Technological University Dublin

ARROW@TU Dublin

Conference papers

School of Marketing

2020-12

Combining BERT with Contextual Linguistic Features for Identification of Propaganda Spans in News Articles

Arjumand Younus

Muhammad Atif Qureshi

Abstract—**Recent endeavours at detection of propaganda in news articles treat this as a fine-grained problem of detecting it within fragments; and hence, transformer based embeddings perform decently in such detection. We build our propaganda detection framework on top of a transformer model simultaneously enriching it with contextual linguistic information of surrounding part-of-speech tags and LIWC categories the word itself belongs to. The evaluation outcomes being encouraging indicate a huge potential for this line of reasoning in natural language processing of news text.**

I. Introduction

Social media has become a major source of news content with this also giving birth to the "fake news" phenomenon whereby a huge amount of misinformation spreads mostly with a malicious intent. A significant but neglected aspect of such malicious content is an associated propaganda which is almost always present in fake news content. As defined by the Institute for Propaganda Analysis [1], the chief purpose of propaganda is to influence the opinion of target individuals, and a core ingredient of such influence is language manipulation. We therefore hypothesize in this paper to take advantage of recent language modelling efforts that generate language representations which take word relationships into account. Moreover, these rich language representations are combined with contextual linguistics such as neighboring part-of-speech tags and linguistic classes of words within a logistic regression model.

The fundamental premise behind our proposed method for propaganda detection in news articles is the observation that surrounding linguistic information of words together with their associated semantics represent a significant amount of information that can be exploited in a supervised learning framework. Through experimental evaluations in which we utilise semantic features from BERT and linguistic features in a logistic regression model correctness of the premise is demonstrated. Our results show an improvement in incremental steps when semantic features extracted from BERT are

first complented with POS tags and then linguistic inquiry and word count categories (Linguistic Inquiry and Word Cound).

II. Task Details and Related Work

This work builds on SemEval-2020 Task 11 whereby systems had to be devised for detection of propaganda spans in articles over a dataset created specifically for this purpose [2]. This task was recently announced, and involved identification of text spans containing propaganda. The task's complicated nature stems from the fact that a propaganda scan can largely vary in length from a single word/token to an entire sentence. Figure 2 shows an example fragment of a news article with the underlined segments being propaganda spans; as is clear the first two comprise a single word while the next two comprise a complete sentence.

Most of existing techniques utilize some form of transformer architecture [3] for language representations and combine them with engineered features within another machine learning architecture. Some approaches also rely on non-contextual word embeddings i.e. global ones along with some linguistic features [4].

III. Methodology

Inspired from the success of transformer-based approaches, we follow a similar methodology. However, we perform word-level classification through fine tuning BERT followed by feeding part-of-speech features and LIWC (Linguistic Inquiry and Word Count) features into a logistic regression model. In the following sections we present details of each of our features followed by a description of how we combine all of them in the final model.

A. Features from BERT

BERT essentially utilises a multi-layer transformer architecture for the tasks of next sentence prediction and masked word prediction on extremely large datasets. Its unique power that makes it perform extremely well in many text classification

Finally, the shocking surge in immigrant displacement of American workers and in the immigrant workforce, which began in January and completely undid the improvement we had begun hopefully to call the "Trump Effect" has stalled.
May job numbers released Friday show native-born Americans big winners and immigrants (legal and illegal, the Labor Department doesn't distinguish) big losers.
But displacement and immigrant workforce growth are still high in absolute terms and could resume anytime.
Only legislation, above all an immigration moratorium, can secure the fruits of the current economic expansion for Americans.

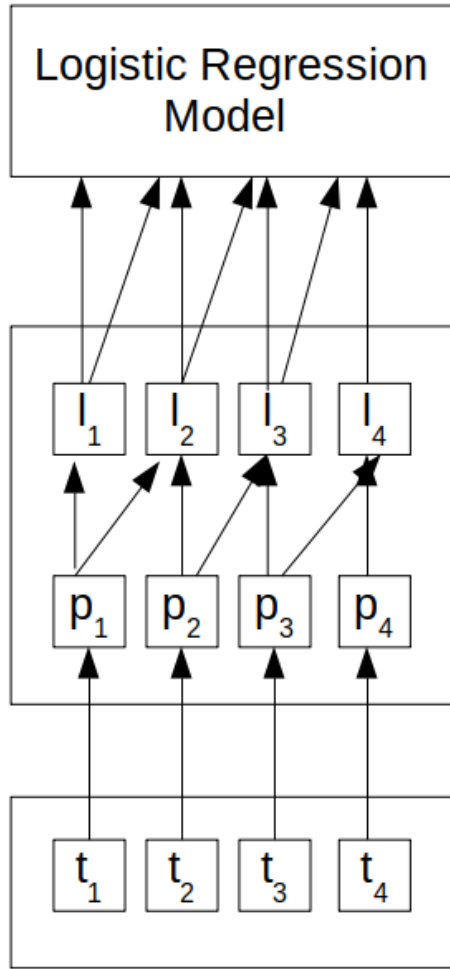Figure 1: Examples of Propaganda Spans in a News Snippet

Figure 2: Architecture of our Propaganda Spans' Detection Framework

tasks comes from the self-attention mechanism through which word representations are enriched with contextual information.

Based on the fine-tuning scheme by Yoosuf and Yang [5] we apply BERT on the token-level where the output from the final classification layer is a propaganda or non-propaganda label for each token. We do not use the labels from this classification layer as output and instead use the vector obtained by BERT's hidden layer which can represent the semantic features.

B. Part-of-Speech Tags as Features

This set of features is basically a linguistic set of features mapping each word to a corresponding part-of-speech tag. We fundamentally differ from the way part-of-speech tags are normally incorporated in that for each word token we also use the tag for the word before and word after a word being considered thereby incorporating linguistic context around that specific word.

C. Linguistic Inquiry and Word Count Features

Linguistic Inquiry and Word Count [6] is a pre-defined lexicon for categorisaton of words into psycholinguistic categories namely social, affective, cognitive, perceptual, biological, relativity, personal concerns and spoken. Similar to part-of-speech tags, we utilise the LIWC category for the word before and word after a word being considered.

Figure 2 shows the model architecture when all the elements are combined; here there are four tokens and t_1 to t_4 represent BERT feature vectors, p_1 to p_4 represent part-of-speech tags and l_1 to l_4 represent LIWC categories. The features from BERT's hidden layers are combined with part-of-speech tags of words and LIWC categories of words. Note that LIWC returns the psycholinguistic category on a per-sentence basis which we distribute for all words in that sentence. As can be seen from Figure 2, for the part-of-speech tags and LIWC categories when fed into the logistic regression model are fed with this information for surrounding words too.

IV. EXPERIMENTS AND RESULTS

We now present details of the dataset used, experimental settings and results of the various experimental settings. Our main motivation is to demonstrate the effectiveness of combining BERT semantic feature vector representation with different contextual linguistic information both in isolation and taking into account surroundin context of a word.

A. Dataset

The dataset comes from the PTC-SemEval20 corpus citeda2020semeval whereby a sample of news articles were retrieved from the period mid-2017 to early 2019. The organizers selected 13 propaganda and 36 non-propaganda news media outlets, as labeled by Media Bias/Fact Check, and retrieved articles from these sources. The dataset is divided into three sets namely training, development and testing. The evaluation measures used are precision, f-measure and recall

Experimental Setting	Precision	Recall	F-Measure
M_B	42.88.	35.67	38.88
M_{BPC}	44.24.	36.34	39.90
M_{BPLC}	51.26.	40.54	45.27
M_{BPL}	43.04.	34.84	38.51

Table I: Results under Various Experimental Settings

but with slight modifications so as to take partial matches of propaganda spans into account.

B. Experimental Settings

Within our framework we perform four different sets of experiments with the first three being the subsequent steps of Figure 2 while the last showing the effect of taking part-of-speech tags and LIWC categories of only the word in consideration rather than surrounding context words. Each experimental setting is defined below:

- M_B: Model with BERT feature vectors only
- M_{BPC}: Model with BERT feature vectors of word being considered and part-of-speech tags of word plus surrouding words
- M_{BPLC}: Model with BERT feature vectors of word being considered, part-of-speech tags of word plus surrouding words and LIWC categories of word plus surrounding words
- M_{BPL}: Model with BERT feature vectors of word being considered, part-of-speech tags of word being considered and LIWC categories of word being considered
  Each setting above is tested with same parameters for the sake of fairness and to ensure correct observation of each algorithm is made specially in relation to incorporation of contextual linguistic information within BERT.

C. Experimental Results

Table 1 shows the results for each setting in terms of precision, recall and f-measure at the character level. Note that the results are shown for the test set as we use the development set for parameter hypertuning.

As can be seen incorporating part-of-speech and LIWC categories of word being considered together with word before and word after that word significantly improves the performance of propaganda spans. In fact the last row of Table 1 shows almost the same performance as when BERT features alone are used. This proves that contextual linguistic aspects together with contextual semantic aspects play a significant role in classification of text data; and to the best of our knowledge this can serve as a significant direction within natural language processing of news pieces.

V. CONCLUSION AND FUTURE WORK

The chief contribution of this work is demonstration of the power transformer models in collaboration with traditional linguistic features, and further to develop the case for incorporation of context in surrounding linguistic information. To the best of our knowledge, this line of reasoning has not been pursued within modern natural language processing community. As future work we aim to incorporate advanced linguistic features and their surrouding context with few examples being named entities, dependency parse trees, readability scores etc.

REFERENCES

[1] I. for Propaganda Analysis, "How to detect propaganda," *Propaganda Analysis*, vol. 1, no. 2, pp. 5–8, 1937.
[2] G. Da San Martino, A. Barrón-Cedeno, H. Wachsmuth, R. Petrov, and P. Nakov, "Semeval-2020 task 11: Detection of propaganda techniques in news articles," in *Proceedings of the 14th International Workshop on Semantic Evaluation, SemEval*, 2020.
[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT (1)*, 2019.
[4] T. Alhindi, J. Pfeiffer, and S. Muresan, "Fine-tuned neural models for propaganda detection at the sentence and fragment levels," *arXiv preprint arXiv:1910.09702*, 2019.
[5] S. Yoosuf and Y. Yang, "Fine-grained propaganda detection with fine-tuned bert," in *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, 2019, pp. 87–91.
[6] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.