ISSN 2509-2817 KiDRS Korea Institute of Design Research Society 통권 31호, 2024. Vol.9 No.2

머신러닝 기반 색채적 특성 분석 및 감성 분류에 관한 연구

사용자 생성 콘텐츠 내 관광지 이미지를 중심으로

Analysis of Color Characteristics and Emotional Classification Using Machine Learning

focusing on Tourist Destination Images based on User-Generated Content

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

교 신 저 자 : 송준우 (Song, Junwoo) Imperial College London
soripe711@gmail.com

https://doi.org/10.46248/kidrs.2024.2.312

접수일 2024. 05. 25. / 심사완료일 2024. 06. 03. / 게재확정일 2024. 06. 12. / 게재일 2024. 06. 30.

Abstract

This study focuses on developing a methodology for classifying tourist destination images based on their color characteristics and emotional categories using machine learning algorithms. By analyzing user-generated content collected through digital media and smartphones, particularly tourist destination images from Instagram, the research examines how the initial impressions and imagery of tourist destinations impact tourists' emotions. Color and text data were analyzed using machine learning algorithms and the BERT language model to interpret the color characteristics and emotional messages conveyed by these images. The research demonstrates that analyzing and classifying the color characteristics and emotions of tourist destination images significantly influence tourists' choices and emotions. It shows that considering these two factors in an interconnected manner, rather than in isolation, enhances their utility. Specifically, color characteristics are crucial elements that can improve the attractiveness of tourist destination images. This study reveals that these images can significantly affect tourists' emotions and experiences, offering new insights and effective strategies for optimizing the tourism experience. Moreover, by systematically analyzing and utilizing the data, it is possible to provide customized suggestions that reflect the preferred images of tourist destinations and user tendencies. This research provides critical insights into image management and marketing strategies for tourist destinations, helping to attract tourists' interest and create positive tourism experiences. Ultimately, the study aims to deepen the understanding of the tourism industry through the color characteristics and emotional classification of tourist destination images, making significant contributions to the enhancement of image management and marketing strategies for tourist destinations.

Keyword

Machine Learning(머신러닝), Color Analysis(색채분석), Sentiment Classification(감성분류)

요약

본 연구는 머신러닝 알고리즘을 활용하여 관광지 이미지를 색채적 특성과 감성적 범주로 분류하는 방법론에 중점을 두었다. 디지털 미디어와 스마트폰을 통해 수집된 사용자 생성 콘텐츠, 특히 인스타그램에서 수집된 관광지 이미지 데이터를 분석하여 관광지의 첫인상과 이미지가 관광자의 감성에 미치는 영향을 연구하였다. 색채 및 텍스트 데이터를 각각 머신러닝 알고리즘과 BERT 언어 모델을 사용하여 분석하였으며, 이를 통해 관광지 이미지의 색채적 특성과 감성적 메시지를 해석하였다. 관광지 이미지의 색채 분석과 감성 분류를 통해 관광자의 선택과 감성에 미치는 영향을 파악하고, 두 요인을 단편적이 아닌 상호 연관하여 고려할 때 활용도가 크다는 것을 보여주었다. 특히, 색채적 특성은 관광지 이미지 개선을 통해 관광지를 보다 매력적으로 만들 수 있는 중요한 요소이다. 이를 통해 관광지 이미지가 관광자의 감성과 경험에 큰 영향을 미칠 수 있음을 밝히고, 관광산업에 대한 새로운 시각과 관광 경험 최적화를 위한 효과적인 방안을 제공한다. 또한, 데이터를 분석하고 이를 체계적으로 활용한다면 관광자가 선호하는 관광지 이미지의 특성과 사용자의 성향을 반영한 맞춤형 제안을 할 수 있다. 본 연구는 관광지 이미지 분석을 통해 향후 관광지의 이미지 관리 및 마케팅 전략에 중요한 시사점을 제공할 수 있을 것으로 기대된다. 궁극적으로, 관광지 이미지의 색채적 특성 및 감성 분류를 통해 관광산업에 대한 이해를 심화시키고, 관광지의 이미지 관리 및 마케팅 전략을 강화하는 데 중요한 기여를 할 것이다.

목차

1. 서론

1-1. 연구의 배경

1-2. 연구의 목적 및 방법

2. 이론적 배경

2-1. 색채와 감성의 관계

2-2. 관광지 이미지의 중요성

2-3. 머신러닝의 특징 및 구조

2-4. 머신러닝과 색채분석

3. 연구 방법

3-1. 데이터 수집 범위 및 방법

3-2. 머신러닝 모델 설계

3-2-1. Overall 분석 프로세스

3-2-2. 색채 분석을 위한 K-Mean Clustering 기법

3-2-3. 감성적 특성 분류 방법(BERTopic)

4. 연구 결과

4-1. 색채적 특성 분석 결과

4-2. 감성적 분류 분석 결과

5. 결론 및 고찰

참고문헌

1. 서론

1-1. 연구의 배경

디지털 시대의 도래와 함께, 사용자 생성 콘텐츠의 소셜 미디어와 인터넷 플랫폼이 급속도로 성장하면서, 사람들은 정보를 얻고 경험을 공유하는 새로운 방법을 모색하게 되었다. 특히 관광 산업에서는 관광지 이미지가 잠재적 방문자의 인식과 선택에 중요한 영향을 미치는 것으로 나타났다. 관광지 이미지는 사진, 동영상 등 시각적 자료를 통해 소셜 미디어 플랫폼상에서 널리 공유되며, 이러한 시각적 자료의 색채는 관광지에 대한 인식과 감성을 형성하는 데 핵심적인 역할을 한다. Valentini et al. (2018)[1] 은 이미지의 색채가 사람들의 감성과 행동에 미치는 영향을 밝혀냈지만, 관광지 이미지에 대한 색채적 분석과 이를 통한 감성 분류에 관한 연구는 상대적으로 미비한 상태이다. 시각적 생동감을 끌어내기 위해 색채의 변화를 주는 것이 관광자(관광 활동을 하는 사람으로써 관광하며 다양한 의사결정을 내리는 주체)에게 더 매력적으로 다가갈 수 있는 방법으로 제시되고 있다. 그러나 현재 관광산업에서는 색채의 영향력을 조사하는 연구가 매우 제한적이며, 특히 디지털 사진에 대한 관심이 증가함에도 불구하고 온라인 사진의 영향에 관한 연구는 여전히 초기 단계에 머물러 있다. 또한 관광 분야에서 색채에 관한 연구는 대부분 관광목적지의 색상에 대한 연구로 한정되어 있으며, 색채의 다른 요소를 함께 고려한 연구 방식은 많이 다루어지지 않고 있다.

1) A. J. Elliot & M. A. Maier, Color psychology: Effects of perceiving color on psychological functioning in humans. Annual Review of Psychology, Vol.65, 2014, pp.95-120.

1-2. 연구의 목적 및 방법

본 연구의 목적은 머신러닝 기술을 활용하여 관광지 이미지의 색채적 특성을 분석하고, 이를 바탕으로 이미지가 전달하는 감성을 분류하는 것이다. 이를 통해 색채가 관광지 선택에 미치는 감성의 영향을 이해하고, 마케팅 및 브랜딩에 중요한 시사점을 제공하고자 한다. 연구 방법으로는 첫째, 대규모 관광지 이미지 데이터셋을 구축하고, 이 데이터셋을 활용하여 색채적 특성을 추출한다. 둘째, 추출된 색채적 특성을 기반으로 머신러닝 모델을 설계하고 학습시켜, 이미지가 전달하는 감성을 분류한다. 본 연구는 관광지 이미지의 시각적 요소인 색채와 관광객의 감성적 반응 사이의 관계를 보다 깊이 이해하는데 기여할 것으로 기대한다.

2. 이론적 배경

2-1. 색채와 감성의 관계

색채는 인간의 감성과 행동에 깊이 영향을 미치는 중요한 요소로, 다양한 연구에서 색채가 사람들의 정서

적 반응에 직접적인 영향을 준다는 사실이 밝혀져 왔다. 색채 심리학은 색상이 인간의 기분, 감정 상태, 심지어 신체적 반응까지도 변화시킬 수 있는 능력을 갖고 있다고 설명한다. Heller(2009)는 색채가 감정과 인지에 미치는 영향을 상세히 조사하였으며, 특정 색상이 행복, 안정, 슬픔, 불안과 같은 다양한 감정 상태를 유발할 수 있음을 지적했다[2]. 예를 들어, 파란색은 일반적으로 평온함과 안정감을 줄 수 있는 반면, 빨간색은 에너지와 위험의 감정을 유발할 수 있다. Elliot과 Maier(2014)는 색채가 인간의 인지 및 행동 과제 성능에 미치는 영향을 탐구하였다[3]. 이들은 특히 빨간색이 경쟁적 상황에서 성능을 저하시키는 반면, 파란색은 창의적 사고를 촉진시킬 수 있다는 결론을 내렸다. 또한, 색채는 마케팅과 브랜딩에서도 중요한 역할을 한다. Gorn et al.(2004)은 소비자의 구매 의사 결정 과정에서 색채가 미치는 영향을 조사하였고, 색상이 브랜드 인식과 제품 선호도에 유의미한 영향을 준다는 것을 발견했다[4]. 이러한 연구들은 색채가 단순히 시각적 요소를 넘어서 인간의 감성과 행동에 깊이 관여한다는 것을 시사한다. 따라서, 관광지 이미지에 있어서 색채의 역할을 탐구하는 것은 관광객의 정서적 반응을 이해하고, 이를 통해 관광지의 매력을 향상시킬 수 있는 효과적인 방법을 제시할 수 있는 중요한 접근 방법이 될 수 있다.

2-2. 관광지 이미지의 중요성

관광지 이미지란 개인이 거주지 이외의 지역에 대하여 가지는 인상 및 잠재적인 방문객들의 인식이다[5].관광지 이미지는 잠재적 방문자의 관광지 선택과정에 중대한 영향을 미치는 요소로 광범위하게 인정받고 있다. 관광지 이미지는 단순히 해당 장소의 시각적 표현을 넘어서, 방문 전 예상되는 경험과 감성을 형성하는 중요한 역할을 한다. 이는 관광객의 기대치 설정, 목적지 선택, 만족도 평가 등 다양한 결정에 직접적으로 영향을 끼친다. Echtner와 Ritchie(1993)에 따르면, 관광지 이미지는 크게 기능적 속성(예: 관광지의 접근성, 편의 시설 등)과 심리적 속성(예: 문화, 분위기 등)으로 구분될 수 있다[6]. 이러한 분류는 관광지 이미지가 관광객의 인식과 태도 형성에 어떻게 기여하는지 이해하는 데 중요한 기반이 된다. 또한, Baloglu와 McCleary(1999)는 관광지 이미지가 개인의 지각, 신념, 감정과 밀접하게 연결되어 있으며, 이러한 요소들이 관광지에 대한 전반적인 태도와 만족도를 형성하는 데 기여한다고 주장한다[7]. 이는 관광지 이미지가 단순히 관광지의 물리적 특성을 반영하는 것을 넘어서, 관광객의 심리적 반응과 연결되어 있음을 시사한다. 최근 연구들은 소셜 미디어와 디지털 플랫폼의 성장으로 인해 관광지 이미지의 전달과 소비 방식이 변화하고 있음을 보여준다. Kim과 Law(2015)는 온라인 리뷰와 소셜 미디어 콘텐츠가 관광지 이미지 형성에 점점 더 중요한 역할을 하고 있으며, 이러한 디지털 콘텐츠가 관광객의 예상과 만족도에 영향을 미친다고 지적한다[8]. 이러한 연구들은 관광지 이미지가 관광객의 인식, 태도, 행동에 미치는 영향을 다각도로 조명하며, 관광지 마케팅과 관리 전략 개발에 있어 관광지 이미지의 중요성을 강조한다. 따라서, 관광지 이미지의 효과적인 관리와 마케팅 전략 개발은 관광지의 경쟁력을 강화하고, 관광객의 만족도와 재방문 의사를 높이는 데 핵심적인 요소이다.

2-3. 머신러닝의 특징 및 구조

머신러닝은 데이터간 패턴과 관계를 학습하고 이를 바탕으로 유의미한 정보를 추출하여 문제를 해결한다.

---

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

3) A. J. Elliot, & M. A. Maier, Color psychology: Effects of perceiving color on psychological functioning in humans. Annual Review of Psychology, Vol.65, 2014, pp.95-120.

4) G. J. Gorn, A. Chattopadhyay, J. Sengupta, & S. Tripathi, Waiting for the web: How screen color affects time perception. Journal of Marketing Research, Vol.41, No.2, 2004, pp.215-225.

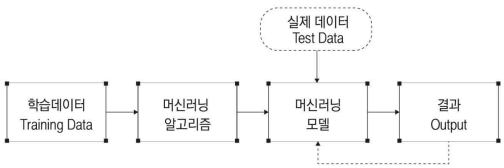
5) J. D. Hunt, Image as a Factor in Tourism Development. *Journal of Travel Research*, Vol.13, No.3, 1975, pp.1-7.

6) C. M. Echtner, & J. R. B. Ritchie, The Measurement of Destination Image: An Empirical Assessment. Journal of Travel Research, Vol.31, No.4, 1993, pp.3-13.

7) S. Baloglu, & K. W. McCleary, A model of destination image formation. Annals of Tourism Research, Vol.26, No.4, 1999, pp.868-897.

8) D. Y. Kim, & R. Law, Smartphones in tourism and hospitality marketing: A literature review, Journal of Travel & Tourism Marketing, Vol.32, No.6, 2015, pp.692-711.

Mohammed et al.(2017)와 Hastie et al.(2009)에 따르면 4가지 과정을 가진다[9]. 첫째, 데이터 전처리 단계에서 원시 데이터로부터 유용한 특징을 추출하고 변환하여 모델의 성능을 향상시킨다 이 과정에는 새로운 특징 생성, 누락된 값 처리, 범주형 데이터의 수치 변환, 데이터의 정규화 및 스케일링, 그리고 차원 축소 등이 포함된다. 둘째, 모델 구축 단계에서는 세 가지 주요 학습 방식인 비지도 학습, 지도 학습, 강화 학습을 통해 다양한 데이터와 적용할 문제에 맞게 모델을 설계한다. 비지도 학습은 레이블 없는 데이터에서 패턴을 학습하고, 지도 학습은 레이블을 사용하여 데이터 내의 규칙과 패턴을 학습한다. 강화학습은 에이전트가 환경과 상호작용하며 최적의 행동을 학습한다. 셋째, 학습 및 평가 단계에서는 모델 아키텍처를 설계하고, 하이퍼파라미터를 설정하여 학습 과정을 제어한다. 학습된 모델은 테스트 데이터로 성능을 검증하며, 필요에 따라 모델의 매개변수를 조정하여 최적의 성능을 달성하기 위해 반복적인 성능 개선 작업을 수행한다. 마지막으로 모델 튜닝 및 개선 단계를 거친다.[10]. [그림 1]은 머신러닝에 대한 기본 구조도를 보여준다. 피드백 루프를 기반으로 예측한 결과와 실제 결과와의 차이를 통해 성능을 향상시키고, 학습한 데이터에만 국한되지 않고 추론을 통해 새로운 데이터에서도 좋은 결과를 내는 모델을 개발하는 것이 중요하다. 이는 머신러닝이 기존 룰 기반(rule-based) 시스템과 차별화되는 중요한 특징이다.



[그림 1] 머신러닝 기본 구조도

2-4. 머신러닝과 색채 분석

본 연구는 대용량 SNS 관광지 이미지 색채 데이터를 분석하기 위해 개인의 주관적인 시각보다 좀 더 객관적이고 시스템적인 과정을 활용하고자 한다. 일반적으로 사용되고 있는 육안에 의한 색채분석은 기계학습을 통한 색채분석 방법에 비해 시간이 많이 소요되며, 단편적으로 색채만 분석하기 때문에 동시에 다른 다양한 정보를 분석하는 것에 한계가 있다. 또한, 이러한 분석은 색채 전문가에 한정되어 이루어지는 경우가 많아, 일반인이 쉽게 접근하거나 이해하기 어렵다는 문제가 있다. 이에 본 연구는 머신러닝을 활용한 색채분석을 통해 관광지 이미지 및 text에서 추출된 색채 데이터와 텍스트 데이터를 활용하여, 객체나 패턴 인식 관련 다양한 분석을 가능하게 한다. 또한 색채 데이터 관련 데이터베이스의 구축으로 일관성, 정확성, 신뢰성을 확보한다. 대량의 이미지 데이터를 활용하여 통계학적 분석을 통해 패턴을 찾아낼 수 있고 관광지 이미지에 대한 색채적 요인을 빠르게 분석하고 결과를 도출할 수 있다. 또한 색채 데이터베이스의 구축은 기존 데이터베이스에 새로운 데이터를 추가하거나 다른 머신러닝 모델의 학습 데이터로 활용할 수 있어, 본 연구에서 매우 중요한 방법적인 과정이다. 이를 통해 다양한 디자인 분야에서 객관적인 색채 추출방식이 가능해진다[11].

3. 연구 방법

3-1. 데이터 수집 범위 및 방법

본 연구에서는 인스타그램을 통해 유럽의 대표 관광지인 영국(런던), 프랑스(파리), 이탈리아(로마), 스페인(바르셀로나)의 4개국의 주요 관광지 이미지 50,000개의 이미지 파일과 54,675개의 게시글, 해시태그, 댓글 기반 텍스트 파일을 수집한다. 데이터 수집은 Meta에서 제공하는 Instagram API[12]와 데이터 크롤링을 이용한 Instaloader 파이썬 라이브러리[13]를 활용하였다. 수집된 데이터는 대부분 분석에 직접 활용할 수 없기 때문에, 데이터 전처리 과정을 통해 분석 가능한 형태로 정제하여 학습 데이터의 품질을 향상시키는 작업이

9) M. Mohammed, M. B. Khan, & E. B. M. Bashier, Machine learning: Algorithms and Applications. Florida: CRC Press, 2017.

10) T. Hastie, R. Tibshirani, & J. Friedman, The Elements of Statistical Learning. New York: Springer, 2009.

11) 김미숙. 관광지 이미지에 대한 색채적 · 정서적 요인에 관한 연구: 머신러닝 방법론을 활용한 색채 분석을 중심으로. 홍익대학교 대학원 박사학위논문. 2024. pp.32-52.

12) Meta, https://developers.facebook.com/docs/, 2024.

13) A. Grünewald, https://instaloader.github.io/, 2024.03.18.

수행된다. 관광지 이미지의 색채적 특성을 분석하고 Text 데이터를 기반으로 분석된 이미지별 감성을 분류한다. 이 과정에서 이미지와 텍스트 데이터를 포함하여 색채를 분석하는 것이 중요하므로, 데이터 전처리는 이미지의 색채 정보와 텍스트 내용을 모두 고려되었다. 이는 데이터의 신뢰성을 높이고 모델의 성능을 향상시키는데 핵심적인 역할을 한다. 본 연구의 데이터 전처리 접근법은 이미지와 텍스트의 복합적인 데이터 특성을 반영하여 더욱 효과적인 분석 결과를 도출하는데 기여한다.

[표 1] 데이터 수집 범위

국가	도시	범위
영국	런던	빅벤, 타워브릿지, 버킹검궁전, 웨스트민스터사원, 런던아이, 대중교통수단
프랑스	파리	에펠탑, 루브르박물관, 노트르담 대성당, 상젤리제 거리, 몽마르트(사크레 쾨르 대성당)
이탈리아	로마	콜로세움, 트레비분수, 바티칸 시티, 판테온, 스페인 계단
스페인	바르셀로나	사그라다 파밀리아 대성당, 구엘공원, 카사 밀라, 카사 바트요, 보케리아 시장

본 연구에서는 사용자 생성 콘텐츠인 인스타그램에서 수집된 이미지들을 머신러닝 모델의 요구에 맞추어 400x500 픽셀로 정규화하고, 연구 목적과 관련 없는 광고, 독사진, 회의 사진 등을 필터링하여 총 12,800개의 적합한 이미지 파일을 확보하였다.

[그림 2] 데이터 이미지 수집 (영국 런던)

[그림 3] 데이터 이미지 수집 (프랑스 파리)

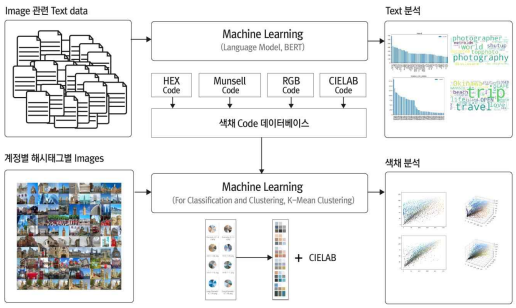
[그림 4] 데이터 이미지 수집 (이탈리아 로마)

[그림 5] 데이터 이미지 수집 (스페인 바르셀로나)

3-2. 머신러닝 모델 설계

3-2-1. Overall 분석 프로세스

본 연구에서는 크게 관광지 이미지와 관련 Text를 분석할 수 있도록 두 가지 머신러닝 기법을 사용한다. 먼저, 머신러닝 모델을 개발하기 위해 비지도 학습에서 클러스터링 기법 중 하나인 K-Mean Clustering을 사용한다. 이 기법은 이미지 내 유사한 특성을 가진 색채 데이터들끼리 클러스터로 나누어 색채 빈도와 트렌드 등을 분석할 수 있다. 두 번째로 관광지 이미지와 함께 수집된 Text 데이터를 분석하기 위해 BERTopic을 사용하였다. BERTopic은 BERT(Bidirectional Encoder Representations from Transformers) 모델을 기반으로 문장 내 단어들의 의미를 이해하고 임베딩 벡터를 생성한다. 생성된 임베딩 벡터는 문장 간의 유사성을 계산하고, 군집화하여 분석하고자 하는 주요 단어들을 추출하는데 사용된다. 그리하여 머신러닝 기반 색채 분석 방법론을 적용하여 관광지 이미지에 대한 색채적 요인과 해당 이미지 관련 Text에 대한 주요 감성 형용사 추출에 대한 분석 시스템을 제안하고자 한다. 인스타그램 내 관광지 이미지와 관련 Text 데이터를 수집하고자 인스타그램에서 제공하는 API 및 관련 라이브러리를 활용하여 Python을 통해 해당 데이터들을 추출하였다. 다음 [그림 6]은 본 연구에서 제시하는 전반적인 머신러닝 방법론을 통한 분석 프로세스이다.

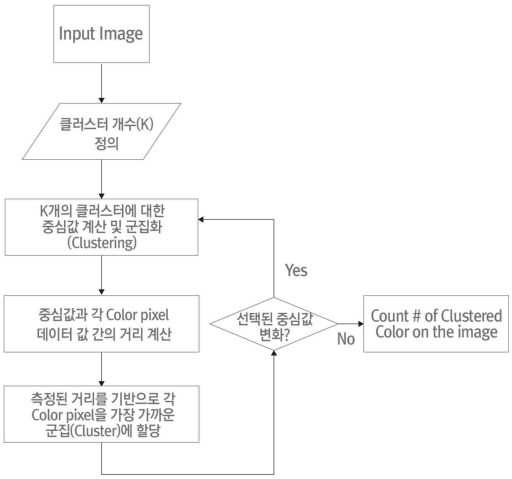


[그림 6] 색채/감성적 요인 분석 Overall 프로세스

3-2-2. 색채 분석을 위한 K-Mean Clustering 기법

본 연구에서는 주어진 이미지 내에서 색채 데이터의 패턴을 탐색하고 숨겨진 의미들을 발견하기 위해 비지도 학습 관련 머신러닝 기법인 K-Mean Clustering을 활용하였다. 이 방법은 데이터를 비슷한 특성을 가진 클러스터(Cluster)로 나누는 기술로, 이러한 클러스터는 데이터 포인트 간의 유사성을 기반으로 형성된다. 이 기법은 계산 효율성이 높아 대량의 이미지 색채를 효과적이고 빠르게 분석할 수 있다.

이 알고리즘은 결과 해석의 용이성이라는 장점이 있으며, 각 데이터 포인트를 하나의 클러스터에 할당하기 때문에 클러스터링 결과의 시각화 및 해석이 비교적 쉽다.



[그림 7] K-Mean Clustering 기법

[그림 7]과 같이 K-Mean Clustering의 프로세스는 다음과 같다.

1) 초기화(Initialization): K개의 클러스터 중심을 무작위로 선택하거나 랜덤하게 배정한다.

2) 할당(Assignment): 각 데이터 포인트를 가장 가까운 클러스터 중심점에 할당한다. 이 과정에서 유클리드 거리 측정 방법을 사용하여 각 데이터 포인트와 중심점 간의 거리를 계산한다.

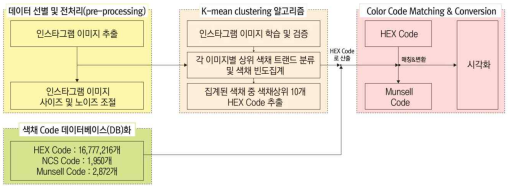
3) 업데이트(Update): 클러스터의 중심점을 업데이트하여, 각 클러스터에 할당된 데이터 포인트들의 평균을 구해 새로운 중심점을 결정한다.

4) 반복(Iteration): 할당과 업데이트 단계를 반복하며, 중심점이 더 이상 크게 변하지 않거나 설정된 중지 기준을 만족할 때까지 과정을 반복한다.

5) 결과(Result): 알고리즘이 수렴하면, 즉 중심점이 더 이상 크게 변하지 않을 때, 데이터 포인트는 K개의 클러스터로 분류되어 최종 클러스터링 결과를 제공한다.

다음 [그림 8]과 같이 K-Mean Clustering을 통해 색채 분석 결과로 산출된 HEX Code를 Munsell Code로 변환하고, 이를 기반으로 색채 비중을 시각화하였다. 이를 위해 Color.org(2020)에서 개발한

API(Application Programming Interface)를 통해 HEX Code(16,777,216개)와 Munsell Code(2,872개)에 대한 색채 코드 데이터를 확보하여 색채 데이터베이스를 구축하였다.

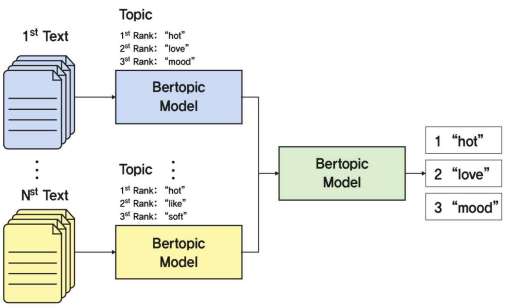


[그림 8] Munsell Code 산출 프로세스

3-2-3. 감성적 특성 분류 방법(BERTopic)

본 연구에서는 사용자 생성 콘텐츠의 대표 소셜미디어인 인스타그램에서 추출한 Text 데이터를 분석하기 위해 BERTopic을 사용하였다. BERTopic은 텍스트의 수치화 및 클러스커링을 통해 주제를 도출하고, 이를 통해 텍스트의 의미와 구조를 파악하여 주요 단어 간의 관계를 규정할 수 있다.

[그림 9]는 Bertopic기반 주요 단어 추출 과정을 나타낸다. 먼저 인스타그램에서 수집 및 정제된 Text 데이터를 삽입한 후, 문서의 내용을 벡터로 변환시켜 Text 데이터의 의미와 구조를 파악한다. 이를 기반으로 TF-IDF(Term Frequency-Inverse Document Frequency)를 통해 Text 내 단어들의 중요도를 분석한다. TF-IDF는 특정 단어가 특정 문서에 나타나는 횟수, 전체 문서에 특정 단어가 나타나는 횟수, 그리고 전체 문서 수에 대한 특정 단어의 문서 빈도의 역수(IDF)를 기반으로 계산된다. 다음으로 Topic Representation을 통해 빈도 및 중요도를 기반으로 각 주요 단어들에 순위(rank)를 매긴다. 마지막으로 Text 데이터별 주요 단어(Topic) 순위를 구한 후 결합시켜 최종 주요단어별 순위를 산출한다.



[그림 9] BERTopic기반 Text별 주요 단어 추출 프로세스

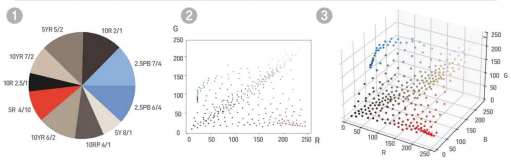
4. 연구 결과

관광지 이미지에서 픽셀 단위의 색채 데이터를 K-Mean Clustering 방법을 통해 분석하고, 관광지의 색채를 Munsell, RGB, L*, a*, b*, C*, HEX Code로 변환하여 데이터베이스화했다. 유럽 대표 관광지에서 추출된 색채 데이터는 그래프와 표를 통해 시각적으로 나타냈으며, L*, a*, b*, C* 값들을 기반으로 한 분석을 수행했다. 이 데이터를 사용하여 관광지들 간의 색채 차이를 분석하였고, 주로 쓰이는 감성 형용사의 빈도를 분석하였다. 이를 통해 관광목적지의 특성 및 정서적 요인과 어떤 연관이 있는지를 확인했다.

4-1. 색채적 특성 분석 결과

K-mean clustering을 통해 관광지별 색채 특성을 분석하였다. 연구 대상지로 선정된 관광지인 영국 런던, 프랑스 파리, 이탈리아 로마, 스페인 바로셀로나의 색채 데이터 분석 결과는 [그림 10]에서 [그림 13]까지 나타내고 있다. 각각의 그림①과 같이 K-Mean clustering을 통해 HEX Code에서 변환되어 추출된 빈도 높은 상위 10개 Munsell Code의 색채 데이터를 기준으로 각각의 그림②, 그림③와 같이 관광지별 색채 분포도를 통해 관광지의 색채 경향을 확인할 수 있었다.

도시	색채	MUNSELL	RGB	CIELAB(L*a*b*)	C*	HEX
런던		5YR 5/2	133/119/108	51.097/3.329/6.518	7.319	#85776c
		10R 2/1	61/51/51	22.346/3.697/1.398	3.952	#3d3333
		2.5PB 7/4	146/178/217	71.22/-3.439/-18.835	19.146	#92b2d9
		2.5PB 6/4	111/146/196	59.425/-2.001/-24.154	24.237	#6f92c4
		10YR 6/2	166/151/137	63.496/3.172/7.615	8.249	#a69789
		10YR 7/2	197/181/163	74.679/2.962/9.106	9.576	#c5b5a3
		10RP 4/1	92/83/82	36.164/2.938/1.599	3.345	#5c5352
		5R 4/10	223/60/42	51.608/50.145/39.551	63.866	#df3c2a
		5Y 8/1	225/216/205	86.84/1.442/5.279	5.472	#e1d8cd
		10R 2.5/1	25/24/23	8.333/0.258/0.711	0.756	#191817

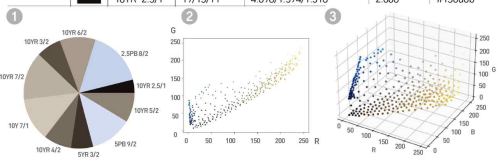


[그림 10] 영국 런던 색채분석

영국 런던의 색채 데이터 L*값의 범위(L*=8~87)은 저명도에서 고명도의 색까지 넓게 분포하고 있는 것으로 확인되었다. 색상 측면에서는 주로 a*값 범위(a*=-3~51)가 플러스 방향 빨간색 영역으로 집중되어 있으며, b*값은 마이너스 방향의 파란색 영역에서 플러스 방향의 노란색 영역으로 고르게 분포(b*=-24~40)하고 있는 것으로 확인되었다. C*값은 다양한 범위(C*=0~64)에 분포하고 있으나, 80%이

상 저채도에 분포되어 있는 것으로 나타났다.

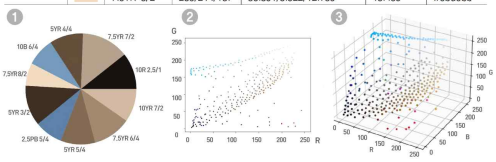
도시	색채	MUNSELL	RGB	CIELAB(L*a*b*)	C*	HEX
파리		10YR 7/2	182/171/155	70.587/1.657.814/	7.986	#b6ab9b
		10Y 7/1	202/190/176	77.69/2.092/6.954	7.262	#cabeb0
		2.5PB 8/2	191/203/228	81.351/-0.347/-11.126	11.131	#bfcbe4
		5PB 9/2	206/216/237	86.053/-0.328/-9.247	9.253	#ced8ed
		10YR 6/2	161/149/133	62.421/2.038/8.135	8.386	#a19585
		10YR 5/2	136/124/110	52.809/2.346/7.541	7.897	#887c6e
		10YR 4/2	111/99/85	42.812/2.498/7.862	8.249	#6f6355
		5YR 3/2	57/47/38	20.305/2.763/6.066	6.666	#392f26
		10YR 3/2	85/73/61	31.989/2.918/7.342	7.901	#55493d
		10YR 2.5/1	19/13/11	4.098/1.594/1.318	2.068	#130d0b



[그림 11] 프랑스 파리 색채분석

프랑스 파리의 색채 데이터 L*값의 범위(L*=4~87)가 넓게 분포하고 있으나, 이 중 70% 이상이 중명도에서 고명도 값에 집중되어 분포하고 있다. 색상 측면에서는 주로 a*값 범위(a*=-0.3~3)가 플러스 방향 빨간색 영역으로 집중되어 있으며, b*값은 마이너스 방향의 파란색 영역에서 플러스 방향의 노란색 영역으로 고르게 분포(b*=-9~9)하고 있는 것으로 확인되었다. C*값은 다양한 범위(C*=2~12)에 분포하고 있으나, 90%이상 저채도에 분포되어 있는 것으로 나타났다.

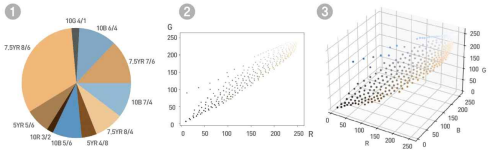
도시	색채	MUNSELL	RGB	CIELAB(L*a*b*)	C*	HEX
로마		7.5YR 7/2	177/157/135	66.076/3.997/11.511	12.185	#b19d87
		7.5YR 6/4	159/132/108	56.824/5.041/13.142	14.076	#9b846c
		10YR 7/2	204/182/158	75.478/4.274/12.235	12.960	#ccb69e
		5YR 3/2	60/46/33	20.267/4.018/8.771	9.648	#3c2e21
		5YR 5/4	124/102/80	44.936/5.259/12.758	13.799	#7c6650
		5YR 4/4	95/75/56	33.626/5.244/11.778	12.893	#5f4b38
		10R 2.5/1	24/18/16	6.036/1.97/1.66	2.576	#181210
		10B 6/4	113/150/183	60.256/-5.462/-17.292	18.134	#7196b7
		2.5PB 5/4	92/126/165	51.404/-3.46/-20	20.297	#5c7ea5
		7.5YR 8/2	235/214/187	86.851/3.522/12.935	13.406	#ebd6bb



[그림 12] 이탈리아 로마 색채분석

이탈리아 로마의 색채 데이터 L*값의 범위(L*=6~87)은 저명도에서 고명도의 색까지 넓게 분포하고 있는 것으로 확인되었다. 색상 측면에서는 주로 a*값 범위(a*=-5~6)가 플러스 방향 빨간색 영역으로 집중되어 있으며, b*값은 플러스 방향의 노란색 영역으로 넓게 분포(b*=-20~14)하고 있는 것으로 확인되었다. C*값은 다양한 범위(C*=2~21)에 분포하고 있으나, 90%이상 저채도에 분포되어 있는 것으로 나타났다.

도시	색채	MUNSELL	RGB	CIELAB(L*a*b*)	C*	HEX
바르셀로나		7.5YR 8/6	235/179/118	77.22/12.634/31.43	33.874	#ebb376
		7.5YR 7/6	205/155/101	67.925/11.65/28.626	30.936	#cd9b65
		7.5YR 8/4	230/194/145	80.737/6.645/23.778	24.689	#e6c291
		10B 7/4	144/179/204	70.996/-6.309/-13.464	14.869	#90b3cc
		10B 6/4	126/163/190	64.978/-6.486/-14.639	16.012	#7ea3be
		10B 5/6	88/143/186	56.918/-7.545/-22.838	24.052	#588fba
		5YR 5/6	162/124/81	55.059/8.986/23.494	25.154	#a27c51
		5YR 4/8	125/78/33	38.051/13.843/27.406	30.704	#7d4e21
		10G 4/1	84/88/84	36.941/-1.834/1.397	2.305	#545854
		10R 3/2	65/35/11	17.466/10.408/16.907	19.854	#41230b

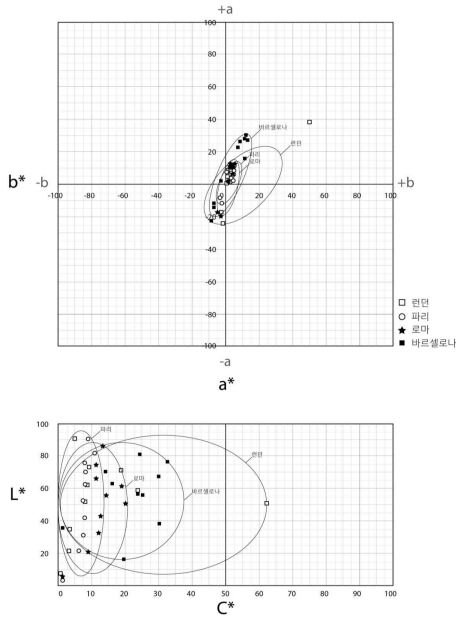


[그림 13] 스페인 바르셀로나 색채분석

스페인 바르셀로나의 색채 데이터 L*값의 범위(L*=17~81)은 저명도에서 고명도의 색까지 넓게 분포하고 있는 것으로 확인되었다. 색상 측면에서는 주로 a*값 범위(a*=-7~13)가 마이너스 방향의 초록색 영역부터 플러스 방향의 빨간색 영역까지 넓게 분포되어 있으며, b*값은 플러스 방향의 노란색 영역으로 넓게 분포(b*=-22~32)하고 있는 것으로 확인되었다. C*값은 다양한 범위(C*=2~34)에 분포하고 있으나, 70%이상 중채도에 분포되어 있는 것으로 나타났다.

다음 [그림 14]와 같이, 관광지의 색채 데이터에서 L*값은 다양한 밝기를 나타내며 저명도부터 고명도까지 폭넓게 활용되고 있었다. a*값의 색채 데이터는 대부분 플러스 방향의 빨강(+a*) 계열로 넓게 분포하고 있었다. b*값의 색채 데이터는 일부 파랑(-b*) 계열도 포함하고 있었지만, 주로 플러스 방향의 노랑(+b*) 계열로 넓게 분포되어 있었다. C*값의 색채 데이터는 주로 저채도에 집중되어 있었다. 이를 통해 관광지 이미지에서 비교적 연한 색채가 많이 사용되거나 색의 선명함이 부족하다는 점을 알 수 있다. 이러한 현상은 도시의 인공물이 주변 환경에 미치는 영향으로 해석될 수 있다. 인공물들이 주변의 색채에 영향을 주어 자연적인 색채보다 저채도의 색채가 더 많이 관찰되는 것을 의미한다.

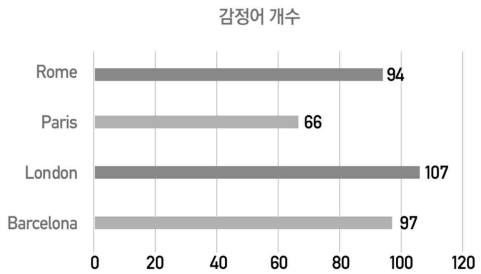
이는 관광지별 관광자들이 관광지 이미지를 통해 느끼는 감성과 연결되어 관광지 선택과 만족도에 상당한 영향을 미치는 것을 시사한다.



[그림 14] 관광지 이미지의 색채 데이터 범위

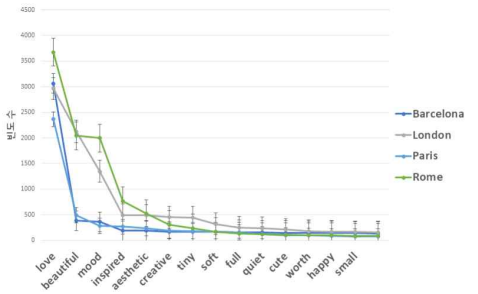
4-2. 감성적 분류 분석 결과

본 연구는 BERTopic을 사용해 이미지별 Text 데이터로부터 주요 감성 형용사들을 추출하여 분석하였다. 먼저, [그림 15]는 관광지별 BERTopic으로부터 추출된 감성어 개수를 말한다. London에서 가장 다양한 감성 형용사가 쓰였다는 것을 알 수 있고, Paris에서 상대적으로 적게 단어가 쓰였다는 것을 알 수 있다.



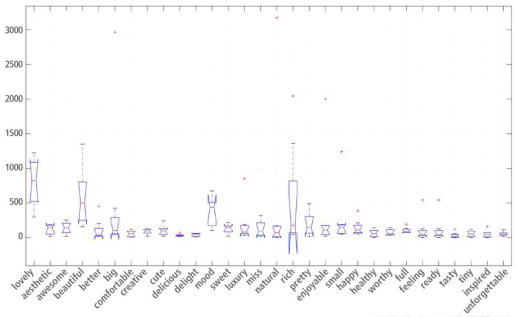
[그림 15] BERTopic 기반 관광지별 감성어 횟수 결과

다음 [그림 16]은 관광지별 상위 14개의 감성 형용사 단어를 기반으로 빈도를 추출하여 시각화한 것이다. 가장 많이 사용된 단어는 “lovely”, “beautiful”, “mood”와 “inspired“가 상대적으로 많이 쓰였고, 그 외 ‘aesthetic’, ‘creative’ 등의 감성과 관련 있는 형용사들이 상위권에 위치하고 있었다.



[그림 16] 관광지별 주요 감성 단어(상위 14개)

다음 [그림 17]과 같이 형용사 ‘lovely’, ‘beautiful’, ‘mood’, ‘rich’의 변동 폭이 크다는 것을 확인할 수 있다. 이는 각 관광지별 사용의 편차가 크다는 것을 의미한다.



[그림 17] 관광지별 주요 감성 단어별 boxplot

이는 관광지에서 사용되는 감성 형용사들의 빈도는 감성적 경험의 공유를 통해 관광지별 감성적 연결고리를 형성하여 관광자들에게 특정 감정과 경험을 암시하고 주된 감성을 표현하는데 중요한 역할을 한다는 것으로 해석된다. 이러한 결과는 관광지 이미지의 표현된 감성이 관광자들의 인식에 영향을 미침을 시사한다.

5. 결론 및 고찰

본 연구는 머신러닝을 활용하여 관광지 이미지를 색채적 특성과 감성적 범주로 분류하는 방법론을 개발하는 데 중점을 두었다. 다음과 같은 구체적인 방법

론을 제안한다.

1) 데이터 수집 및 전처리

대표적인 소셜 미디어인 인스타그램을 통해 유럽 주요 관광지인 영국(런던), 프랑스(파리), 이탈리아(로마), 스페인(바르셀로나)로부터 50,000개의 관광지 이미지와 54,675개의 게시글, 해시태그, 댓글 기반 텍스트를 수집하였다.

2) 색채적 특성 분석

디지털 미디어와 스마트폰을 통해 수집된 사용자 생성 콘텐츠, 특히 인스타그램에서 수집된 관광지 이미지 데이터를 분석하여 관광지의 첫인상을 K-Mean clustering을 통해 분석하였다. 4개 도시에 대해 분석 결과, 명도(L*) 값이 넓게 분포하며, 대체로 중명도에서 고명도에 집중되어 있다. 색상(a*) 값은 대부분 빨간색 계열로 넓게 분포하고, 색상(b*) 값은 주로 노란색 계열로 넓게 분포하는 반면, 일부 파란색 계열도 포함하고 있었다. 채도(C*) 값은 다양한 범위에 걸쳐 분포하나, 대부분 저채도에 집중되어 있어 관광지 이미지에서 연한 색채가 많이 사용되고 색의 선명함이 부족한 특징을 보였다. 이는 도시의 인공물이 주변 환경에 미치는 영향으로, 자연적인 색채보다 저채도의 색채가 더 많이 관찰되는 결과로 해석될 수 있다.

3) 감성적 특성 분류

본 연구는 BERTopic을 사용해 이미지별 Text 데이터로부터 주요 감성 형용사들을 추출하여 분석하였다. 분석 결과, 런던에서 가장 다양한 감성 형용사가 사용된 반면, 파리에서 상대적으로 적은 단어가 사용되었다. 상위 14개의 감성 형용사 중 “lovely”, “beautiful”, “mood”, “inspired”가 가장 많이 사용되었고, “aesthetic”, creative“등의 단어들도 상위권에 위치했다. 특히, 형용사 "lovely", "beautiful", "mood", "rich"는 변동 폭이 커서 각 관광지별 사용의 편차가 큼을 알 수 있다. 이는 관광지에서 사용되는 감성 형용사들이 감성적 경험의 공유를 통해 관광지별 감성적 연결고리를 형성하고, 특정 감정과 경험을 암시하며 주된 감성을 표현하는 데 중요한 역할을 한다는 것을 의미한다. 이러한 결과는 관광지 이미지의 표현된 감성이 관광자들의 인식에 영향을 미침을 시사한다.

그리하여 본 연구는 관광지 이미지의 색채 분석과 감성 분류를 통해 관광자의 선택과 감성에 미치는 영향을 파악하고, 두 요인을 단편적이 아닌 상호 연관하여 고려할 때 활용도가 크다는 것을 보여준다. 특히 색채적 특성은 관광지 이미지 개선을 통해 관광지를 보다 매력적으로 만들 수 있는 중요한 요소이다. 또한, 데이터를 분석하고 이를 체계적으로 활용함으로써 관광자가 선호하는 관광지 이미지의 특성과 사용자의 성향을 반영한 맞춤형 제안을 할 수 있는 핵심 방법론이 될 수 있을 것으로 본다. 이러한 연구 결과는 관광자들이 원하는 경험과 감정에 맞는 이미지를 제시하여 특정 관광지로 유도하는 방안으로 활용될 수 있다. 본 연구는 관광지 이미지의 색채적 특성 및 감성 분류를 통해 관광산업에 대한 이해를 심화시키고, 관광지의 이미지 관리 및 마케팅 전략에 중요한 시사점을 제공할 것으로 기대된다.

참고문헌

1. A. J. Elliot & M. A. Maier, Color psychology: Effects of perceiving color on psychological functioning in humans. Annual Review of Psychology, Vol.65, 2014, pp.95-120.
2. C. Valentini, S. Romenti, G. Murtarelli, & M. Pizzetti, Digital visual engagement: Influencing purchase intentions on Instagram. Journal of Communication Management, Vol.22, No.4, 2018, pp.362-381.
3. A. J. Elliot, & M. A. Maier, Color psychology: Effects of perceiving color on psychological functioning in humans. Annual Review of Psychology, Vol.65, 2014, pp.95-120.
4. G. J. Gorn, A. Chattopadhyay, J. Sengupta, & S. Tripathi, Waiting for the web: How screen color affects time perception. Journal of Marketing Research, Vol.41, No.2, 2004, pp.215-225.
5. J. D. Hunt, Image as a Factor in Tourism Development. Journal of Travel Research, Vol.13, No.3, 1975, pp.1-7.

6. C. M. Echtner, & J. R. B. Ritchie, The Measurement of Destination Image: An Empirical Assessment. Journal of Travel Research, Vol.31, No.4, 1993, pp.3-13.

7. S. Baloglu, & K. W. McCleary, A model of destination image formation. Annals of Tourism Research, Vol.26, No.4, 1999, pp.868-897.

8. D. Y. Kim, & R. Law, Smartphones in tourism and hospitality marketing: A literature review, Journal of Travel & Tourism Marketing, Vol.32, No.6, 2015, pp.692-711.

9. M. Mohammed, M. B. Khan, & E. B. M. Bashier, Machine learning: Algorithms and Applications. Florida: CRC Press, 2017.

10. T. Hastie, R. Tibshirani, & J. Friedman, The Elements of Statistical Learning. New York: Springer, 2009.

11. 김미숙. 관광지 이미지에 대한 색채적 · 정서적 요인에 관한 연구: 머신러닝 방법론을 활용한 색채 분석을 중심으로. 홍익대학교 대학원 박사학위논문. 2024. pp.32-52.

12. Meta, https://developers.facebook.com/docs/, 2024.

13. A. Grünewald, https://instaloader.github.io/, 2024.03.18.