USING LSTM NEURAL NETWORKS FOR FAKE ONLINE USER REVIEW DETECTION

Özlem Uçar[1], Kenan Taşağal[2]
[1] *Trakya Üniversitesi Engineering Faculty, Edirne Turkey oucar@trakya.edu.tr*
[2] *Trakya Üniversitesi Engineering Faculty, Edirne Turkey kenantasagal@gmail.com*

Abstract
Fake online user comments are common problem that affecting other online users on the net. Therefore, it is important to identify fake user comments to highlight real user comments and ideas. In this study, we use a deep learning to identity the fake comments. In our application, %83 percent of success was achieved by using LSTM architecture to identify the fake comments.

Keywords: deep learning, fake review, lstm, text classification.

INTRODUCTION

In the modern era of e-commerce, online reviews are becoming more and more important in the consumer's decision to buy products or services. Unlike ads, online reviews are real consumers' approvals about products. A study by Luca [1] shows that when a product or business has increased the +1 star rating, it increases revenue by 5-9%. Due to the financial benefits associated with online reviews, paid or prejudiced reviewers write fake reviews to mislead a product (or business). Even these fake reviews have been formed by companies that have become a business industry with paid commentators and provide false comment services that affect the user. [2]

Fake user comments have become a common problem in recent days and many studies have been done on this subject. Ott et al.(2011) created a data set that they defined as "gold standard" [3] in their initial studies on the detection of false interpretations, and in their recent work, they achieved the detection of negative counterfeit comments with 86% accuracy using machine learning techniques.

Online fake user comments are generally composed of various types and short texts in terms of content. The present approaches have difficulty adapting to such short texts and are unable to achieve high accuracy in the detection of false interpretations.

The detection of fake user comments is an important application of deep learning. Deep learning can perform feature selection and organization of high-dimensional data and dynamically update parameters via feedback.

DEEP LEARNING

Nowadays, deep learning is widely used in natural language learning and analysis of complex semantics. Prior to deep learning models, training data for machine learning was based on manual presentations of features that did not change according to irrelevant changes in the data. The deep learning approach consists of multiple abstraction structures and multiple processing layers combined to learn representations of the data. [4] An important advantage of deep learning architectures is that it uses effective algorithms for controlled / uncontrolled feature learning or hierarchical feature extraction, instead of handcrafted features. [5]

Traditional machine learning algorithms are linear. However, there is a hierarchy model that varies according to the complexity of the field to be applied in deep learning algorithms. The deep learning process repeats until the final success rate reaches a certain level. [6]

There are many different types of deep learning architectures established by increasing the number of layers in artificial neural networks. One of these architectures, Long Short Term Memory (LSTM) networks, is a type of repetitive neural network with a more complex calculation unit, has achieved strong results on various sequence modeling tasks. [7]

LSTM

LSTM architecture can be interpreted as a different version of computer memory. LSTM architectures provide very good results in speech / text processing. TIMIT data set containing rich content in different dialects was used in frame based sound classification study and 70% success rate was obtained. [8]. A further 83% success rate has been achieved on the same dataset with another LSTM approach that has undergone end-to-end training with appropriate adjustments [9] In another study, an accuracy rate of 84.5% was given on the Verbmobil dataset, which contained extensive speech data for keyword detection [10].

LSTM cells resemble RNN (Recurrent Neural Network) cells. However, the input, output, sigmoid (or tangent) and forget gate in this learning algorithm shown in Figure 1 are not present in the RNN algorithm.

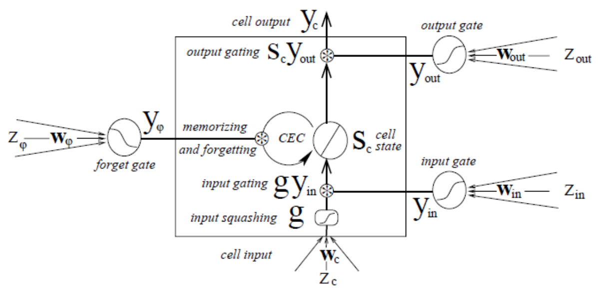

Fig. 1. *LSTM memory block with one cell (rectangle). The so-called CEC maintains the cell state sc,which may be reset by the forget gate. Input and output gate control read and write access to the CEC; g squashes the cell input. See text for details.*

What makes LSTM architechture superior to other architectures is the forget gate in its structure. Forget gate is used to decide how much of the data from the previous cell should be forgotten. With this feature of LSTM, we can see that LSTM remembers important discrepancies to predict the tag of online fake or real comments.

STUDY AND RESULTS

The data set used in our study consists of an authentic and deceptive hotel review for 20 Chicago hotels. The content of the data set is as follows. [11,12,13]

- 400 real positive reviews from TripAdvisor
- 400 deceptive positive comments from Mechanic Turk
- 400 real and negative comments from Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor, Yelp (400).
- 400 deceptive negative comments from Mechanic Turk.

We created a json file to use the data set in our model. If the comment is fake, 1 is defined and if it is real, 0 is defined and we have created the structure in Table 1.

No	Situation	Comment
1341	1	My girlfriends and I stayed 4 nights at the Ta...
144	0	The staff at the Talbott hotel was completely ...
1401	1	We booked our room on Priceline and got a goo...
131	0	Hard Rock Hotel boats that they have the best ...
1440	1	I travelled to Illinois from the UK on busines...
734	1	A disappointment to say the least. I have hear...
1376	1	My wife and I stayed here in the middle of Feb...
770	1	We booked directly with the Intercontinental--...
1015	0	I went to Chicago on business last year and I ...
1488	1	We stayed at this hotel for 4 nights in August...

Table 1. *Dataset list*

We have limited the maximum word length with 400 words to be taken into account in the comments after the text classification for the model to work more effectively on the existing data set.

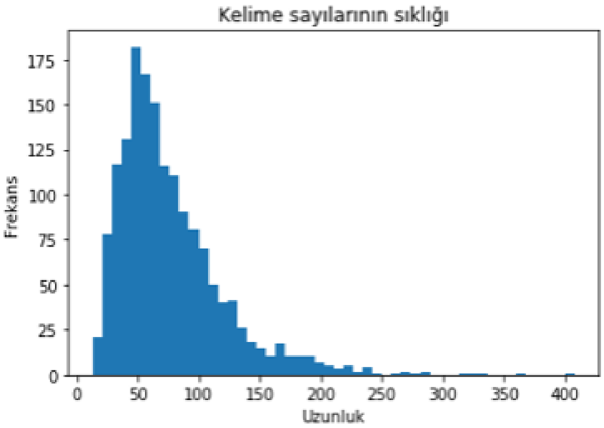

Fig. 2. Word frequency

In our study, we used the Keras library which was written for high level artificial neural networks and could work on Tensorflow. Keras supports both common and recurrent networks and combinations of the two.TensorFlow is an open-source software library for numerical calculation using data flow charts. The flexible architecture allows you to distribute the calculation to one or more CPUs or GPUs on a desktop, server, or mobile device with a single API.The LTSM model that we formed with the stone is given in figure-3.

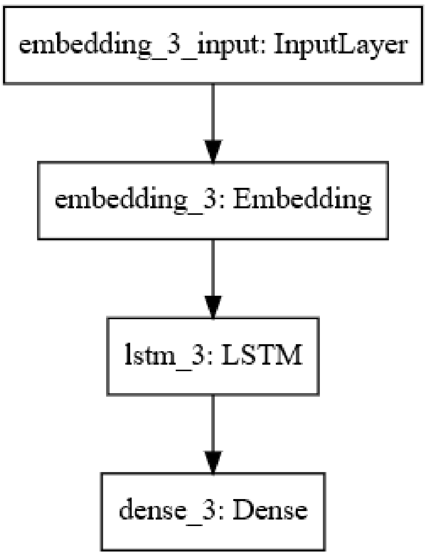

Fig. 3. LSTM Model

The main input of our model is the word sequence matrices obtained from each comment. We used 80% of our data set to train our model. We used the remaining 20% for testing.

The loss and accuracy values obtained from our model after the training are given in the graphs below.

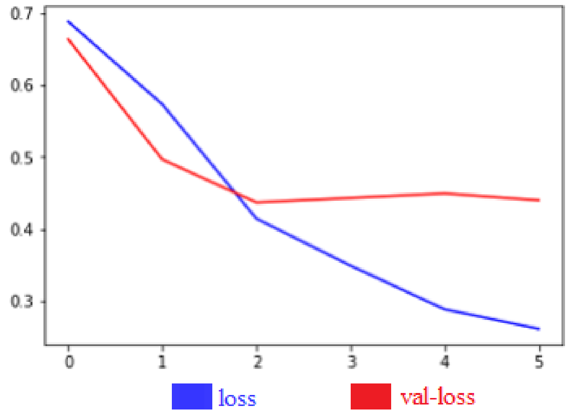

Fig. 4.

The loss value given in Figure 4 is a scalar value that we try to minimize during model training. The lower the loss, the closer our estimates are to actual tags. The problem here is the increase in the possibility of memorization as the number of education increases. In order to control these values, it is necessary to keep the loss value at a certain limit.

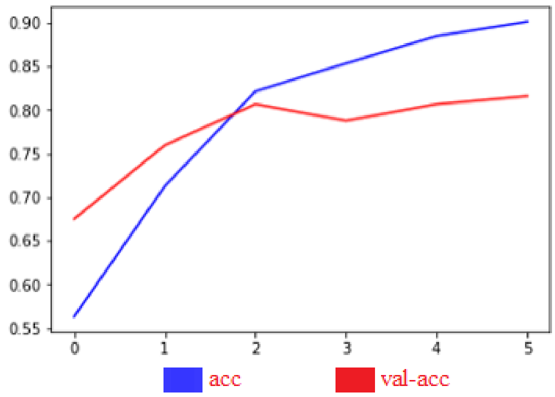

Fig. 5

After training our model 6 times, we reach 90% accuracy rate, but our loss rate is 26%.The Validation Accuracy (val_acc) given in Figure-5 is a measure of how well our model estimates. We have a accuracy rate of 83% after the test.

CONCLUSION

In this study, the LSTM model, which is one of the deep learning architectures, was used to describe online fake comments The analysis of the positive and negative counterfeit interpretations of the data set that we use in the detection of counterfeit interpretations is carried out together.

Our study shows that the proposed model is effective in developing and detecting false interpretations. However, the size of the data set is effective in increasing the determination power of the model.

In our future work, we aim to increase the accuracy rate with larger data sets.

REFERENCE

[1] M.Luca. Reviews, reputation, and revenue: The case of yelp.com. In Working Paper 12-016, Harvard Bus. Sch., 2011.

[2] D. Streitfeld, «The Best Book Reviews Money Can Buy» http://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html. 2018

[3] M. Ott, «Deceptive Opinion Spam Corpus v1.4».http://myleott.com/op_spam/. 2018.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.

[5] H. A. Song and S.-Y. Lee, "Hierarchical Representation Using NMF," in International Conference on Neural Information Processing., 2013, pp. 466–473.

[6] A. Ali Süzen and Kıyas Kayaalp ,"Türkiyede Derin Öğrenme Uygulamaları" 2018.

[7] Kai Sheng Tai, Richard Socher, Christopher D. Manning "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks"-2015.

[8] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," Neural Networks, vol. 18, no. 5–6, pp. 602–610, Jul. 2005.

[9] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645–6649.

[10] S. Fernández, A. Graves, and J. Schmidhuber, "An Application of Recurrent Neural Networks to Discriminative Keyword Spotting," in International Conference on Artificial Neural Networks, 2007, pp. 220–229.

[11] M. Ott, Y. Choi, C. Cardie ve J. T. Hancock, «Finding Deceptive Opinion Spam by Any Stretch of the Imagination» The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1,309-319, ISBN: 978-1-932432-87-9, Stroudsburg, PA, ABD, 2011.

[12] M. Ott, C. Cardie ve J. Hancock, «Estimating the Prevalence of Deception in Online Review Communities» the 21st international conference on World Wide Web, 201-210, ISBN: 978-1-4503-1229-5, New York, NY, ABD, 2012.

[13] M. Ott, C. Cardie ve J. T. Hancock, «Negative Deceptive Opinion Spam» Proceedings of NAACL-HLT 2013, 497–501, Atlanta,Georgia, 2013.