한국ITS학회논문지
제8권, 제3호 (2009년 6월)
pp. 1~10

Markov Chain을 이용한 버스지체시간 예측

The Bus Delay Time Prediction Using Markov Chain

이 승 훈* (Seung-Hun Lee)　문 병 섭** (Byeong-Sup Moon)　박 범 진*** (Bum-Jin Park)

요 약

버스지체시간은 버스노선의 교통여건이 반영되어 나타나는 결과로서 버스도착시간을 예측하는데 있어 중요한 요소이다. 이에 본 연구에서는 다양한 변수를 사용하지 않아도 되는 마코브 체인을 이용하여 분석 정류장간 전이확률행렬표를 생성하고 이를 이용하여 버스지체시간을 예측하였다. 본 연구를 통하여 기존연구의 한계점인 정류장별 계획된 버스도착시간이 존재하지 않은 경우에 대하여 배차시간을 이용한 버스지체시간 산출방법을 제시함으로서 기존연구의 한계점을 극복하였으며, 또한 정류장별 버스지체시간을 예측하기 위해 정의한 정류장간 버스지체의 전이는 동질하다는 귀무가설을 대응표본 T검정을 통하여 채택함으로서 사용한 가정이 95% 신뢰수준에서 유의하다는 것을 확인하였다. 이를 통하여 향후 마코브 체인을 이용하여 버스도착시간 예측이 가능할 것으로 판단된다.

Abstract

Bus delay time is occurred as the result of traffic condition and important factor to predict bus arrival time. In this paper, transition probability matrixes between bus stops are made by using Markov Chain and it is predicted bus delay time with them. As the results of study, it is confirmed a possibility of adapting the assumption which it has same bus transition probability between stops through paired-samples T-test and overcame the limitation of exiting studies in case there is no scheduled bus arrival time for each stops with using bus interval time. Therefore it will be possible to predict bus arrival time with Markov Chain.

Key words: Markov chain, BIS(Bus Information System), transition probability, transition probability matrix, scheduled bus arrival time

Ⅰ. 서 론

1. 연구의 배경 및 목적

버스정보시스템(BIS; Bus Information System)은 지능형 교통시스템(ITS; Intelligent Transportation System)의 한 분야인 첨단 대중교통시스템(APTS; Advanced Public Transportation System)에서 버스를 이용하는 이용자에게는 버스정보를 제공하여 편의성을 높이고, 버스를 운영하는 관리자에게는 버스 운영의 효율성을 높이는 시스템이다 [1].

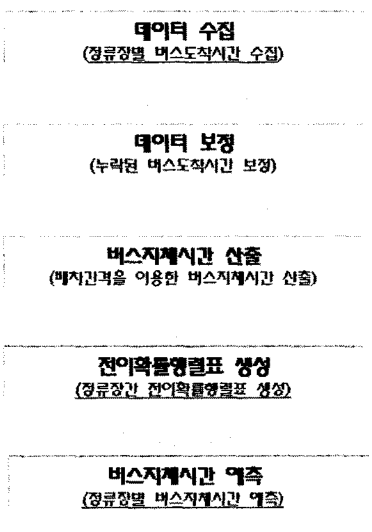
BIS에서 핵심 제공 서비스 중 하나는 버스의 도착시간을 예측하여 버스 이용자에게 그 정보를 제공하

* 주저자 : 한국건설기술연구원 연구원
** 공저자 : 한국건설기술연구원 선임연구원
*** 공저자 : 한국건설기술연구원 연구원
† 논문접수일 : 2009년 3월 10일
† 논문심사일 : 2009년 4월 2일(1차), 2009년 5월 12일(2차)
† 게재확정일 : 2009년 5월 13일

는 것으로 기존에 버스를 이용하는 이용자들이 느꼈던 불편함(버스의 결행, 노선 이탈 등)을 해결하기 위해 버스의 도착정보를 제공함으로서 버스 이용자의 편의를 도모하는 것이다. 결국, 이를 통하여 대중 교통 서비스의 질을 향상시키고, 나아가 승용차의 수요를 대중교통으로 전환시키는 데 궁극적인 목적이 있다.

버스도착시간을 예측하기 위하여 사용하는 대표적인 방법에는 가중이동평균법, 신경망 모형, Kalman Filter 기법, 회귀 모형 등이 있다. 이러한 기존의 예측기법들은 교통량, 링크 통행시간, 도로 기하구조, 평균통행속도, 신호 주기 등 버스의 주행에 영향을 줄 수 있는 다양한 변수를 버스도착시간을 예측하기 위하여 고려하였지만 버스지체시간에 대한 고려는 하지 않았다. 하지만 버스지체시간은 버스노선의 교통여건이 반영되어 나타나는 결과로 버스도착시간을 예측하는데 있어 중요한 요소이다.

본 연구는 기존 연구에서 고려하지 않은 버스지체시간을 예측하는데 목적이 있으며 이를 위해 많은 변수를 고려하지 않아도 되는 마코브 체인을 이용하였다. <그림 1>은 본 연구에서 수행된 연구의 흐름도를 나타낸다.



<그림 1> 연구의 흐름도
<Fig. 1> Study flow

2. 연구의 범위

본 연구는 안양시 흥안로를 통과하는 540번 노선에 대하여 2008년 10월 1일~2008년 10월 14일(2주) 07시~23시 버스도착시간 데이터를 이용하여 전이확률행렬표를 생성하고 이를 활용하여 버스 지체시간을 예측하였다.

또한 찬우물 정류장에서 오뚜기식품 정류장으로 운행(일방향)하는 540번 노선의 5개 정류장(찬우물, 갈현동부대, 인덕원사거리, KT동안양점, 오뚜기식품)만을 대상으로 배차간격 8분[1)]을 적용하였으며, 정류장별 버스지체시간을 예측하기 위하여 정류장간 버스지체시간의 전이는 동질하다(Homogeneous)는 가정을 하였다.

이 구간은 편도 5차선 도로로 시간대(07시~10시, 17시~21시)별로 버스전용차선을 시행하고 있다.

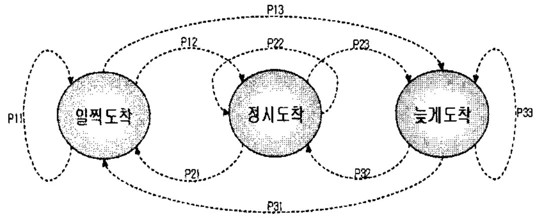
Ⅱ. Markov Chain 개념

마코브 과정(Markovian Process)은 반복되는 상황에 대해 특정 상태에서 다음의 상태로 어떻게 변화할 것인가를 확률적으로 예측하는 기법으로서 어떤 사건이나 현상이 바로 이전 사건이나 현상에 의해서만 결정되는 확률적 과정(Stochastic Process)에 관한 이론을 최초로 전개한 러시아의 수학자 Andrei Andreyevich Markov(1856년~1922년)의 이름을 따라 붙여졌다 [2].

마코브 과정을 전이확률과 함께 마코브 체인이라고 하며 이를 수학적으로 정리하면 다음과 같다. 즉, t번째 기간에 i라는 상태에 있는 것을 X_t 이라고 할 때, 다음의 관계를 만족하는 경우 마코브 체인이라 한다 [3].

$$P\{X_{t+1}=j|X_t=i, X_{t-1}, \ldots, X_0=i_0\}$$
$$=P\{X_{t+1}=j|X_t=i\}$$
$$=P_{ij}$$

1) 안양시 홈페이지에서 540번 버스에 대하여 제공하는 배차간격 적용



<그림 2> 지체-전이 다이어그램
<Fig. 2> Delay-propagation diagram

1. 전이확률행렬표

전이확률(Transition Probability)은 마코브 체인을 이해하기 위한 중요한 요소로서 어떤 반복되는 사건이나 현상이 현재 상태에서 다른 상태로 바뀌는 확률을 말하며, 이 확률을 행렬의 형태로 나타낸 것을 전이확률행렬표(Transition Probability Matrix)라고 한다.

어떤 정류장에서 발생될 수 있는 버스지체시간을 계획된 도착시간에 대하여 <그림 2>와 같이 "일찍도착", "정시도착", "늦게도착"으로 표현이 가능하고 n 정류장에서 상태 i에 있다가 $n+1$정류장에서 j상태에 있게 될 전이확률을 P_{ij}로 정의한다면, 이 정류장에 대한 전이확률행렬표는 다음과 같이 3×3행렬로 나타낼 수 있다.

$$P=[P_{ij}]=\begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & p_{ij} & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix}=\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

$P=[P_{ij}]$: 전이확률행렬표

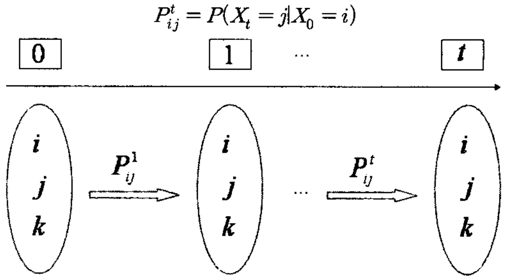
$p_{11},\cdots,p_{mm}$: 상태 i에서 j상태로 바뀌어지는 전이확률($0 \le p_{ij} \le 1$)

m : 상태의 수

예를 들어, 위의 3×3행렬에서 p_{13}은 현재 정류장에서 일찍 도착한 버스가 다음 정류장에 늦게 도착할 확률을 나타내며, p_{33}은 현재 정류장에서 늦게 도착한 버스가 다음정류장에서 다시 늦게 도착할 확률을 나타낸다.

2. Chapman-Kolmogorov 방정식

마코브 체인$\{X_t,\ t \ge 0\}$에서 x_0의 분포를 초기분포



<그림 3> Chapman-kolmogorov 방정식
<Fig. 3> Chapman-kolmogorov equation

라 하고 $p_e=p(X_0=e)$로 할 때, x_0가 나타내는 상태 e를 초기상태라 부른다. 또 상태 i에서 출발한 마코브 체인이 t번째에서 상태 j에 도달할 확률을 i에서 j로의 t-단계전이확률이라고 부르며 P^t_{ij}로 표기한다.

임의의 정수 l과 $m(l,\ m \ge 0)$에 대하여 $P^{(l+m)}=P^{(l)} \cdot P^{(m)}$로 나타내면, 결국 $P^{(l+m)}_{ij}=\sum_{k=0}^{\infty} P^l_{ik} \cdot P^m_{kj}$ $(l \ge 0,\ m \ge 0)$이 성립한다 [3]. 이것이 Chapman-Kolmogorov 방정식이며, 이를 이용하여 t번째 전이확률을 구할 수 있다. $P^{(t)}$을 t-단계 전이확률 P^t_{ij}의 행렬이라고 하면 다음과 같은 식이 성립한다.

$$P^{(l+m)}=P^{(l)} \cdot P^{(m)}$$

예를 들어, 2-단계에 대한 전이확률행렬은 다음과 같다.

$$P^{(2)}=P^{(1+1)}=P^{(1)} \cdot P^{(1)}=P^2$$

따라서 t-단계 전이확률행렬은 1단계 전이확률행렬 P를 t-제곱 한 것과 같다.

Ⅲ. 버스지체시간 예측

1. 버스지체시간 산출

버스지체시간은 계획된 버스도착시간(Scheduled Bus Arrival Time)에 대한 실제 버스도착시간(Actual Bus Arrival Time)의 차를 말한다 [4-6].

하지만 이러한 기존연구의 버스지체시간 산출 방

<표 1> n정류장에서 버스지체시간 산출 방법의 예
<Table 1> Calculation of actual bus delay time

구분	n정류장 버스도착시간	$BAT_{n,i} - BAT_{n,i-1}$	버스 배차간격	버스 지체시간
버스1	17:02:41	-	8분	-
버스2	17:13:20	00:10:39		+00:02:39
버스3	17:19:53	00:06:33		−00:01:27
버스4	17:37:14	00:17:21		+00:09:21

법은 정류장별 계획된 버스도착시간을 필요로 하기 때문에, 정류장별 계획된 버스도착시간이 없는 본 연구의 분석대상 정류장의 경우 버스지체시간을 구하는데 한계가 있다.

따라서 본 연구에서는 계획된 버스도착시간이 없을 경우에 대하여 버스지체시간을 배차간격을 이용하여 산출 하였다. 즉, 버스지체시간은 n정류장에서 i버스의 도착시간, n정류장에서 $i-1$버스의 도착시간, 그리고 분석대상 버스의 배차간격을 이용하여 산출하였다.

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

$AD_{n,i}$: n정류장에서 i번째 버스의 지체시간
$BAT_{n,i}$: n정류장에서 i번째 버스의 도착시간
$BAT_{n,i-1}$:n정류장에서 $i-1$번째 버스의 도착시간
I : 배차간격

버스의 지체시간은 주어진 배차간격에 대한 앞 뒤 차간의 도착시간차로 산출 할 수 있으며, <표 1>은 본 연구에서 적용한 버스지체시간 산출 방법의 예를 나타낸다.

<표 1>에서 버스2와 버스4의 경우는 "+" 버스지체시간 값을 얻었는데 이는 버스가 배차간격보다 늦게 도착한 경우이고, "−" 버스지체시간 값을 얻은 버스3의 경우는 버스가 예정시간보다 일찍 도착한 경우이다.

2. 데이터 보정

모든 시스템이 100% 데이터 수집률을 목표로 구축되지만 실제로 수집된 버스도착시간 데이터를 살펴보면 누락 데이터가 존재한다. 본 연구에서 적용한 버스지체시간 산출 방법은 앞 뒤차간의 버스도착시간 차를 이용하기 때문에, 연구의 수행을 위해서 반드시 누락된 버스도착시간 데이터에 대한 보정이 필요하다. 따라서 누락된 버스의 도착시간을 산출하기 위하여 본 연구에서는 다음 식을 이용하였다.

$$BAT_{n,i} = BAT_{n+1,i} - (BAT_{n+1,i-1} - BAT_{n,i-1}) \text{ or}$$
$$BAT_{n,i} = BAT_{n-1,i} + (BAT_{n,i-1} - BAT_{n-1,i-1})$$

$BAT_{n,i}$: n정류장에서 i번째 버스의 도착시간
$BAT_{n+1,i}$: $n+1$정류장에서 i번째 버스의 도착시간
$BAT_{n+1,i-1}$: $n+1$정류장에서 $i-1$번째 버스의 도착시간
$BAT_{n,i-1}$: n정류장에서 $i-1$번째 버스의 도착시간
$BAT_{n-1,i}$: $n-1$정류장에서 i번째 버스의 도착시간
$BAT_{n-1,i-1}$: $n-1$정류장에서 $i-1$번째 버스의 도착시간

만약, 누락데이터를 가진 버스에 대하여 다음 정류장 또는 전 정류장의 버스도착시간이 존재한다면, 누락된 버스도착시간은 앞차의 정류장간 통행시간과 현재 차량의 다음 정류장 또는 전 정류장의 도착시간을 이용하여 산출할 수 있다.

다음은 본 연구에서 적용된 방법을 이용하여 누락된 버스도착시간 데이터를 보정한 예를 나타낸다.

<표 2>와 같이 A, B의 버스도착시간이 누락되었다고 가정하면, 다음과 같이 버스도착시간을 산출할 수 있다.

A= 17:13:20-(17:02:41-17:02:10)= 17:12:49
B= 17:39:44+(17:26:17-17:24:05)= 17:41:56

<표 2> 누락된 버스도착시간의 예
<Table 2> Example of collected bus arrival time

구분	찬우물	갈현동 부대	인덕원 사거리	KT 동안양점	오뚜기 식품
버스1	17:02:10	17:02:41	17:05:45	17:07:41	17:09:23
버스2	A	17:13:20	17:15:07	17:16:47	17:18:38
버스3	17:18:14	17:19:53	17:22:09	17:24:05	17:26:17
버스4	17:35:59	17:37:14	17:38:17	17:39:44	B

<표 3> 누락된 버스도착시간 데이터 보정의 예
<Table 3> Example of data calibration

구분	찬우물	갈현동 부대	인덕원 사거리	KT 동안양점	오뚜기 식품
버스1	17:02:10	17:02:41	17:05:45	17:07:41	17:09:23
버스2	17:12:49	17:13:20	17:15:07	17:16:47	17:18:38
버스3	17:18:14	17:19:53	17:22:09	17:24:05	17:26:17
버스4	17:35:59	17:37:14	17:38:17	17:39:44	17:41:56

3. 전이확률행렬표 생성

마코브 체인을 이용하여 정류장별 버스지체시간을 예측하기 위해서는 먼저 전이확률행렬표를 생성이 필요하며, 이 전이확률행렬표를 생성하기 위해서는 다양한 값을 가지는 버스지체시간에 대한 분류가 필요하다. 본 연구에서는 버스지체시간을 분류하기 위하여 먼저 전체 버스지체시간에 대한 지체시간 범위를 최대값과 최소값의 차로 산출하였다.

$$DR = AD_{\max} - AD_{\min}$$

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

2008년 10월 1일~2008년 10월 14일(2주)에 대한 일별 버스지체시간 범위를 구하고 이를 평균하여 2주 전체에 대한 평균 버스지체시간 범위를 산출하였다. 산출한 결과, 평균 버스지체시간 범위는 약 19분 27초였으며, 이를 이용하여 버스지체시간의 계급간격을 산출하였다. 즉, 지체시간의 계급간격은 분석 데이터의 범위를 계급의 수로 나누어서 산출하였다 [7].

$$DI = \frac{DR}{k}$$

DI : 지체 간격
DR: 지체 범위
k : 계급의 수

앞에서 산출한 버스지체시간 계급간격을 이용하여 버스지체구간을 설정하면 <표 5>와 같으며, 이를 이용하여 본 연구에서는 산출한 버스지체시간을 분류하였다.

<표 4> 적용한 지체시간의 계급간격
<Table 4> Applied interval of delay time

구분	지체 범위	계급의 수	계산된 계급간격	적용한 지체시간의 계급간격
3 × 3 행렬	19분 27초	3	6분 29초	6분

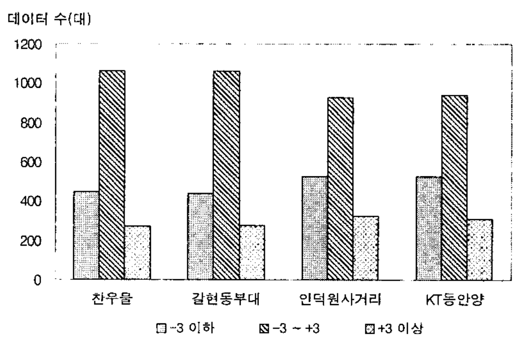
<표 5> 적용한 버스지체구간
<Table 5> Classification of bus delay time

구분	적용한 버스지체구간
3 × 3 행렬	'−3분 이하', '−3분~+3분', '+3분 이상'

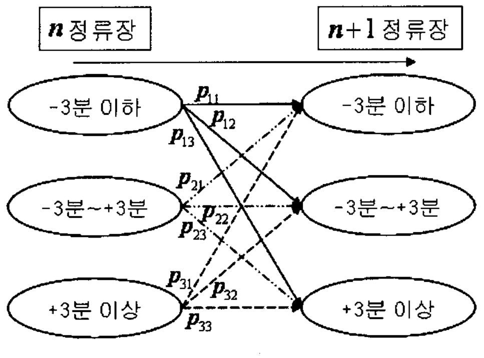
전이확률행렬표 생성에 앞서 그룹별로 분류한 실제 버스지체시간에 대하여 그 분포를 살펴보면 <표 6>과 같다.

<표 6> 정류장별 버스지체 분포
<Table 6> Classification of actual bus delay time

구분	찬우물	갈현동 부대	인덕원 사거리	KT 동안양
−3분 이하	446	437	525	525
−3분~+3분	1,059	1,062	924	940
+3분 이상	271	277	327	311
합계	1,776	1,776	1,776	1,776



<그림 4> 정류장별 실제 버스지체시간 분포
<Fig. 4> Distribution of actual bus delay time at stops



<그림 5> 정류장간 버스지체시간의 전이
<Fig. 5> Propagation of bus delay time

<표 7> 찬우물→갈현동부대
<Table 7> Applied transition probabilities 1

Delay(분)	−3분 이하	−3분~+3분	+3분 이상
3분 이하	0.88080	0.11920	0.00000
−3분~+3분	0.03979	0.94579	0.01443
+3분 이상	0.00000	0.03579	0.96421

<표 8> 갈현동부대→인덕원사거리
<Table 8> Applied transition probabilities 2

Delay(분)	−3분 이하	−3분~+3분	+3분 이상
−3분 이하	0.90405	0.09595	0.00000
−3분~+3분	0.12186	0.81926	0.05887
+3분 이상	0.00000	0.04191	0.95809

<표 9> 인덕원사거리→KT동안양점
<Table 9> Applied transition probabilities 3

Delay(분)	−3분 이하	−3분~+3분	+3분 이상
−3분 이하	0.90486	0.09514	0.00000
−3분~+3분	0.05086	0.94914	0.00000
+3분 이상	0.00000	0.04518	0.95482

<표 10> KT동안양점→오뚜기식품
<Table 10> Applied transition probabilities 4

Delay(분)	−3분 이하	−3분~+3분	+3분 이상
−3분 이하	0.82367	0.17633	0.00000
−3분~+3분	0.04022	0.93416	0.02562
+3분 이상	0.00000	0.08207	0.91793

버스지체시간의 분포는 전반적으로 0분을 중심으로 왼쪽으로 치우치는 경향을 나타냈으며, 이를 통해 분석대상 버스가 적용한 배차간격(8분)에 비하여 전반적으로 일찍 도착하고 있음을 알 수 있다.

3×3 행렬에 대한 버스지체시간의 경우 3가지 버스지체구간으로 분류하였기 때문에 현재 상태에서 다음 상태로의 버스지체의 전이는 3가지 경우, '−3분 이하', '−3분~+3분', '+3분 이상'으로 전이될 수 있다. <그림 5>는 정류장간 버스지체시간의 전이에 대한 개념도를 나타낸다.

만약, n정류장에 '−3분 이하' 데이터가 10대가 존재할 때, $n+1$정류장에서 '−3분 이하' 데이터가 3대가 존재한다고 가정한다면, p_{11}의 값은 3/10=0.3 으로 정의할 수 있다. 이와 같은 방식을 이용하여 생성한 정류장간 전이확률행렬표는 <표 7~10>과 같다.

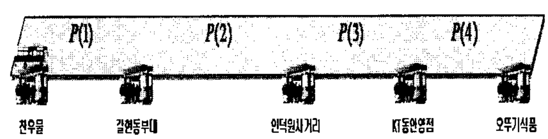
생성한 전이확률행렬표를 살펴보면 '−3분 이하'의 상태에서 '+3분 이상'의 상태로 전이될 확률과, '+3분 이상'의 상태에서 '−3분 이하'의 상태로 전이될 확률값은 모든 정류장 사이에서 '0'의 값을 나타냈다. 이는 본 연구의 분석대상 정류장에서는 '−3분 이하'의 상태에서 '+3분 이상'의 상태로 전이될 확률과, '+3분 이상'의 상태에서 '−3분 이하'의 상태로 전이될 확률이 발생 하지 않는다는 것을 의미한다.

4. 전이확률표를 이용한 버스지체시간 예측

1) 정류장별 버스지체시간 예측

버스정류장간 버스지체시간의 전이는 정류장간 특성(교통량, 도로기하구조, 신호 등)이 다르기 때문에 동질하지 않다(Heterogeneous).

하지만 본 연구에서는 마코브 체인에서 n단계 전이확률을 구하기 위하여 사용하는 Chapman-Kolmogorov



<그림 6> 버스지체 전이가 동질한 경우
<Fig. 6> Bus delay propagation(Homogeneous)

<표 11> 초기 지체상태행렬
<Table 11> Initial delay states matrix

현재 정류장 지체상태	초기 지체상태행렬
−3분 이하	{1 0 0}
−3분~+3분	{0 1 0}
+3분 이상	{0 0 1}

방정식을 하였으며, 이를 위해 버스정류간 버스지체의 전이는 동질하다는 가정을 하였다.

$$P(1)=P(2)=P(3)=P(4)$$

따라서 Chapman-Kolmogorov 방정식을 이용하여 n번째 정류장의 전이확률을 구하면 $P(1)\times P(2)\times\cdots\times P(n-1)=P(1)^{n-1}$과 같으며, 이를 이용하여 n번째 정류장의 버스지체시간은 다음 식을 이용하여 예측할 수 있다.

$$\begin{aligned} ED_n &= P_0\times\{P(1)\times\cdots\times P(n-1)\}\times DS^T \\ &= P_0\times\{P(1)^{n-1}\}\times DS^T \end{aligned}$$

ED_n : n정류장에서 예측한 버스지체시간
P_0 : 초기 지체상태행렬
$P(n)$: n번째 정류장 전이확률행렬
DS^T : 지체상태행렬의 전치행렬

즉, n번째 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, 3×3 행렬의 경우 <표 11>과 같이 표현할 수 있다.

<표 12> 적용된 지체상태 행렬
<Table 12> Applied delay states matrix

구분	지체구간 간격	적용된 지체상태 행렬(분)
3×3 행렬	6분	{-6 0 6}

또한, 지체상태행렬(DS)은 버스지체시간에 대한 예측값을 얻기 위해 필요한 행렬로서 본 연구에서는 버스지체구간의 중간 값을 이용하여 도출하고, 이 행렬의 전치행렬을 버스지체시간을 예측하기 위하여 사용하였다. 연구에 사용한 지체상태행렬은 <표 12>와 같다.

앞에서 설명한 방법을 이용하여 현재 정류장에서 n정류장의 버스지체시간을 예측하였으며, 그 결과는 <표 13>과 같다. ①→④의 의미는 540번 버스가 ①번 정류장(찬우물)에 도착했을 경우 ④번 정류장(KT동안양점)에 버스지체시간을 예측한 경우이다. 예를 들어 ①정류장에 버스가 "−3분 이하"의 지체상태로 도착했을 경우 ④번 정류장에 발생할 버스지체시간은 다음과 같이 구할 수 있다.

$$ED_n = P_0\times\{P(1)^3\}\times DS^T$$

$$=\begin{pmatrix}1\,0\,0\end{pmatrix}\times\begin{pmatrix}0.8808 & 0.1192 & 0\\ 0.0398 & 0.9458 & 0.0144\\ 0 & 0.0358 & 0.9642\end{pmatrix}^3\times\begin{pmatrix}-6\\ 0\\ 6\end{pmatrix}$$

$$=\begin{pmatrix}1\,0\,0\end{pmatrix}\times\begin{pmatrix}0.6968 & 0.2990 & 0.0048\\ 0.0998 & 0.8607 & 0.0396\\ 0.0040 & 0.0981 & 0.8979\end{pmatrix}\times\begin{pmatrix}-6\\ 0\\ 6\end{pmatrix}$$

$$=-4.14824(\text{분})$$

$$=-248.8944(\text{초})\fallingdotseq -249(\text{초})$$

다시 말해 ①정류장에 버스가 "−3분 이하"의 지체

<표 13> 정류장별 예측한 버스지체시간(초)
<Table 13> Estimated bus delay time at stops

구분	①→②	①→③	①→④	①→⑤	②→③	②→④	②→⑤	③→④	③→⑤	④→⑤
−3분 이하	−317	−281	−249	−222	−326	−296	−272	−326	−296	−296
−3분~+3분	−10	−16	−22	−26	−22	−38	−48	−18	−34	−5
+3분 이상	347	334	322	310	346	330	314	344	327	331

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

상태로 도착했을 경우 ④번 정류장에서는 약 −249초 지체가 일어난다는 것을 의미한다.

2) 정류장 간 버스지체 동질성 검증

본 연구에서는 n정류장의 버스지체시간을 예측하기 위하여 마코브 체인의 Chapman-Kolmogorov 방정식을 적용하였으며, 이를 위해 각 정류장 간 버스지체의 전이는 동질하다(Homogeneous)는 가정을 하였다. 실제 본 연구대상의 정류장간 버스지체의 전이가 동질성을 가지는지 검증하기 위해서 2008년 10월 1일~2008년 10월 14일(2주)에서 산출된 정류장간 버스지체시간을 각 시간대(07시~23시)별로 구분하고, 평균 버스지체시간 값을 이용하여 95% 신뢰수준에서 정류장간 대응표본 T검정을 수행하였다.

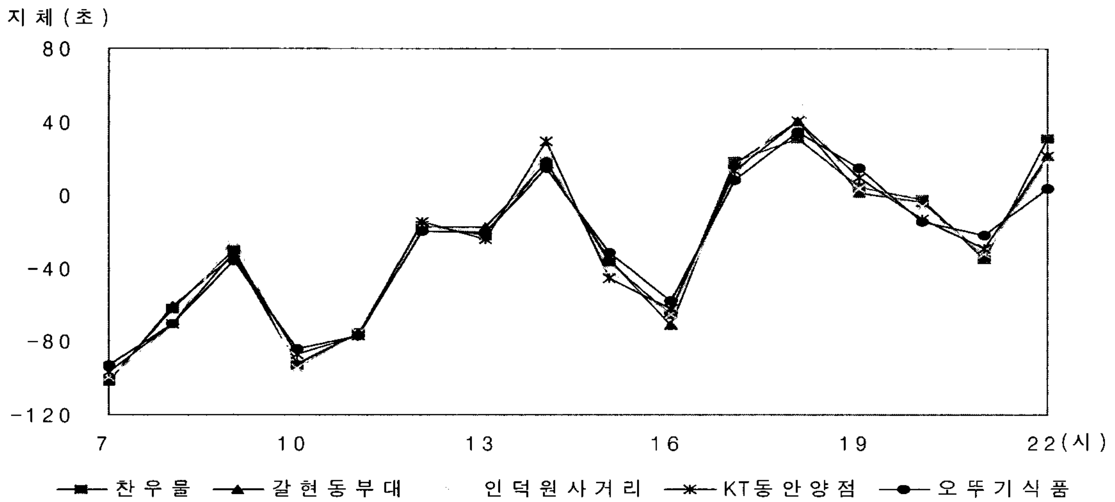
<표 14>와 같이 시간대별로 정류장별 평균 버스지체시간을 산출한 결과, 14시~15시, 17시~19시의 시간대에서 "+" 지체(버스가 배차간격보다 늦게 도착한 경우)값을 나타냈으며, 나머지 시간대의 경우 "−" 지체(버스가 배차간격보다 일찍 도착한 경우)값을 나타냈다.

또한, 정류장별로 전체시간에 대한 평균 버스지체시간 값은 −27초~−28초로 최대 1초의 차이를 나타냈으며, "−" 지체값을 나타냈기 때문에 분석대상 버스는 배차간격보다 평균적으로 일찍 도착한다고 판단 할 수 있다.

<그림 8>은 분석대상 정류장의 시간대별 평균 지체시간의 분포를 나타내며, 정류장별로 매우 유사한

<표 14> 시간대별 평균 버스지체시간(초)
<Table 14> Average bus delay time at individual hour(s)

시간	찬우물	갈현동 부대	인덕원 사거리	KT 동안양점	오뚜기 식품
7	−101	−102	−100	−96	−93
8	−62	−61	−71	−70	−70
9	−30	−33	−23	−31	−35
10	−93	−92	−95	−87	−84
11	−75	−75	−75	−77	−77
12	−17	−17	−17	−14	−19
13	−22	−17	−20	−24	−20
14	18	18	22	29	15
15	−36	−35	−41	−45	−31
16	−65	−71	−65	−62	−58
17	18	17	11	13	8
18	31	40	48	41	34
19	5	1	4	10	15
20	−2	−3	−6	−13	−14
21	−35	−33	−32	−29	−22
22	31	21	19	22	3
평균	−27	−28	−28	−27	−28



<그림 8> 시간대별 평균 지체시간 분포
<Fig. 8> Time series of average delay time(s)

<표 15> 정류장간 대응표본 T검정

<Table 15> T-test between stops

구분	대응 차					t	자유도	유의확률 (양쪽)
	평균	표준편차	평균의 표준오차	차이의 95% 신뢰구간				
				하한	상한			
① - ②	0.43750	4.27346	1.06837	-1.83967	2.71467	0.410	15	0.688
① - ③	0.37500	6.76141	1.69035	-3.22790	3.97790	0.222	15	0.827
① - ④	-0.12500	7.03207	1.75802	-3.87213	3.62213	-0.071	15	0.944
① - ⑤	0.81250	10.40653	2.60163	-4.73275	6.35775	0.312	15	0.759
② - ③	-0.06250	5.37238	1.34310	-2.92524	2.80024	-0.047	15	0.963
② - ④	-0.56250	6.92790	1.73198	-4.25412	3.12912	-0.325	15	0.750
② - ⑤	0.37500	9.37283	2.34321	-4.61943	5.36943	0.160	15	0.875
③ - ④	-0.50000	5.16398	1.29099	-3.25169	2.25169	-0.387	15	0.704
③ - ⑤	0.43750	9.17946	2.29487	-4.45389	5.32889	0.191	15	0.851
④ - ⑤	0.93750	8.07852	2.01963	-3.36724	5.24224	0.464	15	0.649

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

버스지체시간 분포를 나타내고 있다는 것을 확인 할 수 있다. 유사한 분포를 가지는 평균 버스지체시간이 통계적으로도 동질한 유의성을 가지는지 확인하기 위하여, 본 연구에서는 95% 신뢰수준에서 시간대별로 정류장별 평균 버스지체시간에 대한 대응표본 T검정을 수행하였다.

다음은 대응표본 T검정을 수행하기 위해 사용한 귀무가설과 대립가설을 나타낸다.

- 귀무가설: 두 정류장간 시간대별 버스지체시간 분포는 차이가 없다.
- 대립가설: 두 정류장간 시간대별 버스지체시간 분포는 차이가 있다.

앞에서 정의한 가설을 바탕으로 시간대별 평균 버스지체시간에 대하여 대응표본 T검정을 수행한 결과는 <표 15>와 같다.

95% 신뢰구간에서 유의확률(양쪽)이 0.688에서 0.963사이에 나타났으며, 이는 유의수준 0.05보다 큰 값을 가지기 때문에 정류장간 버스지체시간 분포는 차이가 없다는 귀무가설을 채택할 수 있다. 따라서 분석대상 정류장간 시간대별 버스지체시간 분포는 차이가 없다고 할 수 있으며, 이를 통하여 시간대별 정류장간 버스지체의 전이는 동질하다는 가정이 적용가능 할 것으로 판단된다.

Ⅳ. 결 론

본 연구에서는 다양한 변수의 고려 없이 확률적으로 얼마정도의 버스지체값을 발생할 것인지를 예측하기 위하여 마코브 체인을 이용하여 정류장간 전이 확률행렬표를 생성하고 이를 이용하여 정류장별 지체시간을 예측하였다. 특히 기존연구의 한계점인 정류장별 계획된 버스도착시간(Scheduled Bus Arrival Time)이 존재하지 않은 경우에 대하여 배차시간을 이용하여 버스지체시간을 산출함으로서 기존연구의 한계점을 극복하였다.

또한 정류장별 버스지체시간을 예측하기 위해 적용한 정류장간 버스지체의 전이는 동질하다는 가정을 검정하기 위해 대응표본 T검정을 수행하였다. 검정의 결과 95% 신뢰구간에서 두 정류장간 시간대별 버스지체시간 분포는 차이가 없다는 귀무가설을 채택할 수 있었으며, 이를 통하여 정류장간 버스지체의 전이는 동질하다는 가정이 적용가능 할 것으로 판단된다. 결국 이러한 결과는 향후 마코브 체인을 이용하여 버스도착시간을 예측(배차간격+버스지체시간) 하는 데 좋은 기초 자료가 될 것으로 판단된다.

하지만 본 연구는 배차간격이 다른 노선의 경우 버

스지체시간을 예측하기 위하여 새로운 전이확률행렬 표를 생성해야 하며, 기존 연구보다 계산 과정을 최소화하기 위하여 전체 지체시간에 대한 범위를 기준으로 분석하였다. 또한 한달 이상이 아닌 2주간의 자료를 이용하고 배차간격별, 주말/평일별 등 세부적인 변수를 고려를 하지 않은 한계점을 가지고 있다. 그리고 최적의 도착시간을 예측하기 위한 전이확률의 최적화 방법은 고려하지 않고 적용가능성에 대한 부분만을 언급하였다. 이러한 한계점들은 최적의 도착시간을 예측하기 위해 향후 연구에 반영되어야 할 것이다.

참고문헌

[1] 이승훈, *Markov Chain을 이용한 버스도착시간 예측에 관한 연구*, ITS공학 석사학위논문, 과학기술연합대학원대학교, 2009.

[2] 김효정, *마코브 체인을 활용한 고객생애가치 측정 및 고객관계관리 연구*, 일반대학원 경영학 석사학위논문, 서울대학교, 2005.

[3] M. R. Sheldon, *Introduction to Probability Model*, A Harcourt Science and Technology Company, 2000.

[4] W. H. Lin and J. Zeng, "Experimental Study of Real-Time Bus Arrival Time Prediction with GPS Data, Transportation Research Record," *J. Transportation Research Board*, no. 1666, pp. 101-109, 1999.

[5] W. H. Lin and R. L. Bertini, "Modeling schedule recovery processes in transit operations for bus arrival time prediction," *J. Advanced Transportation*, vol. 38, no. 3, pp. 347-365, 2004.

[6] R. Rajat, *Bus Arrival Time Prediction Using Stochastic Time Series and Markov Chains*, Ph. D. Dissertation, New Jersey Institute of Technology, 2005.

[7] 고승곤, 양완연, *일반통계학*, 교우사, 1998.

저자소개

이 승 훈 (Lee, Seung-Hun)

2009년 2월 : 과학기술연합대학원대학교 공학석사 (ITS 공학 전공)
2007년 3월 ~ 2009년 2월 : 한국건설기술연구원 첨단교통연구실 학생연구원
2009년 4월 ~ 현재 : 한국건설기술연구원 첨단교통연구실 연구원

문 병 섭 (Moon, Byeong-Sup)

2002년 2월 ~ 현재 : 한국건설기술연구원 첨단교통연구실 선임연구원

박 범 진 (Park, Bum-Jin)

2003년 3월 ~ 현재 : 한국건설기술연구원 첨단교통연구실 연구원