"SOMETHING SOMETHING HOTA HAI!" AN EXPLAINABLE APPROACH TOWARDS SENTIMENT ANALYSIS ON INDIAN CODE-MIXED DATA

Aman Priyanshu*[1], Sudarshan Sivakumar*[2], Supriti Vijay*[2], Nipuna Chhabra*[2], Aleti Vardhan*[2]

[1]Department of Information and Communication Technology,
[2]Department of Computer Science and Engineering,
Manipal Institute of Technology,
Manipal, India-576104.

Introduction

The increasing use of social media sites in countries like India has given rise to large volumes of code-mixed data. Sentiment analysis of this data can provide integral insights into people's perspectives and opinions. Code-mixed data is often noisy in nature due to multiple spellings for the same word, lack of definite order of words in a sentence, and random abbreviations.

Interpreting a model's predictions allows us to determine the robustness of the model against different forms of noise. In this paper, we propose a methodology to integrate explainable approaches into code-mixed sentiment analysis. By interpreting the predictions of sentiment analysis models we evaluate how well the model is able to adapt to the implicit noises present in code-mixed data.

Key Ideas

Sentiment Analysis of Code-Mixed Data
Sentiment analysis has been an important aspect of natural language processing since its inception. Its implementation on Code-Mixed data has been explored recently. Traditional machine learning approaches such as Support Vector Machines, logistic regression, and random forests have been shown to give significant performance on these datasets [1]. On the other hand, deep learning models such as LSTMs and CNNs perform comparatively better as they efficiently process sequential data [2, 3, 4]. Sub-word and bi-gram based models further improve classification performance[5].

Explainable AI approaches to sentiment analysis
In order to trust a model's predictions, one must be able to interpret the reasons behind its decisions. Several attempts have been made to infer the text classification and sentiment analysis predictions of language models. State of the art model-agnostic explanations like LIME (Local Interpretable Model-Agnostic Explanations) [6] and SHAP (SHapley Additive exPlanations) [7] enable better visualisation and analysis of AI models.

Evaluated Models

Models	SVM	Random Forest	CNN	CNN Glove Embeddings
Train Accuracy	87.46%	96.93%	53.17%	58.02%
Test Accuracy	56.22%	54.16%	54.08%	56.98%

Models	LSTM	LSTM Glove Embeddings	XLM-RoBERTa
Train Accuracy	61.00%	66.08%	85.30%
Test Accuracy	49.48%	59.09%	72.21%

In our venture to integrate explainable AI on code-mixed data, we evaluated multiple models for performance comparison. We provide the resultant performances of each of these in Table 1. Our comparison spanned tree-based random forest, SVMs, LSTMs, CNNs, and a RoBERTa model pre-trained on codemixed data. For the random forest and SVM model, we used a TF-IDF vectorizer as our word representation. While for the LSTM and CNN architecture, we compare randomly initialized word embeddings and pre-trained GloVe embeddings. These GloVe embeddings were pre-trained on the Twitter crawl, which consisted of about two billion tweets.

Data

The SAIL 2017 dataset [8] is a sentiment classification dataset on code-mixed text, with each instance labelled as positive, negative, or neutral in sentiment. This data was collected using the Twitter4j10 API to extract Hindi-English code-mixed data from Twitter. Common Hindi words were collected in Romanized format and then searched using the above API. The data was then annotated manually. It consists of a total of 12,601 instances of text, with corresponding sentiment labels. The dataset is pre-split into three subsets: training set, validation set, and test set.

Metrics

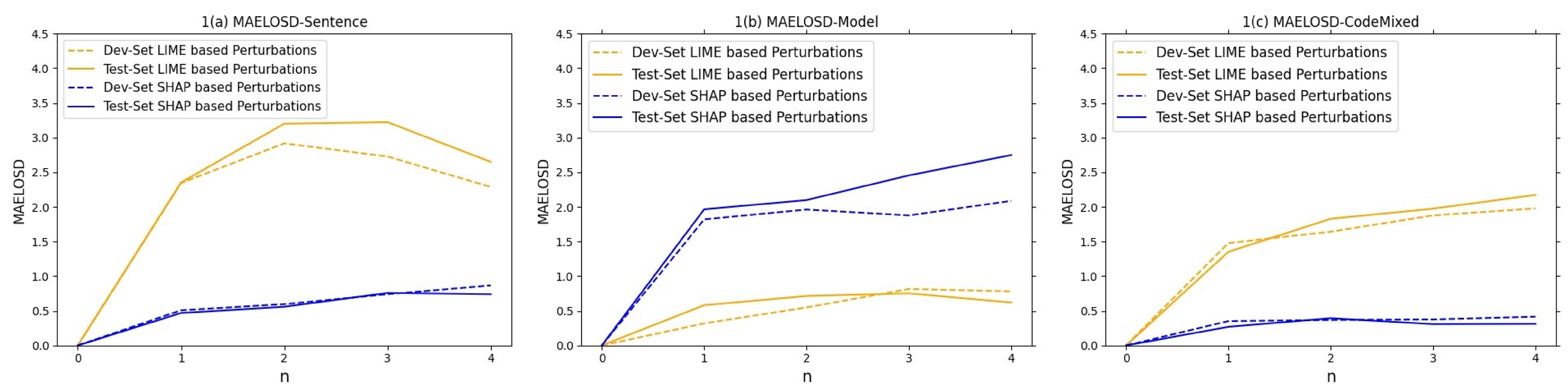
We present a comparison between LIME and SHAP using three metrics, described by Mean Absolute Error of Log-Odds Scores on Deletion (MAELOSD) which is defined as,

$$= \sum_{j=0}^{M} \frac{(|log\_odds_{ini}^{(j)} - log\_odds_{fin}^{(j)}|)}{M} \quad (1)$$

Where ini refers to initial sentences without word deletions, fin refers to the final sentences after the deletion of polarizing words, and M is the total number of sentences. We define polarizing words as those which have been given the highest weights by the explanations of the LIME and SHAP models.

1. MAELOSD of Sentence-Interpreted Polarizing Words (*MAELOSD-Sentence*)
2. MAELOSD of Model-Interpreted Polarizing Words (*MAELOSD-Model*)
3. MAELOSD of Code-Mixed Words (*MAELOSD-CodeMixed*)

Results



Conclusion

Code-mixed data is an integral part of communication in multilingual communities and their culture. The application of state-of-the-art interpretable methods on this assortment of data will pave a path towards the adoption of the same during real-world implementation. Our use of LIME and SHAP, which quantify local and global model explanations, allows us to display their importance and relevance in sentiment analysis of code-mixed data.

References

[1] Pruthwik Mishra, Prathyusha Danda, and Pranav Dhakras. Code-mixed sentiment analysis using machine learning and neural network approaches, 2018.

[2] Dejiao Niu, Zheng Xia, Yawen Liu, Tao Cai, Tianquan Liu, and Yongzhao Zhan. Alstm: adaptive lstm for durative sequential data. In *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 151–157. IEEE, 2018.

[3] Jin Wang, Liang-Chih Yu, K Robert Lai, and Xuejie Zhang. Dimensional sentiment analysis using a regional cnn-lstm model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 225–230, 2016.

[4] Yoon Kim. Convolutional neural networks for sentence classification, 2014.

[5] Ameya Prabhu, Aditya Joshi, Manish Shrivastava, and Vasudeva Varma. Towards sub-word level compositions for sentiment analysis of hindi-english code mixed text, 2016.

[6] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier, 2016.

[7] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017.

[8] Kamal Sarkar. Ju_ks@sail_codemixed-2017: Sentiment analysis for indian code mixed social media texts, 2018.

Contact information:
Name Priyanshu Aman;
aman.priyanshu@learner.manipal.edu