Prediction of User Ratings by Analyzing Reviews

Venkata Kancharla, Viranchi Deshpande and Vishak Lakshman
Department of Computer Science
University of North Carolina, Charlotte

Abstract

Consumers tend to buy a product or use a service which are highly rated and have good reviews. The economic effects the reviews have on the businesses it is often critical for the businesses to properly asses their ratings and plan measures to increase the rating.

It is quite common that not every review is meaningful and also the review doesn't always reflect the rating given. There should be a better mechanism to choose the reviews which are meaningful, analyze the reviews and then assess the business. Our project aims at developing a mechanism using machine learning techniques, to analyze the reviews and predict the rating of the user using the Yelp database. It generates a model which learns from meaningful reviews and ratings, and then can be used to input reviews to get the actual rating from that review.

Introduction

Most of the works [1] in this domain are biased towards providing recommendations for users and only some works tried to address the problem for the businesses. Also, prior works [2], did not consider taking meaningful reviews which reflect the rating given to develop the model. We focused on providing efficient results for the businesses to introspect themselves by developing a model based on meaningful and good reviews.

In our project, we are primarily trying to develop a model to predict the actual user rating based on his review, by analyzing previous reviews and ratings. For this, we considering two models, one which has text-based features and other has numerical features which are extracted from the text. We analyze their performance to see which predicts better.

We have taken the database of Yelp reviews[3]and analyzed the dataset. We used the upvotes(cool, funny and useful) to collect only the reviews that are meaningful and actually reflect the rating.

We performed some other preprocessing steps like, removing stop words, stemming etc. For gathering numerical features out of text, we used AFINN [4] score to convey the valence of the word. Along with this we developed a similar list of words and values, based on our dataset, which gives all the words in all reviews, a value from 0 to 5. For the text-based analysis, we used NB Classifier model and for numerical features we used Logistic regression model. We analyzed the output predictions and presented our views on the outputs.

Dataset

Yelp openly publishes datasets for academic purposes and challenges. This made us to get access to a good dataset upon which we can build our model. The dataset contains details about: User ID, business ID, review ID, review text, rating, upvotes.

Our dataset has 5.4 million entries initially. We analyzed the correlation between the rating and the upvotes to see whether they have any impact on rating, but they have low negative correlation. So, we decide to use them as a metric to gather good reviews[6], i.e. meaningful reviews. To get a good number of entries for our Since it is a large data file, we applied this constraint initially and then loaded the new data set into the project.

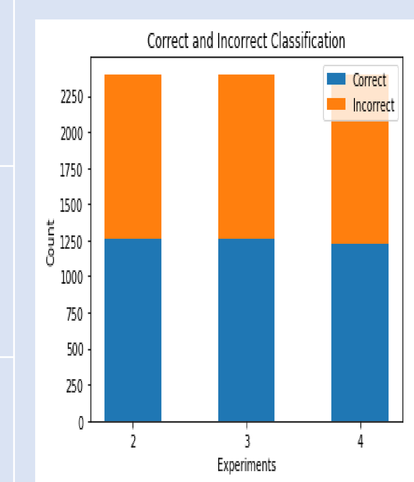
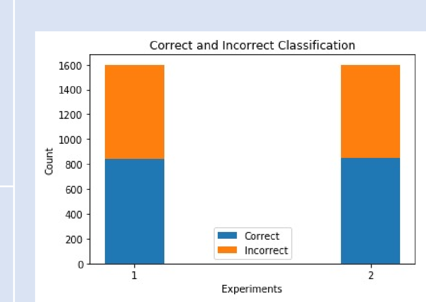
Methods

The reviews in the form of text are subjected to preprocessing and converted into numeric data, so they can be fed as inputs to the predictive models. The texts preprocessing mechanisms involved removal of stop-words, POS tagging and Stemming. The resultant dictionary of words is subjected to Count Vectorization which generates a vector as output with entire vocabulary as length and an integer count of the times the word has occurred in a review.

Since multiple occurrences of the words matter a lot in the classifying the reviews into their respective ratings, we chose the multinomial version of NB classifier. The numeric input data is oversampled before training and after training, the model makes predictions on the test datasets.

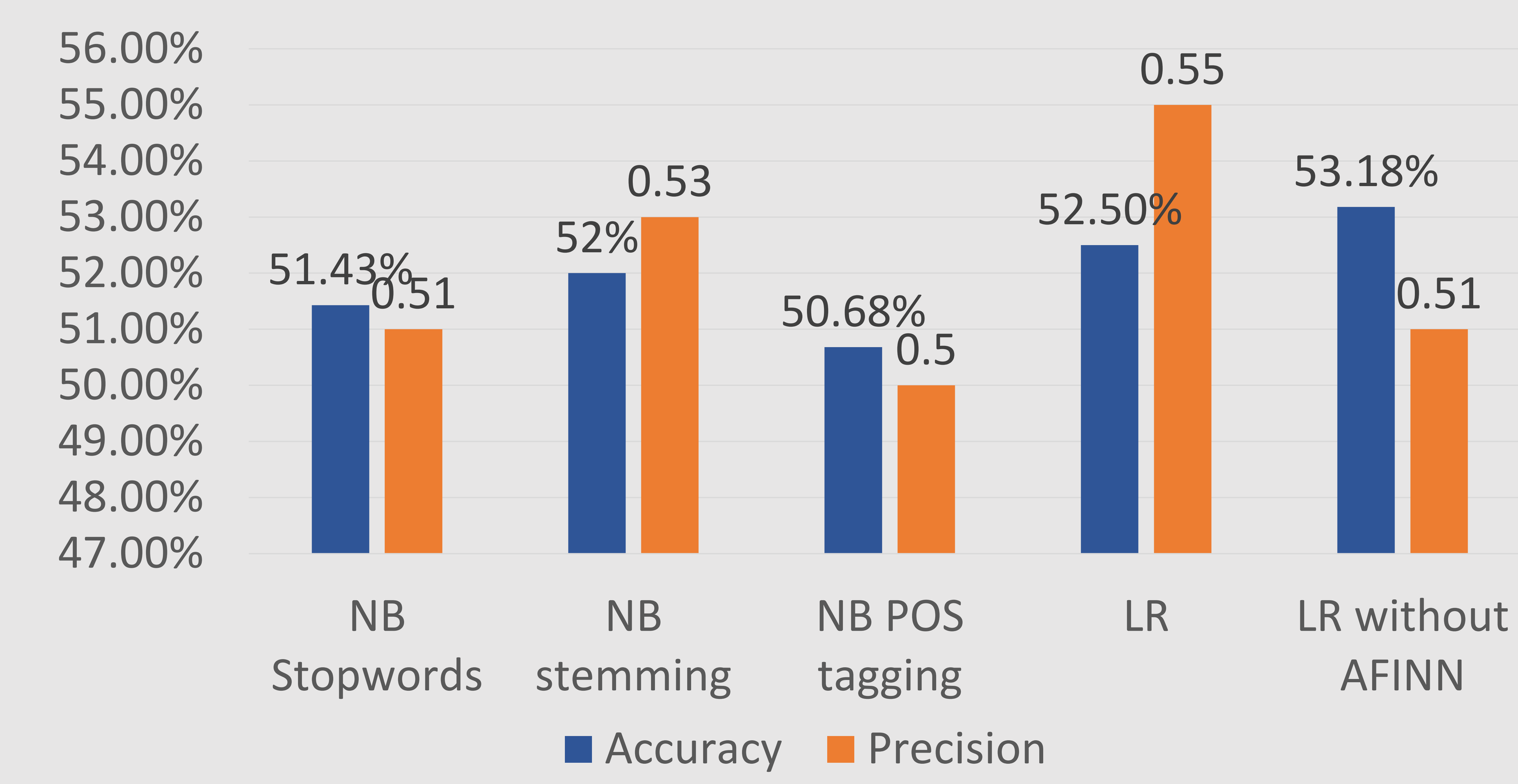
Incase of Logistic Classification, we made use of two derivative features. One is AFINN score which is a list of English words rated for valence with an integer between -5 and +5. The second feature, weighted average (introduced by us) is the weighted average count of each word in the vocabulary. A merged array of these two features is oversampled and fed as input to predictive model for training and testing.

Experiments and Results

EXP NO	CLASSIFIER	CRITERIA	RESULTS COMPARISON
1.	NB	All the punctuations and stop words are removed during text processing.	
2	NB	Stemming is performed along with stop words removal during text processing.	
3	NB	Stop words are removed, POS tagging get only adjectives, adverbs, verbs and then stemming.	
4	Logistic Regression	Both AFINN scores and weighted average as given as input to the model.	
5	Logistic Regression	Only the weighted average is given as input to the model.	

Experiment	Accuracy	Precision	Recall	F1 - Score
NB Stop Words	51.43%	0.51	0.51	0.48
NB stemming	52%	0.53	0.52	0.49
NB POS tagging	50.68%	0.50	0.51	0.49
LR AFINN and Avg Value	52.5%	0.55	0.53	0.53
LR Without AFINN	53.18%	0.51	0.53	0.52

Table 1. Performance comparison of all experiments.



Discussion

It can be observed that the logistic regression model, which is based on numerical features from text, performed better than the Naïve Bayes classification model, which is based on textual data. And also basing the model only on the average values of words obtained using the dataset, proved to be more accurate rather than mixing it with AFINN score. AFINN being a general list, not in context of dataset, might be a reason for this result. But the experiment with AFINN score has high precision.

Overall, the accuracy is not anything higher than 55, but for a multiclass classification with 5 classes, this can be considered as a good prediction. Most of the predictions are near to the actual value, i.e. class 3 is predicted as 2 or 4. With some more experimentation and analysis, a better model with high accuracy can be designed.

Conclusions

Through this project, we have grasped the basics of text processing and the various ways by which the text can be converted into numeric data for building predictive models. Through this exposure, we were able to introduce our own feature that converts raw text data into numeric data which we called it 'weighted score' of text. With the help of these features, we built two robust predictive models in the form of Naïve Bayes and Logistic for classification.

Conducting various experiments on different scenarios and parameters, we have able to achieve a moderate success of 51% across the different test cases. Among those efforts, the logistic regression model of our newly introduced feature against the target variable has produced maximum accuracy of 53%.

Future Work

To extend our work, in future, we are planning to perform more experiments to evaluate the models and parameters. A more sophisticated natural language processing techniques would be more helpful is extracting useful features from the data. We will try to bring in more machine learning models into consideration, like Support vector machines etc. and analyze their performance. Apart from the models, more aspects of the data can be taken into consideration, like a user's previous reviews and ratings, restaurant wise reviews and ratings, etc. for better analysis.

References

1. J Xu, X Zheng, W Ding, 2012, 'Personalized Recommendation Based on Reviews and Ratings Alleviating the Sparsity Problem of Collaborative Filtering', IEEE Ninth International Conference on e-Business Engineering (ICEBE), pp 9-16
2. David Nichols, 2016, 'Learning From Yelp', Harvard University, USA, viewed on 26 February 2018 <http://cs229.stanford.edu/proj2016/report/Nichols-LearningFromYelp-report.pdf>
3. Yelp, 2017, 'Yelp Dataset', Kaggle, USA, viewed on 5 April 2018 <https://www.kaggle.com/yelp-dataset/yelp-dataset>
4. Finn Arup Nielson, 2011, Danmarks Tekniske Universitet, Denmark viewed on 26 February 2018<http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010>
5. Jason Brownlee, 2017, 'How to Prepare Text Data for Machine Learning with scikit-learn', Machine learning mastery, USA, viewed on 22 March 2018 <https://machinelearningmastery.com/prepare-text-data-machine-learning-scikit-learn/>
6. Vaslis Vryoniotis, 2013, 'Machine Learning Tutorial: The Naïve Bayes Text Classifier', DatumBox, USA, viewed on 28 March 2018, <http://blog.datumbox.com/machine-learning-tutorial-the-naive-bayes-text-classifier/>

Acknowledgements

1. Dr. Minwoo Jake Lee, Professor, Department of Computer Science, UNCC
2. Yelp Inc. , https://www.yelp.com

Contact Information

Venkata Kancharla
9519 University Terrace Drive Apt L, Charlotte NC 28262
(704)712-3995