

Research on analysis of articleable advertisements and design of extraction method for articleable advertisements using deep learning

Seoksoo Kim*, Jae-Young Jung**

*Professor, Dept. of Media & Visual Communication, Hannam University, Daejeon, Korea

**Professor, Dept. of AI Bigdata Convergence, Dongyang University, Dongducheon, Korea

[Abstract]

There is a need for and positive aspects of article-based advertising, but as exaggerated and disguised information is delivered due to some indiscriminate 'article-based advertisements', readers have difficulty distinguishing between general articles and article-based advertisements, leading to a lot of misinterpretation and confusion of information. Since readers will continue to acquire new information and apply this information at the right time and place to bring a lot of value, it is judged to be even more important to distinguish between accurate general articles and article-like advertisements. Therefore, as differentiated information between general articles and article-like advertisements is needed, as part of this, for readers who have difficulty identifying accurate information due to such indiscriminate article-like advertisements in Internet newspapers, this paper introduces IT and AI technologies. We attempted to present a method that can be solved in terms of a system that incorporates, and this method was designed to extract articleable advertisements using a knowledge-based natural language processing method that finds and refines advertising keywords and deep learning technology.

▶ **Key words:** Internet newspaper, article advertisement, deep learning, information extraction, media

[요 약]

기사성광고의 필요성과 긍정적인면도 존재하나, 일부 무분별한 '기사성 광고'로 인하여 과장되고 위장된 정보를 전달함에 따라 독자들은 일반기사와 기사성 광고의 구별이 어려움에 따라 많은 정보의 오인과 혼란을 하고 있다. 독자들은 앞으로도 계속 새로운 정보를 획득하고 이러한 정보를 적재적소에 적용하여 많은 가치를 가져와야 하기에, 정확한 일반기사와 기사성 광고를 식별하는 것이 더욱이 중요하다고 판단된다. 따라서 일반기사와 기사성 광고의 구별된 정보를 필요로 하기에, 이러한 일환으로, 인터넷신문에서 이러한 무분별한 기사성 광고로 인한 정확한 정보식별의 어려움이 많은 독자들을 위해, 본 논문에서는 IT기술과 AI기술을 접목한 시스템측면에서 해결할 수 있는 방법을 제시하고자 하였으며, 이러한 방법은 광고성 키워드를 찾아내어 정제해주는 지식기반 자연어처리 방법과 딥러닝기술을 이용한 기사성 광고를 추출하고자 설계 하였다.

▶ **주제어:** 인터넷신문, 기사성광고, 딥러닝, 정보추출, 미디어

-
- First Author: Seoksoo Kim, Corresponding Author: Jae-Young Jung
 - *Seoksoo Kim (sskim@hnu.kr), Dept. of Media & Visual Communication, Hannam University
 - **Jae-Young Jung (jyjung@dyu.ac.kr), Dept. of AI Bigdata Convergence, Dongyang University
 - Received: 2024. 05. 07, Revised: 2024. 05. 31, Accepted: 2024. 06. 03.

I. Introduction

우리는 정보의 홍수속에 살고 있으며, 이러한 정보는 각종 신문사나, 미디어 및 매체관련 회사에서 제작하고 유포하고 있다. 특히 이러한 정보는 온라인에서도 거의 비슷한 형태로 볼 수 있으며, 또한 이러한 정보는 제작자의 의도에 따라 사실들이 약간씩 달리 표현되기도 한다. 이러한 정보의 제작 및 유포의 의도속에 많은 광고들도 포함되어 있어서, 때론 독자로 하여금 일반기사와 광고의 구별이 애매할 때도 있는 것이다. 이렇게 기사와 같은 형태로 광고를 하고 있는 정보를 기사성광고라 말하며, 이러한 기사성 광고는 날로 급증하고 있는 추세이다.

이런 상황에서 기사형 광고는 언론사에게는 수익을 확대할 수 있는 중요한 수단으로, 그리고 광고주에게는 전통적 광고 형식의 낮은 효과를 극복하는 동시에 생산과 유통의 편이성과 비용의 경제성을 높이는 수단으로 인식된 것이다 [1,2]. 더욱이 소비자들도 기사형 광고가 중요한 정보를 요약적으로 전달하는 효율적 수단이라고 인식하고 있다[3].

하지만 이와 같은 긍정적 효과가 메시지 내용을 광고주의 일방적 의견이 아니라 기자에 의해 검증된 정보라고 수용자가 오인한 결과라는 주장이 제기되면서 저널리즘의 기본 원칙을 훼손할 뿐만 아니라 일반적인 윤리 기준에도 부분적으로 위배 된다는 비난이 확대 되었다[1].

기사형 광고를 게재하는 행위는 사실에 바탕을 둔 뉴스 가치가 있는 정보를 보도자료 형태로 제공하여 기사화되도록 하는 PR활동을 의미하는 퍼블리시티로는 볼 수는 없다. 퍼블리시티의 경우, 제품 혹은 서비스의 생산자가 능동적으로 언론사에게 정보를 제공하지만, 기사화의 여부는 전적으로 언론사의 자율적 결정에 따른다. 이에 반해서 기사형 광고는 대가에 의해 기사화가 보장된다는 점에서 명확히 구별되어 진다[4].

요즘, 일반적으로 소비자가 광고에 비해 신문기사나 프로그램에 대해 높은 신뢰도를 보이는 점을 이용해, 광고에 대한 소비자의 신뢰도를 제고하기 위한 방법으로 일반적인 광고형태에서 벗어나 기사나 프로그램 형태로 광고하는 경우가 있으며, 이러한 형태의 광고를 기사성광고라고 하는데, 최근 신문과 잡지에서 양적으로 증가하고 있는 형태로, 시간이 갈수록 광고에 사용된 글씨체, 레이아웃 등이 기사와 구분하기 어렵도록 교묘해지고 있다[5].

최근 기사성광고에 관한 연구와 동향을 살펴보면 다음과 연구(참고문헌 6~10)와 같다.

최근 들어서는 기사형식을 빌린 이른바 '기사형 광고'가 문제가 되고 있다. 재테크 특집, Real Estate 부동산 섹션,

업계 윗계 한 히트상품, 화제의 분양현장, 분양 포커스, Luxury 등의 이름으로 특정 기업 제품이나 부동산, 금융 상품 등을 장점 일변도로 소개한다. 이러한 기사형 광고는 대체로 지면 상단에 'advertorial page' 표기를 넣어 해당 지면이 기사형 광고임을 밝히고 있다. 그럼에도 각각의 기사에 기자 바이라인을 넣어 독자들이 일반 기사로 오인할 수 있게 했다. 신문광고윤리 실천요강은 광고임이 명확하지 않고 기사와 혼동되기 쉬운 편집체제나 표현을 쓰지 못하도록 하고 있다. 이러한 신문 제작방식은 자사와 특정 기업의 영리를 위한 것이라는 지적을 피하기 어렵고, 신문의 신뢰성과 공신력을 훼손할 수 있다[6].

'기사형 광고'는 언론 기사 형식의 광고이다. '기사'가 언론사 기자가 취재한 내용을 사실에 기반하여 전달하는 것이라면, '광고'는 광고주가 상품이나 서비스 등과 관련한 정보를 소비자에게 알리고 구매를 설득하기 위해 전달하는 메시지이다. 기사와 광고는 주제, 목적, 효과가 각기 다르기 때문에 명확히 구분하여 편집해야 한다. 광고가 기사의 형식을 취하면, 독자(소비자)가 이를 기사로 오인하여 합리적인 의사 결정과 선택을 하지 못하는 문제가 발생할 수 있다[7].

기사형 광고에 대한 문제 제기는 오래전부터 계속되었지만 개선되지 않고 있고, 기사형 광고는 오히려 더 정교해지고 있다. 최근 광고대행사가 각 언론사의 특성에 맞게 기사형 광고를 제작하여 포털 사이트에 노출시키는 영업 행위를 했다는 문제가 제기되기도 했다[8].

입법조사처는 기사형 광고가 기사와 광고를 구분하게 한 신문법 위반이라는 점을 지적하며 "광고가 기사의 형식을 취하면, 독자가 이를 기사로 오인하여 합리적인 의사 결정과 선택을 하지 못하는 문제가 발생할 수 있다"고 지적했다. 그러면서 "기사형 광고에 대한 문제 제기는 오래전부터 계속되었지만 개선되지 않고 있다. 기사형 광고는 소비를 끌어낼 수 있을지는 몰라도 결국 언론의 신뢰도를 떨어뜨리는 결과를 가져올 수 있다"고 했다[9].

최근 정부는 의료기관과 의료인의 사상적 광고 방식을 보수하고자 매체를 통한 정보 게재에 대해 강력하게 법적 단속조치를 취하고 있다. 대한치과의사협회는 이를 감시하는 한편, 더욱 강력한 재발 방지가 필요하다고 주장하고 있으며, 정보 게재에 대한 각종 법적 방어막을 마련하는데 있어 다양한 대책 모색을 원하고 있다.[10]

최근에는 이러한 기사성광고로 인해 일부 소비자들은 광고에 대한 부정적인 인식을 보이기도 하며, e-mail에서 제목에 광고라는 표현이 들어가면 기본 알고리즘 처리방식으로 스팸메일로 분류하는 경우도 많이 있다. 특히, 인

터넷신문의 경우 광고에 대한 부정적인 인식은 상당하며, 그 이유는 기본적으로 광고가 많아서 기사를 읽는 것에 대해 방해가 되기 때문이다. 이렇게 집행되는 광고의 내용도 많은 비판을 받기도 하며, 또한 이러한 것으로 인한 광고에 대한 부정적인 경험은 광고회피로 이어지기도 한다 [11]. 미디어 이용자들이 광고를 회피하게 되면 미디어는 존립 기반을 위협당하게 될 수 있으며, 전통적인 미디어 위기와 더불어 현실에서 광고회피의 심화로 인해 기존 광고도 유지하지 못할 수 있을 것이다. 따라서 인터넷신문에 있어서 이용자의 광고회피 연구는 미디어 산업에 유용한 시사점이라고 할 수 있다[12].

이러한 기사성광고는 온라인 또는 오프라인을 막론하고 무분별하게 확대되고 있으며, 독자로 하여금 기사성광고로 인해 정확한 정보를 얻기 위한 여러 어려움이 뒤따르게 된다. 그래서 본 논문에서는 이러한 기사성광고의 여러 가지 상황에 대해 분석하고 고찰하고자 한다. 또한 시스템적으로 기사성광고와 일반기사의 식별을 위해서 기사성광고의 추출방법을 제시하고 있으며, 이는 기사성광고의 특징 및 많이 사용되고 있는 기사성광고의 키워드를 적용하여 사용자 필요에 따라 기사성광고를 추출하는 방법을 설계하고 제시하고자 한다.

특히, 이러한 연구는 인터넷신문에 한정해서 진행하였으며, 인공지능기술의 자연어처리와 딥러닝기술을 이용해 기사성광고의 내용을 추출하고자 한다. 이러한 연구의 목적은 가능한 광고관련 기사를 추출하여 정확한 정보(일반 기사와 기사성광고의 구별)를 원하는 사용자들에게 서비스 하기 위한 방법으로 일부의 도움이 되고자 한다.

II. Related research

1. Newsworthy advertisement

기사성광고란 기사형식을 이용해서 광고를 내는 것을 의미하는 것으로, 신문의 기사나 방송의 보도와 같이 뉴스 스타일의 문장을 광고카피로 이용하여 광고 의뢰인이 광고비를 지불하여 광고지면에 내는 것을 의미한다. 뉴스를 닮은 카피가 광고카피방식으로 채택되면서 광고와 기사의 내용이 구별이 어렵도록 그 경계가 허물어졌다고 볼 수 있다. 또한 기사의 편집에서 본 기사성광고의 분류는 다단편집, 배열, 활자체, 헤드라인카피 등 광고의 구성과 내용면에서 기사형식을 똑같이 가져옴으로써 일반 소비자가 기사로 인식할 가능성이 있는 광고형식을 가지고 있다. 신문에 전면광고를 게재하면서 지면을 상하로 분할하여 상단

에는 특정 광고주의 사업내용과 관련한 기사형태의 글을 싣고 하단에는 일반적인 형태의 광고를 게재한다든가, 잡지에 광고라는 안내도 없이 ‘기획 특집’, ‘△△탐방’, ‘○○체험’ 등의 제목을 지정함으로 광고인지 기사인지 매우 혼동되는 글을 싣는 경우가 점차 증가하고 있다.

Table 1. Characteristics of article-based advertising [5]

| Classification | Feature Content |
|-------------------------|--|
| common characteristics | <ul style="list-style-type: none"> - Describe the pros and cons, reputation, consumer interviews, etc. of a specific product in a format as if covered by a reporter. - Create headlines like newspaper articles with titles such as ‘Acquisition Boom’ or ‘Emergency Diagnosis’ or use words such as ‘coverage’ or ‘editor’ in the advertisement content. |
| unique features | <ul style="list-style-type: none"> - Use section names that indicate the overall nature of the articles featured on the top page - Use the writer’s name and email address displayed at the end of the article |
| magazine advertisement | <ul style="list-style-type: none"> - A symbolic magazine logo is displayed at the end of the advertisement. - Write ‘text (or summary) ○○○, photo ○○, progress △△’, etc. at the beginning. - Use the format of a several-page special article containing several advertisements. |
| Ad serving format | <ul style="list-style-type: none"> - When multiple article-based advertisements form a full-page advertisement - When one article-like advertisement is published as a full-page advertisement - Partial advertisement where article and advertisement are mixed on one page |
| Full-page advertisement | <ul style="list-style-type: none"> - The top part is an article advertisement, and the bottom part is a general advertisement. - Post related article advertisements below the actual article. |

이러한 기사성광고는 광고제품의 품질과 성능에 대해 소비자의 오인을 유발해 합리적인 선택을 방해하기 쉽고, 현행법상 금지되어 있는 광고표현을 사용하기 위한 수단으로 악용될 수도 있거나 부당광고의 우려가 있으며, 의료법상 광고가 금지되어 있는 의료기관의 진료방법을 게재하거나 식품위생법상 위반사항인 체험기를 게재하기 위한 수단으로 이용되는 사례가 있다[5].

상기의 Table 1은 기사성광고의 다양한 특징을 보여주고 있으며, 특징의 분류를 살펴보면 공통적 특징, 독특한 특징, 잡지광고의 특징, 전면광고의 특징, 광고게재 형식에 따른 특징별로 내용을 보여주고 있다.

2. Deep learning

딥러닝은 인공지능학습의 한 분류이며, 응용으로는 음성 인식, 이미지 식별 또는 예측 등 다양하게 이루어지고 있다. 사람의 작업을 대신 수행하도록 컴퓨터를 학습시키는 일종의 머신러닝이며, 본 논문에서는 자연어처리를 활용한 딥러닝의 응용이 한 예이다. 이러한 딥러닝은 데이터에 대한 기본 파라미터를 설정하고 컴퓨터가 여러 처리 계층을 이용해 단어나 패턴을 계속 인식함으로써 스스로 학습하도록 훈련시키는 기술이다.

딥러닝은 기계학습(Machine learning)분류에 포함된 최신 인공지능을 통한 학습방법 중 한 부류이며, 기계학습보다 깊고 복잡하게 구성된 모델이 입력값과 출력값이 포함된 데이터셋으로 학습할 수 있다[13]. 최근 음성인식, 영상 분류 및 추천 시스템은 많은 두각을 나타내고 있으며, 딥러닝 모델의 성능 향상에 대한 연구와 실생활에 적용하기 위한 응용 연구 모두 매우 활발하게 이루어지고 있다. 기존의 전통적인 기계학습방법은 모델의 구조가 단순하기 때문에 입력값도 수동으로 전처리를 통해 단순화해야 적절한 결과를 얻을 수 있지만, 딥러닝은 전처리를 거치지 않아도 복잡한 형태의 입력값을 학습하는 것이 가능하므로 딥러닝 기반 시스템의 일관성 및 성능이 기계학습보다 뛰어나다고 볼 수 있다[14,15].

기계학습이 활발하게 연구되던 1980년대와 2010년대를 고려했을 때, 1980년대에는 하드웨어의 제약 및 과적합(Overfitting) 등 한계점으로 인해 기계학습 모델의 규모를 크게 키우지 못 하였으나[16], 2010년대에 들어서기 시작하면서 고도의 그래픽카드(GPU)를 활용하여 복잡한 구조의 딥러닝 모델을 빠르게 구현하는 것이 가능하였고[17], Dropout[16], Relu[18] 및 Batch normalization[19] 등의 방법을 통해 과적합의 방지 및 딥러닝 모델의 효율적인 학습이 가능해지면서 인공지능 학습 분야는 딥러닝이라는 이름으로 다시 활성화되어, 2019년 현재 세계 각국에서 경쟁적으로 다양한 분야에서 활용할 수 있는 모델들이 활발히 제안되고 있는 실정이다[13,15].

딥러닝은 분석을 통해 해결하고자 하는 문제를 표현함에 있어 사람의 사고방식을 변화시키고 있으며, 과거에는 컴퓨터에게 문제 해결 방법을 알려주는 것이 주목적이었다면 현재는 문제 자체를 해결하도록 컴퓨터를 훈련시키는 것이 관건이라고 볼 수 있다. 이러한 딥러닝의 잠재력은 널리 보급할 수 있고, 우수한 적응력을 발휘할 수 있으며, 데이터가 추가될 때마다 계속해서 진화하며, 경직된 비즈니스 규칙을 기반으로 구축된 예측 시스템보다 더 동적인 예측 시스템을 만들 수 있다는 데 있다. 그러기에 더

이상 모델을 계속해서 수정할 필요가 없으며, 학습만 시키면 되는 것이다[20].

III. The problem with article-based advertising

1. Recognizing the problem of article-based advertising

기사성광고와 같이 위장된 광고는 언뜻 보아서는 기사의 일부인지 아니면 광고인지 구분이 되지 않을 때가 종종 있다. 이러한 광고가 등장하게 된 것은 광고의 매체 상황이 매우 혼잡하여 소비자의 시선과 최대로 관심을 끌 수 있는 방법이기도 신뢰성 확보를 위한 방법으로 기사의 형식을 빌어서 광고를 담게 되었다. 따라서 신문 등 매체 자체의 신뢰성이나 객관성, 더 나아가 언론의 권위나 명성에 편승하여 광고의 신뢰성을 높이고자 한 것이 그 배경이라고 할 수 있다.

이러한 기사성광고들은 일단 소비자의 관심과 호기심을 유발하게 되고 독자가 그 설득 의도를 쉽게 간파하기 어렵기 때문에 설득에 대한 경계심이 줄어들게 되었다. 더욱 심각한 문제는 일반소비자가 그것이 상품판매를 주목적으로 한 광고임을 인지 못하고 마치 신문사에서 어떤 사실을 기사화하는 과정에서 제품을 그 일부로 언급하는 것처럼 잘못 판단하기 쉽다는 점이다. 나아가 권위있는 언론에서 광고와 달리 기사로 소개하는 것이기 때문에 신뢰성이 확보될 가능성이 높기에 소비자를 오도하며 나아가 기만하는 결과가 되는 것이다.

또한 이러한 배경은 경제불황으로 인해 광고비가 삭감되자 광고주들은 광고효과 문제에 집착하게 되고 이러한 형편을 감안한 일부 신문을 비롯한 매체사들이 광고주에게 광고효과 제고 측면에서 여러 가지 편의를 제공하는 사례가 빈번하게 나타나고 있으며, 광고물을 확보하기 위해 이러한 위장광고를 오히려 적극적으로 권유한다는 것은 커다란 문제를 야기시킬 수 있으며, 때에 따라선 신문사 광고국이 이러한 형식의 광고를 직접 제작하여 광고주에게 접근하는 현실은 심히 우려할 상황이다[21].

2. Analysis of article-type advertisements and online article-type advertisements in print media

한편, 아래 Table 2에서와 같이 온라인 기사형 광고는 기사형 광고가 디지털 기술과 결합한 형태이므로 기본적으로 기사형 광고의 형식적 특징을 지닌다. 그러나 동시에

활자체, 레이아웃, 광고자 로고 표시 방법 등과 같은 형식 요인과 제작, 유통, 수용 과정과 같은 비형식적 요인에서 인쇄매체 기사형 광고와 뚜렷하게 구분되는 몇 가지 특징을 지닌다.

Table 2. Characteristics of differences between article-type advertisements in print media and online article-type advertisements[22]

| Division | Print media article type advertisement | Online article advertisement |
|-----------------------|---|---|
| space constraints | There are limitations on ground space. Printing costs increase as publication increases. | Can be posted as per server capacity. There are almost no additional costs associated with posting. |
| Publication burden | You can see at a glance the proportion of news articles and article-type advertisements. If it exceeds a certain percentage compared to news articles, it becomes burdensome. | The structure makes it difficult to determine the proportion of online article-type advertisements. No need to consider the proportion of news articles |
| Check posting history | Third parties can easily check publication history through printed materials | It is difficult to check publication history because the editing results are not fixed and change in real time. |
| Posting format | A structure where the title and body are exposed together. The distinction between news articles and exposure space is relatively clear. | You must click on the exposed title to view the text. It is difficult to distinguish between general news articles and titles as they are listed in the same space. |
| advertising clues | The font is different from the article or a sign such as 'Full PR' appears at the top. Unlike regular articles, there are still signs or clues that it is an advertisement. | All clues about advertising have been removed. Since it uses the same format as digital news, it uses the press company logo, transmission time, scraping function, etc. It is also indicated as 000 reporter, 000 guest reporter, etc. |
| Exposure opportunity | Only exposed on offline platforms owned by media companies such as newspapers and magazines. Only a temporary exposure effect depending on the publication cycle of offline media | It is transmitted from online platforms owned by media companies to external platforms such as portals and is exposed everywhere. Automatically archived and continuously exposed through search function |

3. The ripple effect of article-based advertising

업종별로 높은 빈도를 보인 것은 자격증 교재, 대리점 및 자판기, 영어교재, 건강(성형, 의술, 호르몬, 다이어트, 건강보조식품)가 주를 이루고 있다. 교재류에 속하는 '자격증 교재'와 '영어교재'를 합하면 약 전체광고물의 반을 차지하고 있으며, '의술' 및 '건강관련제품'의 그 뒤를 따르고 있다. 기사체광고는 교재나 건강제품과 같이 이성적이고 고관여제품을 대상으로 하는 경우가 많은데, 이러한 것은 기사형식을 빌어 많은 정보를 제공함으로써 이성적으로 설득하는데 유리하기 때문에 풀이되고 있다[23].

기사성광고에 있어서 긍정적인 면과 부정적인 면이 같이 존재할 수 있으며, 본 연구에서는 어느 정도의 긍정적 측면을 인정하고 있으나 부정적측면이 더욱 심각한 문제라고 생각하고 있다. 이러한 문제를 해결하기 위해서는 적절한 규제나 법규를 재정립할 필요가 있다고 생각하며, 이에 따른 소비자들의 정확한 정보식별에 오인이 없기를 바란다. 또한 이러한 목적 의도를 위해서 인터넷 신문에서의 기사성광고의 추출방법을 같이 제시하고 있다. 아래의 Table 3의 내용은 기사성광고의 긍정적 측면과 부정적 측면의 효과에 대해서 분류하여 설명하고 있다.

Table 3. Effects of article advertising

| Classification | Effect |
|----------------|--|
| Positive side | <ul style="list-style-type: none"> - For advertisers, it is an important source of income in a rapidly declining advertising market. - News-style advertising innovations are of great help in remembering and persuading consumers. - Compared to general advertising, immersion is better. - Intention to purchase products exposed to article-based advertisements increases more than to general advertisements. |
| negative side | <ul style="list-style-type: none"> - You may be confused between general articles and advertising articles. - Newspapers have less trust in the media. - It is difficult to obtain accurate information about consumers. - There are some problems with ethics. - Exaggerated advertising can affect the property, life, and health of consumers. |

IV. Extraction of article-worthy advertisements using deep learning

인터넷 신문에서의 기사성광고의 식별을 도와 더 정확한 일반기사와 광고를 구별하기 위한 방법으로 인공지능기법

의 한 부류인 딥러닝기법을 이용하여 기사의 내용에 기사성광고의 키워드를 찾아 자연어처리를 거쳐 충분한 딥러닝을 거쳐 기사성광고와 일반기사를 구별하여 제시하고자 한다. 이를 위한 방법으로 특정한 광고성키워드를 제시해 훈련을 시키고, 이러한 내용을 지식기반 데이터베이스시스템에 적재하여 다시 정해서 상황분석을 반복함으로써 가능한 일반기사와 기사성광고의 식별도를 높이고자 한다.

아래 Fig. 1에서는 인터넷신문에서의 기사성광고를 추출하기 위한 전체 흐름을 나타낸 것이다. 기사를 제공하는 많은 신문사나 각종 매체회사에서 정보를 만들고 이러한 내용을 인터넷에 유포하게 되면, 이러한 정보를 본 시스템인 지식기반 데이터베이스에 의해 특정 광고성 키워드에 해당하는 기사들을 1차적으로 추출하게 된다. 또한 더 정교한 추출을 위해서 1차에서 걸러진 정보를 딥러닝시스템에 의해서 2차적인 기사성광고를 추출하게 되며, 이렇게 정제된 정보는 사용자에게 기사를 찾을 수 있도록 도와주고자 한다.

4장은 기사성광고를 딥러닝을 이용한 추출하는 방법을 제시하고 있으며, 자세한 내용은 아래의 각 절에서 다루어지고 있다. 즉, 1절에서는 기사성광고의 광고표식과 키워드를 살펴보았으며, 2절에서는 딥러닝과 자연어처리를 이용한 기사성광고 추출방법에 대해서 5단계(데이터 수집 및 준비, 특성 추출 및 전처리, 모델 선택 및 학습, 평가 및 최적화, 배포 및 모니터링)를 거쳐 설명하였으며, 또한 3절에서는 KBMS(지식기반관리시스템)의 자연어 처리를 통한 1차 추출에 대해서 설명을 하였고, 4절에서는 딥러닝을 이용한 기사성광고의 2차 정제방법에 대해서 제시하였다.

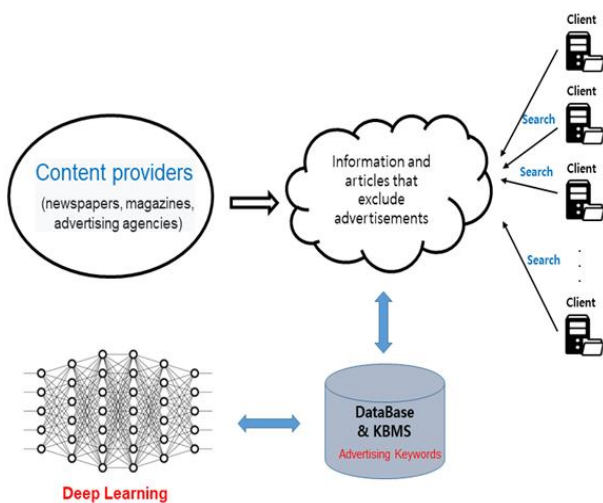


Fig. 1. Overall flow chart for extracting article-worthy advertisements

Fig. 2은 광고관련 특정워드 정제 및 추출 단계의 구성도를 나타내고 있으며, 지식기반관리시스템(KBMS, Knowledge-based management system)에 의해서 이루어지며, 총 2차의 단계로 구성되어 지며, 1차에서는 다시 2개의 단계(1-1, 1-2)로 구분되어 진다. 즉, 1차의 앞 단계(1-1)에서 전면광고, 전면PR, PR특집, 기획광고, 기획특집, 상품정보와 같은 키워드에 의한 정제단계를 거치고, 1차의 뒷단계(1-2)에는 회사명, 제품명, 제품설명, 기능설명, 의학컬럼과 같은 좀 더 자세한 키워드를 주어 정제를 수행하게 된다. 그리고 2차의 단계에서는 이러한 정보들은 다시 딥러닝에 의해서 정제가 이루어지는데, 이는 앞뒤 문맥 및 상황정보등을 가미한 좀 더 자연스럽게 정확한 콘텐츠를 생성하는 단계이다. 이러한 2차단계의 딥러닝방법은 광고 연관성 키워드와 앞뒤의 문맥들의 상황들을 인식하여 좀 더 인식의 정확도를 높이는 단계로써, 광고유발성 키워드에 대한 충분한 학습을 거친 후 맞춤형 온라인기사를 얻도록 하는 위함이다.

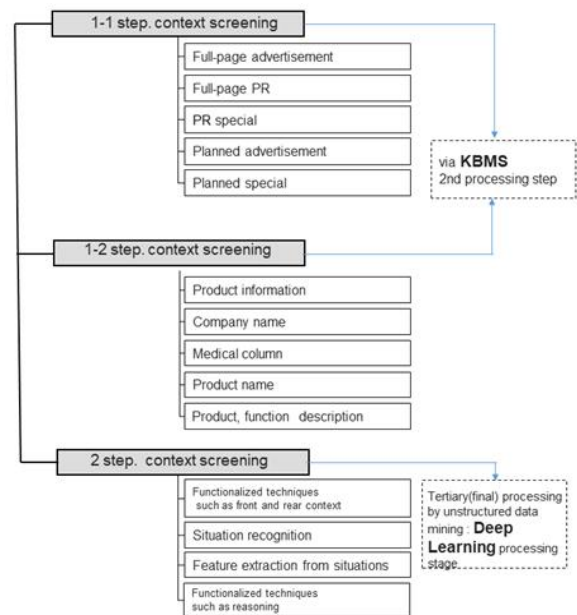


Fig. 2. Advertising-related specific word purification and extraction stage

1. Advertising Markers and Keywords

한국의 기사성광고의 광고표식과 키워드를 살펴보면 다음과 같다. 광고표시를 단 것 중에는 '전면광고'라고 표시된 것이 압도적으로 많으며, '기획특집 PR', '기획특집', '기획 PR', '기획 PR 시리즈'와 같이 유사명칭이 사용된 경우도 많이 있다. '기획 PR'이라는 표시는 의약품(비만치료, 성형, 대머리 치료 등) 광고가 많다. 독자가 보기에 '기획특집', '기

획PR'이라는 용어는 광고라는 의미가 희석되고 대신 신문사에서 제작한 기획기사라는 느낌을 줄 우려가 많다.

'기획특집' 또는 '기획PR'이라는 표시를 단 광고물들은 대부분 광고주가 제작한 광고들이다. 여기서 광고라는 표시 대신 '기획 PR'이라는 문구를 삽입한 광고물은 상대적으로 작은 활자를 사용하고 있어서 각별히 주목하고 있지 않은 한 독자가 표시를 인지하지 못할 가능성도 커 보였다. 더불어 전면광고로 구성된 경우, 대부분 <한방&건강>, <건강>, <Health & Life>, <창업 & 성공>과 같이 특집기사에서 자주 볼 수 있는 것과 비슷한 헤드라인을 사용하고 있다[23].

Table 4는 한국의 광고표식과 관련된 키워드를 몇가지 특징에 기반하여 분류하였으며, 이러한 키워드들을 지식기반관리시스템(KBMS)에 의해서 걸러지게 된다. 우선 전면광고, 전면PR, PR특집, 기획광고, 기획특집, 상품정보과 같은 키워드에 의한 정제단계를 거치고, 다음 단계에는 회사명, 제품명, 제품설명, 기능설명, 의학컬럼과 같은 좀 더 자세한 키워드를 주어 정제를 수행하게 된다. 이렇게 KBMS에 의해서 걸려진 정보들은 다시 딥러닝에 의해서 반복적인 정제가 이루어지는데, 이는 앞뒤 문맥 및 상황정보등을 가미한 좀 더 자연스럽게 고급스런 콘텐츠를 생성하는 단계이다.

Table 4. Ad sign keyword

| Characteristic | Keyword | etc |
|-------------------------------|--|--|
| most used | <Health>, <Oriental Medicine & Health>, <Health & Life>, <Startup & Success> | Full-page advertisement |
| Most advertising keywords | 'Planned PR', 'Planned special feature', 'Planned special PR', 'Planned PR series' | |
| Medicine-related keywords | 'Planning PR' | Plastic surgery, obesity treatment, baldness treatment, etc. |
| Advertiser-generated keywords | 'Planned PR', 'Planned special feature' | use small print |

또한 광고 연관성 키워드와 앞뒤의 문맥들의 상황들을 인식하여 좀 더 추출의 정확도를 높이게 된다. 이러한 기사성광고의 추출방법은 자연어처리를 기반으로한 KBMS와 딥러닝방법을 거침으로써 정확한 정보(일반기사와 광고기사의 구별)을 원하는 독자들에게 좀 더 낫은 서비스인 맞춤형 온라인 기사를 얻도록 하는 방식이다.

2. Refining method using deep learning for article-based advertisements

기사성 광고의 딥러닝을 이용한 정제 방법은 기본적으로 Natural Language Processing(NLP) 기술을 기반으로 한다. 기사성 광고, 즉 "네이티브 광고"나 "스폰서드 콘텐츠"는 일반 콘텐츠와 유사한 형태로 제작되어 사용자가 광고임을 쉽게 인지하지 못하게 하는 광고 유형이다. 이러한 광고를 효과적으로 식별하고 정제하기 위해 딥러닝 기술이 활용될 수 있으며, 주로 다음과 같은 접근 방법이 사용되며, 기사성 광고의 정제는 이와 같은 다양한 딥러닝 기술을 종합적으로 사용하여 수행된다. 사용자의 경험을 향상시키고, 콘텐츠의 품질을 관리하는 데 있어 중요한 역할을 한다.

1) Text Classification

딥러닝 모델을 사용하여 텍스트 데이터가 기사성 광고인지 아닌지를 분류하는 작업이다. 이를 위해, CNN(Convolutional Neural Networks), RNN(Recurrent Neural Networks), LSTM(Long Short-Term Memory) 같은 네트워크가 사용될 수 있으며, 최근에는 Transformer 기반 모델들(BERT, GPT 등)이 더 우수한 성능을 보이고 있다.

2) Content Understanding

기사의 내용을 깊이 이해하기 위해 Natural Language Understanding(NLU) 기술을 사용한다. 이 과정에서는 텍스트의 의미, 문맥, 의도 등을 분석하여 광고성 내용을 식별한다.

3) Feature Extraction

텍스트 데이터에서 기사성 광고를 식별할 수 있는 특징을 추출하는 과정이다. 예를 들어, 특정 키워드의 빈도, 문장 구조, 사용 언어 스타일 등이 이에 해당할 수 있다. 딥러닝 모델은 이러한 특성을 학습하여 광고와 일반 콘텐츠를 구분한다.

4) Supervised Learning

레이블이 지정된 데이터셋(기사성 광고와 아닌 콘텐츠로 구분된)을 사용하여 모델을 학습시킨다. 이 과정에서 모델은 데이터셋 내의 패턴을 학습하여 새로운 콘텐츠가 주어졌을 때 그것이 기사성 광고인지를 예측할 수 있다.

5) Transfer Learning

이미 대규모 데이터셋으로 사전 학습된 모델을 기반으로 소량의 타겟 데이터셋을 추가 학습시키는 방법이다. 이를 통해, 소량의 데이터로도 높은 성능의 모델을 구축할 수 있다.

3. Primary extraction through natural language processing

본 시스템에서는 이러한 기사성 광고를 분류하기 위해 지식기반 시스템과 자연어처리 머신러닝기법을 사용할 수 있다. 먼저 지식기반 시스템을 통해 기사 속에 광고성 특징을 갖는 특정한 단어를 확인하여 기사성 광고로 분류하고자 한다.

자연언어처리는 인간의 언어 현상을 컴퓨터와 같은 기계를 이용해서 묘사할 수 있도록 연구하고 이를 구현하는 인공지능의 주요 분야 중 하나이다. 자연 언어 처리는 연구 대상이 언어이기 때문에 언어 자체를 연구하는 언어학과 언어 현상의 내적 기제를 탐구하는 언어 인지 과학과 연관이 깊다. 또한 구현을 위해 수학적 통계적 도구를 많이 활용하며 특히 기계학습 도구를 많이 사용하는 대표적인 분야이다. 정보검색, 문서 자동 분류, QA 시스템, 신문기사 클러스터링, 대화형 Agent 등 다양한 응용이 이루어지고 있다[20].

먼저 KBMS에서는 “If... then...” 형식으로 기사 속에 특정한 단어가 들어 있는지 확인한다. 트리형식으로 분류를 하고 있기에 머신러닝의 단점인 블랙박스 형식으로 결과를 표출하지 않고 간단히 사용자가 납득할 수 있는 이유가 있으므로 결과를 믿을 수 있다. 하지만 특정 단어를 사용하지 않고 은밀하게 광고를 하는 기사도 있을 수 있기에 이는 자연어처리를 활용하여 분류를 한다.

의사결정을 위한 knowledge work process에 의하여 그 결과물인 구조화된 지식은 지속적으로 축적되고 서로 관계를 가지며 지식베이스를 이루게 된다. 또한 이러한 Knowledge work process는 지식베이스에 저장된 지식에 대하여 저장, 검색, 수정, 네비게이션, 공유, 복사, 삭제, 합병, 분리 등의 작업을 수행해야 하는데 이를 위하여 지식베이스관리시스템이 필요하다[24].

4. Secondary extraction system through deep learning

딥러닝은 Machine learning 분류에 포함된 최신 인공지능 학습 분야의 하나로, 기계학습보다는 더욱 깊고 복잡하게 구성된 모델이 입력값과 출력값이 포함된 데이터셋으로 학습할 수 있다[14]. 2020년 현재를 기준으로 음성인식, 텍스트인식, 영상 분류 및 추천 시스템 등에서 두각을 나타내고 있으며, 딥러닝 모델의 성능 향상에 대한 연구와 실생활에 적용하기 위한 응용 연구 모두 매우 활발하게 이루어지고 있다. 기존의 전통적인 기계학습은 모델의 구조가 단순하기에 입력값도 수동으로 전처리를 통해 단순화

해야 적절한 결과를 얻을 수 있지만, 딥러닝은 전처리를 거치지 않아도 복잡한 형태의 입력값을 학습하는 것이 가능하므로 딥러닝 기반 시스템의 일관성 및 성능이 기계학습보다 뛰어나다[15].

Fig. 3은 기사성광고 추출시스템의 전체 구성도를 나타내고 있으며, KBMS에서 1차 단계에 걸친 광고성기사를 배제하기 위한 광고표식 키워드에 의한 기본적인 정보정제와 “딥러닝”에 의한 고급화된 정보추출을 위한 딥러닝을 통한 2차 정제 및 추출단계를 거침으로써 양질의 콘텐츠를 사용자에게 제공하는 구조이다.

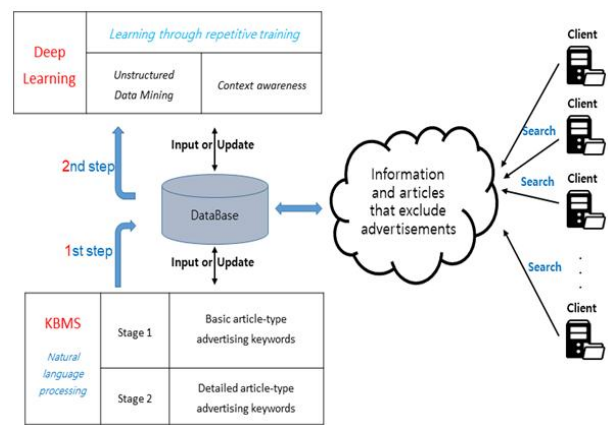


Fig. 3. Overall structure of articleable advertisement extraction system

V. Conclusions

본 논문에서는 기사성광고의 종류와 특징, 파급효과 등의 분석을 통해 알아본 바와 같이, 현재 존재하는 많은 기사에는 일반기사와 기사성광고가 무분별하게 흩어져있고, 이로 인한 소비자의 정보획득에 착오가 발생하게 되어 정확한 기사의 신뢰도는 다소 떨어지게 되며, 이에 따른 반응으로 구독자의 수는 적어지게 되고 광고의 효과와 가치 역시 떨어지게 됨을 알 수 있다.

또한 이러한 무분별한 ‘기사성 광고’로 인하여 우리는 일반기사와 기사성 광고의 구별이 어려운 상황이라서 구독자들은 많은 정보의 오인과 혼란을 하고 있다. 때에 따라 이러한 영향은 해당 신문사의 신뢰성에 치명적인 영향을 끼칠 수 있으며, 미디어 이용자들이 광고를 회피할 수도 있다고 생각한다.

따라서 독자들은 앞으로도 계속 새로운 정보를 획득하고 이러한 정보를 적재적소에 적용하여 많은 가치를 가져와야 하기에, 사용자로 하여금 정확한 일반기사와 기사성

광고를 식별하는 것이 더욱이 중요하다고 판단된다.

그러기에 분명 기사성광고에 대한 부분은 적정하게 규제하여 일반기사와 분류할 필요가 있으며, 소비자입장에서 정보의 착오가 없이 일반기사와 기사성광고를 정확히 구별을 하는 것이 중요하다고 생각한다. 이르기 위해서는 기사성광고의 명확한 법규나 규제에 따른 광고의 발전 및 소비자의 정확한 정보획득에 문제가 없어야 할 것이다.

이러한 부분을 시스템적으로 일부 개선하기 위해서, 본 논문에서는 인공지능시스템의 일부 부류인 자연어처리와 딥러닝기법을 이용하여 필요에 따라 기사성광고를 추출하는 방법에 대해 제시를 하였다.

즉, 이러한 설계의 1차적인 단계는 KBMS의 자연어처리에서 이루어지고 있으며, 이러한 광고성 문구나 특정한 기업 및 제품홍보에 관한 일반적인 키워드를 주어서 필터링하게 하였으며, 또한 좀 더 정확한 정보를 얻기 위한 확장형 키워드를 주어 반복적인 필터링을 거쳐 저장하게 하였다. 그리고 2차적인 단계에서는 이러한 정보를 상황인식을 적용한 자연어처리와 더불어 딥러닝기법을 이용한 좀 더 정확한 정보를 적용하여 기사성광고를 추출할 수 있는 방법에 대해 설계를 하였다.

ACKNOWLEDGEMENT

This work was supported by 2023 Hannam University Research Fund.

REFERENCES

- [1] Bong-seok Yoo, Il-kwon Jeong, "A study on the effectiveness factors of article-type advertisements in news portals" (focusing on adverse effects caused by message exposure environment and perception of deceptiveness), *Information Society & Media*, 2018, Vol. 19, no. 1, 23-45, 2018.
- [2] Van Reijmersdal, E., Neijens, P., & Smit, E. (2005). Readers' reactions to mixtures of advertising and editorial content in magazines. *Journal of Current Issues & Research in Advertising*, 27(2), 39-53, 2005.
- [3] Kim Bong-hyeon. "A study on the form and editing method of article-type advertisements in newspapers: Focusing on content analysis." *"Advertising Research"*, 18(4), 259-288, 2007.
- [4] Newsom, D., Turk, J. Kruckeberg, D., "This is PR - The Realities of Public Relations. CA: Thomson Wadsworth.", 2004.
- [5] Living Economy Bureau Labeling and Advertising Team, Survey on the Status of Article-based Advertisements and Consumer Awareness, November 2001.
- [6] Park Hyeon-gap, "The reality of hate advertising and improvement measures", *Kwanhoon Journal*, Vol. 143, 97-103 (7page). 2017.
- [7] Kim Yeo-ra, "Issues and Points: Problems and Improvement Plans for Article-Type Advertisements," *National Assembly Research Service*, Vol. 1862, 2021
- [8] Kim Do-yeon, "How long are we going to sell 'article-type advertisements' targeting portals?", *Media Today*, 2021, <https://www.mediatoday.co.kr/news/articleView.html?idxno=214142>
- [9] Geum Jun-kyung, "Overflow of article-type advertisements, drawing attention again to the need for punishment", *Media Today*, 2021, <https://www.mediatoday.co.kr/news/articleView.html?idxno=214808>
- [10] Ostrich Desert, "Enforcement of legal crackdown on information posting", 2024, <https://blog.naver.com/m3v239f2/223259024844>
- [11] Cho, C., & Cheon, H. J. "Why do people avoid advertising on the internet?", *Journal of Advertising*, 33(4), 89-97, 2004.
- [12] Ho-jin Choi, Yeon-jun Lee, and Jong-woo Jeon, "Factors influencing hostility toward online article impersonation advertisements and avoidance of advertisements," *Advertising Research*, No. 131, 2021. <http://www.kadpr.or.kr>, Korea Advertising & Public Relations Academy
- [13] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning", *Nature*, Vol. 521, pp. 436-444, 2015.
- [14] Yongjun Kim, Study on improving the recognition rate of unclear license plates based on CNN deep learning image processing, Hannam University master's thesis, 2021.
- [15] Kyu Beom Lee, Hyu-Soung Shin, Seung Hyeon Kim, Dae Mok Ha, Isu Choi, A Study on Automatic Classification of Characterized Ground Regions on Slopes by a Deep Learning based Image Segmentation, *TUNNEL & UNDERGROUND SPACE* Vol.29, No. 6, pp. 508-522, 2019.
- [16] Srivastava N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., Dropout, "a simple way to prevent neural networks from overfitting, *The journal of machine learning research*", Vol. 15, No. 1, pp. 1929-1958. 2014.
- [17] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks", In *Advances in neural information processing systems*, pp. 1097-1105, 2012.
- [18] Nair, V., and Hinton, G. E., "Rectified linear units improve restricted boltzmann machines", In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807-814, 2010.
- [19] Ioffe, S., and Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *arXiv preprint arXiv:1502.03167*, 2015.
- [20] Deep Learning (Definition and Importance of Deep Learning),

2023. https://www.sas.com/ko_kr/insights/analytics/deep-learning.html

- [21] Kim Chung-hyeon, "Current status and problems of newspaper camouflage advertisements," Kwanhoon Journal, Volume 72, No. 74, pp. 146 - 154 (9 pages), 1999.
- [22] Yu Bong-seok and Jeong Il-kwon. "A In-depth Interview Study on the Current Status and Problems of Online Editorial Ads", Korean Journalism Journal, 59(4), p. 234, 2015
- [23] Kang Mi-seon, "The current state of newsworthy advertising and improvement measures: integrated application of surveys, content analysis, and overseas case studies," Korea Advertising Journal, No. 5-1, 2013
- [24] Lee Dong-won, Kim Jae-jeong, Jeong Bang-cheol, Lee Tae-eok, Cho Dong-ho, "Future Knowledge Service Overview and Development Direction", Journal of Information Science Vol. 27, No. 7, pp.18-24, July 2009.

Authors



Seoksoo Kim received the B.S. degrees in Computer Science and Engineering from Kyungnam University, in 1989, and M.S., Ph.D. degrees in Computer Science and Engineering from Sungkyunkwan University,

Korea, in 1991 and 2002, respectively. Dr. Kim joined the faculty of the Department of Multimedia Engineering at Hannam University, Daejeon city, Korea, in 2003. He is currently a Professor in the Department of Media & Visual Communication, Hannam University. He is interested in Multimedia, Web based information system, Augmented reality, artificial intelligence, bigdata, machine learning, and database.



Jae-Young Jung received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Sungkyunkwan University, Korea, in 1989, 1993 and 1997, respectively. Dr. Jung joined the faculty of the Department

of Computer Science at Dongyang University, Yeongju city, Korea, in 1997. He is currently a Professor in the Department of AI Bigdata convergence, Dongyang University. He is interested in artificial intelligence, bigdata, machine learning, pattern recognition and motion tracking.