

제품 출시 초기의 고객 불만 모니터링을 위한 소셜미디어 마이닝 접근법

이승현 · 최재웅 · 윤장혁[†]

건국대학교 산업공학과

A Social Media Mining Approach for Monitoring Customer Complaints in the Early Stage of Product Launch

Seunghyun Lee · Jaewoong Choi · Janghyeok Yoon

Department of Industrial Engineering, Konkuk University

This study suggests a social media mining approach for monitoring critical complaints in the early stage of product launch via word embedding, clustering and sentiment analysis. This approach selectively uses social media product reviews created in the beginning stage of product launch. Word embedding and clustering are used to identify customer complaint factors from reviews. Next, sentence-level sentiment analysis is conducted based on the understanding of review writing pattern, thereby calculating urgency and severity in a quantitative manner. Finally, a customer-stated complaint portfolio map is developed, showing critical complaint factors. In the case study of smartphone ‘Galaxy S10’, the critical complaint factors related to Samsung pay connection error, and fingerprint recognition are identified, which have been rapidly improved since the product launch, indeed. We expect that this approach contributes to identifying critical customer complaints in the early stage, in addition to enabling the prompt response to customer complaints.

Keywords: Customer Complaint, Social Media Mining, Sentiment Analysis, Word Embedding, Clustering

1. 서론

경쟁적 비즈니스 환경에서 기업은 시장 내 경쟁우위 확보를 위해, 품질 관리를 통해 양질의 제품을 고객에게 제공하고자 한다(Herrmann *et al.*, 2000). 제품 출시 이후 제품의 품질을 관리하기 위해서 제품을 실제로 사용한 고객들의 불만사항, 의견을 활용한다(Tang and Yun, 2008). 제품에 대한 고객의 반응은 기업이 제공한 제품의 실제 성능, 기능 및 서비스의 품질에서 비롯되므로, 출시 이후의 고객의 의견은 제품 품질 관리 뿐만 아니라 고객관계관리 등의 측면에서 중요하다(Aguwa *et al.*, 2017). 특히, 고객의 불만 사항은 제품의 잠재적인 위험 요소일 가능성이 있고, 제품 기획 과정에서 놓친 부분, 기술적 결함, 불완전한 제품

서비스 시스템 등과 같은 제품에 대한 피드백을 제공할 수 있다. 제품에 대한 고객 불만을 해결하는 과정에서 제품을 개선하거나 신제품 기획의 새로운 아이디어가 발생할 수 있다(Joung *et al.*, 2019; Misuraca *et al.*, 2020). 고객의 불만과 같은 피드백은 제품 수명 주기 내에서 중요한 역할을 하며, 특히 제품 출시 이후 초기 기간에 나타난 피드백은 후기에 발생할 수 있는 위험을 예방할 수 있고, 제품 개발 및 개선의 방향성과 같은 대응 전략을 조기에 도출할 수 있어 더욱 중요하다(Roch and Mosconi, 2016). 기업은 설문조사, 소셜미디어, 인터뷰 등을 통해 고객들의 불만 사항을 파악할 수 있는데, 특히 포스트, 리뷰, 댓글 등의 소셜미디어 내 고객 의견 데이터를 쉽게 구할 수 있다(Stevens *et al.*, 2018).

논문은 교육부와 한국연구재단의 지원으로 지원을 받아 수행된 「대학혁신지원사업」의 연구결과임. 이 성과는 정부(과학기술정보통신부)의 지원으로 한국연구재단의 지원을 받아 수행된 연구임(2019R1A2C1090228).

[†] 연락저자 : 윤장혁 교수, 05029 서울특별시 광진구 능동로 120 건국대학교 산업공학과, Tel : 02-450-0453, Fax : 02-450-3525,

E-mail : janghyoon@konkuk.ac.kr

2021년 1월 22일 접수; 2021년 3월 5일 수정본 접수; 2021년 4월 16일 게재 확정.

소셜미디어는 사용자가 본인의 경험을 콘텐츠로 생성하거나 다른 사용자와 공유 및 교환할 수 있는 인터넷 기반의 플랫폼이다(Kaplan and Haenlein, 2010). 소셜미디어는 인터넷의 발전에 따라 접근성이 용이해지면서, 많은 고객들이 소셜미디어를 사용하고 있다(Stieglitz et al., 2018). 소셜미디어를 통해, 제품을 실제로 사용한 고객들은 제품에 대한 경험을 공유하고, 리뷰를 작성하고, 잠재 고객들은 리뷰를 읽고 제품에 대한 정보를 얻는다. 따라서, 많은 기업들은 제품에 대한 고객들의 의견, 불만 등을 얻기 위해 소셜미디어에 집중하고 있다(Phang et al., 2015). 최근의 통계자료에 따르면(Hootsuite, 2020), 2020년 2월 기준 1억 4천만여 기업이 페이스북, 인스타그램과 같은 소셜미디어 앱을 사용한다. 미국의 페이스북 사용자 중 58%는 기업의 게시물을 접한 후 해당 기업 및 브랜드에 더욱 관심을 갖게 되었으며, 인스타그램의 경우 다양한 정보를 얻기 위해 기업의 계정을 구독하는 사람들은 월 평균 2.5%씩 증가하고 있다. 다수의 기업들은 소셜미디어 계정을 통해 새로운 제품, 서비스 혹은 이벤트 등을 홍보하여 고객과의 유대관계를 형성하고, 해당 계정에 작성된 고객의 의견에 귀 기울이고 있다.

소셜미디어에서의 고객 의견을 효과적으로 분석하기 위한 학술적 시도 또한 증가하고 있다. 많은 연구자들이 소셜미디어 데이터로부터 고객들이 공통적으로 언급하는 요소를 파악하거나 잠재된 주제를 도출해내기 위해 텍스트 마이닝 기법을 적용하였다. 또한, 소셜미디어 데이터를 작성한 고객의 긍정 및 부정과 같은 정서의 극성을 파악하여 정량화 하기 위해 감성분석을 수행하였다. 예를 들어, 온라인 서비스에 대한 국내/국의 고객의 반응 차이를 조사한 연구(Wu and Chang, 2020), 타겟 제품의 대체 제품을 순위화 하는 방법을 제시한 연구(Liu et al., 2017), 고객 관점의 제품 요소에 대한 만족도와 중요도를 평가하여 제품 기회를 도출한 연구(Jeong et al., 2019)가 있다. 이외에도, 소셜미디어 데이터를 기반으로 고객 만족도 측정(Song et al., 2016), 고객 유형과 고객 만족도의 관계 파악(Chang et al., 2019), 고객 간의 커뮤니티 식별(Xie et al., 2014) 등 다양한 연구가 존재한다.

하지만, 대부분의 소셜미디어 분석 연구들은 다음과 같은 두 가지 보완점이 존재한다. 첫 번째로, 소셜미디어 데이터에 감성분석을 적용하는데 있어서, 고객들이 소셜미디어에 리뷰를 쓰는 패턴에 대한 이해가 필요하다. 소셜미디어 데이터는 하나의 주제에 대해서 일관적인 내용을 언급하기 보다는, 여러 주제에 대한 다양한 의견을 포함하고 있다(Hu et al., 2019). 따라서, 선행연구와 달리, 소셜미디어 데이터에 대한 감성분석의 적용은 문서 단위가 아닌 문장과 같은 세밀한 단위의 분석이 보다 정확한 결과를 제공할 수 있다. 예를 들어, 한 리뷰를 구성하는 문장을 살펴보면 다음과 같을 수 있다. “이 제품을 사용했을 때, 그래도 외관 디자인은 괜찮았다. 하지만 배터리 지속 시간은 정말 최악이었다.” 하나의 리뷰 문서에서 두 가지 제품 요소(디자인, 배터리)가 언급되었으며 각 요소에 대한 감성은 서로 다르다. 해당 리뷰에서, 문서 단위의 감성 분석

보다는 보다 세밀한 문장 수준의 감성분석이 고객 불만에 대한 보다 정확한 결과를 제공할 수 있다. 두 번째로, 제품에 대한 소셜미디어 데이터를 모니터링하는 데 있어서, 제품의 출시 시기가 데이터 수집 과정에서 고려되어야 한다. 소셜미디어 데이터의 분석 결과는 제품 출시 이후의 시간에 따라 내용과 중요성이 다를 수 있다. 특히 제품 출시 이후 초기 단계는 후기 단계에 발생할 수 있는 위험을 최소화 할 수 있고, 고객의 의견을 수렴하기 위한 대응 전략을 조기에 제시할 수 있으므로 더욱 중요하다(Roch and Mosconi, 2016). 따라서, 소셜미디어 데이터를 사용할 경우 데이터가 생성된 시기를 제품 출시일과 함께 고려할 필요가 있다.

본 연구는 위와 같은 소셜미디어 데이터의 특성에 대한 이해를 바탕으로, 제품 출시 초기 단계에서의 고객 불만 분석을 위한 방법론을 제시한다. 우선 분석하고자 하는 타겟 제품을 선정한다. 고객들이 소셜미디어 상에서 타겟 제품에 대해 작성한 리뷰 데이터를 수집한 후, 제품 출시 이후 초기에 작성된 리뷰 데이터를 선별한다. 선별된 리뷰에서 명사구를 추출 및 선별하여 유효 명사구 집합을 구성한 후, 유효 명사구를 벡터로 변환하여 클러스터링 알고리즘을 적용해 제품에 대한 고객의 잠재적 불만 요소를 정의한다. 다음으로, 리뷰 데이터를 이루고 있는 각 문장의 감성(긍정/중립/불만)을 식별한다. 일반적으로 부정적인 리뷰를 살펴보면 처음부터 끝까지 불만이 표출되는 것이 아니라 긍정적인 부분도 함께 언급된다(Hu et al., 2019). 즉, 하나의 리뷰에는 다양한 주제와 의견 및 감성이 포함될 수 있으므로 리뷰 데이터를 문장 단위로 분할하고, 각 문장에 감성분석을 적용하여 문장의 감성을 파악한다. 마지막으로, 불만 문장을 활용하여 잠재적 불만 요소의 불만 정도를 정량적으로 나타낼 수 있는 두 가지 지표(Urgency, Severity)를 계산하고, 정량적 지표를 기반으로 고객 불만 포트폴리오 맵을 형성한다. 고객 불만 포트폴리오 맵을 통해 잠재적 불만 요소 중 가장 개선이 시급한 불만 요소를 식별할 수 있으며, 이에 대한 대응 전략을 도출할 수 있다.

본 연구의 기여점은 다음과 같다. 첫째, 본 연구는 소셜미디어 데이터의 이해를 바탕으로 한, 소셜미디어 마이닝 기반의 고객 불만 모니터링 접근법을 제시한다. 앞서 언급했듯이, 소셜미디어에 작성된 제품 리뷰 문서는 문장별로 다른 주제를 다룰 수 있고, 다른 감성점수를 가질 수 있기 때문에, 본 연구는 문장 단위의 감성분석을 수행한다. 또한, 본 연구는 소셜미디어 데이터를 수집하고 분석하는 데 있어서, 제품의 출시일과 리뷰의 생성일자를 고려하여 초기 불만에 대한 정보를 선별하였다. 이와 같은 접근방식들은 연구자들이 소셜미디어의 제품 리뷰를 다루는 데 있어서, 보다 정확한 고객 이해를 할 수 있도록 도울 수 있다. 둘째, 본 연구는 리뷰 데이터로부터 고객의 불만을 측정할 수 있는 두 가지 지표를 개발하여 제시한다. 본 연구에서 개발한 고객 불만 지표들은 기업이 고객 리뷰 데이터를 모니터링하여 발생한 모든 불만 요소들을 계량적으로 평가하고 우선적 불만 요소를 파악하는 것을 지원할 수 있다.

셋째, 본 연구는 전문가의 개입이 최소화된 연구단계로 구성 되어 있기 때문에, 실제 비즈니스 환경에서 재현되어 고객 불만을 모니터링하는데 활용될 수 있다. 제안된 방법론의 파라미터들은 과학적 기법을 통해 결정되었기 때문에, 다양한 환경에서 재현될 수 있고, 입력 데이터에 맞는 적절한 결과를 제공할 수 있다.

본 논문의 구성으로는 제 2장에서 고객 불만 분석과 소셜미디어 활용에 대한 연구들을 설명하고, 제 3장에서는 본 연구의 연구 절차를 구체적으로 설명한다. 제 4장에서는 제안된 방법론의 사례연구 및 주요 결과를 제시한다. 마지막으로 제 5장에서는 본 연구의 결론 및 추후연구에 대해 기술한다.

2. 관련 연구

고객 만족은 제품의 구매 혹은 소비의 결과로, 고객이 기대했던 정도와 이익 및 비용을 비교하여 나타내는 것이다(Sezgen *et al.*, 2019). 고객은 제품을 사용한 후의 경험과 사용 전의 기대를 비교하여 만족 혹은 불만족을 경험한다. 현대의 고객들은 이와 같은 경험을 소셜미디어 매체를 통해 리뷰로 작성하여 다른 고객들과 공유하기도 한다(Xiao *et al.*, 2016). 따라서 고객들이 소셜미디어에 작성한 리뷰는 고객이 제품이나 서비스를 사용하는데 있어서, 불만 혹은 만족감을 느낀 구체적인 제품 요소들을 담고 있고, 소셜미디어 리뷰 분석은 개선이 필요한 문제점을 도출해낼 수 있다(Trappey *et al.*, 2018). 많은 연

구에서 소셜미디어 분석을 통해 고객의 요구 사항을 파악하고자 하였으며, 나아가 제품의 품질을 정량화하고자 하였다.

고객의 만족도를 파악하기 위해 고객 설문 점수 혹은 리뷰의 평점을 분석하는 다수의 연구들이 수행되었다. 예를 들면 제품에 대한 온라인 평점을 활용하여 제품의 장점과 단점을 식별한 연구(Qian, 2011), 호텔에 대한 평점을 통해 서로 다른 문화권의 고객 만족도 차이를 분석한 연구(Schuckert *et al.*, 2015), 호텔에 대한 리뷰 평점을 사용하여 고객 리뷰가 호텔 점유율에 미치는 영향을 분석한 연구(Viglia *et al.*, 2016) 등이 있다. 하지만 일반적인 경험 및 만족도를 암시하는 리뷰의 평점을 살펴보는 것보다 텍스트 리뷰를 분석하면 더 깊은 결과를 얻을 수 있기 때문에(Xu *et al.*, 2017), 최근에는 소셜미디어의 텍스트 리뷰에 감성분석을 적용하는 연구가 활발히 진행되고 있다(Choi, Yoon, *et al.*, 2020). 감성분석은 많은 소셜미디어 마이닝 연구에서 사용되는 기법으로, 텍스트 데이터로부터 사람들의 부정 및 긍정과 같은 감성을 식별하고 분석할 수 있다(Medhat *et al.*, 2014). <Table 1>은 소셜미디어의 텍스트 리뷰 데이터를 활용한 고객 만족을 분석한 선행 연구 목록이다.

전자상거래 사이트(e.g., Amazon, JD.com, Yelp)뿐만 아니라 Social Networking Service(e.g., Twitter, Facebook), 소셜 뉴스 커뮤니티 사이트(e.g., Reddit) 등 다양한 소셜미디어 플랫폼을 통해 제품 및 서비스에 대한 고객의 텍스트 리뷰를 사용하였다. 고객 만족에 대한 연구는 특정 산업 분야에만 한정되지 않고 전자제품, 식품, 대중교통 등 다양한 분야에서 고객의 감성을 식별하기 위해 감성분석을 적용해왔다.

Table 1. Summary of Literature Review

Social media platform	Dataset	Purpose	Reference
Twitter	Pizza, car, smart phone tweets	Comparison of users' sentiments before and after the launch of three new products	Rathore and Ilavarasan(2020)
	Tweets on various topics	User' sentiment analysis and prediction on real-time events	Yoo <i>et al.</i> (2018)
	Transit agency tweets	Modeling of customers' satisfaction in public transportation systems	El-Diraby <i>et al.</i> (2019)
Amazon	Electronic products reviews	Investigation on whether the online promotions and sentiment of reviews can predict product sales	Chong <i>et al.</i> (2016)
		Investigation on the influence of review sentiment on readership and helpfulness of online reviews	Salehan and Kim (2016)
Facebook	Hospital reviews	Measurement of customer' satisfaction and identification of service attributes related to satisfaction	Zaman <i>et al.</i> (2020)
JD.com	Cell phone reviews	Development of product improvement strategy	Qi <i>et al.</i> (2016)
Reddit	Smart speakers reviews	Identification of time-evolving product opportunities	Choi, Oh, <i>et al.</i> (2020)
Yelp	restaurant reviews	Analysis of the influence of customers' gender and location on restaurant ratings	Micu <i>et al.</i> (2017)
AppStoreHQ	Mobile application services reviews	Presenting a new framework for measuring customer satisfaction for mobile services	Kang and Park (2014)
Booking.com	Hotel reviews	Proposing a heuristic model to analyse and explore marketing insights in customer reviews	Chang <i>et al.</i> (2020)

하지만 <Table 1>에 나열된 연구들을 포함한 대부분의 연구들은 각 리뷰의 감성을 분석하기 위해 감성분석을 사용하였다. 앞서 언급하였듯이 하나의 리뷰는 여러 개의 문장으로 구성될 수 있으므로 다양한 주제 및 감성이 포함될 수 있다. 따라서 리뷰 자체의 감성을 다루기보다는 문장과 같은 세밀한 단위의 감성분석이 요구된다. 또한, 고객으로부터 얻는 정보는 제품의 수명 주기에 따라 내용과 중요성이 다를 수 있으므로, 타겟 제품에 대한 소셜미디어 데이터를 수집함에 있어 제품 출시 날짜와 리뷰 작성 날짜가 고려되어야 한다. 특히 제품 출시 이후 초기 단계에 고객 리뷰 데이터를 분석한다면 제품에 개선이 필요한 부분을 신속히 파악할 수 있고, 후기 단계에 발생할 수 있는 위험을 미리 예방할 수 있을 것이다. 따라서 본 연구는 타겟 제품에 대한 리뷰 데이터를 수집한 후, 제품 출시 이후 초기 단계에 작성된 리뷰를 선별하여 사용한다. 선별한 리뷰를 문장 단위로 모두 분할하여 감성분석을 수행하므로 고객의 감성을 자세히 파악할 것으로 기대된다. 결과적으로 제품 출시 초기 단계의 고객 불만 요소를 식별하고, 해당 불만 요소를 우선순위화 하여 가장 우선적으로 개선 조치가 필요한 불만 요소를 살펴본다.

3. 연구 방법

본 연구의 분석 과정은 <Figure 1>의 절차를 따른다. 1) 먼저, 소셜미디어 데이터를 수집하고 전처리하여 유효한 리뷰와 명사구를 선별한다. 2) 다음으로, 명사구를 벡터로 변환한 후 클러스터링 알고리즘을 적용하여 고객이 언급하는 잠재적 불만 요소를 정의한다. 3) 선별한 리뷰 데이터를 문장 단위로 분할하고, 각 문장의 감성을 식별한다. 불만 문장을 활용하여 각 불만 요소마다 불만 수준을 나타낼 수 있는 정량적 지표를 계산

한다. 4) 정량적 지표를 기반으로 고객 불만 포트폴리오 맵을 형성하고, 기업이 우선적으로 개선해야 할 불만 요소를 식별하여 불만에 대한 대응 전략을 제시한다.

3.1 소셜미디어 데이터 수집 및 전처리

본 연구의 목적은 제품 출시 초기의 모니터링을 위해 초기에 발생한 고객 불만 요소 중 가장 긴급한 요소를 파악하여 대응 전략을 제시하는 것이다. 따라서 본 방법론은 소셜미디어에서 타겟 제품에 대한 리뷰 데이터를 수집하여 분석한다. 소셜미디어 데이터는 웹 스크래핑, 플랫폼별로 구축된 Open Application Programming Interface, 온라인 데이터 웨어하우스(e.g., Google Big Query) 등을 통해 수집할 수 있다. 타겟 제품에 대한 소셜미디어 리뷰 데이터를 수집한 후, 제품 출시일 이후 초기 기간에 작성된 리뷰를 선별적으로 수집한다. 수집한 소셜미디어의 리뷰들은 제품과 관련 없는 내용을 다룰 수 있고, 혹은 길이가 너무 짧아 불만을 파악하는 것이 어려울 수 있기 때문에, 전처리 과정을 통해 양질의 데이터셋을 구축할 필요가 있다.

전처리를 통해 선별된 리뷰 데이터셋을 효과적으로 다루기 위해, 리뷰에 포함된 명사구를 추출하여 내용을 분석한다. 명사구는 Natural Language Toolkit(Loper and Bird, 2002), spaCy (<https://spacy.io/>) 등의 자연어 처리 기반의 키워드 추출 알고리즘을 적용하여 추출할 수 있다. 리뷰 데이터에서 추출한 명사구 중 일부는 제품과 관련된 내용을 나타내지 않는 키워드를 포함할 수 있으므로, 명사구를 선별하는 작업이 필요하다. 예를 들어, 추출된 명사구 중에서 특수문자, 의성어는 전처리 과정을 통해 배제될 수 있다. 또한, 잠재 불만 요소를 나타내기 어려운 명사구도 제외한다. 최종적으로, 추출된 명사구 집합에서 불용어를 제거하여 유효 명사구 집합을 구성한다.

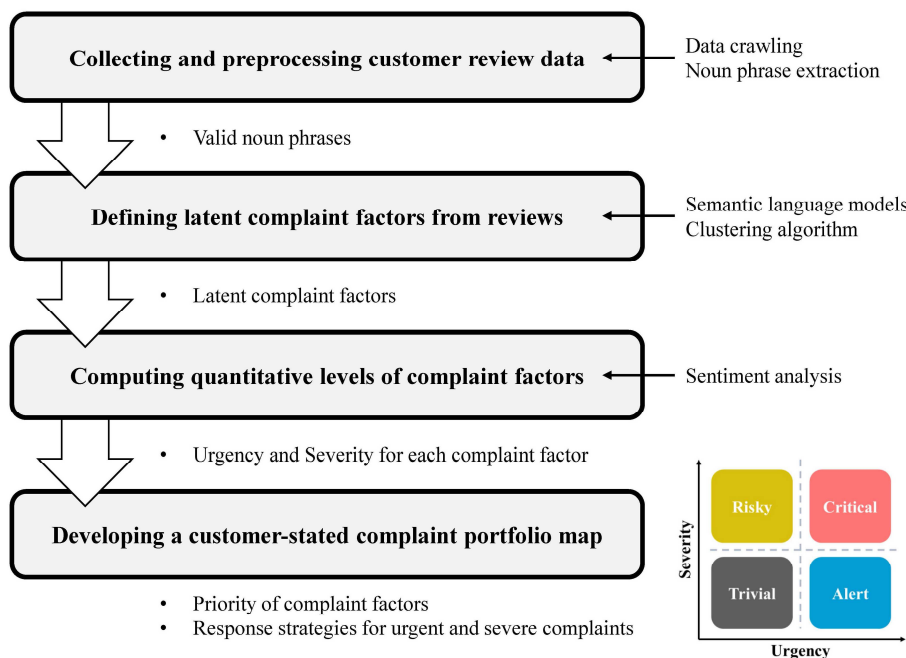


Figure 1. Overall Process of the Proposed Approach

3.2 리뷰 데이터로부터 잠재 불만 요소 추출

본 단계에서는 소셜미디어에서 제품에 대해 고객들이 언급하는 잠재 불만 요소를 정의하기 위해, <Figure 2>와 같이 앞서 구성된 유효 명사구 집합을 대상으로 사전에 학습된 Language model과 클러스터링 알고리즘을 적용한다. 유사한 명사구를 그룹화 하기 위한 단계이므로, 텍스트 형식의 명사구를 벡터화 하는 과정이 요구된다. 따라서, 키워드 임베딩 기술을 활용하여 앞서 생성한 유효 명사구 집합의 모든 명사구를 임베딩 벡터 표현으로 변환한다. 벡터의 유사도를 기반으로 유사한 명사구 그룹을 생성하기 위해 클러스터링 알고리즘을 적용한다. 클러스터링은 비지도 학습 중 가장 대표적인 기법으로, 분류되지 않은 데이터셋을 유사한 개체의 그룹(클러스터)으로 분류하는 기법이다. 클러스터링 알고리즘으로는 K-means, Mean shift, Gaussian Mixture Model, 밀도 기반 클러스터링 등이 있으며, 정확한 클러스터링 결과를 도출하기 위해서는 최적의 클러스터 수를 식별하는 작업이 중요하다(Tibshirani *et al.*, 2001). 이를 위한 방법으로는 Elbow method, Gap Statistic method 등 다양한 기법을 사

용할 수 있다.

클러스터링의 결과로, 유사한 명사구들은 하나의 클러스터로 그룹화 된다. 본 연구는 소셜미디어의 제품 리뷰 데이터에서 타겟 제품과 연관된 명사구를 선별하여 클러스터링에 적용하므로, 도출된 클러스터들은 소셜미디어에서 고객들의 잠재적인 불만 요소로 정의할 수 있다. 각 클러스터는 클러스터의 중심을 나타내는 벡터를 지니기 때문에, 클러스터를 이루는 명사구는 벡터연산을 통해 중심과의 거리를 측정할 수 있다. 따라서 클러스터의 중심과 가까운 명사구를 참고하여 레이블링 과정을 수행하고, 결과적으로 각 불만 요소의 상세한 명칭을 정의한다.

3.3 불만 수준의 계량적 지표 정의

본 단계는 우선 리뷰 데이터로부터 고객의 감성을 탐지하기 위해 감성분석을 수행한다. 소셜미디어의 리뷰에는 문장별로 주제나 감성 수준이 다를 수 있기 때문에, 본 연구에서는 보다 정확한 고객 이해를 위해 리뷰 데이터를 문장 단위로 분할하여 감성분석을 적용한다.

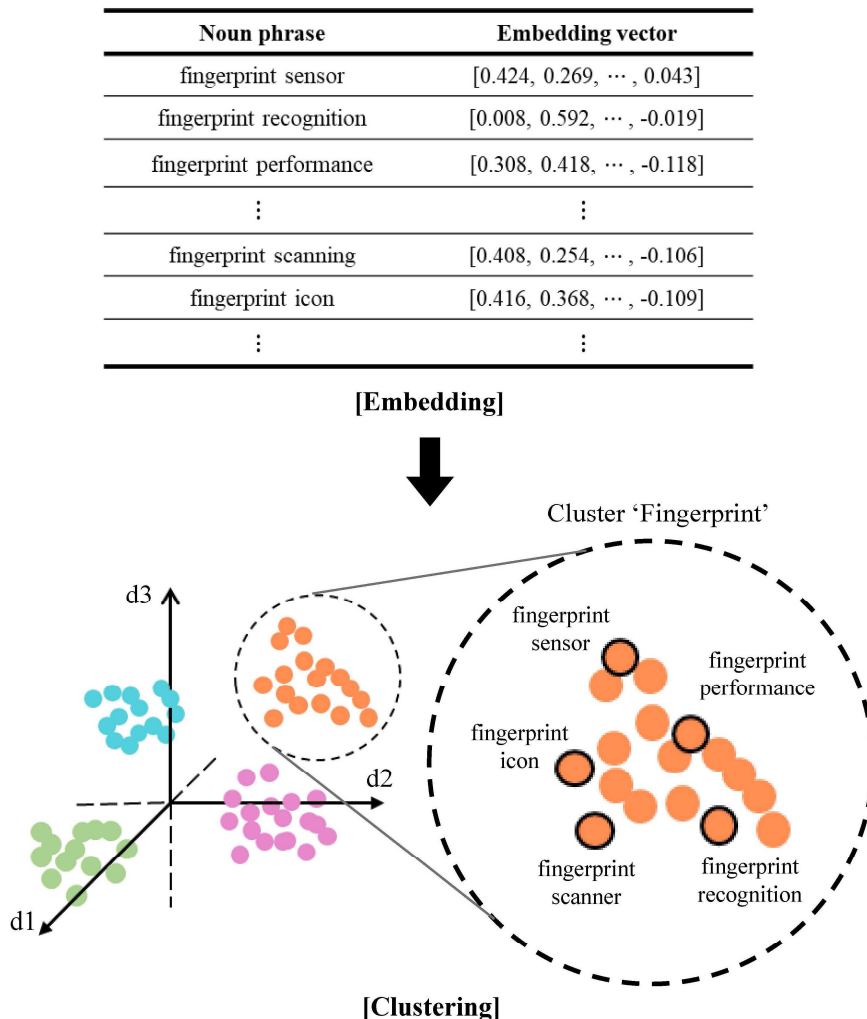


Figure 2. Latent Complaint Factor Extraction Process

Table 2. Example of Sentence-Level Sentiment Analysis Used in the Approach

	Text	Sentiment score	Sentiment result
Review	When I press the capture button the device shakes which is enough to get blurry pictures. However, when I switch to video mode it's doing completely fine.	0.1761	Positive
1st Sentence	When I press the capture button the device shakes which is enough to get blurry pictures.	-0.1027	Negative
2nd Sentence	However, when I switch to video mode it's doing completely fine.	0.2716	Positive

<Table 2>는 감성분석 도구인 VADER (Valence Aware Dictionary and sEntiment Reasoner) (<https://github.com/cjhutto/vaderSentiment>)를 사용하여 스마트폰 사용자가 작성한 실제 온라인 리뷰의 감성분석 결과와 리뷰를 구성하는 문장들의 분석 결과를 나타낸다. VADER는 어휘 및 규칙 기반의 감성분석 도구로, 텍스트를 구성하는 단어의 감성적 특성 뿐만 아니라 특수기호, 대/소문자, 접속사 등을 모두 고려하여 텍스트의 전반적인 감성을 분석한다 (Hutto and Gilbert, 2014). VADER의 분석 결과는 텍스트의 감성을 -1 이상 1 이하의 점수로 나타내며, 감성점수가 0보다 크면 긍정 (Positive) 텍스트, 0이면 중립 (Neutral) 텍스트, 0보다 작으면 불만 (Negative) 텍스트로 판단할 수 있다. 표에 나타난 리뷰 자체의 감성 점수는 0.1761로 긍정적인 감성을 나타내는 것으로 판단된다. 하지만 리뷰를 구성하는 각 문장을 살펴보면, 첫 번째 문장은 사진을 촬영할 때 쉽게 흔들리고 흐릿하게 촬영된다는 부정적인 감성을 표출하고 있다. 반대로 두 번째 문장은 잘 작동되는 동영상 촬영 기능에 대해 긍정적인 감성을 표출하고 있다. 감성 점수를 기준으로 살펴보면, 사진 촬영에 대한 감성 점수 (-0.1027)에 비해 동영상 촬영에 대한 감성 점수 (0.2716)의 절대값이 더 크기 때문에, 리뷰의 감성은 긍정적인 것으로 나타난다. 만약 리뷰 단위로 감성 분석을 수행할 경우 사진 촬영에 대한 고객의 불만은 누락될 것이다. 따라서 본 연구는 고객의 불만이 표출된 텍스트가 누락되는 것을 최소화하기 위해, 선별한 리뷰 데이터를 모두 문장 단위로 분할한 후 세밀한 수준의 감성분석을 적용하여 긍정, 중립, 불만 문장을 식별한다.

다음으로, 3.2절에서 정의한 불만 요소의 불만 수준을 계량적으로 나타내기 위한 두 가지 지표를 정의하며, 앞서 식별한 불만 문장을 기반으로 각 지표를 계산하는 수식을 제시한다. 기존의 다양한 연구자들은 고객이 제품 및 서비스 요소를 언급하는 빈도와 고객이 만족하는 정도를 활용하여 해당 요소의 중요성을 파악하였다 (Choi, Oh, et al., 2020; Jeong et al., 2019; Kim and Kim, 2019). 따라서 본 연구는 불만 요소의 우선순위를 식별하기 위해, 각 불만 요소의 심각한 정도를 정량적으로 측정하고자 불만 요소의 발생 빈도와 고객 불만 수준을 기반으로 두 가지 지표를 개발하여 제시한다.

첫 번째 지표는 Urgency이다. Urgency는 제품 출시 초기 단계에 발생하는 불만 요소의 빈도로, 해당 요소의 긴급한 정도를 의미한다. 제품의 문제점을 초기에 탐지하는 것은 후기에 필요한 비용을 절감할 수 있고, 고객의 만족도를 높일 수 있으므로 기업에게 매우 중요하다 (Deljac et al., 2015). 이때, 고객들로부터 문제점이 언급되는 빈도를 활용한다면 해당 문제의 중요성

을 측정할 수 있다 (Jeong et al., 2019). 따라서, 제품 출시 초기 단계에 발생 빈도가 높은 불만 요소는 우선적으로 해결될 필요가 있다. 임의의 클러스터에 포함된 명사구를 포함하는 문장의 감성이 불만으로 식별될 경우, 이는 해당 클러스터에 불만이 발생한 것으로 판단할 수 있다. 따라서 각 클러스터에 속하는 명사구들이 각 문장에 포함되는지에 대한 여부를 파악하고, 식 (1)과 같이 클러스터를 이루고 있는 명사구가 포함된 전체 문장 개수 대비 불만 문장 개수의 비율을 계산한다. 해당 비율은 각 불만 요소에 대하여 고객들이 실제로 불만을 표출하는 정도를 나타낸다. Urgency 값이 높을 경우 해당 불만 요소가 고객들로부터 자주 언급되어 긴급한 요소임을 의미하고, 낮을 경우 적게 언급되어 비교적 덜 긴급한 요소임을 의미한다.

$$Urgency = \frac{N(S_{neg})}{N(S_{pos}) + N(S_{neu}) + N(S_{neg})} \quad (1)$$

두 번째 지표 Severity는 불만 요소가 발생할 경우 심각한 정도를 나타내는 지표이다. 각 클러스터에 속하는 명사구들이 포함된 문장 중 불만 문장의 감성점수를 종합하여 Severity는 불만 발생 시 얼마나 심각한지를 나타낼 수 있다. 불만 문장의 감성점수는 음수이므로, 식 (2)와 같이 감성점수의 절대값 평균을 계산한다. Severity 값이 높을 경우 해당 요소에 대한 고객들의 불만 정도가 심한 것을 의미하고, 낮을 경우 불만 정도가 약한 것을 의미한다.

$$Severity = \frac{\sum |Sentiment Score(S_{neg})|}{N(S_{neg})} \quad (1)$$

3.4 고객 불만 포트폴리오 맵 형성

본 단계는 앞서 도출한 각 불만 요소의 Urgency, Severity를 기반으로 하여 고객 불만 포트폴리오 맵을 형성한다. 고객 불만 포트폴리오 맵을 통해 여러 불만 요소 중 가장 긴급한 불만 요소를 파악하여 우선적으로 개선되어야 할 요소를 식별할 수 있다. 불만 요소의 Urgency, Severity 두 지표 값이 모두 높을 경우, 해당 요소는 가장 우선적으로 개선이 되어야 함을 의미한다. 반대로 두 지표 값이 모두 낮을 경우, 해당 요소는 우선순위가 낮음을 의미한다. 고객 불만 포트폴리오 맵을 통해 정량적 접근 방법으로 긴급한 불만 요소를 도출하기 위하여, 모든 불만 요소의 Urgency 평균, Severity 평균을 기준선으로 지정하여 맵을 4개 영역으로 분할한다; Critical, Risky, Alert, Trivial (<Figure 3> 참조).

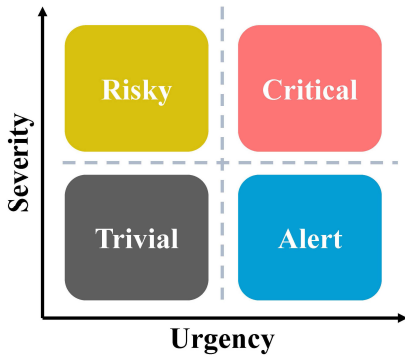


Figure 3. Four Areas of the Customer-Stated Complaint Portfolio Map

(1) Critical

Critical 영역에는 Urgency, Severity 모두 높은 값을 지니는 불만 요소가 포함된다. 이 단계에 속하는 불만 요소가 발생하면 고객들은 매우 부정적인 감성을 느끼고, 또한 다른 요소들에 비해 가장 빈번하게 발생하는 불만 요소이다. 기업은 해당 불만 요소를 즉각 개선하지 않을 경우, 불만이 점점 심해지거나 고객층을 잃을 수 있으므로 기업이 가장 우선적으로 개선해야 할 사항이다. 따라서 본 연구는 가장 긴급한 불만 요소를 식별하기 위해 Critical 영역을 중점적으로 살펴본다.

(2) Risky

Risky 영역에 속하는 불만 요소는 Urgency 값은 낮지만, 높은 Severity 값을 지닌다. 비록 빈번히 발생하는 불만 요소는 아니지만, 발생할 경우 불만의 심각성은 Critical 영역의 요소들과 유사하므로 위험한 요소임을 알 수 있다.

(3) Alert

Alert 영역은 Risky 영역과 반대로, 높은 Urgency 값과 낮은 Severity 값을 지니는 불만 요소를 포함한다. 심각하지는 않더라도 자주 발생하는 요소들이기 때문에, 해당 요소들의 Severity 값을 지속적으로 경계할 필요가 있다.

(4) Trivial

Trivial 영역의 불만 요소는 Urgency, Severity 모두 낮은 값을 지닌다. 이러한 요소들은 기업에게 심각한 영향을 미치지 않는 것으로 판단되며, 앞서 정의한 불만 요소 중 우선 순위가 가장 낮은 요소들로 판단할 수 있다.

본 연구는 4개의 영역 중, 가장 긴급한 불만 요소를 포함하고 있는 Critical 영역을 중점적으로 살펴본다. 구체적으로는 불만 요소를 이루는 명사구들이 포함된 불만 문장을 확인하여 고객들은 해당 불만 요소를 어떻게 포착했는지 자세히 모니터링한다. 다시 말해, 고객 불만 포트폴리오 맵을 형성하여 여러 불만 요소 중 우선적으로 개선되어야 할 요소들을 식별하고, 소셜미디어 상에서 고객들이 해당 요소를 어떻게 언급하였는지를 파악하여 대응 전략을 제시한다.

4. 사례연구 : Samsung Galaxy S10

본 장에서는 2019년 3월에 출시된 삼성 스마트폰 Galaxy S10을 타겟 제품으로 선정하여 앞서 설명한 연구 절차를 타겟 제품에 적용하는 사례연구를 진행한다. Galaxy S10은 국내 뿐만 아니라 해외에서도 출시되어 많은 고객들로부터 사용되었고, 이전의 삼성 스마트폰 Galaxy S 시리즈에 비하여 무선 공유 배터리, 온 스크린 지문 인식, 인공지능 기반의 사용자 맞춤형 빅스비 루틴 등 새로운 기능이 다양하게 탑재되었다. 따라서 Galaxy S10은 다양한 기능을 보유하고 있으며 제품에 대한 많은 고객들의 의견이 소셜미디어에서 활발하게 공유되었으므로 본 연구 방법에 적절한 타겟 제품으로 선정되었다.

4.1 데이터 수집 및 전처리

본 연구는 Reddit (<https://www.reddit.com>)으로부터 Galaxy S10에 대한 고객 리뷰 데이터를 수집하였다. Reddit은 소셜 뉴스 커뮤니티 사이트로, 사용자들이 새로운 글, 정보, 뉴스 등을 작성하여 공유하고 토의하는 것에 초점을 두었다. Reddit은 제품 이름, 브랜드 이름과 같은 주제에 따라 subreddit으로 게시판이 분류되어 있다. 또한, 사용자들은 각 subreddit마다 게시글 혹은 댓글을 작성하며 서로의 의견을 공유할 수 있으므로 제품에 대한 고객들의 다양한 의견을 나타낸다. 따라서 본 연구는 Galaxy S10의 subreddit(<https://www.reddit.com/r/galaxys10>)에 작성된 고객들의 게시글을 리뷰 데이터로 간주하였으며, 2018년 12월 1일부터 2019년 8월 31까지 9개월동안 작성된 25,886개의 고객 리뷰 데이터를 수집하였다.

수집한 리뷰 데이터를 분석하기에 앞서, 본 논문에서 제안하는 연구 방법에 적용할 수 있도록 <Table 3>과 같이 데이터 전처리작업을 수행하였다. 우선, 본 연구의 목적은 제품 출시 초기 단계를 모니터링하는 것이기 때문에, 본 연구는 제품 출시일로부터 6개월간 작성된 리뷰를 초기 단계의 리뷰로 판단하여 선별하였다. Galaxy S10은 2019년 3월 8일에 정식 출시되었으므로 2019년 3월부터 2019년 8월까지 작성된 리뷰를 남기고는 모두 제외하였다. 작성 날짜를 기준으로 선별한 후, 리뷰의 공백 및 특수문자를 포함한 리뷰의 글자 수가 10개 미만인 짧은 리뷰는 유효한 의미 혹은 정보를 나타내지 않는다고 판단하여 제외하였다. 추가적으로 리뷰 내에 작성된 웹 사이트 주소(e.g., Youtube, Amazon, Ebay)를 모두 삭제하여 13,976개의 유효한 리뷰 데이터를 선별하였다. 다음으로는 키워드 추출 알고리즘 spaCy를 사용하여 유효 리뷰 데이터로부터 명사구를 추출하였고, 추출한 명사구에는 의미 없는 불용어가 포함되어 있으므로 이모티콘, 의성어, 분석 대상과 관련 없는 명사구를 제거하였다. 최종적으로 2,831개의 명사구로 이루어진 유효 명사구 집합을 구성하였다.

Table 3. Data Preprocessing

Step	Example
1. Selection of reviews based on posted date	Reviews posted 6 months after product launch date
2. Exclusion of reviews less than 10 in length	“Thanks!”, “And why?”, “[removed]”
3. Elimination of webpage links	“So, after the April update my touch sounds changed to a loud beep instead of subtle click like it used to be, is there a way to change it back? //edit: the new sounds- https://www.youtube.com/~ It’s really bad”
4. Elimination of stop words	Emoticons “:D”, “:-(”, “^^”
	Onomatopoeic words “blah blah”, “hmm”, “haha”
	Irrelevant words “hello”, “hey”, “guys”

4.2 도출된 잠재 불만 요소

고객들이 제품에 대해 소셜미디어에서 언급하는 잠재 불만 요소를 파악하기 위해 앞선 단계에서 추출한 유효 명사구 집합을 의미적 유사도 기반으로 그룹화 하였다. 본 연구는 명사구를 그룹화 할 수 있도록 Sentence-BERT 모델을 사용하여 유효 명사구를 모두 임베딩 벡터로 변환하였다. Sentence-BERT는 BERT 네트워크를 수정한 것으로, 벡터 공간에서 유사한 의미의 문장이 가깝게 배치되도록 조정하는 문장 임베딩 기법이다 (Reimers and Gurevych, 2019). 각 명사구 사이의 벡터 유사도를 기반으로, 유사한 명사구들을 하나의 군집으로 그룹화 하여 잠재 불만 요소로 정의할 수 있으며, 명사구 벡터를 그룹화 하기

위해 K-means 클러스터링 기법을 사용하였다. K-means 클러스터링은 임의의 군집 수를 지정하여 각 데이터를 가장 가까운 군집으로 할당하는 기법으로(Xing *et al.*, 2017), 구현함에 있어 단순하고 효율적이기 때문에 가장 많이 사용되는 클러스터링 방법 중 하나이다(Lu and Zhou, 2016). 본 연구는 통계적 기법을 활용하는 Gap Statistic method (Tibshirani *et al.*, 2001)를 통하여 최적의 클러스터 개수를 27개로 설정한 후에 K-means 클러스터링을 수행하였으며, 각 클러스터의 중심으로부터 가까운 명사구들을 기반으로 레이블링 하여 잠재 불만 요소를 정의하였다(<Table 4>, <Figure 4> 참조). 정의된 불만 요소에 대한 상세한 설명은 <Appendix Table 1>에서 확인할 수 있다.

Table 4. Part of Customer Complaints and their Major Phrases

Complaint	Major phrases
Audio	“sound mode”, “audio output”, “audio sync”, “audio controls”, “audio option”, “sound settings”, “sound device selector”
Battery	“battery threads”, “battery thread”, “battery use”, “battery rating”, “battery usage icon”, “battery icon”, “battery”
Bluetooth	“bluetooth connection”, “bluetooth speaker”, “bluetooth connect”, “bluetooth device”, “bluetooth connections”, “bluetooth settings”, “bluetooth icon”
Camera	“camera mode”, “photo mode”, “picture mode”, “camera modes”, “camera use”, “screen capture”, “camera interface”
Edge	“edge lighting”, “amoled displays”, “edge lighting style”, “amoled display”, “amoled panels”, “settings area”, “edge panel”
Fingerprint	“fingerprint icon”, “fingerprint sensor”, “fingerprint performance”, “fingerprint recognition”, “fingerprint scanner”, “display fingerprint”, “fingerprint scanning”
Protection	“plastic cover”, “glass protector”, “plastic backplate”, “plastic screen protectors”, “plastic sp”, “glass protectors”, “glass screen protector”

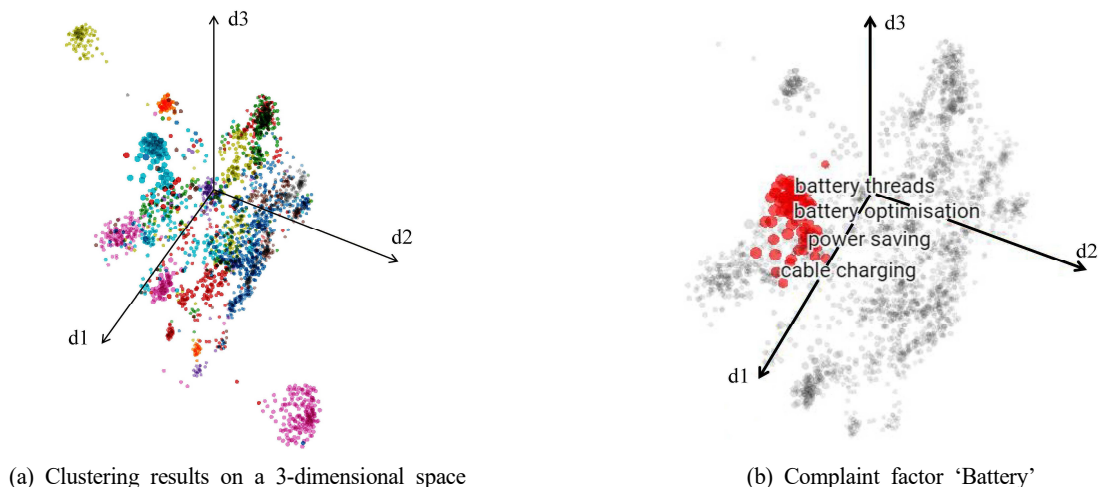


Figure 4. Clustering Using Embedding Vectors of Valid Noun Phrases

4.3 잠재 불만 요소 평가 결과

앞서 정의한 각 불만 요소의 불만 수준을 계량적으로 나타내기 위해 감성분석을 수행하였다. 우선 문장 단위의 감성 분석을 위해 앞서 선별한 13,976개의 유효 리뷰 데이터를 모두 문장 단위로 분할하여 총 54,664개의 문장 집합을 구성하였다. 본 연구는 감성분석 도구 VADER를 사용하여 모든 문장의 감성점수를 측정하였으며, 감성점수를 기반으로 각 문장의 감성(긍정, 중립, 불만)을 식별하였다. 불만 요소를 이루는 명사구가 출현하는 문장은 해당 불만 요소와 연관된 내용의 문장으로 식별할 수 있으므로, 결과적으로 각 불만 요소와 연관된 긍정, 중립, 불만 문장의 개수를 측정하였다.

각 불만 요소에 해당하는 문장의 개수를 활용하여 불만 요소의 두 가지 계량적 지표를 계산하였다. 첫 번째 계량적 지표인 Urgency는 식 (1)을 통해 계산할 수 있다. Urgency는 불만 요소와 연관된 모든 문장 중 불만 문장의 비율을 나타내는 지표이기 때문에, 해당 불만 요소에 대한 고객들의 불만이 실제로 얼마나 많이 발생했는지를 파악할 수 있다. 모든 불만 요소의 Urgency를 측정하고 결과 ‘Samsung pay’, ‘Battery’, ‘Fingerprint’, ‘SNS’, ‘Text and call’ 순서로 가장 높은 Urgency 값을 나타내는 불만 요소를 파악하였다. 다음으로는, 식 (2)를 통해 모든 불만 요소의 두 번째 계량적 지표 Severity를 측정하였다. Severity는 불만 요소와 연관된 불만 문장의 감성점수를 평균계산한 값으로써, 해당 불만 요소에 대한 고객들의 불만 수준이 얼마나 심각한지를 파악할 수 있다. 모든 불만 요소의 Severity를 측정하여 살펴본 결과 ‘Fingerprint’, ‘Protection’, ‘Samsung pay’, ‘Battery’, ‘Wide angle’의 순서로 가장 높은 Severity 값을 나타내는 불만 요소를 식별하였다(<Table 5> 참조).

4.4 고객 불만 포트폴리오 맵을 통한 불만 요소의 우선순위화

각 불만 요소의 Urgency, Severity 값이 모두 높다면, 고객 불만이 자주 발생함과 동시에 불만 수준이 매우 심각함을 의미하므로 해당 불만 요소는 가장 우선적으로 개선되어야 한다. 본 연구는 27개의 불만 요소 중 우선시되어야 할 요소들을 식별하기 위해, 우선 각 불만 요소의 Urgency, Severity 값을 매핑하여 고객 불만 포트폴리오 맵을 형성하였다(<Figure 5> 참조). 다음으로, 불만 요소가 속한 맵의 영역을 기반으로 불만 요소의 우선순위를 식별하였다. 고객 불만 포트폴리오 맵의 네 가지 영역은 Critical, Risky & Alert, Trivial 순서로 불만 요소의 개선에 대한 우선순위를 나타낸다.

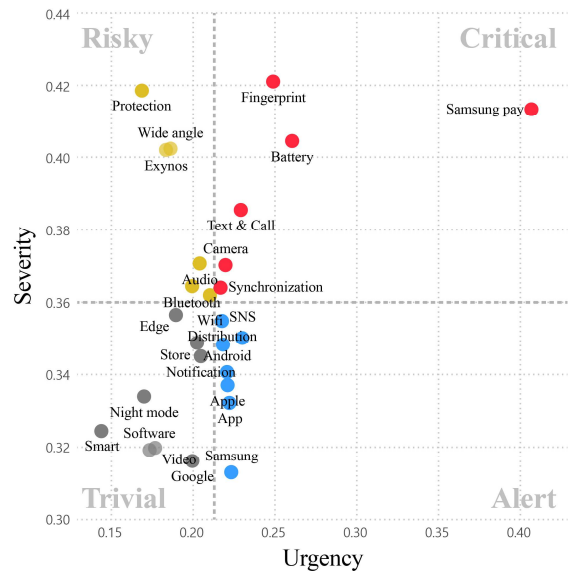


Figure 5. Customer-stated complaint portfolio map

Table 5. Urgency and Severity of Complaint Factors

Complaint	Urgency	Severity	Complaint	Urgency	Severity
Samsung pay	0.4069	0.4134	App	0.2225	0.3322
Battery	0.2608	0.4045	Exynos	0.1835	0.4021
Fingerprint	0.2490	0.4211	Bluetooth	0.1995	0.3645
Text & Call	0.2293	0.3856	Notification	0.2050	0.3451
Camera	0.2201	0.3703	Store	0.2027	0.3488
SNS	0.2302	0.3502	Protection	0.1689	0.4186
Synchronization	0.2169	0.3641	Samsung	0.2235	0.3131
Distribution	0.2180	0.3549	Edge	0.1898	0.3566
Wifi	0.2104	0.3621	Google	0.1998	0.3162
Swipe mode	0.2184	0.3483	Night mode	0.1702	0.3340
Audio	0.2043	0.3707	Video	0.1769	0.3199
Android	0.2210	0.3407	Software	0.1736	0.3193
Wide angle	0.1863	0.4024	Smart	0.1442	0.3245
Apple	0.2212	0.3371	Average	0.2131	0.3600

본 연구는 고객 불만 포트폴리오 맵의 4개 영역 중 가장 긴급한 불만 요소를 포함하고 있는 Critical 영역을 살펴본다. Critical 영역에는 6개의 불만 요소가 포함되었다('Battery', 'Camera', 'Fingerprint', 'Samsung pay', 'Synchronization', 'Text & Call'). 해당 요소들은 Urgency, Severity 두 지표 모두 높은 값을 나타내므로, 발생 빈도가 높으며 심각한 불만 수준을 나타내는 요소이다. Critical 영역의 6개 불만 요소 중에서도 'Samsung pay', 'Fingerprint', 그리고 'Battery'는 특히 높은 Urgency, Severity 값을 나타냈다. 따라서 본 연구는 앞서 언급한 3개의 불만 요소를 위주로 살펴본다.

'Samsung pay'(Urgency : 0.4069, Severity : 0.4134)는 오프라인에서의 결제 단말기뿐만 아니라 온라인에서도 결제가 가능한 삼성의 모바일 결제 서비스인 삼성 페이를 의미한다. 삼성 페이는 MST(Magnetic secure transmission), NFC(Near field communication) 결제방식을 모두 지원하기 때문에 더욱 다양한 곳에서 사용할 수 있다. 불만 요소 'Samsung pay'의 주요 명사구를 살펴보면 'connection error', 'background loading failure', 'error messages' 등으로, 삼성 페이에 대한 불만 원인은 서비스 연결 및 로딩의 실패인 것으로 나타난다. 추가적으로 실제 고객 리뷰 중 'Samsung pay'의 주요 명사구를 포함하는 불만 문장을 살펴보면 "Im not sure why, but it says connection error when i stremot to open the samsung pay app.", "... I installed Samsung pay and it gives: Connection error, unable to connect to Samsung pay temporarily." 등과 같이 삼성 페이를 실행할 수 없는 연결 오류가 주된 불만 원인이다. 삼성 페이의 경우 결제 서비스를 제공하는 소프트웨어이므로, 업데이트를 통해 고객 불만에 대한 신속한 대응이 가능할 것으로 기대된다.

'Fingerprint'(Urgency : 0.2490, Severity : 0.4211)는 Galaxy S10의 지문인식기능과 관련된 불만 요소이다. Galaxy S10의 지문인식은 온스크린 방식으로 디스플레이 내장형 지문인식 기술 중 광학 방식이 아닌 초음파 방식을 적용하였다. 불만 요소 'Fingerprint'의 주요 명사구는 'fingerprint sensor', 'fingerprint performance', 'fingerprint recognition' 등이며, 주요 불만 문장은 "I always find myself too low or too high for the fingerprint sensor to read (or upside down).", "If it's even slightly squashed, the fingerprint recognition will fail." 등과 같다. 지문 인식 장치의 위치가 사용자에게 불편하고 인식 장치가 사용자의 지문을 쉽게 인식하지 않는다는 의견이 많았다. Galaxy S10의 이전 시리즈인 스마트폰 Galaxy S9의 경우 지문 인식 장치가 내장되지 않고 스마트폰 기기 뒷면의 상단에 위치하였다. 사용자가 스마트폰을 한 손으로 쥐었을 때, 손가락을 펴서 쉽게 지문 인식 기능을 사용할 수 있었고, 내장형이 아니기 때문에 지문을 인식하는 속도가 매우 빠른 편이었다. 이에 비해 스마트폰 기기 앞면의 하단에 지문 인식 장치가 내장된 Galaxy S10은 사용하기에도 불편하고 지문 인식률이 매우 낮다는 의견이 불만의 원인이었다. 고객 불만을 해결하기 위해 2019년 4월에 Galaxy S10의 낮은 지문 인식률 관련 업데이트가 진행되었다. 하지만 이후 2019년 10월 영국에서는 등록되지 않은 지문이라도 모두 인식되는 현상이 발견되어 뉴스 기사로 보도되었다(TheSun,

2019). 만약 스마트폰 사용자의 지문이 아니어도 지문인식이 가능하다면, 이는 개인정보, 금융 등을 포함하는 심각한 모바일 범죄 피해를 불러올 수 있다. 이러한 이유로 불만 요소 'Fingerprint'는 'Samsung pay'에 비하여 발생 빈도가 낮더라도 매우 심각한 불만 수준을 지니는 것으로 판단된다.

'Battery'(Urgency : 0.2608, Severity : 0.4045)는 Galaxy S10의 배터리와 관련된 불만 요소이다. 주요 명사구는 'battery threads', 'battery use', 'battery' 등이며, 주요 불만 문장은 "It's a good cpu but a bit disappointed with the battery life.", "My battery life has also been less than expected unfortunately." 등과 같다. 주로 배터리의 수명 시간이 기대했던 것에 비하여 매우 짧다는 의견이다. 삼성은 Galaxy S10의 모델 중 'Galaxy S10 5G'를 출시하며 세계 최초로 5G 스마트폰을 선보였고, 이로써 삼성은 5G 네트워크의 상용화를 선도하였다. 하지만 당시에는 상용화 첫 단계였던 만큼 5G 네트워크의 연결이 원활하지 않았으며, 5G 네트워크와 LTE 네트워크를 빈번히 바뀌가며 사용하게 되는 문제로 인해 배터리가 많이 소모되었다. 또한, 불만 문장이 포함된 리뷰를 살펴보면 "...the main thing that hold me back from buying it is the processor that is Exynos. It's a good cpu but a bit disappointed with the battery life...", "I'm a bit worried for the battery life on the exynos model.. Pls help."처럼 배터리와 엑시노스(Exynos)가 함께 언급되는 경우가 많다. 엑시노스는 삼성의 애플리케이션 프로세서(Application processor, AP)이다. AP는 스마트폰, 태블릿 컴퓨터 등의 이동통신 단말기에 탑재되는 프로세서로 CPU, GPU, 통신 기능 등 다양한 기능을 탑재한 하나의 칩이며 대표적인 모바일 AP는 삼성의 엑시노스, 퀄컴의 스냅드래곤 등이 있다. 북미, 중국, 일본에 출시된 Galaxy S10의 AP는 스냅드래곤이 탑재되었고, 한국을 포함한 기타지역에 출시된 Galaxy S10은 엑시노스 AP를 탑재하였다. Reddit을 포함한 많은 소셜 커뮤니티에서 고객들은 국가 및 지역별로 다른 AP를 탑재한 것, 엑시노스 탑재 모델에 비하여 스냅드래곤 탑재 모델의 배터리 성능이 더 뛰어난 것에 대하여 많은 불만을 표출하였다. Galaxy S10의 배터리는 고객이 위치한 지역에서의 원활한 5G 네트워크 사용 여부와 사용하는 스마트폰에 탑재된 AP에 따라 불만 정도가 다를 수 있다. 이러한 이유로 불만요소 'Battery'는 Urgency 값에 비해 Severity 값이 높은 것으로 판단된다. 분석 결과에 따르면 빠르게 대응할 수 있는 소프트웨어 관련 업데이트가 우선적으로 요구되므로 배터리 소모 시간에 영향을 미치는 요소를 파악하여 연관된 소프트웨어를 개선해야 한다. 추가적으로 이후에 출시할 스마트폰의 배터리 부품에 대한 연구개발이 요구되며, 고객들이 불만을 갖는 엑시노스를 계속 탑재할 것인지 결정해야 할 것이다. 실제로 삼성은 Galaxy S10의 다음 시리즈인 Galaxy S20에 스냅드래곤만 탑재하고 엑시노스를 제외시키며 개선하였다.

5. 결론

본 연구는 제품 출시 초기 단계에서의 고객 불만 요소를 분석하기

위한 소셜미디어 마이닝 방법론을 제시하였으며, 제시한 방법론에서는 클러스터링 알고리즘과 감성분석이 사용되었다. 우선, 잠재 불만 요소를 도출하기 위해 소셜미디어의 제품 리뷰 데이터에 클러스터링 알고리즘을 적용하였다. 다음으로, 제품 리뷰 데이터를 문장 단위로 모두 분할하여 각 문장의 감성을 식별하였고, 앞서 도출된 불만 요소의 불만 수준을 나타내기 위한 계량적 지표 Urgency, Severity를 정의하였다. Urgency는 불만 요소에 포함되는 전체 문장 개수 대비 불만 문장 개수의 비율로 나타냈다. 그리고 Severity는 불만 요소에 포함된 문장 중 불만 문장의 감성점수를 종합하여 나타냈다. 이때 불만 문장의 감성 점수는 음수이므로 절대값의 평균으로 계산하였다. 최종적으로, Urgency와 Severity를 축으로 하는 고객 불만 포트폴리오 맵을 형성하여 불만 요소를 우선순위화 하였고, 가장 개선이 시급한 불만 요소 ‘Samsung pay’, ‘Fingerprint’, ‘Battery’를 식별하였다. 또한, 식별된 불만 요소 3개에 대한 불만 문장을 살펴봄으로써 각 불만 요소에 대한 자세한 고객 의견을 확인할 수 있었다.

본 연구는 다음과 같은 기여점을 가질 것으로 기대된다. 먼저, 본 연구는 소셜미디어 데이터의 이해를 바탕으로 한 소셜미디어 마이닝 기반의 고객 불만 모니터링 접근법을 제시한다. 초기 불만 데이터를 선별하기 위해 제품 출시일을 고려하여 데이터를 수집하였고, 다양한 주제 및 감성이 나타날 수 있는 제품 리뷰 문서를 문장 단위로 분할하여 감성분석을 수행하였다. 이와 같은 접근방식은 제품 리뷰를 활용하여 정확한 고객 이해를 할 수 있도록 도울 수 있다. 다음으로, 본 연구는 리뷰 데이터로부터 고객의 불만 수준을 계량적으로 측정할 수 있는 지표 Urgency와 Severity를 개발하였다. 두 가지 지표는 고객 리뷰 데이터를 모니터링하는 과정에 있어, 발생한 모든 불만 요소들을 대상으로 정량적 평가 및 우선순위화 하는 작업을 지원할 수 있다. 또한, 본 연구의 접근법은 전문가의 개입이 최소화되어 있기 때문에 실제 비즈니스 환경에서의 고객 불만 모니터링에 활용될 수 있다. 제안된 접근법은 분석 목적이나 데이터에 따라서 적절한 결과를 제공할 수 있다. 즉, 식별된 불만 요소, 불만 요소의 정량적 평가 결과, 그리고 고객 불만 포트폴리오 맵은 다양한 비즈니스 환경에서 재현될 수 있다.

소셜미디어의 리뷰 데이터를 활용하는 연구분야는 많은 연구가 지속적으로 수행되어 발전하는 분야이므로, 본 연구는 개선될 수 있는 부분이 존재한다. 우선, 본 연구는 불만 요소를 정의하기 위해 리뷰 데이터의 명사구를 활용하였다. 본 연구는 감성분석을 통해 문장의 감성을 식별하기 때문에, 명사구가 아닌 문장 자체를 사용하여 불만 요소를 정의할 수 있다면 문장의 감성을 더 직관적으로 연구에 사용할 수 있을 것이다. 다음으로, 본 연구는 불만 요소의 불만 수준을 나타내기 위한 지표로 Urgency, Severity를 사용한다. 두 가지 지표는 불만 문장의 개수 및 감성 점수를 활용하여 계산되었다. 추후에 텍스트 데이터 말고도 고객 평점, 이모티콘 등의 추가적인 요소들도 종합하는 고도화된 지표를 사용한다면 더욱 정확한 결과를 얻을 수 있을 것이다.

참고문헌

- Aguwa, C., Olya, M. H., and Monplaisir, L. (2017), Modeling of Fuzzy-based Voice of Customer for Business Decision Analytics, *Knowledge-Based Systems*, **125**, 136-145.
- Chang, V., Liu, L., Xu, Q., Li, T., and Hsu, C. H. (2020), An Improved Model for Sentiment Analysis on Luxury Hotel Review, *Expert Systems*, e12580.
- Chang, Y.-C., Ku, C.-H., and Chen, C.-H. (2019), Social Media Analytics : Extracting and Visualizing Hilton hotel Ratings and Reviews from TripAdvisor, *International Journal of Information Management*, **48**, 263-279.
- Choi, J., Oh, S., Yoon, J., Lee, J.-M., and Coh, B.-Y. (2020), Identification of Time-Evolving Product Opportunities Via Social Media Mining, *Technological Forecasting and Social Change*, **156**, 120045.
- Choi, J., Yoon, J., Chung, J., Coh, B.-Y., and Lee, J.-M. (2020), Social Media Analytics and Business Intelligence Research : A Systematic Review, *Information Processing & Management*, **57**(6), 102279.
- Chong, A. Y. L., Li, B., Ngai, E. W., Ch'ng, E., and Lee, F. (2016), Predicting Online Product Sales Via Online Reviews, Sentiments, and Promotion Strategies, *International Journal of Operations & Production Management*, **36**(4), 358-383.
- Deljac, Ž., Randić, M., and Krčelić, G. (2015), Early Detection of Network Element Outages based on Customer Trouble Calls, *Decision Support Systems*, **73**, 57-73.
- El-Diraby, T., Shalaby, A., and Hosseini, M. (2019), Linking Social, Semantic and Sentiment Analyses to Support Modeling Transit Customers' Satisfaction : Towards Formal Study of Opinion Dynamics, *Sustainable Cities and Society*, **49**, 101578.
- Herrmann, A., Huber, F., and Braunstein, C. (2000), Market-Driven Product and Service Design : Bridging the Gap between Customer Needs, Quality Management, and Customer Satisfaction, *International Journal of Production Economics*, **66**(1), 77-96.
- Hootsuite (2020), <https://blog.hootsuite.com/social-media-statistics-for-social-media-managers/>.
- Hu, N., Zhang, T., Gao, B., and Bose, I. (2019), What do Hotel Customers Complain about? Text Analysis Using Structural Topic Model, *Tourism Management*, **72**, 417-426.
- Hutto, C. and Gilbert, E. (2014), Vader : A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, *Paper Presented at the Proceedings of the International AAAI Conference on Web and Social Media*.
- Jeong, B., Yoon, J., and Lee, J.-M. (2019), Social Media Mining for Product Planning : A Product Opportunity Mining Approach based on Topic Modeling and Sentiment Analysis, *International Journal of Information Management*, **48**, 280-290.
- Joung, J., Jung, K., Ko, S., and Kim, K. (2019), Customer Complaints Analysis Using Text Mining and Outcome-Driven Innovation Method for Market-Oriented Product Development, *Sustainability*, **11**(1), 40.
- Kang, D. and Park, Y. (2014), Review-based Measurement of Customer Satisfaction in Mobile Service : Sentiment Analysis and VIKOR Approach, *Expert Systems with Applications*, **41**(4), 1041-1050.
- Kaplan, A. M. and Haenlein, M. (2010), Users of the World, Unite! The Challenges and Opportunities of Social Media, *Business Horizons*, **53**(1), 59-68.
- Kim, Y.-S. and Kim, M. (2019), 'A Wisdom of Crowds': Social Media Mining for Soccer Match Analysis, *IEEE Access*, **7**, 52634-52639.
- Liu, Y., Bi, J.-W., and Fan, Z.-P. (2017), Ranking Products through Online Reviews : A Method based on Sentiment Analysis Technique and Intuitionistic Fuzzy Set Theory, *Information Fusion*, **36**, 149-161.

- Loper, E. and Bird, S. (2002), NLTK : The Natural Language Toolkit, *arXiv preprint cs/0205028*.
- Lu, X. S. and Zhou, M. (2016), Analyzing the Evolution of Rare Events Via Social Media Data and k-Means Clustering Algorithm, *Paper Presented at the 2016 IEEE 13th International Conference on Networking, Sensing, and Control (ICNSC)*.
- Medhat, W., Hassan, A., and Korashy, H. (2014), Sentiment Analysis Algorithms and Applications : A Survey, *Ain Shams Engineering Journal*, **5**(4), 1093-1113.
- Micu, A., Micu, A. E., Geru, M., and Lixandroi, R. C. (2017), Analyzing User Sentiment in Social Media : Implications for Online Marketing Strategy, *Psychology & Marketing*, **34**(12), 1094-1100.
- Misuraca, M., Scepi, G., and Spano, M. (2020), A Network-based Concept Extraction for Managing Customer Requests in a Social Media Care Context, *International Journal of Information Management*, **51**, 101956.
- Phang, C. W., Kankanhalli, A., and Tan, B. C. (2015), What Motivates Contributors vs. Lurkers? An Investigation of Online Feedback Forums, *Information Systems Research*, **26**(4), 773-792.
- Qi, J., Zhang, Z., Jeon, S., and Zhou, Y. (2016), Mining Customer Requirements from Online Reviews : A Product Improvement Perspective, *Information & Management*, **53**(8), 951-963.
- Qian, J. (2011), Evaluating the Kindle DX e-Book Reader : Results from Amazon, Com Customer Reviews, *Performance Measurement and Metrics*, **12**(2), 95-105.
- Rathore, A. K. and Ilavarasan, P. V. (2020), Pre-and Post-Launch Emotions in New Product Development : Insights from Twitter Analytics of Three Products, *International Journal of Information Management*, **50**, 111-127.
- Reimers, N. and Gurevych, I. (2019), Sentence-Bert : Sentence Embeddings Using Siamese Bert-Networks, *arXiv preprint arXiv:1908.10084*.
- Roch, J. and Mosconi, E. (2016), The Use of Social Media Tools in the Product Life Cycle Phases : A Systematic Literature Review, *Paper presented at the 2016 49th Hawaii International Conference on System Sciences (HICSS)*.
- Salehan, M. and Kim, D. J. (2016), Predicting the performance of Online Consumer Reviews : A Sentiment Mining Approach to Big Data Analytics, *Decision Support Systems*, **81**, 30-40.
- Schuckert, M., Liu, X., and Law, R. (2015), A Segmentation of Online Reviews by Language Groups : How English and Non-English Speakers Rate Hotels Differently, *International Journal of Hospitality Management*, **48**, 143-149.
- Sezgen, E., Mason, K. J., and Mayer, R. (2019), Voice of Airline Passenger : A Text Mining Approach to Understand Customer Satisfaction, *Journal of Air Transport Management*, **77**, 65-74.
- Song, B., Lee, C., Yoon, B., and Park, Y. (2016), Diagnosing Service Quality Using Customer Reviews : An Index Approach based on Sentiment and Gap Analyses, *Service Business*, **10**(4), 775-798.
- Stevens, J. L., Spaid, B. I., Breazeale, M., and Jones, C. L. E. (2018), Timeliness, Transparency, and Trust : A Framework for Managing Online Customer Complaints, *Business Horizons*, **61**(3), 375-384.
- Stieglitz, S., Mirbabaie, M., Ross, B., and Neuberger, C. (2018), Social Media Analytics—Challenges in Topic Discovery, Data Collection, and Data Preparation, *International Journal of Information Management*, **39**, 156-168.
- Tang, X. and Yun, H. (2008), Data Model for Quality in Product Lifecycle, *Computers in Industry*, **59**(2-3), 167-179.
- TheSun (2019), <https://www.thesun.co.uk/tech/10127908/samsung-galaxy-s10-screen-protector-ebay/amp/>.
- Tibshirani, R., Walthers, G., and Hastie, T. (2001), Estimating the Number of Clusters in a Data Set Via The Gap Statistic, *Journal of the Royal Statistical Society: Series B(Statistical Methodology)*, **63**(2), 411-423.
- Trappey, A. J., Trappey, C. V., Fan, C.-Y., and Lee, I. J. (2018), Consumer Driven Product Technology Function Deployment Using Social Media and Patent Mining, *Advanced Engineering Informatics*, **36**, 120-129.
- Viglia, G., Minazzi, R., and Buhalis, D. (2016), The Influence of e-word-of-mouth on Hotel Occupancy Rate, *International Journal of Contemporary Hospitality Management*, **28**(9), 2035-2051.
- Wu, J.-J. and Chang, S.-T. (2020), Exploring Customer Sentiment Regarding Online Retail Services : A Topic-based approach, *Journal of Retailing and Consumer Services*, **55**, 102145.
- Xiao, S., Wei, C.-P., and Dong, M. (2016), Crowd Intelligence : Analyzing Online Product Reviews for Preference Measurement, *Information & Management*, **53**(2), 169-182.
- Xie, H., Li, Q., Mao, X., Li, X., Cai, Y., and Zheng, Q. (2014), Mining Latent User Community for Tag-based and Content-based Search in Social Media, *The Computer Journal*, **57**(9), 1415-1430.
- Xing, K., Hu, C., Yu, J., Cheng, X., and Zhang, F. (2017), Mutual Privacy Preserving k-means Clustering in Social Participatory Sensing, *IEEE Transactions on Industrial Informatics*, **13**(4), 2066-2076.
- Xu, X., Wang, X., Li, Y., and Haghighi, M. (2017), Business Intelligence in Online Customer Textual Reviews : Understanding Consumer Perceptions and Influential Factors, *International Journal of Information Management*, **37**(6), 673-683.
- Yoo, S., Song, J., and Jeong, O. (2018), Social Media Contents based Sentiment Analysis and Prediction System, *Expert Systems with Applications*, **105**, 102-111.
- Zaman, N., Goldberg, D. M., Abrahams, A. S., and Essig, R. A. (2020), Facebook Hospital Reviews : Automated Service Quality Detection and Relationships with Patient Satisfaction, *Decision Sciences*, <https://doi.org/10.1111/deci.12479>.

저자소개

이승현 : 건국대학교 산업공학과에서 2021년 학사학위를 취득하고 건국대학교 산업공학과 석사과정에 재학중이다. 주요 연구관심분야는 Social media mining for business opportunities, Data-driven prognostics and health management이다.

최재용 : 건국대학교 산업공학과에서 2018년 학사학위를 취득하고 건국대학교 산업공학과 석박사통합과정에 재학중이다. 주요 연구관심분야는 Patent intelligence for technology planning, Social media mining for product planning, Machine learning-based decision support systems이다.

윤장혁 : POSTECH 산업공학과에서 학사, 석사 학위를 취득한 후, LG CNS에서 4년 간 재직하였으며, POSTECH 산업경영공학과에서 박사 학위를 취득하였다. 한국지식재산연구원을 거쳐 현재는 건국대학교 산업공학과 정교수로 재직 중이다. 주요 연구분야는 대량 데이터 분석 기반의 Business/technology opportunity identification, Technology (e)valuation, Customer analysis/understanding, Data-driven prognostics and health management이다.

Appendix Table 1. Detailed Description of the Defined Complaint Factors¹⁻

Complaint	Description	1st	2nd	3rd	4th	5th	6th	7th
Samsung pay	Samsung pay connection error	connection error	background loading failure	error messages	camera lags	screen accidental touches	services drain	quality loss
Battery	Battery usage time	battery threads	battery thread	battery use	battery rating	battery usage icon	battery icon	battery
Fingerprint	Fingerprint recognition	fingerprint icon	fingerprint sensor	fingerprint performance	fingerprint recognition	fingerprint scanner	display fingerprint	fingerprint scanning
Text & Call	Text and call notification	text notifications	calls texts	calls messages	texts calls	text verification	messages app	message notifications
Camera	Camera mode switching	camera mode	photo mode	picture mode	camera modes	camera use	screen capture	camera interface
SNS	Social networking service	instagram app	instagram story	instagram notifications	facebook app	facebook notifications	instagram livestream	instagram mode
Synchronization	Auto synchronization	auto sync	auto mod	auto copy	auto rotate	auto syncing	auto reconnect	auto connect
Distribution	Product distribution	delivery time	release time	evaluation times	release date	setup time	standby time	advance exchange
Wifi	Wifi connectivity	wifi signals	wifi signal	wifi settings	wifi connections	wifi usage	wifi status	wifi commands
Swipe mode	Swipe mode	swipe controls	developer settings	swipe mode	center swipe	bixby commands	amoled themes	bixby button
Audio	Audio sound	sound mode	audio output	audio sync	audio controls	audio option	sound settings	sound device selector
Android	Android auto app	android apps	android users	galaxy android users	android software	android user	android devices	android auto app
Wide angle	Wide angle camera	wide angle	wide angle mode	wide lens	wide angle lens	wide font	wide angle lense	wide camera
Apple	Apple's device compatibility	apple products	apple music	apple pay	apple perspective	apple watch	apple's unit	apple airpods
App	Application setting	app icon	app connections	app icons	app launchers	app launcher	app setting	app photo
Exynos	Exyns	exynos model	exynos variant	exynos devices	exynos versions	exynos version	exynos	exynos variations
Bluetooth	Bluetooth connection	bluetooth connection	bluetooth speaker	bluetooth connect	bluetooth device	bluetooth connections	bluetooth settings	bluetooth icon
Notification	Notification error	configuration notification	notification access	notification panel	notification titles	notification content	notification style	notification options
Store	Galaxy store	store collection	store model	store inventory	store warranty	store credit	retail boxes	line store
Protection	Protection material	plastic cover	glass protector	plastic backplate	plastic screen protectors	plastic sp	glass protectors	glass screen protector
Samsung	Samsung account	samsung account	samsung com	samsung service	samsung's system	samsung tech	samsung site	samsung messages
Edge	Edge display	edge lighting	amoled displays	edge lighting style	amoled display	amoled panels	settings area	edge panel
Google	Google compatibility	google search	google lens	google account	google location	google cast	google store	google drive
Night mode	Night mode	night cam	night mode	night mode camera	night pictures	night photos	night sight	account night modes
Video	Video capability	video recording	video messages	video mode	video calls	video recording mode	video files	video enhancer
Software	Software glitch	software info	software information	firmware version	software carrier	device folders	device manager	software locks
Smart	Smart convenience feature	moveable ones	safe mode	smart alert	usefulness review	smart view	smart switch	smart select