

Offline-to-Online Service and Big Data Analysis for End-to-end Freight Management System

Suganya Selvaraj*, Hanjun Kim*, and Eunmi Choi**

Abstract

Freight management systems require a new business model for rapid decision making to improve their business processes by dynamically analyzing the previous experience data. Moreover, the amount of data generated by daily business activities to be analyzed for making better decisions is enormous. Online-to-offline or offline-to-online (O2O) is an electronic commerce (e-commerce) model used to combine the online and physical services. Data analysis is usually performed offline. In the present paper, to extend its benefits to online and to efficiently apply the big data analysis to the freight management system, we suggested a system architecture based on O2O services. We analyzed and extracted the useful knowledge from the real-time freight data for the period 2014–2017 aiming at further business development. The proposed system was deemed useful for truck management companies as it allowed dynamically obtaining the big data analysis results based on O2O services, which were used to optimize logistic freight, improve customer services, predict customer expectation, reduce costs and overhead by improving profit margins, and perform load balancing.

Keywords

Big Data Analysis, Freight Management System, O2O Service, System Architecture

1. Introduction

Recently, freight transportation management companies have been continually forced to make critical decisions seeking to create a highly productive transportation management system. There is a need for novel methods that would facilitate rapid decision making to improve transportation processes with the purpose of reducing costs and overhead, growing profit margins, load balancing, and improving customer service.

In big data, data analysis plays an important role in making better decisions for business development. Data analysis is a process of deriving useful information by evaluating the real-world raw data. Concerning freight transportation management companies, data analysis can be applied to assist both personnel and clients for making a better decision. In the present study, we analyze the real-time freight transportation data for the 4-year period (2014–2017) and categorize the results of the daily and distance-based trends. These results may include various information, such as frequency of requested goods vs.

※ 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.

Manuscript received October 2, 2018; first revision January 8, 2019; second revision July 8, 2019; third revision November 4, 2019; accepted November 25, 2019.

Corresponding Author: Eunmi Choi (emchoi@kookmin.ac.kr)

* Dept. of Financial Information Security, Kookmin University, South Korea (suganya@kookmin.ac.kr, jun8725@kookmin.ac.kr)

** College of Computer Science, Kookmin University, South Korea (emchoi@kookmin.ac.kr)

days, price vs. days, order status vs. days, 4-year comparison of requested goods vs. days, unit price vs. distance for all data items, unit price vs. distance for clusters based on the truck capacity, and unit price vs. distance for clusters considering weekdays and weekends. In the case of freight logistic companies, a freight rate and other trends depend on a requested day of the week. Therefore, we generate analysis results according to days of the week, which can be used by logistic companies to improve customer services by improving profit margins and load balancing. In the present study, we implement three steps in data analysis: data preprocessing, data analysis, and visualization. The data for analysis may be collected from different sources. Consequently, it may contain irrelevant and invalid data. Therefore, we apply data preprocessing to remove such data. R programming is utilized to perform the data analysis. However, when used alone, it is inapplicable to the processing of a large number of data sets as big data algorithms have issues with the continuously increasing volume, variety, and velocity of data [1]. Distributed computing is deemed suitable to address these issues. Therefore, in the present study, we utilize RHadoop to connect functionalities of R and Hadoop distributed systems. Finally, to visualize the analysis results, R libraries are employed.

Online-to-offline or offline-to-online (O2O) is an electronic commerce (e-commerce) model used to interconnect the offline business opportunities by utilizing the Internet. The O2O services can be categorized into four types as follows: (1) expanding business channels from online to offline; (2) expanding business channels from offline to online; (3) services for existing platform-based operators; and (4) services for existing platform-based aggregators. Usually, data analysis is conducted offline. To extend its offline benefits to online and efficiently use big data analysis systems, in the present paper, we propose an end-to-end system architecture to combine online and offline services by applying big data analysis and an O2O service model of the second type (expanding business channels from offline to online). The proposed system can be mainly used by truck management companies for the purpose of dynamically obtaining the big data analysis results based on O2O services to optimize the logistic freight, improve customer services, predict customer expectations, reduce costs and overhead, improve profit margins, and perform load balancing. Moreover, the main practical applications of the proposed system for truck management companies are explained in Section 5.2. In this research, we aim to provide customers and freight transport managers with a tool to make better decisions corresponding to their freight load requests with an appropriate price based on data processing and analysis.

The next section provides an overview on the research work focused on freight transport systems and data analysis. Section 3 discusses a formulation of the freight data specification, analysis environment, and the results of the freight data analysis. Section 4 outlines the definition of the O2O service model of the freight management system. Section 5 describes the proposed end-to-end freight management system architecture and O2O service interaction with freight management services. Section 6 concludes the present research work.

2. Related Work

This section provides a summary on several related works dedicated to freight transportation systems and data analysis approaches.

2.1 Freight Transport

Many freight logistic transport companies attempted to apply O2O and big data to improve their business. In this subsection, we summarize several related research works. *Huochebang* was a provider of an online truck logistics platform, which enabled an O2O truck freight platform providing integrated services for truckers and shippers in China [2]. The report published by the DHL company and entitled “Big Data in Logistics” represented the true value of big data in logistics and classified information exploitation into such categories as the operational efficiency, customer experience, and new business models [3].

In Korea, Uber launched an Uber freight service used to connect drivers who own freight vehicles with Uber services. Connected drivers could easily verify the origin, destination, route, distance, and transportation fee. As the information was communicated by a phone call or message, the Uber freight service benefitted faster payment and allowed avoiding bargain deals [4].

2.2 Data Analysis

Many programming languages can be used for the purposes of data analysis. Most widely used languages are R, Python, SQL, JAVA, Scala, Julia, and MATLAB. Particularly, R and Python are frequently applied to data analysis. Python is a multipurpose language, and R is a free, open-source, powerful, and highly extensible programming language and an environment for the statistical computing and graphics introduced by Ihaka and Gentleman [5] in 1996.

The Apache Hadoop software library is an open-source framework that enables distributed processing of large datasets across clusters of computers using simple programming models [6]. Hadoop provides support for programming in R by introducing RHadoop, an open-source package that contains several R packages that allow users managing and analyzing the data by means of Hadoop in an R environment. RHadoop is mainly related to the R packages, such as ravro, plymr, rmr, rhdfs, rhbase, and others [7]. There are many kinds of possible research ways by using R-Language for data analysis [8]. Rotte et al. [9] explained the usage and development process concerning RHadoop.

Cho et al. [10] suggested an interactive visualization for Big Data analysis using Hadoop and R. Estimation of the fuel consumption efficiency was processed and analyzed based on the digital tachograph (DTG) big data for commercial vehicles by using parallel processing and the MapReduce mechanism [11]. Kabir et al. [12] analyzed the impact of social media based on sentiment analysis and word clouds implemented in R. Optimization of sentiment analysis using machine-learning classifiers was introduced by Singh et al. [13]. Word clustering based on the part-of-speech (POS) feature was applied to perform the efficient twitter sentiment analysis [14]. The SMK-means analysis based on an improved mini batch K-means algorithm using the MapReduce mechanism was proposed for big data by Xiao et al. [15].

Big data preprocessing methods are investigated and formulated with regard to the vehicle driving data using MapReduce techniques [16]. The spatial analysis system was developed based on the map-matching logics defined in the Hadoop MapReduce mechanism [17]. The wavelet histogram generation system based on the MapReduce using Hadoop was developed by Kim [18] for big data in cloud computing. DHL logistic was focused on performing real-time analytics based on big data and artificial intelligence (AI) for the current and future scenarios, such as real-time routing and ad-hoc pickups and instant delivery [19].

3. Data Analysis for Freight Services

In this section, we provide the details about the data analysis process, its environment, and the results of the data analysis conducted on the freight management real-time data.

3.1 Data Specification and Attributes

Initially, in the present study, the real-time freight transportation data for a 4-year period (2014–2017) including 13,790,000 records was analyzed. Table 1 provides the specification of the freight data including such attributes as delivery status, package size, vehicle type, pick-up place, delivery place, delivery start date, delivery end date, and the price. The analysis results were categorized by daily trends and distance-based ones as follows: (i) the goods requesting frequency vs. days; (ii) price vs. days; (iii) order status vs. days; (iv) 4-year comparison of requested goods vs. days; (v) unit price vs. distance for all data items; (vi) unit price vs. distance for clusters based on the truck capacity; and (vii) unit price vs. distance for clusters based on weekdays and weekends.

Table 1. Freight data specification

Entry	Attribute	Type	Description
1	DeliveryStatus	String	Status of dispatch delivery information (completed, canceled, and failure).
2	PackageSize	Float	Requested goods size in Ton.
3	VehicleType	String	Type of vehicle used for the goods transfer (cargo, truck, and so on).
4	PickupPlace	String	Name of goods pickup place.
5	DeliveryPlace	String	Name of goods delivered place
6	DeliveryStartDate	Date	Date and time of goods pickup
7	DeliveryEndDate	Date	Date and time of goods delivery
8	Price	Float	Price for the goods delivery

3.2 Data Analysis Process and the Environment

In the present study, the analysis was performed by means of R library packages in R-Studio, an integrated development environment (IDE) providing a user-friendly interface for R. R-Studio contains R console, R syntax-highlighting script editor, environment history, plots, packages, and other elements [20]. In this experiment, the R library packages, such as readxl, data.table, ggplot2, and others were installed in R-Studio to perform the analysis tasks [21]. Here, Hadoop was installed on two Ubuntu machines to perform distributed computing. Hadoop functionality including MapReduce and Hadoop distributed file system (HDFS) were used to conduct this analysis. The RHadoop package was installed in the middle layer to combine the R programming functionality with the Hadoop one. Rhadoop was used to divide the large analysis tasks by combining R programming with the MapReduce functionality in HDFS. The analysis was performed by a series of tasks, such as importing the real-time data from a semi-structured raw data file, cleaning the raw data by preprocessing, data analysis, and visualizing the results.

In the present study, the freight data were analyzed according to the following three steps, as shown in Fig. 1: the first step was data preprocessing, the second step was data analysis, and the third step was data visualization.

- **Data Preprocessing:** The real-world raw data usually contain noisy, missing, and redundant

information. Therefore, data preprocessing is necessary for transforming the real-world raw data into a clean and complete dataset. Data preprocessing included data cleaning, integration, normalization, and transformation, as well as noise identification and missing value imputation tasks [22]. For this purpose, one or more techniques can be applied during preprocessing according to particular requirements. In this analysis, we applied data cleaning, integration, and transformation, as well as noise identification tasks.

- **Data Analysis:** After the data preprocessing phase, the data analysis phase was executed, as shown in Fig. 1. The dataset was converted into a useful format (tables) using analysis models and other operations. During the data analysis, a statistical model was employed to identify the unit price and the day from the starting date. Here, cluster models were used to find the goods requesting frequency, requested order, and order status by grouping the entire dataset according to a day of the week and unit price vs. distance grouping. Moreover, regression models were used to identify the relationship between prices and the frequency of orders.
- **Data visualization:** Data visualization facilitates data presentation in a graphical format. Here, data visualization was utilized to represent the analyzed freight table data from phase two in the graphical format for better understanding owing to a more organized visualization manner.

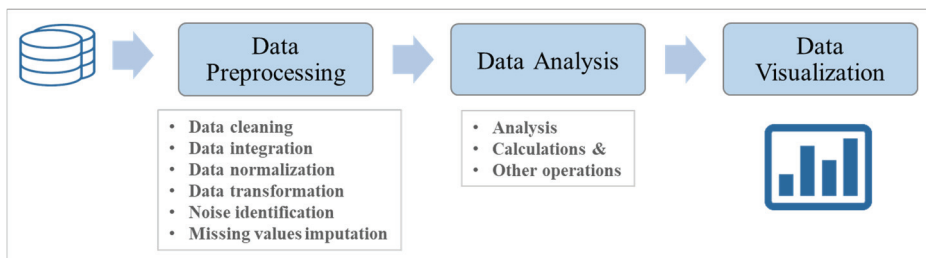


Fig. 1. The process of analyzing the freight big data.

3.3 Discussion of Data Analysis

In the data analysis process, the R project was set up to import the raw freight data as a table. The imported raw freight data were preprocessed to remove irrelevant and invalid data. During the data preprocessing, the percentage values of such data for 2014, 2015, 2016, and 2017 were found equal to 17.4%, 16%, 14.6%, and 13.8%, respectively.

After obtaining the meaningful clean data as a result of data preprocessing, the updated dataset was processed by data analysis using the statistical, cluster, and regression models. Freight data analysis charts were visualized by calculating the aggregate means. Considering the features of the freight data, we analyzed the daily and distance-based trends. These results are discussed in detail in the following subsections.

3.3.1 Daily trends

In this section, different trends, such as the goods requesting frequency, unit price, and order status vs. days are observed for 2017. Additionally, the goods requesting frequency trend vs. days is analyzed for the entire 4-year period (2014–2017).

The analysis results corresponding to the goods requesting frequency vs. days for 2017 depicted in Fig. 2(a) indicate that Sunday is characterized with the lowest number of requests (2%), and the next lowest is Saturday (7% of the total number of requests). Both Monday and Tuesday demonstrate the highest number of requests (20%). Fig. 2(a) clearly represents that the amounts of requested goods in tons differ every day, and generally, weekdays have more requests than weekends.

To represent the comparison of unit prices vs. days (Fig 2(b)), the price per unit and the day of the week need to be derived from the given data, as shown in Table 1. The package size, price, delivery start date, and distance are used to calculate the unit price. The distance is calculated by identifying the latitude and longitude for the pickup and delivery places by using Kakao map application programming interface (API) [23]. The day of the week is specified using the delivery start date. The obtained result shows during a week, Sunday has the highest freight price (18%), and Monday has the lowest one (12%).

The unit price on Sunday is higher than on the other days (Fig. 2(b)) accounts for the results of request goods in tons (Fig. 2(a)), meaning that Sundays has fewer load requests.

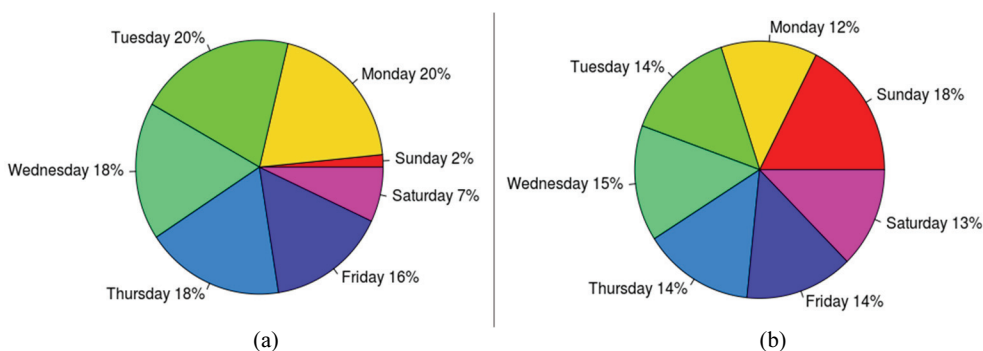


Fig. 2. Comparison of (a) requested goods in tons vs. days and (b) unit price vs. days for 2017.

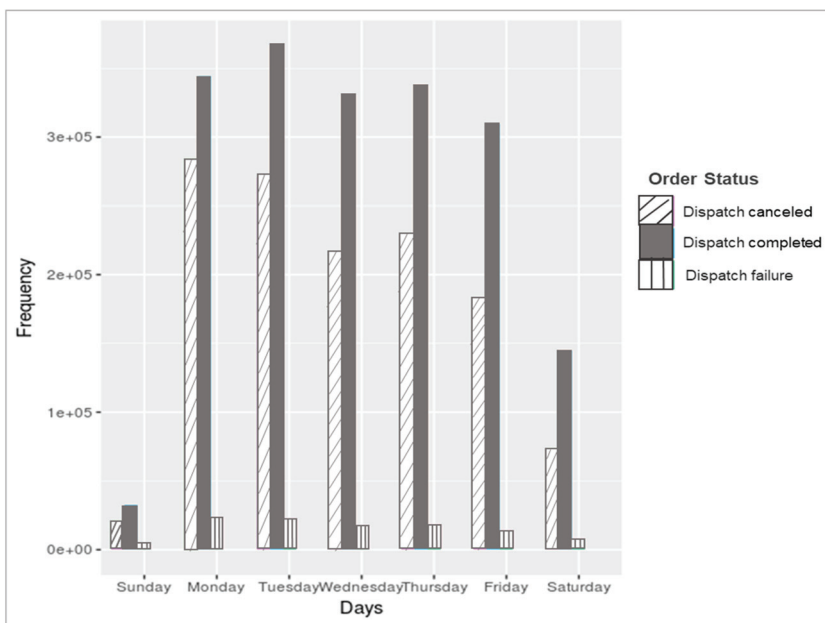


Fig. 3. Comparison of order status frequency vs. days for 2017.

Fig. 3 represents the comparison of order status vs. days for 2017 in detail. There are three order statuses available in the analyzed freight data: dispatch failure, dispatch completed, and dispatch canceled. Fig. 3 shows that the order status “dispatch completed” has the higher frequency compared with the other two statuses. Analyzing the accumulated freight delivery data, companies can make better decisions to improve the quality of customer services.

Fig. 4 represents the comparison of requested goods in tons vs. days for the 4-year period (2014–2017). The obtained results indicate that 2014, 2015, 2016, and 2017 have similar patterns, as the one represented in Fig. 2(a). Therefore, companies can predict the amount of freight to prepare trucks and to control freight allocation in advance.

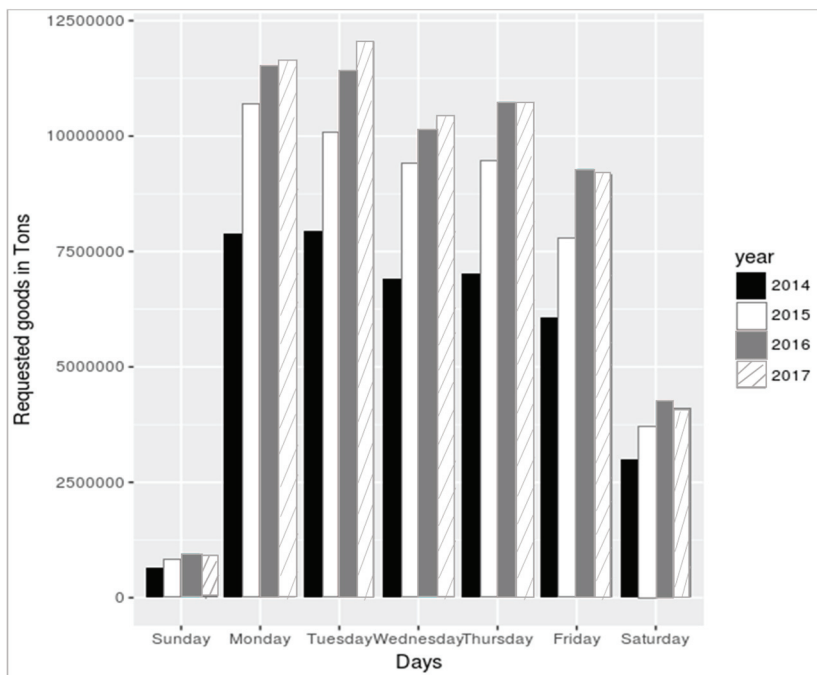


Fig. 4. The comparison of requested goods in tons vs. days for the 4-year period (2014–2017).

3.3.2 Distance-based trends

Fig. 5 represents the results corresponding to the truck capacity. The data clustered according to the truck capacity vs. distance are shown in Fig. 6. The observed clusters can be summarized as follows: “less than 10 tons” (1% of the total number of orders) in Fig. 6(a), “10 to 15 tons” (34% of the total number of orders) in Fig. 6(b), “15 to 20 tons” (14% of the total number of orders) in Fig. 6(c), and “20 to 25 tons” (51% of the total number of orders) in Fig. 6(d). While comparing these clusters, in Fig. 6(a), it can be seen that the frequency of the “lower than 10 tons” requests is rather low, and the minimum unit price is higher than those of the others when the distance increases. Fig. 6(d) indicates that the number of the “20 to 25 tons” requests is higher, and the minimum unit price is lower than those of the others with an increase in the distance and truck capacity. Fig. 6(b) and 6(c) represent medium goods request frequencies. Comparing Fig. 6(a)–(d), it is possible to conclude that the minimum unit price decreases with an increase in the truck capacity.

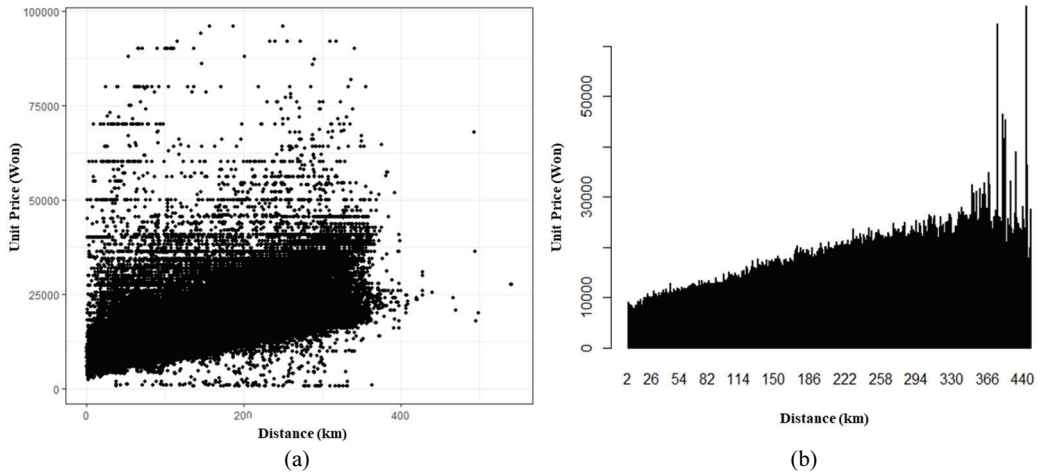


Fig. 5. Unit price vs. distance for the entire dataset: (a) scattered points and (b) regressive plots.

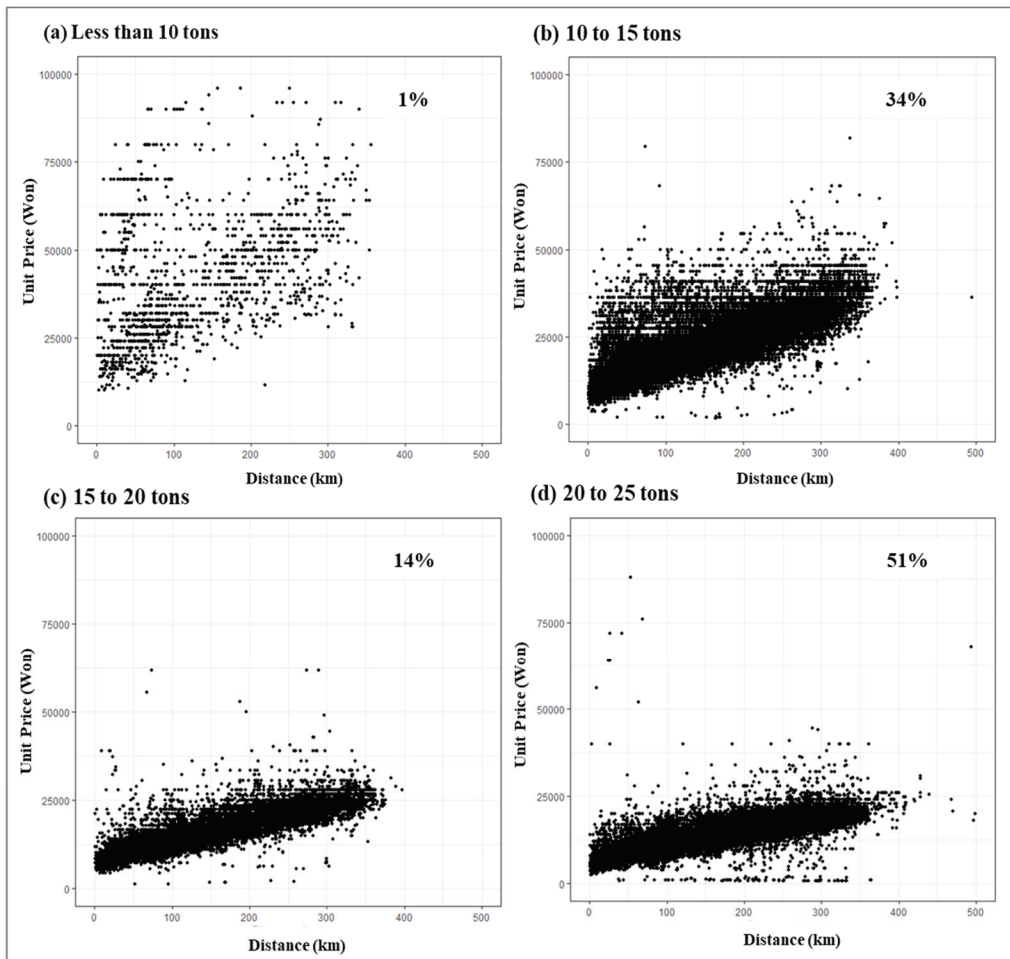


Fig. 6. Clustering the unit price vs. distance on the truck capacity: (a) less than 10 tons, (b) 10 to 15 tons, (c) 15 to 20 tons, and (d) 20 to 25 tons for 2017.

All data are clustered according to weekdays and weekends based on the distance, as shown in Fig. 7. These clusters have almost similar distribution; however, weekends considerably less frequency (7% of the total number of orders) compared with weekdays (93% of the total number of orders), which is similar to the results represented in Fig. 2(a).

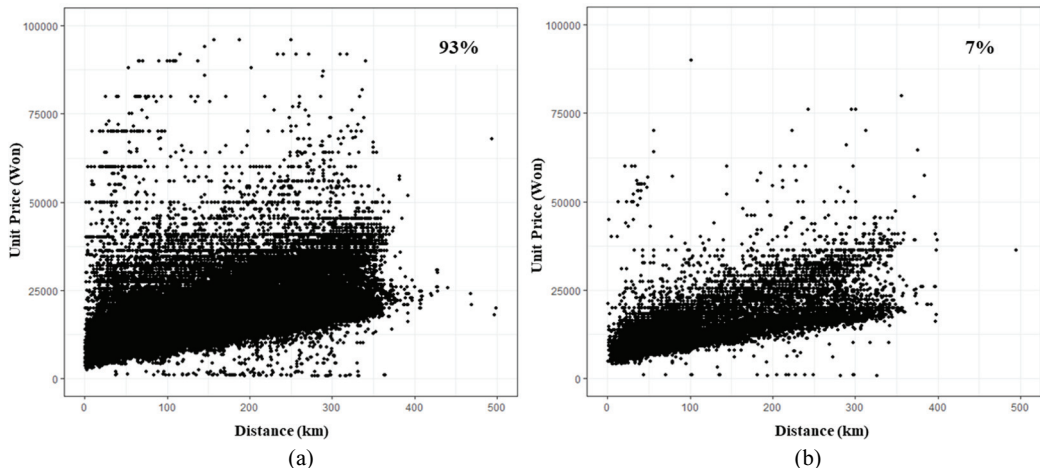


Fig. 7. Unit price vs. distance clusters based on: (a) weekdays and (b) weekends in 2017.

Based on the above results, it can be seen that data preprocessing allows detecting and removing approximately 15% of irrelevant and invalid data for the period from 2014 to 2017. The analysis results corresponding to the unit price on Sunday is higher than those of the other days, which is in line with the results for the lowest goods requests in tons. The requested goods in tons demonstrate similar patterns in the 4-year dataset. The unit price augments linearly with an increase in distances, and the unit price decreases with an increase in the truck capacity. Considering the key aspects of the analysis results may improve load dispatch balancing and the quality of customer services.

4. O2O Service Models of Freight Management Systems

The exponential growth of the mobile Internet plays a critical role in the growth of O2O. Fig. 8 clearly represents the composition of O2O services. The key functionality of O2O is to find customers online and bring them into the physical stores. Customers can purchase products online and receive the products or services at the real-world store [24].

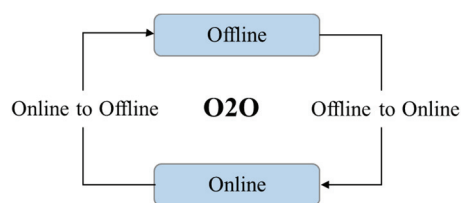


Fig. 8. Composition of an O2O services.

The term O2O meaning “offline-to-online” was introduced in 1999 by Teruyasu Murakami, a member of the board of directors of the Nomura Research Institute, Japan [24]. The O2O marketing model became famous after publishing the research work by Zhang [25] in 2014. The O2O can be applied to many sectors like tourism, real estate, ticketing, foods, rent-a-car, and so on. As shown in Fig. 9, many vendors already adopted the O2O models for their businesses. Various companies, such as Amazon, Walmart, Alibaba, and others efficiently implement O2O for e-commerce [26].

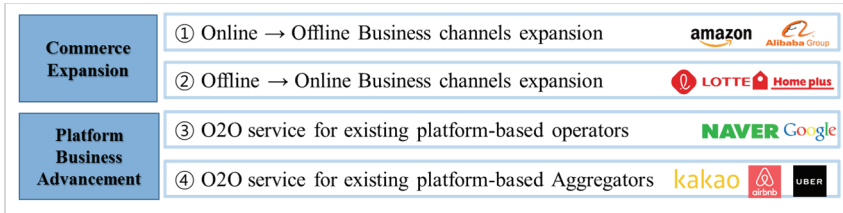


Fig. 9. Types and main operators of an O2O service.

The O2O services are mainly divided into two types: commerce expansion and platform business advancement (Fig. 9). These two types can be further classified according to four subcategories as follows:

- Expanding business channels from online to offline: In 2012, Facebook purchased *Glancee*, a location-based mobile app for social discovery to find the people around [27]. Based on this idea, in 2015, Facebook launched *Place Tips*, an O2O service application, providing the information about stores and attractions for the people who are around the stores [28]. In 2013, NTT Docomo, the leading mobile operator in Japan, launched an O2O e-commerce solution called *Shoppulatto* that was focused on shop and product discovery [29]. Tencent, the leading Internet service provider in China, also entered into the O2O market in September 2013 with introducing *WeChat* application.
- Expanding business channels from offline to online: Walmart launched a service called *Pick Up Today* in 2011 that allowed customers purchasing goods online and picking them up offline. In this case, the product could be picked up on the same day when the order was.
- O2O services for existing platform-based operators: In the United States (US), Google used O2O to link Uber as a transportation option in online restaurant reservation services aiming to expand the linkage with Google Maps. The Naver's mobile messenger LINE with more than 50 million subscribers in Japan utilized the mobile instant messaging (MIM) method, which transmitted discount coupons for LINE subscribers in cooperation with Lawson, the convenience store chain in Japan.
- O2O services for existing platform-based aggregators: Airbnb, a US start-up company providing accommodation sharing services, and Uber, a vehicle-sharing service, played a major role in promoting the O2O service market worldwide.

5. Freight Management End-to-End System Architecture

This section describes a freight management system and its functionality serving to handle freight tasks as upcoming marketing and customer demands.

According to the future trends of freight management systems, such applications, as autonomous vehicles, real-time routing, Internet of Things (IoT) services, big data with AI to predict the real-time demand and to improve customer experience, and Blockchain technologies, will be important [19,30].

To provide end-to-end O2O services to a freight management system, we propose a system architecture that includes several parts. Fig. 10 represents the proposed end-to-end system architecture, which can be useful to build freight management systems facilitating dynamic queries in the big data analysis. The proposed structure implies combining the functionalities of the three following systems: O2O service system, truck management system, and big data analysis system. Here, the O2O service system is operated online, and the truck management and big data analysis systems are implemented offline. In the present study, we focus on O2O of the second type that can be used to expand business channels from offline to online O2O service models.

As shown in Fig. 10, a customer can send a request to the truck management system for fixing a shipment appointment with truck companies by interacting with the O2O service system. After receiving the customer requests, the O2O service system notifies the truck service manager through the truck management system. The truck service manager manually processes this request offline by directly contacting each of truck service companies to ensure the availability of trucks to fix a shipment appointment.

Truck service companies process the request from the truck service manager, and then, the truck management system outputs the list of available trucks. After obtaining the list of available trucks for the particular request, the truck service manager contacts the truck analysis manager to analyze the list of available trucks.

Before fixing an appointment with the truck service company, the truck analysis manager evaluates the list of available trucks using the available analysis data. If the required analysis data are already available, the truck analysis manager responds to the truck service manager. Otherwise, the former requests an analysis expert to obtain the data from truck companies (weekly or monthly), such as the information about shipment, date, price, feedback, and so on.

The analysis expert directly queries the big data analysis system to analyze the log data, and then, the expert shares the obtained analysis results with the truck analysis manager periodically or on the on-demand basis. According to these analysis results, the truck analysis manager can select more reliable performers to improve customer service, to reduce price and overhead, and to increase profit margins.

Once the truck service manager selects an optimal option, the appointment information is sent to the chosen truck company. After receiving the confirmation from the truck company, the truck service manager forwards the appointment response to the customer through the O2O service system.

As shown in Fig. 10, all mandatory actions are logged in to the big data storage, which may contain various information, such as logs of previous appointments, feedback from customers, and so on, collected from the systems.

In the proposed system architecture, big data analysis is processed in four layers as follows: (1) big data layer, 2) analysis programming layer, (3) middle layer, and (4) HDFS. The big data layer deploys the big data storage. The analysis programming layer is used to process the data stored in the big data layer by means of R, Python, Java, and Scala programming languages. HDFS is used for distributed computing so that the tasks are distributed to computers within the cloud system. The middle layer is used to combine functionalities of the analysis programming layer and the HDFS one. There are many software packages available in the middle layer, including Hive, Spark, and the MapReduce mechanism, which can be utilized depending on requirements.

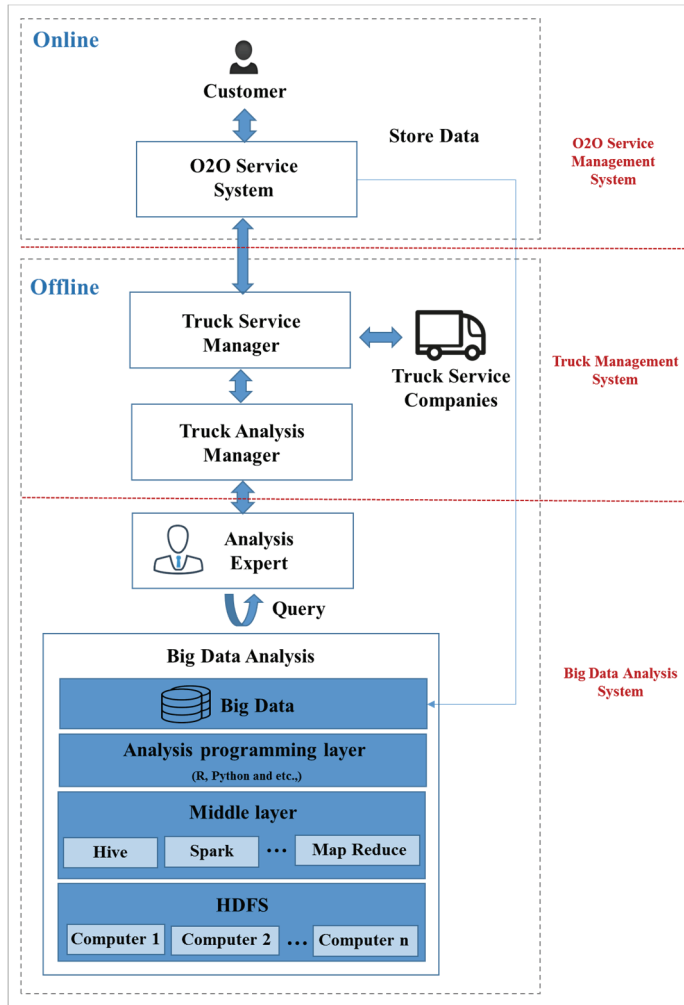


Fig. 10. End-to-end freight management system layout in online and offline services.

5.1 The O2O Service Interaction with the Freight Management System

In this subsection, we discuss the O2O service interaction and activity relationships among various parties and corresponding systems. The interaction regards the influence of customer requests and corresponding decisions made by a freight management company and management system.

Fig. 11 represents the sequential diagram of services for an end-to-end freight management system. The proposed system utilizes these services to fix shipment appointments between customers and truck companies:

- Step 1: All truck companies register their truck details, shipment information, and unit price in a truck management system.
- Step 2: A customer creates a shipment request for the O2O service system providing the input information about location, date, and the weight of goods.
- Step 3: The O2O service system forwards the customer request to the truck management system together with the details obtained at step 2.

- Step 4: The Big Data analysis process is executed by being triggered periodically or on-demand by an analysis expert. After completing each analysis process, the analysis expert forwards the updated parameter weight to the truck analysis manager.
- Step 5: The process to check the availability of trucks is triggered by the truck service manager. Here, after receiving the request from the O2O system, the truck service manager requests truck service companies to check the availability of trucks for the requested location, date, and the weight of goods. If the suitable trucks are available for the particular request, truck service companies return an availability response indicating the price of a shipment. Finally, after step 5, the truck service manager receives the list of available trucks corresponding to the request.
- Step 6: In Step 6, the truck service manager sends a feasibility check request to the truck analysis manager together with the input list of available trucks.
- Step 7: The truck analysis manager analyzes the truck details considering the results of big data analysis aiming to find a feasible option based on goods customer service, the possibility of cost reduction, and other parameters.
- Step 8: The truck analysis manager can always access the real-time analysis results by using the analysis tool and forwards a response with the feasible truck list to the truck service manager. The analysis manager can request the analysis expert for on-demand analysis.
- Step 9: Once the truck manager obtains the feasible response from the analysis manager, the appointment details are forwarded to the chosen truck company.
- Step 10: If the truck company accepts this appointment, it sends a confirmation to the truck service manager.
- Step 11 and 12: The truck service manager sends a confirmation response to the customer through the O2O service system indicating the information about location, date, truck details, and price.
- Step 13 and 14: The log data are saved in the big data storage. The log data may contain the information about previous appointments, feedback from customers, and so on.

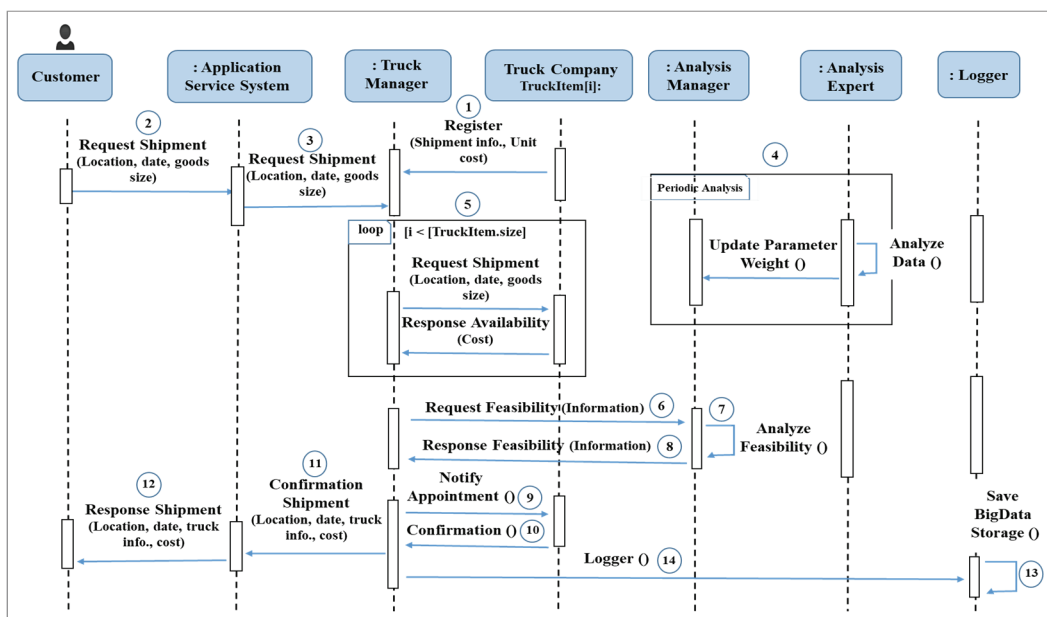


Fig. 11. Sequential diagram of end-to-end freight management system services.

5.2 Features of the Proposed System

The proposed system is deemed useful for the truck management companies to dynamically request the big data analysis results using O2O services to optimize the logistic freight, cost and overhead reductions, grow profit margins and load balancing, predict the customer expectation, and to improve customer services.

- **Optimization of Logistic Freight:** the suggested system can be used to predict the demand by analyzing the daily trends of customer requests and prices derived from the previous data. These results can be useful for truck management companies to optimize the allocation and scheduling processes.
- **Cost and overhead reductions:** the proposed system can be used to obtain a daily observation of trends in the freight data. In this way, the truck management companies can predict the demand and reduce costs and overhead.
- **Growing profit margins and load balancing:** the proposed system is deemed useful to identify the optimal number of trucks to prepare in advance based on analyzing market trends to control freight allocation for load balancing and increasing profit margins.
- **Predication of customer expectations:** the proposed system can be used to predict the customer expectations by analyzing customer requests based on the day of the week and festival seasons based on the previous freight data.
- **Improve customer services:** before responding to a customer appointment request, truck management companies can query the big data analysis tool dynamically to analyze the data, such as feedback and other information corresponding to truck companies. Based on these results, truck management companies can provide better truck services to the customer. These approaches are aimed to improve customer satisfaction and minimize customer attrition.

5.3 Aims of Recent Technologies in the Freight O2O Services

In the continuously evolving world, logistic companies need to perform real-time demand forecasting of planning tasks to improve their business processes. At present, many logistic companies focus on the on-demand real-time analytics to implement routing, ad-hoc pickups, instant delivery, and demand forecasting [19]. Namely, the on-demand real-time analytics receives requests or queries from users or systems and then, outputs the analysis results, which need to be considered in decision-making [31]. To overcome the existing issues, in the present paper, we proposed a system architecture based on big data and AI intended to conduct continuous real-time analytics for ad-hoc demand forecasting, which can be useful for decision making [30].

6. Conclusions

In the present research work, we analyzed and extracted the useful knowledge from the freight big data to facilitate decision-making aiming to improve business processes. Usually, big data analysis is performed offline. To extend offline services to online applications, we proposed the system architecture for freight management systems based on big data analysis and O2O services. This system was deemed useful for freight management companies to improve customer services, reduce costs and overhead, and improve profit margins.

We analyzed the freight data for the 4-year period (2014–2017) and categorized the obtained results into two subcategories, such as daily trends and distance-based ones. The results included the following information: the frequency of requested goods vs. days, price vs. days, order status vs. days, 4-year comparison of requested goods vs. days, unit price vs. distance for all data items, unit price vs. distance for clusters based on the truck capacity, and unit price vs. distance for clusters based on weekdays and weekends. As a result of preprocessing, we detected the percentage values of irrelevant and invalid data in 2014, 2015, 2016, and 2017 equal to 17.4%, 16%, 14.6%, and 13.8%, respectively. The analysis results corresponding to the goods requesting frequency indicated that the lowest number of requests (2%) were executed on Sunday, while Monday and Tuesday were characterized by the highest number of requests constituting 20% of the total. Moreover, the data on requested goods in tons obtained during the considered 4-year period demonstrated similar patterns. It was observed that the unit price augmented linearly with an increase in distances, and the unit price decreased an increase in the truck capacity. Then, we noted that the “20 to 25 tons of goods” transfer requests had a higher frequency (51% of the total number of requests) and the “lower than 10 tons” requests were characterized with the minor frequency (1% of the total number of requests). Considering these results, companies could predict the amount of freight for preparation of trucks to control freight allocation in advance.

As the future research work, we plan to analyze the next level of classification aiming to identify useful patterns in data to further optimize freight logistic processes, improve customer services, predict customer expectation, reduce costs and overhead, and profit margins and load balancing.

Acknowledgement

This work was supported by the Technological Innovation R&D Program (No. A2017-0293) funded by the Ministry of SMEs and Startups (MSS, Korea) and Leaders in Industry-university Cooperation project, supported by the Ministry of Education and National Research Foundation of Korea.

References

- [1] D. Laney, “3D data management: controlling data volume, velocity and variety,” 2001; <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- [2] Crunchbase, “Huochebang: an on-demand platform,” 2008; <https://www.crunchbase.com/organization/huochebang>.
- [3] M. Jeske, M. Gruner, and F. Weiss, “Big data in logistics: a DHL perspective on how to move beyond the hype,” 2013; https://www.dhl.com/content/dam/downloads/g0/about_us/innovation/CSI_Studie_BIG_DATA.pdf.
- [4] Uber Freight, <https://www.uber.com/us/en/freight/>.
- [5] R. Ihaka and R. Gentleman, “R: a language for data analysis and graphics,” *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 299-314, 1996.
- [6] Apache Hadoop, <http://hadoop.apache.org/>.
- [7] Revolution Analytics, “Rhadoop and MapR,” 2014; https://mapr.com/resources/rhadoop-and-mapr/assets/rhadoop_and_mapr.pdf.

- [8] MakeMeAnalyst, "List of Big Companies using R," 2017; <http://makemeanalyst.com/companies-using-r/>.
- [9] A. V. Rotte, G. Patwari, and S. Hiremath, "Big data analytics made easy with Rhadoop," *International Journal of Research in Engineering and Technology*, vol. 4, no. 5, pp. 9-15, 2015.
- [10] W. Cho, Y. Lim, H. Lee, M. K. Varma, M. Lee, and E. Choi, "Big data analysis with interactive visualization using R packages," in *Proceedings of the 2014 International Conference on Big Data Science and Computing*, Beijing, China, 2014, pp. 1-6.
- [11] W. Cho and E. Choi, "DTG big data analysis for fuel consumption estimation," *Journal of Information Processing Systems*, vol. 13, no. 2, pp. 285-304, 2017.
- [12] A. I. Kabir, R. Karim, S. Newaz, and M. I. Hossain, "The power of social media analytics: text analytics based on sentiment analysis and word clouds on R," *Informatica Economica*, vol. 22, no. 1, pp. 25-38, 2018.
- [13] J. Singh, G. Singh, and R. Singh, "Optimization of sentiment analysis using machine learning classifiers," *Human-centric Computing and information Sciences*, vol. 7, article no. 32, 2017.
- [14] Y. Wang, K. Kim, B. Lee, and H. Y. Youn, "Word clustering based on POS feature for efficient twitter sentiment analysis," *Human-centric Computing and Information Sciences*, vol. 8, article no. 17, 2018.
- [15] B. Xiao, Z. Wang, Q. Liu, and X. Liu, "SMK-means: an improved mini batch k-means algorithm based on MapReduce with big data," *Computers, Materials & Continua*, vol. 56, no. 3, pp. 365-379, 2018.
- [16] W. Cho and E. Choi, "Big data pre-processing methods with vehicle driving data using MapReduce techniques," *The Journal of Supercomputing*, vol. 73, no. 7, pp. 3179-3195, 2017.
- [17] W. Cho and E. Choi, "A basis of spatial big data analysis with map-matching system," *Cluster Computing*, vol. 20, no. 3, pp. 2177-2192, 2017.
- [18] J. J. Kim, "Hadoop based wavelet histogram for big data in cloud," *Journal of Information Processing Systems*, vol. 13, no. 4, pp. 668-676, 2017.
- [19] B. Gesing, S. Peterson, and D. Michelsen, "Artificial intelligence in logistics," 2018; <https://www.ibm.com/downloads/cas/XOQW7Q0D>.
- [20] RStudio, <https://www.rstudio.com/products/rstudio>.
- [21] R packages CRAN, <https://cran.r-project.org/>.
- [22] S. Garcia, S. Ramirez-Gallego, J. Luengo, J. M. Benitez, and F. Herrera, "Big data preprocessing: methods and prospects," *Big Data Analytics*, vol. 1, article no. 9, 2016.
- [23] Kakao Developers, <https://developers.kakao.com>.
- [24] Y. M. Kwon, J. Park, H. Lee, and M. Kim, "Beacon-based O2O marketing for financial institutions," *The International Journal of Industrial Distribution & Business*, vol. 5, no. 4, pp. 23-29, 2014.
- [25] J. Zhang, J. "Customer' loyalty forming mechanism of O2O e-commerce," *International Journal of Business and Social Science*, vol. 5, no. 5, pp. 164-169, 2014.
- [26] X. Hu, P. K. Chen, and Z. Lou, "Building a forecasting model of customer demand for O2O dual-channel," in *Proceedings of 2017 International Conference on Applied System Innovation (ICASI)*, Sapporo, Japan, 2017, pp. 721-723.
- [27] J. Bort, "Facebook just did a really smart thing by acquiring Glancee," 2012; <https://www.businessinsider.com/facebook-just-did-a-really-smart-thing-by-acquiring-glancee-2012-5>.
- [28] Facebook, "Place tips: a valuable page tool for local businesses," 2015; <https://www.facebook.com/business/news/place-tips-for-businesses>.
- [29] NTT Docomo Shoppulatto, 2013; https://projekter.aau.dk/projekter/files/77307375/Shopping_Assitant_ICTE.pdf.
- [30] S. Moore, "Gartner identifies top 10 data and analytics technology trends for 2019," 2019; <https://www.gartner.com/en/newsroom/press-releases/2019-02-18-gartner-identifies-top-10-data-and-analytics-technolo>.
- [31] Gartner Glossary: real-time analytics, <https://www.gartner.com/en/information-technology/glossary/real-time-analytics>.

**Suganya Selvaraj** <https://orcid.org/0000-0003-4802-8136>

She received her Master's degree in Master of Computer Applications from Alagappa Chettiyar College of Engineering and Technology, Karaikudi, Tamilnadu, India in 2011. She had been working as a software developer in HCL technologies, India and Tata Consultancy Services, India from 2011 to 2016. Currently pursuing her PhD degree at Kookmin University, South Korea. Her research interests include big data processing, distributed system, Blockchain, and knowledge-based systems.

**Hanjun Kim** <https://orcid.org/0000-0001-8899-6380>

He received a B.S. degree in Industrial Management Engineering from Hankuk University of Foreign Studies in 2015. Pursuing his M.S. in the Department of Financial Information Security, Kookmin University. His interests include Blockchains, big data, big data analysis, ontology, and so on.

**Eunmi Choi** <https://orcid.org/0000-0002-9743-2437>

She is a Professor at Kookmin University, Korea. Her current research interests include big data infra system and analysis, cloud computing, intelligent system, information security, parallel and distributed system, and SW architecture and modeling. Professor Choi received her M.S. and Ph.D. degrees in Computer Science from Michigan State University, the USA in 1991 and 1997, respectively, and B.S. degree in Computer Science from Korea University in 1988 with the top student award. Since March 1998, she had worked as an assistant professor at Handong University, Korea, before joining Kookmin University in 2004. She is the Head of Distributed Intelligent System Lab., in Kookmin University.