Making your tweets more ironic

Kristian Djaković, Rene Kustura, Toma Puljak

University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
{author}@fer.hr

Abstract

This document provides the instructions on formatting the TAR system description paper in LaTeX. This is where you write the abstract (i.e., summary) of the work you carried out within the project. The abstract is a paragraph of text ranging between 70 and 150 words.

1. Introduction

Irony is complicated. In general, irony refers to a clash between expectation and outcome of an expression. (Kreuz, 2020). While this definition looks simple enough, detecting irony in text proves to be a difficult task, not only for models, but for people sometimes as well. This paper tackles the question: "What makes tweets ironic?" by analysing attention values outputed from a classification model.

The attention mechanism allows modeling of dependencies without regard to their distance in the input or output sequence (CITE FROM ATTENTION IS ALL YOU NEED). What's great about attention is that it adds a layer or explainability to the model. Attention values can be used to interpret the focus of a model when completing various tasks. In this paper, attention values are used to gather most ironic tokens from data samples. Most ironic tokens being those with the highest attention values.

For classification purposes, we used a pretrained RoBERTa Transformer model that was trained on the SemEval2018 Irony Detection dataset (REFERENCE TO THE TRANSFORMER AND DATASET). The model uses multi-head attention, thus we used the *attention rollout* method to aggregate attention values across heads for single tokens and connections between token pairs.

Once we gathered the most ironic tokens, we applied them to tweets that the model classified as *non ironic* and observed the change between the prediction certainty. We use this change in prediction certainty as a metric to evaluate how much a token, or token pair, impacts the ironic sentiment of a given tweet.

2. Background

In this section, we provide an overview of the classification model setup, attention aggregation methods and the datasets evaluate our analysis goal.

2.1. Model setup

For classification purposes, we used a pretrained RoBERTa Transformer model that had the best performance in irony detection. The model can be found here: `https://huggingface.co/cardiffnlp/twitter-roberta-base-irony`. RoBERTa is a retrained BERT model with improved performance. The details of the architecture can be found in (CITATION).

2.2. Attention aggregation

To aggregate attention values into a singular output we use the *attention rollout* method. Attention rollout is a recursive method that calculates the product of all attention weights in the attention graph. (CITE) This method will compute a singular output for a single token (node) that we feed into it and if we squeeze the resulting matrix without extraction, we get attention values for two connected tokens in the input sentence.

2.3. Dataset

The dataset used to train the RoBERTa model for irony detection is the SemEval2018 Irony Detection dataset (CITE DATASET) which is presplit into training and testing sets. Each tweet is annotated with a binary label of 0 or 1, representing a non-ironic and ironic label respectively.

3. Analysing the most ironic tokens

In this section, we will cover the process of finding the most ironic tokens from the used dataset and the methods used for evaluating the results.

3.1. Finding the tokens

The first step in finding the tokens is to extract all tweets that were annotated as ironic from the training and testing datasets, keeping them separate so we can test the results of the experiment separately. We then classified those tweets using the RoBERTa model to obtain a confidence score for each label. Sorting the tweets regarding to their ironic label confidence score we obtained a list of the most ironic tweets in the datasets, accoring to the model. A list of the most ironic tweets can be seen in (ADD TABLE 1).

Once we obtained the list of the most ironic tweets, we performed attention aggregation methods outlined in Section 2.2. and sorted the tokens by the attention score in a descending order. That gives us a list of singular and two connected tokens that the model deemed most ironic for each of the most ironic tweets. The obtained list can be seen in (ADD TABLE 2).

3.2. Adding the tokens to non ironic text

Using a similar method to extracting the most ironic tweets, we obtained a list of tweets with the highest non ironic confidence score. We used a subset of this list for analysing the impact of newly introduced ironic tokens. (DEFINE EXACTLY THE SUBSET)

Figure 1: This is the figure caption. Full sentences should be followed with a dot. The caption should be placed *below* the figure. Caption should be short; details should be explained in the text.

To add the ironic tokens into the non-ironic text we simply appended them to the end of the tweet. This was done for each token in the list applied to each tweet in the subset. We are aware that simply appending the token, or tokens, to the end of a tweet may generate something meaningless but ensuring that the newely formed tweet is semanticaly correct is out of the scope of this paper and we leave that for future work.

4. Results

4.1. Single token

We analysed (X) tweets paired with (Y) tokens. A sample of the top (N) and bottom (M) changes in non-ironic confidence scores can be seen in (ADD TABLE), with the full results available at: (LINK TO GITHUB).

The table show that the "most ironic" token was (X) and the "least ironic" was (Y). (ADD MORE COMMENTS ON THE RESULTS)

4.2. Two tokens

We analysed (X) tweets paired with (Y) pairs of tokens. A sample of the top (N) and bottom (M) changes in non-ironic confidence scores can be seen in (ADD TABLE), with the full results available at: (LINK TO GITHUB).

The table show that the "most ironic" pair of tokens was (X) and the "least ironic" was (Y). (ADD MORE COMMENTS ON THE RESULTS)

5. Conclusion

6. Figures and tables

6.1. Figures

Here is an example on how to include figures in the paper. Figures are included in LaTeX code immediately *after* the text in which these figures are referenced. Allow LaTeX to place the figure where it believes is best (usually on top of the page of at the position where you would not place the figure). Figures are referenced as follows: "Figure 1 shows ...". Use tilde (~) to prevent separation between the word "Figure" and its enumeration.

6.2. Tables

There are two types of tables: narrow tables that fit into one column and a wide table that spreads over both columns.

Table 1: This is the caption of the table. Table captions should be placed *above* the table.

Heading1	Heading2
One	First row text
Two	Second row text
Three	Third row text
	Fourth row text

6.2.1. Narrow tables

Table 1 is an example of a narrow table. Do not use vertical lines in tables – vertical tables have no effect and they make tables visually less attractive. We recommend using *booktabs* package for nicer tables.

6.3. Wide tables

Table 2 is an example of a wide table that spreads across both columns. The same can be done for wide figures that should spread across the whole width of the page.

7. Math expressions and formulas

Math expressions and formulas that appear within the sentence should be written inside the so-called *inline* math environment: $2 + 3$, $\sqrt{16}$, $h(x) = \mathbf{1}(\theta_1 x_1 + \theta_0 > 0)$. Larger expressions and formulas (e.g., equations) should be written in the so-called *displayed* math environment:

$$b_k^{(i)} = \begin{cases} 1 & \text{if } k = \text{argmin}_j \|\mathbf{x}^{(i)} - \mu_j\|, \\ 0 & \text{otherwise} \end{cases}$$

Math expressions which you reference in the text should be written inside the *equation* environment:

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} b_k^{(i)} \|\mathbf{x}^{(i)} - \mu_k\|^2 \tag{1}$$

Now you can reference equation (1). If the paragraph continues right after the formula

$$f(x) = x^2 + \varepsilon \tag{2}$$

like this one does, use the command *noindent* after the equation to remove the indentation of the row.

Multi-letter words in the math environment should be written inside the command *mathit*, otherwise LaTeX will insert spacing between the letters to denote the multiplication of values denoted by symbols. For example, compare $Consistent(h, \mathcal{D})$ and $Consistent(h, \mathcal{D})$.

If you need a math symbol, but you don't know the corresponding LaTeX command that generates it, try *Detexify*.[1]

8. Referencing literature

References to other publications should be written in brackets with the last name of the first author and the year of publication, e.g., (Chomsky, 1973). Multiple references are

[1] http://detexify.kirelabs.org/

Table 2: Wide-table caption

Heading1	Heading2	Heading3
A	A very long text, longer that the width of a single column	128
B	A very long text, longer that the width of a single column	3123
C	A very long text, longer that the width of a single column	−32

written in sequence, one after another, separated by semicolon and without whitespaces in between, e.g., (Chomsky, 1973; Chave, 1964; Feigl, 1958). References are typically written at the end of the sentence and necessarily before the sentence punctuation.

If the publication is authored by more than one author, only the name of the first author is written, after which abbreviation *et al.*, meaning *et alia*, i.e., and others is written as in (Johnson et al., 1976). If the publication is authored by only two authors, then the last names of both authors are written (Johnson and Howells, 1974).

If the name of the author is incorporated into the text of the sentence, it should not be in the brackets (only the year should be there). E.g., "Chomsky (1973) suggested that ...". The difference is whether you reference the publication or the author who wrote it.

The list of all literature references is given alphabetically at the end of the paper. The form of the reference depends on the type of the bibliographic unit: conference papers, (Chave, 1964), books (Butcher, 1981), journal articles (**?**), doctoral dissertations (Croft, 1978), and book chapters (Feigl, 1958).

All of this is automatically produced when using BibTeX. Insert all the BibTeX entries into the file `tar2022.bib`, and then reference them via their symbolic names.

9. Conclusion

Conclusion is the last enumerated section of the paper. It should not exceed half of a column and is typically split into 2–3 paragraphs. No new information should be presented in the conclusion; this section only summarizes and concludes the paper.

Acknowledgements

If suitable, you can include the *Acknowledgements* section before inserting the literature references in order to thank those who helped you in any way to deliver the paper, but are not co-authors of the paper.

References

Judith Butcher. 1981. *Copy-editing*. Cambridge University Press, 2nd edition.

K. E. Chave. 1964. Skeletal Durability and Preservation. In J. Imbrie and N. Newel, editors, *Approaches to paleoecology*, pages 377–87, New York. Wiley.

N. Chomsky. 1973. Conditions on Transformations. In S. R. Anderson and P. Kiparsky, editors, *A festschrift for Morris Halle*, New York. Holt, Rinehart & Winston.

W. B. Croft. 1978. *Organizing and searching large files of document descriptions*. Ph.D. thesis, Cambridge University.

F. Feigl, 1958. *Spot Tests in Organic Analysis*, chapter 6. Publisher publisher, 5th edition.

G. B. Johnson and W. W. Howells. 1974. Title title title title title title title title title. *Journal journal journal*.

G. B. Johnson, W. W. Howells, and A. N. Other. 1976. Title title title title title title title title title title. *Journal journal journal*.

Roger J. Kreuz. 2020. What makes something ironic? *The Conversation*, February.