An Active Learning Approach for Emotion Detection

 Colton Aarts [1]

Text classification is the process of determining if a piece of text belongs to one class or another. This process is used in a number of different places. Companies can use text classification to adjust marketing strategies based on the emotions that are displayed in customer feedback. Or psychologists can use classification techniques to determine if patients are displaying symptoms of depression in online text post. One of the main problems that text classification encounters is the cost of creating high quality labels. There are several different problems that cause labelling to be so expensive. These include needing a large number of labelled data, requiring more than one person to label each data point, and requiring the people that are doing the labelling to be experts in their respective fields. All of these contribute to the cost of creating the labels. The most effective way to lower the cost of labelling while preserving the quality of the labels is to reduce the number of labeled data points that are needed to create an effective classifier. One of the techniques to do this is known as active learning. Active learning is the process of purposefully selecting a subset of the available data to be labeled based on the information that the classifier already has. This leads to a process of repeatedly suppling the experts with small batches of data to be labeled. After a hopefully small number of batches the classifier will have achieved the desired performance while requiring fewer labeled examples than a classifier created from a random selection of the data. I will present a current research project where we are working to create an active learning process for minimizing the number of label data points for the creation of a neural network classifier for emotion detection in Tweets.

---

[1]UNBC Computer Science, University of Northern British Columbia, Prince George, B.C., V2N4Z9, Canada (Aarts@unbc.ca).