

밀도기반 비집계 접근법을 이용한 구간통행시간 추정 방법론

Methodology for Estimation of Link Travel Time using Density-based Disaggregated Approach

장 현 호* · 이 승 봉** · 한 등 희*** · 이 영 인****

* 주저자 : 서울대학교 환경대학원 환경계획학과 연구교수
 ** 교신저자 : 서울대학교 환경대학원 환경계획학과 박사수료
 *** 공저자 : 한국도로공사 도로교통연구원 책임연구원
 **** 공저자 : 서울대학교 환경대학원 환경계획학과 교수

Hyunho Chang* · Soong-bong Lee** · Donghee Han*** · Young-Ihn Lee****

* Dept. of Environmental Planning, Graduate School of Environmental Studies, Seoul National University
 ** Dept. of Environmental Planning, Graduate School of Environmental Studies, Seoul National University
 *** Expressway and Transportation Research Institute, Korea Expressway Corporation
 **** Dept. of Environmental Planning, Graduate School of Environmental Studies, Seoul National University
 † Corresponding author : Soong-bong Lee, habanera82@hanmail.net

Vol.16 No.5(2017)

October, 2017

pp.134~143

ISSN 1738-0774(Print)

ISSN 2384-1729(On-line)

<https://doi.org/10.12815/kits.2017.16.5.134>

2017.16.5.134

Received 25 August 2017

Revised 27 September 2017

Accepted 9 October 2017

© 2017. The Korea Institute of Intelligent Transport Systems. All rights reserved.

요 약

고속도로의 경우 해당구간에 버스전용차로, 휴게소, 졸음쉼터 등이 존재할 경우 다수의 통행시간 그룹이 존재할 수 있다. 기존 대부분의 구간통행시간 추정연구에서는 낮은 표본 수집 상태에서 하나의 대표 통행시간(정규분포 가정) 그룹을 가정하고, 특정범위를 벗어난 경우 이상치로 판단하여 제거한 후 구간 통행시간을 추정하였다. 하지만, 고속도로와 같이 해당구간에 버스전용차로, 휴게소, 졸음쉼터 등이 존재할 경우 통행시간 분포는 정규분포가 아닌 쌍봉 또는 다봉 형태를 보일 것이다. 따라서 기존의 추정방법론을 적용할 경우 왜곡된 결과를 초래할 가능성이 있다. 이러한 문제를 해결하기 위하여 첫째, 샘플수가 부족한 상태에서도 신뢰할 수 있으며, 둘째, 다수의 통행시간 그룹 중 일반차로를 이용하면서 휴게소를 이용하지 않은 대표 통행시간 그룹을 선정하고, 선정된 통행시간 그룹의 개별 통행시간 자료를 이용하여 대표 통행시간을 추정할 수 있는 방법론을 제안하였다.

핵심어 : 개별차량 통행자료, 구간통행시간 추정, 대표그룹 중심 탐색법, 밀도기반 비집계 접근법

ABSTRACT

In the case of highway, there may be a large number of travel time groups when there are a bus exclusive lane, a rest area, a sleeping shelter, etc. in the corresponding section. In most of the conventional travel time estimation studies, one representative travel time (assuming normal distribution) group is assumed in the low sample collection state, and if it is out of the specified range, it is determined as outliers and then the travel time is estimated. However, if there is a bus exclusive lane, a rest area, or a sleeping shelter in the relevant section, such as the highway, the distribution of travel time will be in the form of a bi-modal or a multi-modal, rather than a regular distribution. Therefore, applying the existing estimation methodology may result in distorted results. To solve this problem, first, it should be reliable even in the case of insufficient number of samples. Second, we propose a methodology to select the representative time group among a number of time groups and to estimate the representative time using individual time data of the selected time group.

Key words : Individual Travel Data, Link Travel Time Estimation, Exploring Representative Group and Medianoid, Density-based Disaggregated Approach

I. 서론

20세기 말 국내에 ITS(Intelligent Transportation System)가 도입된 이후 교통운영의 효율적인 운영을 위해서는 신뢰성 있는 교통정보를 수집하고 제공하는 것이 매우 중요하다. 현재 고속도로에서는 2015년 이후 하이패스 단말기와 본선 RSE(Road Side Equipment)를 통해서 하루에 약 백만대의 차량주행계적 자료가 수집되고, 누적 저장되고 있다. 기존의 지점 교통정보 수집체계는 특정 지점에 대한 전수화된 교통자료 수집의 큰 의의와 활용도를 갖고 있으나 집계자료 형태로 구간의 통행행태를 파악하기 어려우며, 구간을 대표하는 자료로 활용하기에는 한계가 있다.

본 연구에서는 전국 고속도로를 대상으로 신뢰성 있는 구간 대표통행시간을 산정하기 위해 DSRC(Dedicated short range communication)자료를 활용하고자 한다. DSRC 자료는 3세대 교통자료로 분류되는 경로형 교통자료로서 개별 고속도로 이용자의 연속된 시간과 공간의 연결성을 확인할 수 있으며, 정밀한 통행행태 분석과 교통관리전략 수립에 큰 활용도를 가지고 있다. 하지만, 고속도로의 경우 해당구간에 버스전용차로, 휴게소, 졸음쉼터 등이 존재할 경우 다수의 통행시간 그룹이 존재할 수 있다. 이럴 경우 통행시간 분포는 정규분포가 아닌 쌍봉 또는 다봉 형태를 보일 것이다. 따라서 정규분포를 가정한 기존의 추정방법론을 적용할 경우 왜곡된 결과를 초래할 가능성이 있으므로, 이에 대한 고려가 필요하다. 따라서 본 연구에서는 구간 검지체계에서 기존의 구간 대표통행시간 산정방법론이 갖고 있는 한계점을 살펴보고, 이러한 문제를 개선할 수 있는 방법론을 개발하여 신뢰성 있는 구간 대표통행시간을 산정하고자 한다.

II. 기존연구 검토

1. 통행시간 추정방법론 고찰

구간검지체계에서 신뢰성 높은 구간통행시간을 산정하기 위해서는 검지기에서 수집되는 자료의 정확성과 더불어 신뢰성 높은 대푯값을 추정하는 것이 가장 중요한 문제이다. 해당구간의 링크통행시간 대푯값을 추정하는 문제는 이상치 제거과정과 대푯값 산정으로 구성된다.

이상치(이상점, outlier)란 관측된 데이터 범위에서 비정상적으로 극단값을 갖는 경우나, 비현실적인 값들을 뜻한다. 즉, 극단적인 값을 이상치라고 한다. 어떤 의사결정을 하는데 필요한 데이터를 분석할 경우 이상치가 존재할 경우 의사결정에 영향을 미칠 수 있으므로 제거하는 것이 중요하다. 일반적으로 대푯값을 산정할 때 평균값이 주로 사용된다. 여기서 말하는 평균은 산술평균으로 자료 하나의 영향력은 동일하게 작용하므로, 소수의 크고 작은 이상치가 존재할 경우 크게 영향을 받기 때문에 이상치 자료 제거과정은 매우 중요하다. 이상치 제거과정은 전통적으로 통계적인 방법이 주로 사용되었다.

Box Plot방법은 순서화된 자료에서 4분위수를 구하여 F_L 을 제1사분위수, F_U 를 제 3사분위수라고 하며, 구간($F_L - k \times (F_U - F_L)$), ($F_U + k \times (F_U - F_L)$)의 바깥쪽에 떨어지는 관측치를 이상치로 간주한다. Box Plot방법이 관측치를 분류하는 가장 효율적인 방법은 아니지만, 이상치가 분포의 양쪽 끝에 존재할 경우에 효과적이다.

첨도 통계량(Kurtosis Statistics) 방법은 정규성으로부터 벗어난 측도와 관측치를 검정하는데 사용되며, 계산과 적용이 간편하다. 이 방법의 단점은 인접한 이상치가 존재할 때 가려진다는 것이다.

중위절대편차(Median absolute deviation, 이하 MAD)는 표준편차 대신에 MAD를 이용하여 이상치를 제거하

므로 자료의 분포를 가정하지 않아도 된다. 반면 MAD기법은 자료의 표본수가 낮거나 분산이 큰 경우 대푯값을 왜곡하는 문제가 발생한다. MAD기법은 다음과 같으며, 여기서 tt_i 는 i 번째 통행시간 자료(tt), tt_{med} 는 tt 계열의 중위값, z_i^{MAD} 는 MAD에 의한 표준점수(Z score), 1.4826은 MAD를 정규분포의 표준편차와 동일하게 만드는 계수이다.

$$MAD = 1.4826 \times \text{median}|tt_i - tt_{med}| \tag{식(1)}$$

$$z_i^{MAD} = \frac{tt_i - tt_{med}}{MAD} \tag{식(2)}$$

Transmit 알고리즘은 이상치를 제거하지 않고 15분 수집주기 평활화 기법을 이용하여 평균통행시간을 추정하는 방법이다. 해당 시간대(k) 동안 구간을 이용하는 프로브 차량(n_k)의 시점부(t_{Ai}), 종점부(t_{Bi})의 검지시각 자료를 이용하여 평균 구간통행시간(tt_{AB}) 산정 후, 과거의(동요일, 동시간대) 패턴자료를 계수값(α)을 이용해 평활화한 다음 이전주기 동안 평활화된 통행시간을 이용하여 현재주기의 평활화된 대표 구간통행시간을 추정한다. Transmit 알고리즘은 구간통행시간의 비반복적 특성이 강하고 대표 구간통행시간과 차이가 큰 이상치가 발생하는 경우 평활화 기법만으로 통행시간을 산출하기에는 한계가 있다.

$$tt_{ABk} = \frac{\sum_{i=1}^{n_k} (t_{Bi} - t_{Ai})}{n_k} \tag{식(3)}$$

$$tth''_{ABk} = \alpha \times tth_{ABk} + (1 - \alpha) \times tth''_{ABk-1} \tag{식(4)}$$

TransGuide알고리즘은 Dion and Rakha(2003)에 의해 개발되었으며, 식(5)를 이용하여 사전에 입력된 경계값을 초과하는 경우 이상치를 제거한 후 유효 표본(Stt_{ABi})을 수집한 후 이동평균 기법인 식(6)을 이용하여 구간통행시간을 추정한다. TransGuide 알고리즘은 l_{th} 가 상수값(0.2)으로 설정되어 있어 혼잡의 상태나 구간거리 에 따른 통행시간의 변동량을 고려할 수 없다. 즉, 구간통행시간이 급격히 증가하는 경우 이를 이상치로 판단하게 되며, 최초 수집된 표본에 편향이 발생할 경우 정상치가 제거되고 이상치가 대푯값이 될 수 있기 때문에 실제 구간통행시간을 왜곡시킬 수 있다. 여기서, t_w 는 이동평균의 시간대 길이, tt_{ABi} 는 A-B 구간에서 수집된 유효표본의 평균통행시간이다.

$$Stt_{ABi} = \left\{ \begin{array}{l} t_{Bi} - t_{Ai} | t - t_w < t_{Bi} < t \text{ and} \\ t_{Ai} - t_{Bi} | t - t_w < t_{Ai} < t \end{array} \right\} \tag{식(5)}$$

$$tt_{ABi} = \frac{\sum_{i=1}^{|Stt_{ABi}|} (t_{Bi} - t_{Ai})}{|Stt_{ABi}|} \tag{식(6)}$$

Dion and Rakha(2006)는 TransGuide알고리즘과 Transmit 알고리즘의 단점을 보완한 알고리즘을 제시하였다. 개선된 알고리즘은 자유 교통류 상태에서 통행시간은 정규분포를 따르고, 혼잡교통류 상태에서 log를 취한 통행시간은 정규분포를 따른다는 사실을 이용한다. $t-1$ 시간대까지 평활화된 통행시간과 평활화된 표준편차를 이용하여 t 시간대의 유효 구간통행시간 표본을 결정하고, 산술평균하여 대표 구간통행시간을 추정한다. 그러나 자유교통류 상태에서 이질적 구간통행시간을 설명하는데 한계가 있으며, 동시 교통류 상태에서 구간

통행시간의 변동을 고려하지 않는다. 따라서 기존의 TransGuide 알고리즘과 Transmit 알고리즘의 성능은 개선되었으나, 앞에서 언급한 두 알고리즘의 근본적 문제점을 개선하지는 못하였다.

Korea Expressway Corporation(2008)는 고속도로 DSRC 교통정보시스템에서 이상치를 제거하는 방법론으로 수집주기 내 통행시간의 변동계수(Coefficient of Variation)범위에 따른 상·하한 제거비율을 제시하였다.

Kim et al.(2015)는 DSRC자료를 활용하여, 적정 수집주기 데이터와 해당시간대의 과거데이터인 패턴데이터를 퓨전하여 가공하는 방법론을 제시하였다. 데이터 융합비율은 실시간 데이터 20%, 패턴데이터 80% 비율 적용 시 실제 수집데이터와 추정데이터 오차값이 최소인 것으로 분석되었다.

Jeong et al.(2013)는 통행시간의 이상치 자료를 제거하기 위하여 중위절대편차법과 Voting Rule을 기반으로 하는 복합 필터링 모형을 개발하였다. 본 모형은 복합필터링 모형이 이상치 표본을 선택적으로 제거하여, 통행시간 추정의 정확도 개선에 기여할 수 있음을 확인하였다.

Rim et al.(2009)는 개별차량 주행정보를 이용하여 돌발상황 시 신뢰성 있는 통행시간 예측을 위해 교통류의 특성에 따라 링크를 구분하여 개별적인 통행시간을 산출하는 기법이며, 개별차량의 특성을 반영하기 위해 일정 속도단위로 차량분류군을 구분하여 통행시간을 산출하는 방법을 제시하였다.

Choi and Choi(2008)는 Taxi GPS 자료를 이용하여 구간통행시간을 추정하였고, 승객의 승차차시간 등 주행과 관계없는 불필요한 데이터를 실시간으로 검지하여 제거하는 휴리스틱한 이상치 제거 알고리즘을 개발하였다. 평가를 위해 서울시 간선축에서 번호판조사를 실시하여 비교하였으며, 이상치 제거 알고리즘을 적용한 결과 약 70% 이상치가 제거되었으며, 실제통행시간과의 상대 오차가 73.7% 향상된 것으로 나타났다.

2. 기존방법론의 문제점

기존 통행시간 추정과 관련된 방법론을 고찰한 결과는 다음과 같다.

첫째, 기존 통행시간 추정방법론은 대표 통행시간 그룹을 제외한 나머지는 이상치로 간주하였다. 일반적으로 통행시간 분포는 정규분포를 가정하며, 평균 또는 중위값을 중심으로 신뢰구간을 선정하여 특정범위를 벗어난 경우 이상치로 판단하여 제거한 후 대표 통행시간을 추정하였다. 하지만, 실제 통행시간 분포를 살펴보면 다수의 통행시간 그룹이 존재하였다. 고속도로의 경우에도 해당구간에 버스전용차로, 일반차로, 휴게소/졸음쉼터 등이 존재할 경우 단일봉(Uni-modal)의 형태가 아닌 쌍봉(Bi-modal) 혹은 다봉(Multi-modal)의 형태를 보일 수 있다. 따라서 기존의 통행시간 추정방법론의 1차 단계인 이상치 제거에 대한 접근법은 실제 적용에 있어 불합리한 결과를 초래할 수 있다. 이는 수집되는 통행시간의 분포가 다수의 그룹으로 구성됨으로 정규분포가 아닌 경우가 대다수 발생하기 때문이다.

둘째, 대부분의 연구에서 이상치 제거를 위해 Box-plot, 첨도통계량 등의 통계적 방법을 적용하였다. 이러한 통계적 방법은 정규분포를 가정하기 때문에 분포의 중심에서 벗어난 자료는 이상치로 간주한다. 그러나 고속도로의 구간통행시간은 좌편향과 과분산의 특성을 보이는 경우가 많다.

셋째, Transmit 알고리즘의 평활화 계수와 TransGuide 알고리즘의 이상치 제거 파라미터는 상수로서 현행 시간대에서 교통류 상태를 고려할 수 없다. 즉, 구간통행시간의 변동량에 대한 고려가 미흡하기 때문에 통행시간이 급격히 변화하는 경우 긴 상태회복(State recovery) 시간을 피할 수 없다.

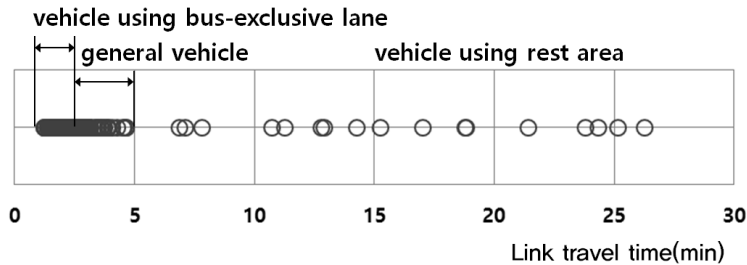
3. 해결방안 도출

기존 구간통행시간 추정 방법론의 문제점은 낮은 표본 수집 상태에서 하나의 대표 통행시간 그룹을 가정

하고, 나머지 통행시간은 이상치로 간주하여 제거한 후 통행시간을 산정하였다. 그러나, 최근의 발달된 구간 검지기기술과 높은 표본수는 기존 통행시간 추정기법과 다른 방법론이 필요함을 의미한다.

<Fig. 1>은 2015년 10월 17일(토요일) 경부선 천안JC~천안IC 구간에 개별 프로브 구간통행시간 분포를 보여주고 있다. 5분 수집기간 동안 207건/5분의 샘플이 수집되었으며, 기존의 낮은 샘플수 문제는 해결되었지만, 해당구간의 특성(버스전용차로, 휴게소)들이 반영되어 구간통행시간 분포가 명확한 그룹의 형태로 분리되었다. 분포의 특징을 살펴보면, 평균 구간통행시간(μ)은 3.31분, 중위값(m)은 2.40분, 표준편차(σ)는 3.87분을 보이는 것으로 분석되었다. 이상치 제거를 위한 상·하한 기준값을 $\mu \pm \sigma$ 로 설정할 경우 통행시간 산정을 위한 유효범위는 -0.56 ~ 7.18분으로 버스전용차로 이용차량과 휴게소 이용차량을 포함하여 대표 통행시간을 추정하게 된다. 따라서 휴게소 이용그룹이 없는 경우는 과소추정, 반대의 경우는 과대 추정되는 문제가 발생한다. 기존의 방법으로 구간통행시간을 산정할 경우 엄밀히 말해 대표통행시간이라 말하기 어렵다.

과대/과소 추정문제를 해결하기 위해서는 우선적으로 버스전용차로, 휴게소를 이용하지 않은 일반차로 이용차량을 대표그룹으로 선정한 후 대표통행시간을 산정할 필요가 있다. 대표그룹의 선정을 위해서는 군집분석(Clustering) 방법 등을 이용하여 결정할 수 있을 것이다.



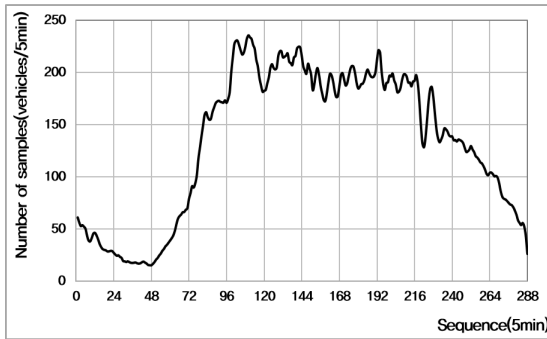
<Fig. 1> Distribution of mixed link travel time for all vehicles

Ⅲ. 통행시간 추정방법론 개발

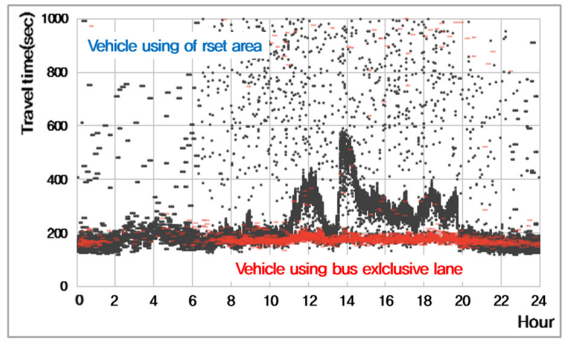
1. 분석자료 검토

본 연구에서는 구간통행시간 산정을 위하여 하이패스 단말기와 노변의 RSE검지기와의 통신으로 수집되는 DSRC 자료를 사용하였다. 신뢰성 있는 구간통행시간 추정을 위해서는 충분한 샘플수 확보가 필수적이다. <Fig. 2>는 경부고속도로 천안JC~천안IC 구간에서 수집된 5분 단위 샘플수(전차종)를 나타내고 있다. 새벽시간대를 제외하고 통행시간 추정을 위한 샘플수는 충분한 것으로 분석되었다. 하지만, 교통량이 적은 노선, 심야시간대, 차종별 분석 시 등 충분한 샘플이 확보되지 않는 경우는 빈번하게 발생할 것이다.

경부고속도로 천안JC→천안IC 구간을 이용하는 차량의 통행시간분포(토요일)는 <Fig. 3>과 같다. 그림에서 검정색은 승용차, 빨간색은 버스를 나타내며, 통행시간 분포는 일반차로 이용차량, 휴게소이용차량, 버스전용차로 이용차량으로 분류되는 것을 볼 수 있다.



〈Fig. 2〉 Temporal variation of samples



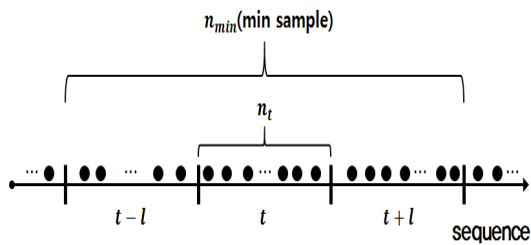
〈Fig. 3〉 Temporal variation of mixed link travel time

2. 구간통행시간 추정방법론 개발

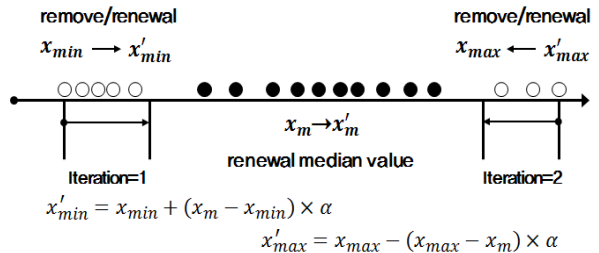
본 연구에서 개발한 구간통행시간 추정모형은 충분한 샘플수가 확보된 상태에서 대표 통행시간 추정을 위한 그룹을 선별하게 되며, 대표 그룹 선별과정에서 다른 통행시간 그룹과 이상치를 점진적으로 제거해가는 과정을 거쳐 대표 구간통행시간을 추정하게 된다.

구간통행시간 추정모형은 고속도로 전 구간에 대해, 차종별 산정이 가능하도록 개발하고자 한다. 앞서 서술한대로 노선별로 통행량 차이가 크며, 시간대별, 차종별로 샘플수의 차이는 매우 크다. 따라서 <Fig. 4>와 같이 최소샘플수 기준($N > n_{min}$ (30건/5분))을 만족하지 않는 경우에는 좌우 시간대($t \pm l$)의 샘플을 추가한다. 비침두시 시간대 최소 샘플수를 만족하기 위하여 수집길이의 연장은 Liu et al.(2008)이 지적한 바와 같이 추정된 구간통행시간이나 통행속도의 신뢰도를 향상시키게 된다.

최소 샘플기준을 만족하게 되면 대표그룹 선정을 위하여 <Fig. 5>와 같이 다른 통행시간그룹과 이상치를 점진적으로 제거하게 된다. 대표그룹이 아닌 자료의 제거는 중위값을 기준으로 좌·우측 밀도를 각각 산정하고, 밀도가 낮은 영역의 끝단에서 가까운 일부 멤버들을 부분적으로 소거하는 과정을 반복하면서 대표그룹 멤버와 중심을 확장해가는 일종의 대표 그룹 중심 탐색법을 이용하였다. 여기서 말하는 밀도라는 개념은 시간의 공간 내에 존재하는 샘플의 개수로 정의하였다. 반복횟수가 증가할수록 밀도는 증가하게 되며 동질성(Homogeneity)이 가장 높은 그룹이 형성된다. 따라서 중위값(x_m)은 수렴하게 되며, 이때 중위값과 평균은 오차범위 내에서 동일하게 될 것이다.



〈Fig. 4〉 Expanding sample size



〈Fig. 5〉 Finding representative group and medianoid

대표그룹 탐색법(Exploring representative group and medianoid, 이하 ERGM)을 적용하기 위해서는 중위값(x_m), 최소값(x_{min}), 최대값(x_{max})이 필요하며, 다음과 같이 산정된다. $x = [x_n | n = 1, 2, \dots, N], N \geq n_{min}$ 에 대하여 x_n 을 오름차순 정렬 후, 초기 통행시간의 중위값, 최소값, 최대값을 산정한다. 중위값을 이용하는 통계기반 이상치 제거기법들은 오름차순 정렬과정을 수반할 수밖에 없다. 따라서 대표그룹 중심 탐색법의 초기 연산부하(load)는 기존의 이상치 제거방법과 동일하다.

ERGM은 반복과정을 거쳐 중위값(x_m)을 갱신하며, 중위값을 기준으로 좌·우측 밀도를 산정하여 밀도가 낮은 공간의 멤버를 소거하면서 대표 그룹의 멤버에 대한 x_m 을 수렴시킨다. 좌·우측 공간의 밀도(ρ_{lt}, ρ_{rt})는 다음 식으로 산정된다. 여기서, N_{lt}, N_{rt} 는 x_m 을 기준으로 좌측공간거리 $d_{lt} = |x_m - x_{min}|$ 와 우측공간거리 $d_{rt} = |x_m - x_{max}|$ 에 포함되는 통행시간의 표본수를 의미하며 $N = N_{lt} + N_{rt}$ 이다

$$\rho_{lt} = \frac{N_{lt}}{|x_m - x_{min}|} \tag{식(7)}$$

$$\rho_{rt} = \frac{N_{rt}}{|x_m - x_{max}|} \tag{식(8)}$$

좌·우측 공간밀도(ρ_{lt}, ρ_{rt})가 산정되면, 식(9)과 같이 밀도가 낮은 곳(ρ_{low})을 선정하고 이에 해당하는 거리공간(d_{low})을 결정한다.

$$\rho_{low} = \min\{\rho_{lt}, \rho_{rt}\} \tag{식(9)}$$

이상치 제거를 위하여 ρ_{low} 에 해당하는 통행시간 샘플 중 일부를 소거하기 위해 ρ_{low} 내에 존재하는 멤버의 집합을 $x_{low} = [x_{low,n} | n = 1, 2, \dots, N_{low}]$ 으로 정의한다. x 는 오름차순으로 정렬되었으므로 모든 $x_{low,n}$ 은 $x_{low,i} \leq x_{low,j}, i < j, \forall i, j \in N_{low}$ 의 조건을 만족한다. x_{low} 가 선정되면, 일정비율(α)을 이용하여 삭제할 멤버를 결정한다. 이상치 제거 시 통행시간자료가 일정공간에 조밀하게 분포할 경우 α 값이 너무 클 경우 한 번에 많은 자료들이 제거될 수 있으며, 너무 작을 경우에는 연산 수행속도가 오래 걸릴 수 있기 때문에 α 값은 황금비(0.6180339887)를 적용하였다. 1개의 멤버를 순차적으로 삭제할 수도 있지만, 대푯값 추정이라는 목적과 연산속도를 고려하여 ρ_{low} 공간의 일부를 순차적으로 삭제한다.

$$\alpha = \frac{\sqrt{5}-1}{2} \tag{식(10)}$$

만약, 밀도가 낮은 곳(ρ_{low})이 ρ_{lt} 인 경우 최소값(x_{min})은 식(11)과 같이 갱신되며, 밀도가 낮은 곳이 ρ_{rt} 인 경우 최대값(x_{max})은 식(12)와 같이 갱신된다.

$$x'_{min} = x_{min} + (x_m - x_{min}) \times \alpha \tag{식(11)}$$

$$x'_{max} = x_{max} - (x_{max} - x_m) \times \alpha \tag{식(12)}$$

갱신된 최소값(x'_{min})과 최대값(x'_{max}), 중위값(x'_m)을 기준으로 반복과정을 거쳐 이상치를 제거하며, 반복과정이 종료조건에 도달하면, 최적 대표 통행시간 그룹과 대표 통행시간(x_o)을 결정하게 된다. 반복과정의 종료조건은 식(13), 식(14)와 같이 최소값(x_{min}), 최대값(x_{max})이 임계조건($x_{threshold} = x_m \times \beta$)을 만족해야 하며, 여기서 β 는 0.05를 적용하였다.

$$x_m - x_{min} < x_{threshold} \tag{13}$$

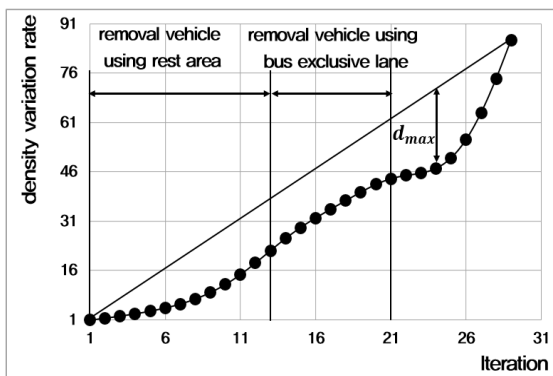
$$x_{max} - x_m < x_{threshold} \tag{14}$$

ERGM는 낮은 밀도 상태의 멤버를 제거하면서 높은 밀도 상태를 형성한다. 따라서 Iteration이 증가함에 따라, 초기밀도($N_0/(x^o_{max} - x^o_{min})$)에 대한 k 번째 밀도(ρ_k)를 이용하여 최적상태의 결과를 도출할 수 있으며, ρ_k 는 식(15)와 같다. 여기서, x^o_{max} 과 x^o_{min} 는 초기 최댓값과 최솟값, x^k_{max} 와 x^k_{min} 은 k 번째 Iteration 수행 후 최댓값과 최솟값이다.

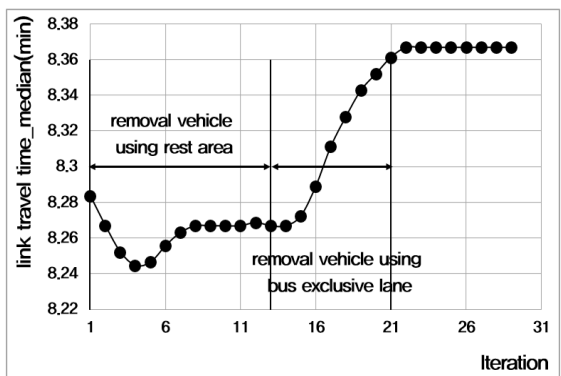
$$\rho_k = \frac{N_k / (x^k_{max} - x^k_{min})}{N_0 / (x^o_{max} - x^o_{min})} \tag{15}$$

<Fig. 6>은 ERGM의 반복연산 횟수 k 에 따른 밀도(ρ_k)의 변화를 보여주고 있다. 반복횟수가 증가함에 따라 ρ_k 는 지수형(Exponential)으로 증가하고 있다. 따라서 $k - \rho_k$ 관계는 전형적인 Elbow 방법(Thorndike, 1953)을 이용하여 최적의 반복회수를 결정할 수 있음을 보여주고 있으며, 최적의 반복회수는 d_{max} 일 때이다. d_{max} 를 만족하는 $k = \{1, 2, \dots, K\}$ 는 식(16)에 의해 산정되며, 여기서 K 는 마지막 반복회수이다.

$$\arg \max \left[\left(\frac{k}{K} \times (\max \{ \rho_k \} - \min \{ \rho_k \}) + \min \{ \rho_k \} \right) - \rho_k \right] \tag{16}$$



<Fig. 6> Determining optimal k value



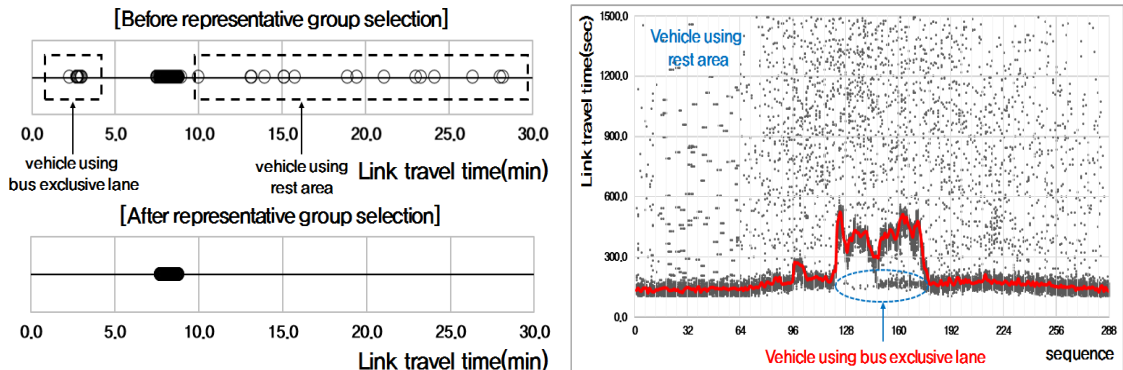
<Fig. 7> Convergence of representative travel time

<Fig. 7>은 반복횟수에 따른 대푯값인 중위값의 변화를 보여주고 있다. 구간통행시간이 긴 휴게소 이용차량의 통행시간이 제거되면서 대푯값은 감소하며, 혼잡시 버스전용차로 이용차량의 통행시간은 일반차로 이

용차량보다 낮기 때문에 낮은 대푯값을 유지한다. 이후 버스전용차로 이용차량의 통행시간이 제거되면서 대푯값은 급격히 증가한 후 수렴한다. 이후 일반차로 이용차량인 대표 구간통행속도 그룹의 멤버가 제거됨에도 불구하고 대푯값은 안정된 상태를 유지하고 있다. 이러한 사실은 ERGM 알고리즘이 구간통행시간 대푯값 추정에 있어 안정적이면서 명확한 대표 구간속도 그룹을 선별하고 있음을 의미한다.

<Fig. 8>은 ERGM 알고리즘을 이용한 대표 구간통행시간 그룹의 선정 전/후를 보여주고 있다. 버스전용차로 이용차량과 상대적으로 구간통행시간이 높은 휴게소 이용차량을 제거한 후 일반차로 이용차량 그룹을 합리적으로 선별하고 있다.

본 연구에서 대표 구간통행시간 자료구축을 위해 개발된 ERGM 알고리즘은 휴게소 이용차량과 버스전용차로 이용차량 등의 다양한 그룹 중 대표 통행시간 그룹을 구성하는 통행시간을 합리적으로 선별하고, 안정적인 대푯값을 추정하는 것으로 판단된다. <Fig. 9>는 경부고속도로 상행선 방향휴게소 구간의 대표 구간통행시간 추정 결과를 보여주고 있으며, 교통류 상황에 따라 통행시간의 동적 특성을 유지하면서 안정적으로 대표 구간통행시간을 추정하고 있는 것으로 판단된다. 또한, 본 연구에서 사용한 DSRC자료는 차종에 대한 정보를 포함하고 있으므로, 구간별 차종별 대표 통행속도 산정 시 유용하게 활용될 수 있을 것이다.



<Fig. 8> Selection result of representative group <Fig. 9> Estimated link travel time

IV. 결 론

구간검지체계에서 신뢰성 있는 정보생성을 위해서는 수집 자료의 정확도, 충분한 샘플 수, 신뢰성 높은 대푯값 추정이 선행되어야 한다.

기존 대부분의 구간통행시간 추정연구에서는 낮은 표본 수집 상태에서 하나의 대표 통행시간(정규분포 가정) 그룹을 가정하고, 평균 또는 중위값을 중심으로 신뢰구간을 선정하여 특정범위를 벗어난 경우 이상치로 판단하여 제거한 후 구간 통행시간을 추정하였다. 하지만, 고속도로의 경우 해당구간에 버스전용차로, 휴게소, 졸음쉼터 등이 존재할 경우 통행시간 분포는 정규분포가 아닌 쌍봉 또는 다봉 형태를 보일 것이다. 따라서 기존의 추정방법론을 적용할 경우 왜곡된 결과를 초래할 가능성이 있다. 이러한 문제를 해결하기 위하여 첫째, 샘플수가 부족한 상태 즉, 통행시간 그룹이 불분명한 조건에서도 신뢰할 수 있으며, 둘째, 다수의 통행시간 그룹 중 일반차로를 이용하면서 휴게소를 이용하지 않은 대표 통행시간 그룹을 선정하고, 선정된 통행시간 그룹의 개별 통행시간 자료를 이용하여 대표 통행시간을 추정할 수 있는 방법론을 제안하였다.

본 연구에서는 대표그룹과 이상치를 점진적으로 제거해가는 대표그룹 중심 탐색법(ERGM)을 개발하였고, 이를 이용하여 구간통행시간을 추정한 결과 교통류 상황에 따라 통행시간의 동적특성을 유지하면서 안정적으로 구간의 대푯값을 추정하는 것으로 나타났다.

본 연구의 한계는 DSRC 샘플이 충분하지 않을 경우 정확한 추정에는 한계가 있다. 예를 들어, 교통량이 많지 않은 일부노선 또는 새벽시간대에 버스 및 화물차는 샘플수가 충분치 않아 정확한 추정은 어려울 것으로 판단된다. 또한, 특정 RSE 통신장애로 인하여 자료가 수집되지 않을 경우 인접한 RSE간 통행시간을 이용하여 통행시간 보정은 가능하지만, 정확한 추정에는 한계가 있을 것으로 판단된다.

향후연구로는 샘플이 충분하지 않은 경우에 안정적으로 대표통행시간을 추정할 수 있는 모형 개발이 필요하며, 고속도로 구간의 정확한 경로 통행시간 예측을 위하여 구간 통행시간 뿐만 아니라 경로 통행시간을 추정할 수 있는 모형 개발이 필요할 것으로 판단된다.

REFERENCES

- Choi K. and Choi Y.(2006), "A heuristic outlier filtering algorithm for generation link travel time using taxi probes in urban arterial," *Journal of the Korean Society of Civil Engineers*, vol. 26 no. 5d, pp.731-738.
- Dion F. and Rakha H.(2003), "Estimation spacial travel time using automatic vehicle identification data," *presented in the 82nd Annual Meeting of Transportation Research Board*, Washington, DC.
- Dion F. and Rakha H.(2006), "Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates," *Transportation Research Part B*, vol. 40 no. 9, pp.745-766.
- Jeong Y., Park H., Kim B. and Kim Y.(2013), "Combined Filtering Model Using Voting Rule and Median Absolute Deviation for Travel Time Estimation," *J. Korea Inst. Intelligent. Transp. Syst.*, vol. 12, no. 6, pp.10-21.
- Kim S., Kim C., Yoo B. and Kwon Y.(2015), "A City Path Travel Time Estimation Method Using ATMS Travel Time and Pattern Data," *Journal of the Korean Society of Transportation*, vol. 33, no. 3, pp.315-321.
- Korea Expressway Corporation(2008), "Development of practical application technology of road traffic information detection system using DSRC".
- Liu Z., Sharma S. and Datla S.(2008), "Imputation of missing traffic data during holiday periods. *Transportation Planning and Technology*," vol. 31, no. 5, pp.525-544.
- Mouskos K. C., Niver E., Pignataro L. J., Lee S. and Antoniou N.(1998), "TRANSMIT system evaluation. Final report," *Institute for Transportation*, New Jersey Institute of Technology, N.J.
- Rim H., Oh C. and Kang J.(2009), "A Novel Method for Estimating Representative Section Travel Times Using Individual Vehicle Trajectory Data," *J. Korea Inst. Intelligent. Transp. Syst.*, vol. 8, no. 6, pp.23-35.
- Thorndike R. L.(1953), "Who belongs in the family?," *Psychometrika*, vol. 18 no. 4, pp.268-276.