EmotionX-SmartDubai_NLP: Detecting User Emotions In Social Media Text

Rahul Venkatesh Kumar, Shahram Rahmanian, Hessa Albalooshi
Smart Services Department
Smart Dubai Government Establishment
Dubai, United Arab Emirates

Abstract

This paper describes the working note on "EmotionX" shared task. It is hosted by SocialNLP 2018. The objective of this task is to detect the emotions, based on each speaker's utterances that are in English. Taking this as multiclass text classification problem, we have experimented to develop a model to classify the target class. The primary challenge in this task is to detect the emotions in short messages, communicated through social media. This paper describes the participation of SmartDubai_NLP team in EmotionX shared task and our investigation to detect the emotions from utterance using Neural networks and Natural language understanding.

1 Introduction

Emotions play a vital part in communication when people interact between each other. The exchange of emotions through text message and blog post in an informal way of writing is a bigger challenge for any machine to understand. Detecting emotions from text is widely used recently in the fields of neuroscience and cognitive services to analyze the consumer behaviors. [6] Emotion detection task is similar to analyzing the sentiment in a text. In this task we aim to detect and recognize types of feelings through the utterance such as "Neutral" "Joy" "Sadness" and "Anger". These four emotions types are related to the facial expression analysis in image recognition field. One of the most colossal challenges in determining emotion is the context dependence of emotions within the text [6]. Another challenge is linguistic co-reference, word sense disambiguation and ambiguity. Here, we describe the method and ideology of detecting the emotion from the text. The regular text classification works by stacking the text representations followed by the learned features. By considering the above discussion, our research model is given in Figure 1.
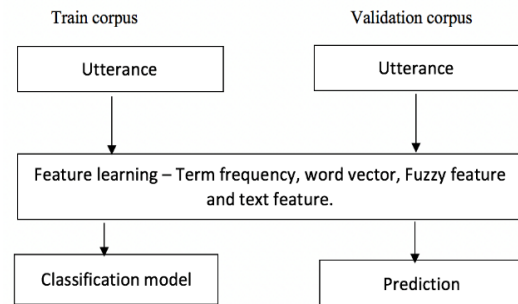

Figure 1: Experimental model

The description of task approach is explained in detail in the following sections 2. Task description, Section 3. Corpus Description, Section 4. Corpus Statistics, Section 5. Methodology Section 6. Feature Engineering, Section 7. Experiment, Section 8. Result Analysis and 9. Error Analysis and Conclusion.

2 Task Description

The given dataset consists of "Speaker", "Utterance" and "Emotion". Utterance text tagged with the emotion information, the objective is to detect the emotion information for the utterance in the validation set. The equation tag ϵ {Neutral, Joy, Sadness, Anger} and (n) represent the total number of target class in the dataset.

$$train\_corpus = \left\{ \left(utterance_1, target_1 \right), \left(utterance_2, target_2 \right), ..., \left(utterance_n, target_n \right) \right\} \quad (1)$$

$$validation\_corpus = \left\{ utterance_1, utterance_2, ..., utterance_n \right\} \quad (2)$$

3 Corpus Description

Corpus is provided by Emotion X SocialNLP 2018 shared task organizers. Training set and validation set both are in the Json format. Input utterance is annotated with target class in the training set. The training data contains total

45

9113 utterance with 4 emotion categories and validation and test data contains 1023 and 5573 utterances without target class. The sample training and validation set are mentioned in table 1.

Emotion	Utterance
neutral	If I do have it, I never used it
joy	Oh cool
sadness	Maybe your love for me doesn't fly
anger	Can't you just not put me in such situations in front of people

Table 1: Training data sample

4 Corpus Statistics

The dataset is primarily the conversation between two speakers. In training set most belongs to neutral utterance (78 percentage) and least belongs to anger utterance (1 percentage) Table 2. In training and test set utterance is mixed with digits, punctuations, and emoji's. Quality of utterance are most challenging task here because of social media content user own text scripts. In training data, we see that longer sentences belong to the neutral category and shorter sentences belong to the anger category. Using Term frequency method, we can see that there are larger number of numeric values and internet slang term like 'lol','haha' and 'idk' are used in the utterances. The utterance word count and target correlation are showed in the figure 2 using violin plot.

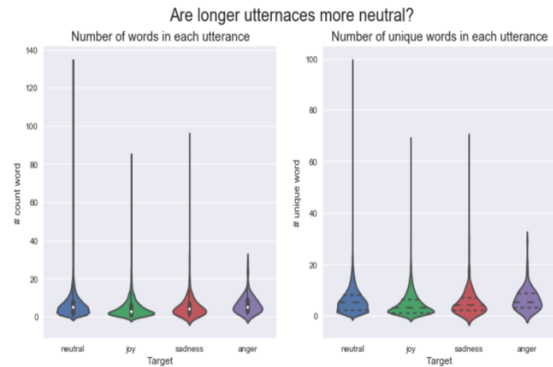|	neutral	Joy	sadness	anger
Training Set Size	7148	1482	389	94
Validation Set Size	825	160	38	9

Table 2: Corpus Count

Figure 2: Word count vs Target

5 Methodology

The figure 3 gives a picture of the architecture that we have currently implemented for this task. We are primarily focused on data pre-processing to improve the quality of utterances and also enhance feature representations. We started our approach with simple term frequency based on CNN+BiLSTM; the same methods are discussed next.
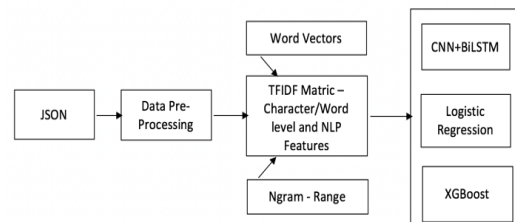

Figure 3: Architecture of our approach

6 Feature Engineering

We have improved our model accuracy on each step of stacking and modifying the features. Our whole approach is based on the architecture model figure 3. Below are the steps followed before building the model. Performance improved after data cleansing.

6.1 Data Normalization

In data normalization step we have mainly focused on cleaning the utterance to improve the performance of the model. We have created the custom list of words to replace the internet slang tokens with proper words and the same method is followed for replacing the emoji's in utterances with corresponding meaning. In addition to this, shorthand's text is replaced with proper abbreviation and mutli spaces with single space. Some of the speakers have empty utterance we have replaced the empty utterance with word "empty line". Sample data set is show in the table 3.

Utterance Before Preprocessing	Utterance After Preprocessing
lol	lots of laughs
do you even have the page liked? :)	do you even have the page liked? smile
how long you been working? :-(	how long you been working? sad
i can't understand all too smart hahahahaha	i cannot understand all too smart smile

Table 3: Utterance after cleaning

6.2 Count Based model

Our first approach is the Concatenates of Word level ngram Term Frequency Inverse Document Frequency(TFIDF) and Character level ngram Term Frequency Inverse Document Frequency(TFIDF) as single feature using sklearn FeatureUnion.

6.3 NLP Features

As a part of feature engineering, we have added hand craft features from the utterances to improve our model. The sample of the features are shown in below points.

- Length of each utterance
- Number of words in each utterance
- Number of stop words in each utterance
- Percentage of unique words each
- Sentiment polarity of each utterance

7 Experiment

As a part of experiment analysis, we ran few best algorithms like Logistic regression, Support Vector with 'rbf' kernel, Multinomial Navi Bayes. In deep learning we tried with Convolutional Neural Network - Bilateral (Long Short-Term Memory) and Convolutional Neural Network - Long Short-Term Memory(LSTM) with fastText word vector. The main advantage of using neural network is that the necessity of heavy lifting on the feature engineering side is minimum. Training set is split into training and validation data with ration 0.2 in below approaches.

(1) The word count based approach is taken as baseline approach, for this we have considered the Concatenates Word level and Character level matrix as feature using Logistic regression with accuracy of 91%. We have used ngram_range = 1,1 in Word level and ngram_range = 1,3.

(2) The second approach we have used is the combination of Word and Character level ngram Term Frequency Inverse Document Frequency(TFIDF) with handcrafted NLP features using GridSearchCV on Logistic regression with accuracy of 77% and XGBClassifier with accuracy of 85%.

(3) The third approach is using neural networks with custom hyper parameters. After prepressing each utterance is given as an input to the network. CNN+BiLSTM – accuracy of 84%. CNN+LSTM with fast Text word vector – accuracy of 86%.

Result of each model run is evaluated using precision, recall, accuracy and F1 score. The result is mentioned in the table 5.

Model	Accuracy %	Precision	Recall	F1-score
Logistic Regression (Char+Word Tfidf)	0.91	0.92	0.92	0.92
Multinomial Naive Bayes (Char+Word Tfidf)	0.85	0.83	0.83	0.78
Xgboost (TFIDF+NLP Features)	0.85	0.89	0.90	0.89
CNN+BiLSTM	0.84.9	0.80	0.85	0.82
CNN+LSTM(fastText wordvector)	0.85.7	0.81	0.86	0.83

Table :5 Cross validation results with different classifiers

8 Result Analysis

The combined feature of Character level with trigram and word level ngram using logistic regression model achieved overall 91.83% accuracy for this multiclass text classification, followed by CNN-LSTM. To classify the text properly we have used custom parameters in term frequency-inverse document frequency and logistic regression. Due to imbalanced data set and social media format contents we have concentrated more on the data preparation and it help us to improved our overall accuracy and our model performance. Our approach for detecting the emotion in the text has been evaluated based on the unweighted accuracy and class wise metrics precision, recall and F1 measure scores are mentioned in the table 6.

Target Class	Precision %	Recall %	F1-score %
neutral	0.84	0.94	0.89
joy	0.82	0.77	0.79
sadness	0.94	0.96	0.95
anger	0.82	0.80	0.81
Average %	0.92	0.92	0.92

Table :6 Target wise performance analysis

We have obtained an overall accuracy of 91.83% using Stacked Tf-idf features logistic regression-based approach for EmotionX" shared task SocialNLP 2018 - Detecting User Emotions in Social Media Text.

8.1 Evaluation Result

The test data contains two files Emotion push and Friends utterances with 50571 unlabeled data and submission of labeled was evaluated by the task organizer final ranking is based on the Unweighted Accuracy mentioned on the below table 7. It was quite disappointment our model didn't perform well on the test data.

Test data	Unweighted Accuracy
Emotion Push	26.55
Friends	25.52

9 Error Analysis and Conclusion

In this paper, a supervised system for we have presented an approach to detect the emotion in speaker utterances which is in social media format. Our experimented methodology, Charterer and Word level ngram stacked feature extracted from utterances. Then the logistic regression with custom parameters is trained using extracted features. Our system is evaluated using the test utterances given EmotionX shared task organizers. We have obtained an overall accuracy of 91.83% in the training set but fails to capture generalized features and performs poorly on the test set. The major drawback is imbalanced data for training set. Another issue dealing with large amount internet slang in dataset. The system could further have improved by replacing the internet slang with proper lexical and experiment with different techniques used on the supervised approach in Machine learning.

ACKNOWLEDGMENTS

References

[1] KP Soman Rahul Venkatesh Kumar, Anand Kumar M,AmritaCEN_NLP@ FIRE 2015 Language Identification for Indian Languages in Social Media Text.

[2] Rongyu Li, Feng Zhou, Jing Wang, Xiaojian Yang, "Application of improved multiple convolution neural network in emotion polarity classification model", Chinese Automation Congress (CAC) 2017, pp. 644-649, 2017.

[3] Yuanye He,Liang-Chih Yu1,K.Robert Lai and Weiyi Liu4YZU-NLP at EmoInt-2017:Determining.

[4] Barathi Ganesh HB, Reshma U, Anand Kumar M and Soman KP,Representation of Target Classes for Text Classification.

[5] Vivek Vinayan, Naveen J R, Harikrishnan NB, Anand Kumar M and Soman KP,AmritaNLP@PAN-RusProfiling : Author Profiling using Machine Learning Techniques.

[6] Emotion Detection and Recognition from Text Using Deep Learning by CY Yam.

[7] Text Based Emotion Recognition: A Survey by Chetan R Chopade.