韓國IT서비스學會誌
第17卷 第2號
2018年 6月, pp.165-177

Journal of Information Technology Services
https://doi.org/10.9716/KITS.2018.17.2.165

Hierarchical Attention Network를 활용한 주제에 따른 온라인 고객 리뷰 분석 모델

장인호* · 박기연** · 이준기***

Analysis of the Online Review Based on the Theme Using the Hierarchical Attention Network

In Ho Jang* · Ki Yeon Park** · Zoon Ky Lee***

■ Abstract ■

Recently, online commerces are becoming more common due to factors such as mobile technology development and smart device dissemination, and online review has a big influence on potential buyer's purchase decision. This study presents a set of analytical methodologies for understanding the meaning of customer reviews of products in online transaction. Using techniques currently developed in deep learning are implemented Hierarchical Attention Network for analyze meaning in online reviews. By using these techniques, we could solve time consuming pre-data analysis time problem and multiple topic problems. To this end, this study analyzes customer reviews of laptops sold in domestic online shopping malls. Our result successfully demonstrates over 90% classification accuracy. Therefore, this study classified the unstructured text data in the semantic analysis and confirmed the practical application possibility of the review analysis process.

Keyword : Deep Learning, Hierarchical Attention Network, Online Market, Online Customer Review, Web Text Analysis

Submitted : May 23, 2018 1st Revision : April 13, 2018 Accepted : June 12, 2018

* 연세대학교 정보대학원 석사
** 연세대학교 정보대학원 석사과정
*** 연세대학교 정보대학원 교수, 교신저자

1. 서 론

상품을 직접 확인해볼 수 없는 온라인 거래의 특성상 이전 구매자의 온라인 리뷰는 잠재 구매자의 구매의사에 큰 영향을 끼친다(Minqing and Bing, 2004; Scaffidi et al., 2007). 실제로 전 세계의 전자상거래 시장을 주도하는 온라인 쇼핑 사이트 아마존(Amazon.com)과 이베이(ebay)는 상품을 구매한 고객의 리뷰를 통해 잠재 고객의 구매의사 결정에 소요되는 시간과 심리적 비용을 줄이고자 하였다. 이처럼 다른 사용자의 제품 구매의사에 결정적인 역할을 하고 있는 온라인 제품 리뷰의 활용은 기업 입장에서 매우 중요한 역할로 인식되고 있으며, 온라인 리뷰 분석에 관한 연구 역시 활발히 진행되고 있다(Korfiatis et al., 2012; Kuan et al., 2015; Mudambi et al., 2010; Singh et al., 2017).

온라인 제품 리뷰에 대한 연구에는 크게 세 가지 방향의 연구들이 진행되고 있다. 첫 번째는 리뷰의 평점, 리뷰 참여자(Reviewer) 수, 텍스트 길이 등을 활용한 리뷰의 유용성 평가에 대한 연구(Dellarocas et al., 2007; Korfiatis et al., 2012; Kuan et al., 2015; Singh et al., 2017)이고, 두 번째는 누가 리뷰를 작성하는가에 관한 온라인 리뷰의 참여성에 대한 연구(Dellarocas et al., 2010; Yoo and Gretzel, 2008)이며, 세 번째는 제품에 대한 온라인 리뷰의 의미적(Semantic) 연구(Kraychev and Koychev, 2011; Somprasertsri and Lalitrojwong, 2010)로, 텍스트 마이닝(Text Mining)을 활용하여 의미 사전(Semantic Lexicon)을 통해 리뷰가 가진 의견(Opinion)을 파악하거나 극성(긍정/부정)을 판단하는 감성 분석(Sentiment analysis)을 실시한다. 세 가지 연구 방향 중 본 연구는 의미적 연구에 해당된다.

기존 의미적 연구들의 한계점은 다음과 같다. 첫째, 감성 분석 위주의 연구들이 진행되었기 때문에 상품에 대한 극성만을 판단할 수 있었을 뿐 제품 구매자들이 남긴 리뷰의 전반적인 생각을 파악하기 어렵다(Lee and Ahn, 2017; Kim et al., 2011). 둘째, 한글 텍스트의 경우 분류기의 성능이 낮아지는 문제로 인해 품사 채택이나 음소, 음절 단위의 분석 등 다양한 방법으로 형태소 분석을 진행하는데, 이 과정에서 정보 손실이 발생하며, 전처리 방법을 선택하고 진행하는 일에 있어서도 시간이 소요된다(Chang et al., 2015; Go and Shin, 2010; Kam and Song, 2012; Lee et al., 2008). 셋째, 구조적 내용 분석이 힘듦으로 사용자들이 준 점수에 의존하고 있는데, 점수 자체가 존재하지 않은 리뷰가 많아 활용도가 적다(Song and Lee, 2011; Yeon et al., 2011).

최근 텍스트 마이닝 분야에서는 온라인 리뷰의 전반적인 생각을 파악하기 위해 딥러닝(Deep Learning) 기법을 활용한 시퀀스-투-시퀀스 학습(Sequence-to-Sequence learning) 연구들(Cho et al., 2014; Sutskever et al., 2014)이 활발히 진행되고 있다. 이러한 딥러닝을 통해 제품 구매자들의 전반적인 생각을 파악하기 위한 자연어 처리(Natural Language Processing, 이하 NLP), 문서 분류(Classification)와 같은 연구들(Bengio, 2009; Glorot et al., 2011) 역시 진행되고 있지만 분류기 활용을 위해 필요한 레이블링(Labeling) 과정에서 작업자가 육안으로 확인하고 주제를 추출하기 때문에 많은 시간과 노력이 소요되는 문제가 있다. 이러한 점을 볼 때 실무에서 딥러닝 기법을 온라인 리뷰 분석에 적용하기에는 아직까지 한계가 있어왔다.

따라서 본 연구에서는 단순 극성과 점수에 의존하고 있는 리뷰 내용 자체 분석을 위해 딥러닝 기법의 계층적 강조 네트워크(Hierarchical Attention Network, 이하 HAN)를 사용하였다. HAN에서는 단어와 문장을 분리하여 단어의 가중치를 추출한 후 문장 인코더(Sentence Encoder)에 삽입하는 순환 신경망(Recurrent Neural Network, 이하 RNN)을 이용하는데, 이 방식은 문장의 길이와 숫자가 일정치 않는 온라인 리뷰 분석에 적합하다. 또한 본 연구에서는 기존 한글 텍스트 형태소 분석 과정에서 소요되는 시간과 정보 손실을

줄이기 위해 품사 선택에 대한 고려를 하지 않고 문장 안의 모든 품사를 분석에 활용하였다. 이렇게 함으로써 주요 명사와 품사를 구별하지 않는 분류가 되어 성능이 떨어질 수 있으나, 딥러닝의 레이어(Layer)를 이용하여 불필요한 조사가 처리될 수 있으므로 전처리 과정의 시간을 대폭 줄임과 동시에 분류기의 성능을 유지시킬 수 있다.

본 논문에서는 주요 온라인 쇼핑몰에서 수집된 총 150,540개의 리뷰 데이터셋을 이용하여 분류 실험을 진행하였다. 분류 결과, 데이터가 정형화되어있지 않았음에도 불구하고, 분류 정확도는 90%를 상회하는 결과를 보였다. 또한, 기존 연구와 달리 리뷰에 포함된 다양한 주제들 역시 추출할 수 있다.

우리는 이 방식을 통해 온라인 리뷰를 참고하는 잠재 구매자들에게는 제품의 상세한 정보를 전달하고, 쇼핑몰 관리자에게는 텍스트 전처리와 레이블링 과정에서 육안으로 주제를 추출할 때 소요되는 시간을 줄여줌으로써 효율적인 온라인 리뷰의 관리 방법이 될 것으로 기대한다.

본 연구의 구성은 다음과 같다. 제 2장에서 본 연구에 활용된 기법들에 대한 이론적 배경을 소개하고, 제 3장에서 연구 방법에 대해 설명하며, 제 4장에서는 연구 결과를 기술한다. 마지막으로 제 5장에서는 본 연구의 시사점과 한계, 그리고 향후 연구 방안에 대해 기술한다.

2. 이론적 배경

2.1 Latent Dirichlet Allocation

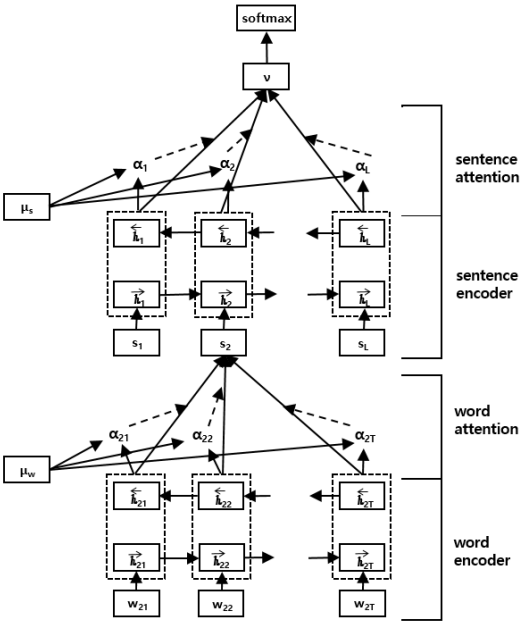
잠재 디리클레 할당(Latent Dirichlet Allocation, 이하 LDA)은 문서 내 어떤 주제들이 존재하는지 파악하게 해주는 확률적 생성 모델로 대표적인 토픽 모델링 기법이다(Blei et al., 2003). 문서 안에 있는 단어의 분포를 분석하여 해당 문서가 어떤 주제들을 가지고 있는지 예측할 수 있다. LDA는 '각각의 문서는 여러 개의 주제를 가지고 있다'와 '주제들은 디리클레 분포(Dirichlet Distribution)를 따른다' 그리고 '단어의 교환성(Exchangeability)'과 같은 특징적인 가정에 기반하고 있다. 특히 '단어의 교환성'은 문헌이 단어의 빈도수만을 가지고 표현을 가능케 하는데 이를 기반으로 단어와 문헌의 교환성을 포함한 혼합 모델이 LDA라고 할 수 있다. LDA를 통해 문헌의 내포된 주제를 찾아내고 데이터가 전반적으로 의미하는 바를 파악하고자 하는 연구가 활발히 진행되고 있다. 특히, Griffiths와 Steyvers(2004)는 PNAS 내 과학 분야의 연구논문들의 전반적인 주제를 파악하기 위해 LDA 기법을 활용하였다.

본 연구에서는 전체 온라인 고객 리뷰의 전반적인 성향 즉, 리뷰의 주된 주제를 파악하기 위해 LDA 모델을 활용한다. 고객 리뷰에는 다양한 제품 관련 정보들이 포함되어 있으므로 리뷰의 전반적인 주제를 파악하기 위한 목적에 적합하다.

2.2 Hierarchical Attention Networks

언어 번역, 문서 분류의 연구에서는 입력 데이터의 연속성을 반영하여 학습이 어려운 DNN(Deep Neural Network)의 단점을 극복하고자, 시퀀스 투 시퀀스 러닝(Sequence to Sequence Learning)을 활용한 연구가 활발히 진행되고 있다. 특히, RNN 중 HAN은 문서의 계층적 특성을 반영하기 위한 모델로서, 문서 분석 연구에 활발히 쓰이고 있다(Yang et al., 2016).

HAN은 단어와 단어뿐 아니라 문장과 문장을 고려하기 위해 두 개의 강조 매커니즘으로 구성되어 있다. 첫 번째 강조 매커니즘은 단어를 처리하기 위한 것으로 단어 인코더(Word Encoder), 단어 가중치(Word Attention)로 구성되어 있다. 두 번째 강조 매커니즘은 문장을 처리하기 위한 것으로, 문장 인코더, 문장 가중치(Sentence Attention)로 이루어져 있다. 단어 가중치를 통해 산출된 값은 문장 인코더의 입력으로 쓰이고 최종적으로 소프트맥스(Softmax) 함수를 통해 문서를 분류



〈Figure 1〉 Online Review Analysis using Hierarchical Attention Network

한다. HAN의 인코더는 기본적으로 Gated Recurrent Unit(이하 GRU) 기반의 양방향 순환 신경망(Bidirectional RNN)을 활용한다(Bahdanau et al., 2014).

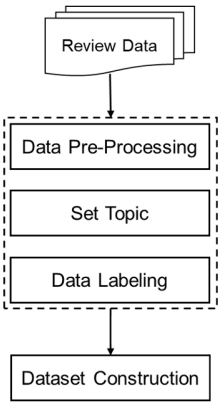
HAN은 강조 매커니즘과 가중치를 사용함으로써 길이가 다양한 온라인 고객 리뷰에 대한 분석을 효과적으로 할 수 있고, 문서들의 계층적 특성을 반영할 수 있으므로, 기존 RNN 방식에 비해 우수한 분류 성과를 보이고 있다(Nikolaos and Belis, 2017).

본 연구에서는 길이가 다양한 온라인 고객 리뷰에 대한 분석을 효과적으로 하기 위해 RNN에서 파생된 강조 매커니즘(Attention Mechanism)의 확장 모델인 HAN을 활용했다(<Figure 1> 참고).

3. 연구 방법

3.1 연구 절차

연구 절차는 다음과 같은 방법을 거쳤다. 우선 국내 온라인 쇼핑몰의 리뷰 데이터를 파이썬(Python)의 수집 라이브러리 중 하나인 셀레니움(Selenium)을 통해 수집했다. 수집한 리뷰 데이터에 대하여 LDA 기법을 적용하여 확률적으로 어떤 주제들이 존재하는지 파악하고, 두 가지 레이블링 기법을 통해 추출한 주제들을 각 리뷰마다 할당하여 데이터셋을 구축하였고, 딥러닝 기법인 HAN을 통해 분류를 진행했다(<Figure 2> 참고).



〈Figure 2〉 Research Model

3.2 온라인 고객 리뷰 데이터의 수집

본 연구는 온라인 리뷰를 의미적으로 분석하기 위해 국내 온라인 쇼핑몰 네 곳에서 리뷰를 수집하였다. 분석 제품은 리뷰가 가장 활발히 작성되는 제품 군 중 하나인 노트북으로 선정하였으며 쇼핑몰에 게시된 2,065개의 노트북 제품의 150,540개의 리뷰 데이터를 수집하였다(<Table 1> 참고).

〈Table 1〉 The Number of Products and Reviews based on Online Shopping Malls

온라인 쇼핑몰	A사	B사	C사	D사	합계
제품 수	300	722	531	514	2,065
리뷰 수	57,206	60,040	26,061	7,233	150,540

Sentence

기존에 쓰던 노트북이 너무 오래 된거라 새로 바꾸기 위해 노트북을 알아보던 도중 상품평이 너무 좋아서 구매하게 되었는데요..

Tokenize

('기존', 'Noun'), ('에', 'Josa'),('쓰', 'Verb'),('던', 'Eomi'), ('노트북', 'Noun'), ('이', 'Josa'), ('너무', 'Noun'), ('오래', 'Noun'), ('된거', 'Noun'), ('라', 'Josa'), ('새로', 'Adjective'), ('바꾸기', 'Verb'), ('위해', 'Noun'), ('노트북', 'Noun'), ('을', 'Josa'), ('알아보던', 'Verb'), ('도중', 'Noun'), ('상품', 'Noun'), ('평이', 'Noun'), ('너무', 'Noun'), ('좋', 'Adjective'), ('아서', 'Eomi'), ('구매', 'Noun'), ('하게', 'Verb'), ('되었', 'Verb'), ('는', 'PreEomi'), ('데요', 'Eomi'), ('..', 'Punctuation')

〈Figure 3〉 Example of Morphology Analysis

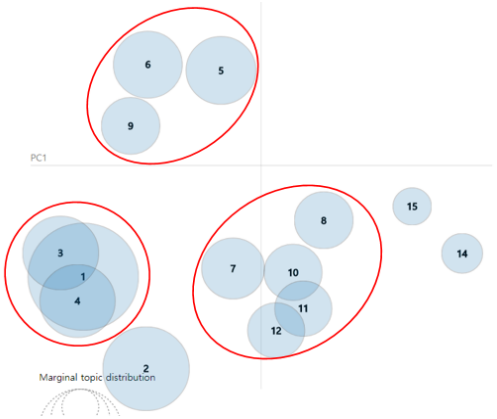
3.3 데이터의 전처리

수집한 리뷰 데이터 및 기사 데이터는 트위터(Twitter) 형태소 분석기를 활용하여 형태소 분석을 시행하였다. 데이터 전처리 과정을 최소화 하고자 품사 선택에 대한 고려를 하지 않고 문장 안의 모든 품사를 분석에 활용하였다. <Figure 3>는 형태소 분석의 예시이다.

4. 연구 결과

4.1 온라인 리뷰의 토픽 모델링

먼저 전체 리뷰 데이터에서 먼저, 15개의 토픽을 추출하여 t-SNE 알고리즘을 통해 도식화하였다. t-SNE는 각 토픽들의 차원을 축소하여 2차원 평면상에 표현할 수 있게 하는 알고리즘으로 <Figure 4>는 추출된 토픽들 간의 유사함을 나타낸다. <Figure 4>에서 총 15개의 토픽들 중 토픽 1, 3, 4와 토픽 5, 6, 9 그리고 토픽 7, 8, 10, 11, 12이 서로 거리가 가까운 것은 각 토픽 별로 나타난 단어들의 분포가 매우 비슷하여 유사한 의미를 지닌다고 볼 수 있다. 따라서 토픽들의 의미 중복을 방지하기 위해 서로 의미가 겹치는 토픽들을 합하여 최종적으로 '소프트웨어 관련', '하드웨어 관련', '외관 관련', '가격 관련', '사은품 관련', '배송 관련', '구입 목적 관련'의 총 7개 토픽으로 선정할 수 있었다. 이를 통해 구매자들이 해당 제품에 대해 가진 성향이 단지 리뷰의 평점에 따른 감성적 요소(좋다/나쁘다) 외에도 다양하게 나타난다는 점을 확인하였다. 각 토픽 별 상위 5개 단어는 <Table 2>와 같다.



〈Figure 4〉 Relative Distance of Topics-Reorganize to 7 Topics

〈Table 2〉 Words by Topics

소프트웨어 관련	하드웨어 관련	외관 관련	가격 관련	사은품 관련	배송 관련	구입 목적
윈도우	메모리	색상	가격	마우스	배송	선물
설치	속도	화이트	저렴	사은품	바로	인강
드라이버	세대	크기	할인	가방	수령	아들
프리도스	하드	실버	만원	파우치	오늘	신랑
부팅	그래픽	흰색	쿠폰	무선	방문	아이

4.2 주제에 따른 데이터의 분류 실험

본 연구에선 온라인 고객 리뷰의 의미적 분석 방법론을 제안하고자 LDA를 통해 총 7개의 토픽들을 찾고, 이를 학습 데이터의 클래스로 활용하여 분류를 진행하였다. 그러나 온라인 리뷰 대부분은 하나의 리뷰에 다양한 주제들이 포함될 수 있는데 가령, “외관은 다른 노트북에 비해 무겁지만 디자인은 투박하지 않아 만족합니다. 윈도우 10이 설치되어 있지 않아 프리도스인 상태였고, 배송은 빨라서 상품에 대해 전반적으로 만족합니다.”의 경우, 하나의 리뷰 안에 제품의 외관, 소프트웨어 상태, 배송에 대한 구매자의 의견들이 함께 포함되어있다. 따라서 이러한 경우, 하나의 리뷰를 대표할 수 있는 토픽을 클래스로 선정해주는 작업이 필요하며, 본 연구에서는 가장 많은 숫자를 보이는 토픽을 대표 토픽으로 선정하는 보팅 레이블링과 모든 토픽에 대한 내용을 포함한 새로운 클래스를 선정하는 멀티 레이블링 기법을 통해 리뷰 데이터에 클래스를 부여하였다.

4.2.1 Voting Labeling

보팅 레이블링은 먼저 문장에 목적 변수를 부여하고 종합한 결과를 바탕으로 리뷰 전체에 최종적으로 목적 변수를 부여하는 방법이다. 즉, 리뷰 내 문장 안에 있는 단어가 LDA를 통해 추출한 토픽들 각각의 말뭉치(Corpus)에 속한 단어라면 문장의 목적 변수는 해당 토픽이 된다. 예를 들어 “배송은 빨라서 정말 좋았습니다.”라는 문장의 경우 ‘배송’이라는 단어가 ‘배송 관련’ 토픽의 말뭉치에 포함된 단어이므로 해당 문장의 목적 변수는 ‘배송 관련’이 된다. 같은 방법으로 모든 문장에 대해 클래스를 부여하고 문장들의 클래스 수를 세어 그 수가 가장 많은 클래스를 리뷰 전체의 클래스로 부여한다(<Figure 5> 참고).

문장들의 클래스 수가 같은 경우에는 어떤 것이 대표 토픽이라고 정할 수 없어서 데이터 셋에서 제외하였다. 그 결과, 학습과 테스트를 위한 총 31,117개 데이터 셋을 구축하였다. <Table 3>는 보팅 레이블링을 통해 구축한 리뷰 데이터의 클래스 별 리뷰의 분포를 나타낸 표이다. 이를 통해 온라인 쇼핑몰에서 노트북 제품의 경우 주로 ‘배송 관련’, ‘노트북 외관 관련’ 논의가 활발한 것으로 확인할 수 있었다.

Example	Label
“다른 노트북에 비해 무거웠습니다.” (Sentence1)	하드웨어 관련 (Topic2)
“디자인은 세련되어 만족합니다.” (Sentence2)	하드웨어 관련 (Topic2)
“윈도우가 설치되어 있지 않아 프리도스인 상태였지만, 배송은 빨라 정말 좋았습니다.” (Sentence3)	소프트웨어 관련 (Topic1)
“다른 노트북에 비해 무거웠습니다. 디자인은 세련되어 만족합니다. 윈도우가 설치되어 있지 않아 프리도스인 상태였지만, 배송은 빨라 정말 좋았습니다.” (Review)	하드웨어 관련 (Topic2)

〈Figure 5〉 Example of Voting Labeling

〈Table 3〉 the Number of Reviews by Class of Voting Labeling

Class	Number of Review
Label 1(소프트웨어 관련)	2,414
Label 2(하드웨어 관련)	4,345
Label 3(노트북 외관 관련)	7,401
Label 4(가격 관련)	4,327
Label 5(사은품 관련)	1,578
Label 6(배송 관련)	9,538
Label 7(구입 목적 관련)	1,514
총 합	31,117

4.2.2 Multi Labeling

이전의 보팅 레이블링의 경우 문장들의 클래스 수를 세는 투표 형식이기 때문에 상대적으로 적은 수의 문장들의 클래스는 배제되므로 리뷰가 포함하는 정보가 소실될 여지가 많다. 따라서 멀티 레이블링을 통해 리뷰에 포함된 문장들의 클래스를 모두 반영하고자 하였다(<Figure 6> 참고).

Example	Label
"디자인이 세련되었고 가격도 저렴했습니다." (Sentence1)	하드웨어, 가격 관련 (Topic2,4)
"프리도스인 상태였지만, 택배는 빨랐습니다." (Sentence2)	소프트웨어, 배송 관련 (Topic1,6)
"무게가 생각보다 무거웠고, 디자인은 그냥 그랬지만 사은품이 좋습니다." (Sentence3)	하드웨어, 사은품 관련 (Topic2,5)
"디자인이 세련되었고 가격도 저렴했습니다. 프리도스인 상태였지만, 택배는 빨랐습니다. 무게가 생각보다 가벼웠고, 디자인은 괜찮지만 가격이 비쌌습니다." (Review)	소프트웨어, 하드웨어, 가격, 사은품, 배송 관련 (Topic1,2,4,5,6)

〈Figure 6〉 Example of Multi Labeling

멀티 레이블링은 리뷰가 문장의 모든 클래스를 포함한다. 따라서 <Figure 6>의 경우 해당 리뷰는 토픽(Topic) 1, 토픽 2, 토픽 4, 토픽 5, 토픽 6가 하나의 고유한 클래스가 된다. 이를 통해 총 42,775개의 데이터 셋을 구축했다. 멀티 레이블링을 통해 구축한 리뷰 데이터 셋의 분포는 <Table 4>와 같다.

멀티 레이블링을 통해 문장 내의 모든 목적 변수 값을 종합하고 이를 바탕으로 총 127개의 클래스로 리뷰가 분류됨을 확인하였다.

4.3 분류 실험 결과

본 연구는 분석을 위해 두 가지의 실험적 방법으로 온라인 고객 리뷰에 클래스를 부여하였다.

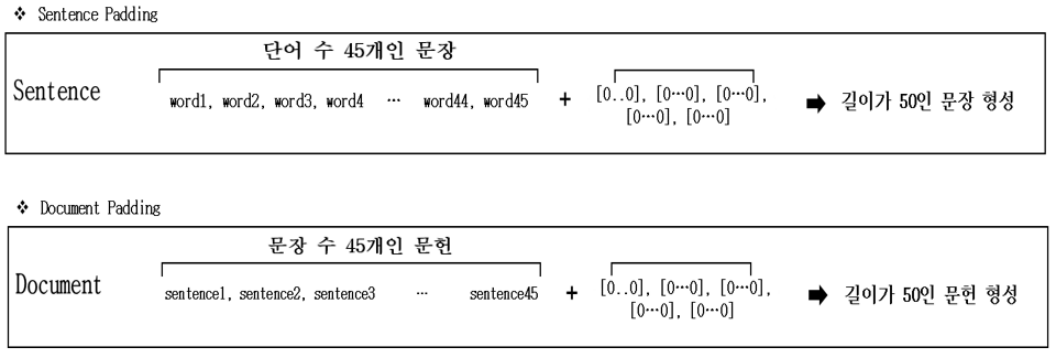
〈Table 4〉 The Number of Reviews by Class of Multi Labeling

Class	Topic	Number of Review
Label 1	1	567
Label 2	1, 2, 3	405
Label 3	1, 2, 3, 4	322
Label 4	1, 2, 3, 4, 5	141
…		
Label 71	2, 3, 4, 6	610
Label 72	2, 3, 4, 6, 7	276
…		
Label 126	6,7	310
Label 127	7	673
총 합		42,775

그리고 목적 변수가 생성된 온라인 고객 리뷰 데이터에 대해 최종적으로 분류 모델에 적용하는 실험을 진행했다.

4.3.1 실험 데이터 셋

리뷰 데이터는 문장 안에 포함된 단어의 수가 최대 50개, 문헌 안에 포함된 문장의 수가 최대 50개가 되도록 제한하여 50×50크기의 매트릭스로 구성했다. 그리고 사전 학습 시킨 64차원의 단어 임베딩 벡터 값을 리뷰 데이터에 적용하여 입력 데이터는 N×50×50×64크기의 4차원 배열의 형태를 갖는다. 또한 문장 안의 단어 수 혹은 문헌 안의 문장 수가 각각 50개 미만이면 나머지 수만큼 0을 채워주는 패딩(Padding) 작업을 실시했다.



〈Figure 7〉 Example of Padding

4.3.2 실험 결과

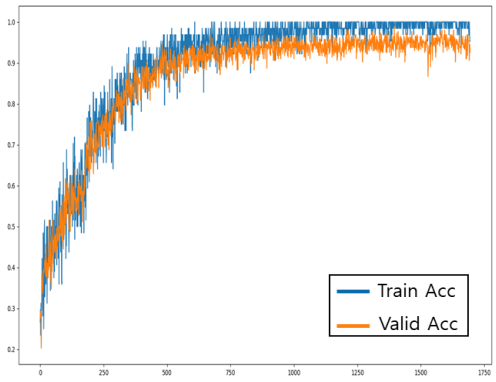
문서 분류 딥러닝 모델 HAN을 활용하여 두 가지 방법으로 클래스를 할당한 온라인 고객 리뷰 데이터를 대상으로 분류 실험을 진행했다.

4.3.2.1 Voting Labeling 데이터 셋 실험 결과

보팅 레이블링으로 구축한 31,054개의 리뷰 데이터 셋으로 분류 실험을 진행했다. 학습 데이터 셋은 24,843개, 테스트 데이터 셋은 6,211개이며 10겹 교차 검증(10-Fold Cross Validation)을 통해 학습 모델을 검증했다. 배치 크기는 64로 하여 총 1,750회 반복 학습하였다.

<Figure 8>, <Figure 9>는 1,750회 모델을 학습시킨 결과이다. <Figure 8>에서 Y축은 정확도, X축은 학습 횟수를 의미하며 상대적으로 위에 있는 그래프가 학습(Train) 데이터 셋의 그래프, 아래 있는 그래프가 검증(Valid) 데이터 셋의 그래프이다. <Figure 8>은 리뷰의 분류 정확도를 학습 횟수에 따라 나타낸 것이다.

<Figure 9>에서 Y축은 손실(Loss), X축은 학습 횟수를 의미하며 상대적으로 위에 있는 그래프가 검증(Valid) 데이터 셋의 그래프, 아래 있는 그래프가 학습(Train) 데이터 셋의 그래프이다. <Figure 9>는 학습 데이터와 검증 데이터 각각에 손실 함수인 크로스 엔트로피(Cross Entropy)를 적용하여 구한 모델의 손실을 학습 횟수에 따라 나타낸 것이다.



〈Figure 8〉 Accuracy Graph of Train and Valid Set (Voting Labeling)

〈Table 5〉 F-Score of Voting Labeling

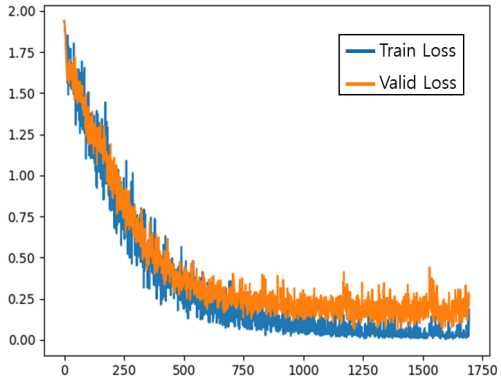
Class	Precision	Recall	F-Score
1	0.91	0.90	0.91
2	0.99	0.85	0.92
3	0.94	0.92	0.93
4	0.86	0.95	0.90
5	0.85	0.97	0.90
6	0.92	1.00	0.96
7	0.89	0.89	0.89
Avg	0.93	0.92	0.92

<Figure 8>과 <Figure 9>에서 나타난 학습 데이터 셋의 그래프와 검증 데이터 셋의 그래프 간의 간극이 작은 것을 볼 때, 보팅 레이블링에서는 과적합 문제가 발생하지 않았음을 확인할 수 있다.

7개의 클래스에 대한 리뷰 분류 실험 결과, 데이터의 분포가 불균형한 클래스임에도 불구하고 리뷰 데이터의 분류 정확도는 95.117%, 손실 값은 0.1674로 준수한 성능을 보였다. 그리고 학습된 모델의 혼동 행렬(Confusion Matrix)과 F스코어(F-Score)를 통해 데이터가 각 클래스마다 잘 분류되었음을 확인했다.

4.3.2.2 Multi Labeling 데이터 셋 실험 결과

멀티 레이블링으로 구축한 41,387개의 리뷰 데이터 셋으로 분류 실험을 진행했다. 학습 데이터 셋은 33,109개, 테스트 데이터 셋은 8,278개 이며



〈Figure 9〉 Loss Graph of Train and Valid Set(Voting Labeling)

10겹 교차 검증을 통해 학습 모델을 검증했다. 배치 크기는 64로 하였으며 클래스가 127개로 그 수가 많기 때문에 총 3,800회 반복 학습하였다.

<Figure 10>, <Figure 11>을 살펴보면, 보팅 레이블링과 유사한 그래프 결과를 보임을 알 수 있다. 또한, 멀티 레이블링에서 역시 과적합 문제가 발생하지 않았음을 확인할 수 있다.

127개의 클래스에 대한 멀티 레이블링 데이터셋의 분류 실험 결과 모델의 분류 정확도는 91.37%, 손실 값은 0.32387로 나타났다. 보팅 레이블링에 비해 그 성능이 약간 떨어졌지만 클래스의 수가 약 18배 더 많은 것을 감안한다면 비교적 준수한 성능을 나타냈다고 판단할 수 있다. 그리고 멀티 레이블링 또한 혼돈 행렬과 F스코어를 클래스별 데이터를 통해 분류가 잘 이루어졌음을 확인하였다.

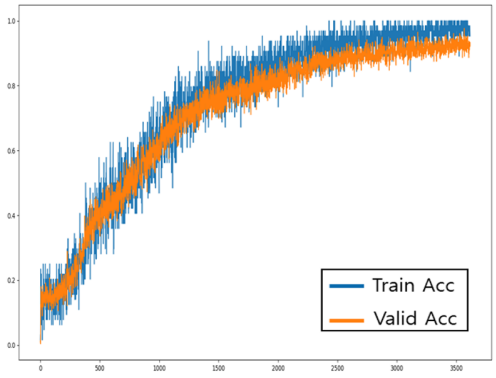
5. 결 론

5.1 연구 결과 요약

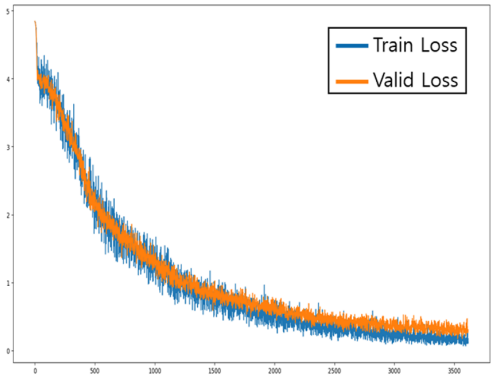
본 연구는 온라인 리뷰에 포함된 다양한 주제를 파악하여 잠재 구매자들에게 제품과 관련된 상세한 정보를 제공하고, 쇼핑몰 관리자에겐 다양한 리뷰의 관리에 필요한 소요 시간을 단축시킬 수 있는 방법론을 제안하고자 하였다. 이를 위해 리뷰에서 논의되는 주제들을 찾고, 선정된 주제에 맞게 리뷰를 분류하였다. 본 연구에서 사용한 데이터는 온라인 쇼핑몰의 노트북을 구매한 고객들이 남긴 리뷰 데이터로, 본 연구에서 주로 '소프트웨어 관련', '하드웨어 관련', '노트북 외관 관련', '배송 관련', '가격 관련', '사은품 관련', '구입 목적 관련'의 총 7개의 주제에 대해 리뷰가 작성된다는 점을 파악했다. 그리고 선정된 주제들을 두 가지 레이블링 기법을 통해 클래스를 부여하였고, 딥러닝 분류 기법 중 하나인 HAN을 활용하여 실험을 진행하였다. 그 결과, 보팅 레이블링에서는 분류 정확도가 95.117%였으며, 멀티 레이블링에서는 분류 정확도가 91.37%로, 두 결과 모두 정확도가 90%를 상회하는 성능을 보임을 확인할 수 있었다.

〈Table 7〉 F-Score of Multi Labeling

Class	Precision	Recall	F-Score
1	1.0	1.0	1.0
2	1.0	1.0	1.0
3	0.58	1.0	0.74
4	1.0	0.75	0.86
…			
26	0.64	0.88	0.74
27	0.91	0.71	0.80
…			
126	1.0	1.0	1.0
127	1.0	1.0	1.0
Avg	0.91	0.91	0.91



〈Figure 10〉 Accuracy Graph of Train and Valid Set (Multi Labeling)



〈Figure 11〉 Loss Graph of Train and Valid Set(Multi Labeling)

5.2 시사점

5.2.1 학술적 시사점

본 연구는 전체 온라인 리뷰에서 논의되는 주제들을 찾아 카테고리화하여 분류한 첫 연구이다. 기존 온라인 리뷰에는 감성 점수를 나타내는 별점 이외에는 딥러닝 학습에 사용하는 데이터의 클래스로 사용할 만한 정보들이 없었기 때문에, LDA로 추출한 토픽 카테고리들을 두 가지 레이블링 방법을 통해 각 리뷰 데이터의 클래스로 사용하였다. 분류 방법에선 단어와 단어 간의 연관성, 문장과 문장 간의 연관성 역시 고려하기 위해 문서들의 계층적 특성을 반영할 수 있는 HAN을 활용하였다. 이런 방식을 채택함으로써 기존에는 학습 데이터의 클래스를 레이블링하는 과정에서 육안으로 확인하는 일이 필요했다면, 그 과정을 LDA 기법과 두 가지 레이블링 방법을 통해 자동화시킬 수 있도록 대체하였다. 또한 리뷰 데이터 사용에 있어 문장 내 모든 품사를 활용하였기 때문에 한글 텍스트를 전처리하는 과정에서 소요되는 시간 역시 줄일 수 있었다는 점에서 학술적 의의를 가진다.

5.2.2 실무적 시사점

기존 한글 딥러닝 연구에서 모델의 성능을 확인하기 위해, 일반적으로 적용하던 정형화된 데이터가 아닌 국내 온라인 쇼핑몰의 리뷰 데이터를 활용하였다. 그리고 딥러닝을 적용한 분류 모델의 정확도가 90%를 상회하는 것을 볼 때, 분류기로써 실무 적용의 가능성을 확인할 수 있었다. 또한, 반자동화된 분석 프로세스를 통해 관리자의 의존도를 낮춰 리뷰 관리의 시간과 비용을 효율적으로 관리할 수 있을 것으로 기대된다.

5.3 한계 및 향후 연구 방향

본 연구는 실험 진행에 하나의 제품군에 한정하여 진행하였기 때문에 다양한 제품에 대한 실무적 활용 가능성을 확인하지 못했으며 시간에 따라 제품의 주제가 변하는 모습을 분석에서 고려하지 않았다는 한계가 있다. 또한 고객 리뷰의 평점을 게시하지 않은 온라인 쇼핑몰이 다수 존재하여 제품에 대한 구매자의 감성을 분석에 포함시키지 못했다. 그리고 본 연구의 두 가지 레이블링 실험을 통해 얻을 수 있었던 분류 정확도의 경우, 90%를 상회한 연구 결과를 통해 실무적인 사용 가능성을 확인하였지만, 다른 분류 모델과의 비교 분석을 실시하지 않아 본 연구의 분류 방법에 대한 우수성을 검증하지 못했다는 한계를 지닌다.

이를 해결하기 위해 먼저 다양한 제품군의 리뷰를 수집하고 시간에 따른 주제의 변화를 분석에 고려하여 연구를 진행한다. 또한 제품에 대한 구매자의 감성 분석을 실시하고, 다른 분류 모델과의 정확도를 비교하여 분류기로서 성능의 우수성을 검증한다면 경영정보로서 더욱 가치 있는 연구가 될 것이다.

References

Bahdanau, D., K.H. Cho, and Y. Bengio, “Neurual Machine Translation by Jointly Learning to Align and Translate”, *ICLR*, 2014.

Bengio, Y., “Learning Deep Architectures for AI”, *Foundations and Trends*, Vol.2, No.1, 2009, 1-127, doi : 10.1561/2200000006(Downloaded September 10, 2017.)

Blei, D.M., A.Y. Ng, and M.I. Jordan, “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, 2003, 993-1022.

Chang, D.S., D.Y. Kim, and Y.S. Choi, “Opinion Mining Based on Korean Phoneme Trigram-Signature”, *KIISE*, 2015, 811-813.

(장두수, 김도연, 최용석, “한글 음소단위 Trigram-Signature 기반의 오피니언 마이닝 기법”, *한국정보과학회 학술발표논문집*, 2015, 811-813.)

Cho, K.H., B.V. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning Phrase representations

using RNN Encoder-decoder for Statistical Machine Translation", *ICLR*, 2014, 1724-1734.

Dellarocas, C., X. Zhang, and N.F. Awad, "Exploring the value of online product reviews in forecasting sales : The case of motion pictures", *Journal of Interactive Marketing* Vol.21, No.4, 2007, 23-45.

Dellarocas, C., G. Gao, and R. Narayan, "Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products?", *Journal of Management Information Systems*, Vol.27, No.2, 2010, 127-158.

Glorot, X., A. Bordes, and Y. Bengio, "Domain Adaptation for Large Scale Sentiment Classification : A Deep Learning Approach", *ICML'11 Proceedings of the 28th International Conference on International Conference on Machine Learning,* 2011, 513-520.

Go, M.S. and H.P. Shin, "Grading System of Movie Review through the Use of An Appraisal Dictionary and Computation of Semantic Segments", *COGSCI*, Vol.21, No.4, 2010, 669-696.

(고민수, 신효필, "감정어휘 평가사전과 의미마디 연산을 이용한 영화평 등급화 시스템", *인지과학*, 제21권, 제4호, 2010, 669-696.)

Griffiths, T.L. and M. Steyvers, "Finding scientific topics", *PNAS*, 2004, 5228-5235.

Jo, Y. and A.H. Oh, "Aspect and sentiment unification model for online review analysis", WSDM '11 Proceedings of the fourth ACM international conference on Web search and data mining, 2011, 815-824. doi : 10.1145/1935826.1935932(Downloaded September 10, 2017).

Kam, M.A. and M. Song, "A Study on Differences of Contents and Tones of Arguments among Newspapers Using Text Mining Analysis", *Journal of Intelligence and Information Systems*, Vol.18, No.3, 2012, 53-77.

(감미아, 송 민, "텍스트 마이닝을 활용한 신문사에 따른 내용 및 논조 차이점 분석", *지능정보연구*, 제18권, 제3호, 2012, 53-77.)

Kim, J.O., S.S. Lee, and H.S. Yong, "Automatic Opinion Classification of Korean Text", *Journal of KISS : Databases*, Vol.38, No.6, 2011, 423-428.

(김진옥, 이선숙, 용환승, "한글 텍스트의 오피니언 분류 자동화 기법", *정보과학회논문지 : 데이타베이스*, 제38권, 제6호, 2011, 423-428.)

Korfiatis, N., 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, 2012, 205-217.

Kraychev, B. and I. Koychev, *Classification of Online Reviews by Computations Semantic Lexicons*, Springer, Third International Conference on Software, Services and Semantic Technologies, 2011.

Kuan, K.Y., K.L. Hui, P. Prasarnphanich, and H.Y. Lai, "What makes a review voted? An empirical investigation of review voting in online review systems", *Journal of the Association for Information Systems*, Vol.16, No. 1, 2015, 48-71.

Lee, J.J. and S.M. Ahn, "Sentiment Analysis of Deep Learning Model Applying Phoneme Unit", *KIISS*, 2017, 113-114.

(이재준, 안성만, "한글 음소 단위를 적용한 딥러닝 모형의 감성분석", *한국지능정보시스템학회 학술대회논문집*, 2017, 113-114.)

Lee, S.W., D.G. Lee, and H.C. Rim, "A Corpus-

based Hybrid Model for Morphological Analysis and Part-of-Speech Tagging", *Journal of the Korea Society of Computer and Information)*, Vol.13, No.7, 2008, 11-18.

(이승욱, 이도길, 임해창, "형태소 분석 및 품사 부착을 위한 말뭉치 기반 혼합 모형", *한국컴퓨터정보학회논문지*, 제13권, 제7호, 2008, 11-18.)

Mikolov, T., I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality", *Advances in Neural Information Processing Systems*, 2013.

Minqing, H. and L. Bing, "Mining and summarizing customer Reviews", KDD '04 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, 168-177, doi : 10.1145/1014052.1014073 (Downloaded September 10, 2017).

Mudambi, S.M. and D.Schuff, "What Makes a Helpful Review? A Study of Customer Reviews on Amazon.com", *MIS Quarterly*, Vol.34, No.1, 2010, 185-200.

Nikolaos, P. and A.P. Belis, "Multilingual Hierarchical Attention Networks for Document Classification", arXiv preprint arXiv : 1707.00896, 2017.

Scaffidi, C., K. Bierhoff, E. Chang, M. Felker, H. Ng, and J. Chun, "Red Opal : product-feature scoring from reviews", EC '07 Proceedings of the 8th ACM conference on Electronic commerce, 2007, 182-191, doi : 10.1145/1250910.1250938(Downloaded September 10, 2017).

Singh, J.P., S. Irani, N.P. Rana, Y.K. Dwivedi, S. Saumya, and P.K. Roy, "Predicting the 'helpfulness' of online consumer reviews", *Journal of Business Research*, Vol.70, No.1, 2017, 346-355.

Somprasertsri, G. and P. Lalitrojwong, "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization", *Journal of Universal Computer Science*, Vol.16, No.6, 2010, 938-955, doi : 10.3217/jucs-016-06-0938 (Downloaded September 10, 2017).

Song, J.S. and S.W. Lee, "Automatic Construction of Positive/Negative Feature-Predicate Dictionary for Polarity Classification of Product Reviews", *Journal of KIISE : Software and Applications*, Vol.38, No.3, 2011, 157-168.

(송종석, 이수원, "상품평 극성 분류를 위한 특징별 서술어 긍정/부정 사전 자동 구축", *정보과학논문지 : 소프트웨어 및 응용*, 제38권, 제3호, 2011, 157-168.)

Sutskever, I., O. Vinyals, and Q.V. Le, "Sequence to Sequence Learning with Neural Networks", *ICLR*, 2014, 3104-3112.

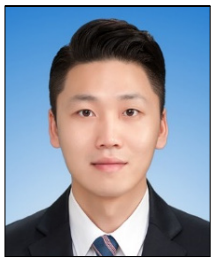
Yang, Z., D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical Attention Networks for Document Classification", *In Proc. of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, 2016, 1480-1489.

Yeon, J.H., D.J. Lee, J.H. Shim, and S.G. Lee, "Product Review Data and Sentiment Analytical Processing Modeling", *JSEBS*, Vol.16, No.4, 2011, 125-137.

(연종흠, 이동주, 심준호, 이상구, "상품 리뷰 데이터와 감성 분석 처리 모델링", *한국전자거래학회지*, 제16권, 제4호, 2011, 125-137.)

Yoo, K. and U. Gretzel, "What motivates consumers to write online travel reviews?", *Information Technology & Tourism*, Vol.10, No.4, 2008, 283-295.

◈ About the Authors ◈



In Ho Jang (norexz@naver.com)

In Ho Jang received his master degree in I.S. from Graduate School of information, Yonsei University in 2018. He is currently working as Deep-Learning Engineer at Wanted Lab. His current research interests include Data Analysis, IoT, and etc.

Ki Yeon Park (daebak-kijeong@daum.net)

Ki Yeon Park is currently a master's course in I.S. from Graduate School of information, Yonsei University. He worked at Hyundai Card Company for more than two years. His current research interests include Data Analysis, IoT and etc.

Zoon Ky Lee (zoonky@gmail.com)

Professor Zoon Ky Lee is currently a Professor at Graduate School of information, Yonsei University. He received his B.S. degree in Computer Science and Statistics from Seoul National University and a Ph.D. degree in Management Information Systems from University of Southern California in 1994. His current research interests include Open Innovation, Service Science Big Data Applications, and E-Transformation.