

이력 자료를 활용한 GPS 교통정보의 효율적인 필터링 방법[†]

An Efficient Filtering Technique of GPS Traffic Data using Historical Data

최진우* / Jin-Woo Choi, 양영규** / Young-Kyu Yang

요약

텔레매틱스 교통정보(해당 도로의 통과 시간이나 속도)를 제공하기 위해 교통자료를 수집하는 다양한 장비들이 존재하지만, 최근 GPS(Global Positioning System) 위성 신호가 민간에 공개된 이후 수신 기술의 발달과 더불어 교통 자료를 수집하는 수단으로써 많은 주목을 받고 있다. GPS를 사용하면 기존의 도로 상에 직접 매설하여 쓰이던 검지기에 비해 투자·유지 비용이 낮고 차량 운행 축을 따라 지속적으로 세밀한 자료를 자동적으로 수집할 수 있다는 장점을 가진다. 하지만, 기존의 타 검지기에 의해 수집된 자료에 적용해 온 교통 정보 필터링 기법들을 GPS 자료에 그대로 적용하기에는 여러 가지 제약사항이 존재한다. 본 논문에서는 GPS로부터 수집되는 교통 자료 중 비정상적인 흐름으로 도로를 통과한 자료를 필터링하여 사용자들에게 보다 정확한 교통 정보를 제공할 수 있는 방법을 제시한다. 먼저 가까운 과거의 자료로부터 사분위수(quartile)와 해당 도로의 가중치에 의한 패턴을 구축하고, 이후 이를 실시간 자료에 적용해 비정상적인 자료를 필터링 해냄으로써 더욱 신뢰성 있는 교통 정보를 생성할 수 있게 되는 것이다. 또한, 서울 강남대로를 대상으로 실제 GPS 수신기를 장착하고 운행한 차량들이 수집한 통행속도 자료를 바탕으로 이력 자료를 활용해 비정상적인 자료를 필터링한 대표 통행 속도가 기존의 필터링 방법들을 통한 값보다 더욱 정확한 결과를 보임을 증명하였다.

Abstract

For obtaining telematics traffic information(travel time or speed in an individual link), there are many kinds of devices to collect traffic data. Since the GPS satellite signals have been released to civil society, thank to the development of GPS technology, the GPS has become a very useful instrument for collecting traffic data. GPS can reduce the cost of installation and maintenance in contrast with existing traffic detectors which must be stationed on the ground. But, there are problems when GPS data is applied to the existing filtering techniques used for analyzing the data collected by other detectors.

† 본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음
(IITA-2008-(C1090-0801-0040))

■ 접수일 : 2008.05.14 ■ 수정일 : 1차 2008.06.25 / 2차 2008.07.16 ■ 심사완료일 : 2008.07.16

* 경원대학교 전자계산학과 박사과정(cjw49)@paran.com

** 교신저자 경원대학교 소프트웨어대학 교수(ykyang)@kyungwon.ac.kr

This paper proposes a method to provide users with correct traffic information through filtering abnormal data caused by the unusual driving in collected data based on GPS. We have developed an algorithm that can be applied to real-time GPS data and create more reliable traffic information, by building patterns of past data and filtering abnormal data through selection of filtering areas using Quartile values. In order to verify the proposed algorithm, we experimented with actual traffic data that include probe cars equipped with a built-in GPS receiver which ran through Gangnam Street in Seoul. As a result of these experiments, it is shown that link travel speed data obtained from this algorithm is more accurate than those obtained by existing systems.

주요어 : 텔레매틱스, 링크통행속도, GPS, 필터링, 사분위수

Keyword : Telematics, Link Travel Speed, GPS, Filtering, Quartile

1. 서 론

우리나라의 교통 혼잡도는 이미 세계 최고 수준에 이르고 있으며, 이로 인한 직-간접적인 사회 손실 비용이 한 해 23조원 이상으로, 국민총생산(GDP)의 2.94%에 달하는 비용이 도로상에서 낭비되어지고 있는 실정이다[1] [2]. 이는 차량의 증가 속도에 대비하여 도로의 확충이나 정비, 교통 신호의 체계화 등이 더딘 발걸음을 해 왔기 때문이다. 하지만, '90년대부터 ITS(지능형 교통 시스템: Intelligent Transportation System) 사업이 꾸준히 진행되어, 여러 부문에서 소기의 성과를 이루었고, 또한 최근에는 도로 위의 ITS 기반 시설과 차량내 텔레매틱스 서비스가 접목되어 우리나라의 고질적인 교통문제의 해결책을 제시해 줄 뿐만 아니라, 자동차를 말 그대로 "움직이는 사무실(mobile office)"로 변신시켜 줄 것이라는 기대가 점차 높아지고 있다[2].

텔레매틱스 서비스 중 사용자 수요가 가장 높게 나타난 핵심 서비스는 단연 차량내 텔레매틱스 단말기에서 제공받을 수 있는 교통정보제공 서비스이다[2]. 이때 교통정보를 수집하는 방법으로 루프검지기, 영상검지기, 비콘, GPS(Global Positioning System) 등의 장비가 활용되어지고 있는데, 최근에는 차량의 운행 축을 따라 지속적인 자료를 수집할 수 있고, 초기 투자비용만으로 양질의 구간 교통 자

료를 수집할 수 있는 GPS를 이용한 방법이 가장 선호되고 있다[3]. 그러나 현재까지는 사생활 침해 문제와 차량 운행 시간이 적은 이유로 인해 일반 차량들에서는 내비게이션 용도로만 쓰이고 있으며, 교통 자료 수집은 주로 버스나 택시 등의 대중교통과 상업용 차량으로만 행해지고 있다[4] [5].

이와 같은 문제로 인하여, GPS를 이용해 수집된 교통 자료는 수집되는 자료의 수가 간헐적이고, 대중교통 특유의 일관성 없는 운전 습성이 자료에 고스란히 드러나는 경우가 많다[4]. 특히 GPS probe로 가장 널리 쓰이고 있는 택시는 공차일 경우, 승객을 기다리는 대기시간으로 인해 통행속도가 승차일 때보다 낮게 수집되는 습성을 가지고 있다.

정보 수집 단위 시간 내에 충분한 자료가 있을 경우, 해당 링크 내에서 이러한 이상 주행 패턴을 보인 차량은 통계적으로 쉽게 제거될 수 있고, 비교적 정확한 통행 시간을 산출할 수 있을 것이다[6]. 다른 검지 장비들에서 수집된 교통정보는 충분한 자료의 수가 확보되어 여러 가지 통계적인 방법으로 이상치 자료에 대한 필터링을 하고 있지만, GPS를 통해 수집된 통행속도 자료에 이를 그대로 적용하기에는 무리가 있다. 본 연구에서는 이를 해결하기 위한 방안으로 GPS probe 택시들로부터 수집된 통행속도 자료를 링크 별로 이력 교통정보를 구축하고, 유용한 통행속도 자료를 구분하기 위한 유효범위를 사분위수(quartile)에 가중치를 고려

한 fence값으로 설정한 뒤, 실시간으로 수집되는 통행속도들을 이 fence값에 의거하여 이상치 자료 인지 판단하고 제거하는 방법에 대해 연구하였다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 배경 이론으로 GPS를 이용한 교통정보의 수집 방법에 대하여 분석한다. 3장에서는 현재까지 교통정보에서 이상치 자료를 필터링하는 방법들에 대해 분석한다. 4장에서는 본 연구에서 실시한 이력 자료를 활용한 GPS 교통정보의 효율적인 필터링 방법에 대해 설명하고, 5장에서는 실험 결과를 알아 본다. 마지막 6장에서는 결론에 대해 언급한다.

2. GPS를 이용한 교통정보 수집

GPS는 처음 미국 국방성에서 군사 용도로 개발 되었지만, 1983년 KAL-007기 격추 사건을 계기로 민간에 신호가 공개되었다. 하지만 오랫동안 민간 GPS 신호에는 SA(Selective Availability)라는 암호코드를 첨가시켜 의도적으로 오차를 발생시켰었다. 그러나 2000년 5월 1일부로 SA를 제거하므로 해서 저가의 수신기를 가지고도 비교적 정확한 위치 정보를 얻어 낼 수 있게 되었다[7].

GPS를 이용한 교통 정보 수집은 GPS 수신기와 무선모뎀을 장착한 프로브 차량이 도로를 운행하며 자신의 위치 정보를 일정 단위 시간 간격으로 교통정보관제센터에 전달하고, 교통관제센터에서는 링크와 노드로 구성된 도로 GIS 자료에 맵 매칭 작업을 통해 모든 프로브 차량의 위치를 재계산하여 차량이 통과한 링크들의 통행 시간 정보 또는 통행 속도를 계산하는 방식으로 이루어진다[3]. 이 중 단말기 수집 시나 통신 시 또는 다른 여러 가지 이유로 생길 수 있는 오류를 제거·보정한 뒤, 각 링크별로 차량들의 평균 속도를 계산하여 링크의 대표 속도를 예측하게 된다. 이 값은 도로를 운행하고 있거나, 운행 예정인 사용자들에게 전달되어 목적지까지 최적의 경로로 이동할 수 있도록 도움을 주게 된다.

GPS를 이용하여 수집된 속도 정보는 구간 검지 체계의 특성상 프로브 차량이 다니지 않는 곳의 자료는 수집할 수 없다는 것과 도심 내 같은 경우에

서는 높은 빌딩 등으로 인해 GPS 위치 측정이 다소 어렵다는 점, 특히 개인적인 프라이버시 문제와 비용 문제 등으로 인해 일반 차량을 프로브 차량으로 사용할 수 없는 한계를 지니고 있다. 이런 이유로 현재까지는 수집 대상이 주로 대중교통이나 상업용 차량에 국한되어 있어 기존의 검지기들을 통해 수집되는 정보보다 수집되는 자료량이 간헐적이다. 또한, 수집되는 속도 정보가 특수한 목적의 차량들의 주행 정보로써 일반 차량들과는 상이한 운전 습관으로 인해 링크마다 포함되는 이상치 자료의 비율이 상이하다. 이에 이런 정보는 적절한 필터링 방법을 통해 걸러 내거나 보완하여야만 교통정보 서비스 시 대표 속도의 정확성을 보장할 수 있다.

3. 기존 교통 정보 필터링 기법 고찰

지금까지 교통 속도 정보 가공 시의 필터링 방법으로는 통계적 기법인 MAD, Trimmed mean, Winsorized mean 등을 이용하는 방법들이 알려져 있는데, 각각의 설명은 다음과 같다.

3.1 MAD(Median Absolute Deviation)를 이용한 방법

수집된 데이터의 중앙값(median)을 이용하여 비정상적으로 크거나 작은 값들을 걸러내는 방법으로써 MAD를 계산하는 기본 식은 다음과 같다.

$$MAD = b \times \frac{1}{N} \times \sum_{i=1}^N |x_i - x_{m,d}| \tag{1}$$

여기서, x_i : 변수 x 의 관측값,

$x_{m,d}$: 변수 x 의 중앙값,

b : MAD를 정규분포에 대한 표준편차와 같도록 만들어주는 수정계수 (correction factor)

위의 식을 통해 MAD를 선정하고, 이상치를 포함한 데이터에 대해 Robust standard units z_i^{MAD} 를 계산한 후 그 값을 제거변수(cut-off value)

z_{out} 과 비교하여 이상치를 판별하게 된다.

z_i^{MAD} 를 계산하는 식은 다음과 같다.

$$z_i^{MAD} = \frac{x_i - x_{med}}{MAD} \quad (2)$$

z_i^{MAD} 이 z_{out} 보다 크면, x_i 는 이상치로 판별되어 제거 대상이 된다. 일반적으로 z_{out} 은 3으로 설정한다.

3.2 Trimmed mean을 이용한 방법

수집된 데이터의 최대값과 최소값 부분을 일정한 퍼센트로 잘라내고 남은 부분만을 이용하여 평균을 취하는 방법으로써, 100% Trimmed mean은 중앙값과 같고, 0% Trimmed mean은 산술평균값과 같게 된다. Trimmed mean, \overline{x}_{ik} 을 구하는 식은 다음과 같다.

$$\overline{x}_{ik} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} x(i) \quad (3)$$

여기서, n : 총 데이터 수,

k : 절삭할 데이터의 수(제거되는 데이터의 수= $2k$)

3.3 Winsorized mean을 이용한 방법

Trimmed mean에서 데이터의 분포를 관찰하고 잘라내어야 하는 부분들을 정하는 것과는 달리 Winsorized mean은 데이터의 중앙에 분포한 값들에 더욱 가중치를 두어 평균을 구하는 방법으로, Trimmed mean에서 자료를 제거하는 절차와는 달리 바로 옆 관측치로 그 값을 대체하여 평균을 구하는 방법이다. Winsorized mean, \overline{x}_{wk} 를 구하는 식은 다음과 같다.

$$\overline{x}_{wk} = \frac{1}{n} ((k+1)x_{k+1} + \sum_{i=k+2}^{n-k-1} x(i) + (k+1)x_{n-k}) \quad (4)$$

여기서, n : 총 데이터 수,

k : 대체할 데이터의 수(대체되는 데이

터의 수= $2k$)

3.4 기존 방법들의 문제점

앞에서 언급한 필터링 방법들은 간단한 절차를 통해 루프검지거나 영상검지기 등, 지점검지체계 수집 장비에서 수집된 속도 정보에서 이상치를 대부분 제거해 냄으로써 현장에서 널리 쓰이고 있는 방법이다. 하지만, 이와 같은 방법들은 모두 통계적인 방법으로 자료를 처리하기 때문에 수집되는 자료의 수가 일정 수준 이상이어야만 정확한 성능을 보이게 되는 한계를 가지고 있다. 또한, 모든 링크를 일괄적인 방법으로 필터링하기 때문에 링크의 특성에 맞게 이상치 자료를 처리하기 어렵다.

따라서 현재 GPS를 통해 수집되는 속도 정보는 위와 같은 필터링 방법들을 바로 적용하기가 어렵고 GPS 교통정보의 특성에 맞는 새로운 필터링 방법을 고안해야 할 필요가 있다.

4. 이력 자료를 활용한 GPS 교통정보의 필터링

본 장에서는 GPS 프로브 차량들을 통해 수집되는 링크 별 이력 속도 자료에서 몇 개의 특징값들로 패턴 DB를 구축한 뒤, 이를 토대로 실시간에 이상치를 판별하여 제거할 수 있는 필터링 방법에 대해 설명한다.

4.1 수집 자료

먼저 본 연구에서 사용한 GPS 교통 자료에 대해 설명한다.



<그림 1> 실측 자료 수집 구간

본 연구에서는 2005년 6월 8~29일 사이에 <그림 1>에 나타난 서울시 강남대로의 일부 링크(한남대교남단→서초우성아파트, 총 연장 2.8Km)를 대상으로 GPS 수신기를 장착한 수집 차량(영업용 택시)들에 의해 수집된 속도 자료를 이용하였다. 자료는 서울시 소재의 수 천대의 프로브 택시들로부터 5분 단위마다 교통 정보를 수집하여 제공 서비스를 하고 있는 리얼텔레콤(주)에서 제공받았다.

수집 장소로 강남대로를 택한 이유는 자료의 수집도가 타 지역에 비해 비교적 높은 지역이라는 것이다. 또한, 프로브 차량으로 운용 중인 차량들이 영업용 택시이기 때문에 시내에서는 승객을 태우기 위한 정차로 인해 정상적인 운행을 하지 않는 경우가 있어 이런 차량에서 수집된 자료는 보통의 다른 자료들과는 상이한 이상치 자료가 된다. 따라서 이 수집 구간이 분석 자료로 활용하기에 매우 적합할 것으로 예상되었다. 자료 수집 구간의 링크 구성은 <표 1>에 나타난 것과 같이 총 5개 링크로 구성되어 있다.

<표 1> 테스트 구간의 링크 구성

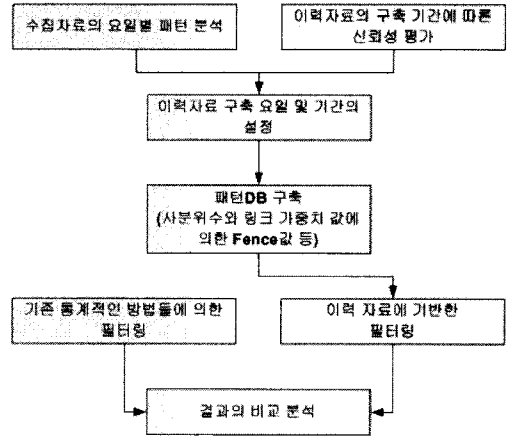
번호	시작노드	종점노드	길이(m)
1	한남대교남단	신사역	610
2	신사역	논현역	548
3	논현역	교보타워사거리	752
4	교보타워사거리	강남역	748
5	강남역	서초우성아파트	554

아울러 본 연구에서는 차량의 흐름이 일정 수 이상 확보될 수 있는 오전 6시에서 오후 10시까지의 자료만을 이용하여 실험을 실시하였다.

4.2 연구의 진행방법

본 연구에서는 보통 5~15분 사이의 교통자료 수집 단위 시간에서는 통계적인 방법으로 필터링하기에는 충분한 수의 자료 확보가 어려운 GPS 교통정보의 특성을 고려하여 가까운 과거의 이력자료를 활용하는 방안을 모색하고 있다. 먼저 본 연구의 진

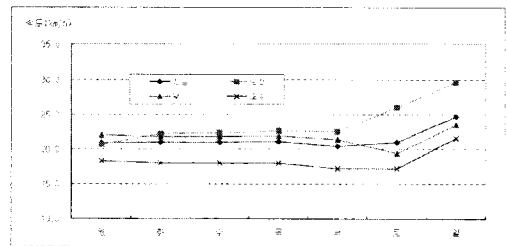
행 절차는 <그림 2>와 같다.



<그림 2> 연구의 진행방법

4.3 수집 자료의 요일별 패턴 분석

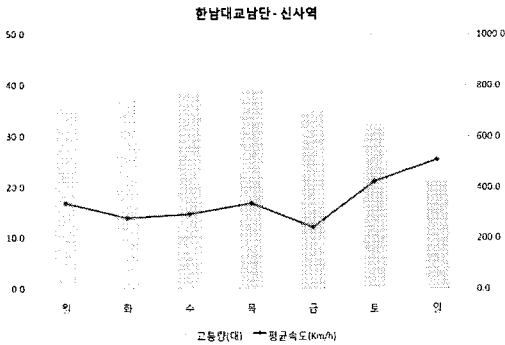
<그림 3>의 서울시 교통국에서 발표한 2005년 서울시 전체 도로의 요일별 평균 속도를 보면, 주말과 공휴일을 제외하면 주중(월~금요일)에는 오전, 낮, 오후 모두에서 비슷한 속도를 보이고 있음을 알 수 있다. [8]



<그림 3> 2005년 서울시 요일별 평균 속도

수집 자료의 분석결과 본 연구에서 수집 장소로 택한 강남대로 또한 이러한 성향이 강하게 나타났는데, <그림 4>에 나타난 것처럼 강남대로의 링크들 중 하나인 '한남대교남단-신사역'의 경우를 보면 요일별 평균 속도와 교통량이 주말에는 아주 상이한 패턴을 보이는 것을 알 수 있다. 아울러 다른 4개 링크도 이와 비슷한 양상을 보였다. 따라서

이를 고려하여 주중과 주말 각각의 이력 자료를 구축해야 할 필요가 있을 것이라 판단된다. 본 논문에서는 일단 주중의 경우(월~금요일)만을 고려하여 연구를 진행하였다.

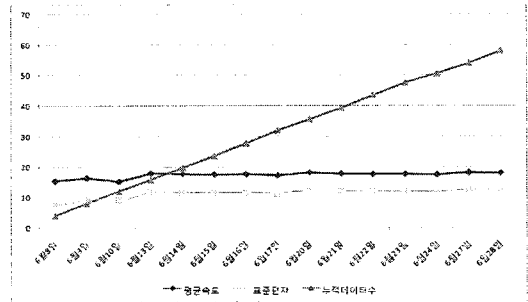


<그림 4> 한남대교남단-신사역의 요일별 프로브 차량의 교통량 및 평균속도(6시~22시)

4.4 이력자료 구축 기간에 따른 신뢰성 평가

본 연구에서 사용된 수집 자료의 경우, 오전 6시~오후 10시 사이 5분 단위 시간 내 한 링크에서 수집되는 자료의 수가 주중 평균적으로 3.88개 정도에 불과했다. 5분 단위 시간 내에 해당 링크를 통과하는 수많은 차량 중 4개도 되지 않는 표본수로 링크의 대표 속도를 산정하기에는 무리가 있을 수 있다. 수집된 개별 자료들을 생성한 프로브 차량들의 정상 운행 여부를 확인하기 어렵기 때문이다. 만일 충분한 수의 자료가 존재한다면 통계적으로 분포를 고려하여 이상치일 가능성이 높은 자료들을 골라낼 수 있겠지만 한정된 프로브 차량의 수로 인해 무수히 많은 링크에서 충분한 자료의 수집이 어려운 상황이다. 따라서 비슷한 교통 흐름을 보였던 가까운 과거 며칠의 이력 자료들을 활용하여 실시간에 수집되는 자료의 이상 여부 확인에 사용할 수 있을 것이다. 문제는 어느 정도 기간의 자료를 구축해야 신뢰성이 보장될 수 있는지에 관한 것이다. <그림 5>에 나타나듯이 “한남대교남단-신사역” 링크에서처럼 이력 자료의 구축 기간이 늘어날수록 평균과 표준편차는 크게 변하지 않으면서도 누적

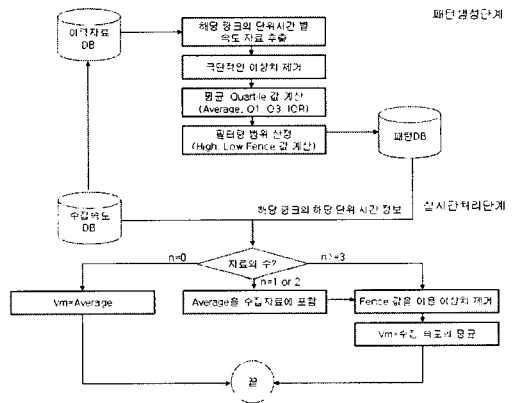
데이터 수는 정비례로 늘어나고 있기 때문에 6월 8~29일 중 월~금요일에 해당하는 15일 동안의 자료를 모두 사용하였다. 단, 마지막 날인 29일의 자료는 테스트를 위해 남겨 놓았다.



<그림 5> 이력자료 구축 기간과 변수들과의 관계

4.5 실시간 필터링을 위한 패턴DB 구축

본 절은 실시간 필터링을 신속하게 처리하기 위해 구축된 이력 자료를 기반으로 생성하게 되는 패턴DB 구축에 관한 내용이다.

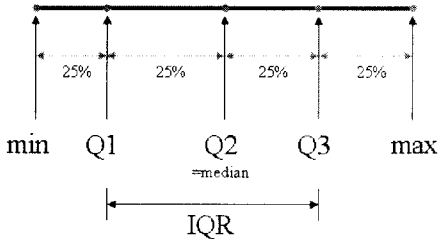


<그림 6> 패턴DB구축 및 실시간 필터링 방법의 흐름도

먼저, 교통자료 수집 단위 시간 내 링크를 통과하는 프로브 차량이 한 대도 없을 경우 혹은 그 수가 1~2대 정도로 매우 적어 대표값을 신뢰할 수 없을 경우에 대비하여 대체할 수 있는 링크별, 단위시간 대별 평균값을 이력 자료로부터 미리 계산해 패턴

DB 에 넣어 둔다.

그 다음 수집 단위 시간 별로 사분위수(David Hoaglin, 1983)를 계산한다. 사분위수는 크기 순서에 따라 늘어놓은 자료를 4등분 하는 수모써 그 의미는 다음과 같다[9].



<그림 7> 사분위수의 개념

◦ Q1, First quartile(=low quartile)

자료의 25%가 그 값보다 작고 자료의 75%가 그 값보다 크게 될 때 그 값을 Q1이라 정의한다.

◦ Q2, Second quartile(=median)

자료의 50%가 그 값보다 작고 자료의 50%가 그 값보다 크게 될 때 그 값을 Q2이라 정의한다.

◦ Q3, Third quartile(=upper quartile)

자료의 75%가 그 값보다 작고 자료의 25%가 그 값보다 크게 될 때 그 값을 Q3이라 정의한다.

◦ IQR, Inter-Quartile Range

Q3-Q1을 IQR이라 정의한다.

다음으로 앞에서 계산된 사분위수 중 Q3, Q1 그리고 전체 자료의 50%가 존재하는 범위를 나타내는 IQR을 기준으로 두 개의 필터링 fence값, High fence(F_{high})와 Low fence(F_{low})를 계산하여 패턴DB에 삽입한다. 이를 구하는 식은 다음과 같다.

$$F_{high} = Q3 + M_{high} \times IQR \quad (5)$$

(단, $F_{high} > Speed_{max}$ 이면, $F_{high} = Speed_{max}$)

$$F_{low} = Q1 - M_{low} \times IQR \quad (6)$$

(단, $F_{low} < Speed_{min}$ 이면, $F_{low} = Speed_{min}$)

여기서,

$Speed_{max}$: 해당 링크 제한속도의 120%값,

$Speed_{min}$: 0,

F_{high} : High fence, F_{low} : Low fence,

M_{high} : High measure, M_{low} : Low measure

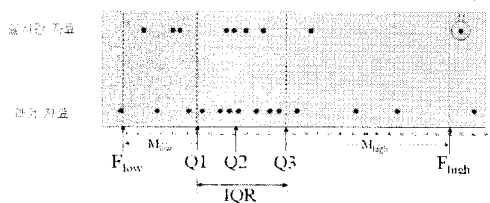
여기서 M_{high} , M_{low} 는 링크 별로 상이한 구조(교차로 신호형태나 횡단보도 등)와 교통 흐름에 유연하게 fence값을 조절할 수 있도록 만들어주는 링크 별 가중치 상수이다. 이 상수를 조정해가며 링크 별로 가장 효율적인 필터링 범위를 산정할 수 있다.

사분위수를 필터링 범위 선정에 쓴 이유는 이력 교통 정보에도 분명 오류 자료가 존재할 것인데, 사분위수는 일반적인 산술평균 같은 값들과는 달리 자료의 분포 정도를 이용해 구하는 값이기 때문에 이를 크게 반영하지 않는 장점이 있기 때문이다.

4.6 실시간 처리

5분의 단위 시간 별로 실시간으로 수집된 프로브 차량의 속도 자료는 해당 링크의 자료 수에 따라 그 값이 보정 또는 필터링 된다. 하나도 자료가 수집되지 않은 링크는 패턴DB에서 이력자료의 평균값으로 대체한다. 또한, 1~2개의 자료만이 수집된 경우에는 실시간 자료와 패턴DB의 평균속도를 조합하여 적어도 3개 이상의 자료가 확보될 수 있도록 조정한다.

3개 이상의 자료가 확보된 링크에서는 패턴 DB의 두 개의 fence 값과 비교되고, 만약 이상치 자료로 판별되었을 경우 즉, High fence보다 높은 속도나 Low fence보다 낮은 속도일 때 이는 제거 대상이 된다. 이러한 일련의 과정을 <그림 8>에 예로써 나타내었다.



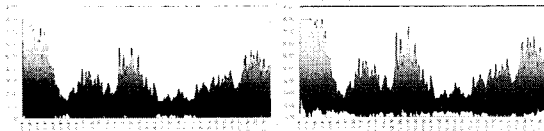
<그림 8> 링크 속도 필터링 과정 예

5. 실험 결과

6월 29일의 자료를 실시간으로 수집되는 속도 자료로 간주하여 3장에서 설명한 3가지 통계적인 방법과 제안하는 방법으로 각각 필터링 과정을 거쳐 링크 별 대표 속도를 산출하였다. 특히, 링크 별로 가중치 상수 M_{high} , M_{low} 를 여러 형태로 조정하며 링크 별로 가장 적합한 값을 산출하였다.

5.1 필터링 범위 산출 결과

<그림 9>는 M_{high} , M_{low} 를 각각 (1.0, 1.0), (2.0, 0.2)로 주었을 때, “한남대교남단-신사역” 링크의 필터링 범위이다.

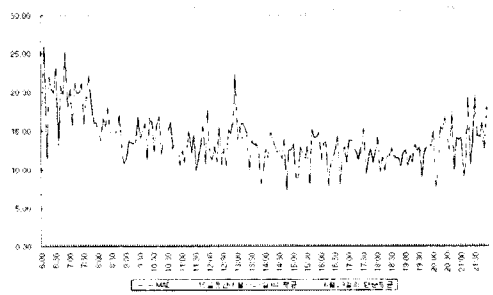


(a) (1.0, 1.0) (b) (2.0, 0.2)
 <그림 9> 1번 링크의 필터링 범위

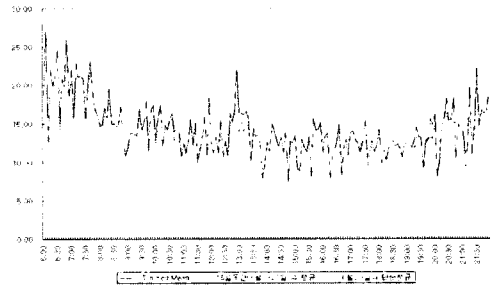
도심 내 위치한 복잡한 링크로 특히 첨두 시간대에서 대체로 필터링 범위가 좁고 낮은 것으로 나타났다. 이는 차량의 흐름이 원활하지 못해 많은 차량들이 정체되고 있는 것이라고 분석할 수 있다.

5.2 각 방법들을 적용한 필터링 결과

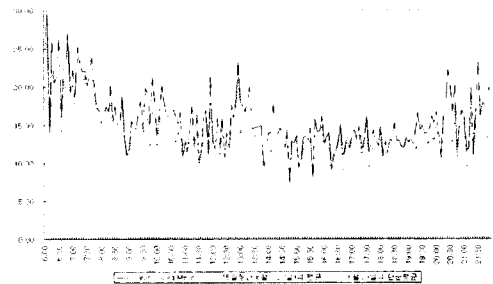
다음 <그림 10~13>에 앞에서 고찰한 3가지 필터링 방법과 본 논문에서 제시한 필터링 방법을 각 5개 링크에 대해 수행하여 전체의 평균 속도를 단위 시간 별로 나타낸 것이다. 비교를 위해 과거 15일 동안의 평균 속도와 필터링 과정을 거치지 않은 단순 평균 속도를 함께 그래프에 나타내었다. 여기서는 가중치 상수 M_{high} , M_{low} 의 값을 (1.5, 0.5)로 모든 링크에 설정하였다. 또한, Trimmed mean과 Winsorized mean 방법의 수준은 10%로 설정하였다.



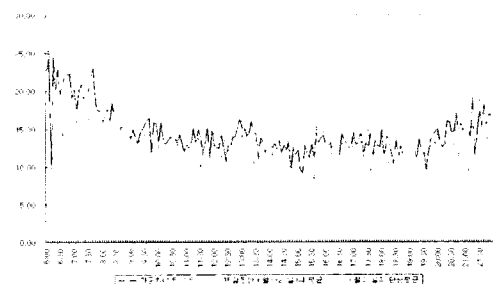
<그림 10> MAD(b=1, Zout=3)를 이용한 필터링 결과



<그림 11> 10% Trimmed mean을 이용한 필터링 결과



<그림 12> 10% Winsorized mean을 이용한 필터링 결과



<그림 13> 가중치(1.5, 0.5)를 적용한 제안 방법의 필터링 결과

시계열 자료인 교통 속도는 어떤 특수한 돌발 상황이 아니면 보통 시간에 따라 서서히 변화하는 흐름을 보이는 데, 원시 자료에는 이상치 자료들의 산재로 인해 몇몇 단위 시간에서 스파크 현상을 목격할 수 있었다. 제안하는 방법의 결과가 타 방법들에 비해 이러한 스파크를 제거하여 시간의 흐름에 따른 교통 흐름을 합당하게 나타낸다고 볼 수 있겠다.

5.3 비교분석

<표 2>는 2005년 6월 29일 단위 시간별로 수집된 속도 자료들을 기존의 3가지의 필터링 방법을 이용해 나온 대표값, 본 논문에서 제안한 필터링 방법을 통해 도출한 대표값을 패턴DB에 있는 15일 동안 취합된 단위 시간별 평균 속도값과 비교하여 나타낸 표이다. 패턴DB 내의 평균값은 비슷한 교통 흐름을 보이는 주중에서 축적된 15일간의 모든 프로브 차량들의 평균 속도이기 때문에 당일 해당 시간, 해당 링크에 어떠한 돌발상황이 없으면 일정 정도의 대표성을 인정받을 수 있을 것이다. 단, 실시간 처리 시 단위 시간 내 자료가 하나도 없을 경우나 2개 이하의 자료만이 수집되었을 경우에는 이 값으로 대표값을 정하거나 조합하기 때문에 3개 이상 자료가 확보된 경우에만 비교하였다.

해당 링크의 대표값 산출 방법은 이상치가 필터링 된 후 나머지 자료들의 산술 평균값으로 설정하였다. 비교 척도는 일반적인 예측 모형의 정확성을 추정하는 수치로써 대표값과의 전체적인 편차를 나타내는 MAE(Mean Absolute Error)를 사용하였다. MAE를 계산하는 식은 다음과 같다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}_i| \quad (7)$$

- 여기서, n : 자료의 수
- y_i : 비교의 기준이 되는 값(여기서는 15일 동안 취합된 단위 시간의 대표 속도)
- \bar{y}_i : 비교 대상이 되는 값(여기서는 각 방법으로 구해진 단위 시간의 대표 속도)

<표 2> 필터링 후 속도 값들과 대표값과의 MAE

번호	MAD	Trimmed	Winsorized	(1.0, 1.0)	(1.5, 0.5)	(2.0, 0.2)
1	1.70	1.83	2.06	1.29	1.28	1.26
2	1.21	1.28	1.37	0.85	0.88	0.90
3	2.58	2.63	3.16	1.35	1.49	1.70
4	2.07	2.26	2.81	1.18	1.16	1.33
5	1.91	1.92	2.40	0.62	0.62	0.64
전체	1.89	1.98	2.36	1.06	1.09	1.17

아울러 제안한 필터링 시스템의 가중치를 각각 (1.0, 1.0), (1.5, 0.5), (2.0, 0.2)로 변화시켜 가면서 해당 링크에 적합한 가중치를 찾으려고 노력하였고, 5개 링크 전체적으로는 (1.0, 1.0)으로 가중치를 주었을 때 가장 적합한 것으로 나타났다.

6. 결론

본 논문에서는 이력 자료에 사분위수를 활용해 GPS로부터 수집되는 교통 정보를 실시간으로 필터링 하는 방법에 관한 연구를 수행하였다. GPS를 통해 수집되는 교통 정보는 개인 프라이버시와 관리의 어려움으로 대중교통만을 통해 수집되고 있고 이런이유로 인해, 아직 타 검지 체계에 비해 수집량이 부족하고 많은 이상치 정보를 포함하는 특성을 가지고 있다. 기존에 교통 정보 필터링에 쓰이던 방법들은 주로 기초적인 통계 기법에 의해 이상 자료를 필터링하는 방법들로 실시간 수집 자료가 많지 않은 GPS 자료에는 적용하기 어려웠다. 이런 점에 착안하여, 본 논문에서는 GPS 교통 정보 특성에 맞게 이력 자료로 패턴을 구성하고, 자료의 분포 정도를 잘 나타내는 사분위수를 통해 해당 링크에 알맞은 필터링 범위를 선정하여 실시간 자료의 오류를 즉시 필터링 할 수 있는 방법을 제시하였다. 이는 기존의 교통 정보를 필터링 하는 방법들보다 더욱 적응성이 높은 방법이라고 할 수 있다. 제안한 방법의 성능은 실제 서울 강남대로의 현장 자료를 통한 테스트 결과에서 기존의 필터링 기법들보다 더욱 우수한 결과를 도출함이 증명되었다.

또한, 본 시스템은 이상 자료를 필터링 하기 위한

기준으로 두 개의 fence 값만 계산해 놓으면, 실시간 자료의 이상 여부를 즉시 판단할 수 있는 방법으로 이루어져 실시간 시스템에 바로 적용하기에도 무리가 없을 것이다.

최근 좀처럼 해결될 것처럼 보이지 않던 우리나라의 고질적인 교통 혼잡 문제가 중앙 정부와 지방자치단체들의 꾸준한 노력과 텔레매틱스나 ITS와 같은 기술의 발전에 힘입어 조금씩 나아지는 추세를 보이고 있다. 앞으로 텔레매틱스 서비스가 본격적으로 활성화되기 시작하면 차량 내에서 보내는 시간이 적지 않은 많은 사람들에게 낭비되는 시간을 줄여줄 수 있고, 도로 위에서 보내는 시간도 헛되어 허비하지 않도록 많은 도움을 가져다 줄 것이다. 본 연구를 통해 텔레매틱스 서비스 중 가장 핵심이 되는 교통 정보 제공 서비스에서 보다 정확한 교통 정보를 생성하는데 도움이 되었으면 하는 바람이다.

향후 연구 계획은 단위 시간에 실시간 자료 결측시 이를 보다 정확하게 해결할 수 있는 방안에 대한 모색과 주말, 공휴일 등 특이한 패턴을 보이는 날에도 적용이 가능하도록 시스템을 보완하는 것이다. 또한, 링크의 특성과 시간대별 차량 흐름을 보다 다양하게 분석하여 시스템이 자동으로 가장 적합한 가중치를 설정할 수 있도록 하는 방법에 대한 연구가 이루어지면 더욱 효율적인 시스템이 될 것으로 생각한다. 마지막으로 결과의 비교분석에서 실제 도로를 주행하는 차량들과의 비교분석을 하지 못한 점은 아쉬움으로 남는다. 좀 더 정확한 비교분석을 위해 향후에는 테스트 차량을 운용하여 실제 도로 위의 차량들의 정보들과 비교분석 해 보는 작업도 진행할 예정이다.

참고문헌

1. 임은모, “이제 텔레매틱스가 자동차문화를 바꾼다”, 진한도서, 2002. 1.
2. 국토연구원, “텔레매틱스시대에 대비한 교통정보 통합기반 필요”, 국토정책Brief, 제72호, 2004. 9.

3. 최기주, 신치현, “GPS와 GIS를 이용한 링크통행 시간 예측기법”, 대한교통학회지, 제16권 제2호, 1998, pp. 157-167.
4. Choi, Keechoo, “Data Fusion Methodology for Link Travel Time Estimation for Advanced Traveler Information System”, KSCE Journal of Civil Engineering, Vol. 3, No. 1, 1993. 3, pp. 1-14.
5. 이영우, “실시간 GPS 수집 자료를 이용한 링크통행시간 선정에 관한 연구”, 대한토목학회 논문지, 제24권 제2D호, 2004. 3, pp. 125-131.
6. 오순미, 이영인, 백승걸, 남궁성, “FTMS 상의 루프검지기 오류자료 판단과정 개선에 관한 연구”, 대한교통학회 제52회 학술발표회, 2006. 6.
7. 성호준, 김기도, “NMEA-0183을 이용한 GPS 애플리케이션 제작”, 마이크로소프트웨어, 2005. 5, 6.
8. 서울특별시 교통국, “2005년도 서울시 차량통행속도 보고서”, 2006.
9. Park, DongJoon, and Hwang, HyunMi, “A Note on Quartile”, Journal of the Korean Society for Quality Management, 1998. 9.
10. Vicek, C., “Faster Real-Time Dispatch Using GPS-based Automatic Vehicle Location”, Proceedings of the IEEE IEE Vehicle Navigation and Information Systems Conference, 1993, pp. 661.
11. 정우진, 이종수, “GPS정보를 이용한 교통상황 판단 알고리즘 개발”, Proceedings of KFIS Fall Conference, 2003. 5, pp. 257-260.
12. Colombo, L., Greenbelt, A., and Sutter, A., “Evaluation of Precise, Kinematic GPS Point Positioning”, Proceedings ION GNSS, 2004. 9.
13. Ko, SY., “Optimal Numbers of probe cars beacon-based Travel Time Data Collection System”, WCTR, 2001.
14. Stanislaw, B., Condie, H., and Sheffey,

A., "ADVANCE Evaluation Detector Travel Time Conversion and Fusion of Probe and Detector Data", University of Illinois at Chicago, 1996.

15. Turochy, R., and Smith, B., "New Procedure for Detector Data Screening in Traffic Management Systems" , Transportation Research Record 1727, 2000.

최진우

2005년 경원대학교 컴퓨터공학과 학사
2007년 경원대학교 전자계산학과 석사
2007~현재 경원대학교 대학원 전자계산학과 박사과정
관심분야 : 텔레매틱스, ITS
e-mail : cjw49@paran.com

양영규

1972년 서울대학교 학사
1974년 서울대학교 환경대학원 환경계획 석사
1984년 Texas A&M University 공간정보처리 박사
2003년~현재 경원대학교 소프트웨어학부 교수
관심분야 : 공간정보처리, 텔레매틱스
e-mail : ykyang@kyungwon.ac.kr