

온라인 리뷰 분석을 통한 상품 평가 기준 추출: LDA 및 k-최근접 이웃 접근법을 활용하여*

이지현

한양대학교 철학과
(jihyun512@hanyang.ac.kr)

정상형

한양대학교 경영학부
(cukirhkwk@hanyang.ac.kr)

김준호

한양대학교 수학과
(junho9857@hanyang.ac.kr)

민은주

한양대학교 파이낸스경영학과
(mej0908@hanyang.ac.kr)

여운영

한양대학교 비즈니스인포매틱스학과
(ywy31730@hanyang.ac.kr)

김종우

한양대학교 경영대학 경영학부
(kjuw@hanyang.ac.kr)

.....

상품 평가 기준은 상품에 대한 속성, 가치 등을 표현한 지표로써 사용자나 기업이 상품을 측정하고 파악할 수 있게 한다. 기업이 자사 제품에 대한 객관적인 평가와 비교를 수행하기 위해서는 적절한 기준을 선정하는 것이 필수적이다. 이때, 평가 기준은 소비자들이 제품을 실제로 구매 및 사용 후 평가할 때 고려하는 제품의 특징을 반영하여야 한다. 그러나 기존에 사용되던 평가 기준은 제품마다 상이한 소비자의 의견을 반영하지 못하고 있다. 기존 연구에서는 소비자 의견이 반영된 온라인 리뷰를 통해 상품의 특징, 주제를 추출하고 이를 평가 기준으로 사용했다. 하지만 여전히 상품과 연관성이 낮은 평가 기준이 추출되거나 부적절한 단어가 정제되지 않는 한계가 있다.

본 연구에서는 이를 극복하기 위해 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 기법으로 리뷰로부터 평가 기준 후보군을 추출하고 이를 k-최근접 이웃 접근법(k-Nearest Neighbor Approach, k-NN)을 이용해 정제하는 모델을 개발하고 검증했다. 제시하는 방법은 준비 단계와 추출 단계로 이루어진다. 준비 단계에서는 워드 임베딩(Word Embedding) 모델과 평가 기준 후보군을 정제하기 위한 k-NN 분류기를 생성한다. 추출 단계에서는 k-NN 분류기와 언급 비율을 이용해 평가 기준 후보군을 정제하고 최종 결과를 도출한다.

제안 모델의 성능 평가를 위해 명사 빈도 추출 모델, LDA 빈도 추출 모델, 실제 전자상거래 사이트가 제공하는 평가 기준을 세 비교 모델로 선정했다. 세 모델과의 비교를 위해 설문을 진행하고 점수화하여 결과를 검증했다. 30번의 검증 결과 26번의 결과에서 제안 모델이 우수함을 확인했다. 본 연구의 제안 모델은 전자상거래 사이트에서 리뷰 특성을 반영한 상품군 별 차원을 도출하는데 활용될 수 있고 이를 기초로 인사이트 발굴을 위한 리뷰 분석 및 활용에 크게 기여할 것이다.

주제어 : 상품 평가 기준, 리뷰 분석, 평가기준 추출, LDA, k-NN

.....

논문접수일 : 2019년 8월 19일 논문수정일 : 2020년 2월 5일 게재확정일 : 2020년 2월 28일
원고유형 : 일반논문 교신저자 : 김종우

* 이 논문은 ‘비즈니스 랩 기반의 빅인텔리전스 경영교육 사업단(CK2)’의 지원을 받아서 수행된 연구임.

1. 서론

Web 2.0 시대의 특징은 소비자가 온라인으로 정보를 얻는 동시에 여러 소셜 플랫폼을 통해 자신의 의견을 생산한다는 것이다. 그 중 대표적인 것이 바로 온라인 쇼핑몰의 상품 리뷰이다. 리뷰는 소비자들의 구매 의사결정에 중요한 요소이며 이커머스 시장의 고성장과 함께 그 양도 방대해지고 있다. 한편, 리뷰는 고객의 잠재된 니즈를 담고 있다는 점에서 기업의 마켓 센싱 측면에서도 중요한 자원이다. 기업은 리뷰로부터 제품 개선 방향에 대한 아이디어를 얻고(Cha et al. 2015) 인사이트를 발굴, 비즈니스적 가치를 창출한다(Jin et al. 2016). 이렇듯 기업 입장에서도 리뷰 활용의 중요성이 커지면서 온라인 리뷰를 분석하는 다양한 기법들이 연구되고 있다.

감성분석 기법인 오피니언 마이닝(Opinion Mining)은 소비자가 상품에 대해 긍정적인지 부정적인지를 분석해준다. 이커머스 사이트에서 고객이 리뷰를 남길 때 글을 쓰는 동시에 상품에 평점을 부여하도록 하고 있다. 이 평점은 고객이 상품에 만족한 정도를 단적으로 보여준다. 오피니언 마이닝은 평점이 없더라도 긍정, 부정, 중립에 해당하는 단어들을 통해 고객의 만족도를 알려주는 것이 가능하기 때문에 자주 활용되고 있다.

문제는 이와 같은 긍정, 부정의 단일 차원 분석은 소비자가 제품의 어떤 점에 만족, 불만족했는지 알려주지 못한다는 점이다. 각기 다른 이유로 제품에 만족한 고객들이 동일한 평점을 준다면 기업은 제품의 장단점을 구체적으로 파악할 수 없다. 이 문제를 보완하기 위한 방안으로 상품에 대한 평가를 사전에 정한 평가 기준 별로 하도록 유도하거나 리뷰를 기초로 평가 기준을

추출하는 것이다. 평가 기준 추출 방법으로는 형태소 분석을 통해 명사를 추출해내거나(Son et al. 2017) 토픽모델링 기법을 통해 각 리뷰의 주제가 될 만한 단어를 추출하는 것이 대표적이다.

하지만 현재의 리뷰로부터 평가 기준을 추출하는 방법은 여전히 상품과 연관성이 없는 단어가 많이 추출되거나 그 양이 너무 많아 수동으로 다시 단어를 걸러야 하는 문제가 존재한다. 무엇보다 그 단어들이 해당 제품의 지엽적인 특징만을 반영할 경우 시장 내에서 상품의 객관적인 위치를 분석할 수 없다는 문제가 있다. 즉 해당 상품의 리뷰의 특징을 기반으로 일반화된 평가 기준 단어를 추출해야 할 필요성이 있는 것이다.

이러한 문제 해결을 위해 본 연구에서는 리뷰에서 상품 평가 기준에 적합한 단어를 반자동으로 추출해내는 모델을 제시한다. 본 모델은 사전에 상위 상품 카테고리 수준에서 사례 베이스를 구축하여 이를 기반으로 하위 상품 카테고리 리뷰 데이터에서 평가 기준으로 적합한 단어를 분류해내는 방식이다. 제시하는 모델은 토픽모델링 기법인 LDA로 리뷰의 특징 단어들을 추출하고 k-최근접 이웃 접근법(k-Nearest Neighbor Approach, k-NN)을 평가 기준 분류에 활용한다. 먼저 상위 상품 카테고리 수준에서 LDA로 상품 평가 기준 후보군을 추출하고 평가 기준으로 적절한지를 판단, 태깅(Tagging) 함으로써 k-NN 사례 베이스를 만든다. 리뷰 데이터 역시 LDA로 평가 기준 후보군이 만들어지면 앞선 사례를 기반으로 평가기준을 생성하고자 하는 하위 상품 카테고리의 평가 기준이 추출된다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 연구를 설명하고, 3장에서는 본 연구의 평가 기준 추출 모델을 제시한다. 4장에서는 모델의 유용성을 확인하기 위한 실험을 서

술한다. 마지막 5장에서는 결론을 제시한다.

2. 관련 연구

2.1 상품 평가 기준 추출

웹의 개방적인 특성 덕분에 소비자들이 온라인상에 텍스트 형태로 남기는 ‘상품 리뷰’의 양이 급격하게 증가하였고, 자연스럽게 상품 리뷰에 담긴 방대한 양의 의견 정보를 자동으로 분석하는 방법이 필요해졌다(Jang et al. 2014). 특히 소비자들이 상품을 구매하고자 할 때 어떠한 평가 기준을 바탕으로 구매를 행하는지에 대한 분석이 중요해졌다. 이에 따라 온라인 리뷰를 분석하여 글쓴이의 의견이나 평가, 태도, 감정 등을 추출해 내는 오피니언 마이닝에 대한 연구가 활발하게 진행되고 있다(Kim et al. 2016).

온라인 리뷰로부터 평가 기준을 추출하기 위해서, 명사를 빈도순으로 정렬하고 상품의 특징으로 판단되는 명사들을 특징으로 추출하는 연구들이 이루어졌다. 특징이라고 판단된 명사들이 속한 리뷰의 평점과 해당 상품의 전체 평점과의 비교를 통해 특징을 분류하는 형태의 연구가 이루어진 바 있다(Lee et al. 2009; Song et al. 2011). 이와 비슷하게 명사를 빈도순으로 정렬한 후, 특징을 추출하고 명사들이 속한 리뷰의 평점을 각 특징에 분배하여 특징 별 점수를 부여한 연구가 있다(Son et al. 2017). 그러나 이러한 연구들은 각 상품에 대한 특징을 추출하는 데 있어서, 주관적인 판단이 개입되었다는 한계를 가진다.

이러한 주관적인 판단이 개입되는 것을 극소화하기 위해 토픽 모델링 기법이 사용되게 된다.

토픽 모델링 기법은 구조화되지 않은 방대한 문헌에서 주제를 찾아내기 위해 사용되는 알고리즘으로 리뷰를 분석할 때에도 적용할 수 있다(Steyvers et al. 2007; Blei, 2012). 한 연구에서는 LDA 기법과 동의어 사전을 활용해 중국어 리뷰를 대상으로 상품 평가 기준을 추출해냈다(Ma et al. 2013). 추출한 상품 평가 기준에 해당하는 리뷰에 대해 해당하는 리뷰에 대해 감성 분석을 실시하여 소비자에게 도움이 될 만한 리뷰를 분류하는 연구이다. 또한 아마존에 존재하는 두 상품의 영어 리뷰를 대상으로 LDA를 실시한 후, 두 상품에 대한 비교 기준을 추출해 내는 연구가 있다(Wang et al. 2018). 그러나 단순히 LDA를 실시하여 토픽을 추출해낸 다음 그대로 토픽들을 사용하기에는 상품 평가 기준으로서 적절하게 활용될 수 없어 이를 정제하는 작업이 필요하고, 이전의 연구에서는 불용어 사전을 활용하거나 또는 주관적으로 걸러내는 등 주관적인 판단이 개입되는 한계를 지니고 있다. 정리하면 기존의 온라인 리뷰로부터 상품 평가 기준을 추출하는 연구들은 기준을 추출할 때 주관이 과도하게 개입되거나, 불용어 사전을 활용한다는 한계를 보인다. 본 연구에서는 이러한 한계를 극복하기 위해 LDA를 실시한 결과에 대해 k-NN을 활용해 기계학습을 실시한다. 이전 연구들과는 다르게 불용어 사전을 활용하지 않고 k-NN을 통해 반자동적으로 상품 평가 기준을 도출해 내어 주관적인 판단을 최소화 했다는 것에 주요한 차이점이 있다.

2.2 워드 임베딩

워드 임베딩(Word embedding)은 단어(word)를 컴퓨터가 인식할 수 있도록 숫자 원소로 이루어

진 수십 또는 수백 차원의 벡터로 변환하는 것이다. 변환된 벡터들은 의미적 또는 문법적 성질이 비슷한 단어들은 다차원의 벡터 공간 내에서 벡터 간의 거리가 작게 표현된다. 이를 통해 단어 벡터들은 단순히 임의적으로 배정된 수치를 나타내는 것이 아니라 의미적 자질들을 표현하는 벡터들이 된다. 이러한 워드 임베딩 기법은 통계적 언어 모델인 n-gram, NNLM, RNNLM 등이 있었으나 신경망 기반 언어 모델로 발전하여 대표적으로 Word2Vec, FastText, ELMO가 활용되고 있다.

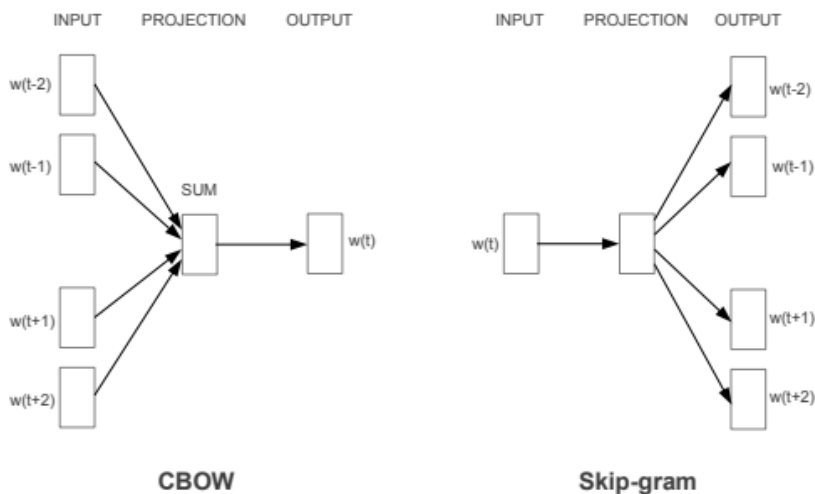
2.2.1 Word2Vec

Word2Vec은 Distributional Hypothesis에 근거한 방법론으로 비슷한 위치에 등장하는 단어들은 그 의미 또한 유사할 것이라는 전제가 내포되어 있다(Mikolov et al. 2013). 주변 단어(window) 크기에 따라 말뭉치를 슬라이딩하여 중심 단어별 주변 단어의 구성에 따라 벡터 값을 업데이트해 나가는 방식으로 훈련을 진행한다. 윈도우 내

에 등장한 주변 단어는 중심 벡터 공간에서 가깝게 벡터 값이 형성되지만, 윈도우 내에 등장하지 않은 주변 단어는 벡터 값이 멀게 형성된다. Word2Vec은 총 2가지 모델로 CBOW(Continuous Bag-of-Words Model)과 Skip-gram(Continuous Skip-gram Model)이 존재한다. CBOW 모델은 주변 단어로 중심 단어가 무엇인지 예측하는 방식으로 학습을 진행된다. Skip-gram 모델은 중심 단어를 통해서 주변 단어가 무엇인지 예측하는 방식으로 이루어진다(<Figure 1> 참조). 하지만 출현 횟수가 적은 단어에 대해서는 임베딩이 제대로 이루어지기 어렵고, 훈련 말 뭉치에 없는 새로운 단어가 출현했을 때 이를 처리할 수 없는 단점이 존재한다.

2.2.2 FastText

FastText는 Word2Vec 방법 중 skip-gram방식을 변형한 방법으로 Word2Vec의 한계점을 극복한 방법론이다. Word2Vec은 단어를 원자적 단위로 취급해서 벡터를 형성하지만 FastText는 단어



<Figure 1> Word2Vec

를 n-gram 조합으로 취급하여 한 단어에 대한 벡터는 n-gram의 합으로 형성된다. 이를 통해 희귀한 단어에 대해 더 좋은 임베딩모델을 형성할 수 있으며, 훈련 말 문치에 존재하지 않았던 단어의 벡터 또한 만들어낼 수 있다. 본 연구에서는 리뷰 데이터를 활용했기 때문에 희귀한 단어들 많았고, 평가 기준 추출 단계에서는 새로운 단어들의 벡터를 형성할 필요성이 있었기 때문에 FastText를 활용했다.

2.3 k-최근접 이웃 접근법(k-Nearest Neighbor Approach, k-NN)

최근접 이웃 접근법(Nearest Neighbor Approach, NN)은 분류되지 않은 새로운 관측 값을 이전에 분류된 값들의 집합 중 가장 가까운 집합으로 분류하는 방법이다(Cover et al. 1967). 거리공간 X 내의 x_i 와 분류 집합 $\{1, 2, \dots, k\}$ 의 한 값 y_i 에 대하여 n개의 쌍 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 이 존재할 때 새로운 x의 분류는 x와 최근접 이웃 $x'_m \in \{x_1, x_2, \dots, x_n\}$ 간 거리에 의존한다. 이때 x와 x_i 의 거리를 구하는 방법에는 유클리드 거리(Euclidean distance), 맨해튼 거리(Manhattan distance) 등이 있으며 본 연구에서는 유클리드 거리를 적용한다. 수식은 다음과 같다.

$$d_E(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \quad (1)$$

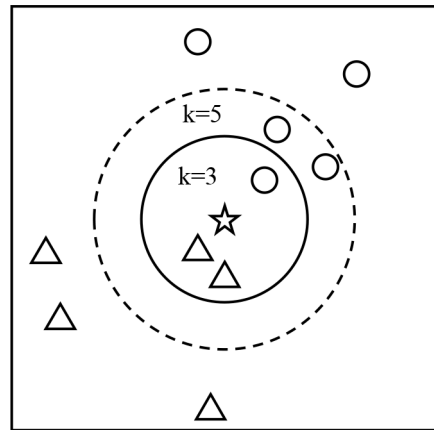
$$d_M(a, b) = \|a - b\|_1 = \sum_{i=1}^n |a_i - b_i| \quad (2)$$

거리함수를 적용하여,

$$d(x'_m, x) = \min d(x_i, x) \quad (i = 1, 2, \dots, n) \quad (3)$$

를 만족하는 x'_m 을 x의 최근접 이웃이라 하고 그때의 y'_m 을 x의 분류로 한다.

k-NN은 고려되지 않은 n-1개의 x'_m 중 k-1개를 추가적으로 고려하는 방법이다(Cover et al. 1967; Patrick et al. 1970). 즉, 새로운 관측 값 x의 주변 $x'_m \in \{x_1, x_2, \dots, x_n\}$ 중 최근접 이웃을 포함한 가까운 k개의 분류를 확인하고 k개 중 가장 많은 y'_m 으로 x를 분류하는 것으로 <Figure 2>과 같이 진행된다.



<Figure 2> k-Nearest Neighbor Approach

<Figure 2>에서는 k=3 일 때 검증 표본(별)의 분류를 삼각형으로, k=5일 때 분류를 원으로 한다. 가장 많은 수의 y'_m 을 x의 분류로 보는 것은 합리적이지만, <Figure 2>과 같이 k 값에 따라 그 결과가 달라지는 경우가 존재하므로 x와 x'_m 간 거리에 따라 가중치를 부여하는 것 또한 합리적이다(Dudani, 1976). 이를 토대로 본 연구에서는 데이터의 불균형을 해결하기 위해 k-NN 적용 시 거리에 따른 가중치를 고려한다. 새로운 x의 근접 이웃들을 거리 순으로 정렬한 것을 $\{x'_1, x'_2, \dots, x'_n\}$ 라 하고, x_j 와 x와의 거리를 d_j 라

하면, 가장 먼 이웃 x_n 과의 거리는 d_n 이 되고, 가중치 w_j 는 다음과 같이 정의할 수 있다.

$$w_j = \begin{cases} \frac{d_n - d_j}{d_n - d_1}, & d_n \neq d_1 \quad (j = 1, 2, \dots, n) \\ 1, & d_n = d_1 \end{cases} \quad (4)$$

본 연구에서 거리 가중 k-NN(Distance Weighted k-NN)은 사전 추출된 평가 기준 후보군에 부여한 태그를 기반으로 새로운 상품의 평가 기준 후보군을 정제하는데 사용된다.

2.4 토픽 모델링(Topic Modeling)

토픽 모델링은 문서의 추상적인 주제를 발견하기 위한 통계적 모델이다. 토픽 모델링은 거대한 문서 집합에도 적용 가능하며 기술의 발전으로 인해 실시간 분석도 가능하다(Blei, 2012). 토픽 모델링 기법은 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA), 계층적 디리클레 프로세스(Hierarchical Dirichlet Process, HDP) 등이 있다.

본 연구에서는 토픽 모델링 기법 중 LDA를 활용하여 상품별 리뷰를 분석하고 리뷰의 잠재된 주제, 주제가 언급하는 내용 및 중요 단어를 파악한다.

LDA는 토픽 모델링 기법 중 하나로 주어진 문서에 대해 어떤 주제가 존재하는지 서술하는 확률적 주제 파악 모형이다(David et al. 2003). 사전 습득 혹은 직접 지정한 주제별 단어 수를 기반으로 문서 내 단어 분포를 분석하여 대상 문서가 어떤 주제들을 가지고 있는지 추정한다. 또한 주제들에 포함된 단어들의 비중을 통해 각 주제가 어떤 내용을 주로 언급하고 있는지 파악할 수 있다.

3. 제안 모델

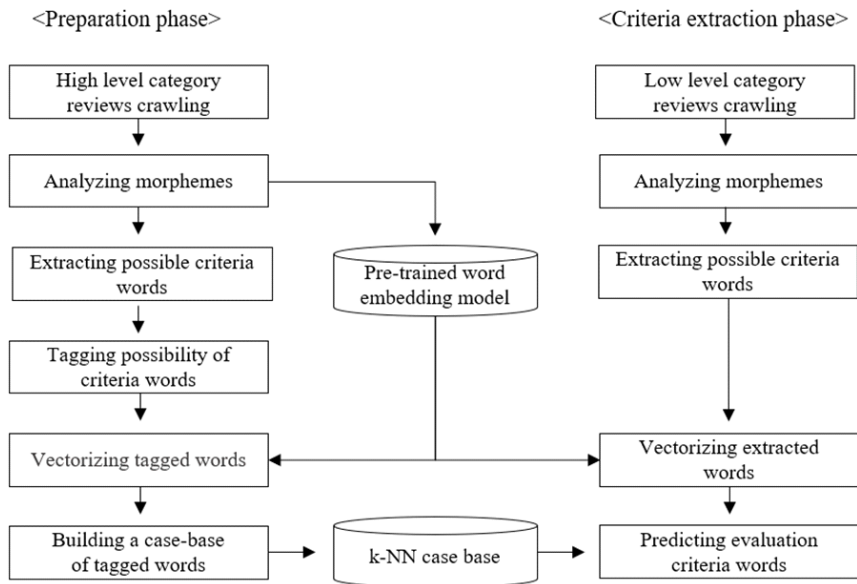
본 모델은 아래 <Figure 3>와 같이 사전 준비 단계(preparation phase)와 평가 기준 추출 단계(Criteria extraction phase)로 구분된다. 사전 준비 단계에서는 k-NN 사례 베이스가 만들어진다. 이 사례베이스를 기반으로 새로운 리뷰 데이터에서 평가 기준으로 적합한 단어들을 분류해내는 것이 평가 기준 추출 단계이다.

이커머스에서는 상품 카테고리를 대분류, 중분류, 소분류 3단계로 구분하는 것이 일반적인데, 본 모델에서는 대분류 별로 사례 베이스를 구축했다. 예를 들어, 국내 이커머스 ‘11번가’의 ‘가습기’ 제품은 대분류인 ‘디지털/가전/컴퓨터’ 카테고리 중에서도 중분류 ‘계절가전’에 속한다. 이 때 ‘디지털/가전/컴퓨터’ 사례베이스는 ‘계절가전’을 포함한 중분류들의 상위 n개 제품의 리뷰 데이터에서 수집된 단어들의 집합이다. 각 단어들은 평가 기준 가능 여부에 따라 태깅된다. 평가 기준에 적합할수록 높은 점수로 태깅되었다. 이후 평가 기준 추출 단계에서 ‘가습기’의 리뷰 데이터를 넣게 되면 주변 사례들을 기준으로 각 단어가 태깅된다. 태깅 점수에 따라 평가 기준으로 사용할 단어를 구분할 수 있다.

3.1 사전 준비단계

3.1.1 데이터 수집 및 전처리

대분류별로 상품 리뷰를 크롤링한다. 대분류 내의 모든 중분류에 대해 ‘많은 리뷰 순’으로 상품을 정렬한 뒤 상위 n개(본 연구에서는 40) 상품의 리뷰를 수집한다. 한국어 형태소 분석기 Konlpy의 Twitter 패키지를 통해 수집한 데이터의 명사, 형용사, 부사, 동사를 추출한다. 추출한



〈Figure 3〉 The Overview of the Proposed Approach to Evaluation Criteria Extraction

단어들로 워드 임베딩(Word Embedding) 모델을 생성하며 이때 FastText 기법을 활용한다. FastText 기법 적용 시 대부분의 파라미터는 기본값을 적용하고 size, min_count, workers, sg 파라미터를 128, 10, 10, 1로 적용한다.

3.1.2 평가 기준 후보군 추출

평가 기준 후보군 추출에는 LDA를 활용한다. 기존 연구에서는 LDA를 통해 특정 주제와 해당 주제를 구성하는 내부 단어를 추출할 수 있었다. 본 연구에서는 특정 주제를 뽑는 것보다 상품 평가 기준이 될 수 있는 단어를 뽑는 데에 중점을 두었다. LDA는 토픽 수, 단어 수, 랜덤 시드(Random Seed) 등의 매개 변수가 변화하면 추출되는 토픽 내 단어도 변화한다. 따라서 매개 변수를 바꿔가며 토픽 내의 출현 단어를 반복 추출하는 과정을 자동화하였고, 여기서 추출된

단어들을 평가 기준 후보군으로 하였다. 평가 기준 후보군 추출은 중분류별로 실시한다.

3.1.3 평가 기준 후보군 정제

평가 기준으로 사용할 수 있는지에 따라 후보군 단어들을 직접 태깅한다. 태그는 모든 상품군에 적용 가능한 일반적인 기준을 2, 특정 상품군에만 적용 가능하거나 하위 기준은 1, 평가 기준이 될 수 없는 단어를 0으로 한다. <Table 1>은 ‘외장 HDD’가 속한 ‘저장 장치’ 중분류의 태깅을 실시한 결과이다. 중분류별로 태깅을 마친 단어들을 대분류 단위로 합치고 k-NN 접근법을 활용하여 k-NN 분류기를 생성한다. 제안 모델은 이 과정을 통해 각 단어의 벡터 값에 따른 태그를 학습할 수 있다. 이는 추후 학습되지 않은 단어에 대해서도 벡터 상에서 가까운 단어들과 그 태그를 확인하여 결과를 도출할 수 있도록 한다.

〈Table 1〉 Middle level category ‘Drives/Storage’ words tagging

| Evaluation criteria words | Tag | Evaluation criteria words | Tag |
|---------------------------|-----|---------------------------|-----|
| Performance (성능) | 2 | Recognition (인식) | 1 |
| Price (가격) | 2 | Cost-effectiveness(가성비) | 1 |
| Color (색상) | 2 | Pink (분홍색) | 1 |
| Effort (수고) | 0 | Brother (동생) | 0 |

이러한 측면에서 단어와 그 분류를 저장해 사전에 존재하지 않는 단어에 대해 결과를 도출할 수 없는 사전분류식과 차별점을 가진다.

3.2 평가 기준 추출 단계

하나의 소분류가 평가 기준 추출의 대상으로 정해지면 리뷰 수가 많은 상위 n개 상품의 리뷰를 크롤링한다. 이후 사전 준비단계와 동일한 과정으로 평가 기준 후보 단어를 추출한다. 워드 임베딩 모델을 사용하여 후보 단어들을 벡터화(Vectorization)하고 사전 준비단계에서 생성된 대분류 k-NN 분류기를 통해 평가 기준 가능성을 판단한다. 아래 <Table 2>는 소분류 ‘외장 HDD’ 평가 기준 추출 결과이다. 태그 2와 1로 분류된 단어를 모두 평가 기준으로 사용할 수 없기 때문에, 리뷰에서 언급된 비율을 기준으로 최종 평가 기준을 선정한다. 태그 2의 경우 모든 상품군에

적용이 가능하기 때문에 기준이 되는 언급 비율은 비교적 낮게 설정되며, 태그 1의 경우 특정 상품군에만 해당하는 평가 기준을 선별하기 위해 기준이 되는 언급 비율을 비교적 높게 설정한다. 본 연구에서는 태그 2에는 1%, 태그 1에는 10%를 기준으로 적용하여 각 소분류별로 7~10개의 평가 기준을 도출한다. 최종 평가 기준 선정에 관한 수식은 다음과 같다.

$$\frac{Word\ Count\ (Tag\ 1)}{Total\ Count\ (Tag\ 1)} > \theta_1, \frac{Word\ Count\ (Tag\ 2)}{Total\ Count\ (Tag\ 2)} > \theta_2 \quad (5)$$

θ_1, θ_2 : Reference ratio criteria for selection

따라서 <Table 2>의 경우 태그 2는 ‘배송’부터 ‘크기’까지 여섯 단어, 태그 1은 ‘용량’부터 ‘포장’까지 네 단어가 남아 총 10개의 평가 기준이 추출된다.

〈Table 2〉 Low level category ‘External HDD’ predicted words and reference ratio

| Tag 2 Predicted words | Reference ratio | Tag 1 Predicted words | Reference ratio |
|------------------------|-----------------|----------------------------|-----------------|
| Delivery (배송) | 0.290828 | Capacity (용량) | 0.289222 |
| Price (가격) | 0.239103 | Speed (속도) | 0.197006 |
| Use (사용) | 0.229682 | Cost-effective (가성) | 0.121557 |
| Design (디자인) | 0.184123 | Package (포장) | 0.114371 |
| Performance(성능) | 0.027918 | Recognition (인식) | 0.071258 |
| Size (크기) | 0.023122 | Function (기능) | 0.048503 |

4. 실험 내용 및 결과

4.1 실험 내용

제안 모델의 성능 평가를 위해 실험을 진행하였다. 비교를 위한 3가지 모델은 명사를 빈도 순으로 정렬해 추출한 평가 기준, LDA 기법을 통해 추출한 주제 단어를 빈도수로 정제한 평가 기준과 11번가에서 실제 제공하는 평가 기준이다. 평가 기준을 추출할 대상으로는 11번가 베스트 100 상품 중 디지털/가전/컴퓨터 분류에 포함되는 상위 10개 소분류¹⁾를 선정했다.

명사 빈도 추출 모델은 리뷰 데이터로부터 명사만을 추출하여 빈도 순으로 정렬하고 상위 n개를 평가 기준으로 선정하는 방법이다. LDA 빈도 추출 모델은 리뷰 데이터에 형태소 분석을 한 뒤 LDA를 진행한다. 진행마다 추출되는 결과가 달라지는 임의성이 존재하므로 반복 진행을 통해 임의성을 제거한다. 추출된 토픽 단어들을 순으로 정렬하여 상위 n개를 추출하고 이를 평가 기준으로 한다. 11번가 제공 평가 기준은 단일상품의 리뷰를 분류한 목록을 수집하고 목록에 포함된 리뷰가 많은 순으로 정렬, 상위 n개를 추출하여 생성한다.

위 두 모델은 단어가 리뷰에서 언급된 빈도를 평가 기준 판단의 주된 척도로 한다. 이는 많이 언급된 단어가 평가 기준이 될 것이라는 가정에 기반하는 것으로 제안하는 모델과 차이가 있다. 제안하는 모델은 두 모델과 달리 사례 베이스를 사용하여 단어의 의미를 나타내는 벡터 값에 따라 평가 기준 여부를 결정한다. 특히, 명사 빈도 추출 모델은 제안하는 모델과 달리 토픽 모델링 기법을 사용하지 않아 문서 내의 주제가 되는 단어를 고려하는 과정이 없다. LDA 빈도 추출 모델은 토픽 단어를 추출한 점은 유사하나, 사례 베이스를 사용하지 않아 각 단어의 벡터 값으로 평가 기준으로 부적합한 단어들을 정제하는 과정이 없다는 점에서 차이가 있다.

제안 모델에서 추출한 상품 평가 기준과 동일한 개수의 평가 기준을 각 모델로부터 추출, 종합 후 중복을 제거한다. 이를 피설문자에게 제공하고 피설문자는 특정 상품군의 평가 기준으로 적합하다고 판단되는 항목을 복수 선택한다. <Table 3>은 ‘외장 HDD’에 대해 제안 모델과 선정한 3개의 모델을 통해 추출한 평가 기준이며 <Figure 4>은 전체 설문 중 ‘외장 HDD’에 해당하는 설문이다.

<Table 3> Extracted evaluation criteria for each model (External HDD)

| Model | Evaluation criteria |
|-----------------|---|
| Suggested Model | 배송, 가격, 사용성, 디자인, 성능, 크기, 용량, 속도, 가성비, 포장 |
| Noun frequency | 배송, 마음, 정말, 제품, 구매, 사용성, 외장하드, 쪽, 디자인, 가격 |
| LDA frequency | 배송, 마음, 가격, 구매, 사용성, 완전, 외장하드, 디자인, 제품, 강 |
| 11st | 제품일반, 사용성, 가격, 품질, 크기, 성능, 양, 디자인, 색상, 구매 |

1) 가습기, USB메모리, 공유기, 메모리카드, 청소기, 휴대폰케이스, 전기요, 선풍기, 외장HDD, 헤어드라이기

해당 제품을 구매할 때 귀하가 고려하시는 기준은 어떤 것입니까?

복수 선택 가능

외장_HDD *

| | | | | |
|------------------------------|------------------------------|-------------------------------|-------------------------------|-----------------------------|
| <input type="checkbox"/> 가격 | <input type="checkbox"/> 마음 | <input type="checkbox"/> 속도 | <input type="checkbox"/> 용량 | <input type="checkbox"/> 포장 |
| <input type="checkbox"/> 가성비 | <input type="checkbox"/> 배송 | <input type="checkbox"/> 썩 | <input type="checkbox"/> 정밀 | <input type="checkbox"/> 품질 |
| <input type="checkbox"/> 구매 | <input type="checkbox"/> 사용성 | <input type="checkbox"/> 양 | <input type="checkbox"/> 제품 | |
| <input type="checkbox"/> 강 | <input type="checkbox"/> 색상 | <input type="checkbox"/> 완전 | <input type="checkbox"/> 제품일반 | |
| <input type="checkbox"/> 디자인 | <input type="checkbox"/> 성능 | <input type="checkbox"/> 외장하드 | <input type="checkbox"/> 크기 | |

<Figure 4> Evaluation criteria of external HDD

설문은 이커머스 이용 경험이 있는 138명에게 진행했다. 10개 상품에 대하여 각각 14~23개의 평가 기준을 제시하고 이를 중복 선택하게 했다. 개인이 선택한 평가 기준과 각 모델별로 추출한 평가 기준의 교집합 수를 기준으로 점수를 부여하고 이를 성능 평가 지표로 사용한다.

$$n(A \cap \text{Model } i) = \text{Score } i \quad (6)$$

A : Criteria selected in the survey

<Figure 5>는 ‘외장 HDD’에 대한 설문 결과이며 <Table 4>는 ‘외장 HDD’에 적용된 각 모델에 점수를 부여한 것이다.

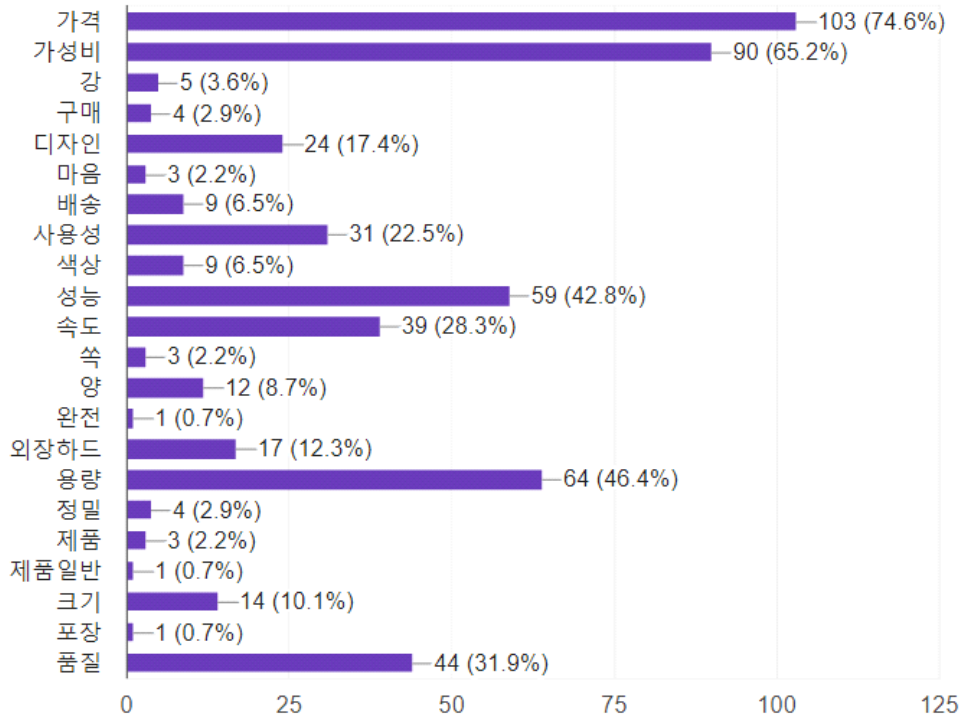
4.2 실험 결과

부여한 점수의 차이를 확인하기 위해 구한 모델 별 평균은 <Table 5>와 같다.

평균 점수를 바탕으로 대응 표본 t-검정을 진행했고 제안 모델과 3개의 모델 간 점수의 차이가 유의한지 확인했다. 제안 모델과 3개의 모델을 검정한 결과는 <Table 6>와 같다.

<Table 4> Scoring for each model

| | Survey results A : 가격, 가성비, 디자인, 성능, 용량, 품질 | Score |
|-----------------|---|----------|
| Suggested Model | 배송, 가격 , 사용성, 디자인 , 성능 , 크기, 용량, 속도, 가성비 , 포장 | 5 |
| Noun frequency | 배송, 마음, 정밀, 제품, 구매, 사용성, 외장하드, 썩, 디자인 , 가격 | 2 |
| LDA frequency | 배송, 마음, 가격 , 구매, 사용성, 완전, 외장하드, 디자인 , 제품, 강 | 2 |
| 11st | 제품일반, 사용성, 가격 , 품질 , 크기, 성능 , 양, 디자인 , 색상, 구매 | 4 |



<Figure 5> Survey results of external HDD

<Table 5> Average for each model

| Small category | Suggested Model | Noun | LDA | 11st |
|-------------------|-----------------|------|------|-------------|
| Humidifier | 4.03 | 1.35 | 1.2 | 3.76 |
| USB flash drive | 3.01 | 1.01 | 1.41 | 2.93 |
| Router | 3.04 | 1.83 | 1.82 | 2.21 |
| Memory card | 2.68 | 0.99 | 1.01 | 2.23 |
| Vacuum cleaner | 3.45 | 1.88 | 1.95 | 3.39 |
| Cell phone case | 1.67 | 1.14 | 1.2 | 2.77 |
| Electric mattress | 2.46 | 1.19 | 1.27 | 3.01 |
| Fan | 3.3 | 1.38 | 1.54 | 2.93 |
| External HDD | 3.14 | 1.2 | 1.22 | 2.18 |
| Hair dryer | 2.75 | 1.42 | 1.41 | 2.52 |

S: Suggested Model, N: Noun frequency, LDA: LDA frequency, 11st: Evaluation criteria of 11st
 *** p < 0.001, ** p < 0.005, * p < 0.01

〈Table 6〉 Paired t-test Results

| Small category | Pair | Mean | Std. Deviation | df | t | p-value |
|-------------------|--------|---------|----------------|-----|----------|----------|
| Humidifier | S - N | 2.6812 | 1.148316 | 137 | 27.3288 | 0.000*** |
| | S-LDA | 2.8333 | 1.225238 | 137 | 27.0668 | 0.000*** |
| | S-11st | 0.2681 | 0.62019 | 137 | 5.0601 | 0.000*** |
| USB flash drive | S - N | 2.0000 | 1.215838 | 137 | 19.2537 | 0.000*** |
| | S-LDA | 1.6014 | 0.99666 | 137 | 18.8073 | 0.000*** |
| | S-11st | 0.0870 | 1.18247 | 137 | 0.8607 | 0.1954 |
| Router | S - N | 1.2101 | 1.364418 | 137 | 10.3813 | 0.000*** |
| | S-LDA | 1.2174 | 1.355286 | 137 | 10.5138 | 0.000*** |
| | S-11st | 0.8261 | 1.089562 | 137 | 8.8743 | 0.000*** |
| Memory card | S - N | 1.6957 | 1.311094 | 137 | 15.1378 | 0.000*** |
| | S-LDA | 1.6667 | 1.315093 | 137 | 14.8338 | 0.000*** |
| | S-11st | 0.4493 | 1.13618 | 137 | 4.6283 | 0.000*** |
| Vacuum cleaner | S - N | 1.5725 | 0.939378 | 137 | 19.5930 | 0.000*** |
| | S-LDA | 1.5000 | 0.861832 | 137 | 20.3718 | 0.000*** |
| | S-11st | 0.0580 | 1.134237 | 137 | 0.5982 | 0.2753 |
| Cell phone case | S - N | 0.5217 | 0.861557 | 137 | 7.0881 | 0.000*** |
| | S-LDA | 0.4638 | 0.790885 | 137 | 6.8635 | 0.000*** |
| | S-11st | -1.1014 | 0.651529 | 137 | -19.7875 | 2.69E-42 |
| Electric mattress | S - N | 1.2754 | 1.684405 | 137 | 8.8623 | 0.000*** |
| | S-LDA | 1.1957 | 1.502938 | 137 | 9.3116 | 0.000*** |
| | S-11st | -0.5435 | 1.167926 | 137 | -5.4466 | 1.15E-07 |
| Fan | S - N | 1.9130 | 1.426856 | 137 | 15.6930 | 0.000*** |
| | S-LDA | 1.7536 | 1.402955 | 137 | 14.6303 | 0.000*** |
| | S-11st | 0.3696 | 1.029113 | 137 | 4.2033 | 0.000*** |
| External HDD | S - N | 1.9420 | 1.667108 | 137 | 13.6349 | 0.000*** |
| | S-LDA | 1.9203 | 1.668383 | 137 | 13.4720 | 0.000*** |
| | S-11st | 0.9638 | 0.988407 | 137 | 11.4129 | 0.000*** |
| Hair dryer | S - N | 1.3261 | 1.22275 | 137 | 12.6939 | 0.000*** |
| | S-LDA | 1.3333 | 1.199839 | 137 | 13.0070 | 0.000*** |
| | S-11st | 0.2246 | 0.916975 | 137 | 2.8674 | 0.002** |

S: Suggested Model, N: Noun frequency, LDA: LDA frequency, 11st: Evaluation criteria of 11st
 *** p < 0.001, ** p < 0.005, * p < 0.01

제안 모델과 명사 빈도 추출 모델, 제안 모델
 과 LDA 빈도 추출 모델 간 대응 표본 t-검정 결

과, 모든 소분류에서 제안 모델이 우수했으며 11
 번가와의 비교에서는 2개 부문을 제외하고 모두

우수했다.

이는 제안 모델에서 평가 기준 후보군을 추출할 때 사용한 LDA 기법이 단순 명사를 추출하는 것보다 우수한 것으로 해석되며, 단순 LDA 기법만을 적용하는 것보다 LDA 기법으로 추출된 평가 기준들을 태그하고 k-NN을 통해 정제하는 것이 우수한 것으로 해석된다.

제안 모델과 11번가 제품 평가 기준 간 검정 결과, 10개 소분류 중 6개 항목²⁾에서 제안 모델이 우수했고 2개 항목³⁾은 기각되지 않았으며 2개 항목⁴⁾은 11번가 제품 평가 기준이 우수했다. 11번가 제품 평가 기준의 경우 자체 사전을 기반으로 리뷰를 분류했기 때문에 등록된 단어와 유사한 언어 혹은 등록되지 않은 언어가 나타날 때 부정확한 결과가 나타난다. 리뷰를 기반으로 평가 기준을 추출한 제안 모델은 리뷰와 사용된 언어에 따라 유동성을 가지므로 11번가 제품 평가 기준보다 우수하다고 해석된다.

기존의 단순 명사 빈도 추출 모델과 LDA 빈도 추출 모델은 단어가 나타나는 빈도에 많은 부분을 의존하고 있다. 때문에 평가 기준으로 적절하지 않은 단어들이 추출된다. 예를 들어 <Table 3>에서 ‘외장 하드’와 같은 직접적인 상품 언급과 ‘마음’, ‘완전’과 같은 감정 표현 단어가 나타나는 한계점이 존재한다. 또한 11번가 제품 평가 기준의 경우 <Table 3>에서 ‘양’과 같이 상품과 관련 없는 단어가 나타나는 한계점이 존재한다. 본 연구의 제안 모델은 이러한 한계점을 해결하여 불필요한 단어를 제외한, 보다 정확한 평가 기준이 추출되도록 한다.

5. 결론

본 연구는 기업의 상품에 대한 평가 기준을 LDA와 k-NN을 이용해 온라인 리뷰에서 추출하는 방안을 제시한다. 기존 사전 기반 추출 방식이 사용자의 의견을 반영하지 못하는 한계를 극복하기 위해 본 연구는 LDA 기법으로 평가 기준 후보군을 추출했다. 리뷰의 명사 단어들을 빈도순으로 정렬해 평가 기준을 추출했던 기존 연구와 달리, 본 연구에서는 k-NN을 이용해 평가 기준 후보군을 정제했다. 제안 모델은 해당 상품군의 리뷰에 대해 형태소 분석을 한 뒤 LDA를 통해 평가 기준 후보군을 추출한다. 이후 사전에 생성한 k-NN 분류기와 언급 비율을 이용해 최종 평가 기준을 추출한다.

제안 모델의 성능을 평가하기 위한 실험을 진행했다. 비교 대상은 명사 빈도 추출, LDA 빈도 추출, 11번가 제품 평가 기준이다. 제안 모델 및 비교 모델로 추출한 평가 기준을 종합하고, 설문을 통해 피설문자가 판단하기에 적합한 평가 기준을 복수 선택하게 했다. 설문 결과를 바탕으로 각 모델에 점수를 부여하고 대응 표본 t-검정을 진행한 결과, 비교 모델보다 제안 모델의 성능이 우수했다. 제안 모델은 사례 베이스를 사용함으로써 언급 빈도는 높으나 평가 기준이 될 수 없는 단어들이 추출되는 문제를 해결하였다. 리뷰 데이터에서 평가 기준 후보군을 직접적으로 추출함으로써 해당 상품에 부합하지 않는 평가 기준을 배제시키는 효과를 볼 수 있었다.

본 연구에서 제안한 모델은 기존 기술에 비해 소비자의 의견을 정확하게 반영하며 보다 적합

2) 가습기, 공유기, 메모리카드, 선풍기, 외장 HDD, 헤어드라이기

3) USB메모리, 청소기

4) 휴대폰케이스, 전기요

한 평가 기준을 제공한다. 기존 기술은 빈도에 의존하기 때문에 모델에 사용하는 리뷰의 질과 양에 따라 추출되는 평가 기준 결과의 차이가 클 수 있다. 제안 모델은 사례 베이스를 기반으로 하여 기업은 원하는 상품군에 대한 객관적인 평가 기준을 얻을 수 있다. 또한, 이커머스의 빠른 성장에 따라 리뷰의 양이 크게 증가하는 상황에서 제안 모델은 리뷰의 신속하고 정확한 파악에 기여할 수 있고 일부 리뷰를 기반으로 학습한 모델을 적용해 해당 상품군 내의 모든 상품에 대한 리뷰를 분석하고 결과를 도출할 수 있다. 새로운 리뷰들에 나타난 평가 기준을 반자동 방식의 모델을 통해 빠르게 도출함으로써 기업은 소비자의 니즈 변화에 대한 마켓 인사이트를 발굴할 수 있고, 객관적인 상품 비교 및 마켓 센싱 또한 가능하다. 따라서 본 모델은 리뷰들을 통합적으로 분석함으로써 제품군 고유의 상품 평가 기준을 추출했다는 점에 의의가 있다.

본 연구의 한계점 및 향후 연구 주제는 다음과 같다. 한계점으로는 k-NN 분류기 생성 시 태깅하는 과정에서 그 기준이 일관되지 못했다는 점과 이 과정이 불가피하여 완전한 자동화를 이루지 못했다는 점이다. 또한, 평가 기준 후보군이 될 수 있는 단어가 제한적이어서 학습 데이터가 부족했고, 이에 본 연구에서는 k-NN만을 사용했다는 한계점이 있다. 향후 분석하는 리뷰의 수를 늘린다면 다른 기계학습 기법까지 시도하여 결과를 개선할 수 있을 것이다.

후속 연구로는 판매자와 제품에 대한 리뷰를 별개로 분석하여, 각각의 평가 기준을 생성해 제품과 판매자에 대한 평가를 구분하는 것이 가능할 것이다. 현재 리뷰는 제품 평가와 함께 판매자의 신뢰도, 이력이 포함되어 있으며 이를 분리한다면 제품만의 심층적인 평가가 가능해진다.

이를 통해 생산자들은 소비자에 대한 한층 더 깊은 인사이트를 얻을 수 있으며, 제품에 대한 소비자들의 평가 기준 또한 보다 더 정확하게 파악할 수 있을 것이다. 나아가 판매자에 대한 단독 평가를 통해 구매에 영향을 미치는 제품 외 요소들을 발견할 수 있다.

참고문헌(References)

- Blei, D. M., A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, Vol.3, Jan(2003), 993-1022.
- Blei, D. M., "Probabilistic Topic Models" *Communications of the ACM*, Vol.55, No.4 (2012), 77-84.
- Bojanowski, P., E. Grave, A. Joulin and T. Mikolov, "Enriching Word Vectors with Subword Information." *Transactions of the Association for Computational Linguistics*, Vol. 5, No.10(2017), 135-146.
- Chae, S. H., J. I. Lim and J. Y. Kang, "A Comparative Analysis of Social Commerce and Open Market Using User Reviews in Korean Mobile Commerce." *Journal of Intelligence and Information Systems*, Vol.21, No.4(2015), 53-77.
- Cover, T. M. and P. Hart, "Nearest Neighbor Pattern Classification." *IEEE Transactions On Information Theory*, Vol.13, No.1(1967), 21-27.
- Dai, X., I. Spasic and F. Andres, "A Framework for Rating Online Reviews: Topic Modelling, Text Classification and Sentiment Analysis." *ACMSE 2017 The Annual ACM Southeast Conference Featuring Multidisciplinary and Interdisciplinary Computing, At Kennesaw*

- State University, Georgia, 2017.
- Dong, R., M. Schaal, M. P. O'Mahony and B. Smyth "Topic Extraction from Online Reviews for Classification and Recommendation." *Proceeding of the Twenty-Third International Joint Conference on Artificial Intelligence*, (2013), 1310-1316.
- Dudani, S. A., "The Distance-Weighted k-Nearest-Neighbor Rule." *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.SMC-6, No.4(1976), 325-327.
- Fix, E. and J. L. Hodges, "Nonparametric Discrimination: Consistency Properties." *Report for the USAF School of Aviation Medicine, Randolph Field Report Number 4*, Texas, 1951.
- Griffiths, T. L. and M. Steyvers, "Finding Scientific Topics." *Proceedings of the National Academy of Sciences*, (2004), 5228-5235.
- Jang, K. R., K. W. Lee and S. H. Myaeng, "Extracting Implicit Customer Viewpoints from Product Review Text." *Journal of KISS : Software and Applications*, Vol.41, No.5 (2014), 376-386.
- Jin, J., P. Ji and R. Gu, "Identifying Comparative Customer Requirements from Product Online Reviews for Competitor Analysis." *Engineering Applications of Artificial Intelligence*, Vol.49, No.3(2016), 61-73.
- Jo, H. S. and S. G. Lee, "Korean Word Embedding Using FastText." *Journal of Korea Information Science Society*, Vol.2017, No.12(2017), 705-707.
- Keller, J. M., M. R. Gray and J. A. Givens, "A Fuzzy k-Nearest Neighbor Algorithm." *IEEE Transactions On systems, Man, and Cybernetics*, Vol.SMC-15, No.4(1985), 580-585.
- Kim, H. W., H. C. Chan and S. Gupta, "Social Media for Business and Society," *Asia Pacific Journal of Information Systems*, Vol.25, No.2(2015), 211-233.
- Kim, M. J., E. J. Song and Y. H. Kim, "A Design of Satisfaction Analysis System for Content Using Opinion Mining of Online Review Data." *Journal of Internet Computing and Services*, Vol.17, No.3(2016), 107-113.
- Kim, S. W. and N. G. Kim, "A Study on the Effect of Using Sentiment Lexicon in Opinion Classification." *Journal of Intelligence and Information Systems*, Vol.20, No.1(2014), 133-148.
- Kwon, J. Y. and M. Y. Lee, "A Study on the Determining Factors of Online Review Helpfulness." *Journal of Korea Intelligent information Systems Society*, Vol.2012, No.12 (2012), 205-211.
- Lee, H. A., W. C. Lee and K. J. Lee, "Automatic Product Feature Extraction for Efficient Analysis of Product Reviews Using Term Statistics." *The KIPS Transactions : Part B*, Vol.16, No.6(2009), 497-502.
- Lee, J. E., H. K. Seo, and K. Y. Han, "Refined IPC Classification System Based on KNN Using Patent Search Results." *Journal of KIISE Academic Publications*, Vol.38, No.2A (2011), 256-259.
- Lee, M., and H. J. Lee, "Increasing Accuracy of Classifying Useful Reviews by Removing Neutral Terms." *Journal of Intelligence and Information Systems*, Vol.22, No.3(2016), 129-142.
- Lee, S. H., J. Cui and J. W. Kim. "Sentiment Analysis on Movie Review through Building Modified Sentiment Dictionary by Movie

- Genre." *Journal of Intelligence and Information Systems*, Vol.22, No.2(2016), 97-113.
- Lim, B. H. and K. H. Um, "A Study of the Comparison of Product Quality Competitiveness of Consumer Electronics among Major Countries." *Korean Corporation Management Review*, Vol.43, No.3(2012), 131-151.
- Ma, B., D. Zhang, Z. Yan and T. Kim, "An LDA and Synonym Lexicon based Approach to Product Feature Extraction from Online Consumer Product Reviews." *Journal of Electronic Commerce Research*, Vol.14, No.4 (2013), 304.
- Mikolov, T., E. Grave, P. Bojanowski and C. Puhersch, "Advances in Pre-Training Distributed Word Representations." *arXiv preprint arXiv: 1712.09405* (2017).
- Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality." *arXiv preprint arXiv:1310.4546* (2013).
- Mikolov, T., K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representations in Vector Space." *arXiv preprint arXiv:1301.3781*, (2013).
- Patrick, E. A. and F. P. Fischer III, "A Generalized k-Nearest Neighbor Rule." *Information and Control*, Vol.16, No.2(1970), 128-152.
- Rasyidi, M. A., J. M. Kim and K. R. Ryu, "Short-term Prediction of Vehicle Speed on Main City Roads Using the K-Nearest Neighbor Algorithm." *Journal of Intelligence and Information Systems*, Vol.20, No.1(2014), 121-131.
- Santosh, D. T., B. V. Vardhan and D. Ramesh, "Extracting Product Features from Reviews Using Feature Ontology Tree Applied on LDA Topic Clusters." *2016 IEEE 6th International Conference on Advanced Computing, IEEE*, Bhimavaram, 2016.
- Sim, J. S. and H. J. Kim, "A Searching Method for Legal Case Using LDA Topic Modeling" *Journal of the Institute of Electronics and Information Engineers*, Vol.54, No.9(2017), 67-75.
- Son, S. B. and J. H. Chun, "Product Feature Extraction and Rating Distribution Using User Reviews." *The Journal of Society for e-Business Studies*, Vol.22, No.1(2017), 65-87.
- Song, J. S. and S. W. Lee, "Automatic Construction of Positive/Negative Feature-Predicate Dictionary for Polarity Classification of Product Reviews." *Journal of KIISE: Software and Applications*, Vol.38, No.3 (2011), 157-168.
- Steyvers, M. and T. L. Griffiths, "Probabilistic Topic Models." *Handbook of Latent Semantic Analysis*, Vol.427, No.7(2007), 424-440.
- Wang, W., Y. Feng and W. Dai, "Topic Analysis of Online Reviews for Two Competitive Products Using Latent Dirichlet Allocation." *Electronic Commerce Research and Applications*, Vol.29, No.13(2018), 142-156.
- Xu, K., S. S. Liao, J. Li and Y. Song, "Mining Comparative Opinions from Customer Reviews for Competitive Intelligence." *Decision Support Systems*, Vol.50, No.4(2011), 743-754.
- Xu, X., X. Wang, Y. Li and M. Haghghi, "Business Intelligence in Online Customer Textual Reviews: Understanding Consumer Perceptions and Influential Factors." *International Journal of Information Management*, Vol.37,

No.6(2017), 673-683.

Yagci, I. A. and S. Das, "Measuring Design-Level Information Quality in Online Reviews." *Electronic Commerce Research and Applications*, Vol.30, No.10(2018), 102-110.

You, E. S., G. H. Choi and S. H. Kim "Study on Extraction of Keywords Using TF-IDF and Text Structure of Novels." *Journal of the Korea Society of Computer and Information*, Vol.20, No.2(2015), 121-129.

Abstract

Product Evaluation Criteria Extraction through Online Review Analysis: Using LDA and k-Nearest Neighbor Approach

Ji Hyeon Lee* · Sang Hyung Jung** · Jun Ho Kim*** ·
Eun Joo Min**** · Un Yeong Yeo***** · Jong Woo Kim*****

Product evaluation criteria is an indicator describing attributes or values of products, which enable users or manufacturers measure and understand the products. When companies analyze their products or compare them with competitors, appropriate criteria must be selected for objective evaluation. The criteria should show the features of products that consumers considered when they purchased, used and evaluated the products. However, current evaluation criteria do not reflect different consumers' opinion from product to product. Previous studies tried to used online reviews from e-commerce sites that reflect consumer opinions to extract the features and topics of products and use them as evaluation criteria. However, there is still a limit that they produce irrelevant criteria to products due to extracted or improper words are not refined. To overcome this limitation, this research suggests LDA-k-NN model which extracts possible criteria words from online reviews by using LDA and refines them with k-nearest neighbor.

Proposed approach starts with preparation phase, which is constructed with 6 steps. At first, it collects review data from e-commerce websites. Most e-commerce websites classify their selling items by high-level, middle-level, and low-level categories. Review data for preparation phase are gathered from each middle-level category and collapsed later, which is to present single high-level category. Next, nouns, adjectives, adverbs, and verbs are extracted from reviews by getting part of speech information using morpheme analysis module. After preprocessing, words per each topic from review are shown with LDA

* Department of Philosophy, College of Humanities, Hanyang University

** School of Business, Hanyang University

*** Department of Mathematics, College of Natural Sciences, Hanyang University

**** School of Finance, Hanyang University

***** School of Business Informatics, Hanyang University

***** Corresponding Author: Jong Woo Kim

School of Business, Hanyang University

222 Wangshimni-ro, Seongdong-gu, Seoul 04763, Korea

Tel: +82-2-2220-1067, Fax: +82-2-2220-1169, E-mail: kjw@hanyang.ac.kr

and only nouns in topic words are chosen as potential words for criteria. Then, words are tagged based on possibility of criteria for each middle-level category. Next, every tagged word is vectorized by pre-trained word embedding model. Finally, k-nearest neighbor case-based approach is used to classify each word with tags.

After setting up preparation phase, criteria extraction phase is conducted with low-level categories. This phase starts with crawling reviews in the corresponding low-level category. Same preprocessing as preparation phase is conducted using morpheme analysis module and LDA. Possible criteria words are extracted by getting nouns from the data and vectorized by pre-trained word embedding model. Finally, evaluation criteria are extracted by refining possible criteria words using k-nearest neighbor approach and reference proportion of each word in the words set.

To evaluate the performance of the proposed model, an experiment was conducted with review on '11st', one of the biggest e-commerce companies in Korea. Review data were from 'Electronics/Digital' section, one of high-level categories in 11st. For performance evaluation of suggested model, three other models were used for comparing with the suggested model; actual criteria of 11st, a model that extracts nouns by morpheme analysis module and refines them according to word frequency, and a model that extracts nouns from LDA topics and refines them by word frequency. The performance evaluation was set to predict evaluation criteria of 10 low-level categories with the suggested model and 3 models above. Criteria words extracted from each model were combined into a single words set and it was used for survey questionnaires. In the survey, respondents chose every item they consider as appropriate criteria for each category. Each model got its score when chosen words were extracted from that model. The suggested model had higher scores than other models in 8 out of 10 low-level categories. By conducting paired t-tests on scores of each model, we confirmed that the suggested model shows better performance in 26 tests out of 30. In addition, the suggested model was the best model in terms of accuracy.

This research proposes evaluation criteria extracting method that combines topic extraction using LDA and refinement with k-nearest neighbor approach. This method overcomes the limits of previous dictionary-based models and frequency-based refinement models. This study can contribute to improve review analysis for deriving business insights in e-commerce market.

Key Words : product evaluation criteria, review analysis, extracting evaluation criteria, LDA, k-NN

Received : August 19, 2019 Revised : February 5, 2020 Accepted : February 28, 2020

Publication Type : Regular Paper Corresponding Author : Jong Woo Kim

저 자 소개



이지현

현재 롯데정보통신 IoT부문에 재직중이다. 한양대학교에서 철학과 경영학을 전공했으며, 한양대학교 빅인텔리전트 경영교육 사업단 주관의 프로젝트 학기제를 통해 (주)한양비즈랩(Business & AI Lab) 인턴 과정을 수료하였다. 주요 관심분야는 빅데이터 마케팅, 기계학습 및 딥러닝 기법의 비즈니스 활용, 디지털 고객경험 등이다.



정상형

현재 한양대학교 경영학부 학생으로 재학 중이다. 한양대학교 빅 인텔리전트 경영 교육 사업단에서 주관하는 프로젝트 학기제 인턴 과정에서 Business AI Lab에서 근무하였다. 주요 관심분야는 빅데이터의 활용, 기계학습 및 딥러닝 활용 및 응용, 데이터 마이닝 등이다.



김준호

현재 취업준비 중이다. 한양대학교에서 수학학사를 취득하였으며 한양대학교 빅 인텔리전트 경영교육 사업단에서 주관하는 프로젝트 학기제 인턴 과정을 수료하였다. 주요 연구 관심분야는 데이터 마케팅, 빅데이터, 딥러닝 등이다.



민은주

현재 (주)SK의 금융/전략Digital 추진그룹에서 재직 중이다. 한양대학교 파이낸스경영학과를 졸업하였고, 한양대학교 빅 인텔리전트 경영 교육 사업단에서 주관하는 프로젝트 학기제 인턴 과정에서 Business AI Lab에서 근무하였다. 주요 연구 관심분야는 금융 산업에서의 기계학습 및 딥러닝 기법 활용, 데이터마이닝 등이다.



여운영

현재 (주)SK에서 Tech. Pre-Sales그룹에 재직 중이다. 한양대학교 일반대학원 비즈니스인포매틱스학과에서 경영학석사를 취득하였다. 주요 연구 관심분야는 자연어처리, 자연어 이해, 딥러닝 기법 활용 및 응용 등이다.



김종우

현재 한양대학교 경영대학 경영학부 교수로 재직 중이다. 서울대학교 수학과에서 학사를 마쳤으며, 한국과학기술원에서 경영과학으로 석사학위를, 산업경영학으로 박사학위를 취득하였다. 주요 연구 관심분야는 데이터마이닝 기법과 응용, 기계학습과 딥러닝, 오피니언 마이닝, 상품추천기술, 지능형 정보시스템, 집단지성, 사회 네트워크 분석, 클라우드 컴퓨팅 서비스 등이다.