

무선인터넷 환경하의 지능형 뉴스 추천 시스템 설계와 유효성 검증

복준영*, 최준연**, 황중선***

*고려대학교 컴퓨터과학기술대학원

**SK 텔레콤(주)

***고려대학교 컴퓨터과학기술대학원

e-mail : jaybok64@korea.ac.kr

Intelligent News Recommendation System Design and Evaluation under Wireless Internet Environment

Jun-young Bok*, Joon-yeon Choi**, Jong-sun Hwang***

*Graduate School of Computer Science & Technology, Korea University

**SK Telecom Corporation

***Dept. of Computer Science & Technology, Korea University

요 약

최근 Device 의 대중화로 모바일 폰 사용자는 언제 어디서라도 정보서비스 및 멀티미디어 동영상 서비스를 쉽게 접할 수 있게 되었다. 하지만, 작은 화면과 불편한 User Interface 를 가진 모바일 Device 의 한계와 양적인 콘텐츠의 증가로 인하여 사용자가 원하는 정보를 Access 하기에는 너무 많은 시간과 비용이 소요되고 있다. 본 논문에서는 사용자의 Accessibility 욕구를 충족하고 시간과 비용을 절감하기 위한 방법으로 무선 인터넷 환경에서의 추천 시스템을 기반으로 방대한 분량의 뉴스 콘텐츠 가운데 개인화 된 뉴스를 선별, 제공해 주는 지능형 추천 시스템을 설계 구현하고자 한다. 본 논문은 인공 지능 분야 중 에이전트의 추천 역할과 기능면에서 Collaborative Filtering 방법이 갖고 있는 계산 속도 문제에 대한 대안으로서, 유사한 선호도의 사용자들 특성에 따라 분류하기 위해 군집분석을 이용했으며, 각 군집별로 뉴스 카테고리를 Weighting 한 후, 뉴스 콘텐츠를 Scoring, Listing 하여 사용자가 읽은 뉴스에 대한 평가를 기록하고 이를 각 군집별로 반영하는 방법을 수행하였다. 추가로 뉴스 콘텐츠의 사용자 만족도를 평가하기 위하여 뉴스를 처음부터 끝까지 읽은 완독률을 측정기준으로 제시하고 일반적인 무선 인터넷 환경하의 뉴스 메뉴체계와 비교하여 설계된 시스템의 성능에 대한 유효성을 검증하였다.

1. 서론

최근 Device 대중화로 인하여 모바일 폰 사용자는 전문적인 정보서비스 및 멀티미디어 동영상 서비스를 쉽게 접하고 활용하고 있다. 그러나, 인터넷상의 정보를 쉽게 취득하고 활용하는 것과는 달리 무선 인터넷 환경에서는 Device 별로 User Interface 의 규격과 제공 기능의 차이로 인한 한계성 및 콘텐츠의 양적 증가로 인한 정보 Access 의 시간과 비용 등 비효율성 문제가 있다.

이러한 상황에서, 사용자 개인별로 사용자가 찾는 정보를 찾아내고, 찾아낸 정보를 그 필요에 가장 부합하는 순서대로 순위화 해주는 개인화 추천 시스템은 훌륭한 해답이 될 수 있다[1]. 유선 인터넷 환경에서 개인화 추천 시스템을 활용하고 있는 아마존과는 다르게 무선 인터넷 환경에서는 개인의 선호도, 취향에 따른 콘텐츠를 찾기 위해서는 속도 및 비용문제를 해결해야 할 필요성이 있다. 이러한 문제를 해결하기 위해 본 논문에서는 방대한 분량의 무선 데이터 뉴스 콘텐츠를 대상으로 지능형 뉴스 추천 시스템을 구현

하고 이에 대한 유효성을 검증코자 한다. 본 논문의 구성은 다음과 같다. 2 장에서는 본 논문에서 구현한 시스템에 관한 관련연구 활동을 알아보고, 3 장에서는 무선 환경에서의 뉴스 추천 시스템 설계 및 유효성 검증을 위한 완독률 등 평가 기준에 대한 내용을, 4 장에서는 실질적인 테스트 수행 내용과 그 결과를 알아 보고, 5 장에서 결론 및 향후 연구과제를 제시하였다.

2. 관련 연구

2.1. 추천 시스템

추천 시스템은 사용자가 어떤 정보에 관심이 있는지, 어떤 정보가 유용한지를 예측하는 일에 초점을 맞춘다. 추천 시스템의 예측은 사용자 개인의 프로파일을 이용함으로써 개인화 된다. 이러한 추천 시스템은 초기에는 주로 내용기반 필터링 기법(content-based filtering technique)을 이용하였으나 최근에는 자동화된 협업 필터링(automated collaborative filtering technique)이 많이 이용되고 있다[2].

일반적으로 추천 시스템은 정보검색 시스템이나 정보 필터링 시스템의 일부분으로 구축되는데, 추천 시스템의 방법으로 협업 필터링 방식을 사용하면 다른 사용자들의 평가 정보를 이용하여 동적으로 연관 링크를 제공할 수 있다는 장점이 있다[3].

본 논문에서는 방대한 분량의 뉴스 콘텐츠 가운데 사용자별 선호도가 높을 것 같은 뉴스만을 선택적으로 제공할 수 있도록 사용자가 원하는 뉴스에 대한 평가 정보를 이용, 필요한 Searching Cost 를 절감하기 위하여 추천 시스템을 사용하였다.

2.2. 협업 필터링 기법

협업 필터링 기법은 추천 시스템에 가장 흔히 사용되는 기법으로서, 자동화된 협업 필터링 기법이 내용기반 정보 필터링 시스템을 보완하는 기법으로 최근 각광을 받고 있다.

협업 필터링 방식의 장점은 아이템에 대한 사용자의 선호도를 수집하고, 같은 취향이나 정보요구를 갖는 사용자들의 선호도와 의견을 바탕으로 특정 사용자의 기호에 맞는 아이템을 추천함으로써 개인화 된 추천을 제공할 수 있다[2]. 그러나, 이는 현재 사용자와 유사한 선호도를 갖는 사용자를 선택하기 위해 전체 사용자와의 해당 아이템별 유사도 계산을 수행해야 하는 과정의 복잡성과 이로 인한 비효율성의 문제가 있다. 또한 사용자 프로파일은 시간이 지남에 따라 그 효율성이 감소하여 사용자에게 오류 정보를 제공할 확률이 높다[3] [6].

2.3. 군집분석

군집분석은 주어진 관찰 값 중에서 유사한 것들을 몇몇의 집단으로 그룹화 하여, 각 집단의 성격을 파악해 데이터 전체의 구조에 대한 이해를 돕고자 하는 분석 방법이다. 대용량의 데이터에서 개별 관찰치를 요약하는 것 보다는 전체를 유사한 관찰 값들의

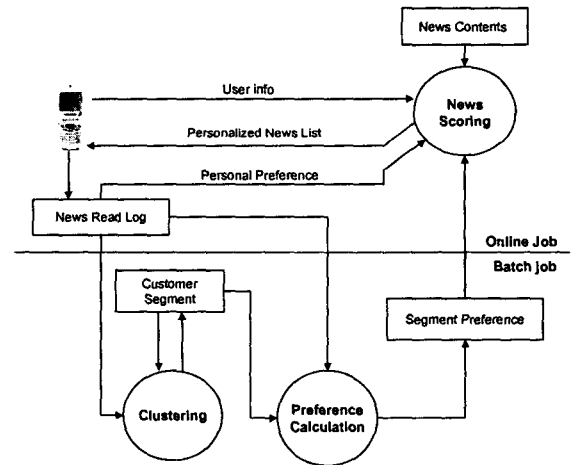
Cluster 로 구분, 복잡한 전체보다는 그를 잘 대표하는 군집들을 관찰함으로써 전체 데이터에 대한 의미 있는 정보를 얻어 낼 수 있다[4]. 군집화를 수행하는 이유는 현재 사용자와 유사한 선호도를 갖는 다른 이웃한 사용자를 선택함으로써 관련 아이템을 예측하기 위해 전체 사용자를 탐색 대상으로 설정하기보다는 사용자가 소속되어 있는 군집만을 탐색하도록 함으로써 검색의 소요시간을 단축하는데 있다[5].

따라서, 본 논문에서는 뉴스 콘텐츠 사용자 개인의 선호도 및 취향 등 효과적인 프로파일 분석을 위하여 군집분석을 사용하였다.

3. 본론

본 논문에서는 무선 인터넷 환경에서 사용자의 선호도와 콘텐츠의 가치변화를 고려한 지능형 뉴스 추천 시스템을 제안하고, 제안된 시스템의 성능을 기존의 콘텐츠 제공방식과 비교 분석하고자 한다. 뉴스 추천 시스템은 성향이 유사한 사용자를 엮어주는 'Clustering', 각 Segment 의 선호도를 계산하는 'Preference Calculation', 최종적인 개별 뉴스의 가치를 계산해주는 'News Scoring' 세가지 처리과정으로 구성되어 있으며, 전체 시스템 구조도는 그림 1 과 같다.

그림 1. 지능형 뉴스 추천 시스템 구조도



전체 시스템은 실시간으로 이루어지는 Online-Job 과 정해진 주기로 작업이 이루어지는 Batch-Job 으로 나뉜다. 'News Read Log'에서는 사용자가 어떤 뉴스를 언제 읽었는지 페이지 단위로 기록이 남으며, 'Customer Segment' 에는 뉴스 선호도 성향이 비슷한 사용자 그룹정보가 남게 된다. 'Segment Preference'에는 각 Segment 별 뉴스 선호도 정보가 계산되어 저장된다. 'News Scoring'에서는 Segment 의 선호도, 개인별 선호도, 뉴스의 속성, 사용자정보를 입력 받아 최종적인 뉴스의 가치를 계산해주어 사용자에게 보여줄 뉴스를 결정하게 된다.

3.1 Clustering

뉴스를 Scoring 하기 위하여 Unsupervised 군집분석을 사용하여 사용자의 인구통계학적 정보, 과거 뉴스를 본 조회 기록, 과거 뉴스 콘텐츠 접속기록을 분석하여 비슷한 성향의 사용자끼리 Grouping 하였다. 첫 방문 사용자는 뉴스 로그 기록이 없으므로, 연령과 성별이 가장 유사한 Segmentation 에 배치하였다.

3.2 Preference Calculation

사용자의 뉴스에 대한 선호도는 뉴스 카테고리별 기준으로 계산된다. 뉴스의 카테고리는 대분류, 중분류, 소분류의 3-depth 로 구성되는데, 모든 카테고리가 소분류까지 갖는 것은 아니며, 대분류나 중분류까지만 존재하는 카테고리도 있다. 그러므로, 본 연구에서는 하위에 child-node 를 갖고 있지 않은 leaf-node 카테고리 75 개를 기준으로 사용자의 선호도를 계산하였다. 또한 시간에 따라 변하는 사용자의 선호도 변화를 반영하기 위해 단기선호도와 장기선호도로 별도로 계산하여 이 값들의 차이를 반영시켰다.

개별 Segment 별로 뉴스 카테고리에 대한 선호도를 계산하는데 뉴스의 속성은 뉴스를 제공하는 콘텐츠 제공자로부터 입력되며, 구성된 항목은 표 1 과 같다.

표 1. 뉴스 속성표

항목	설명	비고
idx	뉴스 일련번호	
ADATE	뉴스등록년/월/일	
ATIME	뉴스등록시/분	
ACODE1	대분류 코드	
ACODE2	중분류 코드	
ACODE3	소분류 코드	
ATTITLE	뉴스 제목	
ASUB_TITLE	뉴스 부제목	null 가능
ATEXT	뉴스 본문	
AREG_TIME	뉴스 추천시간(년/월/일/시/분)	보지 처리

3.3 뉴스 Scoring

각 뉴스 아이템의 개별 사용자에 대한 최종적인 중요도가 결정되는 단계이며, 이는 Score 로 계산된다. 앞에서 구한 Segment Preference 외의 시간(뉴스가 생성된 뒤 지난 시간), 뉴스의 우선순위, 개인의 선호도 변화 등을 참조하여 각 사용자에 대한 개별 뉴스 중요도를 계산하였다. 계산 조건은 뉴스의 우선순위가 높을수록 Score 가 높아지며 해당 사용자가 속한 Segment 의 Category Weight 에 비례한다. 뉴스의 가치는 시간이 지날수록 감소하여 24 시간이 지나면 0 가 되거나 더 이상 추천하지 않는다. 또한 최근 관심도가 증가된 Category 의 Score 는 높아지며 동일한 Segment 사용자가 선택한 뉴스는 같은 Segment 의 타 사용자에게 높은 Score 를 부여하였다. 계산 식은 아래 같이 표현될 수 있다.

$$Score = f_1(news_priority) \cdot f_2(segment_pref) \cdot f_3(time_posted) \cdot f_4(pref_change) \cdot f_5(individual_pref)$$

$$f_3(time_posted) = -\left(\frac{time_present - time_posted}{86400}\right)^2 + 1$$

$$f_4(pref_change) = \log_{10}\left(\frac{user_weight_{id,q,s,days}}{user_weight_{id,q,l,days}}\right) \left((10^{DF} - 1) + 1\right)$$

$$f_5(individual_pref) = 1 + \frac{r_{p,news_id}}{\sum_{1day} r_{p,news_id}}$$

4. 유효성 검증 및 분석

제안된 추천 시스템의 효과를 검증하기 위해, 본 논문에서는 뉴스를 끝까지 읽은 비율을 의미하는 완독률을 측정도구로 도입하였다. 완독률은 하나의 뉴스를 여러 개의 페이지로 나누어 전달하는 무선인터넷의 특성을 이용한 것으로, 사용자가 하나의 뉴스를 어디까지 보았는지에 대한 비율을 의미한다. 본 실험에서는 뉴스 한건을 길이에 따라 3~5 페이지로 나누었으며, 사용자가 매 페이지마다 다음페이지를 요청해야만 전달하게 된다. 뉴스의 첫 페이지를 본 후, 사용자가 추천된 뉴스에 대해 만족한다면 뉴스의 다음 페이지를 요청할 것이란 가정을 하였으며, 이것은 추천시스템이 사용자의 선호도를 정확히 예측한 결과로 해석하고자 한다.

기존의 무선인터넷 환경의 뉴스 서비스 제공방식과 비교하기 위해 사용자가 원하는 뉴스의 카테고리를 선택해서 보는 '분야별 뉴스 보기' 와 최근 뉴스의 순서로 보는 '최신 뉴스 보기'를 비교 대상으로 선택하였다. 실험대상자 2,344 명에 대한 이동통신업체의 무선 데이터 서비스 중 뉴스 콘텐츠의 이용현황을 분석한 결과, 아래 표 3 에서 보이는 것처럼, 본 논문에서 제안된 시스템에 의해 추천된 뉴스 제공방식의 완독률이 기존의 방식보다 높은 것으로 나타났다.

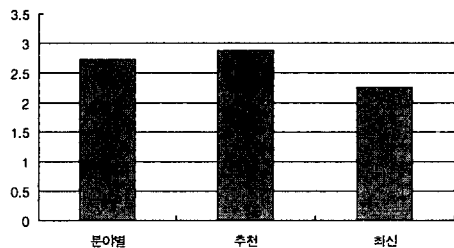


그림 2. 뉴스 완독률 비교 분석

추천뉴스의 완독률이 '분야별 뉴스' 또는 '최신뉴스'보다 높다는 것은, 추천된 뉴스를 제공 받은 경우 뉴스를 끝까지 본 비율이 더 높다는 것을 의미한다. 이것은 사용자의 선호도가 떨어지는 뉴스가 섞여 있는 기존 메뉴에 비해, 선호도가 높은 뉴스만으로 구성되어 있는 '추천뉴스'의 제공방식이 뉴스이용률을 높

일 수 있음을 나타낸다. 이는 PC 환경보다 불편한 User Interface 를 갖고 있는 모바일 폰 에서 개인화 되어 제공되는 콘텐츠가 사용자의 정보 접근성을 높이고, 사용 편리성을 향상시키는 데에 기여할 수 있음을 보여준다.

5. 결론 및 향후 연구 과제

본 논문에서는 협업 필터링 방법과 군집분석을 사용하여 무선환경에서 사용자가 선호할 뉴스만을 선택적으로 제공하는 지능형 뉴스 추천 시스템을 설계하였다. 군집분석을 통해 유사도가 비슷한 사용자를 군집화 하여 개별 특성을 나타내는 대표 사용자로 간주하였다. 군집분석을 사용하여 기존 사용자에게 무선데이터를 제공하는 방식인 일반적인 뉴스 메뉴 추천보다 지능형 뉴스 추천 시스템을 통하여 사용자의 선호도 특성 및 취향을 알기 위한 계산속도가 향상되었다. 또한, 뉴스 콘텐츠 완독률에서 기존 방식을 상회하는 수준의 성능을 보여 주었으며 추천 뉴스에 대한 사용자의 만족도가 향상되는 결과를 보여 주었으며, 이로써 지능형 추천 시스템이 무선인터넷 환경에서 사용성 및 접근성을 높일 수 있음을 알 수 있었다.

계산속도 문제를 해결하기 위하여 군집 분석을 사용하였으며, 뉴스 콘텐츠를 처음부터 끝까지 읽은 완독률 이라는 새로운 무선 데이터 콘텐츠의 사용자 만족도 측정기준을 제시하였다. 이는 실질적으로 사용자 데이터를 무선환경하의 시스템을 통하여 구현한 결과, Rating 되는 사용자의 수가 적어 추천에 대한 사용자 만족도를 측정하는 새로운 기준이 필요하였다.

향후 연구 과제로는, 무선환경하의 User Interface 의 다양성과 이에 따른 한계성의 격차를 줄이기 위하여 유무선 환경에서의 추천 시스템을 비교하는 연구가 필요할 것이며, 세부적으로는 군집을 구성하는 방법에 따라 추천 결과에 어떤 영향을 미치는 지와 다양한 개별 콘텐츠에 대한 사용자 만족도를 측정하는 기준을 개발하는 연구가 지속되어야 할 것이다.

참고문헌

- [1] 이수정, 이형동, 김형주, "사용자 경향에 기반한 동적 추천 기법", 한국정보과학회논문지 소프트웨어 및 응용 제 31 권 제 2 호 논문집, 2004.2
- [2] 정영미, 이용구, "필터링 기법을 이용한 도서 추천 시스템 구축", 정보관리연구, vol.33, no1, 2002
- [3] 김종민, 박영배, "Push 기술을 이용한 쇼핑물 개인 맞춤정보 제공 시스템의 설계," 정보과학회 Vol.27 No.2(2), 2000
- [4] 이경우, 최덕원, "전자상거래에서 상품 추천을 위한 웹 개인화 방안에 관한 연구", 한국경영과학회/대한산업공학회 춘계공동학술대회, 2001
- [5] 최준혁, 김대수, 임기욱, "군집분석과 베이지안 학습을 이용한 웹 도서 동적 추천 시스템", 한국퍼지 및 지능 시스템 학회, Vol.12, No.5, 2002

- [6] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," Accepted for publication at the WWW10 Conference, May, 2001