I started out the assignment by reading some papers on sentiment analysis. I figured out that for features I would need some words representing the opinion of the author.

I started out with finding some positive and negative words online from a general website. I extracted these terms and ran a J48 decision tree classifier on the arff file. I reached 38% accuracy. I used Weka’s attribute feature selection and then I re ran the classifier and I reached a 42% accuracy. I then tried different classifiers and I got the max accuracy of 45% by using SMO. Then I came across a specialized lexicon from Janice Wiebe and Rada Mihalcea. They have a subjectivity lexicon that I downloaded from <http://www.cs.pitt.edu/mpqa/> . And then I re ran my experiments and astonishingly gave me a boost of 5% and I got around 50% accuracy. At the same time accuracy for binary class reached to 74% and for reviewer’s classification to 93%. I was also using some document specific features such as document length, number of positive words, no of negative words, difference between positive words, no of punctuations.etc. For most of the features I was using their tf-idf values.

I then used Stanford parser to get the parts of speech and typed dependencies. I used the count of nouns, verbs, adjectives, adverbs and determiners as features. This raised my accuracy from 50% to 53%. I then tried adding some more positive and negative terms as features but nothing improved any more accuracy.

I compared the features selected by Weka for positive/negative and multi star classifications. There were some features that were only used in one of them and not another. I tried combining all the selected features and that resulted in increasing the accuracy from 54% to 58% for multi star rating and from 74% to 79.8% for positive/negative classification.

I was not able to use typed dependency from Stanford parser. I would have tried it to see if I could have improved the accuracy. I also tried using reviewer’s classification first to increase the accuracy but it didnt result in any increase in the accuracy.

1. Cross validation Accuracy for each experiment?
2. Classifiers used and Why?

J48, SVM, Naïve Bayes.

1. Which were the fastest? Most Accurate? Easiest to use?

Naive Bayesian was the fastest.

SVM SMO was the most accurate one.

Naïve Bayes and J48 were easier to use as they had less parameters.

1. Which one do you prefer and Why?

I preferred SVM-SMO because it was giving better results with respect to Accuracy and classified more instances correctly.

1. Which classification task was easiest and hardest and why?

The classification of reviewers was easiest using the bow features because almost all of the reviewers were using some set of words very often across all of their reviews. This helped in getting good features to classify them.

The classification into star rating was hardest because the classes were very similar and close to each other and the classes differ only by a small set of features and decision even in the real world is also very subjective. For example classes 2 and 3 (star rating) – both have some good points and some bad points about the movies. It’s very subjective to rate it as a 2 star or 3 star and depends on the individual and the weightage of the positive/negative sentiment.

1. Within task were some classes easier to classify then others? Why?

Yes within a task such as the star rating it was easier to classify 4 star or 1 star than to classify 2 star or 3 star. The reason is same they differ only by a small set of features and decision even in the real world is also very subjective. For example classes 2 and 3 (star rating) – both have some good points and some bad points about the movies. It’s very subjective to rate it as a 2 star or 3 star and depends on the individual and the weightage of the positive/negative sentiment.

Whereas clearly class with star 1 and star 4 were easier to classify as the features were more inclined towards either positive or negative opinion sentiment.

Which classifications were most similar and most different and why??

The star rating and the Positive/Negative class classification were similar because in both the cases we were trying to identify the opinion of the text. They differed only in the degree of the sentiment.

The star rating and Reviewers class classification were most different because in one we were trying to classify them based on the opinions and in other we were trying to classify based on the style of the author.

1. What features did you use and why?

I used a set of bag of words features that I selected using the term frequency in the training set and then using weka feature selection to reduce the number of informative features and some other document dependent features such as no of positive words, doc length, no of negative words, no of nouns, difference between positive and negative words, no of superlative words, no of negating words (not, isn’t etc).etc.

1. Did they perform better or worst then expected?

Some of them performed better but some didn’t such count of nouns in the text.

1. Did early experiments guide your thinking for your final submission? How?

Yes early experiments helped me in deciding that I would require a good set of positive and negative sentiment words to increase the accuracy of the classification

1. Which features were most/least helpful? Why?

Doc length was most useful.

Count of NN words was not useful.

1. If you used any external resources which did you use? How did they contribute to the success of your submission?

I used Porter Stemmer – It wasn’t helpful. It reduced the accuracy.

I used Stanford parser, it wasn’t helpful. It didn’t increase the accuracy but took a lot of time to process.

I used a list of positive and negative sentiment words. It was very useful. It helped in increasing the accuracy.