

데이터 마이닝을 이용한 맞춤형 데이트 코스 추천 서비스

송우용*, 김원영*, 김응모*
*성균관대학교 정보통신공학부

e-mail:cowdragon@naver.com

Classified Recommendation Service of Date Course using Data Mining

Woo-Yong Song*, Won-Young Kim*, Ung-Mo kim*
*School of Information and Communication Engineering,
Sungkyunkwan University

요 약

컴퓨터와 통신의 급속한 성장은 방대한 양의 정보를 서로 공유하는 정보화 시대를 출현 시켰고 이러한 많은 양의 다양한 정보로부터 유용한 정보를 얻어 내는 데이터 마이닝이라는 기법이 도입되었다. 데이터 마이닝 기법은 사회 모든 분야에 걸쳐서 사용되고 있으며 이러한 기법으로 산출된 새로운 정보는 각 분야의 의사결정을 하는데 있어서 중요한 요소로 자리 잡고 있다.

본 논문에서는 데이터 마이닝을 통하여 여가 생활의 하나인 데이트를 보다 의미 있는 시간으로 만들기 위한 개인별 맞춤형 데이트 코스 추천 서비스를 제안하고자 한다. 이를 통하여 개인의 정보를 얻기 위한 시간과 노력을 절약하고 개인의 취향과 환경적인 요소를 고려한 특화된 서비스를 제공한다.

1. 서론

최근 현대 사회에서 여가 시간이 크게 늘어남에 따라서 여가 시간의 효율적인 활용에 대한 관심이 높아지고 있다. 현재 여가 생활에 필요한 정보를 얻기 위한 방법으로 인터넷과 매체를 통한 정보의 습득이 주된 수단으로 자리 잡고 있다. 하지만 이러한 정보의 제공은 다수의 사용자들을 위한 공통된 정보를 제공할 뿐 개인의 취향과 환경을 고려한 특화된 정보를 제공하는 데에는 한계가 있다.

본 논문에서는 데이터 마이닝을 이용한 개인별 맞춤형 데이트 코스 추천 서비스를 제안한다. 지금까지의 의사결정은 잡지나 인터넷 정보, TV 방송 등을 통해 얻어진 정보로 결정되는 것이 대부분이다. 이를 통해 얻은 정보는 개인별 취향이 반영되지 않아 그 정보를 이용하는 사용자의 특성과 맞지 않을 수도 있으며 기후나 환경적인 요소를 고려하지 않은 단편적인 정보에 불과하다. 마이닝 기법을 통해 성별, 나이, 직업, 취미 등의 개인정보를 이용하여 유형별 선호 패턴으로 도출하면 각 사용자의 특성에 맞는 서비스를 제공할 수 있게 된다. 또한 기후나 기타 환경적인 요소도 결합시켜 보다 효율적인 의사결정을 할 수 있도록 돕고자 한다.

본 논문은 2장에서 관련 연구를 알아보고 3장에서 제안된 데이트 코스 추천 서비스 모델에 대해 설명하며 4장에서 결론을 맺고자 한다.

2. 관련 연구

2.1 연관규칙

연관 규칙은 데이터베이스에 존재하는 여러 항목(item)들 간의 연관성을 탐색하여 의미 있는 패턴들을 결과적으로 찾아내는 것을 말한다. 연관규칙은 A→B의 형태로 표현한다. 예를 들어, "item A가 구매된 경우, item B도 구매됨"이라고 해석된다[1].

2.2 연속규칙

일정 시간 동안 레코드를 분석하여 순서 패턴(Sequential Pattern)[2] 을 찾아낸다. 시간이 흐름에 따라 발생하는 패턴을 발견해 내는 작업이다. 즉 개인별 트랜잭션 이력 데이터를 시계열적으로 분석하여 트랜잭션의 향후 발생 가능성을 예측하는 작업이다.

2.3 인구 통계학 추천(Demographic-Filtering)

인구 통계학적 추천 기법은 사용자의 인구-통계적 정보, 즉 나이, 성별 직업 등의 자료를 이용하여 해당 그룹의 추천 결과를 사용자에게 제공하는 기법이다. 인구 통계학적 추천에 기반 한 추천에는 같은 연령대로부터의 추천, 영화의 박스 오피스와 같은 불특정 다수로부터의 추천 등이 있다[3].

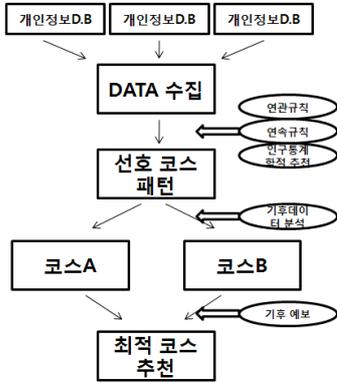
2.4 기후 데이터의 군집 분석

기후데이터를 저장 및 관리하기 위한 데이터를 모델링하고, k-means 기법을 이용하여 국내 기후 데이터를 지역적

특성에 따라 군집함으로써 체계적인 데이터 정보를 제공한다. 이를 통하여 날씨 마케팅에 사용된다[4].

3. 데이트 코스 추천 서비스 모델

본 논문에서 제안하는 추천 서비스 모델은 (그림1)과 같다.



(그림1) 추천 서비스 모델

(그림1)에서 개인별 Survey를 통한 개인정보를 획득한 후 이를 D.B에 저장한다. 이렇게 저장된 정보를 수집하여 연관규칙, 연속규칙, 인구통계학적 추천을 통하여 유형별 선호 패턴 코스를 추출하게 된다. 그리고 기후 데이터 분석에 의한 코스 분류를 통하여 사용자에게 최적의 데이트 코스를 추천하게 된다.

3.1 연관규칙 적용

예를 들어 연령이 20대이며 직업이 학생이고 사는 곳이 강남이며 취미가 영화 관람인 사용자가 추천 서비스를 이용한다. 시스템에서는 같은 연령, 직업, 취미인 사용자들에 의해 검색된 최근 5일간의 데이터를 수집하게 된다.

<표1>의 수집된 데이터를 이용하여 연관 규칙을 적용한다.

ID	Age	데이트코스
100	23	강남CGV, A음식점, B노래방, C음식점
200	21	A음식점, 강남CGV, D백화점, C음식점
300	26	삼성동Mega Box, D백화점, B노래방
400	20	강남CGV, A음식점, F공원
500	22	강남CGV, F공원, B노래방, D백화점

<표 1>

최소 지지도를 3이라고 하였을 때 이것을 만족하는 모든 항목 집합을 찾아 빈발 항목집합을 추출하여 이로부터 연관 규칙을 생성한다.

Item	Count
강남CGV	4
A음식점	3
B노래방	3
C음식점	2
D백화점	3
Mega box	1
F공원	2

Itemset	Count
강남CGV, A음식점	3
강남CGV, B노래방	3
강남CGV, D백화점	2
A음식점, B노래방	1
A음식점, D백화점	1
B노래방, D백화점	2

Itemset	Count
강남CGV, A음식점, B노래방	3

(그림2)

3.2 연속규칙 및 인구통계학적 추천 적용

연속규칙은 시간의 흐름에 따라서 발생하는 패턴을 발견한다. 술집을 다녀온 다음에는 노래방을 가는 경향이 있다. 인구 통계학적 추천은 35살의 회사원인 여성이 휴일에 서비스 검색을 통해 인구통계학적 정보에 의한 데이트 코스를 추천 받는다.

3.3. 기후 데이터의 군집 분석 적용 후 최적 데이트 코스 추천

국내 기후 데이터를 지역적 특성에 따라 군집함으로써 날씨 마케팅에 적용한다. 가령, 비오는 날 막걸리 집을 추천한다거나 화창한 날 유원지 및 공원을 추천할 수 있다. 최종적으로 마이닝된 선호 패턴 코스를 기후에 따라서 최적의 코스로 제공하게 된다.

4. 결론

이 논문에서 제안한 맞춤형 데이트 코스 추천 서비스는 개인별로 선호하는 코스 정보를 도출하도록 하였다. 또한 기후 데이터를 이용하여 날씨에 따른 최적의 시나리오를 추천하게 된다. 이를 통하여 사용자는 적은 시간과 노력으로 의사결정에 큰 도움을 받을 것이며 서비스 제공자 또한 이를 이용한 패키지 상품 등을 개발하여 마케팅에 사용할 수 있을 것이다. 향후 다양한 환경에 적합한 시스템으로 발전하게 위해서는 더 많은 연구가 필요할 것이다.

감사의 글

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

참고문헌

[1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rule," Proc. Int'l Conf. Very Large Data Bases, pp. 487-499, Sept. 1994

[2] Rakesh Agrawal; Ramakrishnan Srikant, "Mining Sequential Patterns", 1p, 1995

[3] M.J. Pazzani. A framework for Collaborative, "Content-based and Demographic Filtering", Artificial Intelligence Review, 13(5-6) : 393-408, 1999

[4] 이양구 외 4명 "날씨 마케팅 적용을 위한 기후 데이터의 군집 분석" 2005. 12. 한국공간정보시스템학회 논문지 제7권 제3호 통권 15호 pp.33-44