Analysis of Research Trends on Body Image Using Text Mining

Su jin Yang, Sunwoo Kim[+], & Seungwoo Seo

Associate Professor, Dept. of Living Culture and Consumer Science, Sungshin Women's University
[+]Research Assistant Professor, Research Institute of Human Ecology, Seoul National University
Master's Course, Department of Applied Artificial Intelligence, Korea University Graduate School of
Computer & Information Technology

Body image is a multifaceted and intricate construct that encompasses individuals' perceptions, evaluations, and emotional responses to their physical appearance and profoundly shapes the development of their self-concept. Body image is influenced by various factors, including sociocultural pressure, personal experience, and individual vulnerability. Extensive research has identified both the positive and negative consequences associated with body image. Notably, body image has been extensively investigated in the clothing and textiles literature due to its substantial influence on consumers' aesthetic preferences and consumption behaviors in the fashion industry. Although a considerable number of studies have examined different facets of body image over the past several decades, few studies have systematically explored the specific aspects and perspectives through a comprehensive literature review. This study aims to fill this gap by thoroughly examining previous research published over the last two decades. Specifically, we elucidate the aspects of body image that have been investigated, the findings, and the contributions to consumers' understanding in the fashion industry.

In the 4th industrial revolution era, triggered by artificial intelligence (AI) and big data, it has been more efficient to collect and analyze unstructured data, including both text and images. In particular, text mining is applicable to multidisciplinary studies based on natural language processing (NLP), which can develop structured data by digitizing and vectorizing unstructured text. Most prominently, NLP has the advantage of excluding respondents' and researchers' prejudices since web-crawled data are not gathered using a controlled experimental or survey methodology. Instead, comments are voluntary and discovered. In academic research, it is essential to analyze the tendency in previous research trends. However, despite the easy access to academic papers around the world through broad Internet search engines such as Google Scholar, the volume of searched literature is too enormous to be reviewed and summarized to capture the overall academic research trends using classic search methods. Thus, there is a significant academic need to examine multidisciplinary research trends using big data analytics. The current research adopts a topic modeling methodology utilizing Python. LDAvis was also applied based on previous literature on topic modeling to determine the number of K (number of topics) and to enhance the power of interpretation.

Topic modeling generally follows three steps: 1) crawling and collecting natural language data, 2) converting the data into computational data through pre-processing and embedding, and 3) classifying

the most appropriate topics using an algorithm. First, to gather relevant information from prior studies, we employed a web-crawling approach that specifically targeted scholarly journals indexed in SCOPUS over the past two decades. SCOPUS is a global academic database platform created by Elsvier Publishing in the Netherlands. It is more suitable for identifying global academic trends outside of the U.S. biased academic evaluation because it selects index journals from a more evenly distributed global region than in the Science Citation Index (SCI). The search terms included in this study for web crawling were "body image" and "body appreciation" in the title of the paper. We crawled data that included the title, author, publication year, journal name, volume, abstract, keywords, and URLs, which can be directly accessed in the crawled data. The search period was from Jan 1, 2004, to March 31, 2023. The final dataset included 5,993 studies with an abstract written in English.

After completing the preprocessing step, including excluding punctuation characters or common words and combining synonyms into a single semantic word, the remaining corpus was subjected to embedding using Term Frequency-Inverse Document Frequency (TF-IDF) analysis. This study adopted Latent Dirichlet Allocation (LDA) as an algorithm for topic modeling, which is used the most in the latest topic modeling research. Hofmann (1999) suggested the concept of topic modeling using probabilistic latent semantic analysis, and Bei et al. (2003) improved this concept by adding unsupervised learning clustering algorithms based on LDA. This approach solves overfitting, which is a consistent problem with AI analysis methods. Furthermore, the interpretability of the results has dramatically improved with advancements in LDAvis, (i.e., a visualization tool developed by Sivert & Shirely, 2014).

TF-IDF methodology standardizes pre-processed words into a "bag of words" (BoW). In this process, words with low TF-IDF rankings are excluded from the analysis to improve the quality of the results. The exclusion thresholds suggested by previous studies vary, such as average, median, and 0.1 or more. This study tried to apply the three suggested thresholds to select the appropriate number of topics. The highest coherence core turned out to be 0.49 by applying the median threshold, producing 11 topics. When applying the median threshold to exclude BoW, the topic modeling was 21,796. The final topic modeling results for 11 topics are summarized in [Tab 1].

[Tab 1] 11 topics of the body image literature

Topic	Name	Example salient keywords
1	Adolescent & education	Adolescent, Student, Physical, Activity, Health, Female, Child, Male, Attitude, Boy, Perception, Diet, Nutritional, Fat, etc.
2	Emotional concerns	Concern, Behavior, Eating, Negative, Low, Shame, Parent, Evaluation, Affect, Emotional, Family, Stress, Disordered, Role, etc.
3	Issues and perspectives	Experience, People, Method, Issue, Bias, Approach, Part, Paper, Perspective, Stimulus, Framework, etc.
4	Double standard	Woman, Gender, Age, Adult, Young, American, Identity, Impact, Black, Culture, Standard, Awareness, Disability, White, Ethnic, Face, etc.

5	Postoperative body image	Patient, Cancer, Breast, Surgery, Life, Quality, Distress, Psychological, Cosmetic, Disease, Mastectomy, Reconstruction, Diagnosis, Postoperative, etc.
6	Women & body concerns	Sexual, Self-esteem, Satisfaction, Symptom, Depression, Relationship, Anxiety, Function, Support, Well-being, Depressive, Mother, Partner, Pregnancy, etc.
7	Scale development	Scale, Factor, Validity, Measure, Correlation, Test, Questionnaire, Sample, Reliability, Population, Evaluate, Item, Instrument, Assessment, etc.
8	Eating disorder	Eat, Disorder, Group, Disturbance, Individual, Risk, Sport, Mental, Difference, Athlete, Healthy, Nervosa, Anorexia, Muscle, Active, Personality, Pathology, etc.
9	Ideal thin appearance	Social Media, Appearance, Model, Comparison, Exposure, Ideal, Thin, Internationalization, Sociocultural, Peer, Instagram, Self-objectification, etc.
10	Positive body image	Appreciation, Intervention, Positive, Exercise, Program, Review, Self-compassion, Improve, Strategy, Acceptance, Cope, Therapy, Motivation, etc.
11	Obesity	Weight, Dissatisfaction, BMI, Size, Overweight, Obesity, Shape, Sex, Ideal, Figure, Silhouette, Desire, Gay, Discrepancy, Gain, Loss, etc.

The 1st and 6th topics were about significant and representative research subjects in the field of body image for adolescents and women. While the 1st topic related to adolescence dealt with education and health-related issues, the 6th topic related to women, including pregnancy and family relationships. While the 2nd topic mainly focused on the relationship between emotional distress and eating disordered behavior, the 8th topic tended to view eating disorder as a medical disease. The 3rd and 7th topics are either comprehensive framework research or methodology development research. The 4th topic related to prevalent double standards according to cultural, ethnic, and sexual groups. The 5th topic discussed psychological symptoms that patients may experience after surgery that can distort their body. The 9th topic is about an ideal thin appearance, and the 10th topic is about having a positive body image, which is likely to be a newer issue that has emerged due to social media. Finally, the 11th topic is about activities to address obesity and weight gain/loss.

Our findings contribute significantly to both academia and industry. Specifically, the results enhance our understanding of consumers in the fashion industry and provide valuable insights that can guide future research directions.