ISSN 2509-2817 KiDRS Korea Institute of Design Research Society 통권 32호, 2024. Vol.9 No.3

BERTopic과 색채분석 기반 감성 추출을 이용한 관광지 추천 시스템 개발

Development of a Tourist Attraction Recommendation System Using BERTopic and Color Analysis-Based Sentiment Extraction

주 저 자 : 김미숙 (Kim, Mee Sook) 배재대학교 디자인학부 겸임교수

교 신 저 자 : 송준우 (Song, Junwoo) SAIT 책임연구원
soripe711@gmail.com

https://doi.org/10.46248/kidrs.2024.3.37

접수일 2024. 08. 25. / 심사완료일 2024. 09. 04. / 게재확정일 2024. 09. 09. / 게재일 2024. 09. 30.
본 논문은 SAIT 과제와 관련이 없습니다.

Abstract

This study aims to develop a personalized tourist destination recommendation system by analyzing the color and emotional factors of destination images. It evaluated and analyzed the emotional responses and color characteristics conveyed by images of natural resorts and urban tourist sites. A survey was conducted to collect users' emotional responses to these images, and a BERTopic model based on BERT was used to analyze text data from social media. The survey results provided a clear understanding of how users react to specific colors and emotional factors. For natural resorts, Green(G) was the most associated color, followed by Green-Yellow(GY) and Blue(B). In the case of urban tourist sites, Yellow-Red(YR) and Yellow(Y) were most commonly associated, followed by Blue-Green(BG) and Blue(B). Additionally, during the system's execution, the Bhattacharyya distance was used to measure color similarity between images, displaying images that match the user's emotions. Therefore, this study developed a recommendation system that considers the impact of destination images on users' emotions, proposing a new approach to better reflect users' emotional needs.

Keyword

Machine Learning(머신러닝), Color Analysis(색채분석), Tourism Recommendation System(관광 추천시스템)

요약

본 연구는 관광지 이미지의 색채적 및 정서적 요인을 분석하여 맞춤형 관광지 추천 시스템을 개발을 목표로 하고 있다. 자연휴양지와 도시관광지의 이미지를 대상으로 각 이미지가 전달하는 정서적 반응과 색채 특성을 평가 및 분석하였다. 설문조사를 통해 사용자들의 관광지 이미지에 대한 감정적 반응을 수집하고, BERT 기반의 BERTopic 모델을 사용해 SNS의 텍스트 데이터를 분석하였다. 설문조사 결과는 사용자들이 특정 색채와 정서적 요인에 대해 어떻게 반응하는지를 명확히 파악하였다. 자연휴양지는 초록(G)계열이 가장 연상되는 색채이며, 그 다음으로 연두(GY) · 파랑(B) 계열 순으로 확인되었다. 도시관광지의 경우 주황(YR) · 노랑(Y)계열이 가장 먼저, 그 뒤로 청록(BG) · 파랑(B) 계열 순으로 연상 결과가 나왔다. 또한 시스템 실행 시, Bhattacharyya 거리를 활용해 이미지 간의 색채 유사도를 측정하여 사용자의 정서에 맞는 관광지 이미지를 사용자에게 보여주었다. 그리하여 본 연구는 사용자의 정서적 요구를 반영할 수 있는 새로운 접근 방식을 제안하기 위해 관광지 이미지가 사용자의 감정에 미치는 영향을 고려한 추천 시스템을 개발하였다.

목차

1. 서론

1-1. 연구의 배경
1-2. 연구의 목적 및 방법

2. 이론적 배경

2-1. 관광지 이미지와 색채의 역할
2-2. 정서와 감정의 이해
2-3. BERTopic 기반 정서 분석과 추천 시스템
2-4. 이미지간 유사도 계산

3. 연구 방법

3-1. 추천 시스템 Overall
3-2. 데이터 수집 범위 및 방법
3-3. 설문 조사 방법
3-4. Text 분석을 위한 BERTopic 기법

4. 연구 결과

4-1. 설문조사 결과

4-2. 추천 시스템

5. 결론

참고문헌

1. 서론

1-1. 연구의 배경

현대 사회에서 관광산업은 경제적, 사회적 가치 창출의 중요한 요소로 자리매김하고 있다. 디지털 기술의 발달로 인해 관광자의 정보 탐색과 의사결정 과정 또한 급격히 변화하고 있다. 특히, 사용자 생성 콘텐츠(User-Generated Content, UGC)플랫폼인 소셜 네트워크 서비스(Social Network Service, SNS)의 확산은 관광산업에 새로운 도전을 제시하고 있다. 인스타그램과 같은 이미지 기반 SNS는 사용자가 관광지를 시각적으로 경험하고 공유하는 주요 매체이다. 관광자들은 이를 통해 여행 전, 중, 후의 경험을 나누고 관광지에 대한 첫인상을 형성하게 된다. 색채는 중요한 시각적 요소로, 특정 관광지에 대한 감정적 반응과 지각된 매력을 결정하는데 중요한 역할을 한다. 관광지 이미지의 색채는 그 장소의 분위기와 특성을 나타내며, 관광자의 의사결정 과정에 큰 영향을 미칠 수 있다. 그러나 관광지 이미지에서 색채가 미치는 영향에 대한 연구는 색채의 일부 요소에만 집중하는 경향이 있다.

1-2. 연구의 목적 및 방법

본 연구의 주요 목적은 관광지 이미지의 색채적 요소와 텍스트 기반의 정서적 요인을 종합적으로 분석한다. 즉 개인의 감정과 경험을 고려한 맞춤형 관광지 추천 시스템을 개발하는 데 있다. 이를 위해 딥러닝 및 자연어 처리 기술을 활용하였다. 그리하여 본 연구는 인스타그램에서 수집된 관광지 이미지와 관련된 텍스트 데이터를 기반으로 색채적, 정서적 요인을 다각적으로 분석하고자 한다.

먼저, 설문조사를 통해 사용자들이 다양한 관광지 이미지에 대해 느끼는 감정과 정서적 반응을 수집한다. 이 설문조사는 관광지 이미지에 대한 주관적 인식을 파악하고, 각 이미지의 색채와 관련된 감정적 반응을 분석하기 위한 데이터를 제공하였다. 수집된 설문조사 데이터를 기반으로 색채적 요인과 정서적 요인 간의 상관관계를 도출하여 관광지 선택에 있어 중요한 정서적 영향을 파악한다.

관광지 이미지를 대상으로 색채분석을 수행하여 주요 색채적 특징을 도출하기 위해 설문조사를 수행하였다. 이미지에서 색채 관련 정서적 요소들을 추출하고, 관광지의 분위기와 특성을 정량적으로 분석한다. 동시에, BERTopic 모델을 활용하여 관련 텍스트 데이터를 분석함으로써 관광지에 대한 정서적 반응과 감정적 표현을 추출하고 이를 색채분석 결과와 결합하였다.

그리하여 본 연구는 관광지 이미지의 색채와 정서적 요소를 종합적으로 고려한 관광지 추천 시스템을 개발하고자 한다. 이를 통해 관광자의 개별적 감정과 경험에 부합하는 맞춤형 관광지 선택을 가능하도록 한다. 이러한 접근은 기존의 텍스트 기반 추천 시스템의 한계를 극복하고, 보다 정교하고 개인화된 관광지 추천을 제공할 수 있는 새로운 가능성을 제시한다.

2. 이론적 배경

2-1. 관광지 이미지와 색채의 역할

관광지 이미지는 관광객의 의사결정 과정에 영향을 미치는 요소로 특히 색채는 관광지 이미지의 첫인상을 결정짓는 중요한 시각적 요소 중 하나이다. 색채는 특정 장소의 분위기와 감성을 전달하는 강력한 도구로 관광지의 매력을 시각적으로 표현하고 잠재 관광자들에게 강한 인상을 줄 수 있다. Valentini et al.(2018)와 Correy(1983)에 따르면 관광지 이미지에서 색채는 자연적이거나 인위적인 환경을 이해하는데 필수적인 정보 통신 수단으로써 그 장소의 분위기와 이미지를 지배하는 핵심 요소로 작용한다[1]. 색채는 관광지의 첫인상을 형성하며 관광자의 감성적 반응을 유도할 수 있는 중요한 시각적 인상을 남긴다. 하지만 현재 관광산업의 관

1) Valentini, C., Romenti, S., Murtarelli, G. & Pizzetti, M., 'Digital visual engagement: Influencing purchase intentions on Instagram', Journal of Communication Management, 2018, Vol.22, No.4, pp.362-381.

광지 이미지에 관한 연구는 시각적 부분 관련 연구가 늘어나고 있지만, 색채 영향력을 조사한 연구는 상대적으로 제한적이다[2]. 또한, 디지털 사진에 대한 관심이 증가하고 있음에도 여전히 온라인 사진의 영향에 대한 연구는 초기 단계에 머물러 있다[3][4]. 특히 인스타그램과 같은 플랫폼에 업로드된 사진들에 대한 색채 관련 연구는 제한적이다. 이로 인해 색채와 정서적인 요소를 함께 고려하여 관광지 선택에 영향을 주는 연구의 분석이 필요하다.

2-2. 정서와 감정의 이해

정서는 인간의 행동과 의사결정 과정에 중요한 요소로, 관광지 선택에서도 중요한 요인이다. 정서는 특정 장소나 경험에 대한 개인의 주관적 반응을 포함하며, 관광객이 특정 관광지 선택에 중요한 역할을 할 수 있다. 이러한 정서적 반응은 관광지에 대한 선호도와 선택에 영향을 미치고 개인의 경험과 밀접하게 연관되어 있다[5][6]. 특히 관광지 이미지와 관련된 정서적 반응은 색채와 같은 시각적 요소로 강화될 수 있다. 그리하여 본 연구는 관광지 추천 시스템 개발을 위해 정서적 요인을 색채적 요소와 결합시켜 관광지 추천을 개인 성향에 맞게끔 추천 결과를 제공하고자 한다.

2-3. BERTopic 기반 정서 분석과 추천 시스템

BERTopic은 BERT(Bidirectional Encoder Representations from Transformers) 임베딩을 활용한 토픽 모델링 기법으로 문서 내에서 주요 주제를 파악하고 의미 있는 정보를 추출하는데 사용된다[7]. BERTopic의 문맥적 임베딩 기능은 텍스트 데이터를 이해하고 문서 관련 토픽들을 자동으로 추출한다. Zhang et al.(2019)는 BERT를 활용하여 사용자 리뷰를 분석하고 이를 기반으로 맞춤형 추천을 제공하는 시스템을 제안하였다[8]. 이 연구에서는 리뷰의 감정적 내용과 사용자 선호도를 반영하여 사용자가 만족할 수 있는 추천을 생성하였다. Nguyen et al.(2024)은 호텔 관련 베트남어 댓글에서 추출된 문맥 벡터를 포함하여 호텔 서비스의 여러 측면(서비스, 인프라, 위생, 위치, 태도)에 대한 평가를 제공하였다. 특히 텍스트 데이터에서 주제와 감정을 효과적으로 추출하고 이를 추천 시스템에 반영할 수 있음을 보여주었다[9]. 본 연구에서는 기존 연구와는 다르게 관광지 추천 시스템을 제안하기 위해 다음과 같이 크게 두 가지 단계를 수행하였다. 먼저, 관광지 이미지별 어떠한 감정을 가지는지에 대한 설문조사를 수행하였다. 두 번째, BERTopic 모델을 통해 해당 이미지 관련 Text 데이터를 기반으로 정서 분석과 토픽 모델링을 진행한다. 이를 앞서 수행한 설문조사에서 분석된 결과와 매핑시켜 사용자 정서와 선호도를 반영한 맞춤형 추천을 제공한다.

2-4. 이미지간 유사도 계산

본 연구에서는 Bhattacharyya 거리(Bhattacharyya distance)를 사용하였다. 각 이미지에 분포하는 색상들에 대한 히스토그램을 이용해 이미지 간의 유사성을 평가할 때 사용되는 지표이다. Bhattacharyya 거리는 두 확률 분포의 겹침 정도를 계산함으로써, 두 분포간의 유사성을 측정할 수 있다. 해당 유사도 계산은 다양한 연구 분야에서 사용되고 있다. 최준호, 조미영, & 김판구(2004)는 이미지 검색 시스템에서 보다 효율적이고 유의미한 검색 결과를 제공하기 위해 Bhattacharyya

2) Yu, C. E., Xie, S. Y. & Wen, J., 'Coloring the destination: The role of color psychology on Instagram', Tourism Management, 2020, Vol.80, pp.104-110.

3) Lee, A. H., 'What does colour tell about tourist experiences?', Tourism geographies, 2020, Vol.25, No.1, pp.136-157.

4) Li, Y. & Xie, Y., 'Is a picture worth a thousand words? An empirical study of image content and social media engagement', Journal of Marketing Research, 2020, Vol.57, No.1, pp.1-19.

5) Boulding, K. E., 'Three Faces of Power', SAGE Publications, Inc., 1990, pp.264.

6) Kotler, P., Haider, D. H. & Rein, I., 'Marketing Places: Attracting Investment, Industry, and Tourism to Cities, States, and Nations'. Free Press, 1993, pp.388.

7) Grootendorst, M.R., 'BERTopic: Neural topic modeling with a class-based TF-IDF procedure', arXiv e-print, 2022, https://arxiv.org/abs/2203.05794.

8) Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., & Artzi, Y., 'BERTScore: Evaluating Text Generation with BERT', arXiv e-print, 2019, https://arxiv.org/abs/1904.09675.

9) Nguyen, M. H., Nguyen, T. T., Ta, M. N., Nguyen, T. M., & Nguyen, K. V., 'RRS: Review-Based Recommendation System Using Deep Learning for Vietnamese', SN Computer Science, 2024, Vol.5, No.5, 492.

거리를 기반으로 이미지 간의 유사성을 측정하여 이미지 검색하는데 활용하였다[10]. 얼굴 인식 관련하여 권성근(2013)은 색채 히스토그램의 분포 특성을 통해 얼굴 인식 관련 정합성을 향상시켰다[11]. 그리하여 본 연구에서도 관광지 이미지 검색 엔진의 성능 향상을 위해 해당 유사도 계산을 활용하였다.

3. 연구 방법

3-1. 추천 시스템 Overall

본 연구는 색채 및 텍스트 기반의 정서적 요인을 종합적 고려한 관광지 추천 시스템을 구현 및 제안한다. [그림 1]는 제안하는 추천 시스템의 전반적인 프로세스를 보여준다.

1 Step: 메타(Meta)에서 제공하는 인스타그램 API 및 파이썬 라이브러리를 통해 “이미지 데이터”와 “텍스트 데이터”를 확보한다. 데이터 전처리와 분석을 수행하여 데이터 퀄리티를 높였다. 이를 기반으로 관광지 이미지별 정서적 요인을 조사하기 위해 설문조사에 사용할 정서 관련 형용사들을 정의하였다. BERTopic의 경우, 확보한 텍스트 데이터를 국가별로 클러스터링 하고 해당 국가와 관련된 주요 정서적 요인을 추출한다.

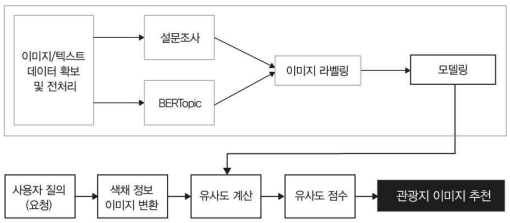
2 Step: 이미지별 정서적 및 색채적 요인을 도출하기 위해 수행한 설문조사 결과와 BERTopic으로부터 얻은 정서적 이미지와의 매핑 후 이미지 라벨링을 수행한다.

3 Step: 라벨링된 이미지를 학습데이터 활용하여 학습 및 모델링 수행한다.

4 Step: 학습된 모델을 통해 사용자 질의에 따라 색상(Pixel) 분석 또는 채도(Chroma) 분석을 선택하여 Color Palette를 생성한다.

5 Step: Color Palette에서 색채를 선택하면 해당 색상/채도와 실제 데이터 베이스에 저장된 관광지 이미지를 기반으로 이미지간 유사도 계산을 수행하고 점수를 얻는다.

6 Step: 유사도 점수를 ranking화 하여 유사도 점수가 상위 3개 안에 들어가는 이미지들을 사용자에게 보여준다.



[그림 1] 추천 시스템 Overall

3-2. 데이터 수집 범위 및 방법

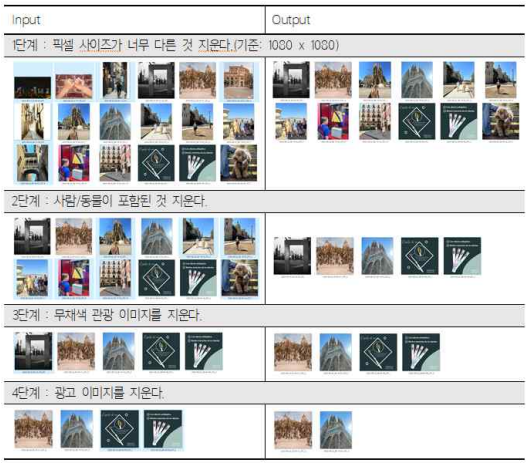
본 연구는 BERTopic과 설문조사를 통해 관광지 이미지의 색채적 요소와 텍스트 기반의 정서적 요인을 종합적으로 분석하여 맞춤형 관광지 추천 시스템을 제안한다. 인스타그램을 기반으로 모델링하였다. 인스타그램 API 및 데이터 크롤링 파이썬 라이브러리인 Instaloader를 통해 4개의 자연휴양지와 4개의 도시관광지로부터 74,000개의 이미지 파일과 81,916개의 텍스트 데이터(게시글, 해시태그 및 댓글 등)를 수집하였다([표 1] 참고). 수집된 이미지와 텍스트에는 추천 시스템 성능 저하를 야기할 수 있는 노이즈가 존재한다. 이를 최소화하기 위해 데이터 전처리를 수행하였다([그림 2],[그림 3] 참고). 다음 [그림 2]는 관광지 이미지 데이터 전처리 과정을 보여준다. [그림 3]은 불용어(Stopwords) 처리 과정을 보여준다. 불용어는 불필요한 단어나 토큰들을 의미한다. 본 연구에서는 BERTopic 라이브러리가 정의하는 불용어를 기준으로 전처리를 수행하였다. 그리하여 불용어로 정의된 부사와 조사를 제거하였고, 이에 더해 연구 목적에 맞지 않은 동사(예, is, would, can등)들 또한 필터링 작업을 수행하였다. 그리하여 41,218개의 이미지와 19,201개의 텍스트를 최종적으로 연구 목적에 맞게 확보하였다.

[표 1] 데이터 수집 범위

분류	관광지명(국가)
자연휴양지	아부다비(아랍에미레이트), 브라이튼(영국), 하와이(미국), 센토사(싱가포르)
도시관광지	두바이(아랍에미레이트), 런던(영국), 뉴욕(미국), 싱가포르(싱가포르)

10) 최준호, 조미영, & 김판구, ‘컬러 분포와 WordNet상의 유사도 측정을 이용한 의미적 이미지 검색’, 정보처리학회논문지, 2004, Vol.11 No.5, pp.509-516

11) 권성근, ‘영상의 색상 분포 정합을 이용한 얼굴 검출 알고리즘’, 멀티미디어학회논문지, 2013, Vol.16, No.8, pp.927-933



[그림 2] 이미지 데이터 전처리 과정

3-3. 설문 조사 방법

본 연구는 소셜미디어에서 사용된 관광지 이미지들에 대한 정서적 요인을 분석하기 위해 2023년 6월 1일부터 7월 1일까지 31일 동안 250명에게 설문조사를 수행하였다. 설문자들의 성비 구성은 여자 165명, 남자 85명이다. 설문지 구성은 [표 2]와 같다. 설문지는 크게 6개 항목(인구통계학, 관광지 이미지, SNS 사용, 정서, 관광행동)으로 되어 있고, 총 32문항으로 구성되어 있다. 그리고 관광지 이미지의 색채 및 정서적 요인이 관광지 선택에 미치는 요인을 분석하기 위해 전처리한 지역별 이미지들을 사용하여 총 64개의 문항을 만들었다. 이미지 선정 관련하여 자연휴양지는 대상지의 대표적인 자연 경관(해변가, 바닷가, 산지, 사막 등)을 중심으로 선정하였고, 도시관광지는 랜드마크(건축물, 타워, 교량 등)을 중식으로 이미지를 선정하였다. 설문 조사 시, 특정 국가나 도시명을 명시하지 않고 진행하였다. 이는 참가자들이 색채와 관련된 고정관념에 영향을 받지 않도록 하기 위함이다. 또한 설문 문항의 위치는 랜덤으로 배치하여 특정 도시가 중복되지 않도록 했다. 결과는 고평가("매우 그렇다"와 "그렇다")와 저평가("전형 아니다"와 "아니다")로 분류되었다. 이 분류는 관광지 선택에 미치는 색채의 정서적 요인 관련 중요성을 평가하는데 사용된다.

	단계	내용
1	Input Data 가져오기	"Exploring the vibrant streets of Barcelona, each corner tells a story of rich history and breathtaking architecture. From the iconic Sagrada Familia to the lively La Rambla, every step unveils a new adventure in this captivating city. #Barcelona #TravelDiaries +"
2	Tokenization	{'Exploring', 'the', 'vibrant', 'streets', 'of', 'Barcelona,', 'each', 'corner', 'tells', 'a', 'story', 'of', 'rich', 'history', 'and', 'breathtaking', 'architecture.', 'From', 'the', 'iconic', 'Sagrada', 'Familia', 'to', 'the', 'lively', 'La', 'Rambla,', 'every', 'step', 'unveils', 'a', 'new', 'adventure', 'in', 'this', 'captivating', 'city.', '#Barcelona', '#TravelDiaries'}
3	품사별 Clustering	①명사(Noun): streets, corner, story, history, architecture, Sagrada Familia, La Rambla, step, adventure, city, Barcelona, TravelDiaries ②동사(Verb): Exploring, tells, unveils ③형용사(Adjective): vibrant, rich, iconic, lively, captivating ④부사(Adverb): each, breathtaking, From, every, new ⑤조사(Particle): of, in, to, From, the, a, this, #
4	불용어 삭제 (부사,조사 제거)	①명사(Noun): streets, corner, story, history, architecture, Sagrada Familia, La Rambla, step, adventure, city, Barcelona, TravelDiaries ②동사(Verb): Exploring, tells, unveils ③형용사(Adjective): vibrant, rich, iconic, lively, captivating
5	연구 목적에 맞게 동사 제거	①명사(Noun): streets, corner, story, history, architecture, Sagrada Familia, La Rambla, step, adventure, city, Barcelona, TravelDiaries ③형용사(Adjective): vibrant, rich, iconic, lively, captivating
6	최종 Output Data	{'streets', 'corner', 'story', 'history', 'architecture', 'Sagrada Familia', 'La Rambla', 'step', 'adventure', 'city', 'Barcelona', 'TravelDiaries','vibrant', 'rich', 'iconic', 'lively', 'captivating'}

[그림 3] 불용어 처리 과정

[표 2] 설문지 구성

	문항		문항수	척도
I	인구통계학적 요인		12	명목척도
II	관광지이미지 요인		10	Likert 5점 척도
III	SNS 사용 요인		15	Likert 5점 척도
IV	색채적 요인	자연휴양지	16	Likert 5점 척도
		도시관광지	16	Likert 5점 척도
V	정서적 요인	자연휴양지	16	Likert 5점 척도
		도시관광지	16	Likert 5점 척도
VI	관광행동 요인		13	Likert 5점 척도
	계		114	

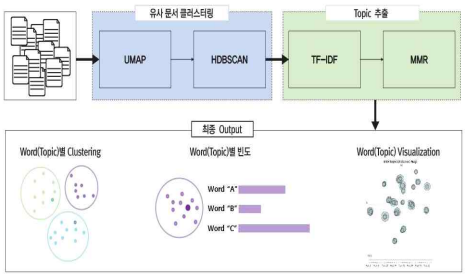
마지막으로 설문조사에 사용된 감정 형용사는 BERTopic을 통해 추출하였다. 전처리 과정으로부터 선정된 관광지 이미지와 텍스트에서 상위 6개의 형용사를 추출하였다. 그 결과, 자연휴양지 이미지에서 추출된 형용사는 'sweet', 'mood', 'awesome', 'beautiful', 'likeable', 'natural'이며, 도시관광지 이미지에서는 'lovely', 'beautiful', 'mood', 'inspired', 'aesthetic', 'creative'가 선정되었다. 이러한 형용사들은 각각 자연휴양지와 도시관광지의 정서적 요인을 평가하는 설문조사에 사용되었다.

[표 3] 관광지 이미지에 대한 정서적 요인 조사에 사용된 관광지 예시 이미지

국가	자연 휴양지	대표 이미지	도시 관광지	대표 이미지
영국	브라이튼		런던	
미국	하와이		뉴욕	
아랍 에미레이트	아부다비		두바이	
싱가포르	센토사		싱가포르	

3-4. Text 분석을 위한 BERTopic 기법

본 연구에서는 자연어 처리에서 우수한 성능을 보이는 BERT 모델을 기반으로 하는 BERTopic을 사용하였다. BERTopic은 문장 내 모든 단어를 동시에 고려하여 양방향으로 학습함으로써 더 좋은 문맥적 이해가 가능하다. 그 결과 문맥상에서 주요 정서적인 요인(주제)을 추출할 수 있다. 즉, BERTopic은 BERT 언어 모델을 활용하여 텍스트 데이터를 수치화하고, 이 수치를 바탕으로 유사한 텍스트를 묶어 주제를 찾아 시각화한다.



[그림 4] BERTopic 프로세스

[그림 4]는 BERTopic의 전반적인 프로세스를 보여준다. 첫 번째, 인스타그램에서 수집 및 정제한 데이터를 삽입한 후, 문서의 내용을 벡터로 변환한다. 이는 텍스트 문서의 단어들을 임베딩하여 문서의 주제나 구조를 파악하는 단계이다. 두 번째 UMAP (Uniform Manifold Approximation and Projection)은 문서 벡터의 차원을 축소하는 데 사용된다. 데이터의 전역적 구조를 잘 보존하면서 차원 축소를 수행하기 때문에 정보의 손실이 없다. 세 번째, HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise)은 데이터의 밀도 기반 클러스터링을 다음과 같은 단계로 수행한다.

1) 밀도 공간 생성: 데이터 공간에서 각 점의 밀도를 계산하여 밀도 공간을 생성한다.

2) 거리 가중치 그래프 생성: 각 데이터 포인트 간의 거리를 기반으로 가중치 그래프를 생성한다. 이 그래프는 데이터 포인트 간의 연결성과 밀도를 나타낸다.

3) 계층적 클러스터링 모델 구축: 연결된 점들 간에 계층적 클러스터링을 수행하여 다양한 밀도 수준에서의 클러스터를 형성한다.

4) 클러스터 계층 구조 축약: 최소 클러스터 크기(Minimum Cluster Size)를 기준으로 클러스터 계층 구조를 축약하여 불필요한 클러스터를 제거한다.

5) 안정적인 클러스터 추출: 축약된 클러스터 계층에서 안정성이 높은 클러스터만을 추출한다. 클러스터의 안정성은 클러스터를 구성하는 데이터 포인트들의 밀도를 기반으로 계산된다.

네 번째, TF-IDF(Term Frequency-Inverse Document Frequency)는 클러스터 내 단어들의 중요도를 분석한다. TF-IDF는 특정 단어가 특정 문서에 나타난 횟수(TF), 전체 문서에 특정 단어가 나타난 횟수(DF), 그리고 전체 문서 수에 대한 특정 단어의 문서 빈도의 역수(IDF)를 기반으로 계산된다. 마지막으로 MMR (Maximal Marginal Relevance)은 정보 검색 및 요약 과제에서 사용된다. 이는 다양한 문서나 문장에서 특정 정보를 추출할 때 정보의 다양성과 중요성 사이의 균형을 유지하는 데 목적이 있다. MMR은 이미 선택한 정보와의 관련성 및 중복성을 고려하여 정보의 다양성을 유지하면서 각 클러스터의 대표 문장 또는 핵심 키워드를 선택하는 데 사용된다.

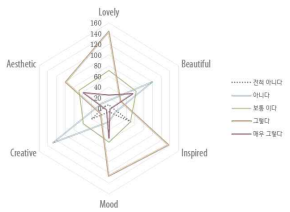
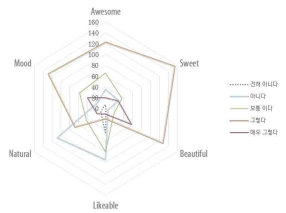
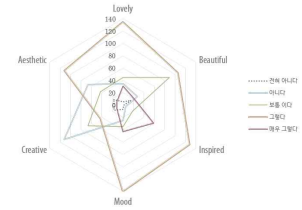
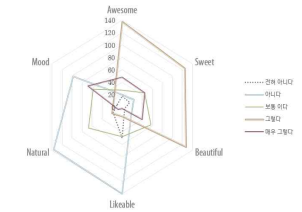
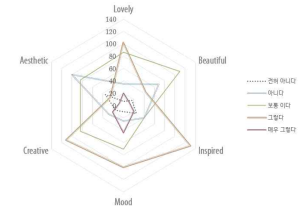
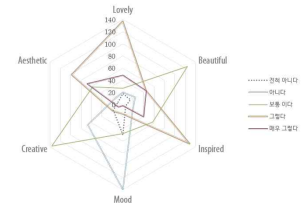
4. 연구 결과

4-1. 설문조사 결과

전체 표본은 250명 중 여성 165명, 남성 85명으로 구성되어있고, 연령은 20대 155명, 30대 63명, 40대

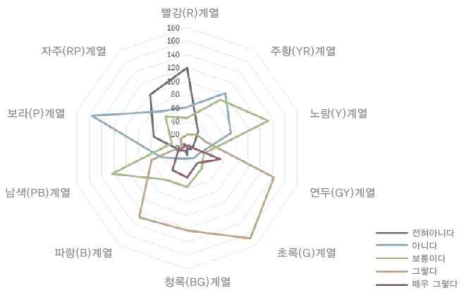
32명으로 가장 활동이 많은 연령대로 구성되어 있다. 기본 조사 결과, 여성(72.7%)이 남성(56.4%)에 비해 국외 관광을 선호하는 것으로 나타났다. 또한, 전체 49.2%가 관광지에 대한 정보를 SNS를 통해 얻는 것을 알 수 있었다. 그리고 전체 54%가 자연관광지에 비해 도시관광지를 더 선호하는 것으로 나타났다. 그리하여 표본 250명의 설문 자료를 기반으로 관광지 이미지별 정서적 요인에 대한 정보를 확보 및 분석할 수 있었다. [표 4]는 자연휴양지와 도시관광지 이미지에서 색채가 불러일으키는 정서적 요인에 대한 분석 결과를 정리한 것이다. 자연휴양지에서는 'sweet'이 공통적으로 긍정적인 평가를 받았다. 반면, 'likeable'과 'beautiful', 'creative'가 주로 부정적인 평가를 받았고, 'natural'은 중립적인 평가를 받았다. 도시관광지에서는 'lovely'가 전반적으로 긍정적인 평가를 많이 받았고, 'inspired'가 그 뒤를 이었다.

[표 4] 정서적 요인에 대한 분석 결과

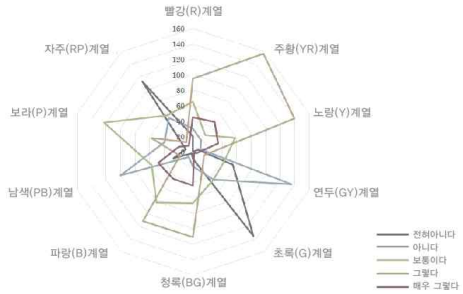
국가	지역	설문 결과
영국	브라이튼	
	런던	
미국	하와이	
UAE	뉴욕	
	아부다비	
	두바이	
싱가포르	센토사	
	싱가포르	

[그림 5]와 같이 자연휴양지의 경우 초록(G) 계열이 가장 연상되는 색채이며, 그 다음으로 연두(GY)·파랑(B) 계열 순으로 확인되었다. 자연휴양지와 가장 상반되는 느낌으로는 빨강(R)·보라(P)·자주(RP) 계열의 색채이다. 이러한 결과는 주로 자연휴양지의 이미지를 형성하는 일반적 환경적 요인인 산지, 숲, 바다, 해안가 등의 자연 자원이 색채를 연상하는데 영향을 준 것으로 파악된다. [그림 6]은 도시관광지의 색채적 요인에 관한 설문 결과이다. 주황(YR)·노랑(Y) 계열이 가장 먼저, 그 뒤로 청록(BG)·파랑(B) 계열 순으로 연상된다고 하였다. 반면에 연상되지 않는 색채는 초록(G)·

연두(GY) 계열이고 자주(RP) 계열도 연상되지 않는 색채로 선택되었다. 도시관광지는 주로 역사 문화와 관련된 건축물 및 시설물, 그 도시를 상징할 수 있는 랜드마크를 중심으로 관광지가 형성되어 있다. 이러한 관광지의 요소들은 대부분 인공적으로 건축된 부분으로, 자연휴양지와는 달리 주로 인공적인 측면에서 연상될 수 있는 색채로 형성되었다.



[그림 5] 자연휴양지 색채적 요인 설문 결과



[그림 6] 도시관광지 색채적 요인 설문 결과

4-2. 추천 시스템

본 연구에서 제안하는 관광지 추천 시스템은 설문조사와 BERTopic 모델링을 기반으로 실제 관광자에게 긍정적 이미지를 끼칠 수 있는 관광지 이미지를 추출하도록 구축되었다. 추천 시스템 사용 프로세스는 다음 [그림 7]과 같다.

Step 1 : 관광지 분류 및 감정 형용사 선택

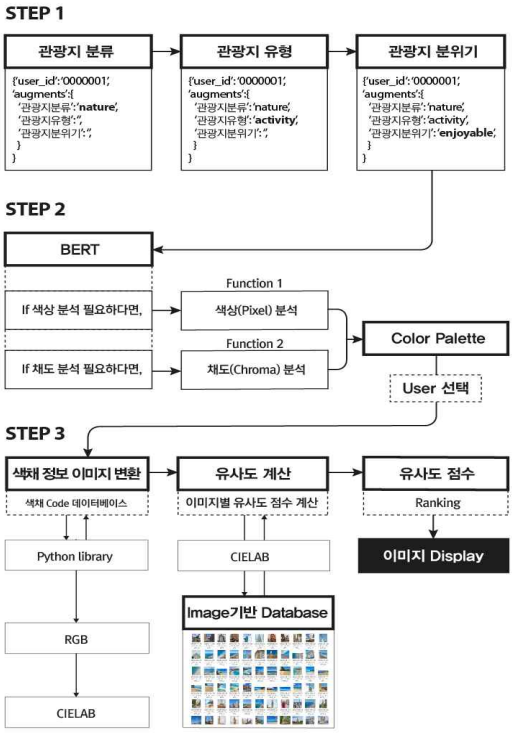
- 사용자는 자연휴양지 또는 도시관광지의 분류, 유형(휴식형, 활동형, 탐방형, 체험형) 그리고 특정 감정 형용사(예: sweet, beautiful)를 선택한다.

Step 2 : BERT 모델을 통한 텍스트 분석 및 색상 팔레트 생성

- 선택된 텍스트 정보는 앞서 설문조사와 BERTopic을 통한 데이터 처리 및 labelling 과정을 거쳐 학습된 BERT 모델에 입력된다. 이를 통해 색상 팔레트가 생성되며, 사용자가 선택한 이미지의 분석이 시작한다.

Step 3 : 이미지 처리 및 유사도 계산

- 이미지 크기 조절, 색상 변환, 유사도 계산 등의 이미지 처리 기능이 수행된다. 이 과정에서 Bhattacharyya 거리를 이용해 이미지간 색상 유사도를 계산한다. 그 결과 유사도가 높은 Top 3 이미지들을 추천한다.

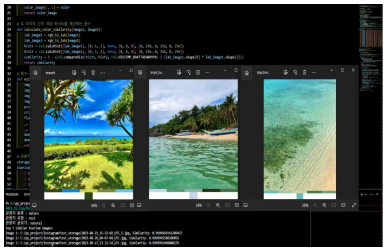


[그림 7] 추천 시스템 사용 프로세스

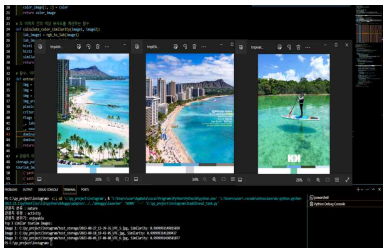
그리하여 [표 5]는 본 연구에서 제안한 추천 시스템의 실행 결과를 보여준다. 해당 추천 시스템에서 사용자는 유사도가 가장 높은 3개의 관광지 이미지 중 하나를 선택할 수 있으며, 원하는 이미지가 없을 경우 재검색이 가능하다. 재검색 시 이전과 다른 새로운 이미지가 추출될 수도 있다. 왜냐하면 비슷한 유사도 점수를 가진 이미지들이 있을 수 있기 때문이다. 마지막으로 한 개의 이미지를 선택하게 되면 이미지 파일명에 장소 정보가 있기 때문에 해당 이미지를 클릭했을 때 장소 정보가 나오도록 하였다.

[표 5] 추천 시스템 실행 결과

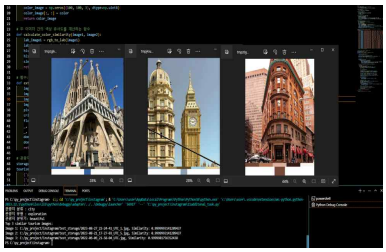
1 번째 테스트 결과 - 자연휴양지(휴식형)

	시스템 입력	내용
Input	관광지 분류	자연휴양지(nature)
	관광지 유형	휴식형(rest)
	관광지 분위기	자연스러운 (natural)
Output		

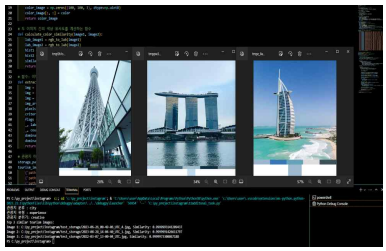
2 번째 테스트 결과 - 자연휴양지(활동형)

	시스템 입력	내용
Input	관광지 분류	자연휴양지(nature)
	관광지 유형	활동형(acticity)
	관광지 분위기	즐거운 (enjoyable)
Output		

3 번째 테스트 결과 - 도시관광지(탐방형)

	시스템 입력	내용
Input	관광지 분류	도시관광지(city)
	관광지 유형	탐방형 (exploration)
	관광지 분위기	아름다운(beautiful)
Output		

4 번째 테스트 결과 - 도시관광지(체험형)

	시스템 입력	내용
Input	관광지 분류	도시관광지(city)
	관광지 유형	체험형 (experience)
	관광지 분위기	창의적인(creative)
Output		

5. 결론

본 연구는 관광지 이미지의 색채적 및 정서적 요인을 분석하여 맞춤형 관광지 추천 시스템을 개발하고 제안하는 데 초점을 두었다. 연구 과정에서는 자연휴양지와 도시관광지의 다양한 이미지를 분석하였다. 각 이미지가 전달하는 정서적 반응과 색채 특성을 설문조사를 통해 평가하였다. 이를 통해 관광지 선택 과정에서 개인의 감정적 요구와 선호를 반영할 수 있는 맞춤형 추천이 가능함을 확인하였다. 특히, BERT 기반의 BERTopic 모델을 활용하여 텍스트 데이터를 심층 분석하였다. 이를 통해 주요 주제와 정서적 요인을 추출하여 사용자 경험을 더욱 개인화할 수 있는 추천 시스템을 구현하였다. 또한, Bhattacharyya 거리를 사용해 이미지 내 색상 히스토그램 비교하였다. 이 방법을 통해 이미지 간 색채 유사도를 정량적으로 계산하였다.

본 연구에서 제안한 추천 시스템은 정서적 요인과 색채적 요인을 기반으로 간단한 결과를 보여주었다. 이를 통해 관광지 이미지 추천 시스템의 기본 개념을 입증하였다. 제안된 시스템의 성능을 높이기 위해는 자연휴양지와 도시관광지의 이미지를 추천할 수 있는 데이터셋 구축이 필요하다. 현재 시스템은 파일럿 수준으로 기본적인 결과값만 제공한다. 그러나 향후 연구에서는 LLM function calling을 위한 finetuning 기법을 통해 더 정교하고 개인화된 추천 시스템으로 발전시킬 계획이다. 이러한 발전은 사용자에게 더 높은 가치를 제공하는 맞춤형 관광지 추천 시스템으로 이어질 것이다.

참고문헌

1. Boulding, K. E., Three Faces of Power, Thousand Oaks, California: SAGE Publications, Inc., 1990
2. Correy, J. F., Visual Perception and Scenic Assesment in Australia, IFLA year book, 1983
3. Kotler, P., Haider, D. H. & Rein, I., Marketing Places: Attracting Investment, Industry, and Tourism to Cities, States, and Nations. New York: The Free Press, 1993
4. 권성근, 영상의 색상 분포 정합을 이용한 얼굴 검출 알고리즘, 멀티미디어학회논문지, 2013, Vol.16, No.8
5. 최준호, 조미영, & 김판구, 컬러 분포와 WordNet상의 유사도 측정을 이용한 의미적 이미지 검색, 정보처리학회논문지, 2004, Vol.11 No.5
6. Grootendorst, M.R., 'BERTopic: Neural topic modeling with a class-based TF-IDF procedure', arXiv e-print, 2022, https://arxiv.org/abs/2203.05794.
7. Lee, A. H., What does colour tell about tourist experiences?, Tourism geographies, 2020, Vol.25, No.1
8. Li, Y. & Xie, Y., Is a picture worth a thousand words? An empirical study of image content and social media engagement, Journal of Marketing Research, 2020, Vol.57, No.1
9. Nguyen, M. H., Nguyen, T. T., Ta, M. N., Nguyen, T. M., & Nguyen, K. V., RRS: Review-Based Recommendation System Using Deep Learning for Vietnamese, SN Computer Science, 2024, Vol.5, No.5, 492.
10. Valentini, C., Romenti, S., Murtarelli, G. & Pizzetti, M., Digital visual engagement: Influencing purchase intentions on Instagram, Journal of Communication Management, 2018, Vol.22, No.4
11. Yu, C. E., Xie, S. Y. & Wen, J., Coloring the destination: The role of color psychology on Instagram, Tourism Management, 2020, Vol.80
12. Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., & Artzi, Y., 'BERTScore: Evaluating Text Generation with BERT', arXiv e-print, 2019, https://arxiv.org/abs/1904.09675.
13. 김미숙. 관광지 이미지에 대한 색채적 · 정서적 요인에 관한 연구: 머신러닝 방법론을 활용한 색채 분석을 중심으로. 홍익대학교 대학원 박사학위논문. 2024