International Journal of Scientific Research in Science, Engineering and Technology (www.ijsrset.com)
© 2020 IJSRSET | Volume 7 | Issue 3 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
DOI : https://doi.org/10.32628/IJSRSET

Mutual Information Approach for Sentiment Analysis Using Deep Machine Learning Convolution Neural Network (CNN) Model

Prof. Neha Khare, Priya Mishra
Takshshila Institute of Engineering and technology, Jabalpur, Madhya Pradesh, India

ABSTRACT

Sentiment Analysis is a field of Natural Language Processing which addresses the problem of extracting sentiment or, more generally, opinion from text. Obtaining deeper insights on that topic can be very valuable for a range of fields such as finance, marketing, politics and business. Previous research has shown how sentiment and public opinion can affect stock markets, product sales, polls as well as public health. This thesis researches the message sentiment polarity classification problem in Twitter aiming to classify messages based on the polarity of the sentiment towards a specific topic, where the tweets and the topics are always given. The dataset analysed and the evaluation metrics considered are provided conference and the 4th task about "Sentiment Analysis in Twitter". This task includes five subtasks, two of which were eventually engaged in this research according to the implemented approach. First, subtask B is a binary classification task, where the goal is to classify messages into two classes, positive and negative regarding the sentiment towards the topic. Following, subtask C where the target is to classify messages in a five-scale sentiment polarity from highly negative, negative, neutral, positive to highly positive, based on the sentiment towards a given topic. We are implemented and experimented with two deep learning models for both subtasks; a Convolutional Neural Network (CNN) and a state-of-the-art Recurrent Neural Network with context attention CNN LSTM model. We compare both models to the baselines of the challenge and show that the CNN LSTM outperforms the other models, in both subtasks, with all evaluation measures but one.

Keywords : Natural Language Processing ,Machine Learning, Convolutional Neural Network (CNN),Sentiment Analysis, LSTM, Word Processing.

I. INTRODUCTION

1. The Sentiment Analysis task

Sentiment Analysis (SA), a kind of Opinion Mining (OM), is a field of Natural Language Processing (NLP) whose goal is to extract the emotion, sentiment or more generally opinion expressed in a human written text. The text mostly derives from social media, product reviews and blogs. While the term opinion or sentiment is quite generic, the field of study attains a number of tasks. Some of these are, identifying the stance on a target or topic, for instance “Climate Change”, extracting the opinion on a product from a review or detecting sentiment polarity in a message. Sentiment Analysis is performed on various linguistic levels. The standard ones are document level, sentence level and aspect or entity level (Appel et al. 2016). Sentiment Analysis has plenty of applications in business, marketing and politics. Determining people’s opinion is key for future planning in many fields. SA can be used to evaluate future business plans based on public opinion on a new product. Trends in product sales can be pre-identified by measuring the sentiment of the customers. Many marketing agencies propose to companies the right direction for advertising a product based on public sentiment, which is extracted from messages in social media or product reviews. In addition, political parties plan their campaigns on public sentiment that can be extracted from text in social networks, blogs and forums.

Sentiment analysis is not an easy project. There are issues that can throw the analysis off and need attention. For instance, tweets can be sarcastic or contain ambiguous words, which often lead to misclassifying the polarity of the tweet. For example:

- "Shut up. And take my money."

which actually refers to a "must buy" product although the message could be classified as negative because it contains negatively charged words. In addition, in a product review there might be a case of a text such as the following, which expresses both a positive and a negative opinion:

- "The tuna was cooked perfectly but the miso dressing wasn't tasty at all".

Also, in Twitter, as well as in product reviews, the polarity if often unclear, for example:

- "Saakashvili is pushing his own agenda here. The Ukrainian economy is growing, although corruption is still a problem"
- "Although some vaccines protect our children, they still have potential to be very toxic".

A Sentiment Analysis system should also handle negation (i.e., "not good"); perform some kind of word sense disambiguation; and in the case of multiple sentiments and sentiment targets, be able to classify them accordingly. If a message has negative sentiment towards a topic, while expressing positive sentiment towards another topic, then the system should classify the message for each topic accordingly. Such a system should also be able to accurately detect irony which is challenging and part of ongoing research.

2. Message Polarity Classification in Twitter

Twitter is a social medium, micro-blogging site where users can post text messages, commonly referred as tweets[1]. The number of monthly active Twitter users in the fourth quarter of 2017 was 300 million[2] while approximately 90% of the tweets are public and can be collected for research without violating user anonymity. Tweets are available in real-time, through Twitter's streaming channel API[3]. Tweets can be filtered both by time and location that they were published.

Messages in Twitter usually include emoticons, misspellings (e.g., "Comeee onnnnn fineee, waaaay too"), slang language (e.g., "hooked on", for being addicted to something, "sick", for something very good) in addition to normal text. Tweets normally include hashtags which many times indicate the topics of the message (e.g., #Yemen, #thankyouobama, #BlackLivesMatter). These deviations in text should be handled or used to gain information regarding sentiment. In Table 1 you can see a few examples of messages in Twitter and their annotated label of sentiment. All annotations were performed on CrowdFlower.[4]

For the 4th SemEval task of "Sentiment Analysis in Twitter" messages should get classified based on the polarity of the opinion expressed in the tweet. The goal of this thesis is to study and examine Topic-based Message Polarity Classification which is described in subtasks B and C of Task 4 in SemEval 2017 (Rosenthal, Farra, and Nakov 2017). The goal of these two subtasks is to classify the sentiment of a tweet, towards a predefined topic. Subtask B employs a two-point scale; positive or negative, assuming that there is no neutral class. By contrast, subtask C employs a more elaborate five-point scale; i.e., highly negative, negative, neutral, positive and highly positive.

The task can be addressed in many different ways, such as using sentiment lexicons, using humans, using ontologies, etc. For the purpose of this thesis we used Machine Learning (ML) techniques to address the problem. We employ a supervised learning model, which gets trained to label messages based on their expressed sentiment polarity. The model is then tested and evaluated for its performance on accuracy and other metrics.

3. Approaches to Sentiment Classification

The existing methods for Sentiment Analysis can be grouped into two main categories.

1. Knowledge-Based
2. Machine Learning

In knowledge-based methods, also called Lexicon-based sentiment classification, the target is to construct or use existing sentiment word lexicons with indicated sentiment labels for the words or phrases in the text. The classification of the text is defined by rules; e.g., a function over thewords, such as the sum of word polarities (Taboada et al. 2011). This approach does not require any training (other than forming a lexicon, if required). However, it requires powerful linguistic resources[5] to extract knowledge from words, which are not always available.

Hu and Liu (2004a) built a lexicon, using only WordNet and a list of labeled seed adjectives. This list contains only positive adjectives (e.g., great, amazing, nice, cool) and negative adjectives (e.g., bad, boring). Their method retrieves and automatically labels the synonyms (same polarity) and antonyms (opposite polarity). This process allows the list to grow into a lexicon. A drawback of this approach is that it is only applicable in languages where WordNet is available. In any case, the knowledge-based method may be difficult due to noise in text data, while manually creating rules to combine information about words obtained from the sentiment lexicons takes time and effort.

On the other hand, Machine Learning requires training a model to predict the polarity of the text. The model is trained with text messages, labeled for their sentiment and represented as feature vectors. The latter conventionally requires text preprocessing using language processing tools like NLTK[6]. Text preprocessing mainly involves tokenization, stemming, tagging, and possibly parsing of the text. The selection of the appropriate features from data is crucial and has proven to be a major issue and is always a key objective for researchers.

Previous work on sentiment analysis has exploited well-known supervised machine learning methods, such as Naive Bayes (Martinez-Arroyo and Sucar 2006) , SVMs (Vinet and Zhedanov 2010), Random Forests (Ho 1995), (Wahid et al. 2017). More recent work uses deep learning models (Goldberg and Hirst 2017), especially Recurrent Neural Networks (RNN) and Convolutional Neural Network (CNN). This thesis also focuses on deep learning models.

II. RELATED WORK

Sentiment Analysis has been an attractive object of study for AI researchers, computational linguists, cognitive scientists and neurobiologist. As mentioned in Section 1.3 one of the most successful approaches for Sentiment Analysis is Nature Language Processing with Machine Learning (ML) techniques. The fundamental requirement for using supervised ML to be able to solve classification and regression problems, is data availability. Thus, while more and more text data become available through blogs, websites and social media, where people can publish their opinion on many subjects, including politics, products, events and business among others, more researchers studied SA. The first papers on SA were focused on reviews for movies (Pang, Lee, and Vaithyanathan 2002), (Turney 2001), (Pang et al. 2002), (Pang and Lee 2005) and (Popescu and Etzioni 2005) or products (Hu and Liu 2004b), (Popescu and Etzioni 2005). Following those studies, other focused on the analysis of sales of products such as books, movies and videogames based on customers opinions (Chevalier and Mayzlin 2006), (Mishne and Glance 2006), (Liu et al. 2007), (Zhu and Zhang 2010).

With the rapid growth of social media like Twitter, more attention was drawn towards social media content as in (Jansen et al. 2009), (Asur and Huberman 2010), (Arias, Arratia, and Xuriguera 2013). In addition to Sentiment Analysis and the impact of people's opinion on sales, extensive research has been done on separate fields of finance and economics. As demonstrated by Lemmon and Portniaguina (2006) and Han (2008) there is a correlation between the sentiment and confidence of the investors and the stock market. Moreover Gilbert and Karahalios (2010) show that "estimating emotions from weblogs provides novel information about future stock market prices.", while Bollen, Mao, and Zeng (2011) explored the fact that national events affect people's emotions and the relationship of their emotions to the value of Dow Jones Industrial Average (DJIA). Due to these findings, more work has been done in the last years on the subject (Oh and Sheng 2011; Zhang, Fuehres, and Gloor 2011; Makrehchi, Shah, and Liao 2013; Si et al. 2013; Smailović et al. 2013; Sprenger et al. 2014; Sprenger et al. 2014). Quoting Mitchell et al. (2013) "Companies should pay more attention to the analysis of sentiment related to their brands and products in social media communication as well as in designing advertising content that triggers emotions."

The ability to measure public opinion on social and political affairs is critical for political parties. The usual methods such as polls are expensive, they may not be accurate and the results are not representative of the public sentiment. Overall polls are unreliable. In addition, getting people's opinion by asking questions is not the best method of collecting useful data. Thus, Sentiment Analysis in social media like Twitter may provide an alternative measure of public opinion and extract useful data (Ceron, Curini, and Stefano 2012; O'Connor et al. 2010; Stieglitz and Dang- Xuan 2012; Zhou et al. 2013). For example, Diakopoulos et al. (2010) present an analysis of ephemeral changes of sentiment in reaction to the first U.S. presidential debate video in 2008.

Further work has been done in other areas of sentiment analysis in social media. Sakaki et al. (2010) propose a method of detecting major events by analysing the stream text in Twitter and at Culotta (2010) propose methods of identifying influenza-related messages. Data from Twitter can be used to analyze public emotion, demography, health characteristics and the "geography of happiness" (Mitchell et al. 2013), a term describing the correlation of sentiment to place. Studying virality in Twitter and the correlation of viral messages with sentiment, Hansen et al. (2011) showed that "news with negative sentiment is more likely to become viral, while in the non-news segment this is not the case".

III. PROPOSED WORK AND RESULTS

Machine Learning ModelThe implementation of the machine learning model is visualised below,

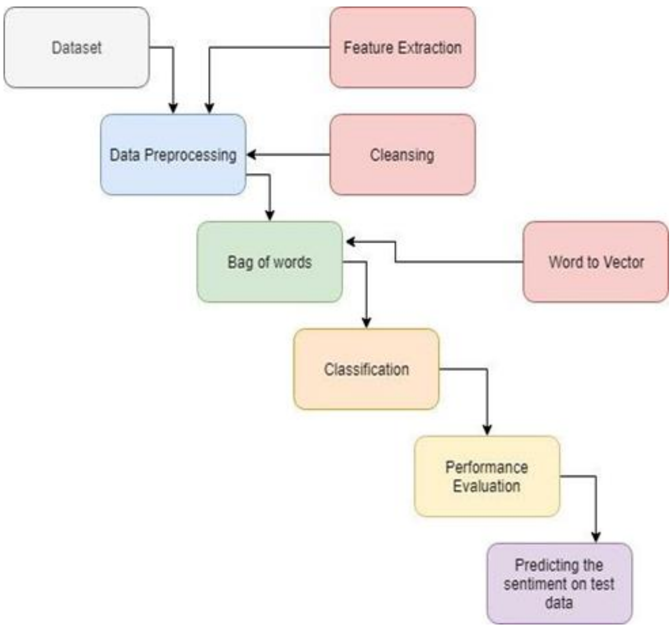

Figure 3.1 Implementation of Machine Learning Model

•Performed the validation with Naïve Bayes, Random Forest, and XGBoost classifiers. The performance metrics for this is tabulated. Initially, the training and testing datasets are split in a 7:3 ratio. This is done to check if the model is built correctly.

•Performed 8-fold cross validation with Naive Bayes, Random forest, and XGBoost on the data set and the performance metrics are tabulated.

1) **Deep Learning Model**

Firstly, the inputs are to be fed to the neural network which is in the English language. The labelled data set which underwent cleansing in a machine learning model is taken here as well for the classification. As neural networks don't understand the English language, the statements undergo word-to-vector representation where each word is represented by its rank which is given according to the number of the frequently repeated word among the other words in the dataset. The input finally obtained is in vector format. This input vector is trained to two consecutive types of neural networks namely, convolutional neural network and long short-term memory network.

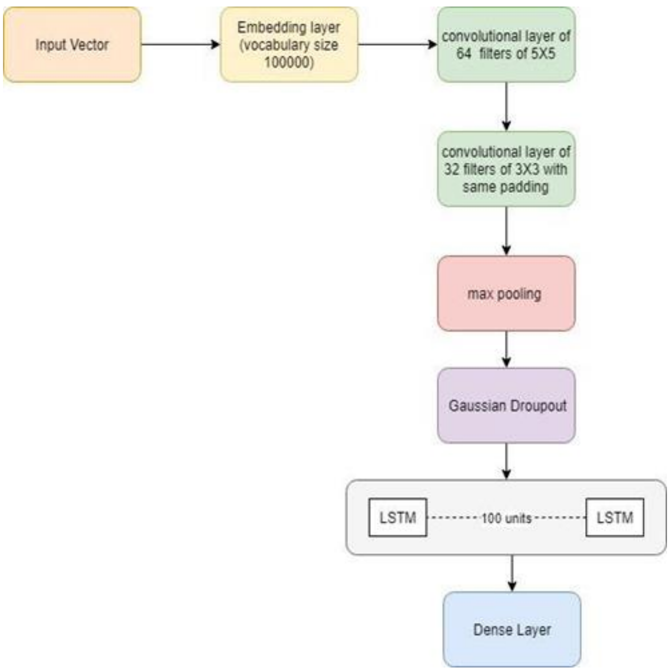

Figure 3.2 Implementation of Deep Learning model

A convolutional neural network which is a layered neural network. The first layer is the embedding layer which embeds the words into low-dimensional vectors. The second layer is the convolutional layer, in this layer multiplication of the input vector and the weight vector is calculated to find the weighted sum. The third layer is the max pooling layer. Here a filter is used to reduce the dimension of input to this layer by replacing values with the maximum among considered values. The fourth layer is the Gaussian Dropout layer, this layer is useful in mitigating overfitting and is only active at training time.

The next type of neural network attached to the model is long short-term memory (LSTM), where there are three gates, these gates are sigmoid associated gates. Finally, a single node dense layer is used to obtain the output of the classification. Meanwhile, the neural network adjusts its network to predict the outcome with at most accuracy. Back-propagation is introduced to calculate the loss and propagate the error back in the network and update the weights accordingly.

Implementation Details:

We will use Anaconda IDE and python 3.7 for implementing above algorithms and we will use Twitter dataset of tweets and to apply prediction model algorithm mentioned above to predict positive and negative tweets.

IV. Experimental Setup

To implement the models, we used LSTM[11] with Keras and Tensor Flow as backend. We trained the neural networks on a GTX 1060 (6GB) GPU. We provide the source code of the LSTM and CNN model.[12]

DATASET:we have used Twitter dataset of 1600000 record and we have predict the positive ,neutral and Negative tweets according to sentimental analysis the

			PDT 2009			for the ball. Man...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all....

20; 7 (3) : 270-276

following dataset is:

	target	ids	date	flag	user	text
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all....

Figure 3.3 Data set of Twitter

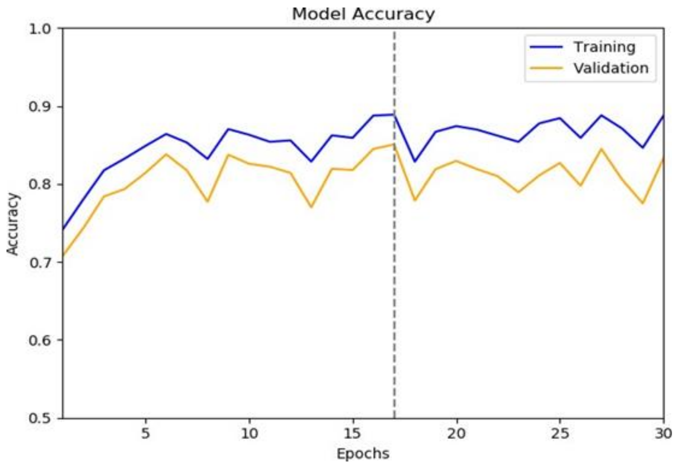

Figure 3. CNN LSTM Accuracy in 30 epochs.

Evaluation at Validation

Before we present the scores that the CNN LSTM model achieved in the task, we explain how we worked during training and the decisions we made based on evaluations over the validation data. In Subtask B we trained the model for 30 epochs, while evaluating its progress by measuring the accuracy on the validation part of the dataset for the same number of epochs. We decided to use accuracy as our main evaluation metric hence we stopped the training after 17 epochs (as shown in Figure 3, the point where the model achieved optimum accuracy). We evaluated and compared the CNN LSTM model with the CNN model for 30 epochs and we observed that the CNN achieves its best performance after the 25th epoch, but, still achieves worse accuracy than the CNN LSTM at its best epoch (17^{th}), though the difference is small and the performance of the CNN is more stable. There is no need to train beyond 30 epochs because the validation metrics decline for both models from that point on. Likewise, for Subtask C we trained the model for 18 epochs considering that this was the lowest point for MAE^{M} during validation. We chose to fine-tune the CNN LSTM model with respect only to macro-averaged MAE, but in future work we would like to study the case of a more balanced performance for both micro and macro averaged MAE. Here the CNN LSTM performs significantly better than the CNN model.

V. Conclusions and Future Work

In this thesis we built a neural network model (called CNN LSTM) for topic-based sentiment classification for messages in Twitter. We re-implemented in PyTorch a Bi-LSTM model (CNN LSTM) based on the work of Baziotis et al. (2017) for a sentiment analysis task that consists of two subtasks and we released it for public use. We managed to outperform the baselines provided by SemEval 2017, while scoring high results in both subtasks. We obtained a test accuracy score of 0.860 in subtask B and regarding subtask C we reduced the macro-average mean absolute error in test data at 0.584. In addition, we built and trained a CNN model (Kim 2014) and compared results obtained from both models. The CNN LSTM performs slightly better than the CNN model in subtask B and much better in subtask C. We added a weighted loss to CNN LSTM, leading to the CNN LSTM and studied the effect in both subtasks. Although this is not a novel addition, we evaluated this component reporting the margin with which it improves the model.

VI. Future work

More research will help deliver even better results in similar tasks, while other types of neural network, such as CNNs are also getting very good results alone or by working together with LSTMs (Cliche 2017).

We did not extensively tune the hyper-parameters of our models. In most cases, we used defaults or hyper-parameter values from previous work. Hence, further improvements may be possible with hyper-parameter tuning, for example using Bayesian Optimization (Snoek et al. 2016). In the CNN LSTM model we "froze" the embeddings, not letting their weights to be updated during training. As a next step, we intend to study employing trainable word embeddings, in order to examine whether better and domain adapted word representations can improve the models. Therefore, future work consists of:

- Extensive fine tuning.
- Trainable embeddings.

VII. REFERENCES

[1]. Al-Rfou, Rami et al. 2020. "Theano: A Python Framework for Fast Computation of Mathematical Expressions." CoRR abs/1605.0.

[2]. Appel, Orestes, Francisco Chiclana, Jenny Carter, and Hamido Fujita. 2019. "A Hybrid Approach to the Sentiment Analysis Problem at the Sentence Level." Knowledge-Based Systems 108:110-24.

[3]. Arias, Marta, Argimiro Arratia, and Ramon Xuriguera. 2019. "Forecasting with Twitter Data."ACM Transactions on Intelligent Systems and Technology 5(1):1-24.

[4]. Asur, Sitaram and Bernardo A. Huberman. 2010. "Predicting the Future with Social Media." Pp.492-99 in 2019 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT '19. Washington, DC, USA: IEEE Computer Society.

[5]. Baziotis, Christos, Nikos Pelekis, and Christos Doulkeridis. 2017. "DataStories at SemEval-2018 Task 4: Deep LSTM with Attention for Message-Level and Topic-Based Sentiment Analysis." Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2018) (1):747-54.

[6]. Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2018. "Twitter Mood Predicts the Stock Market."Journal of Computational Science 2(1):1-8.

[7]. Ceron, Andrea, Luigi Curini, and M. Stefano. 2019. "Tweet Your Vote: How Content Analysis of Social Networks Can Improve Our Knowledge of Citizens' Policy Preferences. An Application to Italy and France." New Media &amp Society 16:1-24.

[8]. Chevalier, Judith A. and Dina Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." Journal of Marketing Research 43(3):345-54.

[9]. Collobert, Ronan et al. 2019. "Natural Language Processing (Almost) from Scratch." CoRRabs/1103.0398.

Cite this article as :

Prof. Neha Khare, Priya Mishra, "Mutual Information Approach For Sentiment Analysis Using Deep Machine Learning Convolution Neural Network (CNN) Model", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 7 Issue 3, pp. 270-276, May-June 2020.

Journal URL : http://ijsrset.com/IJSRSET207370