I2C at SemEval-2022 Task 4: Patronizing and Condescending Language Detection using Deep Learning Techniques

Laura Vázquez Ramos
Escuela Técnica Superior de Ingeniería
Universidad de Huelva (Spain)
laura.vazquez005@alu.uhu.es

Adrián Moreno Monterde
Escuela Técnica Superior de Ingeniería
Universidad de Huelva (Spain)
adrian.moreno521@alu.uhu.es

Victoria Pachón Álvarez
Escuela Técnica Superior de Ingeniería
Universidad de Huelva (Spain)
vpachon@uhu.es

Jacinto Mata Vázquez
Escuela Técnica Superior de Ingeniería
Universidad de Huelva (Spain)
mata@uhu.es

Abstract

Patronizing and Condescending Language is an ever-present problem in our day-to-day lives. There has been a rise in patronizing language on social media platforms manifesting itself in various forms. This paper presents two performing deep learning algorithms and results for the "Task 4: Patronizing and Condescending Language Detection." of SemEval 2022. The task incorporates an English dataset containing sentences from social media from around the world. The paper focuses on data augmentation to boost results on various deep learning methods as BERT and LSTM Neural Network.

1 Introduction

Nowadays, Patronizing and Condescending Language (PCL) (Pérez-Almendros, Espinosa-Anke, and Schockaert 2022) is used to refer to a forced kindness that derives from a perceived superiority towards another person. A subtle form of bullying, being patronized can leave a person feeling infuriated and impotent.

It is a type of behaviour that cuts across generations. An older person can talk down to a younger colleague, but it can just as easily happen the other way around. Men can patronize women at work and vice versa. But what they have in common is power play, with one individual exerting their authority or seniority over another.

People could feel discriminated by condescending comments. Considering the relevance of equality and respect, and it is important to note that nobody should be treated in such a way as to feel intimidated or different. In SemEval-2022: Task 4 Subtask 1, participants must determine whether a phrase presents PCL or not.

The idea behind the proposed solution was to compare three models based on deep learning. The first model is a Long short-term memory (LSTM) neural network (Zhou et al. 2015). The second, a LSTM neural network with embedding pretrained layer. The third one uses BERT (Devlin et al. 2018). The latter yielded the best results. For all the models we have used data augmentation to balance the training dataset.

The F1-score in our final submission (BERT) was 0.4134 on the test dataset. Compared to the results of the winning team, the difference is not extremely large. Nevertheless, our model obtained a good result of accuracy (0.61) compared to other participants.

2 Background

This paper is focused on Subtask 1: Binary classification. The corpus provided to perform subtask 1 (Pérez-Almendros, Espinosa-Anke, and Schockaert 2020) was composed of 10473 documents with 6 features: "id", "doc id", "keyword", "country code", "text" and "label". The dataset was imbalanced with respect to the class "label". Out of 10473 rows, 9470 were from the negative class and 993 from the positive class. English data augmentation was applied to the original dataset to balance them (Shorten, Khoshgoftaar, and Furht 2021).

Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), pages 459 - 463
July 14-15, 2022 ©2022 Association for Computational Linguistics

After being balanced with data augmentation, the rows of the dataset increased to 19401, 9926 (class 1) and 9470 (class 0). In order to train the three proposed models, we only used the "text" and "label" features. The csv file used follows this structure:

- text: "*The pope as well called on the congregation to reach out…*"
- label: *"1"*

3 Related work

In recent years, several people have been researching this topic. A few years ago, an experiment of the impact of age, race, and stereotypes on perceptions of language performance and patronizing speech was published (Atkinson and Sloan 2017). Indeed, a research of different types of behaviours in healthcare settings (as condescending language) was published to show the impact it has in the world (Felblinger 2009). This is a recent problem and there are significant challenges to be overcome.

4 System overview

4.1 Balancing data techniques

Imbalanced data refers to types of datasets where the target class has an uneven distribution of observations. Sometimes, when the records of a certain class outnumber the other class, our classifier may become biased towards the prediction.

Before considering whether to use balancing techniques, we analyzed the data provided and we trained the models with the original dataset to test the results.

For this purpose, we trained both LSTM neural network and BERT models. The results were as expected. Both models reached similar results. To summarize, they obtained an accuracy of 0.91 and a ROC curve of 0.66. Given the obtained results, we decided to apply some balancing techniques to the original dataset.

Random Under-sampling

Random Under-sampling (Prusa et al. 2015) is a technique to remove examples from the majority class. However, this approach can result in the loss of valuable information during model training. The original dataset was extremely imbalanced, so rows of negative class were removed to achieve a balanced dataset.

The new dataset created using this technique totaled 1986 rows, 993 for each class.

Data Augmentation

Due to the results obtained with the under-sampling method, data augmentation was used to balance the data. Among the most common data augmentation techniques, synonym substitution was used. The synonym augmenter (Wordnet, English) (Miller 1995) to create synonym phrases for the minority class was applied. An example of this kind of substitution is:

- Sentence: *"A quick fox jumps over the lazy dog"*
- Synonym sentence: *"A prompt dodger jumps over the lazy domestic dog"*

Finally, the balanced dataset had a distribution of 9926 (class 1) and 9470 (class 0).

Table 1 shows the results obtained after the application of the two balancing methods. As can be seen, data augmentation produced better results.

Model	Under sampling		Data Augmentation	
	Accuracy	ROC Curve	Accuracy	ROC Curve
LSTM Neural Network	0.73	0.73	**0.97**	**0.97**
LSTM Neural Network with embedding pretrained layer	0.70	0.70	0.96	0.96
BERT-base-uncased	0.81	0.81	**0.97**	**0.97**

Table 1: Results obtained using sampling techniques for balancing the dataset

4.2 Models

Based on LSTM Neural Networks and BERT, three different models were implemented.

LSTM Simple Neural Network

LSTM is a special type of recurrent neural network. The main feature of recurrent networks is that information can persist by introducing loops in the network diagram, so they can basically 'remember' previous states and use this information to decide what will be next.

The LSTM was composed of an embedding layer with a size of 200, two bidirectional LSTM layers, a dense layer, a drop layer and, finally, a dense layer with a sigmoid activation.

The parameters used to train the LSTM Neural Network were a batch size of 32 and 10 epochs. Indeed, early stopping was invoked to avoid over-training.

LSTM Simple Neural Network with pretrained embedding layer

A pretrained layer was added to the model described above using GloVe [1] (Pennington, Socher, and Manning 2014, 1532-1543). GloVe is a type of implementation of an inter-contextual model, so that each word that appears in training will have a single vector representation obtained by collapsing all the information available about this word with all its appearances in the data.

Some pretrained word vectors of different sizes were downloaded. Finally, we used a file with a size of 200d. Then we added a weight matrix to the first layer of the recurrent neural network.

BERT: Bidirectional Encoder Representations from Transformers

A BERT-based transformer was used to train our third model. In particular, the model implemented was “BERT-base-uncased”, which consists of 12 transformer layers, 12 self-attention heads per layer, and a hidden size of 768.

A transformer (Vaswani et al. 2017), has an attention mechanism that learns contextual relations between words (or sub-words) in a text. In its vanilla form, transformer includes two separate mechanisms — an encoder that reads the text input and a decoder that produces a prediction for the task. Since BERT’s goal is to generate a language model, only the encoder mechanism is necessary.

The model was fine-tuned with the balancing dataset. Before training, every word was set to lower case. The model was trained with a batch size of 32 and 5 epochs.

5 Experimental setup

To set up our models, some libraries were used. Some of them were “NLTK” (Wang and Hu 2021, 1041-1049), “Keras” [2], “TensorFlow” (Joseph, Nonsiri, and Monsakul 2021, 85-111), “Scikit-learn” (Hao and Ho 2019) and “Pandas” (Stepanek 2020).

To test whether the data was balanced, the original training dataset was used. Once the balancing method was decided, the training data was split into two parts – 80% for training and 20% for test, using a stratify approach.

The stratify parameter makes a split so that the proportion of values in the sample produced will be the same as the proportion of values provided for the parameter to stratify.

During the training phase, we evaluated our models with accuracy, ROC curve, precision, recall and F1-score measures.

Once the organizers provided the test data, the models with the original training dataset were trained without splitting.

6 Results

The two models submitted were LSTM Neural Network with pretrained embedding model and BERT-base-uncased. According to the official metrics (F1-score for the positive class), a result of 0.413 and 0.61 of accuracy was obtained. BERT-base-uncased reached the best results.

After training using under sampling and data augmentation methods, we concluded that data augmentation had the best results.

Table 2 shows a summary of the results obtained during the evaluation phase using data augmentation.

7 Conclusions

In this paper we present our approach and system description on Task 4 (Subtask 1) in SemEval 2022: Patronizing and Condescending Language

[1] https://nlp.stanford.edu/projects/glove/

[2] https://keras.io/

Model	Training			Test		
	Accuracy	ROC Curve	F1-score (class 1)	Accuracy	ROC Curve	F1-score (class 1)
LSTM Neural Network	0.97	0.97	0.98	0.95	0.95	0.94
LSTM Neural Network with embedding pretrained layer	0.96	0.96	0.97	0.96	0.96	0.97
BERT-base-uncased	0.96	0.96	0.96	0.96	0.96	0.95

Table 2: Results obtained during the evaluation phase using data augmentation

Detection towards communities in the media. The main aim was to develop three deep learning models using data augmentation to solve imbalanced problem of the original dataset. We implemented three different models. After training and analyzing each model, an F1-score of 0.41 in the evaluation process for class “1” was achieved. For future works, we think the models could be further improved by training with a bigger dataset or using more balancing techniques.

References

Atkinson, Jaye L. and Robin G. Sloan. 2017. *"Exploring the Impact of Age, Race, and Stereotypes on Perceptions of Language Performance and Patronizing Speech." Journal of Language and Social Psychology 36* (3): 287-305. doi:10.1177/0261927X16662967. https://doi.org/10.1177/0261927X16662967.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. *"BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding."* CoRR abs/1810.04805.

Felblinger, Dianne M. 2009. *"Bullying, Incivility, and Disruptive Behaviors in the Healthcare Setting: Identification, Impact, and Intervention." Frontiers of Health Services Management 25* (4): 13-23. https://www.proquest.com/scholarly-journals/bullying-incivility-disruptive-behaviors/docview/203882663/se-2?accountid=14549.

Hao, Jiangang and Tin Kam Ho. 2019. *Machine Learning made Easy: A Review of Scikit-Learn Package in Python Programming Language. Vol. 44*. Los Angeles, CA: SAGE Publications. doi:10.3102/1076998619832248. https://journals.sagepub.com/doi/full/10.3102/1076998619832248.

Joseph, Ferdin Joe John, Sarayut Nonsiri, and Annop Monsakul. 2021. *"Keras and TensorFlow: A Hands-on Experience." In Advanced Deep Learning for Engineers and Scientists: A Practical Approach,* edited by Kolla Bhanu Prakash, Ramani Kannan, Alex, S. Albert er and G. R. Kanagachidambaresan, 85-111. Cham: Springer International Publishing. doi:10.1007/978-3-030-66519-7_4". https://doi.org/10.1007/978-3-030-66519-7_4.

Miller, George A. 1995. *"WordNet: A Lexical Database for English." Commun.ACM 38* (11): 39–41. doi:10.1145/219717.219748. https://doi.org/10.1145/219717.219748.

Perez-Almendros, Carla, Luis Espinosa-Anke, and Steven Schockaert. 2020. *"Don't Patronize Me! an Annotated Dataset with Patronizing and Condescending Language Towards Vulnerable Communities."*.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. *"GloVe: Global Vectors for Word Representation."Association for Computational Linguistics, oct. doi:10.3115/v1/D14-1162"*. https://aclanthology.org/D14-1162.

Pérez-Almendros, Carla, Luis Espinosa-Anke, and Steven Schockaert. 2022. *"SemEval-2022 Task 4: Patronizing and Condescending Language Detection."Association for Computational Linguistics .*

Prusa, Joseph, Taghi M. Khoshgoftaar, David J. Dittman, and Amri Napolitano. 2015. *"Using Random Undersampling to Alleviate Class Imbalance on Tweet Sentiment Data."*. doi:10.1109/IRI.2015.39.

Stepanek, Hannah. 2020. *Thinking in Pandas: How to use the Python Data Analysis Library the Right Way*. Berkeley, CA: Apress. doi:10.1007/978-1-4842-5839-2. https://library.biblioboard.com/viewer/087e187a-b660-11ea-a44d-0a7fc7c4e64f.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. *Attention is all You Need.*

Wang, Meng and Fanghui Hu. 2021. *"The Application of NLTK Library for Python Natural Language Processing in Corpus Research." Theory and Practice in Language Studies* 11 (9): 1041-1049. doi:10.17507/tpls.1109.09.

Zhou, Chunting, Chonglin Sun, Zhiyuan Liu, and Francis C. M. Lau. 2015. *A C-LSTM Neural Network for Text Classification.*