12/6/2020

Detecting Multi-label Sentiment and Emotions from Bangla YouTube Comments (ICBSLP-2018)

Name of Student: N. I. Md. Ashafuddula
Roll: 18204016
Program: M. Sc. Engineering in CSE Program

ABSTRACT:

Due to the wide range of practical application sentiment analysis has become a key research area in NLP that includes opinion mining, emotion extraction, trend predictions in social media, etc. But in the bangla language, a little research is done. In this paper, researchers presented a set of techniques to identify sentiment and extract emotions from Bangla texts. They built deep learning-based models to classify a Bangla sentence with a three-class (positive, negative, neutral) and a five-class (strongly positive, positive, neutral, negative, strongly negative) sentiment label and also can extract the emotion of a Bangla sentence as any one of the six basic emotions (anger, disgust, fear, joy, sadness, and surprise). They evaluated the performance using a new dataset of Bangla, English, Randomize Bangla comments from different types of YouTube videos. Their proposed approach shows 65.97% and 54.24% accuracy in three and five labels sentiment, which has more accuracy than baseline solutions and existing approaches.

INTRODUCTION:

The advent of online social networking sites such as Facebook, Twitter, and MySpace, has fueled the interest in sentiment analysis research that finds people's opinions, appraisals, evaluations, attitudes, and emotions from the text. Though the sentiment analysis has been widely studied topic in English, it is rarely studied in the context of Bangla language, detecting the only polarity is not enough for analyzing comments in micro-blogging or social sites as they often contain various degrees of sentiment and emotion information.

For Example:

1. "**This video is not up to the mark**" — Negative Sentiment

2. "**What a rubbish video it is!!!**" — Negative Sentiment, but with more negative feelings and expresses disgust

To overcome the above limitations, they aimed to build a multilabel sentiment analyzer (i.e., three and five class sentiments) and an emotion detector for Bangla and Romanized Bangla texts.

OBJECTIVE:

- ❑ To classify a Bangla sentence with a three-class and a five-class sentiment label
- ❑ Extract the emotion of a Bangla sentence as any one of the six basic emotions (anger, disgust, fear, joy, sadness and surprise).

CONTRIBUTION:

- ❑ To work with Bangla sentiment analysis they created a new Dataset of Bangla, English and Romanized Bangla comments from different types of YouTube videos.
- ❑ Their work is the state of the art solution for Bangla language as no prior work has been conducted on identifying emotion from sentences.

- They implemented Deep learing models and different baseline methods to compare experimental results and showed deep learning models outperforms baseline methods to identify emotions from sentences.

DATASET:

Dataset collected from youtube comments and annotate the data then extracted comments from different types of video domains as Table II, using YouTube API version 3.0.

They selected videos in Bangla language based on their popularity (number of views, number of likes or dislikes) from 2013 to early 2018 and limit the number of comments for each video up to 50 to remove redundancy and also exclude the replies. They Used Google translator to detect the language of each comment. Unidentified languages considered as **Romanized Bangla**.

TABLE I: Language distribution of dataset

Language	3 class	5 class	Emotion	Total
Bangla	2797	1208	1006	5011
English	2389	1050	747	4189
Romanized	3724	1628	1137	6489
Total	8910	3886	2890	15689

TABLE II: Domain distribution of Dataset

Domain	3 class	5 class	Emotion
Music Video	1402	571	440
Review Video	1231	553	346
Drama Video	1188	542	352
Funny Video	1080	483	315
Report Video	535	262	140
Sports Video	737	335	232
News Video	1122	468	378
Talkshow Video	1615	672	687

TABLE III: Label distribution in dataset

3 class		5 class		Emotion	
Positive	3104	Strongly Positive	416	Anger/Disgust	823
Neutral	2805	Positive	843	Joy	762
Negative	3001	Neutral	1222	Sadness	272
		Negative	1064	Fear/Surprise	294
		Strongly Negative	341	None	739

The annotation part has been conducted by different native Bengali speakers with various background. They have created a public domain for data annotation purpose and circulate it. Comments from YouTube often contain abusive and vulgar words, slangs and personal attack. Therefore, they ensure that all annotators are adults. They solve the conflict of multiple labels of each sentence by taking majority votes. Table IV shows a sample dataset.

TABLE IV: sample youtube comments

Text	Language	Domain	Classification	Label
শেষের কাহিনী টা এরকম না করলেই ভালো হতো। *(The finishing of the story could be better if it does not happen)*	Bangla	Drama Video	3 class	Negative
vi amazing video make koren oswam *(Bro, You make amazing video. Awesome.)*	Romanized	Review Video	5 class	Strongly positive
They are playing really very nice well-done girls	English	Sports Video	Emotion	Joy
ভাই এতো ভুল ইনফরমেশন দেন কেন!আজেবাজে নিউজ সব *(Bro, why do you so much wrong information! all ridiculous news.*	Bangla	News Video	Emotion	Anger/Disgust
my favorite song is it	English	Music Video	3 class	Positive
2 mohila khali hase ke? mejajta kharap hoia jay. *(Why these two women are laughing? The mood gets worse)*	Romanized	Talkshow Video	5 class	Negative

IMPLEMENTATION PROCESS:

Preprocessing:

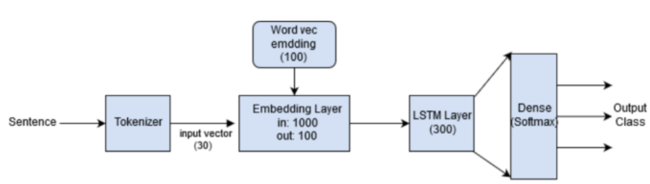
- They Removed noise, duplication, errors, un-necessary information, links, urls, user tags and mentions from comments.
- They Tokenized each sentence and removed stop-words from them.

Word Embedding Representation

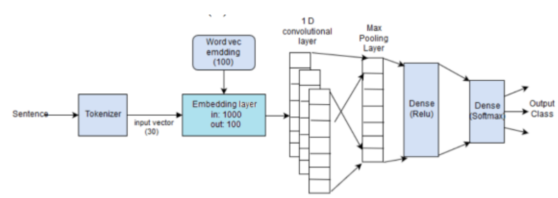
- To represent each word in a sentence as vector, they used Word2Vec.
- They Used both Continuous Bag Of Words (CBOW) and Skip Gram (SG)
- Created a vocabulary of size D from text corpus.
- Each sentence in the corpus is transformed into a one hot encoding vector of length D and feed forward these vectors to a Neural Network.
- Neural network consists of 1 hidden layer with M nodes. Output layer has D nodes where each node denotes probability. (in this paper, M=100 used)

Model Architecture:

- They used Long Short Term Memory (LSTM), Convolutional Neural Network (CNN) and to evaluate performance they used other baseline methods (SVM, Naïve Bayes).



(a) LSTM method

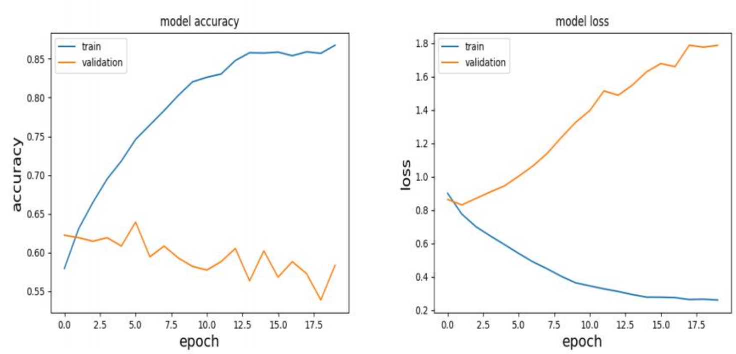


(b) CNN method

Fig[1]: Architecture for sentiment and emotion classification

1) **LSTM:** After necessary prepossessing sentences are passed through a tokenizer to produce a one-hot encoding vector of length 30 as most of the YouTube comments are short. They only consider the top 1000 most frequent words in the vocabulary and skipped the sentences that are more than 30 words long and pad with zeros for shorter comments. Then these vectors are feed into an embedding layer and the weights are initialized with word2vec embedding eights. The output dimension of the embedding layer is 100 as it is the vector length of each word in the word2vec model. The sequence of 30 words is then fed into an LSTM layer. Finally, they added a dense layer with softmax as activation function since each sentence can belong to only one class in this scenario. The number of nodes in the dense layer is equal to the number of classes in a specific problem.

2) **CNN:** After the embedding layer, they added a 1D convolutional layer with 100 filters. The next global max-pooling layer extracts the maximum value from each filter and the output dimension is a just one-dimensional vector with length as same as the number of filters they applied. This vector is directly passed to a dense layer (Relu activation) without any filtering. The final output layer is a softmax layer with several labels as an output node.



(a) Accuracy changing over epoch number
(b) Loss changing over epoch number

Fig[2]: Effect of epoch number in train and validation accuracy and loss

EXPERIMENTAL EVALUATION:

- The Compared the performance of their solutions with baseline SVM and Naive Bayes(NB) approach.

- Fig. 2 shows the accuracy increases for training set as epoch number increases but loss in validation set increase also. Therefore, the problem of over-fitting prevails in their approach.
- Table V shows both LSTM and CNN out performs baseline SVM and NB approach in all classification scenario
- LSTM is slightly better than CNN is most of the cases but CNN is much faster.

TABLE V: Performance measure of different approaches

Method	3 class		5 class		emotion	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
LSTM	**0.659664**	**0.63532**	**0.54242**	**0.5320**	**0.59230**	**0.5290**
CNN	0.6089	0.6052	0.521	0.52086	0.5403846	0.53465
NB	0.60791	0.5947603	.46880290	0.4802	0.5251	0.52473
SVM	0.5918542	0.589046	0.44876	0.465272	0.4926	0.4981

TABLE VI: Performance of different wordvec model

Vectorization model	3 class		5 class		Emotion	
	Accuracy	F1 score	Acc	F1	Acc	F1
No embedding vector	0.614525	0.610142	0.5090	0.5014237	0.539230	0.51725
Continuous bag of words(cbow) (trainable)	0.622	0.62	0.51030	0.50109	0.5403846	0.53465
Continuous bag of words(cbow) (nontrainable)	0.62290	0.60425	0.50606	0.49993	0.53038	0.50461
Skip gram(SG) (trainable)	**0.6592**	**0.6550**	0.53818	0.517733	**0.59230**	0.5290
Skip gram(SG) (nontrainable)	0.639664	0.63532	**0.54242**	**0.5320**	0.5592	**0.53534**

DISCUSSION:

- The highest achievable accuracy for 3 and 5 class sentiment analysis is **65.97% and 54.24%** respectively. For five category emotion detection, the accuracy is 59.23%.
- Between CBOW and SG, Skip-Gram (SG) model provides highest accuracy and F1 score in all the cases.

REPRODUCTION of EXPERIMENT:

- To reproduce the experiment system must need **Nvidia GTX 960M with 4GB memory** to run Deep Learning algorithms.
- Due to the Lackings of **Nvidia Nvidia GTX 960M Graphics,** My laptop is unable to reproduce the experiment.

LIMITATION & OVERCOME:

- Sentiment analysis of short texts from these comments is challenging because of the limited amount of contextual data, the use informal language and the presence of a lot of mistakes in the text. Thus, rule and predefined feature based methods are not applicable for opinion mining from these texts.
  So, they used Deep Learning methods as it has impressive performance in sentiment analysis.
- They used Adam optimizer and categorical cross entropy as loss function. Fig 3 shows the overfitting problem.

To overcome from this they set the epoch number to five in all experiments and use batch size 32.

- Model performs better in detecting emotion from English text and worst performs for randomize Bangla text.
- Model accuracy decreases for news type videos as comments in this domain contain various topics and ambiguous thoughts. However, comments from review type videos score a higher accuracy in their method as they are more polarized.

BENEFIT TO THE COMMUNITY:

Though the sentiment analysis has been widely studied topic in English, it is rarely studied in the context of Bangla language, which is the one of the most widely spoken and culturally rich language, with nearly 250 million of native speakers. The total number of Internet users in Bangladesh has reached 80.829 million at the end of January, 2018[1].

- Sentiment analysis is extremely useful in social media, youtube video monitoring as it allows us to gain an overview of the wider public opinion behind certain topics.
- It gives the ability to quickly understand consumer attitudes and react accordingly.

[1]https://bit.ly/2vQejwT

CONCLUSION:

- Model for multi-label sentiment achieved at least 10% more accuracy than baseline solutions and existing approaches.
- Observed that the performance of proposed approach increases in domain or language specific texts.
- I find this paper as an interesting one because the sentiment analysis has been widely studied topic in English but rarely studied in the context of Bangla language. So, in this perspective, this has a huge impact on analysis sentiment in Bangla language.
- The analysis focus on identifying three class and five class sentiment label which is rare in Bangla language.

FUTURE WORK:

- In future they want to include, multiple aspects and topic information in sentiment and emotion detection.

REFERENCES:

[1] A. Das and S. Bandyopadhyay, “Sentiwordnet for bangla,” Knowledge Sharing Event-4: Task, vol. 2, 2010.

[2] S. Chowdhury and W. Chowdhury, “Performing sentiment analysis in bangla microblog posts,” in Informatics, Electronics & Vision (ICIEV), 2014 International Conference on. IEEE, 2014, pp. 1–6.

[3] M. S. Islam, M. A. Islam, M. A. Hossain, and J. J. Dey, “Supervised approach of sentimentality extraction from Bengali facebook status,” in Computer and Information Technology (ICCIT), 2016 19th International Conference on. IEEE, 2016, pp. 383–387.

[4] A. K. Paul and P. C. Shill, “Sentiment mining from bangla data using mutual information,” in Electrical, Computer & Telecommunication Engineering (ICECTE), International Conference on. IEEE, 2016, pp. 1–4.

[5] R. Feldman, “Techniques and applications for sentiment analysis,” Communications of the ACM, vol. 56, no. 4, pp. 82–89, 2013.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.

[7] Y. Kim, “Convolutional neural networks for sentence classification,” arXiv preprint arXiv:1408.5882, 2014.

[8] C. dos Santos and M. Gatti, “Deep convolutional neural net-works for sentiment analysis of short texts,” in Proceedings of COLING 2014, the 25th International Conference on Computational inguistics: Technical Papers, 2014, pp. 69–78.

[9] Y. Zhang and B. Wallace, “A sensitivity analysis of (and prac-titioners’ guide to) convolutional neural networks for sentence classification,” arXiv preprint arXiv:1510.03820, 2015.

[10] H. Shirani-Mehr, “Applications of deep learning to sentiment analysis of movie reviews,” in Technical Report. Stanford University, 2014.

[11] P. Ekman, “An argument for basic emotions,” Cognition & emotion, vol. 6, no. 3-4, pp. 169–200, 1992.

[12] P. K. Bhowmick, “Reader perspective emotion analysis in text through ensemble based multi-label classification framework,” Computer and Information Science, vol. 2, no. 4, p. 64, 2009.

[13] V. K. Singh, “Sentiment analysis research on bengali language texts,” International Journal of Advanced Scientific Research Development (IJASRD), vol. 02, pp. 122–127, 2015.

[14] M. Al-Amin, M. S. Islam, and S. D. Uzzal, “Sentiment analysis of bengali comments with word2vec and sentiment information of words,” in Electrical, Computer and Communication Engineering (ECCE), International Conference on. IEEE, 2017, pp. 186–190.

[15] A. Hassan, M. R. Amin, N. Mohammed, and A. Azad, “Sentiment analysis on bangla and romanized bangla text (brbt) using deep recurrent models,” arXiv preprint arXiv:1610.00369, 2016.

[16] D. Das and S. Bandyopadhyay, “Developing bengali word net affect for analyzing emotion,” in International Conference on the Computer Processing of Oriental Languages, 2010, pp. 35–40.

[17] R. E. Jack, O. G. Garrod, and P. G. Schyns, “Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time,” Current biology, vol. 24, no. 2, pp. 187–192, 2014.

[18] T. Pranckevičius and V. Marcinkevičius, “Comparison of naïve bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification,” Baltic Journal of Modern Computing, vol. 5, no. 2, p. 221, 2017.