International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 12 (2020), pp. 4151-4155

Accuracy Improvement Technique of Big Data based LBS System

Changbae Mun

[1] *Professor, Dept. of Electrical, Electronic & Communication Engineering, Hanyang Cyber University, Seoul, Korea.*

[1] *ORCID: 0000-0002-3065-3307*

Abstract

The purpose of this study is to construct a methodology by focusing on the technique of rapidly extracting refined data through big data analysis and grafting it into actual applications in terms of accuracy of movement. Currently, as LBS (Location-based service) through various personal communication terminals is activated, the importance of the accuracy improvement technique of LBS is becoming more and more prominent. The system aims to present the most efficient route information by analyzing the movement information of people using various routes. To this end, a system based on map information is constructed, and routing information is analyzed in simultaneously on a server, so that a person who frequently travels the route can provide the same effect as direct guidance. In this study, through the improved intelligent LBS system, users can be provided with more advanced level of location information, and this function can be developed in connection with the existing LBS service.

Keywords: Location based Service, LBS, Apache Spark, Big Data, Recommendation Service

I. INTRODUCTION

Along with the development of Internet services, the location information system and LBS (Location-based service) industries are further developing. The LBS industry is developing by providing a variety of services based on information continuously collected by users' location information. It is increasing at an annual average of 39.77%. The development of intelligent LBS technology is mainly due to the expansion of the service area of various applications related to mobile information. The number of operators participating in the LBS industry has increased, and the development of location information technology and infrastructure industry, and the provision of innovative services due to competition among global internet service companies to introduce LBS are the main factors in the development of domain technology [1]. With the development of technology, various services related to LBS such as entertainment, location information, and surrounding information services such as advertisement services have been introduced. In particular, in the mobile commerce industry, it is used for functions such as various product information, discount coupons, and purchase information recommendation to provide information to users [2]. In particular, with the development of food delivery-related information provision services in Korea, review information of local restaurants and users is provided along with map information. Due to this information, cases in which urban development is newly activated are emerging, and cases of convergence and connection with various tourist information systems are increasing. In this study, a methodology that can more accurately present the route between two points for a specific area was sought, and the movement information of local residents who had relatively high accuracy information about the route information of the area was analyzed. The purpose of this study is to propose a method of applying the results through statistical analysis of data and similarity algorithm to LBS.

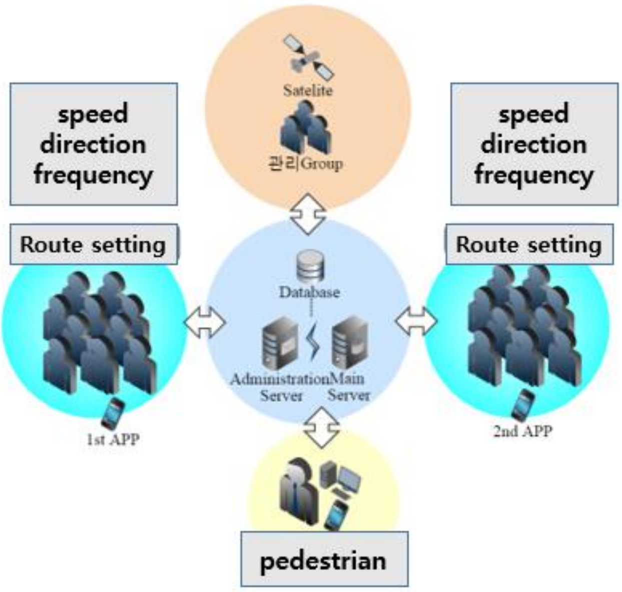

Figure 1. Overall system structure

II. TECHNICAL REVIEW

2.1 Path analysis based LBS system

Representative previous studies on LBS were analyzed. Mun's study collected and analyzed information on users traveling on the same route in a central server. As a result of this study, the method that many people traveled the same route was extracted. In addition, user history information is added to find specific local information of interest to the user in a specific route, and marketing information of a store related to the route. When this information is confirmed, the information is provided to the user in the form of push, text, and notification calls [3]. In a study on a recommendation system through collaborative filtering, a method of providing personalized content was proposed. This function finds major places within the distance set by the user based on the current location, and analyzes the content in units of preference for all users. As a result of this data analysis, content that is predicted to be of high utility to users is provided [4]. A study on the performance improvement technique of the vehicle location tracking system has discussed the method of dividing the space within the area and continuously obtaining the location information of the vehicle [5].

2.2 Functional issues for prior research

In In particular, the research in [5] focused on the presentation of personalized information, a function not used in general navigation. In other words, conventional commercial navigation has a situation where it may be difficult to obtain information when moving the actually suggested route according to climate specifics such as dark roads or rainy weather and user characteristics. In the case of the route to which the user goes for the first time, even if the user checks and understands the map in advance, it is affected by the weather and the accuracy of the map information. In addition, the time required for the user's movement may vary depending on the complexity of the terrain. In particular, it is difficult to visually check the main features of the route information at night or in rainy weather. It is easy to find route information on roads that are the main targets of general navigation, but in the case of alleyways, it is difficult to check information due to a breakdown of streetlights depending on the condition of the road. When the view is not comfortable due to weather such as rain or snow, or there is road control due to construction at a specific location, it is difficult to immediately respond to the information in the navigation system. In order to solve such difficulties, various situations that can occur when the user walks the first road were assumed. And this paper deals with how to solve these individual situations through intelligent information systems. In connection with the location processing platform of the portal map information service, a method of providing refined information dynamically by utilizing the location of each user and providing practical services more accurately based on the user's surrounding information was explored. As an approach to this, in this paper, we analyzed various big data including movement information collected about users. In addition, the LBS platform presented in this study aims to analyze the most efficient route information by collecting movement information of people using various routes. To this end, the structure of a statistical analysis system based on map information was designed, and the movement information of each user was analyzed in real time in the central server. So, through this system, a person who has information about a specific route can get the same effect as guiding you right next door. Since users are provided with more accurate route information through this service, this system can be developed in conjunction with various LBS applications.

III. SYSTEM DESIGN AND VERIFICATION

3.1 Overall functional design of the system

Distributed processing system refers to a system in which a number of systems are connected through a computer network based on network communication and internal interworking algorithms, and system resources distributed in different data centers are connected and integrated. This system has the function of sharing system resources through distribution and increasing efficiency such as speed and capacity.

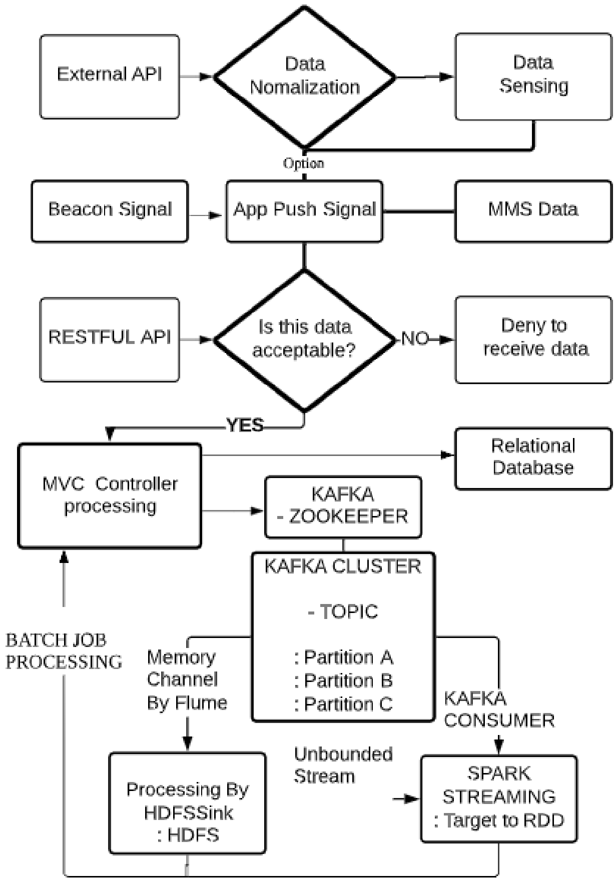

Figure 2. System process configuration [5]

Table 1. Data set and key elements

DATA Level	DATA SET	Included elements
Initial	Latitude and longitude	N/A
Managed	Pedestrian path	XML FILE
Defined	The trajectory of the movement of cars and buses	XML Web Service
Quantitatively Managed	Integrated information system for travel routes	XML External API Web Service

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 12 (2020), pp. 4151-4155

Distributed processing system refers to a system in which a number of systems are connected through a computer network based on network communication and internal interworking algorithms, and system resources distributed in different data centers are connected and integrated. This system has the function of sharing system resources through distribution and increasing efficiency such as speed and capacity. The advantage of a distributed processing system is that even if the target throughput of the entire system increases, it has an efficiency in the individual system throughput distribution method. Hadoop technology, which is widely used in the aspect of data distributed storage, corresponds to the implementation of Hadoop Distributed File System (HDFS) and MapReduce, which can replace Google's file system. Through this HDFS, multiple systems integrated into the system network configured with Hadoop can access data through each node, and support efficient parallel processing in a clustered system environment using distributed processing technology Map Reduce. It is a framework that provides functions for distributed processing of many data sets in a computer cluster. This system is designed so that a single server can have a configuration scaled to thousands of machines. By introducing the data processing architecture to system scheduling at the node level, the streaming data is processed more stably. This study collects and analyzes user's movement data as shown in Figure 1, and excludes main information that can be used as personal information from the information provided by the user. By integrating this process, the entire collection data stream is formed.[5] As the overall schema of the system, the information processing system consists of a function for analyzing and processing collected information and a function for guiding users through a smartphone application. Compensation systems are applied to ensure that people who are familiar with the main branch routes and simple routes of the origin and destination of a particular route act as providers of route information. Most likely, people who have lived in this area for a long time will be the providers of route information. This function activates the overall activity of the system and allows the system to provide more accurate route information. The overall system architecture is as follows.

: SPARK for the purpose of distributed processing of data to collect and analyze route information provided by users

: KAFKA for the purpose of distributed delivery of sets,

: FLUME module for the purpose of performing optional functions in Hadoop

These modules were integrated into the system. Among them, the FLUME module functions to store a large amount of log data generated in the system. In other words, it is used for storing individual information in the information collection function server, has features of high scalability and reliability, and presents a platform for efficiently processing and loading large amounts of data in a real-time streaming server. Flume module is an open source, and it collects information for all structured or unstructured data, including server logs in the system, and can set up compatible syncs with heterogeneous systems through various formats and various networking protocols. A connection with a single channel is connected to multiple storages. In this environment, information linkage is provided from a single source in a complex sink. In addition, system function efficiency is provided through Spring Cloud Data Flow, which can simultaneously process complex data input and output. In particular, it collects the relayed information of the API from the controller of the framework and delivers the data stream based on Spark Stream and Flume. In addition, the Flume module has designed the system structure by storing text logs in HDFS for additional batch processing.

As shown in [Figure 2], the overall structure of the system is composed of information providers, collectors, and managers corresponding to system users as the subjects of each device. Before the user, who is the information provider, moves the route to collect information, the destination information is input in the mobile application. After the user clicks the start button, the user moves the route. At this moment, the central server transmits GPS information of the user's location in units of 2 seconds to the central server. All this information is set and maintained in system properties. The user's movement information is continuously accumulated in the server, and even option data for each movement information is accumulated in the DB as historical data. Through this, the goal of this service is to provide analysis information composed as shown in [Figure 3]. This information is provided to the user as optimal route information. The system analyzes the information provider's movement, and compares each individual's characteristics with information about the weather or road environment. Therefore, even residents residing in similar areas can have multiple route histories, and users of this system make decisions using the recommended route based on the latest information.

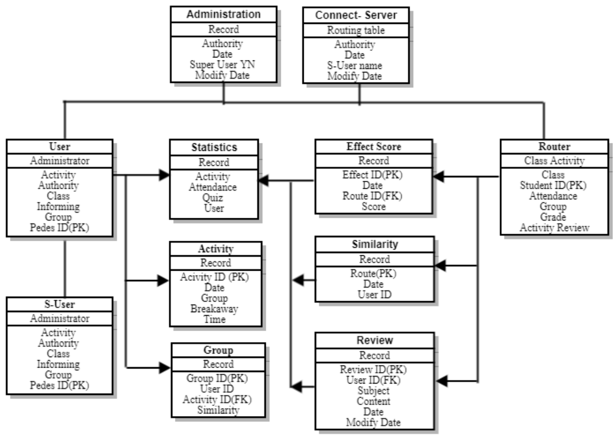

Figure 3. System's main database [5]

The central server is set to transmit GPS information of the user's location in units of 2 seconds to the server. The user's movement information is continuously accumulated in the server, and optional data for each movement information is accumulated. Through this, the goal of this service is to provide the best route to users through analysis information composed as shown in [Figure 3]. Several routes may exist even for residents of the same area according to the movement characteristics of the information provider or the weather or road environment, and users make decisions using the recommended route based on the latest information.

The system configured as described above has three representative functional elements for accurate route information guidance.

- Function 1: Infer the common path of users as a shortcut
- Function 2: Extracting exception information from the user's moving path
- Function 3: Verify temporary exception information and present it to the user

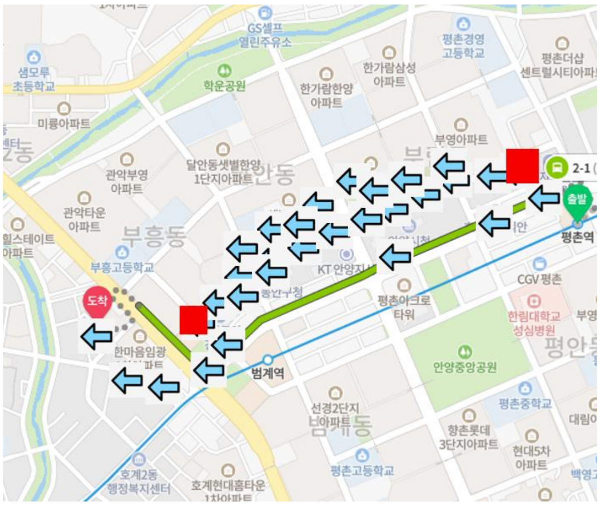

Figure 4. Results of simulation experiment [6]

First, function 1 analyzed the route that users residing in a specific area travel. Through data analysis, detailed paths that commonly move within the path were found. This route is a route that many residents of the region go, and corresponds to a route that has been empirically proven by users. In addition, this route is highly connected to public transportation such as buses and subways, and is likely to be helpful to users in probability. In order to analyze this route, we collected all of the information on the moving route of users on a daily basis. The contact value of the normal distribution of the moving path obtained as a result of this was statistically analyzed. Through this data analyzed by this LBS system, the optimal path collected based on the experiences of users was constructed.

Secondly, function 2 extracts exceptional movement information from the user's movement path. In Function 1, in terms of analyzing the functions of users, it becomes possible to pre-process the data more accurately by extracting the exceptions that occur regularly. For example, movement to specific coordinates, such as going to a beauty salon every two weeks or visiting a mart every three days, causes noise in the path. In order to solve this problem, it goes through a process of organizing and analyzing similar movement information that is common to users. This makes it possible to organize information with more completeness to users. For this information analysis, a multi-label analysis method was introduced.

For the sigmoid function used in the logistic regression applied in function 2, the input data outputs a constant value less than 1 and provides the probability that the corresponding value belongs to one of the two.

Similarly, in the case of multiple binary classification, binary cross entropy loss (BCELoss) is used for binary classification as shown in the following equation, and this technique is included in the cross entropy loss function. By applying the sigmoid function to each of n nodes, binary classification is performed as a model. In the case of exceeding binary classification, n softmax should be applied to each classification for different types of labels. In this case, the final loss function is as follows. Independent Binary Classification is complex, so it is combined into a single matrix, and this value is simultaneously entered into the Softmax function. The softmax function takes a k-dimensional vector of the total number of classes required for classification and estimates the probability for each class. In this study, as a result of this statistical analysis, predictions were obtained from the Softmax function.[6]

Thirdly, function 3 verifies the temporary exception information that occurs less frequently than function 2. The system provides this result to the user. For example, road work that occurred at a specific time or movement is blocked at a specific point on a snowy road. In this case, the moving speed of users is slowed down or the user bypasses at a specific point. The user's moving coordinates are analyzed and the system provides this information to the users. This function is helpful when there is a difference in the movement of users who follow a similar route from the existing route. At this time, the system performs collaborative filtering for users in a similar path. In certain circumstances, there may be cases where user data is insufficient for analysis cases. If the history information is small in terms of system verification, there is a high possibility that the accuracy of the provided information is deteriorated. Therefore, in this type of analysis, a standard is set to calculate the similarity between users only when the minimum number of evaluation personnel is set and 10 or more. Through this, we tried to reduce the exceptional errors that may occur in the experiment. The mobility experiment was conducted through the system implemented in this way, and the following results were obtained. The solid line in the green stone is the movement path of a specific map information service, and the information indicated by arrows is the information of this system. The effectiveness of this system was verified by arranging virtual construction sites and avoidance points at red points.

IV. EVALUATION OF SYSTEM

As a result of this, an average of 5.24 seconds of support was provided when real users moved. In addition, information was provided within 1.46 seconds when the user requested, and 7 cases of service provision error occurred in 10000 requests. And out of the total provided information, 24 route data were accurately matched, and 10 data contributed to improving the user's speed by 10 times. This information is reflected back to the system and used to improve the accuracy of the system. In this study, representative technologies and technology trends were analyzed in relation to LBS-related accuracy improvement techniques. The technical solutions required to improve the accuracy of the LBS system proposed in this study were examined. We intend to provide a functional development direction for the LBS market, which is expected to grow continuously in the future. It is hoped that subsequent research will further develop the LBS system proposed in this study and help to implement a more accurate LBS system.

ACKNOWLEDGEMENT

This work was supported by the research fund of Hanyang Cyber University (2019)

REFERENCES

[1] Korea's Internet Competitiveness, "LBS Industry Trend Report", 2018.

[2] Tiwari, Shivendra, et al. "A survey on LBS: system architecture, trends and broad research areas." International Workshop on Databases in Networked Information Systems. Springer, Heidelberg,

[3] Kushwaha, A., & Kushwaha, V. (2011). Location based services using android mobile operating system. International Journal of Advances in Engineering & Technology, 1(1), 14.News, 2017

[4] Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. IEEE Internet computing, 7(1), 76-80.

[5] Mun, Chang-Bae, and Hyun-Seok Park. "Design of the Intelligent LBS Service: Using Big Data Distributed Processing System." The Journal of the Korea Contents Association 19.2 (2019): 159-169.

[6] Mun, Chang-Bae. " The LBS system with accuracy improvement function using big data." KAIS(Korea Academia-Industrial cooperation Society) Conference proceeding1 1.27 (2020).