Author: Jacob Fuehne

Date: 10/4/2020

Research Paper Summary

Latent Emotion Memory for Multi-Label Emotion Classification, *AAAI Conference on Artificial Intelligence*, 2020

1. Problem definition and the main ideas of the research

Emotion classification is an important task in natural language processing. Identifying multiple emotions in a sentence is typically modelled as a multi-label classification task in existing methods. This paper seeks to find another method to identify multiple emotions in a sentence by leveraging the latent emotion distribution to better classify emotions without requiring external knowledge.

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

This paper mentions that inferring emotions from text is the initial step for future applications such as emotional chatbots, stock market prediction, policy studies, etc. A current challenge in classifying emotions in text is that commonly, there will be multiple emotions expressed in a text. Current methods have two issues mentioned. Firstly, they incorrectly assume that each emotion occurs with equal prior probability, failing to account for prior distributions. And secondly, they don't effectively capture the context information relating to the corresponding emotion.

3. Main research questions and assumptions

This research paper seeks to resolve the issues mentioned by proposing a Latent Emotion Memory network (LEM) for multi-label emotion classification. The assumption here is that by having a latent emotion module, the model can learn the emotion distribution by reconstructing the input via a variational autoencoder. Then by using a memory module, the module can capture context relating to an emotion. Then, by concatenating these two modules into a bi-directional Gated Recurrent Unit (BiGRU), it is assumed that the resulting model will resolve the issues posed previously.

4. Research Methodology

The researchers use the SemEval 2018 task 1C English dataset and the Ren-CECps Chinese dataset to conduct their experiments. The data is used for the latent emotion module, which takes the emotional Bag of Words as input and filters out stopwords, keeping only words that exist in the sentiment dictionary. The latent emotion module is modelled as a set of latent variables, learned using a variational autoencoder. The other component, the emotion memory module, leverages multiple-hop memory to mine rich context information for the corresponding emotion. The latent emotion distribution is concatenated to the feature representation from the memory module and fed into a BiGRU to give a prediction.

5. Experiments

To test the proposed LEM, it is tested with 5 different performance metrics measured: Hamming Loss, Ranking Loss, Micro F1, Macro F1, and Average Precision. The model is compared against 5 popular text classifiers (TextCNN, BiLSTM, RCNN, attLSTM, FastText) using Binary Relevance, the same 5 text classifiers using Joint Binary, and several Multi-Label Emotion Classifiers (ECC, MLLOC, ML-KNN, TMC, EDL, SGM, RERc, DATN). When compared against these classifiers, the LEM outperformed all other classifiers (except RERc in Micro F1 scores on the chinese dataset). The results show that models that are able to integrate relations between emotions perform better, as demonstrated by the performance of the proposed LEM and the runner-up performance of SGM and RERc compared to other methods.

6. Discussion

6.1 Important aspects

- Developing new ways to utilize context in emotion detection

  The paper mentioned problems with previous methods and how those methods were unable to use prior emotional distribution or correlating context in predicting multiple emotions within a text. With this new method, we are a little bit closer to an accurate model for classification of emotion in text.

- Being thorough in analysis of LEM vs RERc and other models

  The paper showed that in the micro F1 scores on the Chinese dataset, the RERc model outperformed the LEM. The paper authors did a very thorough analysis however, and showed that LEM still maintains the best KL of models tested for the distance between the learned distribution and the true distribution.

6.2 Limitations of the paper

- Number of datasets and number of emotions measured

The paper used 2 datasets, SemEval2018 and REN-CECps. While these showed that LEM was better than almost all methods in all categories, the results were different, with different gaps of performance between the two datasets. If given other datasets, with more and less emotion categories than those given, one could expect to see different results than those published (either further supporting the superiority of LEM, or pointing out flaws).

6.3 Questions for presenter

- Do you have any idea why the RERc model outperformed LEM on the REN-CECps dataset?