부정 클릭의 식별 방법을 높이는 방법에 대한 연구

A Study for the Enhancement Method of Fraud Click Identification

홍영란(YoungRan Hong)*, 김동수(Dongsoo Kim)**
yrhong@somansa.com, dskim@ssu.ac.kr

초 록

오버추어 광고를 진행할 때 늘 제기되는 부정클릭 문제는 광고주가 인정할 수 없는 광고비용 청구와 과금이라는 문제를 놓고 늘 제기되어온 것이다. 기존에 부정 클릭을 식별하기 위해 제시된 대표적인 모델은 검색엔진에서 검색한 키워드, 광고를 클릭한 시간, 사이트를 실제 방문한 시간, IP를 기준으로 한다. 이 기준을 가지고 광고를 클릭한 특정 IP가 연속적으로 같은 클릭을 한 경우, 이를 부정 클릭 가능성이 있다고 판단하는 것이다. 그러나 이 방법은 클라이언트 IP를 기반으로 하고 사람이 부정 클릭을 행하는 것을 식별하는 것이기 때문에 자동화 툴 등을 이용하여 부정클릭을 행하는 방법에 대한 식별 방법으로는 부족하다고 할 수 있다.

본 논문에서는 자동화 툴을 이용한 부정 클릭 식별방법을 높이기 위해 세션 정보를 핑거프린팅하는 방법을 제안한다. 제안 모델은 동일한 패턴의 행위에 대한 세션 정보를 핑거프린팅하고, 로그 정보 중에서 동일한 핑거프린트 값을 발견하면 클라이언트 IP와 매칭시킨다. 동일한 IP에서 나온 동일한 핑거프린트 값은 자동화 툴을 이용한 자동 부정 클릭의 판단 기준이 될 수 있다. 이 연구에서는 자동화 툴이 일정한 패턴을 가지고 3단계의 클릭을 하루 3회 실시하도록 하였다. 본 연구방법론은 보다 정교한 형태의 부정 클릭 식별 모델을 만들기 위해 기존의 IP 정보 이외에 세션 정보 등 여러 정보를 이용할 수 있음을 보여주는데 의의를 가진다.

1. 서론

현재 한국의 검색광고 시장에서는 검색 키워드 입력 후 나타나는 검색결과 업체 중에서 소비자가 클릭하는 수에 따라 광고요금을 부과하는 CPC(cost per click:이하 CPC) 방식을 채택하고 있다. 그러나 CPC 방식은 클릭 후의 소비자 행동을 전혀 고려하지 않고 단순히 클릭 수에 따라 과금하는 방식이어서 광고주들의 불만이 매우 높다. 현재의 CPC 방식은 검색어 단가의 결정과정에서 입찰가가 공개적으로 제시되어 과잉경쟁을 유발할 뿐만 아니라 차(次)순위 입찰가 기준 낙찰 원칙을 바탕으로 하고 있어서 경매방식이 지니는 장점을 약화시키고 있다. 그리고 CPC 방식은 무엇보다도 부정클릭에 매우 취약하다는 문제점을 지니고 있다. 실제로 검색광고 광고주들이 가장 많이 경험하고 있는 부정 클릭 사례는 허위클릭과 무효 클릭으로 나타나고 있다[2][4].

부정 클릭은 인터넷 상에서 경쟁업체 사

이트를 집중적으로 클릭하여 경쟁업체가 포털에서 검색되지 않도록 하거나 검색광고의 경쟁을 유발시켜 많은 광고비를 발생시켜 손실을 가져오는 문제점을 야기시킨다.

본 연구에서는 로그의 세션 정보를 핑거프린팅하고, 이 후 유입되는 로그 정보에 동일한 값이 나타날 경우, 이를 부정 클릭의 판단 기준으로 삼을 것을 제안한다. 방문 기록 DB의 세션 정보를 분석하여 동일한 패턴의 클릭 정보와 동일한 시간 간격으로 세션 정보가 처리될 경우, 이를 핑거프린트 하여 부정 클릭의 식별 방법을 높일 수 있다.

2. 관련연구

2.1 관련연구 1

현재 한국의 검색광고 시장에서 사용하고 있는 CPC 무엇보다도 부정 클릭에 매우 취약하다는 문제점을 지니고 있다. 실제로 검색광고 광고주들이 가장 많이 경험하고 있는 부정 클릭 사례는 허위 클릭과 무효 클릭으로 나타나고 있다[2][4].

따라서 CPC 방식의 키워드 검색 광고 및 그 안에서 발생하는 부정 행위에 대한 대응 현황의 검토를 바탕으로 하여 CPC 방식의 키워드 검색 광고에서 발생하는 부정행위 방지 및 분쟁 해결에 대한 접근 방법을 제안하고 특히 '클릭'이라는 가장 기본적인 행동을 유효/무효, 사기의 고의성, 수행 방법 (자동/수동) 등을 기준으로 하여 분류하는 방법이 계속 연구되고 있다[5][1].

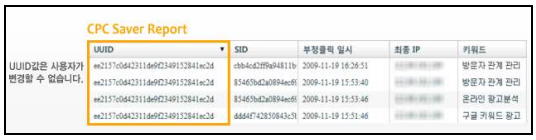
부정 클릭은 인터넷 광고 비용과 밀접한 관련을 갖는다. 다음 <그림1>은 일반적인 CPC의 과금 체계를 다섯 단계로 나누어 보여주고 있다.

<그림 1> CPC 과금 체계

<그림 1>에서 보는 것과 같이 일반적인 CPC 과금 체계는 다음 다섯 가지 단계로 구별된다. 첫 번째는 검색 엔진의 키워드를 통하여 부정 클릭을 발생시키는 단계이다. 두 번째 단계는 사용자 정의를 통한 부정 클릭 판단 기준을 통해 부정 클릭의 발생을 인지하는 단계이다. 세 번째 단계에서는 대부분 광고 관리 등 각 부정 클릭 방지 업체들의 솔루션이 적용되는 단계이다. 네 번째 단계는 설정에 따라 랜딩 페이지를 별도 지정 페이지로 변경하여 부정 클릭을 제지하거나 경고하는 단계이다. 다섯 번째는 마지막 단계로서 검색 엔진 광고의 클릭이 발생할 때 오버추어와 각 포털 별 과금을 실제 업체의 클릭 수와 비교하여 검색 광고의 과금을 하는 단계이다. 이 다섯 단계로 구성되는 부정 클릭 분석 방법은 대부분 로그 분석을 통한 IP 추적이 그 중심 기술이 된다[3][4]. 이 방법은 IP변환 시스템의 적발과 동일 IP가 지속적으로 특정 사이트를 클릭하거나 검색할 때 이를 부정 클릭으로 간주하여 해당 IP를 막는 형식을 취한다.

2.2 관련연구 2

대표적인 부정 클릭 방지 시스템은 동일한 IP로 일정 시간동안 일정횟수 이상 클릭한 행위를 부정 클릭으로 간주하는 방법과 IP 변경 프로그램을 추적하여 부정 클릭으로 간주하는 방법이 있다. 다음은 동일 IP를 추출하여 부정 클릭으로 간주하는 부정 클릭 방지 시스템의 예이다[7].



<그림2> 동일 IP접속을 부정 클릭으로 간주하는 시스템의 경우

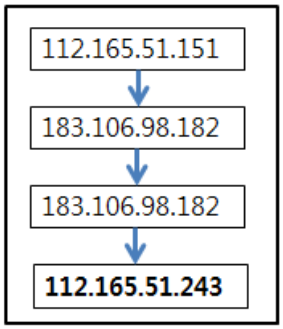
위의 <그림2>에서 보는 것처럼 부정 클릭을 적발하기 위하여 특정 시간대에 동일한 IP가 여러 번 같은 상품 관련 키워드를 검색하거나 특정 사이트의 특정 광고를 클릭한 경우, 처음 방문 정보 중 IP와 키워드를 UUID로 만들고 이후 방문 정보를 계속 UUID로 만든 후, 같은 UUID를 가진 경우를 부정 클릭으로 간주하는 방법이다.

그 다음으로는 IP 변경 프로그램을 추적하여 IP를 변화시켜 클릭하는 움직임을 부정 클릭으로 간주하는 방법을 들 수 있다. <그림3>은 IP 변경 시스템을 이용하여 원래 IP를 추출해냄으로써 부정 클릭을 식별하는 부정 클릭 방지 시스템의 예이다 [8].

IP	키워드	검색엔진	클릭일시
112.165.51.243(11) URL 정보	쿡인터넷	naver.com	2010-09-29 12:30:56
112.165.51.151(10) URL 정보	쿡	naver.com	2010-09-29 11:28:52
112.165.51.151(9) URL 정보	쿡	naver.com	2010-09-29 11:25:02
112.165.51.151(9) URL 정보	쿡	naver.com	2010-09-29 11:12:01
112.165.51.151(8) URL 정보	qook	naver.com	2010-09-29 10:40:23
112.165.51.151(7) URL 정보	쿡	naver.com	2010-09-29 09:27:15
112.165.51.151(6) URL 정보	쿡	naver.com	2010-09-29 08:25:02
112.165.51.151(5) URL 정보 → 쿠키제거후 접속	쿡 가입	naver.com	2010-09-29 08:15:01
112.165.51.151(4) URL 정보 → 쿠키제거후 접속	kt	naver.com	2010-09-29 08:12:23
112.165.51.194(3) URL 정보	qook가입	naver.com	2010-09-29 08:30:01
112.165.51.194(2) URL 정보	qook가입	naver.com	2010-09-28 07:28:15
112.165.51.194(1) URL 정보	qook가입	naver.com	2010-09-28 06:55:32

<그림3> 네이버 클릭 초이스 접속 목록으로 본 IP 정보

IP는 여러 가지로 들어오지만 이것에 대해 IP 변경 추적 프로그램을 돌려보면 아래 <그림4>와 같이 IP가 변경된 IP를 추적해냄으로써 동일 IP가 자동 변경 프로그램을 사용했음을 알 수 있다.



<그림4> IP 추적 프로그램으로 분석한 IP 변경상태

그러나 기존의 연구 방법은 IP 정보만을 기반으로 하기 때문에 자동화 툴을 이용하여 부정 클릭을 실시한 경우에는 식별에 한계점을 갖는다. 여러 개의 가상 머신(VMWare)에 여러 개의 IP를 등록한 후, 자동화 툴로 부정 클릭을 순서를 바꾸어 가면서 실시하게 되면 IP 기반의 부정 클릭 식별 방법은 식별 능력이 떨어질 수 밖에 없다. 본 연구에서 제안하는 세션 정보를 이

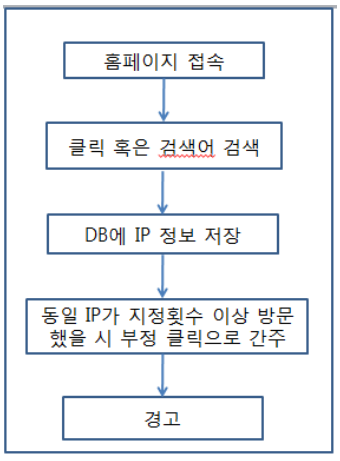
* 숭실대학교 산업정보시스템공학과 박사과정
** 숭실대학교 산업정보시스템공학과 교수

용한 핑거프린트 기법은 동일한 정보를 가진 로그의 핑거 프린트 값이 같다는 점을 이용한다. 데이터 관리에 있어서 이 핑거프린트를 사용할 경우, 필드 해쉬 테이블을 이용하여 다차원 데이터의 저장될 레코드 순서를 빠르게 찾아 저장함으로서 데이터 생성 속도가 향상된다. 또한 해쉬 테이블만을 유지하면 되므로 메모리 사용량이 감소한다. 따라서 해쉬 테이블의 사용으로 데이터의 빠른 검색과 데이터 생성 요청에 빠른 응답이 가능하다는 장점을 가지고 있다[1].

3. 본론

3.1 리서치 프레임워크

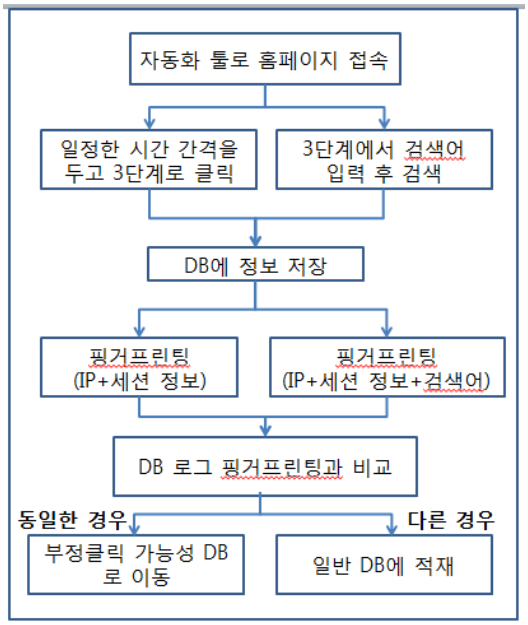
최근 자동화 툴을 이용한 부정 클릭이 늘어나고 있기 때문에 본 연구에서는 자동화 툴을 이용한 부정 클릭 식별을 높이는 방법을 제안한다. 기존에 IP를 이용한 부정 클릭 연구 모델은 다음 <그림5>와 같은 프레임워크를 갖는다.



<그림 5>IP 기반의 부정클릭 식별 프레임워크

<그림5>에서 보는 것처럼 IP를 기반으로 하는 부정 클릭 식별 모델은 변수를 IP 하나로 지정하고, IP 비교 방법을 사용한다. 이 방법은 자동화 툴에 의한 부정 클릭 등 여러 가지 변수를 갖는 다른 형태의 부정 클릭에 대한 변별력을 높이기 쉽지 않다. 본 연구에서는 부정 클릭 식별을 위해 IP 이외에 세션 정보라는 변수를 추가할 것을 제안한다. 또한 기존의 IP 단순 비교 이외에 핑거프린팅 기법을 사용하여 고유 값을 비교할 것을 제안한다.

다음 <그림6>은 추가된 변수와 비교 방법을 적용한 자동 부정 클릭 식별 방법에 대한 전체적인 리서치 프레임워크이다.



<그림 6> 제안하는 리서치 프레임워크

본 연구는 Test Complete라는 자동화 툴을 사용하여 특정 웹 페이지에 자동으로 접속하여 일정시간의 세션을 유지하게 하면서 3단계까지 클릭을 유도하였다. 실험은 하루에 3회씩 총 7일간 배치 작업을 통해 실험을 실시했다. 이어지는 3.2와 3.3에서 위 <

그림6>의 리서치 프레임워크와 실험 내용을 설명한다.

3.2 세션 정보를 이용한 핑거프린트로 부정 클릭 식별 강화

본 연구에서는 변수에 IP이외에 세션 정보를 추가함으로써 자동화 툴을 이용한 부정 클릭의 식별을 강화하는 방법을 제안한다. 자동화 툴은 그 특성상 매크로를 이용하기 때문에 다음 단계로 이동하면서 Depth있게 클릭을 할 경우, 반드시 일정한 시간 간격으로 동일한 위치 정보를 클릭하는 성격을 가진다. 따라서 부정 클릭의 식별 방법을 높이는 1단계로서 한 단계에서 다음 단계로 넘어가면서 클릭을 할 때 대기하는 정보인 세션 정보를 부정 클릭 식별의 중요한 기준이 될 수 있다. 특정 웹 페이지에 자동으로 접속하여 일정시간의 세션을 유지하게 하면서 3단계까지 클릭을 유도하였다. 첫 번째로는 매크로를 이용하여 일정한 패턴으로 웹 페이지에 접속하게 하였다. 단계별 클릭 정보는 <표 1>과 같다.

<표 1> 접속 단계

구분	1 단계	2 단계	3 단계
Depth	홈페이지 접속	링크	국내보안사이트/국외보안사이트
		소개	고객/연락처/연혁/제휴사/회사소개
		검색	디비아이
		소만사 홈	
		제품	메일아이/웹키퍼

위의 <표1>에서 보는 것처럼 3단계에 걸쳐 TestComplete로 사용하여 지속적으로 클릭을 시도하였다. 이 때 1~3단계까지의 세션 유지 시각은 다음과 같다.

<표 2> 접속 단계 세션 유지 시간

구분	1 단계	2 단계	3 단계
세션 유지 시간	5 초	3 초	10 초

처음 클릭하여 페이지에 머무는 시간을 5초로 하고, 다음 단계로 이동하여 머무는 시간을 3초, 마지막인 세번째 단계에서는 페이지에 머무는 시간을 10초로 가정하였다. 이는 일반적으로 사람들이 웹 페이지에 접속을 하여 Depth 있게 클릭을 하는 행위 패턴을 참고한다. <표2>의 세션 유지 시간은 세션 딜레이 시간을 포함하며, 이 시간이 지나도 접속이 되지 않는 경우는 Tool에서 접속을 끊도록 유도했다. 따라서 <그림7>와 같이 DB에 다음과 같은 테이블이 생성된다.

번호	Depth1	Depth2	Depth3	내부경로 URL	클라이언트 IP
1	소만사			www.somansa.com	192.168.1.11
2		링크	국내보안사이	www.somansa.com/korean/link/L	192.168.1.11
3			국외보안사이	www.somansa.com/korean/link/L	192.168.1.11
4		사이트맵		www.somansa.com/korean/link/s	192.168.1.11
5		소개	고객	www.somansa.com/korean/abou	192.168.1.11
6			연락처	www.somansa.com/korean/abou	192.168.1.11
7			연혁	www.somansa.com/korean/abou	192.168.1.11
8			제휴사	www.somansa.com/korean/abou	192.168.1.11
9			회사소개	www.somansa.com/korean/abou	192.168.1.11
10		소만사 홈		www.somansa.com/korean/index	192.168.1.11
11		제품	메일아이	www.somansa.com/korean/prod	192.168.1.11
12			웹키퍼	www.somansa.com/korean/prod	192.168.1.11
13		검색	디비아이	www.somansa.com/korean/searc	192.168.1.11

<그림7> 단계별 DB 테이블

DB 로그를 보면 각 세션 별로 방문 경로는 하나이므로 VisitList table에 데이터가 적재되게 되므로 테이블 단위로 핑거프린팅을 하여 동일한 결과값을 비교할 수 있다.

7일간의 실험 결과는 다음과 같다. 세 개의 단계로 구성된 자동화 툴의 자동 클릭 행위는 모두 DB에 저장되었다. 저장된 정보 중 방문 시각을 제외한 세 단계의 세션 정보만을 핑거프린팅 한 결과, 자동화 툴이 발생시킨 세션 정보 핑거 프린트 값은 모두 같았다. 이를 바탕으로 로그 분석 툴을 이용하여 같은 핑거프린트 값을 갖는 로그의 IP값을 추출하여 비교해준 결과, 모두 TestComplete을 이용하여 클릭을 발생시킨 IP였음을 확인하였다.

5. 결론

본 논문에서는 부정 클릭의 식별 능력을 향상시킬 수 있는 하나의 방법론을 제안하고 이에 대한 간단한 실험을 통해 모델의 유효성에 대해 검증해 보았다.

본 연구는 실제로 다양한 대형 포털의 사이트에 적용되어 실제적인 유효성을 검증하지 못했다는 한계를 갖는다. 그러나 본 논문에서 제안한 것처럼 부정 클릭을 식별해 내는 변수를 IP로만 종속시키지 않고, 세션 정보를 이용한 핑거프린팅 기법을 사용할 경우, 부정 클릭의 식별 방법은 분명 한층 더 발전할 것으로 기대된다. 또한 이후 일정한 패턴으로 움직이는 정보를 DB에 계속 쌓고 대상 데이터를 IP, 클릭한 대상 광고, 세션 정보뿐만 아니라 다른 여러 종류의 정보까지 확대시켜 이를 조합하여 상호 비교하는 정교화된 학습 효과를 가진 엔진으로 발전시킴으로써 지속적으로 부정 클릭 식별 정교화 모델에 대한 연구를 지속할 예정이다.

참고문헌

[1] 김형선, 유병섭, 이재동, 배해영, ‘데이터웨어하우스에서 해쉬 테이블을 이용한 효율적인 데이터 큐브 생성 기법’, 추계학술발표회 32권 2호, 한국정보과학회, 2005.

[2] 오창우, ‘인터넷 검색광고 요금체계의 특징 및 부정 클릭 유형에 관한 연구’, 광고학연구, 제19권, 4호, P7, 한국광고학회, 2008.

[3] 이경전, 이현석, 전정호, ‘CPC 방식의 키워드 검색광고에서의 사기 클릭의 정의와 대응방안 평가’, 한국경영정보학회, P111. 2008.

[4] 방송통신 위원회, 한국전파진흥원,’방송통신 융합환경에서의 시스템적인 광고유통방안에 관한 연구:인터넷 광고를 중심으로’, p.172, 방송통신 위원회, 2009.

[5] 한국 인터넷 진흥원, ‘인터넷 광고 관련 국내.외 법.제도 동향 조사 분석 위탁용역’, 한국인터넷 진흥원, 2009.

[6] 박대윤, 최현주, 네모 도리, 최유진 외, ‘D.I.Y 액세서리 쇼핑몰 절대로 하지 마라”. 정보문화사 2007.

[7] www.cpcgaurd.com

[8] www.logger.co.kr