Consensus Analysis of Amazon Product Reviews

Smit Anand
North Carolina State University
Email: sanand2@ncsu.edu

Mahnaz Behroozi
North Carolina State University
Email: mbehroo@ncsu.edu

Erick Draayer
North Carolina State University
Email: ecdraaye@ncsu.edu

Shama Yazdani
North Carolina State University
Email: syazdan@ncsu.edu

***Abstract*—The online shopping industry has seen remarkable growth in the past decade. As more customers shift from in-store shopping experiences to online browsing, there is greater need for a way to distinguish between high quality products and junk. This report explores the idea of using consensus analysis on the world's number one e-commerce store, Amazon. We begin by discussing one method typically used by consumers when shopping online–relying on previous customer reviews–and possible negative consequences thereof. We then present related literature that further explores the pitfalls of online shopping, and the promise and possible challenges of using consensus analysis to improve online shopping.**

***Keywords*—Consensus Analysis, Amazon, e-shopping, opinion mining, sentiment analysis**

I. Introduction

As we progress through the Information Age, computers become more ingrained in everyday tasks and transform the way we complete them. Perhaps one of the biggest changes has been in how we shop for and buy goods. Large department stores such as Sears, Macy's, and Dick's Sporting Goods are just a few among many that are failing to survive as more and more people look to online retailers for goods. Based on a survey published in Wall Street Journals on June 8, 2016, consumers are moving more towards online shopping, and thus moving towards a customer experience which is in many ways different from the traditional in-store experience. As part of this new experience, consumers are implementing new methods of selection when choosing which products to purchase.

We conducted a small pilot survey to assess the habits of online shoppers, including which online stores they prefer to use, what types of products they buy, and the importance of other users' product reviews in their own decision making process. We found that the most popular online store by far was Amazon, and that most online purchases were in electronics, cell phones, health and beauty products, and books. We also found that the majority of people find user reviews more important in their buying decision than the overall rating of the product. Interestingly, we discovered that consumers place much more weight on user reviews when shopping on Amazon, as compared to Best Buy and Yelp. Based on this information, we decided to focus our system on improving the online shopping experience by optimizing the process of sorting through customer reviews on Amazon.

There are a lot of ways to enhance the costumers purchase experience based on user reviews including: mining the reviews and categorising them based on sentiment analysis, aspect-based sentiment analysis, detecting the fake reviews and extracting them from other honest reviews, etc.

To make this project benefit from user interactions and feed backs, we decided to concentrate more on sentiment analysis or aspect-based analysis approaches. Sentiment analysis is the process of extracting the general ideas, opinions, attitudes, and emotions from written language. Factoring in our decision to take this approach was the recent release of large, organized datasets of user reviews to the public by Amazon. With access to such large-scale data, we believe we'll be able to develop accurate sentiment classifiers.

Our software aims to detect overall opinions from large sets of reviews, such as whether a book has a dissatisfying ending or if an electronic device has a tendency to develop problems after purchase. We hope that shoppers will be able to use our software to quickly determine the general consensus of user reviews, allowing them to make better and safer purchases online. We also want to explore the challenges of implementing sentiment analysis, for possible use in other areas.

The rest of this paper surveys related studies and challenges, describes our approach to gather information about the user tastes in online shopping and the results and our strategy to enhance their experience and finally the conclusion of the report.

II. Related Studies and Challenges

Sentiment analysis has been an active area of research over the past decade and ,still, it has many open challenges which are not solved yet. A vast majority of the research done in this area has been on developing more accurate sentiment classifiers, usually involving supervised machine learning algorithms. Pang and Lee [1], Liu and Zhang [2], and Mohammad [3] give summaries of automatic classifiers, features, and datasets used to detect sentiment in their review surveys.

There are many related tasks that can be done with the help of sentiment analysis. Two of these tasks that are of our interest are: automatic stance detection and sentiment analysis towards aspects of an entity or in other words aspect-based sentiment analysis. Automatic stance detection is the task of automatically determining from text whether the author of the text is in positive, negative, or neutral stance towards a proposition or a target. Most of the work in automatic stance detection focused on two-sided debates [4], for example on congressional debates (Thomas et al., 2006) or debates in online forums (Somasundaran and Wiebe, 2009; Murakami and Raymond, 2010; Anand et al., 2011; Walker et al., 2012;

Hasan and Ng, 2013; Sridhar, Getoor, and Walker, 2014). Mohammad, Sobhani, and Kiritchenko [5] created the first dataset of tweets labeled for both stance and sentiment. More than 4000 tweets are annotated for whether one can deduce favorable or unfavorable stance towards one of five targets ‘Atheism’, ‘Climate Change is a Real Concern’, ‘Feminist Movement’, ‘Hillary Clinton’, and ‘Legalization of Abortion’. Each of these tweets is also annotated for whether the target of opinion expressed in the tweet is the same as the given target of interest. Finally, each tweet is annotated for whether it conveys positive, negative, or neutral sentiment. There are many applications which could benefit from the automatic stance detection, including information retrieval, textual entailment, or text summarization and in particular opinion summarization. Aspect-based sentiment analysis not only deals with stance detection but also tries to consider different aspects of the topic that the author is talking about. For example, a product review may speak positively about the costumer service but may speak negatively about the quality of the product. In this case, aspect-based sentiment analysis will not merge these two attitudes. There is now a growing amount of work in detecting aspects of products in text and also in determining sentiment towards these aspects. In 2014, a shared task was organized for detecting aspect sentiment in restaurant and laptop reviews [6]. The best performing systems had a strong sentence-level sentiment analysis system to which they added localization features so that more weight was given to sentiment features close to the mention of the aspect.

In the review article “Techniques and applications for sentiment analysis” Ronen Feldman focuses on the following major problems and the approaches to solve them in sentiment analysis: Document Level sentiment analysis- In this form of sentiment analysis, the assumption is made that the writer of the document gives an opinion on one major object in the whole document. The two main approaches to this problem is through supervised learning and unsupervised learning. The supervised learning approach assumes that the document is classified into a finite set of classes and all of these classes are available in the training data. A simple approach to this is classifying the document as positive or negative or extending the above approach to add a neutral class to the classifier. The documents can also be placed on a discrete numeric scale like the five star system used by amazon. Any of the common classifier like SVM, Naive Bayes or KNN can be used by this model. In the unsupervised learning approach the semantic orientation (SO) of specific phrases [7] plays an important role. The document is classified as positive if the average SO is above the threshold which is predefined, else the document is classified as negative. Sentence Level sentiment analysis- This approach to used to have a more fine grained view of the opinions expressed in the document as a single document may contain multiple opinions about the same entity [7]. Various assumption of this approach include assuming that the identity of the entity is known and each sentence expresses a single opinion. The second assumption is not a hard one as in the case of a subjective sentence, the sentence may be broken into phrases and then assumed that each phrase express a single opinion. Various methods like supervised learning approach [8], bootstrapping [9] and minimum cuts [10] have been proposed for sentence level sentiment analysis. Ronen has also emphasised that as suggested by research [11], some type of sentences including conditional sentences, sarcastic sentences and interrogative sentences must be handled by different strategies. Aspect based sentiment analysis- An entity is comprised of different attributes and when a review is given about it, it may contain analysis of its different attributes. The writer may not be talking about the entity as a whole every time but instead may be giving his views on different attributes of the entity. Classifying such reviews as a whole would miss the valuable information contained in it. An approach to solving this is to extract all noun phrases (NPs) and then keep just the NPs whose frequency is above some experimentally determined threshold [12],another approach is to reduce the noise in the found NPs [13]. Comparative sentiment analysis- Sometimes user do not directly provide their opinions about a product and they instead provide their opinion in comparison to some other product. Example of suh speech is “Nexus 6P has better features than Nexus 5X”. The goal of the sentiment analysis system in this case is to identify the sentences that contain comparative opinions, and to extract the preferred entity(-ies) in each opinion [7]. Jindal and Liu in their paper [14] found that 98% of comparative opinion can be covered using a relatively small number of words such as comparative adjectives, superlative adjectives and additional phrases.

In the paper titled ”Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank”, they introduced a novel method called recursive neural tensor network which is forked from recursive neural network. To train the model they used a treebank called Stanford sentiment treebank. Stanford sentiment treebank consist of labels for every syntactically plausible phrase in thousands of sentences, allowing them to train and evaluate compositional models [15]. They took corpuses from rottentomatoes.com which consist of movie reviews. The data was originally collected and published by Pang and Lee. The original dataset includes 10,662 sentences, half of which were considered positive and the other half negative [1]. The Stanford Parser (Klein and Manning, 2003) is used to parses all 10,662 sentences. In approximately 1,100 cases it splits the snippet into multiple sentences. They then used Amazon Mechanical Turk to label the resulting 215,154 phrases [15]. Recursive neural network is one of the simplest model of neural network family. It uses a parent child approach to create a tree. Let $R_2 = fn(a, R_1)$ is root of a and R_1 and $R_1 = fn(b, c)$ is root of b and c then using Recursive neural network model we can compute the parent vector using these equations. $R_1 = fn(W \begin{bmatrix} b \\ c \end{bmatrix})$

$$R_2 = fn(W \begin{bmatrix} a \\ R_1 \end{bmatrix})$$

$fn = tanh$ is a standard element wise non linearity. The parent vectors must be of the same dimensionality to be recursively compatible and be used as input to the next composition [15]. MV-RNN: Matrix-VectorRNN – The main idea of the MV-RNN is to represent every word and longer phrase in a parse tree as both a vector and a matrix. When two constituents are combined the matrix of one is multiplied with the vector of the other and vice versa. Hence, the compositional function is parametrized by the words that participate in it.

Each word's matrix is initialized as a $d * d$ identity matrix, plus a small amount of Gaussian noise. Similar to the random word vectors, the parameters of these matrices will be trained to minimize the classification error at each node [15]. RNTN: Recursive Neural Tensor Network – One problem with the MV-RNN is that the number of parameters becomes very large and depends on the size of the vocabulary. It would be cognitively more plausible if there was a single powerful composition function with a fixed number of parameters. The standard RNN is a good candidate for such a function [9]. However, in the standard RNN, the input vectors only implicitly interact through the nonlinearity (squashing) function. A more direct, possibly multiplicative, interaction would allow the model to have greater interactions between the input vectors [15]. The main advantage over the previous RNN model, which is a special case of the RNTN when V is set to 0, is that the tensor can directly relate input vectors. Intuitively, we can interpret each slice of the tensor as capturing a specific type of composition. An alternative to RNTNs would be to make the compositional function more powerful by adding a second neural network layer. However,initial experiments showed that it is hard to optimize this model and vector interactions are still more implicit than in the RNTN [15].

In the paper titled "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis", researchers from HP labs and UIC tried to analyze the variation of sentiment of people from their twitter feed. They proposed a method to analyze the sentiment based on opinion words in context. The approach generally uses a dictionary of opinion words to identify and determine sentiment orientation (positive, negative or neutral). The dictionary is called the opinion lexicon [16]. At the same time analyzing sentiment through opinion analysis based on dictionary is becoming problematic because people on twitter express their emotions in various ways like unstructured words e.g. i lovvveee my phone. If we try to analyze this sentence by matching words with opinion dictionary then we wont find anything since "lovvveeee" is not an English word. Same is the case with emoticons, there is no definite expressed words in English which explains in which scenario what is used. To overcome this method they proposed to compare those words with other tweets where this word has occurred and then analyze the sentiment of the tweet where that word is present then try to map the sentiment of that word and then calculate sentiment of the new tweet. This method generated good result because they followed two step process. First, they perform sentiment analysis at entity level. Second, their technique for polarity assignment is also different since they deal with three class of sentiment (positive, negative, neutral) [16].

In the paper "iFeel: A Web System that Compares and Combines Sentiment Analysis Methods" the authors have made a web application iFeel that is able to detect sentiments in any form of text including the the unstructured social media data [17]. This application is a combination of seven existing methods of sentiment analysis to yield high coverage and F-measure. This methods are SentiWordNet, Emoticons, PANAS-t, SASA, Happiness Index, SenticNet and SentiStrength. iFeel also provides the user to play with the above mentioned methods adjust accordingly and come up with a new combined method. Coverage and F-measure can be optimized according to the type of data being analyzed. The user can increase the weight of one of the methods like SenticNet when analyzing health related text or SASA for politics related text. iFeel serves as a very good platform for comparison of the existing methods too. iFeel allows the user to upload a file and then creates seven asynchronous threads for each of the methods. Since the output of all the methods are in different format, the combined method proposed by the author it greatly useful to combine the results. The Combined-Method in iFeel assigns weight to the output of all the methods corresponding to each line in input. This weight is determined by the F-measure or the weights given by the user. The final output of iFeel is the result of each existing methods in addition to the result of Combined-method. The output also consists of graphs for better visualization of the results.

III. User Survey

We created an online survey using Google Forms. Our survey contained ten questions. Users were asked how often they use three online retailers: Best Buy, Amazon, and Yelp. These three stores were used because they are the only retailers to make their datasets of reviews for product and location available. We also asked participants how important product reviews are to them when deciding whether to buy a product. We then asked them how frequently they shop online and what types of products they generally buy online . Finally, we asked if they had any ideas on how to make product reviews more useful for online shopping. In total we had twenty one participants for our online survey.

A. *Survey Data Analysis*

Our results show that 40% of participants use online shopping "very often", and 45% use online shopping "sometimes" indicating the high number of user group who prefer shopping online and may benefit from our product (See Fig. 1). We

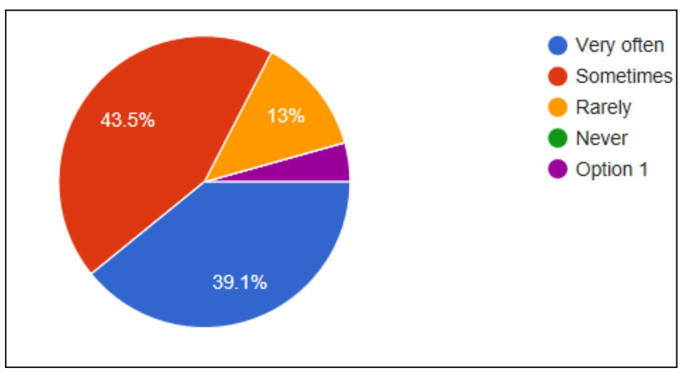

Figure 1: How often do people shop online

found that 75% of participants indicated that they actually buy something online at least once a month, with 15% buying something online at least once a week indicating the high frequency of usage to make online purchasing and therefore increasing the demand of a system that helps them make wiser decisions. Our survey also revealed not only that participants place great importance on product reviews when deciding whether to make a purchase, but that they appear to pay

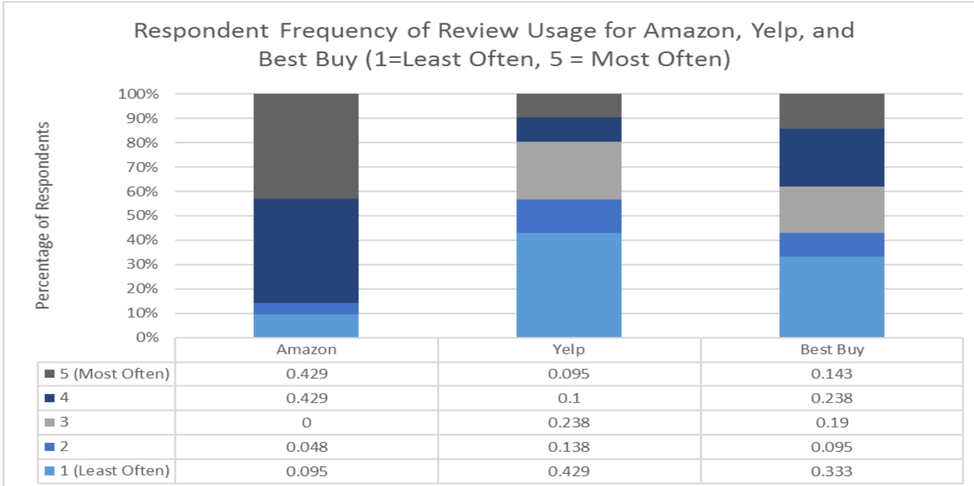

Figure 2: Frequency of review usage among websites.

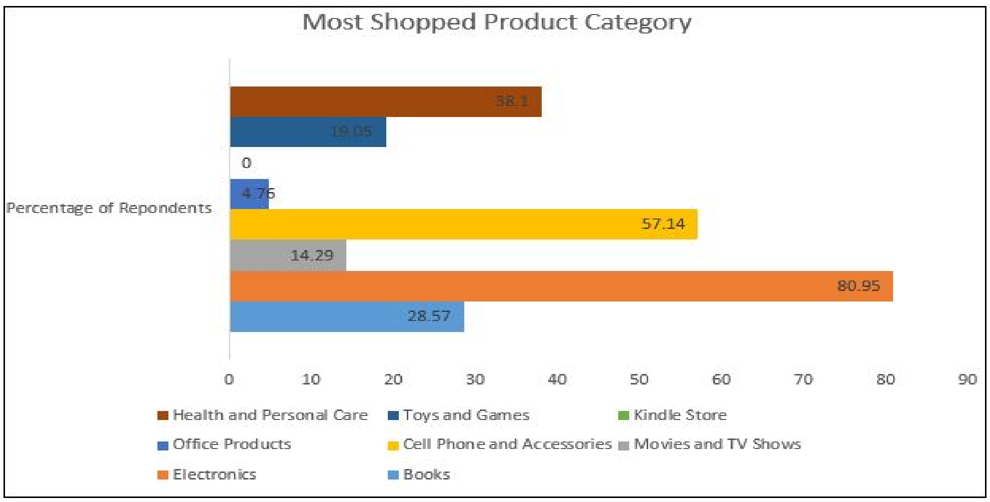

Figure 3: Types of products typically bought

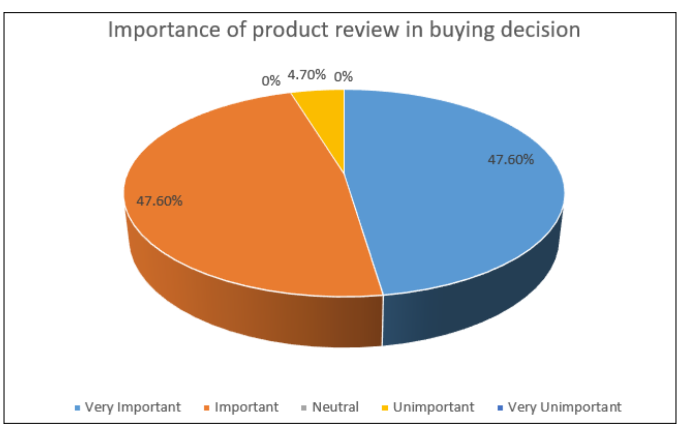

Figure 4: Importance of previous reviews on buying decision

more attention to reviews when browsing through Amazon in comparison to Yelp or Best Buy (See Fig. 2).

Among all the respondents, we found that the most popular products bought online by those polled were: electronics (80% of respondents), cell phone and cell phone accessories (57%), health and personal care products (38%), and books (28%) (See Fig. 3) . Lastly, the survey suggested that for 47% of people product review is very important to their buying decision and other 47% called it important (See Fig. 4)

In the free response section of the survey asking participants to suggest ideas regarding what they expect from a system that helps them make better buying decisions, our participants indicated that they would like to have previous negative comments displayed alongside previous positive comments for easier review. Other suggestions included displaying a summary of the differences between the product as advertised and the actual product experience after purchase, implementing a buyer certification system to lend credibility to reviews, and prioritizing reviews that include relevant photos.

Our survey results highly suggest that our system should focus on helping customers to read product reviews more easily because of the importance that customers place on reviews in their online shopping decisions. The survey revealed that people value customer reviews more when shopping through Amazon as compared to Best Buy and Yelp, so we decided to build our system for use with Amazon. Because survey respondents indicated that most of their online shopping revolves around electronics, we will first explore a software system designed specifically for assessing the reviews on electronics. Then we will adapt our software to work for all types of products.

IV. Conclusion

From our survey we discovered common behaviors among people buying products online. The first is that they prefer to use Amazon when browsing for goods or services. Another attribute we observed among our respondents is that they find Amazon user reviews the most helpful compared to other online companies. Our survey also revealed how important user reviews are to consumers when they are deciding what to buy. Lastly, our survey showed that the type of products people buy online tend to be electronics. From this we concluded that a system catered towards electronics which focuses on extracting overall sentiment from user reviews on Amazon would be the most helpful to online shoppers.

References

[1] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and Trends® in Information Retrieval 2.1–2 (2008): 1-135.

[2] Liu, Bing, and Lei Zhang. "A survey of opinion mining and sentiment analysis." Mining text data. Springer US, 2012. 415-463.

[3] Kiritchenko, Svetlana, and Saif M. Mohammad. "The effect of negators, modals, and degree adverbs on sentiment composition." Proceedings of NAACL-HLT. 2016.

[4] Mohammad, Saif M. "Challenges in sentiment analysis." A Practical Guide to Sentiment Analysis, D. Das, E. Cambria, and S. Bandyopadhyay, Eds. Springer (2016).

[5] Mohammad, Saif M., et al. "A dataset for detecting stance in tweets." Proceedings of 10th edition of the the Language Resources and Evaluation Conference (LREC), Portoroz, Slovenia. 2016.

[6] Pontiki, Maria, et al. "Semeval-2014 task 4: Aspect based sentiment analysis." Proceedings of SemEval (2014): 27-35.

[7] Feldman, Ronen. *T*echniques and applications for sentiment analysis Communications of the ACM 56.4 (2013): 82-89.

[8] Yu, H. and Hatzivassiloglou, V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (2003).

[9] Riloff, E. and Wiebe, J. Learning extraction patterns for subjective eExpressions. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (2003).

[10] Pang, B. and Lee, L. A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on minimum cuts. In Proceedings of the Association for Computational Linguistics (2004), 271–278.

[11] N arayanan, R., Liu, B. and Choudhary, A. Sentiment analysis of conditional sentences. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (Singapore, 2009). Association for Computational Linguistics, 180–189.

[12] Hu, M. and Liu, B. Mining and summarizing customer reviews. In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2004), 168–177.

[13] Popescu, A.-M. and Etzioni, O. Extracting product features and opinions from reviews. In Proceedings of Conference on Empirical Methods in Natural Language Processing (2005).

[14] Jindal, N. and Liu, B. Identifying comparative sentences in text documents. In Proceedings of ACM SIGIR Conf. on Research and Development in Information Retrieval (2006).

[15] Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment treebank." Proceedings of the conference on empirical methods in natural language processing (EMNLP). Vol. 1631. 2013.

[16] Khan, Aamera ZH, Mohammad Atique, and V. M. Thakare. "Combining lexicon-based and learning-based methods for Twitter sentiment analysis." International Journal of Electronics, Communication and Soft Computing Science and Engineering (IJECSCSE) (2015): 89.

[17] Araújo, Matheus, et al. "iFeel: a system that compares and combines sentiment analysis methods." Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014.

V. Chit Words

Following are the chit words provided by students of CSC510 in survey for the project:
dlflaauo
bjhvaaio
dchjeuue
chhwuoeo
ckhhoeiu
cljweeee
dhdkoaoe
bkkwouea
dfdfiooe
fgfrouua
fddzaaee