SENTIMENT ANALYSIS FOR MOVIES REVIEWS DATASET USING DEEP LEARNING MODELS

Nehal Mohamed Ali, Marwa Mostafa Abd El Hamid and Aliaa Youssif

Faculty of Computer Science, Arab Academy for Science Technology and Maritime, Cairo, Egypt

ABSTRACT

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. This paper introduces a developed classification sentiment analysis using deep learning networks and introduces comparative results of different deep learning networks. Multilayer Perceptron (MLP) was developed as a baseline for other networks results. Long short-term memory (LSTM) recurrent neural network, Convolutional Neural Network (CNN) in addition to a hybrid model of LSTM and CNN were developed and applied on IMDB dataset consists of 50K movies reviews files. Dataset was divided to 50% positive reviews and 50% negative reviews. The data was initially pre-processed using Word2Vec and word embedding was applied accordingly. The results have shown that, the hybrid CNN_LSTM model have outperformed the MLP and singular CNN and LSTM networks. CNN_LSTM have reported the accuracy of 89.2% while CNN has given accuracy of 87.7%, while MLP and LSTM have reported accuracy of 86.74% and 86.64 respectively. Moreover, the results have elaborated that the proposed deep learning models have also outperformed SVM, Naïve Bayes and RNTN that were published in other works using English datasets.

KEYWORDS

Deep learning, LSTM, CNN, Sentiment Analysis, Movies Reviews, Binary Classification

1. INTRODUCTION

By definition, sentiment analysis or opinion mining is the use of text analysis, computational linguistics, or Natural Language processing (NLP) in order to get semantic quantification of the studied information[1].

Sentiment analysis aims to indicate the opinion of a specific text (*e.g. a tweet, or a product review*). These indications are used accordingly by decision makers in planning and taking appropriate actions such as marketing decisions, customers hunting or business expansion in specific geographic region.

Due to the massive data evolution and the amount of data being exchanged and produced every second, the urge to comprehend, mine and analyse this data has remarkably increased. And since the regular machine learning techniques and Neural Networks were not sufficient to be obtained on this big data, deep learning was the key of the big data era[2].

Deep learning is a subfield of machine learning and an alternation of neural networks. That is, regular neural network is a single network with an input and output layers in addition to hidden layers in between, were computation is done. Where Deep Neural Networks are, basically, consist of multiple neural networks where the output of one network is an input to the next network and so forth. This concept has overcome the limitation of the number of hidden layers in Neural Networks and made working with big data more feasible[3].

Deep learning networks learns the features on its own[4], that is, it has become apparent as a robust machine learning technique that learns multiple layers of features of the data and induces results of prediction. Deep learning has been recently used in various applications in the field of signal and information processing, especially with the evolution of big data[5][6]. In addition, deep learning networks have been used in sentiment analysis and opinion mining.

This paper has applied sentiment analysis over a dataset of English movies reviews (*IMDB dataset*) using deep learning techniques in order to classify this dataset files into positive and negative reviews.

Since sentiment analysis among English movies reviews has been applied using non-deep learning techniques, which are SVM and Naïve Bayes and using Recurrent Neural Networks[7][8], this paper is introducing four deep neural networks (MLP, CNN, LSTM in addition to a hybrid model CNN_LSTM) and compares the results of these four models with the results of SVM, Naïve Bayes [9]. Moreover, reported results of the proposed models have been compared to the results reported by applying RNTN to an English movies reviews dataset[8].

2. RELATED WORKS

In a published sentiment analysis work the authors have classified online product reviews into positive and negative classes. This paper has applied different machine learning algorithms in order to experimentally evaluate various trade-offs, using approximately 100K product reviews from the web. Three classifiers have been applied (Passive-Aggressive (PA) Algorithm Based Classifier, Language Modelling (LM) Based Classifier, and Winnow Classifier). In addition, n-grams was used for linguistic features extraction[10].

The results have illustrated that when a high order n-gram as features is used with a significant classifier, a comparable result can be achieved, or higher performance than that reported in academic papers can be achieved.

A second work have reported that 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 (imdb.com) movie reviews dataset that consists of 752 negative and 1301 positive reviews [9].

The researchers of a third paper have proposed Recursive Neural Tensor Network (RNTN) model for identifying sentences as positive or negative by using fully labelled parse trees. They used a dataset based on a corpus of 11,855 movie reviews. The RNTN have obtained 80.7% accuracy on fine-grained sentiment prediction across all phrases and captures negation of different sentiments and scope more accurately than previous models[8].

A fourth paper has focused on customer reviews. Products reviews were collected from Amazon.com (11,754 sentences). It proposed a novel deep learning framework named Weakly-supervised Deep Embedding for reviewing sentence sentiment classification with accuracy 87% [11].

Lastly, a work that was published in 2018 performed several deep learning models for a binary sentiment classification problem. They used movie reviews in Turkish from the website www. beyazperde.com to train and test the deep learning models[12]. The whole dataset consists of 44,617 samples including positive and negative reviews. 4000 samples from the dataset were used for testing the models. Two major deep learning architectures were applied, CNN and LSTM. Word embedding were created by applying the word2vec algorithm with a skip-gram model on the used dataset. Experimental results have shown that the use of word embedding with deep neural networks effectively yields performance improvements in terms of run time and accuracy.

3. PROPOSED WORK

As previously mentioned, in this paper 4 deep learning techniques were implemented MLP, CNN, LSTM and SNN_LSTM; The proposed CNN_LSTM is illustrated in Figure 1. As shown, the first step is loading the dataset then applying pre-processing procedures, the result would be processed by a convolutional layer, Maxpooling has been applied afterwards; after that LSTM layer has processed the data and finally the output of classified reviews is produced. Each step of this workflow is elaborated in details in the following subsections. This workflow represents only the workflow of the hybrid CNN_LSTM model, other methods were also applied (MLP, CNN and LSTM) so the number of processing layers may vary according to the applied methodology.
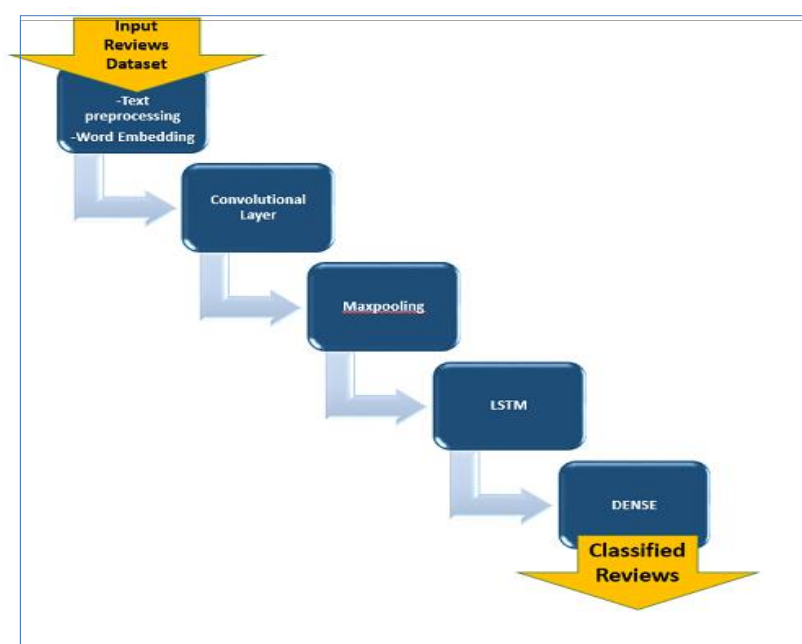

Figure 1. Work flow of the proposed Model CNN_LSTM

3.1 Pre-processing

In this paper, there are more than one method used together in order to extract significant features from the review text files that could be used for training purposes. These features were then used to train the network.

- Tokenization: Initially, data in all documents have been split to formulate a huge vector of words.

21

- Normalization[13]: The data has been normalized before further processing, (i.e. punctuation removal, converting all text into Lowercase and substituting numbers by their word equivalents).

- Word Embedding [14]: word2vec embedding provided by Keras library was applied with vector size of 32. Hence, each word was converted to its corresponding numerical vector to be understood by the network. These vectors are learned as the model trains. The vectors add a dimension to the output array. Hence, (batch, sequence and embedding) are the three resulting dimensions.

- Sequencing [14]: Data has to be passed to any neural network as a binary vector; thus, sequencing is essential in order to turn the numeric array resulting from the word2Vec embedding into a digital vector that would be passed afterwards to the built classification model.

Finally, the maximum review length was caped to 500 words. While, reviews that are longer than that were truncated and documents shorter than that were padded with zero values.

3.2 Convolutional Neural Network (CNN) Layer

Convolutional Neural Networks are like typical Neural Networks, these networks are consisting of neurons that own learnable weights and biases. Some input was received by each neuron; a dot product was performed. CNN consists of one or more convolutional layers after that its followed by fully connected layers like the ordinary multilayer neural network[15].

CNN works using three key concepts (local receptive fields, shared weight and biases, and activation and pooling)[16]. A small region of neurons in the input layer are connected to hidden layer neurons. These small regions are called local receptive fields. The network has neurons with weights and biases. During the training process, the model learns the weight and biases values however these values are the same for neurons in the hidden layer. The step of activation function applies the conversion to the output of each neuron by using Rectified Linear Unit (RLU). RLU is a commonly used activation function. The function of pooling step is to condense the output of the convolutional layer by half by reducing the dimensionality of the features map. Hence, CNN can be used for learning structure in paragraphs of words[17].

3.3 Maxpooling Layer

To enhance and reduce the results of the convolutional layer, Maxpooling mask has been applied[18]. This is done by applying that mask sequentially on the entire data and each time the mask is applied it selects the value with the highest weight only and ignores the rest values which significantly reduces the input to the next layer.

Thus, we have implemented one dimensional CNN with 32 filters provided by Keras library. Based on the given input 500 words vector the output of CNN layer was 500 and then the output was reduced by 50%, producing 250 words vectors after applying the maxpooling layer.

3.4 Long Short-Term Memory (LSTM)

LSTMs are used in the field of deep learning and considered as an artificial recurrent neural network architecture[19]. LSTM consists of (cell, input gate, output gate and a forget gate). Data can be stored or written or read from the cell like information in the memory of a computer [20].

22

The cell takes decisions about what to read or write or erase via opened and closed gates. These gates work on the signals that they receive, and similar to the nodes of NN, they pass or block the data due to its strength or weakness.

For inputs elements of $\{X^{(0)}, \ldots, \{X^{(T)}\}$ Where, $x^{(t)} \in R^{(m)}$ and $h^{(t)} \in R^{(n)}$ as the hidden layer of the Recurrent Neural Networks for time t.

V and W are considered as shared weight matrices and f is considered as a nonlinear activation function.

$$h^{(t)} = f\big(Wh^{(t-1)} + Vx^{(t)}\big) \tag{1}$$

[21] In the proposed CNN_LSTM model, LSTM layer was applied on the output resulting from the maxpooling layer.

3.5 Output Dense

Lastly, the last layer is connected as a dense to a single output node, since the proposed model is handling binary classification problem, sigmoid function has been applied on the e data resulting from LSTM layer in order to produce a fraction value that lies between 0 and 1, and according to this value a review is classified (i.e. 0 for a negative review and 1 for a positive review).

4. EXPERIMENTAL WORK

4.1. Dataset

The used IMDB dataset consists of 50K movie reviews files(25K positive reviews files and 25K negative review files) reviews, all written in English.

Files sizes has ranged from 1kb-15kb. No rating information was included in the text files. Dataset was split to 80% training set and 20% testing set.

4.2. Development Environment

Table 1 summarizes the specifications of the environment used to build and run the proposed model.

Table 1. Development Environment specification

OS	Windwos 10.1
Memory	12.0 GB
Processor	Intel Core i7
Dataset storage location	OS File system
Development tool	Python 3.6
Used Libraries	Keras and Tensor flow

4.3. Data Processing

The proposed model CNN_LSTM has been implemented according to the work flow illustrated in Figure1 in addition to MLP, CNN and LSTM models.

Models has been implemented using Python 3.6 by means of Tensor flow and Keras libraries, each model was implemented, trained and tested separately and results were studied afterwards. Each deep network was trained on 80% of the dataset files and tested using 20% of the dataset. Highest accuracy reported of each model was recorded accordingly, as elaborated in results section.

Number of Epochs has been gradually tuned and then set to 100 epochs, no accuracy enhancement has been reported after exceeding approximately 45 epochs for all implemented models.

5. RESULTS AND DISCUSSION

During the training process, figures 2 and 3illustrate the loss and the accuracy curves of the proposed hybrid CNN_LSTM respectively. Loss curve shows that loss has decreased from 42% at the first epoch to less than 1% by reaching epoch 45.

On the other hand, the accuracy curve illustrates the increase from approximately 60% to 99% after exceeding the 45th epoch.
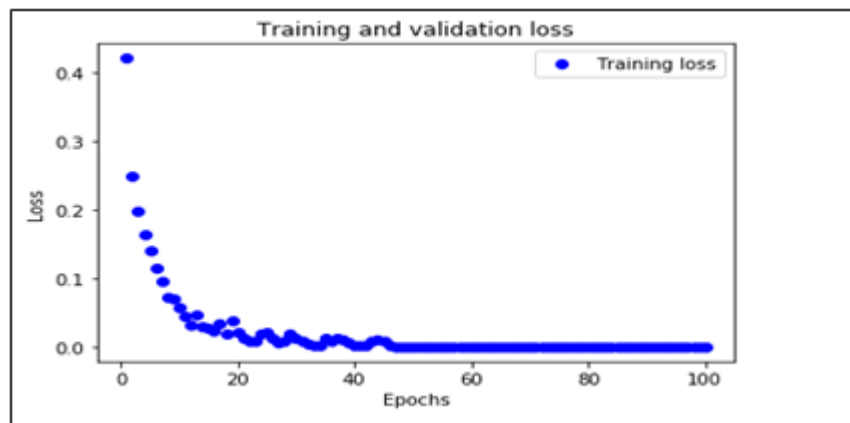

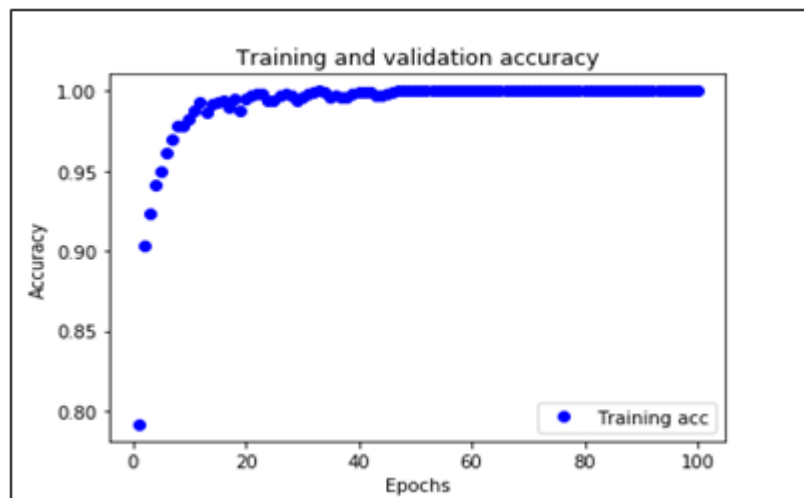Figure 2. Loss curve across the training process over 100 epochs

Figure 3. Accuracy curve of training process over 100 epochs

Moreover, Table 2 illustrates the previously reported results and their referenced papers in addition to the reported accuracies of the proposed deep learning models. While figure 4 summarizes the results achieved by the proposed models in addition to the previously published papers. [7] has applied NB and SVM on 752 negative and 1301 positive reviews with total of 2053, lastly, [8] has used another English movies reviews dataset by Pang and Lee (2005) and consists of 11,855 single sentences extracted from movie reviews.

As shown, the hybrid of CNN_LSTM has outperformed all other deep learning techniques and has also out performed SVM, RNTN, and NB, and has achieved the highest accuracy of 89.2%. This result indicates promising results, considering the significantly higher number of processed reviews, with promising higher accuracy if combined with further enhanced word embedding.

As elaborated in table 2 and figure 4, the reported results proves that all implemented deep learning models in this paper have outperformed the previously reported results using SNM, Naïve Bayes (NB), RNTN.

Table 2. Reported accuracies using the proposed deep learning models Vs reported accuracies that were in published work along with the number of processed reviews.

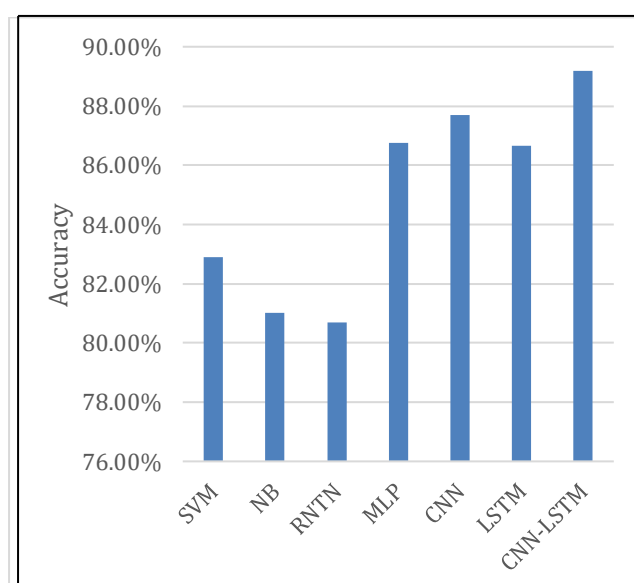| Previous work Models On English Movies reviews Dataset			Proposed Models (50K review files)	
SVM[9]	2035 review files	82.90%	MLP	86.74%
NB[9]		81%	CNN	87.70%
RNTN[8]	11,855 sentences	80.70%	LSTM	86.64%
			CNN-LSTM	89.20%

Figure 4. Reported Accuracies using the proposed deep learning models (MLP,CNN, LSTM and CNN_LSTM) in addition to accuracies of other models reported in previously published works (SVM, NB and RNTN)

6. CONCLUSIONS

This paper has implemented sentiment analysis classifier for IMDB dataset of 50K movies reviews using 3 deep learning networks (MLP, CNN and LSTM) in addition to a hybrid network CNN_LSTM. Word2vector technique was used for words embedding.

The results have shown that CNN_LSTM has outperformed all other implemented deep learning techniques. In addition, the achieved accuracies have been compared to accuracies reported by previously published works that have used other machine learning techniques, and have used English movies reviews dataset; and the result showed that the proposed deep learning techniques (MLP, CNN, LSTM, and CNN_LSTM) have outperformed SVM, Naïve Bayes, RNTN.

REFERENCES

[1]   S. Poria And A. Gelbukh, "Aspect Extraction For Opinion Mining With A Deep Convolutional Neural Network," Knowledge-Based Syst., Vol. 108, Pp. 42–49, Sep. 2016.

[2]   K. Kim, M. E. Aminanto, And H. C. Tanuwidjaja, "Deep Learning," Springer, Singapore, 2018, Pp. 27–34.

[3]   J. Einolander, "Deeper Customer Insight From Nps-Questionnaires With Text Mining - Comparison Of Machine, Representation And Deep Learning Models In Finnish Language Sentiment Classification," 2019.

[4]   P. Chitkara, A. Modi, P. Avvaru, S. Janghorbani, And M. Kapadia, "Topic Spotting Using Hierarchical Networks With Self Attention," Apr. 2019.

[5]   F. Ortega Gallego, "Aspect-Based Sentiment Analysis: A Scalable System, A Condition Miner, And An Evaluation Dataset.," Mar. 2019.

[6]   M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, And E. Muharemagic, "Deep Learning Applications And Challenges In Big Data Analytics," J. Big Data, Vol. 2, No. 1, P. 1, Dec. 2015.

[7]   B. Pang, L. Lee, And S. Vaithyanathan, "Thumbs Up?," In Proceedings Of The Acl-02 Conference On Empirical Methods In Natural Language Processing - Emnlp '02, 2002, Vol. 10, Pp. 79–86.

[8]   A. Y. N. And C. P. Richard Socher, Alex Perelygin, Jean Y.Wu, Jason Chuang, Christopher D. Manning, "Recursive Deep Models For Semantic Compositionality Over A Sentiment Treebank," Plos One, 2013.

[9]   B. Pang, L. Lee, And S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques."

[10]   H. Cui, V. Mittal, And M. Datar, "Comparative Experiments On Sentiment Classification For Online Product Reviews," In Aaai'06 Proceedings Of The 21st National Conference On Artificial Intelligence, 2006.

[11]   Z. Guan, L. Chen, W. Zhao, Y. Zheng, S. Tan, And D. Cai, "Weakly-Supervised Deep Learning For Customer Review Sentiment Classification," In Ijcai International Joint Conference On Artificial Intelligence, 2016.

[12]   B. Ay Karakuş, M. Talo, İ. R. Hallaç, And G. Aydin, "Evaluating Deep Learning Models For Sentiment Classification," Concurr. Comput. Pract. Exp., Vol. 30, No. 21, P. E4783, Nov. 2018.

[13]   M. V. Mäntylä, D. Graziotin, And M. Kuutila, "The Evolution Of Sentiment Analysis—A Review Of Research Topics, Venues, And Top Cited Papers," Computer Science Review. 2018.

[14]   Y. Goldberg And O. Levy, "Word2vec Explained: Deriving Mikolov Et Al.'S Negative-Sampling Word-Embedding Method," Feb. 2014.

[15]   D. Ciresan, U. Meier, And J. Schmidhuber, "Multi-Column Deep Neural Networks For Image Classification," In 2012 Ieee Conference On Computer Vision And Pattern Recognition, 2012, Pp. 3642–3649.

[16]   Y. Kim, "Convolutional Neural Networks For Sentence Classification," Aug. 2014.

[17]   R. Jozefowicz, O. Vinyals, M. Schuster, N. Shazeer, And Y. Wu, "Exploring The Limits Of Language Modeling,".

[18]   N. Kalchbrenner, E. Grefenstette, And P. Blunsom, "A Convolutional Neural Network For Modelling Sentences," Apr. 2014.

[19]   X. Li And X. Wu, "Constructing Long Short-Term Memory Based Deep Recurrent Neural Networks For Large Vocabulary Speech Recognition," Oct. 2014.

[20]   H. Strobelt, S. Gehrmann, H. Pfister, And A. M. Rush, "Lstmvis: A Tool For Visual Analysis Of Hidden State Dynamics In Recurrent Neural Networks," Ieee Trans. Vis. Comput. Graph., 2018.

[21]   Y. Ming Et Al., "Understanding Hidden Memories Of Recurrent Neural Networks,".