

# CRM에서 마코프 모델을 이용한 고객 성향분석에 따른 정보제공

김홍주\*, 이태경\*\*, 서영호\*\*  
동국대학교 전자계산학과

## Information Service by Customer Property Analysis using the Markov Model in the CRM

Hong-Ju Kim\*, Tae-Kyung Lee\*\*, Young-Ho Seo\*\*  
Dept. of Computer Science, Dongguk University\*

### 요약

CRM에서 고객에 대한 log등의 고객 정보를 통해 효율적인 서비스의 필요성이 강조되고 있다. 기존의 서비스는 대상 고객의 관심도와 성향 분석을 통한 것이라기 보다는 무조건적으로 제공되는 구체적이고 체계적인 지식이 결여된 상태이므로 고객의 요구에 정확한 정보의 제공이 어려웠다. 그러므로, 본 논문에서는 고객이 원하는 정보를 정확하게 제공하기 위해 고객이 필요한 정보를 자동적으로 수집, 분류할 수 있는 마코브 모델을 통해서 통계와 분석을 수행하여 고객 정보의 분류와 획득에 의한 정보 서비스를 제공하고자 한다.

### 1. 서론

최근 전자상거래 관련분야의 활성화에서 현재 고객에 대한 log등의 고객 흔적을 통해 효율적인 서비스의 필요성이 강조되고 있는 실정이다. 기존의 서비스는 대상 고객의 관심도와 성향 분석을 통한 것이라기 보다는 무조건적으로 제공되는 구체적이고 체계적인 지식이 결여된 상태에서 고객의 요구에 정확한 정보 제공이 어렵다. 이에 따라 고객이 설정한 관심 정보들 기반으로 생성된 프로파일이 변경되지 않는 한 서비스의 내용은 항상되지 못한다는 문제가 발생한다. 그 결과로 인하여 부정확한 정보를 그대로 제공하였을 경우 원하는 정보를 얻지 못한 고객은 효용성에 대해 신뢰하지 못한다. 따라서 먼저 고객의 관심 문서를 기반으로 구축된 고객 프로파일을 적용하여 여러 가지 정보를 제공함에 따라 효용성이 증가할 수 있을 것이다. 고객은 제공된 정보에 대한 피드백을 수행함으로써 프로파일은 더욱 구체적이고 정확하게 재구축된다. 고객 피드백에 의한 관심도에 대한 강화학습(reinforcement learning) 작업에 대한 구체적인 방법

을 제공하고자 한다. 이를 위하여, 고객이 원하는 정보를 정확하게 제공하기 위해 고객이 필요한 정보를 자동적으로 마코브 모델(Markov model)을 통해서 통계와 분석을 수행하여 고객의 정보의 분류와 획득을 할 수 있고 spam성 정보의 증가를 막을 수 있는 그에 따른 서비스를 제공하자는 의도이다[10][11].

본 논문의 구성은 2장에서 본 연구의 이론적 배경에 대해서 설명하고, 3장에서는 전체 개략 시스템 구성에 대해서 살펴본다. 4장에서는 향후 연구 방향을 검토한다. 마지막으로 5장에서는 결론을 기술한다[1].

#### 1.1 기존 연구

기존의 CRM에서의 성향분석연구인 E-CRM Nexus는 직접적인 질문/대답을 통한 고객선호도 분류, 집단화하고 규칙기반(rule base)을 사용한다. SAS CRM은 고객에 대한 수익성 분석과 신용 리스크 분석을 통해 고객을 세분화하고 고객 그룹별 DW을 이용하여 패턴 분석한다. 사용하는 기술은 Data Warehouse, Data Mining, OLAP을 사용하고 있다.

EC에서의 Agent 연구는 고객의 요구(정보, 제품, 서비스)와 고객에 대한 데이터, 요구, 습관들에 대하여 추론을 통해 고객의 요구를 적절히 충족할 수 있는 신속한 방법을 제공할 수 있도록 하고 있다.

단순 가격 비교에이전트인 BargainFinder, Shopbot은 자동으로 새로운 상점에 대한 규칙을 생성해 내며, 상품 정보를 추출하기 위해 귀납적 학습기법을 이용하고 있다. Jango는 단순한 가격정보 외에도 재고여부, 배달 가격과 방법, 시간, 기타 다른 여러 정보를 구매자에게 같이 제공하여 선택의 폭을 넓혀주고 판매자가 요구하는 사항들의 수용이 가능하지만, 고객 선호도파악이 불가능한 것이 큰 단점이라 할 수 있다.

1.2 연구 목적

본 논문의 목적은 이러한 정보제공에 따른 이탈고객 감소를 위한 고객 성향분석을 통해 고객이 원하는 서비스 제공가능을 가능케하고, 고객 정보의 통합과 SPAM성 서비스를 감소시키는데 있다.

1.3 연구 내용과 범위

마코브 모델을 통해서 고객 성향분석을 하고 고객 정보의 지식베이스화가 가능하도록 한다.

본 논문에서 쓰이는 자료로는 처음 고객이 기록한 프로파일, 로그인 시간, 구매 내역, 접속 경로와 그 접속한 사이트 콘텐츠가 될 수 있도록 한다.

연구의 범위는 마코브 모델에 적용시켜 성향분석을 하고 자료는 학습을 통해서 개인정보에 추가 또는 행동(action)이 이루어진다. 이 행동은 고객측면에서 맞춤 정보(E-mail, 개인화 페이지)를 제공하는 것까지 포함한다[4].

2. 이론적 배경

2.1 강화학습 이론

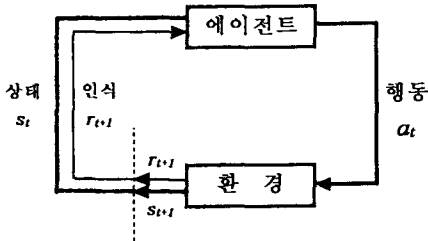


그림1. 에이전트-환경 인터페이스

그림1과 그림2는 에이전트-환경간의 상호작용을 나

타내고 있다.  $t$  ( $t=0,1,2,3,\dots$ )는 각 시간 관계(time step)를 나타내고,  $S$ 는 가능한 상태(state)들의 집합,  $A(S_t)$ 는 상태  $S_t$ 에서 적절한 행동(action)들의 집합을 나타낸다. 그리고  $r_{t+1} \in R$ 은 수치로 표시되며 인식된 자료에 해당된다.

$S_{t+1}$ 는 새로운 상태(다음 상태)를 찾는 것을 나타내고 있다.

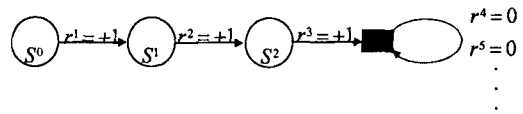


그림2. 에이전트-환경 관계의 상태전이

그림2에서  $S_t$ 는 시간  $t$ 에서의 상태이며,  $r_4, r_5, \dots$ 에서 0 reward 들만 생성되는 상태를 흡수상태라고 한다. 여기서  $T$ 는 맨 마지막 시간 관계를 나타내고 이때  $T$ 의 값은 3이 된다[7][12].

2.2 마코브 성질

이산 확률 과정은  $\{X(n), n=0, 1, \dots\}$ 에서 모든 상태  $i_0, i_1, i_{n-1}, i, j$ 와  $n$ 에 대해서

$$\begin{aligned}
 P(X(n+1)=j | X(0)=i_0, X(1)=i_1, \dots, X(n-1)=i_{n-1}, X(n)=i) \\
 &= P(X(n+1)=j | X(n)=i) \\
 &= P_{ij} \dots\dots\dots\text{식①}
 \end{aligned}$$

상태  $X_0, X_1, \dots, X_{n-1}$ 와  $X_n$ 이 주어졌을 때의 다음 상태  $X_{n+1}$ 의 조건부 분포는 과거의 상태와는 독립이며 오직 현재상태에만 의존함을 의미한다.

식①이 성립하는 성질을 마코브 성질이라고 말하며, 이 확률 과정  $\{X(n), n=0, 1, \dots\}$ 은 마코브 연쇄(Markov chain)라 한다[4][5][6].

2.3 마코브 연쇄

시스템의 상태가 시간에 따라 확률적으로 변해가는 과정을 확률 과정이라고 하고, 이 확률과정 중에서 시스템의 다음 상태가 현재의 상태만 주어지면 과거의 역사와는 무관하게 결정되는 성질을 가지는 확률과정을 마코브 과정(Markov process)이라 한다. 이런 마코브 과정에서 시간의 변화를 연속적으로 고려하지 않고 이산적인 경우로 고려한 경우를 마코브 연쇄(Markov chain)라 한다. 본 연구에서는 이산적인 경우를 사용하므로 마코브 연쇄를 사용한다.

$$P_{ij} \geq 0, i, j \geq 0; \sum_{j=0}^{\infty} P_{ij} = 1, \dots \text{식②}$$

$$i = 0, 1, \dots$$

그러므로 식①에서  $i$ 에 의해 다음 상태  $j$ 가 결정된다. 식②에서  $P_{ij}$ 는 현재 상태가  $i$ 일 때 다음 상태  $j$ 로 transition될 확률( $P_{ij}$ 는 전이 확률)이 된다[3][4][5][6].

2.4 마코브 연쇄 적용 예

(예) 이 번달에 A라는 사람이 상품 B를 구입했을 때 두 달 후에도 구입할 확률은?  
(전체의 80%가 상품을 구입한다고 가정)

이번달\다음	구입	비구입
구입	0.9	0.1
비구입	0.05	0.95

표1. 월별 상품 B의 구입 변화 확률

표1에 대한 상태전이도(state transition diagram)를 나타내면 그림3이 될 것이다.  
(s1 = 구입, s2 = 비구입)

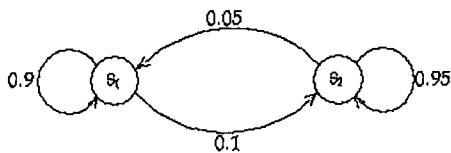


그림3. 상태전이도

그림3에 대한 전이행렬(transition matrix)을 나타내면 아래와 같이 유도할 수 있고,

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

이 것을 확률트리에 대해서 2달후 구매율을 예측하면 그림4와 같이 표현된다.

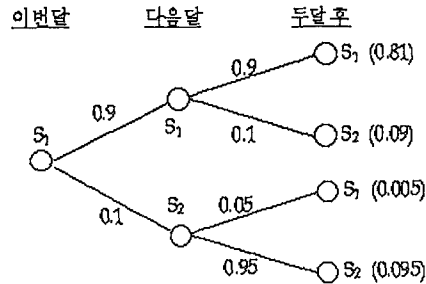


그림4. 확률트리에 의한 2달후 구매율 예측

이 번달에 A라는 사람이 상품 B를 구입했을 때 두 달 후에도 구입할 확률 :  $0.81 + 0.005 = 0.815$ , 두 달 후에도 구입하지 않게 될 확률 :  $0.09 + 0.095 = 0.185$  이라는 결과를 얻을 수 있다.

다음의 전이행렬에 의한 직접 계산을 하는 방법이다. 두 달후의 확률변화는 두 단계의 전이가 반복해서 일어나는 경우이므로 전이행렬을 곱하면 된다. 즉, 두 단계 전이행렬은 아래와 같다.

$$P^2 = \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix} = \begin{bmatrix} 0.815 & 0.185 \\ 0.0925 & 0.9075 \end{bmatrix}$$

만약 80%가 이번달에 구매를 했다면 두 달후의 전체 구매 비율은,  $0.8 \times 0.815 + 0.2 \times 0.0925 = 0.6705$ 이므로 67.05%로 예측할 수 있다.

3. 전체 개략 시스템 구성도

이 장에서는 개략적인 전체 시스템의 구성도를 설계한다. 개인화 에이전트(Personalization Agent), 모니터링 에이전트(Monitoring Agent), 고객 성향 분석 에이전트(Analysis Agent), 구매 도우미 에이전트(Assistant Agent)로 시스템을 구성한다.

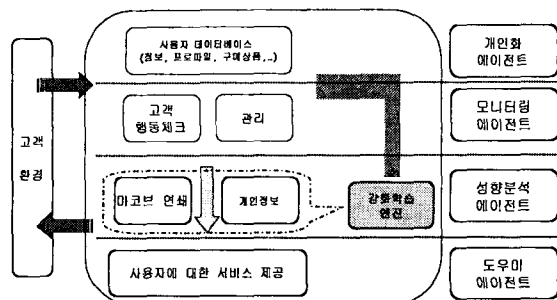


그림5. 전체 시스템 구성도

① 개인화 에이전트(Personalization Agent)

고객이 처음 로그인할 때 고객의 프로파일을 보고, 다음 방문시 인증이 되면 고객 프로파일을 참조하여 광고와 상품을 제시해 주는 고객 내용(Customized content)을 생성해 주는 역할을 한다.

② 모니터링 에이전트(Monitoring Agent)

고객이 방문한 경로와 검색한 키워드, 선택한 광고, 상품 구매등 웹 기반의 모든 행위를 체크하여 Analysis Agent에게 전달한다.

③ 고객 성향 분석 에이전트(Analysis Agent)

고객프로파일에 저장되어 있는 고객의 데이터와 모니터링 에이전트로부터 받은 내용을 분석하여 개인화 에이전트가 그 고객의 취향을 알 수 있도록 전달해 주고 마코브 연쇄를 사용하여 고객이 선호하는 성향을 분석한다. 분석 후 개인화 에이전트에게 전달한다. 모든 작업이 끝나면 모니터링 결과와 구매결과를 분석하여 확률치를 두어 고객 프로파일에 저장한다.

④ 구매 도우미 에이전트(Assistant Agent)

고객의 모든 구매 행위를 보조하고 탐색기능 역할을 포함한다. 개인화 에이전트가 요구하는 상품의 query를 받아 제시하고, 고객 성향 분석 에이전트로부터 전달받은 상품이나 광고를 고객에게 정보제공을 한다.

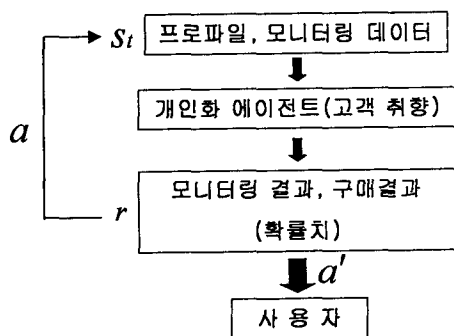


그림6. 성향 분석 과정

그림6은 본 논문에서 가장 핵심이 되는 고객성향의 분석과정을 도식화한 것이다. 고객에 대한 프로파일등의 자료와 모니터링 자료를 개인화 에이전트에게 고객에 대한 성향분석을 하고 그 결과를 다시 전달하고 고객에 대한 확률치를 저장해 둔다[7][11].

4. 향후 연구 방향

향후 연구로는 고객 정보에 대한 효율적인 관리에 대한 다른 방안의 연구와 방대한 고객 정보의 저장공간 효율적인 처리 문제, 다른 연계 분야와의 접목을 연구가 있어야 할 것이다. 고객 입장에서는 E-mail의 spam성이 아닌 다른 유용한 서비스에 대한 구상안이 필요하다.

5. 결론

본 논문은 이탈 고객 감소를 위한 고객 정보의 통합과 고객 성향분석을 통해 고객이 원하는 서비스제공 가능함으로써 앞으로 SPAM성 서비스 감소와 CRM분야에서의 수익성 증대에 도움이 될 것이다.

[참고문헌]

[1] 임양원, 권기훈, 임한규, "서비스 엔진을 이용한 웹 기반 메일 에이전트 시스템의 설계 및 구현", 한국정보처리학회 논문지, 제7권 제2호, 2000.  
 [2] 서두성, "의사결정나무의 평가와 응용", 고려대학교, 석사학위논문, 1998  
 [3] 강정혁, "추이확률의 추정을 위한 확장된 마코브 연쇄 모형", 한국경영과학회지, 제10권 제2호, 1993.10  
 [4] 최기현, 응용확률론 입문, 자유 아카데미, 1991  
 [5] Balkhi, Z. T., Benkherouf, L. , "A markov decision processes formulation for the linear search problem", 한국경영과학회지, 제19권 제1호, 1994  
 [6] 윤복식, "마코프 누적 프로세스에서의 확률적 콘벡스성", 한국경영과학회, 1991  
 [7] Richard S. Sutton and Andrew G. Barto, Reinforcement Learning, 1999  
 [8] 김종호, 마르코프.체인, 동국대학교출판부, 1987  
 [9] 최진갑, 확률과정론 입문, 경문사, 1992  
 [10] 아사오카 토모오, CRM에서 CRE로, 영진Biz.com, 2000  
 [11] 한국NCR CRM컨설팅 그룹, CRM의 이론과 실제, 2001