
Global Report

스마트 운송계획_데이터, 모델 및 알고리즘

1. 개요
2. 스마트운송 분석에서 데이터소스
3. 커넥티드 차량
4. 트래픽흐름 예측: 타임시리즈 예측모델
5. 비즈니스 기회: 트래픽 예측 애플리케이션
6. 결론

1. 개요

현대사회에서 인구 증가는 이동성, 지속가능성 및 교통산업을 통해 사람들을 연결하고 서로 다른 문화를 전파하며, 궁극적으로 진화를 촉진한다는 점에서 중요한 역할을 하고 있다. 지난 2세기 동안 사회 여러 분야에서 혁신정책들이 증가하기 시작하면서 운송시스템은 더욱 정교해졌다. 이로 인해 교통시스템은 더욱 복잡해지면서 장점과 단점 모두를 나타내고 있다. 즉, 교통 관련 도구는 더 저 비용화 되었지만, 교통 혼잡도는 훨씬 더 증가하였다. 따라서 복잡한 운송시스템을 해결하고 처리하기 위해서는 체계적이고 효율적인 교통계획 및 학습이 필수적으로 요구된다. 예측은 이러한 복잡한 학습 작업의 한 사례가 될 수 있다. 이 작업은 주로 공간중속성과 현실 중속성의 복잡성으로 인해 어려움을 겪고 있다. 교통예보의 주요목표는 현재의 교통시스템으로 미래의 흐름을 예측하는 것뿐만 아니라 도로조건과 같은 다른 파라미터를 고려하는 것이다. 이를 통해 충돌사고나 악천후와 같은 돌발 상황이 발생할 경우 새로운 패턴에서 모델을 생성할 수 있다. 전통적인 운송데이터 수집 프로세스 기술은 설문조사와 인구조사를 통해 진행되었지만 해당 데이터를 수집하는 데 시간이 많이 걸리고 비용이 많이 든다. 따라서 사람들은 이제 GPS(Global Positioning System : GLONASS와 함께 완전하게 운용되고 있는 미국 국방부에서 개발한 범 지구위성항법시스템 중 하나로서 무기유도/항법/측량/지도제작/측지/시각동기 등 군용 및 민간용 목적으로 사용되고 있음) 및 GIS(Geographic Information System : 컴퓨터를 이용한 지형/지리정보를 바탕으로 정보 분석/저장/모형화를 통해 계획 및 설계 시 요구되는 제반현상으로 바꾸어, 객관적인 판단을 효율적으로 수행하는 것) 데이터를 포함한 비전통적인 클러스터링 데이터 등 주로 다른 소스의 데이터를 분석에 활용하고 통합하고 있다. 한편, 스마트폰과 스마트 대중교통 카드의 인기가 높아짐에 따라 개인 휴대장치를 통해 수집된 데이터도 점점 더 중요해지고 있다. 수집데이터를 기반으로 운송예측 모델을 구축하는 첫 번째 단계로 사용되었다. 시간에 따라 값을 예측하는 타임시리즈 분석은 이제 가장 인정되고 실행된 예측방법론이 되었다.

스마트 운송계획의 향후연구는 지능형 예측모델에 사용되는 모델, 사용할 수 있는 데이터 원본, 예측모델에 사용할 수 있는 비즈니스 채택 등과 같은 질문을 수용할 수 있는 로드맵이 필요하다. 이 연구에서는 운송계획 문제와 운송기술의 품질과 정확성을 측정하는 메트릭을 해결하는 몇 가지 분석기술과, 모델에 대해 설명한다. 이를 위해 ARMA(Auto Regressive Moving Average : 자동회귀이동평균) 모델, KNN, KD, Holt-Winters의 기하급수적인 스무딩 모델, 랜덤 워크 모델(random walk model), 딥 러닝(deep learning)등 최첨단 시계열 모델을 다룬다. 특히 차세대 스마트 운송시스템으로 주목받고 있는 CVs(Connected Vehicles) 기술은 현대생활에 지능적인 녹색교통을 제공한다. 이 연구에서는 커넥티드카와 스마트 운송계획에 대한 논의를 위한 방법론을 제시하기 위해 연결성, 안전 및 교통관리기법 등을 도입한다.

2. 스마트운송 분석에서 데이터소스

2.1. GPS 및 GIS 데이터

GPS 및 GIS 기술은 운송데이터, 특히 여행데이터를 수집하는 새로운 소스로 부각되고 있다. GPS 데이터는 실시간 공간 및 시간정보(거리/이동속도/여행시간 등)를 디지털 형식으로 이동경로를 보여줌으로써 정보보고의 부담을 줄일 수 있다. GPS 데이터는 두 가지 범주(실시간 추적 GPS와 로깅 GPS)가 있다, 실시간 추적 GPS 장치는 매 초당 주행차량의 좌표를 기록하는데, GPS 데이터를 사용하는 데는 다음과 같은 단점이 있다.

- 불필요한 선택기준으로 인해 통계정보를 상시적으로 나타내지 못한다는 점이다. 불분명한 기준은 결과의 정확성에 영향을 미칠 수 있다. 또한 이 방법은 모든 조사대상자가 GPS 장치를 소유하고 있어야만 가능하기 때문에 수집된 데이터가 완전하지 않을 수도 있다.
- 복잡한 데이터수집 프로세스와 데이터수집장치 추가비용 때문에 데이터수집 비용이 매우 높다[1].

2.2. 트래픽 흐름 데이터소스

센서와 탐지기는 고속도로 현장에 설치하여 운행차량대수와 관련된 데이터를 수집한다. 이러한 데이터에는 교통흐름(볼륨/시간 표시), 차선점유율 및 차량의 평균속도와 같은 특성이 포함된다. 그 후 수집된 데이터를 다양한 방법과 모델링을 통해 분석하여 데이터 기반 솔루션을 파생한다[2].

2.3. 스마트 카드

스마트 카드 데이터는 특정 운송도구를 사용하여 개인의 여행패턴을 분석하는 데 사용된다. 스마트 카드 데이터를 사용하면 데이터가 이동시작 시간, 종료시간 및 이동방향을 표시할 수 있다는 것이다. 관리자는 다양한 목적지의 빈도를 기반으로 트래픽 흐름을 예측하여 적절한 운송일정을 만들 수 있다. 그러나 스마트 카드 데이터는 특정 운송수단에서만 트래픽 데이터를 표시할 수 있기 때문에 데이터의 유연성이 제한되는 단점이 있다[3].

2.4. 휴대전화

환승을 이용하는 여행방법이 점차 복잡해지면서 여행 중에 필요한 모든 세그먼트 데이터를 차량에 반영하기에는 한계가 있다. 휴대전화 데이터는 정확한 사용자의 위치를 제공할 수 있다. 스마트폰은 오랜 기간 동안 개인수준의 위치데이터를 수집하여 왔고, 응답자의 부담이 낮아 개인 여행기록을 저장하는 것이 가능해지고 있다. 휴대전화 데이터는 차량 내 또는 차량 외 세그먼트 여행을 포함하여 여행경로에 대한 고해상도 이미지를 제공한다[4]. 그러나 휴대전화 데이터에도 몇 가지 단점이 있다. 개인 위치데이터의 양이 엄청나게 크기 때문에 데이터정리가 힘들다는 점과, 기기에 전원이 없는 경우 휴대전화에 저장된 데이터와 실제 수집된 데이터가 일치하지 않을 수도 있다. 따라서 충전장비는 데이터수집의 필수 요건이다.

2.5. 세부 통화기록(CDR)

결제를 위해 휴대전화 사업자에 의해 자동으로 수집되는 CDRs(Call Detail Records : 세부 통화기록)은 모바일 기술의 급속한 발전의 결과이다. CRD에는 익명화된 고객의 타임스탬프 및 경도/위도/좌표 정보가 포함되어 있다. 따라서 CRD에서 수집한 정보는 10년에 1~2회 정도 수행되는 기존 여행설문조사보다 더 정기적으로 비용 효율적으로 수집된다. 이에 CDR은 운송모델 분석[5]과 호환되어 개별 여행경로를 효과적으로 캡처할 수 있다[5].

3. 커넥티드 차량

3.1. 커넥티드 차량의 연결성

CVs(Connected vehicles)는 내부 및 외부 환경과 통신할 수 있는 무선연결 지원 차량을 의미한다[6]. 이러한 연결성은 차량의 상호작용을 지원하는 V2X(vehicles-to-X), V2S(vehicle-to sensor-on-board), V2V(vehicle-to-vehicle), V2R(vehicle-to-road infrastructure), and V2I(vehicle-to-Internet) 등 다양한 레벨에서 제공된다. 이러한 수준의 무선연결성은 Blue-tooth, Ultra-Wideband 등과 같은 다른 대안 외에 CV에서 가장 잘 알려진 인프라이다. 무선연결은 vehicle-to-X 연결을 제공하는 솔루션으로 이용된다.

3.2. 커넥티드 차량의 안전성

차량이 다른 V2V 및 V2I와 통신할 수 있는 기능은 도로안전을 향상시킬 수 있는 기회를 제공한다. 커넥티드 카 기술은 운영시간 동안 신뢰할 수 있는 의사결정을 강화하여 모든 측면에 영향을 미칠 수 있다. 자율주행 자동차는 이 기술의 주요 수혜자라고 할 수 있다. 인간은 로봇에 비해 반응시간과 불확실성이 높고, 로봇은 보다 안정적인 행동을 초래하는 결정을 내리기 위해 더 많은 변수를 고려할 수 있다[7].

V2V 및 V2I 통신은 도로 및 차량관련 정보를 제공한다. 운전자의 동작은 수신된 정보에 따라 영향을 받을 수 있으며, 이러한 기능을 통해 운전자 보조기술의 성능을 향상시킬 수 있다. 온보드 센서를 사용하여 시스템은 앞서가는 차량 또는 뒤 따라 오는 차량동작에 따라 가속과 공간을 조정할 수 있다.

3.3. 커넥티드 카의 스마트 교통관리

CVs(Connected vehicles)는 스마트 교통관리를 통해 교통 혼잡을 완화하기 위한 매우 유망한 기술이다. 이 기술은 교통상황에 대한 실시간 데이터를 제공함으로써 데이터품질을 개선하여 교통관리 개선으로 이어지는 결과를 제공한다[8]. CVs와 관련된 차량 안정성에 대한 2가지 유형으로 로컬 안정성과 운행 안정성이 있다. 전자는 운전자의 가속결정에 대한 차량의 반응을 나타내며, 후자는 차량그룹에 대한 운전자의 응답으로서 전체 그룹의 동작을 조사할 수 있다. 교통흐름관리 연구의 주요초점은 트래픽흐름의 운행 안정성을 조사하는 것이다. 이에 따라 커넥티드 카 환경을 갖춘 모델은 자동화 및 연결성으로 지정된 차선세트에서 완전 자율주행 차량을 운영하는 AHS(Automated Highway Systems)를 의미한다. 따라서 AHS의 성능은 차량이동 전략(제어방법)과 TMC(Traffic Management Center)의 결정에 따라 좌우된다.

4. 트래픽흐름 예측: 타임시리즈 예측모델

4.1. ARIMA 모델

ARIMA(Auto-Regressive Integrated Moving Average : 자동회귀 통합 이동평균)는 기록 값을 기반으로 예측할 수 있는 모델 클래스이다. 사용가능한 다른 방법에 비해 트래픽 흐름을 예측하는 가장 정확한 방법 중 하나로 알려져 있다. ARIMA 모델은 AR(Auto Regressive : 자동회귀)부문, MA(Moving Average : 이동평균)부문으로 구성된다. 따라서 ARIMA에서는 타임 시리즈가 고정되어 있어야 한다. 결과적으로 모든 타임시리즈 데이터에서 비 고정성을 제거하는 가장 일반적인 방법은 그것을 차별화하는 것이다. I(통합)는 데이터 값이 현재 값과 이전 값 간의 차이로 대체되었음을 나타낸다. 때로는 타임시리즈의 복잡성에 따라 두 번 이상 처리될 수도 있다. ARIMA 모델의 파라미터는 음수가 아닌 정수인 ARIMA(p,d,q)로 정의된다. p는 자동회귀모델의 순서(시간지연순서)이며, d는 차이가 있는 정도(데이터가 과거 값을 뺀 횟수), q는 이동평균 모델의 순서를 의미한다. 타임시리즈가 계절적 패턴이 있는 경우에는 계절조건을 추가해야하며 이때 계절ARIMA에 대한 짧은 SARIMA가 된다. 계절별 ARIMA 모델은 일반적으로 ARIMA(p,d,q)(P, D, Q)m로 정의된다. m은 각 시즌의 기간 수를 지칭하며, 대문자 P,D,Q는 ARIMA 모델의 계절적 부분에 대한 자동회귀, 이동평균 용어로 지칭된다. ARIMA(p,d,q) 모델은 (식 1)과 같이 나타낼 수 있다.

$$\begin{aligned} p : y_t &= \sum_{i=1}^p \alpha_i y_{t-i} + Z_t \\ d : y'_t &= y_t - y_{t-1} \\ q : y_t &= Z_t + \sum_{j=1}^q \beta_j Z_{t-j} \end{aligned} \dots\dots\dots (식 1)$$

y_t 가 시계열 관측인 경우 y_{t-1} 은 이전의 관측 값, α_p 는 자동회귀 분석계수, β_q 는 이동 평균작업 계수, Z_t 는 백색잡음을 의미한다. (식 1)을 토대로 백 시프트 연산자를 이용한 ARIMA 예측방정식은 (식 2)와 같이 정의된다.

$$\emptyset(B)(1-B)^d y_t = \theta(B) Z_t \dots\dots\dots (식 2)$$

여기서 ϕ 는 p 와 θ 의 다항도(Polynomial degree)를 나타낸다. 교통흐름은 피크 및 비수기 트래픽 조건으로 인해 계절적 패턴이 강하며, 일반적으로 오랜 기간 동안 반복되기 때문에 SARIMA(Seasonal ARIMA : 계절 ARIMA) 모델은 교통흐름 동작을 모델링하기에 적합한 선택이다.

4.2. Kalman 필터링

Kalman 필터 알고리즘(Kalman filter algorithm)은 도시지역의 실시간 트래픽흐름을 예측하는 데 사용할 수 있는 방법 중 하나이다. 이 방법은 수정 센서와 CV 모두에서 구현할 수 있다. CV는 인프라 설치가 필요하지 않기 때문에 연결된 차량에서 구동되는 데이터를 사용하는 것은 고정 센서보다 비용이 훨씬 낮으며, 보다 더 정확한 결과를 얻을 수 있다. 그러나 커넥티드 차량의 아이디어는 실제 이용 가능하기 까지 시간이 오래 걸릴 수 있다. Kalman 필터를 사용하여 트래픽 흐름을 추정하려면 다음과 같은 행렬 값을 지정해야 한다.

- F_t , 상태전환모델 : 다음상태 x_{t+1} 에 이전상태를 매핑
- H_t , 관찰된 공간에 실제상태 공간을 매핑하는 관찰모델
- Q_t , 공정소음의 공분산(Covariance)
- R_t , 관측소음의 공분산(Covariance)

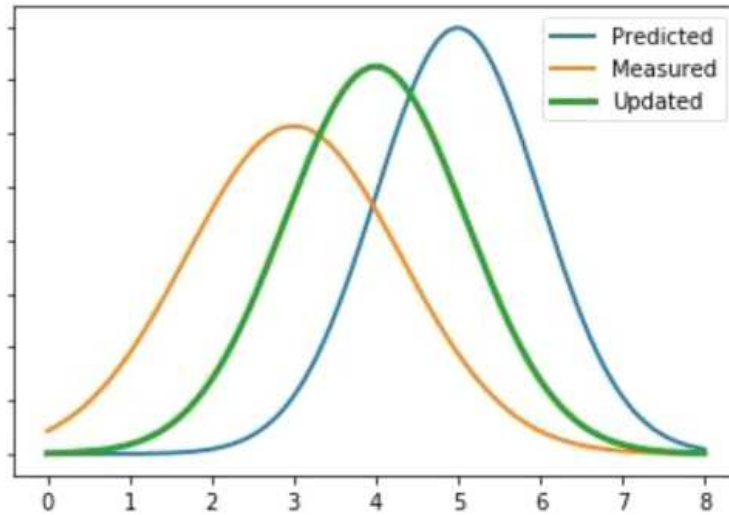
따라서 예측모델은 (식 3)과 같이 작성할 수 있다:

$$x_{t+1} = F_t x_t + w_t \quad \dots\dots\dots \text{(식 3)}$$

여기서 x_t 는 단계 t 의 상태변수이며, 모델에는 0-평균 다변가변형 일반분포 N 에서 가져온 것으로 추정되는 노이즈 벡터 w_t 가, 공변성 Q_t : $w_t \sim \mathcal{N}(0, Q_t)$ 가 포함된다. 시간 t 에서, 실제상태 x_t 의 관측 z_t 는 (식 4)와 같이 계산할 수 있다:

$$z_t = H_t x_t + v_t \quad \dots\dots\dots \text{(식 4)}$$

여기서 v_t 는 변종 R_t 와 제로평균 가우시안 백색 잡음($R_t : w_t \sim \mathcal{N}(0, R_t)$)으로 가정한 관측소음 값을 의미한다. Kalman 필터는 각 작업 단계마다 모델링이 반복적으로 업데이트되고 실시간 트래픽 흐름을 제시하기 때문에 높은 트래픽흐름에 대한 일관되고 적합한 방법을 제시한다. 아울러 다중입력 및 다중출력에도 적용할 수 있다. "예측" 및 "업데이트"라는 2단계 프로세스에서 작동함으로써 예측단계에서 Kalman 필터는 현재 상태변수와 불확실성에 대한 추정치를 생성할 수 있다. Kalman 2단계 프로세스를 [그림 1]에 나타낸다. 예측된 상태추정치는 우선 주 추정치라고도 한다. 업데이트 단계에서 예측상태는 현재 관측정보와 결합되어 상태추정치를 개선한다. 어떤 이유이든 관찰 값을 사용할 수 없는 경우 업데이트될 수 있다. 마찬가지로 여러 개의 독립적인 관찰을 동시에 사용할 수 있는 경우에는 일반적으로 다른 관측 행렬(H_t)을 통해 여러 업데이트 단계를 수행할 수 있다.



★ 자료 : Zahra Karami et al., “Smart transportation planning: Data, models, and algorithms”,
Transportation Engineering 2 (2020) 100013, p.4.

[그림 1] Kalman 2단계 프로세스

4.3. KNN 알고리즘 및 KD 트리

KNN(K-Near Neighbor) 알고리즘은 도시교통 문제를 해결하기 위한 기계학습방법 중 하나로 구현되었다. KNN 알고리즘에는 교육인스턴스 컬렉션이 있다. 각 인스턴스는 일련의 피쳐로 구성되며, 대상과 인스턴스에 가까운 가장 유사한 대상과 연결된다. 유사성은 Euclidean distance와 같은 거리 메트릭을 기반으로 (식 5)와 같이 계산할 수 있다.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots (\text{식 } 5)$$

KNN 알고리즘의 k는 새로운 인스턴스 중 가장 가까운 지점을 가리키며, 대상은 이러한 개체를 찾아 표시된다. 타임시리즈 예측의 경우 각 대상은 타임시리즈의 값 컬렉션을 나타낸다. K는 치수 수를 나타내며, KD트리는 공간을 다른 섹션으로 나누는 다차원 데이터 구조를 의미한다. KNN 알고리즘은 각 턴마다 모든 점을 계산하는 대신 지정된 위치사이의 가장 가까운 점을 계산한다. 타임시리즈 예측방법별 비교결과를 <표 1>에 나타낸다.

<표 1> 타임시리즈 예측방법별 비교결과

방법론	장점	단점	응용프로그램
ARIMA	다른 방법보다 더 정확하다.	대용량의 기록데이터가 필요하다.	- 교통 흐름[18] - 도로 교통 예측 [26] - 도시 도로 이동 시간 예측[27]
Kalman 필터링	재귀예측기록이며, 기록측정 데이터가 필요하지 않다.	단기예측에 적합하며, 정확도가 만족스럽지 못하다.	- 트래픽 흐름 예측[28] - 트래픽 스트림 밀도 추정[29] - 단기 교통 흐름 예측 [13].
Holt Winters' Exponential Smoothing	데이터가 추세와 계절성을 모두 가지고 있을 때 잘 작동한다.	데이터 중 계절성 기간을 찾는 것은 어려운 일이다.	- 승객 흐름 예측[30] - 네트워크 트래픽 모델링[31] - 일일 교통 예측[32]
랜덤	이전 값이 아니라	단기 예측에 적합하다.	- 도시 교통 시뮬레이션 및 최적화[33]

워크 모델	현재 관찰에만 의존한다.		- 지능형 운송 모델링[34] - 모바일 임시 네트워크[35]
KNN	간단한 구현을 통해 데이터의 분포에 대한 사전지식이 거의 또는 전혀 없을 때 적합하다.	대용량 데이터에 대한 성능이 저하하거나, 관련이 없거나 중복된 기능에 매우 민감하다.	- 지능형 교통[22] - 교통량 예측[23] - 이동 시간 예측[24] - 교통 상태 예측[25]

* 자료 : Zahra Karami et al., “Smart transportation planning: Data, models, and algorithms”, Transportation Engineering 2 (2020) 100013, p.6.

5. 비즈니스 기회: 트래픽 예측 애플리케이션

실시간 교통상황을 예측하는 기회를 통해 고부가가치가 있는 운송데이터 분석기업의 발전에 따른 풍부한 비즈니스 기회가 있다. 한 가지 사례로 개별 사용자에게 트래픽 예측결과를 제공하는 것이 있다. 일기예보와 마찬가지로 기업은 교통예측이라는 애플리케이션을 개발할 수 있다. 예를 들어, Google맵은 현재의 실시간 트래픽 흐름만 표시된다. 소비자가 언제 떠날지 계획할 수 있도록 다음 몇 시간 동안 직장에서 집으로 돌아오는 데 필요한 시간을 알고 싶다면 Google맵은 이를 지원할 수 없다. 이러한 서비스를 지원하기 위해서 Google은 이러한 모델을 구현하여 정확한 예상 여행시간과 소비자수요를 충족할 수 있는 가장 빠른 경로를 보여줄 수 있을 것이다. 또한 도시의 24시간 교통상황을 보여주는 일기예보 응용프로그램과 마찬가지로 교통예측이라는 기능을 추가할 수도 있다.

6. 결론

정부와 개인이 교통계획 개발에 대한 발전을 통해 도시교통 상태를 전망하는 것이 보다 더 수월해졌다. 그럼에도 불구하고 분석프로세스와 데이터추출은 상대적으로 높은 정밀도를 요구하고 있어 이를 수용하기에 어려움을 겪고 있다. 모바일 카드 데이터와 스마트 카드 데이터의 조화는 각 개인의 여행활동을 보다 밀접하게 연결시켰다. 그러나 현재의 기술은 높은 수준으로 사용할 수 있도록 오랜 시간이 필요하다. 이전 연구에 대한 포괄적인 검토를 바탕으로 교통예측 방법을 개발하는 것은 여전히 개방형 연구 영역이다. 따라서 이 연구에서는 다양한 데이터소스, 예측모델 및 특성, 향후 운송 산업 분석모델에 대한 비즈니스 기회에 중점을 두고 스마트 운송계획 연구를 위한 로드맵을 제시하였다.

급속한 도시화 및 인구증가에 따라 스마트 운송은 현대사회의 필수 구성요소가 되었다. 이를 위해 기계학습기술 등 여러 산업적 방법을 이용한 광범위한 연구 활동은 새로운 스마트 운송 분야의 길을 열었다. 이 연구에서는 교통트래픽 제어 및 예측을 위한 정보수집, 스마트하고 친환경적인 교통수단 구현을 위한 CVs(Connected Vehicles) EMD 지능형 운송계획에 따른 데이터, 방법론 및 모델을 제시하고 있다. 이 연구결과는 국내 대도시의 도심, 고속도로 및 자동차 전용 도로에서의 원활한 교통흐름을 위한 정책에 벤치마킹 자료로 활용할 수 있을 것이다.

[참고문헌]

- [1] A. Huang, D. Levinson, Axis of Travel: Modeling Non-Work Destination Choice With GPS Data, Transportation Research Part C: Emerging Technologies, 2015.
- [2] Lau, P.L.B. & Marakkalage, S. & Zhou, Y. & UL Hassan, N. & Yuen, C. & Zhang, M. & Tan, U.-X.. (2019). A survey of data fusion in smart city applications.
- [3] Schmöcker, J.-D. & Kurauchi, F.. (2017). Public transport planning with smart card data.
- [4] A. Carrel, P.S. Lau, R.G. Mishalani, R. Sengupta, J.L. Walker, Quantifying transit travel experiences from the users' perspective with high-resolution smartphone and vehicle location

data: methodologies, validation, and example analyses, Transp. Res. Part C 58 (2015) 224–239.

- [5] Vanhoof, M. & Reis, F. & Smoreda, Z. & Ploetz, T.. (2018). Detecting home locations from CDR data: introducing spatial uncertainty to the state-of-the-art.
- [6] N. Lu, N. Cheng, N. Zhang, X. Shen, and JW. Mark, (2014), Connected vehicles: solutions and challenges
- [7] M. Treiber, A. Kesting, D. Helbing, Influence of reaction times and anticipation on stability of vehicular traffic flow, Transport. Res. Rec. 1999 (1) (2007) 23–29.
- [8] A. Talebpour, HS. Mahmassani (2016), Influence of connected and autonomous vehicles on traffic flow stability and throughput.
- [9] A. Emami, M. Sarvi, S.A. Bagloee (2019), Using Kalman filter algorithm for short-term traffic flow prediction in a connected vehicle environment.

<원문제목> Smart transportation planning: Data, models, and algorithms

<원문출처> Transportation Engineering 2 (2020) 100013