

## Design and Implementation of a Big Data Analytics Framework based on Cargo DTG Data for Crackdown on Overloaded Trucks

Bum-Soo Kim\*

\*Researcher, Dept. of Future Technology and Convergence Research,  
Korea Institute of Civil Engineering and Building Technology, Gyeonggi, Korea

### [Abstract]

In this paper, we design and implement an analytics platform based on bulk cargo DTG data for crackdown on overloaded trucks. DTG(digital tachograph) is a device that stores the driving record in real time; that is, it is a device that records the vehicle driving related data such as GPS, speed, RPM, braking, and moving distance of the vehicle in one second unit. The fast processing of DTG data is essential for finding vehicle driving patterns and analytics. In particular, a big data analytics platform is required for preprocessing and converting large amounts of DTG data. In this paper, we implement a big data analytics framework based on cargo DTG data using Spark, which is an open source-based big data framework for crackdown on overloaded trucks. As the result of implementation, our proposed platform converts real large cargo DTG data sets into GIS data, and these are visualized by a map. It also recommends crackdown points.

▶ **Key words:** Big data, Cargo DTG, Spark framework, Preprocessing, GIS

### [요 약]

본 논문에서는 과적 화물차량 단속을 위한 대용량 화물 DTG 데이터 분석 플랫폼을 설계 및 구현한다. DTG(digital tachograph)는 차량운행기록을 실시간으로 저장하는 장치로서, 차량의 GPS, 속도, RPM, 제동유무, 이동거리 등 차량운행 관련 데이터가 1초 단위로 기록된다. 차량 운행 패턴 및 분석을 하기 위해서는 DTG 데이터의 빠른 처리가 필수적이며, 특히 대용량 DTG 데이터를 가공 및 변환하기 위해서는 빅데이터 분석 플랫폼이 필요하다. 본 논문에서는 오픈소스 기반의 빅데이터 프레임워크인 스파크(Spark)를 이용하여 과적차량 단속을 위한 대용량 화물 DTG 데이터의 분석 플랫폼을 구현하였다. 구현 결과, 실제 대용량 화물 DTG 데이터를 GIS 데이터로 변환하여 지도상에 표현하고 단속 추천 지점을 보여준다.

▶ **주제어:** 빅데이터, 화물차 DTG, 스파크 프레임워크, 전처리, GIS

- 
- First Author: Bum-Soo Kim, Corresponding Author: Bum-Soo Kim
  - \*Bum-Soo Kim (bumsookim@kict.re.kr), Dept. of Future Technology and Convergence Research, Korea Institute of Civil Engineering and Building Technology
  - Received: 2019. 11. 25, Revised: 2019. 12. 16, Accepted: 2019. 12. 16.
  - The preliminary version of this paper was published in Proc. of the Korea Society of Computer and Information Summer Conference 2019, Jeju, Korea, pp. 287-288, Jul. 2019.

## I. Introduction

데이터[1]는 네트워킹, 데이터 저장 및 데이터 수집 용량의 빠른 발전으로 물리, 생물 및 의학을 포함한 모든 과학 및 공학 분야에서 급속도로 확대되고 있다. 또한, 숨겨진 가치에 대한 심층적인 이해를 돕고 새로운 가치를 발견할 수 있는 기회를 제공할 수 있도록 데이터셋을 효과적으로 구성하고 관리하는 과정을 내포하고 있다.

최근 건설 분야에서는 과적단속 예방을 위해 화물 DTG 데이터를 활용한 빅데이터 분석 연구가 진행되었다[2]. 이 연구[2]에서는 화물 DTG, 교통량, 검차 및 단속 건수 데이터를 분석하여 과적단속 적정지점을 과적단속 담당자에게 추천함으로써 효율적인 과적단속 업무를 지원할 수 있도록 하였다.

과적차량은 도로법 제77조 및 동법 시행령 제79조 제2항에 근거하여 축하중이 10톤을 초과하거나, 총중량이 40톤을 초과하는 차량(계측기 및 측정오차 등을 감안 1할의 범위 내에서 허용)이거나 차량의 폭이 2.5m, 높이가 4.0m(도로구조의 보전과 통행의 안전에 지장이 없다고 관리청이 인정하여 고시한 도로노선의 경우에는 4.2m 기준), 길이가 16.7m 초과하는 차량을 단속하게 된다.

과적차량은 도로 포장파손 피해 및 교통사고의 주요 요인으로 단속을 위해 고정, 이동식 단속을 실시 중이나, 현재는 과적 검차 및 단속실적만 입력, 관리하고 있어 과적 예방 기능이나 단속지원 기능은 없는 상태이다[2]. 이에 따라, 검차 및 단속정보, 교통량정보 등 연계정보를 분석하여 화물차 이동경로 예측모델 개발 등 운행제한 차량의 단속 업무를 지원하는 체계가 필요하다.

본 논문에서는 과적단속 업무를 지원하기 위해 대표적인 오픈소스 빅 데이터 플랫폼인 스파크(Spark)[3]를 기반으로 하는 대용량 화물 DTG 데이터 분석 플랫폼을 설계 및 구현한다. 화물 DTG 데이터는 DTG 데이터의 특성상 사이즈가 워낙 방대하여 데이터를 변환 및 가공하는 것이 쉽지 않다. 따라서, 스파크를 이용한 빅 데이터 플랫폼에서는 다음과 같은 작업 절차에 따라 수원 국토관리사무소 [4] 영역 내의 화물차량의 이동 경로를 분석한다. 먼저, 본 논문에서는 1) 한국 내 존재하는 DTG 데이터를 얻고, 2) 운행정보가 많은 차량(30분 이상 주행차량)만을 대상으로 삼고, 3) 화물단지로부터 출발/도착한 차량만을 골라 4) 1초 단위 데이터를 10초 단위 데이터로 변환하여 5) 수원 국토관리사무소 영역 반영하는 작업을 수행한다. 또한, DTG 통행량을 바탕으로 화물차량의 이동 경로를 분석하여 주경로를 파악하고, 이 중 회피한 것으로 의심되는 경로를 파악하여 화물차 단속 회피를 차단하기 위한 도주로

를 사전에 봉쇄하는 이동식 단속 지점과 그에 따른 보조 단속지점을 추천한다.

구현 결과, 약 1TB 정도하는 화물 DTG 데이터를 실시간으로 처리되는 것을 확인하였다. 이는 1초 단위의 위치 데이터로서 단순한 가공, 정제뿐만 아니라 공간 정보로서 실시간으로 이해할 수 있도록 직관적으로 표현하는데 그 의미가 있다. 또한, 맵상에 단속 지점과 보조 단속지점을 추천하여 표시하였다. 이는 제안한 시스템이 매우 실용적이고 사용자 친화적이라 확신한다.

## II. Related works

### 1. Spark framework

최근 빅데이터 분석에서는 성능 향상을 위해 하나의 작업을 여러 대의 컴퓨터로 나누어 처리하는 분산처리 시스템을 사용한다[5, 6, 7]. 대표적인 분산처리 시스템으로는 맵리듀스(MapReduce) 방식[8]을 지원하는 아파치 하둡(Hadoop)[9] 환경의 분산처리 프레임워크가 있다[5]. 하둡은 크게 하둡 파일 시스템(HDFS), OS 수준 앱스트랙션(abstractions), 맵리듀스 엔진으로 구성되며, 빅데이터 분석을 위한 여러 확장 프레임워크가 존재한다[5, 10].

스파크(Spark)는 RDD(resilient distributed dataset)라는 고유의 자료구조를 이용해 빅데이터를 분석하는 분산처리 프레임워크이다[3]. 이는 인-메모리(in-memory) 방식으로 데이터를 한꺼번에 읽어들이 분석 작업을 빠르게 처리할 수 있는 장점이 있다. 또한, 스파크는 늦은 수행(lazy-execution)이라는 특성을 가진다. 이는 그 결과를 호출할 때 비로소 동작을 수행하는 것을 뜻한다.

Fig. 1은 도식화된 스파크 프레임워크의 개념이다. 스파크는 다양한 동작 명령어들의 집합을 구성할 수 있기 때문에 단순한 맵리듀스보다 고차원적인 접근이 가능하다. 즉, 수행을 요청하면 미뤄둔 동작 명령어들의 연산을 한꺼번에 수행이 가능하다. 이는 스파크가 빅데이터 처리를 위해 보다 효율적인 업무 배치 여지를 두기 위해서이다. 또한, 스파크는 자바, 스칼라, 파이썬, R 등 프로그래밍에 친화적인 API를 제공하며, SQL, 스트리밍 데이터, 머신러닝 및 그래프 처리를 지원한다. 본 논문에서는 빅데이터 처리뿐만 아니라 통계분석까지 활용할 수 있는 SparkR API를 사용한다.

스파크는 다양한 분야의 분석 환경을 제공한다[3]. 대표적으로 Geo-Spark는 스파크를 이용한 공간 데이터 분석이 가능하다. 즉, Geo-Spark는 공간정보를 보다 쉽게 접근하고 활용할 수 있도록 Spatial-RDD, Spatial SQL의

사용이 가능하며, 다양한 공간 객체의 처리가 가능하다. 또한, 스파크와 마찬가지로 Geo-Spark는 여러 가지 데이터 형식을 제공하며, 일반적으로 사용되는 공간 데이터 파일 포맷을 모두 지원한다. 본 논문에서는 Geo-Spark를 사용하여 화물차량의 GPS와 새주소 도로 형상을 적재 및 처리하며, WGS84 좌표계[11] 기반에서 미터단위의 평면 좌표계로 통일하여 분석한다.

## 2. Freight vehicle pattern

Table 1. The specification of DTG

Description	Type	Unit
Date	String	-
Cumulative mileage(km)	Unsigned integer	km
Cumulative mileage(m)	Unsigned integer	m
Daily mileage	Unsigned integer	km
Car speed	Unsigned integer	km/h
The number of RPM	Unsigned integer	RPM
Brake signal	Unsigned integer	-
GPS x axis	Integer	-
GPS y axis	Integer	-
GPS azimuth	Unsigned integer	-
Acceleration x axis	Integer	m/s <sup>2</sup>
Acceleration y axis	Integer	m/s <sup>2</sup>

DTG은 국내 여객자동차 운수사업법과 화물자동차 운수사업법에 의해 업무용 화물차량 예외 차량을 제외하고 의무적으로 장착해야 되는 기록장치이다. DTG은 화물차량의 안전운전과 분쟁시 활용 및 효율적인 배차관리 등 그 사용 목적이 다양하다. 특히, 교통사고의 주요 원인 중 하나인 운전자의 운전습관을 교정하기 위한 안전운전 교육 자료로 활용되고 있다. 화물 DTG 데이터 분석을 위해 사용하는 DTG 데이터 요소들을 Table 1과 같다. DTG 데이터의 특징은 운행차량의 상황을 파악할 수 있도록 차량 상태정보와 위치 정보가 함께 있는 것이다.

화물차량의 이동 경로를 경로화하기 위한 전처리 작업으로 Han et. al.[12]에서는 DSRC(dedicated short-

range communications) 자료를 개별차량의 경로자료로 변환하고 개별 차량에 대한 경로와 패턴을 생성했다. 이를 활용해 교통운영관리의 고도화 방안과 도로 구간별 경로를 바탕으로 한 국도 우회패턴 도출 등의 결과를 제시하였다. 또한, 회전 교통량, 기종점간의 통행경로, 우회 패턴, 휴게소 이용 행태 등을 분석하여 수행한 결과에 대한 신뢰성 검증을 함께 제시하였다.

화물차량의 이동 경로를 분석하기 위해 Cao et. al.[13]는 수집한 GPS 데이터를 도로 네트워크화 하는 방법론을 제시하였다. 이는 GPS 잡음을 최소화하여 그래프 생성 알고리즘을 활용하였다. 또한, 교통량, 방향성, 도로폭, 연결성 등의 정보를 속성화하여 GPS 신호를 이용하여 도로 네트워크를 구축하였다. Krumm et. al. [14]는 GPS의 신호가 다음 신호로 넘어 갈 때 이동 가능한 도로로의 확률의 순열을 생성하고 이에 Hidden-Markov 모델[15]을 적용하여 맵 매칭 오차를 최소화하였다. Hidden-Markov 모델은 시스템이 은닉된 즉, 결과를 야기하는 직접적인 원인은 확인이 불가능한 상태와 관찰 가능한 결과로 이루어졌다고 간주하는 것이다. 이는 시간의 흐름에 따라 변화하는 시스템의 패턴을 인식하는 작업에 유용하다.

최근 Kim et. al. [2]은 과적단속 예방을 위해 향후 화물차 통행량을 독립 변수로 두어 로지스틱 회귀분석[16]을 사용하여 단속지점을 예측하는 방법을 제시하였다. 본 논문에서는 화물 DTG 데이터의 통행량을 기반으로 화물차의 단속 회피 패턴을 사전에 파악해 단속지점을 추천하는 알고리즘을 구현한다.

## III. The proposed big data analytics framework

Fig. 2는 과적 화물차량 단속을 위한 화물 DTG 데이터 기반의 빅데이터 분석 프레임워크 개념도를 나타낸다. 분석

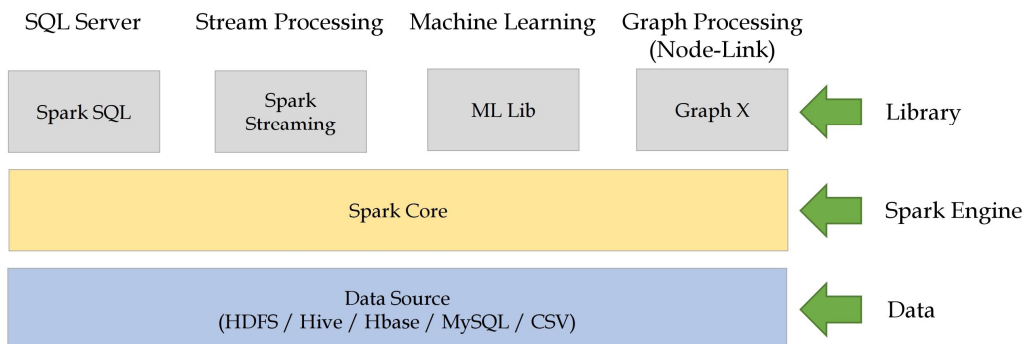


Fig. 1. The concept scheme of Spark framework.

과정은 크게 빅데이터 전처리, 단속회피 패턴 분석, 주/보조 동선 추출과정으로 나눌 수 있다. 본 논문에서 DTG 데이터는 화물차량의 위치와 상태 값을 10초 단위로 변환하여 사용한다. DTG 데이터 상의 화물차량의 위치 값은 GPS를 이용하여 취득한 좌표로 실제 차량 위치와 다를 수 있다. 이는 GPS 신호가 건물에 반사되거나, 차량이 터널, 지하 공간 등 음영 구간에 존재할 때와 차량이 한 자리에 정차해 있더라도 약 10m 이내일 때에서 자주 발생한다. 따라서, 본 논문에서는 맵 매칭 알고리즘[17]을 사용하여 DTG 데이터의 위치 값을 보정한다. 맵 매칭 알고리즘은 기록된 GPS 정보의 이상치나 오류 신호를 제거한 링크데이터를 제공하여 실제 도로와 일치시키는 방법이다. 이러한 오류들을 해결하기 위해 도로 데이터에서는 새 도로주소와 표준노드링크 데이터를 활용한다. 이는 DTG 데이터의 위치정보와 함께 보다 자세한 도로 정보를 제공한다. 즉, 단순히 표준노드링크만을 일치시키는 것보다 지방도 이상의 세밀한 단위의 도로 데이터를 활용하여 화물차량의 이동경로를 파악할 수 있다.

차량 이동경로를 파악하기 위해 DTG 위치 데이터와 도로 데이터들은 다양한 전처리 과정이 필요하다. 먼저, 1) 시간이 지나도 차량의 위치 좌표가 같으면 한 자리에 있다고 판단하여, 중복된 좌표 데이터는 하나로 합친다. 이때, 중복된 데이터 중 가장 이른 시간의 도로진입시간과 가장 나중의 도로진출시간만을 선택하여 사용한다. 2) 또한, 짧은 도로구간(이하 '링크')과 GPS 오차로 인한 맵 매칭 오류를 제거하기 위해 GPS 데이터들이 링크에서 시간 흐름에 따라 연속적으로 나타나는 것이 아니라 불연속적으로

나타나는 데이터들은 제거한다. 3) 또한, 시간이 지나도 중복된 새 도로주소 ID 값의 링크들이 존재하는 차량들은 앞서 중복된 좌표 데이터처럼 하나로 합친다. 4) 마지막으로, 주요 도로가 아닌 고속도로 램프 등의 연결구간이나 짧은 교차로 구간을 제거하기 위해 통과 시간이 1분 미만인 링크는 제거한다. 시간을 1분 미만으로 정한 기준은 화물차량의 전체적인 이동경로를 파악하는데 1분 이상의 데이터만으로도 충분하다는 경험적 판단에 따른다.

산업단지 데이터를 사용하는 이유는 토석 및 폐기물 등 과적 화물차량의 주요 적재물들이 일반 지역보다 많이 생성된다는 직관에 기인한다. 따라서, 산업 단지 권역을 파악하기 위해 특정 주소를 바탕으로 속성정보를 포함한 좌표 정보로 변환시켜 주는 주소정제 솔루션을 이용하여 산업체 주소를 공간 정보화 시킨다. 여기서, 변환된 산업체 공간정보는 Convex Hull 알고리즘[18]을 사용하여 클러스터링하여 권역을 생성한다. Convex Hull은 점 좌표의 밀도를 기반으로 권역을 생성하는 알고리즘이다.

단속 데이터는 분석을 위해 단속 정보를 정제하여 단속 지점을 점과 선으로 표현한다. 단속 정보 중 주소는 주소 정제 솔루션을 사용하여 점 좌표로 변환한다. 주소 정제는 행정동 단위까지 매핑된 데이터를 단속지점으로 사용한다. 또한, 특정 구간의 단속 지점은 선 형상으로 변환한다. 먼저, 차량의 출발지와 도착지의 조합으로 기록된 검문장소를 분해하여 분절된 각각의 데이터와 표준노드링크를 매핑한다. 이를 단속 구간이라 정의한다. 이때, 표준노드링크와의 매핑을 위해 본 논문에서는 국도만을 대상으로 한다.

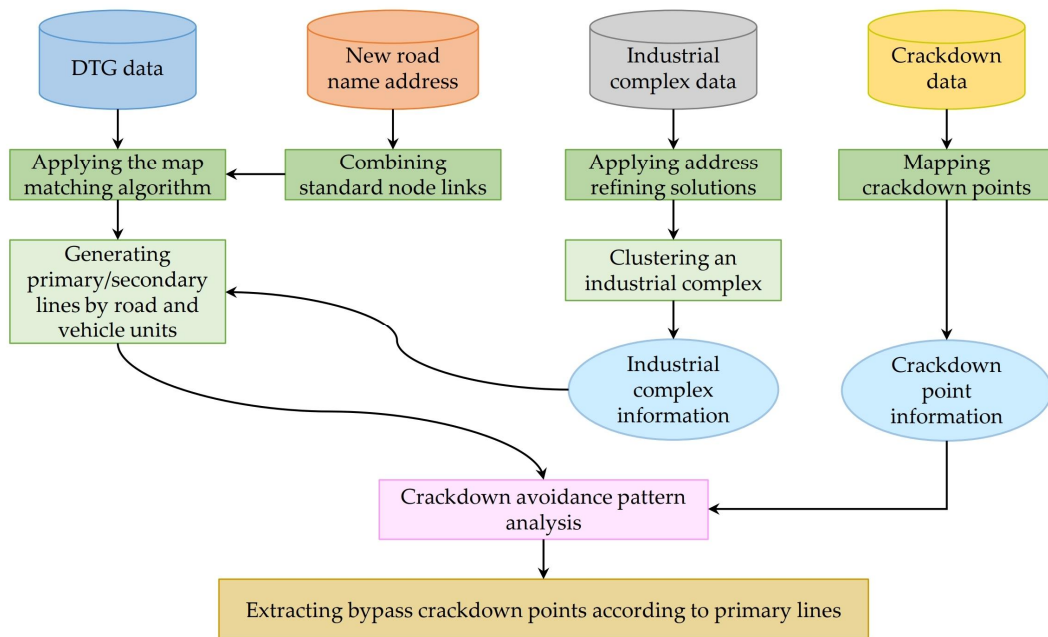


Fig. 2. The concept scheme of a big data analytics framework for crackdown on overloaded trucks.

Table 2. The example of DTG data.

date	code	car	time	day_dist	cum_dist	speed	RPM	breaks	x	y	angle	x_boost	y_boost
20160901	727	33648FCCE3A2F630BF 41329EAE921530	5152700	1	452391	0	426	0	128.6063	38.16277	0	1	1.3
20160901	727	33648FCCE3A2F630BF 41329EAE921530	5153000	0	452391	0	600	0	128.6063	38.16277	0	0	-0.2
20160901	727	33648FCCE3A2F630BF 41329EAE921530	5154000	0	452391	0	599	0	128.6063	38.16277	0	0	0.5
20160901	727	33648FCCE3A2F630BF 41329EAE921530	5155000	0	452391	0	600	0	128.6063	38.16277	0	0	0
20160901	727	33648FCCE3A2F630BF 41329EAE921530	5160000	0	452391	0	598	0	128.6063	38.16277	0	0	0
20160901	727	33648FCCE3A2F630BF 41329EAE921530	5161000	0	452391	0	600	0	128.6063	38.16277	0	0	0
20160901	727	33648FCCE3A2F630BF 41329EAE921530	5162000	0	452391	0	598	0	128.6063	38.16277	0	0	0
20160901	727	33648FCCE3A2F630BF 41329EAE921530	5163000	0	452391	0	600	0	128.6063	38.16277	0	0	0
20160901	727	33648FCCE3A2F630BF 41329EAE921530	5164000	0	452391	0	599	0	128.6063	38.16277	0	0	0

단속회피 분석을 위해서는 먼저 기종간 경로를 추출하는 과정과 회피패턴을 도출하는 과정이 필요하며, 이를 기반으로 과적 단속지점을 도출한다. 기종간 경로 추출은 O/D 기준으로 화물차량을 분류한다. 각 화물차량은 많은 도로를 지나므로 이 중 가장 많이 다니는 새주소 도로의 조합을 주경로로 정의한다. 보조 경로는 주경로와 80%이상 유사한 도로를 지나는 도로로 정의한다. 회피패턴 도출에는 이동식 단속구간을 회피하는 경향을 크게 회피 경로와 정차 지점으로 나눈다. 회피 경로는 우회경로이면서 단속 지점에 따른 화물 차량의 통행량의 변화가 존재하는 지점으로 정의하며, 정차 지점은 휴게소 및 화물단지 인근 상/하차 지점 제외한 차량이 장시간 정차하는 지점으로 정의한다. 과적 단속지점 도출에서는 2개소 이상의 합동 단속 지점을 제시한다. 단속 지점 추천 경로는 회피패턴을 기반으로 추천 지점의 고정화를 방지하기 위해 실시간 화물 교통량을 반영해 과적단속 지점 선정 시 우선순위를 반영한다.

#### IV. Implementation of an index-based system for partial denoising boundary matching

구현을 위해 본 논문에서는 2.1GHz Intel Xeon Silver 4110 2CPU, 288GB 램, 약 2TB SSD 하드 디스크 사양의 두 대의 서버로 클러스터를 구성하여 사용하였다. 데이터 저장장을 위해서는 96TB의 NAS 서버를 사용하였다. 운영체제는 Ubuntu 18.04.1 LTS이며, 빅데이터 처리를 위한 Apache Spark Framework과 데이터베이스로 Postgres SQL과 PostgresGIS를 함께 연동하여 환경을 구축하였다. 하였다. 분석 언어로는 R 패키지 3.5.1와 파이썬(Python) 3.7.3를 사용하였으며, 주소정제 솔루션은 오픈메이트의 Geocoding[19], 맵 서버로는 GeoServer[20]를 설치하였다.

분석에 사용한 DTG 데이터는 최근 2년간 국내에서 제출된 화물차량 DTG 데이터로서, 약 15TB의 크기를 가진다. DTG 데이터는 용량이 매우 크기 때문에 전처리 과정에서 많은 시간이 소요된다. 따라서, 보다 직관적인 전처리를 수행하기 위해서는 DTG 데이터를 시각화하는 것이 효과적이다. Table 2는 DTG 데이터의 예제 값을 나타낸다. 본 구현에서는 이동경로를 파악하는데 큰 어려움이 없는 선에서 10초 단위로 화물차별 일자와 시간, 좌표 값만을 사용한다.

새주소도로와 표준노드링크 정보는 행정안전부에서 제공하는 데이터를 기준으로 필요한 속성 및 데이터만을 골라 이용한다. 본 분석에서는 DTG 데이터와 새주소도로 데이터와의 공간 조인(spatial join)을 하여 20m 이내의 새주소도로 링크 중 가장 가까운 도로를 지나도록 한다. Table 3은 전처리를 거친 DTG 링크 데이터 예제이다. 여기에서 각각의 속성들의 설명은 다음과 같다.

index: 원본 데이터 추적을 위한 순번으로 차량별로 구분하여 1부터 부여

- CAR\_ID: 비식별화된 차량 고유 번호
- DTG\_DT: DTG 일자
- DTG\_TIME: DTG 시간
- RN\_CD: 새주소 도로 고유 번호
- X\_AXIS: X좌표 (미터 단위 평면 좌표계)
- Y\_AXIS: Y좌표 (미터 단위 평면 좌표계)
- fromTime: 도로구간 진입시간
- toTime: 도로구간 진출시간
- lapTime(min): 소요 시간
- distance(m): 직선 이동 거리
- speed(km/h): 구간 평균 속도

본 구현에서는 공공데이터포털[21]에서 얻은 총 65,535개의 산업단지 데이터 중 주소정제 및 권역화하여 총 311개의 산업단

Table 3. The example of processed DTG data.

index	CAR_ID	DTG_DT	DTG_TIME	RN_CD	X_AXIS	Y_AXIS	fromTime	toTime	lapTime (min)	distance (m)	speed (km/h)
1	DB3FA11C289D050B2CE2900166D438A6	20160901	60740	3210136	315441	515471	60740	62010	21.2	2211	6.3
61	DB3FA11C289D050B2CE2900166D438A6	20160901	62110	2012008	313230	515442	62010	62440	7.2	2155	18
77	DB3FA11C289D050B2CE2900166D438A6	20160901	62450	3175029	313467	517584	62440	64240	30	618	1.2
115	DB3FA11C289D050B2CE2900166D438A6	20160901	64300	3175009	314051	517382	64240	64700	7.7	1780	13.9
147	DB3FA11C289D050B2CE2900166D438A6	20160901	65010	3175016	313892	515609	64700	65140	7.3	844	6.9
159	DB3FA11C289D050B2CE2900166D438A6	20160901	65150	3012006	313269	515039	65140	65450	5.2	1338	15.5
177	DB3FA11C289D050B2CE2900166D438A6	20160901	65530	2012013	313847	513832	65450	70910	91	9419	6.2
262	DB3FA11C289D050B2CE2900166D438A6	20160901	70920	3198011	317956	505356	70910	71010	1.7	348	12.5
267	DB3FA11C289D050B2CE2900166D438A6	20160901	71020	2012005	318298	505290	71010	71550	9	1107	7.4
303	DB3FA11C289D050B2CE2900166D438A6	20160901	71600	2012003	318311	504183	71550	72030	8	1227	9.2
331	DB3FA11C289D050B2CE2900166D438A6	20160901	72930	3188063	319070	503219	72030	73320	21.5	419	1.2
355	DB3FA11C289D050B2CE2900166D438A6	20160901	73330	2012003	319334	503544	73320	73640	5.3	1026	11.5
376	DB3FA11C289D050B2CE2900166D438A6	20160901	73650	3198017	318559	504216	73640	73840	3.3	337	6.1
388	DB3FA11C289D050B2CE2900166D438A6	20160901	73900	2012005	318313	504447	73840	74550	11.8	6262	31.8
436	DB3FA11C289D050B2CE2900166D438A6	20160901	74720	3210025	318886	510683	74550	74930	6.3	722	6.8
452	DB3FA11C289D050B2CE2900166D438A6	20160901	81800	3210170	318187	510864	74930	82250	122	482	0.2
465	DB3FA11C289D050B2CE2900166D438A6	20160901	82300	3210027	317808	510566	82250	82530	4.7	1270	16.3
479	DB3FA11C289D050B2CE2900166D438A6	20160901	82810	3210169	317763	511835	82530	82920	6.5	315	2.9

지를 사용하였다. Table 4는 전국 산업단지 분포 결과이다. Fig. 3은 이러한 산업단지와 화물차량 도로를 시각화한 화면이다. 그림을 보면, 주황색으로 산업단지를 구분하였으며, 산업단지 근처로 지나가는 화물차량 도로를 특정하여 표시하였다.

Table 4. Result of industrial complex data.

Region	the number of industrial complex data
Busan City	1
Daegu City	9
Gwangju City	7
Daejeon City	3
Gyeonggi-do	59
Gangwon-do	47
Jeollanam-do	1
Chungcheongbuk-do	96
Gyeongsangnam-do	88

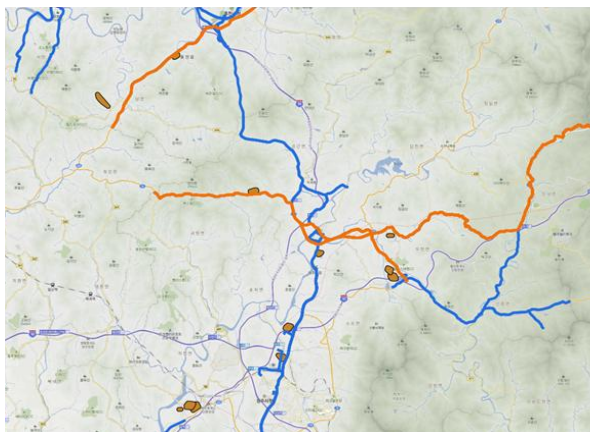


Fig. 3. The visualization of industrial complex data.

한편, 과적단속 지점을 추천하기 위해 본 구현에서는 건설사업정보시스템[22]에서 보유하고 있는 기존 단속 데이터를 활용하였다. Table 5는 단속 데이터의 예제이다. 본 구현에서는 단속 데이터 내용 중 검문일자과 검문장소 소성을 이용하였다. 검문장소는 주소정제 솔루션을 이용하여 좌표값으로 변환하여 이를 단속 지점에 활용하였다.

Fig. 4는 단속 지점을 구간화하여 시각화한 화면이다. 그림을 보면, 보라색 선이 단속 지점을 구간으로 표시한 것이다. 결국, 단속 지점 구간은 DTG 데이터 링크 구간과 산업단지를 지나가는 도로구간과 함께 과적할 가능성이 있는 구간들을 파악하여 우회할 수 있는 후보경로를 도출하는데 사용된다. 본 구현에서는 도출된 우회 예상 경로들을 바탕으로 단속 추천지점을 나타낸다.

Fig. 5는 구현한 빅 데이터 플랫폼을 이용하여 과적단속 추천지점을 지도상에 표현한 화면 예제이다. 그림을 보면, 수원 국토관리사무소 영역을 설정하여 화물 차량 DTG 데이터를 기반으로 과적단속 회피 우회경로를 파악하여 최적 과적단속 지점을 빨간색으로 표시하였다. 이 결과를 기반으로 화물차량 과적단속 업무 담당자들은 별도의 업무 시스템을 이용해 과적단속 의사결정에 활용할 수 있게 된다. 화물차량 과적단속 분석 결과에 따른 과적단속 업무 결과는 향후 연구로 다룰 예정이다. 결론적으로, 본 논문에서 제안한 시스템은 대용량 화물 DTG 데이터를 다루어 과적 화물차량 단속 분석에 매우 효과적이라 사료된다.

Table 5. The example of crackdown data.

INSPECTDATE	OFFICECODE	CHECKPOINTCODE	UPDOWNWAY	INSLOCALE	ROADNUMBER	CARSTAR TPOS	CARENDPOS	CARUPBOARD	CARSTART TIME
201712221348	1010002	88	LD	일반국도6능내리-양평읍 양평군 양서면 도곡리	6		서울~양평		
201712221309	1010002	88	LD	일반국도6능내리-양평읍 양평군 양서면 도곡리	6		성수~양평		



Fig. 4. The screenshot of crackdown data.

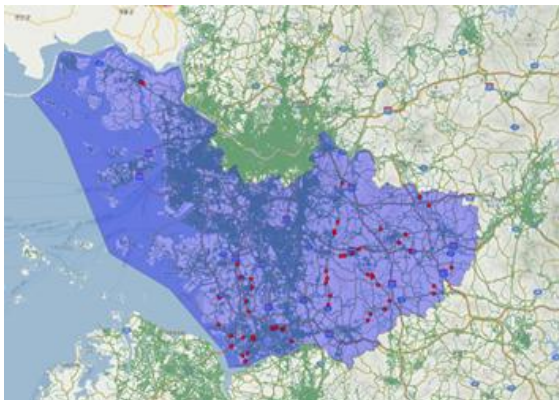


Fig. 5. The screenshot of optimal crackdown points.

## V. Conclusions

본 논문에서는 과적 화물차량 단속을 위한 대용량 화물 DTG 데이터 기반의 빅데이터 분석 플랫폼을 설계 및 구현

하였다. 대용량 DTG 데이터를 실시간으로 가공 및 변환하고 지도상에 표현하기 위해 오픈소스 빅 데이터 플랫폼인 스파크를 활용하였다. 제안한 시스템은 DTG 데이터의 특성을 잘 반영하여 사용자가 쉽게 데이터를 이해하고 사용할 수 있도록 구현하였다. 또한, 국내 최초로 다년간 축적된 건설정보와 공간정보 등을 결합한 도로분야 빅데이터를 기반으로 한 효율적 의사결정 관리기술 개발과 타 기관 확산 기반을 마련하였다는데 의의가 있겠다. 향후 연구로는 본 연구를 기반으로 다른 시설물들의 데이터도 다룰 수 있는 IoT 분석 플랫폼을 설계와 구현할 예정이다.

## ACKNOWLEDGEMENT

This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport(Grant 19SCIP-C146569-02). This work is supported by the digital tachograph data provided by Korea Transportation Safety Authority.

## REFERENCES

- [1] X. Wu, X. Zhu, G. Wu and W. Ding, "Data Mining with Big Data," IEEE Transactions on Knowledge and Data Engineering, Vol. 26, No. 1, pp. 97-107, Jan. 2014.
- [2] T.-H. Kim, B.-S. Kim, and J.-U. Kim, "A Study on the Analysis Method for the Optimal Location of Overstrength Restrictions Using Cargo DTG," In Proc. of the KAIS Fall Conference, pp. 1-3, Nov. 2018(In Korean).
- [3] Apache Spark, <https://spark.apache.org/>
- [4] Suwon National Territory Management Office, <http://www.molit.go.kr/srocm/intro.do>
- [5] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward Scalable Systems

- for Big Data Analytics: A Technology Tutorial," IEEE Access, Vol. 2, No. 8, pp. 652-687, July 2014.
- [6] L. Dagum and R. Menon, "OpenMP: An Industry Standard API for Shared-Memory Programming," IEEE Comput. Sci. & Eng. Vol. 5, No. 1, pp. 46-55, Jan./Mar. 1998.
- [7] G. Malewicz, M. H. Austern, A. J.C. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski, "Pregel: a system for large-scale graph processing," in Proc. of the 2010 ACM SIGMOD Int'l Conf. on Manag. Data, Indianapolis, Indiana, pp. 135-146, June 2010.
- [8] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Communications of the ACM, Vol. 51, No. 1, pp. 107-113, Jan. 2008.
- [9] Apache Hadoop, <https://hadoop.apache.org/>
- [10] Apache Mahout, <http://mahout.apache.org/>
- [11] NGA Office of GEOINT Science (WGS 84), <http://earth-info.nga.mil/GandG/wgs84/index.html>
- [12] D.-H. Han, S.-H. Kim, J.-J. Park, J.-H. Lee, and J.-H. Kim, "A Study on the Traffic Analysis Method Using Vehicle Trajectory Data," Expressway & Transportation Research Institute, 2017.
- [13] L. Cao and J. Krumm, "From GPS Traces to a Routable Road Map," in Proc. of the 17th ACM SIGSPATIAL Int'l Conf. on Adv. in Geo. Inf. Sys., Seattle, Washington, pp. 3-12, Nov. 2009.
- [14] J. Krumm, J. Letchner, and E. Horvitz, "Map Matching with Travel Time Constraints," SAE Technical Paper 2007-01-1102, 2007.
- [15] S. R. Eddy, "Profile Hidden Markov models," Bioinformatics Review, Vol. 14, No. 9, pp. 755-763, Sept. 1998.
- [16] S. Menard, "Applied Logistic Regression Analysis," SAGE Publishing, 2002.
- [17] Y. Lou, C. Zhang, X. Xie, Y. Zheng, W. Wang, and Y. Huang, "Map-Matching for Low-Sampling-Rate GPS Trajectories," in Proc. of the 17th ACM SIGSPATIAL Int'l Conf. on Adv. in Geo. Inf. Sys., Seattle, Washington, pp. 352-361, Nov. 2009.
- [18] G. Branko, "Convex Polytopes, Graduate Texts in Mathematics," Springer, 2003.
- [19] Geocoding XGA Solution, <http://www.openmate.co.kr/>
- [20] OGC GeoServer, <http://geoserver.org/>
- [21] Open Data Portal, <https://www.data.go.kr/>
- [22] Construction CALS System, <https://www.calspia.go.kr/>

## Authors



Bum-Soo Kim received his Ph. D. (2013) degrees in computer science from Kangwon National University. From 2013 to 2017, he was a postdoctoral researcher in Korea Advanced Institute of Science and

Technology (2013 and 2015), Kangwon National University (2014), and Korea University (2016-2017). He is currently a postdoctoral researcher in Department of Future Technology and Convergence Research from Korea Institute of Civil Engineering and Building Technology (KICT). His research interests include time-series data mining, construction bigdata analysis, and data mining applications.