

기계학습 기반 디지털 제품의 온라인 리뷰 분석을 위한 사용자 경험(UX) 모델링 프레임워크

자밀 후세인(Jamil Hussain) 경희대학교 소프트웨어융합학과 연구박사

이승룡 경희대학교 소프트웨어융합학과 교수

1. 머리말

오늘날의 고객은 긍정적이고 사용자 주도적인 경험을 제공하는 대화형 제품을 선호한다. 많은 제품 디자이너와 기업들은 긍정적 사용자 경험(UX, User eXperience)을 제공하는 제품이야말로 성공적인 제품이라고 믿는다[1]. 긍정적인 UX를 만드는 데는 사용자 만족도, 사용 상황, 제품 품질, 사용의 즐거움, 사용 편의성 등 많은 요소를 고려해야 한다. 따라서, 대상 제품, 시스템 또는 서비스 사용자와 고객의 UX를 이해하는 것은 필수적이며, 이를 위하여 많은 연구에서 UX를 평가하는 다양한 방법을 제시하고 있다. 그러나 UX는 본질적으로 주관적이고 역동적이며 상황에 따라 다르기 때문에 UX 평가를 위한 범용적인 방법은 없다. 즉, UX는 사용자가 제품, 시스템 또는 서비스에 대해 느끼는 것으로 경우에 따라 모두 다르다[2].

기존 연구들은 다양한 시나리오를 바탕으로 특정 장소나 실험실에서 설문이나 Report

Grand Techniques(RGT)와 같은 방법을 사용하여 UX를 평가하였다[3][4]. 각각의 시나리오에는 참여자 선택, 할 일 설정, 사용 상황 정의, UX 평가 방법 및 훈련법 선택, 샘플 데이터 수집 비용 등을 포함시킨다. 이러한 설정들은 사용자/고객 경험 데이터를 수집하는 데 필수적이긴 하지만 제약이 있어서 사용자가 제품, 시스템 또는 서비스에 대해 느끼는 반응이 편향되거나 왜곡될 수 있다. 또한 높은 비용으로 인하여 실험 진행이 어려워 설계자 본인이 가진 주관이나 편향에 따라 실험이 진행될 수도 있다.

이같은 문제점을 해결하기 위하여 연구[5][6][7][8]에서는 온라인 제품 분산 플랫폼에서 사용자 제작 콘텐츠(UGC, User-Generated Content) 형태로 사용성과 UX 정보를 추출하였다. UGC는 UX 측면(UX Aspect) 추출을 위해서 중요하며, 제품 감성, 기능 요청, 버그 리포트, 제안 및 불만 사항 등과 같은 관련된 정보를 추출한다. 이러한 플랫폼을 바탕으로 고객들은 사용자 리뷰 형태로 제품 사용 경험, 유용한

제안 사항, 의견 등을 내놓는다[8]. 사용자 리뷰는 언제 어디서나 무료로 접근할 수 있는 자발적이고 통찰력 있는 피드백으로, 제품에 대한 요약 평가일 뿐만 아니라 실제로 느낀 사용자 경험에 대한 자체보고 데이터이다. 기존 연구에 의하면 정보 추출을 위해 활용되는 사용자 리뷰는 제품을 사용하는 동안에 사용자가 느끼는 선호도를 파악할 수 있다고 밝혔다.

수많은 이용자로부터 얻어지는 사용자 리뷰는 여러 상황에서 다양한 사용 의견을 포함하고 있다. 예를들면, 동일한 제품이라 할지라도 사용자마다 체험하는 성능과 경험이 다를 수 있기 때문에 다양한 사용자 경험은 신제품을 개발할 때 중요한 요소이다. 사용자 리뷰는 보통 텍스트와 평점으로 구성된다. 기존 연구들은 주로 온라인 평점을 이용하는데 이는 단순한 숫자 형태의 데이터로 실제적인 맥락이 제대로 반영되지 않는다. 반면, 글로 작성된 사용자 리뷰 텍스트는 기능 요청, 버그 리포트 등과 같은 귀중한 정보를 담은 경우가 많아 제품을 개선하는 데 도움이 된다. 그러나 텍스트는 비정형 형태의 자연 언어로 되어 있기 때문에 텍스트 마이닝 기술을 적용하여 광범위한 UGC 데이터에서 UX 통찰력을 도출하는 방법이 필요하다. 감정 분석 기술은 사용자 리뷰에서 제품에 대한 의견을 찾기 위해 사용되지만[9] 이를 바탕으로 UX 정보를 추출하는 것은 제한이 있다[5].

온라인 리뷰에서 UX를 파악하기 위한 연구는 (1) UX 측면이나 UX 차원(UXD, UX Dimension) 마이닝[7]과, (2) UX 모델링의 두 범주로 구분할 수 있다. 첫 번째 범주의 연구에서는 확률 토픽 모델(Probabilistic Topic Model)인 Latent Dirichlet Allocation (LDA)과 확률 잠재 의미 분석(PLSA, Probabilistic Latent

Semantic Analysis)[8], 워드 임베딩(Word Embedding), 속성 기반 감성 분석이 있다. 이는 추출된 각 UX 속성의 상대적 중요도 분석 등, 다양한 텍스트 마이닝과 기계학습 방법을 사용하여 UX 속성을 추출한다. 연구[8][10][11]에서는 토픽 모델링인 LDA를 사용하여 UX 영역에서 추출된 차원 또는 속성을 검증하기 위해 평점을 매긴 데이터에 회귀 분석을 적용하여 잠재적인 차원을 추출한다. 자연 언어 처리와 기계학습을 이용하여 온라인 평점과 리뷰 내용을 종합하여 분석하면 추출된 UX 속성과 사용자 만족도 사이의 인과관계를 파악할 수 있다. 사전 지식이 없는 경우에 비지도(unsupervised) 모델은 의미론적으로 일관성이 없어서 이해하기 어려운 토픽을 생성할 수 있다[12]. 일부 연구에서는 비지도 모델의 문제점을 해결하고자 토픽 모델링에 분야 지식을 추가하지만, 대부분의 모델은 이를 자동으로 수행할 수 없다. 최근의 연구들은 토픽 모델링에 워드 임베딩 등 다른 어휘 표현 체계를 사용하여 어휘의 지역적 맥락을 고려하고, 동시 발생 정보를 기반으로 어휘 벡터의 차원을 줄이고 있다. 또한, 전역과 지역 맥락을 결합하면 더 일관성 있는 토픽을 제공할 수 있다.

두 번째 범주인 UX 모델링 연구는 속성 수준의 감정 분석과 UX 모델링을 위해 반정형(semi-structured) 데이터와 베이저안 이론(Bayesian Theory)을 사용한 연구가 수행되었다[13]. 여기서는 각각의 속성과 제품 또는 서비스의 상대적 중요도에 따라 하나의 통일된 등급을 생성하기 위해 확률론적 방법을 사용하여 각각의 리뷰에 대해 감정과 제품 속성을 연관시켰다. 연구 [14]에서는 포아송(Poisson), 음의 이항(Negative Binomial), 잠재 집단 포아송(Latent Class Poisson) 회귀 모델을 사용하여 제품 속성에 대

한 사용자 감정이 사용자 만족도에 미치는 영향을 평가하였으며, 실험 결과 음의 이향 회귀 모델을 사용할 경우 가장 높게 영향력을 식별할 수 있는 것으로 판명하였다.

이처럼, 연구들은 각자의 장점을 가진 UX 모델링 방법을 제시하였지만, 리뷰 평점이 가우시안 분포(Gaussian Distribution)의 형태라고 가정하고 있다. 그러나 리뷰 평점은 일반적으로 “J” 모양의 비대칭 분포(J-shaped Asymmetric Distribution)를 형성하고 있다. 즉, 고객은 제품의 다양한 속성에 대해 여러가지 의견을 가지므로 다양한 UXD와 사용자 만족도 사이에 다면적인 연관성이 있을 수 있다. 따라서, 다양한 UXD와 사용자 만족도 간의 복잡한 관계를 처리할 수 있는 모델이 요구된다.

이같은 UX 연구들의 한계점들을 해결하기 위해 본고에서는 온라인 리뷰를 바탕으로 UX를 모델링하는 포괄적인 프레임워크를 제안한다. 제안 프레임워크는 먼저 UX 다기준 품질(UXMCQ, UX Multi-Criteria Quality) 필터를 이용하여 UX 분야와 관련 없는 리뷰를 필

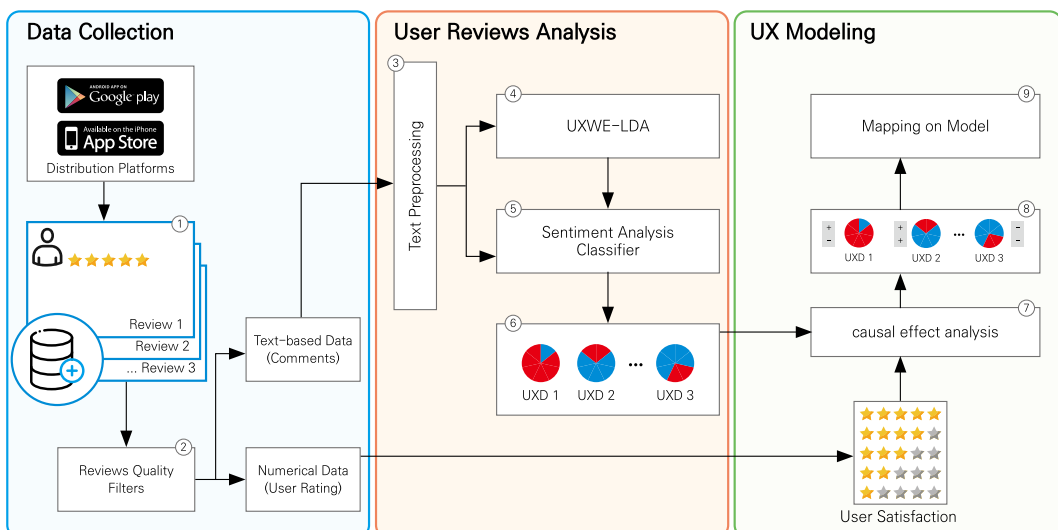
터링한다. 필터링된 리뷰로부터 제안하는 향상된 토픽 추출 방법인 UXWE-LDA(User Experience Word-Embedding-LDA)를 이용하여 UX 차원을 추출한다. 그 후 추출된 UX 차원을 고려하여 리뷰에 대한 감정 분석을 수행한다. 이를 통해 리뷰로부터 핵심적인 UX 속성을 추출할 수 있으며, 이는 제품을 개선하는데 사용된다. 또한 감정 분석 결과를 통해 카노 모델(Kano Model)에 기반하여 제품의 장단점을 알 수 있다. 카노 모델은 UEQ 모델[15]에 기반하여 UX 속성이 어느 범주에 속하는지를 나타내며, 이는 제품 개선에 필수적이다.

2. 제안하는 방법론

제안 방법은 데이터 수집, 사용자 리뷰 분석, UX 모델링의 3단계 과정으로 수행된다.

2.1 데이터 수집

첫 단계는 UX와 관련 있는 사용자 리뷰만을 선정하기 위한 것으로, UX 다기준 품질



[그림 1] 제안하는 UX 모델링 프레임워크 동작 과정

(UXMCQ, UX Multi-Criteria Quality) 필터를 사용하였다. 이는 딥 러닝을 활용하여 각 스팸 탐지, 관련성 파악, 주관성 파악의 세 가지 분류를 수행한다. 이 분류기들의 워드 임베딩을 위하여 BERT3(Bidirectional Encoder Representations from Transformers)와 양방향 LSTM 알고리즘을 사용하였다. 이들은 순차적으로 동작하여 먼저 스팸 리뷰를 제외하고, UX와 관련성이 높은 리뷰를 선택하며 마지막으로 주관적인 리뷰를 선택한다.

스팸 탐지 분류기는 사용자 리뷰가 진실성 있게 작성된 것인지 아니면 허위인지를 탐지한다. 온라인 리뷰는 악의적인 사용자가 특별한 검증이 없어도 사기성 정보를 자유롭게 기술할 수 있기 때문에 자주 악용되는 방법 중 하나이다. 그 결과, 온라인 리뷰 시스템은 개인 또는 전문 스팸 발송자의 대상이 되며, 평가와 리뷰의 내용을 조작하는 기만적인 행위를 수행한다. 스팸 탐지 분류기의 학습을 위하여 Deceptive opinion spam corpus v1.4를 사용했다[16].

관련성 파악 분류기는 리뷰 내용에 UX 관련 정보와 사용성에 대한 정보가 있는지 확인한다. 기준을 정하기 위하여 UX 관련 기사의 체계적인 리뷰 과정을 통해 수집된 3개의 UX 속성(사용자, 제품, 상황)[5]에 기반을 두어 부트스트랩 방법으로 자동 레이블링을 수행한다. 자동 레이블링은 레이블이 지정되지 않은 데이터의 속성 용어와 정확히 일치하는 용어에 대해 수행되며, 윈도우 크기 [+3, -3]으로 생성된 레이블 데이터는 관련성 파악 분류기 학습 데이터로 사용된다.

주관성 분류기는 리뷰가 사용자의 느낌을 쓴 주관적 리뷰인지 혹은 사실 내용만을 작성한 객관적 리뷰인지로 분류한다. 이를 위하여 연구 [17]에서 사용한 실험 데이터인 각각 5,000개의

주관적 리뷰와 객관적 리뷰 데이터를 사용하여 학습을 수행하였다.

2.2 사용자 온라인 리뷰 분석

사용자 리뷰 분석에서는 UX 차원 추출 및 추출된 각 UX 차원에 대한 감정의 지향성 파악을 수행한다.

UX 차원 추출은 제안하는 UXWE-LDA (User Experience Word-Embedding-LDA) 모델을 이용한다. UXWE-LDA는 토픽 모델링 방법 중 LDA와 워드 임베딩을 결합한 방법이다. 주어진 텍스트 말뭉치에서 분야 지식을 자동으로 학습하고, 사전 기반 접근법을 사용하여 추출된 각 주제에 레이블을 UXD로 할당하고자 좀 더 일관성 있는 주제를 추출한다.

이를 위해 먼저 GuidedLDA와 가이드된 종자 어휘(Guided seed word)를 사용하여 주제어와 종자 어휘를 선택한다. 그 후 CluWord[18] 방법과 비슷하게 사전 학습된 워드 임베딩 모델을 기반으로 종자 어휘를 이용한 유사성을 활용하여 의미론적이고 통사론적인 포괄적 전역 문맥을 생성한다. 이후 워드 임베딩과 유사성 탐지 기법인 간결 유사성(Concise similarity)과 PMI를 이용하여 분야 지식과 병합, 더 일관성 있는 UXD 추출을 위한 지역 문맥을 생성한다. 생성된 전역 및 지역 문맥을 이용하여 강화된 분야 지식을 자동으로 병합해서 토픽을 추출하기 위해 Gibbs Sampler 알고리즘을 사용한다. 추출된 토픽 중 상대적으로 중요한 n개의 어휘에만 레이블링을 수행하기 위해 자동 레이블링 알고리즘을 적용한다. 마지막으로, 감정의 지향성을 식별하기 위해 저자의 기존 제안 앙상블 기법[3][19]을 기반으로 감정 분석 모델을 학습하였다.

2.3 UX 모델링

UX 모델링은 각각의 UXD에 대한 감정 지향성에 따른 인과 관계 분석과 UXD 인과 관계를 사용자 만족도 모델에 매핑하는 과정을 수행한다.

각 UXD에 대한 사용자의 긍정적 또는 부정적 감정이 사용자 만족도에 미치는 영향을 측정하기 위해 양방향 LSTM 모델을 사용했다. 기존의 사용자 만족도 모델은 온라인 리뷰에 매겨진 평점이 모든 차원에 대한 감정을 선형 결합한 것으로 가정한다. 그러나 실제로는 다양한 감정이 복합적으로 결합되었을 수 있다. 양방향 LSTM 모델은 기존 만족도 모델에 사용된 가우시안 분포나 회귀 분석 등의 모델이 가진 문제점을 해결할 수 있다. 따라서 제안 방법에서는 양방향 LSTM 모델을 사용하여 각 UXD에 대한 감정 효과를 측정했다. 모델 학습에는 생성된 데이터와 함께 레이블 속성으로 각 리뷰의 평점을 사용하였다.

인과관계 분석 후에 카노 교수진[20]이 개발한 인과 관계를 기반으로 한 이차원 모델인 카노 모델(Kano Model)로 사용자 만족도를 표현하였다. 카노 모델은 사용자 만족도 품질을 당연(Must-be), 일원적(One-dimentional) 또는 성능(Performance), 매력(Attractive) 또는 흥분(Excitement), 무관심(Indifferent), 역(Reverse)으로 분류한다. 이 방법을 통해 제품 설계자는 UEQ 모델의 관점에서 UX 차원의 다

양한 범주를 이해할 수 있으며, 이는 제품 향상에 필수적이다. 본 연구에서는 연구 [11]의 내용과 같이 UXD의 인과 관계를 카노 모델로 매핑했다.

3. 실험 환경 및 결과

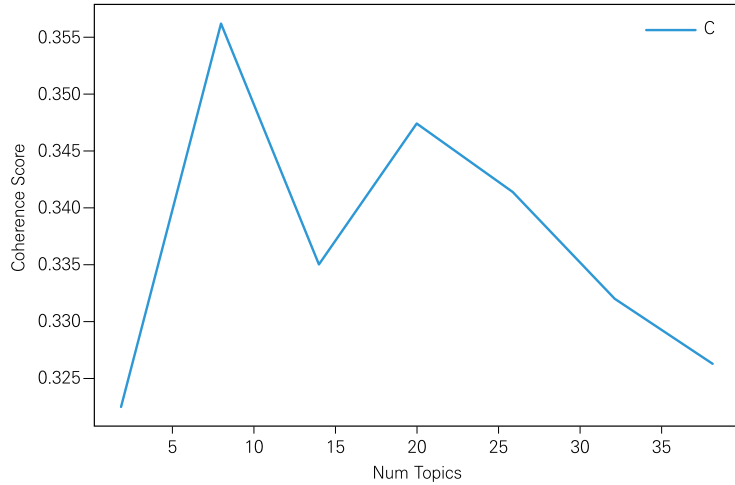
실험은 3가지 다른 방법이 수행되었다. 첫 번째 실험은 스팸 분류, 관련성 파악 분류, 주관성 분류, 감정 분류의 성능을 측정했다. 두 번째 실험은 토픽 일관성(TC, Topic Coherence)을 측정했다. 세 번째 실험은 전문가와 제안 방법과의 성능 비교 실험이다. 실험은 아마존 휴대폰 리뷰 데이터를 바탕으로 성능 측정이 이루어졌으며, 동시에 이를 이용하여 카노 모델에 기반하여 UX 모델을 생성하였다. <표 1>은 첫 번째 실험 결과를 나타낸다.

두 번째 실험은 UXWE-LDA의 매개변수인 상위 종자 어휘(Top seed word) 개수 n , 어휘 유사성 m , 신뢰 점수 u 에 대한 튜닝 실험이 수행되었다. 그 결과 [그림 2]와 같이 10개의 토픽을 가진 상위 30개 어휘는 높은 TC 점수를 나타내었다. 이는 너무 적거나 너무 많은 어휘 종자는 일관된 토픽을 생성할 수 없음을 나타낸다.

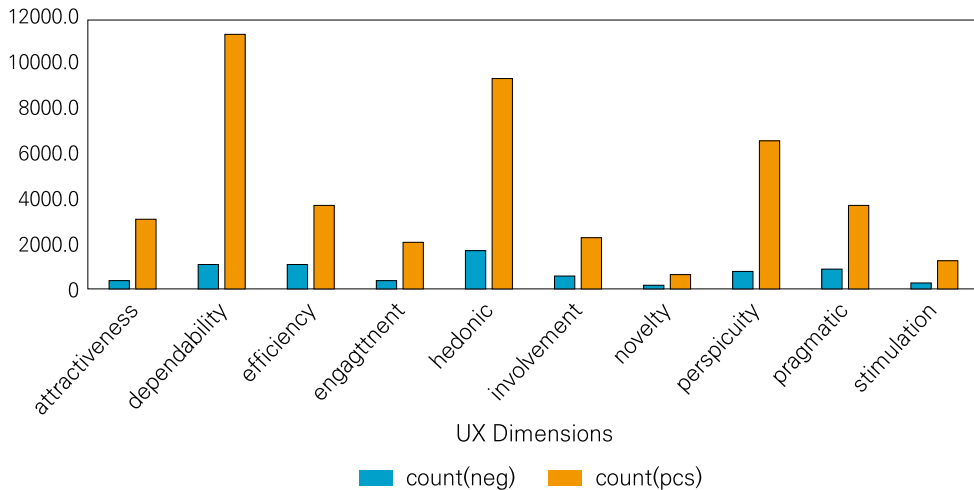
[그림 3]은 추출된 차원과 감정의 지향성을 나타낸다. 이를 통해 제품 설계자는 제품에 대한

<표 1> 제안하는 분류 방법의 성능

	정확도	정밀도	재현율	F1 점수
스팸 분류	98.49%	98.02%	99.00%	98.51%
관련성 파악 분류	97.96%	97.96%	98.36%	97.96%
주관성 분류	95.33%	96.16%	94.35%	95.25%
감정 분류	98.26%	94.57%	96.83%	95.69%



[그림 2] 토픽에 대한 일관성 점수

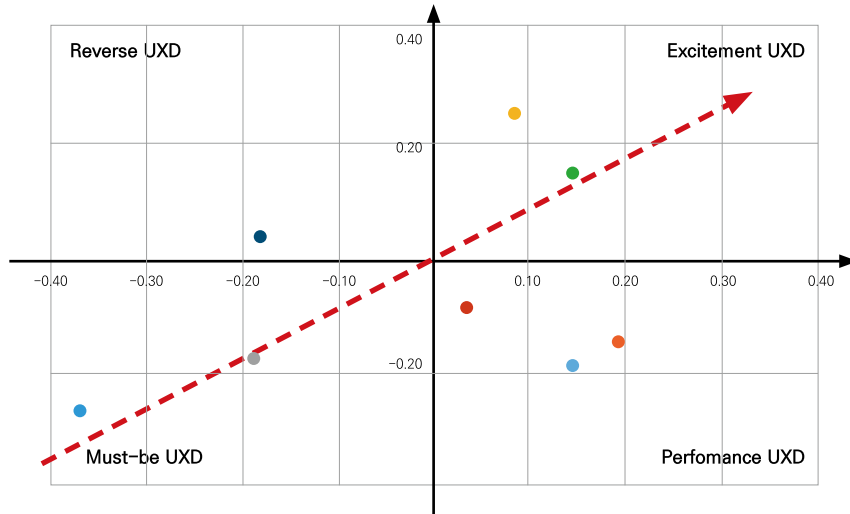


[그림 3] 온라인 리뷰에 대한 각 UXD의 감정의 지향성 결과

사용자의 피드백 정보를 얻고, 이를 바탕으로 제품을 향상시킬 수 있다. 마지막 실험은 제안하는 UXWE-LDA의 성능 측정을 위해 자연 언어 처리 및 텍스트 마이닝 전문가 2인과 시스템의 성능 비교 실험이 수행되었다. 성능 판단은 자카드 계수 유사도(Jaccard coefficient similarity)를 이용하여 UX 차원의 중복성 여부를 검사하였다. 그 결과 두 명의 전문가 및 시스템의 자카드 계수는 각각 0.3, 0.5, 0.4가 나왔다. 이는 시스

템이 전문가와 비교할 만한 성능임을 나타내며 제안 방법의 우수함을 입증한다.

[그림 4]는 인과 효과 모델에 의해 생성된 데이터를 기반으로 아마존 리뷰의 각 UXD 범주를 식별하고 사용자 만족도를 카노 모델에 매핑한 결과를 보여준다. 그 결과 흥분 UXD에는 쾌락과 명쾌함이, 역 UXD에는 실용성이, 당연 UXD에는 열중과 효율성이, 성능 UXD에는 자극, 매력, 의존이 포함된다.



[그림 4] 추출된 차원의 카노 모델로의 매핑

4. 맺음말

소셜 미디어의 발전으로 인하여 사용자는 매일 자신의 의견이 담긴 온라인 리뷰를 작성한다. 이러한 리뷰에는 UX와 관련된 유익한 정보가 포함돼 UX와 UX 모델링에 사용할 수 있다. 본 논문에서는 방대한 양의 온라인 리뷰에서 UX

관련 정보를 자동으로 마이닝하는 포괄적인 프레임워크를 제안하였다. 이를 위하여 UXMCQ로써 UX 분야와 관련된 리뷰 선정, UXWE-LDA를 사용한 UXD 추출, 그리고 사용자 만족도를 카노 모델로 매핑하였다. 실제 전문가와 제안 방법과의 비교 실험 결과 제안 방법이 전문가만큼의 수준을 낸다는 것을 입증하였다. TTA

참고문헌

- [1] F. Pucillo and G. Cascini, 'A framework for user experience, needs and affordances,' *Des. Stud.*, vol. 35, no. 2, pp. 160-179, 2014.
- [2] E. L.-C. Law, V. Roto, M. Hassenzahl, A. P. O. S. Vermeeren, and J. Kort, 'Understanding, scoping and defining user experience,' *Proc. 27th Int. Conf. Hum. factors Comput. Syst. - CHI 09*, p. 719, 2009.
- [3] J. Hussain et al., 'A multimodal deep log-based user experience (UX) platform for UX evaluation,' *Sensors (Switzerland)*, vol. 18, no. 5, 2018.
- [4] E. L. C. Law, P. Van Schaik, and V. Roto, 'Attitudes towards user experience (UX) measurement,' *Int. J. Hum. Comput. Stud.*, vol. 72, no. 6, pp. 526-541, 2014.
- [5] B. Yang, Y. Liu, Y. Liang, and M. Tang, 'Exploiting user experience from online customer reviews for product design,' *Int. J. Inf. Manage.*, vol. 46, no. October 2018, pp. 173-186, 2019.
- [6] S. Hedegaard and J. G. Simonsen, 'Extracting usability and user experience information from online user reviews,' *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '13*, p. 2089, 2013.
- [7] Y. Guo, S. J. Barnes, and Q. Jia, 'Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet??allocation,' *Tour. Manag.*, vol. 59, pp. 467-483, 2017.
- [8] H. Q. Vu, G. Li, R. Law, and Y. Zhang, 'Exploring Tourist Dining Preferences Based on Restaurant Reviews,' 2019.
- [9] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. R. Venugopal, 'Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier,' *World Wide Web*, vol. 20, no. 2, pp. 135-154, 2017.
- [10] F. Xu, L. La, F. Zhen, T. Lobsang, and C. Huang, 'A data-driven approach to guest experiences and satisfaction in sharing,' *J. Travel Tour. Mark.*, vol. 36, no. 4, pp. 484-496, 2019.
- [11] J. W. Bi, Y. Liu, Z. P. Fan, and E. Cambria, 'Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model,' *Int. J. Prod. Res.*, vol. 0, no. 0, pp. 1-21, 2019.
- [12] L. Yao, Y. Zhang, Q. Chen, H. Qian, B. Wei, and Z. Hu, 'Mining coherent topics in documents using word embeddings and large-scale text data,' in *Engineering Applications of Artificial Intelligence*, vol. 64, no. August, Elsevier Ltd, 2017, pp. 432-439.
- [13] M. Farhadloo, R. A. Patterson, and E. Rolland, 'Modeling customer satisfaction from unstructured data using a Bayesian approach,' *Decis. Support Syst.*, vol. 90, pp. 1-11, 2016.
- [14] R. Decker and M. Trusov, 'Estimating aggregate consumer preferences from online product reviews,' *Int. J. Res. Mark.*, vol. 27, no. 4, pp. 293-307, 2010.
- [15] M. Schrepp, 'User Experience Questionnaire Handbook,' pp. 1-11, 2015.
- [16] L. Vanderwende, H. Daumé III, and K. Kirchhoff, 'Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,' in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2013.
- [17] B. Pang and L. Lee, 'A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts,' 2004.
- [18] F. Viegas et al., 'CluWords: Exploiting Semantic Word Clustering Representation for En-hanced Topic Modeling,' pp. 753-761, 2019.
- [19] J. Khan, A. Alam, J. Hussain, and Y.-K. Lee, 'EnSWF: effective features extraction and selection in conjunction with ensemble learning methods for document sentiment classification,' *Appl. Intell.*, vol. 49, no. 8, pp. 3123-3145, 2019.
- [20] K. Matzler, F. Bailom, E. Sauerwein, and H. H. Hinterhuber, 'How to delight your customers,' *J. Prod. Brand Manag.*, vol. 5, no. 2, pp. 6-18, 1996.