

SSN_NLP_MLRG at SemEval-2022 Task 4: Ensemble Learning strategies to detect Patronizing and Condescending Language

Kalaivani Adaikkan and Durairaj Thenmozhi

Department of Computer Science and Engineering
Sri Sivasubramaniya Nadar College of Engineering, Chennai, India
kalaiwind@gmail.com, theni_d@ssn.edu.in

Abstract

In this paper, we describe our efforts at SemEval 2022 Shared Task 4 on Patronizing and Condescending Language (PCL) Detection. This is the first shared task to detect PCL which is to identify and categorize PCL language towards vulnerable communities. The shared task consists of two subtasks: Patronizing and Condescending language detection (Subtask A) which is the binary task classification and identifying the PCL categories that express the condescension (Subtask B) which is the multi-label text classification. For PCL language detection, We proposed the ensemble strategies of a system combination of BERT, Roberta, Distilbert, Roberta large, Albert achieved the official results for Subtask A with a macro f1 score of 0.5172 on the test set which is improved by baseline score. For PCL Category identification, We proposed a multi-label classification model to ensemble the various Bert-based models and the official results for Subtask B with a macro f1 score of 0.2117 on the test set which is improved by baseline score.

1 Introduction

Social media is a wide platform and it grows rapidly. People can communicate with each other and express their opinions easily without any hesitation on social media. Patronizing and Condescending Language (PCL)¹ is language use shows a superior attitude rise towards vulnerable communities in the social media. This effect is unconscious and the intention is trying to help communities like an individual, group of people in need by raising awareness, moving the user audience to action, and standing for the rights. However, these dominant attitudes can lead to discrimination and the user audience is unaware of this diminishing treatment due to its subtlety. Moreover, online social media publications reached more audiences in day-to-day life and we noticed that diminishing treatment of

vulnerable groups leads to greater inequalities. so, PCL can potentially be very harmful, as it feeds stereotypes, routinizes discrimination, and drives to greater exclusion.

Detecting PCL language and its categorization of PCL language on social media have gained a lot of interest recently. The detection of PCL is still an emergent area of study in NLP. To the best of my knowledge, this is the first shared task² to detect the PCL and their categories from the vulnerable communities. The challenge is to detect that PCL is difficult both for humans and NLP systems, due to its subtle nature, and its subjectivity reasoning required to understand this kind of language. SemEval 2022 task 4 presents the problems of detecting PCL and its categorizes of PCL which express the condescending language in English tweets to the NLP community. The PCL shared task consists of two subtasks: Subtask A is to identify the content is PCL or Not PCL. Subtask B is to classify whether the content into Unbalanced power relations, Shallow solution, Presupposition, Authority voice, Metaphor, Compassion, The poorer, the merrier.

This paper describes the systems submitted for SemEval 2022 shared task on PCL Detection by the team SSN_NLP_MLRG. We have participated in the shared task for all the two subtasks. First, we experimented with Bert-based models and we used the ensembling strategies to enhance the performance of the model. Finally, we performed voting to decide the final output. The majority voting on 5 classification models yielded better results than individual systems. The paper is organized as follows: In section 2, we describe the Background work, in section 3, we describe our models, we present the experimental setup in section 4, and compare results in section 5, we provide the conclusion of our work.

¹<https://www.merriam-webster.com/dictionary/patronizing>

²<https://competitions.codalab.org/competitions/34344>

2 Background

In this section, we describe the task provided to the participants and the two subtasks.

2.1 Task Description

The participants were required to produce labels indicating if a paragraph is PCL or Not PCL in the shared of subtask A, and we categorize the PCL Language in the shared task subtask B (Pérez-Almendros et al., 2022).

Subtask A is the binary classification task. Each text content took one of these labels for subtask A as follows: PCL Language: The content shows a superior attitude and language towards a vulnerable community in media. Not PCL: The content is not intended for the PCL Language

Subtask B is the multi-label text classification task. Each text content took into these categories of labels for subtask B as follows: Unbalanced power relations: The author keeps distance from the community or based on the situation they express the will, capacity, or responsibility to help people in need. The author also presents to give something positive to the audience in a more vulnerable situation, especially the author concedes is a right but they do not have any authority to decide to give. Shallow solution: A superficial charitable and simple action by the privileged community which is presented either as life-saving or life-changing or as a solution for a deep-rooted problem. Presupposition. The author assumes a situation as certain without having all the valid information and trustworthy source for it (e.g. a survey of research work). Examples of presupposition such as usage of stereotypes or cliches.

Authority voice: The author stands themselves as a superior power of the group, or advises the members of a community about the specific situation they are living. Metaphor. They can conceal PCL, making a comparison between unrelated concepts. An example of a metaphor is euphemisms. Compassion. The author shows the vulnerable individual or group of people about raising a feeling of pity and compassion from the audience towards them. It is commonly characterized by the use of flowery vulnerable words. The poorer, the merrier. How they spread a positive attribute towards the vulnerable community. People learn to live in vulnerable situations and to admire their values. The typical example of ‘poor people is happier because they don’t have material goods. Table 1 presents

the sample annotated data.

2.2 Related work

The authors (Pérez-Almendros et al., 2020) described a new annotated PCL dataset which is aimed to identify and categorize language that is patronizing or condescending language towards vulnerable communities and used the Bert model to detect and classify the harmful PCL language. Recently, Several works are carried out to detection of offensive language (Kalaivani and Themozhi, 2020a), hate speech (Kalaivani and Themozhi, 2020b), fake news detection, trustworthiness (Atanasova et al., 2018) and fact-checking (Elsayed et al., 2021) prediction is driven towards a particular community. The work (Fiske, 1993) presents a theory of the power of stereotyping and controlling the power of other outcomes. The author (Giles et al., 1993) analyzed the effects of responses and attitudes based on age group towards patronizing and harmful speech. Discourse analysis promises the need to satisfy the teacher’s, student’s textual values that build on techniques and provide a smoother relationship (Huckin, 2002). Margić (2017) examined the communication courtesy or condescending between the native and non-native English speakers. We observed that most of the work is related to the unfair treatment of the particular underprivileged community.

3 Methodology

We used the pre-trained models BERT (Bidirectional Encoder Representations from Transformers). We fine-tune a BERT language model (Devlin et al., 2019) for PCL classification. we also fine-tuned a RoBERTa-base (A robustly optimized Bert pretraining approach) model (Liu et al., 2019) to classify PCL Language and the PCL categories which can be expressed condescension and viewed as an optimized version of BERT. We used two variants of the RoBERTa method that are RoBERTa-large-cased and RoBERTa-base-cased pre-trained models. We also fine-tune the DistilBERT (Distilled version of BERT model) model (Sanh et al., 2019) is the transformer model, which is a lighter and faster variant of BERT. We used the ALBERT (A lite BERT) model (Lan et al., 2019) to fine-tune the system to predict the PCL language. To further explore the performance, we apply the Ensemble strategies to combine the transformers models output to predict the PCL and Category of PCL

Text	Subtask A	Subtask B
1. The scheme saw an estimated 150,000 children from poor families being sent to parts of the British Empire between 1920 and 1974 , by religious orders and charities who said they would lead better lives	1	[1, 0, 0, 1, 0, 0, 0]
2. Durban ’s homeless communities reconciliation lunch	0	[0, 1, 0, 0, 0, 0, 0]

Table 1: Sample annotated paragraph. For subtask A, '0' presents PCL and '1' presents Not PCL. For subtask B, Seven category of PCL are Unbalanced power relations, Shallow solution, Presupposition, Authority voice, Metaphor, Compassion, The poorer, the merrier

language which is based on the majority voting concept. In all cases, we trained the model for 10 epochs. Finally, we got a macro-average f1 score of 0.5172 for subtask A and f1 average score of 0.2117 for subtask B respectively.

4 Experimental setup

4.1 Data description

The dataset for SemEval 2022 Shared task 4 consists of 10,469 paragraphs are split into training, development, and testing sets for subtask A and 993 unique paragraphs, totaling 2,760 instances of PCL, for Subtask B . Don't patronize me dataset offers content from media forums. the training data size is 8,375 contents and the development data size is 2,095 contents and the size of the test data is 3,832 contents. Table 1 presents the split of experiment data. The shared task of subtask A is a binary classification task in which the aim is to build systems able to detect the given paragraph content is PCL or Not PCL. The shared task of the PCL Category classification is a Multi-label classification task in which the aim is to build systems able to classify the PCL category into Unbalanced power relations, Shallow solution, Presupposition, Authority voice, Metaphor, Compassion, The poorer, the merrier.

4.2 Data Preprocessing

We applied down sampling negative instances data augmentation techniques to balance the dataset because the negative instances are 7,581 contents and positive instances are only 794 contents. Preprocessing the text is an important role as the data from social media can be quite noisy and contain a lot of noisy words, excessive use of punctuation, URLs, symbols, misspelling words. We perform data preprocessing by using NLTK libraries. First, we remove the duplication because it affects the system performance. we remove the stop words.

After that, we remove the punctuations, URLs, numerals, emojis and then convert all the upper case English text into lower case text.

4.3 Experimental setting

For both subtasks A and B, We implement the Ensemble model using Simple transformers. We used the colab notebook for implementation purposes with the high-end RAM, GPU for training. For the hyperparameters for the BERT-based five models, we set epochs as 10. For the multilabel classification task, we used simple transformers and a multi-label classification model to predict the PCL Language category. we analyzed the individual classification of all five BERT-based models for both the subtasks. We also examined the final output which is the combination of five models based on a majority voting system for classification.

5 Experimental analysis

This section presents the analysis of the results and submitted official results

5.1 Result Analysis

We experimented with the various transformer model are BERT, DistilBERT, AIBERT, Roberta base, Roberta large, and the ensemble of all five models. We analyzed the comparison scores of various approaches based on the evaluation metrics of precision, recall, and macro average f1 score for the shared task A. Task 1 is a binary classification task that will be evaluated using f1 over the positive class. Task 2 is framed as a multi-label classification problem. For each paragraph, your model will assign a label for each of the seven PCL categories. Then, results for this task will be evaluated using per-class f1, and the final ranking for this subtask will be based on macro-average f1. Table 2 presents the results of subtask A. Table 3 shows the results of subtask B.

Model	Precision	Recall	Macro f1
Run 1:			
BERT model	0.3832	0.6940	0.4938
Run 2:			
Ensemble model	0.4228	0.6656	0.5171

Table 2: Test Results of subtask A

Model	UNB POW	SHAL	PRES	AUTH	MET	COMP	MERR	Avg f1
Run 1:								
BERT	0.3459	0.3376	0.2068	0.1933	0.1212	0.2772	0.0	0.2117
Run 2:								
AIBERT	0.3438	0.3157	0.2056	0.1666	0.1666	0.2677	0.0	0.2094

Table 3: Test Results of subtask B. UNB POW - Unbalanced power relations, SHAL - Shallow solution, PRES - Presupposition, AUTH - Authority voice, MET - Metaphor, COMP - Compassion, MERR - Poorer and Merrier

We submitted two runs for both of the subtasks. For run 1, we submitted the prediction made by the BERT model for subtasks A and B. For run 2, we submitted the prediction made from the ensemble model for subtask A and Albert model for subtask B. We observed that the performance of the Ensemble model achieved good results compared to the BERT model for subtask A and the performance of the BERT model achieved good results compared to the Albert model. Finally, we got the macro f1 score of ensemble model is 0.5172 for the subtask A and the macro f1 average of Bert model is 0.2117 for the subtask B respectively.

6 Conclusion and Future Work

This paper presents the submitted runs for the patronizing and condensing language identification in SemEval 2022 task 4. The results show that the Not-PCL language and PCL language in the dataset receives the same macro f1 scores. We experimented with different approaches such as a BERT model, AIBERT, Roberta base, Roberta large, and Distilbert and Ensemble models. Based on the evaluation, BERT performs well for subtask B to classify the PCL content into Seven categories that express condensing language. Ensemble model performs well for subtask A to detect the content is PCL language or Not PCL language. Our team submission had a macro f1 score of 0.5172 for subtask A and a macro f1 score of 0.2117 for subtask B which are improved by the baseline f1 scores. For future work, we will handle the imbalanced dataset by using external resources and apply the data augmentation techniques to enhance the performance of our model.

References

- Pepa Atanasova, Alberto Barrón-Cedeño, Tamer Elsayed, Reem Suwaileh, Wajdi Zaghouani, Spas Kyuchukov, Giovanni Da San Martino, and Preslav Nakov. 2018. [Overview of the CLEF-2018 checkthat! lab on automatic identification and verification of political claims. task 1: Check-worthiness](#). *CoRR*, abs/1808.05542.
- 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.
- Tamer Elsayed, Preslav Nakov, Alberto Barrón-Cedeño, Maram Hasanain, Reem Suwaileh, Giovanni Da San Martino, and Pepa Atanasova. 2021. [Overview of the CLEF-2019 checkthat!: Automatic identification and verification of claims](#). *CoRR*, abs/2109.15118.
- Susan Fiske. 1993. [Controlling other people: The impact of power on stereotyping](#). *The American psychologist*, 48:621–8.
- Howard Giles, Susan Fox, and Elisa Smith. 1993. [Patronizing the elderly: Intergenerational evaluations](#). *Research on Language and Social Interaction*, 26(2):129–149.
- Thomas N. Huckin. 2002. Critical discourse analysis and the discourse of condescension.
- A Kalaivani and D Thenmozhi. 2020a. [SSN_NLP_MLRG at SemEval-2020 task 12: Offensive language identification in English, Danish, Greek using BERT and machine learning approach](#). In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 2161–2170, Barcelona

(online). International Committee for Computational Linguistics.

- A. Kalaivani and D. Thenmozhi. 2020b. [SSN_NLP_MLRG@HASOC-FIRE2020: Multilingual Hate Speech and Offensive Content Detection in Indo-European Languages using ALBERT](#). In *Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020*, volume 2826 of *CEUR Workshop Proceedings*, pages 188–194. CEUR-WS.org.
- Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. [ALBERT: A lite BERT for self-supervised learning of language representations](#). *CoRR*, abs/1909.11942.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692.
- Branka Drljaca Margić. 2017. Communication courtesy or condescension? linguistic accommodation of native to non-native speakers of english. *Journal of English as a lingua franca*, 6:29–55.
- Carla Pérez-Almendros, Luis Espinosa-Anke, and Steven Schockaert. 2020. Don't Patronize Me! An Annotated Dataset with Patronizing and Condescending Language towards Vulnerable Communities. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5891–5902.
- Carla Pérez-Almendros, Luis Espinosa-Anke, and Steven Schockaert. 2022. SemEval-2022 Task 4: Patronizing and Condescending Language Detection. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Association for Computational Linguistics.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.01108.