ARTICLE IN PRESS

Journal of Business Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

False advertising or slander? Using location based tweets to assess online rating-reliability☆

Amit Poddar[a,*], Syagnik Banerjee[b], Karthik Sridhar[c]

[a] Department of Marketing, Perdue School of Business, Salisbury University, Salisbury, MD 21801, United States
[b] University of Michigan, Flint, United States
[c] Baruch College, The City University of New York, United States

ARTICLE INFO

Keywords:
Big data
Location aware tweets
Data mining
Public policy
Electronic WOM
Online ratings

ABSTRACT

Online reviews have diffused into a large variety of businesses, ranging from physical goods to services. The ubiquity of the reviews and the importance given to them by potential customers makes examining their validity extremely important. While good reviews can boost companies' business, bad reviews can spell their doom. Since online reviews are anonymous, there are cases of both false advertising and slander that can create conflict. In this paper authors provide reasons for online rating bias and demonstrate a way to measure it. Authors mine consumer's location aware tweets from business locations to capture a location's pleasure score and compare the pleasure scores to Yelp ratings to determine how overrated or underrated the venue is. Foursquare and Twitter are mined to extract an emotion score from the location aware tweets using a dictionary called the Affective Norms for English Words (ANEW). Rating biases are found across cities and different types of restaurants and managerial and policy implications discussed.

1. Introduction

2013 was witness to the conclusion of a year-long undercover investigation called "*Operation Clean Turf*" after which 19 companies in New York agreed to stop posting false online reviews for businesses and paid more than $350,000 in fines and penalties (Schneiderman, 2013). The incident brought to light the business of false consumer reviews, which affects consumer expectations and business reputations. Today the large majority of consumers read review websites before making purchase decisions, whether it be for buying a product or a service. Similarly, every organization that sells to or serves a consumer, expects to be reviewed online. Good reviews have become points of pride for businesses and are often displayed as *badges* similar to marks of honor. Hotels and restaurants that are rated highly by travel site "Tripadvisor" or "Yelp" proudly put a sticker outside their business locations demonstrating their popularity. These websites also give out free stickers to businesses which convince first time customers of the trustworthiness of a business, and encourage them to try out a new place. The rating score and sticker thus serve as a seal of approval, a symbol of trust and reliability.

The growth of review sites has also fueled multiple controversies. On one hand, critics believe review sites have become so powerful that even one extra star in a yelp review can increase revenues by 5–9% (Economist, 2015). This has created a cottage industry of *illegal* businesses that capitalize on posting flattering reviews to increase the values of client's businesses (Malbon, 2013). While businesses can pay customers to post reviews and boost revenues, getting a bad review on reputed review sites can also doom a business. Since reviews are anonymous, competitors with the right resources can doom a competing business by posting false reviews or paying one of the aforesaid businesses to post them against a restaurant. Overall, the act of creating or distributing a deceptive review that a reasonable consumer believes to be unbiased is known as "astroturfing." Astroturfing violates New York Executive Law § 63(12), and New York General Business Law §§349 and 350 (Schneiderman, 2013).

This serious problem has prompted some companies (Green, 2014) to sue the rating websites, or in some cases the reviewers, for defamation, so that they can unmask the concerned reviewer or take down libelous reviews. The US Federal Trade Commission (FTC) has updated their guidelines concerning the use of endorsements and testimonials in advertising in 2009 including social media (FTC, October 15, 2009), requiring a hefty $16,000 per day fine for those in violation. Several stakeholders are invested in better understanding the reliability of online reviews. While businesses are concerned about the impact of

☆ Note: This research has not received funding from any external source.
* Corresponding author.
E-mail addresses: axpoddar@salisbury.edu (A. Poddar), syban@umflint.edu (S. Banerjee), Karthik.Sridhar@baruch.cuny.edu (K. Sridhar).

http://dx.doi.org/10.1016/j.jbusres.2017.08.030
Received 15 November 2016; Received in revised form 28 August 2017; Accepted 29 August 2017

reviews on their reputation and effectiveness to attract new customers, consumers are concerned about plausible deception, and policy makers regarding fair and accurate representation. There has been a significant history of lawsuits, interactions and proxy-wars (Goldman, 2014) between Federal Trade commission, businesses, consumers and reviewers which demand the reliability issue be addressed.

In this paper, we present a conceptual framework to elucidate factors that lead to online review bias, namely, geographic location and business specific sub-categories. Then, applying methodological insights from prior literature on temporal congruity, we design and introduce an alternate data source, location based social media posts, as a more objective benchmark to measure the reliability or bias in online ratings of specific venues. We use location based social media posts (twitter) as these posts are immediate (thus not tainted by memory bias) and furthermore may not be formal reviews or reviews at all. In a way tapping into social media posts allow us to electronically eavesdrop on what consumers are saying. By extracting emotions from these electronic eavesdrops, we are able to find the real feelings of these consumers. We use the restaurant twitter postings from the same venues to account for the reliability of the corresponding yelp ratings. Further, we test for differences and identify certain cities and types of restaurants where the biases are more prevalent than others.

1.1. Online review reliability

While there is increasing evidence to businesses that these reviews have direct effects on product sales, for the review websites the sanctity of the reviews is a very serious business as their entire business model is based on having objective, unbiased reviews. The ecosystem of lawsuits and charges have included policy bodies, businesses, review websites like Yelp, as well as consumers that submit reviews. Some companies have started to fight back using technology. As an example, expedia.com allows one to review only if the person reserved and stayed in the hotel, while yelp uses algorithms to flag and remove reviews that depict an inaccurate representation of the business (Mukherjee, Venkataraman, Liu, & Glance, 2013; Streitfeld, 2013). Others have taken the legal route of filing lawsuits to take down false reviews (Green, 2014). A multiple restaurant owner in Mammoth Lakes California has charged Yelp for false advertising the capabilities of their filtering algorithm, for screening out half the reviews on one of his restaurants and not detecting one that allegedly included false statements (Goldman, 2014). Similarly, other small businesses like dentists in Manhattan, pet sitters in Texas and moving and storage companies in Florida have taken legal action in response to negative or low reviews, or sharing of trivial information. Prestigious Pets Inc. filed a $1 million lawsuit against a couple for saying that the company overfed their goldfish (Manatt Phelps & Phillips, 2016). In response, Yelp has posted consumer alerts regarding questionable legal threats, as pop-up boxes on the company's pages, reminding reviewers of their First-Amendment rights to enable a level playing field. On the other side of the table, the FTC has charged AmeriFreight, an automobile shipment broker, for violating section 5 of the FTC Act by failing to disclose that they compensated consumers for submitting online reviews/ratings (Federal trade commission, 2015).

It is evident from the above examples that online rating reliability is of interest to not only businesses and consumers, but also policy bodies and watchdogs that protect and enforce fair and ethical business practices. Prior research in understanding perceptions of online review reliability have explored the effects of argument quality, source credibility, review sidedness, and review consistency, at different levels of involvement and expertise (Cheung, Sia, & Kuan, 2012). Similarly, Luca and Zervas (2015) identify different restaurant characteristics that cause them to use fake reviews. Ney (2013) identifies factors consumers use to assess credibility of online reviews. The problem of unreliable reviews therefore creates an interesting set of questions that we address in this paper.

Prior literature suggests that there are two broad types of online review biases. One is deliberate, led by motives of false claims or slander, popularly known as review fraud, while the other is unconscious bias. The former, suggests that what is considered to be wisdom of the crowds is rather its manipulation, indicating the lack of trustworthiness in online reviews and ratings (Moyer, 2010). Similarly, Das, Lavoie, and Magdon-Ismail (2013) suggest that arbiters of collective intelligence manipulate public opinion towards a goal. The latter, unconscious bias is further caused by lack of cultural consciousness (Coe, 2009) also known as gastronomic bigotry (Simmons, 2014), utilitarian or hedonic nature of product related experience or expectations (Moore, 2015), mood, negativity bias (Chen & Lurie, 2013), self-selection as well as nonresponse bias (Wu, 2013) among many others.

Given recent industry trends, tweeting while eating has become quite popular. 47% of millennials have reported using social media while immersed in eating or drinking. Based on increasing volumes of photographs and digital content from food lovers, a lot of restaurants have also turned their attention to provide better customer service responses via social media. Due to the relevance of dining related usage from restaurants, in this paper we choose to focus on the restaurants industry discussing the context of reviews. Furthermore restaurants are the ideal industry to study as restaurants are an industry which generates most reviews as well as reviews which can very strong. Restaurants are also an industry which provide a large enough sample with enough variability in quality that researchers can test testable propositions with respect to online biases. Based on the above and prior literature, we identify two variables that can demonstrate systematic differences in review biases on restaurant review websites.

1.1.1. Geographic location

In the restaurant industry one of the most important parameters that determines success is location. It can be argued that locations with more consumer populations nearby as likely to do better. However higher consumer density also attracts more competition (e.g. tourist areas). Therefore locations which have more density of restaurants per capita increases competition for consumers, leading to economic incentives for committing review fraud (Luca & Zervas, 2015). The increase in geographic proximity between competitors also leads to an increase in competitive intensity and negative fake reviews. Hence, we propose that.

Proposition 1. Geographic locations higher in business density of restaurants demonstrates greater bias in online ratings.

1.1.2. Business sub-category type

Within a specific business there exits different sub-categories of business types where within each sub-category there exists commonality in the products and services being offered to the consumer and across. For instance, restaurant business is further sub categorized based on the cuisine they serve to the end consumers. Dining experiences emanating from these different restaurant types can be regular or based on sensory or craving related needs. Consumers patronize regular restaurants out of necessity, because they serve more utilitarian purposes evaluated on more cognitive terms. In contrast the latter, can be based on hedonic characteristics like enjoyment, craving or experience. This enjoyment can arise out of cuisine conspicuity in exotic or ethnic cuisine. While cuisine conspicuity can lead to gastronomic bigotry (Coe, 2009) and review biases can arise due to lack of cultural knowledge or appreciation of other types of food or environment (Banerjee, Sridhar, Poddar, & Kumar, 2017), they can also lead to greater disappointment due to tastes not confirming to higher expectations (Banerjee, 2016). Such a rationale can be extended to the sub-categories of business existing under any business type. Hence, we propose:

Proposition 2. Sub-categories of business that elicit emotional expectations demonstrate greater online rating bias.

2. Methodology

In this section we raise several questions on how to detect bias. First, what kind of a data source should be used as a benchmark to detect the bias posed by reviews, surveys or other kinds of solicited primary data? Second, what criteria should be used to evaluate the suitability of the source? Third, what specific variables should be extracted from the chosen data source and used as indicators to assess the rating reliability and why? Fourth, what kind of patterns and solutions can one expect to draw from the findings? In the following section the first three questions are addressed. Then, data is collected and analyzed to demonstrate some sample findings from which certain conclusions and implications can be drawn.

2.1. Location based social media

We considered location-based social media as the preferred choice among secondary data sources recording customer impressions of service experiences for several reasons. When it comes to patronizing a business, many customers don't bother filling up surveys and fall into a “silent majority” category. Even those who do, can be influenced by emotional biases like social desirability bias or corrupted by incentives from businesses themselves. Also, traditional evaluation instruments don't capture a time and location dimension, thus could be tainted by memory bias (eg a consumer going home and then posting a review on the sites website).

While going through the effort to open a reviewer account and filling a form to provide a publicly accessible review is cumbersome, consumers regularly talk to friends and colleagues sharing their experiences without motivated solicitation. Word of Mouth (WOM), which is considered more credible than other forms of paid media, has been gradually absorbed by Social Media in recent years, and become the dominant way in which consumers share their opinions and feelings (Sotiriadis & Van Zyl, 2013; Sparks & Browning, 2011). Social Media listening or monitoring is a popular tool used by brands, who use it to assess their brand sentiments over time (Schweidel & Moe, 2014). Some companies, like Geofeedia, also enable location-based-monitoring to track incidents in and around business establishments.

Several factors help determine the suitability of a social media platform as a data source. One, what type of information does the platform capture? Two, is there a way to validate if the post was connected to the user's personal or physical experience? Three, is the user base of the platform sizable enough and representative of the customers one plans to study? And four, can a researcher or corporation have seamless access to the source in order to ensure an uninterrupted data collection and analysis process at minimal costs? These questions are answered in the following paragraphs.

Social media sites invite sharing different types of information, some, such as Facebook are comprehensive and elaborate, some like Pinterest, Flickr are based on narrow niches including image-sharing, some like Twitter and Vine are geared towards sharing short bursts of immediate raw emotions which can also include embedded urls or videos, while WhatsApp and Telegram serve mainly as substitutes for traditional communication medium. By design, Twitter messages are constrained to 140 characters or less, so consumers share their immediate thoughts and emotions, and there is a limit to the posted content for quick processing.

Twitter also captures latitude and longitude information, so when used from a mobile device, the locus of a user's movement can be traced as the person tweets. An added value of Twitter is it's connectivity with Foursquare, a location based application. Mobile users who use Foursquare, a Location Based Social Network (LBSN) are alerted every time they enter an establishment or business' physical perimeter, so that they can check-in, share their current location with friends in their network to join them. The GPS enabled features of mobile platforms allow user location co-ordinates to be collected and matched with business establishment addresses. Also, Foursquare users can check-in and then tweet about their experiences from within the app. In effect, Twitter was chosen so that one could extract data in combination with Foursquare that provided micro-location Meta data, being the exact venue of physical presence. The act of checking-in validates the information that the user tweeted from inside a particular venue or business at a certain point of time, and hence the feedback or sentiment is first-hand.

Twitter does have a large user base of 313 million active monthly users (Smith August 14, 2016), whereas Foursquare has a base of 60 million (Weber & Novet, August 18, 2015). Though the users' gender is more skewed towards women and younger age groups, younger age groups are more strongly influenced by peer groups (Steinberg & Monahan, 2007) and hence, are more likely to read and write online reviews (Anderson, October 19, 2015). So Twitter and Foursquare's user bases can be considered fairly representative of online review readers and contributors. On access and cost, Twitter's API (application interface) is open to other applications that can build their business atop twitter data. The only real cost of data extraction is the coding skill, otherwise the data is public. This makes it easy for businesses to setup live, streaming data flow into databases that can be used to generate alerts.

The above points answer the question on why Social Media, and why Twitter and Foursquare are suitable for use as a benchmark for determining online review reliability. The third question, and the last to be addressed in this section, is what variables should be used to compare Yelp ratings. Overall, the location validation adds significant value to the shared content. The presence of temporal-contiguity cues, (the location validation indicates the temporal proximity of the consumption experience to the review), reduces the extent of bias in the review (Chen & Lurie, 2013). The content of an unsolicited, instantly shared post from within the premises of a service provider is more likely to be spontaneous and honest as compared a rationalized or monitored review about the same experience later on. Further, the location based tweets are more difficult for competitors to game, as one has to physically be within the premises of the business to count towards it. Also, a key fact is that a tweet from a location may not be a review at all, it could be a spontaneous expression of joy, sorrow or disgust. However, by decoding/inferring a meaning from the tweets, one can associate the subject's presence in that location on the emotion contained in the tweet and thus generate insight into the consumer's feelings about the place. As mentioned in the beginning of the paper, capturing emotions contained in the location aware tweets sent through foursquare is thus akin to eavesdropping on a private conversation and inferring the state of the consumers mind at the particular location. We thus propose to use the emotion contained in those tweets as a proxy for the review of a particular location.

Using data mining, we first propose a method to extract the emotions of the tweet using a validated scale called the ANEW scale. We then calculate average emotion scores for a particular location in the restaurants category. We then compare the emotion scores with the rating provided to a particular restaurant in a popular rating site (yelp). That comparison helps us determine the extent of misleading reviews in that category.

2.2. ANEW scale

Known as the Affective Norms for English Words, the ANEW scale (Bradley & Lang, 1999) was developed to provide normative emotional ratings for a large number of words in the English language. The emotions measured were the same that were identified by the PAD model (pleasure-arousal-dominance) model introduced by Mehrabian and Russell (1974) stemming from environmental stimuli. Pleasure ranges from unhappiness or pain to happiness or ecstasy, arousal ranges from drowsiness to alertness, and dominance ranges from feelings of being cared for to that of being in charge. Prior research was used to generate word lists which were further rated by affective rating systems to develop a list of words and their ratings of emotion. Bradley and Lang (1999) put together a group of 2476 words and then presented them counterbalanced in rows and columns, asking subjects to respond to the words using a self-assessment manikin. The self-assessment manikin is a non-verbal pictorial assessment technique which measures the pleasure, arousal, and dominance embedded in a person's affective reaction to presented stimuli. Subjects select their positions on a scale for three types of feelings represented by pictures: Happy vs. Unhappy, Excited vs. Calm, and Controlled vs. In-control. The positions selected decide the scores each word has in the ANEW scale. At the end of the processing a numerical index for each Tweet was presented, based on the average of the scores from the keywords. Pleasure is the most relevant emotion for marketers as all consumption and more specifically hedonic consumption (eg eating out at restaurants) is driven by the desire of consumers to seek pleasure (Alba & Williams, 2013; Park, 2004). Also research suggests that the pleasure can be considered as a close proxy of likability (Poels & Dewitte, 2008). In this research, we will thus focus on the pleasure perceptions of social media users.

2.3. Data collection and preparation

In phase I we mined data from check-ins at restaurants across six regions in USA including Central Park in New York (NY), Harvard Square in Cambridge (MA), Market Street Twitter office in San Francisco (CA), Capitol Hill in Seattle (WA), Union Station in Chicago (IL) and Union Station in Washington D.C. These specific regions were chosen because of the high volume of check-ins emanating from them on foursquare. We tracked check-ins from April to July. This data allowed us to extract fields like restaurant name, tweet content and time of check in and tweet. Over twenty-five thousand tweets were analyzed which were posted by approximately 14,000 users. Each tweet was divided into its constituent's words and the words were checked against the Anew scale items. When a word was identified, we allotted a numerical pleasure value to that word. At the end of the processing we had an average numerical *pleasure* index for each tweet. This process was repeated for all twenty-five thousand tweets (Fig. 1).

In the second phase each of these restaurants is identified and matched to their respective yelp ratings/scores. Before using the combined database, we did a little bit more data pruning. First we eliminated all data items for which a pleasure score could not be determined. These are tweets that are just about “checking- in” or the tweeter checked in using a foreign language. Further, we eliminated all restaurants which had fewer than 5 tweets from the same location. Similarly, if a restaurant had a yelp listing sans yelp score, it was eliminated. The data pruning ensured we end up comparing restaurants which have ratings/scores using both methods. After this process we were left with 158 unique restaurants. The total unique tweets that were sent from these 158 restaurants in the 6 locations were 6153. The mean number of tweets from each restaurant location is 39, with a minimum of 5 and maximum of 510 (Table 1).

2.4. Data preparation

With the ANEW scores, we prepare the measures in two ways. Firstly, we standardize the pleasure scores and yelp ratings so that a reliability ratio can be calculated, and the proportion of restaurants with higher or lower values can be compared. Secondly, we compute an equivalent pleasure score scaled to the Yelp rating scale, so that absolute differences between them could be tested for statistical significance.

First, for each restaurant we created a pleasure score P(x). This was the average pleasure score received by the restaurant over the period that the data was collected. As mentioned earlier we link each unique restaurant to its respective yelp score Y (x) (noted from the yelp website). To examine if the tweets scores are comparable to the yelp scores, we needed to standardize the scores so that comparison became meaningful. We next calculated a reliability ratio $Z(x) = P_s(x) / Y_s(x)$ for each of the 158 restaurants in the database where $P_s(x)$ and $Y_s(x)$ are the standardized scores of P(x) and Y(x) respectively. If $Z(x) > 1$, it implies consumers get more pleasure than the lower yelp rating suggesting that the restaurant is undervalued. On the other hand, If $Z(x) < 1$, it would suggest that the restaurant is overvalued since customers are getting less pleasure from the restaurant than suggested by the yelp score. It might also be an indicator that the restaurant could be inflating the yelp score by using the help of reputation management companies. Also, the percentage of firms who are significantly divergent from $Z(x) = 1$, would show the extent of the problem as discussed in the beginning of the paper.

Second, we calculated the ANEW score equivalent to a 1–5 scale. The ANEW scale comes with a list of 2476 keywords. The scores can be generated as either a sum of the scores of the matched keywords in a tweet, or an average score of them, which is calculated as the sum score divided by the number of keywords. We used the latter definition, as it seemed more unbiased or less likely to be inflated by number of matched keywords. The lowest possible average score of a tweet is zero, as it's possible for a tweet not to contain any matched keywords at all. The highest possible average score (per keyword) of a tweet is the highest possible score of a keyword, which is 8.82. This is because a tweet cannot contain keywords which have an average score higher than the highest of 2476 keywords. So we converted ANEW scores on a scale from 0 to 8.82 to a scale between 1 and 5 using the following formula.

$$Y = (B - A)*(x - a)/(b - a) + A.$$

a is the lowest value of the scale to be transformed b is the highest value of the scale to be transformedA is the lowest value of the scale to

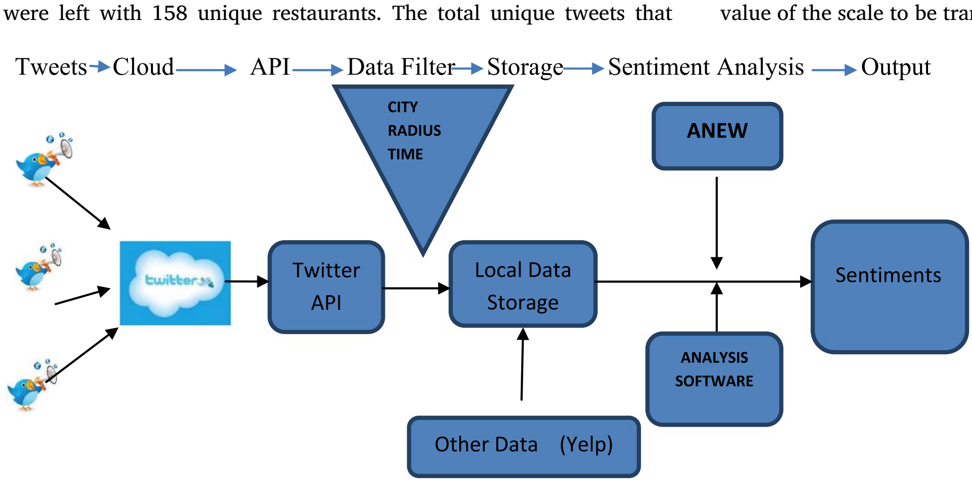

Fig. 1. The data mining, collection, preparation and processing flow chart.

Table 1
Cities and restaurant types by number of restaurants and tweets.

Categories	Variables	Cambridge Harvard Square	Chicago	New York Central Park	San Francisco	Seattle Capitol Hill	Washington DC Union Station	Total
American	# Tweets	167	222	10	34	42	142	617
	# Restaurants	5	8	1	2	2	5	23
Asian	# Tweets	13	7	0	5	5	0	30
	# Restaurants	1	1	0	1	1	0	4
Casual	# Tweets	457	868	525	468	358	328	3004
	# Restaurants	8	12	9	4	4	8	45
European	# Tweets	30	193	64	0	0	38	325
	# Restaurants	1	7	3	0	0	3	14
Fast food	# Tweets	66	404	165	186	47	190	1058
	# Restaurants	3	7	5	5	2	5	27
Healthy	# Tweets	32	288	105	20	13	102	560
	# Restaurants	2	10	6	2	2	6	28
Latin	# Tweets	141	191	63	11	16	99	522
	# Restaurants	5	4	2	1	1	3	16
Total	# Tweets	906	2210	932	724	481	900	**6153**
	# Restaurants	25	50	26	15	12	30	**158**

which the transformation will take place B is the highest value of the scale to which the transformation will take place.

Here ANEW (or the tweet emotion) is the scale to be transformed, to the Yelp scale. So a = 0, b = 8.82, A = 1, B = 5. In other words, 5 on the ANEW/Tweet score will translate to (4*5/8.82) + 1 = 3.26 in 1–5 interval scale.

3. Analysis and findings

Before doing statistical analysis on differences between pleasure scores and yelp scores, we looked at the percentage of overvalued and undervalued restaurants. Overall, we found that among all restaurants, over 75% of the restaurants were classified as overvalued. In other words, based on tweet emotion content, most Yelp ratings appear positively biased. Table 2 provides the percentage breakdown for the different types of restaurants in the same. Latin restaurants were the most overvalued at 88% followed by fast and comfort. One interesting initial finding was that American category restaurants were the most undervalued. 47% of the restaurants were undervalued on yelp as compared to their pleasure scores. That would suggest that while people may not give very glowing remarks to American Cuisine Restaurants on yelp, the analysis of their tweets suggest that they are really happier while eating at these places as compared to other kinds of restaurants.

For different categories of restaurants, we tested for differences between Yelp and the ANEW equivalent scores (on a 1–5 scale). The results for differences between the different kinds of restaurants are shown in Table 3. For American (F (1,44) = 0.31, $p < 0.6$) and European food (F (1, 26) = 0.089, $p < 0.8$), differences were not significant, but for other types of restaurants, namely healthy (F (1,54) = 3.98, $p < 0.06$), casual (F (1,88) = 8.584, $p < 0.01$) Latin (F (1,30) = 24.42, $p < 0.001$) and fast food (F (1,52) = 4.84, $p < 0.04$) categories, Yelp reviews were significantly higher than the pleasure scores from location based twitter streams.

Table 3
Significance testing between Restaurant Types (Twitter vs. Yelp).

Restaurant type	Standardized average pleasure rating (Twitter Score 1–5)	Average Yelp score (1–5)	F	Sig.	N
American	2.766	2.609	0.309	0.581	23
Casual	**2.536**	**3.133**	**8.584**	**0.004**	**45**
European	2.870	2.786	0.089	0.767	14
Healthy	2.574	3.250	3.980	0.051	28
Latin	**2.513**	**3.688**	**24.420**	**0.000**	**16**
Fast food	**2.654**	**3.259**	**4.843**	**0.032**	**27**

Table 4
Significance testing between restaurant locations (Twitter vs. Yelp).

CITY	Standardized average pleasure rating (Twitter Score 1–5)	Average Yelp score (1–5)	F	Sig.	N
Cambridge	2.64	2.72	0.10	0.748	50
Chicago	2.64	2.96	2.15	0.145	100
NYC	**2.49**	**3.41**	**11.18**	**0.002**	**54**
SF	2.65	3.00	1.12	0.299	30
Seattle	2.84	2.92	0.04	0.846	24
DC	**2.64**	**3.47**	**11.53**	**0.001**	**60**

Finally, we decided to test whether there were any score differences between the various cities that constituted our sample. The results of the analysis are shown in Table 4. We find that in New York City (F (1, 52) = 11.18, $p < 0.01$) and Washington DC (F (1, 58) = 11.532, $p < 0.002$) Yelp's review scores were significantly higher than ANEW equivalent scores. In other cities, both sources of review appeared to have similar ratings. This was an interesting finding as recently there was a very high profile case where the Office of New York attorney general's office took actions against several businesses which were

Table 2
Restaurant types by percentages undervalued and overvalued.

Restaurant type	Pleasure rating mean	Pleasure Std. deviation	N	Yelp rating mean	Yelp Std. deviation	Undervalued (%)	Overvalued (%)
American	4.5143	2.98466	23	2.52	1.148	47	53
Asian	5.904	1.229	4	2.94	0.348	50	50
Casual	3.2168	2.17621	45	2.54	1.076	35.5	64.5
European	4.9854	2.52135	14	2.59	1.113	42	58
Healthy	3.359	3.71911	28	3.1	1.673	35.7	64.3
Latin	2.1091	2.16321	16	3.69	0.846	12.5	87.5
Fast and comfort	3.1653	2.03263	27	3.26	1.475	26	74
Total	3.5419	2.70791	158	2.81	1.278	25	75

soliciting fake reviews on yelp in return for monetary consideration in New York City (Nyag Press, 2016 and Sollitto, 2016). The attorney general's office considered these fake reviews as akin to false advertising and the companies agreed to cease the practice and pay fines between $20,000 and $50,000 to settle the case.

Overall, our study demonstrates partial support for both the expectations of finding significant differences, across restaurant and city type. Despite being for different reasons, more competitive cities, and more ethnic restaurants may expect to find more biased reviews and ratings. The former may be deliberate due to competitive intensity, whereas the latter could be unconscious biases due to ethnocentrism and nature of motivation.

4. Discussion and managerial implications

This paper demonstrates how big data, especially location aware tweets can be used for evaluating restaurants and also proposes an alternate rating method using the emotion contained in the tweets. This is done by data mining tweets and performing sentiment analysis using the validated ANEW scale. This section points out various implications of using the data, the method, and the nature of findings on researchers, businesses, consumers and policy bodies.

First, bias detection using the above process including a streaming data source can allow restaurants to more accurately monitor their online reputations and service recovery needs relevant to in-store operations. Currently online reviews and rating websites primarily reflect on data that looks back in time. A high rating on yelp does not reveal how the restaurant is on the current day, it only evaluates how the restaurant has fared since it opened. In contrast, ratings based on location aware tweets can be examined by time periods, or be more current, as required. If there are enough location aware tweets sent from a particular location, the ratings can even be real time. Existing firms who get high traffic of tweets, like Starbucks, can set up alerts of Yelp scores against changing ANEW scores, so that they can be informed when the differences exceed a certain value.

Secondly, these location-based ratings can allow better discovery of new and upcoming restaurants for consumers. New restaurants traditionally have fewer Yelp reviews. As a result, they are typically not easily discoverable by consumers who are looking for new and exciting places to dine. However, ratings based on location aware tweets can allow new restaurants to stand out from the crowd. New restaurants can monitor how location based tweets differ from Yelp ratings and advertise positive differences on their websites. It will help new restaurants establish themselves and consumers discover new and exciting places.

Third, this method for validating online review scores can be used in the policy and legal context for helping settle lawsuits as well as informing the legislative process. Though section 230 of the Communications Decency Act of 1996 (or Title V of the Telecommunications Act of 1996) provides protection against review websites liability for user reviews or their operational filters, the precedent set by the case of Demetriades vs Yelp Inc. (Neuburger, November 19, 2014) demonstrates how such protection can be bypassed using false advertising claims, or in other words, accusing online review websites of advertising trustworthy reviews, but actually having a poor filter. For cases like this, a third party benchmark like location based tweet emotions can illustrate if Yelp scores are significantly different from the existing benchmark, if Yelp's filter is truly poor, or representative, and help settle the case. The same can be done in cases where businesses intimidate their clients from posting poor reviews by helping anti- SLAPP (Strategic Lawsuits Against Public Participation) laws, by revealing how other sources like social media evaluate the same business. Further, such tools can aid Consumer Review Fairness Act, H.R. 5111, by informing what benchmarks can be used to determine defamation, libel or slander.

Fourth, policy watchdogs can use real-time analysis as done above to determine what types of restaurants or cuisines are most vulnerable to inaccurate representations. In our example, Latin or Mexican food restaurants seem to the most overrated on Yelp, compared to their tweet emotion scores. Also, the data can be used to examine which geographic regions are most problematic. The reasons can be many, including competitive intensity which prior researchers consider one of the reasons for fraudulent reviews. Geographic regions with higher competition may demonstrate more inaccurate reviews or representations. From our data, New York and Washington DC seem to be the cities that have most inaccurate representations. These regions can be provided added scrutiny.

Fifth, store managers in areas of high business concentration should have in place systems that can monitor the evaluation of their business locations through various third party review/ratings websites. They need to be proactive in reporting to the website owner the possibility of a fraudulent review or rating. They can also put plug-ins on their business websites that capture and highlight conversations emanating from their business locations and also the overall sentiment/emotion based on the content posted.

4.1. Theoretical implications

Overall, the findings in this paper contribute to determining rating biases and detection of fake reviews (Jiang, Cui & Faloutsos, 2016). We find that cities varying in competitive intensity and cuisine from diverse cultural backgrounds can engender biases in online ratings. While a certain extent of difference can be attributed to competition and considered as deliberate, false self-promotional or slanderous towards a competitor, the bias differences in cuisines can be more subjective, where it is difficult to determine whether genuine ratings by a person unable to acquire an authentic exotic taste can be truly considered fake. Hence, introducing the biases contributed by cuisine differences indicates that the mere existence of the bias is not sufficient to determine astroturfing. The validity of the bias is separate from whether it is deliberate or not. Hence one requires clarity on both validity, as well as intent of demonstrated bias, in order to determine whether the bias amounts to the review being fake rather than misinformed. While it is not possible to measure intent from the existing data, that is a relevant question for future research.

5. Limitations and future research

The study has some limitations that need to be considered by readers to apply it usefully. First, the data set was collected over a period of only three months, which may not be representative of all seasonal or weather effects on moods and emotions. Second, the collected data was from only the United States and even within that from some well recognized touristic areas. This was done to ensure a minimum volume of tweets within a short span of time. The processing was done only in the English language. We also had to delete all tweets that were in foreign languages during data processing, thus we might have missed out on some additional insights. Regional and cultural diversity can also potentially alter the effects. Third other environmental factors need to be accounted for to explain regional patterns of moods emanating from the tweets. Being in-the-moment and spontaneous, factors like local pollution, traffic density and urban environment can affect the moods being tweeted. Fourth, due to the unsolicited nature of the tweets it is difficult to assure the researchers a steady flow of data. It also leaves the study open to vulnerabilities of missing data. For example, there were many consumers who checked in, but did not tweet at all. This could be because there are many consumers who are not as tech friendly, or those who are more private, may not be equally willing to tweet for both good and bad service experiences, leading to another form of the self-selection bias. Finally, the location aware tweets that were used emanated from foursquare. Since the data availability is dependent on foursquare maintaining the connectivity with Twitter through their API, the ability to do this kind of analysis depends on continued access to foursquare data. Every time Foursquare updates their developer guidelines, it interrupts the data mining code from collecting data. If foursquare changes their data availability policy or goes out of business, the ability to do this kind of analysis would be imperiled.

However, the paper does open up an interesting way of viewing Astroturfing related problems that are becoming increasingly relevant for marketers. The third party benchmarks and unsolicited content question how to define the traditional claims regarding false advertising, defamation and libel. If a business claims defamation due to a low review rating, does that claim remain as valid if ratings on other data sources appear similar? To answer this question future studies need to accommodate larger time-span data, with data from more review websites as well as location-based platforms to determine if on-the-spot raves and rants from mobile devices can compare with post-hoc rationalized reviews uploaded from desktops. Further, the scope remains for inter disciplinary research with fields like Geographic Information Systems and Computation Sciences to help extract more specific information from social media posts and map them to visually recognize the spatiotemporal dimensions of their impact. Future studies could also compare content analysis of twitter data with content analysis of Yelp data to determine the impact of immediacy, assess reliability and source credibility. Similarly twitter feed data could also be compared with other online data sources such as facebook, tumbler and wordpress. Comparing online WOM and offline WOM could also add an interesting additional dimension to this research.

References

Alba, J. W., & Williams, E. F. (2013). Pleasure principles: A review of research on hedonic consumption. *Journal of Consumer Psychology, 23*(1), 2–18.

Anderson, M. (2015). *"Young consumers read more reviews & are more likely to write reviews", Search Engine Land, October. 19*. (retrieved from:) http://searchengineland.com/young-consumers-read-reviews-likely-write-reviews-233190.

Banerjee, S. (2016). Exotic or mundane: Designing cuisine and customer rankings from geo-social restaurant data. *Conference proceedings of INFORMS, Nashville*.

Banerjee, S., Sridhar, K., Poddar, A., & Kumar, N. (2017). *Monitoring geo-social posts for situated customer intelligence: A structural interface for mining micro-blogging user generated content.* (Working Paper).

Bradley, M. M., & Lang, P. J. (1999). Affective norms for english words (ANEW): Stimuli, instruction manual, and affective ratings. *Tech. report C-1*. Gainesville: Florida: University of Florida.

Chen, Z., & Lurie, N. (2013). Temporal contiguity and the negativity bias in the impact of online word-of-mouth. *Journal of Marketing Research, 50*(4), 463–476.

Cheung, C. M.-Y., Sia, C.-L., & Kuan, K. K. Y. (2012). Is this review believable? A study of factors affecting the credibility of online consumer reviews from an ELM perspective. *Journal of the Association for Information Systems, 13*(8) (Article 2).

Coe, A. (2009). *Chop suey: A cultural history of Chinese food in the United States.* New York: Oxford University Press.

Das, S., Lavoie, A., & Magdon-Ismail, M. (2013). Manipulation among the arbiters of collective intelligence: how wikipedia administrators mold public opinion. *Proceedings. of CIKM* (pp. 1097–1106). .

Economist (2015). *Five star fakes: Evolving fight against sham reviews.* (Oct 24).

Federal Trade Commission (2009). "Guides concerning the use of endorsements and testimonials in advertising" - *16 CFR Part 255* (October 15. Retrieved from:) https://www.ftc.gov/sites/default/files/documents/federal_register_notices/guides-concerning-use-endorsements-and-testimonials-advertising-16-cfr-part-255/091015guidesconcerningtestimonials.pdf.

Federal Trade Commission (2015). "FTC stops automobile shipment broker from misrepresenting online reviews", February 27th, Press Release (retrieved from:) https://www.ftc.gov/news-events/press-releases/2015/02/ftc-stops-automobile-shipment-broker-misrepresenting-online.

Goldman, E. (2014). "Does Yelp have the 'Most Trusted Reviews'? A court wants to know more", July 30th. *Forbes*. (Retrieved from:) http://www.forbes.com/sites/ericgoldman/2014/07/30/does-yelp-have-the-most-trusted-reviews-a-court-wants-to-know-more/#51ae80be6b7e.

Green, A. (2014). Lincoln city hotel sues for defamation after anonymous person posts scathing trip advisor review. *Oregon Live* (pp. 2014). . (May 21) http://www.oregonlive.com/portland/index.ssf/2014/05/lincoln_city_hotel_sues_for_de.html.

Jiang, M., Cui, P., & Faloutsos, C. (2016). Suspicious behavior detection: Current trends and future directions. *Intelligent Systems, IEEE, 31*(1), 31–39.

Luca, M., & Zervas, G. (2015). *Fake it till you make it: Reputation, competition, and Yelp review fraud. HBS working paper series*.

Malbon, J. (2013). Taking fake online consumer reviews seriously. *Journal of Consumer Policy, 36*(2), 139–157.

Mehrabian, A., & Russell, J. (1974). *An approach to environmental psychology.* Cambridge, Mass: MIT Press.

Moore, S. G. (2015). Attitude predictability and helpfulness in online reviews: The role of explained actions and reactions. *Journal of Consumer Research, 42*(1), 30–44.

Moyer, M. (2010). Manipulation of the crowd: How trustworthy are online ratings? *Scientific American, 303*(1), 26–28.

Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013). *What Yelp Fake Review Filter Might be Doing?* Association for the Advancement of Artificial Intelligencewww.aaai.org (NSF grant # IIS-1111092).

Neuburger, J. (2014). California supreme court denies review of ruling allowing restaurant owner's false advertising claims to proceed against Yelp. *New Media Law and Technology Blog, November 19*. (available at:) http://newmedialaw.proskauer.com/2014/11/19/california-supreme-court-denies-review-of-ruling-allowing-restaurant-owners-false-advertising-claims-to-proceed-against-yelp/.

Ney, J. (2013). *Determining the credibility of online social content during an active purchase decision* (Doctoral Dissertation)Glasgow: Strathclyde Business School, University of Strathclyde.

Park, C. (2004). Efficient or enjoyable? Consumer values of eating-out and fast food restaurant consumption in Korea. *International Journal of Hospitality Management, 23*(1), 87–94.

Phelps, M., & Phillips, L. L. P. (2016). "Yelp Warns Consumers About Legal Threats From Companies", August 11th, Lexology Newsfeed (retrieved from:) http://www.lexology.com/library/detail.aspx?g=de1e0d26-149d-464b-bd73-434ee050ca06.

Poels, K., & Dewitte, S. (2008). Getting a line on print ads: Pleasure and arousal reactions reveal an implicit advertising mechanism. *Journal of Advertising, 37*(4), 63–74.

Press, N.. "A.G Schneiderman announces settlement with Machinima and three other companies for false endorsement", Feb 11th, press release. (2016). http://www.ag.ny.gov/press-release/ag-schneiderman-announces-settlement-machinima-and-three-other-companies-false: (available at: last Downloaded on 8/22/2016).

Schneiderman, M. C. (2013). "A.G. Schneiderman announces agreement with 19 companies to stop writing fake online reviews and pay more than $350,000 in fines", September 23rd, Press release (available at:) http://www.ag.ny.gov/press-release/ag-schneiderman-announces-agreement-19-companies-stop-writing-fake-online-reviews-and.

Schweidel, D. A., & Moe, W. W. (2014). Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research, 51*(4), 387–402 (August, Volucella).

Simmons, A. (2014). Gastronomic Bigotry: Do you think an ethnic restaurant caused your food poisoning? You might be a little bit racist. *Slate*. (retrieved from) http://www.slate.com/articles/life/food/2014/06/ethnic_restaurants_and_food_poisoning_the_subtle_racism_of_saying_chinese.html.

Smith, C. (2016). "By the Numbers: 170 + Amazing Twitter Statistics", August 14th, DMR (retrieved from:) http://expandedramblings.com/index.php/march-2013-by-the-numbers-a-few-amazing-twitter-stats/.

Sollitto, V.. New York attorney general prosecutes fake reviews and validates Yelp's own efforts. (2016). Yelp Press Releasehttps://www.yelpblog.com/2016/02/new-york-attorney-general-prosecutes-fake-reviews-validates-yelps-efforts: (Downloaded on 8/22/2016).

Sotiriadis, M. D., & Van Zyl, C. (2013). Electronic word-of-mouth and online reviews in tourism services: The use of twitter by tourists. *Electronic Commerce Research, 13*(1), 103–124.

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management, 32*(6), 1310–1323.

Steinberg, L., & Monahan, K. C. (2007). Age differences in resistance to peer influence. *Developmental Psychology, 43*(6), 1531–1543.

Streitfeld, D. (2013). Give yourself 5 stars? Online, it might cost you. *New York Times,* A1 (Sep 22).

Weber, H., & Novet, J. (2015). "Foursquare by the numbers: 60M registered users, 50M MAUs, and 75M tips to date", August 18th. *Venturebeat*. retrieved from: http://venturebeat.com/2015/08/18/foursquare-by-the-numbers-60m-registered-users-50m-maus-and-75m-tips-to-date/.

Wu, P. F. (2013). In search of negativity bias: An empirical study of perceived helpfulness of online reviews. *Psychology and Marketing, 30*(11), 971–984.

Amit Poddar is Professor of Marketing at the Franklin Perdue School of Business, Salisbury University, Maryland, USA. He holds a Phd in Marketing from the J Mack Robinson School of Business, Georgia State University in Atlanta. He is also the founding director of the Mid-Atlantic Sales and Marketing Institute (MASMI). He has also served as the chair of the Management and Marketing department at Salisbury University. He has published over 20 peer reviewed articles in *Journal of Business Research, Journal of Interactive Marketing, Journal of Marketing Theory and Practice, Journal of Services Marketing, Journal of Consumer Policy, Journal of Business and Industrial Marketing* among others. He has published papers on online marketing, public policy and trade promotions in B2B settings.

Syagnik Banerjee is an Associate Professor of Marketing at University of Michigan Flint and an Affiliate professor at Michigan Institute of Data Sciences, Ann Arbor. He graduated with a PhD from University of Rhode Island. Broadly, his research focuses on the transformative impact of mobile devices on consumer behavior, effectiveness of marketing practices, and location based market intelligence. Specifically, publications have been on location based advertising, ubiquitous consumption, dual consciousness in virtuo-physical environments, Internet of Things and Geo Social Insights for marketers. He has published in *Journal of Research in Interactive Marketing, International Journal of Mobile Communications, Journal of Marketing Analytics, and Journal of Business Research* among others.

Karthik Sridhar is an Assistant Professor of Marketing at the Zicklin School of Business, Baruch College – The City University of New York. He holds a Ph.D. in Marketing from the School of Management, State University of New York at Buffalo and a B.S. in Electronics Engineering from Anna University, India. His research interests lie in areas relating to consumer shopping behavior and consumer learning, contemporary retail promotions, new product development and health product marketing. He is also explores topics relating to social media marketing, digital marketing and mobile advertising. He is skilled at handling and merging large databases as part of his research. He also has skills in the empirical application of quantitative models to study the impact of various factors such as retail-format, pricing, health perceptions and more recently, direct marketing communications and social media, on consumer choice. He has published in journals such as *Marketing Science, Journal of Retailing* and *Journal of Retailing and Consumer Services*.