

사용자 기기에서 이용한 웹 데이터 분석을 통한 사용자 취향 분석 방법

(An Analysis Method of User Preference by using Web
Usage Data in User Device)

이 승 화 [†] 최 형 기 ^{**} 이 은 석 ^{***}
(Seunghwa Lee) (Hyoung-Kee Choi) (Eunseok Lee)

요약 최근 인터넷 상에 정보가 방대해지면서 사용자의 요구에 맞는 정보 필터링과 개인화 서비스가 매우 중요해지고 있다. 특히 전자상거래 분야에서 상거래를 활성화시키고 정보 제공자에 대한 만족도와 충성도를 높이기 위해, 사용자의 취향을 기반으로 한 정보 추천은 필수적인 요소가 되었다. 기존 추천 시스템은 사용자의 관심 정보를 기술한 사용자 프로파일을 대부분 정보 제공자 측에서 각각 개별적으로 수집하고 이를 기초로 추천 서비스를 제공한다. 따라서 사용자의 정보는 각 정보 제공자 측에서 분산되어 존재하며, 사용자 정보가 부족한 서버에서는 초기에 추천 전략을 세우기 어렵다는 문제가 있다. 또한 사용자 정보를 가지고 있는 서버의 경우에도 사용자가 해당 서버를 주기적으로 방문하지 않았다면, 사용자의 동적인 취향 변화를 반영하기 어렵다. 따라서 본 논문에서는 사용자의 행동을 통합적이고, 지속적으로 관찰할 수 있는 사용자 기기에서, 사용자가 이용한 웹 문서 분석을 통해 사용자의 관심 분야를 추론하고, 이를 다른 정보 제공자가 이용하는 새로운 구조의 추천 시스템을 제안한다. 또한 제안 시스템은 보다 효율적인 프로파일 생성을 위해, 웹 페이지에서 식별된 정보 블록에서 관심 단어를 추출하고, 앵커 태그를 분석하여 사용자의 이동 경로를 추적하는 특징을 포함하고 있다. 이러한 제안 시스템의 특징을 통해, 사용자 정보가 부족한 상점에서도 초기에 개인화 서비스 제공이 가능해지며, 사용자가 평소 이용하는 웹 문서로부터 프로파일을 생성함으로써, 사용자의 동적인 취향 변화를 반영할 수 있다. 또한 정보 블록에서 취향 정보를 추출하는 알고리즘을 통해 보다 빠르고 정확한 프로파일 생성이 가능해진다. 본 논문에서는 최근 구매 활동이 있었던 사용자들의 웹 검색 히스토리와 구매 데이터를 이용하여 제안 시스템의 추천 정확도와 프로파일 분석에 소요되는 시간 측면의 이득을 실험하였으며, 그 결과를 통해 시스템의 유효성을 확인하였다.

키워드 : 정보 필터링, 추천 시스템, 전자상거래, 개인화, 사용자 프로파일

Abstract The amount of information on the Web is explosively growing as the Internet gains in popularity. However, only a small portion of the information on the Web is truly relevant or useful to the user. Thus, offering suitable information according to user demand is an important subject in information retrieval. In e-commerce, the recommender system is essential to revitalize commercial transactions, raise user satisfaction and loyalty towards the information provider. The existing recommender systems are mostly based on user data collected at servers, so user data are dispersed over several servers. Therefore, web servers that lack sufficient user behavior data cannot easily infer user preferences. Also, if the user visits the server infrequently, it may be hard to reflect the dynamically changing user's interest. This paper proposes a novel personalization system analyzing the user preference based on web documents that are accessed by the user on a user device. The system also identifies non-content blocks appearing repeatedly in the dynamically generated web documents,

[†] 정 회 원 : 성균관대학교 정보통신공학부 교수
shlee@ece.skku.ac.kr

^{**} 정 회 원 : 성균관대학교 정보통신공학부 교수
hkchoi@ece.skku.ac.kr

^{***} 중신회원 : 성균관대학교 정보통신공학부 교수
eslee@ece.skku.ac.kr

논문접수 : 2007년 9월 11일

심사완료 : 2008년 12월 22일

Copyright©2009 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨터의 실제 및 레터 제15권 제3호(2009.3)

and adds weight to the keywords extracted from the hyperlink sentence selected by the user. Therefore, the system establishes at an early stage recommendation strategies for the web server that has little user data. Also, user profiles are generated rapidly and more accurately by identifying the information blocks. In order to evaluate the proposed system, this study collected web data and purchase history from users who have current purchase activity. Then, we computed the similarity between purchase data and the user profile. We confirm the accuracy of the generated user profile since the web page containing the purchased item has higher correlation than other item pages.

Key words : Information Filtering, Recommendation System, E-Commerce, Customization, User Profile

1. 서 론

최근 정보통신 기술의 급속한 발전과 함께 인터넷 이용률은 나날이 증가하고 있으며, 인터넷 상에 존재하는 정보도 점차 방대해지고 있다. 그러나 이 중 사용자가 실제로 필요로 하는 정보는 극히 일부이다. 따라서 사용자의 요구에 맞는 정보 필터링과 개인화 서비스가 매우 중요하다. 특히, 전자상거래 분야에서 상거래를 활성화 시키고 정보 제공자(웹 상점)에 대한 만족도와 충성도를 높이기 위해, 사용자의 취향을 기반으로 한 정보(상품) 추천은 필수적으로 요구되고 있다.

전통적인 추천 기법으로, 사용자의 관심과 콘텐츠의 유사도를 비교하는 콘텐츠 기반 추천(Content-based Recommendation)과 유사한 관심을 가진 사용자들 간에 관심 콘텐츠를 상호 추천하는 협업추천(Collaborative Recommendation)이 있다. 이러한 추천 기법들은 모두 사용자의 취향을 기술한 사용자 프로파일(User Profile)을 기초로 한다. 따라서 추천 정보의 품질을 높이기 위해서, 보다 정확한 사용자 프로파일의 생성은 매우 중요하다.

기존 연구와 상용화된 추천 시스템에서 이와 같은 사용자 프로파일의 생성은 대부분 정보를 제공하는 서버에서 이루어진다[1,2]. 따라서 사용자의 취향 정보는 다수의 정보 제공자 측에 분산되어 존재하며, 사용자 정보가 부족한 서버에서는 개인화 서비스가 이루어지기 어렵다는 문제가 발생한다. 이와는 다르게 사용자 기기에서 사용자 취향을 분석하는 연구도 진행되었다[3-5]. 그러나 대부분의 경우에 사용자 기기는 분석된 프로파일 정보를 활용하여 정보 필터링까지 수행한다. 이는 다양한 제약사항(낮은 CPU 성능, 제한된 배터리용량, 제한된 bandwidth 등)을 가지는 모바일 컴퓨팅 환경에서는 적합하지 않다.

사용자 프로파일을 생성하는 방법은 사용자가 제공한 개인 정보나 관심 정보를 이용하는 명시적(explicit)인 방법과 사용자의 행동을 통해 암시적(implicit)으로 취향을 추론하는 방법으로 분류할 수 있다. 사용자의 명시적인 표현을 통해 프로파일을 생성하는 방법은 빠르게 취향을 습득할 수 있지만, 사용자를 번거롭게 하며 동적인 취향

변화를 반영하기 어렵다는 단점이 있기 때문에 암시적인 프로파일 생성방법이 보다 바람직하다고 할 수 있다.

암시적인 프로파일 생성방법의 대표적인 예는 사용자가 이용하는 콘텐츠의 특징을 식별하고, 이를 사용자의 관심 대상으로 분류하는 것이다. 일반적으로 전자상거래 분야에서는 사용자가 기존에 구매한 상품이나 장바구니 정보를 이용하여 상점 측에서 사용자의 취향을 분석한다. 그러나 이는 앞에서 언급한 것과 같이, 각 상점에 사용자 정보가 분산되어 존재하는 형태이며, 사용자의 행동 정보가 일정기간 수집되지 않은 상점에서는 추천 전략을 세우는 것이 어렵다는 문제가 있다.

사용자가 이용하는 콘텐츠가 문서인 경우, 해당 문서에 등장하는 단어의 빈도(term frequency)를 조사하고, 빈번하게 등장하는 단어를 사용자의 관심 키워드로 분류하는 기법이 정보검색 시스템에서 폭넓게 사용되고 있다[6-8].

그러나 최근 대부분의 웹 문서는 사용자의 관심과 크게 관련 없는 많은 부분들(로그, 이동버튼, 광고, 저작권 표시 등)을 포함하고 있다. 이는 하나의 사이트를 제작할 때 공통된 레이아웃이 적용되기 때문이며, 특히 동적으로 생성되는 웹 문서의 경우 이 특징은 더욱 일반적으로 나타난다. 관련연구 [9]에서는 이러한 요소를 '비정보 블록(noncontent blocks)'이라고 정의하였다.

전체 문서 집합에 공통적으로 빈번하게 등장하는 단어의 가중치를 낮추는 Inverse Document Frequency (IDF) 계산[14]을 통해 의미가 적은 단어를 어느 정도 식별할 수 있지만, 사용자의 취향과 크게 관련이 없는 비정보 블록들을 모두 포함하여 문서를 분석하는 것은 프로파일 생성의 속도나 정확도 측면에서 좋지 않은 성능 결과를 나타내며 효율적이지 않다.

정리하면, 기존의 추천 시스템들은 다음과 같은 문제를 가지고 있다. 첫째, 대부분의 시스템들은 서버에서 수집된 사용자의 행동을 기반으로 추천을 수행하기 때문에, 사용자 정보가 각 정보 제공자들에 분산되어 존재하며, 사용자 정보가 부족한 경우에는 추천 전략을 세우는 것이 어렵다. 또한 일정기간 사용자의 방문이 없었던 경우, 동적으로 변화하는 사용자의 취향을 반영하기 어렵다. 둘째, 사용자가 이용한 웹 문서를 기반으로 사용

자 프로파일을 분석하는 경우, 최근 웹 문서는 많은 비정보 블록들을 포함하고 있기 때문에, 일반적으로 널리 사용되는 TF-IDF방법만을 사용한다면 정확도와 분석시간 측면에서 효율성이 낮다.

본 논문에서는 상기의 문제에 대응하기 위해 다음과 같은 특징을 가진 새로운 구조의 추천 시스템을 제안한다. 첫째, 제안 시스템은 사용자의 행동을 통합적이고 지속적으로 관찰할 수 있는 사용자 기기에서 사용자가 이용한 웹 문서를 이용하여 사용자 프로파일을 분석하고 이를 서버에서 활용한다. 둘째, 동일한 서버로부터 얻어진 웹 문서의 구조를 분석하여 반복되는 부분을 비정보 블록으로 식별하여 제거하고, 정보 블록만을 참조하여 사용자 프로파일을 생성한다. 셋째, 사용자의 주요 관심 단어를 식별하기 위해 앵커 태그들을 수집하고, 사용자가 이동한 웹 문서의 경로를 따라가며 해당 앵커 태그와 연결된 핵심 단어들을 추출한다. 그리고 이를 사용자 프로파일에 핵심 단어로 반영한다. 넷째, 사용자 기기에 저장된 문서들을 그것이 포함하고 있는 단어의 유사도를 기반으로 그룹화하여, 다양한 종류의 프로파일을 생성한다.

이러한 제안 시스템의 특징을 통해 사용자 정보가 부족한 상점에서 개인화 서비스가 초기에 가능해지며, 사용자가 평소에 이용하는 웹 문서로부터 프로파일을 생성함으로써, 동적인 사용자 취향 변화를 개인화 서비스에 반영할 수 있다. 또한 정보 블록에서 취향정보를 추출하는 제안 알고리즘을 통해, 전체적인 웹 문서를 분석하는 방법에 비해 빠르고 정확한 프로파일 생성이 가능해진다. 또한 웹 서버는 다양한 종류로 생성된 프로파일 중에서, 자신이 제공하는 콘텐츠와 관련된 프로파일을 참고함으로써, 상품 추천 시에 TV시청이나 뉴스검색 등을 통해 수집된 취향 정보일지라도 내용이 관련되었다면 참조할 수 있는 효과를 얻을 수 있다.

우리는 제안 시스템의 평가를 위해, 최근 구매 활동이 있었던 사용자들의 웹 데이터와 구매 데이터를 수집하여 사용자 프로파일을 추출하고, 구매 데이터와의 관련성을 비교하였다. 또한 웹 문서의 정보 블록을 추출하여 프로파일을 생성하는 제안 시스템과 비정보 블록을 포함한 전체 문서를 분석하는 기존 시스템을 정확도와 소요되는 시간 측면에서 비교하였으며, 실험 결과를 통해 제안 시스템의 유효성과 확장 가능성을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 소개하고, 3장에서는 사용자가 이용한 웹 문서를 분석하는 알고리즘과, 사용자 기기에서 생성된 사용자 프로파일을 서버에서 활용하는 구조의 제안 시스템을 소개한다. 그리고 4장에서는 시스템의 평가를 위한 실험과 그 결과를 정리하며, 5장에서는 결론과 향후과제를 논의한다.

2. 관련연구

앞에서 언급한 것과 같이 최근 정보 필터링과 추천 시스템은 콘텐츠 기반 추천 방법과 협업 추천 방법, 그리고 이를 혼합한 방식이 사용되고 있으며, 이는 모두 사용자의 취향을 기술한 사용자 프로파일을 기초로 한다. 따라서 동적으로 변화하는 사용자 취향을 보다 정확히 파악하고, 이를 사용자 프로파일로 생성하는 것은 매우 중요한 도전 과제이다.

2.1 프로파일 생성

사용자의 관심 정보를 분석하기 위한 다양한 기법들이 연구되고 있다. 가장 보편화된 방법은 사용자가 사이트 방문 초기에 명시적으로 표현한 개인 정보나 관심 정보를 이용하는 것이다[1]. 이러한 방법은 비교적 빠르게 사용자 정보를 얻을 수 있는 장점이 있지만, 사용자를 번거롭게 할 수 있으며 동적으로 변화하는 사용자의 취향은 반영하기 어렵다는 문제가 있다. 이와는 다르게 사용자의 행동을 관찰하고, 이를 통해 암시적으로 취향을 분석하는 방법도 있다[2-4]. 그러나 이는 구매 데이터와 같은 사용자 행동 데이터가 일정기간 수집되어야 하며, 이러한 데이터는 대부분 정보 제공자의 서버에 각각 저장되기 때문에, 사용자가 처음 방문한 상점에서는 개인화 서비스가 이루어지기 어렵다는 문제가 있다. 또한 사용자가 오랜 기간 해당 상점을 방문하지 않은 경우에는, 동적으로 변화하는 사용자의 취향을 반영하기가 어렵다는 문제도 발생한다.

따라서 사용자의 행동을 지속적으로 통합적으로 관찰할 수 있는 사용자의 기기에서, 사용자 프로파일을 분석하고 관리하는 새로운 구조가 필요하다. 게다가 최근에는 유비쿼터스 컴퓨팅 이슈의 등장과 함께, 사용자 주변에 다양한 종류의 컴퓨팅 기기들이 증가하고 있다. 사용자가 각각의 기기들을 사용하면서 발생하는 다양한 행동 정보는 사용자의 대표 기기로 취합되어, 향후 사용자의 취향 분석에 유용한 데이터로 활용될 수 있을 것이다.

2.2 웹문서의 분석

사용자의 취향을 분석하기 위해 본 논문에서 사용한 기본 방법은 사용자가 이용한 웹 페이지들에 등장하는 키워드들을 분석하는 것이다. 이를 위해 Term Frequency * Inverse Document Frequency(TF*IDF)기법이 사용되었다. TF는 특정 단어가 문서에 등장하는 빈도를 나타내며, IDF는 전체 문서 집합에 공통적으로 빈번하게 등장하는 단어의 가중치를 낮추기 위해 사용된다[14,15]. 많은 기존 연구에서 이 기법이 적용되었다[7,8]. 그러나 최근 대부분의 웹 문서들은 동적으로 생성되며, 이동 버튼이나 광고, 저작권 정보와 같이 사용자의 취향과 크게 관련 없는 많은 부분들을 포함하고 있다. 따라

서 이러한 부분을 모두 포함하여 등장하는 단어를 분석하는 것은 처리 시간을 지연시킬 뿐만 아니라 프로파일의 정확도를 낮추게 될 것이다.

최근 웹 페이지에서 비정보 블록들을 제거하고 정보 블록을 식별하기 위한 연구가 활발히 진행되고 있다 [9-13]. 이 연구들은 대부분 데스크탑 PC를 고려하고 만들어진 웹 페이지를 작은 화면의 모바일 기기에 적합하게 변환하기 위한 목적으로 진행되었다.

본 논문에서는 프로파일 생성을 보다 효율적으로 수행하기 위해 웹 문서 분석 과정에 이러한 관련연구와 유사한 접근법을 적용하였다.

이는 HTML로 이루어진 웹 문서를 DOM 트리 구조로 변환하고[16], 동일한 서버로부터 얻어진 페이지들의 각 노드를 비교하며, 반복적으로 등장하는 부분을 비정보 블록으로 식별하는 것이다. DOM 트리의 예가 그림 1에 나와 있다.

또한 기존에 사용자가 선택하는 하이퍼링크와 연결된 문서의 내용을 분석하여, 사용자의 관심 단어를 추출하는 연구도 진행되었다[7]. 이 연구 역시 연결된 문서에 포함된 모든 단어를 분석하는 다소 비효율적인 면을 가지고 있지만, 웹 문서에 등장하는 하이퍼링크를 이용하여 사용자의 이동 경로를 관찰하고, 사용자의 관심 정보를 보다 효과적으로 추출할 수 있는 중요한 단서로 고려한 부분은 본 논문의 제안 시스템과 동일하다.

최근 인터넷 상에 존재하는 대부분의 정보들은 지속적으로 새롭게 갱신되며, 이러한 정보들은 웹 사이트에 새롭게 연결되거나 이메일과 같은 형태로 사용자에게 소개된다. 따라서 이러한 새로운 정보에 대한 하이퍼링크는, 통상적으로 연결된 문서의 내용을 잘 함축할 수 있는 문자열로 이루어져 있다. 이러한 방법은 인터넷 뉴스, 공지사항, 커뮤니티 사이트의 새로운 글, 웹 상점의 새로운 상품 소개 등 다양한 정보에 적용되고 있다. 사용자는 이러한 연결 문서의 제목을 읽고 다음 이동할 문서를 선택하기 때문에 하이퍼링크 문자열에 사용자의 취향을 분석할 수 있는 중요한 정보가 포함되어 있다는

것을 알 수 있다.

다음 장에서는 이러한 기존 관련연구의 다양한 단점을 보완하고, 효율적인 사용자 프로파일 생성을 위해 설계된 제안 시스템을 자세히 설명한다.

3. 제안 시스템

제안 시스템은 사용자가 처음 방문하는(사용자 정보가 부족한) 웹 상점의 개인화 서비스를 지원하기 위해 설계되었으며, 사용자의 동적인 취향 변화를 반영하여 보다 빠르고 정확하게 사용자 프로파일을 생성하는 것을 목표로 한다. 세부적인 특징은 다음과 같다.

- 다수의 웹 서버로부터 사용자가 이용하는 웹 문서를 지속적이고, 통합적으로 분석하기 위해 사용자 기기에 수집된 웹 문서를 이용.
- 비정보 블록을 포함하여 웹 문서 전체를 분석하는 기존 사용자 프로파일 생성방법의 비효율적인 문제를 개선하기 위해 웹 문서에서 식별된 정보 블록을 이용.
- 웹 문서의 하이퍼링크를 분석하여 사용자의 이동을 추적하며, 앵커(anchor) 태그에서 추출된 문자열을 사용자의 주요 관심 단어로 프로파일에 반영.
- 웹 문서의 유사도를 비교하여 문서를 분류하고, 다양한 종류의 프로파일을 생성.

3.1 제안 시스템의 구성과 기본 동작 과정

제안 시스템은 사용자 기기 또는 사용자 측의 프록시 서버에 내장 되어, 사용자가 이용했던 웹 문서를 분석하며, 기본적으로 불용어(stop word) 제거 - 어근 처리(stemming) - 문서 색인(indexing)과정을 거친다. 또한 문서가 포함하고 있는 단어의 셋을 기반으로 문서의 특징을 분류하여 사용자 프로파일을 생성함으로써, 웹 서버는 제공하는 정보와 관련된 프로파일만을 참고할 수 있다. 제안 시스템의 사용자 모듈에 대한 전체적인 구성이 그림 2에 나와 있으며, 각 구성요소의 기능과 동작 과정은 다음과 같다.

- 캐쉬: 사용자가 검색했던 웹 문서가 저장되는 공간을 의미한다.
- HTML 번역기: 입력된 웹 문서의 HTML 태그를 노드로 식별하고, 이를 DOM 트리 구조로 생성하며, 각 노드에 포함된 텍스트를 추출한다. 또한 사용자의 이동을 추적하기 위해, 해당 웹 페이지에 포함된 앵커 태그와 그것에 연결된 문서의 URL을 매핑 테이블로 저장한다.
- 불용어 사전: 여러 문서에 빈번하게 등장하지만 사용자의 프로파일에 크게 영향을 주지 않는 단어나 특수 문자들의 집합을 의미한다. 영어의 경우 접두사나 접미사, 관사, 대명사, be 동사 등을 포함하며, 한글에서는 '~하다', '~이다'와 같은 동사를 포함한다. 또한 웹 문서의 특징 분석에 영향을 주지 않는 일반적인 단어

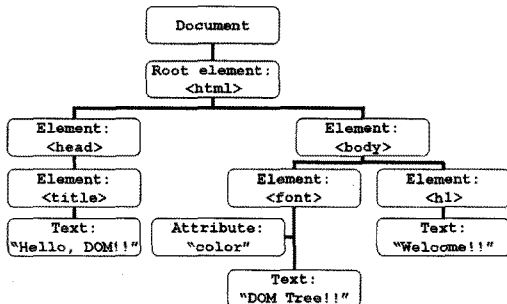


그림 1 DOM 트리의 예

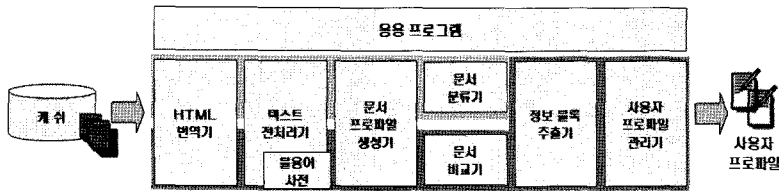


그림 2 제안 시스템의 전체적인 구성

들도 불용어로 처리된다.

- 텍스트 전처리기: 불용어 사전을 기반으로 문서에 포함된 텍스트의 불용어 제거 작업을 수행하며, 어근을 분석하여 기본 명사로 치환할 수 있는 단어를 처리한다.
- 문서 프로파일 생성기: 문서를 특징에 따라 분류하기 위해, 이전 단계에서 정제된 단어들을 기반으로 문서에 대한 메타파일(문서 프로파일)을 생성한다. 문서 프로파일은 단어와 그것의 등장 빈도 셋으로 구성되며, 등장 빈도는 절대적인 값이 아닌 비율을 이용한다.
- 문서 분류기: 생성된 문서 프로파일을 기반으로 캐쉬에 저장된 문서들의 유사도를 비교하고 그룹화한다. 그룹화는 보편적으로 널리 사용되고 있는 k-means 알고리즘을 사용하며, 보다 자세한 동작 과정은 3.2장에서 소개한다.
- 정보 블록 추출기: 웹 문서의 원본 URL을 참조하여, 동일한 서버로부터 얻어진 문서 집합에서 반복적으로 나타나는 비정보 블록들을 식별하고 제거한 뒤, 정보 블록을 추출한다. 이를 통해 보다 빠른 사용자 프로파일 생성이 가능해진다. 자세한 동작 과정은 3.3장에서 소개한다.
- 사용자 프로파일 관리자: 식별된 정보 블록에 등장하는 단어들의 출현빈도를 TF*IDF 기법을 적용하여 계산하며, 이를 기반으로 사용자 프로파일을 생성한다. 사용자 프로파일은 문서 분류기가 생성한 분류의 수만큼 여러 종류로 생성된다.
- 문서 비교기: 개인화 서비스를 수행하는 웹 서버에서 사용자 정보를 요청하는 경우, 보다 관련된 프로파일 정보를 기반으로 응답하기 위해, 웹 서버에서 제공하는 문서와 문서 분류기에 의해 생성된 분류 기준과의 유사도를 비교하고, 적절한 사용자 프로파일을 결정한다. 웹 문서의 분석은 앞에 설명한 모듈들을 이용한다.

3.2 웹 문서의 분류 과정

제안 시스템의 기본 원리는 정보 검색 시스템에서 보편적으로 사용하는 TF*IDF 방법을 사용하여, 사용자가 이용한 웹 페이지에 빈번하게 등장하는 단어를 관심 키워드로 분석하는 것이다. 그러나 사용자가 검색하는 웹 문서들은 광범위한 주제를 가지고 있기 때문에, 단순히 하나의 통합된 사용자 프로파일이 생성된다면, 특정 상

점에서 상품과 프로파일의 관련성을 분석하는 것은 매우 어려운 문제가 될 것이다.

이를 해결하기 위해, 제안 시스템은 사용자 측에 저장된 문서를 분류하여 문서의 특징에 따라 특성화된 여러 종류의 프로파일을 생성하고, 개인화 서비스를 제공하려는 상점과 가장 유사한 문서 그룹에서 추출된 사용자 프로파일을 이용하는 구조를 제안한다.

문서 분류기에 의해 수행되는 문서 그룹화 알고리즘은 이와 같이 사용자 기기에 수집된 웹 문서들을 분석하여, 문서들을 특징에 따라 각각 그룹으로 분류하는 역할을 수행한다. 전체적인 동작과정이 그림 3에 나타나 있다.

각 문서가 포함하는 단어들은 그림 4(a)와 같이 벡터 공간에 표현되며, 해당하는 값(x_{ij})은 TF*IDF를 이용하여 계산된다. 그림 4(a)에서 d 는 문서를 의미하며, t 는 단어를 나타낸다. tf_{ij} 는 문서 i 에서 단어 j 가 나타난 빈도 수를 의미하며, idf_j 는 단어 j 가 전체 문서 집합에서 나타난 비율의 로그 값을 의미한다. N 은 전체 문서 수이다.

문서의 그룹화를 위해, 각 문서는 동일한 구조와 길이의 벡터 셋을 가져야 한다. 이를 위해 Step 3.3에서는 그림 4(b)와 같이 전체 단어 행렬에서 임계치보다 낮은 단어를 제거하는 과정을 거친다. 이후, 각 문서들이 포함하는 단어를 기반으로 문서 프로파일이 생성되며, 각 문서 프로파일 간 유사도를 비교하여 그룹화가 이루어진다. 본 논문에서는 문서들을 효과적으로 분류하기 위해, 사전에 정의된 분류기준이 없어도 각 객체들의 유사도비교를 통해 그룹화를 수행하는 k-means 알고리즘을 이용하였다. 각 문서 간 유사도 비교는 식 (1)과 같은 유클리드 계산법이 사용되었다. 여기서, d 는 문서 i 와 j 간의 거리를 의미한다.

$$d_{i,j} = \sqrt{\sum_{z=1}^m (x_{iz} - y_{jz})^2} \quad (1)$$

이러한 과정을 통해, 사용자가 검색한 웹 문서들은 그것이 포함하고 있는 단어들을 기반으로 그룹지어진다.

3.3 사용자 프로파일 생성 과정

앞에서 언급한 것과 같이, 최근 웹 문서는 사용자의 취향과 크게 관련 없는 비정보 블록들을 다수 포함하고

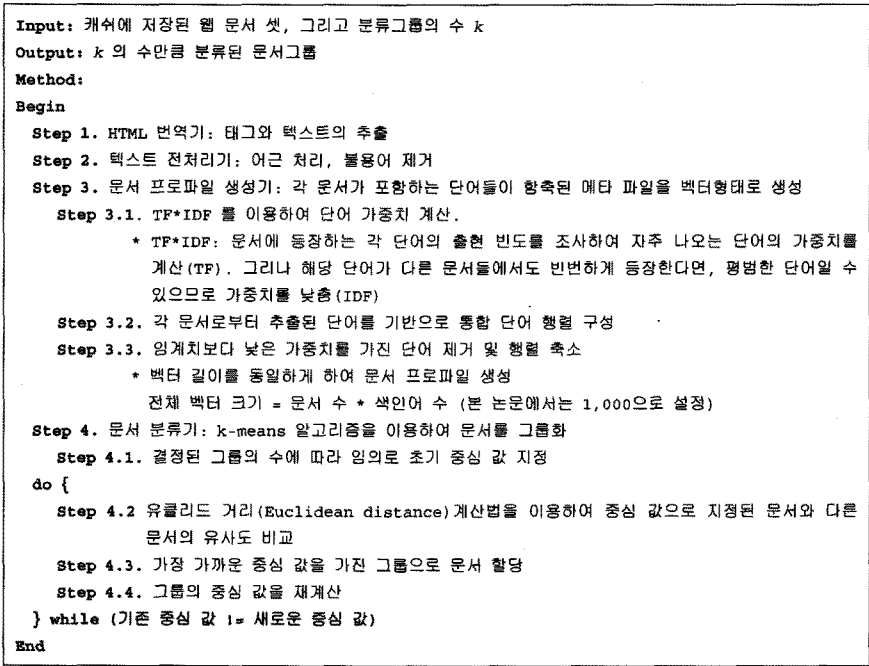


그림 3 웹 문서의 분류 과정

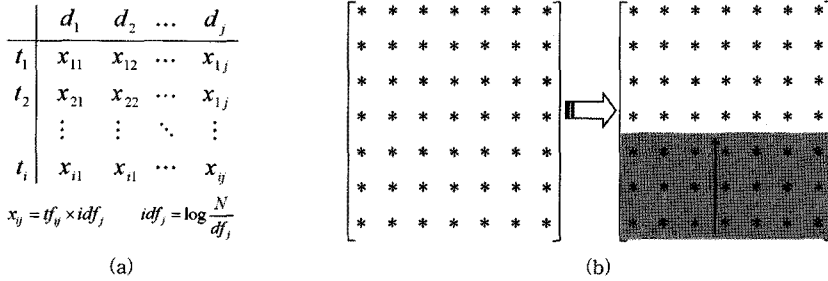


그림 4 웹 문서로부터 추출된 단어 행렬

있다. 이는 이전 단계인 문서의 유사도 비교에는 영향을 주는 요소들이지만, 사용자 프로파일 추출 단계에서는 사용자 취향 분석에 정확도를 낮추는 불필요한 부분이다. 따라서 제안시스템은 웹 페이지에서 이러한 불필요한 요소를 제거하고 정보 블록을 식별하여, 보다 효과적으로 사용자의 관심 키워드를 추출하는 방법을 적용하였다. 전체 동작 흐름은 그림 5와 같다.

정보 블록 식별 알고리즘은 관련연구에서 소개한 S. Debnath et al.,의 연구[9]와 유사하게 동작하며, 본 논문에서는 보다 정확한 사용자 프로파일 생성을 위해 적용되었다. 핵심 아이디어는 동일한 서버로부터 얻어진 문서들에 반복적으로 나타나는 부분을 비정보 블록으로 식별하는 것이다. 이를 위해 제일 먼저 수행되어야 하는 것은 웹 문서를 여러 블록으로 분할하는 것이다.

최근 여러 연구에서 시각적인 요소를 기반으로 웹 문서를 분할하는 방법이 적용되었다[10,11]. 그러나 이는 데스크 탑 환경을 고려하여 작성된 웹 문서를 모바일 기기의 화면에 손실 없이 보여주기 위한 것이며, 상대적으로 비용이 많이 소요된다. 본 논문에서 웹 문서 분할의 목적은 사용자의 관심 단어가 포함된 주요 블록을 검출하는 것이기 때문에, 비교적 단순한 방법인 분할 태그 리스트를 이용한 방법[9]의 적용이 가능하다.

많은 웹 문서 저작자들은 레이아웃을 디자인하기 위해, 대부분 <table>태그를 이용한다. 따라서 분할 태그 목록은 <table> 태그를 우선하여, <table><tr><p><hr><div>과 같은 태그들로 구성된다. 정보 블록 추출기는 분할 태그 목록에 있는 첫 번째 태그를 기반으로 웹 문서를 분할한다. 그리고 분할된 블록

```

Input: 캐쉬에 저장된 웹 문서 중 URL의 앞부분이 동일한 문서 셋
Output: 각 문서의 정보 블록으로부터 추출된 사용자 프로파일
Method:
Begin
  Step 1. HTML 번역기: 태그와 텍스트의 추출 및 구조 분석을 통해 HTML DOM 트리 생성
  Step 2. 정보 블록 추출기: 웹 페이지 분할
    do{
      분할 태그 목록에 있는 i 번째 태그를 기반으로 웹 페이지를 분할
      i++
    } while (다음 분할 태그 ∈ 분할된 블록)
  Step 3. 정보 블록 추출기: 각 블록에 포함된 노드(tag)를 순차적으로 탐색하며, 임계치 이상으로 중복
  되는 구조와 내용을 가진 블록 식별 - 비정보 블록으로 간주하여 제거
  Step 4. 텍스트 전처리기: 추출된 정보 블록에서 단어를 추출하여 어근 처리, 불용어 제거
  Step 5. 사용자 프로파일 관리기: 추출 단어들의 TF-IDF 계산 - 사용자 프로파일 갱신
End
    
```

그림 5 정보 블록 추출 과정

에 분할 태그가 존재하지 않을 때까지 세부 분할을 계속 한다. 동일한 웹 서버로부터 얻어진 각각의 웹 문서들이 여러 블록으로 분할된 이후, 정보 블록 추출기는 이 블록들을 순차적으로 비교하며 임계치 이상으로 중복되는 구조와 내용을 가진 블록을 비정보 블록으로 간주하여 제거한다. 이때 유사도 비교는 각 노드(tag)의 구조와 포함된 텍스트를 위주로 진행하며, 각 노드의 속성 값은 무시된다.

비정보 블록들이 식별된 이후, 사용자 프로파일 관리기는 나머지 블록들에서 추출된 단어를 기반으로 사용자 프로파일을 생성한다. 이때, 보다 사용자의 관심을 반영하는 키워드를 추출하기 위해 핵심키워드 추출 알고리즘을 추가하였다. 이는 웹 문서에 포함된 앵커 태그와 그것에 연결된 링크 문장, URL 정보를 추출하고, 사용자가 이동하는 웹 문서의 경로를 따라가며 사용자가 클릭하는 링크 문장으로부터 핵심 키워드를 추출하는 알고리즘이다. 주요 알고리즘은 그림 6과 같으며, 추출 예가 그림 7에 나와 있다.

그림 7의 사례에서, 사용자가 클릭한 ‘삼성’, ‘노트북’과 같은 키워드는 일반 단어에 비해 높은 가중치로 사용자 프로파일에 반영되며, 이러한 과정을 통해 제안 시스템은 사용자의 관심을 보다 정확하게 분석하게 된다. 생성되는 사용자 프로파일은 식 (2)와 같이 정보필터링 분야에서 일반적으로 사용되는 벡터공간 표현방법(vector space representation)을 이용한다.

$$User\ profile_{type\ id} = (term_1, w_1), (term_2, w_2) \dots (term_n, w_n) \quad (2)$$

여기서, w는 각 단어의 가중치를 의미하며, 이는 TF*IDF를 이용하여 계산된다.

3.4 U-Commerce 시나리오

제안 시스템을 기반으로 사용자 기기에 취합된 사용자의 취향 정보를 웹 서버에서 활용하는 구조가 그림 8에 나타나있다. 시스템은 사용자 기기에서 사용자의 웹 문서를 분석하는 사용자 모듈과 콘텐츠 제공자의 서버에 위치하는 웹 서버 모듈, 그리고 사용자의 정보를 저장하고 관리하는 홈 서버 모듈로 구성된다.

```

Input: 캐쉬에 저장된 웹 문서 중 URL의 앞부분이 동일한 문서 셋
Output: 앵커 태그로부터 추출된 사용자의 주요 관심 키워드
Method:
Begin
  Step 1. 웹 문서로부터 앵커 태그와 그것에 연결된 텍스트, URL 정보 추출하여 매핑 테이블 작성
  Step 2. 캐쉬에 저장된 웹 문서 중 URL의 앞부분이 동일한 문서 검색
    while (같은 서버로부터 얻은 문서 == true){
      매핑 테이블을 참조하여, URL과 연결된 링크 문자열들을 추출
    }
  Step 3. 텍스트 전처리기: 추출된 문자열의 어근 처리, 불용어 제거
  Step 4. 사용자 프로파일 관리기: 처리된 단어를 가중치를 추가하여 사용자 프로파일에 반영
  Step 5. 매핑 테이블 삭제. 알고리즘 종료
End
    
```

그림 6 사용자 관심키워드 추출 과정

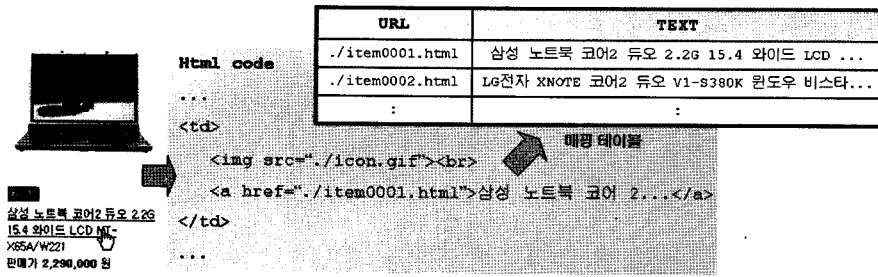


그림 7 웹 페이지의 앵커 태그로부터 사용자의 관심 키워드를 추출하는 예

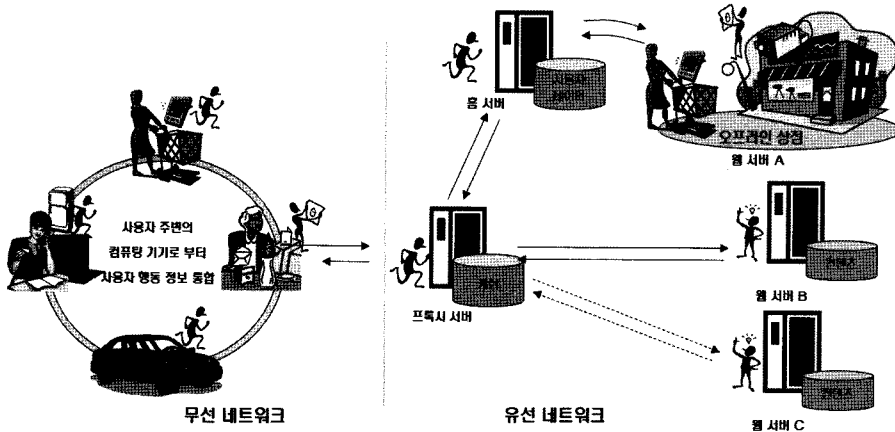


그림 8 사용자 기기에서 추출된 프로파일을 서버에서 활용하는 구조

사용자가 이용하는 콘텐츠는 클라이언트 모듈에 의해 분석된다. 이후, 추출된 사용자 프로파일은 주기적으로 홈 서버로 전송된다. 홈 서버는 사용자 기기의 리소스 문제를 해결하기 위해 설계되었으며, 다수의 사용자 정보를 취합하는 역할을 수행한다. 본 논문에서 홈 서버는 이메일 서버와 같은 신뢰성 있는 서버를 가정하였기 때문에, 프라이버시와 같은 문제는 크게 고려하지 않았다.

사용자 정보가 홈 서버에 전송된 이후, 개인화 서비스를 제공하려고 하지만 사용자의 정보가 부족한 웹 서버에서는 사용자가 요청 정보를 보냈을 때, 자신의 웹 페이지 샘플을 홈 서버로 전송한다. 이를 위해 사용자의 정보를 관리하는 홈 서버의 위치 식별 과정이 필요하다. 홈 서버는 전송된 문서와 다수의 프로파일 간 유사도 비교를 수행한 이후, 가장 관련된 사용자 프로파일을 서버로 전송한다. 서버는 이를 기반으로, 사용자의 취향을 빠르게 식별하여 초기 추천전략을 세운다. 프록시 서버는 향후 추천 콘텐츠의 적용을 수행하기 위해 함께 설계되었다.

4. 시스템 구현 및 평가

본 논문에서는 제안 시스템의 평가를 위해, 최근 구매

활동이 있었던 사용자의 구매 데이터와 웹 문서 데이터를 수집하였으며, 수집된 웹 문서를 비정보 블록을 포함한 전체 문서 분석과 제안 시스템을 이용한 경우로, 각각 사용자 프로파일을 추출하였다. 그리고 추출된 프로파일을 실제 사용자의 구매 데이터 페이지와 비교하여 일치도를 분석하였으며, 정확도 측면과 프로파일 분석에 소요되는 시간 이득 측면의 성능 평가를 수행하였다.

4.1 생성된 프로파일의 정확도 평가

첫 번째 실험은 제안 시스템을 통해 생성된 사용자 프로파일의 유용성 평가를 위해 수행되었다. 이를 위해, 사용자의 캐쉬에서 구매 시점부터 2주 전까지 웹 문서 데이터를 임의로 최대 100건 수집하였다. 이후, 제안 시스템을 이용하여 유사한 문서들을 그것이 포함하고 있는 단어를 기반으로 그룹화 하였다. 문서들은 포함하고 있는 단어를 기반으로 유사한 그룹으로 분류되었으며, 각 그룹별로 사용자 프로파일 생성되었다. 본 실험에서 분류 그룹의 수는 임의로 5개를 지정하였으며, 그룹의 수가 늘어날수록 보다 다양한 종류의 프로파일 생성될 것이다. 생성된 프로파일의 일부가 그림 9에 나타나 있다.

이후, 사용자가 상품을 구매한 쇼핑물로부터, 구매 상

그룹별 프로파일	관심 단어와 가중치					
	$User\ profile_a$	sch	C2250	w2700	기기변경	KTF
	32	32	31	30	30	...
$User\ profile_b$	경기	미들스브러	플타임	노스헝튼	칼링컵	...
	28	27	25	25	24	...
$User\ profile_c$	펀드	투자	미래에셋	은행	수수료	...
	41	41	40	40	40	...
$User\ profile_d$	D40	DSLR	니콘	올림푸스	EasyShare	...
	41	32	29	27	15	...
$User\ profile_e$	건강보험료	보건복지부	인상	진료비	복지부	...
	18	12	10	8	7	...

그림 9 다양한 종류로 생성된 사용자 프로파일의 일부

품 페이지를 포함하여 20건의 상품 페이지를 무작위로 수집하였다. 그리고 사용자 프로파일을 생성하는 방식과 동일한 방법으로 상품 페이지들의 특징을 각각 모델링 하였다. 또한 쇼핑물의 특성을 나타내는 쇼핑물 프로파일도 생성하기 위해, 추출된 페이지 프로파일들을 전체적으로 계산하여 또 다른 통합 프로파일을 생성하였다. 이후, 생성된 쇼핑물 통합 프로파일과 각 그룹별로 생성된 사용자 프로파일의 유사도를 비교하여 가장 적절한 사용자 프로파일을 선택하였다. 각 사용자 프로파일과 쇼핑물 프로파일의 유사도 비교는 식 (3)과 같이 단어의 일치도를 이용하였으며, 실험결과, 쇼핑과 관련된 키워드(예. 가격, 수량, 배송 등)가 포함되어 있는 $User\ profile_a$ 가 가장 유사한 프로파일로 선택되었다. 식에서 Up 는 User Profile, Sp 는 Shop Profile을 각각 의미한다.

$$similarity(U_p, S_p) = \frac{U_p \cap S_p}{S_p} \quad (3)$$

그리고 선택된 $User\ profile_a$ 의 유용성을 확인하기 위해, 각각의 상품 페이지와의 일치도를 다시 계산하였다. 실험 결과, 사용자 프로파일은 사용자가 구매한 페이지의 주요 키워드를 상당 부분 포함하고 있음을 확인할 수 있었다.

이후, 정보 블록을 식별하여 사용자의 관심 단어를 추출하는 제안 시스템의 유용성을 평가하기 위해, $User\ profile_a$ 가 추출된 문서 그룹에 속한 각 문서들을 TF*IDF 방법만을 기반으로 프로파일을 추가 생성하였으며, 이 과정에서 생성된 프로파일과 상품 페이지의 일치도를 계산하여 제안 시스템과 결과를 비교하였다. 각

방식을 통해 추출된 관심 단어의 일부가 그림 10에 나타나 있으며, 일치도 비교 실험 결과는 그림 11에 나타나 있다.

그림 10에 나타난 것과 같이 정보 블록에서 사용자 프로파일을 추출한 제안 시스템의 경우, 보다 의미 있는 단어들 추출될 수 있음을 쉽게 확인할 수 있다. 또한 그림 11에 나타난 것과 같이 사용자가 구매한 상품 페이지와 사용자 프로파일의 단어는 다른 상품 페이지에 비해 일치율이 높으며, 제안 시스템이 TF*IDF만을 이용한 프로파일 생성 방법보다 높은 일치도를 나타냄을 확인할 수 있었다. 또한 실험 결과에 구매 아이템보다 일치율이 높은 상품도 존재하였다. 우리는 이러한 상품을 향후 구매 가능한 상품으로 분석하였다.

결과적으로 본 실험을 통해, 제안 시스템을 추천 시스템에 적용할 경우 사용자가 선호하는 상품이 추천될 수

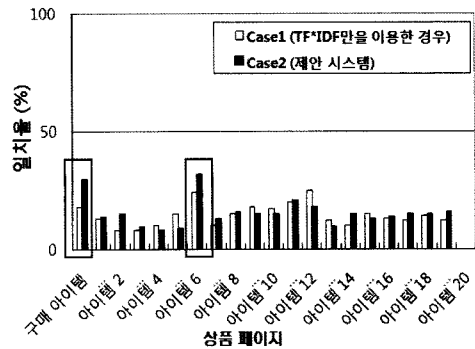


그림 11 사용자 프로파일과 상품 페이지의 비교

문서분석방법	관심단어와 가중치					
	Case 1 (TF*IDF)	A/S	구매	무료	네이버	DSLR
	48	43	35	29	29	...
Case 2 (Proposed System)	D40	DSLR	니콘	올림푸스	EasyShare	...
	41	32	29	27	15	...

그림 10 TF*IDF만을 사용한 경우와 제안 시스템을 통해 추출된 사용자 프로파일의 비교

있다는 가능성을 확인할 수 있었다. 그러나 이는 단순히 키워드의 일치도만을 비교하였기 때문에 실험 이전에 예상했던 수치보다는 낮은 결과였다. 향후 온톨로지(ontology)가 추가되어 동의어를 분석할 수 있다면, 본 실험 결과보다 높은 일치도를 나타낼 것으로 기대된다[17].

또한 상품을 구매하기 이전에 그 상품과 관련된 정보들을 검색한 사용자의 경우에 높은 수치가 나타나지만, 그렇지 않은 경우 일치도가 높지 않다는 것도 확인하였다. 따라서 일반적으로 상품을 구매하기 이전에 충분한 정보 탐색과 신중한 평가를 거친 후 구매를 수행하는 고관여제품(High-involvement product)의 경우, 제안 시스템이 보다 효과적으로 활용될 수 있을 것으로 생각된다.

4.2 프로파일 분석에 소요되는 시간 평가

두 번째 실험은 정보 블록만을 이용하는 제안 시스템의 프로파일 생성과 비정보 블록을 모두 포함한 기존 프로파일 생성 방법의 처리 시간을 비교하기 위해 수행되었다. 이를 위해, 웹 문서 데이터의 양을 변화시키며 각각 분석에 소요되는 시간을 비교하였다. 본 실험에는 정확도 실험에 사용된 사용자의 웹 문서 데이터를 동일하게 사용하였으며, 먼저 적은 수의 문서를 이용하여 프로파일을 생성하고, 문서의 수를 점차 증가시키며 시간 변화를 체크하였다.

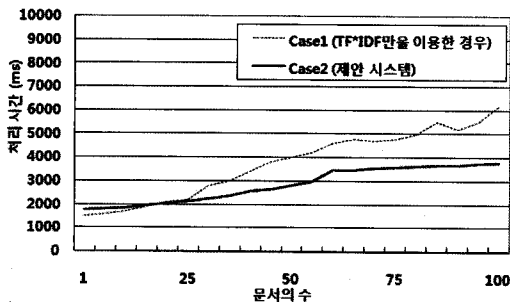


그림 12 프로파일 분석에 소요되는 시간 비교

실험 결과, 제안 시스템은 웹 문서의 DOM 트리를 생성하고 동일한 블록을 식별하기 위해 추가적인 시간이 필요하므로, 문서의 수가 적은 경우에는 기존 방법보다 시간이 다소 더 소요되었다. 그러나 문서의 수가 증가함에 따라 제안 시스템은 기존 방법에 비해 적은 증가율을 나타내었다. 이는 동일한 서버로부터 얻어진 문서의 비정보 블록들을 프로파일 분석에서 제외시키기 때문이며, 본 실험 결과를 통해, 제안시스템은 프로파일의 분석 속도를 효과적으로 단축시킬 수 있음을 확인할 수 있었다. 일반적으로 사용자들은 하나의 서버(동일한 레이아웃을 가지는)에서 다양한 문서들을 이용한다.

5. 결론

전자상거래 분야에서, 사용자의 만족도를 높이고 상거래를 보다 활성화시키기 위해 개인화 서비스는 매우 중요한 요소이다. 개인화 서비스의 핵심은 사용자의 취향 정보를 정확하게 분석하는 것이며, 이를 위해 대부분의 상점에서는 구매 정보와 같은 사용자의 암시적인 행동 정보를 이용한다. 그러나 이러한 분석은 일정 기간의 사용자 행동 정보를 필요로 하기 때문에, 사용자가 처음 방문하는 상점의 경우 추천 전략을 세우는 것이 매우 어렵다. 또한 사용자의 행동 정보가 수집된 경우에도, 오랜 기간 사용자의 방문이 없었다면 동적으로 변화하는 사용자의 취향을 반영하기 어렵다는 문제가 있다.

따라서 본 논문에서는 사용자의 행동을 지속적이고 통합적으로 관찰할 수 있는 사용자의 기기에서, 사용자가 이용한 웹 문서를 기반으로 취향 정보를 추출하고, 이를 상점에서 활용하는 구조의 시스템을 제안하였다. 또한 보다 정확하고 효과적으로 사용자의 관심 단어를 추출하기 위해, 동적으로 생성되는 웹 문서에서 일반적으로 나타나는 비정보 블록들을 식별하여 제거하고 프로파일을 분석하는 기법과, 사용자가 선택하는 하이퍼링크로부터 추출되는 단어에 가중치를 부여하는 기법을 적용하였다.

또한 웹 문서에서 특정 상품과 크게 관련이 없는 모든 웹 페이지를 분석하여 추출된 사용자 프로파일을 추천에 이용한다면 그 정확도는 매우 낮을 것이다. 따라서 제안 시스템은 웹 문서들을 그것이 포함하고 있는 단어들을 기반으로 유사도를 비교하여 그룹화하고, 여러 종류의 프로파일을 생성한다. 이를 통해 개인화를 제공하는 상점에서는 상점의 특성과 가장 유사한 프로파일만을 참고하고 개인화 전략을 세울 수 있다.

본 논문에서는 제안 시스템의 평가를 위해, 최근 구매 활동이 있었던 사용자의 웹 문서 데이터를 이용하여 프로파일을 생성하고 구매 데이터와의 일치도를 비교하였으며, 다른 상품 페이지에 비해 구매 상품 페이지가 높은 일치도를 나타내는 결과를 통해, 시스템의 유효성을 확인하였다. 또한 비정보 블록을 모두 포함하는 기존 방법에 비해 정확도나 소요되는 시간 측면에서 높은 효율성을 나타냄을 확인하였다.

최근 다수의 가전 기기들은 웹을 기반으로 하는 정보 기기(예. 사용자가 원하는 방송을 선택하는 양방향 TV, 요리 방법을 제공하는 주방 기기 등)로 발전하고 있다. 따라서 사용자 기기에 내장된 사용자 모듈 간 상호작용을 통해, 각 기기를 사용하면서 분석된 사용자 취향 정보가 취합된다면, 본 발명의 향후 적용 범위는 개인화 서비스를 필요로 하는 모든 분야에 적용될 수 있을 것이다.

향후 과제로는 실험 결과를 통해 확인된 여러 문제를 보완하기 위해, 온톨로지를 적용하여, 같은 키워드이지만 여러 문맥에서 다른 의미로 쓰이는 다중 의미 문제나 동의어 문제를 해결하는 연구를 진행할 것이다. 또한 멀티미디어 데이터와 같이 문서 분석 기법만으로는 의미를 추출하기 어려운 웹 데이터에 대한 사용자 취향 분석 방법을 연구할 것이며, 사용자의 기기에 수집된 프로파일 정보의 부분별한 유출을 막기 위해, 각 모듈 간 협상(negotiation)과정을 통한 메시지 보안 관련연구도 진행될 것이다. 마지막으로 보다 다양한 환경의 사용자들을 대상으로 실험을 확장하여 시스템의 추가적인 개선 방향을 모색할 것이다.

참고 문헌

[1] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations Item-to-Item Collaborative Filtering," IEEE Internet Computing, pp. 76-80, Jan. 2003.

[2] M. Balabanovic and Y. Shoham, "Fab: Content-based, Collaborative Recommendation," Communication of the ACM, pp. 66-72, Mar. 1997.

[3] 김태훈, 최중민, "사용자 편의의 인터넷 정보검색을 위한 지능형 웹 브라우징 에이전트", 정보과학회논문지(B), 제25권 제7호, pp. 1064-1078, Jul. 1998.

[4] S.-H. Ha, "Digital Content Recommender on the Internet," IEEE Intelligent Systems, pp. 70-77, Mar. 2006.

[5] Z. Yu and X. Zhou, "TV3P: An Adaptive Assistant for Personalized TV," IEEE Trans. on Consumer Electronics, Vol.50, No.1, pp. 393-399, Feb. 2004.

[6] L. Chen, and K. Sycara, "WebMate: A Personal Agent for Browsing and Searching," Proc. 2th international conference on Autonomous Agent, pp. 132-139, 1998.

[7] R. Armstrong, D. Freitag, T. Joachims, and T. Mitchell, "WebWatcher: A Learning Apprentice for the World Wide Web," Proc. AAAI spring symposium series on Information Gathering from Distributed, Heterogeneous Environments, pp. 6-13, Feb. 1995.

[8] 문현정, 김교정, "웹 도큐먼트 기반 연관 지식 추출기법: 생명정보분야에의 적용", 한국 인터넷 정보학회, 제2권 제5호, pp. 9-19, Dec. 2001.

[9] S. Debnath, P. Mitra, N. Pal, and C.-L. Giles, "Automatic Identification of Informative Sections of Web Pages," IEEE Trans. on Knowledge and Data Engineering, Vol.17, No.9, pp. 1233-1246, Sep. 2005.

[10] D. Cai, S. Yu, J.-R. Wen and W.-Y. Ma, "Block-based Web Search," Proc. 27th Ann. international ACM SIGIR conference, pp. 456-463, 2004.

[11] R. Song, H. Liu, J.-R Wen, and W.-Y. Ma,

"Learning Block Importance Models for Web Pages," Proc. 13th World Wide Web conference, pp. 203-211, 2004.

[12] Y. Chen, X. Xie, W.-Y. Ma, and H.-J. Zhang, "Adapting Web Pages for Small-Screen Devices," IEEE Internet Computing, pp. 50-56, Jan. 2005.

[13] T. Laakko and T. Hiltunen, "Adapting Web Content to Mobile User Agents," IEEE Internet Computing, pp. 46-53, Mar. 2005.

[14] S. Robertson, "Understanding Inverse Document Frequency: on theoretical arguments for IDF," Journal of documentation, Vol.60, No.5, pp. 503-520, 2005.

[15] G. Salton, "Introduction to Modern Information Retrieval," Mcgraw Hill, 1983.

[16] <http://www.w3.org/dom/>

[17] S. Aciar, D. Zhang, S. Simoff, and J. Debenham, "Informed Recommender: Basing Recommendations on Consumer Product Reviews," IEEE Intelligent Systems, pp. 39-47, Jun. 2007.



이 승 화

2003년 배재대학교 정보통신공학과(공학사). 2005년 성균관대학교 컴퓨터공학과(공학석사). 2008년 성균관대학교 컴퓨터공학과(공학박사). 2008년~현재 성균관대학교 정보통신공학부 연구교수. 관심분야는 소프트웨어공학, 상황적응형 시스템, 지능형 에이전트 등



최 형 기

1992년 성균관대학교 전자공학과(공학사) 1996년 Polytechnique University 전기전자(공학석사). 2001년 Georgia Institute of Technology 전기전자(공학박사). 2001년~2004년 미국 Lancop Inc. 연구원 2004년~2006년 성균관대학교 정보통신공학부 전임강사. 2006년~현재 성균관대학교 정보통신공학부 조교수. 관심분야는 인터넷보안, 모바일 커뮤니케이션 등



이 은 석

1985년 성균관대학교 전자공학과(공학사) 1988년 일본 Tohoku(동북)대학교 정보공학과(공학석사). 1992년 일본 Tohoku(동북)대학교 정보공학과(공학박사). 1992년~1993년 일본 미쯔비씨 정보전자연구소 특별연구원. 1994년 일본 Tohoku(동북)대학교 Assistant Prof. 1995년~현재 성균관대학교 정보통신공학부 교수. 관심분야는 소프트웨어공학, 유비쿼터스 컴퓨팅, 오토노믹컴퓨팅, 에이전트지향지능형시스템 등