Emotion Mining from Text

Group no. 63

Megha Tyagi
Neha Jhamb
Nitish Srivastava
Raveena Gupta

Introduction

Motivation : Machines have still not become capable of detecting emotions clearly. Emotion mining can be used for finding how a customer feels about a product, making response of chatbot much more expressive etc.

Problem Statement : 8 types of basic emotions - joy, trust, fear, surprise, sadness , disgust, anger, anticipate. We aim to recognize only 5 emotions - anger, fear, joy, surprise and sadness in text.

Dataset and Evaluation

Total Sentences : 50,000

10,000 sentences for each emotion.

Training set : 60%

Validation set : 20%

Test set : 20%

Feature Extraction:

1. Doc2Vec
2. BOW_ADJ

Evaluation metric: Accuracies of the classifiers on whole test dataset.

Analysis and Progress

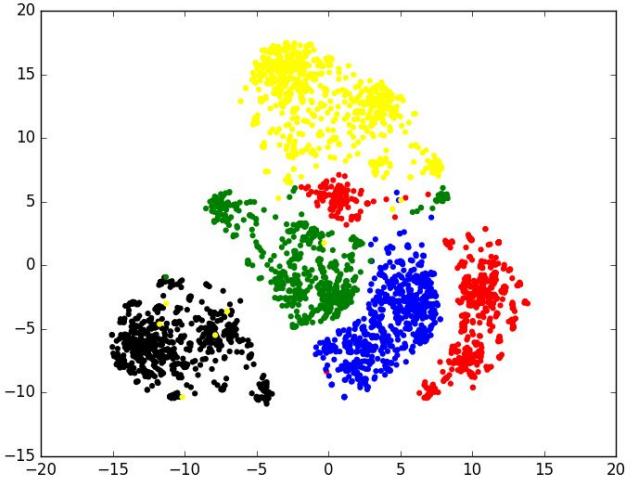
Progress

Implemented 3 models :
Logistic Regression
Linear Discriminant Analysis
Support Vector Machine

Analysis

Learning Method : Supervised Learning
Hyperparameter tuning : Grid search.
Two datasets analysed : We Feel Fine and Affective Text.
Final dataset : We Feel Fine

Scatter plot of We Feel Fine Dataset (T-SNE)



Results

Observations

Model	Best Hyperparameters	Accuracy
Logistic Regression	Penalty = 'l1', C=0.5, max_iter = 100	0.903
Linear Discriminant Analysis	Solver = 'svd', n_components = 20	0.889
Support Vector Machine	Penalty= 'l2', C=0.5, Loss='hinge'	0.895
Linear Discriminant Analysis - Doc2vec	Min_count = 1, size = 1000	0.610

Future Work

Use one other dataset also - ‘Emotion in Text data set’ by CrowdFlower.

Analyse the models on negative sentences such as “not happy”, etc.

Apply features reduction algorithm to reduce the redundant features.

Future Work

Individual task allocation for final evaluation.

Team Member	Task
Megha Tyagi	Extracting features using CNN.
Neha Jhamb	Training SVM, LDA and LR on the above extracted features.
Nitish Srivastava	Training Deep Neural Networks on the above extracted features.
Raveena Gupta	Lemmatization during preprocessing then Training on LR, LDA and SVM.