

■ 論 文 ■

데이터마이닝의 베이지안 망 기법을 이용한 교통수단선택 모형의 설계 및 구축

Design and Implementation of Travel Mode Choice Model
Using the Bayesian Networks of Data Mining

김 현 기

(교통개발연구원 책임연구원)

김 강 수

(교통개발연구원 연구위원)

이 상 민

(교통개발연구원 연구위원)

목 차

- I. 서론
 - II. 기존 연구의 고찰
 - III. 베이지안 망의 설계 및 구축
 - 1. 자료
 - 2. 베이지안 망
 - IV. 베이지안 망을 이용한 교통수단선택 모형의
구축 및 분석
 - 1. 베이지안 망의 구축
 - 2. 교통수단선택 모형의 분석
 - V. 결론 및 향후 과제
- 참고문헌

Key Words : 베이지안 망(Bayesian Network), 베이지안 통계 이론(Bayesian Statistics Theory), 데이
터마이닝(Data Mining), 통행수단선택(Travel Mode Choice), 로짓 모형(Logit Model)

요 약

데이터마이닝(Data Mining)은 대용량의 데이터에 존재하는 관계, 패턴, 규칙 등을 효율적으로 탐색하여 이를
모형화함으로써, 유용한 정보로 추출·변환하는 일련의 과정이다. 특히 베이지안 망(Bayesian Network)은 신
경망, 유전자알고리즘, 퍼지이론 등과 더불어 데이터마이닝의 중요한 기법 중의 하나로서, 베이지안 통계 이론
(Bayesian Statistics Theory)를 적용하여 변수들간의 확률적인 관계를 기호화함으로써, 설명변수들과 종속변
수들간의 인과관계를 파악할 수 있다.

이 연구는 기존에 적용된 바가 없는 데이터마이닝의 베이지안 망을 이용하여 수도권 교통수단선택 모형을 구
축한다. 2002년도 수도권 가구통행실태조사 자료의 사회·경제적 특성과 교통체계 특성을 반영하여 베이지안
망을 이용한 교통수단선택 모형을 설계·구축하여, 각 변수들간의 상관관계와 인과관계를 분석함으로써, 설명변
수인 성과 연령의 구성비가 변화했을 때, 교통수단선택의 변화율(확률)을 예측한다.

이 연구를 통해 현실에서는 내재하나 설명변수간의 복잡한 상관성을 배제하고 설명변수들과 교통수단선택간
의 단순한 직선관계를 가정하는 기존 교통수단선택 모형의 한계를 극복할 수 있는 가능성을 제시한다. 또한 선
택되지 않은 교통수단에 대한 정보의 부족으로 인한 교통수단선택 모형 구축의 어려움을 극복한다. 또한 다양한
교통정책에 따른 교통수단선택의 변화를 실시간으로 시뮬레이션 할 수 있는 방법론을 개발한다.

I. 서론

데이터마이닝(Data Mining)은 대용량의 자료에 존재하는 자료간의 관계, 예기치 않던 패턴, 새로운 규칙 등을 탐색적으로 찾아내고, 이를 모형화해서 유용한 정보로 변환하는 일련의 과정이다. 베이저안 망(Bayesian Network), 군집분석, 인자분석, 회귀분석, 기계학습(Machine Learning), 패턴인식추론(Pattern Recognition Inference) 등이 데이터마이닝을 위해 주로 사용되는 기법이며, 이 중 베이저안 망은 베이저안 통계 이론(Bayesian Statistics Theory)을 통해 변수들간의 확실적인 관계를 기호화함으로써, 설명변수와 종속변수의 인과관계를 파악할 수 있는 통계적 기법이다.

이 연구의 목적은 기존의 연구에서는 적용된 바가 없는 데이터마이닝의 베이저안 망을 이용하여 수도권 교통수단선택 모형을 구축하는 것이다. 2002년 실시된 수도권 가구통행실태조사의 상세한 정보를 이용하여 새로운 개념의 교통수단선택 모형을 설계·구축하고 이를 분석한다.

이 연구를 통해 현실에서는 내재하나 설명변수간의 복잡한 상관성을 배제하고 설명변수와 교통수단선택 간의 단순한 직선관계를 가정하는 기존 교통수단선택 모형의 한계를 극복한다. 또한 선택되지 않은 수단에 대한 정보의 부족으로 인한 교통수단선택 모형 구축의 어려움도 극복 할 수 있다.

논문의 구성은 다음과 같다. 논문의 목적 및 배경을 설명하는 서론 부분에 이어, II절에서는 기존 교통수단선택 모형 구축에 관한 국내·외 연구사례를 살펴본다. III절에서는 이 연구를 수행하기 위한 조사자료와 베이저안 망의 개념을 설명하고, IV절에서는 베이저안 망을 이용한 교통수단선택 모형의 분석 결과를 제시한다. 마지막으로 결론 부분에서는 연구의 요약, 종합 및 향후과제를 담는다.

II. 기존 연구의 고찰

이 절의 목적은 기존 교통수단선택 모형 구축에 대한 사례를 통해 교통수단선택 시 활용되어지는 설명변수를 살펴보고, 이를 베이저안 망을 이용한 분석 시 설명변수로 활용하고자 하는데 있다. 또한 이 연구에서 제시하는 방법론이 기존의 연구와 어떤 차별성을 지니는지에 대해서도 설명하고자 한다.

기존 교통수단선택 모형은 일반적인 로짓 모형과 그 모형의 한계를 극복하는 새로운 모형으로 대별 할 수가 있으나, 모두 설명변수와 교통수단선택과의 관계를 연역적으로 정립하여 구축한다.

이진우(1995)는 교통수단선택 설명변수를 개인특성 변수(범주형 : 성, 결혼 유·무, 학력, 현재 출근 교통수단, 승용차 보유 유·무)와 대안특성변수(실수 : 차내시간, 차외시간, 통행비용(승용차, 택시) 및 통행요금(버스, 지하철))로 구별하고, 직교배열표를 이용한 실험계획법을 이용하여 승용차, 노선버스, 중앙전철 분담율을 제시하는 교통수단선택 모형을 구축하였다. 양영석외 다수(1995)는 개인특성변수(성, 연령, 월평균 소득)와 대안특성변수(차내시간, 차외시간, 통행비용(승용차, 택시) 및 통행요금(버스, 지하철))를 다항로짓 모형에 적용하여, 교통수단(승용차, 택시, 버스, 지하철) 분담율을 제시하였다. 원제무·손기복(1998)은 인간의 판단과 유사한 구조를 갖는 퍼지근사추론(FARM: Fuzzy Approximate Reasoning Methods)을 구축하여, 입력변수로는 버스과 지하철간의 총통행시간의 차이와 총통행비용의 차이를 선정하였으며, 버스와 지하철의 교통수단선택 모형을 구축하였다.

조중래·김채만(1998)은 서울시 가구통행실태조사자료(1996)를 이용하여, 서울과 일산 신도시의 출근 통행에 대하여, 사회·경제적 대안특성변수(성, 연령, 소득, 면허소지 여·부, 차량소유 여·부, 가족수)와 일반변수(통행시간, 통행비용)를 이용하여 다항로짓 모형을 이용한 교통수단선택(승용차, 버스, 지하철 및 택시) 모형의 지역간 이전 가능성을 분석하였고, 출근 통행의 시간 가치 및 탄력성을 분석·비교하였다. 김대웅외 다수(1999)는 도시내에서 비교적 안정된 패턴을 이루고 있는 취업자를 대상으로 대구광역시 사람통행실태자료(1988, 1992)를 이용하여, 취업자가 1일에 행한 복수통행간의 상관관계를 보다 체계적으로 명확히 규명하고, 개인 속성자료(성별, 연령별, 승용차 보유 유·무) 및 대안특성변수(총통행시간과 통행비용)를 기초로 하여 교통수단선택(버스, 택시, 승용차) 다항로짓 모형을 구축하여, 그 모형의 이전 가능성을 검토하였다.

기존의 연구들을 고찰해 볼 때, 기존의 연구들은 대부분 교통수단선택의 분담율을 구하기 위하여 설명변수로 성, 연령, 그리고 각 교통수단의 시간과 비용 변수를 설정하여 고려하였으며, 대부분 로짓 모형을

이용한 분석 방법론을 제시하고 있다.

따라서, 이 연구에서 제시하는 실제의 조사자료를 이용하여 교통수단선택과 설명변수간의 확률적인 인과관계를 설정하고 분석하는 사례는 전무한 형편이다.

한편, 국내·외의 다른 분야에서 베이지안 망을 적용한 기존의 연구를 살펴보면 다음과 같다.

Hailfinder(Abramson et al., 1996)는 미국 콜로라도 북동부 지역의 악기상을 예측하기 위해 개발한 시스템으로, 기상 전문가는 위 지역에서의 날씨를 11개의 전형적인 날씨 형태로 시나리오를 만들고, 56개의 기상요소 노드들과 각각의 상태를 정의하였다. 이 연구에서는 해당지역의 악기상(우박, 폭우, 강풍)의 징후를 3개의 상태(속성)로 분류(악기상이 심한 경우, 뚜렷한 경우, 징후가 없는 경우)하여 실험하였으며, 악기상의 징후가 뚜렷할 경우의 예측값이 44.0(%)으로 가장 높게 나타났다.

또한 걸프전을 전후로 한 세계 오일 시장의 가격변동을 예측하기 위한 시스템인 ARCO1(Abramson & Finizza, 1991)은 오일 가격의 변동에 영향을 미치는 전쟁, 정치적인 상황, 경제적인 요소들, 생산국가들간의 이해관계 등을 고려하여 베이지안 망을 구축하였으며, 이를 오일 가격의 변동을 예측하는데 사용하였다. 이 연구는 1990년대 초 걸프전이 발생하기 전에 기존의 자료와 전문가의 주관적인 평가를 기준으로 모델을 수립하였다.

또한 국내 연구로 정재훈(1999)이 베이지안 망을 적용하여 공군비행장의 기상을 예측하고, 기상 제한치를 고려하여 필요한 비행정보를 비행운영자에게 제공하기 위하여 10년간의 기상 자료를 입력하여 실험한 결과 73.0(%)의 정확도로 기상을 예측하였다. 또한 기상 예측이외에도 비행운영자에게 필요한 여러 가지 정보를 확률로 제공하였다.

차운옥(2002)은 속성들을 평가하는 척도에 따른 네 가지의 속성선택방법을 실제 데이터베이스에 적용하여 베이지안 망을 구축하고 분류 정확도를 구하였다. 그 결과 속성들의 수가 비교적 많은 데이터베이스에서는 속성들의 수를 평균 50(%) 정도 줄였음에도 불구하고, 원래의 속성들을 모두 사용하는 경우와 큰 차이가 없거나 오히려 정확도가 높음을 알 수 있었다.

이처럼 베이지안 망에 관한 연구는 의학진단 시스템에서 시작되어, 농업, 컴퓨터 산업, 기상, 마케팅, 경제 등의 산업의 모든 분야로 확산되고 있다. 그밖에 기상예

측, 소프트웨어 디버깅, 건강진단 등 불확실성이 존재하는 분야에 이용되어 왔다. 그러나 베이지안 망의 설계 과정이 아직 초기 단계에 있으며, 현실에서는 대부분의 전문가나 의사결정자의 주관에 의존하고 있는 실정이다.

III. 베이지안 망의 설계 및 구축

이 절에서는 2002년도 서울시 가구통행실태조사 자료를 이용하여, 국내의 연구로는 처음으로 가구·개인 특성과 교통체계 특성을 반영한 데이터마이닝의 베이지안 망을 교통수단선택 모형에 적용하여, 교통수단 분담율(확률)을 제시하고, 설명변수들의 변화에 따른 장래의 교통수단 분담율을 예측한다.

1. 자료

2002년도 4월 17일(수) 기준인 수도권 가구통행 실태조사 자료가 이 연구에서 사용되었다. 2002년도 수도권 6,548,616 가구 중 조사된 가구수는 197,898 가구이며(유효표본율: 2.5%), 이 중 실제적으로 서울시에서 공개한 53,351 가구(32.89%)에 대한 가구, 개인 및 통행 자료를 이 연구의 분석에 이용하였다.

가구 속성(미취학 아동수, 차량보유 여부 및 현황, 주택 종류, 규모 및 소유형태, 가구수입 등), 개인 속성(성, 연령, 직업, 고용형태, 운전면허보유 및 장애 여부 등), 통행 속성(출생연도, 통행일자, 출발 및 도착 시간, 출발지 및 도착지, 대중교통요금, 승차인원, 주차비지불여부 등)에 대한 속성들이 조사되었다.

2. 베이지안 망

이 절에서는 베이지안 망을 이용한 교통수단선택 모형의 설계 및 구축에 대해서 설명한다. 먼저 베이지안 망의 개념에 대해서 예를 들어 설명하고, 이를 교통수단선택 모형의 구축 시에 활용하는 방법에 대해서 설명한다.

도시 A의 교통수단 분담율에 가장 큰 영향을 미치는 요소가 성이며, 성에 따른 교통수단 선택확률이 <표 1>과 같이 주어졌다고 가정하자. <표 1>은 설명변수인 성(S : 남, 여)에 대한 종속변수인 교통수단(M : 버스, 지하철, 택시, 승용차)에 대한 도시 A의 교통수단 분담율을 보여주고 있다.

〈표 1〉 성에 따른 교통수단의 선택 확률(예)

종속 변수	상태	성(S)		선택 확률
		남자	여자	
교통 수단 (M)	버스	0.1841	0.2030	0.3871
	지하철	0.0932	0.0671	0.1603
	택시	0.0103	0.0112	0.0215
	승용차	0.3289	0.1022	0.4311
선택 확률	-	0.6165	0.3835	1.0000

성에 따른 교통수단 선택의 변화를 분석하기 위해 관측된 도시 A의 조사결과를 이용하여 남자이고 버스를 선택할 결합확률(Joint Probability)인 $P(\text{남자}, \text{버스})$ 는 식(1)과 같은 사전정보¹⁾와 사후정보²⁾로 표현할 수 있다.

$$P(\text{남자}, \text{버스}) = P(\text{남자}) \cdot P(\text{버스} | \text{남자}) = P(\text{버스}) \cdot P(\text{남자} | \text{버스}) \quad (1)$$

여기서,

$$P(\text{남자}) : 0.6145$$

$$P(\text{남자}, \text{버스}) : 0.1841$$

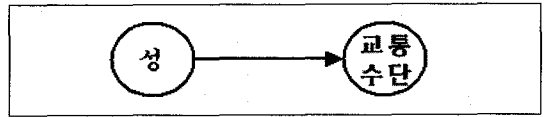
$$P(\text{남자} | \text{버스}) : (0.1841/0.3871) = 0.4756$$

$$P(\text{버스} | \text{남자}) : (0.1841/0.6165) = 0.2986$$

조사자료를 통해 베이저안 통계량(사전정보 및 사후정보)을 구성하는 식(1)을 통해 성별에 따른 교통수단의 확률관계를 정립할 수 있고, 이러한 확률관계를 통해 장래의 성의 구성비 변화에 따른 교통수단선택 확률의 변화를 분석하고 예측할 수 있다. 예를 들어, 장래 남자의 구성비($P(\text{남자})$)가 현재의 61.65%에서 장래에 50.44%로 변화는 경우, 이에 따른 $P(\text{버스})$ 의 변화는 조사자료를 통해 구축되어진 $P(\text{남자}, \text{버스})$, $P(\text{남자} | \text{버스})$, $P(\text{버스} | \text{남자})$ 로 분석할 수 있다.

한편, 이러한 성과 교통수단 선택의 베이저안 종속관계를 보다 명확히 하기 위해 그래프로 표현한 것이 베이저안 망이다.

〈그림 1〉은 성에 따른 교통수단의 변화를 예측하기 위한 베이저안 망을 나타내고 있는데 변수들간

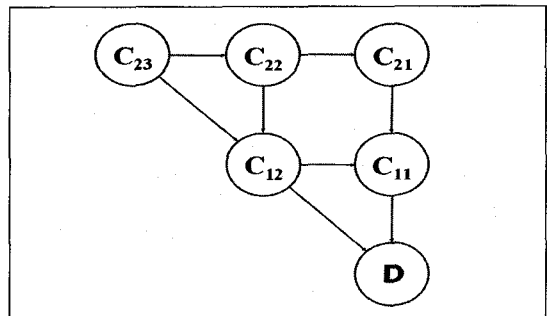


〈그림 1〉 성에 따른 베이저안 망

의 확률적인 관계를 기호화함으로써, 설명(독립)변수인 성과 종속변수인 교통수단간의 인과관계를 모형화하고 있다.

〈그림 1〉에서 성(S)과 교통수단(M)은 베이저안 망에서 노드라고 하며, 일반적으로 설명변수(S)와 종속변수(M)를 의미한다. 그리고 노드간의 종속적인 관계는 노드간의 연결선(Arc)을 통해 확인할 수 있다.³⁾ 즉 〈그림 1〉에서는 성이 교통수단선택에 종속적임을 의미하며, 연결선은 속성들간의 종속적인 관계(Dependent Relationship)를 나타내는 방향성이 있는 비순환 그래프(DAGs: Directed Acyclic Graphs)이다. 남자와 여자를 성 노드의 속성(B_s)이라 한다. 즉 남자, 여자, 버스, 지하철, 택시, 승용차는 B_s 로 나타낼 수가 있으며, 조건부 확률 $P(M | S)$ 를 B_p 로 표기하는 경우 전체적인 베이저안 망은 $B = (B_s, B_p)$ 로 나타낼 수 있다.

이를 일반화한 전형적인 베이저안 망은 〈그림 2〉와 같다. 여기서 C_i 는 성과 같은 설명(종속, 조건)변수가 되며, D 는 교통수단선택과 같은 종속(결정)변수임을 의미한다. 〈그림 2〉에서 각 설명변수간에도 상호 종속적일 수 있어서 C_{23} 의 경우 C_{12} 을 통하여 D 에 영향을 미침을 알 수 있다.



〈그림 2〉 일반적인 베이저안 망

1) Prior Information 또는 Prior Probability Distribution. 여기서는 $P(\text{남자})$ 을 의미한다.
 2) Posterior Information 또는 Posterior Probability Distribution. 여기서는 $P(\text{남자} | \text{버스})$ 또는 추가 정보(Additional Information, Conditional Probability Distribution)인 $P(\text{버스} | \text{남자})$ 를 의미한다.
 3) 노드들간의 종속적인 관계를 나타낼 때 설명변수를 부모노드, 종속변수를 자손노드라 하며, 또한 설명변수도 또 다른 설명변수의 종속변수가 될 수 있다.

이러한 베이지안 망은 대용량의 자료에 존재하는 자료간의 관계, 예기치 않은 패턴, 새로운 규칙 등을 탐색적으로 찾아내고, 이를 모형 화함으로써 유용하게 쓰이고 있다. 특히 베이지안 통계 이론을 통해 변수들간의 확률적인 관계를 기호화함으로써 설명변수와 종속변수와의 인과관계를 파악할 수 있는 장점이 있다.

베이지안 망을 사용하면 실제 데이터베이스에 내재하고 있는 지식을 규칙의 형태로 도출할 수 있으며, 이외에도 차운욱(2002)은 베이지안 망의 특징을 아래와 같이 같이 6 가지로 요약하였다.

① 읽기 용이성(Readability)

망 구조를 그래프 형태로 나타내므로 데이터베이스를 구성하는 속성들간의 관계를 쉽게 나타낼 수 있다.

② 사전지식(Prior Knowledge)

사용자의 선형적 지식을 망 구조를 찾기 위한 탐색 과정이나 구축된 모델을 수정하는데 사용할 수 있다.

③ 낮은 복잡도(Low Complex)

조건부 독립성에 따라 각각의 확률변수의 확률분포는 그 노드의 부모 노드에만 종속되므로 망을 분해할 수 있다. 따라서 변수들간의 관계를 찾아내기 위해 매우 적은 수의 인자만을 필요로 한다. 그래프의 구조를 학습할 때에도 각각의 노드에 대해 적합한 부모 노드들의 집합을 찾으면 되기 때문에 탐색과정이 간단하다.

④ 신뢰도(Confidence Value)

추론결과가 확률형태로 주어지므로, 결과에 대한 신뢰도 측정에 사용할 수 있다.

⑤ 쌍방향 추론(Bi-Directional Inference)

네트워크 구조로부터 "IF A THEN B"와 같은 규칙이 얻어졌을 때, 사실 A가 관찰되면 B를 추론해 낼 수 있고(Deduction), 사실 B가 관찰된 경우, 이것의 원인이 사실 A라는 것을 추론해 낼 수 있다(Abduction).

⑥ 불확실성(Uncertainty) 처리

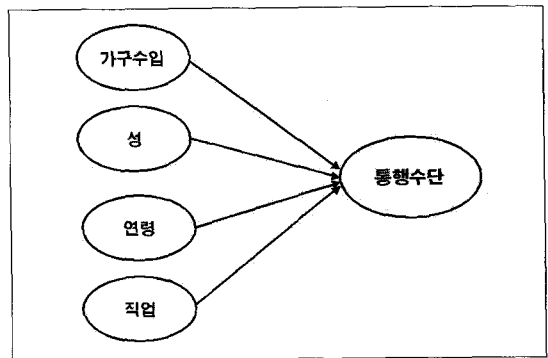
참, 거짓의 두 가지 값이 아닌 믿음의 정도(Degree of Belief)를 확률형태로 표현하므로 불확실성을 효과적으로 처리할 수 있다.

IV. 베이지안 망을 이용한 교통수단선택 모형의 구축 및 분석

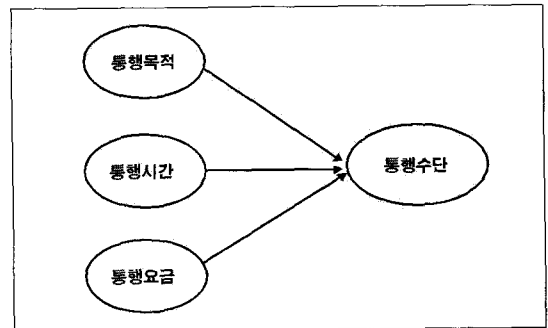
1. 베이지안 망의 구축

이 절에서는 앞서 설명한 수도권 가구통행실태조사 자료와 베이지안 망을 이용하여 교통수단선택 모형을 구축하고, 다양한 시나리오에 바탕을 둔 분석을 실시한다. 이때 기존 문헌을 고찰하여 교통수단에 영향을 미치는 설명변수로 가구수입, 나이, 성, 연령 등 개인의 사회·경제적 특성과 통행요금, 통행시간, 통행목적 등 교통체계 특성을 설정한다. <그림 3>은 개인의 사회·경제적 특성을 반영한 베이지안 망이고, <그림 4>는 교통체계 특성을 반영한 베이지안 망을 나타낸다.

한편, 베이지안 망에서는 각 노드를 그룹화 하여 각 그룹의 확률을 산출할 필요가 있다. 사회·경제적 특성과 통행요금, 통행시간, 통행목적을 반영한 상관관계 분석과 카테고리 분석을 실시하여, 카테고리위와 급간을 <표 2>, <표 3>과 같이 설정하였다.



<그림 3> 가구·개인 특성의 모형



<그림 4> 교통체계 특성의 모형

〈표 2〉 가구·개인 특성을 반영한 카테고리 설정

변수	속성	실측치
통행 수단	버스	45.13
	지하철	22.51
	택시	3.52
	승용차	28.84
연령 (세)	1~7	0.51
	8~24	39.87
	25~44	37.65
	45~64	20.05
	65이상	1.91
가구 수입 (만원)	0~119	36.88
	120~239	18.84
	240~339	2.51
	400이상	41.77
직업	무직	51.68
	1차산업	8.70
	2차산업	27.49
	3차산업	12.13
성	남자	58.52
	여자	41.48

〈표 3〉 통행 특성을 반영한 카테고리 설정

변수	상태	실측치
통행 수단	버스	66.97
	지하철	27.81
	택시	5.12
	승용차	0.10
통행 요금	0~500	23.31
	500~1,000	65.08
	1,000~1,500	7.08
	1,500이상	4.53
통행 목적	개인용무	3.97
	귀가, 퇴근	45.52
	쇼핑, 여가, 픽업	5.31
	통근, 통학	45.20
통행 시간	0~15분	33.11
	16~45분	47.34
	46~105분	18.04
	105분 이상	1.51

2. 교통수단선택 모형의 분석

1) 가구·개인 특성 변화에 따른 수단선택 분석

연령, 가구수입, 직업, 성 등의 가구·개인 특성

변화에 따른 교통수단선택의 변화를 예측하기 위해 조사자료로부터 〈그림 3〉과 같이 교통수단과 연령, 가구 수입, 직업, 성의 확률적인 관계를 베이지안 망으로 구축하였다.

2002년도 수도권 가구통행실태조사 자료의 사회·경제적 특성을 반영하여 분석한 결과, 교통수단 분담율은 버스 45.13%, 지하철 22.51%, 택시 3.52%, 승용차 28.84%으로 나타났다.

또한 조사자료를 이용하여 설정된 베이지안 관계를 통해 연령 또는 연령과 성의 설명변수의 구성비가 변했을 때, 교통수단 선택은 어떻게 변화하는 지를 분석하였다.

따라서, 〈표 4〉, 〈표 5〉에서 제시되는 예측치는 연령 또는 연령과 성의 구성변화에 따른 교통수단의 변화 예측치를 의미한다. 한편, 조사자료를 통해 교통수단선택 모형을 구축하고, 다시 구축된 모형을 기존의 수단선택모형의 결과와 비교하는 기존의 정산 및 검증 작업은 이 논문의 접근방법과 상이함을 밝혀둔다.

(1) 연령의 변화

연령의 변화에 따른 교통수단선택의 확률변화를 예측하기 위해 8~24세의 연령의 구성비(%)를 현재 39.87%에서 34.87%로 5%를 감소시키고, 65세 이상 노인의 연령별 분포를 현재 1.91%에서 6.91%로 5%를 증가시켰을 때, 교통수단 분담율의 변화를 살펴보자. 이때 교통수단 분담율(%)은 버스가 45.13%에서 44.21%로 감소하며, 지하철은 22.51%에서 22.65%로, 택시는 3.52%에서 3.73%로 승용차는 28.84%에서 29.42%로 증가하였다. 따라서, 연령의 분포가 고령화 될 수록 버스를 제외한 지하철, 택시, 자가용의 수단분담 증가율은 증가하고 있다.

(2) 연령과 성의 변화

연령의 변화와 더불어 여성의 구성비(%)를 현재 41.48%에서 46.48%로 5%를 증가시켰을 때, 교통수단 분담율은 버스가 45.13%에서 57.66%, 지하철은 22.51%에서 24.77%로 증가하며, 택시는 3.52%에서 4.35%로, 승용차는 28.84%에서 13.02%로 크게 감소하였다. 따라서, 성 구성비의 변화가 버스, 지하철 등 대중교통 수단의 선택확률에 큰 영향을 미치고 있음을 알 수 있다.

2) 교통체계 특성 변화에 따른 수단선택 분석

통행요금, 통행목적, 통행시간 등의 교통체계 특성에 따른 교통수단선택의 변화를 예측하기 위해 조사자료로부터 <그림 4>와 같이 교통수단과 통행요금, 통행목적, 통행시간의 확률적인 관계를 베이지안 망으로 구축하였다. 즉 조사자료를 이용하여 설계된 베이지안 관계를 통해 통행시간 또는 통행시간과 통행요금의 설명변수 구성비가 변했을 때, 교통수단 선택이 어떻게 변화하는 지를 분석하였다.

2002년도 수도권 가구통행실태조사 자료의 교통체계 특성을 반영하여 분석한 결과, 교통수단 분담율은

버스 66.97%, 지하철 27.81%, 택시 5.12%, 승용차 0.10%으로 나타났다.

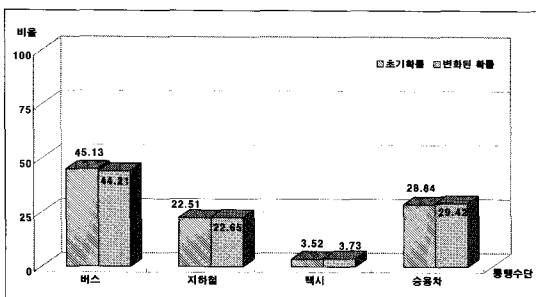
참고로, 앞에서 제시한 가구·개인 특성을 반영한 교통수단 분담율과 교통체계 특성을 반영한 수단 분담율이 차이가 나는 이유는 가구, 개인, 통행의 원시자료(Raw Data)를 병합하여 하나의 총체적인 자료(Aggregative Data)로 만들어 사용할 때, 각각의 베이지안 망 모형에서 고려하는 자료의 형태가 다르기 때문이다. 예를 들어, 원시자료의 통행비용을 고려할 때, 승용차는 조사내용이 없고, 대중교통수단(버스, 지하철)은 이에 대한 정보가 수록되어 있어 가구·

<표 4> 연령 변화에 따른 수단분담율

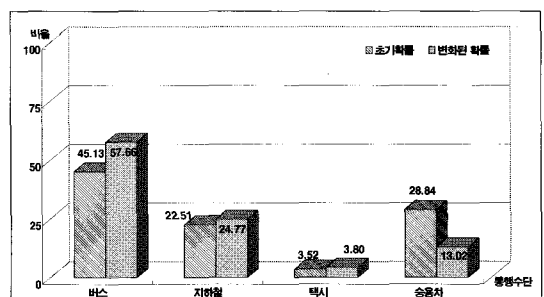
변수	상태	실측치	예측치
통행 수단	버스	45.13	44.21
	지하철	22.51	22.65
	택시	3.52	3.73
	승용차	28.84	29.42
연령 (세)	1~7	0.51	0.51
	8~24	39.87	34.87
	25~44	37.65	37.65
	45~64	20.05	20.03
	65이상	1.91	6.91
가구 수입 (만원)	0~119	36.88	36.88
	120~239	18.84	18.84
	240~339	2.51	2.51
	400이상	41.77	41.77
직업	무직	51.68	51.68
	1차산업	8.70	8.70
	2차산업	27.49	27.49
	3차산업	12.13	12.13
성	남자	58.52	58.52
	여자	41.48	41.48

<표 5> 연령/성 변화에 따른 수단분담율

변수	상태	실측치	예측치
통행 수단	버스	45.13	57.66
	지하철	22.51	24.77
	택시	3.52	4.35
	승용차	28.84	13.02
연령 (세)	1~7	0.51	0.51
	8~24	39.87	34.87
	25~44	37.65	37.65
	45~64	20.05	20.03
	65이상	1.91	6.91
가구 수입 (만원)	0~119	36.88	36.88
	120~239	18.84	18.84
	240~339	2.51	2.51
	400 이상	41.77	41.77
직업	무직	51.68	51.68
	1차산업	8.70	8.70
	2차산업	27.49	27.49
	3차산업	12.13	12.13
성	남자	58.52	53.52
	여자	41.48	46.48



<그림 5> 연령 변화에 따른 수단분담율 경향



<그림 6> 연령/성 변화의 수단분담율 경향

개인 특성의 승용차 부담율이 교통체계 특성의 대중교통수단(버스, 지하철) 부담율로 분산되었기 때문이다.

(1) 통행시간의 변화

통행시간의 변화에 따른 교통수단선택의 변화를 예측하기 위해 16~45분의 통행시간 구성비를 현재 47.34%에서 42.34%로 5%를 감소시키고, 46~105분의 통행시간 구성비를 현재 18.04%에서 23.04%로 5%를 증가시켰을 때, 교통수단 부담율은 버스가 66.97%에서 66.10%로 감소하며, 지하철은 27.81%에서 28.85%로 증가하고, 택시는 5.12%에서 4.95%로 감소하였고, 승용차는 0.10%로 변함이 없었다.

〈표 6〉 통행시간 변화에 따른 수단부담율

변수	상태	실측치	예측치
통행 수단	버스	66.97	66.10
	지하철	27.81	28.85
	택시	5.12	4.95
	승용차	0.10	0.10
통행 요금	0~500	23.31	23.31
	500~1,000	65.08	65.08
	1,000~1,500	7.08	7.08
	1,500이상	4.53	4.53
통행 목적	개인용무	3.97	3.97
	귀가, 퇴근	45.52	45.52
	쇼핑/여가/픽업	5.31	5.31
	통근, 통학	45.20	45.20
통행 시간	0~15분	33.11	34.85
	16~45분	47.34	42.57
	46~105분	18.04	23.99
	105분 이상	1.51	1.59

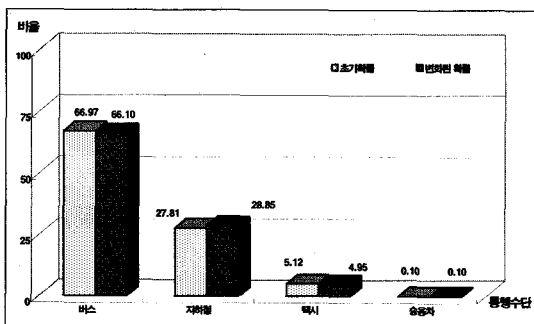
(2) 통행시간과 통행요금의 변화

통행시간의 변화와 더불어 통행요금의 변화를 예측하기 위해 500~1,000원의 통행요금 구성비를 현재 65.08%에서 55.08%로 10%를 감소시키고, 1,000~1,500원의 통행요금 분포를 현재 7.08%에서 17.08%로 10%를 증가시켰을 때, 교통수단 부담율은 버스가 66.97%에서 66.98%로 증가하며, 지하철은 27.81%에서 25.77%로 감소하며, 택시는 5.12%에서 7.11%로, 승용차는 0.10%에서 0.14%로 증가하였다.

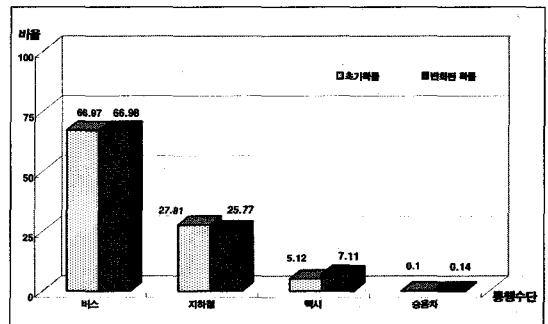
이처럼 교통수단 부담율과 설명변수간의 배이지안망을 구축함으로써 다양한 교통정책 변화에 따른 교통체계의 변화를 기존 조사자료로 부터 보다 빠른 시간 안에 시뮬레이션 할 수 있다.

〈표 7〉 통행시간/통행요금 변화에 따른 수단부담율

변수	상태	실측치	예측치
통행 수단	버스	66.97	66.98
	지하철	27.81	25.77
	택시	5.12	7.11
	승용차	0.10	0.14
통행 요금	0~500	23.31	23.31
	500~1,000	65.08	55.08
	1,000~1,500	7.08	17.08
	1,500이상	4.53	4.53
통행 목적	개인용무	3.97	3.97
	귀가, 퇴근	45.52	45.52
	쇼핑, 여가, 픽업	5.31	5.31
	통근, 통학	45.20	45.20
통행 시간	0~15분	33.11	34.23
	16~45분	47.34	42.34
	46~105분	18.04	23.04
	105분 이상	1.51	1.56



〈그림 7〉 통행시간 변화에 따른 수단부담율 경향



〈그림 8〉 통행시간/통행요금 변화에 따른 수단부담율 경향

V. 결론 및 향후 과제

이 연구에서는 기존에 적용된 바가 없는 데이터 마이닝의 베이지안 망을 적용하여 수도권 교통수단 선택 모형을 구축하였다. 교통수단선택과 연령, 가구수입, 직업 및 성의 확률적인 관계를 베이지안 망으로 구축한 후, 연령 또는 연령과 성의 설명변수가 변화했을 때, 교통수단선택의 변화율(확률)을 예측하였다.

이 연구를 통해 현실에서는 내재하나 설명변수간의 복잡한 상관성을 배제하고 설명변수와 교통수단선택의 단순한 직선관계를 가정하는 기존 교통수단선택 모형의 한계를 극복할 수 있는 가능성을 제시하였다. 또한 선택되지 않은 교통수단에 대한 정보의 부족으로 인한 수단선택 구축의 어려움을 극복할 수 있다. 또한, 다양한 교통정책에 따른 교통수단 선택의 변화를 실시간으로 시뮬레이션 할 수 있는 방법론이 개발되었다.

이 연구의 한계 및 향후 과제는 아래와 같이 요약할 수 있다.

첫째, 각 설명변수의 카테고리 설정함에 있어서 통계적이고 보다 정책적인 접근방법이 필요하다는 것이다. 예를 들어, 이 연구에서는 가구수입의 경우, 기존의 연구를 참고하여 4개의 카테고리로 구분하였으나, 이에 대한 구분은 연구자의 판단(상관관계 및 카테고리분석)에 의해 이루어졌다. 이를 상세하게 분할할 경우, 베이지안 망에서 산출해야 하는 조건부 확률을 계산하는 번거로움이 증가하게 되고, 교통수단 선택의 변화율 계산이 달라지나, 이에 대한 분석은 이루어지지 못했다.

둘째, 이 연구에서는 단순히 종속변수인 교통수단선택과 설명변수인 가구·개인 특성과 교통체계 특성만이 베이지안 망의 설계에 고려되어 졌다. 그러나 각 설명변수들간에도 종속관계나 상호관계가 있을 수 있으며 종속변수에 영향을 미치는 과정과 단계도 다양할 수 있으며 이에 대한 연구가 필요하다.

셋째, 최적의 베이지안 망을 설계하는데 있어서, 망의 구성을 연구자의 판단에 의해 순차적으로 가정했으나, 이에 대한 검증 방법은 존재하지 않는다. 따라서 각 변수간의 상관관계와 종속관계를 고려하는 최적의 베이지안 망 구축에 대한 연구가 필요하다.

참고문헌

1. 김대웅·배영석·이명미(1999), “취업자의 1일 통행수단선택 분석 및 모형의 시간이전성 검토”, 대한교통학회지, 제17권 제5호, 대한교통학회, pp.19~32.
2. 김성인·김현기(2002), “실시간 경로조연을 위한 지능형 시스템의 설계 및 구축”, IE Interface, 제15권 제4호, 대한산업공학회, pp.374~381.
3. 김현명·임용택·이승재(1999), “통합교통망 수단선택-통행배정모형 개발에 관한 연구”, 대한교통학회지, 제17권 제5호, 대한교통학회, pp.75~86.
4. 김현기(1996), “중속적 베르누이 시행에서 추이 확률 추정에 관한 일반적인 역이항 샘플링 방법의 연구”, 고려대학교 일반대학원 산업공학과, 석사학위논문.
5. 김현기(2002), “복잡한 대규모의 도로망에서 실시간 경로 조연을 위한 지능형 시스템의 설계 및 구축”, 고려대학교 일반대학원 산업공학과, 박사학위논문.
6. 김현기의 다수(1999), 통계학 개론, 백산 출판사.
7. 배영석·김기혁·김경식·김연동(1993), “개별행태 접근방법에 의한 교통수단선택 행태분석에 관한 연구 -대구광역시 사례를 중심으로-”, 대한교통학회지, 제14권 제4호, 대한교통학회, pp.47~59.
8. 원재무·손기복(1998), “퍼지근사추론을 이용한 교통수단 선택모형 구축”, 대한교통학회지, 제16권 제1호, 대한교통학회, pp.99~110.
9. 윤대식(1999), “통근통행 이전의 비통근통행 발생여부와 교통수단 선택행태 분석”, 대한교통학회지, 제17권 제5호, 대한교통학회, pp.57~65.
10. 이진우(1996), “선호의식데이터에 의한 교통수단선택모델의 특성”, 대한교통학회지, 제13권 제4호, 대한교통학회, pp.31~45.
11. 전경수(1987), “통행기 중점 및 교통수단선택과 결합된 교통망 평가모형의 구성 및 이용”, 대한교통학회지, 제5권 제2호, 대한교통학회, pp.113~126.
12. 정재훈(1999), “베이지안 네트워크를 이용한 기상예측”, 고려대학교 일반대학원 산업공학과, 석사학위논문.

13. 정현영·김정주(2000), "통근·통학자의 지하철 연계교통수단 선택행태분석", 대한교통학회지, 제 18권 제5호, 대한교통학회, pp.113~126.
14. 조중래·김채만(1998), "출근통행 교통수단 선택행태의 지역간 비교연구 -서울과 일산신도시를 중심으로-", 대한교통학회지, 제16권 제4호, 대한교통학회, pp.75~86.
15. 차운옥(2002), "베이지안 네트워크에 의한 지식 추출 및 성능분석", 한성대학교 정보통신 논문집, 제5권, pp.59~68.
16. Abramson, B. & Finizza, A. J.(1991), "Using Belief Network to Forecast Oil Prices", International Journal of Forecasting. Vol.7, pp.299~316.
17. Abramson, B. & Finizza, A. J.(1995), "Probabilistic Forecasts from Probabilistic Models: A Case Study in the Oil Market", International Journal of Forecasting. Vol.11, pp.63~72.
18. Abramson, B., Brown, J., Edwards, W., Murphy, A. & Robert, L. W.(1996), "Hailfinder: A Bayesian System for Forecasting Severe Weather", International Journal of Forecasting. Vol.12, pp.57~71.
19. James, S. P.(1989), "Bayesian Statistics", Wiley Interscience.
20. Jensen, F. V.(1996), "An Introduction to Bayesian Networks", Springer.
21. Jensen, F. V.(2001), "Bayesian Networks and Decision Graphs", Springer.
22. Jose, M. B. & ADRIAN, F. M. S(1994), "Bayesian Theory", Wiley.
23. Morris, H. and Peg, Y.(1994), "Statistical Analysis for Decision Making", Dryden.
24. Pearl, J.(2000), "Causality Models, Reasoning and Inference", Cambridge.
25. Pham, D. T. & Karaboga, D.(2000), "Intelligent Optimisation Techniques", Springer.
26. Rich, E. N.(1997), "Learning Bayesian Networks", Pearson Prentice Hall.
27. Simon, P. W., Matthew, G. K. & Fred, L. M.(2003), "Statistical and Econometric Methods For Transportation Data Analysis", Chapman & 30. Hall/CRC.

✉ 주 작 성 자 : 김현기

✉ 논문투고일 : 2003. 11. 15

논문심사일 : 2004. 3. 11 (1차)

2004. 4. 13 (2차)

심사판정일 : 2004. 4. 13

✉ 반론접수기한 : 2004. 8. 31