Sarcasm Detection_Bert_LSTM_SVM

Mentor: Zhiyuan Qin
Ratna, Xin Wei Chia, Irene Nandutu
Pod 475_Gigantspinosaurus_Mashed

neuromatch academy

Me :- I am depressed.

Parents :- Don't be.

What is Sarcasm?

- Sarcasm is a form of communication that **conveys the opposite of its literal meaning**.
- It often involves using irony or a mocking tone to express humor, criticism, or contempt.
- Sarcasm is prevalent in social media and online communication, where it can be used for various purposes, including humor, satire, and irony.

Difficulty of Sarcasm Detection in Text

- **Absence of tone and facial expressions (**nonverbal cues)
- Involves **individual, cultural** and **societal** cues
- Involves **contextual** factors that ties in with specific lexical features. E.g. “Yeah, right” (short and informal language)
- Evolving language

⇒ This means that tokens in sarcastic and non-sarcastic are often **closely embedded** in representation space


Donald J. Trump
@realDonaldTrump · Follow
When I looked up to the sky and jokingly said "I am the chosen one," at a press conference two days ago, referring to taking on Trade with China, little did I realize that the media would claim that I had a "Messiah complex." They knew I was kidding, being sarcastic, and just....
8:30 PM · Aug 24, 2019

Questions

1. How do Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT) models overcome these challenges to effectively detect sarcasm in text?

1. How does one model perform better than the others?

Hypothesis

We hypothesise that well-constructed models are able to detect sarcasm when provided with a large dataset

Datasets

News headline dataset from Reddit (taken from Kaggle)

1. Misra, Rishabh and Prahal Arora. "Sarcasm Detection using News Headlines Dataset." AI Open (2023).

Sample size = (a) 5000 (b) The entire dataset (70% Training, 30% Test)

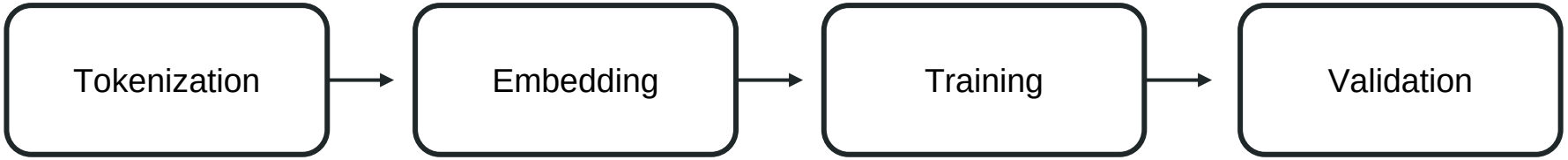
Binary labels: 0 = Sarcastic, 1 = Not sarcastic

Example text:

0: *“scott used to stop breathing nearly 40 times an hour. this device changed his life”*

1: *“courtroom sketch artist has clear manga influences”*

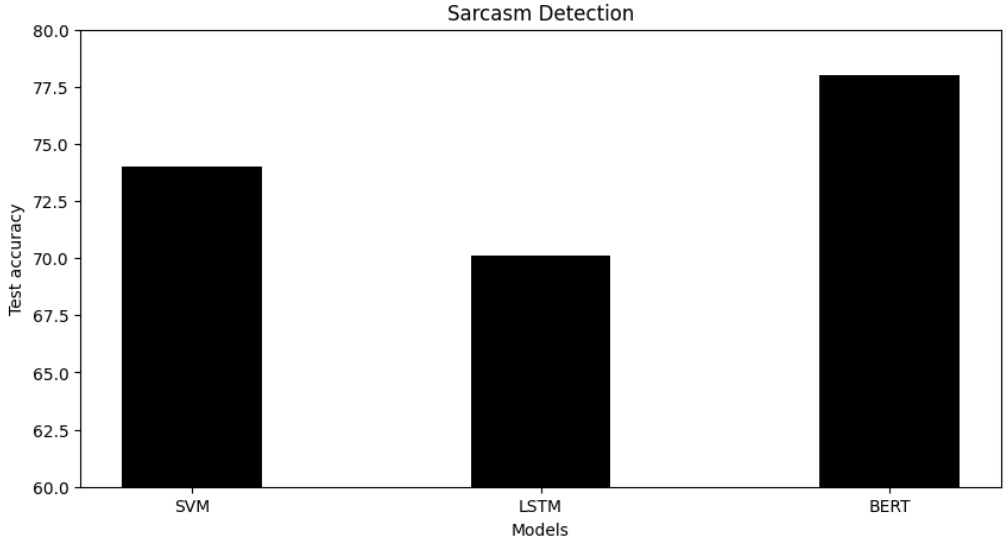
General pipeline



Model Architecture

SVM	LSTM	BERT
Linear kernel **C** = 1 Random state = 42 **Max interactions** = 10000	**Hidden layers** = 2	**Hidden layers** = 8 + single layer binary classifier **Attention heads** = 12 **Immediate size** = 3072

Results



Why did BERT model perform better than other?

1. BERT - context sensitive embeddings
2. Others - context free embeddings
3. Attention
4. Parallel Processing

Data Augmentation

1. GPT-2 for data generation to increase the dataset size

1. The augmented dataset increased the accuracy of SVM, reaching 93%

1. This outcome highlights the significance of data augmentation strategies in the domain of sarcasm detection

1. We hypothesise that the augmented dataset will significantly improve the initial performance of both LSTM and BERT

Discussion (Reasons behind high or low accuracies of models)

More accurate ← Less accurate

Bidirectional Encoder Representations from Transformers (BERT)	SVM	LSTM
1. Includes Attention 2. Parallel Processing of whole sentence	1. Properly defines the boundaries of the embedding space 2. Good at separating vectors in close proximity embedding space	1. For Time Series/Sequence of words 2. Understands the sequences together 3. Sarcastic & Non-Sarcastic tweets: not exactly defined by word sequences but by the kind of words and context

Limitations of the study

- Different pre-processing methods used
- Different hyperparameters used. E.g. epochs, number of layers, etc.
- Insufficient computational resources (GPU and RAM). E.g. BERT was limited to 5000 samples and 8 hidden layers

Future Directions

- Compare the performance of different types of transformers:
  Eg: Different types of BERTs, GPT
- Larger dataset
- Data Augmentation
- Pre-processing: other tokenization and word embedding techniques. E.g. context-free vs context-sensitive
- Explore different parameters: Increase the number of hidden layers, epochs etc.

Acknowledgements

- Dr. William Chin
- Artem Pilzak
- Dr. Sandhya Tripathi

References

Hamada Nayel, Eslam Amer, Aya Allam, and Hanya Abdallah. 2021. Machine Learning-Based Model for Sentiment and Sarcasm Detection. In Proceedings of the Sixth Arabic Natural Language Processing Workshop, pages 386–389, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

P. Verma, N. Shukla and A. P. Shukla, "Techniques of Sarcasm Detection: A Review", *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 968-972, 2021.