MLModeler5 at SemEval-2023 Task 3: Detecting the Category and the Framing Techniques in Online News in a Multi-lingual Setup

Arjun Khanchandani, Nitansh Jain, Jatin Bedi
Computer Science and Engineering Department
Thapar Institute of Engineering and Technology
arjunkhanchandani@gmail.com, *njain_be20@thapar.edu*, *jatin.bedi@thapar.edu*

Abstract

Throughout the ages, we have seen various leaders, dictators, politicians, CEOs, companies, sports agencies, organizations like the UN harness the power of news articles to shape and influence the minds of the people of the world. Various persuasion and framing techniques are employed while writing articles to achieve this aim. It has become necessary now, more so than ever, to identify these techniques and strategies to help understand these articles, reduce the spread of fake news, stop propaganda stories and ensure that people of the world are well informed. To help achieve these tasks, we have implemented RoBERTa and ALBERT models for Subtask 1/2 Task 3 SemEval-2023. Both these models are BERT-based architectures. The results of these models are validated on the dataset provided by the organizers of this task.

1 Introduction

Popular mass media comes together and shapes the perceptions of people and communities on numerous subjects of our society [1]. Monitoring mass media in different languages and understanding and analyzing niche as well as broad domain topics is essential in this day and age.

With the increase in the amount of data being generated and processed, and the proliferation of news as well as information around us in recent times, it is essential to perform in-depth comparisons of how data is being put in front of the public by different mass media of different countries. There is also a need to promote and support the automation of this media analysis process [10].

The main drive behind this task is to foster the development of methods and tools to support the analysis of online media content in order to understand what makes a text persuasive: which writing style is used, what critical elements are highlighted, and which persuasion techniques are used to influence the reader's mind and thinking. There is an essential need to identify and develop methods that identify what techniques are being used in media content.

"The term persuasion comes from the Latin word persuasion, which could be translated as an act of persuading, convincing. Persuasion is a process, where the main motive is talking a recipient round; convincing, mentally influencing, guiding, or directing someone from one place to another." [5].

In a polarized media environment, partisan media outlets intentionally frame news stories in a way to advance specific political agendas[7]. "Even when journalists make their best efforts to pursue objectivity, media framing often favors one side over another in political disputes, thus always resulting in some degree of bias" [4]. Hence, a news framing analysis is helpful because it not only tells us whether a news article is left- or right-leaning (or positive or negative), but also reveals how the article is structured to promote a specific side of the political spectrum.

Framing is a phenomenon primarily studied and debated in the social sciences, where, for example, researchers explore how news media shape debate around policy issues by deciding what elements of an issue to emphasize, and what to exclude. Theories of framing posit that these decisions give rise to thematic sets of interrelated ideas, imagery, and arguments, which tend to cohere and persist over time [2]. In a widely cited definition, [3] argues that "to frame is to select some elements of a perceived reality and make them more salient in a communicating text, in such a way as to promote problem definition, causal

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 1096–1101
July 13-14, 2023 ©2023 Association for Computational Linguistics

interpretation, moral evaluation, and/or treatment recommendation for the item described."

The task focuses on extending the analytical functionalities of media analysis solutions to: automated detection of framing dimensions and persuasion techniques, and visualization of related statistics, etc. The three subtasks cover several complementary elements of what makes a text persuasive - the genre: opinion, report or satire, the framing: what critical elements highlight the rhetoric and the persuasion techniques that are used to influence the reader [9]. This paper focuses on **Subtask 1 - News Genre Categorisation** and **Subtask 2 - Framing Detection**.

This paper presents our system for **Shared Task on Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup @ SemEval 2023**. Two BERT-based architectures - RoBERTa and ALBERT - have been implemented for both Subtask 1 and Subtask 2.

Following are the links to the code for Subtask 1 and Subtask 2

1. Github link for subtask 1

2. Github link for subtask 2.

2 Task Description

The shared task focuses on annotating given new articles by: (a) new category, (b) framing techniques, and (c) persuasion techniques contained in the text. We have been provided with a multilingual dataset in the following 6 languages - **English, French, Italian, German, Polish and Russian** - as training and validation data. We were also provided with 3 more languages - **Spanish, Greek and Georgian** along with the previously mentioned languages as testing data. The aforementioned tasks are mentioned below.

2.1 Subtask 1 : News Genre Categorisation

The first subtask focuses on News Categorisation. It requires us to develop a classifier that can determine whether an article is one of the following - **an Opinion Piece, a Satire Piece or a Piece that Aims at Objective Reporting.**

2.2 Subtask 2 : Framing Detection

The second subtask focuses on Frame Detection. It requires us to develop a multi-label classifier to determine the frames (one or more) used in each article out of a pool of 14 domain-independent framing dimensions which are - **Economic, Capacity and Resources, Fairness and Equality, Legality, Constitutionally and Jurisprudence, Policy Prescription and Evaluation, Crime and Punishment, Security and Defence, Health and Safety, Quality of Life, Cultural Identity, Political and, External Regulation and Reputation.**

3 Data Description

The input for all tasks were news and web articles in a plain text format. Each article appears in one '.txt' file. All the articles had titles on the first row followed by an empty row. The content of the articles starts from the third row.

Articles in six languages were provided, that revolved around a fixed range of topics such as COVID-19, climate change, abortion, migration, the build-up leading to the Russo-Ukrainian war, and events related and triggered by the aforementioned war, and some country-specific local events such as elections, etc. The media selection covers both mainstream media and alternative news and web portals, a large fraction of which were identified by fact-checkers and media credibility experts as potentially spreading mis-/disinformation. Articles whenever possible were retrieved with the Trafilatura library or other similar web-scraping tools, and otherwise were retrieved manually.

News aggregation engines like Google News or the Europe Media Monitor were used for large-scale multi-lingual news aggregation, and online services such as NewsGuard and MediaBiasFactCheck were used for the purposes of fact checking and all the annotations done to the articles to provide us with the labels were done by the organizers. Each language provided in training and validation data has a folder with the articles and the labels associated with each article. The testing data doesn't include the labels.

4 Methodology

4.1 Data Pre-processing

Data Pre-processing means manipulating raw data using various techniques, tools and methods to clean it, normalize it, augment it and overall improve it, so as to ensure best the possible performance and results from the models implemented on the now pre-processed data.

In the context of textual data, data pre-processing includes removing noise from the data like punctuation marks, emojis, links, etc. Various libraries like nltk, spacy and nlpaug were used to preprocess the provided data.

News articles in six different languages were provided for the purposes of training and improving the proposed models. Data pre-processing involved tokenizing the text following which punctuations, white space, individual letters and stopwords were removed. The text was converted to lowercase and then lemmatized. To handle both unbalanced labels and to increase the training data we used nlpaug to augment the data. Parameters used in nlpaug were - “model_path=bert-base-cased”, “action=substitute” and “aug_max=3”.

For subtask 1, two pre-processed data sets were used, one with numbers and one without numbers. For subtask 2, the pre-processed data set included numbers.

4.1.1 Subtask 1

The training and validation articles of the other languages were translated into English using the deep_translator attribute of Google Translator Library. The headlines and articles were translated and pre-processed separately. The training and validation text of each language was pre-processed separately and the final training data formed from the merging of training data of the following languages (English, French, Italian, German, Polish and Russian) and validation data of the following languages (French, Italian, German, Polish and Russian) had a total of 1438 samples (1074 Opinion, 271 Reporting, 93 Satire). Label Encoding was used to convert the labels into numbers. Around 850 samples of Opinion and 70 samples of Reporting were removed to handle the imbalanced nature of the dataset upon which NLP Augmentation was implemented using the nlpaug library. 420 samples of Opinion, 404 samples of Reporting and 184 samples of Satire were left after data preporcessing and data augmentation.

4.1.2 Subtask 2

The training and the validation articles were used from the English language data only. The headlines and articles were pre-processed separately using Python libraries like NLTK and Spacy. We have used MultiLabelBinarizer to convert the comma-separated labels into a numerical binary matrix indicating the presence of a class label. We have used NLP Augmentation using the nlpaug library to increase the size of the training dataset.

4.2 Model Building

4.2.1 RoBERTa & ALBERT

The weights of the RoBERTa [8] and ALBERT [6] layer were initialized using “roberta-base” and "albert-base-v2" pre-trained weights, respectively, with the number of labels equal to three and fourteen as per the subtask 1/2 requirements. The text data needs to be encoded before it is fed into the RoBERTa or ALBERT architecture. The sentences were tokenized and then padded to the maximum length as a part of the encoding process with the maximum length being 512. If the length of the sentence exceeds 512, it is truncated. The encoded sentences were then processed to yield contextually rich pre-trained embeddings which were passed through the RoBERTa transformer (TFRobertaForSequenceClassification) or ALBERT transformer (TFAlbertForSequenceClassification) followed by a Dropout, Flatten and two Dense Layers.

Subtask 1 : A softmax activation function was used for the final dense layer. The softmax activation function gives the probabilities of each class occurring. Softmax assigns decimal probabilities to each class in a multi-class problem. Those decimal probabilities must add up to 1.0. This additional constraint helps training converge more quickly than it otherwise would.

Subtask 2 : A sigmoid activation function was used for the final dense layer. The sigmoid function is used in this multilabel classification problem because the probabilities produced by a sigmoid function are independent, and are not constrained to sum to 1.0. That’s because the sigmoid function looks at each raw output value

Classes	Class Weight 1	Class Weight 2
Opinion	0.446	0.8
Reporting	1.768	0.832
Satire	5.151	1.826

Table 1: *Old and New Class Weights*

separately and thus it's the optimal activation function of a multilabel classification problem.

5 Experimental Setup and Results

In this section, we will discuss the effects of the RoBERTa and ALBERT model on subtask 1 and subtask 2. The parameters used by the overall model for both subtasks are different and will be mentioned in their specific subsections below.

5.1 Subtask 1

The pre-processed data (both with and without numbers) was passed through a classification model, the architecture of which is defined above. Adam Optimiser was used with learning rate equal to 0.00001, loss function equal to SparseCategoricalFocalLoss and batch size equal to 16. Class weights were set while fitting the model on the training data. We focused primarily on 'F1 macro' in this subtask and have set up checkpoints to save the models with the best 'F1 macro' score. The model was run for 20 epochs.

We received a ranking of two in English Subtask 1 Task 3. The F1 macro score was **0.61632** and the F1 Micro score was **0.62963**. This result was obtained using the RoBERTa model. In the official submission, class weights were calculated based on the entire dataset, including the translated text.

Post SemEval Evaluation, the class weights were altered so that it was based on only the training data used to train the model, that was obtained after undersampling and nlp augmentation. It resulted in the changes mentioned in Table 3. The old and new class weights have been mentioned in Table 1.

5.1.1 RoBERTa

Our official submission using the RoBERTa model resulted in a macro F1 score of 0.61632 and micro F1 score of 0.62963 as mentioned in Table 1. Post modifications in class weights, we were able to achieve higher F1 scores both macro and micro on both the test data sets (including and not including numbers) as can be seen in Table 2. We can also observe from Table 2 that the dataset with numbers provides a higher F1 Macro score as compared to the dataset without numbers.

5.1.2 ALBERT

The F1 macro and micro values obtained using the ALBERT model and class weights 1 are 0.35714 and 0.38889 respectively . Post modifications in class weights, i.e. using class weights 2, we achieved a higher F1 macro score of 0.61503 as compared to the results obtained using class weights on the dataset including numbers. In contrast, we achieved a lower F1 macro score of 0.43557 on the dataset not including numbers.

It was also observed that the dataset including numbers generally provided a better result than the dataset not including numbers in both RoBERTa and ALBERT.

5.2 Subtask 2

The pre-processed data was passed through a classification model, the architecture of which is defined above. We used Adam Optimiser with learning rate equal to 0.00001, loss function equal to Binary Cross Entropy Loss and batch size equal to 8. Class weights were set while fitting the model on the training data. We focused primarily on 'F1

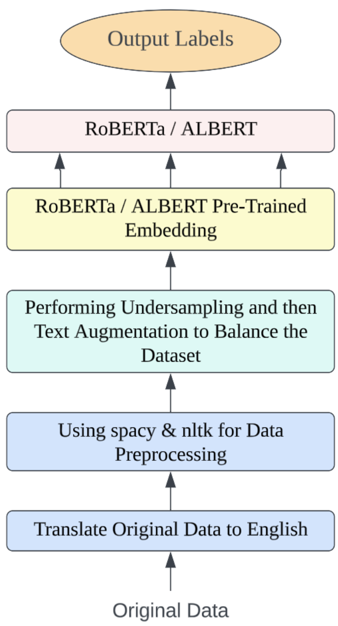

Figure 1: *Basic Architecture for Subtask 1*

Models	RoBERTa		ALBERT	
Metrics	F1 Macro	F1 Micro	F1 Macro	F1 Micro
With Numbers	0.61632	0.62963	0.35714	0.38889

Table 2: *Test Results obtained for Subtask 1 from RoBERTa and ALBERT using class weights 1*

Models	RoBERTa		ALBERT	
Metrics	F1 Macro	F1 Micro	F1 Macro	F1 Micro
With Numbers	0.66077	0.70370	0.61503	0.64815
W/o Numbers	0.63203	0.68519	0.43557	0.64815

Table 3: *Test Results obtained for Subtask 1 from RoBERTa and ALBERT using class weights 2*

Models	RoBERTa		ALBERT	
Metrics	F1 Macro	F1 Micro	F1 Macro	F1 Micro
With Numbers	0.493	0.602	0.410	0.530
W/o Numbers	0.397	0.650	0.357	0.530

Table 4: *Dev Results obtained for Subtask 1 from RoBERTa and ALBERT using class weights 2*

micro' in this subtask. The model was run for 400 epochs.

Text Augmentation-1 - "nlpaug" python library was used for data augmentation and to increase the size of dataset. In text augmentation 1, aug_max = 3 was used in the contextual word embedding function, and each article in the dataset was augmented twice.

Text Augmentation-2 (Post SemEval Evaluation) - In text augmentation 2, aug_max = 4 was used in the contextual word embedding function, and each article in the dataset was augmented thrice.

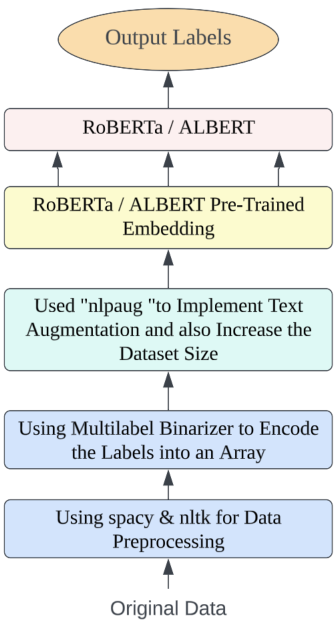

Figure 2: *Basic Architecture for Subtask 2*

5.2.1 RoBERTa

We received a ranking of 13 in English Subtask 2 Task 3. The F1 Micro and Macro obtained on our official submission using Text Augmentation-1 Parameters were 0.47692 and 0.42724 respectively as shown in Table 5.

5.2.2 ALBERT

A submission was made using the ALBERT system post the final deadline using Text Augmentation-1 Parameters. The F1 micro and macro values obtained are 0.35455 and 0.27504 respectively as shown in Table 5.

6 Conclusion

Through this paper we describe our experiment and results over the SemEval-2023 Task 3 dataset consisting of news articles in six different languages. We have proposed the implementation of two strategies - RoBERTa and ALBERT. RoBERTa out performed ALBERT across subtask 1/2 and across both the metrics F1 Macro and Micro. The usage of nlpaug for 'text augmentation' and class weights to handle the unbalanced nature of the data had a vast improvement on the metrics.

The best value of subtask 1 - 0.66077 F1 Macro - was achieved Post SemEval Evaluation using RoBERTa and class weights 2 on the dataset that included numbers and the best value of subtask 2 - 0.47692 F1 Micro - was achieved during the SemEval evaluation using RoBERTa and Text

Models	RoBERTa		ALBERT	
Metrics	F1 Macro	F1 Micro	F1 Macro	F1 Micro
Text Augmentation 1	0.42724	0.47692	0.27504	0.35455
Text Augmentation 2	0.36152	0.44160	0.24821	0.33560

Table 5: *Tests Results obtained for Subtask 2 from RoBERTa and ALBERT method based on Text Augmentation 1 and 2 parameters.*

Augmentation 1 parameters.

In the future, better data pre-processing techniques, fine-tuning of the NLP augmentation method, including increasing the numbers of words being substituted or the number of sentences being formed in the augmentation process, and fine-tuning the hyperparameters of RoBERTa and ALBERT models should be considered for better and more accurate results.

References

[1] Adel Alfatease, Ali M Alqahtani, Khalid Orayj, and Sultan M Alshahrani. The impact of social media on the acceptance of the covid-19 vaccine: a cross-sectional study from saudi arabia. *Patient preference and adherence*, pages 2673–2681, 2021.

[2] Dallas Card, Amber Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, 2015.

[3] Robert M Entman. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58, 1993.

[4] Robert M Entman. Media framing biases and political power: Explaining slant in news of campaign 2008. *Journalism*, 11(4):389–408, 2010.

[5] Monika Hossová-Jakub Glajza. Persuasion techniques in media communication of politicians. *MEGATRENDS AND MEDIA*, page 214.

[6] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*, 2019.

[7] Matthew S Levendusky. Why do partisan media polarize viewers? *American journal of political science*, 57(3):611–623, 2013.

[8] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[9] Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. Semeval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation*, SemEval 2023, Toronto, Canada, July 2023.

[10] Hetti Waluati Triana, Martin Kustati, Yunisrina Qismullah Yusuf, and Reflinaldi Reflinaldi. The representation of women in covid-19 discourses: The analysis of sara mills' critical discourse on media coverage. *Journal of Language and Linguistic Studies*, 17(1):553–569, 2021.