

BIS 자료를 이용한 중장기 버스 통행시간 예측

이주영¹ · 구은모² · 김형주¹ · 장기태^{1*}

¹한국과학기술원 조천식 녹색교통대학원, ²한국철도기술연구원 기술전략실

Long-term Prediction of Bus Travel Time Using Bus Information System Data

LEE, Jooyoung¹ · Gu, Eunmo² · KIM, Hyungjoo¹ · JANG, Kitae^{1*}

¹The Cho Chun Shik Graduate School of Green Transportation, KAIST, Daejeon 34141, Korea

²Korea Railroad Research Institute, Gyeonggi-do, 437-757, Korea

*Corresponding author: kitae.jang@kaist.ac.kr

Abstract

Recently, various public transportation activation policies are being implemented in order to mitigate traffic congestion in metropolitan areas. Especially in the metropolitan area, the bus information system has been introduced to provide information on the current location of the bus and the estimated arrival time. However, it is difficult to predict the travel time due to repetitive traffic congestion in buses passing through complex urban areas due to repetitive traffic congestion and bus bunching. The previous bus travel time study has difficulties in providing information on route travel time of bus users and information on long-term travel time due to short-term travel time prediction based on the data-driven method. In this study, the path based long-term bus travel time prediction methodology is studied. For this purpose, the training data is composed of 2015 bus travel information and the 2016 data are composed of verification data. We analyze bus travel information and factors affecting bus travel time were classified into departure time, day of week, and weather factors. These factors were used into clusters with similar patterns using self organizing map. Based on the derived clusters, the reference table for bus travel time by day and departure time for sunny and rainy days were constructed. The accuracy of bus travel time derived from this study was verified using the verification data. It is expected that the prediction algorithm of this paper could overcome the limitation of the existing intuitive and empirical approach, and it is possible to improve bus user satisfaction and to establish flexible public transportation policy by improving prediction accuracy.

Keywords: bus travel time, bus information system, long-term prediction, self organizing map, weather information

초록

최근 국내에서는 대도시권의 교통혼잡 완화를 위하여 다양한 대중교통 활성화 정책을 시행 중에 있다. 특히 대도시권역에서는 버스정보시스템이 도입되어 버스의 현재위치, 도착예정시간 등에 대한 정보를 제공하고 있다. 하지만 복잡한 도시부를 지나는 버스들의 경우 반복적인 교통혼잡과 버스몰림으로 인하여 정확한 통행시간 정보제공 시 정확도를 확보하는데 어려움이 있다. 기존 버스 통행시간 연구는 링크별 소통정보 제공방식으로 인하여 버스 이용자의 경로 통행시간 정보 제공 시 어려움이 있고, 데이터 기반의 단기 통행방식으로 중장기 정보 제공이 어렵다는 한계가 있다. 이에 본 연구에서는 경로기반의 중장기 버스통행시간 예측 방법론에 대한

J. Korean Soc. Transp.
Vol.35, No.4, pp.348-359, August 2017
<https://doi.org/10.7470/jkst.2017.35.4.348>

pISSN : 1229-1366
eISSN : 2234-4217

Received: 29 May 2017

Revised: 3 July 2017

Accepted: 25 August 2017

Copyright ©
Korean Society of Transportation

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

연구를 실시한다. 이를 위하여 2015년 버스통행정보로 학습데이터, 2016년 자료로 검증데이터를 구성하였다. 학습 데이터를 이용하여 버스통행정보를 분석하여 버스통행시간에 영향을 미치는 요인들을 출발시각, 요일, 그리고 기상 요인 등으로 분류하고, 이들의 특성 값을 자기조직화지도를 활용하여 비슷한 통행 패턴을 가지는 군집으로 분류하였다. 도출된 군집들을 바탕으로 맑음과 우천시에 대한 요일/출발시각 별 버스통행시간 참조 테이블을 구성하였다. 검증데이터를 이용하여 본 연구에서 도출한 버스통행시간의 정확도를 검증하였다. 본 연구의 중장기 예측 알고리즘을 활용하여 기존의 직관적이고 경험적인 접근법의 한계를 극복할 수 있으며, 예측의 정확도 개선을 통한 버스이용자 만족도 향상 및 탄력적인 대중교통 정책 수립이 가능할 것으로 판단된다.

주요어: 버스 통행시간, 버스정보시스템, 중장기 예측, 자기조직화지도, 기상정보

서론

1. 연구의 배경 및 목적

현재 대도시권에서는 교통 혼잡으로 인하여 막대한 사회·경제적 비용이 발생하고 있으며, 이를 해소하기 위해서 다양한 대중교통 활성화 정책을 시행 중에 있다. 이 중 대중교통 체계에 IT기술을 도입하여 버스의 실시간 운행정보를 수집하고 활용할 수 있는 버스정보시스템(Bus Information System, 이하 BIS)이 도입되어 운영 중에 있다. 이는 버스의 현재위치, 도착예정시간 등을 버스 이용자에게 제공하여 대중교통 서비스의 질적 향상과 이용 활성화 등의 효율적인 대중교통 운영을 도모하기 위함이다. 하지만 복잡한 도시부를 지나는 버스 통행시간의 경우 반복적인 교통 혼잡으로 인하여 예측 통행시간 산정의 어려움이 있으며, 이는 이용자들의 서비스 만족도 및 버스 이용 활성화에 어려움을 가져온다(Foote et al., 1998; Lin et al., 2005). 이렇듯 버스의 예상 운행시간과 예상 도착시간은 운영자 및 이용자 관점에서 중요한 요인으로 작용한다. 운영자 측면에서는 교통 흐름에 따라 적절한 배차간격을 조절할 수 있으며, 이용자 측면에서는 다양한 버스정보를 활용하여 효율적인 의사결정이 가능하다(Ma et al., 2007).

버스 이용자 관점에서 예측 정보는 단기와 중장기로 나누어 분류할 수 있다. 단기 통행시간 예측은 실시간 자료를 기반으로 이용자가 버스 통행을 시작하거나 통행 중일 경우 언제 목적지에 도착할지에 대한 예측정보를 제공한다. 반대로 중장기 통행시간 예측은 이용자가 사전에 버스통행을 계획하는 경우 소요시간 산출을 위한 예측 정보로 정의할 수 있다. 중장기 예측을 위한 데이터는 과거 버스 통행 이력자료를 주로 이용하며, 통행계획은 일반적으로 통행 시작 하루 전 또는 일주일 전 등을 기준으로 요일 및 출발시각 단위로 제공이 가능하다. 따라서 변동성이 높은 단기 예측 정보에 비해 일반적인 예측정보를 제공할 수 있으며, 이를 바탕으로 이용자가 본인의 통행과 관련된 의사결정(출발시간 계획, 교통수단 선택 등)을 효율적으로 하는데 활용될 수 있다. 최근까지 국내에서 활용되고 있는 버스정보서비스의 경우 단기적 관점에서 도착예정시간 정보를 제공하는데 중점을 두고 있어, 이용자들의 통행 의사결정에 도움을 주는 중장기 통행시간 예측 연구에는 미흡한 측면이 존재한다. 또한 일반적으로 통행정보의 산출은 링크기반으로 이뤄진다. 이는 운영자 관점에서 실제 통행 수준의 변화를 모니터링 할 수 있다는 장점이 존재하기 때문이며, 이용자에게 정보 제공시에는 단위 링크 통행시간을 산출하고 경로에 해당되는 통행시간의 합을 산출하여 제공한다. 그러나 이와 같은 방법으로 산출되는 통행정보는 실시간 링크소통정보와 실제 통행속도와는 차이가 존재하기 때문에 실제 이용자에게는 편향된 통행시간을 제공할 수 있다(Kim et al., 2013).

이에 본 연구에서는 버스 운영자 및 이용자 입장의 상충 지·정체 구간에 대해 경로기반의 중장기 버스통행시간 예측에 대한 연구를 실시한다. 이를 위하여 입력변수로 버스 통행시간에 영향을 주는 요소인 요일(주중, 주말, 공휴일), 기상요인(강수량 및 기온) 등을 고려하고, 분석방법으로는 머신러닝 비지도학습(unsupervised learning)의 대표적 알고리즘인 자기조직화지도(Self Organizing Map, 이하 SOM)를 활용하여 예측을 실시한다. 이는 사전에 목표 값을 정의하지 않고 데이터 자체의 결합, 연관성 및 유사성 등을 통하여 클러스터를 도출하는 방식으로 학습된 뉴런에 의해 교통패턴의 변화가 심한 경우에도 신뢰성 있는 예측정보 생성이 가능한 장점을 가진다.

2. 국내·외 선행연구 고찰

기존 버스 통행시간 예측의 경우 데이터에 기반한 단기적 버스 예상 운행시간과 예상 도착시간에 대한 연구가 진행되었다. 대표적으로 이력자료평균(historical average model), 회귀모형(regression model), 칼만필터(kalman filter), SVM(support vector machine) 등을 활용한 연구들이 수행되었다.

이력자료평균 모형의 경우 이전 버스 통행의 과거데이터를 기반으로 버스의 예상 운행시간과 예상 도착시간을 산정할 수 있다. 이는 주기적 패턴을 가지는 교통데이터 특성을 활용한 것으로 과거 동일한 시간대의 교통상황이 유사함을 가정하여 예측 도착시간을 산정한다(Smith et al., 1995). Jeong et al.(2004) 연구에서는 평균정차시간을 제외한 정류장간 링크 통행시간에서 과거데이터를 활용한 모델을 제안하였다. 버스의 예상 도착시간의 경우 과거의 동일한 요일 및 시간대의 자료를 활용하였으며, 교통량이 상대적으로 적고 교통흐름이 안정적인 경우에는 신뢰할 수 있는 정보의 생성이 가능하였다. 하지만 교통상태가 혼잡한 경우 예측의 정확도가 낮은 한계를 보였다.

회귀모형의 경우 독립변수와 종속변수 사이의 선형 상관관계를 이용한 방법으로 다중회귀모형(multiple linear regression) 적용 시 다양한 독립변수의 선택(링크길이, 승객 수, 정류장 및 교차로 수, 정류장간 도착시간, 지리적 요인 등)을 통한 회귀모형이 제안되었다(Abdelfattah, et al., 1998; Chien et al., 2002; Patnaik et al., 2004; Ramakrishna et al., 2006). 하지만 독립변수 조합에 따른 상이한 예측결과 및 제한적 자료수집으로 인하여 혼잡 시 낮은 정확도를 보였으며, 범용적 모형에는 한계를 가졌다.

칼만필터의 경우 재귀적 추정 알고리즘으로 새로운 관측치를 활용하여 통행시간 상태변수를 지속적으로 업데이트하는 방법이다. 상태 전이 모형(state transition model)에 의해 예측 및 추정과정이 수행되며, 오차 공분산을 최소화시키는 방식으로 업데이트된다. 이전 시간대의 정보만을 활용하는 장점이 있으며, 실시간 예측에 우수한 정확도를 보였다. 하지만 예측 정확도가 상태 전이 모형에 의존적이며, 각 시간간격에서 통행시간이 급변하는 경우 time-lag 등의 문제가 발생하는 단점을 보였다(Chen et al., 2004; Yang, 2005).

SVM 모형은 지도학습 기반의 선형분류 모델로 주어진 데이터의 공간 경계를 찾는 알고리즘으로 버스 통행시간 예측에 적용되었다. 머신러닝 기법의 일환으로 최근 버스 통행시간 예측에 다양한 연구들이 수행되었다. 하지만 SVM의 경우 이진분류(Binary Classification)에 적합한 알고리즘으로 파라미터 및 커널 선택에 민감한 단점을 가진다(Vanajakshi et al., 2007; Yu et al., 2011).

3. 선행연구와의 차별성

선행연구검토에서 확인할 수 있듯이 버스 통행시간 예측을 위한 다양한 연구가 진행되었으나 대부분의 연구가 단기적 관점에서의 통행시간 예측을 목적으로 수행되었다. 본 연구는 이용자의 통행계획에 활용되는 중장기적 관점의 버스 통행시간 예측 기법을 제시하는데 차별성이 존재한다. 또한 선행연구에서 사용한 방법론 중 이력자료평균 모형의 경우 교통 흐름이 안정적인 경우에만 신뢰할 수 있는 정보가 생성되며, 회귀모형의 경우 회귀모형의 경우 분석구간에 따른 다수의 모형개발이 필요한 단점을 가진다. 칼만필터의 경우 현재 시점의 자료만 활용하기 때문에 예측의 정확도가 낮으며, SVM 모형은 파라미터 및 커널선택에 민감한 단점을 보였다. 버스 통행시간과 관련된 많은 변수들이 존재하기 때문에 기존 방법론들은 많은 변수들을 가정하여 결과를 도출하였고, 이에 따라 실제 데이터와 비교 시 부정확한 예측결과를 보인다. 이에 본 연구에서는 가정사항들을 최소화할 수 있는 비지도학습 기반의 SOM 알고리즘을 활용하여 중장기 버스 통행시간 예측을 실시한다. SOM의 경우 입력패턴에 대한 목표 값을 사전에 정의하지 않고, 데이터 자체의 결합 및 유사성에 근거하여 스스로 학습하는 방식으로 다양한 입력변수를 융합하여 결과 분석이 가능한 장점을 가진다. 이에 본 연구에서는 예측의 정확도 개선을 위해서 버스 통행시간에 영향을 주는 요소인 요일(주중, 주말, 공휴일), 기상요인(강수량 및 기온) 등을 추가적으로 고려하여 분석을 실시한다.

연구방법론

1. 연구의 내용 및 방법

본 연구에서는 먼저 버스 통행시간 예측 관련 다양한 연구 사례를 검토한다. 다양한 예측 모형에 대한 선행연구 검토를 통하여 본 연구에 적용되는 SOM 방법론에 대한 유용성을 검토하고 이에 대한 적용가능성을 분석한다. 이어 서 이력 및 테스트 BIS 자료를 기반으로 SOM을 활용한 증장기 버스 통행시간 예측을 실시하며, 각 단계별 세부절차 및 변수결정, 입력자료 구축, 그리고 클러스터 과정 등을 제시한다. 마지막으로 결론에서는 본 연구의 성과 및 한계, 그리고 향후 연구방향을 논의한다.

2. 버스 통행시간 예측 흐름도

버스 통행시간 예측의 단위는 링크-노드(link-node) 방식과 경로(path) 방식으로 나눌 수 있다. 링크-노드 방식은 개별 링크-노드의 통행시간 및 정차시간을 예측하여 각각을 취합하여 예측 값을 산출하는 방식이다. 하지만 사고 및 유고 등의 돌발상황 발생 시 예측 단위 별로 오차가 발생하고, 링크-노드를 조합한 정보 생성 시 추가적인 오차가 발생한다. 경로 방식은 교통이용자가 원하는 시·종점에 따른 통행시간 정보를 한 번의 예측을 통해 제공하는 방식으로 예측오차 발생 여지가 링크-노드방식에 비해 적다는 장점을 가진다. BIS 원시 데이터 검토 결과 예측 알고리즘 적용 시 노드(정류장)에서 시간 오차가 발생하는 주요 원인으로 작용하므로 버스 통행시간 예측의 경우에는 노드(정류장)를 단위 링크 통행시간에 포함하여 예측하는 경로방식의 예측이 적합하다.

이에 본 연구에서는 비지도학습 기반의 SOM 알고리즘을 이용하여 총 3단계의 과정을 거쳐 버스 통행시간 예측을 수행한다(Figure 1 참조). 먼저 학습에 사용할 입력자료를 수집 및 구축한다. 이를 바탕으로 SOM 알고리즘을 통해 통행시간의 특성을 패턴화하고, 도출된 패턴들을 K-means 클러스터링을 통해 비슷한 통행패턴들을 그룹화하여 최종 클러스터로 구성한다. 이후 날짜와 시간대를 구분하여 버스통행시간 참조표를 구성하여 통행시간 예측에 활용하였다.

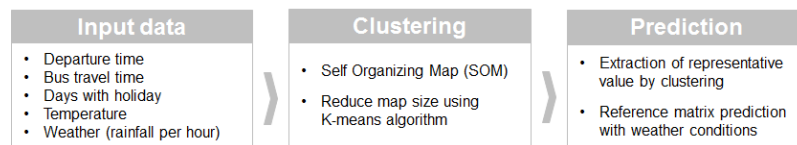


Figure 1. Flow chart for bus travel time prediction

3. 자기조직화지도(Self Organizing Map)

자기조직화지도(SOM)는 주어진 입력 패턴에 대하여 정확한 해답을 미리 주지 않고 자기 스스로 학습하는 비지도학습 알고리즘의 하나로, 비슷한 입력값에 대해 비슷한 출력을 나타낼 수 있도록 연결강도(가중치)를 조정하여 출력층(map)을 스스로 조직화하는 인공신경망이다. 자기조직화지도의 기본요소는 신경망 구조, 뉴런(활성화 함수), 학습방식으로 구분된다. 신경망 구조는 Figure 2와 같으며 입력벡터, 연결강도, 출력층으로 나뉜다. 해당 출력층은 2차원 평면형태로 시각화하여 표현할 수 있다.

본 연구의 자기조직화지도 최적 클러스터 개수는 Garcia et al.(2004)에서 제시된 Equation 1과 같은 식에 결정한다. 이는 샘플 수에 비례하여 충분한 설명력을 확보하기 위한 최적의 클러스터 개수로 학습 데이터들의 특징을 명확히 반영하여 클러스터를 구성할 수 있는 장점을 가진다.

$$TotalSize = 5 \sqrt{\# \text{ of samples}} \quad (1)$$

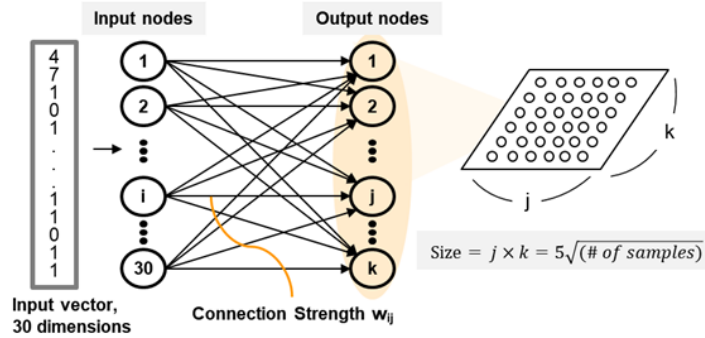


Figure 2. Structure of self organizing map

각 노드는 뉴런(활성화함수)들로 구성되어 있으며 각 뉴런은 연결강도(가중치) 벡터와 입력벡터의 유사성을 계산한다. 그리고 각 뉴런들은 학습할 수 있는 특권을 부여받기 위해 서로 경쟁하려 하는데, 연결강도 벡터와 가장 가까운 뉴런이 승리한다. 이 승자 뉴런만이 출력신호를 보낼 수 있는 유일한 뉴런이며, 이 뉴런 및 인접한 이웃 뉴런들만이 제시된 입력벡터에 대한 학습이 허용된다. 승자 뉴런을 결정하고 난 후 실제 적용되는 신경망의 학습규칙은 Equation 2와 같다.

$$W_{new} = W_{old} + \alpha (X - W_{old}) \tag{2}$$

여기서, W_{old} : 조정되기 이전의 연결강도 벡터

W_{new} : 조정된 후의 새로운 연결강도 벡터

X : 입력패턴 벡터

α : 학습상수

승자의 연결강도는 입력패턴에 가장 유사하기 때문에 연결강도 벡터와 입력패턴 벡터의 차이를 통하여 일정한 비율(학습상수)을 원래의 연결강도에 더하는 방식이다. 반복이 지속될수록 승자뉴런과 함께 연결강도를 조정하는 이웃의 변경 또한 점점 줄어든다. 처음에는 거의 모든 뉴런이 학습을 하나, 점차 훈련이 진행됨에 따라 점점 적은 개수의 뉴런들만 학습을 하며 최종적으로는 승자 뉴런만이 연결강도를 조정한다.

위와 같은 과정을 통해 SOM 학습이 완료되면, K-means 클러스터링 방식을 사용하여 SOM의 결과 값들을 다시 그룹화하여, 지도의 크기 즉 최종 클러스터의 개수를 줄여 일반화된 통행시간 군집을 구축할 수 있다. K-means 클러스터링 수행시 입력 데이터에 적합한 클러스터 개수를 찾기 위해서는 elbow point of WSS(Within group sum of squares) graph를 이용하였다. 클러스터의 개수인 K를 점차 늘려가며 각각의 경우에서 WSS를 계산하여 최소값을 가지는 K를 선택하는 기법이다. K-means 클러스터링 기법 적용시 초기값은 랜덤하게 배정하였으며, 각 K값으로 총 1000번의 클러스터링을 수행하고 이중 가장 좋은 결과를 대표 클러스터 결과로하여 사용하였다. 클러스터 구성이 완료되면 각 클러스터별로 통행시간의 대표 값을 추출하는데, 이때 대표 값은 median 값을 사용한다. 마지막으로 클러스터별 대표값을 기준으로 출발시각/요일로 이루어진 참조표를 구성하여 버스통행시간 예측에 활용한다.

4. 입력변수 설정

SOM을 이용한 버스통행시간 예측을 수행하기 위해서는 충분한 버스 통행시간 정보가 필요하다. 이를 위해 통합 정류장ID, 통합구간ID, 구간통행시간, BIS데이터 테이블을 통합하여 Table 1과 같은 버스통행 자료 테이블을 구성하였다.

Table 1. Table header list of Bus information

Table name	Field name	Types of data	Remark
Integrated bus station ID	StationSEQ	NUMERIC()	Serial number of bus station
	Station_ID	VarChar(9)	Smart card/standard bus station ID
Integrated section ID	F_Station	NUMERIC()	Identical form with serial number of bus station
	T_Station	NUMERIC()	Identical form with serial number of bus station
	Link_ID	NUMERIC()	Serial number of link
Link information	Link_ID	NUMERIC()	Serial number of link
	R_Date	NUMERIC()	Serial number of day
	R_Time	NUMERIC()	Operating time
BIS data	Station_ID	NUMERIC()	Smart card/standard bus station ID
	R_Date	Char(8)	YYYYMMDD
	Route_ID	VarChar(9)	Smart card/standard bus route ID
	Event	Char(2)	Departure/Arrival
	Event_Time	VarChar(10)	HH:MM:SS

중장기적 관점으로 버스의 통행시간은 정류장 체류시간과 노드 통과시간으로 구분되어지며, 노드 통과시간은 이 용객 수가 증가할수록 증가하고 노드 통과시간은 해당 링크의 교통상황에 따라 영향을 받게 된다. 따라서 버스 통행 시간에 영향을 미치는 요인들은 이용객수와 교통상황에 영향을 주는 요인들을 통해 예측할 수 있다. 기존 연구들에서는 통행시간 예측시 날씨(Patnaik et al., 2004; Nookala, L. S., 2006; Faouzi et al., 2010; Qiao et al., 2012), 출퇴근 시간 및 휴일 등의 시간변수(Kown et al., 2000; Xiaoyan Zhang, 2003; Bai et al., 2015) 등을 주요 영향요인으로 이용하였다. 출발시간대 및 휴일 여부에 따라 이용객 수의 차이가 클 것으로 예상되며 첨두시간 변수를 통해 주변 교통상황 반영이 가능하다. 또한 기상요인에 따른 통행량, 통행속도 변화도 영향이 있을 것으로 판단된다. 이에 본 연구에서는 기존 연구들을 참고하여 버스 통행시간에 영향을 미치는 요인들로 요일(주중/주말/공휴일), 기상요인(기온, 강수량), 출발시각(첨두/비첨두)으로 분류하여 적용하였다. 이에 각 요인들을 신경망 입력에 적합한 수치 또는 one-hot 벡터화하여 입력 벡터를 구성하였다. 범주형 변수 값을 가지는 요일과 출발시각은 각각 8개(요일 7개 + 공휴일 1개)와 19개(버스 출발시각 기준 5-23시)로 one-hot 벡터화 하였으며, 기상요인은 수치 값을 그대로 적용하여 2개의 뉴런을 사용하였고, 통행시간 항목을 추가하여 버스타행 한 건당 입력 벡터의 크기는 30 뉴런으로 설정하였다. 위와 같이 구성된 입력 벡터를 0과 1사이의 값으로 정규화하여 클러스터링을 수행하였다.

방법론 적용 및 분석

1. 분석구간 및 자료

본 연구에서는 경기 북부외곽(중산마을2단지)에서 서울 도심(강남역)까지 운행하는 광역버스를 대상으로 분석을 실시하며, 분석 구간의 경우 출·퇴근 시간의 상습 지·정체 구간(일산동부경찰서-강남역)으로 교통상태 변화가 비교적 명확한 구간이다(Figure 3 참조). 중장기 버스 통행시간 예측 시 학습에 활용되는 BIS 이력자료는 2015년 1-9월(9개월), 테스트 자료는 2016년 1-6월(6개월)을 사용하였다. 또한 본 연구에서 사용된 기상자료의 경우 기상청 국가기후데이터센터¹⁾의 시간대별 강수량(mm) 및 기온(°C) 자료를 활용하였다. 시간대별 강수량은 2015년 최대 24mm, 2016년 최대 16mm까지 관측되었으며, 기온의 경우 2015년 -12.9~35.7°C, 2016년 -18~31.9°C 범위로 기록 되어있다. 2016년 데이터의 경우 분석대상 기간이 6월까지 이므로 여름기간이 포함되지 않아 강수량 및 최고

1) 기상청 국가기후데이터센터 통계자료 (<http://sts.kma.go.kr>)

온도가 모두 낮게 관측되었으나, 학습데이터인 2015년 데이터의 기간이 더 넓기 때문에 버스 통행시간 예측 값을 생성하는데 충분한 샘플을 확보할 수 있다.



Figure 3. Case study: metropolitan express bus (M7412)

2. 중장기 버스 통행시간 예측 분석결과

BIS를 통해 도출한 버스통행 데이터를 클러스터링하기 위해 자기조직화지도의 크기를 연구방법론의 Equation 1을 통해 79x18 사이즈의 1,422개 출력뉴런으로 결정하였다. 자기조직화지도 알고리즘 적용시 초기값은 학습샘플들의 가장 큰 고유벡터를 따라 각 뉴런에 배정하는 선형초기화법을 이용하였으며, 학습률은 0.5로 시작해서 0.05까지 점차 감소시켰으며, 반복횟수는 25로 설정하였다. 본 연구에서 적용한 알고리즘은 2단계 클러스터링으로 SOM의 경우 학습데이터를 초기화하는 역할을 하므로 수렴조건은 최대반복횟수 만족시까지로 설정하여 대량을 학습데이터를 초기화시키도록 했으며, 유의미한 통행시간 산출을 위해서는 출력 뉴런의 수를 한번 더 클러스터링하여 비슷한 단위를 그룹화시킬 필요가 있다(Vesanto et al., 2000). 이에 본 연구에서는 K-means 클러스터링을 이용하여 K값을 증가시키며 WSS를 계산하여 K가 100 근처에서 WSS그래프가 완만해지는 것을 확인하였다. 이에 최종 클러스터 수를 100으로 정하였다. 각 클러스터별 대표값을 클러스터 내에 속한 통행시간들의 median으로 설정하여 통행시간 결과를 얻었다. 각각의 클러스터는 입력값으로 사용한 버스 출발시각, 기상상황, 요일에 대해 비슷한 통행시간 특성을 나타내는 항목들을 군집화한 결과이다. 따라서 각 요일별 시간별 기상상황별로 구분하여 중장기 버스통행시간 예측을 위한 참조테이블을 구성할 수 있다.

본 연구에서 학습에 이용한 데이터는 총 9개월간의 버스 통행자료로 버스통행시간에 영향을 미치는 요인들을 출발시각, 요일, 그리고 기상요인 등으로 분류하고, 자기조직화지도를 활용하여 맑음과 우천시에 대한 요일/출발시각별 버스통행시간 참조 테이블을 구성하였다. 따라서 버스 경로 1개당 요일 및 공휴일(8개), 출발시각(19개), 기상요인(2개)로 총 304개의 경우의 수에 대하여 각 경우를 대표하는 클러스터를 배정하였다. 클러스터 배정 기준은 각 경우의 수에 속하는 버스통행 별로 도출된 클러스터들을 비교하여 가장 많이 속하는 클러스터를 대표 클러스터로 배정하였다. 도출된 클러스터 결과 중 월요일과 공휴일 결과를 정리하였다(Table 2, 3 참조).

Table 2. Number of samples by clusters on Monday

(unit: number of samples)

Cluster	Departure time (hour)																			
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
2																				175
3		142																		
23						197														
39	102							171	152	157		179	160	151	167		144			4
59											165									
68						171														
71			168																161	
73				168																
96																				164

Table 3. Number of samples by clusters on holiday

(unit: number of samples)

Cluster	Departure time (hour)																			
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
10		32																		
22								36	38	34		41	34	36	37					3
34						40														
41																				37
51										41										
63			42																	
65				41																
75																			41	
78	27																			
89																39				39
94					43															

대부분의 경우에 대해 출발시각별로 고유의 클러스터에 모든 버스 통행이 속함을 알 수 있다. 이를 이용하여 최종적으로 버스통행시간 참조 테이블을 Table 4, 5와 같이 구성하였다. 통행시간 분석결과 월요일 오전첨두시와 금요일 오후첨두시에 버스 통행시간이 날씨와 관계없이 높았으며, 평일 오전 오후 첨두시에도 다른 시간대에 비해 버스 통행시간이 높았다. 이는 일반적인 도시부 교통패턴과 일치하는 결과로 유의미한 통행시간 분석결과로 판단된다. 또한 우천시 맑은 날에 비해 버스 통행시간이 높은 시간대가 존재함을 확인할 수 있었으며, 토요일 오전의 경우는 우천시 버스 통행시간이 줄어드는데 주말에 우천시 사람들이 외출을 자제하기 때문으로 판단된다. 우천시 참조테이블의 경우 통행시간이 산출되지 않은 시간대가 존재하는데, 이는 23시에 출발하는 버스 샘플수가 적어 통행시간 도출이 불가능하였다. 산출된 참조 테이블을 이용하여 각 요일별, 출발시각, 기상상황에 따른 버스 통행시간 대표값을 결정할 수 있으며, 실제 버스통행시간 정보 제공을 위해 해당 통행시간을 단시간에 계산할 수 있다. 도출된 버스 통행시간 예측 참조표를 통해 각 요일 및 시간, 그리고 강우 여부에 따라 중장기적 관점에서의 예측 통행시간을 확인할 수 있다. 또한 도출된 결과를 이용하여 본 연구에 이용한 M7412버스의 통행 특성도 분석이 가능하다. 강우 여부와 관계없이 많은 시간대의 예측 통행시간이 비슷한 결과를 나타내는데, 본 연구에서는 실제 버스의 과거 통행 이력을 기반으로 구성된 클러스터를 이용하여 통행시간을 산출하였기 때문에 해당 버스가 실제 통행에서도 특정 시간대에는 강우 여부에 크게 영향을 받지 않는다는 것을 확인할 수 있다. 현재 클러스터링에 이용한 데이터

Table 4. Bus travel time reference table on clear day

(unit: min.)

	Departure time (hour)																		
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Sun.	48	41	44	42	44	48	48	48	48	49	48	48	48	48	46	48	44	44	48
Mon.	51	96	77	60	60	60	53	52	56	52	52	61	77	77	56	52	52	60	77
Tue.	53	65	78	61	61	61	58	62	58	61	58	53	62	62	62	58	61	61	58
Wed.	42	62	81	74	71	69	60	52	58	61	70	66	70	76	58	49	49	47	66
Thu.	42	62	80	65	65	65	64	55	59	56	67	68	79	79	61	65	65	56	65
Fri.	59	58	75	68	70	63	74	53	53	64	73	59	74	92	74	63	52	53	68
Sat.	47	59	52	52	56	65	70	68	65	67	68	59	73	65	52	47	56	47	65
Holiday	40	40	41	41	42	46	51	51	51	51	51	51	51	51	46	45	48	45	51

Table 5. Bus travel time reference table on rainy day

(unit: min.)

	Departure time (hour)																		
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Sun.	48	41	44	42	44	48	48	48	48	49	48	48	48	48	46	48	44	44	
Mon.	51	96	77	60	60	60	53	52	56	52	52	61	77	77	56	52	52	60	77
Tue.	53	65	78	61	61	61	58	58	62	58	61	58	53	78	62	58	61	61	
Wed.	42	62	81	74	71	69	60	52	58	61	70	71	70	76	58	49	49	47	
Thu.	42	62	80	65	65	65	64	55	59	56	67	68	79	79	61	65	65	56	
Fri.	59	58	75	68	70	63	74	53	53	64	73	59	74	92	74	63	52	53	68
Sat.	47	59	44	52	56	65	70	68	65	67	68	59	73	65	52	47	56	47	56
Holiday			41			46			51			51							

는 총 9개월간의 자료이므로 강우의 경우에 맑은 날과 차이가 큰 샘플들을 많이 확보하기에는 한계가 존재하므로 추후 연구를 통해 분석 데이터의 양을 늘린다면 강우 여부에 따른 차이도 더 명확하게 드러날 것으로 예상된다. 이에 본 연구에서는 분석결과의 정확도 검증을 위해서 2016년 1-6월(6개월)의 테스트 자료를 이용하여 정확도 검증을 수행하였으며, 현재 중장기 버스 통행시간 예측에 활용되고 있는 평균운행속도 기반의 예측 결과와 비교하였다. 평균 운행속도 기반 예측 방법론은 버스의 시·종점간의 거리와 과거 평균운행속도를 이용하여 버스의 통행시간을 예측하는 방법론으로 현재 주요 서비스에 활용되고 있다. 예측오차는 Equation 3의 식으로 계산되는 MAPE(Mean Absolute Percentage Error)를 사용하였다.

$$MAPE(\%) = \sum_{i=1}^n \left| \frac{(TT_i - \widehat{TT})}{TT_i} \right| \times \frac{100}{n}, \tag{3}$$

여기서, TT_i : i 번째 샘플의 통행시간

\widehat{TT} : 예측 통행시간

n : 전체 샘플 수

테스트 자료를 활용한 예측 통행시간 MAPE 분석결과 맑음 13.76%, 영상우천 18.20%의 오차율을 보였으며, Table 6과 같다. 기존 평균운행속도 기반 방법론에 비해 오차율이 각각 약 8%, 6% 개선되었다. 기존 알고리즘은 기

상상황 구분 없이 버스 통행시간 예측값을 제공하는데 반해 본 연구의 SOM 알고리즘은 기상상황을 구분하여 보다 정확한 버스통행시간 예측값을 제공할 수 있다. 본 연구 방법론을 활용하여 버스 노선별로 클러스터를 구축하여 참조테이블을 구성하였을 시 현재 서비스 중인 버스통행시간 보다 정확한 예측 정보를 제공하여 대중교통 운영자 및 이용자 모두에게 효율성 증대와 정시성에 기여할 수 있을 것으로 판단된다. 이는 버스 이용자가 예측 통행시간을 기반으로 본인의 통행과 관련된 의사결정(출발시간 계획, 교통수단 선택 등)을 하는데 활용되는 것으로 기존의 링크 기반 버스정보서비스의 한계를 보완할 수 있다. 또한 실시간으로 입력되는 버스통행 정보를 주기적으로 학습하여 SOM 알고리즘에 반영한다면, 지속적으로 업데이트된 통행시간 참조표를 활용할 수 있을 것으로 판단된다.

Table 6. Prediction accuracy of SOM algorithm for bus travel time

MAPE(%)	SOM	Bus average speed
Clear	13.76	22.43
Rainy	18.20	24.58

결론

1. 연구의 의의 및 결론

본 연구는 비지도학습 기반의 SOM 알고리즘을 활용하여 상습 지·정체 구간에 대한 버스 통행시간 예측을 실시하였다. 기존의 버스예측 정보는 단기적 관점의 링크소통정보를 제공하는데 중점을 두고 있어 실제 통행속도와의 차이를 반영하기 어려운 한계가 존재하므로 중장기 통행시간 예측을 통한 이용자들의 의사결정을 지원할 수 있어야 한다. 이에 본 연구에서는 기존 버스 통행시간 예측 연구들에 대한 사례 검토를 수행하여 본 연구에 적용되는 SOM 방법론에 대한 유용성을 검토하였고, 각 단계별 세부절차 및 변수결정, 입력자료 구축, 그리고 클러스터 과정 등을 제시하였다. 특히 예측의 정확도 개선을 위해서 통행시간에 영향을 주는 요소인 요일(주중, 주말, 공휴일), 기상요인(강수량 및 기온) 등을 입력변수로 고려하여 분석을 수행하였다. 분석에 사용된 BIS 자료는 경기 북부외곽에서 서울 도심까지 운행하는 광역버스로 출·퇴근 시간의 상습 지·정체를 포함하는 구간으로 교통상태 변화가 비교적 명확한 구간이다. SOM 학습에 사용된 이력자료는 2015년 1-9월(9개월), 테스트 자료는 2016년 1-6월(6개월)을 사용하였다. 기상자료의 경우 기상청 국가기후데이터센터의 시간별 자료를 활용하였다. 요일, 기상, 시각별로 클러스터 분석을 통해 버스통행시간 참조표를 생성하고 이를 활용하여 예측을 수행하였으며, 정확도 검증을 통하여 유효성을 확인하였다. 이는 기존의 단기적 링크기반의 단순 평균을 통한 버스통행시간 예측값의 부정확한 한계를 극복할 수 있는 중장기적 통행시간 예측정보를 제공함으로써 버스이용자의 만족도를 향상시킬 것으로 판단된다. 또한 실시간으로 입력되는 자료를 주기적으로 학습하여 SOM 알고리즘에 반영한다면 지속적인 예측 정확도 개선과 함께 탄력적인 대중교통 정책 수립이 가능할 것으로 판단된다. 이용자 측면에서는 대중교통 신뢰도 향상 및 사용자 맞춤형 정보 제공을 통하여 대기시간 최소화 및 효율적인 의사결정이 가능할 것으로 판단된다.

2. 향후 연구과제

비지도학습의 SOM 알고리즘을 활용한 버스 통행시간 예측은 정확도가 우수하나 전체 버스노선에 적용 시 각 버스 노선별로 참조테이블 구성이 필요하다. 또한 본 연구에는 버스 통행시간에 영향을 미치는 변수로 기상요인을 추가하였지만, 유고상황에 대한 대응전략 및 범용성 확보를 위해 시내 통행노선 및 버스 전용차로 이외의 구간을 이용하는 노선 등에 대한 추가적인 검토가 필요할 것으로 판단된다.

ACKNOWLEDGEMENT

This research was supported by a grant (16CTAP-C097506-02) from Technology Advancement Research Program funded debt by Ministry of Land, Infrastructure and Transport of Korean government.

REFERENCES

- Abdelfattah A. M., Khan A. M. (1998), Models for Predicting Bus Delays, *Transp. Res. Rec.*, 1623, 8-15.
- Bai C., Peng Z. R., Lu Q. C., Sun J. (2015), Dynamic Bus Travel Time Prediction Models on Road With Multiple Bus Routes, *Comput. Intell. Neurosci.*, 2015, 63.
- Chen M., Liu X., Xia J., Chien S. I. (2004), A Dynamic Bus-arrival Time Prediction Model Based on APC Data, *Comput.-Aided Civ. Infrastruct. Eng.*, 19(5), 364-376.
- Chien S., Ding Y., Wei C. (2002), Dynamic Bus Arrival Time Prediction With Artificial Neural Networks, *J. Transp. Eng.*, 128(5), 429-438.
- El Faouzi N. E., Billot R., Bouzebda S. (2010), Motorway Travel Time Prediction Based on Toll Data and Weather Effect Integration, *IET Intell. Transp. Syst.*, 4(4), 338-345.
- Foote P.J., Stuart D.G. (1998), Customer Satisfaction Contrasts Express Versus Local Bus Service in Chicago's North Corridor, *Transp. Res. Rec.*, 1618, 143-152.
- García H. L., González I. M. (2004), Self-organizing Map and Clustering for Wastewater Treatment Monitoring, *Eng. Appl. Artif. Intell.*, 17(3), 215-225.
- Jeong, R., Rilett, L. R. (2004), Bus Arrival Time Prediction Using Artificial Neural Network Model, *Proc. 7th Int. IEEE Conf. Intell. Transp. Syst.*, 988-993.
- Kim H., Kim S., Park S., Jang K. (2013), Assessment of Travel Time Estimates Based on Different Vehicle Speed Data: Spot Speed vs. Sampled Journey Speed in South Korean Expressways, *Proc. 10th Int. Conf. East. Asia Soc. Transp. Stud.*, Taipei, Taiwan.
- Kwon J., Coifman B., Bickel P. (2000), Day-to-day Travel-time Trends and Travel-time Prediction From Loop-detector Data, *Transportation Research Record, Transp. Res. Rec.*, 1717, 120-129.
- Lin J., Wang P., Barnum D. (2008), A Quality Control Framework for Bus Schedule Reliability, *Transp. Res. Part E*, 44, 1086-1098.
- Ma C. Q., Wang Y. P., Chen K. M. (2007), Competition Model Between Urban Rail and Bus Transit, *J. Transp. Syst. Eng. Inf. Technol.*, 7(3), 140-143.
- Nookala L. S. (2006), Weather Impact on Traffic Conditions and Travel Time Prediction (Doctoral dissertation, University of Minnesota Duluth).
- Patnaik J., Chien S., Bladikas A. (2004), Estimation of Bus Arrival Times Using APC Data, *J. Public Transp.*, 7(1), 1-20.
- Qiao W., Haghani A., Hamed M. (2012), Short-term Travel Time Prediction Considering the Effects of Weather, *Transportation Research Record, Transp. Res. Rec.*, 2308, 61-72.
- Ramakrishna Y., Ramakrishna P., Lakshmanan V., Sivanandan R. (2006), Bus Travel Time Prediction Using GPS Data, *Proc. Map India*.

- Smith B.L., Demetsky M.J. (1995), Short-term Flow Prediction: Neural Network Approach, *Transp. Res. Rec.*, 1453, 98-104.
- Vanajakshi L., Rilett L. (2007), Support Vector Machine Technique for the Short Term Prediction of Travel Time, *Proc. IEEE Intell. Veh. Symp.*, 600-605.
- Vesanto J., Alhoniemi E. (2000), Clustering of the Self-organizing Map, *IEEE Trans. Neural Networks*, 11(3), 586-600.
- Yang J.-S. (2005), Travel Time Prediction Using the GPS Test Vehicle and Kalman Filtering Techniques, *Proc. Am. Control Conf.*, 2128-2133.
- Yu B., Lam W. H. K., Tam M. L. (2011), Bus Arrival Time Prediction at Bus Stop With Multiple Routes, *Transp. Res. Part E*, 19(6), 1157-1170.
- Zhang X., Rice J. A. (2003), Short-term Travel Time Prediction, *Transp. Res. Part C*, 11(3), 187-210.