Aspect Based Sentiment Analysis

A. Introduction

Aspect Based Sentiment Analysis is a kind of text analysis that comes under Natural Language Processing.This sort of analysis is used to categorize opinions in the form of text which are aspects and subsequently associate the sentiment with each aspect.The aim of this project is to categorize aspects of the target and the sentiment associated with it.

The Dataset taken for this analysis is split into aspects followed by extracting the information, which is in the form of sentiments.Aspect based sentiment analysis has the potential to emerge to be an essential learning task in both the commercial and academic spheres.

As it deals with the polarity of any sentence, it can be used by organizations to analyse and categorize data, automate tasks like customer support or feedback to get valuable insights.In this project, our aim is to categorise aspects of the target and the sentiment associated with it.

B. Related work

Aspect based Sentiment analysis has attained quite some interest in academia and in industry,particularly for pinning down satisfaction of customers on products and services. Initial study in the field of sentiment analysis aimed exclusively on analyzing the overall sentiment of a given text. [3] [5] The datasets for this task were based on particular review platforms such as Yelp where it is assumed that one specific entity is considered in one review snippet, However the opinion on multiple aspects can be expressed.This task is useful as one can analyze the combined sentiment for individual aspect of a product or service and get more comprehensive understanding of its quality.Another line of research in this field is target dependent sentiment analysis. Targeted sentiment analysis looks into the classification of sentiment polarities with regards to certain target entity mentioned in sentences. [4] [6].Several models have been put forward for recovering sentiment polarity, most of them are composed on LSTM layers, through which sentiment information is retrieved from word embedding. [2] In parallel, within NLP, there have been numerous developments in the field of pre-trained language models, for example Glove and BERT. [1]

The dataset for such analysis is extracted from a question answering platform urban neighbourhoods are discussed by users, where strong baselines where build on logistic regression model and state of the art recurrent neural network. [5]

C. Dataset and Evaluation

C.1. Dataset

For aspect based sentiment analysis task we took the Sentihood dataset,which is extracted from a question answering platform where the urban neighbourhoods are discussed by the participants. The dataset includes Three files which is used for training, testing, and development.Each of these data have 1356, 685 and 341 samples respectively.

C.2. Feature Exatraction

For feature extraction we have constructed auxiliary sentences and used pre-trained model like BERT, Glove and word2Vec for embedding and converted the Aspect based sentiment analysis into a sentence-pair classification task.

C.3. Evaluation Matrix

To analyze the performance of our classification model on a set of test data for which the true values are known, F1 score, precision and recall has been measured.

D. Methodology

We have performed exploratory data analysis on the Sentihood dataset and likewise performed preprocessing on four significant aspects out of twelve, which consists of 48 percent of our training set. In preprocessing we have constructed auxiliary sentences using target entity, aspects and sentiments and concatinated it with the main sentence. Followed by extracting features using BERT, Word2Vec and Glove embedding which are pre-trained models which we used for the ABSA task.

The machine learning models which we used were logistic regression with embedding with BERT, word2vec and glove pre-trained models among which logistic regression with glove embedding embedding is taken as a baseline model and logistic regression, RNN and BERT with BERT embedding are taken as advaced models. Followed by evaluation matrix, confusion matrix, precission recall and F1 score was calculated for evaluating our model

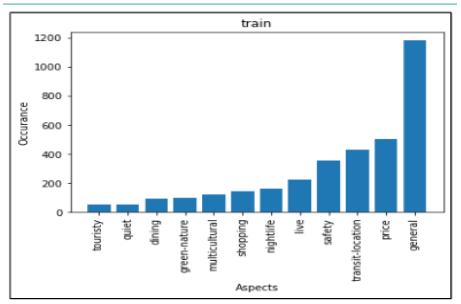

Figure 1. Exploratory data analysis of data showing the frequencies of aspects of 12 aspects, out of which 4 are taken for training

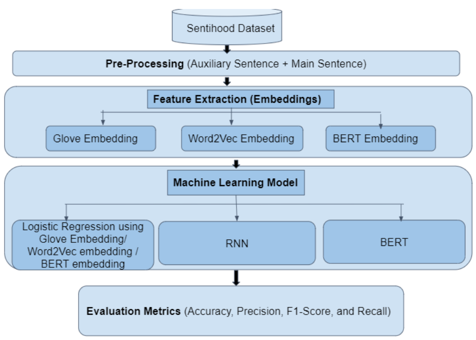

Figure 2. Framework for ABSA task

The challenge we faced was that we had non-contextual embeddings and after using them, we were not able to get a balanced F1 score between both the classes. For overcoming the challenge, we have used the contextual embeddings using bert and found that F1 score of both the classes was closer in 5-Folds validation on train and dev sets. We have shown the results in Figure 8

E. Results and Analysis

On the basis of the performance of the models that have been taken logistic regression glove,word2Vec which are baseline and logistic regression with BERT, RNN and BERT advanced model showed better accuracy, then the baseline. The Accuracy, precision, F1, Recall measures of BERT model was found to be superior to other models. Bert is context based transformer which was used d for fine tuning hence it showed performance of learning and aspect based sentiment task.

From the reviews, plenty of different aspects are procured and trained to the state of the art LTSM-RNN classifier. The accuracy is seen to be around 76 percent which is higher than the existing algorithms.On contrary to our BERT logistic regression model with BERT with 90 percent on test set.

The class-wise analysis is done in appendix

Figure 3. Training loss: BERT

	precision	recall	f1-score	support
0	0.63	0.63	0.63	9901
1	0.44	0.44	0.44	6539
accuracy			0.55	16440
macro avg	0.54	0.54	0.54	16440
weighted avg	0.55	0.55	0.55	16440

Figure 4. Logistic regression using Glove

	precision	recall	f1-score	support
0	0.63	0.63	0.63	4939
1	0.43	0.43	0.43	3245
accuracy			0.55	8184
macro avg	0.53	0.53	0.53	8184
weighted avg	0.55	0.55	0.55	8184

Figure 5. Logistic regression using Word2Vec

	precision	recall	f1-score	support
0	0.92	0.91	0.92	4939
1	0.87	0.88	0.88	3245
accuracy			0.90	8184
macro avg	0.90	0.90	0.90	8184
weighted avg	0.90	0.90	0.90	8184

Figure 6. Logistic regression using BERT

[LibSVM]	precision	recall	f1-score	support
0	0.89	0.95	0.92	9901
1	0.92	0.83	0.87	6539
accuracy			0.90	16440
macro avg	0.91	0.89	0.90	16440
weighted avg	0.90	0.90	0.90	16440

Figure 7. Logistic regression using RNN

	precision	recall	f1-score	support
0	0.99	0.99	0.99	9901
1	0.98	0.98	0.98	6539
accuracy			0.98	16440
macro avg	0.98	0.98	0.98	16440
weighted avg	0.98	0.98	0.98	16440

Figure 8. BERT

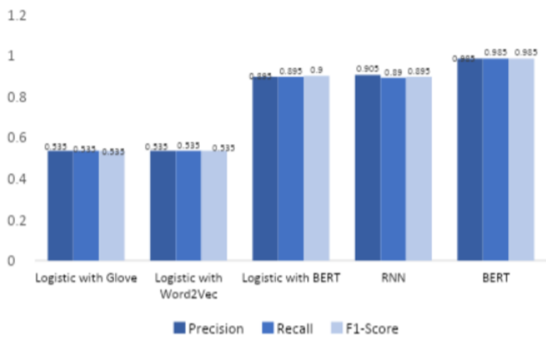

Figure 9. Comparative analysis of the three models

F. Contributions

Each team member did their part of literature review, primarily worked on three separate algorithms. Logistic Regression - By Rajat Talukdar RNN- By Kirti Lakra BERT- By Mann Khatri Each team member did exploratory data analysis and data preprocessing. Along with that each member trained their respective models and evaluated the models using various evaluation metrics.

G. Appendix

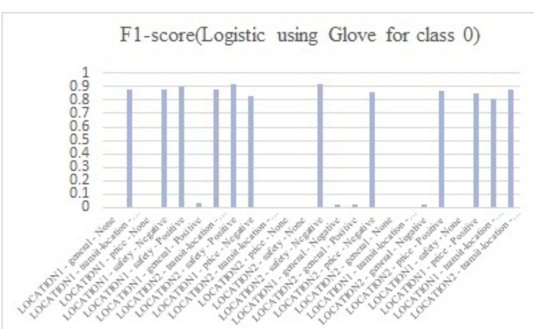

Figure 10. F1 score vs auxiliary sentences of class 0 plot for Logistic regression using Glove embeddings

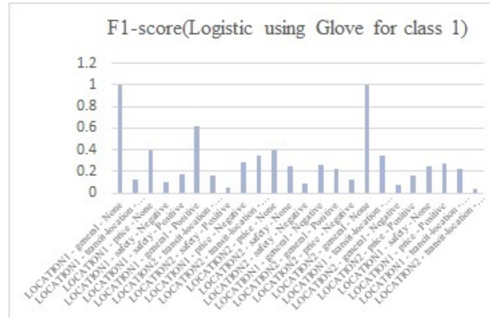

Figure 11. f1 score vs auxiliary sentences of class 1 plot for Logistic regression using Glove embeddings

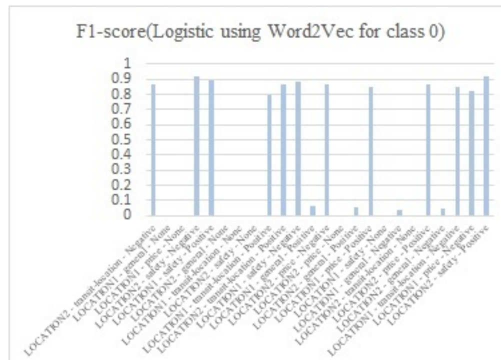

Figure 12. F1 score vs auxiliary sentences of class 0 plot for Logistic regression using Word2Vec embeddings

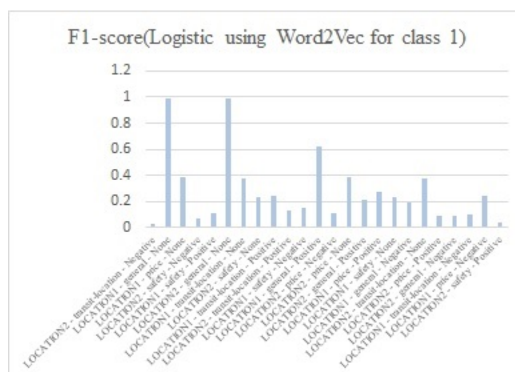

Figure 13. F1 score vs auxiliary sentences of class 1 plot for Logistic regression using Word2Vec embeddings

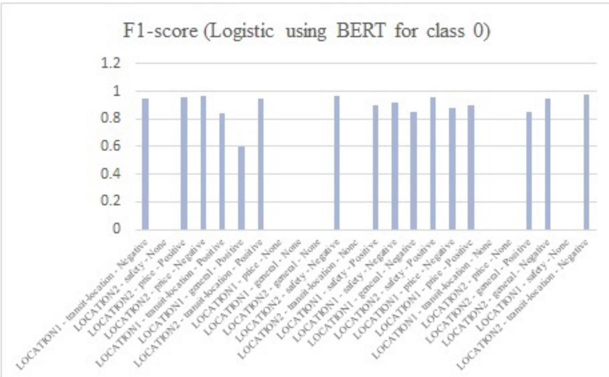

Figure 14. F1 score vs auxiliary sentences of class 0 plot for Logistic regression using BERT embeddings

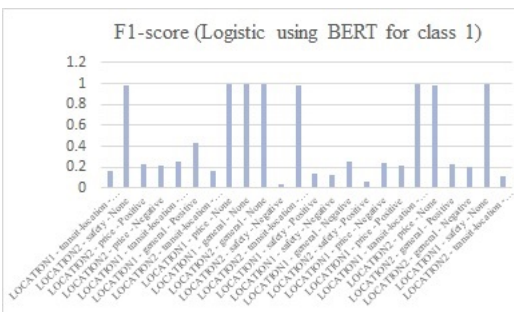

Figure 15. F1 score vs auxiliary sentences of class 1 plot for Logistic regression using BERT embeddings

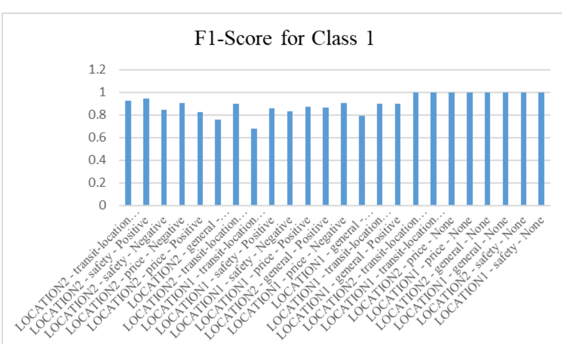

Figure 16. F1 score vs auxiliary sentences of class 1 plot for BERT

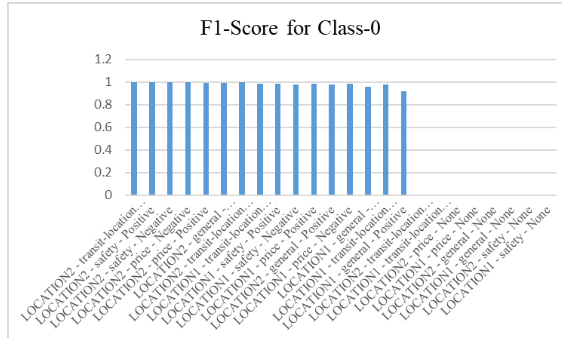

Figure 17. F1 score vs auxiliary sentences of class 0 plot for BERT

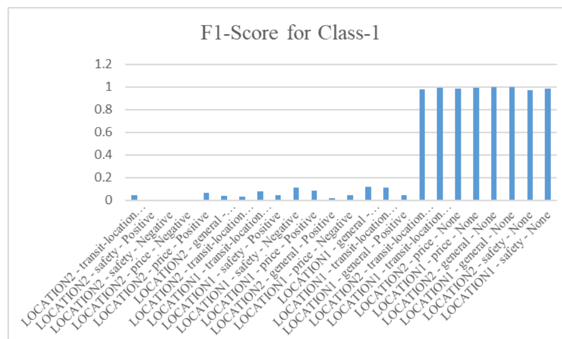

Figure 18. F1 score vs auxiliary sentences of class 1 plot for RNN

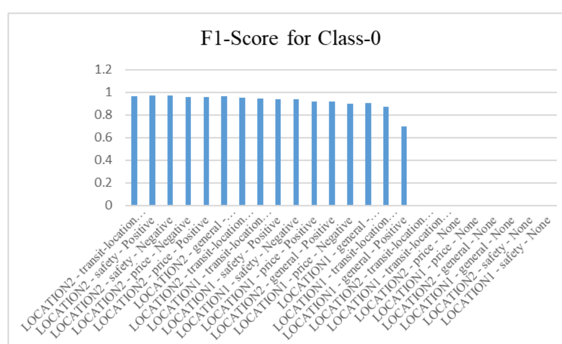

Figure 19. F1 score vs auxiliary sentences of class 0 plot for RNN

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 1

[2] Mickel Hoang, Oskar Alija Bihorac, and Jacobo Rouces. Aspect-based sentiment analysis using bert. In *Proceedings of the 22nd nordic conference on computational linguistics*, pages 187–196, 2019. 1

[3] Minqing Hu and Bing Liu. Opinion extraction and summarization on the web. In *Aaai*, volume 7, pages 1621–1624, 2006. 1

[4] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 151–160, 2011. 1

[5] Marzieh Saeidi, Guillaume Bouchard, Maria Liakata, and Sebastian Riedel. Sentihood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. *arXiv preprint arXiv:1610.03771*, 2016. 1

[6] Meishan Zhang, Yue Zhang, and Duy-Tin Vo. Gated neural networks for targeted sentiment analysis. In *Thirtieth AAAI conference on artificial intelligence*, 2016. 1