

FP-tree를 이용한 사용자 선호도 조사 : 헬스클럽에 적용

홍의찬*, 김원영*, 김응모*
*성균관대학교 컴퓨터공학과
e-mail:cksl12@skku.edu

User Preference survey by FP-tree : Apply to fitness club

Ue-Chan Hong*, Won-Young Kim* and Ung-Mo Kim*
*Dept of Computer Engineering, Sung-Kyun-Kwan University

요 약

오늘날 산업 전반에 걸쳐 데이터베이스 시스템이 사용되고 있다. 이에 따라 데이터베이스의 양은 점차 증가하고 방대한 양의 데이터에서 유용한 정보를 얻는 데이터 마이닝 기법이 중요한 기술로 발전해 나가고 있다. 최근 헬스클럽과 같은 개인의 취미와 여가를 위한 공간의 중요성이 대두되고 있는데 반해 사용자의 증가로 인한 불편이 문제가 되고 있다. 본 논문에서는 헬스클럽을 통해 사용자들의 선호도를 조사하여 불편을 줄이고 나아가 효율적인 시간 분배 방법을 제시해 보고자 한다.

1. 서론

오늘날 데이터베이스 시스템 사용의 보편화는 데이터의 양적 증가를 초래하게 되었다. 그 결과 더욱 다양하고 방대한 데이터가 생성되어 통제하기에 어려움이 있다. 즉, 방대한 데이터는 축적 되었지만 실제로 가치를 지니는 의미 있는 데이터를 추출하는 과정이 복잡해졌다[1]. 본 논문에서는 이러한 관련 지식들을 헬스클럽에 적용해 보고자 한다. 증가하는 사용자의 수요에 맞는 서비스가 이루어지지 않아 사용자의 불편이 늘어나고 있는 헬스클럽과 같은 곳에 효과적인 마이닝 기법을 사용하여 사용자의 선호도를 조사한 후, 보다 나은 서비스 적용 방안을 제안하려 한다. 본 논문은 다음과 같이 구성된다. 2장에서는 실제로 적용된 연관규칙 마이닝 기법을 소개하고, 3장에서는 본 논문에서 제안하는 사용자 선호도 조사를 위한 마이닝 기법에 대해 설명한다. 4장에서는 제안한 시스템에 대한 고찰 및 결론과 향후 발전 방향에 대하여 언급한다.

2. 관련연구

2.1 연관 규칙 마이닝(Association rules minning)

연관 규칙 마이닝은 주어진 데이터 집합에서 흥미로운 연관성이 있는 항목을 찾아내는 방법 중 하나로, 마케팅에서는 손님들의 장바구니에 들어있는 품목간의 관계를 알아본다는 의미에서 장바구니 분석이라고 한다[2][3].

2.2 빈발 패턴 증가 기법(Frequent Pattern Growth)

FP-Growth기법[4][5]은 연관 규칙 마이닝의 알고리즘

중 하나로 후보생성 없이 분할-정복 기법을 사용하여 빈 발향목을 찾아낸다. FP-Growth는 깊이 우선 알고리즘으로 FP-Tree라는 자료구조를 사용하는데 방대한 양의 데이터 마이닝에 적합하다.

3. 사용자 선호도 조사를 위한 마이닝 기법

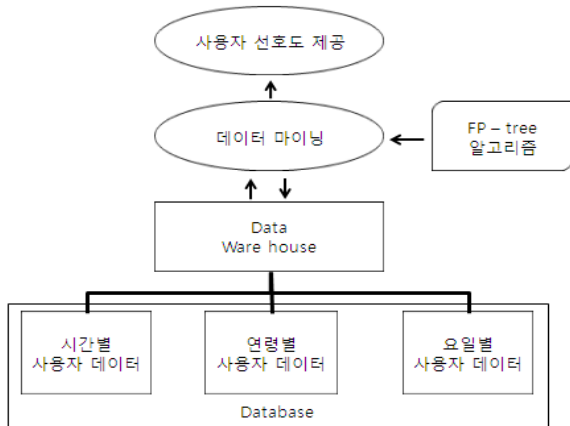
3.1 적용배경

헬스클럽은 특정 시간대에 사용자가 붐비기 때문에 증가하는 사용자에게 대해 서비스가 원활하게 제공되지 않아 사용자들의 불만이 늘어나고 있다. 또한 사용자가 드문 시간에도 헬스클럽의 시설이 무분별하게 가동되어 전기, 수도, 난방 등 오히려 낭비를 하여 큰 손실은 초래한다.

이런 상황에서 시스템 내에 축적되어 있는 사용자 데이터베이스를 이용하여 사용자가 선호하는 요일별 시간대를 추출하여 운영자에게 제공한다면 사용자의 불편은 물론이고 손실되고 있는 자원을 아낄 수 있다.

3.2 사용자 선호도 조사 시스템

사용자 데이터베이스에는 사용자 개인의 데이터가 축적되어 있다고 하자. 헬스클럽의 운영자는 방대한 양의 데이터베이스에서 사용자의 시간별, 연령별, 요일별 데이터를 정제 및 분할하여 사용자 데이터 웨어하우스를 구축한다. 선별된 데이터에 FP-tree 알고리즘을 적용하면 사용자들이 선호하는 요일별 시간대를 추출해 내어 운영계획을 세울 수 있다. 또한 연령대별로 시스템을 적용하여 연령별 사용자가 선호하는 시간을 추출하여 각 연령별 기호에 맞는 서비스를 제공할 수 있다.



(그림 1) 시스템의 구조

(그림 1)은 사용자 선호도 조사 시스템의 전체적인 흐름도이다.

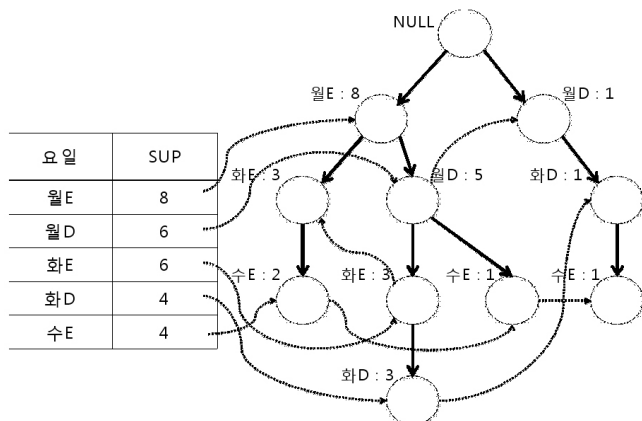
3.3 FP-tree를 이용한 마이닝 기법

본 시스템을 적용하기에 앞서 데이터 웨어하우스를 구축해야 한다. (표 1)는 실제 마이닝에 사용되는 데이터 웨어하우스에 저장된 사용자의 통합 데이터베이스이다.

User ID	Age	List of time table
100	22	월D 월E 화D 화E
200	25	월E 화E 수D 수E
300	31	월D 월E 수E 금D
400	37	월E 화E 목E 금E
500	42	월D 월E 화A 목A 금A
600	46	월A 월D 화D 수E
700	55	월D 월E 화D 화E 목E
800	60	월E 화E 수E
900	63	월D 월E 화D 화E 수D

(표 1) 연령별, 시간별 사용자의 통합 데이터베이스

(표 1)의 Time table에서 시간은 A:7시~10시, B:10시~1시, C:1시~5시, D:5시~7시, E:7시~11시로 구분하였고 토, 일요일은 사용자가 드문 관계로 제외시켰다.



(그림 2) FP-tree

(그림 2)은 (표 1)의 결과를 이용하여 만든 FP-tree이다. 위의 자료를 이용하여 사용자는 선별된 시간대 월요일E, 월요일D, 화요일E를 가장 선호한다는 것을 알아낼 수 있었다.

본 논문에서는 연관규칙을 이용해 사용자의 선호도를 조사하여 운영자에게 효율적인 서비스를 창출해 낼 수 있도록 도와주었다. 더 많은 데이터베이스를 가지고 연령대를 조사한다면 각 연령대가 선호하는 시간대를 알아낼 수 있다. 나아가 성별과 가중치를 고려하여 좀 더 다양성을 갖춘 마이닝 기법의 적용하면 운영자가 헬스클럽을 운영하는 데 큰 도움을 줄 것이다.

4. 결론 및 향후 연구과제

본 논문에서는 연관 규칙 마이닝을 이용하여 사용자가 선호하는 시간대를 운영자에게 제공할 수 있도록 제안하였다. 특정 시간대에 사용자들의 이용이 증가하여 그 시간대를 찾으므로써 운영자는 더 나은 서비스를 제공하게 되고 사용자들의 이용이 적은 시간대에는 낭비되는 자원을 아낄 수 있게 된다. 향후 연구로는 객관성과 다양성을 고려한 데이터 마이닝 기법의 연구가 진행되어야 하고 산업 전반에 걸쳐 마이닝 기법의 적용이 필요하다.

감사의 글

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No. 2009-0075771).

참고문헌

- [1] J. Han, M. Kamber "Data Mining : Concepts and Techniques" Academic Press 2000
- [2] 박중수, 유원경, 홍기형, "연관규칙탐사와 그 응용" 정보과학회지, 98년 9월.
- [3] R. Agrawal, T. Imielinski and A. Swami "Mining association rules between sets of items in large database" In Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data(SOGMOD'93), page 207-216, Washington, DC, May 1993
- [4] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation", Proc. ACM-SIGMOD Int'l Conf. Mangement of Data, pp 1-12, May 2000
- [5] J. Han, J. Pei, Y. Yin and R. Mao, "Mining Frequent Patterns without Candidate Generation : A Frequent-Pattern Tree Approach", Data Mining and Knowledg