QUANTIFYING PRODUCT FEATURE HETEROGENEITY AND ITS IMPACT ON PRODUCT DESIGN DECISIONS

A Thesis in
Industrial Engineering
by
Abhinav Satyendra Singh

© 2016 Abhinav Satyendra Singh

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
May 2016
The thesis of Abhinav Satyendra Singh was reviewed and approved* by the following:

Conrad S. Tucker
Assistant Professor in School of Engineering Design Technology and Professional Programs and Department of Industrial Engineering
Thesis Advisor

Harriet Nembhard
Professor of Industrial Engineering

Janis P. Terpenny
Peter and Angela Dal Pezzo Department Head
Professor of Industrial and Manufacturing Engineering

*Signatures are on file in the Graduate School
The primary objective of designers is to launch products that maximize enterprise value. However, the challenge of accomplishing this objective is understanding what product features (e.g., battery, camera, processor for a mobile phone) to include in a product. This thesis is aimed at reducing the mismatch between what designers deem relevant in the products that they design and what customers deem relevant in their purchasing decision. Product feature preferences are heterogeneous in nature and hence it is difficult for product designers to make decisions about product designs or features while designing next generation products. Typically, searching for product features to include in next generation products is a manual process, often involving numerous iterations with customers. With an increase in the number of customers migrating to online platforms to buy products, acquisition of product feature preference data is viable through online product data streams. In this work, a product feature is defined as representing a product’s form (i.e. pertaining to the shape, scale, material etc.), function (i.e. pertaining to the functional characteristics) or behavior (i.e. pertaining to the performance of a product during its lifecycle). While designing next generation products, designers must be able to assess the current state of product features, in order to minimize the risk of product failure in the market. This thesis illustrates a methodology to quantify the heterogeneity of product feature preferences in customer generated feedback posted online in order to enable product designers extract product features that receive homogeneous sentiments across all customers. Towards this end, this thesis investigates the “must have” and “deal breaker” product features expressed by users of online platforms (e.g., customer review websites or social media networks) in order to inform designers of product features that should be investigated during the next iteration of a product’s launch. Here, the term “product” is defined broadly to include tangible (e.g., cell phone) and intangible (e.g., healthcare service) systems that serve a given function(s). There exists a knowledge gap in terms of what product
features impact real world outcomes. Existing product design literature highlights the risks of aggregating group preferences, and suggest that design teams should instead, focus on maximizing enterprise value by optimizing the features of a product. However, design knowledge about products and product features are influenced by market information, which is dynamic and difficult to acquire. The use of online product review platforms has emerged in the design community as a viable source of product data acquisition and demand model prediction. However, as the heterogeneity of product preferences increases, so does the complexity of understanding which product features should be optimized by the design team to maximize enterprise value. In the process of designing next generation products, designers should be able to assess the relevance of incorporation of features with homogeneous preferences across users called as “must have” or “deal breaker” features based on the product feature preferences mined from online data. These challenges are exacerbated in product preference acquisition techniques that rely on mining online data, as the customer is typically unknown to the designer, which limits the amount of follow up data available to be mined. This work uses natural language processing techniques to acquire product feature preference data from online customer feedback to be combined with statistical analysis in an attempt to enhance the way in which products are evaluated by designers after launching into the market. By quantifying the degree of “must have” and “deal breaker” product feature preferences expressed online, designers will be able to understand what product-features should be omitted from next generation product design optimization models (i.e., “deal breaker” features) and what product features should be considered (i.e., “must have” features). A case study involving customer electronics mined from online customer review websites is used to demonstrate the validity of the proposed methodology. Results obtained from this work are a testimony of its application for analysis of features across different products and services. The authors were able to demonstrate the methodology and extract “must have” and “deal breaker” features for
the product under consideration which would result in optimized decisions on the designers’ end for next generation product design.
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................... viii

LIST OF TABLES ................................................................................................................ ix

ACKNOWLEDGEMENTS ................................................................................................. x

Chapter 1 Introduction ..................................................................................................... 1

Chapter 2 Literature Review .............................................................................................. 4

  2.1 Product Preference Acquisition Techniques ......................................................... 4

  2.2 Mining Customer Preferences from Online Data ...................................................... 6

  2.3 Design Axioms in Customer Preference Modeling ................................................... 8

Chapter 3 Methodology .................................................................................................. 12

  3.1 Text Mining and Feature Extraction ........................................................................ 13

  3.2 Sentiment Analysis for Extracted Features .............................................................. 15

  3.3 Approximation of Sentiment Scores ....................................................................... 16

  3.4 Discovering Homogeneous Product Feature Preferences ....................................... 20

Chapter 4 Case Study: Design Feature Extraction for the Apple iPhone 5 ..................... 25

  4.1 Application of the Methodology .............................................................................. 25

  4.2 Results ..................................................................................................................... 27

Chapter 5 Conclusion and Future Work .......................................................................... 33

  5.1 Summary .................................................................................................................. 33

  5.2 Scope for Future Work ............................................................................................. 33
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Outline of the Methodology</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>Mapping to the component based on features discussed in customers’ reviews</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>Classification of Sentiment Score into Positive or Negative</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>Application of the Methodology</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>Sentiment Analysis output by Alchemy API</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>Keyword Extraction by Alchemy API</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>Keyword Extraction by Alchemy API</td>
<td>28</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1 Customer preference ranking of three features of a product ........................................... 9

Table 2 Changed order preferences on introduction of an additional feature in the product .... 9

Table 3 Uncertainty in optimal product design................................................................. 17

Table 4 Extracted iPhone 5 features .................................................................................. 29

Table 5 Final Feature Classification ...................................................................................... 31
ACKNOWLEDGEMENTS

Firstly, I would like to express my heartfelt gratitude and thanks to my advisor, Dr. Conrad Tucker for constantly motivating and supporting me during the course of Master’s program in the Industrial Engineering Department. Dr. Tucker’s valuable insights and encouragement have helped me indulge into meaningful research. I am grateful to him for being an advisor on each step of the way which developed me into a better researcher and professional under his guidance. I would also like to thank my thesis reader Dr. Harriet Nembhard for taking time to review my thesis and providing suggestions for improvement.

I would like to acknowledge the Industrial Engineering Department for being supportive throughout the duration of the course, the Center for Integrated Healthcare Delivery Systems and NSF Center for Health Organization Transformation (CHOT) for opportunities to enhance my knowledge through research. It would not have been possible to achieve research goals without the help of members of the Design Analysis Technology Advancement (D.A.T.A.) lab. I thank Ajay Bharathi, Matt Dering, Sungwoo Kang, Sunghoon Lim, Qianyu Hu, Shruthi Reddy Bezawada and Owen Shartle for their help.

Most importantly, I would like to thank my parents Satyendra Singh and Aarti Singh for making me the person I am today and supporting me at every step of my life.
Chapter 1
Introduction

Qualitative customer feedback defined as “peer generated product evaluation posted on company or third part websites” is useful in extracting insightful information about the product’s performance in the market [1]. Online reviews and digital media mined from online product data streams have been proven to correlate to real world outcomes [2]. Thus, with the increasing size and complexity of products in the market, it is essential to develop techniques with which designers can gain knowledge about current state of products while designing next generation products. Modeling emerging trends is an efficient way to help new product development teams discover trends in product feature offerings in the market place [3]. A large amount of information is exchanged through online customer review platforms (e.g., Amazon.com, CNET.com) and social media websites (e.g., Twitter and Facebook). Such platforms allow customers to voice their opinion related to the products they use and the product features they like or dislike [4]. Digital word of mouth data, mined from online sources, have been shown to predict real world outcomes [2]. Specifically, in product design, Tuarob and Tucker have shown that social media messages can be used to estimate real world product demand [5]. Product designers must be enabled to extract feature preferences that ensure product market success while developing next generation design of existing products or designing a totally new product. Researchers have focused on eliciting the importance of product features by mining online customer data [6]–[8]. It is however necessary to inform designers about product features that should be consistent in future designs of a product and at the same time product features to be omitted. This work is aimed at informing designers about homogenous product feature preferences mined from online data in order to evaluate product
features that are either “must have” or “deal breaker”. “Must have” product features are product features that need to be included in a product in order for customer to consider the product among their choice set. “Deal breaker” product features are product features that should be omitted or improved in a product to ensure satisfactory functioning and delivery of the product to customers. Decisions based on such insights will ensure better design and incorporation of features which will in turn maximize enterprise value. Such scenarios hold good not only for products but any goods or services provided by enterprises in an attempt to maximize profits. However, in order to achieve this goal, product designers should not aggregate feature preferences but instead discover features that are perceived to be “must have” or “deal breaker” across customers. Engineering design is stated as a process to choose between alternatives or comparison of alternative that depend on values of individual preferences [9]. In this process, any one alternative is chosen based on the individual preference values and the other alternative or alternatives are left unattended. Thus, it is essential to establish a ground rule to evaluate product feature preferences which consider preferences that homogenous across all individual and do not result in aggregation of preferences violating design axioms. Thus designers must be careful while integrating product design knowledge into their decision making, so as not to violate axioms of design [10]. It has been shown that the aggregation of individual preferences is not appropriate in engineering design problems [9]. Customers’ preferences may be heterogeneous in nature and hence, lead to incorrect interpretations of group preferences resulting in non-optimal and unsatisfactory product designs. However, there are some ground truth features which form a vital part of the products and thus, cannot be ignored. Identifying the features to be improved in the generation product design efforts will enable designers to concentrate their resources on optimizing and improving these features. By quantifying the degree of “must have” and “deal breaker” product preferences expressed online, designers will
be able to understand what product-features should be omitted from next generation product design optimization models and what product features should be considered.
Chapter 2

Literature Review

This chapter will discuss previously conducted work and their limitations in area of product preference acquisition techniques, mining customer preferences from online data and design axioms in customer preference modeling. The first section will elaborate on various techniques used by researchers to extract information about customers’ preference for products in the market. Next section will elaborate on extracting relevant information that could be useful to design teams while designing new products or next generation products and final section discusses design axioms in engineering design problems that generally deal with more than one alternatives to choose from for the final product design.

2.1 Product Preference Acquisition Techniques

Product features discussed on online product data streams prove useful during product development stage as they express collective wisdom and if utilized efficiently they can predict real world outcomes [2]. Product designers are able to extract features from abundantly available review data on blogging and social networking websites during product development [11]. The product developed according to customers’ needs are more likely to succeed than those that are an outcome of new technological innovations [12], [13]. A Group Decision Support System in combination with the Lead User Method proposed by Elfvengren et al. has a few limitations concerning the number of participants and hindrance to subtle communication due to limited customer communication [14]. Competition Factors within a group of products were introduced by Wagner and Hansen in combination with interviews, conjoint analysis, and questionnaires, to assess the features needed by key customers of a product [15]. Wang et al. employed a hybrid framework of the AHP and Kano Model to extract customer preferences
and use the results to recognize relationships between design features and market requirements using the Decision Making Trial and Evaluation Laboratory (DEMATEL) [8]. Kano model was integrated with Robust Design Approach in order to make customers feel content and in turn enhance their satisfaction [16]. Further development in the existing model named Analytical Kano Model was introduced by Jiao et al. as an advancement to the qualitative Kano Model that incorporates quantitative measures to customer satisfaction, thus making it more robust [17]. Customer requirements have been acquired by researchers to be used for Quality Function Deployment while designing new products using Fuzzy AHP approaches. These methods were then compared to the House of Quality Matrix and other similar techniques [18], [19]. Efforts by researchers to translate the voice of the customer (VOC) into product design concepts used fuzzy set theory, multi-attribute theory, fuzzy regression, optimization theory and group decision making to combine the subjective and quantitative information for QFD [20], [21]. Such features attracted Wu et al. to extract the key features of sub-compact cars that influence the choice of customers [22]. The multinomial probit model was used to explore the customer preferences relating to products [23]. Kano Method for product design and demand modeling in Decision Based design supported by Discrete Choice Analysis were proposed by Wassenaar et al. [24], [25]. In order to support a real time online recommender system to suggest products to the customers, Niu et al. constructed customer profiles / Profile Tree based on the demographical and behavioral data retrieved from the database of the e-commerce system [26]. Clustering Analysis was demonstrated by Trappey et al. as a tool for market segmentation based on preferences, demographics and other such features to provide services in accordance with customers’ expectations [27].

Efforts in the past have concentrated on mining product preferences from the text available online that is explicitly expressed by the customers[4], [7], [28]. There are some
features that are latent and expressed implicitly by the customers that are not captured by the mining algorithms. Such uncertainty in preferences ultimately leads to uncertainty in designing the product at the designer’s end. This paper outlines the uncertainty and risk involved in engineering design due to the absence of relevant information related to products.

2.2 Mining Customer Preferences from Online Data

Researchers have accomplished review classification and product attribute extraction using natural language processing tools and established classification models for textual data acquired from online sources. There have been attempts at building predictive models for prediction of sentiments of reviews pertaining to products or services [4], [6], [29], [30]. These efforts have not only helped customers but also enabled designers to understand customer requirements in the most appropriate manner [2], [5]. Feguina, Htay and Lynn, Bloom and Argamon have tested various algorithms to extract customer preferences and opinions regarding products from the customer reviews available on the online portals and the other social media [31]–[33]. Dave et al. classified customer reviews as positive or negative in order to quantify the importance of a feature in a product. They used structured reviews for testing and training and scored the reviews for determining whether reviews are positive or negative [29]. Tuarob et al. mined customers’ opinions from social media website related to a variety of cellphones in the market, extracted features and classified them as strong/weak/controversial [34]. Such feature extraction and demand prediction based on mining of online data available in the form of customer reviews, will enable design engineers to design products in accordance to the market trends [35], [36]. Tucker and Kim translated customer requirements into product design targets using Naïve Bayes model for predictive analysis and customer survey [37]. Rai provided a quantification model formulated to rank the features using product attribute ranking
metrics to know which attribute in a product holds importance for a customer [28]. Wang et al. illustrated the advantages of using ‘User Generated Content’ in online reviews, blogs and social networking sites for product design using text mining for feature extraction [11]. In a related approach, a document profile model was proposed by Jin et al. that extends the Apriori algorithm (association rule mining algorithm) in combination with Part-of-Speech Tagger, to enable designers to design products centered to customer preferences [38]. A supervised machine learning approach was used by Lee in order to extract use needs from online customer reviews from Eopinions.com which was further analyzed using unsupervised learning for extracting features and their specifications [39]. Stone et al. used opinions expressed on ‘Twitter’ in the form of tweets for sentiment classification using Support Vector Machine (SVM) and categorized the messages into different product features and levels for preference modeling [40]. In an attempt to extract key features to be incorporated in next generation products Tuarob and Tucker proposed a method to automatically identify the lead users from a pool of users while designing innovative or next generated products [41].

Mining customer preferences from the data available online in the form of social media posts or customer reviews enables designers to assess the existing design and at the same time, plan for future designs. Such quantification may be incomplete as customers do not discuss all of the features relating to a product and more importantly, product feature preferences may be heterogeneous and risk violating design axioms such as individual preference aggregation. This difficulty in discerning the reason behind not discussing certain features leads to uncertainty in product design. Uncertain decisions pertaining to product design may result in failure of a product in the market with the increasing competition. Thus, it is essential to gain knowledge about what product features have the potential to support the product in flourishing in the market.
2.3 Design Axioms in Customer Preference Modeling

Engineering design is a process based on decision sciences with a variety of techniques put into use while making decisions about a product or service design. Techniques like AHP, QFD etc. need to be validated before actual implementation of the decisions obtained on application of such decision making processes. Hazelrigg modeled uncertain behavior of design decisions that were an outcome of various decision analysis techniques and suggested validation of the technique used for making a decision so as to ensure non-violation of design axioms and arrow’s theorem [42]. Extracting features after mining the online customer reviews enables the designers to gain knowledge about the customers’ expectations. The methodologies discussed in the Section 2.1 are implemented to extract the features while Section 2.2 discusses the research conducted in the area of mining online preferences for product design. Researchers claim that the extracted customer preferences help the designers in preference modeling while designing a product by providing them ability to interpret the voice of customers [38]. Decker et al. developed an econometric framework to turn individual preferences from the opinions expressed online to aggregate preferences for product development and improvement processes [7]. In Design Optimization, approaches like QFD and Total Quality Management (TQM) for such aggregated preferences, can lead to erroneous results and are logically inconsistent [43]. Arrow’s Impossibility Theorem (AIT) states that in general, a group utility function does not exist. The multi-attribute value function built on a cardinal approach with respect to features preferred by individuals may not be consistent with the exact expectations of the individuals involved in the design process [43]. Research suggests that any method that requires inputs from a group of individuals to realize group preferences is likely to be flawed [44]. As the number of individuals in a group increase, the number of alternatives for a product’s design
increases based on the aggregated preferences of the customers. Such a situation gives rise to irrationality in design and thus a Customer-Centered view of design is not possible [10].

Table 1 Customer preference ranking of three features of a product.

<table>
<thead>
<tr>
<th>Customer</th>
<th>Feature Preferences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a &gt; b &gt; c</td>
</tr>
<tr>
<td>2</td>
<td>b &gt; a &gt; c</td>
</tr>
<tr>
<td>3</td>
<td>c &gt; b &gt; a</td>
</tr>
<tr>
<td>4</td>
<td>a &gt; c &gt; b</td>
</tr>
</tbody>
</table>

Table 2 Changed order of Preferences on introduction of an additional feature in the product

<table>
<thead>
<tr>
<th>Customer</th>
<th>Feature Preferences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a &gt; b &gt; c &gt; d</td>
</tr>
<tr>
<td>2</td>
<td>b &gt; d &gt; a &gt; c</td>
</tr>
<tr>
<td>3</td>
<td>c &gt; b &gt; d &gt; a</td>
</tr>
<tr>
<td>4</td>
<td>a &gt; c &gt; b &gt; d</td>
</tr>
<tr>
<td>Collective</td>
<td>b &gt; a &gt; c &gt; d</td>
</tr>
</tbody>
</table>

Franssen illustrated the reversal of preferences within a set of features on introduction of a new feature into the feature vector space as shown in Table 2 [45]. With a feature space of 3 in Table 1, the preference order for the product features is transitive and determined by the most preferred feature. In Table 2, it is evident that by introducing a new feature into the set of features, the preference order of customer changes and thus, the collective or global preference
order is changed. Researchers at MIT developed mathematical modeling based on game theoretic principles applicable to decision making in engineering design processes [46]. All the research conducted in this area is to avoid violation of the axioms of design and aggregation of preferences. Ho et al. modified the Nominal Group Technique to obtain customer requirements through discussions and then integrated agreed and individual criteria within the design team, to finally select a set of aggregated preferences for QFD [47]. However, Hazelrigg’s research concludes that designers may achieve sub-optimal solutions if individual preferences are aggregated for an engineering design problem [48]. Hazelrigg discussed proper use of decision sciences in engineering design problems in order to avoid uncertain behavior of other engineering design decision techniques. It was suggested through a mathematical framework that arrow’s impossibility theorem combined with axioms of design could be beneficial while reducing uncertainty in engineering design process [42]. Franssen discusses Arrow’s Impossibility Theorem in Multi-Criteria Decision problems by stating that engineering design is a team effort where each individual’s preferences vary. The problem of what preferences should hold more importance for the best design is exactly where Arrow’s Impossibility Theorem(AIT) becomes relevant in Engineering Design [45]. Hazelrigg’s research on engineering design alternative selection methods suggests that there exists uncertainty in the engineering design process due to the uncertainty in the desired future state of a design and hence, leads to poor outcomes [42]. In a group of rational individuals there exists uncertainty due to intransitive preferences of the designers in a team and thus a sub-optimal design is obtained [10]. Though techniques allow researchers to aggregate individuals’ preferences and design products in accordance to such preferences, it violates the AIT and are not recommended [49]. In essence, many researchers have cited Arrow’s Impossibility Theorem to be a restriction to modeling group preferences to satisfy all individuals. Arrow
asserted that there exists no procedure or voting system which fits best for any decision making process. Moreover, individual preferences cannot be ranked without comparison between each individual’s preferences by any possible procedure [50]–[52].

Customer preferences vary in nature and AIP states that such heterogeneous preferences cannot be aggregated and imposed on a group of individuals who do not have similar attraction towards the preferences considered for the design. The authors of this work propose a methodology that avoids violating Arrow’s Impossibility Theorem in engineering design by mining and quantifying “deal breaker” and “must have” product features that are homogeneously expressed within the product feature space. Ultimately, the work will result in quantifying the heterogeneity of customer preferences and provide vital information to designers regarding product features leading to market failures across various market segments. The paper aims at establishing a technique of providing vital information related to product features to designers using large online customer review or social media data. By analyzing customer opinions from online product data streams, designers can gauge the popularity of features that make their product successful. Concentrating resources on fixing the features that turn out to be “deal breakers” will help organizations maximize productivity and avert market failures.
Chapter 3
Methodology

The methodology presented in this thesis will quantify the degree of “must have” and “deal breaker” product preferences expressed online to enable designers understand what product-features should be omitted from next generation product design optimization models (i.e., “deal breaker” features) and what product features should be considered (i.e., “must have” features). Successful implementation of this approach will provide designers with knowledge of product features that are perceived positively or negatively across all customers. Product features receiving homogeneous positive sentiments will be categorized as being “must have” product features that in their current state are accepted to be benefitting across customers. Product features receiving homogeneous negative sentiments will be categorized as being “deal breaker” product features that in their current state are not acceptable across customers and thus need to be omitted or improved. Designers upon gain this knowledge would be able to concentrate resources in an optimized way on product features that make a product successful in the market while utilizing input that comes directly from customers using that product and thus ensure maximization of enterprise value. Figure 1 shows the general outline of the methodology followed to extract homogenous product feature preferences from online customer feedback data.
In Figure 1, f1, f2…fn are product features as discussed in customer reviews while the product remains constant or the methodology can be used to extract features from products that are similar within a domain. By quantifying the degree of “must have” and “deal breaker” product features expressed online, designers will be able to understand what product-features should be omitted from next generation product design optimization models and what product features should be considered. The steps involved are: 1. Text Mining and Feature Extraction, 2. Sentiment Analysis for Extracted Features, 3. Approximation of Sentiment Scores based on Available Data and 4. Discovery of “deal breaker” and “must have products”.

3.1. Text Mining and Feature Extraction

Social networking and other online platforms provide information related to customers’ experience about certain products. Preprocessed data is made available for sentiment analysis after cleaning residuals. The data available online is mined for features expressed by customers using Natural Language Processing techniques like POS Tagging, stop word removal and
stemming. For example, “iPhone 5 has a great battery life”. The textual data contains many instances like the example mentioned and when provided as an input to the processing tool, reveal keywords from the text corpus. The relevance score of these keywords serve as the basis to determine the initial feature space to start with. The description of a product serves as a means to extract relevant features from textual data available. In the above example, the customer discusses about the feature ‘battery’ while expressing his/her opinion. The extracted features are made available containing the relevant features of the product to be analyzed for the sentiments attached in the review of a customer. The algorithm box below summarizes the process of product feature extraction in large scale online data [53]. Relevance score is obtained for features of a product from the manufacturer’s description. Relevance score is calculated based on the term occurrence and importance of the terms across documents. All the terms are rank ordered according to their magnitude. Relevance score of 0.5 or 50% and above is considered to be a threshold for populating the feature space with keywords from the description. After feature space is created, every review message is analyzed for extracting occurrence of features from this feature space. Thus, finally a dataset is created with text input and the relevant feature extracted from this input. Essentially, all the information that relates to product under consideration is contained in the product’s feature being discussed and the customer’s opinion towards that product feature.
3.2. Sentiment Analysis for Extracted Features

A sentiment can be defined as ‘opinion or feeling of a customer towards the feature of a product’. A positive sentiment for a feature expresses that the customer is in favor of the feature and would like to see the feature as part of the product. A negative sentiment would mean that the customer is not happy if the feature becomes part of the product in the existing state. Such sentiment analysis enables the designers to understand the inclination or disinclination of customers towards the product’s features and design accordingly, while developing new products or advanced version of the same products. For example:

Customer A: ‘Battery backup is probably the best feature of my iPhone’

Customer B: ‘This phone lags too much while opening applications installed from the App Store’

In the reviews above, Customer A discusses the feature ‘Battery’ and a positive sentiment is attached to it while Customer B’s review is mapped back to the speed of the iPhone attached with a negative sentiment which maps to the component “Processor” as per the manufacturer’s
description. Every positive and negative sentiment gets a sentiment score in accordance to the degree of positivity or negativity expressed in the sentence. The positive and negative scores received by the features can help in determining the probability of whether a feature will impress the customer or not and in turn, provide information related to the demand of the product [34]. We cannot simply state that the feature receiving the maximum negative score as a deal breaker for everyone and should be removed from the product. The essential components of a product cannot be omitted from the product and thus being a “deal breaker” or a “must have” has to be further explored by the design team. However, improvement in the current state of the feature is decided by classifying it as a deal breaker or a must have. While establishing a relationship between customer preferences and intent to purchase a product, Haverila discovered that business functionality of the product had a great impact on the repurchasing intent of customers [54]. Thus, the features needed for a product to carry out its functions have to be incorporated into the product. By classifying the features as “must have” and “deal breaker”, designers can focus on improving the features that are “deal breakers” in their current iteration of a product, based on the customer’s opinions. An effort to determine the impact of the addition of attractive features in a product by Chernev brought into light the fact that people have pre-established preferences for a brand or product of their choice. The addition of attractive features in a product enhanced the preferences of customers [55].

### 3.3. Approximation of Sentiment Scores

Consider a product ‘Z’ with features $f(a, b, c, d, e)$. Suppose a customer of ‘Z’ expresses his/her opinion as:

> “Product ‘Z’ is great. I like ‘a’ and ‘c’ but the quality of ‘b’ is somewhat unsatisfactory”.

Another customer expresses his/her opinion as:
“Product ‘Z’ is a great companion. I am in love with the way ‘c’ helps me manage my routine”.

The example illustrates the varying opinions of customers related to a product where customers seem satisfied with the product feature ‘a’ and ‘c’, unsatisfied with product feature ‘b’ and do not mention anything about features ‘d’ and ‘e’. While making inferences from the extracted data and analyzing the reviews of customers, there exists uncertainty when customers do not express their views about some features. In such a scenario, several possible reasons why customers did not express opinions about product feature ‘d’ and ‘e’ can be formulated:

a. The customer assumed that these features will not be included in future products because they are undesirable in the market (e.g., having lead based products)

b. The customer assumed that these features will be included in future products because they are desirable in the market (e.g., seatbelt in a car)

c. The customer is unaware of the existence of feature i in the market (e.g., bendable phone screens that are at their nascent stage)

Table 3 Uncertainty in Optimal Product Design

<table>
<thead>
<tr>
<th>Feature</th>
<th>Customer 1</th>
<th>Customer 2</th>
<th>Customer 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>b</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>c</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>d</td>
<td>?</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>e</td>
<td>?</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Optimal Design</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Such possible reasons lead to uncertainty in engineering design. An added customer will express varying feature combinations and express different opinions regarding the features discussed above. Thus, it becomes difficult to make inferences about the product feature space and the sentiments expressed. With such anomaly in the design process, another issue that
makes the engineering design process difficult is the aggregation of preferences. For product ‘Z’, designers will conclude that the presence of feature ‘c’ is a “must have” as this represents a homogeneous product preference. This aggregation will lead to violation of the Arrow’s Impossibility Theorem if the sentiment relating to a given feature is not homogeneous across all product features, as the design will be centered to satisfying a greater fraction of the population while neglecting the features preferred by lesser number of customers. With the increase in number of customers, the intransitive preferences increase, making the process even more complicated and confounding.

A product is made using a set of sub-assemblies or components each of which serve a different function to the user. Customers while writing reviews about a given product, typically use non-technical language. For example, when a customer tweets, “This phone cannot stay on even for a day. I charge every 2 hours!!!!!!”, it can be mapped to the “Battery” of the phone. Similarly, there are other instances where the customer may not use the word “battery” and instead use phrases like “stay on”, “charge” etc. to discuss about “battery”. John and Tucker proposed a review-feature mapping methodology to quantify the price and demand of re-synthesized products [53]. In essence, a customer may not talk about each and every feature of a product. However, a customer may provide feedback relating to two features that are correlated or either map to the same component of the product. For example:

Customer 1 – “I am not able to fit this phone in my pocket”
Customer 2 – “It is difficult to carry this phone all the time as it is not handy”

Each of customer’s statement maps to the “size” of the iPhone without even mentioning the word “size”. Customer 1 expresses negative sentiment towards size of iPhone comparing it size of his/her pocket. Customer 2 expresses same view about size suggesting difficulties in typing on the screen due to limited reach of the fingers. Figure 2 illustrates how two different reviews
can map back to the same component of a product and a general sense of popularity or unpopularity of the component and its features is established. Thus, correlation determined between the features of the product and the opinions expressed about a specific feature can provide an insight about the correlated feature.

Figure 2 Mapping to the component based on features discussed in customers' reviews

Since customers do not typically specify the exact terminology pertaining to a product feature, related words are clustered by employing WordNet [56]. A product is made up of various sub-assemblies and each sub-assembly serves a different function either alone or in combination with other sub-assemblies. When a customer expresses his/her views about a ‘Sub-Assembly – a’ which correlates to ‘Sub-Assembly – b’ to build ‘Component c’, then the sentiment is mapped through the sub-assemblies building the component. For example, a review related to the picture quality of a photo taken with iPhone and a review related to absence of ‘panoramic view’ in an iPhone map back to the component ‘Camera’. Thus, correlation between the features of a product enables mapping of sentiments expressed about one feature to another.
3.4. Discovering Homogenous Product Feature Preferences

Consider product ‘P’ with feature vector \( f(a, b, c, d, e) \). Suppose we mine reviews posted by two individuals from online customer review data and extract the features discussed by them. Let A and B be two customers who discuss about features \((a, b, c)\) and \((c)\) respectively. The unknown reasons behind the absence of discussion about \((d, e)\) poses risk to understanding whether these features are indeed relevant or desired in the market. Such missing product features do not provide any inference to the designer and thus a predictive model is essential in inferring the customer sentiments about such features.

Suppose customers A and B expressed positive sentiments regarding \((c)\) and customer A expressed negative sentiments regarding \((a, b)\). It is therefore assumed that a positive sentiment indicates that the customer likes the feature more than the other features and thus values it more. A negative sentiment indicates that the customer dislikes the feature and thus, does not value it over the other features. In the example, the sentiment score of \((c)\) is greater than \((a, b)\) because it is rated positive by more number of customers and thus becomes homogeneous.

**Scenario 1:** Sentiment Score\((c)\) > Sentiment Score \((a, b)\)

At the same time, there exists uncertainty while modeling the sentiment for \((d, e)\) because none of the customers discussed the features. Possible conditions are:

**Scenario 2 (a):** Sentiment Score \((c)\) > Sentiment Score \((d, e)\)

**Scenario 2 (b):** Sentiment Score \((c)\) = Sentiment Score \((d, e)\)

**Scenario 2 (c):** Sentiment Score \((c)\) < Sentiment Score \((d, e)\)

The sentiments as expressed by customers are the input for Table 3. However, not all customers discuss all the features and hence there are missing values in the features table. A predictive model based on Expectation Maximization Algorithm using correlation matrices is proposed.
that estimates the missing values for features corresponding to specific customers using the values already available. The EM Algorithm is a general method of finding the maximum likelihood estimate of the parameters from a given data set when the data has missing values. Suppose X1 and X2 are two features with a number of instances and missing values for those instances. We calculate the mean $\mu_1$ and $\mu_2$ for both the variables.

Co-Variance is estimated using Eq. 1

$$\sigma^2_{ij} = \frac{\sum (X_i - \mu_i)(X_j - \mu_j)}{n}$$  \hspace{1cm} (1)

The first estimate of correlation is calculated using $\sigma^2_{12}, \sigma^2_{11}, \sigma^2_{22}$ and Eq. 2

$$\rho = \frac{\sigma^2_{12}}{\sqrt{(\sigma^2_{22})(\sigma^2_{11})}}$$  \hspace{1cm} (2)

A linear model is formulated for finding the missing values in the data set.

$$X_1 = \mu_1 + \frac{\sigma^2_{12}}{\sigma^2_{22}(X_2 - \mu_2)}$$  \hspace{1cm} (3)

$$X_2 = \mu_2 + \frac{\sigma^2_{12}}{\sigma^2_{11}(X_1 - \mu_1)}$$  \hspace{1cm} (4)

Suppose reviewer 1 assigns a negative score of -0.75 in his/her opinion expressed about picture quality and reviewer 2 assigns a positive score of 0.5 to the video quality of the phone. Both the reviews map back to the same component “camera” of the phone. There exists correlation between the features of a product and thus it is assumed that all the reviews that are expressed about features that map back to the same component are correlated [57]. Thus, reviewer 1 will express a negative sentiment towards all the features who have a correlation with the camera component and reviewer 2 will express a positive sentiment towards the features that have correlation with the feature camera. Such estimation of all the sentiment scores attached with features in the feature-customer matrix helps in determining the mean scores of the features. For example, with missing values in the feature column of ‘size’, we calculate the mean to be 0.63 and mean for feature ‘screen’ is calculated to be 0.54. Using the above values and Eq. (1),
we can calculate the covariance between the features and finally the correlation to assign values to all the missing cells based on the interaction between features. These steps are followed until convergence, and the values for all the missing data (that have correlated values with other features with known sentiment scores) are calculated.

After attaining the missing values based on EM Algorithm, the data set is analyzed to count the number of positive and negative sentiments expressed towards a feature. The number of positive and negative values obtained by features determine their classification into “must have” and “deal breaker” features. Features receiving positive sentiments values across all customers are classified as “must have” and features receiving negative sentiment scores across all customers are classified as “deal breakers”. However, there exists heterogeneity in the preferences of customers for features of a product. If a feature has received maximum but not all positive sentiment scores across customers, it is essential to investigate the anomalies based on customer’s sentiment towards other features.

For a given customer ‘i’, expressing negative sentiment towards a feature \( f \in F \) that has received positive sentiment from all other customers:

Let \( p = \max \) (positive sentiment values for ‘i’ across \( F \)), \( n = \max \) (negative sentiment values for ‘i’ across \( F \)), \( \epsilon = \) tolerance constraint for each customer ‘i’, \((s_i \in S)\) = sentiment value for customer ‘i’ assigned to feature \( f \).

\[ \epsilon = \frac{(p + n)}{2} \]

**Figure 3 Classification of Sentiment Scores into Positive or Negative**

If ‘\( s_i \)’ falls within the tolerance limit of the customer’s range of sentiment values across \( F \), it is assumed that customer is indifferent towards the feature. Such instances occur when customer
do not express strong negative or strong positive sentiment towards a feature. For example, “Camera of my iPhone is good enough to click pictures in daylight”, explains that the customer is neutral towards feature ‘Camera’. If ‘s_i’ exceeds ‘ε’ on either the negative or positive side of the customer’s range, then the sentiment is classified into negative or positive respectively. Figure 3 represents the scale of a customer’s sentiment scores across features. ‘ε’ is the threshold or tolerance of the customer’s sentiment scores towards a feature. Transition in customer’s preferences occurs when ‘s’ exceeds ‘ε’ on either side of the tolerance limit. A customer’s sentiment score may exceed ‘ε’ while expressing strong positive or strong negative sentiment towards a feature. In such a situation, a positive or negative sentiment score is assigned to the feature for a customer [58],[59]. Value of ‘ε’ governs the exclusion of noisy data and if the sentiment is strongly negative or positive beyond the epsilon value, it is assigned to the feature. Based on the classification of positive and negative sentiments for the anomalies and approximating the noisy data to fall under one of the respective classification, we may get heterogeneous and homogenous sentiments towards the features.

For a feature ‘f’,

\[ Case 1: S = \text{Positive, } f \text{ is “must have”}. \]

\[ Case 2: S = \text{Negative, } f \text{ is “deal breaker”}. \]

\[ Case 3: S = \text{Positive} \leq \text{Negative or Positive} \geq \text{Negative, } f \text{ is unclassified}. \]

where S is sentiment values assigned by customers.

If all sentiment scores for a feature ‘f’ are positive we classify it as a “must have” and if all the sentiment scores for a feature ‘f’ are negative, we classify it as a “deal breaker”. When the sentiment scores for a feature are a mix of positive and negative values, it remains unclassified. Case 1 results into classification of a feature as “must have” as it receives positive sentiment score across all customers and Case 2 results into classification of a feature as “deal breaker”
as it receives negative sentiment score across all customers. Case 3 does not result into any specific classification for a feature.
Chapter 4

Case Study: Design Feature Extraction for the Apple iPhone 5

4.1 Application of the methodology

To demonstrate the methodology shown in Figure 4, a smartphone design problem is used as an example. The case study will illustrate the difference between aggregating the product feature preferences as expressed by customers and the methodology as proposed by the authors. Product feature preferences as expressed on online product data streams such as Twitter are leveraged in this work. Maximization of enterprise value relies on product sales and it has been shown that a product’s performance is highly correlated to customers’ opinions on product selling websites or social media platforms [34]. It is thus recommended that designers assess current quality characteristics of a product before progressing ahead with design changes in next generation of that product. Following this methodology will ensure that designers focus on product features that are liked across all customers for inclusion and improvement while omitting product features that are disliked across customers. In order to achieve this objective, it is essential to analyze customer feedback posted on online data streams and extract relevant information to provide meaningful insights to product designers. The methodology described in this work leverages text mining and statistical techniques to develop a model that will enable product designers understand customers better and thus make informed decisions. Data acquired from online sources is cleaned, processed and evaluated for product feature extraction. The acquired data is also analyzed for customers’ opinion extraction and evaluated for homogeneity towards product features.
Twitter data related to the iPhone 5 is analyzed to serve as an example for the above methodology. The data contains tweets posted by twitter users about iPhone exclusively [60]. A tweet is a text or a sentence that contains information about the product. For example, “iPhone has amazing camera quality #iPhone5” illustrates clearly that the poster is impressed with the camera quality of iPhone 5. 100 tweets are mined using Alchemy API. Features are considered to be keywords in the sentences expressed by customers. Features are the words that map to the product description and functionality of the product. The text data forms input for the API and features are extracted using the Alchemy API, which uses Natural Language Processing techniques and relevance score for words, based on the text corpus for keyword extraction. All the tweets convey some information regarding the phone and its features. Not all the tweets are able to convey the same features of the phone (and hence the need to infer
sentiments for missing features).

![Figure 5 Sentiment Analysis output by Alchemy API](image)

Figure 5 Sentiment Analysis output by Alchemy API

Figure 5 provides an example of the output in Alchemy API. An instance of a tweet containing “camera” serves as input to the Alchemy API and is analyzed for the product feature sentiment. The output in Figure 4 illustrates the analysis of the opinion or review expressed in the textual content. The feature “camera” received a negative sentiment of magnitude 0.727983 from the customer who posted the tweet. Similarly, all the tweets are analyzed to acquire sentiments and magnitude of the negativity or positivity in the text corpus posted on online product data streams. The sentiment scores serve as the base for final analysis of the product’s features and provide vital information related to the specific feature.

4.2 Results

The features and sentiment scores were extracted using open source Alchemy API built on natural language processing techniques and relevance of the words in the text corpus to process text. The Alchemy API adds semantic annotations to tweets by suggesting the relationship of the keyword with the other parts of the text. The output gives a score for the degree of positivity or negativity based on the words used by the user for a particular feature in a visual form as shown in Figure 6 and Figure 7.
The textual content was filtered for profanity and cleaned before analyzing for positive and negative sentiments towards features. Figure 6 provides information regarding the positivity in the text input to the API. The size of the box denotes the magnitude or severity of the sentiment. For example, a bigger size rectangle denotes higher magnitude of the sentiment expressed.
Figure 6 shows the magnitude of the sentiments expressed in the statements by customers. The size of the rectangle is based on the score received by the sentiment in a statement. In similar manner, Figure 7 provides information about the negativity in the textual content, with the size of the rectangle determining the magnitude of the negative sentiment.

Mining of the tweets related to iPhone 5 revealed the following features:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Battery</strong></td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td><strong>Charger</strong></td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td><strong>iOS</strong></td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td><strong>Phone Case</strong></td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td><strong>Weight</strong></td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td><strong>Camera</strong></td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td><strong>Siri</strong></td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td><strong>Price</strong></td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td><strong>Size</strong></td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td><strong>Screenshot</strong></td>
<td>20</td>
</tr>
</tbody>
</table>

The iPhone 5 users expressed their opinions about the features in Table 4 explicitly. Customers expressed their opinion regarding various features possessed by an iPhone. Mining such tweets suggested that the features listed in Table 4 are the talked about features. Customers explicitly mentioned about camera, size, Wi-Fi, screen and other features of the iPhone. All the listed features serve as the base for iterating towards the optimal feature space. Customers do not generally talk about all the features while expressing their opinions on online digital platforms. Thus, there exist missing values for almost all the features after analysis of the textual data. For
missing values in all such customer rows, the predictive EM Algorithm is employed with various sentiments scores for the variables as input. Using the statistical software package, SPSS (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp) missing values are calculated. The EM Model in SPSS assigned values to the missing cells based on the available data set. At every step, the EM Algorithm in SPSS analyzes the available data and assigns a value to the feature for a particular customer based on the sentiment or value of the other available customers assuming correlation between the features of the iPhone and the opinions expressed by users of the iPhone. The EM Algorithm approximation was obtained based on the available score for a feature as expressed or assigned by a customer. Thus, scores for a given feature were obtained based on the score available for another feature for a customer. All the features are not discussed by a single customer and thus correlation between features enables approximating the scores for related features [57]. Not all the scores for a feature were only positive or only negative and thus the anomalies were classified as positive or negative based on the mean and standard deviation of the scores assigned by customers. ‘ε’ threshold of one standard deviation and mean of the customer’s sentiment scores were calculated. Features with all positive and all negative sentiment scores were classified as “must have” and “deal breaker”. However, not all features received homogenous scores and thus further analysis was performed for classification. All the sentiment scores falling within the ‘ε’ threshold were considered to be sentiment scores by customers who are indifferent towards the features. The sentiment score beyond the ‘ε’ value were assigned to the feature as customer’s sentiment value for that feature. Such instances occurred for customers who expressed strong positive or strong negative sentiment towards a feature. Thus, while all the other sentiment scores for a feature were positive or negative, such anomalies resulted in unclassified features. The “must have” features have positive sentiment
values across all customers as opposed to the “deal breakers” which have negative sentiment scores across all customers. Features that received heterogeneous sentiment values are not classified into any of the two categories and remain unclassified. Some features were not expressed by many customers and thus received negligible number of data points or sentiment scores. Missing values with such data points of negligible magnitude and quantity were approximated to be zero by the EM Algorithm. The exclusion of such features in discussions online can occur either because they are standard features and thus customers do not express anything about those features or because customers do not like the features and thus do not want to comment on them.

### Table 5 Final Feature Classification

<table>
<thead>
<tr>
<th>Must Have</th>
<th>Deal Breakers</th>
<th>Unclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light Weight</td>
<td>Battery</td>
<td>Speaker</td>
</tr>
<tr>
<td>iOS</td>
<td>Screen</td>
<td>Processor</td>
</tr>
<tr>
<td>Wi-Fi</td>
<td>Screenshot</td>
<td>Siri</td>
</tr>
<tr>
<td>Alarm</td>
<td>Cord Length</td>
<td>Front Camera</td>
</tr>
<tr>
<td>Phone Case</td>
<td>Earphones</td>
<td>Size</td>
</tr>
<tr>
<td></td>
<td>Price</td>
<td>Connectivity</td>
</tr>
<tr>
<td></td>
<td>Headphone Jack</td>
<td></td>
</tr>
</tbody>
</table>

The feature space in Table 5 partitions the features mined into two categories. i.e., “Must haves” and “Deal Breakers”. There are some features that are either not expressed or are expressed by negligible number of people and are hence, considered to be irrelevant or noise in the data. Scores for such features are approximated to be zero or negligible by the algorithm due to less number of data points for approximation.
The results presented in this paper are limited to two categories (i.e., “deal breaker” and “must have”) and does not classify the other features into any other category. The ‘deal breaker’ features need special attention from the designers as they can result in the product’s market failure. The ‘must have’ features are the driving force for the product and they attract customers due to the quality and magnitude at which they are implemented in the product. If individual preferences are aggregated for a feature, designers will design for a fraction of the population leaving some individuals unsatisfied. In essence, a “deal breaker” for one individual may be “must have” for another or vice-versa. For example, it is necessary for a phone to incorporate a battery in its body in order to exist. Ignoring the product features that result in existence of the phone will not prove to be an effective strategy and thus improvement in such features is suggested. The existence of a project utility is a must in engineering design if the ultimate objective of the efforts is to make the design rational, maximize demand and enterprise level profit.
Chapter 5
Conclusions and Future Work

5.1 Summary

The results of the methodology illustrate the application of a different approach towards engineering design. Designers should use such an approach to optimize the designs instead of aggregating product feature preferences expressed by customers based on the sentiments. Using such a model takes into account the sentiment of each customer in the model in every iteration and provides the basis for the designer to choose between the ‘Deal Breakers’ and ‘Must Have” features. Arrow’s Impossibility Theorem provides a foundation for understanding the challenges of determining an optimal design in engineering due to the variations that exist in customers’ preferences. While taking such decisions pertaining to the products’ design, many individuals provide preferences for product features in order to maximize the expected benefit. The authors of the paper outlined a methodology for discovering homogeneous product feature preferences discussed in online customer review sources, so as to provide designers with relevant information, while being consistent with the engineering design axioms. By improving the current state of unpopular features from a product (i.e., “deal breakers”), designers will alleviate the risk of product failure in the market segments.

5.2 Scope for Future Work

This concept can be easily extend to any other domain that interacts with users and serves a purpose. Education, healthcare and professional applications are potential implementation of this strategy. In the healthcare domain, while designing computer applications like Nike Run, this concept could highly impact adherence of patients to the treatment protocols. Educational
applications are developed to motivate students and impart practical knowledge. Such applications can take into consideration feedback of students to assess the inclusion or exclusion of features into the applications. Similarly, any product domain could be benefitted by information retrieval from feature preferences. Recently, Apple Inc. decided to remove the headphone jack from next generation iPhone which is currently under development. There are a variety of reasons including technological advance, size issues etc. to take a step in this direction. However, such decisions are also affected by product feature preferences that are mined from customer reviews. In this case study, the headphone jack was determined to be a “Deal breaker”, which is an evidence of how information obtained from online data sources play a vital role in engineering design.

The current methodology is limited to positive and negative sentiments expressed for product features with homogeneous preferences. Also, while setting the threshold for classification of sentiment scores ‘$\epsilon$’ was considered to be one standard deviation from the mean. Future work in this direction could be conducted find the optimal ‘$\epsilon$’ value for classification of the sentiment scores into positive or negative and extract latent features expressed in the statements on online product data streams for stronger and efficient analysis. Furthermore, increasing the scope of the methodology will enable classifying the ‘unclassified’ features in an appropriate manner. Moving in this direction to benefit designers analyzing customer reviews from a different perspective to alleviate the risk of misinterpretation of low product ratings can be pursued by researchers. Customer reviews are accompanied by product ratings, which are a measure of the product’s popularity and quality in the market. These product ratings may not always relate to product’s directly related features like form, function and behavior of the product. In fact, many of these ratings whether low or high are due to service issues with the retailer or online selling website. Customers provide an opinion about
their experience with the product which can be mined to enable designers optimize resources like time and cost by concentrating them on to things that make significant contribution to a product’s success and not worry about low product ratings that are not at all representative of the product’s features. Also, the main contribution of this work was to establish a method that can extract features in order to ensure product success. This work was limited to iPhone, a tangible product and extracted features that should be omitted or included in next iteration of the product’s design. However, this methodology can be extended to any product or service with characteristic product features that serve specific functions. For example, evaluating success of applications on mobile gaming platforms, game designers can analyze relevance of game design features which are considered to be the building blocks of a game. Customers that install games and applications from the game stores are allowed to post comments pertaining to their experience while interacting with the applications. These comments are related to game design features that make a game engaging and motivating. Sentiment analysis of such comments mapped to different aspects of a game will enable game designers to extract “must have” and “deal breaker” features ensuring success of that game in the market. In similar manner, this methodology can be utilized to deliver satisfactory products and services to customers while maximizing enterprise value. Thus, quantification of product feature heterogeneity facilitates informed product design decision making.
References


[34] C. S. T. Suppawong Tuarob, “Fad or Here to Stay: Predicting Product Market Adoption and Longevity Using Large Scale, Social Media Data,” 2013.


