

DSRC와 TCS 정보를 이용한 고속도로 경로통행시간 예측

장현호* · 윤병조**

Chang, Hyun-ho*, Yoon, Byoung-jo**

Forecasting of Motorway Path Travel Time by Using DSRC and TCS Information

ABSTRACT

Path travel time based on departure time (PTTDP) is key information in advanced traveler information systems (ATIS). Despite the necessity, forecasting PTTDP is still one of challenges which should be successfully conquered in the forecasting area of intelligent transportation systems (ITS). To address this problem effectively, a methodology to dynamically predict PTTDP between motorway interchanges is proposed in this paper. The method was developed based on the relationships between traffic demands at motorway tollgates and PTTDPs between TGs in the motorway network. Two different data were used as the input of the model: traffic demand data and path travel time data are collected by toll collection system (TCS) and dedicated short range communication (DSRC), respectively. The proposed model was developed based on k -nearest neighbor, one of data mining techniques, in order for the real applications of motorway information systems. In a feasible test with real-world data, the proposed method performed effectively by means of prediction reliability and computational running time to the level of real application of current ATIS.

Key words : Big data, Motorway, Departure time based path travel time, TCS, DSRC, k -Nearest Neighbor

초 록

출발지 기준 고속도로 경로 통행시간(PTTDP)은 첨단여행자정보시스템(ATIS)의 핵심 정보이다. 이러한 필요성에도 불구하고, 지능형교통체계(ITS)의 예측분야에서 PTTDP에 대한 연구는 성공적으로 극복해야할 핵심 도전과제중 하나로 남아있는 상태이다. 이러한 문제점을 효과적으로 극복하기 위하여, 본 연구에서는 고속도로 IC간 경로통행시간을 동적으로 예측하는 방법론을 제시하고자 한다. 제안된 모형은 고속도로 망에서 TG의 교통수요와 TG간 출발지기준 경로통행시간간의 관계를 기반으로 개발되었다. 모형의 입력 자료로(TCS로 수집되는) 통행수요와(DSRC로 수집되는) 경로통행시간 자료가 이용되었다. 개발 모형은 고속도로 정보시스템에 탑재/운영하기 위하여 Data Mining 기법중 연산속도가 빠른 k -최근린 이웃을 이용하였다. 실제 자료를 이용한 적용 실험에서, 제안된 모형은 예측의 신뢰성과 연산수행속도 측면에서 ATIS에 적용이 가능한 수준의 성능을 보였다.

검색어 : 대용량 자료, 고속도로, 출발지기준 경로통행시간, TCS, DSRC, k -최근린 이웃

1. 서론

첨단여행자정보시스템(Advanced traveller information system, ATIS)는 지능형교통체계(Intelligent Transportation Systems, ITS)의 핵심 시스템 중 하나이다. 그리고 ATIS의 주요한 기능중 하나는 통행시간 정보를 수집/가공/제공하는 것이다. 현재의 ATIS

* 서울대학교 BK조교수, 공학박사 (Seoul National University · netterk@hanmail.net)

** 정회원 · 교신저자 · 인천대학교 도시과학대학 도시공학과 부교수, 공학박사

(Corresponding Author · Incheon National University · bjyoon63@inu.ac.kr)

Received September 11, 2017/ revised September 15, 2017/ accepted November 17, 2017

통행시간 정보는 지점 또는 구간정보를 수집하고 이를 제공함에 있어 Reactive와 Proactive 방식중 하나를 적용하고 있다. Proactive 방식은 시간대(t)에서 시간대(t+1)를 예측을 수행하는 단기 통행시간 예측기법을 적용한 링크기반 통행시간 정보제공을 기본으로 하고 있으며, 예측기법으로는 칼만필터링, 인공신경망, ARIMA 등 다양한 시계열자료 예측기법이 적용되고 있다. 기존의 링크기반 통행시간정보 예측 및 제공은 도시지역과 같이 단거리 통행이 높은 도로구간에서 유용하다. 그러나 고속도로와 같이 단/중/장거리 통행이 혼재하는 경우, 지점정보를 이용한 중/장거리 통행시간 예측에는 한계가 있다. 또한 경로통행시간 정보는 도착지기준 시각으로 수집되기 때문에 출발지기준 시각과는 상당한 시간차(1~6시간)가 발생한다. 따라서 운전자에 원하는 출발지기준 시점에서 경로통행시간을 예측/제공하지 못하고 있는 실정이다.

따라서 본 연구에서는 한국도로공사 폐쇄식 구간에서 수집되는 이력 DSRC (Dedicated Short Range Communication) 경로통행시간 자료와 요금징수시스템인 TCS (Toll Collection System)에서 수집되는 요금소(Tollgate, TG) 유출입 교통 수요자료를 이용하여 단/중/장거리 출발지 시각기준 경로통행시간 예측기법을 개발하였다. 예측기법의 개발은 이력자료와 실시간 자료의 효율적인 활용을 위하여 이력자료기반 모형중 연산수행속도가 빠른 k -최근인 이웃(k -nearest neighbor, KNN) 기법을 기반으로 개발하였다. 모형의 적용성을 고려하여 해당 고속도로 노선에 설치된 TG간의 경로 쌍에 대한 통행시간을 1회 연산에 예측할 수 있도록 모형을 설계함으로써 수행속도를 단축하였다. 정부고속도로를 대상으로 개발모형을 적용한 결과, 개발된 기법의 예측값과 관측값의 오차는 수용가능한 수준으로 분석되었다. 그리고 모형의 연산수행속도는 정보시스템의 데이터 흐름을 충족할 수준으로 분석되었으며, 이는 실제 시스템 적용에 무리가 없음을 의미한다.

2. 기존 연구

본 연구에서 개발된 방법론은 KNN기법을 기반으로 고속도로 TG간 출발지 기준 경로통행시간을 예측한다. 따라서 기존연구 고찰은 기존의 통행시간 예측기법과 KNN 기법을 적용한 연구로 한정하도록 한다.

기존의 예측기법은 실시간 시계열자료를 이용하여 시간대(t)에서 시간대(t+1)의 구간통행시간 상태를 예측하였다. 예측기법으로는 칼만필터링, 인공신경망, ARIMA, 평활화 등의 기법이 주로 적용되었다. 이러한 예측기법은 엄밀히 말해 단일 시간대(single-interval) 통행시간 예측기법으로서 일종의 이동평균(moving average) 방식이다. 따라서 단기 시계열 예측기법의 특성상 시계열적 변동이 큰 상태에서 예측력이 현저히 감소하게 되며, 시간차짐(time lag)을

극복할 수 없는 한계를 가지고 있다. 이상의 기존 단일시간대 예측기법은 통행시간 및 교통량 시계열자료의 시계열적 변동 특성과 시간차짐을 극복할 수 없기 때문에 사실상 다중시간대에 걸치는 상태변화를 예측하기에는 한계가 있다. 이는 실시간으로 수집되는 시계열자료를 주로 이용하기 때문에 장래에 대한 불확실성(uncertainty)을 담보할 수 없기 때문이다.

기존 연구 고찰 결과, 고속도로 중/장거리 경로통행시간의 예측에 관한 국내의 연구는 미비한 것으로 나타났으며, 지속적인 연구개발이 필요한 부분이라 판단된다. 이는 출발지 기준 경로통행시간의 필요성에도 불구하고, 연구의 난이도로 인한 심도 깊은 연구가 수행되고 있지 않기 때문으로 판단된다. 경로통행시간 관련 연구는 과거 이력자료를 이용한 경로통행시간의 ①추정과 ②예측으로 구분할 수 있다. 경로통행시간 추정과 경로통행시간 예측은 장래 교통상태의 불확실성이라는 측면에서 다르다. 경로통행시간 추정은 링크 또는 경로통행시간이 수집된 이후에 실제 경로통행을 추정하는 분야이며, 경로통행시간 예측은 경로통행시간이 수집되지 않은 상태에서 장래의 교통상황을 고려하여 경로통행시간을 예측하는 분야이다. 따라서 링크통행시간으로 경로통행시간을 추정하기 위해서는 1~6시간에 걸치는 링크통행시간을 예측해야 하며, 이는 ITS 통행시간 예측분야에서 해결하지 못한 난제중 하나이다.

①경로통행시간 추정 연구의 경우, Namkoong (2005)은 TCS에서 수집되는 경로통행시간 정보를 이용하여 고속도로 TG간 과거 경로통행시간 정보를 추정하였다. Sim et al. (2013)은 DSRC 통행시간자료를 이용하여 링크기반 경로통행시간과 순 경로통행시간을 비교하였다. 그들의 연구에서 링크기반 경로통행시간과 순 경로통행시간의 차이는 통계적으로 동일하게 분석되었으나, 통행시간의 상태변동에 있어 차이가 발생하였다. ②경로통행시간 예측에 관한 선두적인 연구로서 Namkoong and Noh (1996)은 인공신경망을 이용하여 서울-대전간 통행시간을 예측하였으며, 단일시간대 기법의 적용으로 인한 한계로 인하여 예측오차는 6.5~12.5%까지 나타났다. 그러나 인공신경망의 경우 모형의 연산수행속도가 정보제공시스템에서 요구하는 요구시간을 충족하지 못하기 때문에 실제 적용사례는 보고되고 있지 않은 실정이다. Chang et al. (2010)은 이러한 한계를 극복하고자 이력자료와 버스정류장간 출발지 기준 경로통행시간 현행 자료를 이용하여 출발시각 기준 버스경로통행시간을 예측하였으며, 이력자료를 활용하기 위한 방안으로 KNN기법을 적용하였다.

KNN기법을 적용하기 위해서는 방대한 양의 이력자료가 요구되며, 과거 상태와 현재 상태를 이용한 장래 상태의 예측에 있어 다양한 분야에 이용되고 있다. Robinson (1983), Mulhern and Caprara (1994)는 혼재된 무질서 상태에서 비선형 시계열 예측 문제를 KNN으로 모형화하였다. Mulhern and Caprara (1994)의

연구는 무질서 상태에서 KNN이 장점을 갖는다고 언급하였다. Karlsson and Yakowitz (1987)은 KNN을 기반으로 강우량을 예측하였으며, Mulhern and Caprara (1994)는 시장반응(market response)의 예측에 활용하였다.

교통분야에서 KNN은 교통량 또는 통행시간의 단기예측(5~15 분)에 주로 적용되었다. Davis and Nihan (1991), Smith et al. (2002)은 단일시간대 (single interval) 교통량 예측에 적용하였으며, Qi and Smith (2004)는 유고지속시간을 예측하였다. Oswald et al. (2000)은 KNN 기법을 이용하여 고속도로 교통량을 예측하였으나, 예측오차는 10% 이상으로 나타났으며, 시간적 주기성에 대한 고려가 없었다. 이후 KNN기반의 단일시간대 예측은 다중시간대 예측으로 확장되었다. Chang et al. (2010)은 BMS 이력자료와 실시간 통행시간 자료를 이용하여 버스 정류장간 출발지 기준 경로통행시간을 예측하였으며, 이 연구의 의의는 방대한 양의 이력 정보를 이용하여 다중시간대에 걸친 불확실성을 효과적으로 극복할 수 있는 방안을 제시한 것에 있다고 할 수 있다. Chang et al. (2012a)는 단일시간대 교통량 예측을 다중시간대 교통량 예측으로 확장하였으며, Chang et al. (2012b)는 다중시간대에 걸친 다중 교통량 변수의 누락자료 보정으로 확장하였다. Yoon and Chang (2014)는 극심한 변동을 보이는 단속류 교통량의 예측범위를 예측 오차 10%내에서 다중시간대로 확장하였다.

이상에서 살펴본 바와 같이, (수용가능한 오차수준에서) 출발시각 기준 고속도로 중/장거리 경로통행시간의 예측에 관한 연구는 전무한 실정이라 할 수 있겠다. 그러나 최근의 급속한 Hi-pass 단말기 보급에 따라 DSRC시스템의 경우 양질의 경로통행시간 정보가 수집되고 있으며, 한국도로공사 요금정수시스템인 TCS을 통해 고속도로 망의 정확한 통행수요가 수집되고 있다. 이러한 첨단자료의 예측분야 활용 가능성에도 불구하고, 이를 이용한 출발시각 기준 경로통행시간 정보예측에 관한 연구는 수행되고 있지 않다. 따라서 본 연구에서는 KNN기법을 기법으로 이력 DSRC 경로통행시간 정보와 실시간 TCS 통행량 수요를 이용하여 출발시각 기준 경로통행시간을 예측하고자 한다.

3. 방법론

3.1 개발모형의 기능 정립

3.1.1 유동적인 TG간 경로통행시간

TG간 경로통행시간은 해당 노선에 운영 중인 TG의 개수를 N개라 할 때 $N \times (N-1)$ 개의 경로가 존재한다. 따라서 개발되는 모형은 TG간의 모든 경로 또는 선별된 주요 경로 쌍에 대한 예측이 가능하여야 한다. 또한 개별 TG쌍간에 출발지기준 경로통행시간을 예측하게 될 경우 많은 연산시간이 소요된다. 따라서 1번의 예측으

로 TG쌍간에 경로통행시간 예측이 가능하도록 모형을 설계하여야 한다.

3.1.2 동적 특성의 고려

통행시간자료의 다양성은 주중/주말, 그리고 각 요일별 시간대별 시계열 변동에 포함되어 있다. 따라서 시계열자료에 평활화 과정을 적용할 경우 시계열자료의 변동과 내재적 정보는 희석되게 되며, 추정되는 상태의 의미있는 시계열적 변동은 감소한다. Fig. 1은 경부고속도로 대전TG→서울TG간의 시계열적 변동과 시간대별 (15분 간격) 주기성을 보여주고 있으며 매우 가변적이고 복잡한 초기 결정론적 행태(initial deterministic behavior) 즉, 카오스(chaos)한 상태를 보이고 있다. 따라서 예측기법은 카오스한 시계열적 변동과 특정 시간대별 주기성을 고려할 수 있어야 한다. 그리고 예측된 경로통행시간은 교통수요에 민감하면서도 신뢰성을 담보해야 한다.

3.1.3 장애 상태의 불확실성 극복

고속도로 중/장거리 경로통행은 1~6시간까지 변화한다. 이는 실시간으로 수집되는 경로통행시간 정보는 1~6시간 전에 출발한 차량이 도착한 통행시간이라는 의미이다. 그리고 1~6시간 정도의 시간이면 교통류의 시간적 진화의 측면에서 교통상황은 이미 바뀌어 있을 가능성이 매우 높다. 예를 들어, 비혼잡한 시간인 7시에 출발한 서울TG→대전TG의 평균통행시간이 90분 일 경우, 경로통행시간은 8시 30분에 수집된다. 수집된 정보를 8시 30분에 서울TG를 출발하여 대전으로 향하는 통행자에게 제공할 경우, 이미 고속도로 혼잡이 발생했을 수도 있으며 운행 중에 혼잡을 경험할 수 있다. 따라서 예측의 관점에서 볼 때, 출발시각 기준 경로통행시간을 예측하기 위해서는 1~6시간 동안의 장애 교통상황에 대한 불확실성을 극복해야 한다. 그리고 예측된 경로통행시간은 수용 가능한 오차의 범위 내에서 신뢰성을 담보해야 한다.

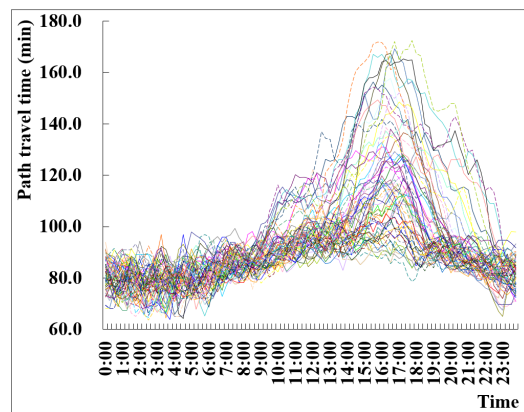


Fig. 1. Variation and Periodicity of Path Travel Time

3.1.4 현장 운영의 용이성

모형의 예측력이 요구수준보다 우수하더라도 실제 시스템에서의 모형운용을 만족해야 한다. 시간 임계적인(time-critical) 정보시스템에서 예측 연산에 할애 가능한 시간은 매우 짧다. 따라서 모형의 예측력이 요구수준 이상을 담보할 경우, 시스템에서 요구되는 연산 수행속도를 만족해야 한다. 이는 기존의 다양한 인공지능망 모형이 제안되었음에도 불구하고 실제 교통정보시스템에서 실용화되지 못하고 있는 이유 중 하나이다. 그리고 현장 전문가들의 경우 고도화된 예측 모형에 대한 이해가 부족한 경우가 많다. 따라서 모형의 구조 및 파라미터 정산 등 모형의 운영에 있어 문제가 발생한다. 반면, 현장 전문가들은 데이터 구조에 전문적 지식을 보유하고 있다. 따라서 개발되는 모형은 파라미터를 최소화하고 더불어 모형의 개선 및 구조변경이 용이해야 한다.

3.2 시간차짐 및 예측 영역

본 연구는 기존의 단일시간대 기반 링크통행시간의 한계를 극복하기 위하여 출발시간 기준 고속도로 경로통행시간을 예측하고자 한다. 현재 고속도로 TG간 경로통행시간 정보인 TCS와 DSRC 통행시간 자료는 도착시간 기준으로 수집되고 있으며, 제공정보 또한 도착시간 기준으로 제공되고 있다. 그러나 고속도로는 중/장거리 통행의 빈도가 높기 때문에 경로통행시간에 비례하여 시간차짐(time lag) 현상이 발생한다. Fig. 2 도착시간 기준 경로통행시간 제공시의 문제점을 보여주고 있다. 7시에 출발하여 도착한 60분의 경로통행시간을 8시에 제공할 경우 오차는 약 35분이 발생하며, 17시에 출발하여 19시 30분에 도착한 150분의 경로통행시간을 19시 30분에 제공할 경우 오차는 약 35분에 달한다. 이와 같이 통행시간은 여러 시간대에 걸쳐기 때문에 기존의 단일시간대 링크 통행시간 예측기법을 적용할 경우 예측오차는 증폭된다. 그러나 링크기반 통행시간 예측기법을 이용하여 이러한 경로통행수집에서

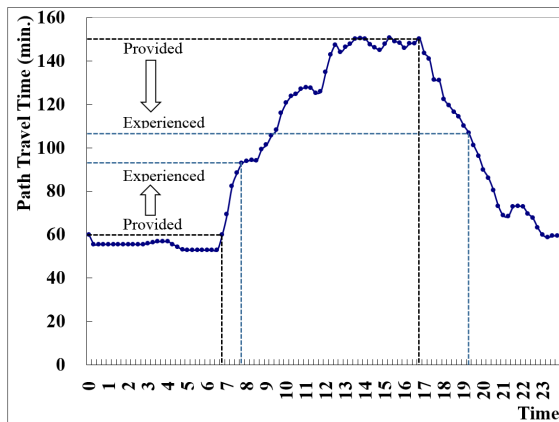


Fig. 2. Limitations of Providing Arrival Time Based Path Travel Time

발생하는 시간차짐을 극복한 연구는 보고되지 않고 있다.

Fig. 3은 경부고속도로 서울-대전 구간의 유입/유출 교통량(대/15분)에 따른 대전TG→서울TG의 경로통행시간이 Chaos한 상태임을 보여주고 있다. 이는 Chaos한 상태에서 우수한 패턴인식을 보이는 KNN의 적용가능성을 의미한다고 하겠다. 따라서 본 연구에서는 (실시간 교통상황을 반영할 수 없는) 지나치게 긴 시간차로 수집되는 경로통행시간 자료가 아닌 고속도로 네트워크의 유입/유출 현행 실시간 교통량 시계열 자료, 이력 유입/유출 교통량 자료 그리고 TG간 이력 출발시간 기준 경로통행시간자료를 이용하여 시간대(t)에서 출발TG→도착TG간 출발시간 기준 경로통행시간을 예측하고자 한다.

3.3 예측모형 개발

3.3.1 모형의 구조

본 연구에서 제안되는 출발시간 기준 고속도로 경로통행시간 모형은 Fig. 4와 같이 ① KNN 정보구축과 ② 예측치 산출로 구성된다. ① KNN 정보구축 과정은 상태거리(state distance) Metric을 이용하여 현행 TCS 통행수요 상태와 유사한 과거 TCS 통행수요 상태를 검색하고, 현행 TCS 통행수요 상태와 가장 유사한 k 개의 과거 TCS 통행수요 상태에 해당하는 과거의 출발 TG(r)과 도착 TG(s)간의 출발시간 기준의 k 개 경로통행시간 표(table)을

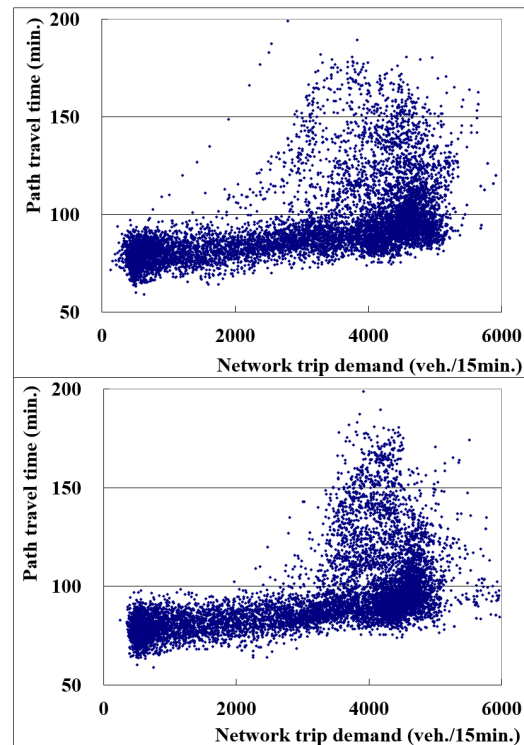


Fig. 3. Relationships Between Network Trip Demand and Path Travel Time

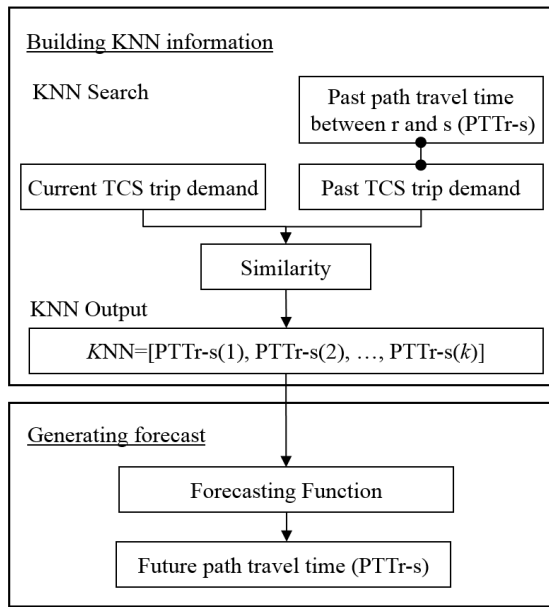


Fig. 4. Structure of the Proposed KNN Model

구축하는 과정이다. 그리고 ② 예측치 산출 과정에서 구축된 k 개 경로통행시간 표(table)와 예측함수를 이용하여 r - s 간 경로통행시간 표를 예측한다. 이상의 변수는 상태벡터를 이용하여 구성된다.

3.3.2 상태벡터의 구성

상태벡터는 입력 상태벡터(input state vector)와 출력 상태벡터(output state vector)로 구성된다. 입력 상태벡터는 과거(past) 상태벡터와 현행(current) 상태벡터인 과거 TCS 통행수요 상태벡터와 현행 TCS 통행수요 상태벡터로 구성된다. 그리고 출력 상태벡터는 과거 TCS 통행수요 상태벡터에 해당하는 기종점(r - s)간 경로통행시간 출력 상태벡터로 구성된다. 따라서 본 연구의 경우 상태벡터는 다음과 같이 정의된다. TCS 통행수요 상태벡터는 과거 유입/유출 통행 수요 상태벡터($q_{h,in}, q_{h,out}$), 시간대(t)에서 TG r - s 간 과거 출발시간 기준 경로통행시간 상태벡터($c_h(r,s)$), 현행 유입/유출 교통량 상태벡터($q_{c,in}, q_{c,out}$), 시간대(t)에서 TG간 출발시간 기준 예측 경로통행시간 상태벡터($c_p(r,s)$), 그리고 출발지 TG(r)와 도착지 TG(s)간 경로통행시간 정보생성을 위한 경로정보 상태벡터($n(r,s)$)로 정의한다.

$q_{h,in} = [q_{h,in}^j(m)], q_{h,out} = [q_{h,out}^j(m)]$: h 번째 이력자료 j 번째 TG의 시간대(t - m)에서 시간대(t)까지 유입/유출 교통량 상태벡터 ($1 \times m$), ($h \in H$). 여기서 H 는 이력자료의 개수이다.

$q_{c,in} = [q_{c,in}^j], q_{c,out} = [q_{c,out}^j]$: j 번째 TG의 current 유입/유출 교통량자료의 시간대(t - m)에서 시간대(t)까지 상태벡터($1 \times m$)

$c_h(r,s)$: TG r - s 간의 h 번째 이력 경로통행시간 상태벡터($r \times s$),

($h \in H$)

$\hat{c}_p(r,s)$: TG r - s 간의 출발시간 기준 예측 경로통행시간 상태벡터 ($r \times s$), ($i \in k$)

$n(r,s)$: 0, 1로 구성되는 TG간 경로집합($r \times s$)

3.3.3 상태간 유사성 척도

본 연구의 제안되는 KNN기반 예측 방법론은 현행 유출입 교통량 상태벡터($q_{c,in}, q_{c,out}$)와 과거 유/출입 교통량 상태벡터($q_{h,in}, q_{h,out}$)의 유사성(similarity)에 기초한다. 다시 말해, 유사성에 기반하여 현행 유출입 교통량 상태벡터($q_{c,in}, q_{c,out}$)와 과거 유/출입 교통량 상태벡터($q_{h,in}, q_{h,out}$)를 탐색하는 과정을 통해 k 개의 과거 경로통행시간 상태벡터($c_h(r,s)$)로 구성된 근집 $c_p(r,s)$ 을 구축한다. 상태간 유사성으로는 Eq. (1)과 같은 L_m 상태거리(state distance)가 주로 이용된다. 본 연구의 경우, $m = 2$ 인 Euclidean 상태거리를 적용하도록 한다.

$$L_m = \left(\sum_i^d |x_i - y_i|^m \right)^{1/m}, \quad d \geq i \quad (1)$$

Eq. (1)의 L_m 상태거리를 이용하여 [$q_{c,in}^j(m), q_{c,out}^j(m)$]와 [$q_{h,in}^j(m), q_{h,out}^j(m)$]간의 상태거리를 표현하면 Eq. (2)와 같으며, 두 상태간의 거리는 u_h 로 정의한다($h \in H$). 여기서 H 는 이력자료의 개수이다.

$$u_h = \left[\sum_{j,m,d} \{q_{c,d}^j(m) - q_{h,d}^j(m)\}^2 \right]^{1/2} \quad (2)$$

여기서, $d = \{in, out\}$

3.3.4 KNN의 자료구조

본 연구에서 근집(neighborhood), 즉 KNN을 구성하는 이웃(neighbor) 해를 탐색/구축하기 위해서는 KNN의 자료구조가 정의되어야 한다. 근집해의 자료구조는 k 개의 이웃해로 구성되며, Table 1과 같이 입력 상태벡터와 출력상태 벡터로 구성된다. 입력 상태벡터는 일반적인 변수기반 통계적 방법론의 독립변수에 해당하며, 출력 상태벡터는 종속변수에 해당한다.

3.3.5 예측함수

근집해가 구축되면 예측함수를 이용하여 TG간 경로통행시간($\hat{c}_p(r,s)$)을 예측하게 된다. 예측함수는 상태거리(u_i)의 역수로 $\hat{c}_p^i(r,s)$ 을 가중 평균한 방법을 이용하도록 하다($i \in k$). 상태간 거리로 가중 평균한 예측함수는 산술평균보다 안정적으로 우수한

결과를 도출하는 것으로 보고되고 있다(Smith et al., 2002; Chang et al., 2012a,b; Yoon and Chang, 2014). 이는 의사결정의 관점에서 상태간 거리가 가까울수록 장래 상태의 결정에 있어 보다 많은 의사결정 권한을 부여하기 때문이다.

$$\hat{c}_p(r,s) = \sum_{i=1}^k \left(\frac{c_i(r,s) \cdot n(r,s)}{u_i} \right) / \sum_{i=1}^k \frac{1}{u_i} \quad (3)$$

본 연구에서는 모형의 구조와 연산과정을 설명하기 위해 행렬 A와 B에 대한 연산은 다음과 같이 정의한다. 추가로, 이러한 정의는 알고리즘의 구축시 매우 유용하다.

$$A \cdot B = \begin{pmatrix} a, & b \\ c, & d \end{pmatrix} \cdot \begin{pmatrix} 1, & 2 \\ 3, & 4 \end{pmatrix} = \begin{pmatrix} a \times 1, & b \times 2 \\ c \times 3, & d \times 4 \end{pmatrix}$$

3.3.6 예측 알고리즘

본 연구에서는 군집(neighborhood)을 구성하는 방법으로 최인접 개수 k 을 정하는 KNN 알고리즘(Smith et al., 2002; Chang et al., 2012a; Yoon and Chang, 2014)을 이용하였다. 본 연구에서 KNN 알고리즘은 3단계로 구성된다. Step 1에서는 Table 1의 KNN 자료구조를 초기화한다. Step 2에서는 Eq. (2)의 상태간 거리 Metric을 이용하여 $q_{h,d}^j(m)$ 과 유사한 k 개의 $q_{h,d}^j(m)$ 을 검색하는 과정을 통해 KNN을 구축한다. Step 2를 통해 KNN이 구축되면 Step 3에서는 Eq. (3)을 이용하여 TG간 출발시간 경로통행시간 표인 $\hat{c}_p(r,s)$ 을 산정한다.

KNN 경로통행예측 알고리즘

주어진 $q_{h,d}^j(m)$, $q_{c,d}^j(m)$, 최인접 개수 k 에 대하여:

(여기서, $h \in H$, H 는 전체 과거 $q_{h,d}^j(m)$ 의 수)

Step 1: 군집해 자료구조의 1에서 k 까지 neighbor를 초기화

Step 2: KNN 탐색 및 구축

Table 1. Structure of Built KNN Information

i	Input state vector		Output state vector	
	Independent variable		Dependent variable	Similarity
1	$q_{1,d}^j(m)$	→	$c_1(r,s)$	u_1
2	$q_{2,d}^j(m)$	→	$c_2(r,s)$	u_2
...
K	$q_{K,d}^j(m)$	→	$c_K(r,s)$	u_K
...
k	$q_{k,d}^j(m)$	→	$c_k(r,s)$	u_k

Note: $d = \{in, out\}$, $1 \leq K \leq k$

모든 이력 상태벡터($q_{h,d}^j(m)$)에 대하여

Step 1-1: h 번째 과거 상태벡터($q_{h,d}^j(m)$)와 현행 상태벡터($q_{c,d}^j(m)$)간의 u_h 계산

Step 1-2: If $u_h < U_{max}$ Then

(여기서, $U_{max} = \max\{u_1, \dots, u_K, \dots, u_k\}$)

Step 1-2-1: U_{max} 에 해당하는 Neighbor를 KNN 군집에서 제거

Step 1-2-2: $q_{h,d}^j(m)$, $c_h(r,s)$, 그리고 u_h 을 KNN 군집에 갱신

Step 1-2-3: 갱신된 KNN 군집에서 U_{max} 을 탐색

Step 3: 예측함수를 이용하여 $\hat{c}_p(r,s)$ 을 산정

4. 적용 및 평가

4.1 실험 설계

4.1.1 자료 구축

본 연구의 사례대상 고속도로는 경부고속도로 서울-대전 구간이며, 경로 쌍은 서울→안성, 서울→대전으로서 4개이다. 수집된 TCS 자료는 15분 집계자료로 유출입 교통량과 출발시간 기준 DSRC 경로통행시간을 구축하였다. 수집 자료의 시간적 범위는 2015년 1월 1일에서 5월 31일이며, 마지막 주일은 경로통행시간 예측에 적용하였다. 따라서 구축된 이력자료는 1월 1일에서 5월 24일까지 총 144일간의 자료이다.

4.1.2 평가 지표

Fig. 1에서 경로통행시간은 60분에서 180분까지 변동하고 있다. 상태의 변동이 큰 경우, 절대오차는 낮은 상태 즉 경로통행시간이 작은 경우를 간과하게 된다. 상태의 변동이 큰 경우, 평균절대값백분율오차(Mean Absolute Percentage Error, MAPE (%))는 모든 상태 요소들을 등가(equivalent)로 평가하면서 직관적으로 이해가 가능하다(Smith et al., 2002; Yoon and Chang, 2014). 따라서 본 연구에서는 경로통행시간의 변동 폭을 고려하여 MAPE를 평가 지표로 설정하였으며, Eq. (4)와 같다. 그리고 미시적인 결과의 분석은 w직관적 이해가 용이한 시계열적 변동을 통해 분석하였다. 본 연구에서 제시한 모형의 파라미터인 k 값을 최적화하는데 있어 오차는 관측 경로통행시간 표와 예측 경로통행시간 표간의 상태간 거리를 이용하였다. 이는 중/장거리 경로통행시간에 우선권을 부여한 것으로서 본 연구에서 제시한 예측모형의 주된 목적인 중/장거리 경로통행시간 예측에 적합하다. 추가로, 실제 적용에 있어 단거리 경로통행시간은 기존의 링크통행시간 예측 기법으로 적용이 가능하다.

$$MAPE(\%) = \left(\frac{1}{n} \sum_{i=1}^N \frac{x_i - \hat{x}_i}{x_i} \right) \times 100 \quad (4)$$

여기서, x_i : 관측치, \hat{x}_i : 예측치, n : 자료의 개수

4.2 파라미터 최적화 및 예측력 분석

4.2.1 적정 k 값 및 오차분석 결과

Fig. 5는 k 값에 따른 관측 경로통행시간과 예측 경로통행시간의 상태간 거리를 보여주고 있다. 상태 거리는 k 값이 증가함에 따라 급격히 감소한 후 다시 서서히 증가하고 있다. 아래로 오목한 U형의 오차관계는 고속도로 TG 통행 수요와 경로통행시간 간에 패턴이 존재함을 의미한다. 그리고 그 패턴의 진화과정에서 패턴간의 경계가 분명하건 분명하지 않건 패턴은 본 연구의 방법론을 통해 인식할 수 있음을 의미한다. 또한 k 값=1인 경우에 비하여 최적의 k 값=5에서 상태간 거리는 51에서 36까지 29.4% 감소하고 있다. 이는 1개의 가장 유사한 패턴이라도 오차를 최소화할 수 없음을 의미하며, 패턴의 중심이 존재함을 의미한다. 추가로 k 값이 14보다 큰 경우 상태거리는 급격히 증가하고 있다. 이는 현행 고속도로 TG 통행수요와 유사한 패턴은 14개까지 존재한다고 할 수 있으며, 이후 다른 패턴이 포함되기 때문이다. 이상의 결과를 기반으로 추가적인 분석을 위하여 최적의 k 값은 5로 설정한다.

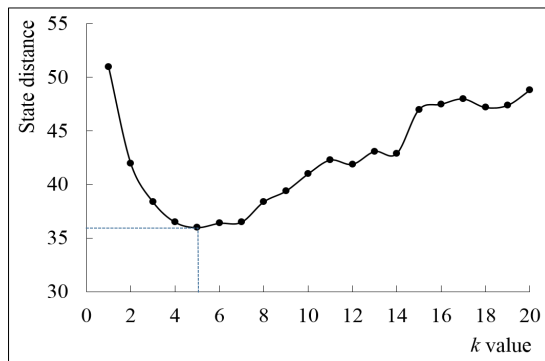


Fig. 5. Effect of k Value on Forecasting Errors

Table 2. Summary of Analysis Results (Error, %)

Day	Direction				Ave.
	Seoul → Anseong	Seoul → Daejeon	Anseong → Seoul	Daejeon → Seoul	
Mon	3.31	4.15	3.07	3.89	3.60
Fri	3.34	4.11	5.12	5.49	4.51
Sat	8.19	5.43	7.68	5.12	6.61
Sun	2.45	4.04	6.63	4.23	4.34
Ave.	4.32	4.43	5.62	4.68	4.76

또한 최적 k 값은 사전에 모의예측과정을 통하여 결정할 수 있으며, 주기적으로 갱신이 가능하다(Smith et al., 2002; Chang et al., 2012a,b; Yoon and Chang, 2014).

Table 2는 요일별 오차분석 결과를 보여주고 있으며, 4개 중/장거리 경로에 대한 관측 경로통행시간과 예측 경로통행시간의 예측 오차(%)는 평균 4.76으로 나타나고 있다. 월요일의 오차가 가장 낮게 그리고 토요일이 가장 높게 나타나고 있다. 이는 월요일이 토요일보다 패턴이 명확히 존재하는 반면, 토요일의 패턴은 월요일에 비해 임의성을 포함하고 있기 때문으로 판단된다. 방향별로는 대전방향의 예측 오차는 2.45~8.19까지 그리고 서울방향의 오차는 3.07~7.68로 나타나 대전방향의 오차 폭이 서울방향보다 넓게 나타났다. 이는 대전방향의 교통상태는 서울방향에 비하여 비반복적 행태가 강하기 때문으로 판단된다. 추가로 중거리보다 장거리의 경우 오차는 증가하고 있지만, 오차의 수준은 기존의 구간통행시간 오차보다 낮게 나타나고 있다. 이는 링크통행시간보다 경로통행시간의 경우 상대적인 변동량이 낮기 때문이다.

4.2.2 시계열 분석

Figs. 6-9는 서울-안성간, 서울-대전간 15분 간격 24시간 출발시각 기준 경로통행시간 예측결과를 보여주고 있다. Figs. 6-9에서 경로의 경우는 중거리와 장거리 각각 2개의 경우[Case1: 서울→안성, Case2: 안성→서울, Case3: 서울→대전, Case4: 대전→서울]이다.

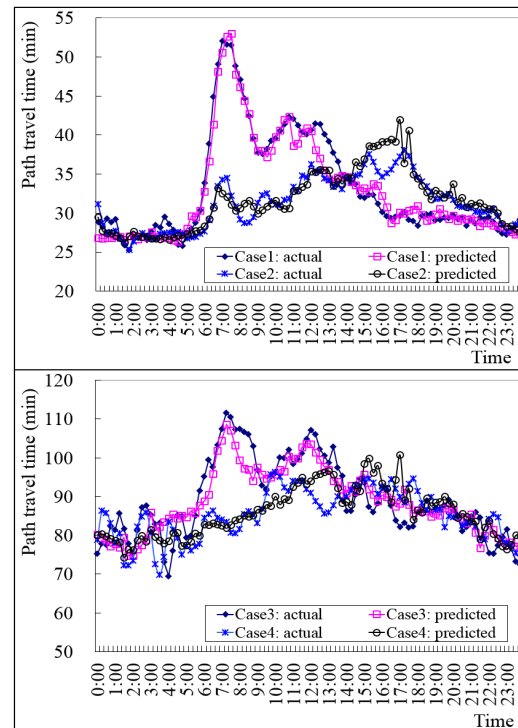


Fig. 6. Time-Series Comparison: Monday

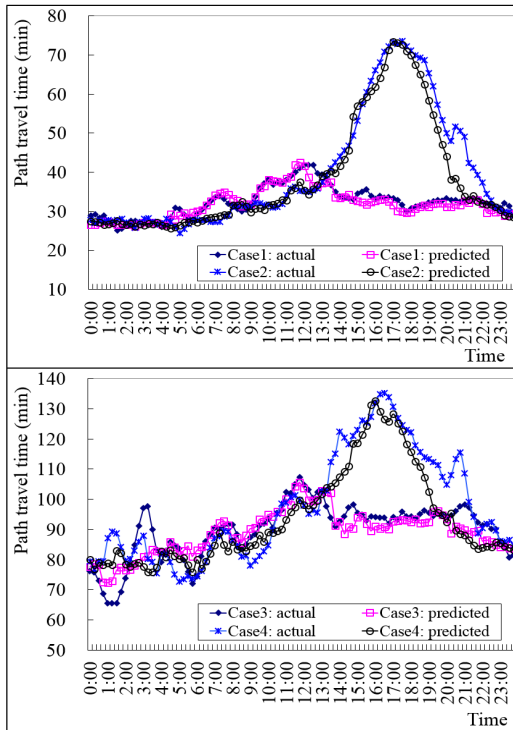


Fig. 7. Time-Series Comparison: Friday

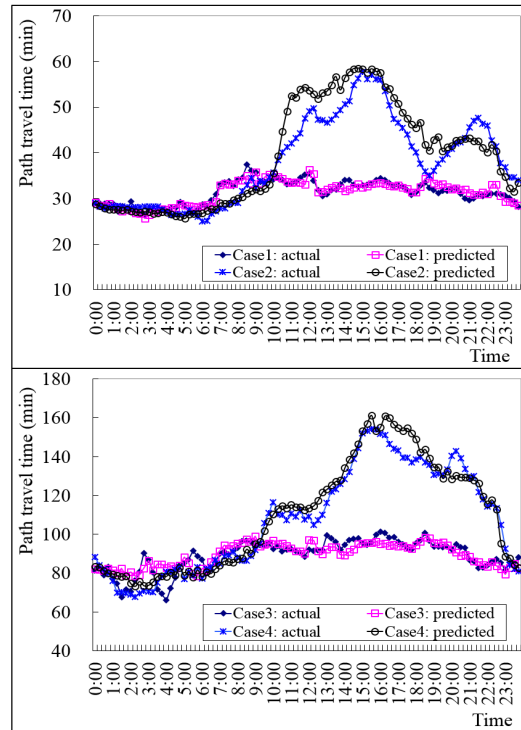


Fig. 9. Time-Series Comparison: Sunday

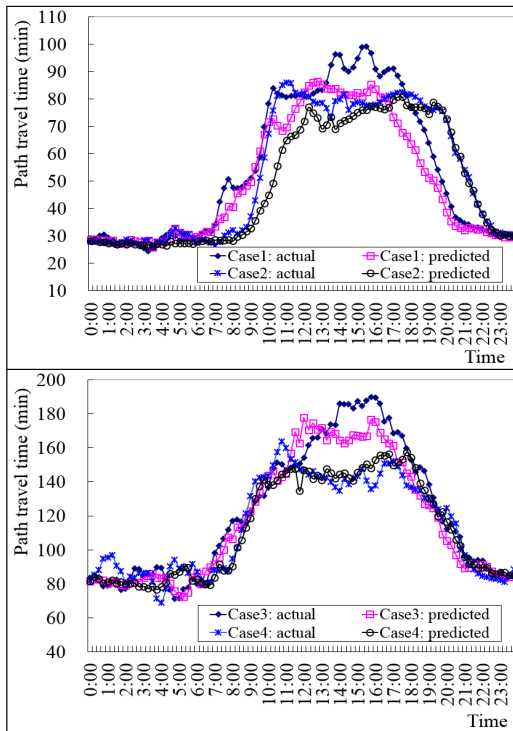


Fig. 8. Time-Series Comparison: Saturday

(경로통행시간이 급격하게 증가한 후 감소하는) 혼잡한 경우에 대한 예측력은 월요일과 금요일이 우수하게 나타났다. 토요일의

경우 다소 낮게 예측되었으나 경로통행시간대비 오차(분)는 크지 않은 것으로 판단된다. 그러나 일요일 안성-서울간의 혼잡발생시 예측 경로통행시간은 관측 경로통행시간에 비하여 최대 10분 정도의 차이를 보이고 있다. 이는 서울 TG에서 안성 TG까지 구간 내에서 혼잡이 형성되었기 때문으로 판단된다.

월요일과 금요일의 서울-대전간 심야시간대의 통행시간 변동은 크게 나타나고 있어 오차는 다소 높게 나타나고 있지만 통행시간대비 오차는 높지 않은 것으로 판단된다. 반면 서울-안성간의 심야시간대 통행시간 변동은 낮게 나타나고 있으며, 이는 휴게소에서의 체류 유무와 체류시간에 따른 변동으로 판단된다. 전반적으로 경로통행시간의 시간처짐을 평균오차 4.76%의 수준에서 극복하고 있다. 따라서 본 연구에서 제시한 방법론은 기존의 단일시간대 링크통행시간 예측기법의 한계를 넘어 효과적으로 다중경로 쌍에 대한 경로통행시간을 예측하고 있다고 판단된다.

5. 결론 및 향후연구

고속도로를 이용하는 운전자가 원하는 정보중 하나는 “출발지에서 도착지까지 경로통행시간이 얼마나 되는가?”이다. 그러나 기존의 단일 시간대 링크기반 통행시간 예측기법은(동적으로 변화하는 장래 교통상황을 반영하지 못 하기 때문에) 사실상 고속도로의 중/장거리 출발시각 기준 경로통행시간의 예측에 적용할 수 없는

한계가 있다. 이는 예측의 관점에서 볼 때 1-6시간에 달하는 장래의 교통상황에 대한 불확실성을 극복하지 못 하기 때문이다. 이러한 문제점을 극복하기 위하여 본 연구에서는 통행 수요와 통행시간의 관점에서 도속도로 네트워크의 유출입 통행 수요의 변동에 따른 경로통행시간의 변동을 이용하여 다수의 TG간 경로 쌍에 대한 출발시간 기준 통행시간을 예측하는 모형을 제안하였다. 제안된 모형은 예측력, 파라미터 최소화, 그리고 모형의 시스템 탑재시 운영자 고려사항 등을 만족하도록 개발되었다. 따라서 현행의 자료 관리시스템하에서 즉각적인 탑재 및 운영이 가능할 것으로 판단된다. 한국도로공사의 DSRC 정보는 수집된 이후 실시간 구간통행시간 정보로 활용되고 있지만, 방대한 양의 이력 DSRC와 TCS 정보는 예측에 활용되고 있지 않다. 최근 ITS 예측분야의 학술연구는 Deep learning 기반으로 진화하고 있다. 대용량 이력자료기반의 Deep learning은 인공지능경망보다는 벡터 양자화 등의 패턴인식에 기인한다. 이러한 의미에서 본 연구에서 개발된 모형은 첨단 예측방법론의 Preprocess로 활용될 수 있을 것으로 판단된다. 전문가의 입장에서 볼 때, 향후 Deep learning이 현장에 도입되기까지는 적어도 수년의 시간이 소요될 것으로 판단된다. 이는 Deep learning의 경우, 방대한 양의 시스템자료를 실시간보다 빠른 속도로 검색/패턴인식/학습 등을 수행해야 하기 때문이다. 그리고 이러한 연산부하로 인하여 현재의 ITS 정보제공시스템에서 요구하는 연산수행속도를 충족할 수 없기 때문이다. 따라서 본 연구에서 제안된 모형은 기존의 ITS 자료관리시스템 및 정보제공체계하에서 실용화가 가능한 유력한 대안 중 하나라고 판단된다. 추가로, Deep learning 기반의 중장기 통행시간 예측은 장래에 발생할 불확실성의 요인들인 교통사고, 유령혼잡(phantom jam), 교통량의 집분산, 심지어 임의적 저속차량으로 인한 혼잡 등을 Deep learning 기법이 예측할 수 있는 경우에 대하여 한정지을 수 있다. 이러한 의미에서 “Deep learning이 과연 장래 상태의 불확실성을 극복할 수 있는가?”는 독자의 판단에 맡기기로 한다.

향후 연구로는 고속도로 본선의 교통량과 DSRC 구간속도자료를 개발모형의 입력 자료로 고려할 경우 예측력은 보다 향상될 것으로 판단된다. 그리고 동적 클러스터링 기법을 적용하면 보다 국부적인 패턴을 추출할 수 있을 것으로 판단된다. 중/장거리 경로통행시간은 수집 자료의 표본수가 낮아 자료의 편차가 크게 나타날 수 있다. 따라서 인접 TG간의 통행시간을 이용하여 중/장거리 통행시간을 산정하는 기법과 결합될 필요가 있다고 판단된다.

References

- Chang, H. H., Lee, Y. I., Yoon, B. J. and Baek, S. G. (2012a). “Dynamic Near-Term Traffic Flow Prediction: System-Oriented Approach Based on Past Experiences.” *IET Intelligent Transport Systems*, Vol. 6, No. 3, pp. 292-305.
- Chang, H. H., Park, D. J., Lee, S. J., Lee, H. S. and Baek, S. G. (2010). “Dynamic Multi-Interval Bus Travel Time Prediction using Bus Transit Data.” *Transportmetrica*, Vol. 6, No. 1, pp. 19-38.
- Chang, H. H., Park, D. J., Lee, Y. I. and Yoon, B. J. (2012b). “Multiple Time Period Imputation Technique for Multiple Missing Traffic Variables: Nonparametric Regression Approach.” *Canadian Journal of Civil Engineering*, Vol. 39, pp. 448-459.
- Davis, G. and Nihan, N. (1991). “Nonparametric Regression and Short-Term Freeway Traffic Forecasting.” *Journal of Transportation Engineering*, Vol. 117, No. 2, pp. 178-188.
- Karlsson, M. and Yakowitz, S. (1987). “Rainfall-Runoff Forecasting Methods, Old and New.” *Stochastic Hydrology and Hydraulics*, Vol. 1, No. 4, pp. 303-318.
- Mulhern, F. J. and Caprara, R. J. (1994). “A Nearest Neighbor Model for Forecasting Market Response.” *International Journal of Forecasting*, Vol. 10, No. 2, pp. 191-207.
- Namkoong, S. (2005). “Progressive Iterative Forward and Backward (PIFAB) Search Method to Estimate Path-Travel Time on Freeways Using Toll Collection System Data.” *Journal of Korean Society of Transportation*, Vol. 20, No. 5, pp. 147-155 (in Korean).
- Namkoong, S. and Noh, J. H. (1996). “Development of Optimal-Path Finding System (X-PATH) using Search Space Reduction Technique Based on Expert System.” *Journal of Korean Society of Transportation*, Vol. 14, No. 1, pp. 51-67 (in Korean).
- Oswald, R. K., Scherer, W. T. and Smith, B. (2000). “Traffic Flow Forecasting using Approximate Nearest Neighbor Nonparametric Regression.” *A research project report for U.S. DOT University transportation center*.
- Qi, Y. and Smith, B. L. (2004). “Identifying Nearest-Neighbors in a Large-Scale Incident Data Archive.” *Transportation Research Report*, 1879, pp. 89-98.
- Robinson, P. (1983). “Nonparametric Estimators for Time Series.” *Journal of Time Series Analysis*, Vol. 4, No 3, pp. 185-207.
- Sim, S. W., Choi, K. C., Lee, S. S. and Namkoong, S. (2013). “An Expressway Path Travel Time Estimation using Hi-pass DSRC Off-Line Travel Data.” *Journal of Korean Society of Transportation*, Vol. 31, No. 3, pp. 45-54 (in Korean).
- Smith, B. L., Williams, B. M. and Oswald, R. K. (2002). “Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting.” *Transportation Research Part C*, Vol. 10, No. 4, pp. 303-321.
- Yoon, B. J. and Chang, H. H. (2014). “Potentialities of Data-Driven Nonparametric Regression in Urban Signalized Traffic Flow Forecasting.” *Journal of Transportation Engineering*, Online.