

Analysis of Public Opinion on the Network of Burst Events based on the PH-LSTM Neural Network

Aiying Ban^{1, 2, *}, Zhiyong Zhang^{1, 2}

¹Information Engineering College, Henan University of Science and Technology, Luoyang, Henan 471023, China

²Henan International Joint Laboratory of Cyberspace Security Applications, Henan University of Science and Technology, Luoyang, Henan 471023, China

*ban168985@163.com

Abstract

Due to the high-speed development of the Internet today, once a burst event occurs, it will attract the general public's attention in a short time and spread rapidly on the Internet, which will lead to the formation of large-scale network public opinion on emergencies. We can accurately judge the development trend of bursts and provide a reference for emergency supervisory departments to deal with public opinion crises by predicting the trend of public opinion on the network of emergencies. To address the issues of a single prediction model's low prediction accuracy and the large influence of social media factors on public opinion trend, an opinion prediction method integrating social media influence and LSTM neural network is proposed, and a PH-LSTM (Post Hot-Long Short-Term Memory) neural network is built. Firstly, the influencing factors of online public opinion are introduced, including the posting hotness and user influence of fused social media influence and the growth rate of events. Secondly, the LSTM neural network is improved to design a PH-LSTM prediction model with three hidden layers, and the PH-LSTM model is used in the quantitative prediction of unexpected events. Model's validity was verified through experiments, and it was confirmed that the model has a good fitting and prediction effect on the trend of online public opinion of unexpected events.

Keywords

Analysis of Online Public Opinion on Burst Events; Social Media Influence; LSTM; Heat of the Posting; User Participation.

1. Introduction

The rapid development of Internet technology has accelerated information exchange and increased the number of Internet users. As of June 2022, China's Internet population had grown to 1.051 billion, with a penetration rate of up to 74.4%. Internet public opinion on emergencies refers to the information containing a variety of emotions, attitudes and opinions released by netizens around the upcoming or occurred emergencies using the Internet as a platform[1]. Once a sudden event occurs, after different online media reports, it will attract the attention of the general public in a short time, and a large number of postings related to the event will be repeatedly commented on and reprinted by netizens and spread rapidly on the Internet. Thus, it will form a sudden event online public opinion.

Many scholars, both domestic and international, have conducted extensive research on the online public opinion of burst events. Dong et al. [2] considered the number of posts posted at each time interval as a time series dataset with multiple small peaks and nonlinear characteristics. They proposed a hybrid back propagation neural network (BPNN) model to

predict the characteristics of such time series datasets. Zhao et al. [3] built an information propagation model for emergencies by combining empirical data with simulation experiments, based on a complex network theory, information propagation theory, and disease propagation theory, and combined it with the study of the murdered Wen ling doctor case. Lv et al. [4] proposed a neural network-based model for predicting the amount of sudden public opinion retweets on campus microblogs using microblog data as an example, which solved the problem of low prediction accuracy of traditional models such as support vector machines. Li et al. [5] proposed a multi-factor propagation model with multi-intelligence modelling, was built on the SEIR model and NetLogo simulation analysis, integrated multiple online opinion propagation features, and extended the existing single online opinion propagation model. Zhao et al. [6] constructed a SIR model of opinion transmission based on the novel coronavirus pneumonia model and microblog information on public health emergencies, taking into account the critical role of opinion leaders in microblogs and users' interest in microblog information. The methods in the literature mentioned above can simulate the propagation trend of online public opinion more accurately and make predictions about it. Still there is an absence of studies that analyze the factors affecting the propagation of online public opinion.

In order to solve the above problems, this paper proposes a PH-LSTM neural network-based model for analyzing the online public opinion of unexpected events. The main contributions of this paper are.

- (1) Comprehensive analysis of the factors affecting the online public opinion of the breaking news, including the posting hotness and user participation of fused social media influence, the event growth rate, and the average of the Baidu index and WeChat index of the breaking news as the current online public opinion value of the breaking news.
- (2) Improving the traditional LSTM neural network to reduce the prediction error sum of squares of the sudden-onset web-based public opinion trend prediction model. Through experimental analysis, the PH-LSTM prediction model proposed in this paper reduces the sum of squared errors by one order of magnitude compared with the other three prediction models. The fitting and prediction effects are significantly improved.

The remainder of this paper is organized as follows: Section 2 describes the work related to the analysis of online emergency opinion trends; Section 3 introduces the influencing factors of online opinion trends; Section 4 gives the methods of online opinion trend prediction; Section 5 presents detailed experimental results and analysis; Section 6 concludes and outlooks the work of this paper.

2. Related Work

Emergencies are a dynamic and changing process, and how to analyze the trend of online public opinion on emergencies is the fundamental online public opinion monitoring. At present, scholars at home and abroad have proposed a large number of analytical methods for research on the prediction of public opinion on the network of unexpected events. In terms of traditional statistics-based online opinion prediction, Haneen et al. [7] used an ARIMA best-fit model to verify the accuracy of predicted values over a relatively long period; Lee et al. [8] used a moving average model to predict futures and argued that the optimal length of the moving average depends on the frequency of structural disruptions; Tratar et al. [9] developed an easy-to-implement exponential smoothing prediction method; Luo et al. [10] developed a prediction model based on a logistic model and used an example of Sina Weibo to demonstrate that the model can effectively predict the development of microblog opinion hotspots in a self-organized state. However, the nonlinear fitting ability of the above models is poor. Based on this, research scholars began introducing machine intelligence learning methods for online public opinion hotness prediction. Li et al. [11] combined data from the Internet with the

Hidden Markov Model to create a modern government public policy and public opinion monitoring and early warning model that can accurately predict the development trend of online hot public opinion. Fang et al. [12] used natural language processing technology to analyze emotions on hot topics and proposed a new prediction method for online public opinion; He et al. [13] built a BP neural network-based prediction model for microblog public opinion based on the characteristics of microblog social platform, which can quantitatively analyze and predict the development trend of public opinion in an efficient way. Cheng et al. [14] used a Logistic curve and BP neural network as the base model to build a variable-weight combination prediction model to fit and predict the changing trend of online public opinion, which improved the accuracy of fitting and predicting the overall trend of public opinion; Sang et al. [15] considered the mutual influence among users in the process of information dissemination and proposed a method based on attention mechanism, aiming to explore the influence of users in social networks. The method aims to explore the influence of users in the process of information dissemination in social networks to predict the trend of information dissemination.

These algorithms have strong self-learning ability and have obvious advantages over traditional single prediction, which has received significant attention from researchers and scholars. However, since machine learning algorithms require a large amount of data, and there is fewer data on online public opinion when an emergency event first breaks out, and its evolution is affected by many factors, machine learning algorithms are prone to the problem of local optimum or overfitting in predicting online public opinion on emergency events. Given, because of this, we comprehensively consider the factors affecting the change of public opinion trends of breaking events and improve the LSTM neural network structure to fit and predict the trend of breaking events online public opinion.

3. Analysis of Factors Influencing Online Public Opinion

The most direct manifestation of Internet users' concern about an emergency event is equal to the number of posts. The more posts related to an emergency event per unit of time, mean the more the emergency event is concerned by Internet users, and Internet public opinion is also affected by this. There are numerous factors influencing the trend of online public opinion on emergencies, such as subjective factors, including the hotness of postings, user participation and the growth rate of the event, and objective factors, including the Baidu index and WeChat index of the emergencies. This paper integrates the subjective and objective factors affecting the trend of online public opinion of unexpected events.

In this paper, we adopt a formal approach to defining a posting as a six-tuple $pt = (i_{pt}, n_{rt}, n_{ct}, n_{lt}, s, c)$. Where i_{pt} denotes the unique identifier of the posting, n_{rt} denotes the number of times the posting has been retweeted, n_{ct} denotes the number of times the posting has been commented, n_{lt} denotes the number of times the posting has been liked, s denotes the source of the posting, and c denotes the content of the posting.

Social media influence reflects the influence of mainstream social media on the Internet. With the promotion of "Internet +" and Web 2.0 technology, Weibo and Zhihu have become the open social mainstream media in Chinese society. Social media influence is calculated as follows.

$$S_{influence} = \frac{V_{RL}}{R_{Alexa}} * (W_{Baidu} + V_{PR}) \quad (1)$$

Where, V_{RL} is the number of backlinks, W_{Baidu} is the Baidu weight, V_{PR} is the Page Rank value, and R_{Alexa} is the Alexa weekly ranking.

Normalization of social media influence weight using a logarithmic function.

$$W_{S_{influence}} = \frac{\log_{10}(S_{influence})}{10} \quad (2)$$

The hotness of a post reflects the hotness of the post. It is obtained by adding up the weights of the number of retweets, comments and likes related to the post in the current period and multiplying them with the product of social media influence, calculated as follows.

$$H_{score} = \sum (\alpha n_{rt} + \beta n_{ct} + \lambda n_{lt}) \bullet W_{S_{influence}} \quad (3)$$

where, α 、 β 、 λ is the weight of the number of retweets, comments, and likes, respectively, and $\alpha + \beta + \lambda = 1$. $W_{S_{influence}}$ denotes the social media influence of the post in the time slice t .

User participation is the ratio of the sum of the number of likes, retweets and comments of the posts related to the event to the number of posts in the current period, and the larger the value is, the greater the user engagement in the breaking event. In t time slices, the user's involvement is given by the following equation.

$$N_{participation} = \frac{1}{N_{pt}} \sum (n_{rt} + n_{ct} + n_{lt}) \bullet W_{S_{influence}} \quad (4)$$

Where N_{pt} denotes the sum of the number of postings in a time slice t .

The event growth rate reflects the increase in the sum of the number of posts and retweets in social media during the current period compared to the previous period.

$$N_{growth} = \frac{(N_{pt} + N_{rt} + N_{ct}) - (N_{pt-1} + N_{rt-1} + N_{ct-1})}{N_{pt-1} + N_{rt-1} + N_{ct-1} + 1} \quad (5)$$

Where, N_{rt} denotes the sum of retweets of posts in time slice t , N_{ct} denotes the sum of comments of posts in time slice t , N_{pt-1} denotes the sum of posts in time slice $t-1$, N_{rt-1} denotes the sum of retweets of posts in time slice $t-1$, and N_{ct-1} denotes the sum of comments of posts in time slice $t-1$.

Baidu index of events is based on massive amounts of Internet user behavior data collected by Baidu in order to study keyword search trends, gain insight into Internet users' interests and needs, monitor public opinion trends, and identify audience characteristics. Searching a keyword of an event in the Baidu search index can reveal the event's trend graph.

The WeChat index of the event comes from the heat of Soyobo, video numbers, public number articles, web pages and some advertising channels. The calculation mainly examines the popularity of the keyword-related content and the importance of the keyword in the related content.

4. Method of Predicting the Trend of Online Public Opinion

LSTM neural network is a new machine learning neural network that improves the internal structure of RNN recurrent neural network. LSTM replaces the cells in the hidden layer with LSTM cells c so that it has a certain memory capability. At the same time, the existence of the input gate and the output gate of the LSTM neural network makes its gradient will not disappear or explode during training. The structure of the LSTM cell is shown in Fig1.

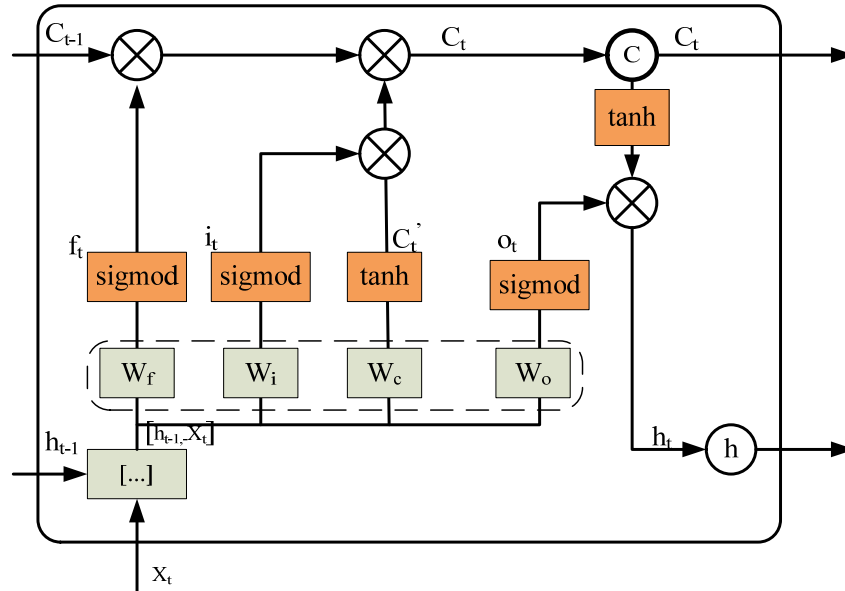


Figure 1. LSTM neural network

In Figure 1, X_t and h_t stand for the input and output vectors of the input layer and the hidden layer, respectively and the sigmoid and tanh symbols represent the sigmoid activation function and the hyperbolic tangent activation function, respectively. LSTM model training uses the standard backpropagation algorithm. Due to the unique internal structure of LSTM, it can achieve higher prediction accuracy in public opinion prediction compared with other algorithms. Therefore, LSTM neural network is chosen as the prediction module of the model in this paper.

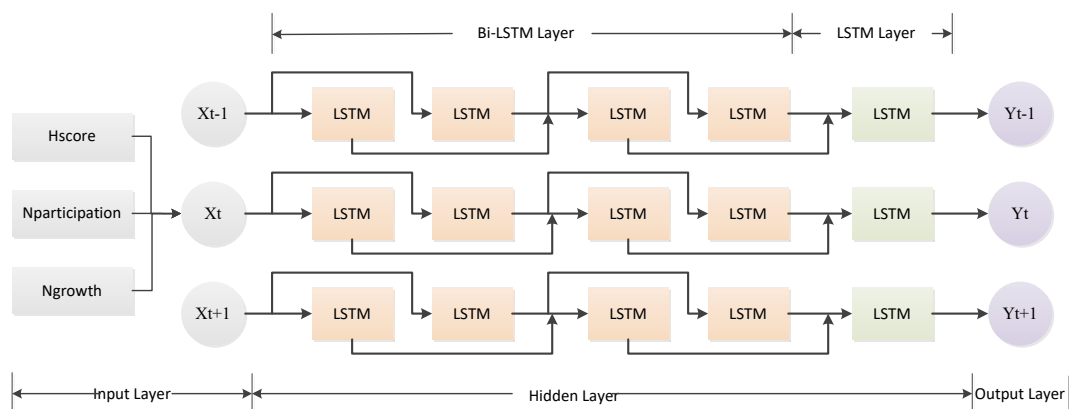


Figure 2. PH-LSTM neural network structure

In this paper, we combine social network data, LSTM network and Bi-LSTM neural network to propose PH-LSTM (post hotspot-long short-term memory) neural network model. Compared with other time-series data such as stock prices and traffic logistics, the amount of sudden-event online opinion data is relatively small, and Compared with other time-series data such as

stock prices and transportation and logistics, the volume of unexpected online opinion data is relatively small, and most of the unexpected online opinion data last about 15-20 days from generation to extinction. Due to this feature, the model comprises of two double LSTMs and one unidirectional LSTM implicit layer in order to avoid the danger of overfitting caused by the sparse training data while maintaining the LSTM network's properties. In Figure 2, the precise structure is displayed.

According to the characteristics of online opinion prediction, the input of the first implicit layer of the model includes three parts: the hotness of the posting, the user participation, and the growth rate of the event, and the output is the online public opinion value of the unexpected event.

The specific formula is as follows.

$$h_t^1 = \sigma(W_1[h_{t-1}^1, H_{score_t}, N_{participation_t}, N_{growth_t}]) \quad (6)$$

In equation (6), h_t^1 denotes the output of the first implied layer at the moment t , W_1 denotes the weight vector of the first implied layer, H_{score_t} denotes the posting hotness at the moment t , $N_{participation_t}$ denotes the user engagement at the moment t , and N_{growth_t} denotes the growth rate of events at the moment t .

The input of the 2nd implicit layer of the model consists of 2 parts: the output of the previous implicit layer at the same moment and the output of the previous time slice of the same implicit layer with the following equation.

$$h_t^2 = \sigma(W_2[h_{t-1}^2, H_t^1]) \quad (7)$$

In equation (7), h_t^2 denotes the output of the 2nd hidden layer at a time t , W_2 denotes the weight matrix of the 2nd hidden layer, and H_t^1 denotes the input from the 1st hidden layer to the 2nd hidden layer at a time t .

The input of the 3rd hidden layer of the model consists of 2 parts, the output of the previous hidden layer at the same moment and the output of the previous time slice of the same hidden layer, as follows.

$$h_t^3 = \sigma(W_3[h_{t-1}^3, H_t^2]) \quad (8)$$

In equation (8), h_t^3 denotes the output of the 3rd hidden layer at a time t , W_3 denotes the weight matrix of the 2nd hidden layer, and H_t^2 denotes the input from the 2nd hidden layer to the 3rd hidden layer at a time t .

The loss function of this model is the sum of squared prediction errors, as follows.

$$MSE = \sum_{i=1}^n (h(x_i) - y_i)^2 \quad (9)$$

In Eq. (9), n is the number of samples, $h(x_i)$ denotes the predicted output of the model at the input sample x_i , and y_i is the label of sample x_i .

5. Experimental Results and Analysis

5.1. Dataset

This paper takes the "assassination of Shinzo Abe" incident that occurred on July 8, 2022 as the research object, and chooses social media with a high influence, extensive coverage, quick update speed, and strong authority, such as Sina Weibo, Zhihu and Today's headlines.

In this paper, we use Python 3.7.0 and Pycharm programming tool to write a distributed crawler to obtain the content posted on social media, and randomly capture more than 12,004 posts posted in online social media from July 1, 2022 to July 31, 2022, which constitute the raw data of this experiment for public opinion analysis. Each data set contains the posting time, content, posting source, online opinion link, number of retweets, likes, and comments on social media. The daily posting data release volume is shown in Figure 3.

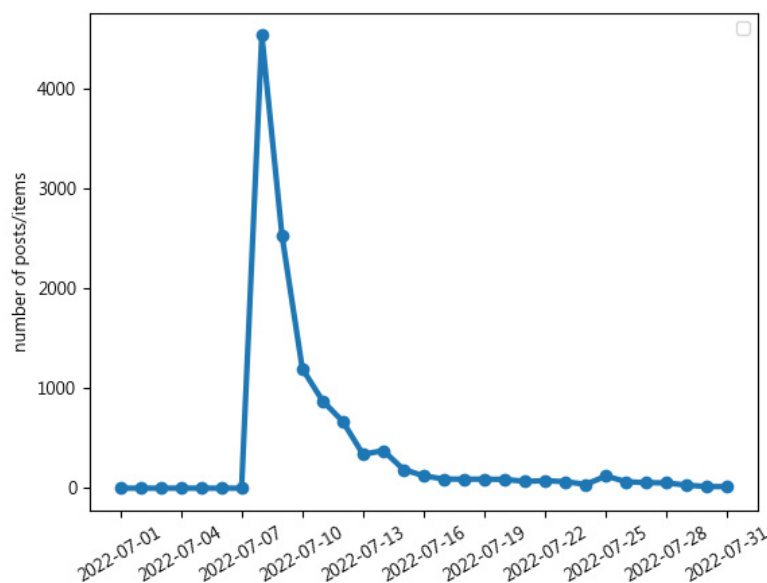


Figure 3. Number of daily postings from July 1, 2022 - July 31, 2022

5.2. Analysis of Experimental Results

The paper compares and analyzes the PH-LSTM prediction model with the BP back propagation neural network, the traditional LSTM long and short-term memory neural network model, and the Bi-LSTM bidirectional long and short-term memory neural network model. BP back propagation neural network is a relatively classical method for opinion prediction, mainly using the gradient descent method to continuously adjust and correct the weights and thresholds of each layer, which is the minimum sum of squared errors between the actual output value and the expected value of the network. In this experiment, the PH-LSTM network structure includes three hidden layers. The first two hidden layers consist of the Bi-LSTM network, and the last one hidden layer consists of the LSTM network.

Table 1. Experimental results data

Model	MSE
BPNN	0.01321
LSTM	0.10812
Bi-LSTM	0.14931
PH-LSTM	0.00495

Through experimental analysis, the mean square error values of each model are shown in Table 1. From the mean square error values, it can be seen that the PH-LSTM model is optimal, the BP direction propagation neural network is second, the LSTM and Bi-LSTM are consistent in order of magnitude, and the model effect is relatively poor.

The four prediction models' relative error curves for predicted and actual values are shown below.

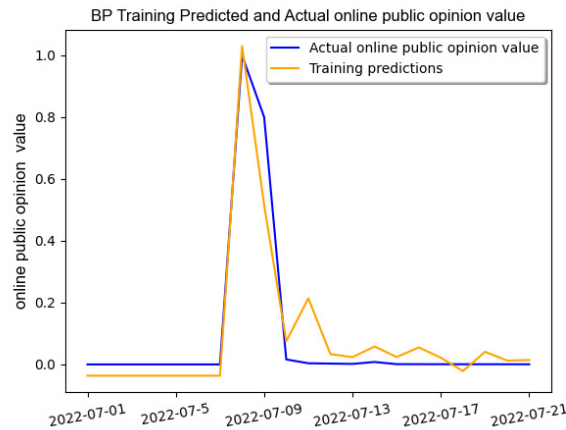


Figure 4. Predicted versus true values of BP neural network on the training set

As can be seen in Figure 4, the BP neural network model was able to fit the trend of the outbreak in general and was able to fit the outbreak point values effectively. Still, the fitted data differed more from the original data after the outbreak point.

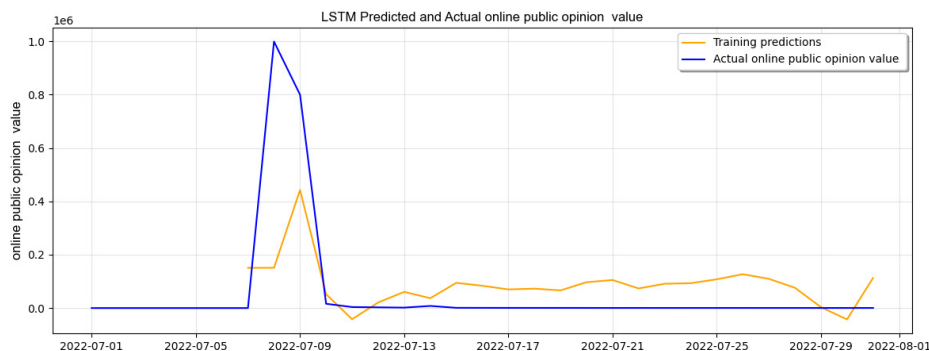


Figure 5. Predicted and actual network opinion values of LSTM neural network

As seen in Figure 5, the LSTM neural network model is basically able to fit the trend of the burst after the burst point, but is not able to fit the bursting point value effectively.

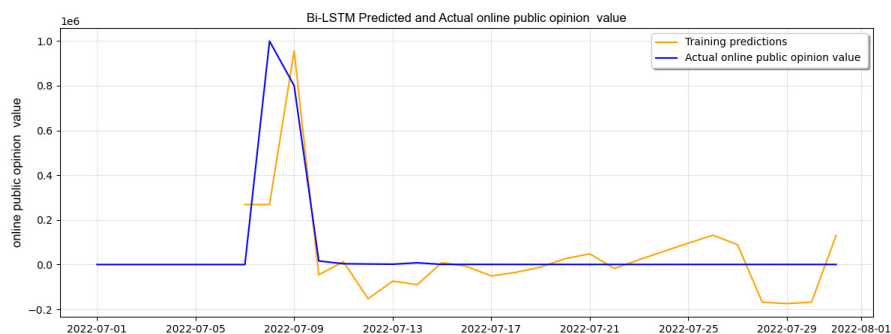


Figure 6. Predicted and actual network opinion values of Bi-LSTM neural network

As can be seen in Figure 6, the Bi-LSTM neural network model as a whole, due to the LSTM model, was able to fit the trend of the burst and was able to fit the bursting point value effectively, but after the burst point, the fit was poor.

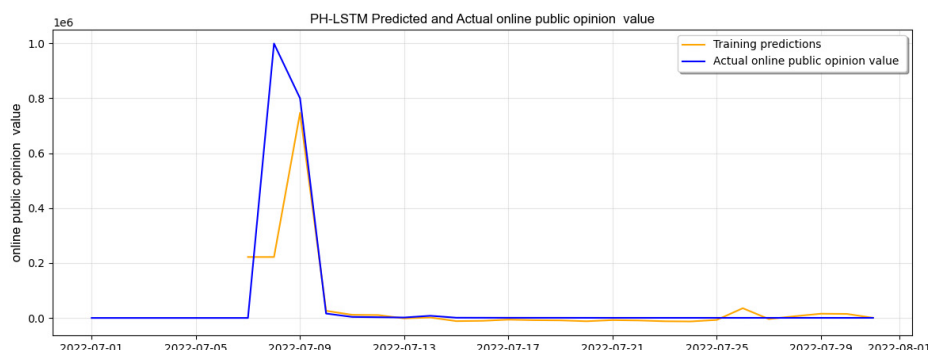


Figure 7. Predicted and actual network opinion values of PH-LSTM neural network

As can be seen in Figure 7, the PH-LSTM neural network model overall outperforms the other three models in being able to fit the trend of the bursts, and the fitted data are broadly consistent with the original data after the bursting point.

6. Conclusion

In this paper, we propose a new prediction method PH-LSTM that integrates social media influence and deep learning, which combines real-time social media network data and authoritative Baidu index and WeChat index for predicting the trend of public opinion on the Internet in emergencies, and the comparison test with the BP back propagation neural network model, LSTM model and Bi-LST model verifies the correctness and superiority of PH-LSTM model. The comparison test with BP back propagation neural network model, LSTM model and Bi-LST model verified the correctness and superiority of the PH-LSTM model. The prediction results help the government to control and guide public opinion information, which is conducive to harmonious and stable social development.

References

- [1] Li G, Chen JH. A review of research on online public opinion on sudden public events [J]. Knowledge of Library and Intelligence, 2014(2): 111-119.
- [2] Dong XF, et al. Small and multi-peak nonlinear time series forecasting using a hybrid back propagation neural network[J]. Information Sciences An International Journal, 2018.
- [3] Zhao JH , Zeng DL , Qin JT , et al. Simulation and modeling of microblog-based spread of public opinions on emergencies[J]. Neural Computing and Applications, 2021, 33(9).
- [4] Lv Z. Prediction of the Forwarding Volume of Campus Microblog Public Opinion Emergencies Using Neural Network. Mobile Information Systems. 2022;2022:3064266.
- [5] Li L, Wan Y, Plewczynski D, Zhi M. Simulation Model on Network Public Opinion Communication Model of Major Public Health Emergency and Management System Design. Scientific Programming. 2022;2022:5902445.
- [6] Zhao J , He H , Zhao X , et al. Modeling and simulation of microblog-based public health emergency-associated public opinion communication[J]. Information Processing & Management: Libraries and Information Retrieval Systems and Communication Networks: An International Journal, 2022 (2): 59.

- [7] Haneen Alabdulrazzaq;Mohammed N. Alenezi;Yasmeen Rawajfih;Bareeq A. Alghannam;Abeer A. Al-Hassan;Fawaz S. Al-Anzi.On the accuracy of ARIMA based prediction of COVID-19 spread . Results in Physics,2021
- [8] Lee, Yoonsuk, Brorsen, et al. Permanent shocks and forecasting with moving averages.[J]. Applied Economics, 2017.
- [9] Tratar LF , B Mojškerc, Toman A . Demand forecasting with four-parameter exponential smoothing [J]. International Journal of Production Economics, 2016.
- [10] Luo TY. Research on the development prediction of microblog public opinion hotspots based on Logistic model[J]. Statistics and Information Forum,2017,32(10):91-95.
- [11] Li Z. Forecast and Simulation of the Public Opinion on the Public Policy Based on the Markov Model. Complexity. 2021;2021:9936965.
- [12] Fang, YJ, Chen Xin, Son, et al. Modelling propagation of public opinions on microblogging big data using sentiment analysis and compartmental models[J]. International journal on Semantic Web and information systems, 2017, 13(1): 11-27.
- [13] He YX, Liu JB, Sun ST. A neural network-based method for predicting microblog public opinion[J]. Journal of South China University of Technology (Natural Science Edition),2016,44(09):47-52.
- [14] Cheng TJ, Wang M. Trend prediction of online public opinion on unexpected events based on variable weight combination[J]. Computer Science,2021,48(S1):190-195+202.
- [15] Sang CY, Xu W, Jia CL, Wen JH. Predicting the evolutionary trend of online opinion events based on attention mechanism in social networks[J]. Computer Science,2021,48(07):118-123.