

# SNS 환경에서의 지능형 실시간 광고 효과 분석 시스템 설계

이아름<sup>o</sup> 방지선 안윤선 김윤희

숙명여자대학교 컴퓨터학과

arsbsp@gmail.com, nin23bix@gmail.com, ahnysun1@gmail.com, yulan@sm.ac.kr

## A Design of Analysis System for Intelligent and Real-time Advertisement

### Effectiveness in SNS Environments

A-Reum Lee<sup>o</sup> Jiseon Bang Younsun Ahn Yoonhee Kim

School of Computer Science, Sookmyung Women's University

#### 요 약

SNS(social networking service)에서의 광고에 대한 개인의 반응은 실시간 생성된다. 이를 통해 개인의 관심 분야의 수집과 분석이 용이해졌으며, 많은 양의 정보를 활용할 수 있게 되었다. 본 시스템에서는 특정 광고에 대한 소셜 네트워킹 서비스에서의 개인의 반응을 데이터로 수집하고, 방대한 양의 반응 데이터를 효율적으로 분석하기 위하여 Hadoop을 통한 분산 처리를 하여 실시간에 가까운 분석을 한다. 또한, 보다 신뢰성 높은 분석을 위해 온톨로지를 이용한 지능형 추론 방법을 도입하여 유사 반응 효과 분석을 가능토록 한다. 시스템이 수집한 데이터를 통해 사용자로부터 학습하여 점차 지능화되는 방향으로 확장되어 정확도와 신뢰도가 높아지고 추론이 가능하도록 설계한다.<sup>1)</sup>

#### 1. 서 론

SNS의 사용자 증가로 인해 사이버 스페이스상의 의사소통과 교류가 확대됨에 따라 SNS의 방대한 데이터를 분석하여 다양한 분야에 활용하는 연구가 활발하다. SNS의 확산과 동향[1]에 따르면 한국 인터넷진흥원이 실시한 2010년 인터넷 이용실태 조사에서 SNS이용자가 65.7%로 높은 부분을 차지하고 있다고 한다. SNS 이용자 수는 지속적으로 증가해왔으며 앞으로도 증가해 갈 것으로 추정된다. 이와 같이 점차 확산되고 있는 SNS는 신속성, 개인성, 정보의 개방성과 같은 특성을 지닌다. 이에 근거하여 SNS를 분석했을 때보다 개인적이고 진실한 정보를 얻을 수 있을 거라고 추측할 수 있으며 실시간으로 빠르게 갱신되는 정보들을 분석할 수 있다는 장점을 가진다. 따라서 SNS를 분석함으로써 실시간으로 유효한 데이터를 통해 광고 효과를 분석할 수 있을 것으로 판단하였다.

SNS의 빅데이터를 실시간으로 분석하기 위해서는 SNS의 정보를 수집하고, 빅데이터를 분석하기 위한 별도의 시스템이 구성되어있어야 한다. 따라서 기존에 존재하지 않던 실질적인 시스템이 개발되어야 하며, 해당 시스템은 기존의 광고 분석 기법에 의거하여 SNS의 정보를 분석하고 처리하여 타당한 결론을 낼 수 있어야 한다. 이를 위해서, 본 논문에서 제안하는 시스템은 하둡을 이용하여 빅 데이터를 빠르게 분석한다. 빅데이터는 그 방대한 양으로 인하여 기존 시스템으로는 분석에 시간 소모가 매우 많이 든다는 단점이 있다. 반면, 하둡을 이용하

면 빅 데이터를 짧은 시간 내에 분석할 수 있기 때문에, 이를 이용하여 광고 효과 분석에 이용하면 분석 시간을 크게 단축할 수 있다.

한편, 제안하는 시스템은 광고효과 측정방법에 있어서도 새로운 접근법을 제시하고 있다. 기존의 광고효과 측정방법은 일일 후 회상 조사(DAR)와 같은 설문조사나 광고 인지도 조사가 대부분이었다[2]. 이는 정확도가 떨어지고 실제 광고 직후 반응을 얻기가 힘들다. 이에, 광고에 대한 반응을 실시간으로 얻을 수 있는 프로그램의 필요성이 대두되었다.

따라서 본 논문에서는 광고 반응 분석과 광고 효과 판단에 있어 신뢰성 높고 정확한 처리가 가능한 시스템을 제작하기 위해, 이전 논문[3]에서 구현했던 실시간 광고 효과 분석 시스템에 온톨로지를 적용하여 의미 데이터의 관계를 구성하고, 의미론적 추론 방법을 활용하여 기존 정보를 통해 새로운 정보를 자동으로 생성시키도록 한다. 이로 인해 기존의 방식보다 더 빠르면서 정확한 처리를 가능하게 하며, SNS 데이터를 단순 매칭이 아닌 의미에 따른 추론 방법을 활용하여 광고에 대한 반응을 분석할 수 있는 실시간 광고 효과 분석 시스템을 설계하고자 한다. 본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 관련 연구들을 살펴보고, 3장에서는 제안하는 시스템의 기능 설계에 대해 설명하며 마지막으로 4장에서 결론을 맺는다.

#### 2. 관련연구

1) 이 논문은 2014년 정부재원(미래창조과학부 여대학(원)생 공학연구팀제 사업)으로 한국연구재단과 한국여성과학기술인지원센터의 지원을 받아 연구되었습니다.

TV광고에 대한 광고 효과 분석 기법으로는 셋톱박스를 설치하여 실시하는 시청률 조사가 있다. 하지만 조사기관이 민간 기업이고, 그 수가 너무 적어 독점에 대한 문제제기가 있어왔다. 또한, 조사 기준이나 절차에 있어서도 대표성이 미흡하고 비일관적이라는 단점이 있었으며, 통계적 오차 범위에서 벗어나 신뢰도에 있어서도 의심 받고 있다[4].

현재 SNS 정보를 수집, 분석하여 보여주는 사이트로써 트윗트렌드[5]와 소셜메트릭스[6]가 존재한다. 하지만, 트윗트렌드의 경우 분석결과가 매우 적어 정확한 분석이 어려웠으며, 소셜 메트릭스의 경우 다양하게 분석이 가능했으나, 여러 광고에 대한 비교는 불가능했다. 또한 두 사이트 모두 분석에 일정시간의 소요가 있었다.

하둠을 이용한 소셜 네트워크 분석에 대한 논문은 2012 년부터 발표되어왔다. Hadoop을 이용한 트위터 메시지 분석 시스템 설계[7]에서는 하둠을 이용하여 소셜 네트워크 데이터를 이용한 분석이 가능함을 제시하고 하둠 시스템과 한글 형태소 분석 프로그램을 도입한 시스템 모델을 설계함으로써 하둠이 SNS 정보를 빠르게 처리할 수 있다는 가능성을 열어 주었다. 하지만 속도 개선 이외의 다른 성능에 대한 언급이 없었고, SNS 분석에 있어 구체적인 결과 분석도 미흡했다. 하둠을 이용한 소셜 네트워킹의 TV광고효과 분석 시스템 설계[8]에서는 한층 더 발전된 형태로, 직접 파싱을 통해 데이터를 전달 받고, 하둠 시스템을 이용하여 광고 효과를 그래프와 수치로 분석하였다. 이 시스템은 하둠을 이용하여 광고효과를 분석하는 프로그램을 처음 만들었다는 점에서 의의가 있었지만 다양한 형태의 분석이 불가능했다.

한편, 온톨로지는 지식을 기반으로 하여 구축할 수 있는 기본 구조를 제공하는 도메인 개념들의 명확한 표현이라고 정의할 수 있다[9]. 온톨로지는 또한 정보시스템의 대상이 되는 자원의 개념을 명확하게 정의하고 상세하게 기술하여 보다 정확한 정보를 찾을 수 있도록 하는데 목적이 있다. 이를 이용하면 자원에 대한 정보의 질을 향상시킬 수 있다. 때문에, 광고 효과 분석에 있어 더 정확한 정보와 예측을 위해서는 온톨로지 기술 접목이 필요하다.

이에 본 논문은 광고 효과를 언급도 이외의 여러 항목에 대해 분석하고, 단일 광고가 아닌 여러 광고를 비교할 수 있도록 한다. 뿐만 아니라, 실시간 광고 효과 분석 시스템에 온톨로지를 적용하여 광고 반응 분석과 광고 효과 판단에 있어 신속하고 정확한 처리가 가능한 분석 프로그램을 만들고자 한다.

### 3. 기능 설계

본 논문에서 제안하는 시스템의 구조는 그림 1과 같다. 전체 시스템은 이전 연구[3]에서 구현하였던, 트위터에서 사용자들의 글 내용을 시간정보와 함께 수집하는 파서, 조건에 맞는 글의 개수를 계산하여 분석결과를 얻는 하둠 프로그램, 이렇게 계산된 정보를 바탕으로 그래프를 보여주는 웹 인터페이스로 이루어지며 프로세싱 과

정에서 온톨로지를 도입하여 시스템을 확장한다.

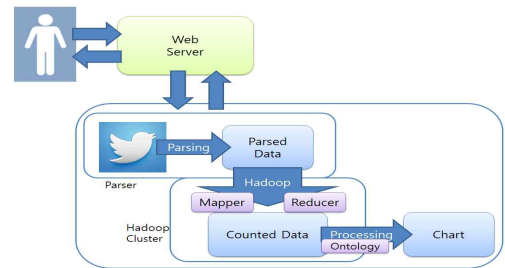


그림 1 시스템 구조도

파서는 트위터 글과 광고정보 사이트 데이터를 수집한다. 파서로는 Jericho, Jsoup HTML Parser를 이용하여 정보를 각각 다른 방식으로 수집, 저장한다. 제안하는 시스템에서 TV광고 정보를 수집하는 사이트는 TVCF라는 사이트로, 1950년부터 현재까지 국내외 TV광고를 카테고리 분류해 놓았다[10]. 각 데이터는 빠른 업데이트를 위해 한 시간에 한 번씩 crontab을 이용하여 파싱을 수행하며 트위터에서는 글쓴이와 내용, 게시시각을, TVCF에서는 광고명과 카테고리, 광고시각을 얻어와 트위터 데이터 키워드와 매핑 한다.

하둠 클러스터는 트위터 데이터의 파싱 결과를 세 가지 시나리오 별로 조건에 맞는 항목의 개수를 계산한다. 여기에서 세 가지 시나리오에 대한 설명은 다음과 같다.

- 1) 가장 효과적인 방송 시간을 알고 이후 활용하기 위해 특정 광고에 대한 시간대별 분석을 실시한다.
- 2) 효과적인 광고인지를 측정하기 위해 광고에 대해 온톨로지 기반 키워드 분석을 통한 긍정적/부정적 반응 정도를 분석한다.
- 3) 동종 광고 중에서 가장 효과적인 광고를 알기 위해 동종 광고의 긍정적/부정적 반응 정도를 분석한다.

한편, 이용자가 광고주인 경우에는 광고 제작 시 활용할 수 있도록 광고를 추천해 주는 기능과, 루머 발생 시 초기 대응을 위해 이를 미리 알려주는 루머알림 기능의 두 가지 기능을 추가적으로 제공한다. 여기에서 광고 추천 기능은, 동종 광고 중 이전까지의 분석 결과가 가장 긍정적이었던 광고를 추천해 주는 형식으로 구현하였으며, 루머 알림 기능의 경우는 온톨로지 디셔너리에 설정된 특정 키워드가 단기간 안에 일정 횟수 이상 언급되었을 시 이를 미리 알려주는 형식으로 확장하여 구현한다.

다섯 가지 기능을 구현하기 위해서 제안하는 시스템에서는 하둠을 이용한다. 하둠은 파싱을 통해 얻은 정보 중에서 조건에 맞는 항목의 개수를 세어준다. 하둠 프로그램은 맵과 리듀스 과정을 통해 이루어지는데, 트위터 데이터를 입력 값으로 넣으면, 맵에서는 시나리오에 따라 시간이나 조건 별 항목의 값이 value가 되고, 조건의 이름이 key값이 되도록 한다. 리듀스 과정에서는 이렇게 생성된 값을 통해 실제 개수의 계산이 이루어진다. 맵리듀스 과정을 통해 얻어진 정보는 key, value 쌍으로 정렬되어 지정된 경로에 저장되도록 하였다.

첫 번째 기능은 시간대별로 광고의 언급도를 분석하여 도표화하는 것이다. 이는 얼마나 많은 사람들이 해당 광

고에 대해서 알고 있는지를 나타내기 위하여 설계하였다. 제안하는 시스템에서는 이를 날짜별, 시간별로 그래프화 하고 두 시간 단위로 표현하여 정확도를 높였다. 시간대별 분석의 결과는 그림 2처럼 동종 광고에 대한 번에 확인하거나 광고 별로 확인 할 수 있게 하였다.

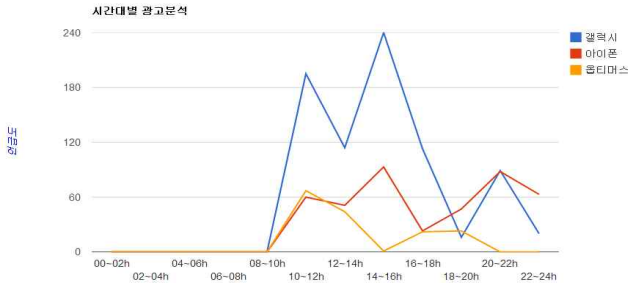


그림 2 시간대별 광고 분석(웹)

두 번째 기능은 트위터에 올라와 있는 글의 선호도를 그래프화 하는 것이다. 선호도는 글의 반응을 토대로 분석되는데, 여기에서 의미가 비슷한 유사데이터를 분석하기 위해 온톨로지를 적용할 필요가 있다. 반응에 대한 온톨로지를 구축하여 유사한 의미, 다양한 견해에 대한 의미론적 방법을 이용한 분석이 가능토록 한다. 또한 긍정, 부정, 중립 분류 중 유사한 의미의 차이를 알 수 있는 유사도 계산법을 적용하여 다양한 분석이 가능하도록 한다. 반응 데이터의 확장을 위하여 SNS에서 실시간으로 수집한 데이터와 기존의 디셔너리에 있는 데이터를 추가하고, 추가된 데이터의 신뢰도를 평가하는 피드백을 통해 정확성을 높일 수 있다. 그림 3은 구현한 앱 시스템의 선호도 분석에 대한 결과로, 긍정, 부정 반응을 원 그래프로 표현하였다.



그림 3 선호도 분석(앱)

세 번째 기능은 동종 광고 중 어떤 광고가 더 선호도, 비선호도가 높은지를 그래프화 하는 것이다. 이를 통해, 같은 업종에서 어떤 광고가 더 긍정적인 반응을 이끌어 냈는지 분석한다.

시스템은 수집된 15.7MB의 총 92,809개 트위터 데이터를 바탕으로 1대의 Master 서버와 3대의 Slave 서버에서 분석하도록 구현되었다. 시스템의 분석 결과로 현 광고에 대한 평가와 광고 방송시기에 대한 적절성을 파악하고 방송 실패를 줄이며 광고를 객관적으로 평가할 수 있다. 또한, 이를 다음 차수의 광고 제작 시 반영하여 더욱

과급력 있는 광고 제작과 효율적인 광고 활용 계획을 수립할 수 있게 한다. 방송 이후에는 광고 목표가 달성되었는가에 대해 객관적 판단 기준이 되어 광고의 질을 높이는 데 일조할 수 있다.

#### 4. 결론 및 향후 연구

본 논문에서는 SNS에서 추출한 데이터로 하둡 분산시스템을 통해 실시간으로 처리하여 원하는 결과를 웹에서 그래프로 보여주는 시스템을 제작하였으며, 온톨로지를 이용한 지능형 추론 방법을 도입하여 유사 반응 효과 분석이 가능하도록 설계를 확장하였다.

이 시스템에 대한 평가는 사용자의 요구가 있을 때, 미리 저장된 데이터에서 얼마나 빠른 시간 내에 데이터를 분석하여 도표화 하는지, 그리고 온톨로지에 구축된 디셔너리로 인해 얼마나 정확도 높은 분석이 가능한 지로 제시될 수 있을 것이다.

본 논문에서 제안한 시스템은, 수집한 데이터를 통해 사용자의 반응으로부터 스스로 학습하여 점차 지능화하는 방향으로 확장되게 된다. 이를 통해 정확도와 신뢰도가 점차 높아지고 추론이 가능한 시스템이 될 것으로 기대된다.

향후에는 트위터 데이터 뿐 아니라 다른 SNS 정보를 이용하여 더 많은 데이터를 통해 사용자 반응을 분석하는 것이 필요하다. 이를 통해 더 정확한 분석이 가능할 것이다. 또한 다섯 가지의 기능 외에 더 다양한 기능을 추가한다면 더욱 발전된 광고 분석 시스템이 제작될 수 있을 것이다.

#### 참고 문헌

- [1] 이진형, “SNS(Social Network Service)의 확산과 동향”, 인사관리 통권271호, 2012
- [2] You-Jae Lee, “Measuring Advertising Effect”, <http://youjae.com/data/cdata3/20/ad13.pdf>
- [3] 방지선, 이아름, 옥윤청, 김윤희, “소셜 네트워크에서의 Hadoop 기반 실시간 광고 효과 분석 시스템 설계”, 한국정보처리학회 KIPS 춘계학술발표대회, 2014
- [4] 배효승, 신소연, 이상우, “IPTV 셋톱박스 로그분석을 통한 시청률 연구”, 방송문화연구 제 24권 제1호, 2012
- [5] <http://tweetrend.com/>
- [6] <http://insight.some.co.kr/>
- [7] 송지훈, 이시진, 박효동, “Hadoop을 이용한 트위터 메시지 분석 시스템 설계”, 한국인터넷정보학회 2012년도 하계학술발표대회 논문집, 2012
- [8] 허서연, 김윤희, “하둡을 이용한 소셜네트워킹의 TV광고효과 분석 시스템 설계, 인터넷정보학회논문지 제14권 제6호, 2013
- [9] 강혜정, 김윤희, “계산 응용 온톨로지를 이용한 실험 환경 확장 기법” 정보과학회논문지: 컴퓨팅의 실제 및 레터, 18.10: pp685-691, 2012
- [10] <http://www.tvcf.co.kr/>