

KNN 알고리즘을 활용한 고속도로 통행시간 예측

신강원* · 심상우** · 최기주*** · 김수희****

Shin, Kangwon*, Shim, Sangwoo**, Choi, Keechoo***, Kim, Soohee****

Expressway Travel Time Prediction Using K-Nearest Neighborhood

ABSTRACT

There are various methodologies to forecast the travel time using real-time data but the K-nearest neighborhood (KNN) method in general is regarded as the most one in forecasting when there are enough historical data. The objective of this study is to evaluate applicability of KNN method. In this study, real-time and historical data of toll collection system (TCS) traffic flow and the dedicated short range communication (DSRC) link travel time, and the historical path travel time data are used as input data for KNN approach. The proposed method investigates the path travel time which is the nearest to TCS traffic flow and DSRC link travel time from real-time and historical data, then it calculates the predicted path travel time using weight average method. The results show that accuracy increased when weighted value of DSRC link travel time increases. Moreover the trend of forecasted and real travel times are similar. In addition, the error in forecasted travel time could be further reduced when more historical data could be available in the future database.

Key words : K-Nearest neighborhood, TCS traffic flow, DSRC link travel time, Weight average method, Path travel time prediction

초록

실시간 자료를 반영한 통행시간 예측 기법은 다양하지만 관련 연구 검토 결과 과거이력데이터가 충분하다면 타 모형에 비해 K 최대근접이웃(K-Nearest Neighbors)의 정확도가 우수하므로 본 연구에서는 이에 대한 적용 방법 도출 및 가능성 평가를 목적으로 한다. 본 연구에서는 KNN의 입력 자료로 TCS 교통량 및 DSRC 구간통행시간의 실시간 및 과거 이력자료, 경로통행시간 이력자료를 활용하였다. 통행시간 예측치는 TCS 교통량 및 DSRC 구간통행시간의 실시간 자료와 유사한 경로통행시간을 탐색한 후 이를 가중평균하여 산출하였다. 예측 기법을 적용한 결과 DSRC 구간통행시간의 가중치가 증가할수록 정확도는 증가하였으며, 이는 실시간 교통상황 변화를 DSRC 구간통행시간이 잘 반영하기 때문이다. 그러나 TCS 교통량을 기반으로 한 경우 역시 정확도의 차이가 크지 않으며, 변화 추이도 유사하게 나타났다. 이러한 결과를 볼 때 향후 대용량의 과거이력자료가 축적될 경우 예측오차는 더욱 감소될 것으로 기대된다.

검색어 : K 최대근접이웃, TCS 교통량, DSRC 구간통행시간, 가중평균, 경로통행시간 예측

1. 서론

교통수요 증가에 따라 도로를 신설·확장하는 공급 확대 방식의 한계로 인해 국내에서는 1990년대부터 기존 도로시설의 운영효율을 극대화하기 위해 첨단기술을 활용한 지능형교통체계 도입과 더불어 실시간 교통정보를 제공하기 시작하였다. 한국도로공사는 1993년

* 정희원 · 경성대학교 도시공학과 조교수 (Kyungseung University · kangwon@ks.ac.kr)

** 정희원 · 교신저자 · 아주대학교 TOD기반 지속가능 도시·교통 연구센터 연구교수 (Corresponding Author · Ajou University · artmania@ajou.ac.kr)

*** 정희원 · 아주대학교 교통시스템공학과 교수 (Ajou University · keechoo@ajou.ac.kr)

**** 한국도로공사 도로교통연구원 선임연구원 (Expressway & Transportation Research Institute · sooheek@ex.co.kr)

Received July 29, 2014/ revised August 25, 2014/ accepted October 14, 2014

고속도로교통관리시스템 구축 이후 실시간 교통정보를 제공하기 시작하였고, 실시간 교통정보의 신뢰성을 향상시키기 위해 2012년 단거리전용통신(Dedicated Short Range Communication, 이하 DSRC)을 활용한 하이패스 교통정보시스템을 구축하였다. 그러나 실시간 교통정보는 제공 시점의 구간별 혼잡상황에 대한 단편적인 정보이기 때문에 시공간적 혼잡상황의 변화에 따라 이용 시점에 제공한 통행시간과 실제 경험한 통행시간의 차이가 발생하게 된다. 이로 인해 고속도로 교통정보에 대한 이용자들의 만족도 및 신뢰도 역시 저하되고 있는 상황이므로 예측정보 제공에 대한 필요성이 증대되고 있다.

한국도로공사는 예측정보의 기반이 되는 이력자료를 저장하기 위해 OASIS (Operation Analysis & Supportive Information System)라는 이력자료시스템을 2008년부터 구축하여 운영하고 있으며, OASIS의 TCS (Toll Collection System) 자료를 활용하여 주요 도시(서울, 부산, 울산, 대구, 대전, 광주, 목포, 강릉)에 대한 출발시간 기준 통행시간을 1시간 단위로 예측하여 제공하는 교통예보지원시스템을 운영하고 있다. 그러나 교통예보지원시스템에서는 시계열 모형인 지수평활화와 ARIMA (Autoregressive integrated moving average)를 이용하기 때문에 실시간 자료를 반영하지 못하고 있으며, 그 결과 통행패턴이 과거와 상이할 경우 정확도가 떨어지는 문제점이 발생하고 있다.

이에 본 연구에서는 실시간 자료를 활용한 다양한 연구 사례를 검토하고, TCS와 DSRC 자료를 기반으로 실시간 자료를 활용한 고속도로 통행시간 예측의 적절한 방법론을 도출하는 것을 목적으로 하며, 그 방법은 다음과 같다.

- 1 단계: 관련 연구 사례 검토
- 2 단계: 서울TG~대전IC 구간의 TCS와 DSRC 자료를 이용하여, KNN (K-Nearest Neighborhood) 모형 구축
- 3 단계: 예측 모형 정확도 검증
- 4 단계: 결론 및 향후 연구 과제 제시

2. 관련 연구 사례

2.1 관련 연구 검토

실시간 자료를 이용한 통행시간 예측 기법은 신경망, 회귀모형, 칼만필터링, KNN 등 다양한 연구가 수행되었다.

Kang and Namkoong (2002)은 TCS 자료 기반의 모듈라 신경망 모형, Lee and Kim (2002)은 지점검지기의 교통량과 TCS 자료를 입력 자료로 하는 인공신경망, Wei and Lee (2007)은 GPS 프로브 및 지점검지기, 유고자료 등을 통합한 인공신경망, Kim and Jang (2013)은 고속도로 DSRC 자료 기반의 인공신경망

을 이용하여 고속도로 통행시간을 예측하였다. Xie et al. (2004)와 Innamaa (2005)는 간선도로와 도시부 고속도로의 통행시간 예측을 위해 다층신경망을 적용하였으며, Jeong and Rilett (2004)는 인공신경망을 이용하여 버스 도착시간을 예측하였고, Lee (2009)는 클러스터링 방법과 인공신경망을 결합하여 버스의 실시간 통행시간을 예측하였다.

Zhang and Rice (2003)과 Rice and Zwet (2004)는 현재 통행시간과 장래 통행시간 사이의 선형관계를 기반으로 한 회귀식을 통해 통행시간을 예측하였고, Yu (2008)은 VDS (Vehicle Detection System)와 AVI (Automatic Vehicle Identification) 이력자료 기반의 다항회귀식을 통해 통행시간을 예측하였다. Ban et al. (2010)은 이력자료의 순간통행시간 합과 현재의 순간통행시간 합의 차이에 대한 파라미터를 회귀식으로 추정하고 이를 반영하여 경로통행시간을 예측하였다.

Lee et al. (2002)는 프로브차량 기반의 자료를 기반으로 한 칼만필터링을 통해 링크통행시간을 예측하였고, 고속도로 및 일반국도의 통행시간을 예측하는데 Oh et al. (2003)은 VDS 자료 기반, Jang et al. (2004)는 AVI 자료 기반의 칼만필터링을 활용하였다. Lee et al. (2005)는 칼만필터링과 데이터 융합기술을 이용한 다주기 통행시간 예측 모형을 개발하였다. Chen et al. (2012)는 SVM (Support Vector Machine)과 칼만필터링을 이용하여 BRT (Bus Rapid Transit) 차량의 통행시간을 예측하였다.

Smith and Demetsky (1997), Lam et al. (2006), Lim and Lee (2011)는 교통량 예측을 위해 KNN 방법을 이용하였으며, You and Kim (2000)은 KNN 방법에 GIS (Geographic Information Systems)와 Machine Learning 기법을 적용하여 시가지도로의 통행시간을 예측하였다. Park et al. (2006)와 Lim et al. (2013)은 KNN 방법을 이용하여 버스의 통행시간과 일반국도의 통행시간을 예측하였다.

이외에 Mehmet and Nikolas (2013)은 병목구간 확인, 교통패턴에 따른 자료 분류, 통계적 혼잡맵을 기반으로 이력자료와 실시간 자료의 융합을 통해 경험적 통행시간을 예측하였다.

2.2 시사점 도출

통행시간 예측과 관련하여 다양한 연구가 진행되었으나 신경망의 경우 학습 과정이 매우 복잡하다는 단점이 있으며, 회귀모형은 다수의 모형(예: 기/종점, 경로 등)을 개발하여야 한다는 문제점이 있다. 반면 KNN의 경우 이러한 문제점을 해결할 수 있으며, 참조할 수 있는 데이터가 충분하다면 타 모형에 비해 정확도가 우수한 장점이 있다(Smith et al., 2002; Nigovski et al., 2005). 또한 한국도로공사는 6년 정도의 고속도로 이력자료를 OASIS에서 보유하고 있어 KNN 방법을 적용하는데 적합한 환경을 보유하고 있다.

기존 연구에서는 실시간 및 이력통행시간을 활용하여 KNN을 적용하였는데, 고속도로의 경우 DSRC 구간통행시간과 더불어 TCS를 통해 진출입 교통량을 수집할 수 있으므로 이러한 부분을 고려할 필요가 있다. 이에 본 연구에서는 실제 경로통행시간을 기반으로 한 TCS 교통량과 DSRC 구간통행시간 자료를 기반으로 KNN을 이용한 통행시간 예측 방법론을 도출하고 적용가능성을 평가한다.

3. KNN 기반의 통행시간 예측

3.1 KNN 기본 구조 및 입력 자료

KNN 모형은 실시간으로 수집되는 입력변수의 상태와 가장 유사한 과거 이력자료 내 동일 변수의 상태를 실시간으로 비교하여 유사한 이웃(근접이웃)의 경로통행시간을 가중평균하여 경로통행시간을 추정하는 구조를 가지고 있다. 이러한 모형의 구조를 고려하여 본 연구에서는 실시간 소통상태 매트릭스, 과거 소통상태 매트릭스, 과거 소통상태별 경로통행시간 매트릭스가 기본 자료로 이용하

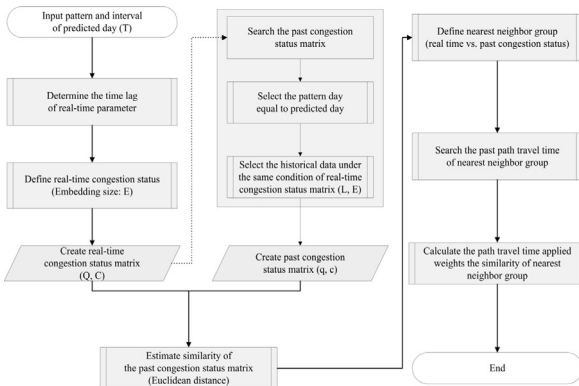


Fig. 1. Flow Chart of KNN Method

Date	Day	Holiday	Interval	Seoul	Suwon	Giheung	Osan
110108	6	1	33	160	44	4	26
110108	6	1	34	136	54	3	15
110108	6	1	35	119	42	1	12
110108	6	1	36	119	58	7	10
110108	6	1	37	139	33	4	16
110108	6	1	38	150	45	5	18
110108	6	1	39	143	35	7	18
110108	6	1	40	124	33	5	11
110108	6	1	41	111	38	7	17
110108	6	1	42	119	38	12	12
110108	6	1	43	129	40	3	20
110108	6	1	44	131	54	7	11
110108	6	1	45	144	31	3	20
110108	6	1	46	143	34	2	16
110108	6	1	47	99	49	10	12
110108	6	1	48	97	40	4	19
110108	6	1	49	117	42	7	13
110108	6	1	50	119	52	8	17
110108	6	1	51	104	50	2	27
110108	6	1	52	109	45	5	17

Fig. 2. Real-time TCS Flow Status Matrix and Embedding Size

며, 실시간 자료와 과거 이력자료의 유사성은 유클리디안 거리를 이용하여 산출하고, 그 흐름도는 Fig. 1과 같다.

3.2 시계열자료 및 상태매트릭스 정의

실시간 교통상황은 각 시간대별 TCS 교통량과 DSRC 구간통행시간의 시계열자료로 구성된 상태매트릭스로 정의되며, 매트릭스의 열의 수는 경로의 요금소 또는 DSRC 구간의 수와 같고, 행의 수는 포함크기(Embedding size, E)와 같다. 예를 들면 현시점 주기 T의 예측정보를 얻기 위해 포함크기 E의 값을 '10'으로 설정할 경우 현시점을 기준으로 이전 10개 주기의 자료가 상태매트릭스에 포함되며, TCS 교통량을 기반으로 할 경우는 Fig. 2와 같다.

3.3 유사성 산출 및 K 근접이웃 결정

실시간 상태매트릭스 작성이 완료되면, 기 정의한 실시간 상태매트릭스와 유사한 과거 상태매트릭스를 탐색하기 위해 후보군을 탐색하게 된다. 탐색 후보군은 예측일과 동일한 요일의 과거이력자료를 선택한 후 선택된 날짜에 수집된 자료 중 실시간 상태매트릭스와 동일 주기를 갖는 과거 상태매트릭스를 일자별로 탐색하여 데이터베이스화하게 된다. 데이터베이스화된 과거 교통상황 매트릭스와 실시간 교통상황 매트릭스의 유사성은 유클리디안 거리로 결정하였으며, TCS 교통량 및 DSRC 통행시간에 대한 산출식은 Eqs. (1) and (2)와 같다.

$$qS_i^d = \left(\sum_{m=1}^M \sum_{j=1}^J \{qS_E^d(i-m, J) - qS_E^r(i-m, J)\}^2 + \sum_{m=1}^M \sum_{j=1}^J \{qS_O^d(i-m, J) - qS_O^r(i-m, J)\}^2 \right)^{1/2} \quad (1)$$

여기서, qS_i^d : 이력자료 중 d번째 날의 i번째 주기의 TCS 교통량의 유사성 ($d = 1, 2, \dots, D$)

$qE^d(i, j)$: d번째 날의 i번째 주기에 j번째 TG 진입 이력교통량 (대/5분)

$qE^r(i, j)$: i번째 주기에 j번째 TG 진입 실시간 교통량 (대/5분)

$qO^d(i, j)$: d번째 날의 i번째 주기에 j번째 TG 진출 이력교통량 (대/5분)

$qO^r(i, j)$: i번째 주기에 j번째 TG 진출 실시간 교통량 (대/5분)

m : 유사성 분석을 위한 포함 주기수

$$cS_i^d = \left(\sum_{m=1}^M \sum_{j=1}^J \{c^d(i-m, J) - c^r(i-m, J)\}^2 \right)^{1/2} \quad (2)$$

- 여기서, cS_i^d : 입력자료 중 d번째 날의 i번째 주기의 DSRC 통행시간의 유사성 ($d=1,2,\dots,D$)
- $c^d(i,j)$: d번째 날의 i번째 주기에 j번째 구간의 입력 구간통행시간 (분)
- $c^r(i,j)$: i번째 주기에 j번째 링크의 실시간 구간통행시간 (분)
- m : 유사성 분석을 위한 포함 주기수

여러 과거 상태매트릭스 중 실시간 상태매트릭스와 가장 유사한 일자의 과거 상태매트릭스를 근접이웃으로 결정해야 한다. 이에 본 연구에서는 최근접 이웃의 유사성(최소 유클리디안 거리) 대비 5% 이내의 유클리디안 거리를 갖는 이웃만을 근접이웃으로 결정하였으며, 이를 K개의 근접이웃으로 결정하였다.

3.4 KNN 기반 통행시간 예측

K개의 근접이웃을 통해 2개(TCS 교통량, DSRC 구간통행시간)의 T주기 경로통행시간이 산출되므로 본 연구에서는 이들을 가중평균하여 최종 경로통행시간의 예측치를 추정하였다. 이 때 TCS 교통량 및 DSRC 구간통행시간의 근접이웃에 대한 경로통행시간은 유사성의 역수를 가중치로 적용하였으며, 산출식은 Eq. (3)과 같다. 유사성의 역수를 가중치로 취한 이유는 유사성이 높을수록 유클리디안 거리가 짧으므로 유사성의 역수를 가중치로 하면 유사성이 높을수록 가중치가 증가하기 때문이다.

$$\hat{P}_i = w \times \frac{\sum_{k=1}^K \left(\frac{1}{qS_i^k}\right) p_k^q}{\sum_{k=1}^K \frac{1}{qS_i^k}} + (1-w) \times \frac{\sum_{k=1}^K \left(\frac{1}{cS_i^k}\right) p_k^c}{\sum_{k=1}^K \frac{1}{cS_i^k}} \quad (3)$$

- 여기서, \hat{P}_i : i번째 주기의 경로통행시간 예측치
- w : 경로통행시간 예측치에 대한 TCS 교통량 가중치
- qS_i^k : TCS 교통량 k 근접이웃의 유사성
- p_k^q : TCS 교통량 기반 k 근접이웃의 경로통행시간
- cS_i^k : DSRC 구간통행시간 k 근접이웃의 유사성
- p_k^c : DSRC 구간통행시간 k 근접이웃의 경로통행시간
- k : 근접이웃의 수($k=1,2,\dots,K$)

4. KNN 기반의 통행시간 예측 평가

경부고속도로 서울~대전 구간을 대상으로 KNN 기반의 예측 모형을 평가하였다. 평가 지표는 RMSE (Root Mean Square Error)를 이용하였으며, 산출식은 Eq. (4)와 같다.

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

입력 자료는 경부선 하행 방면 서울TG~오산IC, 서울TG~청주IC, 서울TG~대전IC의 2011년 1월 1일~12월 31일 TCS 진출입 교통량(대/5분)과 DSRC 링크통행시간(분/5분)을 사용하였으며, 예측은 2011년 1월 1일~6월 30일에 대하여 수행하였다. 근접이웃 탐색을 위한 시간대(E)는 18로 예측목표시점 기준 과거 1시간 30분 동안의 상태매트릭스를 활용하였고, TCS 교통량과 DSRC 통행시간에 대한 경로통행시간 가중치에 따른 영향을 파악하기 위해 본 연구에서는 TCS 교통량에 대하여 3가지 가중치(1.0, 0.8, 0)를 적용하여 예측을 수행하였다. 또한 요일별 통행패턴은 월, 화~목, 금, 토, 일요일 5개 그룹으로 분류하여 분석하였다.

Table 1. Result by Applied KNN Method on Seoul TG ~ Osan IC

Day	RMSE (Min)		
	w=1.0	w=0.8	w=0
Mon	3.6	3.5	2.3
The, Wed, Thu	2.5	2.8	2.0
Fri	2.6	2.5	1.7
Sat	3.5	4.5	2.1
Sun	2.4	2.3	1.6
Average	2.9	3.1	1.9

Table 2. Result by Applied KNN Method on Seoul TG ~ Cheongju IC

Day	RMSE (Min)		
	w=1.0	w=0.8	w=0
Mon	5.5	5.6	3.8
The, Wed, Thu	5.1	5.6	4.4
Fri	3.9	3.8	2.9
Sat	8.3	10.6	6.3
Sun	5.8	5.7	3.9
Average	5.7	6.3	4.3

Table 3. Result by Applied KNN Method on Seoul TG ~ Daejeon IC

Day	RMSE (Min)		
	w=1.0	w=0.8	w=0
Mon	5.7	5.7	4.1
The, Wed, Thu	5.6	5.5	4.8
Fri	4.2	3.8	3.1
Sat	8.9	11.4	6.9
Sun	6.2	6.0	4.2
Average	6.1	6.5	4.6

TCS 교통량 및 링크통행시간을 이용한 예측결과는 Tables 1~3과 같다. 서울TG~대전IC의 경우 TCS 교통량의 가중치가 0.8인 경우를 보면 주중에는 월요일의 RMSE는 5.7분으로 가장 크게 나타났으며, 주말에는 토요일의 예측오차는 11.4분으로 일요일(6.0분)에 비해 크게 나타났다. 이와 같이 일요일의 예측 오차가 토요일에 비해 작게 나타난 것은 하행 방향의 여가 목적 통행이 많지 않아 주중의 교통량과 큰 차이가 없기 때문에 나타난 것으로 판단된다. 그러나 토요일의 예측 오차로 인해 주중(평균 5.0분)보다는 주말(평균 8.7분)의 예측 오차가 크게 나타났다. 또한 토요일을 기준으로 TCS 교통량의 가중치가 0.8인 경우를 보면 단거리 구간인 서울TG~오산IC 구간의 RMSE는 4.5분, 중단거리 구간인 서울TG~청주IC 구간은 10.6분, 장거리 구간인 서울TG~대전IC 구간은 11.4분으로 구간길이가 증가할수록 예측오차도 증가하는 것으로 나타났다.

또한 일자별로 살펴본 결과 유클리디안 거리가 짧을수록 예측오차는 감소하는 것으로 나타났으며, 그 결과의 예는 Figs. 3 and

4와 같다. 이를 보면 어느 정도의 예측 오차를 포함하고 있으나, 대부분 실제 통행시간의 변화 추이와 유사하게 나타났다.

가중치에 따른 영향을 보면 TCS 교통량과 DSRC 구간통행시간의 근접이웃을 가중평균한 경우가 각각의 근접이웃만을 사용한 경우보다 오차가 크게 나타났으며, TCS 교통량만을 사용하는 경우보다 DSRC 구간통행시간만을 사용한 경우의 RMSE가 더 작게 나타났다. 이는 DSRC 구간통행시간의 실시간 자료는 혼잡의 증감과 동일하게 이력자료의 구간통행시간 및 경로통행시간 증감이 이루어지는 반면 실시간 TCS 교통량은 교통량의 증감과 경로통행시간의 증감이 동일하게 변화하지 않기 때문인 것으로 판단된다. 왜냐하면 TCS 교통량은 상·하행으로 분류되어 있지 않기 때문에 진출입 교통량이 어느 방향으로 통행할지 알 수 없을 뿐만 아니라 Figs. 5 and 6과 같이 하나의 TCS 진출입 교통량이 2개의 경로통행 속도(정상류와 혼잡류)에 매칭되므로 교통량이 유사해도 경로통행시간은 차이가 클 수 있기 때문이다.

2011년 1월~6월까지 KNN 기반의 예측 모형과 현재 한국도로

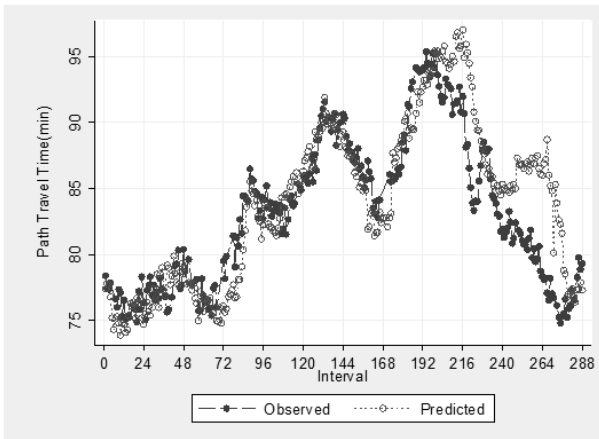


Fig. 3. Predicted Result on Seoul TG~Daejeon IC (Jan 14th 2011)

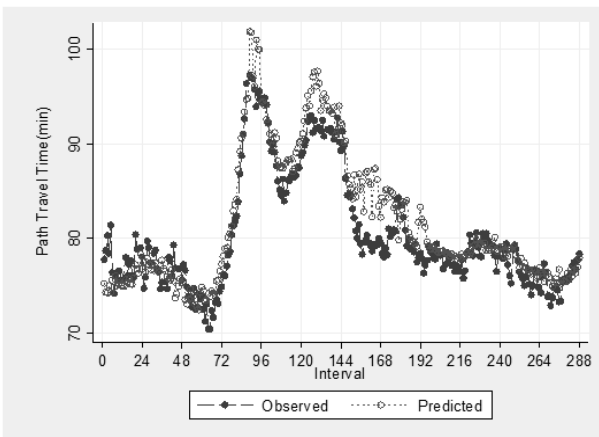


Fig. 4. Predicted Result on Seoul TG~Daejeon IC (Mar 28th 2011)

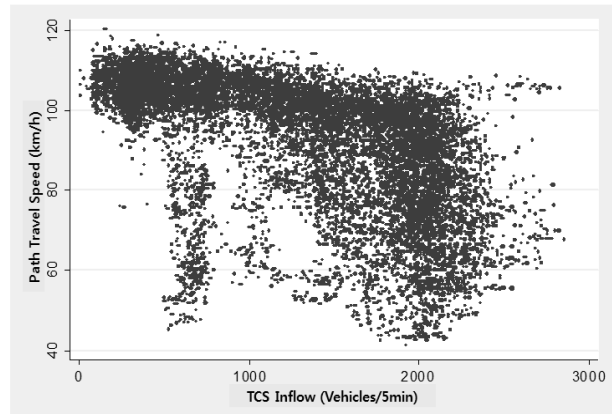


Fig. 5. Relation of TCS Inflow and Path Travel Speed (Seoul TG~Daejeon IC)

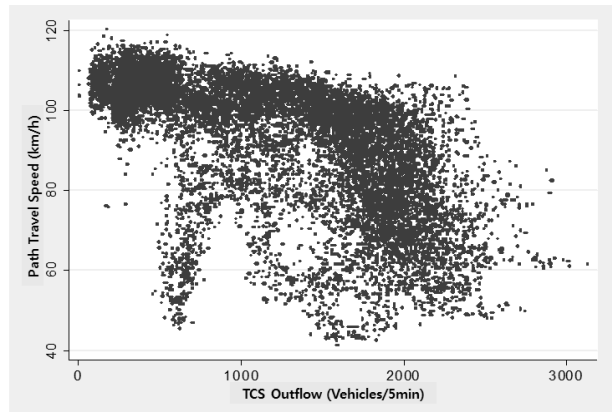


Fig. 6. Relation of TCS Outflow and Path Travel Speed (Seoul TG~Daejeon IC)

Table 4. Comparison Result by Prediction Model on Seoul TG ~ Daejeon IC

Method	MAPE (%)
ARIMA	24.5
Exponential Smoothing	26.7
KNN	7.0

공사 교통예보시스템에서 적용하고 있는 시계열 모형(ARIMA, 지수평활법)의 예측 오차를 비교한 결과는 Table 4와 같이 KNN 기반의 예측모형이 시계열모형의 예측오차에 비해 작게 나타났다. 이는 전술한 바와 같이 시계열모형이 실시간 상황을 제대로 반영하지 못하기 때문이다. 이러한 결과를 볼 때 향후 유사성을 판단할 수 있는 데이터베이스가 좀 더 축적될 경우 예측오차는 더욱 감소할 것으로 기대된다.

5. 결론 및 향후연구과제

본 연구는 TCS와 DSRC 자료를 기반으로 실시간 자료를 활용한 고속도로 통행시간 예측의 적절한 방법론을 도출하는 것을 목적으로 한다. 다양한 예측 기법에 대한 연구 검토 결과를 기반으로 적절한 방안으로 KNN 기반의 예측모형을 선정하였고, 이에 대한 적용가능성을 분석하였다. KNN 기반의 예측 모형 입력 자료는 실시간 교통상황을 반영할 수 있는 TCS 교통량과 DSRC 링크통행시간을 활용하였고, 실시간 및 과거자료에 대한 유사성은 유클리디안 거리를 통해 산출하였다. 유클리디안 거리가 가장 작은 최근접이웃 유사성의 5% 이내를 예측 시 활용할 최대근접이웃의 수로 산정하여 이를 적용한 주요 결과는 다음과 같다.

첫째, 토요일의 예측 오차 증가로 인해 주중보다 주말의 예측 오차가 더 크며, 일요일의 예측 오차가 토요일에 비해 작게 나타난 것은 하행 방향의 여가 목적 통행이 많지 않아 주중의 교통량과 큰 차이가 없기 때문에 나타난 것으로 판단된다. 또한 TCS 교통량의 가중치가 0.8인 경우를 볼 때 평균예측오차가 서울TG~오산IC 3.1분, 서울TG~청주IC 6.3분, 서울TG~대전IC 6.5분으로 경로길이 증가할수록 예측오차가 증가하는 것으로 나타났다.

둘째, 유사성 측정 결과를 볼 때 유클리디안 거리가 짧을수록 예측오차는 감소하는 것으로 나타났는데 이는 현재 상태와 가장 유사한 과거 상황을 기반으로 예측할 수 있다는 것을 의미한다.

셋째, DSRC 구간통행시간에 대한 가중치가 증가할수록 예측오차는 감소하는 것으로 나타났으며, 이는 TCS 교통량의 경우 하나의 TCS 진출입 교통량이 2개의 경로통행속도(정상류와 혼잡류)에 매칭되므로 현재와 상이한 과거 상황이 예측 시 활용될 수 있기 때문이다.

넷째, 현재 한국도로공사 교통예보시스템의 시계열모형과 비교한 결과 KNN 기반의 예측모형의 예측오차가 상대적으로 작게 나타났으며, 이는 KNN 기반의 예측모형이 시계열모형에 비해 실시간 상황을 잘 반영할 수 있기 때문이다.

이러한 결과로 볼 때 KNN 기반의 예측모형을 활용할 경우 실제 통행시간 변화 추이와 유사한 예측통행시간을 추정할 수 있으며, 향후 유사성을 판단할 수 있는 데이터베이스가 좀 더 축적될 경우 예측오차는 더욱 감소할 것으로 기대된다. 그러나 본 연구에서는 TCS 교통량과 DSRC 구간통행시간을 이용하여 KNN 기반의 예측 모형을 개발하였지만 이 중 TCS 교통량의 경우 방향별(상행, 하행) 진출입 교통량을 파악할 수 없는 문제가 있으므로 O/D와 진출입 교통량의 분석을 통해 산출한 방향별 비율로 이를 보완할 필요가 있다. 또한 TCS 교통량과 DSRC 구간통행시간의 가중치에 대한 민감도 분석만 이루어졌으므로 최적의 가중치를 산출하는 방법에 대한 연구도 필요하다. 마지막으로 사고 및 공사 등의 비반복 지체를 고려할 수 있는 변수 설정 및 적용 방안에 대한 추가적인 연구도 필요할 것이다.

감사의 글

본 연구는 한국도로공사의 맞춤형 교통콘텐츠 개발 관련 연구비, 데이터 및 2012년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행됨(NRF-2010-0029446).

본 논문은 대한교통학회 제70회 학술발표회(2014.2.21)에서 발표된 내용을 수정·보완하여 작성된 것입니다.

References

Ban, X. J., Li, Y., Skabardonis, A. and Margulici, J. D. (2010). "Performance evaluation of travel-time estimation methods for real-time traffic applications." *J. of Intell. Transp. Syst.*, Vol. 14, No. 2, pp. 54-67.

Chen, X., Gong, H. and Wang, J. (2012). "BRT vehicle travel time prediction based on SVM and kalman filter." *J. of Transp. Syst. Eng. and Inf. Technol.*, Vol. 12, No. 4, pp. 29-34.

Innamaa, S. (2005). "Short-term prediction of travel time using neural networks on an interurban highway." *Transp.*, Vol. 32, No. 6, pp. 649-669.

Jeong, R. and Rilett, L. R. (2004). "Bus arrival time prediction using artificial neural network model." *Proc. of the 7th Int. IEEE Conf. on Intell. Transp. Syst.*, Washington D.C., USA, pp. 988-993.

Kang, J. and Namkoong, S. (2002). "Development of the freeway operating time prediction model using toll collection system data." *J. of Korean Soc. Transp.*, Vol. 20, No. 4, pp. 151-162 (in Korean).

Kim, H. and Jang, K. (2013). "Short-term prediction of travel time

- using DSRC on highway.” *J. of the Korean Soc. Civil Eng.*, Vol. 33, No. 6, pp. 2465-2471 (in Korean).
- Lam, W. H. K., Tang, Y. F., Chan, K. S. and Tam, M. L. (2006). “Short-term hourly traffic forecasts using Hong Kong annual traffic census.” *Transp.*, Vol. 33, No. 3, pp. 291-310.
- Lee, C., Park, J. and Kho, S. (2002). “Prediction of path travel time using kalman filter.” *J. of the Korean Soc. Civil Eng.*, Vol. 22, No. 5-D, pp. 871-880 (in Korean).
- Lee, E. and Kim, J. (2002). “Development of a freeway travel time forecasting model for long distance section with due regard to time-lag.” *J. Korean Soc. Transp.*, Vol. 20, No. 4, pp. 51-61 (in Korean).
- Lee, Y. (2009). “Freeway travel time forecasting using artificial neural networks with cluster method.” *12th Int. Conf. on Inf. Fusion*, Seattle, USA, pp. 1331-1338.
- Lee, Y., Kim, S. and Youn, J. (2005). “Multi-step ahead link travel time prediction using data fusion.” *J. Korean Soc. Transp.*, Vol. 23, No. 4, pp. 71-79 (in Korean).
- Lim, S. and Lee, C. (2011). “Data fusion algorithm improves travel time predictions.” *IET Intell. Transp. Syst.*, Vol. 5, No. 4, pp. 302-309.
- Lim, S., Lee, H., Park, S. and Heo, T. (2013). “A study of travel time prediction using k-nearest neighborhood method.” *The Korean J. of Applied Statistics*, Vol. 26, No. 5, pp. 835-845 (in Korean).
- Mehmet, Y. and Nikolas, G. (2013). “Experienced travel time prediction for congested freeways.” *Transp. Res. Part B*, Vol. 53, pp. 45-63.
- Nigovski, D., Nishiuma, N., Goto, Y. and Kumazawa, H. (2005). “Univariate short-term prediction of road travel time.” *Proc. of the 8th IEEE Conf. on Intell. Transp. Syst.*, Vienna, Austria, pp. 1074-1079.
- Oh, S., Kim, M. and Baek, Y. (2003). “Development of a freeway travel time estimating and forecasting model using traffic volume.” *J. of Korean Soc. Transp.*, Vol. 21, No. 5, pp. 83-95 (in Korean).
- Park, S., Jeong, Y. and Kim, C. (2006). “Development and evaluation of real-time travel time forecasting model: Nonparametric Regression Analysis for the Seoul Transit System.” *J. of Korean Soc. Transp.*, Vol. 24, No. 1, pp. 109-120 (in Korean).
- Rice, J. and Zwet, V. E. (2004). “A simple and effective method for predicting travel times on freeways.” *IEEE trans. on Intell. Transp. Syst.*, Vol. 5, No. 3, pp. 200-207.
- Smith, B. and Demetsky, M. (1997). “Traffic flow forecasting: Comparison of Modeling Approaches.” *J. of Transp. Eng.*, Vol. 123, No. 4, pp. 261-266.
- Smith, B., Williams, B. and Oswald, K. (2002). “Comparison of parametric and non parametric models for traffic flow forecasting.” *Transp. Res. Part C*, Vol. 10, No. 4, pp. 303-321.
- Wei, C. H. and Lee, Y. (2007). “Development of freeway travel time forecasting models by integrating different sources of traffic data.” *IEEE Trans. Veh. Technol.*, Vol. 56, No. 6II, pp. 3682-3694.
- Xie, C., Cheu, R. and Lee, D. (2004). “Improving arterial link travel time estimation by data fusion.” *Proc. of 84th TRB Annual Meeting*, Washington D.C., USA.
- You, J. and Kim, T. (2000). “Development and evaluation of a hybrid travel time forecasting model.” *Transp. Res. Part C*, Vol. 8, No. 1-6, pp. 231-256.
- Yu, J. (2008). “Real-time travel time estimation model using point-based and link-based data.” *J. of Korean Soc. Road Eng.*, Vol. 10, No. 1, pp. 155-164 (in Korean).
- Zhang, X. and Rice, J. (2003). “Short-term travel time prediction.” *Transp. Res. Part C*, Vol. 11, No. 3-4, pp. 187-210.