

## 지식발견 기반의 고속도로 영업소 분할 교통수요 예측

안병탁\* · 윤병조\*\*

Ahn, Byeong-Tak\*, Yoon, Byoung-Jo\*\*

### Prediction of Divided Traffic Demands Based on Knowledge Discovery at Expressway Toll Plaza

#### ABSTRACT

The tollbooths of a main motorway toll plaza are usually operated proactively responding to the variations of traffic demands of two-type vehicles, i.e. cars and the other (heavy) vehicles, respectively. In this vein, it is one of key elements to forecast accurate traffic volumes for the two vehicle types in advanced tollgate operation. Unfortunately, it is not easy for existing univariate short-term prediction techniques to simultaneously generate the two-vehicle-type traffic demands in literature. These practical and academic backgrounds make it one of attractive research topics in Intelligent Transportation System (ITS) forecasting area to forecast the future traffic volumes of the two-type vehicles at an acceptable level of accuracy. In order to address the shortcomings of univariate short-term prediction techniques, a Multiple In-and-Out (MIO) forecasting model to simultaneously generate the two-type traffic volumes is introduced in this article. The MIO model based on a non-parametric approach is devised under the on-line access conditions of large-scale historical data. In a feasible test with actual data, the proposed model outperformed Kalman filtering, one of a widely-used univariate models, in terms of prediction accuracy in spite of multivariate prediction scheme.

**Key words** : K-nearest neighbors, Non-parametric approach, Multivariate in-and-out model, Forecasting divided traffic demands

#### 초록

고속도로의 주요 영업소 톨부스는 일반적으로 2개 차종(경차포함 승용차, 승용차 이외의 중차량)의 교통수요 변동에 따른 사전 대응방식으로 각 차종에 대하여 운영된다. 이러한 의미에서 2개 차종에 대한 정확한 교통량 예측은 영업소의 첨단 운영에 있어 주요 요소 중 하나이다. 유감스럽게도, 기존 연구로 보고된 현행의 일변량 단기 예측 기법들을 이용하여 2개 차종의 교통량을 동시에 예측하기는 용이하지 않다. 이러한 실용적 학술적 배경으로 인해 수용 가능한 정확도의 수준에서 2개 차종의 장래 교통량 예측은 ITS 예측 분야의 매력적인 연구 주제 중 하나이다. 따라서 본 연구에서는 기존의 일변량 단기 예측기법의 단점을 극복함과 더불어 2개 차종의 교통량을 동시에 예측하기 위한 다중 입출력(Multiple In-and-Out, MIO) 모형을 제시하도록 한다. 제안된 MIO 모형은 대용량 이력자료의 실시간 이용이 가능한 자료 환경에서 비모수 접근법을 기반으로 개발되었다. 실제 자료를 이용한 적용가능 실험에서, 개발모형은 다변량 예측 수준에도 불구하고 폭 넓게 이용되는 일변량 예측모형 중 하나인 Kalman filtering에 비하여 예측 정확도 측면에서 우수하게 나타났다.

**검색어** : 최근인 이웃, 비모수 접근법, 다중 입/출력 모형, 분할 교통수요 예측

\* 인천대학교 대학원 석사과정 (Incheon University · anta19@dodamenc.co.kr)

\*\* 정회원 · 교신저자 · 인천대학교 도시과학대학 도시공학과 부교수, 공학박사 (Corresponding Author · Incheon University · bjoon63@inu.ac.kr)

Received November 3, 2015/ revised November 10, 2015/ accepted December 8, 2015

## 1. 연구의 배경 및 목적

고속도로 영업소(tollgate, 이하 TG)의 효율적인 운영/관리는 첨단 고속도로 운영에 있어 주요한 요소 중 하나이다. 그리고 차량수요 변동에 대한 사전 대응방식의 탄력적인 톨부스(tollbooth, 이하 TB) 개폐운영은 차량의 대기시간 감소 및 영업소의 운영비용 절감에 직접적인 영향을 미치기 때문에 고속도로 TG의 주요한 운영요소이다. TB 개폐운영은 통상적으로 (경차를 포함한) 승용차와 (승용차 이외의) 중차량으로 구분되어 운영된다. 이러한 사전대응 TB 개폐운영에 있어 핵심 정보는 승용차와 중차량의 단기(short-term) 장래 교통량이다. 따라서 승용차와 중차량의 교통량을 수용 가능한 오차의 범위 내에서 동시에 예측할 수 있는 단기 예측기법은 동적 TB 개폐운영에 있어 핵심 요소기술 중 하나이다.

예측 문제는 장래 상태의 불확실성을 최소화하기 위한 문제라 할 수 있으며, 예측 문제를 해결하기 위한 방법으로는 주로 예측모형이 이용되고 있다(Chang et al., 2012). 즉, 예측모형의 최종 목표는 장래 상태의 불확실성을 최대한 극복함으로써 장래 상태를 추정하기 위함이다. 기존의 단기 예측모형은 (소량의 이력) 자료를 이용하여 모형의 구조 및 파라미터 값을 정산하기도 하지만 많은 경우에 있어 현행(current) 자료의 상태변화를 지식을 이용하여 장래 상태를 추정한다(Yoon and Chang, 2014). 그러나 교통량과 같이 국부적 변동이 심한 시계열 자료의 경우, 현행 상태변동 만을 고려하는 예측모형은 예측 실패에 봉착하게 된다(Vlahogianni et al., 2005). 이는 인위적인 수학적 모델링에 의한 현행 상태의 상태진화(state evolution)에 대한 해석만으로 장래 상태의 불확실성을 해결하기에는 한계가 있음(Chang et al., 2012)을 의미한다. 이러한 인류가 발견한 수학적 지식의 한계와 예측 대상의 상태진화에 대한 지식의 한계를 극복할 수 있는 대안 중 하나가 방대한 양의 데이터 자체에 내재된 지식을 탐색/발견/이용하는 지식발견(knowledge discovery) 접근법이다. 추가로, IT기술의 급속한 발달과 더불어 최근에 조성되고 있는 대용량 자료에 대한 온라인 접근성 향상은 지식발견 접근법에 대한 적용 가능성과 더불어 대용량 자료 환경에 적합한 예측 방법론을 요구하고 있다(Yoon and Chang, 2014).

이상에서 언급한 동적 TB 개폐운영의 효과 증대, 일변량 기반 현행 단기 예측기법의 한계, 그리고 대용량 자료기반 예측모형의 시대적 요구사항을 고려하여, 본 연구에서는 대용량 이력자료를 기반으로 승용차와 중차량의 장래 교통량을 동시에 예측하기 위한 다변량 입-출력(MIO, multiple input and output) 모형을 제시하도록 한다. MIO 모형은 최근 이웃( $k$ -nearest neighbors, 이하 KNN) 기반의 데이터 구동형(data-driven) 비모수회귀(non-parametric regression, 이하 NPR) 접근법을 이용하여 대용량의 시계열 과거자

료에 내재된 교통량 상태변화에 대한 경험적 지식을 활용하도록 개발되었다. 그리고 개발된 KNN-NPR기반 MIO 모형은 첨단자료 관리시스템에 탑재되어 구동되도록 설계되었으며, 모형의 구동에 필요한 입력자료는 실시간 자료와 해당 시스템에서 유지 관리되고 있는 방대한 양의 이력자료이다.

개발모형은 한국도로공사의 요금징수시스템인 Toll Collection System (TCS)를 통해 수집된 실제 승용차와 중차량 교통량 자료를 이용하여 적용 가능성을 평가되었다. 비교 모형으로 단일시간대 일변량 예측기법으로 널리 이용되는 칼만필터링(Kalman filtering, 이하 KF)기법과 비교하였으며, 개발모형의 성능은 예측 정확도 측면에서 KF보다 매우 우수한 결과를 보였다. 추가로 개발된 MIO 예측모형은 보다 사전대응적인 TB 개폐운영 및 고속도로 교통류 제어전략에 적용이 가능할 것으로 기대된다.

## 2. 기존연구 고찰

본 연구는 고속도로 TG에서 수집/관리되는 승용차와 중차량의 교통량 자료를 활용하여 TG의 장래 승용차와 중차량 교통량을 예측하기 위함이다. 따라서 기존연구 고찰은 NPR기반 예측 방법론과 교통량 예측에 중점을 두어 수행하도록 하며, 교통 분야에서 포괄적인 예측관련 연구에 관한 고찰연구(Vlahogianni et al., 2004; Vlahogianni et al., 2014; Mori et al., 2015)를 참고하기 바란다.

한 시스템의 상태진화 과정 즉, 상태 재구성(state reconstruction)은 상태이론으로 설명되며, 상태이론은 확률적(stochastic) 상태이론과 초기 결정론적(initial deterministic) 즉 혼돈(chaos) 상태이론으로 구분될 수 있다. 최근접  $k$ -이웃 즉, KNN은 초기 결정론적 상태이론에 기초하며, 모델링에 있어 확률적 상태이론에 기초하여 통계적 지식에 기초하는 모수회귀(parametric regression, 이하 PR)과 달리 구조적으로 모수에 의존하기 않기 때문에 비모수회귀로 불린다(Oswald et al., 2000).

NPR은 모수회귀의 통계적 분포를 가정하지 않는 장점(Oswald et al., 2000)이 있으며, 알려지지 않은 변수를 포함하는 입/출력 변수간의 비선형적 상태진화 관계의 구성에 있어 강력한 기능이 있다(Yoon, 2011; Yoon and Chang, 2014). KNN-NPR은 이력자료에 내재된 과거의 경험적 지식을 기반으로 상태를 재구성한다. 따라서 혼돈(chaotic) 또는 혼재된(Mixed) 상태에 대한 비선형 시계열 예측문제의 경우, NPR를 이용하면 모형의 구조화가 용이하다(Robinson, 1983; Mulhern and Caprara, 1994). 따라서 NPR은 무질서 상태에서 통계기반의 시계열 예측기법보다 장점을 갖는다(Disbro and Frame, 1989; Altman, 1992; Mulhern and Caprara, 1994). 이는 NPR 자체가 이론적으로 초기 결정론적 상태이론에

근거하고 있기 때문이다(Yoon and Chang, 2014).

NPR기반의 교통량 예측에 관한 연구는 고속도로 교통류인 연속류와 신호로 제어되는 교통류인 단속류로 구분하여 시대별로 고찰하면 다음과 같다. 그리고 단속류 교통량 예측(Yoon and Chang, 2014)에 관한 연구를 제외하면 대부분의 연구는 연속류 교통량의 예측에 집중되어있다. Davis and Nihan (1991)은 단기 교통량의 예측에 있어 KNN을 소개하였으며, 점근선적으로(asymptotically) 최소오차에 수렴함을 보였다. 그리고 대용량 이력자료에 내재된 패턴을 다양화할 경우, 예측력은 향상될 수 있음을 언급하였다. Smith and Demetsky (1997)은 KNN기반 예측을 단일 시간대에서 다중시간대로 확장하였다. 그리고 그들의 후속연구(Smith and Demetsky, 1997)에서 KNN은 인공신경망에 비하여 예측력이 우수함을 보였다. 1900년대 NPR을 이용한 연구는 교통량 자료의 수집 정확도와 대용량 이력자료의 확보 어려움으로 인하여 적용 가능성을 제시하는 수준이었다.

2000년대 연구는 방대한 이력자료를 이용한 비교연구, 실제 적용사례, 그리고 다변량 모형으로 확장되었다. Smith et al. (2002)은 대표적인 시계열분석 모형중 하나인 계절 ARIMA모형과 NPR를 비교하였으며, 교통량의 상태진화는 결정론적 행태보다 확률적 행태에 가깝다고 지적하였다. 그러나 두 접근법간 수행성능은 차이가 없으며, 향후 NPR의 잠재력을 언급하였다. 그리고 NPR의 단점 중 하나였던 방대한 양의 이력자료 탐색시간은 실시간보다 매우 빠른 속도까지 향상되었다(Smith and Oswald, 2003). Clark (2003)은 일변량 NPR을 다변량 (교통량, 속도, 밀도) NPR 모형으로 확장하였다.

2010 이후의 연구는 대용량 자료를 이용한 다중시간대 예측, NPR모형의 개선, 단속류 교통량 예측으로 확장되었다. Chang et al. (2012a)는 다중시간대로 확장된 일변량 NPR을 제시하였으며, Chang et al. (2012b)는 다변량 다중시간대 NPR모형을 제안하였다. Chang과 동료들의 연구결과는 Smith et al. (2002)의 결과와 달리 NPR의 성능이 계절 ARIMA모형보다 우수함을 보였으며, “교통량 시스템의 상태진화는 확률적이라기보다 혼돈 또는 혼제에 가깝다.”는 간접적 증거라고 주장하였다. 이후 Yoon and Chang (2014)는 상태변동이 극심한 신호교차로 교통량의 예측에 NPR를 적용하였다. 비교모형으로 적용된 확률적 상태진화기반 모형인 KF와 계절 ARIMA는 예측실패를 피할 수 없는 반면, NPR은 상태의 방향성과 변량을 수용 가능한 오차수준에서 예측할 수 있음을 보였다. 따라서 혼돈 또는 혼제 상태의 교통량 예측에 있어 NPR은 강력한 대안 중 하나라고 언급하였다. Zheng and Su (2014)는 수정된 선형 NPR과 KF모형을 제시하였으며, 두 모형간의 성능차이는 크지 않은 것으로 나타났다. 그러나 충분한 이력자료를 활용하지 않았으며, 15분 지수평활화를 이용하여 가공된 교통량

자료를 실험 자료로 적용하였기 때문에 시계열의 동적 특성이 희석되었다. 따라서 NPR과 KF의 성능차이는 명확하지 않았다. 추가로 Chang et al. (2010)과 Yoon (2011)은 통행시간 자료의 수집에 필연적으로 발생하는 다중시간대 시간 처짐(time-lag)을 극복하기 위한 NPR모형을 개발/적용하였으며, 다중시간대 예측 임에도 불구하고 수용 가능한 오차 내에서 장래 통행시간을 추정하였다.

기존연구 고찰결과, 교통량의 시간적 진화과정은 시계열분석 접근법 연구들의 주장인 확률적 상태보다 혼돈 상태(Altman, 1992; Mulhern and Caprara, 1994; Vlahogianni et al., 2006; Yoon and Chang, 2014)에 가깝다는 직/간접적 주장이 우세하였다. 그리고 NPR은 혼돈이론을 기반으로 하는 강력한 접근법이며, NPR에서 KNN은 데이터에 내재된 패턴인식을 기반으로 혼돈상태의 시공간적 진화과정 해석에 있어 확률상태기반 접근법에 비하여 장점이 있다(Karlsson and Yakowitz, 1987; Davis and Nihan, 1991). 따라서 다양한 패턴이 내재된 대용량 자료를 이용한 KNN-NPR 접근법은 교통량 예측에 있어 모수기반 접근법에 비하여 우수한 성능을 보이는 것으로 보고되고 있다.

### 3. 방법론

본 연구의 KNN-NPR기반 다차중 교통량 예측 방법론은 3개 요소(상태벡터, 결정자, 예측함수)로 구성되며, 이력자료 탐색기반의 KNN-NPR 예측 알고리즘으로 구현된다. 따라서 각 구성요소에 대한 기술 후, 본 연구의 예측 알고리즘을 설명하도록 한다.

#### 3.1 상태벡터

시공간적으로 연속적인 상태의 변화를 이산적인 시계열자료로 구성하기 위하여 상태벡터(state vector)를 이용하며, 본 연구에서 상태벡터는 현행(current), 입력(input), 그리고 출력(output) 상태벡터로 구성된다. 현행 상태벡터( $x_c$ )는 예측시점(prediction point, 이하  $pt$ )의 이전 시간대( $t$ )에서 주어진 lag의 개수( $l$ ) 즉, 시계열자료의 시간대( $t-l+1$ )까지 시간대별 차중( $vt$ )별 교통량( $q^v$ , 대/시간대 길이)을 이용하여 Eq. (1)과 같이 그리고  $vt = [1, 2, \dots, v]$ 로 정의한다. 따라서 현행 상태벡터의 차원은  $v \times l$ 이다. 이력자료에 포함되어 있는  $x_c$ 와 유사한 상태라고 추정되는  $n$ 개의 상태벡터( $x_n$ ) 중  $j$  번째 상태벡터를 Eq. (2)와 같이 입력 상태벡터( $x_j$ ,  $j \in n$ )로 정의하며, 상태벡터의 차원은  $x_c$ 와 동일하다.

$x_c$ 와  $x_j$ 에 대한 출력 상태벡터( $o_j$ )는  $x_c$ 와  $x_j$ 간의 상태간 거리( $d_j$ )와  $x_j$ 에 해당하는 시간대( $t+1$ )의  $vt$ 별 과거 교통량( $q_j^{vt}(t+1)$ )로 정의하며, Eq. (3)과 같다. 그리고  $x_j$  중  $x_c$ 와 가장 유사한  $k$ 개의  $o_j$  즉, KNN의 출력 상태벡터 집합을 예측시점( $t$ )에서  $x_c$ 에

대한 출력 상태벡터( $o$ )로 정의하며, Eq. (4)와 같다. 추가로  $o$ 는 차종별 교통량을 예측하기 위한 예측함수의 입력 값으로 이용된다.

$$x_c = [q_c^{vt}(t), q_c^{vt}(t-1), \dots, q_c^{vt}(t-l+1)] \quad (1)$$

$$= \begin{bmatrix} [q_c^1(t), q_c^1(t-1), \dots, q_c^1(t-l+1)] \\ [q_c^2(t), q_c^2(t-1), \dots, q_c^2(t-l+1)] \\ \dots \\ [q_c^v(t), q_c^v(t-1), \dots, q_c^v(t-l+1)] \end{bmatrix}$$

$$x_j = [q_j^{vt}(t), q_j^{vt}(t-1), \dots, q_j^{vt}(t-l+1)] \quad (2)$$

$$= \begin{bmatrix} [q_j^1(t), q_j^1(t-1), \dots, q_j^1(t-l+1)] \\ [q_j^2(t), q_j^2(t-1), \dots, q_j^2(t-l+1)] \\ \dots \\ [q_j^v(t), q_j^v(t-1), \dots, q_j^v(t-l+1)] \end{bmatrix}$$

$$o_j = [d_j, q_j^{vt}(t+1)] \quad (3)$$

$$= [d_j, [q_j^1(t+1), q_j^2(t+1), \dots, q_j^v(t+1)]]$$

$$o = [o_1, o_2, \dots, o_k] \quad (4)$$

$$= [[d_1, q_1^{vt}(t+1)], [d_2, q_2^{vt}(t+1)], \dots, [d_k, q_k^{vt}(t+1)]]$$

### 3.2 결정자

이력자료에서  $x_c$ 와 유사한  $x_j$ 를 탐색하여  $o$ 를 구축하기 위해서는 일반적으로  $x_c$ 와  $x_j$ 의 상태 유사성을 이용하게 된다. 따라서 유사성을 결정하기 위한 결정자(determinant)가 정의되어야 한다. 결정자로는 상태간 거리인 Minkowski distance (이하, MD), 통계적 기법으로 쌍체검증과 상관계수 등이 이용된다.

본 연구에서는 결정자로 MD를 적용하도록 하며, MD는  $L_m$  거리로 널리 알려져 있다.  $L_m$  거리는  $\mathbb{R}^p$ 에서 Eq. (5)와 같으며,  $m = \{1, 2, \dots, \infty\}$ 에 따라 Manhattan, Euclidean, ..., Maximum 거리로 불린다. 본 연구의 경우,  $m = 2$ 인 Euclidean 거리를 이용하도록 하며,  $x_c$ 와  $x_j$ 의 상태간 거리( $d_j$ )는 Eq. (6)과 같다. 추가로 Euclidean 상태거리는 국부적 상태에 민감하기 때문에 국부적 상태를 간과하지 않으며 교통량 시계열자료와 같이 상태의 변동이 심한 경우에 적합하다(Chang et al., 2012a).

$$L_m = \left( \sum_{i=1}^p |x_i - y_i|^m \right)^{1/m} \quad (5)$$

$$d_j = (x_c, x_j) \quad (6)$$

$$= \left[ \sum_{\alpha=1}^v \sum_{\beta=0}^{l-1} |q_c^\alpha(t-\beta) - q_j^\alpha(t-\beta)|^2 \right]^{1/2}$$

### 3.3 예측함수

KNN-NPR 알고리즘의 이력자료 탐색과정을 통해  $o$ 가 구축되면, 차종별 장래 교통량은 예측함수를 이용하여 추정된다. 예측함수는 산술 평균(straight average, 이하 SA)과 (MD를 이용하는 경우)  $1/d_i$ 의 가중 평균(weighted average by inversed distance, 이하 WAID)이 주로 이용된다. SA는 장래 상태에 대한 의사결정권이 동일한 방법으로서 수학적으로 정확한 방법이지만, 상태간 유사성을 반영할 수 없는 단점이 있다. 반면, WAID는 경험적 방법이지만, 보다 더 유사한 상태 즉, 보다 더 유사한 경험지식에 대한 의사결정권을 고려할 수 있다는 장점이 있다. 그리고 기존 연구의 고찰 결과, SA 계열의 예측함수에 비하여 WAID 계열이 우수한 결과를 도출하였다(Smith et al., 2002; Chang et al., 2010, 2012a,b; Yoon and Chang, 2014). 따라서 본 연구는 WAID를 이용하여 시점( $t+1$ )에서  $vt$ 별 교통량  $q^{vt}(t+1)$ 을 예측하도록 하며, Eq. (7)과 같이  $o$ 의 시점( $t+1$ )에서  $vt$ 별 과거 교통량  $q_i^{vt}(t+1)$ ,  $i \in k$ 은  $1/d_i$ 로 가중 평균된다.

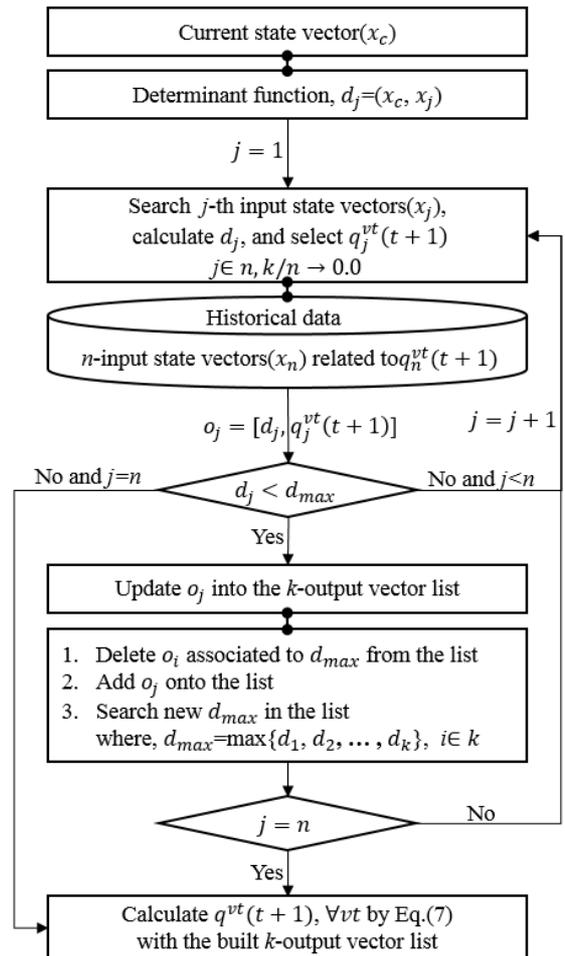


Fig. 1. Prediction Algorithm for Multiple Traffic Volumes

$$q^{it}(t+1) = \frac{\sum_{i=1}^k [q_i^{it}(t+1) \times d_i^{-1}]}{\sum_{i=1}^k d_i^{-1}}, d_i > 0, \forall vt \quad (7)$$

### 3.4 KNN-NPR 예측 알고리즘

데이터 주도형 KNN-NPR 모형의 3개 요소가 정의되면, KNN-NPR 예측 알고리즘으로 구현된다. 본 연구의 KNN-NPR 알고리즘은 입력자료 탐색과정, 출력 상태벡터( $o$ ) 구축과정, 예측과정으로 구성된다. 입력자료 탐색과정은 주어진  $x_c$ 와 결정자인 Eq. (6)을 이용하여 입력자료에 포함되어 있는 모든  $x_j, j \in n$ 에 대한  $o_j$ 를 추출하는 단계이다. 출력 상태벡터( $o$ ) 구축과정은 입력자료 탐색과정을 통해 추출된  $o_j$ , 즉  $o$ 의 구성원에 대한 후보들( $k$ 개의  $o_i$ 로 구성되는)  $o$ 에 상태간 거리를 기준으로 갱신하는 과정이다. 따라서 모든  $x_j$ 에 대한 입력자료 탐색과정과 출력 상태벡터 갱신과정이 완료되며,  $x_c$ 와 가장 유사한  $k$ 개의  $x_j$ 에 대한  $o$ 가 구축된다. 예측과정은 구축된  $o$ 와 Eq. (7)을 이용하여 차종별 장래 교통량을 추정하는 과정이다.

## 4. 개발모형의 적용 및 평가

### 4.1 실험 설계

KNN-NPR모형은 입력자료에 내재된 상태진화의 패턴 다양성에 의존한다. 따라서 본 연구에서는 경부고속도로 서울TG의 TCS 교통량 자료를 이용하여, 2014년의 365일 동안 (경차 포함) 승용차와 (승용차 이외의) 중차량 교통량 자료를 구축하였으며, 집계자료의 시간대 길이는 동적 TB운영에 있어 적정 대응시간을 고려하여 15분으로 설정하였다. 따라서 입력자료의 크기는 35,040=[365(일) \* 96(시퀀스/일)]이다. 그리고 구축된 입력자료 중 12월 26일의 96시간대는 개발모형의 평가에 적용하였다. 그리고 평가대상 요일은 금요일로서 25일 성탄절과 27일 토요일의 징검다리 연휴와 관련이 있기 때문에 반복적 패턴 다양성은 매우 낮으며, 이는 개발모형의 적용평가에 매우 불리한 조건이다.

평가지표는 개발모형의 파라미터 최적화와 예측 성능평가로 구분하여 적용하였다. 평균절대값백분위오차(Mean Absolute Percentage Error(%), 이하 MAPE)는 파라미터 최적화와 예측 성능평가에 적용하며, Eq. (8)과 같다. 여기서,  $x_i, y_i, s$ 는 각각 관측치, 예측치, 표본의 개수이다. 예측 성능평가 지표는 정확도와 신뢰도 지표로 구분하여 적용하였다. 정확도 지표는 MAPE와 Eq. (9)의 절대평균오차(Mean Absolute Error (대), 이하 MAE)를 적용하였으며, 신뢰도 지표는 Eq. (10)의 상대백분위오차(Relative Percentage Error (%), 이하 RPE)의 표준편차인 SDRPE (%)와

차이( $E=x_i - y_i$ )의 표준편차인 SDE (대)를 이용하였다.

$$MAPE = \left( \frac{1}{s} \sum_{i=1}^s \frac{|x_i - y_i|}{x_i} \right) \times 100 \quad (8)$$

$$MAE = \frac{1}{s} \sum_{i=1}^s |x_i - y_i| \quad (9)$$

$$RPE = \left( \frac{x_i - y_i}{x_i} \right) \times 100 \quad (10)$$

비교 모형으로는 KF기법(Kalman, 1960)을 선정하였다. KF모형은 현행 상태를 설명하기 위한 상태공간 방정식을 기반으로 장래 상태를 예측하는 일변량기반 모수적 접근방법이다. 교통의 예측분야에서는 수행속도가 매우 빠르기 때문에 ITS 예측시스템에서 일변량 교통변수(교통량, 속도, 점유율, 통행시간)의 예측의 용도로 널리 이용되고 있다.

## 4.2 평가결과

### 4.2.1 최적 파라미터 값 분석

데이터 구동형 KNN-NPR 모형은 현행 및 입력 상태벡터의 크기를 결정하기 위한 lag의 개수( $l$ )과 현행 상태벡터와 유사한 과거 상태, 즉 입력 상태벡터의 개수인  $k$ 값이 사전에 추정되어야 한다. Fig. 2는  $k$ 와  $l$ 의 280개 시나리오[288=( $k$ 값 1~40)×( $l$ 값 1~7)]에 대한 대상 요일의 96개 시퀀스에 대한 MAPE 변화곡선을 보여주고 있다. 예측오차 곡선은  $k$ 값이 증가함에 따라 최소오차 영역까지 급격히 감소하며, 최소오차 영역을 벗어나면 서서히 오차가 증가하는 오목형태를 보이고 있다. 그리고  $l=2, k=13$ 에서 MAPE는 7.93으로 최소를 보이고 있다. 따라서 적정  $l$ 과  $k$ 값은

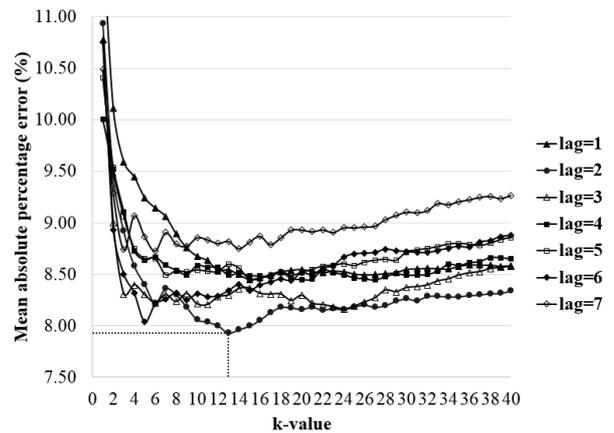


Fig. 2. Analysis of Optimal k-Value and Lag Size( $l$ )

2와 13으로 각각 설정하도록 한다. 그리고  $l=2$ 에서 적정  $k \pm 3$ 의 범위 내에서 오차는 최소오차  $\pm 0.1\%$  내에 수렴하고 있다.

$k$ 값의 증가에 따른 오목 형태의 오차행태는 명확하든 명확하지 않든 상태 간에 경계가 증가함을 간접적으로 의미하며, 이는 교통량의 상태진화가 확률적이기 보다는 혼돈 또는 혼재 상태에 가깝다는 증거이기도 하다(Chang et al., 2012; Yoon and Chang, 2014). 추가로 상태진화가 확률적이라면  $k$ 값의 증가에 따라 오차곡선은 최소오차까지 감소한 후 수렴하게 된다. 이러한 경우, NPR의 예측 상태는 PR의 예측상태와 접근선적으로 동일하다(Davis and Nihan, 1991).

4.2.2 결과 분석

예측모형은 이전 시간대의 상태가 장래 상태에도 변하지 않고 유지된다고 가정하는 사후 방식(Reactive method, 이하 RM)에 비하여 최소한도 우수한 성능을 가져야하며, 이는 예측모형의 최소 수행성능 조건이다. 만일 최소 수행성능 조건을 만족하지 못하는 경우, 해당 모형은 예측실패를 피할 수 없으며, 정확도 측면에서 적용 가능성을 충족하지 못 한다.

Table 1은 평가지표를 이용한 종합 분석결과를 보여주고 있으며, KNN-NPR 모형은 다변량 예측임에도 불구하고 정확도와 신뢰도 지표 모두에서 일변량 예측기법인 KF에 비해 우수한 성능을 보이고 있다. 그리고 개발모형의 KF에 대한 성능향상은 승용차의 경우 최소 13% 이상, 중차량의 경우 53% 이상을 보였다. 특히, KF의 경우 교통량 수준이 낮고 시계열 변동이 심한 중차량의 경우, RM보다 예측성능이 낮은 예측실패를 보이는 것으로 나타났다. 추가로, 평활화가 적용된 KF를 적용할 경우 예측 성능은 다소 증가하겠지만 시계열자체의 동적특성을 상실하게 되며, 이는 Yoon and Chang

(2014) 지적한 바와 같이 시계열변동이 심한 교통량의 예측에는 모수적 접근방법이 적합하지 않을 수 있음을 보여준다.

Figs. 3 and 4는 승용차 교통량에 대한 KF와 KNN-NPR의 예측결과를 시계열 분석(time-series analysis, 이하 TSA)로 보여 주고 있으며, 시계열 자료(time-series data, 이하 TSD)의 변동은 크지 않게 나타나고 있다. 이러한 경우 두 모형의 예측성능은 유사하게 나타난다. 그러나 오전 첨두(a)의 경우, KF는 일정시간 방향성을 상실하면서 과대추정을 보이고 있으며, KNN-NPR의 경우 5%의 오차범위내에서 예측성능을 보이고 있다. 특히 국부적 변동이 지속적으로 나타나는 (b)의 경우, KF는 상태의 방향성을 상실하면서 지그자그 형태의 예측결과를 보이고 있다. 반면, KNN-NPR은 패턴이 존재하는 경우 방향성과 변화량을 정확히 예측하고 있으며, 패턴이 혼재된 경우 접근선적으로 예측오차를 최소화하고 있다.

Figs. 5 and 6은 상대적으로 낮은 중차량 교통량에 대한 KF와 KNN-NPR의 예측결과를 TSA로 보여주고 있으며, TSD의 변동은 승용차에 비하여 크게 나타나고 있다. 두 모형간의 예측력은 극명한 차이를 보이고 있다. 상태변동이 극심한 경우, KF는 장래 상태에 대한 방향성과 변동량에 대한 불확실성을 극복할 수 없다.

따라서 첨두시의 경우(a) 방향성을 방향성과 변동량을 완전히 상실하고 있으며, TSD가 지속적인 변동을 보이는 (b)의 경우 첨두 시에 비하여 일정부분 상태를 유지할 수 있지만 방향성을 상실하였

Table 1. Summary of Analysis Results

VT	Model	Accuracy		Reliability	
		MAPE	MAE	SDRPE	SDE
Car	RM	8.57	57.01	12.77	74.79
	KF	7.67	52.39	11.30	66.53
	KNN-NPR	6.38	45.21	9.43	57.52
	Gain (%)	16.80	13.72	16.54	57.52
Heavy vehicle	RM	16.56	23.86	22.68	33.58
	KF	*20.36	*29.09	*29.20	*41.27
	KNN-NPR	9.49	13.19	13.08	17.72
	Gain (%)	53.40	54.66	55.20	57.20

- 1) VT and KF stands for Vehicle Type and Kalman Filter.
- 2) RM stands for Reactive Method.
- 3) Gain of KNN-NPR is calculated to KF.
- 4) \* means forecasting failure comparing to RM.

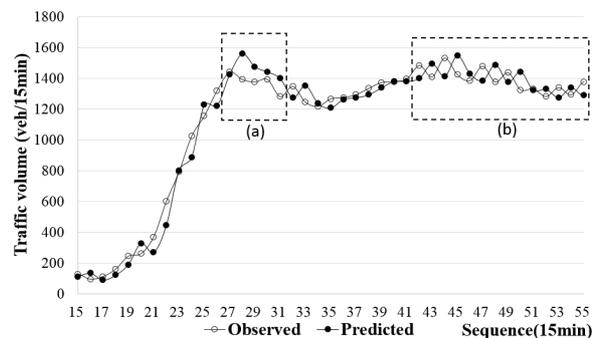


Fig. 3. TSA of KF for Car Volume

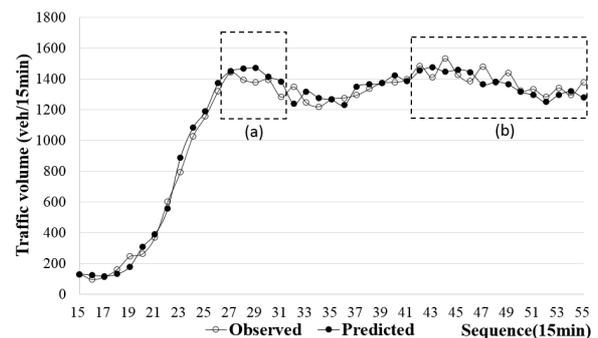


Fig. 4. TSA of KNN-NPR for Car Volume

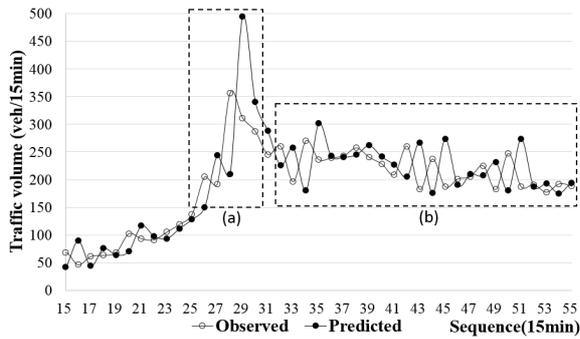


Fig. 5. TSA of KF for Heavy-Vehicle Volume

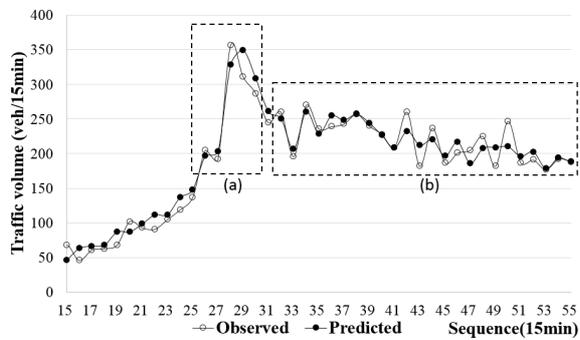


Fig. 6. TSA of KNN-NPR for Heavy-Vehicle Volume

기 때문에 실제 상태와 예측 상태가 지그재그로 어긋나는 현상을 보이면서 예측력이 현저히 저하되고 있다. 이러한 불확실성 극복문제는 현행 자료만을 이용하는 대다수 예측모형의 한계이다(Vlahogianni, et al., 2005; Yoon and Chang, 2014). 반면, KNN-NPR은 첨두시의 경우 급격한 증가와 급격한 감소를 설명하고 있다. 지속적인 변동을 보이는 (b)의 경우, 대부분에 있어 상태의 방향성과 변동량을 설명하고 있으며 일부 방향성을 상실한 경우 평활화 기능을 수행하면서 오차를 최소화하고 있다.

이상의 결과와 같이, 본 연구에서 개발된 KNN-NPR기반의 MIO모형은 TSD의 변동이 극심한 경우, 타 모형에 비하여 우수한 성능을 보이고 있다. 이는 본 연구에서 제안된 MIO모형은 KNN-NPR을 기반으로 개발되었으며, KNN-NPR은 과거 경험 즉, 이력 자료에 내재된 상태변동에 대한 정보를 이용하여 장래 상태의 불확실성 문제는 효과적으로 다룰 수 있기 때문이다. 추가로 KNN-NPR은 초기결정이론을 기초로 혼돈상태를 설명하기 위하여 개발되었다. 따라서 교통량의 상태진화는 비선형적이며, 상태의 진화에 대한 방향성과 변동량은 다양한 패턴의 진화로 구성되는 혼돈상태라고 판단된다. 따라서 교통량의 예측에 이용되는 접근법은 NPR, 인공 신경망과 같은 비선형 모형이 적합하다고 판단된다.

## 5. 결론 및 향후연구

### 5.1 결론

21세기 ITS 정보 수집단을 통해 수집되는 대용량의 이력자료는 DB에 저장된 후 이용되고 있지 않거나 일정시간이 경과하면 시스템의 DB공간을 확보하기 위해 삭제 또는 별도로 저장되고 있다. 그러나 IT기술의 발달과 더불어 대용량 고성능의 첨단자료관리시스템이 도입되고 있다. 이러한 데이터 환경의 변화는 데이터 기반의 전문가 시스템, 사례기반 추론, 데이터 주도형 NPR등의 접근법을 기반으로 하는 다양한 분석방법의 가능성을 증대시키고 있다. 그리고 ITS분야에서 교통량 예측은 동적 TB의 운영과 더불어 교통류 제어의 핵심 입력 자료이다. 따라서 정확한 교통량의 예측은 동적 TB의 운영과 교통류 제어의 효과와 편익을 증가시킨다. 이러한 운영 및 제어에 있어 중차량의 영향은 매우 크다고 할 수 있다.

따라서 본 연구에서는 승용차와 중차량의 교통량을 동시에 예측하는 다변량 모형을 KNN-NPR을 기반으로 개발하였으며, 개발모형의 수행성능은 다변량 예측임에도 불구하고 기존 모형에 비하여 매우 우수하게 나타났다. 이러한 결과는 개발모형이 대용량 이력자료에 포함되어 있는 다양한 상태변화에 대한 패턴과 알려지지 않은 변수에 대한 내재적 정보를 이용하여 장래 상태의 불확실성을 극복하도록 설계되었기 때문이다. 따라서 본 연구에서 개발된 KNN-NPR기반 다변량 예측모형은 교통량뿐만 아니라 기타 교통변수(속도, 통행시간, 점유율)의 예측에도 직접적인 적용이 가능하다. 따라서 개발모형은 보다 사전 대응적인 교통류 상태 모니터링과 장래 교통정보 예측에 적용이 가능할 것으로 사료된다.

### 5.2 향후연구

본 연구에서 개발된 모형의 적용성을 향상시키기 위해서는 다음의 향후 연구가 필요하다. 첫째, 개발 모형은 대용량 이력자료를 기반으로 한다. 따라서 신규도로 등에 적용하기 위해서는 소량의 이력자료로 적용이 가능한 모형으로 전환할 필요가 있다. 둘째, 교통량 이외에 통행시간, 통행속도 등의 교통변수에 대한 적용연구가 필요하다. 셋째, 본 연구에서는 2개 차종(승용차와 중차량)을 적용하였다. 그러나 보다 정교한 교통 제어 및 운영을 위해서는 승용차, 버스, 대/중/소 트럭으로 예측모형을 적용할 필요가 있다. 넷째, 본 연구에서 개발된 방법론은 대용량 이력자료를 기반으로 구동된다. 따라서 신규 TG의 개통과 같은 대량의 이력자료를 확보하지 못 하는 자료의 조건에서 적용에 한계가 있다. 따라서 인접 TG의 이력자료를 활용하는 연구가 필요하다. 마지막으로 개발모형은 단일 시간대 예측에 적합하도록 개발되었다. 따라서 보다 동적인 제어 및 운영에 적용하기 위해서는 다중시간대 예측으로 모형을 개선해야 한다.

## References

- Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression." *The American Statistician*, Vol. 46, No. 3, pp. 175-185.
- Chang, H. H., Lee, Y. I., Baek, S. K. and Yoon, B. J. (2012a). "Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences." *IET Intelligent Transport Systems*, IEEE, Vol. 6, No. 3, pp. 292-305.
- Chang, H. H., Park, D. J., Lee, S. J., Lee, H. S. and Baek, S. K. (2010). "Dynamic multi-interval bus travel time prediction using bus transit data." *Transportmetrica. Taylor & Francis*, 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*, NRC Research Press, Vol. 39, pp. 448-459.
- Clark, S. (2003). "Traffic predicting using multivariate nonparametric regression." *Journal of Transportation Engineering*, ASCE, Vol. 129, No. 2, pp. 161-168.
- Davis, G. and Nihan, N. (1991). "Nonparametric regression and short-term freeway traffic forecasting." *Journal of Transportation Engineering*, ASCE, Vol. 117, No. 2, pp. 178-188.
- Disbro, J. E. and Frame, M. (1989). "Traffic flow theory and chaotic behavior." *New York State Department of Transportation Report FHWA/NY/SR-98/91*, New York.
- Kalman, R. E. (1960). "A new approach to linear filtering and prediction problems." *Journal of Basic Engineering*, ASME, Vol. 82, No. 1, pp. 35-45.
- Karlsson, M. and Yakowitz, S. (1987). "Rainfall-runoff forecasting methods, old and new." *Stochastic Hydrology and Hydraulics*, Springer, Vol. 1, No. 4, pp. 303-318.
- Mori, U., Mendiburu, A., Alvarez, M. and Locano, J. A. (2015). "A review of travel time estimation and forecasting for advanced traveler information systems." *Transportmetrica A: Transport Science*, Taylor & Francis, Vol. 11, No. 2, pp. 119-157.
- Mulhern, F. J. and Caprara, R. J. (1994). "A nearest neighbor model for forecasting market response." *International Journal of Forecasting*, Elsevier, 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.
- Robinson, P. (1983). "Nonparametric estimators for time series." *Journal of Time Series Analysis*, Wiley, Vol. 4, No. 3, pp. 185-207.
- Smith, B. L. and Demetsky, M. J. (1996). "Multiple-interval freeway traffic flow forecasting." *Transportation Research Record*, TRB, Issue 1554, pp. 136-141.
- Smith, B. L. and Demetsky, M. J. (1997). "Traffic flow forecasting: Comparison of Modeling Approaches." *Journal of Transportation Engineering*, ASCE, Vol. 123, No. 4, pp. 261-266.
- Smith, B. L. and Oswald, R. K. (2003). "Meeting real time traffic flow forecasting requirements with imprecise computations." *Computer-Aided Civil Infrastructure Engineering*, Wiley, Vol. 18, No. 3, pp. 201-213.
- 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*, Elsevier, Vol. 10, No. 4, pp. 303-321.
- Turochy, R. E. (2006). "Enhancing short-term traffic forecasting with traffic condition information." *Journal of Transportation Engineering*, ASCE, Vol. 132, No. 6, pp. 469-474.
- Vlahogianni, E. I., Golias, J. C. and Karlaftis, M. G. (2004). "Short-term traffic forecasting: overview of objectives and methods." *Transport Reviews*, Taylor & Francis, Vol. 24, No. 5, pp. 533-557.
- 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*, Elsevier, Vol. 13, No. 3, pp. 211-234.
- Vlahogianni, E. I., Karlaftis, M. G. and Golias, J. C. (2006). "Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume." *Transportation Research Part C*, Elsevier, Vol. 14, No. 5, pp. 351-367.
- 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*, Elsevier, Vol. 43, Part 1, pp. 3-19.
- Yoon, B. J. (2011). "A study on the development of a technique to predict missing travel speed collected by taxi probe." *Journal of the Korean Society of Civil Engineers*, KSCE, Vol. 31, No. 1D, pp. 43-50 (in Korea).
- Yoon, B. J. and Chang, H. H. (2014). "Potentialities of data-driven nonparametric regression in urban signalized traffic flow forecasting." *Journal of Transportation Engineering*, ASCE, Vol. 140, No. 2. DOI: 10.1061/(ASCE)TE.1943-5436.0000662.
- Zheng, Z. and Su, D. (2014). "Short-term traffic forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm." *Transportation Research Part C*, Elsevier, Vol. 43, Part 1, pp. 143-157.