Detection of Constructive and Non-Constructive Reviews

Sivasubramanian Kandaswami, Graduate Student, University of Florida,51447667

***Abstract*— The rise in the amount of goods and services being delivered through a digital platform has created a dependence on reviews available on these platforms, for the customers to make an informed decision. While having access to these reviews is convenient, the drawback is the sheer volume of opinions available. It is vital that people have access to the reviews that are clear, informative and terse. Companies such as TripAdvisor, Amazon are constantly seeking new ways to ensure that their users have access to the detailed descriptions. In this paper, I use the a dataset of reviews from TripAdvisor and make use of Latent Aspect Rating Analysis, (LARA) [1] and other lexical features within the text of the review and other key aspects, I have created a system that detects high quality reviews and also detects the reviews which do not have much useful information to offer.**

I. Introduction

A. *Emergence of Digital Platforms and Opinions*

The last decade has seen the emergence of large digital platforms, built on the backbone of an increasingly connected world, thanks to the internet. There have been major changes in the in the way consume entertainment with the emergence of live streaming technology and ecommerce platforms and with the introduction of these services, we have also seen that time spent on the digital platform mirrors real life.

People share their opinions on many platforms and the abundance of these shared views has introduced a new kind of problem, i.e. the problem of sifting through noisy, opinionated reviews of goods and services in an attempt to identify the most informative, unbiased and comprehensive reviews which answer the types of questions a potential customer is likely to have. The quantity of opinions regarding various goods and services makes it necessary for the owners of these digital platforms to ensure that they have a robust system in place to ensure that the helpful opinions make their way to the top of recommender systems and less helpful reviews are relegated to a lower position since they are less likely to be helpful to any person using them.

B. *Understanding the User's point of view*

There are several approaches have been taken to solve this problem, such as adding a rating to the review, which allows the customer to instantly form an opinion, especially when the aggregate of the ratings is clearly mentioned. While the ratings are very useful for the busy, impatient customer, a more discerning customer would want to be able to access as much information as she probably can, since more information allows the customer to make more informed decisions. This type of customer is willing to invest more time and should hence be provided more detailed reviews.

C. *Features of an ideal review*

The ideal review is a delicate balance between subjective and objective opinions. The subjectivity of a review helps the customer understand the feelings and personal bias of the reviewer and can thus be important. For example, a person who owns a 50$ pair of headphones might be accustomed to listening to good music whereas someone who uses the expensive headphones for the first time is likely to judge the product as being extremely good. Similarly, the objectivity expressed in the review is also important since they provide important information such as facts regarding the product, which tend to stay invariant and such statements are key to understand more about the product/goods or service. An example of this can be the fact that the new Samsung phone has two 'edges' on its display. This is a fact that provides a physical description of the phone, which is an indication of a relevant review.

D. *Context Specific information*

Identifying aspects within reviews which provide more information regarding aspects that are likely questions. Zhai et. Al, extracts features specific to hotel reviews such as the cleanliness of the rooms, the service provided by the staff at the hotel, whether the hotel provides good value for money, the location of the hotel i.e. whether it is near or farther away from common tourist attractions etc. tend to be a common concern for most users of the website and the reviews that address the questions are the ones that are most likely to be helpful to the users. Other hotel specific aspects such as the quality of sleep add to the comprehensiveness of the review and ensure that they provide the users of the website enough information so that they feel confident in making the decision to stay at the hotel.

II. Dataset and Review Samples

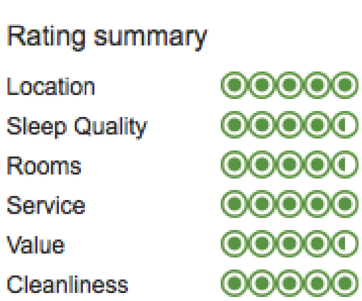
The dataset for this project is a set of reviews in the following format.

The review show above is an example of a review which provides some information which is useful, but does not go into a lot of detail and does not rate the hotel based on auxiliary ratings.

This on the other hand, is an example of a review that not only provides information within the review but also provides auxiliary information, which can be used to gain a better understanding of what it is like to stay at that hotel.

III. EXTRACTING AND PREPARING INFORMATION

For this project, I elected to go with the dataset provided by the website TripAdvisor, which provides reviews for various hotels and also features a multi-faceted rating system, which not only has the traditional overall rating, but also includes individual ratings for cleanliness, location, rooms, service value for money, which are all extremely important factors to a typical customer who is on the lookout for a good hotel to stay at.



In addition to this information, we also obtain the written content of the review, the number of people who have read the review and the number of people who have found that particular review to be helpful. This information was collected and organized using Python. Once the dataset has been created, the next step involves cleaning and preparing the data .There is some amount unwanted data such as data in languages other than English, reviews in which ratings for certain aspects are missing. This unwanted data is purged and data is converted to a format such as JSON, which can easily be parsed using Python. This step is especially important since the ease of use and speed at which data in this format can be read is useful when a large amount of data is being processed.

IV. EXTRACTION OF USEFUL FEATURES

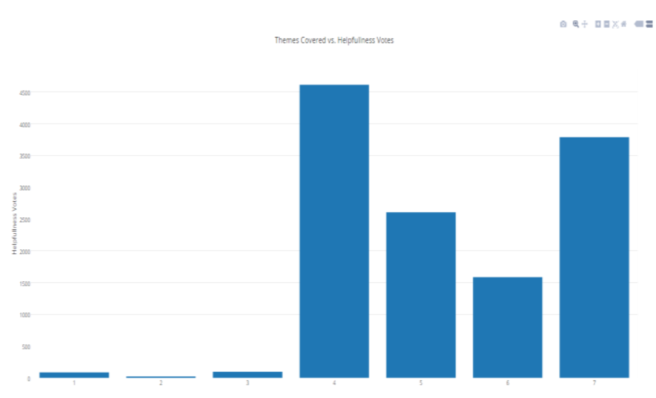
A. *Content of the Review*

The main text contained in the body of the review, is the length of the review. The content of the review provides features such as the level of objectivity and subjectivity of a review. The overall sentiment of the review can also be determined however we refrain from using the sentiment of the review as a classifier since the goal is to determine helpful reviews and such reviews can express a negative sentiment just as likely as a positive sentiment. For this reason, the features that we extract from the text content of the review is limited to the level of subjectivity and objectivity in the review.

A package called 'TextBlob', allows the subjectivity and the objectivity of a tokenized sentence to be calculated using the Natural Language Toolkit (NLTK). A longer review doesn't necessarily mean that it will be a helpful review but longer reviews are more likely to share more information about the subject, which in this case is the hotels they describe.

B. *Number of important themes covered*

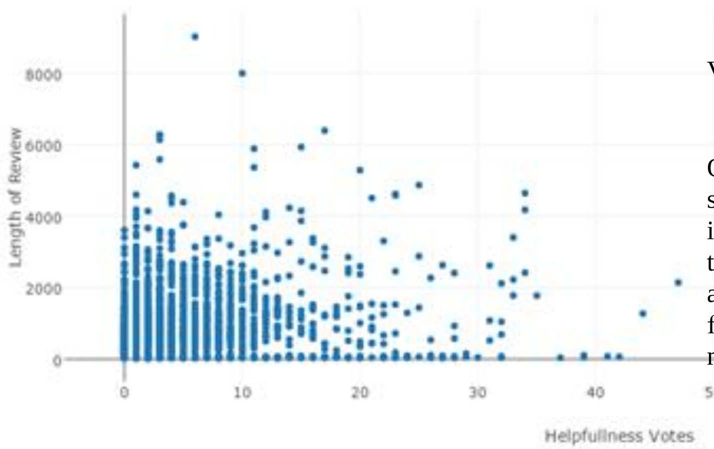
The multi-rating system provides a unique level of information that is traditionally absent from most reviews. We keep track of the number of additional themes that the user has commented on. Reviews which cover a large number of themes are read by more users and fond to be useful by a larger group of users and have a greater chance of being among the most meaningful reviews.



This graph shows the relationship between the number of features covered by a review and its helpfulness. It is evident from this graph that there exists a strong correlation between the number of topics mentioned by the review and its helpfulness to the users. This indicates that the review is comprehensive and covers several key aspects of the hotel that the customers wish to know about. This feature can thus be used as an important factor in determining whether or not a particular review is constructive. Reviews which mention less than three important themes are typically not found to be helpful by the users. We can postulate that a possible reason for this is that most readers of the review found the review to be lacking in certain key details that they felt were necessary for the review to be counted among the most helpful reviews.

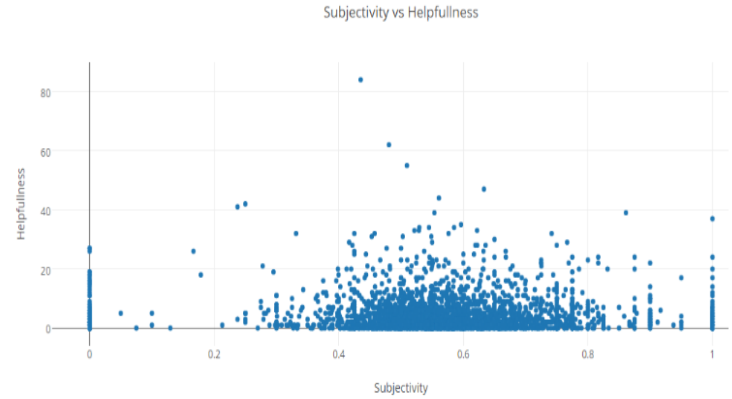
C. *Length of the Review*

The length of the review is a critical feature in determining the utility of a review. Reviews that are too short usually do not provide any meaningful insight into the hotel and are often highly subjective and brief. Some reviews on the other are too long, verbose and might not contain information relevant to the user. Helpful reviews tend to have a larger than average length when compared with to a review which isn't as helpful.



D. *Subjectivity of the Review*

Almost all reviews carry subjective statements. This is because reviews tend to be inherently subjective since it is the expression of how the reviewer's feels about a service or a product. Most helpful reviews maintain a balance between subjectivity and objectivity, since stating objective facts about the service is also an integral part of writing a good review. As we can see from the graph shown below, apart from a few outliers, most of the reviews which have a high number of counts of helpfulness votes are concentrated towards the values where the subjectivity is between 0.4 and 0.7. A score of 1 indicates that the review is highly subjective and a score of 0 indicates that the review is highly objective. Thus the scatter plot being very dense at the center is in alignment with our initial claim that most helpful reviews express a combination of subjective and objective opinions.



E. *Overall Rating*

Reviewers who provide an overall rating allow some users to make up their mind quickly. Most reviews carry an overall rating since it is easy to do even if a reviewer is unlikely to type in content regarding the hotel. The lack of an overall rating for a review almost certainly means that the review will not be helpful to any user. This makes it another important feature in the detection of the unhelpful reviews.

V. OPERATING ON THE FEATURES

A. *Normalization of Features*

Once the features have been identified and extracted, the next step involves preparing them so that they can be provided as input to a classifier. Normalization is performed by identifying the maximum and minimum values of a particular feature across the entire training dataset. Every occurrence of the feature's value is then normalized using the maximum and minimum values determined.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

The formula shown shows the normalization of the value of X into the value X', by subtracting the least value of X in the feature vector distribution from the current value of the feature X and dividing it by the difference between the maximum and minimum values in the distribution. This step is repeated for every feature that will be used to train the classifier. The normalization process ensures that every value is expressed as a number between 0 and 1.

VI. CHOOSING A CLASSIFIER

The next step involves training and testing the dataset and the extracted features using a number of different methods and types of classifiers in order to determine the classifier that is best suited to handle the type of data and the features that we have generated.

A. *Naive Bayesian Classifier*

The 'Naïve' Bayesian classifier is a simple classifier that relies on a bag of words approach. Given a document containing several sentences, these sentences are tokenized to words to generate the bag of words representation.

This classifier does not pay attention to relationships that may exist within the words within the document and instead chooses to treat every token as an independent. Although this may suggest that this might be a big flaw, in practice the bag of words approach fares very well.

$$p(w_i|C)$$

This represents the probability that the ith word in a document occurs in a document that belongs to a class C. The probability that a document, D contains the words in a given class, C is given by

$$p(D|C) = \prod_i p(w_i|C)$$

This leads us to the situation that we are most concerned about i.e. regarding the probability that the document D may be classified as belonging to the class C. This, is given by

$$p(D|C) = \frac{p(D \cap C)}{p(C)}$$

The Naïve Bayesian classifier brings together the probability model described above with a decision rule. The rule often picked is the hypothesis which has the highest probability of being correct. In formal terms, it refers to the maximum a posteriori. A Bayesian classifier can essentially be described as a function in the following manner.

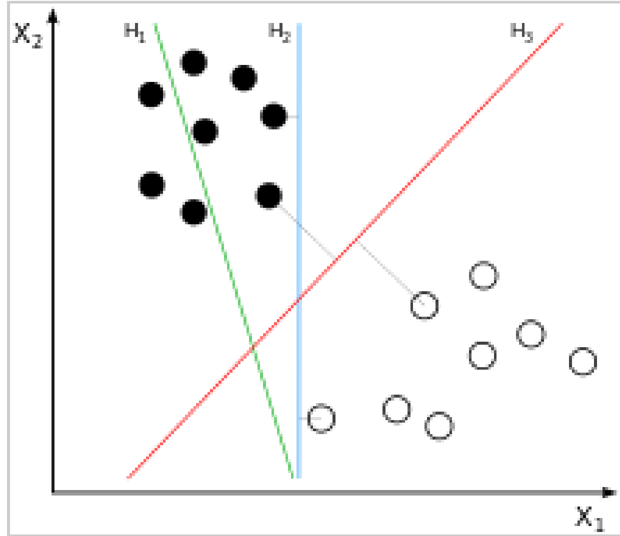
$$\hat{y} = \underset{k \in \{1,\dots,K\}}{\operatorname{argmax}} \ p(C_k) \prod_{i=1}^{n} p(x_i|C_k).$$

This assigns the class label $\hat{y} = C_k$ for some *k*.

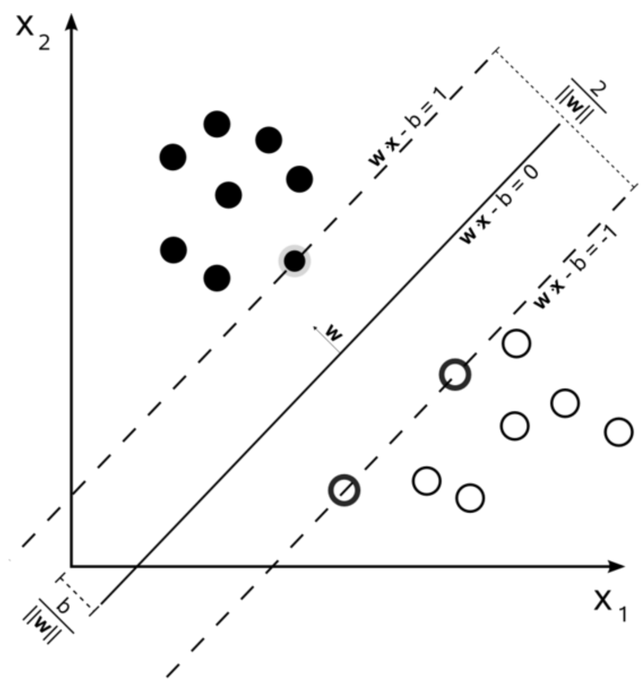
B. *SVM Classifier*

Support vector Machines have their basis in identifying a hypothesis for which the true error is minimum. The SVM creates several hyperplanes in high dimensions and is very adept at classification, among other tasks. SVMs attempt to draw hyperplanes such as H_1, H_2 and H_3 as shown.

SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. The SVM model is a representation of points which mapped in such a way that the samples of the separate categories are divided by a clear gap that is as wide as possible.



In the picture above, H_3 is the best separator, since it separates the classes. With the greatest distance. This, is an example of a linear SVM classifier and H_3 is an example of a maximum margin hyperplane. They are especially good in multidimensional places. When there are several features to be trained and the system is constrained on memory, it proves to be useful. The versatility of SVMs also means that they can be used for a variety of different functions.



The figure above shows the hyperplane with the greatest margin for an SVM trained with two kinds of classes. We can build on SVM using a method called Support vector clustering, (SVC) which builds on kernel functions and is usually used with machine learning which is not supervised.

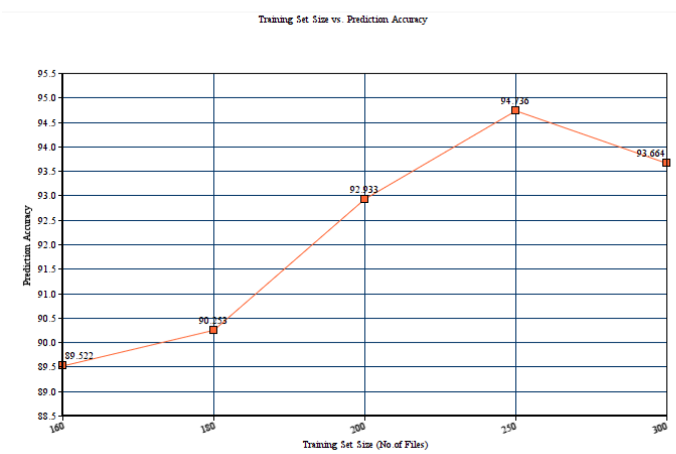
The reason SVMs are popular and are used in text categorization is because of their ability to decrease the requirement for training that involves labelled instances.

VII. RESULTS

After training the classifier using the SVM classifier, with different number of training and test review dataset, the classifier was found to have an average correct prediction rate of about 92%.

Size of Training Set	Prediction Accuracy
160	89.522
180	90.253
200	92.933
250	94.736
300	93.664

It is clear from the code that using the SVM classifier allows us to predict with a high degree of accuracy, whether or not a given review is helpful. With the increase in the size of the training set, the prediction accuracy of the classifier increases.



VIII. FUTURE WORK

Future work in this domain, can include the analyzing the content of the review more deeply and adding the results of the analysis to create new features which can possibly further improve the classifier. Another area of improvement, is in trying other classification methods and experiment to determine whether a better prediction rate can be achieved.

REFERENCES

[1] Wang, Hongning, Yue Lu, and Chengxiang Zhai. "Latent aspect rating analysis on review text data: a rating regression approach." *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010.

[2] Mudambi, Susan M., and David Schuff. "What makes a helpful review? A study of customer reviews on Amazon. com." *MIS quarterly* 34.1 (2010): 185-200.

[3] Ott, Myle, et al. "Finding deceptive opinion spam by any stretch of the imagination." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

[4] Jindal, Nitin, and Bing Liu. "Review spam detection." *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007.

[5] Bird, Steven. "NLTK: the natural language toolkit." *Proceedings of the COLING/ACL on Interactive presentation sessions*. Association for Computational Linguistics, 2006.

[6] Pennebaker, James W., Martha E. Francis, and Roger J. Booth. "Linguistic inquiry and word count: LIWC 2001." *Mahway: Lawrence Erlbaum Associates* 71 (2001): 2001.

[7] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis."*Foundations and trends in information retrieval* 2.1-2 (2008): 1-135.

[8] O'Connor, Peter. "User-generated content and travel: A case study on Tripadvisor. com." *Information and communication technologies in tourism 2008* (2008): 47-58.

[9] Jeacle, Ingrid, and Chris Carter. "In TripAdvisor we trust: Rankings, calculative regimes and abstract systems." *Accounting, Organizations and Society* 36.4 (2011): 293-309.

[10] Joachims, Thorsten. *Making large scale SVM learning practical*. Universität Dortmund, 1999.

[11] Taira, Hirotoshi, and Masahiko Haruno. "Feature selection in SVM text categorization." *AAAI/IAAI*. 1999.

[12] Shanahan, James G., and Norbert Roma. "Improving SVM text classification performance through threshold adjustment." *Machine Learning: ECML 2003*. Springer Berlin Heidelberg, 2003. 361-372.