A Unique Method of Constructing Brand Perceptual Maps by the Text Mining of Multimedia Consumer Reviews

Amir Ekhlassi, University of Tehran, Tehran, Iran
Amirhosein Zahedi, University of Tehran, Tehran, Iran

ABSTRACT

Brand perceptual mapping is a visual technique, it displays how a brand is positioned in the mind of customers, as well as in relation to the competitors. With the rapid growth of e-commerce and the abundance of online consumer-generated content, there is no need for marketers to go through market research in order to understand consumers’ opinions. Therefore, in this study, the authors propose a unique method which allows the building of a perceptual map automatically by mining consumer opinions from in particular online product reviews. The authors employ opinion mining techniques to extract and rank the product aspects that are important to customers, during purchasing digital tablets. Subsequently, they generate a score for each brand in these aspects and build the perceptual map using clustering of the brands by these scores. This proposed method is applied to the online customer reviews for digital tablets obtained from Amazon.com. The experimental results highlight the proposed technique is effective and able to correctly depict the position of a brand in its particular competitive environment.

KEYWORDS
Brand Perceptual Map, Brand Positioning, Consumer Generated Content, Online Product Reviews, Perceptual Mapping, Positioning, Text Mining, User-Generated Content

1. INTRODUCTION

To succeed in our over communicated society, companies must manage their brands through creating a distinct position in the minds of prospective customers. (Ries & Trout, 2010). For performing positioning analysis, marketers widely use perceptual mapping tool (Aggarwal, Vaidyanathan, & Venkatash, 2009). Traditionally, the information needed for the design of a perceptual map would be obtained from comprehensive market research studies. In these methods, you want the customer to score on different aspects of several brands simultaneously (by surveys, interviews or similar techniques). While the existing approaches for perceptual mapping contribute to our understanding of the consumers’ behavior, they are also associated with various drawbacks, such as the limited sample sizes and the complications involved in developing a survey that is able to fully capture the

DOI: 10.4018/IJMCMC.2018070101

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
consumers’ perceptions (Crotts, Mason & Davis, 2009). However, over recent years with the changes in consumer behavior, online shopping is increasingly becoming people’s first choice when shopping (Singh, Irani, Rana, Dwivedi, Saumya, & Kumar Roy, 2017). Along with that, an exponential growth has occurred in the individual’s activities in online channels of communication. Nowadays random conversations about brands are now more credible than targeted advertising campaigns and social circles have become the main source of influence overtaking external marketing communications and even personal preference (Kotler, Kartajaya & Setiawan, 2017). For instance, a recent study on the www.brightlocal.com suggested that 87% of buyers read 10 or less than 10 reviews before trusting a business (Singh et al., 2017).

Online channels of communication contain brand-related information and descriptors that define brand positions in the offline world (Aggarwal et al., 2009). Therefore, analyzing the vast amounts of available web-based information could replace polls, focus groups and other similar techniques used for market research (Marrese-Taylor, Velásquez, Bravo-Marquez, & Matsuo, 2013). However, due to the sheer quantity of this type of data (weblogs, online reviews, discussion boards and other unsolicited forms of consumer opinions), manual tracking of all the available data and processing of the competitors’ activities might be tedious, inaccurate, and rapidly outdated for marketers (Leong, Ewing, & Pitt, 2004). Therefore, a different approach than conventional marketing methods is required for the analysis of large datasets for research purposes (Krawczyk & Xiang, 2015). With this background in mind, an area of increasing attention among retailing researchers and strategists is the positioning of brands based on opinion of the online users (Ailawadi & Keller, 2004). This made academics and practitioners to try to construct algorithms for the effective analyzing of the valuable data of consumer forums, blogs, and product reviews. However, due to the lack of effective methods to extract the key features of these online texts, businesses might be unable to obtain useful information to develop a market structure map.

Technology offers potential solutions for this problem by the distillation of knowledge from huge amounts of unstructured web-based information, which can play a key role in various commercial tasks, such as social media marketing, branding, product positioning, and corporate reputation management (Poria, Cambria, Winterstein, & Huang, 2014). Collaboration of computer scientists and business researchers has often facilitated the dissemination of these tools for business research purposes, thereby resulting in fruitful research initiatives by providing opportunities for the quantitative exploration of new business data sources (Das & Chen, 2007; Feldman, Fresco, Goldenberg, Netzer, & Ungar, 2007; Lee & Bradlow, 2011; Netzer, Feldman, Goldenberg, & Fresko, 2012). To manage the challenges involved in extracting information from consumer forums, researchers have proposed the sentiment analysis techniques, which leverage text mining and natural language processing (NLP) so as to derive meaningful information from the data of customer reviews. Nowadays, data mining is rapidly gaining acceptance as an important technology to give businesses a competitive edge, and there is also growing interest in the text mining technologies to discover the knowledge that is buried in unstructured texts (Leong et al., 2004). The information elicited through the text mining methods from online consumer-generated content (CGC) regarding a product’s attributes and the consumer’s perception towards a brand can be useful in developing a perceptual map for the brand (Lee & Bradlow, 2011; Netzer et al., 2012; Chen, Kou, Shang & Chen, 2015). For example, Chen et al. (2015) have confirmed that the perceptual maps derived from online consumer-generated data are remarkably effective in depicting the market structure such as demonstrating the positioning of competitive brands.

From a managerial perspective, it would be helpful if researchers provided an efficient and easily applicable method for managers, so that they could continually monitor their brands and ensure that this consumer-based positioning was consistent with the marketer-intended positioning (Aggarwal et al., 2009). Nevertheless, application of CGC for exploration of positioning research has only recently been discussed in a handful of studies (Lee & Bradlow, 2011; Netzer et al., 2012; Krawczyk & Xiang, 2015). The challenging gap in this regard is that the large amounts of reviews may lead to information
overload, and therefore, classifying these unstructured data (i.e., big data) in the form of CGC is clearly associated with big data management challenges (Singh et al., 2017). Owing to its qualitative nature, which is very similar to the elicited data from focus groups or in-depth interviews (but on a much larger scale), analyzing this useful source of data, could help marketers to monitor customer opinions, attitudes and preferences to design their products and services accordingly (Netzer et al., 2012). Although various models and algorithms have been proposed by researchers and practitioners in this regard, the current algorithms have several limitations to be widely applied by marketers and managers to derive helpful information from the online opinions of customers about their brands. The main limitation of these models is the inadequate accuracy of these algorithms, which the model presented in this research tries to solve this main limitation without the human intervention.

In this paper, by relying on text mining and NLP, we proposed a unique method for constructing perceptual maps for commercial brands by the automatic mining of online product reviews. We developed a procedure which downloads a number of online reviews for a specific product category and analyzes the collected reviews in an inexpensive, accurate and fast manner in order to extract the consumers’ opinion. The core of the proposed method consists of the aspect finder and opinion ranking algorithms. The aspect finder algorithm mines data and determines the important aspects involved in the decision-making of customers. The opinion ranking algorithm mines data in order to discover the customers’ opinions on the extracted aspects and assigns scores to the different aspects accordingly. The final aspect scores are used to construct a visual brand perceptual map. Of note, the proposed method is not limited to a specific product category and is able to retrieve customer opinions from any online review data. In the present study, we have applied the mentioned methodology to a database of online reviews for digital tablets extracted from www.Amazon.com. According to the experimental results, the proposed technique is effective for the accurate evaluation of commercial brands based on the online reviews of consumers and is also able to correctly depict the position of a brand in its competitive environment.

This study shows that CGC is a free valuable input data for marketers to be used in their brand positioning formation, i.e., a large amount of qualitative unstructured data of CGCs are mined effectively and precisely through the presented method, and the output is a series of easy-to-use, clear and rich structured data, which constructing a perceptual map based on them, leads to a superior brand positioning strategy formation. Essentially, this has a significant impact on marketing research processes of the company. We believe that the previous models in the existing literature of combined text-mining techniques and marketing only represent the tip of the iceberg, and our research adds another dimension to these efforts. The main contribution of this study in comparison to what has already been done is that the unique presented algorithm here is superior to the benchmark competing algorithms developed in existing literature, in that it represents higher accuracy compared to them and also this method is easier to be practically used by marketers in the real-world business.

This paper has been organized in the following manner: Section 2 discusses the literature review, Section 3 describes the methodology of the study, Section 4 discusses the experimental results, and the conclusion has been presented in Section 5.

2. LITERATURE REVIEW

Brand position is the standing of a brand in comparison with its competitors in the minds of customers, prospects, and other stakeholders (Duncan, 2008). The challenge is to select a position that can be realistically supported by the product, the company, and the marketing communication and that can be appreciated by consumers and prospective customers (Duncan, 2008). To overcome this challenge companies need to determine how customers perceive a brand and its competitors. Therefore, one thing market researchers perform is to ask a sample of customers to participate in perceptual mapping. (Duncan, 2008)
Perceptual map is the most widely used tool for positioning analysis (Aggarwal et al., 2009). In a perceptual map, brands within a product category are plotted on a multidimensional space, so that companies are able to supervise the key dimensions that differentiate their brands in the consumers’ minds (Krawczyk & Xiang, 2015; Shocker & Srinivasan, 1979). These maps can be used as the means to understand and visualize the perceptions of consumers towards the features of a product (e.g., price, quality, and attribute rating), identify the market gaps and preemptive positions to discover new product opportunities and positioning strategies, and evaluate the reputation/image of a brand (Trappey, Trappey, Chang, & Chen, 2016; Lilien & Rangaswamy, 2002). To date, numerous empirically applied cases of perceptual maps have been presented; for instance, in lodging industry hotels use perceptual maps to identify the position of their brand among others in order to formulate actionable strategies to (re)position themselves in the marketplace (Krawczyk & Xiang, 2015).

In the literature, several techniques have been proposed for constructing perceptual maps, Techniques such as similarity scaling, factor analysis, discriminant analysis, correspondence analysis, etc. Most perceptual mapping techniques typically rely upon the association between objects and descriptive attributes which are obtained by the traditional market research tools (Ivy, 2001). However, in traditional market research tools (e.g., face-to-face interviews, phone surveys, and online surveys), customers do not always tell marketers what they really think and do. In fact, they are not always able to articulate what they really think and do, even if they want to (Kotler, Kartajaya & Setiawan, 2017). This is despite the fact that, in online platforms of communication individuals freely express their real opinion about a product. This situation has caused that today, online CGC an important element in the decision-making process of customers for purchasing products and affect potential customer’s perception of products and brand image (Gensler, Völkner, Egger, Fischbach, & Schoder, 2015), ultimately influencing the sales success (Moe & Trusov, 2011).

As a single review it may be unlikely to yield important insights from online CGC, because a large body of text must be collected and distilled for relevant findings. This is in contrast to traditional methods which may be able to use a much smaller sample to increase understanding of the phenomena (Malouf, Davidson & Sherman, 2006). But, using CGC in marketing research has been found to yield similar results to those obtained by surveys and interviews with marketing experts (Netzer et al., 2012), and with the advantages which can leverage changes in the traditional marketing activities (Onishi & Manchanda, 2012).

2.1. Analytics of Online Consumer Reviews and Marketing

Online discussions of the users based on their experiences with a brand in the offline environment reveal that the online personality of a brand is closely correlated with its actual market persona (Aggarwal et al., 2009). This issue shows that the online user-generated data provide a new opportunity to observe the market (Urban & Hauser, 2004). Several studies have focused on the efficacy of online customer reviews in various fields (Kim, Pantel, Chklovski, & Pennacchiotti, 2006). In the field of marketing, useful applications of online customer reviews have been studied in particular in recent years. Extensive research has been conducted on the impact of online CGC on sales rates (Netzer et al., 2012). For instance, Liu (2006) predicted box office sales by examining the volume and valence of the messages posted on the Yahoo! movies message board.

Also, in other industries besides online shopping, for example particularly in the lodging industry, several studies have confirmed the practicality of online customer reviews. Li, Ye & Law (2013) elicited the determinants of customer satisfaction in the lodging industry from the analysis and text mining of online hotel reviews. In another research, Sparks, Perkins, & Buckley, (2013) analyzed online travel reviews and observed that online consumer-generated reviews have more influence on the decision-making of travelers than the suggestions provided by travel and tourism agencies.

A number of recent studies have used new tools, such as text mining, to develop new techniques and apply them into marketing fields and particularly in constructing perceptual maps. For instance, Lee & Bradlow (2011) developed a text-mining algorithm for online product reviews, which was
intended to be comprehensive in capturing the product features based on the aggregation of the ‘pros and cons’ of a product. In the mentioned study, the researchers used manual reading to identify 39 distinct clusters of product attributes from the 99 clusters extracted by K-means. Furthermore, (Netzer et al., 2012) used the full text of reviews in order to identify and taxonomize the product features and proposed a hybrid text-mining and semantic network analysis tool for the surveillance of the market structure map. In addition, in the aforementioned study, market structure maps which were derived from the CGC, were compared with the market structure maps derived by traditional approaches. It was established that the text mining approach has the required external validity to reflect not only the opinions and views of forum members, but also the views of a wider population of consumers. Additionally, (Krawczyk & Xiang, 2015) used a text analysis approach to develop perceptual maps from the most frequently used terms in a dataset collected from an online travel agency, reporting that online consumer reviews could represent the level of differentiation between various hotel brands. Therefore, it was concluded that these reviews are an efficient source for understanding the market structure of the lodging industry.

Analysis of reviews is performed by several techniques, including classification, support vector machines, sentiment analysis based on NLP, and regression. Early studies regarding the analysis of online user reviews and marketing aimed to classify the entire documents as containing overall positive or negative polarity. In these systems, the overall attitude of the opinionist was explicitly indicated based on the approaches relying on manually-labeled samples (e.g., movies or product reviews). For instance, Liu, Cao, Lin, Huang, & Zhou (2007) solved the product reviews mining problem through a binary classification method. Therefore, they manually performed an opinion summarization by filtering the reviews of digital cameras on the Amazon website and were able to detect and distinguish only low-quality and high-quality reviews. However, the model proposed in the current research can derive a full range of product attributes from the reviews besides the minimal requirement of human intervention so that they could be easily used for marketing purposes.

In more recent studies, with the advancement of text mining techniques based on NLP, text analysis granularity has been taken down to sentence-level sentiments; such examples are the use of opinion-bearing lexical items to detect subjective sentences or by exploiting the association rule mining for a feature-based analysis of product reviews. As stated by several researchers, sentiments in product reviews have a direct impact on sales (Li & Wu, 2010; Liu, Yu, An, & Huang, 2012). In a study in this regard, Hu, Koh, & Reddy (2014) investigated the interrelationship between the sentiments in product reviews and sales rates. SA was performed on the reviews, followed by examining the effect of sentiments on the sales. According to the results, reviews with moderate sentiment were strongly associated with the sales rate of the product, as compared to the utterly positive/negative reviews. Another important finding in the study by Hu et al. (2014) was that the sentiments in the accessible recent reviews had a substantial impact on the sales rates of the products. Sentiment analysis (SA) approach have made an opportunity for researchers to be able to get structured and qualitative information on the brands in the market, based on the sentiment analysis elicited from large volumes of online texts.

SA is also known as opinion mining (OM), which refers to the use of NLP and text analysis to identify individual opinions, attitudes, and emotions towards different entities, including products, events and their attributes. SA techniques have been widely used for various purposes such as product recommendation (Acirar, Zhang, Simoff, & Debenhem, 2006), sales in the film industry (Duan, Gu, & Whinston, 2008), determining consumer dissatisfaction with online advertising campaigns (Qiu, He, Zhang, Shi, Bu, & Chen, 2010), prediction of the stock market (Bollen, Mao, & Zeng, 2011), prediction of political election results (Tumasjan, Sprenger, Sandner, & Welpe, 2010), product feature extraction and ranking (Zhang, Cheng, Liao, & Choudhary, 2012), mining newspapers and websites to extract the public opinion (Maragoudakis, Loukis, & Charalabidis, 2011), and identification of defects in cars (Abrahams, Jiao, Wang, & Fan, 2012). In the field of marketing, research on semantic orientation and adjective association suggests that SA techniques may also be an abundant source of
information on the online representations of the key dimensions of a commercial brand (Aggarwal et al., 2009).

Sentiment analysis (SA) offers two main approaches, aspect-based and non-aspect based (Liu, 2007). Aspect-based opinion-mining techniques divide input texts into aspects, which are also referred to as ‘features’. The aspect-based approach has been widely applied in developing various models. For instance, the authors of “Mining Reviews for Product Comparison and Recommendation” attempted to generate a product recommendation system by mining the product features from online reviews using the aspect-based SA. Moreover, Singh, Rana, & Alkhawaiter (2015) have used the Part of Speech (POS) Tagging technique in an aspect-based approach, in other to both finding out the sentiment expressed for each attribute of the product as well as a final review of the product. In another aspect-based approach research, the authors of “Designing Ranking Systems for Consumer Reviews: The Impact of Review Subjectivity on Product Sales and Review Quality” proposed ranking mechanisms to classify product reviews by combining econometric analysis with the SA techniques and subjectivity analysis.

In a study entitled “A Holistic Lexicon-Based Approach to Opinion Mining”, the researchers attempted to extract and summarize the consumers’ sentiments from the online product reviews on several electronic appliances. Product features were summarized using the text-mining and machine-learning methods, and the reviews were classified as positive and negative. According to Marrese-Taylor, Velásquez & Bravo-Marquez (2014), the mentioned study is the most reliable research in this area, which has inspired the present study as well.

Although several studies have demonstrated the efficacy of summary statistics for the analysis of online CGC in various marketing fields, they have also highlighted the challenges and limitations associated with the current models. Since the empirical use of the available models is not easy, developing other efficient methods for eliciting meaningful information from the content of online discussions has been shown to be of paramount importance in previous studies.

2.2. Product Attribute Extraction and Ranking

To construct a perceptual map, product attributes and consumer opinions must be extracted from the reviews. Attributes refer to the components or aspects of the product, including the properties, benefits, functions or applications (Liu, 2007). Aspects usually appear as nouns (e.g., battery), noun phrases (e.g., battery life) or nominal combinations (e.g., touch screen) in the text. Opinions are the attitudes or descriptions of the aspects, which normally appear as adjectives or adverbs in the text (Decker & Trusov, 2010; Popescu & Etzioni, 2007).

Opinions can be positive or negative and vary in strength. Based on the previously mentioned assumptions, the problems associated with the extraction of product attributes and opinions from review texts could be classified, as follows (Popescu & Etzioni, 2007):

1. Identifying the product attributes by deriving the most frequently used nouns in the text;
2. Identifying the opinions regarding product attributes by finding the correlating adjective; for instance, in the sentence “The size is too big.”, ‘too big’ as an adjective phrase represents the opinion on the product attribute of size, which is a noun.
3. Determining the polarity of opinions, which can be positive, negative or neutral;
4. Ranking opinions based on their strength; for instance, the adjective ‘horrible’ is a stronger indicator than the adjective ‘bad’.

In addition to the current research, Decker & Trusov (2010), Popescu & Etzioni (2007), and Eirinaki, Pisal, & Singh (2012) have used the aforementioned approach to extract opinions and aspects from product review texts. It is also notable that we have made a unique variation to the algorithm; instead of searching for nouns and finding the correlating adjectives, we searched for the adjectives that would match a correlating noun. This method works better because the algorithms which work with the nouns and the finding the correlating adjectives, might not always accurately working mainly
because of the redundant non-feature nouns which are considered mistakenly among the features when the engine tries to elicit the features only.

In the present study, the aspect finder algorithm was based on the high adjective count (HAC) algorithm, as proposed in feature-based opinion mining and ranking. The main idea behind the algorithm is that the most frequent nouns, for which the reviewers express their opinions, are most likely to be the important and distinguishing features than those for which users do not express such opinions.

In the current research, the HAC algorithm was improved by three different approaches. Initially, the HAC algorithm associated each adjective to its closest noun, which might not be effective since the aspect may not be in the vicinity of the expressed opinion in the text. We solved this problem by using grammatical patterns to extract the required nouns and adjectives. In the next stage, in order to determine the potential features of the proposed method, frequently used nouns were filtered, and the most frequent English nouns were removed. Afterwards, similar to the study entitled “A Holistic Lexicon-Based Approach To Opinion Mining,” the nouns with similar meaning were placed in one group, and the groups with the highest frequency were determined as the potential aspects. Details of the algorithm have been described in Section 3.

3. METHODOLOGY

According to Jobber (2001), there are four steps for creating a perceptual map: (1) Identify a set of competing brands. (2) Identify important attributes that consumers use when choosing between brands. (3) Conduct marketing research and determine customers’ opinion on the attributes. (4) Plot the brands on a two dimensional map.

The mentioned approach was used in the present study to generate the perceptual map. First, we determine the competing brands that we want to plot in the map. Then the product reviews related to these brands are crawled, downloaded and inserted into the review database. After preprocessing the data, an opinion mining engine determines the most important aspects to customers and extracts the related opinions. Similar to previous research studies, we assume that the most important attributes to customers are the ones that they mostly talk about. Therefore, we have determined the most frequently mentioned aspects in the text as the most important attributes to consumers, and then we determine the consumer opinions on these aspects. Thereafter, the extracted opinions are scored using a dictionary-based method and an opinion score is determined for each (brand, aspect) tuple. The final perceptual map is generated from this data. Figure 1 gives the architectural overview of our perceptual mapping system. The detail of the method is described in the following sections.

3.1. Preprocessing the Data

After crawling and downloading the reviews, the data should be preprocessed to be ready for the opinion mining engine. In this step, duplicate data and the reviews that contain missing fields are deleted from the review database. Then a spell correction algorithm searches the data for possible errors and corrects them. Thereafter, using the Stanford POS (parts-of-speech) tagger module (“The Stanford Natural Language Processing Group”, 2018), all of the words in reviews are tagged by their corresponding POS. For example, the sentence: “the camera quality is poor” would be tagged as “<the/DT><camera/NN><quality/NN ><is/VBZ><poor/JJ>”.

3.2. Opinion Mining Engine

The opinion mining engine performs the following tasks:

1. Identifying and extracting aspects that have been commented on in each review.
2. Grouping aspects with the same meaning, as different people may use different words to express the same feature.

3. Determining the most frequent aspects.

The core of this engine consists of the “Aspect Finder” algorithm. The Aspect Finder, which is shown in Figure 2 starts by identifying the adjectives and nouns in each document collection. First, each review is broken down into sentences. As stated in Section 2.2, in each sentence, the words, which have a POS tag of adjective, are determined. Adjectives are considered opinions if they are describing a noun, otherwise they are discarded.

When the algorithm finds an adjective (opinion), it should determine the corresponding noun phrase (aspect). To determine the corresponding noun phrase, we have determined grammatical patterns that are widely used in the English language. These patterns help determine, what noun the adjectives are describing. The patterns are shown in Table 1. If none of the introduced patterns are found in the sentence, then the closest noun to the adjective is considered as the target aspect.

After determining each tuple of (aspect, opinion), the algorithm searches for two more patterns in the sentence: negation words and adverbs before adjectives. There are a number of words such as “not”, “no”, “but”, and “however”, that give an opposite meaning to the sentence. When these words occur before an opinion word, they negate the opinion word. Thus, when we are searching the sentence for adjectives, we also took the remaining context into consideration and if an inversion word appeared, we preserved the negation word along with the opinion. Furthermore, in many English sentences, adverbs occur before the target adjectives and change its strength. For example, “very happy” is a stronger indicator than “happy”. Therefore, the Aspect Finder algorithm, searches for adverbs before the opinions, and saves them alongside the (aspect, opinion) tuple.
Table 1. Grammatical Patterns for Extracting Nouns

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun + vb + Adj</td>
<td>&lt;The/DT&gt;&lt;camera/NN&gt;&lt;is/VBZ&gt;&lt;unclear/JJ&gt;</td>
</tr>
<tr>
<td>Noun + vb + Adj, Adj … and Adj</td>
<td>&lt;The/DT&gt;&lt;camera/NN&gt;&lt;is/VBZ&gt;&lt;unclear/JJ&gt;, &lt;slow/JJ&gt; and &lt;cheap/JJ&gt;</td>
</tr>
<tr>
<td>Noun + vb + adv+ Adj</td>
<td>&lt;The/DT&gt;&lt;screen/NN&gt;&lt;is/VBZ&gt;&lt;very/RB&gt;&lt;bright/JJ&gt;</td>
</tr>
<tr>
<td>Adj+Noun</td>
<td>&lt;unclear/JJ&gt;&lt;camera/NN&gt;</td>
</tr>
<tr>
<td>Adj + Noun Noun</td>
<td>&lt;short/JJ&gt;&lt;battery/NN&gt;&lt;life/NN&gt;</td>
</tr>
<tr>
<td>Adj + Noun Noun … Noun</td>
<td>&lt;Separate/JJ&gt;&lt;micro/NN&gt;&lt;USB/NNP&gt;&lt;port/NN&gt;</td>
</tr>
<tr>
<td>Adj+Noun of Noun</td>
<td>&lt;low/JJ&gt;&lt;quality/NN&gt;&lt;of/IN&gt;&lt;screen/NN&gt;</td>
</tr>
<tr>
<td>Adj+Noun of the/this/that Noun</td>
<td>&lt;short/JJ&gt;&lt;life/NN&gt;&lt;of/IN&gt;&lt;of/DT&gt;&lt;battery/NN&gt;</td>
</tr>
<tr>
<td>Adj+Noun of the Noun … Noun</td>
<td>&lt;Recent/JJ&gt;&lt;version/NN&gt;&lt;of/IN&gt;&lt;the/DT&gt;&lt;DramaFever/NNP&gt;&lt;application/NN&gt;</td>
</tr>
<tr>
<td>Adj+Noun…Noun and Noun … Noun</td>
<td>&lt;bad/JJ&gt;&lt;camera/NN&gt;&lt;and/CC&gt;&lt;battery/NN&gt;&lt;life/NN&gt;</td>
</tr>
<tr>
<td>Adj+Noun,…, Noun and Noun</td>
<td>&lt;perfect/JJ&gt;&lt;size/NN&gt;, &lt;design/NN&gt;, &lt;color/NN&gt;&lt;and/CC&gt;&lt;thickness/NN&gt;</td>
</tr>
<tr>
<td>Adj and Adj Noun</td>
<td>&lt;unresponsive/JJ&gt;&lt;and/CC&gt;&lt;blurry/JJ&gt;&lt;screen/NN&gt;</td>
</tr>
<tr>
<td>Adj,Adj,…and Adj Noun</td>
<td>&lt;unresponsive/JJ&gt;, &lt;small/JJ&gt;&lt;and/CC&gt;&lt;unclear/JJ&gt;&lt;screen/NN&gt;</td>
</tr>
<tr>
<td>Adj,Adj…Adj +Noun</td>
<td>&lt;satisfactory/JJ&gt;&lt;beautiful/JJ&gt;&lt;professional/JJ&gt;&lt;lens/NN&gt;</td>
</tr>
</tbody>
</table>

(Table guideline: Adjective—JJ, Noun singular or mass—NN, Determiner—DT, Coordinating conjunction—CC, Preposition or subordinating conjunction—IN, Proper noun singular—NNP, Adverb—RB, Verb 3rd person singular present—VBZ)

Moreover, in any language, there are certain words that are frequently mentioned in any corpora. For example, the word “time” or “thing” are among the most frequently used English nouns. The Aspect Finder algorithm is based on counting the frequency of nouns. Therefore, it finds many of the frequent words that are not aspects of the product. In order to fix this problem, we used a database of most frequent nouns from the Corpus of Contemporary American English (wordfrequency.info). This corpus has reviewed documents containing about 450 million words and has determined the top 5,000 frequent words in English. We have only selected the 50 top most frequent nouns and have used them in the Aspect Finder to filter the non-aspect frequent nouns. For each noun that the algorithm finds, if it is in the Frequent_English_Nouns set, then it is discarded and no score will be saved for this noun.

As mentioned above, different people may use different words to express the same feature. For example, the resolution of the screen is referred to by terms such as resolution, screen resolution, display resolution, pixels, etc. These nouns have the same meaning and can be a substitute for one another. Therefore, before counting the frequency of each noun, we should group the nouns with the same meaning. Identifying synonyms is a challenging problem and is beyond the scope of this paper. To simplify the problem, we have used a dictionary-based approach. Similar to A holistic lexicon-based approach to opinion mining, the synonyms are determined using WordNet Synsets (Miller, 1995). However, WordNet does not contain all of the aspects of a tablet. Thus, we have filled this gap by adding the missing aspects of the tablet and their synonyms into this dictionary.

After processing all reviews in our collected data, the set of noun phrases, the correlating adjectives, negation, aspect group and adverbs are saved as <aspect, aspect group, opinion, negation,
adverb, review Id> sets in the Aspect-Opinion database. Then the algorithm assigns scores to each of the extracted noun phrases. This score, which is named noun_score in the algorithm, is equal to the frequency of the noun in the Aspect-Opinion database (Popescu & Etzioni, 2007; Hu & Liu, 2004). The noun scores are used to rank the nouns such that, the higher ranked nouns will be the ones having more adjectives describing them. According to the fact that each review belongs to one opinion holder, if a noun is repeatedly mentioned in a review, its frequency will only be increased by one. Therefore, the noun score shows how many opinion holders have mentioned that specific noun phrase in their review and have expressed an opinion on it. The potential aspects are identified by selecting the nouns, which have a score above a particular threshold. The threshold is an argument of the algorithm and can be chosen by experiment. The pseudo code of the Aspect Finder algorithm is shown in Figure 2.

### 3.3. Opinion Scoring

In this step, the Opinion Scoring algorithm calculates the sentiment score for each potential aspect of each brand. For example, given a brand such as Apple and a potential aspect, e.g. battery life, the Opinion Scoring algorithm finds all the Apple product reviews that have assigned an adjective (opinion) to battery life, and calculates the average sentiment score for all those adjectives as the final sentiment score for Apple’s battery life. The algorithm takes three inputs, which are described in the following paragraphs.

The first input to the algorithm is the MPQA subjectivity corpus. MPQA contains a database of words that are classified into positive, negative or neutral. Each word in this database is also tagged
as weak or strong. We calculate the MPQA score for each adjective in the range of [-2, 2]. The MPQA score is calculated using the following formula:

\[
\text{MPQA}(a) = \left( \text{positivity}_a \right) \left( \text{strength}_a \right)
\]

(1)

where \( a \) is the word from the review. The positivity parameter is set to 1 for a positive word, -1 for a negative word and 0 for a neutral word. \( \text{strength}_a \) is the strength of the word in the MPQA corpus. The strength parameter is set to two, whereas, if the word is a strong subject it is set to one.

The second input to the algorithm is the helpful score for the reviews. Amazon.com allows customers to tag a review as helpful or non-helpful. If a review is tagged non-helpful, even for multiple times, then it might be possibly not a valid word or with a strong opinion. Therefore, we can assume that this score represents the reliability of a review. The \( h \_ \text{score} \), which is calculated using Formula 2 demonstrates how helpful the review has been to customers.

\[
\text{h\_score}_r = \frac{\text{Number of users that have rated ras helpful}}{\text{Number of users that have rated ras helpful} + \text{Number of users that have rated ras non-helpful}}
\]

(2)

The third input of the algorithm is the list of opinion words, which are the adjectives used to express an aspect for each brand. For example, in the sentence “the camera is clear and fast”, if camera is a frequent noun, then the opinion words are clear and fast. For each of the adjectives in the opinion word list, the algorithm needs the negation and the adverb associated with the adjective, which are already in the Aspect-Opinion database. The final sentiment score for each brand aspect is calculated using Formula 3.

\[
\text{sentimen score}_{brand,aspect} = \sum_{k=1}^{n} \text{MPQA(adj}_k) \times \text{n\_score}_k \times \text{MPQA(adverb}_k) \times \text{h\_score}_k
\]

(3)

In this formula, “\( n \)” is the number of all adjectives describing the target brand aspect. MPQA (adj) and MPQA (adverb) are the scores of the adjective and its corresponding adverb, “n\_score” is the negativity score and “h\_score” is the helpful score for the review that the adjective belongs to. The pseudo code of the Opinion Ranking algorithm is shown in Figure 3. The output of the Opinion Ranking algorithm is a matrix that contains a score for each aspect of each brand. This matrix is what we need for constructing a brand perceptual map. The following section describes how the map is generated.

3.4. Constructing the Perceptual Map

A perceptual map is usually generated in a two-dimensional matrix. Marketers and scholars use this method to evaluate or determine the position of a certain brand. However, this method fails to compare more than one brand with each other. Yet, by using the clustering method, we are generating a single multi-dimensional perceptual map, which weighs the performance of many brands. To cluster the brands, the matrix of aspect-brand scores are entered into the K-means algorithm, the clusters are determined and the final perceptual map is generated.

The steps of constructing the perceptual map are implemented using the C# language and Microsoft SQL server. In the first step we programmed the software to crawl Amazon.com to download online reviews automatically. In the next step the software performs the preprocessing tasks on the downloaded reviews, applies the Aspect Finder and Opinion Ranking algorithms to the data and outputs the final matrix of aspect scores in brands. The software’s performance is evaluated in the following section.
4. EXPERIMENTAL RESULTS

To evaluate the performance of the software, we applied it to the online reviews downloaded from Amazon.com for tablet products. We chose Amazon.com because among the online retailing platforms, it has the most number of comments of customers. Also, the choice of tablet product reviews was because tablets are a very commonly researched product and the aspects that consumers care about are determined in many previous research papers (Olmsted and Shay 2016, Huang et al. 2011); and therefore, we can easily prove that the proposed algorithm has achieved correct results. The following sections describe the results in detail.

4.1. Determining the Competing Brands

In the first step, major competitors in the category of tablet markets were determined manually based on their market shares (“Statista: Tablet market share by vendor,” 2014). The brands are Apple, Microsoft, Acer, Asus, Neutab, Dell, HP and Samsung. Thereafter, we crawled and downloaded all the consumer online reviews of the tablets with one of the aforementioned brands. 5230 reviews were downloaded on December 2014 and were inserted into the review database. After preprocessing the reviews, 5145 reviews were left usable. Table 2 shows the number of reviews downloaded for each brand.

4.2. Finding the Most Important Aspects

To determine the most important aspects to the customers, the data from the previous step was put to the Aspect Finder. As stated in the algorithm, it was first necessary to look for the adjectives, which have a correlating noun. 38665 sets of <noun, adjective, negation, adverb, reviewId> were detected in the reviews. In the second step, the algorithm groups the nouns that have similar meanings and then counts the frequency of each group to determine the potential aspects. In order to determine the most frequent nouns, according to previous studies (Eirinaki et al., 2012) we set the threshold to 0.5

Table 2. Number of downloaded reviews for each brand

<table>
<thead>
<tr>
<th>Brand Name</th>
<th>Acer</th>
<th>Apple</th>
<th>Asus</th>
<th>Dell</th>
<th>HP</th>
<th>Microsoft</th>
<th>Neutab</th>
<th>Samsung</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of reviews</td>
<td>166</td>
<td>368</td>
<td>1303</td>
<td>110</td>
<td>103</td>
<td>254</td>
<td>196</td>
<td>2645</td>
</tr>
</tbody>
</table>
percent. This means that if a noun is repeated more than 193 (0.005*38665) times, then the noun is considered to be a potential aspect. The resulting potential aspects are shown in Table 3. From the ten potential aspects, only the first one (tablet) is not an aspect of a tablet. Although, opinions on this word can provide an insight about consumer perceptions towards the tablet as a whole, we eliminated this word because it is not an aspect for constructing the perceptual map.

Price, size, battery life, resolution, camera, processor and design are obvious features/benefits of a tablet. The screen aspect, describes the screen of the tablet. It is different from screen resolution. The screen resolution specifically reviews the resolution of the screen while most of the reviews describing the screen aspect evaluate other attributes of the screen. For example, “The screen is quite bright, colorful and contrasty. NO, it’s not a retina screen, but it’s clear and crisp” this shows that the user cares about the clarity and color of the screen.

Moreover, the “application” aspect is mostly about the available applications on the tablet. For example, “With a wide range of available applications from the App Store and also through my company, I have everything I need to stay productive” and “The Android Market doesn’t distinguish between phone applications designed for a small screen and tablet applications” this shows that the users significantly care about the possibility of installing and using applications on the tablet.

Finally, the “review” aspect is about the previous online product reviews. When consumers are buying online, they usually read many of the available online reviews. This study shows that the previous reviews are a determinant factor for consumers while buying tablets online. For example, in the following review the user has clearly mentioned that the previous reviews have affected his/her judgment about the product: “reading the bad reviews about the quality issues and service problems for the TF701t make me think that ASUS is going somehow downhill”. The following review also proves that previous reviews are very important to customers, especially when they do not have enough time to compare various products.

“I originally thought I’d get another Nexus because I’ve been really pleased with it. But then I remembered it’s 20 days until Christmas and I have 6 children to buy gifts for and looked for something that had good reviews, was a brand that I recognized and was in my price range.”

4.2.1. Evaluation of Algorithm

To evaluate the algorithm, we compare the performance of this algorithm to three other algorithms in the area of extracting frequent aspects from reviews. The three algorithms are High Adjective Count (HAC) (Eirinaki et al., 2012), the well-known TF, and TF*IDF. We perform this task the same as Feature-based opinion mining and ranking. We compare the precision of the top-N results given by the four algorithms for various values of N. Precision in this context is defined as follows:

\[ Precision = \frac{\text{number of relevant aspects}}{N} \]  

(4)

Table 3. Potential aspects extracted from online product reviews

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Frequency</th>
<th>Aspect</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tablet</td>
<td>1974</td>
<td>Battery Life</td>
<td>294</td>
</tr>
<tr>
<td>Screen</td>
<td>779</td>
<td>Resolution</td>
<td>278</td>
</tr>
<tr>
<td>Application</td>
<td>554</td>
<td>Camera</td>
<td>247</td>
</tr>
<tr>
<td>Price</td>
<td>364</td>
<td>Processor</td>
<td>246</td>
</tr>
<tr>
<td>Review</td>
<td>309</td>
<td>Design</td>
<td>210</td>
</tr>
<tr>
<td>Size</td>
<td>299</td>
<td>Total</td>
<td>5554</td>
</tr>
</tbody>
</table>
An example of the comparison of the top-10 features for the review dataset between the four algorithms is as follows:

Table 4 shows the comparative results between some of the benchmark competing algorithms developed in existing literature. As can be seen, the proposed method has performed better and represents higher accuracy compared to the similar algorithms previously used in this regard. In the potential aspects found by TF*IDF, which has performed better than the other two, iPad and Samsung are brand names, tablet and device point to the product itself and time is not an aspect of a tablet. It is clear that the potential aspects found by the Aspect Finder are better suited as aspects of a tablet. As it is shown, the accuracy of Aspect Finder algorithm is a higher than the other algorithms. Moreover, when we lower the threshold and compare algorithms in the top 20 nouns, the precision of Aspect Finder is much better than the other algorithms.

4.3. Determining the Opinion Score for Each Frequent Aspect

In this step, the Opinion Scoring algorithm assigns scores to each aspect of the eight selected brands. The final score of each brand in each aspect is shown in Table 5. As is shown in the Table, the maximum average score in all the selected aspects belongs to the Apple brand and it has achieved the highest scores in the aspects of “screen”, “design”, and “processor”. However, Apple has scored very low in the camera aspect. This shows that customers love the Apple brand but they are not much satisfied with the camera. Many negative adjectives such as “insufficient”, “not so great” and “disappointing” have been used to describe iPad’s camera. The following reviews clearly depict this result:

Table 4. Comparing the performance of TF, TF*IDF and HAC algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Top 10 extracted aspects</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect Finder</td>
<td>Tablet, Screen, Application, Price, Review, Size, Display Resolution, Battery life, Camera, Processor, Design</td>
<td>9/10</td>
</tr>
<tr>
<td>HAC</td>
<td>Tablet, Thing, Application, Time, Screen, Product, Device, Price, Year, Review</td>
<td>4/10</td>
</tr>
<tr>
<td>TF</td>
<td>Tablet, Screen, iPad, Application, Device, Android, Time, Samsung, Battery, Thing</td>
<td>4/10</td>
</tr>
<tr>
<td>TF*IDF</td>
<td>Screen, iPad, Application, Device, Battery, Android, Time, Samsung, Thing, keyboard</td>
<td>5/10</td>
</tr>
<tr>
<td>Common Words</td>
<td>Screen, Application</td>
<td>2/10</td>
</tr>
</tbody>
</table>

Table 5. Tablet brand scores in each aspect

<table>
<thead>
<tr>
<th>Brand</th>
<th>Battery life</th>
<th>Size</th>
<th>Processor</th>
<th>Application</th>
<th>Screen</th>
<th>Resolution</th>
<th>Price</th>
<th>Camera</th>
<th>Design</th>
<th>Review</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acer</td>
<td>0.75</td>
<td>0.78</td>
<td>0.94</td>
<td>0.54</td>
<td>0.6</td>
<td>0.46</td>
<td>1.05</td>
<td>1.23</td>
<td>1.05</td>
<td>-0.13</td>
<td>0.73</td>
</tr>
<tr>
<td>Apple</td>
<td>0.95</td>
<td>0.75</td>
<td>1.36</td>
<td>0.77</td>
<td>1.1</td>
<td>0.68</td>
<td>0.82</td>
<td>-0.03</td>
<td>1.35</td>
<td>0.1</td>
<td>0.78</td>
</tr>
<tr>
<td>Asus</td>
<td>1.03</td>
<td>0.65</td>
<td>0.79</td>
<td>0.26</td>
<td>0.93</td>
<td>0.58</td>
<td>0.53</td>
<td>0.98</td>
<td>0.9</td>
<td>0.25</td>
<td>0.69</td>
</tr>
<tr>
<td>Dell</td>
<td>0.28</td>
<td>0.63</td>
<td>0.3</td>
<td>1.04</td>
<td>0.77</td>
<td>0.42</td>
<td>0.37</td>
<td>0.99</td>
<td>1</td>
<td>-0.06</td>
<td>0.57</td>
</tr>
<tr>
<td>HP</td>
<td>0.17</td>
<td>0.12</td>
<td>0.55</td>
<td>1.05</td>
<td>0.58</td>
<td>0.02</td>
<td>0.31</td>
<td>0.16</td>
<td>1.3</td>
<td>0.97</td>
<td>0.52</td>
</tr>
<tr>
<td>Microsoft</td>
<td>0.72</td>
<td>0.48</td>
<td>1</td>
<td>0.62</td>
<td>0.9</td>
<td>0.76</td>
<td>0.49</td>
<td>0.67</td>
<td>0.75</td>
<td>-0.38</td>
<td>0.60</td>
</tr>
<tr>
<td>Neutab</td>
<td>0.4</td>
<td>-0.08</td>
<td>1</td>
<td>0.84</td>
<td>-0.12</td>
<td>0.14</td>
<td>0.66</td>
<td>-0.59</td>
<td>1</td>
<td>0.009</td>
<td>0.33</td>
</tr>
<tr>
<td>Samsung</td>
<td>0.95</td>
<td>0.86</td>
<td>0.73</td>
<td>0.7</td>
<td>0.95</td>
<td>0.66</td>
<td>1.04</td>
<td>0.62</td>
<td>1.14</td>
<td>0.2</td>
<td>0.78</td>
</tr>
</tbody>
</table>
The rear facing camera is pretty low quality. As a camera intended to be used for actually taking pictures, it’s pretty inexcusable.

The quality of these cameras are nowhere near the standards that one expects to see from a piece of technology, especially at the price that Apple charges.

Following Apple, the second top average score belongs to Samsung tablets. Samsung Galaxy Tabs are among the most ‘wished’ for tablets on Amazon.com. This brand has scored first in the “size”, second in “price” and “screen”. The lowest score for Samsung is in the “review” aspect. Although customers love it, many negative reviews have been written about this product. However, some consumers have stated that the negative reviews on Samsung products are bogus. For example, one user said: “Product arrived on time and in perfect condition. Need I say more. All the previous reviews are bogus”. However, classifying reviews as bogus or truthful is outside the scope of this study.

4.4. Constructing the Perceptual Map

The final step of the perceptual map is the part when the maps are generated. We can easily draw two-dimensional maps for each 2-combination of the aspects and show the brands in that map. For example, Figure 4 and Figure 5 are perceptual maps for price-design and camera-review combinations.

As mentioned in section 3.3, we used a K-means clustering method to put the brands in clusters based on all the important aspect scores. In order to determine the best number of K for the clustering task, we used the Silhouette Index. Silhouette Index shows how well each object (brand) lies within its cluster. The best K was determined as 3 and the clustering was performed using Matlab 2010. The results are shown in Table 6.

The first cluster consists of Apple, Samsung and Acer brands. The average sentiment score for all the aspects of these three brands is higher than other clusters. Therefore, we can say that this cluster contains the brands that have the best perception amongst the customers. In order to evaluate the clusters, we have used two lists from Amazon.com: the “mostly bought tablets” and the list of “most wished for” tablets. Each of the two lists contains a hundred tablet models. We have counted the number of tablets in these two lists, which have a brand the same as the brands in the three clusters. As can be seen in Table 6, percentage of the mostly bought and mostly wished for tablets complied with the clustering method. The brands in the first cluster, which has the highest score, has been purchased more than other clusters, and a higher percentage of them are among the most wished for...
Table 6. Results of brands clustering

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Brands in Cluster</th>
<th>Average of aspect sentiment</th>
<th>Number of Most Bought</th>
<th>Percent of Most Bought</th>
<th>Number of Most Wished for</th>
<th>Percent of Most Wished for</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Apple, Samsung, Acer</td>
<td>0.76</td>
<td>35</td>
<td>0.35</td>
<td>35</td>
<td>0.38</td>
</tr>
<tr>
<td>2</td>
<td>Asus, Microsoft, Dell</td>
<td>0.62</td>
<td>7</td>
<td>0.7</td>
<td>8</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>HP, Neutab</td>
<td>0.42</td>
<td>4</td>
<td>0.4</td>
<td>5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

tables. This finding indicated that the proposed technique could effectively construct the perceptual map and accurately hallmark the position of the brand in the perceptual map.

5. CONCLUSION AND DISCUSSION

The availability of online product reviews on the web has provided new opportunities for marketers to become aware of customers’ feedback and perceptions without time-consuming and costly surveys. The process of extracting the consumers’ opinion from text data is called opinion mining, which is currently one of the most challenging research areas in marketing. Opinion mining is concerned with analyzing the opinions of users on a particular subject, as expressed in their natural language. In the current research, we proposed an innovative approach based on opinion mining techniques so as to construct a perceptual map for various brands using online product reviews. The process consisted of four main stages, including the selection and competition of the brands, mining reviews to discover the important aspects of the product to customers using the aspect finder algorithm. As well as determining customer sentiment scores for each aspect of the company brands using the opinion ranking algorithm, and constructing the perceptual map by clustering the brands based on the aspect scores.

The algorithms were applied to the data of the tablet reviews, which were downloaded from Amazon.com. According to the aspect finder algorithm, the important aspects of the products were screen, available applications, market price, quality of the reviews, size, display resolution, battery
life, camera quality, processor, and design. The findings of the study suggested that the proposed technique was highly efficient in hallmarking the most important aspects of various tablet brands. Furthermore, a comparison of the clustering results with the data of the most desired and most purchased products on Amazon.com revealed that the final perceptual model, which was constructed by the applied method, complied with the real opinions of the customers.

In today’s social media context and ever-growing online channels of communication, this method can effectively help practitioners to construct brand perceptual maps based on the data which has a greater richness and more depth, resulting in more accuracy in identifying the positions of the brands. The approach of combining text mining and perceptual mapping to generate a brand positioning map using product reviews has been used in many studies, but what this research adds to theory is that the unique proposed algorithm is superior to some benchmark models because of being more automated and having improved accuracy.

5.1. Managerial Implications

The proposed framework of eliciting a market structure map based on the analysis of online CGC is a powerful marketing tool which could be practically used by marketers and managers in their branding and marketing strategies development. Marketers can take advantage of the architecture presented in Figure 1 to see the whole process of constructing a perceptual map from the reviews data. They can afterwards use the text mining programming algorithm presented in Figure 2 and then with using the formulas number 1,2 and 3 in the programming algorithm of the Figure 3 a matrix is generated that contains a score for each aspect of each brand. This matrix is what we need for constructing a brand perceptual map.

This method helps the manager’s marketing strategies at various levels. Firstly, this method helps the better creation of campaign’s message strategy. The qualitative data elicited from the contents can be considered as a source of customer insight, as a result a company can set up its message strategy based on them in a more relevant way. Hence, using this method allows the managers to better recognize the favorable brand associations to be emphasized on, in the message strategy of the later communication efforts. For instance, in a campaign of new product introduction, managers and brand marketers can potentially take advantage of the reviews to rapidly obtain customer feedback in order to find out whether the predefined message strategy derived the brand image and/or differentiates the brand’s position in the minds of the consumers. Moreover, with the richer, deeper and wider input data which this method gives to the marketers to construct the perceptual maps accordingly, the method helps them to more accurately identify preemptive spaces and unmet positions in the market, and more importantly it is helpful in recognizing the opportunities of the market which may be used in developing a niche marketing strategy.

A comparison of the traditional perceptual maps and those obtained through the proposed method indicates several advantages in the mentioned method. One of these strengths is that the proposed approach constructs the perceptual maps based on the data which has considerable more breadth. Namely, the data comes from a significant number of the reviewers, reviewed brands, and extracted topics from the reviews. Furthermore, along with the automation of the derivation of the market structure, the perceptual maps produced in this method, are similar to those available in the current literature and commonly used by marketing researchers and brand managers. Therefore, marketing practitioners are already familiar with interpretations of these positioning maps and the decision-making process based on the maps.

Another prominent feature of the proposed method is its unique advantage over the traditional positioning maps. The traditional positioning maps are based on the quantitative data that are linear and do not have the required depth. In these maps, researchers determine a priority of the features or attributes that define the dimensions of the map and ask a sample population of consumers to rate the competing brands on a predefined list of product features (Aggarwal et al., 2009). As such, positioning mapping based on consumer surveys urge customers to view the competing brands through
the lenses defined by the researcher and to rate these brands on a predefined scale, which results in researcher/manager bias (Moon, 2017). This problem could be solved since as opposed to a structured questionnaire, online reviews represent an outlet for consumers to express their opinions and reflections on a certain product/service experience in a free and unstructured manner. These natural conversations in the customers’ own environments help them articulate their deepest anxieties and desires. This is because customers are more comfortable and open to tell their fellow customers what they think and do (Kotler, Kartajaya & Setiawan, 2017). Therefore, using the opportunity of listening to these social voices enables researchers to tap into the factors that are indirectly important to the customers.

By using these new technologies, marketers will find that the information of the customers’ reviews can yield valuable input for their marketing planning. For instance, marketers can benefit from a huge source of free text data to discern the time-relative trends affecting the preferences of the customers. Online reviews are time-specific, in that they reflect the perceptions of the current product at the time when the reviews are constructed. Marketing strategies revolving around building an image that is comparative to another specific brand can be measured over time by updating the perceptual maps. As new reviews are posted and through the post-purchasing experiences of the consumers, the company finds the optimal product repositioning clues and afterwards the direction of product redesign.

Practitioners may be able to use the method outlined in the present study as a cost-effective more accurate tool to be applied regularly to contribute to a highly dynamic analysis of the brand image, which has been infeasible in the past with the costly traditional forms of market research. In general, the proposed approach can properly address the managerial questions regarding the most favorable established associations thus far, changes in the brand image over time, and whether the marketing actions strengthen or weaken these changes.

5.2. Limitations and Further Work

Despite all the advantages of the proposed approach, it has limitations as well. For instance, we used basic sentiment analysis techniques based on the MPQA dictionary. Therefore, further research in this regard could be improved by incorporating machine learning techniques. On the other hand, it is recommended that the validation method be generalized to other domains as well (e.g., services and tourism). Also, our model is able to elicit perceptual map for different brands when reviews are depicted in only one sample, and has limitation on comparing brands from two or more sample with-together. As a research idea, we think that combining reviews data and data of other purchases of the customers, along with the customers’ demographic information (that a company might be able to obtain them from the confirmed users of the websites) this could offer novel customer insights into targeted marketing and segmentation strategies. Furthermore, the proposed framework could be extended and adapted to different languages and cultures based on the origin of grammatical rules and cultural interactions.

Our framework can automatically process a large amount of product reviews information efficiently and elicit meaningful results for positioning purposes. However, it requires a significant amount of time in data collection and taxonomy building. Therefore, a future research might attempt to find a method to estimate a minimal number of required reviews to ensure sufficient heterogeneity in the sample selection of online consumer reviews with respect to different product features and the corresponding feature attributes.

Another limitation of the research was the presumption that all the available online reviews were truthful, while as we know, there are numerous false reviews on consumer forums such as Amazon.com. Thus, further research might develop models that are able to discriminate between “valuable information” and fake reviews or extreme reviews which are “error” in the data. This would be useful in having a less noisy data set and having a more accurate result, one could extend our text mining
method on the context of more formal product reviews such as news articles about the product, or blogs and social network pages that professionally review the products. As the language used in these contexts is more organized and tends to follow grammatical standards and rules. In summary, discovering how to deal with detecting and eliminating invalid reviews, could enhance the reliability of the results in further investigations in this regard.
REFERENCES


Amir Ekhlassi is an Assistant Professor at University of Tehran, and Adjunct Professor of marketing in Sharif University of Technology. He has been awarded as a named Branding and Marketing Professor. He has published more than 40 articles and 5 books, including his latest, “Principles of Branding: IMC approach”. He received a PhD from University of Tehran and an MBA from Sharif University of Technology.

Amirhosein Zahedi studies MSc in IT-based entrepreneurship at the University of Tehran. He is a professional guest speaker in marketing related conferences. His research field includes social media marketing, text mining applications in marketing and also gamification related topics.