

# 텍스트 마이닝과 네트워크 분석을 이용한 지역 이미지 변화 분석

정은희\*

## Regional Image Change Analysis using Text Mining and Network Analysis

Eun-Hee Jeong\*

**요약** 소셜미디어 빅데이터는 소비자의 소비형태 뿐만 아니라 지역의 이미지를 파악할 수 있는 많은 정보가 포함되어 있다. 본 논문에서는 국내 포털 사이트인 네이버와 다음의 Blog와 Cafe로부터 '삼척'이 포함된 데이터를 2015년부터 2019년까지 1년 단위로 수집하였고, 텍스트 마이닝과 네트워크 분석을 실시하여 지역 이미지를 형성하는 키워드를 추출하고 지역 이미지 변화를 분석하였다. 연구 결과에 따르면, 2015년 지역 이미지는 '장호항', '동해', '해수욕장' 등 인근 지명이나 장소 등의 이미지 인지적 요소들로 표현되고 있는데, 2016년과 2019년은 지역 내의 특정 장소인 삼척출비치로 이미지 인지적 요소가 변한 것을 알 수 있다. 그리고 지역 이미지와 연관된 키워드들이 삼척을 대표하는 명소인 '장호항', 리조트가 포함하고 있는 것을 보아 지역 이미지 형성에 인프라 시설 요소가 큰 역할을 한다고 볼 수 있다. 네트워크 데이터에 대한 유의성 검증은 부트스트랩 기법을 이용하였고, 2015년, 2016년, 2019년 p-value가 각각 0.0002, 0.0006, 0.0002로 유의수준 5%에서 통계적으로 유의한 것으로 나타났다.

**Abstract** Social media big data includes a lot of information that can identify not only consumer consumption patterns but also local images. This paper was collected annually data including 'Samcheok' from 2015 to 2019 from Blog and Cafe of Naver and Daum in domestic portal site, and analyzed the regional image change after refining keyword which forms the regional image by performing text mining and network analysis. According to the research results, the regional image of 2015 was expressed with image cognitive elements of the nearby place name or place etc. such as 'Jangho Port', 'Donghae', and 'Beach'. However the regional image both 2016 and 2019 were changed with image cognitive elements of 'SamcheokSolbich' which is a special place within region. Therefore as the keywords related to the local image include 'Jangho Port' and Resort, which are the representative attractions of Samcheok, it can be seen that the infrastructure factor plays a big role in forming the local image. The significance test for the network data used the bootstrap technique, and the p-values in 2015, 2016, and 2019 were 0.0002, 0.0006, and 0.0002, respectively, which were found to be statistically significant at the significance level of 5%.

**Key Words** : Bootstrap, CONCOR analysis, Degree centrality analysis, Social Big data, Text mining

### 1. 서론

삼척시 통계연보에 따르면, 2015년 숙박 및 음식점 사업체 수가 1,858개, 주요관광지 방문객 수가 유료관광지에 1,655,596명, 숙박업에 75,057명, 무료관광지에 4,380,164명이고, 해수욕장 개장기간 동안 이용자

수가 3,980,040명이었는데[1], 2016년 숙박 및 음식점 사업체 수가 1,940개, 주요관광지 방문객 수가 유료관광지에 1,780,335명, 숙박업에 676,758명, 무료관광지에 4,550,449명이고, 해수욕장 개장기간 동안 이용자 수가 4,185,998명이다[2]. 2015년 대비 2016

\*Department of Regional Economics, Kangwon National University  
 Received February 13, 2022

Revised February 18, 2022

Accepted March 09, 2022

년을 비교해보면, 숙박 및 음식점 사업체 수가 4% 증가하였고, 주요관광지 방문객 수는 유료관광지 8%, 숙박업 900%, 무료관광지 4%, 해수욕장 이용자 수 5% 증가하였다. 특히 2016년 리조트의 개장으로 인해 숙박업 방문객 수가 급격하게 증가한 것으로 보인다.

초연결사회인 현재, 소셜미디어를 사용하는 사용자들은 여행을 준비할 때는 소셜 미디어상에서 관광지의 정보를 탐색하며 관광 관련 유인물이나 책자 또는 맛집, 포토존 등 주변 여행지역 등에 대해 학습한 후에 여행지에서의 관광 경험을 하거나, 관광 경험 후에는 관광지에 대한 정보를 공유 또는 관광지를 추천하는 등 소셜미디어 커뮤니케이션을 한다[3].

소셜미디어 빅데이터 분석은 소셜미디어 데이터를 전수조사로, 소셜미디어 사용자들의 관심뿐만 아니라 내면에 있는 진심을 파악할 수 있다. 또한, 사용자들의 현재의 심리상태, 소비, 행동에 대한 단기적인 예측이 가능하다고 볼 수 있다. 따라서 이러한 소셜미디어에서 생산되는 빅데이터로부터 이용자들의 의견, 관심, 이슈 등을 분석함으로써 지역에 대한 이미지를 분석할 수 있다.

본 논문에서는 공유되는 소셜미디어 빅데이터를 리조트 운영 이전과 이후로 구분하여 수집하고, 텍스트 마이닝 기법으로 주요 키워드를 분석하고, 분석된 키워드를 활용한 네트워크 분석을 통해 지역 이미지 변화를 분석하고자 한다.

## 2. 이론적 배경

온라인상의 데이터, 즉 유튜브, 페이스북, 블로그, 인스타그램 등에서 수집한 비정형 빅데이터로부터 정보를 추출하고 분석하는 기법에는 텍스트 마이닝, 오피니언 마이닝, 빈도분석, 연결중심성 네트워크 분석, CONCOR 분석 등이 있다.

다양한 학문 분야에서 빅데이터를 이용해 연구를 진행하고 있듯이 관광 부문에서도 소셜미디어에 나타난 빅데이터 분석을 통해 새로운 정보를 찾아내어 트렌드를 파악하고 이를 향후 마케팅에 활용하려는 시도가 이어지고 있다[4].

윤영일, 하동현(2017)은 텍스트를 이용하여 비즈니스호텔에 대한 국민들의 인식을 비즈니스호텔 관

련 소셜미디어 빅데이터로부터 수집하고, 국민들의 호텔에 대한 인식 변화를 Ucinet 6과 NetDraw로 시각화하였다. 국민들은 비즈니스호텔에 대해 합리적이면서 도심형 숙박시설 등으로 인식하는 것으로 나타났다[5]. 황선옥, 민정기(2017)는 리조트 분야 빅데이터를 네이버와 다음의 블로그, 카페 등, 구글 웹과 구글 뉴스, 페이스북에서 수집하고, 수집된 데이터를 텍스트마이닝으로 정제하고, 의미 있는 분석을 위하여 Ucinet6 도구의 NetDraw으로 중심성 분석을 하였다. 그리고 CONCOR 분석 등을 실행하여 데이터들을 군집화함으로써 관광객들이 어떠한 내용에 관심이 있는지, 무엇을 원하는지에 대한 소비성향을 분석하였다[6]. 최우성(2019)은 소셜미디어 빅데이터 분석을 통해 리조트 인식에 관한 연구를 진행하여 사람들이 리조트를 가족 단위로 즐길 수 있는 체험 위주의 시설로 인식하고, 고용창출 등 긍정적인 산업효과가 있는 것으로 인식한다는 것을 도출해 내었고[7], 임종훈, 김영현(2020)은 SNS 데이터를 수집하여 텍스트 마이닝과 의미연결망 분석을 진행하였고, 그 결과를 이용해 익산시 관광 관련 단어들을 검토하였다[4][8].

본 연구에는 소셜미디어 중 블로그와 카페로부터 '삼척'이라는 핵심 키워드가 포함되어 있는 데이터만을 수집하여 텍스트 마이닝과 네트워크 분석을 통해 지역 이미지를 분석하고자 한다.

## 3. 연구설계

본 연구는 다음과 네이버의 블로그와 카페로부터 수집한 비정형데이터로부터 정보를 키워드를 추출하고, 추출된 키워드를 이용하여 지역의 변화를 분석하는데 목적이 있다.

### 3.1 데이터 수집

본 연구는 지역 이미지를 추출하기 위해 핵심 키워드를 '삼척'으로 선정하였고, 데이터의 수집 기간은 리조트 운영 전인 2015년, 운영을 시작한 해인 2016년부터 2019년까지 1년 단위로 구분하여 다음과 네이버의 블로그와 카페로부터 데이터를 수집하

였다. 데이터 수집은 소셜 매트릭스 프로그램 '텍스톰(TEXTOM)'을 활용하였고[9], 수집된 데이터는 표 1과 같다.

표 1. 수집된 데이터  
Table 1. Collected Data

Social Network		Blog	Cafe
2015	Daum	719	992
	Naver	990	780
2016	Daum	718	992
	Naver	990	780
2017	Daum	711	987
	Naver	930	740
2018	Daum	932	992
	Naver	929	850
2019	Daum	730	990
	Naver	660	740
All		8,309	8,843

### 3.2 데이터 정제

블로그와 카페의 텍스트는 다양한 체험과 감성 정보를 담고 있지만, 수집된 데이터에 포함되어 있는 중복된 데이터, 광고성 데이터, 관련성이 낮은 데이터는 제거하고, 비정형적 데이터이므로 효율적인 텍스트 마이닝과 의미연결망 분석을 위해 데이터 정제과정을 다음과 같이 진행하였다.

[1단계] 텍스톰에서 제공하는 Espresso K를 이용하여 텍스트를 분석하기 위한 형태소 분석작업을 수행.

[2단계] 추출된 명사 중 동의어를 재정리하고, 고유명사를 통일하는 데이터 정제 작업을 수행.

정제된 데이터를 텍스트 마이닝을 통해 키워드의 빈도수를 계산하고, 키워드 중에서 주요 단어를 선정하였다. 그런데 2017년부터 2019년까지의 키워드 분석 결과가 유사하여 매트릭스 데이터를 2015년, 2016년, 2019년의 키워드로 생성하고, CONCOR 분석을 진행해 군집화하였다. 그리고 네트워크 분석 결과를 Uninet6의 NetDraw를 이용해 시각화하였다.

## 4. 분석 결과

### 4.1 키워드 분석 및 연결 중심성 분석

'삼척'을 키워드로 수집된 데이터 중 2015년, 2016년, 그리고 2019년의 데이터 중에서 지역을

대표하는 키워드 65개와 키워드 65개에 대한 연결 중심성은 표 2와 같다.

동시 출현 빈도수 비교 결과, 소셜미디어에서 삼척에 대해 언급할 때, 2015년 빈도수는 '장호항', '삼척여행', '동해', '해양레일바이크', '강릉', '환선굴', '정동진', '바다', '태백', '속초', '죽서루', '삼척맛집', '산행', '새천년도로', '바다열차' 순으로 나타났으며, 지명이나 장소들이 많이 포함되어 있다. 즉, 2015년 지역 이미지는 삼척과 인근 지명이나 장소 등의 이미지 인지적 요소들로 표현되고 있다.

연결 중심성이 높을수록 지역 이미지 형성에 영향력이 높은 키워드라 할 수 있는데, 전체 데이터의 연결 중심성 분석 결과, 2015년 연결 중심성은 '정동진', '동해', '장호항', '해양레일바이크', '강릉', '태백', '환선굴', '바다열차', '대금굴', '삼척여행', '해돋이', '무릉계곡', '추암촛대바위' 순으로 나타났다. 특히 '정동진', '동해', '장호항'이 높게 나타났는데, 이 지역들과 관련된 다양한 키워드가 지역 이미지 형성에 큰 역할을 한다고 볼 수 있다.

표 3은 2016년 동시 출현 키워드 빈도수와 연결 중심성 결과이다. 즉, 소셜미디어에서 삼척에 대해 언급할 때 빈도수는 '삼척솔비치', '삼척여행', '장호항', '삼척맛집', '동해', '강릉', '해양레일바이크', '바다', '산토리니', '환선굴', '무견리이끼폭포', '여름휴가', '산행', '가족여행' 순으로 나타났다. 2015년과 비교해볼 때, 지역 내의 특정 장소로 이미지 인지적 요소가 변한 것을 알 수 있다.

연결 중심성은 '삼척솔비치', '삼척여행', '장호항', '삼척맛집', '동해', '강릉', '해양레일바이크', '산토리니', '바다', '정동진', '환선굴', '무견리이끼폭포', '산행', '여름휴가', '속초', '가족여행' 순으로 나타났다. 연결 중심성이 가장 큰 '삼척솔비치'가 지역 이미지 형성에 큰 역할을 한다고 볼 수 있다. 따라서 2015년 삼척 숙박업 이용자 수가 75,057명에서 2016년 676,758명으로 900% 증가한 것이 삼척솔비치 리조트 개장과 무관하다고 볼 수 없다.

표 2. 키워드와 연결 중심성(2015년)

Table 2. Keyword and degree centrality(2015)

No	Keyword	Freq.	degree centrality	Rank
1	Jangho port	953	0.104	3
2	Samcheok travel	634	0.047	10
3	Donghae	629	0.125	2
4	Marine rail bike	493	0.093	4
5	Gangneung	462	0.093	4
6	Hwanseon cave	404	0.073	7
7	Jeongdongjin	348	0.138	1
8	Sea	292	0.027	22
9	Taebaek	242	0.083	6
10	Sokcho	196	0.040	16
11	Jukseoru	183	0.013	35
12	Samcheok restaurant	183	0.013	35
13	Hiking	177	0.014	31
14	New millennium coast road	166	0.014	31
15	Sea train	165	0.071	8
16	Summer vacation	162	0.014	31
17	Mugeonli moss waterfall	160	0.008	43
18	Snorkeling	158	0.016	28
19	Rail Bike	155	0.042	14
20	Daegum cave	150	0.049	9
21	Family trip	149	0.016	28
22	Korea naples	147	0.012	37
23	Rape flower festival	143	0.006	47
24	Haesindang park	140	0.014	31
25	Samcheok beach	140	0.011	39
26	Deokhangsan	136	0.044	13
27	Chuam candlestick rock	136	0.009	41
28	Jeongseon	133	0.024	26
29	Sunrise	124	0.046	11
30	Mureung valley	117	0.045	12
31	Deokpung valley	113	0.007	45
32	Dootasan	109	0.009	41
33	Inwon port	97	0.010	40
34	Isabusaja park	95	0.019	27
35	Samcheok port	95	0.012	37
36	Maengbang beach	81	0.007	45
37	Rose park	79	0.004	51
38	Daegwanlyeong sheep ranch	75	0.041	15
39	Mukho	69	0.037	17
40	Fishing	66	0.005	48
41	Samcheok coast	66	0.004	51
42	Mulhoe	57	0.033	19
43	Nongoldam-gil	57	0.004	51
44	Yukbaeksan	56	0.004	51
45	Osibcheon	54	0.004	51
46	Hill of wind	53	0.036	18
47	Mangsang beach	53	0.025	24
48	Eungbongsan	51	0.031	21
49	Mukho lighthouse	51	0.005	48
50	Daewallyeong samyang ranch	44	0.032	20
51	Mangsang	44	0.025	24
52	Gyeongpodaе	43	0.008	43
53	Jeongwol daeboreum festival	43	0.002	63
54	Water chicken ribs	41	0.004	51
55	Surobinheonwha park	38	0.004	51
56	Tracking	36	0.003	60
57	Yangyang	35	0.005	48
58	Yongwha beach	35	0.003	60
59	Transparent canoe	33	0.004	51
60	Jangho beach	32	0.004	51
61	Gangwonionghab museum	31	0.026	23
62	Coal museum	26	0.016	28
63	Builmakguksu	25	0.002	63
64	Seongwon chicken ribs	24	0.003	60
65	Haeparang-gil	21	0.001	65

표 3. 키워드와 연결 중심성(2016년)

Table 3. Keyword and degree centrality(2016)

No	Keyword	Freq.	Degree Centrality	Rank
1	Samcheok Sol Beach	1554	0.091	1
2	Samcheok travel	689	0.055	2
3	Jangho port	554	0.045	3
4	Samcheok restaurant	449	0.026	4
5	Donghae	360	0.024	5
6	Gangneung	266	0.023	6
7	Marine rail bike	265	0.020	7
8	Sea	221	0.016	9
9	Santorini	186	0.017	8
10	Hwanseon cave	170	0.014	11
11	Mugeonli moss waterfall	156	0.014	11
12	Summer vacation	156	0.012	14
13	Hiking	153	0.014	11
14	Family trip	152	0.011	16
15	Sokcho	137	0.012	14
16	Jeongdongjin	131	0.015	10
17	Samcheok beach	118	0.008	22
18	Korea naples	117	0.010	18
19	Jukseoru	108	0.005	35
20	Snorkeling	105	0.011	16
21	Taebaek	104	0.009	20
22	Dootasan	103	0.009	20
23	Aquaworld	101	0.010	18
24	Min waterfall	90	0.004	42
25	Samcheok coast	89	0.008	22
26	Yukbaegsan	89	0.007	26
27	Rape flower festival	87	0.005	35
28	Yangyang	84	0.006	30
29	Rail bike	83	0.008	22
30	Deokhangsan	81	0.007	26
31	Sashimi	80	0.005	35
32	Sunrise	78	0.008	22
33	Haesindang park	76	0.006	30
34	Daegum cave	73	0.007	26
35	Deokpung valley	69	0.006	30
36	New millennium coast road	66	0.005	35
37	Eungbongsan	66	0.004	42
38	Rose festival	65	0.003	49
39	Mureung valley	63	0.007	26
40	Jeongseon	62	0.006	30
41	Rose park	51	0.003	49
42	Chuam candlestick rock	48	0.005	35
43	Tracking	44	0.005	35
44	Sea train	40	0.006	30
45	Isabusaja park	39	0.003	49
46	Yangyang Sol Beach	37	0.003	49
47	Osibcheon	37	0.002	54
48	Jangho beach	36	0.004	42
49	Highwonchuchupark	36	0.002	54
50	Transparent canoe	35	0.004	42
51	Daegwanlyeong sheep ranch	32	0.005	35
52	Water chicken ribs	31	0.002	54
53	Fishing	31	0.001	60
54	Mukho	29	0.004	42
55	Haeparang-gil	29	0.002	54
56	Builmakguksu	28	0.002	54
57	Surobinheonwha park	26	0.004	42
58	Daegwanlyeong samyang ranch	26	0.001	60
59	Gijuldaligi	26	0.001	60
60	Maengbang beach	25	0.001	60
61	UNESCO	24	0.001	60
62	Jeongwol daeboreum festival	23	0.001	60
63	Mukho lighthouse	22	0.003	49
64	Nongoldam-gil	21	0.004	42
65	Shinumsan	20	0.002	54

표 4. 키워드와 연결 중심성(2019년)  
Table 4. Keyword and degree centrality(2019)

No	Keyword	Freq.	Degree Centrality	Rank
1	Samcheok Sol Beach	1046	0.138	1
2	Samcheok travel	417	0.061	2
3	Donghae	318	0.050	4
4	Jangho port	308	0.053	3
5	Samcheok restaurant	307	0.037	6
6	Gangneung	256	0.039	5
7	Sea	233	0.034	7
8	Family trip	153	0.027	8
9	Haeparang-gil	141	0.026	9
10	Hwanseon cave	141	0.025	10
11	Hiking	127	0.019	14
12	Marin rail bike	124	0.020	13
13	Marin cable car	123	0.021	12
14	Samcheok section	117	0.022	11
15	Rose festival	114	0.014	20
16	Jukseoru	111	0.014	20
17	Hotel	102	0.016	16
18	Deokhangsan	101	0.017	15
19	Summer	91	0.015	17
20	Sokcho	88	0.014	20
21	Samcheok beach	88	0.012	25
22	Pension	86	0.013	23
23	Dootasan	76	0.011	27
24	Sunrise	73	0.015	17
25	Taebaek	73	0.009	33
26	Candlestick rock	70	0.011	27
27	Summer vacation	69	0.012	25
28	Santorini	66	0.013	23
29	Jeongdongjin	65	0.015	17
30	Yangyang	64	0.011	27
31	Samcheok jungang market	61	0.009	33
32	Deokpung valley	59	0.011	27
33	New millennium coast road	55	0.009	33
34	Cafe	55	0.007	41
35	Tracking	51	0.011	27
36	Winter sea	51	0.008	37
37	Daegum cave	51	0.007	41
38	Samcheok coast	51	0.005	47
39	chicken ribs	48	0.008	37
40	Wave	48	0.005	47
41	Mugeonli moss waterfall	46	0.008	37
42	Course walking	45	0.011	27
43	Aquaworld	45	0.007	41
44	Snorkeling	40	0.009	33
45	Rape flower festival	40	0.003	54
46	Isabusaja park	39	0.008	37
47	Fishing	37	0.004	50
48	Sashimi	33	0.006	45
49	Samgojeongmun	33	0.003	54
50	Korea Naples	32	0.007	41
51	Rail bike	32	0.006	45
52	Jeongseon	30	0.004	50
53	Jangho beach camping site	29	0.003	54
54	Sea train	28	0.005	47
55	Cherry blossom	28	0.003	54
56	Maengbang beach	28	0.003	54
57	Camping	27	0.004	50
58	Jeongwol daeboreum festival	26	0.002	61
59	Builmakguksu	22	0.004	50
60	Mureung valley	21	0.003	54
61	Mulhoe	18	0.003	54
62	Ilmidam	17	0.002	61
63	SamcheokJeonboghaemulTtugbaegi	17	0.002	61
64	Transparent canoe	14	0.002	61
65	Gangneung bulgogi	14	0.002	61

표 4는 2019년 동시 출현 키워드 빈도수와 연결 중심성 결과이다. 즉, 소셜미디어에서 삼척에 대해 언급할 때 빈도수는 ‘삼척솔비치’, ‘삼척여행’, ‘동해’, ‘장호항’, ‘삼척맛집’, ‘강릉’, ‘바다’, ‘가족여행’, ‘해파랑길’, ‘환선굴’, ‘산행’, ‘해양레일바이크’, ‘해상케이블카’ 순으로 나타났다. 연결 중심성은 ‘삼척솔비치’, ‘삼척여행’, ‘장호항’, ‘동해’, ‘강릉’, ‘삼척맛집’, ‘바다’, ‘가족여행’, ‘해파랑길’, ‘환선굴’, ‘삼척구간’, ‘해상케이블카’, ‘해양레일바이크’, ‘산행’, ‘덕항산’ 순으로 나타났다. 2016년과 마찬가지로 여전히 ‘삼척솔비치’가 지역 이미지 형성에 큰 역할을 하고 있고, ‘가족여행’과 ‘해파랑길’이 지역 이미지 형성에 대한 역할이 증가하였다. 즉, 가족과 함께 맛집 탐방, 해파랑길 코스 걷기 등을 즐길 수 있는 여행지로 지역 이미지가 형성되었다고 볼 수 있다.

#### 4.2 CONCOR 분석

CONCOR 분석은 전체 키워드를 중심으로 키워드 간의 유사성을 측정하여 패턴이 유사한 키워드들의 상관관계에 따라 군집화하는 분석방법이다. 본 논문에서는 소셜 네트워크 분석 전문 프로그램인 UCINET을 사용하여 연결 중심성 분석 및 군집 분석인 CONCOR를 수행하고, Net Draw를 이용하여 시각화하였다.

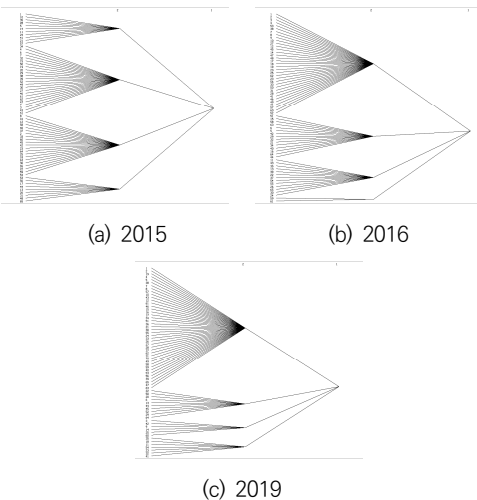


그림 1. 키워드의 덴드로그램  
Fig. 1. Dendrogram of keyword

그림 1은 2015년, 2016년, 2019년의 CONCOR 분석에 따른 군집모형이 모두 Depth 2의 분류에서 동일한 구조적 등위성을 보이고 있다. 따라서 Depth 2에서 4개의 군집모형으로 키워드 간의 관계를 분석할 수 있다.

그림 2는 2015년에 대한 CONCOR 분석의 결과를 Net Draw를 이용하여 시각화한 결과이다.

2015년 CONCOR 분석 결과, 우측상단의 '투명 카누', '스노쿨링', '낚시', '삼척맛집', '삼척해수욕장', '바다' 등은 해양스포츠와 해수욕장을 집중적으로 연결한 군집, 우측하단의 '무건리이끼폭포', '덕풍

계곡', '해파랑길' 등은 산행과 트레킹 중심으로 연결한 군집, 좌측하단의 '동해', '정동진' '환선굴', '해양레일바이크' 등은 인근 지역 관광지와 연결한 군집, 좌측상단의 '장호항', '새천년도로', '삼척해변', '강릉', '속초', '양양', '정선' 등 타지역 지명과 연결한 군집으로 연결망 구조를 형성하고 있다.

2015년 전체적인 삼척의 이미지 형성 구조는 해수욕장, 동해, 강릉과의 연계가 가장 큰 영향을 미치는 요소이며, 산행과 트레킹에 관한 요소가 지역 이미지 형성 구조에 영향을 미칠 수 있는 잠재성을 가진 것으로 해석할 수 있다.

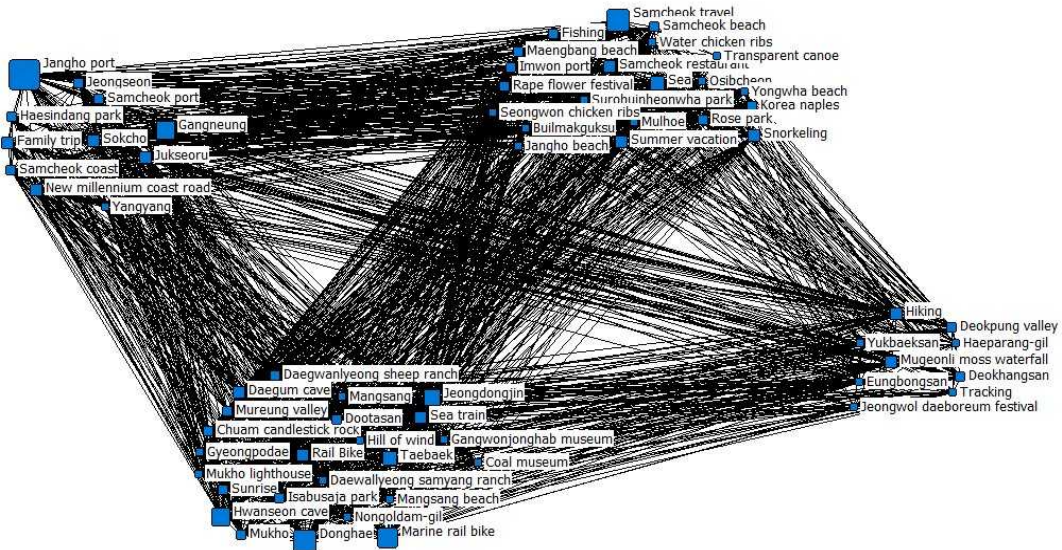


그림 2. CONCOR 분석의 시각화 결과 - 2015년  
Fig. 2. The visualization result of CONCOR analysis - 2015

그림 3은 2016년에 대한 CONCOR 분석의 결과를 Net Draw를 이용하여 시각화한 결과이다.

2016년 CONCOR 분석 결과, 좌측하단의 '삼척솔비치', '삼척여행', '장호항', '삼척맛집', '아쿠아월드', '삼척해수욕장', '낚시' 등은 리조트와 맛집, 지역 관광이 연결된 군집으로, 빈도수가 높은 키워드들로 구성됨에 따라 대표할 수 있는 군집이라고 볼 수 있다. 좌측상단의 '동해', '추암촛대바위', '정동진', '해돋이', '강릉', '대관령양떼마을' 등은 주변 도시 관광지들로 연결된 군집, 우측상단의 '무건리이끼계곡',

'무릉계곡', '두타산', '쉬움산' 등은 산행과 트레킹으로 연결된 군집, 우측하단의 '기줄다리기', '유네스코'는 지역을 대표하는 정월대보름행사로 연결한 군집으로 연결망을 형성하고 있다.

2016년 전체적인 삼척의 이미지 형성에는 리조트인 삼척솔비치, 근처 맛집과 지역 관광과 연계가 가장 큰 영향을 미치는 요소이며, 산행, 트레킹과 인근 도시와의 연계에 관한 요소가 지역 이미지 형성 구조에 영향을 미칠 수 있는 잠재성을 가진 것으로 해석할 수 있다.

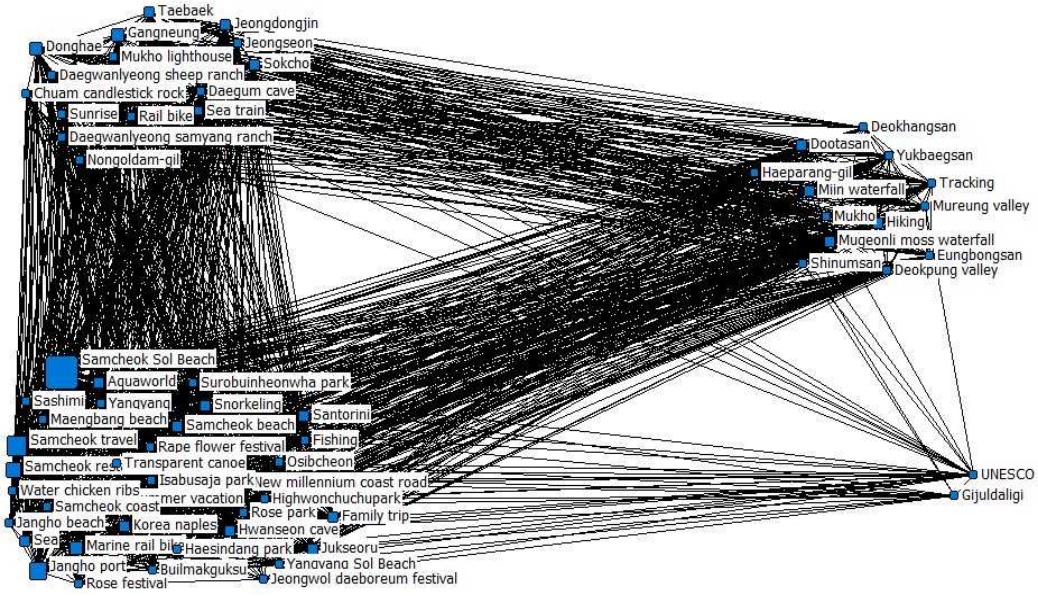


그림 3. CONCOR 분석 시각화 결과 - 2016년  
 Fig. 3. The visualization result of CONCOR analysis -2016

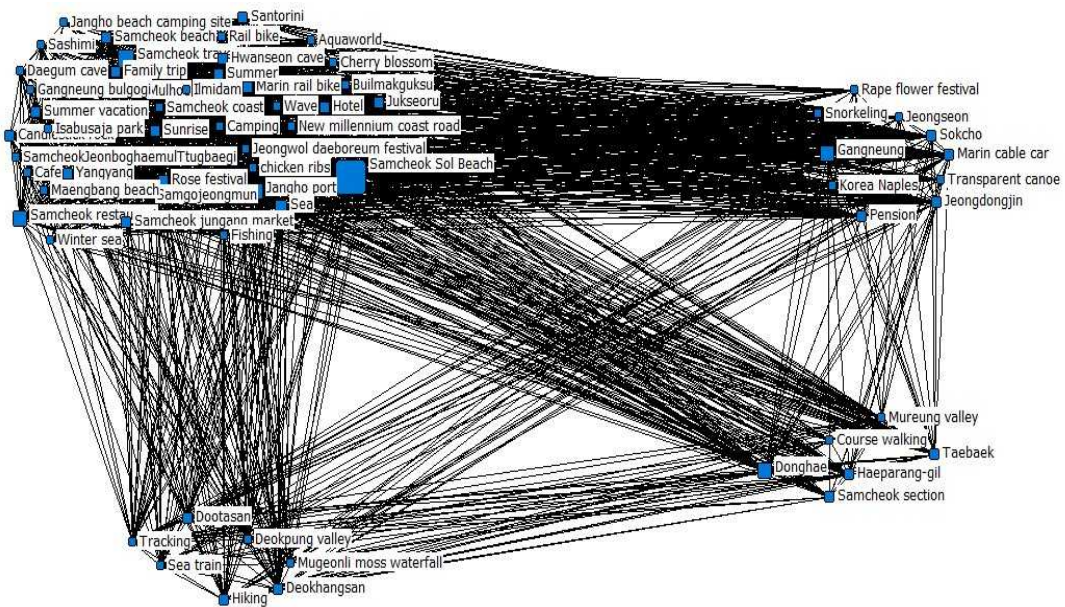


그림 4. CONCOR 분석 시각화 결과 - 2019년  
 Fig. 4. The visualization result of CONCOR analysis - 2019

그림 4는 2019년에 대한 CONCOR 분석의 결과를 Net Draw를 이용하여 시각화한 결과이다.

2019년 CONCOR 분석 결과, 좌측상단의 ‘삼척솔비치’, ‘삼척여행’, ‘삼척맛집’, ‘장호항’, ‘가족여행’, ‘캠핑’ 등은 리조트와 가족과 같이 즐길 수 있는 장소를 연결한 군집으로 빈도수가 높은 키워드들로 구성된 대표할 수 있는 군집으로 볼 수 있다. 좌측하단의 ‘무견리이끼폭포’, ‘덕풍계곡’, ‘두타산’ 등은 산행과 트레킹을 연결한 군집, 우측하단의 ‘동해’, ‘무릉계곡’, ‘해파랑길’, ‘삼척구간’, ‘코스걷기’는 동해와 삼척을 연결하는 군집, 우측상단의 ‘강릉’, ‘정동진’, ‘정선’, ‘속초’, ‘한국나폴리’, ‘투명카누’ 등은 주변 도시와 삼척을 연결한 군집으로 연결망을 형성하고 있다.

2019년 전체적인 삼척의 이미지 형성에는 리조트인 삼척솔비치와 가족과 같이 즐길 수 있는 장소 연계가 가장 큰 영향을 미치는 요소이며, 2016년과 마찬가지로 산행, 트레킹, 인근 도시와의 연계에 관한 요소가 지역 이미지 형성 구조에 영향을 미칠 수 있는 잠재성을 가진 것으로 해석할 수 있다.

### 4.3 유의성 검증

본 연구의 유의성 검증은 부트스트랩 기법을 이용하였다. 부트스트랩 기법에서는 귀무가설을 ‘데이터 모집단의 밀도가 0이다’라고 가정을 하고, 연결망의 노드들을 5000번 무작위 재배열하여 네트워크 밀도를 구한다. 네트워크 밀도에 대한 표본밀도 평균, 추정된 표준오차, 그리고 Z-score 값을 구하고, Z-score에 대한 p-value로 귀무가설의 기각 여부로 가설검증은 진행하였다.

1) 가설 1 : 2015년 데이터 모집단의 밀도가 0이다.

그림 5와 같이 가설 1의 2015년 데이터의 유의성 검증결과에 따라 표본밀도 평균 31.6083, 추정된 표준오차 5.7241, Z-score 5.3577일 때, p-value는 0.0002로 유의수준 0.05 이하이므로 통계적으로 유의하다고 판별할 수 있다.

COMPARE DENSITY W/ HYPOTHESIZED VALUE	
Random seed:	19482
Parameter value is:	0.0000
Density of 15 is:	30.6712
Difference is:	30.6712
Variance of ties for 15:	5634.2703
Classical estimate of SE:	1.1638
Number of bootstrap samples:	5000
Estimated standard error for density of 15:	5.7241
z-score:	5.3577
Average bootstrap density:	31.6083
Proportion of absolute differences as large as observed:	0.0002
Proportion of differences as large as observed:	0.0002
Proportion of differences as small as observed:	1.0000
Running time:	00:00:04
Output generated:	11 2 22 01:46:50
UCINET 6.738 Copyright (c) 2002-2021 Analytic Technologies	

그림 5. 가설1의 유의성 검증결과  
Fig. 5. The Significant test result of Hypothesis 1

2) 가설 2 : 2016년 데이터 모집단의 밀도가 0이다.

그림 6과 같이 가설 2의 2016년 데이터의 유의성 검증결과에 따라 표본밀도 평균 15.8244, 추정된 표준오차 3.5391, Z-score 4.3576일 때, p-value는 0.0006으로 유의수준 0.05 이하이므로 통계적으로 유의하다고 판별할 수 있다.

COMPARE DENSITY W/ HYPOTHESIZED VALUE	
Random seed:	19482
Parameter value is:	0.0000
Density of 16 is:	15.4236
Difference is:	15.4236
Variance of ties for 16:	3313.9340
Classical estimate of SE:	0.8925
Number of bootstrap samples:	5000
Estimated standard error for density of 16:	3.5391
z-score:	4.3576
Average bootstrap density:	15.8244
Proportion of absolute differences as large as observed:	0.0006
Proportion of differences as large as observed:	0.0006
Proportion of differences as small as observed:	0.9996
Running time:	00:00:04
Output generated:	11 2 22 01:49:15
UCINET 6.738 Copyright (c) 2002-2021 Analytic Technologies	

그림 6. 가설 2의 유의성 검증결과  
Fig. 6. The significant test result of Hypothesis 2



3) 가설 3 : 2019년 데이터 모집단의 밀도가 0이다.

```

COMPARE DENSITY W/ HYPOTHESIZED VALUE
-----
Random seed:                19482

Parameter value is: 0.0000
Density of 19 is: 10.5788
Difference is: 10.5788
Variance of ties for 19: 1126.4251
Classical estimate of SE: 0.5204

Number of bootstrap samples: 5000
Estimated standard error for density of 19: 2.2618
z-score: 4.6767
Average bootstrap density: 10.8653
Proportion of absolute differences as large as observed: 0.0002
Proportion of differences as large as observed: 0.0002
Proportion of differences as small as observed: 1.0000

-----
Running time: 00:00:04
Output generated: 12 2 22 00:34:08
UCINET 6.738 Copyright (c) 2002-2021 Analytic Technologies
    
```

그림 7. 가설 3의 유의성 검증결과  
Fig. 7. The significant test result of Hypothesis 3

그림 7과 같이 가설 3의 2019년 데이터의 유의성 검증결과에 따라 표본밀도 평균 10.8653, 추정된 표준오차 2.2618, Z-score 4.6767일 때, p-value는 0.0002로 유의수준 0.05 이하이므로 통계적으로 유의하다고 판별할 수 있다.

따라서 2015년, 2016년, 2019년 데이터를 이용한 네트워크 분석에는 문제가 없음을 확인하였다.

### 5. 결론

초연결사회인 현재, 인터넷, 스마트 폰 등을 이용해 개인 의견을 소셜미디어에 제한 없이 표현하고 있다. 따라서 소셜미디어 빅데이터에는 다양하면서 실질적인 데이터가 포함되어 있다고 볼 수 있다.

본 논문에서는 TEXTOM을 이용해 네이버와 다음으로부터 검색 키워드 '삼척'으로 지역에 관련된 자료를 리조트 운영전인 2015년, 운영을 시작한 2016년부터 2019년까지 1년 단위로 수집하였다. 텍스트 마이닝 기법을 통해 수집된 자료로부터 키워드를 추출한 후, 키워드 분석 결과 2017년부터 2019년까지의 분석 결과가 유사하여 2017년과 2018년을 제외한 2015년, 2016년, 2019년 키워드로 지역 이미지

의 변화를 도출하였다.

빈도수가 높은 키워드 65개 분석 결과, 2015년에는 '장호항(953)', '삼척여행(634)', '동해(629)', '해양레일바이크(493)', '강릉(462)', '환선굴(404)', '정동진(348)', '바다(292)', '태백(242)', '속초(196)' 순으로 나타났고, 2016년에는 '삼척솔비치(1554)', '삼척여행(689)', '장호항(554)', '삼척맛집(449)', '동해(360)', '강릉(266)', '해양레일바이크(265)', '바다(221)', '산토리니(186)', '환선굴(170)' 순으로 나타났다. 그리고 2019년에는 '삼척솔비치(1046)', '삼척여행(417)', '동해(318)', '장호항(308)', '삼척맛집(307)', '강릉(256)', '바다(233)', '가족여행(153)', '환선굴(141)', '해파랑길(141)' 순으로 나타났다.

키워드 분석 결과, 리조트 운영 이전인 2015년에는 삼척을 포함해 동해, 강릉에 관련된 키워드들이 다수 포함되어 있었는데, 리조트가 운영되기 시작한 2016년과 2019년에는 리조트 관련 키워드와 삼척 지역 내의 키워드가 증가하였으며, 특히 '장호비치캠핑장', '캠핑'이 2019년 키워드에 포함되어 있다. 따라서 2015년에 비해 2016년부터 숙박업 이용자 수가 급격히 증가한 것과 비교해 볼 때, 2015년에 비해 2016년부터 삼척이 체류형 형태의 여행지로 지역 이미지가 변경되는 것으로 볼 수 있다.

CONCOR 분석 결과, 지역 이미지는 2015년 해수욕장, 동해를 중심으로 하는 군집이 이미지 형성에 큰 영향을 주었는데, 2016년과 2019년은 삼척솔비치를 중심으로 하는 군집이 이미지 형성에 큰 영향을 주는 것으로 나타났다. 지역 이미지와 연관된 키워드들이 집중된 군집에는 삼척을 대표하는 명소인 장호항, 리조트가 포함하고 있는 것을 보아 지역 이미지 형성에 인프라 시설 요소가 큰 영향을 미치는 것으로 볼 수 있다. 또한, 네트워크 데이터에 대한 유의성 검증은 부트스트랩 기법을 이용하였고, 2015년, 2016년, 2019년 p-value가 각각 0.0002, 0.0006, 0.0002로 유의수준 5%에서 전체 네트워크 데이터들 간의 관계는 통계적으로 유의한 것으로 나타났다.

## REFERENCES

- [1] Samcheok-si, 2016 STATISTICAL YEARBOOK OF SAMCHEOK, Jun. 2016
- [2] Samcheok-si, 2017 STATISTICAL YEARBOOK OF SAMCHEOK, Jun. 2017
- [3] J. H. Im, Y. H. Kim, "Examining Public Opinion on Iksan Tourism Using Social Media Analytics", Tourism Research, vol.45, no.3, pp.427-441, Sept. 2020.
- [4] T. C. Jin, E. H. Jeong, "Big Data Analysis of Social Media on Gangwon-do Tourism", Journal of Korea Institute of Information, Electronics, and Communication Technology, vol.14, no.3, pp.193-200, Jun. 2021.
- [5] Y. I. Yoon, D. H. Ha, "How Business Hotel is Perceived with Big Data Analytics of Social Media", Korean Business Education Review, vol.32, no.3, pp.379-379, Jun. 2017
- [6] U. S. Hwang, J. K. Min, "The Analysis of Consumption Trend of Tourists to Resorts by Big Data Application", Journal of Hotel & Resort, vol.16, no.2, pp.5-26, May. 2017.
- [7] W. S. Choi, "A Study on Awareness of Integrated Resort Using Big Data Analytics of Social Media", Tourism Research, vol.44, no.4, pp.355-371, Dec. 2019.
- [8] J. H. Im, Y. H. Kim, "Examining Public Opinion on Iksan Tourism Using Social Media Analytics", Tourism Research, vol.45, no.3, pp.427-441, Sept. 2020.
- [9] Textorm <http://textom.co.kr/home/main/main.php>

---

## 저자약력

---

정은희(Bun-Hee Jeong)

[중신회원]



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

〈관심분야〉

전자상거래 보안, 빅데이터 분석, 텍스트 마이닝, 계량경제, 경제통계