Feature-based Sentiment Analysis by Distilling Customer Reviews: A Case of Headphones Market

Aneesh Kulkarni, Jiale Qin, Pouria Babvey, Shivi Jain, Wenlei Feng, Yichao Shen

***Abstract*—Predicting what consumers will buy in the future is always a key area of interest for any manufacturer. Existing solutions are based on extrapolating the current products and demand on the market. This will miss the non-linear evolution of both the technology and the demand. Missing the non-linear evolution may lead to missing large market opportunities. This was the case for products like tablets, that didn't exist at all before and became a relevant market segment. Having a more profound insight into the customers feedback on different features of a product can help designers to understand the needs on the market and predict the future evolutionary trends of the existing products. We focused on Headphones market as a case study to develop our feature-brand satisfaction model. We used some of basic text analysis techniques to extract mainly discussed features in the reviews for Headphones products on Amazon. Then, we applied sentiment analysis to the segments discussing each feature separately. Our results show that the existing products on the headphones market have a satisfactory performance on battery performance and material technologies while customers are reluctant about disharmony of headphone sides, Bluetooth connection, sound isolation, and bass response. Finally, we compared the performance of well-known brands on each of these features. The proposed process can automatically find discussed collocations (as features) and can be easily expanded to any other digital product.**

***Index Terms*—sentiment analysis, feature extraction, amazon reviews, data scraping, headphones, customer feedback, prediction model**

Acknowledgment

We thank our teacher, Professor Wentzell, from Stevens Institute of Technology who provided insight and expertise that greatly assisted the research.

I. Introduction

The internet holds a considerable amount of user-generated content describing the opinions of customers on products and services through blogs, tweets and other social media forms. These reviews are valuable for customers making purchasing decisions and companies guiding business activities. However, browsing the extensive collection of reviews to search for useful information is a time-consuming and tedious task. With a huge number of reviews it seems unfeasible for customers to find all the well-informing ones. Moreover, the existent method to highlight the confirmed reviews may not be efficient since it gives more credit to early reviews and overlook the importance of recent reviews.

Consequently, sentiment analysis and opinion mining have attracted significant attention in recent years as they pave the way for the automatic analysis of user reviews and the extraction of information most relevant to users. Sentiment analysis entails several interesting and challenging tasks. One traditional and fundamental task is polarity classification, which determines the overall polarity (e.g., positive or negative) of a sentence or document. However, these tasks are coarse-grained and cannot provide detailed information, such as the aspects on which the users comment. Recently, there has been a shift towards the fine-grained tasks, such as aspect-based (or feature-based) sentiment analysis, which not only involves analyzing the opinionated texts polarity (e.g., positive, neutral, negative) and intensity (e.g., weak, medium, strong, extreme), but also identifying the aspect (or the topic, or target entity) of the opinion[13].

In this way, customers can get the pros and cons of the product easily and save their time of choosing a product. The result of the project can help the online shopping platforms to provide better services to customers. It saves a lot of time of customers, facilitates customer decision-makers, and helps markets to make feedback information more transparent and accessible.

This report is structured as follows: The next section discusses the business understanding; Section 3 describes the process of data gathering whereas section 4 explains the methodology we applied to process the data; Section 5 discusses the main results. Finally, we finish the report with conclusion and future works works.

II. Business Understanding

Our project will benefit both the customer and the vendor as our result will give a score to each extracted feature when compared to all other brands. Customers who are looking for a specific aspect of a product can easily see which brand offers the best on the market. Also, the manufacturers will be able to analyze how their products' features are performing as compared to their competitors.

III. Data Gathering

A. Data Corpus

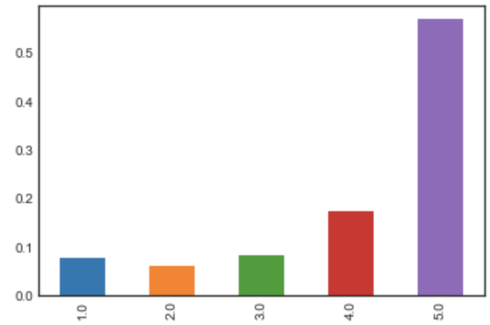
In total we have extracted over 71 thousand reviews for our project. We selected famous brands like *Ailihen, Akg, Bose, Otium, Panasonic, Sennheiser, Shure and Sony* which ensured ample diverse data. We built a review scraper using *Selenium package* and *Pandas* to scrape and organize the data. Once all data was extracted we used *The Natural Language Toolkit 'NLTK'* for data cleaning. We first split all reviews into sentences and later individual words. We then filtered out punctuations and special symbols, stop words and preformed lemmentization to reduce words to base form. We created our own filter using NLTKs collocation function as a base to extract the top 20 features for the product category depending on frequency of occurrence.

B. Cleaning Data

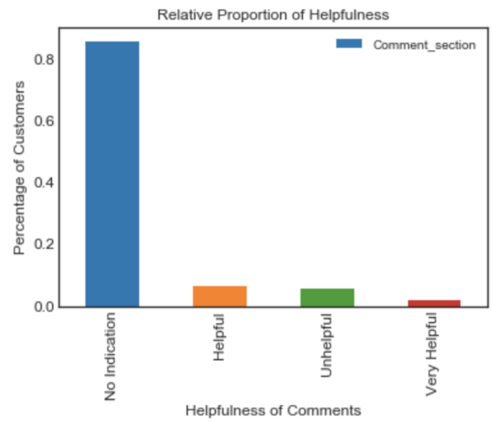
The data we extracted contained column names like helpfulness, rating, reviews, date which we used to determine the importance of a review. Firstly, we transformed some of the variables into numeric variables for a better analysis. The sentences of helpfulness were represented by the count of helpfulness, and the ratings variables were convert from string to numeric variables. The *date-time function* is applied to the date extracted. For the headline and text review part, we used their length to do the work. For further analysis, we divided the help fullness parameter into four meaningful classes - *No Indication, Less Helpful, Helpful, Very Helpful* .

	Helpfullness	Ratings	Reviews	Title	Length_Reviews	Length_Title
0	30	1.0	cans sound fine but the akg branded audio cab...	cans sound fine but poor product quality	1768	41
1	0	5.0	i bought these for my futaba sakura cosplay b...	love these	1312	11
2	0	5.0	holy cow what a set of cans these are some n...	affordable headphones offering great sound qua...	2287	51
3	6	1.0	love this headphone but the left side is loude...	high failure rate	277	17
4	8	5.0	sound leakage nil mildsound isolation epic...	excellent craftsmanship sound and design hits...	321	122

(a) Dataframe



(a) Histogram of Ratings



(b) Histogram of Helpfulness

Fig. 2: Explore for each variable

C. Exploratory Data Analysis

After preparing the data to perform exploratory data analysis, we check the variables in single. We plotted the histogram for ratings and helpfulness respectively and presented in *Figure1*. We can see that, more than 60% of verified customer gave 5 stars to the product. That implies product features have very positive customer satisfaction. And for helpfulness, even though we divided it into 4 classes for the disperse of the data, more than 80% of them belong to the 0 3 class. Therefore, we thought the Helpfulness variable is not usable for our analysis.

Next, we considered the relationship between the variables. Figure 2 is the box plot for ratings and title length b) ratings and review length. We can clearly visible that reviewers gave lengthy reviews for 5 stars. All reviews had the average median word count around 70-80 words. Spread word count observed in 5-star review, which implies customer lengthy reviews when highly satisfied.

Figure 3 shows the distribution of score sorting to negative and positive groups. We set up the criteria that the scoring in reviews which is higher than three stars should be labeled Positive, otherwise labeled Negative. The volume of positive scores is nearly three times as that of negative one, with 74.31% vs 25.69 %, which manifests an imbalance in different labeled reviews data set.

IV. Methods

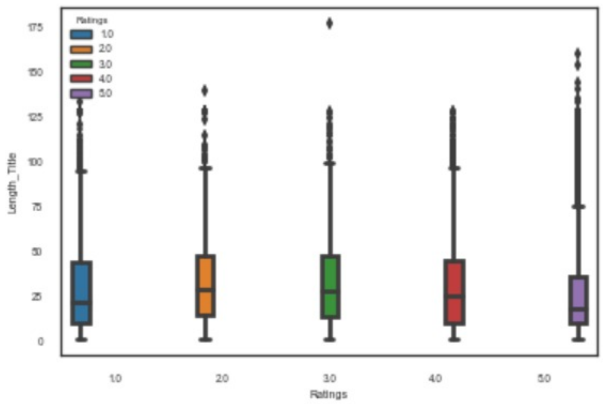
A. Pre-Processing

Pre-processing analyzes the opinions from syntactical point of view and original syntax of sentence is not disturbed. In this phase, the several techniques like POS tagging, Stemming and Stop word removal are applied to data set for noise reduction and facilitating feature extraction. Before passing the content to POS tagger, we cleaned the data, using *Regular expressio*n module by keeping only the alpha words. Also, we replaced the word noise with sound.

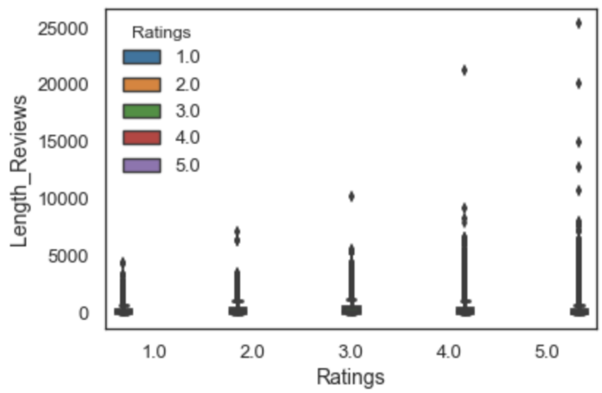
Part of speech (POS) tagging: Parts of speech or POS tagging is a linguistic technique used for product feature extraction as product aspects are generally nouns or noun phrases. A part-of-speech filter only lets through those patterns that are likely to be phrases.

Lemmatization: Lemmatization groups together various inflected forms of word into one. Stemming removes word inflections only whereas, Lemmatization replaces words with their base form. So, we used lemmatization in our pre-processing step.

Stop word removal: The stop words removal reduces dimensionality of the data sets and thus key words left in the review corpus can be identified more easily by the automatic feature extraction techniques. Words to be removed are taken from a commonly available list of stop words in nltk stop-word dictionary.



(a) Box Plot of Title length and Ratings



(b) Box Plot of Review length and Ratings

Fig. 3: Relation between two variables

B. Feature Extraction

Semantic type of features are extracted, that works on semantic orientation (SO). Semantic orientation (SO) technique make use of latent semantic analysis (LSA) and point wise mutual information (PMI), which assigns polarity score to every word or phrase.

C. Feature Selection

Natural language processing-based techniques mainly operate on three basic principles:(a) Noun, noun-phrases, adjectives, adverbs usually express product features. b) Terms occurring near subjective expressions can act as features. (c) P is product and F is feature in phrases like F of P or P has F.

Statistical techniques we used is NLTK Collocation Unigram and Bigrams. A collocation is an expression consisting of two or more words that correspond to some conventional way of saying things. Bigram Collocation method calculates the frequencies of words and their appearance in the context of other words. The collection of words is then filtered to only retain useful content terms. Each bigram of words is then scored according to some association measure, to determine the relative likelihood of each bigram being a collocation.

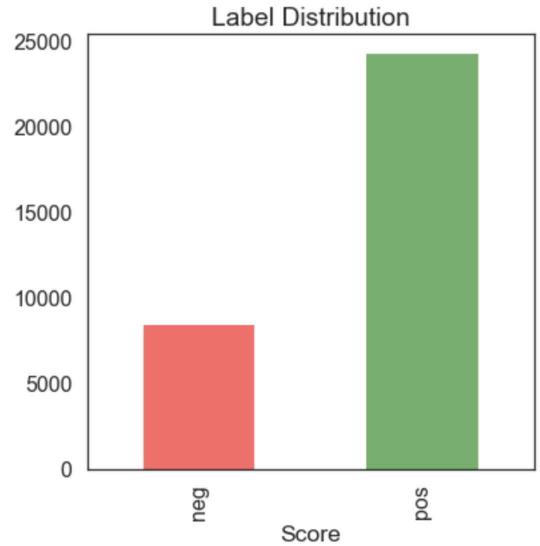


Fig. 4: Positive and Negative score distribution

Class nltk.collocations.BigramCollocationFinder(word_fd, bigram_fd, window_size=2)
*Bases: nltk.collocations.AbstractCollocationFinde*r.

D. Feature Cleaning

Large numbers of unnecessary features (around best 200) are produced during frequent feature set generation phase, which need to be removed. For getting the meaningful aspect, we further filtered the data by applying feature pruning algorithms like Spacy lemmatization.
We removed the collocation that doesnt include useful features by using spacy.token.lemma_ thus filtering the bigrams like (ADP,NOUN), (ADJ, ADJ), (ADJ,NOUN), (ADP,VERB) etc. After applying the filter, we came around with top 20 meaningful aspects for our sentiment analysis. Then, we combine some of the collocations which pertaining to the same feature (like noise cancellation, noise cancelling; or left ear, right ear; or volume control and ear monitor). Then we applied sentiment analysis for the 14 final features.

E. Sentiment Analysis

Sentiment analysis also known as opinion mining. It is the process of determining the emotional tone behind a series of words, used to gain an understanding of the attitudes, opinions and emotions expressed within an online mention.
We used VADER (Valence Aware Dictionary and Sentiment Reasoner) an open source python library to perform sentiment analysis.
Why VADER? With VADER you can perform sentiment classification very quickly even if you dont have positive and negative text examples to train a classifier or want to write custom code to search for words in a sentiment lexicon. VADER is also computationally efficient when compared to other Machine Learning and Deep Learning approaches.
VADER belongs to a type of sentiment analysis that is based

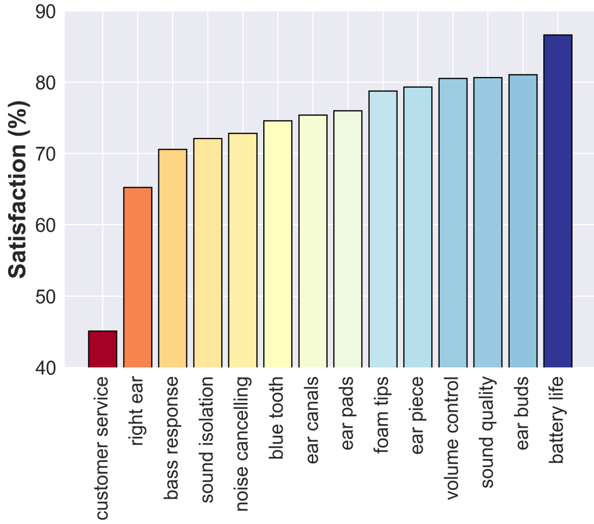


Fig. 5: The overall sentiment analysis for segments of reviews containing each feature: The results show that customers are not satisfied with customer services, balance of sound volume, bass response and sound isolation. It also shows the technology is mature enough in equipment like battery and material parts.

on lexicons of sentiment-related words. In this approach, each of the words in the lexicon is rated as to whether it is positive or negative, and in many cases, how positive or negative. The *vaderSentiment() method* returns the values to represent the amount of negative, positive, and neutral sentiment and also works out the compound sentiment value as a signed value to indicate overall sentiment polarity [11].

Methodology:

Step 1: After finding the meaning collocation, we filtered the reviews brand-wise that contains the specified feature.

Step 2: The filtering results in set of reviews that has key features in it. The reviews are then passed to the VADER module for calculating the sentiments.

Step 3: Each brands sentiment is analyzed with respect to each key feature.

Step 4: The sentiments of each feature of each brand is normalized in range (0,1) for further visualization.

V. Results

After collecting a large proportion of headphones reviews we pre-processed and filtered unrelated bigrams. Finally, we found 14 important features which mainly discussed in the headphones reviews. Then we applied the sentiment analysis to measure the overall customer satisfaction from those features. Our results are summarized in Fig.5. The results show that the customers criticize customer service more than the technological elements. Moreover, the signal processing part of the headphone technology has yet to provide a pleasant experience for users especially for bass response and harmony between two sides. On the other side material technology and batteries provide a satisfactory performance in users experience.

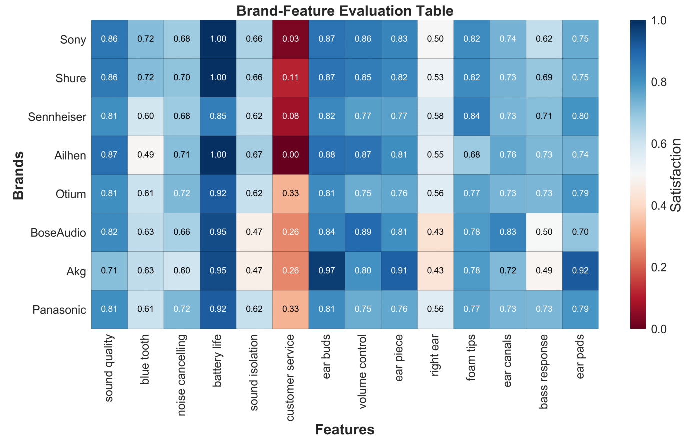


Fig. 6: The brand-feature table: Customers can find the most proper brands considering the feature which has the most correlation with their personal interests

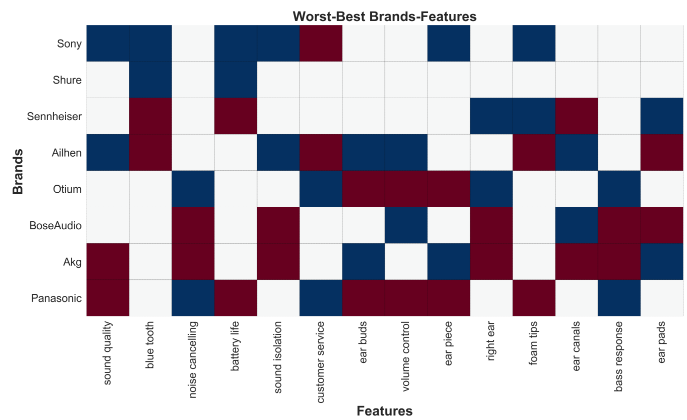


Fig. 7: The Worst-best brands table: The blue intersection represents the brand in the same row is among the two best reviews for the feature in the same column. Conversely, the red intersections show the poor performance of the brand for the features.

The next step of our project was brand-feature extraction table. We analyzed all the reviews for each brand and summarized the results in Fig. 6.

Our analysis show that the features are not strongly correlated among different brands. For example, Panasonic and Otium got the best feedback for the customer service while AKG has the lowest performance for sound isolation. Or Alihen is among the best brands as for battery life while its bluetooth connection has the worst performance. We summarized some these results in Fig. 7.

VI. Conclusion

It is gradually become harder and harder for customers to extract the intended information from the big mass of reviews. For popular products, there are thousands of reviews presented. It is hard for a customer to read all of these reviews to compare different brands of products and then make the purchase decision. Hence, it will be meaningful to present some useful information from the reviews, which are easier for customers to know about this product. We proposed a method to collect and extract useful information from a large amount of reviews. We collected a large proportion of the reviews on Amazon.com for the eight well-known headphones brands. Then, we prepossessed the data to extract discussed features in the reviews. Finally, we apply a sentiment analysis to evaluate the overall customer feedback on each feature. Our results have two main implications. First, designers can understand the overall feedback of customers on the basic elements of the product. It may result in: 1) advancements in the design of inefficient elements in the product 2) development of the more efficient methods to integrate the products 3) providing more understandable information for user (in case it is because of lack of customers knowledge about the product functionality) 4) or even development of the new products based on the customer needs.

For example, *right ear* and *left ear* are among the most negatively discussed bigrams in the headphones reviews. It denotes the volume of the sound are not equalized in both sides. Or at least, some customers have this unsatisfactory experience because of dissimilarity in hearing performance. Thus, designer may add a new degree of freedom for volume control for different years or automatically tune the volume in both sides.

In our case study, our results show that customers are mainly satisfied with battery life, and feel comfortable with ear-buds, ear-pads, and volume control. However, they mostly criticize blue-tooth connections, bass response, sound isolation. The results also show headphones companies have yet to provide a acceptable customer service.

The second implication of our study is a recommendation system for users to choose the most proper product regarding their personal concerns. For example, AKG, and Bose headphones have a poor performance on low-frequency sounds (weak bass response), or Senheiser is among the best brands for all the items except battery life.

VII. Future Works

The process that we developed in the project can be expanded to a tool used by manufacturers. This can be exploited in two ways:

A. A product to be sold either as a stand-alone solution or as a service on-line
B. A service to be provided, along with customization and additional professional services

The next steps of our project are: 1) extending the study to some other digital products 2) drawing a correlation between different features in Word2Vec model to provide a design network based on customer needs 3) applying time-series analysis to the frequency of the features and their sentiment in the reviews to provide a prediction model, and 4) providing an online tool to monitor the reviews instantly.

References

[1] Syed Zohaib Hussain Zahidi, Fadi Aloul, Member, IEEE, Assim Sagahyroon, and Wassim El-Hajj,*"Optimizing Complex Cluster Formation in MANETs Using SAT/ILP Techniques "*; IEEE Sensors journal, VOL. 13, NO. 6, JUNE 2013.

[2] Peng Zhao, Xinyu Yang, Wei Yu, and Xinwen Fu,*"A Loose-Virtual-Clustering-Based Routing for Power Heterogeneous MANETs "*; IEEE Transactions on vehicular technology, VOL. 62, NO. 5, JUNE 2013.

[3] Shengming Jiang, Dajiang He, and Jianqiang Rao ,*"A Prediction-Based Link Availability Estimation for Routing Metrics in MANETs "*; IEEE/ACM Transactions on networking, VOL. 13, NO. 6, DECEMBER 2005.

[4] Srivastava Sumit ,Deepankar Mitra, Devanshi Gupta,*"Proposed Intrusion Detection on ZRP based MANET by Effective K-means Clustering Method of Data Mining "*; International Conference on Reliability, Optimization and Information Technology -ICROIT 2014, India, pp No. 156-160, Feb 6-8 2014.

[5] Lung-Chung Li and Ru-Sheng Liu,*"Securing Cluster-Based Ad Hoc Networks with Distributed Authorities "*; IEEE transactions on wireless communications, VOL. 9, NO. 10, pp-3072-3081, OCTOBER 2010.

[6] Nicklas Beijar,*"Zone Routing Protocol (ZRP) "*; (Networking Laboratory, Helsinki University of Technology).

[7] Preetee K. Karmore et al,*"Detecting Intrusion on AODV based Mobile Ad Hoc Networks by k- means Clustering method of Data Mining "*; (UCSIT) International Journal of Computer Science and Information Technologies, Vol. 2(4), 2011,1774-1779.

[8] Jabas, A.,R.M. Garimella, and S. Ramachandram,*"MANET Mining: Mining Temporal Association Rules "*; in International Sympossium on Parallel and Distributed Processing with Applications. 2008. Sydney, NSW, Australia.

[9] Adnan Nadeem, Michael P. Howarth,*"A Survey of MANET Intrusion Detection & Prevention Approaches for Network Layer Attacks "*; IEEE communications surveys & tutorials, Vol. 15, No. 4, Fourth quarter 2013.

[10] Adnan Nadeem, Michael Howarth,*"A Generalized Intrusion Detection & Prevention Mechanism for Securing MANETs "*; Centre for Communication Systems Research University of Surrey, United Kingdom.

[11] Burchell, Jodie. "Using VADER to Handle Sentiment Analysis with Social Media Text." Standard Error Full Atom, t-redactyl.io/blog/2017/04/using-vader-to-handle-sentiment-analysis-with-social-media-text.html.

[12] Index of /OSMOSES, www.statmt.org/OSMOSES/FeatureEx.pdf.

[13] http://ir.hit.edu.cn/˜car/papers/ieee2015-2.pdf