Proceedings of the ASME 2021
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC/CIE2021
August 17-20, 2021, Virtual, Online

[redacted]

ONLINE PRODUCT REVIEW ANALYSIS TO AUTOMATE THE EXTRACTION OF CUSTOMER REQUIREMENTS

[redacted]

ABSTRACT

The increasing use of online retail platforms has generated an enormous amount of textual data on the user experiences with these products in online reviews. These reviews provide a rich resource to elicit customer requirements for a category of products. The recent research has explored this possibility to some extent. The study reported here investigates the coding of publically available user reviews to understand customer sentiments on environmentally-friendly products. The manual review typically consists of a qualitative analysis of textual content, which is a resource-intensive process. An automated procedure based on Aspect-Based Sentiment Analysis (ABSA) is proposed and explored in this study. This procedure can be beneficial in analyzing reviews of products that belong to a specific category. As a case study, environmentally-friendly products are used. Manual content analysis and automated ABSA-based analysis are performed on the same review data to extract customer sentiments. The results show that we obtain over a 50% classification accuracy for a multiclass classification NLP task with a very elementary word vector-based model. The drop in accuracy (compared to human annotation) can be offset because an automated system is thousands of times faster than a human. Given enough data, it will perform better than its human counterpart in tasks on customer requirement modeling. We also discuss the future routes that can be taken to extend our system by leveraging more sophisticated paradigms and substantially improving our system's performance.

Keywords: Content Analysis, Customer requirements, Natural Language Processing

1. INTRODUCTION

Customer needs analysis has been considered a critical activity in the product development process. It has an impact on the activities downstream and the success of the product. In a typical product development project, customer needs are gathered with various tools such as interviews, focus groups, surveys, and observations [1]. Once the data collection is complete, these are interpreted and prioritized. These ranked customer needs are carried forward to further steps of the product development cycle. Each of these customer needs gathering techniques offer certain advantages and disadvantages, and combinations of these are necessary to gather a complete picture [1, 2].

The ability to purchase various products from online retailers has improved by a great extent in the previous decade. Many of these online platforms provide user-generated ratings for the products they sell. Besides, often, customers leave text-based reviews on the products they purchase and use. Such reviews can be a very rich resource for understanding the perceptions and expectations of the customer from various use situations [3]. While these online reviews can serve as ideal places to elicit customer needs, the literature has mostly overlooked this possibility until recently, as explained by Lee [4]. Most of the studies focus on identifying the product characteristics [5-7] for predicting the sales and success of a product with the help of reviews [8, 9]. The literature shows that the sales of a product are directly correlated to the average user rating of the product [10]. In the past few years, researchers have explored the possibility of automating the collection of consumer sentiments using the publically available textual information [e.g., 52 – 55].

In addition to customer needs assessment, several efforts have been reported to gather information from online reviews. A

[redacted]

prior study has used online reviews to study the influences of informativeness, readability, and subjectivity on the sales of a product [11]. Miao et al. developed a novel ranking mechanism that uses customers' temporal opinion quality and relevance to aid them in their decision-making process [12]. Achark et al. [13], used online reviews to study how the customers valued individual product features and how these features influenced the sales of the product. E.g., a study by Gretzel and Yoo [14] explored how the reviews on *TripAdvisor*, a popular travel review website, influenced people's decision to travel to various places. Qi et al. [44], developed an automatic filtering model to find the most helpful reviews within a list of reviews. Then it utilizes the reviews to develop a Kano model. Recently, Shanshan and Liu [46] proposed a machine learning algorithm that captured data from shopping patterns and experiences to then recommend products and shops to the users. Decker and Trusov [47] used an econometric preference analysis and binomial regression on online reviews for identifying the pros and cons of each product. Liu et al. [45], used a regression system algorithm and regressions to decipher which reviews from a set of Amazon reviews are useful to a designer. They found that the first few reviews were more useful than the rest.

One of the approaches followed for manually extracting customer requirements from a large body of the text of reviews is the qualitative content analysis process. It is a powerful method in finding relevant classifications of information from a variety of sources, including interview transcripts, reports, and online content. However, it is a very time-consuming and labor-intensive process. For identifying customer requirements, it is preferred to have a large number of reviews analyzed; however, the resource commitments prevent designers from utilizing large sample sizes.

Advancements in Natural Language Processing (NLP), especially in Fine-grained and Aspect-based Sentiment Analysis (ABSA), are powerful tools to automate the extraction of useful information from a corpus of text. In this paper, we present an experimental investigation to show that NLP and ABSA are an effective approach to extract insights about specific features of a class of products, called aspects, from massive amounts of online review. As compared to manual content analysis, this process can process a significantly larger number of reviews without misinterpreting them.

The first step in ABSA is to represent words as meaningful vectors, i.e., word embeddings. Words that appear in similar contexts have similar word embeddings. In our prior work, we have explored the impact of training word embeddings using domain-specific data [1]. In this work, we leverage our findings to automate the aspect detection and sentiment analysis for eco-friendly products.

Our contributions in this paper are as follows:

- We present a novel approach to extract aspects from customer reviews (with a case study on eco-friendly products) using five word-embeddings generated by training on varying fractions of domain-specific data.
- We adopt a rule-based sentiment analysis model to extract sentiments of customers towards each aspect and generate a quantitative measure of the sentiment.
- We compare the data generated by these algorithms to a small sample of data generated by a designer using the manual content analysis process to verify its validity.

2. BACKGROUND

As online platforms selling consumer products have become popular, many researchers have attempted to leverage the textual data available in the form of online product reviews. If one can extract useful information on customer needs from these online reviews, resource-intensive customer surveys can be avoided from the product design cycle. This section explores the manual content analysis procedure, the prior attempts on automated customer needs gathering from online reviews, and a literature review on the approach followed in this paper.

Content Analysis and Quantification of Customer Needs

Content analysis is a qualitative research method used to classify textual material into categories to identify the themes within the data. There are different types of content analysis methods; for this project, the method of conventional content analysis is utilized [17-19]. Conventional content analysis is used for studies that aim to describe a phenomenon; in this case, the phenomenon being the perception of consumers about eco-friendly products. Additionally, this approach is appropriate when existing research and theory are limited, as in this case.

Content analysis has four characteristics: objectiveness, systematicity, quantitativeness, and manifestness [20, 21]. Objectiveness requires that the criteria used to categorize content to be precise and avoid evaluative terms like "good-bad," "fair-unfair," etc., as these definitions are dynamic. Systematicity means the content to analyze must be gathered based on a formal and unbiased plan. Quantitativeness refers to the results of the analysis being expressed numerically, such as in ratios or percentages. A manifest analysis requires an emphasis on the message itself and not what shaped the message or the impact it may have.

Content analysis is composed of the following steps: formulating the research question, defining the sampling and data collection methods, and analysis [22, 23]. Once the data is gathered, it is essential to develop a coding scheme so that the content can be traced back to the source. E.g., in a study of consumer preferences for multipurpose products, once the text of a customer review is copied into one datasheet, it is assigned a code number that matches a product listed in a separate datasheet [16]. Coding consistency must be checked throughout the project as an iterative exercise. Iterative checking prevents the creep of inconsistencies in coding when dealing with large contents of data over a period of time [24].

Next, the quantification of measures starts with the development of categories by classifying the content to most efficiently answer the research questions raised. E.g., Viswanathan et al., answers the question 'Why do consumers prefer multipurpose products?' with categories such as multiple functionalities, shortcomings of single-purpose products, and the utility during emergencies [16].

Automated Approaches for Customer Needs Gathering & Prediction

A few researchers, in the past, have used data mining and extraction algorithms to automate data collection from online reviews. Wang and Chen [25] used a data mining and network analysis approach to predict the purchase choices made by customers. Similarly, Burnap et al. [26], predicted the design preferences using a feature learning technique. Ferguson et al. [27], proposed the use of certain cue-phrases to extract information that is useful for designers from online review databases. Singh and Tucker [28] predicted the function, form, and behavior of new products using a machine learning algorithm on product reviews. More recently, Zhou et al. [29], used a machine learning approach to identify customer needs for product ecosystems. Jin et al. [48], suggested a method to extract a few opinionated sentences from a selection of product reviews to indicate the strengths and weaknesses of a product. To assist designers, the algorithm also researches competitive products and their strengths and weaknesses to automate the process of competitor analysis in product development. Wang et al. [49], developed an algorithm that used the inconsistent ordered choice model to measure customer preferences for each product feature. The algorithm then used a sentiment-based importance-performance analysis to categorize product features for guiding product development. Anh et al. [50], used a convolutional neural network model to organize helpful reviews from amazon for the purpose of assisting designers. This method provided a useful guide in conducting review extraction from Amazon. Similarly, Jenitha and Ajitha [51] proposed a technique called "semantic orientation" to organize sentiments. The product aspect is identified, then the sentiment is classified using the identification of connotation, parts of speech, and dictionary definitions.

Aspect-Based Sentiment Analysis (ABSA)

ABSA can be broken down into two key steps – the first step detects the presence of certain aspects in the text, including the extraction of these aspects; the second detects the polarity of the review towards the aspect.

Several topic models have been proposed for the supervised extraction of aspects from a corpus of text [30, 31]. However, accurately annotated datasets are not easily available for specific applications, such as ours. Unsupervised aspect generation emerged as a result [32]. These models use pre-trained word embeddings or train word-embeddings using review datasets. Such approaches are not appropriate for our application since product design does not benefit from generic insights. E.g., it is not useful to extract that eco-friendly product reviews contain the aspect "recyclable." On the other hand, actionable insights are far more relevant. E.g., identifying that customers are interested in the "durability" of eco-friendly products more than the "aesthetics."

Prior work in analyzing customer needs from product reviews uses generic models such as fastText[33] and Latent Dirichlet Allocation (LDA)[31] to extract aspects from review datasets and extract sentiment of users towards these topics [34]. While prior work shows that NLP techniques are a promising technique, the outcome of the approach only generates generic insights from the reviews.

Fang & Zhan [52] used sentiment analysis previously to mine information from product reviews from Amazon. They proposed an algorithm for sentiment classification into positive, negative, and neutral. Their analysis, using the support vector machines algorithm, focused on four major categories of products: beauty, books, home, and electronics. Potdar et al. [53], proposed a bot called "Samiksha" that analyzes products and summarizes the main points in the reviews using a point system for each of the specific features of the product. This bot relied on natural language processing of fetched reviews. Very recently, Joung and Kim [54] used an automated keyword filtering algorithm in LDA to identify product attributes from customer reviews.

Product Category Used for the Case Study

In this study, the market segment, "environmentally (eco-) friendly products," is selected for a case study. Despite the concerns about environmental pollution and global warming, sales of eco-friendly products are susceptible to drop, as stated in market research reports by industry experts and makers of such products [35, 36]. E.g., in recent years, the market for eco-friendly household cleaning products declined at an annual compound rate of 2% [35]. No studies have explored the consumer sentiments that are guiding the sales of these products. Hence this category is selected as a case study.

3. OBJECTIVES

The objectives of this work are:

(1) To develop an efficient process to automatically extract customer sentiments for a specific category of products using product review data.

(2) To perform the customer sentiment extraction on the category of environmentally friendly products and compare the results with those from a manual content analysis process.

4. METHOD

In this study, two separate processes are performed parallelly on the same category of products – a manual content analysis and an automated ABSA-based process. The manual process is employed on a smaller sample (40 products, ~400 reviews) due to the time commitment it demands. The automated process uses a significantly larger sample size. For comparison purposes, the reviews for the same 40 products are used in the automated process.

Manual Content Analysis to Extract Customer Sentiments

Since "eco-friendly products" is selected as a category relevant for this analysis, the first step is to identify a few products that belong to the category. Product selection is made per the criteria laid out in the flow chart in Figure 1. A product can be classified as an eco-friendly one based upon reusability, recyclability, or the ability to conserve resources. E.g., a solar-powered appliance is classified as an eco-friendly product based on its ability to conserve electricity. A metal reusable water bottle is an eco-friendly variant of a disposable plastic water bottle, and therefore is suitable for the content analysis.

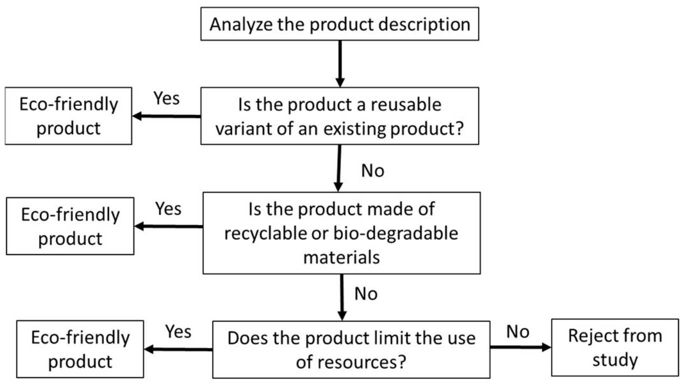

FIGURE 1 – PRODUCT SELECTION FLOWCHART

Once the products are selected based on the above criteria and the additional criteria that each product must have at least ten reviews on *Amazon.com*, product reviews are gathered for analysis. In order to avoid biased reviews, only vendor-verified purchases are used. Reviews are selected based on how well they reflect the measures and are gathered to reflect the overall rating distribution. Once the textual data are collected, content analysis is performed to identify the trends within those reviews.

Next, for each of the products identified, ten reviews are semi-randomly collected from vendor-verified purchases. To clarify this, consider the example of a solar-powered keyboard. For this product, the randomly selected reviews are distributed as six 5-star reviews, one 4-star review, one 3-star review, one 2-star review, and one 1-star review. This distribution is meant to approximately match the 59% 5-star reviews, 13% 4-star reviews, 8% 3-star reviews, 8% 2-star reviews, and 12% 1-star reviews of the product as reported on *Amazon.com* at the time of data collection.

The semi-random process is employed instead of a completely random process to avoid any bias arising from the low sample size in the manual process. E.g., in a completely random selection, most of the reviews selected for a low-rated product may be 1-star reviews, which may not be a true representative sample for the product. The semi-random process ensures that the reviews chosen for the study represent the average rating of the product.

The quantification process begins with the content analysis of the selected reviews. During the preliminary analysis, 14 product reviews are classified into bins that reflect similar consumer sentiments. The sorting is done at a sentence level without paying attention to the origin of the sentence. This is performed to eliminate any bias that arises from the knowledge about the product. Once the categories are finalized, common keywords and key terms that define them are identified. These keywords and key terms are used as guidelines for the analysis of the larger set of 40 products. Table 1 shows the 12 categories (i.e., "Aspects") of user sentiments obtained and the keywords that guide their identification. By the end of the analysis on the first 14 products, the aspect list converged, and the new products were less likely to add more to the list.

TABLE 1. USER SENTIMENT CATEGORIES AND RELEVANT KEYWORDS/TERMS

Number Designation	Aspect	Examples of Keywords
1	Aesthetics	Crisp, Beautiful, Wrinkled
2	Ease of Reprocessing	Wash, Clean, Charge
3	Durability	Wear, Died, Resistant
4	Use Efficiency	Time, Fast, Long
5	Performance	Hold, Well, Glitch
6	Adaptability	Versatile, Outside, Suitable
7	Ergonomics	Comfortable, Easy, Awkward
8	Ease of Storage	Store, Fold, Small
9	Ease of Use	Use, Easy, Convenient
10	Interference	Loud, Taste, Smell
11	Safety	Safe, Drop, Burn
12	Price	Expensive, Cheap, Cost

Figure 2 shows a frequency plot on how often each of the aspects is mentioned within the total number of consumer reviews gathered across the first 14 products. This information helps in establishing the impact each aspect has on consumer preferences.

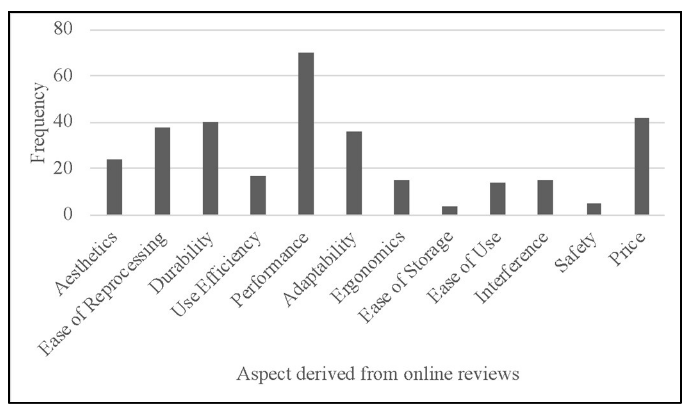

FIGURE 2. FREQUENCY PLOT FOR THE ASPECTS GATHERED FROM MANUAL CONTENT ANALYSIS

<u>*Aspect-Based Sentiment Analysis to Extract Customer Sentiments*</u>

In this paper, we present a novel approach to extract customer sentiments towards specific features of a product, called *aspects*, from review datasets. We use domain-specific aspects that are manually generated by product designers and are features of the product that they are interested in learning about. The trained aspect detection model can then be used to extract the presence of each aspect in a review. Parts of the review corresponding to each aspect are then processed passed to a sentiment analysis model, which outputs a quantitative sentiment code. Our approach is pictorially depicted in Figure 3.

Aspect Detection

Five aspect detection models were trained on different word embeddings for three aspect sets each.

Dataset and Aspect Sets

There are 12 manually assigned aspects for this task, as described in Table 1. The dataset containing 956 reviews favors some aspects (e.g., Performance) more than others (e.g., Interference). Also, the aspects are also ambiguous, especially for NLP models. E.g., for the review "These bags hold a lot," it is unclear if this should be classified under "Efficiency" or "Performance" by the NLP pipeline.

Therefore, we combine some of the manually assigned aspects based on empirical information and intuition on the NLP model to form a reduced aspect set. This process results in three sets of aspects: `aspect_12`, `aspect_6`, and `aspect_6-`. The `aspect_12` set is the original manually assigned aspect set. The sets `aspect_6` and `aspect_6-` refer to aspect sets that have been combined, as shown in Table 2. For `aspect_6`, the results of our predictions were combined after training on the 12 aspects, whereas for `aspect_6-`, the predictions are made on the reduced aspect set.

Word Embeddings

We have trained word-embeddings curated from three different sources, described as follows:

- `gen`: General sentences from the Wikipedia Corpus Dump
- `dom`: Domain related sentences from the Amazon Review Dataset [37]
- `spc`: Domain-specific (i.e., eco-friendly product) Amazon reviews obtained via web scraping

Different word embedding models were explored, and for the scope and purposes of our work, we elected to choose the Word2Vec [56] approach with the CBOW (continuous-bag-of-words) algorithm to generate our word embeddings. This word embedding model has a proven track record in the field of NLP research. Future work will likely see the introduction of other popular word embedding models for comparison of results.

Each of these datasets has different levels of engagement with product review language. Based on our preliminary study, we have found that models trained on larger numbers of tokens, especially those with high proportions of `dom` tokens, perform best. We have used these models to generate the following word embeddings to train the aspect detection model:

- `tot:    dom`
- `tot:gen+dom`
- `tot:gen+dom+spc`
- `lim-nx:100gen+77dom+spc`
- `lim-nx:77gen+77dom+spc`

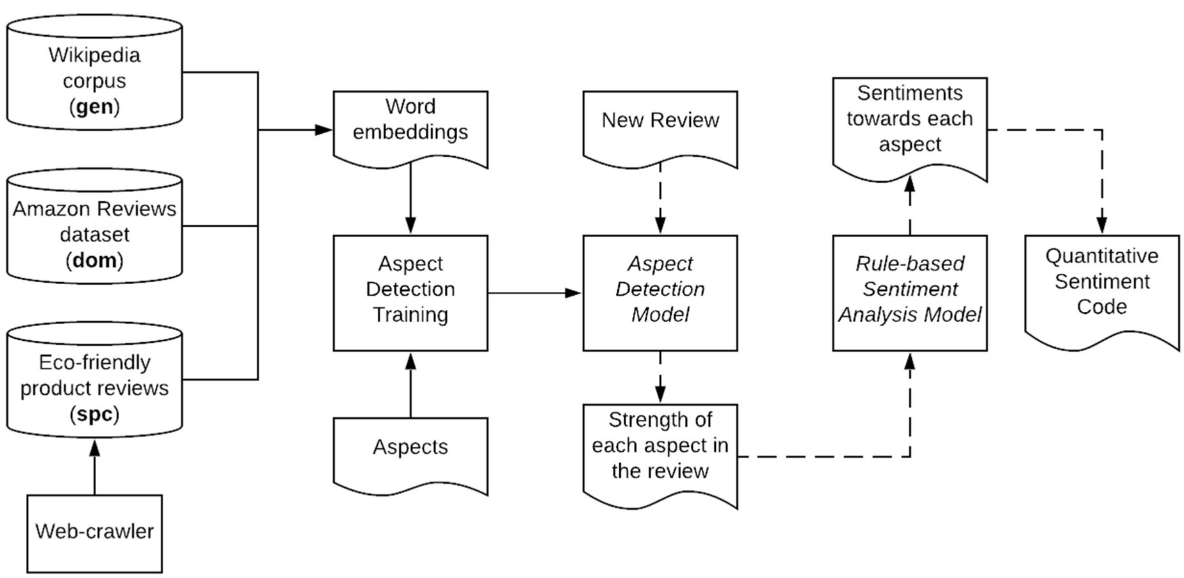

FIGURE 3. BLOCK DIAGRAM DEPICTING OUR NLP APPROACH. THE DASHED LINES REFER TO THE STEPS PERFORMED AT THE TIME OF PREDICTION.

Here, `lim-nx` refers to the fact that the proportions are multiples of x, which is the size of the smallest corpus (i.e., `spc`) - around 6 million tokens. Hence `100gen+77dom+spc` would mean 100*x generic tokens, 77*x domain-related tokens, and x domain-specific tokens. A point to note is the use of 77 as a multiplier. This is because 77*x (around 450 million) is the total size of the `dom` corpus, hence limiting any model involving domain-related tokens to a maximum of `dom` tokens.

Similarly, `tot` refers to word embedding models that have been trained on all the tokens in a corpus. For instance, the `tot:gen+dom+spc` model has been trained on all tokens across the three corpora used by us for these experiments. The training is also done in left-to-right order, i.e., a word embedding model is first created from all tokens in the generic corpus and then fine-tuned by training on all tokens in the domain-related and domain-specific corpora.

TABLE 2. REDUCED ASPECT SETS USED FOR THE AUTOMATED ANALYSIS

Aspects in aspect_12	**Aspects in aspect_6 and aspect_6-**
Performance, Efficiency	Performance
Ease of Use, Ease of Reprocessing, Ease of Storage	Convenience
Ergonomics, Durability, Adaptability	Design
Interference, Safety	Reliability
Aesthetics	Aesthetics
Price	Price

Aspect Detection Model

For each word embedding and aspect set, we trained an elementary aspect detection model that learns sentence similarities between the aspects and the review text to predict the presence of an aspect in the review. In order to accomplish this, a small portion of similarly annotated eco-friendly products (not included in the ~950 products that will be evaluated) is used to seed "representative sentences," i.e., collections of tokens that overwhelmingly point to a particular aspect. Following this, reviews are preprocessed (stop words removed, words tokenized, and lemmatized) and compared with these representative words. Word vectors for each word in a sentence or phrase are averaged to obtain a sentence embedding for similarity comparison. Future work can explore more sophisticated sentence embedding methods, such as the Smooth Inverse Frequency [57] approach. Cosine similarity is used to compute a similarity score, as it works well with high dimension spaces such as those found in document similarity problems [43]. The aspect corresponding to the most similar representative sentence is assigned to the review. To account for subjectivity amongst aspects with fuzzy boundaries, we make a provision where if up to two other aspects show a similarity score above a threshold (currently 90% of the similarity of the assigned aspect), they are also assigned to that review. If any of the aspects match the manually annotated aspect, it is counted as an accurate prediction for the system.

<u>*Rule-Based Sentiment Analysis*</u>

For sentiment analysis, we use VADER [58], a rule-based model for sentiment analysis. VADER uses empirically generated lexical features and generates an intensity of the sentiment expressed in a sentence. The model has been shown to work satisfactorily for microblog-like applications such as Twitter. We pick this model since review language is comparable to microblog language and due to the model's simplicity and availability off-the-shelf.

In this work, the manually annotated sentiments are binary – i.e., only the number of positive and negative references are counted. Our output from VADER is rounded off to enable a meaningful comparison. Using the output of VADER, we automatically generate the code vector for each product.

TABLE 3. CODE COMPARISONS BETWEEN MODELS AND MANUAL ANNOTATION FOR THE PRODUCT "BIODEGRADABLE BAGS AND POTS". POSITIVE VALUES INDICATE AN OVERALL POSITIVE SENTIMENT AND NEGATIVE VALUES INDICATE AN OVERALL NEGATIVE SENTIMENT TOWARDS AN ASPECT.

	Manual	77gen+77 dom_spc	100gen+77 dom+spc	tot:gen+ dom+spc	tot:gen+ dom	tot:dom
Adaptability	1	0.8333333	0.75	0.625	0.777778	0.7
Aesthetics	0	0	0	0	0	1
Durability	-1	-0.555556	-0.6363636	-0.55556	-0.55556	-0.57143
Ease of Reprocessing	0	0.8333333	1	0.714286	0.857143	0.857143
Ease of Storag	0	0.75	0.8	1	0.6	1
Ease of Use	1	0.6	-0.5	-0.5	-0.5	-0.6
Efficiency	1	-1	-1	-1	-1	-1
Ergonomics	0	-0.666667	-0.5	-1	0	0
Interference	0	-0.75	-0.5	-0.66667	-0.5	-1
Performance	1	0	-1	-1	-1	-1
Price	1	-0.75	-0.6666667	-0.83333	-0.66667	-0.66667
Safety	0	0	0	0	0	0

Table 3 shows an example of the automated sentiment codes generated for each aspect (in the `aspect_12` set) by the five-word embedding models juxtaposed against those generated manually. It is evident that automatically generated codes are way more expressive in the values they take on, as well as less sparse than their manually generated counterparts. In certain cases (such as "Adaptability," "Aesthetics," and "Durability"), the automatically generated codes are in agreement with their human-annotated counterparts. On the other hand, they strongly contradict the manually calculated codes in cases such as those of "Price," "Efficiency," and "Performance." Such contradictions could be caused by either incorrect sentiment polarity assignments or incorrect aspect detection for some reviews. Our work focuses on the aspect detection task, given the dearth of existing work, and considering that the results of sentiment analysis are uniform across word embeddings. The latter has hence been relegated to a third-party software, and the discussion of results in the following section will focus on the aspect detection task.

5. RESULTS AND DISCUSSION

Our experiments were run on a High-Performance Computing System provided by the College of Engineering at San Jose State University[2], comprising 36 compute nodes with a total of 1008 compute cores (Intel Xeon E5-2660 v4 processors), 15 GPUs (NVIDIA Tesla P100 12 GB), with 128GB and 256GB RAM available on compute and GPU nodes, respectively.

Table 4 shows the accuracy scores for our system's automatic aspect detection using different word embedding models. The system has been evaluated against a baseline of aspect annotations performed by a single individual human. Our automated aspect extraction system using word embeddings with proven track records has over a 50% accuracy on the `aspect_6` problem set and a near 40% accuracy on the `aspect_12` set. While these results are not close to the accuracy of a human expert, our aspect detection scheme is using a fairly basic algorithm built on sentence similarity. Once this is extended to deep learning models that leverage word embeddings [38], the scores are expected to improve. This is a major area for improvement, and future work will focus on leveraging more advanced natural language processing technologies such as attention-based [39] models, transformer networks [40], and recent advances in transfer learning for language modeling [40, 41]. Since the process of manual annotation is subjective, it will also be interesting to explore how a diverse cohort might annotate the reviews.

TABLE 4. ASPECT DETECTION ACCURACY RESULTS

	aspect_6	**aspect_12**	**aspect_6-**
`lim-nx:77gen+77dom+spc`	**0.52**	0.39	0.15
`lim-nx:100gen+77dom+spc`	**0.51**	0.39	0.15
`tot:gen+dom+spc`	**0.50**	0.38	0.13
`tot:gen+dom`	**0.48**	0.36	0.15
`tot:dom`	**0.46**	0.33	0.14

Further, our dataset suffers from a relatively small size of approximately 950 reviews, as well as being subjectively annotated in a non-standardized manner. This approach is applied on a more standardized dataset for Aspect-Based Sentiment Analysis, such as the Laptop Review Dataset from SemEval [30], which will yield better indicators of the system's performance and accuracy. Finally, while this automated work focuses more on detecting the correct aspect being mentioned in a product review out of a manually predetermined set of aspects, future work will also focus on extracting aspects of products from the reviews themselves, as some unsupervised models do [32]. Hence, the automated system will be able to better model customer requirements directly from existing feedback of similar products. The `aspect_6-` set does not perform well, with extremely low accuracy scores. Since these results are obtained by modeling after reducing the aspects to a set of 6, this could suggest that premature approximation of the aspects' features hurts the predictive capabilities of the system.

6. CONCLUSIONS

Customer needs play a crucial role in the development of new products. Gathering customer needs from traditional sources (customer interviews, surveys, etc.) is a resource-intensive process. The availability of online reviews on consumer goods provides a new possibility of gathering customer needs from them. The manual processes to extract customer needs from a large amount of textual data are labor-intensive. Motivated by this, we propose an automated method with the help of Natural Language Processing and Aspect-Based Sentiment Analysis. In this study, we compare the results from the automated process with the aspects derived by an expert. The results show that the automated process is promising and has approximately 50% accuracy as compared to the human process. Given the small sample size used for this study, the accuracy of the automated process may improve with additional data. Based on the current study, the model proposed here can serve as an efficient way to automatically extract customer needs for a category of products.

Acknowledgments

The authors would like to acknowledge the contributions of Dr. Farshid Marbouti and Mr. Harsimran Singh towards this work.

REFERENCES

[1] Ulrich, K. T., and Eppinger, S. D., 1995, *Product Design and Development*, McGraw Hill, New York, NY.

[2] Otto, K. N., and Wood, K., 2003, *Product Design: Techniques in Reverse Engineering and New Product Development*, Pearson, New York, NY.

[3] Bickart, B., and Schindler, R. M., 2001, "Internet Forums as Influential Sources of Consumer Information," Journal of interactive marketing, 15 (3), pp. 31-40.

[4] Lee, T. Y., 2007, "Needs-Based Analysis of Online Customer Reviews," *in the proceedings of Ninth international conference on Electronic commerce*. Taiwan.

[5] Hu, M., and Liu, B., 2004, "Mining and Summarizing Customer Reviews," *in the proceedings of Tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. Seattle, WA.

[redacted]

[6] Nasukawa, T., and Yi, J., 2003, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing," *in the proceedings of 2nd international conference on Knowledge capture*. Sanibel Island, FL.
[7] Turney, P. D., 2002, "Thumbs up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," *in the proceedings of 40th annual meeting on association for computational linguistics*. Philadelphia, PA.
[8] 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.
[9] Zhang, Z., Ye, Q., Law, R., and Li, Y., 2010, "The Impact of E-Word-of-Mouth on the Online Popularity of Restaurants: A Comparison of Consumer Reviews and Editor Reviews," International Journal of Hospitality Management, 29 (4), pp. 694-700.
[10] Clemons, E. K., Gao, G. G., and Hitt, L. M., 2006, "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," Journal of management information systems, 23 (2), pp. 149-171.
[11] Ghose, A., and Ipeirotis, P. G., 2011, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," IEEE transactions on knowledge and data engineering, 23 (10), pp. 1498-1512.
[12] Miao, Q., Li, Q., and Dai, R., 2009, "Amazing: A Sentiment Mining and Retrieval System," Expert Systems with Applications, 36 (3), pp. 7192-7198.
[13] Archak, N., Ghose, A., and Ipeirotis, P. G., 2007, "Show Me the Money!: Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *in the proceedings of 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. San Jose, CA.
[14] Gretzel, U., and Yoo, K. H., 2008, "Use and Impact of Online Travel Reviews," Information and communication technologies in tourism 2008, pp. 35-46.
[15] Zhou, F., Jiao, R. J., and Linsey, J. S., 2015, "Latent Customer Needs Elicitation by Use Case Analogical Reasoning from Sentiment Analysis of Online Product Reviews," Journal of Mechanical Design, 137 (7), pp. 071401.
[16] Viswanathan, V., and Sangelkar, S., 2016, "Design of Multipurpose Products: A Study on User Expectations," *in the proceedings of 4th International Conference on Design Creativity, ICDC*. Atlanta, GA.
[17] Hsieh, H.-F., and Shannon, S. E., 2005, "Three Approaches to Qualitative Content Analysis," Qualitative health research, 15 (9), pp. 1277-1288.
[18] Weber, R. P., 1990, *Basic Content Analysis*, Sage,
[19] Erlingsson, C., and Brysiewicz, P., 2017, "A Hands-on Guide to Doing Content Analysis," African Journal of Emergency Medicine, 7 (3), pp. 93-99.
[20] Neuendorf, K. A., 2016, *The Content Analysis Guidebook*, Sage, Los Angeles, CA.
[21] Danielson, W., 1963, "Content Analysis in Communication Research," Introduction to mass communications research, pp. 180-206.
[22] Bengtsson, M., 2016, "How to Plan and Perform a Qualitative Study Using Content Analysis," NursingPlus Open, 2 pp. 8-14.
[23] Graneheim, U. H., and Lundman, B., 2004, "Qualitative Content Analysis in Nursing Research: Concepts, Procedures and Measures to Achieve Trustworthiness," Nurse education today, 24 (2), pp. 105-112.
[24] Zhang, Y., and Wildemuth, B. M., 2009, "Qualitative Analysis of Content," Applications of social research methods to questions in information and library science, 308 pp. 319.
[25] Wang, M., and Chen, W., 2015, "A Data-Driven Network Analysis Approach to Predicting Customer Choice Sets for Choice Modeling in Engineering Design," Journal of Mechanical Design, 137 (7).
[26] Burnap, A., Pan, Y., Liu, Y., Ren, Y., Lee, H., Gonzalez, R., and Papalambros, P. Y., 2016, "Improving Design Preference Prediction Accuracy Using Feature Learning," Journal of Mechanical Design, 138 (7).
[27] Ferguson, T., Greene, M., Repetti, F., Lewis, K., and Behdad, S., 2015, "Combining Anthropometric Data and Consumer Review Content to Inform Design for Human Variability," *in the proceedings of ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*.
[28] Singh, A., and Tucker, C. S., 2017, "A Machine Learning Approach to Product Review Disambiguation Based on Function, Form and Behavior Classification," Decision Support Systems, 97 pp. 81-91.
[29] Zhou, F., Ayoub, J., Xu, Q., and Jessie Yang, X., 2020, "A Machine Learning Approach to Customer Needs Analysis for Product Ecosystems," Journal of Mechanical Design, 142 (1).
[30] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., , M., S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., and Hoste, V., 2016, "Semeval-2016 Task 5: Aspect Based Sentiment Analysis.," *in the proceedings of 10th International Workshop on Semantic Evaluation (SemEval)*.
[31] Blei, D. M., Ng, A. Y., and Jordan., M. I., 2003, "Latent Dirichlet Allocation," Journal of machine Learning research, pp. 993-1022.
[32] He, R., Lee, W. S., Ng, H. T., and Dahlmeier, D., 2017, "An Unsupervised Neural Attention Model for Aspect Extraction," *in the proceedings of Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.
[33] Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T., 2017, "Bag of Tricks for Efficient Text Classification," *in the proceedings of Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers.*,
[34] Ayoub, J., Zhou, F., Xu, Q., and Yang, J., 2019, "Analyzing Customer Needs of Product Ecosystems Using Online Product Reviews," *in the proceedings of Proceedings of the ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.*,

[35] Stewart, C., 2016, "3 Reasons Sales of Green Household Products Are Dropping", Accessed on 2018, Available at https://blog.marketresearch.com/3-reasons-green-household-product-sales-are-dropping
[36] Beningo, S., Chambers, M., Ford, C., Notis, K., Liu, M., and Smith-Pickel, S., 2017, "National Transportation Statistics", Published by U. S. D. o. Transportation, Available at https://www.bts.gov/sites/bts.dot.gov/files/docs/browse-statistical-products-and-data/national-transportation-statistics/217651/ntsentire2017q4.pdf.
[37] He, R., and Mcauley, J., 2016, "Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering," *in the proceedings of 25th International Conference on the World Wide Web (WWW'16)*.
[38] Xu, Hu, et al. "Double embeddings and cnn-based sequence labeling for aspect extraction." arXiv preprint arXiv:1805.04601 (2018).
[39] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
[40] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
[41] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
[42] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. In Advances in neural information processing systems (pp. 5754-5764).
[43] Singhal, A., 2001, Modern Information Retrieval: A Brief Overview, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 24 (4): 35–43.
[44] Qi, J, Zhang, Z, Jeon, S, and Zhou, Y., 2016. “Mining Customer Requirements from Online Reviews: A Product Improvement Perspective.” Information & Management 53 (8): 951–63. https://doi.org/10.1016/j.im.2016.06.002.
[45] Liu, Y., Jin, J., Ji, P., Harding, J., and Fung, R., 2013, “Identifying Helpful Online Reviews: A Product Designer’s Perspective.” Computer Aided Design 45 (2): 180–94. https://doi.org/10.1016/j.cad.2012.07.008.
[46] Yi, S., and Liu, X., 2020. “Machine Learning Based Customer Sentiment Analysis for Recommending Shoppers, Shops Based on Customers’ Review.” Complex & Intelligent Systems 6 (3): 621–34. https://doi.org/10.1007/s40747-020-00155-2
[47] Decker, R., and Trusov, M., 2010. “Estimating Aggregate Consumer Preferences from Online Product Reviews.” International Journal of Research in Marketing 27 (4): 293–307. https://doi.org/10.1016/j.ijresmar.2010.09.001
[48] Jin, J., Ji, P., and Gu, R., 2016. “Identifying Comparative Customer Requirements from Product Online Reviews for Competitor Analysis.” Engineering Applications of Artificial Intelligence 49: 61–73. https://doi.org/10.1016/j.engappai.2015.12.005
[49] Wang, A., Zhang, Q., Zhao, S., Lu, X., and Peng, Z., 2020. “A Review-Driven Customer Preference Measurement Model for Product Improvement: Sentiment-Based Importance–performance Analysis.” Information Systems and e-Business Management 18 (1): 61–88. https://doi.org/10.1007/s10257-020-00463-7
[50] Anh, K., Nagai, Y., and Nguyen, L., 2019. “Extracting Customer Reviews from Online Shopping and Its Perspective on Product Design.” Vietnam Journal of Computer Science 6 (1): 43–56. https://doi.org/10.1142/S2196888819500088
[51] Jenitha R., and Ajitha, P., 2016. “Survey of Product Reviews Using Sentiment Analysis.” Indian Journal of Science and Technology 9 (21). https://doi.org/10.17485/ijst/2016/v9i21/95158
[52] Fang, Xing, and Zhan, Justin. 2015. “Sentiment Analysis Using Product Review Data.” Journal of Big Data 2 (1): 1–14. https://doi.org/10.1186/s40537-015-0015-2
[53] Potdar, A., Patil, P., Bagla, R., Pandey, R., and Jadhav, N., 2016. “SAMIKSHA - Sentiment Based Product Review Analysis System.” Procedia Computer Science 78: 513–20. https://doi.org/10.1016/j.procs.2016.02.096
[54] Joung, J., Kim, H., 2021, “Automated Keyword Filtering in Latent Dirichlet Allocation for Identifying Product Attributes From Online Reviews,” Journal of Mechanical Design, 143, 084501
[55] Suryadi, D., Kim, H., 2015, “A Data-Driven Methodology to Construct Customer Choice Sets Using Online Data and Customer Reviews,” Journal of Mechanical Design, 141, 111103.
[56] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* (*NIPS'13*). Curran Associates Inc., Red Hook, NY, USA, 3111–3119.
[57] Arora, S., Liang, Y., & Ma, T. (2019). *A simple but tough-to-beat baseline for sentence embeddings*. Paper presented at 5th International Conference on Learning Representations, ICLR 2017, Toulon, France.
[58] Hutto, Clayton J. and Gilbert, Eric. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.." Paper presented at the meeting of the ICWSM, 2014.