Stance Detection In Online Media Data

Team 27: Meet, Priya, Shaney Waris, Sriza

10 December 2020

Abstract

Stance detection is a task of determining from the text whether the author of the text is in the favour, against, or neutral of the target that can be a person, organization, product, etc. We have used various features into an account for a feature vector like Universal Encoder by Google, BERT Encoder and various other features which will be mentioned below. We have trained various models like Random forest, XGBoost, Multi-Layer Perceptron, SVC, etc with different types of features set and compared the results of each model. What we found is the pre-trained models called Universal Encoder, BERT encoder with other features set gives the accuracy of 85%.

1 Introduction

Our problem statement is stance detection in online media data. Stance detection is also a subproblem of sentimental analysis. Stance detection is an ongoing research field and here we have summarized a few papers that show promising results. Stance detection is the task of automatically determining from the text whether the author of the text is in favor of, against, or neutral towards a proposition or target. The target may be a person, an organization, a government policy, a movement, a product, etc. Major challenges for multi-class stance detection are as follows:

- **Lexical** - The topic of the debate may be explicitly or subtly mentioned by the author and data may contain internet slang/acronyms that are not present in a normal dictionary.
- **Syntactic** - Grammatical structure may not be followed and online media-specific data like hashtags, emojis, replies, etc be present in data.
- **Semantic** - Finding right context and dealing with negation, sarcasm, humor to name a few.

1.1 Motivation

Our motivation to explore this area is to protect our society, especially the lesser educated people of our society because they don't even think twice intentionally/unintentionally before accepting the fake news or rumors and forward it to others. As we have also seen, during election time, social media gets spammed with various fake news in the name of news for political motives. People started creating their fake accounts on various platforms and spread misinformation. It Impacts the thought process of our society negatively, which might result in riots, and create ideological barriers.

2 Related Work

2.1 Twitter Stance Detection - A Subjectivity & Sentiment Polarity Inspired Two-Phase Approach

- Author used two phase model and in the first phase of the model, the tweets are classified into two - positive or negative versus neutral-stance with the Weighted MPQA Subjectivity-Polarity Classification and Wordnet Based Potential Adjective Recognition features. In the second phase, where the positive versus negative stances are classified, was inspired by sentiment classification features. The features that we use in the second phase are as follows: SentiWordNet and MPQA Based Sentiment Classification, Frame Semantics, Target Detection, Word n-grams and Character n-grams.
- Traditional SVM model is used. It surpassed the SOTA model when tested against the benchmark SemEval 2016 stance detection dataset.

2.2 Improved Stance Prediction in a User Similarity Feature Space

- Author used users' historical or topically non-relevant tweets are used to predict stance. The transformation of a bag-of-words (words or interaction elements) feature space to a user similarity space. Propose using the similarity between users as features. Interaction elements in tweets include: user mentions, retweets, replies; links to Web resources (URLs), hashtags.
- The proposed method yields significant improvement in stance prediction.

2.3 Talos Targets Disinformation with Fake News Challenge Victory

- Team SOLAT in the SWEN explored AI techniques esp ML and NLP techniques. After various experimentation, they found out that the best results appeared upon using a 50-50 weighted average on the Gradient Boosted decision Tree model and a deep Convolutional Neural Network model.
- The DL model applied a 1D CNN on the header and body text that was represented by the Google News pre-trained word vector, the output of the CNN was then fed to a MLP which gave a 4 class output(agree/disagree/discuss/unrelated). The other model in the ensemble was the Gradient-Boosted DecisionTrees model that was used to predict the relation between the header and the body.

2.4 Ranking-based Method for News Stance Detection

- Author used ranking based method instead of a direct classification method. Given a news body and its headline, the proposed technique applies a MLP to produce a value for every stance. They have used TF-IDF features.
- The evaluation was based on a weighted two-layer scoring system. In the accuracy evaluation metrics, a 25% weighted score is given to "related"/"unrelated" classification because it's easier, and 75% weighted score is given to "agree", "disagree" and "discuss" classification because it's a bit difficult and more relevant. And a relative score is determined based on this accuracy evaluation metric.
- This method achieves an overall 86.66% relative score, and if unrelated pairs of headlines and bodies are separated, 99% accuracy is achieved. The detection accuracy for "discuss", "agree" and "disagree" is improved by 3.72%, 10.86%, and 187.86%, respectively, which is pretty amazing because "agree" and "disagree" are the most difficult to recognize by previous techniques/methods.

2.5 Fake News Detection with Stance Detection

- Author developed webapp to take the input doc/claim as a text input to the system. This input is then parsed by IBM Watson n Azure APIs with some NLP APIs too to find the most relevant keywords in the input text. Based on the keywords, numerous online articles are found from the event registry on the same issue to find out what their stance is, their stance is found out by feeding these online articles to an ML model (which is again of combination of bag of words, word2vec, TF-IDF, etc) along with some DL techniques and softmax, to classify the result as agree/disagree/neutral. The project has reported 82 percent accuracy.

2.6 Unsupervised Stance Classification in Online Debates

- Author have mentioned the challenges in the online debates which are Limited opinion words and their aspects - polarity pairs, Negation words and Co-references.
- They created heuristics which help in finding the aspects and know their polarity which inturn is used in stance detection.
- They have created a matrix of the aspects-topics-polarity which they use in classification using Integer Linear Programming.

2.7 How Robust Is Your Stance Detection?

- Author combine learning from related tasks (via TL) and MDL, designed to capture all facets of StD(stance detection) tasks, in an in-depth analysis with adversarial attacks, they show that TL(transfer learning) and MDL(multiple dataset learning) for StD generally improves the performance of ML models, but also drastically reduces their robustness if compared to SDL(Single Dataset Learning) models. Adversarial attacks describe test sets aimed to discover possible weak points of ML models, they use and analyze some of these attacks for the task of StD to probe the robustness of their SDL and MDL models.
- The combination of TL and MDL even though a modern method, however results in a loss of performance when compared to SDL performance. A major reason for this was found out to be overfitting on the biases of huge training data in the MDL approach.

3 Methodology

- We have used 4 sets of different features,
  - LaBSE Encoding (Language Agnostic BERT Sentence Embedding).
  - LaBSE Encoding with self implemented features.
  - Universal Encoding
  - Universal Encoding with self implemented features.
- With 4 different Classifiers and analysed the results in each case.
  - RandomForestClassifier
  - MLPClassifier
  - SVC
  - XGBoost
- Self Implemented features are Negated words, Avg per sentence sentiment (+ve, -ve, neutral) using BingLui lexicon, Number of URLs, Lexicon sentiment of hashtags, Length of a tweet without stopwords, No of sentences, Average sentence length, Number of positive, negative, neutral sentiment words using SentimentIntensityAnalyzer, Number of total hashtags, The total number of user tags present, The total number of emoji's present, Number sentiment avg.

4 Results

4.1 Multi-Layer Perceptron

Here, we have used MLP model with four different set of features.
Note: Below, OF stands for ”Other Features” that are self implemented features mentioned in the methodology.

	precision	recall	f1-score	support
0.0	0.16	0.31	0.21	295
1.0	0.05	0.08	0.06	88
2.0	0.34	0.47	0.39	701
3.0	0.93	0.83	0.87	5849
accuracy			0.76	6933
macro avg	0.37	0.42	0.38	6933
weighted avg	0.82	0.76	0.79	6933

Figure 1: MLP with BERT only.

	precision	recall	f1-score	support
0.0	0.16	0.30	0.21	295
1.0	0.33	0.10	0.16	88
2.0	0.35	0.47	0.40	701
3.0	0.92	0.85	0.88	5849
accuracy			0.78	6933
macro avg	0.44	0.43	0.41	6933
weighted avg	0.82	0.78	0.80	6933

Figure 2: MLP with (BERT + OF)

	precision	recall	f1-score	support
0.0	0.16	0.40	0.23	295
1.0	0.02	0.03	0.03	88
2.0	0.29	0.39	0.33	701
3.0	0.92	0.80	0.86	5849
accuracy			0.73	6933
macro avg	0.35	0.41	0.36	6933
weighted avg	0.81	0.73	0.77	6933

Figure 3: MLP with Universal Encoder only.

	precision	recall	f1-score	support
0.0	0.11	0.35	0.17	295
1.0	0.06	0.02	0.03	88
2.0	0.27	0.57	0.36	701
3.0	0.93	0.72	0.81	5849
accuracy			0.68	6933
macro avg	0.34	0.41	0.34	6933
weighted avg	0.82	0.68	0.73	6933

Figure 4: MLP with (UE + OF)

4.2 Random Forest

Here, we have used Ranndom Forest model with four different set of features.

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.47	0.09	0.15	701
3.0	0.85	0.99	0.91	5849
accuracy			0.84	6933
macro avg	0.33	0.27	0.27	6933
weighted avg	0.76	0.84	0.79	6933

Figure 5: RF with BERT only.

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.50	0.08	0.14	701
3.0	0.85	0.99	0.92	5849
accuracy			0.84	6933
macro avg	0.34	0.27	0.26	6933
weighted avg	0.77	0.84	0.79	6933

Figure 6: RF with (BERT + OF)

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.76	0.03	0.06	701
3.0	0.85	1.00	0.92	5849
accuracy			0.85	6933
macro avg	0.40	0.26	0.24	6933
weighted avg	0.79	0.85	0.78	6933

Figure 7: RF with Universal Encoder only.

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.71	0.02	0.03	701
3.0	0.84	1.00	0.92	5849
accuracy			0.84	6933
macro avg	0.39	0.25	0.24	6933
weighted avg	0.78	0.84	0.78	6933

Figure 8: RF with (UE + OF)

4.3 Support Vector Machine

Here, we have used SVM model with four different set of features.

	precision	recall	f1-score	support
0.0	0.16	0.15	0.15	295
1.0	0.00	0.00	0.00	88
2.0	0.48	0.38	0.42	701
3.0	0.89	0.93	0.91	5849
accuracy			0.83	6933
macro avg	0.38	0.36	0.37	6933
weighted avg	0.81	0.83	0.82	6933

Figure 9: SVC with BERT only.

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.02	0.00	0.00	701
3.0	0.84	0.99	0.91	5849
accuracy			0.84	6933
macro avg	0.22	0.25	0.23	6933
weighted avg	0.71	0.84	0.77	6933

Figure 10: SVC with (BERT + OF)

	precision	recall	f1-score	support
0.0	0.23	0.25	0.24	295
1.0	0.00	0.00	0.00	88
2.0	0.49	0.38	0.43	701
3.0	0.90	0.94	0.92	5849
accuracy			0.84	6933
macro avg	0.41	0.39	0.40	6933
weighted avg	0.82	0.84	0.83	6933

Figure 11: SVC with Universal Encoder only.

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.02	0.00	0.00	701
3.0	0.84	0.99	0.91	5849
accuracy			0.84	6933
macro avg	0.22	0.25	0.23	6933
weighted avg	0.71	0.84	0.77	6933

Figure 12: SVC with (UE + OF)

4.4 XGBoost

Here, we have used XGBoost model with four different set of features.

	precision	recall	f1-score	support
0.0	0.20	0.08	0.11	295
1.0	0.00	0.00	0.00	88
2.0	0.42	0.23	0.30	701
3.0	0.87	0.95	0.91	5849
accuracy			0.83	6933
macro avg	0.37	0.32	0.33	6933
weighted avg	0.78	0.83	0.80	6933

Figure 13: XGB with BERT only.

	precision	recall	f1-score	support
0.0	0.16	0.30	0.21	295
1.0	0.33	0.10	0.16	88
2.0	0.35	0.47	0.40	701
3.0	0.92	0.85	0.88	5849
accuracy			0.78	6933
macro avg	0.44	0.43	0.41	6933
weighted avg	0.82	0.78	0.80	6933

Figure 14: XGB with (BERT + OF)

	precision	recall	f1-score	support
0.0	0.26	0.06	0.10	295
1.0	0.00	0.00	0.00	88
2.0	0.57	0.20	0.29	701
3.0	0.87	0.98	0.92	5849
accuracy			0.85	6933
macro avg	0.43	0.31	0.33	6933
weighted avg	0.80	0.85	0.81	6933

Figure 15: XGB with Universal Encoder only.

	precision	recall	f1-score	support
0.0	0.22	0.06	0.10	295
1.0	0.00	0.00	0.00	88
2.0	0.63	0.22	0.32	701
3.0	0.87	0.98	0.92	5849
accuracy			0.85	6933
macro avg	0.43	0.32	0.34	6933
weighted avg	0.81	0.85	0.81	6933

Figure 16: XGB with (UE + OF)

5 Individual Member Contribution

The initial research and information gathering was conducted by all the members in the group. BERT and Universal embedding was done by Meet and Priya and the preprocessing and other lexicon based features were implemented by Sriza and Shaney Waris. The models as mentioned above were then applied to all the embeddings obtained and the results were analysed. The webapp with the help of flask framework is developed by Shaney Waris to detect the stance with an ease.

References

[1] K. Dey, R. Shrivastava, and S. Kaushik, “Twitter Stance Detection — A Subjectivity and Sentiment Polarity Inspired Two-Phase Approach,” in 2017 IEEE International Conference on Data Mining Workshops (ICDMW), Nov. 2017, pp. 365–372, doi: 10.1109/ICDMW.2017.53.

[2] W. Largent, “Talos Targets Disinformation with Fake News Challenge Victory.” http://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html (accessed Dec. 10, 2020).

[3] Q. Zhang, E. Yilmaz, and S. Liang, “Ranking-based Method for News Stance Detection,” in Companion Proceedings of the The Web Conference 2018, Republic and Canton of Geneva, CHE, Apr. 2018, pp. 41–42, doi: 10.1145/3184558.3186919.

[4] “Fake News Detection Details,” Kastan Day, Dec. 25, 2017. https://kastanday.com/fake-news-detection-details/ (accessed Dec. 10, 2020).

[5] S. Ghosh, K. Anand, S. Rajanala, A. B. Reddy, and M. Singh, “Unsupervised stance classification in online debates,” in Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, New York, NY, USA, Jan. 2018, pp. 30–36, doi: 10.1145/3152494.3152497.

[6] “Papers with Code - Stance Detection Benchmark: How Robust Is Your Stance Detection?” https://paperswithcode.com/paper/stance-detection-benchmark-how-robust-is-your (accessed Dec. 10, 2020).

[7] “AI Hub - universal-sentence-encoder-large.” https://aihub.cloud.google.com/u/0/p/products%2F42c1bfd4-8104-450c-a348-29b047d3691c (accessed Dec. 10, 2020).

[8] “Language-Agnostic BERT Sentence Embedding,” Google AI Blog. http://ai.googleblog.com/2020/08/language-agnostic-bert-sentence.html (accessed Dec. 10, 2020).