Journal of information and communication convergence engineering
http://jicce.org

J. Inf. Commun. Converg. Eng. 18(1): 55-60, Mar. 2020 Regular paper

# Product Recommendation System based on User Purchase Priority

**Jinsuk Bang[1], Doyeun Hwang[2], and Hoekyung Jung[2*]**, ***Member, KIICE***

[1]Department of Computer Engineering, Hoseo University, Cheonan 31066, Korea
[2]Department of Computer Engineering, PaiChai University, Daejeon 35345, Korea

## Abstract

As personalized customer services create a society that emphasizes the personality of an individual, the number of product reviews and quantity of user data generated by users on the internet in mobile shopping apps and sites are increasing. Such product review data are classified as unstructured data. Unstructured data have the potential to be transformed into information that companies and users can employ, using appropriate processing and analyses. However, existing systems do not reflect the detailed information they collect, such as user characteristics, purchase preference, or purchase priority while analyzing review data. Thus, it is challenging to provide customized recommendations for various users. Therefore, in this study, we have developed a product recommendation system that takes into account the user's priority, which they select, when searching for and purchasing a product. The recommendation system then displays the results to the user by processing and analyzing their preferences. Since the user's preference is considered, the user can obtain results that are more relevant.

**Index Terms**: Crawling, Product review, Purchase priority, Recommender system

## I. INTRODUCTION

As the amount of data generated on the internet increases, more information is being provided to users. This abundance of information, however, requires users to make a more concentrated effort while selecting products they require [1]. This problem is especially relevant to the field of e-commerce.

In particular, shared opinions on products through Twitter, blogs, and Facebook has influenced consumers regarding their purchase decisions. As a result, consumers make purchase decisions based on other users' purchasing history or product reviews [2]. To solve the issues relating to information overload, a system for recommending product information suitable to the individual is introduced through the findings of this study.

In recent years, several studies have attempted to distinguish between positive and negative opinions by incorporating emotional analysis into natural language processing or machine learning [3, 4]. Furthermore, a study was conducted using statistical analysis to evaluate a product using a user's score of the product, and frequency of characteristic vocabulary, among others [5-7]. Such a recommendation system provides users with a customized recommendation of products or services, and users can obtain information most relevant to them [8]. Furthermore, a technique that can be used for sentiment analysis is text classification. The text classification scheme consists of feature extraction and classification steps. Moreover, it proposes a new method that combines a filter-based global feature selection method and one-sided local feature selection method [9, 10].

Therefore, in this paper, we employ the purchase criteria,

**Received** 03 December 2019, **Revised** 12 February 2020, **Accepted** 13 February 2020
***Corresponding Author** Hoekyung Jung (E-mail: hkjung@pcu.ac.kr, Tel: +82-42-520-5640)
Department of Computer Engineering, PaiChai University, Daejeon 35345, Korea.

Open Access https://doi.org/10.6109/jicce.2020.18.1.55 print ISSN: 2234-8255 online ISSN: 2234-8883

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ⓒ The Korea Institute of Information and Communication Engineering

selected by the user, and product reviews, created by users who purchased the product, to determine products with an excellent evaluation of the purchase criteria, and consequently arrange the top ten products and recommend them to users. The user can select the purchase criteria to be considered, and the information can be customized for an individual by reflecting the selection criteria in the analysis. By using the obtained results for decision making, the user can enjoy a smoother and more efficient purchasing process.

## II. SYSTEM DESIGN

### A. Architectural Design

The data collection layer collects information concerning similar products that the user intends to purchase that includes the URL, product review, and purchase priority. The URL is a link to a web page that directly connects to the corresponding product's purchase page. In this paper, the information concerning the goods refers to the unique characteristics of the products, such as their name, seller, color, and price. In the case of product reviews, articles posted on the list of reviews that can be seen at the bottom of each product purchase page (URL) is referenced.

This article refers to the subjective evaluation of the product by a user who has purchased or used the product. The purchase priority refers to an attribute that affects purchase decisions when a user purchases the product. For example, the mode of delivery could be express or regular, which refers to the speed of delivery. A more detailed explanation is provided in a later section.

In the data storage layer, three tables are managed as a single database called review_db, and the data are fetched in conjunction with Python via MySQL. The name and URL of a product corresponding to a product's information are stored in the product list table, and the product review is stored in the review list table. In the data storage layer, only those product reviews that are to be used in the analysis process are stored in the key review table after preprocessing the product list and review list table.

In the data processing layer, the imported product review data are separated by sentence, and morphological analysis is performed using the Twitter morpheme analyzer. When the morphological analysis is completed, a part-of-speech (POS)-tagger extracts adjectives and verbs that can be further used for emotional analysis.

The reason for extracting adjectives and verbs is that emotions are expressed as affirmatives or negatives because adjectives or verbs that describe the subject (product) correspond to this behavior. Based on the extracted morpheme, positive and negative emotions are analyzed, and the sentence sensitivity score is calculated. Based on the calculated score, a list of ten products is generated and ranked. The user then receives the personalized recommendation information.

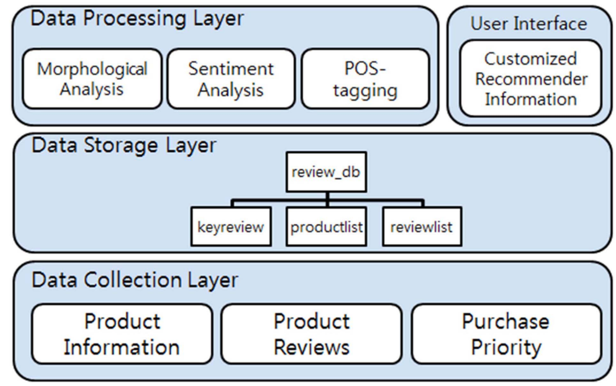


**Fig. 1.** System architecture.

Fig. 1 shows the architecture of the user priority-based product recommendation system proposed in this study.

### B. Data Analysis

Pre-processing is performed prior to full-scale analysis. Since product review data are unstructured data, proper analysis can be conducted smoothly if preprocessing is performed. Preprocessing comprises sentence separation, morphological analysis, and POS tagging.

Through such preprocessing, a small amount of erroneous data, which may degrade the performance of the algorithm, is filtered, and the analysis can be performed efficiently. In addition, the data can be converted into a standard form of data and can easily be managed.

Fig. 2 shows the data analysis process. Preprocessing is performed prior to full-scale analysis. Since product review data are a type of unstructured data that are somewhat challenging to structure, proper analysis can smoothly proceed if preprocessing is performed.

Firstly, the data stored in the DB are retrieved by means of a crawl, and preprocessing is performed. The product reviews containing the user-selected purchase priority criteria used for emotional analysis are imported from the review list table. The sentences are separated based on the punctuation, and morphological analysis is performed using the Twitter morpheme analyzer to extract adjectives and verbs that express emotion.

Secondly, the preprocessed review data are stored in the key review table. We conduct emotional analysis using nouns, verbs, and adjectives extracted in the previous step. Here, the noun is a purchase priority criterion. That is, +1 is counted as a positive value for a purchase priority criterion, and -1 is counted as a negative value. If a plurality is selected, a weight value is calculated.

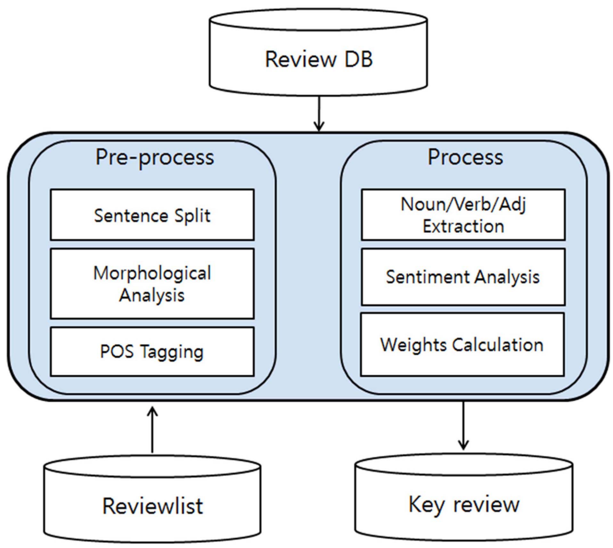



**Fig. 2.** Process of review data analysis.

## III. SYSTEM IMPLEMENTATION

### *A. Crawling*

The experiment was performed using Python. We selected "electric heater" from seasonal household appliances under the category of domestic shopping malls and crawled the information and product reviews associated with the product. In our previous research, a "baby stroller" under the category of childcare articles was selected as the collection target. However, there were limitations to the data collected given that strollers are usually purchased by individuals raising children. Thus, almost no comments were made by men, and it mostly consisted of reviews written by relatively young women in their 20s and 30s. Therefore, in this experiment, the selected product influenced the quality of the product review to create a variety of types of consumers, both young and old. In addition, the site targeted for collection showed a threefold increase in sales volumes as compared to the previous year. This implied that relatively recent review data were collected. As a result of the crawl, 2,000 reviews of approximately 40 items were collected.

The BeautifulSoup web crawler was used to collect product names, product URLs, and product reviews for items listed under the header of "electric heaters." BeautifulSoup is a Python library that extracts data from HTML or XML. We used a parser to determine the structure of a web document and to extract the relevant values by locating the data to be searched. Therefore, the crawl was conducted by referring to the source written in HTML in the URL window of the previously stored product.

Since product reviews, which are qualitative information, are to be quantified and analyzed as quantitative information, an appropriate number of reviews should be collected. If the weight of a review is found to be below a certain level, it will cause an imbalance in the analysis and, thus, is discarded. Therefore, only products with more than twenty "Premium Merchandise Reviews," which consist of long reviews coupled with photographs, are selected and stored in the DB. In addition, when a certain number of reviews are identified, for an online shopping website, the crawl is performed using the Selenium tool because the reviews can only be viewed after a web browser action is performed within the page.

### *B. Pre-processing*

Prior to emotional analysis, preprocessing is performed to facilitate analysis. To calculate the score of the product's characteristics, it is necessary to obtain the score of the sentence to which each characteristic belongs. First, the review data stored in the DB are separated into sentences based on the punctuation and stored in the form of a text file. Next, special characters are removed, including stopwords.

In this paper, the adjectives describing the purchase criteria are considered in combination. For example, if the user needs a low-priced product, keywords such as "cheap" or "low" are added to the keywords to extract the keywords from the review data. By taking adjectives together with nouns, users can more accurately identify products that meet the desired purchase criteria and filter unnecessary data.

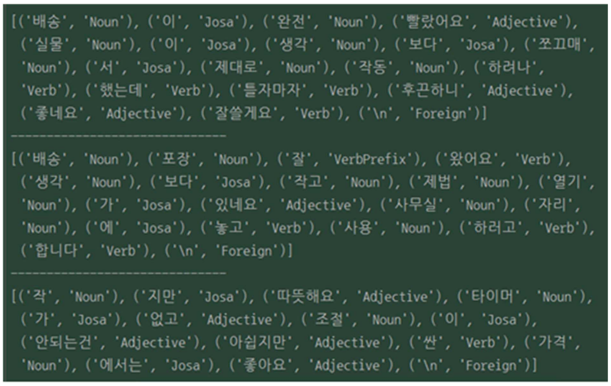
The next step involves POS-tagging through morphological analysis. The KoNLP package is installed and stemming is conducted based on its Twitter class. KoNLPy stands for Korean Natural Language Processing in Python. It is a Python package for processing Korean language information and is necessary for processing and analyzing and available as a bundle of the morpheme analyzer.

The Twitter morpheme is an open-source Korean tokenizer that produces results that are more concise and understandable and has approximately 20 tags, including nouns, verbs, and adjectives. The text file stored in the previous step is read and stored in a list form, and torque-by-sentence processing is performed. Sentences that do not contain adjectives that can judge affirmative or negative in a sentence from a review are excluded from analysis because they are deemed as noise.

Fig. 3 shows the result of stemming product reviews by selecting "price" as the first rank and "shipping" as the second rank.

### *C. Positive and Negative Scoring*

To generate positive and negative product recommendation lists, emotional analysis is performed using "Adjectives" or "Verbs" in which sentences containing the first and second purchase criteria are evaluated with respect to the purchase



```
[('배송', 'Noun'), ('이', 'Josa'), ('완전', 'Noun'), ('빨랐어요', 'Adjective'),
 ('실물', 'Noun'), ('이', 'Josa'), ('생각', 'Noun'), ('보다', 'Josa'), ('쪼끄매',
 'Noun'), ('서', 'Josa'), ('제대로', 'Noun'), ('작동', 'Noun'), ('하려나',
 'Verb'), ('했는데', 'Verb'), ('틀자마자', 'Verb'), ('후끈하니', 'Adjective'),
 ('좋네요', 'Adjective'), ('잘쓸게요', 'Verb'), ('\n', 'Foreign')]
------------------------------
[('배송', 'Noun'), ('포장', 'Noun'), ('잘', 'VerbPrefix'), ('왔어요', 'Verb'),
 ('생각', 'Noun'), ('보다', 'Josa'), ('작고', 'Noun'), ('제법', 'Noun'), ('열기',
 'Noun'), ('가', 'Josa'), ('있네요', 'Adjective'), ('사무실', 'Noun'), ('자리',
 'Noun'), ('에', 'Josa'), ('놓고', 'Verb'), ('사용', 'Noun'), ('하려고', 'Verb'),
 ('합니다', 'Verb'), ('\n', 'Foreign')]
------------------------------
[('작', 'Noun'), ('지만', 'Josa'), ('따뜻해요', 'Adjective'), ('타이머', 'Noun'),
 ('가', 'Josa'), ('없고', 'Adjective'), ('조절', 'Noun'), ('이', 'Josa'),
 ('안되는건', 'Adjective'), ('아쉽지만', 'Adjective'), ('싼', 'Verb'), ('가격',
 'Noun'), ('에서는', 'Josa'), ('좋아요', 'Adjective'), ('\n', 'Foreign')]
```

**Fig. 3.** Result of using Twitter morphological analyzer.

priority. For example, if a sentence reads, "shipping is really fast," a +1 is assigned for a positive priority concerning the word "fast." Conversely, if negative expressions are made, the total score is calculated by subtracting 1 point.

When multiple criteria are selected, such as "price" and "delivery," each standard should be calculated taking into account the priority and weight.

First, we need to determine the sensitivity score of a simple review sentence. A review containing the keyword "price" in the overall review is called $PR = \{pr_1, pr_2, ..., pr_n\}$, a positive and negative rating for the review and the final review score for the purchase-based "price" is PS (PriceScore) calculated using eq. (1) below:

$$KR_1 = \frac{PS}{PR} \tag{1}$$

Likewise, in the case of the second purchase priority "delivery," the review including the keyword is $DR = \{dr_1, dr_2, ..., dr_n\}$, the positive and negative scores for the review DS (DeliveryScore), and the final review score of the keyword "delivery" can be calculated by the following eq. (2):

$$KR_2 = \frac{DS}{DR} \tag{2}$$

The sensitivity score calculated above can be multiplied by the weight, summed, and expressed as a percentage, and the weight for each purchase priority can be obtained. The higher this percentage, the higher the probability of the product being ranked at the top of the product recommendation list. Finally, the review sentiment score (RSS) can be expressed by the following formula (3):

$$RSS = \frac{(KR_1 \times W_1) + (KR_2 \times W_2)}{100} \tag{3}$$

Table 1 shows the list of top ten products with high scores as a result of the emotional score calculation. This is the result of calculating the score by "price" and "shipping" purchase criteria.

**Table 1.** Result of sentiment score calculation

| Ranking | Product Name | Price Sentiment Score | Delivery Sentiment Score | Total Score |
|---|---|---|---|---|
| 1 | 2018 NEW Shinil SEH-G800 | 76.0 | 18.0 | 94.0 |
| 2 | Deawoong multi heater | 72.7 | 20.45 | 93.2 |
| 3 | Swiss DWH-PTC4021M | 81.8 | 9.05 | 90.9 |
| 4 | Daewoong CZ-HT7007C | 59.2 | 29.6 | 88.8 |
| 5 | Batom carbon heater | 66.6 | 22.2 | 88.8 |
| 6 | stand hearter 1490H | 76.9 | 11.5 | 88.4 |
| 7 | Shinil SEH-G800 | 55.5 | 25.0 | 80.5 |
| 8 | Daewoong CZ-6500KH | 62.5 | 15.0 | 77.5 |
| 9 | WINDPIA-1580HS-X2 | 50.0 | 22.9 | 72.9 |
| 10 | Black&Daeker BXSH1801-A | 40.0 | 32.5 | 72.5 |

## IV. ANALYSIS AND DISCUSSION

There have been numerous studies, including a method of measuring the reliability of product reviews [11] and a method of calculating the similarity among users [12].

However, we believe it is crucial to grasp the needs of users before making a recommendation to users. If a recommendation is made without understanding the user's requirements or preferences, the purpose of recommendation systems will not be fully achieved.

Therefore, this paper differentiates the proposed recommendation system from the existing recommendation systems by directly taking into account the purchase priority that influences the decision to purchase a commodity.

As shown in Table 1, we can calculate the weight of a review by dividing the score calculated above by the total number of product reviews. This demonstrates the share of the purchase criteria in the product. In addition, the effect of the weighting on the creation of the product list is observed.

Fig. 4 shows the percentage of purchasing criteria for the top 15 products ranked based on positive and negative scores.

Even though product 11 has a higher weight than the top ten products, it is not ranked in the top ten. The change in the weight value causes this following the ranking of the purchasing criteria. That is the ranking of the final recommendation list changes depending on the weight of the purchase order 1 and 2.

Fig. 5 is a graph showing the price of the top 15 products and the percentage of each purchase criteria for shipping.

Despite the high ratio in Fig. 4, when comparing the 11th ranked product and the top 5 ranked products, despite the low ratio, the emotional score of the "price," the ratio is remarkably high.

That is, since the weight of the second rank is calculated as half the weight of the first rank, the resulting ratio of the

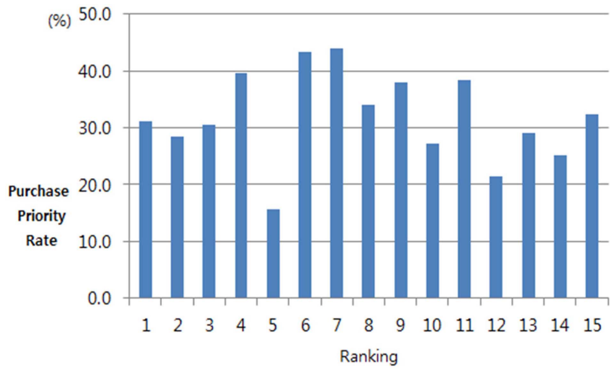



**Fig. 4.** Purchase priority ratios for the top 15 products.

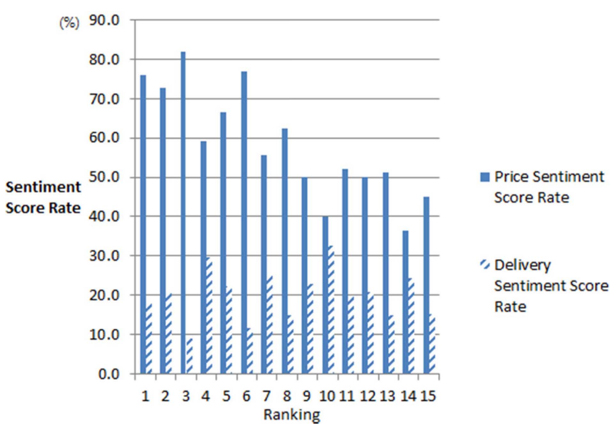



**Fig. 5.** Average error graph including conditions.

entire review is larger but is not ranked as the highest rank. It can be confirmed that the recommendation rank has changed according to the priority of the purchasing criteria selected by the user. This result implies that individual recommendations of personalized products can be made by reflecting the user's preferences and purchase propensity.

## V. CONCLUSIONS

Most consumers, including young individuals (10s and 20s), see product reviews or reviews left by other consumers, and incorporate them into their purchasing decisions. However, the effect of information overload has caused users to spend a significant amount of time and effort in purchasing products. Furthermore, the criteria for selecting products are different. Even though consumers have different purchasing criteria, it is also a problem to receive the same product list from the Internet shopping mall. The ultimate goal of a product recommendation system is to help users make purchasing decisions by recommending products they require at the right place and time. It is necessary to understand why consumers purchase a particular product and take into account their hidden intentions to improve user satisfaction.

Therefore, in this study, the user selects the purchase priority, which is then analyzed, thereby supplementing the existing inefficient purchase decision process in which the product information is consistently provided. The data concerning the reviews of a product are collected by crawling and then stored in a database. Preprocessing is performed by separating the data by sentence and performing morphological analysis. Then, the emotional score is calculated by reflecting the purchase priority selected by the user. At this time, the purchasing intention of the user is appropriately reflected by changing the weights of the first and second order.

The product list reflecting the information of the product, and the purchase priority are rearranged in the order of the emotional score of the product review and provided to the user. This not only reduces the amount of time it takes to purchase a product but also provides customized recommendation information because it reflects the purchase criteria of each user.

In addition, the list of products that have undergone the filtering process is rearranged in the order of the emotional score of a product review. An excellent emotional score implies a high positive score. Therefore, it is possible to increase the satisfaction of the user by allowing the user to quickly receive the product information according to the desired purchase priority without having to read the product list and a large number of product reviews.

## ACKNOWLEDGMENTS

This study was supported by the research grant of PaiChai University in 2020.

## REFERENCES

[ 1 ] A. A. Lenskiy and E. Makita, “Bayesian approach to users’ perspective on movie genres,” *Journal of Information and Communication Convergence Engineering*, vol. 15, no. 1, pp. 43-48, 2017.
[ 2 ] H. G Lee and H. Kwak, “Investigation of factors affecting the effects of online consumer reviews,” *Journal of National Information Society Agency*, vol. 20, no. 3, pp. 3-17, 2013.
[ 3 ] A. Esuli and F. Sebastiani, “Determining term subjectivity and term orientation for opinion mining,” in *Proceedings of 11th Conference of the European Chapter of the Association for Computational Linguistics: EACL*, pp. 193-200, 2006.
[ 4 ] N. Jindal and B. Liu, “Mining comparative sentences and relations,” in *Proceedings of the 21st National Conference on Artificial Intelligence*, vol. 2, pp. 1331-1336, 2006.
[ 5 ] X. Ding, B. Liu, and P. S. Yu, “A holistic lexicon-based approach to opinion mining,” in *Proceedings of the international Conference on Web Search and Web Data Mining*, vol. 1, pp. 231-240, 2008.

DOI:10.1145/1341531.1341561.
[ 6 ] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web," in *Proceedings of the 14th International Conference on World Wide Web*, vol. 1, pp. 342-351, 2005. DOI: 10.1145/1060745.1060797.
[ 7 ] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima, "Mining product reputations on the web," in *Proceedings of the 8th SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 341-349, 2002. DOI: 10.1145/775047.775098.
[ 8 ] B. Xiao and I. Benbasat, "E-commerce Product Recommendation Agents: Use, Characteristic, and Impact" MIS quarterly, vol. 31, no. 1, pp.137-209, 2007. DOI: 10.2307/25148784.
[ 9 ] X. F. Wang and H. C. Kim. "New feature selection method for text categorization," *Journal of Information and Communication Convergence Engineering*, vol. 15. no. 1, pp. 53-61, 2017
[10] J. Y. Choeh, S. K. Lee, and Y. B. Cho, "Applying rating score's reliability of customers to enhance prediction accuracy in recommender system," *Journal of The Korea Contents,* vol. 13, no. 7, pp. 379-385, 2013. DOI:10.5392/jkca.2013.13.07.379.
[11] M. J. Ku and H. C. Ahn, "A hybrid recommender system based on collaborative filtering with selective use of overall and multicriteria ratings," *Journal of Intelligence and Information Systems*, vol. 24, no. 2, pp. 85-109, 2018.
[12] J. Y. Kim, D. Y. Hwang, and H. K. Jung, "Product recommendation system based user purchase criteria and product reviews," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 6, pp. 5454-5462, 2019. DOI:10.11591/ijece.v9i6.

**JinSuk Bang**

JinSuk Bang received her M.S. and Ph. D. degrees in 2002 and 2011, respectively, from the Department of Computer Engineering at PaiChai University, Korea. From 2010 to 2014, she worked for TopTech as an engineer, and from 2014 to 2020, she has worked in the Department of Computer Engineering at Dasom Information. Since 2020, JinSuk has worked in the Department of AI Convergence at Hoseo University, where she now works as a teacher. Her current research interests include software engineering, big data, deep learning, and SmartFactory.

**Doyeun Hwang**

Doyeun Hwang received her B.S. degree from the Department of Business Administration, PaiChai University, Korea, in 2017. She is currently studying in the Department of Computer Engineering at PaiChai University. Her current research interests include text mining, big data, and opinion mining.

**Hoekyung Jung**

Hoekyung Jung received his M.S. and Ph. D. degrees in 1987 and 1993, respectively, from the Department of Computer Engineering, Kwangwoon University, Korea. From 1994 to 1995, he worked for ETRI as a researcher. Since 1994, he has worked in the Department of Computer Engineering at PaiChai University, where he now works as a professor. His current research interests include multimedia document architecture modeling, machine learning, IoT, big data, and artificial intelligence.