

컨텐츠 계층구조를 이용한 평가정보 자동 수집방법

이준훈⁰, 김영지, 문현정, 우용태
창원대학교 컴퓨터공학과

tophuni@ce.changwon.ac.kr, {yjkim, mun, ytwoo}@sarim.changwon.ac.kr

A Method for Implicit Rating Information Collection using Content Hierarchy

Joon-Hoon Lee⁰, Young-Ji Kim, Hyeonjeong Mun, Yong-Tae Woo
Dept. of Computer Engineering, Changwon National University

요 약

전자상거래에서 추천시스템은 사용자들의 관심도에 따라 사용자에게 개인화된 아이템이나 상품을 제안한다. 보통의 추천시스템은 추천의 정확성을 높이기 위하여 사용자로부터 명시적으로 수집한 평가정보를 이용하였다. 그러나 명시적인 평가정보 수집방법은 사용자로부터 충분한 평가정보를 제공받지 못하여 추천이 어려울 수 있다. 최근에는 명시적으로 평가정보를 수집하지 않고 묵시적으로 평가정보를 수집하는 추천시스템에 관한 연구가 활발하게 진행되고 있다. 이러한 묵시적인 평가정보의 장점은 모든 사용자에게 대한 평가정보를 자동적으로 수집할 수 있으며, 사용자는 정보를 이용하는 것 이외의 부가적인 일을 수행할 필요가 없다는 점이다. 본 논문에서는 인터넷사이트에서 계층적으로 구성된 컨텐츠에 대한 사용자의 단계적인 반응도에 따라 자동적으로 평가정보를 수집하기 위한 기법을 제안하고 효율을 측정하였다.

1. 서 론

최근에 전자상거래나 컨텐츠를 유료로 제공하는 상용 사이트가 늘어나면서 인터넷 비즈니스 분야의 경쟁이 치열하게 전개되고 있으며, 경쟁력 강화를 위한 전략의 하나로 eCRM 시스템을 활발하게 도입하고 있다. eCRM 시스템에서 개인화(Personalization)는 사용자별로 차별화된 맞춤 서비스를 제공하기 위한 핵심적인 요소 중의 하나이다. 개인화는 고객의 선호도에 따라 상품이나 컨텐츠를 효과적으로 추천하여 인터넷사이트에 대한 만족도를 높이고, 고객과의 친밀한 관계를 지속시키기 위한 중요한 전략의 하나이다. 추천시스템은 이런 전략을 가능하게 한다.

보통의 추천시스템은 추천의 정확성을 높이기 위하여 사용자로부터 명시적으로 수집한 평가정보를 이용하였다. 그러나 명시적인 평가정보 수집방법은 사용자로부터 충분한 평가정보를 제공받지 못할 수 있고, 평가정보가 지속적으로 업데이트 되기 어려운 문제점으로 인하여 추천효율이 떨어질 수 있다.

최근에는 명시적으로 평가정보를 수집하지 않고 서버 로그나 사용자의 사이트 이용패턴을 분석하여 자동적으로 평가정보를 수집하는 추천시스템에 관한 연구가 활발하게 진행되고 있다. 이러한 묵시적인 평가정보의 장점은 모든 사용자에게 대한 평가정보를 자동적으로 수집할 수 있으며, 사용자는 정보를 이용하는 것 이외의 부가적인 일을 수행할 필요가 없다는 점이다.

본 논문에서는 인터넷사이트에서 계층적으로 구성된 컨텐츠에 대한 사용자의 단계적인 반응도에 따라 자동적으로 평가정보를 수집하기 위한 기법을 제안하고 효율을 측정하였다.

2. 묵시적 평가정보

초기의 협동적필터링에 기반한 추천시스템은 추천의 정확도를 높이기 위하여 주로 명시적인 평가정보를 이용하였다[1]. 그러나 사용자가 직접 평가한 명시적인 평가정보만으로는 추천 효율을 향상시키는데 어려움이 있었다. 즉, 명시적으로 평가정보를 수집하는 방법은 사용자로부터 충분한 평가정보를 제공받지 못할 수 있고, 평가정보가 지속적으로 업데이트 되기 어려운 문제점으로 인하여 추천 효율이 떨어질 수 있다.

최근에는 묵시적인 평가정보를 이용한 협동적필터링 기법의 문제점을 개선하기 위하여 묵시적으로 평가정보를 수집하는 방법에 대한 연구가 활발히 진행되고 있다. 묵시적인 평가정보는 사용자가 인터넷사이트를 이용하는 동안에 남겨진 구매이력이나 서버로그 정보 등을 사용자 선호도의 평가정보로 사용하는 방법을 말한다. 묵시적인 평가정보의 장점은 모든 사용자로부터 평가정보를 자동적으로 수집할 수 있어서 평가정보의 지속적인 업데이트가 가능하다. 또한 사용자들은 컨텐츠를 이용하는 것 이외의 부가적인 일은 하지 않아도 되는 이점이 있다.

Nichols 등은 인터넷사이트에서 정보필터링 프로그램을 위한 묵시적인 평가정보의 이익과 비용에 대해 연구하였다. Nichols는 묵시적인 평가정보 수집을 위해 사용자가 전체 아이템을 읽거나 아이템을 저장, 북마크, 프린트하는 액션을 평가기준으로 분류하였다. 그러나 이 연구에서는 묵시적인 평가정보의 가능성만 제시하였다[2].

Oard와 Kim은 Nichols[2]에 의해 분류된 묵시적인 평가정보를 아이템에 대한 사용자의 행동에 따라 'Examination', 'Retention' 그리고 'Reference'의 3가지의 카테고리로 나누었다[3]. Morita 등은 Usenet News에서 기사를 읽는 시간과 사용자 선호도의 상관관계에 대해 연구하였다. 뿐만아니라 Usenet News 기사를 보면서 발생하는 저장, 전송 등의 행동도 묵시적인 평가정보로 사용할 수 있다는 가능성을 제시하였다[4].

본 연구는 정보통신부 대학기초연구지원사업(과제번호 : 2001-111-2)에 의해 수행되었음.

Konstant는 GroupLens 시스템에서 Usenet News에서 명시적인 평가정보와 Usenet News 기사를 읽는 시간과의 상관관계에 대해 연구하였다. Claypool은 Nichols[2], Oard와 Kim[3]에 의해 분류된 목시적인 평가정보를 확장하여 사용자들의 클릭 스트림(click-stream), 마우스 스크롤 시간, 마우스를 움직이는 시간, 마우스의 클릭 횟수, 콘텐츠를 읽는데 소비하는 시간 등의 목시적인 평가정보와 명시적 평가정보의 상관관계를 분석하였다[4]. Rafter, Smyth와 Bradley는 콘텐츠의 클릭 스트림, 콘텐츠를 읽는데 소비하는 시간, 사용자의 재방문 횟수를 이용한 목시적인 평가정보와 명시적인 평가정보의 상관관계를 분석하였다[5].

최근에 많이 연구되고 있는 목시적인 평가정보 수집방법은 주로 콘텐츠를 이용하는 시간을 평가기준으로 사용하고 있다[4]. 그러나 콘텐츠를 읽는 시간을 이용한 평가정보는 사용자의 네트워크 환경이나 콘텐츠의 길이에 따라 다르게 측정할 수 있는 관계로 평가정보에 대한 기준이 명확하지 않을 수 있다. 그리고 서버로그 정보를 이용하는 방법은 시작시간 정보만을 가지고 있으므로 사용자가 이용하는 시간과 이용하지 않는 시간을 정확히 구분하기 어렵다. 또한 사용 종료시간을 알기 위해서는 이것을 구분하기 위한 부가적인 프로그램이 필요하다. 그리고 사용자가 콘텐츠를 이용하면서 반응하는 행동패턴을 평가정보로 수집하는 방법은 가능성만 제시되었고 구체적인 추천효율에 대한 연구는 진행되지 않았다.

따라서 명시적인 평가정보와 거의 유사한 추천효율을 가지면서 사용자의 콘텐츠에 대한 반응에 따라 자동적으로 평가정보 수집을 할 수 있는 방법이 필요하다. 또한 동적으로 변하는 사용자의 콘텐츠 선호도를 지속적으로 반영할 수 있는 기법들에 대한 연구가 필요하다.

3. 콘텐츠 계층구조를 이용한 평가정보 자동 수집방법

3.1 콘텐츠 계층구조를 이용한 평가정보 자동 수집

본 논문에서는 인터넷사이트에서 계층적으로 구성된 콘텐츠에 대한 사용자의 단계적인 반응도에 따라 자동적으로 평가정보를 수집하기 위한 새로운 방법을 제안하였다. 여기서 콘텐츠에 대한 고객의 반응도는 계층적으로 구성된 콘텐츠에 대한 접속 레벨에 따라 차등적으로 점수화하여 평가정보를 자동으로 수집한다.

일반적으로 전자상거래 사이트의 메뉴는 계층적으로 구성되어 있다. 또한 콘텐츠 역시 제목, 요약정보, 상세정보 그리고 최종적으로 장바구니에 담기식의 여러 단계로 구성된다. 따라서 사용자가 콘텐츠를 이용하는 각 레벨은 해당 콘텐츠에 대한 선호도의 차이를 의미한다. 즉, 사용자는 상품이나 콘텐츠에 대해 요약정보를 열람한 후 관심이 있을 경우 상세정보를 열람하여 더욱 상세한 정보를 얻기를 희망할 것이다. 상세정보를 열람하여 콘텐츠나 상품이 정말 자신에게 유용하거나 흥미가 있는 것이라면 스크랩 또는 프린터를 하거나 상품을 구매할 것이다. 본 논문에서는 사용자가 접속하는 각 레벨을 점수화하여 콘텐츠에 대한 사용자의 관심도를 반영한 평가정보를 자동적으로 수집할 수 있는 방법을 제시하였다.

콘텐츠 계층구조를 이용한 평가정보 자동 수집방법의 구성은 자동 평가정보 수집모듈과 평가정보 데이터베이스로 이루어진다. 자동 평가정보 수집모듈은 사용자가 콘텐츠를 이용시 각 접속레벨 및 행동에 의한 목시적인 평가정보를 자동으로 점수화하여 평가정보 데이터베이스에 저장한다. 그리고 명시적인 평가정보도 사용자로부터 수집하여 평가정보 데이터베이스에 저장한다. 평가정보 데이터베이스에 저장되는 평가정보 데이터는 사용자에게 맞춤형서비스를 제공하기 위한 데이터로 정보를 이용하는 동안에 저장되는 목시적인 평가정보와 사용자가 직접 평가한 명시적인 평가정보가 저장된다. 그림 1은 본 논문에서 제안한 콘텐츠 계층구조를 이용한 평가정보 자동 수집방법의 프로세싱 개념도이다.

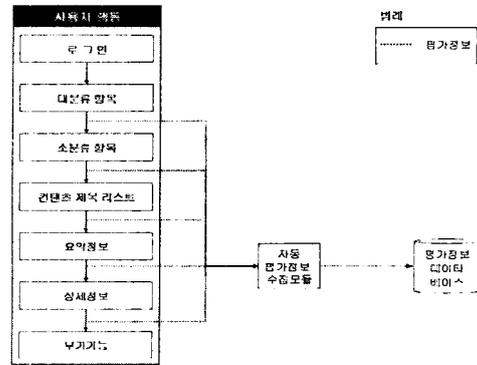


그림 1 평가정보 자동 수집 프로세싱 개념도

3.2 자동 평가정보 수집모듈

본 논문에서는 콘텐츠 계층구조를 이용한 평가정보를 수집하기 위한 자동 평가정보 수집모듈이 있다. 자동 평가정보 수집모듈은 사용자가 콘텐츠를 이용하는 동안에 분류항목 및 콘텐츠에 발생하는 이벤트를 분석하고, 이벤트를 평가정보의 값으로 변환하여 평가정보 데이터베이스에 저장하기 위한 모듈이다.

자동 평가정보 수집모듈은 사용자가 콘텐츠에 대해 이벤트를 발생하였을 경우 우선 이벤트가 명시적인 평가정보인지 목시적인 평가정보인지를 분석한다. 분석 후 명시적인 평가정보이면 명시적인 평가정보를 데이터베이스 입력문에 명시적 평가정보 코드와 평가정보를 넘겨준다. 넘겨받은 평가정보를 다른 부가적인 데이터와 함께 평가정보 데이터베이스에 저장한다. 또한 목시적인 평가정보이면 이벤트의 코드를 분석하여 대분류항목(A-1)이면 평가정보 1점을 데이터베이스 입력문에 넘겨주고 데이터베이스 입력문은 평가정보 1점을 다른 부가적인 데이터와 함께 평가정보 데이터베이스에 저장한다. 이벤트 코드를 분석하여 소분류항목(A-2)이면 2점, 요약정보 열람(A-3)이면 3점, 상세정보 열람(A-4)이면 4점, 부가기능 사용(A-5)이면 5점을 데이터베이스 입력문에 넘겨주어 부가적인 데이터와 함께 평가정보를 평가정보 데이터베이스에 저장한다.

4. 실험 및 결과분석

본 논문에서 실험데이터로 사용한 8,208명의 사용자 중에서 명시적인 평가정보를 제공한 사용자는 1,366명이였다. 이는 전체사용자 중의 13.4%의 사용자만 명시적인 평가정보를 제공한 것이다. 또한 사용자가 이용한 콘텐츠 142,008건 중에서 명시적인 평가정보를 제공한 콘텐츠 건수는 2,089건이였다. 이는 전체 이용 콘텐츠의 1.5%만 명시적인 평가정보가 제공되였다. 또한 각 사용자가 콘텐츠를 본 수는 평균 17.3건 정도였으며 그 중 명시적으로 평가정보를 제공한 건수는 평균 1.5건이였다. 이는 명시적인 평가정보를 사용하는 경우가 나타나 는 문제점으로는 명시적인 평가정보의 희소성으로 인해 추천의 어려움이 발생하는 경우와 평가정보의 지속적인 업데이트가 이루어지지 않아서 추천의 어려움이 발생하는 경우가 나타날 수 있다. 따라서 명시적인 평가정보는 평가정보를 제공하기 위해 강제성을 취하지 않으면 평가정보의 지속적인 업데이트가 이루어지기 어렵다는 것을 보여주고 있다. 또한 지속적인 업데이트가 이루어지지 않으면 사용자들에게 맞춤 정보를 제공하기 어렵다는 것을 보여주고 있다. 또한 명시적인 평가정보를 제공한 사람들 중에서 사용자의 관심도에 상관없이 일괄적인 평가정보를 제공하는 사용자들도 나타났다.

본 연구에서 제안한 방법에 의해 수집된 평가정보와 명시적인 평가정보 사이의 상관관계를 분석하기 위한 실험을 하였다. 실험에서 명시적인 평가정보에 따른 제안한 방법의 평가정보의 상관관계수 값은 0.753이였다. 이는 명시적인 평가정보에 따른 제안한 방법의 평가정보의 상관관계는 밀접한 상관관계에 있음을 알 수 있었다. 이 제안한 방법의 평가정보가 나타내는 결과는 기존에 연구되었던 콘텐츠의 이용시간을 묵시적인 평가기준으로 사용하여 평가정보를 수집하는 것보다 명시적인 평가정보에 더 가깝게 평가정보를 수집할 수 있다. 제안한 방법은 자동으로 평가정보를 수집함으로써 명시적 평가정보의 문제점으로 지적되고 있는 사용자가 평가정보를 제공하지 않는 것이나 지속적인 업데이트 문제를 개선할 수 있을 것이라 생각한다.

다음 그림 2는 명시적인 평가정보에 따른 제안한 방법으로 수집한 평가정보를 나타낸 것이다.

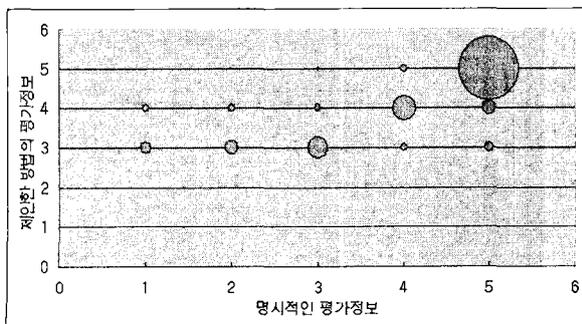


그림 2 명시적인 평가정보에 따른 제안한 방법의 평가정보

5. 결론 및 향후과제

본 논문에서는 인터넷사이트에서 계층적으로 구성된 콘텐츠에 대한 사용자의 단계적인 반응도에 따라 자동적으로 평가정보를 수집하기 위한 기법을 제안하고 효율을 측정하였다.

본 논문에서 제안한 자동 평가정보 수집기법은 명시적인 평가정보와 밀접한 상관관계가 있음을 볼 수 있었다. 또한 명시적인 평가정보의 문제점을 개선할 수 있었다. 향후 본 논문에서 제안한 묵시적인 평가정보 수집방법을 개선할 수 있는 방법에 관한 연구가 필요하다. 또한 사용자들의 불편을 초래하는 명시적인 평가정보 수집과정을 자동화하여 보다 지능적인 개인화 전략을 적용한 효율적인 eCRM 시스템 개발이 가능하리라 예상된다.

참 고 문 헌

[1] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," Proc. of the ACM 1994 Conference on Computer Supported Cooperative Work, pp.175-186, 1994
 [2] D. M. Nichols, "Implicit Rating and Filtering," In Proc. of the Fifth DELOS Workshop on Filtering and Collaborative Filtering, pp.31-36, Nov. 1997
 [3] D. Oard, J. Kim, "Implicit Feedback for Recommender Systems," Proc. of the AAAI Workshop on Recommender Systems, pp.80-82 1998
 [4] M. Morita, Y. Shinoda, "Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval," Proce. of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, pp.272-281, 1994
 [5] J. Konstant, B. Miller, D. Maltz, J. Herlocker, L. Gordong and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News," Communications of the ACM, Vol.40, No.3, pp.77-87, 1997
 [6] M. Claypool, D. Brown, P. Le and M. Waseda, "Inferring User Interest," Technical Report WPI-CS-TR-01-97, 2001
 [7] R. Rafter, B. Smyth and K. Bradley, "Inferring Relevance Feedback from Server Logs : A Case Study in Online Recruitment," Proc. of the 11th Irish Conference on Artificial Intelligence and Cognitive Science, 2000