

Humor Detection and Ranking

Bartol Freškura, Filip Gulan, Damir Kopljar

University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
{bartol.freskura, filip.gulan, damir.kopljar}@fer.hr

Abstract

In this paper, we consider the task of comparative humor ranking in two manner: detecting which tweet of two is more humorous and ranking the given tweets by how humorous they are in three classes. We opted for different approaches based on recent deep neural models in order to eschew manual feature engineering. In evaluation section we experimented with bi-directional LSTMs and CNNs, in combination and separately. For constructing feature vectors we also experimented with *GloVe* word embedding and character embedding. The system was tuned, trained and evaluated on SemEval-2017 Task 6 dataset for which it gives representative results.

1. Introduction

Understanding humor expressed in the text is a challenging natural language problem which has not yet been addressed extensively in the current AI research. Humor is often subjective and relies on vast knowledge base, which is sometimes hard to reason, even for humans. It is also important to say that what is humorous today might not be humorous tomorrow due to the fact that humor can be trend dependent.

2. Related Work

In scientific papers, this section usually (but not necessarily) briefly describes the related research.

3. Our model architecture

The paper should have a minimum of 3 and a maximum of 5 pages, plus an additional page for references.

3.1. Recurrent Neural Networks

3.2. Convolutional Neural Networks

4. Data

4.1. 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 or at the position where you would not place the figure). Figures are referenced as follows: “Figure ?? shows ...”. Use tilde (~) to prevent separation between the word “Figure” and its enumeration.

4.2. Tables

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

4.2.1. 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.

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

4.3. 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.

5. Experiments

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

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

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

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

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

$$f(x) = x^2 + \varepsilon \quad (2)$$

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

Table 2: Wide-table caption

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

$\text{Consistent}(h, \mathcal{D})$ and
 $\text{Consistent}(h, \mathcal{D})$.

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

6. 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

Hvala svima u studiju i režiji.

¹<http://detexify.kirelabs.org/>