

토픽 모델링에 기반한 온라인 상품 평점 예측을 위한 온라인 사용 후기 분석*

박상현** · 문현실** · 김재경***

Online Reviews Analysis for Prediction of Product Ratings based on Topic Modeling*

Sang Hyun Park** · Hyun Sil Moon** · Jae Kyeong Kim***

■ Abstract ■

Customers have been affected by others' opinions when they make a purchase. Thanks to the development of technologies, people are sharing their experiences such as reviews or ratings through online or social network services. However, although ratings are intuitive information for others, many reviews include only texts without ratings. Also, because of huge amount of reviews, customers and companies can't read all of them so they are hard to evaluate to a product without ratings. Therefore, in this study, we propose a methodology to predict ratings based on reviews for a product. In a methodology, we first estimate the topic-review matrix using the Latent Dirichlet Allocation technic which is widely used in topic modeling. Next, we predict ratings based on the topic-review matrix using the artificial neural network model which is based on the backpropagation algorithm. Through experiments with actual reviews, we find that our methodology can predict ratings based on customers' reviews. And our methodology performs better with reviews which include certain opinions. As a result, our study can be used for customers and companies that want to know exactly a product with ratings. Moreover, we hope that our study leads to the implementation of future studies that combine machine learning and topic modeling.

Keyword : LDA, Artificial Neural Network, Topic Modeling, Rating Prediction, Review Analysis

1. 서론

기술과 인터넷의 발전으로 상품 및 서비스를 온라인에서 구매하는 온라인 시장의 규모가 점차 확대되고 있다. 통계청에 따르면 2016년 온라인 쇼핑 거래액은 약 65조 617억의 규모를 보이고 있으며 2010년부터 매년 평균 10% 이상의 성장률을 기록하고 있다(Statistic Korea, 2016). 이러한 온라인 시장의 확대와 더불어 스마트폰 등의 보급으로 상품에 대한 평가 등의 소비자의 참여가 활발해짐에 따라 최근 소비자들은 기업의 일방적인 상품 소개 정보뿐만 아니라 다른 사용자의 사용 후기나 평점, 전문가의 추천에 영향을 받고 있다(Chatterjee, 2001).

소비자들은 상품의 가격, 품질 등의 속성을 나타내는 내부적 요인뿐만 아니라 외부적 요인에 의해 구매 의사결정을 한다(Tsang and Prendergast, 2009). 여러 외부적 요인 중 다른 사람의 평가 또는 의견이 소비자의 구매 의사결정에 영향을 미치는 것을 구전효과(Word Of Mouth, WOM)라 한다. 구전(WOM)은 브랜드, 물품, 서비스 또는 서비스 제공자를 고려할 때 상업적이지 않은 사람이라고 인식되는 사람과 정보를 받는 사람 간의 커뮤니케이션이라고 정의된다(Chatterjee, 2001). 특히 온라인 상에서 이루어지는 구전은 eWOM(electronic Word Of Mouth)으로 정의되고 있으며 이러한 eWOM이 소비자의 구매 의사결정에 미치는 영향 관계에 대한 연구가 다양한 분야에서 진행되어 오고 있다. 예를 들어, Chevalier and Mayzlin(2006)에 따르면 도서 시장에서 사용자의 사용 후기가 상품 판매량에 상당한 영향을 끼치며, Cui et al.(2012)의 연구에서는 새로운 상품의 판매에서 초기의 평점이 판매에 영향을 끼친다고 밝혔다. 또한 Lee et al.(2008)은 인터넷 쇼핑에서 상품의 사용 후기는 다른 소비자들 이 구매를 결정하는 데 영향을 미친다는 연구를 진행하였다. 특히 Park et al.(2007)의 연구에서는 사용자의 리뷰와 함께 평점에 대한 정보가 함께 제공될 때 판매에 미치는 영향이 더 커

진다는 점을 밝혔다.

기존 연구에서는 사용 후기, 평점 등 사용자의 상품에 대한 의견과 판매의 영향 관계를 밝히고 있으며 사용 후기와 평점이 함께 보여질 때 그 영향 관계가 커진다는 것을 밝히고 있다. 특히 사용 후기는 텍스트로 이루어져 사용 후기를 대하는 사람의 태도에 따라 그 영향관계가 달라질 수 있지만 평점의 경우 해당 상품에 대한 수치적 평가로 객관성이 유지된다(Tsang and Prendergast, 2009). 하지만 온라인 상의 상품에 대한 평가는 수치로 되어 있는 평점을 수반하지 않는 경우도 많다(Pang and Lee, 2005). 평점을 수반하지 않은 상품 후기가 많은 경우의 상품을 판매하는 기업 입장에서는 소비자의 객관적인 상품 평가 정보를 얻을 수 없으며 소비자 입장에서도 직접 모든 상품 후기를 읽고 의사결정을 내리는 것이 어려워진다.

따라서 이러한 한계를 극복하기 위하여 본 연구에서는 사용 후기의 토픽을 분석하고 이를 기반으로 사용 후기의 평점을 예측하는 모델링 기법을 제안한다. 온라인에서 사용 후기는 사용자가 해당 상품에 대해 평가하여 자신의 선호를 나타내는 도구로 사용되지만 사용 후기에 나타나는 사용자의 선호를 파악하기에는 어려움이 많다. 따라서 본 연구에서는 먼저 사용 후기의 특성을 분석하기 위하여 사용 후기를 대표할 수 있는 주제를 추출한다. 그리고 이 주제를 통해 사용자의 평점을 예측하는 모델을 제안하여 온라인 쇼핑물 등에 작성되어 있는 사용 후기를 기반으로 한 상품 평점을 예측하려 한다.

사용 후기의 평점 예측 모델링을 위한 기초 데이터로는 구매자만이 사용 후기를 작성할 수 있는 Amazon.com의 사용 후기 및 평점 데이터를 사용하였다. 총 24개의 카테고리 중 5개의 카테고리를 선정하여 각 카테고리별로 평점당 500개의 사용 후기를 랜덤샘플링한 뒤 LDA(Latent Dirichlet Allocation) 기법을 적용하여 토픽이 각 사용 후기를 설명하는 가능성을 측정하였으며 이를 입력 변수로, 평점을 목표 변수로 하는 인공신경망 모형을 구축하였다.

2. 관련 연구

2.1 토픽 모델링(Topic Modeling)

스마트 폰 등의 기술과 SNS(Social Network Service)의 발전으로 사람들은 시간과 장소에 구애 없이 서로 소통하고 정보를 공유할 수 있게 되었다. 하지만 소통과 공유로 생성되는 데이터는 대부분 비정형 데이터인 텍스트, 이미지 등으로 기존의 정형 데이터 분석 기법으로는 분석이 어려운 속성을 지니고 있다. 특히 텍스트는 사용자의 직접적인 감정 표현의 수단으로 지속적으로 SNS, 블로그, 온라인 쇼핑몰 등에서 생성되고 있지만 그 목적과 숨겨진 의미를 파악하기는 어려운 실정이다.

이러한 방대한 문서군에서 각 문서가 다루는 주제가 비슷한 문서들을 자동으로 군집화하여 주제를 파악하기 위한 시도가 토픽 모델링 기법이다(Beli and Lafferty, 2009). 토픽 모델링은 구조화되지 않은 문서 집단에서 토픽을 찾아내기 위한 알고리즘으로 글의 맥락과 관련된 단서들을 이용하여 유사한 의미를 가진 단어들을 클러스터링하는 방식으로 주제를 추론하는 기법이다(Blei and Lafferty, 2012; Kang et al., 2013). 토픽 모델링에 기반한 연구는 다양한 분야에서 이루어져 왔다. 예를 들어 Lu and Zhai(2008)는 블로그나 포럼과 같은 다양한 정보 원천에서 전문가들에 의하여 작성된 수많은 의견을 토픽 모델링 기법에 기반하여 자동적으로 모으는 방법을 제안하였다. 상품과 정치 분야에서 진행된 실험을 통해 연구에서 제시한 방법이 두 분야 모두에서 유용하다는 결론을 내렸다. 또한, Hong and Davison(2010)의 연구에서는 SNS를 다양한 분야의 사람들이 모이는 정보의 원천으로 여겨 SNS에 업로드되는 내용 및 맥락을 파악하는 것이 뉴스 특보의 발견, 친구 추천 및 감성 분석 등의 분야에 큰 도움이 될 것으로 판단하였다. 그에 따라 연구에서 수집한 데이터를 기반으로 토픽 모델링을 진행하였으며 실험 결과 연구에서 제안한 방법이 기존의 분류 모델에 비하여

높은 성과를 보였다.

이와 같이 토픽 모델링은 다양한 분야에서 활용되고 있으며 가장 대표적인 기법은 LDA(Latent Dirichlet Allocation) 기법으로 주어진 문서가 가지고 있는 잠재적인 주제들과 한 단어가 주제에 포함될 가능성을 디리클레 분포를 이용하여 추론하는 기법이다. LDA의 기본적인 아이디어는 여러 가지 토픽들로부터 문서를 모델화하는 것으로 토픽이 주어진 단어들의 분포로 이루어져 있다는 것이다(Beli and Lafferty, 2009). 또한 LDA는 정형화되어 있지 않은 문서들의 사이에서 규칙을 찾아내는 데 문서에서 나타나는 주제들은 서로 다른 비율로 문서에 포함되어 있음을 이용한다. LDA의 전체 프로세스를 요약하면 디리클레 분포를 따르는 사후 확률(posterior probability)을 이용하여 실제 존재하는 단어와 숨겨진 변수들 간의 구조를 파악하는 것이다. 여기에서 실제 존재하는 단어는 문서군에서 실제로 나타나는 단어를 의미하며 숨겨진 변수들은 토픽이 해당 문서에 속해 있는 정도를 나타내는 변수를 의미한다.

Chae et al.(2015)은 LDA 기법을 이용하여 소셜커머스 3개 업체와 오픈마켓 3개 업체의 어플리케이션 사용 후기를 분석하여 그 차이를 비교하였다. 구체적으로 소셜커머스와 오픈마켓 이용자들의 이용 경험을 비교하고 분석한 탐색적 연구로 1만여 개의 사용 후기를 유용성과 편리성 토픽으로 분류한 뒤 감성 분석과 동시출현 단어 분석을 수행하였다. 연구 결과 소셜커머스는 '서비스' 토픽을 다룬 단어들이 이용자들의 긍정적인 이용 경험을 이끌어 내고 있는 반면, 오픈마켓의 경우에는 기술적 문제 및 불편과 관련된 단어들이 부정적 평가와 연관되는 것을 발견하였다. 이를 통해 연구자들은 LDA 기법과 기술수용모델(Technology Acceptance Model, TAM)을 결합하여 사용자의 이용경험을 나타내는 텍스트를 효과적으로 분석할 수 있는 새로운 방법을 제시하였다. 또한 Bisgin et al.(2011)은 미국 FDA에서 승인한 약품의 성능표에 의미가 모호한 표현이 많아 이들을 분류하는 데 전문가의

노력이 필요함에 따라 LDA 기법에 기반한 약품 분류 모델을 제시하였다. 실험 결과 안전과 치료 목적으로 분류된 약품들은 통계적으로 유의한 결과를 보였으며 이를 기반으로 생물 의학 문서에서 안전과 치료의 숨겨진 관계를 발견하였다.

2.2 인공 신경망(Artificial Neural Network)

인공 신경망의 기본 개념은 1940년대에 처음 제안되었으나 1980년대 중반부터 역전파법이 제안되면서 많은 관심을 받고 있다(Sim and Kim, 2008). 인공신경망은 인간 두뇌가 정보를 처리하는 프로세스를 모사한 것으로 두뇌의 가장 기본 단위인 신경세포와 뉴런 간의 연산 과정을 모델화한 것이다. 인공신경망의 학습, 즉 가중치와 편향 등의 파라미터의 자동적인 조정을 위해 대표적으로 사용되는 기법은 역전파 알고리즘이다. 역전파 알고리즘은 인공신경망의 결과와 실제값의 차이를 나타내는 손실함수를 최소화하는 방향으로 파라미터를 조정한다(Kang and Lee, 2015).

인공신경망은 다양한 분야에서 활용되고 있으며 특히 비즈니스 분야에서는 주식 가격 예측, 경영 예측 등의 사례에 사용되고 있다. 인공신경망이 가지는 특징은 먼저 모델 자체가 데이터에 의하여 내부의 파라미터가 자동적으로 조절되는 데이터 기반의 자가 조정 모델이라는 점이다. 또한 인공신경망의 내부는 현실세계에서 일어나는 복잡한 문제를 해결할 수 있는 비선형 모델로 구성되어 있다. Huang et al.(2004)의 연구에서는 신용 등급을 예측하는 데 있어 선형적인 관계를 가지고 있지 않은 14개의 변수를 사용하여 신용 등급을 비교적 정확히 예측하였다. 즉, 인공신경망은 비선형 관계를 가지는 변수로부터 문제를 해결하는 데 뛰어난 결과를 보이고 있다(Tam and Kiang, 1992). 또한, Hwang(2010)은 축적된 데이터와 군집화 기법을 접목한 인공신경망을 이용하여 웹 서비스 분류를 할 수 있는 방법론을 제안하였다.

이와 같은 특징에 따라 인공신경망은 변화하는

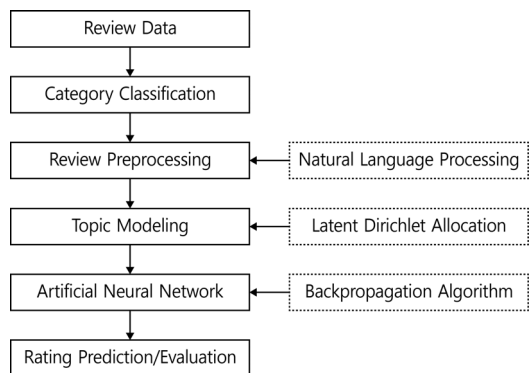
환경에 쉽게 적응이 가능하며, 샘플의 크기, 변수의 수와 데이터의 분포에 덜 민감하고, 잡음(noise) 데이터를 포함하고 있다 하더라도 그 기능을 제대로 발휘한다(Sun et al., 1997). 이러한 장점에 따라 인공 신경망은 다른 기계 학습 기법에 비하여 높은 정확도를 나타낸다. 예를 들어 Jo et al.(1997)은 사례기반 추론과 인공신경망, 판별 분석을 이용하여 은행의 부도를 예측하는 연구를 수행한 결과 인공신경망이 가장 높은 성과를 나타냄을 밝혀냈다. 특히 최근에는 컴퓨터 성능 등의 발달로 인공신경망의 층을 깊게 할 때 발생하는 계산의 복잡성 등의 문제가 일부 해결되어 심층 신경망에 기반한 이미지 분류 등 다양한 분야에서 인공신경망 연구가 활발하게 진행되고 있다(Krizhevsky, 2012).

따라서, 본 연구에서는 토픽 모델링의 한 기법인 LDA 기법을 이용하여 각 사용 후기의 단어가 각 토픽에 속할 가능성을 입력변수로 하고 해당 사용 후기의 평점을 목표변수로 하는 인공신경망 모형을 사용하여 알려져 있지 않은 사용자의 평점을 예측하고자 한다.

3. 온라인 평점 예측 모델

3.1 개요

본 연구에서 제안하는 온라인 평점 예측 모델의 분석 프로세스는 <Figure 1>과 같다.



<Figure 1> Overall Procedure

먼저, 사용 후기 데이터를 각 카테고리별로 분류한다. 본 연구에서는 Amazon.com의 사용 후기 및 평점이 연구대상임에 따라 Amazon.com의 카테고리 분류 중 유아용품, 식품, 건강 및 개인용품, 음악, 스포츠 및 아웃도어 용품 등 5개의 카테고리를 선정하여 연구를 진행하였다. 다음으로 각 카테고리별 사용 후기 원천 데이터를 기반으로 토픽 모델링을 진행하기 위하여 자연어 처리 알고리즘을 사용하여 전처리 과정을 수행하였다. 이렇게 전 처리된 사용 후기를 기반으로 LDA 모델 기반의 토픽 모델링을 수행하여 각 사용 후기의 토픽 가능성을 측정 한 뒤 이를 입력변수로 하고 해당 사용 후기의 평점을 목표 변수로 하는 인공신경망 모형을 역전파 알고리즘을 통해 구축하였다. 마지막으로 인공 신경망의 평점 예측 결과와 실제 값의 차이를 기반으로 분석 모형의 성과를 측정하였다.

3.2 사용 후기 데이터 및 전처리

본 연구에서는 기존의 텍스트 마이닝 관련 연구에서 사용된 Amazon.com의 사용 후기 데이터를 사용하였다. 1996년 5월부터 2014년 7월까지 총 24개 카테고리의 상품에 대한 약 1.5억 개의 사용 후기를 포함하고 있는 데이터로 사용 후기 뿐만 아니라 해당 상품에 대한 메타데이터를 포함하고 있다. <Table 1>은 데이터 예시이다.

<Table 1>에서 ReviewID는 사용 후기를 남긴 사용자의 고유 ID를 의미하며 ASIN 코드는 사용 후기의 대상이 되는 상품의 고유 ID이다. 또한 Helpful은 해당 사용 후기가 얼마만큼 도움이 되었는가를 나타내는 척도로 왼쪽의 숫자는 도움이 되었다고 평가한 사람의 수, 오른쪽은 평가를 한 전체 사람의 수이다. 즉, 예시에서는 3명의 사용자 중 2명의 사용자가 해당 사용 후기가 도움이 되었다고 평가를 한 것이다. 본 연구에서 제안하는 분석 모형은 ReviewText를 토픽 모델링한 결과 값을 입력변수로, Overall 변수가 나타내는 평점이 목표 변수이다. 본 연구에서는 전체 카테고리 분석

<Table 1> Review Data Example

Variable	Example
ReviewerID	A2SUAMIJ3GNN3B
ASIN	0000013714
ReviewerName	J. McDonald
Helpful	[2, 3]
ReviewText	I bought this for my husband who plays the piano. He is having a wonderful time playing these old hymns. The music is at times hard to read because we think the book was published for singing from more than playing from. Great purchase though!
Overall	5.0
ReviewTime	09 13, 2009

에 앞서 대표적인 카테고리에 대한 모형 적용 결과를 분석하기 위하여 유아용품(Baby), 식품(Grocery), 건강 및 개인용품(Health), 음악(Music), 스포츠 및 아웃도어 용품(Sports) 등 5개의 카테고리를 선정하였으며 효율적인 주제어 분석 및 인공신경망의 학습을 위하여 Helpful이 50% 이상인 사용 후기 중 500개의 사용 후기를 랜덤 샘플링하여 평점 예측을 위한 원천 데이터로 사용하였다.

다음으로 이러한 원천 데이터를 토픽 모델링에 사용 가능하도록 변환하는 과정을 수행하였다. 이를 전처리 과정이라 하며 텍스트 마이닝에서는 컴퓨터가 텍스트 데이터를 계산하고 처리할 수 있도록 변환하는 과정을 포함한다. 특히 사용 후기는 비속어 및 줄임말 또는 기호 등의 사용이 많아 적합한 전처리 과정을 수행하지 않으면 분석 결과에 영향을 줄 수 밖에 없다. 따라서 본 연구에서는 먼저 사용 후기에서 단어를 추출한 뒤 불용어를 제거하는 작업을 수행하였다. 불용어는 I, my 등 문서에서 자주 등장하기는 하지만 정보 검색 또는 분석에 영향을 주지 않는 의미 없는 단어를 의미한다(Braun et al., 2005). 이러한 불용어는 분석에 의미가 없음에도 상당한 비율이 존재하여 문서의 주제를 파악하는 데 큰 영향을 미침에 따라 전처리 과정에서 제거한다. 마지막으로 공통 어간을

가지는 단어를 묶기 위해 스템밍(stemming)을 수행하였다. 스템밍은 공통 어간을 가져 같은 의미를 지님에도 다르게 쓰이는 단어들을 하나의 단어로 묶는 과정이다. 예를 들어, bought와 buy는 같은 뜻을 지닌 단어지만 다른 텍스트로 인식되는데 이를 buy 또는 bought로 바꿔 하나의 단어로 묶는 것이다.

3.3 토픽 모델링

전처리가 완료된 사용 후기 데이터를 기반으로 토픽 모델링을 수행하기 위하여 LDA 기법을 적용하였다. 이를 위하여 각 카테고리별로 2,500개의 사용 후기를 무작위 샘플링하여 선정하였다. 또한 데이터의 편향을 방지하기 위하여 5점 척도인 각 평점별로 500개의 사용 후기가 포함되도록 선정하였다.

LDA의 적용을 위해 먼저 Dirichlet 분포(θ)와 편향값(β)를 추정해야한다. 이를 위하여 Minka and Lafferty(2002)가 제안한 기대 전도(Expectation propagation) 기법을 사용하였다. 또한 LDA 기법은 문서에 포함되는 단어들이 토픽 분포에 의하여 생성된 k개의 토픽을 선택하면서 생성이 됨에 따라 적절한 k의 선정, 즉 토픽 수의 선정이 필요하다. 기존 연구에서는 여러 번의 실험을 통하여 가장 적합한 토픽의 수를 선정함에 따라 본 연구에서도 실험을 반복적으로 수행하여 각 카테고리별로 적합한 토픽의 수를 선정하였다(Song et al., 2009).

LDA를 사용 후기에 적용하면 각 토픽당 사용 후기를 설명할 수 있는 비율인 토픽-사용 후기 매트릭스와 각 토픽에 속해 있는 단어들을 파악할 수 있어 이를 기반으로 주제를 파악할 수 있다. 예를 들어, 토픽 1에 포함되는 단어가 album, one, best, much, good 등이라면 토픽 1은 앨범이라는 주제를 나타낸다고 파악할 수 있다. 또한 토픽-사용 후기 매트릭스는 각 사용 후기가 개별 토픽에 속할 확률을 나타낸다. <Table 2>는 토픽-사용 후기 매트릭스의 예시이다.

<Table 2> Topic-Review Matrix

	Topic-1	Topic-2	...	Topic-k
Review-1	0.0873	0.0396		0.07143
Review-2	0.0538	0.0538		0.03846
...				
Review-n	0.03763	0.04838		0.37634

<Table 2>에 나타난 각 사용 후기의 토픽 벡터의 합은 1임에 따라 각 값은 해당 사용 후기가 해당 토픽에 속할 확률로 해석할 수 있으며 그에 따라 가장 높은 확률을 보이는 토픽에 속한다고 할 수 있다. 예를 들어 Review-1을 보면 Topic-1에 속할 확률이 0.0873으로 사용 후기 Review-1은 토픽 Topic-1로 분류가 된다. 본 연구에서는 LDA 기법을 통해 도출된 토픽-사용 후기 매트릭스의 벡터값을 활용하여 해당 리뷰의 평점을 예측하려 한다.

3.4 인공신경망을 활용한 평점 예측

인공신경망은 입력 데이터를 반복적으로 학습하여 데이터에 숨어 있는 패턴을 찾아내는 기법이다. 인공신경망은 일반적으로 입력층, 은닉층, 그리고 출력층으로 구성되며 각 층을 구성하는 뉴런들이 모두 연결되어 있는 완전 연결 모델이다. 본 연구에서는 입력 변수로 토픽-사용 후기 매트릭스를 사용함에 따라 총 k개의 입력 노드를 사용하였다. 또한, 목표 변수가 사용 후기에 대한 평점으로 5점 척도로 측정됨에 따라 출력층에는 5개의 출력 노드를 배치하였다. 다음으로 은닉층의 노드 수는 너무 많은 경우 학습 데이터에만 모형이 적합되는 과적합(overfitting)이 발생되며 너무 적은 경우에는 제대로 된 예측을 할 수 없음을 따라 1개부터 시작해서 늘려나가며 가장 높은 성과를 보이는 노드 수를 선정하였다.

인공신경망의 학습은 인공신경망을 통해 예측된 결과와 실제값의 차이를 최대한 줄이기 위해 인공신경망을 구성하고 있는 계수와 편향을 조절하는 것을 의미한다. 계수와 편향을 조정하기 위해 대

표적으로 사용되는 기법은 역전과 알고리즘(back propagation algorithm)이다(Sim and Kim, 2008; Kang and Lee, 2015). 본 연구에서는 입력층 뉴런에 각 사용 후기에 대한 토픽의 벡터값을 적용시키고 인공신경망을 거쳐 나온 평점 예측값과 실제 평점의 차이를 줄이는 방향으로 가중치와 편향을 조정하는 역전과 알고리즘을 적용하여 모형을 학습하였으며 학습 결과로 구성된 인공신경망을 활용하여 평점을 예측하였다.

3.5 모형의 평가

본 연구에서 제시한 평점 예측 모형은 인공신경망을 활용함에 따라 데이터의 과적합 문제가 발생될 수 있다. 데이터의 과적합은 모형을 구축할 때 사용된 데이터에만 모형이 적합되어 새로운 데이터에는 성과가 낮게 나타나는 현상을 의미한다. 그에 따라 홀드아웃(Hold Out) 방법을 사용하여 모형의 성과를 측정하고자 한다. 홀드아웃은 데이터의 과적합을 막기 위해 사용되는 대표적인 방법으로 전체 데이터를 모형을 구축하기 위한 학습 데이터와 모형의 성과를 측정하기 위한 평가 데이터로 구분하는 방법이다. 일반적으로 학습 데이터와 평가 데이터의 비율을 7:3 또는 8:2로 구분하는데 본 연구에서는 학습 데이터를 80%, 평가 데이터를 20%의 비율로 추출하여 평가에 활용하였다.

인공신경망을 비롯한 기계학습 모형이 분류 문제를 다루는 경우 오분류표를 활용하여 모형의 성과를 측정한다. 본 연구에서 제시한 모형은 평점을 예측하는 모형이지만 평점이 5점 척도로 구성되어 있어 5개의 분류를 예측하는 모형으로 판단할 수 있다. 따라서 본 연구에서 제시한 모형의 성과를 평가하기 위하여 오분류표를 사용한다. 오분류표의 각 값은 <Table 3>과 같다.

분류 모형의 성과 측정을 위해 오분류표를 사용하면 다양한 평가 지표의 측정이 가능하다(Tange et al., 2009; Witten and Frank, 2005). 본 연구에서는 모형의 성과 측정을 위하여 정분류율(Accuracy),

<Table 3> Confusion Matrix

	Actual	
Predict	True	False
True	True Positive (TP)	False Positive (FP)
False	False Negative (FN)	True Negative (TN)

재현율(Recall), 정밀도(Precision), F-1 지표를 사용하였다. 먼저, 정분류율(Accuracy)은 모형이 얼마나 정확하게 분류하였는가를 나타내는 지표로 전체 관측치 중 제대로 분류한 비율로 계산된다.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

다음으로 모형의 완전성(Completeness)을 평가하기 위하여 실제값이 True인 관측치 중 예측값이 적중한 정도를 나타내는 재현율(Recall)을 측정하였다.

$$Recall = \frac{TP}{(TP + FN)}$$

또한 모형이 True로 예측한 관측치 중 실제값이 True인 정도를 나타내는 정밀도(Precision)를 측정하여 모형의 정확성(Exactness)을 측정하였다.

$$Precision = \frac{TP}{(TP + FP)}$$

하지만 재현율과 정밀도는 한 지표의 값이 높아지면 다른 지표의 값이 낮아질 가능성이 높다. 이러한 효과를 보정하기 위하여 두 지표의 조화평균인 F-1 지표도 측정한다.

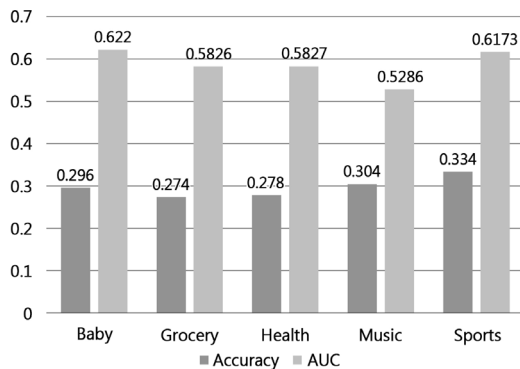
$$F-1 = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

본 연구에서는 분류 모형의 성과를 측정하기 위해 오분류표와 더불어 ROC(Receiver Operating Characteristic) 그래프도 사용하였다. ROC 그래프의 세로축은 민감도, 가로축은 특이도로 구성되

며 무작위 모형(random model)과 비교하여 해당 모형이 얼마나 높은 성과를 나타내는 지 시각적으로 확인할 수 있다. ROC 그래프를 활용한 지표는 AUC(Area Under the ROC curve)가 대표적으로 사용된다(Andrew, 1997). AUC는 ROC 그래프에서 그래프의 아래 면적을 의미함에 따라 기준이 되는 무작위 모형의 AUC는 0.5이다. 따라서 모형의 AUC 값이 0.5보다 크다면 무작위 모형보다 높은 성과를 보인다고 평가하며 1에 가까울수록 높은 성과를 보인다고 해석한다(Beli and Lafferty, 2009).

4. 분석 결과

본 연구에서는 유아용품(Baby), 식품(Grocery), 건강 및 개인용품(Health), 음악(Music), 스포츠 및 아웃도어 용품(Sports) 등 5개의 카테고리별로 모형의 성과를 측정하였다. 먼저, 각 카테고리별로 정분류율과 AUC를 측정한 결과는 <Figure 2>와 같다.

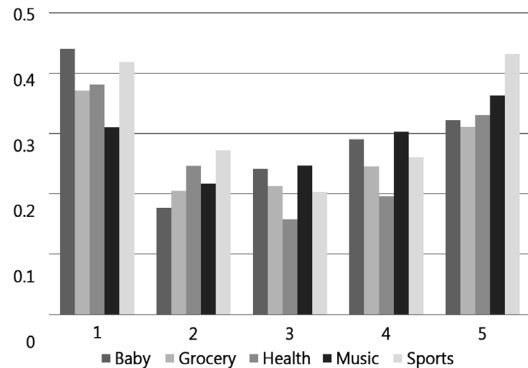


<Figure 2> Comparison of Overall Accuracy and AUC

AUC 관점에서는 모든 카테고리에서 본 연구에서 제시한 평점 예측 모형이 무작위 모형보다 높은 성과를 보인다. 특히, <Figure 2>에 따르면 유아용품의 AUC 값(0.622)이 가장 높은 성과를 보인다. 정분류율(Accuracy) 관점에서는 스포츠 및 아웃도어 용품의 평점 예측 결과가 가장 높은 성과(0.334)를 보이지만 전반적으로 높은 정분류율을 보인다고

판단하기 어렵다. 그 원인을 파악하기 위하여 각 평가 점수별로 F-1 지표를 측정하였다. <Figure 3>은 각 평점에 따라 F-1 지표를 카테고리별로 비교한 성과측정 결과를 나타낸다.

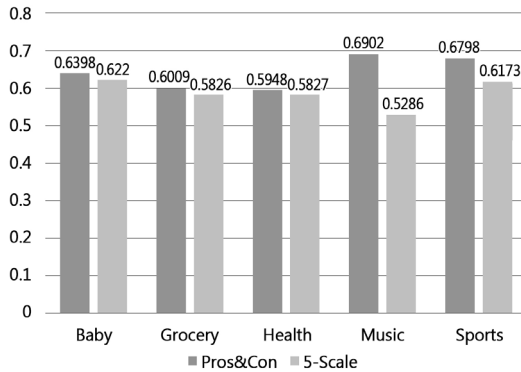
<Figure 3>에 따르면 모든 카테고리에서 1점에 대한 F-1값이 가장 높게 나오고 그 다음으로 5점에 대한 F-1값이 가장 높게 나오는 것을 관찰할 수 있다. 따라서 평점이 극단적인 경우 즉, 선호와 비선호가 확실하게 드러난 경우 사용 후기에 기반한 평점 예측 결과가 더 정확할 수 있다는 결론을 도출하였다.



<Figure 3> F-1 Score Comparison

이러한 결과는 감성 분석(Sentiment Analysis) 연구에서 1점과 2점의 평점을 부정으로 판단하고 4점과 5점을 긍정으로 판단하여 분석을 진행하는 것과 일치한다(Turney, 2002). 즉 사용자의 선호가 불분명한 사용 후기의 경우 평점에 대한 예측이 어렵다는 것을 나타낸다. 그에 따라 본 연구에서도 마찬가지로 1점과 2점의 사용 후기를 긍정으로 4점과 5점의 사용 후기를 부정으로 한 뒤 모형을 적용하여 그 성과를 측정하였다. <Figure 4>는 5점 척도의 분류 모형과 긍정과 부정으로 목표 변수를 수정한 뒤의 AUC 관점에서 비교한 것이다.

<Figure 4>에 따르면 모든 카테고리에서 긍정 및 부정으로 모형을 구축하였을 때 5점 척도를 사용하는 것에 비해 AUC가 증가한다. 따라서 본 연구에서 제시하는 모형은 사용자의 선호가 확실하



<Figure 4> AUC Comparison

게 드러나는 사용 후기를 대상으로 적용하였을 때 더 적합하다는 것을 알 수 있다. 특히 음악 카테고리의 경우 AUC 값이 현저하게 증가하여 음악에서 사용자의 선호가 확실하게 드러나는 사용 후기에 대해 모형이 높은 성과를 나타낸다는 것을 확인할 수 있다.

평가지표에 대한 결과를 종합적으로 살펴보기 위한 각 카테고리별 모형의 성과 측정 결과는 <Table 4>에 제시하였다. 그 결과 본 연구에서 제시한 평점 예측 모형은 평균적으로 0.6411의 AUC 값을 나타내며 모든 카테고리에서 0.5보다 높은 값을 나타내어 무작위 모형보다 우수하다고 판단된다. 특히 <Table 4>에 따르면 스포츠 및 아웃도어 용품의 AUC 값이 가장 높고 F-1 지표 또한 가장 높다. 이는 스포츠 및 아웃도어의 상품들의 사용 후기가 평점을 예측하는 데 좋은 원천 데이터가 됨을 의미하며 또한 다른 카테고리에 비해 많은 정보를 담고 있음을 의미한다.

<Table 4> Results for Evaluation Metrics

Category	Recall	Precision	F-1	AUC
Baby	0.6063	0.6979	0.6489	0.6398
Grocery	0.5654	0.5878	0.6025	0.6009
Health	0.5146	0.6027	0.5552	0.5948
Music	0.7005	0.6798	0.69	0.6902
Sports	0.7647	0.6638	0.7107	0.6798
Average	0.6303	0.6464	0.6415	0.6411

결론적으로 5점 척도에서 불분명한 선호를 나타내는 사용 후기보다 확실한 선호를 드러낸 사용 후기들이 모형의 정확도를 높이는 데 큰 기여를 하였다. 따라서 사용 후기에 기반하여 평점을 예측하는 모형은 그 선호가 확실하게 드러나는 사용 후기를 사용해야 한다고 할 수 있다.

6. 결 론

본 연구에서는 상품 또는 서비스를 구매한 사용자들이 작성한 사용 후기의 평점을 예측하는 모형을 제시하였다. Amazon.com을 대상으로 하여 유아 용품, 식품, 건강 및 개인 용품, 음악, 스포츠 및 아웃도어 카테고리 상품들의 사용 후기 및 평점을 기초 데이터로 사용하였으며 토픽 모델링 알고리즘과 인공신경망 모형을 사용하여 평점 예측 모형을 구축하였다.

실험 결과 각 카테고리별로 성과에 차이는 있으나 그 추세가 비슷함을 확인하였다. 또한 선호가 불분명한 사용 후기로 모형을 학습시키는 것보다 선호가 확실히 드러난 사용 후기로 모형을 학습시키는 것이 더 높은 성과를 보였다. 이러한 결과를 통해 본 연구의 의의는 다음과 같다. 먼저, 온라인 상에 산재해 있는 사용 후기를 기반으로 평점을 예측할 수 있는 모형을 제안하였다. 사용 후기는 텍스트로 이루어져 많은 수의 사용 후기를 검토하고 의견을 파악하는 것은 어려운 일이지만 평점이라는 직관적인 정보가 부여되면 그 해석과 평가가 용이해진다. 하지만 온라인상의 사용 후기에는 평점이 함께 부여되지 않은 경우가 많다. 그에 따라 사용 후기만을 가지고 평점을 예측한다면 소비자 입장에서는 해당 상품에 대한 객관적인 평가 정보를 얻을 수 있으며 기업 입장에서는 본인의 상품에 대한 전반적인 시장의 평가를 알 수 있다.

다음으로 본 연구에서 제시한 방법은 토픽 모델링과 기계학습을 융합한 새로운 분류 방법이다. 기존의 텍스트 마이닝 연구, 특히 감성 분석 분야에서는 이미 매겨진 긍정 또는 부정 점수에 따라

어떤 단어가 많이 쓰이는지를 주로 분석하였다. 따라서 토픽 모델링에 기반하여 평점을 예측하는 방법은 텍스트로 이루어진 사용 후기로부터 기계 학습을 활용하여 소비자의 감성 및 평가를 예측할 수 있는 새로운 시도라 할 수 있다.

하지만 본 연구에서 제안한 방법은 아직 정확도가 높다고 할 수 없다. 그 원인으로 본 연구에서는 선호가 불분명하게 드러난 사용 후기를 지적하였으나 더 높은 정확도를 보이는 모형의 구축을 위해 향후 연구에서는 다음과 같은 부분을 진행할 예정이다.

먼저, LDA 기법에 기반한 토픽 모델링은 사용 후기가 어떤 주제를 나타내는지를 분석하기 위해 사용되는 기법이다. 그에 따라 선호와 관계없는 단어들도 주제어로 선정되는 경우가 존재한다. 이러한 단어는 사용자의 사용 평점에 영향을 주지 않아 평점 예측에 잡음(noise)으로 존재할 가능성이 크다. 따라서, 향후 연구에서는 단어들의 긍·부정을 분석할 수 있는 감성 분석 또는 오피니언 마이닝 등의 기법을 도입하여 정확도를 높이려 한다. 또한, LDA 기법이 토픽 모델링에 대표적으로 사용되는 기법이지만 토픽의 수가 증대됨에 따라 인공신경망에서 고려되어야 하는 변수의 수, 즉 차원이 증가하게 된다. 따라서 데이터 차원 증가에 따른 적절한 인공신경망 모형의 수정이 필요함에 따라 사용 후기 기반의 평점 예측 모형의 파라미터 설정에 관한 연구를 진행할 예정이다. 마지막으로 본 연구에서는 LDA 기법과 역전파 알고리즘 기반의 인공신경망 모형을 사용하였으나 최근의 연구를 반영하여 더 발전된 기법의 탐색을 통해 모형의 성과를 높이는 연구를 진행할 것이다.

References

- Andrew, P.B., "The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms", *Pattern Recognition*, Vol.30, No.7, 1997, 1145-1159.
- Bisgin, H., Z. Liu, H. Fang, X. Xu, and W. Tong, "Mining FDA Drug Labels Using An Un-supervised Learning Technique-Topic Modeling", *BMC Bioinformatics*, Vol.12, No. 10, 2011, 1-8.
- Blei, D.M. and J. Lafferty, *Text Mining : Classification, Clustering, and Applications*, Chapman and Hall/CRC, 2009.
- Braun, L., F. Wiesman, and J. van den Herik, "Towards Automatic Formulation of A Physician's Information Needs", *In Information Retrieval Workshop*, 2005.
- Chae, S.H., J.I. Lim, and J.Y. Kang, "A Comparative Analysis of Social Commerce and Open Market Using User Reviews in Korean Mobile Commerce", *Journal of Intelligence and Information Systems*, Vol.21, No.4, 2015, 53-77.
- (채승훈, 임재익, 강주영, "사용자 리뷰를 통한 소셜 커머스 와 오픈마켓의 이용경험 비교분석", *지능정보연구*, 제21권, 제4호, 2015, 53-77.)
- Chatterjee, P., "Online Reviews : Do Consumers use Them?", *ACR 2001 Proceedings*, 2001, 129-134.
- Chevalier, J.A. and D. Mayzlin, "The Effect of Word of Mouth on Sales : Online Book Reviews", *Journal of Marketing Research*, Vol.43, No.3, 2006, 345-354.
- Cui, G., H.K. Lui, and X. Guo, "The Effect of Online Consumer Reviews on New Product Sales", *International Journal of Electronic Commerce*, Vol.17, No.1, 2012, 39-58.
- Hong, L. and B.D. Davison, "Empirical Study of Topic Modeling in Twitter", *In Proceedings of the First Workshop on Social Media Analytics*, 2010, 80-88.
- Huang, Z., H. Chen, C.J. Hsu, W.H. Chen, and S. Wu, "Credit Rating Analysis with Sup-

- port Vector Machines and Neural Networks : A Market Comparative Study”, *Decision Support Systems*, Vol.37, No.4, 2004, 543-558.
- Hwang, Y., “Facilitating Web Service Taxonomy Generation : An Artificial Neural Network based Framework”, *Journal of Intelligence and Information Systems*, Vol.16, No.2, 2010, 33-54.
- Jo, H.K., I.G. Han, and H.Y. Lee, “Bankruptcy Prediction Using Case-Based Reasoning, Neural Networks, and Discriminant Analysis”, *Expert Systems with Applications*, Vol.13, No.2, 1997, 97-108.
- Kang, B.I., M. Song, and H.S. Jho, “A Study on Opinion Mining of Newspaper Texts based on Topic Modeling”, *Journal of the Korean Library and Information Science Society*, Vol.47, No.4, 2013, 315-334.
- (강범일, 송 민, 조화순, “토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구”, *한국문헌정보학회지*, 제47권, 제4호, 2013, 315-334.)
- Kang, J.E. and M.J. Lee, “Analysis of Urban Infrastructure Risk Areas to Flooding using Neural Network in Seoul”, *Journal of the Korean Society of Civil Engineers*, Vol.35, No.4, 2015, 997-1006.
- (강정은, 이명진, “인공신경망을 활용한 서울시 도시 기반시설 침수위험지역 분석”, *대한토목학회논문집*, 제35권, 제4호, 997-1006.)
- Krizhevsky, A., I. Sutskever, and G.E. Hinton, “Imagenet Classification with Deep Convolutional Neural Networks”, *Advances in Neural Information Processing Systems*, 2012, 1097-1105.
- Lee, J., D.H. Park, and I. Han, “The Effect of Negative Online Consumer Reviews on Product Attitude : An Information Processing View”, *Electronic Commerce Research and Applications*, Vol.7, No.3, 2008, 341-352.
- Lu, Y. and C. Zhai, “Opinion Integration through Semi-Supervised Topic Modeling”, *In Proceedings of the 17th International Conference on World Wide Web*, 2008, 121-130.
- Minka, T. and J. Lafferty, “Expectation-Propagation for The Generative Aspect Model”, *In Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, 2002, 352-359.
- Pang, B. and L. Lee, “Seeing Stars : Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales”, *In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 2005, 115-124.
- Park, D.H., J. Lee, and I. Han, “The Effect of On-line Consumer Reviews on Consumer Purchasing Intention : The Moderating Role of Involvement”, *International Journal of Electronic Commerce*, Vol.11, No.4, 2007, 125-148.
- Sim, H.G. and S.K. Kim, “A Study on Forecasting The Operational Continuous Ability in Battalion Defensive Operations using Artificial Neural Network”, *Journal of Intelligence and Information Systems*, Vol.14, No.3, 2008, 25-39.
- (심흥기, 김승권, “인공신경망을 이용한 대대전투간 작전지속능력 예측”, *지능정보연구*, 제14권, 제3호, 2008, 25-39.)
- Song, Y., S. Pan, S. Liu, M.X. Zhou, and W. Qian, “Topic and Keyword Re-Ranking for LDA-based Topic Modeling”, *In Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 2009, 1757-1760.

- Statistic Korea, *Online Shopping in 2016*, 2016.
(통계청, *온라인쇼핑동향조사*, 2016.)
- Sun, Z., X. Rao, L. Peng, and D. Xu, "Prediction of Protein Supersecondary Structures Based on the Artificial Neural Network Method", *Protein Engineering*, Vol.10, No.7, 1997, 763-769.
- Tam, K.Y. and M.Y. Kiang, "Managerial Applications of Neural Networks : The Case of Bank Failure Predictions", *Management Science*, Vol.38, No.7, 1992, 926-947.
- Tang, H., S. Tan, and X. Cheng, "A Survey on Sentiment Detection of Reviews", *Expert Systems with Applications*, Vol.36, No.7, 2009, 10760-10773.
- Tsang, A.S. and G. Prendergast, "Is A "star" Worth A Thousand Words? The Interplay between Product-Review Texts and Rating Valences", *European Journal of Marketing*, Vol.43, No.11/12, 2009, 1269-1280.
- Turney, P.D., "Thumbs Up or Thumbs Down? : Semantic Orientation Applied to Unsupervised Classification of Reviews", *In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 2002, 417-424.
- Witten, I.H. and E. Frank, *Data Mining : Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Series in Data Management Systems, 2005.

◆ About the Authors ◆



Sang Hyun Park (flypenguin12@khu.ac.kr)

Sang Hyun Park obtained his bachelor degree in Applied Organic Materials Engineering at Inha University. He obtained his M.S. at KyungHee University. His research interests include data mining, big data analysis, AI, and recommender system.



Hyun Sil Moon (pahunter@khu.ac.kr)

Hyun Sil Moon obtained his M.S. and Ph.D. in Management Information Science (MIS), and his B.S. in Business Administration from Kyung Hee University. His current research interests focus on big data analysis, recommender systems, text mining, and social network analysis. He has published numerous papers which have appeared in International Journal of Information Management, Asia Pacific Journal of Information System, Journal of Intelligence and Information Systems, Journal of Information Technology Services, and Journal of Information Technology Applications and Management.



Jae Kyeong Kim (jaek@khu.ac.kr)

Jae Kyeong Kim is a professor at School of Management, Kyunghee University. He obtained his MS and PhD in Management Information Systems (MIS) from KAIST (Korea Advanced Institute of Science and Technology), and his BS in Industrial Engineering from Seoul National University. His current research interests focus on business intelligence, network management, and green business/IT. He has published numerous papers which have appeared in Artificial Intelligence Review, Electronic Commerce Research and Applications, European Journal of Operational Research, Expert Systems with Applications, Group Decision and Negotiations, IEEE transactions on services computing, International Journal of Human-Computer Studies, International Journal of Information Management, Technological Forecasting and Social Change.