

■ 論 文 ■

**기회손실비용을 고려한 버스 운행시격과
링크 통행시간 예측 알고리즘**

An Opportunity Cost Based Headway Algorithm in Bus Operation

이 영 호

(고려대학교 산업공학과 부교수)

조 현 성 · 김 영 진

(고려대학교 대학원 산업공학과)

안 계 형 · 배 상 훈

(교통개발연구원 책임연구원)

목 차

- | | |
|----------------------------|-------------|
| I. 서론 | 2. 2단계 예측기법 |
| II. 최적 운행시격 결정 | 3. 자료융합 단계 |
| 1. 버스 정류장이 하나인 모형 | 4. 전문가 시스템 |
| 2. 버스 정류장이 여러 개인 모형 | 5. 사례 연구 |
| III. 통행시간 예측을 위한 자료융합 알고리즘 | IV. 결론 |
| 1. 온라인 유고상황 감지 | 참고문헌 |

요 약

이 연구는 버스정보 시스템 설계에 필요한 운행시격 결정과 통행시간 예측을 위한 알고리즘 개발을 다룬다. 운행시격 결정 문제는 버스와 같은 대중교통 수단을 운영하는데 중요한 요소 중에 하나이다. 기존 연구는 버스 운행비용과 승객비용의 합을 최소로 하는 운행시격을 찾는 데 초점을 두고 있다. 이때 승객비용이란 승객 대기비용과 승객 교통비용의 합으로 이루어진다. 그런데 우리 나라와 같이 버스회사 수입이 전액 운행수입에만 의존하는 경우엔 이러한 접근 방식이 타당하지 않다. 기존의 방식과 다르게 승객비용으로 승객 이탈비용을 사용하여 버스의 최적 운행시격을 구하는 것이 이 연구의 목적이다. 먼저 정류장이 하나인 경우에 대해 해석적 방법으로 풀고, 정류장이 여러개인 경우에 대해서는 시뮬레이션 기법을 적용한다. 또한 이 연구는 신뢰성이 높고 정확한 통행시간 예측정보를 산출하기 위해 2 단계 예측 기법과 전문가시스템을 이용하는 자료융합 알고리즘을 개발한다. 정확한 정보를 제공하려면 교통정보 수집원을 통해 얻는 자료가 정확해야 하고, 또한 교통상황 변화에 따라 실시간으로 통행시간을 예측하는 것이 필요하다. 이 연구는 AVL(Automatic Vehicle Location)시스템을 이용한 버스정보시스템에서 실시간 데이터와 과거 데이터를 융합하여 통행시간을 예측하는 알고리즘을 개발한다. AVL 데이터를 수집하는 과정에서는 경제성을 고려하여 데이터를 수집한다. 그리고, 버스의 운행관리와 정확한 도착예정시간을 예측하기 위해 AVL시스템을 통해 얻은 데이터의 패턴을 분석하고 유고상황을 감지한다.

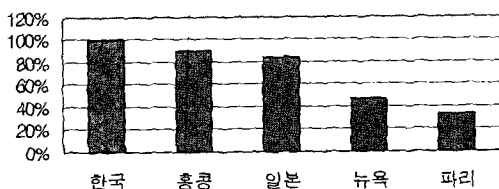
1. 서론

이 연구는 AVL(Automatic Vehicle Location) 시스템을 이용한 버스정보시스템 설계에 필요한 운행시각 결정과 통행시간 예측을 위한 알고리즘을 개발한다. 운행시각은 버스를 운행하는데 결정해야 할 중요한 요소이다. 운행시각은 직접적으로 버스보유대수와 운행비용을 결정한다. 반면에 통행시간에 대한 신뢰성 있는 예측이 운행 시각 결정에 지대한 영향을 끼치므로 이에 대한 연구 결과를 제시한다.

이러한 운행시각을 포함한 버스운행에 대하여 해석적인 방법과 시뮬레이션 기법을 이용한 방법 등 많은 연구가 이루어졌다. 운행시각 결정 문제에 대한 고전적 틀을 제시한 연구로 쟈슨(Jansson)⁷⁾의 연구를 들 수 있다. 그는 승객 대기시간 비용, 승객 통행시간 비용, 버스 운행비용을 포함하는 총사회비용 모형을 세웠다. 이를 통해 제곱근 공식을 이용하여 최적 운행시각을 구하고, 이 최적 운행시각과 혼잡시간 때 수요를 가지고 최적 버스크기를 구하였다. 특히 그가 사용한 운행비의 선형할수 모형은 최근에도 많이 사용한다. 국내 연구로 고승영과 고종섭¹⁾의 버스 최적 운행시각과 보유대수에 대한 해석적 모형에 대한 연구를 들 수 있다. 이들은 요일별, 시간대별로 변하는 시내버스의 이용자 수요 및 버스운행 소요시간을 고려하여 총 교통 비용 모형을 수립하였다. 총 교통 비용모형은 운영자 비용과 이용자 비용의 합으로 이루어 지는데, 이용자 비용에 승객 대기시간을 고려하여 이 비용이 최소가 되는 운행시각을 찾았다.

이와 같이 버스의 운행간격에 대한 연구는 버스가 대중교통 수단이라는 관점에서 버스 운행비용, 이용자의 대기시간 등을 포함하는 총 사회비용을 최소화 하는 것을 목적으로 삼았다. 그러나 <그림 1>에서처럼 버스회사가 전적으로 요금에 수익을 의존하고 있

요금의존도(%)



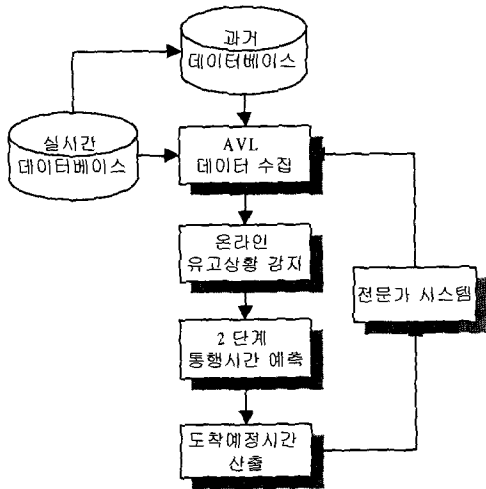
<그림 1> 시내버스 수입에서 요금 의존도²⁾

는 우리나라의 경우엔 기존 연구와는 다른 모형이 필요하다. 즉 버스회사의 관점에서 회사의 이익을 최대가 되게 하는, 혹은 비용이 최소가 되게 하는 운행시각을 찾아야 한다. 운행시각이 길어지면 버스를 운행하는데 드는 비용은 줄어들지만 반면에 그 버스를 이용하려는 승객은 줄어들게 된다. 이러한 이탈은 버스회사 관점에서 기회손실비용을 발생시키게 된다. 따라서 이 연구에서는 버스 운행비용과 승객이탈에 의한 기회손실비용의 합을 총비용으로 사용한다. 이 총비용을 가장 적게 하는 운행시각이 버스회사 관점에서 최적 운행시각이 될 것이다.

또한 이 연구는 운행시각 결정에서 핵심인 구간통행시간 예측을 위한 자료융합 알고리즘을 개발한다. 기존의 교통정보는 대부분 검지기, 비콘, CCTV, 제보자료 등을 이용한다. 하지만 버스가 정류장에 도착할 시간을 예측하기 위해 모든 버스 노선에 검지기, 비콘, CCTV를 설치하는 것은 막대한 비용이 든다. 특히 검지기는 유지 및 보수를 하기가 어렵다. 이러한 비용문제를 해결하기 위한 방법으로 AVL시스템을 이용할 수 있다. AVL시스템은 버스에 장착하므로 유지 및 보수가 용이하다. 그리고, GPS(Global Positioning System)시스템을 통해 운행 중인 버스의 현재 위치와 시각을 구하고, 이를 이용해 버스가 정류장에 도착할 시간을 구할 수 있다.

이 연구는 버스의 도착시간을 예측하기 위해 실시간 데이터와 과거데이터를 융합한다. 자료융합에 대한 기존 연구는 다음과 같다. 넬슨(Nelson)⁹⁾과 라우펠(Rouphail), 타코(Tarko)¹⁰⁾는 온라인 자료융합을 위해 신경망모형을 제안했다. 그리고, 라우펠(Pouphail)과 타코(Tarko)^{11),13)}는 오프라인 자료융합에 회귀분석과 베이지안 이론을 이용하여 링크 통행시간 데이터를 융합하였다. 베이지안 이론은 조사차량의 데이터와 검지기 데이터를 융합할 때와 실시간 데이터와 과거 데이터를 융합할 때 사용하였다.

<그림 2>를 통해 이 연구의 자료융합 알고리즘의 단계를 살펴보면, 먼저 실시간 데이터와 과거 데이터를 통해 AVL데이터를 수집한다. 실시간 데이터는 GPS시스템을 통해 중앙 센터에 보낸 버스의 현재 위치 데이터와 시각 데이터를 말한다. 그리고, 과거 데이터는 실시간 데이터를 누적하여 생성한다. AVL 데이터 수집은 경제성을 고려하여 AVL 데이터의 샘플링 회수를 결정한다. 다음으로, 온라인 유고상황 감지는



〈그림 2〉 자료융합 알고리즘 단계

실시간으로 샘플링한 AVL 데이터와 과거 데이터를 비교하여 유고 상황을 감지하는 단계이다. 그리고, 2 단계 통행시간 예측은 AVL 데이터 예측과 AVL 데이터를 이용한 통행시간 예측으로 나누어 전개한다. 도착 예정시간 산출은 2 단계 통행시간 예측의 결과를 통해 버스가 정류장에 도착할 시간을 산출한다. 마지막으로, 전문가 시스템은 알고리즘의 정확성을 위해 AVL 데이터의 샘플링 변수를 조정하는 단계이다.

이 논문의 구성은 다음과 같다. 먼저 II장에서 최적 운행시각 결정에 관한 문제를 다룬다. 정류장이 하나인 경우에 대해 해석적 모형을 세워 총비용 함수에 대해 분석하고, 정류장이 여러 개인 경우에 대해 몬테카를로(Monte Carlo) 기법을 적용한 시뮬레이션 모형을 세워 분석한다. 그리고 III장에서는 통행시간 예측정보를 산출하기 위해 2 단계 예측 기법과 전문가시스템을 이용하는 자료융합 알고리즘을 개발한다. 끝으로 IV장에서 결론을 맺는다.

II. 최적 운행시각 결정

최적 운행시각을 산출하기 위한 총비용은 버스 운행비용과 승객 비용의 합으로 설정한다. 즉,

$$TC = C_o + C_{ce}$$

여기서, TC : 총비용

C_o : 버스 운행비용

C_{ce} : 승객 비용

버스 운행비용은 버스를 운행하는데 소요되는 비용을 말한다. 그리고 승객 비용은 운행시각을 크게 할 때 발생하는 승객이탈로 인한 기회손실비용으로 정의한다. 먼저 버스 정류장이 하나인 모형에서 비용 함수의 특징에 대해 살펴보고, 정류장이 여러 개인 모형에서 시뮬레이션을 통해 비용함수를 살펴보고자 하자.

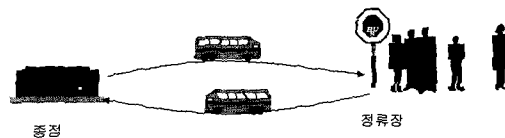
1. 버스 정류장이 하나인 모형

정류장이 하나인 경우란 〈그림 3〉과 같이 버스가 종점에서 출발하여 정류장에서 승객을 싣고 다시 돌아오는 경우이다. 버스는 일정한 운행시각으로 종점을 출발하며 정류장에서 승객을 태우고 다시 종점으로 돌아온다. 승객은 정류장에 도착해서 버스가 있으면 버스를 타고 일정한 시간이 지났는데도 버스가 도착하지 않으면 이탈한다. 종점에서 정류장까지 걸리는 시간은 정규분포를 따르고 승객이 도착하는 시간간격은 지수분포를 따른다고 가정한다. 그리고 승객이 버스를 타는데 걸리는 시간은 승객마다 일정하다고 가정한다.

먼저 버스 운행비용을 살펴보자. 버스를 운행하는데 소요되는 비용은 상당히 많은 항목으로 구성되어 있다. 그러나 여기서는 승객 이탈비용과 비교하기 위해 실제로 운행하고 있는 버스에서 발생하는 비용만을 고려한다. 이 비용은 버스가 실제로 운행한 시간에 비례하는 것으로 가정한다. 대상 시간동안 운행하는 버스의 대수는 대상시간을 운행시각으로 나눈 값이 되며, 이 값에 종점에서 정류장까지 가는 시간과 정류장에서 종점까지 오는 시간을 합한 값을 곱하게 되면 운행비 산출을 위한 총 운행시간이 된다. 따라서 버스 운행비용은 다음과 같이 정의할 수 있다.

$$C_o = (\mu_f + \mu_b) \times \frac{T}{h} \times a_o$$

여기서, h 가 운행시각(분)을 가리키며, T 는 대상 운행시간을 가리킨다. 그리고 μ_f 는 종점에서 정류장



〈그림 3〉 정류장이 하나인 모형

으로 가는데 걸리는 평균시간(분)이며 μ_b 는 정류장에서 종점으로 돌아오는데 걸리는 평균시간(분)이다. a_o 는 운행비용(원/대-분)을 나타낸다.

승객 이탈은 승객이 도착하여 대기 한계시간을 넘도록 버스가 도착하지 않을 때 발생한다. 이 값의 평균에 이 시간대가 발생하는 회수를 곱하면 승객이탈이 발생하는 총시간을 산출할 수 있다. 이 시간대가 발생하는 회수는 바로 버스를 운행하는 횟수이며 이는 대상시간을 운행시격으로 나눈 값과 일치한다. 여기에 승객 도착률과 승객 이탈 기회손실 비용을 곱하면 승객 이탈비용을 산출할 수 있다. 즉 승객 이탈비용은 다음과 같은 식으로 표현할 수 있다.

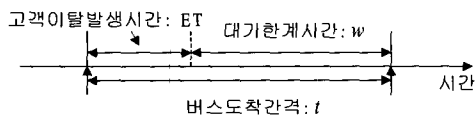
$$C_{ce} = \frac{T}{h} \times E(ET) \times \lambda \times a_c$$

여기서, ET 는 승객이탈이 발생하는 시간이고 λ 는 승객의 도착률, 그리고 a_c 는 승객이탈 기회손실 비용을 나타낸다.

승객 이탈이 발생하는 시간(ET)의 평균값을 구하여 보자. 버스가 도착하는 시간간격을 t 라고 하고 대기한계시간을 w 라고 하자. t 와 w 는 모두 확률변수이다. 이때 t 는 운행시격과 이전 버스가 걸린 통행시간과 이번 버스가 걸린 통행시간을 통해 산출할 수 있다. 즉,

$$t = h + t_i - t_{i-1}$$

여기서, t_i 는 버스 i 가 정류장까지 가는데 걸린 시간이다. 버스가 종점에서 정류장까지 가는데 걸리는 시간이 정규분포를 따른다고 가정하고 모든 버스에 대해 걸리는 시간은 독립이라고 가정하면 t 의 분포를 구할 수 있다. 정류장까지 가는 통행시간의 분산을 σ^2 라고 가정하면 t 의 분포는 정규분포 $N(h, 2\sigma^2)$ 를 따른다. 고객 이탈이 발생하는 시간은 <그림 4>와 같은 상황에서 발생하는데 이를 통해 산출식을 세우면 다음과 같다.



<그림 4> 고객이탈 발생시간

$$E(ET) = \int_0^{\infty} \left[\int_w^{\infty} (t-w)f(t)dt \right] g(w)dw$$

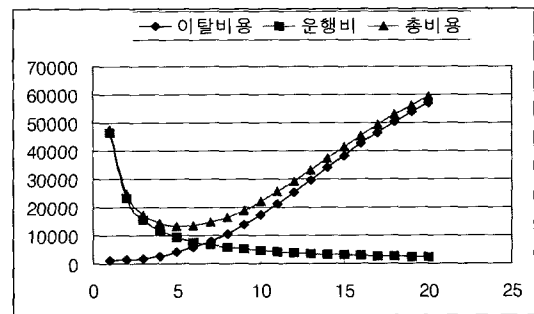
여기서, $f(t)$ 는 t 의 확률밀도함수이고 $g(w)$ 는 w 의 확률밀도함수이다.

해석적 방법을 통해 각 비용함수에 대한 수식을 유도하였지만, 이를 통행 운행시격(h)을 바로 구할 수는 없다. 다음 예를 가지고 비용함수가 가지는 특징에 대하여 살펴보자.

평균승객 발생률(λ)을 0.5분당 한 명으로 가정하고 승객대기한계시간은 균일분포(하한: 6분, 상한: 15분)를 따른다고 가정하자. 종점에서 정류장까지 걸리는 시간(μ_f)은 정규분포 $N(3.1, 0.28)$ (분)을 따르며, 정류장에서 종점까지 가는데 걸리는 시간(μ_b)은 정규분포 $N(3.2, 0.29)$ (분)을 따른다고 가정한다.

비용계수 중에서 운행비용(a_o)은 3,676(원/시)으로 하였는데, 이는 1997년 '서울시 버스개선 기획단'에서 버스개혁 종합대책³⁾의 일환으로 조사한 버스회사의 항목별 운영비 자료에서 인용하였다. 그리고 승객이탈비용(a_c)은 승객 한 명을 태우지 못함으로 인한 기회손실비용으로 간주하여 현재 성인 버스요금인 500(원/명)으로 하였다.

<그림 5>에서 보이는 바와 같이 총비용함수는 완벽한 오목(Convex) 형태를 취하지는 않는데, 이는 이탈 비용곡선이 완벽한 오목함수 형태가 아니라 운행시간간격이 길어질수록 직선에 가까운 형태가 되기 때문이다. 그러나 운행시격에 따라 운행비용은 점차 줄어들며 이탈비용이 점차 증가하는 형태를 가지므로 총비용 함수는 최소값 하나를 가지는 함수라는 것을 알 수 있다. 앞에서와 같이 가정한 상황에서는 최적 운행시격이 5분으로 나타났다.



<그림 5> 정류장이 하나인 경우 비용곡선

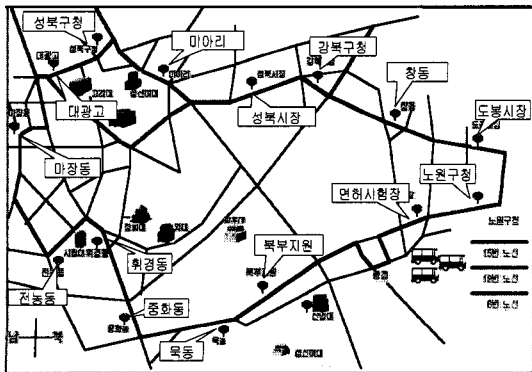
2. 버스 정류장이 여러 개인 모형

정류장이 여러 개인 경우에는 시뮬레이션 방법을 적용한다. 가정은 정류장 하나인 경우와 같으며 단지 정류장이 하나에서 여러 개로 늘어난 것이 차이점이다.

시뮬레이션은 다음 사건을 추적하는 방식(Next-event Time-advance Approach)으로 진행한다. 시뮬레이션에 포함된 사건은 4가지로 이루어져 있다. 첫째는 버스를 운행시격으로 발생시키는 사건이다. 이 사건에서는 버스가 첫 번째 정류장에 도착할 시간을 계산한다. 둘째는 승객을 발생시키는 사건이다. 승객발생이 지수분포를 따르므로 승객도착률 총합으로 승객을 발생시켜 각 정류장으로 분산시킨다. 셋째는 버스가 정류장에 도착하는 사건이다. 정류장에 도착한 버스는 현재 정류장에 있는 승객 중 대기한계시간이 지나지 않은 승객을 태운다. 동시에 대기한계시간이 지나 이탈한 승객의 숫자를 계산한다. 넷째는 버스가 정류장을 출발하는 사건이다. 다음 정류장을 도착할 예정시간을 계산한다.

시뮬레이션 분석에 사용된 자료는 서울시 15번 노선에 대한 것으로 하였다. 15번 노선은 <그림 6>와 같이 종점에서 출발하여 다시 종점으로 돌아오는 순환형 버스이다. 구간 통행시간에 대한 자료는 타코(Tacho)를 통해 얻은 자료이며, 탑승객 수는 서울시 버스개선 기획단에서 버스개혁 종합대책의 일환으로 조사한 자료를 근거로 하였다.

<표 1>은 타코가 설치되어 있는 정류장과 정류장에 도착한 승객수를 나타낸다. 시간은 오후 혼잡시간



<그림 6> 15번 버스 노선도와 주요 정류장

대인 6시에서 8시 사이이며 타코가 설치되어 있지 않은 정류장 승객은 분산시켜 타코가 있는 정류장에 추가하였다. <표 2>는 15번 버스의 타코 자료로 1998년 9월 2일 오후 6시에서 오후 8시 사이 15대 버스가 운행한 자료를 근거로 각 정류장 사이 통행시간에 대해 산출한 자료이다. 정류장 번호는 <표 1>의 순서에 맞추어 종점에서 먼허시험장을 1번으로 하여 순서대로 매겼다. 승객 대기한계시간은 정류장이 하나인 경우와 같은 균일 분포를 가정하였다.

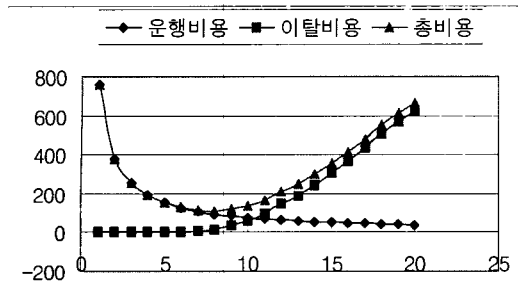
이러한 자료를 근거로 <그림 7>와 같은 결과를 얻을 수 있었다. 그림에서 보는 바와 같이 운행비용은 운행시격이 길어짐에 따라 줄어들는데, 이는 혼잡시간대 2시간 동안 운행하는 버스대수가 줄어들기 때문이다. 또한 이탈비용은 버스 운행시격이 길어짐에 따라 승객이 기다리는 시간이 길어지므로 이에 따라 이탈고객도 많이 발생하게 되어 그림과 같은 형태를 띠게 되었다. 그 결과 총비용이 최소인 최적 운행시격은 8분으로 나타났다.

<표 1> 정류장별 평균 탑승객 수 (단위:명)

정류장	면허장	노원구청	도봉시장	창동
탑승객수	27	13.3	18.5	9.3
정류장	수유역	성북시장	미아리	성북구청
혼잡시간	18.3	4.5	1.8	10.8
정류장	대광고	마장동	진농동	황경동
혼잡시간	3.5	5.5	4.5	3.0
정류장	중화동	목동삼거리	북부지원	종점
혼잡시간	2.3	11.5	2.3	6.5

<표 2> 혼잡시간 때 링크 통행시간 (단위:분)

링크번호	1	2	3	4
평균	8.3	5.8	3.1	10.6
분산	0.52	0.74	0.35	2.83
링크번호	5	6	7	8
평균	9.0	4.9	8.6	6.3
분산	0.86	0.50	0.54	0.64
링크번호	9	10	11	12
평균	2.7	9.4	8.9	12.6
분산	0.24	2.26	4.21	1.83
링크번호	13	14	15	16
평균	4.5	8.4	2.6	5.5
분산	0.55	1.83	0.40	0.41



〈그림 7〉 정류장이 여러 개인 경우 비용합수

III. 통행시간 예측을 위한 자료융합 알고리즘

지금까지는 버스의 최적 운행시격에 대해 설명하였다. 다음으로, 버스 운행시격에서 핵심정보인 정류장에 도착할 시간을 예측하는 알고리즘을 살펴보자. 통행시간 예측 알고리즘에서는 먼저, AVL시스템을 통해 실시간으로 유고 상황을 감지하는 온라인 유고상황감지에 알아본다. 그리고, 버스가 정류장에 도착하는 시간을 예측하는 2단계 예측 기법과 실시간 데이터와 과거 데이터를 융합하는 자료융합 단계를 설명한다. 마지막으로, 샘플링 변수를 조절하는 전문가 시스템에 대해 살펴본다.

1. 온라인 유고상황 감지

온라인 유고상황 감지에서는 AVL시스템을 이용해 실시간으로 유고상황을 감지하여 버스 노선의 이상 유무를 판단한다. AVL 데이터는 GPS시스템을 통해 획득한다. AVL시스템을 장착한 버스는 GPS시스템을 통해 차량의 위치와 시각 데이터를 획득할 수 있다. 이렇게 얻은 데이터는 중앙센터에 보내진다. 중앙센터에서는 샘플링한 AVL데이터를 이용해 버스노선의 교통 상태를 실시간으로 파악한다. 이러한 AVL데이터를 통해 수집한 교통정보는 데이터베이스에 저장되어 버스노선의 과거 링크 통행시간을 구하는데 사용한다.

여기서, AVL데이터를 샘플링하는 방법을 살펴보자. AVL데이터를 샘플링하는 방법에서는 정확성과 경제성을 고려한다. 정확성이란 측면에서 보면, 운행중인 버스의 위치좌표와 시각 데이터를 빈번히 보내면 운행중인 버스의 운행 속도를 정확히 파악할 수 있다. AVL데이터는 무선 통신을 이용해 중앙센터에 보내진

다. 하지만, 중앙센터에 많은 데이터를 보낼수록 통신 비용이 증가한다. 반면, AVL데이터를 긴 시간간격을 가지고 보내면, 통신 비용은 줄어들지만 정확성이 떨어진다. 따라서 통신비용을 줄이면서 정확성을 높이는 샘플링 방법을 결정해야 한다. 이 문제를 해결하기 위해 다음과 같은 샘플링 방법을 이용한다. 첫째, 정확성을 위해서 버스가 정류장에 도착하면 항상 AVL데이터를 샘플링한다. 그 이유는 정류장 단위로 버스정보가 필요하기 때문이다. 예를 들면, 승객이 필요한 정보는 현재 운행 중인 버스가 다음 정류장에 도착할 시간이다. 둘째, 경제성을 위해서 버스가 정상적인 운행상태에 있는 경우와 비정상적인 상태에 있는 경우를 구분하여 샘플링 한다. 정상상태에 있는 경우는 버스가 정류장에 도착할 때만 샘플링을 하여 정류장 도착예정시간을 구한다. 하지만, 버스가 비정상적인 운행 상태에 있는 경우는 샘플링을 빈번히 하도록 한다. 여기서, 비정상상태에 있다는 것은 버스가 심한 정체나 사고 등으로 과거 평균 시간보다 늦게 도착하는 경우를 말한다.

이제, AVL데이터를 이용해 유고상황을 감지하는 방법을 자세히 살펴보자. 먼저, 유고상황을 감지하기 위한 샘플링을 하는 방법을 설명하기 위해 T_i , TP_i , k_i , m_i , σ_i 를 정의하자.

T_i : 링크 i 의 현재 통행시간

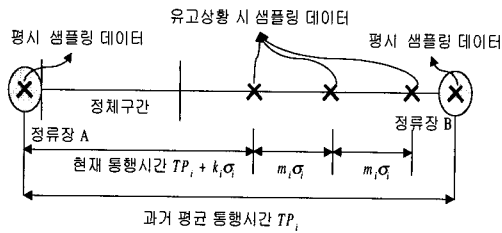
TP_i : 링크 i 의 과거 평균 통행시간

k_i : 링크 i 에서 AVL데이터 샘플링을 시작하기 위한 파라미터

m_i : 링크 i 에서 AVL데이터 샘플링을 계속하기 위한 파라미터

σ_i : 링크 i 에서 과거 통행시간의 표준편차

평상시는 AVL데이터를 버스가 정류장에 도착하면 수집한다. 하지만, 유고상황이 발생한 경우는 정확한 교통정보를 얻기 위해 정류장과 정류장 사이에서도 샘플링을 한다. 그렇다면, 얼마나 자주 샘플링하는 것이 유고 상황에서 정확한 교통정보를 얻을 것인가? 우선 유고상황이 발생했다는 것을 파악하기 위해 과거 평균 통행시간(TP_i)을 활용한다. 〈그림 8〉은 유고상황이 발생했을 경우 샘플링 데이터를 보여 주는데, 운행 중인 버스가 정체 상황으로 인해 과거 평균 통행시간보다 늦게 정류장에 도착하면 샘플링을 시작



(그림 8) 유고상황 발생시 샘플링 데이터

한다. 즉, 현재 통행시간 T_i 가 $TP_i + k_i\sigma_i$ 보다 크면 샘플링을 시작한다. 그리고, 현재 통행시간 T_i 가 $TP_i + k_i\sigma_i + m_i\sigma_i$ 보다 클 때까지도 도착하지 않으면 또 샘플링을 한다. 이와 같이 버스가 다음 정류장에 계속 도착하지 못하면 $m_i\sigma_i$ 간격으로 계속 샘플링을 한다.

2. 2단계 예측기법

2단계 예측기법은 샘플링한 데이터를 이용해 버스가 정류장에 도착할 시간을 예측하는 단계와 샘플링 데이터 자체를 예측하는 단계로 나누어진다.

먼저, 1단계 예측은 현재 운행중인 버스가 유고상황이 발생한 경우에 샘플링한 데이터를 이용한다. 이 샘플링 데이터를 이용하여 도착예정시간을 예측하는 기법이다. 예측 방법은 아래 2차항 회귀식을 이용한다.

$$Y_i = \beta_0 + \beta_1 X_i^2 + \beta_2 X_i + \epsilon_i \quad i = 1, \dots, n$$

i : 현재 버스의 샘플링 데이터 순서

Y_i : 이전 정류장에서 현재 버스 위치 사이의 구간 속도와 과거 평균속도의 차이

X_i : 이전 정류장에서 현재 버스 위치 사이 거리

$\beta_0, \beta_1, \beta_2$: 추정할 모수

ϵ_i : $E(\epsilon_i) = 0, V(\epsilon_i) = \sigma^2$, 서로 다른 i, j 에 대해 $cov(\epsilon_i, \epsilon_j) = 0$. 정규분포를 따르는 오차항

여기서, 2차항에 대한 모형이 1차항 모형에 비해서 우세한 이유는 유고시 현재 버스 위치 사이의 구간 속도와 과거 평균속도의 차이는 구간 거리에 영향을 받기 때문이고, 오차에 대한 정확도를 향상하기 위해서 2차항을 도입하여 보정하였다.

다음으로, 2단계 예측은 샘플링 데이터 자체를 예측하는 것으로, 예측하는데 이용하는 데이터는 앞 버스의 샘플링 데이터이다. 샘플링 데이터를 예측하는 방법은 다음과 같다. 유고가 발생한 후에 첫 번째로 샘플링할 데이터를 미리 예측하기 위해서는 앞 버스의 첫 번째 샘플링 데이터를 이용한다. 마찬가지로 두 번째로 샘플링할 데이터를 예측하기 위해서는 앞 버스의 두 번째 샘플링 데이터를 이용한다. 여기서, 첫 번째 샘플링 데이터를 예측한다는 것은 미래 시간 ($TP_i + k_i\sigma_i$)에 버스가 있을 위치를 예측하는 것이다. 그리고, 두 번째 샘플링 데이터를 예측한다는 것은 미래시간 ($TP_i + k_i\sigma_i + m_i\sigma_i$)에 버스가 있을 위치를 예측하는 것이다. i 번째 샘플링 데이터를 예측하는 회귀식은 아래와 같다.

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \epsilon_{ij} \quad j = 1, \dots, n$$

j : 앞에 지나간 버스의 인덱스

Y_{ij} : 앞에 지나간 버스 j 가 i 번째 샘플링을 했을 때, 이전 정류장에서 버스 j 위치사이의 구간 속도와 과거 평균속도의 차이

X_{ij} : 앞에 지나간 버스 j 가 i 번째 샘플링을 했을 때, 이전 정류장에서 버스 j 위치사이의 거리

β_0, β_1 : 추정할 모수

ϵ_{ij} : $E(\epsilon_{ij}) = 0, V(\epsilon_{ij}) = \sigma^2$, 서로 다른 k, l 에 대해 $cov(\epsilon_{ik}, \epsilon_{il}) = 0$, 정규분포를 따르는 오차항

3. 자료융합 단계

자료융합 단계는 2단계 예측기법을 샘플링 회수에 따라 다르게 적용하는 단계이다. 자료융합 단계는 유고상황이 발생한 링크에 진입한 후에 통행시간을 예측하는 단계와 유고상황이 발생한 링크에 진입하기 전에 통행시간을 예측하는 단계로 나누어진다. 먼저, 유고링크 진입 후 통행시간을 예측하는 단계는 다음과 같다.

[단계 1] 앞에 지나간 버스가 유고 샘플링을 안 한 경우.

1-1. 현재 버스의 유고 샘플링이 1 개인 경우 : 유고

샘플링 데이터 수가 1개이므로 샘플링데이터 발생지점과 바로 전 정류장사이의 구간속도를 이용해 다음 정류장에 도착할 예정시간을 구한다. 다음 정류장까지는 현재 구간속도와 같은 속도로 도착할 것으로 가정한다.

- 1-2. 현재 버스의 유고 샘플링이 2개 이상인 경우 : 샘플링 데이터를 이용하여 거리에 따른 속도차이 곡선을 추정한다. 즉, 1단계 예측을 적용한다. 추정한 곡선을 이용해 다음 정류장에 도착할 시간을 예측한다.

[단계 2] 앞에 지나간 버스가 유고 샘플링을 한 경우 (두대 이상) : 2단계 예측을 적용하여 유고 샘플링 데이터를 예측한다. 예측한 데이터를 포함한 샘플링 데이터를 이용하여 1 단계 예측을 적용한다. 추정한 곡선을 이용해 다음 정류장에 도착할 시간을 예측한다.

다음으로 유고링크에 진입하기 전에 통행시간을 예측하는 방법은 다음과 같다. 여기서는 유고링크를 먼저 통과한 앞 버스의 구간 통행시간을 이용한다. 그리고, 구간속도(정류장A와 정류장B 사이 통행속도) 변화를 예측하여 차량이 정류장A에 있을 때 다음 정류장B에 도착할 시간을 예측한다.

4. 전문가 시스템

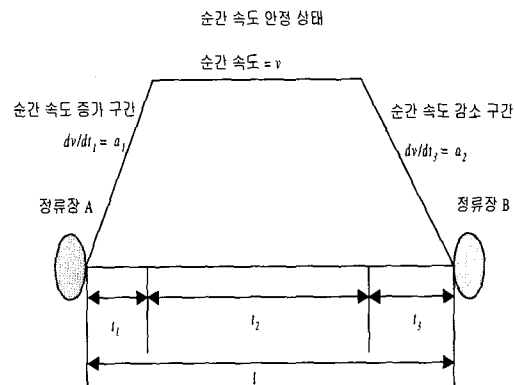
버스가 다음 정류장에 도착할 시간을 예측한 후, 버스가 정류장에 도착하면 실제 도착한 시간과 예측값을 비교할 수 있다. 실제값과 예측값을 비교한 결과 예측의 정확도 떨어진다면 AVL데이터 샘플링 횟수를 늘려 정확도를 높일 수 있다. 여기서, 샘플링 횟수를 늘리려면 샘플링 파라미터인 k_i 와 m_i 을 작게 한다. 여기서, 전문가 시스템은 목표로 설정한 통신 비용 한도 내에서 정확성이 높은 샘플링 파라미터 k_i 와 m_i 을 결정하는 역할을 한다. 그리고, 결정한 샘플링 파라미터 k_i 와 m_i 을 이용하여 샘플링을 하고, 통행시간을 예측한다. 통행시간 예측 후에는 다시 예측값과 실제값을 비교하여 샘플링 파라미터 k_i 와 m_i 을 조정한다.

5. 사례 연구

앞에서 설명한 통행시간 예측 알고리즘을 적용하기 위해 서울 시내버스 15번을 대상으로 설정하였다. 하지만, 실제 AVL 시스템을 버스에 설치하여 데이터를 수집하는 것은 쉽지 않으므로 그 대신에 타코 데이터를 이용하였다. 타코는 이미 버스에 설치된 장비로, 버스가 정류장에 도착한 시간을 알려 준다. 하지만, 타코는 버스가 정류장을 떠나 운행 중일 때 버스의 위치좌표와 시각 데이터를 알려 주지 못한다. 따라서, 타코 데이터를 변환하여 사용하였다. 데이터를 변환하는 과정에서 다음과 같은 가정을 하였다. 첫째는 정류장과 정류장사이에 신호등이 존재하지 않는 링크를 선택하였다. 둘째는 정류장 사이의 순간 속도변화는 Greenshield 모델을 이용하였다.⁵⁾

모델을 설명하면, <그림 9>에서 알 수 있듯이 정류장과 정류장사이의 순간속도 변화는 3 구간으로 나눌 수 있다. 하나는 순간속도가 증가하는 구간이고, 그 다음은 순간 속도가 안정상태에 있는 구간이다. 그리고, 순간속도가 감소하는 구간이다. 순간 속도가 증가하는 구간은 순간속도, 가속률, 통행시간으로 나타낼 수 있다.

이제, 수집한 15번 버스 데이터에 대해 설명한다. 데이터는 서울 시내버스 15번 운영자료를 이용하였다. 이용한 데이터는 15번 버스노선에서 보문동에서 대광고 구간(250m)과 전농2동에서 휘경동(300m) 구간의 데이터이다. 측정 날짜는 1998년 8월 31일 월요일부터 9월 5일 토요일까지이고, 측정 시간은 오전 7시부터 오전 9시까지이다. <표 3>은 해당구간의 통행시간 평균과 표준편차로 이 연구에서 과거 데이



<그림 9> 링크 형태와 순간속도 변화

〈표 3〉 통행시간 평균과 표준편차

항목	보문동-대광고 구간	전농2동-휘경동 구간
평균(분)	1.6	1.3
표준편차	0.70	0.24

터로 활용하였다.

다음으로, 유고 상황이 발생한 경우를 살펴보자. 유고상황은 구간 통행시간이 과거 통행시간 평균에서 2σ를 벗어난 것으로 가정하였다. 그리고, 계속해서 다음 정류장에 도착하지 않을 경우, 1σ간격으로 계속 샘플링을 한다. 즉, 샘플링 파라미터 $k=2$, $m=1$ 로 설정하였다. 데이터를 분석한 결과, 일주일동안 세 번의 유고상황이 발생하였다. 두 번은 보문동에서 대광고 구간에서 발생하였고, 한 번은 전농2동에서 휘경동 구간에서 발생하였다.

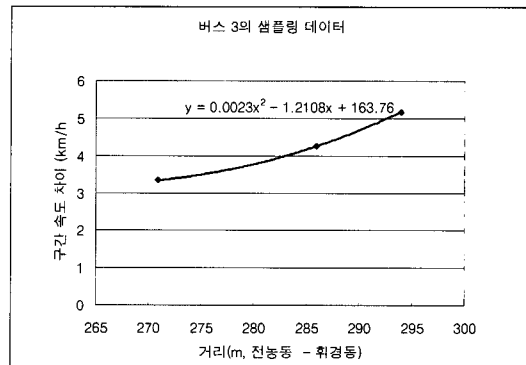
이제, 수집한 데이터를 이용해 1단계 예측을 적용한 예를 살펴보자. 〈그림 10〉은 전농 2동에서 휘경동 구간에서 유고 상황이 발생한 후 3번째로 통과한 버스인 버스 3의 샘플링 데이터이다. 링크통행시간이 $TP_i + k_i\sigma_i(1.3+2\times 0.24\text{분})$, $TP_i + k_i\sigma_i + m_i\sigma_i(1.3+2\times 0.24+0.24\text{분})$, $TP_i + k_i\sigma_i + 2m_i\sigma_i(1.3+2\times 0.24+2\times 0.24\text{분})$ 인 경우에 샘플링한 데이터이다.

〈그림 11〉은 〈그림 10〉의 샘플링한 데이터를 이용하여 2차항 회귀식을 추정한 것이다.

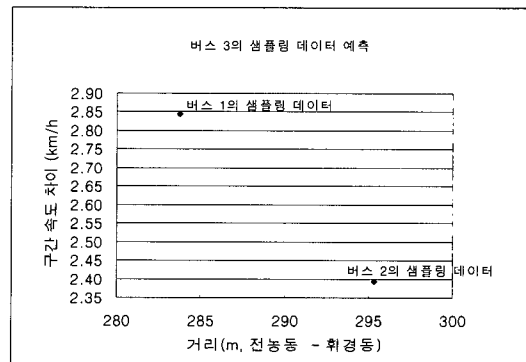
다음으로, 2 단계 예측을 적용한 경우를 살펴보자. 〈그림 12〉는 버스 3의 샘플링 데이터를 예측하기 위한 샘플링 데이터이다. 유고상황이 발생한 후에 전농 2동과 휘경동 구간을 첫 번째로 통과한 버스 1과 두 번째로 통과한 버스 2의 샘플링 데이터이다. 버스 1과 버스 2의 샘플링 데이터는 모두 링크 통행시간이

$TP_i + k_i\sigma_i(1.3+2\times 0.24\text{분})$ 일 때 샘플링한 경우이다.

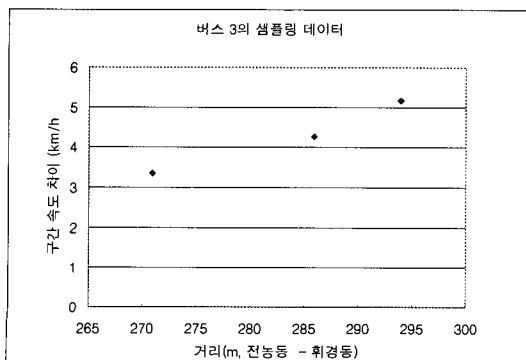
〈그림 13〉은 〈그림 12〉의 샘플링 데이터를 이용하여 추정한 회귀식을 나타낸 것이다. 버스 1과 버스 2의 샘플링 데이터를 이용하여 버스 3이 링크 통행시간이 $TP_i + k_i\sigma_i(1.3+2\times 0.24\text{분})$ 일 때, 버스 3의 위치와 구간 속도를 추정한다.



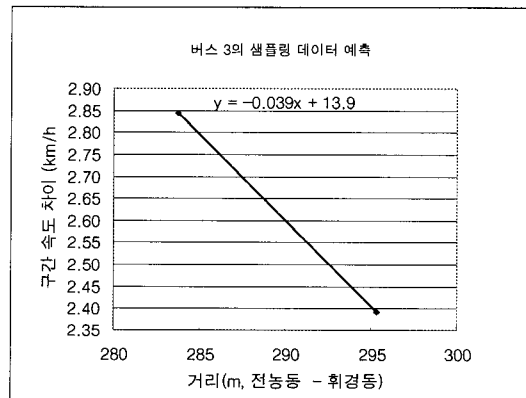
〈그림 11〉 1단계 예측 결과



〈그림 12〉 버스 1과 버스 2의 샘플링 데이터



〈그림 10〉 버스 3의 샘플링 데이터



〈그림 13〉 샘플링 데이터 예측

여기서, 예측 알고리즘의 분석을 위해서 개발한 예측 알고리즘과 시카고 어드밴스 프로젝트의 자료융합 알고리즘을 비교한다.^{11),13)} 이 연구에서 개발한 예측 알고리즘은 '2단계 예측 알고리즘'이라 명한다.

먼저, 시카고 어드밴스 프로젝트의 자료융합 알고리즘에 대해서 살펴보자. 어드밴스 프로젝트에서는 분산 가중치로 이용하여 실시간 데이터와 과거 데이터를 융합한다. 어드밴스 프로젝트에서 사용하는 통행시간 산출식은 다음과 같다.

$$EFTT = \frac{ESTT/\sigma_s^2 + EOTT/\sigma_o^2}{1/\sigma_s^2 + 1/\sigma_o^2}$$

여기서, *ESTT*(Expected Statistics Travel Time)는 과거 통행시간 평균이고, σ_s 는 과거 통행시간 분산이다. 한편 *EOTT*(Expected On-line Travel Time)는 실시간 통행시간 평균이고, σ_o 는 실시간 통행시간 분산이다.

실시간 통행시간(*EOTT*)은 바로 전에 링크를 통과한 4대의 차량의 통행시간 평균으로 하였고, 과거 통행시간(*ESTT*)은 <표 3>의 데이터를 이용하였다.

이 연구에서 사용한 평가척도는 오차율과 등가계수이다. 오차율과 등가계수 다음 식을 통해 구할 수 있다. 등가계수가 1에 가까울수록 실제도착시간과 예측시간이 일치하므로 신뢰도와 같다.

$$\text{오차율}(\%) = \frac{\sum_t |X_t(t) - X_p(t)|}{\sum_t X_t} \times 100$$

$$\text{등가계수} = 1 - \frac{\sqrt{\sum_t [X_t(t) - X_p(t)]^2}}{\sqrt{\sum_t X_t(t)^2} + \sqrt{\sum_t X_p(t)^2}}$$

여기서, $X_p(t)$ 는 *t* 번째 예측 통행시간이고 $X_t(t)$ 는 *t* 번째 실제 통행시간이다.

이제, 알고리즘 분석 결과에 대해 살펴보자. 어드밴스 자료융합 알고리즘과 이 연구에서 개발한 2단계 예측 알고리즘은 세 번의 유고상황에 대해서 버스가 다음 정류장에 도착할 시간을 예측한다. <표 4>, <표 5>, <표 6>는 어드밴스 자료융합 알고리즘과 2 단계 예측 알고리즘의 결과를 비교한 것이다. 유고1의 등가계수의 평균을 보면, 어드밴스 자료융합 알고리즘은 88.5%

인 반면, 2단계 예측 알고리즘은 92.9%의 정확성을 나타내고 있다. 유고2의 경우도 2단계 예측 알고리즘의 등가계수인 95.9%가 어드밴스 자료융합 알고리즘의 등가계수인 72.7%보다 훨씬 높다. 유고3의 결과도 2단계 예측 알고리즘이 96.1%로 83.8%보다 높다. 세 가지 유고 상황에 대해서 모두 2단계 예측 알고리즘이 우수한 것을 알 수 있다.

<표 7>은 세 번의 유고상황에 대해서 오차율과 등가계수의 평균을 비교한 것으로 2단계 예측 알고리즘의 오차율은 8.6%이고, 어드밴스 자료융합 알고리즘은 30.2%이다. 등가계수는 2단계 예측 알고리즘이

<표 4> 어드밴스 자료융합 알고리즘과 2단계 예측 알고리즘 비교(유고1)

유고1	오차율(%)		등가계수(%)	
	어드밴스 자료융합	2 단계 예측	어드밴스 자료융합	2 단계 예측
보문동-대광고 구간 버스 1	14.8	8.2	92.0	95.7
버스 2	36.0	14.4	78.0	91.5
버스 3	22.0	10.9	87.7	91.9
버스 4	7.4	13.9	96.1	92.6
평균	20.1	11.8	88.5	92.9

<표 5> 어드밴스 자료융합 알고리즘과 2단계 예측 알고리즘 비교(유고2)

유고2	오차율(%)		등가계수(%)	
	어드밴스 자료융합	2 단계 예측	어드밴스 자료융합	2 단계 예측
보문동-대광고 구간 버스 1	32.2	0.02	80.8	99.9
버스 2	51.2	10.9	65.6	93.9
버스 3	37.8	12.9	76.7	93.1
버스 4	48.8	5.7	67.7	96.6
평균	42.5	7.4	72.7	95.9

<표 6> 어드밴스 자료융합 알고리즘과 2단계 예측 알고리즘 비교(유고3)

유고3	오차율(%)		등가계수(%)	
	어드밴스 자료융합	2 단계 예측	어드밴스 자료융합	2 단계 예측
전동2동-휘경동 구간 버스 1	28.6	4.7	83.3	96.6
버스 2	16.1	2.1	91.2	98.9
버스 3	37.5	11.1	76.9	92.8
평균	27.4	6.0	83.8	96.1

〈표 7〉 베이지안 이론과 2단계 예측 알고리즘 비교
(유고1, 유고2, 유고3)

항목	오차율(%)		등가계수(%)	
	어드밴스 자료융합	2단계 예측	어드밴스 자료융합	2단계 예측
평균	30.2	8.6	81.5	94.9

94.9%이고, 어드밴스 알고리즘의 81.5%로 2단계 예측 알고리즘의 정확도가 높다.

IV. 결론

이 연구에서 버스를 운행하는데 드는 비용과 기회 손실비용을 비교하여 최적 운행시격과 통행시간 예측 알고리즘에 대해 살펴보았다. 먼저, 운행시격은 버스 회사 수익을 전부 요금에만 의존하는 국내 시내버스 회사에 합리적인 운행관리를 위한 기초를 제공할 수 있을 것으로 기대한다. 사례연구에서 살펴본 결과는 승객이탈이 발생하는 대기한계시간을 균일분포로 가정하고 통행시간을 정규 분포로 한 것이다. 비용함수에 사용되는 모수들과 분포 변화가 운행시격에 미치는 영향을 살펴보는 것이 추후 진행할 과제이다. 또한 현재 이탈비용 모형은 단지 승객 대기시간만을 가지고 수립하였다. 그러나 승객이 교통수단을 선택하게 하는데 미치는 요인은 대기시간 외에 정시성, 편안함 등과 같은 많은 다른 요인들이 있을 것이다. 이러한 다양한 요인을 고려하여 이탈비용모형을 수립하는 연구가 필요하다고 본다. 그리고 링크통행시간이 시간에 따라 변한다고 할 때 운행시격을 달리할 수 있는 연구도 함께 필요하다고 생각한다.

통행시간 예측 알고리즘 부분에서는 2단계 예측 알고리즘이 어드밴스 자료융합 알고리즘보다 더 정확한 것을 알 수 있다. 일주일간 데이터에서 유고상황이 발생한 경우, 어드밴스 자료융합 알고리즘은 81.5%의 정확성을 나타내는 반면, 2단계 자료융합 알고리즘은 94.9%의 정확성을 나타낸다. 여기서 적용한 데이터는 단순한 정체상황을 반영한 것이다. 단순한 정체 상황이기에 때문에 2단계 자료융합 알고리즘의 모든 단계를 적용할 수 없었다. 만일 사고와 같이 교통흐름의 변화를 크게 주는 유고 상황을 반영한다면, 2단계 자료융합 알고리즘은 다른 자료융합 알고리즘과 비교할 때 훨씬 더 정확한 예측 결과를 낼 수 있으리라고 본다.

참고문헌

1. 고승영·고종섭(1998), "버스의 최적운행시격 및 보유대수 모형 개발," 교통학회지, 제16권, 제2호, pp.169~176.
2. 서울시정개발연구원(1997), "서울시 교통센서스 및 데이터베이스 구축".
3. 서울특별시(1997, 7), "시내버스 개혁 종합대책".
4. 최기주(1998), "링크 통행시간 추정을 위한 데이터 퓨전 알고리즘의 개발," 대한교통학회지 제16권 제2호.
5. Sanghoon Bae(1995), "Dynamic Estimation of Travel Time on Arterial Roads by Using Automatic Vehicle Location(AVL) Bus as a Vehicle Probe," Civil Engineering, Virginia Polytechnique.
6. S. K. Chang and Paul M. Schonfeld(1991), "Integration of Fixed and Flexible Route bus System," TRR 1308, pp.51~57.
7. J. O. Jansson(1980), "A Simple Bus Line Model for Optimization of service Frequency and Bus Size," Journal of Transport Economics and Policy, Vol. 14, pp.53~80.
8. Kurt Ker-Tshung Lee and Paul M. Schonfeld (1994), "Real-Time Dispatching Control for Coordinated Operation in Transit Terminals," TRR 1433, pp.3~9.
9. Peter Nelson , Prasad Palacharla(1993), "A Neural Network for Data Fusion in ADVANCE," Presented at the 1993 Transtech Pacific Rim Conference Seattle, Washington.
10. Nagui M. Roupail, Andrzej Tarko, Peter Nelson , Prasad Palacharla(1993), "Travel Time Data Fusion in ADVANCE A Preliminary Design Concept," ADVANCE WORKING PAPER SERIES No. 21.
11. Nagui M. Roupail, Andrzej Tarko(1993), "Travel Time Data Fusion in ADVANCE," ADVANCE WORKING PAPER SERIES No. 28.
12. Nagui M. Roupail, Andrzej Tarko(1997), "Intelligent Traffic Data Processing for ITS

- Application," Journal of Transportation Engineering July/August.
13. Andrzej Tarko(1995), "Data Fusion Algorithm for ADVANCE Release 2.0," ADVANCE WORKING PAPER SERIES No. 48.
14. P. Tisato(1998), "Service Unreliability and Bus Subsidy," Transport Research-A, Vol. 32, No. 6, pp.423~436.