Author: Jacob Fuehne

Date: 9/20/2020

Research Paper Summary

An Analysis of Annotated Corpora for Emotion Classification in Text, *Association for*

Computational Linguistics, 2018

1. Problem definition and the main ideas of the research

Due to the differences in annotation schemata, the domain, and file formats, supervised models for the classification of emotion often use a limited set of resources. This paper seeks to contribute towards solving this problem by conducting a survey of datasets and aggregating them into a common file format with common annotation schema. The paper seeks to experiment with cross corpus classification and gain insight on the differences of models inferred from the data.

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

Emotion detection and classification in text is used in dialog systems, intelligent agents, clinical diagnoses of mental disorders, social media mining and more. A challenge often faced when doing emotion detection and classification is that the set of domains and differences in text are large. Emotions in different mediums are expressed in different ways, as journalists will tend to be more objective, microbloggers will focus on brevity, and authors of tales will likely be more expressive. Transfer across emotion recognition models is challenging and creating original resources from scratch is costly. The paper seeks to contribute to solving these challenges by comparing datasets, exploring how they complement each other, and mapping them to a common format.

3. Main research questions and assumptions

This research seeks to answer the question as to whether one can reuse and/or combine datasets from different domains and different formats for new applications. The assumption here is that by comparing the datasets and mapping labels to more broad labels, one can create a unified dataset using many smaller datasets of different domains and formats. A larger unified dataset is assumed to be able to provide better results than the individual results.

4. Research Methodology

To create an annotated and unified corpus, the authors of the paper use a number of datasets and combine them. A common label set is defined to include joy, anger, sadness, disgust, fear, trust, surprise, love, confusion, and noemo. As the datasets included have different formats of annotation, different emotions are mapped/merged together into labels that are in the common label set. Instead of implement neural methods, the paper opts to use maximum entropy classifiers in scikit-learn with bag-of-words features for the sake of simplicity and reproducibility.

5. Experiments

In the paper, they perform experiments in different settings. They train and test on one corpus in one setting, and in another, they train on an entire corpus, and test on the data of an entire different corpus (for all pairs, including the aggregated corpus). In the findings, it is found that some datasets and domains are more difficult to be modeled than others. In the All vs one tests, it is found that not only is the annotated unified corpus not effective in almost all cases as the authors had hoped, it almost never had the best F1 score.

6. Discussion

6.1 Important aspects

- Attempting to create alternative methods to creating original resources

  As mentioned in the paper, a big hurdle often faced by researchers is finding the right dataset/corpus for their application needs. By creating a new unified dataset and describing methods for combining annotated datasets, one can make emotion detection more accessible.

6.2 Limitations of the paper

- Did not properly address the failures of their results

  While the paper does acknowledge that the annotated unified corpus performed worse than expected, it does not seem to elaborate in depth about possible reasons why or possible ways to fix it. It is understandable that researchers will want to embellish and talk positively of their own work, but in order to truly make a failed experiment meaningful, one must acknowledge the mistakes and allow others to learn from them.

6.3 Questions for presenter

- Do you think there is any simple modification that could have been made to make the annotated unified corpus have better F1 scores?
- Do you think that it is worth pursuing a method to better unify annotated corpora, or do you think it is best to simply use specialty, domain specific corpora?