Journal of The Korea Society of Computer and Information
Vol. 19, No. 2, February 2014
www.ksci.re.kr
http://dx.doi.org/10.9708/jksci.2014.19.2.193

추천시스템에서 구매 패턴 예측을 위한 SOM기반 고객 특성에 의한 군집 분석

조 영 성*, 문 송 철**, 류 근 호***

Clustering Analysis by Customer Feature based on SOM for Predicting Purchase Pattern in Recommendation System

Young Sung Cho*, Song Chul Moon**, Keun Ho Ryu***

요 약

유비쿼터스 컴퓨팅이 생활의 일부가 되어가면서 정보의 양도 급속도로 늘어나고 있으며, 이로 인해 많은 데이터 속에서 정보를 찾아내는 기술이 부각되고 있다. 고객 기반의 협력적 필터링을 이용한 고객 선호도 예측 방법에서는 아이템에 대한 사용자의 선호도를 기반으로 이웃 선정 방법을 사용하므로 아이템에 대한 속성을 반영하지 못할 뿐만 아니라 희박성 문제를 해결하지 못하고 있다. 그리고 비슷한 선호도를 가진 일부 아이템의 정보를 바탕으로 하기 때문에 아이템의 속성은 무시하는 경향이 있다. 본 논문에서는 전자상거래 추천에서 구매 패턴 예측을 위한 고객 특성기반 SOM 학습을 이용한 군집 방법을 제안한다. 제안 방법은 고객의 속성 정보 기반의 유사한 속성의 데이터끼리의 클러스터링을 통해 보다 빠른 시간 내에 고객 성향에 맞는 추천이 가능한 구매 패턴의 추출이 가능하다. 성능 평가를 위해 현업에서 사용하는 인터넷 화장품 아이템 쇼핑몰의 데이터를 기반으로 데이터 셋을 구성하여 기존 시스템과 비교 실험을 통해 성능을 평가하여 효용성과 타당성을 입증하였다.

▸ Keywords : 세분화 기법, SOM, 추천시스템

Abstract

Due to the advent of ubiquitous computing environment, it is becoming a part of our common life style. And tremendous information is cumulated rapidly. In these trends, it is becoming a very important technology to find out exact information in a large data to present users. Collaborative filtering is the method based on other users' preferences, can not only reflect exact attributes of

•제1저자 : 조영성 •교신저자 : 문송철
•투고일 : 2013. 10.24, 심사일 : 2013. 11. 20, 게재확정일 : 2013.12. 3.
* 동양미래대학교 전산정보학부(Dept. of Computer Science, Dongyang mirae University)
** 남서울대학교 컴퓨터학과(Dept. of Computer Science, Namseoul University)
*** 충북대학교 전자컴퓨터공학부(School of Electrical and Computer Science, Chungbuk National University),

user but also still has the problem of sparsity and scalability, though it has been practically used to improve these defects. In this paper, we propose clustering method by user's features based on SOM for predicting purchase pattern in u-Commerce. it is necessary for us to make the cluster with similarity by user's features to be able to reflect attributes of the customer information in order to find the items with same propensity in the cluster rapidly. The proposed makes the task of clustering to apply the variable of featured vector for the user's information and RFM factors based on purchase history data. To verify improved performance of proposing system, we make experiments with dataset collected in a cosmetic internet shopping mall.

▸ Keywords : Segmentation Method, SOM(Self-Organizing Map), Recommendation System

I. 서 론

유비쿼터스 상거래 추천시스템은 신속하고 편리함이 제공되어야 하고 동적인 환경에서 실시간성과 민첩성이 요구된다. 전자상거래에서 구매 촉진을 증진시키는 마케팅 전략으로 데이터마이닝에서 추출한 지식을 적극 활용하는 방법으로 먼저 고객의 특성을 분석하여 고객 분류 작업을 수행하면 효율적인 추천방법을 제공할 수 있다. 신경망의 SOM을 이용한 학습 알고리즘을 이용하여 우수 고객군과 불량고객군 등이 분류가 가능하고 이 고객군의 전략적 포지셔닝 결과를 통하여 마켓팅 프로모션의 대상이 되는 타켓 고객군의 추출이 가능하다. 인터넷 쇼핑몰의 추천 시스템에서 아이템의 연관성과 속성 반영을 위해 세분화 기법을 이용한 데이터 마이닝을 적용한 연구[1-3]가 활발히 진행되고 왔다. 본 논문은 이러한 연구의 계속적 연구의 결과로 전자상거래 추천에서 기존의 추천 시스템의 문제점을 해결하고 구매 가능성이 높은 아이템을 추천하기 위하여 고객의 구매 특성에 의한 군집분석이 가능하고 신속하고 효과적인 추천을 위한 구매 패턴 예측을 위해서 SOM기반 고객 특성에 의한 군집 분석을 제안한다. 본 논문의 구성은 다음과 같다. 제 Ⅱ장은 관련 연구를 다루었으며 제 III 장에서는 제안 추천시스템 설명하며 제 Ⅳ장에서는 실험 및 성능 평가를 실행하며 마지막으로, 제 Ⅴ장에서는 본 논문의 결론과 향후 연구에 대하여 기술한다.

II. 관련연구

2.1 세분화 기법(Segmentation Method)

세분화 기법의 RFM(Recency, Frequency, and Monetary)은 세 가지 요소로 구성되어진다. 첫째, 최근성은 최근에 구매한 고객이 앞으로 구매할 가능성이 높다는 판단 하에 최근 구매일이 가까울수록 높은 점수를 부여한다. 둘째, 빈도성은 일정 기간 동안의 거래 빈도에 따라 고객을 세분화하는 것으로, 빈도가 높은 고객일수록 앞으로 구매할 가능성이 높다는 판단 하에 거래 빈도가 높을수록 높은 점수를 부여한다. 셋째, 총구매액은 일정기간 동안의 아이템 구입에 사용한 총 구매금액에 따라 고객을 세분화하는 것으로, 총 구매금액이 높은 고객이 앞으로 구매할 가능성이 높다는 판단 하에 총 구매금액이 높을수록 높은 점수를 부여한다[2,3]. 고객 데이터는 RFM을 이용하여 세분화가 가능하다. 정5등급(Exact Quintile) 방법은 각 변수 별로순서대로 정렬한 다음 순서대로 정확히 20%씩 5개의 그룹으로 나누어 등급을 부여 한다. 각 요소마다 5개의 세분화 세그먼트로 나누져서 전체 고객은 결국 $5 \times 5 \times 5 = 125$ 개의 세그먼트로 고객을 분류하게 된다[4]. 상위 10% 에 해당하면 5점, 그 다음 20%에 해당하면 4점, 40%에 해당하면 3점, 20% 해당하면 2점 하위 10%에 해당하면 1점을 주는 방식을 사용할 수 있다. RFM Matrix에서 보여주듯이 데이터베이스의 고객정보는 우량고객을 결정하기 위해 정렬 되는데 5등급으로 나눌 때 '555' 셀은 가장 이상적인 고객 집단이 된다. RFM은 가치 있

www.kci.go.kr

는 고객을 추출해 내어 이를 기준으로 고객을 분류할 수 있는 매우 간단하면서도 유용하게 사용될 수 있는 방법으로 알려져 있다. 고객의 진정한 가치를 평가함에 있어 해당 기업의 여건이나 고객에 대한 그 기업의 시각에 따라 R, F, M의 중요성은 달라 질 수도 있다. RFM은 구매 가능성이 높은 고객을 선정하기 위한 데이터 분석 방법 중의 하나이다.

2.2 SOM(Self-Organizing Map)

SOM은 신경회로망(Neural Network)의 일종으로 신경회로망의 중요한 특징 중에 하나는 환경으로부터 학습을 하고, 학습을 통한 수행능력을 가지고 있다. SOM은 외부의 피드백이나 지도가 없이 스스로 학습하여 입력자료에서 의미있는 패턴이나 특징을 발견하는 시스템이다[5]. SOM은 입력벡터를 훈련집합에서 match 되도록 가중치가 조정되는 인공신경세포 (Neuron) 격자에 기초한 자율학습(Unsupervised Learning)의 한 방법이다. SOM의 학습은 경쟁학습을 기초로 한다. 경쟁학습은 출력 뉴런이 활성화되기 위하여 서로 경쟁한 후 어떤 시점에서 단지 하나의 출력 뉴런만을 활성화시키는 것을 의미한다[6]. 경쟁하여 이긴 출력 뉴런을 승자 획득 뉴런이라고 한다. 즉, 개체 벡터에 가장 가까운 중량 벡터를 찾아 개체 벡터 방향으로 이동시키는 과정을 반복하는 데, 이 과정에서 가장 가까운 중량 벡터의 주변 중량 벡터도 함께 개체 벡터 방향으로 이동시킨다. 여기서 가장 가까운 중량 벡터를 모든 중량 벡터와의 경쟁에서 승리하였다는 의미에서 '승자(winner)' 라고 부른다. 학습의 반복은 모든 중량 벡터 값의 변화가 거의 소멸하거나 미리 지정된 최대 한계에 도달할 때까지 계속 진행되며, 이러한 학습과정을 통해 얻은 최종적인 중량 벡터는 개체공간상에서 개체들이 지니는 위상적 지도(topological map)를 반영하게 된다. 그림1은 n차원의 입력 데이터를 표현하는 n개의 입력 노드들과 k개의 분류영역(decision region)을 표현하기 위한 k개의 출력 노드로 구성되어 있다. 모든 입력 노드들은 모든 출력 노드들과 연결되어 있고 연결 가중치(weight)를 갖는다.

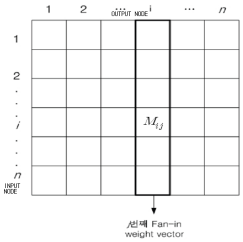

그림 1. 연결 가중치 행렬
Fig 1, Matrix of linked weight

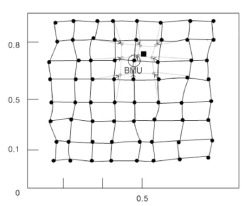

그림 2. 최상 일치 단위
Fig 2. Best Matching Unit

입력노드 i와 출력노드 j를 연결하는 weight M_{ij} 들의 행렬을 보여준다. i번째 행은 입력노드 i로부터 출력노드로 나가는 연결 가중치를 나타내며, j번째 열은 각 입력 노드로부터 출력노드 j로 들어오는 연결 가중치를 나타낸다. 여기서 j번째 열로 나타내어지는 벡터는 j번째 Fan-in weight vector라고 하며 입력 벡터와의 거리 계산에 사용된다[7]. SOM은 학습단계에서 피드백(feedback)과정이 없어 구조적으로 상당히 학습수행이 빨라 실시간 학습처리가 가능한 모델이어서 대량의 구매 데이터를 훈련데이터로 학습할 필요없이 자기조직화가 가능하다. 또한 연속적인 학습이 가능하여 추가적으로 입력되는 구매 데이터에 대한 클러스터링이 가능하다. 다음은 SOM 수행 절차 알고리즘을 나타낸 것이다[8].

표 1. SOM 수행 절차 알고리즘
Table 1. Proceduaral algorithm for SOM

Step 1. K 개의 출력노드를 위상공간 내에 배치한다.
Step 2. 학습률 $\alpha(t)$ 을 초기화한다. 학습률은 0 과 1 사이의 값을 가지며, 시간이 지남에 따라 감소한다.
Step 3. 새로운 x(t) 입터력벡를 입력노드에 제시한다
Step 4. 입력벡터와 모든 출력노드들과의 거리를 계산하여 최소거리를 가지는 승자노드 $m_c(t)$를 찾는다. 이 때 거리는 Euclidian 거리로 계산한다. $\|x(t) - m_c(t)\| = \min \|x(t) - m_i(t)\|$
Step 5. 승자노드와 이웃한 출력노드들의 가중치를 갱신한다.
Step 6. 이웃 노드의 가중치를 갱신하기 위해서 이웃 함수(Neighborhood Function)를 사용한다. $\alpha(t)$는 학습률, 이웃 함수 $\lambda_{ci}(t)$로 Gaussain 함수를 사용하여 이웃 노드를 계산한다. $m_i(t+1) = m_i(t) + \alpha(t)\lambda_{ci}(t)[x(t) - m_i(t)]$ $\lambda_{ci}(t) = \exp(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)})$
Step 7. 2 번으로 가서 반복한다.

III. 고객 특성기반 SOM 학습을 통한 군집을 이용한 추천시스템

이 장에서는 별도의 고객 프로파일을 생성하지 않고 고객 데이터와 구매이력데이터를 이용하여 SOM기반 고객 특성에 의한 이웃고객 군집 생성을 위한 학습 알고리즘을 기술한다.

3.1 고객 특성에 의한 학습을 위한 적용

전자상거래 추천에서 구매 패턴 예측을 위한 고객 특성에 의한 이웃고객 군집을 위해서 고객 특성 분류기는 Kohonen map을 이용하여 고객 특성을 적절한 클러스터로 분류한다.

www.kci.go.kr

고객 특성 분류기는 학습 에이전트 내에 운영되며 고객정보의 고객 특성은 Kohonen map의 입력 유닛에 제시되는 데, 그 고객정보는 고객 특성 값들에 기반하여 이웃고객을 클러스터링 할 수 있다. Kohonen map에 제시되는 데이터 형식은 bit 패턴으로 구성된다. 이 각 bit는 특정고객 특성에 포함되는 하나의 특성 값과 연결된다.

표 2. 고객 특성값의 종류
Table 2. Kinds of customer features

Bit No.	age	gender	occupation	type	region	hobby
0	1~9	F	학생	아토피	서울	경기관람
1	10~19	M	공무원	여드름	인천	독서
2	20~29		전문직	지성	경기	등산
3	30~39		서비스판매직	복합성	충남	미술감상
4	40~49		생산직	민감성	전남	바둑
5	50~59		의료인	건성	경북	운동
6	60~69		교육		경남	사진
7	70~79		농어민		제주	수집
8	80~89		군경		대구	여행
9	90~99		종교인		부산	영화. 음악
10			언론인		광주	요리
11			무직		울산	없음

원래 Kohonen map의 출력 유닛은 이 차원 평면상에 표현하기 위한 것이다. 그러나 전자상거래 추천에서는 그러한 패턴 모양을 추출하는 것이 목적이 아니기 때문에 출력 유닛의 구조를 일 차원으로 구현한다. 일 차원 구조를 갖게 되면 계산 양이 줄어 들기 때문에 실용적인 시스템 구성이 가능하다. 추천 시스템을 구성하는 고객 특성에 의한 학습 에이전트는 고객 특성 분류 기능과 고객 특성 학습 기능을 모두 포함한다. 학습 에이전트의 학습기능은 Kohonen map 신경망에 의해 구현된다. 이 신경망은 입력 데이터를 적절한 패턴 클러스터로 분류하고 그 입력을 계속해서 학습해 나감으로써 빈도수가 높은 입력 데이터를 추출한다.

3.2 고객 특성기반 SOM학습 알고리즘

본 논문에서는 인터넷 쇼핑몰에서 설계된 고객의 특성 변수(나이, 성별, 직업, 지역 등)로 기술되는 고객 정보를 기반으로 구매이력 정보를 조인하여 전자상거래 추천에서 구매 패턴 예측을 위하여 전처리(pre-processing) 과정을 통하여 고객특성 정보와 RFM을 이용하여 고객점수를 생성하여 구매 이력 정보를 고객 등급별로 클러스터링하여 활용한다. 우선 고객 정보를 고객특성별로 클러스터링하고 구매 이력 정보와 조인하여 이웃고객을 클러스터링 하여 추천에 활용한다. 다음 표 3.은 Pseudo code로 작성된 고객 특성기반 SOM 학습 절차 알고리즘이다[8].

표 3. 고객 특성기반 SOM 학습 절차 알고리즘
Table 3. Procedural algorithm of SOM learning based on customer feature

Procedureal Algorithm for customer learning based on SOM
Input : Set of N dimension vector, X
Output : Subset of input data (M subsets)

Begin
Randomly initialize $w_i = (w_{i1}, w_{i2}, \ldots, w_{in})$
for each node
for (t=0; unless a stopping condition is reached; Increase t)
  for (for all input data)
    for (i=0 to M)
      Compute $D_i = //X_t - W_i^{(t)}//$
    endfor
    Find the winner j=i such that $D_i(t)$ *is minimum for over all i*
    Update the winner j (and its neighbors)
  endfor
endfor
End:

SOM은 학습단계에서 피드백(feedback)과정이 없어 구조적으로 상당히 학습수행이 빨라 실시간 학습처리가 가능한 모델이어서 대량의 구매 데이터를 훈련데이터로 학습할 필요 없이 자기조직화가 가능하다. 또한 연속적인 학습이 가능하여 추가적으로 입력되는 구매 데이터에 대한 군집화가 가능하다. 표 4는 이 논문에서 SOM 알고리즘에 의한 고객 데이터 군집

표 4. SOM에 의한 고객 데이터 패턴 클러스터링 결과
Table 4. The SOM Result of Customer Data Pattern by SOM

r: 4.08 f: 5.00 m: 5.00 12명 (96.33)	r: 4.12 f: 4.37 m: 3.87 8명 (82.50)	r: 4.00 f: 3.66 m: 3.29 7명 (73.14)
r: 4.41 f: 2.50 m: 3.25 44명 (63.64)	r: 3.86 f: 2.01 m: 2.01 74명 (54.38)	r: 3.93 f: 1.75 m: 1.90 107명 (44.90)
r: 3.87 f: 1.02 m: 1.15 40명 (32.90)	r: 2.92 f: 1.00 m: 1.64 13명 (27.69)	r: 0.0 f: 0.0 m: 0.0 0명 (0)

www.kci.go.kr

결과를 나타낸 것이다. SOM 알고리즘을 이용하여, 고객들을 9개의 세그먼트로 나누었다. 각 세그먼트는 RFM 분석에 의해 분류된 것으로, 9개의 각각의 세그먼트에 대해서 전략을 세워서 로그인 사용자에게 고객성향에 맞는 각 고객군을 기반으로 추천 전략을 세울 수가 있다[9].

3.3 SOM기반 고객 구매 패턴 분석

전자상거래 추천에서 구매 패턴 예측을 위하여 전처리(pre-processing) 과정을 통하여 고객 데이터와 구매 이력 데이터를 기반으로 고객 구매 패턴 분석에서 고객특성 정보와 RFM factors를 이용하여 고객점수를 생성하고 고객 등급별로 클러스터링하여 활용한다. 다음은 전체 데이터를 대상으로 SOM을 이용한 분석결과와 k-means를 이용한 분석 결과를 비교한 결과이다. 고객을 등급화하기 위해서 RFM factors를 적용하여 고객이 구매한 아이템에 대한 정보를 기반으로 분석 데이터를 생성하였다. 전자상거래에 구매가 발생한 구매 이력 데이터를 중심으로 고객을 세분화 하였으며 각 RFM factors 별 최근성(R), 빈도성(F), 총구매액(M)을 정5등급(Exact Quintile) 방법을 이용하였으며 고객 등급은 10점 단위로 나누어 활용한다. SOM을 이용하여, 구매 데이터를 기준으로 고객을 분석한 결과 고객 점수 20점 이하는 구매가 발생하지 않아 고객들을 9개의 세그먼트로 나누어 졌다. 각 세그먼트는 RFM의 값에 의해 분류된 것으로, 9개의 각각의 세그먼트에 대해서 전략을 세워서 각 고객군에 대한 추천 전략을 세울 수가 있다. 다음 표 5는 SOM에 의한 고객 구매 데이터 패턴 클러스터링 결과를 나타낸 것이다. Level의 구성은 고객점수단위로 Level1(score 〉= 90), Level2(score 〉= 80 and score 〈 90), Level3(score 〉= 70 and score 〈 80),...Level9(score 〈 10)로 구성된다. 고객 패턴 분석 결과는 SOM의 분석 결과가 k-means 분석 결과 보다 더욱 세분화된 결과가 도출됨을 알 수 있다.

표 5. SOM에 의한 고객 구매 데이터 패턴 클러스터링 결과
Table 5. The SOM Result of Customer Purchase Pattern

Level	Output		avg_factor		avg_score	
	k-means	SOM	k-means	SOM	k-means	SOM
1	18	52	3.6	4.69	96.31	96.22
2	12	24	3.58	4.11	81.67	82
3	8	20	3.07	3.71	73.6	73.5
4	69	138	3.25	3.39	63.65	63.71
5	110	238	2.81	2.92	54.45	54.69
6	170	286	2.51	2.53	45.26	44.8
7	42	58	2.37	2.02	33.03	32.86
8	12	16	2.17	1.64	27.75	27.67
9	0	0	0	0	0	0

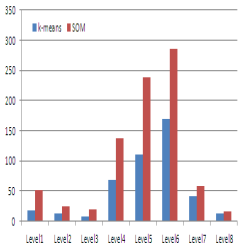

그림 3. 고객수준별 출력건수 비교
Fig 3. The result of Output Count by each level of customer

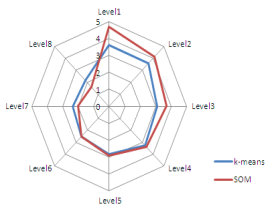

그림 4. 고객수준별 평균요소값 비교
Fig 4. The graph of average factors for RFM by each level of customer

Ⅳ. 실험 및 성능 평가

4.1 실험 환경

구현 및 실험 환경은 윈도우 운영체체하에서 해당 평가하기 위해서 다음과 같은 웹서버 환경을 사용하였다.

Web Server: Apache HTTP Server Version 2.2.8 / WAP 2.0 , j2sdk 1.7.0_11 as Java environement,

server-side: JSP/PHP 5.2.12, JQuery, client-script:XML/XHTML4.0/HTML5.0/CSS3/JAVA SCRIPT, C#.net framework 2.0, jakarta-tomcat5.0.28-http://jakarta.apache.org

4.2 실험 데이터 구성

고객 구매 패턴 예측을 위한 SOM기반 고객 특성에 의한 클러스터링을 이용한 개인화 추천시스템은 윈도우 XP 환경에서 인터넷 화장품 쇼핑몰을 위한 데이터베이스가 구축되었다. 시스템에 대한 평가를 위한 실험데이터의 구성은 쇼핑몰을 이용해 본 경험이 있는 고객 319명의 고객정보와 , 그리고 현재 화장품을 전문적으로 판매하는 인터넷 화장품 쇼핑몰인 P사의 아이템 분류에서 사용하는 화장품 아이템 580개를 대상으로 그들의 추천 1600건의 구매 데이터를 이용하여 2009년 2월 부터 2010 2월까지의 12개월간의 과거의 구매 데이터를 학습 셋으로 사용하였고, 2010년 3월 부터 2010년 5월까지 3개월간의 테스트 셋으로 사용하였다[3].

www.kci.go.kr

4.3 분석 및 성능 평가

추천시스템의 전체적인 성능 평가는 예측 값과 실제 값의 차이를 표시하여 정확성 측면에서 성능을 평가하기 위한 MAE 방식을 사용하였다. MAE는 예측의 정확성을 판단하는데 가장 많이 쓰이는 방법이다. 본 논문에서는 MAE에 대한 실험을 SOM적용 시스템과 k-means 적용시스템을 실험하였다. 우선 첫 번째, 실험으로 MAE에 의해 예측의 성능을 평가하였다. 추천시스템의 예측 값의 정확성을 평가하기 위해 MAE(Mean Absolute Error)를 사용하였고 식3과 같이 산출하였다[10].

$$MAE = \frac{\sum_{i=1}^{N} |\epsilon_i|}{N} \qquad \text{식 1}$$

N은 총 예측 회수를 나타내고, ε_i는 예측 값과 실제 값의 오차를 나타내며 i는 각 예측 단계를 나타낸다. 표 6은 식 1을 이용하여 예측값의 정확성 평가를 수행한 결과이다.

표 6. 제안 및 기존 시스템의 MAE에 의한 성능평가
Table 6. The result for table of MAE by comparing proposal system with existing system

	P_count	Proposal (SOM)	previous(KCA)	Existing
MAE	50	0.23	0.47	0.65
	100	0.13	0.23	0.32
	300	0.04	0.07	0.08
	500	0.03	0.05	0.06

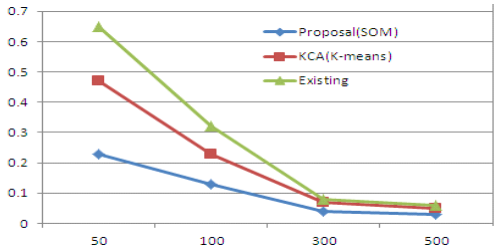

그림 5. 제안 및 기존 시스템의 MAE에 의한 성능평가
Fig. 5. The result for the graph of MAE by comparing proposal system with existing system

다음은 두 번째, MAE 실험으로 고객점수 등급별 MAE에 대한 실험이다. 성능평가는 출력 건수를 고려하여 고객 등급별을 정5등급(Exact Quintile) 방법을 적용하여 전체적인 성능을 평가 할수 있는 척도로 사용하였다.

표 7. 고객 수준별 MAE 성능 비교표
Table 7. The result of MAE by each level by comparing proposal system(SOM) with previous system(k-means)

Level	SOM_Output	K_means_Output	SOM_MAE	k-means_MAE
1	52	18	0.0025	0.00056
2	24	12	0.00292	0.00333
3	158	77	0.00215	0.00273
4	524	280	0.00099	0.00443
5	74	54	0.00149	0.00093

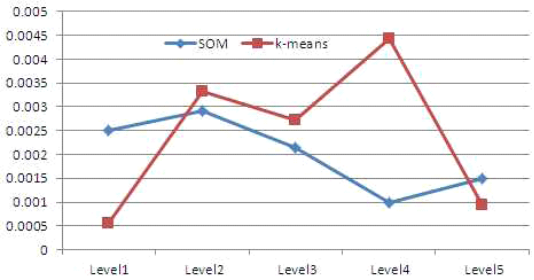

그림 6. 고객 수준별 MAE 성능 비교도
Fig 6. The result of MAE by each level by comparing proposal system(SOM) with previous system(k-means)

첫 번째, 성능평가 결과는 SOM을 적용한 제안시스템이 k-means를 적용한 이전 시스템보다 좋은 MAE를 나타내어 제안시스템의 성능이 향상되었음을 알 수 있었으며 두 번째, 성능평가 결과는 Level1과 Level5를 제외하고는 Level2와 건수가 많은 Level3, Level 4에서는 SOM을 적용한 제안시스템이 k-means를 적용한 이전 시스템보다 좋은 MAE를 나타내어 성능이 향상된 결과를 나타내었다. 기존 시스템은 대규모 데이터베이스의 거래 데이터처리로 인한 추천 아이템을 위한 선호도 계산 작업 등에 많은 시간이 소요되었다. 제안시스템은 전자상거래 추천에서 구매 패턴 예측을 위한 SOM기반 고객 특성에 의한 군집 분석을 이용한 개인화된 추천시스템으로 성능평가에서 k-means를 적용한 시스템 보다 성능이 향상되었다. 전처리 작업으로 SOM기반 고객 특성에 의한 군집 분석을 통한 추천 기법을 적용함으로써 효율적으로 성능을 향상 시킬 수 있었으며 유비쿼터스 상거래에서 요구되는 민첩성과 실시간성을 확보할 수 있었다.

V. 결론 및 향후 과제

대규모 거래가 이루어지는 유비쿼터스 상거래 환경하에서는 민첩성과 즉시성이 확보되어야 하고 고객이 원하는 구매 성향을 반영한 추천 아이템을 제공하기 위해서 고객 구매 패

www.kci.go.kr

턴을 예측하기 위한 고객 특성에 의한 군집이 필요하였다. 기존의 협력적 필터링을 이용한 고객 선호도 예측 방법에서는 비슷한 성향을 가지는 일부 사용자 정보에 근거하여 추천함으로써 나머지 사용자 정보를 무시하는 경향이 있었다. 또한 희박성과 확장성에 문제가 있었다. 이 논문에서는 이러한 문제점을 해결하기 위해서 사용자 평가 자료를 생성하지 않는 묵시적인 방법을 이용하여 구매 가능성이 높은 아이템을 찾기 위해서 전처리 과정을 통하여 고객 속성 분석이 가능한 RFM 기법을 이용하여 고객점수를 등급화 하였으며 SOM기반 고객 특성에 의한 군집을 생성하여 개인화 추천시스템에 고객 구매 패턴 예측을 위해 적용하여 실험한 결과 구매 예측을 위한 구매 패턴 출력 건수도 평균 2배 더 출력됨을 알 수 있었다. SOM을 이용한 추천시스템의 경우 k-means 알고리즘(KCA)을 이용한 이전의 추천시스템 보다 전체적으로 성능이 우수하였고 실시간성이 요구되는 유비쿼터스 상거래에서 구매 패턴 예측을 위한 추천의 성능을 향상되었음을 확인 할 수 있었다. 유비쿼터스 상거래가 증가되는 시점에 시간과 장소에 제약을 받지 않는 모바일 웹앱을 이용한 추천시스템에서 이전보다 향상된 추천시스템으로 개선할 수 있었다. 향후 연구로는 순차 패턴 마이닝을 이용한 개선된 추천방법에 관한 연구가 필요하다.

사사표기

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2012-0000478) 그리고 남서울대학교 학술연구비를 지원받아 수행된 연구임

참고문헌

〔1〕 Cho, Y.S., Moon, S.C., Noh, S.C., Ryu, K.H.: mplementation of Personalized recommendation System using k-means Clustering of Item Category based on RFM. In: 2012 IEEE International Conference on Management of Innovation & Technology Publication (June 2012)

〔2〕 Park, H.B, Y.S., Moon, S.C., Ko, H.H.: Clustering Method of Weighted Preference Using K-means Algorithm and Bayesian Network for Recommendation System. In: Journal of Information Technology & Management, Vol. 20, No. 3, pp. 219-230, (2013)

〔3〕 Cho, Y.S., Moon, S.C., Jeong, S.P., Oh, I.B., Ryu, K.H.: Clustering Method using Item Preference based on RFM for Recommendation System in u-Commerce. In: (eds.) Ubiquitous Information Technologies and Applications. LNEE, vol. 214, pp. 353-362. Springer, Hidelberg (2012)

〔4〕 Wei, J.-T., Lin, S.-Y. and Wu, H.-H.: The review of the application of RFM model. In: African Journal of Business Management, Vol. 4, No. 19, pp. 4199-42060, (2010)

〔5〕 Verdu. S. V. : Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps. In : IEEE Trans. Power System. Vol. 21, No.4, pp. 1672-168 (2006).

〔6〕 Jun S.H.: Technology Marketing using PCA , SOM, and STP Strategy Modeling. In: IJCSI(International Journal of Computer Science Issues), Vol. 8, Issue 1, January 2011

〔7〕 Kate A. Smith, Jatinder N. D Gupta.: Neural Networks in Business: Techniques and Applications. In: The IDEA GROUP PUBLISHING (2001).

〔8〕 T. Kohonen, Self-Organizing Maps. In: Springer (2000)

〔9〕 Cho, Y.S., Moon, S.C. Ryu, K.H.: SOM Clustering Method based on RFM Analysis for Predicting Customer Purchase Pattern in u-Commerce. In: KSCI , The Korea Society of Computer and Information Summer Conference 2013, (Jul, 2013)

〔10〕 Herlocker, J.L., Kosran, J.A., Borchers, A., Riedl, J.: An Algorithm Framework for Performing Collaborative Filtering.In: Proceedings of the 1999 Conference on Research and Development in Information Research and Development in Information Retrival (1999)

www.kci.go.kr

저 자 소 개

조 영 성

1989: 연세대학교 전산학과 공학석사.
2008: 충북대학교 전산학과 공학박사.
1989~2011: 미국CDC 한국지사 SE Manager, ㈜네오아이엔씨(CEO), 기술지도사(중기청),
현재: ㈜ 컴트리 기술연구소(소장), 동양미래대학 산업체 겸임/교수,
관심분야: 유비쿼터스 컴퓨팅 및 GIS, 데이터마이닝,e-Commerce 웹서비스, 정보보안
Email : youngscho@empal.com

문 송 철

1995: KAIST 경영정보공학석사.
2005: 국민대학교 정보관리학 박사.
1998년 : 한보정보통신(주) 이사
2005년 : 가나시스텍(주)대표이사
현 재: 남서울대학교 컴퓨터학과 교수
관심분야: 유비쿼터스 컴퓨팅, 정보보안, MIS, SW Process, e-Commerce
Email :moon@nsu.ac.kr

류 근 호

1976: 숭실대학교 전산학과 이학사
1980: 연세대학교 공학대학원 전산전공 공학석사
1988: 연세대학교 대학원 전산전공 공학박사
1976~1986: 육군군수 지원사 전산실(ROTC 장교), 한국전자통신연구원(연구원), 한국방송통신대 전산학과 (조교수) 근무
1989~1991: Univ. of Arizona Research Staff (TempIS 연구원, Temporal DB)
1986~현재: 충북대학교 전기전자 컴퓨터공학부 교수
관심분야 : 시간 데이터베이스, 시공간 데이터베이스, Temporal GIS, 지식기반 정보검색 시스템, 유비쿼터스 컴퓨팅 및 스트림데이터처리, 데이터 마이닝, 데이터베이스, 보안, 바이오 인포메틱스 컴퓨터공학
Email : khryu@dblab.chungbuk.ac.kr

www.kci.go.kr