

Author: Jacob Fuehne

Date: 10/4/2020

Research Paper Summary

SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text, *Association for Computational Linguistics*, 2019

1. Problem definition and the main ideas of the research

Classifying emotion from ambiguous text is difficult without facial expressions or voice modulations. However, leveraging context can help resolve some of this ambiguity. The goal of SemEval-2019 Task 3 was, given an utterance in a textual dialogue, to use context to infer the underlying emotion from four emotion classes, Happy, Sad, Angry, and Others. This paper is a summary of the methods, resources, and tools used by the 311 teams that participated in SemEval-2019 Task 3.

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

Current machines and digital assistants typically focus on specific tasks, but for a deeper relationship with technology, an intelligence quotient (IQ) is not enough. Research in understanding emotion from context is important to create machines with higher emotional quotients (EQ) that can understand and express emotions. Deciphering emotion from text without facial expressions or voice modulations is difficult, and the remaining resource, context, can completely change the emotion for an expression.

3. Main research questions and assumptions

This paper serves to summarize SemEval-2019 Task 3, with the methods used by the participating teams. They also seek to summarize methods of filtering, processing, and labelling

data, and evaluation metrics. The paper assumes that by gathering together 311 teams, the authors can comment on the preferred methodologies and performance of teams.

4. Research Methodology

The authors describe how they created the test dataset using a random sampling of over 2 million dialogues collected between a user and an agent over one year. The data was filtered to exclude offensive language, multilingual text, and personal information. The authors of the paper noted the models used by the most successful teams, as well as the features and resources. In their paper, they note that most of those teams used BiLSTMs/LSTMs for models, BERT, ELMo, ULMFit for transfer learning, and GloVe for embeddings.

5. Experiments

Since this paper's purpose was to summarize SemEval-2019 Task 3, the authors focus on the description of their data collection techniques, evaluation metrics, and analyzing the performance and preferences of each of the top performing team submissions, including summaries on how their submission works. In the analysis of the different teams, the difference in performance between the top ranking teams is very slim.

The authors also perform a data analysis of the dialogue that was collected. In this analysis, they found that the emotion class happiness is often expressed in fewer words than the sad or angry emotion classes. Also provided was a spiral graph denoting the frequencies of the most popular unigrams in each emotion class, but only words that do not appear in the top 500 most frequent list of "other" emotion class are considered, so as to avoid neutral words from showing up in the spiral graphs. Lastly, the researchers analyzed the emoticons that appeared in the textual dialogues and found that 21% of textual dialogues contain emoticons. One surprising

result from these analyses is in the spiral graph of frequent words, where “loves” is listed as a frequent word in Sad, but “loved” in Happy.

6. Discussion

6.1 Important aspects

- Developing new ways to utilize context in emotion detection

SemEval2019 Task 3 helps to draw interest in using context for emotion detection and fuels innovation in the area by having a competition.

6.2 Limitations of the paper

- Conclusions are not drawn on about why some embeddings are popular

The authors take note of and graph the usage of different techniques utilized by the top performing teams, but they do not make any conclusions or hypotheses as to why such preferences might exist. Instead, they simply acknowledge that these trends exist.

6.3 Questions for presenter

- Did anything surprise you in regards to the words listed or not listed on the most frequent unigram graphs?