

ITS를 위한 차량검지시스템을 기반으로 한 교통 정체 예측 모듈 개발

신원식¹ · 오세도² · 김영진^{2*}

¹한국생산기술연구원 로봇종합지원센터 / ²경희대학교 산업경영공학과

Development of Traffic Congestion Prediction Module Using Vehicle Detection System for Intelligent Transportation System

Won-Sik Sin¹ · Se-do Oh² · Young-Jin Kim²

¹Korea Institute Robot Industry Advancement

¹Department of Industrial and Management Systems Engineering, Kyung Hee University

The role of Intelligent Transportation System (ITS) is to efficiently manipulate the traffic flow and reduce the cost in logistics by using the state of the art technologies which combine telecommunication, sensor, and control technology. Especially, the hardware part of ITS is rapidly adapting to the up-to-date techniques in GPS and telematics to provide essential raw data to the controllers. However, the software part of ITS needs more sophisticated techniques to take care of vast amount of on-line data to be analyzed by the controller for their decision makings. In this paper, the authors develop a traffic congestion prediction model based on several different parameters from the sensory data captured in the Vehicle Detection System (VDS). This model uses the neural network technology in analyzing the traffic flow and predicting the traffic congestion in the designated area. This model also validates the results by analyzing the errors between actual traffic data and prediction program.

Keyword: ITS(Intelligent Transportation System), neural network, VDS(Vehicle Detection System), traffic congestion

1. 서론

ITS는 도로와 차량 등 기존 교통의 구성요소에 첨단 전자, 정보, 통신 기술을 적용시켜 교통시설을 효율적으로 운영하고, 도로 통행자에게 유용한 정보를 제공하는 시스템이다. 이로 인해 교통체계 운영효율성 및 용량의 증가를 통한 교통 혼잡 완화, 여행자 서비스 개선을 통한 운전자의 이동성, 편의성 및 안전성 향상, 교통 시스템의 안전성 제고를 도울 수 있다. ITS

는 ATMS(Advanced Traffic Management System)와 ATIS(Advanced Traveler Information System)를 포함하는 상위 개념이다.

ATMS는 도로상에 차량 특성, 속도 등의 교통 정보를 감지할 수 있는 시스템을 설치하여 교통 상황을 실시간으로 분석하고, 이를 토대로 도로 교통의 관리와 최적 신호 체계의 구현을 꾀하는 것이다. 또한 ATIS는 교통 여건, 도로 상황, 출발지에서 목적지까지의 최단 경로, 소요 시간, 주차장 상황 등 각종 교통 정보를 운전자에게 제공하여 안전하고 원활한 최적 교통을 지

본 연구는 2007년도 경희대학교 연구지원에 의한 결과임(KHU-20070747).

*연락처 : 김영진 교수, 446-711 경기도 용인시 기흥구 서천동 경희대학교 국제캠퍼스 공학대학관 432호,

Fax : 031-203-4004, E-mail : yjkim@khu.ac.kr

투고일(2010년 03월 09일), 심사일(1차 : 2010년 03월 20일, 2차 : 2010년 05월 20일, 3차 : 2010년 08월 28일), 게재확정일(2010년 10월 21일).

원하는 시스템이다.

이 두 가지 시스템을 구축하는 데에 있어 도로상의 차량을 검지하는 차량검지시스템(VDS, Vehicle Detection System)의 통계정보가 중요한 기반이 될 수 있다. 이에 차량검지시스템의 성능적 측면의 개선 연구와 더불어, 축적된 데이터를 이용한 정체 예측 및 자동차의 주행시간 예측에 대한 연구가 여러 방면으로 진행되고 있다.

차량검지시스템의 성능적 측면 개선 연구는 Cheu *et al.*(2002)과 Faouzi(2006)의 연구를 예로 들 수 있다. Cheu *et al.*(2002)은 GPS 데이터를 활용한 자료융합을 통한 주행 속도 검지능력 향상에 대한 연구를 진행하였으며, Faouzi(2006)는 루프검지기 데이터의 베이저안 접근법과 사례적용 결과를 제시하였다. 이러한 연구들은 차량검지시스템의 검출능력과 신뢰도를 높이기 위한 방법론을 제시한 연구들이며, 이는 교통량, 차량속도 등의 정보의 측정에 있어 정확성을 향상시킨 차량 검지 시스템을 만드는 데 기여하였다.

차량검지시스템의 정보를 바탕으로 교통정체 및 주행시간 예측에 관한 연구 또한 활발히 진행되고 있다. Kwon *et al.*(2000)은 출발 시간 범위를 고려한 선형회귀(linear regression) 주행 시간 예측 알고리즘을 제안하였다. 루프검지기를 바탕으로 얻어진 데이터를 이용해 전체 예측 경로 중 출발하고자 하는 시간과의 시차 20분 이내까지는 현재 관측되고 있는 차량 주행 시간을 이용해 소요시간의 계산을 수행하고, 시차가 20분 이상일 경우 하루 전의 차량 주행 평균 시간 기록을 입력으로 한 선형회귀 알고리즘을 통해 주행시간을 예측한다. Chen and Chien(2001)의 연구에서는 도로 네트워크 상에서 운전자가 이용하는 주요 경로형태를 미리 정의하고, 각각의 유형별로 분류 후 주요 경로들을 구성하는 모든 도로 구간에서 각각의 주행 시간을 측정하였다. 이를 통해 일별 혹은 주별 교통 상황의 변동에 따른 영향을 감소시켜 시뮬레이션하는 방법을 사용하였다. Rice and Van Zwet(2004)는 유사 주행 시간패턴을 고려한 최근접(nearest neighbors) 주행 시간 예측 알고리즘을 제안하였다. 즉 현재 관측되는 데이터를 토대로 과거 주행시간 기록 중 현재 날짜와 가장 유사한 일자의 데이터를 탐색하여 혼잡도를 예측하였다.

본 논문은 차량검지시스템을 통해 제공되는 다양한 데이터를 이용하여 특정 지점에서의 정체를 예측하는 시스템을 구축하는 과정을 체계적으로 서술하고 있다. 먼저 정체 예측 시스템을 구축하기에 앞서 도로의 형태를 모델링하여 정체에 영향을 줄 수 있는 변수들을 선정하였다. 이 변수들 중 차량검지시스템에 의해 기 축적된 도로의 통행량 정보와 도로정보데이터(도로의 특성[길이, 교차로 수 등])를 정체예측시스템의 입력변수로 선정하였으며 차량검지시스템에서 검지된 정체발생 횟수를 바탕으로 정체정도를 이산적으로 표현하여 정체예측시스템의 목표변수로 선정하였다. 이러한 패턴을 이용해 다층신경망 정체 예측 시스템을 모델링하고, 학습을 통해 모델의 결합가중치를 계산하였다. 이 모델을 이용해 학습에 관여하지 않은 검증용 데이터를 이용하여 출력변수인 특정지역의 교통

정체 정도를 예측하고 이에 대한 예측성공률을 분석하였다.

2. 기반기술 정리

2.1 차량검지시스템 및 이력자료 이용자서비스

차량검지시스템이란 무인감시카메라(CCTV)또는 센서와 같은 입력수단을 이용하여 차량을 검지하는 시스템이다. 특히 CCTV를 이용한 차량검지시스템은 교통변수 데이터들을 실시간 화면으로 검지할 수 있다. 도로와의 비접촉식으로 설치가 용이할 뿐 아니라 1대의 카메라에서 다양한 교통변수 계측이 가능하므로 적은 비용으로 많은 기능을 수행할 수 있어 근래 들어 도입이 보편화 되고 있다.

미국의 경우 이러한 차량검지시스템을 이용하여 실시간 수집자료를 축적하고 이력자료를 구축함과 동시에, 이력자료를 이용자가 이용할 수 있는 이용자 서비스인 ADUS(Archived Data User Service)까지도 도입되고 있다.

하지만 국내에서는 ADUS와 같은 서비스가 실시되지는 못하고 있으며 아직까지는 학계에서 ADUS의 도입방안을 연구하는 수준에 그치고 있다. 일례로 Kim(2007)의 차량검지기 이력자료 이용자 서비스 도입방안에 대한 연구에서 이력자료 이용자 서비스의 세세한 분석과 더불어 ITS의 발전을 위해서는 ADUS의 도입이 추진되어야 한다는 결론을 도출하고 있다.

본 연구에서는 ADUS의 일환으로 제공된 미국 캘리포니아 주의 차량검지시스템 데이터를 활용하여 정체 예측시스템을 구축하고자 한다.

2.2 신경망 이론(Neural Network)

신경망 이론은 대표적인 Black Box 모델로 여러 분야에 널리 사용되고 있으며 특히 패턴인식과 분류에 강한 성능을 보인다.

본 논문에서 신경망은 기반 데이터에서 모델링된 도로의 특징 패턴을 추출하여 입력값(X)로 이용한다. 이후 역전파 알고리즘을 이용한 학습을 통해 가중치를 확정하여 정체 예측 시스템을 구축하게 된다. 신경망 모델 구현에는 저자들이 MATLAB을 이용하여 자체적으로 개발한 프로그램을 사용하게 되며 일반적인 신경망 모델의 개략도는 <Figure 1>과 같다.

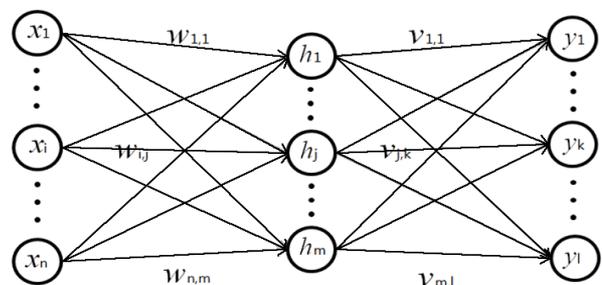


Figure 1. Schematic diagram of neural network

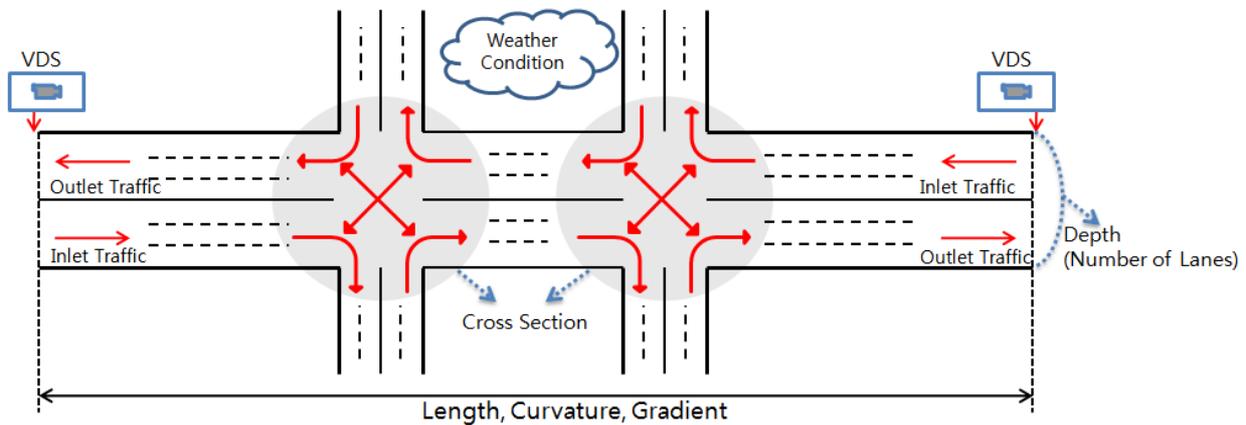


Figure 2. Schematic diagram of road modeling

2.3 도로 모델링 (Road modeling)

본 연구에서는 기존의 도로를 학습할 시스템에 맞추어 특정 변수로 모델링하였고, 이에 대한 개략도는 <Figure 2>와 같다.

각각의 도로는 유입과 유출이 일어나는 지점에서 차량검지 시스템을 통해 차량 대수를 검사하고 이를 데이터베이스에 저장하고 있다. 도로의 중간에는 교차로가 존재하여 차량의 유입과 유출이 일어나고, 이로 인해 정체가 발생된다. 도로의 물리적인 특성이라 할 수 있는 길이와 너비(차선의 수) 또한 정체에 영향을 미치는 중요한 요소이며 도로의 유형 즉 고속도로 및 자동차전용 도로, 외곽도로 등과 같은 도로의 종류 또한 정체에 영향을 미치는 요소이다. 이외에도 도로의 경사, 굽은 정도, 날씨, 사고의 유무, 차선의 수, 측정된 시간(출퇴근 시간, 주말 등)과 같은 수많은 변수들이 고려될 수 있다. 이러한 정체 예측시스템을 구성하는데 의미 있는 변수들에 대한 설명은 <Table 1>과 같다.

Table 1. Description of variables for prediction System

변수	설명
유입 차량 대수	도로로 진입하는 차량 수
유출 차량 대수	도로에서 유출되는 차량 수
교차로 수	도로상의 교차로의 수
도로의 특성	국도, 지방도, 고속도로 등 과 같은 도로의 특성
날씨	당일의 날씨
도로의 경사 및 굽은 정도	평탄한 정도(오르막 또는 내리막)와 직선주요 인지 곡선주요인지에 대한 정보
차선의 수	도로의 차선 수
측정된 시간	측정된 시점이 혼잡도에 영향을 줄 수 있는 특성들(출퇴근시간, 주말, 공휴일연휴)

본 연구는 정체 예측 시스템을 구축하기 이전에 정체에 관

여하는 도로의 변수들을 파악하여 모델링 하였다. 이러한 변수들을 0과 1사이의 값으로 표준화하여 정체 예측 시스템의 입력변수로 사용할 수 있다.

3. 차량검지시스템 데이터를 이용한 학습

3.1 학습을 위한 데이터 선정

본 논문에서는 차량검지시스템을 이용한 자료 중, 미국 캘리포니아 주 Bay Area 근처의 도시인 Santa Clara, San Francisco, Napa, Solano, San Benito, San Bruno, Sonoma, Yolo의 8개의 도시에서 조사된 자료를 기반으로 한다.

이 지역에는 MTC(Metropolitan Transportation Commission, <http://www.mtc.ca.gov/>)가 Highway Capacity Manual과 통계내용을 포함한 문서를 제작하기 위해 각 도로를 동적 차량 조사법으로 조사하며, 이를 통해 각 도로에 대해 AADT(Annual Average Daily Traffic)와 PHV(Peak Hour Volume)등의 데이터를 파악하고 있다. 또한 각 도로에 대한 차량 종류나 해당지역 차량보유 현황, 출퇴근 시 차량 이용 수준까지도 조사하고 있다. 주로 시내 지역을 핵심적으로 조사하지만 지역 전체의 조사도 병행하여 자료를 제공하고 있다. 본 논문에서는 이러한 각 도로의 상태에 대한 문서를 MTC로부터 제공 받아 본 논문의 기초자료로 사용한다.

3.2 입력변수 선정

MTC에서 도로의 평균 차량 대수의 수집 과정과 그 결과에 대한 기초 보고는 MTC PEAK HOUR VOLUME DATA라는 이름의 보고서로 제출된다. 이 보고서는 MTC의 홈페이지인 <http://www.mtc.ca.gov/>에 공개되어 있으며, 이를 기반으로 <Table 2>에 세부적으로 조사된 피크시간/피크 월별(Peak Hour/Peak Month) 상/하행선(Ahead/Back)의 통행차량대수 및 1년 동안의 통과차량대수(AADT)를 정리했다. 특히 이 자료상에 제시되어 있는

Table 2. MTC Peak Hour Volume Data 2007(in part)

District	Route	County	Postmile	Description	Back Peak Hour	Back Peak Mouth	Back AADT	Ahead Peak Hour	Ahead Peak Month	Ahead AADT
12	1	ORA	0.129	DANA POINT, JCT.RTE.5				3800	40500	37500
12	1	ORA	0.780	DANA POINT, DOHENY PARK ROAD	3800	40500	37500	4000	43000	39500
12	1	ORA	1.150	DOHENY BEACH, DEL OBISPO STRET/DANA POINT DRIVE	4000	43000	39500	2350	32000	29500
12	1	ORA	4.320	DANA POINT, CROWN VALLEY PARKWAY	2350	32000	29500	2900	39000	36500

통과 차량대수(AADT)를 바탕으로 각각의 도로에서 출입된 자동차의 대수를 산출해 낼 수 있으며 이 값은(특정 도로에 대한 차량의 유입 및 유출대수)정체에 있어 가장 큰 영향을 줄 수 있는 변수이다.

해당도로의 차선의 수 또한 정체에 큰 영향을 주는 변수이다. 해당 도로의 차선 수는 본 논문에서 인용한 통계적 자료에는 포함되어 있지 않아 구글(Google)사의 구글어스(Google-Earth)를 통하여 조사하였다. 구글어스는 전 세계를 대상으로 위성사진, 항공사진을 조합하여 가상의 지구를 확대, 축소하여 볼 수 있는 시스템이다. 지역별로 해당도로를 검색, 확대하여 차선의 개수를 파악하였다. 조사 시 차선의 수가 줄어들거나 늘어나는 교차로 연결 부위는 고려하지 않았으며 교차로가 끝나는 주로의 차선 수만을 조사하였다.

이렇게 조사된 해당 도로의 유입, 유출대수와 차선의 수를 바탕으로 차선 당 유입차량대수와 차선 당 유출차량대수를 계산할 수 있으며 이 두 변수를 첫째와 둘째 입력변수로 선정했다.

또한 각 도로와 연결되는 다른 도로에 대한 정보를 토대로 도로의 교차로수를 파악하여 해당 도로에 대한 복잡도를 파악하였다. 교차로 수가 많아질수록 원활한 교통의 흐름에 방해가 되기 때문이다. 그러나 교차로의 수는 해당 도로의 길이에 의해서 영향력이 다를 수 있다. 예를 들어 같은 개수의 교차로를 가진 두 도로가 있다 하더라도 도로의 길이가 긴 경우 교차로에 의한 정체가 일어날 가능성도 줄어들게 될 것이다. 그러므로 교차로수를 도로의 길이로 나눈 단위길이 당 교차로수를 세 번째 입력변수로 사용한다.

네 번째 입력변수는 도로의 특성치이다. 도로는 크게 일반도로(시내도로), 외곽도로, 주도로, 고속도로의 4가지로 나눌 수 있다. 이중 일반도로와 외곽도로는 그 특성이 유사하여 단일하게 취급하여 3가지 단계로 분류한다.

본 논문에서는 이와 같은 입력변수를 가진 다층 신경망 모델을 이용한 정체 예측 시스템을 구축해 연구를 진행하였다. 입력변수 중 유입차량대수와 유출 차량대수는 다른 변수에 비해 매우 큰 단위의 값이고, 또 다른 변수인 단위길이 당 교차로 수는 그에 비해 매우 작은 단위의 값이다. 그러므로 학습을 진

행하기 전 정규화 과정을 거쳐야 원활한 학습이 가능하다. 이론적으로 0과 1사이의 값으로 정규화 하는 것이 원활한 학습에 도움이 되므로 해당 입력변수 값을 최대값과 최소값의 차로 나누는 방식으로 정규화 하였다. 또한 유입차량대수와 유출차량대수는 단위가 같고 이들의 관계 또한 정체 시 중요한 요소로 작용할 수 있다. 때문에 이 두 변수들의 상관관계를 보전하기 위해 이 두 입력변수를 통합하여 정규화시켰다. 유입차량대수와 유출차량대수 이외의 나머지 변수들은 각각 따로 정규화시켰다.

신경망 모델에 적용한 입력변수를 요약하면 <Table 3>과 같다.

Table 3. Description of input variables for neural network model

입력 변수	설 명
차선 당 유입차량대수	도로로 진입하는 차량 수÷ 차선의 수
차선 당 유출차량대수	도로에서 유출되는 차량 수÷ 차선의 수
단위길이 당 교차로 수	도로상의 교차로 수÷도로의 길이
도로의 특성	일반도로 및 외곽도로(0.3), 주도로(0.6), 고속도로(0.9)

이외에 고려될 수 있는 입력변수, 즉 교통량에 영향을 줄 수 있는 변수들로는 도로의 경사, 굽은 정도, 날씨, 사고의 유무, 측정된 시간(출퇴근 시간, 주말 등) 등이 있는데 현재로서는 이를 반영할만한 자료가 부족하므로 향후의 연구과제로 남겨둔다.

3.3 목표변수 선정

선정된 입력변수들을 연결시켜 학습할 목표변수의 선별 또한 모델의 예측 정확도에 있어 매우 중요한 요소이다.

각각의 도로에 대한 시간대별 통행시간 또는 평균 통행 소요시간 데이터 중 하나가 존재 한다면 통행 소요시간에 대한 예측 시스템을 구축할 수 있다. 그러나 현재로서는 각각의 도

로별 통행 소요시간에 대해 공개된 데이터가 확인되고 있지 않다. 따라서 그에 대한 대안으로 혼잡정도를 예측하는 시스템을 구축했다. 이에 기반이 된 근거 자료는 도로의 정체 횡수를 MTC는 미국의 캘리포니아주 주요도로의 차량검지시스템을 토대로 매년 각 도로에 대한 정보를 총합해 도로별 통행량을 산출하고 도로와 도로가 만나는 교차점을 기준으로 정체횡수를 측정한다. 측정된 정체횡수는 곧 교차로에서의 정체정도를 나타내며 이 측정값은 1995년부터 지속적으로 수집되고 있다. 이 값은 회계연도 기준으로 분류되어 Observed Traffic Counts라는 이름으로 <Table 4>처럼 발행된다. 본 연구에서는 이러한 연평균 정체 횡수를 바탕으로 한 정체량의 이산값을 목표변수로 설정하여 도로의 정체정도를 예측하는 시스템을 구축했다.

<Table 4>의 Observed Traffic Counts의 데이터 중 정체가 빈번히 일어날만한 지역을 선별하였다. 선별된 지역은 실리콘밸리의 인접 지역이며, 교통의 요지인 Santa Clara, San Francisco, Napa, Solano, San Benito, Sonoma, San Bruno, Yolo이며 도로의 조건은 시내 및 시외 주요 도로와 그 도로끼리 연결되는 주요 교차로이다. 기타 작은 소로나 1차선 도로는 예측시스템의 학습

대상에서 제외했다.

3.4 선정된 데이터에 대한 학습

선정된 데이터에 대해서 신경망을 통한 학습은 저자들이 자체적으로 개발한 MATLAB 기반의 프로그램으로 진행했다. 본 프로그램은 역전파 알고리즘을 바탕으로 작성된 신경망을 운용하기 위한 것으로 향후의 범용성을 위하여 EXCEL로 작성된 데이터를 입력받게 되어 있으며 이를 이용하여 특정 경우에 대한 패턴을 학습하고 반복된 학습의 결과로 수렴된 가중치가 구해진다.

패턴을 구성하는 각각의 입력값은 도로별 이동차량(유입 및 유출차량대수), 교차로의 수, 도로의 길이, 도로의 특성(규모 및 분류)이며 목표값은 도로의 연평균 정체량을 구간별로 나눈 값으로, 연평균 정체량이 없는 경우 0(없음), 1~49의 경우 0.25(적음), 50~150의 경우 0.5(중간), 151~200인 경우 0.75(많음), 201이상인 경우 1(매우 많음)로 설정한 값을 적용하였다.

이 같은 내용을 종합하여 조사된 정보와 학습에 사용될 각각의 패턴들(입력값과 목표값)을 정리하면 <Table 5>와 같이

Table 4. San Francisco Bay Area State Highway System 1990/2007 Observed Traffic Counts(in part)

Fiscal Year	County													
	Alameda	Contra Cost	Lake*	Marin	Napa	San* Benito	Santa Clara	Santa* Cruz	San francisco	Sna* Joaquin	San Mateo	Solano	Sonoma	Yolo*
1990	519	33	0	105	555	0	294	0	43	52	218	2,288	1,064	0
1991	1,242	599	0	0	6	0	589	4	32	183	499	2,507	545	0
1992	602	32	0	191	0	0	804	487	408	207	305	3,245	789	0
1993	708	18	0	112	546	0	368	22	329	297	286	2,378	1,175	0
1994	1,093	698	0	0	128	0	458	0	25	423	551	2,030	561	0
1995	411	49	0	240	174	0	1,329	579	90	701	658	4,453	811	0
1996	655	228	0	125	711	0	312	0	104	612	487	2,097	921	0
1997	1,216	782	0	0	150	0	538	0	6	0	724	1,807	436	0
1998	548	297	0	117	260	0	1,363	0	188	171	1,251	435	682	0

Table 5. Training patterns(in part)

도로번호	교차로수	길이(Km)	유입차량	유출차량	특성	차선	정체량	차선당유입	차선당유출	길이당 교차로수	특성	정체도
9	9	18.42	270	2850	0.9	1	74.00	270.00	2850.00	0.49	0.90	0.50
17	11	22.44	5700	14200	0.9	2	181.00	2850.00	7100.00	0.49	0.90	0.75
25	2	4.12	1800	1800	0.9	1	28.00	1800.00	1800.00	0.49	0.90	0.25
35	6	27.55	180	140	0.9	1	0.00	180.00	140.00	0.22	0.90	0.00
82	39	42.43	4650	2500	0.9	3	0.00	1550.00	833.33	0.92	0.90	0.00
85	23	23.87	4150	4650	0.9	3	157.00	1383.33	1550.00	0.96	0.90	0.75
87	10	14.84	7800	6800	0.9	3	11.00	2600.00	2266.67	0.67	0.90	0.25
101	38	89.38	4800	13400	0.6	4	407.00	1200.00	3350.00	0.43	0.60	1.00
130	9	36.21	1350	20	0.6	2	71.00	675.00	10.00	0.25	0.60	0.50
152	14	56.57	1950	3350	0.6	1	238.00	1950.00	3350.00	0.25	0.60	1.00
156	2	0.97	1150	1150	0.6	1	14.00	1150.00	1150.00	2.06	0.60	0.25
237	16	17.83	3500	5500	0.3	2	0.00	1750.00	2750.00	0.90	0.30	0.00
280	22	33.19	11700	9500	0.3	4	0.00	2925.00	2375.00	0.66	0.30	0.00

정리(53개의 도로 중 13개의 도로 만을 나타내었다) 될 수 있다. <Table 5>의 좌측의 표는 해당 도로에 대한 정보를 정리한 표이며, 우측의 표는 실제 학습패턴을 정리한 표이다.

3.5 학습을 위한 네트워크

본 논문의 예측모델은 51개의 학습 패턴을 역전파 알고리즘을 이용한 다층신경망을 통해 학습한다. 사용되는 신경망은 1개의 은닉층을 갖는 네트워크이며 히든 레이어는 10개의 노드를 갖는다. 최대 반복 학습 회수는 50,000회, 입력 패턴은 4개의 벡터 값(유입량, 유출량, 단위길이 당 교차로 수, 도로의 특성)으로 구성된다. 출력 값은 정체량의 5단계 수준(없음, 적음, 중간, 많음, 매우 많음)이다. 자세한 내용을 <Table 6>에 정리하였다.

Table 6. Variables and parameters for training

Variables				
# Input Variables	4			
Input variables	차선당 유입량	차선당 유출량	단위 길이당 교차로수	특성
Output variable	정체도			
Parameters				
# Hidden layers	1			
# Nodes in HiddenLayer-1	10			
# Epochs	5000			
Step size for gradient	0.5			
Error tolerance	0.01			

Table 7 Inter-layer connection weights

Hidden Layer #1	Input Layer				특성
	차선당 유입량	차선당 유출량	단위 길이당 교차로수		
Node #1	-29.67	22.35	-12.85	-4.91	
Node #2	-29.07	19.21	8.50	-2.92	
Node #3	11.08	6.35	22.09	-27.54	
Node #4	-0.47	-0.20	-0.38	-1.05	
Node #5	-6.17	-7.39	3.14	-8.93	
Node #6	17.78	-19.01	19.36	4.75	
Node #7	-15.38	1.15	-0.22	3.72	
Node #8	32.35	-36.51	6.76	18.02	
Node #9	-0.25	21.76	-3.02	-24.51	
Node #10	-8.33	0.08	1.37	-7.78	

Output	Node #1	Node #2	Node #3	Node #4	Node #5
Output Node	25.30	-19.90	-33.30	0.43	4.32
Output	Node #6	Node #7	Node #8	Node #9	Node #10
Output Node	18.24	8.44	-25.79	17.88	-9.46

3.6 학습에 대한 결과

반복된 학습의 결과로 각각의 내부 층의 변수간의 결합 가중치가 생성된다. 이제까지의 학습에 의해 도출된 가중치를 2개의 표로 나누어 나타낸 결과는 <Table 7>과 같다. <Table 7>의 첫 번째 표는 입력층과 은닉층의 결합 가중치를, 두 번째 표는 은닉층과 출력층의 결합 가중치를 나타낸 표이다.

3.7 테스트 데이터를 이용한 학습결과의 검증

위에서 얻어진 학습의 결과는 스스로 하나의 네트워크를 형성한다. 이 네트워크를 이용하면 미지의 데이터에 대해 목표 변수를 예측하거나 데이터의 동질성을 파악 할 수 있다. 이러한 예측 모델의 신뢰성을 검증하는데 쓰이는 데이터는 총 36개의 패턴으로 이루어져 있으며, 학습에 사용된 것을 제외한 Santa Clara, San Francisco, Napa, Solano, San Benito, Sonoma, San Bruno, Yolo 지역의 데이터를 이용하였다.

학습 패턴에 대한 가중치를 통해 테스트 데이터를 검증할 경우 RMS에러는 약 0.0973, 오분류율은 약 12.89% 정도로 계산되었다. 이는 기존에 주어진 캘리포니아 주에서 수집된 차량검지 시스템의 데이터를 이용한 예측 시스템 개발에 있어 87.11%의 예측성공률을 얻을 수 있음을 의미한다.

이와 같은 결과를 도로의 분류(일반도로 및 외곽도로, 주도로, 고속도로)별로 분석해보면 일반도로 및 외곽도로의 경우 100%의 예측성공률을 보였고 주도로는 80%, 고속도로 86.7%의 예측성공률을 보였다. 이 결과를 정리하면 <Table 8>과 같다.

Table 8. Total result of congestion prediction module

학습 결과	
RMS 오차	0.0261
학습률	94.3%
검증 결과	
RMS 오차	0.0973
예측성공률	87.11%
도로유형별 예측성공률	
일반도로 및 외곽도로	100%
주도로	80%
고속도로	86.7%

결과로 보았을 때 이러한 예측성공률은 양호한 수준이라 판단할 수 있다. 자료의 한정성 때문에 사용한 데이터가 많지 않은 상태에서도 이러한 결과를 보인 것으로, 더 다양한 패턴의 데이터가 확보된다면 보다 나은 검증 시스템을 만들 수 있을 것이다.

4. 토 의

본 연구에서는 일반적인 도로를 교통량에 영향을 미치는 변수

를 이용하여 모델링하고 교통량과 도로특성에 따른 특정위치의 도로의 정체정도를 예측하였다. 이 모델의 학습을 위하여 미국의 캘리포니아 주 지역의 감시 카메라 및 감지센서를 이용하여 측정된 교통 데이터를 이용해 의미 있는 데이터를 확보했다. 그리고 확보된 데이터의 학습을 통하여 각 지점의 정체정도를 분석하고 예측하는 연구를 진행하였다. 그 결과 본 모델은 주어진 데이터에 대해 약 87%의 적중률을 보이고 있다.

본 논문에서는 차량검지시스템에 의해 기 측정된 도로의 통행량 정보와 도로의 특성(길이, 교차로수 등), 정체발생 횟수에 대한 데이터를 기반으로 하여 역전파 알고리즘을 이용한 다층신경망을 통해 학습을 실시하고 이를 기초로 하여 실시간으로 주어지는 데이터를 이용하여 특정지역의 교통 정체 정보를 예측하는 과정을 체계적으로 서술하였다. 또한 학습의 결과와 실제 측정 데이터의 오차를 분석하였다.

선행연구와 차별화된 본 연구만의 특징은 크게 두 개로 나눌 수 있다. 먼저 도로의 일반적인 특징을 모델링하고 정체에 영향을 주는 변수를 분석하였다는 점이다. 그리고 정체예측시스템을 구축함에 있어 실험에서 얻은 데이터가 아닌 도로의 실제 특성과 실제 통행량을 반영한 데이터를 학습해 정체정도를 예측하였다는 점이다. 이렇게 구축된 정체 예측 시스템은 도로상황의 환경 변화(차선확대, 교차로수정) 등 새로운 상황에 적용 가능한 예측 모델이라는 점에서 다른 방법론들 보다 강점을 지닌다. 또한 정체에 대한 정보가 없거나 건설 예정중인 도로 또는 가상의 도로에 대한 정보를 이용하여 정체수준을 예측할 수 있으며 이후 출현하는 데이터에 대해서 지속적인 학습을 통해 가중치를 보정하여 새로운 도로에 대한 정보를 반영한 유연성을 부여 할 수 있는 모형을 구축할 수 있다. 구축한 모델과 선행연구의 방법들에 대한 비교는 <Table 9>과 같다.

5. 향후과제

본 연구는 미국의 캘리포니아 주 지역의 데이터를 토대로 진행됐다. 이는 국내의 데이터가 아직 공개되지 않고 있기 때문이다. 국내에서도 이런 교통량 데이터나 지적 데이터가 충분

히 조사·축적되고 있다. 미국의 사례와 같이 국내에서도 이러한 데이터들을 공개하고, 이러한 데이터를 통해 이러한 교통정체 예측 시스템을 만들 경우 국내에서도 사용이 가능한 시스템을 만들 수 있을 것이다. 또한 연 평균값을 공개하는 미국과는 달리, 시간대 별로 축적되어진 데이터를 공개한다면 시간대 별로 정체 예측을 수행하는 시스템의 구축도 가능하다. 차후에 데이터의 획득이 가능해지면 국내 실정에 맞게 조정된 시스템의 구축이 반드시 필요하다.

또한 이와 더불어 본 연구에서 자료 부족으로 쓰이지 못한 입력변수를 더 반영해 예측 시스템을 구축해야 할 것이다. 이외에 고려될 수 있는 입력변수는 도로의 경사 및 굽은 정도, 날씨, 사고의 유무, 차선의 수, 측정된 시간(평일, 출퇴근시간, 주말, 공휴일)등이며 이를 추가적으로 고려하여 예측 시스템을 구축하고, 더불어 각각의 한 지점이 아닌 다음 지점간의 상호작용을 고려하여 전체적 네트워크의 단위로 학습하여 예측을 수행하는 연구를 진행한다면 더욱 의미 있고 정확한 예측시스템을 구축할 수 있을 것이다.

만약 실시간으로 주어지는 데이터의 획득이 가능하다면 본 논문에서 제시된 방법론을 이용해 획득된 실시간 데이터에 대해 학습을 실시하고 예측 시스템을 구축할 수 있다. 이렇게 구축된 시스템을 통해 새로 제공되는 실시간 데이터에 대해 도로별 정체정도 또는 소요시간에 대한 예측을 수행할 수 있다. 예측된 정보에 맞추어 최적 이동시간, 이동수단, 및 이동경로에 대한 선택권을 부여한다면, 경로상의 시간 절약 및 잉여 도로 공간을 활용할 수 있을 것이다. 또한 실시간 교통정보 및 우회경로정보를 제공함으로써 교통체증의 감소에 도움을 줄 수 있는 ITS구축에 기여할 것이다.

참고문헌

- Chen, M. and Chien, S. (2001), Dynamic freeway travel time prediction using probe vehicle data: Link-based vs. Path-based, *Transportation Research Board 80th Annual Meeting*, 120-129.
- Cheu, R. L., Xie, C., and Lee, D. (2002), Probe Vehicle Population and Sample

Table 9. Comparison with other algorithms

방법	장점	단점
Kwon <i>et al.</i> (2000)	<ul style="list-style-type: none"> 사용자가 지정한 임의의 경로에 대하여 주행 시간 예측가능 	<ul style="list-style-type: none"> 주행 시간 예측의 정확도가 다른 기법에 비해 저하됨 인구의 변화 또는 유입, 유출량의 변화와 같은 환경적 변화에 대응할 수 없음
Rice and Van Zwet (2004)	<ul style="list-style-type: none"> 유사 주행 시간 패턴을 통한 장거리 주행 시 주행 시간 예측 지원 	<ul style="list-style-type: none"> 과거 데이터의 무결성이 보장되어야 하므로 도로의 상태가 변화될 때 데이터를 다시 선별하는 작업 필요
Chen and Chien (2001)	<ul style="list-style-type: none"> 미리 정해진 형태의 도로 구간에 대한 주행 시간 도로의 유형에 따른 예측 가능 	<ul style="list-style-type: none"> 미리 정해진 형태의 주요 경로 형태에 대해서만 예측 가능하여 유연성이 떨어짐
본 연구의 신경망 모델	<ul style="list-style-type: none"> 경로의 특성을 반영한 예측 가능 새로운 도로 패턴에 대해 재학습을 통해 적응가능 	<ul style="list-style-type: none"> 많은 학습데이터 필요 학습 시 상위 방법에 비해 많은 시간 소요

Size for Arterial Speed Estimation, *Computer-Aided Civil and Infrastructure Engineering*, 17(1), 53-60.

Faouzi, N. E. (2006), Bayesian and Evidential Approaches for Traffic Data Fusion: Methodological Issues and Case Study, *Transportation Research Board 85th Annual Meeting*, Paper 06-1510 (Preprinted CDROM).

Kwon, J., Coifman, B., and Bickel, P. J. (2000), Day-to-day travel time trends and

travel time prediction from loop detector data, *Transportation Research Board 79th Annual Meeting*, 120-129.

Kim, H-S. (2007), Introduction plan of Archived VDS Data User Services, *Proceedings of the KOR-KST Conf. 55th*, 449-456.

Rice, J. and Van Zwet, E. (2004), A simple and effective method for predicting travel times on freeways, *Intelligent Transport Systems*, 5(3), 200-207.



신원식

경희대학교 산업공학과, 공학학사
경희대학교 산업경영공학과, 공학석사
한국로봇산업진흥원 재직 중
관심분야: 인공지능, 로보틱스



오세도

경희대학교 산업공학과, 공학학사
경희대학교 산업경영공학과, 공학석사
경희대학교 산업경영공학과, 공학박사과정
관심분야: 데이터마이닝, 인공지능, 진단
시스템



김영진

서울대학교 기계설계학과, 공학학사
서울대학교 기계설계학과, 공학석사
미국 University of California at Berkeley
기계공학과, 공학박사
미국 University of California at Berkeley 부설
Engineering Systems Research Institute 연구원
(Post-Doc)
서울대학교 부설 정밀기계설계 공동연구소
특별연구원
경희대학교 산업경영공학과 정교수
관심분야: 데이터마이닝, 인공지능, 진단
시스템