

사용자 프로파일을 이용한 개인화된 토픽맵 랭킹 알고리즘

박정우^o 김권일 채진기 이상훈
국방대학교

piw2236^o@naver.com, kweon95@gmail.com, nice5556@nate.com, hoony@kndu.ac.kr

Personalized Topic map Ranking Algorithm Using User Profile

Jungwoo Park^o Kweonil Kim Jinki Chae Sanghoon Lee
Korea National Defense University

1. 서론

오늘날 전 세계 수많은 사람들이 매일 자신의 관심영역에 대한 웹서핑을 하면서 개인의 관심주제를 찾고 있으며, 정보 검색시스템을 사용하는 경우에도 그 결과가 방대하기 때문에 사용자가 원하는 정보를 정확하게 얻는 것이 어려운 것이 현실이다. 즉, 오늘날의 검색엔진은 사용자의 관심(Interest)과 배경지식(Context)을 고려하지 않고 검색 결과를 제공하는 단점을 가지고 있다[9]. 그래서 이러한 문제점을 해결하기 위해 컴퓨터가 정보의 의미를 분석가 능케 하는 시맨틱웹(Semantic Web)이 등장하였다.

일종의 개념 네트워크라 할 수 있는 시맨틱웹에서는 RDF, OWL, Topic map 등의 온톨로지(Ontology) 언어들이 있다. ISO 표준인 토픽맵(Topic maps)은 지식층(Knowledge Layer)과 정보층(Information Layer)을 쉽게 연결하기 위해 만들어졌다. 이에 토픽맵은 기존 정보의 형태를 변환하지 않고도 정보자원을 통합하여 서로 다른 토픽(Topics)들의 병합이 가능하다. 또한, 관계(Associations) 기능으로 무한한 정보제공과 보다 광범위한 연계가 가능하다는 장점을 가지고 있다. 하지만 이러한 장점을 가진 토픽맵 또한 개별 사용자의 관심과 배경지식을 고려하지 않고 최초 도메인 전문가에 의해 구축된 토픽맵 상의 토픽(Topics)과 연관되는 관계(Associations), 자원(Occurrences)만을 이용하여 관련 토픽맵 정보를 사용자에게 제공하고 있다.[1,12]

현재의 토픽맵에서 사용자가 선호하는 개인화된 토픽맵 정보를 제공하기 위한 개인화 기능으로 개인 선호항목 설정, 필터링, 범위제한 등 사용자가 직접 관심정보를 사전에 설정하는 기능을 제공하고 있으나 사용자를 위한 개인화된 토픽맵 정보 제공 측면에서 만족스럽지 못하다.[11]

따라서 본 논문에서는 특정 도메인 토픽맵에서 사용자가 원하는 정보를 제공받기 위해 사용자의 토픽 선호도 벡터(Topic Preference Vector)와 토픽맵 지식층의 기본요소인 토픽(Topics)과 관계(Associations)를 이용한 토픽맵 랭킹 알고리즘을 제안한다. 그리고 제안된 알고리즘을 적용한 개인화된 토픽맵 랭킹 시스템을 통해 제공된 정보와 기존 토픽맵에서 제공하는 정보의 효율성 비교를 통해 제안된 알고리즘의 성능을 확인하였다.

2. 개인화된 토픽맵 랭킹(PTR) 알고리즘

2.1 개인화된 토픽맵 랭킹 시스템

본 논문에서 제안하는 개인화된 토픽맵 랭킹 알고리즘을 적용한 시스템은 클라이언트 측에서 작동된다. 그림 1과 같이 개인화된 토픽맵 랭킹 시스템의 절차는 첫 번째, 사용자가 토픽맵 웹 인터페이스에서 로그인후 선호하는 정보들에 대한 클릭 정보를 사용자 프로파일 데이터베이스(User Profile DB)에 수집한다. 두 번째, 사용자가 실시간 토픽(Topic)에 대한 클릭 이벤트를 발생시킬 때 마다 사용자 프로파일 정보를 바탕으로 사용자의 토픽 선호도 벡터(Topic Preference Vector)가 계산되어 데이터베이스에서 유지된다[3]. 세 번째, 사용자 선호도 벡터와 토픽맵 지식층(Knowledge Layer)의 기본 요소인 토픽(Topics)과 관계(Associations)를 이용한 토픽맵 랭킹 알고리즘에 의해 사용자 선호도에 맞는 개인화된 토픽(Topics)정보를 순위화 하여 사용자에게 제공되어 진다.

2.2 개인화된 토픽맵 랭킹 알고리즘 [5,6,7,15,16]

$$W_i = T_i \times \log \left[\frac{(U_i + 0.5)(N - n_i + 0.5)}{(n_i + 0.5)(U - u_i + 0.5)} + k \right] \quad (1)$$

Where, T_i : topic preference vector of topic i , N : size of topics in a specific topic map,

n_i : number of topics associated topic i , U : number of topics in user profile

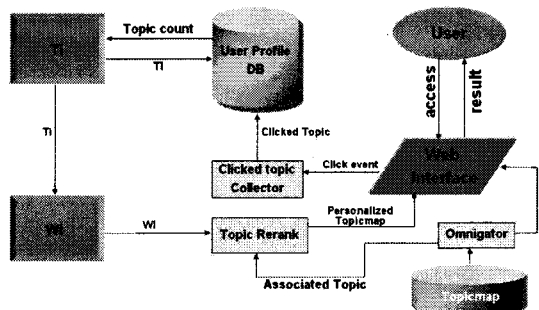
u_i : number of topics in U , associated topic i , k : constant

위의 식(1)에서 표현된 함수의 개인화된 토픽맵 랭킹 알고리즘은 표 1와 같이 표현된다.

(표 1) 개인화된 토픽맵 랭킹 알고리즘(PTR Algorithm)

<ol style="list-style-type: none"> 1. Let be $T = [T_1, \dots, T_n]$ 2. Let T_i be Topic preference vector considering user click counts 3. Let N be the size of topics in a specific topicmap 4. Let n_i be the number of topics in user profile 5. Let U be the number of topics associated topic i 6. Let u_i be the number of topics in U, associated topic i 7. Let W_i be the weight of topic i 8. for each T_i, $i=[0,1, \dots, N]$ 9. $W_i = T_i \times \log \left[\frac{(U_i + 0.5)(N - n_i + 0.5)}{(n_i + 0.5)(U - u_i + 0.5)} + k \right]$ 10. end 11. for each W_i, $i=[0,1, \dots, N]$ 12. T_i. sort() as W_i 13. end 14. Return T

(그림1) 개인화된 토픽맵 랭킹 시스템 아키텍처



3. 성능평가

토픽맵 랭킹 알고리즘을 통해 제공된 토픽 정보의 적절성을 비교하기 위해 기존 토픽맵에서 제공된 정보와 개인화된 토픽맵 랭킹 알고리즘을 통해 제공된 정보를 사용자 입장에서 비교하여 평가 하였다.

3.1 평가 방법

제안한 개인화된 토픽맵 랭킹 알고리즘을 통해 제공된 정보의 적절성 평가를 위해 20명의 대학원생을 선정하여 온토피아에서 제공하는 Italian Opera 토픽맵을 각자 10일간 사용하여 개인별 프로파일을 유지하고, 제안된 알고리즘을 통한 토픽정보와 기존 토픽맵에서 제공되는 토픽정보가 얼마나 개인 선호도에 부합되는지에 관한 평가를 한다. 여기서 평가 대상 토픽은 관계된 토픽이 10개 이상인 토픽으로 한정하여, 사용자들이 직접 적절성을 평가하였다. 그리고 상위 10개의 토픽에 대해서만 Precision과 Recall이 조합된 F-Score를 사용하여 비교평가 하였다. 평가 절차는 다음과 같다. 온토피아에서 제공하는 토픽맵 환경에서 사용자가 선호하는 토픽을 선택하게 되면, 기존 토픽맵에서 제공되는 정보와 개인화된 토픽맵 랭킹 알고리즘을 통해 제공되는 토픽맵 정보에 대해 사용자가 직접 적절성 평가를 0(부적절) 또는 1(적절)로 실시하였다.

3.2 평가 결과

평가 참가자 20명이 직접 평가한 데이터를 상위 10개의 토픽에 대한 평균값을 Precision, Recall, F-Score로 구분하여 비교하였다. Precision 비교에서는 기존 토픽맵에 비해 32.5%의 효율성 향상을 가져왔다. Recall 비교에서 또한 11.3% 효율성 향상을 가져왔다. 이는 제한된 범위의 특정 도메인 토픽맵의 특성상 비교적 작은 수의 토픽들 간의 Recall 비교에서는 기존 토픽맵과 큰 차이를 보이지 않음을 알 수 있다. 하지만 Recall 역시 제안된 알고리즘의 효율성을 보여 주었다. Precision과 Recall을 조합한 F-Score의 비교에서도 역시 23%의 효율성 향상을 보여주어 제안한 알고리즘의 성능을 입증하였다. 이상에서 보여준 결과는 연관 토픽(Association Roles)의 수가 10개 이상인 토픽들을 대상으로 평가 하였다. 하지만 토픽의 수가 방대해지고, 토픽간의 관계(Associations)가 더욱 복잡해진다 면 본 알고리즘에 의한 효율성은 더욱 증가할 것이다.

4. 결론

본 논문에서는 특정 도메인 토픽맵을 이용하는 사용자가 원하는 정보를 찾을 때, 사용자 개인의 관심도에 맞고 보다 개인화된 토픽맵 정보 제공을 목적으로 사용자의 토픽 선호도와 프로파일을 고려하여 토픽들 간 최적의 연관관계를 반영한 개인화된 토픽맵 랭킹 알고리즘을 제안하였다. 본 알고리즘을 통해 토픽맵 사용자는 실시간 토픽 선택 행위에 따라 개인 선호도에 맞게 랭킹된 정보를 제공 받을 수 있게 됨으로써 토픽맵 사용자의 개인화 성능 향상을 가져올 수 있는 장점을 가진다.

참 고 문 헌

- [1] 한국전자거래진흥원, TopicMaps 응용 표준 및 활용 가이드라인 개발, 2003
- [2] Zhongming Ma, Gautam Pant, Olivia R. Liu Sheng, "Interest-based personalized search", ACM Transactions on Information Systems, Volume 25 Issue 1 Article NO5, 2007.
- [3] Feng Qiu, Junghoo Cho, "Automatic Identification of User Interest For Personalized Search", Proceedings of the 15th international conference on WWW, Session: Improved search ranking, pp.727-736, 2006.
- [4] Google personalized search, <http://www.google.com/psearch>
- [5] S. E. Robertson, S. Walker, "Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval", Proceedings of the 17th annual international ACM SIGIR conference, pp.232-241, 1994.
- [6] Stefan Buttcher, Charles L. A. Calrke, Brad Lushman, "Term proximity scoring for ad-hoc retrieval on very large text collections", Proceedings of the 29th annual international ACM SIGIR conference, Poster Session, pp.621-622, 2006.
- [7] Jaime Teevan, Susan T. Dumais, Eric Horvitz, "Personalizing search via automated analysis of interests and activities", Proceedings of the 28th annual international ACM SIGIR conference, Session: User studies, pp.449-456, 2005.
- [8] Kelly D and Teevan J, "Implicit feedback for inferring user preference", SIGIR Forum, 37(2), pp.18-28, 2003.
- [9] Dandan Wang, Darina Dicheva, Christo Dichev, Jerry Akouala, "Retrieving information in topic maps: the case of TM4L", Proceedings of the 45th annual southeast regional conference, pp.88-93, 2007.
- [10] Liu F, Yu C and Meng W. "Personalized Web search by mapping user queries to categories", In Proceedings of CIKM'02, pp.558-565, 2002.
- [11] Steve Pepper, "The TAO of Topic Maps", <http://www.ontopia.net>
- [12] <http://www.frotoma.com>
- [13] Yabo Xu, Ke Wang, Benyu Zhang, Zheng Chen, "Privacy-enhancing personalized web search", Proceedings of the 16th international conference on World Wide Web, pp.591-600, 2007.
- [14] Andrea Ernst-Gerlach, Norbert Fuhr, "Retrieval in text collections with historic spelling using linguistic and spelling variants", Proceedings of the 2007 conference on Digital libraries, pp.333-341, 2007.
- [15] Stephen E. Robertson, Steve Walker, and Micheline Hancock-Beaulieu. Okapi at TREC-7. In Proceedings of the Seventh Text REtrieval Conference, Gaithersburg, USA, November 1998.
- [16] Karen Spärck Jones, Steve Walker, and Stephen E. Robertson. A Probabilistic Model of Information Retrieval: Development and Comparative Experiments (parts 1 and 2). Information Processing and Management, 36(6):779-840, 2000.
- [17] Custom search engine, <http://www.google.com/coop>