Detecting Sarcasm

Leveraging BERT to see the obvious

Eshan Bhatnagar
Mahmoud Ghanem
Michael Kalish

April 17, 2024

Berkeley
UNIVERSITY OF CALIFORNIA

Background

X is a platform on which important and not-so-important information is circulated with the aim to influence public opinion.

Massachusetts Institute
@MI

New discovery shakes scientific community:
MIT researchers discover each other

RCMP Nova Scotia
@RCMPNS

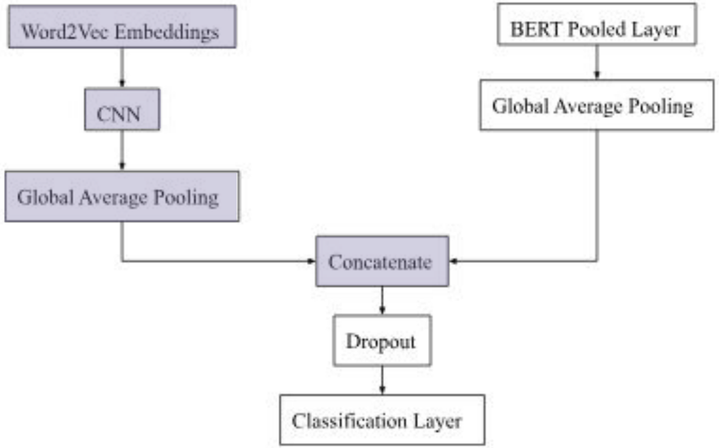
#RCMPNS is responding to a firearms complaint in the #Portapique area. (Portapique Beach Rd, Bay Shore Rd and Five Houses Rd.) The public is asked to avoid the area and stay in their homes with doors locked at this time.

Models

3 baseline models (BERT, BERTweet, ALBERT) + 6 models with concatenated Word2Vec embeddings

Model name	#
BERT	1
BERTweet	2
ALBERT	3
CNNForWord2VecBERT	4
CNNForWord2VecBERTFT	5
CNNForWord2VecBERTweet	6
CNNForWord2VecBERTweetFT	7
CNNForWord2VecALBERT	8
CNNForWord2VecALBERTFT	9

$\left.\right\}$ Baselines



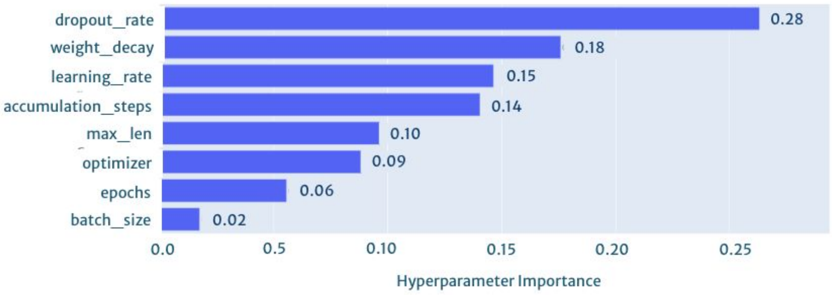
Additional components of fine-tuned models:

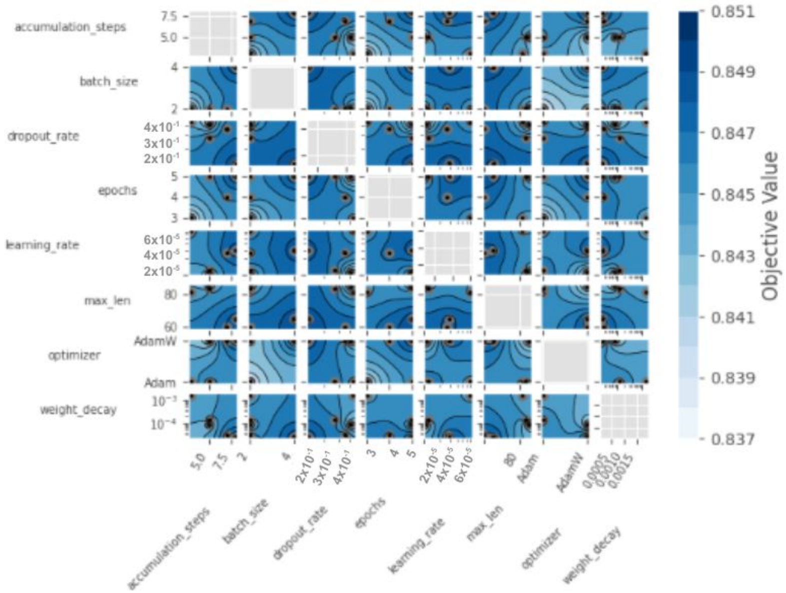
- Attention mask for non-padded tokens
- Frozen BERT Layer

Fine Tuning

An Optuna study with an expansive hyperparameter space to explore over 5 trials. With a million more dollars, we would have loved to have performed more trials.

CNNForWord2VecBERTweet





Berkeley
UNIVERSITY OF CALIFORNIA

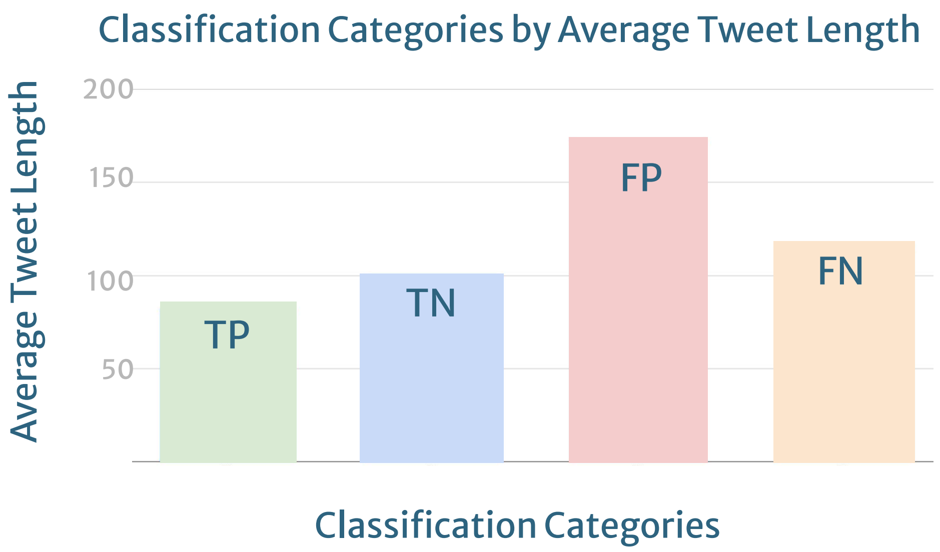
Results

Concatenation of the globally pooled embeddings showed to enhance performance, precision in particular.

Model Name	Precision	Accuracy	Recall	F1-Score
CNNForWord2VecBERTweet	0.951	0.89	0.823	0.882
CNNForWord2VecBERT	0.927	0.896	0.859	0.892
BERTweet	0.899	0.898	0.896	0.897
CNNForWord2VecALBERT	0.827	0.87	0.935	0.878

Error Analysis

Models had trouble in classifying significantly longer tweets



	Predicted 0	Predicted 1
Actual 0	police leaders join effort to reduce incarceration rate	research finds hysterectomy alone associated with increased long-term health risks
Actual 1	daylight saving time yields massive daylight surplus	need for coffee overrides scalding sensation

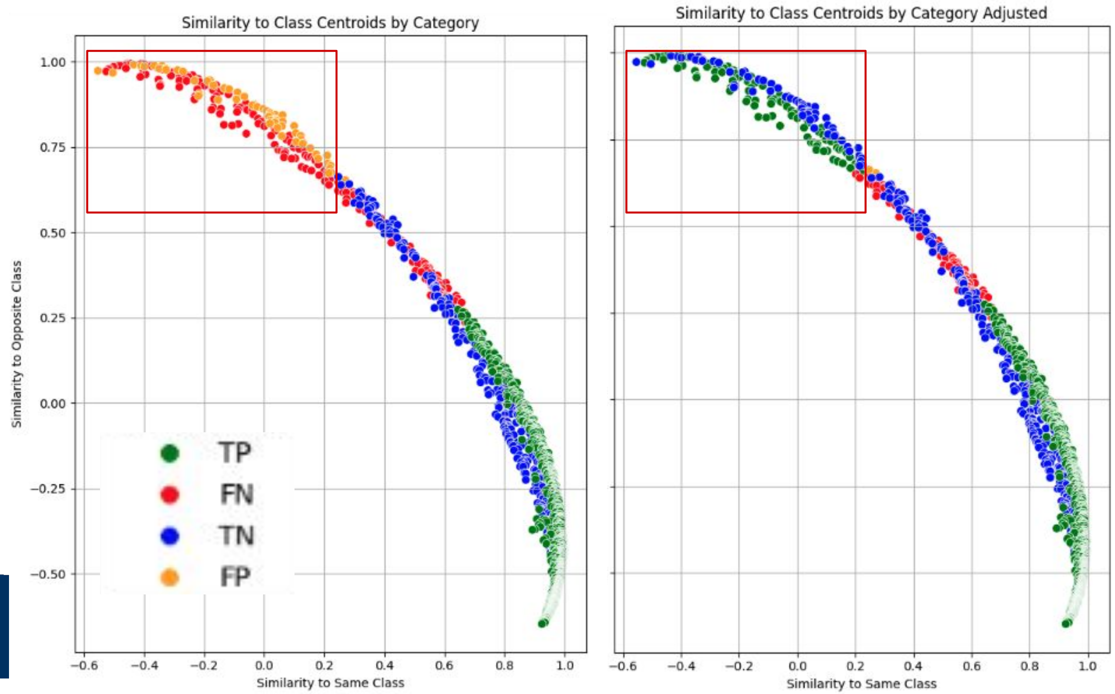
Embedding Analysis

By leveraging the cosine similarity between the embeddings for each input and the centroid of each class embedding, dissimilar and highly unusual tweets subject to change

Adjustment

if SSC < 0.25 and SSO > 0.60
Precision, recall = 1.0, 0.91

Else:
Precision, recall = 0.95, 0.82



Conclusion

Due to the nature of sarcasm, it is foreseeable that determinant factors for sarcasm will change as it is somewhat dependent on the public sentiment and personal circumstance. For this reason, it would be valuable to have a longitudinal study for understanding the classification of sarcasm over time and by/between generations.

But you already know that, don't you.

References

1. A. Abaskohi, A. Rasouli , T. Zeraati , B. Bahrak, *UTNLP at SemEval-2022 Task 6: A Comparative Analysis of Sarcasm Detection Using Generative-based and Mutation-based Data Augmentation*; School of Electrical and Computer Engineering, College of Engineering, University of Tehran
2. T. Sosea, J. Jessy Li, C. Caragea, *Sarcasm Detection in a Disaster Context*; Department of Computer Science, University of Illinois Chicago; Department of Linguistics, The University of Texas at Austin
3. D. Quoc Nguyen , T. Vu, A. Tuan Nguyen, *BERTweet: A pre-trained language model for English Tweets*; VinAI Research, Vietnam; Oracle Digital Assistant, Oracle, Australia; NVIDIA, USA
4. I. Alghanmi, L. Espinosa-Anke, S. Schockaert, *Combining BERT with Static Word Embeddings for Categorizing Social Media*; Cardiff University, UK
5. Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, *ALBERT: A Lite BERT For Self-Supervised Learning Of Language Representations*; Google Research, Toyota Technological Institute at Chicago
6. J. Devlin, M. Chang, K. Lee, K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*; Google AI Language
7. T. Mikolov, K. Chen, G. Corrado, J. Dean, *Efficient Estimation of Word Representations in Vector Space*; Google Inc., Mountain View, CA

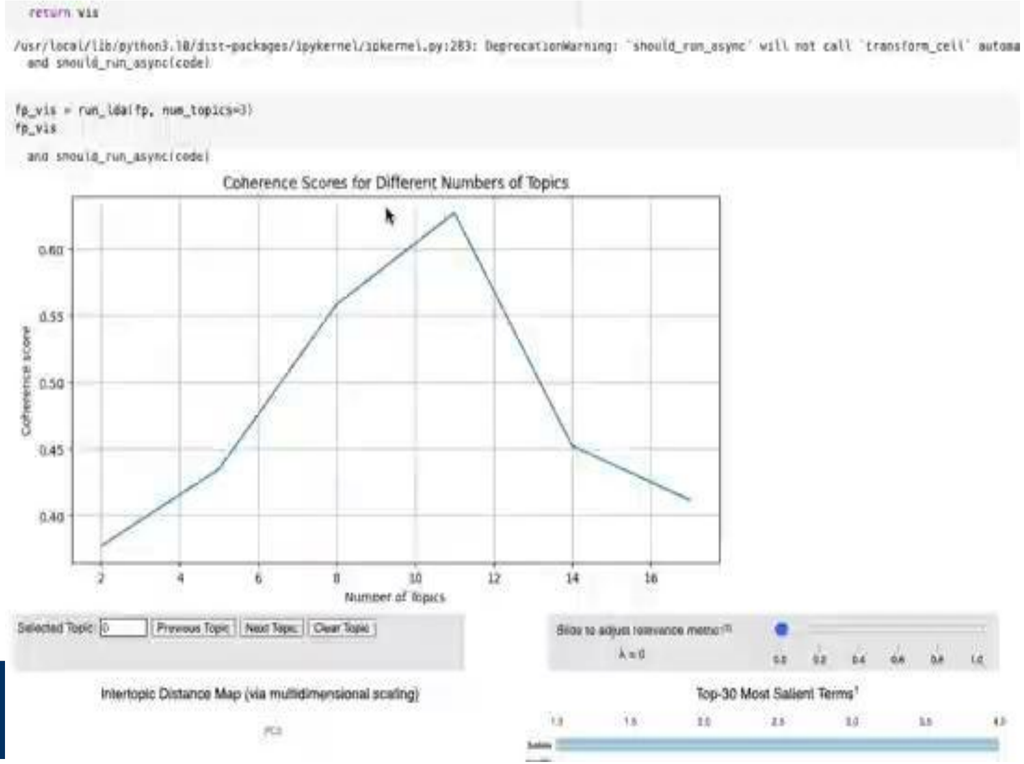
Berkeley
UNIVERSITY OF CALIFORNIA

Error Analysis

False negative and false positive topics are characterized by religion, politics and violence.

Topic modeling

Latent Dirichlet Allocation (LDA)



Appendix 1

Metrics		train_accuracy	train_precision	train_recall	train_f1	train_pr_auc	train_roc_auc
W2v+BERT	Base	0.9536	0.952875	0.9544	0.953637	0.965038	0.9536
BerTweet only	baseline	0.9694	0.969964	0.9688	0.969382	0.977182	0.9694
W2v+BERT	FT	0.9474	0.947579	0.9472	0.947389	0.96059	0.9474
W2v+BERTweet	FT	0.9958	0.994022	0.9976	0.995808	0.996411	0.9958
W2v+AlBERT	Base	0.8942	0.889988	0.8996	0.894768	0.919894	0.8942
W2v+BERTweet	Base	0.9896	0.989992	0.9892	0.989596	0.992296	0.9896
BERT only	baseline	0.987	0.986028	0.988	0.987013	0.990014	0.987
Albert only	baseline	0.6946	0.703301	0.6732	0.687922	0.769951	0.6946
W2v+AlBERT	FT	0.6776	0.679903	0.6712	0.675523	0.757751	0.6776

Metrics		val_accuracy	val_precision	val_recall	val_f1	val_pr_auc	val_roc_auc
W2v+BERT	Base	0.849167	0.908382	0.776667	0.837376	0.898358	0.849167
BerTweet only	baseline	0.87	0.912639	0.818333	0.862917	0.910903	0.87
W2v+BERT	FT	0.8625	0.860697	0.865	0.862843	0.896598	0.8625
W2v+BERTweet	FT	0.853333	0.856902	0.848333	0.852596	0.890535	0.853333
W2v+AlBERT	Base	0.84	0.795652	0.915	0.851163	0.876576	0.84
W2v+BERTweet	Base	0.835833	0.85918	0.803333	0.830319	0.880423	0.835833
BERT only	baseline	0.8325	0.813187	0.863333	0.83751	0.872427	0.8325
Albert only	baseline	0.771667	0.807547	0.713333	0.757522	0.832107	0.771667
W2v+AlBERT	FT	0.67	0.81677	0.438333	0.570499	0.767968	0.67

Berkeley
UNIVERSITY OF CALIFORNIA