CS 224W Milestone

**Detecting Expert Reviewers**

Ryan Foley and Qinghui Ji

**Abstract:**

Online reviews have come to be of great importance for the online retailer and they have been shown to have significant impact in the behavior of customers [Zhu citation?]. Therefore, we present insight into the characteristics of online reviews and describe methodologies by which to classify reviews. In particular, we will examine the differences between reviews from novice and expert reviewers and will demonstrate how to predict from which of the two sources a previously unseen review originates. To do so, we investigate reviews from BeerAdvocate.com which we show to be what we regard as an excellent source for the study. We believe that this paper may have great relevance to the online retail and marketing spaces.

**1. Introduction:**

The greatest advantage internet retailers have over traditional brick-and-mortar stores is the ability to sell more obscure items. On Amazon.com, somewhere between 20 to 40 percent of unit sales fall outside of its top 100,000 ranked products [BHS03]. Traditional retailers are limited by what can be stocked on the shelves and therefore often resort to merely displaying the most popular items. On the other hand, online stores are not limited by shelf space and therefore can present a seemingly infinite list of products. In many markets, they have already begun to replace(quote about Amazon books selling in the long tail from lecture or viral marketing) brick-mortar-stores, such as in the case of Amazon over Borders and Barnes and Nobles. This has happened primarily with the internet retailers ability to sell products more in the long tail.

Although this is a great advantage for the online retailer, actually convincing a potential customer to purchase an obscure product that she or he may have never heard of presents itself as more unusual and difficult task. This challenge has given rise to the creation of advanced recommendation systems. However, another important sales tool of the online retailer is product review. For an online shopper looking at a product that they may have never seen or heard of before, a review from someone who has actually purchased and experienced it is likely persuade the shopper to purchase or not. This effect of the review on the online shopper has been shown to be significant and of great importance to online stores [Zhu and Zang]. Thus, it is essential that reviews be of high quality. However, given the long-tailed nature of online product listings, it is often infeasible for review writer services to be contracted privately, so instead it is far more common for reviews to be written by the public where the only prerequisite is an account with the website. This method allows for reviews of any level of quality to be submitted into a system where the review may be critical to sales. Many online stores attempt to maintain quality review systems by allowing users to either vote up or down a vote and therefore either raise or lower the visibility of a review. However, again the long-tailed nature of online products prevents this method from being a perfect solution because it may often be the case that an obscure product only has a few reviews that are not of high quality and have not been voted on by other users.

Given this dilemma, we present a case study into the differences of inexperienced and expert reviews and the possible classification of a never before seen review as either from an expert or novice. Although it is likely that the nature of a review varies across product categories, we have decided to look at beer reviews from the BeerAdvocate.com because we believe that this data set may provide sufficient variation in products while still remaining in a well defined category. Furthermore, we believe that this data set has a particular advantage over others because many of its reviews have been written by both expert and novice users alike. This will provide us with a greater sample size and of higher quality from both groups from which to extrapolate meaningful insight into the differences between the them.

**2. Related Work:**

- ashtons

- references at bottom of here

- our reaction paper

**3. Data Set:**

In this portion of the paper, we will describe the data set, explore some of its interesting characteristics, and derive features that potentially will be used in our model to predict whether a review is from an expert or novice reviewer.

The data set for our study comes from the BeerAdvocate.com website and in total consists of 1,586,614 reviews of 66,055 different beers from 33,388 active users[[1]](#footnote-0). The variables we will be using to study the activeness of the users are: (1) reviews s/he produced. (2) his/her voting behavior (3) type of beers s/he reviewed.

**3.1 Reviews**

The median number of reviews per users is 3. The first quatile is 1 and the third quatile is 16. The mean is 47.52, due to the fact that there are a few users produce much more reviews than other. For example, the maximum number of reviews produced by an individual is 5817.

First we run a log-log plot on number of reviews per indiviual versus number of people have that many reviews. In Fig 1, the x-axis represents the number of reviews given by an individual and y-axis represents the number of users have that many reviews. We see it appears to be linear, which indicates a power law distribution. The characteristic ‘fat-tail’ of this distribution will be beneficial in that there will be a substantial number of reviewers who have many reviews, which we characterize as experts, as well as those who have few. Also, as we expected, users who have one reviews contribute to almost ⅓ of total users. There are 10,443 users have only one review.

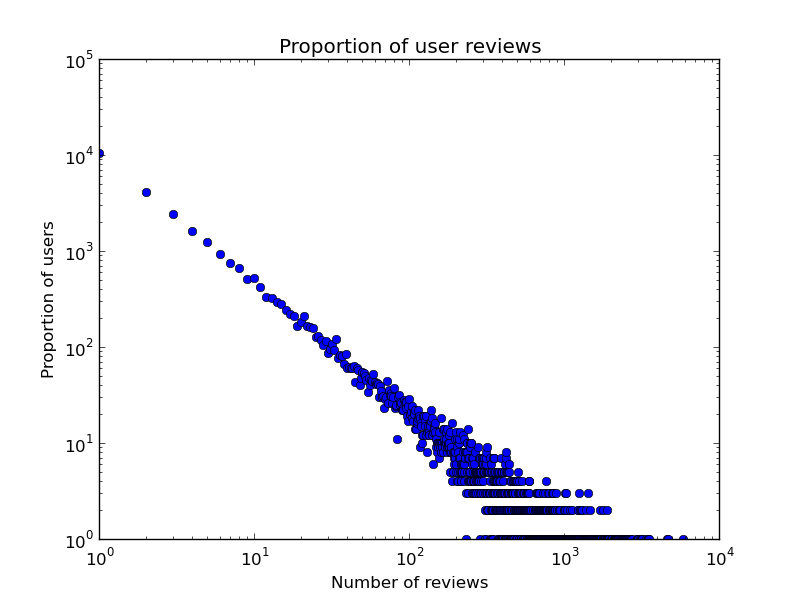
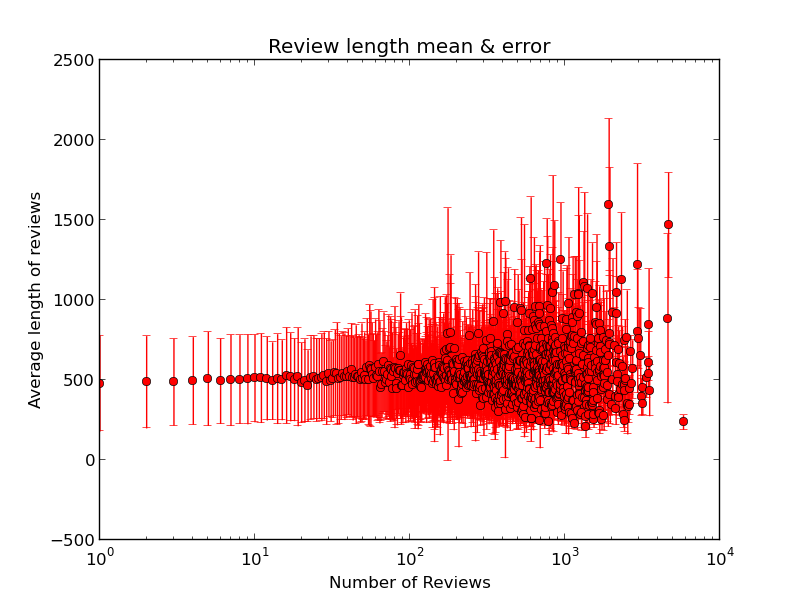


Fig1 - Proportion of user reviews Fig2 - Reviews length (mean +/1 std)

Second, we want to see whether users produce more reviews also produce higher quality reviews. The criteria we use here is the average lengths of reviews. We find the mean by the following method: for example, to find mean for the length of reviews produced by people who only contribute one review, we sum up all the word from those 10,443 reviews and divided by 10,443. Another example, if there are 5 users who individually produce 100 reviews, to find the mean for the length of reviews produced by people who contribute 100 reviews, we sum all the all the words from those 500 reviews and divided by 500. Fig2 shows that people who produce less than 100 reviews have almost a stable number of words imput, and standard deviation too. Since the 3rd quatile is 16 reviews, we know that more than 75% of users produce reviews with average number of words 500.

Once the number of reviews pass 100, number of words start to vary more. Some writes much more than 500 words, like two points reach 1500. Some writes much less. An interesting thing we see in the graph is that the person who contributes most reviews (5817) actually does not write that much, with average words count less than 250.

**3.2 Vote**

We want to see users’ voting behavior is associate with their activeness. First we calculate the mean overall vote from each quatile and the top 1% users:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1st Quartile | 2nd Quartile | 3st Quartile | 4st Quartile | top 1% users |
| mean | 4.02796 | 3.97195 | 3.89705 | 3.80902 | 3.77433 |

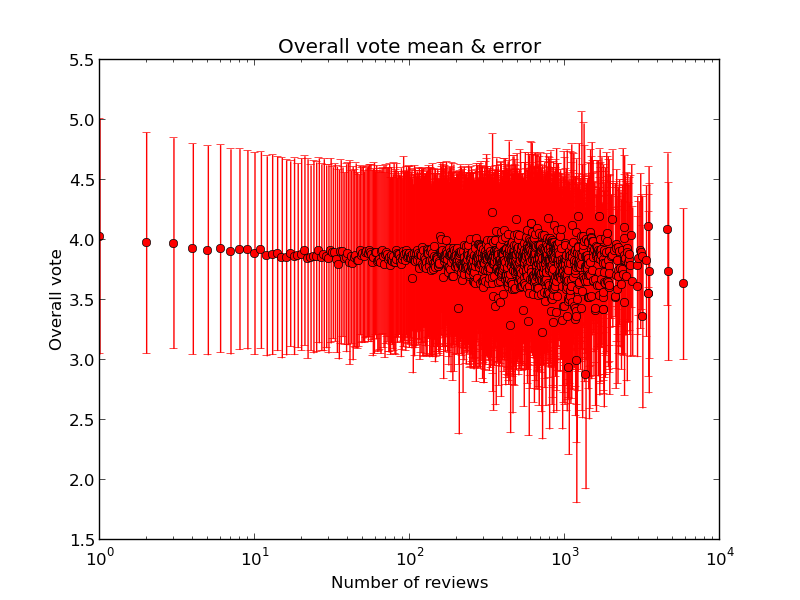


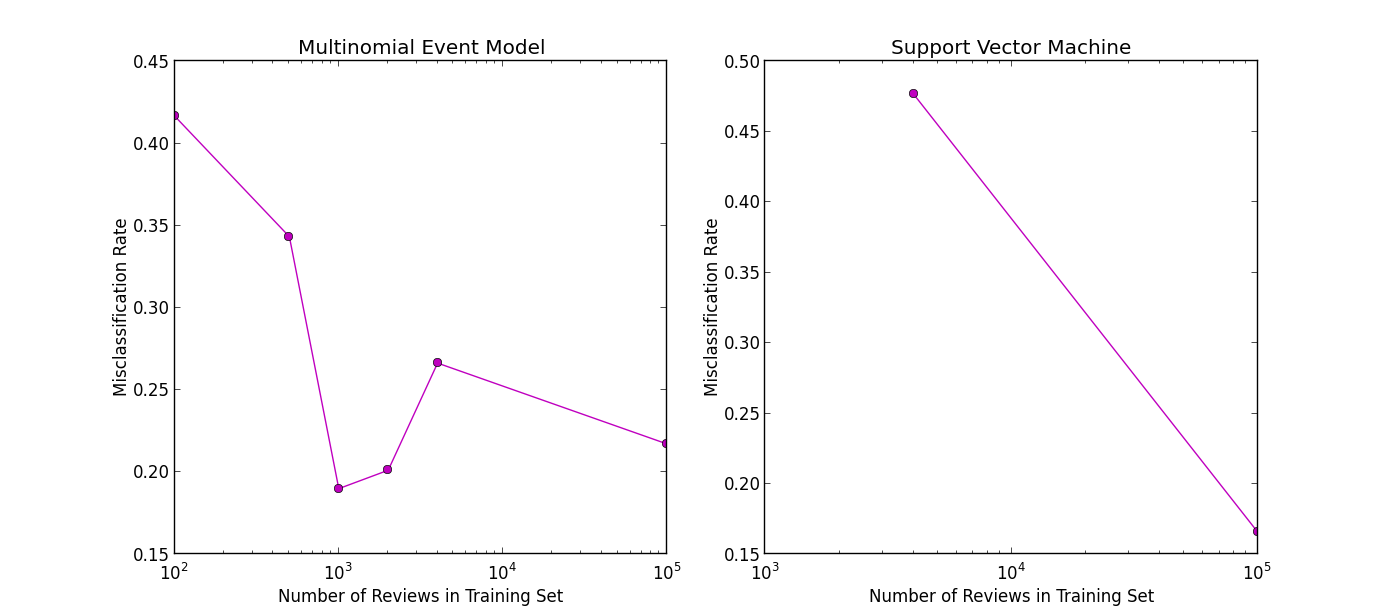
Fig3 Plot of overall vote (mean +/- std)

**3.3 Beer popularity**

**3.4 Timeliness of Review**

**Prediction Task:**

The previous section described potential relevant features in predicting whether a previously unseen review has been written by an expert or novice reviewer. Now we will use these features to develop and test various models to determine which is appropriate for this task. However, before adding in features previously described, we will first compare prediction algorithms using just the preprocessed text of the review as a baseline for our prediction task. The following are comparisons between a Naive Bayes classifier with a multinomial event model and Support Vector Machine with respect to the size of the training corpus:



With a training set size of 100,000 reviews both models did reasonably well, 0.2172 for the multinomial event model and 0.1658 for the SVM. We chose to compare the SVM to the multinomial event model because it is regarded to outperform the multi-variate Bernoulli event model for text classification [Andrew Ng lecture note].

This classification was performed by selecting a random subset of users with an even split between expert and novice users. Here we defined novice users to have 3 or less reviews and experts by having 100 or more. Each model was trained on 70% of the sample and tested on the remaining 30%, being careful that reviewers that were in the training set were not also in the test set in order to avoid training bias. We believe that this segmentation of reviewers between test and training was essential to the prevention of training bias and could have been an easy oversight. Errors were then calculated as average misclassification rates.

**Future Improvements:**

We will be investigating more features of the review data. They include the time of the review with respect to the others and popularity of beers reviewed. With respect to building the model, we will continue with the Support Vector Machine and add in all features found one at time and calculate the improvement from adding each feature into the model. Also, we would like to use a larger sample size and perform k-fold cross validation with 10 folds in order to get a more accurate estimate of model errors. Lastly, we will look at using some regularization with either an L1 or L2 norm in our model.

**Possible References:**

Dynamics of Viral Marketing

Ashton’s stack overflow paper

Feng Zhu & Xiaoquan (Michael) Zhang

Impact of Online Consumer Reviews

on Sales:The Moderating Role of

Product and Consumer

Characteristics

http://www-bcf.usc.edu/~fzhu/ZhuZhang2010.pdf

**possible quotes:**

viral marketing:

We ﬁnd that product purchases that result from recommendations are not far

from the usual 80-20 rule. The rule states that the top twenty percent of the products

account for 80 percent of the sales

1. Active users here are defined as those who entered at least one review. [↑](#footnote-ref-0)