International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

Emotion Recognition in Text Using Deep Learning Approaches: A CNN and LSTM Based Classifier

Deepak Agrawal, Shyam Shankar Dwivedi
Computer science and Engineering Department,
Rameshwaram Institute of Technology & Management, Lucknow
adeepaktop1@gmail.com

Abstract: Analyzing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to develop a functional classifier using deep learning approach and GloVe embedding for accurate and automatic emotion classification of an unknown text. In this project we have proposed a system using CNN and LSTM which can classify short text into five classes, namely: neutral, sad, happy, hate and anger.

Keywords: Public sentiment analysis, CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory), GloVe embedding Sentiment analysis, Emotion detection.

1. Introduction:

Community circles, including blogs, microblogs, forum conversations, and reviews, has grown rapidly during the past ten years. As a result, the web has undergone significant change, enabling billions of people worldwide to engage in a wide range of activities, including sharing, interacting, posting, and manipulating content. Unlike the conventional structured data that is stored in databases, this allows us to be connected and communicate with each other at any time and without regard to geographic borders. The unstructured user-generated data that is produced requires social media mining to develop new computing tools, but it also gives us the chance to study and comprehend people at never-before-seen scales [1, 2, 3, 4, 5, 6, 7]. Social media is a way of information that businesses are using more and more frequently. Conversely, individuals are more than glad to use social media to share details linked their day to day life, expertise, involvement, and point of view with surrounding peoples. By voicing their thoughts and remarks on events that occur in society, they actively engage in such occurrences. Because they are able to share their knowledge and feelings with the public through social media, businesses are compelled to gather more data about their brands, goods, and reputation in order to make informed decisions and continue operating their enterprises successfully.

Emotion recognition is a set of computational methods that automatically extracts and condenses the views from enormous amounts of data that are too large for the ordinary human reader to process. Social media's vast ocean of divisive posts is essential to people's lives because it influences our actions and helps transform industries. Businesses, organizations, and enterprises no longer need to conduct surveys or polls to get opinions about items since there are so many customer evaluations and debates in public forums on the Internet. Additionally, people are no longer limited to enlisting the help of friends and family. Therefore, there are several immediate and practical applications as well as business interests for gathering and evaluating such viewpoints using computational sentiment analysis approaches. Social gatherings, political appointments, healthcare, financial services, consumer products and services, and, more recently, crisis management and natural catastrophes are all covered by these apps.

2. Related Work:

A key area in research for emotion recognition is a categorization of polarity, considers whether a language or document's opinion about a certain characteristic or feature of a target is neutral, adverse, or favourable. Previous studies on its field were conducted by [11, 12], who used various techniques to identify the polarity of movie and product evaluations. The efficacy of utilizing support vector machines and naïve Bayes for emotion identification in movie reviews was initially investigated by Pang et al. [11]. Emotion recognition is often considered a two-class systematize task. When doing this kind of study, emotion identification is basically a difficulty with text categorization, with attribute choosing having a big influence on the classification algorithm's performance. Similar to machine learning and statistics, attribute choosing is the process of choosing a subset of associated with characteristics to build a model. Attribute choosing approaches are selected in text classification to improve generalization by resolving the over-fitting issue, shorten training periods, and simplify the models for better interpretation. Emotion words that denote a certain polarity should be used as feature words in emotion recognition, which requires processing more intricate semantics. Additionally, when comparing various feature selection strategies, researchers found that employing unigram features outperforms other feature types. On the emotion classification test, however, two machine learning techniques perform worse than conventional text classification tasks. The secret to the emotion classification challenge is the same as it is for standard classification tasks: choosing useful characteristics.

3. Methodology:

We propose to use "multi-channel" which originate by grouping of twist kernels (can be said CNN too) and Long-Short-Term-Memory (L.S.T.M.) segments to distinguish small length word sequences (X's Post named as Tweet) into one of five elements of emotion classes instead of the conventional 2 faces (good/bad) or ternary (effective/denial/impartial) items of class.

Research on the classification of brief text sequences into many classes is still very rare. Specifically, few papers explain the efficacy of categorizing brief text sequences (like tweets) into more than three unique types (positive/negative/neutral), exception explained on Bouazizi - Ohtsuki (2017)[41]. Specifically, in two of their publications, Bouazizi-Ohtsuki are the one who obtained correctness of 56.9% or 60.2% on seventh different classes. Our goal is to present a novel strategy that can significantly increase classification accuracy.

CNN:

The most common application of convolutional network from neurons, also referred to as CNNs or ConvNets, in deep learning analysis for visual data. The terms shift invariant and space invariant artificial neural networks (SIANN) are denoted to describe them, because to its translation invariance and shared-weights architecture qualities. They are used with a variety of applications, including medical image analysis, systems used for recommendation, describing characteristics of a picture, natural language processing (NLP), and picture and video recognition.

Regularized versions of CNNs are called multilayer perceptrons. Multi-layer perceptrons are commonly explained as totally inter-connected networks, where each neuron inside in a surface is jointed to all other layer's neurons. The "fully-connectedness" of these type of networks makes vulnerable to data override. One basic and popular regularization tech is to link a magnitude's weight measurement to the loss function. CNNs, however, proceed towards formalization in a different manner. They build together more intricate designs from basic, lesser ones using the data's hierarchical pattern. In connectedness point of view and convolutedness, CNNs are consequently at the down end of its spectrum.

STACKING:

To simplify the underlying processing, networks for convolutional might have local or global stacking layers. By merging the outputs of neuron clusters in one layer into a single neuron in the layer below, stacking layers lower the dimensionality of the data. Tiny clusters are joined together, typically two by two, by local pooling. Global pooling affects every neuron in the convolutional layer.[33][34] Additionally, pooling can determine a maximum or an avg. Maximum stacking used the greatest value of item from every cluster of neurons in the preceding layer. Avg. stacking uses the avg. value from every cluster having neurons in the preceding layer [42].

FULLY CONNECTED LAYERS:

Via fully linked layers, each neuron in one layer can communicate with every other layer's neuron. This neural network is theoretically equal to the normal multi-layer perceptron neural network (MLP). The flattened matrix goes through a fully linked layer to classify the images.

LONG SHORT-TERM MEMORY (LSTM):

Deep learning makes advantage of this artificial recurrent neural network (RNN) architecture [37]. Unlike traditional feed-forward neural networks, LSTM has feedback connections. Both individual data points and whole data sequences (such as audio or video) can be handled by it. (i.e. photos).

Example - L.S.T.M. can be applied to jobs such as intrusion detection systems (IDSs), speech recognition [39, 40], un-segmented, linked recognition of handwriting [38], and network traffic anomaly detection. LSTM networks are perfect for classification, evaluation, and prediction based on time series data since there could be unanticipated delays between important occurrences in a time series. When training conventional RNNs, issues with vanishing and exploding gradients may occur. To address these issues, LSTMs were created.

For example, LSTM can be applied to tasks such as intrusion detection systems (IDSs), voice recognition [39, 40], un-segmented, linked handwriting recognition [38], and network traffic anomaly detection. For categorization, processing, and prediction based on time series data, LSTM networks are ideal since there could be unanticipated lags between important events. Problems with vanishing and exploding gradients may occur when training traditional RNNs. To address these issues, LSTMs were created. In different situations, LSTM perform outstanding, hidden Markov-Technique, and many more sequence techniques of learning because its relative insensitivity to the length of the gap [13].

CHOICE OF NEURAL NETWORK MODEL:

For many NLP applications, RNNs—especially LSTMs—are the best choice since they can "learn" the importance of the sequential data's sequence, including texts and time-series data. CNNs, on the other hand, take characteristics out of data in order to identify them. Previous techniques either employ both methods in the same model (Sosa, 2017) or utilize just one way (Yoo Kim, 2014).

Using components from all of the previously mentioned models, we expand on the concept of building multi-channel networks by letting the model try to figure out for itself which channels help it make better predictions for particular types of input. According to our theory, it can be enabled for the method to join the all benefits of the many routes to provide forecasts that are generally more accurate.

THE SEQUENCE IN WHICH WE ENTER THE INPUT SEQUENCE INTO THE MODEL IS CHOSEN:

An L.S.T.M. CNN replica outperforms a CNN L.S.T.M., pure CNN, or pure LSTM model, as demonstrated by Sosa (2017) [33]. The actual accuracy variations, however, vary depending on the class and stay somewhat close. We employ both strategies in order to benefit from the ability to outcome with meaningful result, our objective is to extract as many features as possible from the limited sequence, therefore we extract features from the LSTM's sequential output (LSTM-CNN) and features to feed into the LSTM (CNN-LSTM).We are able to max-pool and concatenate every channel. The layer from last is a well-connected layer that uses a soft-max activation function to provide a prediction.

The parameters we have used in model are as follows:

1. Separate the data into training and test
2. We use **categorical_crossentropy** Loss function.
3. Batch = 128,
4. Epochs= 50.

International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

5. Relu Activation function.
6. Optimizer we used is 'adam'
7. On each convolution layers 12 filters are used.

4. Result and Discussion:

We have examined the findings in this part using the confusion matrix, test accuracy, and train accuracy. We propose to use "multi-channel", In place of the frequently used binary (effective/denial) or ternary (effective/denial/impartial) categorization, convolutional filter (normally pointed as CNN) and Long-brief-Term Memory (L.S.T.M.) units are used to categorize brief text sequences (in our instance, X's posts) into 1 of 5 emotion classes.

Research on the classification of brief text sequences into many classes is still very rare. Specifically, few research indicate the efficacy of categorizing brief text sequences (like tweets) into more than three unique types (positive/negative/neutral), with the exception explained by Bouazizi-Ohtsuki (2017) [22]. In particular, we have obtained more than 63% characteristics among all others accuracy and precision, while Bouazizi and Ohtsuki only managed 56.9% or 60.2% overall accuracy on seven different classes in two of their articles.

```
Epoch 46/50
37831/37831 [==============================] - 183s 5ms/step - loss: 0.8515 - acc: 0.6504 - val_loss: 0.9102 - val_acc: 0.632
9
Epoch 47/50
37831/37831 [==============================] - 182s 5ms/step - loss: 0.8482 - acc: 0.6508 - val_loss: 0.9107 - val_acc: 0.634
5
Epoch 48/50
37831/37831 [==============================] - 182s 5ms/step - loss: 0.8489 - acc: 0.6546 - val_loss: 0.9097 - val_acc: 0.632
5
Epoch 49/50
37831/37831 [==============================] - 183s 5ms/step - loss: 0.8456 - acc: 0.6558 - val_loss: 0.9101 - val_acc: 0.631
0
Epoch 50/50
37831/37831 [==============================] - 182s 5ms/step - loss: 0.8456 - acc: 0.6539 - val_loss: 0.9117 - val_acc: 0.634
9
```

Fig.1: Training and Validation Accuracy.

From Fig.-1, we can clearly see that the training accuracy is 65.39 % whereas validation accuracy is 63.4 %. We have used 50 epochs for the training process.

	precision	recall	f1-score	support
neutral	0.50	0.34	0.41	1927
happy	0.64	0.68	0.66	3249
sad	0.58	0.71	0.64	3183
hate	0.85	0.66	0.74	865
anger	0.84	0.61	0.71	233
micro avg	0.62	0.62	0.62	9457
macro avg	0.68	0.60	0.63	9457
weighted avg	0.62	0.62	0.61	9457

Fig.2: Test Accuracy

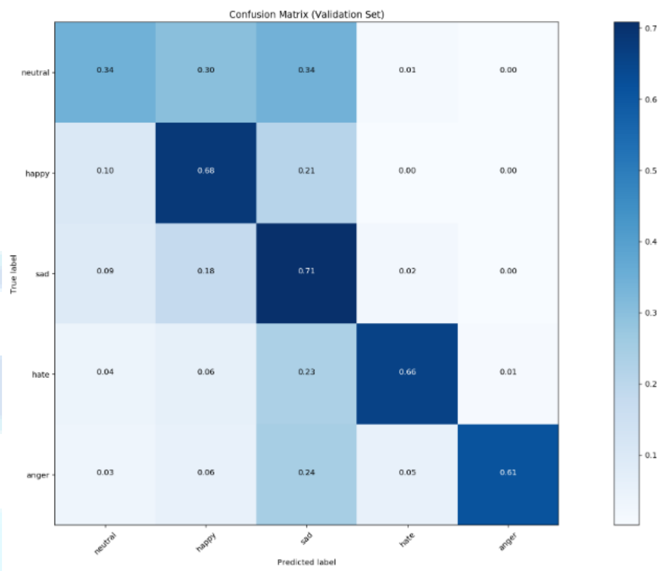

Fig. 3: Confusion Matrix

5. Conclusion:

In present days, social networks are increasingly popular across internet users and produce enormous amounts of data every day. There are some worthwhile uses for this kind of data. We require an efficient approach that can manage real-time analytics and the analysis of massive data sets in order to analyze social data. As a result, guidelines for recognizing emotions have been provided. Text analytics and empathy analysis may offer useful business data to companies. To examine the text's emotions, we used a deep network of convolutions with LSTM. Glove Embedding, which we utilized, outperforms other embedding methods currently in use. There are five categories into which emotions can be divided. Our method attained a 65.39 percent accuracy rate and indicates encouraging outcomes. Thus, we can ultimately say that the deep convolution neural network that uses pre-trained word vectors performs well when it comes to textual emotion recognition.

References:

[1] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. Social media mining: an introduction. Cambridge University Press, 2014.

[2] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. In Proceedings of the 19th international conference on World wide web, pages 641–650. ACM, 2010.

[3] G. Beigi, M. Jalili, H. Alvari, and G. Sukthankar. Leveraging community detection for accurate trust prediction. In ASE International Conference on Social Computing, June 2014.

[4] Jiliang Tang, Xia Hu, and Huan Liu. Social recommendation: a review. Social Network Analysis and Mining, 3(4):1113–1133, 2013.

[5] Hamidreza Alvari, Alireza Hajibagheri, and Gita Sukthankar. Community detection in dynamic social networks: A game-theoretic approach. In Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on, pages 101–107. IEEE, 2014.

[6] Michaela Goetz, Jure Leskovec, Mary McGlohon, and Christos Faloutsos. Modeling blog dynamics. Citeseer, 2009.
[7] Hamidreza Alvari, Sattar Hashemi, and Ali Hamzeh. Discovering overlapping communities in social networks: A novel game-theoretic approach. AI Communications, 26(2):161–177, 2013.
[8] Maura Conway, Lisa McInerney, Neil O'Hare, Alan F. Smeaton, Adam Berminghan, "Combining Social Network Analysis and Sentiment to Explore the Potential for Online Radicalisation," Centre for Sensor Web Technologies and School of Law and Government.
[9] Bing Liu. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.
[10] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1–135, January 2008.
[11] Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on empirical methods in natural language processing, vol 10, 2002. Association for Computational Linguistics, pp 79-86
[12] Turney PD (2002) Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics, 2002. Association for Computational Linguistics, pp 417–424
[13] Khan FH, Qamar U, Bashir S (2016) eSAP: a decision support framework for enhanced sentiment analysis and polarity classification. Inf Sci 367:862–873.
[14] Tellez ES, Miranda-Jiménez S, GraffM, Moctezuma D, Suárez RR, SiordiaOS (2017)A simple approach to multilingual polarity classification in twitter. Pattern Recogn Lett.
[15] AsgharMZ, Khan A, Ahmad S, Qasim M, Khan IA (2017) Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLoS ONE 12(2):e0171649
[16] Phu VN, Tran VTN, Chau VTN, Dat ND, Duy KLD (2017) A decision tree using ID3 algorithm for English semantic analysis. Int J Speech Technol 1-21
[17] Liu Y, Bi JW, Fan ZP (2017) A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm. Inf Sci 394:38–52
[18] Tang D, Qin B, Liu T (2015) Document modeling with gated recurrent neural network for sentiment classification. EMNLP 2015:1422–1432
[19] Tang D, Wei F, Qin B, Liu T, Zhou M (2014) Coooolll: a deep learning system for twitter sentiment classification. In: SemEval@ COLING, 2014. pp 208–212