Detecting Humor

IART - Checkpoint, Group 44

Afonso Caiado, up201806789
Diogo Nunes, up201808546
João Pinto, up201806667

Natural Language Processing: Detecting Humor

Humor poses interesting linguistic challenges to NLP, due to its emphasis on multiple word senses, cultural knowledge, and pragmatic competence. Humor appreciation is also a highly subjective phenomenon, and each person can have a different perception of a joke.

- **Is the intention of the text do be humorous? (0 or 1)**. The first task, is a binary one. We simply have to predict if the text is considered humorous.
- **[If it is intended to be humorous] How humorous do you find it? (1 to 5)**. In this step, by emulating previous humor detection tasks with ratings and classification scores, we have to predict how humorous the text actually is (if it is humorous).

Related Work & References

https://competitions.codalab.org/competitions/27446

https://github.com/04mayukh/YoungSheldon-at-SemEval-2021-Task-7-HaHackathon

---

https://en.wikipedia.org/wiki/Natural_language_processing

https://moodle.up.pt/pluginfile.php/213489/mod_resource/content/1/IART_Lecture7_NaturalLanguageProcessing_2020_21.pdf

https://moodle.up.pt/pluginfile.php/211544/mod_resource/content/0/IART_Lecture5a_Intro_MachineLearning.pdf

https://moodle.up.pt/pluginfile.php/213487/mod_resource/content/0/IART_Lecture5d_MachineLearning_Classification.pdf

Tools & Algorithms

Programming language: Python

Algorithms to implement: We are planning on implementing at least 3 of the following machine learning algorithms:

- Naïve Bayes
- Decision Trees
- Neural Networks
- K-NN
- SVM

Programming Tools: We expect to make use of:

- IPython
- Jupyter Notebook
- NLTK
- Stanza
- Scikit-learn for NLP

Work progress

As of this checkpoint, these are the main topics of work already carried out:

- The notebook structure has been added to the group repository;
- Reading datasets;
- We have started to manipulate and tokenize the datasets.