Opinion Mining on Yelp Reviews

Uma Nagarsekar
Student, USC
[nagarsek@usc.edu]

Nishant Saurav
Student, USC
[saurav@usc.edu]

Siddhartha Singh Sandhu
Student, USC
[sandhus@usc.edu]

Abstract

More and more customers today rely on online reviews. But today, the amount of review information available online is way too much. Customers generally don't have time to go through all these reviews to make their decisions. Hence there is a need to provide them with a concise representation of the reviews. Our project aims to provide customers with a concise representation of yelp reviews to help them decide on a restaurant most suited for them. The major parts of the system include identifying the aspects on which decisions can be made, the opinion information in the review and the mapping between the aspects and the opinions. This report provides an approach to perform these tasks.

1 Introduction

Humans have always relied on opinions. We always consider the experiences of other people. If you want to buy a phone or go out to eat at a good restaurant, you ask your friend who recently bought a phone or a cousin who always loves to try new restaurants. With the advent of the web, now people can post their opinions online which are accessible to almost anyone who wants to read them. Today one can find people sharing their opinions as reviews. These reviews are a summary of the experience of the person with the product or the service. Today almost all retail websites like Amazon, EBay etc. allow their customers to review any product that they sell. Today there are websites only for people to share their experiences like Yelp, Rotten Tomatoes etc. These get us the opinions of not just one or two people but a much larger crowd. A survey conducted showed that 79% of consumers trust online reviews as much as personal recommendations. Sounds like our lives have become easy. Or have they? Previously, all you had to do was call a trusted friend or a cousin. Now, you have to go through hundreds of reviews to figure out how a place is. Today where people don't have time to read newspapers, do people really go through all these reviews? The survey also showed that 67% of consumers read 6 reviews or less. It would be great if people could get a gist of what all the reviews indicate. It comes out as a simple problem of performing sentiment analysis on all the reviews. Then show a rating of 5 stars to indicate the average sentiment of people. This is what most websites do. But there is more to it. Some people are interested in one aspect of the product or service. Your friend might want a phone with an amazing camera, you might want a good speaker. You might be okay with a lousy service as long as you get good food but not your friend. This is where our project comes into picture. The aim of our project is to perform aspect wise sentiment analysis or opinion mining on reviews. For the purpose of this project, we have limited our scope to restaurant reviews on Yelp.

Yelp currently lists the reviews for the businesses. It has been observed that there are at least 50 to sometimes over 1000 reviews for restaurants on Yelp. People generally have the time to read at most 5 to 10 reviews. Our project aims to get the aspects for the restaurant and then display a rating for the aspect. It will save the time that the people would otherwise spend on reading through numerous reviews. It will also give a fair idea of the restaurant as compared to just the first few reviews.

A lot of challenges come into focus when we looked into the problem. The first thing we should be able to do is to identify the aspect. The major challenge here is to identify that different words might refer to the same aspect and club them together as one. The next thing was to identify the sentiment for a specific aspect. People may mention multiple aspects in a single sentence. Getting out the sentiments and correctly relating it to the aspect was challenging. One more problem that we faced was lack of training data. Yelp provides an Academic dataset which we were unable to use for multiple reasons. It had a small set of restaurant reviews and less number of reviews per restaurant. Hence we had to get our own data by scraping the yelp website.

2 Data

Yelp provides students with an Academic Dataset to allow to perform analysis on it. We initially considered using this dataset. Soon we realized that the dataset had few entries for restaurants and very few reviews per restaurant. Since for our approach, the more data we have - the better, we scraped the Yelp website to get the reviews. For the purpose of this project, we chose 17 restaurants spread over a variety of cuisines - Chinese, Indian, French, Egyptian, Japanese, and Italian etc. Each restaurant had about a minimum of 353 to a maximum of 4270 reviews per restaurant. And we have a total of about 29714 reviews. The approach we have used can be scaled with data i.e. it can be applied to less or more data. It only works best with more data.

3 Related Work

A lot of work has been previously done on aspect based opinion mining. This has been mostly centered on product reviews. The task is comparatively easier for a product review. The aspects defining the product are mostly well defined. For example, for a laptop the aspects generally are - resolution, battery life, value, sound, ease of use etc. These rarely change and are almost always used in the same sense. When it comes to food reviews, the aspects are less defined, vary from restaurant to restaurant, depend on the cuisine, type of restaurant etc. The opinion words describing the aspect also are more varied as compared to that of products. Also, the language used in food reviews is more informal, unstructured and erroneous. We have referred some papers defining the task of opinion mining and collaborated their approaches in our technical approach.

We have referred the book "Sentiment Analysis and Opinion mining" by Bing Liu. It gives a general description of the task of Sentiment Analysis and Opinion Mining. The book gives a general overview of various techniques that can be applied to solving the problem of aspect based sentiment analysis. Some of the methods discussed involve Finding Frequent Nouns and Noun, Using Opinion and Target Relations, Using Supervised Learning, Using Topic Models etc. This gave us the overall understanding of our problem. Due to absence of availability of annotated data, we considered the options of Finding Frequent Nouns and Noun, Using Opinion and Target Relations.

Another interesting paper that we referred was "Mining and Summarizing Customer Reviews" by Minqing Hu and Bing Liu. In this paper, they have performed aspect wise sentiment classification on reviews of 5 products from Amazon. Their approach was to use Part of Speech Tagging to get the most frequent nouns and use them as feature words. To get the opinion words, they consider the adjectives in the neighborhood of the feature words. The polarity of the adjective is obtained using WordNet to get synonyms and antonyms and having an initial seed list of oriented words. The major advantage of this approach is that no annotated training data is needed. The con of this technique is that the context information is lost. For sentences like "The food was tasty but the service was lousy", identifying that (food, tasty) and (service, lousy) will be difficult.

We also referred another paper "Automated Summarization of Restaurant Reviews" by Pawar and Mallya. In this paper they have performed aspect based sentiment classification on restaurant reviews from a website we8there.com. They have also considered followed the method of using POS to find the most frequent noun words and bigram words as the features. They have used dependency parsing to obtain the relationships between the feature word and the opinion word. The polarity of the opinion word is determined from WordNet. This paper overcomes the issues of the previous paper by considering the context information. Some of the cons of the approach are that they are considering only three dependencies in their algorithm, they are performing the the analysis on per line basis rather than per review and that they have considered the words appearing in bigrams as independent nouns also. We have tended to these problems in our implementation.

4 Technical Approach

We divided the task in 6 different modules as shown in Figure 1.

4.1 Getting Reviews and Pre-processing

We collected data as discussed earlier and then cleaned it by removing unwanted symbols like smileys, extended exclamation (E.g. !!!! , -----) and tokenized into sentences. We also keep track of which sentence belongs to which review. This association helps us in taking counts of feature words at review level. We write everything in a text file per restaurant which is passed to next module.

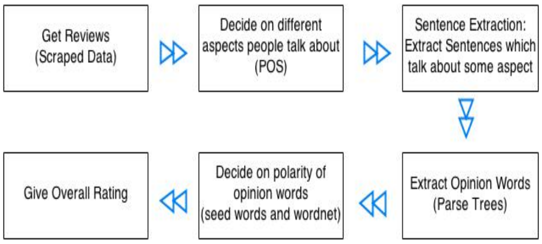

Figure 1. Approach

4.2 Getting aspects people talk about

To get the aspects people talk about we decided to look at all the nouns and noun phrase (bigrams) in review sentences as mostly all aspects are nouns. For example: # give an example sentence. We have created one text file per restaurants of review sentences. We do POS tagging on review sentences using NLTK POS Tagger and looked for NN, NNP, NNPS, NNS tag. We keep track of counts of all aspects which gave us a new file per restaurant, recording counts of aspect seen in review text of that restaurant. But on inspecting this we realized the count is very noisy for example nouns like boyfriend, guy, los angles were having very high count but these words don't tell us anything about the restaurants. We took few corrective measure which gave us improved result.

- We removed geographical words like Los Angeles.
- We took count of all aspects at review level. For example if word boyfriend was occurring 8 times in a review we considered it only once.
- We considered a noun only if it's not a subset of a bigram. For example if we find a bigram Caesar/NN salad/NN in a sentence we don't consider Caesar and salad separately as an aspect.

Even after these processing steps, count for words like boyfriend, guy were quite high and generally people talk about specific food items for example while inspecting pizza restaurant we found that counts for words like pizza, cheese, soda, crust were also very high. Because of this, count of other important aspects like parking, atmosphere were pushed behind. To correct this, for each aspect we counted the number of restaurants it occurs in. From this we made a list of few aspects like parking, crowd, staff which are very important but get suppressed because of other irrelevant high frequency words. Along with this list we picked out frequent aspects from our count list, for each restaurants and used these aspects in sentence extraction module.

4.3 Sentence Extraction

This module get a list of important aspects and iterate over a file which has review sentences, extracting only those sentences which contain important aspects and has at least an adjective. For example – "Our friend recommended this place so we went there for my cousin's bday party". This sentence will not be extracted as it doesn't contain any aspect.

4.4 Extracting Opinion Words

We have used Stanford Dependency Parser to get dependency relation from pruned sentences. We considered only nsubj, amod, neg, advmod, conj_and, conj_but dependency. This gives us a list of opinion words associated with a feature, which is written to a file. We have one such file per restaurant. We have used negation to keep note of opinions which have occurred in negative sense. For example – "Great food but lousy service" gave following dependency.

Amod(food – 2, great -1)
root (Root – 0 , food- 2)
prep(food – 2, but – 3)
amod (service-5 , lousy -4)
pobj(but- 3, service- 5)

Dependencies like amod gives us opinion words related to aspects – food and service.

4.5 Deciding polarity of given word

We create a list of seed words (by taking frequently occurring adjectives in review data set) and assigned weight (0 - 5) to it manually. Importantly we don't assign any opinion word 0 or 5 value it's between 1 and 4. For example for good we assigned weight 3, bad we assigned polarity 1. To handle cases like very good service and very bad service we looked at advmod relation in dependency tree and stored opinion word with +/- sign. In this case for service we write a line as service +good -bad. For phrases like not good service/ not bad service we store it as service -good +bad

After this we expanded our database of positive and negative sentiment words by using WordNet SynSet to get the synonyms and giving the same rating as seed word. While looking for synonyms if a word that's already there in our seed word we ignore it and if a word is encountered more than once as it might me in synonym list of two or more different seed word we just give it the max rating.

4.6 Giving Overall Rating

We calculated overall sentiment to each aspect by averaging the sentiment for that aspect over the review it occurs in. If an opinion word has + or – sign associated with it we add/subtract one from it. For Example if service has opinion words bad in 2 reviews and good in 1 reviews then overall sentiment for service will be 1*2 + 3*1 / (2 + 1) = 1.6 . We map 1.6 to 1.5. Since we get some real number between 0-5 after our calculation we have mapped overall sentiment output for example 1-1.25 mapped as 1 , 1.26- 1.5 mapped as 1.5 , 1.51-1.75 mapped as 1.5 and 1.76- 2 to 2 for each integer from 0-5.

5 Evaluation

Due to absence of any annotated data or baseline rating, we had to perform the evaluation of the system manually. We used unpruned review set (sentence set before sentence extraction phase) to do our evaluation. Initially we decided on giving this to our friends and ask them to give ratings on few chosen aspects. But we couldn't come out with a good evaluation criteria to check the model. Some friend might give 4 star for food and we might have given 3.5 star to it. This made evaluation difficult for us so we decided to just categorize aspects as positive and negative and do the evaluation. Due to time constraints, we evaluated the system for 4 restaurants for a set of 200 reviews. We calculated the accuracy for 4 major aspects widely occurring in all the 4 restaurant files namely - food, service, atmosphere and parking.

Aspects	R 1	R 2	R 3	R 4
Food	47.35	46.15	45.43	46.56
Service	45.23	42.12	44.72	45.15
Atmosphere	42.65	41.23	40.12	41.57
Parking	40.24	38.89	35.42	37.69

Table 1. Evaluation Results

One reason for low accuracy might be the set of seed words we stated with. Because of this some of the adjective are not considered as we lose information. For example "*sunny brunch atmosphere we were craving for* ". We didn't have sunny in our seed word list and even after expanding using synonyms, "*sunny*" didn't appear in opinion word database.

6 Conclusion

Opinion mining especially on restaurant reviews is still a research area in progress. Accuracy of the most of techniques to some extent depends on manual intervention. In our technique we decided on taking frequent nouns as aspect which didn't give encouraging result because of noise in the data and partly attributed to the pronoun resolution which we haven't considered. Sometimes people express sarcasm and complex emotions which are hard to capture. In our current method we are not able to get information from sentences like '*atmosphere just killed it'. 'I will rate it 4/5', 'We will surely be back again'* which have information about different aspects. Starting with a better set of seed opinion words might have increased accuracy. We observed that since we decided on doing manual evaluation it is better to just categorize aspect as positive and negative rather than giving exact ratings between 0-5.

In future we can try topic modelling techniques like LDA and use pronoun resolution to get better set of features. We also observed that people in general talk about specific food items a lot rather than food in general. This give us interesting information about the dishes that are good or bad. We can mine this data to suggest specific food items for restaurants.

Acknowledgements

We would like to thank Kenji Sagae (USC) and Xing Shi (USC) for their support and guidance.

Reference

Liu, B. 2012. *Sentiment analysis and opinion mining*, Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers

Minqing Hu. Bing Liu. 2004. Mining and summarizing customer reviews, Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA

Manoj Pawar. Deepak Mallya. 2010. *Automated Summarization of Restaurant Reviews*

Marie-Catherine de Marneffe. Christopher D. Manning. 2008. *Stanford Dependencies manual.*

Myles Anderseon. 2013. *Study: 79% Of Consumers Trust Online Reviews As Much As Personal Recommendations*