

## DTG 빅데이터 분석 기반 지역 교통지도 확대

조원희<sup>1</sup> · 최은미<sup>1,2\*</sup>

<sup>1</sup>국민대학교 비즈니스IT전문대학원

<sup>2</sup>국민대학교 경영정보학부

whcho@kookmin.ac.kr; emchoi@kookmin.ac.kr

(2015년 1월 15일 접수; 2015년 3월 20일 수정; 2015년 3월 27일 채택)

**요약:** 빅데이터의 적용과 활용 분야는 기업과 공공 분야에 이어서 점차 개인 생활의 영역에 접근을 하여 활용 분야가 확대되고 있다. 본 연구는 최근 활성화된 스마트폰 내비게이션의 교통정보를 이용한 빠른 길 안내 이용 시 상대적으로 교통정보 품질이 취약한 지방지역의 교통지도 커버리지를 확대하는 목적을 가지고 있다. 차량이용이 상대적으로 저조한 지방지역의 교통지도 개선을 위하여 DTG 빅데이터를 활용하여 지역 교통지도를 확대하는 접근법을 적용한다. 국내 모든 상용차량에 대해 의무 장착 대상인 디지털 운행기록계 단말로부터 수집된 많은 양의 GPS위치데이터를 빅데이터 기술을 이용하여 속도 패턴을 산출하고 전자지도 기본 도로 속도 값을 교통 패턴 통계 데이터로 교체 반영하는 방안을 제시하였다. 이를 통해 교통정보 확대 및 실시간 교통정보를 이용한 빠른 길 안내 서비스 품질 개선 효과를 기대할 수 있다.

**키워드:** 빅데이터, 경로탐색, 교통지도, Map coverage, DTG

## Rural Traffic Map Coverage Extension using DTG Big Data Processing

Wonhee Cho<sup>1</sup> and Eunmi Choi<sup>1,2\*</sup>

<sup>1</sup>Graduate School of Business IT, Kookmin University, Seoul, Korea

<sup>2</sup>Dept. of Management Information System, Kookmin University, Seoul, Korea

(Received January 15, 2015; Revised March 20, 2015; Accepted March 27, 2015)

**Abstract:** The utilization and application of big data are being gradually expanded from the business and public sector to personal life area. The aim of this paper is to expand traffic map coverage of rural area and to improve routing service quality of smartphone navigation. In the rural area in which vehicle traffic is relatively low, the approach of expanding local traffic map coverage is applied to improve traffic map quality. We use a large amount of GPS trajectory for calculating a traffic pattern from DTG which is a measuring device required on all commercial vehicles, and we introduce a method that is replacing the map default speed value with traffic pattern for routing calculations in a rural province. This research work gives rural traffic map to be expanded traffic coverage and improve the routing service quality.

**Keywords:** Big data, Route planning, Traffic map, Map coverage, DTG

---

\*교신저자

본 과제(결과물)는 교육과학기술부의 재원으로 지원을 받아 수행된 산학협력 선도대학(LINC) 육성사업의 연구결과이며, 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2011-0011507).

## 1. 서 론

2010년 이후 스마트폰 보급이 활성화되고 핸드폰 화면이 커지면서 스마트폰 용 내비게이션이 발전하였다. 스마트폰 용 내비게이션은 특히 모바일 통신 환경이 제공되므로 교통정보를 이용한 실시간 경로탐색이 가능해져 빠른 길 안내에 대한 서비스가 고도화 되어왔다. 그러나 이용자가 늘어남에 따라 빠른 길 및 통행시간 품질에 대한 개선 요구가 늘어났다. 특히 교통정보가 제대로 수집되지 못하는 지방 지역의 경로 품질 저하로 지방지역 사용자의 불만이 해소되지 못하고 있는 실정이다.

한편 한국에서는 2014년까지 모든 상용차량에 의무 장착되어 서비스 중인 항공기의 블랙박스과 유사한 차량의 운행이력을 기록하는 장치인 디지털운행기록계(DTG)로부터 많은 양의 GPS위치데이터가 수집되고 있다[1]. 상용차량은 이동 특성상 지방지역을 많이 운행하고 있다. 그러나 이러한 GPS 주행 궤적 데이터는 위험운전 등 불량운행 통계 위주로 사용되고 교통정보 수집으로는 거의 활용되지 않고 있다. 이러한 모바일 통신관련 데이터 분석에 빅데이터 Map Reduce기술이 활용되고 있다[2,3].

본 연구에서는 실시간 교통정보가 제대로 수집되지 않는 지역도 이러한 DTG GPS궤적 데이터를 활용하여 지방지역의 패턴교통정보를 수집하고 반영하여 빠른길 및 통행시간품질을 높이는 방안을 제시하였다. 데이터를 필터링하는 전처리과정과 맵매칭을 통해 통행속도를 추출하고 요일별, 시간대별, 구간별 패턴교통정보를 생성하는 방법을 제시하였다. 이 방법을 이용하면 기존에 교통정보가 수집되지 않는 지방지역의 경로탐색시 기존에 전자지도 기본 속성값만 이용하던데서 DTG GPS궤적데이터를 이용한 패턴교통정보를 기반으로 실제 값에 좀더 가까운 도로 소통정보를 이용하게 되어 지방지역 경로 품질 향상을 기할 수 있다.

본 연구는 총 5개의 Section으로 구성되어있다. 2장은 통신형 내비게이션 및 DTG 장비와 데이터에 대한 소개이다. 3장에서는 기존 실시간 경로탐색 시스템의 문제점에 대해 살펴보았다. 4장에서는 이러한 문제를 개선할 수 있는 개선방안 및 빅데이터 MR(Map Reduce, MR)를 이용하여 패턴교통정보를 생성하는 방법을 연구하였다. 5장에서는 결과분석을 하였고, 6장에서는 본 연구의 작업을 요약하였고 향후 연구 과제를 기술하였다.

## 2. 관련 연구

### 2.1 통신형 내비게이션의 경로탐색알고리즘

통신형 내비게이션을 위한 경로탐색 프로그램은 Dijkstra[4]와 A\* 알고리즘[5]을 기본으로 하여 교통정보를 가중치로 반영하여 혼잡지역을 회피하는 경로를 탐색한다. 교통정보 맵에서 노드는 교차로 또는 지점을 의미하고, 링크는 두 노드를 연결하는 선을 말한다. 디지털 교통지도는 노드와 링크의 집합으로 이루어진다[6]. 경로탐색은 교통정보뿐만 아니라 다양한 가중치를 활용하여 탐색하게 된다. 교통정보반영 경로탐색은 교통정보가 존재하는 지역에서는 교통정보 속도를 활용하고, 교통정보가 존재하지 않는 지역에서는 도로의 통행 제한 속도 등 도로 기본 속성을 활용하여 경로탐색을 한다.

### 2.2 DTG 장치 소개

디지털 운행기록계 장치는 한국에서 교통사고 줄이기 운동의 일환으로 2014년 6월까지 모든 상업용 차량에 의무 장착하도록 법제화 되어있다.[7] 디지털 운행기록계는 차량이 운행하는 과정에서 생성되는 차량의 GPS좌표, 속도, 가속도, 엔진회전수, 브레이크신호 등의 차량 운행관련 데이터를 1초 간격으로 6개월 이상 저장하고 교통안전공단에 주기적으로 제출해야한다<sup>1)</sup>. 기본적으로 1초 단위 GPS좌표가 기록되므로 주행 궤적을 분석하여 도로 통행 속도 및 시간을 산출 할 수 있다[8].

### 2.3 DTG 분석데이터 소개

본 연구에서 분석에 사용한 이 데이터는 한국에서 상용차량 관제 서비스를 하는 회사의 실제 데이터이다. 4605대의 2013.11~2014.7월 총 9개월 운행 이력 데이터를 이용하였다<sup>2)</sup>. 여기서 Journey는 차량 한대가 하루 주행한 데이터로 파일 하나가 Journey단위로 생성된다. 이 데이터는 3G모뎀을 통해 모바일 통신으로 서버에 전송되고 이것을 별도 공간에 저장하여 수집하였다. DTG를 의무 장착한 차량은 차종이 화물차등 상

<sup>1)</sup> 교통안전공단 DTG데이터 전송을 위한 web site: <http://etas.ts2020.kr>

<sup>2)</sup> DTG 관제서비스 Gtrac website: <http://tacho.gtrac.co.kr>

용차량으로 시내도로보다는 고속도로/국도/지방도를 많이 운행하는 데이터 특성을 갖고 있다.

## 2.4 기존 경로탐색 시스템 이슈

기존 교통정보를 활용한 빠른길 경로탐색 시스템은 교통정보가 주로 수집되는 수도권이나 광역시, 고속도로 주변의 경로는 품질이 높으나 그 외의 지역, 특히 지방의 국도, 지방도 지역의 경로 품질이 저하되는 경우가 많다. 지방지역의 도로 통행속도 계산 값은 기존의 전자지도 도로 속성 default값을 사용하므로 정체지역 발생 시 품질이 저하 될 수밖에 없다. 더우기 이런 지역의 교통정보 수집은 앞으로도 여의치 않아 개선 여지가 많지 않아 보인다. 이러한 지역은 링크 통행속도가 Map Default 속성 값을 이용하므로 정체 시에도 통행 정체 상황을 반영해주지 못한다[9].

## 3. 교통 지도 커버리지 확대 메카니즘

현재 DTG단말로부터 올라오는 대부분의 GPS위치 좌표는 차량 관제용도 외에는 대부분 버려지고 있다. 개선 방안으로 GPS위치 데이터를 활용하여 지방지역 패턴 교통정보로 활용하는 방법을 제시하고자 한다. 주요 교통정보 수집대상 이외 지역의 GPS위치 데이터는 실시간 교통정보로 활용할 수준은 안되지만 일정시간 축적하여 수집하게 되면 통계정보인 패턴교통정보로 활용할 수 있게 된다. 이 누적된 패턴교통정보를 경로 탐색에 활용하게 되면 상대적으로 취약한 지방지역의 경로 및 통행시간 품질 개선에 도움이 될 수 있다고 본다. 이를 위해 준 실시간으로 GPS위치 데이터를 수집하는 방안으로 대용량데이터를 분석 처리할 수 있는 방안을 제시하였다. Figure 1은 빅데이터 기술을 이용한

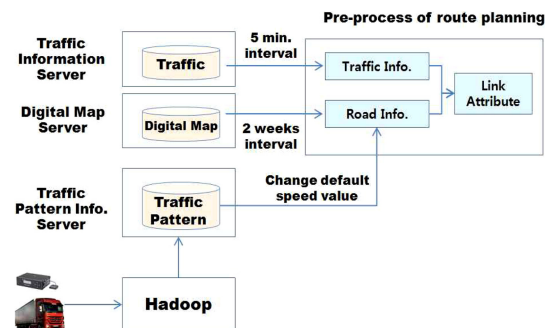


Figure 1. Traffic Information Collection and Service Method

지방지역 교통정보 수집 및 제공 흐름도이다. 지방지역 패턴교통정보가 생성되면 경로탐색용 도로 기본 속성의 도로제한속도 값으로 제공된다[9].

## 3.1 DTG 빅데이터 전처리 단계

DTG 데이터는 이동형 차량 통신 단말이므로 센서 이상, 차량이상, 통신이상 등 실제 수신된 데이터에 오류가 포함되어있을 확률이 높다. 그러므로 오류 detect 및 보정 등의 과정이 필수적이다. 특히 데이터량이 많아 교통정보로 활용할 수 있는 가치가 있는 데이터를 전처리에서 정제하여야 한다.

1단계) 단말 센서 기준값 범위 초과 Detect 및 필터링, 2단계) 통계기법을 이용한 이상치 Detect 및 필터링, 3단계) 패턴생성 필요지역으로 한정하는 필터링의 과정을 거친다.

## 3.2 교통정보 패턴 생성 방법

DTG GPS데이터는 전처리 작업으로 일정부분 데이터를 걸러냈더라도 여전히 방대한 양의 데이터를 처리해야하므로 기존 RDBMS로 처리하기에는 비용 및 시간이 많이 소요된다. 도로 속성 및 고객 통행시간 데이터만 추출해서 전자지도 링크에 매칭하는 기능을 빅데이터 기술을 이용하여 구현함으로써 시간 및 비용 단축할 수 있다. Figure 2는 꺾적 필터링, 맵매칭과 링크별 소통패턴정보를 생성하는 세 단계의 개념도이다.

DTG GPS위치정보는 운송사별, 업종별로 수집이 된다. 1단계 꺾적 Segmenting 단계에서는 종류별로 구분된 위치꺾적을 지역별, 도로별로 필터링 및 분류한다. 방대한 양의 꺾적을 MBR(Most Bounder Rectangle) 단위로 Segmenting 작업을 수행하게 된다. Figure 3을 보면 지역은 일정 규모의 사이즈로 나뉜 직사각형

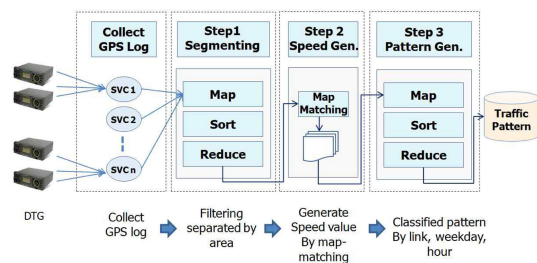


Figure 2. Pattern Generation Method[9]

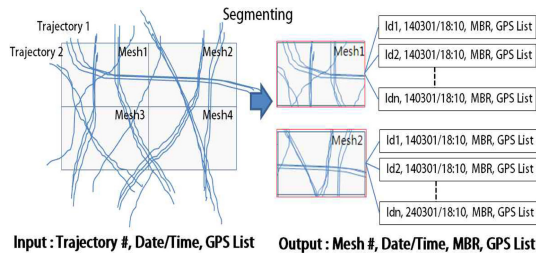


Figure 3. Trajectory Segmentation using MBR

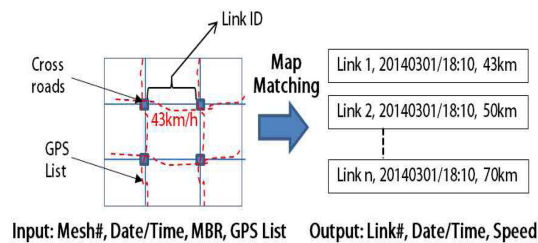


Figure 4. Speed Generation from GPS list

도엽(Mesh)별로 분류하게 된다. 도엽(Mesh)의 우상단과 좌하단의 좌표로 구성된 MBR 값을 기준으로 궤적을 재그룹화 한다. 이 과정을 거치면 Mesh별로 도로번호, 일시, GPS궤적 데이터가 정렬되게 된다.

2단계 속도생성에서는 분류된 GPS궤적을 맵매칭 기법을 이용하여 링크별 속도정보를 생성하게 된다. Figure 4는 도로번호별 속도 생성하는 맵매칭에 대한 그림이다. 맵매칭 방법은 먼저 Mesh내에 있는 모든 Link의 vertex를 기준으로 GPS 좌표와 가장 가까운 vertex 리스트들을 찾아 그 vertex에 해당하는 Link ID를 찾아내는 방법으로 한다. 맵매칭한 결과로 해당 도로번호로 이루어진 link id와 일시, 속도로 분류된다.

3단계 링크패턴 생성은 2단계에서 생성된 링크별 속도정보를 요일별, 시간대별, 도로별 링크패턴으로 통계 처리하여 최종적으로 지방지역 패턴교통정보 DB에 저장된다. 이 패턴교통정보 DB는 주기적으로 경로탐색 서버의 경로탐색용 맵 DB에 전송되어 지방지역을 경유하는 경로 탐색시 활용된다.

## 4. 데이터 처리 결과

### 4.1 전처리 결과

본 연구에 사용한 월별 DTG이동 궤적 로그는 Table 1과 같다. 총 9개월, 660,634개의 로그 파일

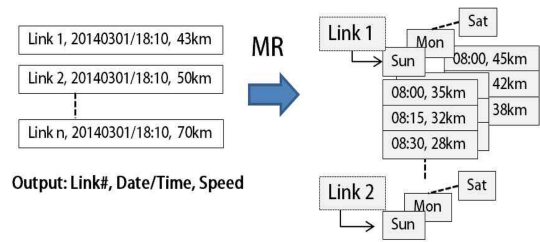


Figure 5. Link Pattern Generation

Table 1. Files and Coordinates Counts

Month	File Count	GPS Coord. Count
1310	69,682	1,517,258,709
1311	45,207	958,038,268
1312	53,001	1,275,556,447
1401	76,506	1,612,633,533
1402	70,849	1,498,286,471
1403	86,204	1,833,835,848
1404	74,511	1,673,004,174
1405	95,464	2,037,652,753
1406	89,210	1,893,590,052
Sum	660,634	14,299,856,255

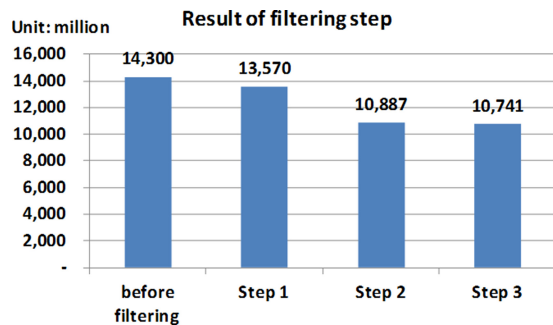


Figure 6. The Amount of Data Filtered

과 14,299,856,255개의 좌표를 사용하였다.

Figure 6은 3단계에 걸쳐 수행한 필터링 결과를 보여준다.

Section 3에서 제시한 방법대로 전처리를 한 결과 1차 필터링에서는 속도 통계에 의미가 없는 1km미만 주행의 파일을 제거하였다. 약 5.1%가 줄어들었다. 2차 필터링에서는 Device 센서 제한값 범위를 벗어나는

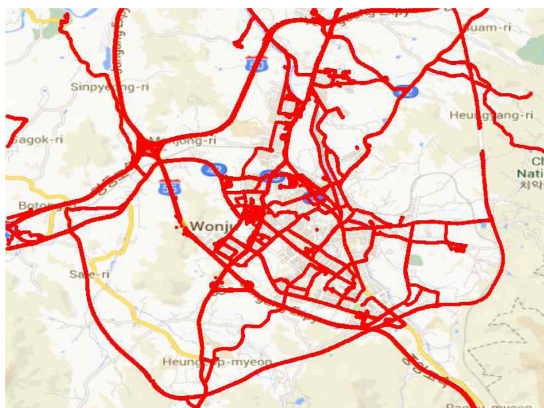
Figure 7. Current Traffic Information Area<sup>3)</sup>

Figure 8. GPS Trajectory of DTG Vehicles

파일들을 제거한 결과이다. Speed 센서 이상, RPM 센서 이상, 브레이크신호이상 그리고 GPS신호불량 등 기계장치 이상을 보이는 데이터를 필터링 하여 합계 23.9%가 감소하였다. 3차 필터링은 통계기법을 이용하여 Outlier들을 필터링하여 약 1%가 추가로 줄어 3차에 걸친 필터링 결과 24.9%가 축소되었다.

#### 4.2 지도 매칭 결과

Figure 7은 현재 한국의 원주 지역 교통정보 제공 현황이다. 이 구간을 샘플로 선정한 것은 고속도로 및 국도가 지나가고 지방시 시내구간도 포함하므로 대표적인 지방 지역 교통 유형을 모두 포함하고 있기 때문이다. 교통정보 제공 현황을 보면 고속도로 구간 위주로 제공하고 있는 것을 볼 수 있다.[7]

Figure 8은 DTG상용차량이 원주지역의 GPS궤적이다. 원주시내 및 국도까지 차량들이 이동한 것을 확



Figure 9. A Sample of GPS Trajectory

인 할 수 있다.

Figure 9는 도로 네트워크에 DTG상용차량의 GPS 좌표 및 통행 속도를 표시한 것이다. 많은 차량들이 다양한 속도로 도로를 지나간 것을 확인할 수 있다.

#### 4.3 교통정보 생성 결과

Table 2는 빅데이터 교통 패턴정보 생성방법을 이용하여 생성한 교통정보 통행속도 사례이다.

표에서 분석된 도로는 가현로라는 지역 도로로서, 차량운행숫자와 요일별, 시간대별 통행속도 평균 사례를 보여주고 있다. 이 도로의 통행제한 속도는 60 km이고 지도 속성에도 도로 평균속도로 60 km로 입력되어있다. 그러나 실제 통행 속도를 보면 제한 속도보다 낮은 것을 확인 할 수 있다. 지역 도로 정보가 교통 맵에 반영이 저조함을 알 수가 있다.

Table 3은 원주지역 교통정보 링크수 및 수집을 통계이다. 분석에 사용된 기본 지도는 2013년 4월 기준으로 수집된 전자지도이다. 원주지역 전체 링크 개수는 총 7,996개이고, 기존 교통정보 수집 링크는 1,307개이다. 본 연구에서 제시한 방법대로 DTG GPS궤적을 활용하게 되면 추가로 1,533개의 교통정보 수집 링크를 확보하게 되어 약 117.3%(1,533/1,307)의 수집을 향상을 기할 수 있다. 다시 말하면 기존에 16.3% 정도였던 교통정보 활용 비율이 35.5%로 늘어나게 되는 것이다. 물론 일부 지역에서는 통과한 차량이 부족하여 오히려 경로에 왜곡현상을 가져올 수도 있으나 이 부분은 오랜 기간 누적하여 통계 보정을 하게 되면 개선할 수 있을 것으로 생각된다.

이와 같이 지역 도로의 교통정보 링크수가 DTG 데

<sup>3)</sup> NAVER지도 교통지도(원주) <http://map.naver.com>

Table 2. Partial Results of Traffic Speed and Counts Extracted (Gahyun-ro road)

Hour	Sun		Mon		Tue		Wed		Thu		Fri		Sat	
	Spd	Cnt	Spd	Cnt	Spd	Cnt	Spd	Cnt	Spd	Cnt	Spd	Cnt	Spd	Cnt
9	44	154	28	558	27	471	29	498	40	118	36	216	38	147
10	30	109	27	392	21	325	34	353	28	352	28	139	29	156
11	51	77	38	158	25	185	32	255	37	261	21	404	30	114
12	40	94	38	164	41	114	33	332	24	153	43	41	33	101
13	51	41	35	79	37	33	31	141	50	72	30	205	64	1
14	46	26	43	55	23	143	6	50	31	252	32	119	4	30
15	16	194	24	241	55	43	39	76	19	233	33	45	16	101
16	17	142	26	208	43	152	49	43	29	174	25	285	30	82

Table 3. Traffic Link Collection Rate

Type	Count	Collection Rate
The number of links on the existing traffic map	1,307	16.3%
The number of links extended on the traffic map	1,533	19.2%
The total number of links	2,840	35.5%
Improvement Rate	117.3%	

이터 분석을 통하여 확보되어 교통정보 활용이 증가하게 된다.

## 5. 결 론

본 연구에서 실시간 교통정보가 제대로 수집되지 않는 지방지역에서 DTG 단말로부터 올라오는 대량의 GPS궤적을 빅데이터 맵매칭과 MR기술을 이용하여 패턴 교통정보 커버리지를 확대하는 연구를 하였다. 차량 통신 단말 특성상 Garbage가 많은 소스 데이터의 품질을 높이는 전처리과정과, 분석하기 용이한 구조로 변환하는 GPS궤적 Segmenting, 이 궤적을 지도와 매칭하여 통행속도를 추출하는 맵매칭기술 그리고 MR기술을 이용하여 대량의 데이터 필터링과 요일별, 시간별 각 도로의 속도를 추출하여 패턴정보를 생성하는 방법을 제시하였다. 그리고 원주지역을 선정하여 교통정보 수집을 개선 효과를 분석하였고 약 117.3%의 추가 수집을 향상할 수 있음을 확인하였다. 대부분의

기존 교통정보를 이용한 빠른길 안내를 서비스하는 경로탐색 엔진은 대도시, 고속도로 및 일부 국도 교통정보만 활용하여 지방지역 경로품질이 비교적 낮았으나 본 연구가 상용 적용되면 지방 지역 교통정보 커버리지가 확대되고 경로품질 개선효과도 기대되어 스마트폰을 이용한 실시간 경로탐색 서비스 활성화에 기여할 것으로 기대된다.

향후 연구로서 실제 실시간 반영을 위하여 실시간 처리 기술을 적용하는 시스템에 대한 연구가 필요하다.

## References

- [1] Seok-June Lee, Chungwon Lee, "Short-Term Impact Analysis of DTG Installation for Commercial Vehicles," The Journal of The Korea Institute of Intelligent Transport Systems, Vol. 11, No. 6, pp.49-59, 2012.
- [2] Jeongrae Kim, Chanki Jeong, "A Study on Phon Call Big Data Analytic," Journal of Information Technology and Architecture, Vol. 10. No. 3, pp. 387-397, 2013.
- [3] Dong Ju Kim, Mincheol Kang, "Evolution of ICT Ecosystem and Mobile Telcos' Counterstrategies," Journal of Information Technology and Architecture, Vol. 10. No. 2, pp. 197-209, 2013.
- [4] Skiena, S. "Dijkstra's Algorithm," Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica, Reading, MA: Addison-Wesley, pp. 225-227, 1990.
- [5] Yao, J., Lin, C., Xie, X., Wang, A. J., & Hung, C. C. "Path planning for virtual human motion using



- improved A\* star algorithm,” In Information Technology: New Generations (ITNG), 2010 Seventh International Conference on IEEE, pp. 1154-1158, 2010.
- [6] Huang, Y. W., Jing, N., & Rundensteiner, E. A. “Effective graph clustering for path queries in digital map databases,” In Proceedings of the fifth international conference on Information and knowledge management, pp. 215-222, ACM, 1996.
- [7] Mandatory to mount DTG on all commercial vehicles, Ministry of Land, Infrastructure and Transport of Korea. Jun. 22, 2010, [http://www.mltm.go.kr/USR/NEWS/m\\_71/dtl.jsp?id=155552574](http://www.mltm.go.kr/USR/NEWS/m_71/dtl.jsp?id=155552574)
- [8] Kee-Choo Choi, Sang-Woo Shim, Dong-Hwan Kim, “A Comparative Study between GPS-based and RFID-based Traffic Information Collection System,” Journal of The Korean Society of Civil Engineers 29.5D, pp. 571-578, 2009.
- [9] Wonhee Cho, Eunmi Choi, “A Traffic Data based User's Route Exploring Mechanism using Hadoop Platform,” Proceedings of KIPS Fall Conference, Vol. 20-2, pp. 176-179, 2013.



#### 조원희 (Wonhee Cho)

1990년 인하대 전산과 (이학사)  
 1992년 인하대 컴퓨터공학 (공학석사)  
 1992년 - 2013년 SKC&C, SK플래닛 IT분야 근무  
 2013년 - 현재 국민대 비즈니스IT대학원 박사과정 재학  
 연구분야: 빅데이터 분석, R, IoT/M2M, ITS, LBS/Telematics  
 Email: whcho@kookmin.ac.kr



#### 최은미 (Eunmi Choi)

1988년 고려대학교 전산과학과 (이학사)  
 1991년 미국 Michigan State University, Computer Science (공학석사)  
 1997년 미국 Michigan State University, Computer Science (공학박사)  
 1997년 - 1998년 Michigan State University, Post-Doc.  
 1998년 - 2004년 한동대학교 전산전자공학부 조교수  
 2004년 - 현재 국민대학교 경영정보학부 정보시스템 전공 교수  
 연구 분야: 분산시스템, 미들웨어, 클라우드 컴퓨팅 및 서비스, 시스템 모델링 및 설계  
 E-mail: emchoi@kookmin.ac.kr