

RFM 분석 기반 고객 구매 패턴을 예측을 위한 SOM 클러스터링 방법

조영성[○], 문송철^{*}류근호^{**}

[○]동양미래대학 전산정보학부

^{*}남서울대학교 컴퓨터학과

^{**}충북대학교 전기전자컴퓨터공학부

e-mail:youngsocho@empal.com[○], moon@nsu.ac.kr^{*}, khryu@dblab.chungbuk.ac.kr^{**}

SOM Clustering Method based on RFM Analysis for Predicting Customer Purchase Pattern in u-Commerce

Young Sung Cho[○], Song Chul Moon^{*}, Keun Ho Ryu^{**}

[○]Department of Computer Info., Dong Yang Mirae University

^{*}Department of Computer Science, Chungbuk National University

^{**}Department of Computer Science, Namseoul University

● 요약 ●

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 내용을 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 유비쿼터스 상거래에서 RFM(Recency, Frequency, Monetary) 분석 기반의 SOM을 이용한 군집방법을 제안한다. 제안 방법은 고객의 구매 데이터 기반의 유사한 속성의 데이터끼리의 클러스터링을 통해 보다 빠른 시간 내에 고객 성향에 맞는 추천이 가능한 구매 패턴 추출이 가능하다.

키워드: RFM기법(RFM Method), 점진적 마이닝(Incremental Mining), 추천시스템(Recommendation System)

I. 서론

유비쿼터스 상거래 추천시스템은 신속하고 편리함이 제공되어야 하고 동적인 환경에서 실시간성과 민첩성이 요구되고 있다. 전자상거래에서 구매 촉진을 증진시키는 마케팅 전략으로 데이터마ining에서 추출한 지식을 적극 활용하는 방법이 있다. 먼저 고객의 유형을 분석하여 고객 분류 작업을 수행한다. 신경망의 SOM을 이용한 알고리즘을 이용하여 우수 고객군과 불량고객군 등이 분류가 가능하고 이 고객군의 전략적 포지셔닝 결과를 통하여 마케팅 프로모션의 대상이 되는 타겟 고객군을 추출이 가능하다. 인터넷 쇼핑물의 추천 시스템에서 아이템의 연관성과 속성 반응을 위해 RFM 분석을 이용한 데이터 마이닝을 적용한 연구[1,2]가 활발히 진행되고 있다. 본 논문은 이러한 연구의 계속적 연구의 결과로 전자상거래 추천에서 기존의 추천 시스템의 문제점을 해결하고 구매 가능성이 높은 아이템을 추천하기 위하여 고객의 구매 패턴 파악이 가능하고 신속하고 효과적인 추천을 위하여 RFM 분석 기반의 고객 구매 패턴을 예측을 위한 SOM 클러스터링 방법을 제안한다.

II. SOM(Self-Organizing Map)

SOM은 신경회로망(Neural Network)의 일종으로 신경회로망의 중요한 특징 중에 하나는 환경으로부터 학습을 하고, 학습을 통하여 수행능력을 가지고 있다. SOM은 외부의 피드백이나 지도가 없이 스스로 학습하여 입력자료에서 의미있는 패턴이나 특징을 발견하는 시스템이다[4]. 즉 입력 벡터를 훈련집합에서 match 되도록 가중치가 조정되는 인공 신경세포 (Neuron) 격자에 기초한 자율학습 (Unsupervised Learning)의 한 방법이다. 전자상거래의 매출 목표를 달성하기 위한 추천을 위해 중요한 고객군/상품군과 비슷한 고객과 아이템을 여러 변수를 한꺼번에 고려하면서 추천 목록을 추출할 때 적합하게 활용이 가능하다.

이 논문에서 SOM을 이용하여 고객 구매 패턴을 예측을 위해서 구매 데이터를 클러스터링한다. 이 네트워크는 경쟁학습을 기초로 한다. 경쟁학습은 출력 뉴런이 활성화되기 위하여 서로 경쟁한 후 어떤 시점에서 단지 하나의 출력 뉴런만을 활성화시키는 것을 의미한다. 경쟁하여 이긴 출력 뉴런을 승자 획득 뉴런이라고 한다.

SOM에서는 뉴런이 N-차원으로 구성된 격자의 노드에 위치한다. 뉴런이 경쟁적 학습과정에 있는 입력 패턴의 클래스에 따라서 선택적으로 조정된다. 조정된 뉴런의 위치는 여러 입력 특징들을 대해서 의미 있는 좌표체계가 격자에 생성되는 방법으로 서로 관련 하여서 순서를 이루게 되는 경향이 있다. 격자에 있는 뉴런의 위치는 입력 패턴의 고유의 특성과 일치하도록 입력 패턴의 지형적인 지도(topological map)를 생성한다. SOM의 개발을 용이하게 하기 위해서는 신경회로망에 적용되어 활성화 되는 상태를 조절하는 매커니즘으로 측면 피드백(lateral feedback)을 사용하는데, 피드백 연결은 뉴런으로부터 떨어진 거리에 의존하여서 활성화 효과나 억제 효과를 생성한다[4]. 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 가진다.

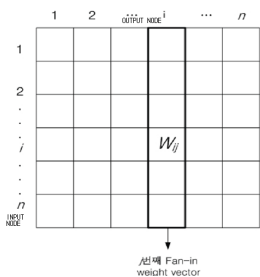


그림1-1. 연결 가중치 행렬

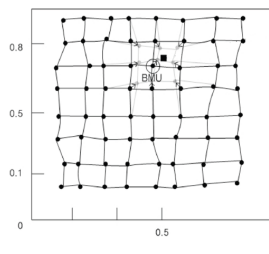


그림1-2. BMU (Best Matching Unit)

표1. RFM분석 기반 SOM Clustering 처리 절차 알고리즘
Table 1. Procedural algorithm for processing by SOM Clustering Method based on RFM Analysis

Algorithm for SOM Clustering	
Input : Set of N dimension vector, X	
Output : Subset of input data (M subsets)	
Begin	
Randomly initialize $w_i = (w_{i1}, w_{i2}, \dots, w_{in})$	
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_i - 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:	

입력노드 i와 출력노드 j를 연결하는 weight w_{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 나타내어지

는 벡터는 j번째 Fan-in weight vector라고 하며 입력 벡터와의 거리 계산에 사용된다. 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 군집화가 가능하다. 표2는 이 논문에서 SOM 알고리즘에 의한 고객 구매 데이터 클러스터링 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.

표2. SOM에 의한 고객 구매 데이터 패턴 클러스터링 결과
Table 2. The SOM Result of Customer Purchase Pattern by SOM

r: 4.03 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 (54.33)	r: 3.93 f: 1.75 m: 1.90 107명 (44.90)
r: 3.37 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)

III. 실험 및 성능 평가

현재 화장품을 전문적으로 판매하는 인터넷 쇼핑몰에서 RFM 분석 기반의 고객 구매 패턴을 예측을 위한 SOM 클러스터링 기법을 이용한 추천의 성능평가를 위해서 고객 319명의 고객 정보와 아이템 580개를 대상으로 그들의 추천 1600건의 구매 데이터를 이용하였다[2]. 추천시스템의 전체적인 성능 평가는 예측 값과 실제 값의 차이를 표시하여 추천시스템의 예측 값의 정확성을 평가하기 위해 MAE(Mean Absolute Error)를 사용하였고 식1과 같이 산출하였다[6].

$$MAE = \frac{\sum_{i=1}^N |\epsilon_i|}{N} \quad \text{식1}$$

N은 총 예측 회수를 나타내고, ϵ_i 는 예측 값과 실제 값의 오차를 나타내며 i는 각 예측 단계를 나타낸다. <표3>는 식2를 이용하

여 예측값의 정확성 평가를 수행한 결과이다.

표3. 제안 및 기존 시스템의 MAE에 의한 성능평가
Table3. The result for table of MAE by comparing proposal system with existing system

	P_count	Proposal	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

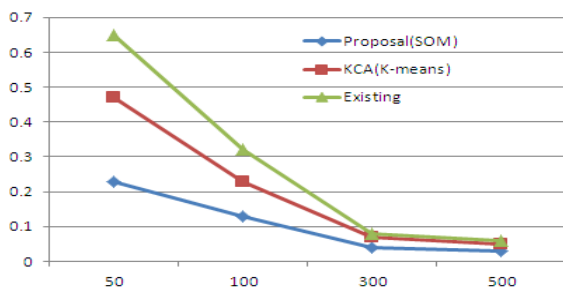


그림 1. 제안 및 기존 시스템의 MAE에 의한 성능평가
Fig. 1. The result for the graph of MAE by comparing proposal system with existing system

실험 결과에서 제안 시스템이 기존 시스템 보다 낮은 MAE의 결과를 얻을 수 있었다.

V. 결론 및 향후 과제

대규모 거래가 이루어지는 유통업체 상거래 환경하에서 고객이 원하는 구매 성향을 반영한 추천 아이템을 제공하기 위해서는 구매 패턴 예측을 위해 고객 군집화가 필요하였다. 기존의 협력적 필터링을 이용한 고객 선호도 예측 방법에서는 비슷한 선호도를 가진 일부 아이템의 정보를 바탕으로 하기 때문에 아이템의 속성은 무시하는 경향이 있었다. 이 논문에서는 아이템의 속성 분석이 가능한 RFM 분석을 바탕으로 기존의 추천 시스템의 문제점을 해결하였고 고객의 구매 패턴 파악이 가능하고 신속하고 효과적인 추천을 위해서 RFM 분석 기반의 고객 구매 패턴을 예측을 위한 SOM 클러스터링 방법을 제안하였다. SOM 클러스터링을 이용한

추천시스템의 경우 기존의 k-means 알고리즘(KCA)을 이용한 추천시스템 보다 전체적으로 성능이 우수하였고 실시간성이 요구되는 유통업체 상거래에서 구매 패턴 예측을 위한 추천의 성능을 향상시킬 수 있었다.

사사표기

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

참고문헌

- [1] Cho, Y.S., Moon, S.C., Noh, S.C., Ryu, K.H.: Implementation of Personalized recommendationSystem using k-means Clustering of Item Category based on RFM. In: 2012 IEEE International Conference on Management of Innovation & Technology Publication (June 2012)
- [2] Cho, Y.S., Moon, S.C., Ryu, K.H.: Personalized Recommendation System using FP-tree Mining based on RFM. In: KSCI, 17th-2 Vol (Feb. 2012)
- [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, Heidelberg (2012)
- [4] 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).
- [5] 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
- [6] Jonathan L. Herlocker, Joseph A. Kosran, Al Borchers, and John Riedl, "An Algorithm Framework for Performing Collaborative Filtering", Proceedings of the 1999 Conference on Research and Development in Information Retrieval, 1999