

Methodology for Identifying Key Factors in Sentiment Analysis by Customer Characteristics Using Attention Mechanism

Kwangho Lee*, Namgyu Kim**

*Graduate Student, Graduate School of Business IT, Kookmin University, Seoul, Korea

**Professor, School of Management Information Systems, Kookmin University, Seoul, Korea

[Abstract]

Recently, due to the increase of online reviews and the development of analysis technology, the interest and demand for online review analysis continues to increase. However, previous studies have not considered the emotions contained in each vocabulary may differ from one reviewer to another. Therefore, this study first classifies the customer group according to the customer's grade, and presents the result of analyzing the difference by performing review analysis for each customer group. We found that the price factor had a significant influence on the evaluation of products for customers with high ratings. On the contrary, in the case of low-grade customers, the degree of correspondence between the contents introduced in the mall and the actual product significantly influenced the evaluation of the product. We expect that the proposed methodology can be effectively used to establish differentiated marketing strategies by identifying factors that affect product evaluation by customer group.

▶ **Key words:** Attention Mechanism, Big Data, Deep Learning, Review Analysis, Text Analytics

[요 약]

최근 온라인 리뷰의 증가와 분석 기술의 발달로 인해 온라인 리뷰 분석에 대한 관심과 수요가 지속적으로 증가하고 있다. 하지만 리뷰 분석을 다룬 기존의 연구는 동일한 어휘라도 각 어휘에 담긴 감정은 리뷰어에 따라 서로 다를 수 있다는 점을 반영하지 못했다는 한계를 갖는다. 따라서 본 연구에서는 고객의 등급에 따라 고객군을 분류하고, 각 고객군별로 리뷰 분석을 수행하여 그 차이를 분석한 결과를 제시하였다. 의류 전문 쇼핑몰인 'M'사의 리뷰에 대한 분석을 수행한 결과, 쇼핑몰 사용도가 높은 고객의 경우 가격적인 요소가, 쇼핑몰 사용도가 낮은 고객의 경우 쇼핑몰에 소개된 내용과 실제 제품의 일치 정도가 제품의 긍/부정 평가에 크게 영향을 미치는 것으로 파악되었다. 제안 방법론은 고객군별로 제품 평가에 중요하게 영향을 미치는 요소를 식별함으로써, 차별화된 마케팅 전략을 수립하는 데에 효과적으로 활용될 수 있을 것으로 기대한다.

▶ **주제어:** 딥러닝, 빅데이터, 리뷰 분석, 어텐션, 텍스트 분석

-
- First Author: Kwangho Lee, Corresponding Author: Namgyu Kim
 - *Kwangho Lee (lkh5021@kookmin.ac.kr), Graduate School of Business IT, Kookmin University
 - **Namgyu Kim (ngkim@kookmin.ac.kr), School of Management Information Systems, Kookmin University
 - Received: 2020. 01. 14, Revised: 2020. 03. 16, Accepted: 2020. 03. 16.

I. Introduction

최근 온라인 기술 발전에 힘입어 온라인 쇼핑 시장은 지속해서 성장하고 있으며, 특히 온라인 결제 환경의 발전은 시장의 성장을 촉진하고 있다. 통계청의 ‘온라인쇼핑동향’에 따르면 온라인 쇼핑 거래액은 2019년 8월 기준 총 11조 2,535 억원에 달하며 전년 동월 대비 21.4% 성장해 2017년 집계 이후 최고치를 기록했다. 대한상공회의소는 국내 소매 유통업 경기전망 지수가 올해 2분기까지 하락하였으나 전 업태 중 온라인 쇼핑 시장은 성장세를 기록할 것이라 전망하였다.

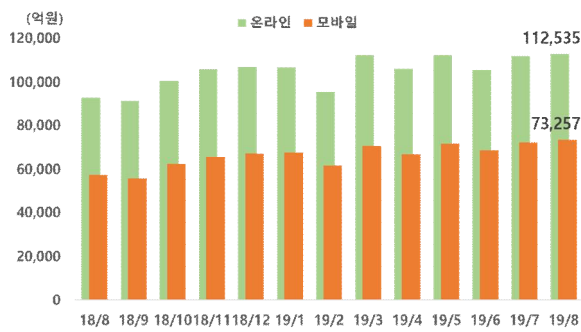


Fig. 1. Trend of Online Shopping Volume

온라인 쇼핑 시장의 성장과 함께 온라인 리뷰의 중요성 또한 증가하고 있다. 시장조사기업 엠브레인 트렌드모니터의 조사에 따르면 전체 응답자의 86.9%가 소비자 리뷰의 필요성에 동의했으며, 70%의 응답자는 소비자 리뷰를 신뢰한다고 답했다. 소비자 리뷰는 소비자 구매의향에 영향을 미치는 주요 요소이며, 구매 과정의 일부로서 작용되고 있다. 이렇듯 소비자 리뷰의 중요성이 증대함에 따라 이와 관련된 연구도 지속적으로 수행되고 있다.

관련 연구의 대표적인 주제로 고객 리뷰를 활용한 감성 분석(Sentiment Analysis)이 있다. 오피니언 마이닝(Opinion Mining)으로도 지칭되는 감성 분석은 제품, 서비스 등에 대한 사람들의 의견, 감성, 평가 및 태도를 전산적으로 분석하는 연구이다[1-6]. 온라인 리뷰는 제품, 서비스에 대한 정성적인 평가가 담겨있는 텍스트 데이터와 평점이라는 정량적 데이터로 구성되는 경우가 많아, 이 두 가지 데이터를 조합하여 분석하는 감성 연구가 활발히 수행되어 왔다. 특히 정량적 데이터인 평점을 활용해 리뷰 텍스트의 감성을 분류하는 감성 문서 분류는 다양한 도메인에서 활발하게 활용되었다.

감성 분석 연구의 초기에는 감성 사전을 구축하여 어휘 기반으로 감성을 파악하는 방식의 연구가 주로 진행되었으나, 도메인별 특성에 따른 성능 차이와 사전 구축에 소

요되는 시간 및 비용 문제로 인해 최근에는 기계학습 기반의 감성 분석이 주를 이루고 있다. 특히, 기계학습 기반 방법론 중에서도 딥러닝의 적용을 통해 좋은 성과를 거둘 수 있다는 사실이 입증되며 다양한 딥러닝 모델을 적용해 분류 성능을 향상시키고자 하는 노력이 이뤄지고 있다[5]. 딥러닝 기반 감성 분석은 사전 구축 등의 작업을 필요로 하지 않으며 딥러닝 모델은 언어적 특징을 데이터로부터 스스로 학습이 가능하다는 특징을 가져 작업할 수 있다는 장점을 갖는다[7-8]. 이로 인해 Convolution Neural Network를 비롯해 Bi-LSTM모델, 최근 각광받고 있는 Attention 메커니즘 등 다양한 딥러닝 알고리즘을 감성 분석에 적용하는 연구가 활발하게 이루어지고 있다.

그러나 이러한 기존의 알고리즘 기반 접근 방법론은 리뷰의 어휘적 특징은 반영하지만, 리뷰를 작성한 리뷰어의 특성은 전혀 반영하지 않는다는 한계를 지니고 있다. 즉 유사한 어휘를 활용하여 리뷰를 작성하더라도 각 리뷰에 담긴 어휘의 감성 수준은 리뷰어에 따라 상이할 수 있음에도 불구하고 기존 연구는 이러한 측면을 분석에 반영하지 못했다.



Fig. 2. Differences in the Emotions of Terms by Reviewers

<그림 2>는 실제 연구 과정에서 획득한 사례의 일부로, 두 리뷰가 서로 매우 유사한 어휘로 구성되었음에도 불구하고 두 리뷰의 평점이 상이하게 나타난 경우를 보이고 있다. 이러한 현상은 동일한 어휘를 활용하여 리뷰를 작성하더라도 리뷰어에 따라 다른 평가를 하고 있음을 증명한다. 이는 리뷰어의 특성을 동일하게 간주하고 어휘의 통계적 분포에만 기반을 둔 감성 분석의 한계를 입증하며, 동시에 리뷰어의 특성을 반영한 리뷰 감성 분석이 이뤄져야 할 필요성을 나타낸다. 실제로 기존 연구에서도 리뷰어의 특성은 리뷰를 읽는 소비자의 구매 의향에 영향을 미치는 요소로 밝혀진 바 있으며[9], 리뷰어에 대한 정보 습득 여부가 리뷰의 유용성에 영향을 미침을 확인한 연구도 수행된 바 있다.

리뷰어의 특성을 반영하여 리뷰 감성 분석을 수행하기 위해서는 리뷰어에 대한 정보가 추가로 요구된다. 하지만

리뷰어에 대한 신상 정보는 대부분 개인 정보로 분류되어 수집이 어려우며, '파워블로거'와 같이 특정 매체가 제공하는 리뷰어에 대한 평판 정보는 일부 리뷰어들에게만 제공되는 제한적인 정보라는 한계를 갖는다. 이러한 측면에서 고객 특성에 대한 정보를 담고 있으면서도 일부가 아닌 모든 리뷰어의 정보를 수집하기에 용이한 정보로 고객 등급이 있다. 쇼핑물의 고객 등급은 쇼핑물에서 정한 기준에 따라 구매 내역 및 리뷰 작성 활동 등에 기반을 두어 책정되며, 일반적으로 쇼핑물 사용 기간이 오래되고 구매 건수가 많아질수록 고객 등급이 높아진다. 따라서 고객 등급이 높을수록 쇼핑물에 대한 활용도 및 이해도가 상대적으로 높으며, 이러한 리뷰어가 작성하는 리뷰는 다른 고객에게 더욱 유용한 정보를 제공할 수 있을 것이라 기대할 수 있다.

이에 본 연구에서는 쇼핑물의 고객 등급을 쇼핑물 사용도로 정의하고, 쇼핑물 사용도라는 리뷰어의 집단적 특성을 반영한 감성 분석을 진행하고자 한다. 구체적으로는 리뷰어를 쇼핑물 사용도에 따라 두 집단으로 구분하고 각 집단별 리뷰의 평점을 예측하는 모델을 생성하고자 한다. 또한 리뷰어 특성에 따라 사용하는 어휘의 차이가 있는지를 확인하기 위해, 쇼핑물 사용도에 따라 구분한 각 집단별 긍정 및 부정 키워드를 도출하여 비교하고자 한다.

본 논문의 구성은 다음과 같다. 다음 2장에서는 감성 분석과 관련된 선행 연구들을 살펴보고, 3장에서는 본 연구의 개요 및 제안 방법론을 제시한다. 이후 4장에서는 실제 쇼핑물 데이터에 제안 방법론을 적용한 실험 결과를 분석하며, 마지막으로 5장에서 본 연구의 기여 및 한계를 요약한다.

II. Related Works

2.1 Influence of Online Reviews

온라인 리뷰는 제품이나 서비스의 소비에 대해 소비 체험 감각 등을 언급한 정보이다[9]. 온라인 리뷰는 엠브레인 트렌드 모니터의 조사 결과에서 86.9%의 소비자가 그 필요성에 동의할 만큼 구매에 있어 이미 큰 영향력을 행사하고 있다. 심지어 소비자가 오프라인으로 제품을 구매하는 경우에도 온라인 리뷰를 참고하기 때문에, 온라인 리뷰는 온라인 구매뿐 아니라 오프라인 구매에도 큰 영향을 미치고 있다고 할 수 있다[10]. 이에 온라인 리뷰가 구매에 미치는 영향에 대한 연구가 지속되어 왔고, 영향력을 구성하는 세부 요소에 대해서도 다양한 측면의 연구가 이뤄지고 있다. 우선 리뷰의 메시지 특성이 구매 의향에 미치는 연구가 진행되었는데, 리뷰의 가독성, 길이, 그리고 별점이 리뷰의 유용성 평가에 중

요한 영향을 미침이 입증되었다[11]. 이외에도 <그림 3>과 같이 리뷰 자체의 특성 외에도 제품의 특성, 리뷰 수용자의 특성, 리뷰어의 특성, 그리고 리뷰 사이트의 특성 등 또한 리뷰의 영향력을 결정하는 요소임이 입증되었다[12].

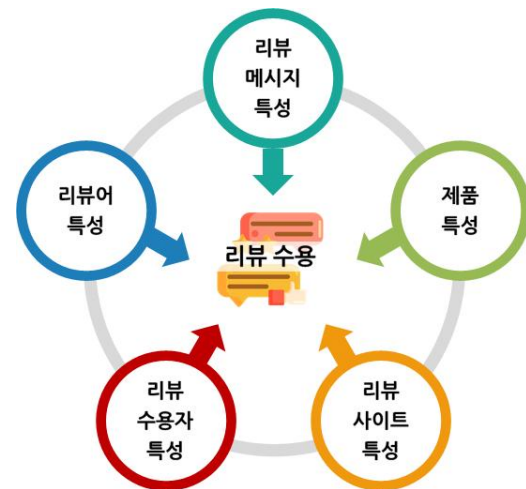


Fig. 3. Review Acceptance Framework

이러한 다양한 요인 중 리뷰어의 특성의 경우 리뷰어에 대한 신뢰도가 향상될 때 리뷰의 유용성이 향상됨을 보여주었는데, 이를 통해 정보의 원천인 리뷰어에 대한 정보가 제공되는가의 여부가 리뷰 수용에 중요한 영향을 미친다는 사실을 알 수 있다. 이러한 사실은 리뷰어 특성이 온라인 리뷰와 그 유용성에 미치는 영향에 대해 고찰한 <표 1>의 선행 연구들을 통해 더욱 자세히 확인할 수 있다[12]. <표 1>에 나타난 바와 같이 이와 같이 리뷰어의 정보는 리뷰의 수용에 영향을 미치는 주요 요소이다. 뿐만 아니라 리뷰어에 따라 제품에 대한 전문성이 달라지기 때문에 리뷰의 내용적 측면에서도 차이가 발생한다고 볼 수 있다. 따라서 리뷰의 내용을 보다 정확하게 분석하기 위해서는 분석 과정에서 리뷰어의 특성을 반영하는 것이 반드시 필요함을 암시한다.

Table 1. Previous Researches on Reviewer's Characteristics[12]

Source	Result
Racherla, et al. (2012) [13]	Online reviews from reputable authors are perceived to be more useful to consumers
Briggs, et al. (2002) [14]	Reliability of information source is the most important factor influencing the adoption of advice online
Forman, et al. (2008) [15]	The author's review is more useful when revealing the reviewer's personal information, such as the real name and location of residence
Josang, et al. (2007) [16]	Trust in sources is more important in online reviews than in offline reviews

2.2 Sentiment Analysis

감성 분석(Sentiment Analysis)은 사람들의 의견, 태도, 그리고 감성을 분석하는 연구이다[1-6]. 감성 분석은 크게 어휘 기반(Lexicon-Based) 방법론과 기계학습 기반(Machine Learning-Based) 방법론으로 나뉜다. <그림 4>는 감성 분석 방법론의 주요 개요이다[17].

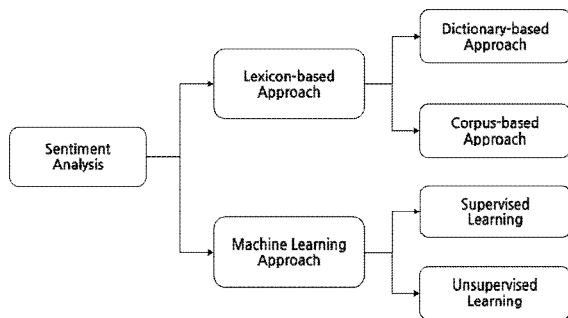


Fig. 4. Sentiment Analysis Methodology[17]

어휘 기반 방법론은 어휘에 대한 사전을 미리 구축하고 이를 활용하여 텍스트를 분석하는 방법이다. 기존의 사전을 재활용하거나, 분석하고자 하는 문서로부터 새로 사전을 구축하여 연구를 진행할 수 있다. 하지만 이 방법론은 사전 구축에 있어 비용 및 시간 소모가 많이 요구되고, 적용되는 도메인에 따라 분석 성능의 차이가 크게 나타난다는 단점을 갖는다. 이에 도메인 지식을 반영하여 감성 분석의 성능을 향상시키려는 연구가 이루어져왔다[18-20]. 다른 한 축인 기계학습 방법론은 크게 지도학습 방법론과 비지도학습 방법론으로 구분된다. 종속변수 값의 존재 여부로 지도학습과 비지도학습으로 구분되며, 감성 문서 분류에는 주로 지도학습 방법론이 사용되고 있다[2]. 지도학습에는 의사결정나무, 선형 분류 기법인 지지벡터 분류기법, 신경망 기법, 규칙기반 분류기법, 확률 기반 분류 기법인 나이브베이즈, 최대 엔트로피 모델 등 기존의 주요 분류 알고리즘들이 사용될 수 있다. 이와 같은 다양한 알고리즘의 적용 및 개선 등을 통해 감성 분류의 성능을 향상시키고자 하는 연구가 활발하게 이루어지고 있다.

2.3 Deep-learning Based Sentiment Analysis

최근 딥러닝 기법이 자연어 처리에 우수한 성능을 보임이 입증됨에 따라, 자연어 처리 분야에서도 딥러닝 기반 연구에 대한 관심이 증가하고 있다[21]. 딥러닝을 활용한 방법론은 데이터로부터 언어적 특성을 스스로 학습이 가능하다는 장점이 있으며, 일반적으로 기존 알고리즘 대비 우수한 성능을 보였다. 기존 텍스트 마이닝 연구의 난제 중 하나로 특성 추출의 이슈를 들 수 있다. 텍스트의 경우 데이터의 구조가 희

소(Sparse)하고, 차원의 수가 크다는 특징이 있어, 이를 효율적으로 처리하기 위해 특성을 추출하는 방안이 필요했기 때문이다. 이러한 이슈에 대한 해결 방안 중 가장 두각을 나타낸 연구로 단어의 특성을 저차원 실수 벡터값으로 나타내어 해결한 워드임베딩(Word Embedding) 연구[7]를 들 수 있다. Milokov가 제안한 워드임베딩의 대표적인 방법론인 Word2Vec은 학습 간 사용된 문서의 문맥 정보를 활용하며 각 단어를 의미 벡터로 구조화한다. Word2Vec을 통해 추출된 단어 벡터를 통해 단어의 의미적/문법적 특징을 벡터 공간 상에 나타낼 수 있으며, 이를 활용하여 유의어 도출, 단어 유추 등의 응용을 수행할 수 있다.

워드 임베딩 연구의 성과에 힘입어 자연어 처리가 용이해짐에 따라, 기존에 이미 제시되었으나 연산 처리 등의 문제로 실효성이 낮은 것으로 평가되던 신경망처리 알고리즘들이 자연어 처리 분야에서 다시 주목받게 되었으며, 이는 감성 분석에도 적용되어 그 효과를 입증했다. Socher et al.(2011)[8]은 언어의 순차적인 성격을 포착할 수 있는 Recurrent Neural Network(RNN)을 활용한 연구를 수행한 바 있으며, 이후 RNN의 장기기억 감소 문제를 해결한 Long Short Term Memory(LSTM) 연구, 그리고 각기 다른 방향의 두 LSTM 층을 붙인 Bi-directional LSTM 등 알고리즘의 개선을 통해 감성 분류의 정확도를 높이려는 노력이 꾸준히 지속됐다. RNN 뿐만 아니라 기존 이미지 처리 분류 분야에서 우수한 성능을 보인 Convolution Neural Network(CNN)을 감성 분석에 도입한 연구[21]도 우수한 성능을 나타낸 바 있다.

이처럼 다양한 알고리즘의 등장이 지속된 가운데 최근에는 기계 번역에서 우수한 성능을 보인 Attention Mechanism을 도입한 연구들이 각광을 받고 있다. 기존의 LSTM을 활용한 연구는 딥러닝 연구에 있어 State of Art의 위치를 차지하고 있음에도 불구하고 일부 한계를 갖고 있기 때문이다. LSTM과 같은 RNN 계열 모델들은 고정된 크기로 입력 데이터를 인코딩하므로, 입력 데이터 간 크기의 차이가 큰 경우 성능이 저하되는 현상을 공통적으로 나타낸다. 한편 Attention Mechanism은 RNN의 각 스텝의 출력을 입력으로 받고 가중치의 합을 계산하여, 각 스텝의 주목(Attention) 정도를 확인한다[22]. 즉 하나의 고정된 컨텍스트 벡터가 아닌 각 단계별 컨텍스트 벡터를 생성하며, 이로 인해 입력 데이터의 크기에 차이가 있더라도 정보 손실을 방지할 수 있다. 이러한 특성으로 인해 Attention Mechanism은 다양한 딥러닝 기법 중 가장 뛰어난 성능을 보였으며 감성 분석에서도 우수한 성능을 보였다. <그림 5>는 Attention Mechanism의 구조를 나타내며, 각 단계별로 컨텍스트 벡터를 생성하는 과정을 보여주고 있다.

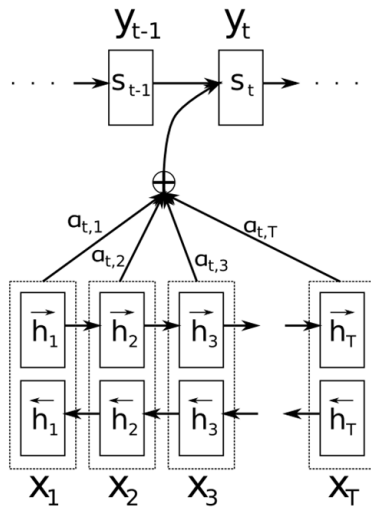


Fig. 5. Attention Mechanism[22]

우수한 성능으로 인해 Attention Mechanism이 주목받음에 따라, 이를 활용하여 변형한 알고리즘도 다수 출현하고 있다. 그 중 대표적인 것으로 문서의 계층적 구조를 활용한 Hierarchical Attention Network(HAN) 모델[23]은 특히 문서 분류 응용에서 우수한 성능을 보였다. HAN 모델은 문서가 문장들의 조합으로 이뤄지고, 문장은 단어들의 조합으로 이뤄지는 계층적 구조를 갖는다는 것에 주목하며, 기존 Attention Mechanism에 계층 구조를 반영하여 2단계로 구성된다. 우선 단어를 처리하는 Word Encoder가 첫 층을 형성하며, 첫 층의 결과가 문장을 처리하는 Sentence Encoder로 전달된다. HAN의 개략적인 구조는 <그림 6>과 같다.

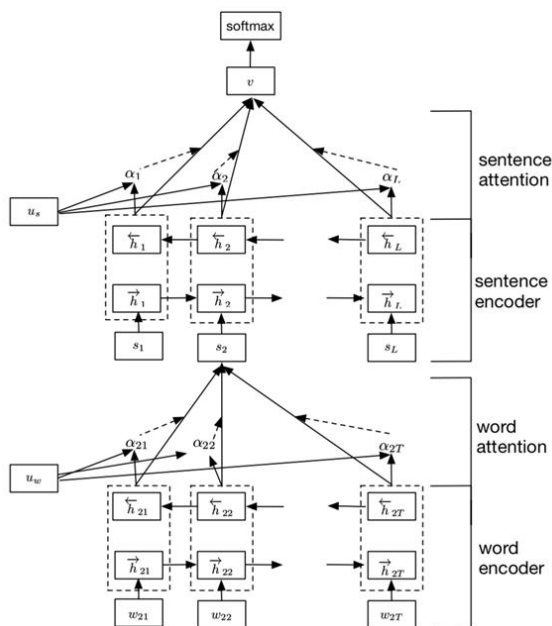


Fig. 6. Hierarchical Attention Network[23]

Attention Mechanism 기반 방법론은 기존 방법론 대비 감성 문서 분류에서 높은 성능을 나타내었으며, 각 단어의 Attention Weight 계산이 가능해 어떤 단어가 분류에 중요하게 작용되었는지를 확인할 수 있다. 이후 딥러닝 기반 알고리즘의 연구는 학습을 위한 데이터 확보와 학습에 많은 시간이 소모된다는 한계를 해결하기 위해 사전에 학습한 언어 모델을 적용하고 분야에 따라 미세 조정(Fine Tuning)을 진행하는 연구들이 이뤄지고 있다. 신경망 기반 사전 학습 언어 모델의 대표적인 사례로는 ELMo(Embeddings from Language Model)[24]가 있다. ELMo는 Bi-LSTM을 기반으로 하여 임베딩을 맥락 정보에 따라 가변화할 수 있게 하였다. 2018년 구글에서 제시된 BERT(Bidirectional Encoder Representations from Transformer)[25] 또한 양방향 학습 언어 모델을 기반으로 제시되었으나 앞서 언급한 Attention Mechanism을 활용해 ELMo의 신호 전달 강도 한계 문제를 해결하였다. 가장 최근에는 XLNet(eXtra Long Network)[26]은 장문 텍스트에 대한 문제도 해결해내어 최근 텍스트 기반 딥러닝 연구에서 각광을 받고 있다.

본 연구에서는 최근 방법론 중 HAN 방법론을 활용하여 감성 분류를 수행하여, 쇼핑물의 고객 등급에 따라 리뷰에서 나타나는 긍정 및 부정 어휘에 어떤 차이가 존재하는지 확인하고자 한다.

III. Methodology

3.1 Research Overview

본 연구에서는 활용 가능한 리뷰어의 특성 정보를 소개하고, 리뷰어의 특성을 반영한 감성 분석 및 주요 키워드 비교 방안을 제시한다. 이를 위한 본 연구의 절차는 <그림 7>과 같다. 본 연구에서는 전술한 바와 같이 쇼핑물 고객 등급을 리뷰어의 특성 정보로 사용한다. 기존 연구에서 리뷰 수용에 영향을 미치는 특성으로 제시된 리뷰어의 개인 정보 등 타 특성들은 일부 리뷰어에 대해서만 공개되어 있는 반면, 쇼핑물의 고객 등급은 모든 리뷰어에 대해 확인이 가능하다는 장점을 갖는다. 또한 쇼핑물의 고객 등급은 쇼핑물에서의 구매 활동 및 리뷰 작성 내역에 따라 등급이 정해지므로, 고객의 등급은 쇼핑물 사용도 및 충성도를 간접적으로 반영한다고 볼 수 있다. 즉 쇼핑물의 고객 등급이 높을수록 쇼핑물에 더욱 친숙하고 높은 활용도를 가진 고객이라고 간주할 수 있다. 따라서 고객의 등급 정보를 활용하여 리뷰어의 특성을 반영한 감성 문서 분류와 리뷰어 고객군에 따른 키워드 비교 연구를 진행하고자 한다.

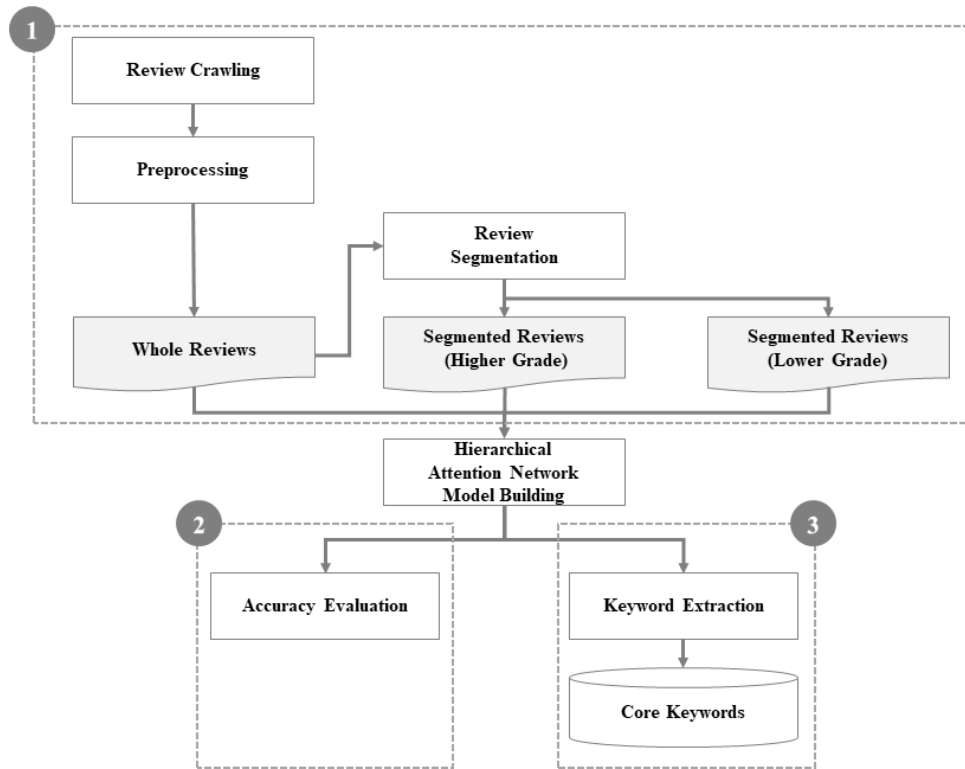


Fig. 7. Research Process

3.2 Data Crawling and Preprocessing

연구 진행의 첫 단계는 데이터 수집 및 전처리이다. 분석을 진행할 쇼핑몰로는 모든 리뷰어의 등급이 공개되는 ‘M’을 선정하였다. 즉 ‘M’사의 단일 제품군의 제품 리뷰를 수집하면서, 이와 동시에 해당 리뷰를 작성한 고객의 쇼핑몰 고객등급을 수집하였다. 추후 키워드 분석 및 감성 문서 분류 분석의 품질 향상을 위해 수집한 리뷰에 대해 전처리 작업을 진행하며, 전처리 단계에서는 조사 및 불용어 등이 제거된다.

3.3 Sentiment Classification

본 단계에서는 사용자별 리뷰에 대해 감성 문서 분류를 수행한다. 즉 평점에 따라 긍정 리뷰와 부정 리뷰를 구분하고, 리뷰 원문으로부터 긍정 및 부정 여부를 예측하는 모델을 구축한다. 다만 본 연구에서는 고객 특성에 따른 리뷰 감성의 차이를 비교하기 위해, 분류 모델 구축에 앞서 고객 등급에 따라 리뷰를 구분한다. 즉 쇼핑몰의 고객 등급에 따라 등급이 높은 고객군의 리뷰(H)와 낮은 고객군의 리뷰(L)로 구분하며, 비교 실험을 위해 이들 데이터를 통합한 전체 고객군의 리뷰(T)도 별도로 관리한다. 이상의 세 가지 리뷰 집합인 T, H, L 각각에 대해 독립적으로 학습을 수행하며, 각 학습에 사용되는 훈련용 데이터(Train Data), 검증용 데이터(Validation Data), 그리고 평가용 데이터(Test Data)의 비율은 6:2:2로 설정한다. 학습은

HAN 모델을 사용하여 수행하며, 문서 분류 학습 과정에서 산출되는 각 단어의 Attention Weight를 통해 각 고객군의 긍정 및 부정 키워드를 파악할 수 있다.

IV. Experiments

4.1 Overview

본 연구의 실험은 남성 의류 쇼핑몰 ‘M’사의 제품 리뷰 데이터에 대해 수행하였다. 여러 제품군의 리뷰를 동시에 사용하는 경우 제품에 따라 리뷰의 어휘가 상이하게 나타날 수 있으므로, 고객에 초점을 두어 어휘 비교를 진행하기 위해 단일 제품군의 리뷰만을 분석에 사용하였다. 즉 단일 제품군인 니트에 대한 리뷰와 5점 척도 기준의 평점 4,780건을 수집하였으며, 해당 리뷰를 작성한 고객의 쇼핑몰 고객 등급을 함께 수집하였다. 해당 쇼핑몰 내 니트 제품군에서 리뷰 수 기준 상위 25개 제품들의 리뷰를 수집하였으며 수집기간은 2015.12.23. ~ 2019.11.11.이다. 수집 대상이 된 제품들의 평균 평점이 약 4점으로 나타났으므로, 긍정 평점과 부정 평점을 구분하는 기준으로 4점을 적용하였다. 구체적으로는 평점이 4점인 데이터는 중립 데이터로 간주하여 분석에서 제외하였으며, 평점 1-3점의 데이터는 부정 데이터로, 5점은 긍정 데이터로 정의하였다. 한편 해당 쇼핑몰의

고객 등급은 총 8단계로 관리되고 있으므로, Lv. 1-4 의 리뷰어는 쇼핑몰 사용도 低(저) 고객군으로, Lv. 5-8 의 고객군은 쇼핑몰 사용도 高(고) 고객군으로 정의하였다. 총 6만여개의 리뷰를 수집하였으나 분석 결과가 데이터 수의 차이에 영향을 받음을 최소화하기 위해, 고객군별, 별점별 데이터의 수를 <표 2>와 같이 동일하게 유지하였다. 약 6만개의 리뷰 중 고객군별 / 별점별 총 4개의 세부 구분상 가장 최소 리뷰수를 가진 사용도 高(고) 고객군의 부정 리뷰 수인 1,195개에 맞추어 개수를 통일하기 위해, 다른 세부 구분상의 리뷰들은 최신성을 기준으로 샘플링하였다.

Table 2. Data Segments

Group	Rating		Total
	1-3	5	
Lower Group (Lv. 1-4)	1,195	1,195	2,390
Higher Group (Lv. 5-8)	1,195	1,195	2,390
Total	2,390	2,390	4,780

4.2 Experiment Design

앞서 소개한 바와 같이 본 실험에 앞서 수집 데이터에 대한 전처리를 진행한다. 비정형 텍스트에 대한 분석은 일반적으로 형태소 분석 단계를 수반하며, 형태소 분석은 감성 분석의 정확도에도 영향을 크게 미친다[5]. 딥러닝 기법을 활용한 분석에서도 형태소 분석기를 활용하는 경우가 많으며, 특히 한국어의 경우 타 언어에 비해 형태소 분석의 중요성이 강조되고 있다[27]. 기존 HAN연구의 경우 영문 데이터를 활용하며 각 문맥에 따른 의미의 중요성을 분석하기 위해 최소한의 전처리를 진행하나, 본 연구의 실험에서는 한국어 데이터를 사용하기 때문에 오타 수정과 같은 기본적인 전처리 수행 후 본격적인 형태소 분석을 진행하였다. 형태소 분석은 Python 환경에서 Konlpy 라이브러리의 Komoran 형태소 분석기를 사용하였다. 분류 모델로 각 문단에서 키워드의 의미를 해석하는 것이 주요 연구 목표 중 하나이므로, 형태소 분석 결과로 획득한 토큰 중 주요 의미를 담고 있는 명사, 동사, 그리고 형용사만을 이후 분석에 사용하였다.

4.2.1 Sentiment Classification

전처리된 문서를 대상으로 하여 감성 문서 분류 실험을 진행하였다. 전체 리뷰와 쇼핑몰 사용도에 따라 구분한 리뷰, 즉 T, H, 그리고 L의 세 개의 리뷰 집합에 대해 각각 HAN 모델을 사용하여 분류 분석을 수행하였으며, 분석에는 Python의 Keras 라이브러리를 활용하였다. 모델의

Batch Size는 16으로, 전체 학습 Epoch 또한 16으로 설정하였으며, 각 용어 및 문장의 임베딩에는 Bidirectional GRU를 사용하였다. 앞서 언급한 바와 같이 각 데이터 셋을 학습용/검증용/평가용 목적으로 6:2:2의 비율로 나누어 분석을 진행하였다. 따라서 세 가지 데이터에 대해 각각 분류 정확도가 산출되며, 이를 통해 전체 리뷰에 대한 분류 정확도, 쇼핑몰 사용도 低(저) 고객군의 리뷰에 대한 분류 정확도, 쇼핑몰 사용도 高(고) 고객군의 리뷰에 대한 분류 정확도를 비교할 수 있다.

4.2.2 Keyword Comparison

구축된 Hierarchical Attention Network 모델을 기반으로 각 리뷰의 단어 및 문장에 대한 분류 간 가중치를 확인한다. 리뷰별로 계산되는 단어 가중치와 문장 가중치를 곱한 값을 각 단어의 가중치로 산출한다. 단, 각 리뷰에서 동일한 단어가 반복되어 나타나는 경우 해당 단어의 가중치를 합산한 것을 최종 가중치로 사용한다. 리뷰별로 최종 가중치를 Sorting한 후 상위 5개의 Keyword를 Counting 한다. 쇼핑몰 사용도에 따른 고객군별로 주요 상위 키워드에 차이가 있는지를 확인/비교한다. 우선 전체 단위에서 키워드 간 특징을 1차로 비교해보며, 다음으로는 해당 리뷰의 긍정/부정 감성에 따라 긍정/부정 단어를 나누어 비교 분석해보는 단계를 거치며, 이 과정은 <그림 8>에 요약되어 있다.

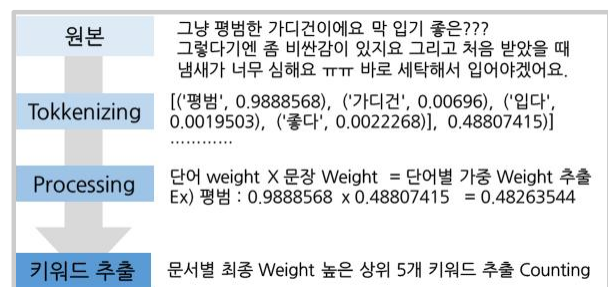


Fig. 8. Keyword processing process

4.3 Experiment Result

4.3.1 Sentiment Document Classification

HAN 모델에 의해 감성 문서 분류 분석을 수행한 결과가 <그림 9>에 제시되어 있다. <그림 9>는 검증을 위한 Test 데이터를 대상으로 하여 각 고객군별로 생성된 문서 분류 모델을 활용해 문서 분류를 시행한 결과이다. <그림 9>에서 정확도는 긍정/부정 문서의 분류에 대한 정확도, 각 문서의 긍정/부정 여부를 얼마나 정확하게 분류하였는지에 대한 성능을 나타낸다. T, H, L 데이터의 검증 정확도는 각각 78.35%, 78.66%, 그리고 76.57%로 데이터간 정확도 차이는 크게 나타나지 않는 것으로 파악되었다. 다만 근소한 차

이이지만 L 데이터, 즉 쇼핑물 사용도 低(저) 고객군의 리뷰에 대한 분류 분석의 정확도가 다른 경우에 비해 다소 낮게 나타난 것에 주목할 필요가 있다. 이는 등급이 낮은 고객의 리뷰에서는 리뷰 어휘와 평점의 관계, 즉 리뷰로부터 평점을 예측할 수 있는 연결성이 높은 등급 고객의 리뷰에 비해 약하게 나타나기 때문인 것으로 해석할 수 있다.

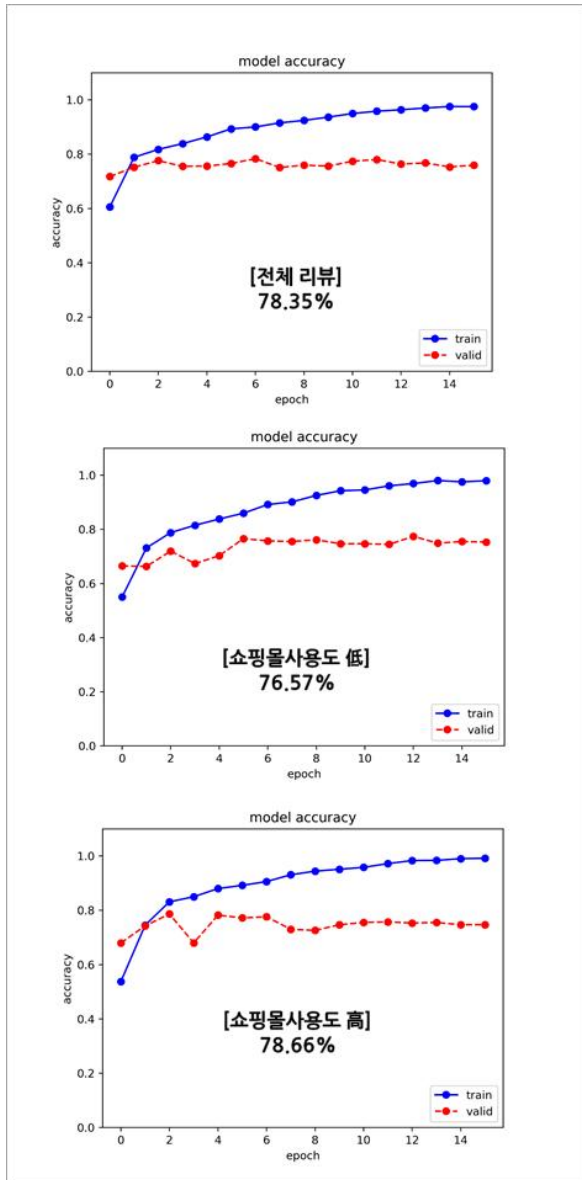


Fig. 9. Accuracy of Sentiment Classification

4.3.2 Keyword Comparison

본 연구의 주요 산출물은 리뷰의 주요 키워드, 구체적으로는 리뷰를 긍정 및 부정으로 구분하는 HAN 모델에서 중요하게 작용한 어휘들의 집합이다. 즉 <그림 9>에서 사용된 예측 모델은 예측 자체를 목적으로 하는 것이 아니라 예측 과정에서 핵심 키워드를 도출하는 것을 목적으로 한

다. <표 3>은 각 문서의 주요 키워드를 1위부터 5위까지 선정한 것으로, 해당 문서를 긍정 또는 부정으로 구분하는 모델에서 Attention Weight가 높은 상위 단어 5개를 발췌한 것이다. 예를 들어 문서 0의 경우 핵심 단어가 ‘심플’, ‘똑같다’, 그리고 ‘기본’이고, 문서 1의 경우 핵심 단어가 ‘초겨울’, ‘덥다’, 그리고 ‘필요’임을 알 수 있다.

Table 3. Core Keywords of Each Document

Doc. ID	1	2	3	4	5
0	심플	똑같다	기본	이랑	사진
1	초겨울	덥다	필요	기준	영덩이
2	변신	기존	평범	먼지	많다
3	관리	차다	생기다	걱정	않다
4	끝다	비싸다	까칠	아쉽다	감
...

위의 <표 3>으로부터 각 고객군별 리뷰의 핵심 키워드를 도출할 수 있으며, 본 실험에서는 특정 단어가 각 문서의 Top 5 키워드 내에 포함된 빈도를 그 기준으로 적용한다. 예를 들어 쇼핑물 사용도 高(고) 고객군의 리뷰는 총 2,390건이 사용되었으므로, 각 문서마다 <표 3>과 같이 Attention Weight 기준 Top 5 키워드를 도출하면 총 11,950개의 단어(중복 포함)로 구성된 표를 구할 수 있다.

일부 단어는 여러 문서에서 Top 5 키워드에 포함되어 나타날 수 있으며, 실제 분석에서 ‘좋다’, ‘이쁘다’, 그리고 ‘입다’라는 단어는 각각 총 181개, 138개, 그리고 137개의 문서에서 Top 5 키워드에 포함되어 나타났다. 이러한 방식으로 쇼핑물 사용도 高(고) 고객군과 쇼핑물 사용도 低(저) 고객군 리뷰의 주요 키워드를 도출할 수 있으며, 각 고객군별로 주요 키워드 40개씩을 선정한 결과가 <표 4>와 <표 5>에 각각 나타나있다.

세부 내용적인 측면에서도 <표 4>와 <표 5>에 나타난 두 고객군별 주요 키워드를 비교하면, ‘좋다’와 ‘이쁘다’ 등과 같이 두 집단 모두에서 주요 키워드로 선정된 것도 있지만, ‘가성비’와 ‘대비’(High Level), ‘색’과 ‘사진’(Low Level) 등과 같이 한 고객군에서만 주요 키워드로 선정된 것도 있다. 두 고객군간 어휘의 차이를 살펴보기 위해, <표 4>와 <표 5>에서 공통적으로 나타난 어휘를 제거한 결과가 <표 6>에 제시되어 있다.

Table 4. Top 40 Keywords of High Level Reviewer

Keywords in Higher Group					
#	Keyword	Count	#	Keyword	Count
1	좋다	181	21	나오다	65
2	이쁘다	138	22	싸다	65
3	입다	137	23	하다	64
4	가격	125	24	사다	64
5	생각	122	25	다니다	63
6	핏	108	26	구매	62
7	같다	107	27	목	60
8	사이즈	105	28	가성비	58
9	재질	102	29	디자인	57
10	색감	98	30	마음	56
11	무난	97	31	기장	56
12	얇다	85	32	예쁘다	54
13	오버	82	33	두껍다	52
14	편찮다	80	34	안	51
15	이너	80	35	편하다	50
16	만족	80	36	따뜻	50
17	옷	77	37	적당	50
18	있다	76	38	크다	49
19	맞다	74	39	겨울	48
20	가디건	71	40	감	47

Table 5. Top 40 Keywords of Low Level Reviewer

Keywords in Lower Group					
#	Keyword	Count	#	Keyword	Count
1	이쁘다	168	21	두껍다	67
2	생각	143	22	옷	65
3	입다	127	23	가격	64
4	좋다	123	24	사다	64
5	색감	120	25	사진	63
6	예쁘다	109	26	길다	63
7	핏	103	27	무난	63
8	같다	93	28	감	61
9	사이즈	92	29	무겁다	57
10	크다	87	30	이너	55
11	색	84	31	따뜻	55
12	재질	83	32	작다	55
13	가디건	81	33	맘	55
14	있다	81	34	디자인	52
15	아쉽다	76	35	귀엽다	52
16	오버	75	36	하다	51
17	들다	73	37	적당	50
18	만족	71	38	니트	50
19	편찮다	68	39	맞다	48
20	얇다	68	40	마음	48

<표 6>을 통해 두 고객군별 키워드를 보다 명확하게 비교할 수 있다. 쇼핑물 사용도 고 고객군의 주요 키워드에 ‘가성비’, ‘대비’ 등의 키워드가 포함된 것으로 보아, 쇼핑물 사용도가 높은 고객일수록 가격에 관련된 어휘가 평가의 긍정 및 부정에 영향을 크게 미침을 확인할 수 있다. <표 4>와 <표 5>에서 쇼핑물 사용도 고 고객군에서는 ‘가격’이 4위 ‘싸다’가 22위인 반면 저 고객군에서는 ‘가격’이 23위, ‘싸다’는 48위에 그친 것도 이러한 해석을 뒷받침한다.

Table 6. Exclusive Keywords in Top 40 Keywords

Higher Group		Lower Group	
Keyword	Count	Keyword	Count
목	60	색	84
가성비	58	아쉽다	76
안	51	들다	73
편하다	50	사진	63
겨울	48	무겁다	57
플라	47	작다	55
살다	45	맘	55
대비	45	귀엽다	52
티	44	니트	50
굿	43	색깔	45
어울리다	43	두께	43
느낌	43	되다	43
색상	43	보풀	42

반면 쇼핑물 사용도 저 고객군의 경우 쇼핑물에서 소개된 제품과 실제 제품과의 대조를 중요한 요소로 여기는 것으로 나타났다. 이는 ‘색’, ‘사진’ 등의 키워드가 주요 키워드에 포함된 것을 통해 파악할 수 있으며, 실제로 이들 키워드가 포함된 원문을 조회한 결과 고객들이 쇼핑물에서 확인한 제품의 색과 실제 제품의 색이 일치하는지 여부에 대해 많이 언급한 것을 확인하였다.

또한 쇼핑물 사용도 고 고객군에서는 ‘작다’ 키워드가 Top 40 키워드에 포함되지 않은 반면, 저 고객군에서는 32위에 위치했고, ‘크다’의 키워드도 사용도 고 고객군은 38위인 반면 저 고객군에서는 10위에 위치했다. 이를 통해 제품을 직접 입어보고 구매할 수 없어서 많이 겪게 되는 의류 실착상의 문제가, 쇼핑물 사용도가 높은 고객군에 비해 낮은 고객군에서 주요 이슈로 작용하고 있음을 확인할 수 있다.

다음으로 긍정 리뷰와 부정 리뷰를 구분하여, 고객군별 어휘에 어떤 차이가 나타나는지 살펴보았다. 5점 리뷰에서 나타난 주요 키워드를 긍정 키워드로 1-3점의 리뷰에서 나타난 주요 키워드를 부정 키워드로 간주하였고, 이를 위의 방법과 동일하게 Top 40 키워드 중 한 고객군에서만 주요 키워드에 포함된 키워드만을 발췌하여 긍정 키워드를 <표 7>에, 부정 키워드를 <표 8>에 요약하였다.

쇼핑물 사용도 고 고객군에서는 긍/부정 키워드에서 모두 ‘가격’과 관련된 키워드가 주요 이슈로 확인되었다. 이들 키워드에 대한 원문 조회 결과 가성비가 우수한 경우 대한 긍정적인 평가가 많았으며, 부정 평가에서 나타난 ‘저렴’이란 키워드는 제품이 저렴하여 구매하였지만 제품을 구매한 이후 만족하지 못하는 상황에서 주로 사용됨을 확인할 수 있었다. 또한 ‘봄’, ‘겨울’과 같이 계절성에 관련 어휘가 다수 등장하였는데, 이를 통해 계절에 맞는 쇼핑을 지속적으로 진행하고 그에 따라 쇼핑물 사용도가 높아졌을 가능성을 유추할 수 있다.

Table 7. Comparison of Positive Keywords

Comparison of Positive Keywords			
Higher Group		Lower Group	
Keyword	Count	Keyword	Count
어울리다	36	귀엽다	51
나오다	35	사진	41
살다	35	무겁다	39
가격	34	색깔	34
감사	33	크다	32
굿	31	춥다	30
요즘	30	넉넉	30
하다	29	실물	29
봄	28	되다	28
겨울	27	처음	27
팔다	26	없다	26

Table 8. Comparison of Negative Keywords

Comparison of Negative Keywords			
Higher Group		Lower Group	
Keyword	Count	Keyword	Count
풀라	46	다르다	35
티	35	마감	32
구매	31	나다	25
느낌	27	일어나다	25
저렴	27	반품	25
생기다	26	심하다	25
빨다	26	들다	24
편하다	26	실망	22
맛	25	기본	22
두껍다	23	사진	22
질	22	이상	21

쇼핑몰 사용도 低고객군에서는 제품의 상태에 대한 이슈가 다수 나타났다. 긍정 리뷰에서는 사진과 실물이 일치하는 경우에 대한 만족이 주로 언급되었으며, 부정 리뷰에서는 제품의 마감 및 완성도에 대한 불만 관련 키워드가 주로 포함되었다. 이는 제품의 지속적인 구매 및 브랜드 신뢰 생성에 부정적인 영향을 줌으로써, 쇼핑몰의 지속적 사용을 저해하는 요소로 작용했음을 짐작할 수 있다.

전체 키워드에 대한 高(高) 고객군과 低(低) 고객군의 키워드 간 차이를 정량적으로 확인하기 위해 각 고객군에서 나타나는 주요 키워드의 순위 차이를 분석했다. <그림 10>은 高고객군 기준 긍정 키워드 상위 10개에 대해 高고객군과 低고객군에서의 순위를 비교한 것이며, <그림 11>은 高고객군 기준 부정 키워드 상위 10개에 대해 高고객군과 低고객군에서의 순위를 비교한 것이다. 긍정 키워드에서 '이쁘다', '핏' 등의 키워드는 두 고객군에서 동일하거나 유사한 순위를 기록한 반면, '입다', '사이즈' 등의 키워드는 각 고객군에서 기록한 순위가 큰 차이를 보인다. 이러한 현상은 부정 키워드에서도 동일하게 발견된다.

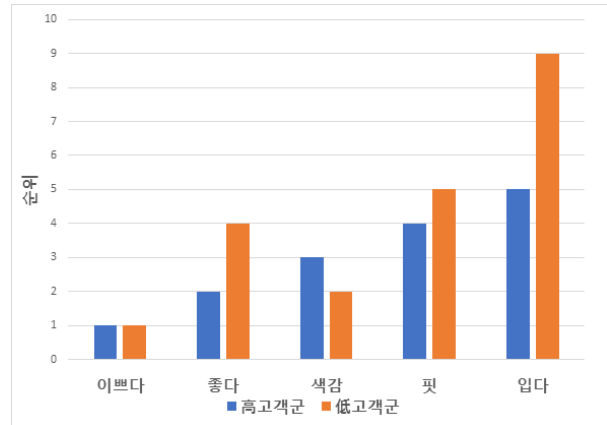


Fig. 10. Rank Comparison - Positive Keywords

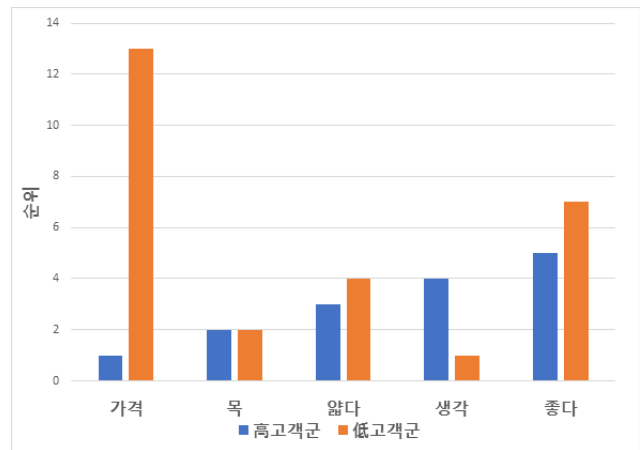


Fig. 11. Rank Comparison - Negative Keywords

각 고객군에서 나타나는 순위의 차이를 통계적으로 파악하기 위해 대상 키워드의 수를 변화시켜가며 키워드 순위 차이의 평균을 비교하여 <표 9>에 요약하였다. 우선 高고객군을 중심으로 상위 긍정 키워드 200개를 도출한 뒤, 이들 키워드의 低고객군에서의 순위를 기록하였다. 단 순위의 차이가 지나치게 크게 나타나는 일부 키워드로 인해 전체 평균이 영향을 받는 현상을 방지하기 위해, 低고객군 순위가 200위를 초과하는 키워드에 대해서는 일괄적으로 201위의 순위를 부여하였다. 다음으로 두 고객군에서의 순위차의 절대값을 구한 뒤, 이들 값의 평균을 분석하였다. 분석 키워드 수는 20개, 50개, 100개, 그리고 200개로 설정하였으며, 부정 키워드의 경우도 긍정 키워드와 동일한 방식으로 실험을 수행하였다. 실험 결과는 <표 9>에 요약되어 있으며, 실험 결과 긍정 키워드에 비해 부정 키워드에서 高고객군과 低고객군간 주요 키워드의 순위차가 크게 발생함을 확인하였다.

Table 9. Analysis of Rank Difference

	# Keywords			
	20	50	100	200
Positive	7.35	20.66	34.16	37.77
Negative	17.8	22.66	33.63	37.89

이상의 분석을 통해 쇼핑물 사용도가 높은 고객군과 낮은 고객군에서 제품의 긍/부정 평가에 중요하게 작용하는 키워드의 차이를 파악할 수 있었다. 이처럼 고객을 특성에 따라 구분하고 각 고객군별로 리뷰 분석을 수행함으로써, 각 고객군이 관심을 갖는 주요 이슈를 파악하고 해당 고객군에 특화된 마케팅 전략을 더욱 효과적으로 수립할 수 있을 것으로 기대한다.

V. Conclusion

본 연구는 리뷰어의 특성을 반영한 리뷰 분석 방법론을 제시하였다. 구체적으로는 고객 등급에 따라 리뷰를 구분하고, 각 리뷰 집합에 대해 Attention 기반 딥러닝 분류를 수행하여 고객군별로 긍/부정 리뷰에 영향을 미치는 주요 키워드의 차이를 확인하였다.

본 연구에서는 전통적인 마케팅적 도구 중 하나인 고객 세분화를 최근 활용도가 높아지고 있는 리뷰 감정 분석에 적용하는 방법론을 제안하였으며, 이는 본 연구의 학술적 기여로 인정받을 수 있다. 또한 결과는 고객군별로 차별화된 마케팅 전략을 수립하는 데에 효과적으로 활용될 수 있으며, 이러한 측면에서 본 연구의 실무적 기여를 찾을 수 있다.

본 연구의 차별점은 다음의 측면에서 찾을 수 있다. Attention을 사용한 다른 연구들이 분류 정확도 향상에 초점을 둔 반면 본 연구는 Attention 학습을 수행하는 과정에서 도출되는 각 단어의 가중치를 활용했다는 점에서 기존 연구들과 차별화된다. 특히 Attention 메커니즘의 특성을 활용하여 각 리뷰를 긍/부정으로 분류하는 데에 크게 기여한 단어를 식별할 수 있었으며, 이에 대한 집계를 통해 각 고객군이 제품에 대해 긍/부정을 평가하는데 영향을 주는 주요 요소를 확인할 수 있었다. 기존의 토픽 모델링 기반 긍정/부정 어휘 추출은 특성화에 기반을 둔 방법인 것에 비해, 본 연구에서 Attention 메커니즘 기반으로 추출한 긍정/부정 어휘는 차별화에 기반을 둔 방법이라는 점에서 새로운 시도로 인정받을 수 있다.

이러한 기여에도 불구하고, 본 연구는 고객 등급이라는 제한적인 정보만으로 고객군을 나누었다는 한계를 갖는다.

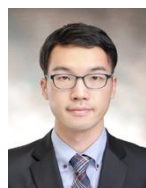
향후 더욱 다양한 측면에서 고객군을 식별하고, 본 연구에서 활용한 리뷰어의 특성 외에, 제품 특성, 리뷰 사이트 특성 등을 다양한 속성들을 활용한 연구가 후속 연구로 진행될 필요가 있다. 또한 본 연구의 최종 산출물로 제시한 고객군별 긍/부정 키워드에 대해서도 추가 분석이 필요하다. 물론 현재의 결과만으로도 두 집단간 차이 및 각 집단의 특성은 확인할 수 있지만, 구체적으로 어떤 측면에서 차이가 나타났는지에 대한 해석을 키워드 빈도에 기반을 두어 사후에 주관적으로 진행하였다. 향후 연구에서는 전체 키워드에 대한 유형을 사전에 구분하고 제안 방법론을 적용함으로써, 방법론의 적용을 통해 도출한 결과에 대한 해석을 보다 체계적으로 내릴 필요가 있다.

REFERENCES

- [1] J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin, "Learning Subjective Language," *Computational linguistics*, Vol. 30, No. 3, pp. 277-308, 2004.
- [2] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, 2008.
- [3] B. Liu, "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions," *The Cambridge University Press*, 2015.
- [4] L. Zhang, S. Wang, and B. Liu, "Deep Learning for Sentiment Analysis: A Survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 8, No. 4, e1253, 2018.
- [5] H. Park, M. Song, and K. Shin, "Sentiment Analysis of Korean Reviews Using CNN: Focusing on Morpheme Embedding," *Journal of intelligence and information systems*, Vol. 24, No. 2, pp. 59-83, 2018.
- [6] P. Jeong and H. Ahn, "A Sentiment Analysis Model for Customer Reviews Considering Sentence Location," *The Journal of Internet Electronic Commerce Research*, Vol. 19, No. 1, pp. 167-186, 2019.
- [7] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," *Advances in Neural Information Processing systems*, pp. 3111-3119, 2013.
- [8] R. Socher, C. C. Lin, C. Manning, and A. Y. Ng, "Parsing Natural Scenes and Natural Language with Recursive Neural Networks," *In Proceedings of the 28th international conference on machine learning(ICML-11)*, pp. 129-136, 2011.
- [9] P. Chatterjee, "Online Reviews: Do Consumers Use Them?," *Advance in Consumer Research*, Vol. 28, pp. 129-133, 2001.
- [10] F. Zhu and X. Zhang, "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of marketing*, Vol. 74, No. 2, pp.

- 133-148, 2010.
- [11] N. Korfiatis, E. G. Bariocanal, and S. S. Alonso, "Evaluating Content Quality and Helpfulness of Online Product Reviews: The interplay of review helpfulness vs. review content," *Electronic Commerce Research and Applications*, Vol. 11, No. 3, pp. 205-217, 2012.
- [12] H. Lee and H. Park, "Investigation of Factors Affecting the Effects of Online Consumer Reviews," *Informatization Policy*, Vol. 20, No. 3, pp.3-17, 2013.
- [13] P. Racherla and W. Friske, "Perceived 'Usefulness' of Online Consumer Reviews: An Exploratory Investigation Across Three Services Categories," *Electronic Commerce Research and Applications*, Vol. 11, No. 6, pp. 548-559, 2012.
- [14] P. Briggs, B. BurFord, A. De Angeli, and P. Lynch, "Trust in Online Advice," *Social Science Computer Review*, Vol. 20, NO. 3, pp. 321-332, 2002.
- [15] C. Forman, A. Ghose, and B. Wiesenfeld, "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, Vol. 19, No. 3, pp. 291-313, 2008.
- [16] A. Josang, R. Ismail, and C. Boyd, "A Survey of Trust and Reputation Systems for Online Service Provision," *Decision support systems*, Vol. 43, No. 2, pp. 618-644, 2007.
- [17] W. Medhat, A. Hassan, and H. Korashy, "Sentiment Analysis Algorithms and Applications: A Survey," *Ain Shams engineering journal*, Vol. 5, No. 4, pp. 1093-1113, 2014.
- [18] J. Song and S. Lee, "Automatic Construction of Positive/Negative Feature-Predicate Dictionary for Polarity Classification of Product Reviews," *Journal of Computing Science and Engineering*, Vol. 38, No. 3, pp. 115-177, 2011.
- [19] E. You, Y. Kim, N. Kim, and S. Jeong, "Predicting the Direction of the Stock Index by Using a Domain-Specific Sentiment Dictionary," *Journal of intelligence and information systems*, Vol. 19 No. 1, pp. 95-110, 2013.
- [20] S. Lee, J. Choi, and J. Kim, "Sentiment Analysis on Movie Review through Building Modified Sentiment Dictionary by Movie Genre," *Journal of intelligence and information systems*, Vol. 22, No. 2, pp. 97-113, 2016.
- [21] Y. Kim, "Convolutional Neural Networks for Sentence Classification," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP)*, pp. 1746-1751, 2014.
- [22] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *International Conference on Learning Representations 2015*, 2014.
- [23] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical Attention Networks for Document Classification," In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pp. 1480-1489, 2016.
- [24] M. E. Peters, N. Mark, I.Mohit, G. Matt, C. Christopher, K. Lee, and Z. Luke, "Deep Contextualized Word Representations," arXiv:1802.05365, 2018.
- [25] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv:1810.04805, 2018.
- [26] Z. Yang, Z. Dai, Y. Yang, C. Jaime, R. S. Russ, and Q. V. Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," arXiv:1906.08237, 2019.
- [27] G. Nam and E. Jo, "Korean Text Sentiment Analysis," *Communication-Books*, 2017.

Authors



Kwangho Lee received the B.A. degree in Business Administration from University of Seoul in 2015, M.S. degree in Management Information Systems from Graduate School of Business IT at Kookmin University in 2020.

Kwangho Lee is interested in text mining, natural language processing, and deep learning.



Namgyu Kim received the B.S. in Computer Engineering from Seoul National University in 1998, M.S. and Ph.D. degrees in Management Engineering from KAIST, Korea, in 2000 and 2007, respectively.

Dr. Kim joined the faculty of the School of Management Information Systems at Kookmin University, Seoul, Korea, in 2007. He is currently a professor in the Graduate School of Business IT at Kookmin University. He is interested in text mining, deep learning, and data modeling.