LING 573 Emotion Classification Project Report

Xena Grant, Jesse R. Mena
Department of Linguistics
University of Washington
xg123@uw.edu, jesserme@uw.edu

Abstract

Stub

1 Introduction

Stub

2 Task Description

Deciphering emotion from social media posts proves difficult with the absence of facial expressions or vocal cues. EmoEvalEs (CodaLab Competition) aims to classify emotion in a set of Spanish tweets. Emotion classes include anger, disgust, fear, joy, sadness, surprise, and other – a category containing neutral or emotionless sentiments. Our primary task is to discern the classification method that yields the highest test accuracy. For our adaption task, wehope to apply the most successful method to another language and to increase accuracy through data augmentation. The dataset contains a collection of tweets during the month of April 2019 that encompasses a variety of topics ranging from entertainment to catastrophes. Anyhashtags in the dataset were

replaced with “HASHTAG” as to not influence the classifier. The dataset is split into development, training,and test partitions. The evaluation process consists of ranking weighted- F1 averages in a multi-class evaluation.

3 System Overview

The following in Figure 1 is an overview of the architecture for our task:

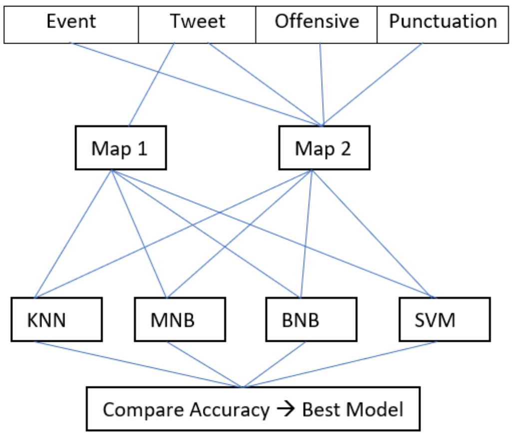

Figure 1

4 Approach

As shown in figure 1, we compared training accuracy across four different classifiers: Multinomial Naïve Bayes, Bernoulli Naïve Bayes, K Nearest Neighbors, and Support Vector Machine. The Naïve Bayes classifiers calculate conditional probability based on either term frequency or binary term presence within each tweet. The K Nearest Neighbors classifier calculates the proximity of a datapoint to others around it, then chooses the class more prevalent among the neighboring datapoints. SVM classifiers find the maximum marginal hyperplane separating the datapoints and then classifies the resulting groupings. We created two models for comparison. The first contains

extra features alongside the content of the tweet: the presence of exclamation marks, the event concerning which the tweet was posted, and whether or not the tweet was judged as offensive. The second model simply contains the cleaned content of the tweet.
For each trial, we created models using the sklearn implementations on the train data and test on the test dataset. Accuracies for each model as well as accuracy for an ensemble combination of the two are included in the chart below (Table 1).

5 Results

Method	Version	Accuracy
SVM	Text only	.577
	All features	.592
	Ensemble	.610
KNN	Text only	
	All features	
	Ensemble	
Multinomial	Text only	
	All features	
	Ensemble	
Bernoulli	Text only	
	All features	
	Ensemble	

Table 1

Performing classification on the content of the tweets alone yielded higher results than when working with the extra features. However, when evaluating both models in an ensemble format, we produced a higher accuracy using SVM.

6 Discussion

Stub

7 Conclusion

Stub

8 References

Stub