Translating Online Customer Opinions into Engineering Characteristics in QFD: A Probabilistic Language Analysis Approach

Abstract

Online opinions provide informative customer requirements for product designers. However, the increasing volume of opinions make them hard to be digested entirely. It is expected to translate online opinions for designers automatically when they are launching a new product. In this research, an exploratory study is conducted, in which customer requirements in online reviews are manually translated into engineering characteristics (ECs) for Quality function deployment (QFD). From the exploratory study, a simple mapping from keywords to ECs is observed not able to be built. It is also found that it will be a time-consuming task to translate a large number of reviews. Accordingly, a probabilistic language analysis approach is proposed, which translates reviews into ECs automatically. In particular, the statistic concurrence information between keywords and nearby words is analyzed. Based on the unigram model and the bigram model, an integrated impact learning algorithm is advised to estimate the impacts of keywords and nearby words respectively. The estimated impacts are utilized to infer which ECs are implied in a given context. Using four brands of printer reviews from Amazon.com, comparative experiments are conducted. Finally, an illustrative example is shown to clarify how this approach can be applied by designers in QFD.

Keywords: Product Review Analysis; Customer Needs; Requirement Understanding; Customer Reviews; QFD; Product Design; Product Engineering Characteristics

1. INTRODUCTION

Requirement understanding plays an important role in product design. Conventionally, customer needs come from questionnaires which are mainly collected by customer investigations (Wang, Kannan and Azarm, 2011). It is often time-consuming and labor-intensive to obtain sufficient customer needs. Nowadays, customer needs and customer opinions are offered in many e-commerce websites, like Amazon.com, CNet.com, etc. A large number of online customer opinions are found in these websites. Many consumers are potentially influenced by online opinions in their purchase, and in the meantime, consumer preferences offered in online opinions are valuable for product designers (Liu, Lu and Loh, 2007; Li, Hitt and Zhang, 2011; Dou *et al.*, 2012). An example of Amazon Kindle DXG customer review, which is one representative type of online opinions, is presented in Figure 1.

[Insert Figure 1.]

As seen from the second and the third paragraph of this review, the customer complains that "…Anything greater than 100 pages becomes quite sluggish. Anything more than 500 pages is virtually unusable… PDF hyperlinks do not work in the kindle DX…" Generally speaking, these comments about the software of Kindle DXG provide helpful suggestions to the designers of Amazon Kindle DXG when they are conceiving to improve the current product model.

To facilitate designers to interpret customer needs, various models are proposed in the design area. One of the prevailing tools is Quality Function Deployment (QFD) which interprets customer needs into engineering characteristics

(ECs). For instance, in the previous example, in QFD, the customer need, “PDF hyperlinks”, could be interpreted as “software” of Kindle DXG, which is one EC. How to interpret customer needs is a vital step in QFD and many methods are introduced in different context (Fung, Chen and Tang, 2006; Kwong et al. 2007; Sener and Karsak, 2011). In these studies, customer requirements from survey data are utilized. These survey data are usually limited and they are often devised as goal-directed and well-formulated questionnaires.

However, online opinions are different with survey-based data. Online opinions are usually presented as free text data, which are submitted by consumers to express their' praise and concerns without the guide of purposive questions. Also, for popular products, a big volume of data are widely distributed in various websites. Distinguishes between survey-based data and online opinions lead to the difference in nature of the input for analyzing consumers' concerns, which makes the technical means in conventional method are technically arguable to exploit valuable customer needs from online opinions for product design in QFD. The valuable information in online opinions and the technical difficulties in conventional methods motive this research to explore an efficient and effective approach for designers to understand customer needs by using QFD and give an immediate response to consumers.

To understand how customer needs are analyzed by designers, a QFD exercise using consumer products is conducted by analyzing online reviews. In particular, customer requirements in online reviews are translated into ECs in QFD by designers manually since one hypothesis we have is that a simple mapping from some specific keywords to ECs is difficult to be built. Without an approach to translate online opinions into ECs, it will be a long haul task to digest online reviews by taking a huge number of online reviews into consideration. In order to help designers to translate online opinions into ECs in QFD, a probabilistic language analysis approach is proposed. In this approach, the weights of keywords and nearby words in online reviews, which contribute which ECs are most relevant to in each sentence, are learned by a unigram language model and a bigram language model. These learned weights are utilized to infer ECs with maximal possibility, which facilitates designers to translate online opinions into ECs in QFD.

The efforts of this research are at least three folds. First, a large number of online opinions are examined as customer requirements for product designers by using QFD. It is one of the first attempts to integrate a large number of online opinions into QFD directly. Second, online opinions in the form of free text, which are a new form of customer requirements, are exploited for product design. Specially, it is one challenging work to analyze unstructured customer needs for requirement analysis through QFD. Finally, this research highlights a possibility to build an approach to alleviate the burdens of product designers to digest online opinions and a probabilistic language analysis approach is proposed to translate customer opinions into ECs in QFD.

The rest of this paper is organized as follows: Section 2 reviews the related work about translating customer needs in QFD and analyzing online user-generated data for product design. In Section 3, an exploratory study is introduced on how customer requirements in online reviews are employed by using QFD. In this exploratory study, online reviews are shown to be manually translated into ECs in QFD. In Section 4, the problem to be studied in this research is defined. Section 5 introduces the technical approach proposed and describes technical details about the probabilistic language analysis approach and the integrated impact factor learning algorithm. Section 6 presents the experimental study and discusses its results. Also, in this section, an illustrative example using printer reviews from Amazon.com is reported to show how the proposed method is utilized by designers to translate online opinions into ECs in QFD for product design. Finally, Section 7 concludes.

2. RELATED WORK

2.1. Translating Customer Needs in QFD

In customer driven product design, after successfully identifying customer needs, designers start to consider how to interpret customer needs to improve their products. Specially, how to translate customer needs into ECs is one important question in QFD. Several contributions are made visible in this area.

Generally, in the design area, researches about translating customer needs into QFD have to cope with the inherent vagueness of human language and subjective judgment in the voice of the customers (Kim et al. 2000). This problem is often seen to be analyzed by introducing the fuzzy set theory into QFD. For instance, to meet customer needs and facilitate information sharing between designers, a market driven design system based on the fuzzy logic was developed (Harding et al. 2001). This system was utilized to translate market information into product specifications. Also, a fuzzy linear regression method was proposed to estimate the uncertainty in the functional relationship between customer needs and ECs for product planning by using QFD (Fung, Chen and Tang, 2006). In their later research, a fuzzy expert system was also proposed to identify important ECs (Kwong et al. 2007). The fuzzy relationship between customer needs and ECs as well as the fuzzy correlation among ECs in QFD were analyzed by this fuzzy expert system. For the maximization of customer satisfaction, a fuzzy multiple objective decision framework was also reported (Sener and Karsak, 2011). Given limited financial budget support, this framework was able to determine target levels of ECs by maximizing the extendibility and minimizing the technical difficulty of ECs at the same time.

Linguistic variables were found to be more appropriate to describe the inputs of QFD (Chen, Fung and Tang, 2006). The method using linguistic variables is different from the previous efforts where the input data were assumed to be precise and treated as numerical data only (Akao, 1990; Griffin and Hauser, 1993; Gustafsson and Gustafsson, 1994). However, linguistic variables were found sometimes difficult to be handled for the subjective assessments (Wang and Xiong, 2011). To ease this problem, an integrated linguistic-based group decision-making approach was proposed to cope with multiple types and multi-granularity linguistic assessments given by multiple decision-makers in QFD planning. This approach processes words in customer needs directly and minimizes the risk of loss of information, without translating linguistic information into various fuzzy numbers. In an uncertain and vague environment, Kano's Model was also reported to be integrated into QFD to quantify customer needs (Mu et al. 2008). A fuzzy multi-objective model was reported to be utilized to balance customer satisfaction and development cost.

2.2. Analyzing Online User-generated Data for Product Design

The value of online user-generated data is widely recognized to reflect customer opinions, but only a few studies are reported in the design area. In these studies, different approaches are proposed to analyze information for product design.

An automatic summarization approach was seen to analyze the topic structure of online reviews (Zhan, Loh and Liu, 2009). This approach was utilized to discover and assemble important topics in online reviews. The final summary of multiple reviews was then clustered by the topic structure, and different clusters were ranked according to the importance of different topics. From the perspective of consumers, they also compare different products. For this sake, a graphical model was adopted to extract the relationship between competing products from customer reviews (Xu et al, 2011). A two-level conditional random field model with unfixed interdependencies was employed to extract the dependencies between relations, entities and words of different reviews. Also, notice that, whether a product is welcome or not is usually reflected by the number of stars in e-commerce websites. Using rough set theory, inductive rule learning, and several information retrieval methods, an integrated system was developed to explore the relationship between the customer reviews and the review ratings (Chung and Tseng, 2012).

To catch the rapid change of customer needs for designers, a two-stage hierarchical process was ever built from online reviews (Lee, 2007). At the first stage, the association rule algorithm was used to cluster related product attributes and customer needs into hyper-edges. At the second stage, hyper-rules were applied on hyper-edges to track consumer needs. To extract customer needs from online reviews, an association rule mining algorithm was utilized on the basis of POS tags (Lee, 2009). A set of POS patterns were learnt from online reviews to bridge customer needs and ECs. However, the patterns were solely built on a sentential-level, such as subject-verb-object (SVO) triples. Different from this association rule method, preferences of product features, which are extracted from online reviews, were regarded as one type of time series data (Tucker and Kim, 2011). The Holt-Winters exponential smoothing techniques were utilized to predict the product preference trend. As observed from previous studies, all of these efforts neglect the quality of online user-generated data. In our latest work, how to identify helpful online reviews in the viewpoint of product designers was discussed (Liu et al, 2013). Based on a close study of how designers actually perceive the helpfulness of online reviews, four categories of features were extracted from online reviews and a classification method was formulated for that problem.

2.3. A Brief Summary

Requirement understanding is a critical step in customer driven product design. Various approaches were proposed to facilitate designers to translate customer needs in QFD. However, in these studies, customer survey data are regarded as customer needs only. Due to time or budget constraints, a rather limited number of customer and market survey data can be obtained manually. These customer survey data often contain formatted tables or targeted interview questions, and not many sentimental expressions are included. It is contrast to a large number of online customer opinions, which are available in different websites. They are presented in the form of free text, and sentimental words are one of the most obvious characteristic of online customer opinions. The difference between them makes online reviews not be processed efficiently and effectively by many existing models in the design area.

In terms of the expression of customer needs, the value of online user-generated data is widely accepted. But only a few researchers tried to analyze a large number of online user-generated data from the viewpoint of product designers. Different models are proposed to explore the value of online user-generated data in some aspects of product design. But, what they neglect is, online customer opinions, as one important type of customer needs, should be digested efficiently and integrated to ECs directly.

What differentiate this effort from existing efforts, several relevant studies were introduced. There is a visible research gap that how to analyze customer needs in online user-generated data in QFD for product design. In this research, the question to be explored is about how to translate online customer opinions into ECs in QFD. The objective of this research is to interpret online customer needs in the perspectives of product designers. It is a challenging and pressing research for product design, especially for understanding customer needs from online reviews and making strategic adjustments to improve products under the circumstance of customer-driven product design.

3. AN EXPLORATORY STUDY

Customer reviews, as one important type of online opinions, contain information about customer needs. To clearly understand how customer needs is utilized by product designers in QFD, an exploratory study is conducted.

One conjecture along with this QFD exercise is whether a simple mapping from some specific keywords to ECs could be built. For instance, online consumers use different words to refer to the same ECs and, sometimes, they also use the same words to imply different ECs in different context. Without an approach to translate online opinions into ECs, it is a time-consuming task to analyze online reviews. Hence, in this exploratory study, this conjecture will be examined. By using QFD, customer requirements from online opinions are analyzed manually. Specifically, these customer requirements from online opinions are translated into ECs.

3.1. Data Collection

In this exploratory study, 770 reviews of four popular color printers (two Epson printers and two HP printers) were selected as examples from Amazon.com and Epson.com. All these four popular color printers are consumer products and there are many targeted customer reviews online. They are “Epson Artisan 810”, “Epson

WorkForce 610”, “HP Officejet 6500”, and “HP Photosmart Premium C309”. For short, they are named as “A810”, “W610”, “H6500”, and “C309” in Table 1, which shows the number of reviews for each model.

[Insert Table 1]

Some statistic information considering the number of words and the number of sentences in each set of reviews is shown in Table 2.

[Insert Table 2]

As noticed from Table 2, on average, there are about 208 words in each review. However, they do not distribute evenly. The maximal number of words in a single review can hit 2,823. Meanwhile, in terms of the number of sentences contained in a single review, it also distributes unevenly with an average number of 11.6 sentences, with a record of 85 as its highest.

3.2. A QFD Exercise Using Consumer Products

Generally, it is hard to invite experienced designers for specific products to evaluate customer needs because of some confidentiality regulations in business. In this exploratory study, the QFD exercise was taken by two service engineers in product design background, who acted as product designers. These engineers were working in Epson Hong Kong and HP Hong Kong. They are very familiar with printers and they had a sound understanding about customer needs. Also, they both had some experience with printer design using QFD, which contributed in building a high quality dataset. In this exploratory study, they were asked to translate them into ECs manually, which is one important step in QFD.

In the first place, a list of ECs was collectively suggested by the two annotators. They are listed in Table 3.

[Insert Table 3]

Then, two annotators were asked to label these printer reviews. Each annotator began to read all of them and distinguish the keywords in each review sentence. Here "keyword" is either a word or a phrase referring to an EC. Occasionally, only one word is utilized to refer to an EC, while, in some other cases, a phrase is employed. If a sentence contains some keywords, they are highlighted in a separate column. The correlated sentiment about the EC is written in the corresponding blank. The linkage value is the customer sentiment linked to certain ECs, which is denoted by "–2", "–1", "0", "1" and "2". A "–2" means the least satisfied and "2" means the most satisfied. In Figure 2, an example of one A810 review analysis is shown. For the conciseness, only nine ECs in Table 3 are presented and some of sentences in this review are illustrated.

[Insert Figure 2]

In this example, the first sentence, "I only have the printer for a few days, but so far I am very pleased", does not contain any keywords associated with ECs, so the keyword column in this line is "-". The seventh line in Figure 2 is "the paper tray feels a bit flimsy, but is easy to remove or insert, and there's no fuss to loading your paper in it." This consumer actually complained about the "Hopper Unit" through the phrase “paper tray”, so the annotators wrote this phrase in the keyword column of the seventh line. If more than one ECs are mentioned in a sentence, the annotators copied this sentence, pasted it into the other line and labeled the second item in a new line. For example, in Figure 2, the third sentence mentions “the actual printing is quiet, and of great quality.” Two ECs are noticed. The word “quiet” is translated into “Noise” and the phrase “great quality” is associated with “Print Quality”. This sentence is repeated in order to clearly label the two keywords. A similar case is observed in the third and fourth line shown in Figure 2. Finally, review annotations were double cross checked by two annotators to avoid any mislabeling.

3.3. Observations and Discussions

It took more than two weeks to finish the QFD exercise using 770 printer reviews for the two annotators, including their cross double check. It implies that they have to read and label these online reviews at a speed of about 100 reviews per day without any interruption. They complained that it is a boring and error-prone task to complete this QFD exercise and they expect an automatic review analysis approach for the sake of efficient customer requirement analysis from online opinions.

Also, it is found that a naive one to one mapping from keywords in online reviews to ECs does not always exist. A specific keyword might be translated into one EC in some review sentences, but this keyword is also possibly translated into another EC in other sentences. They complained that they often need to read sentences containing keywords as well as those sentences in the left hand side and in the right hand side several times, and disambiguate the meaning of these keywords in order to digest customer needs and understand which ECs are referred to. Take the word "paper" as an example. One sentence is "...It obviously needs a more absorbent paper because..." and the other is "...Very easy to swap in alternate papers, always easy to see if paper is left..." In the first sentence, the word "paper" is utilized to refer to "Supported Paper", while the other sentence is referring to "Hopper Unit".

More specifically, in this exploratory study, the number of keywords which are translated into different ECs from the four datasets of printer reviews is shown in Table 4. As seen from Table 4, it is a prevailing case that many keywords are translated into different ECs.

[Insert Table 4]

An interesting relationship between keywords in online reviews and ECs is presented, which implies that a consumer topic in one review sentence may not always be described by a single keyword. It makes that, without an automatic approach for understanding the relationship, to digest online customer opinions, a time-consuming and labor-intensive QFD exercise have to be taken. A clear research gap is that there is not an approach to translate online customer opinions into ECs for designers. However, it is one of essential step to understand customer requirements when QFD or a similar form of requirement analysis is performed for the improvement of current product models.

The phenomenon that the same keywords are translated into different ECs triggers this research to explore how to translate online reviews into ECs in QFD automatically.

4. PROBLEM DEFINITION

The crucial step in QFD is to translate customer needs into ECs. In this research, by assuming that customer needs are reflected in opinions contained in online reviews, the objective is to find a way that automatically relate online customer opinions with ECs in QFD. In order to better define the working context of this research, some notations are introduced at first.

In a typical e-commerce website, such as Amazon.com, each product has a list of reviews, r_1, r_2, ..., r_p. These reviews contain customer opinions, which help designers to identify customer needs. In the product design area, especially in market driven design, customer needs are encouraged to be translated into ECs in QFD. A list of ECs is denoted as $EC = <ec_1, ec_2... ec_n>$. Examples of ECs are demonstrated in Table 3.

In QFD, to understand customer needs and requirements, designers need to identify which ECs are mentioned from online reviews. ECs are usually referred by some keywords in online reviews, and one or multiple ECs can be possibly covered in one single review. Here a keyword W_T points to the most important word or phrase which refers to an EC. For example, in the exploratory study, the designers read customer reviews and highlight the keywords in each sentence to indicate which ECs are referred to.

In the exploratory study, it is found that a naive one to one mapping from keywords to ECs for labeled keywords does not always exist, which makes product designers have to read, analyze and label online reviews manually. However, it is always expensive to read all online reviews and to mark which ECs are pointed to. If only a small number of reviews are considered, a biased conclusion might be drawn regarding customer concerns. Hence, an automatic approach is targeted to relate online reviews with ECs for product design, which help designers to translate customer concerns in an engineering language efficiently. It is a challenging work for product designers.

Accordingly, the problem to be explored in this research is defined as, given a keyword W_T in one customer review r_i, how to translated W_T into a most proper EC ec_k automatically in QFD for product design.

5. METHODOLOGY

In the QFD exercise, there does not always exist an one to one mapping from keywords in online opinions to ECs. However, read the entire set of online opinions and label which ECs are referred to is expensive. An automatic review analysis approach is required to help designers to identify referred ECs efficiently. In this study, a probabilistic language analysis approach is proposed to translate online opinions into ECs.

5.1. A Probabilistic Language Analysis Approach

Notice that, from the labeling results in the QFD exercise, annotators indicates that those words near keywords in either side are helpful to disambiguate which ECs are referred to. Accordingly, context words W_c of a keyword W_T are defined as a set of words involved in a text window around W_T, including left N_L words, right N_R words and W_T itself. The left N_L words constitute a word set W_L, and right N_R words constitute a word set W_R. Accordingly, context words are argued to be helpful for designers to identify which ECs are referred to.

Suppose a keyword W_T, in context W_c, is labeled as one EC ec_p, rather than another EC ec_q. In other words, in context W_c, the possibility that W_T is translated into ec_p is higher than the possibility that W_c is translated into ec_q. Hence, in context W_c, if the possibility where W_T is translated into ec_p is defined as $p(W_c, ec_p)$, $p(W_c, ec_p)$ is argued to be higher than $p(W_c, ec_q)$, denoted as,

$$p(W_c, ec_p) \geq p(W_c, ec_q) \tag{1}$$

For instance, one review sentence is, "... nice feature is that when you plug in a camera memory card in the front ..." After reading this sentence, designers translate the word "card" to "card slot" of the printer, which is one EC in Table 3. However, in the other sentence, "card" might be utilized to refer to "consumable replacement", a different EC from "card slot" in Table 3. Following Equation (1), this example is denoted as:

$$p(W_c, \text{"card slot"}) \geq p(W_c, \text{"consumable replacement"}) \tag{2}$$

In order to infer which ECs are pointed to, the words in the context W_c are considered. However, intuitively, not all the words in the context W_c affect designers' analysis when they analyze which EC is the keyword "card" referring to. For instance, some adverbs, such as, "only", "there", etc., frequently appearing in online reviews. These words are generally regarded as having less effect for designers in translating online reviews into ECs. Hence, after stemming and stop words removal on review sentences with the techniques of language processing, adverbs are filtered out in both the left word set W_L and the right word set W_R.

Without the loss of generality, according to Bayesian rules, the possibility of W_c translated into ec_p, $p(W_c, ec_p)$, is equivalently derived as shown in Model (3).

$$\begin{aligned} & p(W_c, ec_p) \\ &= p(ec_p)p(W_c \mid ec_p) \\ &= p(ec_p)p(W_L, W_R, W_T \mid ec_p) \\ &= p(ec_p)p(W_L \mid ec_p)p(W_R \mid ec_p)p(W_T \mid ec_p) \end{aligned} \tag{3}$$

$p(ec_p)$ is interpreted as the probability that consumers point to ec_p. Empirically, in frequency-count based statistics, it is estimated by the percentage of training samples translating into ec_p. In Model 3, given the context that the keyword W_T is translated into ec_p, $p(W_L \mid ec_p)$, $p(W_R \mid ec_p)$ and $p(W_T \mid ec_p)$ are interpreted as the occurrence possibilities for the left words of W_T, for the right words of W_T, and for W_T itself, respectively.

Now, the problem is to model marginal probabilities, $p(W_T \mid ec_p)$, $p(W_L \mid ec_p)$ and $p(W_R \mid ec_p)$.

5.2. Language Modeling

Two models, a "unigram" model and a "bigram" model, are utilized to estimate the marginal probabilities $p(W_L \mid ec_k)$ and $p(W_R \mid ec_k)$. They are two variants of N-gram model, which is the prevailing method in the field of statistical language modeling. Language models are useful in a broad range of applications, such as speech recognition, machine translation, biological sequence analysis, etc. A N-gram is a contiguous sequence of N tokens from a given sequence. Accordingly, N-gram language models are often regarded as placing a small window over a sequence, in which only N tokens are considered at the same time. The word "unigram" refers to an N-gram of size one and the word "bigram" refers to size two.

(1) The unigram model

The simplest N-gram model is the unigram model and it can be thought of as a window that shows barely one single token at a time. In the unigram model, the probability of hitting an isolated word is calculated, without considering any influence from the words before or after the keywords. The probability of the occurrence of each word merely depends on the word itself. Consequently, the probability of the left context words is modeled as:

$$\begin{aligned} & p(W_L \mid ec_k) \\ &= p(W_{L_1} W_{L_2} \ldots W_{L_{N_L-1}} W_{L_{N_L}} \mid ec_k) \\ &= p(W_{L_1} \mid ec_k) p(W_{L_2} \mid ec_k) p(W_{L_3} \mid ec_k) \ldots p(W_{L_{N_L}} \mid, ec_k) \\ &= \prod_{i=1}^{N_L} p(W_{L_i} \mid, ec_k) \end{aligned} \tag{4}$$

$p(W_{Li} \mid ec_k)$ is the probability, given that W_T is translated into ec_k. $p(W_{Li} \mid ec_k)$ is estimated as the occurrence possibilities of left words W_{Li},

$$p(W_{Li} \mid ec_k) = \frac{c(W_{Li}, ec_k)}{\mid ec_k \mid} \tag{5}$$

$c(W_{Li}, ec_k)$ is the count that W_{Li} is translated into ec_k. $|ec_k|$ is the count of sentences that ec_k is mentioned. Notice that, if $c(W_{Li}, ec_k)$ equals to zero, $p(W_{Li} \mid ec_k)$ will be zero. It makes $p(W_L \mid ec_k)$ equal to zero, which induces the model inapplicable. To avoid this problem, a Dirichlet Priors smoothing method is utilized. In the Dirichlet Priors smoothing method, the probability is parameterized with a prior probability based on the training corpus:

$$p(W_{Li} \mid ec_k) = \frac{c(W_{Li}, ec_k) + \mu p(W_{Li} \mid \boldsymbol{C})}{\mid ec_k \mid + \mu} \tag{6}$$

μ is a constant that tunes the weight of the smoothing item. $p(W_{Li} \mid \boldsymbol{C})$ is the probability that W_{Li} occurs in the training corpus $\boldsymbol{C}$.

(2) The bigram model

The unigram model is often argued that it is not very informative since these are barely the words that form the sequences. The bigram model can be thought of as a window that shows two tokens at a time. All bigrams of a sequence can be found by a window on its first two tokens and by sliding this window, size of two, to the right one token at a time in a stepwise manner. This procedure is repeated until the last two tokens are covered by the window.

Rather than calculating the probability of hitting an isolated word, the bigram model considers the influence from the words before or after each specific word. In bigram model, the probability of each word depends on its own word and the previous word next to it, which means that $p(W_{L_i} \mid W_{L_1}, W_{L_2}, ..., W_{L_{i-1}}, ec_k) = p(W_{L_i} \mid W_{L_{i-1}}, ec_k)$. Accordingly, the marginal probability $p(W_L \mid ec_k)$ can be inferred as:

$$
\begin{aligned}
& p(W_L \mid ec_k) \\
&= p(W_{L_1} W_{L_2} ... W_{L_{N_L-1}} W_{L_{N_L}} \mid ec_k) \\
&= p(W_{L_1} \mid ec_k) p(W_{L_2} \mid W_{L_1}, ec_k) p(W_{L_3} \mid W_{L_1}, W_{L_2}, ec_k) ... p(W_{L_{N_L}} \mid W_{L_1}, W_{L_2}, ..., W_{L_{N_L-1}}, ec_k) \\
&= p(W_{L_1} \mid ec_k) p(W_{L_2} \mid W_{L_1}, ec_k) p(W_{L_3} \mid W_{L_2}, ec_k) ... p(W_{L_{N_L}} \mid W_{L_{N_L-1}}, ec_k) \\
&= p(W_{L_1} \mid ec_k) \prod_{i=1}^{N_L-1} p(W_{L_{i+1}} \mid W_{L_i}, ec_k)
\end{aligned}
\tag{7}
$$

Similarly, in the bigram model, the Dirichlet Priors smoothing method is also utilized to avoid the zero probability problem for $p(W_{Li+1} \mid W_{Li}, ec_k)$:

$$
p(W_{Li+1} \mid W_{Li}, ec_k) = \frac{c(W_{Li+1}, W_{Li}, ec_k) + \mu P(W_{Li+1} \mid \boldsymbol{C})}{c(W_{Li}, ec_k) + \mu} \tag{8}
$$

$c(W_{Li+1}, W_{Li}, ec_k)$ is the count that word W_{Li} and word W_{Li+1} are both mentioned in the context and the keyword is translated into ec_k. $c(W_{Li}, ec_k)$ is the count of frequency that W_{Li} is mentioned in the context.

5.3. Impact Factor

In the previous example, the word “card” is found to refer to “card slot” of the printer and the words in the context W_c are argued to be helpful to understand which ECs are referred to.

However, in the previous example, it is assumed that words closer to “card” possess a greater impact. The impact tends to be weaker for words in a comparatively further distance. For example, compared with “camera” and “memory” which are near the keyword “card”, two words “nice” and “feature” may have a relatively weaker impact for designers to understand consumers’ concern “card slot”. Generally, the impact of W_T and the impacts of words in W_L and W_R should not be the same in terms of helping designers to clear which EC W_T refers to. Those words that are closer to keywords are supposed to have greater impacts and the impacts tend to be weaker for relatively farther words.

Without a loss of generality, α, β and γ are defined as the impact factors for the left words of W_T in the context W_c, the right words, and W_T itself, respectively. By embedding impact factors into Bayesian rules, the possibility of W_c being translated into ec_p, $p(W_c, ec_p)$, is equivalently derived as Model (9).

$$\begin{aligned} & p(W_c, ec_p) \\ &= p(ec_p)p(W_c \mid ec_p) \\ &= p(ec_p)p(W_L \mid ec_p)^{\alpha} p(W_R \mid ec_p)^{\beta} p(W_T \mid ec_p)^{\gamma} \end{aligned} \tag{9}$$

Now, if $p(W_c, ec_p)$ is supposed to be bigger than $p(W_c, ec_q)$ as Model (1) describes, according to Model (9), it is inferred as,

$$\begin{aligned} & p(W_c, ec_p) > p(W_c, ec_q) \\ & \sim \frac{p(ec_p)p(W_L \mid ec_p)^{\alpha} p(W_R \mid ec_p)^{\beta} p(W_T \mid ec_p)^{\gamma}}{p(ec_q)p(W_L \mid ec_q)^{\alpha} p(W_R \mid ec_q)^{\beta} p(W_T \mid ec_q)^{\gamma}} > 1 \\ & \sim \log \frac{p(ec_p)p(W_L \mid ec_p)^{\alpha} p(W_R \mid ec_p)^{\beta} p(W_T \mid ec_p)^{\gamma}}{p(ec_q)p(W_L \mid ec_q)^{\alpha} p(W_R \mid ec_q)^{\beta} p(W_T \mid ec_q)^{\gamma}} > 0 \\ & \sim \alpha \log \frac{p(W_L \mid ec_p)}{p(W_L \mid ec_q)} + \beta \log \frac{p(W_R \mid ec_p)}{P(W_R \mid ec_q)} + \gamma \log \frac{p(W_T \mid ec_p)}{p(W_T \mid ec_q)} + \log \frac{p(ec_p)}{p(ec_q)} > 0 \end{aligned} \tag{10}$$

As mentioned, in frequency-count based statistics, $p(ec_p)$ is estimated by the percentage of training samples translating into ec_p. Accordingly, $\frac{p(ec_p)}{p(ec_q)}$ is approximated as:

$$\log\frac{p(ec_p)}{p(ec_q)}=\log\frac{|ec_p|}{|ec_q|} \tag{11}$$

$|ec_p|$ is the count that ec_p is mentioned in training data. It makes the item $\log\frac{p(ec_p)}{p(ec_q)}$ to be a determinant item, which is derived from training data directly.

$p(W_L \mid ec_k)$, $p(W_R \mid ec_k)$, and $p(W_T \mid ec_k)$ are estimated from the training corpus by the unigram model and the bigram model. Hence, in the unigram model, according to Equation (4), $\alpha\log\frac{p(W_L \mid ec_p)}{p(W_L \mid ec_q)}$ and $\beta\log\frac{p(W_R \mid ec_p)}{p(W_R \mid ec_q)}$ are written as:

$$\begin{aligned}\alpha\log\frac{p(W_L \mid ec_p)}{p(W_L \mid ec_q)}&=\sum_{i=1}^{N_L}\alpha_i\log\frac{p(W_{Li} \mid ec_p)}{p(W_{Li} \mid ec_q)}\\ \beta\log\frac{p(W_R \mid ec_p)}{p(W_R \mid ec_q)}&=\sum_{i=1}^{N_R}\beta_i\log\frac{p(W_{Ri} \mid ec_p)}{p(W_{Ri} \mid ec_q)}\end{aligned} \tag{12}$$

In the bigram model, according to Equation (7), $\alpha\log\frac{p(W_L \mid ec_p)}{p(W_L \mid ec_q)}$ and $\beta\log\frac{p(W_R \mid ec_p)}{p(W_R \mid ec_q)}$ are written as:

$$\begin{aligned}\alpha\log\frac{p(W_L \mid ec_p)}{p(W_L \mid ec_q)}&=\alpha_1\log\frac{p(W_{L1} \mid ec_p)}{p(W_{L1} \mid ec_q)}+\sum_{i=1}^{N_L-1}\alpha_{i+1}\log\frac{p(W_{Li+1} \mid W_{Li},ec_p)}{p(W_{Li+1} \mid W_{Li},ec_q)}\\ \beta\log\frac{p(W_R \mid ec_p)}{p(W_R \mid ec_q)}&=\beta_1\log\frac{p(W_{R1} \mid ec_p)}{p(W_{R1} \mid ec_q)}+\sum_{i=1}^{N_R-1}\beta_{i+1}\log\frac{p(W_{Ri+1} \mid W_{Ri},ec_p)}{p(W_{Ri+1} \mid W_{Ri},ec_q)}\end{aligned} \tag{13}$$

Hence, the unknown parameters in the last inequality of Model (10) are α, β, and γ. α, β, and γ are shared parameters for all reviews. As mentioned, $\log\frac{p(ec_p)}{p(ec_q)}$ is a determinant item, which implies that α, β, and γ have to be tuned to make the inequality in Model (10) to be satisfied for all reviews. Equally, the sum of three parameter-dependent items, $\alpha\log\frac{p(W_L \mid ec_p)}{p(W_L \mid ec_q)}$, $\beta\log\frac{p(W_R \mid ec_p)}{p(W_R \mid ec_q)}$, and

$\gamma \log \frac{p(W_T \mid ec_p)}{p(W_T \mid ec_q)}$ should be tuned as high as possible to make the inequality in Model (10) to be satisfied.

The focus of this research should be on maximizing the sum of three parameter-dependent items by tuning α, β, and γ. According to the probabilistic language analysis approach, an integrated impact learning algorithm is built in order to estimate the impact of W_T and the impacts of words in W_L and W_R from a set of training corpus.

In this integrated impact learning algorithm, according to Model (1) as well as the impact of W_T and the impacts of words in W_L and W_R, given a context W_c, the probability that W_T is translated into ec_p can be compared with the probability that W_T is translated into ec_q. Finally, whether a specific EC ec_k receives the largest possibility is calculated. After that, the EC, which receives the largest possibility, is regarded as the one that W_T refers to.

5.4. An Integrated Impact Learning Algorithm

In this section, an integrated impact learning algorithm is proposed to learn α, β, and γ. By utilizing the learned impact factors, the EC receiving the highest possibility is regarded as the one that W_T is associated with, which accomplishes the task of translating online reviews into ECs in QFD.

As mentioned in Section 5.3, the sum of three parameter-dependent items in the last inequality in Model (10) should be maximized by tuning impact factor α, β, and γ. Accordingly, a function, $Ratio(\alpha, \beta, \gamma)$, is defined as:

$$Ratio(\alpha,\beta,\gamma)=\sum_{ec_p}\sum_{ec_q}\{\alpha\log\frac{p(W_L\mid ec_p)}{p(W_L\mid ec_q)}+\beta\log\frac{p(W_R\mid ec_p)}{p(W_R\mid ec_q)}+\gamma\log\frac{p(W_T\mid ec_p)}{p(W_T\mid ec_q)}\} \quad (14)$$

Hence, the question becomes to tune *α*, *β*, and *γ* to maximize *Ratio*(*α*, *β*, *γ*),

$$max\ Ratio(\alpha, \beta, \gamma) \quad (15)$$

Next, the ratio function *Ratio*(*α*, *β*, *γ*) of the unigram model and the bigram model are defined respectively as followings:

$$\begin{aligned}Ratio_1(\alpha,\beta,\gamma)=\sum_{ec_p}\sum_{ec_q}\{&\sum_{i=1}^{N_L}\alpha_i\log\frac{p(W_{Li}\mid ec_p)}{p(W_{Li}\mid ec_q)}+\sum_{i=1}^{N_R}\beta_i\log\frac{p(W_{Ri}\mid ec_p)}{p(W_{Ri}\mid ec_q)}+\\&\gamma\log\frac{p(W_T\mid ec_p)}{p(W_T\mid ec_q)}\}\end{aligned} \quad (16)$$

$$\begin{aligned}Ratio_2(\alpha,\beta,\gamma)=\sum_{ec_p}\sum_{ec_q}\{&\alpha_0\log\frac{p(W_{L1}\mid ec_p)}{p(W_{L1}\mid ec_q)}+\beta_0\log\frac{p(W_{R1}\mid ec_p)}{p(W_{R1}\mid ec_q)}+\\&\sum_{i=1}^{N_L-1}\alpha_i\log\frac{p(W_{Li+1}\mid W_{Li},ec_p)}{p(W_{Li+1}\mid W_{Li},ec_q)}+\\&\sum_{i=1}^{N_R-1}\beta_i\log\frac{p(W_{Ri+1}\mid W_{Ri},ec_p)}{p(W_{Ri+1}\mid W_{Ri},ec_q)}+\\&\gamma\log\frac{p(W_T\mid ec_p)}{p(W_T\mid ec_q)}\}\end{aligned} \quad (17)$$

Notice that *α*, *β*, and *γ*, which are learned from Model (15), define an optimal determinant function that separates the most appropriate EC with others. Similar to the prevailing machine learning algorithm SVM (Support Vector Machines), which maximizes the margin of hyper-planes and leaving much room in classifying unseen data. Model (15) intends to maximize the probability ratio between ECs and surfaces the one enjoying a higher probability for W_T. In SVM, a normalization term is used in the objective function to avoid the case where parameters used to define the hyper-planes become too large. Likewise, normalization terms also should be applied here. To combine the normalization terms with *Ratio*(*α*, *β*, *γ*), a loss function is then defined as:

$$Loss(\alpha,\beta,\gamma) = -Ratio(\alpha,\beta,\gamma) + C_1\frac{\|\alpha\|^2}{2} + C_2\frac{\|\beta\|^2}{2} + C_0\frac{\|\gamma\|^2}{2} \tag{18}$$

–Ratio(α, β, γ) is introduced in *Loss*(α, β, γ) since *max Ratio*(α, β, γ) is essentially equal to *min –Ratio*(α, β, γ). The corresponding weights of the normalization terms are tuned by C_1, C_2 and C_0 for α, β, and γ respectively.

Similarly, the loss function *Loss*(α, β, γ) for the unigram model and the bigram model are defined as:

$$Loss_1(\alpha,\beta,\gamma) = -Ratio_1(\alpha,\beta,\gamma) + C_1\frac{\sum_{i=1}^{N_L}\alpha_i^2}{2} + C_2\frac{\sum_{i=1}^{N_R}\beta_i^2}{2} + C_0\frac{\|\gamma\|^2}{2} \tag{19}$$

$$Loss_2(\alpha,\beta,\gamma) = -Ratio_2(\alpha,\beta,\gamma) + C_1\frac{\sum_{i=1}^{N_L}\alpha_i^2}{2} + C_2\frac{\sum_{i=1}^{N_R}\beta_i^2}{2} + C_0\frac{\|\gamma\|^2}{2} \tag{20}$$

γ is a scalar, while both α and β are vector parameters:

$$\begin{aligned} \alpha &= (\alpha_1,\alpha_2,\ldots,\alpha_{N_L})^T \\ \beta &= (\beta_1,\beta_2,\ldots,\beta_{N_R})^T \end{aligned} \tag{21}$$

Also, arguably, those words closer to W_T are regarded as having a higher weights in translating W_T into a correct EC. Accordingly, if the impact factor of a word at distance i to the left side of W_T is α_i, α_i is expected to be larger than α_{i+1}. Similarly, the impact factor β_j of a word at distance j to the right side of W_T should be bigger than β_{j+1}. Mathematically, they are denoted as:

$$\begin{aligned} &\forall i \in [1, N_L - 1], \alpha_i > \alpha_{i+1} \\ &\forall j \in [1, N_R - 1], \beta_j > \beta_{j+1} \end{aligned} \tag{22}$$

N_L and N_R are the number of words in W_L and W_R. Two sets of constraints are applied here,

$$
\begin{aligned}
& M \cdot \alpha \leq \mathbf{0} \\
& M \cdot \beta \leq \mathbf{0} \\
& M = \begin{bmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix}
\end{aligned} \tag{23}
$$

The bold "**0**" denotes a zero vector. Since all α_i, β_j, and γ are defined as the impact factors, they are suggested to be nonnegative and not greater than one.

$$
\begin{aligned}
& \mathbf{0} \leq \alpha \leq \mathbf{1} \\
& \mathbf{0} \leq \beta \leq \mathbf{1} \\
& 0 \leq \gamma \leq 1
\end{aligned} \tag{24}
$$

The bold "**1**" denotes a vector consisting of one in all its dimensions. Combining the loss function in Model (18), with the constraints in Model (23) and Model (24), finally, the optimization model is established as,

$$
\begin{aligned}
& \min Loss(\alpha, \beta, \gamma) \\
s.t. \quad & M \cdot \alpha \leq \mathbf{0} \\
& M \cdot \beta \leq \mathbf{0} \\
& \mathbf{0} \leq \alpha \leq \mathbf{1} \\
& \mathbf{0} \leq \beta \leq \mathbf{1} \\
& 0 \leq \gamma \leq 1
\end{aligned} \tag{25}
$$

More specifically, by embedding Model (19) for unigram model or Model (20) for bigram model into the optimization problem (25), a quadratic programming problem is presented, which can be solved by standard optimization methods, such as the gradient descent, etc.

In summary, the integrated impact learning algorithm is described in Algorithm 1. Accordingly, the impact factor is learned from a training corpus. The performance comparison of both unigram model and bigram model are reported in the next section.

Algorithm 1: Impact factor learning Algorithm

for each keywords W_T connecting with multiple ECs in training set **do**
  $S \leftarrow$ {review sentences labeled with W_T},
  $EC \leftarrow$ {all possible ECs for W_T}
  for $S_i \in S$ **do**
    Stemming, stop words removal on S_i
    POS tagging S_i and words filtering with certain *POS*
    $W_L \leftarrow$ {left N_L words of W_T },
    $W_R \leftarrow$ {right N_R words of W_T },
    $ec_p \leftarrow$ {the EC that W_T relates in S_i }
    For $ec_p, ec_q \in EC$ **do**
      Calculate $\log\frac{p(W_L \mid ec_p)}{p(W_L \mid ec_q)}$, $\log\frac{p(W_R \mid ec_p)}{p(W_R \mid ec_q)}$, and $\log\frac{p(W_T \mid ec_p)}{p(W_T \mid ec_q)}$
    End for
  End for
end for
Solve the optimization problem as described in Model (25)
Return α, β, γ

6. EXPERIMENTAL STUDY AND DISCUSSIONS

6.1. Experiment Setup

The four sets of printer reviews, introduced early in Section 3, are adopted in the experimental study, along with the labeled keywords and the corresponding ECs annotated from the two subjects.

The primary objective of this experimental study is to evaluate the performance of the proposed probabilistic language analysis approach, which aims at translating online customer opinions into various ECs automatically. Several parameters are defined in the probabilistic language analysis approach: for example, the number of context words on both sides (N_L and N_R), the constant that tunes the weights for the smoothing item (μ), and the weights for normalization terms (C_1, C_2 and C_0). Thus, to analyze how the performance is influenced by these parameters, different types of experiments are conducted.

All programs were implemented and tested in Java 1.6 on a dual core 2.40GHz personal computer with 4GB memory.

6.2. Results and Discussions

- Experiment 1: performance comparison on the unigram model and the bigram model

Experimental comparison of the unigram model and the bigram model were conducted under the same parameter settings (μ equals to 500, C_1, C_2 and C_0 equal to 50, and N_L and N_R are equal to 25). The results are shown in Figure 3.

[Insert Figure 3]

Notice that, if there are not sufficient words on either side, the corresponding impact factors are set to zero. In Figure 3(a), the result is presented using 258 reviews of A810 as the training data, and in Figure 3(b), the result is based on 210 reviews of HP6500 as the training data.

This result reveal that generally the unigram model outperforms the bigram model. Specially, in Figure 3(b), the performance of the unigram model shows is better than that of the bigram model using different sets of product reviews as testing data. As shown in Figure 3(b), taken H6500 reviews as the training data and W610 reviews as the testing data, only 50% of results obtained by the bigram model are correct, while the unigram model manages to achieve a figure higher than 90%.

[Insert Figure 4.]

In Figure 4, the average accuracies of both unigram model and bigram model, under different numbers of context words are shown (μ equals to 500, C_1, C_2 and C_0 equal to 50). In these experiments, 256 reviews of A810 were utilized as the training data and reviews of other three products are utilized as the testing data. It confirms that the unigram model perform much better than the bigram model, which coincides with the results in Figure 3.

Conversely, a well-known experience is that a higher order N-gram model might perform better. However, notice that higher order models need much more training data and the success of higher order models largely depend on how well they are trained. Specifically, to guarantee a better result, sufficient training data need to be prepared to train these model. In this research, only words which are related with different ECs are utilized as training data. As shown in Table 3 shows, there are less than 40 words in each review dataset. The insufficient training might be one major reason that the unigram model performs better than the bigram model. Thus, the unigram model alone will be applied in the following experiments.

Moreover, as also shown in Figure 4, the performances of two models tends to be consistent when more than 25 context words are involved in both sides (N_L and N_R are equal to 25). A further analysis will be made towards the number of context words and its influence in Experiment 2.

In Figure 5, using A810 reviews as the training data, the impact of different context words on both sides is illustrated in the unigram model. In this figure, γ is denoted by the distance of "0". α_i is denoted by the distance of $-i$. For example, "–5", in the horizontal line, denotes the left fifth word of the keyword W_T. Similarly, β_j, is denoted by the distance of j.

[Insert Figure 5]

It shows the degree of impact on both sides is seen to not be symmetric and it declines significantly when the distance increases, say, more than 10 words on both sides. However, the impact does not change much when the distance is bigger than 15. The impacts of words, when the distance is bigger than 15, are observed not having much influence.

- Experiment 2: performance comparison on different numbers of context words

In Figure 6, using A810 and HP6500 reviews as the training data respectively, performance comparison is made with different numbers of context words. As noted from Figure 6, when only a small number of context words are involved, the accuracies declined relatively low. Accuracy starts to increase when more context word are used. It then gradually becomes stable when the number of context words on each side is bigger than 25. The results presented in this figure show similar patterns with the results in Figure 4. Therefore, about 25 context words are suggested to be chosen on each side, which indicates $N_L = N_R = 25$ is a stable threshold in terms of predicting the meaning of a keyword W_T.

[Insert Figure 6]

- Experiment 3: performance comparison on different μ values

In Figure 7, the performance is compared in terms of using different smoothing values μ in Equation (8).

[Insert Figure 7]

Firstly, by varying μ and using different numbers of context words (N_L and N_R), the average accuracies derived from three product reviews are shown in Figure 7 (a), still using A810 reviews as the training data and others as the testing data. As observed in Figure 7(a), the average accuracy does not change much under different μ values.

However, when reviews of HP6500 are chosen as the training data and other three sets of product reviews as the testing data, were also performed. However, different phenomena were found in Figure 7 (b). The average accuracy is observed to drop gradually when μ is set to be larger than about 800. Similar trends are found with different numbers of context words in this figure. Accuracies drop when a higher value of μ is specified. Therefore, a moderate μ is suggested. For example, in Experiment 1, μ was set to be 500.

- Experiment 4: performance comparison on weights of normalization terms

In Figure 8, performance comparison is mainly performed when different values of normalization terms, C_1, C_2 and C_0 were selected. In this experiment, C_1, C_2 and C_0 are set with the same value C, which means the weights from three normalization terms are supposed to be equal.

[Insert Figure 8.]

As stated previously in Section 5, the normalization terms are utilized to prevent α, β, and γ from growing too large. Typically, smaller weights of the normalization terms impose little impact, and it tends to give rise to an over-fitting problem. As seen from Figure 8, initially, the average performance does not vary much when the weights are relatively small. This is applicable on both sets of A810 or HP6500 as the training data. For example, when C is smaller than 60, the averaged accuracy is around 0.975 and 0.980, respectively. However, it begins to decrease when C gradually increase and it drops greatly on HP6500 reviews when C is greater than 100. Therefore, to remain effective in controlling the growth of α, β, and γ, C_1, C_2 and C_0 is set to 70.

6.3 An Illustrative Example

In this example, 258 Epson Artisan 810 reviews are utilized as the training data and other three sets of printer reviews are taken as evaluating data. As seen from Table 4, in the Epson Artisan 810 review dataset, 33 words are found to be translated into different ECs. To infer the most possible ECs of customer opinions, the context information of the 33 words in Epson Artisan 810 reviews is exploited in the training procedure.

More specifically, in the training step, review sentences of Epson Artisan 810 that contain the 33 words are firstly extracted and the proposed impact factor learning algorithm is applied. Notice that, in this example, only the unigram model is applied since, as reported in Section 6.2, the unigram model performs better than the bigram model. Hence, Equation (6) is employed to estimate the log value of the marginal probability $\log\frac{p(W_L \mid ec_p)}{p(W_L \mid ec_q)}$, $\log\frac{p(W_R \mid ec_p)}{p(W_R \mid ec_q)}$ as well as $\log\frac{p(W_T \mid ec_p)}{p(W_T \mid ec_q)}$ and the ratio function in Equation (16) is to define the loss function in Equation (18). Also, in the training step, all the parameters are set according to the experimental results in Section 6.2. For instance, 25 words are considered in both sides of the keyword W_T, μ is set to be 500 and three weights of normalization terms, C_1, C_2 and C_0, are set to 70. Then, the impact factors, α, β and γ, are inferred by solving the optimization problem in Equation (25).

In the evaluation step, review sentences of the other three sets of printer reviews that contain the 33 words are also extracted. For all of these evaluation sentences, the marginal probability of each considered context word can be inferred, which is also defined in Equation (6). Next, with the learned impact factors, α, β and γ, the ratio function in Equation (14) can be estimated for each pair of possible ECs. Accordingly, which EC receive the largest joint probability in Equation (3) can be deduced and it is regarded as the translated EC from customer opinions.

[Insert Figure 9.]

In Figure 9, taken Epson Artisan 810 reviews as training data, an example is shown about how this approach is utilized by designers to make use of EPSON Workforce 610 reviews for understanding needs and mapping online opinions into ECs in QFD. The following illustrative example is presented to demonstrate how the proposed probabilistic language approach can be utilized by designers in their daily work on how online customer opinions is translated into ECs in QFD for product design. In this example, starting from the review analysis by the probabilistic language approach, customer requirements in online opinions are translated into ECs automatically. It makes the relationship between voice of the customers and ECs be inferred. Next, voice of the customers and ECs are weighted and consolidated respectively to prepare to build the house of quality in QFD. With the inferred relationship, customer requirements in online opinions are integrated in QFD for product design.

[Insert Figure 10.]

In Figure 10, an example is illustrated to show how the proposed approach utilizes online opinions from multiple products for requirement analysis in QFD. In this example, three printers, “Epson WorkForce 610”, “HP Officejet 6500”, and “HP Photosmart Premium C309”, are under comparison. As seen from this figure, online reviews from different products are analyzed by the probabilistic language approach, which translate customer requirements into corresponding ECs. In order to make cross-comparison among multiple products, online customer requirements and ECs from different products are consolidated together. The inferred relationship between customer requirements and ECs is finally combined together. Comparing with the example in Figure 9, product comparison is made efficiently in terms of ECs by analyzing voices of online customers of multiple products.

7. CONCLUSION AND FUTURE WORK

QFD is one widely applied approach in engineering design to understand customer concerns. In QFD, one of the most critical task is to translate customer needs into ECs. Conventionally, customer needs are conventionally analyzed from customer survey or investigation. Yet valuable customer voices in online opinions are not fully exploited, partially, due to several technical challenges in the field of engineering design to handle such a big volume of textual data with rich sentimental information. In addition, these online opinions are generated from time to time in different websites, which impose further burden to analyze affluent online customer concerns. Hence, in this research, an automatic means is provided in accomplishing this crucial task on how to translate online customer opinions into ECs in QFD, which targets at helping designers to analyze online customer opinions efficiently.

To understand how online opinions are actually translated into ECs by designers in QFD, an exploratory study was conducted. From this study, interesting phenomena were found, which include that an one to one mapping from keywords in online opinions to ECs does not always exist. However, reading the entire set of online reviews and labeling which ECs are referred to is time-consuming and labor-intensive. It motivate us to explore an automatic review analysis approach regarding how to translate online customer opinions into ECs efficiently. Through the analysis about the interesting phenomena, a probabilistic language analysis approach was developed. Accordingly, an integrated impact learning algorithm was proposed, which facilitates product designers to translate customer needs expressed in online opinions into product design in QFD and determine which ECs are referred to effectively and efficiently.

One limitation of this research lies in that manually annotated data are relied on. However, only a limited number of manually annotated data are available. Hence, in the future, some promising research problems are to study whether a learning approach with less labeled data can be developed for this problem and whether the proposed approach can be applied to analyze a big volume of consumer data in other types, which will necessarily alleviate the burden of corpus building.

ACKNOWLEDGEMENT

This work was supported by the Fundamental Research Funds for the Central Universities (No. SKZZX2013091).

REFERENCE

Yoji Akao. *Quality Function Deployment: Integrating Customer Requirements into Product Design*. Productivity Press, Cambridge, MA, 1990.

Yizeng Chen, Richard Y. K. Fung, and Jiafu Tang. Rating technical attributes in fuzzy QFD by integrating fuzzy weighted average method and fuzzy expected value operator. *European Journal of Operational Research*, 174(3):1553--1566, 2006.

Wingyan Chung and Tzu-Liang (Bill) Tseng. Discovering business intelligence from online product reviews: A rule-induction framework. *Expert Systems with Applications*, 1(1):1-10, 2012.

Xue Dou, Justin A. Walden, Seoyeon Lee, Ji Young Lee. Does source matter? Examining source effects in online product reviews. Computers in Human Behavior. 28(5), 1555–1563, 2012

Richard Y. K. Fung, Yizeng Chen, and Jiafu Tang. Estimating the functional relationships for Quality Function Deployment under uncertainties. *Fuzzy Sets and Systems*, 157(1):98--120, 2006.

Abbie Griffin and John R. Hauser. *The voice of the customer. Marketing Science*, 12:1--27, 1993.

Anders Gustafsson and N. Gustafsson. Exceeding customer expectations. In *Proceedings of the 6th Symposium on Quality Function Deployment*, pages 52--57, 1994.

J. A. Harding, K. Popplewell, Richard Y. K. Fung, and A. R. Omar. An intelligent information framework for market driven product design. *Computers in Industry*, 44(1):49--63, 2001.

Kwang-Jae Kim, Herbert Moskowitz, Anoop Dhingra, and Gerald Evans. Fuzzy multicriteria models for Quality Function Deployment. *European Journal of Operational Research*, 121(3):504--518, 2000.

C. K. Kwong, Y. Chen, H. Bai, and D. S. K. Chan. A methodology of determining

aggregated importance of engineering characteristics in QFD. *Computers and Industrial Engineering*, 53(4):667--679, 2007.
Ying Liu, Jian Jin, Ping Ji, Jenny A. Harding, Richard Y. K. Fung: Identifying helpful online reviews: A product designer's perspective. *Computer-Aided Design* 45(2): 180-194, 2013
Ying Liu, Wen Feng Lu, and Han Tong Loh. Knowledge discovery and management for product design through text mining - a case study of online information integration for designers. In *ICED'07*, pages 329/1--12, 2007.
Thomas Lee. Needs-based analysis of online customer reviews. In *ICEC'07*, volume 258, pages 311--318. ACM, 2007.
Thomas Lee. Adaptive Text Extraction for New Product Development. In *IDETC'09*, volume 2, pages 769-778. ASME, 2009.
Xinxin Li, Lorin Hitt, Z. Zhang. Product Reviews and Competition in Markets for Repeat Purchase Products. Journal of Management Information Systems archive. 27(4), 9-42, 2011
Li-Feng Mu, Jia-Fu Tang, Yi-Zeng Chen, and C. K. Kwong. A fuzzy multi-objective model of QFD product planning integrating Kano model. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 16(6):793 813, 2008.
Zeynep Sener and E. Ertugrul Karsak. A combined fuzzy linear regression and fuzzy multiple objective programming approach for setting target levels in quality function deployment. *Expert Systems With Applications*, 38(4): 3015–3022, 2011.
Conrad Tucker and Harrison M. Kim. Predicting emerging product design trend by mining public available customer review data. In *ICED'11*, volume 6, pages 43-52
Z. Wang, P. K. Kannan, S. Azarm. Customer-driven optimal design for convergence products. *Journal of Mechanical Design*, 133, 101010.1-101010.13, 2011
Xiao-Tun Wang and Wei Xiong. An integrated linguistic-based group decision-making approach for quality function deployment. *Expert Systems With Applications*, 38(12): 14428–14438, 2011.
Kaiquan Xu, Stephen Shaoyi Liao, Jiexun Li, and Yuxia Song. Mining comparative opinions from customer reviews for competitive intelligence. *Decision Support Systems*, 50:743-754, 2011.
Jiaming Zhan, Han Tong Loh, and Ying Liu. Gather customer concerns from online product reviews - a text summarization approach. *Expert Systems With Applications*, 36(2 Part 1):2107-- 2115, 2009.

82 of 88 people found the following review helpful

Very Poor PDF Support, December 12, 2009

By **M. A. Fine** (Johnson, KS) - See all my reviews

REAL NAME

Amazon Verified Purchase (What's this?)

This review is from: Kindle DX Wireless Reading Device (9.7" Display, U.S. Wireless) (Electronics)

Having owned a kindle and kindle DX, reading is much easier and more "book-like" on the DX as there are fewer page turns (button pushes) and the larger screen more closely approximates the size of the typical paperback. I read two novels on the DX before returning it. I enjoyed the experience. However, the principal reason for my purchase was to use the kindle DX as my pdf medical library.

The pdf support works well for simple, short pdf documents. A 10 page flier displays very nicely and can be quickly flipped through. Anything greater than 100 pages becomes quite sluggish. Anything more than 500 pages is virtually unusable.

PDF hyperlinks do not work in the kindle DX. That means none of the bookmarks that appear in a PDF on your computer will be present. It also means that a pdf table of contents will not function on the kindle DX. There is no good way to navigate pdf files on the kindle DX. Technical manuals, textbooks, or reference materials lose their utility on this device.

The keyboard is awkward. Entering numbers (a very common task) requires the pressing of the alt key before every digit.

I consider this device an important step on the road to digital book technology, but it will quickly be eclipsed by products which offer support for more universal formats and devices offering greater functionality, especially regarding PDF documents. It is ultimately a novelty item and not worth the price paid. I will look to the new crop of e-readers debuting in Q1 2010 which will likely make this product quickly obsolete.

Help other customers find the most helpful reviews

Was this review helpful to you? Yes No

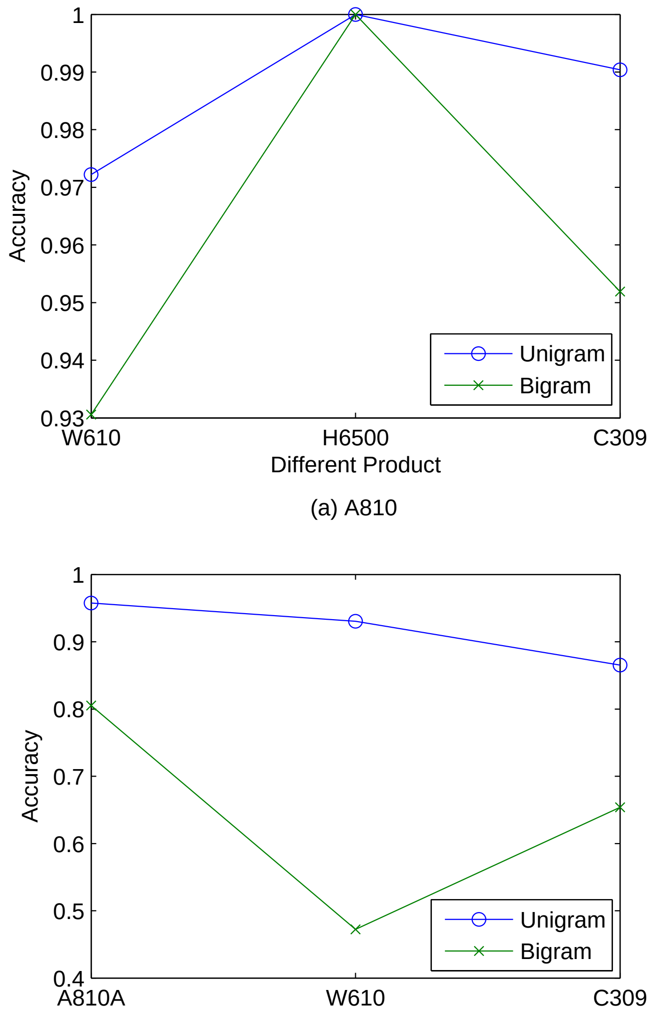
Report abuse | Permalink

Comment (1)

Figure 1. One typical customer review

Review sentences	keyword	Auto Document Feeder	Card Slot	Consumable Replacement	Hopper Unit	Ink Longevity	Noise	Print Quality	Supported Paper	Supplementary Software
I only have the printer for a few days, but so far I am very pleased.	-									
But the paper handling on my 810 has been flawless, if a tiny bit noisy while it pulls in the paper.	noisy						-2			
But the paper handling on my 810 has been flawless, if a tiny bit noisy while it pulls in the paper.	pulls				-1					
The actual printing is quiet, and of great quality.	quiet						1			
The actual printing is quiet, and of great quality.	great quality							2		
The paper tray feels a bit flimsy, but is easy to remove or insert, and there 's no fuss to loading your paper in it.	paper tray				-2					
It can expand to hold legal size paper, and has a separate area for smaller sized papers, usually for photo printing paper.	paper								0	
The package comes with 3 sheets of Epson high quality 4x6 glossy photo paper, so I printed three photos as a test and I could not be more pleased with the photo print quality.	print quality							2		
I have been using an HP Wireless printer up until now(model 5850)and the print quality of the Artisan 810 far outstrips that HP, with much deeper blacks and dark tones, and more rich looking color, which is probably due to superior ink quality more than the printer quality.	quality							2		
Another pre-purchase worry was that there were complaints about the Artisan 800 being a big ink-hog.	-									
Too early to tell, but I 've printed those 3 photos and a fourth on plain paper, and about 25 pages of text(some with graphics)and the ink levels have not budged yet.	ink					1				
So, so far so good on ink usage.	ink					0				
I was surprised and pleased too that the 810 printer comes with an TWO black ink cartridges, so you 'll have an extra one when it runs out.	ink			1						
Additionally, the inks provided are the same capacity as the refills, whereas some printers I have seen came with ink cartridges that have a lower capacity.	ink			2						
A nice feature is that when you plug in a camera memory card in the front, not only can you see and even crop the photos on the printer 's screen, but the photos on the card can also be remotely viewed by your computer on the same wifi!	card		2							

Figure 2. Review analysis



(a) A810

(b) H6500

Figure 3. Performance of the unigram model and the bigram model

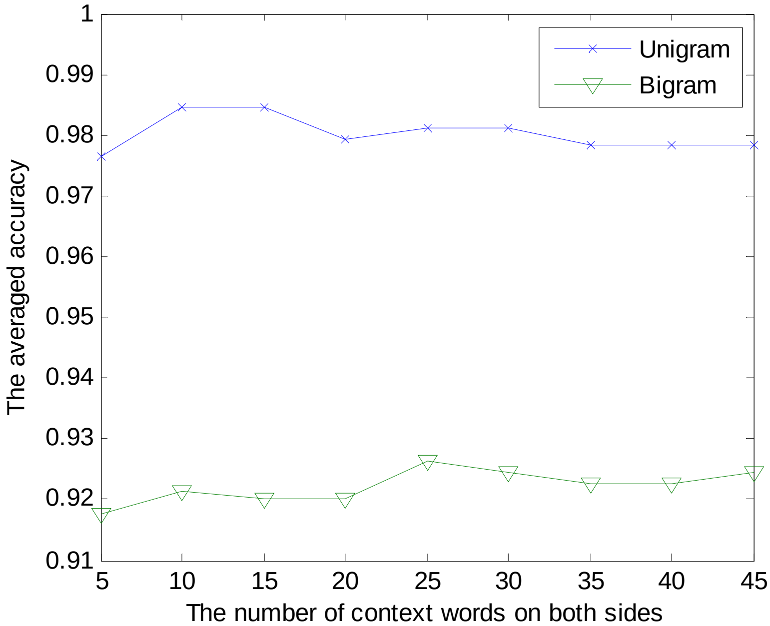


Figure 4. The average accuracy

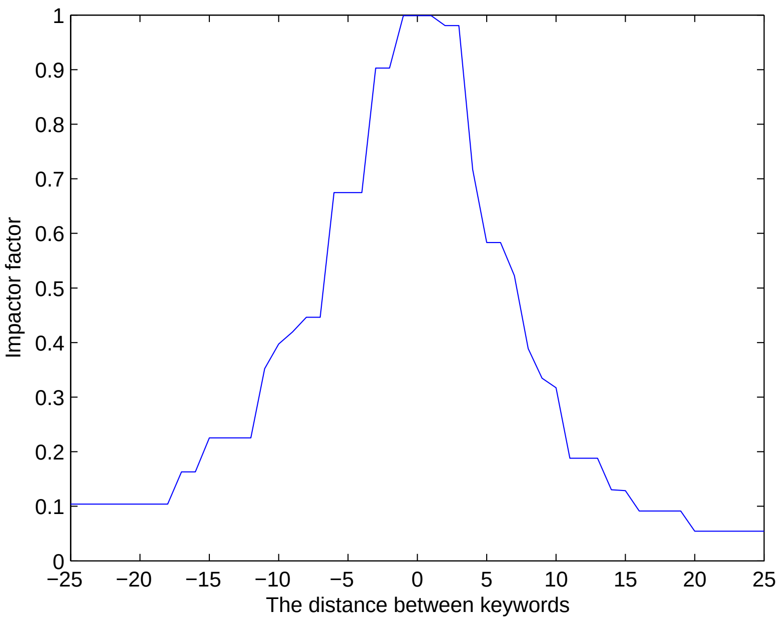
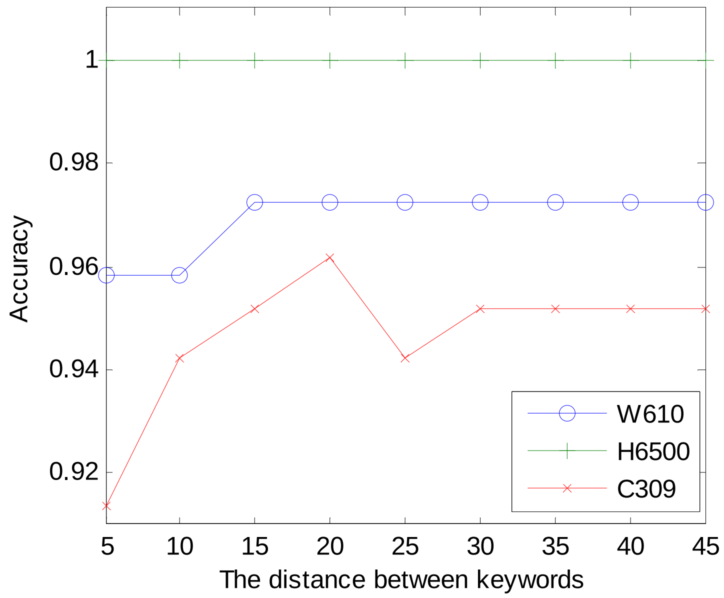
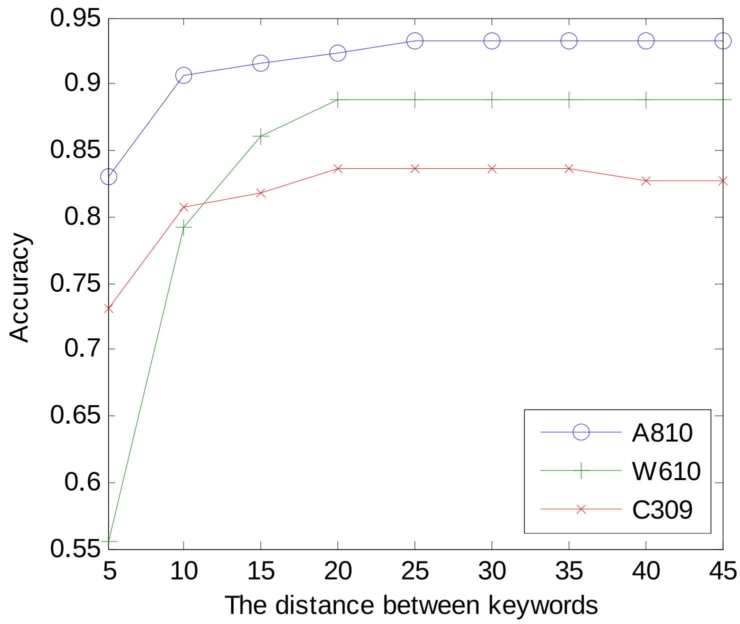


Figure 5. The degree of impact and the distance between keywords

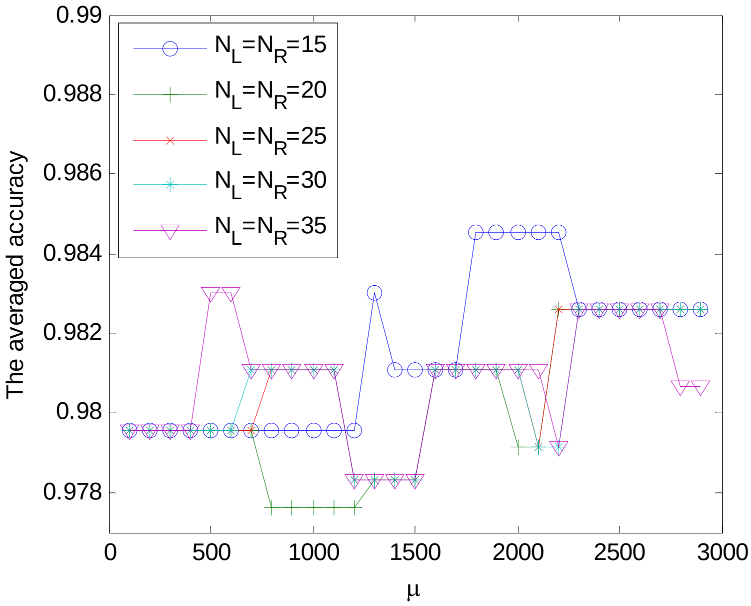


(a) A810

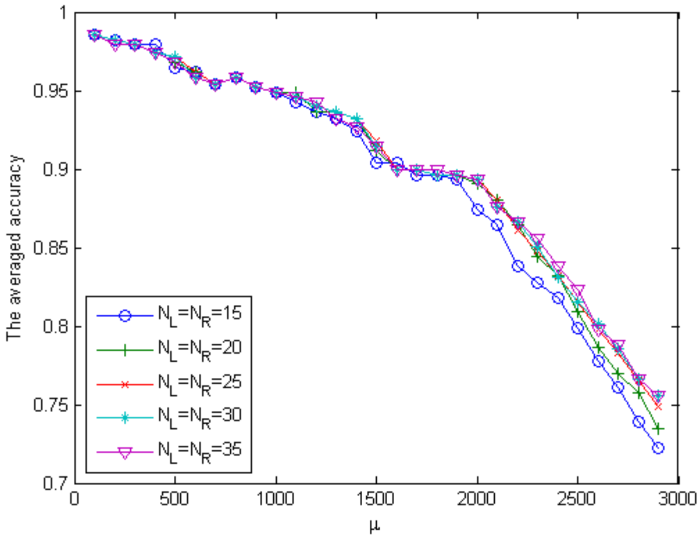


(b) HP6500

Figure 6. The performance comparison focusing on different numbers of context words



(a) 810A



(b) H6500

Figure 7. The performance comparison on different μ

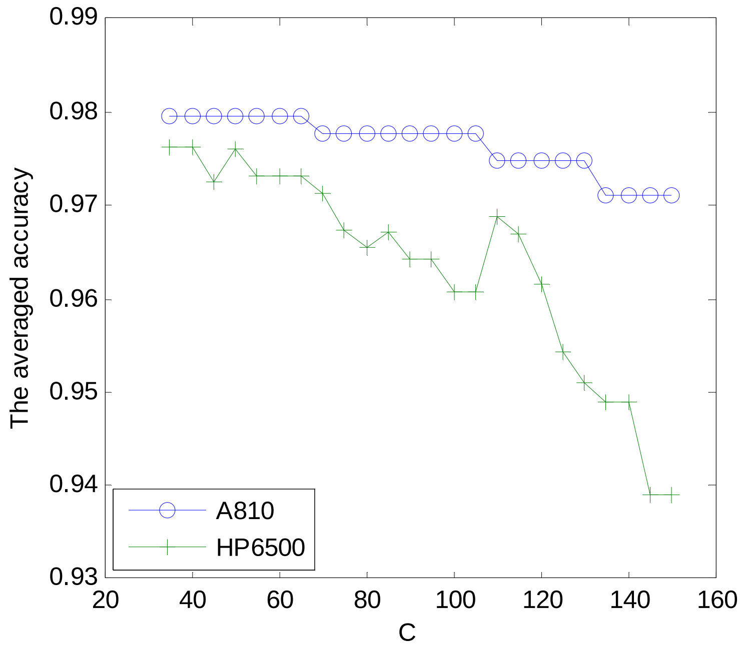


Figure 8. The performance comparison on different *C* values

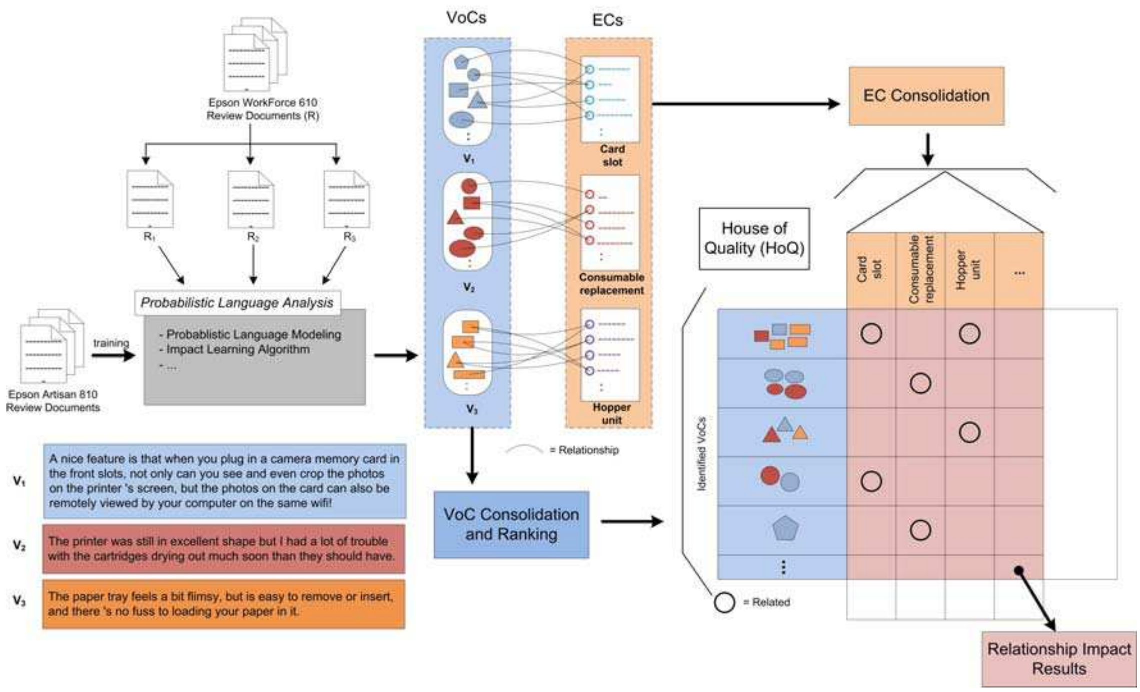


Figure 9. An example to show how the proposed approach is utilized by making use of reviews of one product

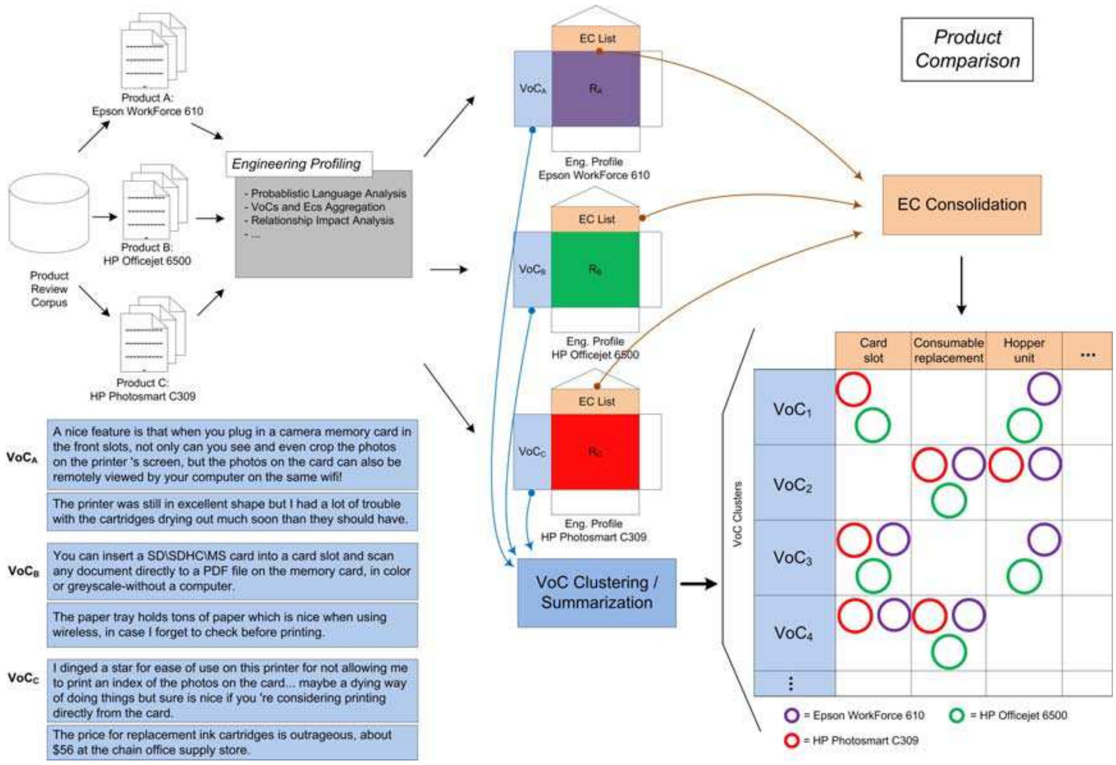


Figure 10. An example to show how the proposed approach is utilized by making use of reviews of multiple products

Table 1. Number of reviews

Printer	A810	W610	H6500	C309	Total
Number of reviews	258	169	210	133	770

Table 2. Statistic information of reviews

	A810	W610	H6500	C309	Avg.
Avg. num of words	237.496	162.637	194.933	235.827	207.723
Avg. num of sentences	12.554	9.631	11.324	12.729	11.560
	A810	W610	H6500	C309	Max
Maximal num of words	2823	1474	1075	1683	2823
Maximal num of sentences	85	78	56	75	85

Table 3. Engineering characteristics

Printer Housing	Power Supply	Fax Setting	Brand
Wifi Integration	Ease of Setup	Ease of Use	Noise
Duplex Printing	Print Quality	Print Head	Package
Software Updated	Scan Software	LCD Panel	Outlooks
Auto Document Feeder	Printing Speed	Hopper Unit	Card Slot
Supplementary Software	Mac Compatible	Ink Longevity	Durability
Consumable Replacement	Supported Paper		

Table 4. Number of words translated into different engineering characteristics

Printer	A810	W610	H6500	C309
Number of words	33	29	24	36