

[연구논문]

**대기행렬 예측 정보 산출 모형에 관한 연구**  
- 남산1호터널을 중심으로 -

**A Model for Predictive Queue Information**  
- In case of Namsan 1 Tunnel -

이 청 원\* · 심 소 정\*\*

목 차

- |                |                  |
|----------------|------------------|
| I. 서론          | IV. 대기행렬 예측모형 검증 |
| II. 이론적 고찰     | V. 결론 및 향후 과제    |
| III. 대기행렬 예측모형 |                  |

**ABSTRACT**

Chung-Won Lee · So-Jung Shim

A traffic queue as well as travel time is a measure of congestion in urban networks. The time queue information based on detectors itself is less useful where the queue change is dynamic due to the time difference among data collection, data provision and drivers' experience time. Since this, a short term prediction mechanism on a queue is sometimes considered implement in an ATIS center software. Two queue prediction algorithms using the Kalman filter method are presented in this paper. One algorithm is a direct queue measure technique and the other is an indirect queue measure technique using the degree of congestion concept from adaptive signal operation. The algorithms were calibrated using a real data from the Namsan tunnel in Seoul, which is the first ATIS project in an urban street in Korea. The result shows the Kalman filter technique is capable to elicit the proper prediction.

\* 서울시정개발연구원 도시교통연구부 연구위원

\*\* 서울시정개발연구원 도시교통연구부 연구원

## 1. 서론

### 1. 연구의 배경 및 목적

교통정보시스템(ATIS)에서 운전자들에게 제공하는 정보들은 다음 세 개의 카테고리로 분류할 수 있다.

- 과거정보(Historical information) : 과거의 시스템 정보
- 실시간 정보(Real time information) : 현재 교통상황에 대해 가장 최근에 업데이트된 정보
- 예측정보(Predictive information) : 현재 교통상황에서 미래의 상황을 예측한 정보

현재 우리나라에서 실시되고 있는 교통정보시스템들에서 제공되는 정보는 대부분 과거정보(historical information) 또는 실시간 정보(current information)이지만, 실제 이용자들은 앞으로 벌어지게 될 네트워크 상황에 따라 통행행태를 결정하게 되기 때문에, 가장 유용하게 사용될 수 있는 정보의 형태는 신뢰성있는 예측정보(Predictive information)이다.

여러 가지 형태로 제공될 수 있는 예측정보 중, 실제 네트워크 이용자는 지체에 관하여 구체적인 수치와 권고 등의 상세한 교통정보 제공에 더욱 민감하게 반응하는 것으로 나타났다<sup>1)</sup>. 지체정보는 구체적으로 통행시간, 통행속도 및 대기행렬로 표현할 수 있는데, 본 연구에서는 이러한 지체를 나타내는 예측정보 가운데 대기행렬을 예측할 수 있는 모형개발을 연구의 대상으로 하였다.

그러면 여기서, ‘왜 대기행렬 예측정보 산출이 필요한가’라는 물음을 떠올리게 되는데, 그에 대한 답으로 두 가지 정도를 들 수 있다.

첫 번째는, 지체정보로서의 대기행렬 정보는 이용자의 경로선택에 큰 영향을 일으키는 것으로 나타났다. 이는 다음의 사례에서 알 수 있는 사실이다.

#### <사례>

- ① 파리의 SIRIUS에서 도로전광표지에 의해 나타나는 운전자의 경로전환 행태 분석 결과<sup>2)</sup>
- ② 남산1호터널 모니터링 결과

<표 1>에서, VMS 메시지가 “대기행렬 1km 이상”인 경우부터 운전자의 통행행태에 변화가 일어나며 대기행렬이 길어질수록 그 변화는 더욱 커진다는 사실을 유추할 수 있다. 이러한 결과를 토대로, 네트워크 이용자에게 제공할 수 있는 지체 정보 중, 대기행렬에 관한 정보가 얼마나 중요한지를 알 수 있으며, 대기행렬의 예측정보를 산출하여 이를 네트워크 이용자에게 제공할 경우, 교통관리(Traffic management) 측면에서 큰 효과를 불러올 것이라는 결론을 내릴 수 있다. 이것이 ‘왜 대기행렬 예측정보 산출이 필요한가’라는 물음의 첫 번째 이유가 된다.

두 번째 이유는, 본 연구는 남산1호터널 교통정보시스템을 대상으로 하는데, 이 대상 네트워크는 대기행렬이 단시간에 매우 급격히 변화하는 특성을 가지고 있다. 이러한 특성 때문에, 현재 남산1

1) 서울시 첨단교통정보체계(ATIS) 구현을 위한 기본연구, 서울시정개발연구원, 2000

2) Yim, Youngbin and Ygnace, Jean-luc, Link Flow Evaluation Using Loop Detector Data: Traveler Responses to Variable Message Signs, TRR 1550, pp.58-64.

<표 1> 대기행렬 정보에 대한 운전자 반응

구분	SIRIUS	남산1호터널
대기 행렬 길이 정보의 영향	평균 도차 분석	남산1호터널
	<ul style="list-style-type: none"> <li>- 대기행렬길이 0.5Hm 미만인 경우에는 평균도차율과 차이가 없음.</li> <li>- 1Hm를 넘어서면 도차교통량이 감소하여 평균도차율보다 낮게 됨.</li> <li>- 특히 3Hm 이상이 되면 평균도차율과의 차이가 급격히 증가함.</li> </ul>	<ul style="list-style-type: none"> <li>- 최초 정제정보가 제공되는 시점(대기행렬길이:약 1Hm)에서 터널이용차량의 변화가 발생시작됨.</li> <li>- 대기행렬길이 2Hm 이상되는 경우에 그 영향이 현저히 증가함.</li> </ul>
대기 행렬 길이 정보의 영향	교통 전환 정도 분석	남산1호터널
	<ul style="list-style-type: none"> <li>- 대기행렬길이 0.5Hm 미만인 경우에는 그 전환비율이 0.01% 미만으로 미약함.</li> <li>- 대기행렬길이 증가함에 따라 그 전환비율이 증가함. (1Hm: 0.01, 2Hm:0.11)</li> <li>- 특히 3Hm 이상이 되면 급격한 증가율을 보임. (3Hm: 0.17, 4Hm: 0.31)</li> </ul>	<ul style="list-style-type: none"> <li>- 대기행렬길이 1Hm의 경우에는 비터널도로로의 전환이 발생하기 시작함.</li> <li>- 2Hm 이상되는 경우에 그 영향이 현저히 증가하고, 3Hm가 되면 딱딱 그 증가율이 완만해짐.</li> <li>- 시스템 개통 후 3개월 이용에 교통지연보다 그 영향이 증가함.</li> <li>- 이러한 변화는 지체가 심한 오후에 더 현저하게 발생함.</li> </ul>

(※위 표의 교통류 전환정도 분석 항목에서, 통행전환율은 파로 SIRIUS의 경우 검지기를 기반으로 조사된 데이터를 기초로 산출한 것이며, 남산1호터널 모니터링의 경우 1호터널과 우회로 각각에 대한 번호판 조사로 얻은 데이터를 기초로 산출한 것을 그대로 인용한 것임.)

호터널 교통정보시스템에서 제공되고 있는 실시간 대기행렬정보를 이용자가 VMS를 통해 접하고, 통행경로를 결정, 터널부로 접근하였을 경우, 터널은 이미 차량의 대기행렬로 꽉 막혀있는 상황이 종종 발생한다.

이러한 상황이 발생하는 경우를 구체적인 예로 살펴보자.

운전자가 도심에서 남산1호터널을 통해 강남방향으로 진행하는 경우, 터널에 대한 정보를 얻을 수 있는 VMS의 위치는

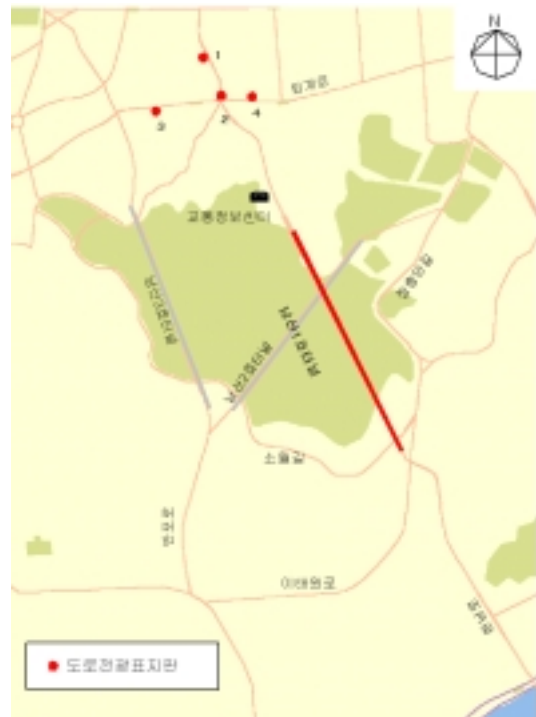
- ① 삼일로 계성초교 앞: 삼일로 이용자에게 정보제공

- ② 퇴계로2가 교차로: 삼일고가 이용자에게 정보제공

- ③ 퇴계로 밀리오레 앞: 퇴계로 이용자에게 정보제공

- ④ 충무로 극동빌딩 앞: 충무로 이용자에게 정보제공

이렇게, 총 네 곳이며, 구체적인 위치는 다음의 도면에 나타나 있다.



<그림 1> 남산1호터널 교통정보시스템 내, 강남방향 정보제공 VMS 위치

그리고, 다음 <그림 2>~<그림 4>은 8월, 남산1호터널 센터의 대기행렬길이 데이터 중, 3일치(8월 1일, 8일, 그리고 29일)를 발췌한 것이고, 각 그림에 해당하는 표는 그 날짜에 해당하는 특정 시간대의 대기행렬길이의 구체적 수치이다.



<그림 2> 8월 1일 대기행렬길이 변화 그래프

<표 2> 8월 1일 특정시각의 대기행렬길이

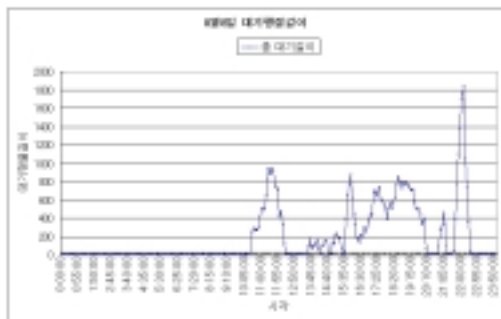
시각	대기행렬길이
11:00	733
11:05	891
11:10	1118
11:15	1351
11:20	1536



<그림 4> 8월 29일 대기행렬길이 변화 그래프

<표 4> 8월 29일 특정시각 대기행렬길이

시각	대기행렬길이
14:45	0
14:50	342
14:55	400
15:00	715
15:05	1048



<그림 3> 8월 8일 대기행렬길이 변화 그래프

<표 3> 8월 8일 특정시각 대기행렬길이

시각	대기행렬길이
21:35	30
21:40	384
21:45	758
21:50	1098
21:55	1391

VMS가 설치된 각 지점에서 남산1호터널에 대한 정보를 제공받고, 터널을 통해 강남방향으로 진행하겠다는 의사결정을 한 후, 터널입구까지 진입하기 위한 통행시간은 평균 5분~10분 사이이다. 8월 1일의 경우, 11시 10분의 대기행렬길이는 1118m이었으나, 10분이 경과한 11시 20분 대기행렬길이는 1536m로 10분사이에 418m가 증가하였다. 이는 운전자가 경로를 결정할 당시, VMS에 ‘터널남단부터 지체’라는 메시지였고, 이를 보고 운전자가 “터널을 원활히 통과할 수 있을 것”이라는 판단 하에 남산1호터널 진입을 결정하고 터널입구까지 도달하였으나, 이미 대기행렬이 증가하여 ‘터널중간부터 지체’인 상황이 발생한 형상이다. 나머지 8월 8일, 29일도 이와 마찬가지로의 경우이다. 그리고 <그림 2~4>의 그래프에서 알

수 있듯이, 대기행렬이 단시간동안 급격히 변화하는 상황이 하루에도 몇 번씩 반복하여 발생한다. 이러한 상황들의 발생 및 반복은 제공된 교통정보를 보고 판단하여, 터널부로 진입한 운전자에게 정보에 대한 불신을 더하게 되고, 교통정보제공을 통한 교통운영의 효과를 저하시킨다. 따라서, 운전자가 터널부의 정보를 제공받을 수 있는 지점에서 터널부의 지체에 관한 예측정보가 절실히 필요하다.

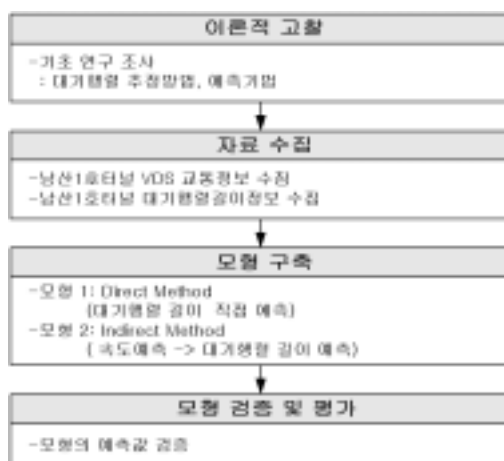
위에서 언급한 두 가지 이유가 ‘왜 대기행렬 예측정보 산출이 필요한가’라는 물음의 답이 되며, 이를 근거로 신뢰성 있는 대기행렬길이 예측정보 산출을 본 연구에서 수행하고자 한다.

## 2 연구의 방법 및 범위

### 1) 연구방법

본 연구는 서울시 남산1호터널 교통정보시스템 네트워크를 대상으로 하여, 네트워크 내에 생성되는 대기행렬의 길이에측을 주요내용으로 한다.

### 2) 연구수행절차



<그림 5> 연구수행도

## II. 이론적 고찰

### 1. 상태공간 모형

상태공간모형은 분석하고자 하는 시스템의 미래현상을 현재의 상태와 미래의 입력값을 이용해 설명할 수 있음을 모델링하는 것이다. 즉 그 시스템의 상태벡터(state vector)는 미래의 값을 예측하기 위해 필요한 현재와 과거의 정보의 최소량으로 정의할 수 있고 출력벡터(output vector)는 상태벡터(state vector)와 미래의 입력벡터(input vector)를 이용해 계산할 수 있다. 상태공간모형은 현재의 상태가 주어졌을 때 시스템의 미래는 현재의 상태에 의존한다는 마코프 성질(Markov property)을 바탕으로 하므로 마코비안 표현(Markovian representation)이라 부르기도 한다.

상태공간 모형 (state-space model)은  $\{X_t\}$ 와  $\{Z_t\}$ 에 대한 다음의 식으로 주어진다.

$$X_{t+1} = F X_t + G \epsilon_t, \quad t=1,2, \dots \quad (1)$$

$$Z_t = H X_t + \eta_t, \quad t=1,2, \dots \quad (2)$$

여기서,  $X_t$  : 시간  $t$ 에서의 상태벡터

$Z_t$  : 시간  $t$ 에서의 관측값벡터

$F$  : 변환행렬 ( $m \times m$  matrix)

$G$  : 입력행렬 ( $m \times g$  matrix)

$H$  : 출력행렬 ( $1 \times m$  matrix)

$\eta_t$  : 왜신벡터(innovation vector), 평균 벡터가 영벡터이고 공분산 행렬이  $Q$ 인  $\{\epsilon_t^i\}$ , ( $g \times 1$  vector)

$\epsilon_t$  : 평균 0, 분산  $\sigma_{\epsilon,i}^2$  인 시간에 따라 서로 독립인 확률과정

위의 (1)과 (2)의 식을 각각 상태공간모형의 상

태방정식 (state equation), 관측방정식 (observation equation)이라 한다.

## 2. Kalman filter

시계열 자료를 상태공간 모형으로 표현한 후, 관측치( $Z_t$ )를 통해 우리가 원하는 최상의 추정량( $X_t$ )을 얻고자 하는데 Kalman filter가 이용된다.

Kalman filter는 새로운 관측치가 얻어지는 경우에, 기존의 추정량을 바로 개선할 수 있게 해주는 방정식들로 구성되어 있다. 상세히 설명하면, Kalman filter는 두개의 part로 구성되는데, 첫 t Step은 Prediction Part로 다음에 관측되어지는 값을 예측하는 과정이고, 두 번째는 Updating Part로 새로운 관측값으로 기존에 예측한 값을 보정하는 과정으로 구성된다.

다음은 Kalman filter의 알고리즘인데, 보다 일

반적인 사례를 다루기 위해 ARMA(p,q) 과정에 있는 데이터의 상태공간모형식을 이용하였다.

알고리즘에서 사용되는 notation들은 다음과 같다.

$\hat{X}_t$  : 현재시점  $t$ 까지의 관측치  $Z_1, \dots, Z_t$ 가 주어졌을 때  $X_t$ 에 대한 MMSLE(Minimum

Mean Square Linear Estimates)

$\hat{X}_{t-}$  : 관측치  $Z_1, \dots, Z_{t-1}$ 이 주어졌을 때  $X_t$ 에 대한 MMSLE

$P_t$  :  $\hat{X}_t - X_t$ 의 공분산 행렬

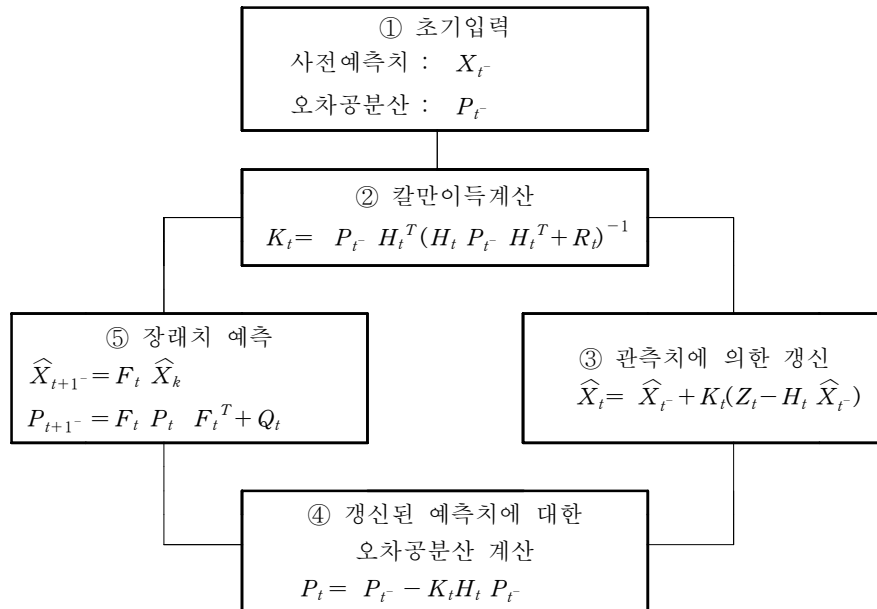
$P_{t-}$  : 예측에러(prediction error)  $\hat{X}_{t-} - X_t$ 의 공분산 행렬

$$P_{t-} = F P_{t-1} F' + G Q G'$$

Prediction part : 예측과정은 다음의 방정식을 통해 이루어짐

$$\hat{X}_{t+1-} = F \hat{X}_t \quad : \text{estimate}$$

$$P_{t+1-} = F_t P_t F_t^T + G Q_t G' \quad : \text{오차공분산}$$



<그림 6> Kalman filter 예측과정

Updating part: 관측치에 의한 갱신 (Measurement update)

$$K_{t+1} = P_{t+1}^- H_{t+1}^T [H_{t+1} P_{t+1}^- H_{t+1}^T + R_{t+1}]^{-1}$$

Kalman gain

$$\hat{X}_{t+1} = \hat{X}_{t+1}^- + K_{t+1} [y_{t+1} - H_{t+1} \hat{X}_{t+1}^-]$$

$$P_{t+1} = P_{t+1}^- - K_{t+1} H_{t+1} P_{t+1}^-$$

위 식에서 보듯이 Kalman filter는 새로운 observation들이 들어올 때마다, filtering equation들을 recursive하게 적용한다. Updating equation은 그 시점에서 가능한 정보로부터 state vector의 optimal estimator를 제공하여, 대부분의 경우 초기치에 상관없이 참값으로 빠르게 수렴하는 경향이 보이고 있다.

### III. 대기행렬 예측모형

#### 1. 남산1호터널 교통정보시스템 네트워크특성

##### 1) 남산1호터널 네트워크 특성

남산1호터널 교통정보시스템 대상 네트워크는 도심(퇴계로)과 강남(한남대교)을 연결하는 하나의 축으로 구성되어 있다. 도심의 삼일로 계성초교 지점을 시작으로, 터널남단에서 한남대교로 연결되는 한남2고가 전까지의 구간이 남산1호터널 교통정보시스템에 해당되는 네트워크이다. 대상네트워크가 가지는 고유의 특성을 다음의 몇 가지로 요약할 수 있다.

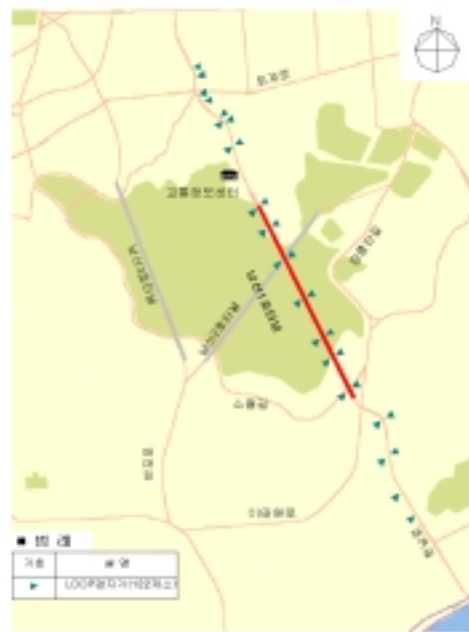
- ① 연속류의 성질을 가지지만, 터널 접근로에 불규칙적인 merging 및 weaving 현상을 보이는 non-homogeneous한 교통류 특성을 가짐
- ② 강남에서 도심으로 진행하는 방향은 거의 모든 시간 소통이 원활하나, 도심에서 강남방향

으로 진행하는 방향은 교통정체가 심각함

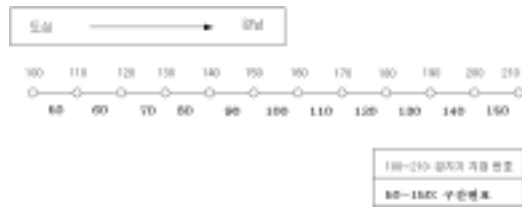
- ③ 특히, 강남방향의 교통상황은 변동패턴이 매우 다양하며, 단시간동안의 변화율이 매우 심함

##### 2) 대상네트워크의 노드-링크

도심의 삼일로 계성초교 지점부터 한남2고가 전까지 루프검지기가 250m 간격으로 매설되어 있으며, 이 루프검지기가 남산터널 교통정보시스템의 정보수집원이 된다. 본 연구에서는 루프 검지기에서 얻어진 데이터(교통량, 점유율, 속도)를 기초 데이터로 활용할 것이며, 본 연구의 구체적인 대상구간은 강남방향으로, 100번 검지기 톨게이트 후방 구름다리 지점에서 210번 검지기 한남2고가까지이다. 대상 네트워크내의 검지기 위치 및 각 지점에 대한 설명은 아래 <그림 7>, <그림 8>과 <표 5>에서 설명하고 있다.



<그림 7> 남산1호터널 교통정보시스템  
검지기 위치



〈그림 8〉 연구대상 네트워크

〈표 5〉 연구대상 네트워크 내의 검지기 지점 및 구간 설명

검지기 지점번호	설 명	구간 번호	구간 길이	관련지점
100	구름다리(돌계이트후방)	50	300	100~110
110	터널북단	60	320	110~120
120	터널 1.25km 지점	70	250	120~130
130	터널 1km 지점	80	250	130~140
140	터널 750m 지점	90	250	140~150
150	터널 500m 지점	100	250	150~160
160	터널 250m 지점	110	279	160~170
170	터널남단	120	334	170~180
180	북한남고가 중간	130	171	180~190
190	한남타워 육교	140	228	190~200
200	외인아파트	150	212	200~210
210	한남2고가			

## 2 대기행렬 예측모형 구축

### 1) 왜 Kalman filter인가?

본 연구에서 활용될 예측기법은 연구대상 네트워크의 특성을 고려하여 선정하여야 하는데, 위에서 설명한 대상네트워크의 특성 중, 특히 ‘교통상황의 변동패턴이 다양하며, 단시간동안의 변화율이 심하다’는 특성은 아래의 세 가지 Kalman filter 고유의 성질과 잘 부합된다.

- ① 상태공간 방정식은 마코프 성질(Markov property)<sup>3)</sup>을 가지고 있어, 교통패턴이 다양

하며 변동이 심한 남산1호터널 네트워크에 적용하기 적절한 예측기법임

- ② Kalman filter는 계산하는데 소요되는 시간이 매우 짧아 단기예측<sup>4)</sup>에 적합한 기법임

- ③ Kalman filter를 적용할 경우, 긴 과거의 데이터를 기억할 필요가 없고, 추차계산에 의해 최적해가 구해지므로, 특히 on-line 계산에 위력을 발휘함

이러한 세 가지 이유를 근거로 대기행렬길이 예측모형에 적용될 예측기법으로 Kalman filter를 선정하였다.

### 2) 왜 5분단위 예측인가?

다음절 구축모형의 구조에서 알게 되겠지만, 예측정보의 가공단위를 5분으로 하였다. ‘왜 5분단위 예측인가’라는 물음에 대한 답은, 특정구간에 대한 교통정보를 가장 마지막으로 접하는 지점에서 특정구간 진입전까지 통행하는데 걸리는 시간과 동일한 시간간격으로 구간의 교통정보를 예측하여 제공하는 것이, 운전자의 정보에 대한 신뢰성을 향상시키는 방법이기 때문이다.

앞 절에서, 남산1호터널 교통정보시스템 내, VMS가 설치된 각 지점에서 터널의 정보를 제공받고, 터널을 통해 강남방향으로 진행하겠다는 의사결정을 한 후, 터널입구까지 진입하기 위한 통행시간은 평균 5분~10분 정도라고 언급한 바 있다. 터널의 정보를 제공받는 지점에서 터널에 대한 5분후의 정보를 제공받고 5분후 터널입구에 운전자가 도착하여 제공받은 정보와 같은 교통상

3)  $X_1, X_2, \dots, X_t$ 가 주어졌을 때  $X_{t+1}$ 의 분포는  $X_t$ 에만 의존함

4) 일반적으로 5분단위 예측을 말함

황을 직접 체험하게 될 때, 정보에 대한 운전자의 신뢰성은 극대화되는 것이다.

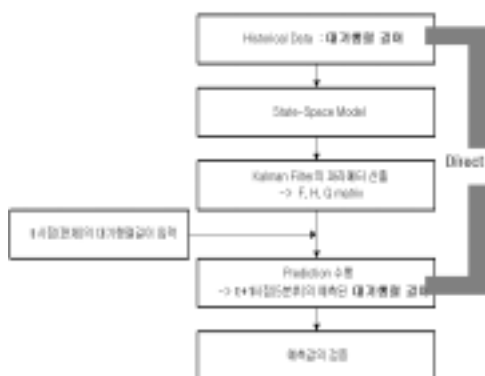
따라서, 본 연구에서는 정보가공 및 제공의 최소단위를 5분으로 하였다.

### 3) 구축모형별 구조

#### ■ 모형 1: Direct Method

모형 1은 대기행렬길이를 입력데이터로 하여 5분후 대기행렬길이 예측값을 최종적인 output으로 산출하는 구조를 가지고 있다. 즉, 입력데이터와 출력데이터가 일대일 관계를 가진다. 이러한 관계 때문에, 모형 1을 직접적 방식(Direct method)에 의한 대기행렬 예측모형이라 하였고, 이렇게 입력데이터와 출력데이터가 동일한 경우, output의 정확도가 높은 것이 일반적인 사실이다.

모형의 구조는 아래 <그림 9>와 같고, 그림에서 보면 일단 대기행렬길이 데이터를 historical 데이터로 하여 Kalman filter 연산에 필요한 파라미터 F, G, Q 매트릭스를 구하고, 구하여진 파라미터를 기초로 대기행렬의 예측치를 산출하는 과정으로 이루어져 있다.



<그림 9> 모형 1의 구조

#### ■ 모형 2: Indirect Method

모형 2는 검지기 데이터 중, 속도 데이터를 모형의 입력데이터로 하여 대기행렬길이의 예측값을 최종적인 output으로 산출하는 구조를 가지고 있다. 따라서 모형 1과는 달리 속도를 대기행렬로 바꾸는 과정이 더 추가된다. 그래서 모형 2를 간접적 방식(Indirect method)에 의한 대기행렬 예측모형이라 하였다.

모형의 구조는 아래 <그림 10>과 같고, 그림에서 보면 일단 검지기의 속도데이터를 historical 데이터로 하여 Kalman filter 연산에 필요한 파라미터 F, G, Q 매트릭스를 구하고, 구하여진 파라미터를 기초로 속도의 예측치를 산출한다. 예측된 속도를 '정체도를 이용한 대기행렬길이 산출 알고리즘'에 입력하여 대기행렬길이의 예측치를 구하는 과정으로 형성되어 있다.



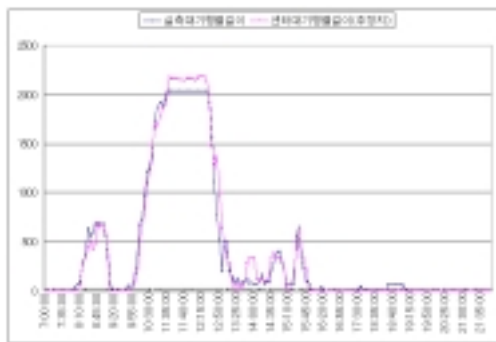
<그림 10> 모형 2의 구조

여기서, 예측된 속도로 대기행렬길이를 산출하는데, 왜 굳이 '정체도를 이용한 알고리즘<sup>5)</sup>'을 사용하려 하는지 의문이 생길 수 있다. 이 의문점에

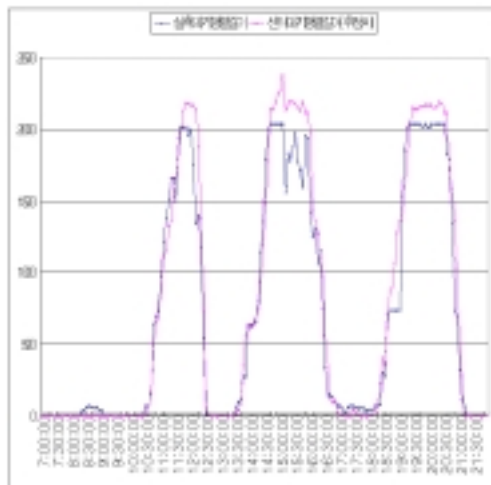
5) 각 링크에 대해 정체도를 구하여 이 정체도가 미리 정해진 임계치 이상의 값을 가질 경우 대기행렬이 생성되었다고 판단하고, 그 길이를 구하는 알고리즘. 여기서 정체도란 링크에서 최고속도에 대해 정체로 인해 저하되는 속도의 상대적인 비(rate)임 (출처: 남산1호터널 서버프로세스 검증자료).

대한 답은, 두 가지로 요약할 수 있다.

첫째, 이 알고리즘에서 추정한 대기행렬 길이가 실측된 대기행렬 길이와 비교한 결과 상당한 추정력을 가지고 있다는 사실을 발견했기 때문이다. 다음 그래프는 현재 보유하고 있는 대기행렬 실측치 데이터(2001년 3월 22일~23일 07:00~22:00)와 센터에서 추정한 대기행렬 길이를 비교한 것이다.



<그림 11> 3월22일 대기행렬길이실측치-추정치 비교



<그림 12> 3월23일 대기행렬길이실측치-추정치 비교

그래프에서 보면 추정치가 실측치에 상당히 유

사한 모양으로 그려지는 것을 알 수 있다. 이것에 좀 더 확실한 통계적 유의성을 신기 위해 ‘추정치와 실측치가 얼마나 유사한가’하는 통계적 검정을 실시하였다. 검정에 사용된 통계적 분석기법은 두 집단, 실측치와 센터추정치의 평균이 같은지를 검정하는 t-test방법이다. 그런데, 실측치와 센터추정치의 표본의 수가 07:00~22:00까지 15시간동안 5분단위, 총180개의 표본을 가지고 있으므로 대표본의 경우이고, 따라서 Z-test가 된다.

#### ■ 가설

귀무가설  $H_0 : \mu_1 = \mu_2$

대립가설  $H_1 : \mu_1 \neq \mu_2$

#### ■ 검정통계량

$$Z = \frac{(\bar{X}_1(t) - \bar{X}_2(t))}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

여기서,  $\bar{X}_1(t)$  : 실측치평균

$\bar{X}_2(t)$  : 추정치평균

#### ■ 기각영역

$$Z < -Z_{\alpha/2} \text{ 또는 } Z > Z_{\alpha/2}$$

#### ■ 가 정

두 표본들이 무작위로 두 개의 모집단들로부터 독립적인 방법으로 추출된다. 표본크기  $n_1$ 과  $n_2$ 들은 충분히 커서  $\bar{X}_1(t)$ 과  $\bar{X}_2(t)$ 들은 정규분포를 이루고 있으며, 따라서 표본분산의 값은 모분산의 값에 훌륭한 근사값을 제공한다.

표본크기가 30이상인 경우 위와 같은 것은 사실이다.

통계적 분석결과, 양측검정에서 유의수준 5%에 대한 기각역은  $R : |Z| \geq z_{\alpha/2} = 1.96$ 인데, 3월 22일 실측치와 추정치의 검정통계량  $Z = -0.132$ 이므로 귀무가설  $H_0 : \mu_1 = \mu_2$ 가 채택된다. 3

월 23일의 검정통계량은  $Z = -0.652$ 이고, 이것 또한 기각역에 해당되지 않으므로 귀무가설  $H_0 : \mu_1 = \mu_2$ 가 채택되고, 결론적으로 5% 유의수준 내에서 실측치와 추정치의 평균이 동일하다는 결론을 내릴 수 있다. 이 테스트의 결과는 다음 절 ‘적용데이터의 수집’에도 한번 더 언급됨을 미리 밝힌다.

본 연구에서 제안한 모형 2의 대기행렬 길이 예측에 ‘정체도를 이용한 대기행렬 산출 알고리즘’을 사용하고자하는 두 번째 이유는, 앞에서 언급했듯이 정체도를 이용한 대기행렬 산출 알고리즘은 현재 남산1호터널 센터에서 대기행렬 추정에 사용하고 있는 알고리즘이다. 따라서, 기존의 알고리즘을 그대로 활용하여 대기행렬길이 예측 모형으로 확장할 수 있는 이점이 있기 때문이다.

### 3. 적용 데이터의 수집

모형 1과 2의 적용에 필요한 데이터는 대기행렬길이 데이터 및 검지기에서 올라오는 속도 데이터이다. 여기서는 필요한 데이터 수집에 대하여 알아보기로 한다.

#### 1) 대기행렬길이 데이터

대기행렬길이 데이터는 실측치를 활용하는 것이 가장 바람직하나, 예측정보를 생성하는데 충분한 실측치를 구하는 것은 불가능하므로, 남산1호터널 센터에서 추정한 대기행렬길이 데이터를 활용하기로 한다. 단, 추정된 대기행렬길이를 실측치 대신으로 사용하기에 앞서, 이것이 실측치를 얼마나 잘 모사하고 있는지 통계적 분석이 선행되어야 할 것이다. 본 연구에서는 선행하는 절에

서, 통계분석으로 ‘실측치와 추정치의 평균값 차이에 대한 검정’을 실시하였고, 그 결과 실측치와 추정치 대기행렬길이의 평균값이 서로 동일하다는 결론을 내릴 수 있었다. 여기서 다시 한번 통계분석 결과를 언급하면 다음과 같다.

통계적 분석결과, 양측검정에서 유의수준 5%에 대한 기각역은  $R : |Z| \geq z_{\alpha/2} = 1.96$ 인데,

3월 22일 실측치와 추정치의 검정통계  $Z = -0.132$ 이므로,

귀무가설  $H_0 : \mu_1 = \mu_2$ 는 채택됨

3월 23일의 검정통계량은  $Z = -0.652$ 이고, 이것 또한 기각역에 해당되지 않으므로

귀무가설  $H_0 : \mu_1 = \mu_2$ 는 채택됨

결론적으로 5% 유의수준 내에서 실측치와 추정치의 평균이 동일하다는 결론을 내릴 수 있음

따라서, 본 연구에서는 모형 1의 적용에 필요한 대기행렬길이 데이터로, 실측된 대기행렬길이 대신에 추정된 대기행렬길이를 활용하기로 한다.

남산센터에서의 대기행렬길이 추정단위는 1분이다. 본 연구에서 활용되기 위해 5분단위로 가공되는데, 5분동안의 산술평균값을 5분의 대표값으로 한다.

#### 2) 속도 데이터

검지기의 속도 데이터는 남산1호터널 교통정보 시스템에서 수집된 데이터를 활용한다. 남산1호터널 센터의 검지기 자료는 polling 단위가 30초이다. 이것이 모형의 예측시간단위인 5분으로 가공되는데, 5분동안의 산술평균값을 5분의 대표값으로 한다.

#### IV. 대기행렬 예측모형 검증

##### 1. 모형적용 시나리오

모형의 적용 및 평가는 두 가지 목적을 가진다. 첫 번째는 모형 1과 2 두 가지가 실제 어느 정도의 예측력을 가지고 있는지를 평가하는 것이고, 두 번째는 두 가지 모형 중, 어느 모형이 예측력이 더 우수한지를 비교, 평가하는 것이다.

모형의 검증 및 평가의 두 가지 목적을 위한 모형검증 시나리오를 정리하면 아래의 <표 6>과 같다.

<표 6> 모형적용 시나리오

구분	Kalman filter의 parameter를 구하기 위해 적용한 historical data	예측치 생성	수행 여부
[1] 모형 1	(1) 8. 20 ~ 8. 24	① 8. 29 (수)	○
		② 3. 22 (목) (실측치)	○
[2] 모형 2	(1) 8. 20 ~ 8. 24	① 8. 29 (수)	○
		② 3. 22 (목) (실측치)	×

시나리오 [1]-(1)-①과 [2]-(1)-①은 2001년 8월 20일 월요일부터 8월24일 금요일까지 5일치 검지기 속도자료 및 대기행렬 추정치를 활용하여 Kalman filter 연산에 필요한 파라미터를 산정, 이를 이용해서 8월29일 수요일 대기행렬길이 예측치를 산출하는 시나리오이다. 시나리오 [1]-(1)-②와 [2]-(1)-②은 구축된 모형 1과 2가 대기행렬 추정치가 아닌 실측치를 입력값으로 넣어 예측치를 산출하려고 할 경우 실측값에 얼마나 잘 맞아 떨어지는지를 보기 위한 것이다. 이때, 시나리오 [1]-(1)-②의 경우에는 보유하고 있는 3월 22일

대기행렬 실측치 데이터를 가공하여 입력데이터로 활용, 모형 1을 적용하는데 별 무리가 없었다. 이때 활용된 3월 22일의 대기행렬 실측치는, 본 연구에서 대기행렬의 시점으로 보는 ‘한남2고가’부터 ‘구름다리(톨게이트 후방)’지점까지 몇 개의 구간으로 나누어 비디오 촬영, 분석하여 얻었다. 시나리오 [2]-(1)-②의 경우, 모형 2의 입력데이터로 활용될 3월 22일 검지기 데이터가 검지기의 불안정으로 대상구간 내, 특정 지점의 검지기 데이터가 완전히 누락되어 있는 등, 몇 가지 데이터상의 문제점으로, 모형 2를 적용하지 못하였다.

##### 2. Kalman filter의 파라미터 산출

파라미터 산정을 위해 5분단위로 가공된 historical 데이터를 SAS 프로그램의 statespace procedure에 입력하여 파라미터 F, G, Q를 구하였다. 모형별, 시나리오별 구해진 파라미터 값은 아래의 행렬형태로 나타난다.

1) 모형 1의 Kalman filter 파라미터 : 8월20일 ~ 24

일 데이터를 기초한 파라미터

$$F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.686 & 0.535 & 1.130 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.182 \\ 1.185 \end{bmatrix}, \quad Q = [19415.75]$$

2) 모형 2의 Kalman filter 파라미터 : 8월20일 ~

24일 데이터를 기초한 파라미터

① 100번지점 검지기

$$F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.118 & -0.193 & 1.282 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 0.890 \\ 0.569 \end{bmatrix}, Q = [7.167]$$

② 110번지점 검지기

$$F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.134 & -0.914 & 1.770 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 0.900 \\ 0.665 \end{bmatrix}, Q = [6.768]$$

③ 120번지점 검지기

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.073 & 0.658 & -1.078 & 1.270 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.089 \\ 1.004 \\ 0.996 \end{bmatrix}, Q = [22.205]$$

④ 130번지점 검지기

$$F = \begin{bmatrix} 0 & 1 \\ 0.015 & 0.932 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.099 \end{bmatrix}, Q = [35.221]$$

⑤ 140번지점 검지기

$$F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.278 & -0.305 & 0.948 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.175 \\ 1.086 \end{bmatrix}, Q = [35.808]$$

⑥ 150번지점 검지기

$$F = \begin{bmatrix} 0 & 1 \\ -0.184 & 1.118 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.258 \end{bmatrix}, Q = [47.449]$$

⑦ 160번지점 검지기

$$F = \begin{bmatrix} 0 & 1 \\ -0.323 & 1.267 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.337 \end{bmatrix}, Q = [44.462]$$

⑧ 170번지점 검지기

$$F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.216 & -0.060 & 1.229 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.266 \\ 1.226 \end{bmatrix}, Q = [57.786]$$

⑨ 180번지점 검지기

$$F = \begin{bmatrix} 0 & 1 \\ -0.214 & 1.135 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.341 \end{bmatrix}, Q = [55.789]$$

⑩ 190번지점 검지기

$$F = \begin{bmatrix} 0 & 1 \\ -0.225 & 1.172 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.286 \end{bmatrix}, Q = [56.067]$$

⑪ 200번지점 검지기

$$F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.477 & 0.128 & 1.333 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.113 \\ 1.043 \end{bmatrix}, Q = [42.695]$$

⑫ 210번지점 검지기

$$F = \begin{bmatrix} 0 & 1 \\ 0.423 & 0.498 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 \\ 1.107 \end{bmatrix}, Q = [45.649]$$

### 3. 시나리오별 모형검증

#### 1) 평가지표

일반적으로 구축된 예측모형의 경우 그 수행결과를 평가하는 척도는 다양하며, 아직까지 절대적인 평가기준은 없으며, 단지 다음과 같은 지표를 토대로 상대적인 평가에 의존한다.

MARE(Mean Absolute Relation Error)

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|x(i) - \widehat{x(i)}|}{x(i)}$$

MAE(Mean Absolute Error)

$$MAE = \frac{1}{N} \sum_{i=1}^N |x(i) - \widehat{x(i)}|$$

이러한 지표중 널리 이용되는 지표는 MARE이다. 본 연구에서는 상기의 제시된 각 지표들을 기준으로 구축된 모형의 효율성을 평가한다.

## 2) 대기행렬길이 예측 평가결과

모형 적용의 대상네트워크인 남산1호터널은 연속류의 특성을 가지는 네트워크이므로, 신호에 의해 제어되는 단속류처럼 대기행렬길이의 작은 변동에 민감하게 반응하지는 않는다. 그리고 생성되는 대기행렬의 길이도 신호교차로의 경우, 평균적으로 최고 200m정도<sup>6)</sup>인 것과는 달리 최고 2,500m까지 형성되는 등, 도시부 신호교차로에서 형성되는 대기행렬 길이와는 비교도 안될 만큼 길다. 그래서, 전체 대기행렬 길이 2500m에 오차 100m~150m 정도는 운전자에게 큰 영향을 미치지 않을 것으로 판단된다.

따라서 본 연구에서는 대기행렬길이 산정 모형의 오차허용범위로 100~150m로 설정하였다. 다음 <표 7>은 현재 보유하고 있는 2001년 3월 22일 대기행렬 실측치 데이터를 기준으로 앞 절에서 설명한 평가지표 값을 산정한 것이다. 여기서 구해진 평가지표 값을 기준으로, 시나리오별로 적용한 결과값을 비교하기로 한다.

<표 7> 평가지표 기준치 (3월 22일 대기행렬 실측치 기준)

	MARE	MAE
대기행렬길이 허용오차 100m인 경우	0.976	100
대기행렬길이 허용오차 150m인 경우	1.464	150

## ■ 시나리오[1] - (1) - ② : 3월 22일 대기행렬길이 실측치-추정치-예측치 비교분석 결과

실측치와 비교했을 때, 남산센터의 추정치가 모형 1에 의한 예측치보다 정확한 값으로 산출되었다.

<표 8> 3월 22일 대기행렬길이 실측치-추정치-예측치 비교분석결과 : 시나리오 [1]-②-②

	MARE	MAE
남산센터 추정치	0.309	53.894
Kalman filter 예측치, 모형 1 (Historical Data 5일치 활용)	1.566	125.247



<그림 13> 5일치 historical data 활용 결과 시나리오 [1]-①-②)

그런데 여기서 위의 그래프를 살펴보면 대기행렬의 길이가 길 때가 짧을 때보다, 예측치의 정확도가 더 높은 것으로 판단되며, 따라서 대기행렬 길이가 500m 이상(10:00~16:00)일 경우에 한해서 실측치-추정치-예측치를 비교해 보았다.

<표 9> 500m 이상 대기행렬이 생기는 경우에 한해 (10:00 ~ 16:00) 실측치-추정치-예측치 비교분석 결과

	MARE	MAE
남산센터 추정치	0.472	110.067
Kalman filter 예측치, 모형 1 (Historical Data 5일치 활용)	1.129	220.691

6) 허낙원, 루프검지기를 이용하여 점유시간 수집주기 및 차종을 고려한 대기행렬 길이 예측 알고리즘, 석사학위논문, 아주대학교, 2001

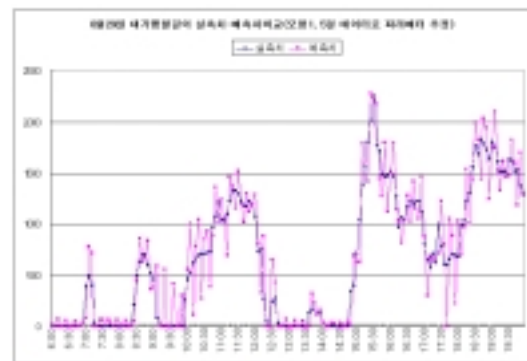
<표 9>에서 보면, 남산센터의 추정치는 500m 이상인 경우 평가지표 MARE가 0.309에서 0.472로 증가하였으나, 예측치는 1.566에서 1.129로 감소하였다. 따라서 본 연구에서 제시한 모형은 대기행렬이 일정한 길이 이상으로 형성되었을 경우 더욱 정확한 값을 산출하는 것으로 나타났다. 이것은 좀더 정확한 예측정보를 제공하는 하나의 방법으로 제시할 수 있다. 다시 말해, ‘대기행렬이 500m 이상인 경우에만 대기행렬 예측정보를 제공하기로 한다’식의 정보제공 방법을 제안할 수 있다.

#### ■ 시나리오[1]-(1)-①, [2]-(1)-①: 8월 29일 모형별 대기행렬길이 예측치-센터추정치 비교 분석 결과

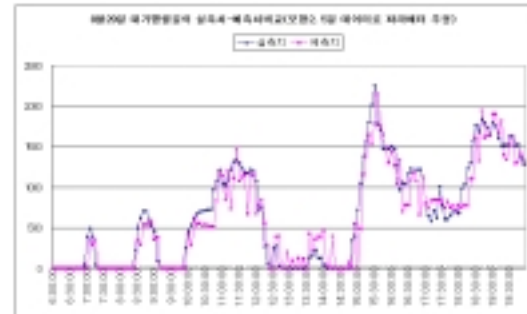
시나리오 [1]-(1)-①, [2]-(1)-①에 의해 비교분석한 결과, 모형 1이 모형 2보다 평가지표 값이 우수하다. 이는 모형 1의 구조가 대기행렬길이를 직접 예측하는 구조를 가지고 있기 때문인 것으로 판단된다. 그리고 모형 1의 경우 시나리오 [1]-(1)-②에서의 결과치보다 MARE 값이 향상된 것으로 나타나는데, 이것은 시나리오 [1]-(1)-②는 8월 historical 데이터를 활용하여 파라미터를 산정, 3월 22일 대기행렬길이를 입력데이터로 하여 예측치를 산출하였기 때문이다. 다시말해, 3월에서 8월 사이에 있을 수 있는 교통상황의 월별, 계절적 변동을 무시한 채, 8월 historical 데이터를 바탕으로 3월의 대기행렬길이를 예측한다는 것은, 상식적으로 생각해도 좋은 결과를 내기는 어려운 것이다.

<표 10> 8월 29일 모형별 대기행렬길이 예측치-센터추정치 비교분석결과 : 시나리오 [1]-(1)-①, [2]-(1)-①

	MARE	MAE
모형1	0.250	155.968
모형2	0.307	153.384



<그림 14> 시나리오 [1]-(1)-①적용결과 (모형1)



<그림 15> 시나리오 [2]-(1)-① 적용결과 (모형2)

## 4. 평가결과 정리

시나리오별로 수행한 결과를 요약정리하면 다음과 같다.

- ① 대기행렬이 500m 이상 형성된 경우에 한해서, 예측치-실측치, 남산센터 추정치-실측치

각각을 비교하였더니, 모형 1의 경우 500m 이상 형성되었을 때가 관측시간 전체에 대한 비교보다 향상된 결과를 보였음. 그러나 남산터널센터 추정치의 경우에는 이와는 상반된 결과를 보임

- ② 이로써, 좀더 정확한 대기행렬 예측정보를 제공하기 위해서 '대기행렬이 500m 이상인 경우에만 예측정보를 제공하기로 한다'는 식의 정보제공 방법을 제안할 수 있음
- ③ 8월20일~24일, 5일치 데이터를 historical 데이터로 활용하여 구한, 모형 1과 2의 대기행렬 예측치는, 대기행렬길이 허용오차 100m인 경우의 평가지표 MARE는 모두 만족하나 나머지 지표인 MAE는 벗어나는 결과를 보임

## V. 결론 및 향후 과제

본 연구는 남산1호터널 교통정보시스템에서 수집되는 교통정보를 이용하여, 남산터널 네트워크에 형성되는 대기행렬길이를 예측하는 알고리즘 개발을 수행하였다.

대기행렬길이 예측을 수행한 결과, 100m~150m 오차허용 범위 내에서 비교적 신뢰할 만한 결과값이 산출되었다. 그러나 이 값의 신뢰성을 더욱 높이기 위해서는 향후, 다음의 몇 가지 작업들이 추가로 수행되어야 할 것이다.

- ① 여기서는 일단, Kalman filter 연산에 필요한 historical 데이터로 8월20일~24일까지 5일치 데이터만을 활용하였으나, 더 많은 historical 데이터를 사용하였을 경우 결과값이 어떻게 나오는지의 수행이 필요함
- ② 만일 historical 데이터를 많이 사용할수록 더

좋은 결과값을 낸다면 오랜 시간 축적된 historical 데이터를 예측에 활용하도록 함

- ③ 방대한 historical 데이터를 활용할 경우, historical 데이터들 사이에서 일정한 패턴을 가지는 것들끼리 grouping하여 예측에 활용한다면 더욱 정확도가 높은 예측치를 산출할 수 있을 것으로 예상됨

## 참고문헌

- 김남선(1998), 「신경망을 이용한 고속도로 여행시간 추정 및 예측모형 개발」, 석사학위논문, 아주대학교
- 김상욱(1996), 「Kalman Filtering 모델을 이용한 고속도로상에서의 교통량 추정」, 석사학위논문, 아주대학교
- 노재현 외(1999), 「고속도로 대기행렬길이 산정모형 개발을 위한 연속류 특성분석」, 대한교통학회지, 제17권, 제2호, pp. 179~191
- 서울시정개발연구원(2000), 「서울시 첨단교통정보체계(ATIS)구현을 위한 기본연구」
- 이종현, 최기현(1994), 「시계열 분석과 그 응용, 자유아카데미」
- 정철훈(1993), 「Kalman Filter에 의한 신호교차로 단기 교통변화예측」, 석사학위논문, 아주대학교
- 한종학(1997), 「인공신경망을 이용한 신호교차로 대기행렬 예측모형에 관한 연구」, 석사학위논문, 한양대학교
- 허낙원(2001), 「루프검지기를 이용하여 점유시간 수집 주기 및 차종을 고려한 대기행렬길이 예측 알고리즘」, 석사학위논문, 아주대학교
- 허문열(1989), 「수리통계학」, 경문사
- 최종욱, 민준영, 남궁성, 이원하(1997), 「ITS와 첨단정보기술」, 참말
- SK씨앤씨, 남산1호터널 서버프로세스 검증자료
- Daganzo et al, 「Using Input-Output Diagram to Determine Spatial and Temporal Extents of Queue Upstream of a Bottleneck」, TRR 1572, pp.140~147
- Daganzo(1997), 「Fundamentals of Transportation and

- Traffic Oprations」
- Diley(1997), 「Travel Time Estimates Using a Series of Single Loop Volume and Occupancy Measurements」, Transportation Research Board 7th Annual Meeting
- Drew(1968), 「Traffic Flow Theory and Control」
- Hoogendoorn(2001), 「Model-Based Multiclass Travel Time Estimation」, 9th World Conference on Transportation Research
- May(1990), 「Traffic Flow Fundamentals」
- McShane, 「Traffic Engineering」
- Newell(1993), 「A Simplified Theory of Kinematic Waves in Highway Traffic, Part 1: General Theory」, Transportation Research, Vol. 27B, No. 4, pp281~287
- Seungjae Lee et al(1998), 「Comparison of Model for Prediction Short-Term Travel Speeds」, Proceedings of 5th World Congress on Intelligent Transport Systems, Seoul, Korea
- Welch and Bishop(2001), 「An Introduction to the Kalman Filter」, UNC-Chapel Hill, TR 95-041
- Yim, Youngbin and Ygnace, Jean-luc, 「Link Flow Evaluation Using Loop Detector Data: Traveler Responses to Variable Message Signs」, TRR 1550, pp.58-64