MLModeler5 @ Causal News Corpus 2023: Using RoBERTa for Casual Event Classification

Amrita Bhatia, Ananya Thomas, Nitansh Jain, Jatin Bedi
Computer Science and Engineering Department
Thapar Institute of Engineering and Technology
nonie.bhatia@gmail.com, *ananyathomas10@gmail.com*
njain_be20@thapar.edu, *jatin.bedi@thapar.edu*

Abstract

Identifying cause-effect relations plays an integral role in the understanding and interpretation of natural languages. Furthermore, automated mining of causal relations from news and text about socio-political events is a stepping stone in gaining critical insights, including analyzing the scale, frequency and trends across timelines of events, as well as anticipating future ones. The Shared Task 3, part of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE @ RANLP 2023), involved the task of Event Causality Identification with Causal News Corpus. We describe our approach to Subtask 1, dealing with causal event classification, a supervised binary classification problem to annotate given event sentences with whether they contained any cause-effect relations. To help achieve this task, a BERT based architecture - RoBERTa was implemented on four different datasets with the main difference being the inclusion and exclusion of both stopwords and various abbreiviations/acronyms present in the dataset. The results of this model are validated on the dataset provided by the organizers of this task. We achieved a rank of 8 with an F1 Score of 0.7475 on Dataset 1 (Removed Stop Words and Replaced Abbreviations).

1 Introduction

The ability to comprehend underlying cause-and-effect associations holds paramount significance across a multitude of disciplines. This knowledge facilitates informed decision making processes. Hence, the investigation of causal relationships assumes a pivotal position in the study of understanding the underlying mechanisms that govern the correlation between actions and their consequences.

With the uncertainty and complexity that characterizes various domains of inquiry, predictive capabilities empower individuals and organizations to navigate the intricate interplay of variables, facilitating better planning, risk management, and mitigation strategies.

The primary objective of this task is to develop a method to classify casual relations and in turn help identify the variable that acts as a reason for another. This will help in anticipating contingencies. The impact of casual analysis is not only seen in the extrapolation of likelihood of events under static conditions but also in events observed in continuously altering conditions, for example, changes induced by treatments or external interventions. [3]

This paper presents our use of the BERT based architecture RoBERTa, in **Subtask 1 Casual Event Classification of CASE 2023 Shared Task 3**.

2 Task Description

The **CASE 2023 Shared Task 3, Event Causality Identification with Causal News Corpus** [4] [5], focuses on the problem of detecting and extracting causal relations in protest event news. A causal relation is defined as a semantic relation between the cause and effect arguments, such that the occurrence of the latter is the subsequent result of the occurrence of the former. For an example to be considered a causal sentence, it has to include at least one cause-effect attribute pair where one argument provides the reason, explanation or justification for the situation described by the other.[1][7]

Subtask 1 - Causal Event Classification [6] is an event detection task which focuses on the automatic identification of the presence of causality in event sentences, i.e, a supervised classification task to detect causality in a given text. The aim is to develop a model to annotate event sentences with binary labels indicating whether

Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text, pages 34–37, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-089-2_004

they contained any cause-effect relations.

3 Dataset Description

The subtask used the second version of the Causal News Corpus (CNC), an annotated news corpus comprising 3,767 event sentences extracted from randomly sampled protest event news from varying sources. For the first subtask, the data consisted of event sentences annotated with binary labels indicating the absence or presence of causality within the span of each text, represented by 0 and 1 respectively.

The training and development dataset consisted of 3075 and 340 annotated text samples respectively. In the testing phase, the development dataset could be incorporated into the training of the model, and an additional test set of 352 un-annotated event sentences was provided.

4 Methodology

4.1 Data Pre-processing

This process entails cleaning of data through numerous methods like removing noise, handling missing data, normalising skewed data, etc. When it comes to text data, removing noise can include the removal of stop words, links, abbreviated words, punctuation and numbers or the conversion of numbers to words and abbreviations to their respective full forms or translating text from one language to another.

For Subtask 1, the training data consisted of 3075 samples with gold labels, the development data consisted of 340 samples with gold labels and the testing data consisted of 352 samples without gold labels.

The training, development and test data was pre-processed using the same methods. Data pre-processing involved removal of links using 'urllib', contractions handling using the 'contractions' library, replaced abbreviations and acronyms, removed punctuations, stop words and words with a length less than 3. Numbers were also converted to words using the library 'num2words'.

A dictionary was created with the most common abbreviations and acronyms present in the dataset. This dictionary was used to replace the respective abbreviations and acronyms from the text.

Libraries like 'nltk' and 'spacy' were also used for further pre-processing.
Four different datasets were created to compare results :

- Dataset 1 - Removed Stop Words and Replaced Abbreviations
- Dataset 2 - Included Stop Words and Replaced Abbreviations
- Dataset 3 - Removed Stop Words and didn't Replace Abbreviations
- Dataset 4 - Included Stop Words and didn't Replace Abbreviations

4.2 Model Building & Experimental Setup

RoBERTa [2] - a BERT based architecture was used for **Casual Event Classification**. The model was initialised using pre-trained RoBERTa weights "roberta-base" [1]. The text data was encoded using RoBERTa Tokenizer. The text was encoded to a length of 100 to generate embeddings that were passed through the model. The architecture included the RoBERTa Transformer layer followed by a Dropout layer, Flatten layer and two Dense Layers.

Since, the task calls for a binary classification, sigmoid activation function was used for the final dense layer. The model was compiled with **Adam** optimizer that had a learning rate of 0.00001, binary_crossentropy loss function and F1 score as metric. The model was run for 20 epochs.

5 Results

The final leaderboard for Subtask 1 was based on **Binary F1 Score**. The **baseline Binary F1 Score** was **0.8191**. Our submission using RoBERTa and Dataset 1 achieved a Binary F1 score of **0.7475**. The organisers also measured Precision, Recall, Accuracy and MCC(Matthews Correlation Coefficient).

It is clear from Table 1 that, Dataset 1 had the highest Recall value as compared to the other datasets. Dataset 2 and Dataset 4 had lower Recall values as compared to Dataset 1 but outperformed

[1] `https://huggingface.co/docs/transformers/model_doc/roberta`

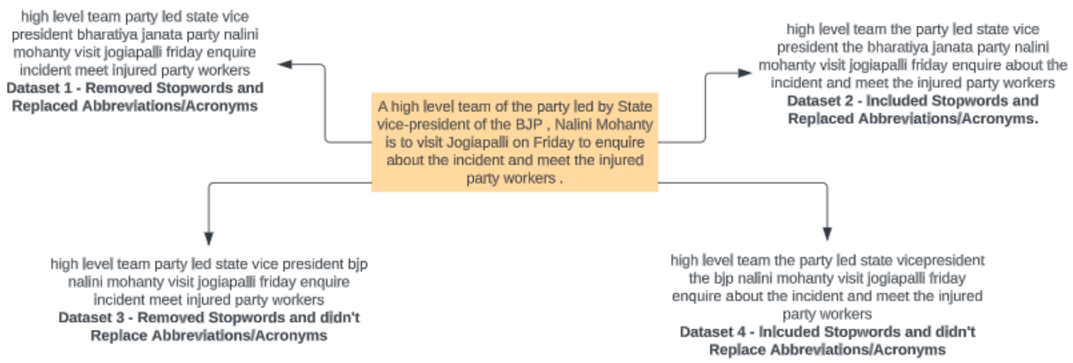

Figure 1: *Datasets Examples*

Models	RoBERTa				
Metrics	Recall	Precision	F1	Accuracy	MCC
Dataset 1	0.87283	0.65367	0.74752	0.71022	0.44829
Dataset 2	0.82081	0.71357	0.76344	0.75	0.50664
Dataset 3	0.73988	0.73142	0.73563	0.73863	0.47725
Dataset 4	0.80346	0.74731	0.77437	0.76988	0.54169

Table 1: *Test Results obtained for Subtask 1 using RoBERTa*

it in all the other metrics.

The only difference between Dataset 2 and 4 was the replacement of abbreviations/acronyms; Dataset 4 had outperformed Dataset 2 on F1 Score by a small margin. This shows that abbreviations/acronyms did not affect the model as massively as the presence or lack of stopwords. Transformers are able to understand context better than other deep learning models like RNN, LSTM, etc and in Casual Event Classification, the presence of the stopwords are generally the way to differentiate between cause and effect.

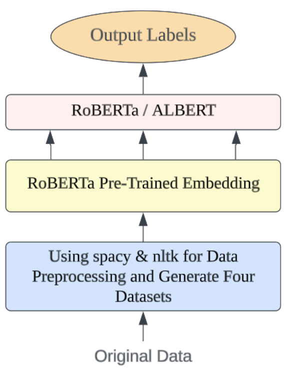

Figure 2: *Overall Methodology*

6 Conclusion

This paper addresses Subtask 1 - Casual Event Classification, which is a binary classification task, with the aim to annotate samples to indicate the presence of causality in a given text.

Our method includes basic data pre-processing techniques to generate four different datasets, which are used as inputs for the RoBERTa model and to the compare the results across five metrics - F1 Score, Recall, Precision, Accuracy and MCC(Matthews Correlation Coefficient).

Our system achieved a rank of eight based on Binary F1 Score i.e., 0.74752 achieved using Dataset 1. The best Binary F1 score value obtained was 0.77347 for Dataset 4, post competition.

Results may be improved by increasing the embedding size to a max length of 500 for RoBERTa. Hyperparameter tuning of the RoBERTa model should also be considered for better results.

References

[1] Biswanath Barik, Erwin Marsi, and Pinar Ozturk. Event causality extraction from natural science lit-

erature. *Research in Computing Science*, 117, 12 2016.

[2] 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. *CoRR*, abs/1907.11692, 2019.

[3] Judea Pearl. Causal inference. In Isabelle Guyon, Dominik Janzing, and Bernhard Schölkopf, editors, *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, volume 6 of *Proceedings of Machine Learning Research*, pages 39–58, Whistler, Canada, 12 Dec 2010. PMLR.

[4] Fiona Anting Tan, Hansi Hettiarachchi, Ali Hürriyetoğlu, Tommaso Caselli, Onur Uca, Farhana Ferdousi Liza, and Nelleke Oostdijk. Event causality identification with causal news corpus - shared task 3, CASE 2022. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, pages 195–208, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics.

[5] Fiona Anting Tan, Hansi Hettiarachchi, Ali Hürriyetoğlu, Nelleke Oostdijk, Onur Uca, Surendrabikram Thapa, and Farhana Ferdousi Liza. Event causality identification with causal news corpus - shared task 3, CASE 2023. In *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, Varna, Bulgaria (Hybrid), September 2023. Association for Computational Linguistics.

[6] Fiona Anting Tan, Ali Hürriyetoğlu, Tommaso Caselli, Nelleke Oostdijk, Tadashi Nomoto, Hansi Hettiarachchi, Iqra Ameer, Onur Uca, Farhana Ferdousi Liza, and Tiancheng Hu. The causal news corpus: Annotating causal relations in event sentences from news. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2298–2310, Marseille, France, June 2022. European Language Resources Association.

[7] Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. The penn discourse treebank 3.0 annotation manual. *Philadelphia, University of Pennsylvania*, 35:108, 2019.