

통행시간 정보 신뢰도 확보를 위한 프로브 대수 산정 방법론

Methods for Determination of Number of Probes Required
for Reliable Travel Time Information



장진환

1. 배경 및 목적

효율적인 실시간 교통정보 정보 제공을 위해서 최근 들어 하이패스 교통정보시스템, UTIS(Urban Traffic Information System) 등 프로브 기반 시스템이 확대 도입되고 있다. 전통적인 지점검지기에 비해 프로브 기반 시스템은 구간 통행시간 실측자료를 수집할 수 있을 뿐만 아니라 구축비용 또한 지점검지기에 비해 저렴하기 때문에 그 효용성이 높다고 할 수 있다.

하지만 프로브 기반 교통정보의 신뢰성을 확보하기 위해서는 일정수준 이상의 표본수가 필요하다. 지점검지기의 경우 통과하는 모든 차량의 데이터를 수집하기 때문에 표본수에 대한 고려가 불필요하지만, 프로브 기반 시스템의 현장자료 수집장치인 RSU(Roadside Unit)는 OBU(On-board Unit)가 부착된 차량(프로브)의 데이터만을 수집하기 때문에 표본수가 적을 경우 자칫 교통정보 신뢰도가 저하될 수 있다.

이러한 과소 샘플수 문제를 해결하기 위해 국도 등에서는 초기 설계단계부터 최소 표본수를 만족할 수 있도록 RSE 설치간격을 조정하거나, 만약 최소 표본수 요건을 만족하지 않을 경우 프로브 기반 시스템 대신 전통적인 지점검지기 기반 시스템을 도입하고 있다. 그러나 현재 최소 표본수 기준이 링크별 교통특성에 대한 구분 없이 5대/5분으로 설정되어 있어 이에 대한 개선이 필요한 실정이다.

통행시간 정보의 신뢰도 확보를 위한 표본수는 해당 링크(RSE와 RSE간의 노선)의 통행시간 분산, 교통류 특성 등에 따라 다르게 나타난다. 교통공학에서도 사전/사후 평가시 최소 표본수 산정에 대한 방법론만 제시할 뿐 최소 표본수에 대해서는 제시하지 않는다. Green 외(2004)의 연구에 따르면 링크 교통특성에 따라 최소 표본수가 1대일수도 있고 수 십대가 될 수도 있다고 주장했다. 또한 야간 등 교통량이 적고 통행시간 패턴이 일정할 경우 해당 수집주기 내에 프로브 데이터가 미수집 되더라도 이전주기 정보를 그대로 제공할 수도 있을 것

이다. 따라서 본 고에서는 프로브 기반 교통정보시스템 도입시 필수적으로 고려해야 하는 최소 샘플 수 산정 방법론을 전통적인 통계적 방법론과 교통정보 시스템 특성을 감안한 휴리스틱 방법론으로 나누어 제시하고자 한다.

II. 통계적 표본수 산정 방법론

널리 알려진 통계이론인 중심극한정리(Central Limit Theorem, CLT)에 의거하여 표본수는 식(1)에 의해 산출된다.

$$n = \left(\frac{z_{\alpha/2} \sigma}{\epsilon} \right)^2 \quad (1)$$

여기서,

- n : 표본수
- $z_{\alpha/2}$: 신뢰수준 $1-\alpha$ 에서의 양측검정 z -값
- σ : 모집단 표준편차
- ϵ : 허용오차

그러나 식(1)의 사용을 위해서는 모집단의 표준편차를 측정해야 한다. 이는 많은 조사비용이 소요되기 때문에 현실에서는 표본수 산정을 위해 주임의 표본 데이터를 수집하여 식(2)와 같이 모집단 표준편차(σ) 대신 표본 표준편차(s)를 사용하고 z -값 대신 t -값을 사용하여 표본수를 산출한다.

$$n = \left(\frac{t_{\alpha/2, n-1} s}{\epsilon} \right)^2 \quad (2)$$

여기서,

- n : 표본수
- $t_{\alpha/2, n-1}$: 신뢰수준 $1-\alpha$, 자유도 $n-1$ 에서의 양측검정 t -값
- s : 표본 표준편차
- ϵ : 허용오차

식(1), (2)는 최소 표본수 산정을 위해 허용오차를 사용한다. 그러나 허용오차는 일정한 값으로 주어지기 때문에 교통류 특성에 따른 오차특성을

고려하지 못한다. 예를 들어 허용오차를 5kph로 설정할 경우, 속도가 100kph일 경우 5%의 오차가 발생하지만 차량 정체시 속도가 20kph일 경우 25%의 오차를 발생시키게 된다. 이러한 문제를 해결하기 위해서는 허용오차 대신 오차율을 사용하는 식(3) 또는 (4)를 이용하여 표본수를 산정할 수 있다. 식(3)은 TTI(1998)에 의해 제시되었고 식(4)는 Srinivasan 외(1996)에 의해 유도되었다. 식(3)은 표준 정규분포 변수(z)를 사용했고 식(4)는 표준 정규분포 누적분포 역함수(Φ^{-1})를 사용하여 나타냈을 뿐 두 식은 본질적으로 동일하다.

$$n = \left(\frac{z_{\alpha/2} c.v}{e} \right)^2 \quad (3)$$

여기서,

- n : 표본수
- $z_{\alpha/2}$: 신뢰수준 $1-\alpha$ 에서의 양측검정 z -값
- $c.v$: 모집단 변동계수(σ/μ)
- e : 상대오차(ϵ/μ)

$$n = \left(\frac{\Phi^{-1}[(1+r)/2]}{\epsilon(\mu/\sigma)} \right)^2 \quad (4)$$

여기서,

- n : 표본수
- $\Phi^{-1}[(1+r)/2]$: 신뢰수준 r 에서의 표준정규분포 누적분포 역함수
- ϵ : 허용오차
- μ : 모집단 평균
- σ : 모집단 표준편차

여기서 유의할 점은 식(1)-(4)를 이용하여 샘플수를 산정할 경우 CLT 기본가정에 의거하여 모집단의 분포가 정규분포이거나 표본수가 30개 이상이어야 한다. 그러나 Green 외(2004)은 통행시간(속도) 분포는 정규분포를 따르지 않는 경우가 존재한다고 하였다. 또한 6차로 이상 고속도로 등 교통량이 많은 도로를 제외하고는 프로브 대수가 5분당 30대 이상인 경우는 드물다. 더욱이 최소 표

본수 산정을 위해 사전 임의 표본조사 후 식(2)를 적용할 경우 표본수를 미리 알 수 없기 때문에 t -값 대신 일반적으로 z -값을 사용하게 된다(Li et al., 2002). 이 경우 특히 샘플수가 적을 경우 t -값과 z -값의 차이가 크기 때문에 t -값 대신 z -값을 사용하는 것은 과소 표본수를 산출할 우려가 있다. 만약 신호교차로가 있는 간선도로의 경우에는 신호등의 영향으로 인해 개별차량 통행시간의 공분산이 0이 아닌 경우 표본평균의 분산은 식(5)와 같이 산출되므로 이러한 경우에는 표본수 산출시 개별차량 통행시간의 공분산도 고려해 주어야 한다(Sen et al., 1997).

$$\frac{\text{var}(\bar{x}) = [n(n-1)]^{-1} \sum_{\substack{i,j \\ i \neq j}} \text{Cov}(x_i, x_j) + \frac{n^{-1} \sum_{i=1}^n \text{var}(x_i) - [(n(n-1))^{-1} \sum_{\substack{i,j \\ i \neq j}} \text{Cov}(x_i, x_j)]}{n}}{n} \quad (5)$$

여기서,

- $\text{var}(\bar{x})$: 표본평균 분산
- n : 표본수
- $\text{Cov}(x_i, x_j)$: 개별 통행시간(x_i, x_j) 공분산
- $\text{var}(x_i)$: 개별 통행시간 분산

앞서 살펴본 바와 같이 일반적인 통계이론을 이용하여 프로브 기반 교통정보시스템 RSE 적정 설치율을 위한 표본수 산정에는 몇 가지 이슈가 존재한다. 따라서 해당 링크가 통계적 표본수 산정을 위한 기본 가정에 위배될 경우 다음 장에서 소개될 휴리스틱 방법론을 적용할 수 있다.

III. 휴리스틱 표본수 산정 방법론

실시간 교통정보제공시스템(Advanced Traveler Information System, ATIS)에서 프로브 데이터 수집 목적은 일반 교통공학에서 목적과는 다소 상이하다. 다시 말해서 일반적으로 교통공학에서 최소 표본수를 산정하는 이유는 교통개선사업 등

에 따른 사업 전/후의 효과를 분석하기 위한 것이므로 최소 표본수 데이터는 수집하는 기간의 모집단을 통계적으로 대표할 수 있어야 한다. 그러나 ATIS에서는 상류부 차량에게 해당 구간의 실시간 교통정보를 제공하는 것이 목적이므로 최소 표본수 데이터는 후방 차량에게 신뢰성 있는 통행시간 정보를 제공할 수 있어야 한다. 따라서 ATIS에서 최소 표본(프로브)수 데이터는 식(6)에 의해 산출되는 오차를 일정수준 이하로 유지할 수 있어야 한다.

$$e_l(\%) = \frac{\mu_{sl,d,a,k} - \bar{x}_{el,a,p,k}}{\mu_{sl,d,a,k}} \times 100 \quad (6)$$

여기서,

- e_l : 링크 l 의 교통정보 오차
- $\mu_{sl,d,a,k}$: k 수집주기에서 링크 l 출발지점(sl)을 출발(d)하는 전체 차량(a) 통행시간 평균(모평균)
- $\bar{x}_{el,a,p,k}$: k 수집주기에서 링크 l 도착지점(el)을 도착(a)하는 프로브 차량(p) 통행시간 평균(표본평균)

식(6)은 프로브 데이터를 산술 평균하여 상류부 차량에게 제공하는 것을 전제로 하고 있다. 그러나 실제 교통정보시스템에서는 신경망, 칼만필터, 패턴매칭, 회귀분석 등 다양한 기법을 이용하여 예측된 정보를 제공한다. 이 경우, 식(7)과 같이 예측된 통행시간을 이용하여 오차를 산출해야 한다.

$$e_l(\%) = \frac{\mu_{sl,d,a,k} - P_{el,a,p,k}}{\mu_{sl,d,a,k}} \times 100 \quad (7)$$

여기서,

- $P_{el,a,p,k}$: k 수집주기에서 링크 l 도착지점(el)을 도착(a)하는 프로브 차량(p) 이용하여 예측한 통행시간

식(7)은 예측된 통행시간과 해당 링크를 출발하는 전체 차량의 통행시간 평균값을 비교하여 오차를 산출함에 따라 운전자들이 실제로 경험하는 오

차와는 다소 차이가 발생할 수 있다. 왜냐하면 일부 비정상(abnormal) 값들은 평균을 증가시키거나 감소시킬 수 있기 때문이다. 따라서 개별 운전자들이 경험하는 통행시간 오차는 식(8)에 의해 산출되어야 한다. Cetin et al.(2005)은 통행시간 오차의 누적 확률을 분석한 결과, 동일 표본수 조건에서 식(8)에 의한 오차가 식(7)에 의한 오차보다 양호한 결과를 도출한다고 주장했다.

$$e_l(\%) = \frac{x_{sl,d,i,k} - P_{el,a,p,k}}{x_{sl,d,i,k}} \times 100 \quad (8)$$

여기서,

$x_{sl,d,i,k}$: k 수집주기에서 링크 l 출발지점(sl)을 출발(d)하는 개별 차량(i) 통행시간

이상에서 살펴 본 휴리스틱 방법론은 통계적 방법론 적용이 이론적으로 불가능할 경우에도 적용 가능하다. 예를 들어 일정 신뢰수준(예. 95%)에서 교통정보 오차가 일정수준(예. 20%) 이하가 되는 최소 표본수를 산출하는 등으로 적용할 수 있을 것이다.

IV. 결론 및 향후과제

연속류 도로인 고속도로와 달리 국도 및 도시부 간선도로의 경우 유출입 지점 등의 영향으로 OBU 보급률이 높음에도 불구하고 링크별 최소 표본수가 만족하지 않아 자칫 제공되는 실시간 통행시간 정보의 신뢰도가 저하될 수 있다. 이에 본 고에서는 최근 들어 활발히 도입되고 있는 프로브 기반 교통정보시스템에서 링크별 최소 표본수를 산정하는 방법론을 고찰하였다.

방법론은 크게 통계적 방법론과 휴리스틱 방법론으로 나누어진다. 통계적 방법론은 널리 알려진 CLT를 기반으로 절대오차와 상대오차를 이용하여 최소 표본수를 산정하고, 휴리스틱 방법론은 교통정보시스템 특성을 고려하여 프로브 데이터를 이용하여 생성된 교통정보가 해당 정보를 받고 해당구간을 통행한 후방 차량의 실제 통행시간과 얼

마나 차이가 발생하는가에 대한 지표를 이용하여 최소 표본수를 산정한다.

서론부에서 언급했듯이 교통정보 신뢰도 확보를 위한 최소 표본수는 해당구간의 통행시간 분산, 교통특성 등에 따라 다양하게 나타나므로 현재처럼 모든 링크에 대해서 일정한 표본수 기준(예. 5대/5분)을 적용하는 데에는 다소 무리가 따르는 것으로 사료된다. 따라서 본 고에서 제시한 다양한 방법론을 활용하여 링크별로 교통정보 신뢰도 확보를 위한 최소 표본수를 산정한 후, 이를 만족할 수 있는 방향으로 시스템이 구축되어야 할 것이다.

참고문헌

- K. K. Srinivasan and P. P. Jovanis (1996), Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network, Transportation Research Record 1537, Transportation Research Board.
- A. Sen, P. Thakuriah, X. Q. Zhu, and A. Karr (1997), Frequency of Probe Reports and Variance of Travel Time Estimates, Journal of Transportation Engineering, 123(4), American Society of Civil Engineers.
- M. Chen and S. I. J. Chien (2000), Determining the Number of Probe Vehicles for Freeway Travel Time Estimation by Microscopic Simulation, Transportation Research Record 1719, Transportation Research Board.
- S Li, K. Zhu, B. H. W. van Gelder, J. Nagle, and C. Tuttle (2002), Reconsideration of Sample Size Requirements for Field Traffic Data Collection with Global Positioning System Devices, Transportation Research Record 1804, Transportation Research Board.
- C. Lee and J. Y. Park (2001), Determining the Optimal Number of Probe Vehicle for ATIS Applications in Urban Networks,

- The 9th World Conference on Transport Research Proceedings, Seoul, Korea.
- R. L. Cheu, C. Xie, and D. H. Lee (2002), Probe Vehicle Population and Sample Size for Arterial Speed Estimation, *Computer-Aided Civil and Infrastructure Engineering* 17, Blackwell Publishers.
- M. W. Green, M. D. Fontaine, and B. L. Smith (2004), Investigation of Dynamic Probe Sample Requirements for Traffic Condition Monitoring, *Transportation Research Record*, 1870, Transportation Research Board.
- M. Cetin, G. F. List, and Y. Zhou (2005), Factors Affecting Minimum Number of Probes Required for Reliable Estimation of Travel Time, *Transportation Research Record*, 1917, Transportation Research Board.
- T. Ishizaka, A. Fukuda, and S. Narupiti (2005), Evaluation of Probe Vehicle System by Using Micro Simulation Model and Cost Analysis, *Journal of the Eastern Asia Society for Transportation Studies*, 6, Eastern Asia Society of Transportation Studies.
- J. M. Ernst, C. M. Day, J. V. Krogmeier, and D. M. Bullock (2012), Probe Data Sampling Guidelines for Characterizing Arterial Travel Time, *Transportation Research Record*, 2315, Transportation Research Board.