ABSTRACT

EXTRACTING PRODUCT DEVELOPMENT INTELLIGENCE FROM WEB REVIEWS

by
Ismail Artun Yagci

Product development managers are constantly challenged to learn what the consumer product experience really is, and to learn specifically how the product is performing in the field. Traditionally, they have utilized methods such as prototype testing, customer quality monitoring instruments, field testing methods with sample customers, and independent assessment companies. These methods are limited in that (i) the number of customer evaluations is small, and (ii) the methods are driven by a restrictive structured format. Today the web has created a new source of product intelligence; these are unsolicited reviews from actual product users that are posted across hundreds of websites. The basic hypothesis of this research is that web reviews contain significant amount of information that is of value to the product design community. This research developed the DFOC (Design – Feature – Opinion – Cause Relationship) method for integrating the evaluation of unstructured web reviews into the structured product design process. The key data element in this research is a Web review and its associated opinion polarity (positive, negative, or neutral). Hundreds of Web reviews are collected to form a review database representing a population of customers. The DFOC method (a) identifies a set of design features that are of interest to the product design community, (b) mines the Web review database to identify which features are of significance to customer evaluations, (c)

extracts and estimates the sentiment or opinion of the set of significant features, and (d) identifies the likely cause of the customer opinion.

To support the DFOC method we develop an association rule based opinion mining procedure for capturing and extracting noun-verb-adjective relationships in the Web review database. This procedure exploits existing opinion mining methods to deconstruct the Web reviews and capture feature-opinion pair polarity. A Design Level Information Quality (DLIQ) measure which evaluates three components (a) Content (b) Complexity and (c) Relevancy is introduced. DLIQ is indicative of the content, complexity and relevancy of the design contextual information that can be extracted from an analysis of Web reviews for a given product. Application of this measure confirms the hypothesis that significant levels of quality design information can be efficiently extracted from Web reviews for a wide variety of product types. Application of the DFOC method and the DLIQ measure to a wide variety of product classes (electronic, automobile, service domain) is demonstrated. Specifically Web review databases for ten products/services are created from real data. Validation occurs by analyzing and presenting the extracted product design information. Examples of extracted features and feature-cause associations for negative polarity opinions are shown along with the observed significance.

EXTRACTING PRODUCT DEVELOPMENT INTELLIGENCE FROM WEB REVIEWS

by
Ismail Artun Yagci

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Industrial Engineering

Department of Mechanical and Industrial Engineering

January 2014

APPROVAL PAGE

EXTRACTING PRODUCT DEVELOPMENT INTELLIGENCE FROM WEB REVIEWS

Ismail Artun Yagci

Dr. Sanchoy K. Das, Dissertation Advisor | Date
Professor of Mechanical and Industrial Engineering, NJIT

Dr. Athanassios Bladikas, Committee Member | Date
Associate Professor of Industrial Manufacturing Engineering, NJIT

Dr. Katia Passerini, Committee Member | Date
Professor of School of Management, NJIT

Dr. Shanthi Gopalakrishnan, Committee Member | Date
Professor of School of Management, NJIT

Dr. Wenbo Cai, Committee Member | Date
Assistant Professor of Mechanical and Industrial Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Ismail Artun Yagci

Degree: Doctor of Philosophy

Date: January 2014

Undergraduate and Graduate Education:

- Doctor of Philosophy in Industrial Engineering,
  New Jersey Institute of Technology, Newark, NJ, USA, 2014

- Master of Business Administration,
  University of Delaware, Newark, DE, USA, 2005

- Master of Science in Hospitality Information Management,
  University of Delaware, Newark, DE, USA, 2004

- Bachelor of Science in Industrial Engineering,
  Istanbul Technical University, Istanbul, Turkey, 1987

Major: Industrial Engineering

Presentations and Publications:

Qi, T., Yagci, A., and Song, M., "Extraction of Key-phrases from Biomedical Full-text with Supervised Learning Techniques," 15th Americas Conference on Information Systems (AMCIS'09), San Francisco, California, August 2009.

I dedicate my Ph.D dissertation work to my family. A special feeling of gratitude to my loving parents, Dr. H. Cahit Yagci, Servet F. Yagci, whose words of encouragement and push for tenacity ring in my ears. My sister, Berna F. Yagci, and my brother, Fethi E. Yagci, my niece, Melisa Avci, and nephew, Berk C. Yagci are very special.

I also dedicate this dissertation and give special thanks to my soul mate, Lily H. Kim, for being there for me throughout the entire doctorate program.

ACKNOWLEDGMENT

The writing of this dissertation has been one of the most significant academic challenges I have ever had to face. Without the support, patience and guidance of my advisor and my dissertation committee members, this study would not have been completed. It is to them that I owe my deepest gratitude.

Foremost, I would like to express my sincere gratitude to my advisor Dr. Sanchoy K. Das, for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank my dissertation committee members: Dr. Athanassios Bladikas, Dr. Katia Passerini, Dr. Shanthi Gopalakrishnan, and Dr. Wenbo Cai, for their encouragement, and insightful comments.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF TABLES

LIST OF TABLES
(Continued)

LIST OF TABLES
(Continued)

LIST OF TABLES
(Continued)

LIST OF TABLES
(Continued)

CHAPTER 1

INTRODUCTION

The disruptive effects of the Internet are progressively affecting all industries, as traditional ways of doing business are being challenged. Product development processes are typically entrenched in engineering design and success has been sought by following traditional methods that depend on experience and deep product knowledge. The belief is that Web reviews are disrupting the product designer's domain by introducing large volumes of unstructured consumer experience reviews into the product knowledge space. Frequently, this knowledge may contradict or be even unknown to the product designer community. This research shows how product designers can utilize opinion mining methods to embrace and exploit this new knowledge. A special application of data mining, opinion mining involves the analysis of customer opinions using product reviews and provides meaningful information including the polarity of the opinions (Jeong, Shin et al. 2011). Available advances in data mining technology make it an amenable research platform for the analysis of Web reviews.

1.1 Research Motivation

In today's business environment, keeping up with fast-paced technological changes and timely response to consumer needs becomes the top priority for businesses to stay competitive in the marketplace. The new product development, whether it is to develop a brand new item or modify an existing product, has become a critical aspect in business and engineering. New product or service development refers to a complete process which begins with idea generation, manufacturing (or service line extension) through product design, and concludes by introducing a new product to the consumers in the marketplace (Li and Moon 2009; Durmuşoğlu and Barczak 2011; Mohabbati, Hatala et al. 2011). Effectively managing the idea generation process for development is one of the most important, however difficult challenges facing product development managers (Aken and Nagel 2004; Van Kleef, Van Trijp et al. 2005).

There are two avenues regarding idea generation, and realization of business or technological opportunities; one involves internal formal R&D process; the other involves market research. However, regardless of where opportunities originate, the consumer is the ultimately the one who makes the final judgment to determine which products become successful. Therefore, understanding consumer opinions and needs, especially from the early stages of the new product development process, provides product development managers the advantage to be able to focus on the product with the highest probability for success in the first instance (Van Kleef, Van Trijp et al. 2005; Mudambi and Schuff 2010). Major companies have realized that effective product development begins with understanding of consumer preferences and needs (Voice of the Customer), and understanding of how customers experience the product. If companies truly intend to listen to their customers, more likely they will receive valuable insights that can drive new products.

In a traditional setting, consumer requirements have been collected through different channels: prototype testing, market survey instruments, field testing methods, and hiring independent assessment companies, which are custom ways for business to determine consumer needs or expectations about products, services, and market dynamics such as those of competitors (Park and Lee 2011). These methods work well to obtain inputs and feedback from the potential consumers; however, they can be costly and time consuming.

Time has changed favorably to businesses with the Web 2.0 platforms such as user-generated content (e.g., discussion forums, consumer review websites, blogs, and various other types of social media) have become tremendously popular (Yang, Wei et al. 2009; Galatis 2011). With growing popularity of social media, online consumer reviews (Web reviews) are an increasingly important part in consumer purchase decisions (Mudambi and Schuff 2010; Galatis 2011). A recent survey revealed that Web reviews are the second most influential ways to affect purchasing behavior after word of mouth (MarketingChart 2008). Another study showed that 86% of public finds customer reviews extremely or very important. There are a series of reports that explained the popularity of Web reviews, one of which is 64% of the individuals viewed the products online and read Web review regardless of where they purchased the products (Ganu, Marian et al. 2010).

Today, hundreds of consumer reviews are available online for a wide range of products and services on several commercial websites such as Amazon.com, KBB.com, DPreview.com, IMDB.com, Cnet.com, ZDnet.com, bizrate.com, ConsumerReview.com, Epinions.com, and RollingStone.com. These websites allow consumers to share their experiences about products/services where reviews were in numerical ratings and/or open-ended free text about the experience on product/services whether they are negative or positive. Just like individuals take advantage of publicly available online reviews, businesses do the same and even further utilize them as a valuable source of information (Dellarocas 2003; Dellarocas 2006). The fact that online consumer reviews are easy for the public to access, as a download, makes it even more beneficial for business to grasp consumer needs or opinions in the early stage of new product development.

With these freely available user-generated contents, the traditional product development methods have been changing in terms of how a business manages consumer expectations, brand positioning, new product development, and other activities accordingly. This trend highlights the importance of Web reviews to businesses in the area of new product development processes as well as defining effective marketing strategies. However, the vast amount of information and its widespread disemination make it challenging for product development managers to find all relevant information, read them, summarize them, and organize them into a usable format for its competitive advantage or drive business critical information about future opportunities and risks (Hu and Liu 2004; Sandhu and Mehta 2011).

1.2 Anatomy of a Web Review

The key data element in this research is a Web review. Web reviews are defined as peer-generated online customer reviews typically recorded on third-party websites (Mudambi and Schuff 2010). Web reviews are not authenticated or validated; that is, there is no guarantee the author of the reviewer is writing an honest review or has even experienced the product or service in question. However, the sheer volume of reviews in the Web, makes them an accepted standard. Structurally, Web reviews are short with length ranging from a single sentence to four or five sentences. Reviews in excess of eight sentences are uncommon and only seen in certain product/service groups.

Figure 1.1 is a description of the anatomy of a Web review in the context of this research and the underlying opinion mining methods applied. These methods deconstruct the Web reviews into sentences, nouns, verbs, and adjectives. The mining strategy then is to identify relationships between these. The overlying research method, as developed here, controls how these relationships are defined and then converts them into research results and conclusions. As shown in Figure 1.1 the example Web review is separated into three sentences, two opinions, three design features, and two causes. To identify and aggregate these elements, an opinion mining method will use one or more lexical resources, for example, Princeton developed WordNet, which is a large lexical database of English. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. This database is then used to determine automatically whether a term that is an expression of opinion has a positive or a negative connotation.

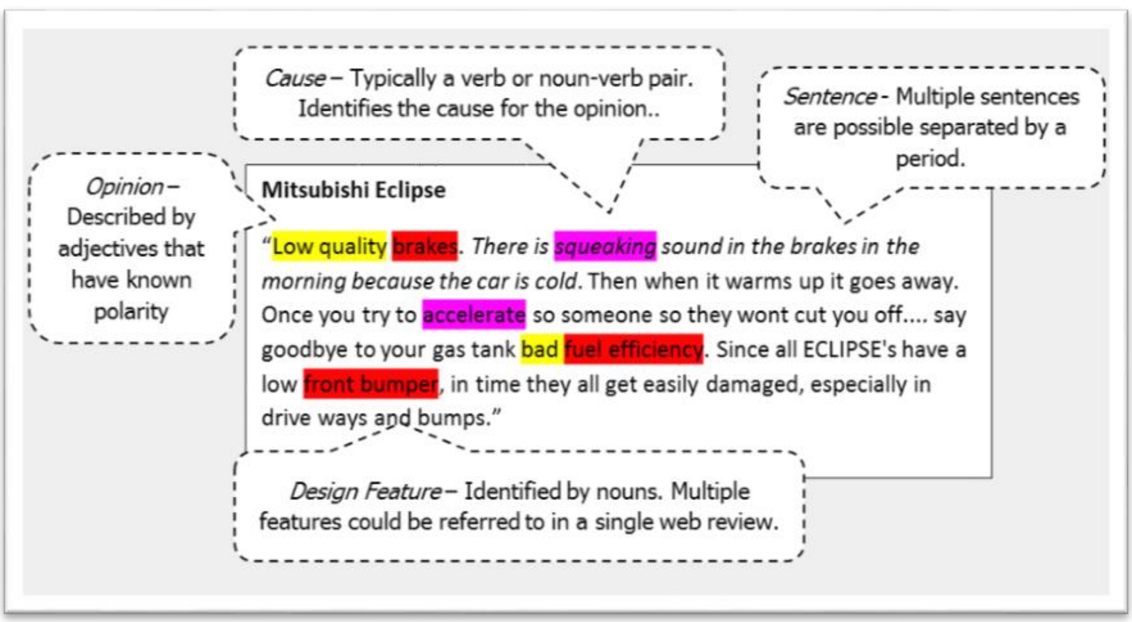


Figure 1.1 Anatomy of a Web review.

1.3 Research Problem Statement

In recent years, the volume of online Web reviews has been growing at a dramatic rate and has become an important data source in the individual decision making process. Presently, these reviews are primarily used by potential buyers to learn the product experience of other consumers, which has significant effects on buying decisions. A second growing use is in marketing where reviews are analyzed to project product sentiment (positive, negative, or neutral). In spite of growing popularity on utilization of online reviews, there is limited research concerning extracting specific product intelligence that can then be used to develop better products that consumers actually want. For example, one may learn from the reviews that the image stabilization feature in a digital camera malfunctions in humid conditions. Frequently, such evaluations are occurring in a small percentage of the user base and are not detected in traditional methods.

The conventional wisdom is that Web reviews are only of consumer interest since they only consist of consumer sentiment or opinion, hence, are used only to influence consumer decisions and develop marketing strategy. There is a need for models and methods to extract the design intelligence from Web reviews such that the extracted knowledge is relevant to the product designers and maps to their portfolio of key product features. Further, these methods must quantify the level of available information to justify any effort that a design group allocates to such analysis. The ideal goal would be to identify specific design features discussed in the Web reviews, evaluate the opinion polarity of these features, and identify the cause of the reflected opinion.

Product development managers are continually challenged to learn what the consumer product experience really is and to learn specifically how the product is performing in the field. Traditionally, they have utilized a variety of methods including prototype testing, customer quality monitoring instruments, field testing methods with sample customers, and independent assessment companies. These methods are limited in that (a) the number of customer evaluations is limited since the methods are cost constrained into a small number of experiments and (b) the methods are driven by a structured format defined by the design community and not the customer community. A method that integrates Web reviews into the product development process would have significant effects on the development process effectiveness.

1.4 Research Objectives

In this dissertation, the author has utilized data mining, more specifically opinion mining, methods to study the following two unexplored problems: how to detect whether or not there is design level information available in Web reviews and how to extract the product development ideas/opportunities from theses reviews. The achieved research objectives in support of the executed activities are:

i. Design and develop the DFOC (Design-Feature-Opinion-Cause Relationship) method for integrating the evaluation of unstructured Web reviews into the structured product design process. This method (a) identifies a set of design features that are of interest to the product design community, (b) mines the Web review database to identify which features are of significance to customer evaluations, (c) extracts and estimates the sentiment or opinion of the set of significant features, and (d) identifies the likely cause of the customer opinion.

ii. Develop an association rule mining procedure for capturing and extracting noun-verb-adjective relationships in the Web review database. This procedure supports the DFOC method and exploits existing opinion mining methods that integrate well-known natural language processing (NLP) algorithms. The mining method deconstructs the Web reviews into sentences, nouns, verbs, and adjectives. It captures feature-opinion pair polarity in reviews

iii. Develop the design-level information quality (DLIQ) measure, which evaluates three components (a) content, (b) complexity, and (c) relevancy. DLIQ is indicative of the content, complexity, and relevancy of the design contextual information that can be extracted from an analysis of Web reviews for a given product. Application of this measure confirms the hypothesis that significant levels of quality design information can be efficiently extracted from Web reviews for a wide variety of product types.

iv. Application of the DFOC method and the DLIQ measure to a wide variety of product classes (e.g., electronic, automobile, service domain). Specifically, Web review databases for ten products/services are created from real data. Validation occurs by analyzing and presenting the extracted product design information.

1.5 Organization of the Dissertation

This dissertation in seven chapters documents the research work and findings. Chapter 2 is an overview of background elements used in the study. In particular, opinion mining and the main tasks are introduced here: document preprocessing, linguistic text processing, classification, and performance evaluation metrics. In addition, association rule mining technique and its evaluation metrics are briefly introduced. Chapter 3 is a presentation of the research framework of the design feature opinion cause (DFOC) extraction method, and its main components: product features, feature-opinion pairs, and feature-causes of negawetive opinion pairs. Also, the results of the first research objective are presented. In Chapter 4, a novel hybrid method used to extract product design intelligence is presented. In Chapter 5, the experimental details are presented on a single product. In particular, the method is presented using one of the datasets as a walkthrough example. Chapter 6 the results of the identification of feature, feature-opinion pairs, and feature-cause pairs of the nine datasets for each product are presented. Chapter 7 is a summary of the method proposed in the study and indicates directions for future works.

CHAPTER 2

LITERATURE REVIEW

This dissertation was partly based on, and closely related to, opinion mining (alternatively, sentiment analysis) utilizing the association rule mining technique. Opinion mining is an interdisciplinary field that relies on information retrieval, text mining, data mining, natural language processing, machine learning, statistics, and computational linguistics. It is the process of extracting knowledge from consumer opinions, sentiments, and emotions toward products and their features. Various terms have been used by researchers to define opinion mining: sentiment classification, sentiment analysis, sentiment extraction, voice of customer analysis (Chung-Hong and Hsin-Chang 2005; Binali, Potdar et al. 2009). The goal of the research concerning opinion mining is the development of techniques, methods, systems, and tools that would be able to process a large amount of opinionated texts (e.g., online consumer reviews, discussion forms, and blogs). Opinion mining systems already have been applied in many areas of organizations, enabling technologies such as automatic (insulting) message detection systems in email and communication applications (Spertus 1997; Hayati and Potdar 2008; Tseng, Sung et al. 2008), political opinion classifiers in politics and government (Efron 2004; Yu, Kaufmann et al. 2008; Sarmento, Carvalho et al. 2009), opinion mining in legal blogs (Conrad and Schilder 2007), recommendation systems in e-commerce (Terveen, Hill et al. 1997; Tatemura 2000), ad classification and sensitivity detection (Giaglis, Kourouthanassis et al. 2003; Ge, Sipei et al. 2010), marketing intelligent systems and product/service benchmarking (Lee and Myaeng 2002; Bonchi, Castillo et al. 2011) in marketing, and others.

In this chapter, the terminology used in opinion mining is defined and general opinion mining is discussed. Further, a brief review of the association rule mining technique is presented.

2.1 Opinion Mining Terminologies

In this section, the basic terminology currently used in the area of opinion mining is defined.

An opinion is a judgment, viewpoint, or belief about something; it is commonly considered subjective and is the result of emotion or interpretation of facts. An opinion about something can be positive or negative; so positive and negative are called opinion polarities (sentiment orientation) (Liu 2010).

An object is any commented on target entity. This can be a product, service, event, organization, or topic.

An opinion holder (reviewer) is a person or an organization that expresses the opinion on an object.

A feature is a set of components of an object that has been commented on in a review. For example, a particular brand of computers is an object. The battery, screen, and memory are all features. An opinion can be expressed on any feature of the object (Liu 2010).

Text review (also called body of text, open-ended text, or comments) is a subjective unstructured text containing complete sentences, short comments, or both, describing opinions of a reviewer regarding a specific object.

Opinion polarity (also called semantic orientation) is an interpretation of the reviewer satisfaction concerning an object or feature in terms of a two-level orientation scale such as either positive or negative.

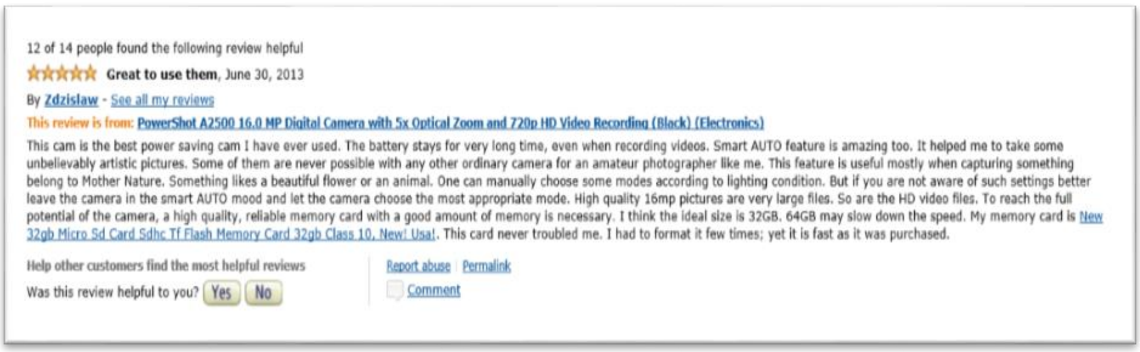


Figure 2.1 A sample review from Amazon.com.
Source: http://www.amazon.com, accessed on 10/12/2012.

2.2 General Opinion Mining Analyses

In this section, the existing and related studies about opinion mining proposed in the literature are presented.

Opinion mining refers to the use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information in documents where opinions were expressed. Opinion mining can be performed (1) at the document level, which is to categorize each whole document as positive, negative, or neutral, (2) on the sentence level, which is to categorize each sentence as expressing positive, negative, or neutral sentiment (e.g., sentiment analysis that is using words but is not extracting representative features), or (3) the feature-level, where each object's feature is graded as positive, negative or neutral.

In the past few years, an increasing number of researchers have begun expressing their research interest on these areas (Kohavi 2001; Hu and Liu 2004; Ding and Liu 2007; Anwer, Rashid et al. 2010; Zhixing 2010), proposed different methods to solve opinion mining problems, and proposed different opinion-oriented information-seeking systems (or algorithms).

In one of the early research studies, (Hatzivassiloglou and McKeown 1997) identified and validated opinion words and their semantic orientations from a large text corpus (Wall Street Journal) by using a non-hierarchical clustering algorithm (Spath 1985). The objective of the study was to identify the orientation of English adjectives and automatically classify them into two groups based on their semantic orientations, namely positive and negative.

As the most well-known example of opinion mining based on a semantic orientation approach, (Turney 2002) presented a simple unsupervised classification approach applying Web-based point-wise, mutual information statistics to determine review-polarity, where mutual information was calculated using Internet hit counts from the reviews of multiple domains, such as automobiles, banks, movies, and travel destinations from epinions.com. The system takes a set of user reviews as inputs, extracts phrases containing adjectives and adverbs, and then produces a classification as an output (recommended-not recommended). To extract phrases, they apply a simple word filter based on POS tags to select adjectives. Previous work has demonstrated adjectives are useful indicators of sentiment (Hatzivassiloglou and McKeown 1997; Hatzivassiloglou and Wiebe 2000). Via similar research, (Bollegala, Weir et al. 2011) focused on a cross domain sentiment classification system using an automatically created sentiment sensitive thesaurus from multiple domains (e.g., books, DVDs, electronics, kitchen appliance reviews from Amazon.com). The objective was to improve classification accuracy in a sentiment classifier. Their system splits labeled and unlabeled reviews into sentences; extracts opinion words (adjective, adverb, verb, and noun) from the sentences using the POS tagging method; creates the sensitive thesaurus using labeled training data; then produces two classes for the unlabeled target data.

(Landauer and Dumais 1997) presented a sentiment classification approach based on a mathematical method, latent semantic analysis (LSA), to analyze the relationships among words and identify the similarity of words in meaning (Menon and Elkan 2011). Their system uses singular value decomposition (SVD) to analyze the statistical relationships among words. The system takes a large text and generates a representation that captures the similarity of words. The proposed system has three major steps. Firstly, it constructs a rectangular matrix A from the large text corpus, where the row vectors represent words and the column vectors represent blocks of text (e.g., sentences, paragraphs, documents). Each cell in the matrix represents the weight of the word in the corresponding block of text. Secondly, it applies the singular value decomposition theorem to the matrix A; calculating the SVD consists of finding the eigenvalues and eigenvectors of AAT and ATA. The eigenvectors of ATA make up the columns of V; the eigenvectors of AAT make up the columns of U. In addition, the singular values in S were square roots of eigenvalues from AAT or ATA. Then, the semantic orientation of two words was calculated by SO-LSA equation—SO-LSA (word) = LSA (word, positive) − LSA (word, negative). Thirdly, it classifies the review as positive or negative based on the SO-LSA calculation.

Following the initial work of (Pang, Lee et al. 2002; Turney 2002) presented a machine learning classification technique with a bag of word as features using movie reviews as data. They experimented with a sentiment classification problem with three well known classification algorithms: Naive Bayes (Robles, Larranaga et al. 2003; Guo 2010), maximum entropy classification (Nigam, Laerty et al. 1999), and support vector machines (Wu, Kumar et al. 2007; Liu 2008).

All of the researchers talked about classifying reviews as positive, negative, or neutral at the document level. There is no doubt that document level sentiment analysis has been useful to businesses in many cases; however, it has failed to detect opinions about features of the products/services.

For example, s/he could be happy overall about his/her camera but s/he might be dissatisfied with battery-life. To businesses, these individual weaknesses and strengths were equally important to know; however, more importantly, understanding individual weakness and strength were even more valuable than the overall satisfaction level of customers. To obtain such detailed aspects, featured-based opinion analysis is needed.

In featured based opinion analysis, three main tasks can be performed: (i) identifying and extracting features of the product, (ii) determining whether the opinion about each feature is positive or negative, and (iii) producing a summary using discovered information. Over the last few years, feature-based opinion mining from consumer reviews has also been examined by a few researchers (Dave, Lawrence et al.

2003; Hu and Liu 2004; Kobayashi, Inui et al. 2004; Popescu and Etzioni 2005; Scaffidi, Bierhoff et al. 2007; Xiaojun, Lin et al. 2010; Zhai, Liu et al. 2010).

(Hu and Liu 2004) proposed feature-based summarization techniques for product reviews based on data mining and natural language processing methods. In the system, the summarization was performed in three steps: (i) mining nouns and noun phrases to indicate product features, (ii) identifying opinion sentences in each review and deciding whether each opinion sentence was positive or negative, and (iii) summarizing the result. To find product features, they used an association mining approach (Amir, Aumann et al. 2005) to find nouns and noun phrases that occurred together in sentences.

(Popescu and Etzioni 2005) proposed the OPINE system, which uses relaxation labeling to identify the semantic orientation of words. OPINE uses KnowItAll, which is an unsupervised, domain independent web-based information extraction system developed by (Etzioni, Cafarella et al. 2005). A set of user reviews from TripaAdvisor.com and Amazon.com was a system input, and a set of feature-opinion pairs for each domain became the output.

To produce the outputs, three major steps were run: first, the system parsed the reviews using MINIPAR parser (Lin 1998), and then extracted the words (i.e., adjective, adverb, and verb as opinion bearing words; noun phrases as product features); second, it was used to determine the polarity of opinions about features by computing the PMI scores and ranking the opinion words based on their strengths; finally, it was used to classify features as positive or negative using the naïve base classifier method (Robles, Larranaga et al. 2003; Kim, Lee et al. 2006).

(Ding, Liu et al. 2008) conducted sentiment analysis to determine whether the opinion expressed on a product was positive or negative by proposing different sentiment orientation calculation algorithms, which counts the number of positive and negative opinion words that were about the product feature in each review sentence. If there were more positive opinion words than negative opinion words, the final opinion on the feature was concluded to be positive and otherwise negative. The opinion words were obtained through a bootstrapping process (Adami, Avesani et al. 2003; Zhixing 2010) using WordNet (Miller 2012).

(Nasukawa and Yi 2003; Yi, Nasukawa et al. 2003) introduced the sentiment analyzer used to extracts opinions about a subject from online text documents (i.e., a digital camera, music reviews). Their sentiment analysis consisted of: (i) a topic specific feature term extraction, (ii) sentiment extraction, and (iii) (subject, sentiment) association by relationship analysis. They extracted only noun phrases from documents and applied feature selection algorithms; they developed and tested two feature term selection algorithms based on a mixture of a language model and a likelihood ratio.

(Su, Xu et al. 2008) introduced a mutual reinforcement approach to manage the feature-level opinion mining problem to discover hidden sentiment associations with Chinese Web pages. More specifically, the approach aggregated product features and opinion words simultaneously and iteratively by combining both their content information and sentiment link information. Then, under the same framework, based on the product feature categories and opinion word groups, they constructed the sentiment association set between the two groups of data objects by identifying their strongest sentiment links.

(Zhuang, Jing et al. 2006) applied some similar strategies to analyze movie reviews from the IMDB website. First, they created a keyword list to identify main feature/opinion words in movie reviews with help from WordNet. Then they applied grammatical rules between feature words and opinions to identify the feature-opinion pairs. Finally, they reconstructed the sentences according to the extracted feature-opinion pairs to generate the summary (e.g., positive or negative). This method has been applied, extended and improved in (Jindal and Liu 2006; Ding, Liu et al. 2008). (Hu and Liu 2004) proposed a bootstrapping approach, which uses a small set of given seed opinion words to find their synonyms and antonyms in WordNet (http://wordnet.princeton.edu).

(Guo, Zhu et al. 2009) examined customer reviews concerning multiple electronics products (i.e., digital cameras, laptops, and cell phones) along with different features—for digital cameras, batteries, memory, picture, and the screen were features; for a laptop, product feature terms became (e.g., battery, OS, processor, and screen).

(Du and Tan 2009) proposed an iterative reinforcement scheme based on the improved information bottleneck algorithm to address feature-based product opinion mining using hotel reviews. Unlike the traditional information bottleneck method (Tishby, Pereira et al. 1999; Chechik, Globerson et al. 2005), feature words and opinion words were organized into categories in a simultaneous and iterative manner by fusing both their semantic information and co-occurrence information.

(Liu, Hu et al. 2005) also were focused on online customer reviews of products. They proposed the opinion observer as a system that compares consumer opinions of multiple products, and visualizations of the results; they designed a supervised pattern discovery system automatically to identify product features from pros and cons in customer reviews.

(Hu and Liu 2004) introduced a technique that used various word features, including occurrence frequency, part-of-speech tagging and semantic orientation of words with help from WordNet. This was to identify a noun word and its nearest opinion words. (Hu and Liu 2004) proposed a technique based on association rule mining to extract product features. The main idea was that people often used the same words when they commented on the same product features. Then frequent itemsets of nouns in reviews were likely to be product features while the infrequent ones were less likely to be product features. This work also introduced the idea of using opinion words to find additional (often infrequent) features.

As presented above, there has been tremendous research concerning sentiment analysis (at the document or sentence level) as well as about feature-based sentiment analysis. However, most of this research concerned consumer marketing, and not much research was done regarding new product development. Launching a new product to a marketplace is according to the business of the design and manufacturing capability, as well as considering what customers need and prefer, and ultimately transferring the customers' opinions into the actual product itself (Bae and Kim 2011).

The customer opinions on certain products/services have been the most influential deciding factor concerning whether a business will be in the marketplace next year or not (Mudambi and Schuff 2010). As was discussed earlier, customer opinions (like or dislike) concerning product features can be captured through many different channels, such as interviews, surveys, and feedback from sales agents and retailers. Therefore, it makes sense that researchers previously focused on how to collect customer opinions including surveys, focus groups, direct customer contact, field intelligence and complaint analysis (Park and Lee 2011) and among many, what were the most useful method in terms of time and cost in order to analyze customer opinions on a product. The faster the economy grows, the shorter the product lifecycle becomes. This prompts business continually to determine how to reduce the data collection time and to remove geographic boundaries by using freely available consumer reviews.

With continuous efforts to make the best use of Web reviews, many researchers have presented outcomes concerning sentiment orientation of opinion words at different granularity levels such as words, sentences, and entities in Web reviews. However, it has not been actively contributing to product development managers who wish to apply the customer opinions to the product design stage. To the best of our knowledge, there was limited research effort in this area and it was definitely emerging as one of the most promising areas of study.

(Park and Lee 2011) focused on how to design and utilize an online customer center in an effort to support new product concept generation. They introduced the decision support system that identifies customer needs and materializes them to develop R&D targets in the new product development process. Their system consists of four stages: (i) extracting consumer reviews from the website (MobilePhoneSurvey.com), (ii) extracting a keywords list of entire documents and their frequencies, (iii) classifying the keywords into several groups based on customers' expressed needs, applying K-means clustering algorithm, and (iv) mapping customer needs with product specifications.

(Jin and Liu 2010) introduced the helpfulness prediction technique, which focused on how to connect customer reviews to product designer ratings in an automated fashion. Their proposed system has two phases: first, the systems create the connection between the customer review and the designer rating with the help of the training set. Then, the system extracts features from four aspects from reviews to aid in prediction, including linguistic features, product features, information quality (accuracy, timeliness, comparability, coverage, and relevance), and information theory.

2.3 Association Rule Mining Analysis

In this section, the overall association rule mining technique is presented. In this dissertation, in conjunction with opinion mining, the association rule mining technique was utilized only in identifying correlations between product features and opinions, and correlations between features and cause for negative opinions.

Association rule mining, which is a widely researched technique in data mining, was first introduced by Agrawal (Agrawal, Imieli et al. 1993; Bin and Zhijing 2003). The original motivation for searching association rules began with the need to analyze supermarket transactions. Association rule-based techniques were often used to determine customer behavior patterns. The classic application of association rules is the market basket data analysis, which targets discovering customer purchasing behaviors specifically, in detecting products (items) that frequently were purchased together. Analysis of transaction data is a commonly used approach to improve the quality of business decisions, and has a wide range of applications in many areas of business practice, such as adjusting store layouts (i.e., placing items optimally with respect to each other), for cross-selling, for promotions, for catalog design and to identify customer segments based on buying patterns (Agrawal, Imieliński et al. 1993; Agrawal and Srikant 1994; Agarwal, Aggarwal et al. 2000; Han, Pei et al. 2000; Mobasher, Dai et al. 2001; Lo 2002)

In the context of opinion mining, association rule mining was used to extract noun phrases as product features (e.g., battery life, hard drive, and picture quality). Both (Hu and Liu 2004) and (Popescu and Etzioni 2005) used association rule mining to extract the frequently occurred noun phrases as potential product features.

In general, the goal of an association rule mining algorithm is to discover associations between data items (or words). It is formally defined as assuming, "$I = (I_1, I_2, \ldots, I_n)$" (Holt and Chung 2007; Ruggieri 2010) is a collection of n different attributes (words), in given database D. Each record T is a collection of a set of attributes of I. That is, $T \subseteq I$ (every element of T is also an element of I). An association rule is an implication of the form $X \rightarrow Y$, where $X \subset I$, $Y \subset I$ were sets of items called itemsets, and $X \cap Y = \emptyset$. This indicates that if X appears in a transaction, Y will be led to appear in the same transaction inevitably. X is called the precondition of the rules, and Y is the result of the rules. The formal definition of association mining can be interpreted in this study as: $W = (W_1, W_2, \ldots, W_n)$ is a collection of n different words, in given sentence-database S, each sentence Si is a collection of a set of words of W. That is, $S \subseteq W$ (every element of S is also an element of W). An association rule is an implication of the form $X \rightarrow Y$, where $X \subset W$, $Y \subset W$ were sets of words called itemsets, and $X \cap Y = \emptyset$. This indicates that if X appears in a sentence, Y inevitably will appear in the same sentence. X is called the precondition of the rules, and Y is the result of the rules.

Association rule mining was applied utilizing two steps: (1) FPGrowth and (2) create association rule. Each rule obtained I accompanied by two meaningful measures, support and confidence. Support was defined as the percentage of documents (sentences) that contain X and Y together to the total number of documents (sentences) in the database, where X was itemset X, and Y was itemset Y. For instance, sup (X U Y) = number of documents that contained the X and Y together divided by the total number of documents. Confidence was defined as the percentage of the number of documents that contained X and Y together to the total number of documents that contained X. Confidence was a measure of strength of the association rules. For example, Conf (X->Y) = sup (X U Y)/ sup (X) (Agrawal, Imieliński et al. 1993; Tan, Steinbach et al. 2006). Support determined how often a rule was applicable to a given document set while confidence determined how frequently items in Y appeared in a document set that contained X.

CHAPTER 3

THE DESIGN–FEATURE–OPINION–CAUSE (DFOC) METHOD

The objective of this dissertation is to study consumers opinions expressed on product features in Web reviews. The basic hypothesis of this research is that Web reviews contain significant amount of information that is of value to the product design community. The conventional wisdom is that Web reviews are only of consumer interest since they only consist of consumer sentiment opinion, and are hence used only to influence consumer decisions. In this chapter, the research steps involve (i) initial characterization of Web reviews (ii) extraction of design intelligence from the reviews and (ii) statistical analysis of the results to confirm the hypothesis.

3.1 Research Definitions

First, some definitions are introduced. These relate to the research approach and method used in this dissertation. Note that because of the fast evolving nature of web-related applications and technology, the same term may have different definitions, and times many terms have the same definition.

i. Web Reviews: Through Web-based consumer opinion platforms (e.g., epinions.com), the Internet enables customers to share their opinions on, and experiences with, goods and services with a multitude of other consumers; that is, to engage in electronic word of-mouth communication (Hennig-Thurau, Gwinner et al. 2004). Web reviews can thus be defined as peer-generated online customer reviews typically recorded on third party websites (Mudambi and Schuff 2010). Web reviews are not authenticated or validated, that is there is no guarantee the author of the review was writing an honest review or has even experienced the product or service in question. However, the sheer volume of reviews in the Web makes them an accepted standard.

ii. Review Database: Each Web review is associated with a unique author, and hence represents the opinions and sentiments of a single consumer. A review database is a collection of hundreds of Web reviews, and therefore represents a population of customers. An effective analysis requires that a significant number of reviews were included in the database. These reviews were collected from multiple Web sources.

iii. Target Product/Service: Each Web review expresses the customer opinion on a specific product or service, which was identified in the platform where the review originates. This identification uniquely identifies the product and the aggregation level. For example consider the auto product Honda (Brand) – Accord (Model) – LX (Sub Model). If the identification states Honda Accord then the review database includes all reviews for the different Accord sub models but does not include reviews for other Honda models.

iv. Opinion/Sentiment: The polarity of a given review that was whether the expressed opinion was positive, negative, or neutral. This polarity can be evaluated at different levels for example in a document (entire review), a sentence, or an entity feature/aspect level. In this research the evaluation focus was at the sentence level and the feature-level. The subject of identifying the polarity in a given document was generally referred to as sentiment analysis (Prabowo & Thelwall, 2009). In advanced sentiment analysis the polarity was defined on an n-point scale, e.g., very good, good, satisfactory, bad, very bad. This research, though, was limited to the basic scale.

v. Design Feature: The basic axiom of product design was governed by the relationship diagram: Functional Requirement >> Design Feature >> Customer Satisfaction. Product designers were thus singularly focused on design features, which were the defining parameters of their specific product. Therefore, a design feature was defined as an attribute or characteristic of the design that was controllable variable for design community and a satisfaction focus for the customer community.

vi. Cause of Opinion: When an opinion or sentiment was expressed then the follow-up query was why the opinion was formed. This "why" was defined as the cause of the opinion, and was central to any design improvement that would focus on improving customer opinion. The cause was contextually at the same level as the opinion. Here the focus was specifically at feature-level causes. Causes could be described as a functionality, physical element, or perceived performance. Note that most Web reviews do not explicitly mention the cause, and frequently only express an opinion. Cause was a justification of action, event, or opinion (Liu, Hu et al. 2005).

3.2 The DFOC (Design-Feature-Opinion-Cause) Relationship

In his classical textbook on design theory Professor Suh of the Massachusetts Institute of Technology (P.Suh 1990; Suh 2000) describes product design as the interplay between “what we want to achieve” and “how we achieve it.” A designer tries to obtain what he/she wants to achieve through appropriate interplay between both sides. Based on this theory a classical design analysis theorem was proposed by (Do and Suh 2001); Figure 3.1). Functional requirements are defined as being equivalent to “what we want to achieve.” These requirements are satisfied by defining or selecting design parameters in the physical domain. Finally, the success or quality of the design is determined by how satisfied the end user (customer) is. Contemporary design practice evolves from this classical approach, but driven by competitiveness puts increased emphasis on the customer satisfaction component.

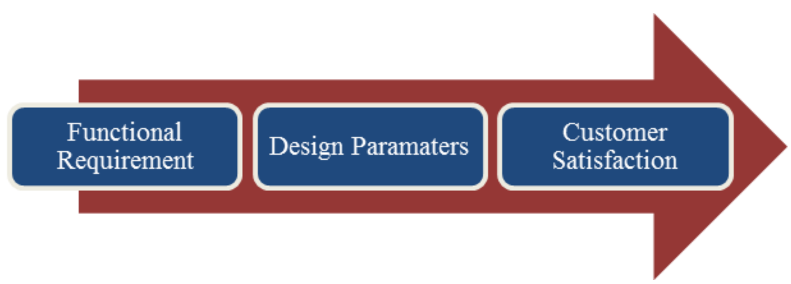


Figure 3.1 Classical design axiom.
Source: (Do and Suh 2001).

Figure 3.2 Customer drive design optimization process.

Figure 3.2 is an illustration of the design view adopted and pursued in this research. This view was driven by classical industrial engineering methods including cause-effect analysis and quality function deployment. In this process, product development managers were constantly challenged to learn what the consumer product experience really is, and to learn specifically how the product was performing in the field. Traditionally, they have utilized a variety of methods including prototype testing, customer quality monitoring instruments, field testing methods with sample customers, and independent assessment companies. These methods were limited in that (i) the number of customer evaluations was limited since the methods were cost constrained into a small number of experiments, and (ii) the methods were driven by a structured format defined by the design community and not the customer community.

Today, the Web has created a new customer evaluation channel which overcomes the above two limitations. Web reviews were unsolicited reviews from actual product users that were posted across hundreds of websites. Presently, these reviews were primarily used by potential buyers to learn the product experience of other consumers, and it was well known that this has a significant impact on the buying decisions. A second growing use was in marketing where reviews were analyzed to project product sentiment (positive to negative). This research integrates Web reviews into the customer evaluation component for Figure 3.2. Thus, this research is an attempt to extract specific product intelligence that can then be used to develop better product designs. For example, one may learn from the reviews that the image stabilization feature in a digital camera malfunctions in humid conditions. Frequently, such issues were occurring in a small percentage of the user base and not detected in traditional methods.

The process developed here was labeled as DFOC or the *Design – Feature – Opinion – Cause Relationship*. That was for a target product design the method first identified a set of design features that were of interest to the product design community. Second, the method mined the review database to identify which of these features were of significance to customer evaluations, third the sentiment or opinion of the set of significant features were extracted and estimated, and fourth DFOC identifies the likely cause of the customer opinion. The DFOC relationship connected the structured design process to the unstructured Web review process. In this chapter, the author demonstrates both the feasibility and utility of the DFOC process.

3.3 The Data Mining Tool

A key analytical tool in this research was data mining or more specifically opinion mining. Sentiment analysis or opinion mining can be described as the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics, and their attributes (Zhang and Liu 2011). They also characterize an opinion as a quintuple, that was [i] the name of the entity (target product), [ii] an aspect (design feature in a target product) of the entity, [iii] the polarity of the opinion about the aspect-entity, [iv] the opinion holder or the author of Web review, and [v] the time when the opinion was recorded. Opinion mining is based on natural language processing was challenging, because it required a deep understanding of the explicit and implicit, regular and irregular, and syntactical and semantic language rules. Opinion mining was a critical part of the DFOC process, in that it was the tool by which the design process was linked with the customer evaluation process. This research does not develop new opinion mining tools, but rather exploits existing methods and applies them in the DFOC process.

Existing tools for opinion mining range from simple Web applications to much more complex toolkits or frameworks. The more well known commercially available and open source tools that are widely used in the research community are:

i. GATE: General Architecture for text Engineering - is an open source tool for text analytics and semantics. It is capable of solving almost any NLP problem. GATE uses supervised machine learning methods, trained on human-annotated data, co-occurrence statistics, and lexicons of positive and negative words, in order to identify problems with products and company services reported on blogs (http://gate.ac.uk).

ii. RAPIDMINER: RapidMiner is one of the world's most widespread and most used open source data mining solutions written in Java. The project was born at the University of Dortmund in 2001 and has been developed further by Rapid-I GmbH since 2007. Provides data integration, analytical ETL, data analysis, and reporting in one single suite. RapidMiner can be used as a flexible data analysis tool, since it provides a wide range of methods from simple statistical evaluations such as correlation analysis as well as dimension reduction and parameter optimization. These methods can be used for various application domains such as text, image, audio and time series analysis (http://www.rapid-i.com).

iii. TRENDMINER: An innovative, portable open-source real-time methods for cross-lingual mining and summarization of large-scale stream media. TrendMiner achieves this through an inter-disciplinary approach, combining deep linguistic methods from text processing, knowledge-based reasoning from Web science, machine learning, economics, and political science. No expensive human annotated data will be required because of the use of time-series data (e.g. financial markets, political polls) as a proxy. A key novelty will be weakly

> supervised machine learning algorithms for automatic discovery of new trends and correlations. Scalability and affordability will be addressed through a cloud-based infrastructure for real-time text mining from stream media (http://www.trendminer-project.eu).

In this research, the decision was made to employ RapidMiner. The decision was based on the available suite of functions and the ease of obtaining a research use license. Since RapidMiner emerged from the YALE data mining environment (Mierswa, Wurst et al. 2006) it was highly amenable for use in the DFOC process. Originally designed to be a rapid prototyping system where data mining implementations could undergo a proof-of-concept using a tool that can easily build, execute, and validate data mining models, before the need to develop a more complex solution. RapidMiner has evolved into an offering with commercial strength features such as Ability to quickly prototype data mining tasks on a graphical user interface and additional functionality specific to text mining. Key details of the RapidMiner application are given in Chapter 4.

3.4 Design-Feature-Opinion-Cause (DFOC) Research Method

The author introduced and proposed a method to extract specific product intelligence utilizing opinion mining and association rule mining techniques, which can be used to develop new products. A feature-based sentiment analysis on sentence level was conducted to investigate product development intelligence, and then an association mining technique was utilized to identify correlations between feature, opinion, and cause. The method in this study was composed of the following steps: data collection and preparation, identifying product features, identifying opinions regarding product features and determining the polarity of each opinion, and identifying cause of each negative opinion associated with features, Figure 3.3 is an illustration of the overall DFOC research method framework.

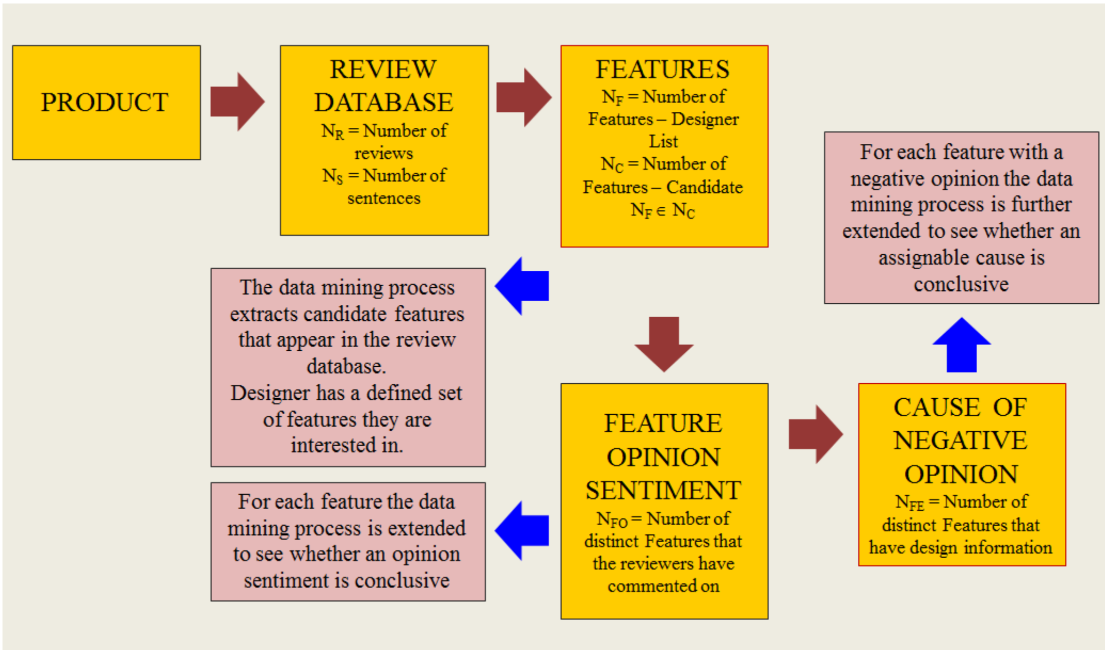


Figure 3.3 The DFOC research method framework.

3.4.1 Step 1 Product

Defining the Product: In the DFOC method, the term product is used to denote both product and service that has been commented on. For example, a particular brand of laptop, digital camera, or restaurant is a product. A product can also have different components. For example, a laptop has set of components (e.g., hard drive, operating system, and keyboard). To simplify the discussion the term product used to represent both product and service, and the word feature used to represent components of products. In general, opinions (view or judgment) can be expressed on anything e.g. product/service or features. A sentence "I don't like this laptop" is an example of opinion on a product. Product features in a sentence can appear either explicitly or implicitly. For example, “The battery life of this laptop is amazing,” in this sentence “battery life” is an explicit feature, while a sentence: “This laptop is too small,” size does not appear in this sentence, so it is an implicit feature. In this study, only explicit features were used.

Product Selection: Before started building the research methodology, first, the author explored the most popular approaches for opinion mining in the literatures and their underlying methodology of identifying and analyzing Web reviews on products and services. Second, to make the discussion and result more concrete, the author identified the ten products for this research from different industries primarily based on the frequency of appearance in the literature reviews and the richness of contents from online consumer reviews. The main criterion for selection was that the products have relatively large number of Web review. However, this method is general enough to be easily adapted to handling other types of Web reviews.

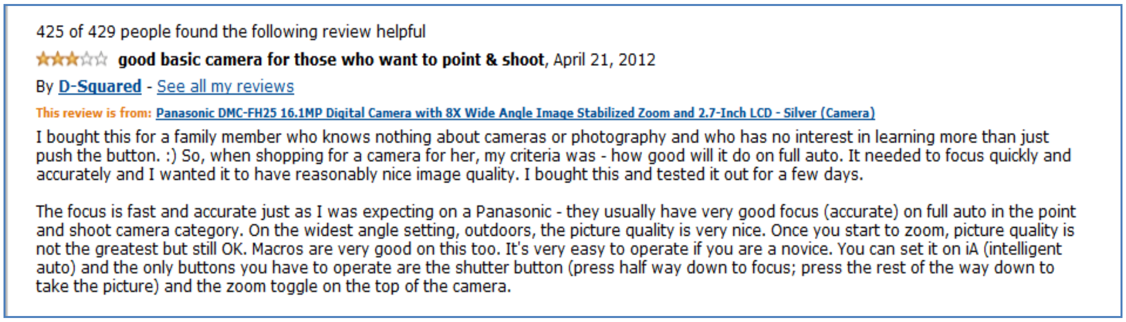
3.4.2 Step 2 Review Database

Data Collection: For this experiment, over 6000 online-reviews were manually collected for ten products and services: sedan, sports car, laptop, digital camera, mobile phone, television, airline customer service, mobile phone service provider, hotel, and restaurant. The Web reviews and their sources used in this study are displayed in Table 3.1.

Table 3.1 Product/Service List and Sources

Product Type	Industry	N_R	Primary Web Source
television	electronics	480	Amazon.com
digital camera	electronics	745	Amazon.com
laptop	computer	312	Amazon.com
mobile phone	mobile	352	Amazon.com
sports car	automotive	671	CarReview.com and KBB.com
sedan	automotive	749	KBB.com
mobile service provider	service provider	847	CustomerServiceScoreboard.com
airline customer service	travel	570	CustomerServiceScoreboard.com
restaurant	food and beverage	821	Yelp.com
hotel	hotel and lodging	528	TripAdvisor.com

Our data was collected over the course of one month during 2012 from various sources. Often, manual collection was necessary in some websites because they disallowed crawling their websites (i.e. Amazon.com; (Hu and Liu 2004). All reviews collected were written in free text format; some reviews were relatively short in one or two sentences, some reviews resemble advertisements or endorsements, and some reviews were written in informal language with poor structure. A typical review has date, user id, some have a numerical rating (like stars), and the body of the review where they share their experiences with products or services (Figure 3.4). The review body where consumers share their opinions was of interest, the remaining parts were called noisy data, which were excluded from this study.


425 of 429 people found the following review helpful

good basic camera for those who want to point & shoot, April 21, 2012

By D-Squared - See all my reviews

This review is from: Panasonic DMC-FH25 16.1MP Digital Camera with 8X Wide Angle Image Stabilized Zoom and 2.7-Inch LCD - Silver (Camera)

I bought this for a family member who knows nothing about cameras or photography and who has no interest in learning more than just push the button. :) So, when shopping for a camera for her, my criteria was - how good will it do on full auto. It needed to focus quickly and accurately and I wanted it to have reasonably nice image quality. I bought this and tested it out for a few days.

The focus is fast and accurate just as I was expecting on a Panasonic - they usually have very good focus (accurate) on full auto in the point and shoot camera category. On the widest angle setting, outdoors, the picture quality is very nice. Once you start to zoom, picture quality is not the greatest but still OK. Macros are very good on this too. It's very easy to operate if you are a novice. You can set it on iA (intelligent auto) and the only buttons you have to operate are the shutter button (press half way down to focus; press the rest of the way down to take the picture) and the zoom toggle on the top of the camera.

Figure 3.4 Example of website product review.
Source: ***http://www.amazon.com****, accessed on 10/12/2012.*

Data Preparation: Data preparation includes data cleaning and data transformation, which were important steps in the opinion mining process, as the quality of data affects the results. To improve the quality of data as well as overall efficiency of the mining process, raw data (sometimes called source data) was preprocessed. There were some commonly used data cleaning tasks that included removing non-textual contents and markup tags (for HTML pages), and removing information about the reviews that was not required for sentiment analysis (e.g., review dates, reviewers' names, and treating punctuation; (Pang, Lee et al. 2002). The reviews before and after the data cleaning process is demonstrated as an example in Figure 3.5 and Figure 3.6, respectively. Figure 3.5 shows the raw data, which is seen on the website. Figure 3.6 shows the data after the preparation process.

0 of 3 people found the following review helpful 4.0 out of 5 stars the price is worth this product but not big enough for drawing room, February 28, 2012 By shan li - See all my reviews Amazon Verified Purchase (What's this?) This review is from: Samsung UN55D8000 55-Inch 1080p 240 Hz 3D LED HDTV (Silver) [2011 MODEL] (Electronics) I brought LG 55' first' this one is fine), then i like to have 3D function so i paid 3,4 hundred more to get the LG 55' 3D one. However, the second one has dead pixel and the remote control is not working properly. Finally, I paid $1976 (including tax and shipping) to get the best band “Samsung“. If you think $2000 is affordable and 3D function is the must and 55' is acceptable (for me, I already used like 2,3 months, I feel 55' is okay for bedroom. but if you use for like 30 square feet or more in the drawing room that is not big enough) i think this is the best. Help other customers find the most helpful reviews Was this review helpful to you? Yes No Report abuse | Permalink Comment

Figure 3.5 Example of raw data.
*Source: **http://www.amazon.com**, accessed on 10/12/2012.*

I brought LG 55' first' this one is fine), then i like to have 3D function so I paid 3,4 hundred more to get the LG 55' 3D one. However, the second one has dead pixel and the remote control is not working properly. Finally, I paid $1976 (including tax and shipping) to get the best band “Samsung”. If you think $2000 is affordable and 3D function is the must and 55' is acceptable (for me, I already used like 2,3 months, I feel 55' is okay for bedroom. but if you use for like 30 square feet or more in the drawing room that is not big enough) I think this is the best.

Figure 3.6 Example of data after preprocessing.

Sentence Splitting: The reviews were split into sentences for accurate feature extraction to increase the chance of correct word grouping as product features. Sentence splitting was a process for segmenting a set of reviews containing several sentences based on punctuation characters. Sentence splitting was necessary because some reviews contained several features, each of which represented different features upon which consumers commented.

Here was a simple example from an automobile database showing how two words could be incorrectly put together if the sentence splitting was not processed. For example, this review has three sentences. “I hate this car. The trunk space is too small. Tiers are torn down very fast.” In this review, the reviewer has opinions about three different features, which were car, trunk space, and tiers. Without sentence splitting, they could be grouped incorrectly as two consecutive words because of the nature of language processing when the method tries to extract noun phrases. For instance, the words car-trunk, and space-tiers can be grouped together. As car-trunk, space-tiers were not in the same sentence, they should not be grouped together as two consecutive words. Therefore, this incorrect word grouping was eliminated by splitting documents into sentences and handling each sentence as an individual review. After the sentence-split, they become three individual sentences (each sentence become a document). For example, “I hate this car,” “The trunk space is too small,” and “Tires are torn down very fast.”

In this research, the reviews were split into sentences to attain accurate feature extraction based on punctuation characters [. ?] as delimiters and saving sentences in the review database, which means each sentence becomes an individual review. After sentence splitting, it was assumed that each sentence contained a single feature upon which the reviewer had an opinion (Jeong, Shin et al. 2011).

3.4.3 Step 3 Product Feature Extraction

The goal of this step was to extract product features that had been commented on in the product reviews.

Before extracting the product features, two properties of product features based on the considerations in Hu et al. (Hu and Liu 2004) were considered. First, the product features were nouns or noun phrases (Zhang, Chen et al. 2006); (Hu and Liu 2004). Second, product features were directly related to opinion-bearing words. One product could have many features. For example, a product such as an automobile could have features (e.g., a transmission, seats, mirror, door handle). Each feature could be expressed with a finite set of words or phrases. For example, a transmission might be expressed as a clutch or a gear. Therefore, it was difficult for a computer to understand such fuzzy phrases and features. POS tagging was applied to extract the product features after several preprocessing steps—removing stop words, stemming, and manual fuzzy matching. Two-gram (bi-gram) was used to extract frequently occurring noun phrases as candidate (potential) product features (Pang, Lee et al. 2002; Pang and Lee 2008). In the next chapter, the detailed steps will be discussed.

3.4.4 Step 4 Product Feature and Opinion Extraction

The goal of this step was to determine whether opinions about the product features were positive or negative. Adjectives were found as effective terms for identifying opinion words in polarity classifications and considered a key piece in the opinion extraction process (Hatzivassiloglou and McKeown 1997; Wiebe, Bruce et al. 1999).

In the study, adjectives were considered opinion words expressed about product features. Adjectives were identified and a synonym of an adjective was replaced utilizing a synonymous set in the WordNet in order to increase term frequency (e.g., magnificent became amazing, or cheapest became cheap; (Miller 2012).

Then, both product features and opinion words together were extracted from the sentences that contained frequent nouns or noun phrases. After extracting the opinions and features, the association mining rule was applied to extract the most frequent feature-opinion pairs. After extracting the pairs, a manual pruning method was used to remove nouns and adjectives that were unlikely to be the product features or the opinion. Finally, the polarities of the opinion words (such as positive and negative) were assigned by applying a heuristic approach.

3.4.5 Step 5 Product Feature and Cause Extraction

The goal of this step was to determine the cause of negative opinions on product features. The causes of distinctive negative features were attempted to be analyzed. If there were no distinctive negative-features, features that had both positive and negative aspects were analyzed to understand what causes negative attitudes toward these features. Verbs were considered as opinion reasoning (cause) along with opinions and features. For example, there was a review saying: "The transmission is bad because the cooler line had rusted out and lost all fluid."

The reviewer had a negative experience concerning the transmission, and then he explained the transmission was bad because the cooler line had rusted out. The objective was not only to identify a negative opinion about the transmission, but also to identify what caused this negative opinion. In this review, it was because of rust in the cooler line.

Verbs were first identified and then the data were normalized to root words using WordNet (Miller 2012) and a custom-made dictionary was created from the collection of reviews in order to increase term frequency (e.g., the word begins, began, beginning becomes begin). After extracting the feature, opinion, and cause from each sentence, the association mining rule was applied to extract feature-cause pairs. After extracting the pairs, a manual pruning method was used to remove nouns and verbs that were unlikely to be the product features or the cause. Finally, the cause of the negative opinion for the features was listed.

3.5 Measuring the Design Level Information Quality in Web Reviews

A key objective of this research was to investigate: What amount of design-level-information is available in Web reviews? The investigative hypothesis then is: *Significant levels of quality design information can be extracted efficiently from Web reviews for a wide variety of classes (e.g., electronic, automobile, service domain), and this information can be directly related to product features of specific interest to the product design community.*

To test this hypothesis a design-level information quality (DLIQ) measure was introduced. DLIQ is indicative of the content, complexity, and relevancy of the design contextual information that can be extracted from an analysis of Web reviews for a given product. The measure is lower bounded at zero and a measure of DLIQ = 100 indicated very high information quality and only 20% of products would typically have a DLIQ measure of 100 or higher. The literature indicates the dominant theme in measuring information is entropy. Entropy is a measure of the uncertainty associated with a random variable. In this context, the term usually refers to Shannon entropy, which quantifies the expected value of the information contained in a message, usually in units such as bits. Common practice is to adopt a binary entropy approach and a function using the logarithmic base 2 is developed. This approach is adopted here to evaluate the DLIQ measure for Web reviews. First, some notation is introduced to describe key parameters in the extracted review database for a specific product.

N_R total number of unique Web reviews in the database;

N_S total number of sentences generating by splitting all reviews in the database;

N_W total number of words authored in all reviews in the database;

N_C total number of noun words identified by the data mining process from all reviews; and

N_F total number of noun words matched by the data mining process to designer listed features.

Additionally, $\mu(\ )_M$ and $\sigma(\ )_M$ represent the mean and standard deviation for the above parameters across a representative population of products. That is if $j \in M$ is the product population, then $\mu(N_R)_M$ is the average N_R for the M products, and $\sigma(N_R)_M$ is the average standard deviation. The three components of the DLIQ are defined as follows. Each component is assigned a weight W in the DLIQ measure. In this study the setting as follows: $W_{Cont} = 0.30$, $W_{Cplx} = 0.30$, and $W_{Relv} = 0.40$.

DLIQ Content Measure - is an evaluation of the total amount of information that is available in the review database. This represents the volume of the raw data that is to be analyzed. Key determinants are the number of reviews and the length of the reviews as measured by the number of words. The rationale is that information sourced from a larger number of unique reviewers coupled with more wordy reviews is more likely to generate valuable information. The DLIQ content measure will increase proportionately as each of these determinants increase, and the base metric is:

$$DLIQ_{Cont}\ (base) = Log_2(N_R) * Log_2\left(\frac{N_W}{2}\right) \tag{3.1}$$

To benchmark this measure against the product population, the $DLIQ_{Cont}$ (benchmark) is calculated for the N_R and N_W levels corresponding to the 80th percentile of the population. This level is considered equivalent to a DLIQ measure of 100.

$$DLIQ_{Cont}\ (benchmark) \\ = Log_2(NORMINV[80\%, \mu\{N_R\}_M, \sigma\{N_R\}_M]) \\ * Log_2\left(\frac{NORMINV[80\%, \mu\{N_W\}_M, \sigma\{N_W\}_M]}{2}\right) \tag{3.2}$$

The NORMINV function above derives the parameters corresponding to the benchmark level for the population. For this study, the population is described by the set of products listed in Table 3.1. The scaled measure is then:

$$DLIQ_{Cont} = W_{Cont} \left\{\frac{Log_2(N_R) * Log_2\left(\frac{N_W}{2}\right)}{DLIQ_{Cont}\ (benchmark)}\right\} \tag{3.3}$$

DLIQ Complexity Measure - A complex Web review will consist of several sentences and will also include many nouns. The more complex the Web review, the more likely it will include valuable information that can be effectively utilized by the design community. Key determinants are the number of sentences per review and the ratio of nouns to words in a review. The rationale is that when reviews refer to nouns in long sentences they are more likely to be discussing specific product features in significant detail. The DLIQ complexity measure will increase proportionately as each of these determinants increase, and the base metric is:

$$DLIQ_{Cplx}\ (Base) = Log_2\left(\left[\frac{N_S}{N_R}\right] * \left[0.1 + \left(\frac{N_C}{N_W}\right)\right]\right) \tag{3.4}$$

To benchmark this measure against the product population, the $DLIQ_{Cplx}$ (benchmark) is calculated for the N_s/N_R and N_C/N_W ratios corresponding to the 80th percentile of the population. This level is considered equivalent to a DLIQ measure of 100.

$$\begin{aligned} DLIQ_{Cplx}\ (benchmark) \\ = Log_2\Bigg(NORMINV\left[80\%, \mu\left\{\frac{N_S}{N_R}\right\}_M, \sigma\left\{\frac{N_S}{N_R}\right\}_M\right] \\ * \left[0.1 + NORMINV\left[80\%, \mu\left\{\frac{N_C}{N_W}\right\}_M, \sigma\left\{\frac{N_C}{N_W}\right\}_M\right]\right]\Bigg) \end{aligned} \tag{3.5}$$

The scaled measure is then:

$$DLIQ_{Cplx} = W_{Cplx}\left\{\frac{Log_2\left(\left[\frac{N_S}{N_R}\right] * \left[0.1 + \left(\frac{N_C}{N_W}\right)\right]\right)}{DLIQ_{Cplx}\ (benchmark)}\right\} \tag{3.6}$$

DLIQ Relevancy Measure - A relevant Web review will identify many of the features that are of specific interest to the product design community. The more such noun features are mentioned the more likely the Web review will include specific product features related to design information. Key determinants are the total volume of nouns in the reviews and the ratio of feature nouns to the total nouns in the review database. The rationale is that when reviews consist of many nouns and include many noun features then the reviewer is more likely discussing their opinion and sentiment of a specific product feature. The DLIQ relevancy measure will increase proportionately as each of these determinants increase, and the base metric is:

$$DLIQ_{Relv}\ (base) = Log_2(N_C) * Log_2\left(\left[1 + \left(\frac{N_F}{N_C}\right)\right]^2\right) \tag{3.7}$$

The square function amplifies the information measure as the ratio N_F/N_C increases. To benchmark this measure against the product population, the $DLIQ_{Relv}$ (benchmark) is calculated for the N_C and N_F/N_C ratios corresponding to the 80th percentile of the population. This level is considered equivalent to a DLIQ measure of 100.

$$\begin{aligned} DLIQ_{Relv}\ (benchmark) \\ = Log_2(NORMINV[80\%, \mu\{N_C\}_M, \sigma\{N_C\}_M]) * Log_2\Bigg(1 \\ + NORMINV\left[80\%, \mu\left\{\frac{N_F}{N_C}\right\}_M, \sigma\left\{\frac{N_F}{N_C}\right\}_M\right]\Bigg) \end{aligned} \tag{3.8}$$

The scaled measure is then:

$$DLIQ_{Relv} = W_{Relv} \left\{ \frac{Log_2(N_C) * Log_2\left(\left[1 + \left(\frac{N_F}{N_C}\right)\right]^2\right)}{DLIQ_{Relv}\ (benchmark)} \right\} \tag{3.9}$$

The composite DLIQ measure is then given by:

$$DLIQ = 100\ \left(DLIQ_{Cont} + DLIQ_{Cplx} + DLIQ_{Relv}\right) \tag{3.10}$$

In this metric, a value of 100 represents a very significant level of information, while a measure of 70 indicates a promising level for design-level information quality. This sets the threshold of whether there are enough product features in the reviews. Therefore, one can move forward with the next step to analyze positive/negative features and the reasoning behind the negative ones. Alternatively, if the score ranges between 50-70, one might move forward with the next step expecting a limited outcome. A score below 50 indicates one must go back and collect additional data.

3.6 The DLIQ for the Product Population

The DLIQ model was applied to all ten products/services in the Web review database. For the television, 480 reviews from Amazon.com were extracted. After splitting the reviews into sentences, 6765 sentences were found. The number of sentences per review (N_S/N_R) was 14.09. The number of words per sentence (N_W/N_S) was 15.89. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 22%.

For the digital camera, 745 reviews from Amazon.com were extracted. After splitting the reviews into sentences, 7455 sentences were found. The number of sentences per review (N_S/N_R) was over 10.01. The number of words per sentence (N_W/N_S) was 15.85. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 12%.

For the laptop, 312 reviews from Amazon.com were extracted. After splitting the reviews into sentences, 3256 sentences were found. The number of sentences per review (N_S/N_R) was over 10.44. The number of words per sentence (N_W/N_S) was 15.01. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 13%.

For the mobile phone, 352 reviews from Amazon.com were extracted. After splitting the reviews into sentences, 1405 sentences were found. The number of sentences per review (N_S/N_R) was over 3.99. The number of words per sentence (N_W/N_S) was 14.79. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 8%.

For the sports car, 671 reviews from two sources, namely CarReview.com and KBB.com, were extracted. Two separate datasets were extracted in order to have a larger number of reviews in one collection for this domain. After merging these two datasets into one, the reviews were split into sentences, producing 6595 sentences. The number of sentences per review (N_S/N_R) was over 9.83. The number of words per sentence (N_W/N_S) was 12.50. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 15%.

For the sedan reviews, 749 reviews from two sources, namely KBB.com and CarReview.com, were extracted. After splitting the reviews into sentences, 3428 sentences were found. The number of sentences per review (N_S/N_R) was over 4.58. The number of words per sentence (N_W/N_S) was 12.09. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 14%.

For the mobile phone service provider, 847 reviews from CustomerServiceScoreboard.com were extracted. After splitting the reviews into sentences, 7200 sentences were found. The number of sentences per review (N_S/N_R) was over 8.50. The number of words per sentence (N_W/N_S) was 16.72. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 11%.

For the airline customer service, 570 reviews were extracted from CustomerServicescoreBoard.com. After splitting the reviews into sentences, 5370 sentences were found. The number of sentences per review (N_S/N_R) was over 9.42. The number of words per sentence (N_W/N_S) was 16.65. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 15%.

For the restaurant, 821 reviews from yelp.com were extracted. After splitting the reviews into sentences, 11718 sentences were found. The number of sentences per review (N_S/N_R) was over 14.27. The number of words per sentence (N_W/N_S) was 15.21. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 18%.

For the hotel, 528 reviews from CustomerServiceScoreboard.com were extracted. After splitting the reviews into sentences, 5788 sentences were found. The number of sentences per review (N_S/N_R) was over 10.96 sentences per review. The number of words per sentence (N_W/N_S) was 15.39. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 10%. The summary stats are demonstrated in Table 3.2.

Table 3.2 Basic Statistics for Ten Products and Services Used in this Study

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
television	480	6765	14.09	107518	15.89	22969	5049	3.40	0.22
digital camera	745	7455	10.01	118173	15.85	24320	2937	3.26	0.12
laptop	312	3256	10.44	48882	15.01	10876	1421	3.34	0.13
cell phone	352	1405	3.99	20778	14.79	5280	664	3.76	0.13
sports car	671	6595	9.83	82426	12.50	20512	2984	3.11	0.15
sedan	749	3428	4.58	41449	12.09	8014	1161	2.34	0.14
mobile phone service provider	847	7200	8.50	120384	16.72	24704	2831	3.43	0.11
airline travel	570	5370	9.42	89405	16.65	17269	2617	3.22	0.15
restaurant	821	11718	14.27	178258	15.21	42116	7578	3.59	0.18
hotel	528	5788	10.96	89104	15.39	21768	2107	3.76	0.10
minimum	312	1405	3.99	20778	12.09	5280	664	2.34	0.10
maximum	847	11718	14.27	178258	16.72	42116	7578	3.76	0.22
average	608	5898	9.61	89638	15.01	19783	2935	3.32	0.14

There are many linguistic theories concerning the average number of sentences per Web review. The length of the review varies on the opinions associated with features in the review: according to some studies, the length of reviews on restaurants with negative opinions was identified as 6.3 sentences, while overall length of review (negative and positive reviews) was identified as 3.78 sentences (Ganu, Marian et al. 2010). Another study shows that the length of review was nine or ten sentences for hotels, and eight sentences for airlines (Khan, Baharudin et al. 2010; Li and Chen 2010).

In this study, the average number of sentences in Web reviews for all products was 9.61 sentences (STDV 3.38), ranging from the highest at 14.27 in television to the lowest at 3.99 in mobile phone reviews. The high range indicates that some product reviews contain few sentences, while others contain more sentences in which reviewers have opinions on the product features. However, overall Web reviews by themselves are valuable whether their length is short or not.

It was observed in this data set that restaurant and television have the highest number of sentences per review, followed by hotels. Alternatively, mobile phone has the least number of sentences per review. Compared with the average of 3.78 sentences per review in the literature (Ganu, Marian et al. 2010), it seems Web reviews collected have enough information about products and services to be utilized for new product development. The statistics of ten products and services are demonstrated in Figure 3.7.

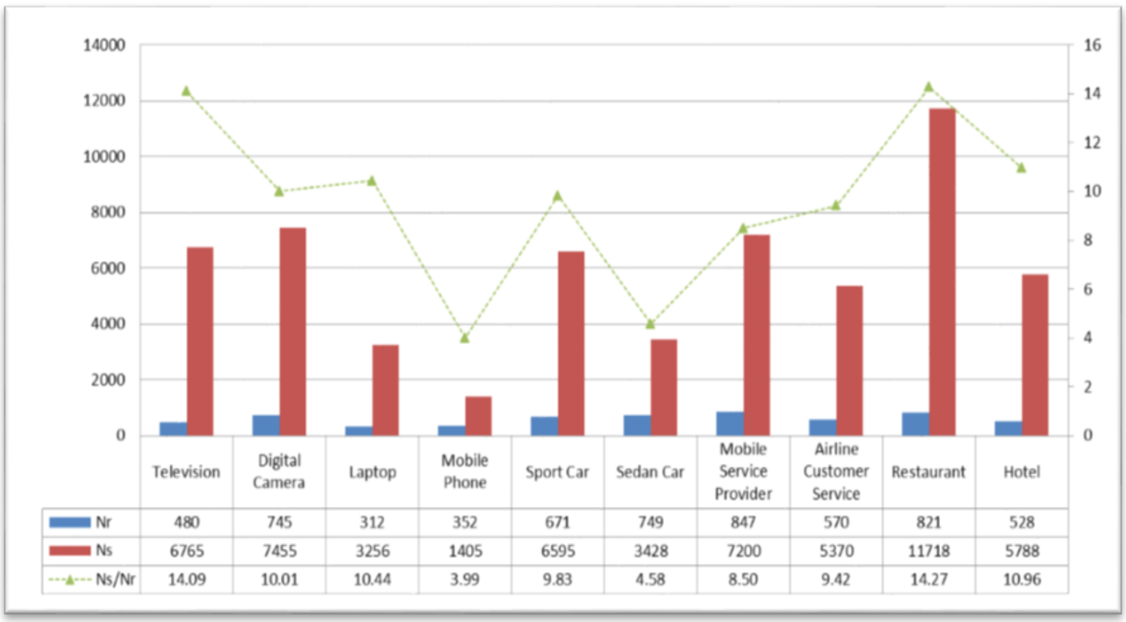
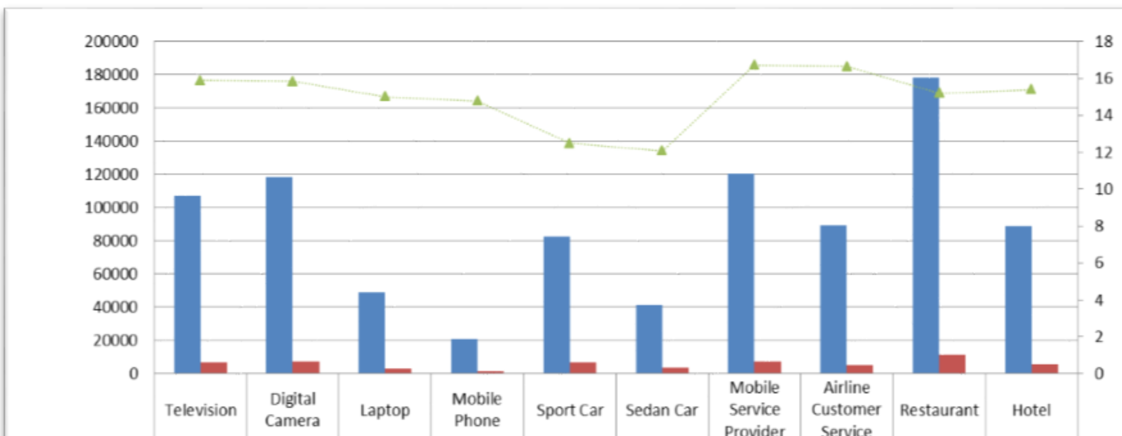


Figure 3.7 Average number of sentence per review.

There are also many linguistic theories about the average number of word per sentence. Some studies reveal eight seems to be a useful number for Web reviews (Ganu, Marian et al. 2010; Khan and Baharudin 2011).

The results show that the average number of words per sentence (N_W/N_S) is 15.01 (STD 1.57). Compared with eight words per sentence, it seems the Web reviews collected have sufficient information to explore further.

In terms of the average sentence length, airline customer service and mobile phone service provider reviews show better results followed by television, mobile phone, and hotel. For the car reviews, the lowest scores among the data sets were identified. However, it is still above the eight words per sentence threshold (Ganu, Marian et al. 2010; Khan and Baharudin 2011). The statistics are demonstrated in Figure 3.8.



	Television	Digital Camera	Laptop	Mobile Phone	Sport Car	Sedan Car	Mobile Service Provider	Airline Customer Service	Restaurant	Hotel
Nw	107518	118173	48882	20778	82426	41449	120384	89405	178258	89104
Ns	6765	7455	3256.00	1405	6595	3428	7200	5370	11718	5788
Nw/Ns	15.89	15.85	15.01	14.79	12.50	12.09	16.72	16.65	15.21	15.39

Figure 3.8 The average number of words per sentence.

In terms of the number of candidate feature nouns per sentence, restaurant and television show the highest results followed by airline customer service, sports car, and laptop. Since the interest is explicitly mentioned features in sentences, the author hypothesized that at least 50% of the sentences have explicit features (either noun or noun phrase). The results explain that most products have at least one explicit feature in sentences to talk about, which supports the hypothesis. The statistics are demonstrated in Figure 3.9.

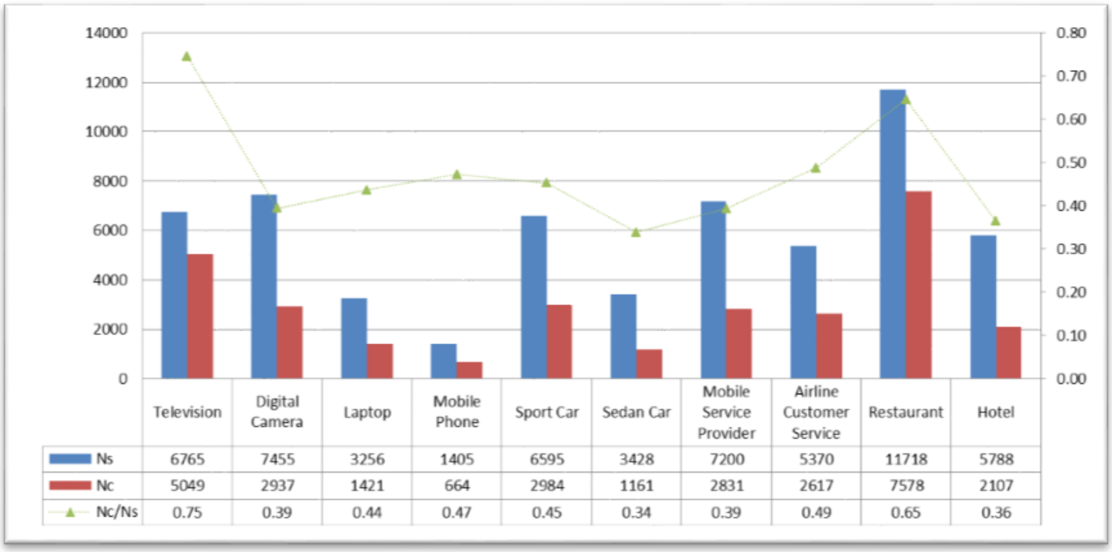
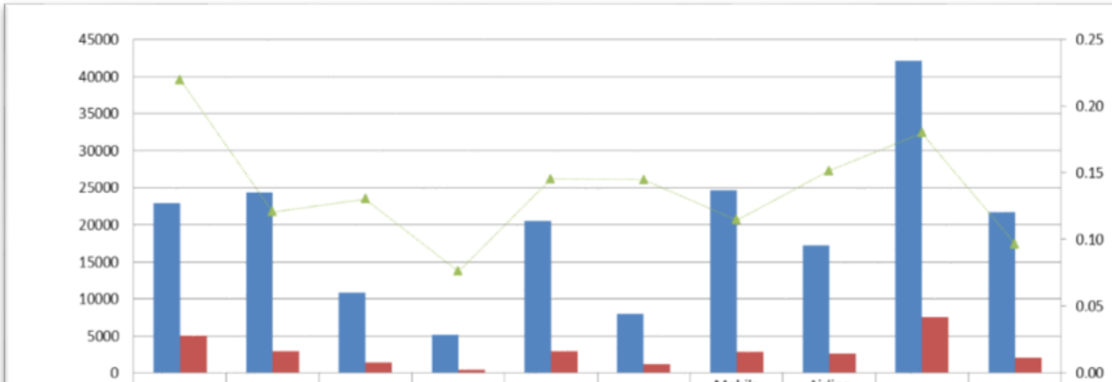


Figure 3.9 The number of candidate feature per sentence.

The average ratio of ten products of feature nouns to the total nouns in the review database (N_F/N_C) was 0.14, which means 14% of candidate features extracted by the tool were found as product features by the product development managers. In this dataset, television shows the highest number (22%), closely followed by restaurants (18%), while mobile phone shows the lowest number with 9%.



	Television	Digital Camera	Laptop	Mobile Phone	Sport Car	Sedan Car	Mobile Service Provider	Airline Customer Service	Restaurant	Hotel
Nc	22969	24320	10876	5125	20512	8014	24704	17269	42116	21768
Nf	5049	2937	1421	391	2984	1161	2831	2617	7578	2107
Nf/Nc	0.22	0.12	0.13	0.08	0.15	0.14	0.11	0.15	0.18	0.10

Figure 3.10 The ratio of features in the total candidate feature in review dataset.

In this dataset, the average DLIQ score for the ten products was 81.2 (STD 16.6). Television shows the highest number (107.1 with STD 2.7), closely followed by restaurant (106.9 with STD 1.4), while mobile phone showed the lowest number (54.1 with STD 18.1. This study shows that all of the scores range above the minimum threshold level (50 or above), which indicates that one move forward with the next data analyzing steps expecting some outcome. The DLIQ scores and their aspects are demonstrated in Table 3.3.

Table 3.3 Design-Level Information Quality Scores

PRODUCT TYPE	CONTENT				COMPLEXITY				RELEVANCY				DLIQ	Est. STD
	Weight	Base	Value	**I(Cont)**	Weight	Base	Value	**I(Cplx)**	Weight	Base	Value	**I(Relv)**	I(Cont) + I(Cplx) + I(Relv)	
television	0.30	153.0	140.0	**27.4**	0.30	2.1	2.1	**30.8**	0.40	6.8	8.3	**48.8**	**107.1**	2.7
digital camera	0.30	153.0	151.2	**29.7**	0.30	2.1	1.6	**23.2**	0.40	6.8	4.8	**28.2**	**81.0**	10.9
laptop	0.30	153.0	120.8	**23.7**	0.30	2.1	1.8	**25.2**	0.40	6.8	4.8	**27.9**	**76.8**	10.5
cell phone	0.30	153.0	112.9	**22.1**	0.30	2.1	0.5	**7.2**	0.40	6.8	4.2	**24.8**	**54.1**	18.5
sport car	0.30	153.0	144.0	**28.2**	0.30	2.1	1.8	**25.6**	0.40	6.8	5.6	**33.0**	**86.8**	10.1
sedan	0.30	153.0	136.9	**26.8**	0.30	2.1	0.4	**6.1**	0.40	6.8	5.1	**29.7**	**62.7**	7.8
mobile phone service provider	0.30	153.0	154.4	**30.3**	0.30	2.1	1.4	**19.8**	0.40	6.8	4.6	**26.8**	**76.9**	12.2
airline travel	0.30	153.0	141.4	**27.7**	0.30	2.1	1.5	**21.1**	0.40	6.8	5.7	**33.7**	**82.5**	11.3
restaurant	0.30	153.0	159.2	**31.2**	0.30	2.1	2.3	**32.5**	0.40	6.8	7.3	**43.1**	**106.9**	1.4
hotel	0.30	153.0	139.7	**27.4**	0.30	2.1	1.9	**27.6**	0.40	6.8	3.8	**22.6**	**77.5**	7.0
MIN				***22.1***				***6.1***				***22.6***	***54.1***	***1.4***
MAX				***31.2***				***32.5***				***48.8***	***107.1***	***18.5***
AVERAGE				***27.5***				***21.9***				***31.9***	***81.2***	***9.3***
STD				***2.8***				***9.0***				***8.2***	***16.6***	***4.9***

The observed result of DLIQ score indicates the collected data of ten products contains a promising level of information quality. This shows that there are enough product features in the reviews. Therefore, one can move forward with the next step to analyze positive/negative features and the reasoning behind the negative ones.

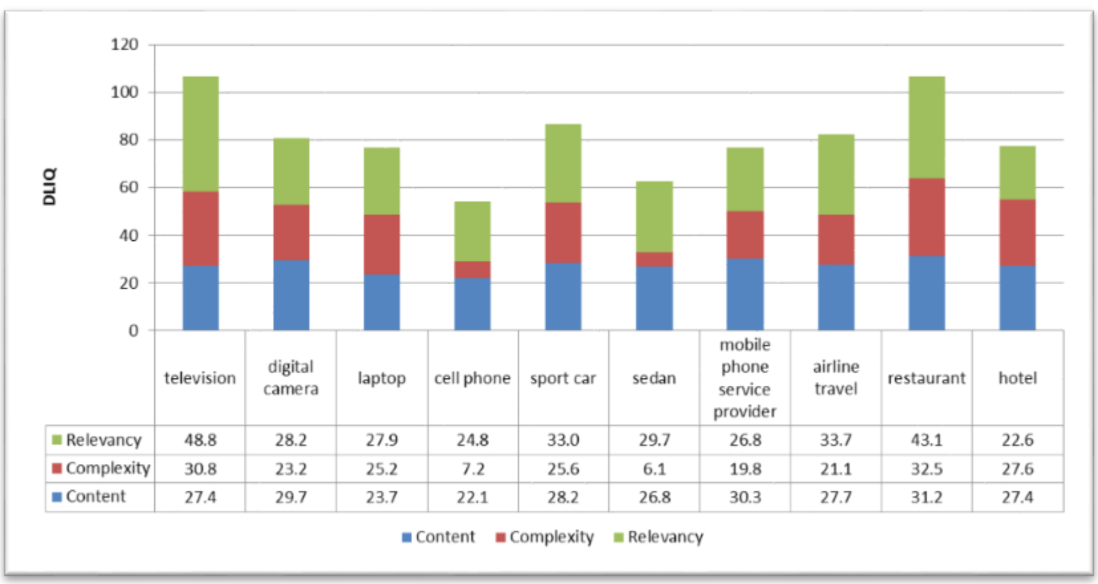


Figure 3.11 Design level information quality (DLIQ) aspect.

3.7 Hypothesis Test

As shown in the previous section, the investigative hypothesis is “Significant levels of quality design information can be efficiently extracted from Web reviews for a wide variety of classes (e.g., electronic, automobile, service domain), and this information can be directly related to product features of specific interest to the product design community.” This investigative hypothesis leads us to focus on two alternative hypotheses.

i. Hypotheses H1: Web reviews on products and services contain a significant level of information.

ii. Hypotheses H2: Web reviews on products and services contain a promising level for design-level information.

In hypothesis H1, the DLIQ score set is equal to or greater than the score of the significant level of information (score = 100). In this study, the interest was testing hypotheses at the 0.05 level of significance by performing a one-tailed t-test. The null and alternative hypotheses were defined as follows:

$$H1_{Null}: \mu < 100$$

$$H1_{Alternative}: \mu \geq 100$$

The test statistic was calculated using the following *t*-test equation:

$$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \tag{3.11}$$

The null hypothesis would be rejected if t ≥ tα; n, and if t < tα; n then the null hypothesis would be accepted. In using this formula, $\mu_0 = 100$. The numbers $\overline{x} = 81.2\ ,\sigma = 16.6, and\ n = 10$ came from the ten data sets. The value t_{α}; n was t = 0.05 ; 10 = 1.812. The test statistic was calculated as follows:

$$t = \frac{81.2 - \ 100}{16.6/\sqrt{10}} = -0.36$$

The obtained result shows that −0.36 < 1.812 (= $\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$ rejection criteria; t is t ≥ 1.812). Thus the calculated test statistic was not in the rejection region. Therefore, the null hypothesis was rejected in favor of the alternative. The conclusion was that the mean is significantly less than the score of 100. Thus, it is proven that the mean DLIQ was less than the score of a significant level of information (DLIQ 100).

In hypothesis H2, the author hypothesized the DLIQ score was equal to or greater than 70 thresholds, inclusive. In this case, the interest was testing the hypothesis at the 0.05 level of significance by performing a one-tailed *t*-test. The null and alternative hypotheses were defined as follows.

$$H2_{Null} = \ \mu\ < 70$$

$$H2_{Alternative} = \ \mu\ \geq 70$$

In using the *t*-test equation (3.11), the μ_0 would be equal 70. The numbers $\overline{x} = 81.2$, $\sigma = 16.6, and\ n = 10$ came from the ten data sets. The value t_α; n was t $_{0.05}$; 10 = 1.812. The test statistic was calculated as follows:

$$t = \frac{81.2 - 70}{16.6/\sqrt{10}} = 2.1$$

The obtained results showed that 2.1 > 1.812 (rejection criteria: t ≥ 1.812). Thus, the calculated test statistic was in the rejection region. Therefore, the null hypothesis was rejected in favor of the alternative hypothesis. It is concluded that the mean was above the score of 70. Thus, it has been proven that the mean of DLIQ was at a promising level for design-level information. In the next chapter, the opinion mining process is explained in detail.

CHAPTER 4

THE OPINION MINING PROCESS

This chapter is a review of the hybrid method of opinion mining for product design intelligence using association mining techniques. As discussed in Chapter 2, opinion mining has been studied extensively in recent years. In this area, three main research directions were explored (e.g., document level, sentence level, and feature-level opinion mining; (Hatzivassiloglou and McKeown 1997; Turney 2002; Hu and Liu 2004; Amir, Aumann et al. 2005). Opinion mining in three level granularities (i.e., document, sentence, and feature) was useful in several ways. For example, individual consumers want to know the opinions about a product from existing users before purchasing it. In marketing, it could help businesses judge the success of an ad campaign or new product launch. However, the majority of existing studies were not addressed much from a new product development perspective. Motivated by feature-level opinion mining, a hybrid method for opinion mining with the association rule mining approach was proposed to identify the correlation among product features, opinions, and further, the cause of negative opinions.

4.1 The Design-Feature-Opinion-Cause Method Architecture

For processing opinion mining and association rule mining tasks, RapidMiner software was used as the mining platform. It is a fully integrated platform for machine learning, data mining, and text mining (Appendix A). Figure 4.1 gives an overview of the proposed hybrid method, and each component in detail subsequently. As can be seen from Figure

4.1, the process includes multiple steps. In step one, the Web reviews for each product were collected and stored in separate folders, then each dataset was split into sentences and stored in the review database; after that, sentences were converted into computer readable format. In step two, product features were identified by utilizing opinion mining techniques. In step three and four, Web reviews were analyzed to identifying opinions about the features and causes of negative opinions by utilizing association mining techniques. Finally, the results of the findings were summarized. The inputs to the system were consumer reviews and the outputs were in the form of a comprehensive report, which includes: (i) the list of product features and list of feature and opinion pairs, and (ii) the list of features that negative comment on and reason pairs.

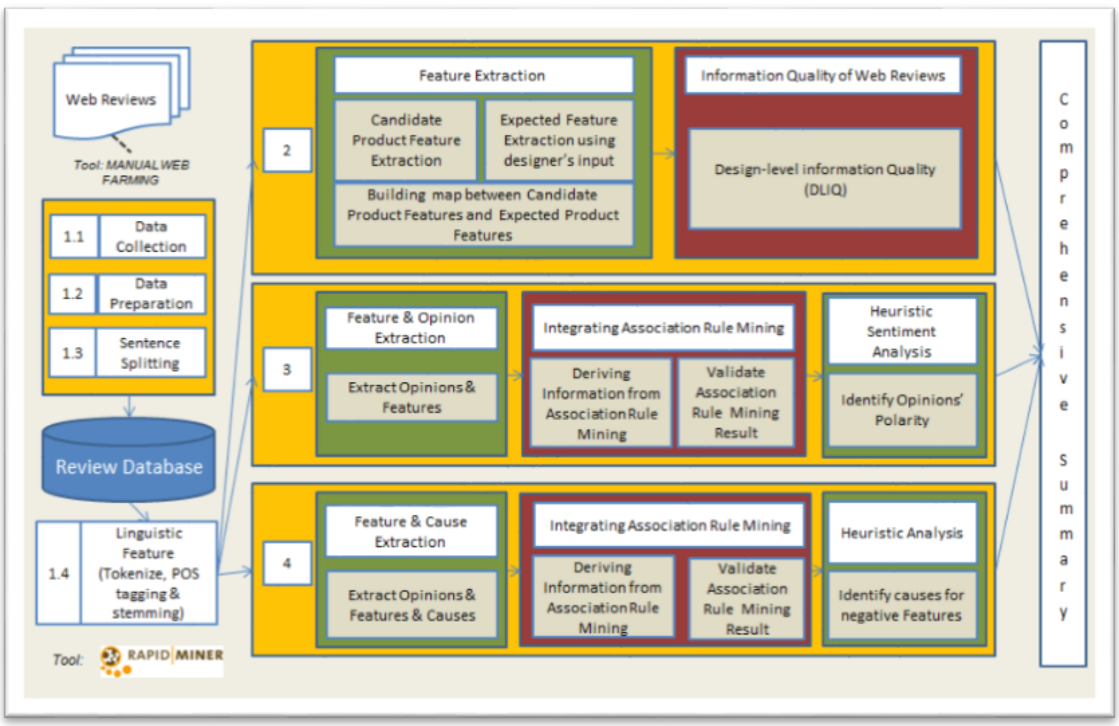


Figure 4.1 Architecture of the design feature opinion cause method. The input of the system was a set of Web reviews and the output of the system was a comprehensive summary of product development intelligence.

4.2 Computational Linguistic Text Processing

After data collection, data preparation, and the sentence splitting process described in Chapter 3, some computational linguistic text processing, also known as natural language processing (NLP), was applied to transform the text data (sentences) into a format that computers can recognize for opinion mining and sentiment analysis (Nasukawa and Yi 2003; Wahl, Winiwarter et al. 2010). Some important tasks of linguistic processing used in the study were lexical analysis, stop word removal, part of speech tagging, and stemming.

Lexical analysis was the process of converting each sentence into a set of words. As a result, each word in the sentence was represented by a single token (Nasukawa and Yi 2003; Guo, Zhu et al. 2009). The list of tokens becomes input for further processing such as feature extraction. For example, in this sentence from the digital camera dataset, "It takes sharp, accurately colored pictures," there were six tokens (e.g., it, takes, sharp, accurately, colored, pictures).

Subsequently, the stop words were filtered out to increase the computation time. Stop words may have little lexical meaning, or may not change the semantics of a sentence but instead serve to express grammatical relationships with other words within a sentence. Common English stop words were (e.g., the, is, at, of, and, to, a, in, which). For example, Kucera et al. (Kucera 1980), who have studied one million words of English text, have found the most common stop words represent approximately 10% of all word occurrence in text documents. Stop word removal was a common preprocessing step in linguistic analysis (Hu and Liu 2004; Jeong, Shin et al. 2011). In this study, to increase the computation time, and to improve the accuracy of extracting product features, the stop words were removed from the word list (Dave, Lawrence et al. 2003).

One special task of linguistic text processing was determining the part of speech of each word in a sentence, known as part of speech tagging (also known as word classes, morphological classes, or lexical tags). POS tagging was the process of assigning a part of speech such as a noun, verb, pronoun, preposition, adverb, adjective or other lexical class marker to each word (token) in a data set. POS tagging was very important to opinion mining analyses because each category has a specific role within a sentence because to extract nouns, adjectives, and verbs, a word's class should be known. Features were usually nouns or noun phrases in the reviews, while user opinions were usually adjectives. Therefore, POS tagging helped in extracting such information from reviews. The most common methods for part of speech tagging are rule-based tagging (Schmitz 2011), transformation base tagging (Wilson and Heywood 2005), stochastic tagging (Brants 2000), PENN Treebank POS Tagging (Schubert and Tong 2003; Liu 2004; Liu 2008; Luole and Li 2011). Among them, PENN Treebank POS tagging is a commonly used method in opinion mining. In this study, to identify product features, opinions, and causes, PENN Treebank POS tagging was used. Table 4.1 shows the common PENN Treebank POS tags. The following examples show some sentences tagged with their part of speech, "Sound/NN quality/NN of/IN Radio/NN spectacular/JJ even/RB with/IN the/EX top/NN down/NN." Each tag represent a POS label (e.g., NN, IN, JJ, RB, EX).

Table 4.1 A Sample of PENN Treebank POS Tags

Tag	Description	Tag	Description
NN	Noun, singular or mass	VB	Verb, base form
NNS	Noun, puller	VBD	Verb, past tense
JJ	Adjective	VBG	Verb, gerund or present participle
EX	Existential	IN	Preposition

Stemming was a process that reduced words by removing suffixes, thereby mapping them to the same root stem. A stemming algorithm was applied to improve word frequency, as words with a common stem tend to bear similar meanings. There were several approaches to stemming, including Lovins, snowball, and Porter stemming. Each has its own strengths and weaknesses. In this study, Porter stemming was applied, which was a very widely used and available stemmer, and was used in many applications (Porter 1980). Porter stemming is an iterative, rule-based replacement of word suffixes intending to reduce the length of the words until a minimum length is reached. For example, the stemming algorithm reduces the words fishing, fished, and fish to the root word fish; the words argue, argued, argues, arguing, and argus reduce to the stem argu, where argu represents argue.

After computational linguistic processing, three analyses were conducted, which were identification of product features (2.1), feature-opinion pairs (2.2), and feature-cause pairs (2.3).

4.3 Product Feature Identification

Feature identification was the process used to gather possible product features from the tagged texts generated by the computational linguistic feature process. A feature was a component of the product that has been commented on in reviews. For example, a particular brand of digital camera has a set of components (e.g., zoom, image quality, battery life). In this step, nouns and noun phrases were considered product features.

The feature extraction process has two steps: extracting candidate features and mapping those candidate features to product features, which product development managers identified. The objective of this process was to identify the product features on which reviewers tend to share their opinions. Locating product features from some sentences was not always possible because of the difficulty of natural language processing. Sometimes, product features were explicitly discussed and were implicitly revealed in the sentences (Ding, Liu et al. 2009). Here are some examples of product features from the reviews of an automobile from KBB.com, “The Bose audio system is amazing and it puts smiles on the face each time I am cruising.” In this sentence, the reviewer seems to be satisfied with the audio system and audio system was the feature on which the reviewer has opinions. In this review, the product feature appears explicitly in the sentence. In another example from KBB.com, “My only problem is the blind spot to the rear on both sides,” this reviewer seemed to talk about the visibility of the mirror, but the word mirror did not exist in the sentence. In this study, only product features that appear explicitly in the sentences were considered. Similar to the research by (Hu and Liu 2004; Popescu and Etzioni 2005), nouns and noun phrases were considered as product features on which reviewers have opinions.

First, based on a word’s POS tag, nouns (NN) and noun phrases (NNS) were identified as product features. Noun phrases (two sequential words) were identified by the n-gram model (e.g., 2-gram; (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009). An n-gram model is a type of probabilistic language model for predicting the next word conditioned on a sequence of previous words. In this study, the 2-gram (called bi-gram) approach was applied to identify noun phrases (e.g., battery life and picture quality).

As a customer review often contains many things that were not directly related to product features, gauging how important a feature was to a document was considered challenging. To overcome this, various statistical forms of the weight calculation have been applied to identify important features such as term frequency (TF), term presence, term frequency-inverse document frequency (TF-IDF). For this purpose, TF weight was used to determine how important product features were to reviewers, The TF is a numerical statistic, which is often used as a weighting factor in opinion and text mining (Pang and Lee 2008; Martineau and Finin 2009; Liu 2010).

Using TF to identify frequent nouns was reasonable, as frequent words were likely to be important product features to the reviewers whether they were negative or positive. The infrequent noun/noun phrases were likely to be less important product features to the reviewers. Each frequent noun or noun phrase in the outcome was a product feature candidate. However, not all candidates were frequent features generated by using TF weight. To remove those unlikely features, a pruning method was applied as a further drill down (Ding, Liu et al. 2009; Weishu, Zhiguo et al. 2010). Pruning specifies either too frequent or less frequent words that should be ignored from the list of the product features. In this study, a word was defined as frequent if it appeared in equal or more than five sentences without any maximum limitation. In addition to the pruning threshold, another constraint on features was placed to filter out unlikely features and increase the accuracy of feature extraction. In this study, a heuristic approach was employed to eliminate unlikely features.

In Web reviews, reviewers often refer the same product features by different words. It was necessary to group them together in order to efficiently analyze product features (Zhang, Jia et al. 2011). For example, image, image quality, picture, and picture quality all refer to the same feature in digital camera reviews and should be grouped together. Otherwise, it was too detailed for product development managers to read, summarize, and analyze all of these product features. In this study, after extracting nouns and noun phrases as features, they were grouped together based on the same, or a similar, meaning to increase term frequency (TF). Then, product development managers validated the list of candidate product features identified by the outcome of the text/data mining tool described above in order to map candidate product features to expected features, because not every candidate feature represented a product feature.

Table 4.2 shows the outcome as a features list (selected), which was identified by product development managers; this list shows the features that reviewers talked about and they have their opinions on, whether positive or negative. The features listed below were no longer candidates; they were the identified features, which were worthwhile for product development managers to investigate further to see if they can utilize the findings in their new product development process.

Table 4.2 Selected Features of Ten Products and Services

Television	3d Mode, Accessories, Auto motion, Battery, Bluetooth, Cabinet Color, Cable box (HD box), Camera, DVD Player, Energy efficiency, game mode, halo_effect, HDMD Receiver
Digital Camera	aperture, Auto flash/Flash, Auto Mode, Battery/Battery Life, Case, CMOS Sensor, exposure, Handheld, HDMI Cable, HDR, Image stabilization, ISO, landscape
Laptop	Adaptor, Apple's Backup Strategy, Battery/Battery Life, Camera, CD/DVD Driver, Command Key, Hard Drive, I/O Ports (Input Output), Internet, Keyboard, Laptop case, Laptop Color, Laptop Size/weight
Mobile Phone	Accessories, Battery, Bluetooth, Camera, Internet Access, Keyboard, Memory size, Operating System, Research in Motion, Screen, Sim Card, Software, Text message
Sport Car	4WD/AWD, AC/Heater, Accelerator Pedal, Air Filter, Alternator/Battery, Audio, Body Color, Brakes, Bumpers, convertible top, cruise control, Cup Holder, Design
Sedan Car	Alternator/Starter, Audio, Battery, Body, Brake, Car price, cruise control, door handle, engine, engine light, Exhaust system, Filter, Fuel Consumption
Mobile Phone Provider	Billing, Cell Phone upgrades, Contract/Contract Termination, coverage, Customer Service, Data Plan, Discount, Insurance, Internet & email, Language Options, Mobile hotspot, Password, Pay phone
Airline Cust. Service	Airport Security/Facilities, Arrival/Departure, Baggage Check-in/Claim, Baggage fee, Bathroom, Boarding, Children/Infants, Connectivity/Transfer, Credit card, Customer Service, disability Access, Economy/Business Class, Flight Attendant
Restaurant	Ambiance/Décor, Appetizers, Asian Food, Asparagus, Bacon, Baguette/Bread, bean, beansprout, brussel sprout, butter, carrot, cauliflower, Cheese
Hotel	Amenities, Bathroom, bed/bedding, Bellman, coffee machine, Customer Service, Facilities, fitting center, Front Desk, Gift Shop, HotelRoom, House keeping, Internet

Next, the product feature and opinion identification process are introduced.

4.4 Product Feature-Opinion Pair Identification

Feature-opinion pair identification was the process to identify associations between product features and the opinions on them from Web reviews. When a feature and its opinion occur in one sentence, they were called a feature-opinion pair. For example, "photos" as a feature and "very good" as its opinion constituted a feature-opinion pair. Unlike the feature extraction process explained above, product feature-opinion pair identification not only considers the noun/noun phrase but also the adjectives as opinion (POS: NN, NNS, JJ, JJR, JJS). The noun-adjective pair was based on the assumption that people use adjectives to evaluate an item/product, and noun and noun phrases represent the features most accurately (Hatzivassiloglou and Wiebe 2000; Turney 2002; Hu and Liu 2004). Based on this assumption, first, the method extracted both nouns and adjectives together from the sentences. Then, the adjectives with similar meanings were normalized using WordNet, a lexical database for English.

Then, associations between feature and opinion were identified by applying the association rule mining technique to determine if and how much features and opinions were related to each other. There were two important basic measures for association rules, support, and confidence. Since the database was large and user concerns were about only those frequently purchased items, usually thresholds of support and confidence were predefined by users to drop those rules that were not of interest or useful. The two thresholds were called minimum support and minimum confidence level respectively. The threshold level to evaluate if a rule was significant was defined based on the confidence of the rule. In this study, to increase the number of association rules between features and opinions, the minimum confidence level value was set to a lower bound (e.g., 0.1 of 1).

For example, in the sports car dataset, 32,976 rules were generated by the algorithm because of the association rule process. The association rule considered the probability that not every pair contained meaningful results. Of this (32,976 rules), 42 interesting association rules were identified, as a result of the pruning process that removed irrelevant opinions for each feature and left only relevance ones for sentient analysis. However, the author presented all interesting rules in details in the next chapter. Below each result, utilizing an example from the sports car database was illustrated.

Table 4.3 shows the result of the association rule for selected two features after applying the algorithm and pruning process, which were seats and interior design. Four interesting rules for seats and two interesting rules for interior design were identified.

Table 4.3 Result of the Association Rule After the Pruning Process

Rules	Feature (Conclusion)	Opinion (Premise)	Support	Confidence
1	seats	uncomfortable	0.0011	0.82
2	seats	not, comfortable	0.0012	0.42
3	seats	not, good	0.0012	0.11
4	seats	comfortable	0.0036	0.32
5	interior design	cheap	0.0015	0.27
6	interior design	not, great	0.0011	0.15

Then, interesting feature-opinion pairs were identified. However, the opinion's polarity of each feature related with each rule was not determined. This step was referred to as a sentiment analysis and was intended to identify opinion polarity on each pair. Unfortunately, the exact algorithm to identify semantic orientation of an adjective for each feature-opinion pair does not exist. Thus, intuition and domain knowledge were required in identifying the right opinion polarity of the features.

To overcome this constraint, manual examining and labeling were done whether each of the adjectives was expressing positive or negative sentiment. For example, in the sports car dataset, the adjectives comfortable and good have positive orientation, while the adjectives uncomfortable, cheap, not, and great show a negative orientation. Often, the opinion information in a sentence was expressed with negative terms such as 'not,' and 'no.' In this case, the orientation of the opinion about the feature was the opposite of the meaning of the corresponding opinion phrase. For example, the opinion "not, comfortable" was considered a negative opinion such as "uncomfortable." Table 4.4 shows the opinion polarity after assigning the label as negative or positive.

Table 4.4 Result of the Association Rule After Assigning Opinion Polarity Label

Rules	Feature (Conclusion)	Opinion (Premise)	Support	Confidence	Opinion Polarity
1	seats	uncomfortable	0.0011	0.82	Negative
2	seats	not, comfort	0.0012	0.42	Negative
3	seats	not, good	0.0012	0.11	Negative
4	seats	comfort	0.0036	0.32	Positive
5	interior design	cheap	0.0015	0.27	Negative
6	interior design	not, great	0.0011	0.15	Negative

After classifying sentiment into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values regarding its feature. For instance, the feature seats shows the sum of all of the positive support values was 0.0036, and the sum of all of the negative values was 0.0034 (0.0011 + 0.0012 + 0.0012) while the feature interior design shows the sum of all of the positive support values was 0, and the sum of all of the negative support values was 0.0026. These two values were utilized to identify overall opinion or impression of product features.

Table 4.5 Aggregated Positive and Negative Support Values

Feature (Conclusion)	Sum of Positive Support	Sum of Negative Support
seats	0.0036	0.0034
interior design	0	0.0026

To identify the overall opinion on features, the opinion polarity (OP) score of features was determined by calculating the difference between the sum of all positive support values and the sum of all negative support values. If the result was positive, then the opinion on the feature was positive, and if the result was negative, then the opinion on the feature was negative. Therefore, it can be assumed that the OP equation was an indicator of whether the consumers feel positive or negative on features. The positive opinion of the feature shows the advantages of the product, and the negative opinion of the feature contains the disadvantages of the product. The equation was developed to identify opinion polarity score is as follows:

$$\text{OP (feature)} = \text{(the sum of positive supports)} - \text{(the sum of negative supports)} \tag{4.1}$$

For example, shown in the table below, opinion polarity score was 0.0001 for seats, then the overall opinion on seats was positive. Alternatively, the opinion polarity score for interior design was −0.0026, making the overall opinion on interior design negative.

Table 4.6 Opinion Polarity Score and Overall Opinion

Feature (Conclusion)	Sum of Positive Support	Sum of Negative Support	Opinion Polarity (OP) Score	Overall Opinion
seats	0.0036	0.0034	0.0001	Positive
interior design	0	0.0026	−0.0026	Negative

Next, the feature-cause pair identification for negative features are introduced.

4.5 Feature - Cause Identification

Feature-cause pair identification was the process to identify associations between product features and the reasons for them from Web reviews. Especially, this study was focused to analyzing the distinctive negative features identified in the feature-opinion identification step: if there were no distinctive negative features, the features that have positive and negative aspects were analyzed to understand what caused a negative attitude toward these features. In this step, verbs were introduced as opinion reasoning along with adjectives and nouns. It was assumed that verbs were considered the core of the sentence, and their meanings were key to understand the meaning of the sentence. For example, here was a review saying, “The cheap plastic on the doors, scratch very easily.” The reviewer had a negative experience about interior design and then he explained the reason (e.g., because a cheap material was used, the door was scratched easily).

To identity the feature cause, first, the method extracted nouns, adjectives, and verbs together from the review database. Second, the verbs were replaced with similar meaning using WordNet and freely available online dictionaries in order to transform them to a base form and to group similar verbs. Third, associations between feature and cause were identified by applying the association rule mining technique to each feature that was identified as a negative feature during the previous step (i.e., the feature-opinion identification step). The feature-cause identification step was an exploration of whether the cause of the features with negative opinions can be identified.

Again, the threshold level to evaluate if a rule was significant was defined based on the confidence of the rule. In this step, to increase the number of association rules between features and cause, each feature was processed individually, and the minimum confidence level value was set to the lower bound (e.g., 0.1 of 1).

In this step, analysis was performed on partitioned data, which filters feature with negative opinions among many identified, then the association rule was run based on these reviews only to capture more feature-cause pairs. It was assumed that unlike identifying opinion polarity in features, capturing feature-cause identification from Web reviews involves more diversity. In other words, the cause of a negative opinion for a feature could vary by personal situation. For example, some people can live in extreme cold weather condition, so they might have negative comments about air conditioning systems. Alternatively, some people live in extremely hot weather, so they might have negative comments concerning an air conditioning system. Even though both of them have negative opinions about air conditioning systems, their causes were different, reflecting different personal situations.

For example, in the sports car dataset, 8221 rules were generated by the algorithm because of the association rule process for the feature - interior design. Of 8221 rules, four interesting association rules were identified because of the pruning process that removes irrelevant opinions for each feature and leaves only relevant ones for sentient analysis. All of the interesting rules are presented in the next chapter in detail. Each result utilizing an example from the sports car database is illustrated.

Table 4.7 shows the results of the association rule for the feature interior design after applying the algorithm and pruning process. Four interesting rules were identified for interior design that indicate cause for negative opinions.

Table 4.7 Result of the Association Rule After Assigning Opinion Polarity Label

Rules	Feature (Conclusion)	Cause (Premises)	Support	Confidence
1	interior design	plastic, cheap	0.0309	0.56
2	interior design	plastic, scratch	0.0137	0.57
3	interior design	weak, plastic	0.0172	0.71
4	interior design	look, cheap	0.0103	0.75
5	interior design	easy, scratch	0.0137	1.00

Then the association rule results were presented, which could be used in developing a new product or modifying an existing product. Regarding interior design, the overall impression was that it can scratch easily, and looks somewhat cheap because of the use of low-quality plastics.

CHAPTER 5

WEB REVIEW DESIGN-FEATURE OPINION CAUSE ANALYSIS - SINGLE PRODUCT

In this chapter, the experimental details on a single product are presented. To present the method and findings on a single product as a walkthrough example, the sports car dataset was chosen from the ten products for the following reasons: (1) the automobile industry was the most lucrative industry, (2) the behavior of consumers played a vital role in creating effects on the purchase of automobiles, which lead to continual modification of car models and its features, (3) the competition has increased in the sector with a host of new players, (4) the sales in the sports car sector increases.

This chapter is organized as follows. Section 5.1 includes a definition of product selection; the second section (5.2) is a description of data collection and preparation; the third section (5.3) of this chapter is a presentation of the product feature extraction process; (5.4) includes a feature opinion sentiment evaluation; (5.5) comprises a feature opinion cause analysis; a Web review DFOC summary is in (5.6); and the last section of the chapter (5.7) is a representation of statistical validation of the method.

5.1 Product Selection

As discussed before, to make the experiment and results more concrete, the ten products were selected for this research from different industries primarily based on the frequency of appearance in the literature reviews and the richness of contents from the Web reviews, one of which was the sports car dataset presented in this chapter. Before further discussion, the following notations used in this research are introduced.

N_R total number of unique Web reviews in the database,

N_S total number of sentences generating by splitting all reviews in the database,

N_W total number of words authored in all reviews in the database,

N_C total number of noun words identified by the data mining process from all reviews,

N_F total number of noun words matched by the data mining process to designer listed features,

N_{FO} number of distinct product feature identified during feature-opinion extraction analysis, and

N_{FE} number of distinct features identified during feature-cause extraction analysis.

5.2 Data Collection and Preparation

Of sports car reviews, 671 were collected from two sources, namely CarReview.com and KBB.com, to conduct the study to increase the number of reviews to coherent analysis. Below is a sample review from CarReview.com. A typical review has the review date, reviewer name, numerical rating (or stars), and the body of the review (such as a summary), where they share their experiences about a product; some have more information (e.g., strengths, weakness, and price; Figure 5.1). The review body where a consumer shares opinions was of interest, while the remaining parts were called noisy data, which was excluded from this study.

Figure 5.1 A sample review from CarReview.com.

To extract the review body, the review was split based on the user-defined keyword these were placed into separate files; the text between the start keyword and the end keyword, both exclusive, was treated as the body of the review (Khan, Baharudin et al. 2009). In sports car, the start keyword was “summary,” and end keyword was “similar product used.” Table 5.1 shows the body of the review as extracted from the review shown in Figure 5.1. In this research, each body of review was considered a review.

Table 5.1 Sample of Body of Review

Great Car and soooo much fun. Everyone comments on the color and how sporty it looks. I have the pearl red with the black top. You just melt in the front seats. Sound quality of radio spectacular even with the top down. Heater works great for those cold evening drives topless. I absolutely say buy this car if you want a zippy fun sexy car. My friend was in awe of the comfort and how easily it handled. He has owned porches and believes this car is funner to drive!

Further on, the body of reviews was split into sentences by defining the splitting point. The text was split into sentences to achieve a finer granularity because the reviews may contain several features, each of which may represent different features on which consumers comment. In sports car, the body of the reviews was split based on punctuation characters (question mark and period) for further analysis. Table 5.2 shows the reviews after the sentence splitting process; the body of the review shown in Table 5.1 becomes nine sentences. As discussed in Chapter 3, the splitting process was necessary because compound reviews may contain several features, each of which may represent different opinion. After sentence splitting, 671 reviews (N_R) become 6595 sentences (N_S). Now it can be assumed that each sentence contains opinions about, at least, a single feature (e.g. "sound quality" in sentence number five, or "heater" in sentence number 6).

Table 5.2 Example of the Sentence Splitting Process

Num (i)	Sentences (S_i)
1	Great Car and soooo much fun.
2	Everyone comments on the color and how sporty it looks.
3	I have the pearl red with the black top.
4	You just melt in the front seats.
5	Sound quality of radio spectacular even with the top down.
6	Heater works great for those cold evening drives topless.
7	I absolutely say buy this car if you want a zippy fun sexy car.
8	My friend was in awe of the comfort and how easily it handled.
9	He has owned Porches and believes this car is funner to drive!

After completing the sentence splitting task, three analyses were processed utilizing the association rule mining algorithm discussed in Chapter 4: (a) identifying product features, (b) feature-opinion pairs, and (c) feature-cause pairs.

5.3 Product Feature Extraction

As discussed previously in Chapter 3, recent research has shown that nouns and noun phrases represent the product features most accurately (Hatzivassiloglou and Wiebe 2000; Turney 2002; Hu and Liu 2004). Identifying such nouns and noun phrases was very challenging but critical for effective opinion mining in many domains. In addition, features were more likely to be discussed by the consumer, which suggests that features should be frequent nouns or noun phrases. The same assumption was made in this study. First, nouns and noun phrases were extracted from the review data as candidate features. However, not all of the frequent nouns were product features. Then, candidate features were matched with product features that the designer identified. Finally, they were grouped together based on the same meaning to increase term frequency.

To extract candidate features, some linguistic feature tasks were applied to transform the text data into a format that the computer could recognize for the opinion/data mining process. First, POS tagging was performed on the collection of sentences. This task generated the POS tag of each word. For example, the sentence, "Sound quality of radio spectacular even with the top down" was tagged as, "Sound/NN quality/NN of/IN Radio/NN spectacular/JJ even/RB with/IN the/EX top/NN down/NN," where NN indicated a noun, VB a verb, JJ an adjective, IN a preposition, and EX an existential.

After determining the POS tag of each word, stop word removal, and Porter stemming (Nasukawa and Yi 2003; Wahl, Winiwarter et al. 2010) were applied to increase the accuracy of the search information and the overall effectiveness of the process. For example, removing some of the most common words from the text such as “a” or “the” improved the processing time for computation because the system processed fewer words; stemming reduced derived words from the original meaning such as brake and brakes; this was for improving frequency (e.g. five reviewers say the brake of the car and 50 plus reviews mention brakes. By stemming, the term frequency can be increased to 55). To extract noun phrases (two consecutive words), the N-gram method was used, specifically the bi-gram method was applied along with POS tagging (e.g. sound quality, head gasket, door handle; (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009). Then the frequency of nouns and noun phrases were filtered based on the POS tag (NN, NNS) by using term frequency weight (TF) discussed in Chapter 4. Pruning was used to filter less frequent nouns that appear in the review collection. Pruning specified either too frequent or less frequent words that should be ignored from the list of the product features. In this study, a noun was defined as frequent if it appeared in five or more sentences without any maximum limitation.

Table 5.3 shows the frequent nouns and noun phrases for sports car. As was shown, there were some frequent nouns, which were not real features. Non-features were distinguished with parentheses.

Table 5.3 Example of Real Feature and non-Features

Real Feature	Non-Feature
wiper	(job)
water pump	(lemon)
power window	(center)
wheel	(owner)
windshield	(classic)
tire	(version)
chrome rim	(smoke)

Since not every candidate feature was a real product feature where ($F \in C$), to filter out non-features and increase the accuracy of feature extraction, an another constraints were implemented. In this study, a knowledge-based heuristic approach was employed; product development managers validated the list of candidate features and marked each candidate feature as a real feature or non-feature.

In addition, different reviewers often referred to the same product features by different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011) and further, similar features were grouped together based on the same meaning to increase term frequency. For example, the features HP, turbo-spool, VTEC, torque, RPM, motor, horsepower, engine, and twin-turbo refer to “engine power.”

Table 5.4 shows 42 distinct product features identified by the product development managers and their corresponding frequencies, which are in descending order. As shown, high frequency indicated reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious that engine power was the feature people talked about most when it came to the sports car, followed by fuel consumption, audio, and the transmission system.

Table 5.4 Distinct Product Features for Sports Car

Num (i)	Feature (F_i – distinct)	TF (F_i- freq)	Num (i)	Feature (F_i – distinct)	TF (F_i- freq)
1	engine power	618	22	design	31
2	fuel consumption	364	23	engine/valve	27
3	audio	258	24	exhaust	23
4	seats	225	25	AC/heater	22
5	transmission systems	147	26	horse power	18
6	wheels and tires	142	27	suspension	18
7	brakes	104	28	leg room	17
8	interior design	91	29	four-wheel drive	16
9	mirror/visibility	87	30	head gasket	13
10	door handle	81	31	oil change	13
11	sales	69	32	bumpers	12
12	windshield/windows	67	33	spoiler	12
13	body color	56	34	engine light	11
14	4WD/AWD	54	35	convertible top	10
15	sun roof/roof	52	36	cruise control	9
16	trunk space	52	37	turning radius	9
17	warranty	51	38	water pump	9
18	alternator/battery	50	39	air filter	7
19	lights	48	40	wiper	7
20	accelerator pedal	40	41	cup holder	5
21	seat belt	34	42	size	5

5.4 Product Feature - Opinion Extraction

During this step, which was feature-opinion identification, the association rule mining approach was applied to identify the correlation between product features and related opinions, where a noun was a feature, and an adjective was an opinion.

To extract nouns and adjectives from the reviews, as explained in the previous section, some linguistic feature tasks such as natural language processing were applied to transform the textual data into a format that a computer could recognize for the opinion mining process such as POS tagging, stop word removing, and stemming (Nasukawa and

Yi 2003; Wahl, Winiwarter et al. 2010). Afterward, POS tagging, stop word removal, and Porter stemming were applied to increase the accuracy of the search information and the overall effectiveness of the process.

After determining the POS tag of each word, the method found the nouns, noun phrases, and adjectives, and only kept nouns and adjectives. To increase term frequency, the synonyms of adjectives were used by utilizing WordNet (Miller 2012). Then, the nouns and noun phrases were replaced using a distinct product feature list identified in the previous section. Table 5.5 shows the extracted features replaced with feature groups. For example, the system searched for and replaced manual transmission with transmission systems based on the matching list.

Table 5.5 Example of Matching List of Feature Groups

Feature Group (F_i – distinct)	Extracted Features (F_i)
transmission systems	manual transmission
transmission systems	stick shift
transmission systems	transmission
serpentine belt	serpentine belt
fuel consumption	fuel mileage
fuel consumption	mile gallon
cruise control	cruise control
transmission systems	clutch
fuel consumption	average MPG
fuel consumption	gas mileage

Then, the association rule mining approach was applied to identify correlations between product features and related opinions. The association rule is one of the most widely used data mining concepts. The goal of an association rule mining algorithm in this study was to discover associations between seemingly unrelated frequent features and opinions in the Web reviews discussed in Chapter 4.

An association rule was illustrated in this example: Engine-power → amazing (support = 0.22%, confidence = 17%). This rule says that 0.22% of customers spoke about engine power and amazing together and those who spoke about engine power, also spoke about amazing 17% of the time. Support and confidence were two important measurements in association rule mining (Agrawal, Imieli et al. 1993). In this work, the minimum confidence (minconf) was set at 10% to observe a larger number of rules. The association rule was discussed in detail in Chapter 4.

Now, the result of the association rule approach is presented. For sports car, overall 32,976 rules were found because of the association rule process. A straightforward visualization of the association rule was to use a scatter plot with two interesting measures on the axes such as confidence and support. It was shown that rules with high confidence had relatively low support.

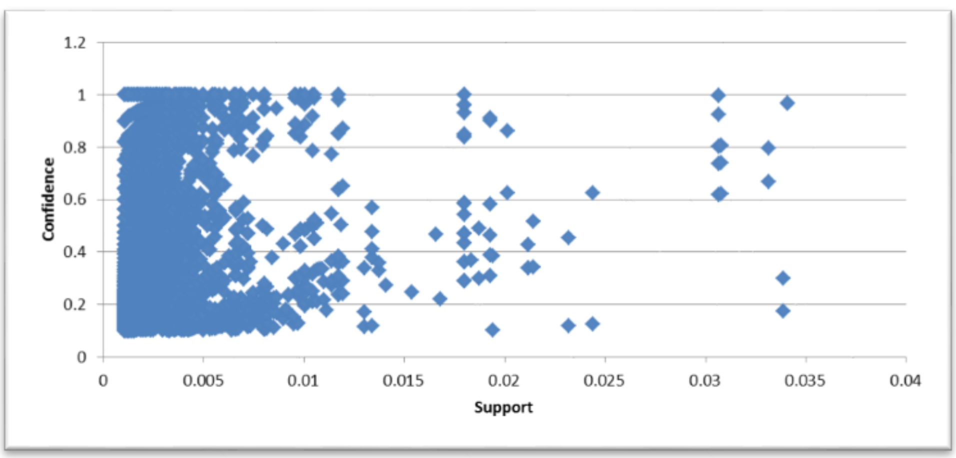


Figure 5.2 Scatter plot for 32,976 association sports car rules.

Even if 32,976 rules were found by the association rule approach, not all of the rules were worthwhile to analyze, meaning that only real product features and related opinions about them were analyzed in this study. For example, rule (#1) “engine power

→ time” and (#6) “engine power → sales” were not interesting rules to analyze. Alternatively, (#3) “engine power → problem” was an interesting rule to analyze in this study.

Table 5.6 Example of the Association Rules for Sports Car

Rules	Feature (Conclusion)	Opinion (Premises)	Support	Confidence
1	engine power	time	0.0017	0.10
2	engine power	reliable	0.0012	0.11
3	engine power	problem	0.0036	0.11
4	engine power	thing	0.0014	0.11
5	engine power	rear	0.0013	0.12
6	engine power	sales	0.0012	0.12
7	engine power	good	0.0063	0.12

Hence, among the 32,976 rules, only 42 trivial/useful rules were given as the results of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). The nontrivial rules were ignored. As listed in Table 5.7, 42 rules were identified and they were categorized to eight distinct product features: audio, door handle, engine power, exhaust, fuel consumption, interior design, mirror/visibility, and seats.

Table 5.7 Selected Association Rules for Sports Car

Rule	Feature (Conclusion)	Opinion (Premises)	Support	Confidence
1	audio	decent	0.0011	0.28
2	audio	easy	0.0011	0.13
3	audio	quality	0.0012	0.20
4	audio	nice	0.0033	0.16
5	audio	amazing	0.0034	0.26
6	audio	great	0.0053	0.10
7	door handle	good strength	0.0011	0.13
8	engine power	not high	0.0011	0.47
9	engine power	not great	0.0015	0.21
10	engine power	noisy	0.0017	0.39
11	engine power	not good	0.0018	0.17
12	engine power	bad	0.0018	0.14
13	engine power	problem	0.0036	0.11
14	engine power	weak	0.0081	0.20
15	engine power	not bad	0.0012	0.23
16	engine power	reliable	0.0012	0.11
17	engine power	strong	0.0018	0.52
18	engine power	quick	0.0020	0.29
19	engine power	nice	0.0021	0.10
20	engine power	amazing	0.0022	0.17
21	engine power	good	0.0063	0.12
22	engine power	great	0.0076	0.15
23	exhaust	intake	0.0013	0.39
24	fuel consumption	not great	0.0018	0.24
25	fuel consumption	high	0.0020	0.22
26	fuel consumption	bad	0.0021	0.17
27	fuel consumption	not good	0.0030	0.28
28	fuel consumption	problem	0.0046	0.14
29	fuel consumption	decent	0.0011	0.28
30	fuel consumption	economic	0.0014	0.86
31	fuel consumption	amazing	0.0017	0.13
32	fuel consumption	low	0.0018	0.22
33	fuel consumption	nice	0.0022	0.11
34	fuel consumption	great	0.0073	0.14
35	fuel consumption	good	0.0091	0.18
36	interior design	not great	0.0011	0.15
37	interior design	cheap	0.0015	0.27
38	mirror/visibility	poor	0.0011	0.21
39	seats	uncomfortable	0.0011	0.82
40	seats	not comfortable	0.0012	0.42
41	seats	not good	0.0012	0.11
42	seats	comfortable	0.0036	0.32

Association rules given in Table 5.7 were explained to demonstrate how the rules in this table should be described.

Rules 8 to 22 contain opinions concerning engine power. Some of the rules seem to have mixed opinions whether they were positive (not bad, reliable, strong, quick, nice, amazing, good, great), with support of (0.0012, 0.0012, 0.0018, 0.0020, 0.0021, 0.0022, 0.0063, 0.0076) respectively, or negative (not high, not great, noise, not good, bad, problem, weak), with support of (0.0011, 0.0015, 0.0017, 0.0018, 0.0018, 0.0036, 0.0081) respectively. The confidences of the rules that correspond to the opinions above were (0.23, 0.11, 0.52, 0.29, 0.10, 0.17, 0.12, 0.15) respectively for positive and (00.47, 0.21, 0.39, 0.17, 0.14, 0.11, 0.20) respectively for negative.

Rules 24 to 35 contain opinions about fuel consumption. Some of the rules seem to have mixed opinions whether they were positive (decent, economy, amazing, low, nice, great, good), with support of (0.0011, 0.0014, 0.0017, 0.0018, 0.0022, 0.0073, 0.0091) respectively or negative (not great, high, bad, not good, problem) with support of (0.0018, 0.0020, 0.0021, 0.0030, 0.0046) respectively. The confidences of the rules that correspond to the opinions above were (0.28, 0.86, 0.13, 0.22, 0.11, 0.14, 0.18) for positive, and (0.24, 0.22, 0.17, 0.28, 0.14) for negative.

Rules 36 and 37 contain opinions about interior design. The rules had negative opinions (not great, cheap), with support of (0.0011, 0.0015). The confidences of the rules that correspond to the opinions above were (0.15, 0.27) for negative.

Rules 39 to 42 contained opinions concerning seats. Some of the rules seemed to have mixed opinions whether they were positive (comfortable), with support of (0.0036) or negative (uncomfortable, not comfortable, not good), with support of (0.0011, 0.0012, 0.0012). The confidences of the rules that correspond to the opinions above were (0.23, 0.11, 0.52, 0.29, 0.10, 0.17, 0.12, 0.15) for positive and (00.47, 0.21, 0.39, 0.17, 0.14, 0.11, 0.20) for negative.

Now, 42 interesting feature-opinion pairs were identified but it was not certain that the opinion's polarity of each feature was related with each rule. Opinion polarity for each feature-opinion pair was identified by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

For example, for the feature seat, the adjective (comfortable) had a positive orientation, so they labeled it as positive, while the adjective (uncomfortable) showed a negative orientation. Therefore, each association was examined and labeled as positive and negative. The following table shows the opinion polarity after assigning the label as negative or positive.

Table 5.8 Result of the Association Rule after Assigning Opinion Polarity Label

Rule	Feature (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
1	audio	decent	0.0011	0.28	Positive
2	audio	easy	0.0011	0.13	Positive
3	audio	quality	0.0012	0.20	Positive
4	audio	nice	0.0033	0.16	Positive
5	audio	amazing	0.0034	0.26	Positive
6	audio	great	0.0053	0.10	Positive
7	door handle	good strength	0.0011	0.13	Positive
8	engine power	not high	0.0011	0.47	Negative
9	engine power	not great	0.0015	0.21	Negative
10	engine power	noisy	0.0017	0.39	Negative
11	engine power	not good	0.0018	0.17	Negative
12	engine power	bad	0.0018	0.14	Negative
13	engine power	problem	0.0036	0.11	Negative
14	engine power	weak	0.0081	0.20	Negative

Table 5.8 Result of the Association Rule after Assigning Opinion Polarity Label

Rule	Feature (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
15	engine power	not bad	0.0012	0.23	Positive
16	engine power	reliable	0.0012	0.11	Positive
17	engine power	strong	0.0018	0.52	Positive
18	engine power	quick	0.002	0.29	Positive
19	engine power	nice	0.0021	0.1	Positive
20	engine power	amazing	0.0022	0.17	Positive
21	engine power	good	0.0063	0.12	Positive
22	engine power	great	0.0076	0.15	Positive
23	exhaust	intake	0.0013	0.39	Negative
24	fuel consumption	not great	0.0018	0.24	Negative
25	fuel consumption	high	0.002	0.22	Negative
26	fuel consumption	bad	0.0021	0.17	Negative
27	fuel consumption	not good	0.003	0.28	Negative
28	fuel consumption	problem	0.0046	0.14	Negative
29	fuel consumption	decent	0.0011	0.28	Positive
30	fuel consumption	economic	0.0014	0.86	Positive
31	fuel consumption	amazing	0.0017	0.13	Positive
32	fuel consumption	low	0.0018	0.22	Positive
33	fuel consumption	nice	0.0022	0.11	Positive
34	fuel consumption	great	0.0073	0.14	Positive
35	fuel consumption	good	0.0091	0.18	Positive
36	interior design	not great	0.0011	0.15	Negative
37	interior design	cheap	0.0015	0.27	Negative
38	mirror/visibility	poor	0.0011	0.21	Negative
39	seats	uncomfortable	0.0011	0.82	Negative
40	seats	not comfortable	0.0012	0.42	Negative
41	seats	not good	0.0012	0.11	Negative
42	seats	comfortable	0.0036	0.32	Positive

After classifying the sentiments of adjectives into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values regarding its feature using the opinion polarity score equation discussed in Chapter 4.

$$OP\ (Fi) = \sum(Support - \ positive) - \sum(Support - negative) \tag{5.1}$$

For instance, for the feature seats, the sum of all of the positive support values was 0.0036 and the sum of all of the negative values was 0.0034 (0.0011 + 0.0012 + 0.0012). These two values were used to identify the overall opinion or impression of the product features. In the feature seat, the overall opinion polarity score was calculated as 0.000 (positive). The opinion polarity score of each feature was determined by calculating the difference between the sum of all positive support values and the sum of all negative support values. If the result was positive, then the opinion on the feature was positive, and if the result was negative, then the opinion on the feature was negative. Therefore, it can be assumed the OP equation was an indicator of whether the consumers feel positive or negative about the features. Table 5.4 shows the summary of the 42 rules for eight features and demonstrates the feature-wise frequencies and final opinion polarity.

Table 5.9 Aggregated Opinion Polarities on Features for Sports Car

Num	Feature (Conclusion)	Sum of Positive Opinions	Sum of Negative Opinions	Opinion Polarity Score (+/-)	Opinion Polarity
1	audio	0.0154		0.0154	Positive
2	door handle	0.0011		0.0011	Positive
3	engine power	0.0244	0.0194	0.0050	Positive
4	exhaust		0.0013	−0.0013	Negative
5	fuel consumption	0.0246	0.0135	0.0111	Positive
6	interior design		0.0026	−0.0026	Negative
7	mirror/visibility		0.0011	−0.0011	Negative
8	seats	0.0036	0.0034	0.0001	Positive

When it comes to sports cars, among eight features (N_{FO}), the features seats, audio, fuel consumption, door handle, and engine represented positive opinions, while the features exhaust, interior design, and mirror visibility represented negative opinions.

For visual interpretation, Figure 5.2 shows summarized user opinions such as positive and negative opinions of the product features in bar chart form. The solid portion in a bar represents positive opinions and the checkered portion represents negative opinions. For example, engine power, fuel consumption, and seats were identified as positive based on the overall opinion polarity. However, these features also have negative aspects as well. For seat as an identified feature, 51% of reviewers who talked about seats expressed positive opinions, while 49% of the reviewers expressed negative opinions. Broadly speaking, this result indicated that car seats could be analyzed for further development or improvement. A similar result can be seen for engine power and fuel consumptions.

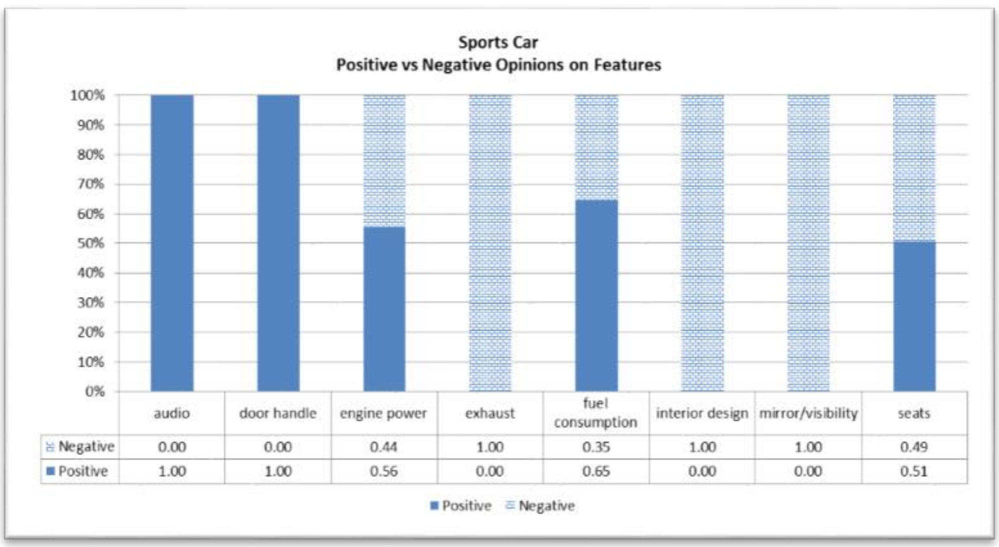


Figure 5.2 Positive and negative opinions expressed on each feature for sports car,

5.5 Product Feature-Cause Extraction

After identifying the association of features and opinions, this study was continued to understand causes of negative opinions by applying the association rule to adjectives, nouns, and verbs, with the noun as a feature and the verb as the cause. The objective of this step was to identify the causes of negative features by inducing a verb along with a noun and adjective.

To extract nouns, adjectives, and verbs from the reviews, some linguistic feature tasks were applied to transform the text data into a format that the computer could recognize for the opinion mining process such as POS tagging, stop word removing, and stemming (Nasukawa and Yi 2003; Wahl, Winiwarter et al. 2010). After determining the POS tag of each word, the method finds the nouns, noun phrase, adjectives, and verbs, and only keeps nouns, adjectives, and verbs. To increase term frequency, a synonym of the adjectives was used and tenses were normalized regarding the verb by utilizing WordNet (Miller 2012) and a custom-made dictionary from the collection of reviews. Then, the noun and noun phrases were replaced by using a distinct product feature list discussed in the previous section.

For sports car, nine association rules for four unique features (N_{FE}) were identified for further analysis to understand the reasons for the negative opinion. The following table presents the meaningful results from the association rule mining. Table 5.10 represents the number of rules identified per feature and its cause for negative opinions.

Table 5.10 Selected Interesting Association Rules for Sports Car

Rules	Features (Conclusion)	Cause (Premises)	Support	Confidence
1	engine power	low	0.0018	0.22
2	engine power	weak	0.0082	0.21
3	interior design	easy scratch	0.0137	1.00
4	interior design	looks cheap	0.0103	0.75
5	interior design	weak plastic	0.0172	0.71
6	mirror/visibility	blind spot	0.0664	1.00
7	seats	little	0.0036	0.41
8	seats	rear	0.0022	0.21
9	seats	weak	0.0053	0.14

Then, the individual association rule results were presented into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of frequencies, and confidence was calculated as the average of confidence, where support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by rule confidence measure helped to identify the rules that might be important for designers/engineers. It took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table. All findings were valuable; however, investigating customer opinions toward the mirrors could be the first priority for this manufacturer.

Table 5.11 Aggregate Association Rules for Sports Car

Features	Sum of Support	Average of Confidence	Sum of Support x Confidence
engine power	0.0099	0.21	0.0021
interior design	0.0412	0.82	0.0338
mirror/visibility	0.0664	1.00	0.0664
seats	0.0111	0.25	0.0026

5.6 Evaluation Results

Since the proposed method showed satisfactory results, which Web reviews contain product design intelligence, the results were evaluated by using design-level information quality metric, which was calculated directly after the product feature extraction. The metric determined whether product features existed in the reviews, and calculated design-level-information quality.

The number of sentences per review (N_S/N_R) was over 9.83. The number of words per sentence (N_W/N_S) was 12.50. The number of candidate feature per sentence (N_C/N_S) was 3.1. The ratio of feature nouns to the total number of candidate feature nouns (N_F/N_C) was 15%, which meant of 20,512 candidate features, 2984 were identified as an indistinct product feature by the product development managers.

Table 5.12 Descriptive Statistic for Sports Car

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
Sports car	671	6595	9.83	82426	12.50	20512	2984	3.11	0.15

Based on the design-level information quality metric, sports car DLIQ score was found upper end of threshold score of 70. This indicates that there were sufficient data in the reviews that consumers post about their experiences with the product features. Concerning sports cars, 42 product features that consumers talked about were identified. Table 5.13 shows the DLIQ score and its three components for sports car, which were content, complexity, and relevancy.

Table 5.13 Design-level Information Quality Score for Sports Car

Product Type	CONTENT				COMPLEXITY				RELEVANCY				DLIQ	Est. STD
	Weight	Base	Value	I(Cont)	Weight	Base	Value	I(Cmplx)	Weight	Base	Value	I(Relv)		
sport car	0.30	153.0	144.0	**28.2**	0.30	2.1	1.8	**25.6**	0.40	6.8	5.6	**33.0**	**86.8**	10.1
				32.5%				29.5%				38.0%	100 %	

Key determinants of content were the number of reviews and the length of the reviews as measured by the number of words, where the value of content I(Cont)—28.2—was determined by using the equations (3.3). Key determinants of complexity were the number of sentences per review and the ratio of nouns in the words in a review, where the value of complexity I(Cplx)—25.6—was determined by using the equation (3.6). Key determinants of relevancy were the total volume of nouns in the reviews and the ratio of feature nouns in the total nouns in the review database, where the value of relevancy I(Relv)—33.0—was determined by using the equation (3.9). Further, the DLIQ score was determined by summing of the content, complexity, and relevancy scores. It can be seen that relevancy value (33.0) has the most dominant effect on the DLIQ score followed by content value (28.1), and complexity value (25.6) respectively. The DLIQ score for sports car was determined as 86.8%, which indicates that it can be moved forward to the next step expecting some degree of outcome of product design intelligence.

5.7 Summary/Conclusion

In this chapter, the method proves that Web reviews on sports cars were critically meaningful resources regarding the richness of content and information quality.

Among 671 reviews collected from CarReview.com and KBB.com, in the product feature extraction process (5.3), 42 distinct product features were identified, which online reviewers often talked about. Feature frequency was presented via the word cloud visualization approach (also called tag clouds), which represents the relative importance of words, in this case, sports car features. As can be seen in Figure 5.3, reviewers talked more frequently about “engine power,” followed by the “fuel consumption,” “audio,” and “seats.”

Figure 5.3 Important features making purchasing decisions for sports car.

Figure 5.3 is an explanation that word size was proportional to the frequency of the feature. The larger the word the greater importance of that feature; smaller words represent features of relatively low import (software used was wordle.net).

In the feature opinion sentiment identification process (5.4), among 32,976 rules identified through the association rule algorithm, 42 rules, which represent the association rules between the opinion (premises) and related features (conclusion), were presented. The 42 rules represent eight distinct features.

Among eight distinct features identified during the feature-opinion step, feature-cause analysis presents product manufacturers four valuable products intelligence information for enhancement, which was how consumers perceived products specifically concerning features and what drove consumers to have negative feedback.

Consumers looked for other people's experience on engine power, interior design, mirror/visibility, and seats when they considered purchasing a sports car. Further, consumers were concerned with low or weak performance on engine power. They disliked cheap looking, plastic-made, weak interior design materials; they disliked a blind spot in the rear mirror, and they disliked tiny back seats.

In addition to features with negative opinions and their causes, the method in this study shows the relative importance of these four features, which aid in manufacturers prioritizing the development or enhancement plan before their next new model release. Product development managers in manufacturing were expected to have greater returns if they enhanced mirror/visibility, specifically the blind spot, as this study reveals the highest sum of support was a confidence score, which was 0.0644, followed by interior design, seats, and engine power respectively.

Presenting this information via graphical form makes it easier for the product managers to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a story with visuals rather than plain words or numbers.

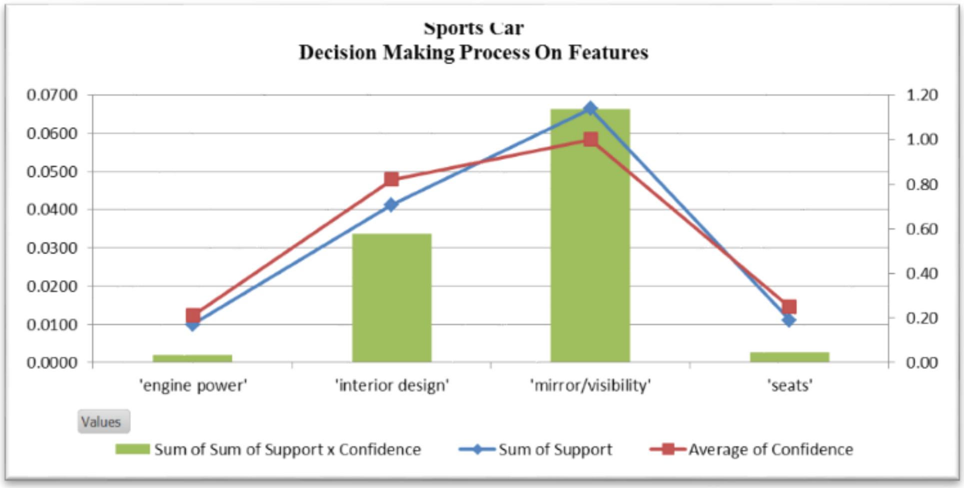


Figure 5.4 Decision making process concerning features for sports car.

Today, new product development, whether to develop a brand new item or modify an existing product, has become a critical aspect in business and engineering. To stay ahead of its competitors regardless of domains, businesses have looked for ways to read consumers' likes and dislikes, one of the most promising sources was Web reviews. They were freely available, extremely widespread, and tremendously rich regarding content.

However, the large amount of information and its widespread location make it challenging for product development managers in business to find all relevant information, read them, summarize them, and organize them into a usable format for competitive advantage or to drive business critical information about future opportunities.

The proposed method aids in the extraction of the design intelligence, which seemingly cannot be identified, which enables product development to extract unknown information from Web reviews and provides the opportunity for new product development to determine consumer needs or expectations about products. The method not only provides critically meaningful product intelligence, but it brings product manufactures the operational efficiency about collecting consumer requirements to avoid time-consuming and labor-intensive traditional ways such as prototype testing, market survey instruments, field testing methods, and hiring independent assessment companies.

5.8 Executive Summary Report

This is an executive summary report, which provides aggregate key results.

Table 5.14 Executive Summary Report for Sports Car

<table>
<tr><td colspan="8">DFOC ANALYSIS - SPORTS CAR</td></tr>
<tr><td colspan="8">WEB REVIEW DATABASE</td></tr>
<tr><td>N_R =</td><td>671</td><td>N_S =</td><td>6595</td><td>N_W =</td><td>82426</td><td colspan="2"></td></tr>
<tr><td colspan="8">DFOC EXTRACTION RATIOS</td></tr>
<tr><td>N_C =</td><td>20512</td><td>N_F =</td><td>2984</td><td>N_C/N_S =</td><td>3.11</td><td>N_F/N_C =</td><td>0.15</td></tr>
<tr><td colspan="8">DESIGN LEVEL INFORMATION QUALITY</td></tr>
<tr><td>$DLIQ_{Cont}$ =</td><td>28.2</td><td>$DLIQ_{Cplx}$ =</td><td>25.6</td><td>$DLIQ_{Relv}$ =</td><td>33</td><td>$DLIQ$ =</td><td>86.8</td></tr>
<tr><td colspan="8">DESIGN INTELLIGENCE</td></tr>
<tr><td colspan="5">Product Features Identified - Notable Interest (TF>5) =</td><td>42</td><td colspan="2" rowspan="2"></td></tr>
<tr><td colspan="5">Product Features Identified - Significant Interest (TF>23) =</td><td>8</td></tr>
<tr><td colspan="8">FEATURE OPINION ANALYSIS</td></tr>
<tr><td>#</td><td colspan="2">Feature</td><td>Support</td><td>+ Opinion</td><td>- Opinion</td><td colspan="2">Polarity</td></tr>
<tr><td>1</td><td colspan="2">Engine Power</td><td>4.4%</td><td>0.0244</td><td>0.0194</td><td colspan="2">Mixed</td></tr>
<tr><td>2</td><td colspan="2">Fuel Consumption</td><td>3.8%</td><td>0.0246</td><td>0.0135</td><td colspan="2">Mixed</td></tr>
<tr><td>3</td><td colspan="2">Audio</td><td>1.5%</td><td>0.0154</td><td>0</td><td colspan="2">Positive</td></tr>
<tr><td>4</td><td colspan="2">Seats</td><td>0.7%</td><td>0.0036</td><td>0.0034</td><td colspan="2">Mixed</td></tr>
<tr><td>5</td><td colspan="2">Interior Design</td><td>0.3%</td><td>0</td><td>0.0026</td><td colspan="2">Negative</td></tr>
<tr><td>6</td><td colspan="2">Mirror/Visibility</td><td>0.1%</td><td>0</td><td>0.0011</td><td colspan="2">Negative</td></tr>
<tr><td>7</td><td colspan="2">Door Handle</td><td>0.1%</td><td>0.0011</td><td>0</td><td colspan="2">Positive</td></tr>
<tr><td>8</td><td colspan="2">Exhaust</td><td>0.3%</td><td>0.0013</td><td>0.0013</td><td colspan="2">Negative</td></tr>
<tr><td colspan="8">FEATURE-CAUSE-OPINION (Negative) ANALYSIS</td></tr>
<tr><td>#</td><td colspan="2">Feature</td><td>Cause</td><td>Support</td><td>Confidence</td><td colspan="2">DFOC Strength</td></tr>
<tr><td rowspan="2">1</td><td colspan="2" rowspan="2">Engine Power</td><td>Low</td><td>0.18%</td><td>22%</td><td colspan="2" rowspan="2">0.21</td></tr>
<tr><td>Weak</td><td>0.82%</td><td>21%</td></tr>
<tr><td rowspan="3">4</td><td colspan="2" rowspan="3">Seats</td><td>Small</td><td>0.36%</td><td>41%</td><td colspan="2" rowspan="3">3.38</td></tr>
<tr><td>Rear</td><td>0.22%</td><td>21%</td></tr>
<tr><td>Damage</td><td>0.53%</td><td>14%</td></tr>
<tr><td rowspan="3">5</td><td colspan="2" rowspan="3">Interior Design</td><td>Easy scratch</td><td>1.37%</td><td>100%</td><td colspan="2" rowspan="3">6.64</td></tr>
<tr><td>Looks Cheap</td><td>1.03%</td><td>75%</td></tr>
<tr><td>Plastic</td><td>1.72%</td><td>71%</td></tr>
<tr><td>6</td><td colspan="2">Mirror/Visibility</td><td>Blind spot</td><td>6.64%</td><td>100%</td><td colspan="2">0.26</td></tr>
</table>

CHAPTER 6

MULTIPLE PRODUCT RESULTS

In this section, the nine datasets for each product are presented. In particular, the datasets are television, digital camera, laptop, mobile phone, sedan, mobile phone service provider, airline customer service, restaurant, and hotel. The common processes (i.e., data preparation and linguistic text processing) are briefly described (section 6.1), and then the results of the identification of features, feature-opinion pairs, and feature-cause pairs are presented for each dataset. Then, the evaluation metric for each dataset is described at the end of this chapter. The tenth dataset (sports car) was discussed in Chapter 5 as a walkthrough example in details; therefore, it is not presented in Chapter 6.

6.1 Data Preparation and Computational Linguistic Feature Process

Web reviews for a particular model of each product were randomly collected from Amazon.com, CarReview.com, KBB.com, CustomerServiceScoreboard.com, Yelp.com, and TripAdviser.com. Respectively, 480, 745, 312, 352, 749, 847, 570, 821, 528 reviews were collected to analyze television, digital camera, laptop, mobile phone, sedan, mobile phone service provider, airline customer service, restaurant and hotel. A typical review had review date, the reviewer’s name, a numerical rating, and the body of the review, which consumers use to share their experiences with products or services. The review body where a consumer shares opinions was of interest while the remaining parts were noisy data, which was excluded, from the study. As discussed in previous chapters, the unnecessary data was eliminated to extract the body of the reviews that need to be analyzed. As discussed in Chapter 3, the splitting process was necessary because compound reviews may contain several features, each of which may represent different opinions. After sentence splitting, the reviews (N_R) become 6765, 7455, 3256, 1405, 3428, 7200, 5370, 11718, and 5788 sentences (N_S), respectively. As discussed in previous chapters, for each dataset individually, nouns, adjectives, and verbs were extracted from the reviews, with nouns as features, adjectives as opinions, and verbs were causes; some computational linguistic text processing tasks such as natural language processing were applied to transform the text data into a format that a computer could recognize for the opinion mining process such as POS tagging, stop word removing, and stemming to increase the accuracy of the search information and the overall effectiveness of the process (Nasukawa and Yi 2003; Wahl, Winiwarter et al. 2010). To increase term frequency, synonyms of adjectives were identified and replaced by utilizing WordNet (Miller 2012) and a custom-made dictionary from the collection of Web reviews. After completing linguistic feature tasks, three analyses for each product and service dataset were performed: identifying (1) product features, (2) feature-opinion pairs, and (3) feature-cause pairs. In the next subsections, findings for each product were presented. The results of three analyses were presented for the remaining part of this chapter.

6.2 Data Analysis of Television

In the feature extraction step, to identify product features, nouns and noun phrases were extracted from the review data as candidate features. To extract noun phrases, the N-gram model method was used, specifically the bi-gram method was applied along with POS tagging (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009).

Since not every candidate feature was a real product feature, another constraint was placed on features to filter out non-features and increase the accuracy of feature extraction by employing a knowledge-based heuristic approach, which was product development managers validating the list of candidate features. As shown in the table, the number of candidate feature per sentence (N_C/N_S) was 3.4. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 22%, which meant of 22,969 candidate features, 5,049 (48 distinct) were identified as indistinct product features by the product development managers (Table 6.1).

Table 6.1 Descriptive Statistic for Television

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
television	480	6765	14.09	107,518	15.89	22,969	5049	3.40	0.22

In addition, different reviewers often referred to the same product features using different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011). To increase term frequency (TF), similar features were grouped together based on the same meaning. For example, the features picture quality, clarity, photo, halo, pixel, image, plasma, HD-quality, and contrast were grouped to "screen resolution/image."

Table 6.2 shows 48 distinct product features identified by the product development managers and their corresponding frequencies in descending order. As was shown, high frequency indicated reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious that screen was the feature consumers talked about most when it came to television, followed by (e.g., sound quality, Internet connection).

Table 6.2 Distinct Product Features for Television

Num (i)	Feature (F_i – distinct)	TF (F_i-freq)	Num (i)	Feature (F_i – distinct)	TF (F_i-freq)
1	screen resolution/image quality	1677	25	wall mount	41
2	speakers/sound quality	306	26	camera	36
3	Internet connection	273	27	warranty	36
4	on/off timer	206	28	router	35
5	DVD player	203	29	motion control	33
6	remote control	192	30	signal strength (antenna/cable)	28
7	3d mode	191	31	PC connection	27
8	keyboard	179	32	power switch	27
9	inputs and outputs	173	33	player	25
10	video and audio	157	34	Bluetooth	24
11	SSG active glass	126	35	battery	23
12	cabinet color	101	36	HDTV display technologies	18
13	price	85	37	cable box (HD box)	17
14	game mode	84	38	smart TV	17
15	HDMI	79	39	starter kit for 3D	17
16	screen size	75	40	auto motion	15
17	smart hub system	75	41	vertical band	11
18	user manual	70	42	voice control function	11
19	accessories	64	43	halo effect	8
20	TV stand	63	44	TV weight	8
21	HDMD receiver	51	45	TV receiver	7
22	power cord	50	46	movie mode	6
23	menu	45	47	energy efficiency	5
24	TV size	44	48	screen menu	5

At this point of the analysis, the results show that consumers shared their experiences concerning 48 distinct features. After the association rule, the mining rule was applied to identify correlations between product features and related opinions, which was the second objective of this study.

Four thousand one hundred forty-five association rules were found by the algorithm, which set the minimum confidence (minconf) at 10% to observe a larger number of rules. Although 4145 rules were found by the algorithm, not all of the rules were worthwhile to analyze, meaning that only product features and related opinions on them were analyzed in this study. Among the 4145 rules, only 40 were given as the result of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). First, interesting feature-opinion pairs were identified; then opinion polarity for each feature-opinion pair was identified by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

For an example of feature screen resolution, the adjective (clear crystal) has a positive orientation, so it was labeled as positive, while the adjective (not bright) shows a negative orientation. Therefore, each association was examined and labeled as positive and negative. The following table shows the opinion polarity after assigning the label as negative or positive.

Table 6.3 Selected F-O Association Rules for Television

Rule	Features (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
1	Internet connection	not work	0.0014	0.17	negative
2	Internet connection	slow	0.0020	0.36	negative
3	Internet connection	easy	0.0014	0.11	positive
4	price	high	0.0010	0.13	negative
5	remote control	not function	0.0010	0.32	negative
6	remote control	not work	0.0013	0.15	negative
7	remote control	easy	0.0023	0.18	positive
8	remote control	nice	0.0016	0.11	positive
9	screen resolution/image quality	bad	0.0032	0.33	negative
10	screen resolution/image quality	not	0.0576	0.31	negative
11	screen resolution/image quality	not amaz	0.0024	0.58	negative
12	screen resolution/image quality	not bright	0.0021	0.81	negative
13	screen resolution/image quality	not clear	0.0013	0.71	negative
14	screen resolution/image quality	not good	0.0054	0.45	negative
15	screen resolution/image quality	not great	0.0034	0.47	negative
16	screen resolution/image quality	not nice	0.0018	0.36	negative
17	screen resolution/image quality	not perfect	0.0011	0.56	negative
18	screen resolution/image quality	poor	0.0011	0.35	negative
19	screen resolution/image quality	terrible	0.0010	0.53	negative
20	screen resolution/image quality	amaz	0.0211	0.58	positive
21	screen resolution/image quality	beauti	0.0035	0.50	positive
22	screen resolution/image quality	bright	0.0071	0.69	positive
23	screen resolution/image quality	bright sharp	0.0010	0.73	positive
24	screen resolution/image quality	brighter	0.0010	1.00	positive
25	screen resolution/image quality	clear	0.0061	0.77	positive
26	screen resolution/image quality	clear crystal	0.0010	0.80	positive
27	screen resolution/image quality	good	0.0216	0.41	positive
28	screen resolution/image quality	good amaz	0.0013	0.63	positive
29	screen resolution/image quality	good great	0.0018	0.61	positive
30	screen resolution/image quality	great	0.0206	0.46	positive
31	screen resolution/image quality	great amaz	0.0018	0.74	positive
32	screen resolution/image quality	impress	0.0015	0.57	positive
33	screen resolution/image quality	natur	0.0014	0.79	positive
34	screen resolution/image quality	Nice	0.0058	0.40	positive
35	screen resolution/image quality	spectacular	0.0010	0.62	positive
36	screen resolution/image quality	vibrant	0.0016	0.72	positive
37	speakers/sound quality	not great	0.0014	0.19	negative
38	speakers/sound quality	fine	0.0013	0.17	positive
39	SSG active glass	free	0.0013	0.14	positive
40	video and audio	bad	0.0011	0.12	negative

After classifying the sentiment of an adjective into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values concerning its feature using the opinion polarity score equation discussed in Chapter 4.

For instance, the feature "speakers" shows a sum of all of the positive support values was 0.0013, and the sum of all of the negative values was 0.0014. These two values were used to identify overall opinion or impression of product features. In the feature speakers, the overall opinion polarity score was calculated as −0.0001 (negative). The opinion polarity score of each feature was determined by calculating the difference between the sum of all positive support values and the sum of all negative support values. The table shows the summary of the association rules for the distinct features and demonstrates the feature-wise frequencies and final opinion polarity.

Table 6.4 Aggregate Opinion Polarities on Features for Television

Num (i)	Features (Conclusion)	Sum of Positive Opinion	Sum of Negative Opinion	Opinion Polarity Score (+/-)	Opinion Polarity
1	Internet connection	0.0014	0.0034	−0.0020	Negative
2	price		0.0010	−0.0010	Negative
3	remote control	0.0039	0.0023	0.0016	Positive
4	screen resolution/image quality	0.0991	0.0804	0.0187	Positive
5	speakers/sound quality	0.0013	0.0014	−0.0001	Negative
6	SSG active glass	0.0013		0.0013	Positive
7	video and audio		0.0011	−0.0011	Negative

When it comes to television, among seven features (N_{FO}), the features screen resolution, remote control, and SSG active glass represented positive opinions, while the features Internet connection, price, sound quality, and video represented negative opinions.

For visual interpretation, Figure 6.1 shows summarized user opinions such as positive and negative on the product features in bar chart form. The checkered pattern portion in a bar represents negative opinions and the solid portion represents positive opinions. For example, remote control and screen resolution were identified as positive based on the overall opinion polarity. However, these features also have negative aspects as well. For remote control as an identified feature, 63% of reviewers who talked about it expressed a positive opinion, while 37% of the reviewers expressed negative opinions. Broadly speaking, this result indicates that the remote control could be analyzed for further development or improvement. A similar result could be seen for screen resolution and sound quality.

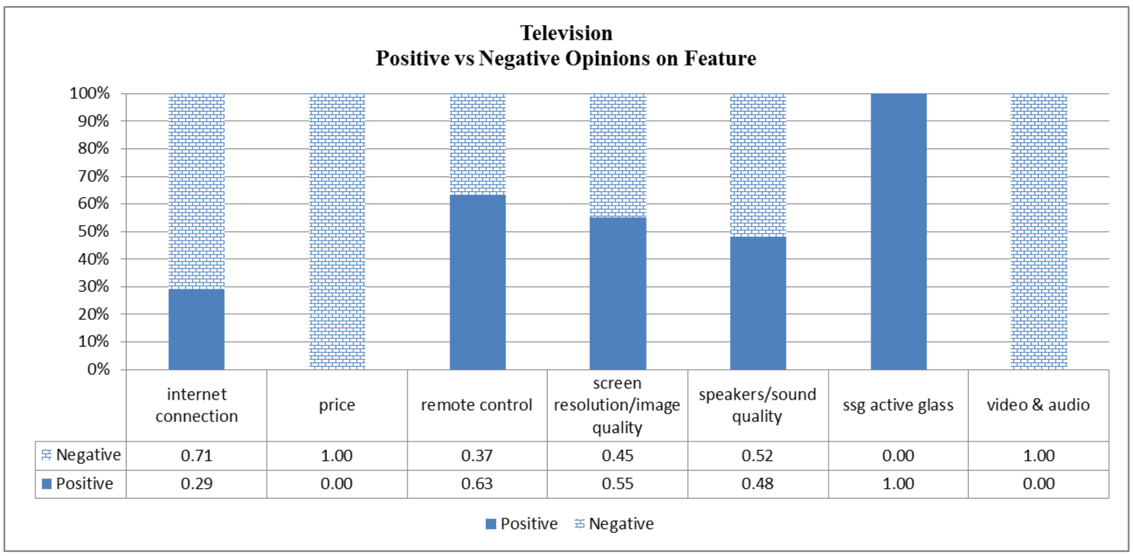


Figure 6.1 Opinion polarities expressed on each feature for television.

In feature-cause extraction, after applying the association rule on adjectives, nouns, and verbs, 17 association rules for five distinct features (N_{FE}) were identified to analyze further to understand the reasons behind the negative opinions. Table 6.5 represents the number of rules identified per feature and its cause for negative opinions.

Table 6.5 Selected F-C Association Rules for Television

Rules	Features (Conclusion)	Cause (Premises)	Support	Confidence
1	price	high	0.0541	1.00
2	price	store match	0.0135	1.00
3	video and audio	blurai	0.0024	0.50
4	video and audio	not come	0.0014	0.12
5	video and audio	not connect	0.0028	0.25
6	video and audio	not play	0.0033	0.39
7	video and audio	not support instant	0.0014	1.00
8	video and audio	not work	0.0033	0.25
9	video and audio	problem play	0.0014	0.75
10	Internet connection	not connect	0.0015	0.31
11	Internet connection	not work	0.0016	0.15
12	Internet connection	slow	0.0020	0.36
13	screen resolution/image quality	noise	0.0011	0.69
14	screen resolution/image quality	not clear	0.0014	0.73
15	remote control	not control	0.0018	0.56
16	remote control	not function	0.0010	0.31
17	remote control	not work	0.0020	0.18

Then, the individual association rule results were grouped into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of frequencies, and confidence was calculated as the average of confidence, where support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by rule confidence measure helped to identify the rules that might be important for designers/engineers, as it took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table.

Table 6.6 Aggregate F-C Association Rules for Television

Features (Conclusion)	Sum of Support	Average of Confidence	Sum of Support x Confidence
price	0.0676	1.00	0.0676
video and audio	0.0161	0.46	0.0067
Internet connection	0.0052	0.27	0.0014
screen resolution/image quality	0.0025	0.71	0.0018
remote control	0.0048	0.35	0.0017

Presenting this information via graphical form makes it easier for people to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a visual story rather than plain words or numbers. For remote control, the problem resides in its functions (not controls, not functions, and not works).

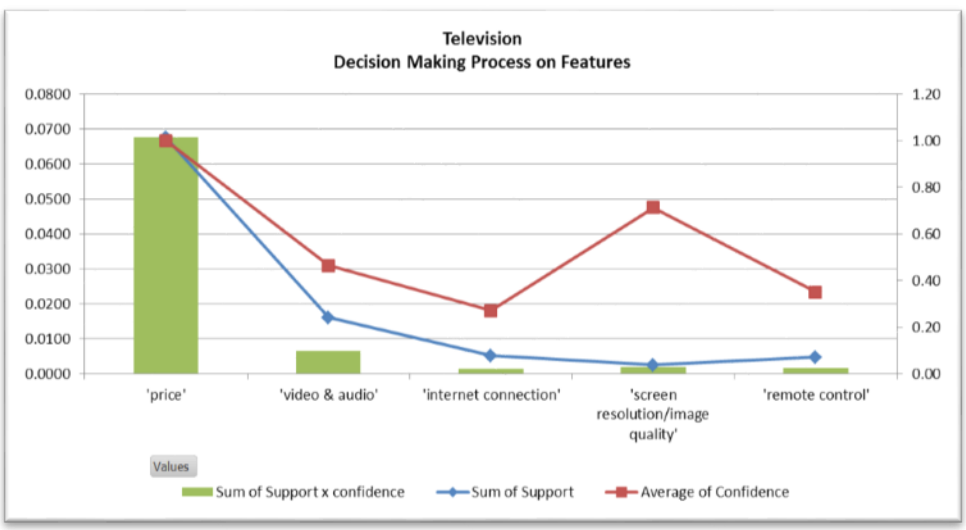


Figure 6.2 Decision making process on features for television.

The method in this study shows the relative importance of these five features, which help the situation when manufactures need to prioritize the development or enhancement plan before their next new model release.

Product development managers in manufacturing are expected to have greater return if they enhance video/audio functionality, as this study reveals the highest sum of support with a confidence score of 0.0067 (directly after price), followed by screen resolution, remote control, and Internet connections respectively.

6.3 Data Analysis of Digital Camera

In the feature extraction step, to identify product features, nouns and noun phrases from the review data were extracted as candidate features. To extract noun phrases, the N-gram model method was used, specifically the bi-gram method was applied along with POS tagging (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009). Since not all candidate features were real product features, another constraint on features was placed to filter out non-features and increase the accuracy of feature extraction by employing a knowledge-based heuristic approach; product development managers validated the list of candidate features. As shown in the table, the number of candidate features per sentence (N_C/N_S) was 3.26. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 12%, which meant of 24,320 candidate features, 2937 (35 distinct) were identified as an indistinct product feature by the product development managers (Table 6.7).

Table 6.7 Descriptive Statistic for Digital Camera

Product Type	Nr	Ns	Ns/Nr	Nw	Nw/Ns	Nc	Nf	Nc/Ns	Nf/Nc
digital camera	745	7455	10.01	118173	15.85	24320	2937	3.26	0.12

In addition, different reviewers often referred to the same product features using different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011). To increase term frequency (TF), similar features were grouped together based on the same meaning. For example, the features megapixel, resolution, and pixel were grouped to "resolution."

Table 6.8 shows 35 distinct product features identified by the product development managers and their corresponding frequencies in descending order. As is shown, high frequency indicated reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious that "picture quality" was the feature consumers talked about most when it came to a digital camera, followed by (e.g., video-quality, battery-life, Powershot).

Table 6.8 Distinct Product Features for Digital Camera

Num (i)	Feature (F_i – distinct)	TF (F_i- freq)	Num (i)	Feature (F_i – distinct)	TF (F_i- freq)
1	picture/picture quality	851	19	SD card	32
2	video/video quality	242	20	exposure	27
3	battery/battery life	200	21	aperture	26
4	Powershot	184	22	warranty	26
5	auto-flash/flash	132	23	image stabilization	21
6	zoom/lens	131	24	LCD screen	19
7	case	104	25	motion	13
8	size	103	26	processor	12
9	light	101	27	HDMI cable	11
10	resolution	100	28	landscape	11
11	screen	93	29	video/movie quality	11
12	storage	85	30	portrait	9
13	ISO	84	31	HDR	7
14	CMOS sensor	81	32	handheld	5
15	auto mode	62	33	menu option	5
16	shutter/shutter speed	58	34	owner manual	5
17	memory size	48	35	video/movie quality	5
18	power switch	33			

At this point of the analysis, the results show that consumers shared their experiences on 35 distinct features. Afterward, the association rule mining rule was applied to identify correlations between product features and related opinions, which was the second objective.

Ten thousand four hundred sixty-four association rules were found by the algorithm, which set the minimum confidence (minconf) at 10% to observe a larger number of rules. Although 10,462 rules were found by the algorithm, not all of the rules were worthwhile to analyze, meaning that only product features and related opinions on them were analyzed in this study. Among the 10,462 rules, only 34 were given as the result of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). First, interesting feature-opinion pairs were identified; then opinion polarity for each feature-opinion pair was identified by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

For example, of the feature picture quality, the adjective clear had a positive orientation, so it was labeled as positive, while the adjective bad showed a negative orientation. Therefore, each association was examined and labeled as positive and negative. The following table shows the opinion polarity after assigning the label as negative or positive.

Table 6.9 Selected F-O Association Rules for Digital Camera

Rule	Features (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
1	auto-flash/flash	bright	0.0011	0.17	positive
2	auto-flash/flash	light low	0.0025	0.13	negative
3	battery/battery life	good life	0.0017	0.94	positive
4	iso	high	0.0024	0.19	positive
5	light	good	0.0074	0.11	positive
6	motion	not slow	0.0010	0.50	negative
7	picture/picture quality	bad	0.0029	0.37	negative
8	picture/picture quality	bright	0.0024	0.36	positive
9	picture/picture quality	clear	0.0050	0.71	positive
10	picture/picture quality	good	0.0241	0.35	positive
11	picture/picture quality	great	0.0220	0.37	positive
12	picture/picture quality	nice	0.0055	0.27	positive
13	picture/picture quality	not good	0.0058	0.40	negative
14	picture/picture quality	not great	0.0044	0.43	negative
15	picture/picture quality	not nice	0.0011	0.30	negative
16	picture/picture quality	not sharp	0.0015	0.50	negative
17	picture/picture quality	perfect	0.0021	0.31	positive
18	picture/picture quality	poor	0.0014	0.36	negative
19	screen	big	0.0014	0.11	positive
20	size	perfect	0.0011	0.16	positive
21	video/video quality	good	0.0104	0.15	positive
22	video/video quality	good HD	0.0021	0.95	positive
23	video/video quality	good light	0.0018	0.24	positive
24	video/video quality	great	0.0107	0.18	positive
25	video/video quality	great easy	0.0010	0.20	positive
26	video/video quality	great HD	0.0024	0.95	positive
27	video/video quality	great light	0.0010	0.25	positive
28	video/video quality	great shoot	0.0011	0.32	positive
29	video/video quality	nice	0.0029	0.15	positive
30	video/video quality	not good	0.0029	0.21	negative
31	video/video quality	not great	0.0026	0.26	negative
32	video/video quality	perfect	0.0010	0.15	positive
33	zoom/lens	big	0.0014	0.11	positive
34	zoom/lens	not great	0.0010	0.10	negative

After classifying the sentiment of adjectives into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values concerning its feature using the opinion polarity score equation discussed in Chapter 4.

For instance, the feature auto-flash shows the sum of all of the positive support values was 0.0011, and the sum of all of the negative values was 0.0025. These two values were used to identify overall opinion or impression of product features. In the feature auto-flash, the overall opinion polarity score was calculated as –0.0014 (negative). The opinion polarity score of each feature was determined by calculating the difference between the sum of all positive support values and the sum of all negative support values. The table shows the summary of the association rules for the distinct features, and demonstrates the feature-wise frequencies and final opinion polarity.

Table 6.10 Aggregated Opinion Polarities on Features for Digital Camera

Num (i)	Features (Conclusion)	Sum of Positive Opinion	Sum of Negative Opinion	Opinion Polarity Score (+/-)	Opinion Polarity
1	auto-flash/flash	0.0011	0.0025	−0.0014	negative
2	battery/battery life	0.0017		0.0017	positive
3	ISO	0.0024		0.0024	positive
4	light	0.0074		0.0074	positive
5	motion		0.0010	−0.0010	negative
6	picture/picture quality	0.0611	0.0170	0.0441	positive
7	screen	0.0014		0.0014	positive
8	size	0.0011		0.0011	positive
9	video/video quality	0.0345	0.0055	0.0290	positive
10	zoom/lens	0.0014	0.0010	0.0003	positive

When it comes to a digital camera, among ten distinct features (N_{FO}), the features battery-life, ISO, light, picture quality, screen, size, video quality, and zoom represented positive opinions, while the features auto-flash and motion represented negative opinions.

For visual interpretation, Figure 6.3 shows summarized positive and negative user opinions on the product features in bar chart form. The checkered pattern portion in a bar represents negative opinions and the solid portion represents positive opinions. For example, remote control, and screen resolution were identified as positive based on the overall opinion polarity. However, these features also have negative aspects as well. For auto-flash as an identified feature, 69% of reviewers who talked about this feature expressed a positive opinion, while 31% of the reviewers expressed negative opinions. Broadly speaking, this result indicates that auto-flash could be analyzed for further development or improvement. A similar result can be seen in picture quality, video quality, and zoom.

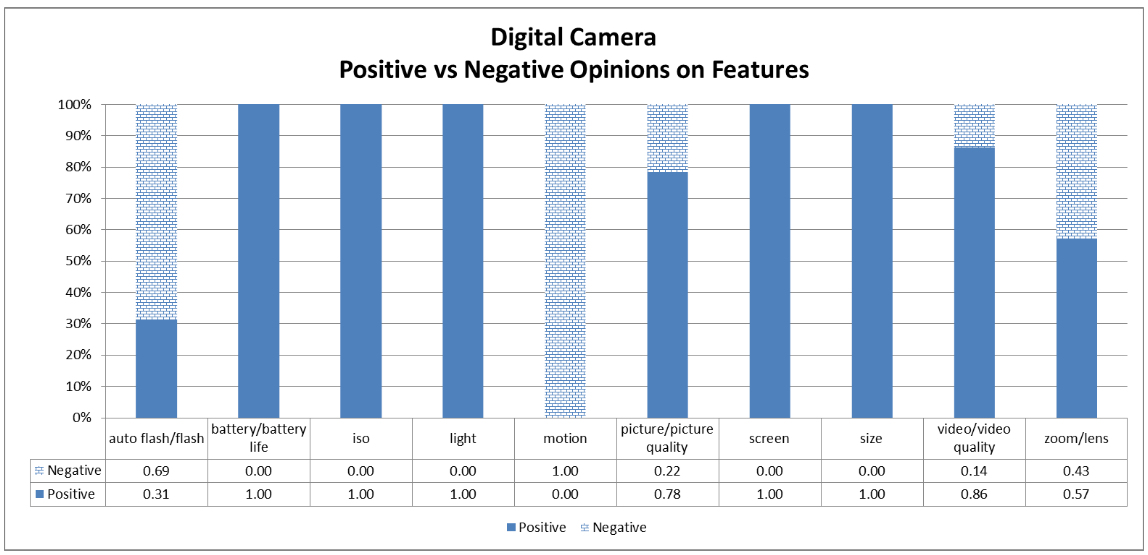


Figure 6.3 Opinion polarities expressed on each feature for digital camera.

In feature-cause extraction, after applying the association rule on adjectives, nouns, and verbs, seven association rules for three distinct features (N_{FE}) were identified to analyze further to understand the reasons of the negative opinion. Table 6.11 represents the number of rules identified per feature and its cause for negative opinions.

Table 6.11 Selected F-C Association Rules for Digital Camera

Rules	Features (Conclusion)	Cause (Premises)	Support	Confidence
1	auto flash/flash	blurry	0.0162	0.50
2	auto flash/flash	corner shadow	0.0116	1.00
3	auto flash/flash	low	0.0580	0.57
4	motion	slow record	0.0833	1.00
5	zoom/lens	make noise	0.0077	1.00
6	zoom/lens	optic low	0.0077	0.80
7	zoom/lens	motor sound	0.0077	1.00

Then, the individual association rule results were grouped into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of frequencies and confidence was calculated as the average of confidence, where support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by the confidence measure helped to identify the rules that might be important for designers/engineers, as it took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table.

Table 6.12 Aggregate F-C Association Rules for Digital Camera

Features	Sum of Support	Average of Confidence	Sum of Support x Confidence
auto flash/flash	0.0858	0.69	0.0527
motion	0.0833	1.00	0.0833
zoom/lens	0.0231	0.93	0.0215

Presenting this information via graphical form makes it easier for people to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a story with visuals rather than plain words or numbers. For auto-flash, the problem seems to be related to the corner shadow, low flashlights, and blurry image.

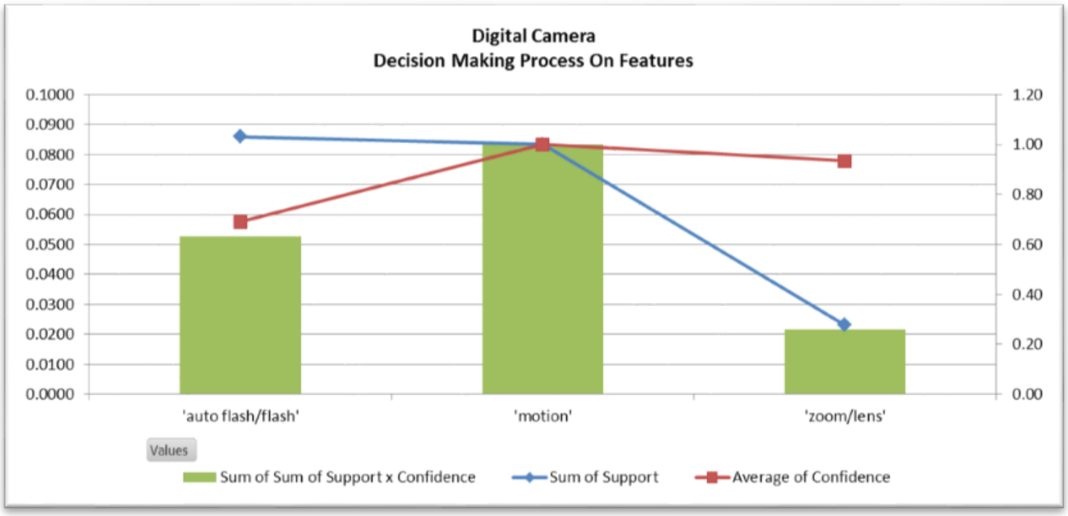


Figure 6.4 Decision making process on features for digital camera.

This study shows the relative importance of these three features, which help the situation when manufactures need to prioritize the development or enhancement plan before their next new model release.

Product development managers in manufacturing are expected to have greater return if they enhance motion functionality as the study reveals the highest sum of support a confidence score of 0.0833, followed by auto flash, and zoom/lens.

6.4 Data Analysis of Laptop

In the feature extraction step, to identify product features, nouns and noun phrases were extracted from the review data as candidate features. To extract noun phrases, the N-gram model method was used; specifically, the bi-gram method was applied along with POS tagging (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009).

Since not all candidate features were real product features, another constraint on features was placed to filter out non-features and increase the accuracy of feature extraction by employing a knowledge-based heuristic approach; product development managers validated the list of candidate features. As shown in the table, the number of candidate features per sentence (N_C/N_S) was 3.34. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 13%, which meant of 10,876 candidate features, 1421 (25 distinct) were identified as an indistinct product feature by the product development managers (Table 6.13).

Table 6.13 Descriptive Statistic for Laptop

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
laptop	312	3256	10.44	48,882	15.01	10876	1421	3.34	0.13

In addition, different reviewers often referred to the same product features using different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011). To increase term frequency (TF), similar features were grouped together based on the same meaning. For example, the features CPU, GHz-processor, and Intel were grouped to "processor."

Table 6.14 shows 25 distinct product features that were identified by the product development managers and their corresponding frequencies in descending order. As is shown, high frequency indicated reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious that "monitor" was the feature consumers talked about most when it came to laptops, followed by (e.g., CD/DVD diver, keyboard, and operating system).

Table 6.14 Distinct Product Features for Laptop

Num (i)	Feature (F_i – distinct)	TF (F_i-freq)	Num (i)	Feature (F_i – distinct)	TF (F_i-freq)
1	monitor	227	14	speakers	28
2	CD/DVD driver	163	15	laptop case	22
3	keyboard	140	16	Apple's backup strategy	21
4	operating system	129	17	price	18
5	processor (or CPU)	103	18	power cord	16
6	memory	94	19	printer	12
7	hard drive	80	20	screen/screen resolution	12
8	battery/battery life	73	21	command key	10
9	I/O ports (input output)	70	22	laptop color	10
10	software	57	23	voice over Internet protocol	6
11	Internet	47	24	VPN connection	6
12	adaptor	36	25	camera	5
13	laptop size/weight	36			

At this point in the analysis, the result shows that consumers shared their experiences on 25 distinct features. Afterward, the association rule mining rule was applied to identify correlations between product features and related opinions.

Three thousand four hundred ten association rules were found by the algorithm, which set the minimum confidence (minconf) at 10% to observe a larger number of rules. Although 3410 rules were found by the algorithm, not all of the rules were worthwhile to analyze, meaning that only product features and related opinions on them were analyzed in the study. Among the 3410 rules, only 38 were given as the result of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). First, interesting feature-opinion pairs were identified; then, opinion polarity for each feature-opinion pair was identified by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

Table 6.15 Selected F-O Association Rules for Laptop

Rule	Features (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
1	battery/battery life	amaz life	0.0019	0.88	positive
2	battery/battery life	good life	0.0019	0.78	positive
3	battery/battery life	great	0.0030	0.10	positive
4	battery/battery life	great life	0.0025	0.90	positive
5	battery/battery life	light	0.0011	0.11	positive
6	CD/DVD driver	perfect	0.0011	0.12	positive
7	hard drive	good memory	0.0011	0.80	positive
8	hard drive	not memory	0.0011	0.29	negative
9	keyboard	amaz	0.0044	0.20	positive
10	keyboard	amaz touch	0.0011	1.00	positive
11	keyboard	big	0.0019	0.27	positive
12	keyboard	easy	0.0039	0.17	positive
13	keyboard	response	0.0011	0.50	positive
14	keyboard	not touch	0.0011	0.80	negative
15	keyboard	not work	0.0011	0.13	negative
16	laptop case	aluminum	0.0011	0.27	positive
17	laptop size/weight	light	0.0011	0.11	positive
18	laptop size/weight	perfect	0.0011	0.12	positive
19	laptop size/weight	big	0.0011	0.15	negative
20	monitor	amaz	0.0036	0.16	positive
21	monitor	amaz quality	0.0011	1.00	positive
22	monitor	big	0.0019	0.27	positive
23	monitor	bright	0.0019	0.70	positive
24	monitor	3D	0.0011	0.21	positive
25	monitor	glossy	0.0025	0.90	positive
26	monitor	good	0.0044	0.11	positive
27	monitor	great	0.0033	0.11	positive
28	monitor	great quality	0.0011	0.57	positive
29	monitor	nice	0.0017	0.13	positive
30	monitor	not great	0.0011	0.24	positive
31	monitor	little	0.0019	0.37	negative
32	monitor	problem	0.0022	0.14	negative
33	operating system	love lion	0.0011	0.80	positive
34	operating system	window good	0.0011	0.36	positive
35	operating system	not work	0.0014	0.16	negative
36	price	worth	0.0028	0.20	positive
37	price	expense	0.0011	0.21	negative
38	processor (or CPU)	dual	0.0014	0.71	positive

After classifying the sentiment of adjectives into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values regarding its feature using the opinion polarity score equation discussed in Chapter 4.

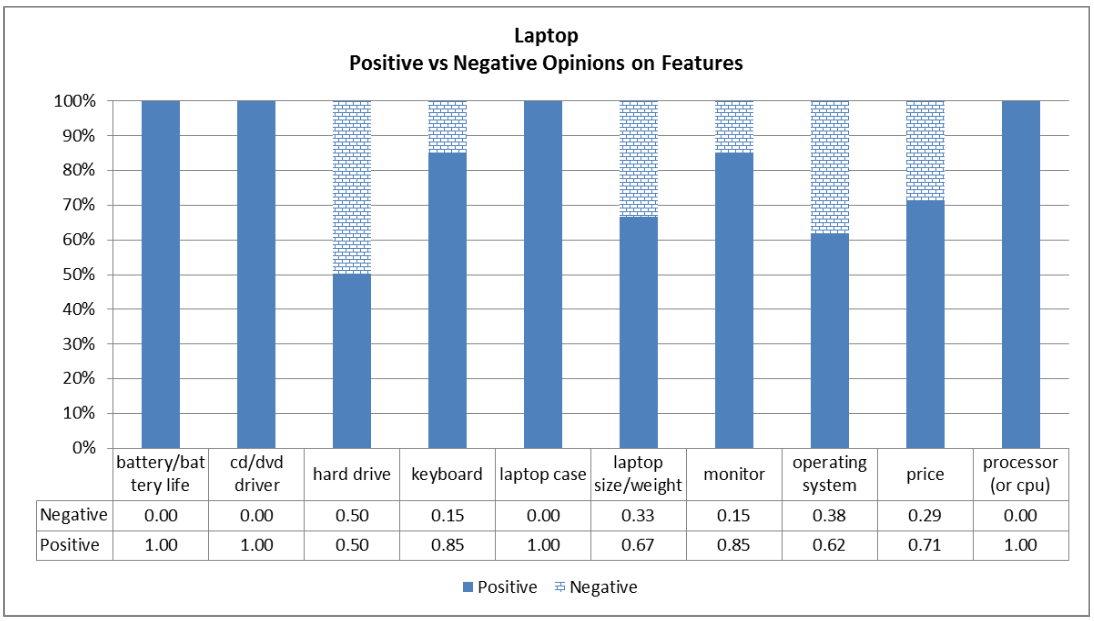
For instance, the feature operating system shows the sum of all of the positive support values was 0.0022, and the sum of all if the negative values was 0.0014. These two values were used to identify overall opinion or the impression of product features. In the feature operating system, the overall opinion polarity score was calculated as 0.0008 (positive). The opinion polarity score of each feature was determined by calculating the difference between the sum of all of the positive support values and the sum of all of the negative support values. The table shows the summary of the association rules for the distinct features and demonstrates the feature-wise frequencies and final opinion polarity.

Table 6.16 Aggregated Opinion Polarities on Features for Laptop

Num (i)	Features (Conclusion)	Sum of Positive Opinion	Sum of Negative Opinion	Opinion Polarity Score (+/-)	Opinion Polarity
1	battery/battery life	0.0105		0.0105	positive
2	CD/DVD driver	0.0011		0.0011	positive
3	hard drive	0.0011	0.0011	0.0000	positive
4	keyboard	0.0124	0.0022	0.0102	positive
5	laptop case	0.0011		0.0011	positive
6	laptop size/weight	0.0022	0.0011	0.0011	positive
7	monitor	0.0238	0.0041	0.0196	positive
8	operating system	0.0022	0.0014	0.0008	positive
9	price	0.0028	0.0011	0.0017	positive
10	processor (or CPU)	0.0014		0.0014	positive

When it comes to laptops, among ten distinct features (N_{FO}), the features battery life, driver, hard drive, keyboard, laptop case, laptop size, monitor, operating system, price, and processor represented positive opinions, while no features represented negative opinions.

For visual interpretation, Figure 6.5 shows summarized user opinions such as positive and negative on the product features in bar chart form. The checkered pattern portion in a bar represents negative opinions and the solid portion represents positive opinions. For example, operating system and keyboard were identified as positive based on the overall opinion polarity. However, these features also have negative aspects as well. For operating system as an identified feature, 62% of reviewers who talked about this feature expressed positive opinions, while 38% of the reviewers expressed negative opinions. Broadly speaking, this result indicates that operating system could be analyzed for further development or improvement. A similar result could be found for keyboard, hard drive, and monitor.



	battery/battery life	cd/dvd driver	hard drive	keyboard	laptop case	laptop size/weight	monitor	operating system	price	processor (or cpu)
Negative	0.00	0.00	0.50	0.15	0.00	0.33	0.15	0.38	0.29	0.00
Positive	1.00	1.00	0.50	0.85	1.00	0.67	0.85	0.62	0.71	1.00

Figure 6.5 Opinion polarities expressed on each feature for laptop.

In feature-cause extraction, after applying the association rule on adjectives, nouns, and verbs, ten association rules for two distinct features

(NFE) were identified to analyze further to understand the reason of the negative opinion. Table 6.17 represents the number of rules identified per feature and its cause for negative opinions.

Table 6.17 Selected F-C Association Rules for Laptop

Rules	Features (Conclusion)	Cause (Premises)	Sum of Support	Average of Confidence
1	hard drive	is not compute	0.0123	1.00
2	hard drive	replace	0.0247	0.50
3	hard drive	slow	0.0185	0.60
4	hard drive	space need	0.0123	1.00
5	hard drive	upgrade	0.1420	0.88
6	operating system	not have windows	0.0024	0.60
7	operating system	not work	0.0039	0.25
8	operating system	problem	0.0024	0.13
9	operating system	switch application	0.0024	0.60
10	operating system	windows prefer	0.0024	0.75

Then, the individual association rule results were grouped into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of frequencies and confidence was calculated as the average of confidence, where support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by rule confidence measure helped to identify the rules that might be important for designers/engineers, as it took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table.

Table 6.18 Aggregate F-C Association Rules for Laptop

Features	Sum of Support	Average of Confidence	Sum of Support x Confidence
hard drive	0.2099	0.80	0.1737
operating system	0.0134	0.47	0.0059

Presenting this information via graphical form makes it easier for people to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a story with visuals rather than plain words or numbers. For hard-drive, based on the findings (e.g., is not compute, replace, slow, space need, upgrade) the problem seems to be related with memory capacity (or memory size).

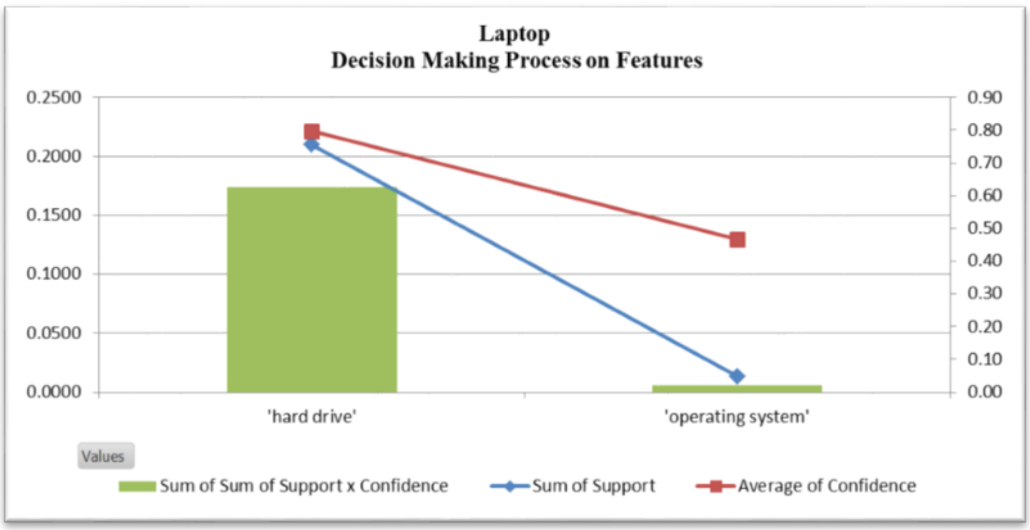


Figure 6.6 Decision making process on features for laptop.

This study shows the relative importance of these three features, which help the situation when manufactures need to prioritize the development or enhancement plan before their next new model release.

Product development managers in manufacturing were expected to have greater return if they enhance the hard drive-related issues as this study reveals the highest sum of support confidence score of 0.1737, followed by operating systems.

6.5 Data Analysis of Mobile Phone

In the feature extraction step, to identify product feature nouns and noun phrases were extracted from the review data as candidate features. To extract noun phrases, the N-gram model method was used, specifically the bi-gram method was applied along with POS tagging (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009). Since not every candidate feature was a real product feature, another constraint on features was placed to filter out non-features and increase the accuracy of feature extraction by employing a knowledge-based heuristic approach; product development managers validated the list of candidate features. As shown in the table, the number of candidate features per sentence (N_C/N_S) was 3.76. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 14.79%, which meant of 5280 candidate features, 664 (17 distinct) were identified as an indistinct product feature by the product development managers (Table 6.19).

Table 6.19 Descriptive Statistic for Mobile Phone

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
mobile phone	352	1405	3.99	20,778	14.79	5280	664	3.76	0.13

In addition, different reviewers often referred to the same product features using different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011). To increase term frequency (TF), similar features were grouped together based on the same meaning. For example, the features button, keyboard, and number were grouped to “keyboard.”

Table 6.20 shows 19 distinct product features identified by the product development managers and their corresponding frequencies in descending order. As is shown, high frequency indicated reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious that “battery” was the feature consumers talked about most when it came to mobile phone, followed by (e.g., Internet, camera, and keyboard).

Table 6.20 Distinct Product Features for Mobile Phone

Num (i)	Feature (F_i – distinct)	TF (F_i-freq)	Num (i)	Feature (F_i – distinct)	TF (F_i-freq)
1	battery	100	11	screen	23
2	Internet	92	12	warranty	22
3	camera/picture	91	13	text	21
4	keyboard	54	14	operating system	19
5	track pad	37	15	Bluetooth	16
6	Wi-Fi	34	16	case	12
7	price	33	17	data plan	7
8	Sim card	32	18	headphone	7
9	accessory	30	19	USB	7
10	navigator	27			

At this point of the analysis, the results show that consumers shared their experiences about 19 distinct features. Afterward, the association rule mining rule was applied to identify the correlation between product features and related opinions, which was the second objective.

Twenty-one thousand six hundred twenty association rules were found by the algorithm, which set the minimum confidence (minconf) at 10% to observe a larger number of rules. Although 21,620 rules were found by the algorithm, not all of the rules were worthwhile to analyze, meaning that only product features and related opinions about them were analyzed in this study. Among the 21,620 rules, only 21 were given as the result of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). First, interesting feature-opinion pairs were identified; then, opinion polarity for each feature-opinion pair was identified by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

For example, for the feature battery, the adjective not bad has a positive orientation, so it was labeled as positive, while the adjectives bad shows a negative orientation. Therefore, each association was examined and labeled as both positive and negative. The following table shows the opinion polarity after assigning the label as negative or positive.

Table 6.21 Selected F-O Association Rules for Mobile Phone

Rules	Feature (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
1	accessory	origin	0.0014	0.38	positive
2	accessory	perfect	0.0018	0.13	positive
3	battery	bad	0.0027	0.26	negative
4	battery	not, bad	0.0023	0.71	positive
5	battery	not, good	0.0027	0.14	negative
6	battery	not, new	0.0023	0.13	negative
7	camera/picture	bad	0.0014	0.13	negative
8	camera/picture	good	0.0069	0.11	positive
9	camera/picture	not, good	0.0041	0.21	negative
10	camera/picture	not, great	0.0023	0.25	negative
11	camera/picture	not, start	0.0018	0.57	negative
12	camera/picture	poor	0.0023	0.71	negative
13	camera/picture	terrible	0.0018	0.31	negative
14	camera/picture	weak	0.0014	0.21	negative
15	Internet	poor	0.0018	0.57	negative
16	Internet	slow	0.0023	0.45	negative
17	Internet	terrible	0.0023	0.38	negative
18	keyboard	full	0.0018	0.25	positive
19	price	good	0.0073	0.12	positive
20	screen	white	0.0014	0.43	negative
21	track pad	amaz	0.0014	0.2	positive

After classifying the sentiment of adjectives into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values regarding its feature using the opinion polarity score equation discussed in Chapter 4.

For instance, the feature battery shows the sum of all of the positive support values was 0.0023, and the sum of all of the negative values was 0.0077. These two values were used to identify the overall opinion or impression of product features. In the feature battery, overall the opinion polarity score was calculated as −0.0054 (negative). The opinion polarity score of each feature was determined by calculating the difference between the sum of all positive support values and the sum of all negative support values.

The table shows the summary of the association rules for the distinct features, and demonstrates the feature-wise frequencies and final opinion polarity.

Table 6.22 Aggregated Opinion Polarities on Features for Mobile Phone

Num (i)	Features (Conclusion)	Sum of Positive Opinion	Sum of Negative Opinion	Opinion Polarity Score (+/-)	Opinion Polarity
1	accessory	0.0032		0.0032	positive
2	battery	0.0023	0.0077	−0.0054	negative
3	camera/picture	0.0069	0.0151	−0.0082	negative
4	Internet		0.0064	−0.0064	negative
5	keyboard	0.0018		0.0018	positive
6	price	0.0073		0.0073	positive
7	screen		0.0014	−0.0014	negative
8	track pad	0.0014		0.0014	positive

When it comes to mobile phones, among eight distinct features (N_{FO}), the features accessory, keyboard, price, and track pad represented positive opinions, while the features battery, camera, Internet, and screen represented negative opinions.

For visual interpretation, Figure 6.7 shows summarized positive and negative user opinions on the product features in bar chart form. The solid line portion in the bar represents positive opinions and the checkered pattern portion represents negative opinions. For example, battery and camera were identified as negative based on the overall opinion polarity. However, these features also have positive aspects as well. For camera as an identified feature, 31% of reviewers who talked about this feature expressed a positive opinion, while 69% of the reviewers expressed negative opinions. Broadly speaking, this result indicates that the camera could be analyzed for further development or improvement. A similar result can be seen for battery.

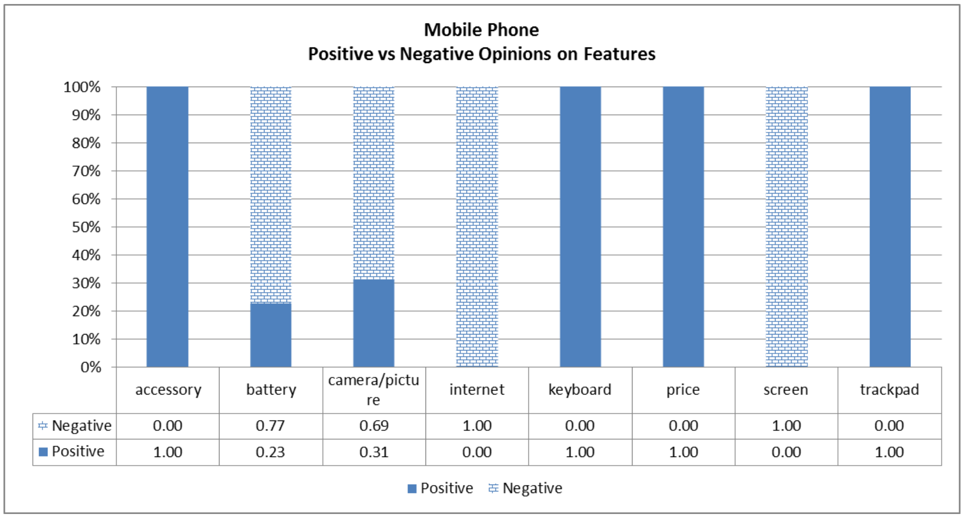


Figure 6.7 Opinion polarities expressed on each feature for mobile phone.

In feature-cause extraction, after applying the association rule on adjectives, nouns, and verbs, seven association rules for four distinct features (N_{FE}) were identified to analyze further as a way to understand the reason of the negative opinion. Table 6.23 represents the number of rules identified per feature and its cause for negative opinions.

Table 6.23 Selected F-C Association Rules for Mobile Phone

Rules	Features (Conclusion)	Cause (Premises)	Sum of Support	Average of Confidence
1	battery	not, new	0.0023	0.13
2	camera/picture	not, good, flash	0.0014	1.00
3	Internet	not, access	0.0018	0.80
4	Internet	not, support	0.0018	0.50
5	Internet	slow	0.0023	0.45
6	screen	side	0.0014	0.43
7	screen	white	0.0014	0.43

Then, the individual association rule results were presented into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of the frequencies, and confidence was calculated as the average of confidence. Support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by rule confidence measure helped to identify the rules that might be important for designers/engineers, as it took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table.

Table 6.24 Aggregate F-C Association Rules for Mobile Phone

Features	Sum of Support	Average of Confidence	Sum of Support x Confidence
battery	0.0023	0.13	0.0003
camera/picture	0.0014	1.00	0.0014
Internet	0.0060	0.58	0.0034
screen	0.0027	0.43	0.0012

Presenting this information via graphical form makes it easier for people to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a story with visuals rather than plain words or numbers. For the Internet, the issue was a connection problem (slow, not access, not support).

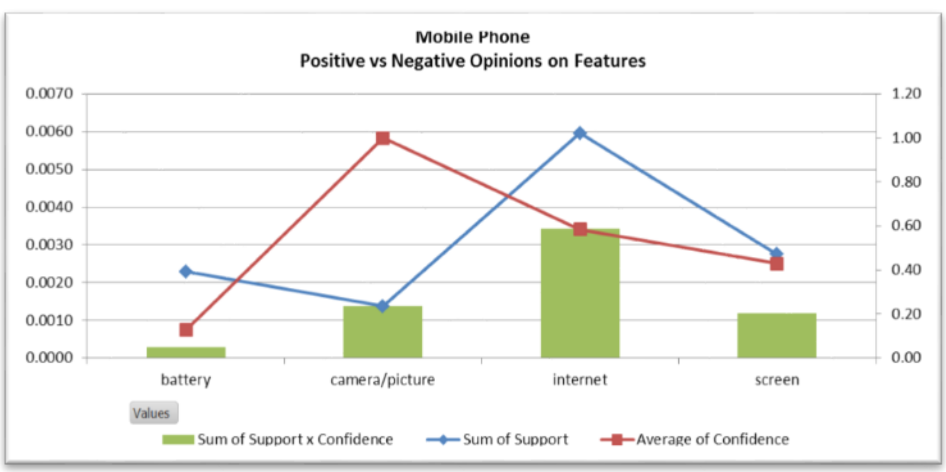


Figure 6.8 Decision making process on features for mobile phone.

This study shows the relative importance of four features that help the situation when manufactures need to prioritize the development or enhancement plan before their next new model release.

Product development managers in manufacturing were expected to have greater return if they address Internet access issues, as this study reveals the highest sum of support was the confidence score of 0.0034, followed by camera, screen, and battery.

6.6 Data Analysis of Sedan

In the feature extraction step, to identify product features, nouns and noun phrases were extracted from the review data as candidate features. To extract noun phrases, the N-gram model method was used, specifically the bi-gram method was applied along with POS tagging (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009). Since not every candidate feature was a real product feature, another constraint was placed on features to filter out non-features and increase the accuracy of feature extraction by employing a knowledge-based heuristic approach; product development managers validated the list of candidate features. As shown in the table, the number of candidate feature per sentence (N_C/N_S) was 2.34. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 12.09%, which meant of 41449 candidate features, 8014 (29 distinct) were identified as an indistinct product feature by the product development managers (Table 6.3.1).

Table 6.25 Descriptive Statistic for Sedan

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
sedan	749	3428	4.58	41,449	12.09	8014	1161	2.34	0.14

In addition, different reviewers often referred to the same product features using different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011). To increase term frequency (TF), similar features were grouped together based on the same meaning. For example, the features cylinder, motor mount, motor, engine, and head gasket were grouped to "engine."

Table 6.26 shows 29 distinct product features identified by the product development managers and their corresponding frequencies in descending order. As is shown, high frequency indicates reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious "fuel consumption" was the features consumers talked about most when it came to sedans, followed by (e.g., engine, transmission, and tires).

Table 6.26 Distinct Product Features for Sedan

Num (i)	Feature (F_i – distinct)	TF (F_i- freq)	Num (i)	Feature (F_i – distinct)	TF (F_i- freq)
1	fuel consumption	405	16	Window	16
2	engine	108	17	cruise control	14
3	transmission systems	85	18	car price	10
4	tire and wheel	79	19	exhaust system	9
5	brake	52	20	serpentine belt	9
6	seat/comfort	50	21	suspension	9
7	alternator/starter	43	22	engine light	8
8	door handle	38	23	maintenance	8
9	oil change	37	24	water pump	8
10	battery	28	25	filter	7
11	heating system	28	26	hood	6
12	audio	25	27	watch and clock	6
13	trunk	22	28	oil filter	5
14	interior size	21	29	spark plug	5
15	body	20			

At this point of the analysis, the results show that consumers shared their experiences on 29 distinct features. Afterward, the association rule mining rule was applied to identify correlations between product features and related opinions.

Seven thousand one hundred seventy-eight association rules were found by the algorithm, which set the minimum confidence (minconf) at 10% to observe a larger number of rules. Although 7178 rules were found by the algorithm, not all of the rules were worthwhile to analyze, meaning that only product features and related opinions on them were analyzed in the study. Among the 7178 rules, only 25 were given as the result of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). First, interesting feature-opinion pairs were identified; then, opinion polarity for each feature-opinion pair was identified by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

For an example of feature picture quality, the adjective clear had a positive orientation, so it was labeled as positive, while the adjective bad showed a negative orientation. Therefore, each association was examined and labeled as positive or negative. The following table shows the opinion polarity after assigning the label as negative or positive.

Table 6.27 Selected F-O Association Rules for Sedan

Rules	Conclusion	Premises	Support	Confidence	Opinion Polarity
1	alternator/starter	repair	0.0017	0.10	negative
2	audio	not, like	0.0012	0.23	negative
3	brake	change	0.0028	0.25	negative
4	brake	mainten	0.0024	0.12	negative
5	engine	not good	0.0012	0.11	negative
6	filter	change	0.0012	0.10	negative
7	fuel consumption	average	0.0031	0.93	negative
8	fuel consumption	bad	0.0012	0.13	negative
9	fuel consumption	cost	0.0021	0.26	negative
10	fuel consumption	issue	0.0014	0.19	negative
11	fuel consumption	low	0.0028	0.38	negative
12	fuel consumption	not good	0.0026	0.24	negative
13	fuel consumption	not great	0.0031	0.37	negative
14	fuel consumption	not like	0.0024	0.45	negative
15	fuel consumption	problem	0.0090	0.20	negative
16	fuel consumption	amaz	0.0024	0.33	positive
17	fuel consumption	good	0.0194	0.32	positive
18	fuel consumption	great	0.0262	0.42	positive
19	fuel consumption	like	0.0050	0.23	positive
20	heating system	cold	0.0019	0.53	negative
21	interior size	plenty	0.0012	0.38	positive
22	seat/comfort	not comfort	0.0012	0.19	negative
23	seat/comfort	comfort	0.0021	0.18	positive
24	trunk	big	0.0014	0.33	positive
25	trunk	space	0.0012	0.42	positive

After classifying the sentiments of adjectives into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values concerning its feature using the opinion polarity score equation discussed in Chapter 4.

For instance, the feature fuel consumption showed the sum of all of the positive support values was 0.0529, and the sum of all of the negative values was 0.0276. These two values were used to identify the overall opinion or impression of the product features. In the feature fuel consumption, the overall opinion polarity score was calculated as 0.0253 (positive). The opinion polarity score of each feature was determined by calculating the difference between the sum of all of the positive support values and the sum of all of the negative support values. The table shows the summary of the association rules for the distinct features and demonstrates the feature-wise frequencies and final opinion polarity.

Table 6.28 Aggregated Opinion Polarities on Features for Sedan

Num (i)	Features (Conclusion)	Sum of Positive Opinion	Sum of Negative Opinion	Opinion Polarity Score (+/-)	Opinion Polarity
1	alternator/starter		0.0017	−0.0017	negative
2	audio		0.0012	−0.0012	negative
3	brake		0.0052	−0.0052	negative
4	engine		0.0012	−0.0012	negative
5	filter		0.0012	−0.0012	negative
6	fuel consumption	0.0529	0.0276	0.0253	positive
7	heating system		0.0019	−0.0019	negative
8	interior size	0.0012		0.0012	positive
9	seat/comfort	0.0021	0.0012	0.0009	positive
10	trunk	0.0026		0.0026	positive

When it comes to sedans, among ten distinct features (N_{FO}) represented positive opinions, while the features alternator, audio, brake, engine, filter, and heating system represented negative opinions.

For visual interpretation, Figure 6.9 shows positive and negative summarized user opinions on the product features in bar chart form. The solid line portion in the bar represents positive opinions and the checkered pattern portion represents negative opinions. For example, fuel consumption and seat were identified as positive based on overall opinion polarity. However, these features have also negative aspects as well. For fuel consumption as an identified feature, 66% of reviewers who talked about fuel consumption expressed a positive opinion, while 34% of the reviewers expressed negative opinions. Broadly speaking, this result indicated that fuel consumption could be analyzed for further development or improvement. A similar result can be seen for seat/comfort.

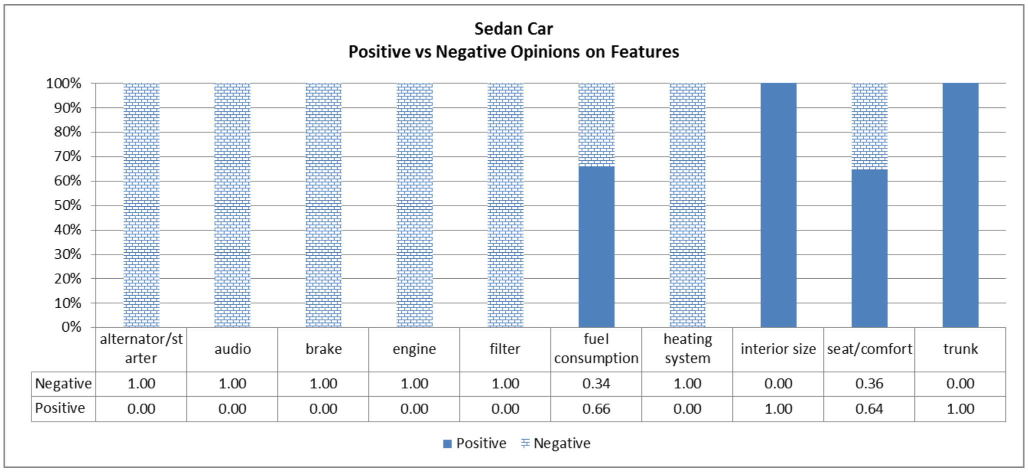


Figure 6.9 Opinion polarities expressed on each feature for sedan.

In feature-cause extraction, after applying the association rule on adjectives, nouns, and verbs, 15 association rules for four distinct features (N_{FE}) were identified to analyze further to understand the reasons for the negative opinions. Table 6.29 represents the number of rules identified per feature and its cause for negative opinions.

Table 6.29 Selected F-C Association Rules for Sedan

Rules	Features (Conclusion)	Cause (Premises)	Support	Confidence
1	alternator/starter	belt change	0.0204	1.00
2	alternator/starter	repair rod	0.0204	1.00
3	alternator/starter	replace belt	0.0612	1.00
4	alternator/starter	replace rod	0.0816	1.00
5	brake	buy strut	0.0235	1.00
6	brake	change	0.1647	1.00
7	brake	front replace	0.0353	0.60
8	brake	need pump	0.0235	1.00
9	brake	need strut	0.0235	1.00
10	brake	replace pump	0.0353	1.00
11	engine	problem mount	0.0164	0.67
12	engine	replace head	0.0164	1.00
13	engine	replace mount	0.0246	1.00
14	engine	replace time belt	0.0164	1.00
15	heating system	cold	0.0019	0.53

Then, the individual association rule results were grouped into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of frequencies, and confidence was calculated as the average of confidence, where support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by rule confidence measure helped to identify the rules that might be important for designers/engineers, as it took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table.

Table 6.30 Aggregated F-C Association Rules for Sedan

Conclusion	Sum of Support	Average of Confidence	Sum of Sum of Support x Confidence
'alternator/starter	0.1837	1.00	0.1837
brake	0.3059	0.93	0.2918
engine	0.0738	0.92	0.0683
heating system	0.0019	0.53	0.0010

Presenting this information via graphical form makes it easier for people to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a story with visuals rather than plain words or numbers. For brakes, the problem was brake related issues (e.g., buy strut, change, front replace, need pump, and need strut).

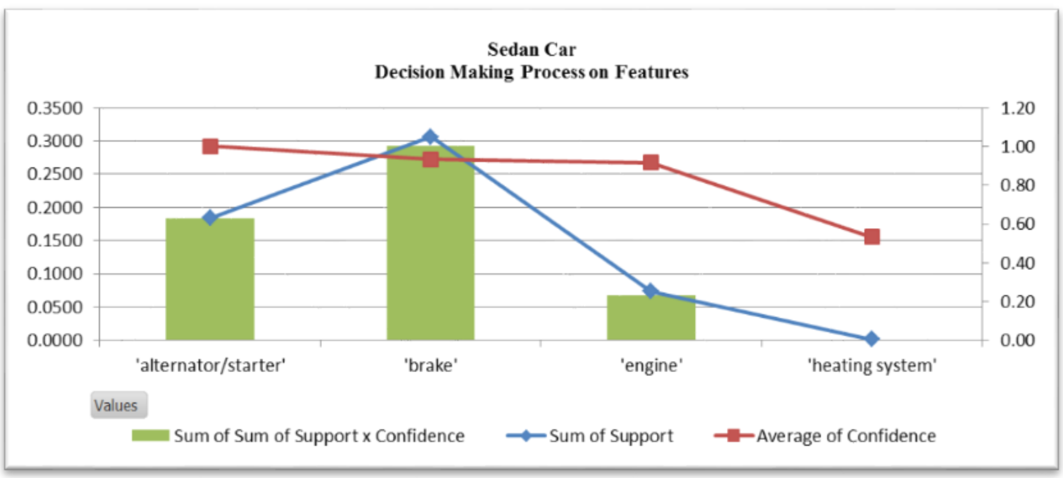


Figure 6.10 Decision making processes on features for sedan.

This study shows the relative importance of these four features, which help the situation when manufactures need to prioritize the development or enhancement plan before their next new model release.

Product development managers in manufacturing were expected to have greater return if they address brake issues, as this study reveals the highest sum of support was a confidence score of 0.2918, followed by alternator, engine, and heating system.

6.7 Data Analysis of Mobile Phone Service Provider

In the feature extraction step, to identify product features, nouns and noun phrases were extracted from the review data as candidate features. To extract noun phrases, the N-gram model method was used, specifically the bi-gram method was applied along with POS tagging (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009). Since not every candidate feature was a real product features, another constraint was placed on features to filter out non-features and increase the accuracy of feature extraction by employing a knowledge-based heuristic approach; product development managers validated the list of candidate features. As shown in the table, the number of candidate feature per sentence (N_C/N_S) was 3.43. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 16.72%, which meant of 120,384 candidate features, 2831 (27 distinct) were identified as indistinct product features by the product development managers (Table 6.31).

Table 6.31 Descriptive Statistic for Mobile Phone Service Provider

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
mobile phone service provider	847	7200	8.50	120,384	16.72	24,704	2831	3.43	0.11

In addition, different reviewers often referred to the same product features using different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011). To increase term frequency (TF), similar features were grouped together based on the same meaning. For example, the features service provider and carrier were grouped to “service provider.”

Table 6.32 shows 27 distinct product features identified by the product development managers and their corresponding frequencies in descending order. As is shown, high frequency indicated reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious “customer service” was the features consumers talked about most when it came to mobile phone service providers, followed by (e.g., contract, billing, and phone plan).

Table 6.32 Distinct Product Features for Mobile Phone Service Provider

Num (i)	Feature (F_i – distinct)	TF (F_i- freq)	Num (i)	Feature (F_i – distinct)	TF (F_i- freq)
1	customer service	1004	15	Sim card	29
2	contract/contract termination	375	16	service charge	28
3	billing	357	17	technical support	22
4	phone plan	209	18	discount	20
5	Internet and email	143	19	password	18
6	text message	132	20	pay phone	16
7	coverage	71	21	sales	14
8	screen/GUI	67	22	service time	13
9	service provider	64	23	phone upgrade	12
10	cell phone upgrades	52	24	voicemail	12
11	data plan	48	25	language options	11
12	insurance	36	26	mobile hotspot	7
13	warranty	35	27	shipping	5
14	phone signal	31			

At this point in the analysis, the results show that consumers shared their experiences on 27 distinct features. Afterward, the association rule mining rule was applied to identify correlations between product features and related opinions, which was the second objective of this study.

Fifteen thousand three hundred sixteen association rules were found by the algorithm, which set the minimum confidence (minconf) at 10% to observe a larger number of rules. Although 15,316 rules were found by the algorithm, not all of the rules were worthwhile to analyze, meaning that only product features and related opinions on them were analyzed in this study. Among the 15,316 rules, only 34 were given as the result of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). First, interesting feature-opinion pairs were identified; then, opinion polarity was identified for each feature-opinion pair by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

For an example of feature picture quality, the adjective clear had a positive orientation, so it was labeled positive, while the adjective bad showed a negative orientation. Therefore, each association was examined and labeled as positive and negative. The following table shows the opinion polarity after assigning the label as negative or positive.

Table 6.33 Selected F-O Association Rules for Mobile Phone Service Provider

Rule	Features (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
1	billing	past	0.0011	0.14	negative
2	contract/contract termination	break	0.0012	0.65	negative
3	contract/contract termination	end	0.0039	0.40	negative
4	contract/contract termination	fee	0.0037	0.22	negative
5	contract/contract termination	switch	0.0012	0.16	negative
6	coverage	not, service, area	0.0011	0.50	negative
7	customer service	bad	0.0118	0.42	negative
8	customer service	complaint	0.0011	0.16	negative
9	customer service	good	0.0048	0.25	positive
10	customer service	great	0.0015	0.18	positive
11	customer service	help	0.0033	0.20	positive
12	customer service	hold	0.0027	0.24	negative
13	customer service	horrible	0.0040	0.39	negative
14	customer service	issue	0.0040	0.16	negative
15	customer service	nice	0.0011	0.23	positive
16	customer service	not, bad	0.0013	0.32	positive
17	customer service	not, good	0.0013	0.31	negative
18	customer service	not, help	0.0013	0.19	negative
19	customer service	not, spoke	0.0012	0.52	negative
20	customer service	poor	0.0032	0.58	negative
21	customer service	problem	0.0049	0.17	negative
22	customer service	resolve	0.0017	0.29	positive
23	customer service	rude	0.0028	0.35	negative
24	customer service	suck	0.0017	0.25	negative
25	customer service	terrible	0.0026	0.45	negative
26	customer service	wrong	0.0013	0.20	negative
27	data plan	unlimited	0.0011	0.18	positive
28	phone plan	cost	0.0012	0.29	negative
29	phone plan	fee	0.0025	0.15	negative
30	phone plan	unlimited	0.0023	0.38	positive
31	service charge	fee, termin	0.0024	0.96	negative
32	service provider	switch	0.0011	0.15	negative
33	text message	unlimited	0.0017	0.29	positive

After classifying the sentiment of each adjective into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values concerning its feature using the opinion polarity score equation discussed in Chapter 4.

For instance, the feature customer service showed the sum of all of the positive support values was 0.0137, and the sum of all of the negative values was 0.0438. These two values were used to identify the overall opinion or impression of the product features. In the feature customer service, the overall opinion polarity score was calculated as −0.0301 (negative). The opinion polarity score of each feature was determined by calculating the difference between the sum of the positive support value and the sum of the negative support values. The table shows the summary of the association rules for the distinct features, and demonstrates the feature-wise frequencies and final opinion polarity.

Table 6.34 Aggregate Opinion Polarities on Features for Mobile Phone Service Provider

Num (i)	Features (Conclusion)	Sum of Positive Opinion	Sum of Negative Opinion	Opinion Polarity Score (+/-)	Opinion Polarity
1	billing		0.0011	−0.0011	negative
2	contract/contract termination		0.0099	−0.0099	negative
3	coverage		0.0011	−0.0011	negative
4	customer service	0.0137	0.0438	−0.0301	negative
5	data plan	0.0011		0.0011	positive
6	phone plan	0.0023	0.0037	−0.0014	negative
7	service charge		0.0024	−0.0024	negative
8	service provider		0.0011	−0.0011	negative
9	text message	0.0017		0.0017	positive

When it comes to the mobile phone service provider, among nine distinct features (N_{FO}), the features text message and data plan represented positive opinions, while the features billing, contract, coverage, customer service, phone plan, service charge, and service provider represented negative opinions.

For a visual interpretation, Figure 6.11 shows positive and negative summarized user opinions on the product features in bar chart form. The solid line portion in the bar represents positive opinions and the checkered pattern portion represents negative opinions. For example, customer service, and phone plan were identified as negative based on the overall opinion polarity. However, these features also have positive aspects as well. For customer service as an identified feature, 24% of reviewers who talked about customer service expressed positive opinions, while 76% of the reviewers expressed negative opinions. Broadly speaking, this result indicates that customer service could be analyzed for further development or improvement. A similar result can be seen in the phone plan.

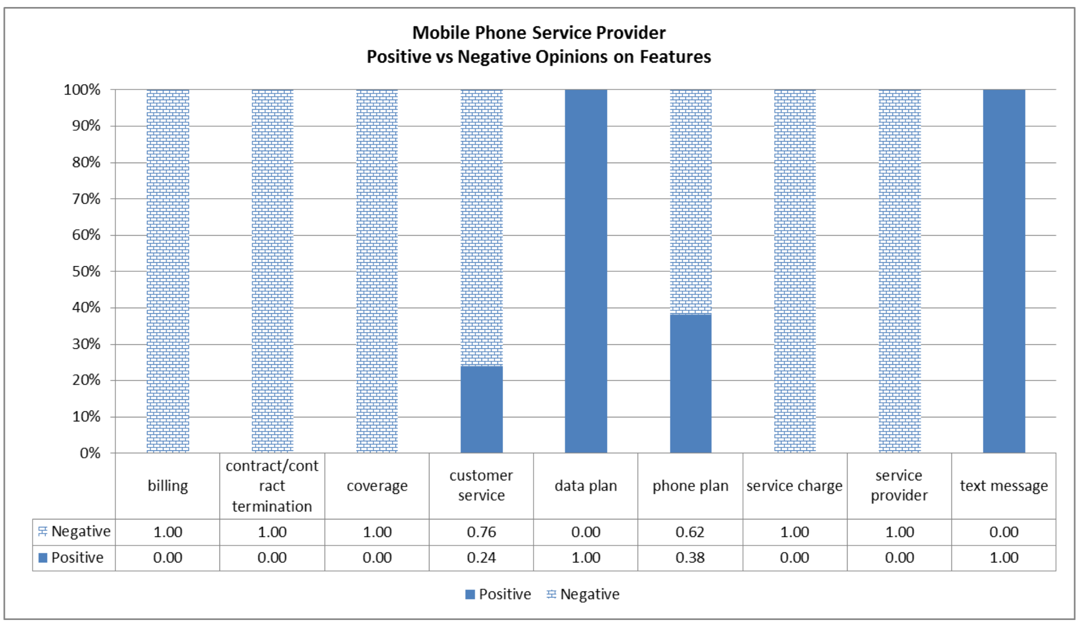


Figure 6.11 Opinion polarities on each feature for mobile phone service provider.

In feature-cause extraction, after applying the association rule on adjectives, nouns, and verbs, 22 association rules for five distinct features (N_{FE}) were identified to be analyzed further to understand the reason of the negative opinions. Table 6.35 represents the number of rules identified per feature and its cause for negative opinions.

Table 6.35 Selected F-C Association Rules for Mobile Phone Service Provider

Rules	Features (Conclusion)	Cause (Premises)	Support	Confidence
1	contract/contract termination	break	0.0057	0.64
2	contract/contract termination	cancel	0.0215	0.38
3	contract/contract termination	not sign	0.0053	0.46
4	contract/contract termination	price	0.0053	0.34
5	customer service	call time problem	0.0014	0.68
7	customer service	hung	0.0018	0.38
8	customer service	do not know	0.0012	0.27
9	customer service	not explain	0.0012	0.34
10	customer service	not help	0.0014	0.19
11	customer service	not resolve	0.0013	0.27
12	customer service	not transfer	0.0013	0.28
13	customer service	not understand	0.0012	0.22
14	customer service	rude	0.0028	0.35
15	customer service	wait	0.0019	0.14
17	billing	fix	0.0013	0.18
18	billing	past	0.0015	0.19
19	billing	credit card	0.0022	0.95
20	coverage	not service area	0.0011	0.50
21	phone plan	call time	0.0021	0.16
22	phone plan	rate	0.0023	0.43

Then, the individual association rule results were grouped into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of the frequencies, and confidence was calculated as the average of confidence, where support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by rule confidence measure helped to identify the rules that might be important for designers/engineers, as it took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table.

Table 6.36 Aggregate F-C Association Rules for Mobile Phone Service Provider

Features	Sum of Support	Average of Confidence	Sum of Support x Confidence
contract/contract termination	0.0377	0.46	0.0160
customer service	0.0170	0.32	0.0054
billing	0.0082	0.55	0.0054
coverage	0.0011	0.50	0.0005
phone plan	0.0043	0.29	0.0013

Presenting this information via graphical form makes it easier for people to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a story with visuals rather than plain words or numbers. For contract, the problems were early termination, and not renew the contract due to price constraint or service quality (e.g., break, cancel, not sign, and price).

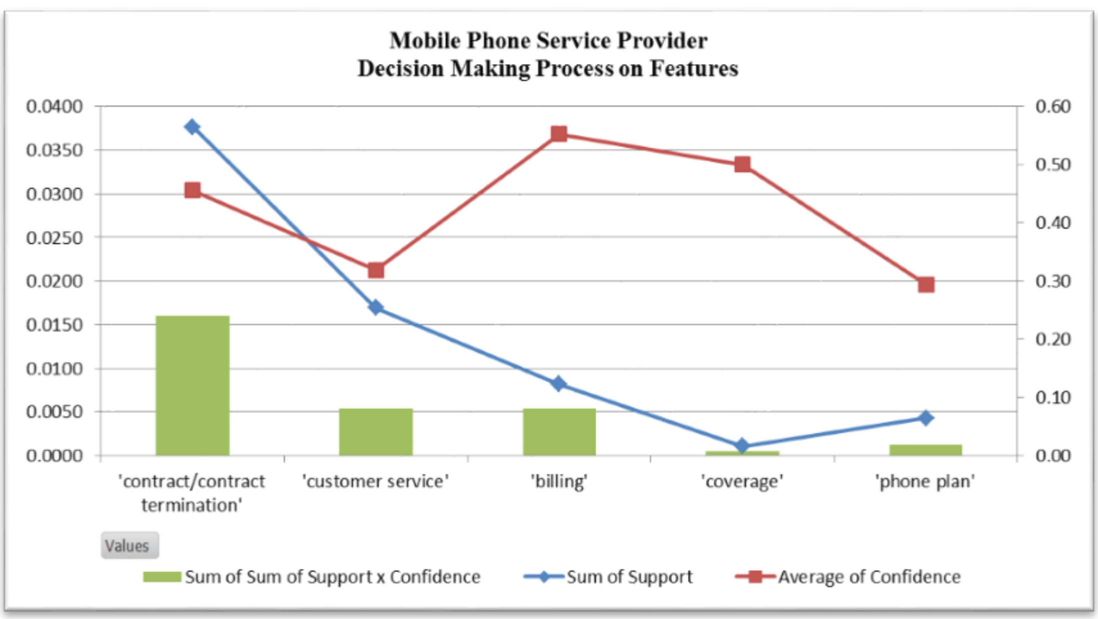


Figure 6.12 Decision making process on features for mobile phone service provider.

This study shows the relative importance of these four features, which helps the situation when manufactures need to prioritize the development or enhancement plan before their next new model release.

Product development managers in manufacturing are expected to have greater return if they address contract-related issues, as this study revealed the highest sum of support was a confidence score of 0.0160, followed by billing, customer service, service area coverage, and phone plan.

6.8 Data Analysis of Airline Travel

In the feature extraction step, to identify product features, nouns and noun phrases were extracted from the review data as candidate features. To extract noun phrases, the N-gram model method was used, specifically the bi-gram method was applied along with POS tagging (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009). Since not every candidate feature was a real product feature, another constraint on features was placed to filter out non-features and increase the accuracy of feature extraction by employing a knowledge-based heuristic approach; product development managers validated the list of candidate features. As shown in the table, the number of candidate feature per sentence (N_C/N_S) was 3.22. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 16.65%, which meant of 17,269 candidate features, 2617 (24 distinct) were identified as an indistinct product feature by the product development managers (Table 6.3.1).

Table 6.37 Descriptive Statistic for Airline Travel

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
airline travel	570	5370	9.42	89,405	16.65	17,269	2617	3.22	0.15

In addition, different reviewers often referred to the same product features using different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011). To increase term frequency (TF), similar features were grouped together based on the same meaning. For example, the features crew, flight attendant, and stewardess were grouped to "flight attendant."

Table 6.1.2 shows 24 distinct product features identified by the product development managers and their corresponding frequencies in descending order. As is shown, high frequency indicated reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious "baggage check-in/claim" was the feature consumers talked about most when it came to airline travel, followed by (e.g., customer service, travel agent, and airport security).

Table 6.38 Distinct Product Features for Airline Travel

Num (i)	Feature (F_i – distinct)	TF (F_i- freq)	Num (i)	Feature (F_i – distinct)	TF (F_i- freq)
1	baggage check-in/claim	515	13	boarding	64
2	customer service	440	14	international travel	24
3	travel agent	383	15	disability access	21
4	airport security/facilities	177	16	in cabin pets	21
5	refund	131	17	credit card	17
6	meal	126	18	economy/business class	14
7	gate/flight changes	120	19	leg room	13
8	seating	120	20	children/infants	10
9	flight attendant	119	21	bathroom	8
10	flight schedule	108	22	missing flight	8
11	arrival/departure	91	23	baggage fee	6
12	reservation	76	24	connectivity/transfer	5

At this point in the analysis, the results show that consumers shared their experiences on 24 distinct features. Afterward, the association rule mining rule was applied to identify correlations between product features and related opinions, which was the second objective of this study.

Two thousand four hundred twenty-eight association rules were found by the algorithm, which set the minimum confidence (minconf) at 10% to observe a larger number of rules. Although 2428 rules were found by the algorithm, not all of the rules were worthwhile to analyze, meaning that only product features and related opinions about them were analyzed in this study. Among the 2428 rules, only 30 were given as the result of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). First, interesting feature-opinion pairs were identified; then, opinion polarity for each feature-opinion pair was identified by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

For example, of feature baggage claim/check-in, the adjective not problem has a positive orientation, so it was labeled as positive, while the adjectives problem and line show a negative orientation. Therefore, each association was examined and labeled as positive and negative. The following table shows the opinion polarity after assigning the label as negative or positive.

Table 6.39 Selected F-O Association Rules for Airline Travel

Rule	Features (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
1	airport security/facilities	line	0.0017	0.13	negative
2	airport security/facilities	problem	0.0014	0.10	negative
3	baggage check-in/claim	line	0.0017	0.13	negative
4	baggage check-in/claim	problem	0.0019	0.14	negative
5	baggage check-in/claim	not problem	0.0013	0.22	positive
6	customer service	bad	0.0042	0.21	negative
7	customer service	complaint	0.0016	0.25	negative
8	customer service	hold	0.0019	0.19	negative
9	customer service	issue	0.0011	0.13	negative
10	customer service	not, answer	0.0014	0.30	negative
11	customer service	not, good	0.0011	0.28	negative
12	customer service	not help	0.0017	0.14	negative
13	customer service	poor	0.0031	0.71	negative
14	customer service	rude	0.0034	0.21	negative
15	customer service	terrible	0.0019	0.38	negative
16	customer service	good	0.0020	0.14	positive
17	customer service	great	0.0023	0.31	positive
18	customer service	help	0.0030	0.11	positive
19	customer service	nice	0.0013	0.13	positive
20	customer service	not poor	0.0013	0.73	positive
21	flight attendant	rude	0.0017	0.10	negative
22	travel agent	issue	0.0011	0.13	negative
23	travel agent	not help	0.0030	0.24	negative
24	travel agent	problem	0.0016	0.12	negative
25	travel agent	rude	0.0053	0.32	negative
26	travel agent	good	0.0019	0.13	positive
27	travel agent	great	0.0011	0.15	positive
28	travel agent	help	0.0069	0.25	positive
29	travel agent	nice	0.0016	0.17	positive
30	travel agent	not problem	0.0011	0.19	positive

After classifying the sentiment of the adjective into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values regarding its feature using the opinion polarity score equation discussed in Chapter 4.

For instance, the feature customer service shows the sum of all of the positive support values was 0.0099, and the sum of all of the negative values was 0.0215. These two values were used to identify overall opinion or impression of product features. In the feature (customer service), the overall opinion polarity score was calculated as –0.0116 (negative). The opinion polarity score of each feature was determined by calculating the difference between the sum of all of the positive support values and the sum of all of the negative support values. The table shows the summary of the association rules for the distinct features and demonstrates the feature-wise frequencies and final opinion polarity.

Table 6.40 Aggregated Opinion Polarities on Features for Airline Travel

Num (i)	Features (Conclusion)	Sum of Positive Opinion	Sum of Negative Opinion	Opinion Polarity Score (+/-)	Opinion Polarity
1	airport security/facilities		0.0031	−0.0031	negative
2	baggage check-in/claim	0.0013	0.0036	−0.0023	negative
3	customer service	0.0099	0.0215	−0.0116	negative
4	flight attendant		0.0017	−0.0017	negative
5	travel agent	0.0125	0.0110	0.0016	positive

When it comes to airline travel, among nine features (N_{FO}), the feature travel agent represents positive opinions, while the features airport security, baggage check-in, and customer service flight attendant represented negative opinions.

For visual interpretation, Figure 6.3.1 shows summarized positive and negative user opinions on the product features in bar chart form. The solid line portion in the bar represents positive opinions and the checkered pattern portion represents negative opinions. For example, travel agent was identified as positive based on the overall opinion polarity. However, these features also have negative aspects as well. For travel agent as an identified feature, 53% of reviewers who talked about customer service expressed positive opinions, while 47% of the reviewers expressed negative opinions.

Broadly speaking, this result indicates that the service level of travel agents could be analyzed for further development or improvement.

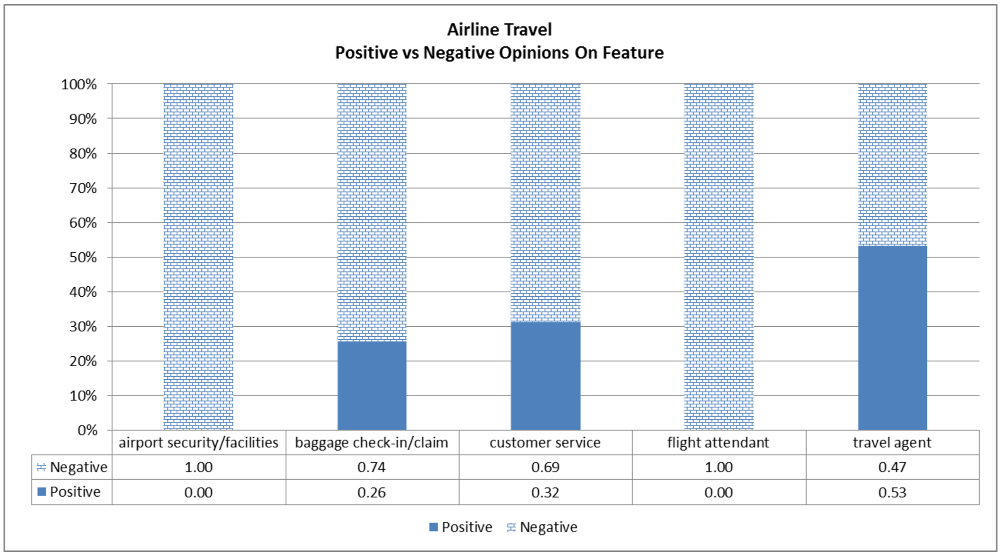


Figure 6.13 Opinion polarities expressed on each feature for airline travel.

In feature-cause extraction, after applying the association rule on adjectives, nouns, and verbs, 16 association rules for five distinct features (N_{FE}) were identified to analyze further to understand the reasons of the negative opinion. Table 6.41 represents the number of rules identified per feature and its cause for negative opinions.

Table 6.41 Selected F-C Association Rules for Airline Travel

Rules	Features (Conclusion)	Cause (Premises)	Support	Confidence
1	airport security/facilities	delay	0.0101	0.45
2	airport security/facilities	line	0.0123	0.13
3	airport security/facilities	miss	0.0134	0.60
4	airport security/facilities	wait	0.0168	0.48
5	flight attendant	not care	0.0041	0.80
6	flight attendant	rude	0.0113	0.42
7	baggage check-in/claim	miss	0.0039	0.23
8	baggage check-in/claim	wait	0.0027	0.11
9	baggage check-in/claim	ruin	0.0011	0.41
10	baggage check-in/claim	brake	0.0013	0.62
11	baggage check-in/claim	damage	0.0022	0.64
12	customer service	hold	0.0019	0.15
13	customer service	wait	0.0034	0.15
14	customer service	ignore	0.0013	0.47
15	travel agent	not get help	0.0011	0.32
16	travel agent	ignore	0.0011	0.41

Then, the individual association rule results were grouped into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of the frequencies, and confidence was calculated as the average of confidence, where support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by rule confidence measure helped to identify the rules that might be important for designers/engineers, as it took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table.

Table 6.42 Aggregated F-C Association Rules for Airline Travel

Conclusion	Sum of Support	Average of Confidence	Sum of Sum of Support x Confidence
airport security/facilities	0.0525	0.42	0.0223
baggage check-in/claim	0.0111	0.40	0.0038
customer service	0.0066	0.26	0.0014
flight attendant	0.0155	0.61	0.0081
travel agent	0.0022	0.36	0.0008

Presenting this information via graphical form makes it easier for people to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a story with visuals rather than plain words or numbers. For baggage check-in/claim, the problem was damage/lost/delayed language (i.e., brake, ruin, wait, and damage).

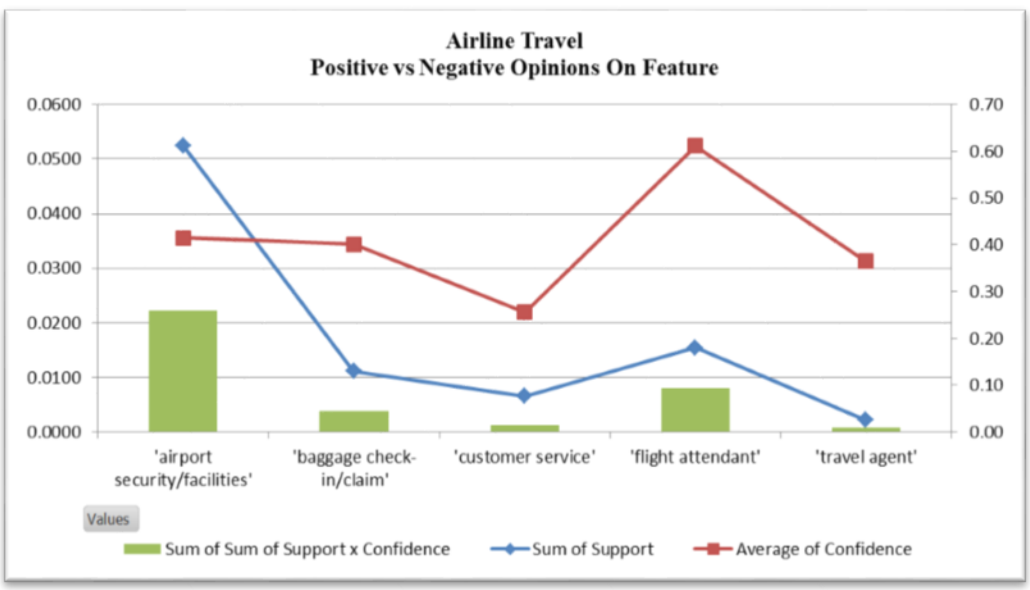


Figure 6.14 Decision making process on features for airline travel.

This study shows the relative importance of these three features, which helps the situation when travel service providers need to prioritize the service quality or service improvement plan before their next new service release.

Management teams in the airline industry were expected to have greater return if they address airport security/facility related issues, as this study reveals the highest sum of support was a confidence score of 0.0223, followed by flight attendants' behavior, baggage related issues, customer service, and travel agents.

6.9 Data Analysis of Restaurant

In the feature extraction step, to identify product features, nouns and noun phrases were extracted from the review data as candidate features. To extract noun phrases, the N-gram model method was used, specifically the bi-gram method was applied along with POS tagging (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009). Since not every candidate feature was a real product feature, another constraint on features was placed to filter out non-features and increase the accuracy of feature extraction by employing a knowledge-based heuristic approach; product development managers validated the list of candidate features. As shown in the table, the number of candidate feature per sentence (N_C/N_S) was 3.59. The ratio of feature nouns in total the total candidate feature nouns (N_F/N_C) was 15.21%, which meant of 178,252 candidate features, 7578 (68 distinct) were identified as an indistinct product feature by the product development managers (Table 6.43).

Table 6.43 Descriptive Statistic for Restaurant

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
restaurant	821	11,718	14.27	178,258	15.21	42,116	7578	3.59	0.18

In addition, different reviewers often referred to the same product features using different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011). To increase term frequency (TF), similar features were grouped together based on the same meaning. For example, the features beef, meat, pork, and steak were grouped to “steak.”

Table 6.44 shows 68 distinct product features identified by the product development managers and their corresponding frequencies in descending order. As is shown, high frequency indicated reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious “seafood” was the feature consumers talked about most when it came to restaurant, followed by (e.g., lunch/dinner, desert, and décor).

Table 6.44 Distinct Product Features for Restaurant

Num (i)	Feature (F_i – distinct)	TF (F_i- freq)	Num (i)	Feature (F_i – distinct)	TF (F_i- freq)
1	seafood	1339	35	tartar	38
2	lunch/dinner	726	36	soup	31
3	desert	635	37	chickens	28
4	ambiance/décor	484	38	pasta	28
5	service	366	39	squash/zucchini	28
6	steak	283	40	potato	26
7	wine and liquor	280	41	mushroom	25
8	staff	258	42	napkin	24
9	menu	241	43	restroom	24
10	Asian food	214	44	bean	21
11	dishes	212	45	beansprout	21
12	baguette/bread	207	46	tomato	20
13	chef	154	47	vegetable	20
14	lunch menu	121	48	vegetarian	19
15	price	119	49	bacon	18
16	sauce	113	50	cheese	18
17	food	110	51	rosemary	18
18	reservation	106	52	asparagus	16
19	drink	101	53	carrot	15
20	nuts	94	54	cauliflower	13
21	lemon/lime	91	55	entre	13
22	fruit	79	56	vanilla	12
23	appetizers	78	57	coconut	11
24	egg	69	58	hoisin plum	11
25	salad	69	59	potato crisps	11
26	drinks	64	60	Brussels sprout	10
27	portion	60	61	dress code	9
28	tea/coffee	60	62	fig bacon	9
29	butter	54	63	cookies	8
30	salt	49	64	cucumber	8
31	prix menu	42	65	eggplant	8
32	yogurt	40	66	lime broth	8
33	paprika	38	67	radish	8
34	Seating	38	68	cinnamon	7

At this point in the analysis, the results show that consumers shared their experiences on 24 distinct features. Afterward, the association rule mining rule was applied to identify correlations between product features and related opinions, which was the second objective of this study.

Twenty-four thousand eight hundred thirty-five association rules were found by the algorithm, which set the minimum confidence (minconf) as 10% to observe a larger number of rules. Although 24,835 rules were found by the algorithm, not all of the rules were worthwhile to analyze, meaning that only product features and related opinions about them were analyzed in this study. Among the 24,835 rules, only 23 were given as the result of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). First, interesting feature-opinion pairs were identified; then, opinion polarity for each feature-opinion pair was identified by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

For example, for the feature desert, the adjectives passion and good chocolate had a positive orientation, so they were labeled as positive, while the adjective salt shows a negative orientation. Therefore, each association was examined and labeled as positive and negative. The following table shows the opinion polarity after assigning the label as negative or positive.

Table 6.45 Selected F-O Association Rules for Restaurant

Rule	Features (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
1	ambiance/décor	eleg	0.0012	0.45	positive
2	baguette/bread	layer, chive	0.0012	0.90	positive
3	baguette/bread	thin	0.0017	0.47	positive
4	desert	good, chocolate	0.0015	0.88	positive
5	desert	passion	0.0011	0.57	positive
6	desert	salt	0.0017	0.49	negative
7	desert	yuzu	0.0011	0.39	positive
8	lunch/dinner	service, amaz	0.0018	0.43	positive
9	lunch/dinner	service, excel	0.0011	0.36	positive
10	lunch/dinner	service, great	0.0014	0.41	positive
11	price	high	0.0010	0.13	negative
12	sauce	basil	0.0013	0.53	positive
13	seafood	amus	0.0030	0.43	positive
14	seafood	chive	0.0015	0.58	positive
15	seafood	crispy	0.0023	0.35	positive
16	seafood	fresh	0.0043	0.40	positive
17	seafood	raw	0.0038	0.52	negative
18	seafood	tender	0.0020	0.49	positive
19	service	impeccable	0.0046	0.74	positive
20	service	top, notch	0.0011	0.49	positive
21	staff	friendly	0.0019	0.46	positive
22	steak	kobe	0.0024	0.95	positive
23	wine and liquor	sommeli	0.0038	0.47	positive

After classifying the sentiment of the adjective into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values regarding its feature using the opinion polarity score equation discussed in Chapter 4.

For instance, the feature desert shows the sum of all of the positive support values was 0.0036, and the sum of all of the negative values was 0.0017. These two values were used to identify overall opinion or impression of product features. In the feature desert, the overall opinion polarity score was calculated as 0.0019 (positive). The opinion polarity score of each feature was determined by calculating the difference between the sum of all of the positive support values and the sum of all of the negative support values. The table shows the summary of the association rules for the distinct features and demonstrates the feature-wise frequencies and final opinion polarity.

Table 6.46 Aggregated Opinion Polarities on Features for Restaurant

Num (i)	Features (Conclusion)	Sum of Positive Opinion	Sum of Negative Opinion	Opinion Polarity Score (+/-)	Opinion Polarity
1	ambiance/décor	0.0012		0.0012	positive
2	baguette/bread	0.0029		0.0029	positive
3	desert	0.0036	0.0017	0.0019	positive
4	lunch/dinner	0.0043		0.0043	positive
5	price		0.0010	−0.0010	negative
6	sauce	0.0013		0.0013	positive
7	seafood	0.0131	0.0038	0.0093	positive
8	service	0.0057		0.0057	positive
9	staff	0.0019		0.0019	positive
10	steak	0.0024		0.0024	positive
11	wine and liquor	0.0038		0.0038	positive

When it comes to the restaurant, among nine distinct features (N_{FO}), the features ambiance/décor, baguette/bread, desert, lunch/dinner, sauce, seafood, service, staff, steak, and wine/liquor represented positive opinions, while the feature price represented negative opinions.

For visual interpretation, Figure 6.15 shows summarized user opinions such as positive and negative on the product features in bar chart form. The solid line portion in the bar represents positive opinions and the checkered pattern portion represents negative opinions. For example, seafood was identified as positive based on the overall opinion polarity. However, these features also have negative aspects as well. For seafood as an identified feature, 77% of reviewers who talked about customer service expressed positive opinions, while 23% of the reviewers expressed negative opinions. Broadly speaking, this result indicates that seafood perception could be analyzed for further service improvement.

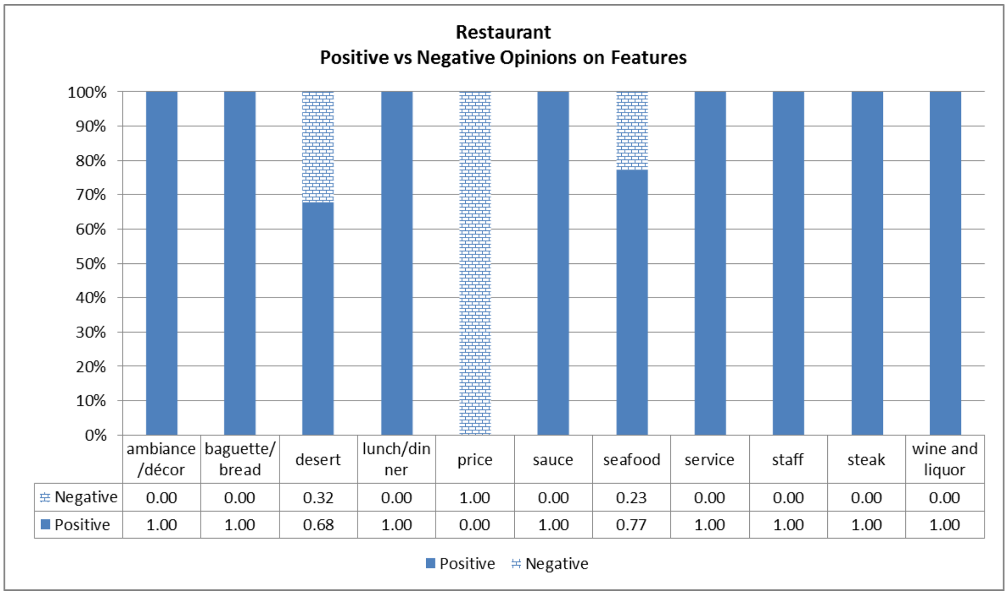


Figure 6.15 Opinion polarities expressed on each feature for restaurant.

In feature-cause extraction, after applying the association rule on adjectives, nouns, and verbs, eight association rules for three distinct features (N_{FE}) were identified to analyze further to understand the reason of the negative opinions. Table 6.47 represents the number of rules identified per feature and its cause for negative opinions.

Table 6.47 Selected F-C Association Rules for Restaurant

Rules	Features (Conclusion)	Cause (Premises)	Support	Confidence
1	price	expensive	0.0281	0.71
2	price	high	0.0449	1.00
3	seafood	escolar	0.0062	0.53
4	seafood	not, cook	0.0107	0.84
5	seafood	not, fresh	0.0038	0.73
6	seafood	overpower	0.0041	0.55
7	seafood	salty	0.0048	0.54
8	desert	shell	0.0011	0.46

Then, the individual association rule results were grouped into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of the frequencies, and confidence was calculated as the average of confidence, where support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by rule confidence measure helped to identify the rules that might be important for designers/engineers, as it took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table.

Table 6.48 Aggregate F-C Association Rules for Restaurant

Feature (Conclusion)	Sum of Support	Average of Confidence	Sum of Support x Confidence
price	0.0730	0.86	0.0650
seafood	0.0296	0.64	0.0198
desert	0.0011	0.46	0.0005

Presenting this information via graphical form makes it easier for people to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a story with visuals rather than plain words or numbers. For seafood, the problems were uncooked fish, not fresh (i.e., escolar, not cook, not fresh, overpower, salty).

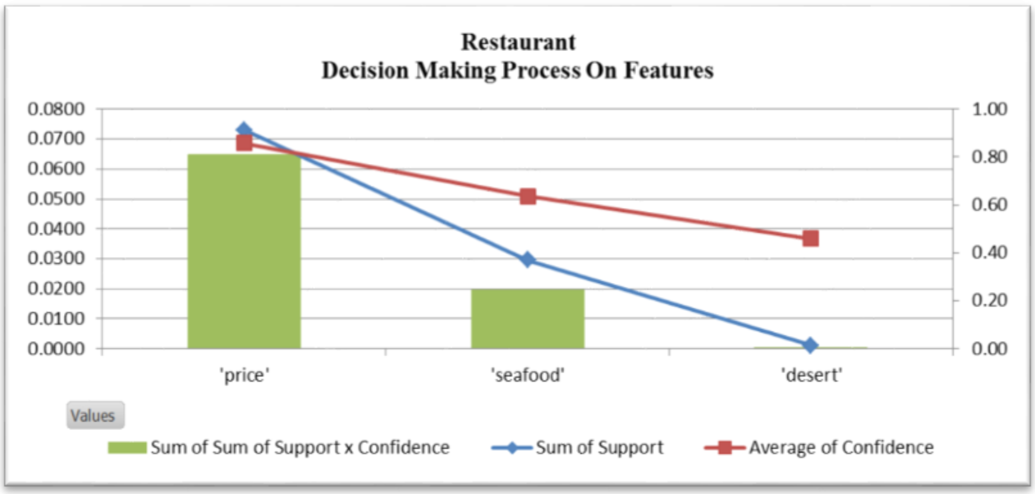


Figure 6.16 Decision making process about features for restaurant.

This study shows the relative importance of these three features, which help the situation when a restaurateur needs to analyze service quality or a service improvement plan.

Restaurant managers in the service domain were expected to have a greater return if they address pricing concerns, as this study reveals the highest sum of support was a confidence score of 0.065, followed by seafood, and dessert respectively.

6.10 Data Analysis of a Hotel

In the feature extraction step, to identify product features, nouns and noun phrases were extracted from the review data as candidate features. To extract noun phrases, the N-gram model method was used, specifically the bi-gram method was applied along with POS tagging (Pang, Lee et al. 2002; Pang and Lee 2008; Weiping and Yuanzhuang 2009). Since not every candidate feature was a real product feature, another constraint on features was placed to filter out non-features and increase the accuracy of feature extraction by employing a knowledge-based heuristic approach; product development managers validated the list of candidate features. As shown in the table, the number of candidate feature per sentence (N_C/N_S) was 3.76. The ratio of feature nouns to candidate feature nouns (N_F/N_C) was 15.39%, which meant of 21,768 candidate features, 2107 (11 distinct) were identified as an indistinct product feature by the product development managers (Table 6.49).

Table 6.49 Descriptive Statistics for the Hotel

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
hotel	528	5788	10.96	89104	15.39	21768	2107	3.76	0.10

In addition, different reviewers often referred to the same product features using different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features (Zhang, Jia et al. 2011). To increase term frequency (TF), similar features were grouped together based on the same meaning. For example, the features coffee maker, gym, fitness center, Internet access, and soap were grouped to "amenities."

Table 6.50 shows 11 distinct product features identified by the product development managers and their corresponding frequencies in descending order. As is shown, high frequency indicated reviewers often talked about that feature. Therefore, they were shown frequently in the reviews. It was obvious that "housekeeping" was the feature consumers talked about most when it came to hotel, followed by (e.g., restaurant, service, front desk).

Table 6.50 Distinct Product Features for Hotel

Num (i)	Feature (F_i – distinct)	TF (F_i- freq)	Num (i)	Feature (F_i – distinct)	TF (F_i- freq)
1	housekeeping	853	7	price	89
2	restaurant	341	8	location	70
3	service	245	9	room service	46
4	front desk	174	10	reservation	32
5	facilities	134	11	loyalty program	6
6	amenities	117			

At this point in the analysis, the results show that consumers shared their experiences on 11 distinct features. Afterward, the association rule mining rule was applied to identify correlations between product features and related opinions, which was the second objective of this study.

Nine thousand six hundred ninety-six association rules were found by the algorithm, which set the minimum confidence (minconf) as 10% to observe a larger number of rules. Although 9696 rules were found by the algorithm, not all of the rules were worthwhile to analyze, meaning that only product features and related opinions on them were analyzed in the study. Among the 9696 rules, only 33 were given as the result of the analysis, which represented the association rules between the opinion (premises) and related features (conclusion). First, interesting feature-opinion pairs were identified; then, opinion polarity for each feature-opinion pair was identified by manually examining and labeling whether each of the adjectives was expressing positive or negative sentiment.

For example, for the feature housekeeping, cleaning and maintenance in hotel room, the adjectives amazing, big, and clean have positive orientations, so they were labeled as positive, while the adjectives bug, noise, and not great show a negative orientation. Therefore, each association was examined and labeled as positive and negative. The following table shows the opinion polarity after assigning the label as negative or positive.

Table 6.51 Selected F-O Association Rules for Hotel

Rule	Features (Conclusion)	Opinion (Premises)	Support	Confidence	Opinion Polarity
1	amenities	free	0.0014	0.16	positive
2	front desk	friendly	0.0024	0.17	positive
3	housekeeping	amazing	0.0077	0.24	positive
4	housekeeping	average	0.0014	0.59	negative
5	housekeeping	bad	0.0018	0.30	negative
6	housekeeping	big	0.0157	0.56	positive
7	housekeeping	clean	0.0200	0.77	positive
8	housekeeping	comfort	0.0133	0.83	positive
9	housekeeping	good	0.0109	0.23	positive
10	housekeeping	great	0.0138	0.24	positive
11	housekeeping	large	0.0017	0.86	positive
12	housekeeping	nice	0.0138	0.45	positive
13	housekeeping	not, good	0.0016	0.23	negative
14	housekeeping	not, great	0.0024	0.43	negative
15	housekeeping	not, nice	0.0027	0.58	negative
16	housekeeping	perfect	0.0014	0.18	positive
17	housekeeping	problem	0.0020	0.17	negative
18	housekeeping	quiet	0.0061	0.64	positive
19	housekeeping	super	0.0016	0.39	positive
20	location	amazing	0.0086	0.27	positive
21	location	beauty	0.0011	0.32	positive
22	location	comfort	0.0017	0.11	positive
23	location	good	0.0054	0.12	positive
24	location	great	0.0111	0.19	positive
25	location	nice	0.0055	0.18	positive
26	location	perfect	0.0011	0.15	positive
27	location	spectacular	0.0014	0.71	positive
28	price	high	0.0018	0.22	negative
29	restaurant	expense	0.0020	0.17	negative
30	restaurant	good	0.0067	0.14	positive
31	restaurant	not, good	0.0013	0.19	negative
32	restaurant	pricey	0.0014	0.28	negative
33	restaurant	worth	0.0014	0.16	positive

After classifying the sentiment of the adjective into positive and negative classes to each opinion, the strength of each class was calculated by summing the support values regarding its feature using the opinion polarity score equation discussed in Chapter 4.

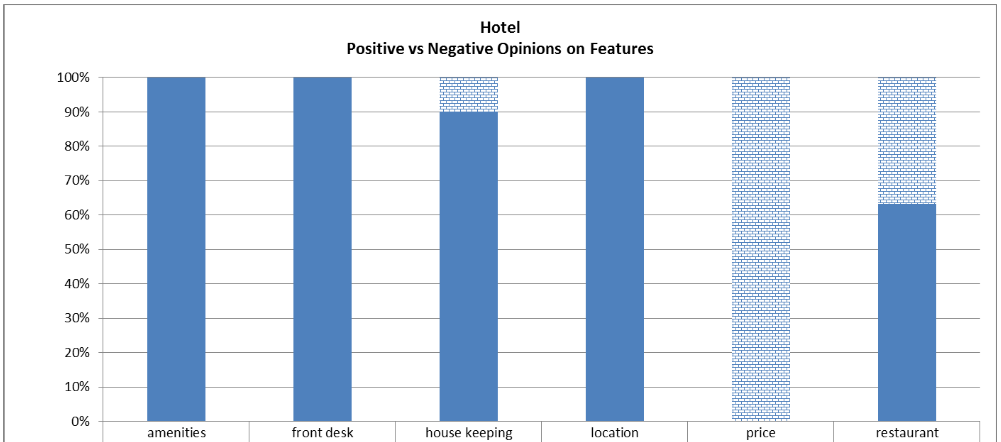
For instance, the feature housekeeping shows the sum of all of the positive support values was 0.1059, and the sum of all of the negative values was 0.0119. These two values were used to identify overall opinion or impression of product features. In the feature housekeeping, the overall opinion polarity score was calculated as 0.0940 (positive). The opinion polarity score of each feature was determined by calculating the difference between the sum of all of the positive support values and the sum of all of the negative support values. The table shows the summary of the association rules for the distinct features and demonstrates the feature-wise frequencies and final opinion polarity.

Table 6.52 Aggregated Opinion Polarities on Features for Hotel

Num (i)	Features (Conclusion)	Sum of Positive Opinion	Sum of Negative Opinion	Opinion Polarity Score (+/-)	Opinion Polarity
1	amenities	0.0014		0.0014	positive
2	front desk	0.0024		0.0024	positive
3	housekeeping	0.1059	0.0119	0.0940	positive
4	location	0.0360		0.0360	positive
5	price		0.0018	−0.0018	negative
6	restaurant	0.0081	0.0047	0.0034	positive

When it comes to the hotel dataset, among six features (N_{FO}), the features amenities, front desk, housekeeping, location, and hotel restaurant represented positive opinions, while the feature price represented negative opinions.

For visual interpretation, Figure 6.17 shows positive and negative summarized user opinions concerning the product features in bar chart form. The solid line portion in the bar represents positive opinions and the checkered pattern portion represents negative opinions. For example, housekeeping was identified as positive based on the overall opinion polarity. However, these features also have negative aspects as well. For housekeeping as an identified feature, 90% of reviewers who talked about customer service expressed positive opinions, while 10% of the reviewers expressed negative opinions. Broadly speaking, this result indicates that housekeeping perception could be analyzed for further service quality improvement.



	amenities	front desk	house keeping	location	price	restaurant
Negative	0.00	0.00	0.10	0.00	1.00	0.37
Positive	1.00	1.00	0.90	1.00	0.00	0.63

Figure 6.17 Opinion polarities expressed on each feature for hotel.

In feature-cause extraction, after applying the association rule on adjectives, nouns, and verbs, 15 association rules for 3 distinct features (N_{FE}) were identified as needing further analysis to understand the reasons behind the negative opinions. Table 6 represents the number of rules identified per feature and the causes for negative opinions.

Table 6.53 Selected F-C Association Rules for Hotel

Rules	Features (Conclusion)	Cause (Premises)	Support	Confidence
1	housekeeping	dark	0.0048	1.00
2	housekeeping	dirty	0.0040	1.00
3	housekeeping	little window	0.0024	1.00
4	housekeeping	flush noise	0.0024	1.00
5	housekeeping	old	0.0143	0.95
6	housekeeping	sound	0.0087	1.00
7	housekeeping	uncomfortable	0.0040	1.00
8	housekeeping	water pressure	0.0032	1.00
9	price	expensive	0.0027	0.15
10	price	high	0.0063	0.34
11	restaurant	busy	0.0045	0.15
12	restaurant	crowded table	0.0013	1.00
13	restaurant	expensive	0.0090	0.36
14	restaurant	overprice	0.0039	0.60
15	restaurant	pricey	0.0065	0.48

Now the individual association rule results were presented into aggregated feature information, which could be applied to product development or improvement processes. Support was calculated as the sum of the frequencies, and confidence was calculated as the average of confidence, where support determined how often a rule was applicable, while confidence determined the strength or reliability of the rule. The rule support multiplied by rule confidence measure helped to identify the rules that might be important for designers/engineers, as it took the confidence value and the support value into account. If the confidence value and the support value were high, the measure concerning rule support multiplied by rule confidence was also high, as presented in the table.

Table 6.54 Aggregate F-C Association Rules for Hotel

Feature (Conclusion)	Sum of Support	Average of Confidence	Sum of Support x Confidence
housekeeping	0.0437	0.99	0.0430
price	0.0091	0.24	0.0026
restaurant	0.0252	0.52	0.0106

Presenting this information via graphical form makes it easier for people to understand the substance of the findings rather than the technical details behind the numbers, as graphs tell a story with visuals rather than plain words or numbers.

For housekeeping, the problems were window size, room lightening, old furniture (or beddings), sound isolation, water pressure (dark, dirty, little-window, flush-noise, old, sound, uncomfortable, water-pressure).

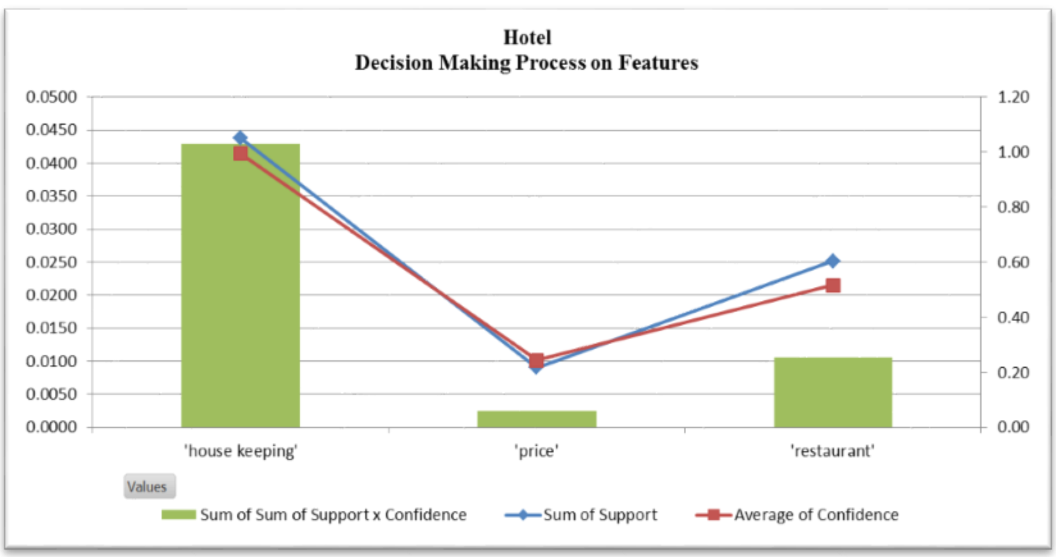


Figure 6.18 Decision making process about features for hotel.

This study shows the relative importance of these three features, which helps the situation when a hotel management team needs to prioritize the service quality or service improvement plan before their next new service release.

Hotel managers in the service domain were expected to have greater return if they address housekeeping issues as this study revealed the highest sum of support was a confidence score of 0.0430, followed by the hotel restaurant, and room-price.

6.11 Summary and Evaluation

In this study, a hybrid approach was introduced, a method for opinion mining of Web reviews using the association rule mining technique, to mine Web reviews, which provides product design intelligence. For each product, 6075 reviews were collected and then transformed into sentences resulting in 58,980 sentences about the ten products. The author were able to operationalize the notations (variables) of the metrics using the datasets. The explanatory notations were total number of reviews (N_R), total number of sentences (N_S), total number of words (N_W), the number of words per sentence (N_W/N_S), total number of candidate features nouns (N_C), total number of features (N_F), the number of candidate feature per sentence (N_C/N_S), and the ratio of feature nouns to candidate feature nouns (N_F/N_C). The descriptive statistics for each dataset were calculated and notations were included in Table 6.55.

Table 6.55 Descriptive Statistics and Calculated Values of All Products

Product Type	N_R	N_S	N_S/N_R	N_W	N_W/N_S	N_C	N_F	N_C/N_S	N_F/N_C
television	480	6765	14.09	107,518	15.89	22,969	5049	3.40	0.22
digital camera	745	7455	10.01	118,173	15.85	24,320	2937	3.26	0.12
laptop	312	3256	10.44	48,882	15.01	10,876	1421	3.34	0.13
mobile phone	352	1405	3.99	20,778	14.79	5280	664	3.76	0.13
sports car	671	6595	9.83	82,426	12.50	20,512	2984	3.11	0.15
sedan	749	3428	4.58	41,449	12.09	8014	1161	2.34	0.14
mobile phone service provider	847	7200	8.50	120,384	16.72	24,704	2831	3.43	0.11
airline travel	570	5370	9.42	89,405	16.65	17,269	2617	3.22	0.15
restaurant	821	11,718	14.27	178,258	15.21	42,116	7578	3.59	0.18
hotel	528	5788	10.96	89,104	15.39	21,768	2107	3.76	0.10

The average number of reviews of all products were 607 reviews with a STD of 189. The average number of sentences of all products was 9.6 sentences with a STD of 3.6. The high STD indicates that some product reviews contained few sentences, while others contained more sentences, with reviewers having opinions on the product features. Considering the average Web review has 6-10 sentences in the literature, in this study, most products were either greater than, or some were close to, the minimum threshold of six sentences (Ganu, Marian et al. 2010; Khan, Baharudin et al. 2010; Li and Chen 2010). This study shows that a relatively high number of N_S/N_R and N_W/N_S yield higher N_F/N_C, which matches the ratio between features and candidate features. Therefore, it could be concluded that rather than the number of reviews by itself, the length of the review and length of the sentence were drivers to determine the information richness of Web reviews. The longer the length of the review and the length of the sentences are, the more likely there would be more features in reviews presenting a better chance of containing the causes of opinions concerning product features.

It was observed that restaurant and television in the dataset have the highest number of sentences per review, followed by hotels. Alternatively, mobile phone has the least number of sentences per review. Compared with the average 3.78 sentences per review in the literature (Ganu, Marian et al. 2010), it seems that Web reviews collected have enough information about products and services to be utilized for new product development.

The method proposed was evaluated by DLIQ metric, which is indicative of content, complexity, and relevancy of the design contextual information. Based on the calculated DLIQ score metric, all of the products' DLIQ scores were found above the minimum threshold score of 50. This indicates that there were sufficient data in the reviews that consumers post their experiences with the product features. The table 5.54 shows design-level information quality score and its three components for all of the products, which were content, complexity, and relevancy. Content evaluates the total amount of information that is available in the review database. Complexity evaluates the number of sentences per review and the ratio of nouns to total words in a review. Relevancy evaluates the total volume of nouns in the reviews and the ratio of feature nouns in the total nouns in the review database.

Table 6.56 Design-Level Information Quality Score for All Products

Product Name	Content	Complexity	Relevancy	DLIQ
television	27.4	30.8	48.8	107.1
digital camera	29.7	23.2	28.2	81.0
laptop	23.7	25.2	27.9	76.8
cell phone	22.1	7.2	24.8	54.1
sports car	28.2	25.6	33.0	86.8
sedan	26.8	6.1	29.7	62.7
mobile phone service provider	30.3	19.8	26.8	76.9
airline travel	27.7	21.1	33.7	82.5
restaurant	31.2	32.5	43.1	106.9
hotel	27.4	27.6	22.6	77.5

The design level information quality (DLIQ) score for the ten products were calculated as 107.1, 81.0, 76.8, 54.1, 86.8, 62.7, 76.9, 82.5, 106.9, and 77.5 for television, digital camera, laptop, mobile phone, sports car, sedan, mobile phone service provider, airline travel, restaurant, and hotel, respectively by using the equations (3.1-3.10) in Chapter 3. Considering the score ranges between 50 to 70 indicates some degree of outcome, in this study, DLIQ scores indicate that there are enough product features in the reviews and, therefore, product designers can move forward with the next step to analyze positive/negative features and reasoning of negative opinions on features.

CHAPTER 7

CONCLUSION

Product development managers are constantly challenged to learn what the consumer product experience really is, and to learn specifically how the product is performing in the field. Today, the Web has created a new source of product intelligence rather than traditional methods (e.g., prototype testing, quality monitoring instruments). These are unsolicited reviews from actual product users who posted across hundreds of websites. Presently, these reviews are primarily used by potential buyers to learn the product experience of other consumers, and it is well known that this has significant effects on buying decisions. A second growing use is in marketing where reviews are analyzed to project product sentiment (positive or negative). In spite of the growing importance of Web reviews, there is limited research focused on extracting specific product intelligence, which could be applied to the manufacturing and service industries concerning product enhancement in a usable format. This dissertation was focused on filling this gap by attempting to extract specific product intelligence that can then be used to develop better product designs. Motivated by the opinion mining research area, the author has presented a feature-based opinion mining system utilizing the association rule mining algorithm to extract product design intelligence.

It analyzed Web reviews from various websites encompassing six products and four service domains to accelerate the product development lifecycle. The method was composed of the following steps: collect data, preprocess data, apply linguistic feature tasks, analyze data by utilizing association rule mining techniques, and summarize the findings. For the experiments, over 6000 Web reviews for ten products and services were collected manually: sedan, sports car, laptop, digital camera, mobile phone, television, airline customer service, mobile phone service provider, hotel, and restaurant. The average number of reviews per product was 607 with a STD of 189.

Data preparation first started with removing noisy data such as review dates, reviewers' names, and treating punctuation as these were nothing to do with consumer experiences in the form of free text. Afterward, sentences were split to extract accurate features to increase the chance of correct word groupings as product features. Sentence splitting was needed because the reviews may contain several features, each of which may represent different features upon which consumers comment. The average number of sentences per review was 5898 with a STD of 2844.

Linguistic feature tasks; tokenization, stop word removal, POS tagging, and Porter's stemming were applied to transform the text data into a format that a computer could recognize for opinion mining (Nasukawa and Yi 2003; Wahl, Winiwarter et al. 2010).

Tokenization is to break a stream of text into words, phrases, symbols, or other meaningful elements called tokens (i.e., a single type of word; Guo, Zhu, et al., 2009; (Nasukawa and Yi 2003; Guo, Zhu et al. 2009). The list of tokens becomes input for further processing such as feature extraction. Stop words were the most commonly used words (e.g., the, an, and), which was to increase the computation time and to improve the accuracy of extracting product features. The PENN Treebank POS-Tagging was applied in the opinion mining framework because nouns and noun phases represent product features, adjectives represent opinions, and verbs represent the causes of opinions. POS tagging assigns each token (words) in reviews as a noun, verb, pronoun, preposition, adverb, adjective, or other lexical class marker.

A stemming algorithm is the process of reducing resultant/derived words to their original meanings. Porter stemming was applied, a very widely used and available stemmer, and was used in many applications (Porter 1980). Porter stemming, is an iterative, rule-based replacement of word suffixes intending to reduce the length of the words and to increase the term frequency; words with higher frequency represent the importance of words.

Three analyses were introduced, which were identification of product features, feature-opinion pairs, and feature-cause pairs.

Based on a word's POS tag, explicitly mentioned nouns and noun phrases were identified as candidate product features. Using TF to identify if the frequent noun was reasonable, as frequent words were likely to be important product features to the reviewers whether negative or positive. Each frequent noun or noun phrase in the outcome was a candidate of product features. However, not all candidates have frequent features generated by using TF weight. To remove those unlikely features, a pruning method was applied as a further drill-down (Ding, Liu et al. 2009; Weishu, Zhiguo et al. 2010). In this study, a word was defined as frequent if it appeared in equal or more than five sentences without any maximum limitation. In this study, after extracting nouns and noun phrases as features, they were grouped together based on the same meaning to increase term frequency (TF). In Web reviews, reviewers often referred to the same product features by different words. It was necessary to group them together in order to efficiently analyze product features (Zhang, Jia et al. 2011). Product development managers validated the list of candidate product features identified by the outcome of the text/data mining tool to map candidate product features to expected features because not every candidate was necessarily a product feature. The number of product features identified was 48, 35, 25, 22, 42, 29, 27, 24, 68, and 11 for TV, digital camera, laptop, mobile phone, sports car, sedan, mobile phone service provider, airline travel, restaurant, and hotel respectively. This confirms that they frequently talked about product features by reviewers, which inspired this study to go further to mine the opinions and opinion polarity. The features were no longer candidates; they were the identified features that were worthwhile for product development managers to investigate further in their new product development process.

Feature-opinion pair identification was the process to identify correlations between product features and the opinions on them from Web reviews utilizing the association mining technique. First, the method extracts both nouns and adjectives together from the sentences. Then, the adjectives with similar meanings were normalized using WordNet, a lexical database for English. Product feature-opinion pair identification not only considers the noun/noun phrase but also the adjectives as opinions. This was to determine if and how much features and opinions were related to each other. There were two important basic measures for association rules, support, and confidence. Since the database was large and users were concerned about only those frequently purchased items, usually thresholds of support and confidence were predefined by users to drop those rules that were not so interesting or useful. The two thresholds were called minimum support and minimum confidence level respectively. The threshold level to evaluate if a rule was significant was defined based on the confidence of the rule. In this study, to increase the number of association rules between features and opinions, minimum confidence level value was set to a lower bound (e.g., 0.1 of 1). The numbers of distinct features that have opinions whether positive or negative were 7, 10, 10, 8, 8, 10, 9, 5, 11, and 6 for TV, digital camera, laptop, mobile phone, sports car, sedan, mobile phone service provider, airline travel, restaurant, and hotel respectively.

Feature-cause pair identification was the process to identify associations between product features and the reasons for them from Web reviews. However, it was focused on analyzing the distinctive negative-features identified in feature-opinion identification. In this step, verbs were introduced as opinion reasoning along with adjectives and nouns. It was assumed that verbs were considered the core of the sentence, and their meanings were key to understand the meaning of the sentence. Similar to adjectives in the feature-opinion identification process, the verbs were replaced with similar meaning using WordNet and a custom-made words list in order to reduce them to a base form of verbs and to group similar verbs, which eventually increase TF. The number of features with causes on negative opinions was 5, 3, 2, 4, 4, 4, 5, 5, 3, and 3 for TV, digital camera, laptop, mobile phone, sports car, sedan, mobile phone service provider, airline travel, restaurant, and hotel respectively. The three analyses were explained in Chapter 6 in detail.

The proposed method was evaluated by introducing the DLIQ measure, discussed in Chapter 3. The DLIQ measure was an evaluation of initial design-level information quality based on the combination of the three factors (i.e., content, complexity, and relevancy), and was calculated directly after identifying product features in the product feature extraction process. Key determinants of content are the number of reviews and the length of the reviews as measured by the number of words. Key determinants of complexity are the number of sentences per review and the ratio of nouns to the total number of words in a review. Key determinants of relevancy are the total volume of nouns in the reviews and the ratio of feature nouns in the total nouns in the review database. The DLIQ score presented 107.1, 81.0, 76.8, 54.1, 86.8, 62.7, 76.9 82.5, 106.9, and 77.5 for TV, digital camera, laptop, mobile phone, sports car, sedan, mobile phone service provider, airline travel, restaurant, and hotel respectively. In this metric, a value of 100 represents a very significant level of information, while a measure of 70 indicates a promising level for DLIQ. A score below 50 indicates additional data is required for analysis. Alternatively, if the score ranges between 50 and 70, one might move forward with next steps (feature-opinion, and feature-cause identification) expecting limited outcome.

The significance of this study was to introduce the hybrid method of opinion mining, which revealed a new source of product development intelligence. This finding was gained based on the methodological approach, which could replace traditional methods in gathering consumer opinions, and even further, could overcome geographical boundary in collecting development opportunity based on local expectation. Eventually, collecting information via the proposed hybrid method of opinion mining will help business, including manufacturing and the service industry, to mitigate product failure risks, to support idea generation, and to reduce product development cycle time and associated time, effort, and ultimately, development cost. Scholarly contributions are

i. Introduced the concept of the Feature-Opinion-Cause chain (unstructured) to web review analysis and related it to the traditional product design process (structured).

ii. Developed and tested an opinion mining based method for effective application of Feature-Opinion-Cause analysis to large web review data sets. Showed that valuable intelligence is extractable.

iii. Developed the Design Level Information Quality (DLIQ) measure which is indicative of the content, complexity and relevancy of the design contextual information that can be extracted from an analysis of web reviews for a given product.

iv. Confirmed and validated that sufficient levels of product design intelligence exists, highlighting the need to include DFOC type analysis in traditional product design. Identified quantitative thresholds and significance ranges for key mining analysis results as related to design intelligence extraction. That is support, confidence, etc.

The results of this research open up multiple new investigations. Possible research topics include:

i. A broad based study of the DLIQ measure across a large portfolio of products. This would characterize the information levels by (i) product and service type (ii) product functional complexity (iii) sales volumes (iv) review volumes and (v) number of significant design features.

ii. As statistical investigation to develop tighter definitions of significant and sufficient levels of support and confidence at both the feature-opinion level and the feature-cause level.

iii. Develop faster opinion mining algorithms which can more reliably extract the target feature sets.

iv. Rules and guidelines for the development of product/service review platforms that emphasize design intelligence as opposed to just peer-to-peer opinion sharing.

APPENDIX A

DATA AND TEXT MINING TOOL - RapidMiner

Figure A.1 shows the graphical user interface (GUI) of the tool, which allows the design of complex data/text mining tasks. RapidMiner's main view consists of three vertical panels. The left panel includes a list of operators and a search box for easy filtering. The blank center panel was the process panel where one could drag and drop operators. Operator parameters were specified on the right panels. Operator parameters also have a very useful help panel below.

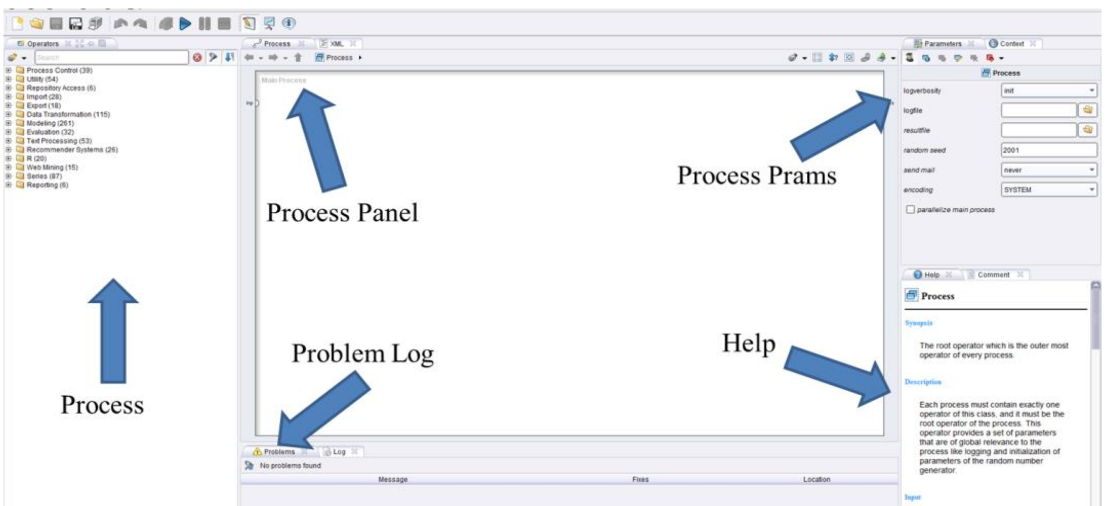
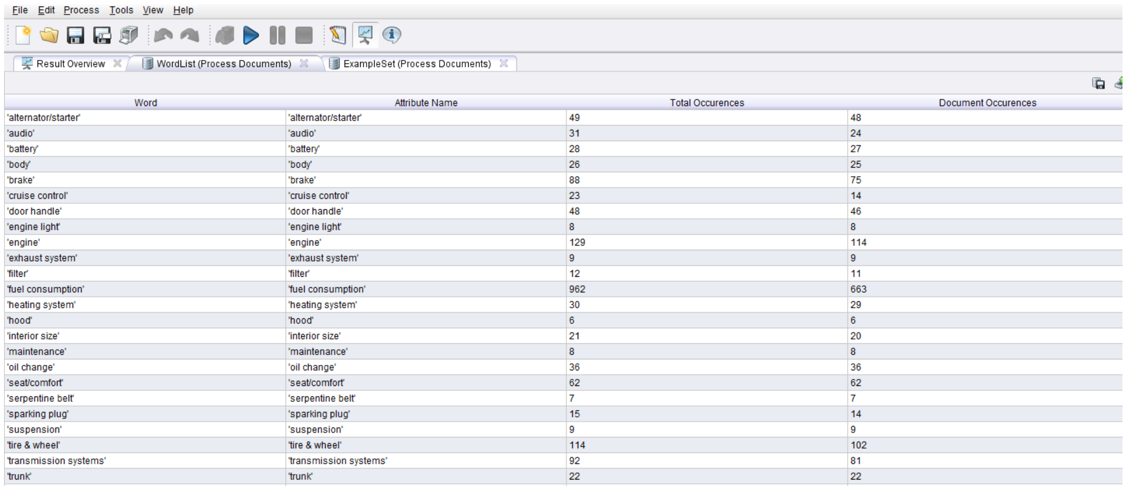


Figure A.1 RapidMiner Graphical User Interface (GUI). The blank central area was the canvas where users graphically build their RapidMiner process. The left side has a menu of operators as well as repositories where processes were stored. The right side has details about the current operator.

Figure A.2 shows the results of text-data processing. The first and second columns (word and attribute name) were word lists from the dataset. The column total occurrence represents the number of words in a sentence, while the column document occurrence represents the number of sentences containing the word.



File Edit Process Tools View Help

Result Overview | WordList (Process Documents) | ExampleSet (Process Documents)

Word	Attribute Name	Total Occurences	Document Occurences
'alternator/starter'	'alternator/starter'	49	48
'audio'	'audio'	31	24
'battery'	'battery'	28	27
'body'	'body'	26	25
'brake'	'brake'	88	75
'cruise control'	'cruise control'	23	14
'door handle'	'door handle'	48	46
'engine light'	'engine light'	8	8
'engine'	'engine'	129	114
'exhaust system'	'exhaust system'	9	9
'filter'	'filter'	12	11
'fuel consumption'	'fuel consumption'	962	663
'heating system'	'heating system'	30	29
'hood'	'hood'	6	6
'interior size'	'interior size'	21	20
'maintenance'	'maintenance'	8	8
'oil change'	'oil change'	36	36
'seat/comfort'	'seat/comfort'	62	62
'serpentine belt'	'serpentine belt'	7	7
'sparking plug'	'sparking plug'	15	14
'suspension'	'suspension'	9	9
'tire & wheel'	'tire & wheel'	114	102
'transmission systems'	'transmission systems'	92	81
'trunk'	'trunk'	22	22

Figure A.2 Results of linguistic processing in RapidMiner. It generates word vectors from a document collection.

Figure A.3 shows a workflow diagram illustrating the process of association rule mining. This process takes in frequent item sets and seeks out any patterns that occur so frequently that they could be considered rules.

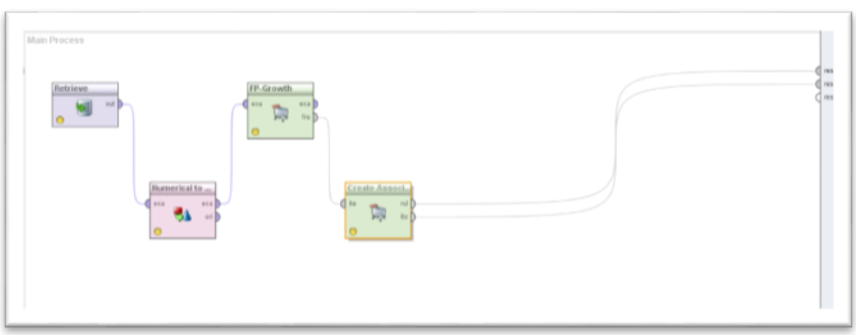
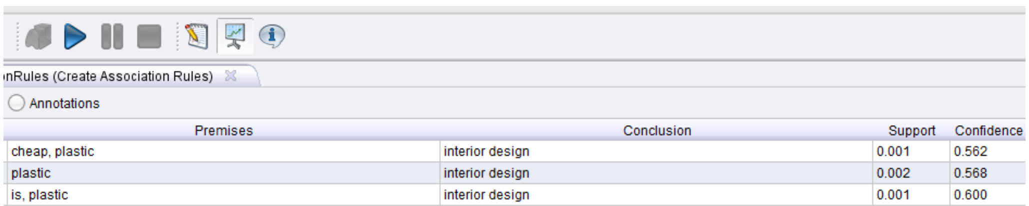


Figure A.3 The process of association rule mining in RapidMiner.

As shown in Figure A.4, many rules were generated that were “interior design” in conclusion, and “cheep-plastic,” “plastic” and “is-plastic” in premises where premises represent adjectives (opinions) and conclusions represent nouns and noun phrases (product features). Figure A.4 is the illustration of association rule results in RapidMiner as an example showing 0.2% of documents contain interior-design and plastic in the database (support), and the percentage of the number of documents that contain interior-design and plastic together to the total number of documents that contain plastic was 56.8%.



Premises	Conclusion	Support	Confidence
cheap, plastic	interior design	0.001	0.562
plastic	interior design	0.002	0.568
is, plastic	interior design	0.001	0.600

Figure A.4 Shows association rule mining results in RapidMiner. Premises represent opinions, conclusions represent product features; support was defined as the percentage of documents that contain premised and a conclusion together with the total number of documents in the dataset; the confidence was the conditional probability that, given premises present in a dataset, a conclusion will also be present.

APPENDIX B

EXECUTIVE SUMMARY REPORTS

This is an executive summary reports for television, digital camera, laptop, mobile phone, sedan, mobile service provider, airline travel, restaurant, and hotel, which provide aggregate key results. The executive summary report for sports car presented on Table 5.14.

Table B.1 Executive Summary Report for Television

DFOC ANALYSIS - TELEVISION							
WEB REVIEW DATABASE							
N_R =	480	N_S =	6765	N_W =	107518		
DFOC EXTRACTION RATIOS							
N_C =	22969	N_F =	5049	N_C/N_S =	3.4	N_F/N_C =	0.22
DESIGN LEVEL INFORMATION QUALITY							
$DLIQ_{Cont}$ =	27.4	$DLIQ_{Cplx}$ =	30.8	$DLIQ_{Relv}$ =	48.8	$DLIQ$ =	107.1
DESIGN INTELLIGENCE							
Product Features Identified - Notable Interest (TF>5) =					48		
Product Features Identified - Significant Interest (TF>85) =					7		
FEATURE OPINION ANALYSIS							
#	*Feature*		*Support*	*+ Opinion*	*- Opinion*	*Polarity*	
1	Internet connection		0.5%	0.0014	0.0034	Mixed	
2	Price		0.1%		0.001	Negative	
3	Remote control		0.6%	0.0039	0.0023	Mixed	
4	Screen resolution/Image quality		18.0%	0.0991	0.0804	Mixed	
5	Speakers/Sound quality		0.3%	0.0013	0.0014	Mixed	
6	SSG active glass		0.1%	0.0013		Positive	
7	Video and audio		0.1%		0.0011	Negative	
FEATURE-CAUSE-OPINION (Negative) ANALYSIS							
#	*Feature*		*Cause*	*Support*	*Confidence*	*DFOC Strength*	
1	Internet connection		Not connect	0.2%	31%	0.1	
			Not work	0.2%	15%		
			Slow	0.2%	36%		
2	Price		High	5.4%	100%	6.7	
			Store match	1.4%	100%		
3	Remote control		Not control	0.2%	56%	2.0	
			Not function	0.1%	31%		
			Not work	0.2%	18%		
4	Screen resolution/Image quality		Noise	0.1%	69%	2.0	
			Not clear	0.1%	73%		
7	Video and audio		Blurry	0.2%	50%	7.0	
			Not come	0.1%	12%		
			Not connect	0.3%	25%		
			Not play	0.3%	39%		
			Not support instant	0.1%	100%		
			Not work	0.3%	25%		
			Problem play	0.1%	75%		

Table B.2 Executive Summary Report for Digital Camera

DFOC ANALYSIS - DIGITAL CAMERA							
WEB REVIEW DATABASE							
N_R =	745	N_S =	7455	N_W =	118173		
DFOC EXTRACTION RATIOS							
N_C =	24320	N_F =	2937	N_C/N_S =	3.26	N_F/N_C =	0.12
DESIGN LEVEL INFORMATION QUALITY							
$DLIQ_{Cont}$ =	29.7	$DLIQ_{Cplx}$ =	23.2	$DLIQ_{Relv}$ =	28.2	$DLIQ$ =	81.0
DESIGN INTELLIGENCE							
Product Features Identified - Notable Interest (TF>5) =					35		
Product Features Identified - Significant Interest (TF>13) =					10		

FEATURE OPINION ANALYSIS					
#	*Feature*	*Support*	*+ Opinion*	*- Opinion*	*Polarity*
1	Auto-flash/flash	0.4%	0.0011	0.0025	Mixed
2	Battery/Battery life	0.2%	0.0017		Positive
3	ISO	0.2%	0.0024		Positive
4	Light	0.7%	0.0074		Positive
5	Motion	0.1%		0.001	Negative
6	Picture/Picture quality	7.8%	0.0611	0.017	Mixed
7	Screen	0.1%	0.0014		Positive
8	Size	0.1%	0.0011		Positive
9	Video/Video quality	4.0%	0.0345	0.0055	Mixed
10	Zoom/Lens	0.2%	0.0014	0.001	Mixed

FEATURE-CAUSE-OPINION (Negative) ANALYSIS					
#	*Feature*	*Cause*	*Support*	*Confidence*	*DFOC Strength*
1	Auto flash/flash	Blurry	1.6%	50%	5.0
		Corner shadow	1.2%	100%	
		Low	5.8%	57%	
5	Motion	Slow record	8.3%	100%	8.0
10	Zoom/Lens	Make noise	0.8%	100%	2.0
		Optic low	0.8%	80%	
		Motor sound	0.8%	100%	

Table B.3 Executive Summary Report for Laptop

DFOC ANALYSIS - LAPTOP							
WEB REVIEW DATABASE							
N_R =	312	N_S =	3256	N_W =	48882		
DFOC EXTRACTION RATIOS							
N_C =	10876	N_F =	1421	N_C/N_S =	3.34	N_F/N_C =	0.13
DESIGN LEVEL INFORMATION QUALITY							
$DLIQ_{Cont}$ =	23.7	$DLIQ_{Cplx}$ =	25.2	$DLIQ_{Relv}$ =	27.9	$DLIQ$ =	76.8
DESIGN INTELLIGENCE							
Product Features Identified - Notable Interest (TF>5) =					25		
Product Features Identified - Significant Interest (TF>18) =					10		
FEATURE OPINION ANALYSIS							
#	*Feature*		*Support*	*+ Opinion*	*- Opinion*	*Polarity*	
1	Battery/battery life		1.1%	0.0105		Positive	
2	CD/DVD driver		0.1%	0.0011		Positive	
3	Hard drive		0.2%	0.0011	0.0011	Mixed	
4	Keyboard		1.5%	0.0124	0.0022	Mixed	
5	Laptop case		0.1%	0.0011		Positive	
6	Laptop size/Weight		0.3%	0.0022	0.0011	Mixed	
7	Monitor		2.8%	0.0238	0.0041	Mixed	
8	Operating system		0.4%	0.0022	0.0014	Mixed	
9	Price		0.4%	0.0028	0.0011	Mixed	
10	Processor (or CPU)		0.1%	0.0014		Positive	
FEATURE-CAUSE-OPINION (Negative) ANALYSIS							
#	*Feature*		*Cause*	*Support*	*Confidence*	*DFOC Strength*	
3	hard drive		Is not compute	1.2%	100%	17.3	
			Replace	2.5%	50%		
			Slow	1.9%	60%		
			Space need	1.2%	100%		
			Upgrade	14.2%	88%		
8	operating system		Not have windows	0.2%	60%	0.6	
			Not work	0.4%	25%		
			Problem	0.2%	13%		
			Switch application	0.2%	60%		
			Windows prefer	0.2%	75%		

Table B.4 Executive Summary Report for Mobile Phone

DFOC ANALYSIS - MOBILE PHONE							
WEB REVIEW DATABASE							
N_R =	352	N_S =	1405	N_W =	20778		
DFOC EXTRACTION RATIOS							
N_C =	5280	N_F =	64	N_C/N_S =	3.76	N_F/N_C =	0.13
DESIGN LEVEL INFORMATION QUALITY							
$DLIQ_{Cont}$ =	22.1	$DLIQ_{Cplx}$ =	7.2	$DLIQ_{Relv}$ =	24.8	$DLIQ$ =	54.1
DESIGN INTELLIGENCE							
Product Features Identified - Notable Interest (TF>5) =					19		
Product Features Identified - Significant Interest (TF>23) =					8		

FEATURE OPINION ANALYSIS					
#	*Feature*	*Support*	*+ Opinion*	*- Opinion*	*Polarity*
1	Accessory	0.3%	0.0032		Positive
2	Battery	1.0%	0.0023	0.0077	Mixed
3	Camera/Picture	2.2%	0.0069	0.0151	Mixed
4	Internet	0.6%		0.0064	Negative
5	Keyboard	0.2%	0.0018		Positive
6	Price	0.7%	0.0073		Positive
7	Screen	0.1%		0.0014	Negative
8	Track pad	0.1%	0.0014		Positive

FEATURE-CAUSE-OPINION (Negative) ANALYSIS					
#	*Feature*	*Cause*	*Support*	*Confidence*	*DFOC Strength*
2	Battery	Not new	0.2%	13%	0.03
3	Camera/Picture	Not good flash	0.1%	100%	0.14
4	Internet	Not access	0.2%	80%	0.34
		Not support	0.2%	50%	
		Slow	0.2%	45%	
7	Screen	Side	0.1%	43%	0.12
		White	0.1%	43%	

Table B.5 Executive Summary Report for Sedan

DFOC ANALYSIS - SEDAN							
WEB REVIEW DATABASE							
N_R =	749	N_S =	3428	N_W =	41449		
DFOC EXTRACTION RATIOS							
N_C =	8014	N_F =	1161	N_C/N_S =	2.34	N_F/N_C =	0.14
DESIGN LEVEL INFORMATION QUALITY							
$DLIQ_{Cont}$ =	26.8	$DLIQ_{Cplx}$ =	6.1	$DLIQ_{Relv}$ =	29.7	$DLIQ$ =	62.7
DESIGN INTELLIGENCE							
Product Features Identified - Notable Interest (TF>5) =					29		
Product Features Identified - Significant Interest (TF>7) =					10		
FEATURE OPINION ANALYSIS							
#	*Feature*		*Support*	*+ Opinion*	*- Opinion*	*Polarity*	
1	Alternator/Starter		0.2%		0.0017	Negative	
2	Audio		0.1%		0.0012	Negative	
3	Brake		0.5%		0.0052	Negative	
4	Engine		0.1%		0.0012	Negative	
5	Filter		0.1%		0.0012	Negative	
6	Fuel consumption		8.1%	0.0529	0.0276	Mixed	
7	Heating system		0.2%		0.0019	Negative	
8	Interior size		0.1%	0.0012		Positive	
9	Seat/comfort		0.3%	0.0021	0.0012	Mixed	
10	Trunk		0.3%	0.0026		Positive	
FEATURE-CAUSE-OPINION (Negative) ANALYSIS							
#	*Feature*		*Cause*	*Support*	*Confidence*	*DFOC Strength*	
1	Alternator/Starter		Belt change	2.04%	100%	18.4	
			Repair rod	2.04%	100%		
			Replace belt	6.12%	100%		
			Replace rod	8.16%	100%		
3	Brake		Buy strut	2.35%	100%	29.2	
			Change	16.47%	100%		
			Front replace	3.53%	60%		
			Need pump	2.35%	100%		
			need strut	2.35%	100%		
			Replace pump	3.53%	100%		
4	Engine		Problem mount	1.64%	67%	6.8	
			Replace head	1.64%	100%		
			Replace mount	2.46%	100%		
			Replace time belt	1.64%	100%		
7	Heating system		Cold	0.19%	53%	1.0	

Table B.6 Executive Summary Report for Mobile Phone Service Provider

DFOC ANALYSIS - MOBILE PHONE SERVICE PROVIDER							
WEB REVIEW DATABASE							
N_R =	847	N_S =	7200	N_W =	120384		
DFOC EXTRACTION RATIOS							
N_C =	24704	N_F =	2831	N_C/N_S =	3.43	N_F/N_C =	0.11
DESIGN LEVEL INFORMATION QUALITY							
$DLIQ_{Cont}$ =	30.3	$DLIQ_{Cplx}$ =	19.8	$DLIQ_{Relv}$ =	26.8	$DLIQ$ =	76.9
DESIGN INTELLIGENCE							
Product Features Identified - Notable Interest (TF>5) =					27		
Product Features Identified - Significant Interest (TF>28) =					9		

FEATURE OPINION ANALYSIS					
#	*Feature*	*Support*	*+ Opinion*	*- Opinion*	*Polarity*
1	Billing	0.1%		0.0011	Negative
2	Contract/Contract termination	1.0%		0.0099	Negative
3	Coverage	0.1%		0.0011	Negative
4	Customer service	5.8%	0.0137	0.0438	Mixed
5	Data plan	0.1%	0.0011		Positive
6	Phone plan	0.6%	0.0023	0.0037	Mixed
7	Service charge	0.2%		0.0024	Negative
8	Service provider	0.1%		0.0011	Negative
9	Text message	0.2%	0.0017		Positive

#	*Feature*	*Cause*	*Support*	*Confidence*	*DFOC Strength*
1	Billing	Fix	0.1%	18%	0.54
		Past	0.2%	19%	
		Credit card	0.2%	95%	
2	Contract/Contract termination	Break	0.6%	64%	1.6
		Cancel	2.2%	38%	
		Not sign	0.5%	46%	
		Price	0.5%	34%	
3	Coverage	Not service area	0.1%	50%	0.05
4	Customer service	Call time problem	0.1%	68%	0.54
		Hung	0.2%	38%	
		Do not know	0.1%	27%	
		Not explain	0.1%	34%	
		Not help	0.1%	19%	
		Not resolve	0.1%	27%	
		Not transfer	0.1%	28%	
		Not understand	0.1%	22%	
		Rude	0.3%	35%	
		Wait	0.2%	14%	
6	Phone plan	Call time	0.2%	16%	0.13
		Rate	0.2%	43%	

Table B.7 Executive Summary Report for Airline Travel

DFOC ANALYSIS - AIRLINE TRAVEL

WEB REVIEW DATABASE

N_R =	570	N_S =	5370	N_W =	89405

DFOC EXTRACTION RATIOS

N_C =	17269	N_F =	2617	N_C/N_S =	3.22	N_F/N_C =	0.15

DESIGN LEVEL INFORMATION QUALITY

$DLIQ_{Cont}$ =	27.7	$DLIQ_{Cplx}$ =	21.1	$DLIQ_{Relv}$ =	33.7	$DLIQ$ =	82.5

DESIGN INTELLIGENCE

Product Features Identified - Notable Interest (TF>5) =	24
Product Features Identified - Significant Interest (TF>119) =	5

FEATURE OPINION ANALYSIS

#	*Feature*	*Support*	*+ Opinion*	*- Opinion*	*Polarity*
1	Airport security/Facilities	0.3%		0.0031	Negative
2	Baggage check-in/claim	0.5%	0.0013	0.0036	Mixed
3	Customer service	3.1%	0.0099	0.0215	Mixed
4	Flight attendant	0.2%		0.0017	Negative
5	Travel agent	2.4%	0.0125	0.011	Mixed

#	*Feature*	*Cause*	*Support*	*Confidence*	*DFOC Strength*
1	Airport security/Facilities	Delay	1.0%	45%	2.23
		Line	1.2%	13%	
		Miss	1.3%	60%	
		Wait	1.7%	48%	
2	Baggage check-in/claim	Miss	0.4%	23%	0.38
		Wait	0.3%	11%	
		Ruin	0.1%	41%	
		Brake	0.1%	62%	
		Damage	0.2%	64%	
3	Customer service	Hold	0.2%	15%	0.14
		Wait	0.3%	15%	
		Ignore	0.1%	47%	
4	Flight attendant	Not care	0.4%	80%	0.81
		Rude	1.1%	42%	
5	Travel agent	Not get help	0.1%	32%	0.08
		Ignore	0.1%	41%	

Table B.8 Executive Summary Report for Restaurant

<table>
<tr><td colspan="8">DFOC ANALYSIS - RESTAURANT</td></tr>
<tr><td colspan="8">WEB REVIEW DATABASE</td></tr>
<tr><td>N_R =</td><td>821</td><td>N_S =</td><td>11718</td><td>N_W =</td><td>178258</td><td colspan="2"></td></tr>
<tr><td colspan="8">DFOC EXTRACTION RATIOS</td></tr>
<tr><td>N_C =</td><td>42116</td><td>N_F =</td><td>7578</td><td>N_C/N_S =</td><td>3.59</td><td>N_F/N_C =</td><td>0.18</td></tr>
<tr><td colspan="8">DESIGN LEVEL INFORMATION QUALITY</td></tr>
<tr><td>$DLIQ_{Cont}$ =</td><td>31.2</td><td>$DLIQ_{Cplx}$ =</td><td>32.5</td><td>$DLIQ_{Relv}$ =</td><td>43.1</td><td>$DLIQ$ =</td><td>106.9</td></tr>
<tr><td colspan="8">DESIGN INTELLIGENCE</td></tr>
<tr><td colspan="5">Product Features Identified - Notable Interest (TF>5) =</td><td>68</td><td colspan="2" rowspan="2"></td></tr>
<tr><td colspan="5">Product Features Identified - Significant Interest (TF>113) =</td><td>11</td></tr>
<tr><td colspan="8">FEATURE OPINION ANALYSIS</td></tr>
<tr><td>#</td><td colspan="2">Feature</td><td>Support</td><td>+ Opinion</td><td>- Opinion</td><td colspan="2">Polarity</td></tr>
<tr><td>1</td><td colspan="2">Ambiance/Décor</td><td>0.1%</td><td>0.0012</td><td></td><td colspan="2">Positive</td></tr>
<tr><td>2</td><td colspan="2">Baguette/Bread</td><td>0.3%</td><td>0.0029</td><td></td><td colspan="2">Positive</td></tr>
<tr><td>3</td><td colspan="2">Desert</td><td>0.5%</td><td>0.0036</td><td>0.0017</td><td colspan="2">Mixed</td></tr>
<tr><td>4</td><td colspan="2">Lunch/Dinner</td><td>0.4%</td><td>0.0043</td><td></td><td colspan="2">Positive</td></tr>
<tr><td>5</td><td colspan="2">Price</td><td>0.1%</td><td></td><td>0.001</td><td colspan="2">Negative</td></tr>
<tr><td>6</td><td colspan="2">Sauce</td><td>0.1%</td><td>0.0013</td><td></td><td colspan="2">Positive</td></tr>
<tr><td>7</td><td colspan="2">Seafood</td><td>1.7%</td><td>0.0131</td><td>0.0038</td><td colspan="2">Mixed</td></tr>
<tr><td>8</td><td colspan="2">Service</td><td>0.6%</td><td>0.0057</td><td></td><td colspan="2">Positive</td></tr>
<tr><td>9</td><td colspan="2">Staff</td><td>0.2%</td><td>0.0019</td><td></td><td colspan="2">Positive</td></tr>
<tr><td>10</td><td colspan="2">Steak</td><td>0.2%</td><td>0.0024</td><td></td><td colspan="2">Positive</td></tr>
<tr><td>11</td><td colspan="2">Wine and liquor</td><td>0.4%</td><td>0.0038</td><td></td><td colspan="2">Positive</td></tr>
<tr><td>#</td><td colspan="2">Feature</td><td>Cause</td><td>Support</td><td>Confidence</td><td colspan="2">DFOC Strength</td></tr>
<tr><td>3</td><td colspan="2">Desert</td><td>Shell</td><td>0.1%</td><td>46%</td><td colspan="2">0.05</td></tr>
<tr><td rowspan="2">5</td><td colspan="2" rowspan="2">Price</td><td>Expensive</td><td>2.8%</td><td>71%</td><td colspan="2" rowspan="2">6.5</td></tr>
<tr><td>High</td><td>4.5%</td><td>100%</td></tr>
<tr><td rowspan="5">7</td><td colspan="2" rowspan="5">Seafood</td><td>Escolar</td><td>0.6%</td><td>53%</td><td colspan="2" rowspan="5">1.98</td></tr>
<tr><td>Not cook</td><td>1.1%</td><td>84%</td></tr>
<tr><td>Not fresh</td><td>0.4%</td><td>73%</td></tr>
<tr><td>Overpower</td><td>0.4%</td><td>55%</td></tr>
<tr><td>Salty</td><td>0.5%</td><td>54%</td></tr>
</table>

Table B.9 Executive Summary Report for Hotel

DFOC ANALYSIS - HOTEL							
WEB REVIEW DATABASE							
N_R =	528	N_S =	5788	N_W =	89104		
DFOC EXTRACTION RATIOS							
N_C =	21768	N_F =	2107	N_C/N_S =	3.76	N_F/N_C =	0.10
DESIGN LEVEL INFORMATION QUALITY							
$DLIQ_{Cont}$ =	27.4	$DLIQ_{Cplx}$ =	27.6	$DLIQ_{Relv}$ =	22.6	$DLIQ$ =	77.5
DESIGN INTELLIGENCE							
Product Features Identified - Notable Interest (TF>5) =					11		
Product Features Identified - Significant Interest (TF>89) =					6		

FEATURE OPINION ANALYSIS

#	*Feature*	*Support*	*+ Opinion*	*- Opinion*	*Polarity*
1	Amenities	0.1%	0.0014		Positive
2	Front desk	0.2%	0.0024		Positive
3	Housekeeping	11.8%	0.1059	0.0119	Mixed
4	Location	3.6%	0.036		Positive
5	Price	0.2%		0.0018	Negative
6	Restaurant	1.3%	0.0081	0.0047	Mixed

#	*Feature*	*Cause*	*Support*	*Confidence*	*DFOC Strength*
3	Housekeeping	dark	0.5%	100%	4.3
		dirty	0.4%	100%	
		little window	0.2%	100%	
		flush noise	0.2%	100%	
		old	1.4%	95%	
		sound	0.9%	100%	
		uncomfortable	0.4%	100%	
		water pressure	0.3%	100%	
5	Price	expensive	0.3%	15%	0.26
		high	0.6%	34%	
6	Restaurant	busy	0.5%	15%	1.06
		crowded table	0.1%	100%	
		expensive	0.9%	36%	
		overprice	0.4%	60%	
		pricey	0.7%	48%	

APPENDIX C

SAME PRODUCT FEATURES BY DIFFERENT WORDS

Different reviewers often referred to the same product features by different words. Therefore, it was necessary to group them together in order to reduce the size of the extracted features, and further, similar features were grouped together based on the same meaning to increase term frequency.

Table C.1 Sample List of Words that is Similar in Meaning – Television

Word	Same Feature Group	Source
player	Video & Audio	Television
Video quality	Video & Audio	Television
Instant video	Video & Audio	Television
video	Video & Audio	Television
speaker	Speakers/Sound Quality	Television
sound	Speakers/Sound Quality	Television
Sound quality	Speakers/Sound Quality	Television
tone	Speakers/Sound Quality	Television
volume	Speakers/Sound Quality	Television
Sound system	Speakers/Sound Quality	Television
Surround system	Speakers/Sound Quality	Television
audio	Speakers/Sound Quality	Television
surround	Speakers/Sound Quality	Television
subwoof	Speakers/Sound Quality	Television
bass	Speakers/Sound Quality	Television
Quality picture	Screen resolution/Image quality	Television
view	Screen resolution/Image quality	Television
electron	Screen resolution/Image quality	Television
resolution	Screen resolution/Image quality	Television
screen	Screen resolution/Image quality	Television
Screen tv	Screen resolution/Image quality	Television
lcd	Screen resolution/Image quality	Television
Color tint	Screen resolution/Image quality	Television
hdtv	Screen resolution/Image quality	Television
glass	Screen resolution/Image quality	Television
hd	Screen resolution/Image quality	Television
Backlight contrast	Screen resolution/Image quality	Television
plasma	Screen resolution/Image quality	Television
Hd quality	Screen resolution/Image quality	Television
contrast	Screen resolution/Image quality	Television
record	DVD Player	Television
Rai movie	DVD Player	Television
movie	DVD Player	Television
Dvd player	DVD Player	Television

Table C.2 Sample List of Words that is Similar in Meaning – Digital Camera

Word	Same Feature Group	Source
zoom	zoom/lens	digital camera
lens	zoom/lens	digital camera
Video quality	Video/Movie quality	digital camera
video	Video/Movie quality	digital camera
Hd movie	Video/Movie quality	digital camera
Hd video	Video/Movie quality	digital camera
Movie quality	Video/Movie quality	digital camera
Movie mode	Video/Movie quality	digital camera
shutter	Shutter/Shutter speed	digital camera
Shutter speed	Shutter/Shutter speed	digital camera
megapixel	Resolution	digital camera
resolution	Resolution	digital camera
pixel	Resolution	digital camera
image	Picture/Picture Quality	digital camera
Image quality	Picture/Picture Quality	digital camera
photograph	Picture/Picture Quality	digital camera
Picture quality	Picture/Picture Quality	digital camera
photo	Picture/Picture Quality	digital camera
picture	Picture/Picture Quality	digital camera
Photo quality	Picture/Picture Quality	digital camera

Table C.3 Sample List of Words that is Similar in Meaning - Laptop

Word	Same Feature Group	Source
Vpn	VPN Connection	Laptop
voiceov	Voice Over Internet protocol	Laptop
sound	Speakers	Laptop
speaker	Speakers	Laptop
software	Software	Laptop
Screen resolution	Screen/Screen Resolution	Laptop
thunderbolt	Screen/Screen Resolution	Laptop
Chip	processor (or CPU)	Laptop
Core	processor (or CPU)	Laptop
Cpu	processor (or CPU)	Laptop
Ghz processor	processor (or CPU)	Laptop
intel	processor (or CPU)	Laptop
processor	processor (or CPU)	Laptop
speed	processor (or CPU)	Laptop
printer	Printer	Laptop
Cost	Price	Laptop
cord	power cord	Laptop
Power cord	power cord	Laptop
linux	Operating System	Laptop
Osx	Operating System	Laptop
Os	Operating System	Laptop
Mac os	Operating System	Laptop
Os x	Operating System	Laptop
vista	Operating System	Laptop
display	Monitor	Laptop
Gui	Monitor	Laptop

Table C.4 Sample List of Words that is Similar in Meaning – Mobile Phone

Word	Same Feature Group	Source
Wi-fi	Wi-Fi	mobile phone
wifi	Wi-Fi	mobile phone
Bluetooth wife	Wi-Fi	mobile phone
email	Text message	mobile phone
text	Text message	mobile phone
software	Software	mobile phone
download	Software	mobile phone
pin	Sim Card	mobile phone
sim	Sim Card	mobile phone
card	Sim Card	mobile phone
Sim card	Sim Card	mobile phone
memory	Memory size	mobile phone
gb	Memory size	mobile phone
keyboard	Keyboard	mobile phone
qwerti	Keyboard	mobile phone
Internet access	Internet Access	mobile phone
internet	Internet Access	mobile phone
Camera bluetooth	Camera	mobile phone
camara	Camera	mobile phone
camera	Camera	mobile phone

Table C.5 Sample List of Words that is Similar in Meaning – Sports Car

Word	Component	Source
awd	4WD/AWD	sports car
Wheel drive	4WD/AWD	sports car
heat	AC/Heater	sports car
intercool	AC/Heater	sports car
supercharge	AC/Heater	sports car
cyl	Engine power	sports car
cylinder	Engine power	sports car
engine	Engine power	sports car
Engine problem	Engine power	sports car
Horse power	Engine power	sports car
hp	Engine power	sports car
motor	Engine power	sports car
piston	Engine power	sports car
rpm	Engine power	sports car
speed	Engine power	sports car
torque	Engine power	sports car
interior	Interior design	sports car
Leather interior	Interior design	sports car
Passing side	Interior design	sports car
Plastic interior	Interior design	sports car
clutch	Transmission Systems	sports car
gear	Transmission Systems	sports car
Manual transmission	Transmission Systems	sports car
transmission	Transmission Systems	sports car

Table C.6 Sample List of Words that is Similar in Meaning – Sedan

Word	Component	Source
alternator	Alternator/Starter	sedan
cd	Audio	sedan
battery	Battery	sedan
brake	Brake	sedan
break	Brake	sedan
Front brake	Brake	sedan
Car price	Car price	sedan
cruise	cruise control	sedan
Cruise control	cruise control	sedan
door	door handle	sedan
handle	door handle	sedan
cylinder	engine	sedan
engine	engine	sedan
Head gasket	engine	sedan
Engine light	engine light	sedan
exhaust	Exhaust system	sedan
filter	Filter	sedan
Average mpg	Fuel Consumption	sedan
fuel	Fuel Consumption	sedan
Fuel mileage	Fuel Consumption	sedan
gas	Fuel Consumption	sedan
Gas milag	Fuel Consumption	sedan
Gas mileag	Fuel Consumption	sedan
Gas price	Fuel Consumption	sedan
ac	Heating System	sedan
Air condition	Heating System	sedan
Chang oil	Oil Change	sedan
Front wheel	Tire & Wheel	sedan
clutch	Transmission Systems	sedan
gear	Transmission Systems	sedan

Table C.7 Sample List of Words that is Similar in Meaning – Mobile Service Provider

Word	Same Feature Group	Source
message	Text message	mobile phone service provider
txt	Text message	mobile phone service provider
text	Text message	mobile phone service provider
text message	Text message	mobile phone service provider
gui	Screen/GUI	mobile phone service provider
screen	Screen/GUI	mobile phone service provider
internet	Internet & email	mobile phone service provider
email	Internet & email	mobile phone service provider
wifi	Internet & email	mobile phone service provider
Internet service	Internet & email	mobile phone service provider
Web	Internet & email	mobile phone service provider
website	Internet & email	mobile phone service provider
service person	Customer Service	mobile phone service provider
service agent	Customer Service	mobile phone service provider
agent	Customer Service	mobile phone service provider
supervisor	Customer Service	mobile phone service provider
call service	Customer Service	mobile phone service provider
sale rep	customer service	mobile phone service provider
clerk	customer service	mobile phone service provider
supervisor	Customer Service	mobile phone service provider
service repres	Customer Service	mobile phone service provider
customer support	Customer Service	mobile phone service provider
call center	Customer Service	mobile phone service provider
custom repres	Customer Service	mobile phone service provider
custom service	Customer Service	mobile phone service provider
Service rep	Customer Service	mobile phone service provider
employee	Customer Service	mobile phone service provider

Table C.8 Sample List of Words that is Similar in Meaning – Airline Travel

Word	Same Feature Group	Source
airport	Airport Security/Facilities	Airline Travel
airwai	Arrival/Departure	Airline Travel
arrive	Arrival/Departure	Airline Travel
backpack	Baggage Check-in/Claim	Airline Travel
bag	Baggage Check-in/Claim	Airline Travel
baggage	Baggage Check-in/Claim	Airline Travel
baggage claim	Baggage Check-in/Claim	Airline Travel
baggage fee	Baggage fee	Airline Travel
bathroom	Bathroom	Airline Travel
board	Boarding	Airline Travel
board flight	Boarding	Airline Travel
answer phone	Customer Service	Airline Travel
answer question	Customer Service	Airline Travel
business class	Economy/Business Class	Airline Travel
breakfast	Meal	Airline Travel
breakfast voucher	Meal	Airline Travel
book	Reservation	Airline Travel
booking flight	Reservation	Airline Travel
agent	Travel Agent	Airline Travel
airline employee	Travel Agent	Airline Travel

Table C.9 Sample List of Words that is Similar in Meaning – Restaurant

Word	Same Feature Group	Source
osetra	Seafood	restaurant
sea bass	Seafood	restaurant
seafood	Seafood	restaurant
seafood dish	Seafood	restaurant
shellfish	Seafood	restaurant
shrimp	Seafood	restaurant
Snapper saffron	Seafood	restaurant
tuna carpaccio	Seafood	restaurant
tuna foie	Seafood	restaurant
yellowtail	Seafood	restaurant
hostess	Staff	restaurant
server	Staff	restaurant
staff	staff	restaurant
staff member	staff	restaurant
beef	steak	restaurant
kobe beef	Steak	restaurant
meat	Steak	restaurant
pork	Steak	restaurant
steak	steak	restaurant
wagyu beef	Steak	restaurant
coffee	Tea/Coffee	restaurant
espresso	Tea/Coffee	restaurant
tea	Tea/Coffee	restaurant
bottle wine	Wine and Liquor	restaurant
champagne	Wine and Liquor	restaurant
cocktail	Wine and Liquor	restaurant
glass wine	Wine and Liquor	restaurant
martini	Wine and Liquor	restaurant
vodka	Wine and Liquor	restaurant
wine	Wine and Liquor	restaurant

Table C.10 Sample List of Words that is Similar in Meaning - Hotel

Word	Same Feature Group	Source
sleeper	Amenities	Hotel
Coffee machine	Amenities	Hotel
Coffee maker	Amenities	Hotel
gym	Amenities	Hotel
Fitness center	Amenities	Hotel
Internet access	Amenities	Hotel
soap	Amenities	Hotel
shampoo	Amenities	Hotel
Gift shop	Facilities	Hotel
Lobby area	Facilities	Hotel
lobby	Facilities	Hotel
Ground floor	Facilities	Hotel
Desk staff	Front Desk	Hotel
Front desk	Front Desk	Hotel
receptionist	Front Desk	Hotel
reception	Front Desk	Hotel
bathroom	House keeping	Hotel
shower	House keeping	Hotel
toilet	House keeping	Hotel
towel	House keeping	Hotel
sofa	House keeping	Hotel
pillow	House keeping	Hotel
Bed sofa	House keeping	Hotel
King size	House keeping	Hotel

APPENDIX D

SYNONYM OF ADJECTIVES

In the study, adjectives were considered opinion words expressed about product features. Adjectives were identified and a synonym of an adjective was replaced utilizing a synonymous set in the WordNet in order to increase term frequency (e.g., magnificent became amazing, or cheapest became cheap). Sample List of synonym of adjectives are presented in the Table D.1

Table D.1 Sample List of Synonym of Adjectives

Synonyms of Adjectives	Adjectives	Synonyms of Adjectives	Adjectives	Synonyms of Adjectives	Adjectives
bloodthirsty	bloodthirstiest	ghastly	ghastliest	weighty	weightiest
bloodthirsty	bloodthirstier	ghostly	ghostliest	wintery	winteriest
unfriendly	unfriendliest	grizzly	grizzliest	beastly	beastlier
unfriendly	unfriendlier	haughty	haughtiest	chintzy	chintzier
foolhardy	foolhardiest	healthy	healthiest	cleanly	cleanlier
slaphappy	slaphappiest	intelligent	clever	courtly	courtlier
sprightly	sprightliest	lengthy	lengthiest	crumbly	crumblier
unhealthy	unhealthiest	naughty	naughtiest	doughty	doughtier
foolhardy	foolhardier	preachy	preachiest	earthly	earthlier
slaphappy	slaphappier	prickly	prickliest	flaunty	flauntier
sprightly	sprightlier	queenly	queenliest	fleshly	fleshlier
unhealthy	unhealthier	raunchy	raunchiest	flighty	flightier
draughty	draughtiest	scraggy	scraggiest	frizzly	frizzlier
friendly	friendliest	scrappy	scrappiest	ghastly	ghastlier
priestly	priestliest	scrawny	scrawniest	ghostly	ghostlier
princely	princeliest	scrubby	scrubbiest	grizzly	grizzlier
scraggly	scraggliest	scruffy	scruffiest	haughty	haughtier
stealthy	stealthiest	shapely	shapeliest	healthy	healthier
stretchy	stretchiest	shrubby	shrubbiest	interest	interest
ungainly	ungainliest	sightly	sightliest	lengthy	lengthier
amazing	breathtaking	sketchy	sketchiest	naughty	naughtier
draughty	draughtier	spindly	spindliest	preachy	preachier
friendly	friendlier	splashy	splashiest	prickly	pricklier
priestly	priestlier	springy	springiest	queenly	queenlier
princely	princelier	squashy	squashiest	raunchy	raunchier
scraggly	scragglier	squatty	squattiest	scraggy	scraggier
stealthy	stealthier	squiffy	squiffiest	scrappy	scrappier
stretchy	stretchier	starchy	starchiest	scrawny	scrawnier
ungainly	ungainlier	stately	stateliest	scrubby	scrubbier
amazing	magnificent	streaky	streakiest	scruffy	scruffier
beastly	beastliest	stringy	stringiest	shapely	shapelier
chintzy	chintziest	stroppy	stroppiest	shrubby	shrubbier
cleanly	cleanliest	swarthy	swarthiest	sightly	sightlier
courtly	courtliest	thirsty	thirstiest	sketchy	sketchier
crumbly	crumbliest	thready	threadiest	spindly	spindlier
doughty	doughtiest	thrifty	thriftiest	splashy	splashier
earthly	earthliest	throaty	throatiest	springy	springier
flaunty	flauntiest	tricksy	tricksiest	squashy	squashier

REFERENCES

Adami, G., P. Avesani, et al. (2003). "Bootstrapping for hierarchical document classification". Proceedings of the twelfth International Conference on Information and Knowledge Management. New Orleans, LA, USA, ACM**:** 295-302.

Agarwal, R. C., C. C. Aggarwal, et al. (2000). "Depth first generation of long patterns". the 6th ACM SIGKDD international conference on Knowledge discovery and data mining Boston, Massachusetts, US ACM SIGKDD.

Agrawal, R., T. Imieli, et al. (1993). "Mining association rules between sets of items in large databases." SIGMOD Rec. **22**(2): 207-216.

Agrawal, R., T. Imieliński, et al. (1993). "Mining association rules between sets of items in large databases". Management of data, Washington, D.C., US, ACM SIGMOD.

Agrawal, R. and R. Srikant (1994). "Fast Algorithms for Mining Association Rules in Large Databases". he 20th International Conference on Very Large Data Bases San Francisco, CA, USA, Morgan Kaufmann Publishers Inc.

Aken, J. E. v. and A. P. Nagel (2004). "Organising and managing the fuzzy front end of new product development". Netherlands, Eindhoven Centre for Innovation Studies**:** 18.

Amir, A., Y. Aumann, et al. (2005). "Maximal association rules: A tool for mining associations in text." Springer Science and Business Media, Inc.

Anwer, N., A. Rashid, et al. (2010). "Feature based opinion mining of online free format customer reviews using frequency distribution and Bayesian statistics". Networked computing and advanced information management (NCM), 2010 sixth international conference.

Bae, J. K. and J. Kim (2011). "Product development with data mining techniques: A case on design of digital camera." Expert systems with applications **38**(8): 9274-9280.

Bin, W. and L. Zhijing (2003). "Web Mining Research". Computational Intelligence and Multimedia Applications (ICCIMA'03), IEEE.

Binali, H., V. Potdar, et al. (2009). "A state of the art opinion mining and its application domains". Industrial technology. ICIT 2009. IEEE international conference, IEEE.

Bollegala, D., D. Weir, et al. (2011). "Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification". Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human

Language Technologies. Portland, Oregon, Association for Computational Linguistics. **1:** 132-141.

Bonchi, F., C. Castillo, et al. (2011). "Social network analysis and mining for business applications." ACM Trans. Intell. Syst. Technol. **2**(3): 1-37.

Brants, T. (2000). "TnT: a statistical part-of-speech tagger". Proceedings of the sixth conference on Applied natural language processing. Seattle, Washington, Association for Computational Linguistics**:** 224-231.

Chechik, G., A. Globerson, et al. (2005). "Information bottleneck for Gaussian variables." J. Mach. Learn. Res. **6**: 165-188.

Chung-Hong, L. and Y. Hsin-Chang (2005). "A classifier-based text mining approach for evaluating semantic relatedness using support vector machines". Information Technology: Coding and Computing.

Conrad, J. G. and F. Schilder (2007). "Opinion mining in legal blogs". Proceedings of the 11th international conference on Artificial intelligence and law. Stanford, California, USA, ACM**:** 231-236.

Dave, K., S. Lawrence, et al. (2003). "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews". Proceedings of the 12th international conference on World Wide Web. Budapest, Hungary, ACM**:** 519-528.

Dellarocas, C. (2003). "The digitization of word of mouth: Promise and challenges of online feedback mechanisms." Manage. Sci. **49**(10): 1407-1424.

Dellarocas, C. (2006). "Strategic manipulation of internet opinion forums: implications for consumers and firms." Management science **52**: 1577-1593.

Ding, X. and B. Liu (2007). "The utility of linguistic rules in opinion mining". Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. Amsterdam, Netherlands, ACM**:** 811-812.

Ding, X., B. Liu, et al. (2008). "A holistic lexicon-based approach to opinion mining". Proceedings of the international conference on Web search and web data mining. Palo Alto, California, USA, ACM**:** 231-240.

Ding, X., B. Liu, et al. (2009). "Entity discovery and assignment for opinion mining applications". Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France, ACM**:** 1125-1134.

Do, S. and N. Suh (2001). "Axiomatic design of software systems." CIRP Annals **49**(1): 95-100.

Du, W. and S. Tan (2009). "An iterative reinforcement approach for fine-grained opinion mining". Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Boulder, Colorado, USA, Association for Computational Linguistics**:** 486-493.

Durmuşoğlu, S. S. and G. Barczak (2011). "The use of information technology tools in new product development phases: Analysis of effects on new product innovativeness, quality, and market performance." Industrial Marketing Management **40**(2): 321-330.

Efron, M. (2004). "The liberal media and right-wing conspiracies: using cocitation information to estimate political orientation in web documents". Proceedings of the thirteenth ACM international conference on Information and knowledge management. Washington, D.C., USA, ACM**:** 390-398.

Etzioni, O., M. Cafarella, et al. (2005). "Unsupervised named-entity extraction from the Web: An experimental study." AAAI.

Galatis, G. (2011). "Business intelligence on 2.0 era a framework desscribe with emphasis on opinion mining.

Ganu, G., A. Marian, et al. (2010). "URSA - User review structure analysis: Understanding online reviewing trends". http://spidr-ursa.rutgers.edu, Rutgers DCS Technical Report No. 668.

Ge, P., L. Sipei, et al. (2010). "Sensitivity-proof content advertising based on two-stage txt cassification". Computational intelligence and software engineering (CiSE), 2010 international conference.

Giaglis, G. M., P. Kourouthanassis, et al. (2003). "Towards a classification framework for mobile location services". Idea group publishing, University of the Aegean, Greece**:** 1-18.

Guo, H., H. Zhu, et al. (2009). "Product feature categorization with multilevel latent semantic association". Proceedings of the 18th ACM conference on Information and knowledge management. Hong Kong, China, ACM**:** 1087-1096.

Guo, Q. (2010). "Research and improvement for feature selection on naive bayes text classifier". Future computer and communication (ICFCC), 2010 2nd international conference.

Han, J., J. Pei, et al. (2000). "Mining frequent patterns without candidate generation". the 2000 ACM SIGMOD international conference on Management of data, Dallas, Texas, US, ACM.

Hatzivassiloglou, V. and K. R. McKeown (1997). "Predicting the semantic orientation of adjectives". ACL '98 Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics archive Stroudsburg, PA, USA, Association for Computational Linguistics

Hatzivassiloglou, V. and K. R. McKeown (1997). "Predicting the semantic orientation of adjectives". Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics. Madrid, Spain, Association for Computational Linguistics**:** 174-181.

Hatzivassiloglou, V. and J. M. Wiebe (2000). "Effects of adjective orientation and gradability on sentence subjectivity". Proceedings of the 18th conference on Computational linguistics. Germany, Association for Computational Linguistics. **1:** 299-305.

Hayati, P. and V. Potdar (2008). "Evaluation of spam detection and prevention frameworks for email and image spam: a state of art". Proceedings of the 10th International Conference on Information Integration and Web-based Applications and Services. Linz, Austria, ACM**:** 520-527.

Hennig-Thurau, T., K. P. Gwinner, et al. (2004). "Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?" Journal of Interactive Marketing **18**(1): 38–52.

Holt, J. and S. Chung (2007). "Parallel mining of association rules from text databases." Journal of Supercomputing **39**(3): 273-299.

Hu, M. and B. Liu (2004) "Mining and summarizing customer reviews." Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 168-177 DOI: 10.1145/1014052.1014073.

Hu, M. and B. Liu (2004). "Mining opinion features in customer reviews". Proceedings of the 19th national conference on Artifical intelligence. San Jose, California, AAAI Press**:** 755-760.

Jeong, H., D. Shin, et al. (2011). "FEROM: Feature extraction and refinement for opinion mining." ETRI Journal **33**(5).

Jin, J. and Y. Liu (2010). "How to interpret the helpfulness of online product reviews: bridging the needs between customers and designers". Proceedings of the 2nd international workshop on Search and mining user-generated contents. Toronto, ON, Canada, ACM**:** 87-94.

Jindal, N. and B. Liu (2006). "Identifying comparative sentences in text documents". Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. Seattle, Washington, USA, ACM**:** 244-251.

Khan, A. and B. Baharudin (2011). "Sentiment classification using sentence-level semantic orientation of opinion terms from blogs". National postgraduate conference (NPC), 2011.

Khan, A., B. Baharudin, et al. (2010). "Sentence based sentiment classification from online customer reviews". Proceedings of the 8th International Conference on Frontiers of Information Technology. Islamabad, Pakistan, ACM**:** 1-6.

Khan, K., B. B. Baharudin, et al. (2009). "Mining opinion from text documents: A survey". Digital Ecosystems and Technologies, 2009. DEST '09. 3rd IEEE International Conference on.

Kim, Y.-g., T. Lee, et al. (2006) "Modified Naive Bayes classifier for e-catalog classification." **4055**, 246-257.

Kobayashi, N., K. Inui, et al. (2004) "Collecting evaluative expressions for opinion extraction." Nara Institute of Science and Technology.

Kohavi, R. (2001). "Mining e-commerce data: the good, the bad, and the ugly". the 7 International conference on Knowledge discovery and data mining San Francisco, California, ACM SIGMOD.

Kucera, H. (1980). "Computational analysis of predicational structures in English". Proceedings of the 8th conference on Computational linguistics. Tokyo, Japan, Association for Computational Linguistics**:** 32-37.

Landauer, T. K. and S. T. Dumais (1997) "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge." Psychological review.

Lee, Y.-B. and S. H. Myaeng (2002). "Text genre classification with genre-revealing and subject-revealing features." SIGIR’02, August 11-15, 2002, Tampere, Finland. ACM: 145-150.

Li, S. and Z. Chen (2010). "Exploiting web reviews for generating customer service surveys". Proceedings of the 2nd International Workshop on Search and Mining User-generated Contents. Toronto, ON, Canada, ACM**:** 53-62.

Li, W. and Y. B. Moon (2009). "A simulation study of mutual influences of engineering change management process and new product development process". Winter simulation conference. Austin, Texas, USA, Winter Simulation Conference**:** 2940-2950.

Lin, D. (1998). "Dependency based evaluation of MINIPAR", University of Manitoba Winnipeg, Manitoba, Canada**:** 14.

Liu, B. (2008). "Opinion mining". New York, NY, USA, Springer.

Liu, B. (2010). "Sentiment analysis and subjectivity", University of Illinois at Chicago, USA.

Liu, B. (2010). "Sentiment Analysis: A multi-faceted problem." IEEE Intelligent Systems.

Liu, B., M. Hu, et al. (2005). "Opinion observer: Analyzing and comparing opinions on the Web". Proceedings of the 14th international conference on World Wide Web. Chiba, Japan, ACM**:** 342-351.

Liu, H. (2004). "MontyLingua: Commonsense-enriched, end-to-end natural language understander for English." Retrieved 3/10/2012, from http://web.media.mit.edu/~hugo/montylingua/.

Lo, V. S. Y. (2002). "The true lift model: a novel data mining approach to response modeling in database marketing." ACM SIGKDD Explorations Newsletter **4**(2): 78-86.

Luole, Q. and C. Li (2011). "Comparison of model-based learning methods for feature-level opinion mining". Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference

MarketingChart. (2008). "Online reviews second only to word-of-mouth in purchase influence." Retrieved 9/29/2012, from http://www.marketingcharts.com/interactive/online-reviews-second-only-to-word-of-mouth-in-purchase-influence-6968/

Martineau, J. and T. Finin (2009). "Delta TFIDF: An improved feature space for sentiment analysis." In Proceedings of the Third AAAI Internatonal Conference onWeblogs and SocialMedia, San Jose, CA.

Menon, A. K. and C. Elkan (2011). "Fast algorithms for approximating the singular value decomposition." ACM transaction knowledge discovery data **5**(2): 1-36.

Mierswa, I., M. Wurst, et al. (2006). "YALE: Rapid prototyping for complex data mining tasks." Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining: 935-940.

Miller, G. A. (2012). "WordNet: Large lexical database of English", MIT.

Mobasher, B., H. Dai, et al. (2001). "Effective personalization based on association rule discovery from web usage data". 3rd international workshop on Web information and data management Atlanta, Georgia, USA, ACM, SIGIR.

Mohabbati, B., M. Hatala, et al. (2011). "Development and configuration of service-oriented systems families". Proceedings of the 2011 ACM Symposium on Applied Computing. TaiChung, Taiwan, ACM**:** 1606-1613.

Mudambi, S. M. and D. Schuff (2010). "What makes a helpful online review? A study of customer reviews on Amazon.com." MIS Quarterly **34**(1): 185-200.

Nasukawa, T. and J. Yi (2003). "Sentiment analysis: capturing favorability using natural language processing". Proceedings of the 2nd international conference on Knowledge capture. Sanibel Island, FL, USA, ACM**:** 70-77.

Nigam, K., J. Laerty, et al. (1999). "Using maximum entropy for text classication", Carnegie Mellon University, Pittsburgh, PA, USA.

P.Suh, N. (1990). "The principles of design ". New York, USA, Oxford University Press.

Pang, B. and L. Lee (2008). "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval. **2:** 1-135.

Pang, B., L. Lee, et al. (2002). "Thumbs up? Sentiment classification using machine learning techniques". Proceedings of the ACL-02 conference on empirical methods in natural language processing Association for Computational Linguistics. **10:** 79-86.

Park, Y. and S. Lee (2011). "How to design and utilize online customer center to support new product concept generation." Expert Systems with Applications **38**(8): 10638-10647.

Popescu, A.-M. and O. Etzioni (2005). "Extracting product features and opinions from reviews". Proceedings of the conference on human languagetechnology and empirical methods in natural language processing. Vancouver, British Columbia, Canada, Association for Computational Linguistics**:** 339-346.

Porter, M. F. (1980, 9/23/2012). "An algorithm for suffix stripping." Retrieved 8/12/2011, from http://tartarus.org/martin/PorterStemmer/def.txt.

Robles, V., P. Larranaga, et al. (2003). "Improvement of naive Bayes collaborative filtering using interval estimation". Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on.

Ruggieri, S. (2010). "Frequent regular itemset mining". Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. Washington, DC, USA, ACM**:** 263-272.

Sandhu, R. and R. Mehta (2011). "Applying opinion mining to organize Web opinions." International Journal of Computer Science, Engineering and Applications (IJCSEA) **1**.

Sarmento, L., P. Carvalho, et al. (2009). "Automatic creation of a reference corpus for political opinion mining in user-generated content". Proceedings of the 1st international CIKM workshop on Topic- Sentiment Analysis for Mass Opinion. Hong Kong, China, ACM**:** 29-36.

Scaffidi, C., K. Bierhoff, et al. (2007). "Red Opal: product-feature scoring from reviews". Proceedings of the 8th ACM conference on Electronic commerce. San Diego, California, USA, ACM**:** 182-191.

Schmitz, S. (2011). "A note on sequential rule-based POS tagging". Proceedings of the 9th International Workshop on Finite State Methods and Natural Language Processing. Blois, France, Association for Computational Linguistics**:** 83-87.

Schubert, L. and M. Tong (2003). "Extracting and evaluating general world knowledge from the Brown corpus". Proceedings of the HLT-NAACL 2003 workshop on text meaning Association for Computational Linguistics. **9:** 7-13.

Spath, H. (1985). "The cluster dissection and analysis theory FORTRAN programs examples". Upper Saddle River, NJ, USA Prentice-Hall, Inc.

Spertus, E. (1997). "Smokey: Automatic recognition of hostile messages." Innovative applications of artificial intelligence (IAAI): 1058–1065.

Su, Q., X. Xu, et al. (2008). "Hidden sentiment association in chinese web opinion mining". Proceedings of the 17th international conference on World Wide Web. Beijing, China, ACM**:** 959-968.

Suh, N. P. (2000). "Axiomatic design: Advances and applications ". New York, NY, USA, Oxford University Press

Tan, P.-N., M. Steinbach, et al. (2006). "Introduction to data mining". Boston, USA, Pearson Education, Inc.

Tatemura, J. (2000). "Virtual reviewers for collaborative exploration of movie reviews". Proceedings of the 5th international conference on Intelligent user interfaces. New Orleans, Louisiana, United States, ACM**:** 272-275.

Terveen, L., W. Hill, et al. (1997). "PHOAKS: a system for sharing recommendations." Commun. ACM **40**(3): 59-62.

Tishby, N., F. C. Pereira, et al. (1999). "The Information bottleneck method". The 37th annual Allerton Conference on Communication, Control, and Computing.

Tseng, C.-Y., P.-C. Sung, et al. (2008). "A novel email abstraction scheme for spam detection". Proceedings of the 17th ACM conference on Information and knowledge management. Napa Valley, California, USA, ACM**:** 1393-1394.

Turney, P. D. (2002). "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews". Proceedings of the 40th annual meeting on association for computational linguistics. Philadelphia, Pennsylvania, USA, Association for Computational Linguistics**:** 417-424.

Van Kleef, E., H. C. M. Van Trijp, et al. (2005). "Consumer research in the early stages of new product development: a critical review of methods and techniques." Food Quality and Preference **16**(3): 181-201.

Wahl, H., W. Winiwarter, et al. (2010). "Natural language processing technologies for developing a language learning environment". Proceedings of the 12th International Conference on Information Integration and Web-based Applications. Paris, France, ACM**:** 381-388.

Weiping, W. and Z. Yuanzhuang (2009). "E-business websites evaluation based on opinion mining". Electronic Commerce and Business Intelligence, 2009. ECBI 2009. International Conference.

Weishu, H., G. Zhiguo, et al. (2010). "Mining product features from online reviews". 2010 IEEE 7th International Conference on e-Business Engineering (ICEBE), Shanghai, China, IEEE.

Wiebe, J. M., R. F. Bruce, et al. (1999). "Development and use of a gold-standard data set for subjectivity classifications". Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. College Park, Maryland, USA, Association for Computational Linguistics**:** 246-253.

Wilson, G. and M. Heywood (2005). "Use of a genetic algorithm in brill's transformation-based part-of-speech tagger". Proceedings of the 2005 conference on Genetic and evolutionary computation. Washington DC, USA, ACM**:** 2067-2073.

Wu, X., V. Kumar, et al. (2007). "Top 10 algorithms in data mining." Knowl. Inf. Syst. **14**(1): 1-37.

Xiaojun, L., D. Lin, et al. (2010). "Opinion mining of camera reviews based on semantic role labeling". Fuzzy Systems and Knowledge Discovery (FSKD), 2010 7th International Conference on.

Yang, C.-S., C.-P. Wei, et al. (2009). "Extracting customer knowledge from online consumer reviews: a collaborative-filtering-based opinion sentence identification approach". Proceedings of the 11th international conference on electronic commerce. Taipei, Taiwan, ACM**:** 64-71.

Yi, J., T. Nasukawa, et al. (2003). "Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques". Data mining, ICDM 2003. Third IEEE international conference, IEEE.

Yu, B., S. Kaufmann, et al. (2008). "Exploring the characteristics of opinion expressions for political opinion classification". Proceedings of the 2008 International Conference on Digital Government Research. Montreal, Canada, Digital Government Society of North America**:** 82-91.

Zhai, Z., B. Liu, et al. (2010). "Grouping product features using semi-supervised learning with soft-constraints". Proceedings of the 23rd International Conference on Computational Linguistics. Beijing, China, Association for Computational Linguistics**:** 1272-1280.

Zhang, G., Y. Chen, et al. (2006). "A study on the relation between enterprise competitive advantage and CRM based on data mining". International technology and innovation conference. ITIC 2006, ITIC.

Zhang, L. and B. Liu (2011). "Identifying noun product features that imply opinions". Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, Oregon, USA, Association for Computational Linguistics. **2:** 575-580.

Zhang, S., W. Jia, et al. (2011). "Product features extraction and categorization in Chinese reviews." ICCGI 2011: The Sixth International Multi-Conferance on Computing in Global Information Technoogy.

Zhixing, L. (2010). "Product feature extraction with a combined approach". Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on.

Zhuang, L., F. Jing, et al. (2006). "Movie review mining and summarization". Proceedings of the 15th ACM international conference on Information and knowledge management. Arlington, Virginia, USA, ACM**:** 43-50.