

단기 통행시간예측 모형 개발에 관한 연구

The study of Estimation model for the short-term travel time prediction

이 승 재* 김 범 일** 권 혁*
(Seung-jae, LEE) (Beom-il, KIM) (Hyug, Kwon)

요 약

최근 몇 년간 도시교통문제의 해결책으로 부각되어온 지능형교통체계(ITS : Intelligent Transport System)의 한 분야로 첨단여행자 정보체계(ATIS : Advanced Travellers' Information System)는 자동차에 장착된 항법장치(CNS)를 통해 운전자에게 원하는 목적지까지 최적경로를 제공하거나 경로에 대한 통행시간 정보를 제공 또는 예측해 주는 시스템이다. 본 연구에서는 이러한 최적경로 제공이나 통행시간 예측에 있어 좀 더 효율적인 통행시간 예측모형을 개발하고자 하였다. 현재까지의 통행시간 예측은 운전자가 통행을 시작할 때의 교통상황에 대한 정보이기 때문에 운전 중에 달라지는 교통상황을 반영할 수 없어 이로 인해 운전자가 경험하는 통행시간과 큰 차이를 발생시킬 수 있다. 본 연구에서는 이러한 불합리적인 예측시스템을 개선시킬 수 있는 예측된(predicted) 통행시간 예측 모형을 개발하고자 하였다. 이를 위해 우선 통행시간 예측모형을 특정링크에 적용시켜 모형들의 예측치와 실제 통행시간을 비교하여 교통량 흐름 패턴에 따라 어느 모형이 적합한지, 또 예측시간이 달라짐에 따라 모형들의 적합도와 침투와 비침투시 예측시간 간격에 따라 예측치와 실측치의 오차를 알아보았다, 이를 통해 선정된 확률과정 모형과 칼만 필터링 예측모형을 서울시의 4개축에 대해서 다시 적용해 보았다. 그 결과 단기통행시간 예측에 있어서는 칼만필터링모형이, 장기 통행시간 예측에 있어서는 확률과정 모형이 통행시간 예측에 있어 우수한 모형임을 밝혀냈다. 마지막으로 서울시 28개 교통축의 5분 후 통행시간 예측에 칼만필터링 모형을 이용하여 오차분석을 적용하여 보았다. 그 결과 칼만필터링 모형이 신뢰할 만한 오차를 보였다.

Abstract

The study of Estimation model for the short-term travel time prediction.

There is a different solution which has predicted the link travel time to solve this problem. By using this solution, the link travel time is predicted based on link conditions from time to time. The predicated link travel time is used to search the shortest path.

Before providing a dynamic shortest path finding, the prediction model should be verified. To verify the prediction model, three models such as Kalman filtering, Stochastic Process, ARIMA. The ARIMA model should adjust optimal parameters according to the traffic conditions. It requires a frequent adjustment process of finding optimal parameters. As a result of these characteristics, It is difficult to use the ARIMA model as a prediction. Kalman Filtering model has a distinguished prediction capability. It is due to the modification of travel time predictive errors in the gaining matrix. As a result of these characteristics, the Kalman Filtering model is likely to have a non-accumulative errors in prediction. Stochastic Process model uses the historical patterns of travel time conditions on links. It is favorably comparable with the other models in the sense of the recurrent travel time condition prediction.

* 회 원 : 서울시립대학교 교통공학과 교수

** 회 원 : 서울시립대학교 교통공학과 석사과정

*** 회 원 : 서울시립대학교 교통공학과 석사과정

† 논문접수일 : 2004년 1월 29일

As a result, for the travel time estimation, Kalman filtering model is the better estimation model for the short-term estimation, stochastic process is the better for the long-term estimation.

Key Words : 단기통행시간예측, 모형별 예측통행시간 비교, 칼만필터링, ARIMA, 확률모형

I. 서론

ITS의 한 분야로서 첨단여행자정보체계(ATIS)는 운전자에게 원하는 목적지까지의 최단경로에 대한 정보를 제공하거나 현재의 교통상황을 반영하여 경로의 통행시간을 인터넷 등을 통하여 시시각각 제공하는 시스템이다. 그러나 지금까지의 이러한 운전자 정보는 실시간(real-time)을 기반으로 한 정보이기 때문에 운전하는 동안 바뀌는 교통상황을 제대로 고려하지 못했다. 이러한 방법은 문제가 된다. 예를 들어 A라는 기점에서 B를 경유해 C라는 곳을 간다고 가정하자. 실시간 분석을 지금 시간을 기준으로 A에서 B까지 15분, B에서 C까지 10분이라는 계산 값을 얻어낸다면 25분이라는 추정시간을 알려줄 것이다. 하지만, B지점에 도달했을 때(출발 10분 후) B에서 C지점까지의 통행시간은 10분이 아닐 수 있다. 처음 출발할 때의 교통량이 시간의 변화에 따라 교통량의 변화로 인해 통행시간이 더 길어지거나 짧아질 수 있기 때문이다. 결국 B에 다다른 시점에서의 교통상황을 고려하지 못하고 15분전의 교통상황을 기준으로 통행시간을 산정 하는 오류가 발생한다.

이런 맥락에서 본 연구에서는 최근의 교통분야 신기술을 이용하여 실시간(real-time) 정보가 아닌 예측된(predicted)교통정보를 제공할 수 있는 교통모형을 개발하고자 한다. '링크'통행시간에 대한 예측은 '경로'통행시간을 예측하기 위해 미리 선행되어야 한다. 본 연구에서는 순간적(instantaneous)이 아닌 실질적(ideal or realistic)인 경로통행시간을 제시하고자 하기 때문이다. 다시 말해, 단순히 운전자가 통행을 시작할 때의 교통상황을 기반으로 경로통행시간을 제시하는 것이 아니라, 운전자가 실제로 경험하게 될 링크별 통행시간을 미리 예측해서 경로통행시간을 제시하고자 한다. 이러한 즉각적인 경

로통행시간과 실질적인 경로통행시간의 차이는 다음의 수식으로 보다 명확하게 설명될 수 있다.

$$C_p^d(t) = c_{a_1}(t) + \dots + c_{a_m}(t + c_{a_1}(t) + c_{a_2}(t + c_{a_1}(t) + \dots + c_{a_{m-1}}(t + c_{a_1}(t) + \dots + c_{a_{m-2}}(t)))$$

$$C_p^n(t) = c_{a_1}(t) + c_{a_2}(t) + \dots + c_{a_m}(t) \quad (1)$$

여기서,

$C_p^d(t)$ 는 시간 t 에 출발하여 경로 p 를 이용할 때의 실질적인 경로 통행시간

$C_p^n(t)$ 는 시간 t 에 출발하여 경로 p 를 이용할 때의 실질적인 경로 통행시간

$c_a(t)$ 는 시간 t 일 때의 링크 a 의 통행시간

위의 식에서 볼 수 있듯이, 현실적 경로통행시간은 운전자가 경로를 이루는 각 링크에 진입하는 시각에서 그 링크통행시간을 계산하여 이들을 합하여 계산한다. 반면, 순간적 경로통행시간은 운전자가 경로를 출발하는 시각에 경로를 이루는 모든 링크에서 나타난 통행시간을 그대로 합하여 계산한다.

결국, 링크통행시간의 예측이 중요한 이유는 운전자에게 제공하고자 하는 현실적 경로통행시간을 계산하기 위해서 경로를 이루는 각 링크의 진입시점에서의 링크 통행시간을 알아야 하기 때문이다. 따라서 본 연구에서는 가장 신뢰성 있는 링크통행시간 예측모형을 제시하고자 한다.

II. 링크통행시간 예측모형의 고찰

링크통행시간 예측의 기본적인 구조는 통행시간 시계열 자료의 구축, 모형의 선정 및 정산(parameter calibration, 모수 추정), 그리고 예측으로 이루어진다. 다시 말해, 과거로부터 현재까지 조사된 통행시

간자료들을 이용하여 링크통행시간의 변화 추세를 가장 잘 설명하는 모형을 구축하고 이를 기반으로 장래의 통행시간을 예측한다. 지금까지 알려진 링크통행시간 예측 모형으로는 확률과정 (Stochastic Process), ARIMA (Autoregressive integrated moving-average), 그리고 칼만필터링 (Kalman Filtering) 등을 들 수 있다.

1. 확률과정을 이용한 통행시간 예측모형

확률과정 중 하나인 마코프 과정(Markov Process)이란 현재의 시점 t 에서 나타난 어떤 시스템의 상태는 과거의 시점 $t-1$ 에서의 상태에만 영향을 받음을 의미한다. 즉, $t-1$ 시점 이전에 이 시스템이 어떤 상태에 머물렀건, 현재의 상태는 한 시점 이전의 과거의 상태에만 영향을 받음을 의미한다. 이러한 성질을 이용하면, $t+1$ 시점의 상태는 t 의 상태를 이용하여 예측할 수 있게 된다. 일반적으로 시스템의 상태를 나타내는 행렬을 S 라 하고, 이 상태가 전이되는 확률을 M 이라 하면, 미래의 상태는 다음과 같은 식으로 표현이 가능하다.

$$S_{t+1} = MS_t \tag{2}$$

이러한, 마코프 과정을 링크통행시간예측 모형에 적용한다면 다음과 같은 식을 구성할 수 있다.

$$\bar{c}_{t+1} = M \bar{c}_t \tag{3}$$

여기서, \bar{c}_t 는 링크통행시간벡터를 나타낸다.

따라서, 마코프 과정을 통해 미래의 통행시간을 예측 할 때는 전이확률(M)을 계산하는 것이 중요하다. 그러나, 마코프 과정은 일반적으로 균형상태 (equilibrium state)로 진행되는 동안의 상태변화를 표현하는 것이기 때문에 통행시간 예측에 이 이론을 적용하는 것은 무리일 것으로 판단된다. 그러나, 장기적인 통행시간의 패턴을 설명하기에는 유용한 모형이다. 가령 하루 동안 지속되는 통행시간의 변화를 모형화 할 수 있어, 30분 후 혹은 60분 후의 통

행시간이 어떻게 변화해 나갈지 예측할 수 있기 때문이다.

2. ARIMA를 이용한 통행시간 예측모형

ARIMA는 시계열 모형 중에서 가장 일반적인 형태라고 볼 수 있다. 시계열 모형 중 대표적인 자기회귀(AR : Auto-Regressive) 모형과 이동 평균법(MA : Moving Average) 모형 그리고 이들을 결합시킨 ARMA 모형을 모두 포괄하기 때문이다. 특히, ARIMA 모형은 단순히 AR 모형과 MA 모형을 결합시킨 ARMA 모형과 구별되는데 이는 원시 시계열 자료를 정상시계열 자료로 전환시키는 차분 과정을 거치기 때문이다. 여기서 정상시계열 자료란 뚜렷한 추세가 없는 즉 평균이 시간 축에 평행하며, 진폭(변동)이 시간의 흐름에 따라 변하지 않고 일정한 자료를 말한다. 이러한 정상시계열 자료의 특성을 통계적으로 기술하면 다음과 같이 정리할 수 있다.

- 시계열 자료 Z_t 의 평균 $E(Z_t)$ 는 시간 t 에 무관한 상수이다.
- 시계열 자료의 분산 $Var(Z_t)$ 는 시간 t 에 무관한 상수이다.
- 시계열자료의 공분산 $Cov(Z_t, Z_{t+h})$ 은 시간 t 에 무관하며 단지 시차(time lag) h 에만 의존한다.

그러나, 대부분의 시계열 자료는 이러한 정상성을 만족시키지 못하는 경우가 많다. 평균이 시간에 따라 변하기도 하고, 분산이 상수가 아니라 증가하거나 감소한다. 그러나, AR 혹은 MA 등의 시계열 모형은 정상 시계열 자료를 기반으로 작성되었기 때문에 비정상 시계열 자료는 정상 시계열로 변환시켜야 한다. 이러한 변환은 대개 분산 상수화와 평균 상수화를 통해 이루어진다. 분산 상수화를 위해서는 시계열 자료를 log 변화시키거나 제곱근을 취한다. 한편 평균 상수화에서는 차분 과정을 거치게 된다. 차분이란 시계열 자료의 차이를 지칭한다. 예를 들어, 시계열 자료가 직선의 추세를 가진다면 1차 차분하여 평균을 상수로 만들 수 있다. 다음과

같이 직선 추세와 정상시계열 자료로 구분할 수 있는 시계열자료를 가정하자.

$$Z_t = \beta_0 + \beta_1 t + I_t \quad (4)$$

여기서,

$\beta_0 + \beta_1 t$ 는 직선추세를 나타냄. (β_0, β_1 = 상수),
 I_t 는 평균이 0인 정상시계열 자료임.

이 경우 1차 차분은 다음과 같은 식으로 표현된다.

$$Z_t - Z_{t-1} = (1-B)Z_t \quad (5)$$

여기서 B 는 다음의 조건을 만족시킨다.

$$B^j Z_t = Z_{t-j} \quad (6)$$

즉, $j=1$ 인 경우 $BZ_t = Z_{t-1}$ 이 된다.

차분식 식(5)에 직선 시계열 자료를 대입하여 풀면 다음과 같이 된다.

$$\begin{aligned} (1-B)Z_t &= \beta_0 + \beta_1 t + I_t \\ &\quad - [\beta_0 + \beta_1(t-1) + I_{t-1}] \\ &= \beta_1 + (I_t - I_{t-1}) \end{aligned} \quad (7)$$

즉, 1차 차분된 $(1-B)Z_t$ 시계열은 평균이 β_1 인 정상시계열이 됨을 알 수 있다. 결국, ARIMA 모형은 차분 과정을 포함시킨 ARMA 모형이라고 볼 수 있다.

비정상시계열 $\{Z_t\}$ 가 있을 때 d 차 차분한 $W_t = (1-B)^d Z_t$ 가 평균수준이 μ 인 정상시계열 ARMA(p,q) 모형을 따를 때 $\{Z_t\}$ 는 ARIMA(p,d,q)를 따른다고 하며, 다음과 같은 식으로 표현된다.

$$\phi(B)(W_t - \mu) = \theta(B)a_t, \quad W_t = (1-B)^d Z_t \quad (8)$$

위의 식에서 좌변은 AR 모형 구조와 맥락을 같이 하며, 우변은 MA모형 구조와 맥락을 같이한다.

따라서 이러한 형태는 ARMA 모형과 구조가 같다고 볼 수 있다. 다만, 차분된 시계열 W_t 를 이용하는 것에서만 ARMA 모형과 차이가 난다. 따라서, ϕ 와 θ 는 각각 AR모형과 MA 모형의 파라미터가 되며, a_t 는 정상시계열자료 (주로 $(0, \sigma^2)$)를 따르는 white noise, 백색잡음)이다. 또한 ARIMA(p,d,q)에서 차분차수 d 이외에 p 와 q 는 AR 모형과 MA 모형의 차수(order)를 의미한다. 예를 들어, $p=2, q=2, d=1$ 이라면 AR 모형의 차수가 2, MA 모형의 차수가 2, 차분차수는 1인 ARIMA 모형이 된다. 이 같은 경우 $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$ 은 $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$ 로 표현된다.

3. 칼만필터링을 이용한 통행시간 예측 모형

칼만필터링의 목적은 무작위 오차 (random error)를 지니는 관측치로 부터 분석하려는 시스템의 상태를 예측하는 것이다. 칼만필터를 위한 상태방정식과 관측치에 대한 추정식은 다음과 같다.

상태방정식(state equation) :

$$x_{k+1} = Ax_k + w_k \quad (9)$$

관측방정식(observation equation) :

$$z_k = Hx_k + v_k \quad (10)$$

여기서,

x_k 는 k 시점에서 시스템의 상태를 나타내고,

z_k 는 관측치,

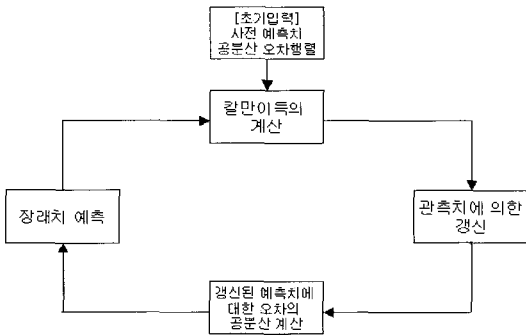
w_k, v_k 는 random error,

A, H 는 상수행렬을 나타낸다.

위의 상태 방정식 식(9)에 의하면, $k+1$ 시점에서의 예측치 x_{k+1} 은 이전 시점의 상태 x_k 와 그 때의 오차항 w_k 를 관련시켜 예측할 수 있다. 또한 현재 시점 k 에서 새로운 관측치 z_k 는 현재에 대한 예측치 x_k 와 현재의 오차항 v_k 를 관련시켜 계산할 수 있다. 칼만 필터에서는 예측치 x_k 와 오차항 w_k 의

분산 P_k 에 대한 초기치, x_k^-, P_k^- 가 주어진 상태 (a priori estimate)에서 상태방정식으로 x_k 의 예측치 \hat{x} 를 구하고, \hat{x} 와 새로운 관측치 z_k 를 이용해 더 나은 x_k 를 다시 예측한다. 즉, 관측치를 포함시켜 새로운 예측치를 찾으므로 보다 현실에 가까워진 예측치를 구하게 된다. 새로운 예측치는 대개 \hat{x} 와 관측치 추정오차의 선형조합 (linear combination)으로 이루어진다. 이때 칼만이득이라 불리는 'K' 상수를 매개 상수로 이용하며, 이 매개상수는 예측 오차를 최소화시키는 방법으로 추정된다. 위에 기술한 방식으로 시점 k를 바꿔가면서 모든 시점에 대한 상태를 예측하면, 전체적인 시점에 대해 예측이 가능하다.

이러한 Kalman Filtering 과정을 순서도로 표현하면 다음과 같다.



〈그림 1〉 칼만필터링 과정

III. 링크 통행시간 예측모형의 비교

이 장에서는 2장에서 고찰한 3개의 링크통행시간 예측모형을 실제 링크에 적용하여 모형들의 예측의 정확성에 대해 검증해 보고자 한다.

1. 평가기준

모형의 정확도를 평가하기 위한 기준으로는 평균 절대오차(MAE)와 상대평균절대오차율(MP)을 사용하였으며, t-검정을 관측 통행시간과 예측 통행시간의 평균 차이에 대한 통계적 유의성을 검토하기 위

〈표 1〉 사례링크에 대한 사항 및 평가방법

링크구간	링크 길이	링크특징
북한남 삼거리→ 한남로터리(55262)	1130m	침두/비침두 구분 명확
북한남 삼거리→ 명동역(24221)	2406m	낮 시간동안 지체 다소 발생
명동역→ 북한남삼거리(24222)	2272m	오후 및 저녁시간 지체발생
한남로터리→ 북한남삼거리(55261)	1130m	전일 거의 소통원활

한 지표로 삼았다.

$$MAE = \frac{1}{N} \sum_i |X_{obs} - X_{prd}| \quad (11)$$

$$MARE = \frac{1}{N} \sum_i \frac{|X_{obs} - X_{prd}|}{X_{obs}} \quad (12)$$

$$MP = MARE \times 100 \quad (13)$$

여기서,

N 은 링크의 총 개수,

t 는 단위시간,

X_{obs} 는 현장 실측자료,

X_{prd} 는 모형으로부터 계산된 예측자료

MAE는 각 시간대 별 절대값 오차의 평균을 의미하므로 값이 작을수록 신뢰도가 높다. MP는 각 시간대 별 절대값 오차를 실제 통행시간으로 나눈 값으로 단위는 %이며, 값이 100을 넘을 수도 있다.

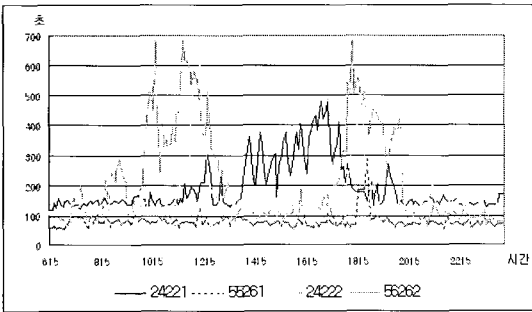
2. 4개 링크에 대한 예측모형 비교

1) 5분 후 예측모형별 비교

링크통행시간 예측모형을 임의의 링크에 적용하여 예측의 정확성을 비교하고자 하였다. 혼잡도로 구간인 남산1호 터널과 이와 인접한 한남로를 사례 링크로 선정하였다. 예측은 평일(2001/12/5,수)과 주말(2001/12/9,일)을 구분하여 5분 단위의 자료를 이

용하여 5분 후에 대한 예측을 하였다. 사례링크에 대한 사항은 아래 표에 정리하였다.

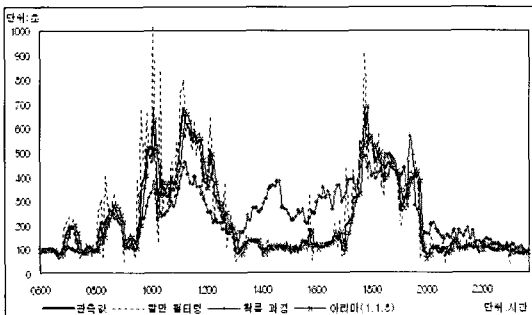
<그림 2>는 사례링크의 오전 6시부터 자정까지의 통행시간 흐름도 이다. 북한남 삼거리→한남로터리 (55262)구간은 첨두 / 비첨두의 구분이 명확하며, 북한남 삼거리→명동역(24221)구간은 오전이나 오후에 비해 낮 시간 동안 다소 지체가 발생한다. 명동역→북한남 삼거리(24222)구간은 12시와 오후 6시경 지체가 한차례씩 발생하고 나머지 시간은 원활한 소통을 보이며, 한남로터리→북한남 삼거리(55261)구간은 전일 거의 지체가 발생하지 않음을 알 수 있다.



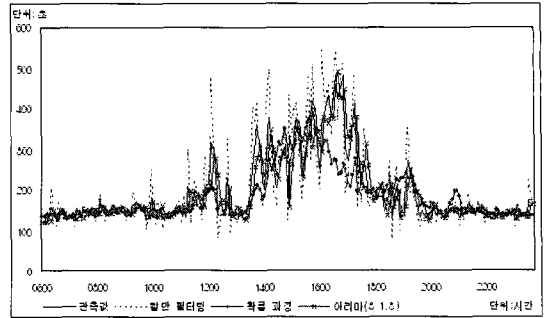
<그림 2> 4개 링크의 통행시간

아래 <그림 3>~<그림 6>의 그래프는 4개 선정링크 각각의 모형별 예측 링크 통행시간과 실제 관측 링크 통행시간을 비교한 그래프이다.

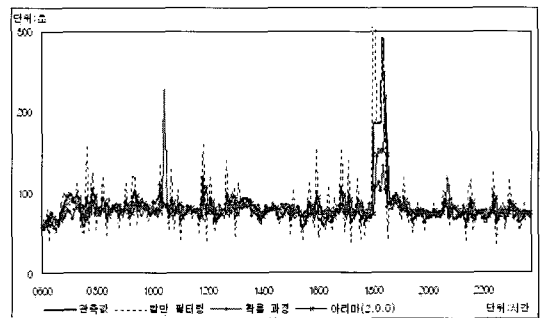
교통류가 가장 안정적인 한남로터리 → 북한남 삼거리 구간은 모든 모형별 오차가 다른 링크에 비해 상대적으로 작게 나오는 반면, 첨두 / 비첨두시



<그림 3> 북한남삼거리→한남로터리 모형별 예측치 비교



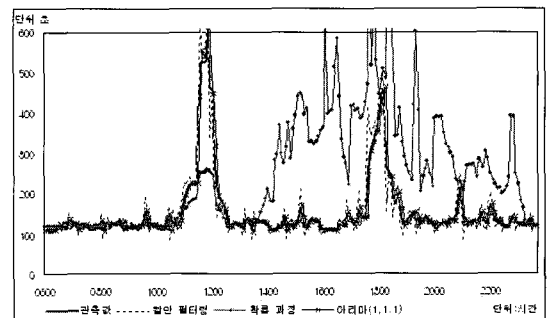
<그림 4> 북한남삼거리→명동역 모형별 예측치 비교



<그림 5> 한남로터리→북한남삼거리 모형별 예측치 비교

구분이 가장 명확하며, 전체적인 교통류가 가장 불안정한 북한남 삼거리 → 한남로터리 구간은 모든 모형별 오차가 다른 링크에 비해 상대적으로 크게 나오고 있음을 알 수 있다.

전체적으로 보면, 아리마 모형과 Kalman Filtering 모형이 비슷한 적합도를 나타내고 있으나, 아리마 모형은 Data의 내용(형태)에 따라 매번 다른 (p,d,q)를 추정해야 하는 단점이 있다. 다시 말해 예측 시



<그림 6> 명동역→북한남삼거리 모형별 예측치 비교

<표 2> 모형 예측치와 실측치의 오차비교

구분	북한남 삼거리→ 한남로터리			북한남 삼거리→명동역		
	MAE (초)	MP (%)	t 값	MAE (초)	MP (%)	t 값
칼만	34.70	15.53	-2.47	25.32	11.49	-2.29
확률	88.05	55.16	-2.47	30.44	12.92	2.13
아리마	35.07	17.14	-0.01	23.51	11.50	0.17
구분	명동역→북한남 삼거리			한남로터리→ 북한남 삼거리		
	MAE (초)	MP (%)	t 값	MAE (초)	MP (%)	t 값
칼만	13.32	7.59	-1.01	11.62	13.60	-1.61
확률	118.21	80.03	-9.17	10.81	11.95	1.56
아리마	14.83	8.28	0.02	10.57	12.34	0.15

주 : 음영부분은 t-검정결과(95%신뢰구간, 유의수준 0.05) 차이있는 부분임

간단위가 달라지면 모형구조도 바뀌어야 한다. 예측의 정교함에서는 다소 앞선다 할지라도 새로운 Data를 이용할 때마다 특히 실시간 정보를 Update 할 때마다 모형식별의 단계를 거쳐야 하므로 실질적으로 통행시간 예측에 적용하기는 어려움이 있다고 할 수 있다.

확률과정모형은 Calibration Data와 Test Data의 통행패턴 차이에 크게 영향을 받으나, Data의 양이 많아서 편차가 작거나, 통행시간이 안정적인 링크에 대한 예측의 신뢰도는 다른 모형들보다 높다. 반면에, 돌발상황이 발생하거나 Calibration Data와 Test Data의 통행시간 편차가 클 경우, 예측의 신뢰도는 매우 낮아지는 단점이 있다.

2) 한 개 링크에 대한 예측시간별 모형별 예측 비교

이 절에서는 한 개의 링크를 선정하여 예측시간을 달리 했을 경우 모형간 실측치와의 오차를 비교하기로 하였다. 선정한 4개의 링크 중 북한남삼거리→명동역 구간의 평일에 대한 5분 후, 10분 후, 15분 후, 30분 후, 60분 후에 대한 모형별 예측치와 실측치를 비교하였다. 선정한 북한남 삼거리→명동

역 구간은 평일과 주말의 통행시간 패턴이 뚜렷이 구분되는 구간이다. <표 3>는 모형별 5분 후, 10분 후, 15분 후, 30분 후, 60분 후의 모형의 예측된 링크통행시간과 실제의 링크통행시간의 오차를 검증한 표이다.

<표 3> 모형별 예측시간의 신뢰성 검증

칼만필터링 모형의 예측결과			
	MAE(초)	MP(%)	t 값
5분 후	25.32	11.49	-2.29
10분 후	28.85	13.17	-1.68
15분 후	23.59	10.58	-1.35
30분 후	28.99	13.28	-0.82
60분 후	39.08	16.98	-0.88

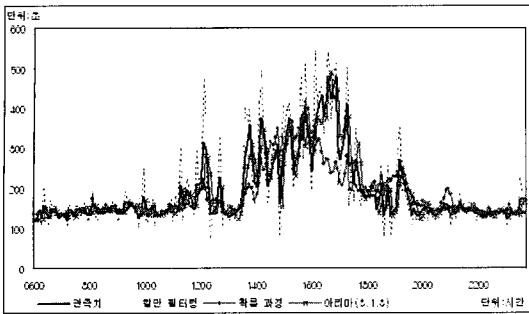
확률과정 모형의 예측결과			
	MAE(초)	MP(%)	t 값
5분 후	30.44	12.92	2.13
10분 후	28.14	11.88	1.61
15분 후	26.52	10.95	1.39
30분 후	24.50	10.35	1.04
60분 후	20.64	8.54	1.07

아리마 모형의 예측결과			
	MAE(초)	MP(%)	t 값
5분 후	23.51	11.50	0.17
10분 후	25.81	12.77	-0.03
15분 후	22.38	10.57	0.36
30분 후	28.43	13.80	0.27
60분 후	34.21	15.97	0.32

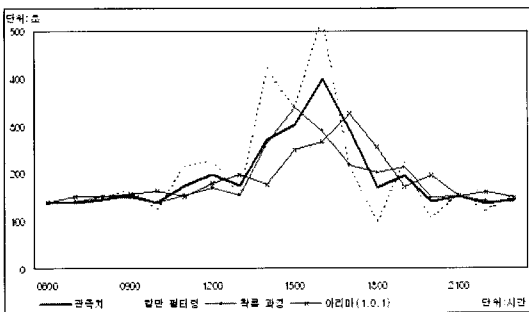
주 : 음영부분은 t-검정결과(95%신뢰구간, 유의수준 0.05)차이있는 부분임

시간 간격별 모형의 예측치와 실측치의 오차비교 결과를 보면 예측 시간간격이 단기 통행시간 예측 일수록 칼만필터링 모형과 아리마 모형의 예측치가 오차율이 적으며 시간 간격이 길어질수록 오차율이 증가하며, 반면 확률과정의 예측치 오차율은 시간이 길어질수록 감소하고 있음을 알 수 있었다.

대표적으로 5분 후(그림 7)와 60분 후(그림 8)의 모형별 실측치와 예측치의 비교 그래프를 보인다.



〈그림 7〉 북한삼거리→명동역구간의 5분후 예측 비교



〈그림 8〉 북한삼거리→명동역구간의 60분후 예측 비교

확률과정모형, 아리마모형, 칼만필터링 모형을 이용한 북한삼거리→명동역 구간 링크통행시간 예측 신뢰성 검증 결과 단시간 예측에서는 칼만필터링과 아리마 모형이 예측력이 우수했고 30분, 60분 후 예측으로 예측시간이 길어질수록 확률과정모형이 칼만필터링과 아리마에 비해 우수한 예측력을 나타냄을 알 수 있었다.

그러나 아리마 모형은 DATA의 형태에 따라 매번 다른(p,d,q)를 선정해야함에 통행시간 예측에 적용하기에는 어려움이 있다.

따라서 본 연구의 목적으로 들어가서 다음 장에서는 통행시간 예측모형으로 확률과정과 칼만필터링 모형을 확대 적용하여 보기로 한다.

IV. 선정된 모형의 적용 및 검증

1. 대표적인 4개 링크에의 적용

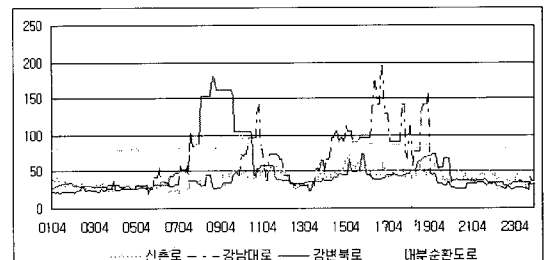
본 장에서는 전 장에서 살펴본 칼만필터링과 확

률과정 통행시간 예측 모형을 서울시 28개 교통축으로 확대 적용하여 보는데 궁극적인 목적이 있다. 28개 교통축 적용에 앞서 우선적으로 4개 교통축의 각각의 한 링크에 대해 확률과정과 칼만필터링 모형을 적용해 본다. 교통축은 도시고속도로인 강변북로와 내부순환도로, 간선도로인 강남대로, 도심부의 신촌로로 선정하였고 이 축들마다 하나의 링크를 선정하여 각각 5분후, 10분 후, 15분 후, 30분 후, 60분 후의 통행시간을 예측하였다. 각 교통축에서 선택된 링크는 다음 표와 같으며 이들 링크의 통행시간패턴(2001, 11, 20, 화요일)은 아래 <그림 9>와 같다.

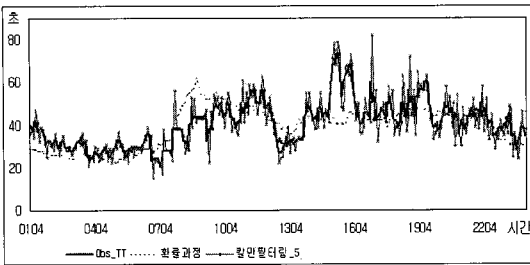
칼만필터링 모형을 이용한 5분후 링크통행시간 예측은 현재 시간 자료 5분 전 자료까지를 사용하였으며, 10분 후 예측을 하기 위해서는 10분 전 자료까지를 사용하여 통행시간을 예측하였다. 15분 후, 30분 후, 60분 후의 통행시간 예측도 방법은 같다. 확률과정 모형은 과거의 평균과 분산을 이용하여 통행시간을 예측하였는데, 2001년 11월 20일(화) 예측을 위해 2001년 9월 4일부터 2001년 11월 13일

〈표 4〉 4개 링크의 교통축 및 시·종점

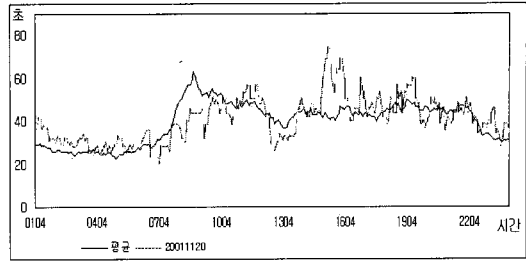
도로 분류	교통축	시점	종점
도시 고속도로	강변북로	상암LPG 주유소	성산대교 북단
도시 고속도로	내부순환도로	마장동 분기점	성동구청
간선도로 (강남)	강남대로	서초 우체국앞	염곡 교차로
간선도로 (도심)	신촌로	신촌역 입구	이대입구



〈그림 9〉 4개 링크의 하루 통행시간



<그림 10> 신촌로 5분 후 예측 결과



<그림 11> 신촌로 과거 평균통행시간과 예측일 통행시간 비교

까지의 화요일의 통행시간 자료를 이용하였다. 확률과정 모형은 실시간 통행시간을 이용하는 것이 아니라 과거 자료를 이용하기 때문에 통행시간 예측이 예측 단위 시간에 관계가 없다.

1) 하루의 통행시간 예측결과의 비교

우선 4개 교통축의 각 선택된 링크의 예측시간 간격별로 확률모형과 칼만 필터링 모형의 예측 시간과 관측된 실측 시간을 비교하였다.

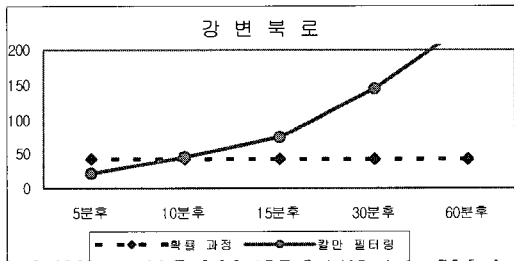
비교결과 확률모형은 과거 데이터와 통행패턴이 비슷할 경우에는 통행시간의 예측치와 실측치가 비슷하지만 통행 패턴이 다를 경우에는 많은 오차를

보인다. 하나의 예로 신촌로를 보자.

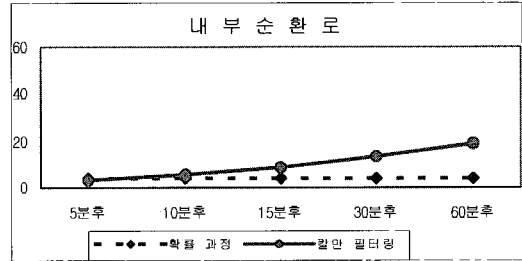
신촌로는 과거의 평균통행시간과 예측일의 통행시간이 거의 비슷하나 오전 9시경과 오후 3시경은 차이가 크다. 이 시간대에 관측치와 예측치의 차이가 크을 <그림 10>과 <그림 11>의 비교에서 알 수 있다.

링크별로 칼만필터링 모형을 이용하여 시간별 예측결과를 관측치와 비교해 보았다. 마찬가지로 신촌로에 대한 결과를 보자.

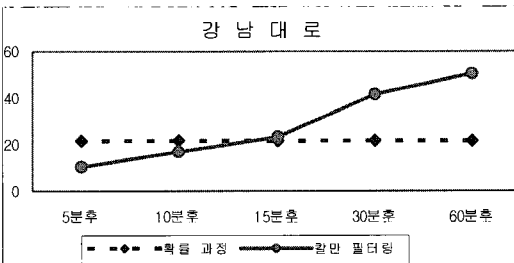
칼만 필터링 모형은 단기 예측인 5분 후, 10분 후 통행시간 예측의 경우 관측 통행시간과 비슷한 패턴을 보이며, 오차율도 적게 나타난다. 따라서 칼만 필터링 모형은 단기 통행시간 예측에 적합한 것을



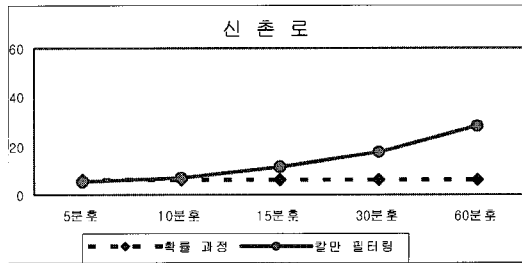
(a) 강변북로



(b) 내부순환로



(c) 강남대로

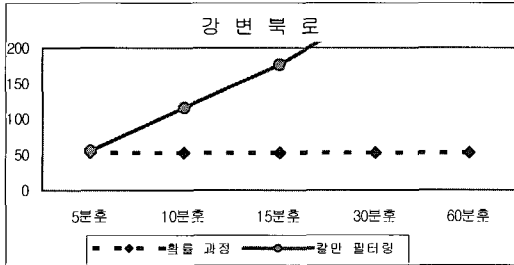


(d) 신촌로

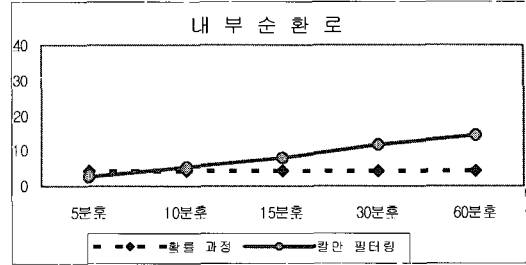
<그림 13> 전체시간대 4개 링크 시간대별 MAE

알 수 있다. 도로상에서 수시로 변하는 현상 (교통 사고, 갑작스런 정체 등)을 다른 모형보다도 아주 민감하게 반영할 수 있을 것임을 알 수 있다. 하지만 장기 통행시간 예측에는 오차가 많이 나타난다.

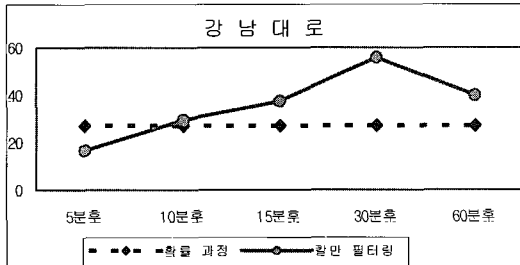
통행 예측 시간이 길어질수록 확률과정 모형의 통행시간 예측력은 높다. 이는 장시간 통행시간 예측은 전반적으로 과거의 평균 통행시간과 비슷하다는 것을 의미한다.



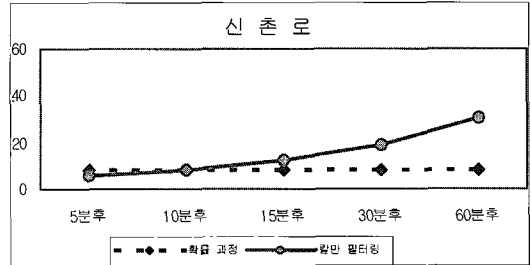
(a) 강변북로



(b) 내부순환로

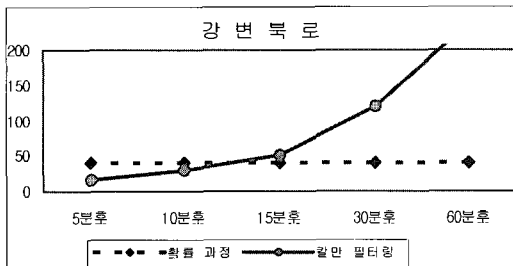


(c) 강남대로

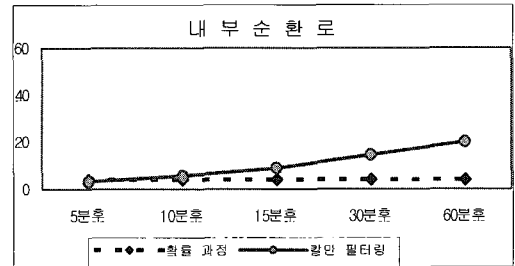


(d) 신촌로

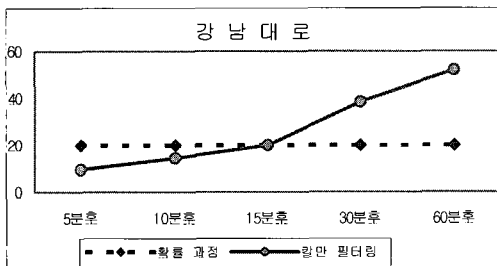
(그림 14) 침두시 4개 링크의 시간대별 MAE



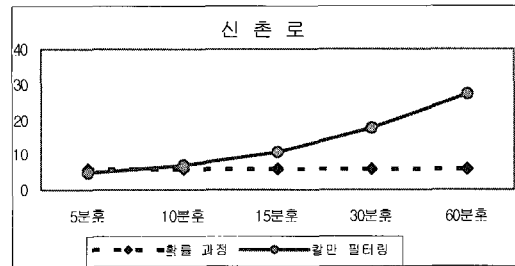
(a) 강변북로



(b) 내부순환로

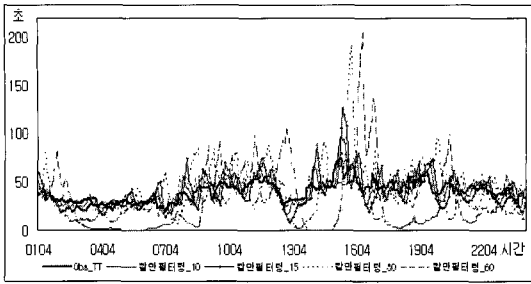


(c) 강남대로



(d) 신촌로

(그림 15) 비침두시 4개 링크의 시간대별 MAE



〈그림 12〉 신촌로 시간별 칼만필터링 예측결과

2) 시간대별 예측결과 비교

이 절에서는 전체시간대, 침두시, 비침두시로 나눠 예측 오차를 비교해 보았다. 침두시간대는 오전 7시와 9시 사이와, 오후 6시와 8시 사이이며, 비침두 시간대는 침두 시간 이외의 시간대이다.

(1) 전체시간대

전체 시간 MAE 결과를 보면 칼만 필터링을 이용한 경우에는 5분 후 예측과 10분 후 예측인 경우에 MAE가 작고, 4개 링크 모두 예측 통행시간이 늘어남에 따라 MAE가 점차 증가한다. 이를 통해 10분-15분 후 예측은 칼만 필터링으로, 30분-60분 후 예측은 확률 과정으로 하는 것이 적합할 것임을 알 수 있다.

(2) 침두시

침두시 MAE 결과를 보면 신촌로, 강남대로, 강변북로, 내부 순환로 모두 예측 통행시간 간격이 늘어남에 따라 MAE가 점점 증가한다. 특히 다른 링크에 비해서 통행 시간의 변동이 큰 강변북로에서 칼만필터링 모형은 다른 링크보다 예측시간간격이 커질수록 MAE가 커지는 것을 확인할 수 있다.

(3) 비침두시

비침두시 MAE 결과를 보면 침두시에 비해 MAE가 대체로 작으나, 침두시와 마찬가지로 모든 링크에서 통행 시간이 늘어남에 따라 MAE가 증가한다. 대체로 전체시간대와 거의 비슷한 패턴을 보인다.

관측 통행시간과 칼만 필터링 모형(5분후)을 이용하여 예측한 통행시간을 비교 분석하였다. <표 5>

의 MAE, MP 결과는 해당 교통축에 포함하는 모든 링크의 평균이다.

MAE는 대체적으로 4.99초와 14.3초 사이의 값을 보이고 있다. MP는 모든 축이 18% 미만의 오차를 보이는 것으로 보아 칼만 필터링 모형은 신뢰할 만 하다고 할 수 있다

〈표 5〉 전체 시간대 4개 교통축 오차 비교

분석 링크	분석 모형	MAE (초)	MP (%)	t 값
신촌로	Stochastic Process	6.37	15.69	2.98
	Kalman Filtering(5분 후)	5.06	12.33	0.16
	Kalman Filtering(10분 후)	7.15	17.47	-0.69
	Kalman Filtering(15분 후)	11.03	27.30	0.20
	Kalman Filtering(30분 후)	17.84	44.53	1.15
	Kalman Filtering(60분 후)	27.80	70.88	4.54
강남대로	Stochastic Process	21.33	32.91	5.92
	Kalman Filtering(5분 후)	10.65	17.49	-1.29
	Kalman Filtering(10분 후)	17.19	27.45	-2.11
	Kalman Filtering(15분 후)	23.21	41.84	-1.22
	Kalman Filtering(30분 후)	41.62	78.93	-0.43
	Kalman Filtering(60분 후)	50.15	84.13	0.12
강변북로	Stochastic Process	42.13	16.48	-1.39
	Kalman Filtering(5분 후)	22.24	7.31	-0.89
	Kalman Filtering(10분 후)	45.42	15.11	-2.32
	Kalman Filtering(15분 후)	73.65	24.73	-1.92
	Kalman Filtering(30분 후)	142.69	51.07	1.03
	Kalman Filtering(60분 후)	238.07	81.09	1.85
내부 순환로	Stochastic Process	3.83	4.72	-4.17
	Kalman Filtering(5분 후)	2.98	3.57	1.76
	Kalman Filtering(10분 후)	5.25	6.36	1.15
	Kalman Filtering(15분 후)	8.38	10.18	1.30
	Kalman Filtering(30분 후)	13.70	16.71	1.60
	Kalman Filtering(60분 후)	19.27	23.29	2.45

주 : 음영부분은 t-검정결과(95%신뢰구간, 유의수준 0.05) 차이있는 부분임

2. 28개축에 칼만필터링 모형의 적용

서울시 28개 주요 교통축 967개 링크에 칼만 필터링 예측 기법을 확대 적용하였다. 기준시간 전 5분 전 통행시간까지를 이용하여 5분 후를 예측하였다.

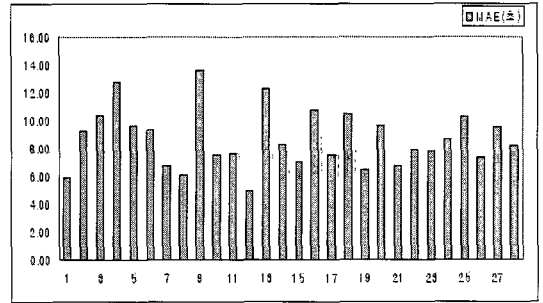
서울시 28개 주요 교통축	
1: 강변북로	15: 성안순환로
2: 강서로	16: 수색로
3: 남부순환로	17: 신촌-계동포길
4: 내부순환로	18: 연서-증산로
5: 도림천-신대방길	19: 울림픽대로
6: 도봉-종암-연주로	20: 원효-대방-시흥대로
7: 동1-영동대로	21: 월계-미아-대학로
8: 동부간선로	22: 자양-송파대로
9: 동작대로	23: 천호대로
10: 둔촌-양재대로	24: 충정-마포-경인로
11: 망우-왕산로	25: 통일-한강-관악로
12: 반포-반포대로	26: 퇴계-광나루-강동대로
13: 서부 간선로	27: 한남-강남-현릉로
14: 선사-테헤란-서초로	28: 한천로

관측 통행시간과 칼만 필터링 모형을 이용하여 예측한 통행시간을 비교 분석하였다. <표 6>의 MAE, MP 결과는 해당 교통축에 포함하는 모든 링크의 평균이다.

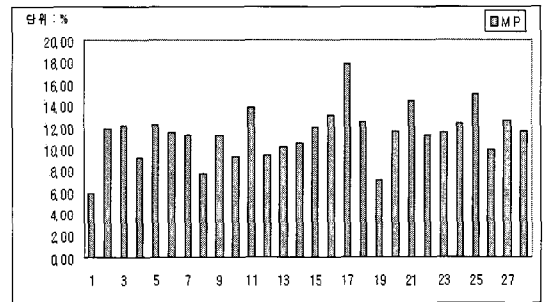
<표 6> 28개 교통축 오차 비교

교통축	칼만필터링		교통축	칼만필터링	
	MAE(초)	MP(%)		MAE(초)	MP(%)
1	5.89	5.969	15	7.09	12.015
2	9.18	11.954	16	10.76	13.141
3	10.39	12.127	17	7.50	17.804
4	12.74	9.269	18	10.46	12.562
5	9.58	12.288	19	6.42	7.256
6	9.28	11.577	20	9.59	11.714
7	6.76	11.291	21	6.79	14.507
8	6.09	7.764	22	7.87	11.225
9	13.55	11.261	23	7.76	11.590
10	7.54	9.405	24	8.67	12.391
11	7.60	13.926	25	10.19	15.073
12	4.99	9.515	26	7.36	10.001
13	12.16	10.259	27	9.50	12.595
14	8.19	10.631	28	8.18	11.689

MAE는 대체적으로 4.99초와 14.3초 사이의 값을 보이고 있다. MP는 모든 축이 18% 미만의 오차를 보이는 것으로 보아 칼만 필터링 모형은 신뢰할 만 하다고 할 수 있다.



<그림 16> 28개 교통축 MAE



<그림 17> 28개 교통축 MP

V. 결론 및 향후 연구과제

본 연구에서는 통행시간 예측모형들의 예측력을 특정 링크들에 대해 다각적인 방법으로 비교·검토하여 보았으며 그 결과 칼만필터링 모형이 단기 통행시간 예측에 적합성이 높다고 판단하였다. 최종적으로 서울시 28개 주요 교통축에 대해 확률과정 모형과 칼만 필터링 모형을 적용해 보았다.

28개 교통축 967개 전체 링크는 5분 후 칼만 필터링 모형의 예측 결과를 분석하였으며, 967개 링크 중에서 도로 특성을 고려한 4개 링크 (도시 고속도로인 강변북로와, 내부순환도로, 도심 지역의 간선도로 중에서 신촌로, 강남 지역의 간선도로 중에서 강남대로)는 좀 더 자세한 분석을 해 보았다. 4

개 링크는 확률과정 모형과, 칼만 필터링 모형을 이용하여 5분 후, 10분 후, 15분 후, 30분 후, 60분 후의 예측 통행시간을 분석하였다. 예측된 통행시간의 평가 지표로는 MAE, MP, t-검정을 사용하였으며, 첨두 시간대, 비첨두 시간대, 전체 시간대에 대해 비교하였다.

칼만 필터링 모형은 5분 후-15분 후 예측과 같은 단기 예측에 그 정확도가 높다. 예측 통행시간이 전체적인 통행 패턴을 잘 반영함은 물론 그 오차값도 확률과정에 비해 작다. 칼만 필터링 모형은 통행시간을 예측할 때 실시간 자료를 이용하여 예측 오차를 보정하기 때문에 오차가 누적되는 것을 최소화시킨다. 이런 알고리즘 특성상 통행패턴이 급격히 변하거나, 유고 혹은 돌발 상황이 발생한 구간에 대한 예측력이 뛰어나다. 하지만 30분 후 예측과 같이 장기 예측의 경우에는 전반적으로 오차가 증가하며, 변동폭이 확대되는 것을 볼 수 있다.

확률과정 모형은 30분 후 예측과 같은 장기 예측에서 칼만 필터링 모형에 비하여 정확도가 높았다. 장기 예측일수록 평균에 근거하여 통행시간을 예측하는 확률과정 모형의 예측력이 더 우수하다고 할 수 있다. 확률과정 모형은 통행시간을 예측할 때 기존 과거자료(실시간 자료를 제외한)를 이용하기 때문에 유고 상황 발생 시에는 예측 정확도가 낮은 한계가 있다.

이상으로 볼 때 통행시간 예측에 있어 단기예측(5분 후-15분 후)에는 칼만 필터링이, 장기예측(30분 후-60분 후)에는 확률과정 모형이 적합하다고 판단된다. 두 모형의 장점을 고려하여 통행시간 예측에 이용한다면 한 가지 예측 모형을 이용하여 통행시간을 예측하는 것에 비해 그 정확도가 높을 것으로 판단된다.

향후 연구과제로는 본 연구를 발전시켜 경로기반 통행시간 예측 모형을 개발하고 실제로 교통정보안내분야 및 정보제공분야에서 다양하게 활용될 수 있는 방법을 연구하도록 한다.

참 고 문 헌

- [1] 과학기술부, Development of the Real Time Traffic Information Analysis Models, 2000
- [2] 서울시립대학교 부설 도시과학연구원, SK 최적 경로 시스템 신뢰도 검증 및 향상 방안 연구, 2002
- [3] 소경숙(2000), GIS를 이용한 교통정보 제공 시스템 개발에 관한 연구, 서울시립대학교 석사학위논문
- [4] 조범철(1998), 동적 기종점표 추정 모형 개발에 관한 연구, 석사학위논문
- [5] MOHINDER S.(1993), ANGUS P., Kalman Filtering: Theory and Practice Using MATRAB

〈저자소개〉



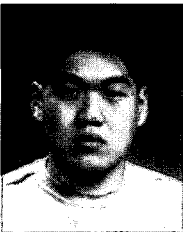
이 승 재(LEE, Seungjae)

런던대학교 토목공학과 교통공학박사
교통개발연구원 책임연구원(1995. 11~1996. 9)
서울시립대학교 도시공학과 전임강사(1996. 10~1998. 9)
서울시립대학교 교통공학과 조교수(1998. 10~2002. 9)
서울시립대학교 건축도시조경학부 교통전공 전공주임교수(2000. 8~2002. 7)



김 범 일(Kim, Beomil)

2002년 : 서울시립대학교 학사
2004년 : 서울시립대학교 석사
현재 : 교통개발연구원



권 혁(Kwon, Hyug)

2003년 2월 : 서울시립대학교 교통공학과 석사 입학
2003년 2월 : 충북대학교 도시공학과 졸업