의사결정나무 기법을 적용한 DSRC 통행속도패턴 분류방안

Study on the Classification Methodology for DSRC Travel Speed Patterns Using Decision Trees

이 민 하*
(Minha Lee)

이 상 수**
(Sang-Soo Lee)

남 궁 성***
(Seong Namkoong)

최 기 주****
(Keechoo Choi)

요 약

본 논문의 목적은 DSRC 기반 통행속도 이력데이터를 활용하여 IC-IC 구간 단위의 통행패턴을 도출하는 것이며, 이를 통해 방대한 이력정보 데이터의 활용도를 높이고, 단순하지만 정확성 높은 방법으로 도로의 통행패턴을 용이하게 파악할 수 있게 하는 것이다. 통행패턴 분류는 의사결정나무 기법을 적용하였고, 월·시간대·구간 단위로 분리된 통행패턴을 생성하여 시·공간이 변화되어도 이에 대응 가능하도록 하였다. 경부고속도로 서울TG-안성IC 구간을 대상으로 의사결정나무 기법을 적용한 결과, 요일 기준으로 (월)(화·수·목)(금)(토)(일) 5개 그룹으로 고정 통행패턴이 분류되었다. 분류 결과를 영동, 중부, 중부내륙 고속도로의 9개 구간에 적용하여 통계적 검증을 수행한 결과 약 93%의 적합도를 갖는 것으로 나타났다. 의사결정나무를 통한 통행패턴 오차를 개선하기 위하여 4개의 추가변수를 도입한 결과, “직전월의 소통상황”을 설명변수로 추가할 경우 통행속도 분산이 약 50% 감소함을 확인하였고, 실제 상황에 적용할 경우 소통 원활 시의 오차가 약 4% 감소되었다.

핵심어 : DSRC, 통행패턴, 의사결정나무, 이력 데이터, 연속류 속도

ABSTRACT

In this paper, travel speed patterns were deducted based on historical DSRC travel speed data using Decision Tree technique to improve availability of the massive amount of historical data. These patterns were designed to reflect spatio-temporal vicissitudes in reality by generating pattern units classified by months, time of day, and highway sections. The study area was from Seoul TG to Ansung IC sections on Gyeong-bu highway where high peak time of day frequently occurs in South Korea. Decision Tree technique was applied to categorize travel speed according to day of week. As a result, five different pattern groups were generated: (Mon)(Tue-Wed-Thu)(Fri)(Sat)(Sun). Statistical verification was conducted to prove the validity of patterns on nine different highway sections, and the accuracy of fitting was found to be 93%. To reduce travel pattern errors against individual travel speed data, inclusion of four additional variables were also tested. Among those variables, ‘traffic condition on previous month’ variable improved the pattern grouping accuracy by reducing 50% of speed variance in the decision tree model developed.

Key words : DSRC, Travel patterns, Decision Tree model, historical data, Free flow speed

* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2010-0029450)

* 주저자 : 아주대학교 ERC 센터 연구원

** 교신저자 : 아주대학교 교통공학과 교수

*** 공저자 : 한국도로공사 도로교통연구원 교통연구실장

**** 공저자 : 아주대학교 교통공학과 교수

† 논문접수일 : 2013년 12월 27일

† 논문심사일 : 2014년 02월 07일

† 게재확정일 : 2014년 03월 19일

I. 서 론

1968년 국내 최초 고속도로인 경인 고속도로의 완공과 더불어 고속 산업화가 진행되는 과정에서 교통 수요가 급격히 증가하였다. 초기에는 도로 건설로 수요를 해소하였으나, 토지 및 예산 문제 등으로 물리적 공급 기반 운영 관리 방식은 한계에 도달하게 되었다. 1990년대부터는 ITS 기술 및 국내 주요 교통축을 토대로 지능형 수요 관리를 시작하였으며, 이로 인해 교통정보의 중요성이 대두되었다. 정부는 이러한 실정에 맞추어 ITS 사업 진행, 프로브 차량 기반 정보수집 시스템 구축 등 첨단 교통시스템 도입을 활발히 진행 중에 있다. 특히 고속도로 교통관리에 필요한 정보 수집은 차량검지기 시스템(VDS: Vehicle Detection System) 및 자동요금징수 시스템(TCS: Toll Collection System), 단거리무선통신(DSRC: Dedicated Short Range Communication) 등을 기반으로 이루어지며, 이를 통해 방대한 양의 고속도로 이력정보 데이터가 수집 및 축적된다.

그러나 그 방대한 양에도 불구하고, 이에 상응하는 활용방안은 부족한 실정이다. 대용량 데이터(big data)의 축적에서 그칠 것이 아니라 이를 활용하여 교통 흐름의 함축정보 및 교통 패턴 도출에 적극적인 개입이 필요한 시점이다. 이러한 패턴 정보는 장래 교통상황 예측 및 운영관리 의사결정에 매우 유용한 정보를 제공하고, 이를 통하여 고속도로 관리시스템의 효율성을 증진시킬 수 있다.

본 연구에서는 현재 방대하게 수집되는 고속도로 DSRC 데이터를 이용하여 구간단위의 통행패턴의 분류 개수를 최소화하면서 일정 수준의 정확성(accuracy)이 보장되는 통행패턴을 도출하는 것을 목표로 하였다. 이는 현재 지속적으로 축적되는 DSRC 데이터가 유연성 있게 패턴에 반영될 수 있게 하기 위함이며, 향후 고속도로 관리자가 고속도로 운영관리 의사결정에 있어 단순하면서도 정확한 방법으로 도로의 통행패턴을 용이하게 파악할 수 있게 하기 위함이다. 특히 구간 단위의 통행패턴 분류는 O/D별 통행속도를 예측하는 방법에 비해 훨씬

간결하면서도, 충분한 샘플 수의 확보가 가능하기에 신뢰도가 높다. 구간 단위의 통행패턴이 고속도로 전 구간에 구축되면 구간별 통행패턴 조합을 통해 모든 O/D에 대한 통행패턴의 산출에 용이하게 활용될 수 있다.

본 연구를 수행하기 위한 분석에는 개별 차량 통행시간을 15분 단위로 집계한 DSRC 통행속도 이력데이터가 활용되었다. 본 데이터를 기반으로 의사결정나무를 활용한 통행패턴 분류와 산출, 통계 오차비교를 통한 통행패턴의 정확성 비교, 적정 이력자료 활용기간 선정을 수행하였다. 특수상황의 경우는 일반 상황과는 상이한 통행 특성을 나타내므로, 이러한 자료를 분리한 후 일반상황에 대한 통행패턴을 도출한다. 이때 특수상황이란, 즉, 공휴일 및 명절과, 강우 및 유고가 발생한 경우를 지칭한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 연구를 고찰하고, 3장에서는 자료수집 및 분석방법론을 선정한다. 4장에서는 통행패턴 분류 및 생성을 수행하여 타 노선에 대하여 검증하고, 그 정확도를 확인한다.

II. 이론적 배경

1. 통행패턴 정의

사전적 의미에서의 패턴이란 ‘일정한 형태나 양식 또는 유형’을 의미한다. 허민국 외[1]는 교통패턴을 ‘한 지점에서 동일한 요일 및 시간대를 기준으로 시간의 흐름에 따라 나타나는 소통상태 변화의 일정한 형태’로 정의하였고, 송상규[2]는 패턴을 ‘교통데이터가 교통특성에 의해 일정한 반복을 나타내는 것을 의미하는 것’으로 정의하였다. Yoshikazu 외[3]는 통행시간 패턴을 ‘이상치를 발생 유형별로 제거한 후 여러 일자의 유사값을 나타내는 실제 통행시간 시계열 데이터의 평균을 적용한 것’으로 정의하였다. 본 연구에서의 통행패턴은 ‘고속도로 구간별로 동일한 시간대에 반복적으로 발생하는 통행속도 추이’로 정의한다.

2. 통행패턴 분류 방법론 및 사례 연구

1) 통행패턴 분류 방법론

송필용[4]은 차종별, 연도별, 월별, 요일별로 구분한 서해대교 일평균 교통량(ADT, Average Daily Traffic) 패턴을 분석하였고, 계절별, 월별 교통량의 패턴은 차이가 있는 것으로 도출하였다. 또한 요일별 특성 중 화·수·목요일의 교통량은 유사성이 존재함을 주장하였다. 조준한 외[5]는 안산시 3개 구간의 단속류 시계열 과거이력자료를 활용하여 유사 통행패턴을 보이는 요일을 5분단위로 그룹핑하는 비교정식 알고리즘을 개발하였다. 최진우 외[6]는 서울 강남대로의 통행속도 이력자료를 바탕으로 분석을 수행하였으며, 평균 속도와 교통량이 평일과 주말은 아주 상이한 패턴을 보이므로 이력자료를 각각 구축해야 할 필요성이 있음을 주장하였다. Lingras[7]는 264개월의 월평균교통량을 각 개별 패턴으로 정의하고, 월 교통량 패턴을 분류하는 연구를 수행하였으며 그 결과 5개 이상의 군집 분류가 되어야만 집단내 오차가 확연히 감소함을 주장하였다. Yoshikazu 외[3]는 모든 누적 통행시간 패턴과 예측 일자의 통행시간 간 제공오차가 작은 이력 데이터를 선정하는 방식으로 패턴을 분류하였다. 누적 패턴 중 예측을 원하는 시간대 전·후 한 시간 동안의 통행시간을 대상으로 가장 유사한 값을 지니는 데이터를 선정하고, 예측 대상이 되는 구간과 공간적으로 가장 인접한 대안을 패턴으로 선정하였다.

2) 통행패턴 생성 사례

이력자료의 수집은 통상 프로브 차량을 기반으로 위성항법장치, VDS, 비콘 등을 활용하여 수집되며, 최근 전 고속도로 구간에 확장 설치가 완료된 DSRC를 활용한 교통정보 수집 역시 이루어지고 있다. 수집된 원시자료는 각 목적에 맞게 다양한 주기 및 교통특성별 이력자료로 생성된다. 한국도로공사[8]는 VDS 및 DSRC 원시자료를 기반으로 지수평활, ARIMA, 베이지안 알고리즘을 적용하여 공휴일이 제거된 요일, 하계휴가철, 명절 패턴을 주요 5개

도시에 대해 생성한다. 천안-논산 고속도로 교통관리시스템[8]은 화·수·목요일을 평일로 지정하고, (월)(평일)(금)(토)(일)(특수 공휴일)로 교통패턴을 분류 및 생성한다.

3. 의사결정나무 방법론

의사결정나무는 의사결정규칙을 도표화하여 관심대상이 되는 집단을 몇 개의 소집단으로 분류하거나 예측하는 분석방법이다. 의사결정나무 알고리즘은 CHAID, CART, QUEST, C4.5 등으로 분류되며 이들은 분리기준, 정지규칙 및 가지치기 등에서 서로 다른 형성과정을 가지고 있다.

Northwestern Univ.[9]는 ADVANCE 프로젝트에서 교통사고자료를 이용하여 돌발상황 처리시간 모형을 개발하였는데, 독립변수로는 대응차량의 대수와 역할을 이용하였다. 신치현 외[10]는 21개월 동안의 교통사고자료를 이용하여 돌발상황 지속시간 의사결정구조를 제시하였다. 하오근 외[11]는 교통사고를 대상으로 돌발처리시간 예측모형 개발을 통한 주요영향인자를 도출하여 의사결정나무에 적용하는 연구를 수행하였다. 김한수 외[12]는 화물자동차 투어유형 선택 행태를 의사결정나무 및 로짓모형을 적용하여 도출하였다. 강진웅 외[13]는 도로이용자가 경로를 우회할 때 적용하는 주된 우회결정요인을 분석하기 위해 의사결정나무와 신경망 모형을 결합하여 사용하였다. Arentze and Timmermans[14]는 통행행태 예측을 위해 1차로 의사결정나무를 이용하여 데이터를 그룹핑하고, 각 그룹 별 대푯값에 해당하는 통행행태를 분류하는 과정을 수행하였다. Sohn과 Shin[15]는 교통사고 심각도 패턴 인식을 통해 잠재적 사고 발생을 예방하기 위하여 관련 변수를 적용한 사고 심각도 추정 분류 모형을 개발하였다.

4. 시사점

기존 통행패턴 분류 방법론 연구에서는 계절 및 월별 통행특성에 차이가 있음을 제시하였다. 또한 대부분의 연구에서 요일을 패턴 분류의 기준으로

삼는 것을 확인하였다. 기존에는 대부분 평일과 주말을 분류하거나 모든 요일을 개별로 분류한다. 주말의 경우 모든 연구에서 분류 필요성을 주장하였다. 한편 평일의 경우 통상적인 분류 기준이 없기 때문에 실시간 데이터와 패턴간의 오차 분석을 통한 분류를 실시한다. 본 연구에서는 15분 단위 DSRC 통행속도 이력데이터를 기반으로 고속도로 구간별 통행속도 추이를 파악하여 통행패턴을 생성하기로 한다.

기존 연구의 경우 통상적 기준 없이 요일 분류를 수행하였으나, 본 연구에서는 통행패턴과 실시간 데이터 간 오차분석을 통해 요일 분류 기준을 선정하였다. 또한 기존에는 특정 대도시 O/D를 기준으로 패턴을 생성하였지만, 본 연구의 경우 인접 IC-IC 단위별로 통행패턴을 도출하였다는 점에서 차이가 있다.

III. 자료수집 및 분석방법론 선정

1. 활용 자료 수집

본 연구에서는 고속도로 DSRC 데이터를 구간 단위로 분할하여 분석에 활용한다. 여기서 구간 단위는 2개의 가장 인접한 IC간을 의미하며, JC가 존재하는 경우 JC로 대체한다. 정부고속도로의 2011년 데이터를 활용하였으며, 구체적인 시·공간적 범위는 다음과 같다.

- 시간적 범위: 2011년 3월 1일 ~ 2011년 11월 30일 06-22시
- 공간적 범위: 경부고속도로 하행 서울TG ~ 안성IC (총 8개 구간)

공간적 범위는 수도권에 해당되는 서울TG~안성IC은 시간대별로 다양한 소통상황을 나타내므로 상대적으로 소통이 원활한 지방부보다 패턴 분류의 필요성이 높은 것으로 판단되어 선정하였다.

분석 구간의 시계열 속도 추이를 확인한 결과, 야간 시간대(22-06시)의 통행속도가 80~120km/h으로 소통원활 상태를 나타내어 패턴의 분류가 필요없다고 판단하였다. 따라서 시간적 범위는 06-22시

로 선택하였다. <그림 1>은 이상치 제거 후 야간 시간대 통행속도의 시계열 추이이다.

<그림 1> 야간시간대(22-06시) 통행속도 시계열 추이
<Fig. 1> Time series of travel speed at night time(PM10-AM06)

패턴분석에 활용되지 않는 데이터는 특수상황의 속도이다. 강우상황의 경우 기상청으로부터 수집한 강수량 데이터를 활용하였는데, 적설량 데이터는 제공되지 않으므로 눈이 오는 상황이 통행패턴에 미치는 영향을 파악할 수 없다. 따라서 겨울철인 1, 2, 12월은 분석에서 제외하였다. 또한 한국도로공사에서 제공한 유고 데이터를 활용하여 유고상황 역시 제외한 후 패턴을 생성한다. 주말의 경우 기존 연구 내용을 기반으로 명백한 고유의 패턴을 갖는 것으로 판단되어, 본 연구에서는 주말은 분석대상에서 제외하였다.

2. 분석 방법론 선정

통행패턴을 도출하기 위해 15분 단위로 집계된 개별차량 통행속도 평균을 대푯값으로 하는 시계열 데이터를 활용한다. 요일 간 유사성이 검증된 요일은 동일한 그룹으로 분류하며, 통행패턴의 생성은 각 구간별 월(月)단위로 수행한다. 이는 구간의 지역적 통행특성에 따라 다양한 변동이 발생하므로 이를 반영한 대푯값 산출이 필요하기 때문이다. 또한 단순히 요일로만 구분된 패턴을 1년 전체에 적용할 경우 월별, 계절별 및 각종 이벤트에 대한 변동성을 고려할 수 없기 때문에 월 단위로 대푯값을 생성한다. 데이터 가공은 특수상황 발생과 일반상

황의 분류를 한 후에 통행패턴을 생성하게 되는데, 가공절차는 <그림 2>와 같다.

<그림 2> 통행패턴 생성을 위한 데이터 가공절차
<Fig. 2> Process of traffic pattern formation

통행패턴 데이터 가공절차는 1단계로 요일별로 분류된 속도 데이터에서 특수상황 시의 속도데이터를 제거한다. 2차로 특수상황을 제거한 데이터를 대상으로 이상치를 제거한다. 이 때 이상치 제거는 한국도로공사 exTMS의 DSRC 데이터 및 국토 관리 시스템에서 적용되는 2σ 를 신뢰구간 범위 이외 값을 제거하는 방식을 채택하여 수행한다.[16][17] 이상치 제거 단계를 포함시킨 이유는 특수상황을 제외하더라도 결측으로 인한 수집자료 부족, 유고 상황 종료시점 차이 발생 등으로 인한 비정상적 통행속도가 나타났기 때문이다. 특수상황 제거 및 이상치 제거가 완료되면 통행속도 평균을 대푯값으로 하는 15분 단위 요일별 통행패턴을 생성한다.

IV. 통행패턴 분류 및 생성

1. 통행패턴 분류

1) 분석 개요

특수상황 데이터 제거 및 이상치 제거를 통해 가공된 15분 통행속도 데이터를 활용하여 의사결정나무에 적용하기 위한 데이터를 구축한다. 평일을 대상

으로 요일별 평균속도 추이 비교를 통해 패턴을 분류한다. 목표변수인 통행속도 데이터를 1시간 단위로 분할하는 과정을 거친 후 요일을 설명변수로 하는 의사결정나무를 적용하였다. 여기서 1시간 단위로 의사결정나무를 적용한 이유는 침두·비침두 등 시간대에 따라 통행속도 추이가 상이하므로 시간대별 요일간 관계를 도출할 필요가 있기 때문이다.

본 분석에서는 목표변수인 통행속도가 연속형 변수이므로, 분산의 감소량을 이용하여 이진분리를 수행하는 CART(Classification and Regression Trees) 알고리즘을 적용하며, 공식은 다음과 같다.

$$V = \frac{1}{n^{(t)}} \sum_{i=1}^{n^{(t)}} (y_i - \bar{y})^2 \quad (1)$$

이는 마디 t 의 목표변수 평균을 그 마디에 속하는 모든 개체의 예측값으로 사용하여 예측오차를 최소화하는 과정으로 볼 수 있으며, 분산의 감소량은 식 (2)와 같이 정의된다. 분산의 감소량을 최대로 하는 설명변수와 그 변수의 최적 분리를 분리점으로 선택한다.

$$\Delta V = V - \frac{n_L^{(t)}}{n^{(t)}} V_L - \frac{n_R^{(t)}}{n^{(t)}} V_R \quad (2)$$

여기서, ΔV : 분산의 감소량

$n^{(t)}$: 부모마디의 관측치 수

$n_R^{(t)}, n_L^{(t)}$: 자식마디의 관측치 수

2) 통행패턴 분류 결과

동일한 시간대에서 요일간 통행속도 변동이 큰 경우에 의사결정나무 가치를 분리함으로써 유사한 통행속도를 갖는 경우 동일한 그룹으로 분류한다. <표 1>에 침두시간의 분류 결과를 나타내었으며 DAY_FLAG는 요일 변수를 의미한다.

주간 시간대(06-22시)에 각기 적용된 요일 간 통행속도 유사성의 분석 결과 총 16개 시간대 중 15개 시간대에서 금요일을 단독 분류하였다. 월요일

의 경우 7개 시간대에서 단독 분류 되었으며, 이는 오전 첨두시간대를 모두 포함한다. 화·수·목요일은 오전 9시~10시의 1개 시간대를 제외한 전 시간에서 동일한 통행속도 그룹으로 분류되었다. 고속도로 소통상황이 원활한 시간대일수록 각 요일을 동일한 패턴으로 분류하는 경우가 빈번하며, 첨두시간대일수록 요일별 통행속도 차이가 발생한다.

〈표 1〉 첨두시간 통행패턴 분류 결과
〈Table 1〉 Travel pattern classification during peak time period

AM 7-8	AM 8-9
<p>노드 0</p> <p>평균 81.304 표준편차 19.877 n 5390 % 100.0 예측 81.304</p> <p>DAY_FLAG 향상=181.631</p> <p>1.0 3.0; 4.0; 5.0; 2.0</p> <p>노드 1 노드 2</p> <p>노드 1: 평균 54.891, 표준편차 27.048, n 1012, % 18.8, 예측 54.891</p> <p>노드 2: 평균 87.416, 표준편차 10.882, n 4378, % 81.2, 예측 87.416</p> <p>DAY_FLAG 향상=0.736</p> <p>5.0 3.0; 4.0; 2.0</p> <p>노드 3 노드 4</p> <p>노드 3: 평균 89.038, 표준편차 9.697, n 1121, % 20.8, 예측 89.038</p> <p>노드 4: 평균 86.858, 표준편차 11.217, n 3257, % 60.4, 예측 86.858</p> <p>- Group 1: Mon - Group 2: Tue·Wed·Thu - Group 3: Fri</p>	<p>노드 0</p> <p>평균 81.304 표준편차 19.877 n 5390 % 100.0 예측 81.304</p> <p>DAY_FLAG 향상=181.631</p> <p>1.0 3.0; 4.0; 5.0; 2.0</p> <p>노드 1 노드 2</p> <p>노드 1: 평균 54.891, 표준편차 27.048, n 1012, % 18.8, 예측 54.891</p> <p>노드 2: 평균 87.416, 표준편차 10.882, n 4378, % 81.2, 예측 87.416</p> <p>DAY_FLAG 향상=0.736</p> <p>5.0 3.0; 4.0; 2.0</p> <p>노드 3 노드 4</p> <p>노드 3: 평균 89.038, 표준편차 9.697, n 1121, % 20.8, 예측 89.038</p> <p>노드 4: 평균 86.858, 표준편차 11.217, n 3257, % 60.4, 예측 86.858</p> <p>- Group 1: Mon - Group 2: Tue·Wed·Thu - Group 3: Fri</p>
PM 6-7	PM 7-8
<p>노드 0</p> <p>평균 90.772 표준편차 12.367 n 5468 % 100.0 예측 90.772</p> <p>DAY_FLAG 향상=35.926</p> <p>5.0 3.0; 4.0; 1.0; 2.0</p> <p>노드 1 노드 2</p> <p>노드 1: 평균 78.942, 표준편차 19.806, n 1121, % 20.4, 예측 78.942</p> <p>노드 2: 평균 93.809, 표준편차 6.859, n 4397, % 79.6, 예측 93.809</p> <p>- Group 1: Mon·Tue·Wed·Thu - Group 2: Fri</p>	<p>노드 0</p> <p>평균 92.526 표준편차 7.955 n 5436 % 100.0 예측 92.526</p> <p>DAY_FLAG 향상=10.396</p> <p>5.0 3.0; 4.0; 1.0; 2.0</p> <p>노드 1 노드 2</p> <p>노드 1: 평균 86.093, 표준편차 9.577, n 1089, % 20.0, 예측 86.093</p> <p>노드 2: 평균 94.140, 표준편차 6.571, n 4349, % 80.0, 예측 94.140</p> <p>- Group 1: Mon·Tue·Wed·Thu - Group 2: Fri</p>

Explanation of terms (노드: node; 평균: average; 표준편차: standard deviation; 예측: prediction; 향상: improvement)

오전 첨두시간대(07-09시)의 경우 월요일과 금요일의 평균 통행속도가 타 요일에 비해 현저히 낮으며, 이로 인해 다른 요일과 분리된 패턴을 형성하게 된다. 이러한 결과를 통해 월요일, 금요일은 명백히 다른 평일과 분리된 통행패턴을 개별적으로 생성할 필요가 있음을 도출할 수 있다. 한편 화·수·목요일은 1개 시간대를 제외하고 항상 유사한 속도 추이를 갖는 것으로 분류되었으므로 이는 동일한 패턴 그룹으로 분류할 수 있음을 의미한다. 따라서 경부고속도로 서울TG~안성IC 구간 대상 패턴 분류 결과, (월)(화·수·목)(금) 3개의 평일 패턴으로 분류되었다. 각 노선 및 구간은 월별·요일별·시간대별로 고유의 패턴을 갖는데, <그림 3>은 경부고속도로 하행 수원IC~기흥IC 구간의 월별 통행패턴 예이다.

〈그림 3〉 경부(수원IC~기흥IC) 요일별 통행패턴 비교
〈Fig. 3〉 Travel Patterns classified by day of week at Suwon IC~Gihung IC on Gyungbu

요일별 통행패턴을 비교한 결과 금요일을 제외한 평일은 명확한 월별 패턴이 나타났지만, 금요일은 월별 패턴의 변동이 다양하게 나타났다. 그리고 8월과 11월에서 가장 낮은 속도 패턴을 나타냈는데 이는 여름철 휴가철과 가을 단풍철 등의 단기성 이벤트가 원인인 것으로 예상된다.

2. 통행패턴 정확성 분석

의사결정나무 모형을 통해 도출된 통행패턴과 개별 일자 데이터의 오차를 이용하여 통행패턴의 정확성을 확인한다. 이 때 개별 일자 속도 데이터를 참값으로 가정하였으며, 정확성 지표는 식 (3)과 같이 MAPE(Mean Absolute Percentage Error, 절대 백분율 오차의 평균)를 적용하였다. MAPE는 MAE(Mean Absolute Error, 오차의 절대값 평균)를 백분율로 나타낸 값으로써 시계열 데이터의 정확성 측정을 위해 주로 사용되는 통계적 수치이다.

$$MAPE(\%) = \frac{\sum \frac{|V_{indiv} - V_{Base\ pattern}|}{V_{indiv}} \times 100}{n} \quad (3)$$

여기서, V_{indiv} : 개별 일자 통행 속도

$V_{Base\ pattern}$: 통행패턴 대푯값

n : 개별 일자 통행 속도 표본 수

주 분석 구간인 경부고속도로 서울TG~안성IC 하행 구간의 통행패턴의 MAPE를 산출한 결과는 <표 2>에 제시되어 있다.

<표 2> 통행패턴 MAPE(서울TG~안성IC)
<Table 2> MAPE of travel pattern
(Seoul TG~Ansung IC downstream)

	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.
Mon	4.01	4.66	5.88	5.91	7.00	5.02	3.55	3.46	3.57
Tue: Wed ·Thu	3.39	5.60	7.35	7.14	6.53	6.77	6.79	4.13	5.24
Fri	3.64	8.18	9.25	10.24	12.42	8.69	8.81	6.54	6.59
Sat	20.45	26.63	20.58	30.86	35.38	44.46	16.23	24.46	15.23
Sun	2.59	4.00	5.65	3.65	9.06	6.92	2.97	5.13	5.98

월별 MAPE 중에서는 3월이 가장 낮고, 여름 휴가철인 8월에 가장 높은 MAPE를 나타낸다. 요일을 기준으로 보았을 때는 토요일 MAPE가 타요일에 비해 2배 이상 높은 값을 나타낸다. 이는 토요일에 발생하는 고속도로 소통상황의 변동이 가장 크고 지·정체가 빈번히 발생함을 유추할 수 있다.

3. 통행패턴 분류 검증

경부고속도로 하행 서울TG~안성IC 구간의 속도 데이터를 활용하여 분류된 패턴 분류 결과가 타 노선에도 적용되는지 확인하기 위하여, 샘플 노선을 선정 후 검증을 수행하였다. 검증은 영동(양지IC~호법JC), 중부(호법JC~대소JC), 중부내륙(여주JC~충주IC)로 총 9개 구간에 대하여 수행하였고 상세 구간명은 <표 3>과 같다.

<표 3> 통행패턴 분류 검증 구간
<Table 3> Sections of travel pattern verification

Sorting	Highway	Section	Length (km)	time
Default	Gyung-bu	SeoulTG~AnsungIC	44.0	06 - 22
Verified Sections	Yeong-dong	YangjiIC-DukpyungIC	10.5	
		DukpyungIC-HobupJC	2.3	
		HobupJC-YichunIC	6.3	
		HobupJC-IljukIC	16.7	
	Jung-bu	IljukIC-DaesoIC	12.9	
		DaesoIC-DaesoJC	4.9	
		YeojuJC-GamgokIC	14.5	
	Jungbu -Neruk	GamgokIC-ChungjuJC	17.3	
		ChungjuJC-BukChungjuIC	9.2	

요일 기준 통행패턴 분류 검증에는 통행속도의 분산을 고려하면서 집단 간 평균의 차이를 검증하는 통계기법인 일원분산분석(one-way Analysis Of Variance, ANOVA)을 적용하였다. 해당 시간대 내에 요일별 속도의 평균 차이 존재 유·무를 검증하기 위해 등분산을 가정한 Scheffe 사후검정법을 적용하였다. 시간대별로 분할된 동일 요일의 15분 주기 평균통행속도에 대하여 유의확률 0.05에서 수행하였으며, 가설은 다음과 같다.

귀무가설(H_0): 요일간 통행속도에는 차이가 없다
대립가설(H_1): 요일간 통행속도에는 차이가 있다.

검증 결과 약 94.4%의 케이스에서 기존 경부고속도로 분류 결과와 동일 패턴 그룹으로 분류됨을 확인하였으며, 이는 <표 4>와 같다.

〈표 4〉 통행패턴 분류 검증 결과
〈Table 4〉 Result of travel pattern verification

Highway route	number of trial	equal(%)	unequal(%)
Young-dong	48	100	0
Jung-bu	48	95	5
Jungbu-neruk	48	88	12
Total	144	94.4	5.6

상대적으로 소통상황이 원활한 검증 시간대에는 전체 평일이 1개 패턴으로 분류되는 비율이 더욱 높게 나타났다. 그러나 검증 노선에서도 마찬가지로 첨두시간에 월, 금요일의 경우 고유의 패턴을 갖는 것으로 나타났으므로, 기존 경부 고속도로 분류대로 월요일과 금요일의 개별 패턴 생성이 필요함을 확인하였다.

4. 의사결정나무를 통한 추가변수 도출

1) 분석 개요

앞절에서 도출된 통행패턴은 월, 요일, 시간대, 구간을 변수로 적용하였다. 그러나 실제로는 이러한 시·공간적 요인뿐만 아니라 매우 다양한 요인들이 교통류에 영향을 미치게 된다. 본 연구에서는 이러한 요인들에 따른 오차를 감소시키는 목적으로 추가변수를 적용한 모형을 검토하였다. 추가변수 도출은 요일별 패턴분류를 토대로 통행패턴 생성 최종 단계에서 새로운 외부 변수를 추가하여 의사결정나무 모형을 정립하여 이를 평가하는 형태로 진행되었다.

2) 추가변수 선정

본 연구에서는 추가변수를 고려하는 과정에서, 고속도로의 시·공간적 특성 외에 통상적으로 교통류 및 교통패턴에 영향을 미칠 것으로 예상되는 요소를 검토하여 추가적인 변수로 선정하였다. 이 때 고려된 추가변수는 패턴 생성 시점 이전 월의 교통상황을 나타내는 ‘직전월 소통상황’, 해당 구간의 ‘강수량’, ‘TCS 교통량’, 1·3·5주/2·4주를 분류한 ‘홀·짝수 주 통행속도’이다.

고려된 변수들에 대하여 독립변수 중요도에 대한 분석을 실시하였고, 이를 토대로 통행패턴 분류에 영향을 미치는 추가변수의 우선순위를 파악하였다. 해당 내용의 결과는 <표 5>와 같다. 고려된 변수중 ‘직전월 소통상황’ 변수가 가장 높은 정규화 중요도를 나타내었고, 다른 변수들은 중요도가 1.3% 이하로 나타나 결과에 미치는 영향이 매우 낮은 것으로 분석되었다.

〈표 5〉 추가 독립변수 중요도
〈Table 5〉 Importance of additional independent variables

Independent variable	Importance	Normalized importance
Traffic condition on previous month	256.513	100.0%
Precipitation	3.385	1.3%
Traffic volume	0.000	0.0%
biweekly travel speed	0.000	0.0%

3) 추가변수 적용 시 위험도 비교

‘직전월 소통상황’을 추가변수로 적용하여 의사결정나무를 생성하였을 때, 위험도가 변화하는 정도를 비교하여 통행패턴 오차 개선에 실제로 영향력이 있을지 판단하였다. 본 분석의 목표변수는 연속형으로써 위험도는 모형에서 분류된 노드의 분산가중합을 의미한다. 설명변수를 다르게 적용할 경우 각각의 모형에서 나타나는 위험도 비교를 통해 동일 그룹(노드, node) 내에서의 분산을 비교함으로써 추가 설명변수가 분류의 정확성을 높이는 지 확인할 수 있다.

〈표 6〉 추가 변수 적용 시 모형 위험도 비교
〈Table 6〉 Importance of additional independent variables

Condition	Risk Estimate	SE of Risk estimate
default	367.734	1.903
including additional variables	189.519	1.319

<표 6>과 같이 추가변수를 적용하지 않은 경우의 위험도 추정값은 약 368로 나타났으나, ‘직전월

소통상황'변수를 추가하는 경우 추정값은 약 190으로 동일 그룹 내 분산이 감소함을 확인하였다.

전체적인 오분류율은 약 50% 감소하였으며, 이에 대한 표준오차는 각각 1.9, 1.3이다.

V. 결론 및 향후 연구과제

본 연구에서는 고속도로 구간 정보의 수집이 가능한 DSRC 기반 통행속도 이력데이터를 활용하여 IC-IC 구간 단위의 통행패턴을 도출하는 방안을 연구하였다. 주요 분석 구간인 경부고속도로 수도권 구간의 DSRC 통행속도 데이터를 의사결정나무 모형에 적용한 결과 요일을 기준으로 (월)(화·수·목)(금)(토)(일) 5개 그룹으로 통행패턴이 분류되었다. 뿐만 아니라 타 노선에의 적용 가능 여부에 대한 일원분산분석 통계검증 결과, 약 94%에서 동일한 패턴으로 분류되는 것으로 나타났다. 의사결정나무 기법을 통하여 통행패턴 분류 오차 개선을 위해 추가 설명변수를 확인한 결과, '직전월 소통상황' 변수가 가장 큰 영향력을 갖는 것으로 나타났다. '직전월 소통상황' 변수를 추가하는 경우 모형에서의 통행속도 분산이 약 50% 감소하였고, 표준오차는 0.58 감소하였다.

DSRC 데이터가 수집되지 않던 과거에는 약 300개에 달하는 O/D의 통행특성 도출이 매우 시간 소모적인 작업이었다. 그러나 본 연구에서 제시한 구간 단위의 통행 패턴이 고속도로 전 구간에 구축되면 구간별 통행패턴 조합을 통해 모든 O/D에 대한 통행패턴의 산출 및 수준 높은 교통정보의 제공이 가능하게 될 것으로 기대된다.

한편 본 연구에서는 수집된 이력데이터 기간의 제약으로 특수상황에서의 패턴 분류가 제외되었고, 주말의 경우 기존 연구 내용을 인용하여 본 연구에서는 제외하였다. 향후에는 장기적인 데이터를 수집하여 해당 부분에 대한 패턴 분류를 실시하고, 이를 본 연구에서 도출된 패턴과의 내용적 관계 정립이 필요하다. 더불어 시간당 강수량, 적설량, 안개 등의 구체적인 기상 상황, 구간별 인근 지역의 콘서트, 운동경기, 지역 축제와 같은 이벤트 발생여부,

차종별 교통량 등과 같은 데이터를 추가 설명변수로 활용하면 더욱 정확성 높은 통행패턴을 구축할 수 있을 것으로 판단된다.

REFERENCES

- [1] M. K. Huh, C. No, W. Kim, and B. S. Son, "Development of an incident detection algorithm by using traffic flow pattern," *Journal of Korean Society of Transportation*, vol. 28, pp.7-15, Dec. 2010.
- [2] S. G. Song, *Development of Long Distance Expressway Travel Time Prediction Models during Holiday Seasons*, Graduate School of Ajou Univ., Suwon, 2010.
- [3] Y. Ohba, "Travel time information calculation method for expressway using toll collection system data," in *Proc. ITS int. Conf. Commun. 1999*, pp.476-477, Tokyo, Japan, Oct. 1999.
- [4] P. Y. Song, *A Study on Structural Behavior Analysis and Maintenance Models of Long-span Bridges Considering Traffic Patterns*, Graduate School of Ajou Univ., Suwon, 2011.
- [5] J. H. Cho, B. S. Kim, S. H. Kim, and W. E. Kang, "Development of a daily pattern clustering algorithm using historical profiles", *Journal of the Korea Institute of ITS*, vol. 10, pp.11-23, Aug. 2011.
- [6] J. W. Choi and Y. G. Yang, "An efficient filtering technique of GPS traffic data using historical data," *Journal of Korea Spatial Information Society*, vol. 10, pp.55-65, Sep. 2008.
- [7] P. Lingras, "Statistical and genetic algorithms classification of highways," *Journal of Transportation Engineering*, vol. 127, No.3, pp. 237-243, June, 2001
- [8] KECRI, *Improvement of Traffic Data Collection, Analysis and Utilization from Vehicle Detection System on Freeway*, 2006

- [9] K. Ozbay and P. Kachroo, *Incident Management in Intelligent Transportation Systems*, Boston, MA: Artech House, 1999
- [10] C. H. Shin and J. H. Kim, "Development of free-way incident duration prediction models," *Journal of Korean Society of Transportation*, vol. 20, pp.17-30, Jun. 2002.
- [11] O. K. Ha, D. J. Park, J. Won, and C. Jung, "The prediction models for clearance times for the unexpected incidences according to traffic accident classifications in highway", *Journal of the Korea Institute of ITS*, vol. 9, pp.101-110, Feb. 2010.
- [12] H. S. Kim, D. J. Park, C. Kim, C. Choi, and K. Kim, "An analysis of choice behavior for tour type of commercial vehicle using decision tree," *Journal of Korean Society of Transportation*, vol. 28, pp.43-54, Dec. 2010.
- [13] J. W. Kang, K. J. Kum, and S. N. Son, "Drivers detour decision factor analysis with combined method of decision tree and neural network algorithm," *Journal of the Korean Society of Road Engineers*, vol.13, pp. 167-176, Sep. 2011.
- [14] T. A. Arentze and H. J. P. Timmermans, *Incorporating parametric action decision trees in computational process models of activity-travel behavior: Theory and illustration*, Advanced OR and AI Methods in Transportation, pp. 567-572, 2005.
- [15] S. Y. Sohn and H. W. Shin, "Pattern recognition for road traffic accident severity in Korea", *Journal of Ergonomics*, Vol 44, pp.107-117, 2001.
- [16] KECRI, *Core Technique Development and Development Plan Report of Hi-pass Transportation Information Systems*, 2011.
- [17] J. K. Kang, "Regional traffic information acquisition by non-intrusive automatic vehicle identification", *Journal of Korea Institute of ITS*, vol. 1, pp.22-32, Oct. 2002.

저자소개

이 민 하 (Lee, Minha)

2013년 : 아주대학교 공학석사 (교통공학 전공)

2013년 2월 ~ 현재 : 아주대학교 TOD기반 지속가능 도시·교통 연구센터(ERC) 연구원

이 상 수 (Lee, Sang-Soo)

2000년 : Texas A&M University 박사 (교통전공)

2001년 ~ 2002년 : 한국건설기술연구원 선임연구원

2002년 ~ 현재 : 아주대학교 교통공학과 교수

남 궁 성 (Namkoong, Seong)

1996년 : 한양대학교 도시공학과 박사 (교통전공)

2002년 9월 ~ 2004년 10월 : University of Virginia, Smart Travel Lab. Research Scientist

2011년 8월 ~ 2012년 12월 : 한국도로공사 교통정보통합활용지원센터(OASIS) 센터장

2008년 1월 ~ 2013년 12월 : 한국도로공사 명절 교통예보관

2013년 1월 ~ 현재 : 한국도로공사 교통연구실장

최 기 주 (Choi, Keechoo)

1992년 : University of Illinois(at Urbana-Champaign) 공학박사 (교통계획)

1992년 ~ 1994년 : 서울시정개발연구원 도시교통연구부 책임연구원

1994년 ~ 현재 : 아주대학교 교통공학과 교수

2010년 ~ 현재 : 아주대학교 TOD기반 지속가능 도시·교통 연구센터(ERC) 소장