

온라인 추천 서비스를 위한 클릭스트림 데이터를 활용한 구매확률 예측 방안

A Study of Predicting Consumer's Purchase Intention Using Clickstream Data for On-line Recommendation Services

김종우(Jong Woo Kim)*, 한송이(Song Yi Han)**, 이홍주(Hong Joo Lee)***,
오재훈(Jae Hoon Oh)****

kjw@hanyang.ac.kr, hsy1001@paran.com, hongjoo@catholic.ac.kr,
jaehoon@nethru.co.kr

초 록

온라인 추천 서비스를 위해서는 고객이 온라인 상에서의 행동의 추적과 이를 기반으로 한 추천 활동이 필요하다. 구체적으로 인터넷 상점에서 온라인 추천을 위해서는 사전적으로 고객의 실시간 행동 데이터로부터 고객의 의도를 파악하는 것이 필요하다. 본 연구에서는 고객의 클릭스트림(Clickstream) 데이터를 이용하여 고객의 구매확률을 예측하기 위한 방안들을 제시하도록 한다. 또한 인터넷 서점의 실제 클릭스트림 데이터를 활용하여 제시한 방법의 유용성을 검토하도록 한다.

1. 서론

인터넷의 발달은 전통적인 시장이 갖는 시간적, 공간적 제약을 극복할 수 있는 해결책을 제시하였으며, 결국 전자상거래라는 새로운 유형의 거래 형태를 등장 시키는 배경이 되었다. 전자상거래의 활성화는 소비자의 욕구에 대한 즉각적인 대응의 중요성을 증가시켰다. 따라서 대중적인 수준이 아닌 개인적인 수준에서의 분석을 통해 잠재

적인 고객을 가려내거나, 구매 의사결정 지원을 위한 추천(recommendation) 서비스와 같은 연구가 점차 활성화 되고 있다 [1,5].

기존의 추천 서비스와 달리 온라인 추천 서비스는 고객의 실시간 데이터를 활용하여 추천을 성과를 향상시키는 데 중점을 두고자 한다[4]. 따라서 온라인 추천 서비스를 위해서는 고객행동의 실시간 추적과 이를 기반으로 한 추천 활동이 필요하다. 그러나

본 연구는 서울시 산학연 협력사업(JP100080)의 연구비 지원으로 이루어 졌음

- * 한양대학교 경영대학 경영학부 교수
- ** 한양대학교 대학원 경영학과 석사과정
- *** 가톨릭대학교 경영학부 조교수
- **** ㈜넷스루 연구소장

현재까지의 추천 기술들은 대부분 현재 시점의 고객 행위보다는 과거 시점의 누적 고객 행위들을 기반으로 추천을 수행한다. 따라서 본 연구에서는 온라인 추천 기술의 일환으로 인터넷 상점에서 고객의 실시간 행동 데이터로부터 고객의 구매 의도를 파악하고자 한다. 본 연구에서는 고객의 클릭스트림(Clickstream) 데이터를 이용하여 고객의 구매확률을 예측하기 위한 방안들을 제시하도록 한다. 또한 인터넷 서점의 실제 클릭스트림 데이터를 활용하여 제시한 방법의 유용성을 검토하도록 한다.

2. 관련연구

2.1 온라인 추천 서비스

고객에게 어떤 상품을 권할 것인가는 현재의 판매동향, 대상 고객의 기호와 감성, 지불 능력 등을 고려해야 하는 중요한 의사결정 문제라 할 수 있다[3]. 온라인 추천 서비스는 고려해야 하는 요건이 다양함에 따라서, 문제를 해결하는 방법 또한 여러 가지 측면에서 연구되고 있다.

대표적인 기존의 추천 기술로는 협업 필터링, 내용기반 추천, 규칙기반 등이 있다 [2,9]. 협업필터링은 선호도가 유사한 사용자들로부터 얻은 정보에 따라 사용자의 관심사를 자동적으로 예측하는 추천방법이다 [8]. 내용기반 추천방법은 각 속성 정보와 사용자의 선호 정보를 비교하여 예측 선호도가 높은 상품을 선별하여 추천하는 방법으로 선호 정보를 어떻게 효과적으로 학습할 것인가에 대해 초점을 맞추고 있다 [9,10]. 규칙기반 추천은 연관 규칙을 생성하고 이를 추천서비스에 활용하는 것으로, 방문경로분석, 장바구니분석을 통해 구매되는 상품의 관련성을 찾아내는 것이다.

2.2 클릭스트림 데이터

클릭스트림 데이터는 사용자의 웹 사이트 검색 이동경로가 기록된 것으로, 경로를 추적하여 사용자의 이동경로간의 관계를 파악하는데 활용된다[5,6]. 사용자의 경로가 실시간으로 기록된 데이터이므로 방문의 목적을 구분하거나 이동경로에 대한 분석 및 예측에 활용된다. 본 연구에서는 패널 집단의 웹사이트 방문 데이터를 수집하는 회사를 통해 국내 한 인터넷 서점의 클릭스트림 데이터의 일부를 수집하였다.

<표 1> 웹 페이지 분류 기준

구분	설명
홈	웹 사이트 메인 페이지
계정	개인정보 관리 및 정보 조회
검색	상품 검색
분류	분류기준에 따른 각 분야별 분류 페이지
상품	상품에 대한 소개 및 정보 제공
정보	이용에 대한 정보 및 고객센터
관측	관측 목적의 페이지
소셜	소셜(Social) 공간
보관	상품 보관 페이지
카트	장 바구니
주문	주문

3. 분석 방안

3.1 분석 데이터

인터넷 상점의 페이지는 몇 가지 유형의 페이지로 구분할 수 있다. 본 연구에서는 <표 1>과 같이 11가지 유형의 페이지로 구분하였다. 클릭스트림 데이터는 명확한 세션 구분이 없으므로 이에 대한 임의적인 결정이 필요하다. 본 연구에서는 행위의 간격이 20분 이내의 것을 한 세션으로 구분하였다. <표 2>은 예시적으로 5개의 세션에 대한 정보를 보여준다. 즉, 첫 번째 세션의 경우, 처음에 홈페이지를 방문하고, 2 차례 검색 페이지를 방문하고, 3 차례 상품 페이지를 방문하는 식으로 고객 행위가 이루어

졌음을 알 수 있다.

가까운 페이지에 가중치를 준다.

3.2 구매 확률 예측 방안

본 연구에서는 다음과 같이 4가지 방법을 활용하여 고객이 해당 세션 내에서의 구매 확률을 실시간으로 예측하도록 하였다.

방안 1: 1-State Model

각 유형에 속한 페이지를 방문했을 때의 구매확률을 구하여, 지금까지 방문한 페이지들의 구매확률을 평균하여 구매확률을 예측한다. 상태 전이를 고려하지 않았으므로 1-State Model이라고 부른다.

방안 2: 1-State with Weight Model

상태 전이를 고려하지 않는 것은 방안 1과 동일하나, 시간적 측면을 고려하여, 현재에 가까운 페이지에 가중치(Weight)를 주어서 구매확률을 구한다.

방안 3: 2-State Model

상태 전이를 고려하여, 한 페이지 유형에서 다른 페이지 유형으로 전환하였을 때 구매확률을 구하고, 1-State Model과 동일하게 포함된 상태 전이의 구매확률을 평균하여 구매확률을 구한다.

방안 4: 2-State with Weight Model

방안 3과 동일하게 2-단계까지의 전환 과정을 반영하나, 방안 2와 같이 현재에

4. 분석 결과

클릭스트림 데이터로부터 만들어진 세션은 총 1842개이며, 이 중에 한 페이지 방문만을 포함하는 세션들을 제외하여 실험에 사용하였다. 가공된 세션 데이터는 훈련 데이터 집합과 테스트 데이터 집합으로 임의로 50:50으로 구분하였다. 훈련 데이터 집합을 사용하여, 페이지 유형별 구매확률과 상태전이 시의 구매확률을 구하였고, 테스트 집합을 활용하여 4가지 예측 방법의 성과를 비교하였다.

<그림 1>은 예시적으로 구매한 특정 세션의 예측 구매확률의 변화를 보여준다. 4가지 방법 중에 상태전이를 고려한 방안 3와 방안 4가 더 좋은 성과를 보이는 것을 알 수 있다. 하지만 시간적인 측면을 고려한 방안(방안 2, 방안 4)이 그렇지 않은 방안(방안 1, 방안 3)과 차이는 크지 않은 것을 알 수 있다.

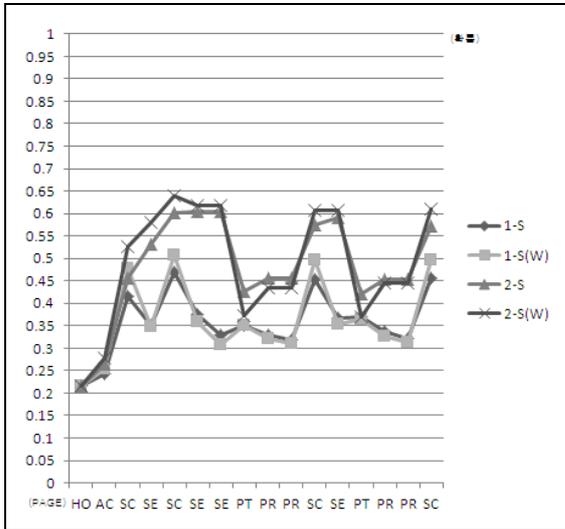
<그림 2>는 구매하지 않은 특정 세션에 대한 구매확률에서의 예측값 변화를 예시적으로 보여준다. 이 세션에 대해서는 4가지 방안으로 확률이 크게 차이가 나지 않는 것을 알 수 있다. 하지만, <그림 1>과 <그림 2>의 확률의 변화를 비교해보면, 제시한 방안을 활용하여 구매 여부를 고객이 3, 4

<표 2> 정제된 데이터의 예

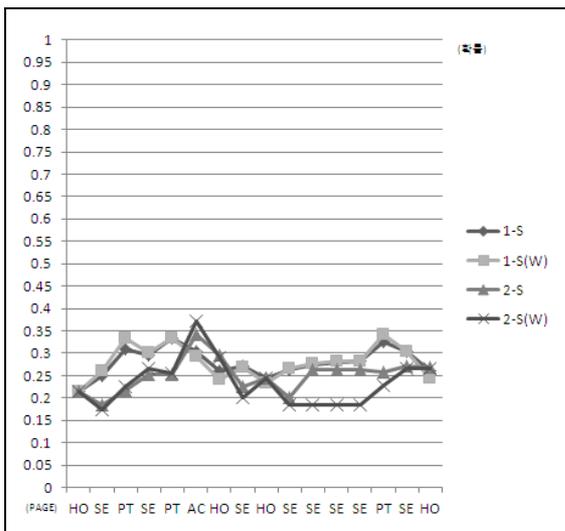
SESSION	PAGE
1	HO SE SE PT PT PT CA SE PT SE PT SE SC
2	HO SE SE PT PT PT PT SE SE SE SE PT
3	OT HO PR SE SE PT HO
4	HO PR AC SC SC SC SO SO HO SE SE PT SE PT PT SC HO AC SE PT SC SE SE SE SC
5	PT SC PT CA CA CA SC CA CA CA SC AC AC AC AC AC AC AC OD

Home=HO; Account = AC; Search = SE; Category = CA; Product = PT; Information = IN; Promotion = PR; Social = SO; Shopping Cart = SC; Order = OD; Other sites = OT.

페이지를 방문한 후에 어느 정도 예측 가능함을 보여준다.



<그림 1> 구매 세션의 구매확률 예측 예



<그림 2> 비구매 세션의 구매확률 예측 예

5. 결론

본 연구에서는 인터넷 상점에서 실시간 고객의 클릭스트림 데이터 분석을 통해 구매확률을 예측하는 4가지 방안을 제시하였다. 실제 인터넷 서점의 데이터를 통해서 실험해본 결과 상태전이를 고려한 경우가 그렇지 않은 경우보다 더 나은 성능을 보임을

알 수 있었다. 향후 연구 과제로는 2-단계 이상의 상태 전이를 고려하는 방법으로의 발전이 필요하고, 보다 빠른 시점에서 구매 여부를 판단할 수 있는 개선된 방법을 고안하는 것이 필요하다.

참고문헌

- [1] 김정욱, 주형진, "사용자 특성이 인터넷 쇼핑을 이용에 미치는 영향에 관한 실증적 연구", 한국경영과학회지, 제27권, 제4호, pp.149-165, 2002.
- [2] 이홍주, "클릭스트림 데이터를 활용한 전자상거래에서 상품추천이 고객 행동에 미치는 영향 분석", 한국경영과학회지, 제33권, 제3호, PP. 59-76, 2008
- [3] 임치환, 정규용, "온라인 추천 서비스를 위한 감성 기반 웹 에이전트 개발", 대한인간공학회지, 제23권, 제3호, pp. 1-12, 2004.
- [4] Adomavicius, G. and A. Tuzhilin, "Personalization Technologies: a Process-oriented Perspective", Communications of the ACM, Vol.37, No.10, pp. 83-90, 2005.
- [5] Alan L. Montgomery, Shibo Li, Kannan Srinivasan, and John C. Liechty, "Modeling Online Browsing and Path Analysis Using Clickstream Data", Marketing Science, Vol.23, No.4, pp. 579-595, 2004.
- [6] Bucklin, J.M. Lattin, A. Ansari, S. Gupta, D. Bell, E. Coupey, J.D.C. Little, C. Mela, A. Montgomery, J. Steckel "Choice and the Internet: From Clickstream to Research

Stream" Marketing Letters, Vol.13,
No.3, pp. 245–258, 2002

- [7] M.W. Kim, E.J. Kim, "Performance Improvements in Collaborative Recommendation Using Multi-Layer Perceptron", LNCS, vol.4234, pp. 350–359, 2006.
- [8] Konstan, J.A., B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News", Communications of the ACM, Vol.40, No.3, pp. 77–87, 1997.
- [9] John S. Breese, David Hecherman, Carl Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, USA, 1998.
- [9] Justin B. and Thomas H., "Unifying Collaborative and Content-Based Filtering", Proceedings of the 21st International Conference on Machine Learning, 2004.
- [10] Konstan, J.A., B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News", Communications of the ACM, Vol.40, No.3, pp. 77–87, 1997.