

Airbnb 숙소 유형에 따른 호스트의 자기소개 텍스트가 공유성과에 미치는 영향¹⁾

Impact of Self-Presentation Text of Airbnb Hosts on Listing Performance by Facility Type

심지환 (Sim, Ji Hwan)	국민대학교 ²⁾
김소영 (Kim, So Young)	국민대학교 ³⁾
정여진 (Chung, Yeojin)	국민대학교 ⁴⁾

〈 국문초록 〉

최근 빠르게 성장하고 있는 숙박 공유경제 시장에서 품질에 대한 불확실성은 사용자의 만족도에 영향을 미치는 위험 요소지만, 이는 시설 제공자가 공개하는 정보를 통해 완화될 수 있다. 그 중 시설 제공자의 본인에 대한 자기소개는 사용자와의 정서적 교류를 통해 심리적 거리를 제거함으로써 공유 성과에 긍정적인 영향을 미친다. 본 연구는 대표적인 숙박 공유경제 플랫폼인 Airbnb에서 호스트의 자기소개가 포함하는 정보의 종류에 따라 공유성과에 미치는 영향을 분석하고, Airbnb의 숙소 유형에 따라 차이를 분석하였다. 이를 위해 호스트가 공개하는 자기소개 텍스트를 문장별로 분리하고 비지도 학습기반의 딥러닝 방법인 Attention-Based Aspect Extraction 방법을 활용하여 각 문장이 포함하는 의미를 추출하였다. 추출된 의미를 토대로 자기소개 텍스트가 포함하는 의미가 공유성과에 미치는 영향과 숙소 유형에 따른 교호작용 효과를 분석하였다. 연구결과, 숙소 유형별로 호스트의 특정 성향이 공유성과에 긍정적인 영향을 미치는 것을 확인하였고, 이를 통해 숙소 유형에 따라 공유성과를 극대화하기 위한 마케팅 전략에 대한 실증적인 함의를 제공한다.

주제어: 공유경제, 자기소개, 특성추출, 텍스트마이닝, 비지도학습

1) 이 논문은 한국연구재단 연구비(NRF-2016R1C1B1010940)와 산림청(한국임업진흥원) 산림과학기술 연구개발사업(2019150B10-1923-0301)에 의해 수행되었다

2) 제1저자, sim2080@gmail.com

3) 제2저자, thdudrnrwk1@naver.com

4) 교신저자, Kookmin University

1. 서론

공유경제시장은 다양한 상품과 서비스를 개인에게 소유권을 이전하는 대신 임시로 공유하는 방식으로 거래하는 시장으로, 2025년까지 연간 약 5,700억 유로의 거래가치가 예상되는 거대한 시장이다. 공유경제의 광범위하고 빠른 성장은 peer-to-peer(P2P) 시장이 소셜 네트워크 플랫폼의 발전을 주도하면서 이루어졌다(Botsman and Rogers, 2011). 특히 여행 및 숙박 산업에서의 P2P 거래 플랫폼은 개인 판매자가 그들 소유의 시설을 시장에 제공할 수 있게 해줌으로써 시장을 크게 변화시켰다.

Airbnb는 대표적인 P2P 숙박 공유 플랫폼으로써 호스트와 게스트가 직접적으로 상호작용하면서 숙박 시설을 대여, 사용할 수 있도록 연결하는 서비스를 제공한다. 온라인 상에서 소개되는 서비스는 눈에 보이지 않고, 특히 숙박 품질은 직접 경험해보기 전에 검증이 어렵기 때문에(Wilson et al. 2012), 이러한 유형의 사회적 교환은 게스트들의 기대와 경험 사이의 괴리가 발생될 위험을 내포하고 있다. 따라서 게스트는 그들이 이용 가능한 최대한의 정보를 통해 숙박 시설에 대한 품질을 파악하여 이러한 불확실성을 해소하고자 한다. 전통적인 e-commerce와 마찬가지로 Airbnb 상의 온라인 평판이나 시설 사진, 평점, 후기 등은 호스트와 숙박시설을 판단하기 위해 게스트가 활용할 수 있는 중요한 정보이다. 이와 함께 호스트의 자기소개(self-presentation)는 해당 시설이 가지고 있는 잠재적 정보를 이해하는데 도움을 주고, 불확실성을 낮추는데 유용하다(Ellison, Heino and Gibbs 2006).

호스트가 공개하는 자기소개는 정보 제공의 역할 뿐 아니라 호스트와 게스트 사이의 정서적 교류를 통해 심리적 거리를 줄이는 역할을 한다. 호스트가 자기소개를 통해 전달하는 주제는 호스트와 게스트 사이

의 신뢰 구축에 영향을 미친다(Zhang et al. 2020). 특히 환영인사, 친근함의 표현, 서비스와 같은 사회적 표현이 많을수록 실제 호스트의 수입에 긍정적인 영향을 미치는 것으로 알려져 있다(Garcia, 2019). 전통적인 e-commerce에서 제품 정보를 통해 전해지는 정서적 교류가 구매에 미치는 영향은 제품의 유형에 따라 상이한 것으로 알려져 있다(Hassanein and Head, 2005). 기술적 제품에 가까울수록 객관적인 제품 정보 이외의 표현은 소비자의 제품평가에 부정적인 영향을 미치는 반면, 유희적 측면이 추구되는 제품은 사회적 현존감(social presence)을 유발하는 표현이 많을수록 긍정적인 반응을 보이는 것으로 밝혀졌다.

Airbnb에서는 숙소 유형이 제품 유형에 해당되는데 이는 집 전체(Entire home), 개인방(Private room), 공유방(Shared room)으로 나누어진다. 이러한 숙소 유형에 따라 게스트의 방문 목적, 성향, 인구통계학적 특성이 상이하기 때문에(Lutz and Newlands, 2018), 호스트와의 정서적 교류와 공유성과와의 관계 역시 숙소 유형에 따라 달라질 가능성이 있다. 따라서 본 연구에서는 숙소 유형에 따라 호스트의 자기소개가 포함하는 내용이 공유성과에 미치는 영향을 분석한다.

구체적으로, Airbnb에 공개되는 호스트의 자기소개 텍스트를 문장별로 나누고, 딥러닝 기반의 특성추출 방법인 Attention-based Aspect Extraction(ABAE; He et al., 2017)을 적용하여 텍스트가 포함하는 특성을 문장별로 추출한다. 이를 통해 추출된 다섯 가지 특성이 공유성과에 미치는 영향을 회귀모형을 통해 검증하고 숙소 유형에 따른 차이를 교호작용 효과를 분석함으로써 탐구한다.

본 연구를 통해 Airbnb의 숙소 소개 웹페이지에서 호스트의 자기소개가 포함하는 내용 중 효과적으로 공유 성과를 높이는 특성을 파악할 수 있다. 또한 숙소 유형에 따라 어떠한 내용이 보다 효과적인지를 비

교분석 함으로써 각 호스트가 공유성과를 극대화 하기 위해서 공개해야 하는 정보의 종류를 판단할 수 있다. 이는 공유경제 시장에서 시설제공자의 마케팅 전략을 수립하는데 중요한 정보를 제공한다.

본 논문의 구성은 아래와 같다. 2장에서는 문헌연구들을 살펴보고, 3장에서 데이터 수집과 전처리, ABAE 모델을 통한 특성 추출 및 회귀모형에 대해 설명한다. 4장에서는 분석 결과를 제시한 뒤 5장에서 결과에 대한 해석 및 본 연구의 의의에 대해 논의한다.

2. 문헌연구

2.1. 공유경제에서의 자기표현의 중요성

공유경제는 활용도가 낮은 자원을 공유하여 소유권이전 없이도 사용자가 해당 자원을 통해 효용을 얻을 수 있는 경제양식을 의미한다(Wilson et al. 2012). 공유경제에서의 거래되는 아이템은 전통적인 e-commerce와는 달리 단기간의 대여서비스, 라이드헤일링(Ride-hailing)과 같은 일시적인 서비스 등이 주를 이룬다. 공유경제에서 나타나는 단기적, 일시적인 P2P 거래는 소비자에게 서비스에 대한 품질의 불확실성을 초래할 수 있다. P2P 거래는 대면 계약이 불가피하기 때문에 제공되는 서비스의 품질이 거래 상대방에 따라 상당한 차이가 있을 수 있고, 항상 서비스 품질이 보장될 수 없다. 또한 P2P 거래에서는 품질을 논할 때 가격적 이점, 제공되는 서비스의 품질과 같이 소비자가 단순하게 느낄 수 있는 경제적 이점 이외에도 거래 상대방과의 심리적 거리의 원근도 평가대상이 될 수 있다. 이러한 이유로 공유경제는 상호의존성 측면에서 거래 당사자 간의 타협 없이는 거래가 성사되기 어려우며, 이러한 과정에서 발생하는 거래 위험은 피할 수 없는 부분이

다(Zhang et al. 2020). 특히 공유경제에서 주로 거래되는 차량 대여 서비스, 숙박 시설 대여 서비스와 같은 경험재(experience goods)의 경우에는 서비스 품질에 대한 불확실성의 해소가 더욱 중요하다.

제품 및 서비스 품질의 불확실성에 대한 연구는 수년간 마케팅 분야에서 연구되어 왔다. 소비자는 제품을 구매하기 전에 제품 자체를 인지하는 과정을 거친다. 이 과정에서 제품에 대한 설명(product description)은 소비자의 관심 또는 흥미를 이끌어내는 핵심적인 역할을 한다(Mou et al., 2019). 특히, e-commerce와 공유경제에서는 대면 거래 없이 오로지 온라인상의 업로드 되어 있는 상품 이미지 및 설명과 같이 전자정보만을 이용하여 품질을 평가해야 하기 때문에 이러한 정보들은 더욱 중요한 역할을 한다(Giuffrida et al., 2017).

공유경제에서는 서비스를 제공하는 주체 또한 서비스 품질평가의 대상이 될 수 있는데, 이 때 판매 주체의 자기표현(self-presentation)은 이용자와의 정서적 교류에 중요한 역할을 하고 있으며, 이는 서비스 품질에 영향을 미칠 수 있다. Goffman(1959)은 자기 자신에 대한 지속적인 전략적 표현으로 자기표현을 설명한다. 이 자기표현은 자아상을 타인에게 전달함에 있어 사회적 상호작용의 중요한 요소임을 강조하고 있다. 개인 프로필은 온라인 상에서 자신을 제시하는 핵심적인 정보를 제공하는 역할을 한다.(Schau and Gilly, 2003; Elison, Hancock, and Toma, 2012). 특히 개인 프로필에서 확인할 수 있는 판매자의 사진, 판매자의 자기표현 패턴 등의 개인 프로필 속성이 P2P 거래에서 신뢰와 평판 형성에 영향을 미친다(Ma et al., 2017).

정서적 교류와 유대감이 구매에 미치는 영향은 제품의 종류에 따라 달라진다. Chang et al.(2012)에 의하면 쾌락적(hedonic) 제품일수록 리뷰어의 전문성과 관계없이 사회적 표현을 통한 강한 유대감으로 제품을

보증할 때 구매의도가 강해진다. Hassanein and Head(2005)는 재미 요소가 포함된 제품은 풍부한 사회적 표현과 추가적인 정보가 긍정적 효과를 가지고 오는 반면, 기능적 측면이 중요한 제품에 대해서는 사회적 표현은 요구되는 정보와 부합하지 않기 때문에 제품 설명에 사회적 표현이 들어가는 것이 부적절하다고 설명하고 있다.

Lutz and Newlands(2018)는 Airbnb의 Shared room과 Entire home 이용자들의 인구통계학적 차이를 검증하였는데, Entire home 이용자들은 소득 수준과 교육 수준이 높고, 파트너와 함께 방문하고, 환경적 요인에 불편함을 크게 느끼는 경향이 있다고 설명한다. 반면에 Shared room 이용자는 상대적으로 젊은 연령층이고, 소득 제약이 있으며 주택에 머무는 동안 대인관계에 불편함을 덜 느끼는 경향이 있다. 방문목적 역시 상이한데, Entire home 이용자는 비즈니스 목적으로 해당 지역을 방문하고, Shared room 이용자는 저비용으로 여행의 목적을 가지고 해당 지역을 방문하는 이용자들이 주를 이룬다. Guttentag et al.(2017)는 Airbnb 이용자들의 동기를 군집화 하였는데, 그 중 새로운 상호작용을 추구하는 이용자 군집이 Shared room에서 머물렀을 가능성이 평균보다 훨씬 높았다. 또한 호스트의 주택을 공유함으로써 해당 지역에서의 진정한 경험을 하고, 현지인들과 교류하는 것을 추구한다는 결과를 보고하였다.

2.2. e-commerce 분야에서의 텍스트 관련 연구

마케팅 분야에서는 텍스트 형식으로 된 피드백의 중요성이 점차 증대되고, 이러한 텍스트 정보를 분석하기 위해 다양한 텍스트 마이닝 기법들이 활용되어 왔다(Ordenez et al. 2014). 텍스트 마이닝은 이용자의 리뷰 혹은 제품에 대한 설명과 같은 비구조적인 텍스

트 상에서 일정한 패턴을 감지하여 소비자의 성향 및 선호도와 판매자의 의도 등을 파악하는 자동화된 프로세스를 의미한다(Younis, 2015). 이 텍스트 마이닝은 사람이 직접 처리할 수 없는 방대한 양의 텍스트를 통해 인사이트를 얻는 데에 그 목적이 있다는 점에서 유용성을 인정받고 있다(Gupta and Lehal, 2009).

텍스트 마이닝은 광범위한 산업에 걸쳐 다양하게 활용되고 있는데, 본 연구는 공유경제에서 활용되고 있는 텍스트 마이닝 사례를 살펴본다. Cheng and Jin(2019)은 Airbnb의 리뷰 상에서 핵심 특성과 감정을 분석하여 이용자들이 시설을 선정함에 있어 중요하게 생각하는 요소로서 지역, 시설, 호스트 등이 있음을 파악하였다. Zhang(2019)b은 다양한 감정이 혼재되어 있는 리뷰 중에서도 부정적 평가가 긍정적 평가보다 더 진정성 있고 신뢰할 만한 정보를 제공한다고 설명한다. 리뷰 텍스트 이외에도 Cheng and Foley(2018)는 Airbnb의 차별금지 정책을 보도한 신문기사에 대한 응답을 실은 논평을 분석하여 이해관계자들에게 의미 있는 정보를 제공하였다.

2.3. Aspect extraction

자연어 처리 기법을 사용하여 의미 있는 특성을 추출하기 위한 aspect extraction 방법은 지난 몇 년간 연구되어왔다. Aspect extraction은 텍스트 데이터에서 aspect 단어를 식별하고 유사한 aspect 단어들을 군집화(cluster)한다. 예를 들면 “I am originally from Pittsburgh but I love living in New York.” 문장에서 “originally”는 “grew”, “born” 등의 단어들과 함께 그룹화되어 “출신지역”의 의미를 갖는 aspect를 형성한다. Aspect extraction에 대한 이전 연구들은 지도학습(supervised learning)과 비지도학습(unsupervised learning) 두 가지 접근법으로 분류할 수 있다.

지도학습 접근법에서 aspect extraction은 순차적으로 입력되는 문자 혹은 단어들을 통해 타겟 카테고리를 예측하는 전형적인 시퀀스 라벨링 문제(sequence labeling problem)로 간주된다(Jin & Ho, 2009). 최근에 딥러닝의 발전으로 심층 신경망, 합성곱 신경망, 순환 신경망을 사용한 aspect extraction 방법론이 제안되었다(Poria et al., 2016; Wang et al., 2016). 이러한 딥러닝 기반 지도학습 모델이 aspect extraction에서 좋은 성능을 제공하는 것으로 알려져 있지만 이를 위해서는 aspect가 부여된 방대한 양의 훈련 데이터가 필요하다는 제약이 있다.

비지도학습 기반 aspect extraction은 훈련 데이터에 대한 제약이 없다는 점에서 실무적으로 유용하다. 주로 토픽 모델링 기법이 적용되는데 가장 일반적인 토픽 모델링 기법인 잠재 디리클레 할당(Latent Dirichlet Allocation; LDA) 모델(Blei et al, 2003)과 이를 확장한 다양한 모델들(Mimno & McCallum, 2008; Roberts, Stewart, & Airoldi, 2016)이 사용되고 있다. 단어들의 동시출현 빈도가 토픽 구성에 있어서 가장 중요한 정보의 근원임에도 불구하고 LDA 모델은 이를 직접적으로 반영하지 않고, 대신 문서로부터 발생하는 단어의 패턴을 확률모형에 의해 간접적으로 파악한다(Mimno et al., 2011). 또한 LDA 모델은 각 문서의 토픽의 분포를 추정해야 하기 때문에 문서의 길이가 짧으면 그 성능이 저하되는 경향이 있다(He et al., 2017; Luo et al., 2019). 이러한 LDA 기반 모형의 한계를 He et al.(2017)과 Luo et al.(2019)은 벤치마크 예제를 활용하여 F1 score를 기준으로 보여주고 있다.

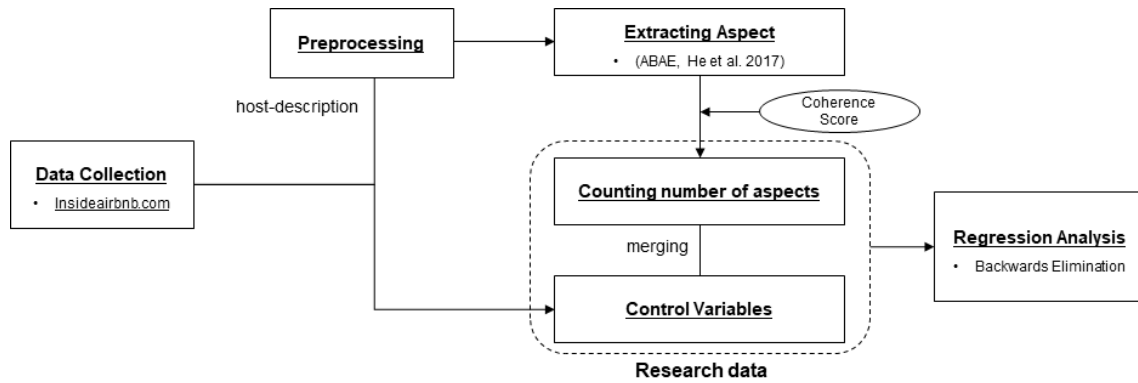
최근 제안된 ABAE 모델(He et al., 2017)은 어텐션 메커니즘(attention mechanism)을 적용한 비지도학습 기반 aspect extraction 방법이다. ABAE 모델은 LDA와 달리 워드임베딩(word embedding)을 사용하여 동시출현단어(word co-occurrence) 분포가 반영된 벡터로 변

환된 단어를 입력값으로 받는다. 어텐션 메커니즘을 활용하여 학습을 하는 동안 aspect 구성과 관련성이 높은 단어만 강조함으로써 효율적으로 aspect 단어를 추출한다. 모델을 통해 각 aspect에 해당하는 벡터가 학습이 되면, 코사인 유사도 행렬(Cosine similarity matrix)을 사용하여 임베딩 공간에서 해당 aspect 벡터와 가장 가까운 단어들을 추출하여 aspect 단어로 사용한다. 확률적 모형으로 토픽을 추출하는 LDA 모델과 달리 ABAE 모델은 워드임베딩을 통한 단어 벡터와 동일한 공간에서 직접적으로 aspect 벡터를 학습시키기 때문에 직접적으로 단어의 동시출현 빈도를 반영하는 aspect를 추출하기 때문에 좋은 성능을 가지는 것으로 알려져 있다. 또한 텍스트의 길이가 짧을 때 LDA가 성공적으로 토픽을 추출하지 못하는 한계를 보완해 준다.

앞서 소개한 문헌 연구를 바탕으로 본 연구는 텍스트 마이닝 기법을 활용하여 host description으로부터 aspect를 추출하여 숙소 유형별로 공유 성과에 미치는 영향을 탐구하고자 한다. Host description은 호스트가 본인의 시설을 광고하기 위한 주요한 텍스트 정보 중 하나이기 때문에, 일반적으로 토픽 추출의 대상이 되는 텍스트에 비해 다양한 내용을 짧은 길이의 글에 담고 있다. 따라서 본 연구에서는 host description의 내용을 문장 단위로 분할한 뒤 각 문장이 담고 있는 의미를 추출하여 그 영향을 분석한다. 이를 위해 짧은 텍스트에서 비교적 특성을 잘 추출한다고 평가되는 ABAE 모델을 사용한다.

3. 연구방법론

본 연구는 데이터 수집, 전처리, aspect 추출, 각 호스트에 대한 추출된 aspect의 요약, 회귀모형을 활용한



〈Figure 1〉 Research Model

aspect의 영향 분석의 과정을 통해 진행한다. 이 일련의 과정은 <Figure 1>에 도식화 되어 있다.

3.1. 데이터 수집 및 전처리

본 연구를 위해서 insideairbnb.com 웹사이트를 통해 2008년 8월부터 2020년 5월까지 미국 뉴욕에 위치한 Airbnb 숙소 리스팅 데이터를 수집하였다. 데이터가 포함하고 있는 host description 변수에는 해당 시설의 호스트에 대한 설명이 문자로 기록되어 있고, 총 50,246개의 리스트를 본 연구의 원본 데이터(raw data)로 활용하였다.

우선 host description에 결측값이 존재하는 리스트를 제거하고 각 host description을 문장 단위로 분리하였다. 영어가 아닌 단어와 문장부호(예를 들면 ‘.’, ‘;’, ‘!’ 등)와 자연어를 분석하는데 의미가 없는 단어인 불용어(예를 들면 ‘is’, ‘if’, ‘not’ 등)를 제거하여 유의미한 토큰을 구성한 뒤 단어들을 기본 사전 단어인 표제어로 추출(예를 들면 ‘has’, ‘had’ -> ‘have’)하였다. 기업형 호스트들을 데이터에서 제외하기 위해 위의 전처리 과정을 진행한 뒤 동일한 호스트가 게시한 시설이 2개 이상이면 제거하였다. 그 결과 데이터에 포함된 총 호스트 ID는 17,598개이고 Host description의 문

장은 총 53,560개 이다. 하나의 Host description은 평균적으로 3개의 문장으로 이루어져 있다. 텍스트 전처리는 파이썬 nltk(Bird et al., 2009) 패키지를 사용하여 진행되었다.

3.2. 텍스트 데이터의 Aspect 추출

Host description의 각 문장이 시설의 어떤 측면을 설명하고 있는지 추출하기 위해 ABAE 기법을 적용하였다. ABAE 모델은 입력 데이터로 문장을 이루는 각 단어의 임베딩 벡터 값을 받는다. 이를 계산하기 위해 파이썬 gensim 패키지를 사용하여 skip-gram(Mikolov et al., 2013) 알고리즘으로 200차원의 단어 벡터를 학습시켰다. ABAE 모델은 사전에 aspect 개수를 정하는 것이 필요하기 때문에 3에서 15까지 aspects의 개수에 대해 모형을 적합시켰다. 각 ABAE 모델은 15 epochs와 32 batch-size, 학습률이 0.001인 Adam optimizer를 사용하여 학습되었다.

최적의 aspect 개수는 topic coherence score(Mimno et al., 2011)를 기준으로 선택하였다. Topic coherence score를 통한 모델 선택 결과는 사람이 직접 판단하는 결과와 높은 상관관계를 가진다고 알려져 있고 topic model의 질을 판단하는데 표준적으로 사용되는 지표

이다(Roberts et al., 2016; Röder, Both, & Hinneburg, 2015). 주어진 z 개의 aspect에서 상위 N 개의 단어를 가지는 집합을 $S^z = \{w_1^z, \dots, w_N^z\}$ 라고 할 때 coherence score는 아래의 공식에 의해서 계산된다.

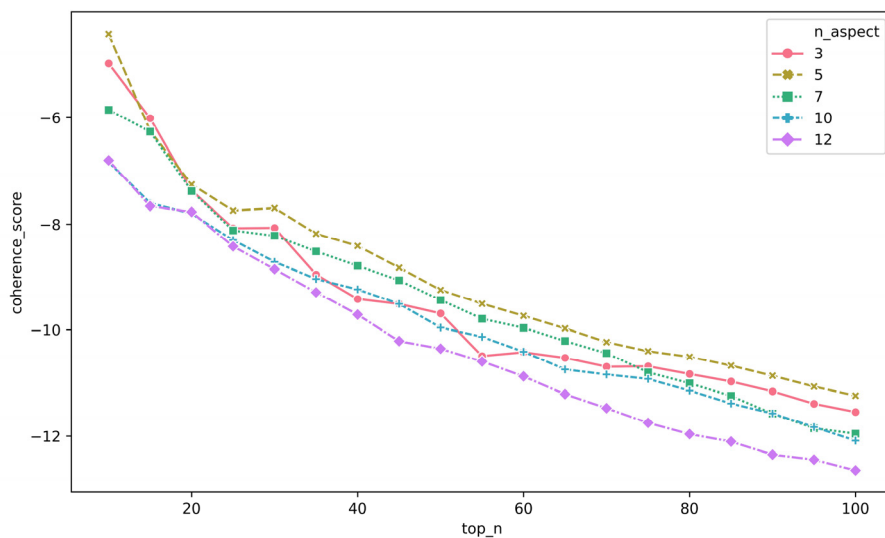
$$c(z; S^z) = \sum_{n=2}^N \sum_{l=1}^{n-1} \log \frac{D(w_n^z, w_l^z) + 1}{D(w_l^z)} \quad (1.1)$$

여기서 $D(w)$ 는 문서 내에서 단어 w 의 빈도수이고 $D(w_1, w_2)$ 는 하나의 문서에 두 개의 단어 w_1 과 w_2 가 동시에 발생한 빈도수를 나타낸다.

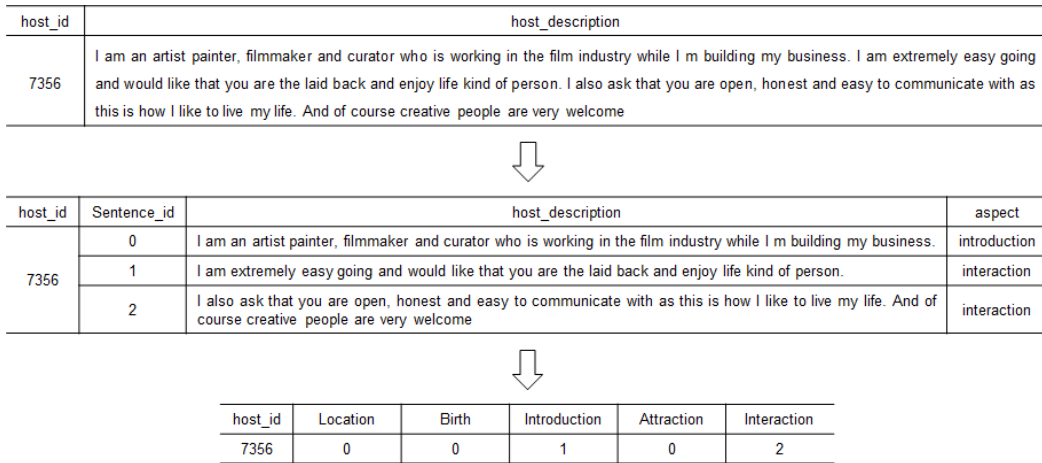
<Figure 2>는 각 aspects의 수에서 상위 단어의 개수($N = 10, 15, \dots, 95, 100$)의 변화에 따른 coherence score를 나타낸 그래프이다. 상위 단어의 개수가 15인 경우를 제외하면 aspect의 수가 5개일 때 가장 높은 일관성 점수를 나타냈다. 따라서 본 논문에서는 5개의 aspect를 사용하여 이후의 분석을 진행한다. 각 aspect에 해당하는 단어는 <Table 1>에 포함되어 있고 단어를 통해 추론한 각 aspect는 Location, Birth, Introduction, Attraction, Interaction으로 명명하였다.

3.3. 회귀모형

본 연구는 호스트 자기소개의 다양한 aspects가 호스트의 마케팅 전략에 미치는 영향을 추정하기 위해서 선형회귀분석 모델을 사용한다. <Figure 3>는 각 host description을 aspect 변수로 변환하는 과정을 설명한다. 개별 host description에서 각 aspect에 해당하는 문장의 개수를 계산하여 다섯 개의 aspect 변수(location, birth, introduction, attraction, interaction)를 생성하여 설명변수로 사용하였다. 숙박 유형(Entire home, Private room, Shared room) 간에 공유 성과를 비교하기 위해서 모델에 두 개의 더미변수($D_{Private Room}, D_{Shared Room}$)를 포함시켰다. 공유 성과는 대여 횟수로 측정하는 것이 이상적이기는 하지만 데이터 수집의 한계를 고려하여 월평균 리뷰의 수를 대리변수(proxy variable)로 사용하여 종속변수로 사용하였다. 리뷰의 수는 숙박 시설의 대여 횟수를 대체하는 성과 측정지표로 기존 문헌들에서 사용되어 왔다(Ye et al., 2011; Zhang, 2019a). 월평균 리뷰의 수가 왼쪽으로 치우친 분포를 나타내기 때문에 모델의 성능 향상을 위해 log변환을 하였다. 통



<Figure 2> Coherence score



〈Figure 3〉 Counting aspects sample

제변수로서 호스트의 응답시간(1시간 이내=1, 몇 시간 이내=2, 1일 이내=3, 1일 이후=4)(*HostResponseTime*), 슈퍼 호스트인지 아닌지(슈퍼호스트=1, 슈퍼호스트가 아님=0)(*Superhost*), 호스트 등록일 이후의 일수(*HostSince*), 숙박 요금(*logPrice*), 숙박의 취소 정책(유통성 있음=0, 보통=1, 엄격=2)(*Cancellation*), 가독성(*Readability*)을 포함한다. 여기서, 가독성은 텍스트 정보를 이해하는데 영향을 주는 변수로서 Flesch 공식

(Flesch, 1979)을 사용하여 계산한다.

Model 1은 모든 설명변수와 통제변수를 포함하였고 Model 2는 Model 1에 숙박 유형에 대한 두 개의 더미변수와 다섯 개의 aspect 변수 간의 교호작용 항을 추가하였다. Model 3은 후진 제거법(backward elimination)을 사용하여 Model 2의 교호작용 항들 중 유의하지 않은 항들을 제거하였다. 세 가지 모형은 아래와 같이 정의된다.

Model 1

$$y_i = \beta_1 HostResponseTime_i + \beta_2 Superhost_i + \beta_3 HostSince_i + \beta_4 Price_i + \beta_5 Cancellation_i + \beta_6 Readability_i + \beta_7 D_{PrivateRoom,i} + \beta_8 D_{SharedRoom,i} + \beta_9 location_i + \beta_{10} birth_i + \beta_{11} introduction_i + \beta_{12} attraction_i + \beta_{13} interaction_i + \epsilon_i$$

Model 2

$$y_i = \beta_1 HostResponseTime_i + \beta_2 Superhost_i + \beta_3 HostSince_i + \beta_4 Price_i + \beta_5 Cancellation_i + \beta_6 Readability_i + \beta_7 D_{PrivateRoom,i} + \beta_8 D_{SharedRoom,i} + \beta_9 location_i + \beta_{10} birth_i + \beta_{11} introduction_i + \beta_{12} attraction_i + \beta_{13} interaction_i + \beta_{14} D_{PrivateRoom,i} * location_i + \dots + \beta_{23} D_{SharedRoom,i} * interaction_i + \epsilon_i$$

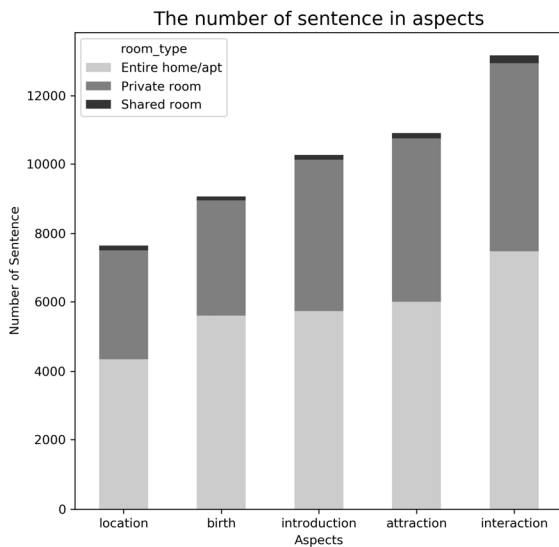
Model 3

$$y_i = \beta_1 HostResponseTime_i + \beta_2 Superhost_i + \beta_3 HostSince_i + \beta_4 Price_i + \beta_5 Cancellation_i + \beta_6 Readability_i + \beta_7 D_{PrivateRoom,i} + \beta_8 D_{SharedRoom,i} + \beta_9 location_i + \beta_{10} birth_i + \beta_{11} introduction_i + \beta_{12} attraction_i + \beta_{13} interaction_i + \beta_{14} D_{PrivateRoom,i} * attraction_i + \beta_{15} D_{SharedRoom,i} * birth_i + \beta_{16} D_{SharedRoom,i} * attraction_i + \beta_{17} D_{SharedRoom,i} * interaction_i + \epsilon_i$$

4. 연구결과

4.1. Aspect 추출 결과

<Table 1>은 ABAE 모델로부터 추출된 5개의 유추된 aspect와 대표 단어들을 나타내고 <Figure 4>는 각 aspects에 해당하는 문장 개수를 숙박 유형과 함께 비교하고 있다. Aspect 변수들의 각 비율을 살펴보면 Interaction(25.78%, 호스트의 상호작용)이 가장 많은 빈도를 나타낸다. 그 다음으로는 Attraction(21.39%, 해당 지역의 문화 및 즐길거리), Introduction(20.10%, 호스트의 직업 및 성별), Birth(17.75%, 호스트의 출생지), Location(14.98%, 해당 지역에 대한 소개 및 교통)의 순서로 이루어져 있다. 숙박 유형별 문장 수를 살펴보면 Shared room은 Entire home과 Private room에 비해서 문장의 개수가 현저히 적다는 것을 알 수 있다. 이는 데이터의 불균형으로 인한 것으로 최종 데이터가 포함하고 있는 호스트의 유형별 수는 Shared room은 240개, Entire home은 10477개, Private room 6881개이다.



<Figure 4> The number of sentence in aspects



<Figure 5> Correlation coefficient between aspects

<Figure 5>는 피어슨 상관 계수를 사용하여 aspect 변수들 사이의 관계를 나타냈다. 대부분의 aspect 간의 양의 상관관계가 존재하고 특히 Location과 Interaction의 상관관계가 크다. Birth는 Introduction과 음의 상관관계를 가지지만 그 크기가 -0.04로 가장 작은 수준이다.

회귀모형의 적합 결과는 <Table 2>에 포함되어 있다. Model 2와 Model 3은 Model 1과 비교했을 때 유의수준 0.05 하에서 통계적으로 유의한 차이가 있기 때문에 숙박 유형에 대한 더미 변수들과 aspect 간의 교호작용이 공유 성과를 설명하는데 의미가 있음을 알 수 있다. Model 2와 Model 3 간에는 유의한 차이가 없는데, Adjusted R^2 는 Model 2가 Model 3보다 약간의 큰 값을 가지지만 베이즈정보기준(Bayes Information Criterion)은 Model 3가 Model 2보다 더 양호하다. Model 3가 포함하는 추정 모수의 개수가 Model 2보다 6개 적기 때문에 본 논문에서는 모수 절약의 원칙에 따라 Model 3를 최종 모델로 사용하였다. Model 3에서 모든 변수의 분산팽창계수(Variance Inflation Factor)의 값이 3보다 작기 때문에 다중공선성 문제는 없다고 여겨진다.

<Table 1> Representative words for inferred aspects

Inferred Aspects	Representative Words
Location	line, train, bus, station, subway, within, airport, block, garden, transportation, distance, car, speed, mn, kitchen, terrace, hour, square, cab, express, top, center, avenue, stop, library, street, wi, minute, wall, stadium, building, front, route, ferry, river, grand, high, ave, union, right, nearby, rooftop, desk, amenity, store, lined, linen, bed, near, metro, foot, dryer, morning, mall, starbucks, fifth, plaza, shower, laundry
Birth	originally, london, grew, california, francisco, san, miami, paris, came, australia, angeles, berlin, moved, germany, florida, decade, france, washington, uk, italy, rio, chicago, usa, brazil, japan, england, since, canada, fell, spain, texas, seattle, orleans, israel, hong, kong, jersey, transplant, zealand, stint, los, hawaii, janeiro, barcelona, rome, lived, india, recently, austin, singapore, greece, coast, north, ago, prague, sydney, carolina, spent, tokyo, rico, virginia
Introduction	woman, im, worker, entrepreneur, social, female, creative, non, health, dancer, engineer, gay, former, employed, male, medium, maker, hardworking, profit, educator, singer, songwriter, healthcare, coach, executive, field, marketing, software, instructor, teacher, digital, technology, therapist, professional, mid, musician, mom, freelance, lawyer, lgbtq, esl, actress, chef, self, violinist, graphic, model, actor, mature, organization, smoker, educated, humanitarian, agency, artist, adventurer
Attraction	cuisine, course, trying, seeing, eating, cultural, especially, activity, food, nature, hiking, music, dancing, outdoors, movie, learning, hobby, biking, wine, sport, outdoor, dance, hike, history, discovering, delicious, skiing, foodie, experiencing, type, listening, expose, adventure, beer, art, cooking, concert, playing, photography, watching, diving, interest, surfing
Interaction	guest, someone, welcoming, anyone, treat, others, expect, comfortable, respect, provide, important, stay, everyone, leave, visitor, host, providing, understand, would, staying, open, believe, comfort, safe, truly, appreciate, come, able, welcome, want, privacy, goal, hope, accommodate, welcomed, cozy, warm, yet, reciprocate, excited, giving, pleasant, affordable, sure, feeling, strive, wish, create, think, respectful, treated, value, give

<Table 2> Model summary and comparison

	Model 1	Model 2	Model 3
R^2	0.3908	0.3915	0.3912
Adjusted R^2	0.3904	0.3907	0.3906
Mean Squared Residual	1.317	1.317	1.317
F statistics	867.5	491.4	664.3
P value	<0.001	<0.001	<0.001
DF	13	23	17
logLik	-29804.22	-29794.92	-29798.68
BIC	59755.07	59834.22	59783.09
deviance	30506.84	30474.59	30487.62
DF residual	17578	17568	17574
Comparison Model 1			
Sum of Square		32.25	19.22
DF		10	4
F statistics		1.8591	2.7698
P value		<0.05	<0.05
Comparison Model 2			
Sum of Square			13.03
DF			6
F statistics			1.2519
P value			0.2762

<Table 3>에 있는 Model 3의 통제변수의 회귀계수 추정치에 의하면 superhost에 대한 여부, Host 등록일 이후의 일수, 숙박의 취소정책, 가독성은 리뷰의 수와 양의 상관관계를 가진다. 반면에 호스트의 응답시간이 길고 숙박요금이 높을 수록 리뷰의 수는 감소한다. Aspect 변수에 대한 계수 추정치가 모두 양의 값을 가지고 있으므로 Entire home에 대해서는 각 aspect의 문장의 수가 증가할수록 공유 성과에 긍정적인 영향이 있음을 추론할 수 있다.

교호작용 항을 통해 Private room과 Shared room에서 각 aspect 변수의 기울기가 Entire home에서와 통계적으로 유의한 차이가 있는지를 알 수 있다. $D_{Private Room}$ 과 Attraction의 교호작용이 음의 값으로 유의하기 때문에 Private room에 대해서 Attraction의 기울기는 Entire home에 대한 Attraction의 기울기보다 완만한 값(0.0319)

임을 알 수 있다. Shared room에 대해서는 Birth와 Interaction의 교호작용 항의 계수가 양수이므로 Entire home에 대한 기울기보다 더 가파르기 때문에 해당 aspect에 대한 서술이 Entire home보다 더 민감하게 공유성과에 긍정적인 영향을 미친다고 할 수 있다. 이와 반대로 Shared room과 Attraction의 교호작용 항이 음의 값을 가지는데 이는 오히려 Entire home보다 공유성과에 좋지 않은 영향을 미치는 것을 나타낸다.

<Table 4>는 이러한 교호작용 효과를 고려하여 계산된 각 숙소 유형에서 aspect에 대한 기울기와 유의성 검정 결과를 나타낸다. 유의수준 0.05 하에서 Entire home의 경우 Location, Birth, Attraction, Interaction에 대한 문장이 공유성과에 양의 영향을 미친다. 특히 Interaction은 0.0997의 가장 큰 기울기를 가진다. Private room의 경우 Entire home과 유사하지만 Attraction이 성

<Table 3> Regression coefficient estimates: Model 3

	Variable	Coef.	(SE.)
	(Intercept)	2,6110 ***	(0.1082)
Control Variables	HostResponseTime	- 0,3840 ***	(0.0091)
	Superhost	1,3850 ***	(0.0283)
	HostSince	0,0001 ***	(0.0000)
	log(Price)	- 0,1243 ***	(0.0189)
	Cancellation	0,3990 ***	(0.0123)
	Readability	0,0001 .	(0.0000)
	Room Type	$D_{Private Room}$	- 0,1047 ***
$D_{Shared Room}$		- 0,2069	(0.1264)
Aspect	Location	0,0394 ***	(0.0106)
	Birth	0,0284 *	(0.0141)
	Introduction	0,0077	(0.0142)
	Attraction	0,0716 ***	(0.0160)
	Interaction	0,0997 ***	(0.0100)
Interaction (Room Type * Aspect)	$D_{Private Room} : attraction$	- 0,0397 .	(0.0233)
	$D_{Shared Room} : birth$	0,2196 .	(0.1219)
	$D_{Shared Room} : attraction$	- 0,1653 .	(0.0870)
	$D_{Shared Room} : interaction$	0,1232 .	(0.0707)

*** p -value < 0.001, ** p -value < 0.01, * p -value < 0.05, . : p -value < 0.1

<Table 4> Comparison of regression slopes of aspects

Aspect \ Room Type	Entire Home Coef. (SE)	Private Room Coef. (SE)	Shared Room Coef. (SE)
Location	0.0394 *** (0.0106)	0.0394 *** (0.0106)	0.0394 *** (0.0106)
Birth	0.0284 * (0.0141)	0.0284 * (0.0141)	0.0281 * (0.1211)
Introduction	0.0077 (0.0142)	0.0077 (0.0142)	0.0077 (0.0142)
Attraction	0.0716 *** (0.0160)	0.0319 . (0.0172)	-0.0937 (0.0856)
Interaction	0.0997 *** (0.0100)	0.0997 *** (0.0100)	0.2230 ** (0.0702)

*** p -value < 0.001, ** p -value < 0.01, * p -value < 0.05, . : p -value < 0.1

과에 미치는 영향이 비교적 작다. Shared room의 경우 Location, Birth, Interaction이 성과에 양의 영향을 미친다.

각 aspect를 기준으로 살펴보면 Location, Birth, Introduction은 숙소 유형 간에 성과에 미치는 영향이 통계적으로 유의한 차이를 보이지 않았다. 그러나 Attraction은 Entire home에서 가장 영향이 컸고 Interaction은 Shared room에서 가장 영향이 강했다. 즉, Entire home에서는 Attraction에 대한 언급이 많을수록, 그리고 Shared room에서는 Interaction에 대한 언급이 많을수록 공유 성과에 긍정적 영향을 미쳤음을 의미한다. 게스트와의 상호작용에 대한 정보가 세 가지 숙소 유형에서 모두 가장 큰 영향을 미쳤는데 이는 특히 호스트나 함께 시설을 사용하는 게스트들과 상호작용이 가장 빈번하게 일어날 수 있는 Shared room에서 가장 영향이 컸다.

종합해보면, 세 가지 시설타입에서 모두 Interaction에 대한 내용이 host description 상에 가장 빈번히 등장하였고 이는 공유성과도 가장 크게 긍정적인 영향을 미쳤다. 그리고 이러한 영향은 shared room에서 가장 크게 나타났다. 하지만 두 번째로 빈번하게 등장한 Attraction은 Entire home 외에는 유의한 영향을 주지 못했다. 시설 주변의 Attraction에 대한 정보들은 Google Map, Tripadvisor 등과 같은 외부 출처로부터

얻을 수 있는 정보임에 반해 호스트와 게스트 사이의 상호작용과 관련된 내용은 host description 외에서는 얻기 힘든 정보들이다. 또한 Shared room의 특성 상 호스트와 게스트 사이 또는 게스트와 게스트 사이의 상호작용이 이용자의 만족도에 크게 영향을 미칠 것이므로 특히 Interaction에 대한 내용이 크게 공유성과에 영향을 준 것으로 판단된다.

5. 결론 및 한계점

본 연구에서는 숙소 유형별로 aspect가 공유성과에 미치는 영향을 비교 분석하였다. Deep learning 기반의 ABAE 모델을 사용하여 host description 상의 aspects를 3~15개의 군집으로 추출하였다. 그리고 가장 우수한 coherence score를 보였던 5개의 aspects(Location, Birth, Introduction, Attraction, Interaction)가 각각의 시설유형에서 공유성과에 미치는 영향을 회귀분석을 통해 검증하였다.

본 연구는 P2P 플랫폼에서도 제공하는 숙소 유형에 따라 효과적인 자기 표현 전략이 다를 것을 발견하였다. 먼저, Entire home, Private room, Shared room에서 모두

동일하게 지역에 대한 소개, 호스트의 출신, 문화 및 즐길거리, 게스트와의 친화의도가 공유 성과에 긍정적인 영향을 주고 있는 결과를 보였다. 반면에 호스트 성격 및 직업에 대한 설명은 크게 영향을 주지 않는다. 한편, Shared room은 친화의도가 Entire home과 Private room에 비해 성과에 더 큰 영향을 미치고 있다. 이는 Shared room을 사용하는 사용자들의 인구통계학적 특성상 현지인들과의 교류를 선호하는 집단일 가능성이 크기 때문에 호스트의 친화의도가 성과에 좋은 영향을 미쳤다는 실증적 증거를 제시했다는 데에 그 의의가 있다.

본 연구의 연구결과는 호스트가 공유성과를 극대화하기 위해 취해야 할 전략에 대한 힌트를 제공한다. 특히, 호스트가 게스트에게 본인의 소개를 함에 있어서 제공하는 숙소 유형에 따라 효과적인 방식이 있고, 이를 적절하게 활용함으로써 공유성과를 극대화할 수 있다. 예를 들어, Shared room을 제공하는 호스트라면 “자발적(spontaneous)”, “열린 마음(open minded)”, “기꺼이 도와주겠다(willing to help).”와 같이 게스트와의 정서적 교류를 의미하는 표현¹⁾을 적극적으로 사용하는 것이 유리할 수 있다. 또한 텍스트의 문장 단위에서 효율적으로 특성을 추출하고 이를 기반으로 성과를 측정하는 일반화된 프레임워크를 제안함으로써 이를 기반으로 다른 분야의 텍스트 분석에 활용될 수 있는 방법론을 제공한다.

본 논문은 두 가지 한계점을 가진다. 먼저, 표본으로 뉴욕이라는 하나의 도시에 대한 Airbnb 데이터를 사용하여 작성되었다. 따라서 본 논문의 결과를 다른 도시로 확장하여 해석하는 데 제약이 있다. 하지만 본 논문이 사용한 방법론은 범용적으로 적용 가능하기

때문에 추가적인 데이터 수집을 통해 이러한 한계를 해소할 수 있다. 두 번째로 Entire home과 Private room에 비해서 Shared room에 대한 샘플 수가 현저히 적다. 따라서 ABAE 모델에 의해서 추출된 aspect가 Shared room의 host description을 충분히 대표하지 못할 수 있고, Shared room에서 회귀계수를 추정하는데 있어서 불확실성이 커진다. 이러한 한계점은 다른 도시의 Airbnb 데이터를 추가함으로써 극복할 수 있다.

1) “I am spontaneous, open minded, love to travel and find it amazing to host people from all over the world, I love different languages and cultures”. “Always willing to help just let me know how”, Host description 일부 발췌.

〈참고문헌〉

[국내 문헌]

1. 김연미, 한진수 (2011). 호텔 웹 광고의 광고속성이 광고태도와 브랜드태도, 구매의도에 미치는 영향. *지식경영연구*, 12(1), 1-16.
2. 이경민, 배채운, 정남호 (2018). 4차 산업혁명 시대의 공유경제 생태계 정책 제안: 우버 (Uber) 사례를 중심으로. *지식경영연구*, 19(1), 175-202.
3. 임현아, 최재원, 이홍주 (2019). 텍스트 분석을 통한 제품 분류 체계 수립방안: 관광분야 App 을 중심으로. *지식경영연구*, 20(3), 139-154.

[국외 문헌]

4. Abrate, G., & Viglia, G. (2019). Personal or product reputation? Optimizing revenues in the sharing economy. *Journal of Travel Research*, 58(1), 136-148.
5. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning research*, 3(5), 993-1022.
6. Botsman, R., & Rogers, R. (2010). *What's mine is yours-How collaborative consumption is changing the way we live*. London: Collins.
7. Chang, K. T. T., Chen, W., & Tan, B. C. Y. (2012). Advertising effectiveness in social networking sites: Social ties, expertise, and product type. *IEEE Transactions on Engineering Management*, 59(4), 634-643.
8. Chang, T., Mills, G., & Wildt, A. R. (1996). Impact of product information on the use of price as a quality cue. *Psychology and Marketing*, 13(1), 55-75.
9. Cheng, M., & Foley, C. (2018). The sharing economy and digital discrimination: The case of Airbnb. *International Journal of Hospitality Management*, 70, 95-98.
10. Cheng, M., & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, 76, 58-70.
11. Ellison, N. B., Hancock, J. T., & Toma, C. L. (2012).

Profile as promise: A framework for conceptualizing veracity in online dating self-presentations. *New Media and Society*, 14(1), 45-62.

12. Ellison, N., Heino, R., & Gibbs, J. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication*, 11(2), 415-441.
13. Garcia, M. N., Munoz-Gallego, P. A., Viglia, G., & Gonzalez-Benito, O. (2019). Be social! The impact of self-presentation on peer-to-peer accommodation revenue. *Journal of Travel Research*, 59, 1268-1281.
14. Giuffrida, M., Mangiaracina, R., Perego, A., & Tumino, A. (2017). Cross-border B2C e-commerce to Greater China and the role of logistics: A literature review. *International Journal of Physical Distribution and Logistics Management*, 47(9), 772-795.
15. Goffman, E. (1978). *The presentation of self in everyday life*. London: Harmondsworth.
16. Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence*, 1(1), 60-76.
17. Guttentag, D., Smith, S., Potwarka, L., & Havitz, M. (2018). Why tourists choose Airbnb: A motivation-based segmentation study. *Journal of Travel Research*, 57(3), 342-359.
18. Hassanein, K., & Head, M. (2005). The impact of infusing social presence in the web interface: An investigation across product types. *International Journal of Electronic Commerce*, 10(2), 31-55.
19. He, R., Lee, W. S., Ng, H. T., & Dahlmeier, D. (2017). An unsupervised neural attention model for aspect extraction. *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 1, 388-397.
20. Jin, W., Ho, H. H., & Srihari, R. K. (2009). A novel lexicalized HMM-based learning framework for web opinion mining. *In Proceedings of the 26th Annual International Conference on Machine Learning*, 465-472.
21. Lutz, C., & Newlands, G. (2018). Consumer segmentation within the sharing economy: The case of Airbnb. *Journal of Business Research*, 88, 187-196.

22. Ma, X., Hancock, J. T., Mingjie, K. L., & Naaman, M. (2017). Self-disclosure and perceived trustworthiness of Airbnb host profiles. *In Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 2397-2409.
23. Mikolov, T., Yih, W. T., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. *In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746-751.
24. Mimno, D., & McCallum, A. (2008). Topic models conditioned on arbitrary features with Dirichlet-Multinomial regression. *In Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, UAI, 411-418.
25. Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 262-272.
26. Mou, J., Zhu, W., & Benyoucef, M. (2019). Impact of product description and involvement on purchase intention in cross-border e-commerce. *Industrial Management and Data Systems*, 120(3), 567-586.
27. Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., & Zaki, M. (2014). Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research*, 17(3), 278-295.
28. Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108, 42-49.
29. Roberts, M. E., Stewart, B. M., & Airoidi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515), 988-1003.
30. Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 15, 399-408.
31. Schau, H. J., & Gilly, M. C. (2003). We are what we post? Self-presentation in personal web space. *Journal of Consumer Research*, 30(3), 385-404.
32. Wagner, W. (2010). Steven Bird, Ewan Klein and Edward Loper: Natural language processing with Python, analyzing text with the natural language toolkit. *Language Resources and Evaluation*, 44(4), 421-424.
33. Wang, L., Liu, K., Cao, Z., Zhao, J., & De Melo, G. (2015). Sentiment-aspect extraction based on restricted boltzmann machines. *In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 1, 616-625.
34. Wang, W., Pan, S. J., Dahlmeier, D., & Xiao, X. (2016). *Recursive neural conditional random fields for aspect-based sentiment analysis*. arXiv preprint arXiv:1603.06679.
35. Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, 27(2), 634-639.
36. Younis, E. M. (2015). Sentiment analysis and text mining for social media microblogs using open source tools: An empirical study. *International Journal of Computer Applications*, 112(5), 44-48.
37. Zhang, J. (2019a). Listening to the consumer: Exploring review topics on Airbnb and their impact on listing performance. *Journal of Marketing Theory and Practice*, 27(4), 371-389.
38. Zhang, J. (2019b). What's yours is mine: Exploring customer voice on Airbnb using text-mining approaches. *Journal of Consumer Marketing*, 36, 655-665.
39. Zhang, L., Yan, Q., & Zhang, L. (2020). A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb. *Decision Support Systems*, 133, 113288.

● 저 자 소 개 ●



심 지 환 (Sim, Ji Hwan)

현재 국민대학교 일반대학원 데이터사이언스학과 박사과정에 재학중이다. 동 대학교 대학원 회계학과에서 석사학위를 취득하였고, 한국기업공헌평가원, 삼일회계법인 감사사업부 등 여러 외부기관에서 보조연구원으로 활동하였다. 주요 관심분야는 딥러닝기반 텍스트 마이닝, 비지도학습, 준지도학습 등이다.



김 소 영 (Kim, Soyeong)

현재 국민대학교 일반대학원 데이터사이언스학과 석사과정으로 재학 중이다. 주요 연구 관심분야는 머신러닝 및 딥러닝, 자연어 처리, 텍스트 마이닝이다.



정 여 진 (Chung, Yeojin)

연세대학교에서 경제학 및 응용통계학 복수전공으로 학사를 취득하였으며, 동 대학원에서 응용통계학 석사학위를 취득한 후 Pennsylvania State University에서 통계학 박사학위를 취득하였다. 현재 국민대학교 경영대학 경영학부 부교수로 재직중이다. 주요 연구분야는 머신러닝, 비모수분포추정, 모형기반 군집분석, hierarchical linear model, 일반화 선형모형이다.

〈 Abstract 〉

Impact of Self-Presentation Text of Airbnb Hosts on Listing Performance by Facility Type

Sim, Ji Hwan^{*}, Kim, So Young^{**}, Chung, Yeojin^{***}

In accommodation sharing economy, customers take a risk of uncertainty about product quality, which is an important factor affecting users' satisfaction. This risk can be lowered by the information disclosed by the facility provider. Self-presentation of the hosts can make a positive effect on listing performance by eliminating psychological distance through emotional interaction with users. This paper analyzed the self-presentation text provided by Airbnb hosts and found key aspects in the text. In order to extract the aspects from the text, host descriptions were separated into sentences and applied the Attention-Based Aspect Extraction method, an unsupervised neural attention model. Then, we investigated the relationship between aspects in the host description and the listing performance via linear regression models. In order to compare their impact between the three facility types(Entire home/apt, Private rooms, and Shared rooms), the interaction effects between the facility types and the aspect summaries were included in the model. We found that specific aspects had positive effects on the performance for each facility type, and provided implication on the marketing strategy to maximize the performance of the shared economy.

Key Words: Sharing economy, Self-presentation, Aspect extraction, Text-mining, Unsupervised learning

* Kookmin University

** Kookmin University

*** Kookmin University