



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

碩士學位論文

온라인리뷰의 효과적
랭킹모델 도출 및 사례분석

濟州大學校 大學院

관광융합소프트웨어학과

이 창 용

2018年 2月

온라인리뷰의 효과적
랭킹모델 도출 및 사례분석

지도교수 김 근 형

이 창 용

이 論文을 工學 碩士學位 論文으로 提出함

2017年 12月

이창용의 工學 碩士學位 論文을 認准함

審査委員長 _____ ①

委 員 _____ ①

委 員 _____ ①

濟州大學校 大學院

2017年 12月

차례

표 차례	IV
그림 차례	V
ABSTRACT	VI
제 1 장 서론	1
제 1 절 연구배경 및 목적	1
1. 연구의 배경	1
2. 연구의 목적	2
제 2 절 연구의 방법 및 구성	3
1. 연구의 방법	3
2. 연구의 구성	4
제 2 장 이론적 배경	6
제 1 절 텍스트 마이닝	6
1. 텍스트 마이닝의 개념	6
2. 텍스트 마이닝의 단계별 과정	7
3. 텍스트 마이닝의 기법	10
1) 자연어처리기법	10
(1) 형태소 분석	11
제 2 절 오피니언 마이닝	12
1. 오피니언 마이닝의 개념	12
2. 오피니언 마이닝의 기법	14
1) 감성분석	14
3. 오피니언 마이닝 관련 기존 연구 고찰	15
제 3 절 온라인리뷰에 대한 선행연구	18
제 4 절 랭킹모델에 대한 선행연구	19

제 3 장 연구 설계	20
제 1 절 연구 데이터 선정	20
제 2 절 연구 모형	20
1. 긍정의견 및 부정의견	20
2. 조정의견	21
3. 리뷰길이	21
4. 평가점수	22
제 3 절 연구 가설	23
제 4 절 변수 측정	25
제 5 절 분석 절차	26
제 4 장 실증분석 및 제언	27
제 1 절 데이터 원천	27
제 2 절 데이터 수집	28
제 3 절 데이터 가공	29
1. 형용사 추출	29
2. 데이터 정량화 및 표준화	30
제 4 절 연구 가설의 검증	31
1. 데이터 가공 결과	31
2. 분석 결과	32
1) 변수 간의 상관분석	32
2) 연구 가설 H1의 단순회귀분석	33
3) 연구 가설 H2의 단순회귀분석	33
4) 연구 가설 H3의 단순회귀분석	34
5) 연구 가설 H4, H5, H6, H7의 다중회귀분석	34
제 5 절 랭킹모델링을 위한 제언	35
1. 랭킹모델의 정의	35
2. 온라인리뷰의 다차원 분석을 위한 랭킹모델의 적용	37

제 5 장	랭킹모델 기반의 분석 사례	38
제 1 절	랭킹모델 기반의 온라인리뷰 분석 방법	38
제 2 절	분석 사례	38
1.	관광지별 긍정/부정/조정 의견 점수에 의한 온라인리뷰 랭킹	38
2.	계절별 긍정의견점수에 의한 온라인리뷰 랭킹	40
3.	계절별 부정의견점수에 의한 온라인리뷰 랭킹	43
4.	계절별 조정의견점수에 의한 온라인리뷰 랭킹	46
5.	연도별 긍정의견점수에 의한 온라인리뷰 랭킹	49
6.	연도별 부정의견점수에 의한 온라인리뷰 랭킹	50
7.	연도별 조정의견점수에 의한 온라인리뷰 랭킹	51
제 6 장	결론	52
제 1 절	연구 요약 및 시사점	52
제 2 절	연구의 한계	54
참고문헌		55

표 차례

<표 3-1> 변수 측정	25
<표 4-1> 온라인리뷰 데이터	29
<표 4-2> 가공된 데이터	32
<표 4-3> 변수 간의 상관분석	32
<표 4-4> 연구 가설 H1의 단순회귀분석	33
<표 4-5> 연구 가설 H2의 단순회귀분석	33
<표 4-6> 연구 가설 H3의 단순회귀분석	34
<표 4-7> 연구 가설 H4, H5, H6, H7의 다중회귀분석	34
<표 5-1> 관광지별 긍정의견점수에 의한 랭킹	38
<표 5-2> 관광지별 부정의견점수에 의한 랭킹	39
<표 5-3> 관광지별 조정의견점수에 의한 랭킹	39
<표 5-4> 계절별 긍정의견점수에 의한 랭킹	41
<표 5-5> 계절별 호감도가 높은 관광지	42
<표 5-6> 계절별 부정의견점수에 의한 랭킹	44
<표 5-7> 계절별 호감도가 높은 관광지	45
<표 5-8> 계절별 조정의견점수에 의한 랭킹	46
<표 5-9> 계절별 호감도가 높은 관광지	47
<표 5-10> 연도별 긍정의견점수에 의한 랭킹	49
<표 5-11> 연도별 부정의견점수에 의한 랭킹	50
<표 5-12> 연도별 조정의견점수에 의한 랭킹	51

그림 차례

<그림 1-1> 연구의 흐름도	5
<그림 2-1> 텍스트 마이닝의 단계별 과정	7
<그림 2-2> 텍스트 마이닝이 가능한 소프트웨어의 개요	9
<그림 2-3> 자연어처리 및 구조화 과정	11
<그림 2-4> 감정 극성 엔트로피(entropy)	14
<그림 3-1> 연구 모형	23
<그림 3-2> 분석 절차	26
<그림 4-1> Tripadvisor의 리뷰 작성 예시	27
<그림 4-2> 데이터 수집 절차	28
<그림 4-3> 형용사 추출 과정	30
<그림 4-4> 데이터 정형화 및 표준화	31
<그림 4-5> 랭킹모델	36
<그림 4-6> 온라인리뷰의 다차원 랭킹모델	37

ABSTRACT

Online reviews, which would be a kind of short text documents, includes experienced story and opinions about product or services or features. The online reviews contains the evaluated results about the products or services in quantitative viewpoints as well as qualitative viewpoints. The quantitative assessment about the products would be represented as a form of star score while the qualitative assessment would be represented as a form of text sentences in online review. The text sentences in the online review includes sensitive assessments with positive or negative opinions about the products which would affect on product purchase behaviors of consumers. In this paper, we analyzed the linear regression relationship between the quantitative assessments and qualitative assessments in the online reviews. We derived implications for modeling ranking of online reviews being based on the result of linear regression relationship.

제 1 장 서론

제 1 절 연구배경 및 목적

1. 연구의 배경

웹2.0 시대를 맞아 SNS가 보편적으로 사용되고 있으며 이를 통하여 많은 온라인리뷰들이 실시간으로 빠르게 생성되고 활용되고 있다. 사람들은 SNS를 통해 자신이 경험한 공간, 제품, 서비스와 같은 콘텐츠들에 대한 의견을 작성하고 공유하며, 이러한 방대한 양의 리뷰 데이터들은 온라인리뷰로써 콘텐츠 이용 결정 여부를 위한 참고자료가 될 뿐만 아니라 콘텐츠 제공자들에게는 사용자의 의견을 파악할 수 있는 매우 가치 있는 정보로써 활용되고 있다(김문지, 송은정, 김윤희, 2015). 온라인리뷰란 인터넷에서 자신이 경험한 제품, 서비스에 관련된 자신의 의견이나 감정을 문자 혹은 이미지 형식으로 다른 소비자에게 전하는 소비자의 경험적 정보로서 사용후기, 구매후기, 댓글 등 여러 용어로 지칭되는 온라인구전의 일종이다(박은주, 정유진, 2013). 온라인리뷰는 리뷰작성자들의 상품 및 서비스의 소비에 대한 경험담이나 의견 등을 포함하며 다른 잠재적 소비자들의 소비행위에 영향을 미친다. 온라인리뷰에 대한 선행 연구 결과들에 의하면 리뷰의 정보 특성에 따라 소비자행동이 달라지고(Dwayne et al., 2001), 정보 특성의 방향성(긍정/부정)이 태도 또는 구매 의도에 영향을 미쳤으며(Peterson & Meria, 2003; Chiou & Cheng, 2003; Park & Yoo, 2006; Kim, 2007), 평가적인 정보 특성(객관성/주관성)이 리뷰 수용자에게 영향을 미치는 것으로 나타났다(Henning et al., 2004; Lee & Lee, 2005; Jeon & Jung, 2006). 이러한 결과는 상품특성에 따라 다르게 나타났는데 리뷰의 정보가 단순히 상품의 객관적 정보 뿐 아니라 소비경험과 개인감정이 포함되어 있기 때문으로 해석되었다(박은주, 정유진, 2013). 그러나 온라인리뷰를 작성할 때는 제품 및 서비스에 대한 작성자의 경험에 대하여 작성자의 주관적이고 감성적인 의견이 텍스트 형태로 표현될 뿐만 아

나라 평점과 같은 정량적인 평가점수도 함께 매겨진다. 즉, 온라인리뷰에는 작성자가 경험한 제품 및 서비스에 대한 평가내용이 포함되어 있고 텍스트 형태의 정성적인 관점과 평가점수 형태의 정량적인 관점이 동시에 표출된다.

온라인리뷰에 관한 선행연구들은 정성적 관점에서 평가한 텍스트 데이터에 대한 분석기술이 주로 이루어졌다(윤홍준, 2010; 장재영, 2009; 홍태호, 2014). 그러나 온라인리뷰를 정확하게 분석하기 위해서는 정성적 관점의 데이터와 정량적 관점의 데이터를 균형적으로 적용하여 분석해야 보다 정확하고 합리적인 분석 결과를 도출할 수 있을 것이다. 지금까지의 연구된 현황을 고찰한 결과, 정성적 관점의 평가와 정량적 관점의 평가에 대한 상관관계를 분석하고 이를 기반으로 제시하는 평가기법에 대한 연구는 전무한 실정이다.

본 연구에서는 온라인리뷰에 포함된 정성적 관점의 텍스트 데이터에 대한 감성 분석 결과와 정량적 평가점수와의 회귀적 관계를 가설검증의 형태로 고찰하며, 이를 기반으로 온라인리뷰를 보다 정확하게 평가할 수 있는 랭킹모델링에 대한 아이디어를 제시한다.

2. 연구의 목적

본 연구는 온라인리뷰에 나타나는 의견의 극성에 대해 보다 정확하고 합리적인 분석결과를 도출하기 위해 정성적인 관점에서 평가한 텍스트 데이터와 정량적인 관점에서 평가한 평가점수 데이터 간의 회귀적 관계를 가설검증 형태로 고찰하고자 한다. 이는 기존에 연구된 온라인리뷰에 대한 정성적인 관점을 위주로 분석하는 기술은 한계가 있기 때문이므로 정량적인 평가에 대한 데이터와의 상관관계를 분석하고 이를 기반으로 한 평가기법을 제시하고자 하는 것이다. 실증분석을 위해 연구 데이터 및 선행연구들을 근거하여 연구 모형 및 연구 가설을 수립한다. 온라인리뷰로부터 형용사를 추출하고 리뷰 데이터와 함께 정량화 및 표준화 작업을 수행한다. 수립된 가설들을 검증하기 위해 오피니언 마이닝 기법으로 가공된 데이터를 가지고 상관분석 및 회귀분석을 실시한다. 그리고 온라인리뷰의 랭킹을 분석할 때 적용될 수 있도록 분석 결과에서 채택된 가설들을 바탕으로 랭킹모델을 제안한다. 랭킹모델은 온라인리뷰 검색 시 효과적으로 검색 결과를

도출하는데 적용될 수 있는 일반화모델이라는 점에서 그 가치와 의의가 있으며 제안된 랭킹모델을 기반으로 온라인리뷰를 분석한 사례를 통해 시사점을 제공하는데 그 목적이 있다.

제 2 절 연구의 방법 및 구성

1. 연구의 방법

본 연구에서는 연구의 목적을 달성하기 위해 문헌적 연구 방법과 실증적 연구 방법을 병행하여 진행하였으며, 구체적인 연구 방법은 다음과 같다.

첫째, 선행연구들을 위주로 하여 기존 온라인리뷰에 대한 오피니언 마이닝 기반 분석기술, 기법에 관한 문헌 연구를 진행하였다. 특히, 가설 검증을 통해 온라인 리뷰의 변수 간 회귀적 관계를 도출하는 과정을 이해하기 위해 사회과학적 관점에서 연구된 선행연구의 흐름을 이해하였다.

둘째, 온라인리뷰와 관련하여 기존 수집된 관광리뷰를 연구 데이터로 선정하고 이를 분석 시 활용 가능한 형태로 변환하기 위해 오피니언 마이닝 기법을 적용하여 가공 처리하였다.

셋째, 수립된 연구 가설들을 검증하기 위해 가공된 데이터를 가지고 상관분석 및 회귀분석을 실시하여 이에 따른 결과를 도출하였다.

넷째, 기존 선행연구의 한계를 해결하기 위해 실증분석 결과에서 채택된 가설을 바탕으로 보다 정확하고 합리적인 리뷰 데이터 분석 결과를 도출하기 위한 랭킹 모델을 제안하고 이를 기반으로 분석한 사례를 제시하였다.

2. 논문의 구성

본 논문은 총 6 개의 장으로 구성되어 있으며 세부내용은 다음과 같다.

제 1 장 서론에서는 연구의 배경 및 목적, 연구의 방법, 본 논문의 구성에 대해 설명한다.

제 2 장 이론적 배경에서는 텍스트 마이닝 및 오피니언 마이닝의 개념, 기법에 대해 이해하고 오피니언 마이닝 관련 선행연구들을 고찰한다.

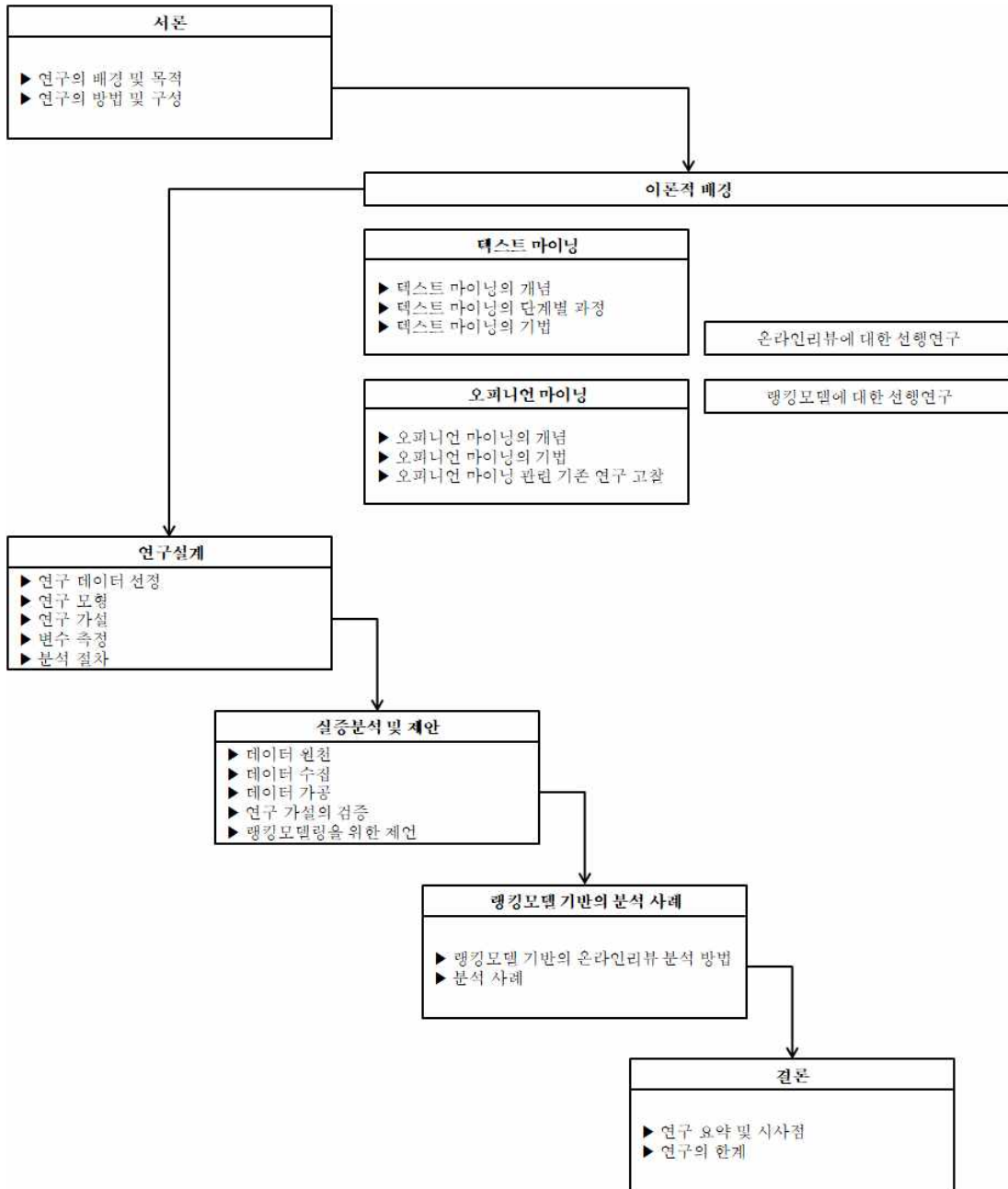
제 3 장 연구 설계에서는 연구 데이터 선정, 연구 모형, 연구 가설, 변수 측정, 분석 절차를 설명한다.

제 4 장 실증분석 및 제안에서는 데이터 수집 및 데이터 가공 과정을 설명하며 수립된 연구 가설에 따른 회귀적 관계를 검증하여 도출된 결과를 바탕으로 랭킹 모델을 제안한다.

제 5 장 랭킹모델 기반의 분석 사례에서는 제안된 랭킹모델을 기반으로 분석한 사례를 제시하여 설명한다.

제 6 장 결론에서는 연구의 요약 및 시사점을 제시하고 본 연구과정에서 나타난 한계점을 설명한다.

<그림 1-1> 연구의 흐름도



제 2 장 이론적 배경

제 1 절 텍스트 마이닝

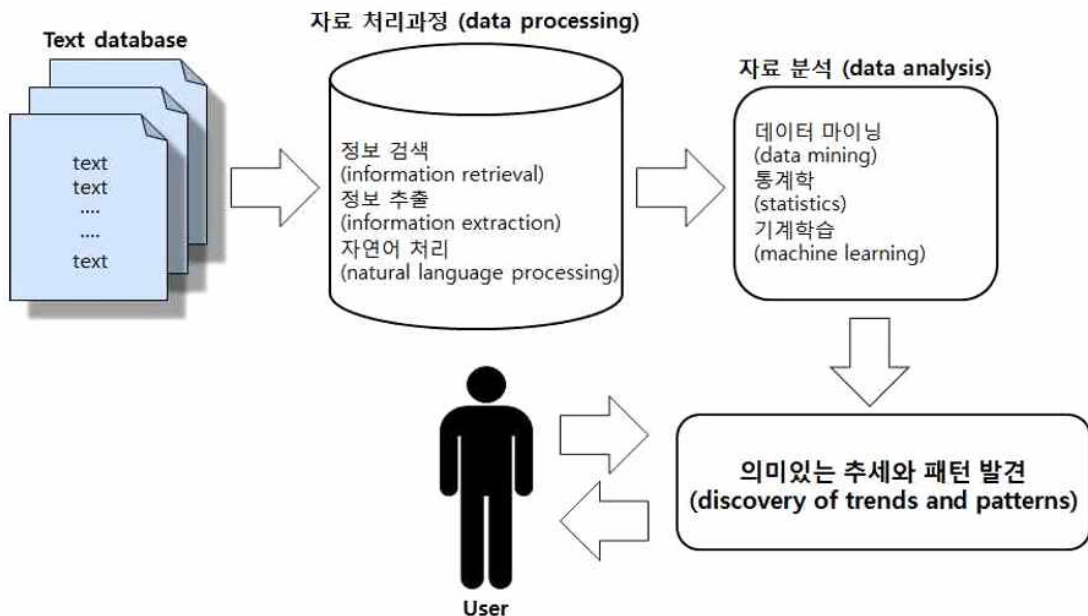
1. 텍스트 마이닝의 개념

데이터는 형태에 따라 정형 데이터와 비정형 데이터로 구분할 수 있는데, 정형 데이터는 일정한 조건 및 형식에 만족하는 데이터를 의미하며 저널논문, 뉴스기사, 책, E-mail 등 비정형 데이터는 조건 및 형식이 없는 데이터로 전체 데이터의 약 80%라는 매우 높은 비중을 차지하고 있다. 비정형 데이터는 정형 데이터에 비해 데이터양이 많다고 할 수 있지만, 그 자체로는 분석이 어렵다. 이에 수많은 연구자들은 다양한 방법을 통해 비정형 데이터를 정형화하여 의미를 제시하였으며, 그 중에서 텍스트와 관련된 연구방법들은 텍스트 마이닝(text mining)이라는 용어를 통해 설명할 수 있다(허필희, 2017). 텍스트 마이닝(text mining)은 다양한 텍스트 정보원천으로부터 자동적으로 정보를 추출함으로써 이전에 알려지지 않았던 새로운 정보를 발견하는 정보처리기술이다(김근형, 2009). 텍스트 마이닝은 데이터마이닝과 비슷한 개념이지만, 좁은 의미의 데이터마이닝은 테이블 구조의 정형화된 데이터구조에 최적화된 기술인 반면, 텍스트 마이닝은 텍스트 문서, e-메일, HTML파일과 같은 비구조화(unstructured) 또는 반구조화(semi-structured)된 데이터를 처리하는 것을 목적으로 한다. 텍스트 마이닝 기술을 이용하면 방대한 양의 데이터 속에서 의미 있는 정보를 추출해 내고, 이를 기반으로 하여 또 다른 정보와의 연계를 발견해 낼 수 있다(조성우, 2011). 온라인 리뷰(online review)는 일종의 텍스트로써 텍스트 마이닝 기술에 의하여 분석 처리될 수 있다. 텍스트 마이닝을 위하여 비구조화된 형태의 문서들이 수집되면 전처리(preprocessing)과정을 거쳐서 텍스트 분석이 수월한 형태로 변환되어야 한다. 전처리 과정에서는 불용어들이 제거되고 자연어처리기술 기반의 형태소 분석 기술을 활용하여 어근, 접두사/접미사, 품사(POS, part-of- speech) 등 다양한 언

어적 속성의 구조를 파악하는 작업이 이루어진다(김근형, 2009).

2. 텍스트 마이닝의 단계별 과정

<그림 2-1> 텍스트 마이닝의 단계별 과정



자료 : 텍스트 마이닝(Text Mining)을 이용한 관광지 이미지 결정요인에 관한 연구(심영석, 2016)에서 인용.

<그림 2-1>은 텍스트 마이닝의 단계별 과정이며 크게 자료처리과정(data processing)과 자료 분석(data analysis)으로 나눌 수 있다. 자료처리과정은 정보 검색(IR), 정보 추출(IE), 자연어처리(NLP) 등의 절차를 통하여 수집한 데이터를 가공 및 정제하는 과정이다. 자료 분석은 텍스트로부터 의미 있는 추세와 패턴 및 지식을 발견하기 위하여 데이터 마이닝, 기계학습, 통계학 등을 활용하는 과정을 말한다(Hotho et al., 2005). 텍스트 마이닝의 자료처리과정(data processing)은 ‘정보 검색 및 추출’ 과 ‘자연어처리’로 구분할 수 있다. 정보 검색은 사용자가 원하는 키워드를 기반으로 원하는 정보가 포함된 텍스트 데이터가 들어있는 문서(document)를 탐색하는 것이다. 정보 검색은 사용 목적에 따라 웹 검색(web

search), 개인 정보 검색(personal information retrieval), 기업이나 기관, 특정 영역 검색(enterprise, institutional, and domain-specific search) 등 세 가지의 형태로 구별된다(Manning et al., 2008). 정보 추출은 토큰화(tokenization), 문장 분할(sentence segmentation), 품사의 배치(part-of-speech assignment)와 독립적 개체들로서의 인식(identification of named entities)을 포함하는 일련의 과정을 거치는 것으로, 정보 검색이 사용자가 필요한 정보와 관련된 텍스트가 들어있는 특정한 문서를 찾는 것인데 반해 정보 추출이란 특정한 문서로부터 구체적인 정보를 정제하는 것을 말한다(Hotho et al., 2005). 텍스트 마이닝의 자연어처리는 구조와 형태가 복잡한 자연어를 컴퓨터로 분석하기 위해 가공하는 단계이며 한국어나 영어와 같이 문맥에 의존하는 자연어를 대상으로 하기 때문에 자료처리과정에서 자연어처리는 필수적이다. 자연어처리는 크게 ‘형태소 분석’, ‘동사 분석’, ‘의미 분석’, ‘화용 분석’으로 구분된다[1]. ‘형태소 분석’은 하나의 문장을 분해 가능한 최소한의 단위로 분리하고 분석하는 것을 말하는데 자연어처리에서 형태소 분석은 어휘 사전을 기반으로 입력문자의 형태소를 분석하는 것이다. 그리고 형태소 분석 후에는 추출된 단어들을 정형화하는 작업이 필요한데 ‘불용어 처리’, ‘어간 추출’ 등이 있다. 불용어 처리는 정관사와 같이 의미 파악이 어려운 단어 및 해당 도메인에서 사용되지 않는 단어를 처리하는 것이며, 어간 추출은 동일한 어간을 가지고 있는 단어는 하나의 단어로 처리하여 자료처리의 효율성을 높이는 것이다(김현정, 2015). ‘동사 분석’은 문장의 어순이나 문법 등이 동사 규칙에 적절한지를 분석하는 것이다. ‘의미 분석’은 문장의 의미가 적절한지를 판단하는 것이며, ‘화용 분석’은 언어의 사회적 기능 등 해당 언어의 실세계를 보는 것으로 맥락상의 적절성을 판단하는 것이다(신명철, 2005). 자료처리과정이 원시자료의 가공 단계라고 한다면 자료 분석은 데이터 마이닝, 기계학습, 통계학 등을 활용하여 의미 있는 결과를 도출하는 과정이다(Jarman, 2011). 자료처리과정이 자료 분석만큼 중요시 되지 않으나 텍스트 마이닝과 같이 비구조화 데이터를 다루는 경우는 자료 분석 못지않게 자료처리과정도 중요하기 때문에 자료처리과정상에서 다양한 자료 분석의 기법들을 적용하여 자료처리과정의 성능을 높이는 연구가 활발히 진행 중에 있다(심영석, 2016). 최근에는 이와 같은 비구조화 데이터의 자료처리와 분석을 용이하게 할 수 있도록 대부분의 주요 통계 패키지과 데이터

마이닝 프로그램들이 텍스트 마이닝 기능을 포함하고 있으며 텍스트 마이닝과 관련한 전처리(preprocess), 연관분석(associate), 군집분석(cluster), 요약(summarize), 분류(categorize)와 같은 다양한 분석 기능까지 제공하고 있다(Davi et al., 2005; Feinerer et al., 2008). 전처리(preprocess)는 데이터를 준비하고 데이터를 해당 소프트웨어에 삽입, 가공 및 정제하는 일련의 전처리 과정이다. 연관분석(associate)은 단어들 간의 동시 발생하는 빈도수를 바탕으로 연관성을 찾아내는 것이며, 군집분석(cluster)은 유사한 객체들끼리 군집하여 동일한 그룹을 묶는 것이다. 요약(summarize)은 텍스트에서 중요한 개념들을 요약한 것으로 일반적으로 빈도수가 높은 단어들을 반환하는 것이며, 범주화(categorize)는 사전 정의된 범주로 텍스트들을 분류하는 것을 의미한다(Feinerer et al., 2008).

<그림 2-2> 텍스트 마이닝이 가능한 소프트웨어의 개요

Software	Function				
	Preprocess (전처리)	Associate (연관분석)	Cluster (군집)	Summarize (요약)	Categorize (범주화)
Commercial					
Clearforest	✓	✓	✓	✓	
Copernic Summarizer	✓			✓	
dtSearch	✓	✓		✓	
Insightful Infact	✓	✓	✓	✓	✓
Inxight	✓	✓	✓	✓	✓
SPSS Clementine	✓	✓	✓	✓	
SAS Text Miner	✓	✓	✓	✓	
TEMIS	✓	✓	✓	✓	
WordStat	✓	✓	✓	✓	
Open Source					
GATE	✓	✓	✓	✓	✓
RapidMiner	✓	✓	✓	✓	✓
Weka	✓	✓	✓	✓	✓
R	✓	✓	✓	✓	✓

자료 : Text Mining Infrastructure in R(Feinerer et al., 2008)에서 인용.

<그림 2-2>를 살펴보면 상업적(commercial)으로 제공되는 텍스트 마이닝 프로그램은 마이닝과 관련된 분석들의 누락이 있는 반면 오픈소스로 제공되는 텍스트 마이닝 프로그램은 다양한 분석들이 가능한 것을 알 수 있다. R의 텍스트 마이닝 패키지는 ‘tm’(text mining package)과 한국어 텍스트 마이닝을 위한 ‘KoNLP’(Korean NLP package) 패키지 등이 있다. 이러한 패키지는 분석이 용이하며 오픈소스라는 다양한 장점들로 인하여 텍스트 마이닝 관련 연구 분야에 많이 사용된다(이동주, 2010).

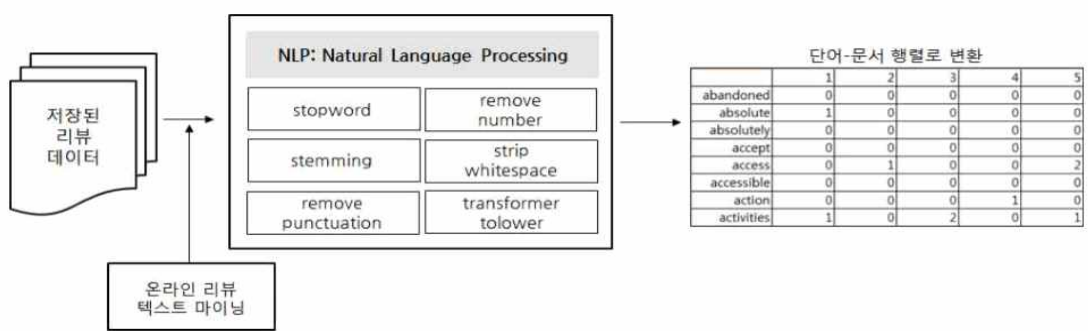
3. 텍스트 마이닝의 기법

앞서 설명한 텍스트 마이닝의 단계별 과정 중 자료처리과정의 자연어처리에서 활용되는 기법을 설명하겠다. 이것은 오피니언 마이닝에서도 기본적으로 사용되는 기법으로써 본 연구에서도 이 기법이 사용되었다.

1) 자연어처리기법

텍스트 마이닝은 텍스트와 같이 문맥에 의존하는 자연어 대상이며 비정형 데이터(unstructured data)이므로 정량적인 분석이 불가능하기 때문에 자료처리과정에서 자연어처리는 필수적이다. 따라서 텍스트로부터 정량적 분석이 가능한 구조화된 데이터(structured data)를 얻기 위해 자연어처리(NLP: Natural Language Processing)과정을 거쳐야 한다(김효원, 2017; 허필희, 2017). 자연어처리는 사람들이 사용하는 형태가 복잡한 언어를 컴퓨터가 이해하고 분석할 수 있도록 가공하는 것이므로 자연어처리가 얼마나 잘 이루어졌는지가 분석 결과에 미치는 영향이 크기 때문에 자료처리과정에서는 중요한 과정중의 하나라고 할 수 있다(김효원, 2017). <그림 2-3>는 자연어처리과정을 나타내며 자연어처리 후, 비구조화 텍스트 데이터를 구조화 데이터로 전환하는 작업을 수행한다(심영석, 2016).

<그림 2-3> 자연어처리 및 구조화 과정



자료 : 텍스트 마이닝(Text Mining)을 이용한 관광지 이미지 결정요인에 관한 연구(심영석, 2016)에서 인용.

자연어처리기법은 [1]에서 설명한 바와 같이 형태소 분석, 동사 분석, 의미 분석, 화용 분석 등 다양한 분석 방법으로 분류할 수 있으며, 이 중에서 본 연구에서 사용되는 형태소 분석에 대해 자세히 설명하겠다.

(1) 형태소 분석

형태소는 문법상 의미를 가지는 문장 구성의 최소 단위이며 형태소 분석은 문서 내의 단어, 구, 절에 해당하는 내용을 가공할 수 있는 데이터 형태로 변환하는 언어분석처리 과정으로 특정 단어와 연관된 “문서들을 신속하게 검색할 수 있도록 인덱스(Index) 파일”을 만드는 과정이다(김근형, 2009).

형태소 분석과 유사한 기법으로 문서 파싱이 있는데, 문장을 형태소 즉 명사, 조사, 부사, 동사원형 등 문법상 의미를 최소 단위로 분해(파싱)하여 주어, 목어, 부사, 동사 등의 언어학 구조를 패턴화하고 문장의 의미를 파악하는 기법이다(장지선, 2011). 문서의 내용이나 특징을 잘 반영하는 단어를 내용어(contentword)라고 한다. 내용어를 추출하기 위해서는 먼저 문장의 형태소를 분석하여 형태소별로 품사(명사, 대명사, 동사, 관사, 형용사, 부사, 접속사, 조사 등)를 결정한다(어동선, 2014). 그리고 조사, 접두사, 접미사 등을 제외하고 분석 시 의미가 있는 품사(명사, 형용사 등)의 형태소로 분류된 단어에 대해서만 추출을 시행하며 때에

따라서는 여러 문서 혹은 모든 문서에서 공통으로 나타나서 별다른 정보를 주지 못하는 불용어(stopword)들이 존재하므로 불용어 사전을 정의하여 형태소 분석 이후 단어를 추출하는 단계에서 불용어들을 제외되도록 한다(어동선, 2014). 이렇게 추출된 형태소들은 분석에 적합한 구조화된 형태의 자료로 변환하는 과정이 필요하다. 일반적으로 구조화된 자료의 형태는 문서를 나타내는 행과 키워드를 나타내는 열로 구성된 하나의 행렬(Matrix)로 표현한다. 이 행렬을 단어-문서 행렬(Term-Document Matrix)라고 한다. 행렬의 요소는 전체 문서에서 각 단어가 출현하는 빈도로 값을 구하여 나타낸다(김호원, 2017).

제 2 절 오피니언 마이닝

1. 오피니언 마이닝의 개념

인터넷에 산재한 정보는 두 가지 형태로 분류할 수 있는데 첫째는 “사실(fact)” 정보이고 둘째는 “의견(opinion)”이다. 사실은 보편적인 현상에 대해서 서술한 객관적인 정보를 의미하며 의견이란 하나의 현상에 대해서 개개인의 주관적인 정보를 의미한다. 그러나 인터넷을 검색하다보면 사실 뿐만 아니라 개개인의 의견 정보가 필요한 경우도 많이 발생한다. 예를 들어, 많은 사람들은 어떤 제품에 대한 주관적 의견(subjective opinions)을 온라인리뷰 형태로 인터넷에 올림으로써 그 제품을 구입하기 원하는 사람들의 선택에 막대한 영향을 주고 있다. 이렇게 인터넷 상에 게재된 수많은 온라인리뷰들을 모두 수작업으로 분석하려면 막대한 시간과 비용이 소요되기에 그 정보들을 자동으로 분석하고 계층화하기 위한 기법이 필요하게 되었다. 그리고 그런 목적들이 오피니언 마이닝(opinion Mining)의 발전을 가져다주고 있다(장재영, 2009). 오피니언 마이닝(Opinion Mining)은 텍스트 마이닝의 한 분야로서 텍스트를 분석하여 글쓴이의 의견이나 평가, 태도, 감정 등을 추출해내는 기법이다(김문지, 송은정, 김윤희, 2015). 오피니언 마이닝은 일정한 틀이 정해져 있지 않은 문서에서 그 문서의 주제를 찾아내는 방법인 텍스트 마이닝(Text Mining)에서 발전되었다. 텍스트 마이닝이 어떠한 문서의 주제를 찾아내고 계층화 한다면(장재영, 2009), 오피니언 마이닝은 텍스트 문서에

나타난 의견의 극성을 분석하는 감성분석이 가장 핵심적이다(김문지, 송은정, 김윤희, 2015). 어떤 의견의 주제가 무엇인지 보다는, 그 의견을 작성한 사람이 주제에 대하여 어떠한 감정을 가지고 있는가, 예를 들어 그 주제에 대하여 ‘좋은 감정을 가지고 있는가?’ 아니면 ‘안 좋은 감정을 가지고 있는가?’를 판단하여 분석한다(장재영, 2009). 현재 감성분석기법에는 크게 의견의 의미방향을 분류하는 분야와 언어적 자원을 구축하는 분야로 구분하여 연구되어져 왔다(김문지, 송은정, 김윤희, 2015). 최근에 연구되고 있는 오피니언마이닝 (opinion mining) 은 웹사이트에 게시되어 있는 온라인 고객리뷰들을 분석 대상으로 하는 텍스트마이닝 (Text Mining) 의 한 분야로서 고객 의견에 대한 긍정 (positive) 의견과 부정 (negative) 의견의 분포 등을 분석할 수 있다(김근형, 김두경, 2015). 따라서 사용자의 감성과 의견을 수치화하여 객관적인 자료로 만들기 때문에 자연어처리를 필수로 수반한다. 양정연 외(2008) 연구에서는 오피니언 마이닝을 다음의 3단계로 나누었다. 첫 번째는 자연어처리기술을 기반으로 부정 의미 수를 계산하여 누적된 정보를 보여준다. 두 번째는 사용자 평가정보를 활용하여 상품 특징별로 점수화 한다. 세 번째는 특징을 나타내는 어휘 빈도를 활용해 전체 정보의 내용을 사용자에게 전달한다(이현경, 김종배, 2017). 현재까지 오피니언 마이닝이 가장 성공적으로 적용되는 분야는 온라인 쇼핑몰에서 사용자의 상품평(product reviews)에 대한 분석이다. 실제 사용자가 작성한 상품평은 하나의 상품에 대해 사용자의 좋고 나쁨에 대한 감정을 표현한 결과이다. 따라서 개개인에 따라 긍정 (positive) 또는 부정(negative)적인 의견으로 나뉘지며 오피니언 마이닝에서는 이러한 감성분석(sentiment analysis)을 통해 상품평을 자동으로 분류 요약하여 유용한 상품 정보로 활용될 수 있다. 오피니언 마이닝 기술은 사용자들의 상품평을 효과적으로 추출하여 요약함으로써 잠재적 소비자들로 하여금 모든 상품평을 살펴보지 않더라도 상품에 대한 다양한 의견들을 쉽게 열람 및 검색할 수 있는 기반 기술을 제공한다(장재영, 2009).

2. 오피니언 마이닝의 기법

오피니언 마이닝은 기본적으로 자연어처리와 같이 텍스트 마이닝의 일부 기법과 과정이 포함되며, 이외에 자연어처리 후 텍스트에 대한 감정의 극성을 분석하는 방법으로서 본 연구에서 사용되지 않았지만 문헌적 연구로 진행했던 부분이므로 선행연구에서 제시하는 감성분석에 관한 기법에 대해 설명하겠다.

1) 감성분석

감성분석은 오피니언 마이닝에서 텍스트로 표현된 감성을 파악하여 해당 텍스트의 감성이 특정한 주제에 대해 긍정인지 부정인지 분석하는 방법이다. 감성분석의 프로세스는 분석하고자 하는 문서 혹은 문장이 긍정인지 부정인지 파악하고 특징을 추출하여 감정의 극성을 찾는 과정을 거친다(Hu et al., 2004). 이를 위해 감성분석과 관련하여 여러 선행연구에서 다양한 기법들을 제시하고 있다.

윤홍준 외(2010)의 연구에서는 상품평의 전체적인 내용이 긍정과 부정 중 어느 쪽으로 기울었는지를 가늠하기 위해 감정 극성 엔트로피(entropy)를 사용하였다.

<그림 2-4> 감정 극성 엔트로피(entropy)

$$H_d = \left(-\frac{p}{s} \log \frac{p}{s}\right) + \left(-\frac{n}{s} \log \frac{n}{s}\right)$$

자료 : 오피니언 마이닝 기술을 이용한 효율적 상품평 검색 기법(윤홍준, 2010)에서 인용.

<그림 2-4>에서 p 는 긍정적인 감정 단어의 출현 횟수, n 은 부정적인 감정 단어의 출현 횟수, 그리고 s 는 감정 단어 전체의 출현 회수를 나타낸다. 엔트로피의 값은 0에서 1 사이의 값이 나오며, 0에 가까울수록 상품평의 극성이 뚜렷함을 의미하므로 사용자의 의견이 일관성 있게 표현된 상품평임을 판단할 수 있다고 설

명한다.

최지은(2016)의 연구에서는 리뷰의 감성 정도가 병원 평가에 미치는 영향을 계량경제학적으로 분석하기 위해 감성분석에는 AFINN 영어 감성 단어 사전을 활용하였으며, 문장마다 총 4개의 감성 단계로 각 문장의 긍정 및 부정 단어의 빈도수를 도출한 후, 감성점수를 계산하였다(Nielsen, 2011). 매우 부정적(-5, -4점), 부정적(-3, -2, -1점), 긍정적(1, 2, 3점), 매우 긍정적(5, 4점) 등 4개 그룹에 해당하는 단어의 빈도수를 각 문장 별로 계산한 후, 그룹 별 감성점수 평균값을 단어의 빈도수에 곱하였다. 짧은 문장 단위의 감성점수를 추출함에 있어서는 보다 덜 세분화된 감성 수준을 활용하는 것이 더 적합하기 때문에 네 개의 그룹에 해당하는 점수를 합산하는 방식으로 문장의 감성점수를 계산하였다고 설명한다(Bromberg, 2013).

3. 오피니언 마이닝 관련 기존 연구 고찰

기존 오피니언 마이닝 관련 연구에서는 대부분 실제 온라인리뷰를 대상으로 감정의 극성을 판단하고 분류하기 위해 감성분석에 대한 기법을 제시하였다.

김근형 외(2009)의 연구에서는 텍스트 마이닝 기술을 사회과학적 관점의 연구에 적용할 수 있도록 하기 위한 방법론을 제시하였다. 온라인리뷰 데이터를 분석하고 고객특성을 도출하기 위하여 텍스트 마이닝의 문서분류기술, 범주화기술, 정보추출 기술 등을 적용하기 위한 방법론을 제안하였으며 방법론 기반의 분석사례를 제시하였다. 장재영(2009)의 연구에서는 오피니언 마이닝 기술을 이용하여 온라인리뷰 형태의 상품평 의견을 자동으로 분류할 수 있는 감성분석 알고리즘을 제안하였다. 이 알고리즘은 온라인 쇼핑몰에 등록된 개별 상품평을 대상으로 긍정의견 및 부정의견을 판단하여 요약된 결과를 제공하는 기능을 수행하였다. 홍태호 외(2014)의 연구에서는 온라인상점(amazon.com)에서의 영화평을 대상으로 문서수준의 용어정보를 추출한 후 정보력이 높은 용어들을 대상으로 문서빈도, 정보획득량, 카이제곱 통계량을 도출하는 오피니언 마이닝에 대한 연구를 하였다. 윤홍준 외(2010)의 연구에서는 대량의 상품평을 검색할 때 검색자의 목적

에 최적화된 검색 우선순위를 부여하는 기법을 제안하였다. 상품평은 감정적이며 주관적인 의견을 포함하고 있기 때문에 상품평 검색은 일반 웹 검색과는 달라야 할 것이라는 아이디어를 기반으로 제시하고 있다. 웹 검색이 사용하는 사용자 검색어뿐만 아니라 상품평 내의 주관적인 의견의 포함 여부 및 감정 극성의 엔트로피 등을 고려하여 상품평의 검색 우선순위를 판단하는 기법을 제안하였다. 양가임 외(2011)의 연구에서는 사용자의 상품평 검색 시 효율적 상품 추천을 위해 오피니언마이닝 기반의 상품평 검색 모델을 제안하였다. 자연어처리기법인 POS 파싱을 통해 상품평의 어휘 정보와 상품의 특성을 추출하고 이를 TF-PMI 알고리즘, 감정 분석 엔트로피를 활용하여 상품평 내용의 의미 극성 정보를 바탕으로 상품 특성의 점수를 계산하는 상품평 검색 기법을 연구하였다. 강선아(2016)의 연구에서는 ‘서울시 무상급식’ 관련 기사의 댓글에 대해 형태소 분석 및 텍스트 마이닝을 활용해 감성 사전을 구축하고 이를 기반으로 댓글의 의미 극성을 분류하는 알고리즘을 제안하였다. 형태소 분석 시에는 서울대 자체 개발 한글 형태소 분석기인 ‘꼬꼬마 형태소 분석기’를 활용하였다. 댓글의 공감정도에 따른 가중치를 부여하여 TF(Term Frequency) 값과 계산된 어휘 극성 값을 바탕으로 구축된 감성 사전을 활용하여 댓글의 긍정/부정을 자동 분류하는 알고리즘을 제안하였다. 김문지 외(2015)의 연구에서는 온라인리뷰 내 콘텐츠 만족도를 도출하는 오피니언마이닝 기반의 콘텐츠 만족도 분석 시스템을 구축하였다. 기존 의견 극성만 분류하는 것에서 보다 세밀한 분류를 위해 HDFS, MapReduce 기반 사전 구축기를 활용하여 수집된 리뷰의 감정 극성에 따른 콘텐츠 별 감성사전을 구축하고 감성분석 알고리즘으로 그 어휘에 따라 계산된 극성 값을 적용하여 만족도를 도출하는 시스템을 구현하였다. 김진옥 외(2011)의 연구에서는 오피니언 마이닝 기반으로 텍스트에 대한 극성 분석에 초점을 맞추어 한글 텍스트의 의견 극성을 판단하고 분류하는 알고리즘을 제안하였다. Bootstrapping 알고리즘, AutoSlog-TS 알고리즘을 적용하여 형용사의 특징에 따른 어휘의 주관성을 분석한 결과로 어휘 사전을 구축하고 텍스트의 감정 극성을 분류하기 위해서 긍정, 중립, 부정 훈련데이터 각각에 대하여 어휘의 빈도수 기반으로 점수가 부여된 긍정, 중립, 부정 패턴을 추출하는 방안을 제안하였다. 김근형 외(2015)의 연구에서는 온라인리뷰 내 형용사들의 유의미한 규칙을 추출하기 위해 품사태깅

(Part-Of-Speech Tagging) 기법을 이용하여 제품의 특징과 고객 의견의 품사들을 분석해 어휘를 추출하고 연관규칙탐사(association rule mining) 기법을 적용하여 제품의 특징과 의견 사이의 연관성을 분석하는 방법론을 제시하였다. 김근형 외(2015)의 연구에서는 오피니언 마이닝을 사회과학연구방법에 적용하기 위한 확장형 오피니언 마이닝 시스템을 설계하였다. 구문 분석기에 의해 생성된 구문구조트리 파일로부터 추출한 어휘에 대해 어휘 온톨로지의 정보를 참조하여 빈발개체 추출 모듈, 5점 척도 기반의 극성계산 모듈을 통해 개념변수의 긍정/부정 정도 값을 계산하고 이에 개념변수들 사이의 상관분석 및 회귀분석을 수행하는 모듈을 적용한 시스템을 구축하였다. 안휘훈 외(2014)의 연구에서는 쇼핑물에 올라온 상품평 중에서 반복되는 의도적으로 남긴 상품평을 추출하기 위해 TF-IDF를 이용하여 각 상품평 별로 나오는 단어의 중요도를 판단하고 TF-IDF에서 얻은 벡터 값을 부여하여 각 상품평의 유사도를 비교해 나온 코사인 유사도 값에 따라 의도적인 상품평을 걸러내는 방법을 제안하였다. 송종석(2010)의 연구에서는 오피니언 마이닝을 이용한 극성 분류 시 분류 성능이 긍정/부정 사전의 어휘적 자원에 많은 영향을 받으므로 어휘적 자원의 수준을 높이는 것이 매우 중요하다고 하였다. 이에 상품평 분류 성능 향상을 위하여 도메인별 제품 특징을 고려한 긍정/ 부정 사전을 자동으로 구축하는 방법을 제안하였다. 제안 방법으로 연구한 결과에서 어휘의 극성이 일반적으로 부여된 상태보다는 도메인의 극성을 고려하여 감성사전을 구축하였을 때 더 높은 성능을 보였으며 접속정보를 활용하는 방법으로 제품 특징을 고려하여 긍정/부정 사전을 구축할 때 성능이 향상되었음을 확인하였다.

앞서 선행연구들을 살펴보면 온라인리뷰 데이터의 극성 값을 감성분석 알고리즘을 활용해 계산하고 이를 토대로 리뷰의 극성을 판단하고 분류하는 연구들이 주로 이루어졌다. 그러나 온라인리뷰라는 것은 텍스트 형태의 주관적이고 감성적인 리뷰 데이터뿐만 아니라 평점과 같이 제품 및 서비스에 대한 정량적인 평가도 포함하기 때문에 이러한 관점들을 균형적으로 고려하여 정성적 평가와 정량적 평가 간의 회귀적 관계를 분석하고 이를 바탕으로 보다 정확하고 합리적인 평가기법을 제시하는 연구들이 이루어져야 할 필요성이 제기된다. 따라서 본 연구에서는 연구 데이터 및 일부 선행연구들을 근거하여 연구 가설을 수립하고 이

를 실증적으로 분석함으로써 제품 및 서비스에 대한 정량적 관점의 평가와 정성적 관점의 평가에 대한 회귀적 영향 관계를 파악하고자 한다. 그리고 분석 결과를 바탕으로 랭킹모델을 도출하며 이를 기반으로 분석한 사례를 제시할 예정이다.

제 3 절 온라인리뷰에 대한 선행연구

기존 오피니언 마이닝 관련 연구 외에도 온라인리뷰의 유용성이나 신뢰성에 관해 가설 검증한 연구들이 존재하였는데, 이에 대한 선행연구들을 살펴보겠다.

이승준(2012)의 연구에서는 온라인리뷰의 영향력이 제품의 특성과 유형에 따라 달라질 수 있다는 견해를 제시하면서 아마존(Amazon.com)이라는 전자상거래에서 취급하는 제품 및 서비스 중 가장 관여도가 높은 제품인 DSLR 카메라 관련 온라인리뷰를 대상으로 구매의사결정에 있어 유용한 것인지에 대해 리뷰의 길이, 평점, 작성 경과일, 실명 인증, 구매 인증 등의 변수로 하여 가설을 설정하고 이에 상관분석 및 회귀분석을 통해 검증하였다. 그 검증 결과에서 리뷰의 길이뿐만 아니라 평점, 인증 정보와 같은 주변적인 정보도 온라인리뷰의 유용성에 영향을 미칠 수 있음을 확인하였다. 박은주 외(2013)의 연구에서는 화장품 구매와 관련하여 소비자가 리뷰를 탐색하는 동기, 리뷰 탐색 과정에서 리뷰의 특성이나 객관성에 대한 신뢰 등이 어떻게 작용하는지에 대해 가설을 설정하고 관련 선행연구를 뒷받침하여 화장품 구매 경험이 있는 대학생들을 대상으로 실시한 설문조사 데이터를 측정모형, 구조모형 분석으로 검증하였다. 그 검증 결과에서 리뷰의 전문성이나 사진, 동영상 같은 시각적인 정보들이 온라인리뷰에 대한 신뢰를 형성하고 이를 매개로 화장품 온라인 구매의도에 영향을 미칠 수 있다는 것을 확인하였다.

제 4 절 랭킹모델에 대한 선행연구

기존 검색 모델의 한계를 해결하기 위해 랭킹을 부여하여 보다 정확하고 합리적인 검색 결과를 도출하고 이에 적용할 방법으로 랭킹모델을 제시하는 것인데, 이에 대한 선행연구들을 살펴보겠다.

황원석 외(2013)의 연구에서는 사용자가 기존 검색 엔진으로 논문 검색 시 많은 양의 논문이 검색 결과로 반환되고 품질이 높은 논문을 찾기 쉽지 않은 문제점을 해결하기 위해 논문들에 높은 랭크를 부여하는 랭킹 방법인 ArtRank를 제안하였다. 이에 논문의 참조 수, 학회의 평판, 논문 간의 상대적 과소평가 방지 등을 모두 고려하기 위해 논문간의 참조관계를 Random Walk with Restart에 이용, 학회의 평판을 측정할 방법으로 임팩트팩터를 제안하고 특히, 논문 데이터의 연도별 특성을 분석하여 이를 랭킹 모델에 적용함으로써 최근 연구 동향을 반영하는 논문들이 높은 랭킹을 부여 받도록 하였다. 오선주(2013)의 연구에서는 소셜 네트워크에서 잠재되어 있는 관계를 인식하여 유용한 관계를 찾기 위한 방안으로서 소셜 네트워크에서 구성원 간 관계를 검색하기 위한 랭킹 방법을 제안하였다. 특히, 온톨로지를 기반으로 개체 간 의미적 관계를 유추하여 확장하고 이를 바탕으로 다양한 랭킹 기준을 융통성 있게 조합하여 검색하고자 하는 관계를 효율적으로 찾기 위한 랭킹모델을 제시하였다.

제 3 장 연구 설계

제 1 절 연구 데이터 선정

기존 선행연구에서 다루었던 온라인리뷰 데이터들의 유형을 보면 주로 온라인 쇼핑몰에 게시된 상품평인 경우가 많았다. 이외에도 관광리뷰를 대상으로 온라인 리뷰에 내포된 감정 극성을 판단하고 분류하는 연구 사례(심영석, 2016)가 있었으나, 보다 정확하고 합리적인 평가기법으로 랭킹모델링을 제시하지 않았다. 이를 근거로 본 연구에서는 관광리뷰를 분석 대상으로 선정하였다.

제 2 절 연구 모형

실증분석에 앞서 연구 가설을 수립하기 위해 연구 데이터 및 일부 선행연구들을 근거하여 온라인리뷰에 따른 정성적 관점의 평가 요인 및 정량적 관점의 평가 요인 변수로 긍정의견, 부정의견, 조정의견, 정량점수, 리뷰길이를 채택하였으며 연구 모형은 <그림 3-1>과 같다. 다음은 각 변수의 채택 근거를 설명한 것이다.

1. 긍정의견 및 부정의견

장재영(2009)은 실제 사용자가 작성한 온라인리뷰는 제품 및 서비스에 대해 사용자의 좋고 나쁨에 대한 감정을 표현한 결과라고 설명한다. 이것은 개개인에 따라 긍정 또는 부정적인 의견으로 구분되며 오피니언 마이닝에서는 이러한 감성 분석을 통해 온라인리뷰를 분류하고 요약하여 유용한 정보로 활용될 수 있다는 것이다. 온라인리뷰는 일반적인 문서보다 길이가 짧고 제품 및 서비스의 여러 속성과 감정적인 의견이 많이 포함되어 있는데, 제품 및 서비스의 여러 속성에 대해 다양한 의견을 표현할 수 있음을 디지털 카메라의 예를 들어 설명하였다. 따

라서 온라인리뷰에 대한 감정의 전반적인 극성도 중요하지만 각 속성에 대한 극성을 판단하는 것도 중요한 것임을 시사하고 있다. 이를 근거로 온라인리뷰에는 속성이나 감정적인 의견이 형용사로 표현되는 경우가 많아 극성을 판단하고 분류하는데 있어 형용사가 그 의미를 가장 잘 반영한다는 점에서 본 연구에서는 긍정의견(긍정 형용사 개수), 부정의견(부정 형용사 개수)을 변수로 채택하였다.

2. 조정의견

윤홍준(2010)은 상품평 검색 시 랭킹 선정 방식 중의 하나로 긍정/부정 등 극성에 따른 감정의 폭을 정량화하기 위해 조정의견을 계산하는 방법으로 감정 극성 엔트로피(entropy)를 제시하였다. 이것은 상품평의 평가가 긍정과 부정 중 어느 한 방향으로 기울었는지 가늠하는 선정 방법으로서 엔트로피의 값에 따라 상품평의 극성이 선명한 정도를 파악하는 것이다. 이를 근거로 조정의견을 계산함으로써 긍정/부정의 극성에 따른 기울기를 파악하여 제품 및 서비스에 대해 리뷰 작성자가 어떤 감정을 더 느꼈는지 정확히 알 수 있을 것으로 판단한다. 따라서 본 연구에서는 조정의견을 변수로 채택하고 앞서 선행연구에서 제시한 기법과 다르게 긍정의견에서 부정의견을 상쇄해서 계산한다. 예를 들어 “성산일출봉은 올라가는데 힘들었지만 정상에서의 경치는 아름다웠다.”라는 리뷰 데이터의 경우, 긍정의견은 1(아름다웠다)이고 부정의견도 1(힘들었지만)이면 조정의견은 0이 된다. 만약, 조정의견 계산 시 음(-)의 정수가 나올 경우, 긍정 보다 부정의 감정을 더 많이 느꼈을 것이고 양(+)의 정수가 나올 경우, 부정 보다 긍정의 감정을 더 많이 느꼈다고 판단할 수 있다.

3. 리뷰길이

이승준(2012)은 온라인리뷰는 제품 및 서비스에 대한 정보들이 포함되어 있다는 점에서 구매의사결정의 중심적인 단서가 될 수 있으며, 문자로 작성되고 읽혀지기 때문에 단어 수가 많이 사용된 긴 리뷰일수록 더 많은 정보를 포함할 것이라고 설명한다. 다수의 선행연구에서는 리뷰에 포함된 정보의 양을 측정하는 방법

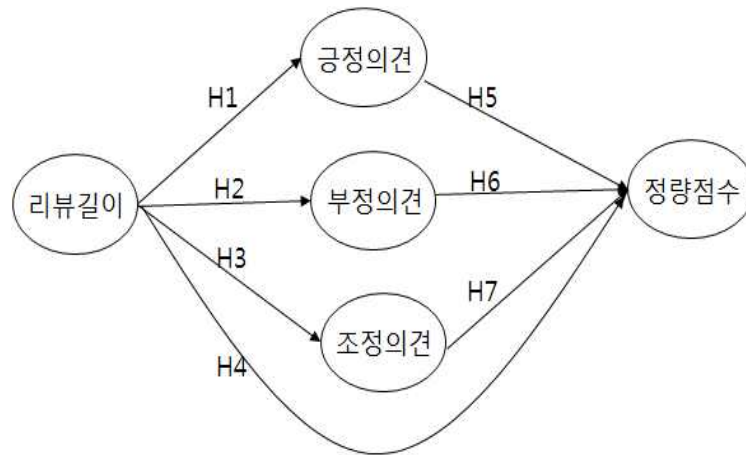
으로 리뷰에 사용된 단어의 수 즉, 리뷰길이를 사용하고 있다. 리뷰길이에 관련하여 Hedonic Marking Hypothesis의 인용에서는 정보 처리의 용이성 즉, 이해하기 쉽게 작성한 리뷰일수록 대상에 대한 평가판단에 긍정적인 영향을 미칠 수 있다고 설명하고 있다. 백현미 외(2011)의 인용에서는 부정적인 의미를 나타내는 단어들이 많이 포함된 리뷰가 더욱 유용하게 지각되는 것으로 나타나 온라인리뷰를 정보 원천으로 사용하는 소비자들에게 부정 편향이 존재함을 확인하였다고 설명하고 있다. 이를 근거로 온라인리뷰에는 제품 및 서비스에 대한 정보와 함께 극성에 따른 감정적인 의견을 포함하는 정도에 따라 리뷰길이에도 영향을 미칠 수 있다는 점에서 본 연구에서는 리뷰길이를 변수로 채택하였다. 앞서 선정된 온라인리뷰 데이터를 보면 길이가 최소 100자 이하, 최대 500자 이상으로 확인되었는데, 이때 리뷰의 길이 정도를 구분하기 위해 250자를 기준으로 정하였다. 예를 들어 250자 이하는 대체로 짧고 250자 이상은 대체로 긴 리뷰인 것으로 보았다.

4. 평가점수

이승준(2012)은 리뷰의 정량점수(평점)는 제품 및 서비스에 대한 리뷰 작성자의 선호를 반영하고 있기 때문에 리뷰의 유용성에 영향을 미치는 요인이 될 수 있다고 설명한다. 정량점수에 관련하여 Chevalier and Mayzlin(2006)의 인용에서는 온라인리뷰의 정량점수는 대체로 긍정적인 경향이 있으며 부정적인 정량점수(1점)의 리뷰가 긍정적인 정량점수(5점)의 리뷰보다 더 큰 영향력을 가질 수 있다고 설명하고 있다. Sridhar and Srinivasan(2012)의 인용에서는 다른 소비자의 정량점수는 개인의 제품 및 서비스에 대한 경험이 정량점수에 미치는 영향을 조절하는 것으로 나타나 온라인리뷰 정량점수의 사회적 영향력을 입증하였다고 설명하고 있다. 백현미 외(2011), Ganetal(2012)의 인용에서는 평균적인 리뷰 정량점수와의 차이가 적은 정량점수의 리뷰가 더욱 유용한 것으로 받아들여지는 것으로 나타났는데, 이는 소비자들이 리뷰에 대한 유용성을 판단하는데 있어서 평균적인 정량점수에 가까운 리뷰를 더욱 유용한 것으로 지각하고 있음을 의미한다고 설명하고 있다. 이를 근거로 제품 및 서비스에 대한 작성자의 경험 및 선호와 함께 온라인리뷰에 표현되는 감정적인 의견의 극성에 따라 정량점수에도 영향을

미칠 수 있다는 점에서 정량점수를 변수로 채택하였다.

<그림 3-1> 연구 모형



제 3 절 연구 가설

제품 및 서비스에 대하여 텍스트 기반의 정성적 평가를 할 때 긍정적 감정이면 텍스트를 정성들여 길게 작성할 것이라는 관점에서 다음과 같이 리뷰길이를 독립변수, 긍정의견을 종속변수로 설정하여 가설 H1을 수립하였다.

H1. 리뷰길이가 긴 리뷰일수록 긍정적일 것이다.

제품 및 서비스에 대하여 텍스트 기반의 정성적 평가를 할 때 부정적 감정이면 불만사항에 대한 구체적인 문제점을 길게 작성할 것이라는 관점에서 다음과 같이 리뷰길이를 독립변수, 부정의견을 종속변수로 설정하여 가설 H2를 수립하였다.

H2. 리뷰길이가 긴 리뷰일수록 부정적일 것이다.

제품 및 서비스에 대하여 텍스트 기반의 정성적 평가를 할 때 긍정적 감정에서 부정적 감정을 상쇄한 조정의견은 리뷰길이와 정비례 관계에 있을 것이라는 관점에서 다음과 같이 리뷰길이를 독립변수, 조정의견을 종속변수로 설정하여 가설 H3을 수립하였다.

H3. 리뷰길이는 조정의견에 영향을 미칠 것이다.

제품 및 서비스에 대하여 텍스트 기반의 정성적 평가를 할 때 제품 및 서비스에 대한 애착이 있으므로 정량적 평가에서 정량점수를 높게 부여할 의지라면 리뷰를 길게 작성할 것이라는 관점에서 다음과 같이 리뷰길이를 독립변수, 정량점수를 종속변수로 설정하여 가설 H4를 수립하였다.

H4. 리뷰길이가 긴 리뷰일수록 정량점수가 높을 것이다.

제품 및 서비스에 대하여 텍스트 기반의 정성적 평가를 할 때 긍정적 감정이면 정량적 평가에서 정량점수를 높게 줄 것이라는 관점에서 다음과 같이 긍정의견을 독립변수, 정량점수를 종속변수로 설정하여 가설 H5를 수립하였다.

H5. 긍정적인 리뷰일수록 정량점수가 높을 것이다.

제품 및 서비스에 대하여 텍스트 기반의 정성적 평가를 할 때 부정적 감정이면 정량적 평가에서 정량점수를 낮게 줄 것이라는 관점에서 다음과 같이 부정의견을 독립변수, 정량점수를 종속변수로 설정하여 가설 H6을 수립하였다.

H6. 부정적인 리뷰일수록 정량점수가 낮을 것이다.

제품 및 서비스에 대하여 텍스트 기반의 정성적 평가를 할 때 긍정적 감정에서 부정적 감정을 상쇄한 조정의견은 정량적 평가의 정량점수와 정비례 관계에 있을 것이라는 관점에서 다음과 같이 조정의견을 독립변수, 정량점수를 종속변수로

설정하여 가설 H7을 수립하였다.

H7. 조정의견은 정량점수에 영향을 미칠 것이다.

제 4 절 변수 측정

연구 데이터 및 일부 선행연구들을 근거하여 변수들을 측정하였다. 각 변수들에 대한 설명과 측정방법은 <표 3-1>에 요약되어 있다.

<표 3-1> 변수 측정

변수	변수 설명 및 측정 방법	선행 연구
긍정의견	리뷰의 긍정 형용사 개수.	장재영(2009)
부정의견	리뷰의 부정 형용사 개수.	장재영(2009)
조정의견	긍정의견에서 부정의견을 상쇄한 평가의견 [긍정의견 - 부정의견].	윤홍준(2010)
리뷰길이	리뷰에 사용된 단어의 수.	이승준(2012)
정량점수	리뷰 작성자가 리뷰의 대상이 되는 제품 및 서비스에 대하여 부여한 평점 [1점 - 5점].	이승준(2012)

제 5 절 분석 절차

제품 및 서비스에 대한 정성적 관점의 평가와 정량적 관점의 평가에 대한 회귀적 영향 관계를 파악하기 위해 실증적인 분석 절차로 <그림 3-2>과 같이 설계하였다. 구체적인 절차는 다음과 같다.

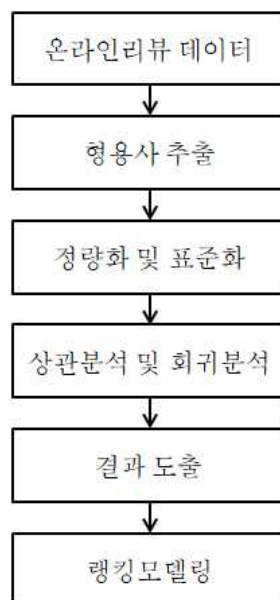
첫째, 기존 수집된 Tripadvisor의 제주 명소 관련 리뷰 데이터로부터 형용사를 추출한다.

둘째, 추출된 형용사를 포함해 리뷰 데이터를 정형화하고 이를 표준화하여 가공 처리한다.

셋째, 수립된 가설들을 검증하기 위해 오피니언 마이닝 기법으로 가공된 데이터를 가지고 상관분석 및 회귀분석을 실시한다.

넷째, 분석 결과를 바탕으로 랭킹모델을 도출하고 이를 기반으로 분석한 사례를 제시한다. 마지막으로 분석 사례를 통한 시사점을 제시한다.

<그림 3-2> 분석 절차



제 4 장 실증분석 및 제안

제 1 절 데이터 원천

본 연구에서는 관광리뷰 중 제주 명소 관련 리뷰 데이터를 분석 대상으로 선정하였다. 이것은 유상욱(2017)의 연구에 사용된 기존 데이터로서 데이터 수집 당시, 세계적인 관광 웹사이트인 Tripadvisor 게시된 256개의 제주 명소 중 리뷰가 있는 195개의 관광지를 대상으로 삼았고 사용자가 관광지별로 리뷰제목, 리뷰내용, 리뷰작성일, 방문날짜, 평점, 계절로 작성한 리뷰를 수집 대상으로 삼았다. Tripadvisor의 리뷰 작성 예시는 <그림 4-1>과 같다.

<그림 4-1> Tripadvisor의 리뷰 작성 예시



자료 : 다차원 순위 모델 기반의 관광후기 분석(유상욱, 2017)에서 인용.

제 2 절 데이터 수집

유상욱(2017)의 연구에 따르면 데이터 수집 당시, Tripadvisor에 게시된 256개의 제주 명소 중 리뷰가 있는 195개의 관광지에 대한 온라인리뷰 데이터를 웹 크롤러 기법을 이용하여 수집하였다고 설명하고 있다. 데이터 수집을 위한 절차로는 <그림 4-2>과 같고 상세 내용은 다음과 같다.

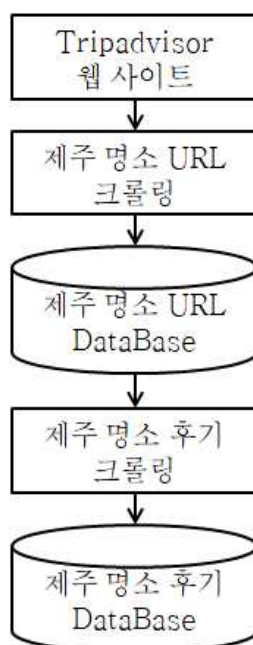
첫째, 웹 크롤러를 이용하여 제주 명소 리스트가 나와 있는 웹 페이지를 Tripadvisor의 웹 서버에 요청한다.

둘째, 응답 받은 웹 페이지에서 리스트의 앵커 태그에 설정된 URL을 크롤링하여 제주 명소 URL Database에 저장한다.

셋째, 제주 명소 URL Database에 저장된 URL을 한 개씩 불러와서 지정된 웹 페이지를 Tripadvisor의 웹 서버에 요청한다.

넷째, 응답 받은 웹 페이지에서 온라인리뷰 데이터만을 크롤링하여 제주 명소 리뷰 Database에 저장한다.

<그림 4-2> 데이터 수집 절차



자료 : 다차원 순위 모델 기반의 관광후기 분석(유상욱, 2017)에서 인용.

본 연구에서는 보다 정확한 실증분석을 위해 당시 수집된 데이터 중에서 리뷰 개수가 30개 이상 존재하는 관광지에 대한 2697개의 온라인리뷰 데이터를 분석 대상으로 선정하였다. 온라인리뷰 데이터의 스키마(schema)는 다음 <표 4-1>과 같이 관광지명, 리뷰제목, 리뷰내용, 일련번호, 리뷰작성일, 평점, 계절 등으로 구성되었다.

<표 4-1> 온라인리뷰 데이터

관광지명	리뷰제목	리뷰내용	일련번호	리뷰작성일	평점	계절
성산 일출봉	두번 가봤는데 너무	다 좋았네요.	1	2016/08	5	여름
성산 일출봉	역시 해돋이 봐야죠	해돋이 보거나 혹은 해가지고나서 주변 야경을	1	2016/05	5	봄
성산 일출봉	힘들지만 곳	부도 좋았고 멋진곳입니다.	1	2016/03	5	봄
성산 일출봉	중귀런너무많음	정말 이것만 보면 예쁜데, 중국인 관광객들이 너무	1	2016/02	5	겨울
성산 일출봉	세해맞이 성산일출봉	하였습니다. 갈때마다 느끼지만 웅장한 자연경	1	2016/12	5	겨울
성산 일출봉	유네스코 자연유산의	올라가는데 20분이면 충분하고 정말 힘들다 싶	1	2016/10	5	가을
성산 일출봉	한해의 끝과 시작		1	2016/12	4	겨울
성산 일출봉	제주도에서 반드시 들	성산 일출봉은 일출 보기로도 유명한 곳이다. 위	1	2016/12	4	겨울
성산 일출봉	가족방문기	유채꽃사진 찍는곳이 돈을받아 눈살이 찌푸려	1	2016/12	5	겨울
성산 일출봉	제주의 절경	제주에서도 이름난 곳이라 찾는 사람도 많네요.	1	2016/03	5	봄

제 3 절 데이터 가공

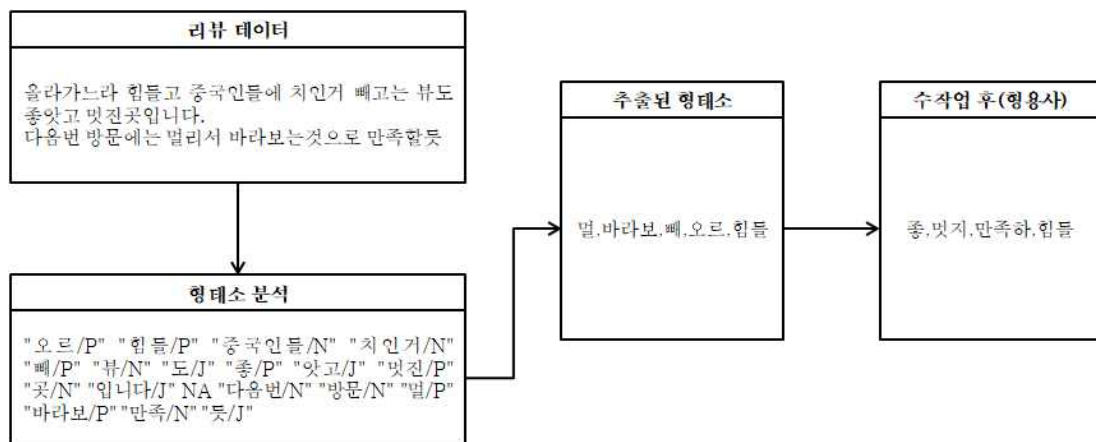
본 연구에서는 실증분석에 활용될 수 있도록 비정형적인 온라인리뷰 데이터를 오피니언 마이닝 기법으로 가공 처리하기 위해 소프트웨어는 R을 이용하였으며, ‘tm’, ‘KoNLP’, ‘stringr’, ‘sqldf’, ‘plyr’ 등의 패키지를 사용하였다.

1. 형용사 추출

온라인리뷰에 표현되는 속성이나 감정적인 의견에 따라 극성을 파악하는데 그 의미를 가장 잘 반영하는 품사로 형용사를 추출하고자 하였다. 먼저 온라인리뷰 데이터의 형태소 분석을 위해 KoNLP 패키지의 SimplePos22() 함수를 사용하였으며, 분석 결과에서 형용사에 해당하는 형태소를 추출하기 위해 stringr 패키지

의 `str_match()` 함수를 사용하였다. `SimplePos22()` 함수는 KAIST에서 개발한 한글 형태소 분석 함수이다. 그리고 추출 결과에서 결측치(NA), 불용어를 제거하기 위해 `lis.na()` 함수, `gsub()` 함수를 사용하였다. 그런데 `SimplePos22()`, `str_match()` 등 함수의 성능 문제로 인하여 형용사 추출 시 형용사가 아닌 다른 품사의 형태소도 함께 추출되거나 형용사를 모두 추출하지 못하는 문제가 있었다. 이런 이유로 본 연구에서는 추출된 형태소 중에서 형용사에 해당하는 형태소만 색출하고 이외 추출되지 않은 형용사를 리뷰 데이터에서 색출하는 일부 수작업을 수행하였다. 이 과정은 다음 <그림 4-3>과 같다.

<그림 4-3> 형용사 추출 과정



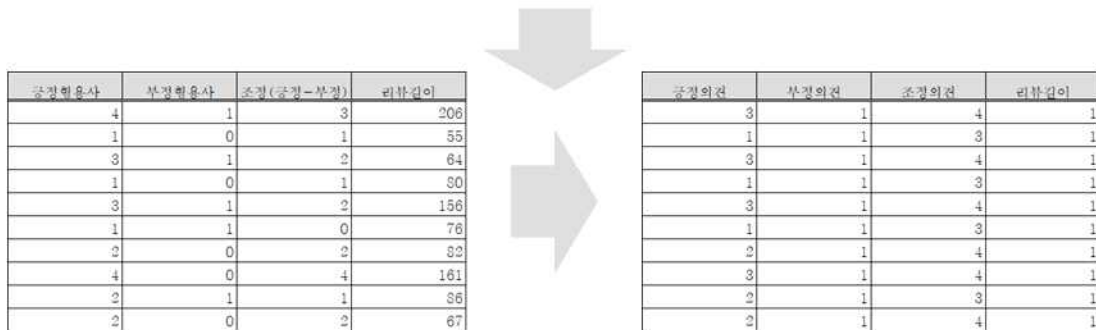
2. 데이터 정형화 및 표준화

온라인리뷰 데이터를 정형화하기 위해 앞서 추출된 형용사들의 극성에 따라 긍정/부정으로 분류하여 각각 긍정의견(긍정 형용사 개수), 부정의견(부정 형용사 개수)을 계산하고 이들을 상쇄하여 조정의견을 계산하였으며 R에서 `nchar()` 함수를 사용하여 각 리뷰의 길이도 계산하였다. 이후 정형화된 데이터들은 표준화된 수치로 변환하였다. 표준화는 각 데이터를 1~5 사이의 값으로 변환하기 위해 각 칼럼별로 그 값의 범위 안에서 기준을 다르게 정하였는데, 예를 들어 긍정의견의 경우 3을 기준으로 3 이하는 기본값으로 하고 4 이상은 3~5 사이의 값으로 변환

하였다. 리뷰길이의 경우 250자를 기준으로 250자 이하는 1, 250자 이상은 2~5 사이의 값으로 변환하였다. 이 과정은 다음 <그림 4-4>와 같다.

<그림 4-4> 데이터 정형화 및 표준화

리뷰내용	형용사
이런 이름으로 부르면 정말 싫어하는데 이정도네요.	별로 형용할 수 없.이런지 안.시원하.좋.힘들
레플레 보거나 혹은 레가직고나서 주변 야경을 볼때가 좋았고 멋진 곳입니다.	아깝지 않 좋.멋져.만족하.힘들
정말 이것만 보면 예쁘데, 중국인 관광객들이 너무 가네요. 갈때마다 느끼지만 웅장한 자연경관에 항상 모	예쁘 웅장하.나르.멋있.비싸
승란가는데 20분이면 충분하고 정말 힘들다 싶을때 정	멋져.힘들 적당하.좋
설악 일출복은 일출 보기로도 유별한 곳이다. 유명한 유재철사진 찍는곳이 도움받아 눈살이 찌푸려집니다	유명하.멋져.매력.좋 좋은살이 찌푸리.아름답
제주에서도 이름난 곳이라 찾는 사람도 많네요. 분에	아름답.이름나



제 4 절 연구 가설의 검증

1. 데이터 가공 결과

본 연구에서는 기존 수집된 데이터 중에서 리뷰 개수가 30개 이상 존재하는 관광지에 대한 2697개의 온라인리뷰 데이터를 분석 대상으로 선정하였다. 관광지명, 리뷰제목, 리뷰내용, 일련번호, 리뷰작성일, 평점, 계절 등의 스키마(schema)로 구성된 온라인리뷰 데이터를 오피니언 마이닝 기법으로 가공 처리하였다. 그 결과 <표 4-2>과 같이 긍정의견, 부정의견, 조정의견, 리뷰길이, 정량점수 등의 스키마(schema)로 구성된 데이터를 도출하였다.

<표 4-2> 가공된 데이터

관광지명	리뷰제목	리뷰작성일	계절	긍정의견	부정의견	조정의견	리뷰길이	정량점수(평점)
성산 일출봉	두번 가봤는데 너무 좋아요	2016/08	여름	3	1	4	1	5
성산 일출봉	역시 해돋이때 봐야죠	2016/05	봄	1	1	3	1	5
성산 일출봉	힘들지만 곳	2016/03	봄	3	1	4	1	5
성산 일출봉	꽃피면너무말아요	2016/02	겨울	1	1	3	1	5
성산 일출봉	새해맞이 성산일출봉	2016/12	겨울	3	1	4	1	5
성산 일출봉	유네스코 자연유산의 경	2016/10	가을	1	1	3	1	5
성산 일출봉	한해의 끝과 시작	2016/12	겨울	2	1	4	1	4
성산 일출봉	제주도에서 반드시 둘러	2016/12	겨울	3	1	4	1	4
성산 일출봉	가족방문기	2016/12	겨울	2	1	3	1	5
성산 일출봉	제주의 절경	2016/03	봄	2	1	4	1	5

2. 분석 결과

본 연구에서는 앞서 수립된 가설들을 검증하기 위해 통계 분석 프로그램으로 SPSS를 이용하여 상관분석 및 회귀분석을 실시하였으며 다음과 같은 분석 결과를 도출하였다. 이에 사용된 데이터는 앞서 가공된 데이터를 활용하였다.

1) 변수 간의 상관분석

연구 모형에 설정된 변수들에 대해 서로의 관계가 어떤 방향이고 어느 정도의 관계를 갖고 있는지 알아보기 위하여 상관분석을 실시하였으며 그 결과는 다음 <표 4-3>와 같다.

<표 4-3> 변수 간의 상관분석

변수	평균	표준편차	1	2	3	4	5
정량점수	4.26	.847	1				
긍정의견	1.83	.847	.243***	1			
부정의견	1.12	.407	-.276***	-.125***	1		
조정의견	3.38	.677	.370***	.765***	-.508***	1	
리뷰길이	1.06	.295	.015	.151***	.152***	.019	1

정량점수는 리뷰길이를 제외한 모든 변수들과 유의한 상관관계를 보이고 있다. 이는 긍정의견, 부정의견, 조정의견 등이 정량점수에 영향을 미칠 가능성을 시사하고 있다. 특히, 긍정의견과는 조정의견이 상관계수가 7 이상으로 비교적 강한

정(+)²의 상관관계를 보이고 있다. 부정의견과는 정량점수, 긍정의견, 조정의견 등이 음(-)의 상관관계를 보이고 있다. 리뷰길이는 긍정의견 및 부정의견과는 유의한 상관관계를 보이고 있으나 정량점수 및 조정의견과는 상관관계의 유의성을 나타내고 있지 않았다.

2) 연구 가설 H1의 단순회귀분석

연구 가설 H1을 검증하기 위하여 리뷰길이를 독립변수, 긍정의견을 종속변수로 두어 단순회귀분석을 실시하였다.

<표 4-4> 연구 가설 H1의 단순회귀분석

변수	B	표준오차	Beta	t-value	Sig.t	공차	VIF
리뷰길이	.434	.055	.151	7.939	.000***	1.000	1.000
R ² = .023, F=63.025							

<표 4-4>와 같이 리뷰길이는 긍정의견에 $p < 0.001$ 수준에서 정(+)²의 영향을 미치고 있음을 알 수 있다. 가설 H1은 채택되었다.

3) 연구 가설 H2의 단순회귀분석

연구 가설 H2를 검증하기 위하여 리뷰길이를 독립변수, 부정의견을 종속변수로 두어 단순회귀분석을 실시하였다.

<표 4-5> 연구 가설 H2의 단순회귀분석

변수	B	표준오차	Beta	t-value	Sig.t	공차	VIF
리뷰길이	.210	.026	.152	8.004	.000***	1.000	1.000
R ² = .023, F=64.070							

<표 4-5>와 같이 리뷰길이는 부정의견에 $p < 0.001$ 수준에서 정(+)²의 영향을 미치고 있음을 알 수 있다. 가설 H2는 채택되었다.

4) 연구 가설 H3의 단순회귀분석

연구 가설 H3을 검증하기 위하여 리뷰길이를 독립변수, 조정의견을 종속변수로 두어 단순회귀분석을 실시하였다.

<표 4-6> 연구 가설 H3의 단순회귀분석

변수	B	표준오차	Beta	t-value	Sig.t	공차	VIF
리뷰길이	.045	.044	.019	1.010	.313	1.000	1.000
$R^2 = .000, F=1.020$							

<표 4-6>과 같이 리뷰길이는 조정의견에 유의한 영향을 미치지 않았다. 가설 H3은 기각되었다.

5) 연구 가설 H4, H5, H6, H7의 다중회귀분석

연구 가설 H4, H5, H6, H7을 검증하기 위하여 긍정의견, 부정의견, 조정의견, 리뷰길이를 독립변수, 정량점수를 종속변수로 두어 다중회귀분석을 실시하였다.

<표 4-7> 연구 가설 H4, H5, H6, H7의 다중회귀분석

변수	B	표준오차	Beta	t-value	Sig.t	공차	VIF
긍정의견	-0.35	.032	-1.035	-1.106	.269	.315	3.174
부정의견	-.235	.049	-.113	-4.787	.000	.568	1.760
조정의견	.424	.045	.339	9.374	.000	.242	4.139
리뷰길이	.090	.053	.031	1.712	.087	.193	.945
$R^2 = .147, F= 117.588$							

<표 4-7>와 같이 부정의견과 조정의견은 정량점수에 $p < 0.001$ 수준에서 정(+)의 영향을 미쳤지만, 긍정의견과 리뷰길이는 정량점수에 유의한 영향을 미치지 않았다. 가설 H6과 H7은 채택되었으나 H4와 H5는 기각되었다.

제 5 절 랭킹모델링을 위한 제언

온라인리뷰에서 언급되는 제품, 서비스, 세부 속성 등에 대한 랭킹(Ranking)은 소비자들에게 단순하면서도 강렬한 정보로서 역할을 할 수 있는데, 텍스트 형태의 온라인리뷰에 포함된 감성적 또는 긍정/부정적 의견과 정량적인 평가점수 등을 종합 적용하여 도출되어야 정확하게 랭킹을 부여할 수 있다. 그래서 오피니언 마이닝 기술은 제품, 서비스, 세부 속성 등의 랭킹을 도출하기 위한 핵심기술이 될 수 있다. 온라인리뷰의 정성적 평가 데이터인 텍스트 데이터에 대한 기존의 오피니언 마이닝 기술(윤홍준, 2010)은 조정의견을 정량적으로 계산하는 것이었다(윤홍준, 2010). 본 연구에서도 조정의견이 평가점수에 유의미한 정비례관계가 있으므로 이러한 분석기술은 일면 타당한 측면이 있다. 그러나 긍정의견과 조정의견에 대한 관계는 정비례 관계를 갖지 않는 것으로 나타났으므로 조정의견만을 가지고 리뷰 데이터를 평가하는 것은 정확하지 않을 수 있다. 본 연구의 결과에 따르면 정성적인 관점의 리뷰 데이터를 분석할 때 긍정의견, 부정의견, 조정의견을 나누어 분석을 제시하는 것이 보다 정확한 분석 결과를 제시하는 것이라 볼 수 있다. 이를 위해 본 연구에서는 온라인리뷰의 랭킹을 분석할 때 적용될 수 있도록 앞서 가설 검증 후 결과를 바탕으로 다음과 같이 랭킹모델을 제안한다.

1. 랭킹모델의 정의

기존의 오피니언 마이닝 기반 기술을 적용하여 온라인리뷰에 대해 정성적 관점에서 분석한 연구들은 한계가 있어 보다 정확하고 합리적인 평가기법이 제시될 필요성이 있다고 앞서 설명하였다. 온라인리뷰와 관련하여 오피니언 마이닝 기반 기술들이 많이 연구되어 왔지만 리뷰 데이터 분석에 대해 구체적인 평가기법을 제시하는 것이 보다 정확한 분석 결과를 도출할 수 있기 때문에 온라인리뷰의 분석 시 적용하기 위한 랭킹모델을 제시하는 것이다. 본 연구에서 제안하는 랭킹 모델은 온라인리뷰의 효과적 검색 결과를 도출하기 위해 이에 적용할 목적으로 리뷰 데이터의 의견 극성을 분류하여 각 의견점수를 계산하고 이에 따른 랭킹을 부여하여 분석한 사례를 제시하기 위한 것이다. 그러므로 리뷰 데이터 분석 시

적용하기 전에 앞서 가설 검증 후 결과를 바탕으로 랭킹모델을 도출하고자 한다.

랭킹모델 도출 시 긍정의견, 부정의견, 조정의견 등 온라인리뷰의 감정 극성에 따른 여러 변수들을 분류하고 이에 따른 점수를 계산하는 모델을 정의하였다. 이때 앞서 가설 검증 후 결과에서 채택된 가설들을 고려하여 다음 <그림 4-5>와 같은 랭킹모델을 도출하였다.

<그림 4-5> 랭킹모델

$$\begin{aligned} \text{긍정의견점수} &= \alpha \cdot \text{긍정의견} + \beta \cdot \text{리뷰길이} \\ \text{부정의견점수} &= \alpha \cdot \text{부정의견} + \beta \cdot \text{리뷰길이} + \gamma \cdot \text{정량점수} \\ \text{조정의견점수} &= \alpha \cdot \text{조정의견} + \gamma \cdot \text{정량점수} \end{aligned}$$

랭킹모델을 기반으로 온라인리뷰마다 긍정의견은 긍정의견점수, 부정의견은 부정의견점수, 조정의견은 조정의견점수를 각각 계산하고 이에 따라 다차원으로 랭킹을 도출하고 분석하는데 활용된다. 각 모델마다 적용되는 변수에 따라서 α , β , γ 등의 가중치를 다르게 조합하고 각 의견점수 계산 시 각각 0.5의 가중치를 부여한다. 긍정의견점수는 앞서 채택된 가설 H1을 바탕으로 도출하였는데, 긍정의견은 리뷰길이와 유의한 영향 관계를 갖고 있다는 점을 토대로 점수 계산 시, 긍정의견과 리뷰길이를 조합하여 계산하게 된다. 긍정의견점수에 따른 랭킹 분석 시 리뷰 데이터마다 긍정적인 감정이 얼마나 내포되어 있는지 가늠할 수 있게 된다. 부정의견점수는 앞서 채택된 가설 H2/H6를 바탕으로 도출하였는데, 부정의견은 각각 리뷰길이, 정량점수와 유의한 영향 관계를 갖고 있다는 점을 토대로 점수 계산 시, 부정의견과 리뷰길이, 정량점수를 조합하여 계산하게 된다. 부정의견점수에 따른 랭킹 분석 시 리뷰 데이터마다 부정적인 감정이 얼마나 내포되어 있는지 가늠할 수 있게 된다. 조정의견점수는 앞서 채택된 가설 H7을 바탕으로 도출하였는데, 조정의견은 정량점수와 유의한 영향 관계를 갖고 있다는 점을 토대로 점수 계산 시, 조정의견과 정량점수를 조합하여 계산하게 된다. 부정의견점수에 따른 랭킹 분석 시 리뷰 데이터마다 부정적인 감정 보다 긍정적인 감정

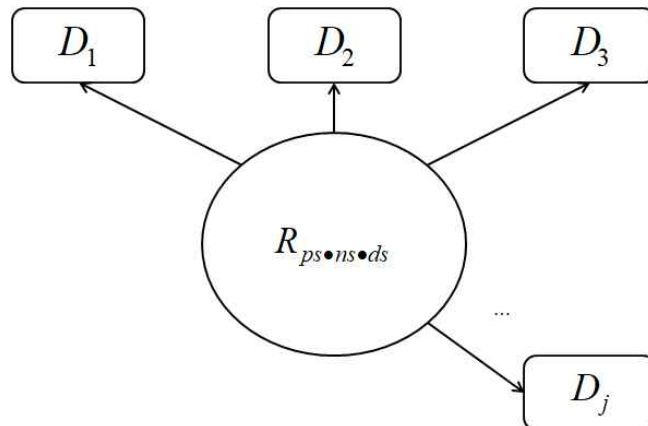
이 얼마나 더 내포하고 있는지 가늠할 수 있게 된다.

2. 온라인리뷰의 다차원 분석을 위한 랭킹모델의 적용

앞서 도출된 랭킹모델이 온라인리뷰 분석 시 적용되어 효과적으로 분석 결과를 도출할 수 있도록 다차원적인 관점으로 랭킹모델을 적용하는 방법을 제시한다. 이를 통해 온라인리뷰의 보다 정확한 랭킹 분석이 이루어질 것이다.

다차원적인 관점으로 온라인리뷰의 랭킹을 분석할 때 관광지, 계절, 연도 등의 차원에서 랭킹모델을 적용하여 분석할 것이다. 다음 <그림 4-6>은 온라인리뷰의 다차원 랭킹모델을 나타내고 있다.

<그림 4-6> 온라인리뷰의 다차원 랭킹모델



$R_{ps \cdot ns \cdot ds}$ 리뷰 데이터마다 랭킹모델을 기반으로 계산된 각 의견점수 속성 ($ps \cdot ns \cdot ds$)의 속성 값(예: $ps_{(1,2,3,\dots,i)}$)에 따른 랭킹 분석결과를 의미한다. D_j 는 분석결과에 따른 차원을 의미한다. 예를 들어, 긍정의견점수 랭킹(R_{ps})을 관광지 차원($D_{\text{관광지명}}$)에서 분석할 수 있고 부정의견점수 랭킹(R_{ns})을 연도 차원($D_{2015/2016}$)에서 분석할 수 있다. 그러면 여러 차원에서 $R_{ps \cdot ns \cdot ds}$ 속성 값 그룹 별로 랭킹을 부여하여 표현하게 된다. 예를 들어, 계절 차원($D_{\text{봄}}$)에서 조정의견

점수(ds) 그룹별로 랭킹을 부여하면 봄에서 관광지(ds_1), 관광지(ds_2), 관광지(ds_3), ..., 관광지(ds_r) 순으로 나타낼 수 있다.

제 5 장 랭킹모델 기반의 분석 사례

제 1 절 랭킹모델 기반의 온라인리뷰 분석 방법

본 연구에서는 앞서 제안된 랭킹모델을 기반으로 온라인리뷰의 랭킹을 도출하고 분석하기 위해 데이터는 앞서 오피니언 마이닝 기법으로 가공된 데이터를 활용하였으며 소프트웨어는 R을 이용하였다. 그리고 랭킹모델을 기반으로 리뷰별로 의견점수를 계산한 후 랭킹을 도출하고 관광지, 계절, 연도 등 다차원으로 분석하고자 하였다. 분석 사례는 다음과 같다.

제 2 절 분석 사례

1. 관광지별 긍정/부정/조정 의견 점수에 의한 온라인리뷰 랭킹

랭킹모델을 기반으로 계산된 의견점수의 평균에 따른 관광지 차원의 온라인리뷰 랭킹을 분석하였다. 상위 10개의 랭킹을 중심으로 파악하였고 분석 결과는 다음 <표 5-1>, <표 5-2>, <표 5-3>과 같다.

<표 5-1> 관광지별 긍정의견점수에 의한 랭킹

순위	관광지명	긍정의견점수(평균)
1	함덕 서우봉 해변	1.71
2	제주 절물 자연 휴양림	1.7
3	에코랜드 테마파크	1.63
4	비자림	1.6

5	월정리해변	1.58
6	송악산	1.58
7	한림공원	1.55
8	산굼부리 분화구	1.55
9	우도	1.53
10	협재 해수욕장	1.53

<표 5-2> 관광지별 부정의견점수에 의한 랭킹

순위	관광지명	부정의견점수(평균)
1	한라산 국립공원	3.49
2	우도	3.44
3	함덕 서우봉 해변	3.39
4	송악산	3.39
5	제주 절물 자연 휴양림	3.37
6	월정리해변	3.37
7	비자림	3.36
8	성산 일출봉	3.35
9	에코랜드 테마파크	3.33
10	협재 해수욕장	3.27

<표 5-3> 관광지별 조정의견점수에 의한 랭킹

순위	관광지명	조정의견점수(평균)
1	함덕 서우봉 해변	4.24
2	제주 절물 자연 휴양림	4.17
3	비자림	4.13
4	송악산	4.08
5	협재 해수욕장	4.07
6	우도	4.04
7	한라산 국립공원	4.03
8	중문·대포해안 주상 절리대	3.96
9	천지연 폭포	3.96
10	에코랜드 테마파크	3.95

관광지별 긍정의견점수에 의한 랭킹에서는 긍정의견점수 모델로 의견점수를 계산하고 이에 관광지별로 의견점수 평균을 계산하여 순위를 매겼다. 그 결과 함덕 서우봉 해변에 대해 가장 긍정적인 의견이 많은 것으로 나타났다. 관광지별 부정의견점수에 의한 랭킹에서는 부정의견점수 모델로 의견점수를 계산하고 이에 관광지별로 의견점수 평균을 계산하여 순위를 매겼다. 그 결과 한라산 국립공원에 대해 가장 부정적인 의견이 많은 것으로 나타났다. 관광지별 조정의견점수에 의한 랭킹에서는 조정의견점수 모델로 의견점수를 계산하고 이에 관광지별로 의견점수 평균을 계산하여 순위를 매겼다. 그 결과 함덕 서우봉 해변이 가장 높은 것으로 나타났다. 모든 랭킹의 의견점수 데이터에서 점수가 전체적으로 근소한 차이를 보이고 있었으며, 긍정의견, 조정의견 랭킹에서 모두 함덕 서우봉 해변이 근소하게 가장 높은 것으로 나타났다. 이것은 다른 관광지와 달리 함덕 서우봉 해변이 리뷰에서 부정적인 감정 보다 긍정적인 감정이 조금 더 많이 내포되어 있음을 알 수 있다. 반면, 한라산 국립공원의 경우 긍정의견 랭킹에서 10위권 밖으로 랭크되어 있고 부정의견 랭킹에서 가장 높은 것으로 보아, 전체 리뷰에서 다른 명소 보다 한라산 국립공원에 대해 부정적인 감정이 많이 드러났기 때문인 것으로 파악할 수 있다.

2. 계절별 긍정의견점수에 의한 온라인리뷰 랭킹

랭킹모델 중 긍정의견점수 모델을 기반으로 계산된 긍정의견점수의 평균에 따른 계절 차원의 온라인리뷰 랭킹을 분석하였다. 상위 10개의 랭킹을 중심으로 파악하였고 분석 결과는 다음 <표 5-4>와 같다. 그리고 보다 정확한 분석을 위해 유상욱(2017)의 연구에서 분석한 <표 5-5>의 계절별 호감도가 높은 관광지 랭킹과 비교하여 분석해보았다.

<표 5-4> 계절별 긍정의견점수에 의한 랭킹

	봄	여름	가을	겨울
순위	관광지명 (긍정의견점수(평균))			
1	한림공원 (1.78)	함덕 서우봉 해변 (1.78)	에코랜드 테마파크 (1.72)	월정리해변 (2)
2	월정리해변 (1.75)	제주 절물 자연 휴양림 (1.75)	제주 절물 자연 휴양림 (1.7)	한림공원 (1.9)
3	송악산 (1.69)	에코랜드 테마파크 (1.71)	함덕 서우봉 해변 (1.65)	제주 절물 자연 휴양림 (1.83)
4	아쿠아 플라넷 제주 (1.67)	비자림 (1.7)	산굼부리 분화구 (1.62)	천지연 폭포 (1.83)
5	제주 여미지 식물원 (1.67)	한라산 국립공원 (1.59)	제주 여미지 식물원 (1.61)	함덕 서우봉 해변 (1.81)
6	서귀포매일올레 시장 (1.65)	우도 (1.56)	비자림 (1.57)	송악산 (1.71)
7	제주 절물 자연 휴양림 (1.62)	동문 재래시장 (1.54)	월정리해변 (1.57)	용머리 해안 (1.69)
8	함덕 서우봉 해변 (1.59)	협재 해수욕장 (1.5)	우도 (1.55)	협재 해수욕장 (1.67)
9	천지연 폭포 (1.59)	산굼부리 분화구 (1.5)	협재 해수욕장 (1.52)	우도 (1.59)
10	에코랜드 테마파크 (1.58)	송악산 (1.5)	아쿠아 플라넷 제주 (1.5)	정방 폭포 (1.57)

<표 5-5> 계절별 호감도가 높은 관광지

	봄	여름	가을	겨울
순위	관광지명 (평점평균)	관광지명 (평점평균)	관광지명 (평점평균)	관광지명 (평점평균)
1	비자림 (4.73)	한라산 국립공원 (5.00)	한라산 국립공원 (4.77)	한라산 국립공원 (4.73)
2	함덕서우봉해변 (4.73)	함덕서우봉해변 (4.70)	함덕서우봉해변 (4.65)	함덕서우봉해변 (4.67)
3	협재 해수욕장 (4.72)	우도 (4.64)	비자림 (4.60)	우도 (4.64)
4	우도 (4.71)	에코랜드 (4.58)	우도 (4.58)	성산일출봉 (4.59)
5	한림공원 (4.67)	주상 절리대 (4.55)	성산일출봉 (4.41)	비자림 (4.53)
6	성산일출봉 (4.58)	용머리 해안 (4.53)	주상 절리대 (4.37)	주상 절리대 (4.25)
7	에코랜드 (4.55)	비자림 (4.52)	협재 해수욕장 (4.31)	쇠소깍 (4.09)
8	쇠소깍 (4.46)	협재 해수욕장 (4.44)	산굼부리 (4.24)	아쿠아플라넷제주 (4.00)
9	한라산 국립공원 (4.44)	천지연 폭포 (4.42)	정방 폭포 (4.23)	테디베어 박물관 (3.95)
10	천지연 폭포 (4.39)	성산 일출봉 (4.39)	천지연 폭포 (4.21)	동문 재래시장 (3.83)

자료 : 다차원 순위 모델 기반의 관광후기 분석(유상욱, 2017)에서 인용.

봄의 긍정의견 랭킹에는 한림공원(1.78)/월정리해변(1.75)/송악산(1.69), 여름의 긍정의견 랭킹에는 함덕 서우봉 해변(1.78)/제주 절물 자연 휴양림(1.75)/에코랜드 테마파크(1.71), 가을의 긍정의견 랭킹에는 에코랜드 테마파크(1.72)/제주 절물 자연 휴양림(1.7)/함덕 서우봉 해변(1.65), 겨울의 긍정의견 랭킹에는 월정리해변(2)/한림공원(1.9)/제주 절물 자연 휴양림(1.83) 순으로 나타났다. 전체 랭킹으로 보면 월정리해변이 2점으로 가장 긍정적인 의견이 많은 것으로 나타났는데, 실제 리뷰 중에서도 긍정의견점수는 대체로 높은 점수인 2점이 넘는 리뷰가 많았으며 주로 ‘너무 예쁘다’, ‘예쁜’, ‘맑다’ 등 긍정적인 의견이 많이 포함되어 있었다. 이처럼 월정리해변이 여름뿐만 아니라 겨울에도 방문하기 좋은 장소임을 알 수 있다. 그리고 월정리해변과 함께 전체적으로 해변이나 공원에 해당하는 명소에 대해 긍정적인 감정이 많은 리뷰가 많다는 것으로 파악할 수 있는데, 특히 제주 절물 자연 휴양림의 경우 전체 랭킹에서 대체로 상위권에 포함되어 있었다. 이것은 실제 리뷰 중에서도 긍정의견점수는 대체로 높은 점수인 2점이 넘는 리뷰가 많았으며 주로 ‘좋았다’, ‘편안하다’, ‘힐링이 되는’ 등 긍정적인 의견이 많이 포함되어 있는 것으로 보아 제주 절물 자연 휴양림은 어느 계절에 가든 방문하기 편하고 좋은

장소임을 알 수 있다. 함덕 서우봉 해변의 경우에도 앞서 관광지 차원의 긍정의견 랭킹에서 가장 높은 점수(1.71)를 기록한 바와 같이 계절 차원의 긍정의견 랭킹에서 마찬가지로 긍정적인 감정이 대체로 많은 장소임을 알 수 있다.

유상욱(2017)의 연구에서 분석한 계절별 호감도가 높은 관광지 랭킹과 비교해보면 결과가 서로 많이 다르게 나타나고 있음을 알 수 있다. 이것은 단순히 평점평균만을 가지고 순위를 분석한 것은 정량적 관점이기 때문에 정성적 관점의 리뷰 데이터까지 고려하지 않은 점을 감안하면 매우 상이한 결과임을 확인할 수 있다. 특히, 결과에서 한라산 국립공원이 평점평균 랭킹으로는 가장 높은 순위에 있지만 긍정의견점수 랭킹에서는 여름 랭킹 외에는 순위권 밖으로 랭크되어 있는데, 이것은 정성적 관점도 고려한 분석 결과이므로 실제 리뷰에서 긍정적인 감정이 많지 않았다는 점에서 정성적 관점까지의 고려 여부에 따라 결과가 매우 상이할 수 있음을 보이고 있다.

3. 계절별 부정의견점수에 의한 온라인리뷰 랭킹

랭킹모델 중 부정의견점수 모델을 기반으로 계산된 부정의견점수의 평균에 따른 계절 차원의 온라인리뷰 랭킹을 분석하였다. 상위 10개의 랭킹을 중심으로 파악하였고 분석 결과는 다음 <표 5-6>와 같다. 그리고 보다 정확한 분석을 위해 유상욱(2017)의 연구에서 분석한 <표 5-7>의 계절별 호감도가 높은 관광지 랭킹과 비교하여 분석해보았다.

<표 5-6> 계절별 부정의견점수에 의한 랭킹

	봄	여름	가을	겨울
순위	관광지명 (부정의견점수(평균))			
1	월정리해변 (3.58)	한라산 국립공원 (3.62)	한라산 국립공원 (3.45)	한라산 국립공원 (3.57)
2	비자림 (3.44)	산굼부리 분화구 (3.5)	우도 (3.44)	용머리 해안 (3.56)
3	우도 (3.43)	우도 (3.47)	월정리해변 (3.43)	월정리해변 (3.5)
4	에코랜드 테마파크 (3.41)	함덕 서우봉 해변 (3.43)	제주 절물 자연 휴양림 (3.4)	송악산 (3.5)
5	성산 일출봉 (3.39)	제주 절물 자연 휴양림 (3.42)	비자림 (3.4)	협재 해수욕장 (3.44)
6	함덕 서우봉 해변 (3.39)	에코랜드 테마파크 (3.38)	송악산 (3.38)	외돌개 (3.43)
7	한림공원 (3.39)	송악산 (3.33)	함덕 서우봉 해변 (3.37)	중문 색달 해변 (3.42)
8	쇠소깍 (3.38)	천제연 폭포 (3.32)	외돌개 (3.36)	우도 (3.41)
9	송악산 (3.38)	쇠소깍 (3.3)	성산 일출봉 (3.33)	한림공원 (3.4)
10	협재 해수욕장 (3.36)	성산 일출봉 (3.29)	제주김녕미로 공원 (3.29)	성산 일출봉 (3.39)

<표 5-7> 계절별 호감도가 높은 관광지

	봄	여름	가을	겨울
순위	관광지명 (평균점수)	관광지명 (평균점수)	관광지명 (평균점수)	관광지명 (평균점수)
1	비자림 (4.73)	한라산 국립공원 (5.00)	한라산 국립공원 (4.77)	한라산 국립공원 (4.73)
2	함덕서우봉해변 (4.73)	함덕서우봉해변 (4.70)	함덕서우봉해변 (4.65)	함덕서우봉해변 (4.67)
3	협재 해수욕장 (4.72)	우도 (4.64)	비자림 (4.60)	우도 (4.64)
4	우도 (4.71)	에코랜드 (4.58)	우도 (4.58)	성산일출봉 (4.59)
5	한림공원 (4.67)	주상 절리대 (4.55)	성산일출봉 (4.41)	비자림 (4.53)
6	성산일출봉 (4.58)	용머리 해안 (4.53)	주상 절리대 (4.37)	주상 절리대 (4.25)
7	에코랜드 (4.55)	비자림 (4.52)	협재 해수욕장 (4.31)	쇠소깍 (4.09)
8	쇠소깍 (4.46)	협재 해수욕장 (4.44)	산굼부리 (4.24)	아쿠아플라넷제주 (4.00)
9	한라산 국립공원 (4.44)	천지연 폭포 (4.42)	정방 폭포 (4.23)	테디베어 박물관 (3.95)
10	천지연 폭포 (4.39)	성산 일출봉 (4.39)	천지연 폭포 (4.21)	동문 재래시장 (3.83)

자료 : 다차원 순위 모델 기반의 관광후기 분석(유상욱, 2017)에서 인용.

봄의 부정의견 랭킹에는 월정리해변(3.58)/비자림(3.44)/우도(3.43), 여름의 부정의견 랭킹에는 한라산 국립공원(3.62)/산굼부리 분화구(3.5)/우도(3.47), 가을의 부정의견 랭킹에는 한라산 국립공원(3.45)/우도(3.44)/월정리해변(3.43)의 리뷰, 겨울의 부정의견 랭킹에는 한라산 국립공원(3.57)/용머리 해안(3.56)/월정리해변(3.5)의 리뷰 순으로 나타났다. 전체 부정의견 랭킹 중 여름, 가을, 겨울에서 한라산 국립공원이 가장 부정적인 감정이 많은 것으로 나타났는데, 앞서 관광지 차원의 부정의견 랭킹에서도 가장 높은 점수(3.49)를 기록한 바와 같이 계절 차원의 부정의견 랭킹에서 마찬가지로 골고루 가장 높은 순위를 기록하였다. 이것은 실제 리뷰 중에서도 부정의견점수는 대체로 높은 점수인 3.5~5점 사이의 리뷰가 많았고 주로 ‘힘들다’, ‘위험하다’, ‘위태롭다’ 등 부정적인 의견이 많이 포함되어 있는 것으로 보아 한라산 국립공원이 경치에 비해 어느 계절이든 여전히 등산하기 힘든 장소임을 알 수 있다. 그리고 전체적으로 우도, 송악산, 한라산 국립공원 등 섬이나 산에 해당하는 명소에 대해 부정적인 감정이 많은 리뷰가 많다는 것으로 파악할 수 있는데, 특히 우도의 경우 전체 랭킹에서 대체로 부정적인 감정이 많은 편에 속한 것으로 판단할 수 있다. 이것은 실제 리뷰 중에서도 부정의견점수는

대체로 높은 점수인 3.5~4.5점 사이의 리뷰가 많았고 주로 ‘비싸다’, ‘변했다’ 등 부정적인 의견이 많이 포함되어 있다는 것을 확인할 수 있는데, 리뷰 내용에서와 같이 우도에서 개발이 많아져서 예전 모습보다 많이 퇴색되어 있다는 점과 관광상품이 대체로 비싸다는 인식이 많다는 점을 알 수 있다.

유상욱(2017)의 연구에서 분석한 계절별 호감도가 높은 관광지 랭킹과 비교해보면 유일하게 한라산 국립공원이 모든 랭킹에서 유일하게 가장 높은 순위를 보이고 있다. 부정의견점수를 계산할 때 평가점수(평균)까지 포함하므로 모든 랭킹이 다소 비슷한 결과가 나올 수 있음을 보이고 있다.

4. 계절별 조정의견점수에 의한 온라인리뷰 랭킹

랭킹모델 중 조정의견점수 모델을 기반으로 계산된 조정의견점수의 평균에 따른 계절 차원의 온라인리뷰 랭킹을 분석하였다. 상위 10개의 랭킹을 중심으로 파악하였고 분석 결과는 다음 <표 5-8>와 같다. 그리고 보다 정확한 분석을 위해 유상욱(2017)의 연구에서 분석한 <표 5-9>의 계절별 호감도가 높은 관광지 랭킹과 비교하여 분석해보았다.

<표 5-8> 계절별 조정의견점수에 의한 랭킹

	봄	여름	가을	겨울
순위	관광지명 (조정의견점수(평균))			
1	한림공원 (4.28)	산굼부리 분화구 (4.5)	제주 절물 자연 휴양림 (4.25)	월정리해변 (4.67)
2	함덕 서우봉 해변 (4.25)	한라산 국립공원 (4.26)	함덕 서우봉 해변 (4.18)	용머리 해안 (4.44)
3	송악산 (4.25)	제주 절물 자연 휴양림 (4.25)	비자림 (4.15)	협재 해수욕장 (4.33)
4	협재 해수욕장	함덕 서우봉	한라산 국립공원	함덕 서우봉

	(4.17)	해변 (4.23)	(4.07)	해변 (4.33)
5	비자림 (4.17)	비자림 (4.14)	월정리해변 (4.07)	산굼부리 분화구 (4.22)
6	제주 절물 자연 휴양림 (4.15)	중문·대포해안 주상 절리대 (4.09)	우도 (4.01)	한림공원 (4.2)
7	제주 여미지 식물원 (4.11)	에코랜드 테마파크 (4.08)	송악산 (4)	섭지코지 (4.17)
8	월정리해변 (4.08)	송악산 (4.06)	협재 해수욕장 (3.96)	정방 폭포 (4.14)
9	우도 (4.06)	우도 (4.04)	외돌개 (3.95)	천지연 폭포 (4.13)
10	천지연 폭포 (4.04)	협재 해수욕장 (4.02)	제주김녕미로 공원 (3.93)	우도 (4.09)

<표 5-9> 계절별 호감도가 높은 관광지

	봄	여름	가을	겨울
순위	관광지명 (평균평균)	관광지명 (평균평균)	관광지명 (평균평균)	관광지명 (평균평균)
1	비자림 (4.73)	한라산 국립공원 (5.00)	한라산 국립공원 (4.77)	한라산 국립공원 (4.73)
2	함덕서우봉해변 (4.73)	함덕서우봉해변 (4.70)	함덕서우봉해변 (4.65)	함덕서우봉해변 (4.67)
3	협재 해수욕장 (4.72)	우도 (4.64)	비자림 (4.60)	우도 (4.64)
4	우도 (4.71)	에코랜드 (4.58)	우도 (4.58)	성산일출봉 (4.59)
5	한림공원 (4.67)	주상 절리대 (4.55)	성산일출봉 (4.41)	비자림 (4.53)
6	성산일출봉 (4.58)	용머리 해안 (4.53)	주상 절리대 (4.37)	주상 절리대 (4.25)
7	에코랜드 (4.55)	비자림 (4.52)	협재 해수욕장 (4.31)	쇠소각 (4.09)
8	쇠소각 (4.46)	협재 해수욕장 (4.44)	산굼부리 (4.24)	아쿠아플라넷제주 (4.00)
9	한라산 국립공원 (4.44)	천지연 폭포 (4.42)	정방 폭포 (4.23)	테디베어 박물관 (3.95)
10	천지연 폭포 (4.39)	성산 일출봉 (4.39)	천지연 폭포 (4.21)	동문 재래시장 (3.83)

자료 : 다차원 순위 모델 기반의 관광후기 분석(유상욱, 2017)에서 인용.

봄의 조정의견 랭킹에는 한림공원(4.28)/함덕 서우봉 해변(4.25)/송악산(4.25), 여름의 조정의견 랭킹에는 산굼부리 분화구(4.5)/한라산 국립공원(4.26)/제주 절물 자연 휴양림(4.25)의 리뷰, 가을의 조정의견 랭킹에는 제주 절물 자연 휴양림(4.25)/함덕 서우봉 해변(4.18)/비자림(4.15)의 리뷰, 겨울의 조정의견 랭킹에는 월정리해변(4.67)/용머리 해안(4.44)/협재 해수욕장(4.33) 순으로 나타났다. 전체 조정의견 랭킹으로 보면 월정리해변이 4.67점으로 가장 높은 것으로 나타났는데, 실제 리뷰들의 조정의견점수는 4.5~5점 사이, 조정의견 수치는 4~5 사이로 대체로 높았다. 게다가 긍정의견에 반해 부정의견이 0인 리뷰가 주로 많았으며 이것은 월정리해변의 리뷰들이 모두 부정적인 감정이 없거나 가장 적고 보다 긍정적인 감정이 더 많다는 것을 알 수 있는데, 이처럼 여러 계절 중 겨울에서 특히 가장 방문하기 좋은 해변인 것을 알 수 있다. 이외에도 한림공원, 산굼부리 분화구, 제주 절물 자연 휴양림 등 오름이나 공원에 해당하는 명소에 대해 부정적인 감정 보다 긍정적인 감정이 더 많은 리뷰가 많다는 것으로 파악할 수 있는데, 실제 조정의견점수는 4.5~4점 사이, 조정의견 수치는 4~5 사이로 대부분 높은 수치로 나타났으며 거의 모든 리뷰의 부정의견이 0인 경우가 많았다. 전체 랭킹을 보면 상위권에 있는 명소들이 주로 자연형 관광지이며 부정적인 감정 보다 긍정적인 감정이 더 많은 리뷰들이 많았고 실제 리뷰에서도 주로 ‘꼭 가봐야 할’, ‘아름다운’, ‘좋았다’ 등 긍정적인 의견이 많이 포함되어 있다는 것을 확인할 수 있는데, 이처럼 다른 명소 보다 주로 자연형 관광지에 대해 좋은 기억을 더 많이 갖고 있다는 것을 알 수 있다.

유상욱(2017)의 연구에서 분석한 계절별 호감도가 높은 관광지 랭킹과 비교해보면 앞서 긍정의견점수 랭킹과 비교했을 때 결과가 비슷한 것으로 보이고 있다. 조정의견점수는 부정적인 감정보다 긍정적인 감정이 얼마나 더 많이 내포하고 있는지 가늠할 수 있는 점수이므로 앞서 긍정의견점수 랭킹과 유사하게 나타날 수 있음을 보여주고 있다. 한편, 모든 랭킹에서 함덕 서우봉 해변이 유일하게 가장 비슷하게 나타나고 있음을 알 수 있는데, 조정의견점수 계산 시 평가점수(평점)도 포함하므로 함덕 서우봉 해변에 대한 긍정/부정의견 기울기가 평점과 유의한 영향을 갖고 있어 모든 랭킹에서 유사하게 나타날 수 있음을 보여주고 있다.

5. 연도별 긍정의견점수에 의한 온라인리뷰 랭킹

랭킹모델 중 긍정의견점수 모델을 기반으로 계산된 긍정의견점수의 평균에 따른 연도 차원의 온라인리뷰 랭킹을 분석하였다. 상위 10개의 랭킹을 중심으로 파악하였고 분석 결과는 다음 <표 5-10>과 같다.

<표 5-10> 연도별 긍정의견점수에 의한 랭킹

	2015	2016
순위	관광지명 (긍정의견점수(평균))	
1	함덕 서우봉 해변 (1.73)	제주 절물 자연 휴양림 (1.71)
2	제주 절물 자연 휴양림 (1.7)	함덕 서우봉 해변 (1.69)
3	월정리해변 (1.7)	에코랜드 테마파크 (1.69)
4	송악산 (1.62)	비자림 (1.64)
5	협재 해수욕장 (1.6)	한림공원 (1.64)
6	산굼부리 분화구 (1.6)	아쿠아 플라넷 제주 (1.61)
7	에코랜드 테마파크 (1.58)	월정리해변 (1.56)
8	우도 (1.56)	제주 여미지 식물원 (1.54)
9	비자림 (1.56)	천지연 폭포 (1.52)
10	중문 색달 해변 (1.52)	우도 (1.51)

2015년 랭킹에서는 함덕 서우봉 해변(1.73)/제주 절물 자연 휴양림(1.7)/월정리해변(1.7), 2016년 랭킹에서는 제주 절물 자연 휴양림(1.71)/함덕 서우봉 해변(1.69)/에코랜드 테마파크(1.69) 순으로 나타났다. 2015년과 2016년의 랭킹에서 각각 함덕 서우봉 해변, 제주 절물 자연 휴양림이 가장 높은 순위로 나타났는데, 앞서 관광지, 계절 차원의 긍정의견 랭킹에서와 마찬가지로 가장 긍정적인 감정이 많

은 관광지임을 확인할 수 있다.

6. 연도별 부정의견점수에 의한 온라인리뷰 랭킹

랭킹모델 중 부정의견점수 모델을 기반으로 계산된 부정의견점수의 평균에 따른 연도 차원의 온라인리뷰 랭킹을 분석하였다. 상위 10개의 랭킹을 중심으로 파악하였고 분석 결과는 다음 <표 5-11>과 같다.

<표 5-11> 연도별 부정의견점수에 의한 랭킹

	2015	2016
순위	관광지명 (부정의견점수(평균))	
1	한라산 국립공원 (3.44)	제주 절물 자연 휴양림 (3.54)
2	송악산 (3.44)	한라산 국립공원 (3.54)
3	월정리해변 (3.4)	우도 (3.52)
4	우도 (3.38)	비자림 (3.44)
5	함덕 서우봉 해변 (3.38)	함덕 서우봉 해변 (3.4)
6	성산 일출봉 (3.35)	아쿠아 플라넷 제주 (3.37)
7	최소각 (3.3)	에코랜드 테마파크 (3.37)
8	에코랜드 테마파크 (3.29)	월정리해변 (3.37)
9	제주 절물 자연 휴양림 (3.28)	성산 일출봉 (3.34)
10	비자림 (3.27)	섭지코지 (3.33)

2015년 랭킹에서는 한라산 국립공원(3.44)/송악산(3.44)/월정리해변(3.4), 2016년 랭킹에서는 제주 절물 자연 휴양림(3.54)/한라산 국립공원(3.54)/우도(3.52) 순으로 나타났다. 이 중에서 한라산 국립공원이 연도 차원의 부정의견 랭킹에서도 골고

루 부정적인 의견이 가장 많은 관광지인 것으로 나타났는데, 앞서 관광지, 계절 차원의 부정의견 랭킹에서와 마찬가지로 가장 부정적인 감정이 많은 관광지인 것을 알 수 있다.

7. 연도별 조정의견점수에 의한 온라인리뷰 랭킹

랭킹모델 중 조정의견점수 모델을 기반으로 계산된 조정의견점수의 평균에 따른 연도 차원의 온라인리뷰 랭킹을 분석하였다. 상위 10개의 랭킹을 중심으로 파악하였고 분석 결과는 다음 <표 5-12>와 같다.

<표 5-12> 연도별 조정의견점수에 의한 랭킹

	2015	2016
순위	관광지명 (조정의견점수(평균))	
1	월정리 해변 (4.3)	함덕 서우봉 해변 (4.25)
2	함덕 서우봉 해변 (4.21)	제주 절물 자연 휴양림 (4.25)
3	제주 절물 자연 휴양림 (4.13)	비자림 (4.18)
4	협재 해수욕장 (4.12)	한림공원 (4.12)
5	송악산 (4.12)	천지연 폭포 (4.03)
6	우도 (4.11)	협재 해수욕장 (4.02)
7	한라산 국립공원 (4.08)	산굼부리 분화구 (4)
8	비자림 (4.08)	송악산 (4)
9	중문·대포해안 주상 절리대 (3.97)	에코랜드 테마파크 (3.98)
10	제주 여미지 식물원 (3.95)	한라산 국립공원 (3.97)

2015년 랭킹에서는 월정리해변(4.3)/함덕 서우봉 해변(4.21)/제주 절물 자연 휴양림(4.13), 2016년 랭킹에서는 함덕 서우봉 해변(4.25)/제주 절물 자연 휴양림(4.25)/비자림(4.18) 순으로 나타났다. 이 중에서 함덕 서우봉 해변이 연도 차원의 조정의견 랭킹에서도 골고루 가장 높은 수치로 나타났는데, 이외에도 월정리해변, 제주 절물 자연 휴양림 등이 대체로 높은 수치로 나타났다. 이것은 앞서 관광지, 계절 차원의 조정의견 랭킹에서도 대체로 높은 순위를 기록했는데, 전체 긍정의견 랭킹에서도 가장 높게 나타난 것으로 보아 다른 명소 보다 긍정적인 감정이 더 많이 내포되어 있음을 알 수 있다.

제 6 장 결론

제 1 절 연구 요약 및 시사점

본 연구에서는 온라인리뷰의 정성적 평가와 정량적 평가의 회귀적 관계를 고찰하였으며 이를 기반으로 온라인리뷰의 보다 정확한 분석을 위한 시사점을 제시하고자 하였다. Tripadvisor에 게시된 2697개의 온라인리뷰를 오피니언 마이닝의 기법으로 변환 및 분석하여 긍정의견, 부정의견, 조정의견, 정량적 평가점수와 상관관계 및 회귀적 관계를 고찰하였다. 그 결과 정성적인 조정의견과 부정의견은 정량적 평가점수에 유의한 영향 관계가 존재하였으나 긍정의견은 유의한 영향 관계가 없었다. 조정의견과 정량적 평가점수 사이에 유의한 영향 관계가 있으므로 지금까지의 오피니언 마이닝 분석 기술은 필요조건이라 볼 수 있다. 그러나 긍정의견과 조정의견에 대한 관계는 유의한 회귀적 영향 관계를 갖지 않는 것으로 나타났으므로 조정의견만을 가지고 텍스트 형태의 리뷰 데이터를 평가하는 것은 정확하지 않을 수 있다. 따라서 정성적 관점에서 리뷰 데이터를 분석할 때 긍정의견, 부정의견, 조정의견을 나누어서 분석하는 것이 필요충분조건이라 볼 수 있다.

본 연구에서는 기존 수집된 리뷰 데이터인 Tripadvisor의 제주 명소 관련 리뷰 데이터를 대상으로 하여 분석 시 활용 가능한 형태로 변환하기 위해 형태소 분석 기법과 같은 오피니언 마이닝의 기법을 일부 적용하여 가공 처리하였으며 이를 가지고 정성적 관점의 텍스트 데이터와 정량적 관점의 평가점수 데이터의 유의한 영향 관계를 가설로 수립하고 이를 검증하기 위해 상관분석 및 회귀분석을 실시하였다. 그리고 분석 결과를 바탕으로 제안된 랭킹모델을 다차원적으로 적용하여 온라인리뷰를 분석한 사례를 제시하였다.

본 연구의 시사점은 다음과 같다.

첫째, 기존 오피니언 마이닝 기술 기반으로 온라인리뷰를 정성적 관점의 텍스트 데이터 위주로 분석한 선행연구와 달리 정량적 관점의 평가점수도 함께 균형적으로 고려하여 이에 회귀적 영향 관계를 가설검증 형태로 고찰함으로써 온라인 리뷰 데이터의 종합적인 관점에서 리뷰 작성자의 의견 감정을 보다 정확하고 합리적인 분석이 가능한 오피니언 마이닝의 분석 기법을 연구하는데 향후 도움을 줄 수 있을 것으로 기대하고 있다.

둘째, 가설 검증의 결과를 토대로 기존의 조정의견을 정량적으로 계산하여 리뷰 데이터의 감정 극성을 판단하고 분류하는 것(윤홍준, 2010)에서 긍정의견, 부정의견 등을 포함해 극성에 따른 여러 변수들을 분류하고 이에 따른 점수를 계산하는 랭킹모델을 제안하고 이를 바탕으로 분석 사례를 제시함으로써 이것이 관광 리뷰뿐만 아니라 상품평과 같은 온라인리뷰 검색 시 이 모델을 적용하여 리뷰의 의견, 속성, 정보에 대해 효과적인 검색결과를 도출하는데 보다 유용한 정보로서 역할을 할 것으로 보고 있다.

셋째, 온라인리뷰의 다차원 분석 시 랭킹모델을 적용하는 방법을 제시함으로써 리뷰 관련 사례 분석 시 이 모델을 다각도로 적용하여 보다 용이하게 분석할 수 있을 것이다.

제 2 절 연구의 한계

본 연구에는 관광리뷰 중 기존 수집된 제주 명소 관련 리뷰 데이터에 한정하여 연구했다는 점에서 다른 지역의 관광지 관련 리뷰에도 정성적 관점 및 정량적 관점으로 유의한 영향 관계가 존재할 수 있다는 것으로 보고 이에 대한 연구가 추가로 이루어져야 할 것으로 보인다. 본 연구에서 채택된 변수들 외에 추가로 다른 요인(예:특정 단어 빈도 수)을 고려하지 않았다. 다른 요인의 변수들을 추가로 고려하면 가설 검증 시 좀 더 다양한 분석 결과를 제시할 수 있을 것으로 판단한다. 그리고 오피니언 마이닝의 기법 중 형태소 분석 기법만 적용하여 데이터 가공을 시도했으나 일부 함수의 성능 문제로 모든 형용사들을 추출하지 못했다는 점이 있었다. 불가피하게 일부 수작업으로 진행하여 형용사에 가까운 형태소들을 색출할 수밖에 없었던 점을 보아 형용사 추출 시 좀 더 정확하게 유의미한 형태소를 추출할 수 있는 함수를 활용할 수 있도록 해야 할 것이다. 이후 극성에 따라 형용사를 분류할 때 수작업으로 진행했던 점에서 간혹 형용사의 극성을 명확하게 분류하지 않았을 수 있다. 이를 위한 감성분석 기법을 추가로 제시할 수 있는 연구가 필요할 것으로 보인다.

참고문헌

- 강선아, "온라인 커뮤니케이션 텍스트 분석을 통한 사회 이슈 긍부정 분류에 관한 연구", 충북대학교 석사학위논문, 2016.
- 김근형, 김두경, "귀납적 사회과학연구를 위한 확장형 오피니언마이닝 시스템 설계", 2015, pp.41-54.
- 김근형, 김두경, "데이터마이닝에 기반한 온라인 고객리뷰 분석 방법론", 2015, pp.65-76.
- 김근형, 오성열, "온라인 고객리뷰 분석을 통한 시장세분화에 텍스트마이닝 기술을 적용하기 위한 방법론", *한국콘텐츠학회논문지*, 제 9 권, 제 8 호, 2009, pp.272-284.
- 김문지, 송은정, 김윤희, "온라인 리뷰 데이터의 오피니언마이닝을 통한 콘텐츠 만족도 분석 시스템 설계", *2015년도 한국인터넷정보학회 추계학술발표대회*, 제 17 권, 제 3 호, 2016, pp.107-113.
- 김진옥, 이선숙, 용환승, "한글 텍스트의 오피니언 분류 자동화 기법", *한국정보과학회논문지*, 제 38 권, 제 6 호, 2011, pp.423-428.
- 김현정, 조남옥, 신경식, "항공산업 미래유망분야 선정을 위한 텍스트 마이닝 기반의 트렌드 분석", *지능정보연구*, 제 21 권, 제 1 호, 2015, pp.65-82.
- 김효원, "텍스트 마이닝을 활용한 웹사이트 특성 비교 분석", *광운대학교 석사학위논문*, 2017.
- 박은주, 정유진, "온라인 리뷰 탐색이 화장품 구매의도에 미치는 영향", *한국생활과학회지*, 제 22 권, 제 2 호, 2013, pp.343-355.
- 송중석, 이수원, "상품평 극성 분류를 위한 특징별 서술어 긍정/부정 사전 자동 구축", *한국정보과학회논문지*, 제 38 권, 제 3 호, 2010, pp.157-168.
- 신명철, 이용훈, 김미영, 정유진, 이종혁, "세종전자사전을 이용한 한국어 부사격의 의미역 결정", *제17회 한글 및 한국어 정보처리 학술대회*, 2005, pp.120-126.
- 심영석, "텍스트 마이닝(Text Mining)을 이용한 관광지 이미지 결정요인에 관한

- 연구", 세종대학교 석사학위논문, 2016.
- 안휘훈, 박병준, "의견 마이닝을 위한 유사 광고성 상품평 추출", *2014년도 대한 전자공학회 하계학술대회*, 제 37 권, 제 1 호, 2014, pp.1593-1596.
- 양가임, 임지연, 김이준, 김응모, "오피니언 마이닝 기술을 이용한 상품평 기반의 효율적 상품 추천 기법", *2011년도 대한전자공학회 하계종합학술대회*, 제 34 권, 제 1 호, 2011, pp.1125-1128.
- 양정연, 명재석, 이상구, "상품특징별 점수화를 이용한 상품리뷰요약 시스템의 설계 및 구현", *지식정보산업연합학회 창립기념 학술회*, 2008, pp.339-347.
- 어동선, "빅데이터를 이용한 텍스트마이닝 기법의 성능 비교", *인제대학교 석사학위논문*, 2014.
- 오선주, "소셜 네트워크에서 관계 랭킹 모델", *한국전자거래학회지*, 제 18 권 제 3 호, 2013, pp.93-105.
- 유상욱, "다차원 순위 모델 기반의 관광후기 분석", *제주대학교 석사학위논문*, 2017.
- 윤홍준, 김한준, 장재영, "오피니언마이닝기술을 이용한 효율적 상품평 검색 기법", *정보과학회논문지*, 제 16 권, 제 2 호, 2010, pp.222-226.
- 이동주, 연종흠, 황인범, 이상구, "꼬꼬마: 관계형 데이터베이스를 활용한 세종 말뭉치 활용 도구", *한국정보과학회논문지*, 제 16 권, 제 11 호, 2010, pp.1046-1050.
- 이승준, "온라인 고객 리뷰의 특성이 리뷰 유용성에 미치는 영향", *홍익대학교 석사학위논문*, 2012.
- 이현경, 김종배, "R 프로그램을 이용한 SNS상의 모바일 기기 사용자 감성분석", *예술인문사회융합멀티미디어논문지*, 제 7 권, 제 5 호, 2017, pp.821-828.
- 장재영, "온라인 쇼핑몰의 상품평 자동분류를 위한 감성분석 알고리즘", *한국전자거래학회지*, 제 14 권, 제 4 호, 2009, pp.19-33.
- 장지선, "텍스트마이닝을 이용한 고객 불만사항에 대한 사례연구", *호서대학교 석사학위논문*, 2011.
- 조성우, "Big Data시대의 기술", *KT종합기술원*, 2011.
- 조자룡, "텍스트마이닝을 이용한 한, 중 빅 데이터 관련 연구동향 분석과 비교",

- 충북대학교 석사학위논문, 2016.
- 허필희, "텍스트 분석 기법을 활용한 이슈 라이프 사이클 분석 삼성전자 갤럭시 노트7 사례를 중심으로", 계명대학교 석사학위논문, 2017.
- 홍태호, 이태원, "온라인 상점 고객의 감성분류를 위한 용어기반 오피니언마이닝", 2014년 한국경영정보학회 춘계학술대회, 2014, pp.619-624.
- 황원석, 채수민, 김상욱, 최호진, "논문 검색 엔진을 한 랭킹 방법", 한국정보과학회논문지, 제 40 권, 제 5 호, 2013, pp.345-357.
- Bromberg, A. (2013). First shot: Sentiment Analysis in R. URL:
<http://andybromberg.com/sentiment-analysis/>
- Chiou, J. S. & Cheng C. (2003). Should a company have message boards on its web sites? Journal of Interactive Marketing, 17(3), 50-60.
- Davi, A., Haughton, D., Nasr, N. Shah, G., Skaletsky, M. and Spack, R. (2005). A Review of Two Text-Mining Packages: SAS TextMining and WordStat, The American Statistician, 59(1), 89-103.
- Dwayne, D. G., Kevin, P. G., & Stephen, W. B. (2001). Generating positive word of mouth communication through customer-employee relationships. International Journal of Service Industry Management, 12(1), 44-59.
- Feinerer I., Hornik K. and Meyer D. (2008). Text Mining Infrastructure in R, Journal of Statistical Software , 25(5), 1-54.
- Henning, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word of mouth via consumer opinion platforms: what motivates consumers to articulate themselves on the internet? Journal of Interactive Marketing, 18(1), 38-52.
- Hotho, A., Nurnberger, A. and Paa, G. (2005). A Brief Survey of Text Mining, Ldv Forum, 1-37.
- Hu, M. and Liu, B.(2004), "Mining and summarizing customer reviews," In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp.168-177, ACM.

- Jarman, J. (2011). Combining Natural Language Processing and Statistical Text Mining: A Study of Specialized Versus Common Languages, Graduate School Theses and Dissertations, University of South Florida.
- Jeon, W. Y. & Jung, H. J. (2006). Effects of online reviews on evaluation and purchase intention of a product in internet shopping: the role of gender differences. *Journal of The Korean Psychological Association*, 7(1), 113-129.
- Kim, C. H. (2007). The influences of internet WOM(word Of Mouth) information valence and contents to WOM effects -focus on cell phone purchasing for student via online WOM-. *Journal of The Korea Association for International Commerce and Information*, 8(4), 23-41.
- Lee, E. Y. & Lee, T. M. (2005). The effects of information characteristics on word of mouth performance in online environment: focused on the moderating effect of consumer knowledge. *Journal of The Korea Advertising Society*, 16(2), 7-33.
- Manning, C.D., Raghavan, P. and Schütze, H. (2008). *Introduction to Information to Information Retrieval*, Cambridge University Press Cambridge.
- Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings*, 93-98.
- Park, C., & Yoo, C. J. (2006). Study on the effects of consumers' word of mouth about brand experiences and responses on brand evaluations at online context. *Journal of The Korean Society of Consumer Studies*, 17(1), 73-93.
- Peterson, R. A. & Meria, C. M. (2003). Consumer information search behavior and internet. *Psychology and Marketing*, 20(2), 99-121.