

데이터마이닝을 이용한 축제 추천 시스템

정종훈*, 박두순*

*순천향대학교 컴퓨터소프트웨어공학과

e-mail : jjhyoro89@naver.com

Festival Recommendation System based on Data Mining

*Jong Hoon Jung, *Doo-Soon Park

*Dept. of Computer Software Engineering, SoonChunHyang University

요 약

현대 사회에서는 시간에 대한 중요성이 커지고 있다. 시간에 대한 중요성이 커지면서 현대인들은 자신의 시간을 관리하며 자신의 시간을 활용해 어떻게 여가생활을 즐길 것인지를 큰 관심사로 두고 있다. 또한 현대인들은 자가용을 이용하여 교외의 축제를 다닐 수 있는 기회가 늘어남으로써 전국의 축제를 언제든지 즐길 수 있게 되었다. 본 논문은 일정관리 어플리케이션을 이용하여 시간을 관리해주며 자신이 원하는 해당 날짜에 축제의 일정을 보여주면서 많은 축제 중 사용자에게 가장 잘 맞는 축제를 추천해주는 추천 시스템이다.

1. 서론

현대 사회에서 시간에 대한 중요성이 커지고 있다. 현대인들은 아무리 시간을 관리하고 절약해도 누구든 시간이 부족하다는 강박으로부터 자유롭지 못하다[1]. 때문에 사람들은 주중에는 회사에서 주말에는 지인들과의 약속 등을 위해 시간을 관리한다. 이렇게 시간에 대한 중요성이 커지면서 현대인들은 자신의 시간을 관리하며 자신의 남은 시간을 활용해 어떻게 여가생활을 즐길 것인지를 큰 관심사다.

또한 여가와 힐링(healing)이 대두되면서 가족단위, 연인단위의 축제를 찾는 사람들이 많아졌다. 현대인들은 자가용과 대중교통 이용하여 교외의 축제를 다닐 수 있는 기회가 늘어남으로써 전국의 축제를 언제든지 즐길 수 있게 되었다. 또한, 자신의 힐링을 위해 혹은 자신의 삶의 질을 높이기 위해 축제를 찾는 사람들도 많아지고 있다.

본 논문에서는 일정관리 어플리케이션을 이용하여 시간을 관리해주며 자신이 원하는 해당 날짜에 축제의 일정을 보여주면서 많은 축제 중 사용자의 조건과 비슷한 조건을 가진 사용자들이 가장 많이 다녀온 축제를 추천해주는 추천 시스템이다.

2. 데이터 마이닝

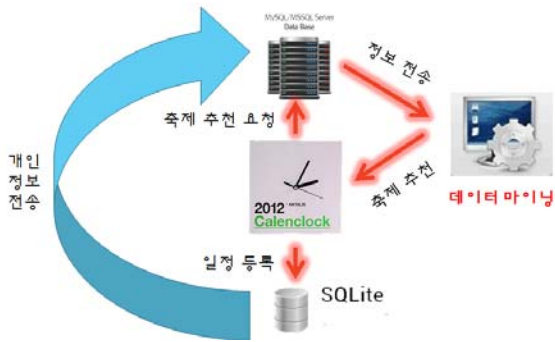
데이터 마이닝(data mining)이란 대용량의 데이터로부터 이들 데이터 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 모형화 함으로써 유용한 지식을 추출하는 일련의 과정

들이다. 통계적인 관점에서는 “대용량 데이터에 대한 탐색적 데이터 분석 (Exploratory Data Analysis)”이라고 말한다[2].

데이터 마이닝의 방법 중에 의사 결정 나무(Decision Tree)와 RFM(Recency Frequency Monetary)에 대해서 간략하게 설명한다. 의사결정나무 기법은 모델 구축 과정의 단순성과 빠른 수행능력 그리고 결과에 대한 해석이 용이함으로 인해 분류 문제에 있어서 일반적으로 많이 활용되는 데이터 마이닝 기법이다[3]. RFM은 고객의 최근 구매 시기(Recency), 구매 빈도(Frequency), 그리고 구매 금액(Monetary amount)의 세 가지 속성을 사용하여 고객의 구매 행태 및 가치를 평가하는 방법이다. 구매 시기는 고객의 최종 구매일로부터 기준일까지의 지난 시간(예를 들어, 날짜 수)이고, 구매 빈도는 일정 기간 동안의 총 구매 또는 평균 구매 건수이고, 구매 금액은 일정 기간 동안의 총 구매 또는 평균 구매 금액을 말한다[3].

3. 축제 추천 시스템

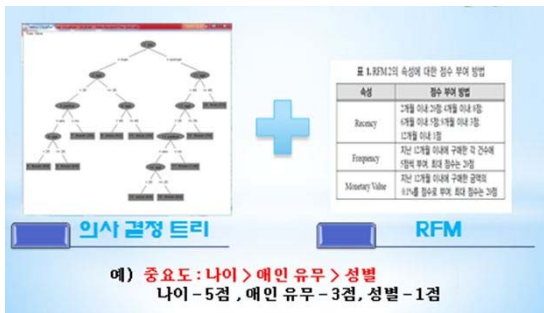
축제를 추천 받기 위해서는 특별한 가입 절차는 존재하지 않는다. 사용자는 일정관리(달력) 어플리케이션을 사용하면서 메뉴속성에서 자신의 정보를 선택하며 자신의 정보를 언제든지 손쉽게 변경할 수 있다. (그림 1)는 추천 시스템의 추천 과정이다.



(그림 1) 축제 추천 과정

사용자가 일정을 등록하고 축제일정을 보면서 자신이 원하는 축제 정보를 보게 되면 해당 사용자의 정보와 축제가 서버 데이터 베이스에 저장하게 된다. 사용자가 축제 추천 요청을 하게 되면 서버는 지금까지의 정보들을 검색하여 가장 많은 사람들이 다녀온 축제를 추천해준다.

(그림2)는 사용자에게 축제를 추천하는 방법이다. 각각의 조건들을 의사결정나무를 이용하여 중요도를 검사하고 각각의 중요도에 따라 RFM 방식으로 점수를 선정하여 다녀온 사용자들의 명수와 속성들의 점수를 곱하여 가장 높은 점수를 가진 축제를 추천하게 된다.



(그림 2) 축제 추천 방법

(그림 3)과 (그림 4)는 초기 데이터베이스에 등록되는 데이터로 주위의 연령대 별 설문을 통해 얻은 데이터들이다.

name	startdate	enddate	adress
2013 진주남강유등축제	2013.10.01	2013.10.13	경남 진주시 남강
제10회 자라섬국악	2013.10.03	2013.10.06	경기 가평군 자라섬
2013 북천 코스모스	2013.09.21	2013.10.06	경남 하동군 직전면
제34회 근로자문화예술제	2013.02.01	2013.11.31	서울 영등포구 KB
제8회 서울국제문화축제	2013.09.29	2013.10.09	서울 중구 남산한남

(그림 3) 축제 데이터 베이스

nan	wonan	20age	30age	40age	50age	partner	nopartner	number
1	0	1	0	0	0	1	0	4
1	0	1	0	0	0	1	0	4
1	0	1	0	0	0	1	0	3
1	0	1	0	0	0	1	0	7
1	0	1	0	0	0	1	0	7
1	0	1	0	0	0	1	0	7

(그림 4) 서버 데이터 베이스

서버 데이터베이스와 어플리케이션의 연동을 위해 XML 파싱을 사용한다. 각 속성 별로 저장된 정보들을 xml파일에 프린트하게 되고 해당 xml파일을 어플리케이션에서 xml파일을 읽어 속성들의 정보를 읽을 수 있다.

(그림 5), (그림 6)는 XML파싱을 통해 얻어진 축제 정보를 리스트화 해주고 해당 축제를 선택하면 축제의 정보를 보여 주는 화면이다. 리스트는 점수의 순위별로 정렬된다. 축제 정보에서는 축제기간과 주소, 설명, 홈페이지가 링크 되어있다.



(그림 5) 축제 리스트 (그림 6) 축제 정보

4. 결론

본 논문은 사용자의 시간을 관리해주며 전국에 존재하는 많은 축제를 중 사용자의 가장 잘 맞는 축제를 추천해주는 시스템이다.

앞으로의 연구 목표로는 달력 기능에 사용자의 편의를 위한 알람 기능을 추가하는 것과 본 논문에서의 사용자의 정보속성 외에 보다 더 많은 정보속성을 이용하여 보다 정확한 축제를 사용자에게 추천할 수 있어야 될 것으로 보인다. 또한 현재의 축제 범위가 전국의 모든 축제를 추천해주지만 사용자의 위치에 따라 가까운 위치의 축제를 추천해 주어 사용자에게 더 알맞은 축제를 추천해 줌으로써 좀 더 완벽한 추천 시스템으로 만들어야 할 것이다.

참고문헌

- [1] 김영선, “현대사회의 시간회소성과 여가변동” 호원논집 12호 2004
- [2] 박차석 “대전보건대학 보건의료정보과” 2009
- [3] 이재식,이진척, “데이터 마이닝 기법을 이용한 모바일게임 구매 추천시스템의 개발” 2010