Development of Road Traffic Analysis Platform Using Big Data

Sung, Hong Ki[1] and Chong, Kyu Soo[2]
[1,2] Korea Institute of Civil Engineering and Building Technology, Goyang, S. Korea

Abstract - *This paper presents the development direction of the driving environment prediction platform based on the road traffic big data. Towards that end, the levels and trends of the development of related big data platform technologies in Korea and overseas were analyzed, and the differentiation and expected effects of the platform that was developed in this study were examined. The suggested road driving environment prediction platform can provide traffic information services distinct from the existing systems, thereby boosting the general users' and experts' access to traffic data. It is expected that it can be used in the information system related industries and in the markets in diverse fields. In addition, the platform can use road traffic public, sensor, and unstructured data to provide meticulous road driving environment information, to improve the drivers' safety, and to boost the reliability of traffic prediction information.*

Keywords: Road Traffic Big Data, Platform, Driving Environment, Vehicles' Sensor Data, Real-time Data

1 Introduction

In the field of road traffic as well, the data scalability has been improved due to the development of related technologies. With the development of the sensor measurement technology for general personal vehicles as well as business vehicles like buses, taxis, and trucks, the scalability has been expanded into the mobile observation technologies (e.g., detection sensors, navigation, black boxes) from the fixed-type observation technologies (e.g., loop detectors, image detectors, CCTVs).[1] Thus, it is expected that the use of big data in the road traffic field will become more active, and it is essential to develop an intelligent IT environment and platforms that can collect, store, and analyze various types of data, such as structured and unstructured data.

Thus, this paper presents the development direction of the driving environment prediction platform based on the road traffic big data, which meets the technology development trend and the social demand, as well as the corresponding expected effects. Towards that end, the levels and trends of the development of related big-data platform technologies in Korea and globals were analyzed, and the differentiation and expected effects of the platform that was developed in this study were examined.

2 Differentiation of Technology

In this section, the development method and direction of the big-data platform that was developed in this study are presented. In particular, it is distinguished from the existing platforms and related technologies through the trend analysis of Korea and globals road traffic big-data platforms.

In Korea road traffic field, platforms for collecting and analyzing public data and related big data have been developed and are being provided. Such platforms, however, can process only the public data provided by the existing systems. A platform for the convergence and analysis of two or more kinds of data has yet to be developed. Thus, this study aimed to develop a platform capable of collecting, combining, and analyzing multi-species big data consisting of structured data (e.g., public data) and unstructured data (e.g., the data obtained by the sensors of individual vehicles, social-media data, etc.). This study thus sought to boost the reliability of the prediction and analysis of road driving situations through the development of a pertinent technology.[2]

In addition, thanks to the recent IT advances, the road traffic field is changing from the information collection system based on the existing fixed-type sensors (e.g., loop and image detectors) to the information collection system based on the mobile-type sensors (e.g., smartphones and navigators). Thus, a platform capable of collecting and analyzing real-time data obtained from individual vehicles' sensors instead of using fragmental data provided by the existing systems was developed. Real-time weather and traffic density data are thus sought to be used to achieve faster and more reliable prediction and analysis of data.[3]

3 Development Methods

The development of a road driving environment prediction platform in this study was divided into the development of collection tools, the development of storage tools, and the development of analysis tools. First, to develop a platform distinct from the existing parallel technologies, the data to be gathered for use in the development of the target platform had to be separated and selected. Towards that end, the target road traffic data were classified according to the collection methods to be used and the contents.

The big-data collection tools that were used for the development of the aforementioned platform were vehicle sensors collecting the road surface temperatures, atmospheric temperatures, precipitation, traffic density, and traffic speed data in real time. The data collection tool was developed according to the type of data. Vehicle sensor data were designed to be collectable through a tool developed for collecting such data through the RESTful-type interface and Kafka. Public data were designed to be collectable through the developed and optimized Flume-based data collection agent. As the collected public data are limited in size and the data are collected in real time at least every 5 minutes, the data were designed to be separable and storable according to the collection purpose, using Flume. Vehicle sensor data were designed to register data in the RESTful way using REST Server and Kafka Broker because they need to be distributable and processable as the number of vehicles increases. Kafka provides a distributed processing function in an environment where large amounts of data are being inputted, thereby preventing data loss.

The big-data storage tool for the development of the target platform was developed so that the collected big data could be classified (according to the data type or use purpose) into HDFS (Hadoop Distributed File System) or NoSQL (Not Only Structured Query Language) for storing in the corresponding databases. <Figure 1> shows the process conceptual diagram of the tool that was developed in this study for collecting and storing traffic big data.

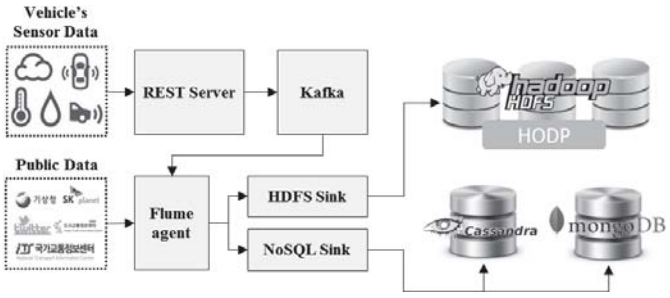

Fig. 1. Process of Data Store Tool

Before being stored in the integrated database, the vehicle sensor and public data have to undergo a pre-treatment process at Spark so that all the collected data will have identical time/spatial storage units. Based on such stored integrated database, the road driving environment is analyzed using the Zeppelin tool. Also, the information display and visualization of the analysis results are provided using the Web-based GIS Map. In the case of processing and handling big data, a phase-matching methodology was developed to construct data with diverse collection types in identical time and space units (5 minutes, standard link). <Figure 2> shows the development conceptual diagram for the traffic big-data analysis tool that was developed in this study.

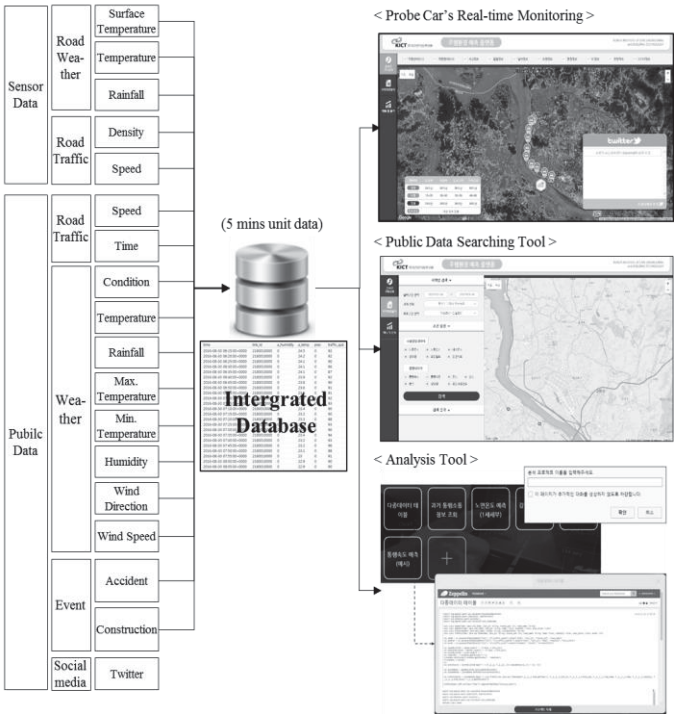

Fig. 2. Concept Map of Road Traffic Analysis Platform

4 Conclusion and Expected Effects

The suggested road driving environment prediction platform can provide traffic information services distinct from the existing systems, thereby boosting the general users' and experts' access to traffic data. It is expected that it can be used in the information-system related industries and in the markets in diverse fields. In addition, the platform can use road traffic public, sensor, and unstructured data to provide meticulous road driving environment information, to improve the drivers' safety, and to boost the reliability of traffic prediction information. In the future, Open API based visualization platforms and a related-platform advanced technology should be developed. In addition, it is necessary to build and apply various algorithms for customized driving environment analysis, and to develop more reliable traffic data platforms through a continuous verification process.

5 References

[1] Korea Embedded Software and System Industry Association, "KESSIA ISSUE REPORT", DEC 2014.

[2] S.R. Kim, M.M. Kang, "Today and the Future of Big Data Analytics Technology", The Korean Institute of Information Scientists and Engineering, vol. 1, pp. 8-17, 2014.

[3] Korea Institute of Civil Engineering and Building Technology, "Development of Driving Environment Prediction Platform based on Big Data", 2016.