

텍스트마이닝을 활용한 여성 스포츠웨어 상품명에 나타난 키워드 분석

Text mining analysis of keywords on product names used in women's sportswear

김 수 민, 이 규 혜*

한양대학교 의류학과 박사과정, 한양대학교 의류학과 교수*

Sumin Kim¹, Kyu-Hye Lee*

PhD Student, Dept. of Clothing & Textiles, Hanyang University, Seoul, Korea

Professor, Dept. of Clothing & Textiles, Hanyang University, Seoul, Korea*

Abstract

As the realm of online shopping expands and diverse products flood the market, the significance of keywords in facilitating product search and information dissemination becomes increasingly vital. This study delves into the compositional characteristics and informational types of keywords associated with women's sportswear. The objective is to discern the prevalent sports mentioned in women's sportswear and to elucidate the nuanced attributes of these terms. Employing text mining techniques, data spanning from 2021 to 2023 was meticulously gathered from Naver Shopping. The data underwent TEXTOM collection, followed by comprehensive analyses including frequency analysis, N-gram analysis, TF-IDF analysis, and matrix analysis. To unravel the intricate network structure and relationships between keywords, centrality metrics (connectivity centrality, betweenness centrality, closeness centrality, Eigenvector centrality, etc.) were employed. Additionally, ego network and CONCOR cluster analysis were conducted. After excluding irrelevant data, an exhaustive analysis of 8,657 product names was performed, revealing the highest frequency keywords in the following order: women's, yoga, sports bra, golf, T-shirt, leggings, Pilates, jogger pants, sports, women, and innerwear. Among the top 100 sports mentioned, yoga, golf, pilates, health, training, fitness, swimming, flying yoga, pole dance, aerobics, running, cycling, mountain climbing, and zumba dance emerged as the most frequently associated. N-gram network analysis brought to light intriguing co-occurrences between terms, with notable pairs including 'women-golf,' 'yoga-pilates,' and 'leggings-yoga.' The significance of this study lies in its provision of reference data for research on product information analysis. By exemplifying the analysis of product names using big data, this research serves as a valuable resource for future studies delving into the intricate realm of product information.

Key words : women's sportswear, text mining, consumer perception, big data

주제어 : 여성 스포츠웨어, 텍스트마이닝, 소비자 인식, 빅데이터

I. Introduction

상품명에 포함된 제품정보는 검색을 하는 과정에서 소비자와 상품을 연결하는 고리로써 온라인 쇼핑의 시작점이 된다. 온라인 쇼핑에서는 스타일, 아이템, 치수, 색상 등 원하는 정보에 해당하는 키워드로 상품을 검색하여 나오는 정보를 확인하고 구매를 결정한다. 상품명에는 일반적으로 스타일, 사이즈, 소재, 색상 등의 정보나 브랜드 또는 아이템 이

름 등의 정보가 제공된다. 상품명은 목적은 제품을 판매하는 것이며. 판매자가 붙인 상품의 명칭은 소비자를 설득해 구매를 자극한다. 상품명은 광고수단으로써 타사 제품 및 브랜드와 차별되어야 하며, 소비자에게 긍정적인 연상을 불러일으켜야 한다(Kim, 2005). 소비자들은 제품의 장식, 실용성, 목적과 형태 뿐 아니라 상품의 이름 역시 구매의 중요 기준으로 삼는다. 시대에 맞는 소비자의 요구를 바탕으로 상품명에 정하기 때문에 사회의 트렌드를 반영하는 경향이 높으며, 매출을 높이는 방향으로 작명이 된다(Hong & Song, 2015). 의류의 상품명에 포함된 정보와 용어를 구체

* Corresponding Author ; Kyu-Hye Lee

Tel : 82-2-2220-1191, E-mail : khlee@hanyang.ac.kr

적으로 분석하고, 소비자의 니즈를 잘 파악한다면, 구매자와 판매자의 효율적인 연결을 위한 방법 모색에 유용한 자료로 사용할 수 있을 것이다.

인터넷 쇼핑물의 상품명의 특징을 보면, 소재나 구성과 관련된 내용은 명사로, 소재나 이미지의 느낌과 관련된 내용은 주로 형용사로 표현된다. 이는 판매자가 주관적으로 선택한 경우가 대부분이므로 같은 정보를 의미한다고 해도 단어의 현대에 따라 검색 결과가 달라질 수 있다(Shim & Kim, 2008). 이처럼 판매자가 사용하는 용어의 다양한 표현 방식 및 불명확한 정보 제공은 소비자와 제품의 연결을 방해할 수 있으며, 부정확한 정보로 인한 잘못된 기대치 형성으로 쇼핑의 만족도를 떨어뜨릴 수 있기 때문에 상품명 선정은 매우 중요한 부분이다. 따라서 패션 브랜드는 소비자와의 효율적인 소통이 가능한 용어 및 표기법으로 product name을 구성하여야 하며, 이를 위해서 product name의 특징을 제대로 파악할 필요가 있다. 온라인 쇼핑의 규모가 커지며 다양한 제품의 빠른 등장에 따라 제품 검색 및 정보 전달을 담당하는 키워드의 중요성은 계속해서 커지고 있다. 또한, 여성 스포츠웨어 시장은 계속해서 성장하고 있으며, 최신 트렌드와 소비자의 선호를 파악하는 것이 중요하며, 어떤 유형의 스포츠웨어를 찾고 있는지 이해하는 것은 제품 개발 및 마케팅 전략에 매우 중요하다. 따라서, 본 연구에서는 여성 스포츠웨어에 등장하는 키워드의 구성 특성과 정보의 종류를 분석해 현재 여성 스포츠웨어에서 가장 많이 언급되는 스포츠는 어떤 것인지 파악하고, 용어의 특징을 구체적으로 파악하는 것이 목적이다. 본 연구의 결과는 관련 업계에서 유용한 정보로 활용될 것으로 사료된다.

II. Theoretical background

상품명은 소비자의 요구를 명확히 파악하고 이름을 정해야 하기 때문에 사회의 트렌드나 요구를 반영하고 있을 가

능성이 높다. 소비자의 니즈를 잘 반영하면서 상품의 정보를 확실하게 전달할 수 있게 제품의 이름을 선정할 필요가 있다. 특히, 온라인에서 제품을 구매하는 경우에는 실제 상품을 직접적으로 확인하지 못하고 구매를 결정해야 하기 때문에 product name의 영향력이 더욱 크다고 할 수 있다. 상품명에 대한 선행연구는 ‘브랜드 네이밍’이라는 이름으로 경영학 분야나 광고학 분야에서 활발하게 이루어져 왔다. 국내 화장품 브랜드의 홈페이지에 나타난 상품명 분석을 진행한 연구를 통해 화장품 제품의 product name을 정할 때는 소비자의 피부 고민을 반영해 상품 네이밍을 한다는 것을 알 수 있었다(Jeong&Lee, 2016). Kang(2023)의 여성 청바지 상품명에 나타난 키워드 분석 연구에서는 텍스트마이닝 기법을 활용해 분석을 하였으며, 식품 상품명과 관련한 Hong and Song(2015)의 연구에서는 product name의 언어적 특징 및 사회문화적 의미에 대해 알아보았다. 이처럼 상품명과 관련해 다양한 선행연구가 이뤄졌다. 상품의 검색은 소비자를 중심으로 이루어지기 때문에 제품의 구매를 원하는 소비자가 원하는 단어를 포함하여야 소비자에게 쉽게 노출될 수 있다.

온라인 쇼핑이 확대되며 패션 브랜드는 소비자들이 구매 과정에서 남긴 검색이나 구매 관련 빅데이터로부터 소비자의 반응과 관심을 분석하고자 노력한다. 이러한 데이터들은 형태가 다양하며 수시로 생성된다. 규모도 방대할 뿐만 아니라 실제 소비자의 생각을 내포하고 있어서 분석할 가치가 있다. 연구에 필요한 정보를 데이터를 추출하기 위해서는 빈도분석, TF-IDF분석, N-gram 네트워크 분석, 연결중심성, 공출현 빈도분석, CONCOR분석, 매트릭스 도출 등의 방법이 활용된다. 본 연구의 결과를 통해 구매가 많이 이루어진 제품의 상품명에는 포함된 스포츠 종목은 어떤 것이 있는지 알아보고, 소비자의 관심을 파악하여 상품명에 반영된 특징을 파악하여 제품 이름 구성 방안에 기초자료를 제시하고자 한다. 소비자가 온라인에 남긴 흔적을 데이터화하여 상품에 대한 소비자의 관심과 반응을 유의미하게 분석하는

텍스트마이닝 연구는 패션 분야에서 활발하게 진행되고 있다. 패션 라이브 커머스 유형별 소비자의 인식 비교 연구에서도 텍스트마이닝이 적용되었고(Gwak & Lee, 2021), 패션 스트리밍 서비스인식에 관한 연구(Kim, 2018), 패션 유튜브 명품 후기 콘텐츠에 대한 소비자의 댓글 분석 연구에서도 텍스트마이닝이 활용되었다(Kim et al., 2022). 온라인 쇼핑 환경 안에서 존재하는 텍스트데이터 분석이 활성화되면서 실제 활용이 가능한 가치정보로 변하고 있다. 데이터는 다양한 형태와 방대한 규모로 수시로 생성되며, 실제 소비자의 생각 역시 내포하고 있기 때문에 이러한 접근 방법의 의미가 크다(Kang, 2023).

III. Method

1. Data collection

국내 1위 포털사이트 네이버의 ‘네이버 쇼핑’에 나타난 목록에서 데이터를 수집하였다. 특정 집단이나 상품을 한정된 플랫폼보다 네이버 쇼핑에서는 폭넓은 사용자를 대상으로 데이터 추출이 가능하기 때문이다. 네이버 쇼핑의 상품 나열 조건은 ‘등록일순’, ‘리뷰 좋은순’, ‘리뷰 많은순’, ‘낮은 가격순’, ‘높은 가격순’으로 정렬할 수 있으나 본 연구에서는 실제 구매가 많이 이루어지는 제품의 상품명을 파악하는 것이 목적이기 때문에 결과의 활용도를 높이고자 조건을 ‘리뷰 많은순’으로 설정하고 수집하였다. 등록된 3백만개 이상의 제품을 ‘리뷰 많은순’으로 나열하고 최근 2년의 상위 10,000개의 상품명을 직접 수집하였으며, 2021~2023년 이외의 데이터 및 중복데이터와 불필요한 데이터는 삭제하였다.

2. Data analysis

텍스트를 활용해 정제 및 분석을 실시하였고, 정제과정에서 제외된 상품명은 총 1,343개로 본 연구에서는 최종적으로

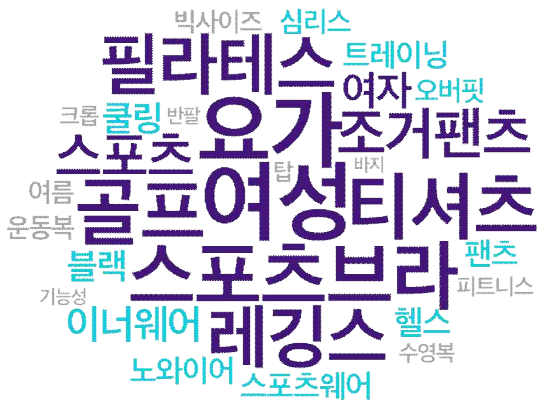
8,657개의 상품명을 가지고 분석을 진행하였다. 이후 텍스트의 분석 툴인 Espresso K를 활용해 상품명과 관련된 유의미한 키워드 추출하였다. 무의미한 숫자나 조사, 감탄사 등의 용어를 삭제하였으며, 복수 표현은 상위빈도의 단어로 통일하였다. ‘요가운동복’, ‘요가웨어’, ‘핫요가’, ‘요가복’ 등은 ‘요가’로 정하였으며, ‘필라테스옷’, ‘필라테스복장’, ‘필라테스복’, ‘필라테스운동복’ 등은 ‘필라테스’로, ‘라운딩룩’, ‘골프복’, ‘GOLF’, ‘골프웨어’, ‘여자골프웨어’, ‘여성골프웨어’ 등은 골프로, ‘여성런닝’, ‘런닝복’, ‘러닝’, ‘조깅복’, ‘여자런닝’, ‘조깅복여자’ 등은 ‘런닝’으로 정제하였다. 여성 스포츠웨어의 특징과 중요 키워드 파악을 위해 TF-IDF분석, 빈도분석, 매트릭스 분석, N-gram분석을 실시하였다. 다음으로 키워드의 연결 관계 및 네트워크 구조를 파악하기 위해 중심성 분석을 실시하였으며, 세부적으로는 연결중심성, 근접중심성, 매개중심성을 분석하였다.

IV. Results

최종 추출된 키워드의 빈도분석 결과, 다음의 순서로 빈도가 높게 나타났다. ‘여성’, ‘요가’, ‘스포츠브라’, ‘골프’, ‘티셔츠’, ‘레깅스’, ‘필라테스’, ‘조거팬츠’, ‘스포츠’, ‘여자’, ‘이너웨어’가 1% 이상 출현 키워드로 나타났으며, 0.5% 이상의 출현 키워드는 28개로 확인되었으며, 0.1% 이상으로 출현한 키워드는 125개로 나타났다. 상위 100위 안에 등장한 스포츠 종목은 빈도가 높은 순으로 ‘요가’(2위), ‘골프’(4위), ‘필라테스’(7위), ‘헬스’(14위), ‘트레이닝’(15위), ‘피트니스’(26위), ‘수영’(27위), ‘플라잉요가’(40위), ‘폴댄스’(45위), ‘에어로빅’(46위), ‘런닝’(48위), ‘사이클’(52위), ‘등산’(64위), ‘줄바덴스’(87위), ‘스피닝’(92위)로 나타났다.

또한, 높은 순위에서 ‘스포츠 브라’, ‘이너웨어’ 등의 언더웨어 제품에 대한 키워드가 등장했으며, ‘쿨링’, ‘노와이어’, ‘심리스’, ‘기능성’, ‘에어’, ‘기모’, ‘소프텍스’ 등 제품의 기능

이나 소재와 관련된 단어 역시 높은 순위에서 나타났다. 길이와 관련한 키워드에서는 크롭이 5부, 7부, 9부보다 2배 이상 많은 것으로 확인되었다. 컬러는 블랙(13위)이 가장 높았으며, 그 뒤를 이어 화이트(31위), 베이지(50위), 네이비(55위), 핑크(75위), 차콜(76위) 등이 100위 안에 나타났다 (Figure 1).

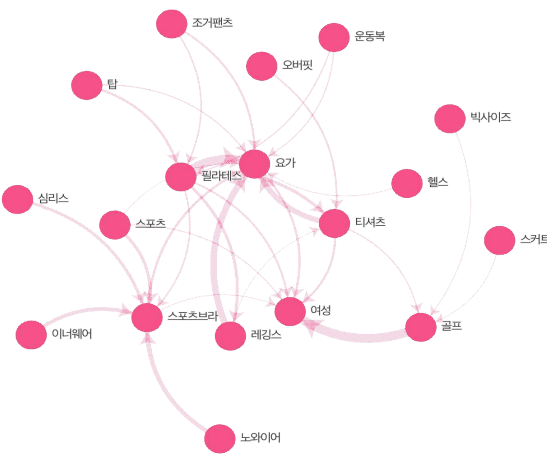


<Figure 1> Frequency analysis

N-gram 네트워크 분석을 통해서 혼용되는 용어 간의 공출현 빈도 분석 결과, ‘여성-골프’, ‘요가-필라테스’, ‘레깅스-요가’, ‘요가-티셔츠’, ‘오버핏-티셔츠’, ‘스포츠브라-이너웨어’, ‘노와이어-스포츠브라’, ‘티셔츠-필라테스’, ‘심리스-스포츠브라’, ‘요가-타입’ 등이 나타났다 (Figure 2). 결과를 통해, 요가복으로는 레깅스와 티셔츠를 많이 착용한다는 것을 알 수 있었고, 스포츠브라에서는 노와이어나 심리스 등 봉제가 없는 디자인을 추구하는 것을 알 수 있었다.

다음으로 중심성 정량화를 위해 연결중심성 분석을 진행하였다 (Table 1). 연결정도중심성 결과, ‘여성’, ‘요가’, 티셔츠’, ‘스포츠’, ‘필라테스’, ‘레깅스’, ‘여자’, ‘블랙’, ‘트레이닝’, ‘스포츠브라’ 등이 상위 키워드로 나타났으며, 근접중심성에서도 같은 결과가 나타났다. 매개중심성에서는 ‘여성’, ‘티셔츠’, ‘필라테스’, ‘스포츠’, ‘레깅스’, ‘블랙’, ‘트레이닝’, ‘스포츠브라’, ‘쿨링’, ‘골프’ 순으로 나타났다. 주요 키워드의 결과를

살펴보면 요가는 빈도 순위가 2위이며, 모든 중심성에서도 2위로 나타나 요가가 여성 스포츠웨어에서 대표 키워드로 확인되었으며, 티셔츠, 필라테스, 스포츠는 중심성에 모두 5위 안에 드는 것으로 확인되었다. 그 중에서도 스포츠브라, 골프, 조거팬츠는 연결중심성, 근접중심성, 매개중심성에서 모두에서 비교적 낮은 순위로 나타났다. 그 이외에 10위권 밖의 키워드 중 여자(11위), 블랙(14위), 트레이닝(16위)는 빈도 순위에서는 10위권 밖으로 나타났지만, 연결중심성과 근접중심성에서는 여자(7위), 블랙(8위), 트레이닝(9위)로 다른 키워드들과 양적 연결과 근거리 연결의 역할이 큰 키워드로 확인되었다. 이처럼 키워드 분석을 통해 여성 스포츠웨어의 특성을 표현할 때 흔하게 사용되는 표현과 스포츠웨어 상품명에 자주 등장하는 스포츠 종목을 파악할 수 있었고, 여성 스포츠웨어 상품명을 구성하는데 있어 이와 같은 특징의 활용이 필요함을 알 수 있다.



<Figure 2> N-gram analysis

출현빈도와 별도로 특정 관계를 보이는 키워드를 분석하기 위해 200개 키워드의 상관계수를 분석하였다. 우선, ‘스포츠브라-노와이어’의 상관인 0.567로 가장 높았고, 2~5위는 ‘필라테스-타입’, ‘심리스-스포츠브라’, ‘필라테스-요가’, ‘주름-스커트’로 나타났다. 그 외 상관관계를 보인 키워드들의 특성을 보면, ‘요가-레깅스’, ‘요가-타입’, ‘필라테스-타입’, ‘오버핏-

티셔츠, ‘반팔-티셔츠’, ‘쿨링-팬츠’ 등 흔히 사용되는 관용구적 표현의 상관이 높은 것으로 나타났다.

<Table 1> Centrality analysis

Frequency		Degree centrality		Closeness centrality		Betweenness centrality	
1	여성	1	.979	1	.980	1	.020
2	요가	2	.969	2	.970	3	.019
3	스포츠브라	12	.884	12	.896	11	.014
4	골프	17	.849	17	.868	16	.012
5	티셔츠	3	.969	3	.970	2	.019
6	레깅스	6	.929	6	.934	6	.016
7	필라테스	5	.944	5	.947	4	.017
8	조거팬츠	23	.728	23	.786	23	.008
9	스포츠	4	.954	4	.956	5	.017
10	여자	7	.924	7	.929	7	.015
11	이너웨어	21	.748	21	.799	20	.009
12	쿨링	15	.869	15	.884	15	.013
13	블랙	8	.909	8	.917	8	.015
14	헬스	16	.854	16	.872	17	.011
15	트레이닝	9	.909	9	.917	10	.014

V. Conclusion

본 연구는 텍스트마이닝을 통해 여성스포츠웨어 상품명의 주요 키워드를 추출하여 키워드에 대한 정보와 단어 간 관계를 분석해 스포츠웨어를 판매하는 패션 브랜드에 상품명의 구성 방안에 대한 기초자료를 제공하는 것을 목적으로 한다. 상품 노출에 있어 다빈도 키워드의 사용은 더 좋은 검색 결과로 이어질 가능성이 높으므로 상품의 노출 빈도를 더욱 증가시킬 수 있다. 분석을 위해서 네이버 쇼핑 플랫폼의 검색창에 ‘여성 스포츠웨어’를 검색해 ‘리뷰많은순’으로 상품명을 나열하여, 10,000개 이상의 상품명을 수집해 키워드 데이터 정제 과정을 거쳐 주요 키워드를 추출하였으며, 의미연결 네트워크 분석을 실시하였다.

그 결과 여성, 요가, 스포츠브라, 골프, 티셔츠, 레깅스, 필라테스, 조거팬츠, 스포츠, 여자, 이너웨어 등이 주요 상위 키워드로 나타났으며, 등장한 스포츠 종목에는 요가, 골프, 필라테스, 헬스, 트레이닝, 피트니스, 수영, 플라잉요가, 폴댄

스, 에어로빅, 런닝, 사이클, 등산, 줌바댄스, 스피닝, 댄스, 테니스 등이 있었다. 이러한 결과를 바탕으로, 실제 구매가 진행되어 리뷰가 작성된 상품명에는 주로 아이템을 대표하는 키워드는 어떤 것이며, 현재 여성이 많이 구매하는 스포츠 종목 의류는 어떤 것이며, 최근 유행하는 스포츠는 어떤 것인지 파악할 수 있었다. 이처럼 많이 사용되는 단어의 특징을 구체적으로 알아낼 수 있다는 점에서 상품명 연구는 의의가 있다고 생각한다. 또한, 빅데이터를 활용하여 제품의 상품명을 분석한 사례를 제시하며 제품 정보 분석 연구에 참고 자료를 제시하였다는 점에서 본 연구에 의미가 있다고 사료된다.

실증연구를 위하여 비정형 데이터를 수집해 활용하였기 때문에, 빠르고 정확하게 최근 2년 동안 구매가 완료된 제품의 상품명 정보를 수집할 수 있었다. 그러나 본 연구는 연구 대상을 국내 소셜미디어와 언론미디어로 한정하였기 때문에, 국외 이슈는 파악하기 어렵다는 한계가 있다. 그러므로 국외 사용자 기반을 대상으로 자료를 수집하여 국내 소비자와 국외 소비자의 특성을 비교분석한다면 더 의미 있는 연구가 될 수 있을 것이다. 또한, 온라인 자료에는 출처가 불분명하거나 품질 수준이 낮은 정보 또는 통제가 불가능한 중의적 표현 등의 부분이 있으므로, 결과를 일반화하는 것에 주의해야한다.

References

- Gwak, H.-Y., & Lee, K.-H. (2021). Consumer perception of types of fashion live commerce: Using text mining. *Journal of Fashion Business*, 25(3), 90-107.
- Hong, M., & Song, H. (2015). A study linguistic features and sociocultural meaning of food product names. *The Journal of Linguistic Science*, 72, 417-438.
- Jeong, J.-W., & Lee, J.-S. (2016). An analysis on the product name for the domestic cosmetics brands. *Journal of the*

Korea Fashion & Costume Design Association, 18(1), 91-103.

- Kang, Y. S. (2023). Text mining analysis of terms and information on product names used in online sales of women's clothing. *The Research Journal of the Costume Culture*, 31(1), 34-52.
- Kim, H. J., Song, K. E., & Hwang, S. J. (2022). Text analysis of consumer comments on youtube luxury product reviews. *Journal of the Korean Society of Costume*, 72(1), 17-28.
- Kim, J-S. (2005). Analysis of product names and naming strategy in German advertisement. *Korean Society for Didactics of German Language and Literature*, 32, 89-110.
- Kim, J-S. (2018). A study on the perception of fashion streaming service using text mining analysis - Focused on PROJECT ANNE -. *Journal of the Korean Society of Fashion Design*, 18(1), 107-118.
- Shim, J.-Y., & Kim, S. Y. (2008) A study for fashion product name of BEST 100 in cyber shopping mall. *The Research Journal of the Costume Culture*, 16(5), 955 - 962.