11-2015

Analysis of Aspects and Star Ratings in Consumer Reviews

Maruthi PRITHIVIRAJAN
Singapore Management University

Vivien LAI
Singapore Management University

Kyong Jin SHIM
Singapore Management University, kjshim@smu.edu.sg

Citation

PRITHIVIRAJAN, Maruthi; LAI, Vivien; and SHIM, Kyong Jin. Analysis of Aspects and Star Ratings in Consumer Reviews. (2015). *ISAAC 2015: International Symposium on Advanced and Applied Convergence: November 12-15, Seoul: Proceedings*. 32-38. Research Collection School Of Information Systems.
Available at: http://ink.library.smu.edu.sg/sis_research/3151

Analysis of Aspects and Star Ratings in Consumer Reviews

Maruthi Prithivirajan, Vivian Lai, Kyong Jin Shim

School of Information Systems, Singapore Management University, Singapore
maruthiprithivi@gmail.com, vivian.lai.2011@sis.smu.edu.sg, kjshim@smu.edu.sg

Abstract

This paper presents an analysis of star ratings in consumer reviews in Yelp, an online social platform for sharing consumer reviews about local businesses. In particular, we analyze consumer reviews about food businesses. We analyze how well or poorly the star ratings (on a scale of one star to five stars) associated with these reviews tally with the sentiment derived from the textual portion of the consumer review.

Keywords: *Consumer reviews, star rating, text sentiment, Yelp*

1. Introduction

The food service industry is one of the most thriving and competitive industries today. The widespread popularity of social review sharing platforms such as Yelp has created a new generation of opportunities as well as problems for businesses in the food service industry. Most of these platforms are free and listing f one's business can be done in just a few clicks. Thus, it makes it extremely easy for low budget local food businesses to create social media presence with low to no marketing budget. The inherent social features on these platforms (such as sharing reviews, complimenting, sending messages, and following other users) naturally create a viral marketing channel via which news about one's food business can spread in the vast social network of users.

While such benefits help food businesses to get online at almost no cost and increase their visibility, having a social media presence can lead to a destructive path. While good news travels fast, it is said that bad news travels faster. Negative reviews (whose authenticity sometimes cannot be easily verified) can be detrimental when it goes viral and it is not mitigated immediately and appropriately. Additionally, in the presence of a large number of reviews, platforms such as Yelp attempt to "filter" consumer reviews. Because it is highly unlikely that a typical user will want to read through hundreds and thousands of consumer reviews, such platforms attempt to pre-filter without the user's input and show "top" reviews in terms of popularity and relevance. However, when such filtering is not done properly [1] and thus creating an imbalance between positive reviews and negative reviews presented to the site's visitors, businesses can suffer.

Hence, as a business, having an online presence in social review sharing platforms can be risky. Thus, businesses must understand pros and cons of popular social review sharing platforms before building a social media presence. Nonetheless, an increasing number of businesses are jumping onto the social media wagon

Manuscript received: 10/13/2015 / revised: 10/13/2015
Corresponding Author: kjshim@smu.edu.sg
Tel:+65-6808-5275
Singapore Management University

J.J. Kang et al. (Eds.): ISAAC 2015, AACL 06, pp. 32~38, 2015

with a hope that good news will spread fast and more money will roll in with more customers leaving good reviews. It has become imperative for any business to stay on top by listening to social conversations and monitor the public sentiment.

2. This Research

This study presents an analysis of star ratings in consumer reviews in Yelp. Specifically, we investigate reviews about food businesses (restaurants). One main aspect of Yelp consumer reviews that we explore is that of the star rating – in particular, we explore the question of whether the star rating (on a scale of one star to five stars) tallies with the sentiment derived from the textual portion of the consumer review (referred to as “review comment”). As shown in Figure 1, the star rating is the most prominent item in a consumer review with bright orange squares at the top. It is often seen before the textual portion of the consumer review is seen and read by the site’s visitors.

Figure 1. A consumer review on Yelp

Given its importance as the first-seen item in a consumer review, it is important that the star rating correctly reflects the consumer’s sentiment as reflected in the review comment. The review comment can contain multiple opinions each addressing a different “aspect” of a consumer’s experience. The different “aspects” can be inferred from a large number of consumer reviews [2], and a sentiment score can be computed for each aspect addressed in a consumer review. For instance, a consumer may have had a positive experience in terms of food (e.g. “The food was very delicious!”) but a negative one in terms of ambience (e.g. “The dining table was too small and the corridor leading up to the bathroom was too dark and noisy!”). In such cases of “conflicting” opinions in terms of sentiment polarity, is the overall star rating reflective of both opinions (e.g. average of the two sentiment scores)? Or, does one of the aspects more strongly influence the overall star rating? This study investigates how well or poorly the star rating aligns with the review comment. Additionally, our research aims to create a framework for identifying one or more “aspects” that strongly influence the overall star rating.

3. Dataset

Yelp has released their massive data of consumer reviews over the years. The dataset used in the present research is from Yelp's Data Challenge held in the year 2014 [3]. The dataset consists of five separate JSON formatted files. They contain data about businesses, user reviews, users, check-ins and tips. The data provided in JSON format were transformed into a transactional format. The tools used for this processing step are: Python 2.7, Microsoft Access, R, and Amazon Mechanical Turk. Food business specific review data were extracted from the "business" dataset by filtering on "categories" column. The extracted food businesses were then mapped to the "reviews" dataset based on the 'business ids" column. After the mapping, our dataset consists of 19 variables and 697,076 reviews. One of the 19 variables is the star rating.

4. Experiments & Results

4.1 Topic Modeling

We performed a manual inspection on 50,000 review comments to investigate generally what were the topics that were frequently mentioned in the reviews. We observed that there were largely four "aspects" addressed in food business reviews: "food", "service", "ambience" and "price". We performed topic modeling using LDA (Latent Dirichlet Allocation) to extract 10 topics with 10 terms from the review comments. The results are shown in Figure 2. The general idea is that similar "terms" are grouped into the same "aspect". For instance, dish or food names will be grouped into the "food" aspect, and frequently co-occurring adjectives or nouns (e.g. delicious, yummy, delicacy) will be grouped into the "food" aspect along with the dish/food names. We observed that the review comments pertained to "food" (e.g. chicken, eat, delicious), "service" (e.g. service, order, time), "ambience" (e.g. place, table), and "price" aspects.

Topic.1	Topic.2	Topic.3	Topic.4	Topic.5	Topic.6	Topic.7	Topic.8	Topic.9	Topic.10
great	great	food	**food**	**food**	good	**food**	**food**	**food**	great
ordered	good	place	good	good	great	great	**service**	just	**service**
maggianos	**place**	**service**	great	**service**	**food**	good	good	**time**	**food**
chicken	**table**	one	maggianos	place	maggianos	really	just	great	**place**
just	also	ive	time	maggianos	**place**	back	like	**place**	**chicken**
restaurant	well	good	really	just	**time**	**delicious**	back	good	like
said	**chicken**	**table**	**delicious**	like	**table**	nice	best	**ordered**	**time**
one	**order**	also	place	**restaurant**	get	**menu**	get	one	**menu**
place	lunch	back	service	back	**order**	much	nice	get	much
well	menu	**eat**	ordered	really	restaurant	**italian**	one	didnt	get

Figure 2. Topic Modeling

4.2 Aspect Based Sentiment Scoring

Amazon Mechanical Turk is a crowd sourcing market place on the Internet. Amazon is one of the many service providers with a huge pool of Turks (workers). Certain measures (e.g. minimum education qualifications of the workers) are in place to ensure the quality of the work output. Tasks are generally referred to as HITs (Human Intelligence Tasks). Four separate sentiment scoring HITs were created, one for each aspect ("Food", "Service", "Ambience", "Price"). In each HIT, five workers were tasked to 1) identify and indicate whether a particular aspect is addressed in the review comment and 2) to assign a sentiment score on a scale of very negative to very positive (see Table 1).

Each and every review comment was given a sentiment score by each of the five workers. The final sentiment score for each review comment is simply the average of the five sentiment scores from the five workers. Each worker assigned a sentiment score to the review on a scale of -2 to 2 where:

Table 1. Sentiment Scores

Score	Meaning
-2	Very negative
-1	Negative
0	Neutral
1	Positive
2	Very positive

Instructions on how the sentiment scores should be assigned were clearly stated to all of the Amazon Mechanical Turk workers. The sentiment scores were logged, and the average sentiment scores were stored in the following way:

Table 2. Sentiment Scoring Example

HITID	Review	Worker1	Score - 1	Worker2	Score - 2	Worker3	Score - 3	Worker4	Score - 4	Worker5	Score - 5	Avg Score
36FFX...	*MEN looking to take your gf or soon to be gf or wife...	A177E....	1	A1T7....	1	A1Y7....	2	A2HM....	2	A3QZ....	-1	1
3R15...	2 1/2 more like it. They put lots of food on the plates...	A177E....	-1	A1T7....	1	A312....	-1	A3HB....	0	ALLP....	0	-0.2
....	....	....	....	....	....	....	....	....	....	....	....	....

To align this scoring scheme (-2, -1, 0, 1, 2) with Yelp's star rating (1, 2, 3, 4, 5), we re-coded the scores from Amazon Mechanical Turk by adding three (e.g. -2 becomes 1, -1 becomes 2, 0 becomes 3, 1 becomes 4, and 2 becomes 5). This way, we can perform a one-to-one comparison between the two platforms.

4.3 Data Exploration

In this section, we report our findings from data exploration. In our dataset, a majority of the consumer reviews are about food businesses specializing in Italian cuisine, followed by Mexican cuisine and so on.

Table 3. Distribution of Food Business Reviews

Restaurant Categories	No. of Reviews
American (Traditional)	70
Bakeries Food Bf & Br	8
Bf & Br Cafes American (Traditional)	74
Bf & Br Fast Food	8
Bf & Br	4
Burgers Fast Food	7
Chinese	52
Fast Food	8
Food Coffee & Tea Internet Cafes Bf & Br	2
Greek	39
Mexican	139
Modern European	3
Pizza	10
Pubs Bf & Br Bars Nightlife	5
Restaurants	3
Italian	568

Further, we segmented out dataset by geographical areas (e.g. different states in the United States) and also by cuisine types (e.g. Italian, Mexican, etc.). As shown in Figure 3, across all of the consumer reviews over all of the cuisine types, 68% of them are positive with star ratings of 4 and 5. In terms of geographical areas, a majority of the reviews are about food businesses in Nevada (76%) followed by Arizona (21%). Figure 4 shows the distribution of sentiment scores (via Amazon Mechanical Turk) by aspects. The overall sentiment is generally neutral or positive.

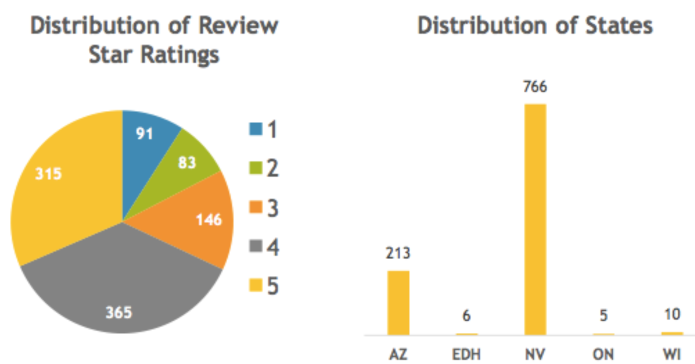

Figure 3. Distribution of Consumer Reviews by Star Rating and by Location

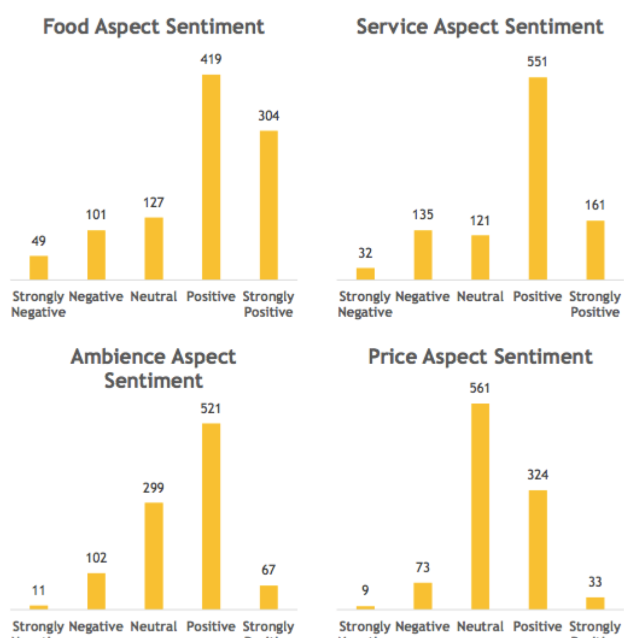

Figure 4. Distribution of Consumer Reviews by Aspects

Next, we investigated the relationship between the overall star rating and aspect-specific sentiment scores. We used partial dependence plots to observe the dependence of the dependent feature (star ratings) with the independent feature (aspect-specific sentiment score). We generated the plots using black box algorithms.

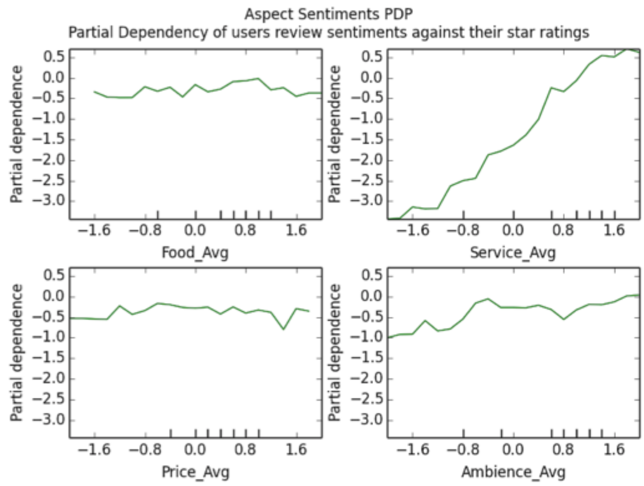

Figure 5. Partial Dependence Plot (all cuisine types & all geographical areas)

As shown in Figure 5, across all cuisine types and all geographical areas, the "service" aspect shows the most noticeable alignment with the overall star rating, more significantly than the other three aspects ("food", "price" and "ambience"). This means that consumer sentiment about the food business's service has the strongest correlation with the overall star rating of the consumer review.

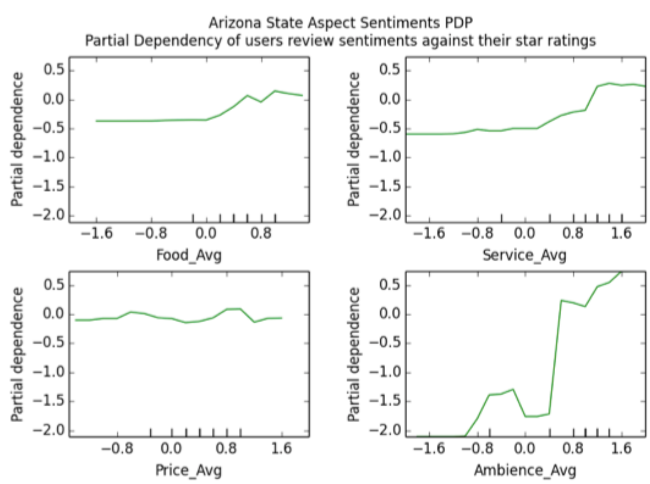

Figure 6. Partial Dependence Plot (all cuisine types & Arizona)

We observe a slightly different pattern in the state of Arizona. Figure 6 shows that "ambience" is seen to have the strongest correlation with the overall star rating.

Figure 7. Partial Dependence Plot (American cuisine & all geographical areas)

Consumer reviews about food businesses specializing in American cuisine in all geographical areas show a yet another pattern (Figure 7). Both “food” aspect and “ambience” aspect have stronger correlation with the overall star rating than the other two aspects.

5. Conclusion & Future Directions

Our findings show that not all but some of the aspects appear to have strong correlation with the overall star rating of the consumer review. Given ample digital data available from online social review sharing platforms, food businesses can today better understand what aspects of their products and services win or lose consumers’ hearts. Future directions include expanding our analysis to the entire Yelp dataset to draw more comprehensive conclusions with regards to the different aspects and their relationship with the overall star rating. We also plan to automate sentiment classification by using machine learning algorithms while still cross checking with manual (human labeled) annotation to ensure high sentiment classification accuracy.

6. Reference

[1] “Yelp -- Is It Hurting Small Business?” *http://www.huffingtonpost.com/anne-phillips/yelp-is-it-hurting_b_4428490.html*

[2] “Yelp Data Challenge”. *https://www.yelp.com/dataset_challenge/dataset*

[3] S. Kiritchenko, X. Zhu, C. Cherry, S. M. Mohammad. “NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews,” Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 437-442, Dublin, Ireland, August 23-24, 2014.