

■ 論 文 ■

신경망을 이용한 고속도로 여행시간 추정 및 예측모형 개발

The Development of Freeway Travel-Time Estimation and Prediction Models
Using Neural Networks

김 남 선

이 승 환

오 영 태

(아주대학교 교통연구센터 연구원) (아주대학교 환경·도시공학부 교수) (아주대학교 환경·도시공학부 교수)

목 차

- I. 서론
 - 1. 연구의 배경 및 목적
 - 2. 연구의 내용 및 방법
 - II. 이론적 고찰
 - 1. 신경망 이론
 - 2. 칼만 필터 기법
 - III. 자료처리방법
 - 1. 이상치 제거 기법
 - 2. Rolling Horizon 기법
 - 3. 여행시간 예측을 위한 칼만 필터링
 - IV. 모형 구축
 - 1. 모형 구축 방안
 - 2. 모형구축의 가정
 - 3. 모형 구조
 - V. 모형의 평가
 - 1. 학습 자료
 - 2. 모형평가 시나리오
 - 3. 모형 평가 기준
 - 4. 모형 평가 결과
 - VI. 결론 및 향후연구방향
- 참고문헌

요 약

본 연구에서는 고속도로 교통관리시스템에서 VDS 교통정보 와 대상지역의 TCS로부터 여행시간을 수집하고, 이들 자료를 토대로 신경망 이론을 이용한 여행시간 추정(Estimation)모형을 구축하였다. 또한, 신경망 이론에 칼만필터기법(Kalman Filter Technique)을 연계하여 단위시간 동안의 여행시간을 예측(Prediction)하여, 고속도로 이용자에게 보다 향상된 실시간 여행시간정보를 제공할 수 있는 여행시간 추정 및 예측 알고리즘을 개발하였다.

신경망 모형의 여행시간 추정 방식과 현재 적용되고 있는 여행시간 산출 방식의 비교/분석을 위해 각 각의 여행시간 산출방식에 의한 평가지표별로 시행한 평가의 결과는 신경망 모형이 제시한 대부분의 지표에서 상대적으로 우수하게 나타났다.

1. 서론

1. 연구의 배경 및 목적

기존의 고속도로 교통관리시스템의 여행시간정보는 공간평균속도(SMS:Space Mean Speed)방법으로 산출되며, 이러한 방식은 통행료징수시스템(TCS:Toll Collection System)의 각 영업소를 차량들이 통과할 경우 모든 차량들이 경험하게 될 영업소(Toll Plaza)에서의 지체시간을 차량검지시스템(VDS:Vehicle Detection System)에서 수집된 교통정보를 기반으로 해석하는 설명이 불가능하다.¹⁾ 이와 같은 상황이 고려된 모형을 개발하기 위해서 본 연구에서는 고속도로를 대상으로 설치된 유도식 루프검지기와 TCS를 기반으로 수집되는 교통정보를 이용하여 연구모형을 개발하였다.

본 연구에서는 고속도로 교통관리시스템에서 VDS 교통정보와 대상지역의 TCS로부터 여행시간을 수집하고, 이들 자료를 토대로 신경망 이론을 이용한 여행시간 추정(Estimation)모형을 구축한다. 또한, 신경망 이론에 칼만필터기법(Kalman Filter Technique)을 연계하여 단위시간 동안의 여행시간을 예측(Prediction)하여, 고속도로 이용자에게 보다 향상된 실시간 여행시간정보를 제공할 수 있는 여행시간 추정 및 예측 알고리즘을 개발하고자 한다.

2. 연구의 내용 및 방법

1) 연구 내용 및 방법

본 연구에서의 모형구축은 신경망 이론과 칼만필터링 기법을 통한 여행시간 추정 및 예측모형을 기반으로 다음과 같은 모형을 구축하고자 한다.

Single(단일)구간 추정모형 구축

추정모형은 임의의 출발지영업소와 인접한 도착지영업소를 모형의 공간적 적용대상으로 하며, 현재시점에서 여행시간을 추정하는 모형이다.

Multiple(다중)구간 추정모형 구축

Multiple구간 추정모형은 임의의 출발지영업소와 임의의 도착지영업소를 모형의 공간적 적용대상으로 하며, 현재시점에서 여행시간을 추정하는 모형이다.

Single(단일)구간 예측모형 구축

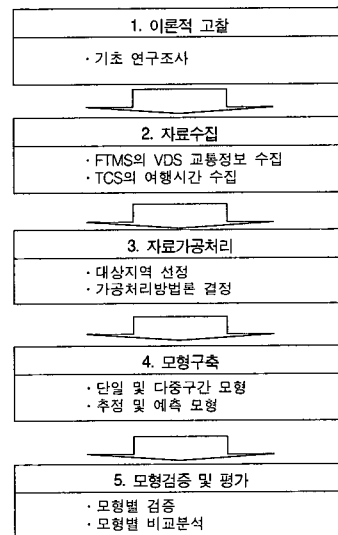
Single구간 예측모형은 임의의 출발지영업소와 인접한 도착지영업소를 모형의 공간적 적용대상으로 하며, 미래시점을 시간적 적용대상으로 하는 예측모형이다.

Multiple(다중)구간 예측모형 구축

Multiple구간 추정모형은 임의의 출발지영업소와 임의의 도착지영업소를 모형의 공간적 적용대상으로 하며, 미래시점에 대한 여행시간을 예측하는 모형이다.

실험 및 검증은 고속도로교통관리시스템과 대상지역의 영업소에서 제공하는 VDS 교통정보 및 여행시간을 학습 자료와 검증 자료로 구분하여 수행한다.

2) 연구수행절차



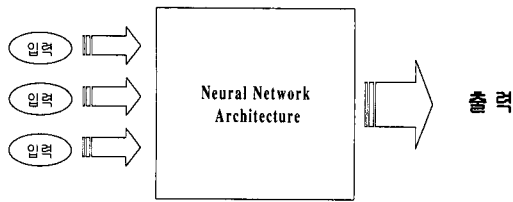
<그림 1> 연구수행절차

II. 이론적 고찰

1. 신경망 이론

1) 신경망을 이용한 시스템 모델링

어떤 실험을 통해 시스템을 규명하기 위해서는 첫째, 센서를 통해 정보를 수집하고 둘째는 모델의 구조를 결정하고 셋째는 모델구조의 모수를 추정하며, 마지막으로 모델구조의 효율성을 평가하는 과정이 필요하다. 신경망을 이용하여 시스템을 규명하기 위해서는 신경망의 입력과 출력의 수, 중간층의 수, 중간노드를 지



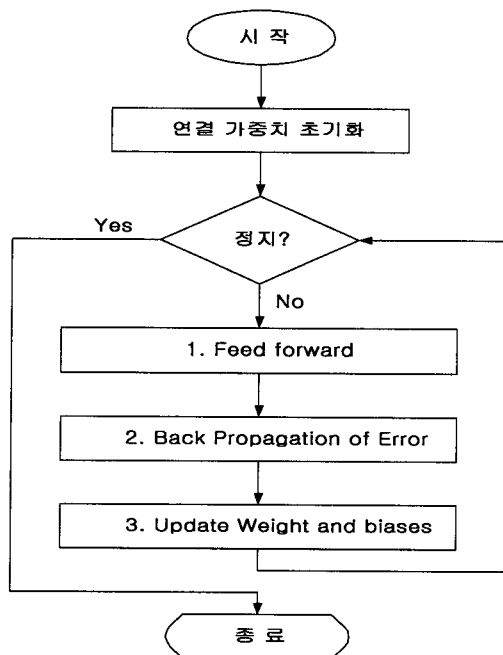
〈그림 2〉 일반적인 신경망 모형

나가는 함수, 그리고 함수의 변수들을 결정해야 한다.¹⁰⁾

본 연구에서는 역전파(Back Propagation)신경망을 이용하여 이상치가 포함되어 있는 VDS데이터의 처리 및 여행시간 추정방법론을 제시한다. 이때, 신경망의 마지막 단계에서 출력되는 결과값은 목적인 고속도로 영업소까지의 여행시간 기대값이 된다.

2) 역전파신경망의 학습알고리즘

역전파신경망의 학습알고리즘의 기본흐름은 다음과 같다.



〈그림 3〉 역전파신경망 알고리즘 흐름도

2. 칼만 필터 기법

1) 개요

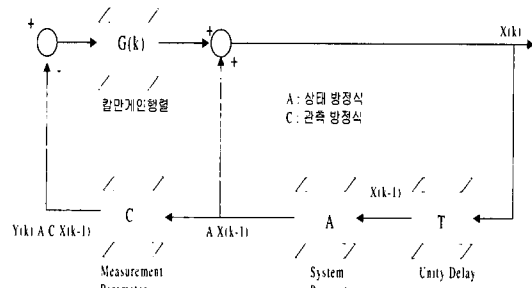
칼만필터에 의한 예측방법은 시간에 따라 입력되는 관측값과 이전시간을 기준으로 예측된 예측값을 비교

하여 평균제곱오차의 최소화 조건을 만족하게 필터를 동적으로 적응시키는 방법이다. 필터링은 신호의 초기 추정값, 오차 공분산(Error Covariance)과 새로운 측정값을 이용하여 이전 상태의 추정값을 새로운 추정값으로 교체한다. 이러한 순환과정은 시스템의 관측 가능한 상태에서 초기 추정값을 더 이상 향상이 불가능한 안정상태에 도달할 때까지 새로운 측정 데이터를 이용하여 개선된다.

2) 칼만필터 모형

(1) 모형의 기본구조

일반적으로 칼만필터의 개인행렬형태는 보통 두 가지 유형 즉, 벡터와 스칼라로 산출하여 적용하고 있으며, 이에 대한 적용 개념도는 다음과 같다.⁵⁾



〈그림 4〉 칼만 필터링의 Block Diagram

〈그림 4〉의 개념도에서 보면, 칼만필터 "G(k)"는 평균오차제곱(MSE: Mean Square Error)와 상태방정식에 의한 화이트노이즈 W(t) 및 관측방정식의 오차항 V(t)의 조합으로 이루어진 칼만필터의 개인행렬(Gain Matrix)이다.

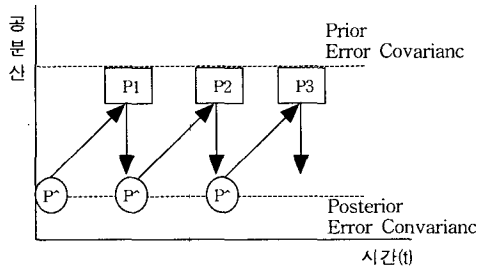
따라서 최초 반복이 수행되기 위해서는 실측 초기값 「Y(0)」 값과 각 오차항의 값이 필요한데, 이들의 값은 초기 값에 한하여 임의적으로 입력이 가능하며, 이에 대한 보정은 다음번 반복을 통하여 즉각적으로 보정된다.

(2) 예측모델의 메카니즘

구축된 칼만필터모형에 대하여 초기치에 의하여 수행되는 반복1은 1차적으로 "Time Update"의 과정을 거쳐 2차적으로 "Measurement Update"를 완료함으로써 종료된다. 그러나 중요한 사항은 반복1이 종료하는 시점에서의 결과가 예측치가 아니라는 점이다.

반복1에서 결정되는 값은 단지 추정치일 뿐이다. 따라서 예측치를 산출하기 위해서는 반복1이 종료된 후, 이들 결과를 다시 "Time Update"시켜야만 예측치를 얻을 수 있다. 이러한 예측치는 Prediction Lead-Interval이 1단계로 만일 2단계 이후의 상황을 예측하기 위해서는 시스템 방정식을 구성하는 상태전이행렬(State Transition Matrix)를 누승하여 적용해야 한다.

이상의 칼만필터모형에 있어서 "Time Update"와 "Measurement Update"의 상호 연계성에 대한 개념도는 다음 <그림 5>와 같다.



<그림 5> "Time Update"와 "Measurement Update"의 상호 연계

III. 자료처리방법

1. 이상치 제거 기법

수집된 데이터중에는 다른 데이터들과 상당한 차이를 보이는 값들이 존재한다. 일반적으로 분석대상이 되는 자료중에서 다른 데이터들과 비교하여 특히 편차가 큰 데이터들을 흔히 이상치라 한다.^{6,11)}

이상치들은 현실적으로 자료를 수집할 때 발생하며, 환경 또는 기기에 의한 여러 가지 오류 등이 원인이 된다. 이러한 이상치들의 발생원인은 크게 다음의 두 가지로 구분할 수 있다.⁹⁾

- 첫째, 자료에 오류값이 포함되어 있는 경우이다.
- 둘째, 모델자체의 원인으로서 선형인 모델을 실제 상황으로 묘사하기 위해 비선형적으로 근사하는 경우에 발생원인에 관계없이 분석자료에 포함되어 있는 이상치들은 자료분석의 정도를 저하시키며 분석결과를 왜곡시키게 된다.

본 연구에서는 자료에 포함되어 있는 이상치를 제거하기 위하여 관리도법(Control Chart Method)에 기반을 두고 1σ(표준편차)를 지표로 삼아 Upper

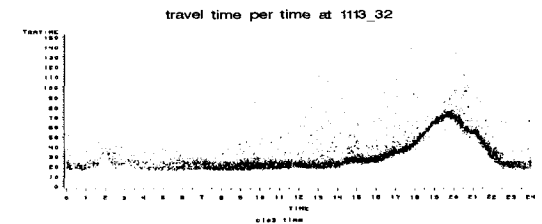
Bound를 설정하여 2단계에 걸쳐 이상치를 제거하였다. 분포상한(Upper Bound)에서만 이상치를 제거한 이유는 TCS 여행시간자료의 Data특성상 분포하한(Lower Bound)에서는 분산이 극히 작고 반면에 상한에서는 상대적으로 광범위한 분산을 갖고 있는 것으로 나타나 자료의 대표값을 결정함에 있어 대부분의 이상영향은 분포상한에서 작용하기 때문이다. 또한 2단계에 걸쳐서 이상치를 제거한 이유는 1회제거시 상한에서의 분포의 분산이 커서 평균(Mean)과 중위수(Median)의 차이가 크게 발생하므로 2회 실시하였다.

수집자료는 5분단위로 하루기준 288개 군으로 관리하며 각 군에 해당하는 원시자료(Raw Data)를 이상치의 제거대상으로 설정하였다.⁴⁾

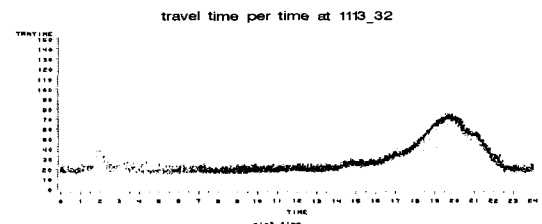
<표 1> 이상치 제거기법

적용목적	실험방법	적용결과	선택방법
■ 대표성 확보	1σ (양측)	정상치와 이상치 모두제거	
	1σ (단측) × 2회	이상치 제거	●
■ 이상치의 영향 최소화	2σ (양측)	이상치 제거 불가능	
	3σ (양측)	이상치 제거 불가능	

이상치를 제거하지 않았을 때의 TCS여행시간자료와 이상치를 제거한 다음의 TCS여행시간자료를 Plotting한 그래프를 비교해 보면 다음과 같이 나타난다.



<그림 6> 이상치 제거전 자료형태

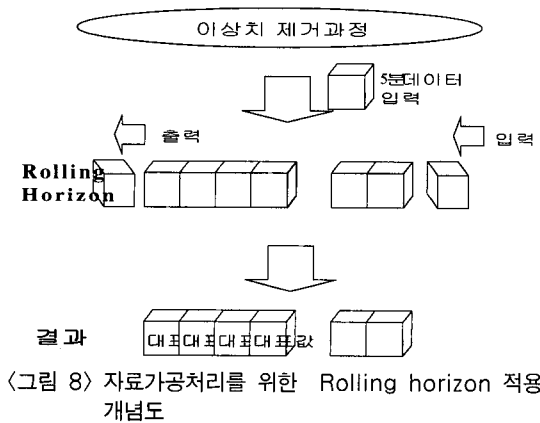


<그림 7> 이상치 제거후 자료형태

2. Rolling Horizon 기법

본 연구에서 이용되는 Rolling Horizon 기법은 여행시간 추정 및 예측시 지점의 과거자료(Historical data)를 이용하여 변화의 추이를 반영하기 위하여 사용하였다. 이는 자료의 관리 측면에서는 Block Moving과 같은 개념으로 적용될 수 있으며, 혹은 Moving Average 개념으로도 응용될 수 있다.³⁾

본 연구에서의 기초자료는 VDS의 속도 및 TCS의 여행시간 등의 개별적 속성을 갖는 자료를 의미하며, 5분단위의 데이터를 제공받아서 제시한 바와 같이 20분 또는 30분 동안의 각 5분 단위의 기초자료를 이용하여 통계적인 대표값을 산정하게 된다. 이는 추정 단위가 5분으로서 추정 및 예측을 위한 시계열적 변동요소를 반영하기 위함이다. 따라서 Rolling은 20분 혹은 30분 동안의 블록에 대한 자료를 기준으로 수행한다. 또한 5분단위 자료가 극소의 교통량으로 인하여 통계적인 교통량의 확보가 어려운 상태일 때는 이상치를 바탕으로 여행시간을 추정하는 오류를 범할 수도 있기 때문에 이를 완화하기 위하여 과거자료를 이용하여 그 영향을 감소시키는 역할도 수행한다. 이와 관련한 개념도가 다음 <그림 8>에 제시되어 있다.



이러한 방법은 기초자료의 산포가 심하여 이들의 경향을 평활화하여 실시간 시스템에 적용시에도 이용할 수 있으므로 이들을 통칭하여 Rolling Horizon 기법이라 할 수 있다.

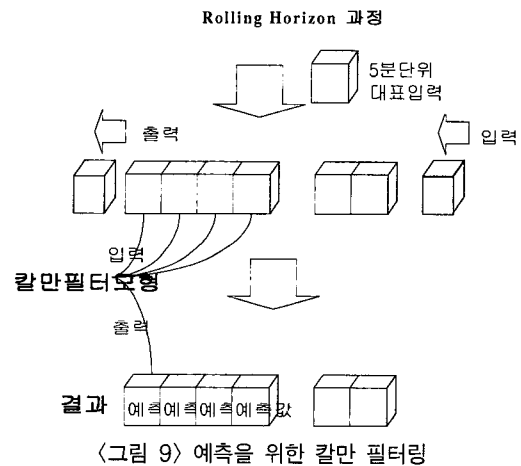
3. 여행시간 예측을 위한 칼만 필터링

본 연구에서의 기초자료를 이용하여 제2절에서 제

시하고 있는 Rolling Horizon 기법에 의해 제공되는 VDS의 속도 및 TCS의 여행시간 정보는 5분 단위로 제공된다. 이들 정보를 칼만필터모형을 통해 5분 후의 교통정보를 예측하여 모형 구축에 이용하고자 한다.

기존의 자료 가공 처리에 의해 제공되는 교통정보를 통해 구축된 모형은 현재의 교통정보를 추정하는데 반해, 이상에서 제시하는 방법에 의해 예측된 정보를 모형구축에 함께 사용한다면, 구축된 모형에 의해 5분 후의 교통정보를 예측할 수 있다.

다음 <그림 9>는 여행시간 예측을 위한 칼만필터모형의 적용 방법을 제시하고 있다.



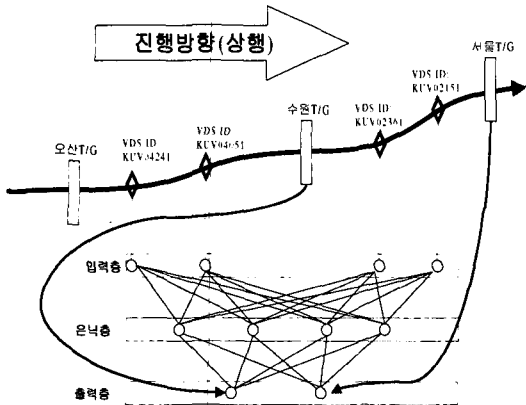
한 예로서 교통상태가 계속적으로 혼잡추세로 교통량이 계속해서 증가하는 추세라 할 때 단지 5분에 대한 정보는 정보제공목적으로서의 정보로서는 의미가 있으나 교통운영 및 제어관리 측면에서는 큰 의미가 없다. 오히려 현재 시점으로부터 15분 이상 30분 이후의 교통상태에 대한 예측 정보를 토대로 도로의 용량상태 및 과포화 유발가능성에 대한 기준을 토대로 본 연구에서 의도한 교통분석이 가능하므로 이에 대한 정보가 더 이용가치가 높은 정보가 된다.

IV. 모형 구축

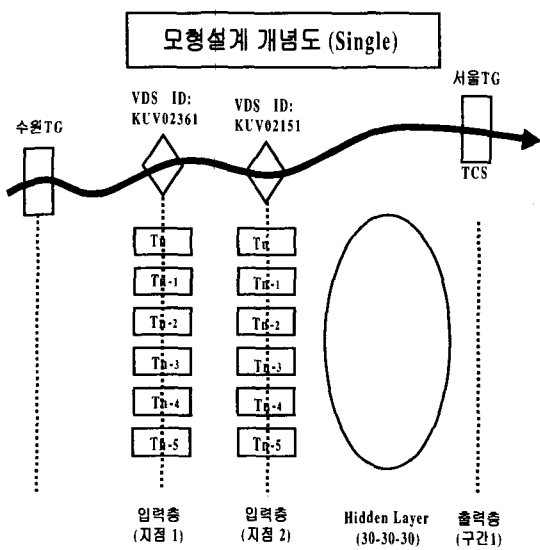
1. 모형 구축 방안

본 연구에서는 Spatial Mapping기반의 신경망모형을 구축하고자 한다. 즉, 구간 교통정보를 현실성 있게 반영하기 위하여 필수적으로 고려되는 구간의

각 지점간 관계성을 정립하여 지점정보를 구간화하고, 각 지점에서의 교통정보가 전이되는 과정을 통하여 구간 여행시간의 추정 및 예측을 수행하도록 모형을 개발하고자 한다.



〈그림 10〉 Spacial Mapping 기반의 신경망 모형 (Multiple 구간)



〈그림 11〉 신경망 모형설계 개념도 (Single 구간)

위의 〈그림 11〉에서 제시된 바와 같이 T/G-T/G간 설치된 각각의 루프점지로부터 속도를 기반으로 다음과 같은 과정에 의하여 여행시간을 추정 및 예측하기 위한 모형을 구축한다. 4개의 VDS를 입력층으로 하고 동일시간대의 2개 TCS를 여행시간값의 출력층으로 하여 그사이에 50-50-50의 은닉층으로 신경망을 구성하였다.

[단계 1] VDS 속도 및 TCS 여행시간 수집

- ① 대상 IC-IC간 설치된 VDS로부터 속도를 입력한다. 이때 기초정보는 FTMS의 30초 Polling Data이므로 이를 Rolling Horizon Method에 의하여 단위시간 마다 대표치를 추출한다.
- ② Rolling Horizon에 의하여 대표치로서 추출된 대상 VDS별 속도와 TCS에 의한 구간여행시간을 1:1 대응시킨다.
- ③ 여행시간 예측인 경우, VDS별 속도를 예측하여 TCS의 예측 Interval 후의 여행시간과 1:1 대응시킨다.

[단계 2] 학습수행과정

학습을 수행하기 위해서는 [단계 1]에서 수집된 TCS 정보와 VDS의 속도에 다양한 환경적 요인이 고려된 총체적인 정보들임을 전제로 한다.

- ① 신경망의 초기 모수를 설정하여 수행한다.
- ② 반복학습을 수행하며, 수렴허용치에 도달 여부를 결정한다.
- ③ 수렴시 학습수행을 종료한다.

2. 모형구축의 가정

- 1) 고속도로상에서 검지되는 루프점지기의 속도정보에 대한 분포는 정규분포를 따른다.
- 2) 단일구간모형에서는 사용자가 Toll Gate와 Toll Gate간의 구간에서 고속도로 휴게소를 경유하지 않는다.
- 3) 고속도로를 이용하는 사용자는 항시운전상태임을 가정하며, 취침 및 차량정비등 그 이외의 경우는 이상치로 판단하여 제거될 수 있다.
- 4) 실험을 위하여 수집된 Data는 평상시 고속도로의 상황을 반영한 것으로 본다.
- 5) 임의이론적 특성상 베이지안 추정방법에 기초한 관계로 적용모형의 메카니즘 역시 VDS자료는 동일시간(도착시간 기준)의 TCS자료를 반영한다.

3. 모형 구조

본 연구에서는 각 IC의 VDS별 속도 정보를 입력 특성으로 하고, 각 IC간의 통행시간을 출력특성으로 하여 신경망 모형을 구축하고자 한다.⁸⁾

본 연구에서는 여행시간 추정과 예측모형으로 구분하였으며, 예측의 기본단위는 5분후, 즉 1 Step을 예측하여 입력층에 반영함으로써 5분후의 여행시간을 예측하였다. 모형의 분류는 출력층의 노드수에 따라 IC~IC간을 Single구간 모형, 여러 IC간을 Multiple구간 모형이라 칭하고자 한다. 또한 대표값, 이상치제거 유무(有無), Rolling시간 및 과거자료 반영 정도등 모형구축에 있어서 여행시간에 영향을 미치는 요인에 대하여 자료의 처리를 서로 달리하여 모형을 세분하였다.

자료의 처리방식에 따라 각 각 다른 실험번호를 부여하여 세부모형으로 구분하였다. 본 연구에서 구축된 모형은 크게 4개의 모형, 세부적으로 64개 모형에 대하여 구축된다.

본 연구에서의 모형은 Single구간 모형과 Multiple구간 모형으로 구분하며, 구축되는 모형의 구조는 다음과 같다.

Single 구간 모형은 IC~IC간의 검지기 자료를 입력자료로 이용하여 여행시간을 추정 및 예측하기 위한 모형이다. 모형은 입력층, 은닉층 3개, 각 은닉층의

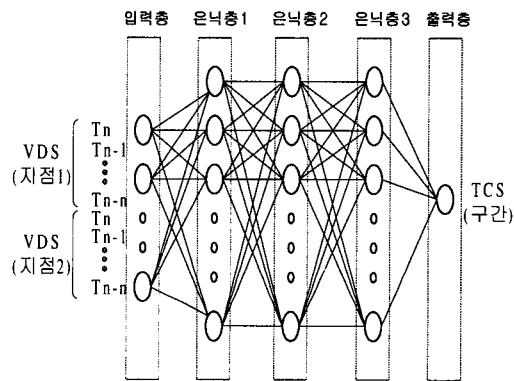
층의 노드(Node) 30, 출력층으로 구성하였다.

Multiple구간 모형은 여러 IC간의 검지기 자료를 입력자료로 이용하여 여행시간을 추정 및 예측하기 위한 모형이다. 모형은 입력층, 은닉층 3개, 출력층으로 구성되며, 각 은닉층의 노드는 조건에 따라 30/50개이며, 출력층 노드는 2개로 구성된다.

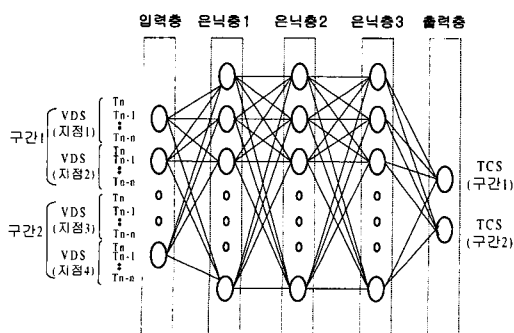
V. 모형의 평가

1. 학습 자료

본 연구에서는 고속도로교통관리시스템의 VDS별 교통정보와 TCS의 여행시간 정보를 수집하여 이용하였다. 수집자료에 대하여 학습에는 6일간의 VDS자료와 TCS자료를 이용하였으며 검증시에는 3일간의 VDS자료가 사용되었다. 또한 평가시에는 검증에 이용된 VDS자료에 대하여 TCS자료를 동일시간자료로 가공하여 사용하였다. 그 상세한 내역은 다음 <표 2>와 같다.



<그림 12> Single 구간 모형



<그림 13> Multiple 구간 모형

<표 2> 수집자료 내역

자료구분	자료특성	수집일자
신경망 학습자료	VDS (30초 Polling)	1998. 11. 13(금)
		1998. 11. 14(토)
		1998. 11. 15(일)
	TCS (1분 단위)	1998. 11. 16(월)
		1998. 11. 17(화)
		1998. 11. 18(수)
검증자료	VDS (30초 Polling)	1998. 10. 16(금)
		1998. 10. 17(토)
		1998. 10. 18(일)
평가자료	TCS (1분 단위)	1998. 10. 16(금)
		1998. 10. 17(토)
		1998. 10. 18(일)

본 연구에서는 기본적으로 신경망 모형구축에 대하여 다음과 같은 경험적인 법칙에 따라 모형에 접근하였다.¹²⁾

First Rule : Hidden = (Input+Output)/2

Second Rule : $2 \times N < \text{Number of facts for learning} < 10 \times N$

신경망 모형 구축을 위해서는 보다 다양한 시행착오(Trial Error)를 통하여 더욱 정교하고 현재의 모형에 알맞는 구성을 갖추는 과정이 필수적이므로 많은 학습과정을 통하여 최적의 학습 파라미터를 선정하고자 노력하였으며, 다음 <표 3>과 같은 결론에 도달하였다.

<표 3> 신경망 학습을 위한 모수

구축 모형	은닉층	은닉층 노드수	학습율	Epsilon	학습 횟수
Single 구간	3	30-30-30	0.7	0.1	10,000
Multiple 구간	3	50-50-50	0.7	0.1	20,000

2. 모형평가 시나리오

본 연구에서의 모형을 평가하기 위해 자료의 속성에 따라 추정과 예측으로 구분되고, 모형 내의 구간의 수에 따라 Single구간 및 Multiple구간으로 구분하여 실시하고자 한다. 각 모형은 다음과 같은 기준에 의해 구분하였다.

- ① 자료의 속성에 따라 추정모형과 예측모형으로 구분한다.
- ② 추정 및 예측하고자 하는 구간의 수에 따라 Single구간과 Multiple구간으로 구분한다.
- ③ TCS자료의 대표값은 최빈값과 평균값으로 구분한다.
- ④ 이상치 제거는 유(有), 무(無)로 구분한다.
- ⑤ TCS자료가 현재시점이므로 TCS의 20분, 또는 30분의 과거자료를 신경망 입력시 반영하기 위하여 Block Moving단위는 5분으로 하며 Rolling시간은 20분과 30분으로 구분하여 실시한다.
- ⑥ “과거자료반영”이란 과거VDS값(5분간 대표치)들, 즉 과거의 History자료를 말한다. 과거Rolling 자료의 적용회수인 반영회수는 3회와 5회로 구분한다.
- ⑦ 실험번호의 구성은
 모형 : N -> 추정, K -> 예측
 지역 : S -> 수원, O -> 오산
 번호 : 001 - 016

이상에서 제시하고 있는 기준에 따라 <표 5>, <표 6>, <표 7> 및 <표 8>과 같이 모형을 평가하고자 한다.

<표 4> Single구간추정모형의 실험계획(수원-서울)

추정구간	대표값	이상치 제거	Rolling 시간 (5분단위)	과거 자료 반영	실험 번호
Single 구간추정모형 (수원-서울)	최빈값 (Mode)	무	20분	3	NS001
			20분	5	NS002
			30분	3	NS003
			30분	5	NS004
			20분	3	NS005
			20분	5	NS006
	평균값 (Mean)	무	30분	3	NS007
			30분	5	NS008
			20분	3	NS009
			20분	5	NS010
			30분	3	NS011
			30분	5	NS012
최빈값 (Mode)	유	20분	3	NS013	
		20분	5	NS014	
		30분	3	NS015	
		30분	5	NS016	
		20분	3	NS015	
		20분	5	NS016	

<표 5> Multiple구간추정모형의 실험계획(오산-서울)

추정구간	대표값	이상치 제거	Rolling 시간 (5분단위)	과거 자료 반영	실험 번호
Multiple 구간추정모형 (오산-서울)	최빈값 (Mode)	무	20분	3	NO001
			20분	5	NO002
			30분	3	NO003
			30분	5	NO004
			20분	3	NO005
			20분	5	NO006
			30분	3	NO007
			30분	5	NO008
			20분	3	NO009
			20분	5	NO010
			30분	3	NO011
			30분	5	NO012
	평균값 (Mean)	유	20분	3	NO013
			20분	5	NO014
			30분	3	NO015
			30분	5	NO016

〈표 6〉 Single구간예측모형의 실험계획(수원-서울)

추정구간	대표값	이상치 제거	Rolling 시간 (5분단위)	과거 자료 반영	실험 번호
Single 구간예측모형 (수원-서울)	최빈값 (Mode)	무	20분	3	KS001
			20분	5	KS002
			30분	3	KS003
			30분	5	KS004
	평균값 (Mean)		20분	3	KS005
			20분	5	KS006
			30분	3	KS007
			30분	5	KS008
	최빈값 (Mode)	유	20분	3	KS009
			20분	5	KS010
			30분	3	KS011
			30분	5	KS012
평균값 (Mean)	20분		3	KS013	
	20분		5	KS014	
	30분		3	KS015	
	30분		5	KS016	

〈표 7〉 Multiple구간예측모형의 실험계획(오산-서울)

추정구간	대표값	이상치 제거	Rolling 시간 (5분단위)	과거 자료 반영	실험번호
Multiple 구간예측모형 (오산-서울)	최빈값 (Mode)	무	20분	3	KO001
			20분	5	KO002
			30분	3	KO003
			30분	5	KO004
	평균값 (Mean)		20분	3	KO005
			20분	5	KO006
			30분	3	KO007
			30분	5	KO008
	최빈값 (Mode)	유	20분	3	KO009
			20분	5	KO010
			30분	3	KO011
			30분	5	KO012
	평균값 (Mean)		20분	3	KO013
			20분	5	KO014
			30분	3	KO015
			30분	5	KO016

3. 모형 평가 기준

일반적으로 구축된 예측모형의 경우 그 수행결과를

평가하는 척도는 다양하며, 절대적인 평가기준은 없다. 단지 다음과 같은 지표를 토대로 상대적인 평가에 의존한다.

1) MARE(Mean Absolute Relation Error)

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|x(i) - \widehat{x(i)}|}{x(i)}$$

2) MAE(Mean Absolute Error)

$$MAE = \frac{1}{N} \sum_{i=1}^N |x(i) - \widehat{x(i)}|$$

3) RMSE(Root Mean Square Error)

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N |x(i) - \widehat{x(i)}|^2}$$

4) EC(Equality Coefficient)

$$EC = 1 - \frac{1}{N} \frac{\sqrt{\sum_{i=1}^N |x(i) - \widehat{x(i)}|^2}}{\sqrt{\sum_{i=1}^N x(i)^2} + \sqrt{\sum_{i=1}^N \widehat{x(i)}^2}}$$

이러한 지표중 널리 이용되는 지표는 MARE이다. 그러나 본 연구에서는 상기의 제시된 각 지표들을 기준으로 구축된 모형의 효율성을 평가한다.

4. 모형 평가 결과

1) 여행시간 추정모형 평가결과

평가결과 Single여행시간 추정모형(수원-서울)은 모두 만족스러운 결과를 보였다. 그러나 모형 NS005~8은 상대적으로 많은 오차를 보이는 것으로 나타났다. 그 이유는 평균값이라는 파라미터가 휴게소 장시간이용차량, 갯길에서의 수면 및 휴식차량, 공공차량 등 일부차량이 전체 여행시간에 미치는 영향이 매우 크게 작용하기 때문이다. 특히 짧은 구간의 경우 여행시간 역시 짧기 때문에 긴 구간에 비해 상대적으로 많은 영향을 받아 대표값의 편이(Bias)가 발생하기 때문인 것으로 판단된다.

Multiple여행시간추정모형(수원-서울)은 모형NO005~8이 상대적으로 많은 오차를 보이는 것으로 나타났다. 그 이유는 일부차량의 여행시간이 각 자료군의

여행시간에 미치는 영향이 매우 크게 작용하기 때문이다. 특히 짧은 구간의 경우 여행시간 역시 짧기 때문에 긴 구간에 비해 상대적으로 많은 영향을 받아 대표값의 편이(Bias)가 발생하기 때문인 것으로 판단된다.

더불어 Single 추정모형(수원-서울)과 Multiple 추정모형(수원-서울)을 구분하여 분석한 이유는 Multiple 구간의 모형구조상 출력층은 구간내의 TCS데이터이므로 설계된 '오산 → 서울' 구간에 대한 신경망의 출력층에는 '오산 → 수원'과 '오산→서울'에 대한 결과가 각각 출력되기 때문에 구간별로 Single모형과 비교하여 보기 위함이다.

Multiple여행시간추정모형(오산-서울)은 모형중 NO004가 상대적으로 많은 오차를 보이는 것으로 나타났다. 그 이유는 대표값으로서 최빈값이 가진 단점으로서 동일한 값을 가진 편이된 자료가 일부지역에 집중되어 있을 때 대표값으로서의 기능을 상실할 수 있다는 점을 보여주며, 다른 한편으로 신경망의 학습을 충분히 하여 모형의 안정화가 필요하다는 점을 의미한다. 또한 모형NO0013~16도 상대적으로 많은 오차를 보이는 것으로 나타났다.

<표 8> Single구간추정모형의 평가결과(수원-서울)

실험번호	MARE	MAE	RMSE	EC	Max Error
NS001	0.1237	1.6613	0.0850	0.9164	8.435
NS002	0.1222	1.6366	0.0861	0.9179	8.659
NS003	0.1240	1.6599	0.0853	0.9162	8.716
NS004	0.1298	1.7950	0.0943	0.9120	8.542
NS005	0.2303	2.8520	0.2962	0.7569	76.827
NS006	0.2090	2.6135	0.2968	0.7616	76.876
NS007	0.2303	2.8405	0.2986	0.7556	76.801
NS008	0.2104	2.6265	0.2990	0.7605	76.885
NS009	0.1286	1.8024	0.0989	0.9140	8.542
NS010	0.1293	1.7969	0.1012	0.9154	8.748
NS011	0.1296	1.7975	0.0984	0.9141	8.811
NS012	0.1311	1.8056	0.1023	0.9143	9.197
NS013	0.1232	1.7048	0.0937	0.9165	7.632
NS014	0.1248	1.7018	0.0954	0.9183	7.849
NS015	0.1252	1.7125	0.0939	0.9162	8.106
NS016	0.1273	1.7275	0.0974	0.9166	8.506

<표 9> Multiple구간추정모형의 평가결과(수원-서울)

실험번호	MARE	MAE	RMSE	EC	Max Error
NO001	0.1448	1.9930	0.0985	0.9062	11.770
NO002	0.1386	1.9169	0.0971	0.9108	8.838
NO003	0.1445	1.9750	0.0965	0.9085	9.274
NO004	0.1615	2.3751	0.1281	0.8848	16.029
NO005	0.2266	2.9955	0.2942	0.7642	76.107
NO006	0.2108	2.9060	0.2963	0.7706	75.913
NO007	0.2250	3.0037	0.2957	0.7647	75.924
NO008	0.2129	2.9477	0.2990	0.7697	75.959
NO009	0.1432	2.0959	0.1133	0.9047	11.383
NO010	0.1422	2.1496	0.1195	0.9047	9.372
NO011	0.1416	2.0597	0.1129	0.9049	10.677
NO012	0.1361	2.0460	0.1165	0.9065	8.938
NO013	0.1743	2.6054	0.1330	0.8883	10.202
NO014	0.1734	2.6412	0.1415	0.8868	9.962
NO015	0.1775	2.6633	0.1369	0.8855	9.457
NO016	0.1815	2.8194	0.1505	0.8807	9.934

<표 10> Multiple구간추정모형의 평가결과(오산-서울)

실험번호	MARE	MAE	RMSE	EC	Max Error
NO001	0.1340	3.7911	0.2000	0.9147	24.817
NO002	0.1317	3.6984	0.1937	0.9203	16.979
NO003	0.1346	3.8038	0.2074	0.9120	27.791
NO004	0.1912	7.2129	0.4775	0.8215	54.438
NO005	0.1196	3.5057	0.1772	0.9288	21.169
NO006	0.1237	3.7324	0.1937	0.9257	18.114
NO007	0.1214	3.5979	0.1895	0.9242	24.678
NO008	0.1240	3.7301	0.1875	0.9284	16.028
NO009	0.1336	4.0092	0.2350	0.9118	25.973
NO010	0.1376	4.2484	0.2582	0.9092	23.333
NO011	0.1329	3.9257	0.2366	0.9115	29.843
NO012	0.1250	3.7861	0.2355	0.9168	25.729
NO013	0.2226	7.2607	0.3735	0.8668	23.299
NO014	0.2239	7.5129	0.4018	0.8645	23.662
NO015	0.2280	7.5176	0.3852	0.8637	23.901
NO016	0.2471	8.8108	0.4625	0.8479	27.452

2) 여행시간 예측모형 평가결과

Single여행시간 예측모형(수원-서울)은 모두 만족스

러운 결과를 보였다. 그러나 모형KS005~8은 상대적으로 많은 오차를 보이는 것으로 나타났다. 그 이유는 평균값이라는 파라미터가 휴게소 장시간이용차량, 갓길에서의 수면 및 휴식차량, 공공차량 등 일부차량이 전체 여행시간에 미치는 영향이 매우 크게 작용하기 때문이다. 특히 짧은 구간의 경우 여행시간 역시 짧기 때문에 긴 구간에 비해 상대적으로 많은 영향을 받아 대표값의 편이(Bias)가 발생하기 때문인 것으로 판단된다.

Multiple여행시간예측모형(수원-서울)은 모형KO005~8이 상대적으로 많은 오차를 보이는 것으로 나타났다. 그 이유는 일부차량의 여행시간이 각 자료군의 여행시간에 미치는 영향이 매우 크게 작용하기 때문이다. 특히 짧은 구간의 경우 여행시간 역시 짧기 때문에 긴 구간에 비해 상대적으로 많은 영향을 받아 대표값의 편이(Bias)가 발생하기 때문인 것으로 판단된다. 또한 모형KO0013~16도 상대적으로 많은 오차를 보이는 것으로 나타났다.

Multiple여행시간예측모형(오산-서울)은 모형NO0013~16이 상대적으로 많은 오차를 보이는 것으로 나타났다. 모형구성이 유사한 모형NO004에 비해 상대적으로 낮은 오차율을 보이는 이유는 Kalman Filtering 알고리즘을 이용한 예측과정에서 Kalman Filter의

예측값이 자료의 패턴을 양성하여 신경망 알고리즘의 설명력에 기여하기 때문인 것으로 판단된다.

〈표 12〉 Multiple구간예측모형의 평가결과(수원-서울)

실험번호	MARE	MAE	RMSE	EC	Max Error
KO001	0.1440	1.9831	0.1002	0.9062	9.6618
KO002	0.1511	2.1246	0.1066	0.9046	9.7090
KO003	0.1577	2.2013	0.1108	0.8972	9.3357
KO004	0.1492	2.1268	0.1094	0.9025	7.9606
KO005	0.2185	2.9297	0.2973	0.7654	76.0332
KO006	0.1959	2.7282	0.2792	0.7846	75.8460
KO007	0.2197	2.9268	0.3004	0.7638	76.1477
KO008	0.1933	2.6754	0.2797	0.7842	75.7432
KO009	0.1445	2.1149	0.1177	0.9032	9.9764
KO010	0.1399	2.0838	0.1210	0.9046	10.2944
KO011	0.1506	2.2613	0.1294	0.8940	10.0350
KO012	0.1416	2.1185	0.1192	0.9063	7.8984
KO013	0.1766	2.6805	0.1423	0.8836	9.1193
KO014	0.1743	2.7135	0.1492	0.8834	9.9586
KO015	0.1791	2.7261	0.1445	0.8822	8.8605
KO016	0.1874	2.9575	0.1618	0.8748	9.2956

〈표 11〉 Single구간예측모형의 평가결과(수원-서울)

실험번호	MARE	MAE	RMSE	EC	Max Error
KS001	0.1249	1.6876	0.0872	0.9159	8.7138
KS002	0.1237	1.6656	0.0887	0.9169	8.8658
KS003	0.1244	1.6732	0.0874	0.9156	8.9362
KS004	0.1251	1.6776	0.0906	0.9151	9.1848
KS005	0.2206	2.7578	0.2983	0.7589	76.8330
KS006	0.1931	2.4657	0.2799	0.7762	76.8814
KS007	0.2216	2.7511	0.3007	0.7571	76.8732
KS008	0.1948	2.4797	0.2823	0.7746	76.9656
KS009	0.1297	1.8233	0.1013	0.9140	8.8396
KS010	0.1311	1.8309	0.1044	0.9145	8.9838
KS011	0.1304	1.8155	0.1013	0.9137	9.0437
KS012	0.1327	1.8359	0.1056	0.9131	9.3059
KS013	0.1246	1.7196	0.0956	0.9168	8.0273
KS014	0.1268	1.7304	0.0982	0.9175	8.1252
KS015	0.1263	1.7295	0.0966	0.9158	8.3673
KS016	0.1289	1.7545	0.1003	0.9155	8.6172

〈표 13〉 Multiple구간예측모형의 평가결과(오산-서울)

실험번호	MARE	MAE	RMSE	EC	Max Error
KO001	0.1433	4.1145	0.2210	0.9074	20.6516
KO002	0.1402	4.0609	0.2285	0.9093	23.9299
KO003	0.1429	4.1907	0.2464	0.8989	24.8448
KO004	0.1386	4.1132	0.2479	0.9025	26.5596
KO005	0.1261	3.7892	0.1976	0.9221	18.6599
KO006	0.1258	3.8294	0.1984	0.9248	15.8462
KO007	0.1285	3.8909	0.2058	0.9193	19.0791
KO008	0.1257	3.8095	0.2001	0.9245	18.3042
KO009	0.1441	4.4074	0.2624	0.9037	21.5727
KO010	0.1382	4.3892	0.2974	0.8971	29.6317
KO011	0.1456	4.7624	0.3300	0.8821	32.2495
KO012	0.1346	4.1055	0.2645	0.9082	24.8047
KO013	0.2194	7.2859	0.3938	0.8631	24.0883
KO014	0.2323	8.1782	0.4483	0.8538	26.3697
KO015	0.2203	7.2839	0.3941	0.8632	24.0554
KO016	0.2307	8.1040	0.4454	0.8545	24.8445

3) FTMS 여행시간

신경망 모형의 여행시간 추정 방식과 현재 적용되고 있는 TCS 데이터 도착시간 기준의 VDS 평균속도에 의한 여행시간 산출 방식과의 비교/분석을 위해 기존의 여행시간 산출 방식을 평가한 결과는 다음과 같다.

<표 14> 기존 FTMS의 여행시간 평가 결과

추정 구간	지역	이상치 제거	대표값	MARE	MAE	RMSE	EC	Max Error
Single 구간	수원	무	최빈값	0.6643	4.9023	0.2341	0.6944	13.9994
			평균값	1.0671	7.4871	0.4158	0.5458	80.8910
		유	최빈값	0.7162	5.3619	0.2709	0.6846	13.9994
			평균값	0.6803	5.0902	0.2561	0.6970	13.8139
Multiple 구간	수원	무	최빈값	0.5937	4.5201	0.2132	0.7289	14.4082
			평균값	0.9904	7.1050	0.4028	0.5695	80.7269
		유	최빈값	0.6363	4.9279	0.2463	0.7213	14.4082
			평균값	0.6045	4.6563	0.2329	0.7322	13.3832
	오산	무	최빈값	0.4933	9.4825	0.5461	0.7020	41.1341
			평균값	0.6574	11.9608	0.5939	0.6879	38.6454
		유	최빈값	0.5226	10.3556	0.6303	0.6943	41.1341
			평균값	0.5029	9.7364	0.5708	0.7163	33.0665

VI. 결론 및 향후연구방향

본 논문은 고속도로교통관리 시스템에서 수집되는 기초 교통정보를 이용하여 고속도로의 영업소간 여행시간 추정 및 예측 알고리즘 개발을 수행하였다. 본 연구에서 목적인 알고리즘은 현장에서 실시간으로 수집된 지점교통정보를 이용하여 Spacial Mapping 개념 기반의 신경망 모형과 동적기반의 칼만필터 모형을 적용하였다. 본 논문이 목적인 개발 모형의 비교우위를 결정하기 위하여 기존에 운영되고 있는 공간 평균속도 추정모형에 의한 결과와 비교 검토하였다. 그러나 본 논문은 제한된 시간·공간적 자료를 바탕으로 작성되었으므로 학습되지 않은 교통상황의 경우에는 적용에 신중할 필요가 있다.

본 연구를 통한 종합적 결론은 다음과 같다.

첫째, 여행시간의 추정 및 예측시 단위 시간 수집 자료 군의 대표치 추출방법은 최빈값이 평균값보다 우수하다.

본 연구에서는 단위 시간당 수집되는 TCS(Toll Collection System) 정보의 산포를 안정화시키기

위하여 Rolling기법을 적용하였으며, 이때 Rolling처리를 위한 기본 단계로서 단위 시간당 수집된 자료의 대표치를 추출하기 위한 방법으로 최빈값과 평균값방법을 적용하였다. 추정모형과 예측모형 모두 자료의 수량이 충분한 경우 최빈값이 우수한 것으로 나타났다. 그러나 자료수가 불충분한 경우 이상치를 제거한 후 평균값을 사용하여도 무방하다.

둘째, 수집된 교통정보에 대하여 이상치를 제거하면 상대적으로 더 우수한 여행시간추정 및 예측결과를 기대할 수 있다.

본 연구에서는 수집된 교통정보에 대하여 이상치의 영향완화기법으로 Rolling기법을 적용하였으며, 이외 통계적 방법을 이용하여 별도의 이상치 제거방법을 적용하였다. 이러한 이상치 제거는 고속도로의 각 영업소를 통과하는 모든 차량들 중 정상적인 고속도로 통행 이외 휴게소이용차량 및 갓길이용 등에 위한 것으로 제거하지 않을 경우 추정 및 예측시 심각한 예측 오차를 유발할 수 있는 요인에 해당한다. 본 연구에서 이 방법을 적용한 결과 이상치를 제거한 경우, 보다 우수한 정확도를 나타냈다.

셋째, 고속도로 교통관리 시스템으로부터 수집되는 VDS 정보를 기반으로 한 이론적 여행시간과 TCS를 통한 실측된 여행시간은 다르다.

VDS에서 계측되는 교통정보는 지점교통정보로서 이들을 공간평균속도로 환산했을 경우와 실제 TCS에서 계측된 각 영업소간 여행시간을 비교할 경우, 이들은 많은 차이를 갖는다. 본 연구에서는 이들에 대하여 MAE, EC, MARE, RMSE, 최대오차 등의 평가지표를 토대로 비교하였다.

넷째, 여행시간 추정 및 예측을 위한 신경망 모형의 경우 VDS와 TCS간의 이론적 차이를 다소 해결할 수 있었다.

본 연구에서는 상기 3항에서 기술한 바와 같은 문제점을 해결하기 위한 방법으로 신경망 모형에 칼만필터모형을 적용하였다. 적용한 결과 칼만필터모형을 적용한 신경망 모형이 VDS정보에 의해 산정된 여행시간보다 상대적으로 더 우수한 결과를 나타냈다.

참고문헌

1. 도철용, 교통공학원론(상), 청문각, 1996.
2. 허문열, 수리통계학, 경문사, 1989.

3. 김원경, 시계열분석, 경문사, 1996.
4. 김성인, 통계적 품질관리, 박영사, 1989.
5. 김규옥, "Kalman-Filtering 모델을 이용한 고속도로상에서의 교통량 추정", 공학석사학위 논문, 아주대학교 대학원, 1996.
6. Bernard Rosner, "Percentage Point for a Generalized ESD Many Oulier Procedure", Technometrics ©, Vol. 25, No. 2, May 1983.
7. S. Cohen and C. Nouveliere, "Travel Time on Urban Controlled Links : A Neural Network Approach", 1997.
8. Optimization of Railway Operations Using Neural Networks, Transpn Res, 1993.
9. G. N Wang, "An Adaptive Hybrid Neural Network Approach to Prediction of Nonstationary Process", Hybrid Intelligent System Application, Edited by Dr. Jay Liebowitz, IOS Press, 1995.
10. Wei C.H and Schonfeld, P.M. An Artificial Neural Network Approach For Estimating Multiperiod Travel Time In Transportation Networks, Presented At The 71st Annual Meeting Of The Transportation Research Board, 1993.
11. L. Fausett, Fundamentals of Neural Networks-Architectures, Algorithms, and Applications, Prentice Hill, 1994.
12. J Lawrence, Introduction to Neural Networks, Design, Theory, and Applications, California Scientific Software Press, 1993