Sentiment Analysis: Multi-level Multi-Class Emotion recognition

Donia Ghazy – Marwan Eid - Mohammed Abuelwafa

The American University in Cairo – CSCE 4602 (Practical Machine Deep Learning)

Introduction

Multi-class emotion recognition is a new field of great interest in Artificial Intelligence. More than ever, people use online platforms to communicate facts, opinions, and emotions with different emotion intensities. Analysing the emotions expressed in online content is an interesting research field due to its applicability in a wide range of fields such as commerce or human-computer communication. Whereas binary sentiment analysis detects whether a text is positive or negative and ternary sentiment analysis detects whether a text is positive, negative or neutral, multiclass sentiment analysis classifies a text into a category of given emotions such as anger or happiness. State-of-the-art binary classification sentiment analysis reaches accuracies of 97%, whereas multi-class sentiment analysis varies between 55% and 60%. The choice of labels, number of classes, and the dataset features regarding balancing the dataset are important factors affecting the accuracy of the multi-class classification. We hypothesize that multi-level classification could lead to better results and adopt a mixture of experts approach where one classifier is trained for positive emotions and another for negative emotions. Using the output of a binary classifier as the input of the two classifiers, we tested our hypothesis by training the classifiers separately, where we parsed the dataset and added noise to the text data. The general architecture of the proposed model is in figure 1.

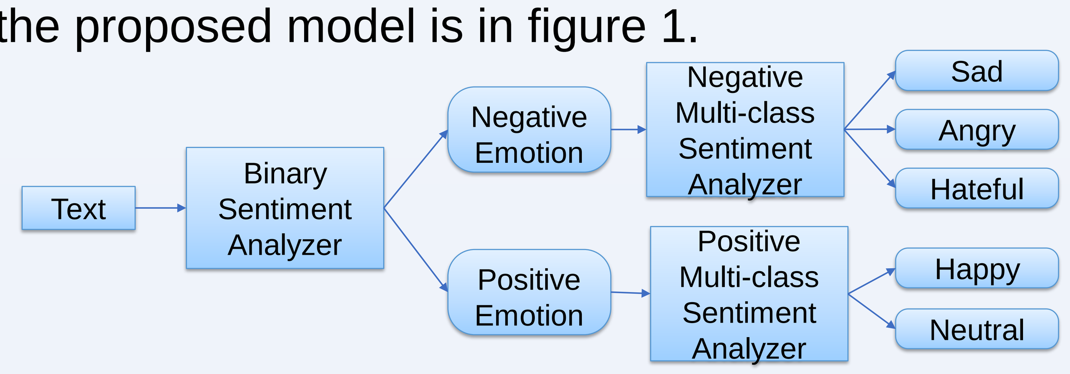

Figure 1

Whereas work of researches on the used dataset had an accuracy of 62.5% [1], our approach reached a 72.45% accuracy. Hence, a multi-level architecture performs better than a single-level architecture in emotion recognition with multi-class classification.

Existing Solutions

Most of previous research used either rule based or statistical machine learning approaches for opinion mining and sentiment analysis. Ibrahim et al. [4] introduced a detailed survey of different techniques used for opinion mining and sentiment analysis. The paper used the document-level sentiment classification which is a binary classification task of labelling a document as expressing either an overall positive or negative opinion.

Othman et al. [4] discussed that both supervised methods like Support Vector Machines or Naïve Bayes and unsupervised methods were used for document-level sentiment classification.

Turney et al. [2] proposed an unsupervised algorithm that uses semantic orientation of the phrases for classification of reviews. The lexicon-based approach determines the polarity or sentiment using some function of opinion words in the document or sentence and attains an average accuracy of 74%.

Another approach is the Recursive Neural Tensor Networks (RNTNs) which have a tree structure with a neural net at each node; that is, nodes are combined into parents using a weight matrix that is shared across the whole network. This model outperformed its preceding methods on several metrics and pushed the state-of-the-art in binary classification from 80% up to 85.4% [3].

Bidirectional Encoder Representations from Transformers (BERT) is another solution that is used in sentiment analysis. It relies on bidirectional LSTM where bidirectional training is performed on sequences instead of unidirectional training. The latest model utilizing this approach pushed the state-of-the-art accuracy of fine-grained sentiment analysis to 55.5% [3].

There are many other proposed solutions that use different approaches such as using Recurrent Neural Networks and Long Short Term Memory, Attention-based Neural Networks, or Fuzzy Convolutional Neural Networks.

Dataset & Architecture

The training and validation testing set is originally composed of 47,288 tweets from twitter with labelled emotions of five classes: neutral, happy, sad, anger and hate. We label 0 for neutral and 1 for happy for the positive classifier and 2 for sad, 3 for hate, and 4 for anger.

Both classifiers used CNNs for feature extraction after performing TF-IDF on the text in order to get the word embedding vectors of the text and classified the five classes using three Fully Connected layers. Using the ReLU activation function and the L2 regularization, with λ = 0.01, the architecture of both classifiers is in table 1.

Table 1

Layer	Out Shape	Params #
Conv1D	(None, 500, 64)	256
Conv1D	(None, 500, 128)	41088
Conv1D	(None, 500, 128)	82048
MaxPooling1D	(None, 250, 128)	0
Conv1D	(None, 250, 128)	82048
MaxPooling1D	(None, 125, 128)	0
Flatten	(None, 16000)	0
Dense	(None, 64)	1024064
Dense	(None, 64)	4160
Dense	(None, 2/3 for +ve/-ve classifier)	195

Experiment

A total of 43924 tweets were used for training, 70%, and validation, 30%. Figure 2 shows the count of tweets for each sentiment.

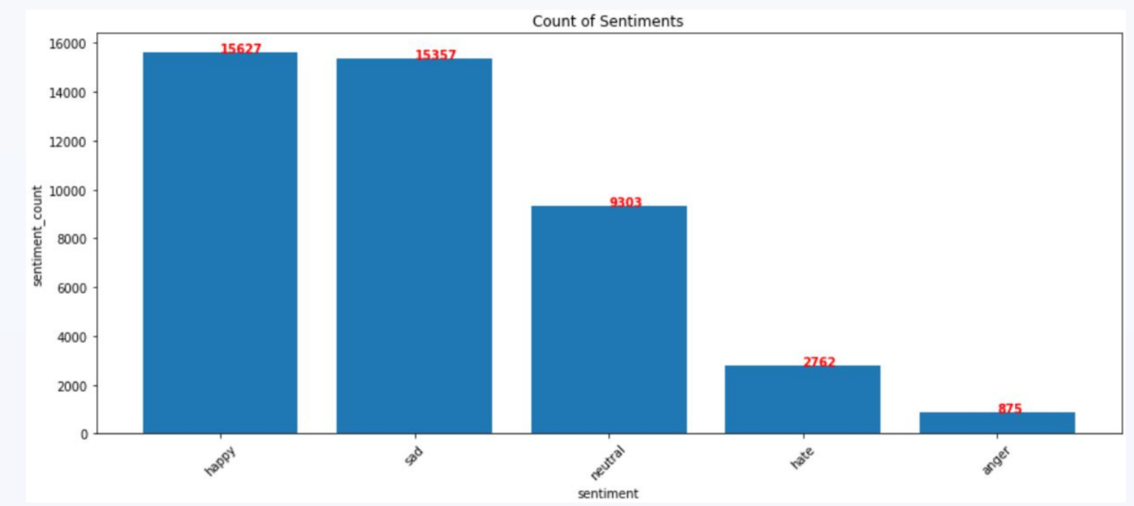

Figure 2

Data was preprocessed by filtering emojis, punctuation, and special characters and changing the letters to lowercase. Bag of Words with TF-IDF scheme was used to convert the text to numerical data. In this approach, the vocabulary of all unique words in all documents is formed and utilized as a feature vector. Words that occur more frequently in one document and less so in other documents are given more importance via TF-IDF.

$$TF = \frac{freq.of\ a\ word\ in\ a\ document}{Total\ no.of\ words\ in\ a\ document}$$

$$IDF = \frac{Total\ no.of\ documents}{No.of\ documents\ containing\ the\ word}$$

For the feature vectors, we chose the most 500 recurring words in the dataset such that the word is not present in more than 80% of the document but is present in at least 4 documents.

Adam optimizer with a 5e-4 learning rate was used; the learning rate was fine and coarse-searched. Training of the positive and negative classifiers is shown in figures 4 and 5, respectively.

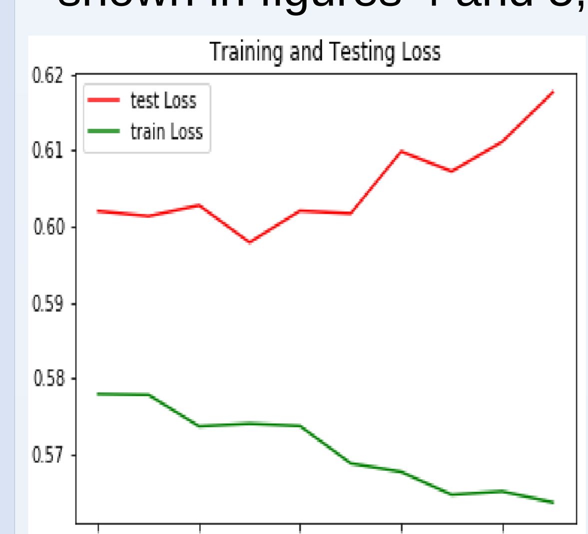

Figure 3

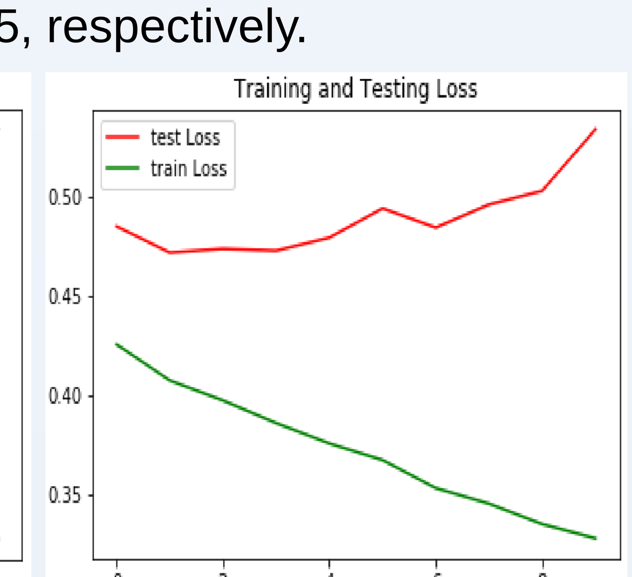

Figure 4

Results

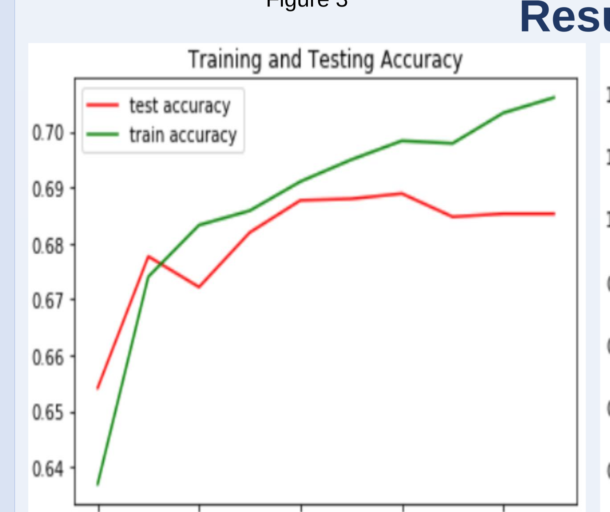

Figure 5

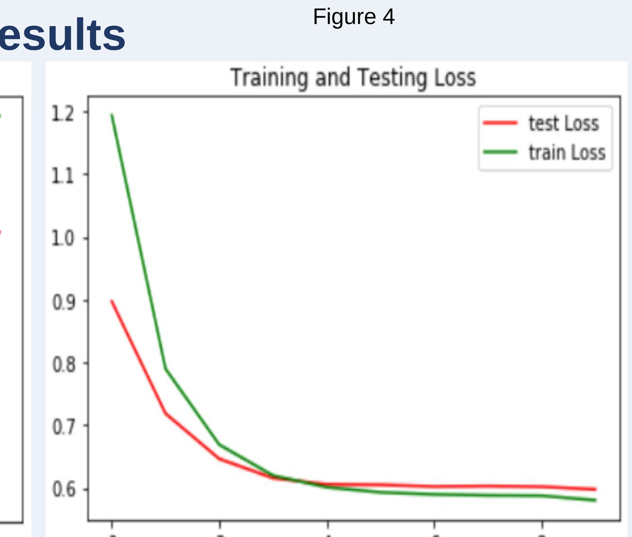

Figure 6

Training and testing accuracies and losses for the positive classifiers were shown in figures 6 and 7, respectively. Training and testing accuracies and losses for the negative classifiers were shown in figures 7 and 8, respectively.

Figure 7

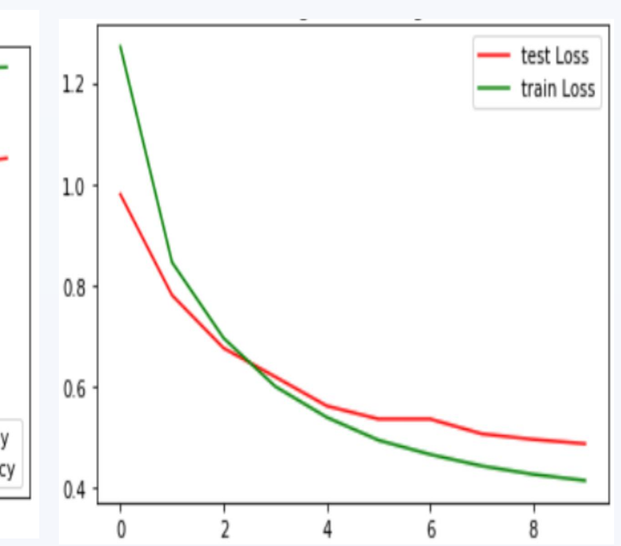

Figure 8

The negative classifier reached a testing accuracy of 87.8%, while the positive classifier reached a 68.5%, resulting in a weighted accuracy of 72.45% for the model.

$$Weighted\ Acc. = \frac{(0.6852 * 7479) + (0.8789 * 1900)}{9379}$$

$$= 72.45\%$$

Conclusion

The analysis of textual data has been in the interest of researchers recently. Multi-class emotion recognition is more complex than binary and ternary classification, where state-of-the-art accuracies vary from 55 - 60%. We proposed a multi-level multi-class architecture that reached a 72.44% accuracy outperforming the highest accuracy of 62.5% on the used dataset.

Future Work

Due to time constraints, we recommend for future work trying a different architecture for the positive emotion classifier as well as trying larger datasets. Further research is needed regarding the dataset features and labels selection.

References

[1] Multi-class Emotion Classification for Short Texts. (n.d.). Retrieved November 11, 2020, from https://tlkh.github.io/text-emotionclassification/

[2] Turney, P.D. (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 (pp. 417–424). doi:10.3115/1073083.1073153.

[3] Mayank, M. (2017, October 07). Best AI algorithms for Sentiment Analysis. Retrieved December 13, 2020, from https://www.linkedin.com/pulse/best-ai-algorithms-sentiment-analysis-muktabh-mayank

[4] Sadegh, M., Ibrahim, R., Othman, Z. A. (2012). Opinion mining and sentiment analysis: A survey. International Journal of Computers Technology, 2(3), 171–178.