

활동기반 교통모형 분석자료 구축을 위한 소셜네트워크 공간빅데이터 활용방안 연구

A Study on the Application of Spatial Big Data from Social Networking Service for the Operation of Activity-Based Traffic Model

김 승 현*

(Seung-Hyun Kim)
(University of Seoul)

김 주 영**

(Joo-Young Kim)
(University of Seoul)

이 승 재***

(Seung-Jae Lee)
(University of Seoul)

요 약

오늘날 우리 주변에는 규모를 가늠할 수 없을 정도로 많은 정보와 데이터가 생산되는 ‘빅데이터(Big Data)’의 시대가 도래 하였으며, 그 중요성이 날로 커지고 있다. 교통분야에서는 전통적인 통행기반교통모형(Trip-Based Model)인 4단계 교통수요추정법의 한계가 드러나고 있으며, 활동기반교통모형(Activity-Based Model)을 이용한 수요 추정 방법이 교통계획에 새로운 패러다임으로 떠오르고 있다. 교통은 사람이나 물류의 공간상의 시간적 이동을 의미한다고 봤을 때 공간데이터와 밀접한 관련이 있다. 따라서 공간정보를 포함하고 있는 SNS를 대상으로 시계열적 공간정보를 추출하고, 이를 현재의 통행기반 교통모형(Trip-Based Model) O/D와 비교·분석하여 그 특성을 파악하고 유용성을 검증하였다. 또한, 활동기반교통모형(Activity-Based Model)의 분석자료를 구축하여 교통시뮬레이터 프로그램을 이용해 시뮬레이션을 수행하고 그 결과를 고찰하였다. 연구결과 다수의 활동기반 교통모형 분석자료를 구축할 수 있었으며, 이번 연구를 통해 교통분야 빅데이터 활용의 기술적 한계를 극복할 수 있는 가능성을 확인하였고, 향후 발전방향을 모색하는 기회가 되었다.

핵심어 : 빅데이터, 공간빅데이터, 소셜네트워크서비스(SNS) 분석, 활동기반교통모형, 통행사출

ABSTRACT

The era of Big Data has come and the importance of Big Data has been rapidly growing. The part of transportation, the Four-Step Travel Demand Model(FSTD), a traditional Trip-Based Model(TBM) reaches its limit. In recent years, a traffic demand forecasting method using the Activity-Based Model(ABM) emerged as a new paradigm. Given that transportation means the spatial movement of people and goods in a certain period of time, transportation could be very closely associated with spatial data. So, I mined Spatial Big Data from SNS. After that, I analyzed the character of these data from SNS and test the reliability of the data through compared with the attributes of TBM. Finally, I built a database from SNS for the operation of ABM and manipulate an ABM simulator, then I consider the result. Through this research, I was successfully able to create a spatial database from SNS and I found possibilities to overcome technical limitations on using Spatial Big Data in the transportation planning process. Moreover, it was an opportunity to seek ways of further research development.

Key words : Big Data, Spatial Big Data, Social Networking Service(SNS) Analysis, Activity-Based Model, Trip chain

† 본 연구는 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015S1A5B8046775)

* 주저자 : 서울시립대학교 교통공학과 박사과정

** 교신저자 : 서울시립대학교 도시과학연구원 융합도시연구센터 연구교수

*** 공저자 : 서울시립대학교 교통공학과 교수

† Corresponding author : Jooyoung Kim(University of Seoul), E-mail trafficplan@naver.com

† Received 12 April 2016; reviewed 23 May 2016; Accepted 4 July 2016

I. 서론

1. 연구의 배경 및 목적

오늘날 인터넷 사용의 급격한 증가, 스마트폰 보급, SNS(Social Networking Service) 활성화 등과 같은 디지털 기술의 발전과 모바일 기기의 보급 등으로 인해 우리 주변에는 규모를 가늠할 수 없을 정도로 많은 정보와 데이터가 생산되는 ‘빅데이터(Big Data)’의 시대가 도래했으며, 그 중요성이 날로 커지고 있다[1].

교통분야에서는 오랜시간동안 우세를 점하고 있었고 도시교통수요 추정에 효과적인 방법으로 여겨졌던 전통적인 통행기반교통모형(Trip-Based Model)인 4단계 교통수요추정법의 한계가 드러나고 있으며, 최근의 컴퓨터 기술의 빠른 성장에 힘입어 활동기반교통모형(Activity-Based Model)을 이용한 수요 추정 방법이 교통계획에 새로운 패러다임으로 떠오르고 있다[2].

교통은 사람이나 물류의 공간상의 시간적 이동을 의미한다고 봤을 때 공간데이터와 밀접한 관련이 있으며, 최근 급부상하고 있는 공간빅데이터(Spatial Big Data)와 접목될 수 있는 분야이다. 따라서 소셜네트워크 서비스(SNS) 상의 공간빅데이터를 추출해 현재 사용되고 있는 통행기반교통모형(Trip-Based Model)의 특성과 비교·분석하고 나아가 활동기반교통모형(Activity-Based Model)의 분석자료를 구축하여 공간빅데이터와 활동기반교통모형을 접목시킨다면 매우 큰 가치가 있으며, 향후 활용성이 클 것으로 생각된다.

2. 연구 내용 및 방법

1) 연구의 내용

본 연구는 공간정보를 포함하고 있는 SNS를 대상으로 시계열적 공간정보를 추출하고, 이를 현재 사용되고 있는 통행기반교통모형(Trip-Based Model)의 OD와 비교·분석하여 그 특성을 파악하였다. 또한, 활동기반교통모형(Activity-Based Model)의 분석자료를 구축하고 교통시물레이터 프로그램을 이용하여 시물레이션을 수행함으로써 가시적인 분석결과를 도출하

고 SNS를 이용하는 개개인의 통행사슬을 추적하였다. 연구의 공간적 범위는 서울외곽순환고속도로를 포함하는 수도권 내를 대상으로 진행하였으며, 시간적 범위는 2015년을 분석연도로 설정하였다.

2) 연구의 방법

본 연구에서는 소셜미디어 관련 연구와 활동기반 교통모형의 국내외 연구동향을 살펴보고 관련 문헌을 고찰하여 SNS상에서 공간빅데이터를 추출하기 위한 고려사항과 관련이론들을 검토하였다. 이를 바탕으로 연구대상 SNS를 선정하고, SNS 공간정보의 시계열적 추출방법을 도출하였으며, JAVA Programming을 이용하여 시계열적 공간정보를 추출하였다. 추출된 공간정보를 기존의 통행기반교통모형의 OD와 비교·분석하여 그 특성을 파악하고 분석 자료로 사용하기에 타당한지 검증을 거친 후 활동기반교통모형(Activity-Based Model)의 분석 자료로 구축하고 교통시물레이션 프로그램을 이용하여 시물레이션을 수행하였다. 또한, 프로그램 구축 과정 및 시물레이션 프로그램 실행 과정상의 문제점 및 이슈사항들을 정리하여 종합 결론을 도출 하였다.

II. 선행연구 고찰

공간빅데이터로서 소셜미디어를 활용하는 것과 관련된 과거 연구의 주요 논문들의 내용을 살펴보면 다음과 같다. Seo(2012)는 ‘SNS 기반의 매시업을 통한 실시간 재해정보 추출 및 표현에 관한 연구’[3]에서 트위터를 기반으로 하여 SNS상에 게시되는 재해 관련 정보들을 실시간으로 추출하여 데이터베이스에 저장함으로써, 재난 상황에서 SNS를 활용할 수 있는 시스템을 제안했다. Park(2013)은 ‘SNS 데이터의 도시정책지표로서의 활용 가능성 연구’[4]에서 트위터를 기반으로 하여 SNS 데이터들 중 주거환경에 대해 언급한 게시물들로부터 정보를 추출하여 분석함으로써, SNS의 도시정책지표로써 활용가능성을 판단하는 연구를 수행했다. Lee(2014)는 ‘공간빅데이터로서 소셜미디어 자료를 활용한 공간정보 추출 및 표현 시스템 개발’[5]의

연구에서 트위터, 페이스북, 플리커를 연구대상으로 선정하여 게시자의 위치를 표출하는 시스템을 개발한 사례가 있다.

한편, 교통에서의 통행사슬과 관련된 연구를 살펴보면 최초로 주목받은 것은 1950년대 말 미국 CATS(Chicago Area Transportation Study, 1959)가 통행은 연쇄적으로 발생하고 상호 관련이 있다는 것에 주목하면서 발전하기 시작하였다[6]. Adler et al. (1979)[7]는 개인이 하루에 발행하는 통행사슬 형태는 사이클(Cycle)수와 스톱(Stop)수에 의해 결정되며 이는 통행의 편리성을 나타내는 지표로 사용될 수 있다고 하였다. Currie et al.(2011)[8]는 멜버른시의 조사데이터를 이용하여 대중교통이용에 대한 통행사슬행태를 연구하였는데 그 결과 혼잡한 도시 내에서 여러 목적지가 연결된 복합연계통행을 할 경우 승용차를 이용한 통행보다는 지하철과 트램의 이용비율이 높게 나타났다는 것을 알았다. 이와 같이 국외에서는 통행사슬의 기본단위인 사이클(Cycle), 스톱(Stop)과 교통행태의 관계를 나타내고자 하는 연구부터 시작하여 사회속성별 특성과 관계를 나타내고자 하는 연구가 다양하게 진행되어 왔다.

국내에서 수행된 통행사슬과 관련된 연구로는 Yoon(1997)[9]은 하루(24시간)단위로 표현되는 통근통행자의 통행패턴을 분석하기 위해 통근통행자의 통행목적수와 사이클(Cycle)수를 네스티드로그 모델로 추정하였다. Choo et al.(2008)[10]은 통행사슬을 집에서 출발하여 다시 집으로 돌아오는 가정기반 통행사슬(Home-based trip chain)과 집 외 장소(예, 직장)에서 출발하여 다시 동일한 집 외 장소로 돌아오는 비가정 기반 통행사슬(Non-home-based trip chain)의 2가지로 구분하여 정리하였으며, 통행사슬의 특성을 분석하고 다중회귀형태의 통행사슬모형을 추정하여 통행목적과 통행사슬과의 연관성을 분석하였다. Bin(2011)[11]은 통행사슬(Trip Chain) 개념을 이용하여 개인이 통행시 통행목적과 통행연계수를 고려하는 행태에 대하여 분석하였다.

활동기반 교통모형의 국내외 연구동향을 살펴보면 MacNally(1999)[12]는 통행수요예측의 전통적인 통행기반 모형은 통행행태를 반영하기에는 부족하

다고 발표하고 통행행태를 연구하기 위한 전통적인 접근법은 하루 동안 또는 며칠 동안 통행행태패턴 분석으로 대체되어야 한다고 논했다. Kim(1994)[13]은 기존 4단계 모형 틀에서는 동적행태 특성을 모형화 하는데 한계가 있음을 소개하며, 목적활동에 근거한 모형의 패러다임 특성을 제시하고 활동목적에 근거한 연구 활동을 3가지로 구분하였다. Kim(2012)[2]은 서울시 강남구를 중심으로 활동기반 모형인 MATSim과 통행기반 모형인 Emme2의 통행배정결과를 비교분석하여 두 모형의 결과는 유사하게 나타났으며, 행위자 기반의 모형을 사용해 동적인 측면에서 4단계 모형을 보완하는 것이 중요하다고 주장하였다. Ali(2015)[14]는 서울시의 스마트카드 데이터를 활동기반 시뮬레이터인 MATSim을 이용하여 시뮬레이션 및 정산과정을 수행하고 스마트카드 자료와 가구통행실태조사 자료를 결합하여 정산과정을 수행한 결과 보다 더 정확하며 신뢰성 있는 결과를 도출 할 수 있었다.

본 연구에서는 공간빅데이터로서 소셜미디어를 활용하는 것과 관련된 과거 연구에서 SNS상에서 공간빅데이터를 추출하기 위한 고려사항과 관련이론들을 검토하였으며, 관련사례를 분석하였다. 또한, 통행사슬과 관련된 연구를 바탕으로 1일간 인당 총통행거리 및 통행 Leg수 분포 등의 특성을 분석하여 구축자료를 기존의 통행기반 OD자료와 비교 및 검증 하였으며, 활동기반 교통모형 연구 결과를 바탕으로 활동기반 시뮬레이터 프로그램을 수행하여 가시적인 결론을 도출 하였다.

Ⅲ. 분석자료의 구축 및 검증

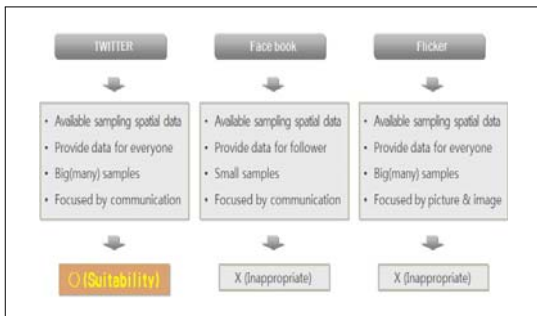
1. 활동기반 교통모형 분석자료 구축

1) 연구 대상 소셜네트워크(SNS) 선정

소셜네트워크 서비스의 종류에는 블로그, 트위터, 페이스북, 플리커, 유튜브, 카카오톡, 카카오토토리, 핀터레스트 등 다양한 것들이 있고 현재에도 계속 새로운 채널들이 개발되고 있다. 본 연구에 앞서 Lee(2014)는 ‘공간빅데이터로서 소셜미디어 자료

를 활용한 공간정보 추출 및 표현 시스템 개발'[5]의 연구에서 트위터, 페이스북, 플리커를 연구대상으로 선정하여 시스템을 개발 한 사례가 있다.

본 연구에서는 앞의 사례를 참고하여 트위터 및 페이스북, 플리커의 3가지 소셜네트워크 서비스를 대상으로 연구의 적절성을 검토한 결과 공간정보 추출이 가능하며 모든 사용자에게 정보를 제공하고, 표본수가 많으며 대화 위주의 SNS인 트위터가 본 연구에 가장 적합한 것으로 판단되어 연구 대상 소셜네트워크 서비스로 선정하였다. 페이스북의 경우 정보제공이 팔로워(Follower)에게만 제공되어 매우 많은 팔로워를 보유하고 있지 않은 경우 정보의 획득이 다소 제한적이며, 플리커의 경우 사진 위주의 SNS로서 공간정보를 추출할 수 있는 게시글의 표본수가 적어 본 연구대상에서 제외하였다.



<Fig. 1> Selection process of SNS for study

2) SNS 공간정보의 시계열적 추출방법

소셜네트워크 서비스로부터 정보를 검색하고 위치정보를 추출해 내기 위해서 연구의 대상이 되는 소셜네트워크 서비스의 API를 분석하는데 관련 홈페이지에 공개되어 있는 기능들 중에서 본 연구에서 개발하려고 하는 시스템에 필요한 변수들을 선정했다. 본 연구에서 필요로 했던 변수들은 첫째로 게시물들의 내용들을 검색할 수 있게 하는 변수, 둘째로 게시물의 작성 날짜와 시간을 나타내는 변수, 셋째로 작성 당시의 위치정보를 나타내는 변수, 넷째로는 검색의 공간적 범위를 제한하는 변수, 다섯째로는 자료의 검증용을 위한 경로상의 거리를 나타

내는 변수 등이 있다. 이러한 트위터의 주요 API 매개변수들을 정리하면 다음의 표와 같다[5].

<Table 1> TWITTER API List

Request variable	Value	Description
TWITTER_CONSUMER_KEY	String (indispensable)	TWITTER's API key
tweet.getID()	String	User's ID
tweet.getText()	String	Contents of posting
word	String	Research word
tweet.getGeoLocation()	GeoLocation	Check the Geoinfo
tweet.getGeoLocation().getLatitude()	Double	Latitude coordination
tweet.getGeoLocation().getLongitude()	Double	Longitude coordination
radius	Double	Radius from point
km	String	Length measure
date	String	Date of posting
time	String	Time of posting

3) 시계열 공간정보 추출 프로그램 개발

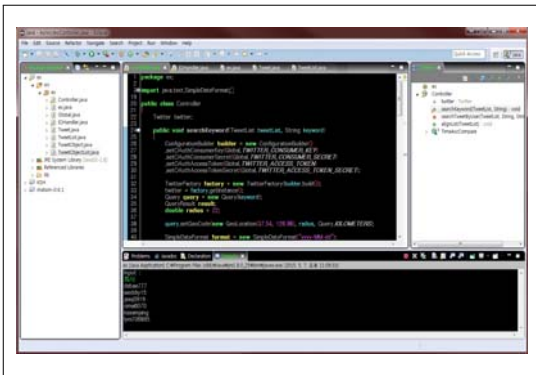
본 연구에서는 소셜네트워크 서비스인 트위터를 대상으로 게시물에 올린 사용자의 위치와 시간을 파악하고 아이디를 통해 하루간에 올린 게시물의 시간변화에 따른 위치의 변화를 파악하여 이를 하루간 통행스케줄이라고 가정하여 시계열적 데이터를 구축하였다. 다만, 회사 근무와 같이 특별히 SNS에 게시물을 올릴 필요가 없는 활동 등의 포착이 어려운 현실적 문제가 존재하나, 이는 향후 다양한 공간 빅데이터를 재가공 및 재조합하여 보완할 필요성이 있으며, 다가올 IoT 시대에는 더욱 많은 시공간상 위치 정보들이 상시적으로 수집됨에 따라 극복 가능할 것으로 판단된다.

위의 가정에 따라 추출된 구글맵기반의 위치좌표를 다음 좌표계 변환 API를 이용하여 변환하는

동시에 다음지도상에서 위치이동간 거리를 추출함으로써 데이터의 검증에 사용하였다. 이를 위해 트위터 Open API를 사용하여 JAVA eclips(LUNA 4.4.1)를 이용해 WINDOWS 기반으로 프로그램을 구축하였다.

〈Table 2〉 Development environment of system

Item	Contents
OS	MS Windows 7 64bit
Language	Java
API	TWITTER Open API Daum Open API
Library	twitter4j TransCoord json-simple-1.1.1
Tool	JAVA eclips LUNA 4.4.1 JDK



〈Fig. 2〉 Expression system of spatial data from SNS(JAVA eclips LUNA 4.4.1)

4) 활동기반 교통모형 분석자료 구축

위의 프로그램을 이용하여 서울특별시 외곽순환고속도로 이내(서울중심에서 반경 22km 이내)를 대상으로 구축된 자료의 형태 및 내용을 살펴보면 person id가 순서대로 부여되는 것을 볼 수 있고, 나이 및 고용형태, 시설구분, 통행목적, 교통수단 등은 개인정보 보호를 위해 공개되고 있지 않거나 구별 할 수 없으므로 임의로 부여 하였다. 사용자의 위치를 표출하는 좌표는 x, y 변수를 통해 표출되고

있으며, 사용자가 게시물을 올린 시간이 end_time 및 dep_time을 통해 표출되고 있다. 또한, 데이터의 검증을 위해 다음지도를 통해 좌표간 거리를 Len 변수를 통해 표출되도록 하였다.

```

1 <person id="1" age="00" employ="yes">
2 <!do include="yes">
3 <cat type="L" Facility="0000" x="313680.4504647055" y="546699.1795414511" end_time="08:18:05" />
4 <leg mode="cat" dep_time="08:19:01"><!-- Len = 0.0 -->
5 </leg>
6 <cat type="L" Facility="0000" x="314207.971031106" y="546697.616642766" end_time="08:47:28" />
7 <leg mode="cat" dep_time="08:47:28"><!-- Len = 3.47 -->
8 </leg>
9 <cat type="L" Facility="0000" x="317123.114123985" y="547074.690642459" end_time="09:13:00" />
10 <leg mode="cat" dep_time="09:13:00"><!-- Len = 2.17 -->
11 </leg>
12 <cat type="L" Facility="0000" x="314484.9517607854" y="546047.9205471985" end_time="09:38:44" />
13 <leg mode="cat" dep_time="09:38:44"><!-- Len = 0.0 -->
14 </leg>
15 <cat type="L" Facility="0000" x="314804.042307685" y="546561.2397417745" end_time="09:45:05" />
16 <leg mode="cat" dep_time="09:45:05"><!-- Len = 0.5 -->
17 </leg>
18 <cat type="L" Facility="0000" x="314484.9517607854" y="546047.9205471985" end_time="10:05:03" />
19 <leg mode="cat" dep_time="10:05:03"><!-- Len = 0.5 -->
20 </leg>
21 <cat type="L" Facility="0000" x="314804.042307685" y="546561.2397417745" end_time="10:39:13" />
22 <leg mode="cat" dep_time="10:39:13"><!-- Len = 0.5 -->
23 </leg>
24 </!do>
25 </person>
26 <person id="2" age="00" employ="yes">
27 <!do include="yes">
28 <cat type="L" Facility="0000" x="310760.562879725" y="552745.403205581" end_time="02:17:39" />
29 <leg mode="cat" dep_time="02:17:39"><!-- Len = 0.0 -->
30 </leg>
31 <cat type="L" Facility="0000" x="309683.4234241204" y="548350.494964655" end_time="03:48:00" />
32 <leg mode="cat" dep_time="03:48:00"><!-- Len = 3.74 -->
33 </leg>
34 </!do>
35 </person>
    
```

〈Fig. 3〉 Sample of construction data for the operation of Activity-Based Model

5) 구축결과

분석자료 구축결과 총 702명의 하루간 통행사슬(Trip chain)을 추출 할 수 있었으며, 그 중 52%인 367명이 1통행 이상 발생한 것으로 집계되었다. 나머지 통행이 발생하지 않은 표본은 하루에 1개의 게시물만 작성한 사용자가거나, 위치변화 없이 한 장소에서 여러 개의 게시물을 올린 사용자로 판단 된다. 총 발생 Leg는 1,285개이며, 1인당 평균 3.5개로 나타났다.

〈Table 3〉 Result of construction data for the operation of Activity-Based Model

Item	Quantity	Ratio
Total sample count	702	100%
Trip generation sample	367	52%
Trip leg	1,285	—
Average trip leg	3.5	—

2. 구축자료의 검증

1) 기존 O/D자료와 특성 비교

2012년 수도권 전수화 보고서[15]에 제시된 서울시 내부 총 목적통행 발생량 및 평균통행거리, 인구를 이용하여 산출한 1일간 인당 총통행거리는 24.1km/일/인당으로 나타났으며, 본 연구에서 집계한 1일간 인당 총통행거리는 26.3km/일/인당으로 나타났다. 따라서 가구통행실태조사 기반의 O/D에 준별 공로최단거리를 적용하여 산출한 전수화 데이터에 비해 9.1% 길게 나타났으나, 기존 O/D자료와 큰 차이가 없어 분석데이터로 활용하기에 타당한 것으로 판단된다.

<Table 4> Total trip length per one day per capita

Item	Total trip length (km/day/capita)	Remark
Seoul Metropolitan OD data(2012.4)	24.1	intra-seoul / purpose trip
This study	26.3	intra-seoul
Difference	2.2	9.1%

2006년 수도권 가구통행실태조사시의 통행 Leg 분포비율[11]과 두 집단 간의 독립성을 검정하는 카이제곱검정 결과 p값이 0.008로 분석되어 전체적인 분포는 유사한 것으로 나타났으며, 8Leg 이상의 복잡한 통행이 가구통행실태조사의 경우 0.3%로 집계되었으나, 본 연구 시에는 10.9%로 집계되어 설문조사 시보다 복잡한 통행을 하는 비율이 다소 높은 것으로 나타났다. 또한 가구통행실태조사의 경우 최대 Leg가 18개로 나타났으나, 본 연구 시에는 43개로 집계되어 사람들의 통행 행태가 다소 복잡한 형태를 나타내는 것으로 분석되었다.

<Table 5> Trip Leg distribution comparison

※ TDS data : Seoul Metropolitan Travel Diary Survey data(2006)

Item	1	2	3	4	5	6	7	8 & Total over (%)	χ^2
TDS data	38.8	38.2	11.9	7.4	2.0	1.2	0.3	0.3	p=0.008
This study	39.2	20.4	12.8	6.5	6.3	2.5	1.4	10.9	

2) 추출자료 특성분석

10 Leg 이상의 통행에 포함된 1km 미만의 통행이 307개 23.9%로 높은 비율로 나타나 단거리 통행을 포함한 복잡한 통행행태가 실제로 이루어지고 있는 것으로 분석되었다. 또한, 1km 미만의 Leg 수 역시 750개로 전체의 58.4%에 달해 단거리 통행이 교통의 큰 부분을 차지하고 있음이 나타났다.

<Table 6> Study on Leg-Trip length distribution

Item	under 1km	1-5 km	5-10 km	10-15 km	15-20 km	20km & over	Total
1	65	39	15	12	5	12	148
2	65	34	21	12	5	14	151
3	65	19	25	6	1	25	141
4	53	22	8	5	3	5	96
5	42	19	31	5	0	13	110
6	27	10	7	0	5	5	54
7	23	1	1	6	0	4	35
8	27	3	0	2	0	0	32
9	76	33	4	1	10	2	126
10 & over	307	56	10	7	2	10	392
Total	750	236	122	56	31	90	1,285

평일-주말간 통행특성을 비교한 결과 평일에 조사된 표본수는 207명 673Leg, 주말에 조사된 표본수는 160명 612Leg로 집계되었다. 1일간 인당 총통행거리는 평일이 높은 반면 평균 Leg수는 주말이 높은 것으로 나타났다. 두 집단의 차이를 검정하는 t검정 결과 p값이 각각 0.253, 0.195로 분석되어 평일과 주말은 통행특성이 다른 것으로 판별되었으며, 이는 평일에는 장거리 위주의 출퇴근 통행이나 업무통행의 비율이 높으며, 주말에는 비교적 단거리며 복잡한 통행패턴을 보이는 쇼핑이나 여가통행의 비율이 높기 때문인 것으로 분석된다.

〈Table 7〉 Weekday-end trip character comparison

Item	Sample	Trip leg	Trip length (km/day/ca)	t-test	Trip leg per capita	t-test
Weekday	207 (56%)	673 (52%)	30.7	F=0.000	3.3	F=0.000
Weekend	160 (44%)	612 (48%)	20.6	p=0.253	3.8	p=0.195
Total	367 (100%)	1,285 (100%)	26.3	-	3.5	-

3) 소결

1일간 인당 총 통행거리 분석결과 본 연구 시에는 26.3km/일/인으로 가구통행실태조사에 의한 전수조사보다 2.2km/일/인 높은 것으로 나타났다. 또한, 가구통행실태조사 데이터와 비교했을 경우 8Leg 이상의 통행이 10.9%로 매우 높게 나타나 실제로는 더 복잡한 통행사슬이 이루어지고 있는 것으로 분석되었다. 1km 미만의 짧은 Leg가 전체의 58.4%로 나타나 매우 높은 비율을 차지하고 있었다. 주말과 평일 분석결과 주말에는 평일에 비해 1일간 인당 총 통행거리는 짧고 인당 평균 Leg 수는 많은 것으로 나타나 사람들의 통행패턴이 평일보다 주말에 다소 복잡한 양상을 보이는 것으로 분석되었다.

본 연구에서 집계된 데이터를 가구통행실태조사에 의한 통행기반 OD자료와 비교·분석했을 때 다소 상의한 특성들도 있었으나, 전체적으로는 비슷한 수준으로 분석되어 활동기반 교통모형의 기초 분석자료로 구축하기에 충분히 유용한 것으로 판단되었다.

IV. 활동기반 교통시뮬레이션 수행

1. 프로그램 선정

SNS에서 추출한 데이터를 기반으로 활동기반 교통모형과 접목시키기 위해 다양한 분석 프로그램을 검토한 결과 MATSim이 본 연구에 가장 적합한 것으로 판단되어 연구대상 프로그램으로 선정하였다.

MATSim은 Multi-Agent Transport Simulation의 줄

임말로서 스위스 ETH Zurich대학 Kay W.Axhausen 교수의 연구실에서 실험프로젝트로 출발하여 지금은 전세계 User들에 의해 활발히 연구되며, 지속적으로 User들에 의해 공동으로 개발되고 있다. Multi-Agent 기술, 동적·확률적모형, 활동기반모형에 기반 한 교통분석 시뮬레이터 프로그램이다[16].

2. 시뮬레이션 분석자료 구축

MATSim을 실행하기 위하여 SNS에서 추출한 데이터를 MATSim의 기초 입력자료인 Plans.xml 파일 형식으로 변환하였다. 변환하는 과정에서 링크배정의 오류를 방지하기 위하여 1km 미만의 Leg는 제외하였다. 1km 미만의 통행의 경우 대다수가 교통수단을 이용하지 않고 도보 또는 자전거 통행일 것으로 판단되므로 교통수요분석상 크게 차이는 없을 것으로 판단되나, 향후 더욱 면밀한 통행형태 분석을 위하여 세부적인 네트워크 보완이 필요할 것으로 판단된다. MATSim 프로그램은 통행배정시 경로(Path)별 링크(Link)를 배정하는데 링크는 경로보다 짧거나 최대한 같아야 한다. 그러나 본 분석에서 사용한 네트워크는 한국교통연구원에서 배포한 전국네트워크를 MATSim 네트워크파일 형식으로 변환한 것으로 세부적인 이면도로까지 네트워크로 구축하기에는 한계가 있다.

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <!DOCTYPE plan SYSTEM "http://www.matsim.org/files/otd/plan_v4.dtd">
3
4 <!--
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
-->

```

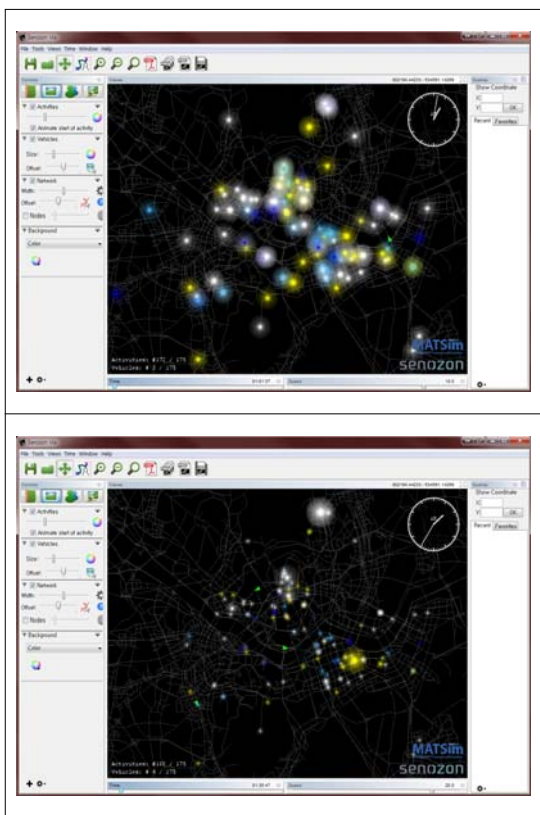
〈Fig. 4〉 Sample of simulation data for the operation of MATSim

3. MATSim을 이용한 시뮬레이션

MATSim 기초 입력자료인 Plans.xml 파일과

config.xml, network.xml, facilities.xml 파일을 이용하여 JAVA eclips LUNA 4.4.1을 이용해 프로그램을 구동하였다. 구동 시 MATSim의 strategy module을 통해 경로선택 모듈을 선택하였으며, 본 연구에서는 Reroute_Dijkstra 모듈을 선택하여 진행하였다. JAVA eclips LUNA 4.4.1을 이용해 네트워크 평행상대에 이를 때까지 iteration 시키고 수렴 후 출력된 결과를 via64 프로그램을 통해 표출하였다. 본 연구에서는 iteration 횟수 100회를 수렴조건으로 설정하였다.

MATSim 실행결과 총 175명의 하루간 통행스케줄이 표출 되었고, 주로 오후 2시~4시 사이의 시간대에 트위터를 사용하는 이용자들의 통행이 활발한 것으로 나타났으며, 서울시 전역에 걸쳐 통행이 발생한 것으로 나타났다. 이용자들의 활동을 나타내는 Activities 역시 서울시 전역에 걸쳐 활발하게 나타났다.



〈Fig. 5〉 Result of simulation using MATSim

V. 결 론

1. 연구의 제약사항 및 한계

본 연구를 수행함에 있어 트위터가 많은 API를 제공하고는 있지만, 기업의 상업적 이윤을 목적으로 API사용에 제한을 두고 있어 많은 양의 데이터를 구축하기에는 한계가 있었다. 또한, SNS 사용자가 하루 간에 올린 게시물의 시간변화에 따른 위치변화를 하루 간 통행스케줄이라고 가정함으로써 회사 근무와 같이 특별히 SNS에 게시물을 올릴 필요가 없는 활동 등의 포착이 어려운 현실적 문제가 존재하였다.

이는 향후 빅데이터의 활발한 활용을 위하여 개인정보보호 및 API 제한 등 빅데이터 사용의 제도적 개선방안이 강구되어야 할 것이며[17], 다양한 공간 빅데이터를 재가공 및 재조합하여 보완할 필요성이 있다. 또한, 다가올 IoT 시대에는 더욱 많은 시공간상의 위치 정보들이 상시적으로 수집됨에 따라 극복 가능할 것으로 판단된다.

2. 결론 및 시사점

API Limit 등의 제도적 제약에도 불구하고 다수의 활동기반 교통모형 분석자료를 구축할 수 있었으며, 교통분석 시뮬레이터인 MATSim을 성공적으로 실행할 수 있었다.

활동기반 교통시뮬레이터 실행을 위한 기초자료 분석결과 가구통행실태조사에 의한 통행량 전수화 방법을 통해 구축된 기존의 O/D 자료에 비해 사람들의 평균통행거리가 길고 단거리 통행이 많이 포함된 복잡한 통행사슬(Trip Chain)을 이루고 있는 것으로 나타났다. 따라서, 최근 들어 도심에서 발생하고 있는 국지적 교통혼잡상황의 대응에 통행기반 교통모형은 한계가 있으며, 단거리 통행까지 잘 묘사할 수 있는 활동기반 모형의 필요성이 다시 한번 확인 되었다. 다만, 현재의 교통분석용 네트워크로는 경로에 따른 링크 배정시 1km 미만의 단거리 통행은 오류가 발생할 소지가 있으므로 더욱 세부적인 네트워크 구축이 필요한 것으로 나타났다.

여러 가지 현실적 제약에도 불구하고 교통분야

에 있어 빅데이터 활용의 기술적 한계를 극복할 수 있는 가능성을 확인하였고, 향후 발전방향을 모색하는 기회가 되었다.

3. 향후 연구과제

앞서 언급한 대로 단거리 통행까지 잘 묘사할 수 있는 활동기반 교통모형을 구축하기 위해서는 보다 세부적인 교통분석용 네트워크 구축이 필요하고, 단순히 사용자의 통행패턴 뿐만 아니라 목적통행 및 수단통행을 구분 할 수 있는 알고리즘의 개발이 필요할 것으로 판단된다. 또한, 기타 SNS 및 모바일 기기로 확대하여 보다 많은 사용자의 통행스케줄 자료를 구축하는 것이 필요 할 것으로 생각된다.

향후에는 다양한 공간 빅데이터를 재가공 및 재조합하여 보다 많고 정확한 활동기반 교통모형 분석데이터를 구축할 필요성이 있고, 아울러 교통분야에서 다양한 빅데이터 활용방안을 모색하는 노력이 필요할 것이다.

REFERENCES

- [1] Kim J. J., Shin I. S. and Han K. J.(2013), "Spatial Big Data," *Surveying & Mapping Magazine*, pp.44-51.
- [2] Kim T. J.(2012), "Application of Activity-based Transport Simulation Model to the Gangnam-gu Area," *University of seoul*.
- [3] Seo T. W.(2012), "A Study of Real-time Disaster Information Extraction and Display using the Mash-up based on SNS," *Pukyong National University*.
- [4] Park J. H.(2013), "A study on the Applicability of SNS Data for the Urban Policy Indicator," *Ewha Womans University*.
- [5] Lee J. Y.(2014), "Development of a spatial information extraction and representation systems utilizing social media as spatial big data," *University of seoul*.
- [6] 近藤勝直(1987), "交通行動分析", 光洋書房.
- [7] Adler T. and Ben-Akiva M. E.(1979), "A Theoretical and Empirical Model of Trip Chaining Behavior," *Transport Research* 13B, pp.243-257.
- [8] Currie G. and Delbosc A.(2011), "Exploring the trip chaining behaviour of public transport users in Melbourne," *Transport Policy*, vol. 18, Issue 1, pp.204-210.
- [9] Yoon D. S.(1997), "Analysis of Urban Workers' Travel Pattern Choice Behavior," *The Korea Transport Institute, Journal of Transportation Research*, vol. 15, no. 4, pp.35-51.
- [10] Choo S. H., Kwon S. N. and Kim D. H.(2008), "Exploring Characteristics on Trip Chaining : the Case of Seoul," *The Korea Transport Institute, Journal of Transportation Research*, vol. 26, no. 4, pp.87-97.
- [11] Bin M. Y.(2011), "A Study on Trip Chain Typed Selection Behavior," *The Korea Transport Institute, Journal of Transportation Research*, vol. 29, no. 3, pp.7-19.
- [12] McNally. M. G.(1999), "Activity-based forecasting model incorporating GIS," *Geographical Systems*, pp.163-187.
- [13] Kim S. K.(1994), "Current state and direction of study of activity focused model," memo.
- [14] Atizaz A.(2015), "Activity-based Modeling Using Smart Card Data," *University of seoul*.
- [15] Metropolitan transportation Authority(2012), "Passenger origin-destination trip volume expansion and future demand forecasting co-work,"
- [16] Kim J. Y.(2014), "(A)simulated annealing algorithm for the activity based travel demand modelling," *University of seoul*.
- [17] Viktor M. S. and Kenneth N. C.(2013), "BIG DATA : A Revolution That Will Transform How We Live, Work, and Think".

저자소개



김 승 현(Kim, Seung-Hyun)

2016년 1월~현재 : 서울시립대학교 박사과정(교통공학전공)
2013년 2월~2015년 8월 : 서울시립대학교 석사 졸업(교통공학전공)
2010년 1월~2013년 1월 : (주)건영이앤씨 근무(교통계획부)
2006년 3월~2010년 12월 : (주)유엔알텍 근무(교통계획부)
1999년 2월~2006년 2월 : 계명대학교 학사 졸업(교통공학전공)
e-mail : comeback883@hanmail.net



김 주 영(Kim, Joo-Young)

2014년 3월~현재 : 서울시립대학교 도시과학연구원 융합도시연구센터 연구교수
2010년 3월~2014년 2월 : 서울시립대학교 박사 졸업(교통공학전공)
2008년 3월~2010년 2월 : 서울시립대학교 석사 졸업(교통공학전공)
2002년 3월~2008년 2월 : 서울시립대학교 학사 졸업(건축도시조경학부)
e-mail : trafficplan@naver.com



이 승 재(Lee, Seung-Jae)

1996년 10월~현재 : 서울시립대학교 교통공학과 교수
2005년 3월~2006년 8월 : 서울시립대학교 부설 교통연구센터장
1995년 11월~1996년 9월 : 교통개발연구원 책임연구원
1990년 10월~1995년 4월 : 런던대학 교통공학 박사 졸업(토목 및 환경공학과)
1988년 3월~1990년 2월 : 서울대학교 환경대학원 석사 졸업(교통학전공)
e-mail : sjlee@uos.ac.kr