

앙상블 경험적 모드 분해법을 이용한 도시부 단기 통행속도 예측

김의진* · 김동규**

Kim, Eui-Jin* , Kim, Dong-Kyu**

Short-term Prediction of Travel Speed in Urban Areas Using an Ensemble Empirical Mode Decomposition

ABSTRACT

Short-term prediction of travel speed has been widely studied using data-driven non-parametric techniques. There is, however, a lack of research on the prediction aimed at urban areas due to their complex dynamics stemming from traffic signals and intersections. The purpose of this study is to develop a hybrid approach combining ensemble empirical mode decomposition (EEMD) and artificial neural network (ANN) for predicting urban travel speed. The EEMD decomposes the time-series data of travel speed into intrinsic mode functions (IMFs) and residue. The decomposed IMFs represent local characteristics of time-scale components and they are predicted using an ANN, respectively. The IMFs can be predicted more accurately than their original travel speed since they mitigate the complexity of the original data such as non-linearity, non-stationarity, and oscillation. The predicted IMFs are summed up to represent the predicted travel speed. To evaluate the proposed method, the travel speed data from the dedicated short range communication (DSRC) in Daegu City are used. Performance evaluations are conducted targeting on the links that are particularly hard to predict. The results show the developed model has the mean absolute error rate of 10.41% in the normal condition and 25.35% in the break down for the 15-min-ahead prediction, respectively, and it outperforms the simple ANN model. The developed model contributes to the provision of the reliable traffic information in urban transportation management systems.

Key words : Short-term prediction of travel speed, Ensemble empirical mode decomposition (EEMD), Artificial neural network (ANN), Intrinsic mode function (IMF)

초 록

단기 통행속도 예측을 위해 데이터 기반 비모수적 기법들을 활용한 다양한 연구들이 수행되고 있다. 그럼에도 교통 신호 및 교차로로 인한 복잡한 동적 특성을 가지는 도시부의 예측 연구는 상대적으로 부족한 실정이다. 본 연구는 도시부 통행 속도를 예측하기 위해 앙상블 경험적 모드 분해법(EEMD)과 인공신경망(ANN)을 이용한 하이브리드 접근법을 제안하는 것을 목적으로 한다. EEMD는 통행속도의 시계열 자료를 고유모드 함수(IMF)와 오차항으로 분해한다. 분해된 IMF는 시간단위의 국지적 특성을 반영하며, ANN을 통해 개별적으로 예측된다. IMF는 원본데이터가 가진 비선형성, 비정상성, 진동 등의 복잡성을 완화하기 때문에, 원래의 통행속도에 비하여 더 정확하게 예측될 수 있다. 예측된 IMF들은 합산되어 예측 통행속도를 표현한다. 본 연구에서 제시된 방법을 검증하기 위하여 대구시의 DSRC로부터 구축된 통행속도 데이터가 활용된다. 성능평가는 도시부 링크 중 특히 예측이 어려운 지점에 대해 수행되었으며, 분석 결과 제시된 모형은 15분 후 예측에 대해 각각 정상시 10.41%, 와해상대시 25.35%의 오차율을 가지며, 단순 ANN 기법에 비하여 우수한 성능을 보이는 것으로 확인된다. 본 연구에서 개발된 모형은 도시교통관리체계의 신뢰성 있는 교통정보를 제공하는 데에 기여할 수 있을 것으로 기대된다.

검색어 : 단기 통행속도 예측, 앙상블 경험적 모드 분해(EEMD), 인공신경망(ANN), 고유모드함수(IMF)

* 서울대학교 공과대학 건설환경공학부 박사과정 (Seoul National University · kyjcw@snu.ac.kr)

** 종신회원 · 교신저자 · 서울대학교 공과대학 건설환경공학부 조교수, 서울대학교 건설환경종합연구소 겸임연구원
(Corresponding Author · Seoul National University · dongkyukim@snu.ac.kr)

Received April 7, 2018/ revised May 8, 2018/ accepted June 2, 2018

1. 서론

2015년 전국 7대도시의 교통혼잡비용은 약 21조원으로 추산되며, 2011년 이후 16% 가량 꾸준히 증가하는 추세이다. 우리나라에서는 지난 20여 년 동안 지능형교통체계(Intelligent Transport System, 이하 ITS)를 전국적으로 구축하였으며, ITS의 중요 하위 시스템인 첨단여행자정보시스템(Advanced Traveler Information System, 이하 ATIS)과 첨단교통관리시스템(Advanced Traffic Management System, 이하 ATMS)을 통해 이용자에게는 실시간으로 예측된 교통정보를 제공하여 최적의 경로 및 수단을 선택하게 하고, 운영자에게는 혼잡 완화를 위한 교통운영전략을 활용할 수 있게 한다. ITS 초기에는 루프검지기, 영상검지기와 같은 지점검지기를 통해 교통정보를 수집하였으나, 검지기 사이의 교통류 정보가 반영되지 않아 정보가 왜곡될 수 있어 최근 단거리 전용 통신(Dedicated Short Range Communication, 이하 DSRC) 등을 활용한 구간 검지체계가 구축되고 있다. 특히 우리나라 대구광역시에는 ATMS 도입을 위해 총 612개의 링크에서 통행시간을 수집하여 신뢰성 있는 교통정보를 제공하고 있다. 하지만 도시부 통행속도는 비정상성(non-stationarity), 비선형성(non-linearity), 진동(oscillation)과 같은 교통류의 동적특성에 신호주기, 교차로 상호작용과 같은 영향요인으로 인해 고속도로보다 더 복잡한 특성을 가지며, 특히 불안정한 교통류 상태변화가 빈번하게 발생한다. 이러한 구간은 ATMS를 통한 혼잡관리가 필요하지만 기존 예측 모형들은 이를 제대로 예측하지 못하거나, 일정 시간 지연된 정보를 예측하여 실용적으로 활용되지 못한다.

경험적 모드 분해(Empirical Mode Decomposition, 이하 EMD)는 신호처리 분야에서 개발된 데이터 적응형(data-adaptive) 기술로 국지적 주기특성을 갖는 고유모드함수(Intrinsic Mode Function, 이하 IMF)로 시계열 자료를 분해하여 복잡한 동적 특성을 각 주기에 따라 개별적으로 나눠 분석할 수 있다. 양상불 경험적 모드분해법(Ensemble Empirical Mode Decomposition, 이하 EEMD)은 EMD에서 분해된 각 IMF에서 서로 다른 주기특성이 중첩되는 모드 믹싱(mode-mixing) 문제를 해결하여, 고유모드함수의 물리적 의미를 더 명확하게 나타내는 개선된 기술이다(Wu and Huang, 2009). EEMD를 통해 분해된 IMF에는 원 자료에 중첩된 정보가 단순한 형태로 분해되기 때문에 분해된 IMF를 개별적으로 예측하고 다시 복원하는 “divide and conquer” 개념의 하이브리드 접근법이 에너지, 수공학, 경제 등 다양한 분야의 시계열자료 예측에 활용되고 있으며, 교통 분야에서도 고속도로 지점속도, 철도승차인원 등의 예측에 활용되었다(Wang et al., 2014; Jiang et al., 2014).

본 연구에서는 EEMD를 이용해 도시부 통행속도 자료를 국지적 주기특성을 갖는 고유모드함수로 분해한 후, 각 고유모드 함수를

인공신경망(Artificial neural network, 이하 ANN)을 통해 개별적으로 예측하고, 다시 재결합하여 통행속도를 예측하는 하이브리드 접근법을 제안한다. 제안된 방법론은 대구광역시 계산오거리-반월당네거리 530m 구간의 90일 DSRC 자료를 이용하여, 일반구간과 교통류 와해 구간에서 그 성능을 검증한다. 본 논문의 2장에서는 단기 교통정보 예측을 위한 다양한 연구와 관련된 기존문헌을 고찰하고, 3장에서는 EEMD, ANN, 하이브리드 접근법의 구조에 대해 설명한다. 이어지는 4장에서는 대구 DSRC 자료를 활용하여 제안된 방법론을 검증하고, 5장에서는 본 연구의 성과와 한계 및 향후 연구방향을 논의한다.

2. 선행연구

최근의 단기 교통정보 예측 연구를 검토해보면 전통적인 모수적 방법론인 autoregressive integrated moving average (이하 ARIMA)나 교통류의 변화를 반영하여 예측을 수행하는 시뮬레이션 모형보다는 데이터 기반 비모수적 방법론이 더 높은 예측력을 갖는 것으로 알려져 있다(Vlahogianni et al., 2014). 따라서 본 연구에서는 비모수적 예측모형과 이를 개선하기 위해 제안된 하이브리드 접근법을 위주로 기존연구를 고찰하였다.

Myung et al. (2012)은 고속도로를 대상으로 차량검지기자료와 요금징수시스템에서 수집된 자료를 통합하여 K 최대근접이웃(K-Nearest Neighbor, 이하 KNN) 기법을 활용한 통행시간 예측 모형을 제안하였다. 단순평균법, 칼만필터링, 단순 패턴매칭기법 등 6개의 방법론과 비교하여 제안된 모형의 우수성을 검토하였다. Kim and Jang (2013)은 5분단위로 집계된 고속도로 24시간의 DSRC 자료를 비선형관계에 적합한 지도학습 방법인 ANN을 이용해 학습하고, 예측하였다. 실측통행시간과 비교한 평균 오차율이 3.95%로 낮게 나타났지만 교통상태 변화가 크지 않은 고속도로 구간의 하루 통행속도로만 검증을 수행하였다. Shin et al. (2014) 역시 KNN을 이용해 이력자료에 내재된 교통류의 시간적 진화행태를 검색하여 통행시간을 예측하는 모형을 제안하였다.

위와 같은 예측모형 적용을 통한 정확도 향상 이외에 교통류 예측에 영향을 주는 다양한 요인에 대한 연구도 진행됐다. Oh and Park (2011)은 고속도로 위 두 개의 연속된 검지기에서 추정된 2~10분 집계 통행속도를 시계열자료 학습을 위한 ANN인 Recurrent Neural Network를 통해 예측하였으며, 엔트로피 개념을 통해 시간적으로 변화하는 통행속도 패턴의 복잡도가 모형의 정확도에 영향을 주는 것을 보여주었다. Asif et al. (2014)는 지지벡터머신(Support Vector Machine, 이하 SVM)을 통해서 도시부 네트워크 5분 집계 통행속도를 예측하였고, 예측정확도를 바탕으로 군집분석을 수행한 결과 링크별로 명확한 군집이 존재함을 보여주

었다. Fusco et al. (2016)은 대규모 도시부 네트워크의 5분 집계 통행속도 예측에 여러 링크의 정보를 활용하는 ANN과 단일링크의 정보만을 활용하는 Seasonal ARIMA를 적용하였으며, 전반적인 정확도는 ANN이 우수하지만 예측이 어려운 비반복적 교통류에서는 단일링크의 정보를 활용하는 Seasonal ARIMA가 더 우수함을 보였다. Chang and Yoon (2018)은 KNN을 이용해 15분단위로 집계된 고속도로 톨게이트의 교통량 자료를 예측하였는데, 다중시간대에 대해서도 효과적인 예측이 가능한 반면, 과거 패턴이 부족한 경우 예측의 방향성을 상실해 큰 오차를 갖는 구간이 존재하는 단점이 있었다. 또한 요일별로 모형의 정확도와 최적 파라미터가 달라지는 것을 통해 같은 구간이라도 패턴의 다양성이나 복잡성이 다르게 존재함을 보였다. 단순 ANN이나 KNN을 활용한 데이터 기반 연구들은 기존 시계열 모형보다는 높은 성능을 보여주나, 교통상태가 불안정하거나 과거패턴이 부족한 경우 정확도에 한계가 있었다. 따라서 다양한 교통류 상태에 활용할 수 있는 강건한 예측모형이 필요하며, 예측이 어려운 비반복적 교통류에서는 단일 링크의 정보를 활용하는 것이 유리할 수 있다.

위에서 제시한 기존 데이터 기반 방법론을 보완하기 위해 원 자료에 대한 전처리를 통해 정확도를 향상시키는 다양한 하이브리드 접근법이 교통 정보의 단기예측에도 활용되고 있다. 이들은 데이터 적용형 방법론인 EMD, EEMD와 비모수적 예측기법인 ANN, KNN, SVM과의 융합을 통해 고속도로 지점속도, 통행속도, 철도 승객 수 등의 단기 교통정보 예측에서 큰 성능향상을 보였다. Hamad et al. (2009)은 고속도로 5분 집계 지점속도 자료를 EMD를 통해 분해한 후 노이즈일 가능성이 있는 고주파 IMF를 제외한 저주파 IMF만을 이용해 ANN을 학습시키고 예측을 수행하였다. Wei and Chen (2012)은 15분 집계의 고속철도 승객 수 자료를 EMD를 통해 분해한 후 IMF와 원 자료의 피어슨 상관계수, IMF의 퍼센트 에너지 등이 높은 유의미한 IMF를 선택하고, 선택된 IMF만을 이용해 역시 ANN을 학습시키고 예측을 수행하였다. 위에 제시된 두 연구 다 기존 ARIMA나 ANN에 비해서 성능이 향상되었지만, 유의미한 IMF를 선택하는 방법이 연구자의 주관적 판단이나 데이터의 형태에 따라 달라지기 때문에 활용에 한계가 있었다. 이러한 IMF의 선택 문제를 해결하기 위해 각 IMF에 대해 개별적인 예측을 수행한 후 예측된 IMF의 결과를 재결합하여 원 자료를 예측하는 “divide and conquer”개념의 방법론이 제안되었다. Wang et al. (2014)은 EMD와 ARIMA를 융합해 고속도로 5분 집계 자료를 예측하였으며, EMD로 분해한 개별 IMF를 각각 예측하고 재결합하여 기존 ARIMA, ANN보다 뛰어난 성능을 보였다. Jiang et al. (2014)은 EEMD와 SVM을 활용해 고속철도 일일 집계 고속철도 승객량 자료를 예측하였으며, 특히 축제나 명절과 같은 비반복적인 영향요인에 대해서도 강건한 특성을 보였다.

단기 교통정보 예측을 위한 비모수적 방법론과 하이브리드 접근법에 대한 기존 연구 검토 결과 비모수적 방법론은 자료의 비선형성을 반영할 수 있지만 비반복적 패턴이나 불안정한 교통상황에 대한 정확도가 떨어지는 단점이 있다. 이를 보완하기 위해 제안된 하이브리드 접근법은 기존 비모수적 방법론의 단점을 보완하며, 그 가능성을 보여주었지만 아직 고속도로에 비해 자료의 복잡도가 높은 도시부 통행속도의 단기 예측에 활용된 사례가 없어 검증이 필요하다.

3. 연구방법론

3.1 Ensemble Empirical Mode Decomposition

EMD는 시계열 자료를 체거름(sifting)이라는 과정을 통해 국소적 주기특성을 갖는 IMF들로 분해한다. 체거름 과정에서 국소적 최대/최소값을 3차원 스플라인 보간(cubic spline interpolation)을 통해 상위/하위막(upper/lower envelop)인 $L_{11}(t)$ 와 $U_{11}(t)$ 로 만들고, 이 두 막의 평균인 $m_{11}(t)$ 를 원 자료 $x(t)$ 에서 빼서 $h_{11}(t)$ 를 얻는다(Eqs. (1) and (2)).

$$m_{11}(t) = [U_{11}(t) + L_{11}(t)]/2 \tag{1}$$

$$h_{11}(t) = x(t) - m_{11}(t) \tag{2}$$

$$h_{12}(t) = h_{11}(t) - m_{12}(t) \tag{3}$$

이 때 $h_{11}(t)$ 가 IMF의 2가지 조건을 만족하면 $h_{11}(t)$ 가 하나의 IMF로 추출된다.

- ① 국소적 최대/최소값의 개수와 원점을 지나는 점의 차이가 0 또는 1인 경우
- ② S변의 연속적 과정동안 같은 상태로 유지된 경우

만약 $h_{11}(t)$ 가 위 조건을 만족하지 않으면 $h_{11}(t)$ 의 상위/하위막의 평균인 $m_{12}(t)$ 와 $h_{12}(t)$ 를 다시 정의하고(Eq. (3)), 위와 같은 과정을 반복하여 두 조건을 만족하는 $h_{1j}(t)$ 를 찾은 후 이를 첫 번째 IMF $c_1(t)$ 로 추출한다. 원 자료 $x(t)$ 에서 첫 번째 IMF를 뺀 나머지 $r_1(t)$ 를 구한 후(Eq. (4)) 위 과정을 $r_1(t)$ 에 반복하여 두 번째 IMF를 추출한다.

$$r_1(t) = x(t) - c_1(t) \tag{4}$$

이러한 과정은 나머진 $r_i(t)$ 가 단조증가/감소함수가 되거나 하나의 국소적 최대/최소값을 가질 때까지 반복하여 여러 개의

IMF를 추출한다. 이 때 마지막 IMF가 추출되고 난 후의 나머지 (residue)인 $R(t)$ 는 자료의 경향성(trend)을 나타내게 된다. 체거름 과정을 마친 최종 결과 원 자료 $x(t)$ 는 n 개의 IMF와 나머지 $R(t)$ 의 합으로 정의된다(Eq. (5)).

$$x(t) = \sum_{i=1}^n c_i(t) + R(t) \quad (5)$$

EEMD는 원 자료에 특정진폭을 갖는 임의의 서로 다른 백색소음을 추가한 후 EMD를 여러 번 수행하고 이들의 양상불 평균을 통해 IMF를 구하는 방식으로 충분한 횟수로 EMD를 수행할 경우 양상불 평균 과정에서 모드믹싱 문제가 해결되고 고유의 주기특성을 갖는 IMF가 생성된다. EEMD에서는 백색소음의 진폭과 EMD 수행횟수가 파라미터가 되며, Wu and Huang (2009)은 일반적으로 자료의 표준편차의 0.2배를 적정 파라미터로 추천했다. 본 연구에서도 원 자료 표준편차의 0.2배를 백색소음의 진폭으로 하고, 300번의 수행횟수를 통해 EEMD를 수행하였다.

EEMD는 체거름이 끝날 때 까지 IMF를 분해하기 때문에 별도로 IMF의 개수를 설정할 필요는 없으나 자료에 대한 사전적 지식이 있다면 분해할 IMF의 개수를 설정할 수 있다. 하지만 복잡한 동적특성을 가진 도시부 교통류의 경우 링크나 요일마다 특성이 다르기 때문에 최적 IMF 개수를 찾는 과정을 생략하고, 링크마다 자동으로 IMF의 개수를 설정하였다.

3.2 Artificial Neural Network

EEMD를 통해 분해된 개별 IMF에 대한 예측은 대표적인 비모수 예측모델인 ANN을 활용하였다. ANN은 역전파 알고리즘(Back-propagation algorithm)을 통해 입력뉴런, 은닉뉴런, 출력뉴런 사이의 연결강도를 지도학습 기반으로 학습하여, 비선형 관계를 모델링할 수 있는 장점이 있다. 본 연구에서는 ANN의 가장 기본적인 형태 다층 퍼셉트론(multilayer perceptron)모형을 사용하였고, 파라미터인 은닉층의 숫자와 학습율은 격자 탐색법을 통해 구했으며, 이 때 탐색결과에 대한 평가는 학습자료에 대한 4분할 교차검증(4-fold cross-validation)을 통해 평균계급근편차가 가장 낮은 파라미터를 선택하였다. 입력뉴런은 시계열 예측의 이력자료 및 실시

간 교통량 자료로 구성되며, 본 연구에서는 예측목표시점 k 분을 기준으로 $k+5$ 분, $k+10$ 분, $k+15$ 분, $k+20$ 분, $k+25$ 분, 하루 전, 일주일 전, 총 7개의 자료를 입력뉴런으로 설정하였다. ANN의 학습과정은 Kim and Jang (2013)에 상세히 제시되어 있다.

3.3 EEMD-ANN 예측모형의 구조

EEMD-ANN 예측모형은 Fig. 1과 같이 EEMD를 통해 원 자료를 n 개의 IMF로 분해한 후 개별 IMF에 대한 단일시간예측(Single step prediction) 후 재결합 하는 구조로 구성된다. 모형의 정확도를 위해 매 시간마다 실시간 현행 교통량 자료와 이력 교통량 자료를 모두 활용하여 EEMD를 수행하고, 개별 IMF에 대한 ANN 모형의 예측을 수행하였다.

3.4 예측모형의 분석 및 평가

본 연구의 예측모형을 평가하기 위해 요일별, 교통류 상태별로 나눠 평가를 수행하였다. 복잡한 도시부 교통류의 경우 요일별 통행패턴 변화로 인한 영향이 커서 평일, 주말, 금요일 등에 따라 다양한 특성을 보이며, 고속도로에 비해 불안정한 교통류 상태를 갖기 때문에 침두-비침두 이외에 국지적인 교통류 왜해가 빈번히 발생한다. 위 두 가지 평가를 통해 기존 단일 예측모형으로 예측하기 어려웠던 다양한 교통류 패턴 및 교통류 왜해구간에 대한 모형의 강건성을 평가하여, 제안된 모형의 개선점과 실용적 가치를 보여줄 수 있다. 모형의 예측력을 평가하기 위한 평가지표는 변동이 큰 자료에 대한 직관적 평가가 가능한 평균절대값백분율오차(Mean Absolute Percentage Error, MAPE (%))를 이용하였다.

$$MAPE(\%) = \left(\frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right| \right) \times 100 \quad (6)$$

x_i : the actual value, \hat{x}_i : the predicted value,
 N : the number of observation

4. 적용 결과

4.1 자료 구축 및 예측방법론 코드

본 연구에서는 대구권 광역시 네트워크의 2016년 4월부터 6월까

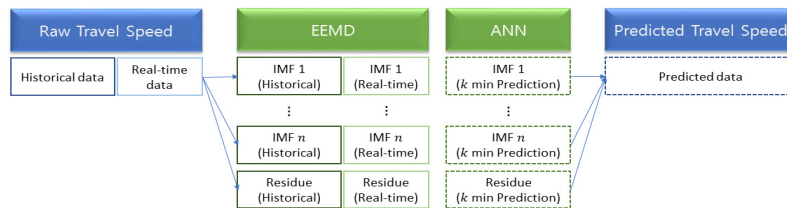


Fig. 1. Proposed EEMD-ANN Model

지의 90일간의 5분 집계 DSRC 통행속도 자료를 사용하였고, 주요 간선도로축인 대구 달구벌대로의 반고개네거리부터 수성네거리까지 총 6개 링크에 대해 분석을 수행하였다. 자료의 전처리 과정에서 이상치와 누락데이터를 보정하였으며, 자료의 이상치는 자료의 표준편차의 1.96배 범위(95% 신뢰구간)를 기준으로 제거하였고, 누락데이터의 경우 누락구간의 전후 3개의 자료(총 30분)를 이용한 단순이동평균법을 통해 보정하였다. 이상치제거와 누락데이터 보정을 거치지 않은 원 자료의 정보는 아래 Table 1에 나타났다. 자료의 전처리와 EEMD-ANN 예측모형의 전 과정은 R 프로그

래밍 언어로 코딩되었으며, EEMD는 libeemd패키지(Luukko et al., 2016), ANN은 nnet패키지를 기반으로 코딩되었다.

4.2 단순 ANN을 이용한 링크별 분석

통행속도 예측의 정확성은 각 링크의 교통류 특성에 따라 다르게 나타난다. 본 연구에서는 대상이 되는 6개 링크에 대해 83일의 자료로 ANN모형을 학습하고 7일 자료로 검증을 수행하여 기존 ANN 모델로 예측시 정확도가 가장 떨어지는 하나의 링크를 선정하여 본 연구에서 제안한 EEMD-ANN 방법론을 적용하였다. 5분

Table 1. Study Site and Data Descriptions, Dalgubul-Daero, Daegu, Korea

Daegu DSRC Data 2016.04.01. - 2016.06.29.				
Link Number	Distance	Quality (%)	Mean Speed (km/h)	Standard Deviation (km/h)
125	970	93.62	30.56	10.44
442	513	96.09	22.21	4.71
546	530	90.72	28.70	5.31
458	595	92.79	43.61	13.48
422	806	98.03	27.20	7.38
540	1,072	93.20	29.21	8.30



Table 2. Preliminary Analysis Using ANN

Link Number	10 min Prediction (MAPE,%)	15 min Prediction (MAPE,%)	20 min Prediction (MAPE,%)
125	5.27	6.68	7.79
442	13.80	17.24	20.08
546	13.79	17.88	21.27
458	6.83	9.96	11.07
422	5.26	6.98	8.88
540	6.24	8.11	9.89

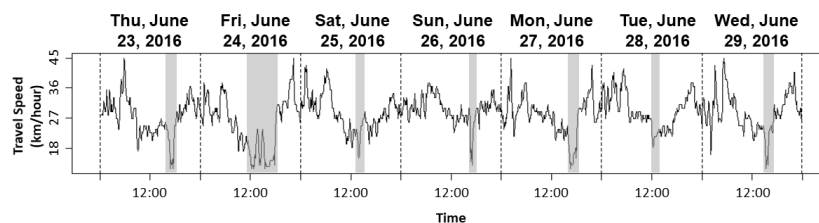


Fig. 2. Time-series of Travel Speed (Link Number 546, June 23-29, 2016)

집계 통행속도 자료를 활용해 10분, 15분, 20분, 총 3개의 주기에 대한 각 링크의 예측결과는 아래 Table 2와 같다. ANN을 통해 사전분석이 수행된 6개의 링크 중 가장 낮은 예측정확도를 보인 구간은 546번 링크인 계산오거리-반월당네거리 구간이었으며, Fig. 2는 제안한 모델의 검증을 수행하는 2016년 6월 23일에서 6월 29일까지의 통행속도 분포이다. 편도 4차선의 간선도로 역할을 수행하지만 530m 링크구간 내에 3개의 신호와 4개의 진출입로가 존재해 첨두-비첨두에 관계없이 하루 종일 불안정한 통행속도 분포를 보인다.

4.3 EEMD-ANN과 ANN의 비교

ANN을 이용한 사전분석 결과 진출입로, 신호 등으로 인해 평시에도 불안정한 교통류 상태를 보여 가장 예측이 어려웠던 DSRC 546번 링크의 1주일 자료에 대해 ANN과 EEMD-ANN 각각을 통해 15분 후 통행속도를 예측하였다. 예측의 정확도는

요일별 정확도와 교통류와해 구간의 정확도 두 가지에 대해 MAPE를 평가지표로 비교하였다. 교통류 와해 구간은 다양한 지표로 정의될 수 있지만 본 연구에서는 Fig. 2의 회색 영역과 같이 하루 중 가장 낮은 통행속도를 갖는 지점의 과거 속도감소 시작지점부터 미래 속도증가 완료지점까지의 구간을 교통류 와해 구간으로 정의하고, 7일 동안 일평균 154분의 와해 구간에 대한 예측을 수행하였다. Table 3은 요일별 교통류 상태별 ANN 모델과 EEMD-ANN모델의 예측력 차이를 MAPE를 지표로 나타낸 것이다. 평시구간에서 EEMD-ANN의 MAPE는 ANN의 MAPE의 46~69% 수준, 교통류 와해 구간에서는 39~99%정도로 수준으로 나타났다. 와해구간 중 특히 낮은 정확도를 보여준 금요일, 일요일은 반복적인 혼잡형태가 아닌 연속되거나 급격한 교통류 와해가 발생하였다. 특히 금요일의 경우 EEMD-ANN을 통해서도 큰 성능향상을 보이지 못했는데, 이는 일반적인 와해상태와 다른 장시간 연속된 지체가 발생해 해당 특성을 제대로 분해하지 못하였기 때문이다. 해당 교통류

Table 3. Comparison of ANN and EEMD-ANN for Travel Speed Prediction

DSRC Link 546, 15min ahead Prediction, 2016.06.23. - 2016.06.29.						
Day of Week	Normal (MAPE, %)		Break-down (MAPE, %)		Total (MAPE, %)	
	ANN	EEMD-ANN	ANN	EEMD-ANN	ANN	EEMD-ANN
Mon	12.22	8.43	58.47	36.29	17.4	11.55
Tue	9.85	6.12	15.19	10.06	10.39	6.52
Wed	12.78	6.83	40.96	32.72	15.83	9.64
Thu	13.15	5.99	28.55	15.86	15.26	7.34
Fri	13.21	9.51	89.64	89.10	36.35	33.61
Sat	11.85	6.5	33.41	12.94	13.8	7.08
Sun	9.21	4.96	99.09	39.51	16.1	7.61
Total	14.21	10.41	51.78	25.35	17.88	11.91

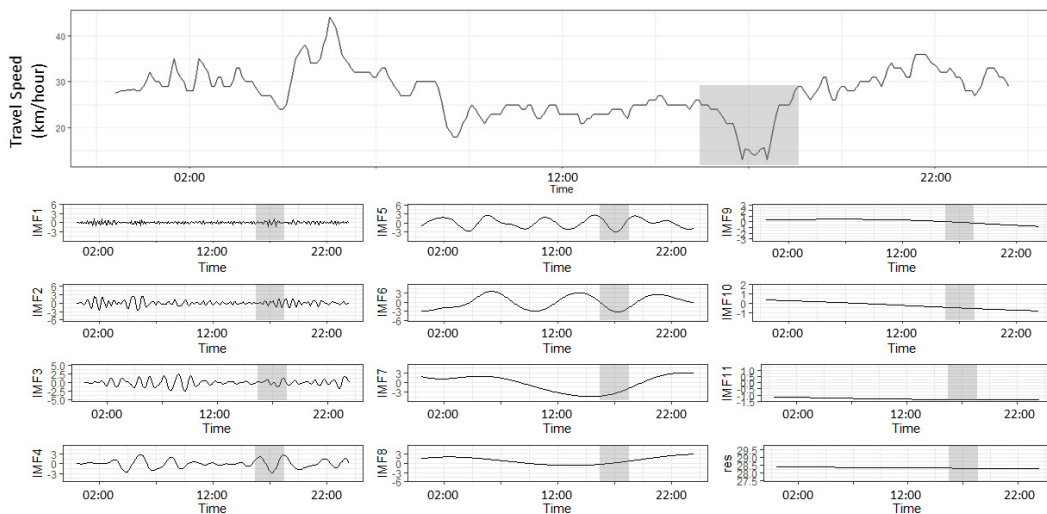


Fig. 3. Travel Speed Data and IMFs Computed using EEMD from Travel Speed Data

상황은 주로 교통사고 시 발생하며, 이에 대응하기 위해서는 사고검지를 통한 별도의 모형이 필요할 것이다. 하지만 이러한 예외상황을 제외하고는 전반적으로는 요일이나 교통류 상태에 관계없이 일관된 예측력 향상을 보여 제안된 예측 모델이 어떤 상황에도 강건하게 활용될 수 있음을 보여주었다. Fig. 3은 EEMD-ANN 모델을 적용했을 때 예측력이 가장 향상된 목요일의 EEMD 수행 결과이다. 총 90일의 원 자료는 Fig. 3과 같이 같이 총 11개의 IMF와 나머지로 분해되었고, 그 중 모델을 평가하는 마지막 7일 중 목요일의 원 자료와 각 IMF의 성분을 아래 Fig. 3에 나타냈으며, 회색영역은 Fig. 2와 마찬가지로 와해구간을 나타낸다. 전체적인 통행속도 자료의 복잡성이 각 IMF에 나뉘어 완화된 형태로 나타난 것을 볼 수 있으며, 특히 와해구간의 특성은 IMF 4-6으로 분리된 것을 확인할 수 있다.

이렇게 분리된 개별 IMF를 예측한 후 다시 재결합하여 원 자료를 예측한 결과는 아래 Fig. 4에 도시하였다. Fig. 4(a)는 목요일 하루 종일, Fig. 4(b)는 평시상태인 10:00~13:20, Fig. 4(c)는 교통류 와해상태인 15:50~19:10의 ANN과 EEMD-ANN을 활용한 예측결과를 비교해서 보여준다. Fig. 4(b)는 평시상태에도 해당 구간에는 불안정한 교통류 상태로 인해 지속적인 통행시간의 변화가 나타나며, ANN은 이를 제대로 예측하지 못하는 반면, EEMD-ANN의 경우 분해된 개별 IMF를 통해 적절히 예측을 수행하는 것으로 나타났다. 이러한 차이는 Fig. 4(c)의 교통류 와해구간에도 마찬가지로 나타나며 예측치의 차이는 더 늘어나게 된다. 이는 교통류 특성 및 상태에 관계없이 실용적으로 활용될 수 있는 제안된 모델의

우수성을 직접적으로 보여준다.

4.4 연산 수행속도

본 연구에서 제안한 EEMD-ANN 모델은 기존 ANN이 예측하기 어려운 불안정한 교통류 상태도 복잡성을 낮춰 예측이 쉬운 IMF로 분해하기 때문에 더 높은 예측 정확도를 보였다. 하지만 EEMD-ANN 모델은 매시간 예측을 수행할 때 새로운 입력 값을 포함한 IMF가 필요해서 EEMD를 매번 수행해야 하므로 많은 연산비용이 필요하다. 본 연구에서 활용한 Rlibeemd와 mnet 패키지는 둘 다 R 프로그래밍 환경에서 멀티코어 연산이 가능하며, 개인용 데스크탑 PC인 Intel (R) Core (TM) i7-8700 CPU @ 3.20GHz의 6코어연산을 기준으로 5분 집계 83일 통행속도 자료에 대해 EEMD는 평균 7초, 12개 IMF에 대한 ANN 모델의 학습과 예측에는 평균 55초가 소요되었다. 이러한 연산비용은 제안한 모델의 단점으로 단일 구간에 대한 실시간 예측에는 문제가 없지만 대규모 도시네트워크에 대한 예측을 위해서는 GPU나 병렬컴퓨팅을 이용한 개선된 시스템 구축이 필요하다.

5. 결론 및 향후 연구과제

본 연구에서는 고속도로에 비해 복잡한 동적특성을 갖는 도시부 도로의 통행속도 예측을 위해 데이터 적응형 신호처리 기법인 EEMD와 비모수적 예측모형인 ANN을 결합한 EEMD-ANN 예측모형을 제안하였다. 통행속도 자료를 국지적 주기특성을 갖는

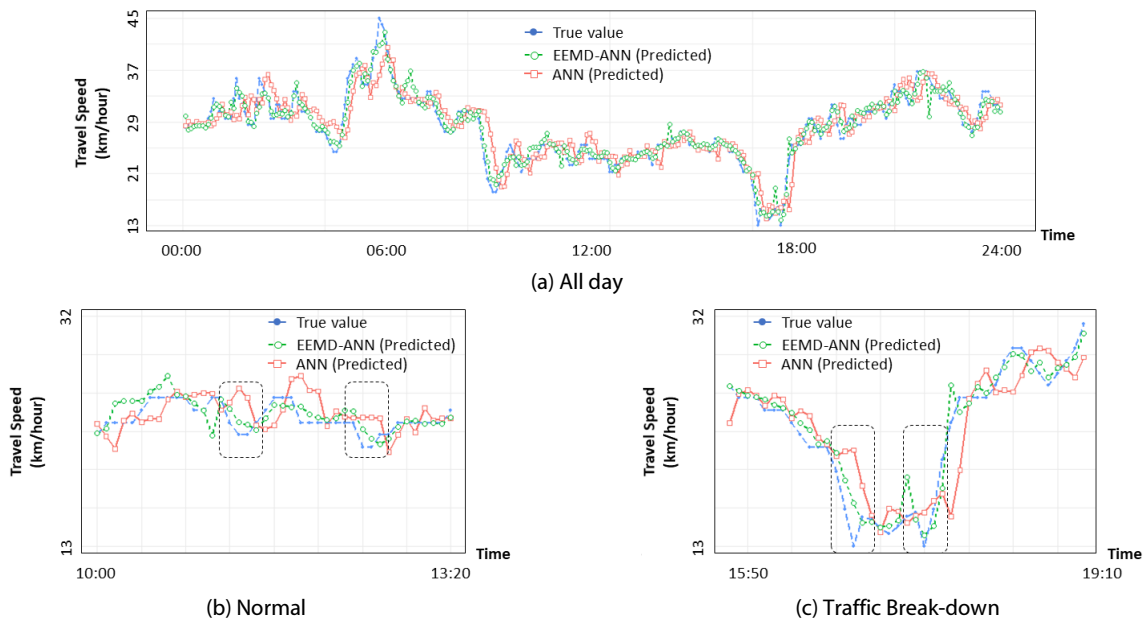


Fig. 4. Comparison of ANN and EEMD-ANN

IMF로 분해하여 복잡성을 완화하고, 분해된 개별 IMF를 예측한 후 다시 복원하는 제안된 모델은 일반적인 침두-비침두 특성과 달리 신호와 진출입로의 영향으로 상시적으로 불안정한 교통류 상태를 보이는 530m 도시부 간선도로에 적용되었다. 해당 구간은 일반적인 ANN모델을 통해서 15분 예측을 기준으로 17.88%의 오차율을 보이며, 일반적인 침두-비침두 특성을 갖는 구간의 6.68~9.96%보다 훨씬 더 큰 오차를 보였다. 하지만 본 연구의 개발모델은 해당 구간에서 요일별로 평균 11.91%의 오차를 보이며, 불안정한 교통류에도 안정적으로 예측이 가능함을 보여주었다. 이는 EEMD의 분해과정이 도시부 통행속도 자료의 복잡성을 효과적으로 완화함을 보여준다. 개발모델의 연산비용은 단일지점에 대한 실시간 예측에는 충분하지만 대규모 도시네트워크에 적용하기 위해서는 개인용 데스크탑 이상의 병렬컴퓨팅이나 GPU를 활용한 코딩이 필요하다. 하지만 이는 지자체나 공공기관 단위의 운영 시스템에서는 충분히 구현이 가능하며, EEMD나 ANN 알고리즘 내에서의 캘리브레이션 과정을 완화함으로써 보완이 가능하므로 실용적으로도 활용이 가능할 것이다.

하지만 본 연구는 몇 가지 한계점이 있고, 이를 개선하기 위한 향후연구가 필요하다. 우선 제안된 방법론은 계산시간의 문제로 하나의 연속된 영역의 도로에 대한 예측에만 활용되었는데, 도시부 광범위한 네트워크나 고속도로에도 적용하여 일반화성능에 대한 추가적인 검증이 필요하다. 또한 제안된 방법론은 단일시간대 예측에만 활용되었는데, 이는 EEMD가 갖는 끝단 부분의 오차(end-effect) 때문에 다중시간대로 확장할 경우 분해과정에서의 오차가 발생하기 때문이다(Wu and Huang, 2009). 다중시간대로 모형을 확장할 경우 단일시간대에 비해 더 적은 횡수로 EEMD나 ANN의 학습을 수행하기 때문에 연산비용을 줄일 수 있다. 그러므로 다중시간대 확장에 필요한 끝단 부분의 오차에 대한 추가적인 연구를 통해 대규모 도시네트워크에 적용하기 위한 효율성 개선이 필요하다. 마지막으로 EEMD는 자료를 분해하는 많은 신호처리 기법 중의 하나로, 통행시간 자료에 더 적합한 신호처리 기법을 찾아서 적용한다면 더 높은 성능향상을 기대할 수 있을 것이다.

감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었습니다(과제번호 17CTAP-C114766-02-000000).

References

Asif, M. T., Dauwels, J., Goh, C. Y., Oran, A., Fathi, E., Xu, M. and Jaillet, P. (2014). "Spatiotemporal patterns in large-scale traffic speed prediction." *IEEE Transactions on Intelligent Transportation*

Systems, Vol. 15, No. 2, pp. 794-804.

Chang, H. and Yoon, B. (2018). "Multiple period forecasting of motorway traffic volumes by using big historical data." *Journal of the Korean Society of Civil Engineers*, Vol. 38, No. 1, pp. 73-80 (in Korean).

Fusco, G., Colombaroni, C. and Isaenko, N. (2016). "Short-term speed predictions exploiting big data on large urban road networks." *Transportation Research Part C: Emerging Technologies*, Vol. 73, pp. 183-201.

Hamad, K., Shourijeh, M. T., Lee, E. and Faghri, A. (2009). "Near-term travel speed prediction utilizing Hilbert-Huang transform." *Computer-Aided Civil and Infrastructure Engineering*, Vol. 24, No. 8, pp. 551-576.

Jiang, X., Zhang, L. and Chen, M. X. (2014). "Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China." *Transportation Research Part C: Emerging Technologies*, Vol. 44, pp. 110-127.

Kim, H. and Jang, K. (2013). "Short-term prediction of travel time using DSRC on highway." *Journal of the Korean Society of Civil Engineers*, Vol. 33, No. 6, pp. 2465-2471 (in Korean).

Luukko, P. J. J., Helske, J. and Räsänen, E. (2016). "Introducing libeemd: A program package for performing the ensemble empirical mode decomposition." *Computational Statistics*, Vol. 31, No. 2, pp. 545-557.

Myung, J., Kim, D. K., Kho, S. Y. and Park, C. H. (2012). "Travel time prediction using k nearest neighbor method with combined data from vehicle detector system and automatic toll collection system." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2011, pp. 51-59.

Oh, C. and Park, S. (2011). "Investigating the effects of daily travel time patterns on short-term prediction." *KSCE Journal of Civil Engineering*, Vol. 15, No. 7, pp. 1263-1272.

Shin, K., Shim, S., Choi, K. and Kim, S. (2014). "Expressway travel time prediction using k-nearest neighborhood." *Journal of the Korean Society of Civil Engineers*, Vol. 34, No. 6, pp. 1873-1879 (in Korean).

Vlahogianni, E. I., Karlaftis, M. G. and Golias, J. C. (2014). "Short-term traffic forecasting: Where we are and where we're going." *Transportation Research Part C: Emerging Technologies*, Vol. 43, pp. 3-19.

Wang, H., Liu, L., Qian, Z., Wei, H. and Dong, S. (2014). "Empirical mode decomposition-autoregressive integrated moving average: Hybrid short-term traffic speed prediction model." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2460, pp. 66-76.

Wei, Y. and Chen, M. C. (2012). "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks." *Transportation Research Part C: Emerging Technologies*, Vol. 21, pp. 148-162.

Wu, Z. H. and Huang, N. E. (2009). "Ensemble empirical mode decomposition: A noise assisted data analysis method." *Advances in Adaptive Data Analysis*, Vol. 1, No. 1, pp. 1-41.