문서 특성에 대한 선호도 기반 웹 검색 개인화

이수정
경인교육대학교 컴퓨터교육과

요 약

웹 상에서 사용자가 원하는 정보를 효율적으로 검색하는데 도움을 주기 위하여 웹 개인화는 사용자에게 흥미있는 웹 문서들을 추출해내는데 초점을 두고 있다. 이를 실현하기 위한 주요 방법들 중 하나는 문서에 포함된 질의어, 링크 및 사용자의 선호어를 이용하는 것이다. 본 연구에서는 이들 요소 외에 사용자들이 웹문서를 선택할 때 중요하게 생각하는 문서 특성들을 설문을 통하여 조사하였다. 설문 결과 문서의 내용이 가장 중요한 특성이었으나, 일부 사용자들에게는 문서에 포함된 이미지와 가독성도 내용과 마찬가지로 중요하게 간주되었다. 이를 바탕으로 각 사용자를 위한 문서의 주요 특성들의 상대적 가중치를 프로필에 유지 관리하고, 검색 결과의 개인화에 반영하는 방안을 제시한다. 제안한 개인화 방법의 성능을 분석한 결과, 일반 검색 엔진에 비해 최대 약 2.3배의 성능 향상을 보였고, 사용자 질의어와 선호어를 모두 이용하여 검색 결과를 산출하는 방법보다 약 1.5배의 성능 향상을 나타내어 그 우수성을 입증하였다.

키워드: 정보 필터링, 웹 개인화, 협력 필터링, 추천 시스템

Web Search Personalization based on Preferences for Page Features

Soojung Lee
Dept. of Computer Education, Gyeongin National University of Education

ABSTRACT

Web personalization has focused on extracting web pages interesting to users, to help users searching wanted information efficiently on the web. One of the main methods to achieve this is by using queries, links and users' preferred words in the pages. In this study, we surveyed from the web users the features of pages that are considered important to themselves in selecting web pages. The survey results showed that the content of the pages is the most important. However, images and readability of the page are rated as high as the content for some users. Based on this result, we present a method for maintaining relative weights of major page features differently in the profile for each user, which is used for personalizing web search results. Performance of the proposed personalization method is analyzed to prove its superiority such that it yields as much as 1.5 times higher rate than the system utilizing both queries and preferred words and about 2.3 times higher rate than a generic search engine.

Keywords: Information Filtering, Web Personalization, Collaborative Filtering, Recommendation System

논문투고: 2010-07-20
논문심사: 2011-03-23
심사완료: 2011-04-04

1. 서론

웹 문서 수가 급증함에 따라 인터넷에서 원하는 정보를 빠른 시간 내에 검색하기가 점차 어려워짐에 따라 거의 90%의 검색 결과가 사용자가 원하지 않는 불필요한 것으로 밝혀졌다. 이 문제를 경감시키기 위해 사용자의 선호도에 부합하는 웹 환경을 조성하여 주는 개인화 작업이 주목받고 있다. 현재 웹 개인화의 가장 유명한 예는 주로 고객이 원하는 상품을 찾을 수 있도록 도움을 제공하는 추천 시스템으로서 Amazon.com과 MovieLens 등이 있다[3]. 웹 개인화의 주요 범주인 웹 검색 결과의 개인화를 구현하기 위한 주된 방법은 협력 필터링과 정보 필터링인데 이들은 모두 사용자가 관심을 보일만한 항목들을 식별하여 정보 과부하를 경감시키고자 하는 시도이다.

협력 필터링은 여러 다른 사용자들의 의견에 따라 웹 사이트를 추천하는 방법이다. 이는 사용자들이 공통된 흥미를 갖고 있음을 가정하여, 가장 인기 있는 웹 문서를 제안하며, 서적, 식품점, 엔터테인먼트 등 다양한 영역에서 사용되었다. 이 방법의 특징은 추천 항목의 내용을 고려하지 않은 채, 유사한 특성을 지닌 타인들이 선호하는 항목들을 추천하는 것이다. 유사성은 cosine 기반, 상관도, 평균 절대오차 및 기타 다른 여러 방법들에 의해 계산한다[8]. 그러나 이 방법의 단점은 새로운 문서에 대해서는 축적된 사용자 선호도 정보가 없기 때문에, 많은 사용자로부터 평가 정보를 필요로 하는 것이다.

이와 달리 정보 필터링은 내용 분석을 토대로, 개인적인 사용자 흥미도 프로필을 구축하는 데 초점을 둔다. [10]에서 제안한 시스템은 사용자의 검색 이력으로부터 선호 범주를 학습한다. 사용자 프로필의 형태는 다양하게 구축되었는데, Daoud 등[5]은 그래프 기반의 프로필을 구축하여 검색 결과를 개인화하는 방안을 제시하였는데, 상호 관련된 개념들을 활성화시키는 방식을 취하였다.

이와 같은 정보 필터링 방법들은 협력 필터링의 단점을 극복할 수 있는데, 아직 평가되지 않은 새로운 문서에 대해 그 내용을 살펴봄으로써 사용자의 흥미 여부를 예측할 수 있기 때문이다. 그러나 프로필에 축적된 사용자의 흥미도 외에 다른 새로운 흥미로운 정보를 발견할 수 없다는 단점이 있다.

사용자별 특징을 가미한 개인화 시스템이 아닌 일반적으로 검색 결과를 개선하기 위해 링크 기반의 순위 계산 기법들이 다양하게 개발되었다. 이들 방법들은 대체로 HITS 알고리즘과 PageRank 알고리즘을 기초로 하였는데, Kleinberg가 개발한 HITS 알고리즘[9]은 검색 결과 문서들과 그들과 근접한 문서들로 구성된 웹 그래프에서 hub와 authority들을 검색 진행 시간 내에 추론한다. 한편 PageRank 알고리즘[4]은 웹 상의 모든 문서들에 대해 미리 그 중요도를 계산하여 순위 벡타를 유지하였다. 이 방법은 보다 많은 수의 유입 링크를 가진 문서는 그 숫자가 적은 문서보다 더욱 중요할 것이라는 아이디어를 기반으로 한 것이다.

위 방법들은 사용자 흥미를 고려하지 않았으므로, 보다 개인화된 검색 결과를 위해 사용자 정보를 추가로 이용한 방법들이 개발되었다. Haveliwala[7]는 Open Directory Project[12]의 주요 주제 16 종류를 기반으로 미리 PageRank 벡타를 형성한 후, 질의어에 적합한 주제를 확률적으로 계산하여, 각 주제에 해당하는 문서들의 순위를 결정하였다. Teevan 등[11]은 사용자 프로필을 이용하여 검색 결과 문서의 순위를 산정하였는데, 사용자 컴퓨터 내에 저장된 각종 파일들을 프로필로 간주하여, 사용자의 흥미도를 간접적으로 파악하였다.

사용자 기기에 저장된 로그데이타를 분석하여 개인화 기법에 활용한 방안으로서 이승화 등[2]의 방법은 로그 데이터로부터 사용자의 관심분야를 추론하고 웹 문서에서 정보 블록만을 식별하여 프로필에 반영함으로써 사용자 정보가 부족한 서버에서 초기 사용자에게 추천 항목을 제시할 수 있도록 하고 사용자의 선호 변화를 반영하였다. 박건우와 이상훈[1]은 사용자의 검색 의도와 관심사에 보다 근접한 검색 결과를 추출하기 위해, 질의어 사용 빈도수와 이에 따른 순위를 데이터베이스로 구축한 후 질의어에 대한 랭킹 정보를 통해 사용자의 주요 관심사를 파악하고 주요 관심사별 커뮤니티를 형성하여 검색을 수행하는 시스템을 제안하였다.

이상과 같이, 사용자의 특성과 웹 검색 행태에 기반하여 검색 결과를 추출하려는 많은 노력들이 진행

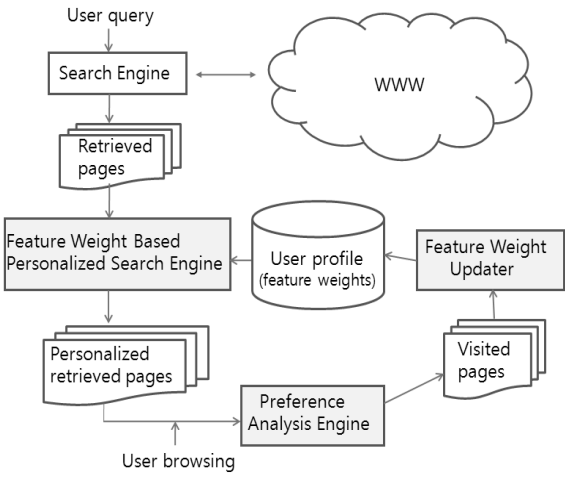
되었으나, 문서 특성에 대한 선호도에 기초하여 웹 검색 결과를 개선하려는 시도는 부재하다. 따라서, 본 연구에서는 사용자들의 문서 특성들에 대한 선호 여부를 조사하고, 이를 기초로 문서 순위를 산출하는 시스템을 개발한다. 문서의 주요 특성들을 선정하기 위하여, 사용자들로부터 6개의 주요 공통 특성들을 설문 수집하였다. 각 특성에 대한 개인별로 다른 선호도를 사용자 프로필에 상대적 가중치로서 유지 보관하였다. 조사 결과, 문서의 내용이 가장 주요한 특성으로 선정되었으나, 문서의 가독성과 문서가 포함하는 이미지 또한 주요 특성임이 밝혀졌다. 이러한 결과에 기초하여, 사용자 프로필에 보관된 각 특성별 가중치 분포를 이용한 검색 결과의 개인화 방법을 개발하였고 그 성능을 설문조사의 실제 데이터에 대한 실험을 통해 증명하였다.

본 논문의 구성은 다음과 같다. 2절에서는 제안된 시스템의 구조와 알고리즘이 제시되며, 3절에서 설문조사 결과와 타 방법과의 성능 비교 결과가 제시된다. 4절에서 논문의 결론을 맺는다.

2. 개인화된 웹 검색 시스템

2.1 시스템 구조

제안된 시스템은 세 개의 주요 프로그램으로 구성된다: 특성 가중치 기반 개인화 검색 엔진(Feature weight based Personalized Search Engine, FPSE), 선호 분석 엔진(Preference Analysis Engine, PAE), 그리고 특성 가중치 관리자(Feature Weight Updater, FWU)이다. (그림 1)에서 일반 검색 엔진의 검색 결과는 FPSE에 전달되며, 이 프로그램은 특성 가중치를 보관하고 있는 사용자 프로필을 참조하여 개인 흥미에 부합되는 문서들에 높은 순위를 부여한다. PAE는 검색 결과에 대한 사용자의 선호 여부를 파악하고, FWU는 선호 문서가 가진 특성들에 대해 그 가중치를 변경한다. PAE는 직접,간접적인 사용자 행동으로부터 흥미 여부를 파악한다. 간접 측정 방법의 개발은 본 연구 주제의 범위에서 벗어나므로 다루지 않는다.



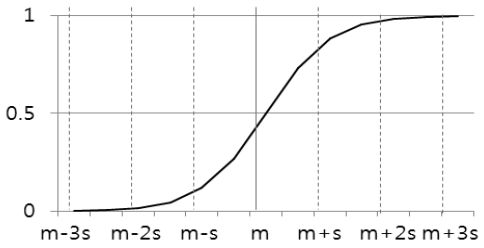
(그림 1) 제안 시스템의 구조

2.2 알고리즘

FPSE의 기본 아이디어는 각 특성별로 검색 결과 문서의 순위를 책정하고, 프로필 내의 각 특성 가중치를 순위에 반영하여 최종 순위를 산출하는 것이다. m_f와 s_f를 특성 f에 대한 전체 사용자들의 선호값 평균과 표준편차라고 하자. 다음 절에서 도출한 문서 특성들에 대해 기술하지만, 특성들 중 이미지, 링크, 길이에 관한 문서 점수를 계산하기 위하여 가우스 분포 $N(m_f, s_f)$를 가정한다. 이는 설문 결과 이들 세 특성에 대하여, 평균에서 크게 벗어나는 특성값을 사용자가 거의 선호하지 않았음에 기인한다. 그러나, 내용 특성에 대해서는 (그림 2)와 같이 로지스틱 함수를 이용하는데, 이 함수가 설문 조사 결과를 가장 잘 대표한다고 판단되었기 때문이다. 즉, 그림에서 x축은 내용 특성값, y축은 내용에 따른 문서 점수를 나타내고, 특성값이 $[m-s, m+s]$ 이내에 존재하면, 문서점수는 일직선상으로 변화하나, 그 밖의 범위에서는 미미한 점수 변화를 일으킨다.

Q를 사용자가 입력한 질의어 집합, P를 Q와 관련된 선호어 집합, 그리고 D를 일반검색엔진에서 산출한 문서 집합이라고 하자. 사용자가 질의할 때 검색 결과문서 d에 대하여, FPSE는 (그림 3)과 같이 작동한다. 웹 문서의 내용 점수는 TF*IDF 방법 [6]에 의해 평가하는데, 6 단계에서 보듯이 질의어 뿐만 아니

라 선호어에 대해서도 이 방법을 적용한다. 단, 선호어에 대해서는 질의어의 0.5배의 영향력을 갖도록 한다. 9단계의 w_f는 (그림 4)에 제시한 FWU에 의해 유지 관리된다.



(그림 2) FPSE의 로지스틱 함수: $m=m_{content}$, $s=s_{content}$

1. $score_{image,d} = N(m_{image}, s_{image})(x)$. x는 d에 포함된 이미지 수.
2. $score_{link,d} = N(m_{link}, s_{link})(x)$. x는 d에 포함된 링크 수.
3. $score_{length,d} = N(m_{length}, s_{length})(x)$. x는 d 문서의 길이.
4. $\forall q \in Q$에 대하여, $TF_{q,d}$는 d에 포함된 q의 개수를 d가 포함한 전체 단어수로 나눈 값.
5. $\forall q \in Q$에 대하여, $IDF_{q,d} = 1+\log(|D|/|\{d \in D | q \in d\}|)$.
6. $QP_d = \sum_{q \in Q \wedge d} TF_{q,d} \cdot IDF_{q,d} + 0.5\sum_{p \in P \wedge d} TF_{p,d} \cdot IDF_{p,d}$
7. $score_{content,d}=1/(1+e^{-(QPd-m)})(x)$, $m=m_{content}$
8. 특성 f에 대하여 $\{score_{f,d} \mid d \in D\}$를 내림차순으로 정렬한 후 $rank_{f,d}$를 d의 특성 f에 대한 순위라고 하자.
9. $SCORE_d = \sum_f w_f (|D|-rank_{f,d})$. w_f는 사용자 프로필에 보관된 특성 f에 대한 가중치.

(그림 3) FPSE 알고리즘

FWU 프로그램은 사용자가 문서에 대한 선호를 보일 때마다 작동된다. 이 프로그램은 사용자의 문서 선호 이유를 추정하고 그 정도를 반영하기 위하여, 내용을 제외한 각 특성값에 대하여 가우스 분포를 가정하여, 특성값에 해당하는 분포 함수값에 비례하도록 특성 가중치를 변경한다. 내용 특성에 대하여는 FPSE에서처럼 로지스틱 함수를 사용하여 가중치를 변경하는데, 단, 문서 d에 대한 QPd 값이 평균($m_{content}$)보다 크다면 내용 특성이 문서 선호 이유라고 판단하여 높은 가중치를 부여하고, 그렇지 않은 경우 가중치를 급격히 떨어뜨리는 방식을 취한다. 자세한 사항은 (그림 4)에 제시하였다. 그림에서 α는 가중치 증가의 최대값을 결정하는 파라미터이다. 5단계에서 최종적으로 모든 특성 가중치를 정규화한다.

1. $w_{image} \leftarrow w_{image} + \alpha \cdot N(m_{image}, s_{image})(x)$. x는 d에 포함된 이미지수. $0<\alpha\leq 1$.
2. $w_{link} \leftarrow w_{link} + \alpha \cdot N(m_{link}, s_{link})(x)$. x는 d에 포함된 링크수.
3. $w_{length} \leftarrow w_{length} + \alpha \cdot N(m_{length}, s_{length})(x)$. x는 d 문서 길이
4. $w_{content} \leftarrow w_{content} + \alpha \cdot 1/(1+e^{-(QPd-m+2s)})(x)$. $m=m_{content}$, $s=s_{content}$.
5. $w_f = w_f / max(w_f)$로 정규화한다.

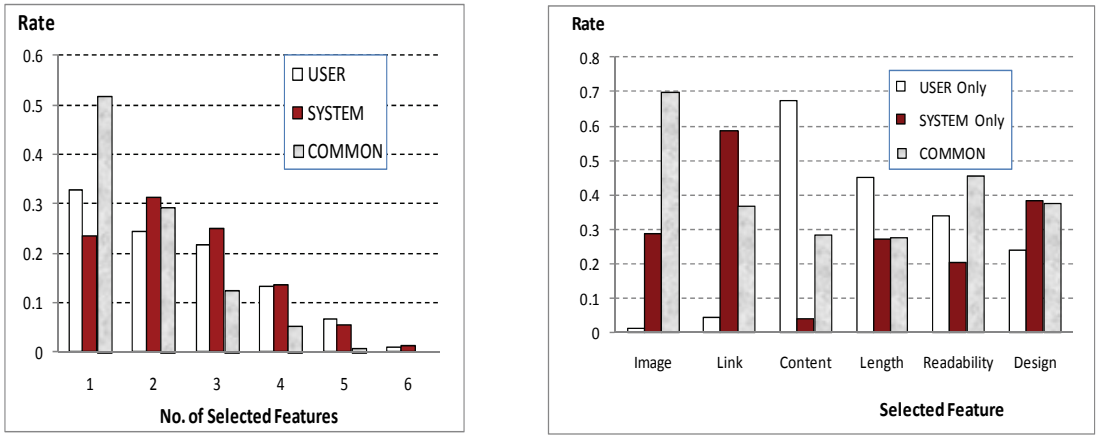
(그림 4) FWU 알고리즘

3. 실험 연구

3.1 실험 배경

실험은 두 단계로 진행되었다: 첫단계에서는 선호 문서에서 주요 특성을 추출하고 둘째 단계에서는 제안된 방법의 성능을 조사한다. 다양한 연령대의 9명의 실험자들이 두 단계의 실험에 참여하였다. 실험자들은 첫단계의 실험에서 임의의 웹 검색 결과 문서들 중 선호 문서들에 대해 그 이유를 작성하였고, 이들 중에서 6개의 주요 공통 항목을 도출하였다. 도출된 항목은 이미지, 링크, 내용, 길이, 가독성과 디자인이다. 둘째 단계에서는 각 실험자가 총 30번의 검색을 하였으며, 각 검색 결과 문서를 대상으로 선호문서를 선택하고, 그 이유를 예비실험에서 도출된 항목들 중에서 하나 이상 선택하였다. 또한 각 검색 질의어와 관련된 선호어들을 나열하였다.

FWU 알고리즘에서 사용하는 α 파라미터값은 1로 하였는데 이는 적은 량의 검색 데이터로 가중치의 급격한 변화가 성능에 미치는 영향을 알아보기 위함이다. 각 특성의 가중치와 평균과 표준편차를 측정하기 위하여 실험 데이터의 1/3을 사용하였고, 나머지는 성능측정에 사용되었다. 성능 비교 대상은 SE, LNK, QRY, Q&P와 제안한 방법이다. SE는 일반검색엔진(http://www.naver.com)의 검색결과를 말하고, LNK는 기존의 여러 알고리즘[4]들에서 주요 주제로 다루었는데, 본 연구에서는 outlink수에 비례하여 높은 우



(그림 5) 선택된 특성 개수의 평균(좌)과 각 특성별 선택 부합률의 평균(우)

선순위를 부여하였다. QRY 방법은 질의어의 TF*IDF 합을 문서 점수로 계산하고 Q&P 방법은 FPSE의 6단계에서 제시한 QP 값에 따라 문서 순위를 결정한다. 각 방법의 문서 점수 계산식을 <표 1>에 요약하였다.

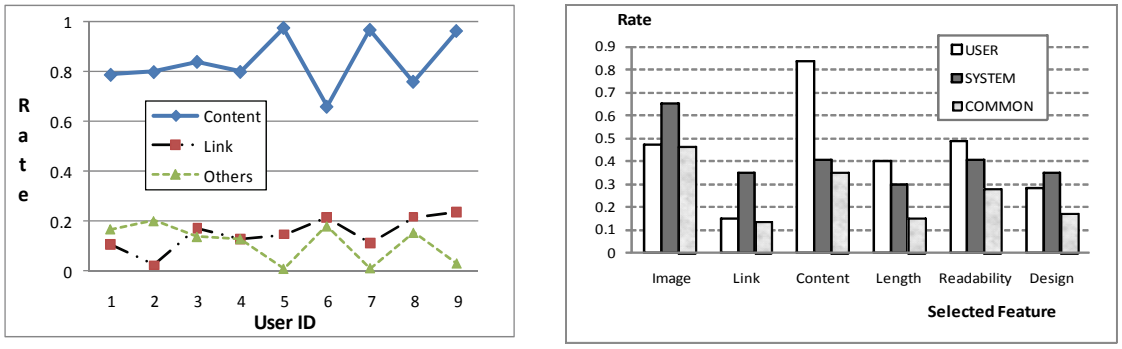
<표 1> 각 방법의 검색 결과 문서에 대한 점수 계산식

방법	문서 d의 점수
LNK	d가 포함한 outlink수 / d 문서 길이
QRY	$QP_d = \sum_{q \in Q \wedge d} TF_{q,d} \cdot IDF_{q,d}$
Q&P	$QP_d = \sum_{q \in Q \wedge d} TF_{q,d} \cdot IDF_{q,d} + 0.5 \sum_{p \in P \wedge d} TF_{p,d} \cdot IDF_{p,d}$

평가 기준으로 기존 연구에서 흔히 사용하는 precision, recall 이외에 추가로 rank rate를 정의한다. 성능 측정에 사용할 문서 범위를 Domain Size(DS)라고 하고, 각 방법에서 산출한 1위부터 DS 순위까지의 문서들을 기준으로 평가한다. Precision P는 N/DS로 측정하는데, N은 DS 문서들 중 실험자의 선호문서수이다. Recall R은 실험자의 전체 선호문서(U로 표기) 중 DS 내에 포함된 문서 수이며, N/U로 측정한다. 또한 성능을 한눈에 편리하게 비교하기 위하여 precision과 recall의 비중을 동일하게 통합한 F1=2PR/(P+R)을 사용한다[10]. 그러나 이들 척도들은 DS 내 포함된 선호문서수만을 나타내고 그 순위는 고려하지 않으므로 본 연구에서는 rank rate를 정의하기로 한다. 즉, DS 순위 내 선호문서들의 순위합을 최적의 문서 순위합에 대한 비율로 나타낸다. 예로서, DS 내 선호문서들의 순위가 2, 5, 7이라면 rank rate = (2+5+7) / (1+2+3) = 2.33이 된다.

3.2 설문 조사 결과

(그림 5(좌))는 실험자들이 하나의 선호 문서 당 선택한 특성개수의 평균을 나타낸다('USER'로 표기). 'SYSTEM'은 시스템에서 정한 객관적인 기준을 의미하고, 'COMMON'은 실험자와 SYSTEM에 의해 공통적으로 선택된 특성개수이다. 그림에서 실험자들은 대체로 단 하나의 특성만을 가장 많이 선택하였고, 그다음으로 두 개와 세 개의 특성들을 많이 선택하였다. 반면에 SYSTEM에 의하면 대개 두 개의 특성이 선택되었다. COMMON 그래프에 따르면 공통적으로 선택된 특성은 한 개로서 월등히 높은 비율을 차지하므로 실험자와 SYSTEM의 기준이 잘 부합되지 않음을 알 수 있다. 실험자의 기준과 시스템의 기준을 비교하기 위하여 (그림 5(우))에 각 특성별 선택 부합률을 제시하였다. 실험자만의 독자적인 기준으로 가장 많이 선택된 특성은 'User Only'에서 알수 있듯이 내용과 문서길이였다. 이는 실험자들이 내용과 길이를 중시하지만 객관적인 통일된 선호기준은 없음을 의미한다. 또한 대부분의 실험자들이 이미지를 선호함을 알 수 있다.

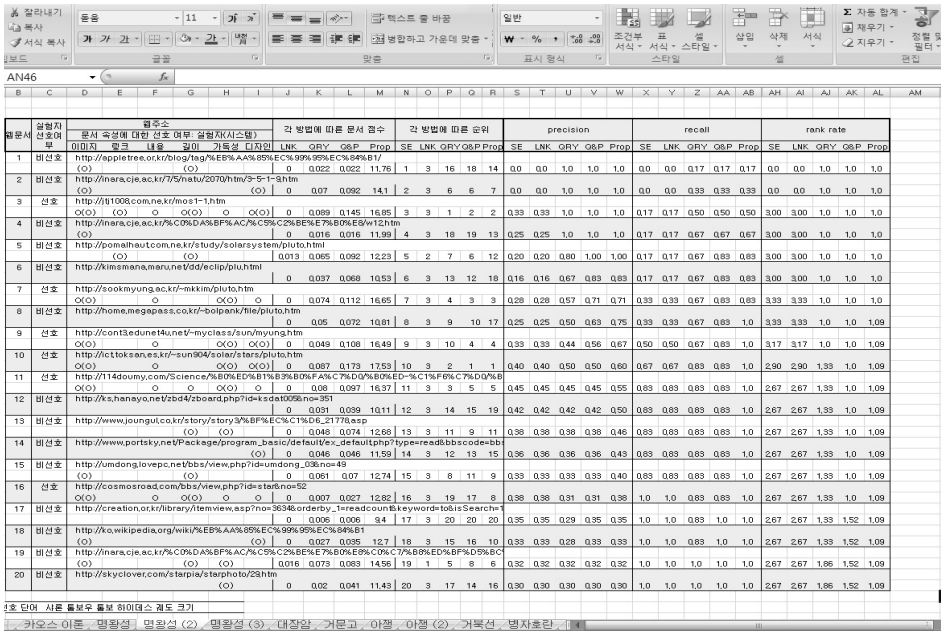


(그림 6) 실험자별 내용, 링크 및 기타 특성 선택률(위)와 특성 선택률 분포(아래)

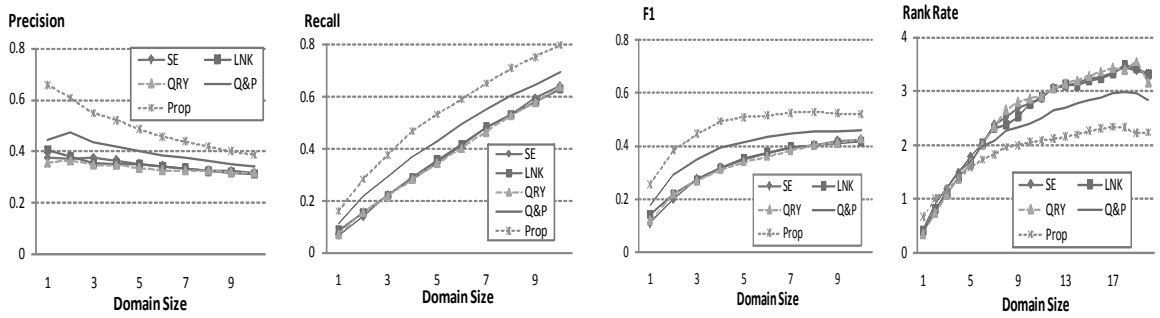
(그림 6)(좌)에 내용, 링크 및 기타 특성이 선택된 선호 문서의 비율을 각 실험자별로 제시하였다. 실험자 5는 거의 모든 선호문서에서 내용을 선호 이유로 선택한 반면 실험자 6은 그 비율이 약 60%에 머뭄을 알 수 있다. 그러나 전체 실험자들이 선호 문서에서 내용을 선택한 비율은 약 83%이고 내용 이외의 특성 때문에 문서를 선택한 비율이 17% 정도이었다. 링크 선호는 뜻밖에도 약 20%를 넘지 못하였다. 선호 문서에 대한 각 특성 선택률은 (그림 6)(우)에 나타냈다. 실험자와 SYSTEM 간에 가장 큰 차이는 내용과 링크 특성에서 보였고, 가독성과 디자인에서 가장 낮은 차이를 보였다. 따라서 문서 내용 측면에서 실험자의 기준과 SYSTEM에서 근거로 한 TF*IDF 기준과는 상당한 차이가 있음을 알 수 있다. 실험자들은 내용 특성을 가장 중시하고, 그다음으로 가독성과 이미지를 중시하였다.

3.3 성능 비교

(그림 7)은 한 실험자의 '명왕성'을 질의어로 한 네



(그림 7) 검색 결과 문서들의 특성에 대한 선호 여부, 문서 점수 및 순위 데이터 예시



(그림 8) Domain size에 따른 성능 비교

이버 검색엔진의 검색 결과 중 첫 20 문서에 대한 각종 데이터 자료이다. 질의어 외에 실험자가 제공한 선호단어는 샤론, 톰보우, 하이데스, 궤도이었다. 그림에서 둘째 열에 문서에 대한 실험자의 선호여부를 표시하였고, 6개 특성에 대한 실험자 및 시스템의 선호여부를 나타냈다('O(O)' 표시는 실험자와 시스템이 모두 선호함을 의미). 앞에서 언급한 공식을 이용하여 각 방법에 따른 문서 점수를 산출하고 순위를 부여하였다.

다양한 domain size에 대해 각 방법의 성능을 (그림 8)에 제시하였다. precision은 domain size가 커짐에 따라 감소하였고, QRY, LNK와 SE 간의 성능차는 거의 없으나 Q&P와 이들 사이의 성능차는 특히 작은 domain size에서 주목할 만하다. 따라서 실험자는 질의어나 링크수 뿐만 아니라 문서 선택에 있어서 선호어 또한 중요하게 고려함을 알 수 있다. 한편 제안된 방법('Prop')은 Q&P에 비해 최대 약 46%, 다른 세 방법들에 비해 약 85%의 precision 성능 향상을 보였다. Recall에 있어서도, Q&P는 LNK와 SE에 비해 각기 최대 약 1.4배와 1.7배 우수하다. Prop은 LNK, SE와 Q&P에 비해 각기 최대 약 1.8, 2.5, 1.5배로 우수함을 보였다. F1 결과에서 Prop은 5~10의 domain size에서 다른 방법들보다 약 22~45% 월등하고 그보다 작은 domain size에서는 SE와 Q&P에 비해 각각 최대 2.3배와 1.5배 우수하였다. Rank rate에 있어서 domain size가 커질수록 Prop은 타 방법들보다 서서히 증가함을 알 수 있다. 이는 제안된 방법이 실험자들의 선호 문서들을 보다 상위 순위에 배치함을 의미한다.

4. 결론

본 논문에서는 문서 특성에 대한 선호 가중치 분포에 의거하여 문서를 추천하는 웹 개인화 방법을 제안하였다. 사용자가 선호하는 문서들을 분석하여 프로필에 각 문서 특성의 가중치를 보관하여, 개인화 알고리즘에 활용한다. 설문조사로부터 사용자가 문서를 선택할 때, 문서의 내용 외에 이미지와 가독성이 영향을 주는 주요 요소임을 알 수 있었다. 이러한 결과를 기초로, 프로필에 보관된 각 특성별 가중치 분포를 이용한 검색 결과의 개인화 방법을 개발하였다. 성능 분석 결과 제안된 방법은 일반 검색 엔진의 결과에 비해 최대 약 2.3배의 성능 향상을 가져왔으며, 질의어와 선호어를 이용한 방법에 비해 최대 약 1.5배 우수하였다. 따라서, 기존 알고리즘에서 사용하는 문서 선호 기준들을 확장시킬 필요가 있으며, 확장된 통합 기준에 의거한 새로운 검색 알고리즘의 개발이 필요함을 알 수 있다.

참 고 문 헌

[1] 박건우・이상훈 (2009). 질의어 패턴 자동분석을 통한 커뮤니티 기반 개인화 검색. 정보과학회 논문지, 36-4, 321-326.

[2] 이승화・최형기・이은석 (2009). 사용자 기기에서 이용한 웹 데이터 분석을 통한 사용자 취향 분석 방법. 정보과학회 논문지, 15-3, 189-199.

[3] G. Adomavicius and A. Tuzhilin (2005). Toward the next generation of recommender systems:

a survey of the state-of-the-art and possible extensions, IEEE Transactions on Knowledge & Data Engineering, 17-6, 734 - 749.

[4] S. Brin, R. Motwani, L. Page, and T. Winograd (1998). What can you do with a web in your pocket. IEEE Data Engineering Bulletin, 21-2, 37-47.

[5] M. Daoud, L.-T. Lechani, and M. Boughanem (2009). Towards a graph-based user profile modeling for a session-based personalized search, Knowledge and Information Systems, 21-3, 365 - 398.

[6] J. Han and M. Kamber (2006). Data Mining: Concepts and Techniques, Morgan-Kaufmann.

[7] Haveliwala, T.H. (2002). Topic-sensitive pagerank. Proc. 11th Intl. World Wide Web Conference, Honolulu, Hawaii, 517-526.

[8] B. Jeong, J. Lee, and H. Cho (2010). Improving memory-based collaborative filtering via similarity updating and prediction modulation, Information Sciences, 180-5, 602 - 612.

[9] J. Kleinberg (1998). Authoritative sources in a hyperlinked environment. Proc. ACM-SIAM symposium on Discrete Algorithms, 668-677.

[10] F. Liu, C. Yu, and W. Meng (2004). Personalized web search for improving retrieval effectiveness, IEEE Trans. Knowl. Data Eng., 16-1, 28-40.

[11] J. Teevan, Dumais, S.T., & Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. Proc. Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 449-456.

[12] http://www.dmoz.org

저 자 소 개

이 수 정

1985 이화여자대학교 졸업(학사)

1990 Texas A&M 대학교 컴퓨터 공학과 졸업(석사)

1994 Texas A&M 대학교 컴퓨터 공학과 졸업(박사)

1994～1998 삼성전자 통신개발실 선임연구원

1998～현재 경인교육대학교 컴퓨터 교육과 교수

관심분야: 컴퓨터교육, 웹마이닝, 개인화된 웹검색

e-mail: sjlee@gin.ac.kr