if the $n+1$rd data corresponding to the state is i, and $H \geq i$, then the predicted value of the $n+1$th data is $\frac{MH}{i+0.5}$. If $H < i$, then the predicted value of the $n+1$th data is $\frac{NH}{i-0.5}$, where M and N are the upper and lower critical bounds of the interval corresponding to the state i, respectively.

3.3. EEMD-ARIMA-Markov modeling

Combining the above, this paper establishes an improved EEMD-ARIMA-Markov model, and the flow of the model is shown in Fig. 1. Firstly, the time series data are decomposed to obtain a series of IMF components [32] and a residual component res. Then the noisy MF components are searched for, and the noisy IMF components are denoised using the new threshold function denoising method to remove the noise. A corresponding ARIMA model is built for the denoised MF component as well as the other IMF components, and an improved Markov chain model is built for the trend component res. The prediction results of each model are summed up to get the final prediction results.

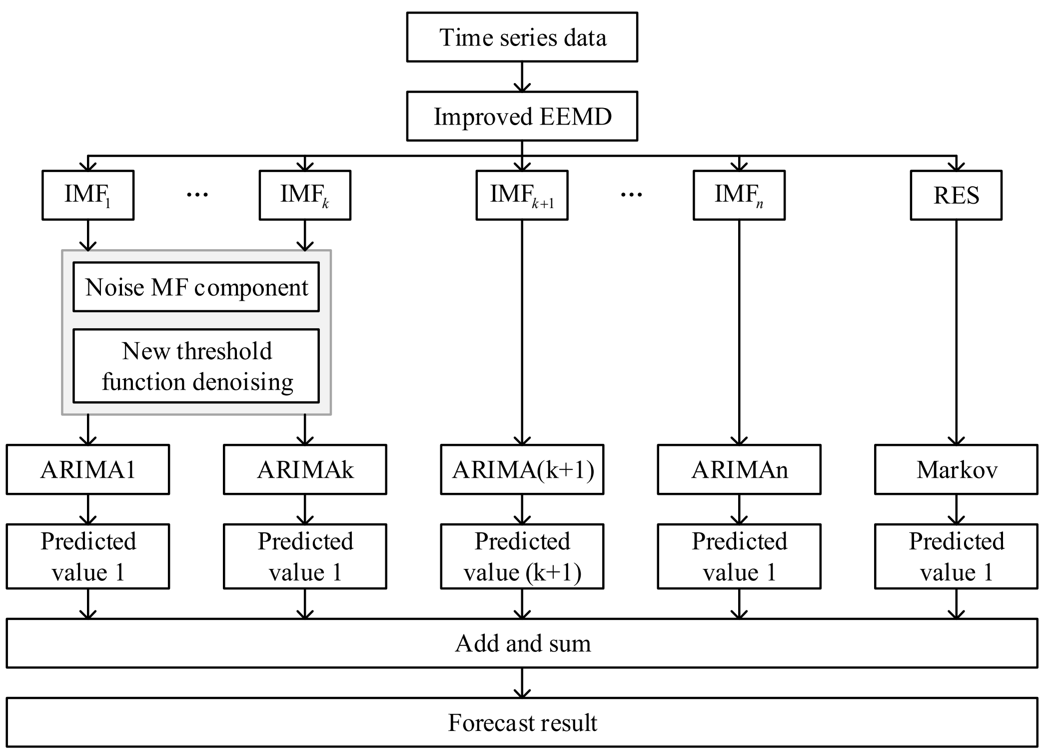


Fig. 1. Flowchart of improving EEMD-ARMI-Markov model

4. Analysis of the efficiency of social media interaction in English language education

4.1. EEMD-ARIMA-Markov model performance analysis

4.1.1. Data set construction. In order to validate the prediction performance of the EEMD-ARIMA-Markov model, experiments are conducted on the following datasets.

1) Stocknet-dataset contains relevant interaction data from January 1, 2020 to January 1, 2021 for 8 accounts in a social platform. Data from January 1, 2020 to August 1, 2020 were used for training, August 2, 2020 to September 30, 2020 for validation, and October 1, 2020 to January 1, 2021 for testing.

2) Astock-dataset contains two main sections, social media posts and interaction data, which cover the interaction data of each post from July 2020 to November 2021 for an official account. The data between July 2020 and December 2020 is used for training, while the data between January 2021 and November 2021 is used for testing.

4.1.2. Experimental setup. In this paper, a lag window of 5 days was used to construct the samples in the experiments, and the batch size for model training was set to 32. The maximum number of text messages per full day was set to 500, and excess text messages were discarded. The word embedding size is set to 150 due to GPU memory limitations. The number of hidden cells of the message embedding layer and the VAE are set to 200 and 400, respectively. The ownership weight matrix in the EEMD-ARIMA-MARKOV model is initialized using the fan-in trick, and the deviation is initialized to zero. In this paper, the model is trained with the Adam optimizer with an initial learning rate of 0.001 and the dropout rate is set to 0.3 to regularize the latent variables.

Meanwhile, in order to verify the prediction performance of the proposed EEMD-ARIMA-Markov model, nine different models were compared with the EEMD-ARIMA-MARKOV model, which are Rand, ARIMA, SVR, LSTM, GRU, TSLDA, HAN, Adv-LSTM, StockNet model. The prediction performance metrics were selected as ACC (Accuracy), MCC (Matthews Correlation Coefficient), P (Precision), R (Recall), and F1 values, which were calculated as shown below:

$$P=\frac{TP}{TP+FP} \tag{21}$$

$$R=\frac{TP}{TP+FN} \tag{22}$$

$$F1=2\cdot\frac{P\cdot R}{P+R} \tag{23}$$

$$MCC=\frac{t_p\times t_n-f_p\times f_n}{\sqrt{(t_p+f_p)(t_p+f_n)(t_n+f_p)(t_n+f_n)}} \tag{24}$$

$$ACC=\frac{n_{acc}}{n_{total}} \tag{25}$$

where t_p, f_n, f_p, and t_n are the number of samples that are true positive, false negative, false positive, and true negative, respectively, n_{acc} is the number of times the model correctly predicts the trend, and n_{total} is the total number of times the model predicts the trend.

4.1.3. Experimental results and analysis. The experimental results of each experimental model on stocknet-dataset are shown in Table 1. From this table, it can be seen that Rand has the lowest values of ACC (49.315%) and MCC (-0.0111), which indicates that stochastic prediction alone is unable to achieve satisfactory prediction accuracy of social media interaction efficiency.The experimental results of SVR are better than those of ARIMA, which is due to the fact that SVR is more adept at fitting social media interaction data than ARIMA. However, both SVR and ARIMA are unable to capture the temporal dependency between the data well, resulting in both having poorer performance compared to deep learning models. However, it can be seen that the complexity of ARIMA and SVR is significantly lower than that of the deep learning based models.LSTM outperforms GRU, but the difference between them is not significant and GRU takes less time than LSTM.The prediction accuracy and MCC of the StockNet model are 55.481% and 0.0222, respectively.In comparison to GRU, the StockNet's ACC is improved by 3.514%, which indicates that the application of VAE can effectively improve the prediction accuracy. The EEMD-ARIMA-MARKOV model proposed in this paper achieves the highest ACC and MCC in the experiments on the stocknet-datasets dataset, which are 57.857% and 0.0301, respectively.Compared with TSLDA, which only uses textual information, the EEMD-ARIMA-MARKOV model proposed in this paper has a significant improvement in ACC and MCC have significant improvement, which can show the contribution of textual information and historical data to the prediction accuracy of social media interaction efficiency trend. Compared with StockNet, the EEMD-ARIMA-MARKOV model achieves a significant improvement of 2.376% and 0.0079 on ACC and MCC. Since social media interaction efficiency prediction is a challenging task, an accuracy of 57.857% is usually considered a satisfactory result for social media interaction efficiency trend prediction. Therefore, the experimental results show that the model in this paper can effectively predict the interactive efficiency trend of social media in English education. Meanwhile, compared with Adv-LSTM, the EEMD-ARIMA-MARKOV model has a significant improvement of 0.941% and 0.0025 on ACC and MCC, respectively, which indicates the effectiveness of orthogonality in enhancing the generalization ability of the model. In addition, the F1 value of the EEMD-ARIMA-MARKOV model is 0.7426, which is better than the experimental results of every other experimental model, which indicates that the EEMD-ARIMA-MARKOV model can more accurately and comprehensively predict the efficiency of social media interaction in English education.

Table 1. Performance on the stocknet-dataset data set

Model	ACC (%)	MCC	P	R	F1
Rand	49.315	-0.0111	0.4912	0.3866	0.5363
ARIMA	49.697	0.0005	0.4929	0.4244	0.5485
SVR	49.946	0.0112	0.5063	0.452	0.5653
LSTM	50.218	0.0112	0.5223	0.4879	0.6106
GRU	51.967	0.0176	0.5249	0.5679	0.6227
TSLDA	54.483	0.0183	0.5378	0.5825	0.6429
HAN	54.696	0.0204	0.5472	0.5981	0.6716
StockNet	55.481	0.0222	0.5532	0.7187	0.6907
Adv-LSTM	56.916	0.0276	0.5579	0.9231	0.7229
EEMD-ARIMA-Markov	57.857	0.0301	0.5629	0.9302	0.7426

The performance analysis results of each model on the Astock dataset are shown in Table 2.The performance of HAN is different from the experimental performance on the stocknet dataset, and its accuracy (54.456%) is better than that of StockNet (53.468%). This is due to the fact that the end-to-end textual attention layer of HAN can process Chinese textual messages in the Astock-datasets dataset more efficiently. However, taken together, the predictive performance of the EEMD-ARIMA-MARKOV model is still better than the other experimental models, which demonstrates the effectiveness of the EEMD-ARIMA-MARKOV model in predicting the efficiency of social media interactions in English education.

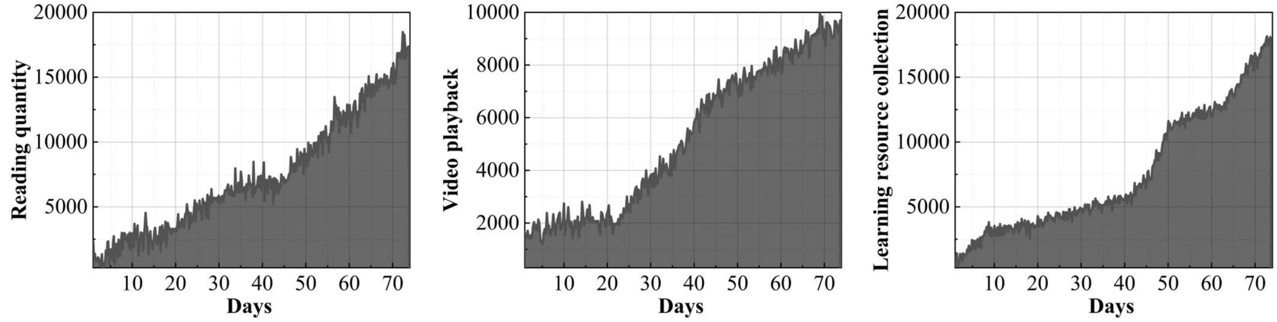
Table 2. Performance on the Astock-dataset data set

Model	ACC (%)	MCC	P	R	F1
Rand	48.007	-0.0003	0.4907	0.4962	0.452
ARIMA	48.895	0.0031	0.494	0.5315	0.4623
SVR	49.967	0.0057	0.4974	0.6237	0.5267
LSTM	51.332	0.0121	0.5051	0.6461	0.571
GRU	52.422	0.0124	0.5122	0.69	0.5731
TSLDA	52.497	0.015	0.5166	0.7526	0.5796
HAN	54.456	0.0193	0.5186	0.7623	0.6067
StockNet	53.468	0.02	0.5319	0.7894	0.6287
Adv-LSTM	55.779	0.0241	0.5319	0.8053	0.6309
EEMD-ARIMA-Markov	56.327	0.0248	0.58	0.8465	0.7204

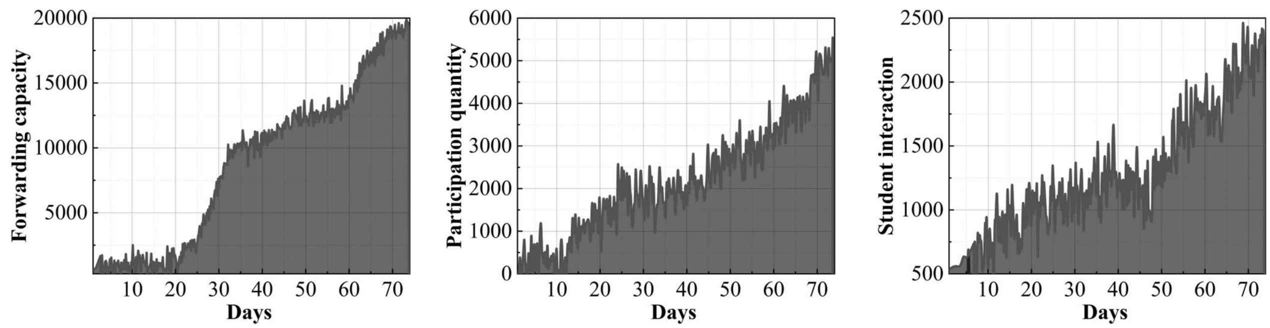
4.2. Results of Social Media Interaction Efficiency Analysis

4.2.1. Data acquisition. The crawling of social media interaction data in English education was conducted using Fiddler 4 tool, Android emulator, and python web crawler. After social interaction data crawling, considering that the reading, playing and forwarding data of related courseware as well as videos etc. in English education change with time, this study conducted several social interaction data collections at different times, each with an interval of at least one week, and compared the data of different times of duplicate resources that appeared in the collected online resources of English education. It was found that the changes in the readership, playback, and forwarding data of the courseware resources after a time interval of more than about half a month were not obvious. Based on the above observations, the span of the uploading time of the educational resources in the social media data set in English education was limited to January 1, 2023 to March 15, 2023, and the video data outside the time span was deleted, and initially 11,543 English education courseware resources were collected and 7162 English video resources. The experiments found that the collected video dataset has the condition of incomplete external features of some videos, such as missing titles, and unavailability of the number of plays and retweets. A small portion of English education related videos could not be played normally after downloading, probably because the files were corrupted due to not downloading properly. A small number of videos without sound may also be due to file corruption during download. Videos with the above three conditions should be removed from the dataset before proceeding to the next step of feature extraction because they cannot be opened normally or key features are missing. After cleaning the video dataset, 6874 videos were finally obtained as the dataset used for building the prediction model.

4.2.2. Descriptive Analysis of Social Data. The results of the trend analysis of each variable of the interaction efficiency of social media in English education between January 1, 2023 and March 15, 2023 in School A are shown in Figure 2, with (a)-(f) representing the results of the time-series analysis of the amount of reading of classroom resources uploaded by teachers, the amount of playing of video resources, the amount of favorites of learning resources, the amount of forwarding of learning resources, the amount of participation in English education activities, and the amount of teacher-student interactions, respectively. Intuitively, it can be seen that the interactive efficiency of social media in English education shows an upward trend with the increase of time, in which the reading volume of English courseware resources reaches a maximum of 18,477 times, and the teacher-student interaction data can reach a maximum of 2,444 items. The distribution of these time-series data is then analyzed from descriptive statistics, and the descriptive statistics of each type of interaction data of social media in English education are shown in Table 3. At 95% confidence level, the Shapiro-Wilk test results of all the social interaction time-series data rejected the original hypothesis, indicating that none of these time-series data have normal distribution characteristics.



(a) Reading quantity (b) Video playback (c) Learning resource collection



(d) Forwarding capacity (e) Participation quantity (f) Student interaction

Fig. 2. The results of the sequential analysis of social interaction data

Table 3. Descriptive statistics of social interaction data

Variable	Minimum value	Maximum value	Degree of bias	Kurtosis	Shapiro-wilk
Reading quantity	554	18477	-0.21	0.89	2.31E-16
Video playback	1192	9980	0.02	-0.94	5.24 E-16
Learning resource collection	528	18147	-0.09	-0.29	2.64 E-16
Forwarding capacity	534	19668	1.44	0.95	1.52 E-16
Participation quantity	72	5543	-0.7	-0.76	2.64 E-16
Student interaction	563	2444	0.5	-0.8	2.34 E-16

4.2.3. Predicted results of social media interaction efficiency. Based on the existing data related to the interaction efficiency of social media in English education in School A, this paper uses the EEMD-ARIMA-Markov time-series prediction model to predict the change of each indicator of the interaction efficiency of social media in English education in the next six months (March 16, 2023-September 16, 2023). The results of the interaction efficiency prediction are shown in Figure 3. The prediction results show that social media has a relatively ideal interaction efficiency in English education in School A. Until September 16, 2023, the reading and playing volume of English courseware resources and video resources uploaded by English teachers in social media can reach 88,629 and 84,571 times, respectively. At the same time, the maximum number of teacher-student interactions can reach 25,288, which indicates that the application of social media in English education is highly feasible and can promote the interaction between teachers and students, which in turn will have a positive impact on students' English education.

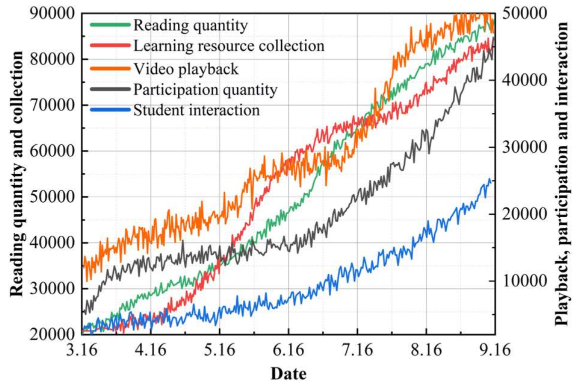


Fig. 3. Interactive efficiency prediction results

5. Conclusion

This paper combines the EEMD model, ARIMA model and the improved Markov model to establish the EEMD-ARIMA-Markov time series prediction model, which is utilized to explore the interaction efficiency of social media in English education. The performance test results show stocknet-dataset dataset, the EEMD-ARIMA-MARKOV model has a significant improvement of 0.941% and 0.0025 on ACC and MCC, respectively, which indicates the effectiveness of orthogonality in enhancing the generalization ability of the model. In addition, the F1 value of the model is 0.7426, which is better than the experimental results of other experimental models. The interaction efficiency of the interactive data of social media in English education in School A shows an increasing trend with time, and the normality analysis shows that none of these time-series data is characterized by normal distribution. The prediction results indicate that social media has a more desirable interaction efficiency in English education in School A. The maximum number of interactions between teachers and students on social media platforms in English education on September 16, 2023 can reach 25,288 interactions. This indicates that social media can promote the interaction between teachers and students in English education in this school, thus promoting students' interest in learning English and learning effect. Social media platforms in the context of the Internet have a positive effect on improving the effect of English education, laying a foundation for better independent learning and active learning in English education.

References

[1] Lin, C. Y., Hsieh, Y. M., Cheng, F. T., Huang, H. C., & Adnan, M. (2019). Time series prediction algorithm for intelligent predictive maintenance. IEEE Robotics and Automation Letters, 4(3), 2807-2814.

[2] Alhnaity, B., & Abbod, M. (2020). A new hybrid financial time series prediction model. Engineering Applications of Artificial Intelligence, 95, 103873.

[3] Chen, J., Li, K., Rong, H., Bilal, K., Li, K., & Philip, S. Y. (2019). A periodicity-based parallel time series prediction algorithm in cloud computing environments. Information Sciences, 496, 506-537.

[4] Han, Z., Zhao, J., Leung, H., Ma, K. F., & Wang, W. (2019). A review of deep learning models for time series prediction. IEEE Sensors Journal, 21(6), 7833-7848.

[5] Parmezan, A. R. S., Souza, V. M., & Batista, G. E. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. Information sciences, 484, 302-337.

[6] Anava, O., Hazan, E., Mannor, S., & Shamir, O. (2013, June). Online learning for time series prediction. In Conference on learning theory (pp. 172-184). PMLR.

[7] Messias, V. R., Estrella, J. C., Ehlers, R., Santana, M. J., Santana, R. C., & Reiff-Marganiec, S. (2016). Combining time series prediction models using genetic algorithm to autoscaling web applications hosted in the cloud infrastructure. Neural Computing and Applications, 27, 2383-2406.

[8] Li, K., Liu, L., Zhai, J., Khoshgoftaar, T. M., & Li, T. (2016). The improved grey model based on particle swarm optimization algorithm for time series prediction. Engineering Applications of Artificial Intelligence, 55, 285-291.

[9] Park, J. M., & Kim, J. H. (2017, May). Online recurrent extreme learning machine and its application to time-series prediction. In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 1983-1990). IEEE.

[10] Perrin, A. (2015). Social media usage. Pew research center, 125, 52-68.

[11] Selwyn, N. (2012). Social media in higher education. The Europa world of learning, 1(3), 1-10.

[12] Van Dijck, J., & Poell, T. (2018). Social media platforms and education. The SAGE handbook of social media, 579-591.

[13] Al Arif, T. Z. Z. (2019). The use of social media for English language learning: An exploratory study of EFL university students. Metathesis: Journal of English Language, Literature, and Teaching, 3(2), 224-233.

[14] Yadav, M. S. (2021). Role of social media in English language learning to the adult learners. International Journal of Linguistics, Literature and Translation, 4(1), 238-247.

[15] Selwyn, N., & Stirling, E. (2016). Social media and education… now the dust has settled. Learning, media and technology, 41(1), 1-5.

[16] Slim, H., & Hafedh, M. (2019). Social media impact on language learning for specific purposes: A study in English for business administration. Teaching English with technology, 19(1), 56-71.

[17] Rowan-Kenyon, H. T., & Alemán, A. M. M. (2016). Social Media in Higher Education: ASHE Higher Education Report, Volume 42, Number 5. John Wiley & Sons.

[18] Tay, E., & Allen, M. (2011). Designing social media into university learning: technology of collaboration or collaboration for technology?. Educational Media International, 48(3), 151-163.

[19] Desta, M. A., Workie, M. B., Yemer, D. B., Denku, C. Y., & Berhanu, M. S. (2021). Social media usage in improving English language proficiency from the viewpoint of medical students. Advances in Medical Education and Practice, 519-528.

[20] Muftah, M. (2024). Impact of social media on learning English language during the COVID-19 pandemic. PSU Research Review, 8(1), 211-226.

[21] Arumugam, N., Farhanahburhanuddin, N., Suppiah, P. C., & Razak, M. I. M. (2020). English vocabulary development among undergraduates: Social media. International Journal of Academic Research in Progressive Education and Development, 9(4).

[22] Han, S. (2022). Flipped classroom: Challenges and benefits of using social media in English language teaching and learning. Frontiers in Psychology, 13, 996294.

[23] Khan, I. U., Ayaz, M. U. H. A. M. M. A. D., Khan, S. A. Q. I. B., & Khan, M. F. (2016). Effect of social media on enhancement of English learning proficiency at university level in Khyber Pakhtunkhwa. Journal of Humanities and Social Sciences, 2(2), 71-84.

[24] Al-Jarrah, T. M., Al-Jarrah, J. M., Talafhah, R. H., & Mansor, N. (2019). The role of social media in development of English language writing skill at school level. International Journal of Academic Research in Progressive Education and Development, 8(1).

[25] Ahmed, S. N., Abbas, F., & Qureshi, A. M. (2021). The use of social-networking sites in English language

education: An exploratory study using SWOT analysis technique. Psychology and Education, 58(1), 4640-4650.

[26] Yu, Z., Yu, L., Xu, Q., Xu, W., & Wu, P. (2022). Effects of mobile learning technologies and social media tools on student engagement and learning outcomes of English learning. Technology, Pedagogy and Education, 31(3), 381-398.

[27] A V Anikeenko & Yu I Naberukhin. (2024). Temperature behavior of the velocity autocorrelation function in large MD models of water. The Journal of chemical physics(12).

[28] Gautam Ratnesh & Sinha Anand Kumar. (2019). Estimation of Autocorrelation function and partial Autocorrelation function of Evapotranspiration for Ranchi district, India, in context of time series Modelling. International Journal of Social Science Research(04),21-31.

[29] Yikui Mo,Shulian Li,Ying Zhao & Lanfang Wang. (2018). Analysis of China Aviation Network Structure Evolution Based on Ward System Clustering. IOP Conference Series: Earth and Environmental Science(6),062039-062039.

[30] Tullio Ceccherini Silberstein,Michel Coornaert & Hanfeng Li. (2024). Expansive actions with specification of sofic groups, strong topological Markov property, and surjunctivity. Journal of Functional Analysis(9),110376-.

[31] Yuchan Jie,Xiaosong Li,Tianshu Tan,Lemiao Yang & Mingyi Wang. (2025). Multi-modality image fusion using fuzzy set theory and compensation dictionary learning. Optics and Laser Technology(PC), 112001-112001.

[32] Lingjie Li & Li Lingjie. (2021). Configurations of Earth's magnetosphere in response to variations in z component Solar-wind flow during the southward IMF. Physica Scripta(1),015007-.