

Voice of Customer Analysis for Internet Shopping Malls

Chi-Hwan Choi¹, Jeong-Eun Lee², Gyeong-Su Park², Jonghwa Na³ and Wan-Sup Cho¹

¹ Dept. of Bio-Information Technology

² Dept. of Business Data Convergence

³ Dept. of Information & Statistics/Business Data Convergence

⁴ Dept. of MIS/Business Data Convergence

Chungbuk National University, Cheongju, Korea

¹chihwan.choi86@gmail.com, { ²masjjong, ²bigonbiz, ³wscho, ⁴cherin }@chungbuk.ac.kr

Abstract

Advances in Internet and Smartphones are generating huge amount of customers' opinion (or VOC: voice of customer) in the online world. VOC is very useful for marketing or product design departments to see the customers' opinion for the products. In this paper, VOC in online review sites has been collected, stored and analyzed in order to find out what kind of factors can affect the customer preference for the electronics products. We have implemented a VOC analysis system that includes ETL (extract, transformation, and loading), natural language processing, positive/negative analysis, and result visualization. Sensitivity adjectives and synonyms are collected and stored as a database for more correct sentiment analysis. Real experiment for VOC data with 39 Giga bytes from well-known review sites in Korea shows that the correctness for the classification of the positive/negative sentences and the causes of the positive/negative are 73% and 70%, respectively. The evaluation has been done based on the manual inspection for the analysis result

Keywords: Sentimental Analysis, Opinion Mining, Voice of Customer, VOC

1. Introduction

As smart devices and wireless Internet have come into wide use, the amount of VOC (voice of customer) data is growing exponentially in the on-line world. Customer has been reflecting their opinion through Internet review sites and SNS data and they become an important source for finding customer preference for products, services, an individual, or policies. **Sentiment analysis** (opinion mining) for online review sites, personal blogs, or SNS data is becoming an active research issue in the business and IT areas [4, 5, 10]. Sentiment analysis aims to determine the attitude of a customer with respect to some topic or the overall contextual polarity of a document [10, 12, 15]. The attitude may be his or her judgment or evaluation (see appraisal theory [13]), affective state (that is to say, the emotional state of the customer when writing), or the intended emotional communication (that is to say, the emotional effect the customer wishes to have on the reader) [8, 10, 12].

⁴ Corresponding author.

In this paper, we analyze customer opinions for widely used electronics for Internet review sites in Korea. To find out which key factors most affect the customers when buying electronics, about 39GB data has been collected from the review sites. Well-defined keywords such as product or brand names are used for filtering out unrelated data from the review sites. Sensitivity adjectives and synonyms are used for intelligent analysis because there are too many abbreviations, synonyms, acronyms, and slang in the Internet.

The analysis shows which electronics is the most frequently referenced in the review sites. The preference can be classified into two categories: *positive* and *negative*. And the causes of the preference can be provided for more deep sentiment analysis. For selected electronics (*e.g.*, TV or washer), it shows the positive/negative trends for a given time interval; it also provides product preferences trends and the manufacturer (brand) preference trends. This result would be very useful for both the product makers and the customers.

Real experiment for VOC data with 39 GB from well-known review sites in Korea has been done for checking performance and correctness of the system. Due to the customized web crawler instead of the open source crawler such as Heritrix[17], web crawling speed has been accelerated significantly. The experiment shows that the correctness for the classification of the positive/negative sentences and the causes of the positive/negative are 73% and 70% respectively. The evaluation has been done based on the sampling and manual inspection for the analysis result.

The most important problem is appropriate handling of abbreviations, synonyms, acronyms, and slang widely used in Internet and smartphones. However, these problems can be moderated as the number of abbreviations, synonyms, acronyms, and slang in the dictionary increases over time.

The paper is organized as follows. In Section 2, we describe related work for sentiment analysis. In Section 3, we propose the system architecture and data life cycle for the analysis. In Section 4, we present real experimental result for well-known review sites. Finally, in Section 5, we conclude the paper.

2. Related Work

In this section, we present current status of the research for sentiment analysis. Active research has been done in the sentiment analysis because of the various commercial and social needs.

An important part of our information-gathering behavior has always been to find out what other people think [2, 4, 10, 12, 13]. With the growing availability and popularity of opinion-rich resources such as various social media, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others [10, 12].

Sentiment analysis, also called opinion mining, is the research area that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [10, 12].

After the year 2000, sentiment analysis becomes a very active research area because it has a wide range of useful applications including commercial and social needs [8, 10, 12, 13]. Business and organizations always want to find consumer or public opinions about their products or services [10, 12]. Individual consumer also wants to know the opinions of existing users of a product before purchasing it [10]. Social media and

Internet is the main source from which we can collect the customers' opinion in a real time style.

Automated sentiment analysis systems are becoming more important because of the proliferation of diverse sites (social media), *e.g.*, review sites, forum discussions, blog and micro-blogs, Twitter, etc. It takes too long time to collect, cleanse, and analysis the review data in the Internet if we adopt manual approach.

Web crawling is the first step for the automated sentiment analysis. Web crawler (also called robot or spider) is a program that finds and downloads web pages automatically from Internet. There are challenges in the crawling. There are tens of billions of pages on the web to be collected. Many new pages emerge or change at any time and anywhere. And there is no central control of all web pages, no easy way to find out how many pages in a website. **Focused crawling** is downloading web pages that are related to a specific topic, and avoid downloading all others [3, 5-7]. **Wrapper based crawling** uses structural information for guiding the crawler [1, 9]. These crawlers are less expensive approach than normal ones.

After crawling, we can do the sentiment analysis. The analysis can be done at three levels: **document level, sentence level, and feature level**. For example, the statement "*Samsung Smart TV has sharp picture, but power consumption is not so good.*" has two aspects, *picture* and *power consumption*, with positive and negative, respectively. This is an example of the aspect level sentiment analysis.

Sentiment words or phrases or idioms are very important in the sentiment analysis. For example, *cool, nice, or wonderful* expresses positive, and *poor or terrible* means negative sentiment. Furthermore, there are full of abbreviations, synonyms, acronyms, slang, and even emoticon in the social media for representing their sentiment. All these words or phrases should be stored in the database called **sentiment lexicon** [11, 19]. There are many challenges in the construction and usage of the sentiment lexicon.

Most sentiment analysis includes restricted natural language processing function. It need not fully understand the semantic of each sentence, but need to understand positive or negative sentiments, related features, and the causes of the sentiment.

In this paper, we use adopted feature level wrapper based crawling for collecting reviews from on line shopping malls. We constructed lexicon dictionary storing sentiment words, abbreviations, synonyms, acronyms, slang for identifying product names, brand names, and sentiments in the reviews. The system has a limited natural language processing capability for understanding products and brand names, sentiment words including abbreviations or synonyms etc.

3. System Architecture and VOC Data Analysis

In this section, we describe the system architecture for the sentiment analysis for on-line shopping mall sites. Then we describe entire data life cycle for the analysis: *data collection, data cleansing and storage, natural language processing, data analysis, and result visualization*.

3.1. System Architecture

Figure 1 shows the system architecture for the sentiment analysis. At first, the review sites' URLs should be selected before the system starts. The next step is the data collection, in which the web crawler starts to collect relevant web pages from the sites and the result is stored in a database or files. The next step is the data cleansing to filter out the pages or data that are not related to the topics we are interested. After filtering, natural language processing

module identifies meaningful words from the sentences collected. If the word corresponds to a product name, a brand name, or senti3ment-related word, it can be a meaningful word. Then the system does analysis to the collected data and finds the positive and negative ratio of the review data. Furthermore, it tries to finds out the causes of the positive or negative opinion for more deep understanding for the sentences of phrases. Finally, the system provides various visualizations (or reporting) for the analysis.

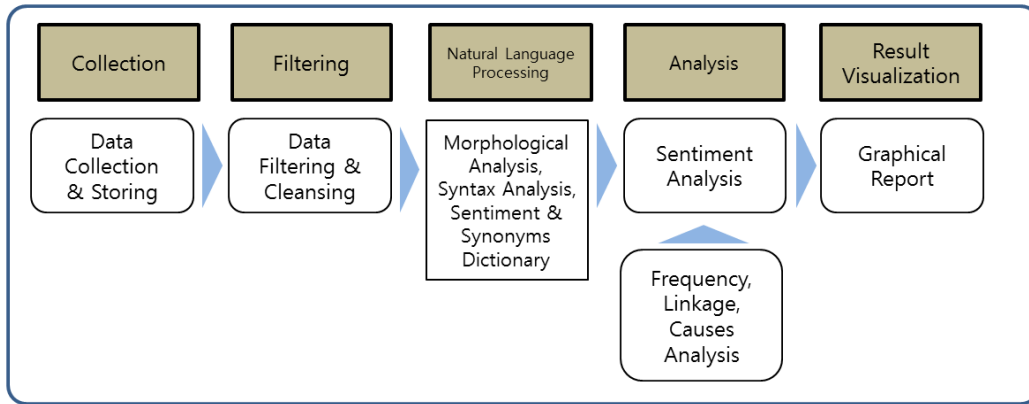


Figure 1. System Architecture – Data Life Cycle

3.2. Data Collection and Storing

The first step is data collection. We tried to collect user’s reviews from web sites by using open source web crawler called Heritrix [17]. However Heritrix-based web crawler can collect only a part of the entire web pages; furthermore, it collects unnecessary data because review sites are very complex and consisting of dynamic server pages such as JSP, ASP, PHP. It takes long time (more than a week in our case) to extract the pages. To overcome this problem, we constructed a customized web crawler for the shopping mall sites. We adopted wrapper-based web crawler [1, 9] because we already have structural information for the on-line shopping mall sites.

Review sites commonly have five-depth hierarchies including main categories and lot of sub-categories (*e.g.*, appliance \ni television \ni led \ni specific TV \ni comments). Figure 2 shows a hierarchy for a typical review site. Even though the open source crawler in our case Heritrix[17] can collect review comments from the web, it may contain a lot of unnecessary web contents and crawling the whole web site is highly time consuming task. In our case, it takes more than 1 week to collect the review data from a shopping mall. In addition, the collected data is not guaranteed whether they meet our interest or not. For this reason we have developed a new web crawler for the review sites we are interested.

Table 1 shows the comparison between open source web crawler and customized web crawler for the performance and accuracy aspects. Open source crawler takes more than 12 days to collect the data in the shopping mall site, but customized crawler takes just 30 hours for the collection. Furthermore, valid data ratios of the open source crawler and customized crawler are 0.67% and 99.98%, respectively. We can see that most of the collected data from the open source web crawler seems to be garbage. This is why we construct a new customized web crawler for the Internet shopping malls.

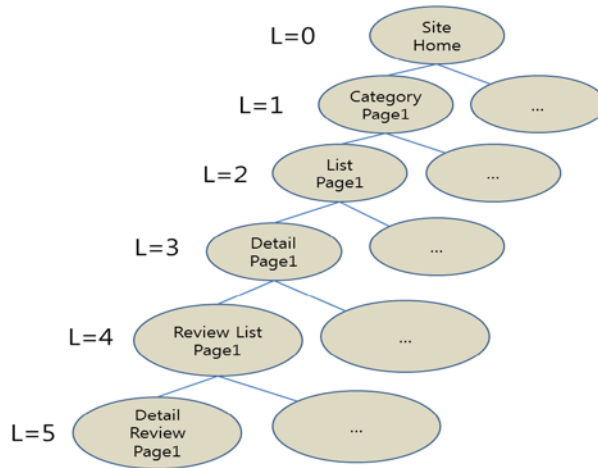


Figure 2. Tree Structure for Shopping Mall Sites

Table 1. The Comparison between Open Source and Customized Web Crawlers

Provision	Open Source Crawler	Customized Crawler
Period	12 days + α	30 hours
Amount	36GB	39GB
The number of valid data / the number of total data	3012/450665	24312/24316
Valid data ratio	0.67%	99.98%

Figure 3 shows the customized web crawler in detail. HTMLs file downloaded from the Internet shopping mall includes useful information (*i.e.*, category, subcategory, product's name, product's specifications and VOCs) for the later sentiment analysis. The HTMLs are processed one by one for storing the information in the database.

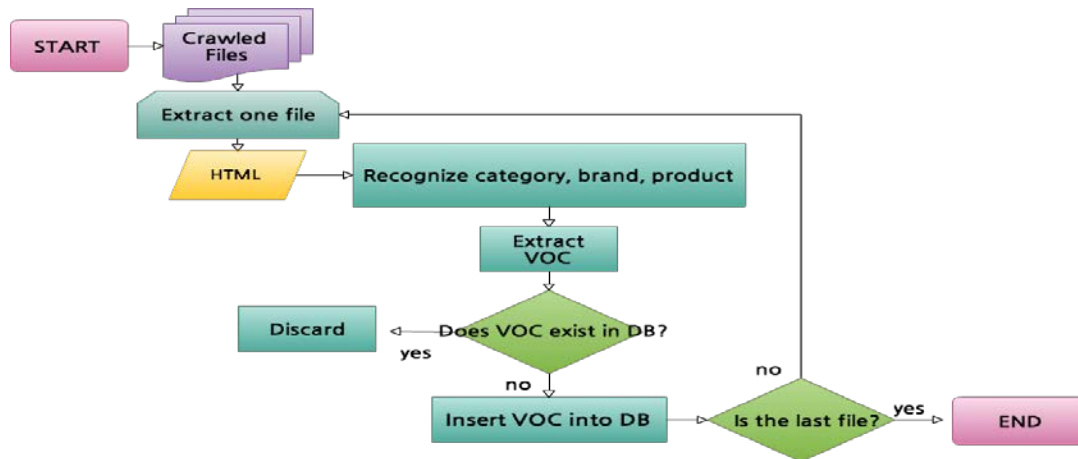


Figure 3. Web Crawler

Extracted review data has the following components as shown in Table 2: *id&nickname, comment, date&time and IP address.*

Table 2. Review Data Format

Provision	Description
ID & Nickname	UserID and user nickname
Comments	Comments up to 1000 characters
Date & Time	Comments time posting
IP address	Comments posting IP address

3.3. Filtering, Cleansing, and Natural Processing

Filtering and cleansing step is essential for the next natural language processing because there are frequent uses of colloquial style, vulgar language, a lot of spelling mistakes, and the text consists of consonants or vowels or emoticon only. Commonly used words have been stored in the dictionary table for processing non-standard words in the Internet.

The next step is morphological analysis. We use the morphological analyzer to determine and mainly extract the sentences having a "clear expression". The role of Korean morpheme is especially important because word-structural changes occur frequently depending on the postposition of the word. We used HANNANUM API [18], a Java version of the morphological analyzer, to categorize the review data in the form of predicates, nouns, postpositions etc. The storage of the review data is VOC(Voice of customers) table. Although HANNANUM morphological analyzer contains dictionary information, the word "Galaxynote" is divided into "Galaxy" + "Note". To solve this kind of error, built-in dictionary information in the analyzer has been reconfigured.

Depending on the level of the analysis of morphological analyzer, it consists of three stages: *text pre-processing*, *morphological analysis*, and *POS tagging*.

- Text pre-processing stage: filtering such as recognizing sentence boundaries, filtering, automatic word spacing, etc. for subsequent morphological analysis.
- Morphological analysis: all possible words are identified for the input sentence.
- POS Tagging: word classes are identified for the input sentence. For example, the sentence "Galaxynote Ice cream sandwich update very good." generates "Galaxynote/n Ice/n cream/n sandwich/n update/v very/av good/a." with corresponding word class for each word.

Another important issue is the synonym table for product names, brand names, and sentimental words. Numerous synonyms are used in the Internet review data. Note that the quality of analysis critically depends on how we collect, organize, and utilize the synonyms and acronyms in the review comments. For instance, "Galnot" and "Galaxynote" indicate the same product. We constructed these kinds of synonyms manually by using more than 3,000 review data as a training set, and the user can add or modify the synonyms in the later upon finding a synonyms. Continuous updating those synonyms can make the result of the analysis clear and precise.

3.4. Sentiment Analysis and Visualization

By using sentiment lexicon (a database of abbreviations, synonyms, acronyms, slang, and even emoticon) and sentiment adjective words, each sentence can be represented by a tuple of five components: *feature, feel, opinion, qualification, sentiments*.

- *Feature* means the characteristics of the product or service such as display, battery, touch, android, ios for the smartphone.
- *Feel* is the modifiers of the features. For example, in the phrase “*thin display*”, “*thin*” is the feel.
- *Opinion* is the sentiment word in the sentence or phrase. For example, in the sentence “*I’m satisfied with galnote’s battery performance*”, the opinion part is “*satisfied*”.
- *Qualification* is the modifier of negative/positive meaning. For example, in the sentence “*battery is not good*”, “*not*” is the qualification.
- *Sentiment* is the final opinion for the sentence. Opinion and qualification are used for determining the final sentiment.

The result of the sentiment analysis is a table of tuples with five attributes as shown at Table 3. Each sentence may correspond to a tuple in the table. Result of the sentiment analysis can be transformed into graphical representation for intuitive understanding. R package [20] provides a reasonable visualization technique for the result.

Example: For the following two sentences, the system constructs two tuples as shown at Table 3. It consists of *Feature, Feel, Opinion, Qualification, and Sentiment* attributes.

Sentence: *Galaxynote Ice cream sandwich update very good*

GalNot has a perfect digital camera !

Tuple 1 : Galaxynote/noun, Ice cream sandwich update/feature, very/qualification, good/opinion.

Tuple 2 : Galnot/feature, high resolution/feature, digital camera/feature

Table 3. Examples of Sentiment Analysis Sentence

Feature	Feel	Opinion	Qualification	Sentiments
<i>Ice cream sandwich update</i>	N/A	good	very	positive
<i>camera</i>	N/A	perfect	N/A	positive

Note that features are classified into seven categories: *Robustness, Display, Design, Price, Ease of Manipulation, Others*. The categories actually depend on the kinds of the product. Here we assume TV and washer for avoiding complexity.

4. Experiment

In this section, we describe real experiment for well-known Internet shopping malls in Korea. 39GB review data has been collected from the sites by using customized web crawler. Inappropriate data has been excluded by filtering and necessary natural language processing has been done by using HANNANUM [17] API.

4.1. System Environments

Table 4 shows the hardware system for sentiment analysis; it consists of one PC for storing and analyzing the data from the shopping malls. Since the data is increasing rapidly as we add new shopping mall sites to the system, we are going to extend the system on top on cloud.

Table 5 shows the software environment. A customized web crawler is used for data collection from the shopping mall, a relational database system (MySQL) for storing the review data, and R-package for sentiment analysis and visualization.

Table 4. Hardware Environment

Provision	Specification
CPU	Intel(R) i7 870
RAM	16GB
HDD	2TB
Network	1000baseT/Full

Table 5. Software Environment

Provision	Version
OS	CentOS 6.3 64bit final
MYSQL	5.0.77
R-package	2.15.0
JAVA	1.6.0
Customized WebCrawler	0.1

4.2. Frequency Analysis

Although the system supports sentiment analysis for various kinds of electronics, we describe just *TV* and *washer* to evade complexity. In the system, user can select target product from more than 100 electronics for the sentiment analysis. Note that specific features in the review sentence are classified into six or seven categories according to the products as explained at Section 3.4.

For TV, Figure 4 (left graph) shows the frequency analysis for the features. In the figure, *robustness* is the most frequently mentioned feature (55.8%). Then *display* (16.3%), *design* (9.3%), *price* (7.8%), and *ease of manipulation* (7%) follow the *robustness*.

For washer, Figure 4 (right graph) shows the frequency analysis. The figure shows that *performance* is the most frequently mentioned feature (48.1%). The next features are *ease of manipulation* (16.9%), *robustness* (16.9%), *design* (14.3%), and the *price* (2.6%), etc.

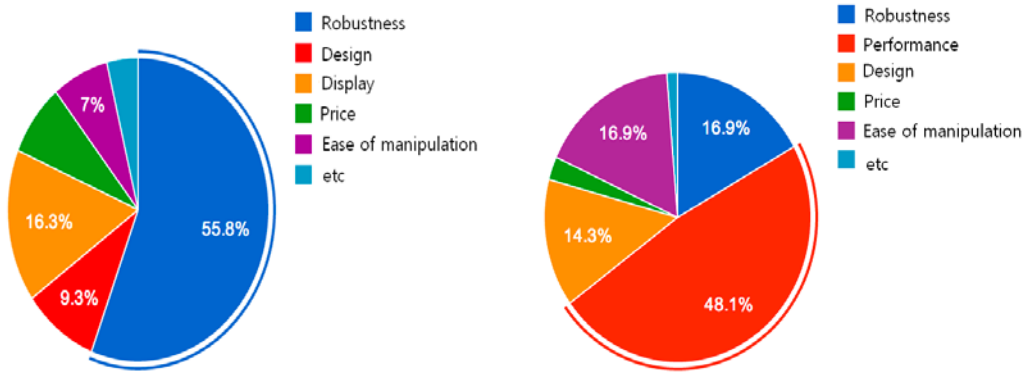


Figure 4. Features Frequency Analysis for TV (left) and Washer (right)

4.3. Sentiment Analysis

Sentiment analysis for the **TV** made by the *company A* shows the following result. Table 6 shows the result and Figure 5 show the visualization of the result.

- **Robustness** feature has positive opinion of 7.8%, negative 11.6%, and neutral 36.4%.
- **Design** feature has positive opinion 5.4%, negative 0.3%, and neutral 3.6%.
- **Display** feature has positive opinion 7.8%, negative 0.7%, and neutral 7.8%.
- **Design** feature has positive opinion 5.4%, negative 0.2%, and neutral 3.7%.
- **Price** feature has positive opinion 3.1%, negative 1.6%, and neutral 3.1%.
- **Ease of manipulation** feature has positive opinion 4.7%, negative 0%, and neutral 2.3%.

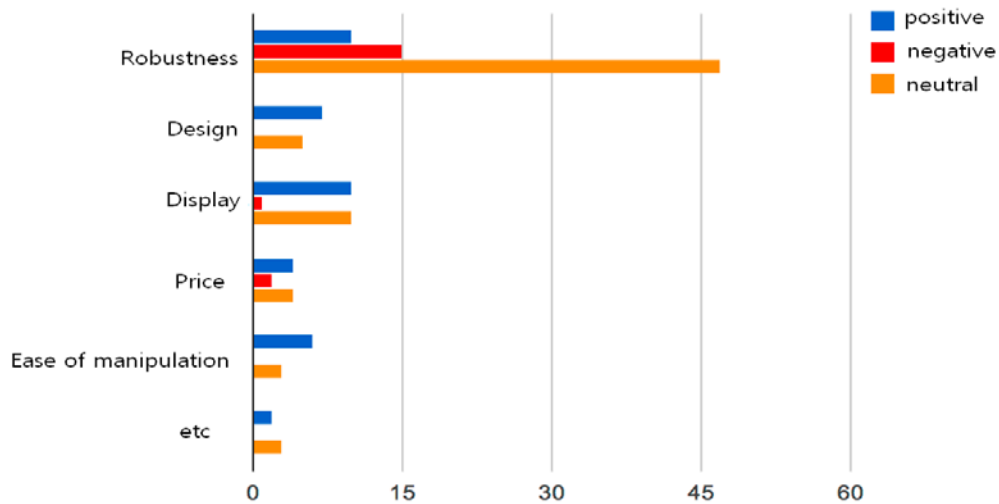


Figure 5. Sentiment Analysis for TV

Table 6. Sentiment Analysis for TV

Feature (frequency)	Ratio of frequency		
	Positive	Neutral	Negative
Robustness (55.8%)	7.8%	36.4%	11.6%
Design (9.3%)	5.4%	3.6%	0.3%
Display (16.3%)	5.4%	3.7%	0.2%
Price (7.8%)	3.1%	3.1%	1.6%
Ease of manipulation (7%)	4.7%	2.2%	0.1%

Sentiment analysis for the *washer* of the *company A* has been analyzed in similar fashion. Table 7 shows the analysis result and Figure 6 shows its visualization.

- **Performance** feature has positive opinion of 33.8%, negative 1.3%, and neutral 13%.
- **Robustness** feature has positive opinion of 10.4%, negative 1.3%, and neutral 5.2%.
- **Ease of manipulation** feature has positive opinion of 14.3%, negative 0%, and neutral 2.6%.
- **Design** feature has positive opinion of 10.4%, negative 1.3%, and neutral 2.6%.
- **Price** feature has positive opinion 1.3 %, negative 0 %, and neutral 1.3%.

This result shows that the customer has interest in the robustness and performance rather than the design or price.

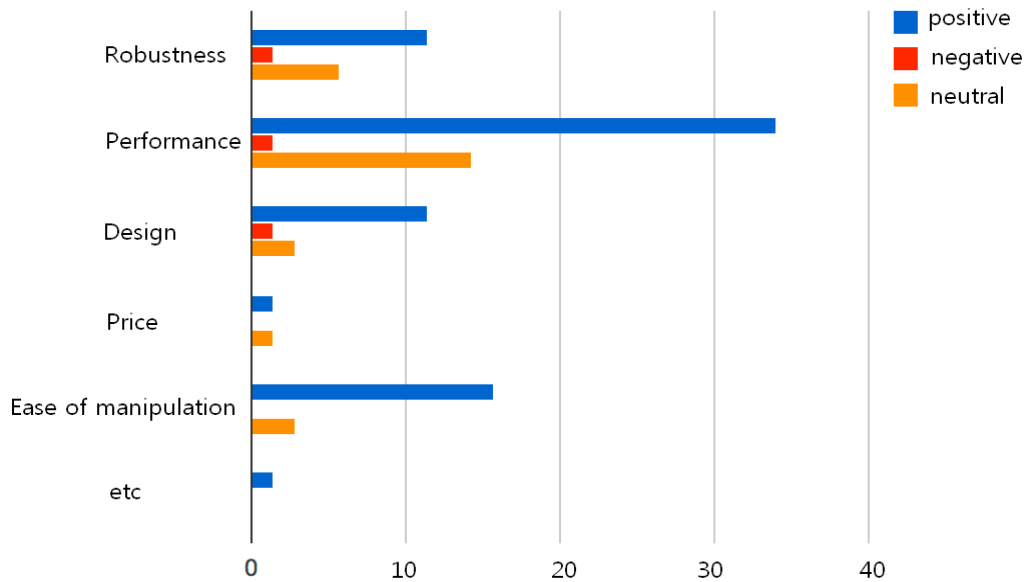


Figure 6. Sentiment Analysis for Washer

Table 7. Sentiment Analysis for Washer

Feature	Ratio of frequency		
	Positive	Neutral	Negative
Performance (48.1%)	33.8%	1.3%	13%
Robustness (16.9%)	10.4%	1.4%	5.2%
Ease of manipulation (16.9%)	14.3%	0%	2.6%
Design (14.3%)	10.4%	1.3%	2.6%
Price (2.6%)	1.3%	0%	1.3%

4.4. Evaluation

In this section, we evaluate the accuracy for the sentiment analysis. The accuracy evaluation can be done by checking the result generated by the system manually. There are 68,085 sentences in the experiment, and we choose 1,000 sentences randomly and check the correctness by manually. The manual checking show 76.2% on average of the sentences are correctly analyzed by the system.

$$accuracy = \frac{\text{TheNumberofCorrectness}}{\text{TheNumberofTotalSentences}}$$

Table 8. Accuracy of Sentiment Analysis about TV and Washer

Provision	Accuracy			
	Positive	Neutral	Negative	Average
TV	86.1%	90.1%	50%	75.4%
washer	60%	80%	90%	77%
Average	73%	85%	70%	76.2%

5. Conclusions

In this paper, we proposed a sentiment analysis system for review sites in Internet shopping malls. Customized web crawler, filtering and natural language processing tools, sentiment analysis and visualization tools have been implemented in the system. Users can see the positive and negative opinions and their causes for various kinds of electronics products such as TVs, washers, PC, and smartphones, *etc.*

Real experiment for VOC data with 39 Giga bytes from well-known review sites in Korea shows that the correctness for the classification of the positive/negative sentences and the causes of the positive/negative are 73% and 70%, respectively.

The most serious problem in the analysis is the quality of the result that comes largely from the shortage of samples (or training set). Another serious problem is handling of abbreviations, synonyms, acronyms, and slang in Internet or smartphones.

However, we believe that these problems can be moderated as the number of synonyms/acronyms in the dictionary increases continuously.

References

- [1] C. Bertoli, V. Crescenzi and P. Merialdo, "Crawling Programs for Wrapper-based Application", Proc. Intl. Conf. on Information Reuse and Integration (IRI), Las Vegas, USA., vol. 13-15, (2008) July, pp.160-165.
- [2] S. Chakrabarti, "Mining the web", Morgan Kaufmann Publishers, San Francisco, (2003).
- [3] S. Chakrabarti, M. van der Berg and B. Dom, "Focused crawling: a new approach to topic-specific web resource discovery", In Proc. Intl. Conf. on WWW, (1999).
- [4] C. H. Chang, M. Kayed, M. R. Girgis and K. F. Shaalan, "A Survey of Web Information Extraction Systems", IEEE Transaction on Knowledge and Data Engineering, vol. 19, no. 10, October, (2006), pp. 1411-1428.
- [5] J. Cho, "Crawling the Web: Discovery and Maintenance of Large-Scale Web Data", Stanford University, Ph. D. Thesis, (2001).
- [6] M. Diligenti, F. Coetzee, S. Lawrence, C. L. Giles and M. Gori, "Focused crawling using context graphs", Proc. Intl. Conf. on Very Large Databases (VLDB), (2000).
- [7] A. M. H. Elyasir and K. S. M. Anbananthen, "Focused Web Crawler", Proc. Intl. Conf. on Information and Knowledge Management (CIKM), (2012).
- [8] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews", Proc. Intl. Conf. on ACM SIGKDD, (2004), pp. 168-177.
- [9] H. H. Kang, S. J. Yoo and D. I. Han, "Modeling Web Crawler Wrappers to Collect User Reviews on Shopping Mall with Various Hierarchical Tree Structure", Proc. Intl. Conf. on Web Information Systems and Mining(WISM), Shanghai, (2009) November.
- [10] B. Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publisher, (2012) May.
- [11] Y. Lu, M. Castekkanos, U. Dayal and C. Zhai, "Automatic Construction of a Context-Aware Sentiment Lexicon: An Opotimization Approach", Tutorial, AAAI, (2011).
- [12] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, (2008) January, pp. 1-135.
- [13] G. Pant, P. Srinivasan and F. Menczer, "Crawling the Web", Springer Berlin Heidelberg Publisher, (2004), pp. 153-177.
- [14] K. R. Scherer, A. Shorr and T. Johnstone, "Appraisal processes in emotion: theory, methods, research", Canary, NC: Oxford University Press, (2001).
- [15] J. H. Yeon, D. J. LEE, J. H. Shim and S. G. Lee, "Product Review Data and Sentiment Analytical Processing Modeling", The Journal of Society for e-Business Studies, vol. 16, no. 4, (2011), pp. 125-137.
- [16] Wikipedia, <http://www.wikipedia.org>.
- [17] Heritrix User Manual, http://crawler.archive.org/articles/user_manual/.
- [18] HANNANUM, <http://sourceforge.net/projects/hannanum/>.
- [19] Sentiment Lexicon, <http://www.stanford.edu/class/cs424p/materials/ling287-handout-09-21-lexicons.pdf>.
- [20] R Project, <http://www.r-project.org/>.

Authors



Chi-Hwan Choi received MS degree in bio-information technology from Chungbuk National University. He is currently working toward the Ph. D. degree in bio-information technology from Chung-buk National University.

His research interests include database systems, bigdata, bioinformatics, data mining, business intelligence, server architecture, and distributed computing platform.



Jeong-Eun Lee received the BS degree in computer science from Chungbuk National University in 2011. She is currently working toward the MS degree in business data convergence from Chungbuk National University.

Her research interests include database systems, bigdata, data mining, and business intelligence.



Gyeong-Su Park received the BS degree in computer science from Chungbuk National University in 2012.

He is currently working toward the MS degree in business data convergence from Chungbuk National University.

His research interests include database systems, bigdata, business intelligence, ERP.



Jonghwa Na received the Ph. D. degree in Statistics from Seoul National University (SNU) in 1994. From 1994 to 1996, he served as researcher in Institute of Statistics of SNU, Korea. He is now a professor in the department of statistics of Chungbuk National University. His research interests include data mining, social network analysis, big data, and statistical modeling & forecasting.



Wan-Sup Cho received the Ph. D. degree in computer science from Korea Advanced Institute of Science and Technology (KAIST) in 1996. From 1987 to 1990, he served as a research engineer in the Electronics and Telecommunications Research Institute(ETRI), Korea. He is now a full professor in the Management Information Systems of Chungbuk National University. His research interests include database systems, bigdata, business intelligence, and ERP.

