Pattern Recognition Letters 138 (2020) 397–402

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

User personality prediction based on topic preference and sentiment analysis using LSTM model

Jinghua Zhao[a], Dalin Zeng[b,*], Yujie Xiao[c], Liping Che[a], Mengjiao Wang[d]

[a] School of Business, University of Shanghai for Science and Technology, Shanghai, China
[b] School of Management Engineering, Shandong Jianzhu University, Jinan, China
[c] School of Marketing and Logistics Management, Nanjing University of Finance and Economic, Nanjing, China
[d] School of Economic and Management, Tongji University, Shanghai, China

ARTICLE INFO

Article history:
Received 27 April 2020
Revised 9 July 2020
Accepted 27 July 2020
Available online 28 July 2020

Keyword:
Attention-based LSTM
LDA
Big-Five

ABSTRACT

Based on the original text information, this paper converts the users' theme preferences and text sentiment features into attention information and combines different forms with the LSTM (Long Short-Term Memory) model to predict the personality characteristics of social network users. Finally, the experimental results of multiple groups' show that the Attention-based LSTM model proposed in the paper can achieve better results than the currently popular methods in the recognition of user personality traits and that the model has good generalization ability.

1. Introduction

Social platforms provide users with a place to communicate, exchange, and express their views with others. The rise of social networks and the exponential growth of the volume of data have created opportunities for text research. Academic research can identify users' personality traits from the user corpus, social platform user characteristics, and other information. As a research problem aimed at identifying users' personality traits, personality recognition has received great attention in the past few years. First, personality traits, as a clear internal factor, are closely related to the topics that users are keen to discuss, which makes it possible to identify the personality traits of social network users based on user generated content (UGC). Second, with the in-depth development of the mobile Internet, the influence of networks in people's daily lives continues to deepen, the cooperation between different platforms is becoming more frequent, and cross-platform interaction is increasing. For example, Facebook accounts can be used to log in various shopping platforms, such as Sephora, Amazon, etc., which provides a broader application scenario for identifying users' personality traits based on social network data. However, most user propensity analyses focus on the relationship between user behavior and user personality. For example, Pennebaker et al. [1] found that people with high extraversion use more positive emotional words while those with high neuroticism tend to use more negative emotion words. Gosling et al. [2] demonstrated the relationship between 11 behavioral characteristics of users on Facebook and their personality, and confirmed the correlation between them.

In recent years, since it has been determined that there is indeed a relationship between user behavior and user personality, how to determine a user's true personality based on information and use that personality to predict behavior and guide business decisions has become a problem of practical value. Personality analysis and personality prediction in user portraits have received great attention from academia. The fundamental purpose is to understand the user's behavior and mine the user's interest in the social network. With the knowledge of the user's personality, the social networking company can provide better service and enhance the user experience. In the field of user personality prediction, the main research method is to mine the information generated by users on social networking sites and use machine learning to predict user personalities. Back et al. [3] judged the personality traits of users through the linguistic features in their email addresses and pointed out that they can judge the personalities of individuals such as their neuroticism, openness, pleasantness, conscientiousness, and narcissism with a certain precision. Qiu et al. [4] proved that the linguistic characteristics of Twitter can accurately determine the agreeableness and neuroticism personality traits of Twitter users. Globeck et al. [5,6] crawled and modeled the text information shared by users on Facebook and used ZeroR and Gaussian process regression methods to predict the scores of different types of personalities for users, which can reach

* Corresponding author.
E-mail addresses: zhaojinghua@usst.edu.cn (J. Zhao), zengdalin@sdjzu.edu.cn (D. Zeng), yujiexiao@nufe.edu.cn (Y. Xiao), cheliping_psy@sina.com (L. Che).

https://doi.org/10.1016/j.patrec.2020.07.035
0167-8655/

an 11% -18% error rate. Quercia et al. [7] used the basic information of users on Twitter: the number of fans, the number of followers and the number of states to predict the personality characteristics of users. Ahmad et al. [8] aimed to predict user traits based on social media by gathering certain information pertaining social network usage behavior based on features such as likes, number of comments, number of friends, etc. Bai et al. [9] analyzed the correspondence between the microblog information and personality information of more than 400 microblog users. First, the users' scores on the Big Five personality traits were obtained through a questionnaire. Second, the personalized data of 29 dimensions in four categories for microblog users were selected. Pearson's correlation coefficient was used to measure the correlation between the users' big five personality traits and the 29 dimensions and the users' personalities were predicted by the single-task regression model, the incremental regression model, and the multitask regression model, respectively. Lima et al. [10] used groups of data to make personalized predictions for users. Different from previous research, it directly extracted the meta-attributes related to the Big Five personality traits from the text and transformed the essence of the problem into five sets of two-class classification problems, which are solved by using semisupervised methods. The experiment finally uses the Naive Bayes, SVM, and multilayer neural networks for testing and the results show that this method can improve the precision of personality prediction for Twitter users.

Wan et al. [11] used the actual personality scores of 131 users, their microblog data, and text to complete the prediction of the personalities of microblog users using machine learning methods. In general, the research on the personality prediction of social network users mostly uses machine learning methods and the research object is the real data sets of social networking sites. However, further improvement is needed in model building and feature selection to improve the prediction precision.

To explore the effective methods of user personality prediction, this paper takes the Big Five model theory and LDA theme model as a priori knowledge, introduces an attention mechanism and builds an Attention-based LSTM model that combines thematic and emotional features. Based on the original text information, this model converts the users' theme preferences and text sentiment features into attention information. It is combined with the LSTM model in different forms to predict the personality characteristics of social network users.

2. User personality prediction analysis framework

The attention mechanism was first used in image processing tasks. By adding attention information, the model highly focused on some important information of the input image so that the model parameters could be learned and adjusted according to this information. The success of the attention mechanism in the imaging field has made its neural network model one of the research hotspots in recent years. In the natural language processing task, Bahdanau et al. [12] proposed a recurrent network model based on the attention mechanism and applied it to automatic translation tasks. By using in-network computing at each time step, the attention mechanism allows the model to pay close attention to specific word information so that it can learn and adjust the parameters according to specific words during translation, thereby achieving better results than traditional methods. Furthermore, the proposed model also verifies the effectiveness of the attention mechanism in natural language processing tasks. Yin et al. [13] proposed a convolutional neural network based on the attention mechanism. This model constructs the attention matrix in the convolutional layer and the pooling layer, respectively. Through the attention information of different network layers, it can learn the interdependence between sentence pairs. The proposed model also verifies

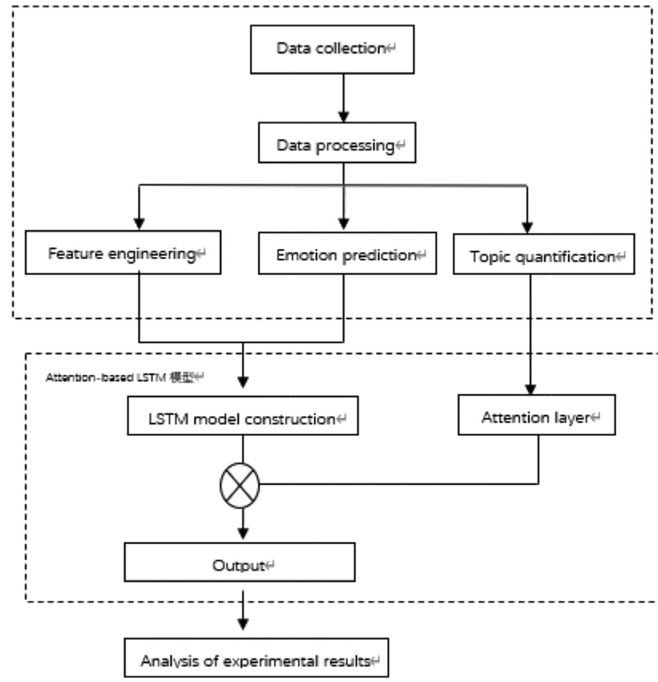

Fig. 1. Framework of the user personality prediction.

the effectiveness of the attention mechanism and convolutional neural network in natural language processing tasks. Wang et al. [14] combined the attention mechanism at the word content level into the LSTM network model so that each neuron highly focused on the emotional information of a specific target during the training process through the calculation and update of the attention matrix. In this way, the contextual words that are important to the sentiment polarity of the target word in the sentence are excavated and, finally, the good emotion polarity recognition results are obtained from multilingual data sets in different fields. The above studies have fully demonstrated the effectiveness of the attention mechanism in the field of natural language processing. Therefore, based on the attention model, the paper proposes a personality prediction algorithm.

The Attention-based LSTM proposed in the paper is mainly composed of the following 4 parts.

(1) Input matrix. The correlation method is used to transform the words in the corpus into word vectors and splice them into two-dimensional matrices.
(2) LSTM model. The user corpus is input into the LSTM model in order and the output result is used as one of the inputs of the Attention layer. The matrix output after the operation is the user's five personality prediction values.
(3) Quantification of theme preference. This section focuses on how to transform the subject information extracted by LDA into an input matrix. LDA is used to extract the topic probability distribution to complete the topic distribution extraction.
(4) Introduction of attention information. When constructing the Attention layer, this model needs to consider the input position of the two attentions about emotional information and topic information, the form of the calculation method, and the matrix.

In general, the purpose of the paper is to explore the effect of Attention-based LSTM on the personality prediction of social network users and to improve the precision of the model for such tasks. The overall framework is shown in Fig. 1. It is mainly divided

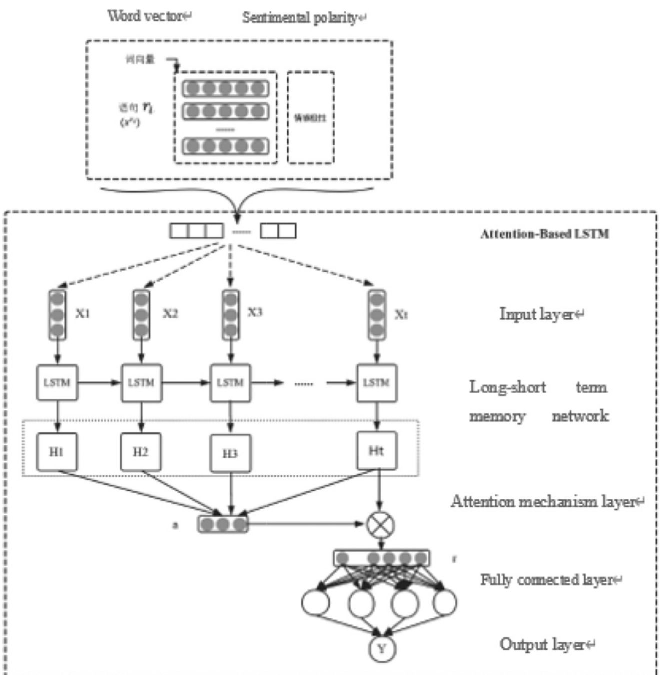

Fig. 2. Structural diagram of the Attention-based LSTM model.

into three parts: data processing, Attention-based LSTM model construction, and the analysis of the experimental results.

3. Analysis of the model structure

Since the data set used in this paper is a Facebook corpus containing user personality tags, this model is a supervised deep learning model.

The Attention-based LSTM model built in the paper needs to consider the introduction and calculation of the attention layer based on the LSTM model; therefore, the structure of this model is more complicated. The structural diagram is shown in Fig. 2. During the construction of the model, the original text information needs to be input into the LSTM model together. When introducing the attention mechanism, it is necessary to consider the position of the attention mechanism and the corresponding calculation method.

3.1. Quantitative analysis of themes

The paper uses the LDA model to convert the topics that users follow/favor into input vectors. The first step is to select the user's expected information. If the topic information is extracted only for a certain microblog of a social network user, the text is too short and the extracted main body information is limited or invalid. Therefore, it is reasonable to use a user's text for a while to analyze and extract the network topics that the user cares about/favors and add them as Attention to the LSTM model. The second step is to list all the subject information. The existing and mature topic corpus can be used to list topic information, but considering that the existing corpus is too large, there is much irrelevant information. Therefore, the paper adopts the data used in this research to set the number of topics as 30 and the number of subject words under each topic as 20, which are used to establish the LDA topic model.

The above is the topic distribution information of a certain user's corpus. For all users N in this experiment, a topic matrix of N * 30 can be obtained after subject prediction by the LDA model. The topic probability distribution of most users is not 0, which means that even if each user has personal preferences regarding topic information, the topics involved in their daily expectations are more extensive. To strengthen the strong connection between the user and the topic and weaken the weak connection between the user and the topic, the probabilities for each user are calculated as follows and shown in formula 1.

$$P(n,\ i) = \begin{cases} P(n,\ i) & P(n,\ i) > \propto \\ 0 & P(n,\ i) < \propto \end{cases} \tag{1}$$

After analysis and experiment, $\propto$ in this section is set to 0.15.

3.2. Build the attention mechanism layer

Considering the characteristics of different information and the size of the input matrix, this model processes the emotional information together with the original text information into a two-dimensional matrix and inputs it into the LSTM model. Then, the topic information is used as the Attention mechanism layer, together with the output matrix of LSTM.. The result after the operation is the value predicted by the user's personality.

According to the word vector processing method, the paper uses Word2vec's Skip-gram method to process the original text information. In other words, if the text T contains R sentences, the text is divided into R regions as $T = \{r_1,\ r_1,\ \ldots r_i,\ r_j,\ r_k,\ \ldots,\ r_R\}$. A region is equivalent to a sentence and the region contains several words in the form of $r_i = \{w_1^{r_i},\ w_2^{r_i},\ \ldots,\ w_I^{r_i}\}$, $r_j = \{w_1^{r_j},\ w_2^{r_j},\ \ldots,\ w_J^{r_j}\}$, and$r_k = \{w_1^{r_k},\ w_2^{r_k},\ \ldots,\ w_K^{r_k}\}$. Each vocabulary, such as $w_1^{r_i}$, exists in the form of a word vector. Therefore,$r_i \in R^{d*|V|}$, where $|V|$ represents the sentence length, that is, the number of words contained in the sentence; and d represents the dimension of a word vector, which is $r_i \in R^{301*|100|}$ in the problem. The text information of the processed users is input into the LSTM and the output matrix is $H = \{h_1,\ h_2, \ldots, h_N\}$, where $H \in R^{d*N}$, and$N$ is the number of users. Because the standard LSTM model uses the same state vector for each step of the prediction, it cannot fully learn the details of the sequence coding during the prediction. To better capture the effective information in the text data and grasp the core key features, this paper adds the topic information $\propto$, where $\propto \in R^N$, to the classification method, and its calculation formula is shown in 2.

$$r\ = H \otimes \propto \tag{2}$$

r is the combination matrix of theme $\propto$ and LSTM output matrix H.

The tanh function is used to transform the subject information and the output matrix of the model. The calculation is shown in Eq. (3), where W_r and W_N are the matrix r and the weight matrix of matrix h_N, respectively. Through the following model training steps, the components of each weight matrix can be adjusted to adjust the influence of the different vectors on the classification results so that the model can fully mine the information of the sentence and improve the precision of the prediction of the user's personality traits.

$$h^* = \tanh(W_r r + \ W_N h_N) \tag{3}$$

After going through a fully connected layer and Softmax layer, the output of $h^*$is the value predicted by the user's personality.

In the above, the paper introduced the Attention mechanism layer of the Attention-based LSTM model in detail. Its performance will be verified in the experiment in the next section. Additionally, many other parameters in many models have no clear settings in the current deep learning field. These properties of deep neural networks are often set based on experts' experience; therefore, this

Table 1
Topic distribution of a user.

Topic ID	Topic probability
0	0.063979336343144743
2	0.19344804518287365
6	0.049013217087340168
7	0.31535985300654678
8	0.074829314949487347
14	0.04697730038468321
15	0.04443453569838469
18	0.09128154138884592
28	0.08502085435624978

paper determines the number of hidden layers and number of neurons of LSTM by building experiments, and finally gets the optimal result through parameter adjustment (Table 1).

4. Experiments and results

4.1. Corpus collection and evaluation indicators

The thesis uses real Facebook data to construct the model. The experimental data are derived from two parts: one is the corpus, that is, the microblog information and the score of the five personalities of the users themselves; and the other is the result extracted based on the LDA model. The skip-gram method of word2vec is used to generate the word vector and take it as part of the input. The classification of corpus information and personality traits is shown in Table 2. Another part of the input—emotional words can be extracted from the LDA model. Thus, the input matrix of the Attention-based LSTM model is obtained.

The second is the input information of the Attention layer. LDA topic model training is conducted on the whole Facebook corpus and the number of topics is set to 30. For the obtained topic dictionary, matching the corpus of user n with it can obtain a 30-dimensional probability vector, which represents the topic distribution of user n for these 30 topics. Using the theme quantitative analysis method to process the theme distribution vector can obtain the theme distribution information of all users and this information is used as the input of the attention layer.

This experiment divides the data into a training set and a test set at a ratio of 7:3. The training set has a total of 6893 data and the test set includes 3024 data. All the experimental content of this paper is based on the Python language using Pycharm as the development tool. The experiment uses the precision, recall and F1 score to evaluate the performance of the model. The calculation method is shown below. Table 3 shows the classification matrix.

Here, TP represents the number of positive examples that are classified correctly (true positives), FN represents the number of negative examples that are classified incorrectly (false negatives), FP represents the number of positive examples that are classified incorrectly (false positives), and TN represents the number of positive examples that are classified correctly (true negatives).

According to the numbers of TPs, FNs, FPs, and TNs, the corresponding precision, recall and F1 Score (F1) can be calculated.

Formula 4-7 to formula 4-9 show the calculation method of the recall rate, precision rate, and F1, respectively. In a binary classification system, precision refers to the proportion of correctly classified samples with respect to the total number of samples, which reflects the classification performance of the classification model for the overall sample, and the value range is [-1, 1]. The larger the value is, the better the classification performance of the model. The specific calculation is shown in Equation 4-7. The recall rate evaluates the overall data prediction of a category and its calculation is shown in formula (8).

Table 2
Corpus and personality of some Facebook users.

User id	corpus	Extroversion	Neurotic ism	Agreeableness	Conscientiousness	openness
1112	likes the sound of thunder.	N	Y	N	N	Y
1112	is so sleepy it's not even funny that she can't get to sleep.					
1112	is sore and wants the knot of muscles at the base of her neck to stop hurting. On the other hand, YAY I'M IN ILLINOIS! <3					
1112	likes how this new song sounds today					
1112	is watching cousin play computer games on the television. Also, sleepy.					
1113	should spend less time writing chapters for the story she'll never finish, and try reading more books. Preferably, during the daytime, and not at 3:00 AM.	N	Y	N	Y	N
1113	saw a nun zombie, and liked it. Also, Tentacle Man + Psychic Powers = GREAT Party.					
1113	Doesn't mind coloring with markers as much as she thought. Yay for coloring vividly creepy villainesses! :D					
1114	was about to finish a digital painting before her tablet went haywire. Is now contemplating the many ways she wishes to exact her revenge on faulty technology.	Y	Y	N	N	N
1114	is celebrating her new haircut by listening to swinger music and generally looking like a doofus.					
1114	has a crush on the Green Lantern.					

Table 3
Classification matrix.

Real results	classification results	
	Positive examples	negative examples
Positive examples	TP	FN
negative examples	FP	TN

Table 4
Training precision and time of different hidden layers.

Hidden layers	Precision	Time consumption/s
1	55.31%	45
2	57.95%	65
3	55.13%	88

$$\text{Precision} = \frac{TP}{TP+FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP+TN} \tag{9}$$

$$\text{F1} = \frac{2TP}{2TP+FN+FP} = \frac{2{*}\ \text{Precision} *\ \text{Recall}}{Precision +\ \text{Recalll}} \tag{10}$$

F1 is the harmonic mean of the recall and precision. The recall rate reflects the classification model's ability to recognize positive examples. The higher the recall rate is, the stronger the model's ability to recognize positive examples. The precision rate reflects the model's ability to distinguish negative examples. The higher the precision rate is, the stronger the model's ability to distinguish negative examples. The F1 score is a combination of the two with a value range of [0, 1]. The higher the F1 measurement is, the more robust the classification model. The specific calculation is shown in formula 4-9. In the sentiment binary classification problem mentioned in this article, the comprehensive evaluation of the precision and recall rate by the F1 score is more important.

4.2. Determination of the number of hidden layers of LSTM

When the number of hidden layers is different, Attention-based LSTM will have different learning effects and output sequences. The purpose of this section is to find the number of hidden layers that makes the classifier perform the best. There are three experiments in this section. The variable of the experiment is the number of hidden layers. The When the other variables remain unchanged and the number of hidden layers is 1, 2, and 3, the precision and time consumption of the model are calculated, respectively. The number of neurons in the hidden layers of the three experiments is 128. The results are shown in Table 4.

It can be found from the table that the number of hidden layers has a strong correlation with the time consumption. The more hidden layers that exist, the more time it takes to train the same data. This is because when the number of hidden layers increases, the number of neurons to be trained in the model increases exponentially. The increase in the parameters will inevitably bring about increased resource and time consumption. However, there is no strong correlation between the number of layers and the precision, that is, an increase in the number of layers does not increase the precision of the model. Therefore, it is not that more hidden layers are better.

It can be seen from the table that when the number of hidden layers is two layers, the time consumption is moderate and the most important thing is that the precision of the two-layer neural network is the highest. Therefore, the following LSTM models use two hidden layers.

Table 5
Training precision and time for different numbers of neurons.

Number of neurons	Precision	Time consumption/s
128-128	57.95%	65
128-64	56.09%	64
256-128	55.13%	86

Table 6
LSTM parameter setting table.

Hyperparameter	value
Number of hidden layers	2
Number of neurons	128
Dropout	0.5
Learning rate	0.001

4.3. Determination of the number of LSTM neurons

The experiment in the previous section determined the number of layers of the model by adjusting the single variable of the number of hidden layers. However, for the convenience of the experiment, the premise of the experiment is that the number of neurons in each layer is 128. Therefore, in this section, after determining the number of hidden layers of the model, the number of neurons in each layer is determined through experiments to make the Attention-based LSTM model better. Because LSTM has two hidden layers, it is necessary to conduct an experiment of the number of neurons in the two-layer network. Three sets of experiments were designed and the numbers of neurons were (128, 128), (256, 128), and (128, 64), respectively. The precision and time consumption of each group of experiments were calculated and the respective results are shown in Table 5.

It can be obtained from the table that the number of neurons has little effect on the training time, and so this experiment takes the set of parameters with the highest precision, that is, the number of hidden layers is 128.

4.4. Results and analysis of the experiments

(1) Parameter setting
Through the above experiments, the test parameter settings of the LSTM part of this model can be determined, as shown in Table 6.
The above three groups of experiments determine the basic overview of the LSTM model used in this paper. In addition to the above two groups of experiments, this paper also conducted experiments on the selection of the activation function and the selection of the dropout rate. In this paper, dropout with a rate of 0.5 is connected behind the two hidden layers. Due to space limitations, the two sets of experiments are not repeated here.
The summary is as follows. First, the embedded word vectors are trained based on Word2Vec in the input layer and sentiment analysis features are added. Second, the LSTM model has 2 hidden layers and each hidden layer has 128 neurons. The LSTM output is combined with the Attention layer and calculated. The output layer of the model is a fully connected layer with 1 neuron and then, through the Softmax activation function, the user's personality category is finally output.

(2) Analysis of the results
To prove the effectiveness of this model, this section compares the proposed model with the traditional machine learning method that is widely used at present. The comparison indicators include the precision, recall and F1 score. The

Table 7
forward classification results of attention-based LSTM and other machine learning methods.

	Precision	Recall	F1-measure
Bayesian Network	55.11%	51.43%	63.97%
Random forest	50.12%	59.31%	67.18%
Support Vector Machine (SVM)	57.26%	53.45%	62.35%
Attention-based LSTM	57.95%	65.78%	72.2%

selected models are as follows: a Bayesian network, a random forest, and a support vector machine (SVM). In this section, tenfold cross validation method is adopted. Each time, the data sets are randomly classified according to 7:3. The training set is trained for ten times, and the test set is tested for ten times. The average result of the ten times is taken as the final result of the model. The experimental results are shown in Table 7.

Table 7 shows the indicators of each classifier in the prediction of the big five personalities. It can be seen that attention-based LSTM achieves good performance in the three metrics of the accuracy, recall rate, and F1 score. Because the user's personality is divided into five dimensions and each dimension has two values (yes/no), the user's personality is divided into a total of 32 categories. In this paper, after learning the text information, the prediction accuracy is improved from 1/32 to 57.95%, which proves the good classification effect of the Attention-based LSTM model. In addition, except for the fact that the accuracy is not much different from the support vector machine, the remaining indicators of Attention-based LSTM are more than five percentage points better than those of the other classifiers, showing better performance.

This experiment proves that the Attention-based LSTM model constructed in this paper can better predict a user's personality. Additionally, it has a good generalization ability and is more accurate than other methods.

Conclusions

The paper combined an attention mechanism and an LSTM network to propose a deep network model Attention-based LSTM for user personality prediction. The feature information of the sentences extracted by other algorithms is used as the attention information of the LSTM network, which can make the model pay close attention to a specific feature during the training process and can pay close attention to the word information that is important to a specific feature to mine out more hidden information. Then, the LSTM network, which receives the sequential input of words, acquired the dependency of the target entity in the sentence and identified the differences between different users in the use of text, which is helpful to better identify a user's personality.

Declaration of Competing Interest

None.

Acknowledgements

This work was supported by the National Natural Science Fund (71672128), the National Key Research and Development Program of China (2018YFC0830400) and Natural Science Fostering Foundation (1F-19-303-001).

References

[1] W Pennebaker J, E Francis M, J Booth R, Linguistic inquiry and word count: LIWC 2001[J], Mahway: Lawrence Erlbaum Assoc. 71 (2001) (2001) 2001.
[2] D Gosling S, A A Augustine, S Vazire, et al., Manifestations of personality in online social networks: self-reported facebook-related behaviors and observable profile information[J], Cyberpsychol. Behav. Soc. Netw. 14 (9) (2011) 483–488.
[3] D Back M, C Schmukle S, B Egloff, How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses[J], J. Res. Personal. 42 (4) (2008) 1116–1122.
[4] L Qiu, H Lin, J Ramsay, et al., You are what you tweet: personality expression and perception on twitter[J], J. Res. Personal. 46 (6) (2012) 710–718.
[5] Golbeck J, Robles C, Edmondson M, et al. Predicting personality from twitter[C],2012:149–156.
[6] Golbeck J, Robles C, Turner K. Predicting personality with social media[C], 2011:253–262.
[7] D Quercia, M Kosinski, D Stillwell, et al., Our twitter profiles, our selves: predicting personality with twitter[J], in: IEEE Third International Conference on Privacy, Security, Risk, and Trust, 2012, pp. 180–185.
[8] Zaaba Ahmad, Lutfi, et al., Personality prediction of malaysian facebook users: cultural preferences and features variation[J], Adv. Sci. Lett. (2017).
[9] Bai S, Hao B, Li A, et al. Predicting big five personality traits of microblog users[C], 2013:501–508.
[10] A-C Lima, N De Castro L, A multi-label, semi-supervised classification approach applied to personality prediction in social media[J], Neural Netw. 58 (5) (2014) 122–130.
[11] D Wan, C Zhang, M Wu, et al., Personality Prediction Based on All Characters of User Social Media Information[J], Commun. Comput. Inf. (2014).
[12] Meng F, Lu Z, Tu Z, et al. Neural transformation machine: a new architecture for sequence-to-sequence learning[J], 2015.
[13] W Yin, H Schütze, B Xiang, et al., ABCNN: attention-based convolutional neural network for modeling sentence pairs[J], Comput. Sci. (2015).
[14] Y Wang, M Huang, X Zhu, et al., Attention-Based LSTM for Aspect-Level Sentiment Classification[C], 2017, pp. 606–615.