

텍스트 마이닝을 통한 관광명소 유형별 동서양 관광객들의 관광 만족도 요인에 대한 비교 연구

A Text Mining Approach to Identifying Destination Attractiveness Factors
Influencing Customer Satisfaction in the Tourism Industry:
A Cross-Cultural Comparison between East Asian and Western Reviewers

INDEX

1. 연구 배경
2. 연구 목적
3. 연구 방법
4. 연구 결과
5. 결론
6. 참고문헌

연구 배경

- 국제관광산업의 성장
- 관광 산업에서 온라인 리뷰 데이터 증가와 그에 따른 여행 리뷰 분석의 중요성 대두
 - 관광 산업에서 관광객들의 온라인 리뷰 사용의 증가
 - 데이터 수집의 접근성, 속도와 단순성, 인간 피험자에 대한 비침입성을 포함하여 많은 이점 존재 (Lu & Stepchenkova, 2015)
- 목적지의 중요한 요소인 '명소' 관련 리뷰 데이터 분석 연구 부족
 - 기존의 관광 연구들은 호텔, 레스토랑 리뷰 데이터를 이용하여 머신러닝 기법 적용



따라서, 본 연구에서는 리뷰데이터를 유저의 **문화적 배경**에 따라 분리하고 여행 목적지 **명소 유형별로**
여행 리뷰어의 만족도 요소를 텍스트 마이닝 기술을 통해 **비교 분석**하려고 함



1.4 billion

Total international tourist arrivals



USD 1.7 trillion

Total international tourism exports
(International tourism receipts + passenger transport)
(UNWTO, 2019)

연구 목적

본 연구는 관광객들이 관광지를 선택하는데 중요하다고 생각하는 관광지 특성이 문화적 영향(서양과 동양)과 관광지 명소 유형에 따라 어떻게 다르게 나타나는지 관광객 리뷰 데이터를 통해 알아보려고 함.

연구 문제

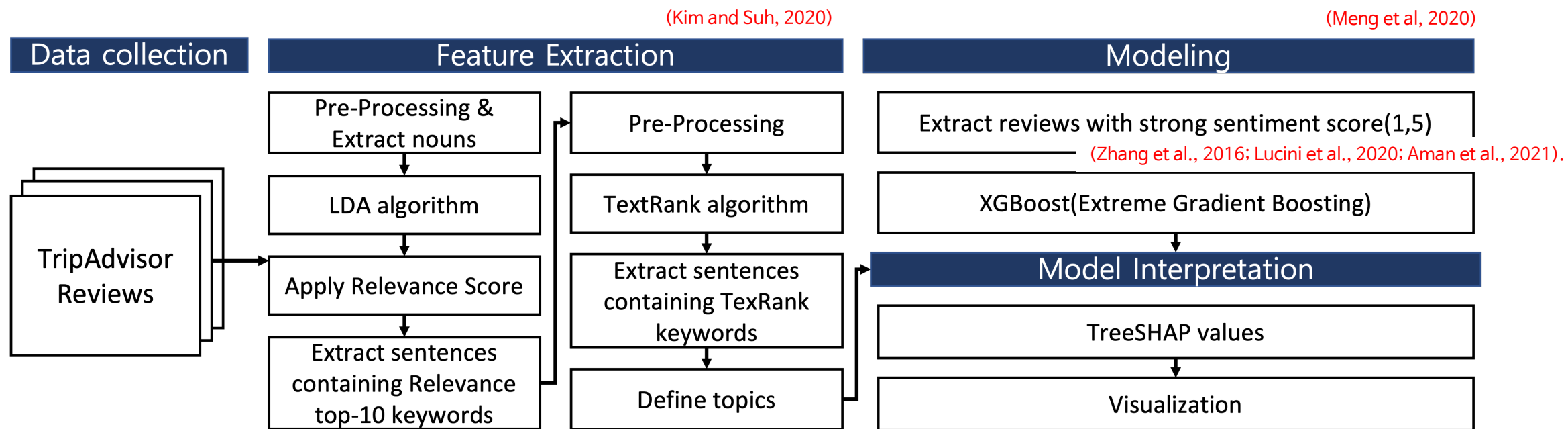
연구 문제1. 온라인 여행 리뷰에 언급된 주요 관광지 특성들은 무엇인가?

연구 문제2. 발견된 주요 관광지 특성에 대해 명소 유형별(쾌락적 장소, 실용적 장소) 그리고 동서양 관광객 사이에 차이가 존재하는가?

연구 문제3. 관광지 특성과 여행객의 특징(동양/쾌락적 장소, 동양/실용적 장소, 서양/쾌락적 장소, 서양/실용적 장소) 사이에 어떤 연관성이 존재하는가?

연구 문제4. 어떤 관광지 특성이 각 명소 유형별 동서양 관광객의 관광 만족도에 중요 결정 요인으로 작용하였나?

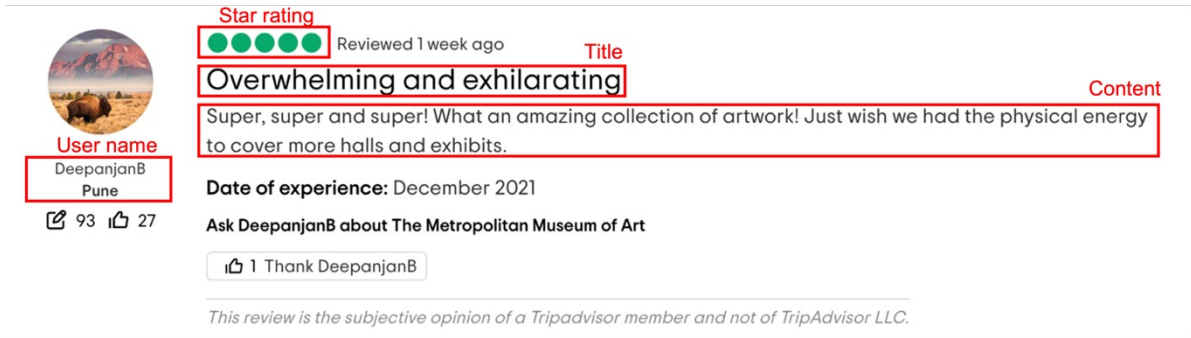
연구 프레임워크



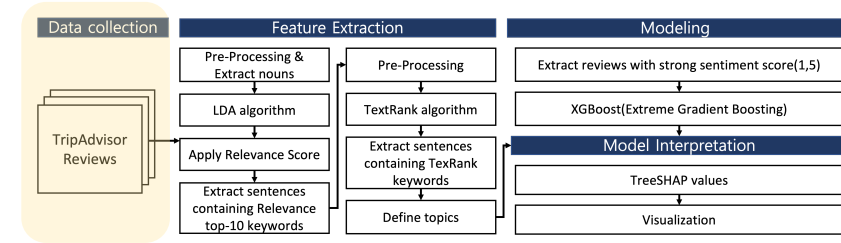
- 1) 카이제곱 검정
- 2) CA분석

데이터 수집

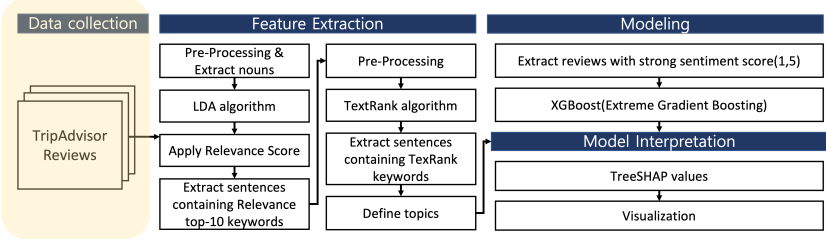
- 웹사이트: TripAdvisor.com
- 분석 툴: Python 3.8.2 version
- 번역기: Google Translate API version 4.0.0
- 기간: 2021년 12월 31일까지의 모든 리뷰 데이터



1) 유저 이름 2) 유저 국가 3) 별점 4) 리뷰제목 5) 리뷰 내용



- 목적지 : 미국 (UNWTO, 2020)
 - 대표적인 국제적인 나라
 - 동시에 2019년 다양한 문화권의 관광객들이 방문하는 여행지 Top3에 선정
 - (+상대적으로 다른 나라의 리뷰에 비해 아시아인들의 리뷰 내용이 많아 동서양 비교 분석을 하기위한 충분한 데이터 확보가 가능했음)
- 동서양 선택 (Hofstede, 1991; Reisinger et al., 2009; Kort, 2005)
 - 동양: mainland China, Hong Kong, Macao, Mongolia, South Korea, Taiwan, and Japan
 - 서양: US, Canada, Australia, UK
- 명소 유형 (Hlee et al., 2021; Byun and Jang, 2015)
 - 쾌락적장소: 아름다운 경관과 관련된 장소 ex.해변, 캐니언, 국립공원
 - 실용적 장소: 역사 박물관, 미술관



[수집된 데이터]

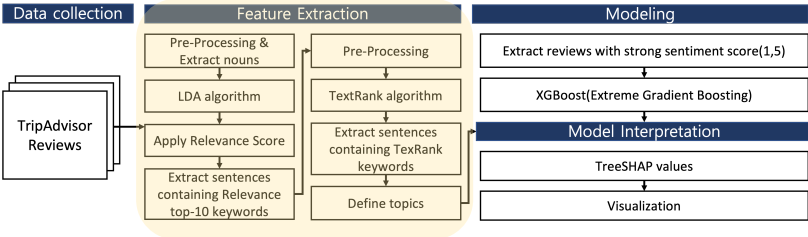
변수	동양	서양	총 리뷰수
명소 유형			
실용적 장소	7,779	7,780	15,559
쾌락적 장소	7,845	7,845	15,690
			31,249
Rating			
1-star	73	130	203
2-star	199	261	460
3-star	1,676	1,071	2,747
4-star	5,034	3,385	8,419
5-star	8,642	10,778	19,420
			31,249

[전처리 과정]

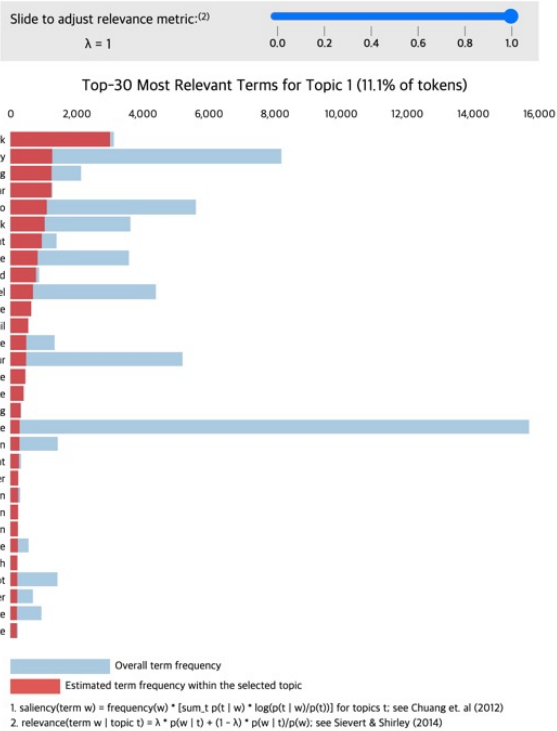
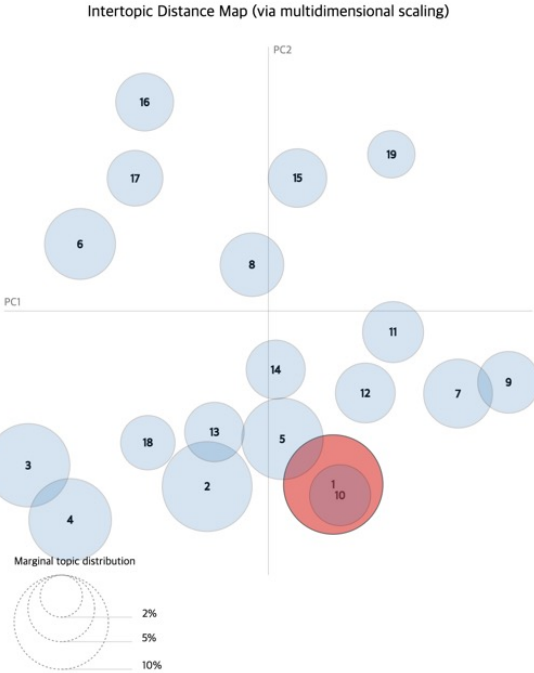
- Kumar et al.(2020)에서 사용한 전처리 과정 참고
 - 1) 텍스트 데이터를 **토큰화**한 후 각각 2개, 3개의 단어를 묶는 bigram, trigram을 사용하여 구문 단어를 생성
 - 2) 품사 태깅(POS-tagging)을 사용하여 품사를 각 단어에 적용하고 명사 추출
여행 리뷰어들에게 중요하다고 인식되는 주요 특징만을 추출하기 위해 **‘명사’** 만 추출 (Lucini et al., 2020)
 - 3) 리뷰에 등장하는 모든 대문자는 모두 **소문자로 치환**
 - 4) 모든 **관사(‘a/an’, ‘the’)**를 포함하는 모든 불용어는 gensim에서 제공하는 stopwords기능을 이용하여 제거
 - 5) **인칭대명사** (man, men, guy, boy, woman, women, girl, lady etc.), **목적지를 나타내는 명사들**(waikiki, hawaii, honolulu, new_york, san_diego, america, Washington_dc, metropolitan etc.) 제거
 - 6) 이후 **길이가 1인 단어**, 언급 횟수가 **2%미만인** 단어는 의미 없는 단어로 판단하여 제거
 - 7) 단어에 **표제화(lemmatization)** 적용하여 단어의 의미가 변하지 않고 본래의 의미를 유지하도록 함.
 - 8) 단어의 **정규화** (dollar→ money, mile/km/m-→ distance, min→minute etc..) 비슷한 의미의 단어가 중복되거나 생략되지 않도록 함.

관광객이 언급한 선호 관광지 특성 탐지

Kim and Suh, 2020



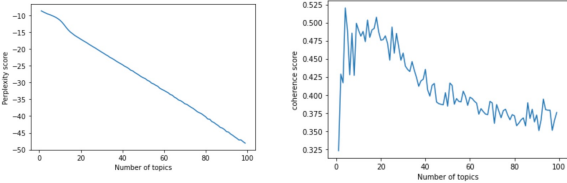
Selected Topic:



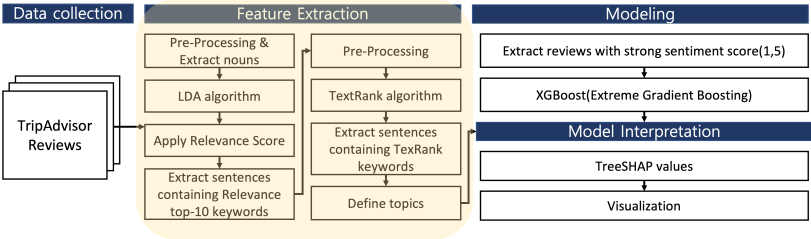
Topic 1 예시>

Rank	LDA keywords	Relevance keywords	TextRank keywords	
			words	weight
1	park	park	(parking, lot)	0.22893
2	scenery	car	(car, park)	0.12674
3	parking	parking	(walk,)	0.12236
4	car	point	(point,)	0.07481
5	photo	road	(scenery,)	0.06319
6	walk	drive	(road,)	0.06307
7	point	trail	(time,)	0.05977
8	minute	website	(beach,)	0.05408
9	road	walk	(minute,)	0.05317
10	travel	hike	(drive,)	0.05234

Perplexity, coherence score 반영하여 총 18개의 토픽 추출



관광지 선호 특성에 관련된 연구



Factor	Description	Reference
Packages	pre-arranged packages by intermediaries and principals	(Cooper et al., 1998; Buhalis, 2000)
Accessibility	entire transportation system comprising of routes, terminals, and vehicles	(Ritchie and Zins, 1978; Cooper et al., 1998; Buhalis, 2000; Reitsamer et al., 2016)
Amenities/ Infrastructure	accommodation and catering facilities, retailing, other tourist services	(Cooper et al., 1998; Buhalis, 2000; Reitsamer et al., 2016)
Activities	all activities available at the destination and what consumers will do during their visit	(Cooper et al., 1998; Buhalis, 2000)
Scenery	the natural form and landscape of a destination	(Ritchie and Zins, 1978; Milman and Abraham, 1995; Reitsamer et al., 2016)
Climate	amount of sunshine, temperature, winds, precipitation	(Ritchie and Zins, 1978; Hu et al, 1993; Milman and Abraham, 1995)
Safety	level of personal and material safety	(Ritchie and Zins, 1978; Milman and Abraham, 1995)
Local community/ Attitudes towards tourists	natural and experiential resources in a destination and refers to a connection with local people & the warmth of reception by the local population, ease of communication, willingness to provide information, and a lack of hostility towards tourism activities.	(Ritchie and Zins, 1978; Kim et al., 2012; Reitsamer et al., 2016)
Price level	the value received for money spent on major services, food, lodging and transportation within the region	(Ritchie and Zins, 1978; Milman and Abraham, 1995)
Season	seasonal preference of a destination	(Hong-bumm, 1998)
Family oriented	suitability for families with children	(Milman and Abraham, 1995; Hong-bumm, 1998)
Information/ Credibility	destination-related details and images (a factor that gives credibility)	(Hong-bumm, 1998)
Festivals/special events	festivals/ special events information of a destination	(Shafiee et al., 2021)

라벨링 결과

TextRank keywords	Example	Label	prior study
(parking, lot), (car, park), (walk, road),	...15 minutes or more from the parking lot... ...Walk 10 to 15 minutes from the parking lot (sneakers are desirable because they walk in the sand)	Parking	/
(tour, (guide,), (site,), (lot, information), (group,))	...There are tours and bus tours going by plane, but bus tours are tough if they cannot afford time... ...Well worth the tour and very well organised... ...They have guided tours in english... ...Purchase the tour tickets, then obtain tour group time & explorer around before the meeting time...	Packages	Y
(bus, (city,), (ride,) (shuttle,), (stop,), (access,))	...The free bus is always moving, so you can easily move to a superb view site... ...As the hotel was close from the starting bus stop, we were able to ride a lot, but this bus route is popular and can be cared immediately, so it could no longer get ride from the bus stop in the center of Waikiki...	Accessibility	Y
(food, drink), (price, (service,), (bike,), (town, (quality,), (ticket,), (minute,), (money,), (price, admission), (advance,))	...Rental services equipments like snorkels \$7-15 and lockers \$7 are available... ...Prices are also high... ...Food is expensive and if you're not camping think a Cabin cost more than... ...The price was also reasonable... ...Weather was fine and sunshine was very strong... ...I saw the sunset, but the weather was not good and I could not say the best, but it was still very beautiful... ...If you are bad at the weather, please be careful about clothing etc...	Price Level	Y
(morning, (night,), (sunset,), (time,), (sun,), (weather,), (evening,), (lot, people))	...There were many people because it was a summer vacation period, but I should go absolutely... ...Summer is super busy... ...I went twice, but it is a good place to go many times, but it is painful for summer...	Weather	Y
(summer, (time, period), (ticket, month), (architecture, (winter,), (design,), (phone, (spring,), (museum,))	...There are also supermarkets that can be accommodated and shopping... ...Second stop: Looking for a restaurant to eat, although the road (in the wrong road) came to the parking lot of the town, there is a wooden house, there is a network, but in the end, it is still on the restaurant, and eats a Chinese meal...	Tourist Season	Y
(hotel, (shopping,), (beach,), (distance,), (restaurant, (street,), (shop,))		Facilities	Y

(morning, (night,), (sunset,), (time,), (sun,), (weather,), (evening,), (lot, people))

(summer, (time, period), (ticket, month), (architecture, (winter,), (design,), (phone, (spring,), (museum,))

(hotel, (shopping,), (beach,), (distance,), (restaurant, (street,), (shop,))

(security, check), (story, (camera,), (hand,), (bag, (event,), (luggage,))

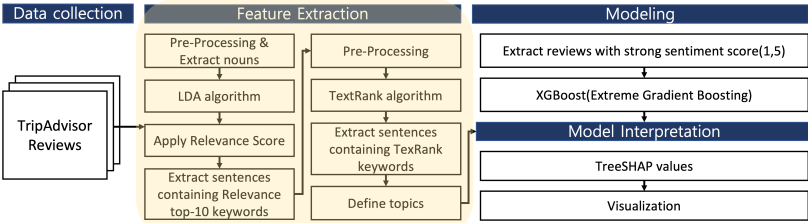
(lot, fish), (snorkel, (watch, video), (swim,), (water,), (place,), (purchase, ticket), (sea,), (variety,), (location,), (lot, activity), (variety, fish), (image,), (art, museum), (age), (interest,), (staff, member))

(photo, (point,), (tour, guide), (drive,))

(scenery, (travel,), (foot,), (history,))

(family, (time, year), (child,), (number, people), (adult,))

(tourist, (spot,), (movie,), (crowd,))



...weater was fine ana sunsune was very strong...
...I saw the sunset, but the weather was not good and I could not say the best, but it was still very beautiful...
...If you are bad at the weather, please be careful about clothing etc...
...There were many people because it was a summer vacation period, but I should go absolutely...
...Summer is super busy...
...I went twice, but it is a good place to go many times, but it is painful for summer...
...There are also supermarkets that can be accommodated and shopping...
...Second stop: Looking for a restaurant to eat, although the road (in the wrong road) came to the parking lot of the town, there is a wooden house, there is a network, but in the end, it is still on the restaurant, and eats a Chinese meal...
...It costs admission fee, but it is safe to put the luggage and enter the sea alone...
...The valuables can be deposited in the locker with a paid, so I was able to swim with confidence even if I was an individual...
...I can swim with tropical fish and you can swim with tropical fish without going to a deep place...
...You can perform a snorkeling and other water activities!...
...Learn about activities and entertainment in advance, and you will have a good time...
...I love outdoor activity, so I strongly recommend everyone to stop by the Red Rock Canyon...
...This is a bucket list activity...
...I was crazy and taking a lot of photos...
...We stopped at all the roadside parking lots to enjoy the scenery and take photos...
...Great art and beautiful place our favorite part was finding the artwork featured in "Ferris Bueller's Day Off" and taking a few photos...
...The scenery spread in front of you is too magnificent, and it is as much as you lose words...
...However it was an interesting historical tour and we enjoyed the scenery very much, which is beautiful...
...The courtyard is serene and many people sat along the outskirts as they took a rest and appreciated the scenery...
...There are many Junior Ranger Programs that children can learn about the Earth...
...The whole family went together...
...Thanks to the parasol, the scenery and family photo shoots were also enjoyable...
...I personally, would never bring young children...
...Stunning you've seen it in the movies, so you already know what to expect, and yet, it is absolutely stunning to watch...
...This is one of those places that you've seen many times in pictures, on tv, on the internet, but you still go WOW when you see it with your own eyes...

Weather

Y

Tourist Season

Y

Facilities

Y

Safety/Security

Y

Activities

Y

Camera
(Taking pictures)

Y

Scenery

Y

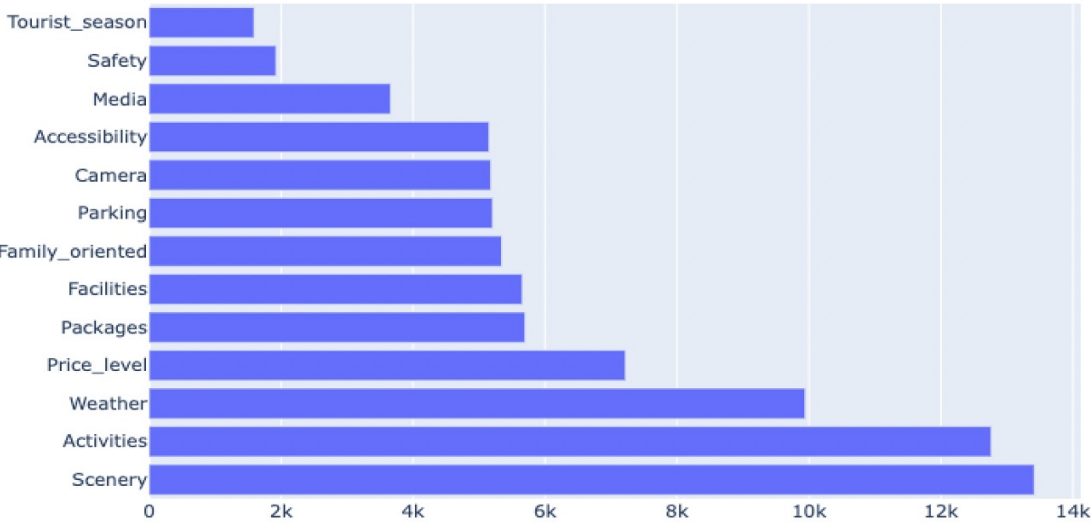
Family Oriented

Y

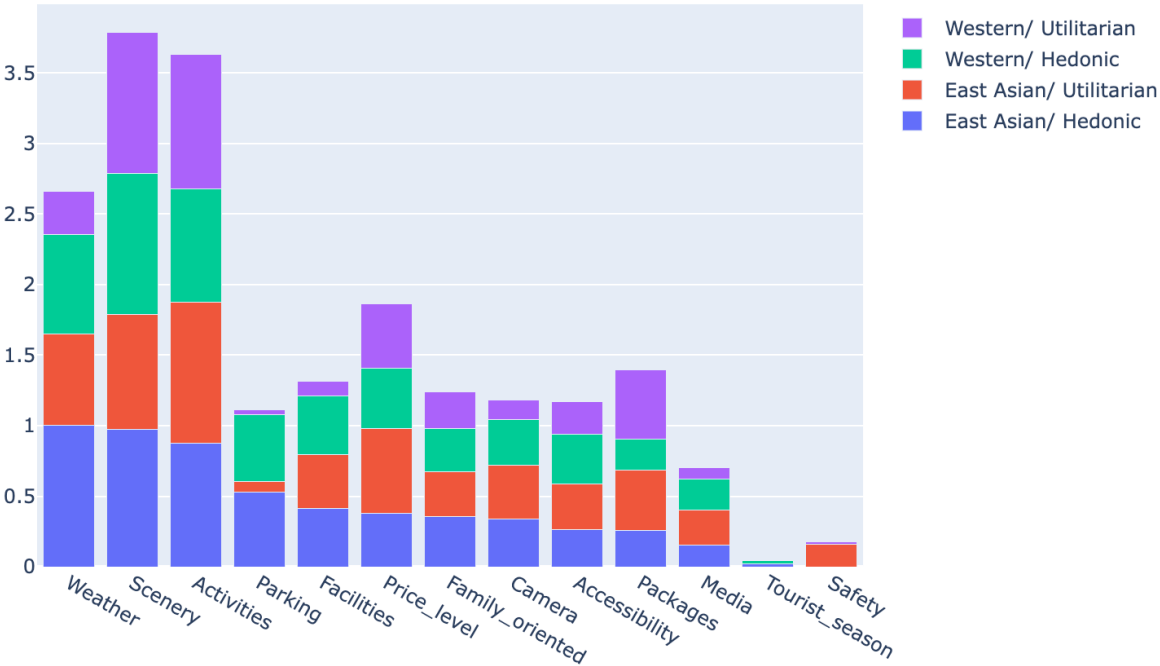
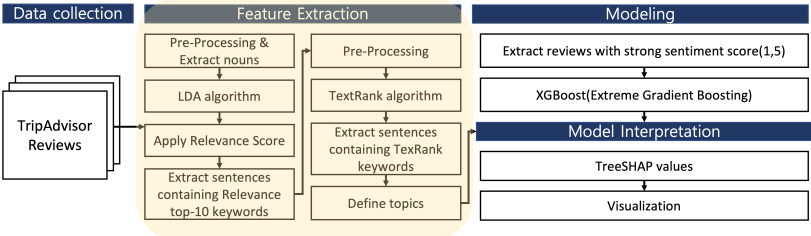
Media

/

명소 유형별 동서양 관광객이 언급한 특성 분포



[Figure 5. Distribution of total features]



- 서양/실용적 장소: Packages, Family oriented, Accessibility
- 서양/쾌락적 장소: Parking, Facilities, Accessibility
- 동양/실용적 장소: Packages, Camera, Facilities
- 동양/쾌락적 장소: Parking, Facilities, Price level

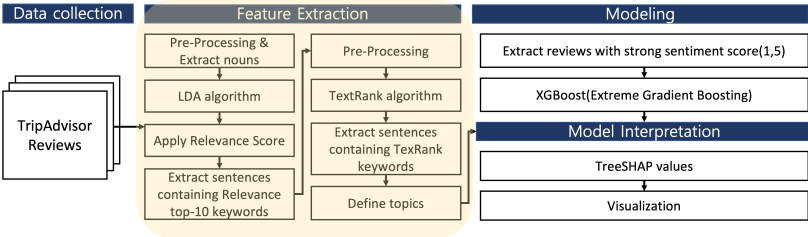
명소 유형별 동서양 관광객이 언급한 특성 분포 차이 & 연관성

[카이제곱 결과표]

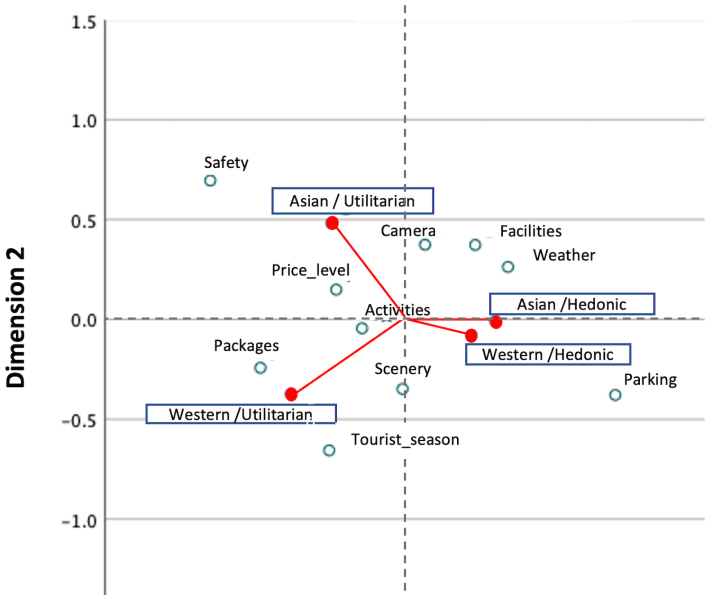
Features	Culture			Attraction type		
	χ^2	df	p value	χ^2	df	p value
Parking	4.669	1	.031*	1798.331	1	0**
Packages	46.817	1	0**	413.053	1	0**
Accessibility	50.111	1	0**	6.366	1	.012**
Price level	11.159	1	.001**	77.420	1	0**
Weather	274.904	1	0**	1054.777	1	0**
Tourist season	42.589	1	0**	8.635	1	.004**
Facilities	21.872	1	0**	277.829	1	0**
Safety/Security	16.571	1	0**	316.817	1	0**
Activities	22.219	1	0**	19.709	1	0**
Camera	23.64	1	0**	48.931	1	0**
Scenery	292.945	1	0**	145.156	1	0**
Family oriented	1.1811	1	.181	13.713	1	0**
Media	1.826	1	.178	0.009	1	.929

* $p < .05$, ** $p < .01$

[Table 6. Result of Chi-square Test]



[대응분석(CA) 결과표]



쾌락적 장소에서 Facilities, Weather, Parking, Camera에 대한 특성이 높으며, 동양인과 서양인 사이의 관련성이 높게 나타남.
실용적 장소에서 Safety, Price level, Packages, Tourist season에 언급 많이 보이며, 동양인과 서양인 사이의 관련성이 약하게 나타남.

명소 유형별 동서양 관광객의 만족도에 영향을 미치는 결정 요인 발견

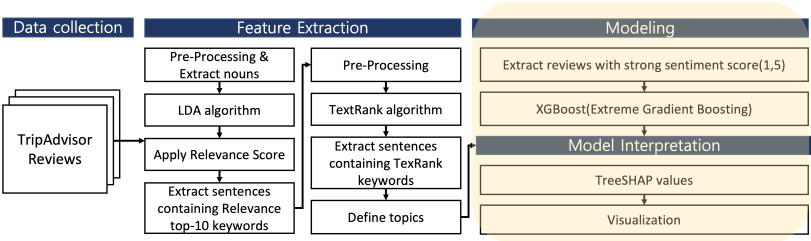
- 독립변수 : 토픽 distribution
- 종속변수: 별점 (0,1)
(Kolomojets and Dickinger, 2023)

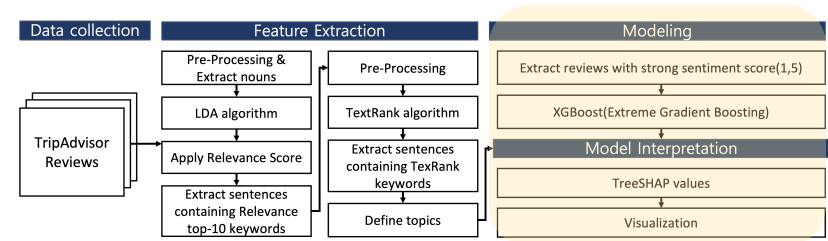
- 데이터 셋 분리 후 모델 적용
 - 1) 동양 / 쾌락적 장소
 - 2) 동양/ 실용적 장소
 - 3) 서양/ 쾌락적 장소
 - 4) 서양/ 실용적 장소

[XGBoost 모델 검증 결과표]

- 긍정, 부정 데이터 셋의 차이가 커 긍정 데이터 셋이 맞춰 오버샘플링 진행
- grid search로 파라미터 조정 (논문에 파라미터 제시)

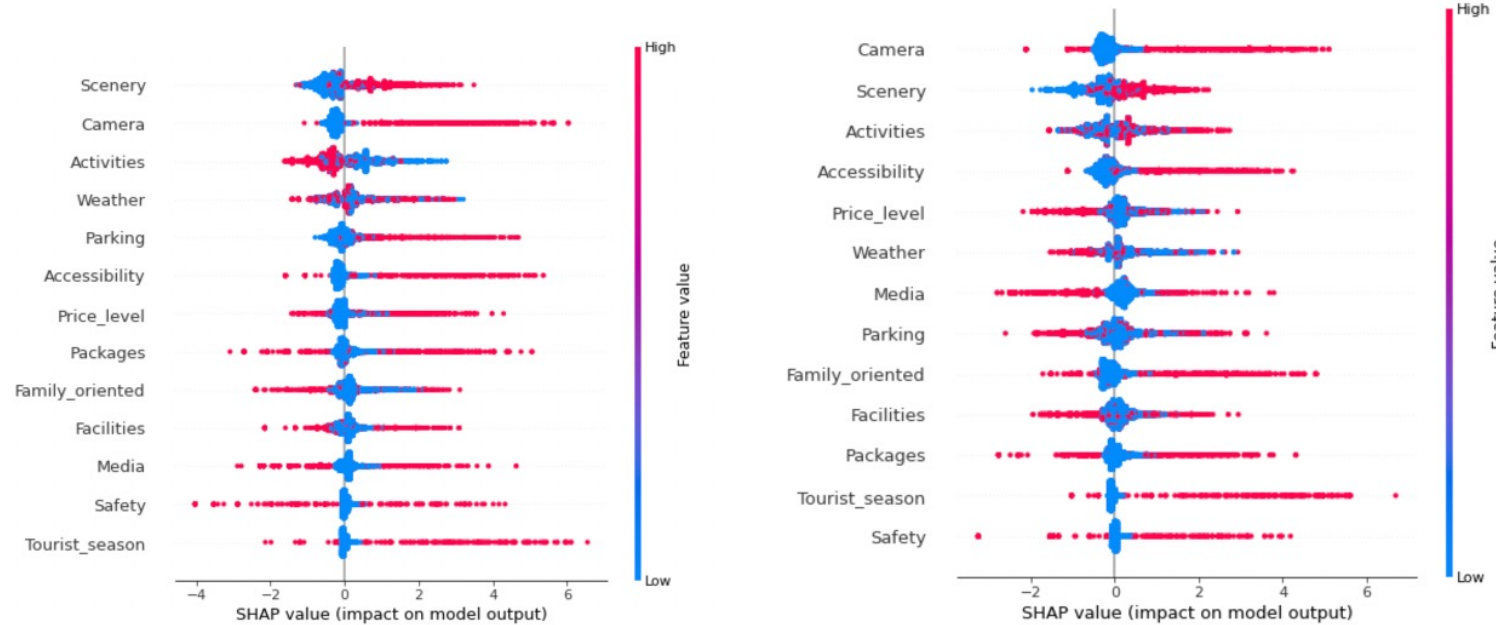
데이터 셋	F1	Precision	recall	accuracy
동양/쾌락적 장소	78.68%	86.41%	72.22%	80.53%
동양/실용적 장소	80.22%	84.12%	76.66%	81.07%
서양/쾌락적 장소	80.73%	89.85%	73.29%	80.73%
서양/실용적 장소	85.09%	84.20%	86.01%	85.09%





명소 유형별 동서양 관광객의 만족도에 영향을 미치는 결정 요인 발견

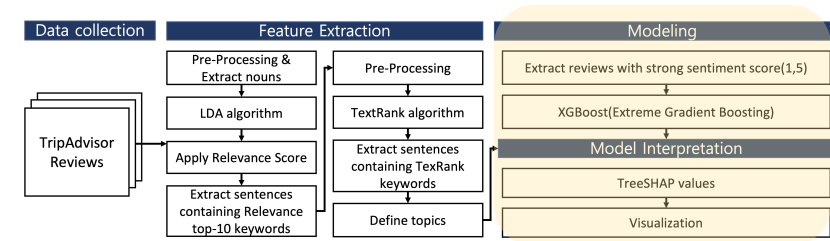
[TreeSHAP value]



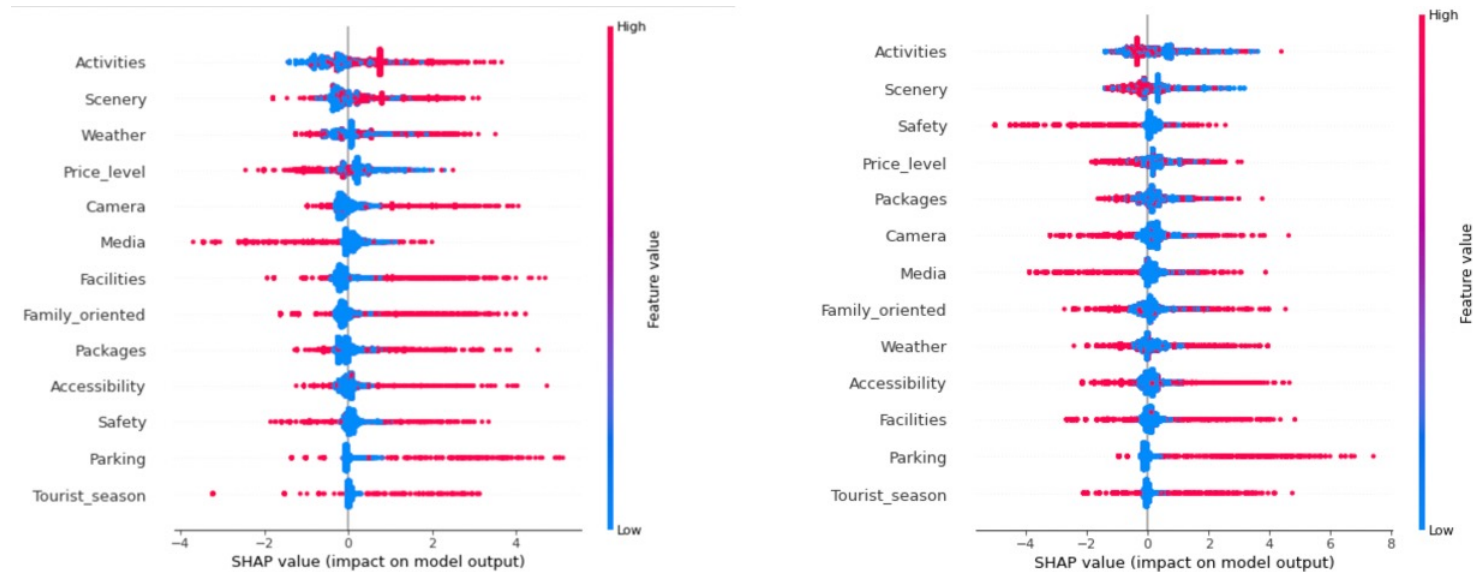
[Figure 8. SHAP feature contribution in the Hedonic place (East Asian (left) vs. Western(right))]

- 쾌락적 장소에서 **공통적**으로 Camera, Activities, Scenery의 영향력이 크게 나타남.
- 상대적으로 **동양인의 만족도**에 영향력이 큰 특성: Weather, Packages
- 상대적으로 **서양인의 만족도**에 영향력이 큰 특성: Camera, Accessibility, Price level, Media
 - 동양인들은 Activity, Safety에서 부정적인 감정이 나타남.
 - 서양인의 경우 Parking에서 부정적인 감정이 나타남.

명소 유형별 동서양 관광객의 만족도에 영향을 미치는 결정 요인 발견



[TreeSHAP value]



[Figure 9. SHAP feature contribution in the Utilitarian place (East Asian (left) vs. Western(right))]

- 실용적 장소에서는 **공통적**으로 Activities의 영향력이 크게 나타남.
- 또한 동양인과 서양인의 만족도에 영향을 미치는 특성이 비슷하게 나타남.
- 상대적으로 **서양인**들에게 Safety에 대한 영향력이 크게 나타남.
 - 동양인의 경우 Price level에서 부정적인 감정이 나타남.
 - 서양인의 경우 Activities, Scenery, Price level에서 부정적인 감정이 나타남.

시사점

- 관광 관련 매니저 또는 마케터가 자신의 지역을 홍보할 경우 지역의 특성과 고객 타겟에 맞는 적절한 홍보 대책을 세우는데 도움을 줌.
- 빈도를 고려한 분석 결과와 감정을 고려한 분석한 결과를 동시에 제시함으로써 두개의 수치를 동시에 고려한 비즈니스 전략 매트릭스로 활용이 가능
- 이전 연구에서 잘 사용하지 않았던 ‘명소’ 리뷰 데이터를 이용함으로써 매력적인 관광지 속성에 대한 새로운 아이디어 제시 및 관광지 속성 이해 확장 가능.
- 최신 리뷰 데이터를 이용한 목적지에 대한 문화적 차이 분석과 동시에 관광지 명소 유형을 분석한 연구는 드물었다는 점에서 의의가 있음.

한계점

- 웹사이트 특성상 국적을 하나만 선택할 수 있기 때문에 이중 국적을 가진 사람들을 탐지할 수 없어 리뷰어의 문화적 배경을 정확하게 분리 불가능
- 서양인의 경우 목적지인 미국은 자국의 고향일 수도 또는 거리상 가깝다고 할 수 있음. 하지만 동양인의 경우 목적지인 미국은 거리가 상당히 있는 타국이므로 이러한 특징으로 인해 관광객들 사이에 차이가 존재할 가능성 있음
- 서양인의 경우 목적지인 미국은 같은 언어를 쓰지만, 동양인의 경우 다른 언어를 씀. 따라서 이로 인한 관광객들 사이에 차이가 존재할 가능성 있음
- 리뷰의 특성으로 여러 문장으로 이루어진 리뷰 하나에 별점 하나를 부여하는 방식임. 따라서 부정적인 감정과 긍정적인 감정이 섞여 있는 리뷰의 경우 타겟 감정으로 정확한 감정을 부여할 수 없었음

따라서 이후의 연구에서는 이러한 점을 보완하여 연구 진행 가능함

Moscardo, G. (2004). East versus West: a useful distinction or misleading myth. *Tourism (Zagreb)*, 52(1), 7-20.

Kang, M., & Moscardo, G. (2006). Exploring cross-cultural differences in attitudes towards responsible tourist behaviour: A comparison of Korean, British and Australian tourists. *Asia Pacific Journal of Tourism Research*, 11(4), 303-320.

Kozak, M. (2001). Repeaters' behavior at two distinct destinations. *Annals of tourism research*, 28(3), 784-807.

Ekinci, Y., & Hosany, S. (2006). Destination personality: An application of brand personality to tourism destinations. *Journal of travel research*, 45(2), 127-139.

Hudson, S., & Ritchie, J. B. (2009). Branding a memorable destination experience. The case of 'Brand Canada'. *International Journal of Tourism Research*, 11(2), 217-228.

Hlee, S., Lee, H., Koo, C., & Chung, N. (2021). Will the relevance of review language and destination attractions be helpful? A data-driven approach. *Journal of Vacation Marketing*, 27(1), 61-81.

Byun J and Jang SS (2015) Effective destination advertising: matching effect between advertising language and destination type. *Tourism Management* 50: 31-40.

Lu, W., & Stepchenkova, S. (2015). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. *Journal of Hospitality Marketing & Management*, 24(2), 119-154.

Hofstede G. 1991. *Cultures and Organizations: Software of the Mind*. McGraw-Hill: London.

REISINGER, Yvette; MAVONDO, Felix T.; CROTTS, John C. The importance of destination attributes: Western and Asian visitors. *Anatolia*, 2009, 20.1: 236-253.

KIM, Eunhoe; SUH, Yuhwa. A Method of Calculating Topic Keywords for Topic Labeling. *Journal of the Korea Society of Digital Industry and Information Management*, 2020, 16.3: 25-36.

Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63-70).

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, 75, 550-568.

감사합니다😊