Twitter corpus about technology for sentiment and emotion analysis with automatic and manual annotation

Building Language Resources

Oihane Cantero, Julen Etxaniz and Jose Javier Saiz

1   Description

This corpus is a monolingual resource of unique English tweets about technology, with each tweet paired with an adversarial sentence. Both original and adversarial tweets are manually and automatically annotated according to sentiment polarity (positive, negative, neutral) and emotions (joy, anger, sadness, optimism). For automatic annotation, the "TweetEval" evaluation framework[1] was used, which covers both sentiment polarity and stance detection for Twitter-specific data, among other classification tasks. The objective is to compare and evaluate both types of annotation.

2   Application

The aim of this resource is to compare the annotations of tweets about technology and their adversarial sentences. It will also be interesting to compare these annotations when we do it by hand or automatically. The results will show whether the models are affected in their performance on classification tasks when faced with adversarial samples. If performance is significantly compromised by this phenomenon, improvements in the criteria used to generate the models may be proposed.

3   Methodology

First, we will take the 140 tweets from the Twitter API via the Tweepy package. We will exclude retweets, quotes and replies to get better tweets. We will also exclude tweets that have links to reduce spam tweets. Initially, a sample of the first 20 instances will be shared among the annotators. Only one label per tweet is allowed. We will annotate the tweets depending on their sentiment (positive, negative, neutral) and emotion (joy, anger, sadness, optimism).

These inital tweets will be used to calculate Inter Annotator Agreement and make modifications to the guidelines and scope if necessary. These tweets will also be labelled using automatic models. Two HuggingFace models that are finetuned for each task will be used for evaluation. We will compare the results with our annotations to have an initial idea of the accuracy of our models.

Thereafter each annotator will be given a different set of 40 tweets for annotation. The same methodology will be followed, the manual annotations will be used to evaluate the accuracy of the models.

---

[1] Barbieri, Francesco & Camacho-Collados, José & Espinosa-Anke, Luis & Neves, Leonardo. (2020). TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. 1644-1650. 10.18653/v1/2020.findings-emnlp.148.

Then, we will create some adversarial tweets automatically to measure the impact in the accuracy of the models. The aim of these adversarial ecamples is to confuse the models while maintaining the manual label. Therefore, if we generate the adversarial examples correctly, there is no need to annotate them. For example, substituting words with their synonyms is a good option. We will add these adversarial examples to our final resource.

Finally, we will compare the results and extract some conclusions. We will visualize the percentage of tweeets from each class in the manual and automatic annotations. We will also visualize the most common words of each class in wordclouds.

4 Inter Annotator Agreement

We will calculate the inter annotator agreement between us three and the automatic results for the 20 shared tweets. We might also calculate the agreement between each of us and the automatic results for the 40 tweets we will annotate alone.

We will compute the Kappa scores to see if we agree in the annotations of both sentiment and emotions, and if we agree with the automatic annotations.

5 Data Statement

We will make a short version of the data statement, a summary of the long version that will be presented in the final paper. It covers the following topics:

A. CURATION RATIONALE

The corpus is a monolingual test set intended to evaluate text classification models for their ability to estimate sentiment and emotion. This is based on their performance against regular documents and adversarial phenomena. It contains a sample of 140 tweets and their adversarial pairs.

B. LANGUAGE VARIETY

Data extraction was made using the BCP 47 identifier for English (ISO 639-1 en). Since this tag does not address any specific variety of English and Twitter is used worldwide, we can assume that the extracted sample includes a number of varieties of English.

C. SPEAKER DEMOGRAPHIC

Because we can not infer any precise information about the authors from the text alone, we must rely on general Twitter demographics, which include a majority of speakers which are male (70.4%) and ranges between 18 and 49 years old (78.7%)[2]. Tweet metadata sometimes provides geolocation information, which for our sample suggests that a significant number of them originate in the United States. Beyond that, our demographic sample is only restricted by language and topic of interest.

D. ANNOTATOR DEMOGRAPHIC

The annotation was made by three Spanish MSc students with training in linguistics and informatics.

E. SPEECH SITUATION

---

[2] Iqbar, Mansoor. (2022). Twitter Revenue and Usage Statistics.
https://www.businessofapps.com/data/twitter-statistics/

Tweets are characterised by being short, spontaneous, time-sensitive texts. This means that the content of tweets is only relevant to the time in which they were written.

F. TEXT CHARACTERISTICS

The sample was extracted between 1st February and 3rd February 2022. It includes tweets made before 1st February. The Twitter context "Interests and Hobbies Vertical: Technology'' (domain and entity ID: 65.848920371311001600) was used as the thematic parameter. This context includes "top-level interests and hobbies groupings" about "technology and computing", as indicated in the description of the metadata context.

6 License

The resource and code are publicly available at https://github.com/juletx/twitter-sentiment. The code is licensed under the MIT open source license.

All non-code materials provided are made available under the terms of the CC BY 4.0 license (Creative Commons Attribution 4.0 International license).