

Instructions for *ACL Proceedings

Anonymous ACL submission

Abstract

This paper presents our system design for SemEval-2024 Task 4: Multilingual Detection of Persuasion Techniques in Memes in subtask 1. We tried to fine-tune models such as BERT, GPT2 and RoBERTa on dataset of subtask and we got the best results with GPT2. Our submit ranked 17th in subtask 1 among 33 teams.

1 Introduction

We refer to propaganda whenever information is purposefully shaped to foster a predetermined agenda. Memes consist of an image superimposed with text. The role of the image in a deceptive meme is either to reinforce/complement a technique in the text or to convey itself one or more persuasion techniques. In subtask 1, Given only the “textual content” of a meme, we must identify which of the 20 persuasion techniques, organized in a hierarchy, it uses.

For this problem, GPT2 was chosen as the base model after experiments on GPT2, BERT and RoBERTa. After that, the model was fine-tuned on the given data set and after doing prediction analysis and comparing them with true labels, the threshold of sensitivity was changed manually to get best results. In addition, we tried to fine-tune model on SemEval-2023 Task 3 dataset which is similar to the given dataset for this task. However, the results didn’t improve.

Regarding the noticeable change in scores just by changing the threshold of predicting labels, we realized the importance of prediction analysis and the easy tricks comes after actually understanding the behavior of model and it’s problems.

2 Background

2.1 Dataset Description

The dataset consists of 7000 samples for training and 500 samples for validation. each sample contains three fields:

- id: Each sample has an id which is used to identify the related image of that meme. However, images are not used in this subtask.
- text: this field is the textual content of the meme, as a single UTF-8 string. While the text is first extracted automatically from the meme, it has been post-processed to remove errors and formatted in such a way that each sentence is on a single row and blocks of text in different areas of the image are separated by a blank row.
- label: it is a list of valid technique names used in the text. There is 22 techniques in this dataset at all. However, only 20 of them are used for subtask 1. There is hierarchy of persuasion techniques and only leaf nodes are used in the labels.

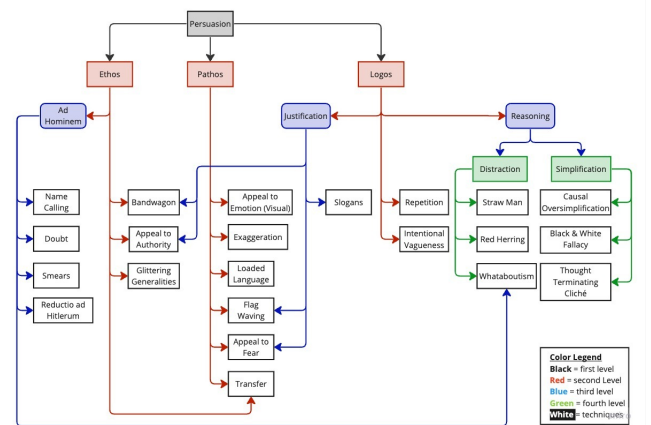


Figure 1: hierarchy of persuasion techniques

- link: the link of the meme in social networks. some samples may have no link and there is "null" keyword for them instead of the link.

2.2 Related Works

Before this SemEval2024 event task, similar datasets have been collected in previous articles.

```
{
  "id": "66730",
  "text": "WHEN THE POWER OF LOVE IS GREATER THAN THE
  LOVE OF POWER, THE WORLD WILL KNOW PEACE",
  "labels": [
    "Loaded Language",
    "Black-and-white Fallacy/Dictatorship",
    "Slogans"
  ],
  "link": "null"
},|
```

Figure 2: An example in the training set

The article "Detecting Propaganda Techniques in Memes" (Dimitar Dimitrov et al., 2021) establishes a novel multi-label, multimodal task of automatically detecting propaganda techniques in memes. creating a dataset of 950 annotated memes covering 22 propaganda techniques, the authors provide a crucial resource for training and evaluating future detection models. In addition, by creating a dataset of 950 annotated memes covering 22 propaganda techniques, the authors provide a crucial resource for training and evaluating future detection models.

The article "SemEval-2023 Task 3: Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup" (Jakub Piskorski et al., 2023) provides a publicly available dataset of annotated news articles, along with code and evaluation metrics. These resources serve as a valuable starting point for future research and development in multilingual news analysis tasks.

2.3 Task evaluation and ranking

The labels have a hierarchical taxonomy and the task description states that if the ancestor node of a technique is predicted, only a partial reward is given and this is a hierarchical multilabel classification problem. The metric used for evaluating submits is hierarchical f1 score (Svetlana Kiritchenko et al., 2006). Hierarchical f1 score is a way of adapting the F1 score metric to be used for classification tasks with hierarchical structures. These structures involve classes having parent-child relationships, forming a kind of tree-like organization. Regular F1 score is designed for flat classifications with no such hierarchy.

3 System overview

3.1 Model Architecture

We first used three models BERT, RoBERTa and GPT2 to be fine-tuned on training set. After eval-

uating model with previously explained metrics, the results on the GPT2 model were better than other models. Therefor, we continued the training process with this model. After fine tuning the model, we finally submitted the results with the GPT2 model.

3.2 Fine tuning on extra dataset

After training the model on the introduced dataset, we tried to use similar datasets from previous papers and SemEval events. The Semeval-2023 Task 3 dataset contained paragraphs from various news and articles tagged with 19 propaganda techniques.

After cleaning the data and removing uncommon tags between SemEval2023 task 3 dataset and our task dataset, we had 3445 new samples for training the model.

3.3 Adjusting the prediction threshold

The model generates continues values as possibility of classes instead of explicitly predicting presence or absence of a persuasion technique in input text. There is a Threshold for discretization of probability values. If probability is greater than the threshold, the output would be 1 otherwise, it would be 0. By adjusting this threshold for model, we were able to improve the results to some extent.

The first line of the file must be

```
\documentclass[11pt]{article}
```

To load the style file in the review version:

```
\usepackage[review]{acl}
```

For the final version, omit the review option:

```
\usepackage{acl}
```

To use Times Roman, put the following in the preamble:

```
\usepackage{times}
```

(Alternatives like txfonts or newtx are also acceptable.)

Please see the \LaTeX source of this document for comments on other packages that may be useful.

Set the title and author using `\title` and `\author`. Within the author list, format multiple authors using `\and` and `\And` and `\AND`; please see the \LaTeX source for examples.

By default, the box containing the title and author names is set to the minimum of 5 cm. If you need more space, include the following in the preamble:

Command	Output	Command	Output
<code>\`a</code>	ä	<code>\c c</code>	ç
<code>\^e</code>	ê	<code>\u g</code>	ğ
<code>\`i</code>	ì	<code>\l</code>	ł
<code>\.I</code>	İ	<code>\~n</code>	ñ
<code>\o</code>	ø	<code>\H o</code>	ö
<code>\'u</code>	ú	<code>\v r</code>	ř
<code>\aa</code>	å	<code>\ss</code>	ß

Table 1: Example commands for accented characters, to be used in, e.g., BibTeX entries.

`\setlength\titlebox{<dim>}`

where <dim> is replaced with a length. Do not set this length smaller than 5 cm.

4 Document Body

4.1 Footnotes

Footnotes are inserted with the `\footnote` command.¹

4.2 Tables and figures

See Table 1 for an example of a table and its caption. **Do not override the default caption sizes.**

4.3 Hyperlinks

Users of older versions of L^AT_EX may encounter the following error during compilation:

```
\pdfendlink ended up in different
nesting level than \pdfstartlink.
```

This happens when pdfL^AT_EX is used and a citation splits across a page boundary. The best way to fix this is to upgrade L^AT_EX to 2018-12-01 or later.

4.4 Citations

Table 2 shows the syntax supported by the style files. We encourage you to use the natbib styles. You can use the command `\citet` (cite in text) to get “author (year)” citations, like this citation to a paper by ?. You can use the command `\citep` (cite in parentheses) to get “(author, year)” citations (?). You can use the command `\citealp` (alternative cite without parentheses) to get “author, year” citations, which is useful for using citations within parentheses (e.g. ?).

A possessive citation can be made with the command `\citeposs`. This is not a standard natbib command, so it is generally not compatible with other style files.

¹This is a footnote.

4.5 References

The L^AT_EX and BibTeX style files provided roughly follow the American Psychological Association format. If your own bib file is named `custom.bib`, then placing the following before any appendices in your L^AT_EX file will generate the references section for you:

```
\bibliography{custom}
```

You can obtain the complete ACL Anthology as a BibTeX file from <https://aclweb.org/anthology/anthology.bib.gz>. To include both the Anthology and your own .bib file, use the following instead of the above.

```
\bibliography{anthology,custom}
```

Please see Section 5 for information on preparing BibTeX files.

4.6 Appendices

Use `\appendix` before any appendix section to switch the section numbering over to letters. See Appendix A for an example.

5 BibTeX Files

Unicode cannot be used in BibTeX entries, and some ways of typing special characters can disrupt BibTeX’s alphabetization. The recommended way of typing special characters is shown in Table 1.

Please ensure that BibTeX records contain DOIs or URLs when possible, and for all the ACL materials that you reference. Use the `doi` field for DOIs and the `url` field for URLs. If a BibTeX entry has a URL or DOI field, the paper title in the references section will appear as a hyperlink to the paper, using the `hyperref` L^AT_EX package.

Acknowledgements

This document has been adapted by Steven Bethard, Ryan Cotterell and Rui Yan from the instructions for earlier ACL and NAACL proceedings, including those for ACL 2019 by Douwe Kiela and Ivan Vulić, NAACL 2019 by Stephanie Lukin and Alla Roskovskaya, ACL 2018 by Shay Cohen, Kevin Gimpel, and Wei Lu, NAACL 2018 by Margaret Mitchell and Stephanie Lukin, BibTeX suggestions for (NA)ACL 2017/2018 from Jason Eisner, ACL 2017 by Dan Gildea and Min-Yen Kan, NAACL 2017 by Margaret Mitchell, ACL 2012 by Maggie Li and Michael White, ACL 2010 by Jing-Shin Chang and Philipp Koehn, ACL 2008 by Johanna D. Moore, Simone Teufel, James Allan, and

Output	natbib command	ACL only command
(?)	\citep	
?	\citealp	
?	\citete	
(?)	\citeyearpar	
?’s (?)		\citeposs

Table 2: Citation commands supported by the style file. The style is based on the natbib package and supports all natbib citation commands. It also supports commands defined in previous ACL style files for compatibility.

226	Sadaoki Furui, ACL 2005 by Hwee Tou Ng and	
227	Kemal Oflazer, ACL 2002 by Eugene Charniak and	
228	Dekang Lin, and earlier ACL and EACL formats	
	written by several people, including John Chen,	229
	Henry S. Thompson and Donald Walker. Addi-	230
	tional elements were taken from the formatting	231
	instructions of the <i>International Joint Conference</i>	232
	on <i>Artificial Intelligence</i> and the <i>Conference on</i>	233
	<i>Computer Vision and Pattern Recognition</i> .	234
	A Example Appendix	235
	This is an appendix.	236