

의사결정나무와 신경망 모형 결합에 의한 운전자 우회결정요인 분석

Drivers Detour Decision Factor Analysis with Combined Method of Decision Tree and Neural Network Algorithm

강진웅	Kang, Jin Woong	정회원 · 명지대학교 공과대학 교통공학과 석사과정 (E-mail : gd173@hanmail.net)
금기정	Kum, Ki Jung	정회원 · 명지대학교 공과대학 교통공학과 교수 (E-mail :kjkum@mju.ac.kr)
손승녀	Son, Seung Neo	정회원 · ITS Korea (E-mail :son3003@nate.com)

ABSTRACT

This study's purpose is to analyse factors of determination about detouring for making a standard model in regard of unfavorableness and uncertainty when unspecified individual recipients make a decision at the time of course detour. In order to achieve this, we surveyed SP investigation whether making a detour or not for drivers as a target who take a high way and National highway. Based on this result, we analysed detour determination factors of drivers, establishing a combination model of Decision Tree and Neural Network model. The result demonstrates the effected factors on drivers' detour determination are in ordering of the recognition of alternative routes, reliable and frequency of using traffic information, frequency of transition routes and age. Moreover, from the outcome in comparison with an existing model and prediction through undistributed data, the rate of combination model 8.7% illustrates the most predictable way in contrast with logit model 12.8%, and Individual Model of Decision Tree 13.8% which are existed. This reveals that the analysis of drivers' detour determination factors is valid to apply. Hence, overall study considers as a practical foundation to make effective detour strategies for increasing the utility of route networking and dispersion in the volume of traffic from now on.

KEYWORDS

drivers detour decision factor, datamining, decision tree algorithm, neural network algorithm

요지

본 연구는 불특정 다수의 도로이용자들이 경로우회 시 갖는 의사결정과정속에 내포된 비선형성과 불확실성을 고려한 정도 있는 모형구축으로 주요 우회결정요인을 분석하는 것이 주요 목적이다. 이를 위하여 고속도로 및 국도를 이용하는 운전자를 대상으로 우회여부에 관련된 SP조사를 실시하였고, 조사결과에 대하여 의사결정나무와 신경망이론의 결합된 모형을 구축하여 운전자 우회결정요인을 분석하였다. 분석결과 운전자 우회여부결정에 영향을 미치는 요인은 우회도로 인지여부, 교통정보 신뢰도 및 이용빈도, 경로전환빈도, 나이순으로 나타났다. 또한 오분류표를 통한 기존 모형과의 예측력의 비교결과 결합된 모형의 오분류율이 8.7%로 기존 모형인 로짓모형 12.8%, 의사결정나무 단독 모형 13.8%와 비교했을 때 가장 예측력이 높은 것으로 나타나 운전자 우회결정요인 분석에 관한 모형의 적용 타당성을 확인할 수 있었다. 본 연구의 결과는 향후 교통량 분산효과와 도로망 효율 증대를 위한 효과적인 우회관리전략 수립 시 기초 자료로 활용가능하리라 사료된다.

핵심용어

우회결정요인, 데이터마이닝, 의사결정나무모형, 신경망모형

1. 서론

1.1. 연구의 배경 및 목적

현재 우리나라는 친환경 녹색성장 기조에 맞춰 지정 체 혼잡구간 개선에 주력함으로써 도로신설보다 기존

노선을 효율적으로 이용하는 쪽으로 투자패턴을 전환하는 추세에 있다. 또한 첨단교통정보체계를 구축하여 도로의 수요관리시행을 통한 운전자 경로전환유도로 교통량을 분산함으로써 지정체구역의 혼잡완화 및 지체감소

의 효과를 높이고 있다. 도로는 불특정 다수에게 제공되는 공공시설이며 이를 이용하는 운전자의 우회여부결정은 개개인 각자의 주관적인 판단에 의해 이루어진다. 지금까지 많은 운전자의 우회행태에 관련된 다양한 연구가 수행되어 왔으나, 운전자의 의사결정과정 속에 포함된 비선형성과 불확실성의 해석은 지속적으로 연구되어야 할 분야이다.

따라서 본 연구는 기존 우회결정요인을 추출하는 모형들의 단점을 보완하여 운전자가 우회결정시 갖는 의사결정과정을 분석하기 위해 SP조사를 실시하였다.

이에따라 그 데이터 속에 존재하는 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화함으로써 유용한 지식을 추출하는 데이터 마이닝 기법의 의사결정나무와 신경망 이론의 결합모형 구축 및 모형의 적용타당성을 검토하고, 그에 따른 운전자 우회결정 요인을 분석하는데 주목적을 두었다.

1.2. 연구의 내용 및 수행절차

본 연구에서는 운전자가 우회결정시에 영향을 주는 요인들을 선정하기 위하여 기존의 연구 및 문헌 등을 통해 대상요소들을 검토하여 각 특성별 변수들을 선정하였다.

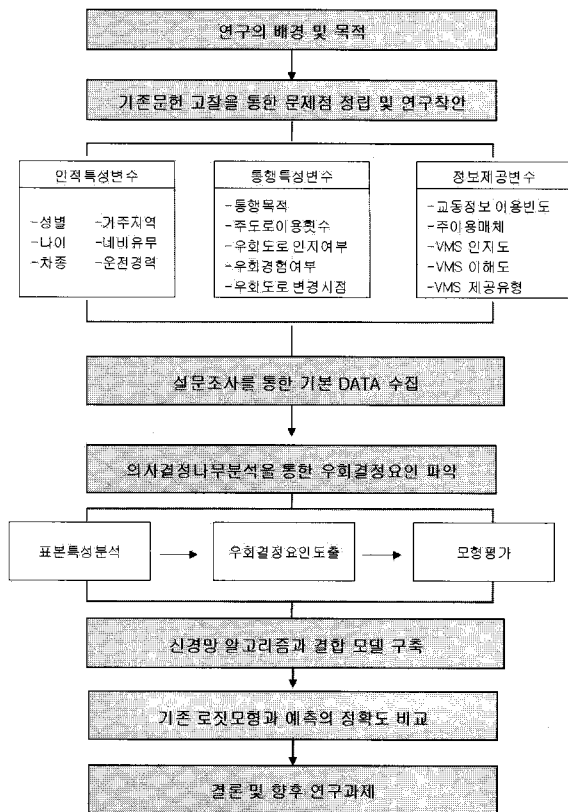


그림 1. 연구의 수행과정

선정된 변수들을 바탕으로 SP조사 기법을 활용하여 우회결정에 영향을 주는 요인을 개인특성변수, 통행특성변수, 정보제공변수로 구분하여 설문조사 data를 구축하였고, 데이터마이닝 기법인 의사결정나무 모형을 통하여 운전자가 우회결정시에 주요 영향을 미치는 요인들을 추출하여 그 요인들간의 교호관계에 대해서 분석하였다.

또한 의사결정나무 모형을 통해 분석된 요인들을 신경망모형에 입력변수로 사용하여 의사결정나무모형과 신경망 결합에 의한 운전자 우회결정 모형을 구축하였다. 모형의 정확도 및 타당성을 검토하기 위해 오분류표를 이용하여 타 모형간의 정확도를 비교하였다. 이에 본 연구의 수행절차는 그림 1과 같다.

2. 선행연구 및 이론 고찰

기존 연구고찰에서는 선행된 우회결정요인 및 데이터마이닝과 관련된 연구를 개략적으로 살펴보고, 데이터마이닝 기법인 의사결정나무와 신경망 모형에 대한 이론적 고찰을 통한 본 연구의 적용성을 검토하였다.

2.1. 데이터마이닝 기법에 대한 기존 연구

데이터마이닝의 활용분야는 매우 다양하며, 현재에도 그 활용분야가 점차 늘어나고 있다. 그 중에서 다양한 의사결정과 향후 대처에 대한 문제에서 많이 활용되고 있다.

특히 모형의 상호 장단점을 보완하기 위한 결합모형에 관련된 연구들이 지속되고 있다.

이진천(2000)은 입력데이터의 패턴 분석 모형을 통해 예측모형을 선정하는 결합모형을 구축하였다. 판별모형이 입력되는 데이터를 분석하여 더 좋은 적응률을 보일 수 있는 모형을 선정하였다. 선정된 각 모형이 예측을 수행하는 구조의 모형으로서, 결합모형의 타당한 평가를 위해 단일모형을 별도로 구축하여 예측결과를 비교 평가한 결과 단일모형보다 정확도가 8.6% 향상된 것으로 분석했다.

김연지(2003)은 데이터마이닝을 이용하여 개인의 신용을 가장 잘 예측할 수 있는 모형을 로지스틱 회귀분석, 의사결정나무, 신경망 분석으로 디자인하고, 신경망 모형의 해석상의 어려움을 극복하기 위하여 의사결정나무로 재해석하여 기존의 의사결정나무 모형과의 유사성을 찾고자 하였다.

김중호 외(2010)은 유전자 알고리즘과 신경망 이론의 결합에 의한 신호교차로 위험도 예측모형을 개발하고, 예측값과 실측값의 비교 분석을 실시하였고 그 결과 결합모형이 기존모형에 비해 우수한 연구결과로 나타났다.

이처럼 현재 많은 연구에서 각각의 알고리즘의 단점을 보완하고자 단일기법보다는 변형된 알고리즘의 결합을 통하여 예측의 정확도를 향상시키는 연구가 이루어지고 있다.

2.2. 운전자 우회요인분석에 대한 기존 연구

김일평(2008)은 우회결정요인 분석을 통한 교통정보 제공방안에 관한 연구에서 설문조사를 통하여 우회행태에 대한 기초분석을 실시하였으며, 선호도 분석을 통하여 우회결정요인을 도출하고, 결과를 바탕으로 정보제공매체의 특성을 고려한 정보제공 전략 제안을 수립하였다.

장정아(2005)는 FTMS와 RTMS의 시행효과를 SP기법을 이용하여 분석하였고, 우회경로에 대한 운전자의 인식정도에 따라 교통정보 제공에 따른 우회율을 분석하였다.

서임기(2006)는 교통정보시스템 구축 전 지역, 구축 지역, 구축 활성화 지역으로 분류하여 교통정보의 신뢰도와 만족도에 의한 경로변경에 대한 분석을 실시하였고, 개별형태 모형인 로짓모형으로 교통정보가 경로변경에 영향을 미치는 정도를 규명하였다.

최기주 등(2004)은 SP조사를 바탕으로 로짓모형을 정립하여 우회율에 영향을 미치는 변수들을 도출하였고, VMS에서 제공되는 교통정보 메시지에 따라서 우회율이 크게 달라지는 것과 VMS 순응도에 대해 분석하였다.

운전자 우회요인분석에 대한 기존 연구를 검토한 결과 우회여건, 운전경력, 정보제공 메시지, 우회도로 인식여부 등 다양한 우회요인들이 분석되었고, 그에 따른 다양한 교통정보제공 방안에 관하여 연구가 진행되었다.

하지만 대부분의 논문에서 우회결정모형 정립 시 로짓모형 등 단독모형을 통한 모형 구성으로 분석방법의 한계성이 나타났다. 따라서 본 연구에서는 많은 기존문헌에서 우회결정모형을 정립하는데 사용되었던 로짓모형 등 단독모형의 단점을 보완하고자 데이터마이닝의 대표적 기법인 의사결정나무모형과 신경망모형의 결합모형 구축으로 결정 요인들간의 교호관계 파악과 모형의 정확도 향상을 위하여 신뢰성이 높은 주요 우회결정요인들을 선정하고자 한다.

2.3. 의사결정나무와 신경망모형의 이론적 고찰

데이터마이닝이란 대용량의 데이터로부터 그 속에 존재하는 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화함으로써 유용한 지식을 추출하는 일련의 과정을 일컫는다.

즉 과거에 알지 못했던 새로운 데이터의 규칙을 발견하고, 미래에 실행 가능한 정보를 추출해 내 의사결정에 이용하는 과정을 의미한다.

데이터 마이닝의 기본 개념은 통계학에 그 바탕을 두고 있으며, 데이터마이닝 중 예측모형으로 축적된 자료를 활용하여 미래의 현상이나 나타날 행동 등에 대해 예상하는 대표적인 모형으로 의사결정나무모형과 신경망모형 등의 방법론이 있다.

그 중에서도 의사결정나무모형은 적용의 용이성, 해석의 편리성 등을 갖추어 데이터마이닝의 주요 방법론으로 사용되고 있다.

의사결정나무모형은 의사결정규칙을 도표화하여 관심대상이 되는 집단을 몇 개의 소집단으로 분류하거나 예측을 수행하는 분석방법으로써 존재하는 자료로부터 사전에 알려지지 않는 특징 및 성향 등의 정보들을 추출하는 것으로, 인공지능분야의 기계학습이론에 기초를 두고 있다.

의사결정나무모형의 장점은 두 개 이상의 변수가 결합하여 종속변수에 어떻게 영향을 주는지를 쉽게 알 수 있다는 것이다. 이 모형은 선형성(Linearity), 정규성(normality), 등분산성(equal variance)등의 가정을 필요로 하지 않는 비모수적(non-parametric) 방법으로, 순서형 또는 연속형 변수는 단지 순위(rank)를 통해서만 분석에 영향을 주기 때문에 이상치(outlier)는 민감하지 않다는 장점을 가지고 있다(진현정, 2010).

기존 우회결정모형에 주로 적용되었던 회귀분석이나 로짓분석과 같은 모수적(parametric)모형에서 교호효과(interaction effects)를 모형에 포함시켜 분석하는 것이 어렵다. 그러나 이 모형은 이러한 효과를 효율적으로 추출해낸다는 특징이 있다. 또한 분석의 과정 및 결과가 나무구조에 의해 표현되기 때문에, 분류 또는 예측을 목적으로 하는 다른 방법들에 비해 연구자가 분석과정을 쉽게 이해하고 설명할 수 있다는 장점을 지니고 있다(진현정, 2010). 본 연구에서는 운전자들의 우회 시 주 영향을 미치는 변수들을 추출하는 것을 목표로 모형을 사용하였다.

신경망모형은 인간의 두뇌구조를 모방한 지도학습 방

법으로써, 여러 개의 뉴런들이 상호 연결되어 입력값에 대한 최적의 출력값을 예측하는 모형이다.

복잡하고 비선형적이며 관계성을 갖는 다변량을 분석할 수 있는 방법으로써 어떤 기존의 통계적 모형보다도 정확한 예측을 제공한다(강병서, 2010).

기본적으로 신경망 모형은 인간의 신경학적 뉴런과 작동 원리가 비슷한 노드(node)와 층(layer)로 구성된다. 노드는 입력물, 가중치, 합계 활동 함수, 출력물 가치 등으로 구성 되는데 가중치에 의해서 계산된 입력변수들은 활동함수(f)를 통해서 새로운 값으로 변환되어 출력된다. 출력된 계산물은 다음 노드에 입력물이 된다.

층은 입력층(각 입력변수에 대응되는 노드로 구성되고, 노드의 수는 입력변수의 개수와 같음), 은닉층(입력층으로부터 전달되는 변수값들의 선형결합을 비선형함수로 처리하여 출력층 또는 다른 은닉층에 전달), 출력층(목표변수에 대응되는 노드로써, 분류모형에서는 그룹의 수 만큼 출력노드가 생성)로 구성되며, 학습 페리다임에 근거한 인공신경망은 입력물 데이터를 기초로 가중치를 통해서 의사결정을 지원한다.

신경망모형은 장점과 단점을 갖고 있다. 복잡하고 다양한 자료를 용이하게 해를 찾으며, 질적변수와 양적변수에 관련없이 모든 변수가 분석 대상이 될 수 있는 장점이 있다. 또한 입력 변수들 간 비선형 조합이 가능하여 예측력이 우수하다. 반면 결과에 대한 분류와 예측 결과만을 제공하기 때문에 결과 생성의 원인과 이유를 설명하기가 곤란하다.

따라서 본 연구에서는 분류와 예측에 효과적이며, 독립변인들간의 교호작용을 명확하고 쉽게 이해가 가능한 의사결정나무의 모형의 결과를 신경망이론의 입력변수로 사용하여 결과생성의 원인과 이유를 설명하고, 모형의 예측률을 높일 수 있는 모형을 구축하였다.

3. 평가모형 정립

3.1. 기존 모형의 한계성 보완

기존 로짓모형은 선형성을 가정함으로써 생기는 회귀모형의 단순성을 가지고 있어, 운전자의 우회결정에 영향을 미치는 변수들간의 복잡한 비선형성을 예측함에 있어 유용성 측면에서 한계를 지니고 있다. 또한 각 입력변수의 영향이 다른 변수에 종속 되어있지 않는다고 가정함으로써, 일부 변수들 간의 유용한 교호작용을 탐색하는 것이 어렵다.

따라서, 본 연구에서는 이러한 기존 로짓모형의 한계

성을 보완하고자 의사결정나무 분석을 통한 주요 우회요인간의 교호작용 및 비선형성의 용이한 해석과 신경망모형의 결합을 통한 높은 예측력 확보로 설명력있는 모형을 구축하였다.

3.2. SP모형 설계

본 연구에서는 가상의 VMS 정보제공(지체 및 돌발)상황에서의 우회여부를 설문조사 함으로써, 운전자의 우회결정모형을 구축하는 방법론을 채택하였다. SP모형을 위한 주요 설계항목은 다음과 같다.

- 조사대상자 : 고속도로 및 국도 이용자
- 선택상황 : 우회여부선택
- 속성변수 : 개인특성변수, 통행특성변수, 정보제공변수

표 1. 속성변수의 유형

변수	척도	변수 값	
개인 특성 변수	성별	이산형	남, 여
	나이	연속형	20대, 30대, 40대, 50대 이상
	거주지역	범주형	수도권, 광역시, 기타
	차종	범주형	트럭, 승용차(중·대형), 경차, 승합차
	운전경력	연속형	6개월 미만, 6개월~1년 미만, 1년이상~5년 미만, 5년 이상
통행 특성 변수	네비게이션 사용유무	이산형	유, 무
	통행목적	범주형	통근(출퇴근, 등하교), 업무, 여행(쇼핑), 기타
	주도로 이용횟수	연속형	주 1~3회, 4~5회, 6회 이상
	우회도로 인지여부	이산형	알고 있다, 모른다
정보 제공 변수 (교통 상황)	우회경험 여부	범주형	없음, 가끔 이용, 수시 이용
	우회경로 변경시점	이산형	출발전, 운전중
	교통정보 이용빈도	범주형	매번 이용, 가끔 이용, 보통, 거의 이용 안함, 전혀 이용 안함
	주이용매체	범주형	VMS, 라디오, 인터넷, ARS, 이용 안함
	교통정보 신뢰도	범주형	신뢰 보통, 불신
	VMS 인지도	이산형	인지O, 인지X
VMS 제공유형	VMS 이해도	이산형	쉽다, 어렵다
		범주형	1. 지체정보(지체유무) 2. 지체정보(지체유무+시간) 3. 돌발정보(사고발생) 4. 돌발정보(사고발생+지체시간)

3.3. 의사결정나무와 신경망 결합모델 생성

의사결정나무와 신경망 결합모델은 의사결정나무모형의 장점인 결과 해석의 용이성과 교호작용 또는 비선형성을 통해 규칙이 생성된 변수를 선정하고 예측력에 큰 장점을 갖는 신경망 모형에 결합시킴으로써, 두 모형간의 장단점의 상호보완을 통해 보다 정도 높은 예측결과를 얻고자 한다.

3.3.1. 의사결정나무 모형정립을 통한 주요변수 선정

의사결정나무는 하나의 나무구조를 이루며, 마디(node)라고 불리는 구성요소들로 이루어져 있다. 그 마디 중 가장 상단에 놓여 있는 마디인 뿌리마디(root node) 또는 부모 마디(parent node)라고 불리는 마디부터, 자식마디라고 불리는 중간마디(internal node), 끝마디(terminal node)로 구성되어 있다.

의사결정나무분석에 사용되는 알고리즘의 선택은 분석의 기본 목적과 자료의 구조에 의해서 분석 방법이 구분된다. 주요 알고리즘인 CHIAID(chi-squared automatic interaction detection)은 목표변수(종속변수)가 질적 변수이거나 양적변수이며, 예측 변수는 질적 변수인 경우에 사용된다.

CHIAID에서는 분리기준으로 목표변수가 질적변수인 경우에 카이제곱 통계량을 사용하는데 카이제곱 통계량이 크며 이에 대한 확률(p)이 $\alpha=0.05$ 보다 작은 경우 부모마디는 자식 마디를 형성하게 된다.

하지만 언제나 CHIAID 알고리즘이 최적분리를 보장해 주는 것이 아니기 때문에 본 연구에서는 CHIAID의 수정된 분석 방법으로 예측변수에서 모든 가능한 분리를 고려하여 최적분리를 찾을 수 있도록 보완한 Exhaustive CHIAID 알고리즘을 바탕으로 모형을 구축하였다.

독립변인들간의 분리(splitting)와 병합(merging)기준은 .05수준이며, 정지규칙으로 최대트리깊이는 3, 본 연구의 샘플수를 고려하여 부모마디(parent node)의 사례수는 50, 자식마디(child node)의 사례수는 10으로 지정하였다.

3.3.2. 의사결정나무와 신경망 결합모델 구축

의사결정나무 모형을 통해 최고의 이익비율을 기준으로 각 마디를 형성하는데 사용된 주요 우회결정요인을 신경망모형의 입력변수로 설정하였다. 또한 모형의 과대적합(Over fitting)이 일어나면 표본 내에서는 뛰어난

적합도를 가져올 수 있지만 표본 밖의 데이터에서는 적합도가 떨어지는 문제를 발생시킬 수 있기 때문에 이를 확인하기 위하여 전체 데이터 중 80%를 모형을 수립하는 훈련데이터로, 나머지 20%를 검증데이터로 분리하여 사용하였다.

결합모형을 구축하는 신경망모형은 가장 많이 사용되는 다층(Multilayer Perception)신경망으로 입력변수를 가지고 입력층에서부터 출력층에 이르기까지 가중치를 반복적으로 부과하는 과정을 거쳐 모형을 만들어 분석을 하였다.

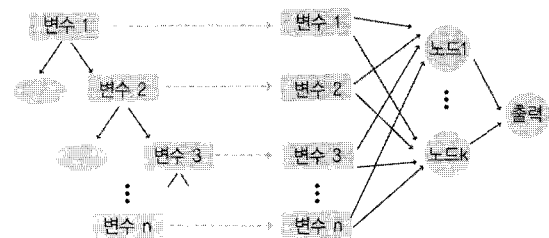
모형을 형성하는 알고리즘으로는 목적함수가 연결강도에 대하여 비선형 함수인 것의 최적화를 위해 가장 널리 이용되는 역전파 알고리즘을 사용하였다.

역전파 알고리즘은 1단계(주어진 연결강도를 이용하여 예측값을 계산), 2단계(실제 출력값과 예측값 사이의 오차를 계산), 3단계(오차를 은닉층과 입력층으로 역전파 시켜서 연결강도를 새로 조절하는 단계)를 거치면서 초기 연결강도를 이용하여 계속 반복 후 연결강도의 값이 일정하게 유지되면 반복을 멈추는 학습방법이다.

은닉층과 출력층에 제시되는 값을 변형시켜 출력하는데 사용하는 활성화 함수로는 다층 신경망에서 가장 많이 사용되며 그 유의성 또한 뛰어난 것으로 알려져 있는 시그모이드 함수를 사용하였다. 시그모이드 함수는 각 변수들을 [0,1]의 범위로 변환하여 처리하는 함수이다.

이를 바탕으로 본 연구에서 구성된 결합모델의 형성 과정은 표 2와 같다.

표 2. 의사결정나무와 신경망모형의 결합



의사결정나무 모형	신경망모형
<ul style="list-style-type: none"> - 분류와 예측에 효과적 - 이해하기 쉬운 규칙생성 - 변수들의 교호작용 파악 	<ul style="list-style-type: none"> - 높은 예측률 제공 - 결과에 대한 설명력 부족
의사결정나무와 신경망 결합모형	
<ul style="list-style-type: none"> - 의사결정나무로부터 생성되는 변수를 신경망모형의 입력변수로 적용 - 결정변수의 규칙생성 및 변수들의 교호작용 해석으로 모형의 설명력 강화 및 예측의 정확도 향상 	

표 2의 설명과 같이 결합모형의 구축으로 서로의 장 단점을 보완하여 기존 로짓모형과 달리 우회결정요인을 보다 구체적이고 정확하게 밝힐 수 있다.

즉, 의사결정나무모형을 통해 우회 의사결정과정 시 영향을 주는 의미있는 변수들의 파악과 그들의 인과관계를 설명하여, 신경망이론의 결과에 대한 해석의 난이성의 문제를 해결하고 모형의 정확도 또한 높이고자 한다.

4. 실증분석 및 결과

4.1. 자료설명

고속도로 및 국도 이용자를 대상으로 현장면접조사 및 E-mail 조사를 통한 인적 분포는 다음과 같다.

표 3. 개인특성 변수 유형 조사결과

구분	빈도	%	
성별	남	204	58.6
	여	144	41.4
나이	20대	72	20.7
	30대	138	39.7
	40대	93	26.7
	50대 이상	45	12.9
지역	서울	122	35.1
	수도권	127	36.5
	광역시	61	17.5
	중소도시	38	10.9
운전경력	6개월 미만	31	8.9
	6개월~1년	62	17.8
	1년~5년	131	37.6
	5년 이상	124	35.6
차종	트럭	14	4.0
	승용차(중, 대형)	236	67.8
	경차	68	19.5
	승합차	30	8.6
네비유무	유	306	87.9
	무	42	12.1

설문결과 성별의 경우 남성이 58.6%, 여성이 41.4%의 비율로 나타났으며, 나이대별로는 30~40대의 연령이 66.4%로 가장 많았다. 지역은 서울 및 수도권이 71.6%로 가장 많았고, 운전경력은 1년~5년 37.6%, 차종은 승용차가 67.8%로 가장 비중이 높았다. 네비게이션을 보유한 운전자들은 87.9%로 대다수가 네비게이션을 보유하고 있는 것으로 분석되었다.

표 4. 통행특성 및 정보제공 변수 유형별 조사결과

구분	빈도	%	구분	빈도	%		
통행목적	통근	195	56.0	변경결정시점	출발전	115	33.0
	업무	103	29.6		운전중	233	67.0
	쇼핑	43	12.4	주이용매체	VMS	169	48.6
	기타	7	2.0		라디오	71	20.4
주도로이용횟수	주1~3회	271	77.9		인터넷	99	28.4
	주4~5회	65	18.6		ARS	5	1.4
	주6 이상	12	3.4		이용안함	4	1.1
우회도로인지여부	인지	223	64.1		교통정보신뢰도	매우신뢰	116
	미인지	125	35.9	조금신뢰		102	29.3
경로전환빈도	매우 자주	51	14.7	보통		82	23.6
	약간 자주	107	30.7	조금불신		44	12.6
	보통	97	27.9	매우불신		4	1.1
	거의 안함	78	22.4	VMS시인성		좋음	255
전혀 안함	15	4.3	나쁨		93	26.7	
교통정보이용빈도	매번 이용	10	2.9	VMS이해도	쉽다	259	74.4
	가끔이용	148	42.5		어렵다	89	25.6
	보통	126	36.2	우회영향메세지	지체유무	10	2.9
	거의 이용 안함	57	16.4		지체유무+시간	143	41.1
전혀 이용 안함	7	2.0	사고발생		37	10.6	
			사고발생+시간		158	45.4	

운전자 통행특성을 분석한 결과 통행목적은 통근이 56.0%로 가장 높게 나타났으며, 주요 도로 운행횟수는 주당 1~3회 운행하는 운전자가 77.9%로 분석되었다. 우회도로를 인지하고 있는 운전자는 64.1%로 확인되었으며, 도로이용자의 48.6%가 VMS를 통해 교통정보를 제공받는 것으로 나타났다. 또한 운전자들의 교통정보 신뢰도는 62.6%로 비교적 높았으며, 사고발생에 따른 지체정보 메시지를 받았을 때 우회할 확률이 45.4%로 가장 영향을 주는 메세지 유형으로 확인되었다.

4.2. 의사결정나무 분석결과

설문조사를 바탕으로 의사결정나무를 구축한 결과는

그림 2와 같다. 분석결과를 살펴보면, 노드분할에 있어서 기준이 되는 Pearson 카이제곱 통계량에 따라 운전자 우회여부결정에 가장 영향을 미치는 요인은 우회도로 인지여부로 나타났고, 그 다음으로 교통정보 신뢰도 및 이용빈도, 경로전환빈도, 나이순으로 나타났다.

이에 분석결과를 구체적으로 살펴보면 전체 표본 집합에서 우회도로를 인지하고 있는 운전자 중 78.8%가 우회를 선택하는 것으로 파악되었다. 반면 우회도로를 인지하지 못한 운전자의 경우 82.8%가 우회를 선택하지 않는 것으로 나타나, 운전자 우회결정요인 중 우회도로 인지여부가 결정적인 변수인 것으로 파악되었다.

우회도로를 인지하고 있으며 교통정보를 매우 신뢰하는 운전자의 90.8%가 우회를 선택하는 것으로 나타났고, 이 경우 경로전환을 자주하는 운전자의 우회선택 확률은 97.3%로 분석되었다. 또한 우회도로를 인지하며 교통정보를 신뢰하는 수준이 높지 않거나 보통인 경우 78.1%가 우회를 선택하였고, 이 중 교통정보를 보통이상의 빈도로 이용하는 운전자의 우회선택확률이 85.6%

로 나타났다. 반면, 우회도로를 인지하고 있더라도 교통정보를 신뢰하지 않는 경우 우회하지 않을 확률이 100%로 분석되었다.

우회도로를 인지하지 못한 운전자들 중 교통정보를 불신하는 96.1%가 우회를 선택하지 않았고, 그 중 20대 운전자의 우회하지 않을 확률이 100%로 나타났다.

이상의 분석을 통해 우회도로인지 여부와 신뢰도는 매우 밀접한 관계를 갖고 있는 것을 확인했고, 특히 우회도로를 인지하며 교통정보를 신뢰하는 경우 우회확률이 매우 높고, 경로전환빈도와 교통정보 이용빈도가 높을수록 우회확률이 높은 것으로 분석된다.

반면 우회도로를 인지하지 못하는 운전자의 경우 우회를 하지 않을 확률이 높고, 특히 교통정보 신뢰도가 낮은 20대의 경우가 우회하지 않을 확률이 가장 높은 것으로 나타났다.

결과의 통계적 진단을 확인해보면, 의사결정나무모형의 적합도를 나타내는 분류 정확도는 86.2%로 나타났으며, 오분류가 나오게 될 위험도는 추정값이 0.138로

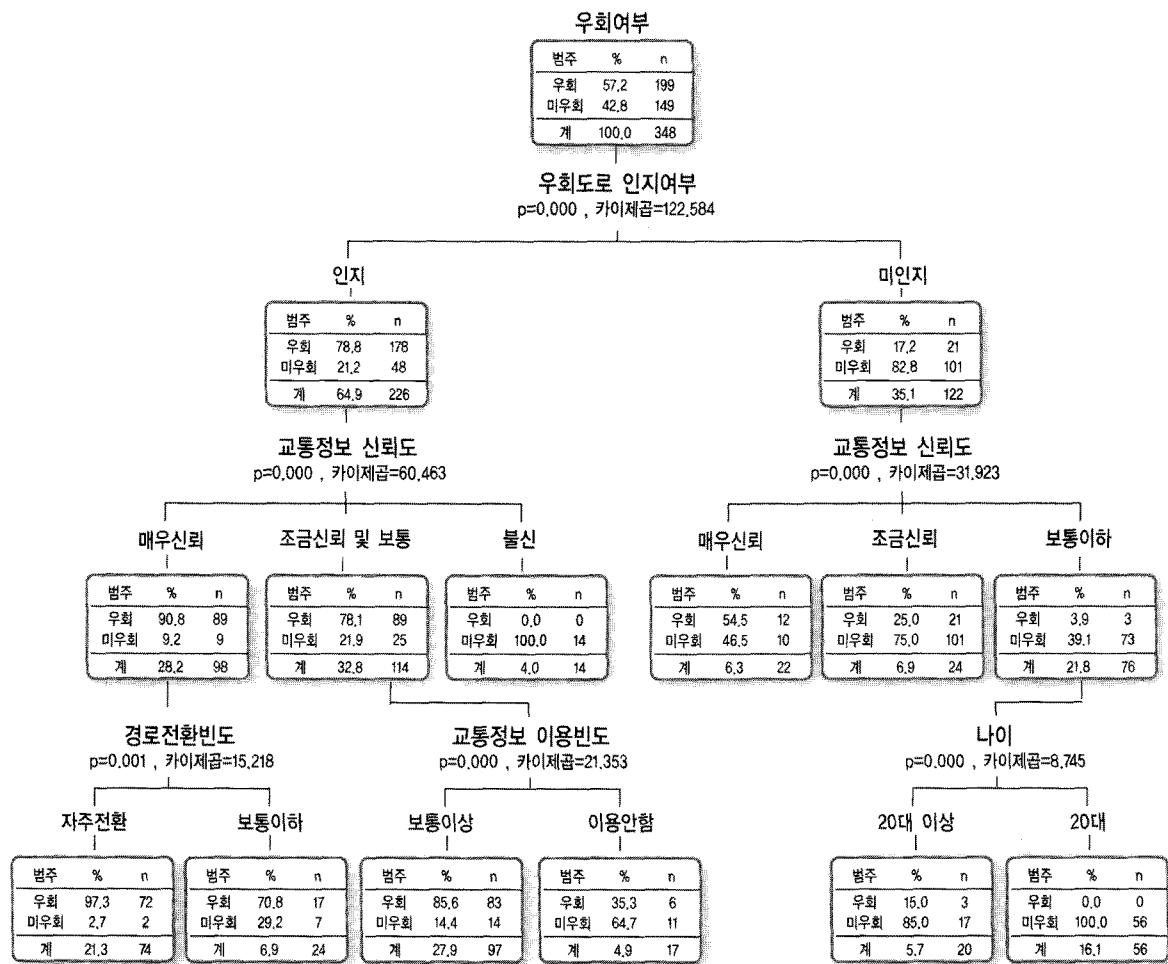


그림 2. 의사결정나무 분류 결과표

나와 적합한 모형으로 분석되었다.

4.3. 의사결정나무와 신경망결합모형 결과

의사결정나무 분석에 따른 분류결과로 나타난 운전자 우회여부결정 주요요인인 우회도로 인지여부, 교통정보 신뢰도, 교통정보 이용빈도, 경로전환빈도, 나이를 신경망 모형의 입력변수로 선정하여 다층신경망모형을 구축화한 모형의 결과는 다음과 같다.

그림 3은 모형의 훈련 후의 네트워크 다이어그램을 보여 주고 있다. 네트워크 다이어그램에서 시냅스 가중값(synaptic weight)은 주어진 레이어와 그 다음 레이어 사이에서의 관련성을 시각적으로 보여주는데, 레이어간 굵은 실선으로 연결된 것이 가중치가 높은 것으로 결과 해석을 한다(강병서, 2010).

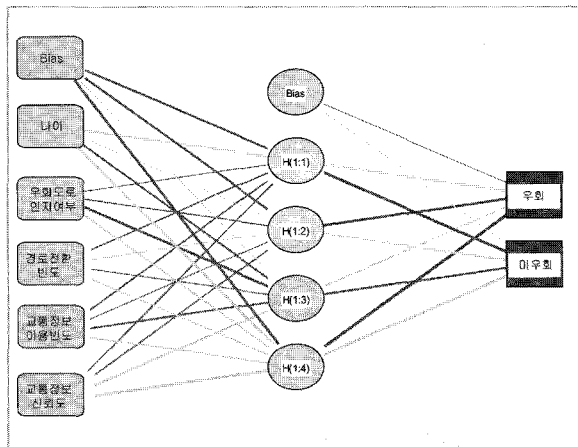


그림 3. 네트워크 다이어그램(Synaptic weight)

결과로 나타난 네트워크 다이어그램을 해석하면 우회를 선택하는 경우 우회도로 인지여부, 교통정보 이용빈도, 교통정보 신뢰여부가 가중치가 높은 것으로 분석되어 중요한 변수로 판단된다.

구축한 모형의 성능을 판단하는 ROC(Receive Operating Characteristics) 곡선을 통해 민감도와 특이도를 나타낸 결과는 그림 4와 같다.

ROC곡선은 좌상의 꼭지점에 가까울수록 평가모형이 우수하다는 것을 가르키며 곡선아래의 면적(AUC:area under the ROC curve)가 넓을수록 그 모형의 설명력이 높음을 알 수 있다. 일반적으로 AUC가 0.9~1.0일 경우 매우 정확한 모형으로 판단하는데, 본 모형의 AUC 값은 0.938로 우수한 모형이라고 할 수 있다.

또한 훈련데이터와 검정데이터를 구분하여 분할표본 타당성 검사를 실시한 결과 훈련표본의 오분류가 나오게 될 위험도 추정값이 0.123이고 검정표본의 오분류

가 나오게 될 위험도 추정값은 0.087로 두 데이터에서 얻는 결과의 차이가 적어 모형의 안전성이 있다고 판단된다.

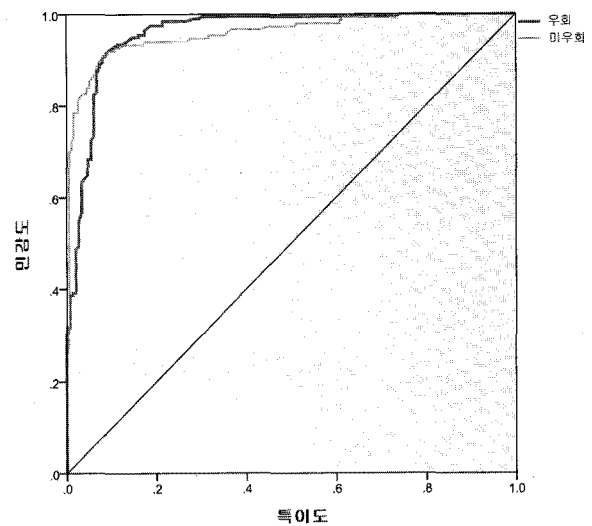


그림 4. ROC 곡선

5. 모형의 적합도 비교

본 연구에서는 의사결정나무와 신경망 결합모형의 객관적 평가를 위하여 우회결정문제를 추정하는 전통적인 방법인 로짓모형 및 의사결정나무 단독모형과의 결과값 비교를 실시하였다. 각 모형의 적합도는 오분류표(Misclassification Table)을 통해 비교하였다.

오분류표(Misclassification Table)란 목표변수의 실제 범주와 모형에 의해 예측된 분류범주 사이의 관계를 나타내는 표라고 할 수 있다. 즉, 목표변수의 범주별로 이를 제대로 분류한 빈도와 그렇지 못한 빈도를 함께 제시한 표이다

이는 목표변수의 범주가 c 개인 경우 $c \times c$ 개의 셀로 이루어진 표 형식을 취한다. 이 때 분류표의 대각에 위치하는 셀은 목표변수의 원래 범주가 제대로 분류된 것을 뜻하고, 비대각에 위치하는 셀은 분류가 잘못된 부분의 빈도를 말한다(염현섭, 2005).

본 연구에서는 오분류율(Error Rate)은 실제값과 예측값이 차이가 날 확률, 정분류율(Accuracy)은 실제값과 예측값이 일치할 확률을 말하며, 민감도(Sensitivity)는 실제 미우회자를 미우회자로 예측할 확률을 뜻하고, 특이도(Specificity)는 실제 우회자를 우회자로 예측할 확률을 의미한다. 오분류표(Misclassification Table)를 통한 각 모형별 적합도 비교결과는 표 5와 같다.

		예측집단		
		0	1	
실제집단	0	n11	n12	n1+
	1	n21	n22	n2+
		n+1	n+2	Total: n

- 오분류율(Error Rate)	→	$\frac{n_{12} + n_{21}}{n}$
- 정분류율(Accuracy)	→	$\frac{n_{11} + n_{22}}{n}$
- 민감도(Sensitivity)	→	$\frac{n_{22}}{n_{2+}}$
- 특이도(Specificity)	→	$\frac{n_{11}}{n_{1+}}$

그림 5. 오분류표

모형의 적합도 비교결과 의사결정나무 단독 모형의 경우 오분류율이 13.8%로 기존 로짓모형의 오분류율인 12.6%보다 예측력이 낮았으나 의사결정나무와 신경망 결합모형의 오분류율은 8.7%로 예측력이 가장 높은 것으로 나타나 모형의 적합도가 우수한 것으로 분석되었다.

표 5. 모형별 적합도 비교

변수	Logit				주 설명변수
	추정치	표준편차	Wals	유의확률	
지역	.405	.175	5.323	.021	
우회도로 인지여부	2.082	.381	29.841	.000	
경로전환빈도	1.030	.204	25.423	.000	1. 우회도로인지여부
정보이용유무	.605	.220	7.536	.006	2. 신뢰도
신뢰도	1.059	.198	28.529	.000	3. 정보이용유무
VMS사인성	1.702	.474	12.877	.000	4. 경로전환빈도
우회영향 메시지	-.387	.191	4.119	.042	5. 나이
상수항	-11.484	1.513	57.606	.000	
오분류율 (Error Rate)	12.6%				13.8% 8.7%
정분류율 (Accuracy)	87.4%				86.2% 91.3%
민감도 (Sensitivity)	85.2%				77.9% 84.4%
특이도 (Specificity)	88.9%				92.5% 95.8%

6. 결론 및 향후연구과제

본 연구는 운전자의 우회사결정과정 속에 내포되어 있는 비선형성과 불확실성을 밝혀내기 위해 의사결정나무와 신경망 이론을 결합한 모형을 구축하여 우회결정요인 분석을 실시하였다. 그 결과, 기존 연구에서 우회결정요인 분석 시 전통적으로 사용되었던 로짓모형과 의사결정나무 단독 모형과의 비교를 검토한 결과 다음과 같은 결론을 얻을 수 있었다.

첫째, 의사결정나무와 신경망 결합모형의 적합도 결과 오분류율이 8.7%로 기존 모형인 로짓모형 12.6%, 의사결정나무 단독 모형 13.8% 보다 낮은 것으로 분석되어, 운전자 우회결정요인 분석에 관한 모형의 적용 타당성을 확인하였다.

둘째, 의사결정나무와 신경망 결합모형을 통해 기존 로짓모형에서 설명할 수 없었던 우회결정 영향변수간의 교호관계를 용이하게 설명하였다.

이 결과에 따라, 모형의 분석결과 운전자의 우회결정에 영향을 미치는 다양한 요인들 가운데 우회인지 여부와 신뢰도가 가장 영향력 있는 변수로 나타났다.

특히 우회도로를 인지하고 있고, 교통정보를 신뢰하는 경우, 우회경험여부와 정보이용여부에 따라서 우회를 선택하는 것으로 나타났다.

이러한 결과를 바탕으로, 교통량 분산효과와 도로망 효율 증대를 극대화하기 위한 우회관리전략 수립 방안으로 신뢰성 있는 정확한 정보제공이 무엇보다 중요하며, 특히 우회도로를 인지하지 못하는 도로이용자들을 위한 교통정보 서비스 제공방안이 필요할 것으로 판단된다.

향후 연구과제로는 본 연구의 실증분석에 사용된 348명의 숫자는 연구를 위한 관측치로는 크게 부족하다고 볼 수 없으나, 의사결정나무모형의 자식마디로 갈수록 각 범주에 속하는 운전자의 수가 줄어들기 때문에, 현장에 적용 가능한 신뢰성있는 분석을 위해서 보다 많은 양의 표본수가 필요할 것으로 판단된다.

또한 본 모델의 적용성을 높이기 위하여 실제 도로상황에 맞는 시뮬레이션을 실시하고, 운전자 우회상황에서의 실측값과 비교하는 등의 더욱 정밀도 있는 연구가 진행된다면, 본 연구에서 제외되었던 변수인 운전경력이나 교통정보 이용빈도, 우회메세지 등 다른 요인들에 대한 영향력을 세부적으로 알 수 있다.

그리고 의사결정나무와 신경망의 결합에 의한 모형구축이 운전자 우회결정요인 분석에 적용 타당성을 확인

한 만큼 본 모형을 바탕으로 다양한 교통분야에서 운전자의 의사결정과정의 불확실성을 확인할 수 있는 연구가 진행되어야 할 필요가 있는 것으로 판단된다.

참고 문헌

강병서, 김계수(2010), SPSS17.0 사회과학 통계분석, 한나래 아카데미.

김연지(2003), *의사결정나무를 이용한 신경망의 해석*, 박사학위논문, 한국과학기술연구원.

김일평(2008), *우회결정요인 분석을 통한 교통정보 제공 방안에 관한 연구*, 박사학위논문, 홍익대학교.

김중호, 신재만, 박제진, 하태준(2010), *유전자 알고리즘과 신경망 이론의 결합에 의한 신호교차로 위험도 예측모형 개발에 관한 연구*, *대한토목학회논문집 D*, Vol.30 No.4, pp.351~360

서임기, 박상명, 이병주, 남궁문(2006). *교통정보의 신뢰도와 만족도가 경로선택에 미치는 영향*. *대한토목학회 학술대회*, Vol.2006.No.10, pp.3063~3066.

염현섭(2005), *신용평점표와 혼합모형을 이용한 부실기업 예측모형*, 석사학위논문, 연세대학교.

이진천(2000), *입력자료 판별에 따른 데이터 마이닝 기법 선택을 통한 시스템의 성능 개선*, 석사학위논문, 아주대학교.

장정아, 문병섭, 최기주(2005), *고속도로에서의 우회(국도)교통정보 제공에 따른 경로전환 효과분석*, *대한토목학회논문집 D*, Vol.25,no.2, pp.221~226.

진현정(2010), *의사결정나무모형을 이용한 소비자 그룹 분석*, *한국농업경제연구지*, Vol.51.No.1, pp.19~40.

최기주, 장정아, 김성현, 김종식(2004), *VMS 우회정보 제공에 따른 우회율 분석(국도 3호선을 중심으로)*, *대한토목학회논문집 D*, Vol.24.No.6, pp.879~880.

접 수 일 : 2011. 4. 20
심 사 일 : 2011. 4. 21
심사완료일 : 2011. 8. 9