

전자상거래 추천을 위한 RFM기반의 점진적 빈발 패턴 마이닝 기법

조영성[○], 문송철^{*}, 류근호^{**}

[○]동양미래대학 전산정보학부

^{*}남서울대학교 컴퓨터학과

^{**}충북대학교 전기전자컴퓨터공학부

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

RFM based Incremental Frequent Patterns mining Method for Recommendation in e-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 ●

기존의 연관규칙을 이용한 추천시스템은 점진적으로 증가하는 트랜잭션 데이터를 처리하기 위해서 기존에 처리한 데이터를 재 처리하는 비효율성의 문제가 있다. 본 논문에서는 전자상거래에서 RFM(Recency, Frequency, Monetary)기반의 점진적 빈발 패턴 마이닝을 이용한 추천기법을 제안한다. 제안 방법은 새로운 트랜잭션 데이터가 추가 되었을 때 보다 빠른 시간 내에 연관 규칙을 추출이 가능하다.

키워드: RFM기법, 점진적 마이닝, 추천시스템

● Abstract ●

A existing recommendation system using association rules has the problem, which is suffered from inefficiency by reprocessing of the data which have already been processed in the incremental data environment in which new data are added persistently. We propose the recommendation technique using incremental frequent pattern mining based on RFM in e-commerce. The proposed can extract frequent items and create association rules using frequent patterns mining rapidly when new data are added persistently.

Keywords: RFM Method, Incremental Mining, Recommendation System

I. 서론

유비쿼터스 상거래 추천시스템은 신속하고 편리함이 제공되어야 하고 동적인 환경에서 실시간성과 민첩성이 요구되고 있다. 기존의 데이터 마이닝을 이용한 추천시스템은 특정시점의 데이터베이스 전체를 대상으로 하고 있으며 연관규칙 탐사를 위해 과도한 데이터베이스 연산처리 부하를 수반한다. 또한 연관규칙의 탐사시점 이후에 발생하는 트랜잭션은 전혀 탐사 대상으로 고려할 수 없기 때문에 시간적 제약을 받아서 연관규칙 탐사를 필요로 하는 실시간 탐사의 동적 환경하에서는 연관규칙의 처리에 어려운 문제점을 갖고 있다. 인터넷 쇼핑물의 추천 시스템에서 아이템의 연관성

과 속성 반영을 위해 RFM 분석을 이용한 연관규칙을 적용한 연구[1,2]가 활발히 진행되고 있다. 본 논문은 이러한 연구의 계속적 연구의 결과로 전자상거래 추천에서 연관규칙의 탐사 처리에 문제점을 해결하고 구매 가능성이 높은 아이টে를 추천하기 위하여 고객의 구매 패턴 파악이 가능하고 신속하고 효과적인 추천을 위하여 RFM기반의 점진적 빈발 패턴 마이닝 기법을 제안한다.

II. 점진적 마이닝

연관규칙 탐사는 대규모로 축적되어 있는 트랜잭션 데이터베이스

스를 바탕으로 지지도와 신뢰도를 이용하여 연관성이 강한 항목들을 찾아내는 것으로 정의된다. 점진적 마이닝 기법은 기존의 연관 규칙 마이닝 기법을 확장한 것으로서 지속적으로 증가하는 데이터 간의 규칙을 유지하는 기법으로 이전에 탐사된 규칙을 유지하면서 데이터가 증가된 이후의 규칙을 탐사하는 방법이다[3]. 데이터(D)로부터 주어진 시간(t)에 규칙탐사를 수행하고 분석된 시점(t') 이후부터 추가적으로 저장된 데이터(D')에 대해서 새로운 규칙을 탐사하기 위해서 이전의 규칙을 유지한다[4]. 점진적 마이닝 기법에서 갱신이 일어날 경우 이전에 발견된 규칙을 이용하여 갱신된 데이터베이스의 크기를 줄이기 때문에 점진적으로 추가된 데이터를 가지고 분석을 하기 위해 I/O 비용과 계산 비용을 최소화할 수 있다. 다음<표1>은 이 논문에서 사용하는 용어에 대한 설명이다.

표1. 용어에 대한 설명
Table 1. Notations description

Notation	Description
DB	원본 데이터 셋
db	새로 추가된(점진적) 데이터 셋
UD	갱신된 데이터 셋
d	db안의 트랜잭션의 수
s	db안의 minimum support
L_k	DBUDb 의 Frequent K-itemsets
$X, Support_d$	db의 아이템X에 대한 Support count
$T-tree_{DB}$	원본 데이터베이스(old_DB)
$T-Tree_{db}$	새로 추가된 데이터베이스(new_DB)
$T-tree_{UD}$	갱신된 데이터베이스(now_DB)

DB로부터 기본 빈발 패턴 생성 모델인 $old_DB(T-tree_{DB})$ 를 생성하고 Pruning과정을 생략하는 것이 점진적 마이닝 알고리즘의 첫 번째 단계이다. 새로 추가된 db의 정보를 추가하기 위하여 새로운 $new_db(T-Tree_{db})$ 를 구축한다. $old_DB(T-tree_{DB})$ 의 크기를 줄이고 효과적인 메모리 사용을 위하여 db로부터 구축된 $old_DB(T-tree_{DB})$ 에서는 다음의 식1을 기반으로 새로운 데이터가 추가될 때 생성되는 빈발패턴들에 대한 효과적인 전처리 방법인 pre-minimum support[5] 개념을 사용하여 Pruning작업을 한다. 이 논문에서는 인터넷 쇼핑몰에서 운영되는 구매이력 데이터베이스(sale)를 기반으로 가장 많은 분포를 이루는 아이템 RFM 점수가 반영된 연관규칙 탐사를 위하여 점진적 빈발 패턴 마이닝을 적용한다. 고객 군집내의 각 고객들의 아이템 구매 정보를 일정기간(예: 1주일)동안 수집하여 추가된 데이터(new_db)를 누적시킨다. 그리고 고객군집으로부터 아이템 구매 연관규칙을 추출하여 연관규칙 테이블에 저장한다. 그리고 주기적으로 점진적 마이닝을 위한 연관규칙 마이닝의 처리 단계를 반복한다. 다음은 추가되는 데이터에 대한 연관규칙 처리를 위한 RFM기반 점진적 빈발 패턴 마이닝 처리 절차 알고리즘을 나타낸 것이다.

표2. RFM기반 점진적 빈발 패턴 마이닝 처리 절차 알고리즘
Table 2. Procedural algorithm for processing by Incremental Frequent Patterns mining Method based on RFM

```

Input : sale, old_DB(T-treeDB), new_db(T-Treedb),
Minimum Support threshold.;
Output : now_db(T-treeUD), AR_tbl;

Begin
if (not exist (now_db))
    now_db ← CRFM(sale);
    // Association Rule Mining with Minimum Support threshold
    using FP-tree based on RFM
    save (AR_tbl) ← Association_Rules_set(now_db);
end if;
if (not exist (new_db)) exit;
while (!= EOF(new_db))
{ now_db(T-treeUD) ← update (old_db ∪ new_db);
  // create now_db(T-treeUD) using FP-tree
  save (AR_tbl) ← Association_Rules_set(now_db);
  // create the Association Rules table for recommendation }
End;
    
```

III. 실험 및 성능 평가

현재 화장품은 전문적으로 판매하는 인터넷 쇼핑몰에서 RFM 기반의 점진적 빈발 패턴 마이닝 기법을 이용한 추천시스템의 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. <표5>는 식2를 이용하여 예측값의 정확성 평가를 수행한 결과이다.

표5. 제안 및 기존 시스템의 MAE에 의한 성능평가
Table5. The result for table of MAE by comparing proposal system with existing system

	P_count	Proposal	Existing
MAE	50	0,47	0,65
	100	0,23	0,32
	300	0,07	0,08
	500	0,05	0,06



그림 6. 제안 및 기존 시스템의 MAE에 의한 성능평가
Fig. 6. 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 전자상거래 환경하에서 연관규칙 탐사를 위해 과도한 데이터베이스 연산처리 부하를 수반하였다. 또한 연관규칙의 탐사시점 이후에 발생하는 트랜잭션은 전혀 탐사 대상으로 고려할 수 없었기 때문에 시간적 제약을 받아서 연관규칙 탐사를 필요로 하는 실시간 탐사의 동적 환경하에서는 연관규칙의 처리에 어려운 문제점을 갖고 있었다. 이 논문은 점진적으로

증가하는 트랜잭션 데이터의 효율적인 데이터 처리 방법을 제시하였으며 메모리 사용이 효과적인 RFM기반의 빈발 패턴 마이닝을 적용하여 연관규칙 생성시 후보집합을 생성하지 않아 기존의 Apriori 알고리즘보다 전체적으로 성능이 우수하고 탐색비용을 감소시키는 효과를 유지 할 수 있었다.

참고문헌

- [1] Jin Byeong Woon, Young Sung Cho, Keun Ho Ryu, "Personalized e-Commerce Recommendation System using RFM method and Association Rules", KSCI, 15th-12 Vol, pp 227-235, Dec, 2010
- [2] Young Sung Cho, Keun Ho Ryu, "Personalized Recommendation System using FP-tree Mining based on RFM, KSCI, 17th-2 Vol, Feb., 2012
- [3] J. Han, Y. Cai, and N. Cercone, "Data driven discovery of quantitative rules in relational databases," IEEE Transactions on Knowledge and Data Engineering, pp.29-40, 1993.
- [4] Eun Hee Kim, and Keun Ho Ryu, "A Design of false alarm analysis framework of intrusion detection system by using incremental mining method", KIPS, 13th-C Vol No3, Jun., 2006
- [5] Xin. Li, Zhi-Hong Deng, and Shiwei Tang, 2006, "A Fast Algorithm for Maintenance of Association Rules in Incremental Databases," ADMA, pp. 56-63, 2006
- [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