# LTIatCMU at SemEval-2020 Task 11: Incorporating Multi-Level Features for Multi-Granular Propaganda Span Identification

**Sopan Khosla*   Rishabh Joshi*   Ritam Dutt*   Alan W Black   Yulia Tsvetkov**
Language Technology Institute
Carnegie Mellon University
{sopank,rjoshi2,rdutt,awb,ytsvetko}@cs.cmu.edu

## Abstract

In this paper we describe our submission for the task of Propaganda Span Identification in news articles. We introduce a BERT-BiLSTM based span-level propaganda classification model that identifies which token spans within the sentence are indicative of propaganda. The "multi-granular" model incorporates linguistic knowledge at various levels of text granularity, including word, sentence and document level syntactic, semantic and pragmatic affect features, which significantly improve model performance, compared to its language-agnostic variant. To facilitate better representation learning, we also collect a corpus of 10k news articles, and use it for fine-tuning the model. The final model is a majority-voting ensemble which learns different propaganda class boundaries by leveraging different subsets of incorporated knowledge.[1]

## 1 Introduction

Propaganda (Bernays, 1928) is "the deliberate and systematic attempt to shape perceptions, manipulate cognition, and direct behavior to achieve a response that furthers the desired intent of the propagandist" (Jowett and O'donnell, 2018). The rise of digital and social media has enabled propaganda campaigns to reach vast audiences (Glowacki et al., 2018; Tardáguila et al., 2018), influencing or undermining the democratic processes, amplifying the effects of misinformation, and creating echo-chambers.

As a tactic of manipulation, propaganda is considered most effective when it is veiled: the readers are not able to identify it, but their opinions are shaped according to the propagandist's hidden agenda (Barrón-Cedeño et al., 2019); it is thus very difficult to identify propaganda automatically. However, automatic identification and analysis of propaganda in news articles and on social media are essential to understand propaganda at scale and develop approaches to countering it (King et al., 2017; Starbird, 2018; Field et al., 2018).

Prior research on propaganda detection focused primarily on identifying propaganda at a document level, due to the dearth of finer-grained labelled data (Rashkin et al., 2017; Barrón-Cedeño et al., 2019). This has resulted in classifying *all* news articles from a particular source as propaganda (or hoax, disinformation, etc.), which is often not the case, and which can obfuscate a finer-grained analysis of propagandistic strategies (Da San Martino et al., 2019; Horne et al., 2018).

Recently, Da San Martino et al. (2019) carried out a seminal task of fine-grained propaganda detection and curated a dataset consisting of about 550 news articles. The dataset contains word-span level annotations provided by high-quality professionals along with additional information about the propaganda technique employed in the span. The Span Identification (SI) sub-task of SemEval 2020, Task 11, (Da San Martino et al., 2020) employs this aforementioned corpus and requires participants to detect propaganda spans in news articles.

In this paper, we describe our solution to the SI task. We propose a BERT-BiLSTM based multi-granularity model, inspired from (Da San Martino et al., 2019). The model's parameters are jointly

---

[1]Our final ensemble attains $4^{th}$ position on the test leaderboard. Our final model and code is released at https://github.com/sopu/PropagandaSemEval2020.
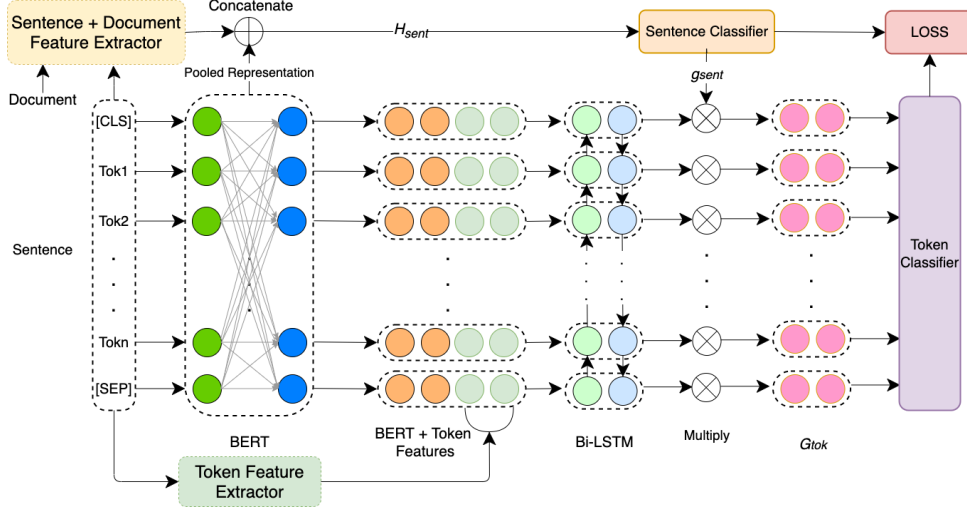
Figure 1: Overview of our Multi-granular BERT-BiLSTM model. The sentence is passed into BERT and the token representations are concatenated with token features before being passed to a BiLSTM layer. The pooled representations are concatenated with sentence and document level features before getting the logits of the sentence containing propaganda. This gating value is then multiplied with the BiLSTM outputs of each token before being used to predict token labels. The model is jointly trained using loss from both token and sentence classification (Details in Section 2).

optimized on the tasks of sentence-level and word-level propaganda detection. We fine-tune BERT-BiLSTM end-to-end on the joint-loss to obtain competitive scores on the SI task. Furthermore, we explore the benefits of incorporating word, sentence and document-level syntactic and affective features extracted from dictionaries like LIWC (Tausczik and Pennebaker, 2010) and Empath (Fast et al., 2016). We find that incorporating these features leads to significant improvements in performance (discussed in Section 3). We also leverage articles from multiple well-known propaganda news sources to fine-tune BERT in an unsupervised manner. Furthermore, in order to tackle the problem of unbalanced data, we use a weighted cross-entropy loss. We present our model and additional features in Section 2, and show our results and ablation analysis in Section 3.

## 2 Model

In this section, we describe the modeling aspects we have adopted for the task. The overview of our model is given in Figure 1.

### 2.1 Multi-granular BERT-BiLSTM

We implement a model that leverages sentence-level propaganda detection to guide the task of detecting propaganda word-spans (Da San Martino et al., 2019). Both stages have their separate classification layers $L_{sent}, L_{tok}$ and the feature representation learnt at the sentence-level $H_{sent}$ influences $H_{tok}$.

To classify each sentence, we mean-pool the token representations output from BERT, concatenate it with the sentence ($\boldsymbol{F}_{sent}$) and document ($\boldsymbol{F}_{doc}$) level features, and pass it through a fully-connected layer to obtain the logits $\boldsymbol{p}_{sent}$. $[;]$ denotes the concatenation operator.

$$\boldsymbol{H}_{sent} = [\text{Pool}(\text{BERT}([\text{CLS}], w_1, w_2, \ldots, w_n, [\text{SEP}])); \boldsymbol{F}_{sent}; \boldsymbol{F}_{doc}], \qquad \boldsymbol{p}_{sent} = \text{FC}(\boldsymbol{H}_{sent}). \quad (1)$$

For word-span level predictions, we input the token representations output from BERT and concatenate them with word-level features for each token ($\boldsymbol{F}_{word}$) before passing them into a BiLSTM:

$$\boldsymbol{H}_{tok}^* = [\text{BERT}([\text{CLS}], w_1, w_2, \ldots, w_n, [\text{SEP}]) \oplus \boldsymbol{F}_{word}], \qquad \boldsymbol{H}_{tok} = \text{BiLSTM}(\boldsymbol{H}_{tok}^*). \quad (2)$$

where $\oplus$ represents the token-wise concatenation operator and $\boldsymbol{H}_{tok}$ is the list of contextual representations for each token output by the BiLSTM. $\boldsymbol{F}_{word}$, $\boldsymbol{F}_{sent}$, and $\boldsymbol{F}_{doc}$ are described in Section 2.2.

The sentence-level representation is passed through a masking gate that projects $\boldsymbol{p}_{sent}$ to a single dimension and applies sigmoid non-linearity. To control the information flow, each element of the token-level representation $\boldsymbol{H}_{tok}^i$ is multiplied with the gate to obtain gated token representations $\boldsymbol{G}_{tok}$. This biases the token-level model to ignore samples strongly classified as negative in the sentence-level task, allowing it to selectively learn additional information to improve performance on token-level prediction:

$$\boldsymbol{g}_{sent} = \sigma(W_g\,\boldsymbol{p}_{sent} + b_g), \qquad \boldsymbol{G}_{tok}^i = \boldsymbol{g}_{sent} * \boldsymbol{H}_{tok}^i, \qquad \boldsymbol{p}_{tok}^i = \text{FC}(\boldsymbol{G}_{tok}^i). \tag{3}$$

We use a cross-entropy loss with sigmoid activation for sentence-level classification, whereas the token-level classification is trained using a cross-entropy loss with softmax activation. The losses $L_{sent}, L_{tok}$ are jointly optimized using hyper-parameter $\alpha$ to control the strength of both losses:

$$\boldsymbol{L} = \alpha \boldsymbol{L}_{sent} + (1-\alpha)\boldsymbol{L}_{tok}. \tag{4}$$

## 2.2 Additional Features

We incorporate additional lexical, syntactic, linguistic, and topical features at word, sentence and document levels to better inform the model.

**Syntactic Features:** Inspired by prior work that leveraged syntactic features effectively (Vashishth et al., 2018; Kumar et al., 2020), and motivated by the observation that many propaganda spans are persuasive or idiomatic phrases, we extract phrasal features from constituency parse trees to explicitly incorporate structural syntactic information. We encode the path from a word to the root in the parse tree as a $d_c$-dimensional embedding.[2] Stanford CoreNLP (Manning et al., 2014) was used to extract the constituency parses as well as part-of-speech tags that were also used as features.

**Affective and Semantic Features**: In addition to structural cues, prior work has shown that propaganda is marked with affective and emotional words and phrases (Gupta et al., 2019; Alhindi et al., 2019). Motivated by that, we append to word embeddings of the $i^{th}$ token (after BERT) features extracted using affective lexicons, including the LIWC (Tausczik and Pennebaker, 2010) dictionary, NRC lexicons of Word Emotion, VAD and Affect Intensity (Mohammad and Turney, 2013; Mohammad, 2018a; Mohammad, 2018b). We also assign a score to each token that corresponds to the frequency of the word in the propaganda spans as opposed to the non-propaganda ones. For example, words like 'invader', will have a high score, as it is salient to propaganda spans. Furthermore, we incorporate semantic class features. These include named entities (such as Person, Place, Temporal Expressions) as identified by the CoreNLP NER tagger (Manning et al., 2014) and finer-grained topical categories from Empath (Fast et al., 2016) (such as Government, War, Violence).

These word-level features ($\boldsymbol{F}_{word}$) are concatenated to the token-level BERT representations.

**Sentence-level Features** ($\boldsymbol{F}_{sent}$): We also encode a sentence using the BERT-large-case model fine-tuned on news articles from different sources to get a 1024 dimensional vector.

**Document-level Features** ($\boldsymbol{F}_{doc}$): Similarly, we obtain a 1024-dimensional embedding of the document, by averaging across BERT embeddings obtained for each sentence in the document.

The motivation for incorporating sentence and document features is to inform the model about the overall topical content of the article. We hypothesize that these features are especially helpful in detecting a specific type of propaganda called "repetition" where certain events are mentioned several times in the document. We append these to the pooled BERT representation to improve sentence-level classification.

## 2.3 Unsupervised Fine-tuning

Neural language models like BERT (Devlin et al., 2019) unlock their real power from the large amounts of data they are pre-trained on. The original BERT models are trained on BookCorpus and Wikipedia datasets that are essentially a text-based knowledge source. Therefore, the vanilla BERT learns language properties from an objective dataset and thus missing the nuances of persuasive and metaphorical language extensively used in propaganda (Da San Martino et al., 2019). To alleviate this issue we pre-train the original BERT models on a large collection of news articles collected from different sources. To ensure equal representation, we scraped approximately 10k articles from propaganda websites mentioned

---

[2]We set the value of $d_c$ to 30 in our experiments.

in Da San Martino et al. (2019), e.g., Lew Rockwell, SHTFplan.com, and 10k articles from trusted non-propagandist sources like CNN and New York Times.[3] We scraped articles from 2018 to mid-2019 to ensure that the articles follow the same topical distribution as that of the training articles. We provide the unlabeled data, split to sentences, and train BERT on masked LM and next sentence prediction losses (Devlin et al., 2019). We leverage the pipeline provided by the Huggingface transformers library.[4]

## 2.4 Class-Imbalance

Since our corpus suffers from class imbalance in both sentence classification and word-level span identification task, we associate higher weight to the loss incurred by samples of the minority class. Following Khosla (2018), we calculate these weights as inverse of the normalized frequency of samples of each class and plug them into the cross-entropy loss. Namely $L = -\frac{1}{N} \sum_{n=1}^{N} w_n * \text{loss}_n$, where $w_n$ is the weight associated with the loss for each sample.

## 3 Experimental Setup

We used the BERT model from hugging-face library which was then fine-tuned on the corpus. Since the test-set was not made available during the competition, we hold out a small part of the training data as dev-set which was used to tune the models. The submitted model was chosen based on its Span-level Normalized F1-Score on the validation set provided by the competition organizers.[5] The hyperparameter choices are provided in Table A1 in the Appendix. We run all our experiments on Nvidia Geforce GTX 1080 Ti GPUs with 12GB memory.

## 4 Results

In this section, we present our method's results on the official validation set. Table 1 shows the performance of different multi-granularity BERT and BERT-BiLSTM. We report the mean of 5 independent runs (with different seeds). We find that BERT-large variants perform significantly better than BERT-base with BERT-large-cased performing the best. This suggests that case-based signals might be important for identifying spans as the writer might use capitalization to put more emphasis on propaganda information.

We also analyze the importance of word, sentence and document-level features (as detailed in Section 2.2) to our model. For brevity, we only present the results on multi-granularity BERT-large-uncased BiSLTM (**MGU-LSTM**) and BERT-large-cased BiLSTM (**MGC-LSTM**) models due to their superior performance. Initial few rows in Table 3 depict the results for concatenating word-level features to word-level BERT representations. We find that concatenating Affective, LIWC, (**A**) and Syntax (**X**) features increases the performance of **MGU-LSTM** and **MGC-LSTM** models by 0.70 and 0.24 F1 respectively. Further adding NER and Empath (**N**) based features only seems to help **MGC-LSTM** attaining 0.18 additional points. This suggests that the embeddings learnt by the two models might differ in the kinds of features they represent. **MGU-LSTM** might already encode most of the information present in C (dimensions highly correlated with C). Nevertheless, testing this hypothesis requires an in-depth analysis of individual features and the word-level representations which is out of scope for this paper.

Our experiments with sentence (**S**) and document-level (**D**) features (Table 3) suggest no clear pattern as adding sentence-level features to **MGC-LSTM-AXN** increases F1 by 0.1 but degrades **MGU-LSTM-AX**'s performance from 44.11 to 44.07 F1. However, adding document-level features seem to benefit **MGU-LSTM-AX-S** but show no significant improvement ($p = 0.05$) in **MGC-LSTM-AXN**.

Finally, we present the contributions of the weighted cross-entropy loss (**W**) and the unsupervised fine-tuning (**F**) performed on articles from propaganda news sources in Table 2. Adding inverse-frequency based weighting gives a boost of 0.66 F1 points to **MGC-LSTM-AXN-S**. This is expected as the corpus is highly imbalanced towards non-propaganda words. Fine-tuning on news articles also provides a significant performance jump indicating the advantages of learning domain-informed contextualized

---

[3]https://www.lewrockwell.com/,https://vdare.com/,http://www.shtfplan.com/,https://www.cnn.com/,https://www.nytimes.com/
[4]https://github.com/huggingface/transformers
[5]The official competition metric. We use the script provided by the organizers to calculate Span-level F1.

representations. Our best performing (single) model is a multi-granularity BiLSTM architecture with BERT-large-cased embeddings fine-tuned on news articles, incorporating Affect, LIWC, Syntax, NER, Empath and sentence features, and optimized using a weighted cross-entropy loss.

| Model | Span-level F1 |
| --- | --- |
| BERT-base-uncased (BERT-BU) | 41.38 |
| BERT-base-cased (BERT-BC) | 42.12 |
| BERT-large-uncased (BERT-LU) | 43.12 |
| BERT-large-cased (BERT-LC) | **43.56** |
| BERT-LU BiLSTM (**MGU-LSTM**) | 43.41 |
| BERT-LC BiLSTM (**MGC-LSTM**) | **43.86** |

Table 1: Multi-granularity model variants. The BERT based models are the baselines.

| Model | F1 |
| --- | --- |
| **MGC-LSTM-AXN-S** | 44.38 |
| + Weight (**W**) | 45.04 |
| + Fine-tune (**F**) | 44.56 |
| **+ W + F** | **45.45** |

Table 2: Results for Weighted cross-entropy loss (W) and unsupervised fine-tuning (F) on news articles.

| Features | F1 |
| --- | --- |
| **Word** | |
| **MGU-LSTM** | 43.41 |
| + (Aff + LIWC) (**A**) | 43.69 |
| + A + Syn (**X**) | **44.11** |
| + AX + (NER + Emp) (**N**) | 43.81 |
| **MGC-LSTM** | 43.86 |
| + (Aff + LIWC) (**A**) | 43.87 |
| + A + Syn (**X**) | 44.10 |
| + AX + (NER + Emp) (**N**) | **44.28** |
| **Sentence** | |
| **MGU-LSTM-AX** + Sent (**S**) | 44.07 |
| **MGC-LSTM-AXN** + Sent | **44.38** |
| **Document** | |
| **MGU-LSTM-AX-S** + Doc (**D**) | 44.27 |
| **MGC-LSTM-AXN-S** + Doc | **44.35** |

Table 3: Results for different word, sentence and document-level features. We only show the scores on MGU-LSTM and MGU-LSTM here for brevity (**Aff** = Affective features, **Syn** = Syntactic features, **Emp** = Empath features).

**Ensemble** We created a majority-voting ensemble of a subset of the model variants discussed above. We observe that we obtain the best results when the ensembled models are the most dissimilar. This pattern has also been shown to be beneficial in stacking where the same machine-learning problem is tackled with different types of learners. We find that an ensemble of 7 models namely **BERT-LC-AXN-S-D-W** (seed 3), **MGC-LSTM-AX-S-W** (seed 3), **MGU-LSTM-AX-S-W** (seed 1), **MGC-LSTM-AXN-S-W** (seed 3), **MGC-LSTM-AXN-S-W-F** (seed 1), and **MGC-LSTM-AXN-S-W-F** (seed 3) performs substantially better than individual models improving the performance on validation-set to **48.86**. We illustrate an example of detecting propaganda spans by our model in Appendix. The spans are highlighted in bold. We observe that our model is not only able to detect short spans like loaded language or named calling but also large spans corresponding to repetition and slogans that rely on contextual information.

**Final Submission** We submit the results from the ensemble followed by post-processing like merging disjoint spans separated by by 1-2 words, removing stop-words or stray characters like quotes from the beginning and end of spans, as well as label words frequently labelled with the 'loaded language' label. Our final submission scored **49.06** F1-Score on the official dev-set and **47.66** on the official test-set.

## 5 Conclusion

This paper describes our $4^{th}$ place submission to the Span Identification (SI) subtask of SemEval 2020 Task 11. Our approach is based on a multi-granularity BiLSTM with BERT embeddings. We explore the contributions of several word, sentence and document-level features concatenated with token and pooled embeddings output from BERT. Our work also highlights the importance of tackling class-imbalance in the corpus and learning domain-informed representations through unsupervised fine-tuning of BERT on latest news articles. We submit a majority-voting ensemble of multiple models with potentially dissimilar decision boundaries to make robust predictions on the official validation and test-set.

# References

Tariq Alhindi, Jonas Pfeiffer, and Smaranda Muresan. 2019. Fine-tuned neural models for propaganda detection at the sentence and fragment levels. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 98–102, Hong Kong, China, November. Association for Computational Linguistics.

Alberto Barrón-Cedeño, Israa Jaradat, Giovanni Da San Martino, and Preslav Nakov. 2019. Proppy: Organizing the news based on their propagandistic content. *Inf. Process. Manag.*, 56(5):1849–1864.

Edward L Bernays. 1928. *Propaganda*. New York: H. Liveright.

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5636–5646, Hong Kong, China, November. Association for Computational Linguistics.

Giovanni Da San Martino, Alberto Barrón-Cedeño, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020. SemEval-2020 task 11: Detection of propaganda techniques in news articles. In *Proceedings of the 14th International Workshop on Semantic Evaluation*, SemEval 2020, Barcelona, Spain, December.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4647–4657.

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in russian news: a computational analysis of intricate political strategies. *arXiv preprint arXiv:1808.09386*.

Monika Glowacki, Vidya Narayanan, Sam Maynard, Gustavo Hirsch, Bence Kollanyi, Lisa-Maria Neudert, Phil Howard, Thomas Lederer, and Vlad Barash. 2018. News and political information consumption in mexico: Mapping the 2018 mexican presidential election on twitter and facebook. *The Computational Propaganda Project*.

Pankaj Gupta, Khushbu Saxena, Usama Yaseen, Thomas Runkler, and Hinrich Schütze. 2019. Neural architectures for fine-grained propaganda detection in news. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 92–97, Hong Kong, China, November. Association for Computational Linguistics.

Benjamin D Horne, Sara Khedr, and Sibel Adali. 2018. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. In *Twelfth International AAAI Conference on Web and Social Media*.

Garth S Jowett and Victoria O'donnell. 2018. *Propaganda & persuasion*. Sage Publications.

Sopan Khosla. 2018. Emotionx-ar: Cnn-dcnn autoencoder based emotion classifier. In *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pages 37–44.

Gary King, Jennifer Pan, and Margaret E Roberts. 2017. How the chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, 111(3):484–501.

Gaurav Kumar, Rishabh Joshi, Jaspreet Singh, and Promod Yenigalla. 2020. Amused: A multi-stream vector representation method for use in natural dialogue. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 743–751, Marseille, France, May. European Language Resources Association.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. 29(3):436–465.

Saif M. Mohammad. 2018a. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.

Saif M. Mohammad. 2018b. Word affect intensities. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miyazaki, Japan.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937, Copenhagen, Denmark, September. Association for Computational Linguistics.

Kate Starbird. 2018. Medium, Oct.

Cristina Tardáguila, Fabricio Benevenuto, and Pablo Ortellado. 2018. Fake news is poisoning brazilian politics. whatsapp can stop it. *The New York Times*, 17.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

Shikhar Vashishth, Rishabh Joshi, Sai Suman Prayaga, Chiranjib Bhattacharyya, and Partha Talukdar. 2018. RESIDE: Improving distantly-supervised neural relation extraction using side information. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1257–1266, Brussels, Belgium, October-November. Association for Computational Linguistics.

## A    Hyper-parameters

| Hyper-parameter | Values |
|---|---|
| batch-size | 8, 16 |
| epochs | 20 |
| learning-rate | $3e-5$ |
| $\alpha$ | 0.9, 0.5 |
| l2 regularization ($\beta$) | 0.0001 |
| early-stopping patience | 9 |
| early-stopping criterion | Span-level F1 |
| seeds | 1, 2, 3, 12, 123, 1234, 12345 |

Table A1: Hyper-parameters

## B    Example of propaganda detection

Tudor Simionov was killed when he tried to prevent a gang from **gatecrashing** a party in Mayfair. The Page One headline in the Evening Standard said it all: 'First Stab Victims Of The New Year.' Welcome to 2019. ... He died at the scene, in Park Lane, after being attacked by up to ten men — including, allegedly, the 26-year-old son of **Finsbury Park firebrand Captain Hook**, the **Islamist hate preacher** currently serving life ... But a visible police presence may have deterred the gang who killed Tudor Simionov. ... **No doubt London's two-bob chancer of a Mayor Sadiq Khan would blame the absence of police** on the **streets on 'austerity' or the 'Tory cuts'.** Maybe if he hadn't **splurged £2.3 million of taxpayers' money on a vainglorious New Year fireworks display,** he could have diverted more money to the Met to put extra bobbies on the beat. Some of this largesse was frittered away turning the London Eye into a blue and yellow EU flag, as part of a **cynical, self-serving propaganda stunt aimed at bolstering the Stop Brexit campaign and burnishing the image of the Mayor himself.**... What's most likely to deter immigrants from coming here is not Khan slandering Leave voters as **knuckle-scraping, racist bigots,** but the news that someone can **be stabbed to death on one of London's busiest, best-known and most upmarket streets** .... And if that **chilling reality** doesn't put people off moving here from overseas, then perhaps they may be dissuaded by the **paralysis on London's roads** ... **Rough sleepers and aggressive beggars** are a permanent fixture in the West End. **Yet while innocent blood runs in the gutters,** and police budgets are restrained, Khan indulges in **fatuous propaganda stunts,** ... **London's evening paper has a splash headline marking the first stabbings of the year.**

The first of many more to come, no doubt. **Makes you proud to be British.**