**Sentiment Analysis On User Reviews - Yelp**

Karan Desai, Kirty Vedula

Rutgers University

**Abstract**

Our project is on implementing sentiment analysis on user reviews obtained from Yelp’s academic dataset consisting of restaurant reviews for this purpose. The main of the project is to analyze sentiments on the user reviews and devise our own five-scale sentiment rating and try to match it to the star rating that the user provided. Following the n-gram approach for the analysis, we are analyzing adjective and adverbs that occur in particular sequences for this purpose. After calculating a weighted score for each review, we try to predict the users’ ratings by treating multi class classification using five different classifiers and compare and contrast their results. Here, we try to determine how accurate the users are, in giving the ratings by comparing their star based rating to with our calculated results.

Reviews that are very ambiguous or belong to notorious users are obviously given a very less weight in the average calculation. Because we are able to get a good accuracy for our sentiment analysis scores it gives a much more precise calculation and a fair review system. Possible future work can include working on business IDs obtained from the dataset. The overall business ratings are no more just the average of all given ratings but rather a weighted average based on the examination of reviews by our approach and the user’s accuracy ratings that we calculate.

**I. Motivation**

Today’s businesses rely highly on how users or customers have reviewed them on various websites. Bad reviews can be a turn off to new customers. Conversely, positive reviews can grow your client list and reinforce customer loyalty. Most importantly they diverge to profit or loss. In this age of smartphones it only takes less than thirty seconds to check the review of a product before buying it or of a restaurant before dining in. This means one is highly relying on the users who review those businesses and products. We are also relying on the honesty and integrity of the user. There are a lot of factors to consider here. Some sites in this category are Zagat, Tripadvisor, Edmunds, Kbb and OpenTable.

Nielsen Study commissioned by Yelp concluded that 82% users visit the site before they intend to buy. Not only that but 44% users read the review text, 26% see the ratings , 17% see the number of users while 14% see reviews from family and friends. 85% of consumers use the internet to find local businesses. It has other material like how businesses on Yelp profit more than those who don’t.

A study by Maritz research that included 3404 people found that one in four people believe the information available on ratings sites is unfair. And while the older, highly visited sites were generally perceived as more trustworthy, more than a third of visitors were still cautious of information on these sites. There may be a credibility crisis on the horizon for online review sites. If the lack of confidence in customer reviews continues, these sites could become obsolete and it is worse for consumers and businesses.

Yelp is a business review site that helps people find cool places to eat, shop, drink, relax and play, based on the informed opinions of it’s user community. It uses a five-point star rating system with user reviews. F

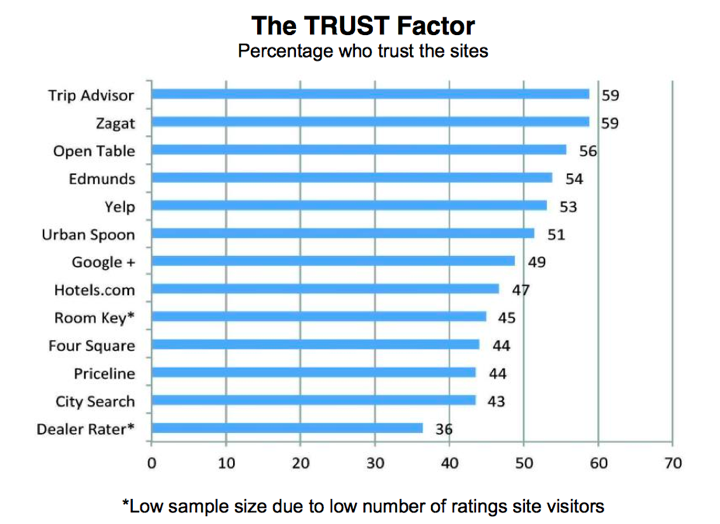


Figure : Maritz research about percentage of consumer who trust different websites

**II. Problem Description**

Summarizing the applications of our project work

1. Generating ratings from user reviews in systems where ratings may be missing
2. Rating the user by analyzing his reviews and rating
3. Filtering certains kinds of spam based on ambiguous reviews
4. Being able to give weight to user reviews for final average calculations.

We are proposing a method that will try to eliminate such unfairness from user reviews. Yor the scope of this project we will work with Yelp’s restaurant subset of data. First we will do sentiment analysis on user reviews to predict their Likert scale ratings. Then we match these to the ones we already have to see how accurate our predictions were. We also give an accuracy rating to the users based on his how their five-star rating matches to our calculated ratings. Depending on these factors the business’s average rating takes into account the individual review weights as well as weights from their users/reviewers and not simply just the average of all reviews. This comes as a part of natural language processing and artificial intelligence where the agent is developed over multiple iterations for predicting the user’s ratings basing on the words given in the reviews.

We are doing the sentiment analysis and giving a five point based rating (0 to 5) while most analyzers are only based on positive or negative sentiment. We are using a Bi-gram approach for sentiment analysis that is based on adjectives and adverbs occurring together or adjectives occurring by themselves. A more detailed explanation of the methodology is the next section.

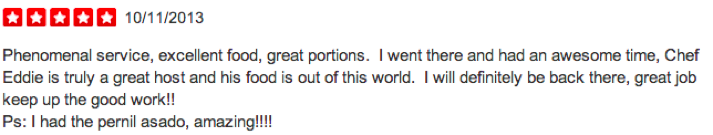


Figure : Yelp Sample Review with 5 Star Rating System

3. Motivation, Choice of AI Approach, and Prior Work on this kind of problem (with references) – Novelty of your approach or routine useful application, with reason for using which AI approach – 1 to 3 pages

**(each one’s part)**

4. Implementation – Choice of your specific representation and programming language and methods – system diagram (input/output) and reasoning methods diagram and flow of control diagram – problems encountered and how overcome – distribution of work among team and role of each participant - 2 to 4 pages

**Implementation**

The steps that need to be done for the implementation of this project are listed in the steps below

1. Acquire Dataset

2. Split the dataset into only a user-review set

3. Use Parts-of-Speech Tagger to extract adjectives and adverbs

4. Arrange text in order to analyze bi grams. Stars-to-Words mapping only

5. Sentiment Analysis input preparation – Get

6. Analyze the text and come up with a Likert (5 point scale) rating

7. User ID’s and Business ID’s can be in separate locations indexed by review order

8.

**Data Set and Text Processing**

On request one can acquire a dataset from Yelp. This dataset is a specially crafted academic dataset. It has three kinds of objects in the dataset Business Objects, Review Objects and User Objects. We only need information from the User Objects for now. These objects have the review text we are looking for. There are approximately 330,000 reviews that we are going to user for this project. We will alter divide them for classification. The following is the exact JSON based data that we get in these objects.

|  |
| --- |
| DATA SET FORMAT  {  'type': 'review',  'business\_id': (the identifier of the reviewed business),  'user\_id': (the identifier of the authoring user),  'stars': (star rating, integer 1-5),  'text': (review text),  'date': (date, formatted like '2011-04-19'),  'votes': {  'useful': (count of useful votes),  'funny': (count of funny votes),  'cool': (count of cool votes)  }  } |

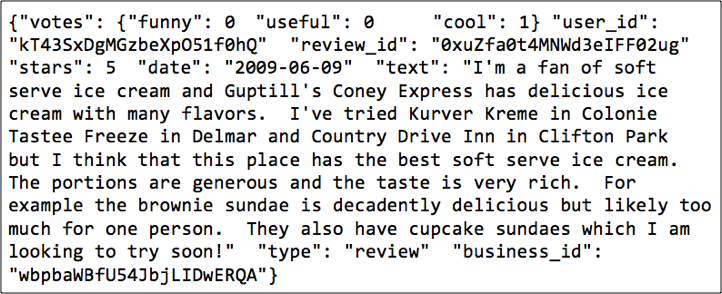


Figure : Raw Data Snapshot

We use python as our main programming language. The version of python being used is 2.7.5. Most of our dataset related processing is CSV(comma separated values) format that can be read easily in any programming languages like python or using softwares like Microsoft excel. From this dataset we extract the fields: business\_id, user\_id, stars and text. These are all maintained in CSV formats.

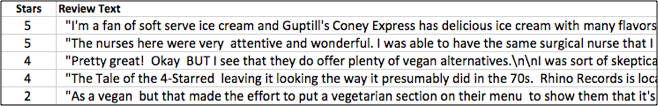


Figure. Data after extraction

**NLTK and POS-Tagger**

The second step is using a POS-Tagger to identify the adjectives occurring in the review. The process of classifying words into their parts of speech and labeling them accordingly is known as part-of-speech tagging, POS-tagging, or simply tagging. Parts of speech are also known as word classes or lexical categories in such scopes. We also need to identify the adverbs surrounding these adjectives in the text. NLTK (Natural Language Toolkit) provides this functionality of extracting the parts of speech. We used the latest version NLTK 3.0. The following is an example from the POS-Tagger that NLTK uses.

|  |
| --- |
| POS-Tagger  >>> import nltk  >>> text=nltk.word\_tokenize("This restaurant is very good .I love it here")  >>> nltk.pos\_tag(text)  [('This', 'DT'), ('restaurant', 'NN'), ('is', 'VBZ'), ('very', 'RB'), ('good', 'JJ'), ('.I’, 'NN'), ('love', 'NN'), ('it', 'PRP'), ('here', 'RB')]  Bi-gram  ('very', 'RB'), ('good', 'JJ'), |

Figure: POS-tagger processes a sequence of words, and attaches a part of speech tag to each word

Once we have an output of this kind we select words with tags ‘JJ’ and ‘RB’. ‘JJ’ stands for adjective and ‘RB’ stands for adverbs. We are going to ignore other parts of speech for now. Now we have a csv file with user’s star ratings, the words from the POS tagger and the user ratings.

**Opinion Lexicons and Sentiment Analysis**

Opinion Lexicons or Dictionaries have to be obtained in order for us to do sentiment analysis. We got two text files with list of negative and positive words. Both of them have approximately 7000 words and should be enough for our scope. Having more words obviously implies more accuracy. We also have two text files that have lists of negative and positive adverbs. For the Sentiment Analysis we consider the Bi-gram approach (n-gram with n=2). What this means is that we analyze sentiments in pairs of two words wherever there is an adverb-adjective pair. This is very important step for sentiment analysis. Consider a scenario where we found the adjective *“good”.* If the two preceding it said *“not”(a negative adverb)* the sentiment is inversed. On the other hand if the word preceding the adjective is *“too” (a positive adverb)* then it means that the positive sentiment of the adjective is boosted. So based on this logic our python program analyses the keywords and calculated a sentiment score. For each positive sentiment in the sentence we add 1 (+1) to the total score and for each negative we subtract 1(-1). For adverbs at correct positions, positive adverbs mean multiplied by two (x 2) while negative adverb means inverting the sub score (x -1). The final sentiment likert score is obtained by first dividing the score by the number of words and then normalizing it to the 5 scale. Thus



Figure . Data Snapshot after Sentiment Analysis

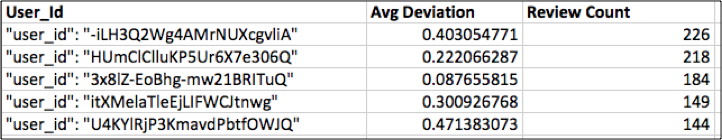


Figure. User Based deviations from actual rating

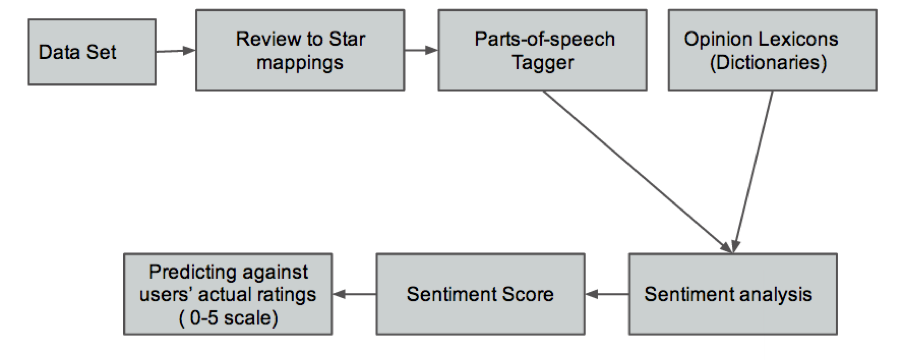
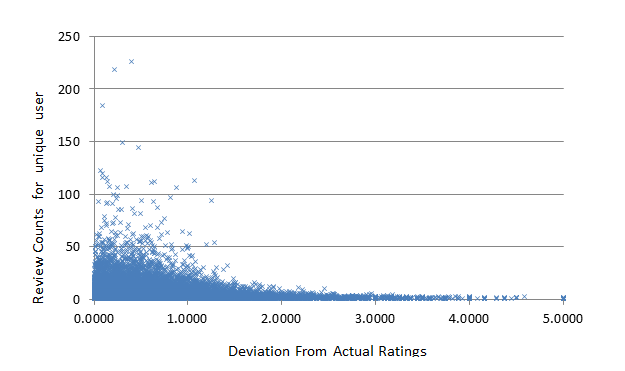


Figure: Flow Chart: Predicting User Ratings from Reviews

5. Preliminary results as presented last week and how this influenced changes in project –either formulation, implementation and results/ outcomes – 2 to 3 pages

6. Final results – 2 to 4 pages

7. Discussion – comparison to your own preliminary results and prior research results on this kind of problem, effectiveness (or not) of collaboration)- 1 to 3 pages



geneal accuracy of sentiment analysis

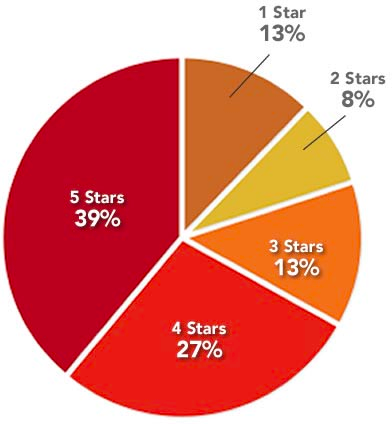


Figure . Yelp Rating Distribution ( Complete Set of All kinds of businesses )

8. Future possible extensions of this project if you had more time (6- 8 weeks)

and possible impact beyond this course – ½ to 1 page

Business IDs -

SVMs -

3. Individual summary of your Contributions to the Project and what you could have done differently either as individual or as a group ( ½ page to 1 page)

Machine learning techniques are often used for data analysis and decision-making tasks such as classification of categories, estimating probabilities, and data mining. However, implementing and comparing different machine learning techniques to choose the best approach can be challenging.

**ACKNOWLEDGEMENT**

This project would not have been possible without an excellent data set provided by Yelp. It has a huge potential for experiments. We would also like to thank Casimir Kulikowski, Rutgers University for his help regarding various topics on the matter.

**REFERENCES**

Yelp Survey <http://officialblog.yelp.com/2012/04/search-engine-lands-recent-local-consumer-review-survey-looked-at-the-way-consumer-behavior-has-changed-since-2010-intere.html>

Maritz research <http://www.maritzresearch.com/~/media/Files/MaritzResearch/Press/Maritz-Research-2013-Online-Customer-Review-Study-Release-FINAL-9-18-13.ashx>

Opinion Lexicon -<http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

Stanford’s NLTK - <http://nltk.org/>

Sentiment Analysis of Twitter Data <http://www.cs.columbia.edu/~julia/papers/Agarwaletal11.pdf>