

데이터 마이닝 기반 추천 시스템에 관한 연구 분석

최은혜, 김성수*, 정태선*
아주대학교 대학원 소프트웨어특성화학과
*아주대학교 컴퓨터공학과 교수
e-mail : gracechoi@ajou.ac.kr

Analysis of recommendation system based on data mining

Eun-Hye Choi, Sung-Soo Kim*, Tae-Sun Chung*
Graduate School of Software, Ajou University
*Professor of Computer Engineering, Ajou University

요 약

온라인 서비스와 스마트 기기의 발달로 언제 어디서나 인터넷에 접속할 수 있는 시대가 도래되었다. 수많은 콘텐츠와 서비스가 쏟아져 사용자 입장에서 자신이 선호하는 콘텐츠를 자신이 원할 때 전달받는 것이 필요해졌다. 즉, 사용자의 선호도에 따라 콘텐츠를 추천하는 시스템이 현재 실생활에서도 활용되고 있다는 뜻이다. 이를 근거로 대용량의 데이터를 다루는 마이닝 기법 기반의 추천 시스템인 협업 필터링 추천기법과 내용기반 추천기법의 개념과 문제점들을 분석해 보았다.

1. 서론

최근 현대사회는 IT 기술의 발달로 인터넷을 통해 다양한 정보들을 얻을 수 있게 되었다. 하지만 넘치는 정보의 홍수 속에서 사용자가 필요로 하는 정보만을 분류하기 위해서는 어느 정도의 노력이 필요하다. 한편 온라인 상에서 사용자의 검색 패턴들을 분석해 편의를 제공하는 것이 추천 시스템 (recommendation system)이다. 추천 시스템이란 사용자가 아직 평가하지 않은 아이템들 중 사용자의 과거 정보를 사용하여 가장 선호할 것 같은 아이템을 사용자에게 권하는 시스템이다 [1]. 이는 자동적인 정보 제공을 위해서 예측 기능이 필요하며, 이를 위하여 데이터 마이닝 기법이 사용될 수 있다 [2]. 즉, 사용자에 대한 다양한 정보를 모아 기록하고, 그 기록을 토대로 여러 개의 보기를 사용자에게 추천을 함으로써 좀 더 쉽게 사용자 스스로가 원하는 것을 선택하게끔 도움을 주는 것이다.

본 논문에서는 2 장 데이터 마이닝에 대한 개념에 대해 살펴보고 3 장 추천 시스템과 현존하는 기법들, 4 장 결론을 내린다.

2. 관련연구

데이터 마이닝은 대용량의 데이터 속에서 유용한 정보나 패턴, 규칙 등을 찾아내는 기술을 의미한다. 데이터 마이닝은 다양한 분야에서 활용되고 있다. 친화성 분석, 분류, 예측 등 다양한 마이닝 기법을 바탕으로 [3] 정보 간의 연관성을 파악함으로써 이익을 극대화시킬 수 있었다. 군사분야에서는 미사일의 정확도에 영향을 주는 요인들을 찾기 위해, 천체 분야에서는 행성과 성운을 식별하기 위해, 제조업에서는

품질관리를 위해, 그리고 의학연구자들은 암의 재발 가능성을 예측하기 위해 데이터 마이닝을 사용한다.

3. 추천 시스템

사용자가 생각하는 순위나 선호도를 예측하여 최적의 아이템을 필터링하는 일종의 정보 검색 시스템이다 [4]. 추천 대상은 게임, 책, 영화, 음악 등 정보화 사회 속에서 다양하고, 현재 실생활에서도 활발히 이용되고 있어 필요성이 높아지고 있다.

3.1 협업 필터링 추천기법

협업 필터링 추천기법은 가장 많이 사용되고 있는 방식으로, 많은 사용자들로부터 얻은 기호정보(taste information)에 따라 사용자들의 관심사들을 자동적으로 예측하게 해주는 방법이다. 협력 필터링 접근법의 근본적인 가정은 사용자들의 과거의 경향이 미래에서도 그대로 유지 될 것이라는 전제에 있다. 유사한 취향의 사용자들에게 서로 아직 구매하지 않은 상품을 교차 추천하거나 분류된 사용자의 취향이나 생활 형태에 따라 관련 상품을 추천하는 형태의 서비스를 제공한다. 일반적으로 협력 필터링에 대한 두 가지 이슈는 초기 평가자 문제와 콜드 스타트 문제이다. 초기 평가자 문제는 어느 사용자에게도 평가 받지 못한 새로운 상품이 있을 때이고 콜드 스타트 문제는 지금까지 한 번도 평가를 하지 않은 사용자들이 있을 경우이다. 두 가지 경우 모두 상품에 대한 추천을 할 수가 없게 된다. 그 외에도 협업 필터링의 단점은 기호 가치들에만 의존함으로 콘텐츠 기반의 필터링과 비교하여 실제 상품의 내용과 어떤 관계인지 알지 못한다는 것이다 [5].

3.2 내용기반 추천기법

내용기반 추천 기법은 정보 검색이나 정보 필터링 연구에서 발전된 방식으로 멀티미디어 데이터에서 속성들을 분석, 분류하고 이를 기존 사용자의 선호도 기반으로 추천해주는 방법이다. 이 기법은 정보검색에 기반을 두고 있으므로 이 분야에서 자주 사용되는 기법인 가중치 기법, 적합성 피드백, 불리언 검색, 확률검색 모형 등을 활용한다 [6]. 유용한 특징을 추출하는 것은 쉽지 않을뿐더러 사용자가 이미 평가한 상품과 유사한 상품만을 제공하는 경향이 있다 [7-9].

4. 결론

네트워크 기술의 발달로 현대인의 일상생활은 보다 편리해 졌으며 기존의 생활구조를 획기적으로 변화시키고 있다. 스마트폰을 켜서 사용자를 확인만 하면 책과 음악, 영화 등을 그 자리에서 바로 추천 받을 수 있다. 본 논문에서는 데이터 마이닝을 이용한 추천 기법 두 가지를 소개하였다. 만약 정확성이 낮아 추천 시스템으로부터 사용자 입장의 필요한 정보를 제공받지 못한다면 사용자는 더 이상 추천 시스템을 사용하지 않게 될 것이다. 이를 개선하기 위해 기존의 연구는 물론 향후 추천의 정확성 향상을 위한 데이터 마이닝 기법 연구가 필요하다.

참고문헌

- [1] Adomavicius, G., Tuzhilin, A., "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering, TKDE, Vol.17, No.6, pp.734-749, 2005.
- [2] 이선아, 강성원, "소프트웨어 진화 작업을 위한 데이터 마이닝 기반 추천 시스템", 정보과학회지 제 31 권 제 10 호, 2013.
- [3] Han, J., M. Kamber, "Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems)", 2011.
- [4] 전련화, 김연정, 김보현, 이민수, "소셜 네트워크 기반의 개인 맞춤형 정보추천 기법", 정보과학회 논문지: 컴퓨팅의 실제 및 레터, 제 19 권 제 12 호, 2013.
- [5] Huang, Z., H. Chen, D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering", ACM Transactions on Information Systems, Vol.22, p116~142, 2004.
- [6] 김현주, 이광석, "관련성 분포정보를 이용한 사용자 선호도 추천시스템", 한국지식정보기술학회 논문지, No.1, 제 5 권 제 5 호, 2010.
- [7] 안명환, 권준희, "개념 계층 모델을 이용한 온톨로지 기반 상황 인식 추천 서비스", 인터넷정보학회, 제 8 권 제 5 호, 81-89 쪽, 2007.
- [8] 한동조, 박대영, 최기호, "사용자 상황을 이용한 추천 서비스 시스템의 필터링 기법에 관한 연구", 한국학회 논문지, 제 8 권 제 1 호, 119-126 쪽 2009.
- [9] 정영미, 이용구, "필터링 기법을 이용한 도서 추천 시스템 구축", 정보관리연구, 제 23 권 제 1 호, 1-17 쪽, 2002.