

베이지안 네트워크와 신경망을 이용한 구매패턴 분석

A Purchase Pattern Analysis Using Bayesian Network and Neural Network

황정식 · 피수영 · 손창식 · 정환목

Jeong-Sik Hwang, Su-Young Pi, Chang-Sik Son, Hwan-Mook Chung

대구가톨릭대학교 컴퓨터정보통신공학부

요약

소비자의 구매패턴을 분석하기 위해서는 문화적 요인, 사회적 요인, 개인적 요인, 심리적 요인 등을 고려해야 한다. 소비자의 구매에 이르는 심리나 내부 상태까지 깊이 고려하여 분석하게 되면 소비자의 구매행동을 정확하게 예측하거나 구매요인을 추출할 수가 있기 때문에 기업에서는 소비자의 기호에 맞는 상품개발에 효과적으로 이용할 수가 있다. 이러한 요인들은 직접적인 정보처리로 분석하기 어렵기 때문에 불확실한 정보를 취급하는 기술이 필요하다. 베이지안 네트워크는 불완전한 자료나 관측이 불가능한 자료에도 대응이 가능하며 구매자의 구매에 이르는 심리나 내적상태등과 같은 관측이 곤란한 요소도 다룰 수 있다.

따라서 본 논문에서는 상품구매에 따르는 소비자의 구매행동 패턴을 분석하기 위해 판매자의 노하우와 소비자의 구매의식을 조사하여 이 데이터를 바탕으로 베이지안 네트워크를 구성하고 구매패턴을 예측하는 방법을 제안하였다. 베이지안 네트워크를 이용하여 불필요한 속성을 가진 데이터를 제거하고 코호넨의 SOM을 이용하여 소비자의 구매패턴을 분석하였다.

Abstract

To analyze the consumer's purchase pattern, we must consider a factor which is a cultural, social, individual, psychological and so on. If we consider the internal state by the consumer's purchase, Both the consumer's purchase action and the purchase factor can be predicted, so the corporation can use effectively in suitable goods development in a consumer's preference. These factors need a technology that treat uncertain information, because it is difficult to analyze by directly information processing. Therefore, bayesian network manages elements those the observation of inner state such as consumer's purchase is difficult. In addition, it is interpretable about data that the observation is impossible.

In this paper, we examine the seller's know-how and the way of consumer's purchase to analyze consumer's purchase action pattern through goods purchase. Also, we compose the bayesian network based on the examined data, and propose the method that predicts purchase patterns. Finally, we remove the data including unnecessary attribute using the bayesian network, and analyze the consumer's purchase pattern using Kohonen's SOM method.

Key Words : 베이지안 네트워크, 구매패턴, SOM

1. 서 론

실세계에서의 문제는 매우 복잡하고 예측하기 어려우며 다양한 상황들이 발생한다. 데이터 마이닝 연구 분야에서는 고객의 프로필 데이터나 구매이력 데이터 등에 대해 하나의 목적변수와 설명변수를 정하여 결정목이나 신경망 등의 데이터 마이닝 기술을 이용하여 목적변수를 가장 잘 설명할 수 있는 모델을 선택 한 후 선택된 모델을 이용하여 소비자의 행동 패턴을 분석하고 있다. 그러나 소비자의 구매행동 예측에서는 소비자의 기호가 다양해져 감에 따라 기존의 고객 프로필 데이터나 구매경력 데이터만으로는 소비자의 구매심리나 행동을 분석하기에는 불충분하다[1, 2].

접수일자 : 2005년 4월 1일

완료일자 : 2005년 6월 2일

본 논문은 산업자원부에서 실시한 지역혁신특성화사업의 지원을 받아 작성하였음.

소비자의 구매패턴을 분석하기 위해서는 문화적 요인, 사회적 요인, 개인적요인, 심리적 요인 등을 고려해야 한다. 소비자의 구매에 이르는 심리나 내부 상태까지 깊이 고려하여 분석하게 되면 소비자의 구매행동을 정확하게 예측하거나 구매요인을 추출할 수가 있기 때문에 기업에서는 소비자의 기호에 맞는 상품개발에 효과적으로 이용 할 수가 있다. 판매자는 소비자에게 맞는 정확한 상품을 예측하고 기업의 이익을 최대화하는 마케팅 전략을 세우기 위해서 소비자의 구매에 이르는 행동을 파악하는 설문조사를 실시한다. 설문조사는 비교적 간단하고 대규모 조사가 실시하기 쉽지만 응답작업의 부담이 커서 응답을 하지 않는 경우가 많고 심리적으로 대답하기 어려운 문항에 대해서는 신뢰성이 낮아 질 수 있다. 이로 인해 소비자의 구매심리나 내적 상태를 이끌어내기 힘들다.

따라서 이러한 설문조사의 성질을 고려하여 구매자의 패턴을 분석하는 수단으로 베이지안 네트워크를 이용한다. 베이지안 네트워크는 응답이 없는 데이터 즉, 관측이 곤란한 요소를 다루는 것이 가능하고 추측되는 가설의 확신도를 실

제의 설문자료를 바탕으로 검증이 가능하며 전문가의 지식을 네트워크 구조로 표현이 가능하다. 또한 불완전한 자료나 관측이 불가능한 자료에도 대응이 가능하며 구매자의 구매에 이르는 심리나 내적상태 등의 관측이 곤란한 요소도 다룰 수 있다.

본 논문에서는 상품구매에 따르는 소비자의 구매행동 패턴을 분석하기 위해 판매자의 전문지식과 소비자의 구매의식 설문조사 자료를 바탕으로 베이지안 네트워크를 이용하여 모델링 하였다. 베이지안 네트워크를 이용하여 불필요한 속성을 가진 데이터를 제거한 후 코호넨의 SOM을 이용하여 동일한 패턴에 대해 다른 형태의 출력력을 하는 출력패턴을 학습시켜 소비자의 구매행동 패턴을 분석하고 예측한다.

2. 관련 연구

2.1 베이지안 네트워크

베이지안 네트워크는 확률변수들 간의 의존관계를 네트워크 구조를 사용하여 문제의 대상을 표현하는 확률모델이다 [3, 4]. 베이지안 네트워크는 실체(entity)들 간의 인과관계를 나타내는 확률 모델로서 결손치(missing value) 즉, 관측이 곤란한 요소를 다루는 것이 가능하고 추측되는 가설의 확신도를 실제 데이터를 바탕으로 검증을 할 수가 있으며, 전문가의 지식을 네트워크 구조로 표현하는 것이 가능한 장점이 있다.

변수는 노드로, 변수간의 의존관계는 원인으로부터 결과가 되는 변수로서 방향을 가지는 유향링크로 그림 1과 같이 나타낼 수 있다. 노드를 연결하는 아크는 직접적인 인과관계를 나타내는 것으로 생각할 수 있으므로 때로는 인과 네트워크 (causal network)라고 부르기도 한다[6].

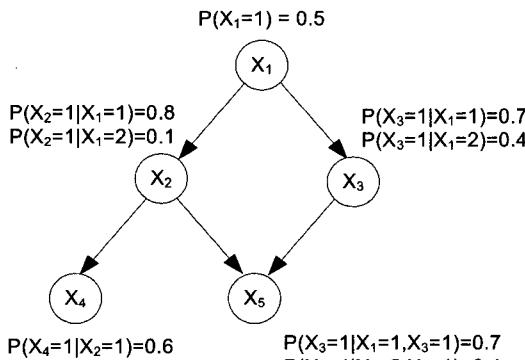


그림 1. 베이지안 네트워크 예

n 개의 확률변수 X_1, \dots, X_n 를 Bayesian Network를 이루고 있는 노드라고 가정하고 조건부 독립이라는 가정하에 이 네트워크의 모든 노드에 대한 결합 확률은 식 (1) 같이 표현할 수 있다. Bayesian Network가 의미하는 부분순서(partial order)와 일관성을 유지하는 연쇄법칙 순서를 사용하는 모든 변수의 결합 확률을 구하는 연쇄법칙에 조건부 독립을 적용함으로써 식 (1)을 쉽게 유도할 수가 있다.

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

베이지안 네트워크를 이용하여 모델을 구축 할 때에는 3 단계를 거쳐서 모델을 구축한다. 1단계에서는 그래프구조를 설정하는 단계로 확률변수로 무엇을 선택할 것이지 즉 변수 선택의 단계이다. 2단계는 선택 된 변수들 간의 인과관계 값을 정하는 단계이다. 3단계는 확률추론을 하는 추론단계이다.

2.2 Kohonen Map

코호넨 맵은 비교사(Unsupervised) 학습 방법으로 스스로 n-차원의 입력 데이터들을 분류하여 2차원으로 나타낸다. 신경망의 일종으로 자율(upsupervised)학습과 경쟁학습방법을 이용하며 BPN(back propagation network)과는 달리 출력 노드의 위상(topology)자체가 의미가 있다. 코호넨 맵은 두 개의 층인 입력층(Input Layer)과 경쟁층(Competitive Layer)인 2차원의 격자로 이루어져 있다. 모든 연결은 첫 번째 층에서 두 번째 층의 방향으로 되어 있으며 두 번째 층은 완전연결(Fully Connected)되어 있다.

코호넨 네트워크(Kohonen network)라고도 불리는 이 신경망을 사용하기 위해서는 층 내의 뉴런 연결강도 벡터가 임의 값을 가지면서 적합하게 최적화 되어야 하고 연결강도 벡터와 입력벡터의 값이 통상 0에서 1사이로 정규화(normalized)되어야 한다[5, 7].

그림 2는 SOM(Self-Organizing Maps)의 구조를 나타낸다.

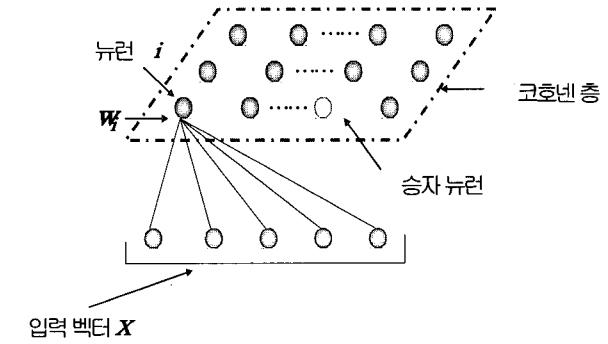


그림 2. SOM의 구조

모든 입력 벡터 \mathbf{X} 는 출력 노드인 코호넨 층과 연결되어 있고 연결 가중치(weight)를 가진다. 초기의 연결 가중치는 일반적으로 0에서 1사이에 임의의 값으로 할당한 후 유clidean 거리를 이용하여 입력 벡터 \mathbf{X} 와의 유사성을 계산한다. 계산한 결과 중에서 가장 가까운 거리를 가진 뉴런이 승자뉴런이 된다. 승자뉴런이 결정되면 학습규칙에 따라 뉴런의 연결강도를 조정한다. 코호넨 맵의 학습규칙은 다음과 같다.

[Step 1] 연결강도 W 를 초기화한다.

(0에서 1사이의 임의의 값으로 설정)

[Step 2] 연결강도를 변경시킬 범위(radius)를 설정하고 학습율 α 을 결정한다.

[Step 3] 입력 벡터 \mathbf{X} 를 입력하여 유사도 D 를 계산하고, D 가 가장 작은 뉴런을 승자뉴런으로 선정한다.

$$D = (W - \mathbf{X})^2 \quad (2)$$

[Step 4] 승자뉴런으로부터 반경 r 범위 내에 있는 뉴런의 연결강도를 변경한다. $k+1$ 학습 단계에서 연결강도 W_{k+1} 는 다음과 같다.

$$W_{k+1} = W_k + \alpha (X - W) \quad (3)$$

[Step 5] 규정된 반복횟수만큼 학습이 진행되면 반경 r 과 학습률 α 을 감소시킨 다음 학습과정을 반복한다.

학습이 진행됨에 따라 뉴런들 간의 반경은 점차 줄어들어서 작은 개수의 뉴런들이 학습을 하게 되며 최종적으로는 승자뉴런만이 연결강도를 조정하게 된다. 이러한 과정이 끝나면 또 다른 입력벡터가 들어오게 되고 위에서 기술한 학습단계를 계속 반복하다가 모든 뉴런들이 변화가 없을 때 종료하게 된다.

3. 베이지안 네트워크를 이용한 노드감축

3.1 전체 시스템 구성도

본 논문에서는 3단계를 거쳐서 소비자의 구매 패턴 분석을 한다. 첫 번째 단계는 소비자에게 맞는 정확한 상품을 예측하고 기업의 이익을 최대화하는 마케팅 전략을 세우기 위해서 소비자의 구매에 이르는 행동을 파악하는 설문조사를 실시한다. 두 번째 단계는 설문조사를 통해 얻은 자료와 판매자의 지식을 바탕으로 베이지안 네트워크를 이용하여 모델을 구성한 후 변수를 감축하여 감축된 베이지안 네트워크를 최종적으로 구성한다. 세 번째 단계는 잠재변수에 영향을 미치는 변수들 즉, 감축된 베이지안 네트워크를 코호넨 맵을 이용하여 소비자의 구매행동 패턴을 분석한다. 전체 시스템 구성도를 그림 3과 같이 나타낸다.

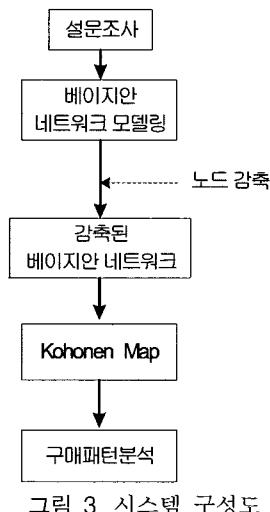


그림 3. 시스템 구성도

3.2 노드 감축

노드 감축은 특정 변수에 대한 확률분포를 계산하려고 할 때, 대상이 되는 변수를 X 라 하고, 그 외의 변수는 n 개 있으면 Y_i 라고 한다. X 와 Y 는 각각 $x_1, \dots, x_m, y_1, \dots, y_m$ 의 m 개의 상태 값을 가지는 확률변수라고 가정하면 X 의 사후확률은 X 이외의 변수 Y_i 에 대해서 있을 수 있는 모든 상태를 평균화하는 변화에 의해 식(4)과 같이 구해진다.*수식(4,5,6)수정*

$$\begin{aligned} P(X=x_j) &= \frac{P(x_j, Y_1, \dots, Y_n)}{\sum_{k=1}^m P(x_k, Y_1, \dots, Y_n)} \\ &= \alpha P(x_j, Y_1, \dots, Y_n) \end{aligned} \quad (4)$$

여기서 a 는 정규화 정수이다. $P(x_j, Y_1, \dots, Y_n)$ 는 다음과 같이 나타낸다.

$$\begin{aligned} P(x_j, Y_1, \dots, Y_n) &= \sum_{y^1 \in Y_1} \cdots \sum_{y^n \in Y_n} P(x_j | Y_p) \prod_{i=1}^n P(y^i | pa(Y_i)) \end{aligned} \quad (5)$$

식 (4)에서 부모를 가지지 않는 최상위의 변수를 Y_T 라고 하면 이 항은 다른 변수로부터 영향을 받지 않기 때문에 재귀적으로 전체 합에서 제외될 수 있으므로 식(4)은 다음과 같이 식(6)으로 변형될 수 있다.

$$\begin{aligned} P(X=x_j) &= \alpha \sum_{y \in Y_T} P(y) \\ &\quad \sum_{y^1 \in Y_1} \cdots \sum_{y^n \in Y_n} P(X | Y_p) \prod_{Y_i \neq Y_T} P(y^i | pa(Y_i)) \end{aligned} \quad (6)$$

식 (6)에 의해서 계산의 반복횟수를 줄일 수 있고 조건부 독립성을 이용하여 관련되지 않는 부분을 재귀적인 계산에서 제외시킬 수 있다. 위와 같이 베이지안 네트워크의 성질을 이용하여 적은 계산량으로 효율적으로 노드를 감축 할 수가 있다.

4. 모의실험

4.1 베이지안 네트워크를 이용한 노드 감축

디지털카메라 이용자의 구매성향을 분석하기 위해 설문조사를 실시하였다. 12문항으로 구성한 설문조사를 바탕으로 디지털 카메라를 구입할 때 제조회사에 따라 구매행동에 차이가 있는지 파악하고자 한다. 즉 잠재변수를 제조회사로 설정하여 베이지안 네트워크를 구축하여 모델링 했으며 CPT(Conditional probability Table) 값은 설문지자료의 응답 빈도분포에 따라 설정했다.

설문조사는 주로 디지털카메라를 많이 이용하는 20대와 30대 총 150명을 대상으로 실시했다. 12개의 항목을 대상으로 제조회사를 잠재변수로 하여 베이지안 네트워크를 이용하여 감축한 결과 3개의 항목이 선택되었다. 구축된 베이지안 네트워크에서 불필요한 노드를 감축시킨 결과 그림 4와 같은 네트워크 구조로 표현할 수 있다.

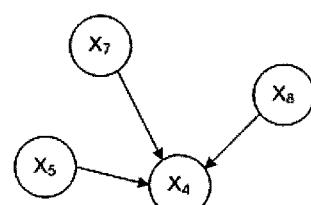


그림 4. 노드감축 후 베이지안 네트워크 구조

위 그림에서 X_4 는 잠재변수 제조회사(6가지)이며 X_5 , X_7 , X_8 은 각각 디지털카메라의 선택기준(6가지), 가격대(5가지) 그리고 화소 수(5가지)를 나타낸다. 베이지안 네트워크를 이용한 노드 감축으로 불필요한 노드를 제거한 후 구성된 베이지안 네트워크 구조를 바탕으로 비교사 학습기법인 코호넨의 SOM을 이용하여 동일한 패턴에 대해서 다른 형태의 출력력을 할 수 있는 출력패턴을 학습시켜 구매패턴을 분석하고 예측할 수가 있다.

4.2 SOM(Self-Organizing Maps)

베이지안 네트워크를 이용하여 감축한 자료를 대상으로 SOM의 입력패턴 수를 3개의 노드(선택기준, 가격대, 화소수), 클러스터 수를 6개의 노드(Canon, Olympus, Kenox, Nikon, Sony, Etc)로 구성하였다. 그리고 뉴런 수는 2차원 상에 20*20 매트릭스로 구성하였다. 뉴런의 초기 가중치(initial weight)는 0에서 1사이에 랜덤 값(random value)으로 초기화하였고, 초기 학습률(initial learning rate)은 0.999, 학습율 인수(learning rate factor)는 0.99로 설정하였다. 여기서 학습율 인수는 네트워크의 학습율을 매 단계마다 증가시킨다. 학습범위(learning radius)에서는 초기의 학습 범위는 4.0으로 학습범위 인수는 0.995로 설정하였고, 출력순서는 임의로 지정하였다.

실험에 사용된 150개의 데이터를 바탕으로 학습시킨 결과 학습 횟수가 1,444회일 때 최적의 상태를 얻을 수 있었다. 아래의 그림 5는 SOM의 최종 훈련 상태를 나타낸다. 여기서 Current Learning Rate는 초기의 학습율 값을 0.999, 학습 인수를 0.99로 초기화했을 때의 학습율을 의미하고, Current Learning Radius는 초기의 학습 범위를 4.0, 학습범위 인수를 0.995로 주어졌을 때의 학습 범위를 나타낸다.

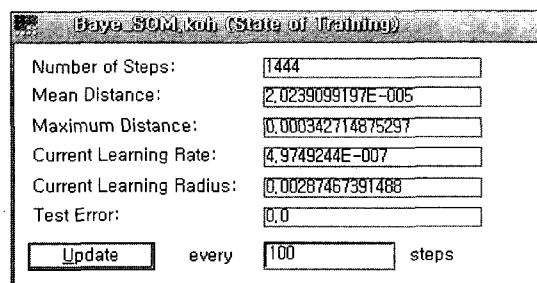


그림 5. SOM의 최종 훈련 상태

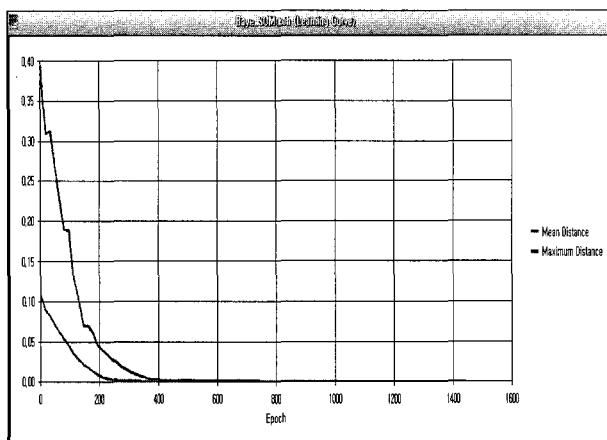


그림 6. 학습 곡선

그림 6은 학습곡선을 나타낸 것으로서 Mean Distance와 Maximum Distance는 각각의 모든 오브젝트들 간의 평균거리와 최대거리를 나타낸 것이다.

그림 7은 SOM을 2차원 평면 즉, 20*20으로 구성하였을 때의 총 뉴런 400개의 연결 가중치의 일부를 나타낸 것이다. 그림 7에서 첫 번째 열은 400개의 뉴런의 인덱스를 의미하고, 두 번째열은 출력력을 의미하는 잠재변수인 제조회사를 의미한다. feature5, feature7, feature8은 감축된 3개의 노드 즉 선택기준, 가격대, 화소수를 의미한다.

연결 가중치에 의해서 400개의 뉴런들이 잠재변수인 제조회사의 6개의 노드 값으로 출력된다. 그림 8은 최종적인 SOM 맵을 시각적으로 나타낸 것으로서 적절하게 학습이 되어 6개의 클러스터로 분류됨을 알 수가 있다.

| Bayes SOM Kohl (Neuron Weights) | | | | |
|---------------------------------|---------|----------------|----------------|----------------|
| | | feature5 [] | feature7 [] | feature8 [] |
| 100 | Nikon | 1.000 | 0.000 | 2.000 |
| 101 | Cannon | 2.044 | 4.046 | 3.000 |
| 102 | Cannon | 1.922 | 4.077 | 3.000 |
| 103 | Etc | 1.028 | 4.002 | 3.000 |
| 104 | Etc | 1.000 | 4.000 | 3.000 |
| 105 | Etc | 1.001 | 3.958 | 3.000 |
| 106 | Olympus | 1.000 | 3.076 | 3.000 |
| 107 | Olympus | 1.000 | 3.000 | 3.000 |
| 108 | Olympus | 0.964 | 3.000 | 3.005 |
| 109 | Etc | 0.195 | 3.000 | 3.028 |
| 110 | Etc | 0.007 | 2.998 | 3.003 |
| 111 | Etc | 0.180 | 2.824 | 3.002 |
| 112 | Olympus | 0.964 | 2.036 | 2.992 |
| 113 | Olympus | 1.000 | 2.000 | 2.991 |
| 114 | Olympus | 1.000 | 1.997 | 2.061 |
| 115 | Olympus | 1.001 | 1.504 | 2.251 |
| 116 | Kenox | 1.000 | 1.001 | 2.000 |
| 117 | Kenox | 1.000 | 1.000 | 2.000 |
| 118 | Kenox | 1.001 | 0.996 | 2.000 |
| 119 | Nikon | 1.123 | 0.320 | 1.999 |
| 120 | Nikon | 1.090 | 0.047 | 2.000 |

그림 7. 최종 가중치 변화량

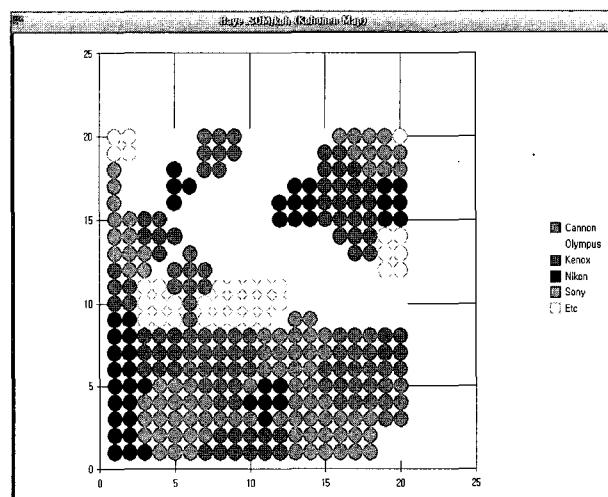


그림 8. 최종 SOM 맵

최종 실험 결과, 뉴런을 2차원 공간으로 20*20 매트릭스로 구성하고 뉴런간의 학습범위를 4로 설정하였을 경우 6개의 클러스터가 적절하게 분류됨을 알 수 있었다. 그림 9는 SOM의 최종 실행결과를 나타낸 것으로 사용자로부터 X_5 , X_7 , X_8 항목에 대한 질의에서 X_5 는 기능적인 면, X_7 는

| | Test_Output(1~150).dat | | | | | | | | | | | | | | | |
|----|------------------------|------------------|------------------|--------------------|---------------------|-------------------|------------------|-------------------|-----------------|---------------|-----------------|---------------|---------------|--------------|-------------|-----------|
| | feature5 [input] | feature7 [input] | feature8 [input] | Cannon [output[1]] | Olympus [output[2]] | Kenox [output[3]] | Sony [output[4]] | Nikon [output[5]] | Etc [output[6]] | Koh_Canon [I] | Koh_Olympus [I] | Koh_Kenox [I] | Koh_Nikon [I] | Koh_Sony [I] | Koh_Etc [I] | Error [I] |
| 13 | 1.000 | 3.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 14 | 1.000 | 3.000 | 2.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 15 | 1.000 | 2.000 | 2.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 16 | 3.000 | 1.000 | 5.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 17 | 2.000 | 2.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 18 | 2.000 | 3.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 19 | 3.000 | 3.000 | 2.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 20 | 1.000 | 3.000 | 2.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 21 | 1.000 | 2.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 22 | 3.000 | 3.000 | 2.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 23 | 2.000 | 1.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 24 | 2.000 | 3.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 25 | 2.000 | 3.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 26 | 2.000 | 2.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 27 | 1.000 | 2.000 | 2.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 28 | 4.000 | 2.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 29 | 2.000 | 3.000 | 3.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 30 | 3.000 | 4.000 | 5.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

그림 9. 최종 실행 결과

30에서 40만원, X_8 은 300에서 400만 화소라는 응답을 얻었으면 잠재변수 X_4 (제조회사)는 Olympus라는 결과를 얻을 수 있다.

5. 결론 및 향후 연구과제

소비자의 구매행동 패턴을 분석하기 위해서는 소비자의 구매에 따르는 심리상태나 내부 상태까지 깊이 분석하여야 한다. 소비자의 구매성향을 정확하게 파악하고 기업의 이익을 최대화하기 위해서는 마케팅 전략을 세우고 소비자의 구매에 따르는 행동을 조사할 필요가 있다. 판매자는 기업의 이익을 최대화하는 마케팅전략을 세우기 위해 소비자의 구매동향을 조사하는 설문조사를 실시하고 있다. 설문조사는 비교적 간단하고 대규모 조사도 실시하기가 용이하지만 문항이 많은 경우에는 응답자의 부담이 커서 응답을 회피하는 문제점과 어려운 문항에 대해서는 신뢰성이 떨어지는 문제점이 있다.

따라서 본 논문에서는 응답이 없는 데이터 즉, 관측이 곤란한 요소를 다루는 것이 가능하고 추측되는 가설의 확신도를 실제 데이터를 바탕으로 검증이 가능한 베이지안 네트워크를 이용하여 모델링 하였다. 판매자의 노하우와 소비자의 구매의식 조사 데이터를 바탕으로 베이지안 네트워크를 구성하였고 불필요한 속성을 가진 데이터를 제거하여 구매패턴을 분석하는데 정확도를 높일 수가 있었다. 감축한 자료를 대상으로 코호넨의 SOM을 이용하여 동일한 패턴에 대해서 다른 형태의 패턴을 학습시켜 소비자의 구매패턴을 예측할 수가 있었다. 모의실험을 실행한 후의 결과를 보면 감축 이후에 좋은 분류 결과를 얻을 수 있음을 알 수 있었다.

소비자의 구매패턴을 분석하는데 있어서 베이지안 네트워크는 노드 수가 많아지면 복잡해지고 정확도가 떨어지는 단점이 있으므로 이를 보완한 베이지안 네트워크에 대한 연구가 필요하다.

참 고 문 헌

- [1] 村上 外, "ベイジアンネットワークによる消費者行動分析", 電子情報通信學會, 2004.
- [2] 後藤秀夫, 市場調査ケーススタディ 改訂新版, 日本マーケティング教育センター, 1996.
- [3] David Heckerman, "A Tutorial on Learning Bayesian Networks", Technical Report MSR-TR-95-06, 1995.
- [4] Thomas Dean et al, Artificial Intelligence Theory and Practice, Addison Wesley, 1995.
- [5] Kohonen, T, Self-Organization Maps, Series in Information Science, vol. 30, Springer-Verlag Berlin Heidelberg New York, 1997.
- [6] Jensen, F. V., Bayesian Networks and Decision Graphs, Springer-Verlag Berlin Heidelberg New York, 2001.
- [7] Utsugi A., "Hyperparameter selection for self organizing maps", Neural Computation, Vol. 9, No. 3, 623-635
- [8] Neal, R. M., Bayesian Learning for Neural Networks, Springer-Verlag Berlin Heidelberg New York, 1995.
- [9] Remco R. Bouckaert, "Conditional Dependence in Probabilistic Networks", RUU-CS-92-34, 1992
- [10] Fabio Gagliardi Cozman, "Generalizing Variable Elimination in Bayesian Networks", Workshop on probabilistic Reasoning in Artificial Intelligence, Atibaia, Brazil, 2000
- [11] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.b. Bayesian Data Analysis, Chapman & Hall, 1995
- [12] Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2001

- [13] Stuart Russell, Peter Norvig, *Artificial Intelligence A Modern approach*, Prentice Hall, 2003
- [14] K.P. Murphy, Y. Weiss, M.I. Jordan, "Loopy belief propagation for approximate inference: An empirical study", *Proceedings of the Fifteenth conference on Uncertainty in Artificial Intelligence*, 1999
- [15] Gregory F. Cooper, "The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks", *Artificial Intelligence* v.42 no.2 pp393~405, 1990



손창식(Chang-Sik Son)

2000년 : 대구가톨릭대학교 전자정보
공학부(공학사)
2002년 : 대구가톨릭대학교 대학원
전자계산학과(이학석사)
2002년 ~ 현재 : 대구가톨릭대학교 대학원
컴퓨터정보통신공학부(박사과정)

관심분야 : 퍼지추론 및 튜닝시스템, 신경망, 감성처리
Phone : +82-53-850-2741
E-mail : iisman1@cu.ac.kr

저자 소개



황정식(Jeong-Sik Hwang)

2004년 : 대구가톨릭대학교 컴퓨터정보
통신공학부(공학사)
2004년 ~ 현재 : 대구가톨릭대학교 대학원
컴퓨터정보통신공학부(석사과정)

관심분야 : 인공지능, 확률 추론, 베이지안 네트워크
Phone : +82-53-850-2741
E-mail : icsman@cu.ac.kr



정환목(Hwan-Mook Chung)

10권 4호 참조

Phone : +82-53-850-2741
Fax : +82-53-850-2741
E-mail : hmchung@cu.ac.kr



피수영(Su-Young Pi)

2000년 : 대구가톨릭대학교 컴퓨터정보
통신공학부(공학박사)
2000년 ~ 현재 : 대구가톨릭대학교
실용컴퓨터 전담교수

관심분야 : 데이터마이닝, 신경망, 지능형에이전트, 유비쿼터
스 컴퓨팅 등
Phone : +82-53-850-3699
E-mail : sypi@cu.ac.kr