

# A Study of Travel Time Prediction using K-Nearest Neighborhood Method

Sung-Han Lim<sup>a</sup> · Hyang-Mi Lee<sup>a</sup> · Seong-Lyong Park<sup>b</sup> · Tae-Young Heo<sup>c,1</sup>

<sup>a</sup>Highway & Transportation Research Division, Korea Institute of Construction Technology

<sup>b</sup>ITS & Road Environment Division, Ministry of Land, Infrastructure and Transport

<sup>c</sup>Department of Information and Statistics, Chungbuk National University

(Received September 6, 2013; Revised October 10, 2013; Accepted October 15, 2013)

---

## Abstract

Travel-time is considered the most typical and preferred traffic information for intelligent transportation systems(ITS). This paper proposes a real-time travel-time prediction method for a national highway. In this paper, the K-nearest neighbor(KNN) method is used for travel time prediction. The KNN method (a nonparametric method) is appropriate for a real-time traffic management system because the method needs no additional assumptions or parameter calibration. The performances of various models are compared based on mean absolute percentage error(MAPE) and coefficient of variation(CV). In real application, the analysis of real traffic data collected from Korean national highways indicates that the proposed model outperforms other prediction models such as the historical average model and the Kalman filter model. It is expected to improve travel-time reliability by flexibly using travel-time from the proposed model with travel-time from the interval detectors.

Keywords: Travel-time prediction, nonparametric method, K-nearest neighbor, intelligent transportation system.

---

## 1. 서론

통행시간 정보는 교통정보 중에서 가장 대표적이고 이용자 선호도가 높은 정보이다. 최근 IT 기술의 발달로 인하여 교통정보를 차내에서 수신할 수 있는 개인 단말기 및 스마트 폰의 확대에 의해 보다 신속한 교통정보의 제공이 가능해짐에 따라 도로 이용자들은 현재 시점에 출발했을 때 목적지까지 걸리는 시간 즉, 미래의 시공간적 교통상태 변화가 반영된 통행시간을 원하고 있다.

통행시간 정보는 일반적으로 지점검지기 또는 구간검지기 자료로부터 산출되고 있는데, 이 자료들을 각각 이용한 통행시간 예측은 한계를 가지고 있다. 지점검지기 자료는 구간 내 일부 지점에서 수집된 공간표본자료(Spatial sample data)이기 때문에 구간 전체의 교통특성을 반영하는데 한계가 있다. 따라서 시간평균속도를 공간평균속도로 변화하는 과정에서 오차가 발생하게 된다. 또한 지점검지기로부터 산출된 통행시간은 ‘통행시간을 산출하는 대상차량과 통행시간을 제공받는 대상차량간의 출발지점 차

---

<sup>1</sup>Corresponding author: Associate Professor, Department of Information and Statistics, Chungbuk National University, Cheong-Ju, Chungbuk 361-763, Korea. E-mail: [theo@cbnu.ac.kr](mailto:theo@cbnu.ac.kr)

이'인 시간 처짐(time lag) 현상이 발생하게 된다. 이로 인해 지점검지기자료는 통행시간 정보 제공 측면에서 볼 때 과거정보이며, 시간 처짐에 따른 교통상태 변화로 인해 통행시간 오차가 발생하게 된다. 반면, 구간검지기 자료는 시점과 종점에 설치된 두 개의 검지기를 통해 수집된 자료로 구간의 동적 교통상황이 정확히 반영된 정보이다. 이로 인해 지점검지기와 같이 시간평균속도를 공간평균속도로 추정할 때 발생하는 통행시간 오차는 발생하지 않지만, 구간검지기 역시 현재 시점을 도착시점으로 하는 과거의 통행시간이기 때문에 시간 처짐 현상은 발생한다. 이처럼 통행시간을 생성하는 자료에는 시간 처짐 현상이 발생하고 있어 이용자들이 원하는 교통정보를 생성하기 위해서는 반드시 예측이 전제되어야 한다 (Lim, 2011; Lim과 Lee, 2011).

일반국도는 고속도로와는 달리 통행시간 생성 시 신호교차로나 유·출입 지점 등에 의한 지체시간이 영향을 많이 받게 된다. 지점검지기자료를 이용하여 통행시간을 생성 할 경우 링크구간 이외에 신호교차로나 유입과 출입 등의 노드에 대한 지체시간을 추가로 구해야 하는 문제가 발생하여 지점검지기 자료를 이용하는 데는 한계가 있다.

Lee와 Lee (2002) 연구에 따르면, 단일 교차로의 구간 통행시간은 적색신호의 유무에 따라 통행시간 분포가 두 개의 봉 형태를 나타내지만, 다구간의 구간 통행시간의 분포는 일정 통행시간 범위 내에 밀집되는 경향을 보였는데, 이는 통행시간에 신호교차로에 의한 지체 시간이 미치는 영향이 교차로 수가 증가할수록 상쇄되기 때문이라고 판단하였다. 이와 같이 구간검지기 자료를 이용할 경우 지체시간들이 모두 포함된 교통정보이기 때문에 지점검지기보다 정확한 교통정보를 생성할 수 있다.

따라서 본 연구에서는 구간검지기 자료를 활용하여 일반국도를 대상으로 실시간 시스템에 현실적으로 적용 가능한 통행시간 예측 방법을 개발하고자 한다.

## 2. 선행연구 고찰

신뢰성 있는 통행시간 정보 제공을 위해 다양한 예측 방법 연구가 진행되어 왔다. 연구에 적용된 통행시간 예측 방법을 시계열 모형, 회귀모형, 패턴인식 기반 방법, 교통류 기반 방법으로 구분했을 때, 시계열 모형 중 칼만 필터 방법과 패턴인식 기반 방법을 적용한 연구가 가장 활발히 이루어져 온 것으로 판단하였다. 그 이유는 칼만 필터 방법은 반복적으로 새로운 관측 값을 즉시 수용할 수 있어 온라인 시스템에서의 적용이 유리하며, 패턴인식 기반 방법은 실시간 변동성이 큰 교통자료를 예측할 경우 시계열 모형에 비해 정확도가 높기 때문인 것으로 판단된다.

가장 많이 적용되고 있는 시계열 모형과 패턴인식방법 등을 비교 평가한 연구 결과를 살펴보면, Smith와 Demetsky (1997)는 이력자료평균방법, 시계열모형, 신경망모형, 그리고 K 최대근접이웃(K-Nearest Neighbor; KNN) 방법을 이용하여 향후 15분 동안의 교통량을 예측한 결과 KNN 방법의 예측 정확도가 가장 높은 것으로 분석되었다. 또한 Lam 등 (2006a, 2006b)의 연구에서는 단기교통량 예측 방법 중 KNN 방법이 일별 시간대별 교통량 변화를 가장 잘 반영하는 것으로 분석되었으며, You와 Kim (2005)은 실시간 시스템에 적용하기 위한 통행시간 예측 방법을 비교 평가하였는데, KNN 방법이 실시간 통행시간 예측 방법으로 가장 적합한 것으로 분석되었다. Tam과 Lam (2009)은 기존 실시간 통행시간 예측 방법을 비교 평가한 결과, KNN 방법이 대부분의 평가기준에서 만족할 만한 성능을 보이는 것으로 분석되었다. Lim (2011)은 KNN 방법, 이력자료평균방법, 가중이동평균방법을 이용하여 향후 30분 동안의 교통량 및 속도를 예측한 결과 KNN 방법의 예측 정확도가 가장 우수한 것으로 분석되었다. Lim과 Lee (2011)는 단기 통행시간 예측을 위한 교통류 모형에 입력변수로 활용되는 교통량 예측을 위해 KNN 방법을 적용하였는데 평균절대백분율오차(Mean Absolute Percentage Error; MAPE)가 2~17%로 나타나 예측 성능이 우수한 것으로 분석되었다. 또한 KNN 방법은 결측값 대체

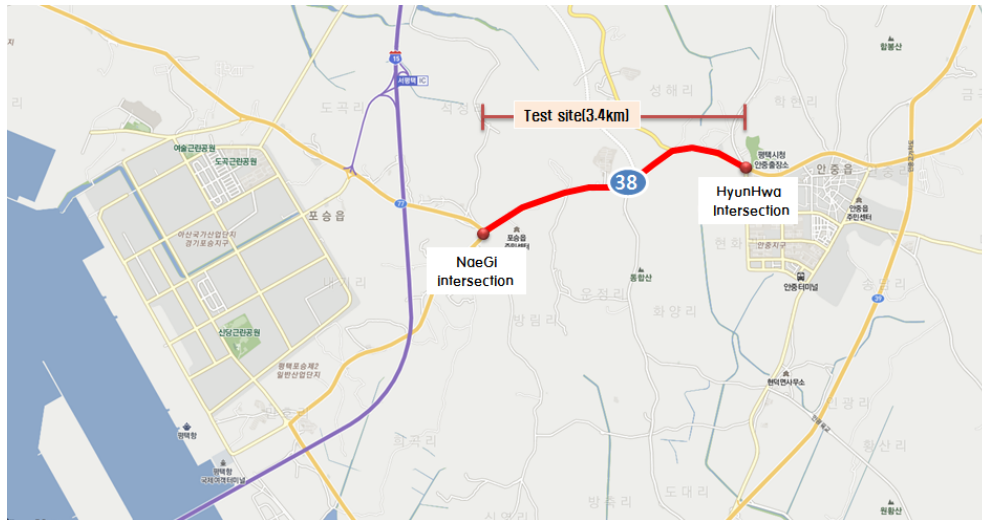


Figure 3.1. Test site

및 상호정보 추정에도 활용되며, 유전자 알고리즘을 이용한 KNN 방법도 연구되어 왔다 (Cha와 Huh, 2008; Park과 Huh, 2010; Park 등, 2011).

KNN 방법은 학습과정이 간단하고 예측 정확도가 높아 분류 및 예측 문제 해결을 위한 방법론으로 널리 이용되고 있으며, 특히 복잡한 비선형 문제에 있어 예측능력이 우수한 것으로 평가 받고 있다. 또한 데이터에 대한 특별한 가정과 모수 추정 과정이 필요 없으며, 연산시간이 짧아 실시간 시스템에 적합한 방법이다. 따라서 본 연구에서는 통행시간 예측 방법으로 KNN 방법을 적용, 평가하였다.

### 3. 일반국도 통행시간 특성 분석

#### 3.1. 자료수집

일반국도를 도로용량편람 기준으로 구분하면 주로 자동차전용도로와 다차로도로, 도시 및 교외간선도로로 구분할 수 있는데, 일반국도의 통행시간 특성을 분석하기 위해 교차로나 유출입 지점이 많은 도시 및 교외 간선도로 유형의 일반국도 구간을 분석하는 것이 타당할 것으로 판단된다. 이에 본 연구에서는 Figure 3.1에서와 같이 일반국도 38호선 중 현화교차로에서 내기삼거리 구간을 대상으로 자료 수집 및 분석을 수행하였다. 분석 대상 구간은 약 3.4km 구간에 차량검지기(Vehicle Detection System; VDS) 1대와 시·중점에 구간검지기인 Dedicated Short Range Communication(DSRC) 장비가 설치되어 있는 구간이다. 자료 수집은 VDS 자료와 DSRC 자료의 속도 자료이며, 2013년 1월 7일~2013년 2월3일까지 4주간의 데이터를 수집하여 분석하였다. 통행시간(속도) 대푯값은 평균값을 적용하였으며, 시간집계간격은 통행시간 추정 및 예측에서 많이 사용하고 있는 5분으로 설정하였다.

#### 3.2. 통행시간 특성 분석

앞서 서론에서도 언급 했듯이 일반국도의 경우 신호교차로 등이 많기 때문에 구간의 시점과 종점 사이에 설치된 지점검지기 속도를 평균한 값과 구간검지기에서 수집된 실제 속도 간에는 차이가 발생하게 된다.

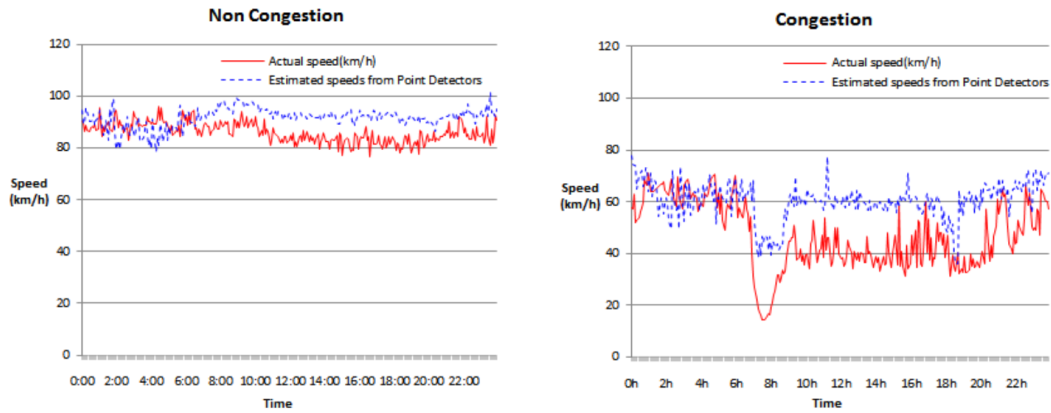


Figure 3.2. Errors of speed caused by point detectors

Table 3.1. Result of the travel time error on departure time based and arrival time based

	Total (%)	Non-congestion (%)	Congestion (%)
Min.	0	0	0
Max.	56.13	45.49	42.09
Average	10.07	9.12	16.09

이는 지점검지기에서 수집되는 공간표본자료를 이용하여 구간 통행속도를 추정할 경우 오차가 발생하기 때문인데, 이를 실제 데이터로 비교 분석하였다. 구간검지기 자료는 동적인 통행시간을 반영한 자료이기 때문에 실제 운전자가 통행한 시간이라고 볼 수 있다. 이에 구간검지기 자료의 출발시각 기준 통행속도를 실제속도로 간주하여 분석하였다.

Figure 3.2와 같이 자료를 분석한 결과 지점검지기 자료를 이용할 경우 전체 시간대에서 평균 약 22km/h 정도 과대 추정되는 것을 알 수 있었으며, 교통정보가 더욱 중요한 정체 시에는 지점검지기 평균속도와 실제속도 간에 약 34km/h의 큰 차이를 보이는 것을 알 수 있었다. 이는 지점검지기 속도를 이용하여 구간 통행시간을 추정하는 데에는 한계가 있음을 의미하는 것이다.

그러나 구간검지기 또한 시간 처짐이라는 문제점을 가지고 있어 이 시간 처짐 현상을 실증적으로 비교하기 위해 일한국도 구간검지기인 DSRC 데이터를 이용하여 통행시간 분포를 비교하였다.

현장에서의 실측자료를 토대로 분석하였을 때, 구간검지기 출발시각기준과 도착시각기준 사이의 시간 처짐 현상은 소통원활 시간대를 보면 비교적 시간 처짐 현상의 폭이 매우 좁아 무시할 정도가 된다. 그러나 정체가 발생하여 대상구간을 통행하는데 소요되는 시간이 커질수록 시간 처짐의 폭이 커지는 것을 알 수 있으며, 혼잡시간대를 기준으로 할 때 통행시간이 급격히 증가함에 따라 비교적 큰 시간 처짐이 발생한다.

2013년 1월 7일부터 2013년 2월 3일까지 자료를 이용하여 출발시각기준과 도착시간기준의 오차를 일별로 통행시간의 최대, 최소, 평균의 평균값을 비교한 결과 Table 3.1과 같이 분석되었으며, 정체시에 오차의 평균값은 비정체시보다 약 7% 더욱 큰 것을 알 수 있다. 또한 최대값과 최소값의 차이는 정체시가 42.09%로 가장 낮은 변화를 보였는데, 이는 통행시간의 이는 출발시각기준과 도착시각기준의 시간 처짐으로 인한 통행시간 오차는 단위 시간당 교통상태(통행시간) 변화량에 비례하며, 구간 검지기 통행시간 만으로도 높은 통행시간 정보 제공이 가능함을 의미한다.

## 4. K-근접이웃 기반 통행시간 예측 방법

### 4.1. K-근접이웃 방법

KNN 방법은 과거 이력자료의 시간적 변화 패턴을 이용하여 미래의 값을 예측하는 방법으로, 과거의 통행시간 이력자료 집합에서 현재 시간대의 통행시간과 거리가 가장 가까운, 즉, 교통상태가 가장 유사한 이웃(Nearest neighbor)을 탐색하여 미래의 통행시간을 예측하는 방법이다. 이 때 거리 측정 지표로는 유클리디안 거리(Euclidean distance)가 이용되며, 이웃의 수는 한 개가 아닌 여러 개의 이웃(예를 들면  $K$ 개)이 이용되는 것이 일반적이다 (Tam과 Lam, 2009). 따라서  $K$ 개의 이웃을 이용하고자 하는 경우, 유클리디안 거리가 짧은 순으로  $K$ 개의 이웃이 결정되며 예측 교통변수는 유클리디안 거리에 따른  $K$ 개 이웃의 가중치를 이용하여 산정되며, 경우에 따라 예측 정확도를 높이기 위해 유클리디안 거리 계산 시 하나의 시간대 자료를 이용하기보다는 일련의 시간대 단위의 자료를 이용하기도 한다.

### 4.2. KNN 방법 적용

KNN 방법은 과거의 교통패턴을 이용하여 미래의 값을 예측하는 방법으로서, 과거 교통패턴 중 현재 교통패턴과 가장 유사한 것을 찾는 것이 핵심이다. 동일한 교통패턴이 일(day) 별로 반복되지 않을 경우 현재의 교통패턴과 가장 유사한 교통패턴은 과거 동일 시간대에서 발생하지 않을 가능성이 크기 때문에, 과거 이력자료에서 이웃을 탐색할 때 동일 시간대 자료를 대상으로 하기보다는 전체 시간대 자료를 대상으로 해서 탐색하는 것이 정확도를 높일 수 있다.

과거 동일 시간대 자료만을 대상으로 해서 현재 교통패턴과 유사한 교통패턴을 찾기 위해서는 매우 많은 양의 과거 자료를 수집하고 처리해야 하는 단점이 있다. 따라서 과거 이력자료에서 이웃을 탐색할 때 전체 시간대 자료를 대상으로 하고, 예측하고자 하는 시간대 기준으로 최근 2주일간 수집된 이력자료를 활용하도록 하였다.

동일한 교통상태라 할지라도 어떤 교통상태를 경유해 왔는지에 따라 다음 시간대의 교통상태가 달리 나타나기 때문에 통행시간을 보다 정확히 예측하기 위해서는 일정 기간 동안의 통행시간 자료를 입력하여 교통상태 변화과정이 반영되도록 할 필요가 있다. 본 연구에서는 본 과업에서는 30분에서 3시간까지 30분 단위(30분, 60분, 90분, 120분, 150분, 180분)로 비교 대상 기간을 변화시켜 가며 방법의 성능을 분석하고자 하였으며, 5분 자료가 이용되므로 비교 대상 자료 수( $n$ )는 6, 12, 18, 24, 36이 된다.

본 연구에서 제안하는 KNN 방법은 다음과 같다.

$$TT(t+1) = \frac{\sum_{k=1}^K [TT^k(h+1) \times w^k(t)]}{\sum_{k=1}^K w^k(t)}$$

$$w^k(t) = \frac{1}{d^k(t)}$$

$$d^k(t) = \left[ \sum_{i=1}^n (TT(t+1-i) - TT^k(h+1-i))^2 \right]^{\frac{1}{2}}, \quad k = 1, 2, \dots, K$$

여기서,  $TT(t+1)$ 는  $t+1$ 시간대의 예측통행시간,  $TT^k(h)$ 는 과거 이력자료 중  $k$ 번째 이웃으로 탐색된  $h$ 시간대의 이력통행시간,  $w^k(t)$ 는  $k$ 번째 이웃에 해당하는 가중치,  $d^k(t)$ 는  $k$ 번째 이웃과의 유클리디안 거리,  $K$ 는 이웃의 수,  $n$ 은 유클리디안 거리 계산을 위한 비교 단위 자료의 수를 나타낸다.

**Table 5.1.** Characteristics of three types of test road

Type	Motorway	Multi-lane Highways	Urban and Sub-urban Arterial
Test site	- National Highway 77 (MunBal IC → SeongDong IC)	- National Highway 3 (GalHyun IC → GalMa Tunnel)	- National Highway 38 (HyunHwa Intersection → NaeGi Intersection)
Length	6.7km	3.2km	3.4km
Road Number	8	4	4
Speed limit	90km/h	80km/h	80km/h
Traffic Quality	Non congestion	Non-Recurrent Congestion	Recurrent Congestion

## 5. 평가 및 검증

### 5.1. 대상구간 선정 및 자료 수집

일반국도는 다양한 도로 특성들이 존재하며 모든 구간에 적합한 방법을 적용하기는 쉽지 않기 때문에 대표구간을 선정하여 분석하였다. 앞서 언급했듯이 통행시간은 연속류나 단속류에 따라 영향을 받으므로 연속류 시설과 단속류 시설로 유형을 구분하고 있는 도로용량편람을 기준으로 자동차전용도로, 다차로도로, 도시 및 교외 간선도로 3가지 유형의 도로를 대상구간으로 선정하였다.

또한 통행시간을 예측하는 것은 정체가 발생하는 구간에 대해 의미가 크기 때문에 도로유형별로 정체가 발생하는 구간을 선정하여 분석하였으며, 분석대상구간 개요는 Table 5.1과 같다. 자동차전용도로의 경우 정체가 발생하는 구간이 없어 정체가 발생하지 않음에도 적용가능성을 살펴보기 위해 분석을 수행하였다.

분석 대상구간의 자료는 구간검지기인 DSRC자료를 이용하였으며, 4주 데이터를 수집하였다. 분석자료 중 앞의 2주 데이터는 학습데이터로 이용하고 뒤의 2주를 검증데이터로 이용하여 분석하였다.

### 5.2. 교통상태에 따른 통행시간 예측 결과

본 연구에서는 KNN 방법을 이용하여 일반국도 통행시간 예측 방법의 적정성을 평가하고자 하였으며, 예측한 통행시간을 출발시각기준 통행시간을 기준값으로 이용하여 평가하였다. 예측 방법은 교통상태가 정체 일 때 의미가 있으므로 평균통행시간을 기준으로 평균×1.5 미만, 평균×1.5 이상, 평균×2.0 이상, 평균×3.0 이상으로 4개의 교통상태를 구분하여 분석하였다. 평균×1.5 미만은 비정체 상태, 평균×3.0 이상은 정체 상태를 의미하며 평균×1.5 이상, 평균×2.0 이상은 지체에서 정체, 정체에서 지체로 교통상태가 변화하는 상태를 의미한다고 볼 수 있다.

평가지표로는 평균절대백분율오차와 변동계수(Coefficient of Variation; CV)를 적용하여 평가하였다. 평균절대백분율오차는 기준값과 예측값의 차이를 기준값에 대한 비율로 나타내는 지표이며, 변동계수는 표준편차를 평균으로 나눈 값으로 값이 클수록 상대적인 차이가 큰 것을 의미한다.

분석 결과, Table 5.2에서와 같이 정체가 발생하지 않는 자동차전용도로는 평균절대백분율오차는 2.814, 변동계수는 0.816으로 분석되었다. 비반복정체가 발생하는 다차로도로는 각각 교통상태별로 6.301, 11.038, 10.878, 21.258의 값을 나타냈으며, 정체가 심해질수록 평균절대백분율오차가 증가하였다. 특히 비정체시보다 정체 시에 약 3배 정도 오차가 증가하였다. 그러나 전체값은 6.602로 평균×1.5 미만인 비정체 상태의 6.301과 크게 차이가 나지 않는데 이는 정체가 발생한 시간대가 적기 때문으로 판단된다. 그러나 반대로 변동계수의 경우 전체적으로 변동은 1.311 정도로 크며, 정체로 갈수록 변동이

**Table 5.2.** Results of travel-time prediction for various road types

Road Classification	Congestion Classification		MAPE(%)	CV
Motorway	All		2.814	0.816
	↑	(Congestion) $\geq$ Average $\times$ 3	N/A	N/A
		$\geq$ Average $\times$ 2	N/A	N/A
		$\geq$ Average $\times$ 1.5	N/A	N/A
	(Non Congestion)	$<$ Average $\times$ 1.5	2.814	0.816
Multilane Highways	All		6.602	1.311
	↑	(Congestion) $\geq$ Average $\times$ 3	21.258	0.512
		$\geq$ Average $\times$ 2	10.878	0.874
		$\geq$ Average $\times$ 1.5	11.038	1.043
	(Non Congestion)	$<$ Average $\times$ 1.5	6.301	0.921
Urban and Sub-urban Arterial	All		11.021	0.909
	↑	(Congestion) $\geq$ Average $\times$ 3	9.830	0.807
		$\geq$ Average $\times$ 2	10.114	0.724
		$\geq$ Average $\times$ 1.5	15.484	0.887
	(Non Congestion)	$<$ Average $\times$ 1.5	10.969	0.860

적은 것으로 분석되었다.

반복정체가 발생하는 도시 및 교외간선도로는 전체와 각 교통상태별로 비슷한 평균절대백분율오차를 보였다. 평균 $\times$ 1.5 이상인 교통상태에서 15.484로 가장 높은 값을 나타냈는데, 이는 비정체에서 정체 또는 정체에서 비정체로 변화하는 과정으로 KNN 방법의 예측력 평균절대백분율오차가 높은 것으로 판단된다. 변동계수 또한 교통상태별로 크게 차이가 없는 것으로 분석되었다.

### 5.3. 결과비교

본 장에서는 KNN 방법의 결과와 타 방법을 통해 예측한 통행시간 결과를 상호 비교평가 함으로써 통행시간 예측방법의 적정성을 평가 하고자 하였다. 본 연구에서는 대표적으로 많이 적용되고 있는 이력자료평균 방법과 칼만필터 방법을 선정하여 비교평가 하였다.

이력자료평균 방법은 과거 많은 연구에서 폭 넓게 적용되고 있으며, 적용이 용이할 뿐 아니라 실시간 자료를 다루는 국내외 ITS 분야에서 자료 예측 또는 결측자료 처리 기법으로 자주 이용되고 있으며, 본 연구에서 제안하는 이력자료 평균 방법은 다음과 같다.

$$TT(t+1) = \frac{TT^1(t+1) + TT^2(t+1) + \dots + TT^n(t+1)}{n}$$

여기서,  $TT(t+1)$ 는  $t+1$  시간대의 예측통행시간,  $TT^n(t+1)$ 는 최근  $n$ 주 전의  $t+1$  시간대 이력통행시간을 나타낸다.

칼만필터 방법은 새로운 관측치를 즉시 수용할 수 있고 참값에 빠르게 수렴시킬 수 있는 장점이 있어 온라인 시스템에서의 적용이 유리하며, 국내외 ITS 센터에서 많이 활용되고 있다. 본 연구에서는 Jang 등 (2004)의 연구에서 적용된 칼만필터 방법을 참고하여 통행시간 예측 방법을 구성하도록 하였다.

예측 방법별로 MAPE 분석결과, 자동차전용도로는 이력자료평균 방법이 1.563, 칼만필터 방법이 3.207, KNN 방법이 2.814로 비정체 구간에서는 방법별로 크게 차이가 없는 것으로 분석되었다. 비반복 정체인 다차로도도에서는 모든 교통상태에서 KNN 방법이 우수한 것으로 분석되었으며, 정체시에는 이력자료평균 63.863인데 반해 KNN은 21.258로 KNN 방법이 약 3배정도 오차가 적은 것으로 분

**Table 5.3.** Evaluation of the travel-time prediction model based on MAPE

Road Classification	Congest Classification		HDA	KF	KNN
Motorway	All		1.563	3.207	2.814
	(Congestion)	$\geq \text{Average} \times 3$	-	-	-
	↑	$\geq \text{Average} \times 2$	-	-	-
		$\geq \text{Average} \times 1.5$	-	-	-
	(Non Congestion)	$< \text{Average} \times 1.5$	1.563	3.207	2.814
Multilane Highways	All		10.682	8.052	6.602
	(Congestion)	$\geq \text{Average} \times 3$	63.863	26.969	21.258
	↑	$\geq \text{Average} \times 2$	28.686	14.698	10.878
		$\geq \text{Average} \times 1.5$	27.016	14.717	11.038
	(Non Congestion)	$< \text{Average} \times 1.5$	9.526	7.605	6.301
Urban and Sub-urban Arterial	All		12.218	12.751	11.021
	(Congestion)	$\geq \text{Average} \times 3$	12.731	22.319	9.830
	↑	$\geq \text{Average} \times 2$	21.189	21.599	10.114
		$\geq \text{Average} \times 1.5$	21.896	24.573	15.484
	(Non Congestion)	$< \text{Average} \times 1.5$	11.950	12.434	10.969

**Table 5.4.** Evaluation of the travel-time prediction model based on CV

Road Classification	Congest Classification		HDA	KF	KNN
Motorway	All		1.572	0.847	0.816
	(Congestion)	$\geq \text{Average} \times 3$	-	-	-
	↑	$\geq \text{Average} \times 2$	-	-	-
		$\geq \text{Average} \times 1.5$	-	-	-
	(Non Congestion)	$< \text{Average} \times 1.5$	1.572	0.847	0.816
Multilane Highways	All		1.734	1.401	1.311
	(Congestion)	$\geq \text{Average} \times 3$	0.153	0.526	0.512
	↑	$\geq \text{Average} \times 2$	0.776	0.859	0.874
		$\geq \text{Average} \times 1.5$	0.502	0.919	1.043
	(Non Congestion)	$< \text{Average} \times 1.5$	1.286	1.047	0.921
Urban and Sub-urban Arterial	All		1.073	1.037	0.909
	(Congestion)	$\geq \text{Average} \times 3$	0.633	0.442	0.807
	↑	$\geq \text{Average} \times 2$	0.763	0.603	0.724
		$\geq \text{Average} \times 1.5$	0.815	0.609	0.887
	(Non Congestion)	$< \text{Average} \times 1.5$	0.942	0.927	0.860

석되었다. 반복정체구간인 도시 및 교외 간선도로 또한 모든 교통상태에서 KNN 방법이 우수한 것으로 분석되었다.

CV값 분석결과 자동차전용도로는 칼만필터와 KNN에서 각각 0.847과 0.816 값으로 크게 차이가 없는 것으로 분석되었는데, 이력자료평균은 1.572로 칼만필터와 KNN 방법보다 변동이 큰 것으로 분석되었다. 다차로도로는 평균 $\times 1.5$  이상부터 정체시까지 이력자료평균 방법이 변동이 적은 것으로 분석되었으나, 비정체시와 전체시간대로 비교했을 때는 0.921과 1.311로 KNN 방법이 우수한 것으로 분석되었다. 마지막으로 도시 및 교외간선도로는 지.정체가 발생할 때는 칼만필터 방법이 우수한 것으로 분석되었으나, 전체시간대로 비교했을 때는 이력자료평균방법이 1.073, 칼만필터 방법이 1.037, KNN 방법이 0.909로 KNN 방법이 우수한 것으로 분석되었다.



## 6. 결론

본 연구는 일반국도 실시간 시스템에 적합한 통행시간 예측 방법을 개발하는 것이 목적으로 KNN 방법을 개발하였으며, 일반국도의 실제데이터를 이용하여 통행시간 예측하여 분석하였다. 개발된 KNN 예측 방법의 적정성을 평가 하고자 대표적으로 적용되고 있는 이력자료평균 방법과 칼만필터 방법을 비교 평가 하였다.

분석 결과, MAPE 분석에서는 모든 도로유형, 교통상태에서 KNN 방법이 우수한 것으로 분석되었으며, CV 분석결과에서는 도로유형별, 교통상태별로 다른 결과를 보였다. 특히 정체가 발생하는 다차로 도로와 도시 및 교외 간선도로에서는 정체시에 각각 이력자료평균과 칼만필터 방법이 변동이 적은 것으로 분석되었다. 그러나 전체시간대를 비교했을 때는 모든 도로유형에서 KNN 방법이 우수한 것으로 분석되어, KNN 방법이 일반국도의 통행시간을 예측 방법으로 적합한 것으로 판단된다. 본 연구를 통해 개발된 일반국도 통행시간 예측 방법을 적용하여 통행시간 정보 제공 시 기존 통행시간과 예측 통행시간을 탄력적으로 적용함으로써 통행시간 정보의 신뢰성을 향상시킬 수 있을 것으로 판단된다.

그러나 KNN 방법은 패턴인식 기반 방법의 일종으로 과거 이력자료와 미래에 관측될 자료의 패턴이 얼마나 유사한지에 따라 방법의 성능이 좌우된다. 따라서 돌발 상황 발생 시에는 과거 이력자료에서 유사한 패턴을 찾기가 어려워 방법의 성능이 떨어질 가능성이 크다. 앞서 분석한 다차로도로구간의 경우 비 반복정체가 발생하는 구간으로 정체 시에 반복정체구간인 도시 및 교외간선도로보다 예측력이 떨어지는 것으로 분석되었다. 따라서 향후 통행시간 예측 방법에 돌발 상황 감지 알고리즘이 추가로 반영될 경우 통행시간 예측 성능을 개선시킬 수 있을 것으로 판단된다.

## References

- Cha, W. O. and Huh, M. Y. (2008). k-nearest neighbor-based approach for the estimation of mutual information, *Communications for Statistical Applications and Methods*, **15**, 17–26.
- Jang, J., Beak, N., Kim, S. and Byun, S. (2004). Dynamic travel time prediction using AVI data, *Journal of Korean Society of Transportation*, **22**, 169–175.
- Kim, J., Rho, J., Park, D. and Namkoong, S. (2006). Comparative analysis of link travel times: Departure time based vs arrival time based, *Journal of Korea Spatial Planning Review*, **48**, 71–86.
- Lam, W. H. K., Tang, Y. F., Chan, K. S. and Tam, M. L. (2006a). Short-term hourly traffic forecasts using Hong Kong annual *Traffic Census*, *Transportation*, **33**, 291–310.
- Lam, W. H. K., Tang, Y. F. and Tam, M. L. (2006b). Comparison of two non-parametric models for daily traffic forecasting in Hong Kong, *Journal of Forecasting*, **25**, 173–192.
- Lee, Y. and Lee, J. (2002). A study on link travel time estimation methodology for traffic information service (Determination of an adequate sample size), *Journal of Korean Society of Transportation*, **20**, 55–67.
- Lim, S. (2011). *Travel Time Prediction Simultaneously using Point and Interval Detector Date*, A dissertation submitted for the degree of Doctor of Philosophy, The University of Seoul.
- Lim, S. and Lee, C. (2011). Data fusion algorithm improves travel time predictions, *IET Intelligent Transport Systems*, **5**, 302–309.
- Ministry of Land, Infrastructure and Transport (2005). Highway Capacity Manual.
- Park, C. S. and Huh, K. (2010). Optimal k-nearest neighborhood classifier using genetic algorithm, *Communications for Statistical Applications and Methods*, **17**, 17–27.
- Park, S.-H., Bang, S.-W. and Jhun, M.-S. (2011). On the use of sequential adaptive nearest neighbors for missing value imputation, *The Korean Journal of Applied Statistics*, **24**, 1249–1257.
- Smith, B. L. and Demetsky, M. J. (1997). Traffic flow forecasting: Comparison of modeling approaches, *Journal of Transportation Engineering ASCE*, **123**, 261–266.

- Tam, M. L. and Lam, W. H. K. (2009). *Short-term Travel Time Prediction for Congested Urban Road Networks*, Transportation Research Board 88th Annual Meeting 2009 Paper 9-2313.
- You, J. and Kim, T. J. (2005). *Towards Developing a Travel Time Forecasting Model for Location-based Services: A Review*, In Reggiani, A., and L. A. Schintler, Eds. *Methods and Models in Transport and Telecommunications*, Springer, Berlin, Heidelberg, 45-61.

# K 최대근접이웃 방법을 이용한 통행시간 예측에 대한 연구

임성한<sup>a</sup> · 이향미<sup>a</sup> · 박성룡<sup>b</sup> · 허태영<sup>c,1</sup>

<sup>a</sup>한국건설기술연구원 도로교통연구실, <sup>b</sup>국토교통부 첨단도로환경과

<sup>c</sup>충북대학교 자연과학대학 정보통계학과

(2013년 9월 6일 접수, 2013년 10월 10일 수정, 2013년 10월 15일 채택)

---

## 요약

통행시간은 교통정보 중에서 가장 대표적이고 이용자 선호도가 높은 정보이다. 본 연구에서는 일반국도를 대상으로 실시간 시스템에 적용 가능한 통행시간 예측 방법을 개발하고자 하였다. 통행시간 예측방법으로 비모수적 접근 방법인 K 최대근접이웃 방법을 적용하였다. K 최대근접이웃 방법은 데이터에 대한 특별한 가정이 필요 없고, 모수 추정 과정이 필요 없어 실시간 교통관리시스템에 적합하다. K 최대근접이웃 방법의 우수성을 평가하기 위해 교통 분야에서 많이 적용되고 있는 이력자료 평균방법과 칼만 필터방법을 선정하여 평균절대백분율오차와 변동계수를 통해 평가하였다. 평가 결과 K 최대근접이웃 방법이 이력자료 평균방법과 칼만 필터방법에 비해 우수한 것으로 분석되었다. 통행시간 정보 제공 시 본 연구에서 개발된 방법을 통해 도출된 통행시간과 구간검지기로부터 관측된 통행시간을 탄력적으로 적용함으로써 통행시간 정보의 신뢰도를 향상시킬 수 있을 것으로 기대된다.

주요용어: 통행시간 예측, 비모수적 방법, K-최대근접이웃, 지능형교통체계.

---

<sup>1</sup>교신저자: (361-763) 충북 청주시 흥덕구 개신동, 충북대학교 정보통계학과, 교수. E-mail: theo@cbnu.ac.kr