Author: Jacob Fuehne

Date: 10/11/2020

Research Paper Summary

Automatically Identifying Complaints in Social Media, *Association for Computational Linguistics*, 2019

1. Problem definition and the main ideas of the research

Complaining is a basic speech act used to express a negative mismatch between reality and expectations towards a state of affairs, product, organization or event. While computational approaches towards detecting negative sentiments are plentiful, little to no research has been done in specifically analyzing text for this mismatch of expectations. This paper seeks to serve as the first extensive analysis of complaints in computational linguistics.

2. Significance of research study (Importance and Challenges of research problem)

The ability to detect complaints is critical for companies and organizations to improve customer experience and address issues, as well as future researchers in psychology and linguistics for analyzing and researching complaints. Previously, research has focused on associating text as positive or negative, or on assigning more specific emotions to a text. However, the paper points out that while complaints are certainly correlated with negative emotions, not every negative sentence is a complaint and not every complaint expresses negative emotions.

3. Main research questions and assumptions

This paper seeks to address automatically identifying complaints independently of the overall sentiment. In their research, they create and make publicly available an expert annotated corpus on complaints across several domains. The assumption here is that by having experts annotate a corpus, they can demonstrate methods for a predictive model to identify complaints in text. They seek to encourage future research in complaints by linguistics and psychologists by making their dataset publicly available.

4. Research Methodology

Previously, there were no publicly available datasets with annotated complaints as defined in linguistics. The authors resolve this by creating a new dataset using the Twitter API and selecting several domains, so that they could analyze transferability across domains. They created a binary annotation task for identifying if a tweet contains a complaint or not. A variety of generic and proven models and features are implemented to perform an extensive analysis of what is best suited for the task. Sentiment models such as MPQA, NRC, V&B, VADER, and the Stanford model are compared, with features such as unigrams, LIWC, word2vec, and POS tags. Also implemented are the neural models MLP and LSTM. Lastly, the authors implement what they consider complaint specific features such as requests, intensifiers, downgraders, temporal references, and pronoun types. The complaint specific features are inspired by linguistic aspects of the complaints defined in previous research.

5. Experiments

The researchers used their various models to perform an extensive evaluation of which model performs best, and which linguistic features are important in classifying complaints. In their results, they found that using bag-of-words with all features achieved the best predictive performance, outperforming even the neural models. As pointed out by the authors, one possibility for this was the small training data size. In the experiments for cross domain transferability, it was shown that in all but one case, adding out of domain data helps performance. The domain in question was apparel, which was pointed out to be qualitatively very different from other fields, as the types of issues in the complaints were different.

6. Discussion

6.1 Important aspects

- Publication of annotated corpus

As the first extensive study in methods for identifying complaints, it is very important and beneficial to future researchers that the authors made their annotated data public.

- Testing various linguistic features

Performing a comprehensive evaluation of accuracy across generic and proven models, while testing old features and complaint specific researchers provides a good foundation for future researchers to build on. It shows which features are most important, and allows analysis into why certain types of models may perform better than others.

6.2 Limitations of the paper

- VERY small data set

  Going through the corpus on their github, not only is it not labelled clearly, but there are only a little over 3000 tweets across all domains. While certainly still beneficial to the community, the small sample size means it has limited usefulness in a number of SOTA models.

- Better explanation of EasyAdapt vs Pooling needed and how it affects results

Due to the noticeable, but inconsistent improvements in cross domain transferability with EasyAdapt vs Pooling, the authors should have investigated/discussed why this difference.

6.3 Questions for presenter

- Can you explain how EasyAdapt is different from Pooling? They both combine in and out of domain data, but EasyAdapt shows small to significant improvements in most cases.
- Why might Pooling be better than EasyAdapt in some cases?