Ambivalence and Electronic Word of Mouth

Emergent Research Forum (ERF)¹

Mehmet Akgul
McMaster University
akgulm@mcmaster.ca

Ali Reza Montazemi
McMaster University
montazem@mcmaster.ca

ABSTRACT

Electronic word of mouth (eWoM) communications are online consumer-generated reviews that affect other consumers’ perceptions of adopting pertinent services. The major component of eWoM is the sentiment portrayed in form of text by the sender of the eWoM to enlighten the receiver of eWoM about the nature of a focal goods/services. It is customary to have the sender of eWoM message to also provide his/her overall attitude in form of a bipolar measure (e.g., star-rating). However, research into ambivalence spawned from the observation that traditional bipolar measures of attitude fail to distinguish between ambivalence and indifference. This paper explores the possible discrepancies that arise between using the sentiment of the eWoM text message versus overall attitude denoted by the star rating within the context of eWoM for restaurants in Yelp.com.

Keywords

Electronic word of mouth, eWoM, ambivalence, ambivalent attitude

Introduction

Electronic word-of-mouth (eWoM) reviews are online consumer-generated reports posted on a service providers’ or third party’s websites (Mudambi and Schuff, 2010). The significance of eWoM reviews is its effect on consumers’ service adoption (Godes and Mayzlin, 2009, Li and Hitt, 2008). One of the most widely used measurement of the eWoM message content has been use of star rating provided by the reviewer to show his/her attitude towards the reviewed product/service. This rating is then used to assess the usefulness of attributes such as message valance (i.e., positive, negative or neutral) on the receiver of the eWoM message to adopt goods/services. For example, the star rating of the eWoM message on Yelp.com has been used to categorize the message as positive, negative or neutral (e.g., Chen and Lurie, 2013): rating 1-2 used as negative, rating 3 as neutral, and 4-5 rating as positive. Such a measurement assume that the star rating accurately measure the sentiment of eWoM message that is questionable (van Harreveld et al, 2015). Furthermore, it is assumed that for example a rating between 1-2 has no positive content/sentiment or rating between 4-5 have no negative content/sentiment. We believe that such assumptions are flawed. Instead, we propose that sentiment of the eWoM message (text) provides a fertile ground to assess the nature of the message. After all, it is the content of the message that is expected to evaluate the characteristics of the focal product/service. In turn, communication of the message as eWoM is expected to persuade the receiver to decide (positive/negative) in adopting the pertinent goods/services.

In this research, we focus on the nature of reviews in which the reviewers’ rating is not compatible with the written reviews’ sentiment. We are interested in exploring the circumstances under which such variations occur and whether it can be justified in terms of reviewers’ ambivalent attitude. To that end, we

¹ Authors are equal contributors
apply exploratory methodologies commonly used in the extant literature (e.g., Luca and Zervas, 2016) to shed light on our research question.

**Theoretical Foundation**

Research into ambivalence spawned from the observation that traditional bipolar measures of attitude (e.g., a semantic differential ranging from “good” to “bad”) fail to distinguish between ambivalence and indifference (van Harreveld et al, 2015). On such bipolar measures, both respondents who are torn between strong opposing evaluations and those who simply do not care will tick the midpoint of the bipolar scale (Klopfer and Madden, 1980), even though their evaluations are fundamentally different. Van Harreveld et al. (2015) contend that:

“Evaluation is one of the most pervasive concepts in psychology. Not only is nearly all cognition and perception evaluative in nature (Markus & Zajonc, 1985), evaluations often take place quickly and without requiring much cognitive effort (Bargh, Chaiken, Govender, & Pratto, 1992). Evaluations serve important functions (e.g., Katz, 1960), such as preparing and guiding our behavior (Allport, 1935). They are based on a relatively stable set of associations that together form an attitude (Cunningham, Zelazo, Packer, & van Bavel, 2007). When associations are of the same valence, i.e., either positive or negative, evaluations form quickly and are seemingly effortless guides of human behavior (Armitage & Conner, 2000; Bargh et al., 1992). Often, however, attitudes are made up of positive and negative associations. Such coexistence of positive and negative associations within one attitude is what we call ambivalence. (P.2)"

Associative structure of ambivalence is based on positive and negative association weights (objective ambivalence) (van Harreveld et al., 2015). Thus, existence of ambivalence is expected for the evaluation of services in form of eWoM reviews. Generally, it is assumed that the reviewer rating (e.g., start rating) of the reviews convey the reviewers’ sentiment portrayed in the message/review (objective ambivalence). To our knowledge, such an assumption is based on anecdotal evidence. To investigate our research question, we focus on the reviewers’ ambivalent attitude within the context of experience services. Experience services are characterized by the attributes that cannot be objectively evaluated, rather need to be experienced and subjectively evaluated by the consumers prior to adoption (Xiao and Benbasat, 2007). Examples of experience services include restaurant services (Luca and Zervas, 2016). Given that eWoM reviews are comprised of other consumers’ experiences, they better match the subjective information required for reducing consumers’ uncertainty towards evaluating experience services adoption decision processes (Fang et al., 2011). Our analysis investigates restaurants’ reviews from Yelp, which is the dominant review site for restaurants (Luca and Zervas, 2016).

**Methodology**

We apply sentiment analyses to assess the content of reviews. Sentiment analysis, or opinion mining, is an active area of study in the field of natural language processing that analyzes people’s opinions, sentiments, evaluations, attitudes, and emotions via the computational treatment of subjectivity in text (Hutto and Gilbert, 2014; Liang et al., 2015). There are two approaches in the sentiment analysis: machine learning and lexical methods (Annett and Kondrak, 2008; Hutto and Gilbert, 2014). In the machine learning case, the documents are first investigated and categorized into sentiments by experts. Then the system by using the labeled documents is trained to learn the relationship between the documents and the sentiments. If a new document is provided to the system, the system finds out the class (sentiment) of the new one. The main drawback of this method is that it requires experts to label the documents, which are time-consuming and may not be available all the time. In the lexical methods, various lexicons containing various words and their sentiments (e.g. positive, negative, and neutral) are being used in the determination of the sentiment of a document. In this case, no training is necessary and the system can immediately provide the sentiment of a given text. However, the success of such methods is highly dictionary dependent. We use the lexicon-based approach since there is no expert labeled data available for the Yelp restaurant reviews.

As indicated before, our analyses is based on Yelp.com restaurants” reviews. The data is publicly available and can be downloaded from https://www.yelp.com/dataset. The dataset include business profile, review,
reviewer’s profile and photo information in JSON format. For the purpose of this study we used the business profile and review data files. Business profile data file include id, name, address, Yelp-star rating, price range and category of the business. Review data file contains the consumers’ reviews about a specific business in text format, the reviewer-rating, associated business id and date of the reviews. We selected to compare restaurant reviews in three cities across Canada. The reason for selecting different cities to assess our analyses is that ambivalent attitude may vary across different cultures (van Harreveld et al. 2015). We only selected the texts in English and businesses with more than one review.

We apply the VADER sentiment analysis tool to measure the sentiments of the reviews. VADER is a python tool, developed by Hutto and Gilbert (2014) that enable the assessment of the sentiment of a sentence and its intensity. While it can determine the sentiment of long texts, it is suggested to use the tool at the sentence level sentiment determination and then use their average for analyses. The output of VADER is a continuous variable between -1 and +1 where -1 represents highly negative sentiment and +1 represents highly positive sentiment respectively. Such interpretation is inline with previous research related to ambivalence assessment (van Harreveld et al., 2004). Next, we provide our preliminary findings from Yelp restaurant reviews for three cities in Canada.

**Preliminary Analyses**

To compute the sentiment score for a review, we first calculated the sentiment score for each sentence and then took the average of the sentiments and assigned it as the sentiment score of the review. To determine if there are any differences in the reviewer-rating and his/her sentiment scores, we first converted the reviewer-rating from a scale of [1 to 5] to [-1 to 1] by using the linear transformation function f(x) = (x-3)/2. The function maps the reviewer-ratings [1 to 5] into the [-1 to1] interval. Next, we applied the paired t-test to find out if there is a significant difference in the means of the reviewer star ratings (R) and the sentiment scores of the pertinent reviews (S). We found that there is a significant difference in the means of the R and S (p < 0.001) when considering all the restaurants in the three cities.

Next, we applied the same procedure as described above and grouped the restaurants into two: GP-A contains the restaurants in which R and S are aligned (i.e., no significant difference between R and S) and GP-NA contains the restaurants where R and S are not aligned (i.e., significant difference between R and S). Such grouping enabled us to observe the differences of the restaurants in each group. Here we report initial findings from three cities in Canada: Toronto, Montréal, and Calgary. The number of restaurant for each group is shown in Table 1. Table 2 presents the mean values for Yelp-star rating, number of reviews and price range for the three cities.

<table>
<thead>
<tr>
<th>City</th>
<th>Dependent Variable</th>
<th>GP-A</th>
<th>GP-NA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toronto, ON</td>
<td>Overall Rating</td>
<td>3.41</td>
<td>3.47</td>
<td>6881</td>
</tr>
<tr>
<td></td>
<td>Number of Reviews</td>
<td>33.74</td>
<td>79.84</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Price Range</td>
<td>1.80</td>
<td>1.86</td>
<td></td>
</tr>
<tr>
<td>Montréal, QC</td>
<td>Overall Rating</td>
<td>3.51</td>
<td>3.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of Reviews</td>
<td>17.72</td>
<td>59.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Price Range</td>
<td>1.93</td>
<td>1.98</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Number of Restaurants in the Two Groups for the Three Cities
Ambivalence and eWoM

We analyzed the differences of the dependent variables individually for each group (depicted in Table 3) using ANOVA analysis and found that Yelp-star rating, the number of reviews and price range is significantly different between the two groups for all three cities.

<table>
<thead>
<tr>
<th>City</th>
<th>Dependent Variable</th>
<th>F-Score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toronto, ON</td>
<td>Yelp-Star Rating</td>
<td>F(1,6879) = 12.97</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Number of Reviews</td>
<td>F(1,6879) = 479.21</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Price Range</td>
<td>F(1,6879) = 15.53</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>Montréal, QC</td>
<td>Yelp-Star Rating</td>
<td>F(1,3072) = 351.76</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Number of Reviews</td>
<td>F(1,3072) = 205.15</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Price Range</td>
<td>F(1,3072) = 4.71</td>
<td>p &lt; 0.04</td>
</tr>
<tr>
<td>Calgary, AB</td>
<td>Yelp-Star Rating</td>
<td>F(1,2470) = 67.79</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Number of Reviews</td>
<td>F(1,2470) = 227.09</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Price Range</td>
<td>F(1,2470) = 6.14</td>
<td>p &lt; 0.02</td>
</tr>
</tbody>
</table>

Table 3 ANOVA results

ANOVA analyses of the mean values for Yelp-star rating, number of reviews and the price range show that Yelp-star ratings for the restaurants in all cities are significantly higher for GP-NA than for GP-A. Furthermore, the restaurants in GP-NA have more reviews than the restaurants in GP-A for all three cities. Their price range is also significantly different between the two groups of GP-NA and GP-A for the three cities.

Concluding Remarks

Our preliminary analyses show that reviewer-ratings and the review sentiment are significantly different for 40.3% of the restaurants. This can be explained by the extant literature on ambivalent attitude. Ambivalence goes from the objective state to the more subjective state when individuals experience their positive and negative associations as conflicted. Such conflict can result from various sources, such as engaging in introspection or having to trade off the positive and negative associations and come to one unequivocal evaluation. Having to make a binary choice can render the two evaluations incompatible and lead ambivalence to be experienced as negative arousal, uncertainty, and feelings of regret. Our preliminary analyses show that restaurants’ price and popularity (i.e., number of review) have effects on consumers’ ambivalent attitude on their eWoM reviews. Our future analyses will shed more light on the nature of factors affecting ambivalent attitude within the context of eWoM.

<table>
<thead>
<tr>
<th>City</th>
<th>Overall Rating</th>
<th>3.40</th>
<th>3.66</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Reviews</td>
<td>17.98</td>
<td>37.38</td>
</tr>
<tr>
<td></td>
<td>Price Range</td>
<td>1.88</td>
<td>1.95</td>
</tr>
</tbody>
</table>

Table 2 Mean values for Yelp-star rating, number of reviews and price range for each city
REFERENCES