

베이저안 네트워크를 이용한 개인화 된 상품 추천 에이전트

A Personalized Recommender Agent Using Bayesian Network

박진희¹, 정환묵²

¹ 대구가톨릭대학교 컴퓨터정보통신공학부

E-mail : aimajor@cu.ac.kr

² 대구가톨릭대학교 컴퓨터정보통신공학부

E-mail : hrmchung@cu.ac.kr

요 약

소비자가 최적의 상품을 선택하기 위해서는 충분한 상품정보를 파악하여 상품정보를 일일이 조사해야하는 번거로움이 생긴다. 이러한 문제점을 해결하기 위하여 여러 가지 상품추천방법이 제안되고 있으나 상품추천 과정에서 고객의 기호 변화를 다루는 연구가 부족하다.

본 논문에서는 소비자의 기호 변화에 적응하는 개인화 된 상품 추천을 위하여 베이저안 네트워크를 모델링하여 상품 구매에 따르는 선호도를 분석하고, 추천된 상품에 대한 사용자의 행동으로 관심 정도를 측정하여 추천 리스트를 제공한다.

Key Words : 베이저안 네트워크, 개인화 추천시스템

1. 서 론

소비자의 기호가 다양화되고 소비자층에서 개인소비자 기호에 맞는 상품의 공급이 요구되고 있다. 따라서 고객의 프로필 데이터나 구매 경력 데이터만으로는 충분히 분석할 수 있는 데이터가 부족하기 때문에 소비자의 구매에 이르는 심리나 소비심리 내부 상태의 깊은 분석이 요구되고 있다. 따라서 앙케이트 조사를 통해 소비자의 구매에 이르는 행동을 분석하곤 한다. 그러나, 심리적으로 대답하기 어려운 문항에 대해서는 신뢰성이 낮아질 수 있고, 이로 인해 소비자의 구매 심리나 내적 상태를 이끌어 내기 힘들다 [1].

베이저안 네트워크는 관측이 곤란한 요소를 다루는 것이 가능하고 추측되는 가설의 확신도를 실제 데이터를 바탕으로 검증 할 수 있는 장점이 있다.

본 논문에서는 베이저안 네트워크를 이용하여 상품 구매에 따르는 선호도를 분석하고 사용자의 행동을 관측하여 기호의 변화에 대해 적응하는 상품추천 서비스를 제공한다.

입력의 불확실성을 해결하고 개인화된 정보를 처리하기 위한방법으로 퍼지 입력변수와 비 퍼지화 된 입력변수를 베이저안 네트워크 변수로 적용한다. 베이저안 네트워크에서 출력된 상품 목록은 사용자의 행동정보를 이용하여 개인화된 상품서비스를 제공한다.

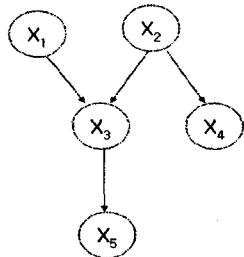
2. 관련연구

2.1 베이저안 네트워크

베이저안 네트워크는 변수사이의 확률적 의존성을 정의하는 방향성 그래프(DAG: Directed Acyclic Graph)이다[3]. 베이저안 확률 추론은 불충분한 정보를 가진 환경을 표현하고 추론하는 대표적인 기법들 중의 하나이다. 그것은 어떤 사실이 관측되었을 경우 환경을 적절히 표현하는 가설이 어떤것인지 추론하며, 베이저안 네트워크는 환경을 표현하는데 사용된다. 네트워크의 노드는 랜덤변수를 표현하며 아크는 변수들 간의 의존성을 표현한다. 아크는 직접적인 인과관계를 나타내는 것으로 생각 할 수 있으

므로 때로는 인과 네트워크(causal network)라고 부르기도 한다[4].

변수는 노드로, 변수간의 의존관계는 원인으로 부터 결과가 되는 변수로서 방향을 가지는 유향링크로 [그림 1]과 같이 나타낼 수 있다.



[그림 1] 베이저안 네트워크 구조의 예

n 개의 확률변수 X_1, \dots, X_n 를 베이저안 네트워크를 이루고 있는 노드라고 가정하고 조건부 독립이라는 가정하에 이 네트워크의 모든 노드에 대한 결합 확률은 식(1) 같이 표현할 수 있다.

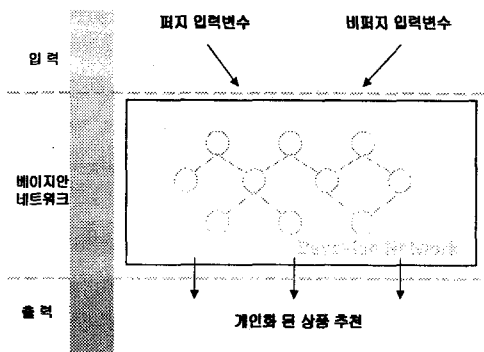
$$P(X_1, \dots, X_n) = \prod_{j=1}^n P(X_j | Pa(X_j)) \quad (1)$$

베이저안 네트워크를 이용하여 모델을 설계 할 때에는 3단계를 거쳐서 모델을 설계한다. 1단계에서는 변수 선택의 단계이다. 2단계는 선택된 변수들 간의 인과관계 값을 정하는 단계이다. 3단계는 확률추론을 하는 추론단계이다[2].

3. 개인화 된 상품추천 에이전트

3.1 상품추천 에이전트의 구조

상품추천에이전트의 구조를 [그림 2]에 나타낸다.



[그림 2] 상품추천 에이전트의 구조

입력의 불확실성을 처리하고 개인화된 정보처리를 위해 베이저안 네트워크 입력에 퍼지입력 변수와 비퍼지화 입력변수를 적용시킨다.

입력은 사용자의 나이, 소득, 구매목적등과 같은 프로필 정보를 이용하여 퍼지 및 비퍼지 입력변수를 생성한다. 베이저안 네트워크에서 출력되어진 개인화 된 상품목록은 사용자의 클릭 횟수와 사용자가 정보를 관찰하는 시간에 따라 관심의 정도를 계산하여 사용자에게 기호에 맞는 상품추천 리스트를 작성한다.

3.2 평가방법

3.2.1 베이저안 네트워크 방법

새로운 인스턴스를 적절한 클래스로 분류하기 위한 베이저안 접근법(Bayesian approach)은 가장 큰 확률값인 v_{MAP} 을 할당하는 것이다. 속성값 $\langle a_1, a_2, \dots, a_n \rangle$ 인 인스턴스가 주어졌을 때 $v_j (v_j \in V)$ 의 확률값 중 가장 큰 확률값인 v_{MAP} 는 식 (2)과 같이 표현한다.

$$v_{MAP} = \arg \max_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n) \quad (2)$$

또한, 베이저안 정리에 의해 식(3)과 같이 표현할 수 있다.

$$v_{MAP} = \arg \max_{v_j \in V} \frac{P(v_j | a_1, a_2, \dots, a_n) P(v_j)}{P(v_j | a_1, a_2, \dots, a_n)} \quad (3)$$

$$= \arg \max_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n) P(v_j)$$

베이저안 분류에서 속성 값 $\langle a_1, a_2, \dots, a_n \rangle$ 간에는 상호조건부 독립이라는 가정을 하고 있다. 이 가정에 의해 목표 값이 주워졌을 때 속성이 $\langle a_1, a_2, \dots, a_n \rangle$ 인 인스턴스의 확률은 목표값이 주워졌을 때 개별적인 속성들의 확률 값을 곱한 것과 같다.

$$P(a_1, a_2, \dots, a_n | v_j) = \prod_{i=1}^n P(a_i | v_j) \quad (4)$$

식(4)에 의해 식(3)을 다음과 같이 나타낼 수 있다.

$$v_{NB} = \arg \max_{v_j \in V} P(v_j) \prod_{i=1}^n P(a_i | v_j) \quad (5)$$

식(5)의 v_{NB} 는 베이저안 분류기에 의해 구한 출력 값이다. 이와 같이 베이저안 학습 방법은 단순히 샘플데이터의 빈도수로부터 각각 $P(v_j)$ 항과 $P(a_i | v_j)$ 항을 구할 수 있다[5].

3.2.2. 선호도 계산 방법

베이저안 학습을 통해 출력된 값 즉, 개인화

된 상품추천에 가중치를 곱하여 식(6)와 같이 선호도를 계산한다.

$$I_{ij} = \sum_k \alpha_k IV_{ij}^k \quad (6)$$

- I_{ij} : 고객*i*의 제품*j*에 대한 선호도
- IV_{ij}^k : 고객*i*의 제품*j*에 대한 k 번째 측정 항목의 값
- α_k : 각 측정항목의 가중치($\sum \alpha_k = 1$)

각각의 측정 항목의 가중치를 부여하기 위해 순서대로 가중치를 부여하는 순위법을 식(7)과 같이 사용하였다.

$$w_j = \frac{n - r_j + 1}{\sum (n - r_k + 1)} \quad (7)$$

- w_j 는 j 번째 기준에 대한 정규화 된 가중치
- n 은 고려의 대상인 기준의 개수
- r_j 는 기준의 순위

각각의 기준은 $n - r_j + 1$ 의 가중치를 가지며, 정규화하기 위해 모든 가중치의 합인 $\sum n - r_k + 1$ 로 나누어주게 된다.

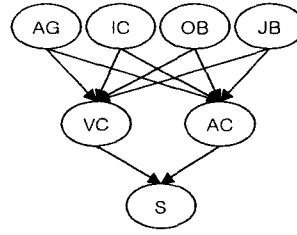
4. 모의실험

해외여행 상품의 선호도를 분석하기 위해 한국관광공사의 2005년 국민 해외여행 실태조사 자료를 이용하였다. 해외여행을 하고자 할 때 어느 도시를 여행할 것인가와 여행 중 주요 활동에 대한 성향을 분석하기 위하여 양케이트 조사의 데이터로 베이지안 네트워크를 모델링 하였다.

여행 도시로 전체 응답에서 상위 8개 도시를 이용하였고, 해외 여행시 활동으로는 전체 응답의 상위7개의 활동을 선택하였다. 사용자의 선호도를 분석하기 위해서는 연령, 연소득, 직업, 여행목적을 이용하였다.

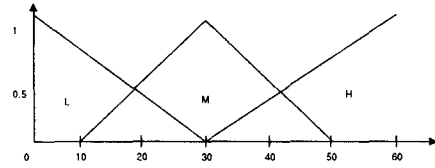
연령은 10대, 20대, 30대, 40대, 50대, 60세 이상으로 구분하고, 연소득은 1천만원, 3천만원, 4천만원 이상으로 구분하였으며, 직업은 전문/관리직, 행정/관리직, 사무/관리직, 판매종사자, 서비스종사자, 농림축수산업, 생산/노무직, 가정주부, 학생, 기타로 구분하고, 여행목적은 여가/휴양, 방문, 사업 전문 활동, 종교 및 순례, 연수, 봉사활동으로 구분하였다.

해외여행실태 조사에 대한 데이터를 이용한 베이지안 네트워크는 [그림 4]와 같이 구성하였다.

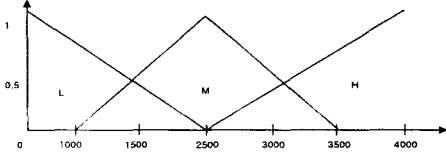


[그림 4] 베이지안 네트워크 구조

(a) 나이



(b) 연봉



[그림 5] 퍼지 소속함수

[표 1] 퍼지데이터의 속성값

변수	설명	속성값
AG	나이	L, M, H
IC	소득	L, M, H

[표 2] 비퍼지 데이터의 범위

변수	설명	범위
OB	여행목적	1~8
JB	직업	1~11
VC	방문도시	1~10
AC	주된 활동 사항	1~3
S	여행스케줄	1~50

사용자의 데이터로 연령, 소득, 직업, 여행목적의 항목에 대해서 각 23세, 2500만원, 사무/관리직, 여가/휴양을 입력하면 여행도시는 상해, 예측되는 주된 활동으로는 '문화답사-쇼핑-도시관광'이라는 선호 경향을 추론 할 수 있다.

위의 분석결과로 검색한 추천 리스트는 [표 3]과 같고 사용자에게 행동에 의해 선호도를 측정함으로써 개인의 기호를 분석한다.

[표 3] 추천 상품에 대한 선호도 측정

분류	번호	추천 상품	Click	Time	선호도
문화답사 A	1	a ₁	1	15	1
	2	a ₂	1	30	2
	3	a ₃	3	55	4
	4	a ₄	1	15	1
쇼핑 B	1	b ₁	1	30	2
	2	b ₂	4	90	6
	3	b ₃	2	15	2
	4	b ₄	3	85	4
도시관광 C	1	c ₁	1	30	2
	2	c ₂	2	55	3
	3	c ₃	2	15	2
	4	c ₄	0	0	0

with Bayesian Networks”, Technical Report MSR-TR-95-06, Microsoft Research, March, 1995.

- [4] F.V. Jensen, Bayesian Networks and Decision Graphs, Springer-Verlag Berlin Heidelberg New York, 2001.
- [5] TOM M. Mitchell, "Machine Learning" McGraw-Hill, 1997.

5. 결론

소비자가 최적의 상품을 선택하기 위해서는 상품에 대한 정보를 일일이 조사해야한다. 이러한 번거로움을 해결하기 위해 여러 가지 해결 방안이 제시 되었지만 상품 추천 과정에서 수시로 변화하는 고객의 기호에 맞추기에는 어려움이 있다.

본 논문에서는 퍼지를 이용해 전처리 단계를 한 후 베이지안 네트워크를 통하여 개인화 된 정보를 추출하는 방법을 제안하였다.

사용자의 행동 정보를 이용하여 개인의 기호에 따라 상품 추천리스트를 제시하고 사용자 선호도를 계산하여 개인화된 상품추천 서비스를 제공한다. 사용자를 대신하여 유저모델이 추천된 각 상품에 대해 평가를 함으로써 사용자의 기호와 목적에 보다 적합한 추천 목록이 제공된다.

향후 연구로는 사용자의 기호와 행동에 대한 다양한 정보로 사용자를 모델링하기 위한 연구가 필요하다.

6. 참고문헌

- [1] 村上 外, 베이지안네트워크による消費者行動分析, 電子情報通信学会, 2004.
- [2] 황정식, 피수영, 손창식, 정환목, “베이지안 네트워크와 신경망을 이용한 구매패턴 분석”, 한국퍼지 및 지능시스템학회, Vol. No. pp.306-311, 2005.
- [3] D. Heckerman. "A tutorial on learning