

Article

Customer Complaints Analysis Using Text Mining and Outcome-Driven Innovation Method for Market-Oriented Product Development

Junegak Joung ¹, Kiwook Jung ¹, Sanghyun Ko ² and Kwangsoo Kim ^{1,*}

¹ Department of Industrial and Management Engineering, Pohang University of Science and Technology, 77 Cheongam-ro, Nam-gu, Pohang 790-784, Korea; june30@postech.ac.kr (J.J.); kiwook@postech.ac.kr (K.J.)

² Korea Institute of Defense Analyses, 37 Hoegi-ro, Dongdaemun-gu, Seoul 02455, Korea; francescokoh@gmail.com

* Correspondence: kskim@postech.ac.kr; Tel.: +82-54-279-8234

Received: 1 December 2018; Accepted: 19 December 2018; Published: 21 December 2018



Abstract: The rapid increase in the quantity of customer data has promoted the necessity to analyse these data. Recent progress in text mining has enabled analysis of unstructured text data such as customer suggestions, customer complaints and customer feedback. Much research has been attempted to use insights gained from text mining to identify customer needs to guide development of market-oriented products. However, the previous research has a drawback that identifies limited customer needs based on product features. To overcome the limitation, this paper presents application of text mining analysis of customer complaints to identify customers' true needs by using the Outcome-Driven Innovation (ODI) method. This paper provides a method to analyse customer complaints by using the concept of job. The ODI-based analysis contributes to identification of customer latent needs during the pre-execution and post-execution steps of product use by customers that previous methods cannot discover. To explain how the proposed method can identify customer requirements, we present a case study of stand-type air conditioners. The analysis identified two needs that experts had not identified but regarded as important. This research helps to identify requirements of all the points at which customers want to obtain help from the product.

Keywords: customer needs; job concept; clustering analysis; latent needs; text analysis

1. Introduction

Considering customer needs during the product development process is a useful method to minimize the risk of developing inappropriate products [1]. Customer needs have helped to direct research and development (R&D) to launch new products and services [2]. Customers' suggestions and complaints can generate ideas to determine product concepts. Moreover, considering customer requirements during new product development (NPD) can increase the number of novel ideas and thereby improve the quality of innovation [3]. Empirical study has demonstrated that identifying customer needs affects development activities and new products [4]. Therefore, identifying customer requirements offers a starting point for effective and efficient planning of companies' overall development activities such as R&D and launch of new products or services.

Many studies [5–14] have analysed unstructured text data such as customer suggestions, customer complaints and customer feedback to identify customer needs for the product development. However, the previous research used keyword extraction, regression analysis, clustering method, association rule mining, classifier, latent Dirichlet allocation and the Kano model; they only identified a limited set of customer needs based on features of a product. Customer requirements that are identified

in this way are only a few of the requirements [15,16]. Previous studies focused on only customer needs that are directly associated with an execution step of a product, whereas several steps such as order, installation, monitoring and arrangement can be involved when customers use a product [17]. Customer needs in the pre-usage and the post-usage steps of a product should be considered.

To solve the problem, we use the Outcome-Driven Innovation (ODI) method, which identifies real customer needs in various steps of product use by customers; the purpose is to develop innovative products or services [18]. The method considers a job map, which identifies latent customer needs based on the concept of a job; and a format of customer requirement, which captures unambiguous customer needs. Many firms (e.g., Bosch, Microsoft, Kroll Ontrack, Hussmann and Abbott Medical Optics) have used the ODI method to innovate products and services.

However, the ODI method to identify customer needs lacks objectivity because advance information is totally dependent on managers who know about customers [19]. The ODI method needs supporting tools to derive customer requirements, not from subjective opinions but from real customer feedback. In these circumstances, text mining can help managers to analyse customer feedback quickly [20]. Consequently, to help managers to obtain objective evidence for customer requirements, tools to mine textual data are essential.

This paper suggests a method to identify customer requirements from customer complaints by applying text mining to the ODI method. This research fills a gap in text mining literature for product development by providing a method to explore customer needs in various steps of product use by customers. This study also supports the ODI method by providing a method to analyse large amounts of customer complaints.

The proposed method first collects customer complaints from customer service centres, then feature candidates are extracted. The extracted feature candidates are selected in terms of the concept of job and clustering. Features are used to construct a similarity matrix, which is used in clustering analysis to identify customer requirements. Ultimately, customers' true requirements are identified and they guide subsequent product activities including product concept, technology development and project prioritization. A case study of stand-type air conditioners is given to explain how the proposed approach is used.

The remainder of this paper is organized as follows. Section 2 reviews the theoretical background of the proposed method. Section 3 explains the proposed method to capture customer requirements and Section 4 presents a case study of stand-type air conditioners. Section 5 presents conclusions and future work.

2. Theoretical Background

2.1. Text Mining for Market-Oriented Product Development

Development of smartphones and proliferation of electronic word-of-mouth channels have expanded opportunities for customers to express their opinions [21–23]. Web harvesting techniques enable companies to collect or save customer reviews or complaints on web pages by automatically scanning many of them [24,25]. Recent progress in text mining has enabled analysis of the unstructured text data of customers [26–28]. Therefore, companies can analyse customer needs from a large amount of the textual data in a short time by using text mining and this analysis can reduce the time to develop products to respond to customer needs [20].

Much research for product development has been performed to identify customer needs from customer reviews or complaints by using text mining. Customer concerns were identified by extracting salient topics, then ranking them based on automatic summarization of customer reviews of a target product. The identified customer concerns provided preferred features of the product and the reasons for these preferences [5]. Decker and Trusov [6] used text-mining techniques to identify the pros and cons of product attributes, then used negative binomial regression to estimate the influence of each attribute. Park and Lee [7] constructed a 'voice of customer' vector by extracting

product features from customer complaints about a mobile phone, then used clustering analysis to segment customers according to product features that they complained about. Then the authors used co-word analysis to inspect each customer group by identifying useful keywords. The result was used to devise specifications of the new product. Aguwa et al. [8] used text mining and association rule mining to identify customer needs from qualitative and quantitative customer data by transforming customer input to engineering input. A probabilistic Naïve Bayes-based classifier has been proposed to automatically identify customer requirements by developing preferred features of a product's feature [9]. Aguwa et al. [10] used text mining to extract significant attributes of a product and to develop a real-time system to monitor customer feedback by learning association rules; the system includes a unique model that uses fuzzy logic to identify negative and positive feedback. Liang et al. [11] used the topic model of latent Dirichlet allocation (LDA) to identify product features that customers frequently mentioned, then identified product problems by exploring the relationship of product features using association rule mining. Qiao et al. [12] developed a new LDA model to identify critical product defects by defining keywords related to specific product. Wang et al. [13] used sentiment analysis with regression model to identify the impact of features of washing machine such as type, colour, display and energy-efficiency. Min et al. [14] developed a review-based Kano model that identifies customer needs by evaluating customer satisfaction with product attributes.

Prior studies [5–14] identified customer needs on the basis of product features by analysing unstructured text data provided by customers but this approach is insufficient to represent customers' true requirements. When a customer uses a product, several steps are involved, such as ordering, installing, using, monitoring and arranging [29]. Existing research has tended to overlook latent customer needs in the pre-usage and the post-usage steps by focusing on a step of usage. However, customer requirements are not derived in the various product usage steps, so the combination of preferred product features in a usage step cannot be considered as a final product outcome that customers genuinely want. Therefore, existing studies result in imitative products, because customers tend to seek unusual product features that existing competitors already offer [30].

To identify genuine customer needs, an alternative model must be developed to analyse unstructured text data of customers from new perspective. This paper suggests a method to identify customers' true needs from customer complaints by using the concept of job in the ODI method.

2.2. Outcome-Driven Innovation Method

ODI identifies customer requirements by applying the concept of job, which is defined as the goal that the customers want to achieve, or problems that they try to solve in a given situation [18,30]. Christensen, et al. [17] argued that customer needs identified by considering only existing characteristics of the product were poor indicators of customer behaviour and that companies must identify customer needs by assessing how well customers get a job done while using products or services. Identifying customer needs based on the features of a product is a point-in-time solution that can change over time but job-based customer needs have a long-term focus that is stable over time. The concept of job also has no racial and religious features. Therefore, by using the concept of job, companies can easily comprehend what jobs customers are trying to complete, regardless of time, race, religion, or region. Furthermore, definite measurements such as speed and predictability can be used to evaluate how well customer requirements are actualized based on a specific job.

Initially, the job-based approach observed customers' use of products or services in detail [31]. Customers' reason for using products or services is to complete a job. Customers conduct several actions to achieve the job [29]. For example, "to lower indoor temperature," customers define goals on how much to reduce the temperature, investigate items that can perform the job, then install or use them. After executing these steps, customers monitor whether the product operates well and arrange the product. The concept of job helps companies to identify customer needs in various steps of product usage, so it is distinct from the existing research that concentrates on customer needs during one step of product execution [29,32–34].

The ODI method elicits customer needs by conducting in-depth interviews. To conduct such interviews effectively, ODI experts require advance information of the needs, which is produced by experienced managers who have interacted with many customers [19]. However, the experience of managers is subjective. Therefore, the ODI method needs supporting tools to derive customer requirements from empirical evidence by analysing large volumes of unstructured text data. Our study proposes a method that uses text mining to analyse unstructured text data from customers to support the ODI method.

3. Proposed Method

3.1. Overall Research Framework

The overall process of customer complaints analysis for market-oriented product development (Figure 1) collects customer complaints regarding target product, then extracts feature candidates from unstructured contents of customer complaints. To apply the concept of job to analysis of the complaints, these candidates are first identified based on the diverse jobs of the target product. Then secondary candidates that have a great effect on clustering are selected. Collected customer complaints are grouped using a clustering method on the basis of significant features that are associated with various job steps. Analysis of the clustering result can identify customer requirements with a structure of customer needs.

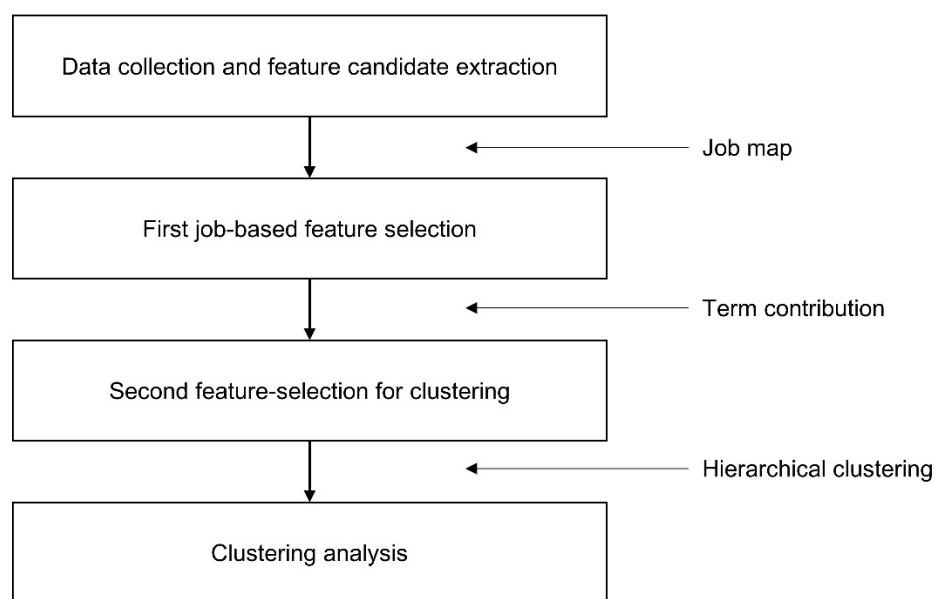


Figure 1. Overall research process.

3.2. Data Collection and Feature Candidate Extraction

Customer complaints of target product are collected by the company's internal channels such as web sites or mobile applications of service centres, or external channels such as third-party review sites (i.e., Epinions, Amazon customer reviews). The collected data should contain enough number of complaints in various steps of product usage of customers. The criteria of suitable data for analysis can be roughly identified by searching keywords related to the various steps of product usage.

From the collected complaints, we extract single keywords, multiple keywords and action-object (AO) or subject-action (SA) combinations, which all become candidates for the feature that represents the free-form text of customer complaints. To extract feature candidates, a language parser such as Stanford parser or Korean NLP Package is first used to automatically tag parts of speech such as noun, verb, adjective, or adverb. Then standard pre-processing is conducted [35]: steps include transforming all text to lowercase; tokenizing customer complaints; eliminating meaningless stop words (e.g., 'she,'

‘the,’ ‘but,’ ‘what’); lemmatizing words (e.g., to convert ‘foreign substances’ and ‘generated’ to the root forms ‘foreign substance’ and ‘generate’); removing words that occur either too frequently or very rarely. A single keyword is identified by nouns and multiple keywords are identified by n-gram modelling, which models sequences of natural language by using statistical properties. An AO or a SA is identified using a matrix that represents the co-occurrence of keywords and verbs.

3.3. First Job-Based Feature Selection

We select first job-based features from extracted feature candidates (e.g., single keyword, multiple keyword, AO, SA) by considering high relativeness to a job map. The job map represents the process of product usage of customers separated from technical solutions and is composed of universal process steps [29]: defining what the job needs; gathering and locating required inputs; organizing and preparing the components in accordance with circumstance; confirming that the task is ready to be performed; executing the job; monitoring the result of execution; giving modifications; ending the job. For example, the job map of “lowering indoor temperature” is expressed by the form of AOs that present specific jobs under the universal job steps [33] (Figure 2); an object is expressed by a single keyword or multiple keywords. Therefore, the high relativeness is assessed by identifying that each feature candidate semantically corresponds to AOs of the job map. To conduct semantic processing, we use WordNet, which is a large hierarchical generic database of English words; or a standard Korean dictionary, which is the Korean version of WordNet. This job-based feature selection helps to analyse customer complaints in various jobs from beginning to end.

Then the first job-based feature is determined by considering the dependency of innate features such as synonyms regardless of clustering algorithms, because feature dependency degrades clustering accuracy [36]. If the object in an AO overlaps single keywords or multiple keywords, the first features will remove duplicated keywords and retain AOs that represent customer complaints in detail.

After selecting the first job-based features, customer complaints that do not contain the first feature set are considered as the first outlier set, which is analysed. Many of these first outliers are customer complaints that are not related to various job steps but were collected as a result of the inaccuracy of a parser.

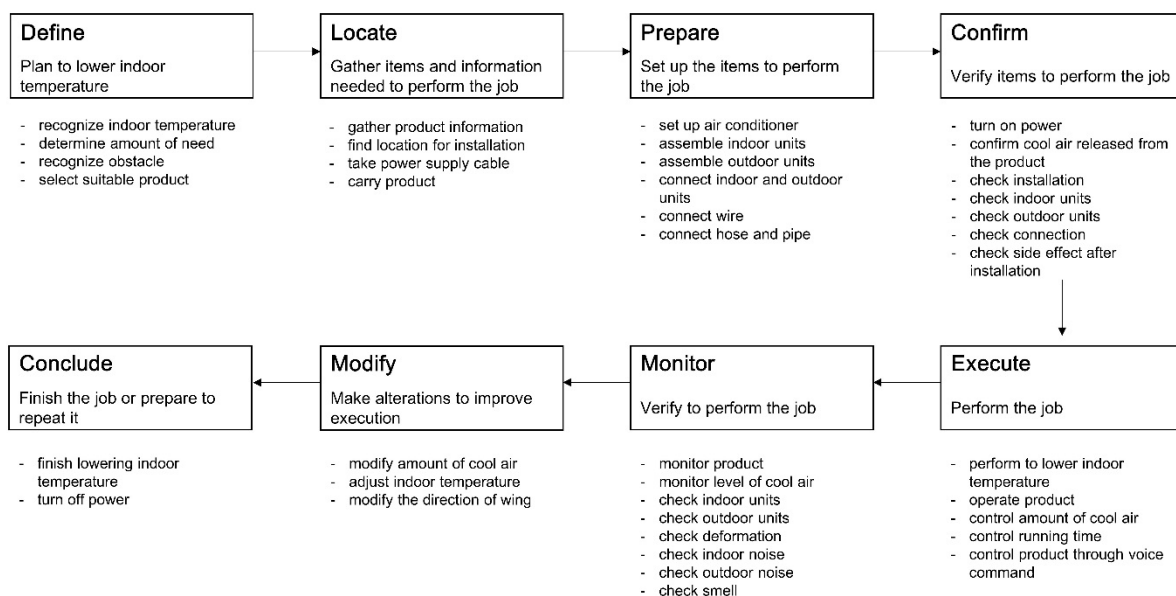


Figure 2. Example of the job map of “lowering indoor temperature.”

3.4. Second Feature-Selection for Clustering

After the first feature set related to the job process is finalized, the second feature set is selected to increase the accuracy of clustering among the first feature set by applying a dimensionality-reduction

technique to the text data. These techniques can be divided into feature extraction and feature selection [37,38]. Feature extraction generates a set of new features with reduced dimensionality from the original features; the goal is to identify the most important influences. Examples of the technique include principle component analysis (PCA) [39] and word clustering [40]. However, the new features created by feature extraction techniques may not have a clear physical meaning, so the clustering results may be difficult to interpret. In contrast, feature selection chooses a small subset of the original feature set by considering how the subset affects the clustering result; examples include document frequency (DF) [41], entropy-based ranking (En) [42] and term contribution (TC) [37]. Compared to feature extraction, these techniques provide better readability and interpretability of the clustering results, because their physical meanings are not lost.

Therefore, this paper uses TC to select second features from the first features (Figure 3). TC is suited to interpret the clustering results in text data and does not have high computational complexity. First, by calculating term frequency-inverse document frequency (TF-IDF) value between customer complaints and first features, a matrix is represented. Then the second features are selected using TC and a matrix that represents TF-IDF values between customer complaints and second features is constructed. TC is calculated based on the similarity of customer complaints and therefore gives a high value to features that have a major influence on similarities among many customer complaints. The similarity of customer complaints is calculated by cosine similarity:

$$\text{sim}(c_i, c_j) = \sum_f f(f, c_i) \times f(f, c_j) \quad (1)$$

where c_i, c_j are two customer complaints and $f(f, c_i), f(f, c_j)$ represents the TF-IDF [43] weight of feature f in customer complaint c .

TC is calculated as [37]

$$TC(f) = \sum_{i,j \cap i \neq j} f(f, c_i) \times f(f, c_j) \quad (2)$$

The second feature set is determined by TC rankings.

After the second feature selection for clustering, customer complaints that do not include the second features are regarded as second outliers; these are trivial complaints that are not as important as the main customer needs based on the job.

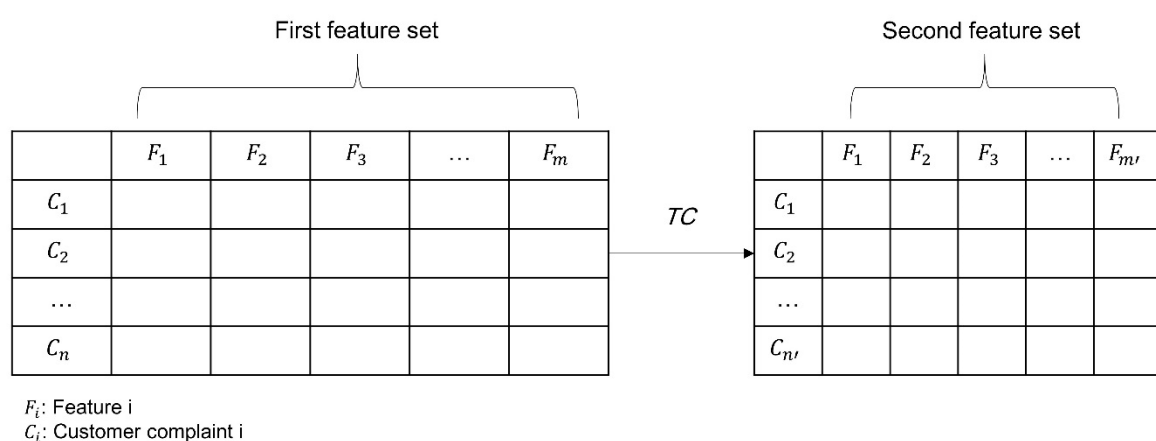


Figure 3. Process to select the second feature set from the first feature set.

3.5. Clustering Analysis

Semantic similarities of customer complaints are calculated using cosine similarity based on the second feature vector and its synonyms in diverse contexts [44], then the clustering algorithm is performed on the similarity matrix. While a clustering result is statistically significant in representing the original data, various clustering methods such as spectral clustering, k -means clustering and

hierarchical clustering can be applied according to the properties of the data. Spectral clustering is effective if a clustering result is expected to have a small number of clusters and k -means clustering is appropriate if experts know the number k of clusters. Hierarchical clustering is suitable if the clusters have different sizes. This paper uses hierarchical clustering because clusters in case data are expected to have different sizes.

After clustering has been performed, the number of clusters is determined by considering whether or not the whole cluster includes customer complaints about a diverse job process and whether or not each cluster includes related customer complaints about a single job. In each cluster, the criteria can be easily identified by looking for representative features, which are discovered in customer complaints over a certain level.

Lastly, ODI experts analyse the clustering results (Figure 4) to elicit customer requirements based on the structure of customer needs [45]. When they analyse the clustering results, ODI experts follow rules for structuring customers' statements of need (Table 1). Well-formatted requirements include the type of improvement (minimize, increase) and a unit of measure (time, likelihood, number, frequency, amount, risk). ODI experts distinguish between requirements and solutions, clarify vague statements and eliminate duplicates. The structure of customer needs in ODI helps to capture concrete requirements of customers.

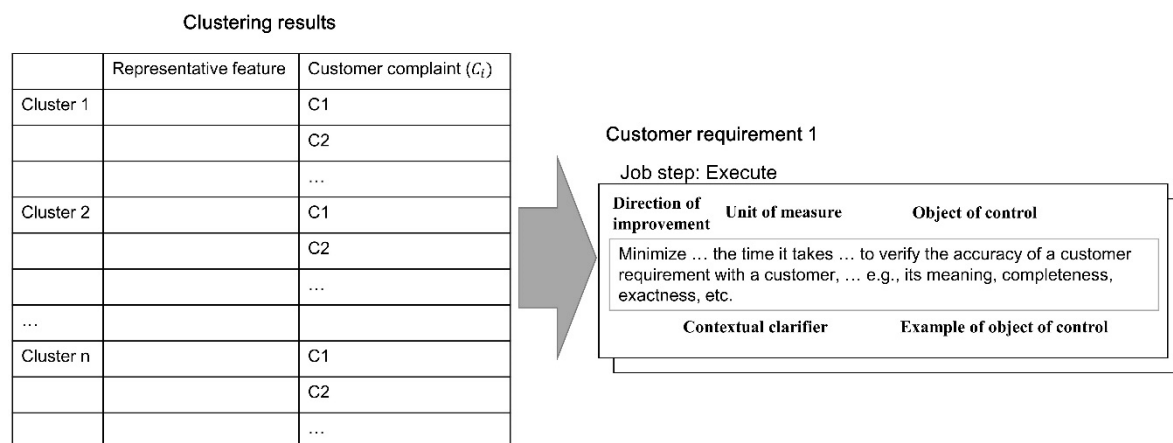


Figure 4. A process to derive customer requirements.

Table 1. Rules for structuring customer requirements in outcome-driven innovation (ODI) method (from [45]).

Rules For Structuring Customer Requirements	
1.	Needs statements must be free from solutions and specifications—and stable over time.
2.	Needs statements must not include words that will cause ambiguity or confusion, for example, certain adjectives and adverbs, pronouns, process words, jargon, acronyms.
3.	Needs statements must be specific without sacrificing brevity.
4.	Needs statements must follow the rules of proper grammar.
5.	Do not use different terms to describe the same item or activity from statement to statement; be consistent in language.
6.	Needs statement must have a consistent structure, content and format.
7.	Needs statements must relate to the primary job of interest and not to ancillary jobs.
8.	Needs statements must be introduced with only one of two words: minimize (90%) or increase (10%).
9.	Needs statements must contain a metric (time, likelihood, number) so performance can be measured.
10.	Examples added to the end of a statement for purposes of clarification must be similarly and consistently formatted.
11.	Needs statements must be usable in all downstream activities, for example, questionnaires, for deployment.

4. Empirical Analysis: The Case of Stand-Type Air Conditioners

4.1. Collecting Data and Extracting Feature Candidates

The proposed method was used to analyse customer complaints in 2013 recorded from web sites or mobile applications of customer service centres in a large Korean company with annual sales of 56 billion South Korean won (~ US\$ 47 million) that manufactures electronic appliances. In that interval, the electronics company wanted to identify customers' true needs for innovative product development by adopting a new text mining approach to analyse customer complaints. The company provided complaints about several appliances. A case study of stand-type air conditioners was selected for two reasons. First, this appliance's database system to record customer complaint data was well-established, so it is suitable for extraction of metadata by using queries. Second, the air conditioner market is becoming increasingly competitive as an increasing number of new products is introduced to satisfy unmet needs, so this company is confronted with a situation in which it should develop new and competitive products.

To occupy the market in these circumstances, innovative technology and products must be developed to attract customers. We collected 2362 customer complaints in the period 2013/01–2013/12 in Korea from customer database systems by using a search query. Most customers expressed complaints in voice calls, so the volume of textual complaints was not too high. However, these complaints were expected to discover customer needs in the various steps of product usage because they contained complaints about product preparation, installation, execution, or monitoring steps.

Parts of speech such as noun, verb, adjective, or adverb were tagged by the Korean NLP package, the hannanum parser, which is a POS tagger that has been developed by the Semantic Web Research Centre at KAIST since 1999. After text pre-processing, rare words that occurred 10 times or fewer, or very common words that occur 500 times or more were excluded due to low discriminatory power [46]. 312 single keywords were easily extracted from the POS-tagged nouns and 73 multiple keywords were extracted using the frequencies of bigrams and trigrams; 62 AOs or SAs were extracted by developing a matrix that notes co-occurrence more than 100 times between keywords and verbs.

4.2. Selecting First Job-Based Feature

In this step, the first job-based feature set was selected by a high relativeness of the job map, then the first feature set was determined by considering the dependency of features. We identified 63 features from 447 feature candidates by semantically comparing AOs of the job map of "lowering indoor temperature" by consulting the standard Korean dictionary. These 63 features included 48 keywords such as 'installation (설치)', 'operation (작동)', 'power (전원)', 'noise (소음)' and 'voice (음성)' and 15 of AOs or SAs such as 'cool-air weak (냉기가 약하다)'. A total first feature set of 54 was chosen by removing features that had synonym relationships.

After the first job-based feature was selected, 83 customer feedbacks that the first feature set did not represent became the first outliers among the 2362 customer complaints. The first outliers included complaints such as dry function and its smartphone control that are not associated with the core jobs of an air conditioner. Some of the first outliers were caused by the hannanum parser, which sometimes mistook nouns for verbs.

4.3. Selecting Second Feature for Clustering

The second feature for clustering was selected from the 54 in the first feature set by using TC ranking. After constructing the first feature vector that represents customer complaints based on TF-IDF, each TC value in the first feature set was calculated (Table A1, Appendix A). Some features that have a major effect on the similarity of customer complaints and the clustering result had higher TC values than others. The second feature set was determined by ranking TC values from the first feature set. Although the difference of TC value was large in the top rank, the second feature for clustering contained many crucial features, which will identify the requirements of customers within various

features that represent job steps. In the end, the 54 features in the first feature set were reduced to 33 in the second feature set and the second feature vector to indicate customer complaints was built for clustering based on TF-IDF.

After selecting the second feature for clustering, 12 customer feedbacks that the second feature set did not represent were considered as second outliers. Analysis determined that these were composed of unimportant customer questions such as requirements for notification of gas leak and function modification alarms in the ‘monitor’ job.

4.4. Analysing Clusters

After the first and second outliers had been eliminated, hierarchical clustering was performed by constructing a 2267×2267 similarity matrix of customer complaints; it was composed of cosine similarity calculated using second feature vectors. Various linkage methods such as single linkage, complete linkage, average linkage, weighted linkage and centroid linkage can be used for clustering. A linkage method was selected that uses cophenetic correlation, which indicates a high correlation of original data. The average linkage method had the highest cophenetic correlation of 0.86, which is high enough to demonstrate the reliability of the clustering process. This value means that clustering result explains 86% of the variation in the data and was therefore used for hierarchical clustering.

We determined the number of clusters by identifying representative features that present diverse job steps and analogous customer complaints about a single job. The feature was considered ‘representative’ if it represented 70 percent of customer complaints in each cluster. Therefore, 14 clusters included the diverse job steps and analogous complaints in the single job; they were analysed to elicit customer requirements.

The ODI experts then identified analogous customer complaints in each cluster and elicited 22 customer requirements by using the structure of customer needs in the ODI (Table A2, Appendix B). The most commonly-mentioned job step was an ‘Execute’ job step with 33.4% of the total. In cluster 1, 2 and 3, we identified that most customers complained when the air conditioner’s ability to emit cool air did not meet customers’ expectation. The second most-frequently reported job step was a ‘Monitor’ job step (32.3%). In cluster 9, 12 and 13, we identified most customers complained about the noisiness of the air conditioner. The third most commonly-mentioned job step was a ‘Modify’ job step (11.6%); ‘Prepare’ and ‘Confirm’ job steps (4%) were the final job steps identified. In cluster 5, we identified that some customers complained about side effects and monitoring after installation.

Product managers who had worked in customer service centre for ten years identified 22 customer needs and agreed on primary customer needs. The managers also regarded customer needs in ‘Prepare’ and ‘Confirm’ job steps as valuable unexpressed needs that the existing method cannot discover.

Previous studies that analyse customer reviews or complaints identified partial needs of customers because the studies focused on visible problems based on product attributes in a step during which customers use the product [29]. To identify the full set of problems, we analysed customer complaints by using the concept of job in the ODI method. This concept helps to identify customer needs in various steps, including preparation, installation, execution and monitoring of the product. This job-based analysis can identify customer latent needs in the pre-execution steps and the post-execution steps; prior research cannot discover these needs. Job-based analysis can present these needs precisely by using a format of customer requirement in the ODI method [45]. As a result, this research helps to identify requirements of all the points at which customers want to obtain help from the product.

5. Conclusions and Future Study

This research suggests an ODI-based method that uses text mining to identify customer’s true needs from customer complaints. This study identified not only customer needs in the execution step of product but also latent needs of customers in the pre-execution and post-execution steps of product. The discovery of latent needs is distinct from the existing research. The proposed method also provides ODI experts with supporting tools to analyse a large number of customer complaints.

These tools provide a clustering that can present complaints about various jobs, so experts can derive customer needs from the clustering result. These needs derived from customer data can effectively provide advance information that can help experts to conduct in-depth interviews; the data can also reduce the need to use experienced managers in the interview process. Companies can gain new insights by combining the knowledge from analysis of these complaints with the previous knowledge of managers [47].

Customer complaints analysis in the case of the Korean air conditioners offers managerial insights to practitioners. Decision makers can obtain directions of innovative product development from the customer perspective by allocating resources related to customers' true requirements. Needs analysts can use the proposed method as a tool to perform analysis of a large volume of unstructured contents of complaints based on the job. Developers can identify the specification of customer viewpoint that the final product should satisfy.

However, this study has some limitations, which provide directions for further research. First, this paper analysed complaints from end users to identify customer needs; the analysis focus on five job stages (e.g., prepare, confirm, execute, monitor, modify) even though the job map consists of eight stages. Therefore, future analysis to capture all customer requirements that include purchase motivation of the product and delivery will be represented.

Second, a customer database system to record customer complaints in diverse contexts through various channels should be constructed in advance, so that the suggested framework can be applied. In the case study, the proposed method was applied to enough complaints listed in a constructed a customer database but analysis of such a small amount of biased customer complaints may not clearly identify the latent needs, so the reliability of the result is not guaranteed. In the end, the proposed framework combines automatic method and expert judgment but it is not fully automated. Therefore, the output of the method is static analysis and requires effort from a human expert. This method is expected to be improved to be fully automated for dynamic real-time analysis.

Author Contributions: Conceptualization, J.J. and K.K.; Data curation, K.J. and S.K.; Investigation, K.J.; Methods, J.J.; Supervision, K.K.; Writing—original draft, J.J.; Writing—review & editing, J.J.

Acknowledgments: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry Science, ICT and Future Planning (MSIP) (NRF-2016R1A2B4008381).

Conflicts of Interest: The authors have no conflict of interest.

Appendix A TC Ranking of First Feature Set and the Difference of TC Values

Table A1. TC ranking of first feature set and the difference of TC values.

First Feature	TC Value	Rank	Difference
F1	13,503.58	1	374.76
F2	13,128.82	2	2810.48
F3	10,318.34	3	299.43
F4	10,018.91	4	2550.85
F5	7468.06	5	768.88
F6	6699.18	6	361.87
F7	6337.31	7	3223.71
F8	3113.6	8	716.27
F9	2397.33	9	389.86
F10	2007.47	10	282.49
F11	1724.98	11	212.23
F12	1512.75	12	76.25
F13	1436.5	13	0.08
F14	1436.42	14	0.05
F15	1436.37	15	87.29
F16	1349.08	16	53.33
F17	1295.75	17	86.7

Table A1. Cont.

First Feature	TC Value	Rank	Difference
F18	1209.05	18	157.99
F19	1051.06	19	61.32
F20	989.74	20	334.91
F21	654.83	21	36.28
F22	618.55	22	59.48
F23	559.07	23	4.36
F24	554.71	24	3.75
F25	550.96	25	23.27
F26	527.69	26	22.75
F27	504.94	27	9.94
F28	495	28	25.43
F29	469.57	29	29.43
F30	440.14	30	17.36
F31	422.78	31	16.66
F32	406.12	32	23.06
F33	383.06	33	162.32
F34	220.74	34	5.14
F35	215.6	35	8.09
F36	207.51	36	1.73
F37	205.78	37	13.6
F38	192.18	38	4.26
F39	187.92	39	1.09
F40	186.83	40	0.34
F41	186.49	41	7.44
F42	179.05	42	9.55
F43	169.5	43	33.1
F44	136.4	44	19.27
F45	117.13	45	7.05
F46	110.08	46	7.19
F47	102.89	47	7.18
F48	95.71	48	4.33
F49	91.38	49	1.03
F50	90.35	50	7.06
F51	83.29	51	3.32
F52	79.97	52	22.49
F53	57.48	53	27.15
F54	30.33	54	30.33

Appendix B The Result of Clustering Analysis on the Stand-Type Air Conditioner

Table A2. The result of clustering analysis on the stand-type air conditioner.

Number of Clusters (Ratio)		Customer Requirements	Representative Feature	Job Step
Cluster 1 (18.6%)	1	Increase the amount of releasing cool air during device operation	Cool air-weak (냉기 약하다)	Execute
	2	Increase the range of releasing cool air during device operation		Execute
Cluster 2 (9.5%)	3	Minimize the time to achieve target indoor temperature during device operation	Cool (시원한) Temperature (온도)	Execute
Cluster 3 (12.0%)	4	Minimize the frequency of at which device stops releasing cold wind	Wind (바람) Temperature (온도)	Execute
	5	Minimize the time to take to release cold wind after turning on device		Execute
Cluster 4 (5.5%)	6	Minimize the time to monitor cooling capacity of the device	Temperature (온도)	Monitor

Table A2. Cont.

Number of Clusters (Ratio)		Customer Requirements	Representative Feature	Job Step
Cluster 5 (4.0%)	7	Minimize the likelihood of side effect after installation	Installation (설치)	Prepare
	8	Minimize the time to identify side effect after installation		Confirm
Cluster 6 (6.2%)	9	Minimize scratches on the surface of the device	Scratch (기스)	Monitor
	10	Minimize cracks on the surface of the device		Monitor
	11	Minimize discoloration on the surface of the device		Monitor
Cluster 7 (7.0%)	12	Minimize the likelihood of malfunction after voice command	Voice (음성)	Execute
	13	Increase speech recognition rate during voice command mode		Execute
Cluster 8 (4.9%)	14	Minimize the frequency of hiatus of device operation	Power supply (전원)	Execute
	15	Minimize number not to run power button		Execute
Cluster 9 (5.2%)	16	Minimize noise of parts which decide the direction of the wind during device operation	Noise (소음), wing (날개)	Monitor
Cluster 10 (5.6%)	17	Minimize the frequency not to function the direction of the wind in modifying the wind	wing (날개)	Modify
Cluster 11 (5.3%)	18	Minimize the frequency which device emits stench during device operation	Smell (냄새)	Monitor
Cluster 12 (4.7%)	19	Minimize indoor noise during device operation	Indoor (실내), Noise (소음)	Monitor
Cluster 13 (5.4%)	20	Minimize outdoor noise during device operation	Outdoor (실외), Noise (소음), Operation (작동)	Monitor
	21	Minimize the time to monitor loudness level of noise during device operation		Monitor
Cluster 14 (6.0%)	22	Minimize noise in performing specific function	Noise (소음)	Modify

References

1. Lengnick-Hall, C.A. Customer contributions to quality: A different view of the customer-oriented firm. *Acad. Manag. Rev.* **1996**, *21*, 791–824. [\[CrossRef\]](#)
2. Nambisan, S. Designing virtual customer environment for new product development: Toward a theory. *Acad. Manag. Rev.* **2002**, *27*, 392–413. [\[CrossRef\]](#)
3. Rigby, D.; Zook, C. Open-market innovation. *Harv. Bus. Rev.* **2002**, *80*, 80–81. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Atuahene-Gima, K. An Exploratory Analysis of the impact of market orientation on new product performance: A contingency approach. *J. Prod. Innov. Manag.* **1995**, *12*, 19. [\[CrossRef\]](#)
5. Zhan, J.; Loh, H.T.; Liu, Y. Gather customer concerns from online product reviews—A text summarization approach. *Expert Syst. Appl.* **2009**, *36*, 2107–2115. [\[CrossRef\]](#)
6. Decker, R.; Trusov, M. Estimating aggregate consumer preferences from online product reviews. *Int. J. Res. Mark.* **2010**, *27*, 293–307. [\[CrossRef\]](#)
7. Park, Y.; Lee, S. How to design and utilize online customer centre to support new product concept generation. *Expert Syst. Appl.* **2011**, *38*, 10638–10647. [\[CrossRef\]](#)
8. Aguwa, C.C.; Monplaisir, L.; Turgut, O. Voice of the customer: Customer satisfaction ratio based analysis. *Expert Syst. Appl.* **2012**, *39*, 10112–10119. [\[CrossRef\]](#)
9. Wang, Y.; Tseng, M.M. A Naïve Bayes approach to map customer requirements to product variants. *J. Intell. Manuf.* **2015**, *26*, 501–509. [\[CrossRef\]](#)
10. Aguwa, C.; Olya, M.H.; Monplaisir, L. Modeling of fuzzy-based voice of customer for business decision analytics. *Knowl.-Based Syst.* **2017**, *125*, 136–145. [\[CrossRef\]](#)
11. Liang, R.; Guo, W.; Yang, D. Mining product problems from online feedback of Chinese users. *Kybernetes* **2017**, *46*, 572–586. [\[CrossRef\]](#)

12. Qiao, Z.; Zhang, X.; Zhou, M.; Wang, G.A.; Fan, W. A domain oriented LDA model for mining product defects from online customer reviews. In Proceedings of the 50th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 4–7 January 2017.
13. Wang, Y.; Lu, X.; Tan, Y. Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines. *Electron. Commer. Res. Appl.* **2018**, *29*, 1–11. [[CrossRef](#)]
14. Min, H.; Yun, J.; Geum, Y. Analysing dynamic change in customer requirements: An approach using review-based Kano analysis. *Sustainability* **2018**, *10*, 746. [[CrossRef](#)]
15. Leonard, D. The limitations of listening. *Harv. Bus. Rev.* **2002**, *1*, 155–171.
16. Füller, J.; Matzler, K. Virtual product experience and customer participation—A chance for customer-centred, really new products. *Technovation* **2007**, *27*, 378–387. [[CrossRef](#)]
17. Christensen, C.M.; Anthony, S.D.; Berstell, G.; Nitterhouse, D. Finding the right job for your product. *MIT Sloan Manag. Rev.* **2007**, *48*, 38.
18. Ulwick, A.W. *What Customers Want: Using Outcome-Driven Innovation to Create Breakthrough Products and Services*; McGraw-Hill: New York, NY, USA, 2005.
19. Ulwick, A.W. *Jobs to Be Done: Theory to Practice*; IDEA BITE Press: New York, NY, USA, 2016.
20. Menon, R.; Tong, L.H.; Sathiyakeerthi, S.; Brombacher, A.; Leong, C. The needs and benefits of applying textual data mining within the product development process. *Qual. Reliab. Eng. Int.* **2004**, *20*, 1–15. [[CrossRef](#)]
21. Bradley, G.L.; Sparks, B.A.; Weber, K. Perceived prevalence and personal impact of negative online reviews. *J. Serv. Manag.* **2016**, *27*, 507–533. [[CrossRef](#)]
22. Hennig-Thurau, T.; Gwinner, K.P.; Walsh, G.; Gremler, D.D. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *J. Interact. Mark.* **2004**, *18*, 38–52. [[CrossRef](#)]
23. King, R.A.; Racherla, P.; Bush, V.D. What we know and don't know about online word-of-mouth: A review and synthesis of the literature. *J. Interact. Mark.* **2014**, *28*, 167–183. [[CrossRef](#)]
24. Johnson, P.A.; Sieber, R.E.; Magnien, N.; Ariwi, J. Automated web harvesting to collect and analyse user-generated content for tourism. *Curr. Issues Tour.* **2012**, *15*, 293–299. [[CrossRef](#)]
25. Liu, B. *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data*; Springer Science & Business Media: Berlin, Germany, 2011.
26. Mankad, S.; Han, H.S.; Goh, J.; Gavirneni, S. Understanding online hotel reviews through automated text analysis. *Serv. Sci.* **2016**, *8*, 124–138. [[CrossRef](#)]
27. Miguéis, V.L.; Nóvoa, H. Exploring online travel reviews using data analytics: An exploratory study. *Serv. Sci.* **2017**, *9*, 315–323. [[CrossRef](#)]
28. Ordenes, F.V.; Theodoulidis, B.; Burton, J.; Gruber, T.; Zaki, M. Analysing customer experience feedback using text mining: A linguistics-based approach. *J. Serv. Res.* **2014**, *17*, 278–295. [[CrossRef](#)]
29. Bettencourt, L.A.; Ulwick, A.W. The customer-centered innovation map. *Harv. Bus. Rev.* **2008**, *86*, 109. [[PubMed](#)]
30. Ulwick, A.W. Turn customer input into innovation. *Harv. Bus. Rev.* **2002**, *80*, 91–97. [[PubMed](#)]
31. Leonard, D.; Rayport, J.F. Spark innovation through empathic design. *Harv. Bus. Rev.* **1997**, *75*, 102–115. [[CrossRef](#)]
32. Bettencourt, L. *Service Innovation: How to Go from Customer Needs to Breakthrough Services*; McGraw Hill Professional: New York, NY, USA, 2010.
33. Lim, J.; Choi, S.; Lim, C.; Kim, K. SAO-based semantic mining of patents for semi-automatic construction of a customer job map. *Sustainability* **2017**, *9*, 1386. [[CrossRef](#)]
34. Oestreicher, K.G. Segmentation & the jobs-to-be-done theory: A conceptual approach to explaining product failure. *J. Mark. Dev. Compet.* **2011**, *5*, 103.
35. Boyd-Graber, J.; Mimno, D.; Newman, D. Care and feeding of topic models: Problems, diagnostics and improvements. In *Handbook of Mixed Membership Models and Their Applications*; CRC Press: Boca Raton, FL, USA, 2014; pp. 225–255.
36. Mitra, P.; Murthy, C.A.; Pal, S.K. Unsupervised feature selection using feature similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 301–312. [[CrossRef](#)]

37. Liu, T.; Liu, S.; Chen, Z.; Ma, W. An evaluation on feature selection for text clustering. In Proceedings of the 20th International Conference on Machine Learning, Washington DC, USA, 21–24 August 2003; pp. 488–495. [\[CrossRef\]](#)
38. Alelyani, S.; Tang, J.; Liu, H. Feature selection for clustering: A review. *Data Clust. Algorithms Appl.* **2013**, *29*, 110–121.
39. Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer: Berlin, Germany, 2002.
40. Slonim, N.; Tishby, N. Document clustering using word clusters via the information bottleneck method. In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece, 24–28 July 2000; pp. 208–215.
41. Yang, Y.; Pedersen, J.O. A comparative study on feature selection in text categorization. In Proceedings of the Fourteenth International Conference on Machine Learning, Nashville, TN, USA, 8–12 July 1997; pp. 412–420.
42. Dash, M.; Liu, H. Feature selection for classification. *Intell. Data Anal.* **1997**, *1*, 131–156. [\[CrossRef\]](#)
43. Salton, G. Automatic text processing: The transformation. *Anal. Retr. Inf. Comput.* **1989**, *14*, 15.
44. Joung, J.; Kim, K. Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data. *Technol. Forecast. Soc. Chang.* **2017**, *114*, 281–292. [\[CrossRef\]](#)
45. Ulwick, A.W.; Bettencourt, L.A. Giving customers a fair hearing. *MIT Sloan Manag. Rev.* **2008**, *49*, 62–68.
46. Wu, H.C.; Luk, R.W.P.; Wong, K.F.; Kwok, K.L. Interpreting TF-IDF term weights as making relevance decisions. *ACM Trans. Inf. Syst.* **2008**, *26*, 13. [\[CrossRef\]](#)
47. Lam, S.K.; Sleep, S.; Hennig-Thurau, T.; Sridhar, S.; Saboo, A.R. Leveraging frontline employees' small data and firm-level big data in frontline management: An absorptive capacity perspective. *J. Serv. Res.* **2017**, *20*, 12–28. [\[CrossRef\]](#)



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).