

사용자의 의도를 반영한 문맥 광고 랭킹 개선 기법*

정다운, 하종우, 심규선, 이상근
고려대학교 정보통신대학 컴퓨터통신공학부
e-mail:{daounjung, okcomputer, bluesks, yalphy}@korea.ac.kr

Improving Contextual Advertising Ranking by Reflecting User Intention

DaOun Jung, JongWoo Ha, Kyu-Sun Sim, SangKeun Lee
Division of Computer and Communications Engineering, Korea University

요 약

최근 몇 년간 정보 검색 분야에서 문맥 광고에 관한 연구가 활발히 연구되고 있다. 하지만 기존의 관련된 연구들은 대부분 웹페이지 내용만을 활용하여 유사한 광고를 찾고자 하였다. 그럼으로써 동일한 웹페이지를 접속하는 다양한 의도를 가진 사용자들이 동일한 광고를 보게 된다는 한계가 존재하였다. 본 논문에서는 웹페이지의 내용뿐만 아니라 각각의 사용자들의 웹페이지 방문 의도를 웹 페이지 방문 히스토리로부터 추출하여 이를 활용한 기법을 제안하고자 한다. 또한 실험을 통하여 본 논문에서 제안된 기법이 사용자 방문 의도를 반영함으로써 기존 기법에 비해 성능이 향상되었음을 보여준다.

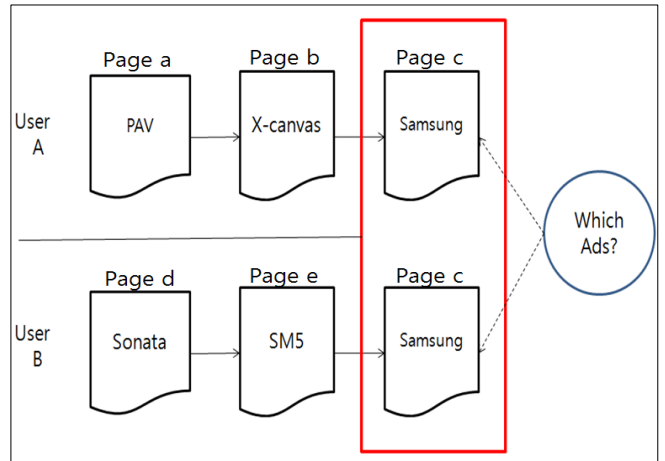
1. 서론

웹 검색 광고는 크게 스폰서 광고와 문맥 광고로 나눌 수 있다. 스폰서 광고는 키워드 검색을 기반으로 검색 키워드에 대해서 광고주가 미리 구매한 키워드 리스트와 매칭 되는 광고를 검색 결과에 보여 주는 방식을 일컫는다. 이에 반해 문맥 광고는 웹페이지의 내용을 분석하여 웹페이지의 내용과 연관된 광고를 보여주는 방식이다. 이미 많은 인터넷 서비스 업체들이 웹 광고를 통해서 막대한 수익을 올리고 있으며 그 비중은 점차 늘어나고 있다. 그러므로 사용자에게 필요한 광고를 웹 페이지에 부착하는 것은 웹 광고의 수익성을 올리기 위한 중요한 문제이다[1].

현재의 문맥 광고 시스템에서는 웹 페이지의 내용을 분석하여 이와 가장 연관도가 높은 광고를 제공하고 있다. 하지만 이러한 기존의 방법은 웹 페이지의 내용만을 이용하여 연관도가 높은 광고를 찾기 때문에, 모든 방문자에게 동일한 광고만 제공할 수 밖에 없다는 한계가 있다. 같은 페이지를 방문하더라도 방문자의 웹 페이지 방문 의도는 다를 수 있어 광고의 효과성을 높이기 위해서는 웹 페이지 방문자의 방문 의도에 맞게 개인화된 광고가 제공되어야 한다.

예를 들어, 그림 1과 같이 삼성과 관련된 웹 페이지(Page c)를 방문하기 전에 사용자 A는 PAV 텔레비전과 X-canvas 텔레비전에 관련한 웹 페이지를 각각 방문하였

고 사용자 B는 Sonata 자동차와 SM5 자동차에 관련한 웹 페이지를 방문하였다고 하자. 현재의 문맥 광고 시스템에서는 Page c의 내용만을 이용하여 광고를 제공하기



(그림 1) 사용자의 웹 페이지 방문 의도

때문에 사용자 A와 사용자 B에게 삼성과 관련한 동일한 광고만을 제공할 것이다. 하지만 Page c에 대하여 사용자 A에게는 삼성과 관련한 텔레비전 광고를 사용자 B에게는 삼성과 관련한 자동차 광고를 제공해 줄 수 있다면 사용자 방문 의도에 적합한 광고 일 것이다.

본 논문에서는 웹 페이지 방문자의 히스토리를 이용하여 사용자의 웹 페이지 방문 의도를 파악하고 사용자 의도에 적합한 개인화된 광고를 제공할 수 있는 기법을 제안한다.

2. 방문자 의도 분석 방법

본 논문에서 제안하는 기법을 활용하기 위해서는 사용자

* 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임 (No. 009-0077925)

웹 페이지 방문 히스토리를 분석하여야 한다. 기존의 개인화된 검색(Personalized Search) 연구에서는 웹 페이지 상의 Java Script를 이용하여 개인의 데스크 탑의 웹 브라우저로부터 방문 히스토리를 얻어 올 수 있다고 가정한다 [2]. 하지만 실제 사용자의 페이지 방문 데이터를 얻는 데에는 기존 논문에서 기술된 바와 같이 현실적으로 한계가 있다. 그러므로 본 논문에서는 개개인의 웹 페이지 방문 히스토리를 검색 엔진의 검색 로그의 세션 데이터로 가정하고 사용자의 의도를 파악하고자 하였다. 검색 로그에서 하나의 세션 로그 데이터 또한 그림 2와 같이 각각의 사용자의 웹페이지 방문 히스토리를 확인하는데 이용할 수 있기 때문이다. 본 논문에서는 동일한 세션 아이디를 가지는 로그 데이터는 동일한 사용자의 로그 데이터라고 간주한다.

Session ID	time	URL
0000d815a2154c4b	2006-05-27 오후 5:00:16	http://www.banjo.com/shopping/shopdisplayproducts.asp
0000d815a2154c4b	2006-05-27 오후 5:00:42	http://www.acutab.com/
0000d815a2154c4b	2006-05-27 오후 5:01:05	http://www.jacktuttle.com/banjo.html
0000d815a2154c4b	2006-05-27 오후 4:59:19	http://www.bluegrassbanjo.org/
0000daebae1649c0	2006-05-26 오전 8:32:59	http://forum.fish.com/
0000db2d603a4935	2006-05-30 오전 5:49:00	http://www.nwa.com/mall/
0000db2d603a4935	2006-05-30 오전 5:49:13	http://www.nwa.com/hk

(그림 2) MSN 검색 로그 데이터

본 논문에서 웹 페이지와 광고간의 유사도를 측정하여 매칭하는 기법은 기존 논문에서 제안하는 의미 매칭 방법을 이용하였다[3]. 기존 논문에서는 식 1에서와 같이 현재 페이지와 광고들 간의 매칭에 ODP와 같은 분류 트리와 로치오(Rocchio) 분류기[4]를 이용하고 있다.

$$Score(p_i, a_i) = \alpha \cdot TaxScore(Tax(p_i), Tax(a_i)) + (1 - \alpha) \cdot KeywordScore(p_i, a_i) \quad (1)$$

식 1에서 p는 광고를 붙이려는 페이지, a는 웹 광고를 의미한다. Tax(p)와 Tax(a)는 페이지와 광고가 로치오 분류기를 이용하여 분류 트리 내에서 분류된 결과 노드(주제)를 의미한다. 즉, 이 결과 노드는 각각의 페이지와 광고가 로치오 분류기를 이용해 분류 트리 상에서 정해진 주제(Topic)를 의미한다. 그러므로 TaxScore는 웹 페이지와 광고의 분류된 주제 간의 연관된 정도를 측정하는 점수를 의미한다. 여기서 주제 간의 연관도는 분류 트리 내에서 노드들 간의 거리를 이용하여 측정한다. KeywordScore는 웹페이지와 광고를 벡터 스페이스 모델에 적용하여 웹 페이지와 광고간의 코사인 유사도를 측정 한 값이다.

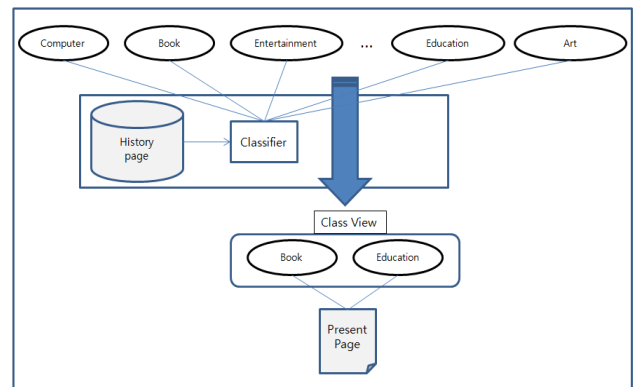
본 논문에서는 이러한 매칭 방법을 이용할 때 웹 페이지 방문 히스토리 데이터로부터 미리 사용자가 관심 있어 하는 주제를 파악하여 위와 같은 페이지와 광고의 매칭에

있어서 작은 클래스 뷰(주제 집합)를 만들어 주고자 한다. 클래스 뷰란, 방문 페이지의 주제를 분류기를 이용하여 파악하고 분류된 주제들로 하나의 작은 주제 집합을 구성한 것이다. 이러한 방식으로 미리 만들어 놓은 클래스 뷰를 이용하게 되면 페이지에 적합한 광고를 찾을 때 클래스 뷰에 속하는 광고에 대해서만 페이지와 매칭을 하기 때문에 사용자의 방문 의도에 부합하는 광고들을 찾아 낼 수 있다. 즉 식 (1)에서 매칭하는 광고의 개수 i의 크기를 줄이는 동시에 ODP의 트리 노드 중에서 유저의 의도와 부합하는 노드만을 사용한다.

본 논문의 제안 기법은 구체적으로 다음과 같다.

1) 클래스 뷰 생성

사용자의 웹페이지 방문 히스토리를 이용할 수 있다 하더라도 사용자의 히스토리가 모두 현재 웹 페이지(광고를 부착하려는 웹 페이지)와 관련 있는 웹 페이지라고 보장할 수 없다. 그러므로 이러한 각각의 세션마다 현재 웹 페이지와 연관된 방문 페이지를 찾을 필요가 있다. 본 논문에서는 연관된 방문 페이지를 찾는 과정에서 문서간의 유사도를 측정하는 방법 중 하나인 코사인 유사도(Cosine Similarity)를 이용하였다. 이러한 코사인 유사도는 0에서 1사이의 값을 가지게 되는데, 특정 기준 값(S)를 설정하여 기준 값 이상의 유사도를 가지는 방문 페이지를 연관된 페이지로 간주하였다. 즉, 현재 웹 페이지와 방문 페이지 간의 코사인 유사도가 기준 값 이상 되는 방문 페이지에 대해서 로치오 분류기를 이용하여 ODP 내의 새로운 클래스 뷰를 만들었다.

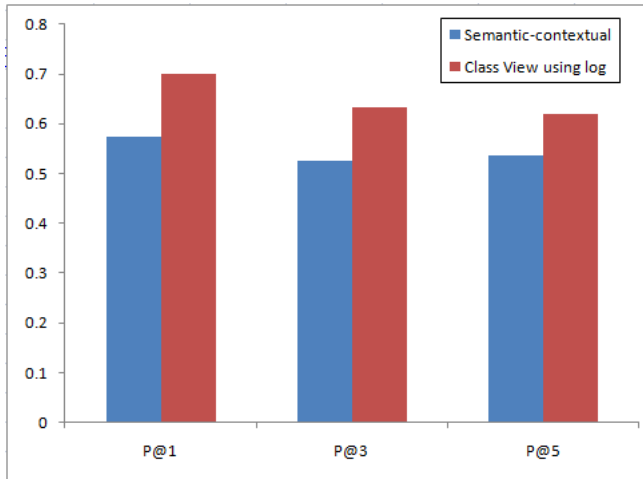


(그림 3) 클래스 뷰 적용 과정

2) 클래스 뷰를 적용한 광고 랭킹

위의 과정을 거쳐서 만들어진 클래스 뷰는 사용자의 방문 의도를 반영하는 주제들로 구성되게 된다. 그렇기 때문에 방문 페이지들로부터 만들어 놓은 클래스 뷰에 속하는 광고만 검색한다면 사용자의 방문 의도가 반영된 광고들을 랭킹 할 수 있다. 그림 3과 같이 분류 트리 상의 각각의 노드(Computer, Book,...)마다 연관된 많은 광고들이 존재 할 것이다. 각각의 노드들에 연관된 무수히 많은 광고들 모두와 현재 페이지를 매칭하기 보다는 사용자의 의도를 반영시켜 클래스 뷰(주제 집합)에 속하는(그림 3에서는 Book, Education) 광고들과 현재 페이지를 매칭 시킨다면

사용자마다 방문의도에 맞게 개인화된 광고를 제공할 수 있다. 클래스 뷰와 연관 있는 광고들과 현재 페이지와의 매칭에는 기존 논문[3]의 기법을 적용하였다. 사용자 방문 히스토리로부터 웹 페이지 방문 의도에 맞는 광고 주제를 미리 정하는 것만으로도 실험을 통하여 향상된 성능을 확인 할 수 있었다.



(그림 4) 상위 k개 정확도 실험 결과

3. 성능평가

본 연구에서는 사용자의 페이지 방문 히스토리 데이터를 검색 로그 세션으로부터 얻을 수 있다고 가정하였다. 실험에서 사용한 검색 로그 데이터는 2006년 5월 한 달 동안 MSN Live Search 검색 엔진을 이용한 사람들의 실제 검색 로그 데이터이다[5]. 이 검색 로그 데이터로부터 방문 페이지가 3개 이상인 세션들을 시간 순서대로 랜덤 샘플링하여 총 40개의 세션을 실험 데이터로 선정하였다. 실험에 사용한 광고는 구글 광고 검색으로부터 약 11,000개를 모아서 이용하였다.

실험은 각각의 세션들의 마지막 방문 페이지를 현재 페이지로 가정하고 이전 방문 페이지들을 사용자 히스토리 데이터로 이용하였다. 40개의 세션은 평균적으로 약 3.2개의 방문 히스토리를 가지고 있었으며 코사인 유사도를 통하여 평균 약 2.5개의 연관된 방문 페이지를 찾을 수 있었다. 연관된 페이지를 찾는 기준 값(S)은 실험을 통해서 S=0.3으로 설정하였다.

실험은 기존 논문의 의미 매칭 방법을 이용한 문맥 광고 랭킹과 본 논문에서 제안하는 히스토리 로그 데이터를 이용한 클래스 뷰 광고 랭킹의 성능 비교로 이루어졌다. 그림 4는 상위 k개의 정확도(Precision)를 평가한 것이다. 정확도 평가 방식은 기존 논문에서 사용했던 방식을 이용하였다[3]. 기존 논문에서는 페이지에 매칭되는 광고에 대하여 페이지와 광고의 의미적 유사성을 적합함, 조금 적합함, 적합하지 않음과 같이 3단계로 나누어 평가한다. 본 논문의 실험에서는 기존 논문과 같은 정확도 측정 방식(상위 k개 정확도 평가)을 이용하되, 적합성 판단 기준을

적합함, 적합하지 않음 2가지로 나누어 기존 논문과 본 논문의 제안 기법의 정확도를 평가하였다. 문맥 광고에서는 하나의 웹 페이지에 표시할 광고의 개수가 1-5개로 제한되기 때문에 광고 매칭에 있어서 상위 k개의 정확도는 매우 중요하다. 상위 한 개의 광고, 상위 세 개의 광고, 상위 다섯 개의 광고들에 대해서 각각 비교 하였을 때 평균 정확도는 그래프와 같이 로그를 이용한 제안 기법이 더 향상되었음을 확인할 수 있었다. 상위 한 개의 광고에 대해서는 기존 논문의 정확도에 비해 약 21%로 향상된 정확도를 얻을 수 있었다. 상위 세 번째, 다섯 번째까지의 광고에 대해서도 각각 20%, 15% 향상된 정확도를 얻을 수 있었다.

4. 결론 및 향후 연구

본 연구에서는 사용자들의 웹 페이지 방문 의도를 광고 매칭에 이용하였다. 사용자의 페이지 방문 의도를 파악하기 위해서 사용자의 웹 페이지 방문 히스토리를 이용하고 이를 활용하여 클래스 뷰(주제 집합) 생성을 통한 문맥 광고 매칭 기법을 제안하였다. 성능 평가를 통하여 기존의 기법 보다 최대 약 20% 이상 향상된 성능을 확인할 수 있었다.

향후 연구 방향은 사용자 히스토리 데이터를 효과적으로 이용하기 위한 세션화에 대한 연구를 진행할 예정이다. 또한 문맥광고에서 로그 데이터를 다양하게 활용한 기법들에 대해서 연구하고자 한다.

참고문헌

- [1] C. Wang, P. Zhang, R. Choi, and M. D'Eredita, "Understanding consumers attitude toward advertising", In Eighth Americas Conference on Information System, pp. 1443-1148, 2002.
- [2] Micarelli, A., Gasparetti, F., Sciarrone, F., and Gauch S.: Personalized Search on the World Wide Web. In Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York (2007) this volume
- [3] A. Z. Broader, M. Fontoura, V. Josifovski, and L. Riedel, "A semantic approach to contextual advertising", In Proceedings of Association for Computing Machinery SIGIR, 559-566, 2007.
- [4] J. Rocchio, "Relevance feedback in information retrieval", In The SMART Retrieval System-Experiments in Automatic Document Processing, pp. 313-323, 1971.
- [5] Microsoft Live Labs. Accelerating search in academic research, 2006.
<http://research.microsoft.com/ur/us/fundingopps/RFPs/Se arch 2006 RFP.aspx>.