

# 빅데이터 분석을 통한 유명인 모델의 광고효과 예측 모형 개발

김유나<sup>1</sup>, 한상필<sup>2\*</sup>

<sup>1</sup>서울예술대학교 광고창작과 교수, <sup>2</sup>한양대학교 광고홍보학과 교수

## Development of a Prediction Model for Advertising Effects of Celebrity Models using Big data Analysis

Yuna Kim<sup>1</sup>, Sangpil Han<sup>2\*</sup>

<sup>1</sup>Professor, Department of Creative Advertising, Seoul Institute of the Arts

<sup>2</sup>Professor, Department of Advertising and PR, Hanyang University

**요약** 본 연구는 소셜 빅데이터에 기반을 둔 유명인과 브랜드의 이미지 유사도가 광고효과를 예측할 수 있는 결정변수가 될 수 있는지를 파악하기 위해, 광고효과 예측모형을 생성하고 빅데이터 분석기법인 기계학습 방법을 통해 그 타당도를 검증하는 것을 목적으로 하였다. 이를 위해 SNS상의 키워드 네트워크 구조에 기반하여 유명인-브랜드 이미지 유사도를 정량화하고, 학습 데이터를 통해 이미지 유사도를 독립변수로, 광고효과 데이터를 종속변수로 하는 다중회귀모형을 반복 실시하여 광고효과 예측모형을 생성하였다. 이렇게 생성된 예측모형의 정확도를 판단하기 위해 예측 데이터에서 얻은 광고효과 예측값과 비교 기준으로서의 서베이값을 비교한 결과, 타당도를 판단하는 기준치인 75%의 분류 정확도를 보였으므로 본 광고효과 예측 모델링의 타당성은 확보된 것으로 입증되었다. 본 연구는 유명인-브랜드 이미지 유사성 구조를 소셜 네트워크 구조로 설명하고 그 효과를 기계학습을 통한 예측 모델링으로 검증하여 빅데이터 기반 모델링 연구에 새로운 방법론적 대안과 방향을 제시하였다.

**주제어** : 유명인-브랜드 이미지 유사성, 광고효과 예측모형, 기계학습(머신러닝), 소셜 빅데이터, 연상 네트워크 이론

**Abstract** The purpose of this study is to find out whether image similarity between celebrities and brands on social network service be a determinant to predict advertising effectiveness. To this end, an advertising effect prediction model for celebrity endorsed advertising was created and its validity was verified through a machine learning method which is a big data analysis technique. Firstly, the celebrity-brand image similarity, which was used as an independent variable, was quantified by the association network theory with social big data, and secondly a multiple regression model which used data representing advertising effects as a dependent variable was repeatedly conducted to generate an advertising effect prediction model. The accuracy of the prediction model was decided by comparing the prediction results with the survey outcomes. As for a result, it was proved that the validity of the predictive modeling of advertising effects was secured since the classification accuracy of 75%, which is a criterion for judging validity, was shown. This study suggested a new methodological alternative and direction for big data-based modeling research through celebrity-brand image similarity structure based on social network theory, and effect prediction modeling by machine learning.

**Key Words** : Celebrity-brand image similarity, Advertising effect prediction model, Machine learning, Social big data, Association network theory

This article is a summary of the first author's doctoral thesis.

\*Corresponding Author : Sangpil Han(sphan@hanyang.ac.kr)

Received May 25, 2020

Accepted August 20, 2020

Revised July 19, 2020

Published August 28, 2020

## 1. 서론

광고 모델은 광고효과를 좌우하는 핵심 구성 요인이면서 광고비의 많은 부분을 차지하는 요소이다[1]. 지금까지 연구자들은 유명인이 가지고 있는 이미지와 브랜드 이미지의 적합성을 주요 연구 주제로 다루었으나[2], 실제로 유명인 모델과 브랜드의 이미지 적합성에 대한 실체가 무엇인지, 어느 정도로 적합해야 하는지, 그리고 이미지 적합성을 고려하여 유명인을 선정했을 때 실제로 광고효과를 담보할 수 있을지에 대한 근거들은 여전히 부족한 실정이다. 최근에는 유명인의 사생활이나 개인적인 활동들이 인터넷상에서 쉽게 노출되면서, 어디까지가 유명인의 공적인 이미지이고 어디까지가 일반인으로서의 사적 이미지인지 구분하기 어려울 정도로 복합체의 형태를 띠고 대중에게 전달되고 있다. 따라서, 유명인의 이미지 연구에서 이미지를 고정된 형태의 '표상'으로 보는 것을 넘어, 생활 맥락 속에서 변화무쌍하게 나타나는 '복합체'로서의 측면을 고려한 새로운 접근이 필요하다고 하겠다.

본 연구의 목적은 첫째, 최근 주목받는 소셜 빅데이터를 활용하여 유명인과 브랜드에 대한 소비자 담론으로 유명인 모델의 이미지 실체를 파악하고, 둘째, 유명인-브랜드 이미지 유사도 연구를 통해 복합체로서의 유명인 모델에 대한 이미지 구조를 새롭게 정립하며, 셋째, SNS 상의 키워드 구조를 연상 네트워크 구조로 접근해 볼 수 있다는 점에 착안하여 유명인과 브랜드의 이미지 유사도 산출 방식을 도출하는 것이다. 마지막으로, 유명인-브랜드 이미지 유사도를 독립변인으로 하여 실질적으로 광고효과를 예측할 수 있는지, 유명인 모델에 대한 광고효과 예측모형을 생성하고 그 타당성을 검증하고자 한다.

## 2. 문헌 연구

### 2.1 유명인 모델과 브랜드의 이미지 유사성

유명인(celebrity) 모델은 일반적으로 특정 분야에서의 성공을 기반으로 대중들에게 이름을 널리 알린 인물을 의미한다[3]. 유명인 모델에는 가상의 인물도 모두 포함되는데, 최근에는 SNS(Social Network Service)상에서 막대한 팔로워(follower)를 보유하면서 트렌드를 이끌고 유명세를 얻고 있는 인플루언서(influencer)도 유명인 모델에 속한다고 하겠다.

이러한 유명인 모델이 가진 이미지와 모델이 광고하는 브랜드 이미지 간의 유사도가 높을수록 광고효과가 높아

진다는 연구 결과들이 있다[4]. 하지만 기존의 연구들은 브랜드 이미지를 측정하는 데 있어서 사전에 형성해 놓은 이미지 셋(set)을 설문 조사를 통해 평가하는 방식이기 때문에 측정에 한계를 갖는다[5]. 또한, 오늘날같이 SNS 커뮤니케이션이 활발한 쌍방향 네트워크 시대에는 고정된 '표상으로서의 이미지' 외에 서사적 맥락을 고려한 '복합체로서의 이미지'로 유명인 모델과 브랜드 이미지를 확장해서 보는 접근이 필요하다고 하겠다.

#### 2.1.1 표상으로서의 이미지

이미지를 구성하는 단위로서 '연상'의 중요성을 강조하는 연구들은 기본적으로 이미지를 '표상'의 관점으로 분석한다. 이미지의 형태는 그 연상들이 어떻게 조직화되어 있는지에 따라 달라지는데, 이러한 연상의 조직화는 인지심리학에서 논의되는 지식의 네트워크 구조로 설명될 수 있다[6,7]. 연상 네트워크 모델은 특정 기억과 지식이 노드(node)와 링크(link)의 집합으로 형성되어 있다고 설명하며, 이러한 구성 요인들은 네트워크 형태로 연결되어 있다고 주장한다[8,9]. 또한, 특정 연상이 얼마나 강력한 브랜드 지식을 이루고 있는지는 연상의 노드 크기와, 해당 연상이 얼마나 브랜드와 강력하게 연결되어 있는지를 나타내는 링크의 거리와 강도로 결정된다고 말한다.

네트워크 분석은 본래 사회학에서 시작된 방법론으로, 사람들 간의 연결 관계를 표현한 네트워크 구조를 분석하고 이를 기반으로 다양한 사회현상을 해석하기 위해 사용되어져 왔으며[10], 최근에는 브랜드 연상 구조를 파악하는 데도 활용되고 있다. 이러한 맥락에서 SNS상에서 생성되는 소비자 여론 역시 수많은 네트워크를 통해 확산되는 경향을 보이므로 네트워크 분석으로 그 의미에 접근해 볼 수 있겠다. 따라서 본 연구에서는 소셜 빅데이터를 통해 수집한 유명인 모델의 이미지와 브랜드 이미지의 키워드를 연상의 기본 단위로 여기고, 연상 네트워크 이론에 근거하여 유명인-브랜드 간 이미지 유사도 분석하고자 한다.

#### 2.1.2 복합체로서의 이미지

대상에 대한 이미지는 그 배후에 존재하는 사건을 통해서야만 실체가 명확하게 드러나는 법이다. 대상의 의미를 사건의 맥락 속에서 파악하는 방식은 구조주의의 의미론적 접근에 근거해서 설명할 수 있다. 대상의 이미지를 구조적 관점으로 연구한다는 것은 전후좌우의 맥락 속에서 존재하는 서사적 접근을 통해 그 이미지의 실체

에 다가감을 뜻한다. 또한, 단일한 정체성이 있다기보다, 시간의 흐름에 따른 행동과 상황에 대한 대응 방식, 그리고 타인과의 대화 등이 서사적 정체성을 만들어 나갈 수 있음을 의미한다[11].

최근에는 트위터, 인스타그램, 페이스북 같은 소셜미디어를 통해서 유명인에 대한 소식과 정보를 실시간 탐색할 수 있게 되면서, 유명인이 가진 모델로서의 이미지 외에 일상생활 속 개인적인 모습도 많이 접하게 되었다. 이러한 시대적 상황에 따라 이제는 유명인의 이미지를 표상적 차원에서만 보는 것에서 나아가 생활 맥락 속의 서사적 관계성을 고려한 복합체적 이미지로 접근할 수 있는 여건이 갖춰지게 된 것이다. 여기서 복합체적 이미지로 접근한다는 것은 특정 개체가 갖는 이미지를 '개체' 자체, 개체의 행동 결과로 얻어지는 '역할', 개체의 행동이 서사적 구조 속에서 드러나는 '사건'의 연결성 속에서 해석하는 것을 말한다.

따라서 본 연구에서는 유명인 모델의 개체로서의 이미지를 '일반인으로서의 이미지'로 규정하고, 유명인으로서의 사회적 위상에 따라 형성되는 이미지를 '유명인으로서의 이미지'로 규정하며, 서사적 구조로 드러나는 유명인의 행동을 '사건'으로 규정하여 그 실체를 파악해 보고자 한다.

## 2.2 빅데이터를 통한 효과 예측 방법

빅데이터 분석은 무수하고 다양한 데이터들의 관계를 컴퓨팅 기술을 활용하여 빠르고 동시적으로 분석하여 새로운 가치를 창출하는 데이터 처리 방식이다[12]. 빅데이터를 활용한 분석 방법은 마이닝(mining) 기법과 소셜 네트워크 분석, 클러스터로 분류되며, 마이닝은 텍스트 마이닝과 데이터 마이닝으로 분류된다. 텍스트마이닝은 비정형 데이터를 자연어처리 기법을 이용하여 유용한 정보를 추출 혹은 가공하는 것이고, 데이터마이닝은 빅데이터 내에 존재하는 데이터 간 유사성, 패턴, 관계 등을 파악하여 유의한 정보를 도출하여 의사결정에 활용하는 기술을 말한다. 데이터마이닝은 크게 분류(classification), 추정(estimation), 예측(predication), 연관분석(association analysis), 군집 분석(clustering) 등의 목적으로 수행된다.

최근에는 빅데이터를 활용한 예측 모델링에 관심이 높아지면서 기계학습(machine learning) 기법에 대한 활용이 늘고 있다. 기계학습은 인공지능 기술의 하나로, 기계가 데이터로부터 여러 번의 시행착오를 거치면서 프로그램 스스로 학습하여 성능을 향상하는 기술이다. 기계

학습은 지도 학습(Supervised Learning) 그리고 비지도 학습(Unsupervised Learning)으로 분류된다. 지도학습은 학습 데이터에 레이블이 있는 경우로 문제와 이에 대응되는 답을 다량으로 제시한 후에 특정 문제에 대한 답을 추출하는 방식을 말하고, 비지도 학습은 학습 데이터에 레이블이 없는 경우로서 정답이 제시되지 않은 상태에서 문제만으로 학습을 진행하는 방식을 말한다. 예측 모델링을 위해 활용되는 기계학습 방법들은 분석 목적에 따라 여러 가지로 나뉘지만 주로 다중선형회귀, 의사결정나무, 인공신경망, 딥러닝 등의 다양한 기법들이 활용된다[13-15].

## 3. 연구 방법

### 3.1 연구 문제 및 제안모형의 정의

본 연구는 유명인 이미지를 복합체적인 관점으로 접근하여, 유명인-브랜드 이미지 유사성을 소셜미디어 내의 키워드 네트워크에 기반하여 산출한 후, 이미지 유사성을 독립변수로 하여 유명인의 광고효과 예측모형을 구축하는 것이 목적이다. 따라서 해당 목적을 달성하기 위해 다음의 연구 문제를 도출하였다.

- 연구 문제1. 소셜 네트워크상에서의 소비자 담론에 근거한 유명인과 브랜드의 이미지 유사도는 어떻게 산출하는가?  
 연구 문제2. 유명인 모델-브랜드 이미지 유사도를 독립변수로 한 유명인 광고효과 예측모형은 설명력이 있는 타당한 모형인가?

이러한 연구문제를 해결하기 위해, 다음과 같은 두 단계의 절차를 거쳐 분석 모형을 설계하였다. 첫째, 유명인 모델과 브랜드 사이에 공통 키워드가 얼마나 유사한 패턴으로 나타나는지 정량화할 수 있도록 유명인-브랜드 유사도 알고리즘을 산출한다. 둘째, 유명인 모델-브랜드 이미지 유사도가 광고효과에 영향을 주는 주요 변수가 되는 것을 파악하기 위해 기계학습 방법을 통해 유명인 모델의 광고효과 예측 모델링을 실시한다.

### 3.2 빅데이터 수집 모듈

본 연구에서 소셜 빅데이터를 수집하기 위한 채널은 정제된 용어를 사용하여 수집과 분석이 용이한 온라인 뉴스 사이트로 한정하였다. 본 예측 모델링을 위해서 광

고효과에 대한 공신력 있는 과거 지표들이 준비되어야 하므로, 분석 대상으로서 유명인 모델은 한국광고종합연구소(KAA)에서 실시한 광고 모델 선호도 조사 결과(2014년 5월)를 참고하여[1], 유명인 여자 모델 3명(전지현, 김연아, 김태희)을 선정하였다.

또한, 당시 자료가 최신성이 떨어진다는 단점과 선정된 유명인 모델은 공통으로 전자제품 카테고리에 국한(김연아 - 삼성 스마트에어컨 Q9000, 전지현 - 삼성 지펠 T9000, 김태희 - LG 디오스 김치톡톡) 되어 있었기 때문에, 2015~2018년도 화장품 광고에 출연했던 유명인 모델과 광고(전지현 - 헤라, 한가인 - 동인비, 김태리 - 오뚜기, 이성경 - 라네즈, 송혜교 - 설화수)를 추가하여, 최종적으로 총 7명의 여자 유명인 모델을 분석 대상 모델로 선정하였다. 또한, 정확한 예측력 학습을 위해 유명인 별로 ROI가 가장 높았던 광고를 추가로 선정하였다(김연아 - 맥심 화이트골드 커피믹스, 전지현 - SKT 텔레콤 LETA, 김태희 - LG 디오스 정수기 냉장고).

이때, 11건의 선정된 광고물 중에서 광고 집행 전 소셜 데이터가 충분하지 않은 2개의 케이스(신규 모델이나 신규 브랜드를 대상으로 하여 소비자 여론이 충분치 않은 경우)를 제외하고 총 9건의 광고물(맥심 화이트골드, 삼성 스마트, LG 디오스 김치톡톡, 설화수, 라네즈, SK텔레콤, 삼성 지펠, 헤라, 동인비)로 최종 분석하였다.

### 3.3 빅데이터 전처리 모듈

수집된 문서에서 유의한 분석 키워드를 산출하기 위해서는 불필요한 정보를 제거하고 비정형 데이터를 정형 데이터 구조로 변환하는 데이터 전처리 과정이 수반되어야 한다. 따라서 빅데이터 전처리 모듈에서는 수집된 문서에서 목적에 부합하는 분석 키워드를 추출하기 위해 형태소 분석을 통해 불필요한 정보인 가비지(garbage)를 제거하는 데이터 정제 작업을 수행하였다. 또한, 정제된 비정형 데이터를 정형 데이터로 바꾸는 분석을 하였다.

### 3.4 빅데이터 통합 적재 모듈

빅데이터 통합 적재 모듈에서는 유명인과 브랜드를 각각 검색한 결과 산출된 키워드 중에서 공통으로 나타나는 키워드를 하나의 논리적인 저장 구조로 통합하여 저장한다. 이때, 모델의 표상적 이미지 외에 사건 맥락을 고려한 복합체적 이미지를 함께 고려하여 데이터화 하기 위해 '개체-역할-사건' 구조에 근거하여 데이터를 적재하였다. 또한, 소셜데이터의 특성에 따라 복합체적 이미지

는 무수하게 다양한 키워드들로 추출되기 때문에, 개별 키워드를 일반화 가능한 분류 기준으로 유형화하여 예측 모델링에 투입하였다.

### 3.5 이미지 유사도 산출 모듈

이미지 유사도 산출 모듈에서는 앞서 선정한 공통 키워드의 분류 기준에 따라 유명인-브랜드 유사도를 측정하게 된다. 우선 유명인과 브랜드의 1차 연관어 중 상위 20개 키워드를 도출하고, 각 키워드에 대해 1차 연관어와 독립되게 추출한 2차 연관어 중 상위 20개 키워드를 추가로 추출하여 공통 키워드 셋(set)을 준비하였다. 그리고 이 공통 키워드가 전체 문서에서 얼마나 자주 출현했는지, 동시 노출된 횟수를 기준으로 유사도 산출 식을 도출하였다.

이때, 유명인-브랜드 유사도 점수를 산출하는데 연산 네트워크 이론의 기본 개념을 활용하였다. 즉, 유사도 강도를 산출하기 위해 네트워크 링크의 굵기와 거리를 고려해야 하는데, 굵기는 노출 빈도(frequency)로, 거리는 공통 키워드 도출까지의 깊이(depth)로 계산하였다. 예측모형에 투입할 독립변수를 도출하기 위해, 분류 기준별 유명인-브랜드 유사도 점수는 분류 기준 내 공통 키워드들의 유사도 점수 총합으로 계산하였고, 유명인-브랜드의 이미지 유사도 점수는 각 공통 키워드들에 대한 링크 강도의 대푯값(평균)으로 결정하였다.

### 3.6 광고효과 학습 모듈

광고효과 학습 모듈에서는 기계학습 기법을 사용하여, 이미 광고효과에 관한 결과가 나와 있는 기존 집행 광고를 대상으로 유명인-브랜드의 이미지 유사도에 따른 광

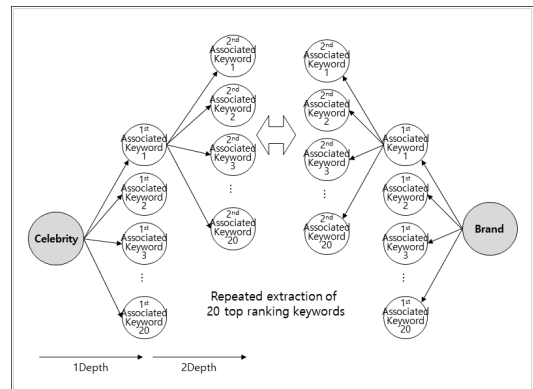


Fig. 1. Network structure for celebrity- brand image similarity

고효과를 학습시키게 된다.

유명인 모델의 광고 집행 전 광고효과를 예측할 수 있는 모델링을 하기 위해서는 기존 집행 광고의 성과가 어떠한지를 알 수 있는 종속변수가 필요하다. 가장 유용한 종속변수는 실제 행동 지표나 공신력 있는 기관에서 발표한 검증된 지표이겠지만, KAA에서 발표했던 효과 지표들이 특정 광고 모델에만 국한되어 산출되었다는 한계가 있으므로, 본 연구에서는 해당 광고에 대해 서베이를 진행하여 소비자 기억에 근거한 광고효과를 측정하여 이를 종속변수로 대체하여 활용하였다. 따라서 본 연구에서는 2가지 유형의 데이터가 분석에 활용되었는데, 하나는 과거에 유명인이 출연했던 광고물에 대한 소셜 빅데이터(독립변수)이고, 다른 하나는 각 광고물의 효과를 측정하는 서베이 데이터(종속변수)이다.

요약해 보면, 학습 모듈에서는 학습 데이터 set(전체 9개 광고 중 7개의 광고)에 기반하여, 모델과 브랜드에서 공통으로 나타나는 키워드에 대해 분류 기준별로 유사도 점수를 산출한 뒤, 유명인-브랜드 이미지 유사도를 독립변수로, 광고효과를 측정하는 서베이 데이터를 종속변수로 하는 광고효과 예측 모형(다중회귀모형)을 학습시켰다.

본 분석에 활용된 9건의 광고물은 기계학습에 기반한 예측 모델링을 수행하기에는 부족한 분량이므로, 총 9건의 광고물에서 2건의 광고물을 예측 데이터 세트로 구성할 수 있도록 모든 경우의 수를 선정하여 데이터를 증식(총 분석 케이스 36건)하는 과정을 포함하였다.

### 3.7 광고효과 예측 모듈

광고효과 예측 모듈은 학습 모듈에서 산출한 유명인-브랜드 이미지 유사성이 광고효과를 설명하는 회귀모형식이 예측 데이터에서도 적용될 수 있는지를 검증하는

과정을 말한다. 본 예측 모듈에서는 나머지 20%의 예측 데이터(9개 광고물 중 나머지 광고물 2건에 해당하며, 모델×브랜드 최대 조합을 생성하기 위해 총 36건의 증식된 데이터를 분석에 활용)를 사용하여 위의 학습 모형으로 예측이 가능한 것인지를 검증하였다. 이때, 유명인 모델의 광고 효과 값을 정량적으로 예측해 내기보다, 해당 유명인 광고가 효과가 있는 것인지 아닌지를 파악하는 분류 예측으로 분석하였다.

예측의 정확도를 산출하는 과정은 다음과 같다. 우선 학습 모듈에서 생성한 광고효과 예측 모델링을 통해 2건의 광고물에 대한 예측값을 산출하고, 그중 예측값이 높게 나타난 유명인 광고를 예측 모델링에 의해 추천한다. 그리고 이 결과가 서베이를 통해 얻은 추천 결과(앞의 2건의 광고물에 대한 서베이 결과 중 광고효과가 높은 것을 추천한 것)와 일치하는지를 판별하여 분류 예측의 정확도를 산출한다. 이러한 예측의 정확도를 통해 본 예측이 얼마나 타당한지를 판단하게 되는데, 분류 예측에 대한 정확도 판단 기준은 75% 이상 예측 정확도가 산출되었을 때 예측값이 타당하다고 결론 내린다. 전 분석 과정은 R 3.5.1 버전 프로그램을 활용하였다.

## 4. 연구 결과

### 4.1 연구 문제1: 이미지 유사도 분류 기준 선정 결과

유명인과 브랜드의 이미지 유사도를 나타내는 공통 키워드를 통해 유명인 모델의 광고 효과를 예측하기 위해서는, 각 키워드들을 유사한 개념의 그룹으로 분류한 후에 예측 알고리즘에 투입해야 한다. 따라서, 유명인-브랜드 유사도 점수를 예측 회귀모형에 투입하기 위해서, 유사도 키워드 분류 기준으로 Table 1과 같이 '개체, 역할, 개인 관계, 집단관계, 미분류'의 다섯 가지 유형을 추출하였다. 해당 분류 기준은 세상에 널리 알려진 '유명인의 이미지' 뿐만 아니라 '일반인으로서의 이미지'와 '사회적 역할로서의 이미지', 그리고 '단체 혹은 개인 관계로 형성되는 이미지'를 추가로 포함하고 있으며, 이들은 추후 예측 모델링의 독립변수로 투입되었다.

### 4.2 연구 문제2: 유명인-브랜드 이미지 유사도의 광고효과 예측 의사결정

Table 1의 7개 학습 데이터에 대해, 소셜 빅데이터를 통해 도출한 유명인-브랜드 이미지 키워드의 분류 기준

Table 1. Definition of classification criteria for celebrity-model image similarity

criteria	Definition
individual	Keywords for personal privacy, personality, lifestyle, etc., not keywords for images as celebrities (eg : marriage, scandle, anecdote, etc.)
Role	Keywords related to social role or image as a celebrity (eg. : Works, roles in the play, etc.)
Group relation	Keywords mentioned in various groups or collective relationships such as countries and political parties
Personal relation	Keywords mentioned in personal relationships, such as the name of a specific individual
Others	Keywords not included to any of the classifications quoted above

별(개체, 역할, 집단관계, 개인 관계, 미분류) 유사도 점수를 독립변수로 하고 서베이를 통한 광고효과 점수를 종속변수로 하여 회귀모형을 도출하고, 독립변수에 해당하는 5개의 키워드 분류 기준별로 회귀계수를 산출하였다. 분석 결과, 전체 회귀모형에 대한 R 값은 0.8891로 나타났으며, 도출된 회귀식의 통계적 유의성은 높게 나타나 광고효과 예측 의사결정 모델은 유효한 것으로 판정되었다.

또한, 이 규칙을 2개의 예측 사례에 적용한 결과, ‘김연아’의 ‘커피믹스’ 광고의 예측값이 ‘스마트에어컨’ 광고의 예측값보다 낮게 나타났으므로, 본 연구를 통해 ‘김연

아’의 ‘커피믹스’ 광고효과가 더 낮을 것으로 추천되었다. 이런 방식으로 예측 모델링을 통해 제안된 광고물과 서베이를 통해 제안된 광고물을 최종 비교하여 본 회귀식의 예측 타당성을 검증하였다. 나머지 예측 데이터(36개의 증식된 데이터)에 대해서도 동일한 방식의 분석을 총 36번 반복하였다. 이 과정은 80%의 학습 데이터를 바탕으로 생성한 회귀모형에 나머지 20%의 예측 데이터를 적용하여 산출한 추천 결과와 서베이에 의한 추천 결과와 비교하여 예측 정확도를 판정하는 과정이다. 본 예측은 분류 예측을 기반으로 수행되었으므로, 예측 정확도는 분류 결과에 대한 정(正)·오(誤)에 대한 비율로 판단하였다. Table 2와 같이 총 36번의 분석 중에서 27번의 결과가 정확하게 일치함으로써 75.0%의 예측 정확도를 보였다. 이는 분류 정확도의 75% 기준에 부합하는 것이므로 본 모델의 예측 타당성은 높다고 하겠다.

Table 2. Judgment on predictive modeling through comparison of predicted results by survey and modeling

No.	AD 1		AD 2		Prediction by survey	Prediction by modeling	Judgment
	Celebrity	Brand	Celebrity	Brand			
1	Kim, Yuna	Maxim	Kim, Yuna	Smart Aircon	AD 2	AD 2	T
2	Kim, Yuna	Maxim	Kim, Tae-hee	Kimchi-toktok	AD 1	AD 1	T
3	Kim, Yuna	Maxim	Song, Hye-kyo	Sulwhasoo	AD 1	AD 2	F
4	Kim, Yuna	Maxim	Lee, Sung-kyung	Laneige	AD 1	AD 1	T
5	Kim, Yuna	Maxim	Jun, Ji-hyun	SK Telecom	AD 1	AD 1	T
6	Kim, Yuna	Maxim	Jun, Ji-hyun	Zipel	AD 1	AD 2	F
7	Kim, Yuna	Maxim	Jun, Ji-hyun	Hera	AD 1	AD 2	F
8	Kim, Yuna	Maxim	Han, Ga-in	Donginbi	AD 1	AD 2	F
9	Kim, Yuna	Smart Aircon	Kim, Tae-hee	Kimchi-toktok	AD 1	AD 1	T
10	Kim, Yuna	Smart Aircon	Song, Hye-kyo	Sulwhasoo	AD 1	AD 1	T
11	Kim, Yuna	Smart Aircon	Lee, Sung-kyung	Laneige	AD 1	AD 1	T
12	Kim, Yuna	Smart Aircon	Jun, Ji-hyun	SK Telecom	AD 1	AD 1	T
13	Kim, Yuna	Smart Aircon	Jun, Ji-hyun	Zipel	AD 1	AD 1	T
14	Kim, Yuna	Smart Aircon	Jun, Ji-hyun	Hera	AD 1	AD 1	T
15	Kim, Yuna	Smart Aircon	Han, Ga-in	Donginbi	AD 1	AD 1	T
16	Kim, Tae-hee	Kimchi-toktok	Song, Hye-kyo	Sulwhasoo	AD 2	AD 2	T
17	Kim, Tae-hee	Kimchi-toktok	Lee, Sung-kyung	Laneige	AD 1	AD 1	T
18	Kim, Tae-hee	Kimchi-toktok	Jun, Ji-hyun	SK Telecom	AD 1	AD 1	T
19	Kim, Tae-hee	Kimchi-toktok	Jun, Ji-hyun	Zipel	AD 2	AD 2	T
20	Kim, Tae-hee	Kimchi-toktok	Jun, Ji-hyun	Hera	AD 2	AD 2	T
21	Kim, Tae-hee	Kimchi-toktok	Han, Ga-in	Donginbi	AD 1	AD 2	F
22	Song, Hye-kyo	Sulwhasoo	Lee, Sung-kyung	Laneige	AD 1	AD 1	T
23	Song, Hye-kyo	Sulwhasoo	Jun, Ji-hyun	SK Telecom	AD 1	AD 1	T
24	Song, Hye-kyo	Sulwhasoo	Jun, Ji-hyun	Zipel	AD 2	AD 1	F
25	Song, Hye-kyo	Sulwhasoo	Jun, Ji-hyun	Hera	AD 2	AD 2	T
26	Song, Hye-kyo	Sulwhasoo	Han, Ga-in	Donginbi	AD 1	AD 1	T
27	Lee, Sung-kyung	Laneige	Jun, Ji-hyun	SK Telecom	AD 2	AD 2	T
28	Lee, Sung-kyung	Laneige	Jun, Ji-hyun	Zipel	AD 2	AD 2	T
29	Lee, Sung-kyung	Laneige	Jun, Ji-hyun	Hera	AD 2	AD 2	T
30	Lee, Sung-kyung	Laneige	Han, Ga-in	Donginbi	AD 1	AD 2	F
31	Jun, Ji-hyun	SK Telecom	Jun, Ji-hyun	Zipel	AD 2	AD 2	T
32	Jun, Ji-hyun	SK Telecom	Jun, Ji-hyun	Hera	AD 2	AD 2	T
33	Jun, Ji-hyun	SK Telecom	Han, Ga-in	Donginbi	AD 1	AD 2	F
34	Jun, Ji-hyun	Zipel	Jun, Ji-hyun	Hera	AD 2	AD 2	T
35	Jun, Ji-hyun	Zipel	Han, Ga-in	Donginbi	AD 1	AD 2	F
36	Jun, Ji-hyun	Zipel	Han, Ga-in	Donginbi	AD 1	AD 1	T

※ AD1: 광고물1, AD2: 광고물2, T: True, F: False

## 5. 결론

본 연구의 목적은 소셜 빅데이터에 근거하여 유명인 모델과 브랜드의 이미지 유사도를 산출하고, 이러한 이미지 유사도에 근거하여 유명인 모델의 광고효과 예측모형을 구축하는 것이다. 이러한 연구 문제를 해결하기 위해 데이터마닝 절차에 따라 6가지 모듈(빅데이터 수집 모듈→ 빅데이터 전처리 모듈→ 빅데이터 통합 적재 모듈→ 이미지 유사도 산출 모듈→ 광고효과 학습 모듈→ 광고효과 예측 모듈)로 빅데이터 분석을 실행하였다.

유명인-브랜드 이미지 유사도로 광고효과를 예측하기 위해, 일단 유명인과 브랜드에서 공통으로 나타난 키워드를 ‘개체, 역할, 개인 관계, 집단관계, 미분류’의 이미지 유사도 분류 기준으로 유형화하여 이미지 유사도 점수를 산출하였다. 또한 이렇게 도출된 이미지 유사도가 광고효과 예측 변수가 될 수 있는지를 알아보기 위해, 이를 독립변수로, 광고효과를 종속변수로 하는 예측 모델링을 기계학습을 실시하여 학습시켰다. 이때, 유명인 광고의 광고효과를 정량적으로 예측해 내기보다, 해당 유명인 광고가 효과가 있는지를 판단하는 분류 예측을 실시하였다. 따라서, 학습 데이터로 도출한 예측 모델링을 통해 얻은 유명인 광고의 광고효과 예측결과를 서베이 결과와 비교하여 예측의 정(正)·오(誤)를 판정한 결과, 총 36번의 분석 중에서 27번의 결과가 정확하게 일치함으로써 75.0%의 정확도로 예측의 타당성을 확보하였다. 즉, 유명인 모델과 브랜드의 이미지 유사도는 광고효과를 예측하는데

타당한 변수임이 입증되었으므로, SNS상의 다양한 유명인의 이미지 실체를 탐색하여 이를 기반으로 광고효과를 예측하는 것은 가능하다는 결론을 내릴 수 있다고 하겠다.

본 연구가 갖는 학문적 및 실무적 의의는 아래와 같다. 첫째, 본 연구는 최근 관심을 받는 소셜 빅데이터를 활용하여, 유명인 모델을 대상으로 실시간 생성되는 소비자 담론에 근거한 유명인-브랜드 이미지 유사도를 산출하였다는 데서 방법론적 의의가 있다. 둘째, 유명인의 일상에서 일어나는 다양한 사건들을 서사적 구조로 여기고, 모델의 이미지를 표상적인 관점뿐만 아니라 복합체적인 관점도 포함하여 그 실체를 다차원적으로 파악하였다는 데 이론적 의의를 갖는다. 셋째, 본 연구에서는 SNS에서의 소비자 여론 구조가 브랜드 연상 네트워크 구조와 유사하다는 점을 반영하여, 이에 대한 이론적 배경을 기반으로 유명인과 브랜드에 공통으로 출현하는 키워드의 링크 강도를 측정하여 유명인-브랜드의 실제적인 이미지 유사도를 정량화할 수 있었다. 넷째, 본 연구는 최근 빅데이터 분야에서 주목받는 기계학습 기법을 사용하여 유명인 모델의 광고효과 예측 알고리즘을 개발하고, 이를 통해 유명인 모델을 선정하기 위한 최적의 의사결정에 기여했다는 발견에서 실용적인 시사점을 찾을 수 있다.

더불어 빅데이터 분석에 대한 정교함을 높이기 위해 다음과 같은 후속 보완 연구들이 필요하다고 본다. 첫째, 본 연구에서 활용한 데이터는 온라인 뉴스 사이트에 한정되어 있지만, 소비자의 진성 여론이 다양한 소셜미디어에서 생성되고 있으므로 분석 채널을 확대하여 데이터를 수집하는 노력이 필요하겠다. 둘째, 본 연구에서 사용한 예측 기준변수로 서베이를 통한 광고효과 점수를 활용하였으나, 좀 더 공신력 있는 효과 지표를 확보하면 예측 모델링의 신뢰도를 더 높일 수 있을 것으로 예상된다. 셋째, 본 연구의 분석 대상으로 선정한 유명인은 일부 중복 출연이 통제되지 못한 상황에서 분석에 투입되었는데, 중복 출연 케이스를 좀 더 엄격하게 통제하게 되면 유명인 효과에 대한 객관성을 확보하는 데 도움이 될 것으로 기대된다. 넷째, 본 예측 모델링에 투입된 독립변수로서의 유명인-브랜드 이미지 유사도 분류 기준은 다소 부족한 데이터 세트로 인해 유명인-브랜드에 대한 심층적 이미지 구조를 도출하는데 다소 한계가 있었으므로, 향후 연구에서 다양한 소셜미디어에서 형성되는 사회적 담론을 활용한다면 좀 더 의미 있는 예측 변수 개발이 가능할 것으로 판단된다.

이상으로 소셜 빅데이터를 이용한 유명인 모델의 광고 효과 모델링은 광고 모델에 관한 연구 분야에 새로운 지

평을 열어주었다는 점에서 학문적 시사점을 갖는다. 하지만, 빅데이터 분석 방법에 근본적으로 내포된 내적 타당성 문제와 유용한 결과 도출에 이르기까지 많은 과거 데이터 확보가 필요하다는 비효율성의 문제는 앞으로 빅데이터 연구가 진화·발전하는데 있어서 극복해야 할 과제로 남아 있다.

## REFERENCES

- [1] W. S. Kyeong. (2014, July). Mode's double appearances and advertising effects, *KAA Special Report, July+August*, 10-13.
- [2] C. Amos, G. Holms & D. Strutton. (2008). Exploring the relationship between celebrity endorser effects and advertising effectiveness. *International Journal of Advertising*, 27(2), 209-234.
- [3] H. H. Friedman & L. Friedman. (1979). Endorser effectiveness by product type. *Journal of Advertising Research*, 19(5), 63-71.
- [4] I. Y. Nam. (2002). The degree of match-up and celebrity endorser's advertising effectiveness : considering various product types. *Korean Journal of Broadcasting and Telecommunication Studies*, 16(1), 171-199.
- [5] C. R. Knott & M. St. James. (2004). An alternate approach to developing a total celebrity endorser rating model using the analytic hierarchy process. *International Transactions in Operational Research*, 11, 87-95.
- [6] J. D. Roedder, L. Barbara, K. Kim & A. B. Mong. (2006). Brand concept maps: a methodology for identifying brand association networks. *Journal of Marketing Research*, 43(November), 549-563.
- [7] J. R. Anderson & G. H. Bower. (1980). *Human Associative Memory: A Brief Edition*. Hillsdale, N.J.: L. Erlbaum Associates.
- [8] D. A. Aaker. (1996). *Building strong brands*. New York: The Free Press.
- [9] K. L. Keller, S. Brian & T. Alice. (2002). Three Questions You Need to Ask About Your Brand. *Harvard Business Review*, September, 81-86.
- [10] K. H. Choi & J. A. Yoo. (2015). A reviews on the social network analysis using R. *Journal of The Korea Convergence Society*, 6(1), 77-83. DOI : 10.15207/JKCS.2015.6.1.077
- [11] M. J. Sirgy. (1982). Self-image/product image congruity and advertising strategy. *Development in Marketing Science*, 5(February), 129-13.
- [12] A. Labrinidis & H. Jagadish. (2012). Challenges and opportunities with big data. *Proceedings of the LDB*

*Endowment*, 5(12), 2032-2033.

- [13] H. W. Byeon. (2015). The factors of participating in a smoking cessation program using integrated method of decision tree and neural network algorithm. *Journal of The Korea Convergence Society*, 6(2), 25-30.  
DOI : 10.15207/JKCS.2015.6.2.025
- [14] M. Che, S. Mao & Y. Liu. (2014). Big Data: ASurvey. *The journal Mobile Networks and Applications*, 19, 171-209.
- [15] P. Russom. (2011). *Big data analytics*. TDWI best practices report [Online], fourth quarter.  
<https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf>

김 유 나(Yuna Kim)

[정회원]



- 2000년 2월 : 이화여자대학교 수학과(이학사)
- 2004년 8월 : 고려대학교 소비자 및 광고 심리학 (문학석사)
- 2017년 6월 : 한양대학교 광고홍보학 (문학박사)
- 2012년 ~ 2020년 1월 : 대흥기획 빅데이터컨설팅팀 팀장/센터장
- 2020년 3월 ~ 현재 : 서울예술대학교 광고창작과 조교수
- 관심분야 : 디지털마케팅, 디지털트랜스포메이션 전략, 빅데이터 마케팅, 온·오프 통합 브랜드 전략
- E-Mail : yuna.kim@seoularts.ac.kr

한 상 필(Sangpil Han)

[정회원]



- 1985년 2월 : 한양대학교 신문방송학 (문학학사)
- 1990년 10월 : 미국 일리노이주립대학교 커뮤니케이션(문학박사)
- 1991년 3월 ~ 현재 : 한양대학교 광고홍보학과 교수
- 관심분야 : 국제광고, 디지털광고, 광고 전략
- E-Mail : sphan@hanyang.ac.kr