

고속버스 통행시간 예측의 정확도 제고를 위한 입력자료 분석기간 선정 연구

Analysis Period of Input Data for Improving the Prediction Accuracy of Express-Bus Travel Times

남 승 태	Nam, Seung-Tae	한국도로공사 건설처 차장 (E-mail : naset@ex.co.kr)
윤 일 수	Yun, Ilsoo	정회원 · 아주대학교 교통시스템공학과 부교수 · 교신저자 (E-mail : ilsooyun@ajou.ac.kr)
이 철 기	Lee, Choul-Ki	아주대학교 교통시스템공학과 부교수 (E-mail : cklee@ajou.ac.kr)
오 영 태	Oh, Young-Tae	아주대학교 교통시스템공학과 교수 (E-mail : ytoh@ajou.ac.kr)
최 윤 택	Choi, Yun-Taik	한국도로공사 도로교통연구원 원장 (E-mail : cytk_2000@ex.co.kr)
권 건 안	Kwon, Kenan	교통안전공단 운영지원처 차장 (E-mail : ids2nao@ts2020.kr)

ABSTRACT

PURPOSES : The travel times of expressway buses have been estimated using the travel time data between entrance tollgates and exit tollgates, which are produced by the Toll Collections System (TCS). However, the travel time data from TCS has a few critical problems. For example, the travel time data include the travel times of trucks as well as those of buses. Therefore, the travel time estimation of expressway buses using TCS data may be implicitly and explicitly incorrect. The goal of this study is to improve the accuracy of the expressway bus travel time estimation using DSRC-based travel time by identifying the appropriate analysis period of input data.

METHODS : All expressway buses are equipped with the Hi-Pass transponders so that the travel times of only expressway buses can be extracted now using DSRC. Thus, this study analyzed the operational characteristics as well as travel time patterns of the expressway buses operating between Seoul and Dajeon. And then, this study determined the most appropriate analysis period of input data for the expressway bus travel time estimation model in order to improve the accuracy of the model.

RESULTS : As a result of feasibility analysis according to the analysis period, overall MAPE values were found to be similar. However, the MAPE values of the cases using similar volume patterns outperformed other cases.

CONCLUSIONS : The best input period was that of the case which uses the travel time pattern of the days whose total expressway traffic volumes are similar to that of one day before the day during which the travel times of expressway buses must be estimated.

Keywords

expressway, express-bus, travel time prediction, DSRC, analysis period

Corresponding Author : Yun, Ilsoo, Associate Professor
Department of Transportation System Engineering,
Ajou University, 206, World cup-ro, Yeongtong-gu,
Suwon-si, Gyeonggi-do, 443-749, Korea
Tel : +82.31.219.3610 Fax : +82.31.215.7604
E-mail : ilsooyun@ajou.ac.kr

International Journal of Highway Engineering
http://www.ksre.or.kr/
ISSN 1738-7159 (print)
ISSN 2287-3678 (Online)
Received Mar. 2, 2014 Revised Mar. 3, 2014 Accepted Aug. 1, 2014

1. 서론

1.1. 연구의 배경 및 목적

최근 Information Technology(IT) 및 Intelligent Transport Systems(ITS)에 힘입어 고속버스 이용자들이 필요로 하는 각종 교통정보서비스에 대한 요구가 증가되고 있다. 다양한 요구 중에서 고속버스 이용자들에게 가장 필요한 것은 목적지까지의 경로별에 따른 요일별, 시간대별 통행시간에 대한 정확한 정보로 조사되었다. 특히 고속버스의 경로별에 따른 주말과 평일의 통행시간과 시간대별 통행시간에 대해 많은 이용자들이 궁금해 하고 있는 것으로 판단된다. 실제로 도로이용자의 88.9%가 출발 4시간에서 48시간 전에 출발시각을 결정하고 82.2%가 출발시각 결정 시 교통상황 예측정보를 활용한다고 한다(KOTI, 2010).

고속버스 이용객 또는 마중객이 고속버스 통행시간을 알고 싶은 경우는 크게 두 가지로 구별될 수 있다. 첫 번째는 이용객이 여행을 계획하거나 고속버스 표를 예매할 때 자신의 출발지에서 목적지까지 얼마나 걸릴지를 사전에 알고 싶어 하는 경우이다. 두 번째는 이용객이 고속버스를 타고 이동하고 있는 중에 또는 마중객이 마중을 갈 때 이용객이 타고 있는 고속버스가 언제 도착하는지를 알고 싶어 하는 경우이다. 전자의 경우 보통 off-line 통행시간예측에 해당하며 주로 이력자료(historic data)를 이용하여 예측하는 경우가 일반적이다. 후자의 경우는 on-line 또는 실시간 통행시간예측에 해당되며, 실제 통행시간이 기상상태, 교통량, 유고 여부 등에 따라 상당히 가변적이라는 특징이 있다. 따라서 GPS를 이용하여 실시간으로 차량의 위치를 추적하고 또한 VDS 등을 통해 실시간으로 수집한 고속도로 교통상황을 충분히 활용하여야 한다. 본 연구에서는 첫 번째 경우에 해당하는 off-line 통행시간예측의 예측 신뢰도를 높이는 것으로 연구의 범위를 한정하고자 한다.

이에 본 연구에서는 고속도로 버스전용차선이 운행 중인 서울에서 대전구간의 고속버스의 운행자료를 수집하여 요일별, 시간별에 따른 운행특성을 분석함으로써 현행 고속버스 통행시간 예측모형의 오차량 범위와 이러한 오차를 최소화 할 수 있는 고속버스 통행시간 예측모형 입력자료의 분석기간(analysis period)을 선정하고자 한다. 즉, 본 연구는 고속버스의 하이패스 단말기로부터 추출된 Dedicated Short Range Communication(DSRC) 자료를 활용하여 서울~대전구간의 고속버스 통행시간 예측모형에 대한 입력자료의 분석기간을 산정함으로써 예측의 정확도를 높이는데 그 목적이 있다.

고속버스 통행시간 자료는 시계열적 자료로서 계절요인 및 월별요인 등이 내재되어 있어 분석에 어려움이 있다. 따라서 본 연구는 고속버스 통행시간 예측에 사용될 입력자료의 분석기간을 충분히 과학적이고 합리적으로 선정함으로써 고속버스 통행시간 예측의 정확도를 제고하는 데 기여할 것으로 판단된다.

1.2. 연구의 범위 및 방법

2007년 하이패스 전국 확대 이후 고속버스에 하이패스 단말기가 부착되어 운행되고 있다. 따라서 본 연구의 시간적 범위는 이러한 단말기로부터 수집된 2008년 10월부터 2012년 3월까지의 DSRC 자료를 집계·분석하는 것으로 설정하였다. 공간적 범위는 상대적으로 자료 수집이 용이하고 충분한 샘플을 확보할 수 있는 서울과 대전 간을 운행하는 고속버스 경로로 설정하였다.

본 논문의 연구는 선행연구 고찰, 고속버스 운행자료 조사, 고속버스의 운행특성 분석, 고속버스 통행시간 예측모형의 입력자료의 적정 분석기간 평가 순으로 수행하였다.

선행연구 고찰 항목에서는 기존 고속버스 통행시간 예측모형 분석자료 및 분석기법에 대해 기술하였다. 고속버스 운행자료 조사항목에서는 고속버스 운행자료를 1시간 단위 통행시간 자료로 생성하고, 출발영업소와 도착영업소에 대한 코드별 분류작업과 고속버스 통행시간의 중위수 등을 산출하여 날씨 등 추가정보 관련 테이블을 결합하고 통행시간을 보정하였다. 또한 서울과 대전 간 고속버스 통행시간에 대한 이상치를 제거하여 자료의 신뢰성을 향상하였다.

고속버스의 운행특성 분석 항목에서는 서울과 대전 구간의 고속버스 통행시간을 분석하였다. 전체 운행일수를 각 행선별로 산출하고, 요일별, 시간대별 결측률 및 통행시간을 분석하고, 요일과 시간대별의 두 가지 내용을 혼합하여 결측률 및 통행시간을 분석하였다. 또한 경로별 결측치 보정 전과 보정 후의 통행시간을 비교·분석하였다.

마지막으로 통행시간 예측모형의 입력자료 분석기간 산정항목에서는 서울~대전구간의 DSRC 자료를 활용한 고속버스 통행시간 예측모형의 입력자료를 선정하는 방법을 평가하여 통행시간의 정확도를 제고하고자 하였다.

2. 이론적 고찰

2.1. TCS 기반 통행시간 예측

현행 고속버스 통행시간 예측은 Toll Collection System(TCS) 자료를 기반으로 하고 있다. 즉, TCS에 기록되는 입·출구 교통량, 입·출구 통행시간, 수납집계 교통량, 실시간 교통량, 차종간 비율 자료들을 이용하여 교통예보지원시스템의 예측 모델에 의해 예측교통량 및 예측통행시간을 생성한다(KEC, 2009).

한국도로공사가 운용 중인 교통예보지원시스템은 Fig. 1에서 보인 바와 같이 과거 5년간의 교통자료를 이용하여 교통량, 통행시간 등 다양한 교통정보를 예측하고 분석할 수 있는 정보시스템이다(KEC, 2010). 이 시스템에서 생성되는 예측모형은 각 P-CODE(요일별 패턴)에 해당하는 20주(혹은 20일)의 통행 패턴을 주 자료로 활용하고 시계열 분석방법인 지수평활법과 자기회귀이동평균(Auto Regressive Integrated Moving Average; ARIMA) 모형을 이용하여 1시간 단위로 통행시간을 예측하고 있다(KEC, 2010). 여기서 통행 패턴은 하루 단위로 구축되는 24시간 통행시간의 조합을 의미한다.

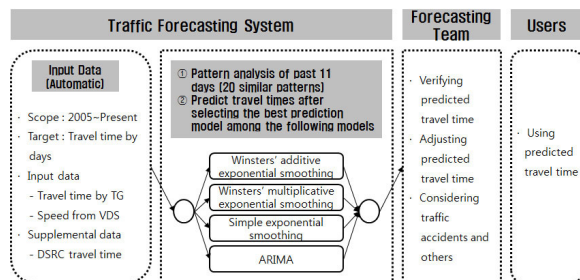


Fig. 1 Prediction Process for Travel Times

ARIMA 모형은 시계열의 발생과정에 확률과정 이론을 도입하여 모형화한 방법으로서 거의 모든 시계열 자료에 사용할 수 있고 시간의 흐름에 따라 빠르게 변동하는 시계열에도 적용할 수 있다는 장점이 있으나, 생성된 모형에 대한 의미 해석이 어렵다는 단점이 있다(Yoon, 1995). ARIMA 모형은 일반적으로 $ARIMA(p, d, q) (P, D, Q)$ 로 표현한다. 자기회귀(autoregressive, AR) 요소와 이동평균(moving average, MA) 요소를 고려한 모형이다. ARIMA 모형은 예측과 관련하여 계절적 패턴과 비계절적 패턴을 동시에 포함하고 있으므로 비계절 모형(p, d, q)과 계절모형(P, D, Q)의 조합인 형태로 표현할 수 있다. ARIMA 모형도 지수평활법과 마찬가지로 먼 과거보다는 최근시점에 가까운 관측값에 더 많은 비중을 두고 있으며 단기예측에 적합하다(Joo, 2012). 참고로, 본 연구에서는 고속버스 통행시간 예측을 위하여 $ARIMA(1,0,0)(1,0,0)$ 모형이 사용되었다(KEC, 2010).

지수평활법은 모형을 설정하지 않고 과거의 관측치만의 평균을 이용하여 예측치를 구하는 방법이다. 지수평활법의 종류로서 단순지수평활법, 이중지수평활법, 삼중지수평활법, 윈터스계절지수평활법 등이 있다. 이 중에서 단순지수평활법(simple exponentially smoothing)은 시계열에 평활 및 분해법을 적용하여 불규칙 변동을 제거하고 관심이 되는 변동만을 식별하는 방법으로서, 추세와 계절 효과가 없고 시간에 따라 일정한 계열에 적합하다(Yoon, 1995). 지수평활 모형은 단기예측을 실시하는 데에 가장 많이 이용되며 정확성이 높고, 모형의 설정이 쉬우며, 모형의 정확성을 시험하기 쉬운 장점이 있다(Park and Kim, 2005).

한국도로공사에서 운용 중인 교통예보지원시스템의 예측흐름은 SPSS 예측모형을 적용하는데, 우선 TCS 자료를 이용하여 과거 요일패턴을 탐색하고 동일패턴 최근 20개를 추출하여 지수평활법과 ARIMA 모형 중 최적선택 후 통행시간을 예측하게 된다. 예측결과는 대도시구간의 2일간(시간단위)에 대한 통행시간 예측값을 제공하고 있다(KEC, 2010).

2.2. 선행연구 고찰

선행연구 고찰을 위해 DSRC 자료를 사용하였거나, 통행시간의 예측 정확도를 높이기 위해 추진되었던 연구를 검토하였다.

Park and Kim(2005)은 DSRC 방식에 의해 수집된 프로브 자료에 대한 분석을 실시하여 확인된 프로브 자료의 특성을 통해, 통행시간 예측의 신뢰도 확보를 위해 다수의 프로브 차량을 확보해야 한다는 주장은 재검토되어야 한다고 제시하였다. 또한 대기행렬검지기와 프로브 자료의 상관관계 분석을 통해 검지기 속도가 높으면 프로브 자료의 변동이 커지므로 검지기 속도가 높을 때는 실시간 수집자료나 전시간대의 자료를 활용하여 통계적으로 통행시간을 산출하는 방법 역시 문제가 있다고 주장하였다.

Sim et al.(2007)은 Radio Frequency Identification(RFID)을 활용하여 RFID방식에 적합한 링크 통행시간 추정기법을 개발하였다. RFID는 리더기를 중심으로 통행시간이 생성되기 때문에 이에 적합하도록 교통정보생성구간을 생성하고 통행시간을 추정하였다. 그러나 이러한 결과는 소수의 테스트 차량의 분석결과이므로 보다 정확한 평가를 하기 위해서는 번호판 조사 기법과 같은 대단위 조사를 통해 추가적인 평가가 이루어질 필요가 있다.

Lee(2009)는 단거리 DSRC 평가모듈을 개발하고 교통류 시뮬레이션 소프트웨어인 CORSIM(version 5.0)에 이식하여 다양한 교통상황을 반영할 수 있는 평가모듈을 개발하고 방법론을 제시하였다. 개발된 평가모듈의 효과적으로 통행시간을 설정하여 통신시간의 이론값과 각 프레임 구조별 통신시간을 통계적으로 검증하여 평가모듈의 유효성을 입증하였다. 기존의 교통정보수집 및 제공시스템 평가방법은 현장실험을 통한 제한된 시스템 평가방법이었다면 본 연구에서 개발된 평가모듈은 교통류 시뮬레이션 상에서 다양한 교통상황을 반영할 수 있는 방법이라고 할 수 있다.

Kang and Namkoong(2002)은 TCS에서 수집되는 톨게이트 간 통행시간 자료를 기반으로 TCS통행시간 속성과 시계열적 패턴을 규명하고, 이를 바탕으로 모듈러 신경망모형(Modular Neural Network Model)을 이용한 통행시간 예측모형을 개발하였다. 본 연구에서 개발된 모형은 예측범위가 고정적인 대부분의 시계열모형과는 달리 최소의 입력(3개)을 가지면서 하나의 신경망으로 학습한 최대/최소의 예측시간범위 내에서 그 크기에 상관없이 거의 동일한 수준의 예측을 보이는 장점을 가지고 있다.

Yoshikazu et al.(2000)은 TCS 자료를 활용하여 통행시간을 예측하는 모형을 제시하였다. 예측방법은 패턴매칭법을 이용하였고, 현장에서 수집된 자료를 활용하여 검증하였다. TCS 자료만을 이용한 예측기법으로 다른 검지장치의 보완 없이 현장에서 쉽게 적용할 수 있는 장점이 있다. 그러나 패턴매칭 모형은 방대한 자료를 필요로 하며 현재 통행시간 또는 과거 통행시간 패턴이 미래에도 주기적으로 반복된다는 가정을 전제해야 하는 한계가 있다. 또한 정확한 패턴분류가 선행되어야 하며 타 모형에 비해 교통상황이 급격히 변화하는 경우 패턴분류의 어려움으로 인해 모형이 이를 설명하기 어렵다. 또한 실제 적용상에 많은 시간이 요구된다는 점과 시간 처짐현상이 발생한다는 문제점이 있다.

Olivier and Faouzi(2006)는 인접구간 통행료자동지불시스템(Electronic Toll Collection, ETC) 자료를 이용하여 단거리 영업소간 통행시간을 추정하였다. ETC에 의한 통행시간 추정 시 혼잡시간대 ETC 자료가 충분하지 않은 경우 통행시간 추정치의 신뢰구간이 커지는 경향이 있는데 이 경우 인접 영업소까지 통행시간 자료를 이용하여 대상구간의 통행시간을 계산하여 해당 문제를 해결하였다. 도시부와 같이 단거리 통행이 잦은 지역에서는 이와 같은 방법이 가능하지만 장거리 통행이 많은 도시외곽지역에서는 적용하기가 어려운 단점이 있다.

2.3. 선행연구와의 차별성

기존의 통행시간 예측 연구는 TCS, GPS(Global Positioning System), RFID 등의 구간검지체계를 이용하여 대부분 진행되어 왔다. 특히 우리나라 고속도로 구간의 경우 기중점 영업소 간에서 실제 소요된 통행시간을 알 수 있는 TCS자료를 많이 이용하고 있다.

하지만 TCS자료를 입력자료로 하는 연구는 구간의 길이가 길어질수록 휴게소 체류차량, 갓길 주행 등에 의한 비정상적인 통행시간을 갖는 자료가 많고 차종별 분류가 명확하지 않아 예측의 정확도가 떨어질 수 있다. 또한, 고속도로의 차종분류체계에서도 버스는 화물차량과 함께 3종으로 분류되어 고속버스 고유의 운행특성을 분석하는데 한계가 있다.

이러한 특성으로 인하여 기존의 통행시간 예측 등의 연구에 고속버스의 명확한 운행특성이 반영되지 않아 이상치 제거의 필요성이 있어 본 연구에서는 고속버스만의 운행자료를 통해 고속버스의 운행특성을 분석하여 예측의 정확도를 제고하고자 하였다.

따라서 고속버스 DSRC 전수자료를 분석하여 구간 추정값이나 샘플조사에 머물렀던 고속버스 운행자료에 대한 명확한 특성 분석을 실시하였다.

3. 고속버스 운행자료 수집

3.1. 고속버스 운행자료 수집

고속버스 운행자료는 각 고속버스에 장착된 차량단말장치(On-Board Equipment, OBE)에서 DSRC를 사용하여 노변장치(Road Side Equipment, RSE)를 통해 산출된 자료이다. 즉, 차량단말장치를 차내에 설치한 고속버스가 고속도로 주요 지점에 설치한 노변장치를 통과할 때 통과한 고속버스의 식별번호 및 통과시간이 기록되며, 본 연구에서는 이러한 자료를 고속버스 운행자료라고 명명하였다.

서울에서 대전간의 고속버스 통행시간 자료의 생성을 위해 1시간 단위로 서울에서 대전에 도착하는 버스를 추출하고, 대전 도착 버스의 서울 출발시간을 추출하여, 대전 도착시간에서 서울 출발시간을 빼주었다. 대전에서 서울간의 경우에도 동일한 방법을 적용하였다. 1시간 단위 통행시간 자료가 생성되면 여기에 날짜코드, 시간정보, 지역별 날씨(강수량, 강우량)정보를 결합하여 결측값을 보정하고 통행시간의 이상치를 보정하였다.

맑은 날씨를 기준으로 자료를 생성하였으며, 생성된 서울~대전구간 통행시간 자료의 기간은 2008년 10월 10일에서 2012년 3월 30일까지로 1,205일간의 자료를 탐색하였다.

3.2. 이상치 및 결측값 보정

이상치(outlier)란 관측된 자료의 범위에서 많이 벗어난 아주 작은 값이나 아주 큰 값을 말한다. 어떤 의사결정을 하는데 필요한 자료를 분석할 경우 이렇게 이상한 값들에 의해서 의사결정에 영향을 미칠 수 있으므로 전처리과정(pre-process)을 통해 제거할 필요가 있다 (Kang and Jeng, 2013).

승용차와 같이 버스전용차로를 이용할 수 없는 차량들은 교통혼잡 등으로 인하여 통행시간이 쉽게 와해될 수 있기 때문에 다소 복잡한 전처리과정이 필요할 수 있다. 하지만, 서울~대전구간의 고속버스 통행시간은 버스전용차로 설치 등으로 인해 대부분 일정한 값을 나타내고 있었다. 따라서 본 연구에서는 서울과 대전구간의 고속버스 통행시간에 대한 이상치를 통행시간의 평균 $\pm 2\sigma$ 범위 밖에 존재하는 통행시간으로 정의하였으며, $\pm 2\sigma$ 범위 이상 혹은 이하인 경우는 이상치로 간주하여 제거하였다. 이러한 근거는 특정일의 1시간 동안 서울에서 출발하여 대전에 도착한 고속버스의 통행시간이 정규분포를 따르는 지 KS-검증(Kolmogorov-Smirnov test)을 시행한 결과 신뢰도 95%에서 p값이 0.15 이상으로 나타나 정규분포를 따르는 것으로 분석되었다. 고속버스의 통행시간이 정규분포를 따르는 경우 이상치 제거를 위하여 $\pm 2\sigma$ 또는 $\pm 3\sigma$ 를 많이 사용한다. $\pm 3\sigma$ 를 사용할 경우 원래 자료의 약 99.7%가 포함되고 $\pm 2\sigma$ 를 사용할 경우 95.4%가 포함된다(Olsson et al., 2010). 본 연구에서는 두 가지 기준 중에서 상대적으로 이상치 제거성능이 강한 $\pm 2\sigma$ 를 사용하였다.

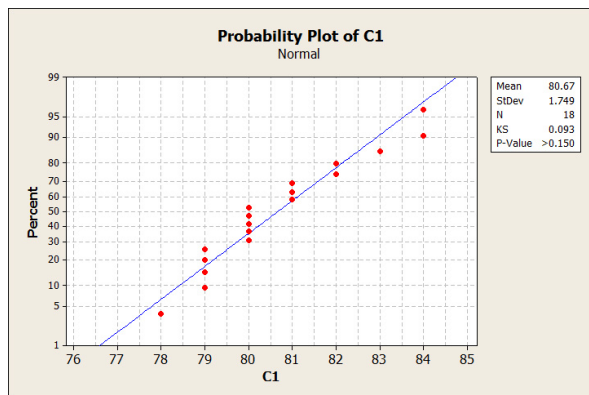


Fig. 2 Normality Test Using KS-Test

이상치 제거 후 결측값의 보정은 두 단계를 거쳤는데 우선 시간대별 결측률이 80% 이상인 경우 해당일의 통행시간의 중위수(median)로 보정하였다. 여기서 결측률이란 시간대별로 전체 자료의 수에서 고속버스 통행시간이 존재하지 않는 자료의 수의 비율을 말한다. 여기서 결측률이 80% 이상인 시간대는 대부분 고속버스가 운행하지 않는 심야시간대인 것으로 조사되었다. 이러한 심야시간대라도 시간대별 통행시간의 대푯값이 필요하므로, 상대적으로 이상치에 영향을 적게 받는 중위수를 이용하여 대푯값을 설정하였다. 여기서 높은 결측률을 보이는 심야시간대는 고속버스 통행시간 예측과정에서 포함되지 않았음을 밝힌다. 시간대별 결측률이 80% 미만인 경우에 해당하는 결측값은 선형보간법을 사용하여 보정하였다. 시간대별 결측률이 80% 미만인 경우는 짧은 시간 동안 자료를 수집하지 못하는 경우이므로 선형보간법으로 결측시간대를 보완하였다. 선형보간법이란 결측값 바로 이전의 값과 마지막 결측값 다음 값의 선형 값으로 대체하는 방법이다.

4. 고속버스 운행특성 분석

4.1. 요일별 결측률 및 통행시간 분석

Table 1에서 보인 바와 같이 대전에서 서울로 운행하는 상행(northbound, NB) 노선의 요일별 결측률은 전체 평균이 21.8%이다. 화요일이 23.1%로 가장 높은 결측률을 보였으며, 월요일은 19.1%로 가장 낮은 결측률을 보이는 것으로 분석되었다.

Table 1. Northbound Missing Rates and Travel Times by Days

Days	Total Days	Missing Rates		Average Travel Time (min.)	
		Missing Days	Rates	Before Pre-Process	After Pre-Process
Mon.	173	33	19.1%	82.2	80.2
Tue.	173	40	23.1%	81.8	80.5
Wed.	172	39	22.7%	81.9	80.7
Thu.	170	39	22.9%	82.1	80.7
Fri.	172	38	22.1%	86.7	83.6
Sat.	173	37	21.4%	83.4	81.3
Sun.	172	37	21.5%	89.5	84.5
Total	1,205	38	21.8%	83.9	81.6

이상치 제거 및 결측값 보정을 포함한 전처리 전 요일별 통행시간의 경우 전체 평균이 83.9분이었으며, 금요일

일과 일요일의 평균 통행시간이 다른 요일에 비해 높은 것으로 나타났다.

전처리 후 통행시간은 전체 평균이 81.6분으로 다소 낮아졌으며, 보정 전과 마찬가지로 금요일과 일요일의 평균 통행시간이 다른 요일에 비해 높은 것으로 나타났으나 전처리 전 보다는 편차가 줄어드는 것으로 분석되었다.

Table 2와 같이 서울에서 대전으로 운행하는 하행 (southbound, SB) 노선의 요일별 결측률의 경우 전체 평균이 22.2%이고, 목요일이 22.9%로 가장 높은 결측률을 보였으며, 토요일은 20.8%로 가장 낮은 결측률을 보이는 것으로 분석되었다.

Table 2. Southbound Missing Rates and Travel Times by Days

Southbound		Missing Rates		Average Travel Time (min.)	
Days	Total Days	Missing Days	Rates	Before Pre-Process	After Pre-Process
Mon.	173	39	22.5%	80.9	80.1
Tue.	173	39	22.5%	80.4	79.8
Wed.	172	39	22.7%	80.3	79.8
Thu.	170	39	22.9%	80.6	79.9
Fri.	172	38	22.1%	82.5	81.0
Sat.	173	36	20.8%	84.3	82.2
Sun.	172	37	21.5%	79.3	78.8
Total	1,205	38	22.2%	81.2	80.2

전처리 전 요일별 도착 통행시간의 경우 전체 평균이 81.2분이었으며, 금요일과 토요일의 평균 통행시간이 다른 요일에 비해 높은 것으로 나타났다. 전처리 후 통행시간은 전체 평균이 80.2분으로 다소 낮아졌으며, 전처리 전과 마찬가지로 금요일과 토요일의 평균 통행시간이 다른 요일에 비해 높은 것으로 나타났으나 역시 보정 전 보다는 편차가 줄어드는 것으로 분석되었다.

4.2. 시간대별 결측률 및 통행시간 분석

Table 3에서 보인 바와 같이 대전에서 서울구간 상행 노선의 시간대별 결측률은 1시에서 5시 사이에 높은 결측률을 보이고 있는 것으로 나타났다. 참고로 높은 결측률을 보이는 상기 시간대는 고속버스 통행시간 예측에서 제외되었다. 통행시간의 경우 이상치 제거 및 결측값 보정 후 16시에서 18시대가 오래 소요되는 것으로 분석되었다.

Table 3. Northbound Missing Rates and Travel Times by Days

Northbound		Missing Rates		Average Travel Time (min.)	
Hours	Total Hours	Missing Hours	Rates	Before Pre-Process	After Pre-Process
0	1,205	337	28.0%	82.1	80.4
1	1,205	1,046	86.8%	96.5	80.4
2	1,205	1,174	97.4%	104.5	80.4
3	1,205	1,202	99.8%	272.0	80.4
4	1,205	1,197	99.3%	148.9	80.4
5	1,205	994	82.5%	93.2	80.0
6	1,205	6	0.5%	79.4	79.1
7	1,205	4	0.3%	81.3	79.2
8	1,205	6	0.5%	80.8	79.5
9	1,205	5	0.4%	81.5	80.2
10	1,205	13	1.1%	82.2	80.3
11	1,205	7	0.6%	81.1	80.2
12	1,205	5	0.4%	81.7	80.4
13	1,205	6	0.5%	83.2	81.3
14	1,205	4	0.3%	83.6	82.6
15	1,205	6	0.5%	86.1	84.6
16	1,205	5	0.4%	88.5	86.9
17	1,205	13	1.1%	88.9	86.7
18	1,205	5	0.4%	87.2	85.4
19	1,205	4	0.3%	85.6	84.0
20	1,205	8	0.7%	87.0	84.4
21	1,205	47	3.9%	85.4	82.4
22	1,205	59	4.9%	83.1	80.6
23	1,205	152	12.6%	81.2	79.6
Ave.	1,205	263	21.8%	96.0	81.6

Table 4에서 보인 바와 같이 서울에서 대전구간 하행 노선에서도 상행 노선과 마찬가지로 1시에서 5시대 사이의 결측률이 80% 이상인 것으로 나타났다. 통행시간의 경우 시간대별로 차이가 거의 없이 일정하게 유지되는 것으로 분석되었다.

Table 4. Southbound Missing Rates and Travel Times by Times

Southbound		Missing Rates		Average Travel Time (min.)	
Hours	Total Hours	Missing Hours	Rates	Before Pre-Process	After Pre-Process
0	1,205	222	18.4%	78.7	79.6
1	1,205	1,188	98.6%	111.1	80.0
2	1,205	1,199	99.5%	108.2	80.0
3	1,205	1,205	100.0%	-	-
4	1,205	1,195	99.2%	102.9	80.0
5	1,205	1,185	98.3%	99.9	79.8
6	1,205	28	2.3%	80.2	79.7

7	1,205	12	1.0%	81.0	80.2
8	1,205	5	0.4%	81.8	80.8
9	1,205	2	0.2%	82.9	81.2
10	1,205	6	0.5%	83.3	81.9
11	1,205	5	0.4%	83.0	81.5
12	1,205	9	1.2%	82.1	80.8
13	1,205	5	0.5%	81.2	80.4
14	1,205	15	0.1%	81.2	80.3
15	1,205	6	0.3%	81.1	80.5
16	1,205	1	0.1%	81.5	80.8
17	1,205	4	0.3%	82.1	81.0
18	1,205	7	0.6%	81.2	80.3
19	1,205	18	1.5%	80.5	79.7
20	1,205	11	0.9%	80.2	79.5
21	1,205	12	1.0%	80.2	79.3
22	1,205	14	1.2%	79.9	78.9
23	1,205	52	4.3%	79.3	78.7
Ave.	1,205	267	22.2%	85.4	80.2

4.3. 요일·시간대별 결측률 및 통행시간 분석결과

요일별 고속버스의 운행특성을 정리하면 상행선의 경우에는 요일별로 큰 차이는 없지만 일요일은 교통량 증가로 인해 통행시간이 매우 큰 편차를 보였다. 최대 통행시간이 166분, 최소 통행시간이 70분으로 96분의 차이를 보였다. 하행선의 경우에도 요일별로 큰 차이는 없었지만 토요일은 최대 도착 통행시간이 172분, 최소 통행시간이 71분으로 101분의 큰 차이를 나타내는 것으로 분석되었다.

시간대별 고속버스의 운행특성을 정리하면 상행선의 경우에는 오전 12시까지는 최대 통행시간이 81분~97분이었으나 오후 13시부터 오후 22시까지는 107분~166분으로 오전보다는 오후시간이 26분~85분 오래 걸리는 것으로 분석되었다. 하행선의 경우에는 오전시간대의 통행시간이 오후시간대보다 오래 걸리는 것으로 분석되었다. 따라서 시간대별 운행특성을 정리하면, 상행선은 오전, 하행선은 오후 시간대의 통행시간이 적게 걸리는 것으로 분석되었다.

이에 따라 요일과 시간대별의 고속버스 운행특성을 함께 정리하면, 평균적인 통행시간은 약 79분~90분으로 나타났고, 금요일(상행선)을 제외한 모든 요일이 시간대별로 비슷한 통행시간을 나타내는 것으로 분석되었다.

또한 본 연구에서는 전처리 후의 값을 대상으로 분석하였는데, 전처리 전의 값과는 상당히 큰 차이를 나타냈다. 요일별 상행선의 통행시간을 살펴보면 전처리 전의 최대 통행시간은 402분, 최소 통행시간은 73분이고, 전

처리 후의 최대 통행시간은 130분, 최소 통행시간은 74분으로 나타났다.

5. 고속버스 통행시간 예측모형의 입력자료 분석기간 산정

5.1. 예측모형의 분석기법

본 연구에서는 한국도로공사가 운영 중인 교통예보지원시스템에서 사용하고 있는 통행시간 예측모형을 준용하여 예측모형 입력값들의 분석기간을 다양하게 변화시켜 고속버스 통행시간 예측의 신뢰도 제고를 위한 적절한 분석기간을 찾고자 한다. 이를 위해서 교통예보지원시스템에서 사용하고 있는 대표적인 예측모델인 ARIMA 모형과 지수평활방법을 본 연구에서 적용하였다. 참고로 현재의 교통예보지원시스템은 20주를 기본 분석기간으로 사용하고 있다.

고속버스의 경우 일반적인 개인 교통과 달리 사전에 결정된 정기적인 운행 스케줄에 의해 운행되는 특성을 갖고 있어 기존 승용차 위주의 예측에 사용되던 20주를 기본 분석기간으로 하는 교통예보지원시스템 방식의 적합성에 대한 검토가 필요할 것으로 판단되었다.

이를 검증하기 위하여 기존의 예측 기준을 포함하여 총 6개의 시나리오를 설정하여 각 시나리오 별 예측의 정확도를 평가하는 과정을 거쳐 적정 분석기간을 선정하기로 하였다.

5.2. 분석자료 선정방안 및 평가기준

현재 교통예보지원시스템에서는 통행시간 예측 시 과거 20주의 통행시간 패턴을 이용하여 예측한다. 하지만 6월 첫 주 토요일을 예측하는 경우 1월에서 5월까지 자료를 사용하게 되는데, 6월 예측 시 1월 자료를 꼭 사용할 필요가 있는지에 대해 의문점이 생길 수 있다. 따라서 본 연구에서는 좀 더 정확한 통행시간 예측을 위하여 통행시간 예측에 사용하는 자료 분석의 시간적 범위를 다양하게 탐색하고자 하였다.

우선 분석기간은 서울과 대전구간의 주말로 선정하였으며, 토요일인 2012년 6월 2일과 6월 9일, 일요일인 6월 3일과 6월 10일을 대상으로 하였다.

분석기간에 대해 Table 5와 같이 총 6가지의 시나리오를 생성하여 그 결과를 비교하였다.

우선 첫 번째로는 기존에 사용되는 것과 동일한 방법으로 예측일을 기준으로 20주의 통행시간 패턴을 이용

하여 통행시간을 예측하는 것이다. 예를 들어 2012년 6월 2일 토요일의 20주 패턴이란 전 주 토요일인 2011년 5월 19일부터 2011년 12월 31일 까지 20주의 토요일 자료를 사용하는 것을 말한다. 두 번째로는 예측일 기준 10주의 패턴을 사용하여 통행시간을 예측하고, 세 번째로는 예측일에 해당하는 월의 작년 자료를 반영하는 것으로서 예를 들어 2012년 6월 2일을 예측한다면 2011년 6월의 자료를 사용하는 것이다. 네 번째로는 예측일에 해당하는 월의 작년 통행시간 자료와 최근 5주 자료를 반영하는 것으로서 예를 들어 2012년 6월 2일을 예측한다면 2011년 6월의 자료와 최근 5주 자료를 함께 사용하는 것이다. 다섯 번째로는 예측일 전일의 교통량과 비슷한 교통량(오차율 ±10만대 이내)을 보이는 전일을 갖는 날들의 통행시간 패턴 10개를 이용하여 통행시간을 예측하는 것이고, 마지막 여섯 번째도 예측일 전일의 교통량과 비슷한 교통량(오차율 ±20만대 이내)을 보이는 전일을 갖는 날들의 통행시간 패턴 10개를 이용하여 통행시간을 예측하는 것이다.

마지막 두 가지 case의 경우를 포함한 이유는 다음과 같다. 고속버스의 경우 경부선 버스전용차로 운영 등으로 인하여 출발일자 교통량에 민감하게 영향을 받을 수밖에 없는 운행성격을 갖고 있다고 볼 수 있다. 이러한 판단아래 분석기간의 적합성과 더불어 통행시간은 교통량과의 관련성이 높은 것으로 알려지고 있다. 따라서 전 날의 교통량의 패턴들이 유사한 기간과의 비교를 통하여 가장 적합한 분석기간을 선정하는 방법론을 선택하였다.

Table 5. Scenarios for Analysis Periods

Cases	Scenarios
Case 1	Past 20 weeks data from the day to be predicted
Case 2	Past 10 weeks data from the day to be predicted
Case 3	Same month data of the last year as that of the day to be predicted
Case 4	Same month data of the last year as that of the day to be predicted + recent 5 weeks data
Case 5	10 patterns whose traffic volumes of the previous day are similar to those of the previous day of the day to be predicted (error rate ±100,000 vehicles)
Case 6	10 patterns whose traffic volumes of the previous day are similar to those of the previous day of the day to be predicted (error rate ±200,000 vehicles)

모형의 사후평가는 예측값과 실제값의 차이인 예측 오차의 크기에 따라 예측의 질을 평가하는 것으로 여러 가

지 평가척도가 있다. 그 중 몇 가지를 살펴보면 평균비율오차, 평균절대비율오차, 불균등계수, 자누스 계수 등이 있다(Lee, 1999). 본 연구에서는 모형평가기준으로 일반적으로 널리 사용되는 평균절대비율오차 방식을 사용하였다. 평균절대비율오차(Mean Absolute Percentage Error, MAPE)는 대상 집단의 오차의 범위를 산정하는데 있어 중요한 방법이며 Eq. (1)와 같이 계산된다.

$$MAPE = \frac{100}{T} \sum_{t=0}^T \left| \frac{z_t - \hat{z}_t}{z_t} \right| \quad (1)$$

전체 MAPE는 모형에 사용된 자료에 의해 산출된 예측결과와 전체 오차율을 나타내는 것이고, 예측일 MAPE는 예측모형을 이용하여 도출한 예측일 전체 예측시간 간격(1시간)에 대한 정확도를 나타내는 것이다. 따라서 전체 MAPE 및 예측일의 MAPE 오차율이 적을 수록 정확도가 높은 것을 의미하는 것이다.

MAPE의 경우 실제값에 비해 예측값이 과도하게 작거나 또는 클 경우 계산된 MAPE 값이 왜곡되는 현상이 존재한다(Yoon, 1995). 하지만, MAPE를 적용해 본 결과 Table 6에서 보인 바와 같이 예측값과 실제값에 큰 차이가 없어 적용하는 데 무리가 없는 것으로 판단되었다.

5.3. 예측모형의 시간적 범위에 따른 적합도 평가

예측일 MAPE란 예측모형을 이용하여 도출한 예측일에 대한 정확도를 말한다. Table 6의 ARIMA 모형 예측일 MAPE 비교현황을 보면 예측일 전일과 비슷한 교통량을 보이는 전일을 갖는 특정일의 통행시간 패턴을 사용하는 case의 오차율이 적었다. 하지만 이러한 방식의 단점은 예측일 전일의 교통량이 확보된 이후에 예측의 통행시간 패턴을 예측할 수 있다는 것이다.

Table 6. Comparison of MAPE Using ARIMA Model

(unit : %)

Cases	Case				Saturday				Ave.
	June 2, 2012		June 9, 2012		June 3, 2012		June 10, 2012		
	NB	SB	NB	SB	NB	SB	NB	SB	
1	5.51	4.01	1.57	2.15	6.32	0.90	2.35	1.14	2.99
2	5.26	4.54	1.53	2.34	7.92	0.52	4.83	1.16	3.51
3	5.61	2.37	2.29	4.64	2.36	1.14	3.74	1.75	2.99
4	5.03	2.41	1.61	2.87	5.05	0.71	3.99	1.17	2.86
5	5.16	2.31	1.45	3.29	1.16	0.59	3.05	1.32	2.29
6	5.25	1.62	1.67	2.31	1.05	0.66	3.01	1.05	2.08

Table 7에서 제시된 지수평활 모형의 예측일 MAPE 비교현황을 보면 예측일 전일과 비슷한 교통량을 보이는 전일을 갖는 패턴을 사용하는 case의 오차율이 적었다.

Table 7. Comparison of MAPE Using Exponential Smoothing Model (unit : %)

Cases	Case				Saturday				Ave.
	June 2, 2012		June 9, 2012		June 3, 2012		June 10, 2012		
	NB	SB	NB	SB	NB	SB	NB	SB	
1	5.68	4.80	1.58	2.88	14.54	0.58	1.73	1.14	4.12
2	5.39	4.27	1.52	2.25	17.36	0.68	6.86	1.09	4.93
3	6.24	1.12	3.01	3.39	3.31	3.61	2.73	3.53	3.37
4	5.27	1.75	1.30	3.18	10.35	0.61	3.92	1.08	3.43
5	5.23	1.70	1.36	3.66	1.08	0.61	2.22	1.16	2.13
6	5.33	1.47	1.81	3.19	0.91	1.42	1.33	1.01	2.06

분석결과 모형의 종류(ARIMA, 지수평활)에 따라 오차율의 차이는 크게 나타나지 않아 모형이 오차율에 미치는 영향은 적은 것으로 분석되었고, 고속버스 통행시간 예측모형의 분석기간 산정에 있어 예측일의 MAPE 오차율이 적을수록 정확도가 높은 것을 의미하므로 예측일 전일과 비슷한 교통량을 보이는 전일을 갖는 패턴을 입력자료로 사용하는 경우인 Case 5와 Case 6이 가장 적합한 것으로 분석되었다.

6. 결론 및 향후 연구과제

본 연구는 DSRC 자료를 활용하여 고속버스의 통행시간에 대한 예측의 정확도를 높이는데 그 주안점을 두었다. 따라서 고속버스에 부착되어 있는 하이패스 단말기로부터 2008년 10월 10일에서 2012년 3월 30일 까지 총 42개월 동안 수집된 서울에서 대전구간의 DSRC 자료를 집계하여 고속버스의 운행특성을 분석하였고, 교통예보지원시스템에서 사용하고 있는 예측모형을 준용하여 예측모델 입력값들의 분석기간을 다양하게 변화시켜 고속버스 통행시간 예측의 신뢰도 제고를 위한 적절한 분석기간을 결정하였다.

이와 같은 DSRC 자료를 이용하여 교통예보지원시스템에서 사용하는 예측모델 입력값들의 분석기간을 다양하게 변화시켜 고속버스 통행시간 예측의 신뢰도 제고를 위한 적절한 분석기간을 결정하였는데, 총 6가지의 경우로 분석기간 자료 선정방안을 마련하여 토요일인 2012년 6월 2일, 9일과 일요일인 6월 3일, 10일을 대

상으로 ARIMA방법과 지수평활방법에 따라 각각의 오차율을 분석하였다.

분석기간에 따른 적합도를 평가한 결과 전체 MAPE (평균절대비율오차)는 전반적으로 모든 방안들이 낮게 나타났으나, 예측일 MAPE는 예측일 전일과 비슷한 교통량을 보이는 전일을 갖는 패턴을 이용하는 경우가 다른 방안들에 비해 낮게 나타났다. 예측일에 해당하는 전년 동월 자료를 사용하는 경우에는 오차율이 크게 나타나 이를 사용하기에는 부적합한 것으로 판단되고, 예측일 전일과 비슷한 교통량을 보이는 전일을 갖는 패턴을 사용하는 경우가 전반적으로 오차율이 작게 나타났다.

따라서 DSRC 자료를 활용하여 통행시간을 예측하는 경우의 입력자료 분석기간은 예측일 전일과 비슷한 교통량을 보이는 전일을 갖는 패턴을 사용하여 모델링을 하는 것이 가장 적합한 것으로 나타났다. 이러한 결과를 토대로 향후 고속도로 구간의 교통정보예측 시의 입력자료는 전국교통량을 사용하도록 검토되어야 할 것이다.

본 연구의 분석자료가 서울~대전구간으로 한정되었기 때문에 추가적인 자료를 수집하여 재검증을 함으로써 본 연구의 신뢰성을 확보하고, DSRC 통과시간을 이용한 통행시간 산출 알고리즘을 개발하여 IC↔IC 구간 예측으로 통일된 예측정보 제공에 관한 연구를 지속적으로 추진해야 할 것으로 사료된다.

감사의 글

본 연구는 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구(2012R1A1A1044620)입니다.

References

- Joo, I., A Case Study on Crime Prediction using Time Series Models, Journal of Korea Security Science Association, Vol. 30, pp. 139-169, 2012.
- Kang, J. and Namkoong, S., Development of The Freeway Operating Time Prediction Model Using Toll Collection System Data, Journal of Korean Society of Transportation, Vol. 20, No. 4, pp. 151-162, 2002.
- Kang, K. and Jeng, W., Easy Statistics, Orae, 2013.
- Korea Expressway Corporation, Practical Development of Expressway Traffic Condition Detection System, 2009.
- Korea Expressway Corporation, Study on the Enhancement of the Traffic Condition Forecasting Support System, 2010.
- Lee, D., Understanding Prediction Methods, Korea Information Industry, 1999.
- Lee, H., Development of A Microscopic Simulation Model for Traffic Information Collection & Service System Based On

- Dedicated Short Range Communication, Master Thesis, Ajou University, 2009.
- Olivier, D. M. and Faouzi, N. E. E., Innovative Processing of Toll Collection Data, LICIT Report No. 0604, 2006
- Olsson, C., Eriksson, A. and Hartley, R., Outlier Removal Using Duality. In CVPR, 2010.
- Park, E. and Kim, H., Identifying Some Relationships between the Probe and the Loop Detector Data for Better Estimation of Link Travel Time, in Proc. the Korea Institute of Intelligent Transport System, pp. 209-214, 2005.
- Sim, S., Choi, K., Lee, K., Link Travel Time Estimation and Evaluation of Applicability of Traffic Information collection Based RFID Probe Data, the Journal of the Korea Institute of Intelligent Transport Systems, Vol. 6, No. 2, pp. 15-25, 2007.
- The Korea Transport Institute, Survey Results of Traffic Condition Forecasting Information, 2010.
- Yoon, Y., Understanding Prediction Methods, Jayou Academy, 1995.
- Yoshikazu, O., Hideki, U. and Masao, K., Travel Time Prediction Method for Expressway Using Toll Collection System Data, Proceedings of 7th ITS World Congress, Torino, 2000.