개념 망을 통한 전자 카탈로그의 시맨틱 검색 및 추천

Semantic Search and Recommendation of e-Catalog Documents through Concept Network

이재원(Jae-won Lee)*, 박성찬(Sungchan Park)*, 이상근(Sangkeun Lee)*,
박재휘(Jaehui Park)*, 김한준(Han-joon Kim)**, 이상구(Sang-goo Lee)*

초 록

현재까지, 사용자의 요구에 맞는 카탈로그 문서를 제공하기 위해 널리 사용되고 있는 패러다임은 키워드 검색 혹은 협업적 필터링 기반 추천이다. 일반적으로 사용자의 질의어는 짧기 때문에, 사용자의 요구(질의어, 선호도)에 적합한 카탈로그 문서를 제공하는 것은 쉽지 않다. 이를 극복하기 위해 다양한 기법들이 제안되었으나, 이전 연구들은 색인어 매칭을 기반으로 하고 있다. 기존 베이지안 신념 망을 이용한 방법은 사용자의 요구 및 카탈로그 문서들을 연관성이 높은 개념들로 표현하였다. 하지만 개념들이 카탈로그 문서에서 추출된 색인어로 구성되어 있기 때문에 개념간의 관계 정보를 잘 표현하지 못하였다. 이에 본 연구는 베이지안 신념 망을 확장하여, 사용자의 요구 및 카탈로그 문서들을 웹 디렉토리에서 추출한 개념(혹은 카테고리) 망으로 표현한다. 개념 망을 이용함으로써, 사용자의 요구와 카탈로그 문서간의 개념 매칭도를 계산하는 것이 가능하다. 즉, 사용자의 질의어와 카탈로그 문서의 색인어가 일치하지 않을지라도, 개념적으로 관련성이 높은 문서를 검색하는 것이 가능하다. 또한 사용자간의 개념적 유사도를 계산함으로써, 시맨틱 기반의 협업적 필터링 추천이 가능하다.

ABSTRACT

Until now, popular paradigms to provide e-catalog documents that are adapted to users' needs are keyword search or collaborative filtering based recommendation. Since users' queries are too short to represent what users want, it is hard to provide the users with e-catalog documents that are adapted to their needs(i.e., queries and preferences). Although various techniques have beenproposed to overcome this problem, they are based on index term matching. A conventional Bayesian belief network-based approach represents the users'needs and e-catalog documents with their corresponding concepts. However, since the concepts are the index terms that are extracted from the e-catalog documents, it is hard to represent relationships between concepts. In our work, we extend the conventional

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 육성지원사업(NIPA-2010-C10 90-1031-0002)의 연구결과로 수행되었음.

이 논문은 2010 한국전자거래학회 학술대회에서 "전자 카탈로그의 시맨틱 검색 및 추천을 위한 개념 망 구축"의 제목으로 발표된 논문을 확장한 것임.

* 서울대학교 전기컴퓨터 공학부

** 서울시립대학교 전자전기컴퓨터 공학부

2010년 05월 14일 접수, 2010년 06월 24일 심사완료 후 2010년 06월 24일 게재확정.

Bayesian belief network based approach to represent users' needs and e-catalog documents with a concept network which is derived from the Web directory. By exploiting the concept network, it is possible to search conceptually relevant e-catalog documents although they do not contain the index terms of queries. Furthermore, by computing the conceptual similarity between users, we can exploit a semantic collaborative filtering technique for recommending e-catalog documents.

키워드 : 개념 망, 베이지안 신념 망, 시맨틱 검색, 시맨틱 추천
Concept Network, Bayesian Belief Network, Semantic Search, Semantic Recommendation

1. 서 론

현재까지, 사용자의 요구(질의어, 선호도)에 맞는 카탈로그 문서를 제공하기 위해 널리 사용되고 있는 패러다임은 키워드 검색 혹은 협업적 필터링 기반 추천이다. 검색 서비스는 사용자의 질의어와 관련성이 높은 카탈로그 문서를 제공하는 서비스이며, 협업적 필터링 기반 추천 서비스는 질의어 대신 선호도 정보를 이용하여 유사한 사용자(즉, 선호도가 비슷한 사용자)가 선호하는 카탈로그 문서를 제공하는 서비스이다. 일반적으로 사용자의 질의어(혹은 선호도 정보)는 짧기 때문에, 사용자의 요구에 적합한 카탈로그 문서를 제공하는 것은 쉽지 않다. 이를 극복하기 위해 다양한 기법들이 제안되었으나, 이전 연구들은 색인어 매칭을 기반으로 하고 있다. 예를 들어, 검색 서비스의 경우 사용자의 질의어와 카탈로그 문서의 색인어가 일치하지 않는 경우, 개념 적으로(혹은 시맨틱) 관련성이 높은 카탈로그 문서일지라도 사용자에게 제공되는 것이 불가능하다. 본 연구는 사용자의 질의어(혹은 선호도 정보)와 카탈로그 문서(혹은 사용자)간의 색인어 매칭이 아닌 개념 매칭을 이용하기 위해 베이지안 신념 망(Bayesian belief network; BBN)을 기반으로 한다.

BBN 기반 검색 모델[1, 2]은 사용자의 요구 및 카탈로그 문서를 연관성이 높은 개념들로 표현한다. 또한 BBN 기반 검색 모델은 사용자의 요구 및 카탈로그 문서를 연관성이 높은 개념들로 표현하기 위한 명료한 확률적 방법론을 제공하며, 키워드 검색에 비해 만족스러운 검색 결과를 제공하는 것으로 알려져 있다[1, 2]. 하지만, 기존 BBN 기반 검색 모델은 카탈로그 문서에서 추출된 색인어를 개념들로 간주하므로 개념간의 관계 정보 및 도메인 지식을 이용하는데 한계가 있다. 일반적으로 사용자는 자신의 도메인 지식을 기반으로 질의어 및 선호도 정보를 생성하므로, 검색 및 추천 시스템 등이 사용자의 질의어 및 선호도 정보를 이해하기 위해서는 방대한 양의 도메인 지식베이스를 이용해야 한다[3]. 현재 많은 도메인 지식베이스가 존재하지만, 최근 연구[5~7]들은 웹 디렉토리와 같은 분류 지식베이스를 도메인 지식베이스로 사용하고 있다. 또한 분류 지식베이스를 이용함으로써, 카테고리간의 관계를 이용하는 것도 가

능하다.

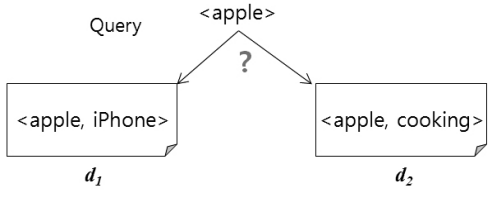
이에 본 연구는 베이지안 신념 망을 확장하여, 사용자의 요구 및 카탈로그 문서들을 웹 디렉토리와 같은 분류 지식 베이스에서 추출된 개념 망으로 표현한다. 특히 ODP(open directory project) 웹 디렉토리는 지금까지 알려진 분류 지식베이스 중에서 가장 크고 포괄적인 범위를 다루기 때문에 검색 및 추천 시스템이 사용자의 질의어(혹은 선호도 정보)를 이해하는데 용이하다[4]. 기존 BBN 기반의 모델에서는 카탈로그 문서로부터 추출된 색인어들이 개념으로 이용되었지만, 본 연구에서는 ODP 웹 디렉토리의 각 카테고리들이 개념으로 이용한다. 비록 본 연구에서는 ODP를 도메인 지식 베이스로 이용하고 있으나, 개념간의 관계 정보 및 개념을 표현하기 위한 설명이 존재하는 지식베이스라면 아무런 제약 없이 적용이 가능하다. 개념 망을 이용함으로써, 사용자의 질의어와 카탈로그 문서간의 개념 관련도를 계산하는 것이 가능하다(시맨틱 검색). 또한 사용자들이 서로 다른 카탈로그 문서를 선호할지라도, 개념적으로 유사한 카탈로그 문서를 선호하는 사용자를 알아내는 것이 가능하다. 사용자간의 개념적 유사도를 계산함으로써, 시맨틱 기반의 협업적 필터링 추천이 가능하다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 질의어 확장 및 협업적 필터링을 위한 이전 연구들에 대하여 설명한다. 제 3장은 확장된 베이지안 신념 망 및 모델링 방법을 제시한다. 제 4장은 개념 망을 이용한 시맨틱 검색 및 추천 방법을 확률 식으로 제시한다. 제 5장은 논의 및 향후 과제에 대해 기술한다.

2. 관련 연구

2.1 질의어 확장

짧은 색인어로 구성된 질의어는 색인어의 모호성으로 인하여 사용자의 요구를 잘 표현하지 못한다. 예를 들어, <그림 1>에서 사용자가 "*apple*"라는 색인어로 구성된 질의를 한 경우, 기존 검색 시스템은 사용자가 어떤 도메인의 카탈로그 문서를 원하는지 알 수 없기 때문에 "*apple*"이라는 색인어를 포함하고 있는 카탈로그 문서 d_1과 d_2를 사용자에게 제공한다. "*apple*"이라는 키워드 자체는 "*computer company*"라는 의미 및 "*fruit*"이라는 이중의 의미(즉, 모호성)를 가지고 있으므로, 사용자의 요구에 적합하지 않은 문서들도 제공된다.



〈그림 1〉 색인어의 모호성 예시

이와 같은 질의어의 모호성을 제거하기 위해 질의어 확장과 관련된 연구들이 진행되었다[5~7]. 즉, 사용자 프로파일에서 추출된 색인어를 이용하여, 주어진 짧은 질의어를 확장한다. 사용자의 프로파일은 명시적으로 사용자가 지정할 수 있고, 암시적으로 과거 접근 로그(즉, 검색, 클릭 등)에서 추출된 색인어 혹은 로그와 관련된 카테고리 명을 이용할 수 도 있다. <그림 1>의 예시에서 사용자가 과거에 "*iPhone*" 관련 카탈로그 문서를 접근했

었다고 가정하면, 사용자의 질의어(“*apple*”)는 프로파일에서 추출된 색인어인 “*iPhone*”으로 확장된다. 확장된 질의어 “<*apple*, *iPhone*>”가 검색 엔진에 제공되고, 검색 엔진은 d_1을 사용자에게 제공하게 된다.

질의어 확장과 관련된 연구들은 질의어를 확장하기 위한 방법에 연구 초점을 맞추고 있기 때문에 기존 키워드 검색의 단점을 완전히 극복하지 못하였다. 즉, 확장된 질의어를 카탈로그 문서가 포함하고 있지 않은 경우, 여전히 검색이 불가능하다. 본 연구의 질의어 확장 방법은 이전 연구들과 비슷하지만, 질의어와 카탈로그 문서간의 색인어 매칭이 아닌, 개념 매칭을 이용한다. 비록 카탈로그 문서가 확장된 질의어를 포함하고 있지 않더라고, 사용자 질의에서 추출된 개념들 혹은 관련성이 높은 개념들을 포함하고 있다면 검색 결과로서 제공될 수 있다.

2.2 협업적 필터링 기반 추천

많은 상업적 사이트(특히, 온라인 쇼핑몰)에서 사용되는 추천 시스템의 목적은 사용자가 향후 구매할 가능성이 높은 카탈로그 문서를 사용자가 쉽게 찾을 수 있도록 도와주는 것이다. 많은 추천 알고리즘들 중 특히, 협업적 필터링(collaborative filtering; CF)기반 추천은 가장 성공한 기법들 중 하나이다[8, 9]. 이는 선호도가 비슷한 사용자들이 선호했던 상품(혹은 카탈로그 문서)을 엑티브 사용자(active user)에게 추천하는 방법(사용자 기반 CF) 및 이전에 구매한 상품과 비슷한 것을 엑티브 사용자에게 추천하는 방법(아이템 기반 CF)로 나뉜다[10, 11]. 아이템 기반 CF[12, 13]는 사용자 기반 CF[14~16]의 단점인 확장성 및 실시간 서비스 문제를 극복하고자 제안되었으나, [17]에 따르면 사용자 기반 CF가 아이템 기반 CF에 비하여 정확한 추천 결과를 제공하는 것으로 알려져 있다.

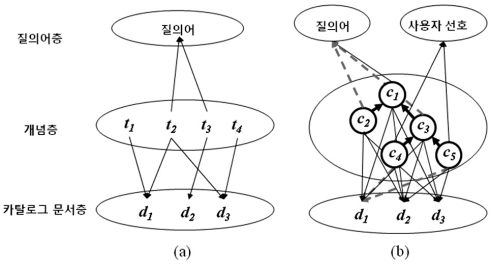
CF 기반 추천 알고리즘들이 많은 상업적 사이트(예 : *Amazon.com*, *Last.fm* 등) 적용되고 있음에도 불구하고 기존 CF 기반 추천 알고리즘은 치명적인 두 가지 단점이 있다 희박성 문제(sparsity problem), 콜드 스타트 문제(cold start problem). 희박성 문제는 데이터베이스에 상품(혹은 카탈로그 문서)의 수가 증가할수록 상품들에 대한 사용자의 로그 밀도가 감소하는 것을 의미한다[18]. 기존 CF 기반 추천 시스템들은 사용자간의 유사도를 계산하기 위해 이들이 과거에 접근했던 상품들간의 매칭 정보를 이용한다. 즉, 만약 사용자들이 다른 상품을 접근했다면, 해당 사용자들이 비슷한 선호도를 갖는다고 하더라도 다른 선호도를 갖는 사용자들로 간주된다. 콜드 스타트 문제는 새로 추가된 사용자 및 상품에 대한 충분한 정보가 존재하지 않을 경우, 추천이 어렵게 되는 것을 의미한다.

본 논문은 사용자의 선호도 및 카탈로그 문서(혹은 상품)들을 관련성이 높은 개념들로 표현함으로써, 희박성 문제를 해결하고자 한다. 특히, 사용자의 선호도를 개념 망으로 표현함으로써, 서로 다른 상품을 접근할지라도 개념적으로 유사한 선호도를 갖는 사용자들이라면 유사도를 계산하는 것이 가능하다.

3. 확장된 베이지안 신념 망

본 절에서는 ODP와 같은 분류 지식베이스

를 이용하여, 기존의 베이지안 신념 망(BBN) 기반 모델[2]을 확장한다. <그림 2>는 기존 BBN 기반 모델과 확장된 베이지안 신념 망(extended Bayesian belief network; EBBN) 기반 모델의 구조를 보여준다.



〈그림 2〉 (a) BBN과 (b) EBBN 기반 모델

BBN 기반 모델과 EBBN 기반 모델의 가장 큰 차이점은 개념층의 구조이다. BBN은 개념층이 카탈로그 문서로부터 추출된 색인어로 구성된 반면, EBBN의 개념층은 웹 디렉토리로부터 추출된 카테고리로 구성되어 있다. 카테고리를 이용함으로써(개념간의 관계 정보를 이용함으로써), 질의어층과 문서층 사이의 매칭 확률을 높이는 것이 가능하다. 예를 들어, <그림 2>(b)에서 사용자의 질의어가 개념 c_2, c_3로 표현되고, 카탈로그 문서 d_1이 개념 c_4, c_5로 표현된다고 가정한다. c_3가 c_4, c_5의 상위 개념이라는 것을 이용하면, 주어진 질의어와 카탈로그 문서간의 관련도를 계산하는 것이 가능하다. 하지만, 이와 같은 계층 관계 정보를 이용하지 않으면, 주어진 질의어와 카탈로그 문서간의 관련도 혹은 사용자간의 유사도를 계산하는 것이 불가능하다.

3.1 개념층 모델링

본 절은 웹디렉토리와 같은 분류 지식베이스에서 추출된 카테고리를 이용하여 EBBN 기반 모델의 개념층을 모델링하는 방법에 대하여 제시한다. 분류 지식 베이스의 각 개념(카테고리)들은 해당 개념으로 분류된 여러 개의 웹 페이지 제목 및 짧은 요약문을 가지고 있다. 이와 같은 제목 및 짧은 요약문을 본 연구에서는 서술(description)이라고 명한다. 본 연구는 각 개념에 분류된 서술들로부터 추출된 색인어를 이용하여 해당 개념의 시맨틱을 표현한다. 형식적으로 하나의 개념은 다음과 같이 모델링 된다. 본 논문에서 굵은 글씨체는 벡터를 의미한다.

<정의 1> 분류 지식 베이스가 n개의 개념으로 구성되었다고 가정하면, 분류 지식 베이스는 $C = \{c_1, \cdots, c_i, \cdots, c_n\}$와 같은 개념 집합이다. 개념간의 관계 정보를 이용하기 위해, 개념 c_i의 시맨틱은 c_i 및 c_i의 하위 개념에서 추출된 서술들의 평균 색인어 벡터(term vector)로 표현한다.

$$\boldsymbol{c}_i = \frac{1}{|D^{c_i}|} \sum_{\boldsymbol{d}_j^{c_i} \in D^{c_i}} \boldsymbol{d}_j^{c_i} \qquad (1)$$

$$where \ \ D^{c_i} = \{\boldsymbol{d}_1^{c_i}, \cdots, \boldsymbol{d}_j^{c_i}, \cdots, \boldsymbol{d}_m^{c_i}\} \cup \bigcup_{\forall c_s < c_i} D^{c_s}$$

식 (1)에서 $c_s < c_i$는 c_s가 i 번째 개념 c_i의 하위 개념임을 의미하고, D^{c_i}는 c_i와 c_i의 하위 개념들에 속하는 서술들의 집합을 의미한다.

$\boldsymbol{d}_j^{c_i}$는 개념 c_i에서 j번째 서술을 의미한다. 각 서술 $\boldsymbol{d}_j^{c_i}$는 가중된 색인어 벡터로 표현된다 : $\boldsymbol{d}_j^{c_i} = \langle w_i, \cdots, w_k, \cdots, w_{|V|} \rangle$, w_k는 k번째 색인어의 가중치, V는 색인어 집합을 의미한다. 각 색인어의 가중치 w_k는 TF(term frequency)와 IDF(inverse document frequency)

에 의해 계산한다. 색인어 가중치를 계산할 때, 서술들에서 불용어(관사, 전치사, 접속사 등)들을 제거하며, 포터스 알고리즘(Porter stemming algorithm)[19]을 이용하여 어미가 변형된 색인어들은 어근의 형태로 변형한다.

3.2 카탈로그 문서층 모델링

EBBN 기반 모델에서 카탈로그 문서들은 분류 지식 베이스의 개념들로 표현된다. 본 절은 카탈로그 문서를 개념들로 표현하는 방법에 대하여 제시한다. 카탈로그 문서를 분류 지식베이스에서 추출된 개념들로 표현하면 다음과 같다.

$$\boldsymbol{d_x} = \langle Pr(\boldsymbol{d_x} \mid \boldsymbol{c_1}), \cdots, Pr(\boldsymbol{d_x} \mid \boldsymbol{c_i}), \cdots, Pr(\boldsymbol{d_x} \mid \boldsymbol{c_n}) \rangle$$

여기서 확률 $Pr(\boldsymbol{d_x}|\boldsymbol{c_i})$는 개념 $\boldsymbol{c_i}$와 카탈로그 문서 $\boldsymbol{d_x}$간의 관련도를 나타낸다. 실제적으로, 카탈로그 문서 및 개념들은 여러 개의 색인어로 이루어져있으므로 카탈로그 문서와 색인어간의 관련도를 나타내는 확률 $Pr(t_k|\boldsymbol{d_x})$, 색인어와 개념간의 관련도를 나타내는 확률 $Pr(t_k|\boldsymbol{c_i})$을 이용하여 확률 $Pr(\boldsymbol{d_x}|\boldsymbol{c_i})$를 정의한다. 조건부 확률의 정의와 전확률 법칙에 의하여 확률 $Pr(\boldsymbol{d_x}|\boldsymbol{c_i})$는 다음과 같이 정의된다.

$$Pr(\boldsymbol{d_x} \mid \boldsymbol{c_i}) =$$

$$\frac{1}{|Pr(\boldsymbol{c_i})|} \cdot \sum_{t_k} Pr(\boldsymbol{d_x}|t_k) \cdot Pr(\boldsymbol{c_i}|t_k) \cdot Pr(t_k)$$

베이지안 이론을 위의 수식에 적용하면, 다음과 같다.

$$\frac{1}{|Pr(c_i)|} \cdot \sum_{t_k} \frac{Pr(\boldsymbol{d}_x) Pr(t_k|\boldsymbol{d}_x)}{Pr(t_k)} \cdot \frac{Pr(\boldsymbol{c}_i) Pr(t_k|\boldsymbol{c}_x)}{Pr(t_k)}$$

$$\cdot Pr(t_k) = \sum_{t_k} \frac{Pr(\boldsymbol{d}_x)}{Pr(t_k)} \cdot Pr(t_k|\boldsymbol{d}_x) \cdot Pr(t_k|\boldsymbol{c}_x) \quad (2)$$

확률 $Pr(\boldsymbol{c_i})$, $Pr(\boldsymbol{d_x})$, $Pr(t_k)$는 개념, 카탈로그 문서, 색인어에 대한 사전 확률(prior probability)이다. 계산 과정을 단순화 시키기 위해, 확률 $Pr(\boldsymbol{c_i})$가 모든 개념들에 대하여 동일하다고 가정한다. 즉, $Pr(\boldsymbol{c_i})$ = 1/총 개념들의 수. 유사하게 각각의 확률 변수에 대하여 동일하다고 가정하면 식 (2)는 다음과 같이 비례식으로 표현된다.

$$Pr(\boldsymbol{d}_x \mid \boldsymbol{c}_i) \propto \sum_{t_k} Pr(t_k|\boldsymbol{d}_x) \cdot Pr(t_k|\boldsymbol{c}_x) \quad (3)$$

식 (3)에서 확률 $Pr(t_k|\boldsymbol{d_x})$은 색인어와 카탈로그 문서 사이의 관련도를 표현한다. 카탈로그 문서의 속성들을 고려하지 않을 경우, 하나의 카탈로그 문서는 V-차원의 색인어 벡터로 표현이 된다(V는 색인어의 전체 집합). 확률 $Pr(t_k|\boldsymbol{d_x})$은 카탈로그 문서에 대한 색인어 벡터에서 색인어 t_k의 가중치를 의미하며 다음과 같이 정의된다.

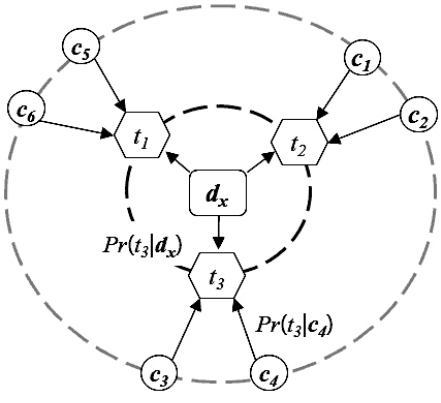
$$Pr(t_k \mid \boldsymbol{d}_x) = \frac{w(\boldsymbol{d}_x, t_k)}{\sum_{t_j} w(\boldsymbol{d}_x, t_j)} \quad (4)$$

식 (4)에서 분자 $w(\boldsymbol{d}_x, t_k)$는 문서 $\boldsymbol{d_x}$에서 k번째 색인어의 가중치를 의미하며, 분모 $\sum_{t_j} w(\boldsymbol{d}_x, t_j)$는 $\boldsymbol{d_x}$에서 모든 색인어의 가중치 합을 의미한다.

카탈로그 문서와 유사하게 분류 지식베이스의 각 개념은 V-차원의 색인어 벡터로 표현이 된다. 확률 $Pr(t_k|\boldsymbol{c_i})$은 개념에 대한 색인어 벡터에서 색인어 t_k의 가중치를 의미하며 다음과 같이 정의된다.

$$Pr(t_k \mid \boldsymbol{c_i}) = \frac{w(\boldsymbol{c_i}, t_k)}{\sum_{t_j} w(\boldsymbol{c_i}, t_j)} \quad (5)$$

식 (5)에서 분자 $w(\boldsymbol{c_i}, t_k)$는 개념 $\boldsymbol{c_i}$에서 k 번째 색인어의 가중치를 의미하며, 분모 $\sum_{t_j} w(\boldsymbol{c_i}, t_j)$는 $\boldsymbol{c_i}$에서 모든 색인어의 가중치 합을 의미한다. 다음 예시는 확률 $Pr(\boldsymbol{d_x}|\boldsymbol{c_4})$ 구하는 과정을 설명한다.



〈그림 3〉 확률 $Pr(d_x|c_4)$의 예시

본 예시에서 카탈로그 문서 $\boldsymbol{d_x}$는 색인어 t_1, t_2, t_3로 구성되어 있으며, 개념 $\boldsymbol{c_4}$는 색인어 t_3로 구성되어 있다고 가정한다. 카탈로그 문서 $\boldsymbol{d_x}$와 색인어 t_3사이의 관련도는 확률 $Pr(t_3|\boldsymbol{d_x})$로 표현되며, 색인어 t_3와 개념 $\boldsymbol{c_4}$사이의 관련도는 확률 $Pr(t_3|\boldsymbol{c_4})$로 표현된다. 두 확률을 곱함으로써, 카탈로그 문서 $\boldsymbol{d_x}$와 개념 $\boldsymbol{c_4}$사이의 관련도를 계산하는 것이 가능하다. 식 (3)과 같이 모든 색인어 (t_1, t_2, t_3)에 대하여 위의 과정을 반복하면, 카탈로그 문서 $\boldsymbol{d_x}$는 개념들로 다음과 같이 표현된다.

$$\boldsymbol{d_x} = <Pr(\boldsymbol{d_x} \mid \boldsymbol{c_1}), \cdots, Pr(\boldsymbol{d_x} \mid \boldsymbol{c_6})>$$

각 개념들은 상하위 관계를 포함하고 있으므로, 카탈로그 문서 $\boldsymbol{d_x}$는 개념 망으로 표현된다.

3.3 질의어층 모델링

EBBN 기반 모델에서 질의어층은 사용자의 질의어와 사용자 선호도 정보로 구성되어 있다. 본 절은 질의어와 사용자 선호도 정보를 개념들로 표현하는 방법에 대하여 제시한다.

일반적으로 짧은 색인어로 구성된 질의어는 카탈로그 문서층과 비슷한 방법으로 모델링이 가능하다. 질의어도 카탈로그 문서들과 마찬가지로 짧은 색인어로 구성되어 있기 때문이다. 즉, 사용자의 질의어 $\boldsymbol{q}$를 하나의 카탈로그 문서로 간주하면, 식 (3)을 이용하여 다음과 같이 표현할 수 있다.

$$Pr(\boldsymbol{q} \mid \boldsymbol{c_i}) \propto \sum_{t_k} Pr(t_k \mid \boldsymbol{q}) \cdot Pr(t_k \mid \boldsymbol{c_x}) \quad (6)$$

식 (6)을 이용하여, 질의어 $\boldsymbol{q}$와 개념 $\boldsymbol{c_i}$간의 관련도를 계산하는 것이 가능하며, 질의어도 n개의 개념들로 표현된다. 즉,

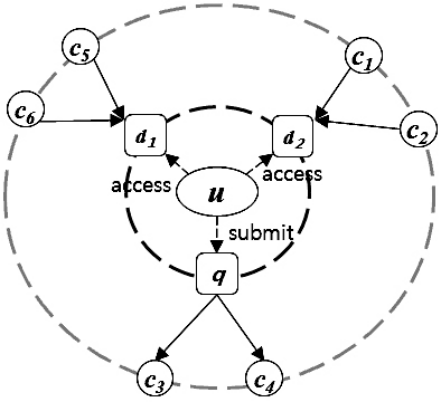
$$\boldsymbol{q} = <Pr(\boldsymbol{q} \mid \boldsymbol{c_1}), \cdots, Pr(\boldsymbol{q} \mid \boldsymbol{c_i}), \cdots, Pr(\boldsymbol{q} \mid \boldsymbol{c_n})>$$

각 개념들은 상하위 관계를 포함하고 있으므로, 질의어 $\boldsymbol{q}$는 개념 망으로 표현된다.

사용자 선호 $\boldsymbol{p_u}$ 역시, 개념 망으로 표현이 가능하다. 즉,

$$\boldsymbol{p_u} = <Pr(\boldsymbol{p_u} \mid \boldsymbol{c_1}), \cdots, Pr(\boldsymbol{p_u} \mid \boldsymbol{c_i}), \cdots, Pr(\boldsymbol{p_u} \mid \boldsymbol{c_n})>$$

확률 $Pr(\boldsymbol{p_u}|\boldsymbol{c_i})$는 사용자의 선호 $\boldsymbol{p_u}$와 개념 $\boldsymbol{c_i}$간의 관련도를 나타낸다. 사용자의 선호도 정보는 사용자의 로그에서 추출된 카탈로그 문서로 표현될 수 있다. 식 (3) 및 식 (6)처럼 조건부 확률의 정의 및 전확률 법칙을



〈그림 4〉 사용자 선호도 및 질의어의 개념 매핑 예시

적용하면 다음과 같이 비례식으로 표현된다.

$$Pr(\boldsymbol{p_u}|\boldsymbol{c_i}) \propto \sum_{\boldsymbol{d_x} \in D} Pr(\boldsymbol{d_x}|\boldsymbol{p_u}) \cdot Pr(\boldsymbol{d_x}|\boldsymbol{c_i}) \quad (7)$$

확률 $Pr(\boldsymbol{d_x}|\boldsymbol{p_u})$은 사용자의 선호도가 주어졌을 때, 카탈로그 문서 $\boldsymbol{d_x}$를 선호하는 정도를 나타내며, 다음과 같이 계산된다.

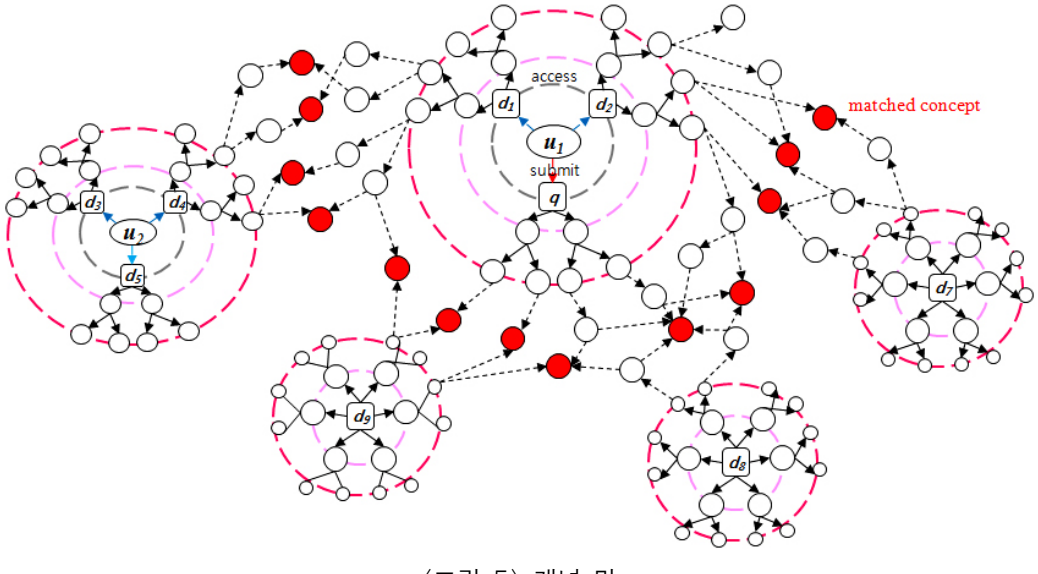
$$Pr(\boldsymbol{d_x}|\boldsymbol{p_u}) = \frac{access(u, \boldsymbol{d_x})}{\sum_{\boldsymbol{d_x}} access(u, \boldsymbol{d_x})} \quad (8)$$

식 (8)에서 분자 $access(u, \boldsymbol{d_x})$는 카탈로그 문서 $\boldsymbol{d_x}$에 대한 사용자 u의 접근 횟수이며, 분모 $\sum_{\boldsymbol{d_x}} access(u, \boldsymbol{d_x})$는 모든 카탈로그 문서에 대한 사용자 u의 접근 횟수이다.

<그림 4>는 사용자 u의 선호도 정보 및 질의어를 개념들로 표현한 예시이다. 본 예시에서 사용자가 카탈로그 문서 $\boldsymbol{d_1}$과 $\boldsymbol{d_2}$를 접근했었다면, 사용자 선호는 식 (7)에 의해 개념 $\boldsymbol{c_1}$, $\boldsymbol{c_2}$, $\boldsymbol{c_5}$, $\boldsymbol{c_6}$로 표현된다. 또한 사용자의 질의어 $\boldsymbol{q}$는 $\boldsymbol{c_3}$, $\boldsymbol{c_4}$로 표현된다.

4. 개념 망 및 그 활용

현재까지 많은 검색 및 추천 방법들이 제안되었지만, 이전 연구들은 검색 혹은 추천의



〈그림 5〉 개념 망

한 측면에 초점을 맞춘 정보 추출(information retrieval) 모델이었다. 그러나 본 연구는 개념 망을 이용하여 시맨틱 검색 및 추천을 위한 통합된 정보 추출 모델을 제안한다. 본 연구는 사용자의 질의어, 선호도 정보, 카탈로그 문서를 모두 분류 지식 베이스(웹 디렉토리)의 개념(카테고리)으로 표현하고 있다. 분류 지식 베이스는 개념간의 상·하위 관계를 가지고 있으므로, 이를 이용하면 <그림 5>와 같이 개념 망을 구축할 수 있다.

사용자 (u_1)의 질의어 $\boldsymbol{q}$ 및 카탈로그 문서들 ($\boldsymbol{d_7}$, $\boldsymbol{d_8}$, $\boldsymbol{d_9}$)간의 개념 매칭을 이용하여 시맨틱 관련도를 계산한다. 이 경우, 질의어와 카탈로그 문서간의 색인어가 일치하지 않아도 시맨틱 관련도(relevance degree)를 계산하는 것이 가능하다.

또한 사용자 u_1와 u_2의 로그에서 추출된 카탈로그 문서와 매핑된 개념을 각 사용자 선호의 시맨틱으로 가정하면, 두 사용자간의 시맨틱 유사도(similarity degree)를 계산하는 것이 가능하다. 비록 두 사용자가 다른 카탈로그 문서를 접근하였어도, 사용자간의 시맨틱 유사도를 계산할 수 있다. 시맨틱 유사도를 이용하여 시맨틱 기반의 협업적 필터링 추천이 가능하다.

4.1 시맨틱 검색

본 절에서는 사용자 질의어 $\boldsymbol{q}$와 카탈로그 문서간의 확률적으로 개념 관련도를 계산하는 방법을 제시한다.

Wong[20]은 카탈로그 문서의 랭크를 구하기 위한 두 가지 확률적 방법을 제안하였다 : $Pr(\boldsymbol{q}|\boldsymbol{d_x})$와 $Pr(\boldsymbol{d_x}|\boldsymbol{q}) \cdot Pr(\boldsymbol{q}|\boldsymbol{d_x})$은 q에 대한 $\boldsymbol{d_x}$의 정확도를 의미하며, $Pr(\boldsymbol{d_x}|\boldsymbol{q})$은 q에 대한 $\boldsymbol{d_x}$의 재현율을 의미한다. 그러나 Wong은 적절한 정규화 과정을 거치면 $Pr(\boldsymbol{q}|\boldsymbol{d_x})$와 $Pr(\boldsymbol{d_x}|\boldsymbol{q})$은 동일한 값을 갖을 수 있다고 증명하였다. 그러므로 어떤 조건부 확률식을 사용하는지는 중요하지 않다. 본 논문은 질의어 $\boldsymbol{q}$가 조건으로 주어지는 것이 좀 더 자연스러우므로, $Pr(\boldsymbol{d_x}|\boldsymbol{q})$을 랭크 함수로 이용한다.

조건부 확률의 정의 및 전확률 법칙에 의하여 시맨틱 검색의 랭크 함수는 다음과 같다.

$$Pr(\boldsymbol{d_x} \mid \boldsymbol{q}) = \frac{Pr(\boldsymbol{d}_x \cap \boldsymbol{q})}{Pr(\boldsymbol{q})}$$

$$= \frac{1}{Pr(\boldsymbol{q})} \cdot \sum_{\forall \boldsymbol{c_i}} Pr(\boldsymbol{q} \cap \boldsymbol{d_x} \mid \boldsymbol{c_i}) \cdot Pr(\boldsymbol{c_i}) \qquad (9)$$

<그림 2>(b)에서와 같이 질의어층과 카탈로그 문서층을 사이에 개념층을 위치시킴으로써, 질의어층과 카탈로그 문서층은 논리적으로 상호 독립적인 관계를 갖게 된다. 그러므로, 식 (9)는 다음과 같이 변경한다.

$$\frac{1}{Pr(\boldsymbol{q})} \cdot \sum_{\forall \boldsymbol{c_i}} Pr(\boldsymbol{q} \mid \boldsymbol{c_i}) \cdot Pr(\boldsymbol{d_x} \mid \boldsymbol{c_i}) \cdot Pr(\boldsymbol{c_i})$$

사전 확률 $Pr(\boldsymbol{q})$와 $Pr(\boldsymbol{c_i})$가 각각 확률 변수 $\boldsymbol{q}$와 $\boldsymbol{c_i}$에 대하여 동일한 값을 갖는다고 가정한다면 시맨틱 검색을 위한 랭크 함수는 다음과 같다.

$$Pr(\boldsymbol{d}_x \mid \boldsymbol{q}) \propto \sum_{\forall \boldsymbol{c_i}} Pr(\boldsymbol{q} \mid \boldsymbol{c_i}) \cdot Pr(\boldsymbol{d_x} \mid \boldsymbol{c_i}) \qquad (10)$$

4.2 시맨틱 기반 협업적 필터링 추천

본 절은 사용자의 선호도 정보가 주어졌을 때, 선호도가 비슷한 사용자를 찾는 방법 및

비슷한 사용자가 접근했던 문서를 엑티브 사용자에게 추천하는 방법을 제시한다. 조건부 확률 $Pr(\boldsymbol{d}_{u'x}|\boldsymbol{p}_u)$는 사용자 u의 선호도 정보($\boldsymbol{p}_u$)가 조건으로 주어졌을 때, u'이 접근했던 $\boldsymbol{d}_{u'x}$가 얼마나 사용자 u의 선호도에 적합한 문서인가를 확률적으로 나타낸다. 여기서, u는 엑티브 사용자이며, u'은 u와 선호도가 비슷한 사용자이다. $\boldsymbol{d}_{u'x}$는 오직 u'만 접근한 x번째 카탈로그 문서이다.

조건부 확률의 정의 및 전확률 법칙을 적용하면 확률 $Pr(\boldsymbol{d}_{u'x}|\boldsymbol{p}_u)$는 다음과 같다.

$$Pr(\boldsymbol{d}_{u'x}\,|\,\boldsymbol{p}_u)=\frac{Pr(\boldsymbol{d}_{u'x}\cap\boldsymbol{p}_u)}{Pr(\boldsymbol{p}_u)}$$

$$=\frac{1}{Pr(\boldsymbol{p}_u)}\cdot\sum_{u'\in U}Pr(\boldsymbol{d}_{u'x}\cap\boldsymbol{p}_u\,|\,\boldsymbol{p}_{u'})\cdot Pr(\boldsymbol{p}_{u'})$$

$$=\frac{1}{Pr(\boldsymbol{p}_u)}\cdot\sum_{u'\in U}Pr(\boldsymbol{d}_{u'x}\,|\,\boldsymbol{p}_{u'})\cdot Pr(\boldsymbol{p}_u\,|\,\boldsymbol{p}_{u'})$$

$$\cdot Pr(\boldsymbol{p}_{u'})\propto\sum_{u'\in U}Pr(\boldsymbol{d}_{u'x}\,|\,\boldsymbol{p}_{u'})\cdot Pr(\boldsymbol{p}_u\,|\,\boldsymbol{p}_{u'})\quad(11)$$

식 (11)에서 U는 비슷한 사용자 집합을 의미하며, 계산 과정을 단순화하기 위해 사전 확률 $Pr(\boldsymbol{p}_u)$와 $Pr(\boldsymbol{p}_{u'})$는 확률 변수 $\boldsymbol{p}_u$와 $\boldsymbol{p}_{u'}$에 대하여 동일하다고 가정한다. 확률 $Pr(\boldsymbol{d}_{u'x}|\boldsymbol{p}_{u'})$은 사용자 u'이 카탈로그 문서가 $\boldsymbol{d}_{u'x}$에 접근할 확률을 의미하며 접근로그를 이용하여 다음과 같이 계산된다.

$$Pr(\boldsymbol{d}_{u'x}\,|\,\boldsymbol{p}_{u'})=\frac{access(\boldsymbol{u'},\,\boldsymbol{d}_{u'x})}{\sum_{\boldsymbol{d}_{u'x}}access(\boldsymbol{u'},\,\boldsymbol{d}_{u'x})}\quad(12)$$

식 (12)에서 분자는 사용자 u'의 카탈로그 문서 $\boldsymbol{d}_{u'x}$에 대한 접근 횟수를 의미하며, 분모는 사용자 u'의 카탈로그 문서 전체에 대하여 총 접근 횟수를 의미한다.

더욱이, 식 (12)에서 확률 $Pr(\boldsymbol{p}_u|\boldsymbol{p}_{u'})$는 사용자간의 확률적 유사도를 의미한다. 본 연구에서는 사용자의 선호도 정보를 개념 망으로 표현하고 있으므로, 사용자간의 개념적 유사도는 다음과 같이 정의된다.

$$Pr(\boldsymbol{p}_u\,|\,\boldsymbol{p}_{u'})=\frac{Pr(\boldsymbol{p}_u\cap\boldsymbol{p}_{u'})}{Pr(\boldsymbol{p}_{u'})}$$

$$=\frac{1}{Pr(\boldsymbol{p}_{u'})}\cdot\sum_{\boldsymbol{c}_i\in\boldsymbol{C}}Pr(\boldsymbol{p}_u\,|\,\boldsymbol{c}_i)\cdot Pr(\boldsymbol{p}_{u'}\,|\,\boldsymbol{c}_i)$$

$$\cdot Pr(\boldsymbol{c}_i)\propto\sum_{\boldsymbol{c}_i\in\boldsymbol{C}}Pr(\boldsymbol{p}_u\,|\,\boldsymbol{c}_i)\cdot Pr(\boldsymbol{p}_{u'}\,|\,\boldsymbol{c}_i)\quad(13)$$

식 (13)에서 확률 $Pr(\boldsymbol{p}_u|\boldsymbol{c}_i)$는 사용자의 선호 $\boldsymbol{p}_u$와 개념 $\boldsymbol{c}_i$간의 관련도를 나타내며, 이는 식 (7)에서 이미 설명을 했다. 식 (7)에서 $\boldsymbol{d}_x$는 사용자 u가 접근했던 카탈로그 문서를 의미한다.

식 (11)에 식 (3), 식 (7), 식 (13)를 적용하면 카탈로그 문서 $\boldsymbol{d}_{u'x}$를 사용자 u에게 추천하기 위한 확률 $Pr(\boldsymbol{d}_{u'x}|\boldsymbol{p}_u)$은 다음과 같다.

$$Pr(\boldsymbol{d}_{u'x}\,|\,\boldsymbol{p}_u)=\sum_{u'\in U}Pr(\boldsymbol{d}_{u'x}\,|\,\boldsymbol{p}_{u'})\cdot Pr(\boldsymbol{p}_u\,|\,\boldsymbol{p}_{u'})$$

$$=\sum_{u'\in U}Pr(\boldsymbol{u}_{u'x}\,|\,\boldsymbol{p}_{u'})$$

$$\times\left[\sum_{\boldsymbol{c}_i\in\boldsymbol{C}}\left\{\sum_{\boldsymbol{d}_{ux}\in\boldsymbol{D}}Pr(\boldsymbol{d}_{ux}\,|\,\boldsymbol{p}_u)\cdot Pr(\boldsymbol{d}_{ux}\,|\,\boldsymbol{c}_i)\right\}\right.$$

$$\left.\times\left\{\sum_{\boldsymbol{d}_{u'x}\in\boldsymbol{D}}Pr(\boldsymbol{d}_{u'x}\,|\,\boldsymbol{p}_{u'})\cdot Pr(\boldsymbol{d}_{u'x}\,|\,\boldsymbol{c}_i)\right\}\right]$$

$$=\sum_{u'\in U}Pr(\boldsymbol{d}_{u'x}\,|\,\boldsymbol{p}_{u'})\times$$

$$\left[\sum_{\boldsymbol{c}_i\in\boldsymbol{C}}\left\{\sum_{\boldsymbol{d}_{ux}\in\boldsymbol{D}}Pr(\boldsymbol{d}_{ux}\,|\,\boldsymbol{p}_u)\cdot\right.\right.$$

$$\left(\sum_{t_k} Pr(t_k \mid \boldsymbol{d}_{ux}) \cdot Pr(t_k \mid \boldsymbol{c}_i)\right)\Bigg\}$$

$$\times \Bigg\{ \sum_{\boldsymbol{d}_{u'x} \in \boldsymbol{D}} Pr(\boldsymbol{d}_{u'x} \mid \boldsymbol{p}_{u'}) \cdot$$

$$\left(\sum_{t_k} Pr(t_k \mid \boldsymbol{d}_{u'x}) \cdot Pr(t_k \mid \boldsymbol{c}_i)\right)\Bigg\}\Bigg] \qquad (14)$$

식 (14)는 시맨틱 기반 협업적 필터링 추천을 위해 확장된 베이지안 신념 망을 이용한 추천 모델이다. 기본적으로 식 (14)는 확률 $Pr(\boldsymbol{d}|\boldsymbol{p})$, $Pr(t|\boldsymbol{d})$, $Pr(t|\boldsymbol{c})$로 구성되어 있으며, 이러한 확률들은 학습 데이터로부터 미리 계산될 수 있는 값들이다.

5. 논의 및 향후 과제

본 연구는 사용자의 질의어, 선호도, 카탈로그 문서의 시맨틱을 도출하기 위해 분류 지식베이스로부터 추출한 개념을 이용하였다. 도출된 개념을 이용하여, 사용자의 질의어 및 카탈로그 문서의 간의 색인어 불일치 문제를 해결한 시맨틱 검색 모델을 제시하였다. 또한 사용자의 선호도 정보 역시 개념으로 표현함으로써, 협업적 필터링 기반 추천 알고리즘의 치명적인 단점인 희박성 문제를 해결하였다. 특히, 이전의 정보 추출 연구들이 검색 혹은 추천의 한 측면에 초점을 맞춘 모델을 제시한 반면, 본 연구는 검색 및 추천을 시맨틱 공간에서 수행할 수 있는 정보 추출 모델을 제시하였다.

향후 과제로는 본 연구의 우수성을 실험적으로 증명해야 한다. 실험을 위해 고려해야 할 사항을 정리하면 다음과 같다.

시맨틱 검색 및 추천 시스템의 성능을 평가하기 위해서는 고려해야 하는 사항은 검색 결과가 사용자의 요구에 얼마나 적합한지(혹은 관련성이 있는지)를 판단할 기준이 있어야 한다. 일반적으로 사용자 요구와 검색 결과 사이의 관련성을 판단하기 위한 방법은 두 가지로 분류된다 : 사용자 기반 판단 방법[4, 21, 22]과 로그 기반 판단 방법[7, 23, 24].

사용자 기반 판단 방법은 많은 사용자들이 검색 결과에 대한 직접적인 판단을 수행한다. 이 방법은 실험을 수행하는데 있어서 많은 비용을 요구한다. 더욱이, 실험에 참가하는 사용자들은 자신들이 테스트를 받고 있다는 것을 알고 있기 때문에 실험 결과가 바이어스(bias) 될 수 있다[25].

로그 기반 판단 방법은 사용자 기반 판단 방법에 비해 비용 및 바이어스를 줄일 수 있다. 하지만, 이 방법은 많은 양의 사용자 질의 및 접근 로그를 필요로 한다. 실제적으로 구글 및 야후! 같은 검색 엔진들은 사용자 정보 보호를 위해 이와 같은 로그를 배포하지 않는다. 그러므로 우리를 포함한 대부분의 연구자들이 질의 및 접근 로그를 수집하는 것도 쉽지 않다.

그러므로 본 연구의 우수성을 판단하기 위해서는 사용자 기반 판단 및 로그 기반 판단 방법 이외의 다른 방법을 제안해야 하며, 향후 과제로 남겨둔다.

참 고 문 헌

[1] Baeza-Yates, R. and Ribeiro-Neto, B., Modern Information Retrieval, Addison

Wesley, 1999.

[2] Ribeiro, B. A. and Muntz, R., “A Belief Network Model for IR,” In Proceeding of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR ’96), 1996, pp. 253-260.

[3] Lenat, D. and Guha, R., Building Large Knowledge Based Systems, Addison Wesley, 1990.

[4] Chirita, P. A., Nejdl, W., Paiu, R., and Kohlschutter, C., “Using ODP Metadata to Personalized Search,” In Proceeding of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR ’05), 2005, pp. 178-185.

[5] Vogel, D., Bickel, S., Haider, P., Schimpfky, R., Siemen, P., Bridges, S., and Scheffer, T., “Classifying Search Engine Queries using the Web as Background Knowledge,” SIGKDD Explorations Newsletter, Vol. 7, No. 2, 2005, pp. 117-122.

[6] Shen, D., Sun, J., Yang, Q., and Chen, Z., “Building Bridges for Web Query Classification,” In Proceeding of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR ’06), 2006, pp. 131-138.

[7] Liu, F., Yu, C., and Meng, W., “Personalized Web Search for Improving Retrieval Effectiveness,” IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 1, 2004, pp. 28-40.

[8] Cosley, D., Lawrence, S., and Pennock, D., “REFEREE : An Open Framework for Practical Testing of Recommender Systems Using Researchindex,” In Proceeding of 28th International Conference on Very Large Databases(VLDB 2002), 2002, pp. 35-46.

[9] Pazzani, M. and Bilsus, D., “Learning and Revising User Profile : the Identification of Interesting Web Sites,” Machine Learning, Vol. 27, No. 3, 1997, pp. 313-331.

[10] Li, Y., Lu, L., and Xuefeng, L., “A Hybrid Collaborative Filtering Method for Multiple Interests and Multiple Content Recommendation in E-Commerce,” Expert Systems with Applications, Vol. 28, No. 1, 2005, pp. 67-77.

[11] Wang, J., Vries, A. P., and Reinders, M. J., “Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion,” In Proceeding of 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2006), 2006, pp. 501-508.

[12] Deshpande, M. and Karypis, G., “Item-based Top-n Recommendation Algorithm,” ACM Transaction on Information Systems, Vol. 22, No. 1, 2004, pp. 143-177.

[13] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J., “Item-based Collaborative Filtering Recommendation Algorithms,”

In Proceeding of 10th International Conference on World Wide Web(WWW 2001), 2001, pp. 285-295.

[14] Herlocker, J. L., Konstan, J. A., Borchers, A., and Riedl, J., "An Algorithmic Framework for Performing Collaborative Filtering," In Proceeding of 22th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR 1999), 1999, pp. 230-237.

[15] Jin, R., Chai, J. Y., and Si, L., "An Automatic Weighting Scheme for Collaborative Filtering," In Proceeding of 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR 2004), 2004, pp. 337-344.

[16] Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P., and Riedel, J., "GroupLens : An Open Architecture for Collaborative Filtering of Netnews," In Proceeding of the ACM Conference on Computer Supported Cooperative Work (CSCW 1994), 1994, pp. 175-186.

[17] Mild, A. and Natter, M., "A Critical View on Recommendation Systems," Working Paper Series, 2001, pp. 1-16.

[18] Mobasher, B., Jin, X., and Zhou, Y., "Semantically Enhanced Collaborative Filtering on the Web," Lecture Notes in Computer Science, Vol. 3209, 2004, pp. 57-76.

[19] Porter, M. F., "An Algorithm for Suffix Stripping," Readings in Information Retrieval, 1997, pp. 313-316.

[20] Wong, S. and Yao, Y., "On Modeling Information Retrieval with Probabilistic Inference," ACM Transactions on Information Systems, Vol. 13, No. 1, 1995, pp. 38-68.

[21] Pitkow, J., Schutze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., Adar, E., and Breuel, T., "Personalized Search," Communication of the ACM, Vol. 45, No. 9, 2002, pp. 50-55.

[22] Pretschner, A. and Gauch, S., "Ontology-based Personalized Search," In Proceeding of the 11th IEEE International Conference on Tools with Artificial Intelligence, 1999, pp. 391-398.

[23] Speretta, M. and Gauch, S., "Personalized Search based on User Search Histories," In Proceeding of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, 2005, pp. 622-628.

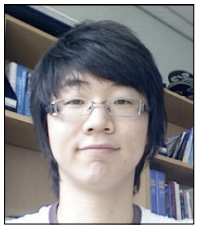
[24] Sieg, A., Mobasher, B., and Burke, R., "Web Search Personalization with Ontological User Profiles," In Proceeding of the 16th ACM Conference on Information and Knowledge Management (CIKM '07), 2007, pp. 525-534.

[25] Xu, S., Bao, S., Fei, B., Su, Z., and Yu, Y., "Exploring Folksonomy for Personalized Search," In Proceeding of the 31th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR '08), 2008, pp. 155-162.

저 자 소 개

이재원 (E-mail : jwlee@europa.snu.ac.kr)
2003년 숭실대학교 정보통신전자 공학부 졸업 (학사)
2004년~현재 서울대학교 컴퓨터공학부 대학원 석·박사통합 과정
관심분야 시맨틱 검색, 개인화 검색, 데이터베이스

박성찬 (E-mail : baksalchan@europa.snu.ac.kr)
2005년 서울대학교 컴퓨터공학부 졸업 (학사)
2007년 서울대학교 컴퓨터공학부 대학원 졸업 (석사)
2007년~현재 서울대학교 컴퓨터공학부 대학원 박사과정
관심분야 데이터베이스, 데이터마이닝



이상근 (E-mail : liza183@europa.snu.ac.kr)
2005년 한국과학기술원 전산학과 졸업 (학사)
2005년~현재 서울대학교 컴퓨터공학부 대학원 석·박사통합 과정
관심분야 상황인지, 추천 시스템, 시맨틱 검색, 데이터베이스

박재휘 (E-mail : jaehui@europa.snu.ac.kr)
2005년 한국과학기술원 전산학과 졸업 (학사)
2006년~현재 서울대학교 컴퓨터공학부 대학원 석·박사통합 과정
관심분야 데이터베이스, 정보검색

김한준 (E-mail : khj@uos.ac.kr)
1994년 서울대학교 계산통계학과 졸업 (학사)
1996년 서울대학교 전산과학과 대학원 졸업 (석사)
2002년 서울대학교 컴퓨터공학부 대학원 졸업 (박사)
2002년~2002년 서울대학교 공과대학 Post-Doc
2002년~현재 서울시립대학교 전자전기컴퓨터공학부 부교수
관심분야 텍스트마이닝, e-비즈니스 기술, 정보검색, 데이터베이스

이상구 (E-mail : sglee@europa.snu.ac.kr)
1985년 서울대학교 계산통계학과 졸업 (학사)
1987년 Northwestern University, Computer Science (석사)
1990년 Northwestern University, Computer Science (석사)
1992년~현재 서울대학교 컴퓨터공학부 교수
관심분야 e-비즈니스 기술, 시맨틱 웹, 온톨로지, 데이터베이스