

# 마케팅 분야의 머신러닝 연구 동향 분석

## Research Trends in Machine Learning Applications in Marketing

김혜진 • Kim, Hye-jin, 이명구 • Lee, Myounggu

최근 ICT 기술의 발달로 고성능 컴퓨팅과 데이터 저장 기술이 발전하고 다양한 데이터가 생성, 공유, 저장되는 것이 가능해짐에 따라 머신러닝을 활용하여 마케팅 인사이트를 얻는 사례가 크게 늘어났다. 본 연구는 마케팅 연구자들에게 머신러닝 기법의 기초적인 개념들을 소개하고 어떠한 마케팅 연구에 적용될 수 있는지 정리하여 머신러닝 연구를 활성화하는 데 목적이 있다. 먼저, 머신러닝의 기초적인 개념들과 머신러닝 방법론의 종류를 간략하게 소개한다. 그리고, 마케팅과 경영학의 최우수 저널에서 머신러닝을 다루는 연구를 찾아내어 여섯 가지 유형으로 정리하였는데, 그 유형들은 다음과 같다. 1) 머신러닝을 이용하여 데이터에서 마케팅 변수를 추출하여 다른 변수와의 관계를 살펴보는 연구, 2) 머신러닝을 이용하여 새로운 데이터를 활용하는 연구, 3) 새로운 방법론이나 현존하는 방법론의 개선을 제안하는 연구, 4) 머신러닝을 이용하여 마케팅 현상의 패턴을 설명하는 연구, 5) 머신러닝을 이용하여 마케팅 현상을 예측하고, 예측력 향상에 영향을 미치는 변수들을 살펴보는 연구, 6) 마케팅 문제에 대한 해결책을 제시하고 인과관계를 규명하는 연구이다. 마지막으로, 마케팅 분야에서 머신러닝 활성화를 위한 제언과 마케팅 연구자들을 위한 함의를 제공한다.

핵심주제어: 머신러닝, 기계학습, 딥러닝, 토폭모델링, 잠재 디리클레 할당, 자연어 분석, 감성 분석

김혜진 | 한국과학기술원(hyejinkim@kaist.ac.kr), 제1저자 및 교신저자

이명구 | 한국과학기술원(myounggulee@kaist.ac.kr), 공동저자

## ABSTRACT

Machine learning methods have gained popularity in marketing academia due to the proliferation of ICT technology. This research aims to introduce marketing academics to methods in machine learning and how they can be applied to research in marketing. We first introduce basic concepts and methodologies in machine learning. Then we examine a list of literature from top marketing and business journals, identify a list of papers that apply machine learning, and categorize them into six types of research. (1) research which extracts marketing variables from data and examines the relationship with other variables, (2) research which uncovers ways to analyze new types of data, (3) research which proposes new methodologies or ways to improve current methodologies, (4) research which aims to describe patterns in marketing phenomena rather than identifying relationships between variables, (5) research which aim to predict marketing phenomena and identify variables that aid prediction, and (6) research which aims to tackle marketing problems and identify causal relationships. We conclude the study with a summary of the advantages and disadvantages of machine learning methods and implications for marketing academics.

**Keywords:** Machine Learning, Deep Neural Networks, Topic Modeling, Latent Dirichlet Allocation, Natural Language Processing, Sentiment Analysis

**Hye-jin Kim** | Korea Advanced Institute of Science and Technology (KAIST), First Author, Corresponding Author  
**Myounggu Lee** | Korea Advanced Institute of Science and Technology (KAIST)

## 1. 서론

최근 고성능 컴퓨팅, 클라우드 컴퓨팅, 데이터 저장기술 등 최신 ICT 기술의 발달과, 여러 플랫폼의 등장으로 사용자 생성 콘텐츠(User Generated Content: UGC), 기업의 거래 데이터, 멀티미디어 데이터 등 다양한 데이터가 생성되고 공유되는 것이 가능해짐에 따라, 머신러닝을 활용하여 마케팅 인사이트를 얻는 사례가 크게 늘어났다. 예를 들어, 미국의 온라인 비디오 스트리밍 업체인 넷플릭스(Netflix)는 사용자의 온라인 활동을 분석하고 다른 사용자의 행동과 비교하는 알고리즘에 따라 영상 콘텐츠를 추천하고 있으며(Kumar et al. 2020), 엔터테인먼트 업체인 월트 디즈니(Walt Disney)는 자사의 놀이공원을 이용하는 고객의 세분화된 데이터를 수집하여 요구 사항을 예측하고, 놀이기구의 이용 상황에 따라 개인화된 혜택을 제공하고 있다(Marr, 2018). 글로벌 경영 컨설팅 회사인 맥킨지앤컴퍼니는 인사, 공급망 관리 등 기업의 다양한 활동 영역 중에서 머신러닝을 포함한 고급 분석 기법의 도입의 가치가 가장 큰 영역으로 마케팅 및 세일즈를 제시했으며, 그 규모가 2019년 이후 최대 6조 달러에 이를 것으로 분석하였다(McKinsey and Company, 2019). 마케팅 영역에서 다루는 빅데이터의 범위가 고객 및 구매 내역에 대한 자사 데이터(1st Party Data)에서부터 공급사, 협력사 등 타사 데이터를 포함하는 결합 데이터(2nd Party Data), 데이터 전문 기관 등 외부 조직에서 획득한 제3자 데이터(3rd Party Data)에 이르기까지 방대하기 때문에, 점점 더 많은 기업들이 다른 회사에서 획득한 고객 수준의 정보를 보완하여 데이터의 가치를 높이고자 노력하고 있다(Schneider et al. 2017).

마케팅 학계에서는 1950년 이래로 분석의 계량적인 측면을 강화하려는 시도가 끊임없이 이루어져 그 결과로

여러 계량경제학적 분석 방법론과 통계적 방법을 이용한 연구가 이루어졌는데(Sheth and Kellstadt 2020), 최근에는 산업 트렌드와 컴퓨터 공학 등 타분야에서의 방법론 발전에 발 맞추어 머신러닝을 이용하는 연구가 기하급수적으로 늘어나고 있다. 예를 들어, 주요 경영/경제 관련 학술지에 발표된 인공지능과 관련된 마케팅 연구는 2014년 이전까지 매년 10편 이하로 비교적 미미한 수준이었으나, 2015년부터 그 수가 급격히 증가하는 추세를 보이며 2019년에는 51편의 논문이 발표되었다(Mustak et al. 2021). 따라서 지금까지 진행되었던 연구를 살펴보고, 앞으로 마케팅 분야에서 머신러닝을 이용한 연구가 더 활발하게 진행되기 위해 해결되어야 할 이슈들에 대해서 짚어보는 것이 필요하다.

본 리뷰는 대략 2010년대 이후 활발히 이루어진 마케팅 학계의 머신러닝 연구에 초점을 맞추어, 머신러닝이 주로 어떠한 주제에서, 어떠한 문제를 해결하기 위해 활용되었는지 살펴보고자 한다. 따라서 머신러닝의 주요 방법론에 대해서 깊게 설명하기보다는, 마케팅과 경영학의 최고 수준 저널에 게재된 논문 중 머신러닝과 빅데이터를 다루는 논문들의 내용들을 살펴보고, 공통적인 트렌드를 제시하여 향후에 국내의 마케팅 연구자들이 추가적인 연구를 하는 데 도움이 되고자 하는데 목적을 두고 있다.

머신러닝은 “컴퓨터가 경험을 통해, 명시적인 규칙을 프로그래밍하지 않고도 학습할 수 있도록 하는 연구 분야”로 정의할 수 있다(“Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed”; Samuel 1959). 좀 더 구체적으로, 머신러닝 분야의 개척자 중 한명인 Tom Mitchell에 의하면, “어떤 컴퓨터 프로그램이 특정한 작업 T에 대해서 경험 E가 늘어남에 따라 성능에 대한 척도 P가 개선된다고 할때, 작업 T와 그에

대한 성능 척도 P에 대해서 경험 E를 통해 학습한다고 정의한다.” (“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.”: Mitchell 1997) 여기서 T, P, E에 대한 예를 들자면, 컴퓨터가 사자, 기린, 독수리가 각각 그려진 사진을 보고 어떤 동물인지 분류하는 것을 T라고 했을 때, 성능 척도 P는 각 사진을 정확하게 분류할 확률, E는 동물이 그려진 사진(인풋)과 사진에 그려진 동물이 무엇인지 정답을 표기한 레이블(아웃풋)을 한 쌍으로 하여 구성된 데이터셋이다. 여기서 기계가 학습한다는 것은 사자는 갈기가 있고, 기린은 목이 길고, 독수리는 날개가 있다는 규칙을 알려주지 않고도 컴퓨터가 충분한 사례들을 경험하면서 사자, 기린, 독수리의 패턴을 스스로 배워가는 것이다. 실제 머신러닝 문제에서 많은 경우에 작업 T는 어떤 함수로 나타내어지고(이 함수는 파라미터로 설명할 수 있는 형태(parametric form)로 명확하게 제시될 수도, 아닐 수도 있다), 학습이 이루어지는 것은 그 함수의 정확도, 또는 정확도를 측정할 수 있는 어떤 성과지표(P)를 높이는 과정이며, 인풋 아웃풋으로 구성되어 있는 샘플 데이터셋이 E로 주어진다(Jordan & Mitchell 2015).

머신러닝 분야는 컴퓨터 공학을 위주로, 인접한 학문인 통계학 등의 지원을 받아 발전해 왔는데, 광범위한 종류의 데이터를 다루고 다양한 문제를 해결하기 위해 여러 알고리즘이 개발되어 왔다(Murphy 2012). 모든 방법론에 대해서 자세하게 설명하는 것은 본 연구의 목적과는 맞지 않기 때문에, 다음 장에서는 머신러닝의 여러 방법론에 대한 기초적인 내용들에 대해 설명한다.

## II. 이론적 배경: 머신러닝 기법의 종류

### 1. 지도학습(Supervised Learning)

지도학습은 머신러닝에서 가장 광범위하게 사용되는 학습 기법이다. 서론에서 제시한 동물 분류 작업이 지도학습의 한 예시라고 할 수 있으며, 이 외에도 스팸메일 분류, 이미지에서의 얼굴 인식, 음성 인식 등도 지도학습의 대표적인 예라고 할 수 있는데, 이는 모두 어떠한 인풋  $x$ , 아웃풋  $y$ 의 쌍인  $(x, y)$ 들에 대해서,  $x$ 와  $y$ 의 관계를 가장 잘 나타내는 함수  $f(x)$ 를 학습하여 향후에 어떤 새로운  $x'$ 가 주어졌을 때  $y'$ 를 예측하거나 분류하는 것을 핵심으로 한다.  $(x, y)$ 의 쌍은 정답이 있기 때문에  $f(x)$ 를 추정할 때 얼마나 잘 했느냐에 따라 피드백을 받게 된다. 예를 들어, 어떤 인풋  $x$ 에 대해  $y^* = f(x)$  값을 얻었을 때 실제  $y$ 값과 예측된  $y^*$ 의 차이인 에러( $e = y - y^*$ )를 얻을 수 있기 때문에 그 에러를 최소화할 수 있는  $f(x)$ 를 찾는다. ‘지도(supervised)’ 학습이라는 이름은 이러한 측면을 반영한 것이다.

인풋인  $x$ 는 숫자로 이루어진 벡터가 될 수도 있고, 어떤 문서, 이미지, 오디오, 비디오, 행동 데이터 등이 될 수도 있는데, 대부분의 경우 이러한 데이터는 향후에 숫자로 이루어진 벡터나 행렬로 변환되어 알고리즘의 인풋으로 사용된다. ICT 기술의 발달로 여러 종류의 데이터 수집이 가능해졌기 때문에, 이러한 데이터를 분석하기 위해 전처리하는 과정도 중요해졌으며, 마케팅에서도 다양한 데이터가 활용되고 있다. 향후 장에서 마케팅 연구에서 그동안 사용되었던 다양한 데이터들에 대해서 소개할 것이다. 마찬가지로, 아웃풋  $y$ 도 다양한 형태가 있을 수 있지만, 주로 숫자로 이루어진 벡터의 형태를 취하며, 예측하는 값이 범주형 변수(categorical variable)라면 분류(classification) 작업, 연속형 변수(continuous

variable)라면 회귀(regression) 작업이 된다. 전자는 주어진 메일이 스팸메일인지 아닌지를 판단하는 것을 예로 들 수 있고, 후자는 과거 데이터를 기반으로 하여 주식 가격을 예측하는 것을 예로 들 수 있다.

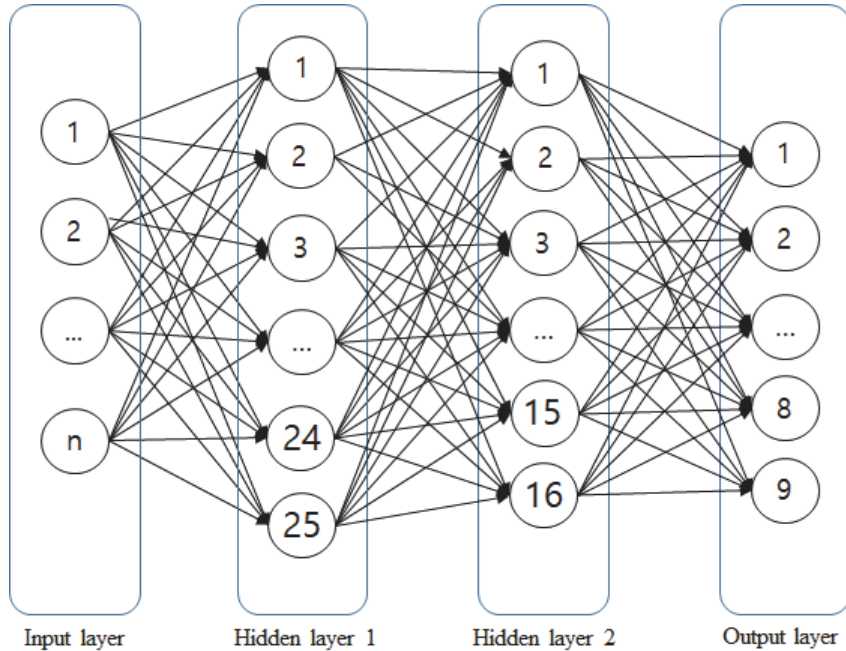
머신러닝에서는 모델을 구축하기 위해 데이터셋을 크게 세 가지 용도로 나누게 된다. 먼저, 모델을 학습하기 위해, 다시 말해서 함수  $f(x)$ 를 알아내기 위해 쓰이는  $(x, y)$  데이터셋을 학습 데이터(training data)라고 한다. 이때, 정확도(accuracy), 재현율(recall), 정밀도(precision), F1(재현율과 정밀도의 조화평균) 등의 지표로 모델을 평가하게 된다. 머신러닝에서는 주어진 데이터에 대해서 정답을 얼마나 잘 예측할 수 있을지가 중요한데, 학습 데이터에 적용했을 때에는 최고의 성과를 보이지만 다른 데이터셋에 적용시키면 전혀 예측을 못하는 '과학습(overfitting)'이 일어날 위험이 있다. 예를 들어, 학습 데이터의  $(x, y)$  쌍들이 있을 때 그 쌍들만 있는 곡선을 구현한다면 학습 데이터에 대해서는 100퍼센트의 정확도를 보이지만, 새로운 데이터에 대해서 예측을 하고자 할 때는 예측이 완전히 어긋나는 문제가 생길 수가 있다. 따라서 학습에 쓰인 데이터 외에도, 모델이 처음 접하는 데이터셋에 대해서 모델의 최종적인 성능을 평가하게 되는데, 이를 평가 데이터(test data)라고 한다. 때로는 학습 데이터의 일부를 따로 떼어서, 학습에는 사용하지 않되 학습하는 도중의 모델 성과를 살펴보고 모델을 좀 더 정교하게 만든다든가(fine-tuning), 모델 선택(model selection)에 사용할 수 있는데, 이를 검증 데이터(validation data)라고 한다.

한편, 지도학습 분야에서 최근 매우 각광받고 있는 분야는 딥러닝(deep learning)이다. 딥러닝은 지도학습에만 적용되는 것은 아니지만 지도학습 분야에서 가장 큰 발전을 이루었다. 간단하게 설명하자면, 딥러닝은 인간의 뇌의 신경망을 본뜬 인공 신경망(artificial neural

network; 이하 ANN)을 기본으로 하는데, 인풋층(input layer)에서 아웃풋층(output layer)으로 이어지는 중간에 여러 개의 은닉층(hidden layer)이 있고, 각각의 은닉층은 여러 개의 노드(또는 뉴런)으로 이루어져 있다. 이때, 인풋층에서 아웃풋 층까지 데이터가 전달되는 알고리즘을 전방 전달 신경망(feed-forward network)이라고 한다. 각 층의 노드는 크게 두개의 함수로 이루어져 있다. 첫번째 함수는 이전 층에서의 아웃풋(들)을 인풋으로 받아서 각각에 가중치를 곱한 합을 아웃풋으로 내보낸다. 여기서 가중치들은 딥러닝 알고리즘의 최적화 과정을 통해 조정이 되는 파라미터(parameter)들이다. 첫번째 함수를 통한 아웃풋은 두번째의 활성화 함수(activation function)를 통해 일정한 수준 이상으로 활성화가 되어야 다음 층으로 넘어가게 된다. 그리하여 모든 층을 지나 최종 아웃풋을 얻게 되는데 이것과 실제 아웃풋의 차이인 오차를 최소화하는 과정이 딥러닝의 핵심이다. 이를 위해 역전파 알고리즘(backpropagation)을 통해 이전 층의 각 노드들이 최종 오차에 얼마나 기여를 하였는지 계산을 하게 되는데, 이를 바탕으로 각 노드들의 함수에서 가중치가 조정되어 오차를 줄이는 방향으로 진행이 된다. 가중치를 조정하는 방법으로는 경사 하강법(gradient descent)을 이용한다. 그리고 전방 전달 과정과 역전파 알고리즘을 학습 데이터에 대해서 여러 번 반복하면 오차를 최소화하는 모델을 도출할 수 있게 된다. <그림 1>은 인풋층과 아웃풋층, 그리고 노드를 각각 25개, 16개를 가지고 있는 은닉층 두개를 가지고 있는 ANN의 구조를 나타내고 있다.

딥러닝은 그동안 복잡한 신경망에 대한 계산의 한계를 극복하는 방법들이 개발되고(Hinton et al. 2006, Hinton and Salakhutdinov 2006, Srivastava et al. 2014) 병렬 컴퓨팅과 GPU(Graphics Processing Unit)의 발전으로 컴퓨터의 계산 능력이 급속도로 향상

〈그림 1〉 ANN의 구조 예시



되면서 비약적인 발전을 이루게 되었으며, 대표적인 딥러닝 모델에는 합성곱 신경망(Convolutional neural network; 이하 CNN)과 순환신경망(Recurrent neural network; 이하 RNN)이 있다.

## 2. 비지도학습(Unsupervised Learning)

지도학습의 학습 데이터가  $(x, y)$ 의 짝으로 이루어져 '정답'을 포함하는 데 반하여, 비지도학습의 학습 데이터는 인풋인  $x$ 만을 포함하고 아웃풋  $y$ 에 대한 정보는 없다. 비지도학습은 정답이 없는 데이터에 대한 패턴을 찾아내거나 어떤 정보를 추출하는 데 목적을 두고 있다. 대표적인 예는 차원축약(dimension reduction) 문제이다. 이는 마케팅에서도 자주 사용되는 요인분석(factor analysis)이나 다차원척도분석(multidimensional scaling)과 그 목적을 같이 한다. 또다른 대표적인 예로, 군집분석(clustering analysis)에서는 현재 데이터

의 인풋  $x$ 를 이용하여 데이터를 여러 개의 이질적인 집단으로 나누되, 각 집단 내는 동질성이 유지되도록 하는데, 이것은 마치 마케팅에서 세분시장을 나누는 것과 같다. 그리하여 그 규칙을 이용하여 새로운 데이터가 주어졌을 때 어느 한 집단으로 분류할 수 있게 된다. 온라인 쇼핑몰의 장바구니 데이터를 이용할 때 많이 사용되는 협업필터링(collaborative filtering) 알고리즘 또한 비지도학습의 일종이다. 쇼핑몰에서 비슷한 품목을 구매한 사람들을 분류하여 그 사람들이 향후에 어떠한 제품을 구매할 것인지 예측하고, 알맞은 제품을 추천해 주는 것이다.

## 3. 강화학습(Reinforcement Learning)

강화학습은 컴퓨터가 일련의 행동에 대한 결정을 내리도록 학습시키는 것을 말한다. 알파고가 바둑을 두도록 훈련시키거나 자율주행차를 훈련시키는 것이 대표적인 예

라고 할 수 있다. 지도학습에서의 학습 데이터가 인풋, 그리고 인풋의 정답인 아웃풋의 쌍으로 이루어진 데 반하여, 강화학습에서의 학습 데이터는 어떠한 정답이 정해져 있지 않은 대신에 각 단계에서 행동을 한 후에 보상(reward) 또는 비용(cost)에 대한 피드백을 얻게 된다. 보상은 각 단계에서 이루어지고 컴퓨터는 여러 단계에 걸친 결정을 내려야 하기 때문에 현재의 행동이 미래에 어떤 가치를 갖는지에 예측할 수 있어야 한다. 따라서 누적 보상을 반영한 목적 함수를 최적화하는 것이 학습의 핵심이다. 강화학습은 다음 단계의 상태가 현재 상태에만 영향을 받는 마르코프 특성(Markov property)을 갖는다는 가정을 따르는 경우가 많다.

### III. 마케팅 분야에서 머신러닝 기법을 이용한 연구 현황 분석

#### 1. 연구 방법

본 연구에서는 최근 마케팅 분야에서 머신러닝 기법이 어떻게 적용되었는지 파악하기 위해 여러 문헌들을 검토하였다. 문헌을 파악하기 위해 빅데이터와 정보의 관점에서 마케팅 분야 내 '디지털 트랜스포메이션(Digital Transformation)' 영역의 문헌을 살펴보았던 Miklosik & Evans (2020)의 키워드 검색 방법론을 참고하였다. 효과적인 문헌의 검토를 위해 대상 저널과 연구 목적에 부합하는 다양한 키워드 조합을 생성하였으며, Science Direct, ACM Digital Library, Emerald와 같은 유명 과학 데이터베이스에서 키워드 검색을 실시하였다. 연구의 초기에는 마케팅 분야의 최고 수준 저널인 Journal of Marketing, Journal of Marketing Research, Marketing Science, International Journal of

Research in Marketing, Management Science와, 경영정보시스템 분야에서 잘 알려진 Information Systems Research, MIS Quarterly를 검토하였다. 연구 목적에 부합하는 키워드로는 'machine learning in marketing', 'machine learning applications in marketing', 'big data', 'big data marketing', 'deep learning in marketing', 'topic modeling', 'sentiment analysis' 등을 사용하였으며, 마케팅에서 머신러닝 연구가 비교적 활발하게 수행되기 시작했던 2015년과 그 이후의 기간만을 대상으로 하였다. 우선 모든 키워드 검색 결과를 잠재적인 문헌 목록에 포함시켰고, 논문의 제목, 키워드 및 초록을 개별적으로 검토하여 관련성 있는 문헌의 목록을 재구성하였다. 데이터베이스 내 키워드 검색 과정에는 'AND', 'OR', 'NOT' 등의 부울(Boolean) 연산자가 적극적으로 활용되었다. 문헌 검토로부터 얻은 새로운 키워드와 각 논문의 참고 문헌을 통해 2015년보다 앞선 기간에 나온 논문을 포함하여 기존의 리스트를 보완하였으며, 이 과정에서 Journal of Business Research, Journal of Machine Learning Research, IEEE Intelligent Systems 등에 발행된 논문을 추가적으로 검토하였다. 문헌 분석은 2020년 9월부터 2021년 1월 까지 약 5개월 간 진행되었으며, 머신러닝의 마케팅 분야에의 적용이라는 연구 주제와 부합하는 총 83건의 문헌을 대상으로 하였다.

본 연구에서는 마케팅 분야에서 머신러닝을 적용한 연구의 목적을 크게 여섯 가지로 분류하였다. 1) 머신러닝을 이용하여 데이터에서 마케팅 변수를 추출하여 다른 변수와의 관계를 살펴보는 연구, 2) 머신러닝을 이용하여 새로운 데이터를 활용하는 연구, 3) 새로운 방법론이나 현존하는 방법론의 개선을 제안하는 연구, 4) 머신러닝을 이용하여 마케팅 현상의 패턴을 설명하는 연구, 5) 머신러닝을 이용하여 마케팅 현상을 예측하고, 예측력

향상에 영향을 미치는 변수들을 살펴보는 연구 6) 마케팅 문제에 대한 해결책을 제시하고 인과관계를 규명하는 연구이다. 첫 번째부터 세 번째 유형은 주로 데이터의 취합과 분석, 또는 방법론에 관한 내용이고, 네 번째부터 여섯 번째 유형은 주로 연구의 목적에 따른 분류를 표방하고 있다. 물론, 모든 연구는 하나의 목적이나 하나의 방법론만을 염두에 두고 있지는 않으므로 두 개 이상의 유형에 해당될 수도 있으나, 연구 개요에서 가장 뚜렷하게 드러나는 내용을 바탕으로 유형을 분류하였다. 다음 절에서는 각각의 유형에 해당하는 연구에 대해서 정리해 보기로 한다.

## 2. 머신러닝 기법을 이용한 연구의 유형 분류

유형 1: 머신러닝을 이용하여 데이터에서 마케팅 변수를 추출하여 다른 변수와의 관계를 살펴보는 연구

첫번째 유형은 머신러닝을 이용하여 데이터에서 변수를 추출하고, 일반적인 통계적 기법을 이용하여 그것과 다른 마케팅 변수와의 관계를 보는 것이다. 이 경우는 데이터의 규모가 너무 크거나 복잡하여 전통적인 측정 방법, 즉 사람의 손으로 레이블링하는 식으로 정보를 추출하기 어려웠던 것을 머신러닝으로 보완한 경우가 많다. 본 유형의 연구 중 가장 많은 비중을 차지 하는 것은 사용자 생성 콘텐츠, 그 중에서도 주로 텍스트를 이용한 연구이다. 사용자 생성 콘텐츠는 금전적인 대가를 받는 전문가가 아닌 일반 대중들이 생성하고 제작하여 주로 인터넷을 통해 유통되고 전달되는 미디어 콘텐츠를 말한다(Daugherty et al. 2008). 마케팅 연구에서 데이터로 사용되는 사용자 생성 콘텐츠 예로는 소셜 네트워크 서비스나 소셜 미디어, 온라인 제품 리뷰, 블로그 등이

있다. 온라인 구전(electronic word of mouth, 이하 e-WOM)도 이 범주에 속한다.

먼저, 사용자 생성 콘텐츠에서 추출한 변수들과 시장 성과(market outcome)와의 관계에 대한 연구가 활발히 이루어졌다. 이 중 가장 대표적인 연구는 리뷰의 양(volume)을 측정하고 감정(sentiment)과 같은 변수들을 분류하는 연구이다. Tirunillai & Tellis (2012)는 사용자 생성 콘텐츠가 주식시장 성과와 관련이 있는지, 관련이 있다면 사용자 생성 콘텐츠의 어떠한 지표가 가장 큰 영향을 끼치는지에 대해서 살펴보았다. 이를 위해 사용자 생성 콘텐츠에 관한 여러 지표를 도출하고 다변량 시계열 모형을 이용하여 주식 시장 성과와의 관계를 알아보았다. 특히, 리뷰의 긍/부정의 방향성(valence)을 추출하기 위해 나이브 베이지안(Naive Bayesian)과 서포트 벡터 머신(Support Vector Machine; 이하 SVM)을 사용하였다. Homburg et al. (2015)는 소비자가 온라인 커뮤니티에서 기업의 적극적인 참여에 어떻게 반응하는지 살펴보기 위해 문장별로 SVM을 이용하여 감정을 측정하였다.

Marchand, Hennig-Thurau & Wiertz (2017)는 e-WOM의 종류인 소비자 리뷰와 마이크로 블로거가 제품에 끼치는 영향력이 동등하지 않을 것이라는 가설을 가지고, 1,300만개 이상의 트위터 데이터와 17,000개 이상의 아마존 리뷰 데이터를 수집하여 신작 게임의 출시 전과 후의 매출에 끼치는 영향을 살펴보았다. 그 결과로, 소비자 리뷰의 양은 제품 출시 후 10주까지, 그리고 리뷰에 드러난 긍/부정의 감정은 6주까지 영향력을 행사한데 반하여, 마이크로 블로거의 양은 출시 첫 주 매출에만 영향을 끼치는 것으로 나타났다. Liu, Lee & Srinivasan (2019)은 소비자가 리뷰 읽기 - 검색 - 구매의 연쇄적인 의사결정 과정에서 리뷰 콘텐츠를 어떻게 활용하는지 알아보기 위해서 영국의 대형 온라인 쇼핑몰



의 구매 행동 데이터를 분석하였다. 리뷰에서 6개의 속성(feature)과 방향성을 추출하기 위해 부분 딥러닝(partial-deep learning) 방법을 사용하였고, 속성들이 구매 전환율에 끼치는 영향을 분석하였다.

한편, 잠재 디리클레 할당(Latent Dirichlet Allocation: 이하 LDA)을 적용한 토픽 모델링도 활발히 적용되었다. LDA는 자연어 처리 기법 중 문서에 나타난 주제를 분석하는 방법론이다. 여기서 문서라 함은 마케팅 연구에 적용하면 제품 리뷰, SNS의 게시물 등을 일컫는다. 각 문서는 드러나지 않은(latent) 여러 개의 토픽(주제)으로 이루어져 있고, 각 토픽은 여러 단어의 집합이다. 여기에서 각 토픽은 어떤 확률 분포에 따라 단어를 생성한다고 가정하는데, 실제로 관찰된 단어를 바탕으로 베이지 추론을 이용하여 가장 가능성이 높은 토픽을 찾아내는 과정이다(Blei et al. 2003). Hollenbeck (2018)은 온라인 여행 플랫폼 내 평판이 독립형과 체인형의 호텔 분류에 따라 수익에 어떠한 영향을 미치는지 알아보기 위해 트립어드바이저(TripAdvisor)와 프라이스라인(Priceline)에서 호텔의 평점과 리뷰를 수집하고, 호텔 점유세 데이터를 통해 호텔의 월별 수익을 추정하였다. 더불어 LDA를 통해 각 리뷰를 토픽별로 분류하고, 텍스트 데이터 기반의 독립변수를 생성하였다. Wang & Chaudry (2018)는 고객의 온라인 리뷰에 대한 관리자의 답변이 후속 리뷰에 어떻게 영향을 미치는가에 대해 연구하였다. 이를 위해 개별 고객에게 제공하는 맞춤 응답(response tailoring)을 측정하였는데, LDA를 이용해 리뷰와 답변에 대해 10개의 토픽과 해당 확률을 도출하였고, 코사인 유사도(cosine similarity)로 답변이 본 리뷰와 얼마나 유사한지 측정하였다.

사용자 생성 콘텐츠에서 추출한 속성들이 소비자 선택에 미치는 영향과 관련한 연구도 이루어졌다. Ghose, Ipeiritis & Li (2012)는 소비자들이 제품을 평가할

때 이용하는 다차원적인 요소들을 고려하여, 여러 소스의 사용자 생성 콘텐츠에 대해서 텍스트 마이닝, 이미지 분류, 소셜 위치 태깅(social geotagging) 등의 다양한 방법으로 호텔을 설명하는 속성을 추출하고, 소비자의 지불액 대비 최고의 가치를 제공하는 제품을 추천하는 순위 시스템을 개발하고자 하였다. 연구에서는 여행 목적이나 호텔의 특성 등에 의해 나타나는 소비자의 이질성도 고려하여 이산선택 구조 모형(discrete choice structural model)을 적용하였는데, 이 모델은 호텔의 위치나 서비스 기반 특성 등의 경제적 가치를 추정할 수 있으며, 모델에서의 추정치들을 바탕으로 하여 특정 호텔을 이용하는 고객의 평균적인 효용을 도출하여 순위 시스템을 구축하였다.

사용자 생성 콘텐츠의 활성화 요인을 살펴본 연구도 활발히 이루어졌다. Chae et al. (2017)은 일명 인플루언서 마케팅이라고 하는 시드 마케팅 캠페인(seed marketing campaign)의 파급효과에 대해서 살펴보았다. 영향력 있는 시드 소비자에 의한 WOM이 초점 제품, 브랜드 및 제품 카테고리 수준에 대해서 비시드 소비자의 WOM에 미치는 다양한 파급효과를 살펴보았다. 이를 위해 장기간에 걸친 대규모의 데이터를 분석했는데, 시드 마케팅이 초점 제품의 WOM에는 주로 긍정적인 영향을 끼치지만, 동시에 같은 브랜드 내 다른 제품과 같은 카테고리 내 다른 브랜드의 WOM에는 부정적인 영향을 끼친다는 결과를 보고하였다. Jalali & Papatla (2019)는 브랜드들이 트위터 상에서 더 많이 리트윗되기 위한 전략을 알아보려고 프로모션 관련 단어, 브랜드 관련 단어, 이벤트 관련 단어, 시간 관련 단어 등 특정 주제와 관련된 단어의 위치를 살펴보았는데, 이러한 관련 단어들이 포스팅의 앞 부분에 나올수록 리트윗에 영향이 있다는 결과를 보고하였다. Berman et al. (2019)는 2016년 미국 대선을 중심으로, 생방송 토론회 도중

에 그에 대한 내용을 반영하는 트위터 게시물들이 어떠한 패턴으로 생성되고, 어떤 내용들이 더 활발하게 공유되는지 살펴보기 위해 트윗에서 사용된 단어를 분석하는 등 여러 자연어 분석 처리 방법을 사용하였다. Tirunillai & Tellis (2017)는 오프라인 텔레비전 광고가 온라인 채팅, 또는 사용자 생성 콘텐츠의 여러 지표에 어떠한 영향을 미치는지 살펴보았는데, 준실험(quasi-experiment)을 통해 광고가 실시되는 기간 동안 실험 대상 기업인 HP, 그리고 관련 경쟁사들을 합친 합성(counterfactual) 브랜드의 온라인 대화에 차이가 있는지 비교하였다. 이때, 소비자 리뷰 및 평가의 지표로 양, 감정, 극단성(polarity) 등을 측정하였으며, 특히 감정을 측정하기 위해 SVM과 나이브 베이즈 알고리즘을 사용하였다. 그 결과, 오프라인 텔레비전 광고는 온라인 채팅에 짧지만 긍정적인 영향을 미친다는 사실을 발견하였다. Adamopoulos, Ghose & Todri (2018)는 딥러닝 방법론 중 하나인 딥워크(Deep Walk)를 이용하여 사용자의 드러나지 않은 성격 특성이 소셜 미디어 플랫폼의 WOM의 활성화에 어떠한 영향을 끼치는지 살펴보았다. 그 결과, 메시지 발신자와 수신자의 성격 특성이 동질적일수록 WOM의 효과가 증가하였고, 발신자와 수신자의 성격 특성이 특정하게 조합되면 WOM의 효과에 영향을 미친다고 보고하였다.

앞서 살펴본 사용자 생성 콘텐츠는 대부분 텍스트 분석을 바탕으로 이루어졌는데, 텍스트 데이터는 그 외의 연구 문제에도 이용되었다. Homburg, Ehm & Artz (2020)는 미국 상장 기업의 연례 보고서에서 주주들에게 보내는 서신을 수집하여 마케팅 우수성(marketing excellence)을 표방한 서신을 분류하였는데, 여기에는 빈도 기반의 텍스트 분석과 SVM이 사용되었다. 저자는 마케팅 우수성을 표방하는 기업만을 포함한 포트폴리오가 연간 최고 8.58%의 초과수익률을 달성하였다고 보고하였다. B2B 세일즈 전략을 다룬 Singh, Marinova

& Singh (2020)의 연구에서는, 이메일로 이루어지는 세일즈 협상 상황에서 특정한 텍스트 내용이 어떠한 세일즈 전략을 의미하는지 분류하였고, 이를 통해 어떠한 전략이 구매자의 주의를 환기시켜 효과적인 거래로 이어지는지에 대하여 살펴보았다.

## 유형 2: 머신러닝을 이용하여 새로운 데이터를 활용하는 연구

두번째 유형은 첫번째 유형처럼 데이터에서 변수를 추출하는 것을 목적으로 하지만, 특히 머신러닝이 활성화되기 이전에 변수를 추출하기 어려웠던 새로운 형태의 데이터, 즉 비정형 데이터를 다루는 연구이다. 그 중에서 가장 활발하게 연구된 분야는 이미지이다. Li & Xie (2020)는 사진이 있는 소셜미디어 게시물이 사진이 없는 게시물보다 인기가 많은지, 그리고 어떤 사진 특성들이 인기를 견인하는지, 그 효과가 플랫폼이나 제품 특성별로 차이가 있는지를 살펴보았다. 이를 위해 텍스트의 감정, 토픽 등을 추출하였고, 이미지와 관련해서는 색감, 얼굴 인식과 감정 분석, 이미지 품질 등을 측정하였다. 한편, 이미지 중 인간의 얼굴을 분석하여 마케팅적 의미를 찾아내는 연구 또한 이루어졌다. Xiao & Ding (2014)은 광고에 나타난 모델의 얼굴이 소비자의 반응에 미치는 영향에 대해서 연구하였다. 이를 위해 피험자에게 얼굴만을 달리한 여러 광고를 보여주고 그들의 반응을 측정하는 실험을 시행하였다. 결과는 얼굴의 여러 특성이 소비자의 반응에 영향을 끼치며, 소비자 세그먼트 별로 선호하는 얼굴의 특성이 이질적이라는 것을 보였다. 이외에 Peng et al. (2020)는 전자상거래에서 외형이 매력적이지 않은 판매자가 평범한 판매자보다 더 많은 관심과 신뢰도를 유발하여, 외형의 매력도가 매출에 미치는 영향이 U자가 된다는 가설을 확인하였다. 저자는 이

미지 처리 기술을 이용하여 얼굴의 주요 특징을 추출하였으며, SVM 회귀, 랜덤 포레스트 회귀, CNN 등 여러 머신러닝 방법을 이용하여 얼굴 특징과 인간 평가자의 매력도 점수를 학습시켰다. 이를 독립변수로 하여, 종속 변수인 성과에 미치는 영향을 회귀분석과 이항 선택 모델(binary choice model)로 분석하였다. 또한, Zhou, Lu & Ding (2020)는 온라인 상에서 얼굴을 공개할 때 나타날 수 있는 프라이버시 문제를 해결하기 위해 얼굴의 윤곽 이미지 만을 제시하는 방법을 제안했는데, 얼굴 전체가 아닌 윤곽 이미지 만을 공개하더라도 타인이 얼굴을 인식하여 추론하는 데에는 큰 차이가 없다는 것을 보여주었다. Liu, Dzyabura & Mizik (2020)은 소셜미디어에 브랜드와 관련되어 공유된 사진을 통해 브랜드 이미지를 추출하는 방법을 제시하였다. 이를 통해 플리커(Flickr)에서 태그가 달린 사진을 수집하여 CNN 을 통해 훈련시켰고, 인간이 평가한 인스타그램 사진으로 정확도를 평가하였다.

또한, Li, Shi & Wang (2019)은 이미지 분석에서 나아가 동영상 분석을 시도하였다. 온라인 크라우드펀딩 사이트인 킥스타터(Kickstarter)의 데이터를 수집하여, 비디오 동영상이 포함된 프로젝트에 대해서, 동영상에 사람 및 악기가 포함된 경우 프로젝트 성공에 영향이 있을지에 대한 가능성을 제시하였다. 비디오 분석을 위해서는 CNN을 이용하였다. 또한 Lu, Xiao & Ding (2016)은 의류 매장에서 적용할 수 있는 비디오 기반 추천 시스템을 제안하였는데, 얼굴 인식과 과거 구매 내역을 이용하여 협업 필터링을 적용하여 추천이 이루어졌고, 단순 협업 필터링보다 성능이 더 좋음을 밝혀내었다.

그 외의 데이터 활용과 관련하여, Pfeiffer et al. (2020)는 랩의 가상현실 상황과 실제 슈퍼마켓에서 피험자가 물건을 구매하려고 물건을 검색(search)할 때의 아이트래킹(eye-tracking) 데이터를 활용한 support vector

machine을 이용하여 목표 지향, 또는 탐색적 동기를 예측할 수 있음을 보였다. 결과적으로 가상현실에서 80%, 그리고 실제 현실에서 85%의 정확도로 검색 동기를 예측할 수 있었다. Ghose, Li & Liu (2019)는 매장 내 고객의 위치 궤적을 이용한 상품 추천 전략의 효과를 살펴보고자 하였다. 이를 위해 실제 쇼핑몰에서 필드실험을 실시하였는데, 고객을 궤적에 따라 유사한 패턴을 가진 클러스터로 분류하고 그 특정 클러스터에 속한 사람이 자주 방문하는 상점에서 모바일 기기를 통한 추천을 진행하였다. 그리하여 궤적 기반의 모바일 타겟팅이 다른 전략에 비해 쿠폰 사용 확률이 높고, 더 빨리 사용하며, 거래 금액 또한 높다는 결과를 도출하였다.

### 유형 3: 새로운 방법론이나 현존하는 방법론의 개선을 제안하는 연구

본 유형은 방법론의 측면에서 특정 문제 해결을 위해 이전까지 사용했던 방법론과 다른, 머신러닝 기법을 제안하거나 이전의 방법들을 개선하고자 한 연구이다. 이 경우, 각 연구에서 제안한 방법론의 유용성과 우월성을 검증하기 위해 다양한 방법을 사용한다. Culotta & Cutler (2016)은 트위터에서 브랜드들의 소셜 네트워크를 분석하여 브랜드 인식을 추론하는 방법을 제안하였다. 이를 위해 트위터 상에서 200개가 넘는 브랜드의 소셜 네트워크를 추출하여 특정 속성을 대표하는 브랜드와의 비교를 통해 속성들의 점수를 매기고, 얻어진 결과를 설문조사 방법론과 비교하여 검증하였다. 본 방법론은 인간의 개입이 필요없이 자동화된 과정을 사용하여 효과적으로 브랜드 인식을 추출할 수 있다는 장점을 가지고 있다. Timoshenko & Hauser (2019)은 마케팅 인사이트를 얻기 위하여 사용자 생성 콘텐츠를 분석할 때, 유용한 정보가 들어있지 않거나 중복적인 내용이 들

어있는 경우 효과적인 분석이 어렵다고 주장하였다. 이에 저자들은 CNN을 이용해 유의미한 문장을 식별하는 방법을 제안하였다. 그들은 고객의 니즈를 파악하기 위한 목적의 인터뷰나 포커스그룹의 결과와 사용자 생성 콘텐츠를 분석한 결과를 비교하였는데, 사용자 생성 콘텐츠가 고객 니즈를 나타내는 소스로서 가치가 있으며, 머신러닝 기법으로 데이터를 분석하는 것이 고객 니즈를 파악하는 효율을 향상시킨다고 보고하였다.

Trusov, Ma & Jamal (2016)은 소비자의 온라인 서핑 데이터를 바탕으로 사용자 프로필을 분류하고, 정보가 제한적일 때에도 고객 프로필을 예측할 수 있는 접근법을 제안한다. 여기에서 사용된 모형은 토픽 모델의 일종인 상관관계 토픽 모델(Correlated topic model: CTM)인데, 텍스트로 이루어진 문서가 여러 개의 토픽으로 이루어진 것처럼, 사람의 웹사이트 방문 기록을 바탕으로 사용자 프로필을 분류할 때 여러 역할, 또는 목표의 조합으로 간주하는 것이다. 그 예로, 사회적 역할을 수행하기 위해 페이스북(Facebook.com), 쇼핑 역할을 수행하기 위해 아마존(Amazon.com)을 방문하는 것을 들 수 있다. 이 때, 본 모형은 사용자 간의 이질성을 고려하기 위해 인구통계 등과 같은 설명 변수를 포함하고, 웹사이트 방문 경향의 변화를 설명하기 위해 시간에 따른 역할의 진화 또한 고려한다. 그 결과로, 정보가 제한적인 상황에서도 효과적으로 사용자 프로필을 분류할 수 있었다. Hu et al. (2019)는 그루폰(Groupon)과 같은 일간 딜 웹사이트에서 딜에 대한 소비자의 탐색과 학습 과정을 설명하기 위해 베이지안 접근법의 일종인 동적 디리클레 학습 모델(dynamic Dirichlet learning model)을 제안한다. 모델에 의하면, 소비자는 첫번째 단계에서 제한된 정보를 가지고 가격을 검색할 것인지의 여부를 결정하고, 두번째 단계에서는 가격 검색을 한 후 구매 여부를 결정한다. 저자들은 그루폰 웹사이트에서

수집한 딜의 텍스트 및 이미지 데이터를 분석하여 임의의 딜에 대한 품질을 측정하고, 모델을 적용하여 구매를 유발하는 요인에 대해 살펴보았다. Yoganarasimhan (2020)은 사용자의 검색 및 클릭 기록을 기반으로, 검색 순위를 개인화하여 검색 품질을 향상시키는 시도를 하였다. 본 연구의 알고리즘은 크게 세 모듈로 이루어졌다. (1) 향후 모듈에 인풋으로 들어갈 속성 추출, (2) Normalized Discounted Cumulative Gain (NDCG) 기반 LambdaMART 알고리즘을 이용하여 검색 품질을 최적화하는 순위 시스템 도출, (3) Wrapper 기법을 이용한 알고리즘을 최적화하는 속성 선택 등이다. 본 연구에서는 러시아의 검색 엔진인 Yandex의 사용자 데이터를 이용하였는데, 사용자의 검색 히스토리를 여러 세션에 걸쳐 사용한 개인화 결과가 그렇지 않은 것에 비해 효과가 더 좋았다고 보고하였다.

패널 데이터 분석을 통해 소비자 선호나 선택을 효과적으로 측정하는 방법 또한 제안되었다. 전통적으로 패널데이터의 분석은 다항 로짓(multinomial logit)이나 프로빗(probit) 모형을 적용하여 이루어졌다. 그런데 특정 제품 카테고리의 구매 결정은 그 카테고리 내의 과거 구매와 다른 연관된 제품 카테고리의 과거 구매 이력과 관련이 있다. 과거의 분석방법은 제한된 갯수의 카테고리만을 고려하였고, 시간의 흐름에 따른 복잡한 관계를 반영하지 못하였다. Xia, Chatterjee & May (2019)는 조건부 제한 볼츠만 머신(conditional restricted Boltzmann machine)을 이용하여 방대한 소비자 패널 데이터셋의 분석에 시간의 흐름과 서로 다른 제품 카테고리를 반영할 수 있음을 보였다. 컨조인트 분석과 같은 분해적 선호 추출법(decompositional preference elicitation method)은 제품의 속성이 복잡해지거나 속성이 많아지게 되면 소비자의 선호를 추출하기가 어렵다는 단점이 있다. Huang & Luo (2016)의 연구에서

는 복잡한 속성을 가지고 있는 제품에 대한 소비자 선호를 이해하기 위해서 퍼지 SVM 능동적 학습(fuzzy SVM active learning) 알고리즘을 제안한다. 이를 위해 초기의 부분가치(initial partworths) 값은 협업 필터링을 이용하여 비슷한 선호를 가진 사람의 것으로 지정을 한 후, 응답자의 응답에 따라 변화하는 일련의 적응형 질문들을 통해 고려군(consideration set)과 최종 선택 제품에 대한 정보를 얻는다. 동시에, 이를 이용하여 즉석에서 개인별 부분가치를 도출해 낸다. 제안된 알고리즘은 기존 방법에 비해 고차원 속성 제품에 대한 선호 측정 문제에 적합하며, 응답자 측면에서도 질문의 양이 적기 때문에 인지적 부담을 덜어줄 수 있다는 장점이 있다. Liu & Toubia (2018)는 소비자의 선호가 검색 활동에 어떻게 반영되는지, 그리고 검색 결과 콘텐츠 선호도와와의 관계는 어떠한지 파악하고자 하였다. 그리하여 소비자의 검색 쿼리가 검색 결과와 의미적으로 연관이 있다는 데에 착안하여 계층적 이중 LDA(hierarchically dual latent Dirichlet allocation; HDLDA) 모형을 제안하였다. 본 방법론은 랩 실험과 필드 실험을 통해 검증되었다.

Ansari, Li & Zhang (2018)은 제품 속성, 사용자 평점, 제품 설명 태그 등을 이용하여 제품을 잠재 토픽으로 분류하고 토픽들을 통해 사용자의 선호를 설명하는, 지도학습 기반의 토픽 모델을 제안하였다. 이때 데이터의 규모와 양을 고려한 Stochastic variational Bayes 기법을 통해 추정의 속도와 정확도를 향상시켰는데, 이는 실제 추천 상황에서 유용하다. 본 방법론을 실제 영화 데이터에 적용하였을 때, 벤치마크 모델들보다 예측 정확도가 높고, 소비자 선호에 대한 흥미로운 통찰을 제공하였다. Danaher et al. (2020)는 여러 브랜드가 속한 제품 카테고리 내에서 멀티미디어 광고 매체에 대한 개인별 수준의 매출 효과를 살펴보기 위해 2년 간

에 걸친 대규모의 패널데이터를 이용하였다. 저자는 브랜드 내, 매체 내 상관관계와 개인별 수준의 광고 반응 추정을 반영하는 토빗(Tobit) 모델을 고안하였는데, 그러한 복잡성에서 기인하는 복잡한 임의효과(random effect)를 해결하기 위해 variational Bayes를 사용하여 모델을 추정하였다. 그 결과 특정 브랜드의 이메일과 카탈로그 광고는 다른 브랜드에 부정적인 영향을 끼치지 않지만, 유료 검색은 자사의 판매에만 영향을 끼친다는 것을 발견하였다. 또한, 여러 브랜드를 소비하고 옴니채널(omni-channel) 쇼핑을 즐기는 고객들이 멀티미디어 광고에 더 큰 영향을 받는다고 보고하였다.

더불어, 토픽모델링의 한계점을 고려하여 개선된 방식들도 함께 고려되었다. Toubia et al. (2019)의 연구는 고객의 선호에 영향을 끼치는 엔터테인먼트 상품의 속성은 추출하기 어렵다는 것에 착안하여 이를 보완한 guided LDA 기법을 제안하였는데, 연구자가 설정한 테마를 기반으로 하여 텍스트에서 새로이 나타난 속성들도 반영될 수 있도록 하였다. 본 연구에서는 guided LDA를 통해 도출된 속성들이 기존의 모델들보다 소비자 선택을 잘 설명한다고 보고하였다. 한편, 기본적인 LDA의 한 가지 특징 내지 한계점은 문서 데이터셋 전체에 대해서 주제의 분포가 고정되어 있다고 가정하고 시간에 따른 변화를 허용하지 않는다는 것이다. Dew, Ansari & Li (2020)는 베이지안 프레임워크를 기본으로 하여, 기존에 많이 연구되었던 횡단적 이질성(cross-sectional heterogeneity)과 달리, 잠재변수(latent variable)가 시간이 지남에 따라 변할 때의 이질성을 찾아낼 수 있는 새로운 방법론을 제시한다. 여기에서 데이터의 분포 가정은 다양하게 적용할 수 있어서, 개별 수준 모델(individual-level model)은 다항 로짓이 될 수도 있고, 토픽 모델이 될 수도 있다. 본 연구에서는 한 가지 사례로, 아마존의 제품 리뷰의 주제가 시간이 지남에 따라

다르게 강조될 수 있다는 가정을 바탕으로 하여 LDA보다 발전된 동적 토픽모델을 제시한다. 또다른 연구는 Zhong & Schweidel (2020)의 연구인데, 소셜 미디어 콘텐츠 내용의 근본적인 변화를 감지하는 방법을 제시하였다. 예를 들어 2015년 9월에 발생한 자동차 배기가스 테스트 스캔들 당시 폭스바겐 자동차 브랜드와 관련된 소셜 미디어 게시물이 어떻게 변화하였는지를 밝혀낼 수 있다. 저자는 소셜 미디어의 토픽이 논의되는 빈도가 시간이 지남에 따라 변경될 수 있다고 가정하는 LDA-MLC(Latent Dirichlet Allocation with Multiple Latent Changepoints) 기법을 사용하여, 콘텐츠가 변화하는 지점을 감지해냈다.

한편, 동적 구조 모델(dynamic structural modeling)을 다룬 연구에서 정책함수(policy function)와 상태 전환 프로세스(state transition process)를 추정하는 경우에 동적 프로그래밍(dynamic programming)을 이용하는 경우가 많은데, 최근에는 구조 모델 내의 복잡한 수치해석학적(numerical) 이슈들을 처리하는 데에 머신러닝이 활용되고 있다. 한 예로, Liu, Montgomery & Srinivasan(2018)은 은행 계좌에서 초과 인출이 발생할 확률을 예측하는 것을 목표로, 소비자의 반응(소비자의 부주의, 모니터링 비용, 계좌 해지 등의 행동)을 통합하는 동적 구조 모델을 개발하였다. 저자들은 은행의 초과 인출 수수료 정책과 관련한 반사실적 분석(counterfactual analysis)을 수행하기 위해 최신 병렬 컴퓨팅 기술을 도입하였고, 정책에 영향을 받지 않는 파라미터를 효과적으로 추정하였다. 또한, 개인별로 특정 날짜에 계좌 초과 인출이 일어날 가능성과 그에 따른 소비자의 계좌 유지 및 은행의 수익성에 미치는 영향을 예측하였고, 이를 통해 여러 수수료 책정 체계의 효과를 비교할 수 있었다. 그 결과 새로운 수수료 체계가 은행의 수익을 증가시킬 수 있다는 것을 보여주었다. Kim,

Sudhir & Uetake (2019)는 영업사원이 다차원적인 성과 보상 체계 하에서 고객 개인 정보를 가지고 수행하는 다양한 대출 유치와 상환 활동을 동적 구조 모델로 구현하였다. 추정은 두 단계로 이루어지는데, 이중 1단계의 정책함수(policy function) 추정을 위해 영업사원의 이질성과 영업사원만이 알고 있는 고객의 정보를 반영할 수 있는 랜덤 포레스트 기법을 사용하였다. 그 결과로, 영업사원에는 잠재적인 두 개의 세그먼트가 있는데, 대출 유치를 더 잘하는 “수렵자” 유형과 대출 상환에 더 효율적인 “농부” 유형이 있다는 것을 보여주었다. 한편, Dzyabura & Hauser (2019)의 연구는 동적 프로그래밍을 활용한 새로운 추천 시스템을 소개한다. 소비자들은 제품 검색을 하면서 제품 속성에 대한 선호도가 중치를 배우고 업데이트하게 되는데, 기존의 추천 시스템들은 소비자의 선호도가 가중치가 업데이트되면서 검색된 모든 제품에 대한 효용과 아직 검색되지 않은 제품들에 대한 잠재적인 효용 또한 바뀌는 것을 반영하지 못하였다. 연구에서 제안한 추천 시스템은 이러한 가중치 업데이트를 반영하며, 다양한 속성을 가진 제품을 추천하고, 소비자가 과소평가하는 제품, 또는 소비자의 기존 인식을 바꿀 수 있는 제품들을 추천하여 소비자의 향후 검색이 더 효과적으로 이루어지도록 한다. 이때 소비자의 최적 검색 경로는 동적 프로그램의 벨만 방정식(Bellman Equation)에 대한 해법과 맥을 같이 하는데, 소비자가 검색을 통하여 속성에 대한 선호도가 중치를 업데이트할 수 있도록 재귀 구조(recursive structure)를 조정하였다.

한정된 자원을 배분하는 결정에 강화학습에서 많이 연구된 멀티암드밴딧(multiarmed bandit: MAB) 문제를 적용하는 여러 연구도 이루어졌다. Schwartz, Bradlow & Fader (2017)은 효과적인 광고를 선택하기 위한 테스트를 진행할 때 여러 광고 버전을 균등하게 노출하고

향후에 가장 효과가 좋은 광고로 전환하는 기존의 방법 (earning vs. learning)과 달리 학습과 수익창출을 동시에 진행하는 방법론(earning while learning)을 제안하였다. 관찰되지 않는 이질성(서로 다른 웹사이트에 등장하는 광고의 효과 차이 등) 때문에 기존의 MAB문제보다 고려할 내용들이 더 많고 복잡한 문제가 되는데, 톰슨 샘플링(Thompson Sampling; 또는 randomized probability matching) 방법을 적용하여 필드 실험을 통해 검증한 결과 대조군 정책보다 8% 더 높은 고객 획득률을 보였다. Misra, Schwartz & Abernathy (2019)는 불완전한 수요 정보를 가지고 대규모 제품들의 가격을 실시간으로 책정하는 문제에 대해서 MAB에 미시경제 선택이론을 반영한 방법론을 제안한다. 기존의 가격 변화에 대한 매출을 모두 테스트하여 최적의 가격을 산출하는 ‘균형 실험법(balanced experiment)’와는 다르게, 학습하는 동안 수익을 극대화한다는 목표를 갖는 동적 최적화 문제를 구성하였다. 저자들은 기존의 필드 가격실험을 기반으로 한 시뮬레이션에서 한 달 동안의 수익이 43%, 연간으로는 4%까지 증가하였다고 보고하였다. 이와 관련하여, Hansen, Misra & Pai (2021)의 연구에서는 경쟁자들이 자동화된 알고리즘을 이용하여 가격 조정을 한 결과로 담합을 형성할 수 있다는 가능성을 보여준다. 이때 판매자는 동적 가격실험을 통해 수요 곡선과 각 제품의 수익 극대화 가격에 대해 학습하여 가격을 도출하는데, 온라인 시장의 복잡성과 규모를 고려할 때 기업을 독점기업, 또는 고정된 경쟁업체의 가격에 가장 잘 반응하는 과점 기업이라고 가정한다. 일련의 시뮬레이션을 실행한 결과, 가격 실험의 정보 가치(본 논문에서는 signal-to-noise ratio, SNR로 표현)가 증가함에 따라 경쟁자들의 가격의 상관성이 높아지고, 경쟁 시장의 가격보다 높은 가격(supra-competitive price), 즉 담합가격을 형성하게 되었다.

데이터 분석을 위한 여러 기법들을 비교하는 연구도 있었다. Vermeer et al. (2019)는 e-WOM을 추출하기 위한 다양한 감성 분석, 사전(dictionary) 기반 및 머신러닝 기술을 비교하여 그 성과를 평가하였다. 그 결과로, e-WOM을 분류하는 데 있어서 단어 세기(word count) 등의 감성 분석이나 사전 접근 방식보다 그 외의 머신러닝 알고리즘들의 분류 성과가 더 좋다고 보고하였다. 그 밖에, 머신러닝 기법을 이용하여 데이터에 적용할 수 있는 후보 모델들을 선택하는 연구도 있었다. Schwartz, Bradlow & Fader (2014)의 논문에서는 분류 및 회귀 트리(Classification and Regression Tree; CART) 및 랜덤 포레스트 등의 머신러닝 알고리즘을 적용하여, 실무에서 관찰 가능한 고객 행동 패턴을 기반으로 다양한 상황에서 적용할 수 있는 모델을 제안할 있는 일반적인 규칙을 도출하였다. 또한 실제 데이터와 시뮬레이션 데이터를 사용하여 본 방법론의 가능성을 검증하였다. Simester, Timoshenko & Zoumpoulis (2020)는 기업이 신규 고객을 확보할 때 프로모션 대상을 정교화하기 위해 현장 실험 결과를 어떻게 사용할 수 있는지 알아보았다. 저자는 미국의 멤버십 기반 대형 소매 업체와 수행한 대규모의 현장 실험을 통해 커널 회귀(kernel regression), k-NN(k-Nearest Neighbor), SVM을 포함하는 7가지 기계 학습 모형을 구성하고, 각 방식의 타겟팅 성능을 평가하였다. 두 번째 현장 실험을 통해, 모델 기반의 라쏘 회귀(Lasso regression) 및 유한 혼합 모형(finite mixture model)이 거리 기반의 k-NN, 커널 회귀, 위계적 클러스터링(hierarchical clustering)과 분류 기반의 SVM 등의 기법보다 더욱 나은 성능을 보임을 확인하였다.

유형 4: 머신러닝을 이용하여 마케팅 현상의 패턴을 설명하는 연구

세번째 유형의 연구는 변수들 간의 상관관계나 인과관계를 규명하기 보다는 마케팅 현상을 묘사하고 설명하는데 초점을 맞춘 연구이다. 다시 말해서, 주로 기술적 연구(descriptive research)에 해당한다. Moon, Kim & Iacobucci (2020)는 텍스트 마이닝 기법을 이용하여 온라인 리뷰를 긍정적인 진짜 리뷰, 긍정적인 가짜 리뷰, 부정적인 진짜 리뷰, 부정적인 가짜 리뷰의 네 가지로 분류하였다. 그 결과로 가짜 리뷰는 감정적 과장, 특정 대명사 사용(1인칭 대명사 사용 감소), 콘텐츠의 세부 정보 감소 등의 특징이 있음을 발견하였다.

Dzyabura, Jagabathula & Muller (2019)는 소비자 선호를 측정할 때, 상대적으로 친숙한 제품이라도 현실에서 제품을 평가할 때와 온라인 설명을 바탕으로 평가할 때의 불일치가 발생하는 현상에 대해서 알아 보았다. 일련의 참가자들이 두 개의 컨조인트 실험(온라인과 오프라인, 순서는 랜덤)을 완료하도록 하여 두 실험 결과 간의 통계적으로 유의한 차이가 있는지 살펴보았다. 온라인과 오프라인 실험 양쪽 모두 참여한 사람들의 숫자가 온라인 실험만 참가한 사람들의 숫자에 비해 적은 것을 고려하여 개인의 오프라인 부분가치를 추정하기 위한 계층적 베이저안(hierarchical Bayesian) 접근법과 머신러닝 알고리즘인 k-NN의 두가지 기법을 제안하였다. 그 결과 온라인과 오프라인 실험에서 얻은 부분가치의 괴리가 발견되었으며, 향후 온라인 실험을 보완하는 과정이 포함되어야 한다고 주장하였다.

데이터를 통하여 시장 구조를 설명하고자 하는 연구도 있었다. Gabel, Guhl & Klapper (2019)가 제안한 P2V-MAP는 확장 가능한 시장구조 분석 및 시각화 방법론이며, 머신러닝과 자연어처리 분야의 SG/t-SNE 알고리즘을 사용한다. 잠재 속성 추출 및 시장 장비구니 데이터 기반 모델링에는 Skip-gram(SG) 모형이 사용되었고, Barnes-Hut t-distributed stochastic

neighbor embedding (t-SNE) 알고리즘을 이용하여 시장 구조 표현을 위한 2차원 지도를 생성하였다. 저자들은 독일의 147개 대형 매장을 12개월 간 관찰한 데이터를 이용하여 본 방법론의 유용성을 검증하였다. 또다른 접근법으로, Ringel & Skiera (2016)는 수백개의 제품으로 이루어진 대규모 시장에서 경쟁 구조를 효과적으로 시각화할 수 있는 방법을 찾고자 하였다. 그리하여 DRMAS(Decomposition and Reassembly of Markets by Segmentation)이라는 새로운 모델을 개발하였는데, 이것의 기본 원칙은 전체 시장을 하위 시장으로 분해하고 상대적 위치를 결정한 후 지역 하위 시장 구조를 전체 시장으로 다시 통합하는 것이다. 본 모델의 유용성을 보이기 위해 클릭스트림 데이터를 사용하여 1,000개가 넘는 제품으로 이루어진 대형 내구재 카테고리에서 시장 구조를 시각화하는 사례를 소개하였다.

Klosterman et al. (2018)는 인스타그램의 이미지와 텍스트를 분석하여 소비자의 브랜드 인식을 설명하는 연구를 시행하였다. 게시물에 포함된 이미지를 수집하여, 이미지 클러스터링 분석을 통해 총 20개의 클러스터를 도출하고, 각 클러스터별로 자주 출현하는 레이블의 빈도를 측정하였다. 이것을 바탕으로 시각화 도구를 통해 각 태그의 연결성을 확인하였으며, 이를 통해 여러 클러스터들을 비교할 수 있다는 것을 보여주었다.

유형 5: 머신러닝을 이용하여 마케팅 현상을 예측하고, 예측력 향상에 영향을 미치는 변수들을 살펴보는 연구

본 유형의 연구는 머신러닝을 통해 특정 현상을 예측, 분류하고 예측의 정확도에 영향을 미치는 변수들이 무엇인지 알아보는 데에 초점을 맞춘 연구이다. 유형 1의 연구와 다른 점은 통계적인 인과관계를 규명하는 것보다는



예측력 향상에 초점을 맞추었다는 것이다. 주로 정답을 알고 있는 데이터를 이용한 지도학습 위주의 연구가 많다. Liu, Singh & Srinivsan (2016)은 트위터의 텍스트 데이터를 분석하여 텔레비전 시리즈와 미식축구 경기의 시청률을 예측해 보았다. 분석 결과, 트윗의 양이나 긍/부정의 감정 같은 표면적인 지표들보다 트윗이 담고 있는 정보나 시의적절성이 예측력을 개선한다고 보고하였다. 또한, 트위터 외의 다른 온라인 데이터(구글 트렌드, 위키피디아 페이지, 영화전문 사이트인 IMDB 리뷰 등)은 시청률을 잘 예측하지 못한다는 것을 발견하였다. Mejjia, Mankad & Gopal (2019)은 레스토랑의 온라인 리뷰에서 제공되는 정보로 레스토랑의 위생을 측정하고 위생 위반 여부를 예측하는 연구를 진행하였다. 이를 위해 미국 뉴욕시 위생점검 점수 데이터를 수집하고 리뷰 사이트인 옐프(Yelp)의 레스토랑 리뷰 데이터를 이용하여 위생 관련 사건을 구성하였는데, 시의 위생 점검에서 A 등급을 받은 후 리뷰에서 위생상태가 퇴화되는 것으로 나타난 식당을 판별하여 식당들의 도덕적 해이를 발견할 수 있었다.

Rutz, Sonnier & Trusov (2017)는 온라인의 유료 검색 광고의 효과를 측정하기 위해 두 단계의 분석을 시행하였다. 첫째, 고객의 인식과 클릭 성과를 측정하기 위해 광고 두 쌍을 평가하는 실험을 통해 데이터를 약 15,000개 수집하였고, Elo 알고리즘을 이용하여 각 광고에 대한 상대적인 인식과 클릭 성과를 예측하는 방식으로 고객의 인식과 실제 클릭 행위 사이의 관계를 규명하고자 하였다. 둘째는 새로운 광고의 효과를 살펴보기 위해 bag-of-words와 LDA 등의 텍스트 분석과, 벌점 회귀(penalized regression) 모델인 VANISH 모델을 베이저안 접근법으로 적용하여 온라인 검색광고의 효과를 높이는 변수들이 무엇인지 살펴보았다.

Netzer, Lemaire & Herzenstein (2019)는 대출

신청 시 차용인이 작성한 텍스트를 통해 채무 불이행을 예측하고자 하였다. 이를 위해 나이브 베이즈, LDA 등의 방법론을 이용하였고, 대출을 위해 작성하는 서류의 텍스트가 자신의 상태와 특징을 나타내며, 향후 상환을 예측하는데 효과적이라는 것을 보였다. Jindal (2020)은 광고와 R&D가 기업의 파산 생존률에 어떤 영향을 미칠까라는 주제로, 기업의 파산 생존률을 예측하기 위해 여러 통계적 기법과 더불어 랜덤 포레스트 기법을 이용하였는데, 이 기법을 적용하면 다른 기법들보다 예측률이 5.05%p 향상된 것으로 나타났다.

#### 유형 6: 마케팅 문제에 대한 해결책을 제시하고 인과 관계를 규명하는 연구

본 유형의 연구는 머신러닝 기법을 통해 타겟팅 등 마케팅 활동을 최적화하는 방안을 제시하는 연구이다. 이때, 제시한 방안이 실제로 효과적인지 판단하기 위해 실험을 이용하는 연구가 근래에 활발히 이루어지고 있는데, 최근에는 실제 소비자를 대상으로 한 필드실험이 많이 진행되었다. 즉, 머신러닝을 통해 최적화된 활동, 즉 처치효과를 기존의 방법, 또는 아무 처치도 하지 않은 대조군과 비교하여 어떠한 방법이 더 나은지 살펴보는 것이다.

Fong et al. (2019)는 모바일의 타겟 프로모션이 제품 또는 유사제품 구매를 촉진하지만 검색 활동을 줄임으로써 관계가 적은 제품의 구매에는 부정적인 영향을 끼치지 않을까라는 가설을 바탕으로 효과적인 타겟팅 방법을 강구하였다. 공변량 공간(covariate space)을 예측 오류를 최소화하는 트리로 분할하는 코질 트리(causal tree) 기반의 머신러닝 방법론을 사용하여 홍보 제품 판매와 총 판매량을 동시에 최적화하는 타겟팅 방법의 효과를 필드실험을 이용하여 살펴보았다. Luo,

Lu & Li (2019)는 전자상거래의 구매율을 높이기 위해 장바구니 포함 품목에 대해서 가격 인센티브 또는 재고 희소성 알림 등의 방법을 통한 전자상거래 장바구니 타겟팅(e-Commerce Cart Targeting; ECT)의 효과를 살펴보았다. 이를 위해 랜덤 포레스트를 이용하여 각 고객별 가장 효과적인 타겟팅 방법을 찾았고, 그 효과를 확인하기 위해 22,000명 이상의 모바일 이용자를 대상으로 한 현장실험을 실시하였다. 한편, 최근에는 신규 고객 유치보다는 현 고객 유지를 위한 마케팅 활동(customer retention program)의 중요성이 대두되었다. 일반적으로는 고객의 이탈율(churn probability) 또는 유지 오퍼(retention offer)에 얼마나 잘 반응할지를 예측하여 타겟팅을 하는 접근법이 이용되었다. Lemmens & Gupta (2020)의 연구에서는 수익 기반의 손실함수를 설정하여 한 고객 당 고객 유지 활동의 금전적 가치를 예측하고자 하였다. 최적화를 위해서는 경사 하강법을 기반으로 하는 확률적 경사 부스팅(stochastic gradient boosting; SGB) 알고리즘을 사용하였고, 필드 실험을 이용하여 방법론을 검증하였다. 그 결과 제안된 모델이 비교 대상인 다른 모델들보다 고객유지 활동의 수익성이 증가하는 것을 보였다.

그동안 주로 예측 문제에 적용되었던 랜덤 포레스트 기법 등은 인과관계 추론을 위해서도 많이 사용되고 있다. 이는 컴퓨터 공학 분야에서 흔히 쓰이는 지도학습 방법으로, 큰 정교화 과정 없이 즉시 사용 가능하며 안정적인 성능을 보인다고 알려져 있다(Athey & Imbens 2017). 특정 마케팅 활동의 효과를 살펴보기 위하여 실험에서 처치효과를 살펴볼 때, 처치를 받지 않은 대조군의 종속변수를 예측하거나 실험군과 짝을 지을 대조군을 찾아내는 데에 이러한 알고리즘의 예측 기능을 이용하고, 이를 통해 반사실적 분석을 시행할 수 있다. Rafieian & Yoganarasimhan (2020)은 모바일 인앱(in-app)

광고 생태계에서 개인화된 정보를 사용하여 타겟팅을 하는 이익과 개인정보 침해의 손해의 균형을 찾고자 하였다. 그리하여 개인정보를 이용하지 않는 상황기반(context-based) 타겟팅, 과거 개인 정보를 사용하는 행동기반(behavioral-based) 타겟팅 등을 비교하였는데, 특정 타겟팅 기법을 이용하여 노출된 광고와 실제로 노출되지 않은, 다른 타겟팅 기법을 이용한 광고의 효과(click-through rate; CTR)를 XGBoost를 활용한 머신러닝 기법을 이용하여 각각 예측하였고, 그 후에 분석적 경매 모델(analytical auction model)을 이용하여 각 국면(regime) 하에서 경쟁관계, 그리고 광고주와 광고 네트워크의 수익을 분석하였다. Ascarza (2018)는 이탈 위험이 가장 높은 고객을 식별하는 이탈 예측 모델을 사용하는 것이 고객 관리 측면에서 최선의 선택이 아니라는 것을 주장하는데, 개입을 했을 때와 안 했을 때의 이탈 가능성의 차이가 큰 고객을 선택하는 방법과 비교하였다. 이때 후자의 고객을 예측하기 위해 일종의 랜덤 포레스트 기법을 적용하였고, 두 개의 현장 실험 데이터를 이용하여 고객 유지 캠페인의 효과를 시뮬레이션 하였다. Guo, Sriram & Manchanda (2021)는 미국 내 제약회사들이 의사들에게 지불하는 마케팅 비용을 공시하도록 한 정책의 효과를 살펴보았다. 이를 위해 코절 랜덤 포레스트를 이용하여 정책의 전과 후에 해당하는 의사-약품 쌍을 매칭하여, 준실험 이중차분법(Difference-in-Difference; DID)을 적용하였다. 그 결과로, 전체적인 비용에는 통계적으로 유의한 증가가 없었지만, 고가의 약품과 처방전을 대량으로 작성하는 의사에게 비용이 집중되었다는 것이 나타났다.

뿐만 아니라, 코절 포레스트(causal forest)의 유연한 구조를 이용하여 모집단 수준이 아닌, 이질적인 개별 유닛 수준에서의 처치효과를 살펴볼 수 있게 되었다. Chen et al. (2020)은 간암 고위험 환자의 예방적 검진 지

원(outreach) 활동의 효과를 알아보기 위해 무작위 필드실험을 실시하였다. 그 결과, 단순 지원 활동 실험군과 환자 내비게이션(patient navigation) 지원 활동 실험군의 주효과가 대조군과 비교하여 개선된 것으로 나타났다. 그러나, 두 실험군 사이에는 통계적으로 유의한 차이가 없었다. 하지만, 코질 포레스트 알고리즘을 이용하여 환자들이 개별적 특성에 따라 지원 활동에 다르게 반응하며, 환자 수준의 타겟팅을 이용하면 의료 시스템의 수익 개선에 도움이 된다는 결과를 밝혀내었다.

#### IV. 논의 및 결론

본 연구에서는 마케팅에서 머신러닝을 적용한 여러 연구에 대해서 연구 목적에 따라 여섯 가지 유형으로 나누어 살펴보았다. 2010년대에 들어서면서 국제적으로 다양한 주제의 머신러닝을 활용한 마케팅 연구가 이루어져 왔다는 것을 알 수 있는데, 이에 비해 국내 연구는 부족한 실정이다(강우성, 정환, 김영찬 2015). 과거에 인공신경망과 로짓 모형을 사용하여 대형 TV 등 내구재에 대한 소비자의 구매의도를 예측한 연구(한상만, 박승배, 정남호 2000)가 있었으나, 2015년 이후를 살펴보면 신경망 모델을 사용하여 가맹사업의 매출을 예측하고 모델 간의 예측력의 차이를 살펴보았던 연구(김주영 2018), 토픽 모형 분석을 활용하여 시장을 세분화하는 새로운 접근법을 제시했던 연구(김종대, 이유석 2019), 빅데이터 분석을 통해 기부연구의 흐름을 확인했던 연구(오민정 2020) 등 제한적인 분야의 연구에서만 머신러닝이 활용되었다. 따라서 국내에서 머신러닝을 이용한 연구가 활성화되기 위해 몇가지의 연구 방향을 제시한다.

첫째는 오디오, 이미지, 비디오, 텍스트, 위치 등의 비정형 데이터 활용이다. 이러한 데이터들은 정보 추출과

정을 통해 정형 데이터로 변환이 되어야 하고, 그 과정에서 기술이 요구되기 때문에 진입장벽이 높은 편이다. 하지만 반대로 생각하면, 그동안 연구가 많이 되지 않았기 때문에 많은 기회가 있다. 데이터 수집을 위해 크롤링(crawling)을 통해서 온라인 상에 공개되어 있는 사용자 생성 콘텐츠를 활용할 수 있으며, 최근 정부에서 개방하고 있는 공공데이터를 활용하거나 필요한 데이터를 요청하는 것도 가능하다.

둘째는 변수들 간의 인과관계를 규명하는 것이다. 전통적으로, 마케팅 연구, 나아가 사회과학 연구자에서는 변수들 간의 인과관계를 확립, 즉 가설을 설정하고 샘플 데이터와 통계적인 모형을 통해 가설을 검증하고 그것을 이론으로 만드는 데 초점을 맞춰왔다. 다시 말해서,  $x$ 와  $y$ 의 관계를 잘 설명하는 함수를 설정하고(예: 선형회귀식), 함수의  $\beta$ 값을 추정하는 것이다. 반면에, 머신러닝에서도  $x$ ,  $y$ , 그리고 그 둘의 관계를 설명하는 함수가 있지만 대부분의 머신러닝 알고리즘은  $x$ ,  $y$ 의 관계 중  $\beta$ 에 대해서는 큰 관심을 두지 않고, 회귀식의  $\beta$ 를 도출할 수 있다고 하더라도 이 추정량은 일치성(consistency)을 만족시키지 않기 때문에 통계학에서 지향하는 추정량의 특성과 어긋난다. 관계식 또한 선형이 아닌 여러 형태를 띠 수 있다(Mullainathan & Spiess 2017). 반면에, 머신러닝의 가장 큰 장점이라고 하면 굉장히 많은 양의  $x$ 를 다룰 수 있고,  $y$ 값을 예측하는 데 있어서 뛰어난 성능을 보이는 것이다. 대부분의 머신러닝 알고리즘들이 비록 블랙박스의 특징을 가지고 있지만 앞선장에서 소개한 연구 유형 6에서 살펴보았듯이 실험군과 대조군을 머신러닝으로 매칭하여 DID 등의 기법을 이용하여 인과관계를 규명할 수도 있고, 자연실험(natural experiment)의 일환으로 특정 이벤트의 효과를 살펴볼 수도 있다. 또한, 코질 트리나 포레스트 알고리즘을 이용하여 전체 처치효과 뿐만 아니라 개별 유닛에 대한 처

치효과도 살펴볼 수 있다.

머신러닝을 적용하는 데 있어서의 이슈는 규모 있는 학습데이터를 구축하는 것과 데이터의 품질 문제이다. 첫째, 컴퓨터가 학습을 하기 위해서는 학습데이터를 구성해야 하는데, 머신러닝 알고리즘들은 대개 많은 양의 데이터를 요구하고 어느 정도 성과에 도달하기 위해서 적게는 수만개에서 많게는 수백만개의 데이터포인트를 필요로 한다(Halevy et al. 2009). 문제는, 특히 지도 학습의 경우에 컴퓨터가 달성할 수 있는 성과는 인간이 어떠한 작업을 했을 때의 성과를 목표로 하기 때문에 인간이 레이블링한 학습데이터를 사용하는 경우가 많으며, 큰 규모의 데이터를 인간이 직접 레이블링하는 것은 어려움이 따를 수 밖에 없다. 한가지 해결책으로는, 다른 상황에서 이미 학습된 모델 파라미터를 초기값으로 사용하여 초반에 모델이 어느 정도 안정이 될 때까지의 데이터 요구량을 줄이고 적은 규모의 데이터셋을 구성하여 다시 학습시키거나 정교화하는 방법이 있다. 이러한 기법을 전이학습(transfer learning)이라고 한다. 전이 학습은 이미지 분류 등의 작업에서 큰 성과를 보였다(Krizhevsky et al. 2017, Simonyan & Zisserman 2014). 하지만, 이 또한 훈련 데이터를 적게나마 구축해야 한다는 단점이 있다. 또한, 이미 학습된 모델에서 쓴 데이터와 연구에서 적용하고자 하는 데이터의 괴리가 큰 경우 이러한 방법을 사용하는 것이 덜 효과적이다.

둘째는 학습 데이터 자체의 품질이다. 머신러닝 알고리즘은 데이터를 가장 잘 설명하는 패턴을 찾는 것이기 때문에 학습 데이터의 품질이 굉장히 중요하다. 예를 들어, 온라인 상의 가짜 제품 리뷰(fake review)를 분류하는 머신러닝 연구는 사람이 평가하여 레이블링한 내용이 정확하다는 전제를 두고 시작한다. 만약에 사람의 데이터 분류 능력에 의문이 있다면 컴퓨터가 이를 바탕으로 학습한 모델도 정확도가 떨어질 수 있다. 최근에는

인간이 효과적으로 잘 식별할 수 없는 작업들이 컴퓨터에 의해서 더 잘 분류된다는 연구 또한 발표되었다(Clarke et al. 2020). 또한 이미 저장되어 있는 데이터도 데이터의 일부 정보가 없거나, 부정확하거나, 중복되거나, 시기에 뒤떨어지는 등의 문제가 있다. 이 밖에도 데이터가 편향되어 있다면 모델의 성과와 결과의 일반화에 영향이 있을 수 있다. 따라서, 학습 데이터를 구축함에 있어서 데이터를 선별하고 전처리하는 과정이 중요하다(Obermeyer & Emanuel 2016).

앞선 장에서 살펴보았듯이, 머신러닝은 연구자들에게 새로운 방법론을 제공하는 동시에 그전에 여러 이유에서 분석이 불가능했던 새로운 데이터를 다루고, 새로운 문제를 해결할 수 있는 가능성을 주었지만 한계 또한 있다고 할 수 있다. 마케팅 연구자들이 머신러닝의 장점과 단점을 파악하여 적절한 연구 주제에 적용한다면 그전에 얻지 못한 새로운 인사이트를 얻을 수 있을 것이다. 본 리뷰가 아직 머신러닝을 접해보지 못한 마케팅 연구자들이 새로운 방법론을 적용하는 데에 출발점이 되기를 바란다.

〈최초투고일: 2021년 1월 22일〉

〈수정일: 1차: 2021년 2월 16일〉

〈게재확정일: 2021년 2월 18일〉

## 참고문헌

- 강우성, 정환, 김영찬(2015), 계량마케팅 연구에 대한 고찰, *마케팅연구*, 30(1), 105-144.
- 김종대, 이유석(2019), 토픽 모형 분석을 활용한 시장 세분화 접근법: 영화 시장을 중심으로, *마케팅연구*, 34(4), 53-72.
- 김주영(2018), 신경망(Neural Network)모델들의 비교를 통한 가맹사업 매출예측분석, *마케팅연구*, 33(3), 73-90.

- 오민정(2020), 빅데이터를 통해 살펴본 기부연구, *마케팅연구*, 35(4), 53-74.
- 한상만, 박승배, 정남호(2000), 인공지능경망과 로짓모형을 이용한 내구재의 구매의도 예측에 관한 비교연구, *마케팅연구*, 15(3), 71-92.
- Adamopoulos, Panagiotis, Anindya Ghose, and Vilma Todri (2018), "The Impact of User Personality Traits on Word of Mouth: Text-mining Social Media Platforms," *Information Systems Research*, 29(3), 612-640.
- Ansari, Asim, Yang Li, and Jonathan Z. Zhang(2018), "Probabilistic Topic Model for Hybrid Recommender Systems: A Stochastic Variational Bayesian Approach," *Marketing Science*, 37(6), 987-1008.
- Ascarza, Eva(2018), "Retention Futility: Targeting High-risk Customers Might Be Ineffective," *Journal of Marketing Research*, 55(1), 80-98.
- Athey, Susan, and Guido W. Imbens(2017), "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives*, 31(2), 3-32.
- Berman, Ron, Shiri Melumad, Colman Humphrey, and Robert Meyer(2019), "A Tale of Two Twitterspheres: Political Microblogging During and After the 2016 Primary and Presidential Debates," *Journal of Marketing Research*, 56(6), 895-917.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan(2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Chae, Inyoung, Andrew T. Stephen, Yakov Bart, Dai Yao (2017), "Spillover Effects in Seeded Word-of-mouth Marketing Campaigns," *Marketing Science*, 36(1), 89-104.
- Chen, Yixing, Ju-Yeon Lee, Shrihari(Hari) Sridhar, Vikas Mittal, Katharine McCallister, and Amit G. Singal(2020), "Improving Cancer Outreach Effectiveness Through Targeting and Economic Assessments: Insights from a Randomized Field Experiment," *Journal of Marketing*, 84(3), 1-27.
- Clarke, Jonathan, Hailiang Chen, Ding Du, Yu Jeffrey Hu (2020), "Fake News, Investor Attention, and Market Reaction," *Information Systems Research*, forthcoming. DOI: 10.1287/isre.2019.0910.
- Culotta, Aron, and Jennifer Cutler(2016), "Mining Brand Perceptions from Twitter Social Networks," *Marketing Science*, 35(3), 343-362.
- Danaher, Peter J., Tracey S. Danaher, Michael Stanley Smith, and Ruben Loaiza-Maya(2020), "Advertising effectiveness for multiple retailer-brands in a multimedia and multichannel environment," *Journal of Marketing Research*, 57(3), 445-467.
- Daugherty, Terry, Matthew S. Eastin, and Laura Bright(2008), "Exploring Consumer Motivations for Creating User-generated Content," *Journal of Interactive Advertising*, 8(2), 16-25.
- Dew, Ryan, Asim Ansari, and Yang Li(2020), "Modeling Dynamic Heterogeneity Using Gaussian Processes," *Journal of Marketing Research*, 57(1), 55-77.
- Dzyabura, Daria, and John R. Hauser(2019). "Recommending Products When Consumers Learn Their Preference Weights," *Marketing Science*, 38(3), 417-441.
- Dzyabura, Daria, Srikanth Jagabathula, and Eitan Muller (2019), "Accounting for Discrepancies between Online and Offline Product Evaluations," *Marketing Science*, 38(1), 88-106.
- Edlich, Alex, Greg Phalin, Rahil Jogani, and Sanjay Kaniyar (2019), "Driving Impact at Scale from Automation and AI," (accessed January 17, 2021), <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/driving-impact-at-scale-from-automation-and-ai>
- Fong, Nathan, Yuchi Zhang, Xueming Luo, and Xiaoyi Wang (2019), "Targeted Promotions on an E-book Platform: Crowding Out, Heterogeneity, and Opportunity Costs," *Journal of Marketing Research*, 56(2), 310-323.
- Gabel, Sebastian, Daniel Guhl, and Daniel Klapper(2019), "P2V-MAP: Mapping Market Structures for Large Retail Assortments," *Journal of Marketing Research*, 56(4), 557-580.
- Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li(2012),

- “Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-generated and Crowdsourced Content,” *Marketing Science*, 31(3), 493-520.
- Ghose, Anindya, Beibei Li, and Siyuan Liu(2019), “Mobile Targeting Using Customer Trajectory Patterns,” *Management Science*, 65(11), 5027-5049.
- Guo, Tong, S. Sriram, and Puneet Manchanda(2021), “The Effect of Information Disclosure on Industry Payments to Physicians,” *Journal of Marketing Research*, 58(1), 115-140.
- Halevy, Alon, Peter Norvig, and Fernando Pereira(2009), “The Unreasonable Effectiveness of Data,” *IEEE Intelligent Systems*, 24(2), 8-12.
- Hansen, Karsten T., Kanishka Misra, and Mallesh M. Pai (2021), “Frontiers: Algorithmic Collusion: Supra-competitive Prices via Independent Algorithms,” *Marketing Science*, 40 (1), 1-12.
- Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh (2006), “A Fast learning Algorithm for Deep Belief Nets,” *Neural Computation*, 18(7), 1527-1554.
- Hinton, Geoffrey E., and Ruslan R. Salakhutdinov(2006), “Reducing the Dimensionality of Data with Neural Networks,” *Science*, 313(5786), 504-507.
- Hollenbeck, Brett(2018), “Online Reputation Mechanisms and the Decreasing Value of Chain Affiliation,” *Journal of Marketing Research*, 55(5), 636-654.
- Homburg, Christian, Laura Ehm, and Martin Artz(2015), “Measuring and Managing Consumer Sentiment in an Online Community Environment,” *Journal of Marketing Research*, 52(5), 629-641.
- Homburg, Christian, Marcus Theel, and Sebastian Hohenberg (2020), “Marketing Excellence: Nature, Measurement, and Investor Valuations,” *Journal of Marketing*, 84(4), 1-22.
- Hu, Mantian, Chu Dang, and Pradeep K. Chintagunta(2019), “Search and Learning at a Daily Deals Website,” *Marketing Science*, 38(4), 609-642.
- Huang, Dongling, and Lan Luo(2016), “Consumer Preference Elicitation of Complex Products Using Fuzzy Support Vector Machine Active Learning,” *Marketing Science*, 35(3), 445-464.
- Jalali, Nima Y., and Purushottam Papatla(2019), “Composing Tweets to Increase Retweets,” *International Journal of Research in Marketing*, 36(4), 647-668.
- Jindal, Niket(2020), “The Impact of Advertising and R&D on Bankruptcy Survival: A Double-Edged Sword,” *Journal of Marketing*, 84(5), 22-40.
- Jordan, Michael I., and Tom M. Mitchell(2015), “Machine Learning: Trends, Perspectives, and Prospects,” *Science*, 349(6245), 255-260.
- Kim, Minkyung, K. Sudhir, and Kosuke Uetake(2019), “A Structural Model of a Multitasking Salesforce: Multidimensional Incentives and Plan Design,” *Management Science, Forthcoming*.
- Klostermann, Jan, Anja Plumeyer, Daniel Böger, and Reinhold Decker(2018), “Extracting Brand Information from Social Networks: Integrating Image, Text, and Social Tagging Data,” *International Journal of Research in Marketing*, 35(4), 538-556.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2017). “ImageNet Classification with Deep Convolutional Neural Networks,” *Communications of the ACM*, 60(6), 84-90.
- Kumar, V., Divya Ramachandran, and Binay Kumar(2020), “Influence of New-age Technologies on Marketing: A Research Agenda,” *Journal of Business Research*, 125, 864-877.
- Lemmens, Aurélie, and Sunil Gupta(2020), “Managing Churn to Maximize Profits,” *Marketing Science*, 39(5), 956-973.
- Li, Xi, Mengze Shi, and Xin Shane Wang(2019), “Video Mining: Measuring Visual Information Using Automatic Methods,” *International Journal of Research in Marketing*, 36(2), 216-231.
- Li, Yiyi, and Ying Xie(2020), “Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement,” *Journal of Marketing Research*, 57(1), 1-19.
- Liu, Jia, and Olivier Toubia(2018), “A Semantic Approach for Estimating Consumer Content Preferences from Online

- Search Queries,” *Marketing Science*, 37(6), 930-952.
- Liu, Liu, Daria Dzyabura, and Natalie Mizik(2020), “Visual Listening In: Extracting Brand Image Portrayed on Social Media,” *Marketing Science*, 39(4), 669-686.
- Liu, Xiao, Alan Montgomery, and Kannan Srinivasan(2018), “Analyzing Bank Overdraft Fees with Big Data,” *Marketing Science*, 37(6), 855-882.
- Liu, Xiao, Dokyun Lee, and Kannan Srinivasan(2019), “Large-scale Cross-category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning,” *Journal of Marketing Research*, 56(6), 918-943.
- Liu, Xiao, Param Vir Singh, and Kannan Srinivasan(2016), “A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing,” *Marketing Science*, 35(3), 363-388.
- Lu, Shasha, Li Xiao, and Min Ding(2016), “A Video-based Automated Recommender (VAR) System for Garments,” *Marketing Science*, 35(3), 484-510.
- Luo, Xueming, Xianghua Lu, and Jing Li(2019), “When and How to Leverage E-commerce Cart Targeting: The Relative and Moderated Effects of Scarcity and Price Incentives with a Two-stage Field Experiment and Causal Forest Optimization,” *Information Systems Research*, 30(4), 1203-1227.
- Marchand, André, Thorsten Hennig-Thurau, and Caroline Wiertz (2017), “Not All Digital Word of Mouth Is Created Equal: Understanding the Respective Impact of Consumer Reviews and Microblogs on New Product Success,” *International Journal of Research in Marketing*, 34(2), 336-354.
- Marr, Bernard(2018), “27 Incredible Examples of AI and Machine Learning in Practice,” (accessed January 17, 2021), <https://www.forbes.com/sites/bernardmarr/2018/04/30/27-incredible-examples-of-ai-and-machine-learning-in-practice/?sh=56eb44007502>
- Mejia, Jorge, Shawn Mankad, and Anandasivam Gopal(2019), “A for Effort? Using the Crowd to Identify Moral Hazard in New York City Restaurant Hygiene Inspections,” *Information Systems Research*, 30(4), 1363-1386.
- Miklosik, Andrej, and Nina Evans(2020), “Impact of Big Data and Machine Learning on Digital Transformation in Marketing: A Literature Review,” *IEEE Access*, 8, 101284-101292.
- Mitchell, T. M.(1997). *Machine Learning*, New York: McGraw-Hill Higher Education, 2-5.
- Moon, Sangkil, Moon-Yong Kim, and Dawn Iacobucci(2020), “Content Analysis of Fake Consumer Reviews by Survey-based Text Categorization,” *International Journal of Research in Marketing*, forthcoming. DOI: 10.1016/j.ijresmar.2020.08.001.
- Mullainathan, Sendhil, and Jann Spiess(2017), “Machine Learning: An Applied Econometric Approach,” *Journal of Economic Perspectives*, 31(2), 87-106.
- Murphy, Kevin P.(2012), *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Mustak, Mekhail, Joni Salminen, LoicPlé and JochenWirtzd (2021). “Artificial Intelligence in Marketing: Topic Modeling, Scientometric Analysis, and Research Agenda,” *Journal of Business Research*, 124, 389-404.
- Netzer, Oded, Alain Lemaire, and Michal Herzenstein(2019), “When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications,” *Journal of Marketing Research*, 56(6), 960-980.
- Obermeyer, Ziad, and Ezekiel J. Emanuel(2016), “Predicting the Future—Big Data, Machine Learning, and Clinical Medicine,” *The New England Journal of Medicine*, 375 (13), 1216.
- Peng, Ling, Geng Cui, Yuho Chung, and Wanyi Zheng(2020), “The Faces of Success: Beauty and Ugliness Premiums in e-Commerce Platforms,” *Journal of Marketing*, 84(4), 67-85.
- Pfeiffer, Jella, Thies Pfeiffer, Martin Meißner, and Elisa Weiß (2020), “Eye-tracking-based Classification of Information Search Behavior Using Machine Learning: Evidence from Experiments in Physical Shops and Virtual Reality Shopping Environments,” *Information Systems Research*, 31(3), 675-691.
- Rafieian, Omid, and Hema Yoganarasimhan(2020), “Targeting and Privacy in Mobile Advertising,” *Marketing Science*, Forthcoming.

- Ringel, Daniel M., and Bernd Skiera(2016), “Visualizing Asymmetric Competition among More Than 1,000 Products Using Big Search Data,” *Marketing Science*, 35(3), 511-534.
- Rutz, Oliver J., Garrett P. Sonnier, and Michael Trusov(2017), “A New Method to Aid Copy Testing of Paid Search Text Advertisements,” *Journal of Marketing Research*, 54(6), 885-900.
- Samuel, Arthur L.(1959), “Some Studies in Machine Learning Using the Game of Checkers,” *IBM Journal of Research and Development*, 3(3), 210-229.
- Schneider, Matthew J., Sharan Jagpal, Sachin Gupta, Shaobo Li, Yan Yu(2017), “Protecting Customer Privacy When Marketing with Second-party Data,” *International Journal of Research in Marketing*, 34(3), 593-603.
- Schwartz, Eric M., Eric T. Bradlow, and Peter S. Fader (2014), “Model Selection Using Database Characteristics: Developing a Classification Tree for Longitudinal Incidence Data,” *Marketing Science*, 33(2), 188-205.
- Sheth, Jagdish, and Charles H. Kellstadt(2020), “Next Frontiers of Research in Data Driven Marketing: Will Techniques Keep Up with Data Tsunami?,” *Journal of Business Research*, forthcoming. DOI: 10.1016/j.jbusres.2020.04.050.
- Simester, Duncan, Artem Timoshenko, and Spyros I. Zoumpoulis (2020), “Targeting Prospective Customers: Robustness of Machine-learning Methods to Typical Data Challenges,” *Management Science*, 66(6), 2495-2522.
- Simonyan, Karen, and Andrew Zisserman(2014), “Very Deep Convolutional Networks for Large-scale Image Recognition,” *arXiv preprint arXiv:1409.1556*.
- Singh, Sunil K., Detelina Marinova, and Jagdip Singh(2020), “Business-to-Business E-Negotiations and Influence Tactics,” *Journal of Marketing*, 84(2), 47-68.
- Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov(2014), “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *The Journal of Machine Learning Research*, 15(1), 1929-1958.
- Timoshenko, Artem, and John R. Hauser(2019), “Identifying Customer Needs from User-generated Content,” *Marketing Science*, 38(1), 1-20.
- Tirunillai, Seshadri, and Gerard J. Tellis(2012), “Does Chatter Really Matter? Dynamics of User-generated Content and Stock Performance,” *Marketing Science*, 31(2), 198-215.
- Tirunillai, Seshadri, and Gerard J. Tellis(2017), “Does Offline TV Advertising Affect Online Chatter? Quasi-experimental Analysis Using Synthetic Control,” *Marketing Science*, 36(6), 862-878.
- Toubia, Olivier, Garud Iyengar, Renee Bunnell, and Alain Lemaire(2019), “Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption,” *Journal of Marketing Research*, 56(1), 18-36.
- Trusov, Michael, Liye Ma, and Zainab Jamal(2016), “Crumbs of the Cookie: User Profiling in Customer-base Analysis and Behavioral Targeting,” *Marketing Science*, 35(3), 405-426.
- Vermeer, Susan A.M., Theo Araujo, Stefan F. Bernritter, and Guda van Noort(2019), “Seeing the Wood for the Trees: How Machine Learning Can Help Firms in Identifying Relevant Electronic Word-of-mouth in Social Media,” *International Journal of Research in Marketing*, 36(3), 492-508.
- Wang, Yang, and Alexander Chaudhry(2018), “When and How Managers' Responses to Online Reviews Affect Subsequent Reviews,” *Journal of Marketing Research*, 55(2), 163-177.
- Xia, Feihong, Rabikar Chatterjee, and Jerrold H. May(2019), “Using Conditional Restricted Boltzmann Machines to Model Complex Consumer Shopping Patterns,” *Marketing Science*, 38(4), 711-727.
- Xiao, Li, and Min Ding(2014), “Just the Faces: Exploring the Effects of Facial Features in Print Advertising,” *Marketing Science*, 33(3), 338-352.
- Yoganarasimhan, Hema(2020), “Search Personalization Using Machine Learning,” *Management Science*, 66(3), 1045-1070.



Zhong, Ning, and David A. Schweidel(2020), “Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model,” *Marketing Science*, 39(4), 827-846.

Zhou, Yinghui, Shasha Lu, and Min Ding(2020), “Contour-as-Face Framework: A Method to Preserve Privacy and Perception,” *Journal of Marketing Research*, 57(4), 617-639.