2016 First IEEE International Conference on Computer Communication and the Internet

Text Sentiment Analysis Based on Long Short-Term Memory

Dan Li, Jiang Qian
School of Science
Beijing University of Posts and Telecommunications
Beijing, China
e-mail: lidan18811719039@163.com, jqian104@126.com

***Abstract*—With the rapid development of Internet and big explosion of text data, it has been a very significant research subject to extract valuable information from text ocean. To realize multi-classification for text sentiment, this paper promotes a RNN language model based on Long Short Term Memory (LSTM), which can get complete sequence information effectively. Compared with the traditional RNN language model, LSTM is better in analyzing emotion of long sentences. And as a language model, LSTM is applied to achieve multi-classification for text emotional attributes. So though training different emotion models, we can know which emotion the sentence belongs to by using these emotion models. And numerical experiments show that it can produce better accuracy rate and recall rate than the conventional RNN.**

Keywords-sentiment analysis; RNN; LSTM

I. INTRODUCTION

As people post more and more text information on Internet, it has been a great challenge to distinguish whether the information is useful or not. As a result, it's very urgent to develop language models to dig out valuable information, Specifically for product reviews, mining useful product review information may help merchants to develop sales strategy effectively. However, the traditional method is very limited, as it's unable to deal with huge amounts of data timely. Consequently, establishing an efficient sentiment system is getting more important. So we need language models if we want to analysis text sentiment, and RNN is a good model.

While feed-forward networks are able to take into account only a fixed context length to predict the next word, recurrent neural networks (RNN) can take advantage of all previous words. Traditional RNN language model is going further in model generalization: instead of considering only the several previous words (parameter n) the recursive weights are assumed to represent short term memory. More in general we could say that RNN sees text as a signal consisting of words. Long Short-Term Memory (LSTM) neural network is different type of RNN structure. It structure allows to discover both long and short patterns in data and eliminates the problem of vanishing gradient by training RNN. LSTM approved themselves in various applications and it seems to be very promising course also for the field of language modeling [3].

II. RELATED RESERTCH

In the study of text sentiment analysis, the sequential relationship between words is of critical importance. Mikolov [1] proposed a language model known as Recurrent Neural Network (RNN), which is publicly recognized as pretty suitable to process text sequence data. RNN consists of three modules, which are input layer, hidden layer and output layer. In RNN, the input layer at time 't' together with the hidden layer at time 't' are aggregated as a new input layer to calculate the hidden layer at time 't'. With such a loop structure, the hidden layer successfully reserves all information in previous words, which improves the performance of identifying the sequential relationships between words [1]. So RNN is a network that contains loops, and it allows information to be persistent.

In theory, the RNN language model could cover the time order structure of the whole text, and deal with long-term dependence problem. In practice, however, RNN could not learn the knowledge successfully. When the interval between the relative information of texts and the current location to be predicted becomes large, some problems will come out. As there are too many unfold layers in the back propagation through time optimization algorithm(BPTT), which leads to history information loss and gradient attenuation while training. To overcome this difficulty, some researchers put forward a strategy named Long Short-Term Memory (LSTM), which leads to better experimental results in some application scenarios.

LSTM through deliberate design to avoid long-term dependence, in practice, remember the long term information is the default behavior of LSTM. At present, LSTM network is the most widely used one, it replaces RNN node in hidden layer with LSTM cell, which is designed to save the text history information. LSTM uses three gates to control the usage and update of the text history information, which are input gates, forget gates and output gates respectively. The memory cell and three gates are designed to enable LSTM to read, save and update long-distance history information. The structural diagram is shown in Figure 1 [2].

Figure 1 provides an illustration of an LSTM memory block with a single cell. An LSTM network is the same as a standard RNN, except that the summation units in the hidden layer are replaced by memory blocks, as illustrated in Figure 2 [2].

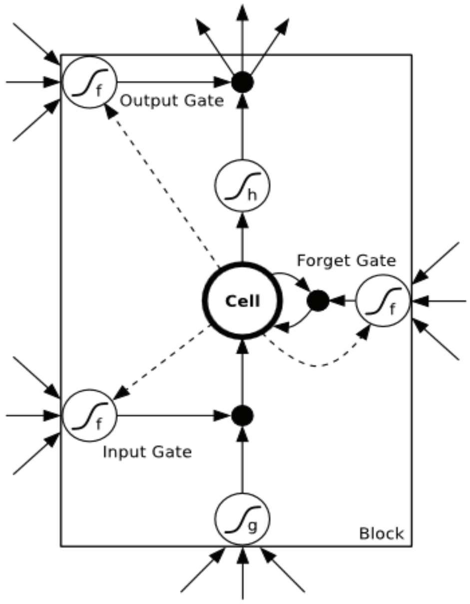

Figure 1. The structural diagram of LSTM cell.

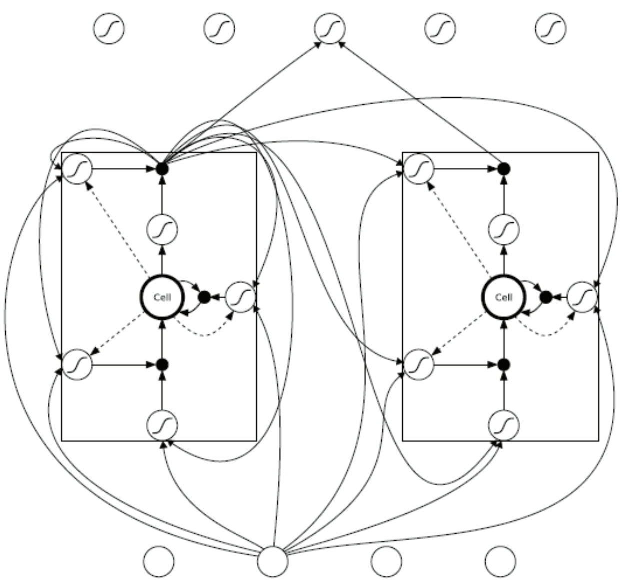

Figure 2. A LSTM network.

The cells state of LSTM is key, the vertical line run through in the middle, only a small amount of information interacted. LSTM remove or add ability from information to the state of the cell through the 'gate' structure, and the 'gate' is a way to make information selectively. The first step, forget gates determine which information from the cell state should be discarded. The second step, input gates determine the new information that is stored in the cell state. The third step, update the old cell state, using the above input gates and forget gates information to calculate the updated value of the cell state. Finally, the output gates determine the value of the output, which is based on the state of the cell [8].

The calculation process of LSTM mainly includes 4 steps. 1)Calculate the values of forget gate and input gate 2)Update the state of LSTM cell. 3)Calculate the value of output gates. 4)Update the output of the whole cell [2][3][4][5]. The detailed formula is shown as below.

Input Gates:

$$a_l^t = \sum_{i=1}^{I} w_{il} x_i^t + \sum_{h=1}^{H} w_{hl} b_h^{t-1} + \sum_{c=1}^{C} w_{cl} s_C^{t-1}$$
$$b_l^t = f(a_l^t)$$

Forget Gates:

$$a_\emptyset^t = \sum_{i=1}^{I} w_{i\emptyset} x_i^t + \sum_{h=1}^{H} w_{h\emptyset} b_h^{t-1} + \sum_{c=1}^{C} w_{c\emptyset} s_C^{t-1}$$
$$b_\emptyset^t = f\left(a_\emptyset^t\right)$$

Cells:

$$a_c^t = \sum_{i=1}^{I} w_{ic} x_i^t + \sum_{h=1}^{H} w_{hc} b_h^{t-1}$$
$$s_c^t = b_\emptyset^t s_c^{t-1} + b_l^t g(a_c^t)$$

Output Gates:

$$a_l^t = \sum_{i=1}^{I} w_{iw} x_i^t + \sum_{h=1}^{H} w_{hw} b_h^{t-1} + \sum_{c=1}^{C} w_{cw} s_C^t$$
$$b_w^t = f(a_w^t)$$

Cell Outputs:

$$b_c^t = b_w^t h(s_c^t)$$

In the above functions, $g(z)$ is the sigmoid function, and $h(z)$ is the tanh function.

III. The LSTM Network Process of Text Sentiment Analysis

From the perspective of language model, the RNN with LSTM can be regarded as an improved model of the conventional RNN language model. They both calculate the error of each model via putting the text statements as the input sequence. The smaller error indicates higher degree of confidence of text statement under this model [6]. But generally the RNN model with LSTM is more effective to overcome the sequence information attenuation problem when the text sequence information is rather long. So we would apply the RNN with LSTM on the text sentiment analysis.

We deal with both English and Chinese texts. For an English sentence, it is naturally split into participles for each word. However, Chinese word does not have the natural participle as English word has. So word segmentation is necessary before the Chinese text sentiment analysis [7]. For a sentence as a Chinese text word sequence, we first apply a word segmentation technique (here we used CRF++-0.58) to transform the sentence into participles. Then following the natural order, from left to right, of the participles in the sentence, traversal one by one, through the forward calculation of LSTM. The output results are the probabilities of the word at time 't', giving the vocabulary sequence before time 't'. Finally the error value of the sentence is measured by the joint distribution probability of all words. Higher joint distribution probability will result in lower value of text statement error.

The detailed process is introduced as follows:

- In the training phase, the training data are divided into several categories, according to their emotional labels. Then the LSTM models are trained for each category of data, resulting in several LSTM models, each for the corresponding emotional reviews.
- To forecast the emotion of a new input review, the LSTM models obtained in the training phase are evaluated on the new input review, giving error values. The model giving the smallest error value is assigned as the emotional category to the new input review.

The main process of the training phase is shown in the figure 2 below, where the data are classified into three categories as an illustration: positive, negative and neutral.

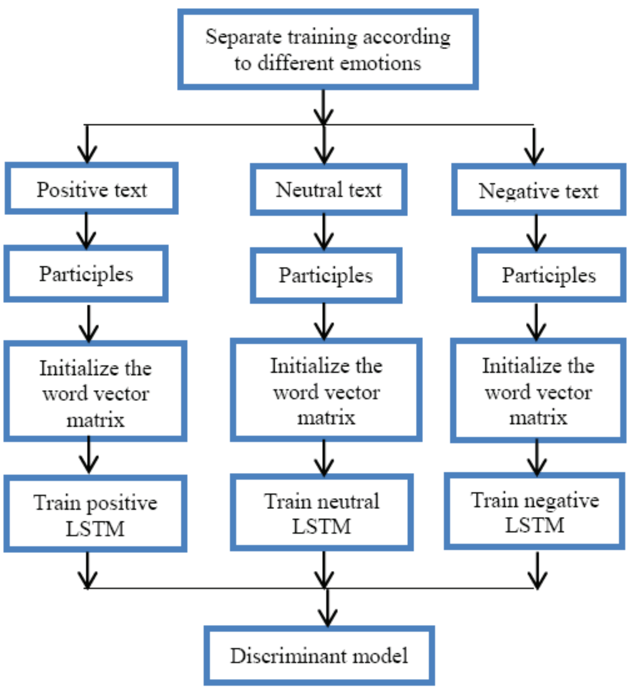

Figure 2. The flow diagram of emotion classification.

Compared with the conventional RNN language model, RNN with LSTM could cover a long sentence completely, and does perform better in many experiments, especially for structures with conjunctions, such as 'although…but…', 'not only…but also…', 'however', 'what is more', 'in addition' and so on.

IV. THE RESULTS AND ANALYSIS

A. The Experimental Data Set

We used four sets of data: comments from the website JD.COM, travel comments from ctrip and English movie reviews (There are two kinds of data of JD.COM). Among them, comments from JD.COM and travel comments from ctrip are in Chinese.

After necessary data cleaning, one kind of comments from JD.COM are manually classified into three categories: positive emotion, neutral emotion and negative emotion. The number of total sample data set is 39000, with proportion1:1:1 for the three categories. A random sample of 3000 for each class are chosen as testing sets, while the others are used as training sets. The other kind of comments from JD.COM are classified into two categories: positive emotion and negative emotion. The number of positive training data is 19493, the number of negative training data is 23955, and the number of positive test data is 10000, the number of negative test data is 8000.

The travel comments from ctrip and English movie reviews are classified into two categories manually: positive emotion and negative emotion. For data set from ctrip, the number of training data is 6000 for each category, and the number of test data is 2000 for each category. For English movie reviews, the number of training data is 12500 respectively, and the number of test data is 12500 respectively.

B. The Experiment Results Analysis

The training process uses BPTT algorithm to train the weights. The number of unfold state layer would affect the training accuracy. More unfold state layers generally lead to better results, but also cost higher computational complexity. In the meantime, the congenital advantage of LSTM structure requires less number of unfold layers to achieve comparable results with the conventional RNN. So we set the number of unfold layer in BPTT as 10 in our experiment.

We apply both the RNN with LSTM and the conventional RNN on each data set. Sentiment analysis results are shown in the following tables.

DATA SET FROM JD.COM:

TABLE 1 RESULTOF LSTM

	Positive	Negative	Neutral	Accuracy rate	Recall rate
Positive (3000)	2857	20	123	95.62%	95.23%
Negative (3000)	39	2661	300	93.07%	88.7%
Neutral (3000)	92	178	2730	86.58%	91%

TABLE 2 RESULTOFCONVENTIONAL RNN

	Positive	Negative	Neutral	Accuracy rate	Recall rate
Positive (3000)	2825	41	134	94.83%	94.16%
Negative (3000)	45	2642	313	91.7%	88.06%
Neutral (3000)	109	198	2693	85.73%	89.77%

TABLE 3 RESULTOF LSTM

	Positive	Negative	Accuracy rate	Recall rate
Positive (10000)	9015	985	93.96%	90.15%
Negative (8000)	580	7420	88.28%	92.75%

TABLE 4 RESULTOFRNN

	Positive	Negative	Accuracy rate	Recall rate
Positive (10000)	8635	1365	92.76%	86.35%
Negative (8000)	674	7326	84.29%	91.57%

PUBLIC DATA SETS FROM CTRIP:

TABLE 5 RESULTOF LSTM

	Positive	Negative	Accuracy rate	Recall rate
Positive (2000)	1752	228	88.80%	87.6%
Negative (2000)	221	1779	87.77%	88.95%

TABLE 6 RESULTOFRNN

	Positive	Negative	Accuracy rate	Recall rate
Positive (2000)	1735	265	87.89%	86.75%
Negative (2000)	239	1761	86.92%	88.05%

PUBLIC DATA SETS OF ENGLISH MOVIE REVIEW:

TABLE 7 RESULTOF LSTM

	Positive	Negative	Accuracy rate	Recall rate
Positive (12500)	10568	1932	89.42%	84.54%
Negative (12500)	1251	11249	85.34%	89.99%

TABLE 8 RESULTOFRNN

	Positive	Negative	Accuracy rate	Recall rate
Positive (12500)	10413	2087	88.35%	83.3%
Negative (12500)	1373	11127	84.21%	89.02%

These results show that RNN with LSTM can lead to better accuracy rate and recall rate than the conventional RNN. Specifically, RNN with LSTM can identify many text instances with the structure like 'although…but…', 'not only…but also…', 'however' and so on. It identifies some long statements better than the conventional RNN. Figure 3 are some examples from the English movie reviews, which are correctly identified by RNN with LSTM, while the conventional RNN fails.

Positive	**Negative**
i loved this movie from beginning to end. i am a musician and i let drugs get in the way of my some of the things i used to love (skateboarding, drawing) but my friends were always there for me. music was like my rehab, life support, and my drug. it changed my life. i can totally relate to this movie and i wish there was more i could say. this movie left me speechless to be honest. naturally, along with everyone else, i was primed to expect a lot of hollywood fantasy revisionism in they died with their boots on over the legend of custer. just having someone like errol flynn play custer is enough of a clue that the legend has precedence over the truth in this production. and for the most part my expectations were fulfilled(in an admittedly rousing and entertaining way).	This german horror film has to be one of the weirdest i have seen. i was not aware of any connection between child abuse and vampirism, but this is supposed based upon a true character. like i said, a very strange movie that is dark and very slow as werner pochath never talks and just spends his time drinking blood. foolish hikers go camping in the utah mountains only to run into a murderous, disfigured gypsy. the prey is a pretty run of the mill slasher film, that mostly suffers from a lack of imagination. the victim characters are all-too-familiar idiot teens which means one doesn't really care about them, we just wonder when they will die! not to mention it has one too many cheesy moments and is padded with endless, unnecessary nature footage. however it does have a few moments of interest to slasher fans, the occasional touch of spooky atmosphere, and a decent music score by don peake.

Figure 3. Examples from the English movie reviews.

Our experiments are on classification problems with two or three categories. But the model can be easily extended to text sentiment analysis with more categories.

V. CONCLUSION

In this paper, an improved RNN language model is put forward—LSTM, which successfully covers all history sequence information and performs better than conventional RNN. It is applied to achieve multi-classification for text emotional attributes, and identifies text emotional attributes more accurately than the conventional RNN.

For RNN model, scientists put forward an idea named 'Attention' which may lead to better results. This is to make RNN select information from larger information set in each step. Besides, 'Attention' is not the only development in RNN area. Scientists have proposed Grid LSTM which also showed excellent performances in some application scenarios. In future work, we could try these improvement programs or use different models combination to improve the performance of text sentiment analysis.

REFERENCES

[1] Sepp Hochreiter and Jurgen Schmidhuber, LongShort-Term Memory, Neural Computation, pages 12-91,1997.

[2] Alex Graves, Supervised Sequence Labelling with Recurrent Neural Networks, Studies in Computational Intelligence, pages 385-401, 2011.

[3] Daniel Soutner and Luděk Müller, Application of LSTM Neural Networks in Language Modelling, Lecture Notes in Computer Science, pages 105-112,2013.

[4] Martin Sundermeyer, Zoltan Tuske, Ralf Schluter, Hermann Ney,Lattice Decoding and Rescoring with Long-Span Neural Network Language Models, ICASSP, pages 661-665, 2014.

[5] Martin Sundermeyer, Hermann Neyand Ralf Schluter, From Feedforward toRecurrent LSTM Neural Networks for Language Modeling, IEEE/ACM Transactions on, 23(3):517-529, 2015.

[6] Jiang Yang, Shiyu Peng, Min Hou, Chinese comment text orientation analysis based on the theme of emotional words. The fifth national youth computational linguistics seminar, pages 569-572, 2010.

[7] Xin cheng Hu, the semantic relationship between classification research based on LSTM, Harbin Institute of Technology, pages 18-42, 2015

[8] http://www.open-open.com/lib/view/open1440843534638.html.