

An Effective Method for Blocking Illegal Sports Gambling Ads on Social Media

Ji-A Kim*, Geum-Boon Lee**

*Teacher, Seokang High School, Gwangju, Korea

**Professor, SW Convergence Education Institute, Chosun University, Gwangju, Korea

[Abstract]

In this paper, we propose an effective method to block illegal gambling advertisement on social media. With the increase of smartphone and internet usage, users can easily access various information while sharing information such as text and video with a large number of others. In addition, illegal sports gambling advertisements are also continue to be transmitted on SNS. To avoid most surveillance networks, users are easily exposed to illegal sports gambling advertisement images by including phrases in the images that indicate illegal sports gambling advertisements. In order to cope with these problems, we proposed a method to actively block illegal sports gambling advertisements in a way different from the conventional passive methods. In this paper, we select words frequently used for illegal sports gambling, classifies them into three groups according to their importance, calculate WF for each word using weighted formula by degree of relevance and frequency, and then sum the WF of the words in the image. Blocking, warning, and passing were determined by cv , the total of WF . Experimenting with the proposed method, 193 out of 200 experimental images were correctly judged with 96.5% accuracy, and even though 7 images were illegal sports gambling advertisements. Further research is needed to block 3.5% of illegal sports betting ads that cannot be blocked in the future.

▶ **Key words:** Illegal sports gambling ads, Word Frequency, word Weighting, Blocking method, Three Groups

[요 약]

본 논문은 SNS 상의 불법 스포츠 도박 광고를 차단하기 위한 효과적인 방법을 제안한다. 스마트폰 사용과 인터넷 이용률의 증가로 사용자가 글이나 영상 등의 정보를 다수의 사용자들과 공유하면서 다양한 정보를 쉽게 접할 수 있게 됨에 따라 불법 스포츠 도박 광고 또한 SNS 내에서 지속적으로 전송되고 있다. 네트워크상의 감시를 피하기 위해 이미지 속에 불법 스포츠 도박 광고임을 나타내는 문구들을 포함시켜서 광고를 함으로써 이용자들은 불법 스포츠 도박 광고 이미지에 쉽게 노출되어 있다. 이러한 문제에 대응하기 위해 기존의 소극적인 방식과는 다른 방법으로 불법 스포츠 도박 광고를 적극적으로 차단할 수 있는 방법을 제안한다. 본 논문에서 제안하는 방법은 불법 스포츠 도박에 자주 사용되는 단어들을 선정하고 중요도에 따라 3개의 그룹으로 나누고 그룹 가중치와 빈도 별 가중치 공식을 이용하여 각 단어마다 WF 를 산정한 후, 해당 이미지 내의 단어들의 WF 의 총합인 cv 로 차단, 경고, 통과를 판정하였다. 제안한 방법으로 실험을 한 결과 200개의 실험 이미지 중 193개가 정확한 판정을 받아 96.5%의 정확도를 보였으며, 7개의 이미지가 불법 스포츠 도박 광고임에도 통과 판정을 받았다. 향후 차단되지 못한 3.5%의 불법 스포츠 도박 광고를 차단시키기 위한 연구가 필요하다 하겠다.

▶ **주제어 :** 불법 스포츠 도박 광고, 단어 빈도, 그룹 가중치, 차단 방법, 3 그룹

- First Author: Ji-A Kim, Corresponding Author: Geum-Boon Lee
- *Ji-A Kim (kimj1213@naver.com), Seokang High School
- **Geum-Boon Lee (goldpalm@chosun.ac.kr), SW Convergence Education Institute, Chosun University
- Received: 2019. 12. 04, Revised: 2019. 12. 24, Accepted: 2019. 12. 26.

I. Introduction

스마트폰의 보급률이 높아지면서 동시에 인터넷 이용률도 크게 증가하였다. 우리나라 1,952만 가구 중 1,943만 가구(99.5%)가 인터넷 접속이 가능하게 되었으며, 스마트폰(94.1%), 데스크탑(61.2%), 그리고 노트북(31.6%) 등을 인터넷 접속기기로 사용하고 있다. 사람들은 인터넷에 접속하여 메신저(95.1%), SNS(68.2%), 게임(57.1%), 동영상(77.5%), 신문, 잡지 읽기(89.6%) 등의 활동을 하는데, 그 중에서도 청소년은 스마트폰을 이용하여 메신저(99.9%), SNS(92.8%), 인터넷쇼핑(91.5%) 등의 활동을 하는 것으로 나타났다[1]. 사용자가 텍스트나 영상 등의 정보를 다수의 사용자와 공유하면서 타인과 네트워크를 형성하는 서비스인 SNS는 다양한 정보를 쉽게 얻을 수 있는 반면 원치 않는 정보도 쉽게 접할 수 있는 문제가 있다. 원하지 않는 정보에는 청소년에게 유해가 될 수 있는 매체가 다수 포함되어 있다. 자극적인 정보로 선정적인 사이트를 홍보하는 광고물과 주로 불법 스포츠 도박 사이트를 홍보하는 광고물들이 SNS의 특성을 이용해 불법적으로 전송 된다는 점이다. 청소년 도박문제 실태조사 결과, 재학 중 청소년의 6.4%가 도박문제 위험집단으로 나타났다. 그 중에서위험군이 4.9%, 문제군이 1.5%로 매년 증가되고 있는 실정이다. 특히 온라인 돈내기 게임을 경험한 경우 PC(25.8%)보다는 스마트폰(74.2%)을 통해 접속했다는 응답이 3배 가량 높아 스마트폰이 청소년의 돈내기 게임 이용 접근성을 증가시키고 있음을 보여주고 있다[1-3]. 이러한 문제 상황을 인식하고 불법 스포츠 도박 광고를 저지하려는 시도가 SNS 내에서도 이뤄지고 있다. 불법 광고 매체 발견 시 신고를 부탁하는 문구나, 커뮤니티 규정을 보여주는 형태의 대응으로 이용자들이 불법 스포츠 도박 광고에 대처할 수 있도록 하고 있다. 그러나 접수된 신고에 제한하여 문제를 해결하거나, 관리자가 상시 모니터링 함으로써 방대한 양의 정보를 확인해야 하는 소극적인 대응이 현실이다. 반면에 불법 광고업자들은 불법이라는 위험성과 메신저의 편의성 때문에 대부분 따로 홈페이지를 제작하지 않고 다양한 인터넷 서비스를 이용해 불법 스포츠 도박을 조장하고 있다. 자신의 홈페이지 보다 기존의 페이스북과 같은 SNS를 활용함으로써 유해사이트로 차단되지 않고, 특정 단어에 의해 제재될 수 있는 텍스트 전송보다 홍보 문구를 삽입해서 이미지화한 다음 그 영상을 댓글로 올리거나 전송하는 방식을 선호한다.

따라서 이러한 불법 스포츠 도박 광고의 특징을 보았을 때 모니터링하는 기존의 방식보다 새로운 불법 스포츠 도박을 원천적으로 차단할 수 있는 방법의 필요성이 있다. 본 논문에서는 SNS에 전송되는 불법 광고에 삽입된 문구

를 분석하여 불법 스포츠 도박 광고에 사용된 특정 단어를 수집하고, 사용 빈도에 따라 3개의 그룹으로 분류한 다음 각 그룹마다 제안하는 가중치 공식을 적용하여 임계값을 기반으로 차단 여부를 판단하는 방법을 제안한다. 제 2장에서는 관련 연구에 대해 기술하고, 제3장에서는 단어를 수집하고 분류하는 기준에 대해 설명하고 그룹별 가중치 공식을 제시하며, 제4장에서는 제안한 방법을 이용해 실험한 결과를 보여주며, 제5장에서는 결론을 맺는다.

II. Related Works

1. Research of Illegal Sports

정보통신망을 통하여 사람들이 사행행위를 할 수 있게 서비스를 제공하는 행위는 불법사행행위에 속한다. 불법 스포츠 도박을 이용하는 사람들은 배팅 금액에 제한이 없고, 보다 다양한 운동경기에 대해 배팅이 가능하며 그리고 공익사업으로 사용되는 돈이 없기 때문에 정부 허가 사업에서의 배당액보다 높다는 점 때문에 높은 기대치를 갖고 불법 스포츠 도박을 더 많이 이용하고 있다[4]. 또한 불법 스포츠 도박은 해외 서버 운영 등의 이유로 국가의 통제가 어려운 환경에서 불특정 다수가 높은 사행성과 중독성으로 금전적 피해와 심리적 제반 문제에 노출될 수 있다는 점에서 문제가 되고 있다.[5]. 최근에는 불법 스포츠 도박을 홍보하기 위해 다양한 유인책을 사용하는데 카페, 인터넷 방송, 블로그, SNS 등을 이용해 참여자를 모집하거나 문자 메시지를 이용해 불특정 다수의 참여를 유도하고 기존에 불법 스포츠 도박을 하던 이용자가 지인에게 홍보를 하는 방법도 존재한다[6]. 특히 컴퓨터만 있으면 누구나 쉽게 접근이 가능한데다 가입 절차가 간단해서 참여가 지속적으로 늘고 있는 불법 스포츠 도박은 수사기관의 단속망을 피하기 위해 해외 서버를 운영하기도 하는데 지속적인 홍보를 막기가 어려워 최근에는 10대 청소년들 사이에서까지 유행처럼 번지고 있다[7].

불법 온라인 도박 사이트를 포털 사이트에 검색하도록 하였을 때 네이버를 통해 114개, 구글 112개, 다음 10개, 네이버 5개의 결과가 검색되었고, 그 밖에도 SNS를 통해 사이트에 대한 홍보가 이루어지는 경우가 있었다. 특히, 지인과의 관계성에 기초한 홍보가 이루어지는 페이스북에서는 거의 대부분 불법 스포츠 도박 광고였고[8], 불법 스포츠 도박을 이용하는 청소년들이 대부분 고등학교 2학년 때 처음으로 불법 스포츠 도박을 접하였고, 주로 인터넷, 휴대폰 어플을 이용해서 불법 스포츠 도박을 이용하며 대부분 본인 확인 인증 절차 없이 불법 스포츠 도박을 이용하고 있다는 연구를 발표했다[6].

2. Harmful Contents Blocking Methods

불법 스포츠 도박 광고를 포함한 유해 콘텐츠 차단을 위한 노력들이 있어 왔다. 메타 정보 및 텍스트 기반 차단 기술은 제목, 제작 정보, IP 주소, 또는 멀티미디어의 부가 설명에 포함된 텍스트 정보를 기반으로 하여 자주 등장하는 특정 키워드를 이용하여 유해 멀티미디어를 찾아내어 차단하는 방법이다. 해쉬 및 DB 기반 차단 기술은 유해 멀티미디어에 대한 해쉬값을 계산, 그 값을 DB화하여 유해 멀티미디어 차단에 활용하는 방법이다.[9]. 또한 안드로이드의 유해 어플리케이션에 대해 안드로이드의 로컬 프록시를 이용해 웹사이트의 접속과 차단되지 않은 웹 페이지의 유해 콘텐츠 차단 방법이 있으며 [10], 웹사이트들이 하이퍼링크로 연결 관계를 형성한다는 특성을 이용하여 유해사이트를 판별하는 방법이다. 기존 유해사이트에 관한 웹 페이지를 수집하고 URL을 수집한다. 추출된 정보를 이용하여 유방향 그래프를 생성, 그래프에 존재하는 순환을 찾아 낸 뒤에 이 순환에 포함되는 웹사이트는 유해사이트로 판별하는 방법이다[11]. 구글 검색을 통해 개인 정보 판매 게시글을 차단하는 기법도 제시되었는데 먼저 판매 게시글에 대한 단어 형태를 분석하고, Google Custom Search API를 이용하여 자주 사용하는 단어가 DB에 저장되도록 설계한 뒤, 검색한 결과를 바탕으로 차단하는 방법이다[12].

위와 같이 유해 콘텐츠를 차단하기 위한 많은 연구들이 계속되고 있지만 대부분 유해 콘텐츠 중에서도 선정적인 콘텐츠에 집중하고 있었고 또 어플리케이션이나 웹사이트를 차단하는 방법에 초점을 둔 연구가 많으며, 텍스트 추출 중심의 연구로 이미지화 된 텍스트를 분석하는 방법은 아직 제시되고 않고 있다.

3. Weighting Techniques

가중치를 부여하는 방법은 해결하고자 하는 문제만큼 다양하다. 표 1은 다양하게 고안된 가중치 공식 및 특징을 보여준다[13].

Table 1. Word Frequency Weight

Weight Formulas	Description
$1(\text{if } tf > 0), 0$	1 if a word appeared, 0 otherwise
tf	equation applying word frequency as it is
$1 + \log(tf)$	equation to reduce the difference in word influence using log
$1 + \log(1 + \log(tf))$	equation applying log of log
\sqrt{tf}	equation applying root
$(1-w) + w \times \frac{tf}{\max - tf}$	equation limiting the range of weights (min, max)
$\frac{tf+3}{4}$	equation to make root and weight gradients similar

그리고 추출된 단어의 가중치를 결정하기 위해 경험적인 방법으로 가중치를 부여하는 방식도 있으며, 스팸 메일 필터링에서와 같이 확률적으로 계산하는 방법도 있다[14-16].

III. The Proposed Method for Blocking Illegal Gambling Ads

청소년에게 유해한 콘텐츠나 유해 매체물을 차단하기 위한 방법은 끊임없이 연구되고 있다. 기존의 연구는 음란 유해사이트와 관련된 연구가 대부분이고 주로 웹 사이트를 차단하거나 어플리케이션을 차단하는 방식을 이용했다. 그러나 불법 스포츠 도박을 홍보하는 방식이 텍스트 방식에서 영상 속에 문구를 삽입해 홍보하거나 동영상 속에 문구를 삽입해 홍보하는 방식으로 발전해 감에 따라 이에 대한 차단방법이 미비한 것이 사실이다.

최근에는 SNS를 통해서 불법 스포츠 도박 광고 횟수가 빈번해지고 있다. SNS 상에서 홍보하기 위해 이미지 속에 홍보하고자 하는 문구를 삽입하여 영상으로 게재하는 방법을 사용한다. 또한 모든 정보를 한 눈에 확인할 수 있는 타임라인에서 확인 할 수 있는 게시글에 홍보 이미지를 게재하지 않고 인기 있는 게시글의 코멘트에 홍보 이미지를 게재하는 교묘한 방법을 주로 사용하고 있다. 따라서 눈에 쉽게 띄지 않아 감시가 어려워지고, 특정 단어를 이용해 필터링을 하는 것도 힘들어 진다.

따라서 본 논문에서는 불법 스포츠 도박임을 광고하는 문구가 삽입된 이미지를 사전에 처리하는 빈도별 가중치 공식을 이용하여 불법 스포츠 도박 광고 차단 방법을 제안하고자 한다. 그림 1은 제안하는 차단 시스템을 보여준다.

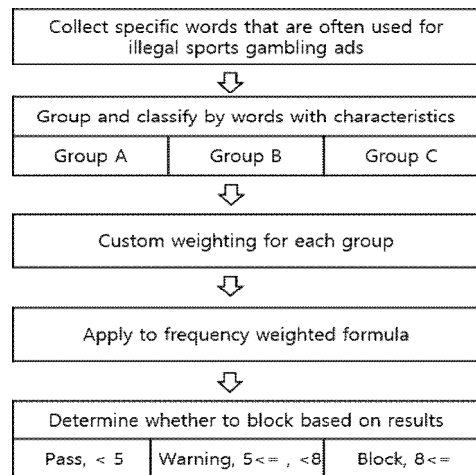


Fig. 1. Blocking Architecture for Illegal Ads of Sports Gambling

1. Imaged Words Analysis

SNS 상의 대부분의 불법 스포츠 도박 광고는 영상물로 이루어져 있다. 불법 스포츠 도박 광고 관계자들은 특정 단어가 필터링 되는 것과 같은 감시망을 피하기 위해 이미지에 속어와 같은 말을 담아서 페이스북 사용자들에게 홍보를 하고 있다. 주로 자극적인 글이나 재미있는 사진 등 사용자들이 관심을 끌 수 있는 글을 찾고, 해당 글에 불법 스포츠 도박 광고 영상을 코멘트로 올리는 방식을 사용한다. 사용자들은 자신들의 타임라인을 통해 친구나 가족, 지인이 반응한 글을 확인할 수 있는데, 이 때 코멘트에 있는 불법 스포츠 도박 광고를 보게 될 수 있다. 코멘트는 직접 확인을 하지 않는 이상 매 순간마다 노출이 되어있는 것이 아니므로 위와 같은 방법으로 자연스럽게 광고를 하면서도 감시자들의 눈을 피할 수 있다. 이는 불법 스포츠 도박 광고업자들이 페이스북의 장점이자 단점인 쉬운 접근성을 파악하고 이를 악용하는 것으로 판단된다. 그들이 이미지 속에 삽입하는 문구는 가족방 홍보, 진행되고 있는 게임의 종류, 진행 중인 이벤트 등이 있다.

불법 스포츠 도박을 운영하거나 이용하는 관계자들은 일반인들이 이해하기 어려운 은어를 사용하기도 하는데, 이는 해당 이미지가 도박 광고인지 아닌지를 구분하여 차단 여부를 결정하는데 중요한 근거가 될 수 있다. 일상생활에서 쓰는 표현과 다른 중의적인 의미를 지니는 경우도 있으나 일상생활 용어들도 다수 존재한다. 따라서 불법 스포츠 도박 광고를 판별하는 시스템을 구축하기 위해 불법 스포츠 도박 광고 시 가장 빈번하게 사용되는 단어들을 선별했다. 불법 스포츠 도박 광고물에서 자주 언급되는 문구는 '카톡', '문의', '이벤트', '사다리', '충전', '가족방', 'com', '스포츠', '루틴', '마틴', '첫충', '가입', '단폴', '달팽이', '다리다리', '매충', '파워볼', '배팅', '놀이터', '포인트', '돌발', '메이저', '공원', '지급' 등 12개 단어이다.

단어들을 분석한 결과 '카톡'처럼 불법 스포츠 도박과 직접적인 연관이 없을뿐더러 일상생활에서 자주 쓰이는 단어들이 있는 반면 '이벤트'처럼 사행성을 조장하는데 쓰일 수 있으면서 일상생활에서 사용되는 단어들이 있고, '사다리' 같은 단어는 일상생활에서 쓰이지만 도박에서 사용하는 단어의 의미와는 전혀 다른 의미를 가진 단어이다. 그리고 불법 스포츠 도박 관계자와 이용자들만 알 수 있는 '가족방'은 일상생활에서 전혀 쓰지 않는 단어이다.

이렇게 수집하여 분석된 단어들은 각 특성에 맞게 분류하는 작업이 필요하다. 단어마다 사용되는 빈도 및 중요도가 상이하기 때문에 모두 같은 기준을 적용할 수 없으므로, 본 논문에서는 3개의 그룹으로 분류한다.

1.1 A Group

A 그룹은 중요도가 가장 높은 단어를 분류한 그룹이다. 불법 스포츠 도박 광고에서 자주 쓰이는 단어 중 관련도가 가장 높은 단어인 관계자 간에 이해할 수 있는 은어들을 해당 그룹으로 분류한다. A 그룹으로 분류된 단어는 '가족방', '루틴', '마틴', '첫충', '단폴', '다리다리', '매충', '파워볼', '공머니', '슈어맨', '다음드' 등 11개의 단어이다.

해당 단어들은 불법 스포츠 도박 광고에서만 쓰이는 용어들로 일상생활에서는 사용하지 않는 단어들이다. 따라서 불법 스포츠 도박 광고로 추정되는 이미지에 해당 단어가 등장하면 불법 스포츠 도박 광고로 간주하여 차단한다.

1.2 B Group

B 그룹은 일상생활에서 사용하는 단어를 중의적인 표현으로 사용하는 단어나 의미는 같지만 불법 스포츠 도박과 밀접하게 관련 있는 단어들을 중심으로 분류한다. B 그룹으로 분류된 단어는 '이벤트', '사다리', '충전', '스포츠', '달팽이', '배팅', '놀이터', '포인트', '돌발', '메이저', '공원', '지급' 등 12개 단어이다.

해당 단어들은 특징에 따라 구분된다. 첫 번째는 일상생활에서도 사용되는 용어들을 가져다가 다른 의미로 사용하고 있는 경우로 '사다리', '스포츠', '달팽이', '놀이터', '돌발', '공원' 등이 있다. 두 번째는 일상생활에서 사용되는 용어와 의미가 같지만 도박과 밀접한 관련이 있는 경우로 '충전', '배팅', '포인트', '메이저', '지급' 등이 있다.

1.3 C Group

C 그룹은 A 그룹과 B 그룹처럼 불법 스포츠 도박과 밀접하게 관련 있는 단어는 아니지만 자주 쓰이는 단어들을 중심으로 분류한 그룹이다. C 그룹으로 분류된 단어는 '카톡', '문의', 'com', '가입', '상담' 등 5개 이다.

해당 그룹의 단어들은 일상생활에서 같은 뜻으로 사용되는 단어들이지만 불법 스포츠 도박 광고에 자주 등장하는 단어들이다. C 그룹의 단어들은 일상생활에서 사용할 때의 의미와 불법 스포츠 도박에서 사용할 때의 의미가 같다. 표 2는 그룹 별 단어 분류를 보여준다.

Table 2. Words by Groups

Group	Words					
A	가족방	루틴	마틴	첫총	단폴	다리다리
	매총	파워볼	꽁머니	슈어맨	다음드	
B	이벤트	사다리	스포츠	총전	달팽이	배팅
	지급	포인트	놀이터	돌발	메이저	공원
C	카톡	문의	가입	com	상담	

2. The Proposed Weighting Technique

A 그룹은 스포츠 도박과 가장 밀접한 관련이 있는 단어들로 분류되어 B 그룹과 C 그룹의 단어보다 중요도가 높다. 반면에 C 그룹의 단어들은 등장하는 빈도수가 가장 높다. 따라서 차단 알고리즘 설계 시 모든 그룹에 동일한 가중치를 부여하기 어렵고, 그룹마다 다른 중요도와 빈도수에 따른 가중치를 적용해야 한다.

본 논문에서는 100개의 불법 스포츠 도박 광고 이미지를 분석하여 그룹 별 그룹 가중치를 산출한다. 중요도가 가장 높은 그룹인 A 그룹의 가중치는 $\alpha = 50$, B 그룹은 $\alpha = 40$, 중요도는 낮지만 빈도수가 높은 C 그룹의 가중치는 $\alpha = 10$ 이다. 이렇게 중요도에 따라 그룹 별로 부여한 α 값은 빈도 별 가중치 공식에 적용된다.

$$WF = \alpha \times \frac{wf}{\max wf} \tag{1}$$

여기서 단어마다 고유의 WF값을 산출하여 불법 스포츠 도박인지 아닌지 판단할 수 있는 기준으로 사용한다. 수식 (1)에서 α 는 그룹 가중치, wf 는 단어 빈도수, $\sum wf$ 는 그룹 내 단어의 전체 빈도수이다. 여기서 WF를 구할 때 광고 이미지 내의 각 단어가 속한 그룹마다 다른 그룹 가중치를 적용해서 그 중요도와 빈도에 따라 값의 차이가 존재한다. 즉, 도박 관련 여러 단어들이 같은 사용 빈도로 등장해도 해당 그룹의 가중치에 따라 WF값이 달라질 수 있는 것이다. 표 3은 단어 빈도수와 그룹별 빈도수 평균을 나타낸다.

Table 3. wf & Mean by Groups

Group	Word Frequency (wf)						Mean
A	가족방	루틴	마틴	첫총	단폴	다리다리	31.45
	48	45	43	40	37	33	
B	매총	파워볼	꽁머니	슈어맨	다음드		32
	31	31	16	14	8		
B	이벤트	사다리	스포츠	총전	달팽이	배팅	32
	58	57	48	47	35	26	
C	지급	포인트	놀이터	돌발	메이저	공원	45.8
	25	21	20	20	17	10	
C	카톡	문의	가입	com	상담		45.8
	64	57	52	51	5		

본 논문은 단순히 단어의 빈도에 초점을 맞추다기보다 해당 단어가 불법 스포츠 도박과 얼마나 관련 있는지, 또 얼마나 중요한 단어인지에 대해 초점을 맞추고 관련성이나 중요도가 높은 단어가 등장하면 불법 스포츠 도박 광고 일 확률이 높아진다는 점을 이용하여 단어의 빈도수 외에 그룹 가중치 α 를 적용한다.

Table 4. WF Table

Group	WF					
A	가족방	루틴	마틴	첫총	단폴	다리다리
	6.94	6.50	6.21	5.78	5.35	4.77
B	매총	파워볼	꽁머니	슈어맨	다음드	
	4.48	4.48	2.31	2.02	1.16	
B	이벤트	사다리	스포츠	총전	달팽이	배팅
	6.04	5.94	5.00	4.90	3.65	2.71
C	지급	포인트	놀이터	돌발	메이저	공원
	2.60	2.19	2.08	2.08	1.77	1.04
C	카톡	문의	가입	com	상담	
	2.79	2.49	2.27	2.23	0.22	

3. Determination of Blocking

표 4의 WF 표를 기준으로 광고 이미지 내의 각 단어마다 WF를 산정하고 이를 합산하여 불법 스포츠 도박 광고인지 아닌지를 판정하는 기준으로 삼는다. 이때 중복되는 단어는 중복해서 더하지 않는다.

$$cv = \sum_{i=1}^n WF_i \tag{2}$$

한 이미지 내의 WF의 총합은 식 (2)와 같이 cv로 정의하고 cv가 8 이상이면 차단 판정을 하며, 5이상 8미만이면 경고 판정을, 5미만일 때는 통과 판정을 하도록 했다.

IV. Experimental Results

본 실험은 불법 스포츠 도박 광고 이미지 내에서 도박과 연관된 단어를 선정, 중요도 별로 A, B, C 그룹으로 분류하고 그룹 가중치 공식인 α 를 적용한 WF 수식을 고안하여 관련 이미지에 대해 적용하였다. 그룹 가중치 α 를 위해 사용된 이미지는 200개의 실험 이미지로 161개의 스포츠 도박 관련 광고 이미지와 일반 이미지 39개를 사용하였다. 한 이미지 내의 홍보 단어들을 그룹별로 분류하고, 그룹 별 정의된 가중치를 적용하여 각 단어들의 WF를 구한다. 식 (2)와 같이 WF들의 총합인 cv를 구하여 판정을 위한 임계값으로 사용하였다.

여기서 차단 판정을 받으면 해당 이미지를 게재할 수 없도록 차단하고, 경고 판정을 받으면 이미지를 게재할 때 경고 메시지를 출력하도록 한다. 그리고 통과 판정을 받으면 이미지를 게재할 수 있도록 한다.



(a) Gambling ads Image (b) ads Image

Fig. 2. Test Ads Images

그림 2(a)는 실험에 사용된 불법 스포츠 도박 광고 이미지이고, 그림 2(b)는 일반광고 이미지이다. 그림 2(a)에 포함된 단어를 표 4의 WF 표와 대조한 결과 ‘가입’, ‘이벤트’, ‘단폴’, ‘배팅’, ‘매충’, ‘카톡’이라는 6개의 스포츠 도박 관련 단어를 확인할 수 있다. 제한한 빈도 별 가중치 공식을 바탕으로 연산한 각 단어의 WF_i $i = 1, \dots, 6$ 의 합을 구하면, $WF_1 = 2.27$, $WF_2 = 6.04$, $WF_3 = 5.35$,

$WF_4 = 2.71$, $WF_5 = 4.48$, $WF_6 = 2.79$ 이다. 따라서 WF 의 총합인 cv 는 23.64로 차단 여부의 임계값인 8 이상이므로 해당 이미지는 차단 판정을 받고, 이미지는 게재되지 않는다. 그림 2(b)는 일반 광고 이미지로 해당 이미지 내의 단어들과 WF 표를 비교분석 했을 때 일치하는 단어가 없으므로 해당 이미지의 cv 는 0이다. 따라서 불법 스포츠 도박 광고 이미지가 아니라고 판단하여 통과 판정을 받고 이미지는 게재된다.

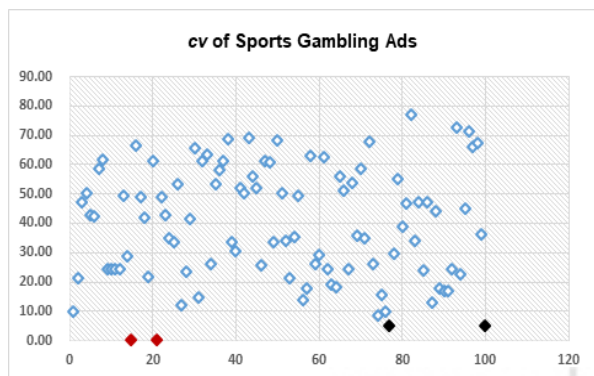


Fig. 3. cv Distribution of Illegal Sports Gambling Images

그림 3의 차트는 본 논문에서 제안하고자 하는 불법 스포츠 도박 광고 이미지를 그룹별 가중치와 빈도별 가중치를 적용한 WF 의 총합으로 산출된 cv 분포로 보여주고 있다. 100개의 불법 스포츠 도박 광고 이미지의 cv 를 보여주고 있으며 96개의 이미지는 파란 표식으로 차단 판정을 받았고, 빨간색으로 표시된 2개의 일반 광고 이미지는 통과 판정을 받았다. 그리고 2개의 검은색 표식은 불법 스포츠 도박 광고 이미지임에도 불과하고 낮은 cv 로 통과 판정을 받았다.

본 논문은 100개의 불법 스포츠 광고 이미지로 표 2, 표 3, 표4와 같이 빈도별, 그룹별 가중치를 산출하였으며, 200개의 관련 이미지를 통해 제안한 방법으로 실험한 결과 일반 광고 이미지 39개는 모두 통과 판정을 받았고 불법 스포츠 도박 광고 161개 중 153개의 이미지가 차단 판정, 1개의 이미지가 경고 판정, 그리고 7개의 이미지가 통과 판정을 받았다. 통과 판정을 받은 7개의 이미지는 불법 스포츠 도박 광고임에도 불구하고 관련된 어휘가 등장하지 않았거나, 적은 수의 어휘가 사용되어 낮은 cv 로 차단이나 경고 판정을 받지 않게 된 것이다. 결국 빈도 별 가중치 공식을 이용해 불법 스포츠 도박 광고를 차단한 실험은 96.5%의 성공률을 보였다.

V. Conclusions

불법 스포츠 도박 광고는 SNS에서 많이 이루어지고 있는데, 특히 페이스북의 댓글에서 쉽게 발견할 수 있다. 불법 스포츠 도박 광고임을 나타내는 문구들을 포함시켜 광고를 함으로써 이용자들은 불법 스포츠 도박 광고 이미지에 쉽게 노출되어 있었다. 불법 스포츠 도박이 스마트 폰의 이용 접근성을 증가 시키고 나아가 청소년 도박을 심각하게 부추기고 있는 문제를 해결하고자 기존의 관리자에 의한 모니터링 방법이나 게시된 콘텐츠의 사후 삭제와는 다른 방법으로 원천적으로 불법 스포츠 도박 광고를 차단할 수 있는 방법을 제안하였다. 제안하는 방법은 불법 스포츠 도박에 자주 사용되는 단어들을 선정하고 중요도에 따라 3개의 그룹으로 나누고 그룹별, 빈도 별 가중치 공식을 이용하여 각 단어마다 WF 를 산정한 후, 해당 이미지 내의 단어들의 WF 의 총합인 cv 로 차단, 경고, 통과를 판정하였다.

본 논문에서 제안한 방법으로 실험을 한 결과 96.5%의 정확도를 보였다. 그러나 3.5%는 불법 스포츠 도박 이미지임에도 단어의 수가 적어 차단 판정에 필요한 임계치에 도달하지 못하여 통과판정을 받았다. 제안하는 방법은 온

라인 내기 게임과 온라인 카지노, 블랙잭 등 불법 스포츠 도박 광고를 통한 도박 사이트의 접근을 차단함으로써 온라인 불법 도박에 대한 한 해결책이 될 것이다. 향후 불법 스포츠 도박 이미지내의 단어의 빈도수에 대한 가중치외의 특징 분석을 통해 불법 스포츠 도박을 근절하기 위한 연구를 지속해야 할 것이다.

REFERENCES

- [1] Korea Internet & Security Agency Summary Report, "2017 Internet usage survey," , pp. 1-6, Feb. 2018.
- [2] Korea Center on Gambling Problems, "2018 Gambling problem survey," pp. 1-4, Dec. 2018.
- [3] National Information Society Agency, "2017 Survey on smartphone dependence," pp. 1-9, Feb. 2018.
- [4] J. Park, S. Choi, and S. Oh, "An effectiveness Analysis of the punishment strengthening policy on illegal acts connected with sports toto," Journal of Korea Society and Administrative Studies, Vol. 25, No. 3, pp. 35-57, Nov. 2014.
- [5] J. Chon, N. Kim, and S. Park, "An Exploration into the college student's adherence behavior of illegal sports gambling," Journal of Korea Alliance for Health, Physical Education, Recreation, and Dance, Vol. 54, No. 6, pp. 95-105, Nov. 2015.
- [6] S. Kim and J. Oh, "A survey on the illegal sport gambling of male middle and high school students," Journal of Korea Alliance for Health, Physical Education, Recreation, and Dance, Vol. 26, No. 3, pp. 23-39, Dec. 2017.
- [7] D. Kim, K. Park, and C. Lee, "The study of illegal gambling industry," Journal of Korean Association of Public Safety and Criminal Justice, Vol. 21, No. 1, pp. 9-42, 2012.
- [8] W. Yun, "Study about operation methods of unlawful internet gambling: suggesting a new research method by using observation survey," Journal of Korean Association of Public Safety and Criminal Justice, Vol. 25, No. 3, pp. 139-165, 2016.
- [9] S. Han, B. Choi, and J. Lim, "A survey on the illegal sport gambling of male middle and high school students," Journal of Korea Institute of Information Security and Cryptology, Vol. 22, No. 3, pp. 19-24, 2012.
- [10] I. Kim and M. Yang, "A study for blocking harmful contents through a local proxy on android," Journal of Korea Society of IT Services, Vol. 12, No. 2, pp. 103-118, 2013.
- [11] B. Kim and S. Lee, "Improvement of methods for discriminating harmful web sites by using link relations between web sites and constructing white-list," Conference of Korean Institute of Information Scientists and Engineers, Vol. 28, No. 12, pp. 238-240, 2018.
- [12] G. Lee and Y. Bae, "The techniques to block distribution for the illegal information through google searching-based on posting the sale of personal information," Journal of Korea Institute of Digital Forensic , Vol. 12, No. 2, pp. 77-92, Dec. 2013.
- [13] J. Lee, B. Choi, and Y. Chung, "Comparative evaluation of term weighting methods in automatic document classification," Conference of Korea Society for Information Management, pp. 41-44, Aug. 2000.
- [14] J. Kim, "A plan to block illegal gambling advertisements in social network services of teenagers," Master's thesis, Chosun University Graduate School of Education, 2017.
- [15] S. Kang, "Junk-mail filtering by mail address validation and title-content weighting," Journal of Korea Multimedia Society, Vol. 9, No. 2, pp. 255-263, Feb. 2006.
- [16] S. Kim, and H. Park, "Normalized term frequency weighting method in automatic text categorization," Conference of The Institute of Electronics and Information Engineers, pp. 255-258, Nov. 2003.

Authors



Ji-A Kim received the B.S. degree in Computer Engineering from Chosun University and M.S. degree in Information and Computer Science Education from Chosun University, Korea, in 2015 and 2018,

respectively. She is currently a teacher in Seokang High School, Gwangju, Korea, 2018. She is interest in information and computer education, mobile game addiction, and Social Network Services of teenagers.



Geum-Boon Lee received the M.S. and Ph.D. degrees in Computer Engineering from Daejeon University and Chosun University, Korea, in 2002 and 2010, respectively. Dr. Lee joined the faculty of Chosun University,

Gwangju, Korea, in 2018. She is currently a professor in SW Convergence Education Institute, Chosun University. She is interested in artificial intelligence, pattern recognition, bigdata analysis, and deep learning.