International Journal of Software Engineering and Its Applications
Vol. 9, No. 1 (2015), pp. 191-200
http://dx.doi.org/10.14257/ijseia.2015.9.1.17

Web Application for Sentiment Analysis Using Supervised Machine Learning*

Bryan Nii Lartey Laryea[1], Chi-Hwan Choi[2], In-Sun Jung[3], Kyung-Hee Lee[4] and Wan-Sup Cho[5]

[1,3] Dept. of Management Information System
[2] Dept. of Bio-Information Technology
[4] Dept. of Business Data Convergence
[5] Dept. of MIS/Business Data Convergence
Chungbuk National University, Cheongju, Korea
[1]bry_lar@yahoo.com,
{[2]chihwan.choi86, [3]517insun, [4]lee.kyunghee, [5]wscho63}@gmail.com

Abstract

Sentiment analysis is now in focus of companies to extract information from customers' reviews. Usually the analysis classification is positive, negative and neutral. In this research we focus on the reviews for electronic products. The demographic and technical expertise level of consumers and reviewers are diverse hence there is difference in the way they review a product. Some review contains technical solutions, improvised ways of tackling problems, frustrations, joy etc. This differences in review calls for a wider classification scheme to contain these differences. We thereby introduced a five classification scheme namely positive, negative, advice, no sentiment and neutral at the sentence. We crawled data from amazon.com and used open source natural language processing tools to get the sentiment out of the review.

***Keywords**: Sentiment analysis, Machine learning, Naïve Bayes, Support vector machine (SVM), Corpus, Ling Pipe Classifier*

1. Introduction

The ever increasing use the Internet for e-commerce has produced vast large consumer generated contents such as reviews. It is becoming imperative for organizations to collect these data and useful insight can be mined from the reviews. As prices of storage space is decreasing it serves as the right step to collect review from web for analysis to know what users are saying about a particular product. Organization can use these user feedbacks to make their upcoming products better or improve their customer service. According to a study of 1046 individuals who have had experiences with the customer service of a mid-sized company, The vast majority of participants who have seen reviews claimed that that information did impact their buying decisions. This was true of both positive reviews (90%) as well as negative reviews (86%) [14].

Online reviews is now the extension of the conventional word of mouth approach used by consumers to recommend products to people they know. Before online reviews word of mouth reach is a little beyond the number of acquaintances a consumer knows. Amazon let users write reviews and indicate their satisfaction level using a 5 star rating. There is a place for identify if they are verified buyer on amazon and also their location. Walmart goes a step further by specifying the age group, usage duration, gender and who

* This paper summarizes a part of a master's thesis
[4] Corresponding author.

ISSN: 1738-9984 IJSEIA

recommended the product [16]. TigerDirect takes a different approach by incorporating rating for value, features, quality and performance [17].

Rating of user depends on the overall performance and experience. Some reviews are very short and some are very verbose covering hundred plus lines. The more detailed review shows the technical knowledge of the reviewer. As there is not a strict language system in reviewing, a lot of users use slangs and texting words and emoticon.

Due to the popularity of reviews there has been a lot of Sentiment Analysis and text mining applied to reviews such as movies [3, 7]. These techniques usually places into three categories namely positive, negative and neutral. The positive is favourable experience, negative for bad experience and neutral is no sentiment. However analyzing reviews on the sentence level, the sentences can be placed in more than three categories. Users with more technical knowledge gives more information such as how to circumvent problems face, recommendation to improve a product. Some sentences just give background information of where the product were bought.

In this paper we will focus on consumables. We will apply a five classification scheme namely recommendation or advice, negative, positive, neutral and no sentiment to these sentences. Our primary data is from Television reviews of LG and Samsung.

2. Literature Review

Sentiment analysis has gained much popularity in recent times due to the abundance of online reviews and corporations are anxious to find out what consumers are saying about their products. There is also a corresponding amount of research on sentiment analysis.

In literature [3], Pang *et al.*, used movie reviews as data, and machine learning algorithms such as Support Vector Machine (SVM), Naïve Bayes and maximum entropy classification to train a model which far out performed human- produced baselines. They applied a classifier to detect where a movie review is positive or negative. The Naïve Bayes highest score was 81.5% accuracy, Maximum entropy highest score was 81.0% accuracy and SVM highest was 82.9% accuracy.

In 2004, literature [7] improved on their first paper [3] by using efficient techniques for finding minimum cuts in graphs in subjective text. The technique labels the sentences as either subjective or objective and selecting the subjective text and applying standard machine learning classifier to it. This technique achieved a higher accuracy as compared with their paper in 2002.

In literature [1], they formulated the HL-SOT algorithm, which was developed based on generalizing an online-learning algorithm H-RLS [5]. This can be generalized to make it possible to perform sentiment analysis on target texts that are a mix of reviews of different products.

3. System Architecture

In this section, we describe system architecture and steps in the data life cycle for the sentimental analysis. Figure 1 shows data source (amazon.com), storage (MySQL), text analysis and mining, and visualization of the result on the Web.

Figure 1. System Architecture for the Sentimental Analysis

In the next section, we formally propose the five classification scheme, feature detection, and geotagging the reviews.

3.1. Classification Scheme

Online reviews are usually classified by the reviewer so we will rather classify the sentence level. A 5 star review can also have some sentences which expresses a bad experience using a product. By reading through thousands of reviews some patterns of sentences emerged rather the three broad classification system. Consider the following sentences:

1. This TV is amazing and I'm glad I got it.
2. You will need a flat surface that is about 27 inches wide for this 32 inch TV.
3. The picture is excellent, however, the volume seems a bit muffled.
4. I was looking at a $600 Samsung with 120 Hz refresh rate, then I found this one.
5. The television cuts on, cuts off, randomly.
6. The machine will still rotate your clothes every so often to avoid wrinkles.

The above sentences show that reviews can be more than positive and negative and neutral. Hence we introduce a five class scheme namely positive, negative, advice, neutral and no sentiment. The first and sixth sentence will be placed in the positive category as they clearly state a positive feeling, the second would be place as an advice as it shows what surface the TV should be placed on. The third sentence has both positive and negative sentiment hence it would fall in the neutral category. The fourth would be placed in no sentiment category as there is no mood attached to it. Finally the fifth would be placed in the negative category as the reviewer wrote about a negative experience. Once we have our classification scheme set up we will proceed to the data collection.

(1) **Data Collection.**

We created a crawler to get the review from amazon.com using Jsoup API [19]. We collected data from the electronic section. We used the television, washing machines, refrigerators, notebooks and monitors products from Samsung and LG as our primary source of data. The information collected were the review text and the location of the reviewer. The crawled data was put in a MySQL database for later use.

(2) **Sentence Extraction**

As our focus was on the sentence level we had to split all the reviews into sentences before going to the next phase. We created a mentioned table for this.

(3) **Location Extraction**

On amazon.com, reviewers are given an option to state their location in the location field. The location can be the state, country, county. We used this location data when stated, as an input to get the geographical location of the review. However the data input from the reviewer does not conform to a standard. For example California was represented as CA, Cali and Calif *etc*. Atlanta is usually written as ATL, New York as NYC or NY, Washington DC as Dc. There are a lot of abbreviations, come conform to the approved state naming convention. To overcome this problem we performed some data cleansing to standardize the input to query for. This is done before the data is stored in the database.

(4) **Supervised Learning.**

Reviewers are not under any obligation to adhere to strict grammar rules hence the reviews sentences are unstructured. Many use emoticons, slangs or bad grammar. In one review the reviewer was so enthused about the product that she wrote “'Very goooood ~It made my husband happy.'” There were numerous spelling errors. This compelled us not to use the grammar context sentiment analysis and the SENTIWORDNET [11] approach. Also is very had to detect product reviews using grammar context as different products have different criteria to be described as good. For example a quiet washing machine or notebook is a positive sentiment whiles a quite TV or audio system is a negative sentiment. Additionally there are schemes to detected negative and positive but no scheme to detect recommendations are advices. These reasons compelled us to opt for a machine learning algorithm. We used LingPipe’s Classifier API to train our classifier.

3.2. Feature Detection

There are many researches on extracting the topics of a review. In literature [13], Titov *et al.*, formulated a statistical model which is able to discover corresponding topics in text and extract textual evidence from reviews. In literature [4], Zhai, Z., *et al.*, have developed a semi-supervised technique to identify clusters of features, *i.e.*, sets of synonyms that are likely to refer to the same product features. Their technique needs a fixed number of clusters, with a starting set of features. Our technique is a little similar of Zhai, Z., *et al*. In our approach we started with a list of product features. We used WordNet to get the Synset from this features to get more words. For each extracted sentence we converted it into lowercase before we used Stanford Part-Of-Speech (POS) tagger. Stanford POS tagger detects uppercased words as a noun, so in the case where a reviewer is shouting by typing all sentences in capital cases, the tagger might see all the words as nouns. The tagged nouns are looked up in the feature list created for validation. We also incorporated compound nouns detection, which is usually a feature in a feature. In case a compound noun is detected and a part can be found in the feature list it automatically detected as a feature.

Feature Detection Algorithm

1. Select relevant features of a product and search for the synset in WordNet.
2. Convert sentences into lowercase.
3. Parse sentence into POS tagger
4. Detect compound nouns
5. Detect single word nouns
6. If single word noun in detected compound word delete the noun
7. Else if single word noun not in detected compound noun add to the feature detection list.
8. Return the compound nouns and other nouns

3.3. Web Application

Having data in the database is not enough to get meaning from collection. To complete the process, visualization of the data is needed. There have been several attempts to do this. Liu *et al.*, [6] proposed a framework for analyzing and comparing consumer opinions of competing products where users can see the performance of competition products. Oelke *et al.*, [9] proposed a scalable alternative in order to aggregate large numbers of products and features, clustering similar users. Similarly, the system by Miao *et al.*, [8] visualizes the sentiment expressed in product reviews over time. Positive and negative opinions were aggregated over time and displayed with different charts. Wu *et al.*, [15] proposed the OpinionSeer, for reviews where uncertain sentiment through time is visually represented and aggregated. In this paper we present an application to visualize the aggregated data.

4. Experiment Results

In this research we collected full text reviews and tokenized the reviews into sentences to analyze them and get the sentiment. In this section we will discuss the outcome of our experiment.

4.1. Web Crawler

The web crawler designed collected the data from the site and stored them into the database. Table 1 summarizes the total data collected. Note that LG had no notebook with review.

Table 1. All Data Collected

Product	Total Reviews	Samsung	LG
Television	35571	28778	6793
Monitor	1179	1010	161
Notebook	6013	6013	0
Washers	275	110	568
Refrigerator	189	133	56

4.2. Sentence Extraction

This is one of the main task in this research as we are interested in getting the sentence by sentence sentiment of a review text. To undertake this task we used LingPipe Sentence Boundary detector. This process was done after the crawler completed its task. Many reviewers do not write properly usually, hence the sentence boundary detected cannot detect the boundary of sentences which does not adhere to grammar rules. Table 2 summarizes the total number of sentences in a mentioned table for each category

Table 2. Sentences Extracted

Product	Total Reviews	Samsung	LG
Television	288230	28778	6793
Monitor	8428	1010	161
Notebook	68371	6013	0
Washers	3433	110	568
Refrigerator	2333	133	56

4.3. Location Extraction

We were able to retrieve the location of the place mentioned by creating a database with all cities in the USA and querying against it to get the geolocation. As we were only interested in only cities in United States other location outside could not be found. Table 3 shows the data collection success rate.

Table 3. Locations Extracted

Product	Total Non-Empty	Location Found	Success Rate
Television	15664	13117	83.7
Monitor	493	397	80.5
Notebook	1991	1641	82.4
Washers	127	106	83.5
Refrigerator	73	63	86.3

4.4. Sentiment Classification

Having identified the classification of reviews and choosing supervised learning to train our data, we used Ling Pipe Classifier API to build our train our classifier. The API was able to get results closer to results achieved by Bo Pang, *et al.* [3]. We used a sample data from our database as the training data. We wrote manually read the sentences and put them in their correct class. After we use the Classifier API to train the model. Table 4 gives a detailed account of the training dataset. We did not train the classifier with data from monitor review sentence as televisions and monitors are similar. Classifier is employed during the sentence extraction phase before storing it in the database. Figure 2 gives a detail of the accuracy (70%) of our classifier when used on random dataset.

Table 4. Data Trained

Product	POS	NEG	ADV	No sent	NEU	Total
Television	1,577	1,057	311	579	316	3840
Notebook	961	447	234	763	72	2,477
Washers	256	200	92	265	14	827
Refrigerator	216	338	53	63	13	875

```
BASIC POLARITY EVALUATIION NGRAMS OF CUSTOM SET 8

Data Directory=C:\Users\Bryanic\workspace\pySentiment\neg\hand

Evaluating.
  # Test Cases=426
  # Correct=299
  % Correct=0.7018779342723005
pos
THe Custom copus is pos
```

Figure 2. Basic Classifier Evaluation

4.5. Web Application.

We chose Vaadin a variant of Google GWT and also opens source. We used Vaadin's charts, tables and a google Map plugin by Tapio to in the visualization of the data stored in the database. It is divided into three parts, namely dashboard, category search and the sentiment map. The dashboard (Figure 3) gives a quick view of the aggregation of reviews over time, the products with the top review, proportion of star rating for each brand in each product category. The category section is to check for product and get their review text and each sentence sentiment analysis (Figure 4). The sentiment map (Figure 5) shows the map of USA to map out location with a search feature to map out case counts of searched feature and sentiment. The markers are place at each state capital to show the number of cases. The results showing the sentiments are editable to train the classifier when the sentiment classification is wrong. This helps trains the classifier to be more accurate.

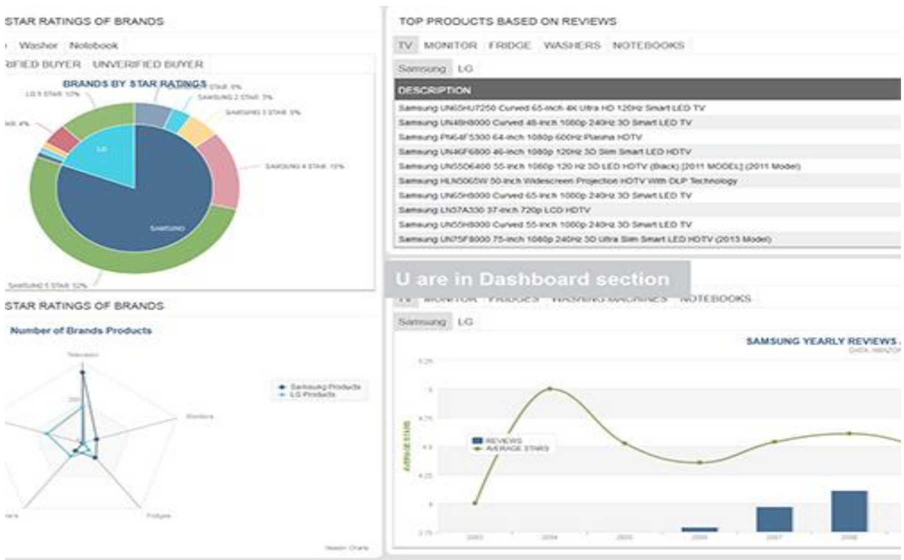

Figure 3. Dashboard showing the Year to Year Average Rating of the Manufacturers, Top Products based on Average Review Points, Proportion of Stars Rating of Manufacturers

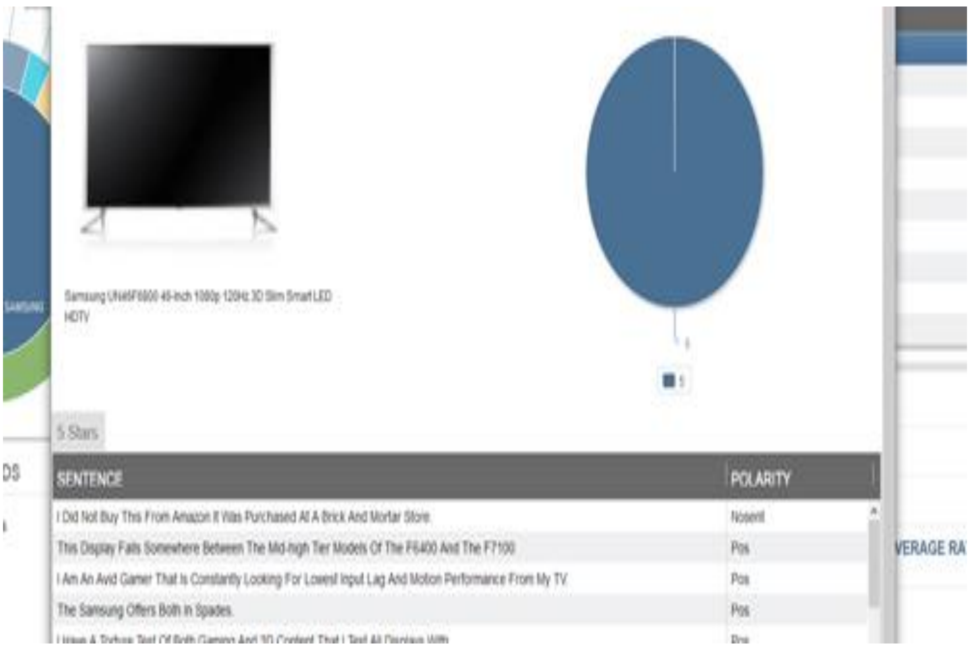

Figure 4. A Window showing a Product Sentences Sentiments

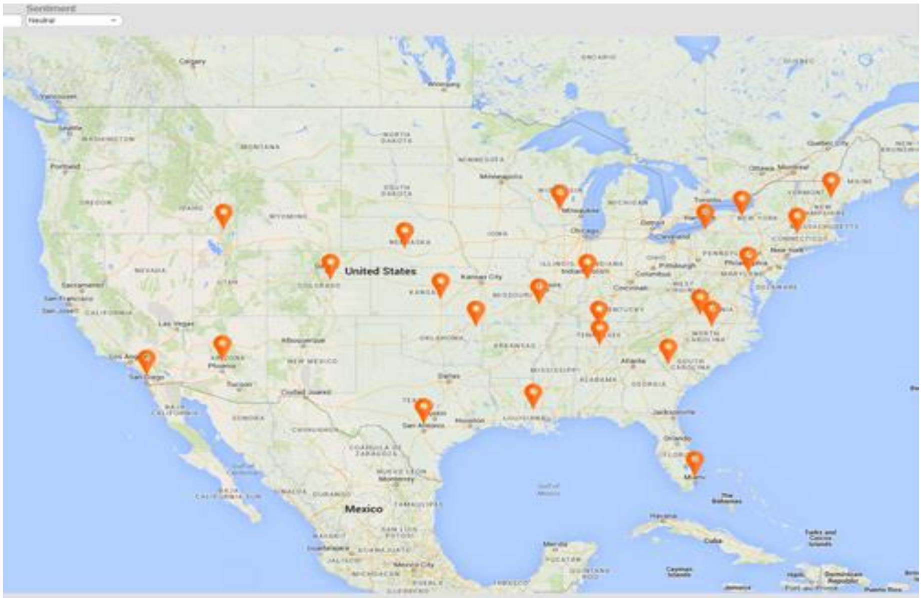

Figure 5. Sentiment Map showing Cases of Neutral Search Feature Sentiment

5. Conclusion

In this paper, we presented a novel five classification scheme and feature level at the sentence level of review collected from amazon.com. This approach was undertaken using wholly open source tools. It integrates natural language processing techniques to get the results. We also showed how we got the locations to the reviews for the sentiment map and the data for the training of our classifier. The disproportionate number of training set for each class made the classifier perform better for positive, negative and no sentiment classes and very poor for the neutral and advice classes. Also the lack of strict grammar made caused some reviews not to be tokenized properly. Future work should focus on training the classifier more and making the distribution proportionate to maximize the efficiency.

Acknowledgements

This research was supported by the MSIP(The Ministry of Science,ICT and Future Planning), Korea, under the "SW master's course of a hiring contract" support program (NIPA-2014-HB301-14-1011) and ITRC(Information Technology Research Center) support program (NIPA-2014-H0301-14-1022) supervised by the NIPA(National IT Industry Promotion Agency).

References

[1] W. Wei and J. Atle Gulla, “Sentiment Learning on Product Reviews via Sentiment Ontology Tree”, Proceedings of the 5th International Joint Conference on Natural Language Processing, AFNLP, **(2011)**, pp. 327–335.

[2] M. Cataldi, B. Andrea, T. Ilaria and M.-A. Aufaure, “Good Location, Terrible Food: Detecting Feature Sentiment in User-Generated Reviews”, Social Network Analysis and Mining, Springer, vol. 3, no. 4, **(2013)**, pp. 1149-1163.

[3] B. Pang, L. Lee and S. Vaithyanathan, “Thumbs up? Sentiment Classification using Machine Learning Techniques”, EMNLP Proceedings, **(2002)**.

[4] Z. Zhai, B. Liu, H. Xu and P. Jia, “Clustering product features for opinion mining”, Proceedings of the 4th ACM International Conference on Web Search and Data Mining, ACM, **(2011)**, pp. 347–354.

[5] N. Cesa-Bianchi, G. Claudio and L. Zaniboni, “Incremental Algorithms for Hierarchical Classification”, Journal of Machine Learning Research, vol. 7, **(2006)**, pp 31–54.

[6] B. Liu, M. Hu and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web", Proceedings of the 14th International Conference on World Wide Web, ACM, WWW, **(2005)**, pp. 342–351.
[7] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", ACL Proceedings, **(2004)**.
[8] Q. Miao, Q. Li and R. Dai, "AMAZING: A sentiment mining and retrieval system", Expert Systems with Applications, vol. 36, no. 3, **(2009)**, pp 7192–7198.
[9] D. Oelke, M. C. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, L. E. Haug and Janetzko, "Visual opinion analysis of customer feedback data", IEEE VAST, **(2009)**, pp. 187-194.
[10] J. Murach, Murach's MySQL, **(2012)**.
[11] S. Baccianella, A. Esuli and F. Sebastiani, "SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining", **(2010)**.
[12] S. Siersdorfer, "How Useful are Your Comments? - Analyzing and Predicting YouTube Comments and Comment Ratings", Proceedings of the 19th international conference on World wide web, **(2010)**.
[13] Titov and McDonald, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization", Proceedings of ACL-08: HLT, **(2008)**, pp. 308–316.
[14] Dimensional Research, "A Survey of Customer Service From Mid-Size Companies", **(2013)**, pp. 10.
[15] W. Y1, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou and H. Qu, "OpinionSeer: interactive visualization of hotel customer feedback", IEEE Transactions on Visualization and Computer Graphics, vol. 16, no. 6, **(2010)**, pp. 1109-1118.
[16] Walmart, http://www.walmart.com/.
[17] TigerDirect, http://www.tigerdirect.com/.
[18] Nicolas Fränkel, Learning Vaadin,2nd Edition, **(2014)**.
[19] Jsoup API Cookbook, http://jsoup.org/cookbook/.
[20] Vaadin Book, https://vaadin.com/book.
[21] WordNet, http://wordnet.princeton.edu/.
[22] Stanford Part-Of-Speech Tagger, http://nlp.stanford.edu/software/tagger.shtml.

Authors

Bryan Nii Larte Bryan, had his Bachelors degree in Information Studies and Geography from the University of Ghana. He is currently in his last semester of his master's degree from Chungbuk National University. His main interests are database systems, micro controller programming, distributed computing and open source technologies.

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.

In-Sun Jung, received BA degree in Chonbuk national university division of international trade. She is currently studying toward the Ph. D. degree in Management Information System from Chung-buk National University.

Her research interests include database systems, Bigdata, data mining, business intelligence, OLAP, Social Network, data Quality.

Kyung-Hee Lee, received the Ph. D. degree in computer science from Chungbuk National University in 2004. From 2008 to now, she serverd as a research engineer in the Business Data Covergence of Chungbuk National Univeristy.

Her research interests include database systems, bigdata analysis algorithm and technology, data mining and distributed computing platform.

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.