Application of neural network based on long and short term memory in rumor detection

Zhiyan Wang[*1], Mingzhao Li[*2]

*1School of Artificial Intelligence, Jilin International Studies University, Changchun 130000, China
E-mail:zhiyanwang@163.com

Abstract. In this paper, we optimize the variable parameters in the long-term and short-term memory neural network through the slap swarm algorithm, so that it can more accurately adapt to the actual needs of rumor detection data. According to an input text segment, the short-term memory neural network can processes the Chinese word segmentation, and then convert it into binary number form by hot coding method, so as to map word vector. In the experiment, we design the model carefully, train it repeatedly in the word vector data set, and optimize the parameters of the model through the bottle ascidian optimization algorithm. The practice shows that it has strong applicability to rumor recognition.

Keywords: Rumor detection,Intelligent optimization algorithm, Salp swarm algorithm

1. INTRODUCTION (HEADING 1)

Rumor detection is one aspect of natural language application. Natural language can automatically correct and fill in search keywords. Through natural language processing, language translation can be carried out between different languages. In real life, speech recognition can be used to recognize speech into words quickly. In addition, natural language processing also plays a key role in social public opinion monitoring. Through the processing and application of natural language, we can have a good dialogue with chat robot, so as to achieve the application of human-computer interaction. In addition, through the file retrieval and web crawler function, we can realize the company information query and statistics.

Pathak et al. Conducted a series of discussions and studies on rumor detection and other related issues[1]. Luo et al. Proposed a new post based enhanced representation method to deal with rumor events. The accuracy reached 80.09%[2]. Xu et al. Developed a new topic driven rumor detection framework to detect whether microblog posts are rumors[3]. Mukiri et al. Briefly reviewed different rumor source identification technologies[4]. Wang et al. Proposed a rumor event detection algorithm based on SD and tsdts, and the result is better than the latest rumor event detection algorithm[5]. Agarwal et al. Established two new sequential game models to detect rumors[6]. Alzanin et al. Summarized the research on rumor detection in social media based on hybrid machine learning[7]. Jiang et al. Proposed a new rumor propagation debunking (RSD) model to explore the interaction mechanism between rumor propagation and debunking process by using ordinary differential equation (ODE) system[8]. Bodaghi reveals two different communication modes through the macro analysis of the spread process of rumors[9]. Aker et al. Proposed a neural network model based on internal attention to detect rumors, which is highly competitive compared with the baseline on the newly released rummareval2019 dataset[10].

2. RELATED KNOWLEDGE

2.1. Salp swarm algorithm

The Salps belongs to the neobranchiaceae, and its body is a barrel like transparent organism. It generates power through the process of sucking in and discharging seawater, which makes the body swim forward. This mechanism helps them to move quickly and hunt together in the ocean. Inspired by this kind of swarm intelligence behavior, mirjalili et al. proposed SSA based on stochastic population optimization algorithm in 2017.

Compared with the grid search method, the optimization algorithm has better exploration and development ability. It can obtain the approximate optimal solution of the problem by simulating the navigation ability and foraging behavior of marine organisms in nature, so it has strong optimization ability and high optimization accuracy. However, the algorithm will fall into the possibility of local extremum when dealing with complex problems.

The theoretical basis of SSA algorithm is the bottle ascidian chain, which divides the bottle ascidian individuals into two categories: the only leader and some followers. The leader leads the followers to move in the n-dimensional space to find the food source (i.e. the optimal solution). N represents the total number of variables contained in the solution of a given problem; The positions of leaders and followers are represented by two-dimensional matrix X. Leaders update their positions according to formula (1)

$$x_j^1 = \begin{cases} F_j + c_1(c_2(ub_j - lb_j) + lb_j), c_3 \geq 0.5 \\ F_j - c_1(c_2(ub_j - lb_j) + lb_j), c_3 < 0.5 \end{cases} \quad (1)$$

Among them, x_j^1 represents the position of the bottle ascidian (leader) numbered 1 in the j-dimensional space; F_j is the position of food source in the j-dimensional space; ub_j And lb_j denote the upper and lower bounds of the j-th position value respectively; c_2 and c_3 is a random number evenly distributed in [0,1], which is used to adjust the change trend of leader's position; c_1 is the most important parameter of SSA algorithm, which represents the parameter to measure exploration and development in search space.Its definition is shown in equation (2).

$$c_1 = 2e^{-(\frac{4l}{L})^m} \quad (2)$$

Where l is the current number of iterations and l is the maximum number of iterations. After the leader's position is updated, the followers update their positions according to equation (3): the followers update their positions based on Newton's law of motion. The followers update their positions according to Newton's motion theorem:

$$x_j^i = \frac{1}{2}at^2 + v_0 t \quad (3)$$

When i≥ 2, it means the position of the ith follower in the j-th dimension. t is the time and v_0 is the initial velocity,

$$a = \Delta v / t$$
$$\Delta v = v' - v_0$$
$$v' = \frac{x_j^i - x_j^{i-1}}{t} \quad (4)$$

Because the optimization time t is iterative, the dispersion between iterations is equal to 1, so t = 1, set the initial speed v_0 = 0, then

$$a = x_j^i - x_j^{i-1} \quad (5)$$

The location update of followers can be expressed as:

$$x_j^i = \frac{1}{2}(x_j^i + x_j^{i-1}) \quad (6)$$

Among them, NP is the population size, 2 ≤ i≤ NP; x_j^i Denotes the position of followers numbered I in the j-th dimensional space. Because the larger the number of followers, the worse the corresponding solution, so, x_j^2, x_j^3..., x_j^i represents the position of the first to I-1 followers in the j-dimensional space.

In this paper, the parameters of the long-term and short-term memory model are optimized by the SSA algorithm, so that the long-term and short-term memory model has a higher interpretability to the data. The long-term and short-term memory network is a kind of cyclic neural network, which can effectively solve the problem of long-term dependence in cyclic neural network. The long-term and short-term memory network can more effectively solve and process sequential time events with long time interval or delay time. In natural language processing, long-term and short-term memory networks also have good recognition effect. Therefore, by optimizing the parameters of the long-term and short-term memory network, its performance is better, so as to form a high-quality algorithm model.

2.2. Natural Language Processing

The one-hot coding： According to the word dimension generated by the construction dictionary, we use the hot coding method to initialize the sentence coding construction. To transform sentences into data dictionary expressions, the principle of construction is to transform every sentence in the language into a data matrix. The number of rows in the matrix is the length of the construction dictionary, and the number of columns in the matrix is the number of words in the sentence. When the words in the dictionary match the words in the sentence, the corresponding position in the dictionary vector will be assigned to one, otherwise it will be assigned to zero.

Language Dictionary: language can be defined as the lexical expression of communication between people, through which their thoughts and actions are expressed. With the development and iteration of human society, this kind of lexical expression gradually evolves and advances, most of which keep their original appearance, and a few are abandoned and pruned, Therefore, in order to enable more people to quickly understand and master the language, and carry out barrier free communication, the different languages in the expression of words are integrated into a whole, and maintained and managed. When we come across a new word that we can't understand in the process of learning, our habit is to consult the language dictionary to understand and digest it. Once we really digest and understand the word, we will naturally put it into our own language dictionary for further communication. Based on this theory, a language dictionary is formed.

N-element model construction method: N-element model is the process of constructing the nearest two words or words together to express semantics, where n is the number of word units in the combination.

Word embedding method: the language model matrix generated by the exclusive hot coding method, because of its large dictionary length in reality, causes its matrix expression dimension to be high, which is not conducive to the construction of language model and the extraction of dimension features. In this case, some people put forward the word embedding method to deal with it, which is a process of mapping the high-dimensional data in the dimension table to the low dimensional space without destroying its semantics and structure. The word mapping process of word embedding method is not meaningless and random conversion process, but the process of mapping words into the feature space according to a certain feature angle.

Error minimization: for the error problem in the algorithm, we use gradient descent method to reduce the minimum error of the experiment. The gradient descent method can ensure that the error will disappear exponentially with the length of time.

Word vector representation: in the process of natural language processing, it is an important step to transform a word or word into a word vector according to a certain rule form. Word vector representation is the form of data string according to word emotion and part of speech. Word vector can not only express meaning, but also express the similarity between words. In real life, word vector has been verified by many ways, which greatly improves the performance of natural language processing, and has a high application in grammar analysis and emotion analysis.

Word vector similarity: in the process of word vector representation, the similarity can be calculated by the cosine between vectors. Obviously, if the meaning of two words is the same or similar. The higher the cosine of the two vectors. The data generated by word vector representation belongs to multidimensional array. In the experiment, in order to make the data express its semantics more efficiently and perfectly, the data can be dimensionally reduced. For the dimensionally reduced data, we can show it through visualdl tool. In the display diagram, we find that words with similar semantics have higher density in the word vector space.

Semantic similarity: in the process of actual semantic analysis, we need to calculate not only the similarity between words, but also the similarity between sentences. Through semantic similarity calculation, we can express the semantics of sentences well. There is a method to express semantics, which is called bag of words. When using bag of words to Express semantics, there are three main steps, First of all, we need to do word segmentation for the sentence. The words that have been segmented are searched and matched through the dictionary, and then they are transformed into word vectors. Secondly, according to the bagging principle of the bag of words model, the word vectors are loaded into the bag of words by superposition to form a sentence vector. Finally, the cosine value between the two sentence vectors is calculated by cosine calculation, so as to judge the similarity of the two sentences. In this process, we use bowencoder to transform the bag of words.

Text semantic similarity calculation is also an important step in natural language processing. Among them, the text semantic matching problem has great practicability in real life. For example, in machine interaction, the robot will quickly match the questioner's question with the known question database, so as to find the most matching question, and then answer according to the answer of the question. Practice has proved that this is a very feasible scheme.

3. ALGORITHM DESIGN

3.1. Named entity recognition

named entity recognition is a key step in the process of natural language processing. By processing and identifying named entities, we can easily extract important information from the article or text description questions, and can also facilitate machine interaction and self-help response. Moreover, through the identification of named entities, it can perform syntactic analysis or machinetranslation function well. The recognition of named entities is the pre-processing stage of natural language processing data, and its recognition effect directly affects the effect of various downstream applications. Name the entity task. There are two common solutions, one is the long-term memory learning network or the combination of Gru and conditional random field. The underlying text information is extracted by the cyclic neural network model. Then, the conditional random field model is used to carry out organic training, to discover and identify the relationship between the bottom layers. The other is to train, test and verify the pre training model directly, so as to label the entity directly.

3.2. Algorithm optimization process

The optimization process of SSA algorithm is as follows:

Algorithm2.1: **SSA**

N: Population size of SALP;
D: The dimension of the problem;
lb: Lower boundary;
ub: Upper boundary;
The population is initialized randomly in the upper and lower bounds $X=(X_1, X_2, \ldots, X_N)$;
The initial population x is evaluated and the fitness is obtained *fitness*;
The optimal solution is obtained from the initial population FoodPosition And its fitness FoodFitness ;
while *Termination conditions not met* ***do***
  Update with Eq. (2) c_1
  for $i=1$ *to* N
    if $i <= N/2$
      Use Eq. (1) to update the candidate's location V_i;
    else
      Use Eq. (6) to update the candidate's location V_i;
    end if
    Based on the upper and lower bounds of variables, V_i is modified;
    if $f(V_i) < fitness_i$
      $X_i = V_i$;
      $fitness_i = f(V_i)$;
      $trial(i) = 0$;
    else
      $trial(i) = trial(i) + 1$;
    end if
  end for
  Update the best solution *FoodPosition* ***and its fitness*** *FoodFitness;*
end while

Algorithm design: we integrate SSA algorithm into LSTM model (short-term and long-term memory network) to optimize LSTM model parameters to improve its algorithm performance. The framework consists of four important stages: the first stage is to standardize the data samples into [- 1,1]; The second stage is to determine the key parameters of LSTM model performance by executing SSA optimization algorithm, which is used to determine the performance of LSTM model; The third step is to estimate the parameter value with LSTM model based on 10 fold test set, and then obtain the optimal parameter pair through F; Finally, the optimal parameters are used to initialize the LSTM model, and the test data are predicted based on 10x cross validation. In the process of obtaining the best parameters, the fitness is the average classification accuracy avgac obtained by LSTM through 10 fold cross validation analysis, which can be calculated by Eq. (7):

$$fitness = avgACC = \frac{\sum_{i=1}^{10} testACC_1}{10} \qquad (7)$$

4. EXPERIMENTAL DESIGN

4.1. Named entity recognition

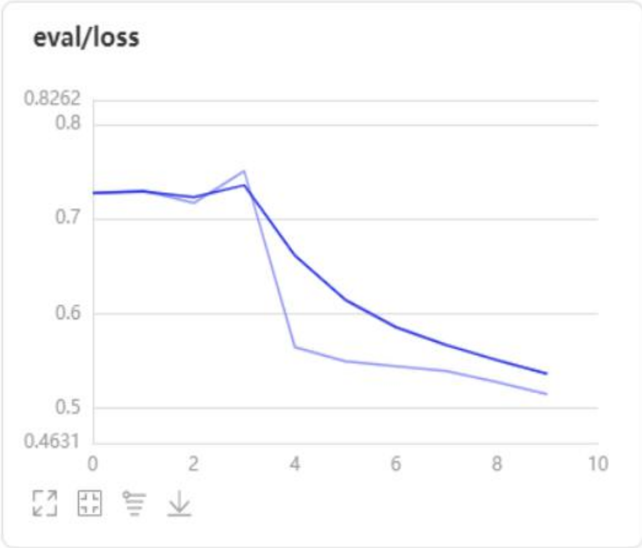
In order to verify the accuracy of the semantic matching model more effectively, we use lcqmc, a commonly used authoritative data set, to do matching verification. Lcqmc data set is based on the question base provided by Baidu know. It is an authoritative public data set. The process of semantic matching is also a process of classification to some extent. According to the user's questions, the training data set is identified one by one in the material library. In this way, the semantic matching problem can be transformed into a binary classification problem .

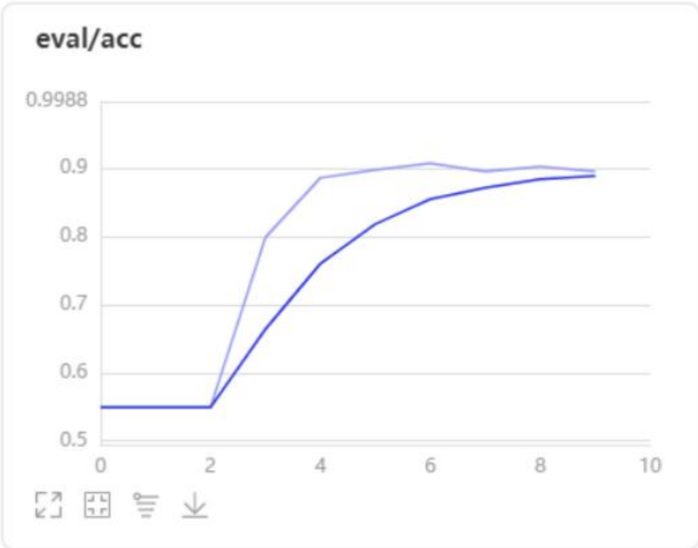
In the experiment, we use ernie-gram model to train, evaluate and predict the data. In these three processes, we need to divide the lcqmc data set into training set, verification set and test set in a planned way. The training set mainly trains the parameters of ernie-gram model to find an optimal parameter, The model will adjust its parameters dynamically according to the data characteristics and application scenarios of the training set to achieve the effect of accurate classification.

Verification set is mainly used to evaluate and verify the trained model, and to verify the advantages and disadvantages of the model, convergence rate and convergence effect. In practice, the verification set is mainly used to adjust the super parameters of the model, and the model can be optimized by adjusting and selecting the super parameters. The test set is based on the test results of the model, combined with various statistical indicators to evaluate the indicators, so as to verify the generalization ability of the model.

There are many ways of word vector transformation. In the experiment, we use paddlenlp. Embedding. Token embedding to transform word vector.

5. CONCLUSION





As can be seen from the figure, with the increase of the number of iterations, the loss rate decreases gradually, while with the increase of the number of iterations, the accuracy rate increases gradually.

REFERENCES:

[1] Ajeet Ram Pathak,Aditee Mahajan,Anusha Nair. Analysis of Techniques for Rumor Detection in Social Media. Procedia Computer Science16 April 2020Volume 167 (Cover date: 2020)Pages 2286-2296.

[2] Yongcong Luo,Jing Ma,Chai Kiat Yeo. BCMM: A novel post-based augmentation representation for early rumour detection on social media. Pattern Recognition12 January 2021 Volume 113 (Cover date: May 2021) Article 107818.

[3] Fan Xu,Victor S. ShengMing,wen Wang. Near real-time topic-driven rumor detection in source microblogs. Knowledge-Based Systems21 August 2020Volume 207 (Cover date: 5 November 2020)Article 106391.

[4] Raja kumari Mukiri,B. Vijaya Babu,Prediction of rumour source identification through spam detection on social Networks- A survey. Materials Today: ProceedingsAvailable online 28 April 2021In press, corrected proof.

[5] Zhihong Wang,Yi Guo, Rumor events detection enhanced by encoding sentimental information into time series division and word representations. Neurocomputing28 January 2020Volume 397 (Cover date: 15 July 2020)Pages 224-243.

[6] Puneet Agarwal,Ridwan Al Aziz,Jun Zhuang, Interplay of rumor propagation and clarification on social media during crisis events - A game-theoretic approach. European Journal of Operational ResearchAvailable online 18 July 2021In press, corrected proof.

[7] Samah M. Alzanin,Aqil M. Azmi, Detecting rumors in social media: A survey. Procedia Computer Science15 November 2018Volume 142 (Cover date: 2018)Pages 294-300.

[8] Meiling Jiang,Qingwu Gao,Jun Zhuang, Reciprocal spreading and debunking processes of online misinformation: A new rumor spreading–debunking model with a case study. Physica A: Statistical Mechanics and its Applications1 December 2020Volume 565 (Cover date: 1 March 2021)Article 125572.

[9] Amirhosein Bodaghi,Jonice Oliveira,The characteristics of rumor spreaders on Twitter: A quantitative analysis on real data. Computer Communications1 July 2020Volume 160Pages 674-687.

[10] Ahmet Aker,Alfred Sliwa,Kalina Bontcheva, Rumour verification through recurring information and an inner-attention mechanism.Online Social Networks and Media29 August 2019Volume 13 (Cover date: September 2019)Article 100045.