



Original Article

Analysis on the Type of and Interest in Home Training Video Contents Using the YouTube Platform

Hyo-Jun Yun¹, Jae-Hyeon Park¹ and Minsoo Jeon^{2*}

¹Korea National Sport University

²Dankook University

Article Info

Received 2022.11.17.

Revised 2022.12.24.

Accepted 2022.12.26.

Correspondence*

Minsoo Jeon

minsoo1144@dankook.ac.kr

Key Words

Home training,
Video contents,
Topic modeling,
Type analysis,
Interest analysis,
YouTube

이 논문은 2021년 대한민국 교육부와 한
국연구재단의 지원을 받아 수행된 연구임
(NRF-2021S1A5B5A17049950).

PURPOSE The purpose of this study is to analyze the type of and interest in home training video contents using the YouTube platform. **METHODS** Web crawling was performed using Python and a total of 3,937 sets of video information (title, content, number of views, upload date) were obtained, 3,155 of which were finally selected for the study material. Overlapping and unrelated content were excluded. The data of text underwent 3 stages of preprocessing, the TF and TF-IDF of the keywords were calculated to identify the main keywords, and the LDA algorithm was applied in the topic modeling to successfully identify the types. In order to understand the level of interest by type, the number of views was subdivided into the percentage of the assigned type. **RESULTS** First, the types of home training videos were classified into bare whole body exercise for aerobic and muscular power strengthening, Pilates exercise for core and upper body strengthening, upper body exercise using tools, lower body line exercise, posture correction and upper body stretching exercise for pain relief, hip-up exercise, dance and tabata exercise for diet, diet and lower body correction stretching exercise for diet, and bare body exercise for core and lower body strengthening. Second, it was found that the proportion and interest were high in the contents of bare whole body exercise for aerobic and muscular power strengthening, dance and tabata exercise for diet, diet and lower body correction stretching exercise for diet. **CONCLUSIONS** The findings of this study may provide baseline data about the development of the active online home training videos in the market.

서론

중국 우한에서 최초 시작된 코로나바이러스감염증-19(이하 코로나 19)가 전 세계에 급격하게 확산됨에 따라 2020년 3월 11일 세계보건기구(WHO)는 코로나 감염병에 따른 세계적 대유행(Pandemic)을 선언하였다. 코로나19의 확산으로 사회적 거리두기, 재택근무, 홈스쿨링 등 일상의 교류가 단절되고 코로나19가 장기화됨에 따라 감염병에 대한 불안과 공포와 함께 사람들과의 물리적 접촉제한, 경제적 어려움 등 일상에 직접적인 영향을 받고있는 실정이다 (Ministry of Health and Welfare, 2020). 국내 경기도민 1000명을 대상으로 조사한 결과, 코로나19로 인한 스트레스와 불안감이

중증질환이나 자연재해보다 더 높은 것으로 나타났으며, 소셜 빅데이터 분석결과 코로나 초기에서는 확산, 안전 등의 단어가 주요 단어였으나, 코로나19가 장기화됨에 따라 힘들다, 위기와 같은 부정적 심리작용을 나타내는 단어가 주요 단어로 나타났다(Kim et al., 2020). 이는 국민이 인지하고 있을 코로나19에 대한 불안감이 고조된 유추할 수 있다.

스포츠 영역에서 코로나19의 영향은 도쿄올림픽 개최가 연기되는 사상 초유의 사건이 발생하였으며, 세계 프로스포츠 리그가 중단 및 무관중 개막을 하는 사태를 맞이하게 되었다(Yun et al., 2021). 또한, 생활체육 및 운동 양식의 변화도 적지 않다. 국민생활체육조사 결과에 따르면, 최근 5년간 생활체육 참여율은 꾸준히 증가(2015년 56%, 2016년 59.5%, 2017년 59.2%, 2018년 62.2%, 2019년 66.6%)하는 추세였으나, 2020년 60.1%로 감소하는 것으로 나타났다(Ministry of Culture, Sports and Tourism, 2020). 특히, 보디빌딩, 요가, 필라테스, 수영 등 실내종목에서 참여율이 감소하였으

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

며, 이는 감염병 확산 예방을 위한 국가 정책으로 사회적 거리두기가 강화됨에 따라 생활체육대회 개최가 제한됨과 더불어 헬스장, 탁구장, 스크린골프장 등 실내체육시설 이용제한으로 풀이된다.

헬스장과 같은 실내체육시설 이용제한 등 생활체육환경 변화는 생활체육 참여자의 홈트레이닝에 대한 관심과 참여를 불러일으켰다. 홈트레이닝은 집을 뜻하는 홈(home)과 운동을 의미하는 트레이닝(training)의 합성어로 시공간의 제약으로부터 자유로워 센터를 방문하지 않고 스스로 주도적 운동이 가능하여 비교적 접근이 용이하다는 장점이 있다(Kwon & Nam, 2022; Seo & Kim, 2020). 통신사에서 제공하는 IPTV AI 서비스의 2020년 발화량 분석결과, 홈트레이닝 관련 발화량이 전년 대비 87% 증가하였으며, 요가(113.7%), 스트레칭(102.6%), 필라테스(96.5%)와 관련된 발화가 전년도 대비 크게 증가하였다(Ko, 2020). 또한, 아령(211.8%), 요가밴드(58%), 홀라후프(16.1%)와 같은 홈트레이닝 관련 용품의 판매량이 증가(Cha, 2020)하였다는 점에서 홈트레이닝에 대한 관심이 증가한 것을 유추할 수 있다.

홈트레이닝 수요가 증가함에 따라, 온라인 홈트레이닝 동영상 콘텐츠가 증가하고 있다. 1인 미디어의 홈트레이닝 동영상 콘텐츠뿐만 아니라 시도체육회에서 온라인 홈트레이닝 동영상 콘텐츠를 제작하여 비대면 운동 교실을 운영하고 있다(Lee, 2020a). 문화체육관광부와 국민체육진흥공단에서는 코로나로 인해 침체된 스포츠산업을 지원하기 위해 비대면 스포츠 시장을 육성하겠다고 밝혔다. 이는 코로나로 침체한 민간 체육시설업체가 조속히 극복할 수 있도록 새로운 사업 방식을 제공하는 한편, 점차 비대면 선호가 강해지는 미래 사회에 선제 대응할 수 있는 환경을 구축하는 것을 목표로 비대면 스포츠 융합인력 양성, 민간 체육시설업자의 비대면 사업전환 지원, 비대면 스포츠콘텐츠 제작 및 유통망(플랫폼) 구축 등 3개 분야에 예산 총 55억을 지원하겠다는 입장이다(Kim, 2020). 이러한 정부의 방침 아래 온라인 홈트레이닝 시장의 규모가 더욱 커질 전망이다.

콘텐츠 공유 플랫폼 중 가장 대표적인 것이 유튜브(youtube)이다. 미국 Conscore의 자료에 의하면 2012년도에 5월 한 달 동안 미국 온라인 동영상 서비스를 이용한 1억 8천만 명 중 1억 5천만 명이 유튜브를 이용한 것으로 나타났으며 1인 평균 이용 시간은 약 8시간에 달하는 것으로 나타났다. 국내 동영상 플랫폼 중에서 유튜브의 점유율은 85.6%로 아프리카TV와 네이버TV의 점유율이 2~3%인 점인 것 고려하면, 국내에서도 가장 많이 활용되는 대표적인 동영상 공유 플랫폼인 셈이다(Shin, 2020). 특히, 유튜브 홈트레이닝 채널은 방대한 양의 영상이 있어 선택이 폭이 넓으며, 자신이 원하는 영상을 찾아 트레이닝할 수 있고, 스마트폰, 스마트 TV, 컴퓨터 등 다양한 디바이스로 접근이 가능한 장점이 있다(Jeon & Pan, 2020).

온라인 홈트레이닝 콘텐츠 관련 사전연구를 살펴보면, 홈트레이닝 참여자에 관한 연구(Choi, 2020; Choi et al., 2017; Lee, 2020b), 홈트레이닝 산업에 대한 연구(Cho, 2020; Jeon & Pan, 2020; Kim et al., 2019; Sheen, 2017; Shin, 2019)가 주류를 이루고 있으며, 온라인 홈트레이닝 콘텐츠 현황을 분석한 연구로는 Lee & Kwak(2020)은 유튜브 홈트레이닝 채널을 중심으로 채널의 특징과 구독자, 조회 수, 댓글 등을 분석하였다. 그러나 이들 연구에서는 구독자 수를 근거로 상위 8개의 채널만을 한정하여 분석을 실시하였다. 이는 온라인 홈트레이닝 콘텐츠의 전반적인 현황과 유형을 분석하지 못한 점에서 일반화하기 어려우며, 분석방법에서 내용분석을 적용하고 있어 분석에 대한 주관성을 완전히 배제하기 힘들다는 제

한이 있다. 반면, 토픽모델링은 텍스트마이닝 기법 중 하나로 텍스트내의 주제를 객관적으로 토출하는데 유용한 알고리즘이다(Yun et al., 2021).

따라서 이 연구는 토픽모델링을 적용하여 온라인 홈트레이닝 콘텐츠의 유형을 파악하고 나아가 유형별 관심도를 분석하는 것이 목적이다. 이는 온라인 홈트레이닝 수요자의 니즈를 파악하고 공급자에게는 콘텐츠 제작에 기초자료로 활용되고자 하는 목적이다.

연구방법

연구절차

이 연구의 목적을 달성하기 위해 수행된 연구절차는 다음 <Figure 1>과 같다.

연구자료

연구자료 선정을 위해 우선 유튜브 검색엔진에 홈트레이닝과 홈트의 키워드로 검색하였을 때 노출되는 동영상 콘텐츠의 자료를 수집하였다. 또한, Lee & Kwak(2020)에 연구에서 보고된 홈트레이닝 관련 유튜브 채널 상위 8개 채널의 동영상 콘텐츠를 수집하였다. 수집대상은 동영상 콘텐츠의 제목, 내용, 조회 수, 업로드 날짜를 대상으로 수집하였다. 자료수집을 위해 python 3을 활용하여 웹크롤링 자동화 프로그램을 개발하였으며 이때 selenium의 webdriver를 사용하였다. 최초 연구자료로 수집된 동영상 콘텐츠 수는 3937개이었으나, 중복되는 콘텐츠와 연구자가 콘텐츠의 제목과 내용을 읽어보고 홈트레이닝과 관련이 없는 콘텐츠를 제거하였다. 따라서 최종적으로 동영상 콘텐츠의 수는 3,155개이다.

텍스트전처리

동영상 콘텐츠 유형을 분석하기 위해 사용된 데이터는 최종연구자료 선정된 동영상 콘텐츠의 제목 데이터이다. 수집단계에서 콘텐츠의 제목과 내용을 모두 수집하였지만, 제목만을 한정된 이유는 내용은 채널에 대한 설명이 서술되어 있어 내용을 토대로 유형을 분석할

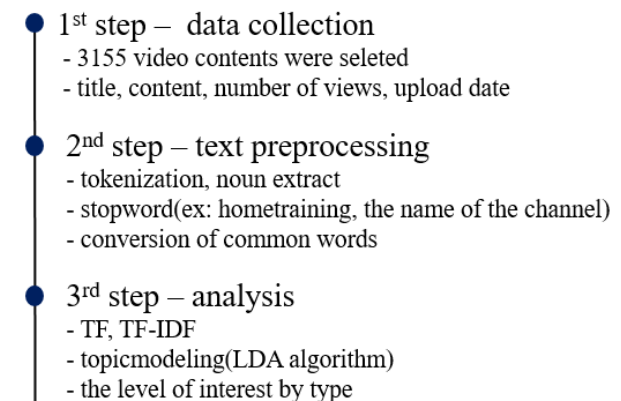


Fig. 1. Research procedure

시 채널이 영향을 미칠 수 있을 방지하기 위함이었다. 제목 데이터는 텍스트 자료로 분석에 소요되는 시간과 정확성을 높이기 위해서 텍스트전처리 단계가 필수적이다.

따라서 이 연구에서는 3단계에 걸쳐 텍스트전처리를 실시하였다. 첫째, 선정된 자료를 단어별로 토큰화(tokenization)를 실시하였으며 이후 명사단어들만을 추출하였다. 토큰화는 기본적으로 Okt 형태소분석기를 사용하였으며, 타바타, 런지, 집볼, 팔뚝살 등 단어사전에 없는 단어들은 사전을 수정하여 적용하였다. 둘째, 불필요한 단어를 제거하였다. 이 연구에서는 검색 키워드로 사용한 홈트레이닝과 채널의 이름을 불용어 처리하였으며 편, 주 등과 같이 의미 없는 단어도 불용어 처리하였다. 셋째, 초보자와 초보, 솔더와 어깨, 엉덩이와 힙과 같이 같은 의미이지만 다르게 표현된 단어들을 공통단어로 변환시켰다.

자료처리

텍스트 자료의 주요 키워드를 알아보기 위해 TF(Term Frequency)와 TF-IDF(Term Frequency - Inverse Document Frequency)를 산출하였다. TF는 단어의 빈도를 의미하며, TF-IDF는 단어의 빈도와 역문서의 빈도를 곱한 값으로 단어 간의 상대적 중요도를 평가하기 위해 고안된 방법이다(Yun et al., 2019).

동영상 콘텐츠 유형을 분류하기 위해 토픽모델링을 적용하였다. 토픽모델링의 알고리즘은 1990년 Deerwester 등이 제안한 LSA(Latent Semantic Analysis) 알고리즘이 시초이며 이후 Hofmann(1999)는 LSA 알고리즘의 확장된 알고리즘으로 pLSA(Probabilistic Latent Semantic Analysis) 알고리즘을 제안하였다. 그러나 pLSA 알고리즘은 새로운 문서에 대한 확률을 계산할 수 없으며 많은 매개변수로 인해 수식이 복잡해질 수 있으며 훈련데이터에 지나치게 맞춰지는 과적합(overfit) 현상이 나타날 수 있는 단점이 있다(Blei et al., 2003). 이러한 단점을 보완하고자 Blei et al.(2003)은 LDA(Latent Dirichlet Allocation) 알고리즘을 제안하였다. 이 알고리즘은 자료 차원 축소의 유용성과 의미적으로 일관성 있는 주제를 생성할 수 있다. 따라서 이 연구에서는 토픽모델링 중 LDA 알고리즘을 적용하였다.

토픽모델링은 사전에 설정한 토픽 수에 따라 산출되는 결과가 상이하기 때문에 전체 문서 집합으로부터 몇 개의 토픽이 존재하는지를 결정해야 한다. 토픽 수를 결정하기 위해 지수로 Perplexity와 Coherence가 대표적이지만 Perplexity index는 내재적으로 학습 성능 정도를 의미하고 있을 뿐 그 결과를 해석하기 난해하다는 단점을 가지고 있다(Chang et al., 2009). 반면, Coherence는 Newman et al.(2010)에 의해 처음 제안된 방법으로 토픽모델링 결과로 산출된 각각의 주제들이 포함된 상위 단어들이 얼마나 높은 유사도를 가지느냐에 따라 평가되는 방법이다. 즉, coherence score가 높을수록 토픽모델링으로 산출되는 각각의 주제가 의미론적으로 유사한 단어들로 구성되었다고 해석할 수 있으며 이 연구에서는 coherence score를 기반으로 토픽 수를 결정하였다.

토픽모델링은 알고리즘을 통해 기계학습으로 이뤄지지만, 각 토픽의 의미를 해석하고, 토픽 명을 설정하는 과정이 더 중요하게 여겨진다(Chandra et al., 2016; Quinn et al., 2010). 따라서 토픽 명을 부여하기 위한 방법의 하나로 관련 학문 분야 전문가들의 의견을 수렴하여 토픽 명을 부여하는 과정이 뒤따라야 한다(Chandra et al., 2016;

Guo et al., 2016). 이 연구에서 도출된 토픽모델링 결과와 원문을 토대로 본 연구진 3인의 회의를 거쳐 최종적으로 토픽을 명명하였다.

동영상 콘텐츠 유형의 관심도를 알아보기 위해 조회 수 데이터를 활용하였다. LDA 알고리즘은 한 문서에 여러 개의 주제가 존재함을 가정하는 모형으로 한 문서에 할당된 토픽의 퍼센트가 산출된다. 즉 이 연구에서 하나의 동영상 콘텐츠가 하나의 유형으로 분류되는 것이 아닌 유형별 할당된 퍼센트가 산출되는 것이다. 따라서 유형별 관심도를 파악하기 위해서 콘텐츠에 조회 수를 할당된 퍼센트로 분배하여 산출하였다. 가령, 어떤 콘텐츠의 조회 수가 100회이고, 3가지 유형으로 분류된 콘텐츠의 할당 퍼센트는 1유형 50%, 2유형 30%, 3유형 20%라고 가정하자. 이 경우 1유형의 조회 수는 50회, 2유형의 조회 수는 30회, 3유형의 조회 수는 20회로 분배하여 유형별 관심도를 산출하였다. 이 연구에서 모든 자료처리는 python 3을 활용하였다.

연구결과

주요 키워드 분석

주요 키워드를 알아보기 위해 TF와 TF-IDF를 산출하였으며 다음 <Table 1>은 상위 40개의 단어를 나타낸 표이다. TF는 운동, 전신, 다이어트, 살, 엉덩이가 순으로 높게 나타났으며, TF-IDF는 운동, 전신, 다이어트, 뱃살, 살 순으로 높게 나타났다. 다음 <Figure 1>은 TF를 기반으로 워드클라우드 시각화 결과이다.

토픽모델링 결과

이 연구에서 토픽 수를 설정하기 위해 토픽 수를 5부터 15까지 설정하여 Coherence Score를 산출하였으며, 그 결과 토픽 수를 9개로 설정하였을 때 Coherence Score가 0.579로 가장 높게 나타났다. 따라서 토픽 수를 9개로 설정하여 토픽모델링을 실시하였다.

토픽점유율을 살펴보면, 토픽1이 33.0%로 가장 높게 나타났으며, 토픽 7이 11.3%로 두 번째로 높게 나타났다. 토픽별 주요 키워드를 살펴보면, 토픽1은 운동, 전신, 유산소, 근력, 체지방, 초보, 뱃살, 복부, 빼기, 맨몸으로 구성되어있으며, 토픽2는 필라테스, 코어, 상체, 강화, 기구, 걷기, 팔뚝, 플랭크, 끝, 수업으로 구성되어있었다. 토픽3은 어깨, 등, 밴드, 폼롤러, 팁, 팔, 제기, 팔뚝살, 등살, 동작으로 구성되어있으며, 토픽4는 살, 허벅지, 지방, 다리, 체력, 안쪽, 땀, 파워, 몸



Fig. 2. Wordcloud result

Table 1. TF and TF-IDF result

Rank	TF		TF-IDF		Rank	TF		TF-IDF	
	Keyword	Value	Keyword	Value		Keyword	Value	Keyword	Value
1	운동	1868	운동	153.5	21	맨몸	123	빼기	41.1
2	전신	443	전신	107.6	22	초보	122	순환	39.7
3	다이어트	375	다이어트	85.6	23	복부	122	복부	39.7
4	살	274	뱃살	76.1	24	어깨	121	칼로리	39.3
5	엉덩이	251	살	75.8	25	빼기	114	필라테스	39.3
6	유산소	237	유산소	74.6	26	칼로리	113	허리	38.0
7	루틴	220	복근	67.2	27	코어	122	아침	38.0
8	뱃살	215	엉덩이	66.9	28	순환	110	라인	37.4
9	근력	213	체지방	62.5	29	집	100	어깨	37.1
10	스트레칭	202	몸	59.7	30	비만	97	근육	33.7
11	하체	202	근력	57.5	31	스쿼트	94	코어	33.6
12	체지방	199	요가	56.7	32	상체	94	하루	32.5
13	복근	189	하체	56.2	33	지방	93	상체	32.3
14	몸	170	루틴	54.4	34	힙업	91	지방	31.8
15	필라테스	167	스트레칭	54.3	35	근육	90	효과	31.6
16	허벅지	154	허벅지	52.7	36	식단	90	비만	31.3
17	라인	136	타바타	51.7	37	등	88	힙업	30.5
18	타바타	134	맨몸	46.2	38	효과	87	스쿼트	30.2
19	요가	134	초보	45.5	39	밴드	83	식단	28.8
20	허리	127	집	42.9	40	하루	82	옆구리	28.7

Table 2. Topic modeling result

Topic	Percent	Keyword (probability distribution)				
Topic 1	33.0%	운동(0.321)	전신(0.073)	유산소(0.049)	근력(0.042)	체지방(0.038)
		초보(0.037)	뱃살(0.036)	복부(0.029)	빼기(0.025)	맨몸(0.025)
Topic 2	5.2%	필라테스(0.104)	코어(0.104)	상체(0.077)	강화(0.052)	기구(0.029)
		걷기(0.018)	팔뚝(0.032)	플랭크(0.015)	끝(0.014)	수업(0.014)
Topic 3	6.4%	어깨(0.160)	등(0.074)	밴드(0.053)	폼롤러(0.036)	팁(0.035)
		팔(0.033)	제거(0.033)	팔뚝살(0.032)	등살(0.028)	동작(0.027)
Topic 4	9.7%	살(0.149)	허벅지(0.072)	지방(0.053)	다리(0.034)	체력(0.030)
		안쪽(0.027)	땀(0.021)	파워(0.021)	몸매(0.021)	셀룰라이트(0.020)
Topic 5	7.6%	몸(0.150)	스트레칭(0.131)	아침(0.050)	통증(0.038)	자세(0.036)
		목(0.030)	헬스(0.022)	자극(0.021)	마사지(0.018)	완화(0.017)
Topic 6	9.2%	엉덩이(0.014)	루틴(0.108)	힙업(0.060)	탄력(0.054)	라인(0.052)
		덤벨(0.035)	여름(0.020)	충간소음(0.019)	근살(0.019)	애플(0.018)
Topic 7	11.3%	다이어트(0.176)	타바타(0.069)	근육(0.055)	가슴(0.041)	효과(0.039)
		댄스(0.026)	호흡(0.020)	시간(0.014)	동시(0.014)	이용(0.013)
Topic 8	10.2%	하체(0.100)	순환(0.072)	여자(0.051)	식단(0.040)	옆구리(0.037)
		골반(0.036)	교정(0.034)	하루(0.033)	집중(0.030)	소모(0.030)
Topic 9	7.5%	복근(0.138)	허리(0.081)	스쿼트(0.067)	버피(0.035)	방법(0.028)
		기초(0.023)	테스트(0.023)	준비(0.020)	런지(0.019)	마운틴(0.015)

Table 3. Result of interest analysis by topic

Topic	M	SD	Md	Q	Skewness
Topic 1	115080	574399	3699	9671	10.1
Topic 2	13954	72880	479	1211	12.2
Topic 3	18616	109631	571	1641	16.8
Topic 4	37237	220240	966	2463	17.0
Topic 5	18922	85822	659	1757	8.8
Topic 6	26158	118093	957	2421	9.6
Topic 7	41619	242602	1063	3202	13.0
Topic 8	30954	151481	1067	2747	11.2
Topic 9	21473	105822	677	1892	13.0
Total	324014	1404121	11556	30514	8.1

매, 셀룰라이트 구성되어있었다. 토픽5는 몸, 스트레칭, 아침, 통증, 자세, 목, 헬스, 자극, 마사지, 완화로 구성되어있으며, 토픽6은 엉덩이, 루틴, 힘업, 탄력, 라인, 덤벨, 여름, 층간소음, 근살, 애플로 구성되어있었다. 토픽7은 다이어트, 타바타, 근육, 가슴, 효과, 댄스, 호흡, 시간, 동시, 이용으로 구성되어있으며, 토픽8은 하체, 순환, 여자, 식단, 옆구리, 골반, 교정, 하루, 집중, 소모로 구성되어있었다. 토픽9는 복근, 허리, 스쿼트, 버피, 방법, 기초, 테스트, 준비, 런지, 마운틴으로 구성되어있었다.

토픽1에 구성된 키워드와 토픽1에 높은 퍼센트로 할당된 동영상 콘텐츠의 제목과 내용을 토대로 토픽 1을 유산소 및 근력 강화를 위한 맨몸 전신운동으로 명명하였으며 토픽2는 코어 및 상체강화를 위한 필라테스운동으로 명명하였다. 토픽3은 밴드, 폼롤러 등 같은 도구를 활용한 상체운동으로 명명하였으며, 토픽4는 허벅지, 종아리 라인과 관련된 콘텐츠로 구성되어있어 하체라인 운동으로 명명하였다. 토픽5는 통증 완화를 위한 자세교정 및 상체스트레칭운동으로 명명하였으며, 토픽6은 힙업운동으로 명명하였다. 토픽7은 다이어트를 위한 댄스 및 타바타운동으로 명명하였으며, 토픽 8은 다이어트를 위한 식단 및 하체교정 스트레칭운동으로 명명하였다. 토픽 9는 코어 및 하체강화를 위한 맨몸운동으로 명명하였다.

동영상 콘텐츠 유형별 관심도 분석

콘텐츠의 조회 수를 기반으로 동영상 콘텐츠 9가지 유형별 관심도를 분석하였다. 그 결과 다음 <Table 3>과 같다. 전체적으로 홈트레이닝 관련된 동영상 콘텐츠의 조회 수는 평균적으로 324,014회로 나타났다으며 중앙값은 11,556회로 산출되었다. 유형별 관심도를 살펴보면 평균을 기준으로 토픽1, 토픽7, 토픽8, 토픽4, 토픽6, 토픽5, 토픽9, 토픽3, 토픽2 순서로 관심도가 높은 것으로 나타났다. 중앙값을 기준으로는 토픽1, 토픽8, 토픽7, 토픽4, 토픽5, 토픽9, 토픽5, 토픽3, 토픽2 순서로 관심도가 높은 것으로 나타났다. 왜도값을 살펴보면, 모든 유형에서 ± 2 범위에서 벗어나 정상분포는 만족하지 않은 것으로 나타났다.

논의 및 결론

이 연구는 유튜브 플랫폼의 등록된 홈트레이닝 동영상 콘텐츠 자료를 활용하여 홈트레이닝 동영상 콘텐츠의 유형과 관심도를 파악하고자 하였다. 이 연구에서 도출된 결과를 바탕으로 논의는 다음과 같다.

주요 키워드를 알아보기 TF와 TF-IDF를 산출한 결과 TF의 순위와 TF-IDF의 순위는 차이가 있는 것으로 파악되었으며, 이는 TF-IDF가 단어 간의 상대적 중요도를 평가하기 위해 고안된 방법이기 때문이다(Yun et al., 2019). 운동과 전신 키워드의 TF값의 차이는 1425로 나타나지만, TF-IDF값의 차이는 45.9로 비교적 적은 편이다. 이는 운동이라는 키워드는 빈도가 높게 나왔지만 많은 콘텐츠에서 사용하고 있는 단어이기 때문에 TF-IDF 값이 비교적 작게 나타난 것으로 해석할 수 있다. TF-IDF 값을 기준으로 상위권에 도출된 단어들을 보면 다이어트, 뱃살, 살, 체지방 등 비만과 관련된 단어들 이 많이 나타났으며, 이는 홈트레이닝 동영상 콘텐츠의 주요 목적은 운동을 통한 체중 감량으로 판단할 수 있다. Kim(2022) 연구에 따르면 홈트레이닝 관련 언론 미디어 보도량은 코로나 이전 영양섭취와 다이어트 대한 보도량이 많은 것으로 보고되고 있으며, 이는 홈트레이닝이 다이어트를 위한 하나의 수단으로서 많이 이뤄졌음을 시사함을 의미한다. 반면, 코로나 이후 홈트레이닝 안전과 산업에 대한 보도량이 증가하였지만, 사회적 거리두기로 인한 신체활동량 감소는 체중증가라는 결과를 만들었으며 이는 홈트레이닝 동영상 콘텐츠의 꾸준한 관심대상이었을 것으로 판단된다.

토픽모델링의 결과를 토대로 홈트레이닝 동영상 콘텐츠 유형을 분류하면, 유산소 및 근력 강화를 위한 맨몸 전신 운동, 코어 및 상체 강화를 위한 필라테스운동, 도구를 활용한 상체운동, 하체 라인 운동, 통증 완화를 위한 자세교정 및 상체 스트레칭운동, 힙업운동, 다이어트를 위한 댄스 및 타바타운동, 다이어트를 위한 식단 및 하체교정 스트레칭운동, 코어 및 하체강화를 위한 맨몸운동으로 분류할 수 있었다. Kim(2021)는 유튜브 홈트레이닝 콘텐츠의 조회 수 특성에 관한 연구에서 운동유형을 체중감량, 대근육발달, 유연성 및 자세교정의 3가지로 콘텐츠를 유형화하였으며, 후속연구에는 세부화된 기준이 마련되길 언급한바 이 연구에서는 객관적인 방법으로 9가지로 유형으로 분류한 점에서 의의가 있을 것이다.

유형별 점유율은 코어 및 하체강화를 위한 맨몸운동과 다이어트를 위한 댄스 및 타바타 운동이 가장 높게 나타났다. 두 유형은 공통점은 맨몸 전신운동으로 체중 감량의 밀접한 관련이 되어 있으며, 앞서 설명한 바와 같이 홈트레이닝의 주요관심 대상은 체중 감량으로 판단된다. 이는 유튜브 홈트레이닝 콘텐츠의 소비 동기와 빈도의 관계를 분석한 연구(Choi, 2020; Oh et al., 2019; Roh & Kim, 2020; Yun & Kim, 2021)들의 결과를 지지해준다. 특히, 타바타 운동(tabata training)은 고강도 인터벌 트레이닝 중 하나로 짧은 시간의 고강도 운동(최대산소섭취량 170%)과 짧은 휴식을 여러 차례 반복하는 유무산소성 운동방법이다(Tabata, 2019). 타바타 운동은 특별한 기구 없이 맨몸으로 수행(Ahn et al., 2022)할 수 있어 홈트레이닝에 적합한 운동으로 부각되고 있다.

동영상 콘텐츠 유형을 살펴보면, 필라테스운동, 하체 라인 운동, 힙업운동, 하체교정 스트레칭 등 여성 위주의 운동으로 구성되어 있는 것을 확인할 수 있다. 이는 이 연구에서 연구자료를 선정할 때 Lee & Kwak(2020)의 연구를 기반으로 홈트레이닝 관련 유튜브 채널 상위 8개 채널을 대상으로 동영상 콘텐츠를 수집하였고, 상위 8

개 채널의 중 7개가 여성 유튜버가 운영하는 관리하는 채널이었기 때문일 가능성이 있다. 그러나 Kim(2021) 연구에 따르면 유튜버의 홈트레이닝 콘텐츠의 74%가 여성 크리에이터인 것으로 보고하고 있으며, Lee & Kwak(2020)의 연구에서도 홈트레이닝 채널은 여성, 다이어트, 미용 위주의 콘텐츠를 제공하고 있어 홈트레이닝 관련 동영상 콘텐츠는 남성보다 여성 중심의 콘텐츠가 많은 것으로 판단된다.

동영상 콘텐츠 유형별 관심도 분석결과를 살펴보면, 왜도값이 ± 2 범위에서 벗어난 것으로 확인할 수 있었다. 이는 유튜브 콘텐츠 자료의 기술통계를 보고한 연구(Kim, 2021)에서 보고한 왜도값 13.45와 일치한다. 왜도가 ± 2 범위에서 벗어난 점은 정상분포가 아닌 정적편포에 해당하며, 조회수가 극단적으로 높은 동영상 콘텐츠로 인해 발생한 현상으로 유추할 수 있다. 따라서 유형별 관심도를 파악하기 위해서 평균을 기준으로 하는 것보다 중앙값을 기준으로 하는 것이 타당하다고 판단된다. 중앙값을 기준으로 유형별 관심도를 평가하면 유산소 및 근력 강화를 위한 맨몸 전신 운동, 다이어트를 위한 식단 및 하체교정 스트레칭운동, 다이어트를 위한 댄스 및 타바타운동 순으로 관심도가 높게 나타났다. Kim(2021) 연구에 따르면, 체중감량을 목적으로 한 콘텐츠가 대근육발달, 유연성 및 체형교정 운동을 목적으로 한 콘텐츠보다 조회수가 더 높은 것으로 보고하고 있어 연구의 결과를 지지해준다. 다만, 이 연구는 LDA 알고리즘을 적용하여 유형을 분류하고 조회 수를 유형별로 재분배하였기 때문에 제시한 유형별 평균 및 중앙값 조회 수가 하나의 콘텐츠 조회 수를 대표하는 값으로 해석하기 어렵다. 구체적으로 하나의 콘텐츠가 어느 하나의 유형으로 정의할 수 없고 유산소 및 근력 강화를 위한 맨몸 전신 운동 유형의 콘텐츠의 경우, 평균적으로 115,080회, 중위적으로 3,699회 조회 수가 나타났다고 해석하기 어렵다는 것이다. 유형별 조회 수의 기술통계는 관심도의 서열적인 측면에서 의미가 있을 것이다.

이 연구는 홈트레이닝 동영상 콘텐츠의 유형과 관심도를 파악하는 것이 주요 목적이며, 다음과 같은 결론을 도출할 수 있었다. 첫째, 홈트레이닝 동영상 콘텐츠의 유형은 유산소 및 근력 강화를 위한 맨몸 전신 운동, 코어 및 상체 강화를 위한 필라테스운동, 도구를 활용한 상체운동, 하체 라인 운동, 통증 완화를 위한 자세교정 및 상체 스트레칭운동, 힙업운동, 다이어트를 위한 댄스 및 타바타운동, 다이어트를 위한 식단 및 하체교정 스트레칭운동, 코어 및 하체강화를 위한 맨몸운동으로 분류할 수 있다. 둘째, 유산소 및 근력 강화를 위한 맨몸 전신 운동, 다이어트를 위한 댄스 및 타바타운동 유형, 다이어트를 위한 식단 및 하체교정 스트레칭운동의 콘텐츠 비중과 관심도가 높은 것으로 나타났다. 이는 홈트레이닝의 주요 목적이 체중 감량으로 판단된다. 이 연구는 홈트레이닝 동영상 콘텐츠의 유형과 관심도를 분석하기 위해 유튜브 플랫폼만 활용하였다는 점에서 제한점이 있지만, 국내 동영상 플랫폼 중에서 유튜브의 점유율은 85.6%인 점을 감안하면 의미는 적지 않을 것으로 판단된다. 또한 콘텐츠의 조회 수가 업로드된 기간에 영향을 받을 가능성을 완전히 배제할 수 없을 것이다. 이 연구는 자료를 수집 과정에서 콘텐츠를 시청하는 연령대와 성별 등 인구통계학적 특성을 파악하기 어려운 부분이 있었다. 향후 연구에서는 연령대와 성별 등 인구통계학적 특성에 따른 시청자의 콘텐츠의 유형 관심도를 파악하는 연구가 이루어지길 기대한다.

CONFLICT OF INTEREST

논문 작성에 있어서 어떠한 조직으로부터 재정을 포함한 일체의 지

원을 받지 않았으며 논문에 영향을 미칠 수 있는 어떠한 관계도 없음을 밝힌다.

AUTHOR CONTRIBUTION

Conceptualization: HY Yun, Data curation: HY Yun, MS Jeon, Formal analysis: HY Yun, Funding acquisition: HY Yun, Methodology: HY Yun, MS Jeon, Projectadministration: JH Park, Visualization: MS Jeon, Writing-original draft: HY Yun, MS Jeon, Writing-review&editing: JH Park

참고문헌

- Ahn, S. Y., Park, W. I., & Hong, K. S. (2022). Effects of 8-week online-based Tabata home training on body composition, health-related physical fitness, and mental health in male adolescents. *Asian Journal of Physical Education of Sport Science*, 10(2), 69-81.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Cha, M. Y. (2020. 3. 9.). Increase in single people due to COVID-19 increase in sales of hiking and biking products. *Asia Economy*. Retrieved from <https://view.asiae.co.kr/article/2020030907493407240>
- Chandra, Y., Jiang, L. C., & Wang, C. J. (2016). Mining social entrepreneurship strategies using topic modeling. *PloS one*, 11(3), e0151342.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems*, 22.
- Cho, T. S. (2020). The effect of quality of service on exercise satisfaction and reuse of home training applications. *Korean Journal of Physical Education*, 59(4), 223-234.
- Choi, C. H. (2020). Investigation on the differences in participation motive, satisfaction and intention to continuous participation by the types of sports: Focusing on potentials of home training via the media from the perspective of social distancing. *The Korean Journal of Sport*, 18(3), 479-488.
- Choi, S. H., Lee, W. O., Kim, H. J., Won, J. S., Lee, J. H., Lee, Y. J., & Kim, J. W. (2017). Home training system: focusing on immersion, perceived competition, and learning motivation. *Korean Journal of the Science of Emotion & Sensibility*, 20(3), 119-130.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.
- Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, 93(2), 332-359.
- Hofmann, T. (1999, August). Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 50-57.
- Jeon, H. M., & Pan, Y. H. (2020). A classification of type and user interaction model of online home training service in the social distancing environment. *Design Convergence Study*, 19(4), 15-29.
- Kim, D. K., Han, J. W., & Choi, K. K. (2019). The relationship among home training product's perceived value, intention to continuous use and repurchase intention. *Journal of Sport and Leisure Studies*, 77, 121-130.
- Kim, J. H. (2021). A study on the characteristics of views of youtube home training content: Focusing on exercise type and optimal video duration. *Korean Journal of Sport Management*, 26(5), 22-41.
- Kim, J. H., Oh, J. H., Kim, T. Y., Yoo, J. K., Choi, H. J., Lee, D. G., ... & Park, W. I. (2020). *Corona 19 from the perspective of Gyeonggi-do residents*. Issue & Diagnosis, 1-25.
- Kim, K. S. (2022). Analysis on home training issue network and topic modeling before and after corona using media big data. *Korean Journal of Convergence Science*, 11(2), 39-54.
- Kim, K. Y. (2020. 8. 5.). Ministry of Culture, Sports and Tourism prepares a plan to nurture the non-face-to-face sports market and supports 5.5 billion won. *Yonhap News*. Retrieved from <https://www.yna.co.kr/view/AKR20200805047800007?input=1195m>
- Ko, S. S. (2020. 12. 30.). AI service corona effect Home Training and karaoke popularity. *Naeil*. Retrieved from http://www.naeil.com/news_view/?id_art=372228
- Kwon, J. Y., & Nam, S. B. (2022). How has the home training trend changed before and after the COVID-19 pandemic? *Journal of Sport and Leisure Studies*, 90, 279-293.
- Lee, B. S. (2020a. 6. 4.). Incheon Sports Association - Incheon Sports Science Center, home training video production and provision. *Joongboo*. Retrieved from <http://www.joongboo.com/news/articleView.html?idxno=363424547>
- Lee, H. J. (2020b). *A narrative study on home-training for adult men and women*. Master's thesis, Sogang University.
- Lee, S. H., & Kwak, J. H. (2020). Analysis of YouTube channel related to home training through un contact generation. *The Korea Journal of Sports Science*, 29(6), 181-191.
- Ministry of Culture, Sports and Tourism (2020). *2020 National living sports survey result report*. Retrieved from https://www.mest.go.kr/kor/s_policy/dept/deptView.jsp?pCurrentPage=1&pType=07&pTab=01&pSeq=1442&pDataCD=0417000000&pSearchType=01&pSearchWord=
- Ministry of Health and Welfare (2020). COVID-19 national mental health survey. Retrieved from <https://www.khealth.or.kr/kps/rsrhBusnRept/view>
- Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010, June). The 2010 annual conference of the North American chapter of the association for computational linguistics. In *Proceedings of Human Language Technologies*, 100-108.
- Oh, J. H., Oh, J. W., & Cho, K. M. (2019). Research on consistent use intention of home-training program on personal media service youtube based on post-adoption model. *Journal of the Korea Convergence Society*, 10(2), 183-193
- Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1), 209-228.

- Roh, S. H., & Kim, M. S. (2020).** Theoretical paradigm of home training experience in the 4th industrial revolution: Focused on Technology Acceptance Model(TAM). *Journal of Sport and Leisure Studies*, 81, 93-108.
- Seo, J. H., & Kim, E. Y. (2022).** A study on the design direction of smart mat for the efficient home fitness in the Post-COVID-19 era. *Bulletin of Korean Society of Basic Design & Art*, 23(2), 255-273.
- Sheen, H. R. (2017).** *A study on branding strategy and platform development for vitalize of home workout*. Master's thesis, Hongik University.
- Shin, E. W. (2019).** Design and Realization of smart health system for home training. *Journal of Digital Contents Society*, 20(3), 451-457.
- Shin, H. S. (2020).** Use of YouTube text in the debate education: Based on college students' evaluation about using YouTube. *The Korean Society for Thinking and Communication*, 13(1), 7-42.
- Tabata, I. (2019).** Tabata training: One of the most energetically feective high-intensity intermittent training methods. *The Journal of Physiological Sciences*, 69, 559-572.
- Yun, H. J., Park, J. H., & Yoon, J. W. (2019).** Introduction of topic modeling for extracting potential information from unstructured text data: Issue analysis on news article of dementia-related physical activity. *Korean Journal of Sport Science*, 30(3), 501-512.
- Yun, H. J., Park, J. H., Yoon, J. W., & Jeon M. S. (2021).** Issue analysis applying topic modeling on sports field due to spread of COVID-19. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 23(1), 11-20.
- Yun, S. U., & Kim, G. (2021).** Study on the use factors of YouTube-based home training content. *Journal of Digital Convergence*, 19(2), 345-355.

유튜브 플랫폼 홈트레이닝 동영상 콘텐츠의 유형 및 관심도 분석

윤효준¹, 박재현¹, 전민수²

¹한국체육대학교

²단국대학교

[목적] 이 연구의 목적은 유튜브 플랫폼에 홈트레이닝 동영상 콘텐츠의 유형과 관심도를 분석하는 것이다.

[방법] 연구의 목적을 달성하기 위해 python을 활용한 웹크롤링을 실시하여 3937개의 콘텐츠 정보(제목, 내용, 조회 수, 업로드 날짜)를 수집하였으며, 중복되는 콘텐츠와 관련 없는 콘텐츠를 제외하고 최종적으로 3155개의 콘텐츠를 연구자료로 선정하였다. 텍스트 자료는 3단계의 전처리 거쳤으며, 주요 키워드를 확인하기 위해 키워드의 TF와 TF-IDF를 산출하였고, 유형을 파악하기 위해 토픽모델링 중 LDA 알고리즘 적용하였다. 유형별 관심도를 파악하기 위해 콘텐츠의 조회수를 할당된 유형의 페센트로 재분배하였다. 모든 자료처리는 python 3을 활용하였다.

[결과] 첫째, 홈트레이닝 동영상 콘텐츠의 유형은 유산소 및 근력 강화를 위한 맨몸 전신 운동, 코어 및 상체 강화를 위한 필라테스운동, 도구를 활용한 상체운동, 하체 라인 운동, 통증 완화를 위한 자세교정 및 상체 스트레칭운동, 힙업운동, 다이어트를 위한 댄스 및 타바타운동, 다이어트를 위한 식단 및 하체교정 스트레칭운동, 코어 및 하체강화를 위한 맨몸운동으로 분류되었다. 둘째, 유산소 및 근력 강화를 위한 맨몸 전신 운동, 다이어트를 위한 댄스 및 타바타 운동 유형, 다이어트를 위한 식단 및 하체교정 스트레칭운동의 콘텐츠 비중과 관심도가 높은 것으로 나타났다.

[결론] 이는 현재 활성화되고 있는 온라인 홈트레이닝 시장 발전에 기초자료로 활용될 수 있을 것이다.

주요어

홈트레이닝, 동영상 콘텐츠, 토픽모델링, 유형분석, 관심도분석