

Restaurant Rating: Industrial Standard and Word-of-Mouth A Text Mining and Multi-dimensional Sentiment Analysis

Qiwei Gan

Edinboro University of Pennsylvania
ggan@edinboro.edu

Yang Yu

Rochester Institute of Technology
yyu@saunders.rit.edu

Abstract

AAA Restaurant Diamond Rating Guidelines (which is regarded as industry standards) rate a restaurant in three aspects: food, service, and décor/ambience. Drawing upon extant literature, we argue that special contexts and pricing are two other major aspects in restaurant rating in addition to aforementioned three aspects. We tested our hypotheses based on our text mining and sentiment analysis of 268, 442 customer reviews of 7,508 restaurants on Yelp.com, a form of digital word-of-mouth. Results from fitting a multilevel model showed that the sentiments about each of these five aspects alone explained about 28% of the explainable between-restaurant variances, and 12% of the explainable within-restaurant variances of the restaurants' star ratings. With other level and control variables, the multilevel model can explain more than 53% between-restaurant variances and 28% within-restaurant variances.

1. Introduction

Dinners have been using various sources to find a "perfect" restaurant to meet their dining needs, such as restaurant ratings on magazines or other publications, TV ads, recommendations from friends, etc. On the other hand, restaurants have been using these same sources to reach diners and try to win their business. Among these sources, AAA (America Auto Association) has been publishing its *Diamond Rating Guidelines* for restaurants for decades, which is regarded as industrial standards to evaluate restaurant quality [1]. Meanwhile, word of mouth has been another important source for diners and restaurants. Over recent years, online reviews on restaurants, a digital form of word of mouth [2], become prominent for both diners to select restaurants and for restaurants to improve their offerings. Many studies have found that online reviews affect retail sales and customer purchasing decisions [3, 4]. Therefore, it is interesting to link

these two important sources and compare the industrial standards and what customers are really caring about. However, to our knowledge, there is no much literature about this. The purpose of this paper is to bridge the industrial standards and the word of mouth of rating restaurants.

AAA Diamond Rating Guidelines for restaurants is one of the most widely known among many professional ratings/reviews on restaurants, and has been used in many studies [5-8]. The AAA guidelines rate restaurants based on three aspects: food quality, service quality, and décor/ambience quality. However, AAA and most other professional ratings/reviews focus more (or solely) on quality of restaurants and less on customer experiences. The online customer reviews, on the other hand, focus more on many normal customers' experiences (rather than inspectors). Some restaurants on Yelp.com, a popular online customer review website, have hundreds of customer reviews. These many customer reviews provide a good venue to explore what customers are really caring for and value when they select dining in restaurants. In addition to the aforementioned three aspects (food, service, and décor/ambience) in AAA guidelines, we argue in this paper that two other aspects, pricing and special contexts, are also important in rating restaurants among customers, based on extant literature and reading on many actually reviews. Pricing refers to customers' evaluation on the prices of dining in a restaurant as compared to the services they actually received. In other words, it refers to the degree whether the customer thinks dining in a restaurant worth the money he/she paid. Special contexts refer to whether a restaurant is good for a special purpose, e.g., romantic, family, birthday, etc.

We employ text mining and multi-dimensional sentiment analysis techniques to analyze online customer reviews texts on restaurants to test our hypotheses. Many studies have formally incorporated and tested the impact of the textual content of online user reviews [4, 11-13]. However, only a few

considers multi-dimensional features of online user reviews [3], even less on sentiments of multi-dimensional features. Our method first extracts the five aforementioned aspects (dimensions) from review texts and then analyzes the customers' sentiments about them. Our data consist of 268, 442 customer reviews on 7, 508 restaurants on Yelp.com and we fit a multi-level model [14] to test our hypotheses. Most analysis on online reviews in the extant literature is at review level and does not consider the clustering of reviews under a certain business (e.g. restaurant). In the sense that many reviews are about a same restaurant, these reviews are not independent. Therefore, some statistical testing methods (e.g. ordinary least square regression) may produce biased results. Multi-level models, on the other hand, consider both between and within group variances [14] and more appropriate in the context of online reviews on restaurant. Our results show that the sentiments about each of these five aspects alone explained about 28% of the explainable between-restaurant variances, and 12% of the explainable within-restaurant variances of the restaurants' star ratings. With other level and control variables, the multilevel model can explain more than 53% between-restaurant variances and 28% within-restaurant variances.

The paper proceeds as follows. We first review the literature on restaurant ratings, especially the AAA Diamond Rating Guidelines and then on text mining and sentiment analysis on online reviews. Next, we develop our hypotheses based on the literature. Then, we describe the methods used in this paper, including data collection, text processing, sentiment analysis, and a multilevel statistical model to test our hypotheses. Results are presented next and discussed. We conclude the paper by discussing the implications, limitations, and future work.

2. Literature review

2.1. Rating restaurants

The hospitality industry has the tradition of evaluating service quality of businesses. Researchers and practitioners have been using professional ratings on restaurants, such as AAA Diamond Rating Guidelines [1], Michelin Red Guide [17], Mobil Travel Guide [18, 19]. These professional ratings focus on restaurants and provide detailed criteria to evaluate many aspects of restaurants besides service quality. AAA Diamond Guidelines classify restaurants with its diamond system, with one

diamond as basic, to five diamond as world-class dining. Due to its widely usage and consistent criteria, AAA Diamond Guidelines have been regarded one of the industry standards.

Specifically, the AAA Diamond guidelines focus on three aspects of restaurants: food quality, service quality, and décor/ambience quality. It has very detailed criteria for a restaurant to be rated in each diamond level for these three aspects. For example, for food quality, it specifies criteria of evaluating presentation of food, ingredients, and menu preparation for each diamond level. Although AAA claims that the diamond ratings represent the combination of these three aspects, it seems that the ratings still neglect some important aspects mentioned in other literature. Drawing up the extant literature, pricing of restaurants has been highlighted, and found to affect sales, customer satisfaction, and restaurant ratings [18, 20, 21]. The pricing is different from prices which are measured by ranges of dollar amounts, for example, one can classify average price below \$20 as inexpensive, between \$20 and \$35 as medium, and above \$35 as expensive. The pricing here refers to the degree on whether the customer thinks dining in a restaurant worth the money he/she paid. Meanwhile, many customers dine out with purposes, either because of convenience, a special occasion (birthday, graduation, dating, etc.), or other purposes. Many popular review website (e.g. Tripadvisor.com and Yelp.com) provide such information as "a place good for..." This type of information is valuable for customers to select restaurants. We characterize this type of information as special contexts. However, to our knowledge, no prior literature considers this type of information in rating restaurant. We believe it is necessary to include it in this paper.

In addition, AAA ratings are based on its own inspectors' evaluations without considering the feedbacks from general customers. The professional inspectors do have experiences and knowledge of evaluating restaurants. However, only one or a few inspectors rate a certain restaurant at a single or a few time points. Online customer reviews, on the other hand, consist of many customers' feedback about a same restaurant at many different time points. Although one customer may focus only on a few aspects, the aggregation of many customer reviews on a same restaurant may represent a more complete picture of the restaurant, and thus is complementary to the professional ratings/reviews. To bridge the professional ratings/reviews and online customer reviews, we investigate the three aspects (food,

service, and décor/ambience) of restaurants from professional ratings and two additional aspects (pricing and special contexts) in the context of online customer reviews. Next we will explain how these five aspects are explored in online customer reviews using text mining and sentiment analysis techniques.

2.2. Sentiment analysis on online review

Sentiment analysis is the computational detection and study of opinions, sentiments, emotions, and subjectivities in text [24, 25]. In plain language, sentiment refers to the attitudes towards something, negative, neutral, or positive. As mentioned in the previous section, we investigate the five aspects of restaurants in online user reviews. The sentiments about these five aspects reflect customers' post-purchase experiences. For example, that a customer mentioned that "this restaurant is too pricy" (the price exceeds expectation) reflects his/her negative sentiments about the pricing of the restaurant.

Over recent decade, sentiment analysis and related approaches has gained great popularity in both research and practice, due to the advancement of machine learning methods in natural language processing and information retrieval, the increasing availability of large datasets, and the realization of many commercial applications [26-28]. Most the sentiment analysis studies in the extant literature focus on the overall sentiment of customer about a product or service. However, more recent development in sentiment analysis is to investigate multi dimensions of the sentiments. This makes sense that one may feel positive to the food quality but negative to the service quality of a restaurant. We employ this approach in this paper. To accomplish this goal, the sentiment analysis technique is often split into two consecutive tasks: detecting which text segments (e.g., sentences) contain the dimensions (the five aspects of restaurant in this paper), and determine the polarity and strength of the sentiment of each of these dimensions [26]. We will detail our approach in the method section.

3. Hypotheses

Online user review website is a popular outlet for customers to share their post-purchase experiences. They express their sentiments about various aspects of a business (e.g. restaurant). The more positive sentiment to a restaurant, the more satisfied a customer is, and the higher the customer will rate the restaurant. On the other hand, the more negative

sentiment to a restaurant, the less satisfied a customer is, and the lower the customer will rate the restaurant. In addition, the sentiment to a restaurant is based on the following five aspects: food, service, décor/ambience, pricing, and special contexts. Therefore, we hypothesize the following.

H1: The more positive a customer's sentiment about the food of a restaurant, the higher the customer rates the restaurant.

H2: The more positive a customer's sentiment about the service of a restaurant, the higher the customer rates the restaurant.

H3: The more positive a customer's sentiment about the décor/ambience of a restaurant, the higher the customer rates the restaurant.

H4: The more positive a customer's sentiment about the pricing of a restaurant, the higher the customer rates the restaurant.

H5: The more positive a customer's sentiment about the special contexts of a restaurant, the higher the customer rates the restaurant.

4. Methods

4.1. Data Collection

We test our hypotheses on the data from the Yelp Dataset Challenge, which is provided by Yelp, Inc. This dataset consists of 335,022 consumer review texts on 15,585 businesses throughout the city of Phoenix AZ. The comprehensive dataset includes the reviewer's explicit feelings specified by a rating from 1 (negative) to 5 (positive) stars about the business reviewed, the date that the review has been posted, the detailed information about the reviewer (user name, elite user status, fans, etc.), and the detailed information about the business being reviewed (name, address, category, hours, etc.). In addition, each review was subject to being labeled useful, funny, or cool by other reviewers.

The full Yelp dataset contains over 300,000 reviews about various businesses, e.g. restaurants, hotels, financial services, etc. To increase the reliability of our information measures, we filtered this dataset by choosing 268, 442 reviews on 7,508 restaurants in several major categories (Coffee & Tea, Italian, American, Bars, Fast Food, Mexican, and Chinese) and also at least 100 words in each

review. For each review, besides the review content, we also have the details information of this review, such as other consumers' votes, the reviewer's id, review post time and the target restaurant id. The Yelp dataset is informative due to its complete supportive information. For example, restaurant id could be matched with business data set to get the restaurants-related information, including the physical attributes such as address, city, neighborhood, longitude, latitude, etc., and also review related attributes such as the number of reviews the restaurant has been received. In addition, the reviewer's id could be matched with reviewer data set to get the review poster's related information, such as the overall votes the reviewer has been received, the name, the elite status of this reviewer (yes or no), how many fans she/he has and the friends list of her/him, etc.

4.2 Sentiment Analysis

We perform sentiment analysis in the following steps. To recognize the five aspects of restaurant hidden in the review content, we designed a semi-supervised machine learning approach, which includes two major tasks, topic detection and sentiment classification. With this approach, we could detect the five aspects of restaurant in user reviews that are highly correlated with the positive and negative opinions, which could help us understand both the overall sentiment scope as well as the drivers behind the sentiment. In the first task of detecting the five aspects of restaurants, we followed the idea of McCallum and Nigam [29]. We start working with a set of keywords that are known to cue for certain aspect and then use those keywords to bootstrap a classifier within a naïve Bayes framework.

First, to create reliable word list for five aspects (food, service, décor/ambience, pricing, and special contexts), we combined multiple sources. We randomly select 2000 reviews, invited three experts who are familiar with restaurant industry to read and manually pickup words/terms related to five aspects. We then pickup high frequency nouns from AAA Restaurant Diamond Rating Guidelines and add them into the previous generated word lists after the experts tagged them to each of the five aspect.

Second, we randomly select 1000 reviews (6153 sentences). By using the five words lists of the five aspects, we count the frequency of words which fall into a list then assign each sentence one of five aspects. We use the highest frequency list name as

the aspect and assign a sentence as "others" if there is no single word fall into any one of those five lists.

Third, to help determine the polarity direction in some terms of the text, we use AFINN sentiment lexicons, and another extended hand-built list with 68 words. AFINN is a sentiment lexicon containing English words manually labeled by Finn Arup Nielsen [30]. Words were rated between minus five (negative) and plus five (positive). This lexicon is based on the Affective Norms for English Words lexicon (ANEW) proposed by Bradley and Lang. However, AFINN is more focused on the language used in microblogging platforms. The word list includes slang and obscene words as also acronyms and web jargon. In addition, in order to consider the specific context (restaurant reviews), we extended this 2,477 English words lexicon by adding 68 words manually picked from sample reviews. These words are specific to the context and not included in the AFINN list. For example, "delicious" is not in the AFINN list, however, it conveys a positive sentiment in the context of restaurant. We then calculate sentiment scores for the five aspects based on the word/lexicon lists.

We also consider emotional signals besides explicit sentiment lexicons. Our classification will only work on content in English because both our words list and training data is English-Only. There are multiple emoticons that can express positive emotion and negative emotion. For example, ":)" and ":-)" both express positive emotions, and ":(" express negative emotions. We also consider the emoticons in our extended list. The full list of emoticons can be found in Table 1.

Table 1. Emotion Mapping	
Emoticons mapped to :)	Emoticons mapped to :(
:)	:(
:-)	:(
:)	:(
:-P	:-e
^^ or ^^ or ^^	:(
(^0^ or ^o^	X-<
=)	:[
:-D or :D or =D	>:O or >:-O or >:o or >:-o

Reviews contain very casual language. For example, some arbitrary number of letters would be shown in words such as "Awwwful", "Awfuuuul", "Ruuuude", etc., or misspellings such as "restarant" or "restaraunt". To deal with repeat letters scenario, we use preprocessing so that any letter occurring more than two times in a word is replaced with two

occurrences. In the samples above, these words would be converted into the token “awwful”, “awfuul” and “Ruude”. This preprocessing make it easier for error correction algorithms to recognize and correct. We apply error correction algorithms such as automatic repeat request (ARQ) or minimum Hamming distance on them to correct these words.

4.3. Operationalization of variables

The main focus of this paper is the sentiment about the five aspects of restaurants hidden in the review texts. The results of sentiment analysis are the five sentiment scores about the five aspects of restaurant for each review. The sentiment scores are AFINN scores weighed by the proportion of the number of sentences of an aspect, given by the following equation. The AFINN scores are the sum of the sentiment values of the words in the aspect sentences which are also in the extended AFINN word list.

$$Sentiment_{ij} = AFINN\ Score_{ij} * \frac{\# of\ sentences\ of\ Aspect_{ij}}{\sum \# of\ sentences\ of\ Aspect_{ij}}$$

$i = 1 \dots N\ review;$
 $j = food, service, decor, context, price$

The dependent variable is the restaurant star rating given by the reviewer. Weighted Sentiment scores of the five aspects are the main independent variables. In addition, we add several control variables: the number of reviews that a restaurant received, whether a restaurant is a fast-food restaurant, the number of reviews that a reviewer has

ever posted on Yelp.com, and the reviewer’s past rating behavior operationalized as the average stars of all the reviews this reviewer ever posted on Yelp.com. The reason to add “fast food” as a control variable is that it has been pointed out in many hospitality studies that fast food restaurants are slightly different from normal restaurants. The descriptive statistics of the variables are given in Table 2, which provides the mean, standard deviation, min, max, and correlations of variables.

4.4. Multi-level Modeling

We employed two-level multi-level models to test our hypotheses [14]. Our reviews data are nested within restaurants. Reviews about the same restaurants may share some common characteristics and may not be independent. If ordinary least square regression models are used to test the hypotheses, it violates the independent assumption and may provide misleading results [14].

The multi-level models take this into consideration of both between and within group variances (for more information about multi-level modeling, see [14]). In our context, our model has two levels, review level (level 1) and restaurant level (level 2), and reviews are nested within restaurant level. We fit a series of multi-level models below to justify the use of multi-level models and test our hypotheses.

Table 2. Descriptive Statistics of Variables

	Mean	SD	Min	Max	1. Star	2. Service	3. Food	4. Price	5. Décor	6. Context	7. Review Count	8. Reviewer Count	9. Fast Food	10. Behavior
1	3.77	1.22	1	5	1.00	0.21	0.30	0.15	0.04	0.13	0.13	-0.03	0.04	0.47
2	0.40	0.83	-10	31	0.21	1.00	0.15	0.15	0.05	0.17	-0.01	-0.01	0.03	0.12
3	1.07	1.27	-11	32	0.30	0.15	1.00	0.15	0.05	0.25	0.06	-0.03	-0.01	0.17
4	0.39	0.78	-15	31	0.15	0.15	0.15	1.00	0.04	0.21	0.04	0.02	-0.02	0.09
5	0.04	0.28	-6	14	0.04	0.05	0.05	0.04	1.00	0.06	-0.02	0.01	0.00	0.02
6	0.29	0.81	-15	32	0.13	0.17	0.25	0.21	0.06	1.00	0.03	0.01	-0.03	0.09
7	161.65	182.26	3	1124	0.13	-0.01	0.06	0.04	-0.02	0.03	1.00	-0.04	-0.14	0.05
8	47.31	84.92	1	721	-0.03	-0.01	-0.03	0.02	0.01	0.01	-0.04	1.00	0.04	-0.01
9	0.20	0.40	0	1	0.04	0.03	-0.01	-0.02	0.00	-0.03	-0.14	0.04	1.00	0.01
10	3.76	0.61	1	5	0.47	0.12	0.17	0.09	0.02	0.09	0.05	-0.01	0.01	1.00

Model 1

Level 1:

$$ReviewStar_{ij} = \beta_{0j} + r_{ij},$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

where $r_{ij} \sim iid N(0, \sigma^2), \mu_{0j} \sim iid N(0, \tau_{00})$

$i = 1, 2, \dots i^{th} review$

$j = 1, 2, \dots j^{th} restaurant$

Model one is an unconditional model. At level 1, we express that a review's star rating about a restaurant is the sum of an intercept (β_{0j}) for the restaurant and a random error (r_{ij}) associated with the i^{th} review for the j^{th} restaurant. At level 2, we express that the restaurant intercepts as the sum of an overall mean (γ_{00}) and a series of random deviations from that mean (μ_{0j}). This specification enables us to examine two variance components—one for the variation between restaurants (τ_{00}) and the other for the variation among reviews within the restaurant (σ^2). The unconditional model also enables us to test if the two variance components are statistically significant from zero and serves as a benchmark model when we add more variables later.

Model 2

Level 1:

$$\begin{aligned} ReviewStar_{ij} = & \beta_{0j} \\ & + \beta_{1j}Sentiment_Food_{ij} \\ & + \beta_{2j}Sentiment_Service_{ij} \\ & + \beta_{3j}Sentiment_Decor_{ij} \\ & + \beta_{4j}Sentiment_Price_{ij} \\ & + \beta_{5j}Sentiment_Context_{ij} + r_{ij} \end{aligned}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\begin{aligned} & \text{where } r_{ij} \sim iid N(0, \sigma^2), \mu_{0j} \sim iid N(0, \tau_{00}) \\ & i = 1, 2, \dots i^{th} \text{ review} \\ & j = 1, 2, \dots j^{th} \text{ restaurant} \end{aligned}$$

In model two, we add five level 1 variables to model one, that is, the reviewer's sentiments about the five aspects of the restaurant. This model enables us to test how the reviewer's sentiments alone explain the between and within restaurant variances of the star ratings. We also tested a variation of model 2, with food, service, and décor aspects added first, and then pricing and special context added later to test the difference caused by the two "new" aspects.

Model 3

Level 1:

$$\begin{aligned} ReviewStar_{ij} = & \beta_{0j} \\ & + \beta_{1j}Sentiment_Food_{ij} \\ & + \beta_{2j}Sentiment_Service_{ij} \\ & + \beta_{3j}Sentiment_Decor_{ij} \\ & + \beta_{4j}Sentiment_Price_{ij} \\ & + \beta_{5j}Sentiment_Context_{ij} \end{aligned}$$

$$\begin{aligned} & + \beta_{6j}Reviewer_Count_i \\ & + \beta_{7j}Reviewer_Behavior_i \\ & + r_{ij} \end{aligned}$$

Level 2:

$$\begin{aligned} \beta_{0j} = & \gamma_{00} \\ & + \gamma_{01}Review_Count_j \\ & + \gamma_{02}Fast_Food_j + \mu_{0j} \end{aligned}$$

$$\begin{aligned} & \text{where } r_{ij} \sim iid N(0, \sigma^2), \mu_{0j} \sim iid N(0, \tau_{00}) \\ & i = 1, 2, \dots i^{th} \text{ review} \\ & j = 1, 2, \dots j^{th} \text{ restaurant} \end{aligned}$$

In model three, we add two control variables, the number of reviews that a reviewer has ever posted on Yelp.com, and the reviewer's past rating behavior to model 2. We also add two level 2 variables, the number of reviews that a restaurant received on Yelp.com, and whether the restaurant is a fast food restaurant. As in model 2, we also tested a variation of model 3, with food, service, and décor aspects added first, and then pricing and special context added later to test the difference caused by the two "new" aspects.

5. Results

We used SAS® Proc Mixed to fit our multi-level models and the results are presented in Table 3. As shown in Table 3, both of the variance components are statistically significant from zero in the unconditional model (model 1). In addition, the between restaurant variance accounts for 21% of the total variance ($0.3274/(0.3274+1.2671)$), which indicates that there is a fair bit of clustering of star rating within restaurant, and an OLS analysis may yield misleading results [14].

In model 2, after adding reviewer's sentiments about the five aspects of the restaurant to model 1, the between restaurant variance has been reduced by 28% ($((0.3274-0.2357)/0.3274)$), and the within restaurant variance has been reduced by 12% ($((1.2671-1.1211)/1.2671)$), suggesting that the sentiments about the five aspects of the restaurant play an important role in explaining the variance of star ratings. In addition, we compared the model fit with 3 aspects (food, service, décor) and 5 aspects (food, service, décor, pricing, special context) in the model. Table 3 also shows that with pricing and special context added to the model, the model fit improved, indicating the importance of the two aspects.

In model 3, after adding level 2 variables and control variables into model 2, both between restaurant variance and within restaurant variance are further reduced from unconditional model, by 53% and 28%, respectively, suggesting that the model has a good explanation power. Similar to model 3, the model fit improved as two extra aspects (pricing and special context) are added in the model.

Table 3. Effects of reviewer's sentiments on star rating			
	Model 1	Model 2	Model 3
Variance			
<i>Between restaurant</i>	0.3274***	0.2357***	0.1554***
<i>Within restaurant</i>	1.2671***	1.1211***	0.9179***
Fixed effect			
<i>Intercept</i>	3.5671***	3.2322***	0.4582***
<i>Sentiment_Service</i>		0.2086***	0.1611***
<i>Sentiment_Food</i>		0.2192***	0.1694***
<i>Sentiment_Decor</i>		0.0344***	0.0204***
<i>Sentiment_Context</i>		0.1189***	0.0891***
<i>Sentiment_Price</i>		0.0938***	0.0738***
<i>Reviewer_Count</i>			-0.0001***
<i>Reviewer_Behavior</i>			0.7799***
<i>Review_Count</i>			0.0013***
<i>Fast_Food (0 vs 1)</i>			-0.1434***
Model Fit			
<i>AIC (5 aspects)</i>	838,018	782,652	729,374
<i>AIC (3 aspects)</i>	N/A	782,954	729,545
<i>BIC (5 aspects)</i>	838,032	782,666	729,388
<i>BIC (3 aspects)</i>	N/A	782,968	729,559

Notes: N=261, 243 reviews.

AIC = Akaike's Information Criterion; BIC = Bayesian information criterion

5 aspects=food, service, décor, pricing, and special context

3 aspects= food, service, and decor

*p < .05. **p < .01. ***p < .001.

In addition, all the coefficients of the sentiments about the five aspects are positive and statistically significant in both model 2 and model 3, supporting H1, H2, H3, H4, and H5. As the sentiments of the five aspects of restaurant are of the same scale, we can compare the relative importance of these five aspects. The sentiment about food has the largest coefficient, followed by service, special context, pricing, and décor, in descending order, suggesting their relative importance. The coefficient of number of reviews that a reviewer has ever posted (reviewer_count) is negative and statistically significant, suggesting that the more reviews a reviewer posted, the lower star she/he would rate. The coefficient of number of reviews that a restaurant has received (review_count) is positive and statistically significant, suggesting that the more reviews a restaurant receives, the higher its star rating would be. In the end, the coefficient of fast food

restaurant is negative and statistically significant, suggesting that if a restaurant is a non-fast food restaurant, its star rating would be lower as compared to a fast food restaurant, holding other things equal.

6. Discussion and future study

The results from fitting a series of multi-level models highlighted relationships between reviewers' sentiments about the five aspects of restaurants and restaurant star ratings. Customers rate a restaurant based on these five aspects: food, service, décor/ambience, pricing, and special contexts. Food, service, and décor/ambience are often used as industrial standards, while our finding suggests special contexts and pricing are two other important aspects used by customers to rate restaurants. In addition, special context and pricing are relatively more important than décor/ambience.

Our findings contribute to the existing literature in several ways. First, we found special context and pricing are two additional aspects to industrial standards that customers care for in rating restaurants. Second, our methods contribute to the existing literature in that we demonstrate how text mining and multi-dimensional sentiment analysis can be applied to hospitality research. Third, the use of multi-level models contributes to the literature of online user reviews.

Nonetheless, this paper is not without limitations. The five aspects of restaurants investigated in this paper are not the only aspects used by customers in rating restaurants. In facts, these five aspects are "theory-driven" which are derived in the existing literature. It is interesting to find out if other aspects play an important role in rating restaurants. Our future work will focus on generating a "data-driven" list of aspects to rate restaurants, which are derived by purely mining review texts.

7. References

- [1] American Automobile Association, Approval Requirements and Diamond Rating Guidelines, AAA Publishing, Heathrow, FL, 2009.
- [2] Dellarocas, C., "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms", *Management science*, 49(10), 2003, pp. 1407-1424.
- [3] Archak, N., Ghose, A., and Ipeiritos, P.G., "Deriving the Pricing Power of Product Features by

- Mining Consumer Reviews", *Management science*, 57(8), 2011, pp. 1485-1509.
- [4] Duan, W., Gu, B., and Whinston, A.B., "The Dynamics of Online Word-of-Mouth and Product Sales—an Empirical Investigation of the Movie Industry", *Journal of retailing*, 84(2), 2008, pp. 233-242.
- [5] Verma, R., Stock, D., and McCarthy, L., "Customer Preferences for Online, Social Media, and Mobile Innovations in the Hospitality Industry", *Cornell Hospitality Quarterly*, 53(3), 2012, pp. 183-186.
- [6] Torres, E.N., Adler, H., Lehto, X., Behnke, C., and Miao, L., "One Experience and Multiple Reviews: The Case of Upscale Us Hotels", *Tourism Review of Aiest - International Association of Scientific Experts in Tourism*, 68(3), 2013, pp. 3-20.
- [7] Edelman, J.R., Lee, Y.L., Lee, S.H., and Caldicott, J., "Developing a Taxonomy of 'Award-Winning' restaurants—What Are They Actually?", *Journal of Quality Assurance in Hospitality & Tourism*, 12(2), 2011, pp. 140-156.
- [8] Namasivayam, K., "The Impact of Intangibles on Consumers' quality Ratings Using the Star/Diamond Terminology", *Foodservice Research International*, 15(1), 2004, pp. 34-40.
- [11] Ghose, A., and Ipeirotis, P.G., "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics", *Knowledge and Data Engineering, IEEE Transactions on*, 23(10), 2011, pp. 1498-1512.
- [12] Sonnier, G.P., Mcalister, L., and Rutz, O.J., "A Dynamic Model of the Effect of Online Communications on Firm Sales", *Marketing Science*, 30(4), 2011, pp. 702-716.
- [13] Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M., "Mine Your Own Business: Market-Structure Surveillance through Text Mining", *Marketing Science*, 31(3), 2012, pp. 521-543.
- [14] Singer, J.D., "Using Sas Proc Mixed to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models", *Journal of Educational and Behavioral Statistics*, 23(4), 1998, pp. 323-355.
- [15] Parasuraman, A., Zeithaml, V.A., and Berry, L.L., "A Conceptual Model of Service Quality and Its Implications for Future Research", *Journal of marketing*, 49(4), 1985,
- [16] Cronin Jr, J.J., and Taylor, S.A., "Measuring Service Quality: A Reexamination and Extension", *The journal of marketing*, 1992, pp. 55-68.
- [17] Antun, J.M., and Gustafson, C.M., "Menu Analysis: Design, Merchandising, and Pricing Strategies Used by Successful Restaurants and Private Clubs", *Journal of nutrition in recipe & menu development*, 3(3-4), 2005, pp. 81-102.
- [18] Cotter, M.J., and Snyder, W.W., "How Mobil Stars Affect Restaurant Pricing Behavior", 1996,
- [19] Snyder, W., and Cotter, M., "The Michelin Guide and Restaurant Pricing Strategies", *Journal of Restaurant & Foodservice Marketing*, 3(1), 1998, pp. 51-67.
- [20] Abrate, G., Capriello, A., and Fraquelli, G., "When Quality Signals Talk: Evidence from the Turin Hotel Industry", *Tourism Management*, 32(4), 2011, pp. 912-921.
- [21] Ye, Q., Li, H., Wang, Z., and Law, R., "The Influence of Hotel Price on Perceived Service Quality and Value in E-Tourism: An Empirical Investigation Based on Online Traveler Reviews", *Journal of Hospitality & Tourism Research*, 38(1), 2014, pp. 23-39.
- [22] Oliver, R.L., "Effect of Expectation and Disconfirmation on Postexposure Product Evaluations: An Alternative Interpretation", *Journal of Applied Psychology*, 62(4), 1977, pp. 480.
- [23] Oliver, R.L., "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions", *Journal of marketing research*, 1980, pp. 460-469.
- [24] Liu, B., "Sentiment Analysis and Subjectivity", *Handbook of natural language processing*, 2(2010), pp. 627-666.
- [25] Li, N., and Wu, D.D., "Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast", *Decision Support Systems*, 48(2), 2010, pp. 354-368.
- [26] Pang, B., and Lee, L., "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", in *Book A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts*, Association for Computational Linguistics, 2004, pp. 271.
- [27] Tong, S., and Koller, D., "Support Vector Machine Active Learning with Applications to Text Classification", *The Journal of Machine Learning Research*, 2(2002), pp. 45-66.
- [28] Das, S.R., and Chen, M.Y., "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web", *Management science*, 53(9), 2007, pp. 1375-1388.
- [29] McCallum, A., Nigam, K., Rennie, J., and Seymore, K., "A Machine Learning Approach to Building Domain-Specific Search Engines", in *Book A Machine Learning Approach to Building Domain-Specific Search Engines*, Citeseer, 1999, pp. 662-667.
- [30] Nielsen, F.Å., "A New Anew: Evaluation of a Word List for Sentiment Analysis in Microblogs", *ESWC2011 Workshop on 'Making Sense of*

Microposts': Big things come in small packages 718
in CEUR Workshop Proceedings, 2011, pp. 93-98.