Empirical Evaluation of Automated Sentiment Analysis as a Decision Aid

Research-in-Progress

Ozgur Turetken
Ryerson University
Toronto, Canada
turetken@ryerson.ca

Sameh Al Natour
Ryerson University
Toronto, Canada
salnatour@ryerson.ca

Abstract

Research has consistently shown that online word-of-mouth (WOM) plays an important role in shaping customer attitudes and behaviors. Yet, despite their documented utility, explicit user scores, such as star ratings have limitations in certain contexts. Automatic sentiment analysis (SA), an analytics technique that assesses the "tone" of text, has been proposed as a way to deal with these shortcomings. While extant research on SA has focused on issues surrounding the design of algorithms and output accuracy, this research-in-progress examines the behavioral and interface design issues in regards to SA scores as perceived by their intended users. Specifically, in an online context, we experimentally investigate the role of product (product category) and review characteristics (review extremity) in influencing the perceived usefulness of SA scores. Further, we investigate whether variations in how the SA scores are presented to the user, and the nature of the scores themselves further affect user perceptions.

Keywords: Sentiment analysis, online search behavior, Web 2.0, text mining

Introduction

People's decisions are influenced by others' opinions (Keller and Staelin 1987; Mudambi and Schuff 2013). Research has consistently shown that word-of-mouth (WOM) plays an important role in shaping consumer attitudes and behaviors (Buttle 1998; Chen and Xie 2008; Kumar and Benbasat 2006; Pavlou and Dimoka 2006). Hence, providing customers with tools that quickly capture the general sentiment of others' expressions, helps to reduce the effort in obtaining a critical input to the decision-making process. This is especially important in online consumer contexts where WOM, which often takes the form of product/service reviews, is considered an essential element of the online marketing mix (Chen and Xie 2008), as well as a significant input to the online purchasing decision-making process (Pavlou and Dimoka 2006), and potentially a determinant of online sales (Chevalier and Mayzlin 2006).

In a recent six-month survey (Bulbul et al. 2014), 74% of respondents indicated that online WOM is a major influencer when making purchase decision - surpassing the influence of store visits (69%) and company websites (59%). Similarly, in their most recent annual "Global Trust in Advertising" survey, Nielsen (2015) found that while the most credible source of product information remains to be friends and family, 66% of respondents trust consumer opinions posted online, which a larger proportion than those who trust traditional advertising. In another survey looking at how customers use online reviews to make decisions regarding local businesses (BrightLocal 2015), 92% of customers indicated that they read online reviews, and 80% trust them as much as personal recommendations. More than half stated that they read more than 3 reviews before making a decision.

To reduce the complexity associated with sifting through the potentially hundreds of available reviews, websites use review scores that are intended to provide an overall evaluation of the product/service across a single, or multiple dimensions (Zhu and Zhang 2010). This approach, however, has a number of shortcomings. Despite their documented utility, explicit user scores such as star ratings can have limitations in certain contexts. For example, more often than not, long reviews are accompanied by only a single rating assigned (Lak and Turetken 2014). In such a case, the challenge is deciding which part of the review the rating refers to, and what exactly to read. Meanwhile, many other useful sources such as blog posts, news websites, and twitter, which contain relevant information and have been shown to be effective mediums of WOM (Akehurst 2009; Chen, Fay, and Wang 2011), do not contain scores. This makes it difficult to incorporate such valuable sources of information into the decision-making by a typical consumer's.

There are additional difficulties associated specifically with the way review scores are employed today. First, these scores are only available if the online vendor allows for their inclusion when the customer submits the review, and only when the customer chooses to provide them. Second, these scores are often intended to capture ratings of functional aspects of the product/service, rather than qualitative assessments such as customers' attitudes and sentiments. Finally, it is difficult to integrate reviews from different sources that may use different rating scales.

Automatic sentiment analysis (SA), an analytics technique that assesses the "tone" of text, has been proposed as a way to deal with these shortcomings (Turetken and Olfman 2013). Through SA, sentiments of individuals expressing their opinions can be quickly identified, hence lowering the need for labor intensive manual analysis. In online contexts, SA has been used to mine user-generated content, such as social networking posts (e.g., Mostafa 2013) and blogs and forums (e.g., Pang and Lee 2008; Ye et al. 2009); often producing results that are not significantly different from manually identified sentiments (Jansen et al. 2009). Yet, a research gap that remains is whether SA scores have utility for decision-making, and if so, how they should be structured and presented to improve the decision-making process and outcomes.

To fill this research gap, we critically investigate SA effectiveness as a decision support tool. Specifically, we attempt to answer the following research questions:

1) What is the role of review characteristics, such as their overall sentiment of (predominantly positive or negative) in affecting the perceived utility of SA scores?
2) Do SA scores have varying utility for different product categories? If yes, in which product contexts is the inclusion of SA scores most useful?
3) What are the most effective means of representing SA scores, both in terms of physical representation and the number of dimensions?

To answer the first question, we examine whether varying the proportion of positive/negative reviews can enhance/reduce the perceived usefulness of SA scores. To answer the second question, we examine whether providing SA scores has equal utility for experience (e.g., restaurants) versus search (e.g., electronics) products. To answer the last question, we examine the representation of SA scores in different forms such as stars, emoticons, and number scores, so we can evaluate the most effective means for designing SA tools and their output. Furthermore, since sentiments are essentially the affective parts of opinions, they rarely are fully captured by an overall unidimensional score (Kim and Hovy 2004). Hence, we also investigate whether using SA scores that are multidimensional in nature increases their utility.

The answers to these questions have both theoretical and practical implications. First, this research will help us understand how SA scores are perceived and processed by customers, and integrated into their decision-making. Second, from a practical perspective, the proposed research offers actionable guidelines on how SA scores can be integrated into web interfaces, and how they should be presented.

The remainder of this paper is organized as follows. The following section offers a review of the literature on sentiment analysis, especially in consumer contexts. Section 3 presents the research model and develops the hypotheses. The research method is described in section 4, which is followed by some concluding remarks.

Literature Review: Sentiment Analysis

The new wave in Internet use, commonly referred to as Web 2.0 or the social web, revolutionized the way people create, disseminate, and use information (Turetken and Olfman 2013). This ushers in a variety of opportunities for people to engage in near-real time use and generation of information. Naturally, the resulting amount of information is even larger than before; therefore the information overload problem exists in a more severe form. In essence, the sheer amount of information that is available to users is simply beyond their capabilities to process, i.e. filter, analyze, consume, and understand (Poston and Speier 2005; Turetken and Sharda 2005). Despite technological and managerial advances to alleviate this “mega-problem”, it is only getting more complicated manifesting itself in new domains, and with domain-specific challenges and opportunities (Baek, Ahn, and Choi 2012; Park and Lee 2009; Zhu and Zhang 2010).

To deal with this information overload problem, some have proposed prototype systems, often using clustering techniques, which expose users to larger and richer information collections while reducing the potential for information overload (e.g., Bai 2011; Huang, Tu, Fu, and Amanzadeh 2013; Liu, Li, Zhou, and Xiong 2011; Ye, Zhang, and Law 2009). Studies testing such systems found that the while strength of people’s opinions can be changed by reading relevant the resultant information (Turetken and Sharda 2005), there is only a weak support for the effectiveness of clustering information content (Schuff, Turetken, and Louis 2011).

Just as clustering provides cues into the content of a large collection of text by indicating semantic similarities among various subsets of that content, sentiment analysis (SA) provides cues about the “valance” or “tone” of text by identifying semantic patterns and classifying text based on those patterns. In essence, SA algorithms use the frequency and positioning of sentiment words and their location in a piece of text to determine the sentiment rating of that text (Dang, Yulei, and Chen 2010). Hence, through SA, sentiments of individuals (content providers) expressing their opinions can quickly be identified, lowering the need for labor intensive manual analysis.

SA technology is grounded in the fact that sentiment words and their respective polarity in sentences can be identified with high accuracy, to around 82% (Hatzivassiloglou and McKeown 1997). SA systems vary in their unit of analysis (sentence, paragraph, whole document, etc.) and the precision of their output ranging from binary to completely analog values (Feldman 2013). SA is also influenced by the “target” of the analysis (e.g., a product or an office candidate), the topic discussed, and the characteristics of the text to be analyzed (Carvalho et al. 2011). While predominantly, the sentiments detected are expressed as a one dimensional score (positive to negative), there have also been attempts to identify a variety of sentiments such happy, fearful, disgusted, etc. typically using lexicon-based approaches (Liu, Lieberman, and Selker 2003) with extensions such as fuzzy logic (Subasic and Huettner 2001).

The results of SA (typically aggregated over large collections) can be helpful to make better decisions in significantly less time, by letting users more efficiently digest reviews and other relevant content (Huang et al. 2013). In fact, some went as far as suggesting that SA can be applied in domains ranging from the evaluation of consumer products, services, healthcare, and financial services to the analysis of social events and political elections (Liu 2012). Liebmann et al. (2012) used SA to summarize unstructured e-commerce data used by financial analysts and investors, and found that while SA results were helpful to explain the behavior of both groups, the influence of SA was group-dependent. Others have used them to mine web 2.0 content, such as tweets (e.g., Mostafa 2013) and blogs and forums (e.g., Pang and Lee 2008; Ye et al. 2009). The findings suggest that, in general, SA output can be used as a reliable method in analyzing attitudes towards products and global brands (Mostafa 2013). In Zhao et al. (2012), categorization of tweets into the categories of “angry”, “disgusting”, “joyful” and “sad” led to the identification of certain mood patterns and abnormal events according to those patterns.

This brief review reveals that other than a few studies (e.g. Liebmann et al. 2012; Zhao et al. 2012) the major focus of SA research has been on algorithm development and testing (e.g. Dang, Yulei, and Chen 2010). These tests typically show high overall accuracy, but without deep analysis of where the shortcomings of these algorithms are. More importantly, while SA has been hailed as a tool to assist decision-makers, scant research has been conducted to examine its promised efficacy at doing so. In contrast, for the most part, previous research has taken the value of SA for information users (consumers) for granted. We believe this is a gap in the research on SA, because the way sentiments influence decision makers is more nuanced (e.g. Pavlou and Dimoka 2006) than the way it has been treated in the literature. As discussed in the next section, this research aims to address those shortcomings.

Research Model and Hypotheses

Given the objectives of this research, we adopt a behavioral and an empirical approach, rather than a design science one. Hence, we are more concerned with understanding, from the user's perspective, the impact of SA and its design on decision-making (a human-computer interaction view). We hence anchor our theoretical framework in multidisciplinary theories, such as theories of information processing (e.g., Christensen-Szalanski 1980; Pirolli and Card 1999) and media richness (Dennis and Kinney 1998). Collectively, these theories help us evaluate the potential effects of SA utilization in consumer contexts, and guide our inquiry as to how SA outputs should be designed and represented. The latter goal of understanding the role of design of a technology on perceptions and behavior of its users, has been argued should be major focus of the field's research (Benbasat and Barki 2007; Banbasat and Zmud 2003).

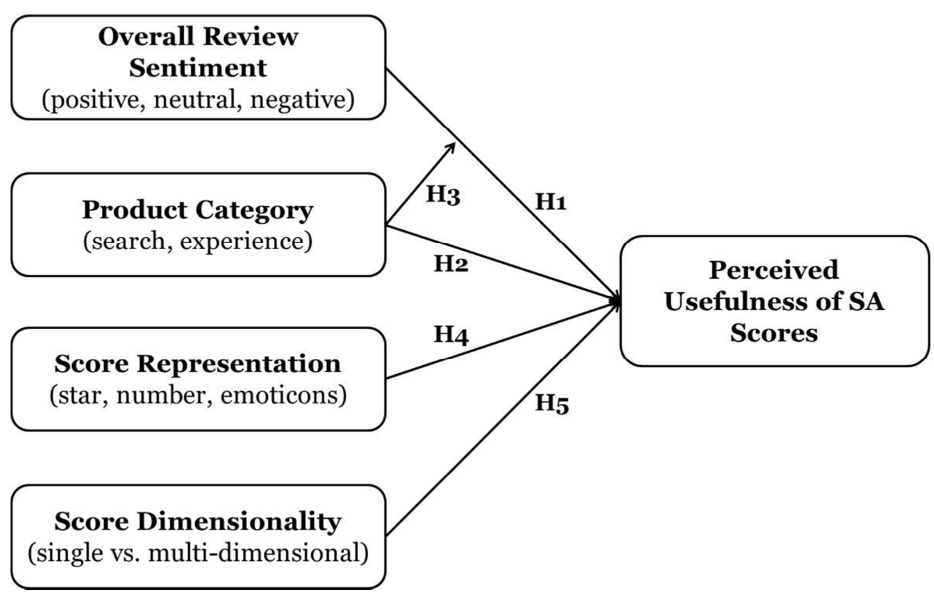

Figure 1. Research Model

The research model is depicted in Figure 1. It focuses on the effectiveness of SA in supporting consumer decisions, where a consumer is broadly defined as anyone who is interested in purchasing a product or service. Specifically, we examine the effects of review characteristics and product category type (Nelson 1970) on consumer perceptions. Subsequently, we examine the role of different visual and qualitative representations of SA scores on their utility.

Based on theories of information processing, specifically the accuracy-effort framework (Payne, Bettman, and Johnson 1993), it is easy to see how SA scores can act as a signaling mechanism in so much as they indicate the overall sentiment of another's review without the need to read it. They, hence, allow customers to reduce the amount of processing needed (effort) without sacrificing accuracy. This proposition has received preliminary empirical support (Lak and Turetken 2014). Further to that, we propose that the inclusion of SA scores has an additional important effect. Specifically, they amplify the overall valence (positive, negative) of textual reviews by biasing customers to focus on extreme reviews.

Past research has shown that the nature of ratings, whether extreme or not, can play an important role in determining customers' perceptions of review usefulness (Mudambi and Schuff 2013), as do overly positive and overly negative reviews (Schlosser 2011). Past research has also indicated that the variance of ratings and their strength influence product sales (Chevalier and Mayzlin 2006; Clemons, Gao, and Hitt 2006). Specifically, the variance of ratings and the strength of the most positive quartile of reviews were found to play a significant role in determining which new products sales grow the fastest. Similar research has also indicated that low-involvement consumers tend to conform to the perspective of negative reviews regardless of their quality (Lee, Park, and Han 2008).

Guided by the information foraging theory (Pirolli and Card 1999), we propose that SA scores are considered more useful when the overall sentiment of the reviews is positive/negative compared to when they are neutral. Information foraging theory (Pirolli and Card 1999) argues that information seeking behavior in humans is similar to food foraging behavior in animals. The theory assumes that individuals, when possible, will modify their strategies of acquiring information, or the structure of the environment, to maximize their rate of gaining valuable information. It argues that cues in the immediate environment of information presentation will let out a scent about the nature of information. This scent will then direct the user to choose and pursue that source of information, or ignore it for a more promising one. In the context of this study, SA scores will act as the "scent", directing customers' focus to more extreme reviews, and consequently, increasing the perceived utility of these scores.

H1: SA scores are perceived to be more useful when the overall sentiment of the reviews is overly positive/negative compared to when they are neutral.

This study also investigates whether the type of product affects the perceived usefulness of SA scores. Specifically, we examine whether SA scores when included as a part of reviews for search products/services (for which characteristics are easily evaluated before purchase) are perceived to be equally useful to when they are included as a part of reviews for experience ones (characteristics are difficult to ascertain before purchase).

Past research has indicated that reviews are not equally influential in all product/service contexts (e.g., Bei, Chen, and Widdows 2004; Mudambi and Schuff 2013; Zhu and Zhang 2010). In general, reviews are found to be more helpful when they concern an experience product/service than a search one (Bei et al. 2004). This can be attributed to the often sufficiency of retailer/manufacturer product information when making decisions about search products, and therefore the need for consumer reviews is minimized. Alternatively, experience products are hard to describe, and their quality is difficult to determine without actual use (Nelson 1970). Hence, the need for first-hand experience communicated through reviews is amplified (Bei et al. 2004).

Similarly, we propose that the type of product (search vs. experience) will influence customers' evaluations of SA scores. As suggested by information forging theory, information seekers seek to maximize the rate at which they gain valuable information (Pirolli and Card 1999). In the case of search products, the information seeking returns are maximized by primarily focusing on official retailer/manufacturer product information, since they are often complete, clear, and easy to find. In the case of search products, first-hand experience can only be attained through consumer reviews, and since the information seeker is seeking to maximize the rate of acquiring that information, he/she is more likely to focus on SA scores and perceive them to be useful. Hence, we propose that the effects of SA scores is diminished for search products, and amplified for experience products.

H2: SA scores are perceived to be more useful for experience products compared to search product.

Past research has also shown that product category type moderates the effects of review extremity on the helpfulness of a review. For experience goods, reviews with extreme ratings are less helpful than reviews with moderate ratings (Mudambi and Schuff 2013). Hence, in addition to its direct effects, we also propose that product category type moderates the effects of overall review sentiment. Specifically, guided by research that has shown that formal information resources, such as retailer/manufacturer websites, are of equal importance to that of customer reviews when shopping for search products (Bei et al. 2004), we expect that the previously discussed effect of overall review sentiment to be more pronounced in the case of experience products compared to search ones. Again such an effect can be explained via information forging theory (Pirolli and Card 1999), where users are expected to maximize the rate of gaining valuable information by focusing on extreme reviews, especially in product contexts where first-hand user experiences are more important.

H3: Product category moderates the effects of the overall sentiment of the reviews, where overly positive/negative reviews will be perceived to be more useful for experience products compared to search products.

This study also investigates if the representation of SA scores and their qualitative nature affects their utility. Specifically, first, we examine different forms for representing SA scores and measure their effects on their perceived utility. Our investigation is guided by findings based on media richness theory (e.g., Dennis and Kinney 1998), which indicate that the form in which a message is encoded can have significant effects on how it is perceived and processed. Of relevance is research suggesting that visual cues, such as emoticons (representations of a facial expression) (Huang, Yen, and Zhang 2008), are best suited to augment the meaning of textual messages (e.g., Walther 2006). Hence, we propose that the format for representing SA scores has an impact on their perceived utility.

H4: SA score physical representation will affect perceptions of their perceived usefulness.

While displaying a single overall score can quickly assist the user in determining whether the review is worthy of reading, representing SA scores as multidimensional based on the different sentiments included in the review (and which the current SA technology can accomplish) is more informative. Particularly, it aids the customer to efficiently understand the reasoning behind the negative/positive review, likely without the need to read the review in detail. In other words, a single review can have multiple dimensions, and each one of these dimensions can be assigned its own SA score (hence, resulting in multiple sentiments for each review). Subsequently, such a multidimensional score will significantly increase accuracy by allowing the customer to gain an understanding of the review's essential elements, without adding too much additional effort. Hence, we propose that offering a refined, and therefore, a more informative SA score based on the different sentiments included in the review, will increase the perceived utility of SA scores.

This view is consistent with research in psychology, which has suggested that human sentiments are multidimensional. Examples include the five major sentiments that all humans express and understand, namely happiness, sadness, anger, disgust, and fear. Moreover, it has been shown that the effects of these sentiments are similar regardless of culture (Ekman 1992). Such sentiments can guide human behavior, even though different sentiments with opposite valence could have similar consequences (e.g., both happiness and anger increase confidence, feeling of power, and decreases sensitivity to risk; Lerner and Tiedens 2006), and therefore can lead to similar behaviors (e.g., aggressive pricing of a product). As mentioned earlier, in recent research (Zhao, Dong, Wu, and Xu 2012), categorization of tweets into "angry", "disgusting", "joyful" and "sad" led to the identification of certain mood patterns and abnormal events according to those patterns. We hence propose that multidimensional scores will affect customers more profoundly, and hence, perceived to be more useful.

H5: Multidimensional SA scores are perceived to be more useful.

Method

A mixed factorial experimental design with 3 (review sentiment: positive, neutral, negative) x 2 (product category: experience or search) x 3 (SA score representation: star rating, numbered score, emoticons) x 2 (SA score dimensionality: single score or multidimensional score) conditions will be used to test the research model. Review sentiment will be the within-subjects factor. Participants will be randomly assigned to the experimental groups.

Experimental Task

Prior to starting the shopping task, participants will receive a training session, during which they will familiarize themselves with the experimental interface and the tools available on the website. Participants will be given the task of browsing three websites that sell laptop computers (for search products) or travel packages (for experience products) and indicate their top three product choices. Additional product categories may be identified in our pilot tests for increased generalizability. Three website are used to manipulate the within-subjects factor.

To control for differences in subjects' preferences, all subjects will be told that they are shopping for a fictional friend. The laptop needs and the general vacation preferences of this fictional friend will be presented to subjects so they can read them before starting the shopping task. This approach has been successfully used in prior similar studies (e.g., Al-Natour, Benbasat, and Cenfetelli 2011). To increase participants' motivation, and enhance the need to look carefully at product reviews, participants will be informed that a number of performance-based cash prizes will be distributed. These will be adjudicated by a panel of judges based on the subjects' written justification of their product choices; which they will submit at the end of the task.

The experimental websites will be identical except in relation to the manipulated factors. All product information and product reviews will be real. This can be accomplished by creating affiliate websites that stream actual data from real websites. Specifically, for search products, and through functionality offered by Amazon Web Services, we will create our own store that features Amazon products of our choosing. We further can limit the number of available products and whether and how the search results, and product and review information, are displayed. For the experience product, a similar type of an affiliate website that uses the Kayak hotel and vacation search engine will be created (the Kayak search engine is used by many travel websites).

Upon completing the shopping task, all participants will be asked to complete an online questionnaire using Qualtrics that measures the variables (e.g., perceived usefulness) with instruments adapted from the literature. The questionnaire will also collect background data (e.g., online purchase experience, disposition to trust online reviews, domain expertise … etc.), which will be used as control variables and to ensure that there are no difference amongst the groups.

Treatment Groups

To manipulate the first factor concerning the overall sentiment of reviews, we will increase the proportion of the included reviews that are positive in nature (i.e., receiving a score of more than 3 out of 5) for the overall positive sentiment condition, and the reviews that are generally negative in nature (receiving a score of less than 3) for the negative. Alternatively, for the neutral condition, the included reviews will be chosen randomly.

Product category will be manipulated via the type of product the participants are tasked to shop for. The two products, namely laptops and travel packages were chosen as they have been used in prior research to represent search and experience products respectively (e.g., Al-Natour, Benbasat, and Cenfetelli 2008; Bei et al. 2004). Furthermore, both products are relatively expensive and of equivalent value, ensuring a high level of involvement on the part of the participants. Finally, both products are consistently ranked as top-selling product categories online (eMarketer 2014).

SA Score Representation is manipulated by representing SA scores in three different ways: 1) star ratings, 2) numbered scores, and 3) emoticons. To ensure that the representations are equivalent, the number and the star ratings will be on a scale of 1 to 5, or 1 star to 5 stars, respectively. To manipulate the last factor (score dimensionality), participants will either see a multidimensional sentiment score (the dimensions will depend on the product category) or a unidimensional score. Table 1 depicts some examples of different types of rating scores currently used by large e-commerce companies.

Sample

All participants will complete the task online, and will be recruited through a marketing panel company (to enhance the results' generalizability). Email invitations will be sent to participants with instructions on how to complete the task and links to the fictional websites. They will be provided an honorarium for their participation, in addition to the performance-based cash prizes that will distributed.

Three hundred and sixty participants will be recruited for this study (20 per cell). This will result in power values of 97.9 for the interaction effects, and 99.2 for the main effects, assuming an effect size of 0.25.

The effects of the experimental factors will be tested through ANOVA (ANCOVA will be performed to test for the effects of the control variables). To test for the moderation effects, we will analyze the relevant two way interactions (e.g., the interaction between overall review sentiment and product category to test H3).

Table 1. Sample SA Scores

Type of SA Score	Possible Representation
Start Rating	 Source: Yelp.com
Number Rating	 Source: Zomato.com
Emoticon Rating	 Source: Facebook.com
Multiple Dimensions SA Score (note: the dimensions do not represent sentiments as intended for this study)	 Source: TripAdvisor.com

Pilot Testing

A number of pilot tests will be conducted. One pilot study will test whether a basic SA system (unidimensional number-based representation) is perceived to be more useful than the traditional star-based review systems currently available online. Based on the results from the first pilot study, we will conduct a second pilot study to pre-test the instrument and the three proposed representations. In regards to the latter, we seek to find whether all three representations convey equivalent amounts of information. Based on the results of the pilot studies, modifications to the experimental design maybe required.

Concluding Remarks

This research in progress investigates the effects of SA scores for varying types of products, reviews, and score designs. In contrast to extant research on SA that has primarily adopted a design science approach, this research focuses on behavioral and HCI issues to understand the determinants of SA score utility, and the product and review context in which they are most useful. The potential applications of SA tools are plentiful. Yet, to rise to their potential, the output of these tools should be effectively-designed, and they should be applied in the correct contexts.

To answer the three research question posed, this paper proposes an experiment that examines whether varying the proportion of positive/negative reviews affects the perceived utility of SA scores. To investigate the role of context, we study products of various types. To understand the effects of their design on their perceived utility, we examine multiple forms of representing SA scores, and compare unidimensional and multidimensional ones.

The research in progress has the potential to make both theoretical and practical contributions. First, this research can help us understand how customers use SA scores, and the effects of these on their decision-making. Specifically, this study can allow a better understanding of how SA scores are incorporated into the decision-making process, and used by consumers in a variety of contexts. It also has the potential to add to our understanding of how online users process visual and textual cues and content, and evaluate multidimensional evaluative information. In a broader sense, it can enhance our understanding of information-seeking behavior online.

From a practical perspective, the research can offer website designers insights into how to develop SA tools to increase their utility. Specifically, it can offer guidelines regarding their representation, their nature, and contexts in which they can be most useful. While this study will be conducted in a commercial setting, our findings should apply to information users and service providers beyond consumers and businesses. For example, voters, policy makers, public administrators, and non-profit organizations are potential beneficiaries of this research program.

References

Akehurst, G. 2009. “User Generated Content: The Use of Blogs for Tourism Organisations and Tourism Consumers,” *Service Business* (3:1), pp. 51-61.

Al-Natour, S., Benbasat, I., and Cenfetelli, R. 2008. “The Effects of Process and Outcome Similarity on Users' Evaluations of Decision Aids,” *Decision Sciences* (39:2), pp. 175-211.

Al-Natour, S., Benbasat, I., and Cenfetelli, R. 2011. “The Adoption of Online Shopping Assistants: Perceived Similarity as an Antecedent to Evaluative Beliefs,” *Journal of the Association for Information Systems* (12:5), pp. 347-374.

Baek, H., Ahn, J., and Choi, Y. 2012. “Helpfulness of Online Consumer Reviews: Readers’ Objectives and Review Cues,” *International Journal of Electronic Commerce* (17:2), pp. 99-126.

Bai, X. 2011. “Predicting Consumer Sentiments from Online Text,” *Decision Support Systems* (50:4), pp. 732-742.

Bei, L. T., Chen, E. Y., and Widdows, R. 2004. “Consumers’ Online Information Search Behavior and the Phenomenon of Search vs. Experience Products,” *Journal of Family and Economic Issues* (25:4), pp. 449-467.

Benbasat, I., and Barki, H. 2007. “Quo vadis TAM?” *Journal of the Association for Information Systems* (8:4), pp. 211-218.

Benbasat, I., and Zmud, R. W. 2003. “The Identity Crisis Within the IS Discipline: Defining and Communicating the Discipline's Core Properties,” *MIS Quarterly* (27:2), PP. 183-194.

BrightLocal. 2015. “Local Consumer Review Survey,” Available at: https://www.brightlocal.com/learn/local-consumer-review-survey. Last accessed: May, 2016.

Bulbul, C., Gross, N., Shin, S. and Katz, J. 2014. “When the Path to Purchase becomes the Path to Purpose,” Available at: http://think.storage.googleapis.com/docs/the-path-to-purpose_articles.pdf. Last accessed: May, 2016.

Buttle, F. A. 1998. “Word of Mouth: Understanding and Managing Referral Marketing,” *Journal of Strategic Marketing* (6:3), pp. 241-254.

Carvalho, P., Sarmento, L., Teixeira, J., and Silva, M. J. 2011. “Liars and Saviors in a Sentiment Annotated Corpus of Comments to Political Debates,” in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, Portland, OR, pp. 564-568.

Chen, Y., and Xie, J. 2008. “Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix,” *Management Science* (54:3), pp. 477-491.

Chen, Y., Fay, S., and Wang, Q. 2011. “The Role of Marketing in Social Media: How Online Consumer Reviews Evolve,” *Journal of Interactive Marketing* (25:2), pp. 85-94.

Chevalier, J. A., and Mayzlin, D. 2006. “The Effect of Word of Mouth on Sales: Online Book Reviews,” *Journal of Marketing Research* (43:3), pp. 345-354.

Christensen-Szalanski, J. J. 1980. “A Further Examination of the Selection of Problem-Solving Strategies: The Effects of Deadlines and Analytic Aptitudes,” *Organizational Behavior and Human Performance* (25:1), pp. 107-122.

Clemons, E. K., Gao, G. G., and Hitt, L. M. 2006. “When Online Reviews Meet Hhyperdifferentiation: A Study of the Craft Beer Industry,” *Journal of Management Information Systems* (23:2), pp. 149-171.

Dang, Y., Zhang, Y., and Chen, H. 2010. “A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews,” *IEEE Intelligent Systems* (25:4), pp. 46-53.

Dennis, A. R., and Kinney, S. T. 1998. “Testing Media Richness Theory in the New Media: The Effects of Cues, Feedback, and Task Equivocality,” *Information Systems Research* (9:3), pp. 256-274.

Ekman, P. 1992. An argument for basic emotions. Cognition and emotion, 6(3-4), pp. 169-200.

eMarketer 2014. “Total US Retail Sales Top $4.5 Trillion in 2013, Outpace GDP Growth,” Available at: http://www.emarketer.com/Article.aspx?R=1010756. Last accessed: May, 2016.

Feldman, R. 2013. “Techniques and Applications for Sentiment Analysis,” *Communications of the ACM* (56:4), pp. 82-89.

Hatzivassiloglou, V., and McKeown, K. R. 1997. “Predicting the Semantic Orientation of Adjectives,” in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics (eds.)*, Somerset, NJ, pp. 174-181.

Huang, A. H., Yen, D. C., and Zhang, X. 2008. “Exploring the Potential Effects of Emoticons,” *Information and Management* (45:7), pp. 466-473.

Huang, S. W., Tu, P. F., Fu, W. T., and Amanzadeh, M. 2013. “Leveraging the Crowd to Improve Feature-Sentiment Analysis of User Reviews,” in *Proceedings of the 2013 International Conference on Intelligent User Interfaces,* Los Angeles, CA, pp. 3-14.

Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A. 2009. “Twitter Power: Tweets as Electronic Word of Mouth,” *Journal of the American Society for Information Science and Technology* (60:11), pp. 2169-2188.

Keller, K. L., and Staelin, R. 1987. “Effects of Quality and Quantity of Information on Decision Effectiveness,” *Journal of Consumer Research* (14:2), PP. 200-213.

Kim, S. M., and Hovy, E. 2004. “Determining the Sentiment of Opinions,” in *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland, Article 1367.

Kumar, N., and Benbasat, I. 2006. “The Influence of Recommendations and Consumer Reviews on Evaluations of Websites,” *Information Systems Research* (17:4), pp. 425-439.

Lak, P., and Turetken, O. 2014. “Star Ratings versus Sentiment Analysis -- A Comparison of Explicit and Implicit Measures of Opinions,” in P*roceedings of 47th Hawaii International Conference on System Sciences*, Waikoloa, HI, pp. 796-805.

Lee, J., Park, D. H., and Han, I. 2008. “The Effect of Negative Online Consumer Reviews on Product Attitude: An Information Processing View,” *Electronic Commerce Research and Applications* (7:3), pp. 341-352.

Liebmann, M., Hagenau, M., and Neumann, D. 2012. “Information Processing in Electronic Markets: Measuring Subjective Interpretation Using Sentiment Analysis,” in *Proceedings of the Thirty Third International Conference on Information Systems*, Orlando, FL.

Liu, B. 2012. “Sentiment Analysis and Opinion Mining,” *Synthesis Lectures on Human Language Technologies* (5:1), pp. 1-167.

Liu, H., Lieberman, H., and Selker, T. 2003. “A Model of Textual Affect Sensing Using Real-World Knowledge,” in *Proceedings of the 8th International Conference on Intelligent User Interfaces*, Miami, FL, pp. 125-132.

Liu, X., Li, K., Zhou, M., and Xiong, Z. 2011. “Collective Semantic Role Labeling for Tweets with Clustering,” in Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, pp. 1832-1837.

Mostafa, M. M. (2013). “More than Words: Social Networks’ Text Mining for Consumer Brand Sentiments,” *Expert Systems with Applications* (40:10), pp. 4241-4251.

Mudambi, S. M., and Schuff, D. 2010. “What Makes a Helpful Review? A Study of Customer Reviews on Amazon.com,” *MIS Quarterly* (34:1), pp. 185-200.

Nelson, P. 1970. “Information and Consumer Behavior,” *Journal of Political Economy* (78:2), pp. 311-329.

Nielsen, A. 2015. “Global Trust in Advertising,” Report available at nielsen.com.

Pak, A., and Paroubek, P. 2010. “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” in *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, Valletta, Malta, pp. 1320-1326.

Pang, B., and Lee, L. 2008. “Opinion Mining and Sentiment Analysis,” *Foundations and Trends in Information Retrieval* (2:1-2), pp. 1-135.

Park, D. H., and Lee, J. 2009. “eWOM Overload and its Effect on Consumer Behavioral Intention Depending on Consumer Involvement,” Electronic Commerce Research and Applications (7:4), pp. 386-398.

Pavlou, P. A., and Dimoka, A. 2006. “The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation,” *Information Systems Research* (17:4), pp. 392-414.

Payne, J. W., Bettman, J. R., and Johnson, E. J. 1993. *The Adaptive Decision Maker*, New York, NY: Cambridge University Press.

Pirolli, P., and Card, S. 1999. "Information Foraging," *Psychological Review* (106:4), pp. 643-675.

Poston, R. S., and Speier, C. 2005. "Effective use of Knowledge Management Systems: A Process Model of Content Ratings and Credibility Indicators," *MIS Quarterly* (29:2), pp. 221-244.

Schlosser, A. E. 2011. "Can Including Pros and Cons Increase the Helpfulness and Persuasiveness of Online Reviews? The Interactive Effects of Ratings and Arguments," *Journal of Consumer Psychology* (21:3), pp. 226-239.

Schuff, D., Turetken, O., and Louis, R. D. S. 2011. "Mitigating Selective Filtering's Polarizing Effect on Web 2.0 Content," in *Proceedings of the Seventeenth Americas Conference on Information Systems*, Detroit, MI.

Subasic, P., and Huettner, A. 2001. "Affect Analysis of Text Using Fuzzy Semantic Typing," *IEEE Transactions on Fuzzy Systems* (9:4), pp. 483-496.

Turetken, O., and Olfman, L. 2013. "Introduction to the Special Issue on Human-Computer Interaction in the Web 2.0 Era," *AIS Transactions on Human-Computer Interaction* (5:1), pp. 1-5.

Turetken, O., and Sharda, R. 2005. "Clustering-Based Visual Interfaces for Presentation of Web Search Results: An Empirical Investigation," *Information Systems Frontiers* (7:3), pp. 273-297.

Walther, J. B. 2006. "Nonverbal Dynamics in Computer-Mediated Communication or: (And the Net: ('S with You, :) and You :) Alone," in *The Sage Handbook of Nonverbal Communication*, V. Manusov and M. L. Patterson (eds.), Thousand Oaks, CA: Sage Publications, pp. 461-480.

Ye, Q., Zhang, Z., and Law, R. 2009. "Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches," *Expert Systems with Applications* (36:3), pp. 6527-6535.

Zhao, J., Dong, L., Wu, J., and Xu, K. 2012. "Moodlens: An Emoticon-based Sentiment Analysis System for Chinese Tweets," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Beijing, China, pp. 1528-1531.

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing* (74:2), pp. 133-148.