How Effective is Incongruity? Implications for Code-mix Sarcasm Detection

Aditya Shah and Chandresh Kumar Maurya

Department of Computer Science and Engineering
Indian Institute of Technology Indore

December 18, 2021

Overview

Introduction

- Sarcasm is defined as a sharp remark whose intended meaning is different from what it looks like. For example, "*I am not insulting you. I am describing you.*"
- Sarcasm usually involves ambivalence (also known as *incongruity* which means words/phrases having contradictory implications.
- Though English is used as a way to communicate and exchange messages, majority of the people still use the mother language to express themselves on social media [1].
- According to one study [2], more than 50% posts on Twitter are written in a language other than English.
- *Code-switching* (also known as *code-mixing*) is a writing style in which the author uses words from different languages either in the same sentence (called *intra–sentential*) or different sentences (called *inter–sentential*) switching. Example: *"he said kal karte hai kaam""* (Gloss: he said tomorrow we'll do the work).

Introduction

Though many studies exist to detect sarcasm in unimodal and multimodal data [3, 4, 5, 6, 7], methods to detect sarcasm in code-mix data are limited and have not been explored much [8, 9, 10].

The Problem

To detect sarcasm in code-mix data through incongruity capturing

Challenges

- Ambiguous words
- Variable lexical representation
- Word-level code-mixing
- Reduplication
- Word-order

Contributions

- Propose a deep learning based architecture along with sub-word level features to capture incongruity for sarcasm detection.
- Evaluate the performance of the proposed model on the Hindi-English (Hinglish) code-mix Twitter data that we collected. We further analyze existing multilingual models on the same. Our code+data will be available on [1].
- We will release the benchmark sarcasm dataset for Hinglish language to facilitate further research on code-mix NLP.

[1] https://github.com/likemycode/codemix

Literature Survey

- **Learning Representation for Code-Mix Data**
  - Multilingual representation learning have been explored for code-mix data representation in [11, 12, 13, 14].
  - Character-level representations have been utilized to address the out-of-vocabulary (OOV) issue in [14],
  - Hand-crafted features were used in [15] for handling low-resource scenarios.
  - Fine-tuning multilingual models like mBERT has shown to yield good results for various NLP tasks in [12], and surprisingly outperforms cross-lingual embeddings.
  - A centralized benchmark for Linguistic Code-switching Evaluation (LinCE) is released in [13, 12].
- **Sarcasm Detection in Code-Mix Data**
  - In [9], the author experiments with FastText [16] and Word2Vec embeddings on two kinds of data: (1) Hinglish (Hindi-Eng) tweets, and (2) Hinglish+English tweets and achieves best F1 score of 79.4%.
  - [8] finds that switching feature is a good indicator for irony/sarcasm/hate speech detection. However, none of these works handles *incongruity* explicitly or implicitly which has shown to achieve impressive results in sarcasm detection [6].

Model Architecture

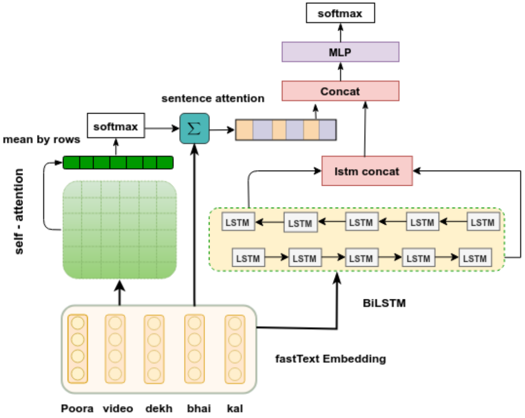


Figure 1: The model architecture

Dataset

- The code-mix dataset used by [9] is highly imbalanced with just 10% of sarcastic tweets and rest non-sarcastic.
- So, we create a dataset using TweetScraper built on top of scrapy [2] to extract code-mix hindi-english tweets.
- We pass search tags like #sarcasm, #humor, #bollywood, #cricket, etc., combined with most commonly used code-mix Hindi words as query.
- All the tweets with hashtags like #sarcasm, #sarcastic, #irony, #humor etc. are treated as positive.
- Non sarcastic tweets are extracted using general hashtags like #politics, #food, #movie, etc.
- The balanced dataset comprises of 166K tweets. We preprocess and clean the data by removing urls, hashtags, mentions, and punctuation in the data.

[2] https://github.com/jonbakerfish/TweetScraper

Baselines

- **Attention BiLSTM** [9]: The text features are extracted using word2vec and FastText which is fed to Series CNN, Parallel CNN, LSTM, Bi-LSTM and Attention Bi-LSTM.
- **Multilingual Models**: To showcase the competitiveness of the proposed approach, we also compare with the state-of-the-art multilingual models like XLM-RoBERTa[3] and mBERT[4] from Huggingface library [17].
- Specifically, we first fine-tune these models on the preprocessed code-mix corpus for mask language modeling task. Next, we use trained model by attaching a dense layer on top of it for detecting sarcasm in the code-mix tweets.

[3] https://tinyurl.com/ydseww9d
[4] https://tinyurl.com/2dafn48n

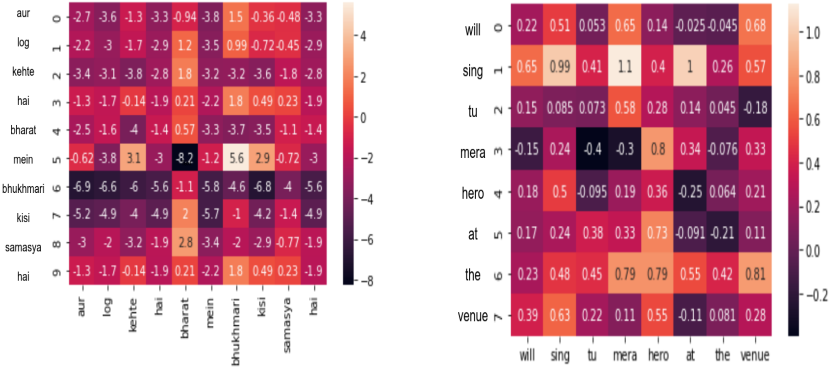
Experimental setup

- We use a train/valid/test split of 65:15:20.
- Categorical cross-entropy loss is minimized using adam optimizer.
- 15 epochs
- Learning rate of 5e-4 with step wise learning rate scheduler.
- FastText embedding size is 100 and the number of hidden units in BiLSTM and MLP layers are 256.
- Apply dropout of 0.4 along with gradient clipping of 0.3.

Results

Table 1: Comparative evaluation of the proposed approach.

Model	**Recall**	**Prec.**	**Acc.**	**F1**	**Params**	**GPU**	**Train time**
Attn. BiLSTM	77.34	81.24	80.21	79.34	21M	68 MB	0.8Hr
XLM-RoBERTa	86.17	91.48	89.04	88.75	278M	575 MB	8 Hr
mBERT	83.20	**94.55**	**89.17**	88.51	167M	483 MB	7 Hr
SelfNet (Ours)	**88.12**	88.25	89.04	**88.89**	35M	80 MB	1 Hr



(a) Sarcastic (b) Non-sarcastic

Figure 2: Incongruity visualization via attention matrix

Ablation Study

Table 2: Ablation Study

Models	Recall	Prec.	Acc.	F1
Self Matching Net	81.68	81.55	81.7	81.68
with XLM-RobertA	87.71	87.73	87.81	87.81
with mBERT	86.94	86.72	86.85	86.94
SelfNet (Ours)	**88.12**	**88.25**	**89.04**	**88.89**

Conclusion

- In the present work, we propose the significance of incongruity in order to capture sarcasm in code-mix data.
- Our model effectively captures incongruity through FastText sub-word embeddings to detect sarcasm in the text.
- Empirical results on code-mix sarcasm data show that our approach performs satisfactorily compared to the multilingual models while saving memory footprint and training time.
- In future, we plan to work on a generalized model for other code-mix NLP tasks (NLI, NER, POS, QA etc) as well as test other code-mix languages like English - Spanish, English - Tamil, English - French etc.

References I

Brenda Danet, Susan C Herring, Susan C Herring, et al.
The multilingual Internet: Language, culture, and communication online.
Oxford University Press on Demand, 2007.

Lichan Hong, Gregorio Convertino, and Ed Chi.
Language matters in twitter: A large scale study.
In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, 2011.

Aditya Joshi, Vinita Sharma, and Pushpak Bhattacharyya.
Harnessing context incongruity for sarcasm detection.
In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 757–762, 2015.

References II

Aditya Joshi, Pushpak Bhattacharyya, and Mark J Carman.
Automatic sarcasm detection: A survey.
ACM Computing Surveys (CSUR), 50(5):1–22, 2017.

Paula Carvalho, Luís Sarmento, Mário J Silva, and Eugénio De Oliveira.
Clues for detecting irony in user-generated contents: oh...!! it's" so easy";-.
In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 53–56, 2009.

Tao Xiong, Peiran Zhang, Hongbo Zhu, and Yihui Yang.
Sarcasm detection with self-matching networks and low-rank bilinear pooling.
In *The World Wide Web Conference*, pages 2115–2124, 2019.

References III

Yitao Cai, Huiyu Cai, and Xiaojun Wan.
Multi-modal sarcasm detection in twitter with hierarchical fusion model.
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2506–2515, 2019.

Srijan Bansal, Vishal Garimella, Ayush Suhane, Jasabanta Patro, and Animesh Mukherjee.
Code-switching patterns can be an effective route to improve performance of downstream nlp applications: A case study of humour, sarcasm and hate speech detection.
arXiv preprint arXiv:2005.02295, 2020.

Akshita Aggarwal, Anshul Wadhawan, Anshima Chaudhary, and Kavita Maurya.
" did you really mean what you said?": Sarcasm detection in hindi-english code-mixed data using bilingual word embeddings.
arXiv preprint arXiv:2010.00310, 2020.

- Sahil Swami, Ankush Khandelwal, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava.
  A corpus of english-hindi code-mixed tweets for sarcasm detection.
  arXiv preprint arXiv:1805.11869, 2018.

- Genta Indra Winata, Samuel Cahyawijaya, Zihan Liu, Zhaojiang Lin, Andrea Madotto, and Pascale Fung.
  Are multilingual models effective in code-switching?
  arXiv preprint arXiv:2103.13309, 2021.

- Simran Khanuja, Sandipan Dandapat, Anirudh Srinivasan, Sunayana Sitaram, and Monojit Choudhury.
  Gluecos: An evaluation benchmark for code-switched nlp.
  arXiv preprint arXiv:2004.12376, 2020.

References V

Gustavo Aguilar, Sudipta Kar, and Thamar Solorio.
Lince: A centralized benchmark for linguistic code-switching evaluation.
arXiv preprint arXiv:2005.04322, 2020.

Genta Indra Winata, Chien-Sheng Wu, Andrea Madotto, and Pascale Fung.
Bilingual character representation for efficiently addressing out-of-vocabulary words in code-switching named entity recognition.
arXiv preprint arXiv:1805.12061, 2018.

Gustavo Aguilar, Suraj Maharjan, Adrian Pastor López-Monroy, and Thamar Solorio.
A multi-task approach for named entity recognition in social media data.
arXiv preprint arXiv:1906.04135, 2019.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov.
Bag of tricks for efficient text classification.
arXiv preprint arXiv:1607.01759, 2016.

References VI

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew.
Huggingface's transformers: State-of-the-art natural language processing.
CoRR, abs/1910.03771, 2019.

Thank You!