

상품명 및 리뷰를 기반으로 한 브랜드-트렌드 연관성을 통한 이커머스 경쟁력 강화

신기영^o, 정현영^o
한양대학교^o, 디자인노블
skyworldgo@hanyang.ac.kr^o, jhyoung@designovel.com

Enhancing E-commerce Competitiveness through Brand-Trend Association Based on Product Names and Reviews

Ki-young Shin^o, Hun-young Jung^o
Hanyang University^o, Designovel

요 약

본 연구는 브랜드가 시장 트렌드를 파악하고 이를 활용하여 경쟁 우위를 확보하고 성장하는 방법을 탐구하고 있다. 이를 위해 세 가지 핵심 요소를 고려하였다. 첫째, 시장의 트렌드 정보를 파악하기 위해 검색 포털 사이트의 검색어 랭킹 정보를 활용하였다. 둘째, 브랜드 상품과 트렌드의 연관성을 분석하기 위해 상품 타이틀과 리뷰 데이터를 활용하였다. 셋째, 각 상품의 브랜드 중요성을 추정하기 위해 리뷰 수, 리뷰 길이, 표현의 다양성 등을 고려했다. 연구 결과, 브랜드는 시장 트렌드를 더욱 정확하게 이해하고 파악함으로써 경쟁 우위를 확보하고 성장할 수 있는 기회를 제공함을 확인하였다. 더불어, 이를 통해 브랜드는 소비자의 요구를 더욱 효과적으로 충족시키고 고객 경험을 개선하는데 기여할 수 있을 것으로 기대된다.

주제어 : 워드 임베딩 (Word Embedding), 시계열 분석 (Time Series Analysis), Bag of Words (BoW) 모델

1. 서론

현대 비즈니스 환경에서 브랜드의 성공은 빠르게 변화하는 소비자 취향과 요구사항에 대응하는 능력에 달려 있다. 특히 건축 및 인테리어 분야는 초기 B2B(비즈니스 투 비즈니스)로 여겨졌지만, 최근에는 B2C(비즈니스 투 컨슈머) 소비자들의 관심과 수요가 높아지고 있다. 이러한 변화로 인해 한국과 미국을 포함한 다양한 시장에서 가구 단위의 인테리어 공사와 관련된 시장에서 중요한 트렌드와 브랜드 정보에 대한 수요가 급증하고 있다.

이로 인해 건축 및 인테리어 분야의 사업자들은 더욱 더 적극적으로 인터넷을 통해 최신 트렌드와 각 브랜드의 제품 정보를 찾고, 고객들의 요구를 충족시키기 위해 노력해야 한다. 브랜드는 이러한 변화된 환경에서 경쟁 우위를 유지하고 성장하기 위해서는 시장 트렌드에 대한 신속하고 정확한 이해가 필수적이다.

이 논문은 브랜드와 시장 트렌드 간의 관계를 조사하기 위해 시장 트렌드 정보, 브랜드 상품의 트렌드 연관성, 상품의 중요성 세 가지 주요 요소를 고려한다. 이를 통해 건축 및 인테리어 분야의 브랜드와 사업자들은 경쟁에서 성공하고 변화하는 환경에서 기회를 잡을 수 있을 것이다.

2. 관련 연구

2.1. 이커머스 상품명과 리뷰 및 시계열 분석

이커머스와 관련하여 키워드 및 트렌드, 리뷰 등을 분석하는 방법으로 고객 또는 사업자의 경쟁력을 갖출 수 있도록 하는 다양한 논문들이 출판되고 있다. 최근 연구 동향을 중심으로 몇가지 논문의 연구를 참조하고자 한다. 첫 번째 논문은 소셜 미디어와 금융 시장의 상호작용을 탐구하여 투자와 정보 전달의 관계를 밝히고 있다는 점에서 온라인의 정보가 충분히 트렌드 및 의사결정에 영향력이 있다는 것을 보여주고 있다[1]. 또한 뉴스 기사 데이터를 활용하여 기업의 비즈니스 트렌드 모델링을 다루며, 기업의 경영 전략에 대한 통찰력을 얻을 수 있다는 논문을 통해서[2] 일련의 데이터를 중요하게 다루는 것이 충분한 의미가 있음을 알 수 있다. 다른 논문은 전자 상거래 플랫폼의 검색 결과 품질 향상을 위한 솔루션을 논의하며, 이는 전자 상거래 분야의 검색 및 추천 시스템 등 오랫동안 연구되어온 분야로서 방법론이나 정당성에 대하여 충분히 합의된 수준의 기술적 방법론이 존재한다는 점을 알 수 있다[3]. 뿐만 아니라 주요 패션 전자 상거래 기업의 입찰 전략 최적화를 다루며, 입찰 전략의 효율성 향상에 관한 실용적인 정보를 제공한다. 이 점에서 온라인 커머스를 운영하는 사업자가 일련의 정보를 의미있게 활용할 수 있음을 유추할 수 있다[4]. 최근 생성형 모델이 대두된 이후 커머스 또한 대화

형태로 진행할 수 있다는 점이 검토되고 있다는 점에서 이미지 및 텍스트 등 데이터를 활용하여 원하는 상품을 찾을 수 있도록 도움을 주는 행위[5]에 대한 다양한 연구들을 살펴볼 수 있다.

2.2. 시계열 데이터 기반 트렌드 추출

시계열 데이터는 일반적으로 추세, 계절성, 잔차의 요소를 포함하고 있어 트렌드를 분석하기 위해서는 데이터 상에서 계절성과 잔차의 요소를 제거하고 트렌드만 분리해낼 필요가 있다.

트렌드 분리에서 가장 기본적으로 사용되는 방법은 이동평균(Moving Average)로 계절성의 주기를 커버하는 충분히 큰 k 에 대해서 $-k \sim k$ 시점의 데이터를 평균하는 것으로 계절성의 영향과 잔차의 임의성을 희석시켜 트렌드만 남기는 방식이다.

X11 모델[6]은 이동평균으로 구한 트렌드를 제거한 계절성-잔차 데이터에 다시 이동평균을 적용하여 계절성 데이터를 분리하고 이를 다시 원 데이터에서 제거한 트렌드-잔차 데이터에 이동평균을 적용하는 과정 전체를 반복하는 것으로 더 정교하게 트렌드 데이터를 추출한다. SEATS[7] 방법론은 ARIMA[8] 시계열 모델링 기법과 스펙트럼 분석을 반복적으로 적용하는 것으로 트렌드를 분리하고, STL [9] 모델은 데이터를 구간별로 나누어 구간별로 최소자승회귀(Least Squares Regression) 방법으로 구간별 추세를 구하고 이를 smoothing으로 연결하는 방식으로 전체의 트렌드 요소를 분리한다.

2.3. 단어의 벡터 표현

고전적으로 단어-문서 등장 여부를 가지고 계산하는 Bag of Word나 단어-문서 등장 빈도수에 기반하는 TF-IDF [10] 방법의 경우 단어와 문서, 혹은 단어와 토픽과의 연관성 정보를 표현하나 단어의 문법적, 의미적 정보를 표현하는 것이 부족했다.

대량의 말뭉치와 신경망 기법이 발달하면서 단어의 문법적, 의미적 정보를 벡터화 하여 표현하는 기법이 발달하였다. 첫 번째로 사용된 것은 Word2Vec [11] 로 목표 단어 주변 단어에 기반하여 목표 단어를 찾는 CBOW 방식과, 반대로 목표 단어에 기반하여 주변 단어를 찾는 Skip-gram 방식이 있다. 주변단어를 직접 접근하는 Word2Vec와 조금 다르게 GloVe [12] 는 두 단어 벡터 표현의 내적이 두 단어의 co-occurrence가 되도록 하여 전체 말뭉치의 확률 정보를 사용하였다.

위 방법론들은 일정 window size 내에서 단어를 bag-of-word로 간주하여 주변 정보를 정밀하게 사용하지 않는다. Context2Vec [13]나 ELMo [14] 의 경우 주변 단어의 sequence를 LSTM, 혹은 Bi-LSTM으로 받아 목표 단어를 찾는데 사용하여 대량의 말뭉치에 기반한 단어 표현을 생성하였다.

3. 접근 방법

브랜드의 시장 트렌드 추종 정도를 알기 위하여 다음과 같이 세가지 요소를 고려하였다. 첫 번째 요소는 시장의 트렌드 정보이고 두 번째 요소는 브랜드 상품과 트렌드의 연관성, 세 번째 요소는 각 상품의 브랜드 중요도이다.

이 세가지 요소를 파악하기 위해 사용된 정보는 다음과 같다. 검색 포털 사이트의 검색어 랭킹 정보를 사용하여 시장 트렌드 정보를 파악하고, 검색어와 브랜드 상품의 상품명 및 리뷰데이터에 기반하여 브랜드 상품별 트렌드 연관성을 계산하며, 각 상품의 리뷰 숫자, 길이 및 표현의 다양성에 기반하여 브랜드에서의 상품의 중요성을 추정하였다.

3.1. 시장 트렌드

시장의 트렌드 흐름에 대한 정보를 알기 위해 검색 포털 사이트에서 제공하는 키워드 검색량에 대한 데이터를 사용하였다. 이 때 계절성, 주기성에 영향 받는 부분을 제외하기 위하여 이전해 동월 데이터와 비교하여 변화량을 계산하는 방식을 사용하였다. 신규 트렌드에 대한 키워드는 검색량 데이터가 충분히 길지 않아 기존의 시계열 분석 방법을 키워드 검색량을 직접 적용하기에는 적절하지 않아 키워드 검색량을 직접 사용하기 보다는 키워드의 카테고리 검색어 랭킹의 변화를 사용하였다.

키워드 k 에 대한 트렌드 점수(TR)의 계산 방식은 다음과 같다.

$$TR(k) = \frac{1}{|M|} \sum_{m \in M} \{Rank(k, m, y) - Rank(k, m, y-1)\}$$

이 때 y 는 트렌드 계산의 기준이 되는 해이고 m 은 기준 월이 되며 Rank는 m, y 시점에 대한 k 의 랭킹이 된다.

3.2. 브랜드 상품의 트렌드 연관성

상품의 타이틀과 리뷰데이터를 사용하여 상품의 트렌드 연관성을 계산하기 위해 상품 타이틀 및 리뷰 데이터를 하나의 문서로 간주하고 이문서의 Bag of Tokens에서 트렌드 키워드의 발생 확률을 계산하여 연관성으로 사용하였다.

이 때 키워드와 문서 토큰 사이에서 발생하는 사전 불일치 문제를 해결하기 위해 직접 일치로 비교하지 않고 문서 토큰의 발생빈도 및 토큰과 키워드의 유사도를 사용하였다.

트렌드 키워드 k 와 브랜드 상품 p 의 연관성(SR)은 다음과 같다.

$$SR(k, p) = \frac{1}{|D_p|} \sum_{w \in D_p} TP(w, D_p) * S(w, k)$$

이 때 D 는 상품 p 의 타이틀 및 리뷰 문서를, TP 는 문서 D 에 등장하는 토큰 w 의 출현확률을, S 는 두 토큰 w 와 k 의 유사도를 의미한다.

두 토큰 w 와 k 의 유사도를 계산하기 위해 word2vec[11]의 Skip-gram 모델을 사용하였다. 수집한 리뷰데이터에서 모델을 학습하여 각 단어 벡터 표현의 cosine 값을 유사도로 사용하였다.

3.3. 브랜드에서의 상품의 중요성

상품이 소속 브랜드에서 많은 비중을 차지할수록 리뷰가 많이 달리고 리뷰에 사용된 표현의 정보량이 많을 것으로 간주하여 브랜드의 전체 리뷰 수 대비 상품의 리뷰 수 비율과 리뷰 전체의 shannon entropy[15] 값을 사용하여 다음과 같은 식으로 정의하였다.

$$BI(p) = \frac{Count(R_p)}{\sum_{p \in B} Count(R_p)} \times \sum_{w \in D_p} p(w) \log_2 \frac{1}{p(w)}$$

이 때 R 은 상품의 개별 리뷰로 상품 p 에 대한 모든 리뷰와 상품 타이틀 텍스트를 합하면 D 가 된다.

4. 실험 조건 및 결과

키워드 랭킹 정보를 위해 N사 포털에서 제공하는 키워드 랭킹을 2021년 8월 ~ 2023년 8월 구간에 대해서 2월, 5월, 8월, 11월을 기준월로 삼아 총 8개 달에 대해 카테고리별 키워드 랭킹 500위까지 정보를 사용하였다. 이 때 기준 카테고리로 사용한 것은 다음과 같다: 가구/인테리어, 거실가구, 침실가구, 주방가구, 홈데코, 아웃도어가구.

이 실험 조건에서 계산된 트렌드 점수 상위 4개 키워드와 하위 4개 키워드 결과는 표 1과 같다.

인테리어 상품 및 리뷰 데이터 수집을 위해 0사의 인테리어 상품 비교사이트에서 정보를 수집하였다. 각 브랜드별 상품에 대한 리뷰데이터 수집 결과는 다음과 같다: 72개 브랜드, 135개 상품, 275,774개 리뷰.

이 중 4개 이상의 상품이 수집되어 트렌드 분석 대상으로 선정한 브랜드의 데이터는 다음과 같다: 7개 브랜드 46개 상품, 90,287개 리뷰.

표 1. 트렌드 점수 상위/하위 4개 단어

키워드	트렌드 점수
가정용방음부스	488.75
서빙카트	479.00
렉산폴리카보네이트	448.25
복층렉산	448.00
아치형전신거울	-303.75
와이드전신거울	-334.25
등받이삼각쿠션	-346.00
침대메트리스	-360.50

표 2. 브랜드 별 트렌드 점수 산출 결과

브랜드 번호	상품 수	리뷰 수	트렌드 점수
1	5	10.2K	0.0093
2	5	10.4K	0.0077
3	10	19.1K	0.0076
4	4	7.9K	0.0039
5	13	22.8K	0.0024
6	4	9.2K	-0.0038
7	5	10.7K	-0.0031

word2vec 모델을 학습하는데 있어서 135개 상품, 275,774개 리뷰 정보 전체를 사용한다. 상품 정보 및 리뷰 통합 문서 분석에 사용되는 토큰나이저로는 K사에서 개발한 한국어 GPT[16]의 토큰나이저를 사용하였다. 토큰 출현 빈도수를 계산하는데 있어서 상품 타이틀에 나오는 단어가 리뷰에 나오는 단어보다 더 중요하기에 상품 타이틀의 토큰 출현 횟수를 리뷰 개수 만큼 배수로 계산하였다. 이렇게 사용된 리뷰 문서 약 276K 개 문서의 총 토큰 수는 약 15M 로 토큰 사전의 크기는 약 23.9K 개가 된다.

이렇게 구축된 데이터에 기반하여서 각 브랜드별 트렌드 점수를 계산하여 표 2와 같이 산출하였다.

5. 결론

이 연구를 통해 브랜드는 시장 트렌드를 더욱 정확하게 이해하고 파악할 수 있으며, 이를 통해 경쟁 우위를 유지하고 성장할 수 있는 기회를 확인하였다. 또한, 이를 통해 브랜드는 소비자의 요구를 더욱 효과적으로 충족시키고 고객 경험을 개선하는데 기여할 수 있을 것으로 기대한다.

그러나 이번 연구에는 몇 가지 제한사항이 있다. 첫째, 트렌드 정보의 정확성과 신뢰성은 검색 포털 사이트의 데이터에 의존하고 있으며, 이 데이터의 한계나 왜곡 가능성을 고려해야 한다. 둘째, 상품의 트렌드 연관성 및 브랜드 중요성을 계산하기 위한 방법론에는 여전히 발전의 여지가 남아 있으며, 더욱 정교한 모델링과 분석이 필요하다. 마지막으로, 이 연구에서 다루지 못한 다른 요인들도 브랜드의 성공에 영향을 미칠 수 있으므로, 미래 연구에서는 이러한 다양한 변수들을 더 포함시키는 방향으로 확장되어야 한다.

이러한 한계를 고려하면서도, 본 연구는 브랜드와 시장 트렌드 간의 관계를 탐구하고 브랜드의 경쟁력을 높이는 데 중요한 인사이트를 제공하였으며, 미래의 연구 및 실무 적용을 위한 기초를 마련하는데 기여할 것으로 기대한다.

참고문헌

- [1] PANDEY, Aditya; FATHIYA, Haseeba; PATEL, Nivedita. Cross-Domain Shopping and Stock Trend Analysis. arXiv preprint arXiv:2212.14689, 2022.
- [2] ARSLAN, Muhammad; CRUZ, Christophe. Unlocking Insights into Business Trajectories with Transformer-based Spatio-temporal Data Analysis. arXiv preprint arXiv:2306.10034, 2023.
- [3] WU, Fanyou, et al. Some Practice for Improving the Search Results of E-commerce. arXiv preprint arXiv:2208.00108, 2022.
- [4] PROVODIN, Danil; JOUDIUX, Jérémie; DURYEV, Eduard. Learning Optimal Bidding Strategy: Case Study in E-Commerce Advertising. arXiv preprint arXiv:2304.00999, 2023.
- [5] LI, Jianri, et al. Designovel's system description for Fashion-IQ challenge 2019. arXiv preprint arXiv:1910.11119, 2019.
- [6] BELL, William R.; HILLMER, Steven C. Issues involved with the seasonal adjustment of economic time series. *Journal of Business & Economic Statistics*, 2.4: 291-320. 1984,
- [7] GÓMEZ, Víctor; MARAVALL, Agustín. Seasonal adjustment and signal extraction in economic time series. *A course in time series analysis*, 202-247. 2001,
- [8] BOX, George EP, et al. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [9] RB, CLEVELAND. STL: A seasonal-trend decomposition procedure based on loess. *J Off Stat*, 6: 3-73. 1990,
- [10] SPARCK JONES, Karen. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28.1: 11-21. 1972,
- [11] MIKOLOV, Tomas, et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
- [12] PENNINGTON, Jeffrey; SOCHER, Richard; MANNING, Christopher D. Glove: Global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. p. 1532-1543. 2014.
- [13] MELAMUD, Oren; GOLDBERGER, Jacob; DAGAN, Ido. context2vec: Learning generic context embedding with bidirectional lstm. In: *Proceedings of the 20th SIGNLL conference on computational natural language learning*. p. 51-61. 2016.
- [14] PETERS, Matthew E., et al. Deep contextualized word representations. *NAACL-HLT*. arXiv, 2018.
- [15] Shannon CE. A mathematical theory of communication. *The Bell system technical journal*. Jul;27(3), pp. 379-423.1948.
- [16] Ildoo Kim, et al. KoGPT: KakaoBrain Korean(hangul) Generative Pre-trained Transformer. 2021.