CS583 Project Report: Sentence Level Sentiment Analysis

Team Members: Amruta Nanavaty (UIN: 665473328)

Sabita Acharya (UIN: 676636765)

INTRODUCTION:

The primary goal of this project is to classify the expressed opinion in a given sentence into positive, negative or neutral class. This was achieved by designing features for the different supervised classification techniques and building those classifier models on given training dataset. The performances of those models were evaluated by testing them on the provided test dataset and by calculating the precision, recall and F-score for respective target classes. We experimented with different features like unigrams, bigrams *etc.* along with different classification techniques like Support Vector Machines and NaiveBayes Multinomial in order to accomplish the task. We have also tried to test the impact of varying size of training dataset on the performance of the classifier model. Major assumption for the project was that we have ignored sentences comprising of mixed opinions.

APPROACH:

1. Manually tagged the training dataset.
2. Divided the dataset into training dataset of size 3700 (1500 positive, 700 negative and 1500 neutral) and testing dataset of size 900 (300 positive, 300 negative and 300 neutral).
3. Determined and formed list of corpus dependent stop-words. [Majority of the dataset comprises of hair care related sentences.]
4. Determined and generated lexicon of opinion words specific to hair care domain.[1]
5. Extracted following features: [NOTE: For each feature term frequency as well as tf-idf were computed]:
    a. Unigrams [comprises of the unigrams extracted from the corpus. In addition to that it also includes the specific lexicon generated for the hair-care domain]
    b. Bigrams [Extracted descriptive 2-word phrases based on POS tags][2]
    c. Unigrams + Bigrams [Combined features obtained from a and b]
    d. Opinion lexicon provided on Professor's website.

While generating these features we have omitted those unigrams and bigrams occurring in the stop list as they do not contribute towards the sentiment analysis.

6. Feature Representation: We conducted experiments on three different representations of above features.

   a. Represented each sentence in training set as a feature vector of the term frequency of individual unigrams (bigrams, unigrams+bigrams) as features. In this case the feature vector size will be equal to the size of the unigram list.

   b. Represented each sentence in training set as a feature vector of the tf-idf of individual unigrams (bigrams, unigrams+bigrams) as different features in the feature vector. Again feature vector size will be equal to length of unigram list.

   c. Represented each sentence in training set as a feature vector of 3 features namely S^{+} , S^{-} and $S^{=}$. Here '+' indicates positive class, '– 'indicates negative class and '=' indicated neutral class. In this method, each term was given a score with respect to positive, negative and neutral class. When a training sentence is encountered each feature represents the summation of the scores with respect to their class.

7. Three supervised classification techniques were tested and evaluated :

   a. Support Vector Machines

   b. Naïve Bayes

   c. Naïve Bayes Multinomial

OBSERVATION:

Usage of combined unigrams and bigrams features with removal of corpus specific stop-words on Support Vector Machine classification algorithm gave best results in our generated unknown test data as well as on the test set provided by the TA. The unigram list as mentioned above also included the hair-care specific opinion words. The bigram phrases were produced from specific grammar rules [e.g. adjective (JJ) followed by noun (NN)].

NOTE: On TA provided dataset, we obtained best precision, recall and f-score for positive class but average results for negative class. We also had to retrain the model since we observed that reduction in the size of training set gave better results for TA provided test set. We also observed that the logical reasoning behind labeling training set and test set were different which could be the possible reason for lower f-score for negative class. Second possible reason which could have affected our results was that the training data set majorly comprised of hair-care related sentences or normal (no domain specific) sentences whereas the test set comprised of hair-care, camera and normal sentences.

RESULTS:

Tools used: Eclipse and WEKA.

[A] Classifier Name: Support Vector Machine (SVM)

POSITIVE CLASS:

Precision = 0.772

Recall = 0.795

F-score = 0.784

NEGATIVE CLASS:

Precision = 0.75

Recall = 0.14

F-score = 0.235

NEUTRAL CLASS:

Precision = 0.944

Recall = 0.992

F-score = 0.968

[B] Classifier Name: NaiveBayes Multinomial

POSITIVE CLASS:

Precision = 0.351

Recall = 0.697

F-score = 0.467

NEGATIVE CLASS:

Precision = 0.224

Recall = 0.605

F-score = 0.327

NEUTRAL CLASS:

Precision = 0.918

Recall = 0.616

F-score = 0.737

REFERENCES:

1. Amy Pogue (Year 2008). Opinion words for hair-care domain. Retrieved from http://www.words-to-use.com/words/hair-care/
2. Mullen. Introduction to Sentiment Analysis. Retrieved from http://www.lct-master.org/files/MullenSentimentCourseSlides.pdf