Sentiment Analysis Using Deep Learning Approach

Peng Cen[1], Kexin Zhang[1] and Desheng Zheng[1, *]

Abstract: Deep learning has made a great breakthrough in the field of speech and image recognition. Mature deep learning neural network has completely changed the field of natural language processing (NLP). Due to the enormous amount of data and opinions being produced, shared and transferred everyday across the Internet and other media, sentiment analysis has become one of the most active research fields in natural language processing. This paper introduces three deep learning networks applied in IMDB movie reviews sentiment analysis. Dataset was divided to 50% positive reviews and 50% negative reviews. Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) neural networks are two main types, which are widely used in NLP tasks, while Convolutional Neural Networks (CNN) is often used in image recognition. The results have shown that, CNN network model can achieve good classification effect when applied to sentiment analysis of movie reviews. CNN have reported the accuracy of 88.22%, while RNN and LSTM have reported accuracy of 68.64% and 85.32% respectively.

Keywords: Deep learning, RNN, LSTM, CNN, sentiment analysis.

1 Introduction

Sentiment analysis or opinion mining is the analysis of people's opinions, emotions, and attitudes from written expressions [Liu (2012)]. Nowadays, Individual users have shown great interest in opinions about products and services on the web, and this information largely affects user decisions. In addition to individuals, analyzing consumer sentiment is an important way for companies to be aware of their products and services being perceived [Pang, Bo and Lillian (2008)]. Due to the enormous amount of data and opinions being produced, shared and transferred everyday across the internet and other media, Sentiment analysis has become vital for developing opinion mining systems. The purpose of sentiment analysis is to indicate the views expressed in a particular text and to identify positive and negative opinions in the text, which are particularly important for decision makers' choices and plans.

Artificial intelligence has become an area with many practical applications and active research topics, and is booming. Artificial intelligence systems need to have the ability to acquire knowledge themselves, which is the ability to extract patterns from raw data. This ability is called machine learning [Anitescu, Atroshchenko, Alajlan et al. (2019)]. Deep

---

learning is a kind of machine learning. It is a technology that can make computer systems improve from experience and data. It is one of the ways to lead to artificial intelligence [Li, Zhang, Suk et al. (2014)]. In the development of the past few decades, deep learning has borrowed a lot of knowledge from the human brain, statistics and applied mathematics. In recent years, with the increasing amount of data and the rapid development of computers, there are varieties of big data issues. Then, the practicality of deep learning is gradually improving such as in the fields of speech and image recognition and so on. In addition to these, deep learning is also used in natural language processing, where text sentiment analysis is an important application area.

For this task, this paper uses the deep learning technology to analyze the sentiment of the movie reviews in the IMDB dataset, and divide them into negative and positive categories. It is necessary to classify movie reviews. For researchers, classification can be based on the relevance of the sentiment and rating of the film reviews. For the user, it is a recommendation tool for movie selection, which helps the user make a choice. For film companies, this information can be used for marketing decisions and finding customers [Han and Kim (2017)].This paper is introducing three neural network models (CNN, RNN, LSTM) and compares the results of these three models with the result of SVM [Pang, Lee and Vaithyanathan (2002)] and RNTN [Socher, Perelygin, Wu et al. (2013)].

2 Related work

Sentiment analysis is a subject of in-depth study in natural language processing. We reference some relevant papers to compare the results and discuss what we learned and applied from them.

In a published sentiment analysis work, SVM performed better classification with 82.9% accuracy compared with the Naive Bayes that achieved 81% on positive and negative movie reviews from the IMDB movie reviews dataset that consists of 752 negative and 1301 positive reviews [Pang, Lee and Vaithyanathan (2002)].In another study, Recursive Neural Tensor Network(RNTN) were used to identify sentences as positive or negative. They used a dataset based on 11,855 movie reviews. RNTN achieved an accuracy of 80.7% in sentiment prediction of all phrases [Socher, Perelygin, Wu et al. (2013)]. In [Janssen (2001)], it is proposed that a text cannot be analyzed in isolation. The meaning of a sentence is closely related to the words before and after it. In Lin et al. [Lin, Horne, Tino et al. (1996)], it is stated that Recurrent Neural Networks (RNN) model can deal with the short-term dependence in a sequence of data, but there is a problem of gradient explosion in dealing with the long-term dependence. These long-term dependencies have a significant impact on the meaning and overall polarity of the document. In order to solve this long-term dependence problem, the LSTM model is proposed. The researchers have investigated CNN and basic RNN for relation classification in Vu et al. [Vu, Adel, Gupta et al. (2016)]. They report higher performance of CNN than RNN and give evidence that CNN and RNN provide complementary information: while the RNN computes a weighted combination of all words in the sentence, the CNN extracts the most informative ngrams for the relation and only considers their resulting activations. Dragoni et al. [Dragoni, Poria and Cambria (2018)] presented OntoSenticNet, a commonsense ontology for sentiment analysis based on SenticNet, a semantic network of

100,000 concepts based on conceptual primitives. Multimodal sentiment analysis is a very actively growing field of research. Majumder et al. [Majumder, Hazarika, Gelbukh et al. (2018)] presented a novel feature fusion strategy that proceeds in a hierarchical fashion, first fusing the modalities two in two and only then fusing all three modalities. On multimodal sentiment analysis of individual utterances, this strategy outperforms conventional concatenation of features by 1%, which amounts to 5% reduction in error rate. Cambria et al. [Cambria, Poria, Hazarika et al. (2018)] coupled sub-symbolic and symbolic AI to automatically discover conceptual primitives from text and link them to commonsense concepts and named entities in a new three-level knowledge representation for sentiment analysis. In particular, they employ recurrent neural networks to infer primitives by lexical substitution and use them for grounding common and commonsense knowledge by means of multi-dimensional scaling.

In some other papers, Preethi, G. et al. [Preethi, Krishna, Obaidat et al. (2017)] explored a new application of Recursive Neural Networks (RNN) with deep learning system for sentiment analysis of reviews. The proposed RNN-based Deep-learning Sentiment Analysis (RDSA) recommends the places that are near to the user's current location by analyzing the different reviews and consequently computing the score grounded on it. Qian et al. [Qian, Huang, Lei et al. (2016)] proposed simple models trained with sentence-level annotation, but also attempt to generating linguistically coherent representations by employing regularizers that model the linguistic role of sentiment lexicons, negation words, and intensity words. Results show that their models are effective to capture the sentiment shifting effect of sentiment, negation, and intensity words, while still obtain competitive results without sacrificing the models' simplicity. Nguyen et al. [Nguyen, Shirai and Velcin (2015)] built a model to predict stock price movement using the sentiment from social media. this paper shows an evaluation of the effectiveness of the sentiment analysis in the stock prediction task via a large scale experiment. adopted the neural network method to demonstrate that the machine learning method is useful for studying the differences and commonalities of different quantizing methods of quantum correlation [Li, Zhu, Zhu et al. (2019)]. They found the link between the geometric and discord for special canonical initial states based on neural network method [Li, Zhu, Zhu et al. (2019)]. Deep learning can be used to find relevance, which is a good inspiration for sentiment analysis.

3 Network structure

In the above, we introduced the development prospect and related work of deep learning. CNN, RNN and LSTM are the most widely used network structures in deep learning. They are applied in many fields such as text classification and speech recognition. Next, then the principle of CNN and LSTM will be introduced.

3.1 CNN

Convolutional neural network is a feedforward neural network, and it is also the most mature field of deep learning algorithm application [Wang, He, Sun et al. (2019)]. It has strong characterization learning ability [Dragoni, Poria and Cambria (2018)]. The network structure consists of three layers, a convolutional layer, a pooling layer, and a

fully connected layer. As shown in Fig. 1. First, the convolutional neural network convolves the input word vector sequence, generates a feature map, and then uses max pooling on the feature map to get the feature corresponding to the kernel [Long and Zeng (2019)]. Finally, the splicing of all features is a fixed-length vector representation of the text. In practical applications, we use multiple convolution kernels to process sentences and stack convolution kernels with the same window size into a matrix, which can perform operations more efficiently [Hassan and Mahmood (2018); Bawden, Sennrich, Birch et al. (2017)].
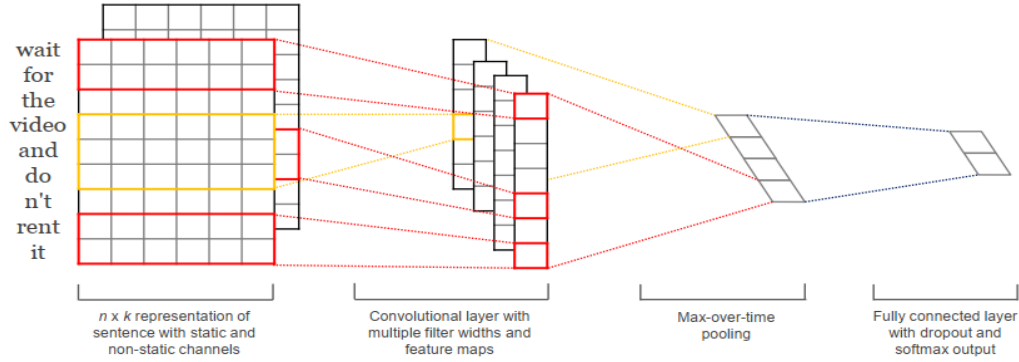

Figure 1: CNN network structure [Kim (2014)]

Eq. (1) represents the operation of each convolution layer in the forward propagation process, where conv represents the convolution operation function, W is the weight matrix corresponding to each layer, X is the current input vector, and b is the per-layer paranoid.

$$V = \mathrm{conv}(W, X) + b \tag{1}$$

The prediction probability of all categories is calculated by the Eq. (2).

$$Y = \phi(V) \tag{2}$$

Eq. (3) represents the mean square error function, where T is the actual value and P is the predicted value, and the result can be used to analyze the training level of the neural network.

$$E = \sum \frac{1}{2}(T - P)^2 \tag{3}$$

Eq. (4) represents the chained derivation rule, which is used to update the weight matrix during the back propagation of the neural network. This is very important for optimizing the neural network.

$$\frac{d_z}{d_t} = \frac{\partial_z}{\partial u} \frac{d_u}{d_t} + \frac{\partial_z}{\partial_v} \frac{d_v}{d_t} \tag{4}$$

3.2 RNN

Recurrent Neural Network (RNN) is a Recurrent Neural Network that takes sequence data as input, recursively recurses in the evolution direction of the sequence, and all nodes (cyclic elements) are linked by chain. Loop neural network has Shared memory, parameter and Turing complete (Turing completeness), so study on the nonlinear characteristic of the sequence has certain advantage. Cyclic neural network is applied in Natural Language Processing (NLP), such as speech recognition, Language modeling, machine translation and other fields, and is also used in various time series forecasting. A Convolutional Neural Network (CNN) is introduced to deal with computer vision problems involving sequence input.

3.3 LSTM

Before introducing LSTM, the recurrent neural network should be introduced. Because LSTM is developed from RNN. In recent years, the recurrent neural network has been widely used in the field of natural language processing such as language analysis and text classification [Sak, Senior and Beaufays (2014)]. The schematic diagram of the cycle structure is shown in Fig. 2.
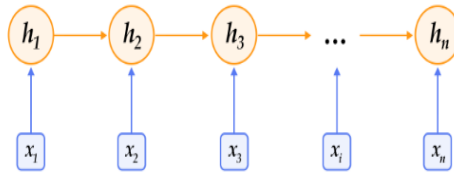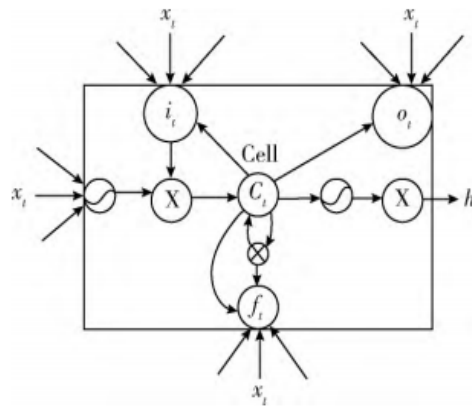

Figure 2: RNN network structure

Figure 3: Long Short-term Memory Cell [Kim (2014)]

Assume that at the n-th moment, the current input of the neural network is x_n, and the hidden value of the previous time is h_{n-1}. The current hidden value h_n is calculated by the Eq. (5). Where W_{nh} is the matrix parameter input to the hidden layer, W_{hh} is the matrix parameter of the hidden layer to the hidden layer, b_n is the bias vector parameter of the hidden layer, and σ is the sigmoid function.

$$\mathrm{h}_n = f(x_n, h_{n-1}) = \sigma(W_{nh_{xn}} + W_{hh}h_{n-1} + b_n) \tag{5}$$

In order to solve the problem in processing long sequence data and the phenomenon of gradient disappearance and gradient explosion, Hochreiter and Schmidhuber proposed LSTM (long short term memory) in 1997 [Chen and Wang (2017)].

LSTM is a time-cycle neural network, an advanced version of RNN, which is now more widely used in industry [Miedema (2018)], as shown in Fig.3. This is the model structure of the LSTM. It contains LSTM's unique "door" structure. The i_t, f_t, c_t, o_t in the Fig. 3 represents the input gate, the forgetting gate, the cell memory unit, and the output gate.

The input gate determines how much of the previous sample is stored in memory. The output gate regulates the amount of data transferred to the next layer, the forgetting gate controls the rate of loss of stored memory and determines what information is discarded from the state of the cell. LSTM's weight is changing through the input gate, the forgetting gate and the output gate, thus avoiding the problem of gradient disappearance or gradient explosion. Input gate, forgetting gate, output gate and state cell are calculated by the following Eq. (6) to Eq. (10).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{cj}c_{t-1} + b_i) \tag{6}$$

$$\widetilde{C}_t = f_t \Theta c_{t-1} + i_t \Theta \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{7}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{8}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f \tag{9}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_0 \tag{10}$$

$$h_t = o_t \Theta \tanh(c_t) \tag{11}$$

3.4 Model parameters

In this paper, CNN model has 128 input neurons, 2 convolution layers with peer layers 512 neurons, and finally 1 output layers. LSTM network model has 128 LSTM block and output layers with 2 neurons. This paper use tanh function as activation function. Tanh's effect will be very good when the characteristic difference is obvious, and the characteristic effect will be continuously expanded during the cycle. As shown in Eq. (12).

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{12}$$

Due to Adagrad is an algorithm that adaptively assigns different learning rates to each parameter, Adagrade works well with deep learning models. This paper use Adagrad function as CNN and LSTMs' optimization function. As shown in Eq. (13).

$$w^{t+1} = w^t - \frac{\eta}{\sqrt{\sum_{i=0}^{t}(g^i)^2}} \tag{13}$$

Learning rate and steps are important parameters that affect the output result. We start

with 0.001 and gradually increase it to 0.1. Finally, there is a good result when learning rate is 0.001 and steps is 200.

4 Experiment

In this paper, we use Baidu's PaddlePaddle deep learning framework to conduct experiments. It is a very popular deep learning framework [Akimov, Kosivets and Volkanin (2017)]. We did three experiments including RNN, LSTM, CNN to try to come up with a better model for dealing with movie review analysis. The experimental environment is shown in Tab. 1 below.

Table 1: Development environment specification

Experimental environment	
Memory	8 GB
Processor	Intel Core i5
Development tool	Python 3.6
Used Libraries	PaddlePaddle

4.1 Data set

To demonstrate which model has a better effect on sentiment analysis, we used the public IMDB dataset. It contains 50,000 reviews from the online movie database, of which 25,000 are used for training and another 25,000 are for testing. In order to ensure better fairness, negative and optimistic comments each account for 50% [Tang, Qin and Liu (2015)]. As shown in Tab. 2.

Table 2: IMDB reviews

Train data		Test data	
Neg	12500	Neg	12500
Pos	12500	Pos	12500

After confirming that the IMDB dataset is used, we have some pre-processing on the IMDB dataset. We will use the html,',', '.' in the file. The punctuation marks are removed to achieve interference-free text that is not yet recognized by the computer. So we also need to convert our text into a one-dimensional vector [Shen, Wang, Wang et al. (2018)] by word2vev.

4.2 Experimental results

We did experiments based on the network model and data set above. The three models reached a steady state after 180 steps. The accuracy of each model is shown in the Fig. 4, and the loss rate is shown in the Fig. 5. The experimental results show that CNN achieved the highest accuracy and the lowest accuracy in the experiment. The accuracy rate is over 88.22% and the loss value is kept at 0.3. The experimental results show that CNN has a

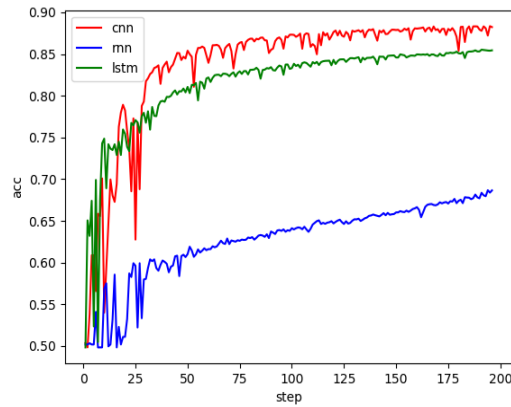good effect on the problem of text classification.

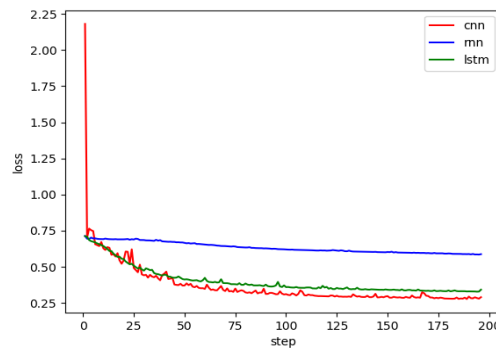**Figure 4:** Network model accuracy comparison

Figure 5: Network model loss rate comparison

The following Tab. 3 describes the accuracy of the previous people's emotional comment analysis and the accuracy of our experiment in this paper. It can be seen from the Tab. 3 that CNN and LSTM are better than SVM and RNTN. Among them, CNN is 88.22%. Therefore, CNN also has a good effect in the field of text classification.

Table 3: Comparison of five different network models

Network model	Accuracy rate	Dataset
SVM [Pang, Lee and Vaithyanathan (2002)]	82.90%	2035 Review Files
RNTN [Socher, Perelygin, Wu et al. (2013)]	80.70%	11855 Sentence
RNN	68.64%	IMDB text
CNN	88.22%	IMDB text

5 Conclusion

In this project, we used three deep learning networks (CNN, RNN, LSTM) to classify movie reviews from the IMDB dataset into positive and negative categories. Models that are often used for sentiment analysis are RNN and LSTM, CNN is often used in image recognition. Through experiment, We found that the CNN neural network still has good performance for processing a sequence of data, achieving an accuracy of 88.22%.In addition, the achieved accuracies have been compared to accuracies reported by previously published works that have used other machine learning techniques, and the result showed that the proposed deep learning techniques (CNN, RNN, LSTM) have outperformed SVM, RNTN. For future work, we hope to create our own integration through the method of superposition model, set up better models in the experiment, and improve the effect of movie review or other emotional analysis. We also know that the data preprocessing greatly affects the model recognition and feature extraction. Therefore, it is the focus of our future work to find better data preprocessing methods.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Akimov, A.; Kosivets, R.; Volkanin, A. (2017): Traffic forecasting using paddlepaddle. *CEUR Workshop Proceedings*, vol. 1990, pp. 102-111.

Anitescu, C.; Atroshchenko, E.; Alajlan, N.; Rabczuk, T. (2019): Artificial neural network methods for the solution of second order boundary value problems. *Computers, Materials & Continua*, vol. 59, no. 1, pp. 345-359.

Bawden, R.; Sennrich, R.; Birch, A.; Haddow, B. (2017): Evaluating discourse phenomena in neural machine translation. *arXiv preprint arXiv:1711.00513.*

Cambria, E.; Poria, S.; Hazarika, D.; Kwok, K. (2018): Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. *Thirty Second AAAI Conference on Artificial Intelligence.*

Chen, J.; Wang, D. (2017): Long short-term memory for speaker generalization in supervised speech separation. *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4705-4714.

Dragoni, M.; Poria, S.; Cambria, E. (2018): Ontosenticnet: a commonsense ontology for sentiment analysis. *IEEE Intelligent Systems*, vol. 33, no. 3, pp. 77-85.

Han, Y.; Kim, K. K. (2017): Sentiment analysis on social media using morphological sentence pattern model. *In 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 79-84

Hassan, A.; Mahmood, A. (2018): Convolutional recurrent deep learning model for sentence classification. *IEEE Access,* vol. 6, pp. 13949-13957.

Janssen, T. M. (2001): Frege, contextuality and compositionality. *Journal of Logic, Language and Information*, vol. 10, no. 1, pp. 115-136.

Kim, Y. (2014): Convolutional neural networks for sentence classification. a*rXiv preprint arXiv:1408.5882.*

Li, R.; Zhang, W.; Suk, H. I.; Wang, L.; Li, J. et al. (2014): Deep learning based imaging data completion for improved brain disease diagnosis. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 305-312.

Li, X.; Zhu, Q.; Meng, Q.; You, C.; Zhu, M. et al. (2019): Researching the link between the geometric and rènyi discord for special canonical initial states based on neural network method. *Computers, Materials Continua,* vol. 60, no. 3, pp. 1087-1095.

Li, X. Y.; Zhu, Q. S.; Zhu, M. Z.; Huang, Y. M.; Wu, H. et al. (2019): Machine learning study of the relationship between the geometric and entropy discord. *Europhysics Letters*, vol. 127, no. 2, pp. 20009.

Lin, T.; Horne, B.G.; Tino, P.; Giles, C.L. (1996): Learning long-term dependencies in narx recurrent neural networks. *IEEE Transactions on Neural Networks*, vol.7, no. 6, pp. 1329-1338.

Liu, B. (2012): Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1-167.

Long, M.; Zeng, Y. (2019): Detecting iris liveness with batch normalized convolutional neural network. *Computers, Materials & Continua*, vol. 58, no. 2, pp. 493-504.

Majumder, N.; Hazarika, D.; Gelbukh, A.; Cambria, E.; Poria, S. (2018): Multimodal sentiment analysis using hierarchical fusion with context modeling. *Knowledge Based Systems*, vol. 161, pp. 124-133.

Miedema, F. (2018): *Sentiment Analysis with Long Short-Term Memory Networks.* Vrije Universiteit Amsterdam.

Nguyen, T. H.; Shirai, K.; Velcin, J. (2015): Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, vol. 42, no. 24, pp. 9603-9611.

Wang, N.; He, M.; Sun, J.; Wang, H.; Zhou, L. et al. (2019): IA-PNCC: noise processing method for underwater target recognition convolutional neural network. *Computers, Materials & Continua*, vol. 58, no. 1, pp. 169-181.

Pang, B.; Lee, L. (2008): Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval,* vol.2, no.1-2, pp. 1-135.

Pang, B.; Lee, L.; Vaithyanathan, S. (2002): Thumbs up? sentiment classification using machine learning techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10, pp. 79-86.

Preethi, G.; Krishna, P. V.; Obaidat, M. S.; Saritha, V.; Yenduri, S. (2017): Application of deep learning to sentiment analysis for recommender system on cloud. *International Conference on Computer, Information and Telecommunication Systems*, pp. 93–97.

Qian, Q.; Huang, M.; Lei, J.; Zhu, X. (2016): Linguistically regularized LSTMs for sentiment classification. *arXiv preprint arXiv:1611.03949.*

Sak, H.; Senior, A.; Beaufays, F. (2014): Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *arXiv preprint arXiv:1402.1128.*

Shen, D.; Wang, G.; Wang, W.; Min, M. R.; Su, Q. et al. (2018): Baseline needs more love: on simple word-embedding-based models and associated pooling mechanisms. *arXiv preprint arXiv:1805.09843.*

Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C. D. et al. (2013): Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631-1642.

Tang, D.; Qin, B.; Liu, T. (2015): Document modeling with gated recurrent neural network for sentiment classification. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1422–1432.

Vu, N. T.; Adel, H.; Gupta, P.; Schütze, H. (2016): Combining recurrent and convolutional neural networks for relation classification. *arXiv preprint arXiv:1605.07333.*