Prediction of Hot Topics of Agricultural Public Opinion Based on Attention Mechanism LSTM Model

Lifang Fu, Northeast Agricultural University, China
Feifei Zhao, Northeast Agricultural University, China

ABSTRACT

In order to timely and accurately analyze the focus and appeal of public opinion on the internet, a LSTM-ATTN model was proposed to extract the hot topics and predict their changing trends based on tens of thousands of news and commentary messages. First, an improved LDA model was used to extract hot words and classify the hot topics. Aimed to more accurately describe the detailed characteristics and long-term trend of topic popularity, a prediction model is proposed based on attention mechanism long short-term memory (LSTM) network, which named LSTM-ATTN model. A large number of numerical experiments were carried out using the public opinion information of “African classical swine fever” event in China. According to the results of evaluation indices, the relative superiority of LSTM-ATTN model was demonstrated. It can capture and reflect the inherent characteristics and periodic fluctuations of the agricultural public opinion information. Also, it has higher convergence efficiency and prediction accuracy.

KEYWORDS

Agricultural Public Opinion, Attention Degree Prediction, Attention Mechanism, Hot-Topics Classification, Long Short-Term Memory (LSTM) Network, Public Opinion Prediction, Topic Model, Topic Popularity

INTRODUCTION

With the rapid development of social informatization, a variety of news, information and comment information have been rapidly and widely disseminated on the Internet. Among them, network public opinions involving agriculture and agricultural products are also frequent. Different from other types of public opinion, agricultural public opinion is often closely related to the safety of food and supplies in people’s daily life, which is more likely to cause widespread and sustained attention, and even lead to the agglomeration effect of negative emotions such as complaints and panic. This maybe has a negative impact on agricultural industry development and social stability. It is necessary to improve the monitoring ability of agricultural public opinion, quickly and accurately understand the hot topics of public opinion. By analyzing the hot topics and emotional attitudes of the people in public opinion, the government can guide public opinion effectively and prevent the spread of negative emotions and false public opinion.

The information of agricultural public opinion is complex and diverse, and the rapid spread of the Internet makes the information rise exponentially. How to quickly and accurately collect and sort out the massive information relevant agricultural public opinion events, dig out and focus on the hot
spots, analyze and predict its development trend is an important hot research issue, which has received continuous attention from researchers in various fields presently.

This paper aims to solve this problem by means of machine learning and deep learning techniques. Firstly, the text information related to public opinion is collected by web crawler. Then, LDA (Latent Dirichlet Allocation) topic model is used to analyze the keywords and their probability distribution in public opinion, and the different hot topics are extracted by word co-occurrence analysis. Finally, based on attention mechanism, a LSTM (Long-Short Term Memory) model is proposed to predict the popularity of different hot topics in the near future. This paper provides a technical method to study the law of changing and developing of public opinion.

In this paper, to address the massive agricultural public opinion text information, a method is proposed for keyword extracting and hot topics analysis. And the LSTM model is improved by combining the attention mechanism to accurately predict the heat change of various topics. The research work in this paper can improve the efficiency of dealing with massive information in public opinion analysis, help to discover public opinion topics timely and accurately, and scientifically predict the heat change of each topic, so as to improve the efficiency of public opinion supervision.

The structure and content of the paper are as follows: the second part discusses the relevant research literature of public opinion analysis in detail and comprehensively. In the third part, the methods and specific process of keywords and hot topics extraction of public opinion are given, and the heat changing prediction model LSTM-ATTN is proposed based on attention mechanism. The fourth part is the application and experimental study of the model based on the public opinion information about African swine fever events in China in the past two years. The second part discusses the relevant research literature of public opinion analysis in detail and comprehensively, and makes a detailed comparison and comment on the application of various deep learning methods and models in public opinion analysis. In the third part, the methods and specific steps of hot topic extraction and hot topic prediction of public opinion are given, and the hot topic prediction model LSTM-ATTN based on attention mechanism is proposed. The fourth part is the application and experimental study of the model based on the public opinion information about African swine fever events in China in the past two years. The fifth part obtains the research results of this paper through the comparative analysis and discussion of various evaluation indexes of the prediction models.

RELATED RESEARCH WORKS

In the related research of network public opinion, the discovery and analysis of public opinion hotspots (Claudia et al., 2019) have always been the important research topic focused by scholars at home and abroad.

The relevant research contents mainly include public opinion hot spot detection, topic extraction and classification, and public opinion heat analysis and prediction. The core research contents mainly include the discovery of public opinion hot words, topic extraction and classification, and prediction of public opinion popularity.

In the research of public opinion hotspots detection and topic classification, the commonly used methods in the past are word frequency analysis, improved clustering algorithm (Pohl D et al., 2012), and classification algorithm based on machine learning (Sriram et al., 2010). Nowadays, with the development of natural language technology, many related methods have been applied in the field of public opinion topic discovery. Commonly used methods are text clustering algorithms (Wang et al., 2020), neural networks combined with topic models (Xiao et al., 2019), and improved LDA topic models (Zheng et al., 2019). The above-mentioned researches have focused on mining and classifying different events from large amounts of text information. However, due to the particularity of agricultural public opinion, a certain public opinion event may contain many different topics which public concerned. Therefore, the above method is not suitable for the analysis of agricultural public opinion.
In the study of the heat change of public opinion hot topics, the main research problems include the measurement of hot topics, the evolution trend analysis and the prediction of topic heat. The research methods adopted mainly include three types. The first type is to use the previously determined index system to quantitatively analyze the heat of hot topics. The second kind is using neural network, Logistic equation, autoregressive integrated moving average (ARIMA) model to predict the heat of hot topics. The third type is to use the theories of system dynamics, complex networks and social networks to construct the evolution model of public opinion hot topics. Some scholars used Baidu index (Chen et al.,2017), media attention (Wang et al.,2008) and web page popularity (Zhanget al. (.,2017) as public opinion indicators to establish a popularity evaluation model. Some scholars also used regression models (Lian et al.,2018), machine learning (Yang et al.,2019) and neural networks (Sun et al.,2018) to predict the popularity of public opinion.

In the research of prediction methods, as its unique advantages, the attention mechanism attracts more and more scholars ' attention. The attention mechanism mimics the human attention mechanism, which can select the information that is more critical to the current task target from a large amount of information. At present, it is widely used in deep learning fields such as natural language processing (Rush et al.,2015), speech recognition (Graves et al., 2013) and image recognition (Mnih et al.,2014). Minh et al. (2014) introduced the concept of visual attention combined with the recurrent neural network model to save a lot of computational cost. Wang et al. (2016) proposed a neural network model that uses an attention mechanism to significantly improve the relationship extraction effect. On the prediction problem of time series, Qin et al.(2017) proposed a two-stage attention loop neural network model (DA-RNN) to solve the problem of target time series prediction with multiple exogenous sequences. Du et al. (2018) proposed an attention mechanism-based LSTM-CNN network to extract and classify the latent patterns between time series.

In the research field of prediction methods, attention mechanism has also been successfully applied to the model such as LSTM, which improves the prediction accuracy. Du et al. (2018) used an attention-based LSTM model to automatically capture long-term dependencies between sequences. Ye et al. (2021) introduced the attention mechanism into the double-layer residual network to improve the signal prediction accuracy. According to the characteristics of short-time power load prediction, Gong (2021) introduced an attention mechanism to calculate the weight of hidden layers to improve the prediction accuracy. Sang et al. (2021) used the attention mechanism to mine the dependencies between users to predict the real information transmission process. However, presently the application of such deep learning methods and models in the prediction of hot topics of public opinion is still rare.

Based on the above analysis of the current research literature, this paper proposes a prediction model LSTM-ATTN based on attention mechanism to extract hot topics and predict heat change of agricultural public opinion. Firstly, a LDA model is constructed to find out hot word and key topics of agricultural public opinion, and then seven different hot topics with high attention are extracted. Aiming at the temporal data variation characteristics of topics popularity, LSTM model with attention mechanism is constructed to predict short-term changing of topics heat. The predicting model can capture the local change characteristics and internal rules of public opinion information more effectively, improve the reliability and accuracy of hot topic prediction.

ANALYZING OF AGRICULTURAL OPINION BASED ON LSTM-ATTN MODEL

The Process of Predicting the Popularity of Hot Topics

The main purpose of this article is to predict hot topics popularity of agricultural public opinion. The specific process is shown in Figure 1 bellow. The improved LDA model is used to extract hot words and analyze hot topics of agricultural public opinion information, divide reasonable topic categories and calculate the popularity of each topic. In the heat prediction stage, the LSTM model with the attention mechanism is used to make the final decision of the model more concentrated, and
to assign the main decision weight to the historical data that really helps the final prediction. Thereby improving the accuracy of model prediction.

Figure 1. Flow chart of agricultural public opinion prediction

Keywords and Hot Topics Extracting Based on Improved LDA Model

LDA model is a three-layer Bayesian model proposed by Blei (Bleiet al., 2003). The model can solve the polysemy problem when mining the implicit semantics of documents. At present, LDA is one of the most widely used topic models. The core idea of LDA is to introduce an intermediate variable ‘topic’ between ‘document’ and ‘word’. Theme is composed of some keywords. A document can contain multiple different topics. The probability distribution between documents and topics, topics and words can be described by multiple distributions.

Algorithm Flow of LDA Model

The algorithm flow of LDA model is shown in Figure 2. Important parameters in this model include the following: $\alpha$ is the Dirichlet prior probability parameter representing the multinomial distribution of topics within each public opinion document. $\beta$ is the Dirichlet prior probability parameter representing the multinomial distribution of keywords within a topic. Parameter $\theta_m$ represents the topics distribution of public opinion documents. $\phi_k$ represents the word distribution of the $k$-th topic. Parameter $z_{m,n}$ is the $n$-th topic and $w_{m,n}$ is the $n$-th word in public opinion document. The process of public opinion analysis using LDA model can be summarized as Algorithm 1.
Algorithm 1. Topics extraction and classification of public opinion information

Input: The Dirichlet prior probability parameters $\alpha$ and $\beta$. Word Vectors of Public Opinion Text

Output: All keywords and their probability distributions $\phi_k$, all the hot topics and their probability distributions $\theta_m$

For each iteration process do

Randomly assigns a topic to each word with a prior probability distribution $\alpha$.

Calculate the probability that the $i$-th word belongs to the $k$-th topic by using Gibbs Sampling, where

$$\theta_{mk} = \frac{n_{m,i}^{(k)} + \alpha_k}{\sum_{k=1}^{K} (n_{m,i}^{(k)} + \alpha_k)}$$

$$\phi_{kt} = \frac{n_{t,i}^{(k)} + \beta_t}{\sum_{t=1}^{V} (n_{t,i}^{(k)} + \beta_t)}$$

If not the last word, then calculate the posterior probability of the word belongs to the $k$-th topic

$$P(z_i = k | Z_{-i}, W) \propto \frac{n_{m,i}^{(k)} + \alpha_k}{\sum_{k=1}^{K} (n_{m,i}^{(k)} + \alpha_k)}$$

$$\frac{n_{t,i}^{(k)} + \beta_t}{\sum_{t=1}^{V} (n_{t,i}^{(k)} + \beta_t)}$$

When the iteration termination condition is satisfied, output the vectors $\phi_k$ of hot words and probability distributions

Output the vectors $\theta_m$ of hot topics in the public opinion

End

Popularity of Public Opinion Hot Topics

For the vectors $\theta_m$ of hot topics extracted in Algorithm 1, calculate its popularity according to the degree of attention of each topic. With the pretreated text information of public opinion, the number of documents for each hot topic in a given time period is counted. When the number of documents is more, the hot topic is more popular. The calculation formula is:
\[
\theta_z^d = \frac{z^d}{D_t}
\]

where \( \theta_z^d \) represents the probability that the \( d \)-th document belongs to the \( z \)-th topic. \( D_t \) is the number of documents in the time period.

**The Popularity Prediction of Hot Topics Based on LSTM-ATTN Mode**

**The Neuron Structure of LSTM Model**

LSTM (Long Short Term Memory Network) neural network is proposed to solve the problem of gradient disappearance (Hochreiter & Schmidhuber, 1997). By improving the RNN neural network, the long-term dependence problem can be better solved (Ah C.T., 1998). Compared with RNN, LSTM can mine more data features in longer time series data. The structure of this model is shown as figure 3, the core of LSTM is cell state, which is updated by three gates of neurons, named output gate, forget gate and input gate. LSTM model control the transmission state by gated state, realize selective memory of information, retain important information, and forget not important information.

Figure 3. LSTM neuron structure
The Prediction Model LSTM-ATTN Based on Attention Mechanism

Attention mechanism was first used in machine translation system by Bahdana (Bahdana et al. 2014), and was widely used in the research field of natural language processing lately. Attention mechanism can select important information and give higher weight to important information, so as to give more attention to key information and neglect non-important information. In this paper, it is necessary to predict the future heat value according to the historical data of topic popularity. Attention layer is added to the structure of LSTM model, so that the process of data feature extraction is more focused on the key data with large volatility. The improved model structure is shown in Figure 4.

In the encoding step, LSTM unit combined with attention mechanism is used to encode the time series data of the heat for given hot topics, from which the temporal dependence of the predicted value and the historical data is analyzed. In the decoding part, the attention mechanism is introduced to adaptively select the output value of the predicted topic heat, so that these predicted values can reflect the trend and periodicity of the future topic heat time series.

Algorithm Flow of LSTM-ATTN Model

The prediction model based on the attention mechanism is defined as LSTM-ATTN model. The processing steps of the model are shown in Algorithm 2.
EXPERIMENTAL RESULTS AND DISCUSSION

Since 2019, there have been several outbreaks of African swine fever in various parts of China, causing sustained attention and discussion among a large number of people and forming influential public opinion on various Internet. This paper uses python web crawler technology to obtain news reports and microblog texts related to African swine fever epidemic during June 1, 2019 and October 31, 2019. A total of 25169 documents was obtained by crawler technology. The information was filtered and preprocessed, and 49 documents with missing data were deleted, with 25120 public opinion documents as the analysis data set.

Algorithm 2. Prediction of hot topics popularity

| Input: Standardized hot topics heat time series data $\theta^t_z$ |
| Output: The predicted hot topics heat $\hat{\theta}^t_z$ of future period |

Step 1: Normalize topics heat time series data by using MinMaxScaler method, divide training set and test set and put into layer 1.

Step 2: Encode the internal input and calculate the hidden layer state values

$\hat{h} = \{h_1, h_2, \ldots, h_t\}$ for each time node

Step 3: calculate the attention weight $\hat{l}^i_t$ for each time node by using attention mechanism in the hidden layer, where:

$\hat{l}^i_t = \frac{v_d^i \text{tanh} \left( W_d \left[d_{t-1}^i; s_{t-1}^i \right] + U_d h_t \right)}{\sum_{i=1}^{T} \beta^i_t h_t}$

Step 4: Normalized the attention weight $\hat{l}^i_t$ by using softmax function, calculate the context vector by weighted summation of all moments, where:

$\beta^i_t = \frac{\exp \left( \hat{l}^i_t \right)}{\sum_{i=1}^{T} \exp \left( \hat{l}^i_t \right)} \quad c_t = \sum_{i=1}^{T} \beta^i_t h_t$

Step 5: Update the state value of the decoder hidden layer $d_t = f \left( \theta_{t-1}^i, c_{t-1}, d_{t-1} \right)$

Step 6: Predict the popularity of hot topics by full connection regression $\hat{\theta} = v^\theta_d \left[ d_t^i; c_t^i \right] + b^\theta_d + b^\theta$

End

Topic Classification and Heat Calculation of Agricultural Public Opinion

Data Processing and Text Analysis

Chinese language has certain particularity, each word in the sentence does not indecently. It is necessary to use other tools to segment Chinese texts. It is necessary to use other tools to segment Chinese texts. There are many methods for Chinese word segmentation. Among them, the jieba word segmentation library is a relatively common method at present. The jieba word segmentation has its own dict.txt dictionary. The analysis results include the number of entries and part of speech. At the same time, it also supports adding a custom dictionary and defining the word frequency. Based on this, this paper uses python jieba to segment public opinion documents in Chinese.

When analyzing the words after word segmentation, only the words with actual meanings need to be saved. In natural language processing, the words that appear frequently in Chinese but have no
actual meaning are generally eliminated. In the experiment of this article, not only should these basic stop words be eliminated, some meaningful and frequently occurring words will still interfere with the topic division, so these words need to be manually added to the stop word list. The experimental data in this article uses the Internet public opinion of African swine fever, so every public opinion document has words such as “Africa” and “swine fever”. These frequently appearing words will make each topic contain these words and affect the topic division effect, these words need to be added to the stop word list. Here, this article uses the stop vocabulary of Harbin Institute of Technology to filter meaningless words in the segmented text. The sk-learn vectorization tool Counter Vectorizer is used to vectorize the training corpus for the text after word segmentation and stop word removal.

**Topic Extraction and Classification**

It is necessary to determine the number of topics before using the LDA model to extract hot topics. According to the relevant research work, the perplexity is used to determine the number of topics. The perplexity represents the uncertainty of the subject of the document, so the lower the perplexity, the better the topic generalization. Set the number of topics from 1 to 14 to extract topics from the public opinion information document and calculate the corresponding perplexity results as shown in Figure 5 below. Finally, the number of topics with the lowest perplexity is determined to be 7 based on the calculation results.

![Figure 5. Perplexity of the number of topics from 1 to 14](image)

After determining the number of topics, use the method in the python-based machine learning library sk-learn to induct topics form the preprocessed data. Since the time information of each document is needed for subsequent analysis of topic popularity, the time information of the document is added to the LDA modeling. The number of topics is set to 7, and the hyperparameters are determined by the model default values. After word segmentation to remove stop words, a total of 58085 independent words were obtained for model training. After the model training is completed, the document-topic vector with document time information is saved for subsequent analysis. Summarize seven topics according to the eight most probable words under each topic.

As shown in Topic 1, the three most weighted words are “vaccine”, “virus”, and “research and development”. It can be inferred that this topic discusses topics related to vaccine development, so the topic is summarized as “vaccine research and development”. In topic 4, based on the subject terms’ “pork”, “price”, “live pig”, and “increasing”, it can be inferred that this topic discusses the topic of pork prices, so the topic is summarized as “pork price increase”. According to this method, the seven topics are summarized as shown in Table 1.
The time information of the document was entered when using LDA modeling for topic discovery. Therefore, based on the document-topic probability matrix with time information obtained based on the LDA training results, the popularity trends of different topics can be obtained. The document-topic vector obtained is divided into time windows in units of days, and the topic intensity values of different topics in each day are calculated as topic popularity as shown in Table 2. And the daily popularity of each topic is used as the input data of LSTM-ATTN’s popularity prediction.

### Public Opinion Popularity Prediction Based on LSTM-ATTN

#### Topic Hot Time Series Acquisition and Preprocessing

The experimental topic popularity time series data comes from the previous part of the LDA topic model. Seven topics and documents with time information are obtained through topic model training—the topic probability distribution is divided into time windows in days, and the daily popularity of each topic is calculated to obtain Table 2, which is used as the input of LSTM-ATTN heat prediction. Due to space limitations, topic 1 is selected to discuss the LSTM-ATTN modeling process, and the remaining topics only show the model results. After obtaining the topic popularity time series with day as the time window through the LDA part, it is convenient for data processing, and the data is normalized. Data normalization can not only improve the convergence speed of the model, but also improve the accuracy of the model. In this paper, the method of MinMaxScaler function is used.
to scale the data to between [0,1]. The normalized data is divided into a training set and a test set. Here, the first 0.75 data is used as the training set, and the last 0.25 data is used as the test set. That is, the data from June 1 to September 23 is used as the training set, and the data from September 24 to October 31 is used as the test set.

**Time Series Forecast of Topic Popularity**

In the topic prediction part, the LSTM model that introduces the attention mechanism is used to predict the topic popularity time series. The model is implemented based on the TensorFlow deep learning framework. The loss function used is the mean square error (MSE), and the optimizer uses adaptive moment estimation (Adam). The values of the hyperparameters obtained through a series of experiments are as follows: the batch size is set to 1, and the number of hidden layer units in the LSTM layer is set to 64. In order to compare the performance of the LSTM-ATTN model, the commonly used time series prediction models (Ah C.T., 1998), and Long Short-Term Memory Network (LSTM) are used as the comparison models.

The comparison between the predicted values and actual values of different models is shown in Figure 6. The black line is the true value of the topic’s popularity. The predicted values of the LSTM-ATTN model, LSTM model, RNN model, and ANN model correspond to the green, purple, red, and blue lines respectively. It can be seen from the curve between the predicted value of the model and the actual value that the LSTM-ATTN predicted straight line is closer to the true value.

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic1</th>
<th>Topic2</th>
<th>Topic3</th>
<th>Topic4</th>
<th>Topic5</th>
<th>Topic6</th>
<th>Topic7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun. 1st</td>
<td>0.09</td>
<td>0.18</td>
<td>0.05</td>
<td>0.10</td>
<td>0.25</td>
<td>0.26</td>
<td>0.08</td>
</tr>
<tr>
<td>Jun. 2nd</td>
<td>0.16</td>
<td>0.07</td>
<td>0.02</td>
<td>0.10</td>
<td>0.30</td>
<td>0.25</td>
<td>0.09</td>
</tr>
<tr>
<td>Jun. 3rd</td>
<td>0.14</td>
<td>0.25</td>
<td>0.03</td>
<td>0.19</td>
<td>0.17</td>
<td>0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>Jun. 4th</td>
<td>0.05</td>
<td>0.11</td>
<td>0.05</td>
<td>0.13</td>
<td>0.46</td>
<td>0.16</td>
<td>0.04</td>
</tr>
<tr>
<td>Jun. 5th</td>
<td>0.07</td>
<td>0.37</td>
<td>0.03</td>
<td>0.15</td>
<td>0.21</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Jun. 6th</td>
<td>0.08</td>
<td>0.24</td>
<td>0.06</td>
<td>0.23</td>
<td>0.22</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>Jun. 7th</td>
<td>0.04</td>
<td>0.12</td>
<td>0.03</td>
<td>0.24</td>
<td>0.38</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>Jun. 8th</td>
<td>0.08</td>
<td>0.17</td>
<td>0.04</td>
<td>0.15</td>
<td>0.21</td>
<td>0.07</td>
<td>0.27</td>
</tr>
<tr>
<td>Jun. 9th</td>
<td>0.10</td>
<td>0.12</td>
<td>0.06</td>
<td>0.16</td>
<td>0.15</td>
<td>0.30</td>
<td>0.07</td>
</tr>
<tr>
<td>Jun. 10th</td>
<td>0.10</td>
<td>0.31</td>
<td>0.05</td>
<td>0.13</td>
<td>0.18</td>
<td>0.19</td>
<td>0.04</td>
</tr>
<tr>
<td>Jun. 11th</td>
<td>0.28</td>
<td>0.13</td>
<td>0.04</td>
<td>0.17</td>
<td>0.17</td>
<td>0.16</td>
<td>0.05</td>
</tr>
<tr>
<td>Jun. 12th</td>
<td>0.30</td>
<td>0.05</td>
<td>0.02</td>
<td>0.17</td>
<td>0.15</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>Jun. 13th</td>
<td>0.38</td>
<td>0.09</td>
<td>0.14</td>
<td>0.07</td>
<td>0.13</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td>Jun. 14th</td>
<td>0.38</td>
<td>0.08</td>
<td>0.03</td>
<td>0.26</td>
<td>0.15</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>Jun. 15th</td>
<td>0.38</td>
<td>0.09</td>
<td>0.03</td>
<td>0.24</td>
<td>0.14</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>Jun. 16th</td>
<td>0.24</td>
<td>0.13</td>
<td>0.03</td>
<td>0.21</td>
<td>0.23</td>
<td>0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>Jun. 17th</td>
<td>0.43</td>
<td>0.14</td>
<td>0.02</td>
<td>0.06</td>
<td>0.05</td>
<td>0.14</td>
<td>0.16</td>
</tr>
<tr>
<td>Jun. 18th</td>
<td>0.34</td>
<td>0.14</td>
<td>0.05</td>
<td>0.10</td>
<td>0.18</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>Jun. 19th</td>
<td>0.27</td>
<td>0.24</td>
<td>0.05</td>
<td>0.12</td>
<td>0.09</td>
<td>0.10</td>
<td>0.13</td>
</tr>
<tr>
<td>Jun. 20th</td>
<td>0.12</td>
<td>0.36</td>
<td>0.14</td>
<td>0.14</td>
<td>0.08</td>
<td>0.04</td>
<td>0.11</td>
</tr>
</tbody>
</table>
The root means square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), $R^2$ are used as evaluation indicators to measure the performance of model prediction. The evaluation indicators for calculating the predicted values of different models are shown in Table 3. As shown in the table, the minimum values of RMSE, MAE and MAPE are 0.024, 0.017, and 11%, respectively. Through the comparison of the three indicators, LSTM-ATTN has the smallest error and therefore the highest prediction accuracy. As to the indicator $R^2$, the LSTM-ATTN model reach the highest value of 91.98%. Therefore, the LSTM-ATTN model captures the inherent change characteristics and periodic fluctuations of public opinion time series data effectively, and give accurately prediction of the hot topics’ popularity.

In order to evaluate the versatility of the LSTM-ATTN model that introduces the attention mechanism, the other six topic popularity time series are respectively predicted and related indicators are calculated as shown in Table 4.

As shown in Table 4, the RMSE of the predicted value of the model LSTM-ATTN on the six topics is the smallest, which are 0.032, 0.036, 0.048, 0.054, 0.074, 0.044. On the six topic popularity sequence data, the MAE and MAPE values of LSTM-ATTN are also the smallest among the four models. Therefore, the LSTM-ATTN model proposed in this paper has good versatility in predicting
topic popularity, and can accurately capture the changing characteristics and internal laws of topic popularity sequence, and improve the reliability and accuracy of public opinion topic popularity prediction.

**CONCLUSION**

The outbreak of African swine fever across China in 2019 has generated widespread public opinion on the Internet. In view of this epidemic event, this paper carried out the analysis and research of agricultural public opinion. Using web crawler technology, a large number of news and microblog information related to the event are collected. Using deep learning model and technology, the hot topics of the public opinion events were extracted and the changing trend of hot topics are analyzed in depth. Firstly, the improved LDA model is used to analyze and extract keywords in a large number of text information. Through word co-occurrence analysis and frequency statistics, seven different hot topics are summarized, and the probability distribution of hot topics is given. A long-term and short-term memory network prediction model LSTM-ATTN was proposed based on attention mechanism,
by which the popularity of the seven hot topics and the changing trend were calculated. At the last of the paper, a large number of comparative experiments are carried out. The experimental result results show that the model LSTM-ATTN can quickly and accurately capture the key features and periodic laws in the time series data of topic heat. The algorithm has high convergence rate and gives more accurate prediction results.

In the case study about African classical swine fever” event in China. Seven relatively independent hot topics of opinion were extracted and classified, including “epidemic notification”, “pork safety” and “pork price rise” etc. These hot topics reflects the focus and hot issues concerned by the public in the event. In order to describe and predict the changing trend of topics popularity, the model named LSTM-ATTN was proposed by instructing attention mechanism into the LSTM neural structure. the paper proposed a new LSTM network by introduction which model.

According to the results of a large number of numerical experiments, the prediction accuracy and performance of several models were compared and discussed. Using the evaluation indexes, such as RMSE, MAE and MAPE, the results illustrated for the most of hot topics, the LSTM-ATTN model give relatively accurate prediction with lower RMSE, MAE and MAPE. This means LSTM-ATTN model can more effectively capture the inherent change characteristics and periodic fluctuations law of the time series data of agricultural public opinion. On the other hand, it has higher converging efficiency and prediction accuracy. The paper provides an effective model for analyzing of online public opinion, which is helpful to explore and grasp the focus of public concern about some food safety events.

However, as the topics popularity of public opinion was greatly affected by external news, it often fluctuates violently at particular times. The prediction model proposed in the paper cannot describe and predict such local violent fluctuations perfectly. In future research work, more efforts should be done to improve the predicting accuracy and stability of the model.

ACKNOWLEDGMENT

This research was funded by the project of Philosophy and Social Science Research in Heilongjiang Province [grant number 20SEB080].
REFERENCES


Fu Lifang, an associate professor, received her doctorate from Harbin Institute of Technology and is currently working at Northeast Agricultural University. Her main research interests are agricultural public opinion and big data research in the field of agriculture.

Zhao Feifei, a postgraduate student, is currently studying at Northeast Agricultural University, whose main research direction is data mining in the field of agriculture.


