한국정보전자통신기술학회논문지(jkiiect)'20-12, Vol.13 No.6
논문 20-13-06-489 http://dx.doi.org/10.17661/jkiiect.2020.13.6.489

토핑 모델링을 활용한 동해안 관광의 변화 분석

정은희*

The Analysis of Changes in East Coast Tourism using Topic Modeling

Eun-Hee Jeong*

요 약 4차혁명이 진행되고 있는 초연결사회에선 다양한 IT기기를 통해 데이터량이 증가하고 있고, 이렇게 생산된 데이터를 분석하여 새로운 가치를 창출 할 수 있다. 본 연구에서는 빅카인즈에서 2017년부터 2019년까지 중앙지, 경제지, 지역조합지, 주요방송사 등에서 "(동해안 관광 또는 동해안 여행) 그리고 강원도"라는 키워드로 기사를 총 1,526건을 수집하였다. 수집된 1,526건의 기사를 분석하기 위해 R언어로 구현된 LDA 알고리즘을 이용하여 토픽 모델링을 수행하였다. 2017년부터 2019년까지 각각의 년도별 키워드를 추출하고, 각 년도별로 빈도수가 높은 키워드를 분류하여 비교하였다. Log Likelihood와 Perplexity를 이용하여 최적의 토픽 수를 8로 설정한 후, 깁스 샘플링 방법으로 8가지의 토픽을 추론하였다. 추론된 토픽들은 강릉과 해변, 고성과 금강산, KTX와 동해북부선, 주말바다여행, 속초와 통일전망대, 양양과 서핑, 체험관광, 교통망 인프라이다. 추론된 8개의 토픽의 비중을 이용해 동해안 관광에 대한 기사들의 변화를 분석하였다. 그 결과, 통일전망대와 금강산의 비중은 큰 변화가 없는 것으로 나타났고, KTX와 체험관광의 비중은 증가하였고, 그 외의 토픽들의 비중은 2017년에 비해 2018년에 감소하였다. 2019년에는 KTX와 체험관광의 비중은 감소하였으나, 나머지 토픽들의 비중은 큰 변화가 없는 것으로 나타났다.

Abstract The amount of data is increasing through various IT devices in a hyper-connected society where the 4th revolution is progressing, and new value can be created by analyzing that data. This paper was collected total 1,526 articles from 2017 to 2019 in central magazines, economic magazines, regional associations, and major broadcasting companies with the keyword "(East Coast Tourism or East Coast Travel) and Gangwon-do" through Bigkinds. It was performed the topic modeling using LDA algorithm implemented in the R language to analyze the collected 1,526 articles. It was extracted keywords for each year from 2017 to 2019, and classified and compared keywords with high frequency for each year. It was setted the optimal number of topics to 8 using Log Likelihood and Perplexity, and then inferred 8 topics using the Gibbs Sampling method. The inferred topics were Gangneung and Beach, Goseong and Mt.Geumgang, KTX and Donghae-Bukbu line, weekend sea tour, Sokcho and Unification Observatory, Yangyang and Surfing, experience tour, and transportation network infra. The changes of articles on East coast tourism was was analyzed using the proportion of the inferred eight topics. As the result, the proportion of Unification Observatory and Mt. Geumgang showed no significant change, the proportion of KTX and experience tour increased, and the proportion of other topics decreased in 2018 compared to 2017. In 2019, the proportion of KTX and experience tour decreased, but the proportion of other topics showed no significant change.

Key Words : Bigdata, East coast tour, Latent Dirchlet Allocation, Text mining, Tour change, Topic modeling

*Department of Regional Economics, Kangwon National University (jeongeh@kangwon.ac.kr)
Received December 07, 2020 Revised December 10, 2020 Accepted December 15, 2020

1. 서론

빅데이터 기술은 초연결사회, 즉, 4차 산업혁명으로 모바일기기의 확산, SNS의 활성화 등으로 데이터양이 기하급수적으로 증가하기 시작하였을 뿐만 아니라 사물인터넷의 발달로 인한 데이터양 또한 기하급수적으로 증가하고 있다. 따라서 증가한 데이터로부터 새로운 가치와 혁신을 창출하기 위한 여러 가지 방안들이 제시되고 있다[1].

빅데이터에 대한 동향을 파악하는데 사용되는 유용한 기법 중의 하나가 토픽 모델링 기법이다. 박자현, 송민은 1970년부터 2012년까지 문헌정보학 주요학술지의 발표논문초록을 수집하여 LDA(Latent Dirichlet Allocation)기반의 토픽 모델링을 통해 국내 문헌정보학 분야의 연구 동향을 분석하였고[2], Flora는 웹 블로그의 콘텐츠 동향을 토픽 모델링을 통해 분석하였다[3,4].

본 연구에서는 (“동해안 관광” 또는 “동해안 여행”) 그리고 “강원도”를 키워드로 지정하여 강원도 관광과 여행과 관련된 기사를 2017년~2019년까지의 중앙지 10곳, 경제지 8곳, 강원도 지역종합지 2곳, 주요방송사 5곳, 전문지 2곳에서 기사를 수집하여 토픽 모델링 기법으로 강원도 동해안 여행 및 관광의 변화를 분석하고자 한다.

2. 관련연구

2.1 선행연구

토픽 모델링 기법은 빅데이터를 분석 중에서 동향을 분석하는데 다양하게 활용되고 있는 기법이다.

노설현은 토픽 모델링을 활용하여 4차 산업혁명과 인공지능의 핵심기술 분야와 경제, 산업, 교육, 문화 전반에서 일어나고 있는 변화, 그리고 정부의 대응 정책들을 폭넓게 나타내었다[4,5].

황서이, 김문기는 오픈소프트웨어인 R을 활용하여 토픽 모델링과 의미연결망 분석을 사용하였다. 인공 지능분야들을 다룬 학술논문을 1985년 ~ 2018년까지 총 1,691편의 서명, 주제어, 초록을 대상으로 인공지능 분야의 연구 동향을 파악하기 위해 연구토픽과 토픽간 분석을 하였다[4,6].

조경원, 우영운은 2016년부터 2019년까지의 한국학술지인용색인(KCI)에서 “4차 산업” 단어를 포함하고 있는 논문 제목을 수집하고, LDA 분석으로 토픽을 추출하는 등, 텍스트 마이닝 기법을 이용하여 4차 산업과 관련된 키워드의 빈도수, IDM(Intertopic Distance Map), 토픽 및 토픽별 빈도수 상위 단어, 대표 토픽들의 연도별 변화 등을 분석하였다[7].

이와같이 토픽 모델링 기법이 문헌정보학의 연구 동향, 인공지능 분야 연구 동향, 4차산업 연구 동향 웹 블로그 콘텐츠 동향 등의 분석에 활용되고 있다. 본 논문에서는 토픽 모델링 기법을 이용해 동해안 여행 및 관광의 변화에 대해 분석하고자 한다.

2.2 LDA 알고리즘

본 논문에서는 수집된 데이터에서 키워드를 추출하고, 추출된 키워드로부터 토픽을 도출하기 위해 LDA 알고리즘 기반 토픽모델링 기법을 사용한다.

그림 1은 LDA 알고리즘의 절차를 설명한 것으로 이다. 여기서, D는 말뭉치 전체 문서 수, K는 전체 토픽 수, N은 d번째 문서 내 단어 수를 의미한다. 네모칸은 해당 횟수만큼 반복을 의미하고, 동그라미는 변수를 의미한다[7].

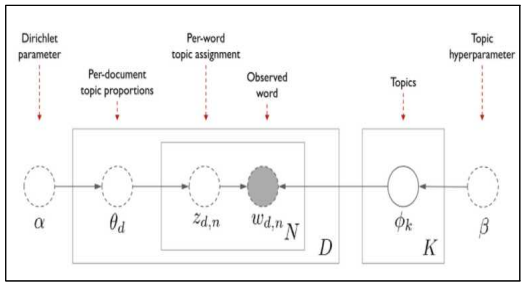

그림 1. LDA 알고리즘 절차
Fig. 1. The procedure of LDA Algorithm

[1단계] 하이퍼 파라미터인 α와 β의 값과 전체 토픽의 개수 K를 설정한다.

[2단계] 문서(D)의 단어를 토픽(K)별로 할당한 후에, d번째 문서가 가진 토픽의 비중인 θ_d를 계산하고, k번째 토픽에 해당하는 비중인 ϕ_k를 계산한다.

[3단계] θ_d와 ϕ_k의 결과값을 이용하여 d번째 문서 n번째 단어가 어떤 토픽에 해당하는지 할당($z_{d,n}$)해주고, d번째 문서에 등장하는 n번째 단어를 할당($w_{d,n}$)해준다.

LDA 분석에서는 토픽 수 K는 사용자가 지정하므로 좀 더 정확한 분석을 위해서는 최적의 토픽 수를 지정해야 한다. 이때, 최적의 토픽 수를 찾을 때 Perplexity 지표를 사용한다.

우선 깁스 샘플링으로 구한 ϕ와 θ을 활용해 전체 문서, 모든 단어의 발생확률 $p(w)$을 로그로 취한 결과가 식(1)과 같다[9].

$$\log\{p(w)\} = \sum_{d=1}^{D}\sum_{j=1}^{V} n^{jd} \log\left[\sum_{k=1}^{K} \theta_k^d \phi_k^j\right] \quad (1)$$

식 (1)을 이용한 perplexity 지표는 식(2)과 같이 계산한다.

$$perplexity(w) = \exp\left[-\frac{\log\{p(w)\}}{\sum_{d=1}^{D}\sum_{j=1}^{V} n^{jd}}\right] \quad (2)$$

여기서 p(w)는 클수록 좋은 추정이므로 exp(-log(p(w))는 작을수록 좋은 것이므로, 토픽 수 K에 따라 perplexity를 구한다. 이때, perplexity가 가장 작을 경우의 K를 최적의 토픽 수로 설정하면 된다[9].

3. 연구방법

3.1 자료수집

본 논문에서는 강원도 동해안 여행과 관련된 자료를 수집하기 위해 빅카인즈(www.kinds.or.kr)를 이용하였고, 좀 더 다양한 동해안 여행에 관련된 자료를 수집하기 위해 검색어로 "(동해안 관광, 동해안 여행) AND 강원도"을 사용하였고, 검색 기간은 2017년 1월 1일부터 2019년 12월 31일로 하였다. 그 결과, 총 1556건의 기사가 검색되었고, 그중에서 중복되거나 강원도 관광과 관련성이 낮은 30건의 기사를 제외한 1,526건을 분석하였다.

"(동해안 관광, 동해안 여행) AND 강원도"로 검색하였을 때, 2017년 481건, 2018년 478건, 2019년 597건의 기사가 검색되었고, 100건 이상의 기사가 검색된 곳은 강원도민일보(363건), 강원일보(251)건으로 총 2곳이었다. 수집된 기사를 정치, 경제, 사회, 문화 등의 분야별로 분류했을 때, 지역(928), 문화(418), 정치(358) 순으로 나타났다. 그림 1은 수집된 기사의 통합분류 결과를 나타낸 것이다.

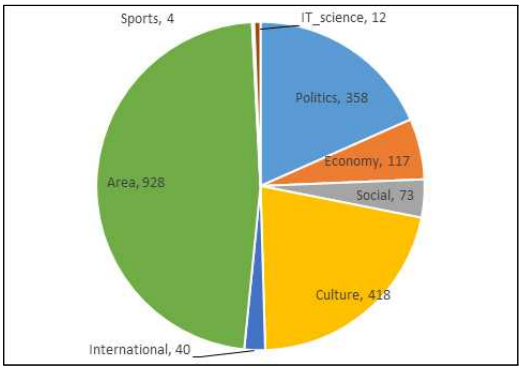

그림 1. 분야별 현황
Fig. 1. Status by field

3.2 데이터 처리 및 분석

빅카인즈를 이용하여 수집된 기사들을 LDA 분석을 시작하기 전에, 말뭉치(corpus)로 읽어들인 후, 명사 중심으로 데이터를 추출하고, 불용어를 제거하는 전처리 과정을 실행하여 LDA 알고리즘을 실행할 수 있는 데이터 형식으로 변환시켰다.

LDA 분석은 R 언어에서 제공하는 LDA 함수 중인 lda.collapsed.gibbs.sampler() 함수를 이용하였다. LDA 분석을 실시하기 위해 먼저 토픽의 수를 결정해야 하는데, 본 논문에서는 Log Likelihood와 Perplexity를 이용하여 최적의 토픽 수를 결정하였다. 그림 2는 수집된 데이터를 이용하여 토픽의 개수 별로 perlexity를 계산한 결과이다. 즉, 토픽의 수가 2개에서 15개로 변화할 때, perplexity 값의 변화를 측정하였다. 그 결과, Perplexity값의 변화가 가장 작은 경우 8를 토픽 수로 설정하여 LDA 분석을 하였다.

또한, 나머지 Dirichlet parameter인 알파(alpha)와 에타(eta)는 각각 0.02, 0.04, 0.06, 0.08, 0.1로 반복 실험하였다. 파라미터값에 따라 추출된 토픽들을 비교하여 토픽들의 구분이 좀 더 명확해지는 0.04로

Dirichlet parameter로 설정하였다.

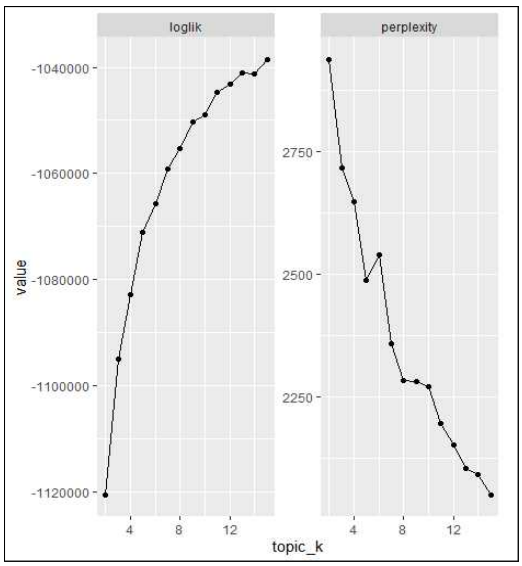

그림 2. 토픽 수에 따른 Perplexity 값
Fig. 2. The perplexity value according to topic count

4. 연구결과

4.1 키워드 빈도 분석

2017년~2019년까지의 "(동해안 관광, 동해안 여행) AND 강원도"에 관련된 총 1,526건의 기사들의 내용 중에서 강원도 여행에 관련이 높은 기사들을 추출하였다. 추출된 기사들 중에서 빈도수가 가장 높은 키워드들을 각각의 년도별로 정리한 결과가 표 1과 같다.

표 1에서 설명하고 있듯이, 강원도 동해안 관광 및 여행에 대한 상위 20위 내의 키워드들은 2017년 해변(해수욕장) 474건, 강릉 447건, 올림픽 372건, 개통 313건, 평창 292건 순으로 나타났고, 2018년 열차(기차) 352건, 강릉 278건 고성 215건, 평화 200건, 축제, 178건 순으로 나타났다. 그리고 2019년에는 해변 346건, 고성 311건, 강릉 282건, 바다 248건, 금강산 225건 순으로 나타났다. 이외에도 3년 동안 상위 20위에 포함되는 키워드들이 KTX, 통일전망대, 양양, 서핑 등의 키워드들이 나타났다.

동해안의 대명사라고 할 수 있는 해변, 강릉, 바다는 상위 10위 내에 위치하고 있고, 특히 키워드 중에서 고성은 2017년 13위였는데, 2018년 금강산관광에 대한 관심으로 3위로 빈도수가 높게 나타났고, 2019년에는 고성 산불로 인해 빈도수가 2018년도보다 더 높게 나타난 것으로 볼 수 있다.

표 1. 상위 20 키워드
Table 1. Top 20 Keyword

Rank	2017		2018		2019	
	Nouns	Freq.	Nouns	Freq.	Nouns	Freq.
1	Beach	474	Train	352	Beach	346
2	Gangneung	447	Gangneung	278	Gosung	311
3	Olympic	372	Gosung	215	Gangneung	282
4	Opening	313	Peace	200	Sea	248
5	Pyeong-chang	292	Festival	178	Mt. Gumgang	225
6	Seoul	283	Make up	176	Forest fires	162
7	Yangyang	267	Mt.Gumgang	171	Sokcho	147
8	Festival	266	Olympic	169	Downtown	130
9	Train	253	Beach	168	Donghae	122
10	Sea	218	Sea	163	Business	122
11	Section	195	Unification Observatory	131	Train	122
12	Make up	165	Business	130	Plan	116
13	Gosung	150	Downtown	129	Surfing	113
14	Samcheok	148	Samcheok	125	Yangyang	112
15	KTX	146	Connect	125	Start	109
16	Sokcho	138	North and South	122	Course	106
17	Facility	137	Opening	119	Damage	101
18	Highway	133	Experience	119	Festival	100
19	Operation	131	Seoul	117	Make up	98
20	Hold	127	DMZ	114	Propel	98

그림 3은 2017년부터 2019년까지의 추출된 키워드들의 빈도수가 높은 상위 100개를 이용해 시각화시킨 워드 클라우드이다. 그림 3에서 설명하고 있듯이 빈도수가 높은 단어일수록 다른 키워드에 비해 크게 표현되었다. 그림 3을 통해, 동해안 관광 및 여행에 대한 기사에서 가장 많이 언급되었던 단어들을 해변, 강릉, 열차, 고성, 바다, 올림픽, 바다, 축제, 금강산, KTX 등이 있음을 확인해 볼 수 있다.

그림 3. 상위 100개 키워드에 대한 워드 클라우드
Fig. 3. Word cloud by top 100 Keyword

4.2 토픽 분석

2017년~2019년까지의 동해안 관광 및 여행에 관련된 저널의 기사 또는 방송 데이터에서 중심이 되는 토픽을 추출한 결과가 표 2와 같다.

본 논문에서는 perplexity의 변화가 가장 작을 때인 8일 때를 토픽의 수로 설정하였고, 토픽별 15개의 단어들의 연관성을 중심으로 토픽의 이름을 설정하였다.

토픽 1은 강릉, 해변, 최대, 피서객, 환경, 축제 등의 키워드로 구성되어 있다. 동해안 여행 및 관광 관점에서 강릉 해변, 해변 축제, 최대 피서객 방문, 동계올림픽 개최로 인한 평화 축제 등으로 연관 지어 볼 때, 강릉과 해변으로 동해안 관광 토픽을 추론할 수 있다.

토픽 2는 고성, 금강산, 올림픽, 개최, 개통, 참여 등의 키워드로 구성되어 있다. 평창올림픽을 계기로 금강산 관광에 대한 관심이 높아졌음을 알 수 있으며, 평창올림픽 개최, 금강산 관광 등으로 연관 지어 볼 때, 동해안 관광 토픽으로 고성과 금강산으로 추론할 수 있다.

토픽 3은 열차, KTX, 시설, 문화, 서울, 동해북부선, 활성 등의 키워드로 구성되어 있다. 동계올림픽을 개최함에 따라 서울과 강릉간의 KTX가 구축되었고, 더 나아가 남북간의 철도를 연결하고자 하는 동해북부선에 대한 관심이 증가하였음을 알 수 있다. 열차, 서울-강릉간 KTX 개통, 동해북부선 추진 등으로 연관 지어 볼 때, 동해안 관광 토픽으로 KTX와 동해북부선으로 추론할 수 있다.

표 2. 토픽 모델링 분석 결과
Table 2. The analysis result of topic modeling

No	Topics Name	Topics	Freq. (%)
1	Gangneung and beach	Gangneung, Beach, Plan, Use, Maxium, A summer vacationer, Environment, Wind, Peace, Festival, Citizen	11.63
2	Gosung and Mt.Gumgang	Gosung, Mt.Gumgang, Opening, Course, Hold, City, Event, Olympic, Participation, Downtown	10.00
3	KTX and Donghae -Bukbu line	Train, KTX, Facility, Culture, Goods, Seoul, Uses, Donghae-Bukbu line, Propel, Vicinity, Resource, Activation	9.17
4	Weekend Sea tour	Sea, Pyeongchang, Festival, Cohort, Seoul, Link, Ready, Show, Food, Weekend, Busan	8.41
5	Sokcho and Unification Observatory	Sokcho, Operation, Economy Unification observatory, Wave, North korea, Goverment, Person, World, Yangyang,	8.09
6	Yangyang and Surfing	Olympic, Yangyang, Winter, Development, Location, Surfing, Prople, Route, Change, Open,	7.52
7	Experience tour	Business, Furtherance, Samcheok, Experience, National highway, Start, Uljin, Visit, Observatory	6.77
8	Transportation network infra	Highway, Connect, Donghae, Sights, Section, Nature, Popularity, Expansion, Transportation network, Infra, Famous	6.69

토픽 4는 바다, 서울, 평창, 주말, 공연, 음식 등의 키워드로 구성되어 있다. 강릉-서울간 KTX 개통으로 인해 이동 시간이 단축됨에 따라 지역 축제 및 공연, 맛집 탐방 등 주말여행에 대한 관심이 높아졌음을 알 수 있다. 동해안 관광 토픽으로 주말 바다 여행으로 추론할 수 있다.

토픽 5는 양양, 속초, 북한, 통일전망대, 정부, 세계 등의 키워드로 구성되어 있다. 대한민국의 최북단에 위치하고 있는 고성 통일전망대가 속초여행의 필수코스로 추천되고, 금강산의 봉우리를 육안으로 볼 수 있는

통일전망대와 DMZ에 대한 관심이 높아졌음을 알 수 있다. 토픽 5의 키워드들을 기반으로 동해안의 관광 토픽으로 속초와 통일전망대로 추론하였다.

토픽 6은 올림픽, 양양, 겨울, 서핑, 개장 등의 키워드로 구성되어 있다. 평창 동계올림픽 개최로 인한 KTX 개통과 양양고속도로가 양양을 서핑 메카로 만드는 기폭제 역할을 하였다고 할 수 있다. 양양은 21개의 해변 중에서 14개 해변에 서핑 숍이 있고, 매년 서핑페스티벌이 개최될 정도로 유명한 곳이기도 하며, 겨울에 서핑을 즐기는 서퍼들도 있다. 토픽 6의 키워드들로 동해안의 관광 토픽으로 양양과 서핑으로 추론하였다.

토픽 7은 사업, 조성, 삼척, 체험, 방문, 전망대 등의 키워드로 구성되어 있다. 어촌체험, 낚시, 해상케이블카, 해양레일바이크, 석회동굴 등의 직접 체험할 수 있는 관광인 체험관광으로 동해안 관광 토픽으로 추론하였다.

토픽 8은 고속도로, 구간, 연결, 확충, 교통망, 인프라 등의 키워드로 구성되어 있다. 올림픽이 개최되는 평창과 강릉시, 그리고 서울 간에 교통망 인프라가 구축되어 동해안 강릉시의 접근성을 향상 시켰고, 양양고속도로 개통으로 고속도로를 확충하여 양양, 속초 지역의 접근성을 향상 시켰다. 장기적인 관점에서 동북아~유라시아를 잇는 글로벌 교통망을 구축할 것으로 기대하며 동해안 관광 토픽으로 교통망 인프라로 추론하였다.

그림 4는 2017년부터 2019년까지의 강원도 동해안 관광 또는 여행에 대한 8개의 토픽들에 대한 비중의 변화를 나타낸 것이다.

전체적으로 2017년에 비해 2019년에 토픽들에 대한 비중이 감소하는 것으로 나타났다. 강릉과 해변은 2017년에 비해 2018년은 감소하였으나, 2018년과 2019년은 큰 변화 없이 관심이 지속 되었다. 하지만, 속초와 통일전망대, 고성과 금강산은 큰 변화가 없으나, 동계올림픽으로 인해 KTX와 동해북부선에 대한 관심은 2018년 증가하였으나 2019년에 급격하게 감소한 것으로 나타났다.

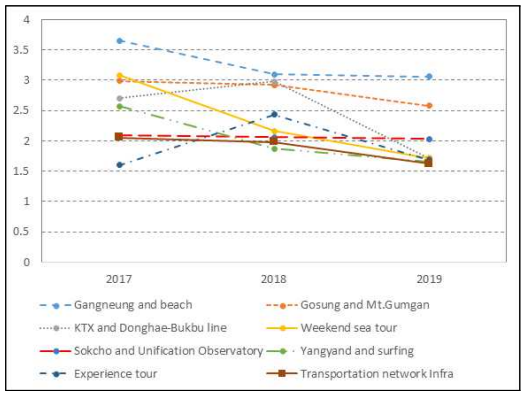

그림 4. 토픽의 변화
Fig. 4. Topic variance

5. 결론

본 연구에서는 (“동해안 관광” 또는 “동해안 여행”) 그리고 “강원도”를 키워드로 지정하여 강원도 관광과 여행과 관련된 기사를 2017년~2019년까지의 중앙지를 비롯한 경제지, 강원도 지역종합지, 주요방송사, 그리고 전문지에서 총 1556건의 기사를 수집하여, 그 중에서 중복되거나 강원도 관광과 관련성이 낮은 30건의 기사를 제외한 1,526건을 분석하였다.

본 연구 주제인 강원도 동해안 관광 및 여행에 대한 기사에서 가장 많이 언급되었던 단어들을 해변, 강릉, 열차, 고성, 바다, 올림픽, 바다, 축제, 금강산, KTX 등이었고, 추론된 토픽은 “강릉과 해변”, “고성과 금강산”, “KTX와 동해북부선”, “주말 바다여행”, “속초와 통일전망대”, “양양과 서핑”, “체험관광”, “교통망 인프라”이다.

2017년부터 2019년까지의 동해안 관광 또는 여행에 대한 토픽들은 전체적으로 비중이 감소하는 것으로 나타났으나 고성과 금강산, 속초와 통일전망대는 큰 변화가 없었다. 특히, KTX와 동해북부선, 체험관광은 2018년 증가하였다가 2019년 크게 감소하는 것으로 나타났다. 강릉, 해변, 양양, 서핑은 2017년에 비해 2018년에는 감소하였으나, 2019년에는 큰 변화가 없는 것으로 나타났다.

2018년 평창 동계올림픽의 전후로 수집된 기사를 토픽 모델링 분석한 결과, 올림픽 개최를 위한 교통망 인프라 구축으로 인한 접근성 향상으로 동해안 관광이

변화한 것으로 보인다. 특히, 동계올림픽이 개최되던 해에 금강산, DMZ 등에 대한 관심이 증가한 것으로 나타났다.

REFERENCES

[1] S.Y. Kim, J.H. Do, B.R. Kim, "Bigdata", KISTEP, 2018-11, pp.1-33 2018.

[2] J.H. Park and M. Song, "A study on the research trends in library & inormation science in Korea usning topic modeling", Journal of the Korea Society for Information Management, vol. 30, No. 1, pp.7-32, 2013.

[3] S.T. Floar, "Blogger-link-topic model for blog mining," Pacific-Aisa Conference on Knowledge Discovery and Data mining, pp.28-39, 2017

[4] S.H. Noh, "Analaysis of Issues Related to the Fourth Industrial Revolution Based on Topic Modeling," Journal of Digital Contents Society, vol.21, no.3, pp.551-560, Mar. 2020

[5] S.H. Noh, "Analaysis of Issues Related to Artificial Intelligence Based on Topic Modeling," Journal of Digital Convergence, vol.18, no.5, pp.75-87, May. 2020

[6] S. I. Hwang, M. K. Kim, "An Analysis of Aritifical Intelligence(A.I.) related Studies' Trends in Korea Focused on Topic Modeling and Semantic Network Analysis," Journal of Digital Contents Society, vol.20, no.9, pp.1847-1855, Sep. 2019

[7] K.W. Cho, Y.W. Woo, "Topic Modeling on Research Trends of Industry 4.0 Using Text Mining," Journal of the Korea Institute of Information and Communication Engineering, vol.23, no.7, pp.764-770, July. 2019.

[8] D.H. Kim, D.S. Seo, S.Y. Cho, P.S. Kang, "Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec," ScienceDirect, Vol. 477, pp.15-29, Mar. 2019.

[9] https://ratsgo.github.io/from%20freq uency%20to%20semantics/2017/06/01/LDA/

저자약력

정 은 희(Eun-Hee Jeong) **[종신회원]**

- 1998년 2월 : 관동대학교 일반대학원 전자계산공학과 (공학석사)
- 2003년 2월 : 관동대학교 일반대학원 전자계산공학과 (공학박사)
- 2003년 9월 ~ 현재 : 강원대학교 글로벌인재학부 지역경제전공 교수

〈관심분야〉

전자상거래 보안, 데이터 마이닝, 텍스트 마이닝, 빅데이터 분석