대한산업공학회/한국경영과학회 2002 춘계공동학술대회
한국과학기술원(KAIST) 2002년 5월 3일-4일

전자상거래 개인화 추천을 위한 다차원척도법의 활용

Application of Multidimensional Scaling Method for E-Commerce Personalized Recommendation

김종우*, 유기현*, Robert F. Easley**
{jwkim, khyoo94}@stat.cnu.ac.kr, Robert.F.Easley.2@nd.edu
* 충남대학교 통계학과
** Management Department, Mendoza College of Business, University of Notre Dame, USA

Abstract

In this paper, we propose personalized recommendation techniques based on multidimensional scaling (MDS) method for Business to Consumer Electronic Commerce. The multidimensional scaling method is traditionally used in marketing domain for analyzing customers' perceptional differences about brands and products. In this study, using purchase history data, customers in learning dataset are assigned to specific product categories, and after then using MDS a positioning map is generated to map product categories and alternative advertisements. The positioning map will be used to select personalized advertisement in real time situation. In this paper, we suggest the detail design of personalized recommendation method using MDS and compare with other approaches (random approach, collaborative filtering, and TOP3 approach).

1. 서론

고객 맞춤(customization) 또는 개인화(personalization) 서비스는 인터넷 상점이나 정보 서비스 제공자의 주요한 성공 요인의 하나로 인식되고 있다[2,4,10,11,15]. 이러한 서비스를 하기 위한 상품 추천 기술은 고객입장에서는 원하는 상품을 찾기 위한 노력과 시간을 줄여주고, 인터넷 상점 입장에서는 고객의 충성도(loyalty) 증대, 광고 수익의 증대, 타겟 홍보의 이익을 가져다 줄 수 있다. 현재 상용화된 추천 시스템들은 주로 협동적 필터링(collaborative filtering)이나 규칙기반(rule-based) 기법을 활용하고 있고, 이외에도 선호도 점수법(Preference Scoring), 신경망, 베이지만 네트워크, 계층 회귀분석 등 다양한 기법들의 적용이 연구되고 있다[1,3,8,11,12,13,16].

하지만 기존의 마케팅 이론들을 접목한 추천기술에 대한 연구들은 상대적으로 소홀한 형편이다. 본 논문에서는 다차원척도법(Multidimensional Scaling: MDS)에 기반을 둔 상품 추천 기술을 제시한다. 학습 대상 고객들의 구매 데이터를 기반으로 집단화 하고, 다차원 척도법을 활용하여 고객 집단과 대안 상품 광고를 이차원 평면에 투영한다. 이렇게 생성된 포지셔닝 맵을 활용하여, 인터넷 상점의 고객에게 실시간으로 개인화된 광고를 제공한다. 본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대하여 검토하고, 3장에서는 다차원척도법 기반의 상품 추천 방안을 제시한다. 4장에서는 제시한 방법의 타당성을 검토하기 위한 실험 데이터에 대한 분석 결과를 제시한다. 5장에서는 결론을 제시한다.

2. 관련연구

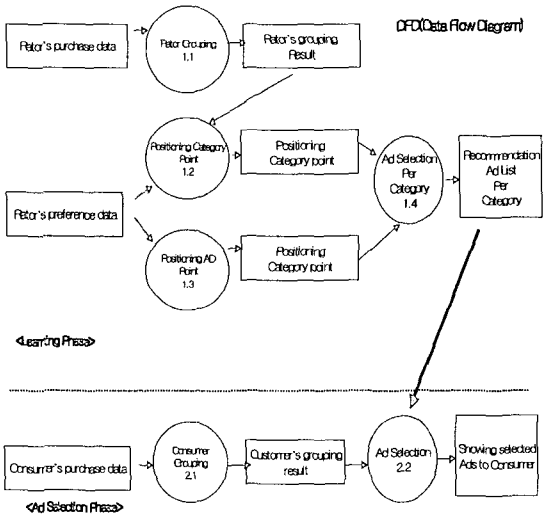
2.1 상품추천기술

상품추천을 위해서 가장 많이 활용되고 있는 협동적 필터링 기법은 고객의 광고에 대한 평가 정보들을 기반으로 하여, 유사한 선호도를 갖는 고객들의 평가를 활용하여 개인화된 광고를 제시한다[8,9,12]. 하지만 협동적 필터링은 원래 인터넷 뉴스의 그룹필터링을 위해서 고안된 기법으로 상품 필터링을 위한 특화된 기법은 아니며, 이론적인 배경이 빈약한 약점을 가지고 있다. 이외에도 인공지능 부분

에서 활용되고 있는 매칭기법을 활용하여 상품을 추천하는 규칙 기반 추천 기술 등이 있다[13,15].

2.2 다차원척도법

다차원척도법은 경영학, 정치학, 경제학, 심리학, 사회학, 지리학, 교육학 등 다양한 분야에서 활용되고 있으며, 특히 마케팅 분야에서 포지셔닝 맵(Positioning Map)을 구성하는데 주로 활용되고 있다[5,7,8,14,17,18]. 포지셔닝 맵을 구성하기 위한 수단으로서의 다차원척도법은 응답자의 평가에 의해 중요한 인지차원을 객관적으로 측정할 수 있으며, 이러한 평가차원에 대해 상표와 소비자를 동일공간에 나타낼 수 있어 유용한 정보를 제공한다. 따라서 마케팅 분야에서 다차원척도법을 이용한 상표 인지도와 선호도를 분석하는 유용한 도구로 활용되고 있다.



[그림 1] MDS 기반 추천 기법

3. MDS 기반 추천기법

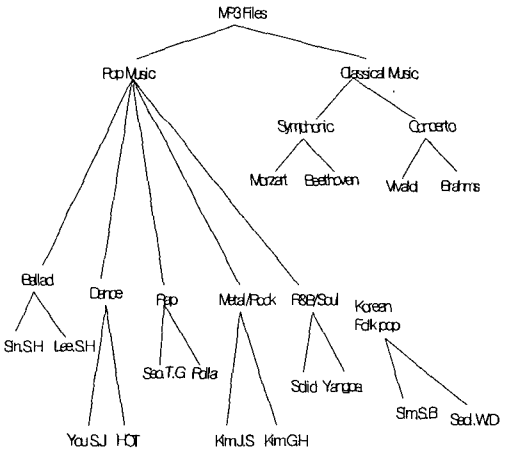
3.1 개요

본 논문에서 제시하는 MDS 기반 추천기법은 크게 (1) 학습단계와 (2) 광고 선정 단계로 이루어진다. 학습단계에서는 학습대상 고객의 구매정보를 기반으로 학습대상 고객을 그룹핑(rator grouping)(1.1), 카테고리 포지셔닝(positioning category point)(1.2), 광고 포지셔닝(positioning ad point)(1.3)을 하여 각각의 카테고리에 할당될 대안 광고들을 선정(AD selection per Category)(1.4) 한다.

3.2 학습단계

3.2.1 학습대상 고객 그룹핑

학습대상 고객의 구매 데이터를 기초로 학습대상 고객을 그룹핑 한다. 고객 그룹핑을 위해서는 사전에 정의된 상품군에 대한 계층구조가 정의되어 있어야 한다. 예를 들어, MP3 음악 파일에 대한 상품군의 계층구조가 [그림 2]과 같이 정의될 수 있다. 학습대상 고객들을 그룹핑 하기 위해서는 그룹핑을 하기 위한 수준(level)을 결정해야 한다. 수준을 2로 한 경우에는 팝음악, 클래식음악 두 개로 고객 그룹핑이 이루어지고, 수준을 3으로 결정한 경우에는 8개(발라드, 댄스, 랩, 메탈/락, R&B, 한국가요, 심포니, 콘체르토)로, 수준을 4로 결정한 경우에는 16개로 고객 그룹핑이 이루어진다. 그룹핑 수준이 결정되면, 학습대상 고객의 그룹핑은 고객의 구매 데이터를 기반으로 이루어 진다. 고객의 상품별 구매데이터를 상품군의 구매 데이터로 매핑하기 위해 상품-상품군에 대한 소속 관계가 필요한데, 이를 위해 상품-상품군 소속 행렬(D)를 다음과 같이 정의한다.



[그림 2] MP3 음악 파일에 대한 상품군 계층구조의 예

Definition 1
상품-상품군 소속 행렬 **D**는 다음과 같이 정의된다.

$$D_{l,m} = \begin{cases} 1 & \text{상품 } m \text{이 상품군 } l \text{에 속한 경우} \\ 0 & \text{그렇지 않은 경우} \end{cases} \quad \text{(식 1)}$$

여기서, $m \in \mathbb{M}$(상품 집합), $l \in \mathbb{L}$(상품군 집합)이다.

고객의 특정 상품의 구매 여부에 대한 정보도 행렬로 정의하면 다음과 같다.

Definition 2
고객 상품 구매 행렬 B^L는 다음과 같이 정의된다.

$$B^L_{m,j} = \begin{cases} 1 & \text{상품 } m \text{을 고객 } j \text{가 구매한 경우} \\ 0 & \text{그렇지 않은 경우} \end{cases} \quad \text{(식 2)}$$

여기서, $j \in \mathbb{J}$(학습대상 고객 집합)이다.

행렬 $\mathbb{D}$, B^L을 활용하여, 학습대상 고객 집합 $\mathbb{J}$를 구매가 많은 상품군을 기준으로 그룹핑할 수 있다. 다음 Definition 3은 이를 위한 식이다. 두 개의 상품군에 대하여 구매 횟수가 동일한 경우에는 평균 구매 횟수가 적은 상품군에 할당되도록 정의하였다. 즉, 두개의 상품군에 대한 해당 고객의 구매횟수가 동일한 경우에는 두 상품군 구매 비율(=상품군에 속한 상품에 대한 구매 횟수/전체 구매 횟수)이 적은 쪽으로 할당된다.

Definition 3
각 상품m의 구매 비율 $\mathbb{W}$와 학습대상 고객의 상품군 점수행렬 $\mathbb{H}$는 다음과 같이 정의된다.

$$W_m = \frac{\sum_{k=1}^{R} B^L_{m,k}}{R} \quad \text{(식 3)}$$

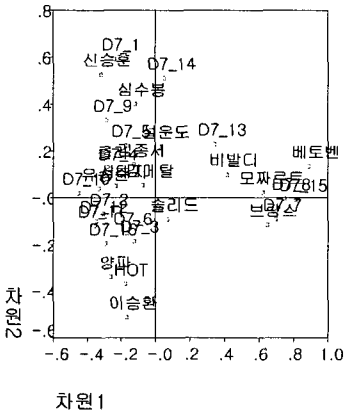
$$H = DB^L - DW \quad \text{(식 4)}$$

학습대상 고객 j는 다음과 같이 상품군 l'에 할당된다.

$$j \in G(l'),\ H_{l',j} = MAX_{\text{모든 } l \text{에 대하여}}(H_{l,j}) \text{인 경우} \quad \text{(식 5)}$$

3.2.2 포지셔닝 맵(positioning map) 작성

학습대상 고객의 대안 광고에 대한 선호도 데이터를 기초로 해서, 다차원척도법(MDS)을 이용한 포지셔닝 맵을 생성한다. 포지셔닝 맵에는 3.2.1에서 생성된 고객그룹 $G(l)$에 대한 위치와 대안광고들의 위치가 표시된다. [그림 3]은 수준을 4로 하여 16개의 고객 그룹을 정하고, 각 고객 그룹과 대안 광고에 대한 포지셔닝 맵을 작성한 것이다. D7_1, D7_2 등이 대안 광고에 대한 위치를 표시한다. 포지셔닝 맵에 위치를 기준으로 하여 각 학습대상 고객 그룹 $\mathbf{G}(l)$에 대안광고들이 선정된다. 즉, 각 고객 그룹에 해당하는 위치에 근접하는 대안광고들이 차례로 선정된다.



[그림 3] 포지셔닝 맵 작성 예

Definition 4
각 학습대상 고객 그룹 l에 대한 n개의 선정 광고 집합 $\mathbf{S}(l)$은 다음과 같이 정의된다.

$\mathbf{S}(l)$ = {a| DISTANCE(point($\mathbf{G}(l)$),point(a)) 값이 작은 n개의 a} (식 6)

여기서, n은 선정하고자 하는 광고의 수이고, point($\mathbf{G}(l)$), point(a)는 각각 포지셔닝 맵에서 $\mathbf{G}(l)$, 대안 광고 a의 좌표값이다.

3.3 광고 선정 단계

3.3.1 해당 고객의 고객 그룹에 할당

광고를 제시하고자 하는 고객에 대하여 3.2.1에서 이루어진 학습대상 고객 그룹핑과 유사한 절차를 거쳐 해당 고객을 사전에 정의된 $\mathbf{G}(l)$중 하나에 할당한다.

Definition 5
광고선정 대상 고객들의 상품 구매 행렬을 B^T라 하면 임의의 고객 j는 다음 (식5),(식6),(식7)에 의해 $\mathbf{G}(l)$ 중 하나에 할당된다.

$$W_m = \frac{\sum_{k=1}^{R} B^T_{m,k}}{R} \quad \text{(식 7)}$$

$$H^T = DB^T - DW^T \quad \text{(식 8)}$$

$$j \in G(l'),\ H^T_{l',j} = MAX_{\text{모든 } l \text{에 대하여}}(H^T_{l,j}) \quad \text{(식 9)}$$

3.3.2 포지셔닝 맵(positioning map)을 참조한 광고 제공

고객 j가 3.2 절에서 생성된 특정 고객그룹 $G(l)$에 할당되면, Definition 4에 의해서 선정된 $S(l)$이 해당 고객에게 제시된다.

4.비교실험

앞 절에서 제시한 다차원척도법(MDS)을 이용한 광고선정기법의 효과성을 평가하기 위하여 비교 분석을 수행하였다. 고객분류를 위한 고객그룹의 개수를 2개, 8개, 16개, 32개의 경우로 나누어 실험하였고, 기존의 방법인 협동적 필터링(Collborative Filtering)과 TOP3 방법(학습대상 전체의 선호도 점수 중 평균이 가장 높은 광고를 3개 선정), 임의선정과도 비교하였다.

4.1 실험데이터

본 논문에서 분석을 위해서 사용한 데이터는 웹 설문조사를 통해서 의해 학력수준, 직업, 기호에 따라 폭넓게 수집되었다. 설문조사는 두 가지 유형의 제품(MP3, Sporting and Leisure Goods)에 관해 제품구매여부와 광고선호도 점수를 알아보기 위해 설계되었다. 본 논문에서는 MP3 제품에 관한 결과만을 가지고 실험을 수행하였다.

MP3의 계층적 구조(hierarchical category structure)는 [그림 2] 에 나타나 있다. MP3 제품은 32개의 제품으로 이루어져 있고, 16명의 가수 카테고리(artist category)에 각각 2개의 MP3 제품을 구성하며, 16명의 가수 카테고리는 8개의 장르에 걸쳐 2개의 카테고리로 나뉘어져 있다.

4.2 분석결과

[표 1]은 MP3 제품들 중 3개의 광고를 선택하게 하여 4가지 광고추천 방법을 적용한 결과이다. 전체 351명의 데이터는 임의로 학습 데이터(210명), 테스트 데이터(141명)로 구분하였다. [표 1]의 평균점수는 테스트 데이터로 구분된 141명의 응답자가 16개의 광고 중 대안광고로 제시된 3개의 광고에 대해 5점 척도로 점수를 부여 한 것을 산술평균한 점수이다. 광고추천기법간의 차이가 과연 추천기법간의 차이 때문인지의 여부를 알기 위해 통계적 기법인 분산분석(ANOVA test)을 실시하였다. MDS 방법을 적용한 경우가 협동적 필터링이나 TOP3 방법보다 전체적으로 평균이 높았으며, MDS(16)의 경우는 두 방법보다 5% 유의수준에서 통계적으로 유의하였다.

5. 결론

본 논문에서 MDS를 활용한 개인화 광고 제공 기법을 제시하였다. 학습 대상 고객들의 구매 데이터와 광고별 선호도 정보를 활용하여 포지셔닝 맵을 작성한다. 포지셔닝 맵에는 상품 카테고리별 위치와 대안 광고별 위치가 표시된다. 상품 카테고리별 위치는 학습 대상 고객의 구매 데이터를 활용하여 상품 카테고리에 고객들을 할당하여, 할당된 고객들의 선호도들을 종합하여 결정한다. 대안 광고별 위치는 전체 학습 대상 고객의 해당 광고에 대한 선호도 점수를 바탕으로 결정한다. 실시간에 특정 고객에게 광고를 제공하기 위해서는 해당 고객의 구매 데이터를 참조하여 상품 카테고리에 할당하고, 학습 단계에서 생성된 포지셔닝 맵을 참조하여 개인화된 광고가 선정된다.

추후 연구과제로는 구매횟수에 따른 제시

[표1] 광고선정기법간 비교표

광고추천기법	광고선호도 점수		분산분석을 통한 다중비교						
	평균	표준편차	Random	CF	MDS(2)	MDS(8)	MDS(16)	MDS(32)	TOP3
Random	2.7778	0.7566	-	0	#0.000	#0.000	#0.000	#0.000	#0.000
CF	3.4894	0.8697		-	0.187	0.102	#0.033	0.211	0.363
MDS(2)	3.6265	0.9657			-	0.750	0.413	#0.010	#0.026
MDS(8)	3.6596	0.8605				-	0.617	#0.004	#0.011
MDS(16)	3.7116	0.8758					-	#0.001	#0.002
MDS(32)	3.3593	0.8166						-	0.733
TOP3	3.3948	0.9450							-

유의수준 0.05에서 유의 함.

한 기법의 효과 분석이 필요하다. 또한 실험분석 결과에서 볼 수 있듯이 여러 가지 경우의 카테고리 수의 결정에 따라서 광고에 대한 선호도 점수가 차이가 났는데 고객들을 그룹화할 때 어느 수준의 카테고리 세분화가 고객만족도를 최대로 할 수 있을 것인지에 대한 연구가 필요하다.

참고문헌

[1] 김종우, 이경미, 김영국, 유관종, "인터넷 상점에서의 실시간 개인화 광고 제공 기법," *경영정보학연구*, 제9권, 제4호, 1999.
[2] 김종우, 이경미, "인터넷 상점에서 개인화 광고를 위한 장바구니 분석 기법의 활용," *경영과학*, 제17권, 제3호, 11월, 2000, pp. 19-30.
[3] 이건창, 정남호, "데이터 마이닝 기법과 지능형 에이전트 기법을 결합한 인터넷 쇼핑몰의 설계 및 구현에 관한 연구," *정보기술응용연구*, 제1권, 제2호, 10월, 1999, pp. 113-137.
[4] 이재규, 최형림, 김현수, 이경전, *전자상거래원론*, 법영사, 1999.
[5] 이재창, 박정섭, "다차원축척(Multidimensional scaling)기법," *응용통계*, 제1권, 제1호, pp.61-79.
[6] 최용석, *SAS다차원척도법*, 자유아카데미, 서울, 1995.
[7] 허명회, *SAS최적척도법:사례분석을 중심으로*, 자유아카데미, 서울, 1994.
[8] 황병연, "개선된 추천을 위해 클러스터링을 이용한 협동적 필터링 에이전트 시스템의 성능," *정보처리논문지*, 제7권, 제5S호, 5월, 2000, pp. 1599-1608.
[9] Agrawal, R., Arning, A., Bollinger, T., Mehta, M., Schafer, J. and Srikant, R., "The Quest Data Mining System," *Proceeding of the 2nd International Conference on Knowledge Discovery in Databases and Data Mining*, Portland, Oregon, August, 1996.
[10] Allen, Cliff, Deborah Kania, and Yaeckel Beth, *Internet World Guide to One-to-One Web Marketing*, John Wiley & Sons, Inc., New York, 1998.
[11] Ansari, Asim, Skander Essegaier, and Rajeev Kohli, "Internet Recommendation Systems," *Journal of Marketing Research*, Vol. XXXVII, August, 2000.
[12] Breese, John S., David Heckerman, and Carl Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Technical Report*, MSR-TR-98-12, Microsoft Research, October, 1998.
[13] BroadVision, "A BroadVision One-to-One White Paper," *White Paper*, BroadVision, 1996.
[14] Johnson, R.A. and Wichern, D.W., *Applied Multivariate Statistical Analysis*, Prentice-Hall, New Jersey, 1992.
[15] Kim, Jong Woo, Kyung Mi Lee, Michael, J. Shaw, Hsin-Lu Chang, Mathew Nelson, Robert F. Easley, "A Preference Scoring Technique to Personalized Advertisements on Internet Storefronts." revised to *Mathematical and Computer Modeling*, 2002.
[16] Kim, Jong Woo, Byung Hun Lee, Michael J. Shaw, Hsin-Lu Chang, Mathew Nelson, "Application of Decision Tree Induction Techniques to Personalized Advertisements on Internet Storefront," forthcoming *International Journal of Electronic Commerce*, Vol. 5, No. 3, Spring, 2001.
[17] Kruskal, J.B. and Wish, M., *Multidimensional Scaling*, Sage University Paper Series on Quantitative Applications in the Social Sciences, Sage Publications, Beverly Hills and London, 1978.
[18] Young, F.W., Multidimensional Scaling: History, Theory, and Appcialtions(ed. R.M.Hammer), Lawrence Erlbaum Associates, New Jersy, 1987.