

대리자를 이용한 군집화 기반 사용자 적응적 추천 모델*

류상현, 송창환, 장현수, 엄영익
성균관 대학교 전자전기컴퓨터공학과
e-mail:{shryu, eerien, jhs4071, yieom}@ece.skku.ac.kr

User Adaptive Recommendation Model Based on User Clustering using Proxies

Sanghyun Ryu, Changhwan Song, Hyunsu Jang, Young Ik Eom
Dept. of Electrical and Computer Engineering, SungKyunKwan University

요 약

사용자 적응형 추천 시스템의 목적은 사용자의 선호도와 행동 정보 등을 분석, 분류하여 그를 바탕으로 각 사용자가 필요로 하거나 선호 할 만한 서비스를 사용자에게 추천하여 사용자 편리성을 높이는 것이다. 그러나 기존의 추천 시스템은 새로운 사용자의 등장이나 새로운 서비스의 등장 시 분석에 많은 시간을 필요로 하거나, 과특성화와 희귀성이라는 특성으로 인한 추천 서비스 단순화 등의 문제점을 안고 있다. 본 논문에서는 새로운 사용자 등장 시 결정 트리를 이용한 분류로 분석시간을 줄이고, 새로운 아이템의 등장 시 분석시간의 감소와 다양한 사용자 중심적인 추천을 위해 대리자를 이용한 사용자 군집화와 추천을 수행하는 새로운 모델을 제시한다. 또한 제안된 모델을 분석하여 위의 문제점들이 어떻게 해결되는지 설명한다.

1. 서론

최근 정보통신기술의 급속한 발전과 함께 사용자 중심의 서비스를 제공하는 사용자 적응형 추천 시스템의 개발이 활발히 이루어지고 있다. 이러한 사용자 적응형 추천 시스템의 목적은 사용자의 선호도와 행동 정보 등을 분석, 분류하여 이를 바탕으로 각 사용자에게 적합한 서비스를 구성하고 제안하는 것이다.

이러한 시스템에서 사용자의 정보를 분석하기 위한 방법으로는 명시적 방법과 암시적 방법이 있다. 명시적 방법은 사용자가 추천 시스템에 등록할 때 자신의 관심사항이나 개인정보를 직접 입력하는 것이고, 암시적 방법은 사용자의 행동 정보를 수집하여 이를 기반으로 사용자의 관심사항을 추론하는 것이다. 이후 추론된 사용자의 관심사항을 기반으로 서비스 추천이 이루어진다. 대표적인 추천방법의 종류로는 사용자의 관심 정보와 과거에 사용했던 서비스의 유사도를 비교하여 추천하는 콘텐츠 기반 추천방법(Content-based Recommend Method)[1]과 유사한 취향을 가진 사용자들 간에 상호추천을 수행하는 협력적 추천방법(Collaborative Recommend Method)[2]이 있다.

위에서 언급된 추천방법에는 다음과 같은 단점들이 존재한다.

첫째로 새로운 사용자가 나타날 경우, 사용자를 분석하

는데 시간이 소모된다. 둘째로 콘텐츠 기반 추천방법에서 나타나는 과특성화(Overspecialization)로 인해 추천 서비스의 제한이 발생한다. 셋째로 협력적 추천방법 새로운 서비스가 등장 할 때 문제점이 나타난다. 새롭게 등장한 서비스가 다른 사용자에게 추천되기 위해서는 다른 사용자들의 평가가 미리 이루어져야 한다. 또한 이러한 추천방법에는 희귀성(Sparsity)이라는 특성이 존재하여 일정량의 사용자 평가를 내리지 않으면 추천 되지 않는다.

이러한 문제점들을 해결하기 위하여 콘텐츠 기반 추천방법과 협력적 추천방법을 합쳐 사용하는 혼합형 추천방법(Hybrid Recommend Method)이 나타났다.

본 논문에서는 위에서 설명한 추천 모델들의 단점들을 보완하기 위해 결정 트리를 이용한 분류와 대리자를 이용한 사용자 군집화와 추천을 수행하는 모델을 제안한다. 이 모델은 크게 결정 트리를 이용한 분류와 대리자를 이용한 사용자 군집화와 추천이라는 두 부분으로 이루어진다. 이를 이용하여 사용자의 간단한 참여만으로 다른 추천방법보다 정확하게 사용자를 분석하고 추천 할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 관련된 연구들을 소개하고 3장에서는 제안 모델의 구조와 전체적인 동작과정에 대해서 설명한다. 4장에서는 결론 및 향후 연구를 설명한다.

2. 관련 연구

전통적인 추천방법의 종류로는 사용자의 관심 정보와

* 본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원 사업의 연구결과로 수행되었음 (IITA-2009-(C1090-0902-0046))

과거에 사용했던 서비스의 유사도를 비교하여 추천하는 콘텐츠 기반 추천방법(Content-based Recommend Method)[1]과 유사한 취향을 가진 사용자들 간에 상호추천을 수행하는 협력적 추천방법(Collaborative Recommend Method)[2]이 있다. 이런 추천방법들에는 다음과 같은 단점들이 존재한다.

첫째로 새로운 사용자가 나타날 경우, 일반적으로 사용자는 자신의 선호도를 세밀하게 정의하고 분류하는 것을 꺼린다. 사용자가 명시적으로 자신의 선호도를 정의한다면 쉽게 해결 될 수 있겠지만 그렇지 않다면 추천 시스템이 사용자를 분석하는데 적지 않은 시간이 소모된다. 이러한 단점은 시스템 전체에 대한 신뢰를 떨어뜨리고, 사용자에게 적합한 추천이 이루어지지 못하게 되는 요인이 된다.

둘째로 콘텐츠 기반 추천방법에서 나타나는 과특성화(Overspecialization)로 인해 추천 서비스의 제한이 발생한다. 과특성화란 사용자의 과거 정보를 바탕으로 하여 사용자 선호도를 분석하기 때문에 발생하는 특성이다. 이러한 특성은 사용자가 이용했던 서비스와 관계가 먼 서비스이거나 사용자가 사용했던 서비스와 흡사한 서비스들이 추천되지 않는 원인이 된다. 예를 들어, 이러한 단점을 가진 추천시스템은 일정 기간 문서 작성 작업만 수행한 사용자에게 문서작성과 거리가 먼 서비스들은 추천되지 않거나 새로운 문서작성에 대하여 추천하지 않는다.

셋째로 협력적 추천방법 새로운 서비스가 등장 할 때 문제점이 나타난다. 새롭게 등장한 서비스가 다른 사용자에게 추천되기 위해서는 다른 사용자들의 평가가 미리 이루어져야 한다. 이러한 평가가 이루어지기 위해서는 많은 시간을 필요로 하고 사용자들의 참여 역시 필요하다. 이러한 추천방법에는 희귀성(Sparsity)이라는 특성이 존재한다. 희귀성이란 서비스가 추천되기 위해서는 일정량 이상의 사용자 평가가 필요 하지만 그 양이 채워지지 않는다면 아무리 좋은 평가를 받은 서비스라도 다른 사용자에게 추천 되지 않는다는 것을 의미한다. 또한 다른 사용자들이 좋지 않게 평가했지만 특정 사용자에 한해 요구되는 서비스의 경우, 협력적 추천 기법으로는 그 서비스가 특정 사용자에게 추천되기는 어렵다.

이러한 문제점들을 해결하기 위하여 콘텐츠 기반 추천방법과 협력적 추천방법을 합쳐 사용하는 혼합형 추천방법(Hybrid Recommend Method)이 나타났다. 혼합형 추천방법은 다음과 같이 분류된다.

첫째로 콘텐츠 기반 추천방법과 협력적 추천방법의 결과를 이용하여 추천 하는 방법이다[3]. 이 방법은 다시 두 가지로 분류할 수 있다. 첫 번째는 각각의 결과를 합치는 방법이고 둘째는 두 결과 중 어떤 결과가 현재 상황에 더 적합한지를 결정하여 추천하는 방법이다.

둘째로 협력적 추천 방법에 콘텐츠 기반 추천방법의 특성을 더하는 방법이다. 이에 해당하는 대부분의 방법은 전통적인 협력적 추천방법에 콘텐츠 기반 추천방법의 사용자 정보 분석법을 이용한다. 이 방법은 협력적 추천 방

법의 단점 중 하나인 희귀성을 해결 할 수 있다고 알려져 있다.

세 번째는 가장 많이 사용되는 방법으로 콘텐츠 기반 추천방법에 협력적 방법의 특성을 더하는 방법이다.

마지막으로 독립된 추천 모델을 개발하는 방법이 있으며[4], 최근 이러한 방향으로 많은 연구가 진행되고 있다.

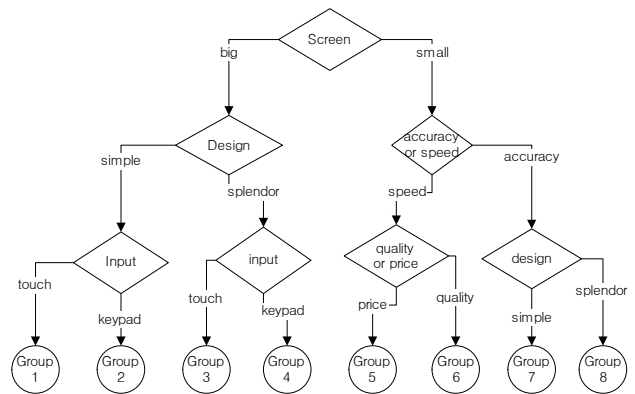
위의 방법들보다 더 확장된 추천 시스템을 설계하기 위하여 다음과 같은 주제의 연구가 요구된다[5].

- 좀더 넓은 사용자의 이해
- 모델 기반의 추천방법의 확장
- 추천의 다차원화
- 다양한 기준의 등급화
- 강제성의 제한
- 유연성의 증가
- 추천 시스템의 효과에 대한 평가방법 개선

본 논문에서는 결정 트리 모델을 이용하는 분류와 대리자를 이용한 사용자 군집화와 추천을 수행 하는 사용자 적응적 추천모델을 제시하여 기존 추천방법들의 단점을 극복하고자 하였다.

3. 제안 모델

제안 모델은 결정 트리를 이용한 분류와 대리자를 이용한 사용자 군집화와 추천이라는 두 부분으로 나누어진다. 결정 트리를 이용한 분류에서는 새로운 사용자 등장 시 발생하는 문제를 해결한다. 대리자를 이용한 사용자 군집화와 추천에서는 새로운 서비스의 등장 및 추천의 단순화 문제를 해결한다.



(그림 1) 제안 모델의 기본 구조

3.1 결정 트리를 이용한 분류

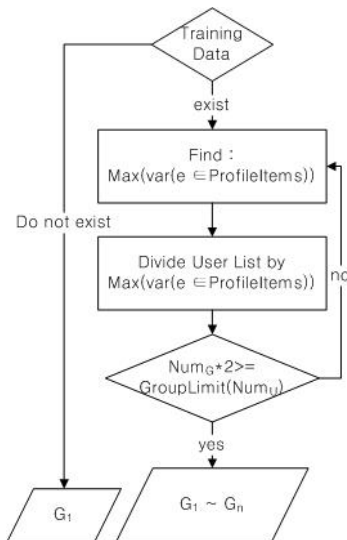
결정 트리를 기반으로 하는 구조는 여러 추천 모델의 공통적인 단점으로 꼽히는 새로운 사용자의 선호도 분석에 소요되는 시간을 감소시킨다.

결정 트리의 구성을 위해서는 사용자의 선호도 정보를 담고 있는 트레이닝 데이터가 필요하다. 트레이닝 데이터가 존재 할 경우에 결정 트리는 다음과 같은 과정으로 생

성된다. 결정 트리의 각 분기는 트레이닝 데이터에서 사용자들의 선호도 정보들 중 두 분류로 나누기에 가장 적절한 자료를 선택하여 분기한다. 사용자의 수에 따라 이러한 분기를 일정 횟수 반복하고 분류가 완료된 사용자들은 각각 하나의 그룹을 구성한다.

결정 트리에서 각 분기 항목의 선택은 매우 중요하며 분기 항목 선택 방법에 따라 전체적인 사용자 그룹화의 성능이 결정된다. 본 논문에서는 사용자 선호도 정보의 각 항목들의 분산도를 계산하고, 분산도가 가장 높은 항목을 분기 항목으로 설정한다.

트레이닝 데이터가 존재하지 않는다면 임계값 이상의 사용자가 확보될 때까지 결정 트리의 구성을 미룬다. 사용자의 수가 임계값을 초과하면 결정 트리를 구성하고 주기적으로 결정 트리를 재구성한다. 사용자가 늘어날수록 결정 트리의 재구성 주기를 점차적으로 줄인다. 주기적인 결정 트리의 재구성은 사용자들의 분류 정확도를 높인다. 또한 다음 절에서 설명할 사용자의 주기적인 이동과 더불어 사용자에게 대한 지속적인 분석과 사용자 적응적 추천을 가능하게 한다.



(그림 2) 트리 초기 생성 과정

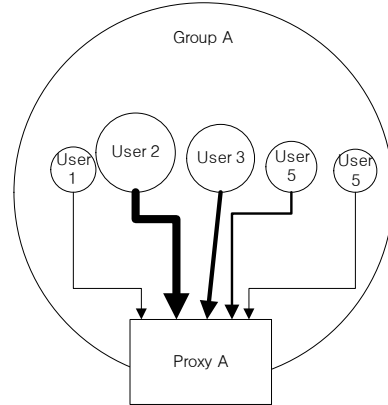
그림 2는 트리의 초기 생성 과정을 나타낸다. 초기 트레이닝 데이터가 존재 한다면, 프로파일의 항목 중 가장 높은 분산치를 찾아 그 값으로 나눈다. 그 후 그룹을 한번 더 나눈 그룹의 수가 임계 그룹수보다 작다면 다시 위의 과정을 반복하고 크다면 그대로 그룹을 생성한다.

3.2 대리자를 이용한 사용자 군집화와 추천

이 절에서는 결정 트리를 이용하여 분류된 그룹들 간의 추천과 사용자의 이동에 대하여 다룬다. 각각의 그룹은 비슷한 선호도를 지닌 여러 사용자로 구성된다. 이런 여러 사용자들의 선호도를 통합하기 위해서 그룹 내에 대리자가 존재한다.

이 대리자는 그룹 내 사용자들의 선호도 정보를 통합

하여 자신의 선호도를 생성한다. 그러나 이 과정에서 그룹 사용자들의 선호도를 단순히 더하면 해당 그룹에 속하지만 실제로 사용하지 않는 사용자들의 특성까지 모두 더해지게 된다. 이 같은 문제를 방지하기 위해 사용자의 활동량에 따라 가중치를 두어 대리자 선호도에 영향을 미치게 된다. 즉, 서비스 평가 활동을 활발히 하는 사용자는 활동이 적은 사용자보다 대리자 선호도에 더 큰 영향을 미치게 되는 것이다.



(그림 3) 그룹 내의 영향도 차이

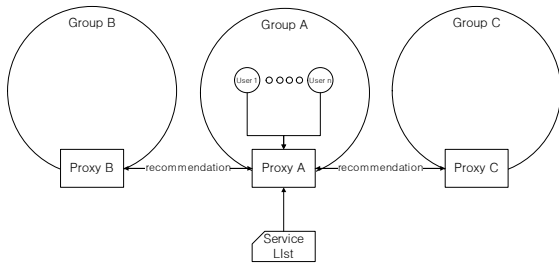
그림 3에서 사용자들의 활동 정도는 원의 크기 정도로 표현되어있고 영향력의 크기는 선의 굵기 정도로 표현된다.

이렇게 생성된 대리자 선호도는 다음과 같이 사용된다. 첫째로 새로운 아이템의 등장 시 대리자는 그 아이템에 대하여 자동적으로 대리자 선호도 기반의 평가를 내린다. 임시적 평가치는 사용자들이 서비스를 이용하고 평가를 내림으로서 영향도를 줄이고, 대리자의 선호도 역시 재조정된다. 즉, 대리자는 선호도를 사용자들의 평가에 따라 지속적으로 재조정하고, 결과적으로 그룹 내 사용자들의 선호도 변화에 적응한다.

둘째로 다른 그룹의 대리자들에게 필요한 추천 서비스를 요청하거나 요청 받을 때 이용한다. 사용자가 필요한 서비스를 요청하였을 때, 그룹 내에 해당 서비스에 대한 추천 정보가 없다면 자신과 가장 가까운 단말노드들을 순회하며 필요한 서비스를 검색한다. 반대로 다른 대리자가 필요한 서비스를 요청하였을 때, 대리자는 요구 대리자의 선호도와 비교하여 유사도가 임계 수치 이상일 때 자신이 속한 그룹에서 사용되거나 추천된 서비스를 추천한다.

또한 대리자는 사용자 추천 서비스의 다양성 증가를 위해 자신의 선호도 항목 중 하나를 반대로 설정하고 루트노드부터 순회한다. 결과적으로 현재의 선호도와 단 하나의 항목에서 차이를 보이는 그룹으로 이동하게 된다. 이런 그룹은 사용자에게 추천 되지 않으나 사용자가 선호할만한 서비스를 지니고 있을 수 있고 이런 그룹을 순회함으로써 해당 서비스들을 추천 받을 수 있다.

대리자를 이용한 사용자 군집화의 다른 특징은 사용자



(그림 4) 제안 모델의 대리자를 이용한
군집화와 추천

의 이동이다. 대리자는 항상 자신이 속한 그룹의 사용자들의 선호도와 자신의 선호도를 비교하여 선호도 차이 정도를 확인한다. 이러한 차이 정도가 임계 수치 이상으로 커지는 경우 대리자는 사용자를 그룹에서 제외시킨다. 해당 사용자는 새로 등장한 사용자와 마찬가지로 자신에게 적합한 그룹을 찾아서 결정 트리의 첫 번째 분기로 돌아가 분류를 위한 질문에 답하고 새로 속할 그룹을 결정한다. 이러한 과정에서도 다시 같은 그룹으로 결정 된다면 사용자는 현재 그룹에서 가장 적은 이동비용으로 이동 할 수 있는 그룹으로 이동된다.

3.1절에서 설명한 대로 각각의 그룹들은 주위 그룹들과 친밀한 관계를 맺고 있다. 그 중 가장 작은 이동 비용으로 이동 할 수 있는 그룹들은 가장 비슷한 선호도를 나타내는 이웃 그룹들을 의미한다. 이런 방법으로 사용자는 유사한 그룹을 이동하여 선호도에 적합한 그룹을 찾아 서비스를 추천받게 된다.

4. 모델 분석

컨텐츠 기반 추천 방법은 새로운 사용자가 등장하면 사용자에게 대한 정보가 충분히 수집되기 전까지는 추천 서비스를 제공할 수 없다. 이 문제를 해결하기 위해 제안 모델에서는 새로 등장한 사용자를 이미 구성되어 있는 결정 트리에 따라 적절한 그룹으로 분류하여 추천서비스를 제공한다. 따라서 제안 모델은 컨텐츠 기반 추천 방법에 비해 사용자의 적은 명시적 활동으로도 추천 서비스를 제공할 수 있다.

이와 유사하게 협력적 추천방법에서는 새로 등장한 서비스에 대하여 평가가 부족해 사용자에게 추천되지 않는 문제점이 발생한다. 이를 위해 그룹 대리자는 사용자를 대신하여 새로 등장한 서비스를 평가하고 추천함으로써 이 문제를 해결한다. 또한 대리자를 통한 추천에서 현재의 선호도와 단 하나의 항목이 다른 그룹을 이용해 추천하여 서비스 추천의 범위를 넓힌다. 이로써 그룹 내에서 발생하는 과특성화를 해결할 수 있다. 이러한 그룹단위의 추천은 적은 비용으로 사용자에게 적합한 서비스를 추천할 수 있는 장점이 있다.

제안 모델의 한 그룹 내에는 많은 사용자들이 포함 되어 있고, 사용자들 중 그룹에 적절하지 않은 사용자가 포함되어 있을 수 있다. 즉, 한 그룹 내에서 소수의 다른 취

향을 가지는 유저들이 존재 한다면 그러한 유저들은 자신들이 원하는 서비스의 추천을 받지 못하게 된다. 이러한 상황을 해결하기 위하여 제안 모델은 지속적으로 사용자를 이동시키고 트리를 재구성한다.

5. 결론 및 향후 연구

본 논문에서는 기존의 대표적인 추천 시스템에 대하여 알아보고 기존 추천 시스템의 단점을 보완한 모델을 제안하였다. 제안 모델은 결정 트리를 이용한 사용자 분류와 대리자를 이용한 사용자 군집화와 추천으로 구성된다. 결정 트리를 이용한 사용자 분류에서는 결정 트리의 분기를 이용한 분류를 사용하여 새로운 사용자 분석에 걸리는 시간을 감소시켰다. 대리자를 이용한 사용자 군집화와 추천에서는 사용자가 새로운 서비스에 대해 평가하지 못해 나타나는 문제점을 대리자가 서비스를 미리 평가하여 해결하였다. 또한 지속적인 사용자 분석을 위해 결정 트리의 재구성과 대리자를 이용한 사용자의 재 이동을 이용하였다.

향후 연구 과제로는 결정 트리의 분기 항목 선택과 그룹 내 대리자와 사용자 사이의 수학적 관계를 정의하고 제안 모델을 구현하여 그 효율성을 실험적으로 확인하는 연구를 수행할 것이다.

참고문헌

- [1] M. Balabanovic and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation," Comm. ACM, Vol. 40, No. 3, 1997.
- [2] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, "Recommending and Evaluating Chices in a Virtual Community of Use," Proc. Conf. Human Factors in Computing Systems, 1995.
- [3] T. Tran and R. Cohen, "Hybrid Recommender Systems for Electronic Commerce," AAAI 2000, Technical Report WS-00-04, 2000.
- [4] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as Classification: Using Social and Content-Based Information in Recommendation," AAAI 1998, Technical Report WS-98-08, 1998.
- [5] G. Adomavicius and A. Tuzhilin "Toward the Next Generation of Recommender Systems", IEEE Transactions on Knowledge and Data Engineering Vol. 17, No. 6 June 2005.