

■ 論 文 ■

## 버스지체시간을 활용한 버스도착시간 예측

### The Bus Arrival Time Prediction Using Bus Delay Time

이 승 훈

(한국건설기술연구원  
첨단교통연구실 연구원)

문 병 섭

(한국건설기술연구원  
첨단교통연구실 선임연구원)

박 범 진

(한국건설기술연구원  
첨단교통연구실 연구원)

#### 목 차

- |  |  |
|--|--|
| <p>I. 서론</p> <p>1. 연구의 배경 및 목적</p> <p>2. 연구의 범위</p> <p>II. 버스지체시간 예측</p> <p>1. 데이터 보정</p> <p>2. 버스지체시간 산출</p> <p>3. 전이확률행렬표 생성</p> | <p>4. 버스지체시간 예측</p> <p>III. 버스도착시간 예측 및 평가</p> <p>1. 버스도착시간 예측</p> <p>2. 예측한 버스도착시간 평가</p> <p>IV. 결론</p> <p>참고문헌</p> |
|--|--|

Key Words : 버스도착시간, 버스정보시스템, 전이확률행렬표, 평균절대오차, 평균절대상대오차  
 Bus Arrival Time, BIS; Bus Information System, Transition Probability Matrix, MAE;  
 Mean Absolute Error, MARE; Mean Absolute Relative Error

#### 요 약

버스도착시간은 배차간격에 맞춰 차고지에서 출발한 버스가 해당정류장에 도착하는 시간을 말하며, 승하차 시간, 신호 주기, 버스전용차로의 유무 등 여러 가지 교통여건으로 인하여 정류장에 도착할 때는 어느 정도의 오차를 발생시킨다. 본 연구에서는 다양한 교통여건을 반영하는 버스지체시간을 산출하여 정류장별 버스지체시간을 예측하고, 이를 이용하여 정류장별 버스도착시간을 예측하였다. 그 결과 본 연구의 조건과 같은 경우, 7×7 행렬과 9×9 행렬을 이용하여 버스도착시간을 예측하였을 때 분석대상도시에서 기존에 사용 중인 가중이동평균법을 이용한 버스도착시간예측방법 보다 높은 정확도를 얻을 수 있었다.

It is occurred bus arrival time errors when a bus arrives at a bus stop because of a variety of traffic condition such as traffic signal cycle, the time to get on and off a bus, a bus-only lane and so on. In this paper, bus delay time which is occurred as the result of traffic condition was estimated with Markov Chain process and bus arrival time at each bus stop was predicted with it. As the result of the study, it is confirmed to improve accuracy than the method of bus arrival time prediction with existing method (weighed moving average method) in case predicting bus arrival time using 7 by 7 and 9 by 9 matrixes.

## 1. 서론

### 1. 연구의 배경 및 목적

버스도착시간(Bus Arrival Time)은 배차간격에 맞춰 차고지에서 출발한 버스가 해당정류장에 도착하는 시간을 말하며, 승하차 시간, 신호 주기, 버스전용차로의 유무 등 여러 가지 교통여건으로 인하여 정류장에 도착할 때는 어느 정도의 오차를 발생시킨다.

버스도착(통행)시간을 예측하기 위해 초기에는 원리가 단순하고 구조가 간단한 회귀 모형을 이용하였으나 (Abdelfattah 외, 1998), 이 방법은 예측의 정확도를 높이기 위해 많은 변수를 고려해야 하고 특별한 교통상황이 발생했을 경우, 이에 대한 대처 능력이 떨어지는 단점을 가지고 있다. 그리고 인간 두뇌와 같은 구조의 신경망 모형(Chien 외, 2002)과 반복적 알고리즘인 Kalman Filter 기법(Shalaby 외, 2003) 또한 버스도착시간을 예측하기 위해 사용되었지만 이들 방법 역시 버스도착시간을 예측함에 있어 결과에 대한 설명이 어렵고 자료가 부족할 경우 처리가 어려운 한계점을 가지고 있다.

최근에는 앞에서 언급한 방법보다 단순하고 정확도가 우수한 가중평균이동법(김영찬 외, 2005)이 버스정보시스템(Bus Information System)을 도입한 여러 지자체(안양시, 제주특별자치도 등)에서 주로 사용되고 있다. 하지만 이 방법은 버스도착시간의 예측값에 대한 오차 때문에 전 정류장과 전전 정류장을 출발한 버스의 경우 버스도착시간 정보 대신에 “전 정류장 출발”, “전전 정류장 출발” 등으로 표시하여 정보를 제공하고 있다. 또한, 지자체별로 버스전용차로의 적용이 확대되면서 버스도착시간 예측을 위한 교통여건에 관한 변수들의 영향이 줄어들어 기존의 예측방법으로는 정확도를 향상시키기 어려워졌다.

본 연구의 목적은 버스노선의 교통여건이 반영되어 나타나는 버스지체시간을 사용하여 정류장별 버스지체시간을 예측하고, 이를 이용하여 정류장별 버스도착시간을 예측하는 것이다. 이를 위해 본 연구에서는 마코브 체인(Markov Chain)을 버스지체시간 예측을 위해 사용하였고, 적용된 방법론을 검증하기 위하여 분석대상도시에서 사용하고 있는 방법(가중이동평균법)의 결과와 비교하였다.

<그림 1>은 본 연구에서 수행된 연구의 흐름도를 나타낸다.



<그림 1> 연구의 흐름도

### 2. 연구의 범위

본 연구에서는 A시에서 운영 중인 한 노선(이 구간은 편도 5차선 도로로 시간대별 버스전용차선을 시행중에 있음)에 대하여 2주간의 07시 ~ 23시 데이터를 이용하여 4가지 행렬(3×3, 5×5, 7×7, 9×9 행렬)별로 전이확률행렬표를 생성하고 이를 활용하여 정류장별 버스 지체시간을 예측하였으며, 이 예측된 버스지체시간과 배차간격을 이용하여 정류장별 버스도착시간을 예측하였다.

또한 예측한 버스도착시간의 정확도를 살펴보기 위하여 현재 분석대상도시에서 사용 중인 가중평균이동법과 MAE, MARE값을 비교분석하였다.

## II. 버스지체시간 예측

### 1. 데이터 보정

본 연구는 앞 뒤차간의 버스도착시간 차를 이용하여 버스지체시간을 산출하였기 때문에 누락된 버스도착시간이 존재할 경우 이에 대한 보정이 먼저 필요하다. 따라서 누

락데이터를 가진 버스에 대하여 다음 정류장 또는 전 정류장의 버스도착시간이 존재한다면, 누락된 버스도착시간은 앞차의 정류장간 통행시간과 현재 차량의 다음 정류장 또는 전 정류장의 도착시간을 이용하여 산출하였다.

$$T_{n,i} = T_{n+1,i} - (T_{n+1,i-1} - T_{n,i-1})$$

or  $T_{n,i} = T_{n-1,i} + (T_{n,i-1} - T_{n-1,i-1})$  (1)

- $T_{n,i}$  : n 정류장에서 i 번째 버스의 도착시간
- $T_{n+1,i}$  : n+1 정류장에서 i 번째 버스의 도착시간
- $T_{n+1,i-1}$  : n+1 정류장에서 i-1 번째 버스의 도착시간
- $T_{n,i-1}$  : n 정류장에서 i-1 번째 버스의 도착시간
- $T_{n-1,i}$  : n-1 정류장에서 i 번째 버스의 도착시간
- $T_{n-1,i-1}$  : n-1 정류장에서 i-1 번째 버스의 도착시간

2. 버스지체시간 산출

기존연구에서는 버스지체시간을 산출하기 위하여 계획된 버스도착시간에 대한 실제 버스도착시간의 차를 이용하였다.(Lin 외., 1999; Lin 외, 2004; Rajat, 2005) 하지만 본 연구의 분석대상 도시의 경우 정류장별 계획된 버스도착시간이 존재하지 않기 때문에 기존방법을 이용하여 버스지체시간을 산출하기에는 한계가 있다. 따라서 본 연구에서는 계획된 버스도착시간 대신 분석 대상버스의 배차간격을 이용하여 버스지체시간을 산출하였다. 즉, 버스지체시간은 주어진 배차간격에 대한 앞뒤차간의 도착시간차로 산출하였다.(이승훈 외, 2009)

$$AD_{n,i} = (T_{n,i} - T_{n,i-1}) - I \quad (i \geq 2)$$
 (2)

- $AD_{n,i}$  : n 정류장에서 i 번째 버스의 지체시간
- $T_{n,i}$  : n 정류장에서 i 번째 버스의 도착시간
- $T_{n,i-1}$  : n 정류장에서 i-1 번째 버스의 도착시간
- $I$  : 배차간격

3. 전이확률행렬표 생성

전이확률행렬표(Transition Probability Matrix)를 생성하기 위해서는 먼저 다양한 값을 가지는 버스지체시간에 대한 분류가 필요하다. 따라서 버스지체시간을 분류하기 위하여 본 연구에서는 전체 버스지체시간에 대한 지체시간 범위를 최대값과 최소값의 차로 산출하였다.

$$DR = AD_{max} - AD_{min}$$
 (3)

- $DR$  : 실제 버스지체시간의 범위
- $AD_{max}$  : 실제 버스지체시간의 최대값
- $AD_{min}$  : 실제 버스지체시간의 최소값

식(3)을 이용하여 2주간의 데이터에 대한 일별 버스지체시간 범위를 구하고 이를 평균하여 2주 전체에 대한 평균 버스지체시간 범위를 산출한 결과, 평균 버스지체시간 범위는 약 19분 27초였으며, 이를 이용하여 버스지체시간의 계급간격을 산출하였다. 즉, 지체시간의 계급간격은 분석 데이터의 범위를 계급의 수로 나누어서 산출하였다.(고승영 외, 1998)

$$DI = \frac{DR}{k}$$
 (4)

- $DI$  : 지체 간격
- $DR$  : 지체 범위
- $k$  : 계급의 수

앞에서 산출한 버스지체시간 계급간격을 이용하여 버스지체구간을 설정하면 <표 2>와 같으며, 이를 이용하여 산출한 버스지체시간을 분류하였다.

<표 1> 적용한 지체시간의 계급간격

구분	지체 범위	계급의 수	계산된 계급간격	적용한 지체시간의 계급간격
3×3 행렬	19분 27초	3	6분 29초	6분
5×5 행렬		5	3분 53초	4분
7×7 행렬		7	2분 47초	3분
9×9 행렬		9	2분 10초	2분

<표 2> 적용한 버스지체구간

구분	적용한 버스지체구간
3×3 행렬	'-3분 이하', '-3분 ~ +3분', '+3분 이상'
5×5 행렬	'-6분 이하', '-6분 ~ -2분', '-2분 ~ +2분', '+2분 ~ +6분', '+6분 이상'
7×7 행렬	'-7.5분 이하', '-7.5분 ~ -4.5분', '-4.5분 ~ -1.5분', '-1.5분 ~ +1.5분', '+1.5분 ~ +4.5분', '+4.5분 ~ +7.5분', '+7.5분 이상'
9×9 행렬	'-7분 이하', '-7분 ~ -5분', '-5분 ~ -3분', '-3분 ~ -1분', '-1분 ~ +1분', '+1분 ~ +3분', '+3분 ~ +5분', '+5분 ~ +7분', '+7분 이상'

5×5 행렬에 대한 버스지체시간의 경우 5가지 버스지체구간으로 분류하였다. 따라서 현재 상태에서 다음 상태로의 버스지체의 전이는 5가지 경우, ‘-6분 이하’, ‘-6분 ~ -2분’, ‘-2분 ~ +2분’, ‘+2분 ~ +6분’, ‘+6분 이상’으로 전이될 수 있다.

만약,  $n$ 정류장에 ‘-6분 이하’ 데이터가 10대가 존재할 때,  $n+1$ 정류장에서 ‘-6분 이하’ 데이터가 3대가 존재한다고 가정한다면,  $P_{11}$ 의 값은  $3/10=0.3$ 으로 정의할 수 있다. 이와 같은 방식을 이용하여 생성한 정류장간 5×5 행렬의 전이확률행렬표는 <표 3, 4, 5, 6>과 같다.

<표 3> 5×5 행렬(찬우물 → 갈현동부대)

Delay	-6분 이하	-6분 ~ -2분	-2분 ~ +2분	+2분 ~ +6분	+6분 이상
-6분 이하	0.83413	0.16587	0.00000	0.00000	0.00000
-6분 ~ -2분	0.01414	0.93255	0.05330	0.00000	0.00000
-2분 ~ +2분	0.00000	0.05049	0.92061	0.02891	0.00000
+2분 ~ +4분	0.00000	0.00000	0.11935	0.85492	0.02573
+6분 이상	0.00000	0.00000	0.00000	0.01786	0.98214

<표 4> 5×5 행렬(갈현동부대 → 인덕원사거리)

Delay	-6분 이하	-6분 ~ -2분	-2분 ~ +2분	+2분 ~ +6분	+6분 이상
-6분 이하	0.53458	0.46542	0.00000	0.00000	0.00000
-6분 ~ -2분	0.00668	0.85176	0.14156	0.00000	0.00000
-2분 ~ +2분	0.00000	0.07802	0.88295	0.03903	0.00000
+2분 ~ +4분	0.00000	0.00000	0.21020	0.72774	0.06206
+6분 이상	0.00000	0.00000	0.00000	0.09399	0.90601

<표 5> 5×5 행렬(인덕원사거리 → KT동안양점)

Delay	-6분 이하	-6분 ~ -2분	-2분 ~ +2분	+2분 ~ +6분	+6분 이상
-6분 이하	0.90278	0.09722	0.00000	0.00000	0.00000
-6분 ~ -2분	0.01924	0.94137	0.03938	0.00000	0.00000
-2분 ~ +2분	0.00000	0.06671	0.90797	0.02532	0.00000
+2분 ~ +4분	0.00000	0.00000	0.01639	0.92850	0.05511
+6분 이상	0.00000	0.00000	0.00000	0.10853	0.89147

<표 6> 5×5 행렬(KT동안양점 → 오뚜기식품)

Delay	-6분 이하	-6분 ~ -2분	-2분 ~ +2분	+2분 ~ +6분	+6분 이상
-6분 이하	0.55272	0.44728	0.00000	0.00000	0.00000
-6분 ~ -2분	0.03404	0.90001	0.06595	0.00000	0.00000
-2분 ~ +2분	0.00000	0.13808	0.80128	0.06064	0.00000
+2분 ~ +4분	0.00000	0.00000	0.04653	0.91599	0.03748
+6분 이상	0.00000	0.00000	0.00000	0.10995	0.89005

생성된 전이확률을 살펴보면 같은 조건의 버스지체구간 값을 가지는 행렬값(정중앙 행렬값을 포함하여 왼쪽에서 오른쪽 아래로 대각선 방향의 행렬값)을 중심으로 높은 행렬값을 나타냈으며 대각선 축을 중심으로 멀어질수록 0값을 나타내는 동질한 형태의 분포를 나타냈다.

4. 버스지체시간 예측

본 연구에서는 마코브 체인에서  $n$ 단계 전이확률을 구하기 위하여 사용하는 Chapman-Kolmogorov 방정식(Sheldon, 2000)을 사용하였으며, 이를 위해 버스정류간 버스지체의 전이는 동질하다는 가정을 하였다. 이 가정을 통하여  $n$ 번째 정류장의 전이확률을 구하면  $P(1) \times P(2) \times \dots \times P(n-1) \times P(n) = P(1)^n$ 과 같으며, 이를 이용하여  $n$ 번째 정류장의 버스지체시간은 식(5)를 이용하여 예측 할 수 있다.

$$ED_n = P_0 \times \{P(1) \times \dots \times P(n)\} \times DS^T \quad (5)$$

$$= P_0 \times \{P(1)^n\} \times DS^T$$

$ED_n$  :  $n$ 정류장에서 예측한 버스지체시간(초)

$P_0$  : 초기 지체상태행렬

$P(n)$  :  $n$ 번째 정류장 전이확률행렬

$DS^T$  : 지체상태행렬의 전치행렬

즉,  $n$ 번째 정류장에서 예측한 버스지체시간은 초기 지체상태행렬( $P_0$ ), 전이확률행렬( $P(n)$ ), 지체상태 전치행렬( $DS^T$ )의 곱으로 나타낼 수 있다. 초기 지체상태는 현재 정류장에서 지체의 상태를 의미하며, 5×5 행렬의 경우 <표 7>과 같이 표현할 수 있다.

<표 7> 초기 지체상태행렬 예(5×5 행렬)

현재 정류장 지체상태	초기 지체상태행렬
-6분 이하	{1 0 0 0 0}
-6분 ~ -2분	{0 1 0 0 0}
-2분 ~ +2분	{0 0 1 0 0}
+2분 ~ +6분	{0 0 0 1 0}
+6분 이상	{0 0 0 0 1}

<표 8> 적용된 지체상태 행렬

구분	지체구간 간격	적용된 지체상태 행렬(분)
3×3 행렬	6분	{-6 0 6}
5×5 행렬	4분	{-8 -4 0 4 8}
7×7 행렬	3분	{-9 -6 -3 0 3 6 9}
9×9 행렬	2분	{-8 -6 -4 -2 0 2 4 6 8}

<표 9> 정류장별 예측한 버스지체시간 예(초)\_5x5 행렬

구분		①	①	①	①	②	②	②	③	③	④
		→ ②	→ ③	→ ④	→ ⑤	→ ③	→ ④	→ ⑤	→ ④	→ ⑤	→ ⑤
- 6분 이하	Total	-441	-406	-375	-348	-369	-294	-240	-457	-435	-373
	Week	-450	-423	-397	-374	-369	-295	-242	-445	-415	-348
	Weekend	-413	-363	-325	-295	-371	-295	-241	-480	-480	-407
-6분~-2분	Total	-231	-221	-212	-203	-208	-180	-156	-235	-231	-232
	Week	-229	-219	-208	-199	-209	-182	-160	-235	-229	-234
	Weekend	-234	-226	-217	-207	-207	-177	-150	-236	-232	-231
-2분~+2분	Total	2	2	1	-1	0	-2	-4	-3	-8	-4
	Week	3	4	4	3	-7	-13	-19	-7	-13	-10
	Weekend	0	-2	-3	-6	13	19	22	-2	-3	7
+2분~+6분	Total	416	363	317	277	365	282	221	460	441	449
	Week	410	353	305	264	361	276	213	457	438	458
	Weekend	425	376	334	296	363	282	225	464	448	441
+6분 이상	Total	245	248	250	251	263	272	273	266	287	267
	Week	248	253	256	257	264	273	273	268	289	268
	Weekend	240	240	240	240	263	271	273	270	293	270

\* ① : 찬우물, ② : 갈현동 부대, ③ : 인덕원사거리, ④ : KT동안양점 ⑤ : 오투기식품

또한, 지체상태행렬(DS)은 버스지체시간에 대한 예측 값을 얻기 위해 필요한 행렬로서 본 연구에서는 버스지체 구간의 중간 값과 구간 범위를 이용하여 도출하고, 이 행렬의 전치행렬을 버스지체시간을 예측하기 위하여 사용하였다. 예를 들어 5x5 행렬의 경우 “-2분 ~ +2분” 상태의 중간 값인 0을 기준으로 4분 간격의 지체구간 범위를 적용하여 {-8, -4, 0, 4, 8}의 행렬 값을 적용하였다. 연구에 사용한 지체상태행렬은 <표 8>과 같다.

앞에서 설명한 방법을 이용하여 현재 정류장에서 n 정류장의 버스지체시간을 예측한 결과는 <표 9>와 같다.

### III. 버스도착시간 예측 및 평가

#### 1. 버스도착시간 예측

버스도착시간을 예측하기 위하여 본 연구에서는 예측한 버스지체시간, 배차간격, 그리고 전 버스의 도착시간을 이용하여 n 정류장에서 i 버스의 버스도착시간을 예측하였다.

$$\widehat{T}_{n,i} = T_{n,i-1} + I + ED_{n,i} \quad (6)$$

$\widehat{T}_{n,i}$  : n 정류장에 i 번째 버스의 도착예정시간

$T_{n,i-1}$  : n 정류장에 i-1 번째 버스의 도착시간

I : 배차간격

$ED_{n,i}$  : n 정류장에 i 번째 버스의 예측된 지체시간

<표 10> 버스도착시간 예측 예

구분	정류장1	정류장2	정류장3	정류장4	정류장5
버스1	17:02:10	17:02:41	17:05:45	17:07:41	17:09:23
버스2	17:12:49	-	-	-	A

- 배차간격 : 8분

- 정류장5에서 버스2의 예측된 지체시간 : 1분

예를 들어, <표 10>과 같이 배차간격이 8분이고, 예측한 버스지체시간이 1분이면 정류장5에서 버스2의 도착시간(A)은 다음과 같이 산출 할 수 있다.

$$\begin{aligned} \widehat{T}_{n,i} &= T_{n,i-1} + I + ED_{n,i} \\ &= 17:09:23 + 00:08:00 + 00:01:00 \\ &= 17:18:23 \end{aligned}$$

#### 2. 예측한 버스도착시간 평가

##### 1) 가중이동평균법

예측한 버스도착시간 평가에 앞서 기존 버스도착시간 예측방법인 가중평균법을 살펴보면 다음과 같다. 가중이동평균법은 해당 구간을 통과한 이전 차량의 통행시간을 가중 평균하여 통행시간을 예측하는 방법으로, 단순하고 구현이 용이하며 신뢰도가 높다는 장점이 있다.(이승훈, 2009)

$$\overline{RT}_{Link} = \frac{\alpha_1 \times RT_1 + \alpha_2 \times RT_2 + \dots + \alpha_n \times RT_n}{\alpha_1 + \alpha_2 + \dots + \alpha_n} \quad (7)$$

$$\alpha_n = \frac{T_n - T_a}{T_w}$$

$\overline{RT}_{Link}$  : 버스구간 평균 통행시간  
 $RT_n$  : n번째 차량의 버스구간 통행시간  
 $\alpha_n$  : 가중치  
 $T_n$  : n번째 차량의 정보수집시간  
 $T_a$  : 정보가공시간  
 $T_w$  : 정보가공 주기

2) 평가척도

본 연구에서는 예측한 버스도착시간을 평가하기 위하여 정확도 평가에서 오차율과 절대값을 모두 알아보기 위하여 일반적으로 많이 쓰이는 평균절대오차(MAE; Mean Absolute Error)와 평균절대상대오차(MARE; Mean Absolute Relative Error)를 평가지표로 정하고, 실제 수집한 버스의 도착시간(같은 노선에 대한 1주간의 데이터 비교분석)과 예측한 버스도착시간을 비교하였다.(변상철, 2006)

MAE는 추정치와 실측치 간의 절대적인 오차를 의미하며, 다음과 같이 실제 수집한 버스도착시간과 예측한 버스도착시간의 차를 이용하여 산출할 수 있다.

$$MAE = \frac{\sum_{i=1}^N |T_i - \hat{T}_i|}{N} \quad (8)$$

$T$  : 수집한 버스의 도착시간  
 $\hat{T}$  : 예측한 버스의 도착시간  
 $N$  : 데이터 수

여기서, MAE 값은 작은 값을 가질수록 절대적인 오차가 작다는 것을 의미하며, 결국 MAE 값이 작을수록 정확도가 높다는 것을 의미한다.

그리고 MARE는 추정치와 실측치 간의 상대적 오차를 나타내는 것으로, MAE 값과 마찬가지로 값이 작을수록 상대적 오차가 작다는 것을 의미한다. MARE 값은 식(9)를 이용하여 산출할 수 있다.

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|T_i - \hat{T}_i|}{T_i} \quad (9)$$

$T$  : 수집된 버스도착시간  
 $\hat{T}$  : 예측된 버스도착시간  
 $N$  : 데이터 수

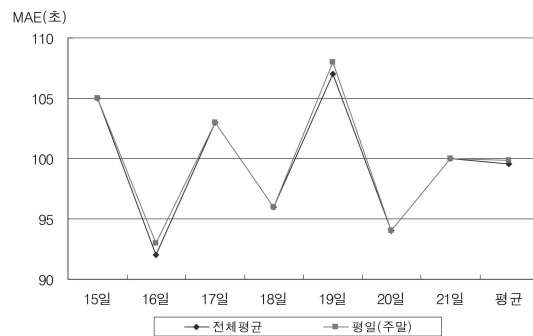
3) 전체평균 및 평일(주말) 구분

본 연구에서는 기본적으로 평일과 주말 구분 없이 평균 전이확률행렬표를 생성하고 버스도착시간을 예측하였다. 전체평균 및 평일(주말)구분에서는 평일과 주말을 구분하여 전이확률행렬표를 생성하고, 이를 이용하여 버스도착시간을 예측하였다. 그리고 MAE 값을 도출하여 전체평균 전이확률행렬표를 이용한 경우와 MAE 값을 비교하였다.

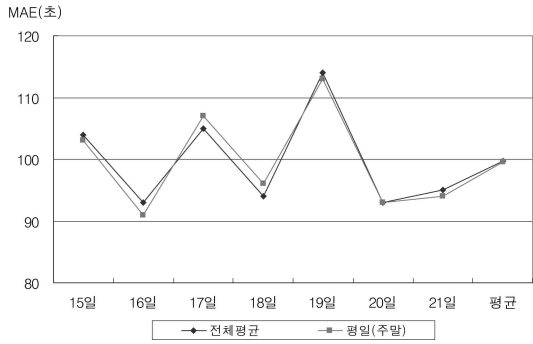
각 행렬별로 MAE 값을 비교한 결과, 세부 요일에서 차이가 조금 있었지만 전반적으로 유사한 값을 얻을 수 있었다. 결과적으로 본 연구의 조건과 같은 상황에서는 평일 및 주말을 구분하여 전이확률행렬표를 생성할 필요 없이 전체평균 전이확률행렬표를 사용하여 버스도착시간을 예측하여도 정확도에 미치는 영향은 작을 것으로 판단된다.(일별 1~2 초의 MAE값의 차이가 있었음)

<표 11> 전체평균 및 평일(주말)구분 MAE

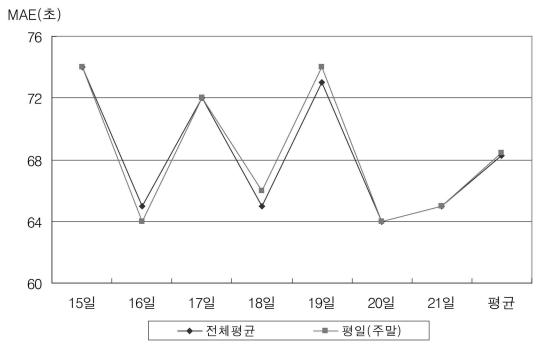
구분	3×3 행렬		5×5 행렬		7×7 행렬		9×9 행렬	
	전체 평균	평일/주말 구분	전체 평균	평일/주말 구분	전체 평균	평일/주말 구분	전체 평균	평일/주말 구분
15일	105	105	104	103	74	74	69	68
16일	92	93	93	91	65	64	56	56
17일	103	103	105	107	72	72	68	68
18일	96	96	94	96	65	66	60	61
19일	107	108	114	113	73	74	70	71
20일	94	94	93	93	64	64	59	59
21일	100	100	95	94	65	65	57	57
평균	100	100	100	100	68	68	63	63



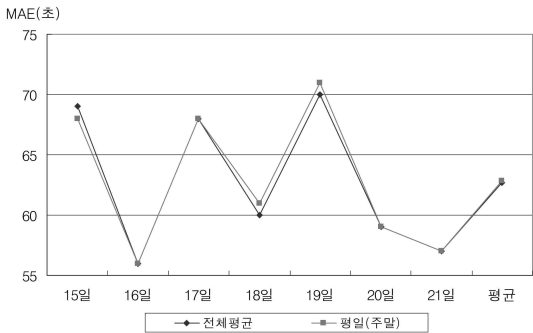
<그림 2> 3×3 행렬 전체평균 및 평일(주말)



<그림 3> 5×5 행렬 전체평균 및 평일(주말)



<그림 4> 7×7 행렬 전체평균 및 평일(주말)

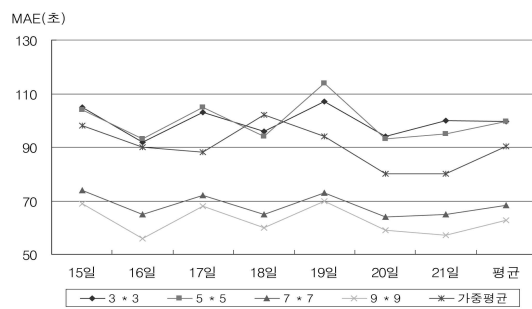


<그림 5> 9×9 행렬 전체평균 및 평일(주말)

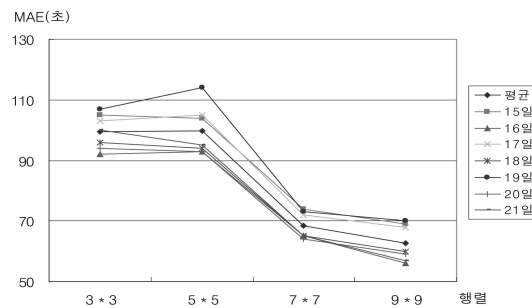
4) 전체평가

전체평가에서는 행렬별/정류장별 평균 전이확률표를 이용하여 예측한 버스도착시간과 실제 수집된 버스도착 시간에 대하여 MAE와 MARE 값을 비교하였다. 전체 평가에 대한 MAE와 MARE 값은 <표 12>와 같다.

분석결과, 전반적으로 계급의 수가 많은 행렬일수록 낮은 MAE와 MARE 값을 나타냈으며, 7×7 행렬과 9×9행렬을 이용한 방법은 기존에 적용되고 있는 가중이동평균법보다 낮은 MAE와 MARE 값을 얻었다. 이는 7×7 행렬과 9×9행렬을 이용한 방법이 기존 방법에 비해 버스도착시간을 예측하는 데 있어 정확도가 높다는 것을 의미한다.



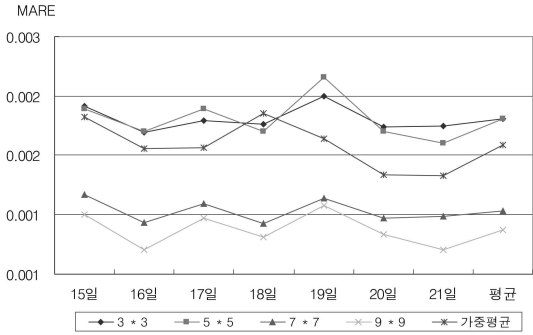
<그림 6> 전체 평가(MAE)



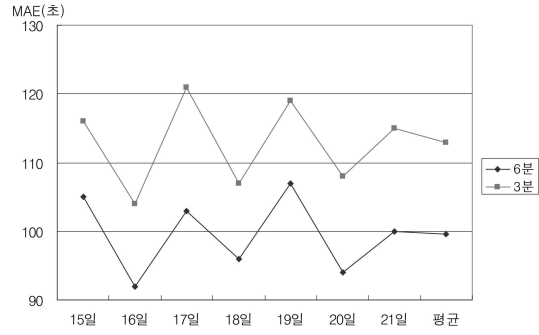
<그림 7> 전체 평가(행렬별 MAE)

<표 12> 전체평가 MAE, MARE

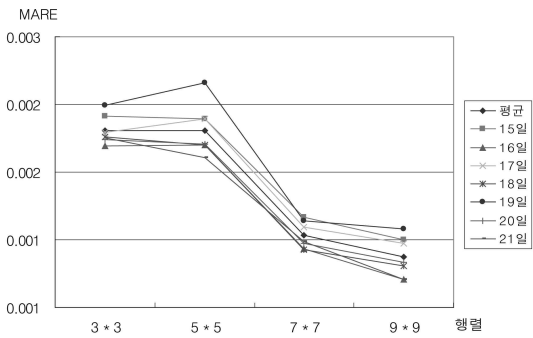
구분	3×3 행렬		5×5 행렬		7×7 행렬		9×9 행렬		가중이동평균	
	MAE	MARE	MAE	MARE	MAE	MARE	MAE	MARE	MAE	MARE
15일	105	0.002132	104	0.002115	74	0.001534	69	0.001401	98	0.002056
16일	92	0.001953	93	0.001961	65	0.001346	56	0.001165	90	0.001847
17일	103	0.002034	105	0.002116	72	0.001476	68	0.001378	88	0.001853
18일	96	0.002008	94	0.001959	65	0.001340	60	0.001248	102	0.002080
19일	107	0.002197	114	0.002329	73	0.001513	70	0.001462	94	0.001914
20일	94	0.001990	93	0.001964	64	0.001377	59	0.001265	80	0.001671
21일	100	0.002000	95	0.001884	65	0.001389	57	0.001163	80	0.001662
평균	100	0.002045	100	0.002047	68	0.001425	63	0.001297	90	0.001869



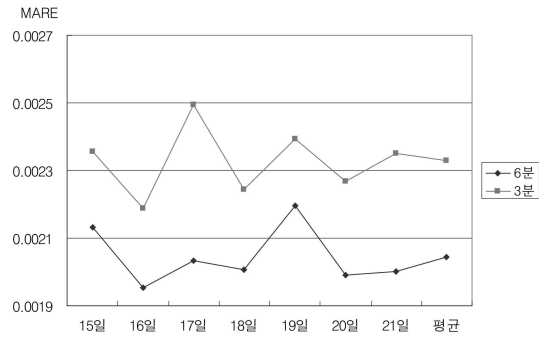
<그림 8> 전체 평가(MARE)



<그림 10> 지체구간 간격별 MAE



<그림 9> 전체 평가(행렬별 MARE)



<그림 11> 지체구간 간격별 MARE

5) 버스 지체구간 간격별 평가

앞의 전체평가에서 계급수가 많아질수록 정확도가 높아지는 것을 확인할 수 있었으며, 계급의 수가 많아진다는 것은 버스지체시간 계급간격이 좁게 설정되어 있다는 것을 의미한다. 결국 버스지체시간 계급간격이 정확도에 영향을 미친다고 가정할 수 있으며, 이에 대한 검증을 위해 버스 지체구간 간격별 평가에서는 3x3 행렬에 대하여 다른 조건은 동일하게 하고 버스지체시간 계급간격을 다르게 하여(6분->3분) MAE와 MARE 값을 비교해 보았다. <표 13>은 산출한 3x3 행렬의 버스지체시간 계

<표 13> 3x3 행렬 버스지체시간 계급간격별 평가

구분	3분		6분	
	MAE	MARE	MAE	MARE
15일	116	0.002355	105	0.002132
16일	104	0.002189	92	0.001953
17일	121	0.002496	103	0.002034
18일	107	0.002245	96	0.002008
19일	119	0.002393	107	0.002197
20일	108	0.002269	94	0.001990
21일	115	0.002352	100	0.002000
평균	113	0.002328	100	0.002045

급간격별 MAE, MARE 값을 나타낸다.

다른 조건은 모두 동일하게 하고 버스지체구간 간격을 6분에서 3분으로 설정하여 MAE와 MARE 값을 비교한 결과, 오히려 3분으로 범위를 좁게 설정한 방법의 경우가 6분을 적용한 방법보다 높은 MAE와 MARE 값을 나타냈다. 이는 버스도착시간예측에 있어 버스지체시간 계급간격만이 정확도 향상에 영향을 미치는 것이 아니라는 것을 의미한다.

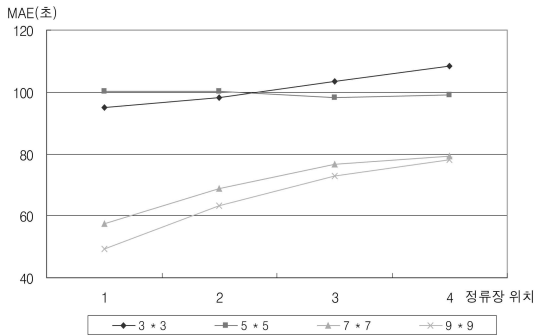
6) 정류장 위치별 평가

정류장 위치별 평가에서는 예측한 지점에서 정류장 위치가 멀어질수록 예측의 정확도 차이가 있는지 확인하기 위하여 4가지 행렬(3x3, 5x5, 7x7, 9x9 행렬)에 대한 정류장 위치별 MAE와 MARE 값을 비교해 보았다. 비교 결과, 5x5 행렬을 제외하고는 정류장 위치가 예측한 지점에서 멀어질수록 높은 MAE와 MARE 값을 나타냈다. 이는 본 연구에 적용된 방법이 전반적으로 예측한 지점에서 정류장 위치가 멀어질수록 버스도착시간 예측의 정확도가 떨어진다는 것을 의미한다. <표 14>는 산출한 정류장 위치별 MAE, MARE 값을 나타낸다.

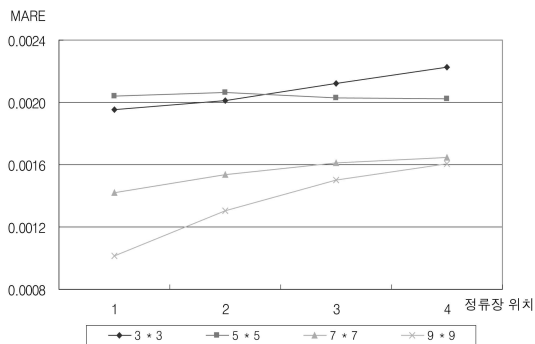


<표 14> 정류장 위치별 MAE, MARE

정류장 위치	3×3 행렬		5×5 행렬		7×7 행렬		9×9 행렬	
	MAE	MARE	MAE	MARE	MAE	MARE	MAE	MARE
1	95	0.001955	100	0.002043	57	0.001420	49	0.001014
2	98	0.002014	100	0.002061	69	0.001538	63	0.001306
3	103	0.002124	98	0.002029	77	0.001610	73	0.001503
4	108	0.002227	99	0.002022	79	0.001645	78	0.001605



<그림 12> 정류장 위치별 MAE



<그림 13> 정류장 위치별 MARE

렬을 이용하여 버스도착시간을 예측한 결과, 분석대상도 시에서 기존에 사용 중인 가중이동평균법을 이용한 버스도착시간예측방법 보다 정확도가 높았다. 이는 가중이동평균법을 이용한 방법의 경우 데이터를 가공하기 위해 10분, 15분 등의 Time Lag을 발생시키는 반면, 본 연구에서 사용한 방법은 가공시간 동안 발생할 수 있는 상황들을 모두 포함하여 확률적으로 접근하였기 때문으로 얻은 결과라 판단된다. 그리고 7×7 행렬과 9×9 행렬에서 기존방법 보다 좋은 결과를 얻었으므로 11×11 행렬 등 행렬수를 계속해서 늘이는 방법과 계급 간격(1분 등)을 줄이는 분석은 분석 과정만 복잡해질 뿐 크게 의미가 없을 것으로 판단된다.

결과적으로 본 연구의 방법(7×7, 9×9 행렬 이용)을 이용하여 버스도착시간을 예측한다면 기존 방법(가중이동평균법)보다 높은 정확도로 버스도착시간을 버스이용자에게 제공할 수 있을 것이다. 결국 신뢰성 있는 버스도착시간 정보제공을 통하여 대중교통에 대한 신뢰성을 높이고, 더 나아가 대중교통 활성화에 기여 할 수 있을 것으로 예상된다.

참고문헌

1. 고승곤·양완연(1998), “일반통계학”, 교우사.
2. 김영찬 외 (2005), “안양 BIS 통행시간 예측 알고리즘 개선 및 개발 최종보고서”, 서울시립대학교.
3. 변상철 (2006), “통행특성을 반영한 일반국도 단기 통행시간 예측기법 연구”, 교통공학 박사학위논문, 서울시립대학교.
4. 이승훈 (2009), “Markov Chain을 이용한 버스도착시간 예측에 관한 연구”, ITS공학 석사학위논문, 과학기술연합대학원대학교.
5. 이승훈·문병섭·박범진 (2009), “Markov Chain을 이용한 버스지체시간 예측”, 한국ITS학회, 한국ITS학회 논문지, 제8권, 제3호, pp.1~10.

IV. 결론

버스도착시간을 예측하기 위하여 기존 연구에서는 가중이동평균법, 신경망 모형, Kalman Filter 기법, 회귀 모형 등의 방법을 주로 사용하였다(김영찬 외, 2005; 이희중 외, 2005; Chien 외, 2002; Shalaby, 2003). 하지만 이들 방법은 많은 변수를 고려해야 하거나 결과에 대하여 설명하기 어려운 한계점을 가지고 있다. 이에 본 연구에서는 많은 변수를 고려하지 않으면서 버스도착시간을 예측할 수 있는, 확률적 접근방법인 마코브 체인 과정을 이용하여 버스도착시간을 예측하였다.

본 연구의 조건과 같은 경우에서 7×7 행렬과 9×9 행

6. 이희중 · 류승기 (2005), “광역버스 통행시간예측을 위한 신경망 모형의 적용”, 대한토목학회, 대한토목학회 정기학술대회.
7. Abdelfattah, A. M., and Khan, A. M. (1998), “Models for Predicting Bus Delays”, Transportation Research Record 1623, Transportation Research Board, pp.8~15.
8. Chien, I. J., Ding, Y., and Wei, C. (2002), “Dynamic Bus Arrival Time Prediction with Artificial Neural Networks”, Journal of Transportation Engineering, Vol. 128, Issue 5, pp.429~438.
9. Lin, W. H. and Zeng, J.(1999), “Experimental Study of Real-Time Bus Arrival Time Prediction with GPS Data”, Transportation Research Record, Journal of the Transportation Research Board, No. 1666, pp.101~109.
10. Lin, W. H., and Bertini, R. L.(2004), “Modeling Schedule Recovery Processes in Transit Operations for Bus Arrival Time Prediction”, Institute for Transportation, Journal of Advanced Transportation, Vol. 38, No. 3, pp.347~365.
11. Rajat, R.(2005), “Bus Arrival Time Prediction Using Stochastic Time Series and Markov Chains”, Dissertation for the Degree of Doctor of Philosophy in Transportation, New Jersey Institute of Technology.
12. Shalaby, A., and Farhan, A. (2003), “Bus Travel Time Prediction Model for Dynamic Operations Control and Passenger Information System”, Transportation Research Board, 82nd Annual Meeting, Washington D.C.
13. Sheldon, M. Ross(2000), “Introduction to Probability Model”, A Harcourt Science and Technology Company.

✉ 주 작성자 : 이승훈

✉ 교신저자 : 이승훈

✉ 논문투고일 : 2009. 9. 7

✉ 논문심사일 : 2009. 12. 3 (1차)

2010. 1. 20 (2차)

2010. 1. 25 (3차)

✉ 심사판정일 : 2010. 1. 25

✉ 반론접수기한 : 2010. 6. 30

✉ 3인 익명 심사필

✉ 1인 abstract 교정필