

## 온라인 후기에 내재된 고객의 감성분석과 LQI 차원별 호텔 서비스 품질 평가\*

사공 원\*\* · 하 성 호\*\*\* · 박 경 배\*\*\*\*

〈목 차〉	
I. 서론	IV. 분석 결과
II. 이론적 배경	4.1 데이터 수집 및 정제
2.1 서비스 품질	4.2 Word2vec 결과
2.2 온라인 후기	4.3 감성점수 계산
2.3 감성 분석	4.4 분석 결과
III. 분석 방법	V. 결론
3.1 후기 분석 설계	참고문헌
3.2 투숙 후기: TripAdvisor.com	<Abstract>
3.3 후기 데이터 자연어 처리과정	
3.4 핵심어 분류	
3.5 감성 분석	

### I. 서론

컴퓨터의 보급과 인터넷의 발달로 현대인들의 쇼핑, 영화, 음악 감상, 게임 등 다양한 형태의 온라인 상에서의 활동은 폭발적인 데이터를 생산하게 되었고, 이른바 빅데이터 시대를 맞이하게 되었다. 온라인에 존재하는 많은 데이터 중 특정 제품이나 서비스 및 브랜드에 관해 고

객들의 자발적 참여로 작성되어진 구매후기 정보는 산업 분야를 막론하고 그 활용방안에 관하여 연구자들과 기업인들에게는 보다 깊이 있는 소비자 이해를 위한 새로운 과제와 기회를 창출하게 되었다. 특히, 서비스 산업군 중 대표 주자인 호텔산업의 경우, 다양한 호텔 상품 구매경로(예: 온라인 여행사, 자사 홈페이지, TripAdvisor 등)에 표출된 여행객들의 투숙 후

\* 이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2015S1A5A2A01012887)

\*\* 경북대학교 일반대학원 경영학부 석사, [wsakong@empas.com](mailto:wsakong@empas.com)

\*\*\* 경북대학교 경영학부 교수, [hsh@mail.knu.ac.kr](mailto:hsh@mail.knu.ac.kr), 교신저자

\*\*\*\* 경북대학교 일반대학교 경영학부 박사수료, [iamkbpark@knu.ac.kr](mailto:iamkbpark@knu.ac.kr)

기 정보는 경영진들에게 큰 관심사일 것이다.

문화관광체육부의 통계자료(2012)에 따르면 2000년 중반부터 현재까지 국내에 입국하는 외국인들의 수는 매년 꾸준히 증가해 왔다. 외국인들의 방문 이유는 약 80%가 여행이었고 나머지는 사업 또는 전문 활동을 위한 비즈니스 목적이었다. 그들이 이용한 대표적인 숙박시설로는 호텔이 전체 75.8%의 높은 비율을 차지하였으며, 유스호스텔/게스트하우스/여관/모텔이 13.4%를 차지하였다. 방문 비율이 가장 높은 지역은 전체 17개의 시·도 중에서 80.4%를 차지한 서울이었고, 그 뒤를 제주(18.0%), 경기(13.0%), 부산(8.0%)이 차지하였다. 최근 꾸준히 증가하는 중국 여행객들의 한국 방문 추세를 반영해 볼 때 내한 방문객 수는 앞으로도 증가할 것으로 예상된다.

국내 호텔 경영진의 입장에서는 이러한 외국인들의 방문 증가 추세에 대응하여 수익 창출 및 증가를 위해서 한국인 여행객들뿐만 아니라 외국인 여행객들의 요구와 그들이 표출한 투숙 후기를 바탕으로 호텔 서비스에 대한 평가를 보다 구체적으로 수행할 필요성이 있다. 따라서 본 연구는 외국인들의 방문율이 높은 국내 도시들 중, 그 비율이 압도적인 서울을 선정하여, 이 지역에 위치한 국내 계열 호텔(A호텔, B호텔)과 외국 계열 호텔(C호텔, D호텔)에 관한 후기를 분석에 활용하였다.

온라인에서 제품을 구매할 때, 고객은 기존 고객이 남긴 온라인 후기를 통하여 제품에 대한 불확실성을 낮추고자 한다. 긍정적 후기가 많을수록 제품에 대한 구매 의도는 올라가게 된다는 선행 연구들을 토대로 여행객들이 남긴 호텔 숙박 후기를 이용하여 국내 호텔의 시설

및 서비스 품질에 대해 평가하고자 하였다. 이를 위해 감성분석을 활용하여 여행객들이 남긴 투숙 후기가 비록 전체적으로는 긍정적이라 할 지라도 그 안에 숨겨진 아쉬웠던 점이나 불편했던 점 그리고 서비스 품질의 개선점 등을 도출하였다.

본 연구는 감성 어휘에 대한 분류를 위하여 호텔 서비스 품질 모형 중, Lodging Quality Index(LQI) 모형의 5가지 차원을 기준으로 삼았다. 이것은 호텔별로 서비스 품질을 비교 평가함에 있어 비록 호텔이 보유한 시설은 다를 수 있으나 서비스 품질을 측정하는 기준은 동일하게 적용할 수 있기 때문이다. 또한, 수집한 후기들을 기간별로 분석하여 내외국인 여행객들의 호텔 서비스에 대한 감성이 시간에 따라 어떻게 바뀌는지 알아보고 이를 바탕으로 호텔 서비스 품질의 변화를 살펴보고자 하였다. 또 여행객들의 감성 변화를 토대로 국내 호텔 서비스에 대한 고객 만족도와 개선점에 대해 설명해 보고자 하였다.

## II. 이론적 배경

### 2.1 서비스 품질

기본적으로 서비스는 제공하는 사람과 제공 받는 사람 사이에서 일어나는 일시적인 행위의 개념으로, 무형성(intangibility), 이질성(heterogeneity), 소멸성(perishability), 불가분성(inseparability)이라는 4가지 특징을 가진다.

무형성은 서비스가 물리적 제품이 아닌 행위로서, 서비스가 제공되기 전까지는 서비스에 대

한 평가를 할 수 없음을 뜻한다. 이질성은 동일한 서비스라도 제공받는 사람, 시간, 장소에 따라 달리 평가됨을 의미하며, 소멸성은 무형의 서비스가 제공됨과 동시에 사라지며 제공받은 서비스를 따로 저장하거나 보관할 수 없는 성질을 의미한다. 불가분성이란 서비스의 서로 다른 평가 요소들이 동일한 서비스를 통해 동시에 측정됨을 의미한다(Parasuraman et al., 1985; Ladhari, 2009).

일반적으로 3차 산업으로 분류되는 서비스 산업에서는(홍동문, 2003) 서비스 기업이 소비자에게 제공하는 서비스 품질 자체가 기업 경쟁력에 큰 영향을 미치므로 지난 수십 년간 이에 관해 다수의 연구가 진행되어 왔다. Rathmell(1966)은 서비스를 무형의 제품이라고 설명하였으며, Smith(1976)는 국부론을 통하여 재화 생산과 소비만을 경제활동으로 한정 지었던 기존 경제학에 다른 유형의 경제활동인 서비스를 소개하였는데, 재화가 생산되어 소비되기까지의 과정에서 필요한 서비스의 중요성을 강조하였다. 또한, Berry(1980)는 유형성 즉, 객관적인 형태를 가진 제품과 달리 서비스는 무형의 행위 또는 노력이라고 소개하였고, Kotler(1988)는 서비스가 제공받는 사람에게 전해지면서 소유될 수 없는 무형의 행위와 노력이라고 설명하였다.

특히, Parasuraman et al.(1988)은 서비스 품질이 가속화 되는 규제 완화와 함께 심화되는 경쟁사회에서 기업이 갖추어야 할 차별적 요소라고 설명하였으며, Saleh et al.(1991)은 서비스 품질이 특히 호텔 산업의 지속성을 결정하는 중요한 요인이라고 설명하며, 고객의 최초 방문은 다양한 원인이 있을 수 있지만 지속적

인 방문은 기업과 직원의 서비스 제공 능력에 의해 결정된다고 주장하였다. 더 나아가 Said et al.(2013)은 서비스 품질을 서비스 산업에서 기업의 경쟁 이익과 관련된 핵심요소라고 설명하였다.

일반적으로 제품 구매 시, 고객들은 제품의 형태, 내구성, 제품의 색상, 감촉, 성능 등을 경험하고 제품의 품질을 평가한다. 하지만, 서비스의 경우 무형적인 이유로 인하여 기업 입장에서 고객이 서비스의 품질에 대하여 어떻게 평가하는지를 이해하기란 매우 힘들다(Parasuraman et al., 1985).

### 2.1.1 SERVQUAL

본 연구에서는 호텔 요소 별 서비스 품질을 측정하기 위해 SERVQUAL의 차원들을 인용하였다. Parasuraman et al.(1985)은 서비스 산업에서 제공되는 서비스의 품질은 형태와 장소에 관계없이 공통적으로 설명되는 10개의 차원, 즉 신뢰성(Reliability), 대응성(Responsiveness), 경쟁성(Competence), 접근성(Access), 친절성(Courtesy), 의사소통(Communication), 신용성(Credibility), 안전성(Security), 이해성(Understanding/Knowing the customer), 유형성(Tangibility)으로 측정이 가능하다고 설명하였다.

서비스 품질 측정을 위한 각 차원들을 살펴보면, 먼저, 신뢰성은 서비스 품질에 대한 믿음과 서비스 품질의 일관성을 설명한다. 신뢰성은 기업이 제공하는 서비스가 시작부터 적절하였는지, 제공되는 서비스에 대하여 기업이 정한 가격과 서비스 제공 기록이 정확한지 그리고 고객과 약속한 시간에 정확히 제공되었는지

를 측정한다. 둘째, 대응성은 직원이 얼마나 준비된 자세에서 의욕을 가지고 서비스를 제공하는지를 설명한다. 이러한 대응성은 고객에게 서비스를 빠르게 제공하는 것뿐만 아니라 잘못 제공된 서비스에 대하여 직원이 얼마나 빨리 관리자에게 보고하는지와 그러한 사실을 고객에게 얼마나 빨리 알려주는지를 측정한다.

셋째, 경쟁성은 기업과 직원이 고객에게 얼마나 풍부한 지식과 뛰어난 기술로 서비스를 제공하는지를 설명한다. 경쟁성과 관련된 지식과 기술에는 고객을 대면할 때 필요한 지식과 기술, 고객의 요구를 수행하는 데 필요한 지식과 기술, 그리고 고객에게 서비스를 제공하기 위해 필요한 시설들에 대한 기업의 관리 능력이 포함된다. 넷째, 접근성은 고객이 서비스를 요구하거나 서비스를 제공받는 과정이 얼마나 간편한지를 설명한다. 고객이 전화상으로 서비스를 요구하는 행위가 얼마나 간편한지, 고객이 서비스를 제공받기 위해 기다리는 시간이 연장되지 않도록 관리하는 기업의 능력, 서비스를 제공받음에 있어 시간에 대한 제약과 서비스 제공 시설의 위치 편리성이 포함된다.

다섯째, 친절성은 직원의 정중함, 친절, 고객에 대한 존중, 이해심, 고객을 대면할 때의 친근한 자세가 포함되며, 여섯째의 의사소통은 직원과 고객 사이에 서비스에 대한 요구 및 설명에 대한 정보가 얼마나 정확하게 제약 없이 전달되는지를 설명한다. 직원과 고객 사이에서 전달되는 서비스, 서비스 비용, 고객의 요구사항에 대한 처리 과정 등의 정보가 언어적 제약 없이 정확하게 전달되는 능력이라 할 수 있다. 일곱째, 신용성은 기업의 이름과, 평판, 그리고 직원의 인성 등으로부터 발생하는 서비스에 대한

신뢰도라 할 수 있고, 여덟째의 안전성은 고객이 위협이나 의심받는 상황으로부터 얼마나 자유로울 수 있는지를 나타내는 기업의 보안성과 관련되어 있으며, 물리적 또는 금전적인 위험으로부터 기업이 고객의 안전을 책임지는 행위를 일컫는다.

아홉째, 이해성은 기업이나 직원이 고객의 요구를 이해하기 위해 노력하는 자세를 말하는데, 고객의 특별한 요구에 대한 기업이나 직원의 학습 태도, 개별적 관심, 정기적으로 방문하는 고객을 관리하는 능력이 포함된다. 마지막으로 유형성이란 무형적인 서비스가 아닌 시설이나 장비, 기계 등 유형적인 요소들의 외관에 대하여 설명한다. 물리적 시설과 직원의 용모, 서비스의 제공에 이용되는 장비나 기계, 또한 같은 시설 내의 다른 고객들까지 서비스의 유형성에 영향을 미친다.

이에 머물지 않고, Parasuraman et al.(1988)은 요인 분석을 이용하여 기존에 제안한 10가지의 서비스 품질 측정 차원들 중 유사 또는 일부 중복되는 차원들을 통폐합하여 보다 간결한 5가지 차원의 개선된 SERVQUAL 모형을 소개하였다. 이것은 신뢰성(Reliability), 대응성(Responsiveness), 유형성(Tangible), 확신성(Assurance), 공감성(Empathy) 들이다. 신뢰성, 대응성, 유형성 차원은 기존의 차원들과 동일하지만, 확신성 차원은 이전 SERVQUAL의 의사소통, 신용성, 안전성, 경쟁성, 친절성을 포함하는 차원으로, 직원의 보안성, 친절성, 지식 및 기술보유 수준, 신뢰성, 의사소통법 등을 측정하는 차원이다. 공감성 차원은 기존 10개의 차원들 중 이해성과 접근성을 통합한 것으로, 고객의 요구 및 문제 해결을 위한 접근 방법을 측

정한다.

### 2.1.2 LODGSERV

SERVQUAL은 외식, 의료, 관광 및 호텔 산업과 같이 다양한 서비스 산업을 대상으로 포괄적 성격의 서비스 품질을 측정하는 모형이라고 할 수 있다. 여러 분야의 산업에 걸쳐 제공되는 서비스가 기본적인 공통점을 가지고 있다 하더라도, 세부적으로 제공되는 서비스의 종류, 장소, 시간과 목적에 차이가 있기 때문에 보다 정교한 서비스 품질을 측정하기 위해서는 해당 산업에 특화하여 서비스 품질을 측정하는 모형이 필요하다.

따라서 서비스 품질을 보다 정교히 측정하기 위해 많은 연구가 진행되었고(장은진·김정근, 2011; 양흔 등, 2013; 홍종의, 2013), 각각의 연구가 제안하고 품질 측정에 이용한 척도는 조금씩 달랐지만(Saleh et al. 1991), 본 연구에서는 근간이 되는 SERVQUAL의 차원들을 인용하면서 호텔 요소별 서비스 품질을 측정하는 LODGSERV와 LQI 모형의 차원들을 이용하였다.

Akbaba(2006)는 호텔 산업의 서비스 품질 측정에 SERVQUAL의 10개 차원을 이용해도 되지만, 보다 정확한 분석을 위해서는 호텔 산업의 특징을 제대로 반영한 모형이 필요함을 강조하였다. LODGSERV는 5가지 차원으로 이루어진 SERVQUAL 모형으로, 호텔산업의 서비스 측정에 특화된 모형이다(Knutson et al., 1990). LODGSERV에서 신뢰성은 사전에 약속된 서비스를 일관성 있고 정확하게 제공하는 능력을 의미하고, 대응성은 지원과 기업의 고객 을 도우려는 의지와 신속하게 서비스를 제공하

는 능력을 지칭한다. 유형성은 호텔의 물리적 장비, 시설, 직원의 용모 등과 관련되고, 공감성은 감정이입차원으로 호텔이나 직원의 고객에 대한 개별적인 관심도를 나타낸다. 마지막으로 확신성은 종업원의 지식과 문제 해결에 필요한 기술보유 정도 및 고객이 신뢰할 수 있도록 만드는 직원의 정중한 태도를 가리킨다.

### 2.1.3 Lodging Quality Index (LQI)

Carman(1990)과 Cronin et al.(1992)은 SERVQUAL이 서비스의 종류와 관계없이 서비스 품질의 공통적인 요소들을 이용한 평가모형으로 다양한 서비스 산업에서 이용되고 있지만 이를 특정 산업에 적용시킬 경우 서비스 품질을 측정하는 차원들을 수정하고 재정의 할 필요성이 있다고 주장하였다. 많은 서비스 품질 측정 모형들이 SERVQUAL의 차원을 그대로 인용한 점에 대해 Van Dyke et al.(1997)는 서로 다른 서비스 산업의 서비스 품질 측정 시 모형 내에 이용되는 특정 차원을 제거하거나 새로운 차원을 추가하여 해당 산업에 맞춤형 차원들을 고안할 필요성이 있음을 주장하였는데, 이는 각 서비스 분야의 특성을 정확히 분석하기 위함이었다. 또한, Nadiri et al.(2005)은 SERVQUAL의 5가지 차원들 중 유형성 차원을 제외한 나머지는 비유형성(Intangibles)차원 하나로 설명하는데 이러한 차원들을 이용함은 적절치 못하다고 지적하였다.

여러 선행 연구들을 토대로 Getty et al.(2003)은 호텔 서비스 품질 측정에 특화된 LQI 모형을 소개하였는데, Parasuraman et al.(1985)이 사용한 기존의 10가지 차원을 수용하고 수정하여 LQI 나름의 새로운 서비스 품질

측정 차원을 개발한 것이다. 세분화된 호텔 산업의 서비스 품질을 측정하기 위해서는 기존 서비스 품질의 10가지 차원을 활용하되 새로운 잠재적인 차원을 개발해야 한다는 주장이었다.

LQI 모형은 신뢰성(Reliability), 대응성(Responsiveness), 유형성(Tangibility), 믿음성(Confidence), 의사소통(Communication)의 5가지 차원으로 이루어져 있다. 신뢰성은 기존 SERVQUAL의 신뢰성과 신용성으로 구성되며, 대응성은 기존의 대응성이, 유형성은 기존의 유형성이, 믿음성은 기존의 경쟁성, 친절성, 안전성, 접근성들로 이루어져 있으며, 의사소통의 경우 기존의 이해성과 의사소통으로 이루어져 있다.

각 차원에 대한 구체적인 설명을 살펴보면(Ladhari, 2012), 신뢰성은 호텔이 고객과의 약속을 얼마나 잘 수행하는지를 나타내는 호텔의 신뢰도와 신용도를 평가하는 차원이고, 대응성은 직원이 고객의 요구를 얼마나 효율적이며 신속하게 대응해 주는지에 대해 평가하는 차원이라 할 수 있다. 유형성은 호텔의 시설이나 외관 등 유형적인 측면을 평가하는 차원이고, 믿음성의 경우 직원의 친절성 및 호텔의 안전성을 평가하는 차원이며, 의사소통이란 고객의 요구에 대한 직원의 이해도 및 호텔의 적절한 조치를 평가하는 차원이라 할 수 있다.

Getty et al.(2003)은 LQI 모형의 차원들 중 신뢰성의 경우 인지된 만족도, 추천의도와 밀접한 관련성이 있으며, 대응성의 경우 고객의 감정적인 만족도와 밀접한 관련성이 있다고 설명하였다. 그리고 유형성의 경우 종합적인 서비스 품질, 인지된 만족도, 감정적인 만족도, 추천 의도, 호텔에 대한 고객의 충성도와 관련성이 높

으며, 의사소통은 종합적인 서비스 품질, 인지된 만족도, 추천 의도, 충성도와 프리미엄 상품에 대한 비용지불의도와 밀접한 관련성이 있다고 설명하였다.

## 2.2 온라인 후기

비상업적인 목적으로 사람들 간의 입에서 입으로 정보를 전달하는 행위를 구전커뮤니케이션(Word-of-Mouth, WOM)이라 한다. 오프라인에서의 구전커뮤니케이션의 경우, 비용, 시간 대비 효과, 효율성의 문제로 인하여 현실적인 마케팅이 불가능하였으나, 최근 인터넷을 포함한 각종 정보통신 기술의 발전은 소셜네트워크 서비스(Social Network Service, SNS), 블로그와 같은 각종 온라인 미디어의 급격한 성장을 통해 구전커뮤니케이션의 성장을 촉발하고 있다. 구전 커뮤니케이션은 대인 영향과 더불어 고객의 제품 구매 결정에 매우 중요한 영향 요인이다(김진화 등, 2011).

1대 1로 전해지던 정보의 전달이 온라인상에서는 1대 N의 형태로 전달되면서 정보의 전달, 확산 및 공유되는 속도가 빨라지게 되었는데(김보영 등, 2016), 이를 온라인 구전(electronic WOM)이라 한다(성연신 등, 2001; 오걸, 2015). 온라인 구전은 판매자, 제품, 서비스에 관한 모든 정보를 고객에게 전달하는 의사소통이며(Litvin et al., 2008), 비용효율적인 마케팅 수단으로 많이 이용되고 있다.

### 2.2.1 온라인 후기에 대한 정의

온라인 구전의 대표적인 형태는 온라인 후기인데 온라인에서 고객들이 제품에 대한 후기를

보는 이유는 제품의 품질에 대한 불확실성을 해소하기 위함이고(Chen et al., 2001), 제품에 대한 평점을 매길 수 있는 척도가 제공될 때 소비자들이 온라인 후기를 작성하려는 의도가 비로소 강해진다(Chen et al., 2005). Dabholkar et al.(2006)은 판매자 입장에서 평점을 매길 수 있는 척도를 제공하는 것은 많은 소비자들이 제품에 대한 후기를 작성하도록 유도하여 결국 제품에 대한 홍보 효과를 가져다준다고 설명하였다.

Connor(2010)는 사용자가 제품이나 서비스에 대하여 자신의 의견이나 중요하다고 생각하는 사실과 관련 경험을 자유롭게 표현한 형태를 온라인 후기라고 설명하였으며, 구태희(2010)는 제품이나 서비스에 대하여 구매하였거나 사용해본 경험이 있는 사람들이 자신의 경험, 의견 및 평가 등을 온라인 상에 텍스트 데이터 형식으로 기록하는 행위라고 설명하였다. 또한 Connors et al.(2011)은 사람들이 타인의 구매의도에 결정적인 정보를 제공하기 위하여 온라인 후기를 작성하는 경향이 있다고 주장하였다.

이정현(2013)에 의하면 온라인 후기는 사용자가 제품 구매시점부터 투자한 시간적, 금전적인 노력을 포함하며, 이를 바탕으로 사용자의 제품 및 서비스에 대한 입장이 고스란히 반영되기 때문에 잠재적 고객에게는 제품에 대한 중요한 정보원천이 된다고 설명하였다. 더불어, 온라인 후기는 실제 제품을 사용한 경험이 있는 사용자들이 제품에 대하여 상업적인 목적이 아닌 자발적인 제품 이용에 대한 경험을 제공한다는 점에서 다른 잠재적 고객들이 이를 통해 제품이나 서비스에 대한 간접적인 경험을

하게 된다고 설명하였다. Liu(2013)은 이러한 방대한 양의 온라인 후기를 “군중의 지혜(Wisdom of crowd)”라고 표현하였다.

### 2.2.2 온라인 후기의 효과

온라인 후기는 후기 내용의 성향에 따라 긍정적인 효과와 부정적인 효과를 띄게 되는데, 이를 온라인 후기의 방향성이라 한다(Brister, 1991). 제품에 대한 칭찬이나 유용한 점에 대한 설명과 같이 긍정적 성향의 후기의 경우 긍정적인 효과를 불러일으키는 반면, 제품에 대한 불만족 사항이나 사용 시 불편했던 경험 등의 부정적 성향의 후기인 경우 부정적인 효과를 초래하게 된다(오걸, 2015).

성연신 등(2001)은 온라인 후기의 효과에 대한 연구에서 단순 정보제공을 목적으로 하는 후기의 경우에는 대체로 긍정적 성향을 띄었으나 감정평가와 관련한 후기에서는 부정적 성향을 띄는 후기들이 많았다고 설명하였다. 박은아(2007)는 온라인 후기가 제품의 특성과 같은 객관적 사실에 관한 내용일 경우 긍정적 성향의 후기들이 잠재적 고객에게 미치는 영향이 더 높았으며, 반대로 후기가 주관적 평가와 같은 내용일 경우 잠재적 고객들은 부정적 성향의 후기에 더 관심을 보였다고 온라인 후기의 방향성에 대하여 주장하였다.

이정현(2013)은 호텔에 관한 후기 내용 분석을 통해 고객이 지각하는 호텔의 가치 차원에 대하여 설명하였다. 긍정적 성향을 띤 후기와 부정적 성향을 띤 후기 간의 내용 비교를 통하여 고객이 지각하는 기능적, 정서적 호텔 가치에는 어떤 것들이 포함되는지를 설명하며 고객이 남긴 후기의 성향이 부정적일수록 자신의

경험이나 느낀 점을 상세하게 표현한다고 하였다. 조승연 등(2015)은 온라인 후기를 이용한 추천시스템의 개발을 통하여 고객의 제품 평가 기준과 구매의도에 영향을 미치는 요소를 설명하였다.

## 2.3 감성 분석

### 2.3.1 감성분석의 정의

온라인에 존재하는 텍스트 데이터를 수집, 분석하여 데이터 안에 남아있는 작성자의 감성의 종류와 정도, 감성이 일어난 대상에 대하여 분석하는 방법을 감성분석(Sentiment Analysis 또는 Opinion Mining)이라 한다(Wiebe, 2004; Wang, 2011; Yu, 2013). Liu(2010)에 의하면 감성분석이란 텍스트 데이터 안에 내재된 고객의 감성이나 의견 등에 대한 기계적 분석이며, 상품, 서비스, 기관, 특정 사건, 이슈 등에 관해 후기, 블로그, 트위터(Twitter)나 페이스북(Facebook)과 같은 소셜네트워크서비스 등에 남긴 텍스트를 수집, 분석하여, 사람들의 감정, 태도, 평가, 의견 등을 분석하는 연구이다(Liu, 2012).

또한 Pang et al.(2008)은 텍스트에서 단어를 계산하여 텍스트 내에 담겨진 의견, 감정, 주체성을 분석하는 기법이라고 감성분석을 설명하였고, Balahur et al.(2010)은 “사람들의 의견, 감정, 감정, 평가, 믿음에 대한 언어적 표현”이라 정의 내렸다. Fan et al.(2013)은 기업의 고객에 대한 이해를 도와 현대 기업들이 보다 나은 전략을 세우고 고객에게 서비스를 제공하는 분석이라 하였으며, Feldman(2013)은 기업의 최종 의사결정 단계에 특히 많이 활용되는 분석

이라고 설명하였다.

감성분석은 최근 많은 주목을 받고 있는데 이는 급증하는 인터넷 사용자와 발전한 인터넷 기술로 온라인 상의 정보가 하루가 다르게 늘어나며, 많은 온라인 사용자들이 이러한 정보를 기반으로 제품이나 서비스의 구매를 결정하고 기업의 가치를 판단하기 때문이다. 후기를 포함한 온라인 정보는 단순히 다른 고객의 구매 의도에 영향을 미칠 뿐만 아니라, 후기 정보 안에 담겨진 작성자의 감성을 통하여 제품에 대한 고객들의 만족도를 판매자에게 제공한다(Liu, 2013). 이러한 감성분석은 이용자의 단순 감성 뿐만 아니라, 특정 대상에 대한 이용자의 평가, 태도, 느낌 등의 의견까지도 이해하는 데 목적이 있다. 감성분석은 다양한 분야의 기업과 사회 여러 분야에서 활용되어지며, 다양한 방법론에 관해서 다수의 연구가 진행되고 있다.

### 2.3.2 감성분석 방법

전자상거래의 대중화로 많은 이용자들이 특정 제품에 대한 평가나 느낀 점 등을 온라인에 후기 형태로 남기게 되고, 후기에 표출된 기존 고객들의 감성은 잠재적 고객들의 구매 의사 결정에 큰 영향을 미치고 있다. 이러한 이유로 기업에게는 고객들이 남긴 후기 내용을 분석하여 그들의 감정, 평가, 태도 등을 정확히 이해하는 것이 중요한 과제로 대두되고 있다(Shi et al., 2011).

감성분석 방법은 크게 Supervised learning 과 Unsupervised learning으로 분류된다. 후기 텍스트를 작성할 때 작성자들이 느낀 정도나 방식에 따라 표현하는 다양한 어휘들을 감성사전을 이용하여 유의어의 집단으로 분류하는 기



법을 Supervised learning이라 하고, 후기 텍스트 내의 감성이 단순히 긍정적인 성향인지, 부정적인 성향인지를 판단하는 기법을 Unsupervised learning이라고 한다.

Turney(2002)는 문장의 성향을 나타내는 단어들을 형용사와 부사로 정하고, 문장 내의 단어들에 대한 형태소 나열 방법을 이용하여 후기의 성향이 긍정인지 부정인지를 분류하였다. Pang et al.(2002)은 영화 후기 데이터를 먼저 주제별로 분류한 뒤, Naïve Bayes, Maximum Entropy, Support Vector Machine 등의 기계학습 분류기를 이용하여 긍정적 성향과 부정적 성향을 판별하였다. Liu(2010)는 분류기를 이용하여 SNS 상의 텍스트 메시지의 감성이 긍정 혹은 부정적 성향인지에 대하여 알아보았다.

Nasukawa et al.(2003)은 Sentiment Lexicon을 감성사전으로 이용하여 후기 데이터의 긍정적 또는 부정적 성향에 대해 판별하고 특정 성향과의 관련성에 대해 분석하였으며, Liu et al.(2004)은 WordNet을 이용하여 후기 데이터의 대상에 대한 특성과 텍스트의 성향을 감성별로 분류하였다. Liu et al.(2007)은 문장의 구조에 대한 분석을 통해 긍정과 부정으로 판단하기 힘든 의미론적으로 객관적인 단어들이 어떤 성향의 단어들과 연결되는지를 판정하였고 Opinion words를 감성사전으로 이용하였다.

### 2.3.3 감성분석 수준

일반적으로 텍스트의 형태에 따라 감성분석을 크게 세 가지 수준, 즉 문서 수준(Document Level), 문장 수준(Sentence Level), 관점 수준(Entity and Aspect Level)으로 나누어 설명할 수 있다(Pang, 2008; Liu, 2012). 문서 수준이란

텍스트 문서 전체를 하나의 단위로 분석하는 수준으로서 문서 전체에서 나타나는 글쓴이의 종합적인 감성을 파악해 낸다. 하지만, 문서 전체가 일관성 있게 특정 관점에서 작성된 경우에 이용되며 복합적인 관점이 포함된 문서에는 부적합하다.

문장 수준은 텍스트 문서를 문장 단위로 나누어 각 문장에서 표현되는 감성을 분석하는 것이다. 주관적인 관점이나 표현을 포함하는 문장들만을 따로 분류하여 문장 별로 내포하는 의미가 긍정인지, 부정인지, 중립적인지를 판단하는 것이다. 하지만 객관적인 사실을 설명하는 문장조차 글쓴이의 의견이나 감정을 간접적으로 표현하는 경우가 있어 이 부분에 대한 연구는 보완되어야 하는 상황이다.

관점 수준은 Feature Level이라고도 불리는 데 문서 수준이나 문장 수준의 분석과는 달리 텍스트에 포함된 글쓴이의 감정, 평가, 의견 등이 무엇을 대상으로 발생했는지를 알아낼 수 있는 분석이다. 관점 수준의 가장 큰 특징은 비록 문서나 문장에서 느껴지는 글쓴이의 감정이 긍정적일지라도 내용 속에 언급된 특정 대상에 대한 감정은 부정적일 수 있는데 이러한 감정을 개별적으로 분석해 낼 수 있다.

### 2.3.4 감성사전

감성분석을 할 때 많이 이용하는 감성사전으로 Opinion Lexicon, MPQA, SentiWordNet을 들 수 있다. Opinion Lexicon은 Liu(2010)가 관점수준의 감성분석을 하기 위해 만든 감성사전으로, 2004년부터 이루어진 감성분석에서 긍정적인 의미를 나타내는 단어 2006개와 부정의 의미를 가진 단어 4783개를 추출하여 구성하였

다. MPQA는 Wiebe et al.(2005)이 만든 감성 사전으로 하나의 단어라도 문법적인 사용 용도에 따라 의미하는 감성 및 감성지수가 달라지는 것에 주목하여 각 단어에 대해 서로 다른 관점에 따라 감성점수들을 부여하여 만든 사전이다. Wilson et al.(2005)은 MPQA가 하나의 단어에서 절까지, 형태와 길이가 매우 다양한 감성 표현방식에 대해 감성지수를 측정하는 방식이라고 설명하였다.

본 연구에서 사용하는 감성사전인 SentiWordNet은 WordNet을 기반으로 만들어진 감성사전이다. WordNet은 Miller(1995)가 만든 단어 의미네트워크로서, 118000개 이상의 서로 다른 단어 형태와 90000개 이상의 다양한 의미, 그리고 166000개의 단어 쌍으로 이루어져 있다. 또한 유의어(Synonymy)와 반의어(Antonymy)를 포함하는 영어 단어들의 관계에 대한 정보를 전자적으로 제공한다.

SentiWordNet은 WordNet의 단어들 중 감성을 표현하는 단어들을 모아 만든 사전으로, 사전 안에 포함된 단어들이 문장 내에서 이용되는 방식과 의미 별로 단어의 성향이 긍정인지, 부정인지, 또한 성향의 강도 정보가 태그되어 있다(Esuli, 2006). SentiWordNet에 포함된 단어들의 성향에 대한 강도는 긍정일 경우 0에서 1사이의 값으로 계산되며, 값이 1에 가까울수록 매우 긍정적이라 볼 수 있다. 반대로 부정적인 감성점수는 -1에서 0사이의 값으로 표현되며, -1에 가까울수록 부정적인 성향이 강해진다. 0은 단어가 객관적으로 사용되었음을 의미

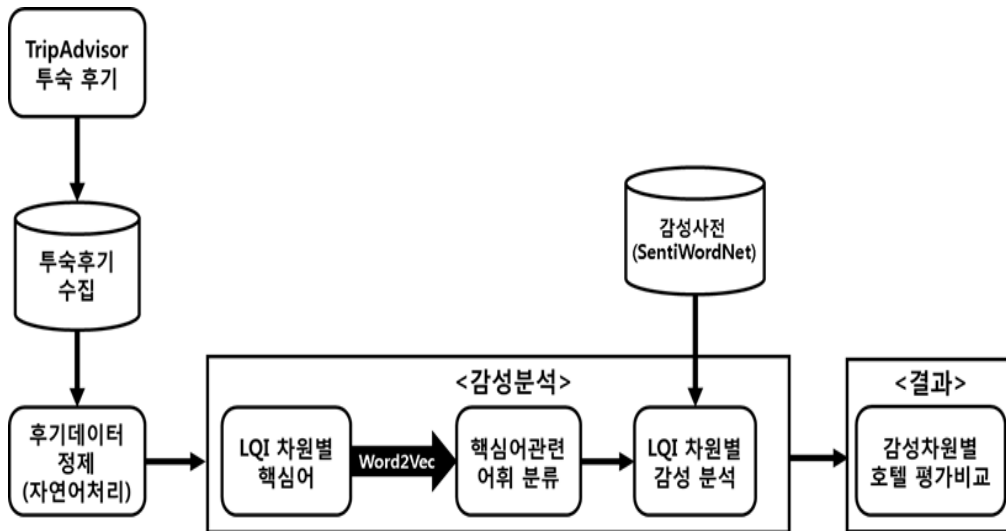
한다. 즉, 주관적이지 않은(텍스트를 작성할 때 작성자의 주관이 반영되지 않은) 경우를 의미하므로 감성점수는 0으로 표현된다.

### III. 분석 방법

#### 3.1 후기 분석 설계

<그림 1>은 본 연구의 프레임워크를 도식화한 것이다. 본 연구는 온라인에서 내외국인 여행객들이 호텔을 대상으로 남긴 영어로 작성한 숙박 후기를 수집한다. 수집한 후기는 텍스트 데이터이므로 텍스트 정제를 위한 일반적인 자연어 처리 작업을 시행한다. 정제가 끝난 후기에서 word2vec 알고리즘을 이용하여 LQI 차원별 어휘들을 추출하고 단어 간의 cosine 유사도 값을 구한 후, SentiWordNet 감성사전에서 관련 어휘들의 감성점수를 구한다. 감성분석을 위해 본 연구에서 계산하는 감성점수는 어휘들의 cosine 유사도 값과 감성사전의 감성점수를 곱하여 구한다.

온라인 숙박 후기 데이터 수집을 위하여 R 프로그래밍 언어를 이용한다. R 프로그래밍 언어는 오픈소스 소프트웨어를 제공하는 회사인 R Foundation의 R Development Core Team이 통계적 계산을 위해 개발한 소프트웨어이다. 본 연구에서는 후기 데이터를 수집하여 데이터베이스에 저장하기까지의 과정은 R 프로그래밍



<그림 1> 연구 모형

언어에서 제공하는 rvest 패키지를 이용하며, word2vec 알고리즘 구현을 위하여 tmcn.word2vec 패키지를 이용한다.

자연어 처리 작업은 RapidMiner 소프트웨어를 이용한다. RapidMiner는 사용자 친화적인 데이터마이닝 소프트웨어로서 다양한 기계학습, 텍스트마이닝, 예측 분석 등을 지원하며 GUI 방식으로 보다 편리하고 빠르게 데이터 분석을 진행할 수 있는 장점을 제공한다. word2vec의 cosine 유사도 값과 SentiWordNet의 감성점수를 곱하여 LQI 차원별 핵심어의 감성점수를 계산할 때는 C언어를 이용한다.

### 3.2 투숙 후기: TripAdvisor.com

본 연구에서는 온라인 후기의 원조 격이자 여행 및 투숙 관련 후기 제공 사이트로서 역사와 공신력에서 대표 주자라 할 수 있는 TripAdvisor를 선정하여 분석에 필요한 온라인

숙박 후기를 수집하였다. TripAdvisor는 Expedia, Priceline, Agoda 등과 함께 대표적인 온라인 여행 중개업체로서 전 세계 많은 여행객들이 이용하고 있다. TripAdvisor를 방문하는 여행객들은 매일 약 3억 4,000만 명에 달하며, 이들에게 전 세계 45개국 490만 개 이상의 숙박시설, 음식점 및 관광명소에 대한 2억 2,500만 건 이상의 평가 정보를 제공하고 있으며, 투숙 후기 작성 시 TripAdvisor 사이트 자체적으로 숙박시설의 서비스 품질을 평가하는 척도를 제공한다.

### 3.3 후기 데이터 자연어 처리 과정

기본적으로 후기 데이터들은 시제나 문법에 맞지 않게 작성되고 대소문자 구별이 무시되거나 특수 기호를 포함하여 작성된 경우가 많으며 인터넷 용어 및 초성만을 이용하여 작성된 경우도 있다. Mathiak et al.(2015)은 감성분석

에 사용되는 데이터의 양이 방대하기 때문에 Stemming 단계를 통하여 단수 형태와 복수 형태처럼 문법적 의미가 같은 단어는 하나의 단어로 통일시켜야 하며, Stopwords 단계를 통하여 빈번하게 사용되지만 중요한 의미를 내포하지 않는 단어들은 제거함으로써 분석에 저해되는 요소를 줄여야 한다고 설명하였다. 또한, Xiang et al.(2015)은 텍스트 데이터에 대한 자연어 처리 작업을 통해 불필요한 단어의 제거 및 정제는 분석 결과에 대한 내용 타당성을 확보한다고 설명하였다. 따라서 본 연구에서는 후기 분석을 위해 Tokenization, Normalization, Stemming, Stopwords 등의 일반적인 정제 과정을 거친다.

Tokenization 단계는 문단, 문장, 단어들을 이루는 알파벳을 제외한 각종 부호 및 기호들, 알파벳들을 새롭게 조합하여 만든 이모티콘 등을 제거하는 작업이다. 온라인에서 작성된 텍스트의 경우 각종 부호 및 기호들을 이용하여 만들어진 의미 없는 문자열을 포함하고 있는 경우가 많다. 이러한 기호들을 제거하지 않은 채 데이터를 분석할 경우 사전적 의미를 갖고 있는 단어들의 문장 내 비율을 정확하게 분석해 내기 힘들 뿐만 아니라 단어들 간의 사전적 의미 관계에 대한 분석 결과는 정확도가 떨어진다(Manning, 1999).

Normalization 단계는 사용자가 온라인에 남긴 후기 데이터에서 Tokenization 단계를 거친 단어들을 대상으로 단어들이 올바른 철자로 이루어져 있는지를 판별하는 작업이다. 동일한 단어일지라도 대소문자나 띄어쓰기에 따라 서

로 다른 단어로 인식되어지기 때문에 같은 단어일 경우 동일한 형태로 변형시켜주어야 한다. Stemming 단계는 시제나 단수/복수 등, 경우에 따라 형태가 달라지는 단어들에 대하여 어근을 찾아서 어근이 같은 경우는 동일한 단어로 간주하여 분석한다. Stopwords 단계는 문장에서 많이 사용하지만 중요한 의미가 없는 단어들을 제외하는 작업이다. 영어의 관사, 접속사, 전치사, 대명사처럼 문장 내에서의 빈도수는 높지만, 특별한 의미가 없는 단어들을 제거하는 작업이다(Xiang et al., 2015; Mathiak et al., 2015).

### 3.4 핵심어 분류

한편, 감성 어휘를 분류하는 작업은 LQI의 5가지 차원을 기준으로 진행한다. 앞서 언급한 바와 같이 LQI는 신뢰성, 대응성, 유형성, 믿음성, 의사소통의 5가지 차원으로 이루어져 있는데, 각각의 차원에 대한 핵심어를 Getty et al.(2003)이 제시한 “LQI 측정 설문 문항” (<표 1>)에서 언급하는 감성 어휘들을 중심으로 정한다. 이를 기준으로 삼아 내외국인 여행객들이 호텔들의 서비스 품질에 대해 후기에 남긴 어휘들의 감성분석을 실시함으로써 LQI 차원별로 감성의 정도를 분석하고 호텔 서비스 품질에 대한 고객 만족도를 비교해 보기로 한다. <표 2>는 “LQI 측정 설문 문항”에서 추출한 핵심적인 감성 어휘들을 LQI 차원별로 나열한 것이다.

<표 1> LQI 차원별 26가지 측정 항목 (굵은 단어는 핵심 감성어)

유형성(Tangibility)	
1	The front desk was visually appealing
2	The employees had <b>clean, neat</b> uniforms
3	The restaurant's atmosphere was <b>inviting</b>
4	The shops were <b>pleasant</b> and attractive
5	The outdoor surroundings were visually <b>attractive</b>
6	XYZ hotel was bright and <b>well lighted</b>
7	XYZ hotel's interior and exterior were <b>well maintained</b>
8	XYZ hotel was <b>clean</b>
신뢰성(Reliability)	
9	My reservation was handled <b>efficiently</b>
10	My guestroom was ready as <b>promised</b>
11	TV, radio, air-conditioning, lights, and other mechanical equipment worked <b>properly</b>
12	I got what I paid for
대응성(Responsiveness)	
13	Employees responded <b>promptly</b> to my requests
14	<b>Informative</b> literature about the XYZ hotel was provided
15	Employees were <b>willing to</b> answer my questions
16	Employees responded <b>quickly</b> to solve my problems
17	Room service was <b>prompt</b>
믿음성(Confidence)	
18	Employees <b>knew about</b> local places of interest
19	Employees treated me with <b>respect</b>
20	Employees were <b>polite</b> when answering my questions
21	XYZ hotel provided a <b>safe</b> environment
22	The facilities were <b>conveniently</b> located
의사소통(Communication)	
23	Charges on my account were <b>clearly</b> explained
24	I received undivided <b>attention</b> at the front desk
25	Reservation staff attempted to ascertain my <b>particular</b> needs
26	Employees <b>anticipated</b> my needs

<표 2> LQI 차원별 핵심 감성어휘

LQI 차원	핵심어
유형성	visually appealing, appealing, clean, neat, inviting, pleasant, attractive, bright, well lighted, well maintained
신뢰성	efficiently, promised, properly, ready
대응성	prompt, promptly, informative, willing to, quickly
믿음성	know about, respect, polite, safe, convenient
의사소통	clearly, attention, particular, anticipated

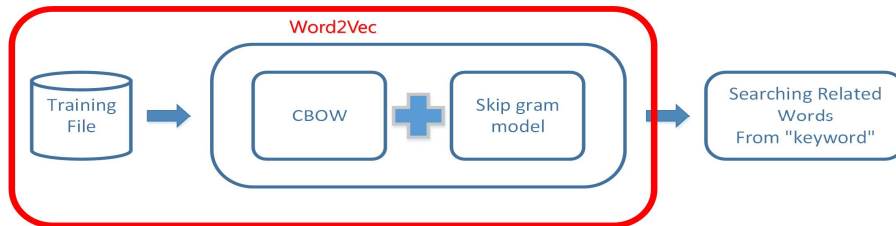
### 3.5 감성분석

#### 3.5.1 Word2vec

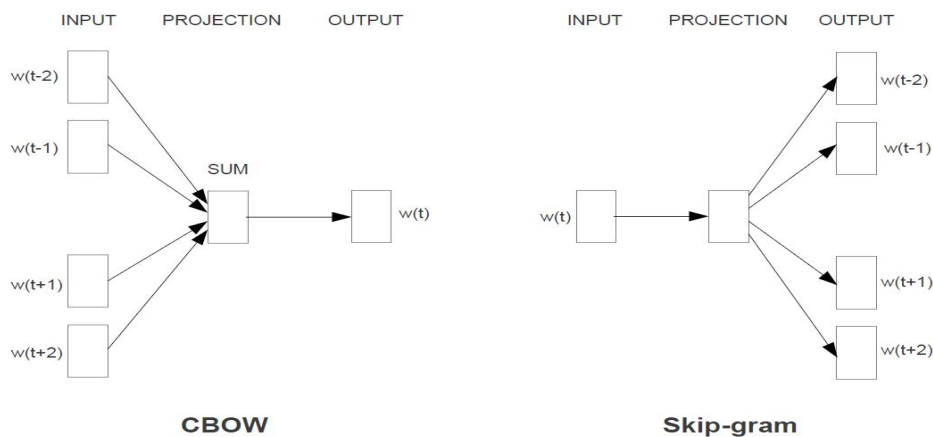
감성분석에서는 word2vec 알고리즘을 이용하여 LQI 차원별로 핵심 감성어에 대한 유의어 및 관련성이 높은 어휘들을 추출한다. Word2vec은 Google의 Mikolov et al.(2012)이 개발한 자연어 기계학습 알고리즘이다(<그림 2>). Word2vec은 Neural Language Network 알고리즘 중 최신 모형으로 word2vec 알고리즘을 분석에 이용할 경우, 먼저 텍스트 데이터 전체에 대한 학습(Training)을 거쳐 텍스트 안의 모든 단어들을 벡터화 한다. 단어 간의 벡터화는 CBOW (Continuous Bag of Words) 알고리

즘과 CSG (Continuous Skip Gram) 알고리즘을 통하여 형성된다.

CBOW 알고리즘은 텍스트 내의 단어들의 앞뒤 위치 및 거리를 계산하며, CSG 알고리즘은 텍스트 안의 모든 단어들을 구분 짓고 앞뒤에 배치된 단어들 간의 관련성을 로그확률함수를 이용하여 계산하며 단어 간의 유사성/관련성을 cosine 유사도라는 값으로 표현하는데 cosine 유사도는 0에서 1사이의 값을 가지고 1에 가까울수록 두 단어 사이의 거리가 가까움을 나타내며 관련성이 높음을 의미한다 (Mikolov et al., 2012). <그림 3>은 두 알고리즘의 메커니즘을 도식화한 것이다.



<그림 2> word2vec 알고리즘 구조



<그림 3> CBOW와 CSG 알고리즘의 구조

본 연구는 수집한 후기들을 호텔 별로 하나의 텍스트 문서로 만들고 word2vec 알고리즘으로 학습시킴으로써 후기들로 이루어진 문서 내에서 단어 간의 관련성을 추출한다. 후기에서 사용된 단어 간의 관계를 계산함으로써 후기 내용에서 관련성이 높은 단어들의 유추를 가능하게 한다.

### 3.5.2 감성점수

LQI 차원별 핵심 감성어와 cosine 유사도가 높은 어휘들의 감성의 정도는 감성사전인 SentiWordNet을 이용하여 계산한다. SentiWordNet은 문장 내에서의 단어의 위치와 의미에 따른 감성의 정도를 각각 수치로 표현하는데 이를 감성점수라고 한다(Esuli, 2006).

SentiWordNet에는 전체 15가지의 감성점수가 저장되어 있는데, 긍정을 나타내는 감성점수는 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875 중 하나의 값을 가지게 되며, 값이 1에 가까울수록 긍정적인 성향이 높음을 의미한다. 부정을 나타내는 감성점수는 -0.125, -0.25, -0.375, -0.5, -0.625, -0.75, -0.875 중 하나의 값을 가지게 되며, 마찬가지로 -1값에 가까울수록 부정적

인 성향이 높음을 의미한다. 동일한 단어라도 문장 내 사용 용도에 (예를 들어, 동일한 단어가 형용사로 사용될 수 있고 부사로도 사용될 수 있음) 따라 다른 감성점수를 가질 수 있으므로 본 연구에서는 이런 경우 감성점수의 평균값을 구하여 사용한다.

최종적으로 LQI의 각 차원별 핵심 감성어의 감성점수는 word2vec 알고리즘이 도출한 cosine 유사도의 합계를 word2vec 알고리즘이 알려준 관련 단어들의 개수로 나눈 후 SentiWordNet을 이용한 평균값을 곱하여 계산한다. 감성점수 계산 방법을 수식으로 표현하

$$\text{면, } \frac{\sum_{i=1}^N C_i \times S}{N} \text{ 와 같으며, 이때 } C \text{는 cosine}$$

유사도,  $N$  은 word2vec이 알려준 관련 유사 단어 개수,  $S$ 는 SentiWordNet의 단어별 감성점수 평균값을 의미한다. SentiWordNet에 저장되어 있는 감성점수를 cosine 유사도로 가중평균함으로써 대상이 되는 후기에서 LQI 감성어에 근접할수록(유사할수록) 긍정 감성은 더 큰 긍정으로, 부정 감성은 더 큰 부정으로 가중해주는 효과를 얻을 수 있다.

<표 3> 호텔별로 수집한 후기 개수

호텔 \ 년도	2015	2014	2013	2012	2011	2010
서울 A 호텔	92	88	34	48	40	21
서울 B 호텔	235	266	201	116	57	32
서울 C 호텔	176	181	196	94	42	25
서울 D 호텔	223	148	147	143	83	27

<표 4> 호텔별 word2vec 분석 결과

호텔 명	년도	후기 개수 (영어)	후기 내의 전체 단어 개수	벡터화 된 단어 개수	학습률 (alpha)
A 호텔	2010~2011	61	2356	196	0
	2012~2013	82	2806	228	0
	2014~2015	180	6907	416	0
	소계	323	12069	840	
B 호텔	2010~2011	89	3160	232	0
	2012~2013	317	15138	665	0.0000001
	2014~2015	501	24748	898	0
	소계	907	43046	1795	
C 호텔	2010~2011	67	2842	222	0
	2012~2013	290	11448	552	0
	2014~2015	357	13664	591	0
	소계	714	27954	1365	
D 호텔	2010~2011	110	4383	292	0
	2012~2013	290	14764	694	0.025
	2014~2015	371	15485	714	0.006375
	소계	771	34632	1700	

## IV. 분석 결과

### 4.1 데이터 수집 및 정제

본 연구에서는 TripAdvisor 사이트로부터 호텔별 투숙 후기들을 수집하기 위하여 R 프로그램을 이용하였다. R 프로그램 내의 rvest 패키지를 이용하여 수집된 투숙 후기들은 해당 사이트의 구조를 분석하여 TripAdvisor 데이터베이스에서 데이터를 직접 수집하는 방식으로 코딩하였으며 후기별로 작성자 ID, 후기 제목, 후기 내용을 수집하였다. 각 호텔별 투숙 후기들의 개수는 <표 3>과 같다.

본 연구에서는 2010년~2015년 동안의 국내 대표 호텔인 A호텔과 B호텔, 그리고 대표적인 외국 호텔인 C호텔, D호텔에 남겨진 숙박 후기들을 수집하였다. 수집된 후기를 호텔별로 하나

의 텍스트 파일로 묶어 RapidMiner 프로그램을 이용하여 텍스트 정제 작업을 진행하였다. 특히, SentiWordNet 사전의 단어들과 정제된 단어 간의 형태를 맞추기 위하여 WordNet 기반의 Stemming 작업을 실시하였다.

### 4.2 Word2vec 결과

정제작업이 끝난 후기 파일을 word2vec 알고리즘을 이용하여 분석할 때 R 프로그램의 tmcn.word2vec 패키지를 활용하였다. 이 패키지는 프로그래밍 언어인 C를 이용하여 작성되었으며 텍스트 파일에 대한 학습을 통해 파일 내의 단어들을 벡터로 처리하였다. 학습이 끝난 후 패키지가 제공하는 distance 함수를 이용하여 핵심 감성어와 cosine 유사도가 높은 관련 단어들을 추출할 수 있었다. 호텔별 후기 파일



에 word2vec 알고리즘을 적용시킨 결과로서 후기 파일을 구성하는 전체 단어의 개수, 벡터화된 단어 개수와 word2vec 알고리즘의 학습율은 <표 4>와 같다.

<표 4>에서 “후기 내의 전체 단어 개수” 칼럼은 호텔별로 수집한 영어 후기들을 2년 단위로 나눈 후 (2010년-2011년, 2012년-2013년, 2014년-2015년), 나뉜 후기들을 별도의 파일로 묶고 각각의 파일에 대해 자연어 처리 작업을 거친 결과, 파일 안에 남아 있는 단어의 개수를 보여주고 있다. “벡터화 된 단어 개수” 칼럼은 자연어 처리가 끝난 후기 파일에서 word2vec 알고리즘을 이용하여 학습된 단어들로서 후기 내에서 빈도수가 높고 관련 어휘들을 도출 가능한 단어들의 개수를 의미한다. “학습률(alpha)”은 각 후기 파일들에 대한 word2vec 알고리즘의 학습률을 의미하는데 학습률의 값이 0에 가까울수록 완벽하게 학습이 진행되었다고 말할 수 있다. 각 파일에 대한 학습률이 대부분 0에 근접하므로 모든 후기에 대한 word2vec 알고리즘의 학습이 잘 수행되었음을 알 수 있다.

word2vec의 학습이 끝난 후 LQI 차원별 핵심 감성어와 관련성이 높은 어휘를 추출하는 방법은 <표 5>를 이용하여 설명할 수 있다. 예를 들어, 2010년~2015년 사이에 작성된 A호텔의 후기들을 word2vec 알고리즘으로 학습시킨 후, LQI의 유형성 차원의 핵심 감성어인 “clean”과 믿음성 차원의 핵심 감성어인 “respect”에 대하여 cosine 유사도가 높은 어휘를 각각 40개씩 추출해볼 수 있다.

유형성 차원의 핵심 감성어인 “clean”에 대해 cosine 유사도가 가장 높은 어휘는 “amenities”로 나타났으며(0.999926), 그 뒤를 이어 “friendly”의 cosine 유사도가 0.999394를 기록하였다. 이 단어 역시 “clean”과 매우 관련성이 높다고 말할 수 있다. cosine 유사도의 값이 1에 가까울수록 두 단어 사이의 관련성이 높다. 마찬가지로, 믿음성 차원의 핵심 감성어인 “respect”의 기준에서 가장 관련이 높은 단어는 “friendly”로 나타났고(0.983350), “comfortable”이 그 뒤를 이어 높은 cosine 유사도를 나타내었다(0.982344).

#### 4.3 감성점수 계산

LQI 차원별 핵심 어휘에 대한 감성점수를 구하는 방법은 <표 6>, <표 7>을 이용하여 설명할 수 있다. 유형성 차원의 핵심 감성어인 “clean”과의 관련 어휘들 중에서 SentiWordNet 사전에 감성점수가 표시되어있는 단어들은 “amazing”, “big”, “friendly”, “great”, “help”, “helpful”, “light”, “love”, “wonderful” 등이다 (<표 6>). SentiWordNet 사전에 있는 단어들은 문법별 사용 용도에 따라 감성점수가 다르므로 평균값을 계산하여 단어의 감성 정도를 표현하였다. 그리고 난 후에 cosine 유사도와 단어별 평균 감성점수를 곱하고 더한 후 관련 단어들의 개수로 (“clean”의 경우는 9) 나누어 “clean”의 감성점수로 삼았다. <표 7>은 믿음성 차원의 핵심 감성어인 “respect”에 대해 감성점수를 동일한 방식으로 계산한 결과이다.

<표 5> 핵심 감성어 “clean”과 “respect”에 대해 cosine 유사도가 높은 어휘 40개

핵심어 : clean			핵심어 : respect		
순서	단어	cosine 유사도	순서	단어	cosine 유사도
1	amenities	0.999926	1	friendly	0.983350
2	friendly	0.999394	2	comfortable	0.982344
3	amazing	0.999346	3	speak	0.978475
4	city	0.998964	4	reception	0.974719
5	big	0.998786	5	trying	0.974041
6	include	0.998726	6	get	0.973301
7	front	0.998296	7	speaking	0.971923
8	door	0.997544	8	professional	0.971730
9	concierge	0.997445	9	polite	0.971677
10	customer	0.997429	10	nice	0.969497
11	food	0.997242	11	thing	0.969244
12	breakfast	0.997165	12	decent	0.969224
13	cleanly	0.997144	13	coffee	0.968268
14	bed	0.996031	14	overprice	0.968215
15	floor	0.995705	15	problem	0.967793
16	bridge	0.995050	16	excellent	0.967538
17	bedroom	0.995318	17	money	0.966381
18	helpful	0.989860	18	feel	0.966109
19	rooms	0.989460	19	example	0.964058
20	marina	0.989379	20	recently	0.963642
21	price	0.989318	21	why	0.962933
22	give	0.989092	22	modern	0.962854
23	pool	0.988659	23	meet	0.962677
24	locate	0.988483	24	love	0.962410
25	love	0.988433	25	course	0.962396
26	lot	0.987474	26	worth	0.961913
27	help	0.987338	27	impress	0.961509
28	gym	0.987336	28	prefer	0.960779
29	great	0.987021	29	provide	0.959951
30	light	0.986903	30	maybe	0.959516
31	guest	0.986317	31	old	0.955358
32	seem	0.985076	32	main	0.952318
33	spacious	0.979630	33	cover	0.948970
34	top	0.979611	34	expect	0.948541
35	toilet	0.979234	35	renovate	0.941141
36	think	0.977782	36	spotless	0.937603
37	side	0.977051	37	pretty	0.935435
38	window	0.976348	38	date	0.931688
39	wonderful	0.975527	39	bright	0.929450
40	take	0.975421	40	disappoint	0.925156

<표 6> 핵심 감성어 “clean”의 감성점수 계산표

순서	관련 단어	cosine 유사도	단어별 감성점수(평균값)	cosine유사도 × 단어별 감성점수(평균값)	clean에 대한 최종 감성점수
1	amenities	0.999926	0	0	0.394686
2	friendly	0.999394	0.344	0.343791	
3	amazing	0.999346	0.5	0.499673	
4	city	0.998964	0	0	
5	big	0.998786	0.675	0.67418	
6	include	0.998726	0	0	
7	front	0.998296	0	0	
8	door	0.997544	0	0	
9	concierge	0.997445	0	0	
10	customer	0.997429	0	0	
11	food	0.997242	0	0	
12	breakfast	0.997165	0	0	
13	cleanly	0.997144	0.125	0.124643	
14	bed	0.996031	0	0	
15	floor	0.995705	0	0	
16	bridge	0.995050	0	0	
17	bedroom	0.995318	0	0	
18	helpful	0.989860	0.25	0.247465	
19	rooms	0.989460	0	0	
20	marina	0.989379	0	0	
21	price	0.989318	0	0	
22	give	0.989092	0	0	
23	pool	0.988659	0	0	
24	locate	0.988483	0	0	
25	love	0.988433	0.575	0.568348	
26	lot	0.987474	0	0	
27	help	0.987338	0.291	0.287315	
28	gym	0.987336	0	0	
29	great	0.987021	0.341	0.336574	
30	light	0.986903	0.135	0.133231	
31	guest	0.986317	0	0	
32	seem	0.985076	0	0	
33	spacious	0.979630	0	0	
34	top	0.979611	0	0	
35	toilet	0.979234	0	0	
36	think	0.977782	0	0	
37	side	0.977051	0	0	
38	window	0.976348	0	0	
39	wonderful	0.975527	0.75	0.731645	
40	take	0.975421	0	0	

<표 7> 핵심 감성어 “respect”의 감성점수 계산표

순서	관련 단어	cosine 유사도	단어별 감성점수(평균값)	cosine유사도 × 단어별 감성점수(평균값)	respect에 대한 최종 감성점수
1	friendly	0.983350	0.344	0.338272	0.249608
2	comfortable	0.982344	0.219	0.215133	
3	speak	0.978475	0	0	
4	reception	0.974719	0	0	
5	trying	0.974041	0	0	
6	get	0.973301	0	0	
7	speaking	0.971923	0	0	
8	professional	0.971730	0.028	0.027208	
9	polite	0.971677	0.4	0.388671	
10	nice	0.969497	0.594	0.575881	
11	thing	0.969244	0	0	
12	decent	0.969224	0.392	0.379935	
13	coffee	0.968268	0	0	
14	overprice	0.968215	0.5	0.484107	
15	problem	0.967793	0.375	0.362922	
16	excellent	0.967538	1	0.967538	
17	money	0.966381	0	0	
18	feel	0.966109	0.151	0.145882	
19	example	0.964058	0	0	
20	recently	0.963642	0	0	
21	why	0.962933	0	0	
22	modern	0.962854	0.06	0.057771	
23	meet	0.962677	0	0	
24	love	0.962410	0.575	0.553385	
25	course	0.962396	0	0	
26	worth	0.961913	0.2	0.192382	
27	impress	0.961509	0	0	
28	prefer	0.960779	0.312	0.299763	
29	provide	0.959951	0	0	
30	maybe	0.959516	0	0	
31	old	0.955358	0.107	0.102223	
32	main	0.952318	0	0	
33	cover	0.948970	0	0	
34	expect	0.948541	0	0	
35	renovate	0.941141	0.167	0.15717	
36	spotless	0.937603	0	0	
37	pretty	0.935435	-0.5	-0.467717	
38	date	0.931688	0	0	
39	bright	0.929450	0.208	0.193325	
40	disappoint	0.925156	-0.25	-0.231289	

#### 4.4 분석 결과

<표 8> 서울지역 호텔에 대한 LQI 차원별 감성분석 결과 (2010년-2011년)

2010년 - 2011년 LQI 차원		A 호텔	B 호텔	C 호텔	D 호텔
유형성 (Tangibility)	(visually)appealing				
	clean	0.28579	0.396626	0.396816	0.210863
	neat				
	inviting				
	pleasant				
	attractive				
	bright				
	well lighted well maintained				
신뢰성 (Reliability)	efficiently				
	promised		0.192057	0.240999	
	properly				
대응성 (Responsiveness)	prompt		0.250572		
	promptly			0.350195	
	informative				
	willing to				
믿음성 (Confidence)	know about				
	respect	0.421089		0.154816	0.285514
	polite		0.472186	0.339292	0.48312
	safe				
의사소통 (Communication)	conveniently				
	clearly	0.219312	0.285924	0.366681	0.338889
	attention				
	particular				
	anticipated				

2년 단위로 후기들을 나누어 시간의 변화에 따른 호텔별 감성점수의 변화를 분석하였다. <표 8>은 2010년부터 2011년 사이에 작성된 서울지역 호텔 숙박 후기를 LQI 차원별로 감성 분석한 결과이다. 먼저 유형성 차원을 살펴보면 서울지역 호텔의 “clean”에 대한 감성점수를 확인할 수 있다. 분석 대상이 된 호텔들에서 모두 긍정적인 감성점수가 나왔지만 B호텔과 C호텔의 감성점수가 보다 높게 나타나 이 호텔들에 대한 고객들의 긍정적인 감성 표현이 높았다고 할 수 있다.

신뢰성 차원의 “promised”는 B호텔과 C호텔

에서 감성점수를 확인할 수 있었는데 C호텔이 B호텔보다 좋은 점수를 받았다. 이는 C호텔의 대고객 서비스가 사전에 약속한 서비스와 차이가 적었음을 의미한다. 투숙객들에 대한 직원의 친절성 (“respect”)에서는 A호텔이, 예절성 (“polite”)에서는 D호텔이 가장 높은 긍정 점수를 받았다. 의사소통 차원의 “clearly”는 직원이 얼마나 여행객들로부터 요구를 전달받고 분명하게 정보를 전달했는지를 설명하는 핵심어로서 C호텔, D호텔, B호텔, A호텔의 순으로 긍정적인 감성이 높았다.

<표 9> 서울지역 호텔에 대한 LQI 차원별 감성분석 결과 (2012년-2013년)

2012년 - 2013년		A 호텔	B 호텔	C 호텔	D 호텔
LQI 차원					
유형성 (Tangibility)	(visually)appealing				
	clean	0.376003	0.390784	0.311735	0.331606
	neat				
	inviting				
	pleasant	0.381079	0.405047		
	attractive				
	bright				
	well lighted				
신뢰성 (Reliability)	well maintained				
	efficiently	0.217091			0.319883
	promised		0.362558	0.208703	
대응성 (Responsiveness)	properly				
	prompt		0.144407	0.071941	
	promptly		0.144263	0.125702	
	informative				
믿음성 (Confidence)	willing to				
	know about				
	respect	0.235677		0.400001	
	polite		0.188888	0.192122	0.234692
	safe	0.237704	0.334597	0.239274	
의사소통 (Communication)	conveniently				
	clearly				
	attention				
	particular		0.496273	0.406989	0.472769
	anticipated				

<표 9>는 서울지역 호텔들에 대해 2012년부터 2013년 동안 작성된 후기들에 대한 감성분석 결과이다. 2012년-2013년 동안의 A호텔과 D호텔의 “clean”에 대한 감성점수가 이전 기간에 비해 많이 향상되었고, 나머지 호텔 2곳은 비슷한 수준을 유지하고 있음을 알 수 있었다. B호텔의 신뢰성 차원의 “promised”가 이전 기간보다 많이 향상되었다. 직원의 친절성과 예절성 또한 4곳 모두 변화가 심했는데 A호텔의 친절성은 낮아진 반면 C호텔의 친절성은 많이 개선되었다. B호텔, C호텔, D호텔의 예절성은 다소 낮아진 것으로 나타났다.

<표 10>은 2014년부터 2015년 사이의 보다

최신 후기들을 분석한 결과이다. “clean”에 대하여 B호텔의 점수가 많이 높아졌고 전반적으로 호텔 4곳의 청결도는 높게 평가되었다. “pleasant”의 경우 A호텔, B호텔, C호텔 모두 긍정적으로 평가되었으며, 신뢰성 차원의 “promised”는 B호텔과 C호텔의 경우 분석 기간 동안 전반적으로 향상된 모습을 보였다. 2014년부터 2015년 동안에는 A호텔, B호텔, C호텔들 모두 고객들로부터 서비스의 신속성 (“prompt”)에 대하여 높은 긍정의 점수를 받았다. 또, A호텔의 예절성을 제외한 모든 호텔들에서 직원의 친절성과 예절성이 상당히 긍정적인 평가를 받았음을 확인할 수 있었다.

<표 10> 서울지역 호텔에 대한 LQI 차원별 감성분석 결과 (2014년-2015년)

2014년 - 2015년		A 호텔	B 호텔	C 호텔	D 호텔
LQI 차원					
유형성 (Tangibility)	(visually)appealing				
	clean	0.329487	0.479495	0.259279	0.386147
	neat		0.132029		
	inviting				
	pleasant	0.307664	0.248736	0.254032	
	attractive				
	bright				
	well lighted well maintained				
신뢰성 (Reliability)	efficiently	0.312595			0.448426
	promised		0.350671	0.378145	
	properly				
대응성 (Responsiveness)	prompt	0.36444	0.370558	0.371789	
	promptly				
	informative				
	willing to				
믿음성 (Confidence)	know about				
	respect	0.445101	0.325342	0.366298	
	polite	0.150099	0.484688	0.377952	0.465031
	safe conveniently				
의사소통 (Communication)	clearly		0.401926	0.414047	
	attention				
	particular				
	anticipated				

## V. 결론

본 연구에서 LQI 차원별로 핵심 감성어를 선정하고 Neural Language Network 기반의 word2vec 알고리즘을 이용하여 온라인 숙박 후기에 담겨진 고객의 호텔 서비스에 대한 감성을 도출하며 호텔 서비스 품질에 대한 평가를 해보았다. 이를 통해 동일한 지역 (서울 지역) 내 호텔 간의 서비스 품질을 비교해 보았으며 또 작성 시기별로 후기들을 일정 단위(2년)로 나누어 분석함으로써 호텔별 서비스 품질이 시

간의 흐름에 따라 어떻게 변화하였는지 살펴보았다.

여행객의 대다수가 방문하는 서울지역의 4개 호텔들에 대해 감성분석을 수행한 결과를 보면 서비스에 대해 불만을 표현하는 여행객들의 비율은 낮았던 것으로 보인다. 특히 객실 이용 후 국내 계열 호텔들과 외국 계열 호텔들의 청결도에 대한 여행객들의 감성점수는 상당히 높았으며, 직원의 서비스 제공과 관련한 친절성과 예절성, 여행객들의 요구사항에 대한 직원들의 해결 능력은 충분히 긍정적인 평가를 받았

다. 하지만 서비스의 신속성이나 효율성, 여행객과 직원 사이의 명확한 정보의 전달과 관련한 평가는 인식했던 것으로 드러났다. 이는 분명 국내 호텔들이 앞으로 많은 여행객들을 확보하기 위해 개선해야할 사항일 것이다.

본 연구는 호텔산업의 서비스 품질을 측정하는데 특화된 LQI를 활용하였고, 각각의 LQI 차원을 측정하는 핵심 감성어들과 관련성이 높은 단어들을 word2vec 알고리즘으로 도출하여 SentiWordNet 사전에서 이들 어휘의 감성점수를 읽은 후, 각 LQI 차원들에 대한 감성점수를 계산하였다. Word2vec은 Neural Language Network으로서 단어 간의 semantic 관계를 바탕으로 학습이 이루어진다. 따라서 전체 문서의 크기와 문서를 구성하는 단어들의 배열과 빈도에 따라 결과에 차이가 나타난다. 관련 어휘들의 cosine 유사도는 어떤 문서를 학습하느냐에 따라 그 결과 값이 달라질 수 있으며 문서의 수가 작은 경우 결과 값을 얻기 힘들 수도 있다. 더욱이 작은 양의 후기를 2년 단위로 나누어 분석할 경우 결과 값에 일부 차이가 존재할 수 있다.

어휘별로 하나의 감성점수를 부여하고자 SentiWordNet의 해당 어휘의 감성점수를 평균하였는데, 단어의 문법적 사용 용도에 따라 동일한 단어라도 서로 다른 감성점수를 가질 수 있어 이를 일괄 평균한 것은 본 연구의 한계라고 할 수 있다. 그럼에도 불구하고, 호텔 경영자들에게 고객들의 주요 관심사와 중점 평가 사항에 대한 식견을 제공하는 결과를 도출한 것은 만족할 만하다. 공통적으로 청결성, 친절성, 예절성은 호텔의 기본적인 서비스 품질이기에 고객들의 기대는 당연하고 다양한 후기에서

“clean”, “respect”, “polite”이라는 단어와 이들과 관련한 여러 유의어 파악이 중요하고, 이를 통해 고객의 시각에서 바라보는 호텔 평가 척도에 관한 통찰을 얻을 수 있게 된 것은 본 연구를 통해 얻은 수확이다.

본 연구가 영문 후기만을 분석하였기에 향후 한글 후기와 영문 후기들을 모두 포함한 비교 분석으로 확장해 보는 것은 호텔의 서비스 품질에 대한 내외국인들의 시각 차이를 파악하고 보다 총체적이고 효과적인 서비스 품질 측정으로 확장하는 계기가 될 것으로 사료된다.

## 참고문헌

- 구태희, “온라인 구매후기 작성동기가 패션제품 재구매의도 및 추천 의도에 미치는 영향,” 한국의류산업학회지, 제12권, 2호, 2001, pp. 188-193.
- 김보영, 호이결, 송니은, “제품 속성별 온라인 구전의 방향성이 중국 소비자의 구매의도에 미치는 영향,” 인터넷전자상거래연구, 제16권, 제3호, 2016, pp. 123-143.
- 김진화, 변현수, 이승훈, “온라인 리뷰를 활용한 사용자 이해 및 서비스 가치증대,” 정보시스템연구, 제20권, 제2호, 2011, pp. 21-36.
- 박은아, “온라인 사용후기가 구매의사결정에 미치는 효과: 후기방향성, 브랜드인지도, 제품유형에 따른 차이,” 광고학연구, 제18권, 5호, 2007, pp. 7-22.
- 성영신, 유형열, 장인숙, “Word of Mouth: On-line 상의 소비 구전정보 연구,” 한



- 국광고학회 연차학술대회, 2001, pp. 3-12.
- 양흔, 최상민, 문태수, “한국 의료관광 서비스품질과 이용의도에 관한 실증연구,” 인터넷전자상거래연구, 제13권 제3호, 2013, pp. 185-205.
- 오걸, “노트북 구매과정에서 온라인 사용후기가 소비자의 구매의도에 미치는 영향,” 경희대학교 석사학위논문, 2015.
- 이정현, "온라인 호텔이용후기의 질적 내용분석에 의한 고객가치 연구," 한국콘텐츠학회논문지, 제13권, 10호, 2013, pp.533-546.
- 장은진, 김정균, “후기게시판 신뢰 요인 연구: 온라인 쇼핑몰 후기게시판을 중심으로,” 정보시스템연구, 제20권, 제4호, 2011, pp. 233-254.
- 조승연, “고객 온라인 구매후기를 활용한 추천 시스템 개발 및 적용,” 한국경영정보학회, 2015.
- 홍동문, “중국의 WTO 가입과 한·중 서비스산업구조의 비교분석,” 한국지역발전학회, 지역발전연구, 제3권, 2호, 2003 pp. 375-404.
- 홍중의, “공공서비스 평가를 위한 다측면 프레임워크 개발,” 한국산업정보학회지, 제18권, 제6호, 2013, pp. 71-82.
- Mei W., “한·중 서비스산업의 국제경쟁력 비교 분석,” 부경대학교 석사학위논문, 2014.
- Akbaba, A., "Measuring service quality in the hotel industry: A study in a business hotel in Turkey," *International Journal of Hospitality Management*, Vol. 25, No. 2, 2006, pp. 170-192.
- Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Goot, E. v. d., Halkia, M., et al. , “Sentiment analysis in the news,” *International Language Resources and Evaluation*, 2010.
- Berry, L.L., "Services Marketing Is Different," *Business*, Vol. 30, 1980, pp. 24-28.
- Brister, J., “Word of mouth communication and their effects in consumer network,” *Advances in Consumer Research*, Vol. 18, 2004, pp. 155-169.
- Carman, J.M., “Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions,” *Journal of Retailing*, Vol. 66, No. 1, 1990, pp. 33-55.
- Cronin, Joseph, J., Jr. ,Taylor, S. , “Measuring Service Quality: A examination and Extension,” *Journal of Marketing*, Vol. 56 No. 4., 1992, 55-68.
- Crotts, J., Peyton R., and Boyd D., “Measuring Guest Satisfaction and Competitive Position in the Hospitality and Tourism Industry,” *Journal of Travel Research*, Vol. 48, No. 2, 2009, pp. 139-51.
- Ding, X., and Liu, B., “The utility of linguistic rules in opinion mining”, *SIGIR*, 2007
- Esuli, A., Sebastiani, F., “SentiWordNet: A publicly available lexical resource for opinion mining,” *Language Resources and Evaluation(LREC)*, 2006.
- Feldman, R., ”Techniques and Applications for Sentiment Analysis,” *Communications of the ACM*, Vol. 56, No. 4, 2013, pp. 82-89.
- Getty, J. and Getty, R., “Lodging quality index

- (LQI): assessing customers' perceptions of quality delivery," *International Journal of Contemporary Hospitality Management*, Vol. 15, No. 2, 2003, pp. 94-104.
- Hu, M., and Liu, B., "Mining and summarizing customer reviews," *International Conference on Knowledge Discovery and Data Mining*, 2004.
- Huang, C.H., "An application of LODGSERV to the comparison of leisure travelers' and hoteliers' service quality perspectives," 2011.
- Keith, K., Simmers, S. , "Measuring hotel service quality perceptions: the disparity between comment cards and lodgserv," *Academy of Marketing Studies Journal*, Vol. 17, No. 2, 2013, pp 119-131.
- Knutson, B., Stevens, P., Wullaert, C., Patton, M. and Yokoyama, R., "LODGSERV: a service quality index for the lodging industry," *Hospitality Research Journal*, Vol. 15, 1990, pp. 277-84.
- Kotler, P., Cox, K. K., "*Marketing management and strategy (4th Ed.)*," New Jersey-Englewood Cliffs: Prentice-Hall., 1988.
- Ladhari, R., "A review of twenty years of SERVQUAL research," *International Journal of Quality and Services Sciences*, Vol. 1, No. 2, 2009, pp. 172 - 198.
- Ladhari, R. , "The Lodging Quality Index: An Independent Assessment of Validity and Dimensions," *International Journal of Contemporary Hospitality Management*, Vol. 24, 2012, en prensa.
- Litvin, S.W., Goldsmith, R.E., Pan, B., "Electronic word-of-mouth in hospitality and tourism management," *Tourism Management*, Vol. 29, No.3, 2008, pp. 458 - 468.
- Liu, B., "Sentiment analysis and subjectivity," *Handbook of Natural Language Processing*, N. Indurkha and F.J. Damerau, eds. 2010.
- Liu, B. , "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, Vol. 5, No. 1, 2012, pp. 1-167
- Liu, Y. , Yu, X., An, A., and Huang X., "Riding the tide of sentiment change: sentiment analysis with evolving online reviews," *World Wide Web*, 2013, pp. 1-20.
- Manning, C. D., Schütze, H, "*Foundations of statistical natural language processing*," Cambridge Massachusetts: MIT Press., 1999.
- Mathiak, B., Eckstein,S., "*Five steps to text mining in biomedical literature*," Data mining and text mining for bioinformatics european workshop, 2004.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J., "*Efficient estimation of word representations in vector space*," ICLR Workshop, 2013.
- Nadiri, H., Hussain, K., "Perceptions of service quality in North Cyprus hotels,"

- International Journal of Contemporary Hospitality Management*, Vol. 17, No. 4, 2005, pp. 469-80.
- Nasukawa, T., Yi, J., "Sentiment analysis: Capturing favorability using natural language processing," *Proceedings of the 2nd international conference on Knowledge capture*, ACM, 2003.
- O'Connor, P., "Managing a hotel's image on TripAdvisor," *Journal of Hospitality Marketing & Management*, Vol. 19, No. 7, 2010, pp. 754 - 772.
- O'Connor, P., Frew, A.J., "An evaluation methodology for hotel electronic channels of distribution," *International Journal of Hospitality Management*, Vol. 23, No. 2, 2004, pp. 179 - 199.
- Ortigosa, A., Martin, J.M., Carro, R.M., "Sentiment analysis in Facebook and its application to e-learning," *Computers in Human Behavior*, Vol. 31, 2014, pp. 527-541.
- Pang, B., Lee, L., and Vaithyanathan, S., "Thumbs up? Sentiment classification using machine learning techniques," *Empirical Methods in Natural Language Processing(EMNLP)*, 2002, pp. 79 - 86.
- Pang, B., Lee, L., "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, Vol. 2, No. 1-2, 2008, pp. 1135.
- Parasuraman, A., Zeithaml, V.A., and Berry, L.L., "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing*, Vol. 49, No. 3, 1985, pp. 41-50.
- Parasuraman, A., Zeithaml, V.A., and Berry, L.L., "SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality," *Journal of Retailing*, Vol. 64, 1988, pp.12 - 40.
- Rathmell, J.M., "What Is Meant by Services?," *Journal of Marketing*, Vol. 30, No. 5, 1966, pp. 32-6.
- Saleh, F. and Ryan, C., "Analysing service quality in the hospitality industry using the SERVQUAL model," *Services Industries Journal*, Vol. 11 No. 3, 1992, pp. 324-43.
- Said, A., Shuib, A., Ayob, N., Yaakub, N., "An evaluation of service quality from visitors' perspectives: The case of Niah National Park in Sarawak," *International Journal of Business and Society*, Vol. 14 No. 1, 2013, pp. 61 - 78.
- Smith, A., "An inquiry into the Nature and Causes of the wealth of nations," Clarendon Press, Oxford, 1976, pp. 330~349.
- Turney, P., "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," *Proc. of the ACL*, 2002.
- Wang, H., Zhang, D., Zhai, C., "Structural topic model for latent topical structure analysis," *ACL*, 2011, pp. 1526 - 1535.
- Wiebe, J., Wilson, T., Bruce, R., Bell, M., Martin, M., "Learning subjective

language,” *Computational Linguistics*, Vol. 30, No. 3, 2004, pp. 277 - 308.

Xiang Z, Schwartz Z, Gerdes J H, “What can big data and text analytics tell us about hotel guest experience and satisfaction,” *International Journal of Hospitality Management*, Vol. 44, 2015, pp. 120-130.

Yu, H., Hatzivassiloglou, V. “Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences,” *Empirical Methods in Natural Language Processing*, 2003, pp. 129 - 136.

Wilson, T., Wiebe, J., and Hoffmann, P., “Recognizing contextual polarity in phrase-level sentiment analysis,” *Human Language Technology and Empirical Methods in Natural Language Processing*, 2005.

Wiebe, J., Wilson, T., Cardie. C., “Annotating expressions of opinions and emotions in language,” *Language Resources and Evaluation*, Vol. 39, 2005, pp. 165-210.

#### 사공 원 (Sakong, Won)



영남대학교 수학과를 졸업하였으며, 경북대학교에서 경영학 석사학위를 취득하였다. 연구 관심분야는 Neural Net Language Model 과 스포츠 데이터 분석이다.

#### 하 성 호 (Ha, Sung Ho)



한국과학기술원에서 박사 학위를 취득하고 현재 경북대학교 경영학부 교수로 재직 중이다. 국내외 저널의 편집위원을 맡고 있으며 e-비즈니스, 데이터마케팅을 주로 연구하고 있다.

#### 박 경 배 (Park, KyungBae)



미국 University of Texas at Dallas에서 석사, Virginia Tech. 및 경북대학교 경영학부에서 박사수로 하였다. 주요 연구 분야는 텍스트마케팅, 토픽모델링 등이다.

<Abstract>

## **Hotel Service Quality Evaluation Based on LQI using Sentiment Analysis of Online Reviews**

Sakong, Won · Ha, Sung Ho · Park, KyungBae

### **Purpose**

With the increasing number of foreign travelers visiting Korea, it is a heavy question to evaluate service quality of typical domestic hotel companies. Our research aims to evaluate service quality of domestic hotels in Korea from the perspective of foreign travelers in order to provide the quality improvements that call attention for the hotel management.

### **Design/Methodology/Approach**

In this paper, topics of sentiment followed Lodging Quality Index(LQI) dimensions classifying lodging service quality appropriately. Also, we employed word2vec algorithm which calculates similarity and affinity among the vocabularies accurately. To calculate sentiment of each dimension, we adopted scores from SentiWordNet.

### **Findings**

From the result, we found the number of foreign travelers particularly satisfied with cleanliness, politeness, and problem solving skills. In contrast, it has also been found out that both promptness of services and efficiency of communication do not fulfill the requirements of travelers.

**Keywords:** LQI, online review, sentiment analysis, service quality, word2vec

\* 이 논문은 2016년 8월 19일 접수, 2016년 9월 5일 1차 심사, 2016년 9월 24일 게재 확정되었습니다.