

시계열 분석을 활용한 고속도로 교통류 예측

윤병조*

Forecasting of Motorway Traffic Flow based on Time Series Analysis

Yoon, Byoung-Jo*

Abstract: The purpose of this study is to find the factors that reduce prediction error in traffic volume using highway traffic volume data. The ARIMA model was used to predict the day, and it was confirmed that weekday and weekly characteristics were distinguished by prediction error. The forecasting results showed that weekday characteristics were prominent on Tuesdays, Wednesdays, and Thursdays, and forecast errors including MAPE and MAE on Sunday were about 15% points and about 10 points higher than weekday characteristics. Also, on Friday, the forecast error was high on weekdays, similar to Sunday 's forecast error, unlike Tuesday, Wednesday, and Thursday, which had weekday characteristics. Therefore, when forecasting the time series belonging to Friday, it should be regarded as a weekly characteristic having characteristics similar to weekend rather than considering as weekday.

키 워 드: 고속도로, 교통류 예측, 시계열 분석

Key Words: Motorway, Forecasting of Traffic Flow, Time Series Analysis

1. 서론

1.1 연구의 배경 및 목적

아직 발생되지 않은 도로의 교통수요를 예측하고자 하는 것은 사업에 대해 투자를 하기 위해 결정의 중요도를 결정하고 도로의 계획을 수립하거나 이를 유지하고 관리하는 것에 대한 정책적인 의사결정을 확인하기 위하여 활용될 수 있는 중요한 판단 기준이라고 할 수 있다. 이 중 고속도로의 교통수요 자료는 교통주체의 의사 변화에 따른 효율적인 고속도로 설계 및 유지관리를 수행하기 위해 매우 중요한 자료로 활용되고 있다. 고속도로 차량 지정체의 주된 원인은 차량의 대기행렬과 교통사고이다. 따라서 고속도로의 경우, 교통수요의 변화에 따른 차량의 대기 시간 감소, 교통사고 감소, 그리고 예산 절감율도모하고 있다.

고속도로의 교통량을 예측한다는 것은 예측시간대에 따라서 단일시간대나 다중시간대 등에 대한 예측기법의 종류들을 활용하게 되며 예측기법으로는 시계열 분석이나 비모수 회귀식 등을 활용하게 된다. 이에 투입되는 변수인 예측대상은 통행시간, 교통량 등으로 구분되어진다. 고속도로에서 발생하는 심각한 돌발상황에 의한 지체시간은 1시간 이상인 경우가 많으므로 예측대상시간은 long-term 예측 이어야 한다.

이에 분석에 활용되는 ARIMA 모형은 임시적

이거나 데이터 중심의 방법이라기 보다 확률적인 접근방법과 수리통계의 기본 원리에 근거하여 활용되는 시계열 예측 방법이다. 단변량 ARIMA 모형은 투입된 변수가 적으므로 먼 과거 이력자료보다는 현행상태와 근접한 과거이력자료를 활용하여 분석하는 것이 일반적이다. 즉 장기 예측값은 단기 예측값에 비해 신뢰성이 적다. 또한 ARIMA 모형은 계절적인 변동이나 랜덤한 패턴을 보이는 반복적인 변동이 존재할 때 분석에 용이하다.

ARIMA는 여러 반복적인 패턴을 가정하고 시계열 자료를 분석하기 때문인데, 과거이력자료를 여러 패턴으로 나눠 분석한 뒤 최종적으로 남은 미세한 변동은 백색잡음으로 간주하여 정규분포를 따른다고 가정한다. 또한 ARIMA 모형은 정상성(stationarity)을 필요로 하기 때문에 여러 데이터 처리가 필요하다.

본 연구에서는 SPSS 통계 패키지를 이용한 ARIMA 예측을 수행하여 장래의 고속도로 교통량을 추정하였다.

1.2. 연구의 범위 및 방법

본 연구에서 개발되는 교통량 예측모형은 고속도로의 장래의 교통수요를 예측하기 위해 활용될 수 있으며, 이에 관련된 연구의 범위 및 내용은 다음과 같다.

첫째, 본 연구의 예측모형은 고속도로의 방대

* 인천대학교 도시과학대학 도시공학과 교수, 교신저자 (bjyoon63@inu.ac.kr).

한 과거자료를 이용하여 수행된다.

둘째, 예측모형은 장래 5분 동안의 교통량을 예측한다. 즉, 시간대(t)에서 장래 5분 동안의 교통량을 예측한다.

셋째, 연구의 공간적 범위는 고속도로 영업소이며, 사례대상 영업소는 서울외곽순환고속도로 영업소이다.

넷째, 수집자료의 범위는 2016년 1년 동안의 자료를 이용한다.

2. 기존 문헌고찰

2.1 모수 예측모형 연구

모수 모형은 기존에 알려져 있는 확률분포를 기반으로 하여 모수를 추정하는 과정이 있는 모형을 통칭한다. 가장 널리 알려져 있는 모수기반 분석방법으로 단순선형회귀분석(Simple Linear Regression Model)이 있다. 단순선형회귀분석은 독립변수와 종속변수간의 관계를 선형으로 표현하는 과정에서 필연적으로 발생하는 오차(error)에 대하여 평균은 0이고 분산은 특정한 상수값을 가지는 정규분포를 따른다는 가정하에 분석한다. 여기서 오차의 분포가 기존에 알려진 정규분포를 따를 것이라는 가정을 기반으로 모형을 수립하기 때문에 모수모형이라고 일컫는다.

유정훈(2010)은 TransCAD를 이용하여 지점의 영향권을 설정하였고, 고속도로의 교통량을 종속변수로 한 다중회귀모형을 이용하여 장래의 수요를 분석하고자 하였다. 독립변수로는 더미변수(Dummy variable)로 설정한 영향권 인구밀도, 도로연장, 영향권밀도, 개통년수, 교통량분산 등이 반영되었으며 SAS로 분석한 결과 모형의 적합성은 결정계수와 유의확률이 매우 좋게 나왔다. 하지만 관측교통량과 예측교통량의 오차율은 -30%에서 30%까지의 범위에서 나타났으며 평균 약 18.5% 정도로 최근에 진행된 연구에 비해 만족할 수준의 예측력을 나타냈다고 볼 수 없다. 연구자도 입력자료가 가지는 불확실성과 분석당시의 여건과 다르기 때문에 발생하는 오차가능성을 언급하였다.

박영환(2002)는 칼만필터링(Kalman Filtering; KF)을 이용하여 장래의 국토교통량을 예측했다. 칼만필터링은 Kalman(1960)의 “A New Approach to Linear Filtering and Prediction Problems” 연구를 통해 연구자들에게 알려지기 시작했다. 개발된 시기인 1960년대는 컴퓨터의 성능이 매우 기초적이었기 때문에 연산능력을 알고리즘적으로 보완하기 위하여 재귀적 방법을 활용한 관측갱신 알고리즘으로 부각되었다. 현재까지 다양한 형태로 발전해 왔으나 오차를 특정

한 분포를 나타내는 잡음으로 간주하여 제거하는 방법이므로 모수적 방법이라고 볼 수 있다. 박영환(2002)의 연구에서는 칼만필터링을 이용한 관측치와 예측치의 비교를 결정계수(R^2)로 나타냈으며 결정계수는 장래 교통량을 예측한 지역별로 약 81%부터 55%까지 다양하게 나타났다.

유상록(2013)은 시계열 분석(Time Series Analysis)을 이용하여 해사교통량을 예측하였다. 시계열모형 중 ARIMA 모형은 자기회귀(Auto Regression, AR) 모형과 이동평균(Moving Average, MA) 모형, 그리고 과거 데이터의 추세(Momentum)를 반영한 모형으로 데이터의 연도, 계절, 월, 일, 시간 등 시계열적으로 나타나는 패턴을 사전에 정의한 모형들로 분류하여 장래를 예측하는 방법이라 할 수 있다. 유상록(2013)의 연구에서 ARIMA 모형은 추정된 계수에 대한 검정, 확률적인 백색잡음이 유의미성을 나타내지 않고 신뢰한계선 내에 존재하는지 등에 대한 진단을 하였으며 진단결과 단순계절 모형을 최종 모형으로 결정하고 장래 해상교통량을 예측하였다. 하지만 ARIMA 모형은 시계열적으로 사전에 정의한 패턴과 뚜렷한 유사성과 유의미한 일관성을 보여야 한다는 번거로움이 있으며, 분석 결과 통계량이 유의미한 결과로 나타나도, 해당 패턴이 실질적인 의미를 지니는지에 대한 충분한 고찰이 필요하다.

홍다희(2017)은 공간통계기법을 이용하여 도시부 도로 교통량을 추정하는 연구를 진행하였다. 해당 연구에서 활용한 공간통계기법은 회귀크리깅(Regression Kriging)으로 설명변수로 설명가능한 추세를 회귀식으로 추정하고, 크리깅을 이용하여 잔차(residual)을 산정하여 보정하는 방법이다. 추정된 회귀모형에 잔차의 공간적 상관성을 고려한 가중치를 보정하므로써 보다 정확한 예측값을 산출할 수 있다는 장점이 있으나, 해당 연구에도 언급한 것처럼 설명변수인 교통량이 적을 경우 예측력이 떨어진다는 것을 알 수 있다. 즉, 잔차를 랜덤한 확률분포로 고려하지 않고 가중치로 고려함으로써, 교통량이 적은 구간에서는 상대적으로 잔차가 차지하는 비율이 크기 때문에 이를 이용한 가중치가 예측값에 큰 영향을 주는 한계가 있다.

이처럼 모수기반의 연구들을 살펴보면 예측오차가 낮지 않은 편임을 알 수 있으며 분석 과정에 필요한 가정의 유의미성을 발견해야 된다는 번거로움이 존재한다. 또한 모형을 설정하면서 수집한 자료에 대한 모형의 적합성이나 검정이 완료되었다고 하더라도, 향후 갱신된 자료에 적합한지에 대한 확인이 필요할 때 검정의 복잡성으로 인해 유동적인 대처능력이 떨어진다고 볼

수 있다. 이는 실시간으로 갱신되는 교통류 자료에 대한 적용의 한계 의미하며 처리속도 측면에서도 좋은 성능을 기대하기 어렵다고 볼 수 있다.

2.2 비모수 예측모형 연구

김혜정(1995)은 퍼지논리를 이용한 교통량 예측기법 연구를 진행하였고 교통 신호 제어시스템에 적용하였다. 퍼지모형은 몇 가지의 퍼지 명제로부터 다른 근사적인 퍼지 명제를 유도하는 근사 추론 방법이다. 현행 퍼지 명제는 IF에 해당되며, 장래상태는 THEN에 의해 근사적으로 예측된다. 김혜정(1995)은 4지 교차로를 대상으로 퍼지주기를 활용한 예측기법을 이용해 교통량의 정제 정도에 따른 신호주기를 설정하는 방법을 고안한 것이다. 해당 연구에서는 발전시킨 퍼지논리를 사용하여 변경된 신호주기가 교통혼잡을 완화시키는 것으로 나타났다. 이처럼 퍼지논리는 장래의 정확한 교통량을 예측한다기 보단 근사적인 값을 도출하여 단계를 구분하는데 주로 활용이 되고, 정확한 예측을 위해서는 주로 다른 예측 기법과 함께 사용되는 편이라고 볼 수 있다.

비모수기반 예측 이론으로 인공신경망 분석 방법도 존재한다. ANN모형은 경험적 학습을 통한 수행과정 인식이라는 방법을 통해 변수간의 관계를 나타낼 때 사용한다. 변수간의 관계가 존재할 것으로 보이나 그 관계에서 유의미한 수식이나 통계적 분석 방법을 설정하기 어려울 경우 충분히 많은 관측값을 활용할 수 있을 때 효과적으로 사용할 수 있는 분석방법이다. 즉 연구자가 자료에 대한 패턴파악이 어려울 경우 약식으로 사용할 수 있는 패턴인식 방법이라고 볼 수 있다.

강원의(2004)는 인공신경망 이론 중 다층 퍼셉트론(multi-layer perceptron; MLP)을 이용하여 국도의 교통량을 예측하는 연구를 진행했다. 다층 퍼셉트론은 단층 퍼셉트론의 XOR 문제나 비선형 자료에 적용할 수 없는 한계를 극복한 개념으로 중간에 은닉층(hidden layer)을 설정한 것이다. 때문에 비선형 모형에 적용할 수 있다는 장점이 있으나 강원의(2004)는 분석방법의 한계로 많은 매개변수(parameter)를 사용하게 됨에 따라 발생하는 과적합(overfitting)문제를 제시하였다. 이를 극복하기 위하여 전방향(feedforward) 네트워크가 아닌 오류 역전파 알고리즘에 대한 적용성에 대해 분석했으며 매개변수에 대한 다양한 모형선택기준(model selection gorithm)을 사용하였다. 하지만 해당 연구에서 학습 자료를 통해 선정된 최적의 모형들이 검증 자료에 대한

최적의 성능을 나타내지 못했으며, 학습자료와 검증자료를 공통적으로 만족시킬 수 있는 정보 평가 기준이 필요하다고 하였다.

노정현(2002)은 붓스트랩(bootstrap)기법을 활용하여 TCS 데이터로부터 차종별 교통량 추정모형을 구축하였다. 붓스트랩은 모수적인 방법과 비모수적인 방법이 동시에 존재하고 분포를 가정하고 수행하는 모수기반 붓스트랩과 많은 양의 데이터를 기반으로 표본을 재생성하여 값을 추정하는 것을 반복수행하는 비모수기반 붓스트랩으로 나뉘게 된다. 노정현(2002)은 비모수기반 붓스트랩을 수행하였으며 기존 연구 검토를 통해 표본의 무작위 추출반복횟수를 100회를 수행할 경우 유의한 결과를 도출할 수 있음을 확인하였다. 최근 Efron(2016)에 따르면 약 1000회 이상의 붓스트랩 표본을 추출해야 exact한 값을 얻을 수 있다고 나타났다.

비모수기반 예측 방법론에 대한 다른 연구로 k 최근린 이웃 비모수 회귀분석 방법이 있다(k nearest neighbor nonparametric regression, KNN-NPR). KNN은 k 최근접 군집분석 알고리즘으로 알려져 있고 NPR은 모수기반이 아닌 비모수 기반의 회귀분석으로 알려져 있지만, KNN-NPR 모형을 구축하여 활용한 연구자들이 많고자 하는 의미는 기존에 정형화 된 의미가 아닌 거리를 기반으로 하여 데이터의 적합도를 판단한 장래 예측 추세 확인이라는 거시적인 개념을 나타내는 의미로 사용했다.

백승걸(2003)은 비모수회귀식을 기반으로 하여 다중시간 고속도로 교통량을 예측하였다. 비모수회귀식은 과거관측치가 smooth 하지 않을 경우에 많은 관측자료를 확보하여 분석할 수 있다면 적용이 적합하다고 나타났다. 그리고 기존의 비모수회귀식을 이용한 연구에서 나타난 한계를 극복하기 위해 요일이나 교통류의 상황변화 주기성을 고려한 분석을 실시했다.

Gong(2002)는 KNN-NPR과 연구자가 제안한 개선된 KNN-NPR 모형을 이용한 교통량 예측을 수행하였으며 이를 인공신경망 모형 및 칼만필터링과 비교하였다. 예측오차를 비교한 결과 개선된 KNN-NPR 모형이 가장 정확한 예측값을 나타냈고, KNN-NPR, 칼만필터링, 인공신경망 모형 순으로 정확하다고 나타났다. 하지만 당시 데이터 처리 기술은 현재와 달리 매우 느렸고 상당한 양의 데이터를 처리할 수 있는 시스템이 구축되지 않은 상태였기 때문에 데이터 처리 속도가 저조하였고 시스템을 활용하기 위한 비용이 높아 이를 연구의 한계와 향후 연구의 방향으로 언급하였다.

하지만 10년이 지난 Chang(2012)의 연구에서 KNN-NPR 모형을 이용하여 장래 교통량을 증분

히 예측가능하도록 모형을 구축하였다.

KNN-NPR 모형은 상태벡터(state vector)와 결정자(determinant), 예측함수로 구성했으며 KNN을 이용하여 단기예측 벡터를 생성하고 이를 이용하여 예측 알고리즘을 형성, 이후에 여러 시간 단계마다 출력 벡터를 갱신하는 반복적인 프로세스를 제공하도록 하였다. 분석자료는 일주일에 7일, 20주 동안 1일 15분 단위의 96회 간격으로 자료를 수집하여 총 13,440건의 관측자료를 활용하였고 해당 자료를 활용하여 시간대와, 요일을 나누어 분석한 결과 평균 절대 백분율 오차(Mean absolute percentage error, MAPE)를 비교한 결과 평균적으로 KNN-NPR이 6.92%, ARIMA 모형이 13.07%의 오차율을 나타내 KNN-NPR의 우수성을 언급하였다.

2.3 예측방법에 따른 특성 비교

앞서 기존 연구를 검토하며 크게 두 가지로 분류하여 고찰하였다. 하나는 모수적 모형을 사용한 연구들이고, 다른 하나는 비모수적 모형을 사용한 연구들이다. 여기서 모수적 모형을 사용하는 상당수의 연구들은 연구자의 가정을 전제로 한다. 즉, 교통량 데이터 상태공간의 패턴과 구조적인 부분에 대해 심층적인 탐구가 이뤄지는 것이 아니라, 자료가 수집되는 공간의 외적인 요소에서 예측오차를 줄일 수 있는 가정과 조건이 추가되는 것이다.

예를들어 예측할 교통량 정보의 정확도를 높이기 위해서는 연구자의 가정에 따라 다량의 설명변수들이 투입되며 투입되는 변수들은 예측교통량의 경향성이나 추세성을 설명할 수 있도록 분석된다. 특히 분석 방법도 한 가지 방법인 경우 뿐만아니라 여러 통계 분석 모형을 적용한 뒤 최종적으로 가장 설명력이 높은 설명변수와 모형을 선택하게 된다. 이러한 모수 모형을 적용하는 많은 연구자들이 수행하는 연구 절차는 연구자가 충분히 체감할 수 있는 설명변수를 선택할 수 있다는 것과 과거에 충분히 검토된 모형을 사용함으로써 확률적으로 최대한 타당하다고 판단할 수 있는 예측 모형을 만들어 결론을 도출해내기에 유리하다는 장점이 있다.

하지만 통계적 모형을 수식적으로 재검토하는 것이 아니라 여러 분야의 연구에서 가정을 검토하기 위해 적용하는 것이 대부분이었기 때문에 교통량의 장래상태를 예측하는 방법에 있어서 예측 성능이나 시간적, 공간적으로 교통량 상태가 변화하는 모습을 예측하는데 중점을 두고 있지 않다는 한계를 가지고 있다.

이에 비해 비모수모형을 적용한 연구들은 설명변수와 교통량 장래상태의 인과관계보다 분석

성능, 예측 성능, 시공간적 상태 변화에 따른 대응 성능 등의 향상에 연구 초점이 맞춰져 있다. 때문에 예측 오차를 감소하기 위하여 설명변수와 교통량 장래상태간 인과관계의 가정을 재검토하는 것이 아니라, 설계한 모형을 수식적으로 재검토하게 된다. 또한 기술적으로 수행가능한 최적의 수행횟수만큼 분석을 반복하여 최적의 예측 값을 도출해내는 방법을 사용한다. 이처럼 모수기반의 분석 방법론과 비모수기반의 분석 방법론의 특징을 비교한 것을 Table 1에 나타냈다.

Table 1. 예측방법론 특성 분류

연구 종류	연구 초점	예측오차 감소방안
모수모형 기반 연구	<ul style="list-style-type: none"> • 통계적 모형을 이용한 연구자 가정 검토 • 설명변수와 종속변수의 인과관계 검증 • 정책적 시사점 도출 	<ul style="list-style-type: none"> • 다량의 설명변수를 투입한 뒤 유의한 변수만 선택 • 여러 분석 모형을 사용한 뒤 가장 적합한 모형 선택
비모수모형 기반 연구	<ul style="list-style-type: none"> • 분석 성능 향상 • 예측 성능 향상 • 시공간적 변화에 대응 가능한 모형 도출 	<ul style="list-style-type: none"> • 분석의 반복 수행을 통한 최적의 예측값 도출 • 모형의 수식적 재검토

2.4 ARIMA 모형 고찰

ARIMA 모형은 임시적이거나 데이터 중심의 방법이라기 보다 확률적인 접근방법과 수리통계의 기본 원리에 근거하여 활용되는 시계열 예측 방법이다. 단변량 ARIMA 모형은 투입된 변수가 적으므로 먼 과거이력자료보다는 현행상태와 근접한 과거이력자료를 활용하여 분석하는 것이 일반적이다. 즉 장기 예측값은 단기 예측값에 비해 신뢰성이 적다. 또한 ARIMA 모형은 계절적인 변동이나 랜덤한 패턴을 보이는 반복적인 변동이 존재할 때 분석에 용이하다.

ARIMA는 여러 반복적인 패턴을 가정하고 시계열 자료를 분석하기 때문인데, 과거이력자료를 여러 패턴으로 나눠 분석한 뒤 최종적으로 남은 미세한 변동은 백색잡음으로 간주하여 정규분포를 따른다고 가정한다. 또한 ARIMA 모형은 정상성(stationarity)을 필요로 하기 때문에 여러 데이터 처리가 필요하다.

본 연구에서는 SPSS 통계 패키지를 이용한 ARIMA 예측을 수행하여 장래의 고속도로 교통량을 추정하였다.

3. 자료수집 및 분석방법

3.1 자료 수집

교통량 자료는 고속도로에서 수집된 2016년 1

월 1일 0시부터 2016년 12월 31일 24시까지의 5분 간격의 통행량 자료를 활용하였다. 2016년 교통량 자료는 2016년이 2월이 29일까지 있는 윤년이기 때문에 366일 자료라는 점이며 따라서 구축된 교통량 자료의 크기는 105,408건 (12 sequence/hour × 24 hour × 366일)이다.

Figure 1은 예측의 정확도를 파악하기 위한 검증자료로 활용되는 교통류 모습을 그래프로 나타냈다.

3.2 예측모형의 정확성 평가척도

MAPE(Mean absolute percentage error)는 실측값과 예측값의 상대적인 비율을 고려할 수 있어 유사성과 편차를 동시에 고려할 수 있다는 특징을 가지고 있으나 예측되는 교통량의 크기가 작으면 작을수록 비율이라는 척도의 특성상 오차의 크기가 상대적으로 크게 나타날 수 있기 때문에 MAPE 뿐 아니라 MAE(Mean absolute error)도 확인할 필요가 있다.

또한 보다 정확한 예측값 평가를 위해서 예측값이 편중되어 있는 정도를 확인할 수 있는 상대백분율오차(Relative percentage error)의 평균(MRPE)을 산정함으로써 예측값의 균형성을 확인한다. 정확성은 실측값을 중심으로 예측값이 벗어난 오차의 절대 크기와 상대적 비율을 확인한다면, 균형성은 실측값을 중심으로 예측값이 ± 방향으로 쏠려있는 정도와 오차의 분포가 확산된 정도를 확인한다고 볼 수 있다.

$$MAE(veh) = \sum_{i=1}^n \frac{|x_i - y_i|}{n}$$

$$MAPE(\%) = \sum_{i=1}^n \frac{|x_i - y_i|}{n \times x_i} \times 100.0$$

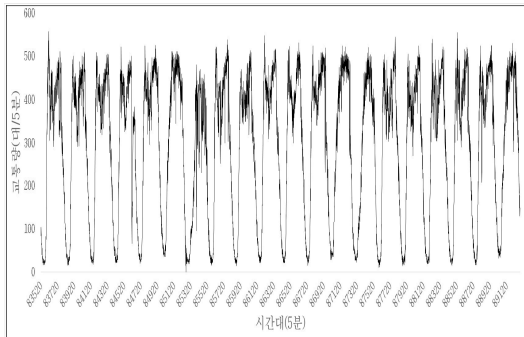


Figure 1. 수집된 고속도로 교통량 자료 모습

$$RPE(\%) = \sum_{i=1}^n \frac{x_i - y_i}{x_i} \times 100.0$$

$$MRPE(\%) = \sum_{i=1}^n \frac{x_i - y_i}{n \times x_i} \times 100.0$$

3.3 ARIMA

시계열은 크게 정상적(Stationary)시계열과 비정상적(Nonstationary) 시계열로 나누어지며 ARIMA모형은 정상적시계열에 대해 모형의 식별, 추정 등의 과정을 거쳐 예측을 하게 된다. ARIMA를 수행하는 일련의 과정은 비정상적시계열을 정상적시계열로 변환시켜 수행해나가는 과정이라고 볼 수 있다.

분석하고자 하는 대상이 되는 고속도로 교통량 자료의 시계열적 특성은 가시적으로는 대부분 일주일을 주기로 반복되는 정상적시계열로 여러 예측모형을 활용하여 분석할 시 어느정도 통계적 유의성을 가지는 모형이 개발될 수 있을 뿐만 아니라, 계절성을 갖고 있다. 이처럼 일정한 시간 간격으로 일정한 형태를 반복하고 있을 때 계절시계열(Seasonal Time Series)이라고 한다. Smith and Demetsky(1997)은 ARIMA모형을 이용하여 고속도로 15분 단위 교통량을 예측하였으며, ANN과 NPR보다 예측력은 낮게 분석되었다. 이후 Smith et al. (2003)은 SARIMA와 NPR을 이용하여 15분 집계 고속도로 교통량을 추정하였으며, 정확도 측면에서는 SARIMA가 다소 우수하나 예측력의 차이가 없었으며 다량의 이력 자료가 이용될 경우 NPR의 예측력이 우수할 수 있음을 향후 연구로 지적하였다.

4. 분석 결과

4.1 ARIMA 예측 모형

고속도로 1일치 교통량을 ARIMA모형으로 예측한 모습을 Figure 2에 나타내었다. ARIMA 모형의 경우 0시부터 6시까지의 교통량 예측이 거의 유사한 흐름으로 예측된 것을 볼 수 있는데, 앞서 언급한 것처럼 ARIMA는 smooth한 예측값을 산출하는 경향이 있기 때문에 상대적으로 교통량과 변동폭이 적은 새벽시간대에 적합했다. 다만 과다예측을 하게 되어 실측된 새벽교통량에 비해 높은 교통량을 예측하고 있는 것을 확인할 수 있다.

예측의 주목적이라고 할 수 있는 침투 시간대에 대해서도 어느정도 유의미한 예측을 수행하고 있다. 다만 전체적인 시간대에 대하여 급격한 변동은 반영하지 못하는 평균적인 추세의 시계열 예측이 수행되었음을 확인할 수 있다.

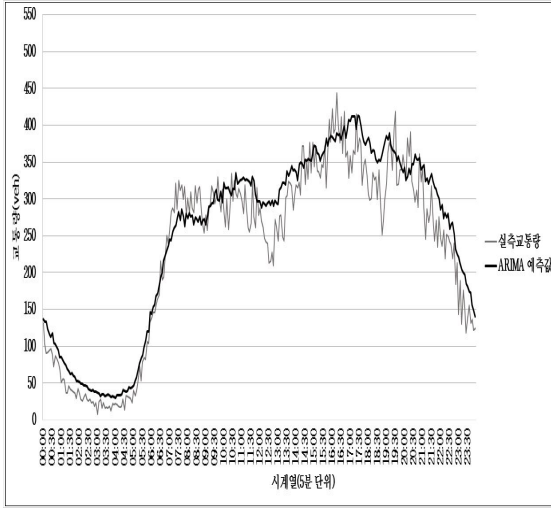


Figure 2. ARIMA 모형 적용 교통량 예측

Table 35. ARIMA 모형 적용 교통량 예측오차

시간대	MAPE(%)	MAE(veh)	MRPE(%)	SDRPE(%)
전체	23.53	28.81	-20.61	34.96
06시 이후	12.12	32.38	-8.23	18.29
07시~09시	9.11	27.67	8.21	29.45
18시~20시	13.97	43.06	-11.61	16.03

Table 2는 예측된 1일치 교통량에 대해 시간대별로 예측오차를 나타낸 표이다. 전체 시간대에 대한 MAPE는 23.53%로 일반적으로 알려져 있는 모수모형이나 비모수 모형 중 활발히 연구되고 있는 KNN-NPR 모형에서 제시되는 평균적인 MAPE가 약 10% 정도임을 감안하였을 때 약 13%p 높게 나타나 예측오차가 심한 편임을 알 수 있다. 하지만 오전 첨두시간대에 대한 MAPE는 약 9.11%로 전체적인 시간대에 비해 오전 첨두시간대에 대한 예측이 잘 수행되었음을 확인할 수 있다. 또한 저녁 첨두시간대에 대해서도 13.97%로 전체적인 MAPE에 비해 낮은 오차를 나타내고 있다. 이는 6시 이후 MAPE가 전체 시간대 MAPE와 비교했을 때 12.12%로 낮은 것을 보았을 때, 새벽 0시부터 6시 사이의 시간대에서 MAPE가 급격히 높아지는 원인이 있다고 볼 수 있다. MAPE는 실측값이 작을수록 크게 나타나는 경향이 있는데 이유는 상대오차이기 때문이다.

교통량이 적은 새벽시간대는 상대적으로 MAPE가 크게 나타날 수밖에 없고, 새벽시간대의 MAPE가 반영된 전체 시간대에서는 오차가 크게 나타나는 것이다. MAE를 확인하면 분명하

게 확인할 수 있는데 전체 시간대의 MAE는 약 28.81대고 오전 첨두시간대의 MAE는 27.67대로 큰 차이를 나타내고 있지 않다. 특히 MAPE가 13.97%로 나타났던 저녁 첨두시간대의 MAE는 43.06으로 MAE는 전체 시간대에 비해 오차가 더 커진 것을 확인할 수 있다.

ARIMA 모형은 일반적으로 시계열의 데이터 특성에 영향을 특히 많이 받는 모형으로 참조하는 과거이력자료의 크기가 클수록 예측값은 평균에 가까워진다는 단점이 있으며 이에 따라 여러 보정을 통해 장기예측을 수행하지만 일반적으로 단기 예측에 활용된다.

고속도로 교통량은 요일에 따른 교통류 특성이 분명히 나타나기 때문에 요일에 따라 교통량 예측 특성이 나뉘질 수 있다고 판단했다. 따라서 고속도로 교통량을 보다 정확하게 예측하기 위해 요일별 예측을 수행하였다.

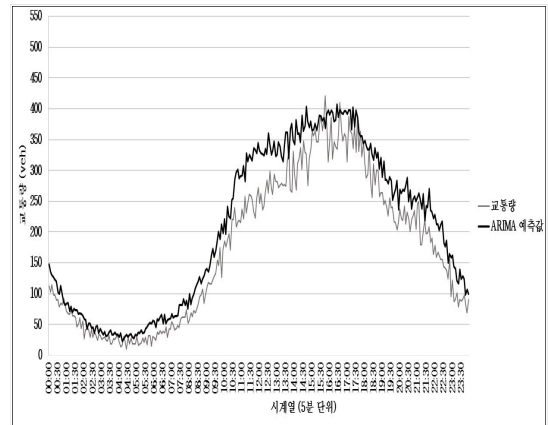


Figure 3. 요일별 ARIMA 모형 적용 교통량 예측 (일요일)

Table 3. 요일별 ARIMA 모형 적용 교통량 예측오차 (일요일)

시간대	MAPE(%)	MAE(veh)	MRPE(%)	SDRPE(%)
전체	31.2	34.8	-30.9	186.5
06시 이후	25.7	41.7	-25.5	179.5
07시~09시	41.3	26.9	-41.3	195.0
18시~20시	16.6	41.6	-16.5	169.8

4.2 요일별 ARIMA 예측 수행

Figure 3부터 Figure 7까지, Table 3부터 Table 7까지 요일별로 ARIMA 모형을 활용하여

고속도로 교통량 예측을 수행한 모습이다. 우선 Figure 3과 Figure 7의 일요일, 금요일 예측 결과를 살펴보면 출근시간의 교통량 증가 특성이 나타나지 않는 교통량 모형을 확인할 수 있다. 일요일 교통량 예측결과에 대해서는 MAPE가 약 31.2%로 여러 예측 결과나 본 연구에서 수행한 ARIMA 모형들에 비해서 높은 예측 오차를 나타냈다. 특히 MRPE가 -30.9%로 예측치에 비해 실측값이 낮게 나타났다는 것을 수치적으로 확인할 수 있으며 Figure 3을 보더라도 예측값이 전반적으로 실측값에 비해 높은 값을 보이는 예측을 수행했음을 확인할 수 있다. 또한 오전첨두시간의 예측오차가 전체 시간대의 예측오차에 비해 낮게 나타나는 평일에 비해, 일요일은 오전첨두시간대의 예측오차가 전체 시간대에 비해 MAPE가 약 10%point 높게 나타났다.

일요일 과거 이력자료를 활용한 장래 교통량 예측의 경우 오차가 높은 이유를 확인하기 위해 예측의 대상이 되는 2016년 1월 24일에 대한 특징을 발견하고자 조사하였으나, 큰 특징은 확인할 수 없었고 간략하게 1월 24일에 큰 한파가 있었다는 사실만 확인할 수 있었다. 때문에 과거 이력자료로 참조되었던 일요일들에 비해 예측대상이 되는 일요일의 실측 교통량이 낮게 나타났다고 추측할 수 있다.

Figure 4와 Table 4는 화요일의 교통량 예측 값을 나타낸 것이다. 화요일의 경우 일요일의 교통량 예측에 비해 좋은 예측력을 보이는 것을 알 수 있다. 특히 오전 첨두시간의 MAPE는 6.2%로 낮았고, MAE는 약 14.6대로 일요일 예측에 비해 실측 교통량에 비해 벗어나지 않는 교통량을 예측했음을 확인할 수 있다. 특히 오전 첨두시간의 교통량을 잘 설명하면서 출근 시간의 정체를 미연에 방지할 수 있는 예측 모형을 제시하고 있다고 판단한다.

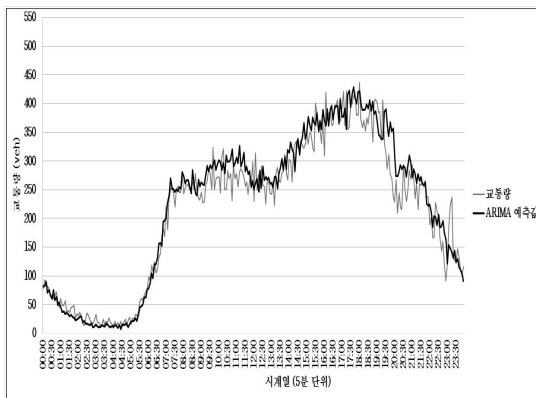


Figure 4. 요일별 ARIMA 모형 적용 교통량 예측 (화요일)

Table 4. 요일별 ARIMA 모형 적용 교통량 예측오차 (화요일)

시간대	MAPE(%)	MAE(veh)	MRPE(%)	SDRPE(%)
전체	14.0	21.7	0.2	90.8
06시 이후	10.7	26.5	-5.7	95.7
07시~09시	6.2	14.6	-4.1	93.4
18시~20시	13.3	42.5	-8.4	98.7

Figure 5와 Figure 6, Table 5와 Table 6은 각각 수요일과 목요일 교통량을 예측한 모습을 나타낸 것이다. 화요일과 마찬가지로 일요일에 비해 예측오차가 낮은 것을 확인할 수 있으며 특히 화요일, 수요일, 목요일과 같은 평일특성을 가지는 요일중 수요일이 가장 좋은 예측력을 나타내고 있다. 오전 첨두시간만 비교하였을 때, 화요일이 MAPE가 6.2%, MAE가 14.6대, 수요일은 MAPE가 6.1%, MAE가 16대, 목요일은 MAPE가 6.7%, MAE가 19.4대로 MAPE는 수요일이 가장 우수하다. 하지만 상대오차를 산정하는 변수의 값이 클수록 MAPE가 상대적으로 작게 나올 수 있다는 점을 고려했을 때 평일에 해당되는 요일별 오차 차이는 큰 의미를 가진 차이라고는 할 수 없을 것이다. 하지만 주말특성을 가진 일요일에 비해 평일특성에 대한 ARIMA 분석의 정확도가 더 높게 나타났다는 것은 확인가능하다.

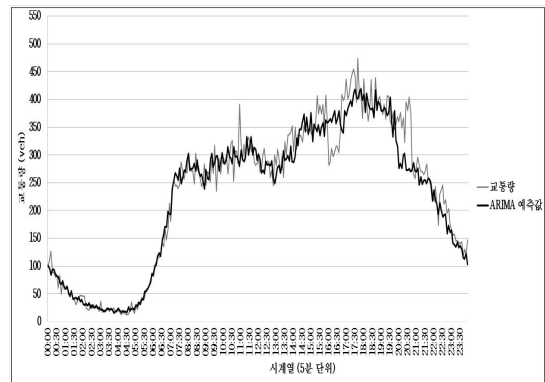


Figure 5. 요일별 ARIMA 모형 적용 교통량 예측 (수요일)

Table 5. 요일별 ARIMA 모형 적용 교통량 예측오차 (수요일)

시간대	MAPE(%)	MAE(veh)	MRPE(%)	SDRPE(%)
전체	10.8	20.7	0.0	102.2
06시 이후	8.5	25.7	2.1	99.5
07시~09시	6.1	16.0	-1.7	103.0
18시~20시	5.8	22.4	2.0	99.2

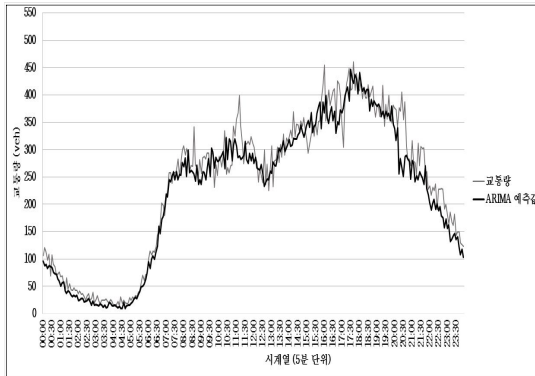


Figure 6. 요일별 ARIMA 모형 적용 교통량 예측 (목요일)

Table 6. 요일별 ARIMA 모형 적용 교통량 예측오차 (목요일)

시간대	MAPE(%)	MAE(veh)	MRPE(%)	SDRPE(%)
전체	12.8	22.4	9.5	115.5
06시 이후	9.1	26.7	5.0	119.5
07시~09시	6.7	19.4	5.8	118.4
18시~20시	4.3	16.6	0.4	123.7

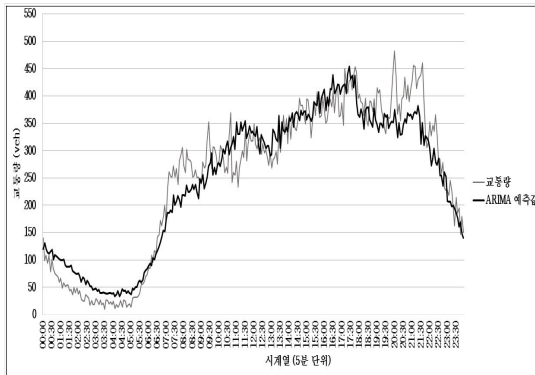


Figure 7. 요일별 ARIMA 모형 적용 교통량 예측 (금요일)

Table 7. 요일별 ARIMA 모형 적용 교통량 예측오차 (금요일)

시간대	MAPE(%)	MAE(veh)	MRPE(%)	SDRPE(%)
전체	29.3	32.0	-18.3	140.2
06시 이후	11.1	34.4	3.5	110.3
07시~09시	18.6	50.0	18.4	95.0
18시~20시	8.4	32.9	5.6	107.7

Table 8. 요일별 전체시간대 ARIMA 모형 적용 교통량 예측오차

요일	MAPE(%)	MAE(veh)	MRPE(%)	SDRPE(%)
일요일	31.2	34.8	-30.9	186.5
화요일	14.0	21.7	0.2	90.8
수요일	10.8	20.7	0.0	102.2
목요일	12.8	22.4	9.5	115.5
금요일	29.3	32.0	-18.3	140.2

Figure 7과 Table 7은 금요일 교통량에 대한 예측을 나타낸 것이다. 금요일에 대한 ARIMA 모형의 예측은 금요일이 평일임에도 불구하고 평일특성보다는 주말특성을 가지고 있는 것으로 나타났다. 화요일, 수요일, 목요일에 대한 ARIMA 모형 예측오차 특성과 다르게 일요일에 대한 ARIMA 모형 예측오차와 비슷한 모습을 나타냈기 때문이다. 따라서 금요일은 평일을 마무리하는 시점으로 교통량은 평일특성보다 주말특성에 따른다고 판단된다.

4.3 요일별 ARIMA 예측오차 특성 비교

Table 8은 요일별로 전체시간대에 대한 ARIMA 모형 예측오차를 나타낸 것이다. 표를 참고하면 일요일과 금요일의 예측오차가 유사하고, 화요일, 수요일, 목요일의 예측오차가 유사한 것을 확인할 수 있다. MAPE는 일요일, 금요일이 약 30%를 나타냈으며, 화요일, 수요일, 목요일은 약 10%에서 14%까지의 값을 나타내고 있다. MAE 또한 일요일과 금요일은 약 30대인 것에 비해 평일특성을 가지는 화요일, 수요일, 목요일은 약 20대의 MAE를 나타내면서 확연하게 다른 양상을 나타내고 있다. 따라서 향후 고속도로 교통량을 예측하기 위해서는 주말특성과 평일특성이

나닌다는 점을 고려하여 과거 이력자료를 요일 구분 없이 활용하지 않고, 요일 특성이 반영된 자료로 구분하여 예측하거나 이에 대한 보정계수를 산출하여 활용해야 좋은 예측 모형이 개발될 것으로 판단된다.

5. 결론

고속도로 교통량 자료를 활용하여 교통량에 대한 예측을 ARIMA 모형을 이용해 수행한 결과 예측오차를 줄이기 위해서는 요일별로 구분해야 된다는 점을 확인하였으며, 특히 평일특성과 주말특성이 구분됨을 예측오차를 통해 확인하였다.

예측결과 평일특성은 화요일, 수요일, 목요일에서 두드러지게 나타났으며, 일요일은 MAPE와 MAE를 비롯한 예측오차가 평일특성을 지닌 요일에 비해 각각 약 15%point, 약 10대 정도 높게 나타나 주말특성으로 분류해야 됨을 확인하였다.

또한 금요일은 평일임에도 불구하고 예측오차가 평일특성을 지닌 화요일, 수요일, 목요일과 다르게 일요일 예측오차와 유사한 모습을 보이며 높게 나타났다. 따라서 금요일에 속하는 시계열을 예측하고자 할 때는 평일로 고려한다기 보다는 주말과 유사한 특성을 가지는 주말 특성으로 간주해야 할 것이다.

전반적으로 ARIMA 분석모형은 주말에 비해 평일에 대한 예측력이 좋게 나타났으며 이는 오전 첨두시간의 출근 교통량의 두드러짐 유무에 따라 구분된다고 볼 수 있다. 화요일, 수요일, 목요일 교통량은 일요일, 금요일 교통량에 비해 오전 첨두시간의 출근 교통량의 급격한 유입이 두드러지게 나타났으며 이처럼 패턴을 구분지을 수 있는 요인이 내재되어 있는 교통량 자료일 경우 예측을 위한 모형 판별이 수월하기 때문이라고 고려된다.

따라서 향후 연구를 통해 요일별 특성을 보다 세분화 하여 구분하고 예측을 수행할 수 있는 기준을 수립해야 할 것이며 좋은 예측 모형이어도 자료의 상태에 따라 예측력이 낮게 나타날 수 있다는 점을 고려하여 자료를 분석하기 위해 자료를 가공하는 과정에서 활용할 수 있는 보정계수나 기준을 수립하는 것이 필요할 것이다.

REFERENCES

Altman, N.S.(1992), "An introduction to kernel and nearest neighbor nonparametric regression", *The American Statistician*, vol. 46, pp. 175-185.

Chang, H. H., Yoon, B. J (2018). "High-Speed Data-Driven Methodology for Real-Time Traffic Flow Predictions: Practical Applications of ITS". *Journal of Advanced Transportation*.

Chang, H., Lee, Y., Yoon, B., and Baek, S. (2012), "Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences", *IET: Intelligent Transport Systems*, vol. 6(3), pp. 292-305.

Clark, S.(2003), "Traffic prediction using multivariate nonparametric regression", *Journal of Transportation Engineering*, vol. 129(2), pp. 161-168.

Davis, G. and Nihan, N. (1991), "Nonparametric regression and short-term freeway traffic forecasting", *Journal of Transportation Engineering*, vol. 117, pp. 178-188.

Gong, X., & Wang, F (2002), "Three improvements on KNN-NPR for traffic flow forecasting." In *Intelligent Transportation Systems, Proceedings The IEEE 5th International Conference*, pp. 736-740

Hobeika, A. G., & Kim, C. K (1994), "Traffic-flow-prediction systems based on upstream traffic", In *Vehicle Navigation and Information Systems Conference*, pp. 345-350

Ishak, S., and Al-Deek, H. (2002), "Performance evaluation of short-term time series traffic prediction model", *Journal of Transp. Engineering*, vol. 128(6), pp. 490-498.

Ishak, S., and Alecsandru, C. (2004), "Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings", *Journal of Transportation Engineering*, vol. 130(4), pp. 452-465.

Kalman, R. E. (1960), "A new approach to linear filtering and prediction problems", *Journal of basic Engineering*, 82(1), 35-45.

Karlsson, M., and Yakowitz, S.(1987), "Rainfall-runoff forecasting methods, old and new", *Stochastic Hydrology and Hydraulics*, vol. 1, pp. 303-318.

Oswald R. K., Scherer W.T. and Smith B.L. (2000), "Traffic flow forecasting using approximate Nearest Neighbor Nonparametric regression", *Research project*

- report for U.S. DOT University transportation center.
- Smith, B.L., and Demetsky, M.J.(1995), "Short-term traffic flow prediction: neural network approaches", Transportation Research Record, vol. 1453, pp. 98-104.
- Smith, B.L., and Demetsky, M.J.(1996), "Multiple-interval freeway traffic flow forecasting", Transportation Research Record, vol. 1554, pp. 136-141.
- Smith, B.L., and Demetsky, M.J.(1997), "Traffic flow forecasting: comparison of modeling approaches", Journal of Transportation Engineering, vol. 123(4), pp. 261-266.
- Smith, B.L., and Oswald, R.K.(2003), "Meeting real time traffic flow forecasting requirements with imprecise computations", Computer-Aided Civil and Infrastructure Engineering, vol. 18(13), pp. 201-213.
- Smith, B.L., and Ulmer, J.M.(2003), "Freeway traffic flow rate measurement: Investigation into impact of measurement time interval", Journal of Transportation Engineering, vol. 129(3), pp. 223-229.
- Smith, B.L., Williams, B.M., and Oswald, R.K.(2002), "Comparison of parametric and nonparametric models for traffic flow forecasting", Transportation Research Part C, vol. 10, pp. 303-321.
- Yoon, B. et al. (2014), "Potentialities of data-driven nonparametric regression in urban signalized traffic flow forecasting", Journal of Transportation Engineering, 140(7), 04014027.

논문투고일: 6월 19일
논문게재일: 6월 30일