TCS 자료를 이용한 고속도로 통행시간 예측

The Travel Time Forecasting on Freeway using TCS Data

-Interim Report-

남궁성	윤일수	조범철
(한국도로공사 도로연구소 교통연구실 책임연구원)	(한국도로공사 도로연구소 교통연구실 책임연구원)	(한국도로공사 도로연구소 교통연구실 위촉연구원)

목 차

I. 서론
II. 교통물류정보시스템(TLIS) 소개
1. 교통물류정보시스템의 개요
2. 국내 유사 시스템 현황
III. 고속도로 통행시간 특성 분석
1. 자료의 수집
2. TCS 통행시간 자료의 특성
IV. 통행시간 자료구축을 위한 전처리
1. 전처리의 필요성
2. 이상자료 제거 및 대표값 추출
3. 잡음 제거
V. 예측모형 선정 및 수행
1. 예측 모형 선정을 위한 고려사항
2. 예측모형 선정
3. Kalman Filtering
4. Artificial Neural Network
5. Recurrent Neural Network
6. Pattern Matching
7. 모형 수행 결과
VII. 결론 및 향후 연구과제

I. 서 론

한국도로공사는 TCS(Toll Collection System), FTMS(Freeway Traffic Management System) 등 다양한 정보 제공원을 보유/운영 중에 있다.

지금까지는 교통정보 수집시스템을 포함하여 과학적인 교통관리를 위한 기반시설 확충에 중점을 두어 왔으나, 2000년대에는 수집된 정보의 활용가치를 극대화하고자 하는 고도화/실용화 단계에 들어설 것으로 전망된다. 따라서 다양한 고속도로 교통정보 수집원이 개발되고 그 수집범위가 전국으로 확대되어가고 있는 현 상황을 비추어 볼 때, 고도화 단계에 이르러 통행시간 예측기술과 같은 정보가공 및 분석기술이 매우 필요할 것이다. 더욱이 국민 소득수준의 향상으로 고속도로 이용고객의 서비스 요구수준이 높아짐에 따라 통행시간 예측정보와 같은 고급정보의 수요가 증가할 것이다.

본 연구는 한국도로공사가 수립한 '교통물류정보시스템 개발계획('98. 6)' 중 제 2단계 개발계획에 의거하여 동시스템의 주요 기능인 교통정보 분석 및 가공을 위한 통행시간 산출 및 예측 기술 개발을 목표로 하고 있다.

II. 교통물류정보시스템(TLIS) 소개

1. 교통물류정보시스템의 개요

교통물류정보시스템(TLIS, Traffic and Logistics Information System)은 전국 고속도로 1996km에 있는 톨게이트 139개소로부터 수집되는 통과차량들의 차종(현재 경승용차를 포함하여 6개 차종으로 분류하고 있음), 교통량, 그리고 고속도로 진출입시간 자료를 이용하여 톨게이트간 차종별 통행량 및 통행소요시간 등 다양한 정보를 이용자 및 고속도로 운영자에게 제공하기 위해 개발된 시스템이다. TLIS는 '98년 12월에 1단계 개발이 완료되었으며, 주요 제공 정보는 다음과 같다.

- 실시간 고속도로 소통상황
- 고속도로 차종별 Real-time Dynamic OD
- 출발-도착 톨게이트 간 고속도로 통행시간
- 톨게이트 폐쇄정보
- 시간대별, 일별, 요일별, 월별 고속도로 교통통계자료
- TG, IC, 휴게소 등 교통시설물 정보

2. 국내 유사 시스템 현황

현재 국내에서 교통정보를 제공하는 기관은 한국도로공사(FTMS), 건설교통부(건설교통종합정보 센

타), 경찰청과 도로교통안전관리공단(교통정보서비스 센타, KORITIC), 서울시와 서울경찰청(올림픽도로 교통정보 시스템, OHTIS), (주)LG인터넷 교통정보사업부(ROTIS), 교통방송 및 이동전화 5개사에서 무료 또는 유료로 제공되고 있다. 상기 시스템들은 루프 및 영상 감지기, CCTV, 비콘 및 Probe Car 등을 이용하여 교통정보를 수집/가공/처리하여 서비스 지역의 실시간 통행시간 정보를 제공하고 있으나, 현재 통행시간 예측 정보를 제공하고 있는 시스템은 거의 없는 상태이다.

III. 고속도로 통행시간 특성 분석

1. 자료의 수집

기본적으로 TLIS에서는 15분단위로 집락된 TCS 교통자료, 5분단위 VDS 자료 및 기타 관련자료가 수집/저장된다. 본 연구에서는 이러한 집락자료와는 별도로 정밀한 분석으로 위하여 다음과 같이 TCS 원시자료(해당 영업소를 통과한 모든 차량의 전수자료)를 TLIS를 통해 수집하였다.

본 연구에서는 경부선 내의 5개 영업소 또는 Tollgate (서울, 수원, 천안, 청주, 대전)를 대상으로 2월 18일부터 3월 31일까지 42일간의 자료를 수집하였다.

이 자료는 해당 톨게이트(폐쇄식)를 통과하는 모든 차량에 대하여 수집되며, 7개 필드로 구성되어 있다. (즉, 출구영업소 번호, 출구영업소 통과 월/일, 출구영업소 통과 시/분, 입구영업소 번호, 입구영업소 통과 월/일, 입구영업소 통과 시/분, 차종) 여기서 출구영업소 통과시각과 입구영업소 통과시각간의 차이가 해당구간(영업소간) 통행소요 시간이 된다.

2. TCS 통행시간 자료의 특성

수집된 42일간의 (1999. 2. 18. ~ 3. 31.) 자료에 대한 기본 통계량은 다음 표와 같다.

<표 1> 기본 통계량 (단위:분)

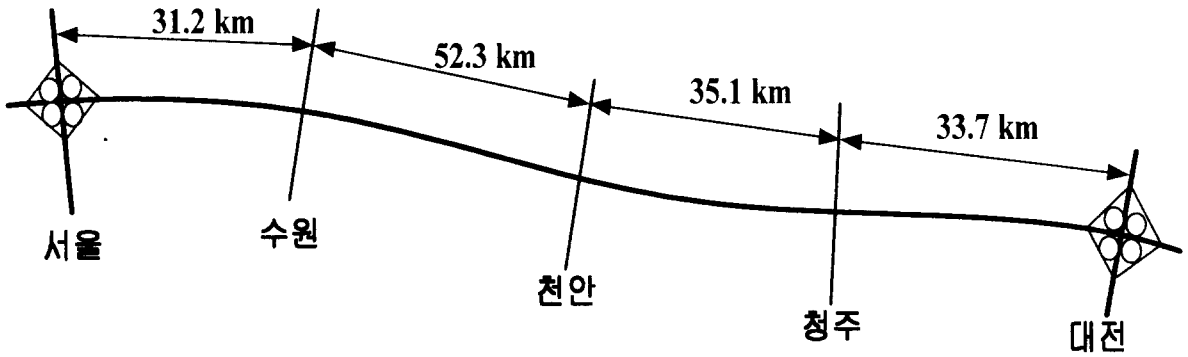
구 분	서울→수원	서울→천안	서울→청주	서울→대전
거리(km)	31.2	83.5	118.6	152.3
평 균	10.0	50.3	76.1	99.9
분 산	78.2	506.2	742.7	1048.1

TCS 자료를 통한 구간통행 소요시간은 실제 고속도로를 통과한 전체 차량에 대한 정확한 출발도착시간과 통행시간 정보로써 다음과 같은 특성을 같는다.

① 다양한 잡음(Noise)이 포함되어 있다. 여기서 노이즈라 함은 차량고장, 사고, 휴게소 체류 등으로 인해 통행소요시간이 상대적으로 큰 데이터와 이와 반대로 정상적인 속도 이상으로 운행하여 비정상적으로 통행소요시간이 짧은 데이터를 말한다.

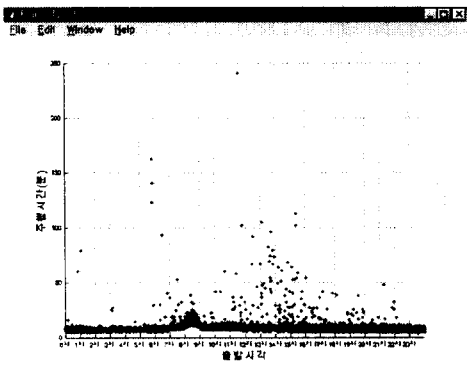
② 위의 <표 1>에서 나타난 바와 같이 대상구간(영업소간)의 거리가 멀수록 데이터의 분산이 커진다.

③ 차량이 출구영업소를 통과하는 시점에서 비로소 TCS 시스템에 자료가 저장된다. 즉, TCS자료가 실시간으로 수집된다 하더라도

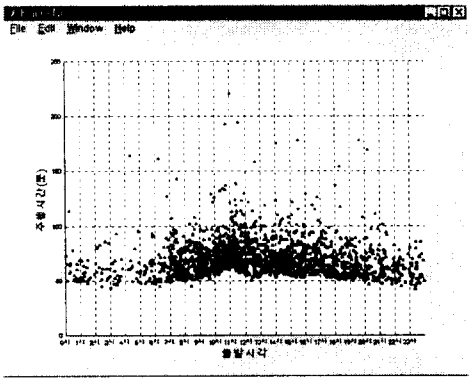


< 그림 1 > 여행시간 예측을 위한 경부고속도로의 대상 구간

현재 시각 기준으로 하여, 동일한 구간을 이용하고자 하는 차량(또는 운전자)입장에서 보면 그 데이터는 이미 당해구간 통행소요시간만큼 앞서서 출발한 차량이 경험한 시간이라는 것이다. 이와 같이 TCS자료를 기반으로 한 통행소요시간 데이터는 '시간처짐'이라는 특성을 갖는다.



(1) 서울→수원간 통행시간 (10.5Km)



(2) 서울→대전간 통행시간 (131.8Km)

<그림 2> 거리에 따른 통행시간 자료의 분포

위의 <그림 2>는 거리별 통행시간의 분포를 출발시각을 기준으로 나타낸 것으로 <표 1>에서도 보는 바와 같이 거리가 멀어질수록 분산이 커지고 있음을 알 수 있다.

IV. 통행시간 자료구축을 위한 전처리

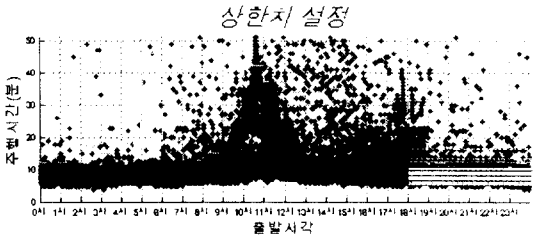
1. 전처리의 필요성

통행시간 산출 및 예측을 위해서는 비정상적인 이상데이터(잡음; noise)를 제거한 다음, 하나의 시간간격(Sampling Rate : 1분 기준)내에 존재하는 여러 데이터중에 하나의 대표값을 산출함으로써 단일시계열로 만들어 주기 위한 전처리 과정이 필요하다.

본 연구에서 사용한 전처리는 이상자료 제거, 대표값 추출, 잡음 제거 순으로 수행되었다.

2. 이상자료 제거 및 대표값 추출

이상자료 제거는 해당 구간의 자료에 대하여 42일치 자료에 대한 요일별 시간대별 누적상대도수비를 계산하여 95%와 99%에 해당하는 값을 산출하였고, 99% 값이 충분히 이상자료를 제거하지 못함에 따라, 최종적으로 95% 값을 상한치로 설정하였다. 그리고 이 상한치보다 큰 값을 이상자료로 정의하고 제거하였다. <그림 3>은 각각 99%와 95%를 상한치로 설정한 결과를 나타내고 있다.



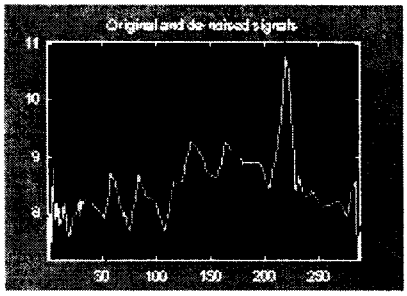
(서울→수원, 금요일 자료)

<그림 3> 상한치 제거의 결과 비교 (99%,95%)

대표값 추출을 위하여 상한치가 제거된 자료를 5분 단위로 집락하고, 집락된 단위 시간간격내의 대표값으로 평균값을 취하였다.

3. 잡음 제거

상기 과정을 통해 산출된 시계열 자료에 대하여 잡음을 제거하기 위하여 먼저 Wavelet 분석을 수행하여 시계열 자료의 특성을 파악하였으며, 42일치 시계열 자료에 대하여 Wavelet De-noising을 통해 잡음을 제거하였다. 수집된 통행시간 시계열 자료에 대하여서는 분해되는데, 최초의 자료는 저역 통과필터를 통과한 근사 신호(A1)와 고역통과필터를 통과한 세부신호(D1)로 분할된다, 그리고 다시 차단주파수가 변경된 저역, 고역통과필터를 거쳐서 A2와 D2로 분할되며 이러한 과정을 반복하여 결국 4개 대역의 신호로 분해한 다음, 고주파 성분중 백색잡음(white noise)이라고 판단되는 성분을 제외한 나머지 신호를 합하여 원신호로 복원함으로써 주어진 단일시계열의 잡음을 제거하였다.



(서울→수원, 1999. 2. 18.)
사용 Wavelet : DB3, 분석 level : 3
De-noising 방법 : Automatic Soft Thresholding
<그림 4> Wavelet De-noising 예

<그림 4>은 웨이블렛을 통해 잡음이 제거된 신호와 원시계열 자료를 비교하고 있다.

V. 예측모형 선정 및 수행

1. 예측 모형 선정을 위한 고려사항

고속도로 TCS 교통정보는 앞에서 설명한 바와 같이 과거에 출발지에서 출발하여 목적지에 도착한 차량의 정보만을 알 수 있기 때문에 자료 수집에 있어 시간처짐이 발생한다. 이러한 시간처짐을 극복하는 것은 통행시간 산정 및 예측에 있어 매우 중요한 요소이다. 따라서 본 연구에서의 예측은 기본적으로 다단계(Multi- Period) 예측을 필요로 한다.

본 연구에서 단거리와 장거리 구간에 따라서 시간 처짐 현상과, 시간과 공간에 따른 통행시간 등을 비롯한 TCS 시계열 자료의 특성이 다르게 나타나기 때문에 통행시간 예측 대상구간을 서울톨게이트→수원톨게이트간과 같은 단거리 구간과 서울톨게이트→대전톨게이트와 같은 장거리 구간으로 구분하여 예측방법을 적용했다. 또한 예측시간간격으로는 실용성을 고려하여 단거리 구간에서는 one-time·period 예측과 multi-period 예측을 동시에 시험하였고, 장거리 구간에서는 multi-period 예측만을 적용하였다.

이는 고속도로 이용자가 톨게이트에 진입하여 필요로 하는 정보를 제공하기 위해서는 현재 도착한 차량의 자료를 이용하여 여행시간만큼의 시간을 앞서서 예측하여 제공하여야 하기 때문이다. 서울-수원 구간의 평균 여행시간은 약 10여분이고 서울-대전 구간의 평균 여행시간은 약 90여분이므로 서울-수원간은 5분 15분 30분 1시간 2시간 3시간을, 서울-대전간은 1시간, 2시간, 3시간을 예측시간간격으로 설정하였다.

2. 예측모형 선정

본 연구에서는 교통정보 예측과 관련된 문헌조사와 TCS 교통시계열 특성, 교통물류정보제공시스템의 요구기능을 종합적으로 검토하여 다음과 같은 예측 방법들을 시험하였다.

① Kalman Filter
② Artificial Neural Network ; Back-Propagation
③ Recurrent Neural Network
④ Pattern Matching

3. Kalman Filter

칼만필터는 최초 동적선형시스템에서 관측이 가능한 정보를 이용하여 미지의 상태를 추정하는 예측방법론이다. 칼만필터에서는 상태식과 관측식은 시스템의 특성을 최대한 반영하여 설계하여야 한다. 본 방법론에서 구축된 상태 방정식은 통행시간이 AR 모형을 따른다는 가정 하에서 아래와 같이 구성될 수 있다.

또한 상태방정식을 구성하는 방법으로 아래와 같이 파라메타를 갱신하는 방법과 여행시간 자체를 갱신하는 방법 등 2가지를 사용하였다.

- 상태식 :

$$\begin{bmatrix} TCS(t+1) \\ \theta(t+1) \end{bmatrix} = \begin{bmatrix} \theta(t) & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} TCS(t) \\ \theta(t) \end{bmatrix} + \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}$$

- 관측식 1 : 파라메타를 갱신하는 방법

$$\frac{TCS(t)}{\theta(k-1)\ TCS(k-1)} = [\,0 \quad 1\,] \cdot \begin{bmatrix} TCS(t) \\ \theta(t) \end{bmatrix} + v(t)$$

- 관측식 2 : 여행시간을 갱신하는 방법

$$TCS(t) = [\,1 \quad 0\,] \cdot \begin{bmatrix} TCS(t) \\ \theta(t) \end{bmatrix} + v(t)$$

여기서,

$TCS(t)$: 시간 t에 도착한 차량의 여행시간

$\theta(t)$: 시간 t에서 시계열 모형의 파라메타

$w(t)$, $v(t)$: 시간 t에서 백색잡음.

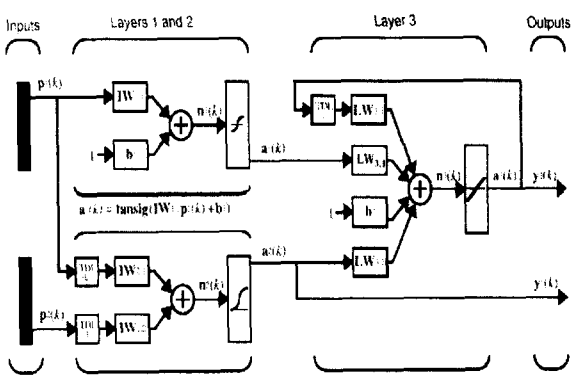
파라메타를 갱신하는 방법은 서울-대전구간에서 1시간, 2시간, 3시간이후 예측과 서울-수원구간 5분, 15 분 이후를 예측하는 경우에 사용되었고, 여행시간으로 직접 갱신하는 방법은 서울-수원 구간에서 30분, 1시간, 2시간, 3시간 이후를 예측하는 방법에서 사용되었다. 파라메타를 갱신하는 방법은 급변하는 단거리 장기예측에서 불안정한 결과를 보이기 때문이다.

4. Artificial Neural Network

인공 신경망 중 가장 많이 사용되는 학습규칙은 역전파 학습방법으로 입력치와 목표치간의 에러를 최소화시켜주는 연결강도로 학습/갱신하는 방법이다. 본 연구에서는 이러한 역전파 학습법을 사용하였는데, 먼저 여행시간에 대해서 정규화 과정을 거친 후, 입력값으로 현재 시간간격과 이전 시간간격(5분전)의 통행시간, 목표값으로는 각각 5분, 15분, 30분, 1시간, 2시간 그리고 3시간 후의 여행시간으로 각각 사용하였다. 학습자료로는 2월 18일과 2월 19일의 양일 자료가 사용되었고, 3월 15일~28일의 2주간의 자료가 모형검증을 위한 예측자료로 사용되었다.

5. Recurrent Neural Network

신경망 이론 중에서 역전파 학습방법은 단순히 현재의 입력에 대한 출력으로 그전의 입력값에 대해서 영향을 받지 않기 때문에 시계열적으로 변동이 없는 패턴을 예측할 때에는 적합하다. 반면 Recurrent Neural Network은 현재의 출력값이 현재의 입력값과 이전 시간대의 출력값으로 결정됨으로써 변동이 심한 시계열 패턴을 잘 모사한다.



<그림 5> Recurrent Neural Network 구조

학습자료와 검증자료는 위의 역전파 학습방법을 이용한 인공신경망과 동일하나 대신, 네트워크는 시계열적인 영향을 고려하기 위해 위의 < 그림 5 >과 같이 회귀적인 형태를 가진다.

6. Pattern Matching

Pattern Matching을 이용한 예측은 현재 입력된 시계열 부분패턴을 과거 시계열 자료의 패턴과 비교하여 가장 유사한 패턴을 검출하고, 검출된 패턴의 다음 패턴을 현재 입력패턴의 다음 패턴으로 가정하고 이를 예측치로서 취하는 방법이다. 이는 수집된 TCS 자료를 먼저 15분 단위로 환산하고 이를 여행시간의 변화율로 계산하였다. 2월 18일부터 3월 14일까지의 여행시간 자료는 패턴별로 분류하여 패턴 검색용으로 저장하고 이후 3월 15일부터 28일까지의 자료는 입력자료로 사용하였다. 본 연구의 Pattern Matching 알고리즘은 아래와 같다.

① 여행시간의 변화율로 전환
② 기존의 저장된 패턴들과 입력패턴을 비교 ;
- 오차의 합과 오차의 최대값을 계산
- 오차의 합과 그 최대값이 임계치 이하일 경우 검색된 패턴 ID를 Index집합에 저장
③ 임계값 보정 : Index 집합이 공집합일 경우로 모든 결과가 임계치 이상일 경우
- 보정된 임계값으로 위의 과정 2를 반복
④ Index 집합 내에서 최소값을 검색
⑤ 변화율을 여행시간으로 재전환

7. 모형 수행 결과

제안된 각 모형은 TCS 자료를 통해서 추정된 여행시간을 이용하여 서울-대전 구간은 1시간, 2시간, 3시간 이후의 여행시간을 서울-수원 구간은 5분, 15분, 30분, 1시간, 2시간, 3시간 이후의 여행시간을 예측했으며, 각 모형의 결과를 평가/비교하기 위하여 99년 3월 15일~28일의 2주간의 자료를 동일하게 검증자료로 사용하여 예측하였으며, 모형간 예측도의 평가를 위한 지표는 아래와 같다.

① Mean Average Relative Error

$$MARE = \frac{1}{N}\sum_{i}\left|\frac{(t_i - \hat{t}_i)}{t_i}\right|$$

② Root Relative Square Error

$$RRSE = \sqrt{\frac{1}{N} \times \sum_i \left(\frac{t_i - \hat{t}_i}{t_i} \right)^2}$$

여기서,

t_i : 시간대 i에서의 실제 여행시간

$\hat{t}_i$: 시간대 i에서의 예측 여행시간

N : 전체 관측/예측 시간간격 개수

< 표 2 > 서울-대전 여행시간 예측결과

		1H	2H	3H
MARE	K-F	6.38	6.92	9.93
	ANN	6.10	7.30	8.0
	RNN	6.23	7.14	7.81
	P-M	6.67	7.38	12.49
RRSE	K-F	8.71	9.46	14.11
	ANN	8.30	9.90	10.9
	RNN	8.19	9.18	10.0
	P-M	9.32	9.94	17.27

< 표 3 > 서울-수원 여행시간 예측결과

		5M	15M	30M	1H	2H	3H
MARE	K-F	0.93	2.49	4.35	9.47	13.00	18.67
	ANN	2.05	5.61	6.80	7.62	9.23	9.83
	RNN	1.80	3.28	5.47	8.14	11.06	13.37
	P-M	2.57	2.57	4.50	8.95	8.38	15.05
RRSE	K-F	2.52	4.83	8.10	18.78	27.55	38.47
	ANN	3.12	7.53	9.23	10.60	12.30	12.86
	RNN	3.76	5.62	7.99	11.87	14.15	16.02
	P-M	5.00	5.00	8.42	16.9	17.39	29.14

위의 결과에서 알 수 있듯이 단기예측에서는 칼만필터가, 장기예측에서 인공신경망 모형이 우수한 결과를 보이고 있다. 특히, 인공 신경망 모형은 예측간격이 길어져도 그 오차가 다른 모형에 비해서 덜 커지는 것을 볼 수 있어 장기추정에 있어서 안정적인 모형이라 할 수 있다. 이와 반대로 칼만필터와 패턴매칭방법은 장기예측으로 갈수록 예측오차가 급속하게 증가하는 경향을 보이고 있다.

또한 구간거리에 따른 예측결과를 보면 긴 구간(서울-대전)에 대해서는 Recurrent Neural Network 모형이 전반적으로 우수한 결과를 보였으며, 짧은 구간(서울-수원)에 대해서는 역전파 학습방법을 이용한 Neural Network이 우수한 결과를 보였다. 이는 서울-수원 구간의 통행소요시간이 주기적이고 일정한 패턴을 가지는데 반해, 서울-대전 구간은 비반복적이고 변동적인 패턴을 가지기 때문인 것으로 판단된다.

VI. 결론 및 향후 연구과제

TCS자료는 고속도로 이용차량의 유출입 전수자료로서 최근 외국에서도 전자통행료 징수시스템(ETCS)의 구축과 함께 유료도로 이용차량 정보를 교통정보로서 활용하고자 하는 시도가 이루어지고 있다. 본 연구에서는 TCS자료를 이용하여 고품질의 교통정보를 제공하고자 하는 것으로 차량검지기(VDS)가 설치되어 있지 않은 구간에 대한 교통정보 수집을 가능하게 함은 물론, 고속도로상의 출발/목적지간 소요시간 예측정보 제공에 있어 매우 유용하게 활용될 것으로 판단된다. 다만, 본 연구는 현재 계속 진행 연구중인 과제로서 본 고에서는 TCS자료를 이용한 통행시간 산출 방법 및 몇 가지 예측방법의 적용결과를 시험적으로 제시하였다.

이후에 진행될 연구내용으로서는, 첫째, 통계학적 방법 및 혼돈이론(Chaos theory) 등을 이용한 TCS 통행시간 시계열자료의 비선형 패턴분석 및 예측치 사전평가방법 개발. 둘째, 공사, 사고, 계절변동, 관련 외부환경의 변화(유가변동, 통행료 인상 등)를 고려한 직관적 예측방법(Judgmental Forecasting)의 도입. 셋째, 주어진 시계열 입력패턴의 특성에 따른 예측방법 선정 모형의 개발 등이 있다. 마지막으로 경로단위의 신뢰성 있는 통행시간 예측을 위한(즉 다단계 예측) 장기예측 방법 개발 및 다양한 예측모듈의 통합에 관한 연구가 진행될 예정이다.

참고문헌

- 강정규(1996), 칼만필터를 이용한 도시고속도로 교통량예측 및 실시간 O-D 추정, 대한교통학회 회 제 14호
- Duane Hanselman and Bruce Little field. (1998), Matering MATLAB 5.0, Prentice Hall.
- Grewal M.S and Andrews, A.P.(1993) Kalman Filtering :Theory and Practice, Prentice Hall.
- Howard Demuth and Mark beale, (1992), Neural Network User Manual, MathWorks.
- Betty Bin Xia(1997), Similarity Search in Time Series Data Sets, In M.Sc. Thesis, Simon Fraser University.