

# 3218IR at SemEval-2020 Task 11: Conv1D and Word Embedding in Propaganda Span Identification at News Articles

<b>Dimas Sony Dewantara</b> Faculty of Computer Science Universitas Indonesia Depok, Indonesia Indonesian Institute of Sciences Jakarta, Indonesia dimas.sony91@ui.ac.id	<b>Indra Budi</b> Faculty of Computer Science Universitas Indonesia Depok, Indonesia indra@cs.ui.ac.id	<b>Muhammad Okky Ibrohim</b> Faculty of Computer Science Universitas Indonesia Depok, Indonesia okkyibrohim@cs.ui.ac.id
--	--	---

## Abstract

In this paper, we present the result of our experiment with a variant of 1 Dimensional Convolutional Neural Network (Conv1D) hyper-parameters value. We describe the system entered by the team of Information Retrieval Lab. Universitas Indonesia (3218IR) in the SemEval 2020 Task 11 Sub Task 1 about propaganda span identification in news articles. The best model obtained an F1 score of 0.24 in the development set and 0.23 in the test set. We show that there is a potential for performance improvement through the use of models with appropriate hyper-parameters. Our system uses a combination of Conv1D and GloVe as Word Embedding to detect propaganda in the fragment text level.

## 1 Introduction

News articles can have many forms of presentation. A news article often tends to influence the reader to have some particular view following the author's agendas. The method usually used is to add some bias to the article being written (Baeza-Yates, 2018). On a different level, news management sometimes has lacking neutrality shapes information by emphasizing positive or negative aspects purposefully (Jowett and O'Donnell, 2006). Identifying bias is not easy. Often the writer is not realized that he has included his bias in his writing. In another case, the writer deliberately introduces bias into his writings, intending to influence his readers. The last condition is known as propaganda. Propaganda is a shape of opinion or action by individuals or groups deliberately designed to change the perspective or actions of other individuals or groups regarding predetermined ends (Jackall, 1995).

Da San Martino et al. (Da San Martino et al., 2020) organized the SemEval 2020 Task 11: Detection of Propaganda Techniques in News Articles. This propaganda refers to whenever information is shaping purposefully to foster a predetermined agenda. This research focuses on subtask 1 about span identification that indicated the location of the text fragment of propaganda in news articles.

Propaganda span identification is a problem of detecting text fragment in a sentence or paragraph containing propaganda techniques. The purpose of this task is to detect the presence of words or phrases that cause bias in the information on purpose or called propaganda. This study proposes a deep learning approach, specifically 1 Dimensional Convolutional Neural Network (Conv1D) combined with word embedding (GloVe) in detecting text fragment related to propaganda. Conv1D, as part of Convolutional Neural Network (CNN), is one of the deep learning approaches that has been widely used. Convolutional Neural Network is widely used in computer vision, especially in image classifications where data is processed in the form of 2 dimensions (Hussain et al., 2018). Whereas in text processing, 1-dimensional Convolutional Neural Network (Conv1D) is used in relation to its suitability for the characteristics of text that only has 1 dimension in the form of word sequence. Conv1D was chosen because it has been known to have excellent performance in the natural language processing (NLP) task such as sentiment analysis and classification of text (Kim, 2014)(Kuttala et al., 2018) especially in the categorization of text fragments (Hughes et al., 2017) and classification in word and character level (Mo et al., 2018). Word Embedding is a language modelling and feature learning technique in (NLP) that maps words into a low

---

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

dimensional continuous space (Li et al., 2015). Word embedding as input representation has been known to increase the performance of tasks in NLP through mapping words in the form of a low-dimensional vector that allows a more expressive and efficient representation by maintaining the contextual similarity between words (Naili et al., 2017). GloVe or global vector is one type of word embedding that is known to have an excellent performance. GloVe combines global matrix factorization and local context window methods. Rather than training on the entire sparse matrix or individual context windows in the corpus, GloVe model efficiently captures statistical information by training only on the nonzero elements in a word-word co-occurrence matrix. Which then becomes a solution to the problem of computational costs and context constraints (Pennington et al., 2014).

In our work, we explain the motivation of using Conv1D and word embedding (GloVe) in the related work section. In the system overview section, we describe the system that we used in our submission. We then describe the variant of the model with Conv1D that we use to reach the best performance of our classification model in the experimental setup section. We also describe our results on the development and test data, as well as a short analysis in the experiment result and analysis section.

## **2 Related Work**

Some research has been done in detecting propaganda in news articles. Previous research on the language of news media in the context of political fact-checking and fake news detection trying to compare the language of real news with that of satire, hoaxes, and propaganda by finding linguistic characteristics of untrustworthy text (Rashkin et al., 2017). Other research proposes a model to automatically assess the level of propagandistic content in an article based on different representations, from writing style and readability level to the presence of specific keywords (Barrón-Cedeño et al., 2019). Another research is trying to utilize BERT in classifying propaganda techniques in news articles (Hua, 2019).

However, the detection of propaganda related to this task tends to be challenging because they have to be done to the level of a text fragment, and this kind of task at least was first competed on SemEval. In other words, this task still leaves opportunities in terms of performance improvement or analysis related to the possibility of using other approaches than previously done. Conv1D in several previous studies has been shown to have an excellent performance related to the classification of texts in word and character level (Mo et al., 2018) and also in text fragment level (Hughes et al., 2017). This can be achieved by utilizing the process of convolution and pooling in the word, character and text fragment level by pick out stand out features like tokens or token sequences in a way that is not fixed in their position in the input sequence (Goldberg, 2016). The use of word embedding as word vector representation has also been shown to have better performance than the use of one-hot vector directly (Zhang and Wallace, 2017). These facts motivate us to explore the potential of using Conv1D combined with word embedding in text classification, especially at text fragment level by utilizing automatic feature extraction in news articles related to propaganda.

## **3 System Overview**

### **3.1 Dataset**

The dataset used in this task is provided by SemEval 2020 Task 11. This dataset consists of the corpus with about 550 English news articles in which fragments containing one out of 18 propaganda techniques have been annotated. This dataset consists of pairs of files where the first file contains articles where rows separate each sentence. While the second file is labelled with a separated line tab format containing the article id, begin and end offset to show the location of word fragments that indicate propaganda as shown in Figure 1.

### **3.2 Methodology**

We use the text fragment classification approach as a solution for SemEval 2020 sub-task 1 span identification. The classification process involves two classes, text fragments that indicate propaganda and text fragments that do not indicate propaganda. The workflow that we use starts with data preparation, model architecture determination, training and testing. We conducted a series of experiments on the model that



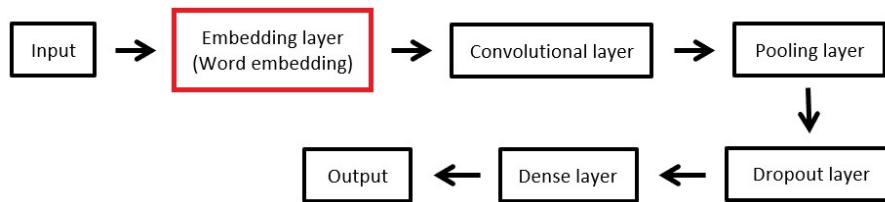


Figure 4: Model architecture

Convolution layer used in this research is Conv1D. Conv1D creates a convolution kernel that is convolved with the input over a single spatial dimension to produce outputs. The number of different convolution layers and also variations of filter values and kernel sizes will be tested in the experimental phase related to hyper-parameters tuning. The initial values for these hyper-parameters will follow the values in the example application that are used as the model baseline. The pooling layer used in this research is MaxPooling1D. This layer performs pooling operations with maximum values in the single spatial dimension data in producing output. The pooling layer is usually applied in pairs with the convolution layer. MaxPooling1D has several parameters such as pool size, strides and others. Different value of pool size will be used as one of the hyper-parameters. Dropout layer is used as a method to overcome the possibility of overfitting that often occurs when using the deep learning approach. The dropout layer randomly decreases the number of neuron units in the network to reduce connections at each iteration in the training process. The dense layer in the model functioning as a classifier by only containing 1 unit of neural network with sigmoid activation function. Classification is done to distinguish which text fragment are related or not related to propaganda.

We will use macro average F1-scores as evaluation measurements in this set of experiments according to the type of measurement evaluation used by the organizer of this task.

## 4 Experiment Setup and Analysis

This section described the setup and analysis for the approaches used in this research, such as data, standard parameters settings, hyper-parameters tuning, and analysis of the experiment results.

### 4.1 Experiment Setup

The data we use in the training and testing process is data that has been through the process of loading data and data preparation that is ready to be processed by the model. There are 16670 pairs of sentences with the label which are then separated for training for 13338 pairs and validation for 3335 pairs.

In this research, we try to focus only on hyper-parameters tuning related to convolution layer. Several hyper-parameters will be tested. Values of these hyper-parameters will be given in a range of slightly above the values at the model baseline. These hyper-parameters including the number of convolution layer (convolution and pooling layer as pair), number of filters, size of the kernel and pool size. With fixed parameter values such as learning rate of 0.001, dropout value of 0.5, batch size of 32, the number of epochs 5 times with binary crossentropy as a loss function and Adam as an optimizer. All scenarios in this experiment were carried out using a development dataset to determine which scenario produced the best performance. Scenario with the best performance will be used in processing the testing dataset to produce output for submission.

### 4.2 Experiment Result and Analysis

Based on the experiments, the scenario with the best performance is shown in Table 1. This table also contains the results of several other scenarios, including the baseline scenario.

Sc 16 scenario as the best performing scenarios uses 3 layers of convolution layers with the same number of filters and variations on kernel size and pooling size. This scenario shows a pattern of increase in the kernel size and a decreasing pattern in the pooling size from the initial layer to the next layer. The next scenario variations after Sc 16 that using 3 convolution layers with the different number of

Scenario (Sce)	Conv filters	Kernel size	Pooling size	F1-score (dev set)	Recall (dev set)	Precision (def set)
Number of conv layers: 1						
Scce 1 (Baseline)	64	3	3	0.1768	0.1258	0.2973
Scce 2	64	3	4	0.1557	0.1054	0.2979
<b>Scce 3</b>	<b>64</b>	<b>4</b>	<b>3</b>	<b>0.2004</b>	<b>0.1616</b>	<b>0.2638</b>
Number of conv layers: 2						
Scce 4	64, 64	3, 3	3, 3	0.1494	0.0925	0.3877
Scce 5	64, 64	3, 4	3, 4	0.0868	0.0485	0.4087
Scce 6	64, 64	4, 3	4, 3	0.0862	0.0492	0.3461
Scce 7	64, 64	4, 3	3, 4	0.0610	0.0330	0.3993
Scce 8	64, 64	3, 4	4, 3	0.1787	0.1274	0.2993
Scce 9	64, 128	3, 3	3, 3	0.1972	0.1552	0.2701
Scce 10	64, 128	3, 4	3, 4	0.1634	0.1137	0.2896
<b>Scce 11</b>	<b>64, 128</b>	<b>3, 4</b>	<b>4, 3</b>	<b>0.2081</b>	<b>0.1812</b>	<b>0.2443</b>
Number of conv layers: 3						
Scce 12	64, 64, 64	3, 3, 3	3, 3, 3	0.2178	0.1983	0.2416
Scce 13	64, 64, 64	3, 4, 5	3, 4, 5	0.2058	0.1672	0.2675
Scce 14	64, 64, 64	5, 4, 3	5, 4, 3	0.2095	0.1805	0.2494
Scce 15	64, 64, 64	5, 4, 3	3, 4, 5	0.2223	0.2018	0.2475
<b>Scce 16</b>	<b>64, 64, 64</b>	<b>3, 4, 5</b>	<b>5, 4, 3</b>	<b>0.2403</b>	<b>0.2399</b>	<b>0.2406</b>
Scce 17	64, 128, 256	3, 3, 3	3, 3, 3	0.2028	0.1654	0.2621
Scce 18	64, 128, 256	3, 4, 5	5, 4, 3	0.1308	0.0797	0.3630
Scce 19	128, 128, 128	3, 4, 5	5, 4, 3	0.1413	0.0898	0.3321
Scce 20	256, 256, 256	3, 4, 5	5, 4, 3	0.0640	0.0355	0.3245

Table 1: Scenario Results

convolution filters has shown a decrease in the value of F1-score achieved. This also happens in scenarios where the number of convolution layers exceeds the value of 3. Separation of scenario based on the number of convolution layers used, the best-performing scenario almost has the same pattern. This pattern includes the increasing of filters number, increasing kernel size number and decreasing pooling size number for each additional layers. Because the same pattern does not apply to Scce 16 scenarios with 3 convolution layers, it can be concluded that the pattern used in this scenario is the most optimal pattern related to the experimental arrangements in this research.

Finally, the scenarios with the best performance in the development set are then applied to the test set, which then outputs are used in the submission of the SemEval 2020 sub-task 1. From this scenario submission, it is known that the F1-score obtained in the test set is 0.2347 with a ranking of 32 from 36 participants with the highest F1-score is 0.5155.

## 5 Conclusions

Based on the results obtained from this research, it was concluded that Conv1D combined with word embedding could be used as a model in propaganda span identification problem. The best results are generated from scenarios that use several convolution layers with the same number of filters but have different kernel sizes and pooling sizes for each layer. This scenario also shows an increasing pattern in the kernel size and decreasing pattern in the pooling size from the initial layer to the next layer. This pattern can be used as consideration related to hyper-parameters tuning at the convolution layer.

Furthermore, the limited combination of hyper-parameters values used in this research leaves the possibility of achieving better performance by using combinations that have not yet been tested. Other parameters whose values are made static are also likely hiding the potential for improved performance.

## 6 Acknowledgements

The authors gratefully thank Universitas Indonesia for the International Publication (PUTI Prosiding) Grants No. NKB-877/UN2.RST/HKP.05.00/2020 Year of 2020.

## References

- Ricardo Baeza-Yates. 2018. Bias on the web. *Communications of the ACM*, 61:54–61, 05.
- Alberto Barrón-Cedeño, Israa Jaradat, Giovanni [Da San Martino], and Preslav Nakov. 2019. Proppy: Organizing the news based on their propagandistic content. *Information Processing Management*, 56(5):1849 – 1864.
- 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, September*.
- Yoav Goldberg. 2016. A primer on neural network models for natural language processing. *J. Artif. Int. Res.*, 57(1):345–420, September.
- Yiqing Hua. 2019. Understanding BERT performance in propaganda analysis. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 135–138, Hong Kong, China, November. Association for Computational Linguistics.
- Mark Hughes, Irene Li, Spyros Kotoulas, and Toyotaro Suzumura. 2017. Medical text classification using convolutional neural networks. *Studies in Health Technology and Informatics*, 235, April.
- Mahbub Hussain, Jordan Bird, and Diego Faria. 2018. A study on cnn transfer learning for image classification. In *18th Annual UK Workshop on Computational Intelligence , UKCI 2018*, June.
- R. Jackall. 1995. *Propaganda. Main Trends of the Modern World*. NYU Press.
- Garth Jowett and Victoria O’Donnell. 2006. *Propaganda and persuasion*. Thousand Oaks, Calif. : Sage, 4th ed edition. Includes bibliographical references (p. 369-391) and indexes.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, August.
- Radhika Kuttala, Bindu K R, and Latha Parameswaran. 2018. A text classification model using convolution neural network and recurrent neural network. *International Journal of Pure and Applied Mathematics*, 119:1549–1554, January.
- Yitan Li, Linli Xu, Fei Tian, Liang Jiang, Xiaowei Zhong, and Enhong Chen. 2015. Word embedding revisited: A new representation learning and explicit matrix factorization perspective. In *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI’15*, page 3650–3656. AAAI Press.
- Kyoungyun Mo, Jaesun Park, Myeongjun Jang, and Pilsung Kang. 2018. Text classification based on convolutional neural network with word and character level. *Journal of the Korean Institute of Industrial Engineers*, 44:180–188, June.
- Marwa Naili, Anja Habacha, and Henda Ben Ghezala. 2017. Comparative study of word embedding methods in topic segmentation. *Procedia Computer Science*, 112:340–349, 12.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. volume 14, pages 1532–1543, 01.
- 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.
- Ye Zhang and Byron Wallace. 2017. A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 253–263, Taipei, Taiwan, November. Asian Federation of Natural Language Processing.