WILEY | Hindawi

Research Article

Adaptive Attention with Consumer Sentinel for Movie Box Office Prediction

Kaicheng Feng and **Xiaobing Liu**

School of Economics and Management, Dalian University of Technology, Dalian 116024, China

Correspondence should be addressed to Kaicheng Feng; 11311062@mail.dlut.edu.cn

To improve the movie box office prediction accuracy, this paper proposes an adaptive attention with consumer sentinel (LSTM-AACS) for movie box office prediction. First, the influencing factors of the movie box office are analyzed. Tackling the problem of ignoring consumer groups in existing prediction models, we add consumer features and then quantitatively analyze and normalize the box office influence factors. Second, we establish an LSTM (Long Short-Term Memory) box office prediction model and inject the attention mechanism to construct an adaptive attention with consumer sentinel for movie box office prediction. Finally, 10,398 pieces of movie box office dataset are used in the Kaggle competition to compare the prediction results with the LSTM-AACS model, LSTM-Attention model, and LSTM model. The results show that the relative error of LSTM-AACS prediction is 6.58%, which is lower than other models used in the experiment.

1. Introduction

The movie box office, as an indicator of the level of film development, has attracted great attention from all walks of life. At present, the prediction of the movie box office has become one of the hottest research by scholars [1]. Linear regression and nonlinear regression models used to construct social media-driven movie box office prediction models were proposed [2]. A new method of movie box office prediction based on two-level and twice proxy variables was proposed [3], which can predict the first week's box office by using some preindicators obtained before the movie is released. A single influencing factor of the movie box office was mainly analyzed [4]. They analyzed the influence of celebrity effect on box office. They concluded that celebrity influence is positively related to box office. The competition factors with similar movie release time on the standard regression framework were tested, and a more simplified empirical model was proposed [5]. A BRP feedback neural network was proposed to solve movie box office prediction and classification problems [6]. The prediction model using the BP neural network has the following

shortcomings. (1) Binary is used in the discretization of the model to quantify the various influencing factors of the movie box office [7]. These variables are not processed according to the actual situation, and the differences between different influencing factors cannot be fully expressed. (2) When using BP neural network for training, it is easy to fall into the problem of local minimization [8].

LSTM [9] is a time recurrent neural network. A movie box office prediction based on the LSTM model was proposed [10]. This model can solve the BP neural network using only simple Boolean coefficient values. It can map as many movie box office influence factors as possible in the input and output. However, its analysis of film sale influencing factors is not comprehensive, and the predicted results still have large relative errors.

Tackling the current movie box office prediction problem, this paper proposes an adaptive attention LSTM model with consumer sentinel. Compared with traditional LSTM, this model proposes an attention with consumer sentinel. On the one hand, it can fully consider the impact of movie consumer information on the movie box office and improve the model input. On the other hand, adaptive attention can

capture the input affective information more vigorously, thereby further improving the prediction accuracy. Specifically, the model is based on the LSTM model injecting the adaptive attention (AAM) with consumer sentinel. Consumer sentinel can identify the influence of the influencing factors of the movie box office from more dimensions and solve the long-standing problem of ignoring consumer information in box office forecasting. The use of LSTM takes into account the random volatility and long-time span of the movie box office. LSTM also remembers the information for a long time to predict the movie box office. Injecting an adaptive attention can capture affective input information, which provides a guarantee for the accuracy of the movie box office prediction results. The proposed model provides a certain reference value for film investors in film risk control, and it can have a certain planning value for film release schedule and has practical application prospects. The contribution of this paper can be summarized as follows. (1) To improve the movie box office prediction accuracy, this paper proposes an LSTM model with an AAM and consumer sentinel (LSTM-AACS). It can better capture consumer characteristics, thereby improving prediction accuracy. (2) The LSTM-AACS model is applied to the prediction of the movie box office and achieves good results. The results show that the relative error of LSTM-AACS prediction is 6.58%, which is lower than other models used in the experiment.

2. Literature Review

There are many factors including investment, director, actors, and sequel and play a role in promoting and guiding the film's box office. In [11], six independent variables of film investment were selected; film quality, director, actors, film sequel, and piracy. They established a linear regression model for influencing factors and movie box office. A semiparametric method was proposed to deal with random effects in a nonparametric way [12]. The example of comparing the reviews of movie critics uses the logit model of the adjacent category and the logit model of the related baseline category. Although this method eliminates the influence of extreme data, it also makes the use of data information insufficient.

The above research provides an important reference when selecting factors affecting the movie box office. Then, they used the Sawhney and Eliashberg model to predict the cumulative number of viewers of the movie after a few weeks of its release [13]. Its practical significance is that, during the life cycle of movie release, movie theaters can dynamically adjust the projection strategy. For example, movie producers can expand or reduce the number of theaters showing the movie, change the projection period, and so on. However, this method has the following shortcomings. (1) When using the multiple linear regression algorithm to predict the cumulative audience in the first week, few film influencing factors (number of film copies, user ratings, number of theaters, and audience age) are considered, and it did not consider the special attributes of the movie to attract the audience. This led to an excessively large prediction error for the first week. (2) This kind of error will accumulate when the

diffusion model is used to predict the number of viewers in the next few weeks, which will affect the final prediction accuracy.

Based on the multilayer neural network algorithm, multiple movie attributes that affect the box office were combined [14]. They proposed a movie box office classification model and used the classification accuracy as the main index to evaluate the classification performance of the model. They achieved good classification results. However, this method uses binary discrete numbers to quantify the various influencing factors of the movie box office, which is obviously a vague processing method. These variables are not quantified according to the actual situation, so they cannot fully reflect the different variables in the influencing factors. In addition, the classification of the movie box office in the output layer of the prediction model is also vague, making the classification of each movie box office level too large. Such classification is of little relevance value for film investors and movie theaters to control the cost of film production and screening. A multimodal deep neural network for movie box office revenues prediction was proposed [15]. A CNN was built for extracting features from movie posters. Then, a multimodal deep neural network was built to leverage both movie poster features and other movie-related data for movie box-office revenues prediction. In addition, the features of CNN learned from movie posters were analyzed. However, the research did not focus on building more multimode DNNs, nor did it merge audio and video data related to movies. In [16], a hybrid social recommender system utilizing a deep autoencoder network is introduced. The proposed approach employs collaborative and content-based filtering, as well as users' social influence. The social influence of each user is calculated based on his/her social characteristics and behaviors on Twitter. For the evaluation purpose, the required datasets have been collected from MovieTweetings and Open Movie Database. However, the dataset used in this study is not comprehensive enough and may have limitations in prediction accuracy.

The LSTM-AACS model used in this paper uses the LSTM model based on the adaptive attention. A lot of work has been proposed for the LSTM model based on the attention. The attention-based LSTM model was proposed for financial time series prediction [15], and the model prediction can be intuitively understood through the attention vector. In addition, their focus on time and factors makes it easy for people to understand why certain trends are predicted when accessing a given time series table. They also modified the loss function of the attention model using weighted classification crossentropy. However, there is a shortcoming that the error is small in the long-term forecast, and the performance in the short-term forecast is not ideal, with high errors. A forecasting framework was established to predict the opening prices of stocks [16]. They processed stock data through a wavelet transform and used an attention-based LSTM neural network to predict the stock opening price, with excellent results. However, simply considering the impact of historical data on price trends is too singular and may not be able to fully and accurately forecast the price on a given day. An attention-based long

short-term memory network for aspect-level sentiment classification was proposed [17]. The attention mechanism can concentrate on different parts of a sentence when different aspects are taken as input. However, its flaw is that different aspects are input separately, and it does not realize modeling of multiple aspects simultaneously with the attention mechanism. An attention-based LSTM network is proposed for cross-language sentiment classification [18]. They use bilingual bidirectional LSTM to model the sequence of words in the source and target languages. Based on the particularity of sentiment classification tasks, they proposed a hierarchical attention model that was jointly trained with LSTM network. The model has achieved gratifying results on the benchmark dataset with Chinese as the source language and English as the target language. However, the problem is that the performance of the model is not evaluated on more datasets and more language pairs. An attention-based LSTM model for the task of hashtag recommendation was proposed [19]. They adopted the architecture of LSTM to avoid hand-crafted features. Their model incorporates topic modeling into the LSTM architecture through an attention mechanism and takes over the advantages of the both. Through evaluations run on a large dataset from Twitter, they have demonstrated that the proposed method outperforms competitive baseline methods effectively. However, the present work does not consider the use of other types of data in microblogs for hashtag recommendation [20].

The main problems above are as follows. (1) It performs well in short-term prediction, but the effect is not ideal in long-term prediction. (2) The input data of the model is not comprehensive, which leads to the prediction results only in a certain dataset to achieve high prediction accuracy. (3) The influencing factors of the results in the prediction problem are not considered comprehensively, such as ignoring user information and resulting in low prediction accuracy. Based on the above problems, we propose an AAM for movie box office prediction with consumer sentinel. With consumer sentinel, it can solve the problem of ignoring consumer groups in previous predictions. AAM can capture effective input information well. Finally, the LSTM model based on the above two algorithms is used to predict the movie box office and compare with other models. Experiments show that the prediction accuracy of the AAM for the movie box office prediction model with consumer sentinel is better than other models used in the experiment.

3. Adaptive Attention Mechanism with Consumer Sentinel

3.1. Framework Design. The framework is shown in Figure 1. It can be seen that this paper adds consumer information to the previous movie box office influencing factors and injects an AAM into the LSTM neural network (its structure is shown in the blue box, and consumer sentinel are input into the model as features and then combined with the attention mechanism to train the LSTM model). This improves the prediction accuracy.

3.2. Determination of Influencing Factors

3.2.1. Factors of a Movie. This paper uses the statistical analysis of the historical box movie office data in China combined with the actual situation of the movie market. The paper selects the director, actor, film genre, nation, and release data as the film's own influencing factors (as the film's information input). This paper then assigns different weights to each factor. The calculation method will be explained in detail in Section 4.1.

3.2.2. Consumer Groups. Based on the consideration of a movie's own influencing factors mentioned in Section 3.2.1, this paper adds the age information of movie consumer groups. This is because every movie must have its audience. For example, military subjects are more suitable for viewing by teenagers and above, while cartoons have more children as the audience. Generally, elderly people rarely go to the cinema to watch movies and so on. The age information of consumers is used as input information, and weights are assigned to jointly predict the final box office of the movie.

3.3. Long- and Short-Term Memory Network Layer. LSTM is an improved RNN (Recurrent Neural Network) model that solves the problems of gradient explosion or gradient disappearance during RNN training. Different from the single tanh loop structure in standard RNN, LSTM is a special network with three "gates" [21,22]. They are the forget gate, input gate, and output gate. The forget gate is responsible for choosing to forget invalid information in the past. The input gate is responsible for determining that useful new information is stored in the cell state. The output gate determines the output information. The process of the memory module for status update and information output is as follows:

(1) The core of LSTM is cell: cell state is the memory transmission belt of the entire module that changes over time. The conveyor belt itself cannot control which information is memorized. The forget gate, input gate, and output gate play a controlling role.

(2) Forget state information: select the input x_t at the current moment and the memory unit state information h_{t-1} at the previous moment, and then use the sigmoid function to output a value of [0, 1] to indicate the degree to which historical information needs to be retained:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right), \tag{1}$$

(3) Update the status information and store useful new information in the cell state. First, calculate the value of the input gate. The function of the input gate is to control how the current data input affects the state value of the memory unit. Then, calculate the candidate memory unit information $\tilde{C}_t$ at the current time t, which contains the new information to be added. Finally, merge the old cell state $C_{t-1} * f_t$

Three input data set formats:
(1) Box office data
(2) Box office data + movie
information
(3) Box office data + consumer
information

Movie information

Director
Actor
Genre
Nation
Data

Normalized

Director weight$_i$
Actor weight$_i$
Genre weight$_i$
Nation weight$_i$
Data weight$_i$

Consumer information

Age

Box office data

Box office mixed dataset   LSTM – AACCS

X_1 X_2 X_3 X_i Input layer

LSTM   LSTM   LSTM   …   LSTM   LSTM layer

H_1 H_2 H_3 … H_i Self-attention mechanism layer

$\hat{\alpha}_t$

$\otimes$

X_i

h_{t-1} →   LSTM   → h_t

h_i

V → H_t

C_t

$\hat{\alpha}_t$

γ Full connected layer
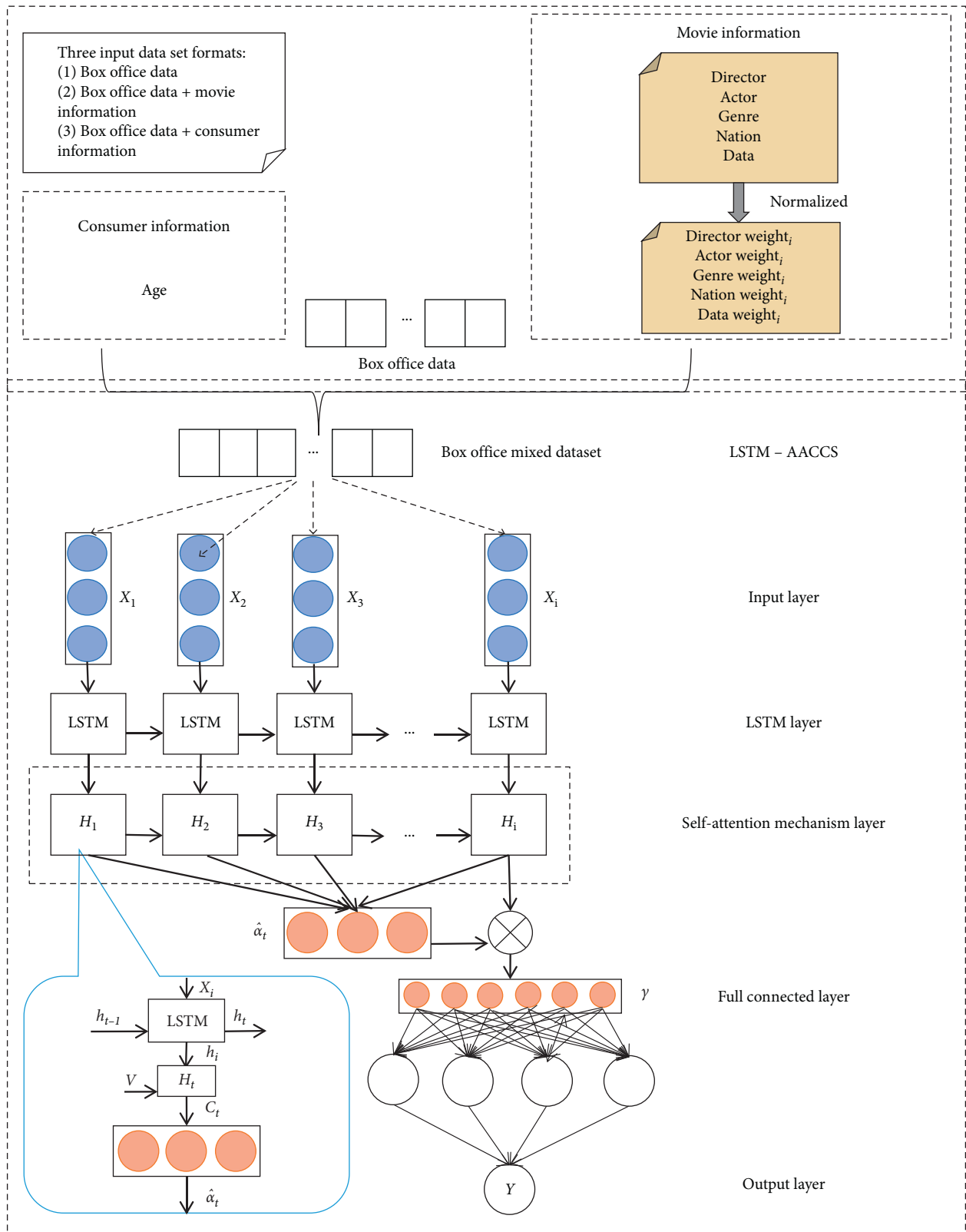
Y Output layer

FIGURE 1: Overall framework of the model.

(used for forgetting) with the new candidate information $i_t * \widetilde{C}_t$ to determine the updated information:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i),$$
$$\widetilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \qquad (2)$$
$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t.$$

(4) Output information: first determine which part of the state will be outputted. Finally, obtain the memory unit output information at the current time after the value of the output gate and the state information of the memory unit undergo tanh transformation:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \qquad (3)$$

$$h_t = o_t * \tanh(C_t). \qquad (4)$$

3.4. Adaptive Attention Mechanism Layer. This paper adds an AAM layer [22] to the method, which can better capture the affective information in the movie box office data and grasp the core data information. It overcomes the problem of the standard LSTM model using the same state vector in each step of the prediction, which results in the inability to fully learn the detailed information of the sequence encoding during the prediction. The specific extension method is based on the original LSTM model, adding two formulas:

$$g_t = \sigma(W_i x_t + W_h h_{t-1}),$$
$$s_t = g_t \odot \tanh(C_t), \qquad (5)$$

where W_i and x_t are the input of LSTM and $W_h \in R^{d \times d}$ is the parameter matrix that the subsequent model needs to train. C_t is a memory cell, called sentinel gate. It is similar to the input gate, forget gate, and output gate in LSTM. The structure of the formula is similar to (4). The vector $\hat{c}$ in the AAM can be expressed as follows:

$$\hat{c} = \beta_t s_t + (1 - \beta_t) c_t, \qquad (6)$$

where $\beta_t \in [0, 1]$ can be regarded as a sentinel gate in the true sense controlling the degree to which the model pays attention to C_t. The representation of β_t is as follows:

$$\beta_t = w_h^T \tanh(W_s s_t + (W_g h_t)). \qquad (7)$$

At the same time, the attention distribution α_t of the K areas of the AAM has also been expanded to $\hat{\alpha}_t$. The method is to splice an element after z_t:

$$\hat{\alpha}_t = \text{soft max}([z_t; \beta_t]), \qquad (8)$$

where $\hat{\alpha}_t$ has $k + 1$ elements, and the expression of z_t is as follows:

$$z_t = w_h^T \tanh(W_v V + (W_g h_t) 1^T). \qquad (9)$$

Finally, the probability distribution of the matrix can be expressed as follows:

$$\gamma = \text{soft max}(W_p(\hat{c} + th_t)), \qquad (10)$$

where W_p is the parameter matrix that the subsequent model needs to train. The output variable h_t finally passes through a fully connected layer and softmax classifier, which finalizes the prediction of movie box office.

4. Experiment and Discussion

4.1. Normalization of Impact Factors. This section will elaborate on the factors that affect the movie box office and give the corresponding definitions. At the same time, the quantification process of each attribute of the movie box office data will be given to prepare for the construction of the LSTM-AACS training set.

4.1.1. Director. Define movie box office influence of director i index as

$$\text{Dir}_i = \frac{\left(\sum_{j=1}^m \sum_{k=1}^5 b_{jk}\right)}{m}, \quad m = \min\{5, m\}, \qquad (11)$$

where i means director number, j represents the jth movie filmed by director i, k indicates the week of the release, m means the m movies with the closest release time from the current time among all the movies filmed by director i, and b_{jk} represents the box office during the kth week of the latest jth movie. Furthermore, the box office influences weight DirectorWeight$_i$ of the film directed by director i which can be obtained as follows:

$$\text{DirectorWeight}_i = \frac{(\lg(\text{Dir}_i/\text{Dir}_{\min}))}{(\lg(\text{Dir}_{\max}/\text{Dir}_{\min})),} \qquad (12)$$

where

$$\text{Dir}_{\max} = \max\{\text{Dir}_1, \text{Dir}_2, \ldots\},$$
$$\text{Dir}_{\min} = \min\{\text{Dir}_1, \text{Dir}_2, \ldots\}, \qquad (13)$$

where i represents the number of the director and Diri represents the influence of the ith director.

4.1.2. Actor. Define box office influence of actor i index as

$$A_i = \frac{\sum_{j=1}^m u_{ij}\left(\sum_{k=1}^5 b_{jk}\right)}{m}, \quad m = \min\{5, m\}, \qquad (14)$$

where i means actor number, j represents the jth movie filmed by the actor i, k indicates the week of the release, m means the m movies with the closest release time from the current time among all the movies filmed by actor i, b_{jk} represents the box office during the kth week of the latest jth movie, and u_{ij} is the participation coefficient of the jth movie actor i recently participated in, and it is defined as follows:

$$u_{ij} = \begin{cases} 1 - \dfrac{(n-1)}{10}, & n \in [1, 5], \\ \\ 0.5, & n \in (5, +\infty), \end{cases} \qquad (15)$$

where n is a positive integer, indicating the order of the actor i in the jth movie. Furthermore, the box office influence weight ActorWeight$_i$ of the film directed by actor i can be obtained as follows:

$$\text{ActorWeight}_i = \frac{(\lg(A_i/A_{\min}))}{(\lg(A_{\max}/A_{\min}))}, \tag{16}$$

where

$$\begin{aligned} A_{\max} &= \max\{A_1, A_2, \ldots\}, \\ A_{\min} &= \min\{A_1, A_2, \ldots\}, \end{aligned} \tag{17}$$

where i represents the number of the actor and A_i represents the influence of the ith actor.

4.1.3. Movie Genre. Define movie box office influence of movie genre i index as

$$G_i = \sum_{j=1}^{m} \sum_{k=1}^{5} b_{jk}, \tag{18}$$

where i means genre number ($i = 1,2, \ldots, 9$), k indicates the week of the release, m represents the week of screening of genre i, j represents the jth movie belonging to genre i, and b_{jk} represents the box office of the jth movie with content genre i in the kth week of its release. The box office influence weight GenreWeight$_i$ of the film of genre i can be obtained as follows:

$$\text{GenreWeight}_i = \frac{(\lg(G_i/G_{\min}))}{(\lg(G_{\max}/G_{\min}))}, \tag{19}$$

where

$$\begin{aligned} G_{\max} &= \max\{G_1, G_2, \ldots\}, \\ G_{\min} &= \min\{G_1, G_2, \ldots\}. \end{aligned} \tag{20}$$

4.1.4. Nation. Define box office influence of movie nation i index as

$$N_i = \sum_{j=1}^{m} \sum_{k=1}^{5} b_{jk}, \tag{21}$$

where i means nation number($1 \le i \le 5$), the value of i from 1 to 5 corresponds to Europe, America, Japan, Korea, Hong Kong, and Taiwan, Mainland China, and other regions, k indicates the week of the release, m represents the total number of movies in the distribution nation i belonging to the area, j represents the jth movie in the distribution nation i, and b_{jk} represents the box office during the kth week of the release of the jth movie in the distribution nation i. The box office influence weight NationWeight$_i$ of the film directed by nation i can be obtained as follows:

$$\text{NationWeight}_i = \frac{N_i}{\sum_{j=1}^{5} N_j}, \tag{22}$$

where i represents the serial number of the issuance area, N_i represents the influence of the issuance area i, and N_j represents the influence weight of the issuance area j.

4.1.5. Data. Define box office influence of release data i index as

$$D_i = \sum_{j=1}^{m} \sum_{k=1}^{4} b_{jk}, \tag{23}$$

where i means data number($1 \le i \le 4$), the value of i from 1 to 4 corresponds to the Lunar New Year file, the 51st file, the summer file, and the eleventh file. k indicates the week of the release, m represents the total number of movies with the release date in schedule i, j represents the jth movie belonging to data i, and b_{jk} represents the box office data generated during the kth week of the release date of the jth movie with the release date in schedule i. The weight DataWeight$_i$ measures the box office influence of the type on the movie attributable to that type:

$$\text{DataWeight}_i = \frac{(\lg(D_i/D_{\min}))}{(\lg(D_{\max}/D_{\min}))}, \tag{24}$$

where

$$\begin{aligned} D_{\max} &= \max\{D_1, D_2, \ldots\}, \\ D_{\min} &= \min\{D_1, D_2, \ldots\}, \end{aligned} \tag{25}$$

where i represents the serial number of the data and D_i represents the influence of the data i.

4.1.6. Consumer Group. This paper is divided into 4 age groups: under 18, 18–45, 46–69, and over 69. Define box office influence of movie nation i index as

$$N_i = \sum_{j=1}^{m} \sum_{k=1}^{5} b_{jk}, \tag{26}$$

where i means age group number($1 \le i \le 4$) the value of i ranges from 1 to 4 corresponding to ages under 18 years old (excluding 18 years old), 18–45 years old, 46–69 years old, and over 69 years old (excluding 69 years old), k indicates the week of the release, m represents the total number of movies in age group i, j represents the jth movie in the distribution age i, and b_{jk} represents the box office during the kth week of the release of the jth movie in the distribution age i. The box office influence weight AgeWeight$_i$ of the film indexed by age group i can be obtained as follows:

$$\text{AgeWeight}_i = \frac{(\lg(\text{Age}_i/\text{Age}_{\min}))}{(\lg(\text{Age}_{\max}/\text{Age}_{\min}))}, \tag{27}$$

where

$$\begin{aligned} \text{Age}_{\max} &= \max\{\text{Age}_1, \text{Age}_2, \ldots\}, \\ \text{Age}_{\min} &= \min\{\text{Age}_1, \text{Age}_2, \ldots\}, \end{aligned} \tag{28}$$

where i represents the serial number of the age group and Age$_i$ represents the influence of the age group i.

TABLE 1: Specific error table.

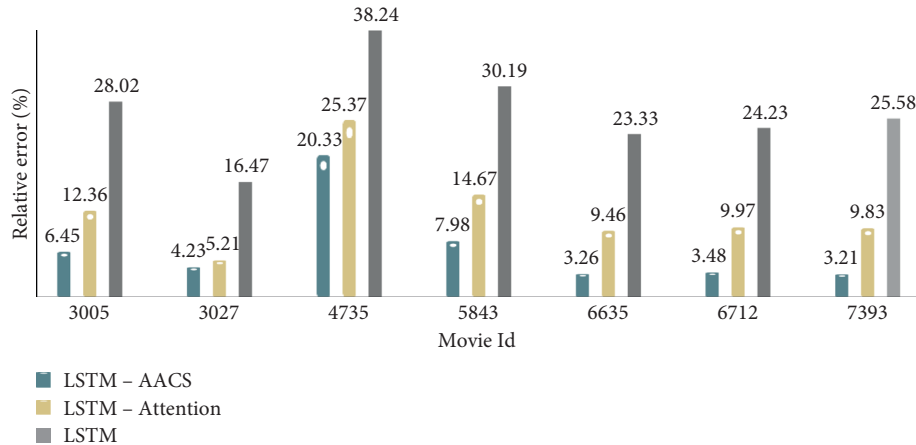| Method	Average relative error (%)
LSTM	28.54
LSTM-attention	11.45
LSTM-AACS	6.58

FIGURE 2: Relative error comparison.

4.2. Model Parameters.

In the LSTM-AACS model, we set the dropout rate to 0.5. During training, a small batch stochastic gradient descent method is used to reduce the training loss, and the minibatch is set to 64. When analyzing the prediction results, the relative error is used for analysis. This paper uses the movie box office prediction dataset in the Kaggle competition. After obtaining the movie box office prediction data, the calculation formula for the relative error of the prediction result is as follows:

$$\text{Relative error} = \frac{\text{prediction} - \text{real}}{\text{real}} \times 100\%. \tag{29}$$

4.3. Comparison and Analysis of Experimental Results

4.3.1. Error Comparison.

In the experiment, the results are analyzed by crossvalidation. This paper randomly takes out the first 3000 pieces of data from 10,398 pieces of data as the training set, and the last 7398 pieces of data as the test set. Learning is done 30 times when training the model, and then ten crossvalidation rounds are applied. Finally, the average relative error of these two models is shown in Table 1.

It can be seen from Table 1 that the average relative error of LSTM time series and LSTM-Attention, using ten crossvalidation, is higher than the relative error of the model proposed in this paper. It shows that the LSTM-AACS model is better than the LSTM model and the general LSTM model with attention for movie box office prediction.

Randomly selecting the prediction results of several movies from the test set, Figure 2 shows the comparison of partial relative errors of the two models under the training set.

From the seven randomly selected movies, it can be seen that the relative error predicted by the LSTM-AACS model in the test set is lower than the relative error predicted by the LSTM model and the LSTM-Attention model. The prediction results of the LSTM-AACS model is relatively more accurate, and the performance is improved.

4.3.2. Result Comparison.

In order to make the model have both long-term and short-term prediction capabilities, we compare the long term prediction capabilities of the LSTM-AACS model, the LSTM-Attention model, and the LSTM time series model. We also compare and analyze the movie box office data in the Kaggle competition. Additionally, we choose the box office data of a Maoyan movie to predict the short term box office. Considering the classic movies of previous years, this paper chooses Dangal; My People, My Country; Wolf Warriors II and Fast & Furious 7 as the movies to predict their cumulative box office. Using these movies, this paper compares the actual value, predicted value, absolute difference, and relative error of the three models, respectively. The specific results are shown in Table 2.

As can be seen from Table 2, the relative error of the LSTM-AACS model in predicting the cumulative box office of the above four movies is lower than the relative error of the prediction results of the LSTM model and the LSTM-Attention model. This proves the feasibility of the LSTM-AACS model proposed in this paper in predicting movie box office. This also proves that the LSTM-AACS model can have a better evaluation effect on movie investors.

TABLE 2: Movie box office prediction results of previous classic movies.

Movie name	Real (100 million)	LSTM			LSTM-attention			LSTM-AACS		
		Prediction (100 million)	Absolute error (100 million)	Relative error (%)	Prediction (100 million)	Absolute error (100 million)	Relative error (%)	Prediction (100 million)	Absolute error (100 million)	Relative error (%)
Dangal	12.99	14.31	1.32	10.16	13.98	0.99	7.62	13.18	0.19	1.46
My People, My Country	31.71	35.42	3.71	11.70	33.67	1.96	6.18	32.59	0.88	2.78
Wolf Warriors II	56.92	60.21	3.29	5.78	58.66	3.10	5.45	58.34	0.52	0.91
Fast & Furious 7	24.27	27.66	3.39	14.00	25.78	1.51	6.22	25.57	1.3	5.36

5. Conclusion

Tackling the problems of ignoring consumer factors and low prediction accuracy in movie box office prediction, this paper proposes an adaptive attention movie box office prediction model with consumer sentinel. The experimental results show that the introduction of consumer data into the prediction model can improve the prediction accuracy on the basis of a movies own influencing factors. Compared with a single LSTM model and an LSTM model with an attention mechanism, the LSTM model with AAM has better prediction capabilities for movie box office prediction. In the future, the model can be further optimized by enriching the characteristics of expert experience, introducing more consumer characteristics, and adding movie reviews as an influencing factor.

Data Availability

The data used to support the findings of the study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] P. Nagamma, H. R. Pruthvi, K. K. Nisha, and N. H. Shwetha, "An improved sentiment analysis of online movie reviews based on clustering for box-office prediction," in *Proceedings of the International Conference on Computing, Communication & Automation*, pp. 933–937, Noida, India, May 2015.

[2] T. Liu, X. Ding, Y. Chen, H. Chen, and M. Guo, "Predicting movie Box-office revenues by exploiting large-scale social media content," *Multimedia Tools and Applications*, vol. 75, no. 3, pp. 1509–1528, 2016.

[3] F. W. Derrick, N. A. Williams, and C. E. Scott, "A two-stage proxy variable approach to estimating movie box office receipts," *Journal of Cultural Economics*, vol. 38, no. 2, pp. 173–189, 2014.

[4] R. A. Nelson and R. Glotfelty, "Movie stars and box office revenues: an empirical analysis," *Journal of Cultural Economics*, vol. 36, no. 2, pp. 141–166, 2012.

[5] J. Prieto-Rodriguez, F. Gutierrez-Navratil, and V. Ateca-Amestoy, "Theatre allocation as a distributor's strategic variable over movie runs," *Journal of Cultural Economics*, vol. 39, no. 1, pp. 65–83, 2015.

[6] J. Du, H. Xu, and X. Huang, "Box office prediction based on microblog," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1680–1689, 2014.

[7] L. Zhang, J. Luo, and S. Yang, "Forecasting box office revenue of movies with BP neural network," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6580–6587, 2009.

[8] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: a search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

[9] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," in *Proceedings of the 13th Annual Conference of the International Speech Communication Association*, Portland, OR, USA, September 2012.

[10] C. Yang, W. Jiang, and D. Shao, "Movie box office prediction algorithm based on LSTM model," *Data Communication*, vol. 5, p. 9, 2019.

[11] P. Marshall, M. Dockendorff, and S. Ibáñez, "A forecasting system for movie attendance," *Journal of Business Research*, vol. 66, no. 10, pp. 1800–1806, 2013.

[12] J. Hartzel, A. Agresti, and B. Caffo, "Multinomial logit random effects models," *Statistical Modelling: An International Journal*, vol. 1, no. 2, pp. 81–102, 2001.

[13] D. Barman, N. Chowdhury, and R. K. Singha, "To predict possible profit/loss of a movie to be launched using MLP with back-propagation learning," in *Proceedings of the 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS)*, pp. 322–325, Kolkata, India, December 2012.

[14] R. Sharda and D. Delen, "Predicting box-office success of motion pictures with neural networks," *Expert Systems with Applications*, vol. 30, no. 2, pp. 243–254, 2006.

[15] Y. Zhou, L. Zhang, and Z. Yi, "Predicting movie box-office revenues using deep neural networks," *Neural Computing and Applications*, vol. 31, no. 6, pp. 1855–1865, 2019.

[16] H. Tahmasebi, R. Ravanmehr, and R. Mohamadrezaei, "Social movie recommender system based on deep autoencoder network using Twitter data," *Neural Computing and Applications*, pp. 1–17, 2020.

[17] S. Kim and M. Kang, "Financial series prediction using attention LSTM," 2019, http://arxiv.org/abs//1902.10877.

[18] J. Li, S. X. Pan, L. Huang, and X. Zhu, "A machine learning based method for customer behavior prediction," *Tehnicki Vjesnik-Technical Gazette*, vol. 26, no. 6, pp. 1670–1676, 2019.

[19] L. Wang, Z. Hao, X. Han, and R. Zhou, "Gravity theory-based affinity propagation clustering Algorithm and its Applications," *Tehnicki Vjesnik-Technical Gazette*, vol. 25, no. 4, pp. 1125–1135, 2018.

[20] J. Niresh, N. Archana, and G. Anand Raj, "Optimisation of linear passive suspension system using MOPSO and design of predictive tool with artificial neural network," *Studies in Informatics and Control*, vol. 28, no. 1, pp. 105–110, 2019.

[21] X. Fan, H. Lin, L. Yang et al., "Phonetics and ambiguity comprehension gated attention network for humor recognition," *Complexity*, vol. 2020, Article ID 2509018, 2020.

[22] N. Wang, L. Yang, Y. Zheng, X. Cai, X. Mei, and H. Dai, "A tri-attention neural network model-based recommendation," *Complexity*, vol. 2020, Article ID 3857871, 2020.