

## 대용량 이력자료를 활용한 다중시간대 고속도로 교통량 예측

장현호\* · 윤병조\*\*

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

### Multiple Period Forecasting of Motorway Traffic Volumes by Using Big Historical Data

#### ABSTRACT

In motorway traffic flow control, the conventional way based on real-time response has been changed into advanced way based on proactive response. Future traffic conditions over multiple time intervals are crucial input data for advanced motorway traffic flow control. It is necessary to overcome the uncertainty of the future state in order for forecasting multiple-period traffic volumes, as the number of uncertainty concurrently increase when the forecasting horizon expands. In this vein, multi-interval forecasting of traffic volumes requires a viable approach to conquer future uncertainties successfully. In this paper, a forecasting model is proposed which effectively addresses the uncertainties of future state based on the behaviors of temporal evolution of traffic volume states that intrinsically exists in the big past data. The model selects the past states from the big past data based on the state evolution of current traffic volumes, and then the selected past states are employed for estimating future states. The model was also designed to be suitable for data management systems in practice. Test results demonstrated that the model can effectively overcome the uncertainties over multiple time periods and can generate very reliable predictions in term of prediction accuracy. Hence, it is indicated that the model can be mounted and utilized on advanced data management systems.

**Key words :** Motorway, Multiple time-period, Forecasting, Big data, k-Nearest Neighbor-NonParametric Regression (KNN-NPR)

#### 초 록

고속도로 교통류 제어는 기존의 Reactive 방식(실시간 대응)에서 Proactive 방식(사전 대응)으로 발전하고 있다. 첨단 고속도로 교통류 제어의 핵심 입력자료 중 하나는 여러 시간대에 걸치는 장래 교통량 상태이다. 다중 시간대 교통량 예측을 위해서는 장래 상태의 불확실성을 극복해야 한다. 이는 예측 시간대의 확장에 따라 장래 상태의 불확실성은 증가하기 때문이다. 따라서 다중 시간대 교통량 예측을 위해서는 장래 상태의 불확실성을 효과적으로 극복할 수 있는 실행 가능한 방안이 필요하다. 본 연구에서는 대용량 이력자료에 내재된 교통류 상태의 시간적 진화 행태를 이용하여 장래 상태의 불확실성을 효과적으로 극복함으로써 다중 시간대 장래 교통량 상태를 예측하는 모형을 제시하도록 한다. 개발 모형은 현행 교통량의 상태 진화를 기반으로 대용량 자료에 내재된 과거 상태를 추출하고, 이를 이용하여 장래 상태를 예측한다. 추가로, 개발된 모형은 실제 적용을 고려하여 자료관리시스템에 적합하도록 설계되었다. 적용결과, 개발모형은 다중 시간대에 걸치는 불확실성을 효과적으로 극복함으로써 우수한 예측력을 보였으며, 첨단자료관리시스템에 실제 적용이 가능하다고 판단된다.

**검색어 :** 고속도로, 다중시간대, 예측, 대용량자료, k-최근린 이웃 비모수회귀

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

\*\* 정회원 · 교신저자 · 인천대학교 도시과학대학 도시공학과 부교수, 공학박사 (Corresponding Author · Incheon National University · bjuoon63@inu.ac.kr)

Received September 11, 2017/ revised October 13, 2017/ accepted November 17, 2017

## 1. 서론

첨단 고속도로 교통류 제어에 있어 장래 교통상황을 예측하는 것은 매우 중요하며, 이를 위해서는 장래 교통수요 예측은 필수적이다. 그러나 기존의 교통량 예측은 시간대( $t$ )까지의 시계열자료를 이용하여 시간대( $t+1$ )의 상태를 예측한 단일시간대 예측이라고 할 수 있다. 예측기법으로는 상태공간(state space, SS) 분석, 인공신경망(artificial neural network, ANN), 시계열분석(ARIMA), 평활화 등 다양한 시계열자료 예측기법이 이용되고 있다. 그러나 이러한 단일 시간대 예측기법은 다중시간대에 걸쳐 변화하는 교통량을 예측할 수 없는 단점이 있다. 이는 실시간으로 수집되는 교통량 시계열 자료를 이용하여 교통량을 단기로 예측하였기 때문이다.

국내에 ITS가 도입된 이후, 교통정보 수집기술과 더불어 방대한 양의 교통자료를 수집·활용할 수 있는 기반이 조성되었다. 대부분의 수집된 교통자료는 실시간 교통정보제공 및 교통운행을 위한 목적으로 활용되어 왔으며, 이력자료는 효율적으로 관리·활용되고 있지 않는 실정이었다. 그러나 최근에는 이력자료를 유자관리하기 위한 첨단자료관리시스템이 도입되기 시작하면서 이력자료의 활용 가능성은 매우 높아지고 있다.

따라서 본 연구에서는 (다중시간대에 걸친 사전대응적인 고속도로 교통류 제어 및 운영을 위하여) 기존 단일 시간대 예측기법의 한계를 극복한 다중시간대 교통량 예측기법을 개발하였다. 개발된 모형은 실시간 교통량 시계열 자료와 방대한 양의 이력 교통량자료를 이용하도록 설계되었다. 그리고 자료관리시스템에 탑재되어 운영이 용이하도록 설계되었다. 개발모형은 방대한 양의 이력자료와 검색지원시스템이 제공될 경우 효과적인  $k$ -최근린 이웃 비모수 회귀( $k$ -nearest neighbor non-parametric regression, KNN-NPR)를 기반으로 개발하였다.

## 2. 기존 연구

본 연구는 KNN-NPR를 기반으로 다중시간대 교통량을 예측하고자 한다. 따라서 기존연구 고찰은 기존의 예측기법과 KNN-NPR 기법을 적용한 연구에 대하여 수행하였다.

예측기법은 실시간 현행 시계열자료를 이용하여 시간대( $t$ )에서 시간대( $t+1$ )의 상태를 예측하는데 이용되고 있다. 예측기법으로는 ANN, (수행속도가 빠른) SS, (계절형) ARIMA, 평활화 등의 기법이 주로 적용되고 있다. 이러한 예측기법은 엄밀히 말해 단일 시간대 통행시간 예측기법이다. 따라서 단기 시계열 예측기법의 특성상 시계열적 변동이 큰 상태에서 예측력이 현저히 감소하게 되며, 시간차임(time-lag)을 극복할 수 없는 한계를 가지고 있다. 이러한 한계를 극복하고자 Chang et al.(2003)은 시뮬레이션을

이용하여 단기 구간통행시간을 예측하였으나, 수행속도가 낮고 기본적으로 단기 구간교통량을 예측하여 입력값으로 적용하였다. 이상의 기존 단일시간대 예측기법은 통행시간 및 교통량 시계열자료의 시계열적 변동특성과 시간차임을 극복할 수 없기 때문에 사실상 다중시간대 이후의 상태 예측에는 한계가 있다(Chang et al., 2010). 그리고 이들 모형은 실시간으로 수집되는 자료를 이용하도록 설계되었기 때문에 방대한 양의 정보를 포함하고 있는 이력자료의 활용에 적합하지 않은 단점이 있다. 또한 ARIMA의 경우, 평균값에 기반하여 시계열 변동을 분석하기 때문에 변동의 극단치를 (심지어 지속적인 변동 조차) 설명할 수 없는 단점이 있다.

교통량 상태의 시간적 진화는 카오스 상태에 가깝다(Yoon and Chang, 2014). 따라서 교통량의 시계열 변동은 속도 및 통행시간보다 임의성을 수반한 변동이 매우 크게 나타난다. 따라서 단기 교통량 예측에는 다양한 예측기법의 적용(Vlahogianni et al., 2014)이 보고된 반면, 다중 시간대 교통량 예측은 매우 제한적이다. Vlahogianni et al.(2005)은 인공신경망을 기반으로 그리고 Stathopoulos and Karlaftis(2003)은 SS를 이용하여 다중시간대 단속류 교통량을 예측하였다. 두 연구에서 단일 시간대의 경우 10% 이내의 수용 가능한 오차를 보였으나, 다중 시간대의 경우 오차는 20% 이상으로 급격히 증가하였다. 이는 두 모형의 경우, 장래 상태의 불확실성 즉, 상태의 진화방향과 변동량을 설명할 수 없음을 의미한다. 실제 적용에 있어, 인공신경망의 경우 이력자료를 활용한 신경망의 학습에 필요한 시간이 요구되기 때문에 시간 임계적(time-critical)인 ITS 시스템 적용사례는 보고되고 있지 않다. 추가로, Chang et al.(2010)은 다중시간대에 걸치는 장래 상태의 불확실성을 극복하고자 이력자료와 버스정류장간 출발지 기준 경로통행시간 현행 시계열 자료를 이용하여 출발지 기준 버스경로통행시간을 예측하였으며, 현행 시계열자료와 이력자료를 활용하기 위한 방법으로 KNN-NPR기법을 적용하였다.

회귀식은 (모수)회귀식과 비모수회귀식으로 구분할 수 있다. (모수)회귀식은 설명변수에 의한 영향, 즉 파라미터가 장래에도 동일하게 종속변수에 영향을 미친다는 가정을 전제로 한다는 문제를 갖는다(Oswald et al., 2000). 이에 비해 NPR 기법은 현재 조건과 유사한 과거 관측치를 탐색하여 장래 상태를 추정하는데 적용이 용이하다(Smith et al., 2003). 일반적인 (모수)회귀식 모형과 달리 NPR 모형은 입력과 출력자료간의 관계를 결정하는 것으로서 파라미터의 계산 없이 새로운 관측 자료를 쉽게 모형에 추가할 수 있다. KNN-NPR 기법은 과거의 패턴 즉, 경험을 바탕으로 장래의 통행시간을 결정하는 경험기반 의사결정과정이라고 할 수 있다(Smith et al., 2003).

KNN-NPR 기법을 적용하기 위해서는 방대한 양의 이력자료가 요구되며, 과거 상태와 현재 상태를 이용한 장래 상태의 예측에

있어 다양한 분야에 이용되고 있다. Robinson(1983)과 Mulhern and Caprara(1994)는 혼재된 무질서 상태에서 비선형 시계열 예측 문제를 NPR을 기반으로 모형화하였으며, 무질서 상태에서 NPR이 장점을 갖는다고 지적하였다. Karlsson and Yakowitz(1987)은 KNN-NPR을 기반으로 강우량을 예측하였다. ITS 예측분야에서 KNN-NPR은 교통량 또는 통행시간 예측에 주로 적용되었다. Davis and Nihan(1991), Oswald et al.(2000), Smith et al.(2002)은 KNN-NPR을 교통량 예측에 적용하였으며, 예측오차는 10% 이상 높게 나타났으며 이는 소량의 이력자료를 이용하였기 때문이다. 또한 Oswald et al.(2000)의 경우, 요일 등 시간적 주기성에 대한 고려가 없었다. 그리고 Qi and Smith(2004)는 유고지속시간을 예측하였다. Chang et al.(2010)은 BMS 이력자료와 실시간 통행시간 자료를 이용하여 다중시간대에 걸치는 버스 정류장간 출발지 기준 경로통행시간을 예측하였다. 이상의 연구는 대용량 이력자료의 활용이 가능한 환경에서 KNN-NPR의 다중시간대 예측에 대한 가능성을 언급하였다. 최근 들어 첨단자료관리시스템의 도입과 더불어 대용량 이력자료의 활용이 가능해졌으며, 이러한 자료 환경의 변화는 이력자료기반 예측에 있어 새로운 기회를 제공해주고 있다.

### 3. 예측모형 개발

#### 3.1 개발시 고려사항

##### 3.1.1 시계열 변동 및 주기성 고려

교통상태의 복잡성은 주중/주말, 그리고 각 요일별 시간대별 패턴과 특성을 시계열 변동에 포함되어 있다. 따라서 시계열자료에 평활화 과정을 적용할 경우 시계열자료의 변동과 정보는 희석되게 되며, 추정되는 상태의 시계열적 변동 또한 낮아지게 된다. Fig. 1은 수원 톨게이트(이하 TG)의 고속도로 유입방향에 대한 시계열적 변동과 시간대별(15분 간격) 주기성을 보여주고 있으며, 확률적(stochastic) 상태라기 보다는 일정 패턴이 내재되어 있는 혼돈

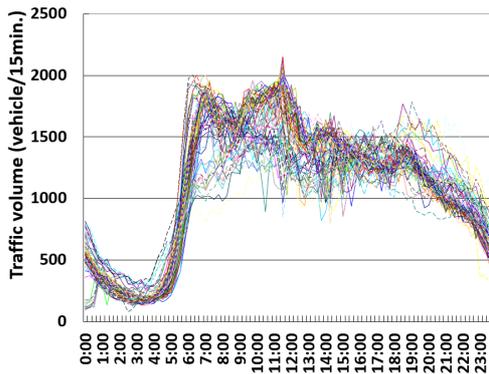


Fig. 1. Time-Series Variation of Traffic Volume

(chaos)상태, 즉 초기결정상태(initial deterministic condition)에 가깝다 할 수 있겠다. 따라서 예측모형은 시계열적 변동뿐만 아니라 특정 시간대의 반복적 주기성을 고려할 수 있어야 한다.

##### 3.1.2 유동적인 시간대 예측 범위

예측 교통량 자료는 시계열자료의 시간대 길이와 이용 목적에 따라 다양한 시간대에 대한 예측이 가능해야 한다. 본 연구는 다중시간대에 걸친 TG의 교통수요를 예측하는데 있다. 따라서 개발되는 모형은 단일 시간대에서 복수개의 시간대까지 적용이 가능한 모형의 구조로 설계되어야 한다. 단일시간대의 예측은 예측의 정확도와 시계열자료의 변동에 따른 민감도를 고려해야 함과 동시에 시간처짐을 효과적으로 극복해야 한다. 그리고 다중시간대 예측에 있어서는 예측오차를 최소화하면서 추세를 반영할 수 있어야 한다.

##### 3.1.3 추정의 불확실성 문제

임의성을 수반하는 동적인 교통상태의 특성을 고려하면 예측상태에 대한 불확실성은 예측/추정 기법의 근본적인 문제라 할 수 있다. 모형이 지나치게 정적이거나 민감하지 않은 경우 교통상태의 동적특성을 고려할 수 없으며, 지나치게 민감한 경우 교통상태의 동적특성을 고려할 수 있으나 (국부적으로) 예측오차는 증폭될 수 있다. 따라서 개발되는 모형은 요구되는 추정오차의 범위에서 교통량 상태를 예측하면서 안정적이도록 개발되어야

##### 3.1.4 적용의 용이성

모형의 추정력이 요구수준을 만족하면 모형의 연산수행 속도와 더불어 운영에 있어 적용성을 고려해야 한다. 모형의 예측력이 뛰어나더라도 시스템 측면에서 연산수행 시간이 예측정보를 활용할 수 있는 주어진 시간적 범위를 초과하거나, 운영측면에서 지속적으로 모형의 파라미터를 조정해야 한다면 실제 모형의 적용성은 떨어진다. 따라서 파라미터의 조정이 시스템적으로 간편하면서 모형의 연산속도가 빠른 모형을 개발하여야 한다.

#### 3.2 접근법 및 예측 영역

KNN-NPR모형의 수행과정은 일반적으로 현행 시계열 상태와 상태간 거리인  $l_p$  거리를 이용하여 현행 시계열 상태와 이력 시계열 상태의  $l_p$  거리가 가까운 과거의 시계열 상태를 탐색하는 과정을 통해서  $k$ 개 최근인 이웃과  $k$ 개 출력상태로 구성되는 군집을 구성하는 과정과 구축된 군집과 예측모형을 이용하여 장래 상태를 예측하는 Fig. 2 과정으로 구성된다.

본 연구는 기존의 단일시간대 예측기법이 아닌 다중시간대 예측을 수행하고자 한다. 따라서 시간대(t)에서 시간대(t+1)의 상태를 예측하는 기존의 예측기법과는 달리 다중시간대 예측을 위한 예측

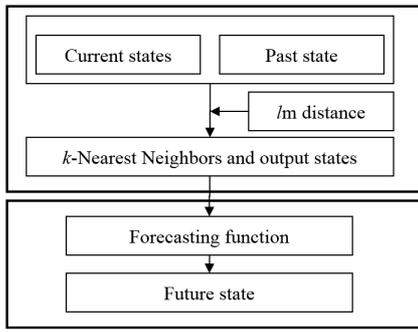


Fig. 2. Structure of KNN-NPR Model

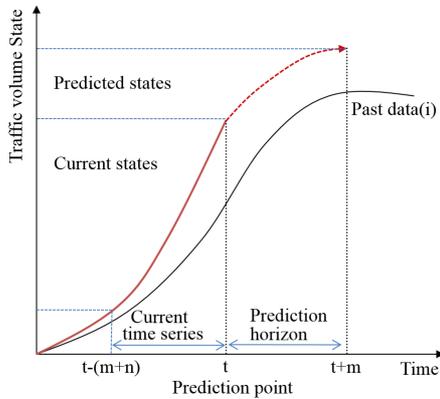


Fig. 3. Prediction Horizon

영역을 Fig. 3과 같이 설정하였다. 예측하고자 하는 장래 시간대 (t+m, m ≥ 1)에 대한 현행 시계열은 [t-(m+n) ~ t]로 정의하며, n은 m 시간대에 대한 시계열의 크기를 조정하기 위한 파라미터이다. 일반적으로 예측하고자 하는 장래시간대의 시점인 m이 증가할수록 현행 시계열 자료의 시간적 길이는 증가한다.

### 3.3 다중시간대 예측모형 개발

KNN-NPR모형은 상태벡터, 상태간 거리 Metric, k개의 이웃으로 구성되는 군집의 자료구조, 예측함수, 그리고 군집을 구축하기 위한 탐색 알고리즘으로 구성된다.

#### 3.3.1 상태벡터의 구성

상태벡터는 입력 상태벡터와 출력 상태벡터로 구성되며, 입력 상태벡터는 현행 상태벡터, 과거 상태벡터, 그리고 출력 상태벡터로 구성된다. 따라서 본 연구에서 상태벡터는 다음과 같이 ① 현행 교통량 상태벡터( $q_c$ ), ② 이력 교통량 상태벡터( $q_h$ ), 그리고 ③ 장래 교통량 상태벡터( $q_p^k$ )로 정의한다.

- ①  $q_c = [q_c(j)], t-(m+n) \leq j \leq t$ : 현행 자료의 시간대(t-(m+n))에서 시간대(t)까지 현행 교통량 상태벡터

- ②  $q_h^i = [q_h^i(j)], \tau-(m+n) \leq j \leq \tau$ : i 제 이력자료의 시간대 ( $\tau-(m+n)$ )에서 시간대( $\tau$ )까지 이력 교통량 상태벡터, 여기서  $\tau$ 는 과거 시간의 흐름(running time index)이다.

- $q_p^k = [q_p^k(K)], 1 \leq K \leq k$ : k개의 이웃으로 구성되는 시간대 t+m의 예측 교통량 출력 상태벡터

#### 3.3.2 상태거리 Metric

현행 상태벡터( $q_c$ )와 유사한 이력 상태벡터( $q_h$ )를 탐색하여 장래 상태벡터( $q_p$ )로 구성된 군집을 구성하기 위해서는 탐색 알고리즘을 이용하게 되며, 탐색의 과정에서  $q_c$ 와  $q_h^i$  간의 상태거리는 Eq. (1)과 같은  $l_p$  거리를 주로 이용한다.  $l_p$  거리는  $p=1$ 인 경우 Manhattan 거리,  $p=2$ 이면 Euclidean 거리로 알려져 있다. 본 연구에서는 상태벡터간 거리가 증가할수록 상태거리가 증가하는 Euclidean 거리를 이용하도록 한다. Euclidean 거리는 Manhattan 거리와 같이 상태벡터 간의 절대거리가 같은 경우라도 상태벡터의 요소간 거리를 증폭하기 때문에 보다 민감하게 상태 유사성을 설명할 수 있다.

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

$l_p$  거리를 이용하여 상태  $q_c$ 와  $q_h^i$  간의 거리를 표현하면 Eq. (2)과 같으며, 두 상태간의 거리는  $u^i$ 로 정의한다.

$$u^i = \left[ \sum_j |q_c(j) - q_h^i(j)|^2 \right]^{1/2} \quad (2)$$

#### 3.3.3 군집해의 자료구조

군집을 구성하는 이웃을 탐색/구축하기 위해서는 k개의 이웃을 저장하기 위한 자료구조가 요구된다. 자료구조는 k개의 유사 이웃에 대한 입력 상태벡터와 출력 상태벡터로 구성되며, Table 1과 같다.

Table 1. Data Structure of KNN

i	Input state vector		Output state vector	
	Independent variable	Dependent variable	Dependent variable	Nearness
1	$q_h^1(j), q_c(j)$	→	$q_h^1(t+m)$	$u_1$
2	$q_h^2(j), q_c(j)$	→	$q_h^2(t+m)$	$u_2$
...	...		...	...
K	$q_h^K(j), q_c(j)$	→	$q_h^K(t+m)$	$u_K$
...	...		...	...
k	$q_h^k(j), q_c(j)$	→	$q_h^k(t+m)$	$u_k$

Note:  $q_h^K(j) \in q_h^i, q_h^K(t+m) \in q_h^i, t-(m+n) \leq j \leq t, 1 \leq K \leq k, 1 \leq m, \forall j, k, m$

### 3.3.4 예측합수

군집해가 구축되면 예측합수를 이용하여 장래 시간대( $t+m, m \geq 1$ )의 교통량을 예측하게 된다. 예측합수는 종속변수를 산술평균한 방법과 상태간 거리의 역수로 가중 평균한 방법이 주로 이용된다. 그러나 산술평균 모형은 상태간 거리를 고려할 수 없는 단점이 있다(Yoon and Chang, 2014). 이러한 단점을 극복하기 위하여 상태간 거리의 역수로 가중 평균한 방법을 많이 이용하며 산술평균 보다 우수한 결과를 보였다(Smith et al., 2002; Yoon and Chang, 2014). 따라서 본 연구에서는 상태간 거리의 역수로 가중 평균한 모형을 이용하도록 하며, 이는 상태간 거리가 가까울 수록 장래 상태의 결정에 있어 보다 많은 의사결정 권한을 부여한다고 할 수 있다.

$$\hat{q}_p(t+m) = \sum_{i=1}^k \left( \frac{q_h^i(t+m)}{u_i} \right) / \sum_{i=1}^k \frac{1}{u_i} \quad (3)$$

### 3.3.5 예측 알고리즘

본 연구에서는 군집을 구축하는 방법으로 최인접 개수  $k$ 를 정하는 KNN 알고리즘을 이용하였다. KNN 알고리즘은 ① 주어진 대응량 자료에서 이력 상태벡터( $q_h^i(j), q_h^i(j) \in q_h$ )을 탐색하여 현행 상태벡터  $q_c(j)$ 와 상태거리가 가장 가까운  $k$ 개의 이웃해 ( $q_h^k(t+m), q_h^k(t+m) \in q_h$ )를 구성하는 탐색과정과 ② 예측과정으로 구성되며 다음과 같다.

주어진 장래 시간대( $m$ ), 최인접 개수  $k$ ,  $q_c(j)$ ,  $q_h^i(j)$  대하여:

- ① 군집해 자료구조의 1에서  $k$ 까지 이웃을 초기화
- ② 모든 이력 상태벡터( $q_h(j)$ )에 대하여
  - ㉠  $i$ 번째 이력 상태벡터( $q_h^i(j)$ )와 현행 상태벡터( $q_c(j)$ )간의  $u_i$ 을 Eq. (2)를 이용하여 계산
  - ㉢ If  $u_i < U_{\max}$  Then  
(여기서,  $U_{\max} = \max\{u_1, \dots, u_k, \dots, u_k\}$ )
    - i)  $U_{\max}$ 에 해당하는 이웃을 제거
    - ii)  $u_i$ 의 독립변수, 종속변수, 그리고 상태거리를 군집에 갱신
    - iii) 갱신된 군집에서  $U_{\max}$ 을 탐색
- ③ Eq. (3)를 이용하여  $q_p(t+m)$ 을 산정

## 4. 적용 및 평가

### 4.1 실험 설계

#### 4.1.1 자료 구축

이용 자료는 한국도로공사의 TCS (Toll Collection System)에

서 수집된 TG의 교통량 자료이다. 본 연구의 사례 대상은 수원 TG의 고속도로 유입 방향이며, 수집된 TCS 자료는 15분 간격으로 집계/구축하였다. 구축자료의 시간적 범위는 2016년 1월 1일에서 10월 31일이며, 마지막 1주일의 교통량 예측에 적용하였다. 따라서 구축된 이력자료는 1월 1일에서 10월 31일까지 총 305일간의 15분 교통량 자료이며, 전체 자료의 크기는 29,280 (=305일×96시간/대/일) 이다. 그리고 교통사고, 도로점유공사, 폭우, 폭설 등 비탄복적 상황에 관련된 경우는 분석시 이력자료에서 제외하지 않았다. 이는 본 연구에서 개발된 모형의 접근법인 KNN 경우, 이력자료에 내재된 다양한 패턴에 대한 강력한 패턴인식을 기반으로 하기 때문이다(Davis and Nihan, 1991; Yoon and Chang, 2014). 이는 실제 모형의 적용에 있어 장점이기도 하다.

예측의 범위는 시간대( $t$ )를 기준으로 장래 4개 시간대( $m=4$ ), 즉 (15분, 30분, 45분, 60분)이며, 10월 마지막 일주일의 모든 시간대에 대하여 예측을 수행하였다.

### 4.1.2 평가 지표

모형의 예측력을 평가하기 위하여 평균절대값백분율오차(Mean Absolute Percentage Error, MAPE (%))를 이용하였다. MAPE는 교통량과 같이 상태의 변동이 큰 경우 직관적 평가에 적합하다(Smith et al., 2002). 또한 MAPE는 적정  $k$ 값의 산정에 이용하였다. 그리고 추정 결과의 세부적인 시계열 분석은 시계열적 변동을 통해 분석하도록 한다.

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

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

## 4.2 파라미터 및 예측력 분석

### 4.2.1 적정 $k$ 값 및 오차분석 결과

Fig. 4는  $k$ 값에 따른 MAPE (이하, 오차) 분석결과를 보여주고 있다.  $k$ 값이 증가함에 따라 오차는 감소 후 증가하고 있다. 즉,  $k$ 값과 오차의 관계는 아래로 볼록한 관계를 보인다. 이러한  $k$ 값과 오차의 관계는 교통량 상태의 시간적 진화과정에 있어 패턴간 경계가 명확해진 명확하지 않던 패턴이 존재함을 의미한다. 또한 가장 유사한 1개의 과거자료로 장래를 예측할 시 오차를 최소화하지 않음을 알 수 있다. 이는 교통량 상태의 시간적 진화는 패턴 내에서도 일정한 변동을 수반하며, 그 변동은 유사한 패턴에 보다 의사결정권을 부여하면서 보다 안정적으로 장래 상태의 불확실성을 극복할 수 있음을 의미한다.

평균 최적  $k$ 값 ( $k_o$ )은 9로 분석되었으며, 이때 오차는 약 8.8%로

나타나고 있다. 토요일과 일요일의 경우,  $k_0$  값은 평균보다 낮은 수준인 5와 6으로 각각 분석된 반면, 월요일과 금요일의 경우 10% 오차 내외에서  $k_0$  값은 9와 11로 각각 분석되었다. 월요일과 금요일의 경우, 오차곡선은  $k_0$  값의 근처에서 크게 벗어나지 않고 있다. 이는 유사패턴의 변동성이 타 요일에 비하여 높기 때문이다. 토요일과 일요일의 오차곡선은  $k_0$  값을 전/후로 급격히 상승하고 있다. 이는 다른 요일에 비하여 이력자료에 내재된 패턴 다양성이 크며, 패턴간 차이가 상이할 수 있음을 의미한다. 그리고 화-목요일의  $k_0$  값은 8~10에서 결정되고 있으며, 오차는 월요일과 금요일보다 낮게 나타나고 있다. 이는 화-목요일의 경우 월요일과 금요일 정도의 패턴은 내재되어 있으나, 시계열적 변동이 월요일과 금요일만큼 심하지 않음을 의미한다. 또한  $k_0$  값의  $\pm 3\sim 5$  범위 내에서 오차는 크게 변하지 않고 있다. 이는 사전에  $k_0$  값이 월기반 요일기반으로 분석된 후 적용할 수 있음을 의미한다. 장래 예측 시간대( $m$ )별 적정 시계열 개수는 평균 6으로 분석되었으며, 시간대별로  $\pm 1$ 의 차이를 보였으나 예측오차의 차이는 미비하였다. 그리고 적정 시계

열 개수는 요일별로 동일하게 나타났다. 이는 장래 예측영역의 시간대 개수 및 요일별로 특정 시계열의 길이가 존재함을 의미하며, 시간대의 길이가 15분 일 경우, 그 길이는 약 90분 정도임을 의미한다. 따라서 실험분석에 의해  $t+m$  별 적정  $n$  값이 결정되면 지속적으로 적용할 수 있음을 의미한다.

Table 2는  $k_0$  값에서 요일별 오차분석 결과를 보여주고 있다. 예측 영역이 증가함에 따라 평균 오차는 t+1 예측시 7.61%에서 t+4 예측시 9.62%까지 증가하고 있다. 이러한 오차의 행태는 t2부터 예측오차가 급격히 증가하는 기존의 연구결과(Stathopoulos and Karlaftis, 2003; Vlahogianni et al., 2005)와 상이한 행태를 보이고 있다. 이는 대용량 과거자료를 이용한 KNN-NPR이 다중시간대에 걸치는 장래 상태의 불확실성을 오차 10%내에서 극복하고 있음을 의미한다. 예측오차는 토요일과 일요일이 평균 7.6%로 가장 낮았으며, 화-목요일은 평균 8.8%, 월요일과 금요일은 각각 9.8%와 10.0%로 가장 높게 나타났다.

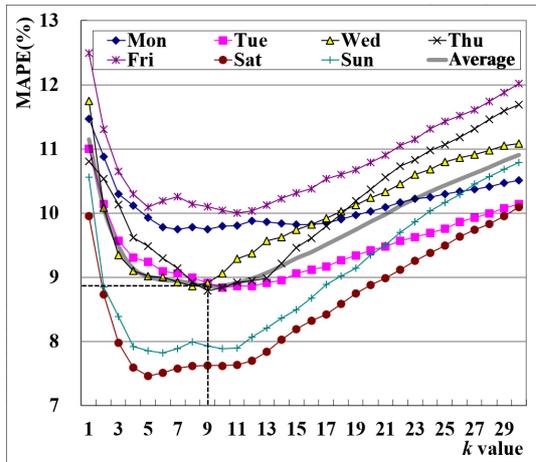


Fig. 4. Effects of  $k$  Value on Prediction Errors

Table 2. Summary of Analysis Results

Day	$k_0$	Prediction Error				
		Time-step ahead				Average
		t+1	t+2	t+3	t+4	
Mon	9	8.71	10.05	9.93	10.33	9.76
Tue	10	7.80	8.74	9.13	9.68	8.84
Wed	8	7.87	8.90	9.18	9.50	8.86
Thu	9	7.61	8.86	9.08	9.62	8.79
Fri	11	8.45	9.77	10.53	11.27	10.00
Sat	5	6.67	7.72	7.73	7.73	7.46
Sun	6	6.17	7.99	7.95	9.18	7.82
Average	9	7.61	8.86	9.08	9.62	8.79

#### 4.2.2 시계열 분석

Figs. 5~8은 요일(월, 금, 토, 일)별로 다른 시계열 형태에서 예측결과를 보여주고 있다. 월요일의 경우(Fig. 5), 오전에 급격한 교통량 변화가 4번 발생하는 다중 첨두를 보이고 있으며, 오전 비첨두는 일반적인 패턴에서 벗어나는 경우이다. 급격한 교통량의

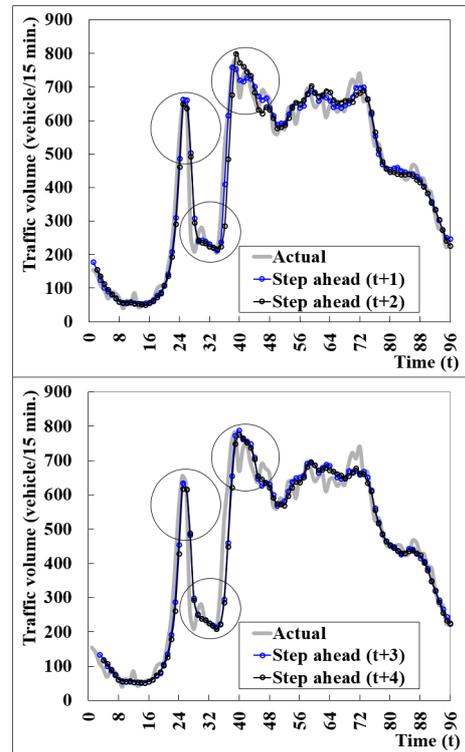


Fig. 5. Time-Series Comparison: Monday

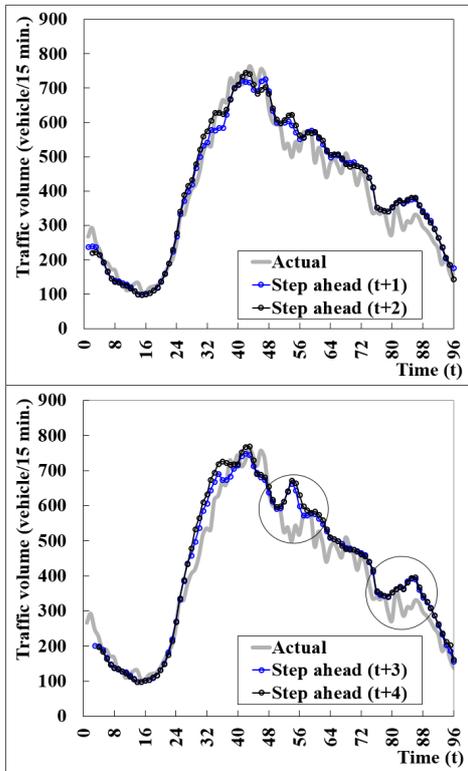


Fig. 6. Time-Series Comparison: Friday

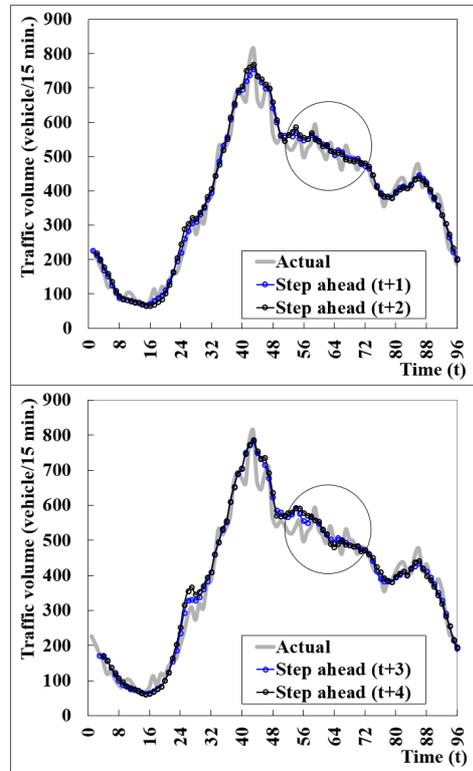


Fig. 8. Time-Series Comparison: Sunday

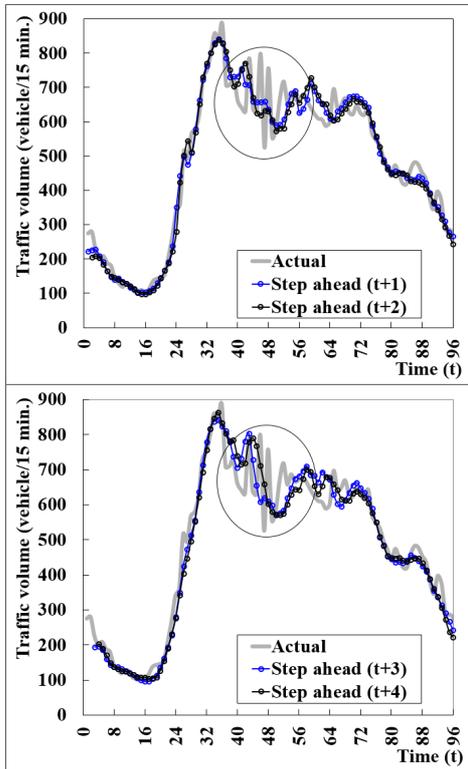


Fig. 7. Time-Series Comparison: Saturday

변화시, 예측정도는 기존 단일 시간대 예측기법에서 발생하였던 시간차집이 발생하지 않고 있으며 시계열적 추세를 반영하고 있다. 따라서 비반복적인 상태에서도 본 연구에서 개발된 모형은 강력한 패턴인식 기능을 가진다고 할 수 있다. 반면 금요일의 경우(Fig. 6), 예측시간대 t+3과 t+4의 15와 20시 근처에서 급격한 변화시 오치는 다소 증가하고 있다. 이는 과거 패턴의 부재로 인하여 상태의 방향성을 상실하였기 때문이다. 따라서 충분한 이력자료를 통해 패턴 다양성이 확보될 경우 극복할 수 있을 것으로 판단된다. 그리고 토요일의 9~12시(Fig. 7)와 일요일의 12~16시(Fig. 8)의 경우 시계열적 변동이 매우 크게 나타나고 있지만, 기존 단일시간대 예측기법의 지그재그 현상으로 발생하는 예측실패를 평활화를 통해 극복하고 있다.

#### 4.2.3 연산 수행속도

본 연구에서 예측한 총 회수는 일반 데스크탑 PC에서 80,640회 (= 7일 × 96시간대/일 × 4개 시간대 × 30개 k값)의 수행에 총 448초가 소요되었다. 따라서 1번의 예측에 소요된 시간은 약 0.005556초이다. 도로공사 TG의 개수는 약 450개로서 유입/유출 방향으로 900개소에 달한다. 따라서 모든 TG별 방향별 4개 시간대에 대한 예측시 연산시간은 개략적으로 20.0016초가 소요되었다.

그러나 ITS 정보시스템과 같은 시간 임계적인 시스템의 입장에서 연산시간 20초는 기존의 모형에 비하여 상대적으로 느린 연산속도이다. 따라서 20개 연산장치에서 병렬연산을 수행할 경우 1.0초에 예측이 가능함으로 시스템의 자료흐름과 동기화가 가능할 것으로 판단된다. 또한 주요 TG를 대상으로 연산을 수행할 수 있을 것으로 판단된다.

## 5. 결론 및 향후연구

본 연구에서는 기존 단일 시간대 예측기법의 한계를 극복하고자 다중시간대 교통량 예측기법을 개발하였다. 개발된 모형은 실시간 현행 교통량 자료와 이력 교통량 자료를 이용하도록 설계되었다. 그리고 첨단자료관리시스템에 탑재 및 운영이 용이하도록 이력자료에 내재된 행태를 이용하는 KNN-NPR을 기반으로 개발되었다. 모형의 적용 결과, 예측치는 관측치의 시계열적 추세와 변화를 잘 반영하였으며, 이는 교통량 상태 진화의 방향성과 변동량을 효과적으로 설명하였음을 의미한다. 그리고 비반복적 상황에서도 강력한 패턴인식 능력으로 장래상태를 추정하는 것으로 나타났다. 따라서 장래 상태의 불확실성에 기존의 모형에 비하여 효과적으로 대응하고 있다고 할 수 있겠다. 따라서 다중시간대 예측에도 불구하고 평균오차는 7.61~9.62%로 나타나 기존의 다중시간대 교통량 예측의 연구보다 우수하게 나타났다. 개발모형은 파라미터를 최소화하였으며, 개발모형의 연산속도는 시스템 탑재가 가능한 것으로 분석되어 모형의 적용성은 우수하다고 판단된다.

향후 연구로는 다양한 교통량 시계열자료에 대한 보다 적용연구가 필요하다. 장래 혼잡 시뮬레이션, 교통사고분석 등을 위해서는 예측 시간대를 본 연구의 4개(1시간)에서 12개(3시간)까지 확장하는 연구가 필요하다. 추가로, 첨단자료관리시스템과 더불어 첨단 검색엔진이 지원되지 않는 기존의 자료관리시스템에 적용을 위해서는 기존 자료관리시스템하에서 과거 자료의 탐색시 연산을 획기적으로 감소시킬 수 있는 방안에 대한 연구가 필요할 것으로 판단된다.

## References

Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression." *The American Statistician*,

Vol. 46, pp. 175-185.  
 Chang, H. and Lee, S. (2003). "A study on link travel time prediction by short term simulation based on CA." *Journal of Korean Society of Transportation*, Vol. 21, No. 1, pp. 91-102 (in Korean).  
 Chang, H., Park, D., Lee, S., Lee, H. and Baek, S. "Dynamic multi-interval bus travel time prediction using bus transit data." *Transportmetrica*, Vol. 6, No. 1, pp. 19-38.  
 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.  
 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.  
 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.  
 Stathopoulos and Karlaftis (2003). "A multivariate state-space approach for urban traffic flow modeling and prediction." *Transportation Research Part C*, Vol. 11, No. 2, pp. 121-135.  
 Vlahogianni, E. I., Karlaftis, M. G. and Golias, J. C. (2005). "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach." *Transportation Research Part C*, Vol. 13, No. 3, pp. 211-234.  
 Vlahogianni, E. I., Karlaftis, M. G. and Golias, J. C. (2014). "Short-term traffic forecasting: Where we are and where we're going." *Transportation Research Part C*, Vol. 43, No. 1, pp. 3-19.  
 Yoon, B. and Chang, H. (2014). "Potentialities of data-driven non-parametric regression in urban signalized traffic flow forecasting." *Journal of Transportation Engineering*, Online.