

제품 사용 기간을 반영한 기계학습 기반 사용자 평가 변화 예측 모델

부현경* · 김남규**

Machine Learning-based model for predicting changes in user evaluation reflecting the period of the product

Boo Hyunkyung · Kim Namgyu

〈Abstract〉

With the recent expansion of the commerce ecosystem, a large number of user evaluations have been produced. Accordingly, attempts to create business insights using user evaluation data have been actively made. However, since user evaluation can change after the user experiences the product, it is difficult to say that the analysis based only on reviews immediately after purchase fully reflects the user's evaluation of the product. Moreover, studies conducted so far on user evaluation have overlooked the fact that the length of time a user has used a product can affect the user's product evaluation. Therefore, in this study, we build a model that predicts the direction of change in the user's rating after use from the user's rating and reviews immediately after purchase. In particular, the proposed model reflects the product's period of use in predicting the change direction of the star rating. However, since the posterior information on the duration of product use cannot be used as input in the inference process, we propose a structure that utilizes information about the product's period of use using an auxiliary classifier. As a result of an experiment using 599,889 user evaluation data collected from the shopping platform 'N' company, we confirmed that the proposed model performed better than the existing model in terms of accuracy.

Key Words : Deep Learning, Auxiliary Classifier, Review Analysis, Customer Experience

I. 서론

인터넷의 발전과 장기화된 코로나19(COVID-19)로 인한 디지털전환 가속화로 일상생활에 많은 변화가 일어나고 있다. 특히 비대면 소비문화가 확산하면서 온라

인 쇼핑 규모는 꾸준히 증가하고 있다. 통계청 조사에 따르면 2021년 온라인 쇼핑 연간 거래액은 180조 845억 원으로 2018년 거래액과 비교했을 때 64.5% 증가하였다[1, 2]. 이에 온라인 쇼핑 분야에서는 새벽 배송의 'ラスト마일 딜리버리(Lastmile Delivery)'[3]나 실시간 소통이 가능한 '라이브 커머스(Live Commerce)'[4]를 도입하는 등, 커머스의 생태계가 다양한 방향으로 확장되

* 국민대 비즈니스IT전문대학원 박사과정(제1저자)

** 국민대 비즈니스IT전문대학원 교수(교신저자)

고 있다. 최근에는 실제 구매 후기를 영상으로 공유하여 후기에서 판매까지 연결한 ‘리뷰 커머스(Review Commerce)’[5]가 급부상하고 있다.

이와 같은 커머스 생태계의 확장으로 방대한 양의 사용자 리뷰가 생산되면서 이를 활용하고자 하는 시도가 더욱 적극적으로 이루어지고 있다. 기업에서는 사용자 리뷰를 고객 만족도 지표로 두어 기업의 제품이나 서비스 등의 개발 혹은 개선에 사용해 기업의 경쟁력 강화를 위한 수단으로 활용한다[6, 7]. 사용자 리뷰는 사용자의 제품 및 서비스의 구매 결정에 기여 가능한 강력한 도구로 인식되고 있으며, 기존 사용자의 후기 혹은 추천을 통해 상품을 구매하는 경향이 뚜렷하게 존재하기 때문이다[8].

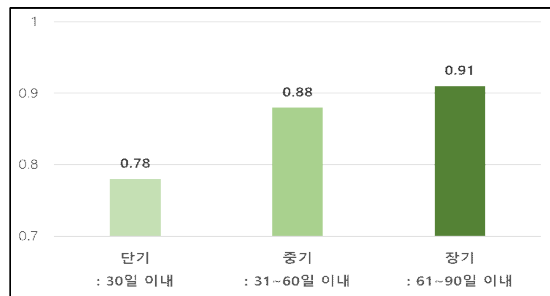
이에 따라 네이버(Naver), 아마존(Amazon) 등 다양한 쇼핑 플랫폼의 리뷰와 사용자 평가 데이터를 활용한 감성 분석, 별점 예측, 그리고 상품 추천 등의 연구가 활발히 수행되고 있다[9-12]. 하지만 사용자의 평가는 사용자의 제품 경험에 따라 충분히 변할 수 있음에도 불구하고 기존의 많은 연구는 사용자의 구매 직후의 리뷰만을 사용하여 수행되었다. 제품 구매 후 구매자가 소유자가 되면서 구매한 제품의 가치를 구매, 사용 중, 사용 후의 과정마다 상이하게 평가함을 감안하면[13] 구매 직후의 리뷰만을 활용한 분석은 제품 사용 후 사용자의 변화된 제품 평가를 반영하기 힘들다는 큰 한계를 가질 수 있다.

<그림 1>은 ‘N’사의 리뷰 시스템에서 특정 사용자가 구매 직후 평가와 제품 사용 후 평가를 작성한 예이다. 구체적으로 <그림 1(a)>는 구매 직후에 작성된 사용자 평가이고, <그림 1(b)>는 해당 사용자가 제품을 일정 기간 사용한 후에 작성한 평가이다. <그림 1>에서 나타난 바와 같이 제품의 구매 직후와 실제 사용 이후의 평가가 상이하게 나타날 수 있으므로 기업들은 사용자들로부터 구매 직후의 평가 외에도 제품 사용 후 평가를 수집하여 분석하려는 방안을 모색하고 있다.



<그림 1> ‘N’사 리뷰 시스템에서의 사용자 평가 예

이처럼 사용자의 실제 사용 경험이 반영된 평가의 중요성이 대두되면서 사용 전/후의 평가 데이터를 활용한 분석이 시도되고 있으며, 특히 최근에는 구매 직후, 즉 제품 사용 전의 리뷰와 별점을 분석하여 실제 제품 사용 이후의 평가가 변화하는 방향을 예측하여 다양한 의사결정에 활용하기 위한 연구[14]가 수행된 바 있다. 하지만 이 연구는 제품 사용 기간이 평가에 미치는 영향을 간과하고, 이를 ‘장기 사용 리뷰’로 획일화하여 분석했다는 한계를 갖는다.



<그림 2> 제품 사용 기간에 따른 사용자 평가 변화

제품 사용 기간에 따라 평가가 상이하게 나타나는 현상은 <그림 2>를 통해 확인할 수 있다. <그림 2>는 ‘N’사에서 제공하는 사용자 평가 데이터 중 생활용품, 디지털/가전, 홈트레이닝 등을 포함한 14개의 제품

카테고리의 제품 평가 154,909건에 대해, 구매 직후에 부여한 별점과 일정 기간 사용 후에 부여한 별점의 차이를 비교한 결과이다. 제품 사용 기간은 구매 후 첫 리뷰를 작성한 날로부터 사용 후 리뷰를 작성한 날까지 소요된 기간으로 측정하였으며, 분석 결과 구매 직후와 사용 후 별점의 차이는 구매 기간에 따라 각각 0.78, 0.88, 그리고 0.91로 다르게 나타남을 확인하였다.

이 결과는 사용자의 제품 구매 직후의 평가와 실제 제품 사용 후 평가의 차이가 사용 기간에 따라 상이하게 나타남을 보여주며, 이는 제품 사용 후 사용자의 평가를 예측하기 위한 연구에서 제품 사용 기간에 대한 고려가 반드시 포함되어야 함을 의미한다. 즉, 사용자의 제품 사용 경험에 따라 사용자의 평가는 변화할 수 있으므로, 구매 직후에 작성된 사용자 평가 데이터를 활용하여 향후 사용자 평가 변화를 예측하는 방안을 제시하고자 한다.

본 연구에서는 구매 직후 평가인 사용자 별점과 리뷰 그리고 제품의 카테고리 정보를 활용하여 제품 사용 후 평가의 변화를 예측하는 모델을 생성한다. 제안 모델은 제품 사용 기간을 보조 분류기[15]로 학습에 반영하여 예측 모델의 정확도를 향상시키고자 하며, 이는 본 모델의 가장 큰 특징이라고 할 수 있다.

본 논문의 이후 구성은 다음과 같다. 먼저 제2장에서는 본 연구와 관련된 선행 연구를 소개하고, 제3장에서는 사용자 구매 직후 평가인 리뷰와 별점, 제품 카테고리 정보 그리고 범주화된 제품 사용 기간을 활용한 보조 분류기(Auxiliary Classifier) 기반의 사용자 평가 변화 예측 방법론을 제안한다. 또한 제4장에서는 제안 방법론을 실제 운영 중인 리뷰 시스템으로부터 수집한 데이터에 적용하여 실험을 수행한 결과를 소개한다. 마지막 장인 제5장에서는 본 연구의 기여와 한계를 정리하고 향후 연구 방향을 제시한다.

II. 관련 연구

2.1 사용자 만족도 관련 연구

사용자 만족도는 재이용, 재방문 등의 리텐션(Retention)에 영향을 미치기 때문에, 제품, 서비스 등 다양한 재화에 대해 사용자의 만족도를 측정하기 위한 연구가 다양하게 진행되었다[16]. 예를 들어 고객 만족도와 고객 재구매 의도에 영향을 주는 웹사이트 품질 요인을 분석하고 고객 리텐션을 위한 모델을 제안한 연구[17], 모바일 서비스(Mobile Service) 분야에서 고객 만족도와 고객 리텐션의 상호 관계를 비교한 연구[18], 그리고 공급자 관점이 아닌 소비자 관점에서의 쇼핑 동기를 이용해 온라인 쇼핑 구매 만족도와 지속적 이용 의도를 분석한 연구[19] 등이 있다.

특히 온라인 사용자 평가와 사용자 구매 의도 간의 상호 관계를 파악한 연구가 활발하게 수행되고 있다. 예를 들어 국내 여행 사이트 방문객의 설문 조사를 통해 온라인 리뷰가 사용자의 구매 의사결정에 영향을 줄 수 있음을 확인한 연구[20]와 사용자 평점보다는 사용자 평가의 개수가 구매 의도에 더 큰 영향을 미친다는 결과를 보인 연구[21], 인터넷 쇼핑을 통해 얻은 경험이 구매 만족에 유의한 영향을 미치고 이 만족이 재구매 의도에도 긍정적인 영향을 주는 것을 밝힌 연구[22], 쇼핑 성향에 따른 구매 만족도의 차이를 제시하고 쇼핑 성향에 따른 마케팅 전략 수립 방안을 제시한 연구[23], 그리고 온라인 서비스 리뷰 이용자의 감정반응요인과 서비스 지속이용의도의 인과 관계를 분석한 연구[24]가 있다.

이와 더불어 이커머스, 호텔, 레스토랑 등을 평가하는 다수의 플랫폼에서 작성된 사용자 리뷰와 별점을 적극적으로 활용하여 사용자 만족도를 분석하고자 하는 시도가 활발하게 이루어지고 있다. 구체적인 예로는 사용자 리뷰 분석을 통해 얻은 제품 속성을 회귀 분석과 결합하여 고객 만족도에 영향을 미치는

요소를 분석한 연구[25], Amazon.com의 온라인 리뷰 데이터를 분석하여 제품에 대한 사용자의 사용 경험 변화를 파악한 연구[26], 그리고 사용자 리뷰와 별점을 결합하여 새로운 만족도 지표를 제안한 연구[27] 등을 들 수 있다.

추천 시스템 분야에서도 사용자 만족도 향상을 위한 연구가 활발히 수행되고 있다. 대표적인 예로는 제품 추천 결과에 대한 고객 만족도를 극대화하기 위해 사용자별 프로필과 제품 평가 점수를 기반으로 고객 등급을 세분화한 연구[28], 동일 사용자가 이용하는 다수의 기기 행동 패턴을 이용해 상품 정보 추천 기법을 제안한 연구[29], 그리고 추천 성능 향상을 위해 사용자 리뷰가 얼마나 유용했는지를 나타내는 지표인 리뷰 유용성을 활용한 연구[13] 등을 들 수 있다. 한편, 편향된 추천 결과는 사용자 만족도 하락에 영향을 줄 수 있음을 지적하고[30], 추천 알고리즘을 평가할 수 있는 다양성 지표를 제시한 연구[31]도 수행된 바 있다.

2.2 DNN(Deep Neural Network) 관련 연구

컴퓨터 자원의 발전으로 방대한 연산량 처리가 가능하게 되면서, 머신러닝(Machine Learning)의 한 분야인 딥러닝(Deep Learning) 기술을 활용한 이미지 객체 인식, 자연어 처리 등의 연구가 폭발적으로 증가하고 있다. 딥러닝은 심층 신경망(Deep Neural Network, DNN)의 구조를 가지며, 은닉층을 구성하는 노드 간 가중치 갱신을 통해 학습하는 알고리즘이다. 이때 심층 신경망은 입력층(Input Layer)과 출력층(Output Layer) 사이에 두 개 이상의 은닉층(Hidden Layer)을 갖는 인공 신경망(Artificial Neural Network, ANN)을 의미한다. 이러한 심층 신경망을 응용한 주요 알고리즘으로는 합성곱 신경망(Convolution Neural Network, CNN), 순환 신경망(Recurrent Neural Network, RNN), LSTM(Long

Short Term Memory), GRU(Gated Recurrent Unit) 등이 있다.

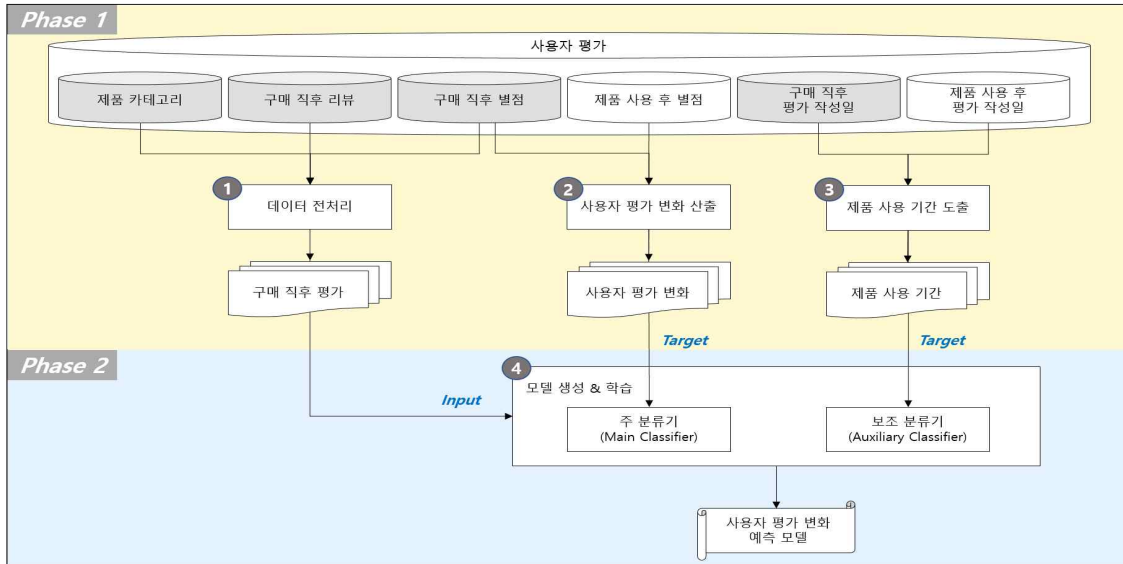
최근에는 이러한 딥러닝 알고리즘을 응용하여 다양한 문제를 딥러닝 기반 모델로 해결하기 위한 시도가 활발히 이루어지고 있다. 대표적인 예로 분류 성능 향상을 위해 은닉층에 별도의 지도 학습을 추가하는 보조 분류기 구조를 제안한 연구[15]를 들 수 있다. 해당 연구에서 제안된 보조 분류기는 학습 과정에서 깊은 신경망 구조에서 발생하는 기울기 소실(Gradient Vanishing) 문제를 해결하며, 추론 모델에서는 학습에 사용되었던 보조 분류기는 제외된다.

이처럼 보조 분류기는 모델의 학습에만 사용되고 실제 추론에서는 사용되지 않기 때문에, 추론 과정에서는 사용할 수 없는 정보를 학습 과정에 사용하기 위한 목적으로 보조 분류기를 활용할 수 있다. 예를 들어 추천 시스템 분야에서 사용자-아이템의 최적 임베딩 벡터(Embedding Vector)를 도출하기 위해 아이템 기반으로 구성된 사용자의 원-핫 벡터(One-hot Vector)를 보조 분류기의 예측값으로 활용한 연구[32], 텍스트 분류 분야에서 계층적으로 구성된 카테고리 분류의 정확도 성능을 높이기 위해 최종 예측 레이블의 상위 레이블 정보를 보조 분류기에 적용한 연구[33], 그리고 이미지 캡셔닝의 성능 향상을 위해 이미지 내 객체의 탐지 및 분류에 보조 분류기를 적용한 연구[34]가 있다.

III. 제안 방법론

3.1 제안 방법론 개요

본 절에서는 구매 직후에 작성된 사용자 평가 데이터 및 제품 사용 기간 정보를 활용하여 제품 사용 후 평가 변화를 예측하는 방법론을 소개한다. 제안 방법론의 전체적인 과정은 <그림 3>과 같다.



<그림 3> 제안 방법론 개요

제안 방법론은 크게 모델의 입력과 출력에 사용할 데이터를 가공하는 Phase 1, 그리고 보조 분류기를 활용한 사용자 평가 변화 예측 모델을 생성하고 학습하는 Phase 2로 구성된다. 또한 그림에서 음영으로 표시된 데이터는 제품 사용 이전에 생성되는 데이터를, 그리고 흰색으로 표시된 데이터는 제품 사용 이후에 생성되는 데이터를 나타낸다. 구체적으로 Phase 1에서 단계 (1)은 사용자의 구매 직후 리뷰를 전처리하고 텍스트 인코딩을 수행하는 과정, 그리고 제품 카테고리를 인코딩하는 과정으로 구성된다. 단계 (2)는 사용자의 제품 사용 후 별점과 구매 직후의 별점의 차이를 산출하고, 산출한 별점차를 이용해 사용자의 평가 변화를 산출하는 과정이다. 단계 (3)은 사용자의 구매 직후 평가 작성일과 제품 사용 후 평가 작성일의 차이를 계산하여 제품 사용 기간을 도출하는 과정이다.

이후 Phase 2에서 단계 (4)는 단계 (1) ~ (3)에서 도출한 데이터를 이용해 사용자 평가 변화 예측 모델을 구축하는 과정이다. 구체적으로 본 모델은 구매 직후

리뷰와 함께 구매 직후 별점 및 제품 카테고리를 입력으로 받아서 사용자 평가 변화를 목표(Target) 변수로 예측하는 학습을 수행한다. 또한 제안 모델은 이러한 주 분류뿐 아니라 보조 분류도 동시에 수행하는데, 보조 분류기의 경우 주 분류와 동일한 입력을 받아 제품 사용 기간을 목표로 예측하는 학습을 수행한다. 이후 전처리를 거친 구매 직후 리뷰, 구매 직후 별점, 그리고 제품 카테고리를 입력으로 사용자 평가 변화를 추론한다.

본 장의 이후 부분은 제안 방법론의 각 단계에 대해 가상의 예를 사용하여 설명하고, 제안 방법론에 실제 데이터를 적용한 성능 평가 결과는 4장에서 제시한다.

3.2 데이터 전처리

본 절에서는 사용자 평가 데이터의 구성 항목을 소개하고, 구매 직후 리뷰의 전처리와 제품 카테고리를 구조화하는 과정(단계 1)을 설명한다.

먼저 <표 1>은 사용자 평가 데이터의 예로, 리뷰 ID, 사용자ID, 제품 카테고리, 구매 직후 평가 작성일, 구매 직후 리뷰, 구매 직후 별점, 제품 사용 후 평가 작성일, 제품 사용 후 리뷰, 그리고 제품 사용 후 별점으로 구성된다. 이러한 항목 중 본 연구에서는 제품 카테고리, 구매 직후 리뷰, 구매 직후 별점, 구매 직후 평가 작성일, 제품 사용 후 평가 작성일, 그리고 제품 사용 후 별점 항목만을 사용한다.

<표 1> 사용자 평가 예

항목	내용
리뷰ID	리뷰1
사용자ID	사용자1
제품 카테고리	화장품
구매 직후 평가 작성일	2019-07-16
구매 직후 리뷰	배송빠름제품도 기대했던그자체임... (생략)
구매 직후 별점	4
제품 사용 후 평가 작성일	2019-09-10
제품 사용 후 리뷰	항 잔잔하니 나쁘지 않네요 지속력은 약한편... (생략)
제품 사용 후 별점	3

구매 직후 리뷰
배송빠름 제품도 기대했던 그자체임Wn
봉투는 상품예안맞게 크기만함Wn
봉투는 안살거임K

텍스트 정규화 결과
배송빠름 제품도 기대했던 그자체임 봉투는 상품예안
맞게 크기만함 봉투는 안살거임

<그림 4> 텍스트 정규화 예

텍스트 정규화 결과
배송빠름 제품도 기대했던 그자체임 봉투는 상품예안
맞게 크기만함 봉투는 안살거임

토큰화 결과
[배송, 빠름, 제품, 도, 기대, 했, 던, 그, 자, 체, 임, 봉, 투, 는, 상, 품, 예, 안, 맞, 게, 크, 기, 만, 함, 봉, 투, 는, 안, 살, 거, 임]

정수 인코딩 결과
[2287, 6621, 6459, 6117, 4158, 5859, 1267, 7872, 1185, 7147, 7436, 7136, 2418, 7641, 5760, 2683, 6896, 6812, 6166, 5400, 4683, 6150, 7837, 2418, 7641, 5760, 3135, 6519, 5377, 7136]

<그림 5> 리뷰 정수 인코딩 예

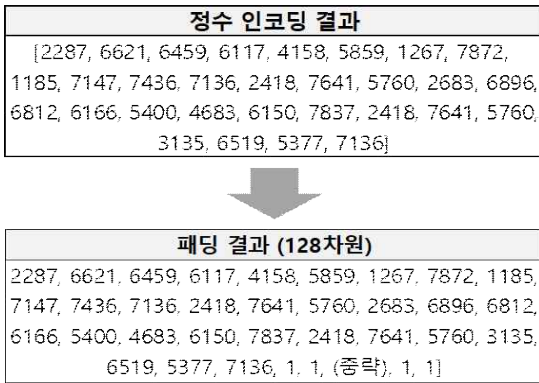
<그림 3>의 단계 (1)에서는 <표 1>의 '구매 직후 리뷰'에 대해 리뷰내 HTML 태그, 특수문자, 이모티콘, 한글 자·모음, 영문자, 숫자를 제거하고 이 과정에서 발생한 다수 공백과 줄바꿈 기호를 한 개의 공백으로 치환하는 텍스트 정규화(Text Normalization), 그리고 정규화된 텍스트 인코딩과 제품 카테고리에 대한 인코딩을 수행한다. 텍스트 정규화는 <그림 4>와 같이 수행한다.

다음으로 <그림 5>의 예와 같이 전처리된 리뷰를 예측 모델에 입력 가능한 형태인 숫자로 변환하는 정수 인코딩(Encoding)을 수행한다. 이 과정은 텍스트 정규화를 통해 도출된 리뷰 텍스트를 BERT 토큰라이저(Tokenizer)를 이용하여 토큰화(Tokenizing)하는 단계와, 이를 통해 도출된 토큰들을 BERT 토큰라이저

에 등록된 단어 사전의 인덱스로 변환하는 단계로 구성된다.

이후 정수로 인코딩된 리뷰 벡터의 크기를 동일하게 맞추는 패딩(Padding) 작업을 진행한다. 인코딩된 리뷰의 토큰 개수가 미리 설정된 최대 차원 수보다 적을 경우, 마지막 인코딩 값 이후부터 '1'의 값을 채우는 방식으로 패딩을 수행한다. <그림 6>은 리뷰 벡터를 128차원으로 맞추기 위해, 마지막 토큰인 7,136의 이후 빈 공간을 모두 '1'로 채우는 패딩의 예이다. 단계 (1)에서는 리뷰의 정규화 및 인코딩뿐 아니라 제품 카테고리의 인코딩도 이루어진다. 우선 제품 카테고리의 경우, 각 리뷰의 대상이 되는 제품이 어떤 카테고리에 속하는지의 정보를 원-핫 인코딩(One-hot

Encoding) 형태로 표시한다. 원-핫 인코딩이란 어떤 정보를 단 하나의 요소만 '1'의 값을 갖고 나머지 모든 요소는 '0'의 값을 갖는 벡터로 인코딩하는 방식이다. 리뷰 5건에 대해 각 리뷰의 대상 제품의 카테고리를 원-핫 인코딩으로 표현한 예는 <표 2>에서 확인할 수 있다.



<그림 6> 리뷰 벡터 패딩 예

<표 2> 리뷰별 제품 카테고리의 원-핫 인코딩 표현 예

리뷰ID	제품 카테고리	원-핫 인코딩
리뷰1	화장품	[0, 0, 0, 1]
리뷰2	생활용품	[0, 1, 0, 0]
리뷰3	디지털/가전	[1, 0, 0, 0]
리뷰4	디지털/가전	[1, 0, 0, 0]
리뷰5	홈트레이닝	[0, 0, 1, 0]

3.3 사용자 평가 변화 산출

본 절에서는 구매 직후 별점과 제품 사용 후 별점을 이용하여 사용자 평가 변화를 산출하는 과정(단계 2)을 설명한다. 우선 사용자 평가의 변화를 산출하기 위해, 구매 직후 별점과 제품 사용 후 별점의 차이를 계산한다. <표 3>은 5건의 리뷰에서 나타난 사용자 평가 변화의 산출 예이다.

<표 3> 리뷰별 사용자 평가 변화 산출 예

리뷰ID	구매 직후 별점	제품 사용 후 별점	별점차	평가 변화 방향	레이블
리뷰1	4	3	-1	하락	1
리뷰2	3	5	2	상승	0
리뷰3	4	5	1	상승	0
리뷰4	4	2	-2	하락	1
리뷰5	2	5	3	상승	0

예를 들어 <표 3>의 '리뷰1'의 경우, 구매 직후 별점이 4점이고 제품 사용 후 별점이 3점이므로 별점 차이는 '-1'로 나타난다. 이는 실제 제품 사용 후 사용자의 평가가 구매 직후에 비해 하락하였다는 것을 의미한다. 즉 사용자 평가 변화는 '상승'과 '하락'을 의미하는 이진 변수로 구성되며, 본 연구에서는 이를 레이블 인코딩(Label Encoding)을 통해 '상승'인 경우는 '0'으로 '하락'인 경우 '1'로 표현한다.

3.4 제품 사용 기간 도출

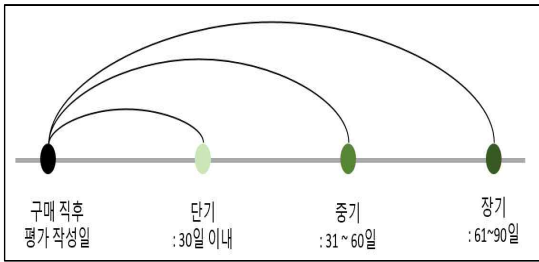
본 절에서는 구매 직후 평가 작성일과 제품 사용 후 평가 작성일을 이용하여 제품 사용일을 계산한 후, 이를 30일 단위로 구간화하여 제품 사용 기간을 산출하는 과정을 설명한다. 우선 5건의 리뷰 각각으로부터 제품 사용일을 산출한 예는 <표 4>와 같다.

<표 4> 리뷰별 제품 사용일 산출 예

리뷰ID	구매 직후 평가 작성일	제품 사용 후 평가 작성일	제품 사용일
리뷰1	2019-07-16	2019-09-10	56
리뷰2	2019-11-07	2019-11-17	10
리뷰3	2021-03-03	2021-04-09	37
리뷰4	2021-12-22	2022-03-15	83
리뷰5	2022-01-02	2022-01-30	28

다음으로 단계 (2)는 <표 4>에서 산출한 제품 사용일을 30일 단위로 구간화하여 제품 사용 기간을 도출

한다. 즉 본 연구에서는 <그림 7>의 예처럼 제품 사용 기간을 제품 사용일에 따라 ‘단기’, ‘중기’, 그리고 ‘장기’로 구분한다. 예를 들어 <표 4>의 리뷰 5건에 대해 <그림 7>의 기준에 따라 제품 사용 기간을 도출한 결과는 <표 5>와 같다. 3개의 구간으로 구분된 제품 사용 기간은 원-핫 인코딩으로 표현하여 이후 분석에 사용한다.



<그림 7> 제품 사용 기간의 구간화

<표 5> 리뷰별 제품 사용 기간 도출 예

리뷰ID	제품 사용일	제품 사용 기간	원-핫 인코딩
리뷰1	56	중기	[0, 1, 0]
리뷰2	10	단기	[1, 0, 0]
리뷰3	37	중기	[0, 1, 0]
리뷰4	83	장기	[0, 0, 1]
리뷰5	28	단기	[1, 0, 0]

3.5 사용자 평가 변화 예측 모델 생성

본 절에서는 본 연구의 핵심인 사용자 평가 변화 예측 모델의 구조를 소개한다. 단계 (1) ~ (3) 과정을 통해 구조화된 데이터는 사용자 평가 변화 예측 모델의 입력 및 목표 변수로 사용되며 그 예는 <표 6>과 같다.

제안 모델은 DNN 모델에 주 분류기와 보조 분류기가 추가 연결된 구조로 구성되며, 제안 모델의 학습 구조는 <그림 8>에서 확인할 수 있다. 보조 분류기는 제품 사용 기간을 예측하고 주 분류기는 본 제안 방법론을 통해 최종적으로 분류하고자 하는 사용

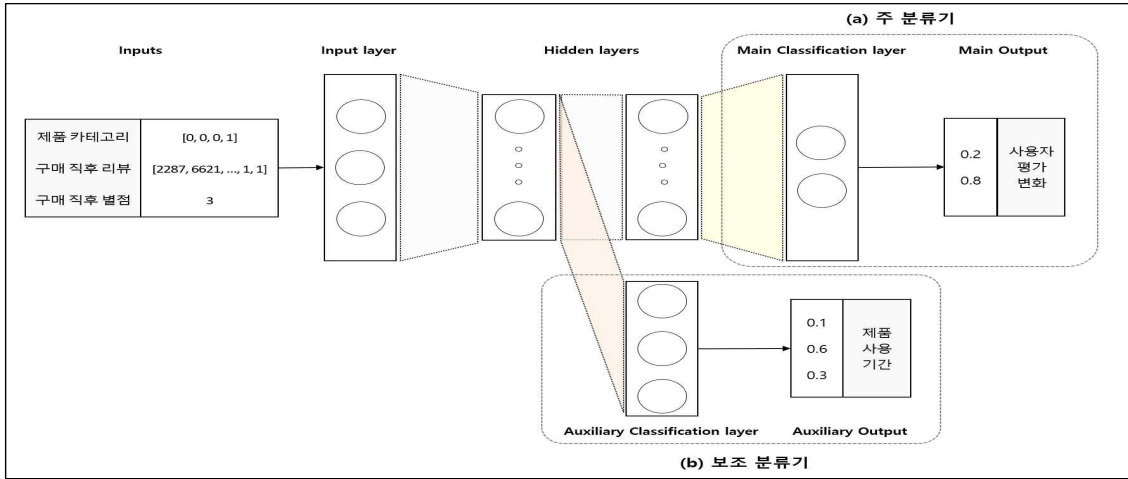
<표 6> 사용자 평가 변화 예측 모델의 입력과 목표 변수 예

항목	역할	내용
제품 카테고리	입력	[0, 0, 0, 1]
구매 직후 리뷰	입력	[2287, 6621, ..., 1, 1]
구매 직후 별점	입력	3
제품 사용 기간	목표 (보조)	[0, 1, 0]
사용자 평가 변화	목표 (주)	1

자 평가 변화를 예측한다. 특히 본 연구에서는 제품 사용 기간이 제품의 평가에 미치는 영향에 주목한다. 하지만 제품 사용 기간은 최초의 제품 구매 시 확정되지 않는 사후 정보의 성격을 갖는다. 따라서 제품 사용 기간은 사용자 평가 변화 예측 모델의 추론 시 입력으로 사용될 수 없다.

이에 따라 본 연구에서는 제품 사용 기간을 모델의 학습에만 반영하고 추론 시에는 입력으로 요구하지 않기 위해 보조 분류기를 활용한 구조를 제안한다. 구체적으로 본 연구에서는 사용자 평가 변화를 예측하는 모델의 학습 과정에서 제품 사용 기간 정보를 보조 분류기의 목표 변수로 설정하여 동시에 학습을 수행한다. 추론 단계에서는 보조 분류기가 필요하지 않으므로, 실제 추론은 제품의 카테고리, 사용자의 구매 직후 리뷰와 별점을 입력으로 받아서 구매 후 사용자 평가 변화의 방향을 예측하는 방식으로 이루어진다.

<그림 8>은 보조 분류기가 ‘장기’, ‘중기’, ‘단기’의 세 개의 클래스로 구분되는 제품 사용 기간을 예측하고, 주 분류기가 ‘상승’, ‘하락’의 두 개의 클래스로 구분되는 사용자 평가 변화를 예측하는 예이다. <그림 8>의 예에서 제품 카테고리, 구매 직후 리뷰, 그리고 구매 직후 별점의 입력값으로 각각 [0, 0, 0, 1], [2287, 6621, ..., 1, 1], 3이 사용되었다. 또한 <그림 8(a)>의 주 분류기의 출력은 사용자 평가 변화를 나타내는 ‘상승’, ‘하락’ 클래스 각각에 대한 예측 확률로 0.2, 0.8이 산출되었고, <그림 8(b)>의 보조 분류기 출력은

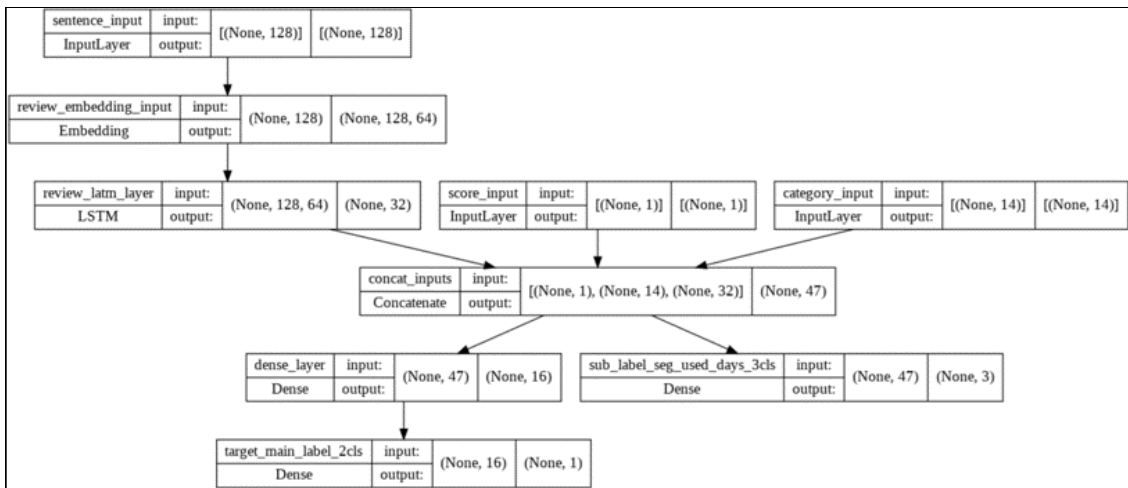


<그림 8> 제안 모델의 학습 구조

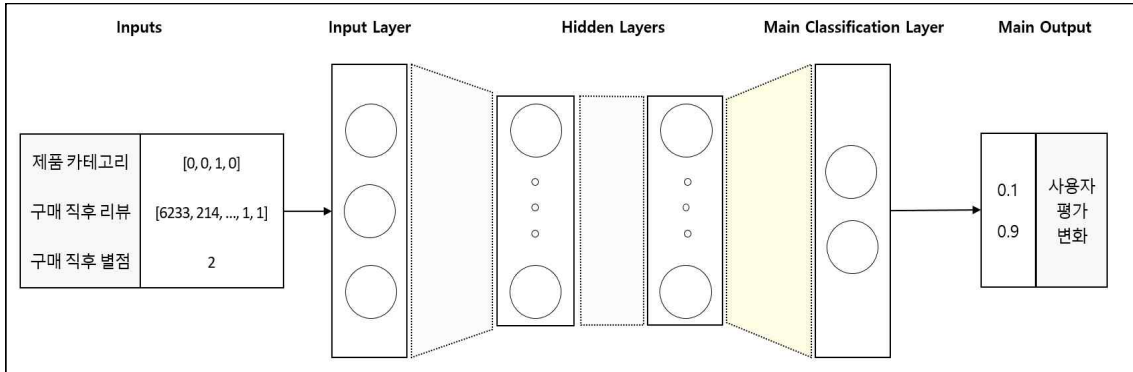
제품 사용 기간을 나타내는 ‘단기’, ‘중기’, ‘장기’ 클래스 각각에 대한 예측 확률로 0.1, 0.6, 그리고 0.3이 산출되었다.

이때 제안 모델은 주 분류기를 통한 예측과 보조 분류기를 통한 예측이 동시에 수행되는 종단간 (End-to-End) 학습 과정을 거치므로, 전체 손실(Total Loss)은 주 분류기와 보조 분류기의 손실을 모두 받

영하여 계산된다. 구체적으로 모델의 전체 손실은 <식 1>과 같이 주 분류 손실(Main Loss)과 보조 분류 손실(Auxiliary Loss)의 가중 합으로 산출된다. 이때 가중치에 따라 전체 손실 값 다르게 계산되며, 전체 손실 값은 모델의 성능에 영향을 미친다. 따라서 가중치 값을 다양하게 조절하며 모델 성능을 분석하는 것이 반드시 필요하다.



<그림 9> 학습 모델의 내부 구현



<그림 10> 제안 모델의 추론 구조

$$Total\ Loss = \alpha * (Auxiliary\ Loss) + \beta * (Main\ Loss) \quad (1)$$

(α : 보조 분류의 손실 가중치, $0 \leq \alpha \leq 1$,
 β : 주 분류의 손실 가중치, $0 \leq \beta \leq 1$)

한편 <그림 9>는 <그림 8>의 개념적 구조를 본 연구의 실험을 위해 실제로 구현한 구조이다. 해당 구조에서 보조 분류기의 목표 변수인 제품 사용 기간은 ‘단기’, ‘중기’, ‘장기’의 클래스로 구성되므로 출력 크기를 3으로 지정하였고, 주 분류기의 목표 변수인 사용자 평가 변화는 ‘상승’, ‘하락’의 클래스로 구성되므로 출력 크기를 1로 지정하였다. 전술한 바와 같이 제안 모델의 목적은 구매 직후의 사용자 평가 데이터를 활용하여 제품 사용 후 사용자 평가의 변화를 예측하는 것이므로, 제품 사용 기간이 필요한 보조 분류기는 모델의 학습에만 사용하고 추론 시에는 주 분류기만 사용한다. 이러한 구조는 <그림 10>에서 확인할 수 있다.

IV. 실험

4.1 실험 개요

본 장에서는 제안 방법론의 성능 평가를 위해 수행

한 실험의 결과를 제시하고 분석한다. 먼저 본 연구에서는 Python 3.7을 통해 실험 환경을 구축하였고 구체적인 H/W와 S/W 환경은 <표 7>과 같다.

<표 7> 실험 환경

H/W	
GPU	Tesla V100 PCIE 32GB
CPU	Intel Xeon Core 16
Memory	192G
S/W	
Python	3.7.15
Tensorflow-gpu	2.9.2
Keras	2.9.0
CUDA	11.2

또한 실험 데이터는 쇼핑 플랫폼인 ‘N’사로부터 14개의 대분류 카테고리에서 제품을 무작위로 선정해 해당 제품에 대한 사용자 평가 599,889건의 데이터를 수집하였다. 14개의 대분류 카테고리는 생활용품, 디지털/가전, 홈트레이닝, 음식, 자동차공구, 패션잡화, 인테리어, 화장품, 유아동, 가구, 캠핑용품, 강아지, 건강의료용품, 그리고 고양이로 구성된다. <표 8>은 사용자 평가 변화 방향별 데이터 수를 요약한 결과이다. 표에서 ‘동일’은 구매 직후 별점과 제품 사용 후

별점이 동일한 경우를 의미하고, '상승'은 제품 사용 후의 별점이 구매 직후 별점에 비해 상승한 경우를, '하락'은 그 반대의 경우를 나타낸다.

<표 8> 실험 데이터

별점 변화	데이터 수 (건)	구성비
하락	57,288	9.55%
동일	475,695	79.30%
상승	66,906	11.15%

4.2 실험 데이터 구축

본 절에서는 실험 데이터 구축의 과정 및 결과를 소개한다. 먼저 3.2절의 과정에 따라 데이터 전처리를 수행한 결과 사용 가능 데이터는 493,166건으로 추려진다. 추려진 데이터에 대한 별점 분포는 <표 9>와 같다. 아래 <표 9>에서 가장 빈도가 높은 데이터는 사용 전후의 별점이 모두 5점인 데이터로, 전체의 68.59%를 차지하고 있음을 확인하였다. 실제로 별점

이 5점인 데이터의 경우 불성실 리뷰가 다수 존재함이 관련 업계에서 꾸준히 지적됐으며, 본 실험에 사용한 데이터 또한 이러한 현상이 발견되었다.

따라서 본 실험에서는 사용 전후의 별점이 모두 5점인 데이터를 실험에서 제외하기로 하였으며, 모든 별점에 대해 일관된 기준을 적용하기 위해 사용 전과 후의 별점이 동일한 데이터를 모두 실험에서 제거하였다. 그 결과 모든 데이터의 별점 변화는 '동일'이 아닌 '상승' 또는 '하락'의 값만 갖게 된다. 한편 구매 직후 별점이 1점인 경우의 평가 변화는 '상승'만 가능하고, 반대로 구매 직후 별점이 5점인 경우의 평가 변화는 '하락'만 가능하다. 따라서 구매 직후 별점이 1점 또는 5점인 데이터 역시 실험에서 제외하였다. 이러한 과정을 통해 필터링된 사용 전후의 별점 분포는 <표 10>과 같다.

또한 <표 10>에서 각 셀에 해당하는 데이터 분포에 큰 불균형이 존재하는 것을 확인할 수 있다. 이때 불균형 데이터를 모델 생성 및 학습에 그대로 사용하는 경우 다양한 문제가 발생할 수 있으므로, 본 실험

<표 9> 제품 사용 전후의 원 별점 분포

		제품 사용 후 별점					총합계
		1	2	3	4	5	
구매 직후 별점	1	1,116 (0.23%)	307 (0.06%)	486 (0.1%)	260 (0.05%)	657 (0.13%)	2,826 (0.57%)
	2	472 (0.1%)	545 (0.11%)	1,005 (0.2%)	539 (0.11%)	619 (0.13%)	3,180 (0.64%)
	3	996 (0.2%)	1,418 (0.29%)	9,632 (1.95%)	6,591 (1.34%)	5,436 (1.1%)	24,073 (4.88%)
	4	932 (0.19%)	1,209 (0.25%)	9,426 (1.91%)	39,146 (7.94%)	32,009 (6.49%)	82,722 (16.77%)
	5	1,803 (0.37%)	1,453 (0.29%)	8,520 (1.73%)	30,332 (6.15%)	338,257 (68.59%)	380,365 (77.13%)
	총합계	5,319 (1.08%)	4,932 (1%)	29,069 (5.89%)	76,868 (15.59%)	376,978 (76.44%)	493,166 (100%)

<표 10> 제품 사용 전후의 필터링된 별점 분포

		제품 사용 후 별점					총합계
		1	2	3	4	5	
구매 직후 별점	2	472 (0.78%)	0 (0%)	1,005 (1.66%)	539 (0.89%)	619 (1.02%)	2,635 (4.34%)
	3	996 (1.64%)	1,418 (2.34%)	0 (0%)	6,591 (10.87%)	5,436 (8.96%)	14,441 (23.81%)
	4	932 (1.54%)	1,209 (1.99%)	9,426 (15.54%)	0 (0%)	32,009 (52.77%)	43,567 (71.85%)
	총합계	2,400(3.96%)	2,627 (4.33%)	10,431 (17.2%)	7,130 (11.76%)	38,064 (62.76%)	60,652 (100%)

에서는 언더 샘플링(Under sampling) 기법을 적용해 데이터 불균형 문제를 해소한다. 구체적으로 전체 셀 중 데이터의 수가 가장 적은 경우, 즉 별점이 2점에서 1점으로 변화한 경우의 데이터 수인 472건에 맞추어, 나머지 유형에서도 472건씩의 데이터를 랜덤 샘플링(Random sampling)을 적용해 추출하였다. 이를 통해 최종 데이터 5,664건을 확보하였으며, 이를 다시 6:2:2의 비율로 분할하여 학습(Train), 검증(Validation), 그리고 테스트(Test) 데이터 세트를 구축하였다. 구체적인 실험 데이터 수는 <표 11>에서 확인할 수 있다.

<표 11> 실험 데이터 구성

데이터 세트	상승(0)	하락(1)	전체
학습	1,698	1,698	5,664
검증	567	567	
테스트	567	567	

4.3 입력 및 목표 변수 생성

본 절에서는 실험에서 사용할 입력 및 목표 변수의 도출 과정과 결과를 소개한다. 먼저 <표 11>에서 추출한 실험 데이터에 대해 3.3절과 3.4절의 과정에 따라 사용자 평가 변화와 제품 사용 기간을 도출한다.

<표 12>는 이 과정을 통해 도출된 결과 일부로, 제품 카테고리, 구매 직후 리뷰 및 별점, 사용자 평가 변화, 그리고 제품 사용 기간으로 구성된다. 또한 <표 13>은 <표 12>의 값을 모델이 사용할 수 있는 형태로 변환한 결과 일부이다. 해당 데이터는 사용자 평가 변화를 예측하는 모델의 학습 및 평가 데이터로 사용된다.

4.4 모델 성능 평가

본 절에서는 제안 모델인 제품 사용 기간을 활용한 보조 분류기 기반 사용자 평가 변화 예측 모델의 성능을 비교 모델들과 함께 평가한 결과를 소개한다. 본 제안 방법론의 성능 평가를 위한 실험 설계는 <그림 11>과 같다. 먼저 본 실험에서는 제안 모델의 성능을 평가하기 위해 두 개의 비교 모델을 설정하였다. 먼저 비교 모델 (A)는 보조 분류기를 활용하지 않은 단순 DNN 모델이다. 그리고 제안 모델 (C)는 본 연구에서 제안하는 방법론으로, 보조 분류기의 출력 값으로 제품 사용 기간을 이용하는 모델이다. 또한 비교 모델 (B)는 제안 모델 (C)와 동일 구조이나, 비교 모델 (B)의 보조 분류기는 제품 사용 기간을 예측하는 대신 주 분류기와 동일한 사용자 평가 변화를

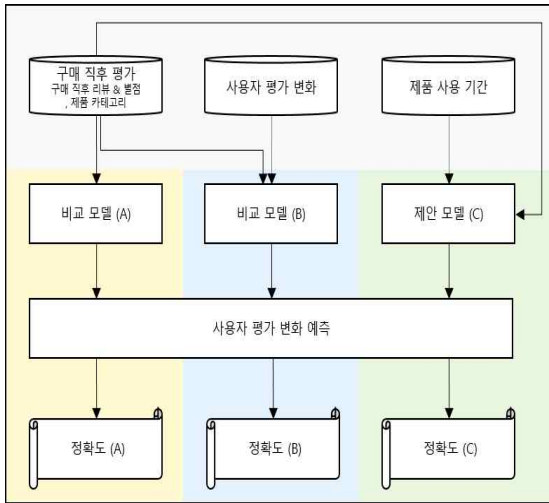
<표 12> 원본 입력 변수와 도출된 목표 변수 (일부)

제품 카테고리	구매 직후 리뷰	구매 직후 별점	사용자 평가 변화	제품 사용 기간
생활	가격이 정말 메리트 있어요	4	상승	중기
디지털/가전	가격은 더 저렴한데 무지 가슴기 ... (생략)	3	상승	장기
디지털/가전	안방에 틀어놓고 자는데 잘모르겠어요	2	상승	장기

<표 13> 변환된 입력 변수와 목표 변수 (일부)

제품 카테고리	구매 직후 리뷰	구매 직후 별점	사용자 평가 변화	제품 사용 기간
[0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]	[730, 4102, 2016, 6122, 7659, 3868, ... (생략)]	4	0	[0, 1, 0]
[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]	[729, 1698, 3992, 7828, 5850, ... (생략)]	3	0	[0, 0, 1]
[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]	[3135, 6305, 6896, 4793, 6855, ... (생략)]	2	0	[0, 0, 1]

예측한다. 이때 비교 모델 (B)는 제안 모델의 성능 향상이 단순히 보조 분류기를 추가한 것에 기인하는지, 아니면 보조 분류기에서 제품 사용 기간을 예측하는 학습을 수행하는 것에 기인하는지 확인하기 위해 설계되었다.



<그림 11> 전체 실험 구조

성능 평가는 <표 11>에 요약된 바와 같이 학습 3,396건, 검증 1,134건, 그리고 테스트 1,134건의 데이터를 사용하여 수행되었다. 특히 학습 모델의 경우 손실 가중치의 최적값을 찾기 위해 손실 가중치를 다양하게 변경하며 반복 실험을 진행하였다. 구체적으로, 주 분류의 손실 가중치는 1로 고정하고 보조 분류의 손실 가중치를 0부터 1까지 0.1 단위로 늘려가며 실험을 수행하였다. 이때 비교 모델과 제안 모델 모두 모델 과적합 방지를 위해 Early Stopping을 적용하였다. 또한 배치 1회당 사용되는 데이터 수인 배치 사이즈(Batch Size)와 보조 분류의 손실 가중치를 제외한 나머지 실험 파라미터는 고정하였다. 실험에 적용한 파라미터는 <표 14>에 요약되어 있다. 그리고 실험을 통해 얻은 모델의 최적 손실 가중치와 정확도가 포함된 성능 평가 결과는 <표 15>에 요약되어 있으며 모델 성능 비교 결과는

<그림 12>에 나타나 있다.

<표 14> 실험 파라미터

조건	설정값	설명
Batch Size	64, 128, 256, 512	배치 1회당 사용 데이터 수
Activation	relu	활성화 함수
Optimizer	adamax	최적화 함수
Learning Rate	0.001	학습률
Epoch	50	학습 횟수
Early Stopping	monitor='val_loss', patience=10	학습 조기 종료 조건
Auxiliary Loss	0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1	보조 분류의 손실 가중치

앞서 전술한 바와 같이 보조 분류기를 활용한 모델의 경우 주 분류의 손실 가중치는 1로 고정된 상태에서 보조 분류의 손실 가중치 변경을 통해 실험을 수행하였다. 이 결과 비교 모델 (B)와 제안 모델 (C)의 보조 분류의 최적 손실 가중치를 도출할 수 있었으며 그 결과는 <표 15>에서 확인할 수 있다.

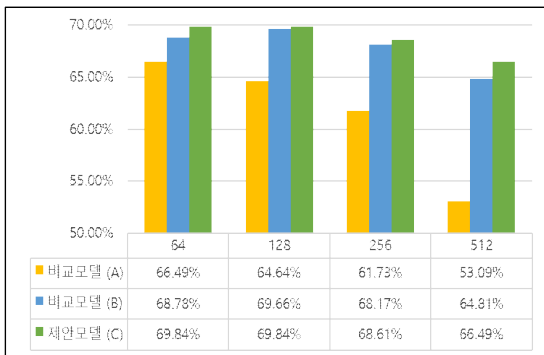
<표 15> 모델 성능 평가 결과

	비교 모델 (A)		비교 모델 (B)		제안 모델 (C)	
	Acc.	Aux. Loss	Acc.	Aux. Loss	Acc.	Aux. Loss
Batch Size						
64	66.49%	0.4	68.78%	0	69.84%	
128	64.64%	0.4	69.66%	0.2	69.84%	
256	61.73%	1	68.17%	0.6	68.61%	
512	53.09%	0	64.81%	0.7	66.49%	

먼저 비교 모델 (B)의 경우 각 배치 사이즈 '64', '128', '256', '512'에 대해 보조 분류의 최적 손실 가중치는 '0.4', '0.4', '1', '0'일 때 성능이 가장 높았으며 제안 모델 (C)의 경우 각 배치 사이즈 '64', '128', '256', '512'에 대해 보조 분류의 최적 손실 가중치는 '0',

'0.2', '0.6', '0.7'일 때 성능이 가장 높게 나왔다. 비교 모델 (A)는 보조 분류기가 없는 단순 DNN 모델이므로 보조 분류의 손실 가중치는 따로 없다.

그리고 모델 성능 비교 결과인 <그림 12>를 통해 제안 모델 (C)와 비교 모델 (A)를 비교했을 때 제안 모델 (C)가 우수한 성능을 보임을 확인하였다. 그리고 보조 분류기를 활용한 모델인 제안 모델 (C)와 비교 모델 (B)를 비교했을 때는 제안 모델 (C)가 비교 모델 (B)의 정확도보다 소폭 상승했음을 확인할 수 있었는데 이는 동일한 모델 구조에서 보조 분류기의 예측값으로 제품 사용 기간을 이용한 것이 모델 성능 향상에 긍정적인 영향을 미쳤다고 해석할 수 있다.



<그림 12> 모델 성능 비교 결과

V. 결론

최근 커머스 생태계의 확장으로 방대한 양의 사용자 평가가 생산되고 있으며, 이에 따라 이를 활용하여 인사이트를 창출하려는 시도가 더욱 활발하게 이루어지고 있다. 하지만 사용자의 평가는 사용자의 제품 사용 경험에 따라 충분히 변화할 수 있으므로 구매 직후의 리뷰만을 활용한 분석은 제품에 대한 사용자의 평가를 온전히 반영한 것으로 인정받기 어렵다. 이에 본 연구에서는 구매 직후에 작성된 사용자 평가

데이터를 기반으로 사용자 평가 변화를 예측하는 방안을 제시하였다. 특히 이 과정에서 보조 분류기 구조를 통해 제품 사용 기간 정보를 활용하였으며, 실험 결과 제안 모델이 예측 정확성 측면에서 우수한 성능을 보임을 확인하였다.

본 연구에서는 구매 직후의 사용자 평가를 바탕으로 제품 사용 후의 사용자 평가 변화를 예측하고, 이 과정에서 제품 사용 기간을 반영하기 위해 보조 분류기를 활용한 모델을 제안하였다. 이러한 시도는 기존의 사용자 평가를 다룬 연구에서는 이루어지지 않았던 것으로, 제품 사용 기간을 평가에 반영하는 새로운 방안을 제시했다는 점에서 본 연구의 학술적 기여를 인정받을 수 있다.

또한 제안 모델을 통해 기업은 구매 직후의 사용자 평가로부터 향후 제품 사용 후 사용자 평가의 변화를 예측할 수 있다. 예를 들어 제안 방법론을 적용하여 기업은 현재 판매 중인 제품에 대한 사용자의 향후 만족도를 사전에 예측할 수 있으며, 이를 기반으로 만족도 하락이 예측되는 상품일 경우 이를 대체할 수 있는 제품을 추가로 판매하는 등의 판매 전략을 선제적으로 수립할 수 있다. 또한 예측을 통해 향후 만족도가 상승할 제품만을 선별하여 판매 촉진을 위한 제품 기획전 등을 수행해볼 수도 있다. 이처럼 제안 방법론은 사용자 반응에 대한 선제 대응이 필요한 분야에 널리 활용될 수 있으며, 이러한 측면에서 본 연구의 실무적 기여가 인정된다.

다만 본 연구에서는 사용자 평가의 변화 방향을 상승과 하락의 두 가지로 구분하여 실험을 수행하였으나, 향후 상승과 하락뿐 아니라 유지를 포함한 세 가지 클래스를 구분하는 모델로 확장하여 모델의 현실성을 높일 필요가 있다. 또한 향후 연구에서는 사용자 평가 변화의 방향성 예측에서 더 나아가, 사용자의 평가가 얼마나 변화했는지를 정량적으로 예측하는 모델이 고안될 필요가 있다.

참고문헌

- [1] 통계청, 2022년 1월 온라인쇼핑 동향, 2022, www.kostat.go.kr
- [2] 통계청, 2018년 1월 온라인쇼핑 동향, 2018, www.kostat.go.kr
- [3] 강선·현병환, “온라인 쇼핑몰충성도에 미치는 배송서비스 품질의 핵심 요인 연구: 쿠팡의 로켓배송 중심,” 경영학 연구, 제36권, 제3호, 2021, pp.49-77.
- [4] 온가신·이혜미, “라이브 커머스는 어떻게 소비자들의 마음을 사로잡았나? 플랫폼 및 BJ 정보원천 특성이 중국 소비자들의 라이브 커머스 구매의도에 미치는 영향과 성별의 조절효과,” e-비즈니스 연구, 제21권, 제5호, 2020, pp.215-234.
- [5] 김경은 (2022.05.22), 리뷰-라이크-콘텐츠로 상품 판다... ‘커머스’ 생태계의 진화, 아주경제, <https://www.ajunews.com/view/20220521111312217>
- [6] 김도훈·차경진, “AI 키즈폰의 소비자리뷰 분석을 통한 제품개선 전략에 대한 연구,” 한국전자거래학회지, 제24권, 제2호, 2019, pp.71-89.
- [7] 이주민·방정혜, “화장품 회사의 빅데이터분석을 통한 브랜드컨셉 개발 사례분석,” 지식경영연구, 제21권, 제3호, 2020, pp.215-228.
- [8] 이재성·김재영·강병욱, “RFM 다차원 분석 기법을 활용한 암시적 사용자 피드백 기반 협업 필터링 개선 연구,” 지능정보연구, 제25권, 제1호, 2019, pp.139-161.
- [9] Wu, S., Zhang, Y., Zhang, W., Bian, K. and Cui, B., “Enhanced review-based rating prediction by exploiting aside information and user influence,” Knowledge-Based Systems, Vol.222, 2021.
- [10] Wassan, S., Chen, X., Shen, T., Waqar, M. and Jhanjhi, N. Z., “Amazon product sentiment analysis using machine learning techniques,” Revista Argentina de Clínica Psicológica, Vol.30, No.1, 2021, pp.695-703.
- [11] 한영경·김철민·박광호, “온라인 고객 리뷰에 대한 텍스트마이닝을 활용한 고객가치제안 방법,” 산업경영시스템학회지, 제44권, 제4호, 2021, pp.85-97.
- [12] LIU FAN·이병현·최일영·정재호·김재경, “E-커머스 사용자의 평점과 리뷰 유용성이 상품 추천 시스템의 성능 향상에 미치는 영향 분석,” 지능정보연구, 제28권, 제1호, 2022, pp.311-328.
- [13] 박경진, “제품의 구매가치와 소유가치 인식 변화에 대한 연구,” 한국과학예술통합학회, 제34권, 2018, pp.119-131.
- [14] 안지예·김남규, “딥러닝을 활용한 고객 경험 기반 상품 평가 변화 예측 방법론,” 한국IT서비스학회지, 제21권, 제4호, 2022, pp.75-90.
- [15] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D. and V. Rabinovich, A., “Going deeper with convolutions,” In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp.1-9.
- [16] 박연자, “고객만족과 고객감동이 고객충성도에 미치는 영향에 관한 실증연구; 재구매와 고객유지에 대한 중점 연구,” 한국산업정보학회논문지, 제24권, 제1호, 2019, pp.109-123.
- [17] Kumar, V. and Ayodeji, O. G., “E-retail factors for customer activation and retention: An empirical study from Indian e-commerce customers,” Journal of Retailing and Consumer Services, Vol.59, 2021.
- [18] Alshurideh, M., Masa’deh, R. M. D. T. and Alkurdi, B., “The effect of customer satisfaction upon customer retention in the Jordanian

- mobile market: An empirical investigation," *European Journal of Economics, Finance and Administrative Sciences*, Vol.47, No.12, 2012, pp.69-78.
- [19] 전병호 · 최재용 · 김재영, "온라인 쇼핑 가치 특성이 구매만족과 재이용의도에 미치는 영향 연구," (사)디지털산업정보학회 논문지, 제13권, 제1호, 2017, pp.147-158.
- [20] 윤영선, "온라인 리뷰가 온라인 쇼핑행동에 미치는 영향," *국제회계연구*, 제52권, 2013, pp.139-156.
- [21] 안서원, "구매 여정에서 나타나는 고객 평점과 후기의 영향," *마케팅연구*, 제36권, 제4호, 2021, pp.135-155.
- [22] 유순순 · 손조기 · 전상택, "중국인터넷쇼핑몰에서 총체적 체험과 재구매의도 관계에서 쇼핑가치의 조절효과 연구," *e-비즈니스연구*, 제21권, 제5호, 2020, pp.89-112.
- [23] 이지용 · 김용선, "화장품 쇼핑성향 유형에 따른 화장품 온라인 리뷰 신뢰도, 구매 만족도, 구전행동 영향관계," *한국미용학회지*, 제24권, 제2호, 2018, pp.243-255.
- [24] 채수인 · 권두순 · 박복원 · 박동철, "스마트 관광에서의 항공·호텔 온라인 리뷰 이용자의 감정반응요인이 지속이용의도에 미치는 영향," (사)디지털산업정보학회 논문지, 제17권, 제4호, 2021, pp.209-229.
- [25] Wang, Y., Lu, X. and Tan, Y., "Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines," *Electronic Commerce Research and Applications*, Vol.29, 2018, pp.1-11.
- [26] 박정은 · 양동욱 · 김하영, "인공지능 스피커의 세대별 온라인 리뷰 분석을 통한 사용자 경험 요인 탐색," *한국융합학회논문지*, 제12권, 제7호, 2021, pp.193-205.
- [27] Balbi, S., Misuraca, M. and Scepti, G., "Combining different evaluation systems on social media for measuring user satisfaction," *Information Processing & Management*, Vol.54, No.4, 2018, pp.674-685.
- [28] Jiang, Y., Shang, J. and Liu, Y., "Maximizing customer satisfaction through an online recommendation system: A novel associative classification model," *Decision Support Systems*, Vol.48, No.3, 2010, pp.470-479.
- [29] 권준희, "이커머스 환경에서 구매와 공유 행동을 이용한 기기 중심 개인화 상품 정보 추천 기법," (사)디지털산업정보학회 논문지, 제18권, 제4호, 2022, pp.85-96.
- [30] 이지윤 · 박규동 · 조민수, "추천 시스템의 편향 및 사용자 만족도 탐구: 체계적 문헌고찰 방법을 중심으로," *한국디지털콘텐츠학회 논문지*, 제23권, 제9호, 2022, pp.1823-1832.
- [31] 백주린, "E-Commerce 데이터를 활용한 개인화 추천 다양성 증대 방법 비교," *한국컴퓨터정보학회논문지*, 제27권, 제9호, 2022, pp.59-68.
- [32] 권유진 · 최민석 · 조운호, "부가 정보를 활용한 비전 트랜스포머 기반의 추천시스템," *지능정보연구*, 제28권, 제3호, 2022, pp.119-137.
- [33] 김윤하 · 김남규, "오토인코더 기반 심층 지도 네트워크를 활용한 계층형 데이터 분류 방법론," *지능정보연구*, 제28권, 제3호, 2022, pp.185-207.
- [34] Çaylı, Ö., Kılıç, V., Onan, A. and Wang, W., "Auxiliary Classifier based Residual RNN for Image Captioning," In *2022 30th European Signal Processing Conference (EUSIPCO)*. IEEE., 2022, pp.1126-1130.

■ 저자소개 ■



부 현 경
Boo Hyunkyung

2020년 3월-현재
국민대학교 비즈니스IT전문대학원
비즈니스IT전공 박사과정
2018년 2월 국민대학교 비즈니스IT전문대학원
전자정부전공(이학석사)
2015년 2월 강릉원주대학교 정보통계학과
(이학사)

관심분야 : Deep Learning, Text Mining
E-mail : hkboo@kookmin.ac.kr



김 남 규
Kim Namgyu

2007년 3월-현재
국민대학교 경영정보학부 및
비즈니스IT전문대학원 교수
2007년 2월 KAIST 경영공학과(공학박사)
2000년 2월 KAIST 경영공학과(공학석사)
1998년 2월 서울대학교 컴퓨터공학과(공학사)

관심분야 : Deep Learning, Text Mining
E-mail : ngkim@kookmin.ac.kr

논문접수일 : 2022년 12월 30일
수정접수일 : 2023년 1월 17일
게재확정일 : 2023년 1월 21일