

개인화된 뉴스 서비스를 위한 소셜 네트워크 기반의 콘텐츠 추천기법

홍명덕

인하대학교 컴퓨터정보공학과
(hmdgo@eslab.inha.ac.kr)

가명현

인하대학교 컴퓨터정보공학과
(gagaman7777@eslab.inha.ac.kr)

오경진

인하대학교 컴퓨터정보공학과
(okjillo@eslab.inha.ac.kr)

조근식

인하대학교 컴퓨터정보공학과
(gsjo@inha.ac.kr)

.....

세계에는 수많은 사람들이 살아가고 있고, 사람들의 일상으로부터 매일, 매 시간 단위로 새로운 뉴스가 발생한다. 발생하는 뉴스는 예정된 일과 예상하지 못한 일들을 포함하고 있다. 발생하는 뉴스의 거대한 양과 이를 전달하는 수많은 미디어들로 인해 사람들은 뉴스 콘텐츠를 이용하는데 많은 시간을 소비하게 된다. 하지만 미디어에 시시각각 나타나는 속보와 실시간 이슈의 대부분이 가십 기사로 이루어져 있어 사용자가 자신의 성향에 맞는 뉴스를 선별하고, 뉴스로부터 정보를 획득하는 것은 쉽지 않은 일이다. 또한 사용자의 관심사가 시간에 따라 변하기 때문에 뉴스 제공에 있어 사용자의 변하는 관심사를 반영하는 것이 요구된다. 본 논문에서는 사용자의 최근 관심사를 기반으로 사용자 선호도에 맞는 뉴스를 제공하기 위한 콘텐츠 기반의 추천 기법 및 시스템을 제안한다. 사용자의 최근 선호도를 파악하기 위하여 소셜 네트워크 서비스인 Facebook 사용자의 정보와 최근 게시글을 이용하여 동적으로 사용자 프로파일을 생성하여 이를 뉴스 서비스에 활용하고, 사용자 선호도에 적합한 뉴스를 추출하기 위해서 뉴스 콘텐츠의 분석을 요구한다. 뉴스 콘텐츠 분석을 위해 미디어에서 제공되는 뉴스의 카테고리를 사용하고, 뉴스 방송원고의 분석 및 주요 키워드 추출을 통해 뉴스 프로파일을 생성한다. 사용자 프로파일과 뉴스 프로파일 간의 유사도 측정을 위해서는 두 프로파일 간 형식의 일치화가 요구되므로 사용자 프로파일을 뉴스 프로파일과 동일한 형태로 생성한다. 사용자가 시스템에 접속하면 시스템은 사용자 프로파일에 명시된 선호도를 기반으로 뉴스 프로파일과의 유사도를 측정하고, 사용자 선호도에 가장 적합한 뉴스들을 제공하게 된다. 또한 사용자에게 제공된 뉴스 프로파일과 다른 뉴스 프로파일들 간에 유사도를 측정하여 유사도가 높은 관련된 뉴스들을 제공하게 된다. 제안한 개인화된 뉴스 서비스의 성능을 평가하기 위해 사용자에게 추천된 뉴스에 대한 사용자 평가와 시스템 예측값의 오차를 기반으로 6Sub-Vectors 벤치마크 알고리즘과 성능 평가를 수행하였고, 실험 결과를 통해 제안한 시스템의 우수성을 입증하였다.

.....

논문접수일 : 2013년 08월 16일 논문수정일 : 2013년 09월 09일 게재확정일 : 2013년 09월 17일

투고유형 : 국문급행 교신저자 : 조근식

1. 서론

세계에는 수많은 사람들이 살아가고 있고, 사람

들의 일상으로부터 매일, 매 시간 새로운 뉴스가 발생한다. 뉴스는 날씨, 스포츠, 경제, 과학, 범죄 등 다양한 분야를 포함하고 있으며, 예정된 일이나 예

* 이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2011-0015484).

상하지 못한 일들을 다루고 있다. 사람들은 TV나 인터넷 포털사이트로부터 다양한 뉴스를 접하고 있으며, 뉴스를 확인하는데 많은 시간을 소비하게 된다. 이것은 사람들이 현재 무엇이 일어났는지, 가까운 미래에 어떠한 일들이 일어날 것인지를 알기 원하는 것을 의미한다. 또한, 뉴스에 대한 내용을 지인들과 공유하고 의견을 나누고자 하는 정보 교환에 대한 사용자 수요가 있음을 의미한다(Powers and El-Nawawy, 2009). 소셜 네트워크 서비스의 등장은 뉴스 등을 통한 정보의 순환과 소비가 즉시적인 특성을 갖게 하는 계기가 되었으며, TV 및 포털 사이트로부터 게재되는 뉴스뿐만 아니라, 뉴스 미디어가 즉시적으로 커버하지 못하는 범위의 뉴스들이 소셜 네트워크를 통해 실시간으로 전파되고 있다. 하지만 수많은 뉴스 콘텐츠들로 인해 사용자가 원하는 뉴스만을 선별하는 것은 쉽지 않다. 이는 대다수의 뉴스 미디어가 사용자의 관심사와는 별개로 속보 및 실시간 이슈로 사용자들에게 뉴스를 전달하고 있기 때문이다. 즉, 기존의 뉴스 미디어가 보유한 시스템들은 사용자의 관심사를 파악하지 못하고 뉴스를 푸싱서비스의 형태로 제공하고 있음을 의미한다.

거대한 규모의 뉴스 콘텐츠와 다양한 미디어로부터 사용자의 선호도에 적합한 뉴스를 제공하기 위해서는 개인화된 뉴스 서비스가 요구된다. 본 논문의 접근에서 뉴스 도메인은 다음과 같은 특징을 가지고 있다. 첫째, 뉴스의 타이틀과 방송 원고 사이에는 높은 상관관계가 있다. 둘째, 뉴스는 카테고리로 분류되어 있으며, 일부 카테고리는 계층구조를 가지고 있다. 셋째, 사용자는 연속된 뉴스 및 최신의 뉴스를 보기를 원한다. 뉴스 도메인과 관련된 이러한 특성 때문에 뉴스 도메인에서는 전형적인 협업적 필터링 방법을 이용하여 사용자에게 뉴스를 제공하는 것보다는 사용자의 선호도를 파악하고 이에

맞는 뉴스를 제공하는 콘텐츠 기반의 추천기법이 요구된다.

사용자의 선호도를 파악하기 위해서는 사용자 프로파일이 요구된다. 하지만 사용자는 자신의 관심사를 명시적으로 표시하는 것에 부담을 느끼고 있으며, 사용자의 선호도 및 관심사는 시간에 따라 변하기 때문에 자동적으로 뉴스 도메인에 맞는 사용자 프로파일을 생성하고, 지속적으로 업데이트 가능한 모델을 만드는 것이 필요하다(Liu et al., 2010). 또한 사용자 프로파일에 적합한 뉴스를 추천하기 위해서는 뉴스 콘텐츠가 포함하고 있는 정보로부터 뉴스 프로파일의 생성이 요구된다. 뉴스 도메인의 특징에서 언급된 바와 같이 뉴스는 카테고리 및 방송 원고를 포함하고 있기 때문에 이를 분석하여 뉴스 프로파일을 생성할 필요가 있다. 생성된 뉴스 프로파일은 관련 뉴스를 추출하는데도 유용하게 사용된다.

본 논문에서는 개인화된 뉴스 서비스 제공을 위한 콘텐츠 기반의 뉴스 추천 시스템을 제안한다. 사용자 프로파일은 소셜 네트워크 사용자의 정보와 최근 게시글을 이용하여 작성한다. 뉴스 프로파일은 방송사의 웹 사이트로부터 추출한 뉴스 정보를 이용하여 구축한다. 사용자가 시스템에 접속하면 사용자 프로파일을 기반으로 사용자가 선호할만한 뉴스를 추천하고, 관련 뉴스들을 제공한다. 제안한 시스템을 통해 사용자는 개인화된 뉴스 서비스를 제공받을 수 있게 된다.

본 논문의 구성은 다음과 같다. 제 2장에서는 배경 지식 및 관련 연구를 기술한다. 제 3장에서는 개인화된 뉴스 서비스를 위한 콘텐츠 기반 추천 시스템을 설명한다. 제 4장에서는 제안한 시스템을 위한 실험 데이터 및 결과를 나타낸다. 마지막으로 제 5장에서는 제안하는 연구의 결론을 맺고, 향후 연구를 기술한다.

2. 배경 지식 및 관련 연구

2.1 사용자 프로파일

개인화된 서비스를 제공하기 위해서는 사용자의 정보가 필수적으로 요구되며, 이를 위한 개인화 시스템들은 사용자 프로파일 생성이 필수적으로 요구된다(Peter and David, 2013; Hong et al., 2012; Abel et al., 2013; Zhang, 2012). 사용자 프로파일에 포함되는 정보는 개인화가 필요한 도메인에 따라서 다르지만, 구축된 사용자 프로파일의 정확도에 따라 개인화된 서비스의 품질이 결정된다고 할 수 있다. 초기 포털 사이트 및 웹 사이트들의 경우 사용자들의 관심사를 카테고리 분류하여 명시적으로 입력을 받았지만, 최근 정보보호와 관련하여 사용자들로부터 명시적으로 선호하는 분야에 대한 정보를 획득하는 것은 쉽지 않다. 따라서 많은 개인화 시스템들에서는 사용자의 소셜 네트워크로부터 자동적으로 시스템에 적합한 정보를 추출하고, 사용자 프로파일을 생성하기 위한 기법들을 연구하기 시작하였다(You et al., 2013; Thay et al., 2013; Sean et al., 2012).

본 논문에서는 소셜 네트워크 서비스 중 하나인 Facebook의 사용자들로부터 정보를 획득하고, 뉴스 서비스에 적합한 사용자 프로파일을 생성한다.

2.2 벤치마크 알고리즘

SME.SK에서는 개인화된 뉴스 서비스를 제공하기 위해 콘텐츠 기반의 뉴스 추천 기법을 제안하였다(Kompan and Bieliková, 2010). 슬로바키아에 가장 큰 뉴스 포털인 SME.SK의 뉴스 콘텐츠와 사용자 모델을 이용하여 개인화된 뉴스 추천시스템을 구축하였다. 제안된 시스템은 뉴스 콘텐츠의 유사도 측정, 사용자 모델링, 이전 두 단계를 기반으로 뉴스

추천의 세 단계를 포함한다. 뉴스 콘텐츠 유사도를 측정하기 위해 각 뉴스를 벡터 모델로 표현되며, 이를 위해 뉴스에서 불용어 등을 제거하는 등의 전처리과정을 거치고, 뉴스의 특징벡터를 6개의 항목으로 분류하여 추출하고, 이를 기반으로 각 뉴스간의 유사도를 측정한다. 특징 벡터를 위한 6가지 부분은 다음과 같다. 첫째, 뉴스의 타이틀이다. 타이틀은 뉴스의 한줄 요약과 같다. 둘째, 뉴스 콘텐츠에 나타난 타이틀 단어의 단어 빈도수이다. 타이틀이 추상적인 개념이 아니라면 본문에 나타난 빈도로 유용한 단어로 사용할 수 있다는 개념이다. 셋째, 이름과 장소를 추출한다. 넷째, 뉴스에 나타나는 주요 키워드를 의미한다. 이는 전통적인 TF-IDF(단어빈도수-문서빈도의 역) 방식을 이용하여 추출한다. 다섯째, 포털 사이트에서 제공하고 있는 뉴스 카테고리를 이용하여 가중치와 함께 표현한다. 여섯째, CLI(Coleman-Liau Index)는 텍스트의 이해도에 대한 정보를 나타낸다(Coleman and Liau, 1975). 이것은 유사도 계산을 위해 중요한 부분은 아니지만 결과를 재 정렬하기 위한 중요한 요소이다.

시스템은 뉴스 추천을 위해 사용자 프로파일과 뉴스 사이의 유사도를 다음과 같은 식 (1)을 통해 계산한다.

$$\text{similarity} = \frac{\sum_{j=1}^m \sum_{i=1}^n a_{ji} b_{ji}}{\sqrt{\sum_{j=1}^m \sum_{i=1}^n a_{ji}^2} \sqrt{\sum_{j=1}^m \sum_{i=1}^n b_{ji}^2}} \quad (1)$$

두 번째 단계로 사용자 모델링은 사용자가 방문하고 클릭한 뉴스의 쿠키 등과 같은 사용자의 암묵적인 피드백으로부터 생성된다.

마지막 단계로, 사용자에게 뉴스를 추천하기 위해서는 두 가지 요소를 활용한다. 첫 번째 요소는 방문했거나 이미 추천 받은 것과 유사한 뉴스 리스트이다. 두 번째 요소는 사용자가 방문했지만 이전에 추천되지 않은 뉴스와 유사한 리스트이다. 시스

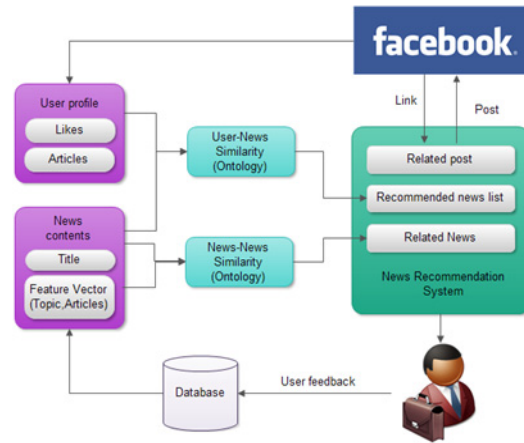
템은 사용자에게 추천된 각 뉴스와 유사한 뉴스 리스트를 추출하고, 그 리스트를 통해 사용자에게 추천되지 않은 뉴스를 선별한다. 최종적으로 사용자에게 추천되는 리스트는 추천된 뉴스와 유사한 뉴스 중 사용자가 방문하지 않았던 뉴스가 우선적으로 추천리스트에 포함되고, 그 다음 이전에 방문했지만 추천되지 않은 뉴스와 유사한 리스트 중에서 유사도가 높은 뉴스를 사용자에게 제공한다.

실험 결과는 SME.SK에서 제안한 6Sub-Vectors를 이용한 뉴스간의 유사도 측정과 이를 기반으로 한 콘텐츠 기반 추천 시스템의 우수성을 보여주었다.

3. 소셜 네트워크 기반의 콘텐츠 추천 시스템

본 논문에서 앞서 제 2장에서 소개한 관련 연구들을 기반으로 생성된 사용자의 프로파일과 뉴스의 프로파일을 이용하여 사용자에게 개인화된 뉴스를 추천하고 그 추천된 뉴스와 관련된 다른 뉴스를 추가로 추천해 주는 시스템을 제공하고자 한다. <Figure 1>은 본 논문에서 제안하는 소셜 네트워크 기반의 뉴스 추천 시스템의 구조도이다. <Figure 1>의 시스템 구조도와 같이 사용자의 프로파일에 따라 뉴스를 추천하며, 뉴스의 프로파일을 통해 유사한 뉴스를 추천해 준다.

뉴스를 추천하는 과정은 <Figure 1>과 같이 사용자-뉴스, 뉴스-뉴스 추천으로 구분한다. 사용자-뉴스 추천은 현재 사용자가 선호하는 뉴스를 추천하는 것이고, 뉴스-뉴스 추천은 현재 시청중인 뉴스와 관련성이 높은 다른 뉴스를 추천하는 것이다. 두 추천 모두 각기 다른 방법을 통해 유사성(Similarity)을 계산하여 뉴스를 추천한다. 첫 번째로 사용자-뉴스 추천을 위해 사용자의 Likes와 Interest 정보를 뉴스의 Topic 정보와 유사성을 구하고 사용



<Figure 1> News Recommendation Structure

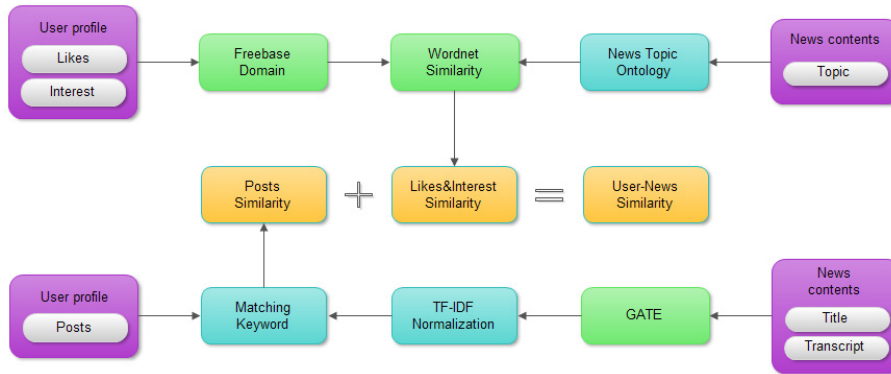
자의 Posts를 뉴스의 Transcript와 유사성을 구하여 두 유사성을 이용하여 상위 뉴스를 추천한다. 두 번째로 뉴스-뉴스 추천을 위해 각 뉴스의 Title, Topic, Transcript 간의 유사성을 이용하여 상위 뉴스를 추천한다.

3.1 사용자-뉴스 추천

본 논문에서 사용자의 프로파일을 생성하기 위해 Facebook으로부터 사용자 정보를 수집한다. 수집된 사용자 정보를 기반으로 뉴스를 추천하기 위해 <Figure 2>와 같이 2가지 방법으로 진행된다. 첫 번째로 사용자의 Likes와 Interest 정보와 뉴스의 Topic 정보 간에 유사성을 이용하고, 두 번째로 사용자의 Posts 정보와 뉴스의 Title과 Transcript 정보 간의 유사성을 이용한다.

각 단계는 다음과 같은 순서로 진행이 된다.

첫 번째로 사용자의 Likes와 Interest와 뉴스의 Topic의 유사성을 구하기 위해서 사용자의 Likes와 Interest를 단어 단위로 구분하여 Freebase에 MQL (Metaweb Query Language)을 이용하여 조회하고, 각 단어 별 메타정보를 JSON(JavaScript Object



<Figure 2> User-News Recommendation Structure

Notation) 형식으로 가져와서 Freebase 도메인 정보를 획득한다(Bollacker et al., 2008). 그리고 뉴스의 Topic을 미리 정의된 온톨로지를 통해 상위 Topic으로 변환한다. 이렇게 각각 사용자의 Likes와 Interest로부터 구한 Freebase domain과 뉴스의 상위 Topic을 식 (2)와 같이 Likes의 각 단어 W_k 별로 Topic과의 유사도를 WordNet을 통해 구하고, 각 단어 별 유사도 중 가장 큰 값을 유사도 값으로 결정한다. 여기서 식 (2)의 l 는 Likes와 Interest이고, t 는 Topic을 나타낸다.

$$likes_sim_{l,t} = \max(W = \{k | k = 1 \text{ to } likes_words_l\}, \quad (2)$$

$$wordnet_sim(W_k, topic_t))$$

실제 시스템을 구현 시에 실시간 성능을 보장하기 위해 Freebase domain과 뉴스의 Topic의 유사성은 사전에 WordNet을 통해 유사도를 구해두고 진행한다.

두 번째로 사용자의 Posts와 뉴스의 Title, Transcript의 유사성을 구하기 위해서 사용자의 Posts를 단어 단위로 구분한다. 그리고 뉴스의 Title, Transcript를 GATE¹⁾를 이용하여 형태소 단위로 나누

고(Cunningham, 2002), 그 중 명사만 추출하여 식 (3)의 TF-IDF 가중치 기법을 통해 단어 중요도를 구하고 식 (4)의 정규화 과정을 거쳐서 뉴스의 대표 단어를 선정한다. 여기서 식 (3)과 식 (4)에서 i 는 뉴스이고, j 는 뉴스에서 나오는 단어이다.

$$TF-IDF_{ij} = tf_{ij} \times idf_i \quad (3)$$

$$TFIDF_nor_{i,j} = \frac{TF-IDF(t_i, d_j)}{\sqrt{\sum_s TF-IDF(t_s, d_j)^2}} \quad (4)$$

이렇게 각각 사용자의 Posts로부터 구한 단어와 뉴스의 Title, Transcript로부터 구한 핵심 단어들을 식 (5)와 같이 유사도를 구한다. 여기서 식 (5)의 p 는 Posts이다.

$$posts_sim_{p,i} = \frac{posts_words \sum_{k=1}^W \text{if} \begin{cases} 1, W_k = TFIDF_nor_i \\ 0, otherwise \end{cases}}{count(TFIDF_nor_i)} \quad (5)$$

앞서 구한 두 가지 유사도를 각기 다른 중요도($\alpha + \beta = 1.0$)를 식 (6)과 같이 계산하여 사용자-뉴스 유사도를 구하여 그 중 Top-N개의 뉴스를 높은 순으

1) GATE : <http://gate.ac.uk/>.

로 선정하여 사용자에게 뉴스를 추천할 수 있다. 여기서 식 (6)의 u 는 사용자이고, n 은 뉴스이다.

$$news_sim_{u,n} = (a \times likes_{\sim l,t}) + (\beta \times posts_{p,i}) \quad (6)$$

본 논문에서는 a 를 0.5, β 를 0.5로 지정하여 기존 관심사인 Likes와 최근 관심사인 Posts의 중요도를 동등하게 평가하여 진행하였다.

3.2 뉴스-뉴스 추천

앞서 제 3.1절에서 사용자-뉴스 추천을 하고, 추천된 뉴스와 관련된 다른 뉴스를 추천하기 위해 <Figure 3>와 같이 2가지 방법을 통해 두 뉴스간의 유사성을 구하여 추천하고자 한다.

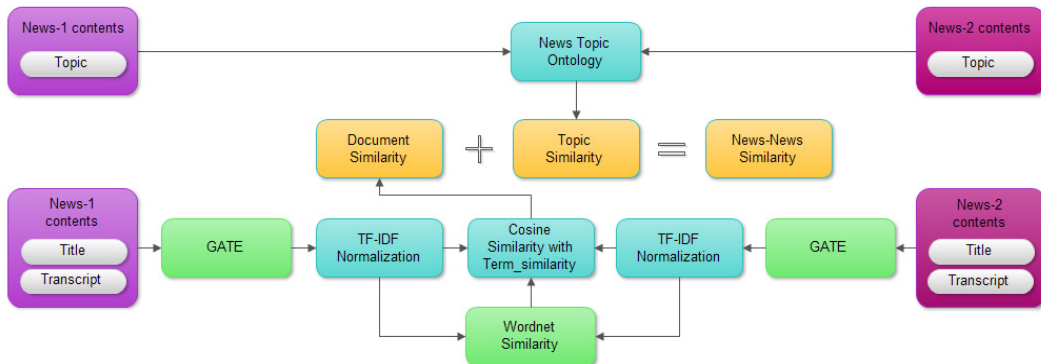
첫 번째로 각 뉴스의 Topic 간의 유사성을 구하기 위해 사전에 정의된 온톨로지를 통해 Topic의 상위 온톨로지 유사성을 구한다. 만약 같은 온톨로지에 속한다면 식 (7)과 같이 1.0의 유사성을 갖게 되며, 만약 같은 온톨로지는 아니지만 상위 온톨로지가 서로 같거나 혹은 상위-하위 온톨로지가 같은 경우에는 0.7의 유사성을 갖게 된다. 그 외에 온톨로

지가 매칭 되지 않는 경우에는 두 Topic간의 유사성이 없다고 판단하여 0.0의 유사성을 갖게 된다. 여기서 식 (7)의 n 과 m 은 유사성을 비교할 서로 다른 뉴스이다.

$$topic_sim_{n,m} = \begin{cases} 1.0, & \text{if } (topic_n = topic_m) \\ 0.7, & \text{if } (s \in topic_n, t \in topic_m | s = t) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

두 번째로 각 뉴스의 Title과 Transcript 간의 유사성을 구하기 위해 뉴스의 Title과 Transcript를 GATE를 이용하여 형태소 단위로 나누고 그 중 명사만 추출하여 식 (3)의 TF-IDF 가중치 기법을 통해 단어 중요도를 구하고 식 (4)의 정규화 과정을 거쳐서 뉴스의 대표 단어를 선정한다. 이렇게 구한 뉴스 별 핵심 단어들을 식 (8)의 WordNet 유사성을 포함한 코사인 유사도(Cosine similarity)를 통해 두 뉴스간의 유사도를 구한다.

$$doc_sim_{n,m} = \frac{\sum_k TFIDF_{nor_{k,n}} \times TFIDF_{nor_{k,m}} \times term_sim_{n,m}}{\sqrt{\sum_k TFIDF_{nor_{k,n}}^2} \times \sqrt{\sum_k TFIDF_{nor_{k,m}}^2}} \quad (8)$$



<Figure 3> News-News Recommendation Structure



<Figure 4> Information collection using Facebook FQL

앞서 구한 두 가지 유사도를 각기 다른 중요도($\alpha + \beta = 1.0$)를 식 (9)와 같이 계산하여 뉴스-뉴스 유사도를 구하여 그 중 Top-N개의 뉴스를 높은 순으로 선정하여 사용자에게 연관된 뉴스를 추천할 수 있다.

$$news_sim_{n,m} = (\alpha \times topic_sim_{n,m}) + (\beta \times doc_sim_{n,m}) \quad (9)$$

본 논문에서는 α 를 0.6, β 를 0.4로 지정하여 Topic의 중요도를 좀 더 높게 평가하여 진행하였다.

4. 구현 및 실험

본 논문에서 제안하는 개인화된 뉴스 서비스를 위한 소셜 네트워크 기반의 콘텐츠 추천 시스템을 위하여 Intel Core i7-2600 3.40GHz, RAM 6GB, MS Windows 7, MS-SQL 2005 환경에서 ASP.NET with C#, HTML, CSS, JavaScript, Node.js로 구현하였다.

먼저 Facebook으로부터 사용자의 프로파일을 생성하기 위해 <Figure 4>와 같이 Facebook의 FQL (Facebook Query Language)을 이용하여 사용자의 프로파일 정보-Facebook id, user name, likes, interest, gender, birthday, location, posts(articles) 등-정보를 JSON 형식으로 가져와서 처리한다. 사용자의 프로파일 정보는 사용자가 플랫폼을 이용할 때마다 최신 정보로 갱신하기 위해 본 연구의 웹 사이트에서 사용자가 로그인할 때 마다 실시간으로 수집하여 Database를 갱신한다.

다음으로 사용자에게 제공될 뉴스정보-URL, Type, Topic, Air date, Title, Summary, Transcript 등-는 PBS²⁾로부터 2012년 01월 02일부터 2013년 03월 07일 까지 총 1727편의 뉴스 정보를 [Figure 5]와 같이 웹크롤러(WebCrawler)를 Windows 기반의 C# Form으로 구현하여 수집하고 이를 Database화 하였다.

2) Public Broadcasting Service : <http://www.pbs.org/>.



<Figure 5> Webcrawler for Collect News of PBS

본 논문에서 제안하는 뉴스 추천 시스템은 <Figure 6>와 같은 단계로 구현되었다. <Figure 6>(a)와 같이 처음 사용자가 웹사이트에 접속하여 Facebook 계정을 통해 로그인하여 사용자의 정보를 자동으로 수집하고, <Figure 6>(b)와 같이 사용자-뉴스 유사도를 이용하여 Top-5개의 뉴스를 추천하며, 추가로 뉴스 추천을 요구하는 경우에는 5개씩 추가로 추천한다. 또한 별도로 사용자가 원하는 뉴스를 검색할 수 있도록 검색 기능을 제공하며, 추천된 뉴스들의 핵심 단어들을 Tag-Cloud를 통해 제공된다. 사용자는 제공된 뉴스를 선택하여 해당 뉴스를 <Figure 6>(c)와 같이 시청할 수 있으며, 제목 하단에 위치한 평점 버튼을 1~5사이로 선택하여 추천된 뉴스의 선호도를 입력할 수 있도록 제공하였다. 또한 하단에 Related News에 뉴스-뉴스 유사도를 통해 추천된 뉴스를 선택하여 연관된 뉴스를 탐색하거나 Facebook에 관련된 글을 작성할 수 있도록 하였다.

사용자로부터 수집된 뉴스 평점 정보를 이용하여 본 논문에서 제안하는 시스템의 성능을 비교하기 위해 PBS 뉴스 1727개를 이용하여 Facebook의 47명의 유저를 대상으로 사용자 별 평균 101개의 뉴스 평점을 수집하였다.

$$Pr_i = \frac{(sim_i - MinSim) \times 4}{(MaxSim - MinSim)} + 1 \quad (10)$$

사용자의 평점과 예측된 유사도 값을 비교하기 위해 식 (10)과 같이 Karypis의 Numerical prediction을 이용하여 예측 값을 구하였다(Lekakos and Caravelas, 2008). 여기서 Pr_i 는 i 번째 뉴스의 예측 값이며, 이를 구하기 위해 i 번째 뉴스의 최소 유사도 값(MinSim)과 최대 유사도 값(MaxSim)을 이용하여 식 (10)과 같이 계산하였다. 계산된 예측 값과 수집된 사용자의 평점 값으로 절대평균오차(Mean Absolute Error; MAE)를 구하였다.

Microsoft News Recommendation Project

Home
About
Logout

Ontology-based Semantic Interoperability between Social Network Service and Video for Personalized News Service
(개인화된 뉴스 서비스를 위한 SNS 및 비디오의 온톨로지 기반 의미적 상호운용성 확보)

You can use this app, if you can login on Facebook. ? You can get personalized news service after login with Facebook.
(본 앱을 사용하기 앞서 페이스북으로 로그인하셔야 진행 가능합니다.)

f Login with Facebook

The information obtained is only for the purpose of generating news recommendation in our news service. The information include your name, age, birth, Likes, Interests, Friends list, and parts of recent news feed.
이 웹사이트에서 수집한 사용자의 페이스북의 프로파일 정보는 연구 목적으로만 사용됩니다. (이름, 나이, 생일, Likes, Interes, Friends list, 최근 뉴스피드 일부)

[Public Broadcasting Service \(PBS\)](#) is a copyright holder for all news information offered at here.
이 웹사이트에서 제공되는 모든 뉴스 정보는 [Public Broadcasting Service \(PBS\)](#)가 저작권을 보유하고 있습니다.

(a) User login via Facebook Account

SEARCH NEWS (뉴스검색)

| | 2013년 6월 | | | | | | | 2013년 6월 | | | | | | |
|------|----------|----|----|----|----|----|----|----------|----|----|----|----|----|----|
| | 일 | 월 | 화 | 수 | 목 | 금 | 토 | 일 | 월 | 화 | 수 | 목 | 금 | 토 |
| Date | 26 | 27 | 28 | 29 | 30 | 31 | 1 | 26 | 27 | 28 | 29 | 30 | 31 | 1 |
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| | 30 | 1 | 2 | 3 | 4 | 5 | 6 | 30 | 1 | 2 | 3 | 4 | 5 | 6 |

Selected date: 2013-06-01 Selected date: 2013-06-01

Keyword: Search Close

AirDate Desc v

TAG CLOUD

[haley_barbour](#) [al_gore](#) [heather_timmons](#) [richard_blanco](#) [kirby_dick](#)
[louise_erdrich](#) [password](#) [ian_james](#) [suzi_parker](#) [ahmet_davutoglu](#)
[ivory](#) [fred_hof](#) [harper_reed](#) [stuart_bowen](#) [jim_lehrer](#) [jeff_flake](#)
[david_ferry](#) [brad_meltzer](#) [kenneth_cuccinelli](#)
[lori_montgomery](#)

RECOMMENDED NEWS (추천 뉴스)

| | | | | |
|---|---|--|--|---|
| | | | | |
| Budget Plan Offers Laid Out in Negotiations, But Deal Remains Elusive | News Wrap: Computer Maker Dell Goes Private After Decades of Public Trading | The Mind of a Rampage Killer | Logistical Challenges of Combating Islamic Militants in Saharan Africa | Citing Financial Woes, U.S. Postal Service Announces End of Saturday Delivery |

(b) News Recommendation by User-News Similarity



(c) Watching News and News Recommendation by News-News Similarity

<Figure 6> News Recommendation System

본 실험 평가를 위해 일반적으로 사용되는 TF-IDF와 제 2.2절에서 소개한 벤치마크 알고리즘인 6Sub-Vectors 그리고 본 논문에서 제안하는 방법을 비교하였다. 성능 비교를 위해 비교하려는 알고리즘 별로 추천하는 뉴스의 Top-N을 5부터 25까지 5씩 증가해서 절대평균오차를 <Table 1>와 같이 구하였다. 그 결과 본 논문에서 제안하는 방법이 0.8212~0.8946 정확도로 다른 알고리즘들이 대부분 0.9의 정확도를 갖는 것에 반하여 우수함을 보였으며, 모든 알고리즘들이 Top-N이 낮을 때가 높을 때 보다 좋은 성능을 보이고 있는데 이는 사용자가 추천 받은 뉴스가 많아질수록 시청하는 비율이 낮

고 관심도가 상대적으로 낮아짐에 따라 평가 비율이 낮기 때문으로 보여진다.

5. 결론 및 향후 연구

일반적으로 제공되는 대부분의 뉴스 추천 시스템은 사용자가 명시적으로 선호하는 뉴스 카테고리를 입력하고 그에 따라 뉴스를 추천하는 방식이다. 이와 같은 추천은 사용자가 선호하는 관심 분야를 명시적으로 입력해야 함은 물론이거니와 시간의 변화에 따라 변화되는 개인의 선호하는 관심분야가 바뀌는 경우 매번 명시적으로 입력해야 하는 불편

<Table 1> News Prediction Accuracy(MAE)

| Method \ Top-N | 5 | 10 | 15 | 20 | 25 |
|----------------|--------|--------|--------|--------|--------|
| TF-IDF | 0.9253 | 0.9276 | 0.9319 | 0.9427 | 0.9803 |
| 6Sub-Vectors | 0.8926 | 0.9005 | 0.9047 | 0.9128 | 0.9426 |
| our method | 0.8212 | 0.8374 | 0.8595 | 0.8672 | 0.8946 |

함이 존재한다. 본 논문에서 제안하는 방법은 소셜 네트워크 사용자로부터 사용자가 소셜 네트워크에 남긴 정보를 바탕으로 하여 뉴스 도메인에 대한 선호도를 추출하여 뉴스를 추천한다. 사용자 프로파일 일을 동적으로 생성하고 지속적으로 업데이트하기 위하여 소셜 네트워크 서비스인 Facebook 사용자의 정보를 추출하고, 사용자가 명시한 Likes 정보를 추출한다. Likes 정보는 Freebase의 객체 정보를 기반으로 뉴스 도메인에 해당하는 카테고리 매핑되고, 사용자의 Likes 정보가 바뀌면 시스템은 자동으로 사용자 프로파일을 업데이트할 수 있다. 제안된 시스템은 사용자에게 뉴스를 추천하는 것으로 끝나는 것이 아니라 추천된 뉴스와 연관성이 높은 뉴스를 추가로 제공함으로써 사용자에게 정보 획득의 기회를 확대시킨다. 또한 뉴스를 제공하는 입장에서는 추천된 뉴스를 보고 있는 사용자에게 연관된 다른 뉴스를 추천해 줌에 따라 사용자에게 뉴스 노출 시간을 증가시켜 다른 서비스를 제공할 기회가 확대될 수 있다.

향후 연구로 사용자의 소셜 네트워크 게시물로부터 단순히 단어만 추출하기보다 자연어 처리 기법을 적용하여 게시물의 의미를 이해하고(Ek et al., 2011; Rao et al., 2013), 그 이해를 바탕으로 하여 보다 정확한 뉴스를 추천하는 연구도 필요하다. 또한 추천 시스템이 공통적으로 직면하는 문제인 추천의 실시간성을 보장하기 위한 연구(Li et al., 2011)가 요구된다.

참고문헌

- Abel, F., E. Herder, G. J. Houben, N. Henze, and D. Krause, "Cross-system user modeling and personalization on the Social Web," *User Modeling and User-Adapted Interaction*, Springer Netherlands, Vol.23, No.2-3(2013), 169~209.
- Bollacker, K., C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase : a collaboratively created graph database for structuring human knowledge," *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, (2008), 1247~1250.
- Cunningham, H., "GATE-a general architecture for text engineering," *Journal of Computers and the Humanities*, Vol.36, No.2(2002), 223~254.
- Coleman, M. and T. L. Liao, "A computer readability formula designed for machine scoring," *Journal of Applied Psychology*, Vol.60, No.2(1975), 283~284.
- Ek, T., C. Kirkegaard, H. Jonsson and P. Nugues, "Named Entity Recognition for Short Text Messages," *Procedia-Social and Behavioral Sciences*, Vol.27(2011), 178~187.
- Hong, K., H. Jeon and C. Jeon, "UserProfile-based personalized research paper recommendation system," *Proceeding of 8th International Conference on Computing and Networking Technology*, (2012), 134~138.
- Kompan, M. and M. Bieliková, "Content-Based News Recommendation," *Proceedings of International Conference on E-Commerce and Web Technologies*, (2010), 61~72.
- Lekakos, G. and P. Caravelas, "A hybrid approach for movie recommendation," *Multimedia tools and applications*, Vol.36. No.1-2(2008), 55~70.
- Li, L., D. Wang, T. Li, D. Know, and B. Padmanabhan, "SCENE : a scalable two-stage personalized news recommendation system," *Proceedings of the 34th international ACM SIGIR conference on Research and develop-*

- ment in Information Retrieval*, (2011), 125~134.
- Liu, J., P. Dolan, and E. R. Pedersen, "Personalized News Recommendation Based on Click Behavior," *Proceeding of international conference on Intelligent user interfaces*, (2010), 31~40.
- Miller, G. A., "WordNet : a lexical database for English," *Communications of the ACM*, Vol. 38, No.11(1995), 39~41.
- Peter, B. and N. C. David, "Preface to the Special Issue on Personalization in Social Web systems," *User Modeling and User-Adapted Interaction*, Springer Netherlands, Vol.23, No. 2-3(2013), 83~87.
- Powers, S. and M. El-Nawawy, "Al-Jazeera English and global news networks : clash of civilizations or cross-cultural dialogue," *Journal of Media, War and Conflict*, Vol.2, No.3(2009), 263~284.
- Rao, J., A. Jia, Y. Feng, and D. Zhao, "Personalized News Recommendation Using Ontologies Harvested from the Web," *Proceedings of International Conference on Web-Age Information Management*, (2013), 781~787.
- Sean, V., K.-J. Oh, and G.-S. Jo, "A User Profile-based Filtering Method for Information Search in Smart TV Environment," *Journal of Intelligence and Information Systems*, Vol.18, No.3(2012), 97~117.
- Thay, S., I. Ha, and G.-S. Jo, "Incorporating Social Relationship discovered from User's Behavior into Collaborative Filtering," *Journal of Intelligence and Information Systems*, Vol.19, No. 2(2013), 1~20.
- You, T., A. N. Rosli, I. Ha, and G.-S. Jo, "Clustering Method based on Genre Interest for Cold-Start Problem in Movie Recommendation," *Journal of Intelligence and Information Systems*, Vol.19, No.1(2013), 57~77.
- Zhang, L., *Webpage Personalization and User Profiling*, Computational Advertising Workshop at SAMSI, 2012. Available at http://www.samsi.info/sites/default/files/liang_august2012.pdf(Downloaded 23 September, 2013).

Abstract

Content-based Recommendation Based on Social Network for Personalized News Services

Myung-Duk Hong^{*} · Kyeong-Jin Oh^{**} · Myung-Hyun Ga^{***} · Geun-Sik Jo^{****}

Over a billion people in the world generate new news minute by minute. People forecasts some news but most news are from unexpected events such as natural disasters, accidents, crimes. People spend much time to watch a huge amount of news delivered from many media because they want to understand what is happening now, to predict what might happen in the near future, and to share and discuss on the news. People make better daily decisions through watching and obtaining useful information from news they saw. However, it is difficult that people choose news suitable to them and obtain useful information from the news because there are so many news media such as portal sites, broadcasters, and most news articles consist of gossipy news and breaking news. User interest changes over time and many people have no interest in outdated news. From this fact, applying users' recent interest to personalized news service is also required in news service. It means that personalized news service should dynamically manage user profiles. In this paper, a content-based news recommendation system is proposed to provide the personalized news service. For a personalized service, user's personal information is requisitely required. Social network service is used to extract user information for personalization service. The proposed system constructs dynamic user profile based on recent user information of Facebook, which is one of social network services. User information contains personal information, recent articles, and Facebook Page information. Facebook Pages are used for businesses, organizations and brands to share their contents and connect with people. Facebook users can add Facebook Page to specify their interest in the Page. The proposed system uses this Page information to create user profile, and to match user preferences to news topics. However, some Pages

* Department of Computer and Information Engineering, Inha University

** Department of Computer and Information Engineering, Inha University

*** Department of Computer and Information Engineering, Inha University

**** Corresponding author: Geun-Sik Jo

School of Computer and Information Engineering, Inha University

100 inharo, Nam-gu, Incheon 402-751, Korea

Tel: +82-32-875-5863, Fax: +82-32-875-5863, E-mail: gsjo@inha.ac.kr

are not directly matched to news topic because Page deals with individual objects and do not provide topic information suitable to news. Freebase, which is a large collaborative database of well-known people, places, things, is used to match Page to news topic by using hierarchy information of its objects. By using recent Page information and articles of Facebook users, the proposed systems can own dynamic user profile. The generated user profile is used to measure user preferences on news. To generate news profile, news category predefined by news media is used and keywords of news articles are extracted after analysis of news contents including title, category, and scripts. TF-IDF technique, which reflects how important a word is to a document in a corpus, is used to identify keywords of each news article. For user profile and news profile, same format is used to efficiently measure similarity between user preferences and news. The proposed system calculates all similarity values between user profiles and news profiles. Existing methods of similarity calculation in vector space model do not cover synonym, hypernym and hyponym because they only handle given words in vector space model. The proposed system applies WordNet to similarity calculation to overcome the limitation. Top-N news articles, which have high similarity value for a target user, are recommended to the user. To evaluate the proposed news recommendation system, user profiles are generated using Facebook account with participants consent, and we implement a Web crawler to extract news information from PBS, which is non-profit public broadcasting television network in the United States, and construct news profiles. We compare the performance of the proposed method with that of benchmark algorithms. One is a traditional method based on TF-IDF. Another is 6Sub-Vectors method that divides the points to get keywords into six parts. Experimental results demonstrate that the proposed system provide useful news to users by applying user's social network information and WordNet functions, in terms of prediction error of recommended news.

Key Words : Content-Based Recommendation System, Personalized News Service, User Profile, Social Network

저 자 소개



홍명덕

Received a B.S. degree Computer Science, Seoul Digital University, Korea, in 2008, and a M.S. degree in Information Engineering from Inha University, Korea, in 2011. He worked for Kuwoo Information Technology as a researcher 2005 to 2008. He is currently a Ph.D. Candidate in Information Engineering of Inha University, Korea. His research interests include Recommender System, Semantic Web, Ant Colony Optimization and Meta-Heuristic.



오경진

Received a B.S. degree Computer and Information Engineering from Inha University, Korea, in 2006, and a M.S. degree in Information Engineering from Inha University, Korea in 2008. He is a Ph.D. Candidate in Information Engineering of Inha University, Korea. His research interests include Data Mining and Semantic Web.



가명현

Received a B.S. degree Computer and Information Engineering from Inha University, Korea, in 2012. He is currently a M.S. Candidate in Information Engineering of Inha University, Korea. His research interests include Recommender System and Data Mining.



조근식

Is a Professor in Computer and Information Engineering, Inha University, Korea. He received the B.S. degree in Computer Science from Inha University in 1982. He received the M.S. and the Ph.D. degrees in Computer Science from City University of New York in 1985 and 1991, respectively. He has been the General Chair and/or Technical Program Chair of more than 20 international conferences and workshops on artificial intelligence, knowledge management, and semantic applications. His research interests include knowledge-based scheduling, ontology, semantic Web, intelligent E-Commerce, constraint-directed scheduling, knowledge-based systems, decision support systems, and intelligent agents. He has authored and coauthored five books and more than 200 publications.