

# 지역신문기사 자료와 토픽모델링을 이용한 해변 관련 계절별 현안분석\*

유무상\*\* · 정수연\*\*\* · 김건후\*\*\*\* · 손철\*\*\*\*\*

## Seasonal analysis of Beach-related Issues using Local Newspaper Articles and Topic Modeling\*

Mu-Sang Yoo\*\*, Su-Yeon Jeong\*\*\*, Geon-Hu Kim\*\*\*\*, Chul Sohn\*\*\*\*\*

**국문요약** 본 연구의 목적은 2004년부터 2017년까지의 해변과 해수욕장을 키워드로 하는 지역신문 기사를 이용하여 계절별 현안을 분석하는 것이다. 분석을 위해 오픈소스 프로그램을 기반으로 한 토픽모델링과 시계열회귀분석을 수행하였다. 토픽모델링 분석 결과 계절별 토픽은 봄 35개, 여름 47개, 가을 36개, 겨울 35개가 도출되었다. 모든 계절에서 공통적으로 도출된 주제는 해수욕장, 축제·행사, 사건사고 및 환경문제, 관광지, 개발·분양, 행정·정책, 날씨로 나타났다. 시계열회귀분석 결과 봄에는 35개의 토픽 중 5개의 상승 토픽과 2개의 하락 토픽이 도출되었다. 여름에는 47개의 토픽 중 6개의 상승 토픽과 3개의 하락 토픽이 도출되었다. 가을에는 36개의 토픽 중 4개의 상승 토픽과 3개의 하락 토픽이 도출되었다. 겨울에는 35개의 토픽 중 3개의 상승 토픽과 3개의 하락 토픽이 도출되었다. 그리고 각 계절별로 상승 토픽과 하락 토픽에 해당하지 않는 토픽은 중립 토픽으로 구분하였다. 본 연구를 통해 해변과 같이 계절별로 용도가 다른 경우에 지역현안에 대한 분석을 위해 계절별 토픽모델링을 진행한다면 더욱 유용한 결과를 도출하고 이에 따른 세부적인 진단이 가능하다고 판단된다.

**주제어** 토픽모델링, 시계열회귀분석, 해수욕장, 해변, 지역신문

**Abstract:** The purpose of this study is to analyze the seasonal issues using the local newspaper articles with the keyword beach from 2004 to 2017. Topic modeling and Time series regression analysis based on open source programs were performed for analysis. Topic modeling results showed 35 topics in spring, 47 topics in summer, 36 topics in autumn and 35 topics in winter. The common themes were 'beaches', 'festivals and events', 'accident and environmental

\* 본 연구는 해양수산부가 지원하는 강원씨그랜트사업에 의해 이루어진 것이며, 2018년 한국지역학회 전기 학술대회에서 발표한 논문을 수정·보완한 것임.

\*\* 강릉원주대학교 공간정보협동과정 석사과정(주저자: yms7y2@nate.com)

\*\*\* 강릉원주대학교 지역개발학과 석사수료(공동저자: cobutta@gwnu.ac.kr)

\*\*\*\* 강릉원주대학교 공간정보협동과정 박사수료(공동저자: plan3308@gmail.com)

\*\*\*\*\* 강릉원주대학교 도시계획부동산학과 교수(교신저자: csohn@gwnu.ac.kr)

issues', 'tourism', 'development and sale', 'administration and policy' and 'weather'. Time series regression analysis showed in the spring, 5 Hot-Topics and 2 Cold-Topic were found out of the 35 topics. In the summer, 6 Hot-Topics and 3 Cold-Topic were found out of the 47 topics. In the autumn, 4 Hot-Topics and 3 Cold-Topic were found out of the 36 topics. In the winter, 3 Hot-Topics and 3 Cold-Topic were found out of the 35 topics. And for each season, topics that do not fall into the Hot-Topic and Cold-Topic are classified as Neutral-Topic. In this study if seasonal uses are different such as beaches are deemed that seasonal topic modeling for analysis of regional issues will yield more useful results and enable detailed diagnosis.

**Key Words:** Topic Modeling, Time Series Regression, Beaches, Local News Paper

## 1. 연구의 배경 및 목적

해변 및 해수욕장(이하 해변이라 한다)은 아름다운 경관과 함께 다양한 레저스포츠 기회를 제공하는 공간이다. 해변이 제공하는 경관과 레저스포츠는 해당 자치단체가 가진 주요한 관광자원이기 때문에 각 자치단체는 해변을 매력 있는 장소로 유지하기 위해 지속적으로 관리해야 한다.

해변을 가진 자치단체가 해당지역을 효과적으로 관리하기 위해 필요한 것은 해변과 관련한 현안을 지속적으로 발굴하고 대응방안을 고안하는 것이다. 현안 발굴은 설문조사나 해변 이용자를 포함한 다양한 이해관계자의 의견수렴 등을 통해 이루어질 수 있다.

그러나 설문조사나 의견수렴과 같은 현안분석방법은 기획하는 담당자의 가치판단과 사안의 경중에 대한 주관적 판단에 따라 그 범위가 좁혀질 수 있다는 단점이 있다. 반면 최근 문헌 빅데이터를 이용한 텍스트마이닝 방법은 분석자의 주관을 최소화하면서 방대한 문헌을 분석하여 문헌들의 내용을 효과적으로 분석할 수 있는 기술적 환경을 제공한다.

텍스트마이닝 방법 가운데 하나인 토픽모델링(Topic Modeling)은 비정형 문서 내에서 주제를 도출하기 위하여 사용되는 방법론이다. 토픽모델링에 사용되는 문서는 블로그, 논문, SNS, 뉴스 데이터 등 다양하다(김창식 등, 2017).

본 연구에서는 텍스트마이닝 방법 가운데 하나인

토픽모델링으로 지역신문에 적용하여 강원도 지역 해변에 대해 계절별로 부각되고 있는 현안을 분석하고 이를 통해 텍스트마이닝을 통한 해변 관련 계절별 현안분석 가능성을 평가하고자 한다.

## 2. 선행연구

비정형 텍스트 데이터의 분석은 텍스트마이닝 기법을 통해 이루어질 수 있으며 이 기법 중 하나인 토픽모델링을 활용한 연구는 다양한 분야에서 활발히 진행되어 왔다. 또한 시계열회귀분석은 토픽모델링 결과를 시간의 흐름에 따라 시각적으로 표현할 수 있는 대표적인 분석기법으로 토픽모델링과 함께 활용되어 왔다(윤혜진 등, 2017). 최근 국내외에서 텍스트 데이터의 분석을 위해 토픽모델링을 사용한 연구를 살펴보면 다음과 같다.

김창식 등(2017)은 국내에 출판된 정보시스템분야 저널의 연구동향을 조사하기 위하여 2002년부터 2016년까지의 관련 논문 초록 1,245편을 대상으로 토픽모델링과 시계열회귀분석을 수행하였다. 분석 결과 도출된 20개의 토픽 중에서 고객충성도, 소통혁신 등 4개의 상승토픽과 시스템구축 등 2개의 하락토픽을 도출하였다.

김태경 등(2015)은 토픽모델링을 이용하여 인터넷과 모바일 환경에서 금융과 IT가 융합된 핀테크(Fin-

tech) 세부 기술에 대해 Hot topic과 Cold topic을 이용하여 유망기술과 쇠퇴기술을 도출하였다. 분석 결과 추출한 20개의 토픽 중에서 모바일결제, 인증/보안 등 5개의 Hot topic과 주택담보대출, 신용카드결제 등 5개의 Cold topic을 도출하였다.

박상연(2016)은 방송 미디어와 소셜 미디어간의 연관성을 알아보기 위하여 인기 드라마 5개와 이를 대상으로 작성된 네이버 블로그를 수집하여 토픽모델링 분석을 통한 콘텐츠 분석을 수행하였다. 분석 결과 10개의 토픽을 도출하였고 이 토픽들은 시청률과의 상관관계가 있으며 시청률이 높고 낮음에 따라 바뀌는 토픽의 차이가 존재함을 밝혔다.

박자현 등(2013)은 문헌정보학 주요 4개 학술지(정보관리학회지, 한국문헌정보학회지, 한국도서관·정보학회지, 한국비블리아학회지)의 발표 논문 초록을 수집하고 토픽모델링 분석을 수행하여 국내 문헌정보학 분야의 연구동향을 규명했다. 분석 결과 정보관리학회지는 정보학 관련 연구주제가 도서관 관련 연구 주제보다 많이 출현하였고 한국문헌정보학회지 등 3개의 학회지는 이와 반대되는 결과를 보였다고 밝혔다.

우현지 등(2017)은 제주도를 대상으로 Open API를 이용해 제주도의 GPS 좌표 참조 트윗을 대상으로 토픽모델링을 활용해 트위터 데이터의 지리적 분석을 통한 공간분포 패턴을 분석하였다. 분석 결과 지리적 위치와 트윗 내용은 관련이 있으며 토픽모델링 분석은 지역 분포 특성의 직관적 확인에 있어 유용하게 활용될 수 있음을 보여주었다.

윤혜진 등(2017)은 지역관광에 관한 논문의 초록에 대해 토픽모델링과 시계열회귀분석을 활용하여 지역관광 분야의 토픽을 추출하고 토픽의 연도별 추세분석을 수행하였다. 분석 결과 도출된 토픽은 20개이며 모두 상승추세를 보이는 상승토픽에 해당된다고 밝혔다.

Chandelier et al.(2018)은 프랑스에서의 맹수(늑대)의 출현과 이로 인해 벌어지는 지역 거주민과 맹수의 충돌에 대한 1994-2014년 사이 지역신문과 전국신문의 보도내용을 구조적 토픽모델링(Structural Topic

Modeling)을 통해 분석하였다. 분석 결과는 맹수와 직접적으로 접촉하는 지역의 신문은 이 문제에 대해 인간중심적 입장의 보도를 한 반면 전국단위 신문은 맹수 측면을 고려한 보도에 치중한 것을 보여주었다.

Cheng et al.(2018)은 생태-환경-빈곤과 관련한 연구의 부족을 지적하며 1981년부터 2017년까지의 생태, 환경 및 빈곤을 키워드로 하는 4,335개의 간행물을 수집하였다. 토픽모델링을 통해 생물 다양성(Species diversity)과 경제와 환경개발(Economic and environmental development), 건강위험(Health risk), 도시빈곤(Urban poverty) 등 9개의 주요토픽을 도출하였다. 이를 통해 향후 관련 연구의 의사결정에 대한 이론적 기초를 제공할 수 있다고 밝혔다.

Li et al.(2018)은 토픽모델링을 통해 의미 없는 단어들도 도출되는 경우는 단어의 가중치가 동일하여 발생하는 현상이며 이에 대한 해결책으로 Entropy Weighting(EW)을 통해 보다 의미 있는 단어들도 도출할 수 있다고 밝혔다.

Yang et al.(2011)은 1829-2008년 사이 미국 텍사스 주에 발간된 지역신문의 기사에 대해 토픽모델링을 적용하여 지역신문이 지역의 역사에 대해 가지고 있는 정보를 줄 수 있는지 분석하였다. 분석 결과는 토픽모델링이 자동적으로 추출해낸 토픽들이 지역의 역사를 드러내는데 적절하나 역사가의 추가적 해석 없이는 추출된 토픽들의 상대적 가치를 해석하기 어렵다는 점을 보여주었다.

동 연구들은 연구동향, IT, 관광, 생태환경 등 다양한 영역에서 토픽모델링을 활용한 정보의 추출이 가능한 것을 보여준다. 본 연구에서는 이들이 제시하는 토픽모델링을 활용하여 강원도에 소재한 해변과 관련하여 지역 현안을 파악하고자 한다.

### 3. 분석 대상 및 방법

#### 1) 분석대상

지역신문을 이용하여 해변과 관련한 계절별 현안분석을 수행하기 위해 강원도의 지역신문인 강원도민일

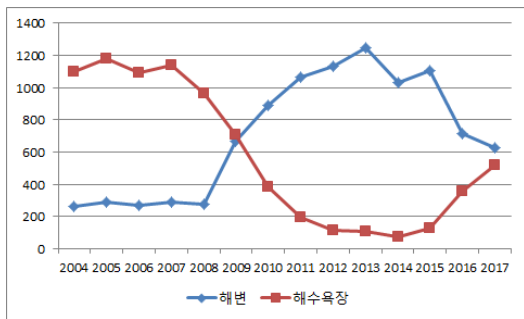
보와 강원일보를 대상으로 2004년부터 2017년까지의 기사 중 해수욕장과 해변을 키워드로 하는 기사를 분석대상으로 선정한다.

해변에 대해 전통적으로 이용해오던 명칭인 해수욕장은 여름철에만 한시적으로 이용된다는 개념이 강하여 점차 그 이용이 축소되고 있다. 국립공원관리공단도 2011년에 한려해상, 태안해안, 다도해해상, 변산반도 등 국립공원에 있는 74개의 해수욕장을 해변으로 명칭 변경을 진행했다(환경부, 2011). 그러나 여전히 해수욕장과 해변의 표기는 혼용이 빈번하다. 2014년 ‘해수욕장의 이용 및 관리에 관한 법률’이 제정·시행되면서 최근에는 해수욕장의 표기가 다시 증가하고 있는 추세다. 해수욕장의 명칭 변경과 법률 제정 등의 이유로 시기별로 단어사용이 달라짐에 따라 본 연구에서는 포괄적인 분석을 위해 키워드를 해변과 해수욕장으로 나누어 분석에 이용하였다. <그림 1>은 해변과 해수욕장으로 검색된 연도별 기사 추세를 나타낸다. 계절의 구분은 3~5월은 봄, 6~8월은 여름,

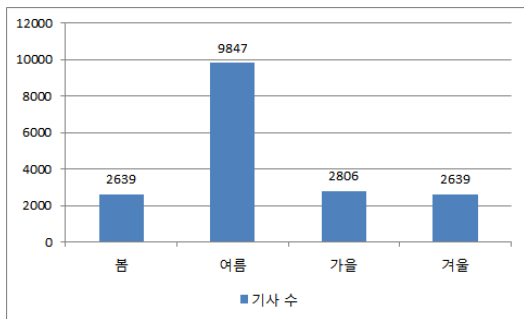
9~11월은 가을, 12~2월은 겨울로 분류하였다. 해변과 해수욕장으로 검색된 기사는 계절적인 영향으로 인해 6~8월에 해당하는 여름이 다른 계절에 비해 약 3배 이상의 높은 기사건수를 기록했다. 계절별 기사수는 봄 2,639건, 여름 9,847건, 가을 2,806건, 겨울 2,639건이다. <그림 2>는 계절별 기사 건수를 나타낸다.

## 2) 분석방법 및 절차

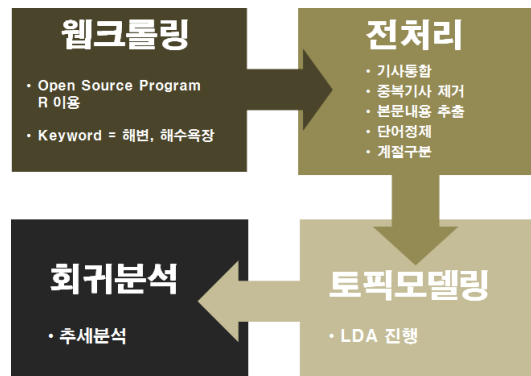
본 연구는 ‘해변’과 ‘해수욕장’을 키워드로 검색한 17,931건의 강원도 지역신문 기사를 대상으로 토픽모델링과 시계열회귀분석을 통해 강원도 지역의 해변과 관련한 지역현안을 계절별로 파악하였다. 분석은 웹크롤링, 전처리, 토픽모델링, 시계열회귀분석 단계로 수행하였다. 첫째, 지역신문 기사 수집을 위한 웹크롤링 작업은 오픈소스 프로그램인 R을 통하여 진행하였다. 둘째, 분석을 위한 전처리 작업으로는 각 언론사별 기사를 하나의 데이터로 통합한 뒤 서로 중복되는 기사를 제거하였다. 그리고 기사 작성 월을 기준으로 계절을 구분하였다. 셋째, 토픽모델링은 웹크롤링과 전처리를 통해 구축한 지역신문 기사 중 본문 내용만 추출하여 단어정제를 진행하였고 토픽모델링을 진행하였다. 넷째, 토픽모델링의 결과를 활용하여 연도별 시계열회귀분석을 진행하였다. 분석과정의 흐름은 <그림 3>과 같다.



<그림 1> 키워드 간 연도별 기사 추세



<그림 2> 계절별 기사 건수



<그림 3> 분석과정의 흐름

## 4. 분석 결과

### 1) 토픽모델링 분석 결과

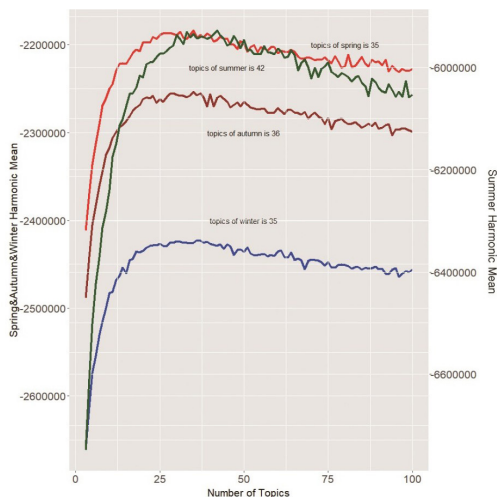
본 연구에서 오픈소스 프로그램인 R을 활용하여 Blei et al.(2003)에 의해 소개된 잠재디리클레할당(Latent Dirichlet Allocation, LDA)모델을 통해 계절별로 구분하여 분석을 수행하였다. 토픽모델링을 위한 R 패키지는 'lubridate', 'tm', 'topicmodels'을 활용하였다.

토픽모델링에 앞서 토픽의 개수는 우선적으로 고려되어 값이 주어져야 한다. 토픽 수가 잘못 추정되면 과적합(Overfitting) 등의 문제로 모형의 객관성이 결여되기 때문에 최적의 토픽 수를 선택하여야 한다(김태경 등, 2016). 토픽 수는 확률분포의 perplexity 값의 변화가 최소가 되는 지점의 값 혹은 토픽 수를 임의로 선정하여 도출된 결과가 주제를 가장 잘 설명하고 있다고 판단되는 값과 우도 값의 harmonic mean이 최대가 되는 값을 사용할 수 있다(이수상, 2016). Ponweiser(2012)는 harmonic mean을 이용한 토픽 수 선정 기법을 최적의 토픽 수 결정 기법 중 하나라고 제안하였다. 본 연구에서는 비교적 단순하고 효율적인 계산방식을 적용한 harmonic mean 값이 최대가 되는 지점을 각 계절별 토픽 수로 선정하였다.<sup>1)</sup> 봄에 해당하는 토픽 수는 35, 여름에 해당하는 토픽 수는 42, 가

을에 해당하는 토픽 수는 36, 겨울에 해당하는 토픽 수는 34로 선정하였다. <그림 4>는 그래프로 나타낸 각 계절별 토픽 수의 선정을 위한 harmonic mean 값을 보여준다.

<표 1>에서 <표 4>는 2004년부터 2017년까지의 지역신문 기사를 계절별로 나누어 분석한 토픽모델링 결과를 나타낸다. 표에 명시된 키워드(Keyword)는 빈도를 기준으로 토픽(Topic)의 설명력을 높일 수 있는 단어들을 저자가 임의 선정하였으며 주제(Theme)는 키워드를 대표할 수 있는 것으로 선정하였다. 계절별로 토픽의 수와 토픽의 주제는 다소 차이가 있으나 공통적으로 해수욕장, 축제·행사, 사건사고 및 환경문제, 관광지, 개발·분양, 행정·정책, 날씨로 나타났다. 계절별로 봄에는 사건사고 및 환경문제와 관광지에 대한 토픽이 다수 도출되었으며 여름에는 해수욕장과 축제·행사에 대한 토픽이 다수 도출되었다. 그리고 가을에는 관광지와 개발·분양에 대한 토픽이 다수 도출되었고 겨울에는 축제·행사와 행정·정책에 대한 토픽이 다수 도출되었다.

<표 5>는 계절별 지역 토픽을 분류하여 살펴본 결과를 나타낸다. 강릉, 고성, 삼척은 사계절 모두 관광지와 관련된 토픽이 도출되었고 동해는 겨울을 제외한 계절에서 축제·행사와 관련된 토픽이 도출되었다. 속초는 봄과 가을에만 관광지와 관련된 토픽이 도출되었고 양양은 사계절에서 관광지, 개발·분양, 축제·행사가 도출되었다. 계절별로 지역 토픽을 분류한 결과 공통적으로 계절과는 상관없이 지역에서는 주로 관광지와 축제·행사에 관련된 토픽이 도출되었다. 이는 관광지와 축제·행사에 관한 내용이 지역별로 세분화됨을 나타낸다.



<그림 4> 계절별 harmonic mean

〈표 1〉 봄(3월~5월) 토픽모델링 결과

No	Theme	Topic	Keyword
T32	해수욕장	운영	개장, 물놀이, 안전관리, 관광객, 영업 ...
T01	축제·행사	삼척	맹방, 유채꽃축제, 벚꽃길, 축제장 ...
T15		구성방안	관광객, 운영, 체험, 계획, 활성화, 유치 ...
T24		동해	축제, 망상, 엑스포, 이벤트, 프로그램 ...
T08	사건사고 및 환경문제	화재	산불, 주민, 진화, 불길, 강풍, 공포, 피해 ...
T09		사건	경찰, 인근, 구조, 사고, 신고, 치료, 안전 ...
T10		자연보존	소나무, 주변, 공간, 식재, 역할, 보호 ...
T11		바다쓰레기	바다, 쓰레기, 바닷속, 포구, 작업 ...
T13		해안침식	침식, 해안, 방파제, 설치, 심각한, 복구, 피해 ...
T14		사건	상품권, 혐의, 가짜, 입건, 수사, 불구속 ...
T17		발전소	주민, 공사, 요구, 화력발전소, 설치, 갈등 ...
T34		철조망	철거, 철책, 군부대, 경계, 펜스, 국방부, 구간 ...
T03	관광지	강릉	경포, 정동진, 바우길, 경포호, 오죽헌 ...
T07		고성	화진포, 송지호, 통일전망대, 해양박물관 ...
T16		촬영지	영화, 드라마, 촬영, 감독, 배우, 시청률 ...
T20		삼척	해양레일바이크, 케이블카, 장호항, 환선굴 ...
T26		동해	동해항, 묵호항, 망상오토캠핑리조트, 무릉계곡 ...
T29		속초	대포항, 외옹치, 영랑호, 케이블카, 아바이마을 ...
T30		삼척, 동해	추암, 촛대바위, 이사부, 대명리조트, 쓸비치 ...
T35		양양	낙산, 하조대, 낙산사, 남대천, 송이, 물치 ...
T18	개발·분양	해양관광개발	사업, 개발, 조성, 추진, 육성, 구축, 투자 ...
T19		강릉 개발	강릉, 사업, 계획, 규모, 개발, 관광지, 추진 ...
T22		해안도로 개발	구간, 사업, 조성, 설치, 국도, 정비, 해안도로 ...
T31		분양	건물, 상가, 인근, 분양, 규모, 공급, 이용 ...
T06	행정·정책	정책제안	문제, 필요, 대책, 변화, 방안, 규제, 역할, 지원 ...
T21		예산심의	의원, 의회, 지원, 예산, 사업, 심의, 지적, 조례 ...
T23		의견	참석, 의원, 시장, 시청, 개최, 회의, 의회, 의장 ...
T25		선거	후보, 일자리, 공약, 활성화, 제시, 유치, 경제 ...
T27	날씨	기온	날씨, 주말, 휴일, 기온, 더위, 초여름, 최고기온 ...
T05	대회	마라톤	황영조, 코스, 참가, 출발, 기록, 우승, 풀코스 ...
T12		아이스하키	북한, 선수단, 숙소, 여자, 훈련, 아이스하키 ...
T33		축구	금강대기, 전국, 청소년, 축구, 스포츠, 경기장 ...
T28	자연경관	경관	마을, 코스, 자연, 호수, 자랑, 다양한, 모습 ...
T02	기타	화합	우리, 세계, 사람, 도시, 시대, 자연, 국민, 미래 ...
T04		지역 구분	부산, 경북, 충남, 전국, 전남, 강원도, 포항 ...

〈표 2〉 여름(6월~8월) 토픽모델링 결과

No	Theme	Topic	Keyword	
T03	해수욕장	범죄	경찰, 혐의, 성범죄, 몰카, 발생, 사건, 신고 ...	
T15		운영	개장, 계획, 편의시설, 설치, 무료, 준비 ...	
T30		단속	요금, 단속, 대상, 실시, 근질, 집중, 불법, 행위 ...	
T31		민원	주민, 철거, 공사, 불편, 불법, 민원, 지적, 문제 ...	
T35		관광객 현황	피서객, 주말, 집계, 지난달, 증가, 피서철, 예약 ...	
T36		단속	음주, 경찰, 행위, 호객행위, 요금, 무질서, 금지 ...	
T37		쓰레기문제	쓰레기, 수거, 청소, 처리, 환경, 관리, 배출 ...	
T02		축제·행사	고성	고성, 축제, 체험, 화진포, 송지호, 조개잡이 ...
T07	속초		설악산, 대포항, 청초호, 음악대향연, 음악축제 ...	
T20	동해		망상, 추암, 엑스포, 지역, 개최, 국제, 크루즈 ...	
T22	동계올림픽		올림픽, 풍창, 평창동계올림픽, 홍보관, 유치 ...	
T33	삼척		삼척, 페스티벌, 공연, 이벤트, 음악회, 하이원 ...	
T38	강릉		공연, 경포, 축제, 페스티벌, 음악회, 예술축전 ...	
T39	동해		망상, 축제, 페스티벌, 행사, 공연, 프로그램 ...	
T40	광복절		일본, 역사, 독도, 전시회, 광복, 태극기 ...	
T42	할인행사		포켓몬, 서비스, 무료, 할인, 이벤트, 상품, 행사 ...	
T43	청소년활동		청소년, 학생, 프로그램, 캠프, 장애인, 봉사활동 ...	
T46	산림박물관		산림, 선정, 우수, 전국, 산림박물관, 해양수산부 ...	
T10	사건사고 및 환경문제		물놀이사고	파도, 구조, 물놀이, 병원, 발견, 구조대, 신고 ...
T16			해온상승	오징어, 수온, 적조, 출현, 어획량, 결과 ...
T19			재난사고	메르스, 환자, 세월호, 확진, 대책, 우려 ...
T26			철조망	철거, 철책, 군부대, 해안, 경계, 협의, 설치 ...
T27			안전사고	물놀이, 안전, 안전사, 안전사고, 예방, 안전요원 ...
T23	관광지	금강산	금강산, 북한, 관광, 전쟁, 비무장지대, 북측 ...	
T25		고성	고성, 화진포, 송지호, 통일전망대, 해양박물관 ...	
T34		강릉	강릉, 경포, 주문진, 정동진, 경포호, 대관령 ...	
T41		삼척	맹방, 이사부, 해양레일바이크, 환선굴, 장호항 ...	
T04		양양	양양, 호텔, 리조트, 콘도, 신축, 객실, 가구 ...	
T13	개발·분양	관광지	사업, 조성, 계획, 추진, 설치, 관광객, 관광지 ...	
T28		해양심층수	사업, 조성, 개발, 일자리, 활성화, 해양심층수 ...	
T06	행정·정책	강릉	강릉, 관광객, 운영, 홍보활동, 계획, 활성화 ...	
T17		갈등	주민, 문제, 예산, 추진, 요구, 협의, 우려 ...	
T18		의견	참석, 개회식, 격려, 개최, 대회의실, 축사 ...	
T21		의견	캠페인, 회원, 참석, 개장식, 참가, 국회의원 ...	
T32		정책도모	관광, 지역, 개발, 변화, 필요, 문제, 방안, 지적 ...	
T45		의견	참석, 의원, 의장, 개최, 의회, 회의실, 격려 ...	
T47		의견	참석, 의원, 의회, 의장, 개최, 격려, 부의장 ...	
T01		날씨	기온	더위, 주말, 무더위, 날씨, 폭염, 기온, 열대야 ...
T05	태풍		태풍, 피해, 파도, 집중호우, 발생, 폭우, 대책 ...	
T09	대회	마라톤	전국, 참가, 경기, 주최, 스포츠, 종목, 일반부, 마라톤 ...	
T24	자연경관	경관	바다, 소나무, 코스, 명소, 절경, 유명, 낭만 ...	
T12	피서	도로상황	차량, 국도, 구간, 고속도로, 정체, 교통 ...	
T14		숙박·편의시설	캠핑장, 예약, 야영장, 텐트, 샤워장, 이용료 ...	
T44		휴가	여름, 가족, 시간, 더위, 피서지, 일상, 여름휴가 ...	
T08	기타	벤처마킹	서핑, 미국, 양양, 투표, 레고랜드, 서피비치 ...	
T11		지역구분	부산, 전국, 충남, 전남, 경북, 전북, 국내 ...	
T29		농촌체험	미울, 농산물, 옥수수, 농특산물, 농촌체험 ...	

〈표 3〉 가을(9월~11월) 토픽모델링 결과

No	Theme	Topic	Keyword
T07	해수욕장	운영	운영, 여름, 선정, 설치, 결과, 우수, 관리, 요금 ...
T08		관광객	관광객, 피서객, 증가, 감소, 기간, 방문객, 분석 ...
T10	축제·행사	봉사활동	쓰레기, 활동, 수거, 봉사활동, 환경정화, 실시 ...
T13		동해	동해, 망상, 엑스포, 개최, 박람회, 전시, 참석 ...
T23		양양	캠핑장, 서평, 힐링, 운영, 송이, 조형물, 양양 ...
T31		고성	축제, 행사, 체험, 고성, 고성명태, 이벤트, 운영 ...
T03	사건사고 및 환경문제	해안침식	파도, 해안, 침식, 대책, 피해, 발생, 지역, 계획 ...
T20		철조망	철거, 철조망, 철책, 군부대, 설치, 주민, 경계 ...
T33		물놀이사고	발견, 파도, 신고, 구조, 물놀이, 사고, 실종 ...
T35		폭행사건	폭행, 혐의, 사건, 피해자, 가해자, 불구속, 기소 ...
T01	관광지	촬영지	영화, 작품, 드라마, 문화, 촬영, 배우, 소설 ...
T02		낙시	낙시, 방파제, 항구, 포인트, 바다낙시, 오징어 ...
T06		북한	북한, 역사, 고구려, 전쟁, 북쪽, 통일공원, 분단 ...
T14		고성	화진포, 송지호, 비무장지대, 파랑길, 통일전망대 ...
T15		강릉	경포, 송정, 바우길, 오죽헌, 강문, 경포호, 허균 ...
T17		삼척	추암, 코스, 바다열차, 촛대바위, 해양레일바이크 ...
T18		속초	설악산, 단풍, 설악문화제, 아바이마을, 대포항 ...
T05		개발·분양	발전소
T19	호텔개발		계획, 조성, 추진, 호텔, 개발, 시설, 부지, 건립 ...
T25	해양심층수		발전, 조성, 사업, 해양심층수, 활성화, 녹색성장 ...
T27	아파트		아파트, 분양, 평형, 수도권, 부동산, 접근성 ...
T29	크루즈	러시아, 크루즈, 동해항, 경제자유구역, 동북아 ...	
T04	행정·정책	갈등	주민, 공사, 업체, 시설, 문제, 방치, 중단, 반대 ...
T12		의전	참석, 의원, 의장, 의회, 격려, 간담회, 방문 ...
T26		우수마을	마을, 주민, 지원, 소득, 우수마을, 농산물, 계획 ...
T30		활성화	동해안, 활성화, 지역, 유치, 전략, 인프라, 연계 ...
T11	날씨	태풍	태풍, 시간, 통제, 바람, 대관령, 운행, 지역 ...
T09	대회	사진공모전	디자인, 수상, 선정, 사진, 금상, 공모전, 풍경 ...
T16		낙시	감성돔, 대회, 상금, 주최, 전국, 낙시대회, 손맛 ...
T36		마라톤	금강산, 마라톤, 통일전망대, 코스, 출발, 구간 ...
T21	기타	교육	학생, 대학, 학교, 여성, 교육, 교사, 총장, 수업 ...
T22		강원도	춘천, 원주, 평창, 정선, 영월, 화천, 횡성 ...
T24		현안	지역, 필요, 문제, 발전, 사회, 변화, 가치, 역할 ...
T28		낙산지구	양양, 낙산사, 낙산지구, 낙산월드, 조산리 ...
T32		관계	엄마, 아빠, 삼촌, 아이, 나라, 보도, 신경, 반응 ...
T34		회상	사람, 우리, 생각, 마음, 모습, 자리, 고향, 기습 ...



〈표 4〉 겨울(12월~2월) 토픽모델링 결과

No	Theme	Topic	Keyword
T17	해수욕장	관광객현황	관광객, 증가, 분석, 감소, 집계, 통계, 피서객 ...
T03	축제·행사	해돋이	행사, 새해, 소망, 희망, 일출, 첫날, 소원, 떡국 ...
T06		올림픽 준비	공연, 실사단, 예정, 평가단, 개막식, 준비 ...
T10		동계올림픽	올림픽, 강릉, 평창, 경기장, 개최, 성공, 참가 ...
T28		동해	망상, 추암, 엑스포, 무릉계곡, 묵호항, 개최 ...
T29		불우이웃	지역, 활동, 노인, 전달, 장학금, 성금, 봉사 ...
T31		동해안 해돋이	새해, 일출, 동해안, 명소, 인파, 경포, 동해 ...
T32		강릉 해돋이	해돋이, 강릉, 정동진, 경포, 새해, 일출, 축제 ...
T35		운영	축제, 대상, 제공, 홍보, 운영, 프로그램, 개최 ...
T02		사건사고 및 환경문제	낙산사 화재
T04	세계		지진, 일본, 미국, 사건, 동영상, 전투, 해일 ...
T05	해상사고		발견, 경찰, 해상, 인근, 발생, 파도, 구조, 실종 ...
T18	철조망		철거, 철책, 철조망, 경계, 설치, 계획, 구간 ...
T21	태안 기름유출		태안군, 자원봉사, 기름유출, 기름띠, 작업, 유출 ...
T33	해안침식		파도, 침식, 해안, 대책, 현상, 피해, 원인, 방지 ...
T09	관광지	삼척	코스, 팔경, 보존, 삼척, 가치, 문화유산, 문화 ...
T11		고성	화진포, 금강산, 송지호, 건봉사, 별장, 대진항 ...
T22		삼척·동해	삼척, 동해, 새천년도로, 장호, 맹방, 대금굴 ...
T24		강릉	정동진, 선교장, 코레일, 오죽헌, 관광, 코스 ...
T26		촬영지	영화, 드라마, 촬영, 작품, 작가, 주인공, 배경 ...
T07	개발·분양	해양심층수	사업, 추진, 조성, 일자리, 해양심층수, 발전 ...
T12		항구개발	동해항, 동해안권, 경제자유구역, 건설, 동북아 ...
T14		호텔개발	사업, 계획, 조성, 개발, 추진, 호텔, 예정, 공사 ...
T34		속초 아파트	아파트, 평형, 분양, 조양동, 대포동, 조양권 ...
T01	행정·정책	의견	의원, 참석, 의회, 의장, 예산, 요구, 방안 ...
T08		발전방향	축제, 관광, 개발, 계획, 활성화, 차별화, 조성 ...
T13		정책제안	지역, 관광, 개발, 변화, 필요, 문제, 역할, 환경 ...
T19		주민의견	주민, 지역, 의견, 협의, 요구, 입장, 우려 ...
T20		개발정책	지역, 관광, 개발, 변화, 역할, 규제, 가치 ...
T25		문제지적	설치, 지적, 계획, 예산, 불법, 대책, 방안, 확보 ...
T27		평가	선정, 우수, 평가, 운영, 결과, 부문, 우수기관 ...
T16		날씨	관광객, 폭설, 인파, 제설작업, 관계자, 구간 ...
T15	기타	자연경관	바다, 겨울, 시간, 모습, 자연, 풍경, 절경, 주변 ...
T23		지명	충남, 원주, 평창, 홍천, 횡성, 영월, 제주, 부산 ...
T30		명칭변경	마을, 명칭, 지명, 이름, 변경, 최근, 주민 ...

〈표 5〉 계절별 지역 토픽 분류

Season	Region	No	Theme
봄	강릉	T03	관광지
		T19	개발·분양
	고성	T07	관광지
	동해	T24	축제·행사
		T26	관광지
	동해, 삼척	T30	관광지
	삼척	T01	축제·행사
		T20	관광지
속초	T29	관광지	
양양	T35	관광지	
여름	강릉	T38	축제·행사
		T34	관광지
		T06	행정·정책
	고성	T02	축제·행사
		T25	관광지
	동해	T20	축제·행사
		T39	축제·행사
	삼척	T33	축제·행사
		T41	관광지
	속초	T07	축제·행사
양양	T04	개발·분양	
가을	강릉	T15	관광지
	고성	T31	축제·행사
		T14	관광지
	동해	T13	축제·행사
	삼척	T17	관광지
	속초	T18	관광지
양양	T23	축제·행사	
겨울	강릉	T32	축제·행사
		T24	관광지
	고성	T11	관광지
	삼척	T09	관광지

2) 시계열회귀분석 결과

연도에 따른 토픽모델링 결과의 변화 추세를 파악하기 위하여 오픈소스 프로그램인 R을 활용하여 시계열회귀분석을 진행하였다. 진행에 앞서 연도 변수는 회귀계수 추정 값의 변동을 줄여 모형의 예측 오류를 개선하기 위해 정규화를 사용하여 값을 재구성하였다. 시계열회귀분석 결과를 통해 각 계절별 토픽의 14년간 연도별 추세를 파악할 수 있었다. 또한 시계

열회귀분석의 유의확률 값이 통계적으로 유의미하며 회귀계수 값이 양수로 나타나는 증가추세의 토픽을 Hot-Topic으로 하고 음수로 나타나는 감소추세의 토픽을 Cold-Topic으로 구분하였다. 그리고 유의확률 값이 통계적으로 유의미하지 않고 증감추세가 무의미한 중립토픽을 Neutral-Topic으로 구분하였다.

〈표 6〉은 봄과 여름의 시계열회귀분석 결과를 나타내고 〈표 7〉은 가을과 겨울의 시계열회귀분석 결과를 나타낸다. 분석 결과 봄에는 해수욕장-운영(T32), 축제·행사-구성방안(T15), 관광지-강릉(T03), 개발·분양-해양관광개발(T18), 날씨-기온(T27), 대회-아이스하키(T12), 축구(T33)가 통계적으로 유의미한 결과를 나타냈으며 여름에는 해수욕장-범죄(T03), 민원(T31), 쓰레기문제(T37), 축제·행사-삼척(T33), 사건사고 및 환경문제-물놀이사고(T10), 재난사고(T19), 개발·분양-양양(T04), 행정·정책-정책도모(T32), 피서-숙박·편의시설(T14)이 통계적으로 유의미한 결과를 나타냈다. 가을에는 축제·행사-양양(T23), 고성(T31), 관광지-낙시(T02), 강릉(T15), 삼척(T17), 행정·정책-의전(T12), 대회-낙시(T16)가 통계적으로 유의미한 결과를 나타냈고 겨울에는 축제·행사-동계올림픽(T10), 사건사고 및 환경문제-세계(T04), 개발·분양-항구개발(T12), 행정·정책-개발정책(T20), 평가(T27), 기타-명칭변경(T30)이 통계적으로 유의미한 결과를 나타냈다.

〈그림 5〉는 연도별 봄의 Hot-Topic을 나타낸다. Hot-Topic은 관광지-강릉(T03), 대회-아이스하키(T12), 축제·행사-구성방안(T15), 날씨-기온(T27), 해수욕장-운영(T32)으로 나타났다. 〈그림 6〉은 연도별 봄의 Clod-Topic을 나타낸다. Cold-Topic은 개발·분양-해양관광개발(T18), 대회-축구(T33)로 나타났다. 〈그림 7〉은 연도별 봄의 Neutral-Topic을 나타내며 Hot-Topic과 Cold-Topic을 제외한 나머지 토픽으로 나타났다. 〈그림 8〉은 연도별 여름의 Hot-Topic을 나타낸다. Hot-Topic은 해수욕장-범죄(T03), 개발·분양-양양(T04), 사건사고 및 환경문제-물놀이사고(T10), 피서-숙박·편의시설(T14), 사건사고 및 환경문제-재난사고(T19), 축제·행사-

〈표 6〉 봄, 여름 시계열회귀분석 결과

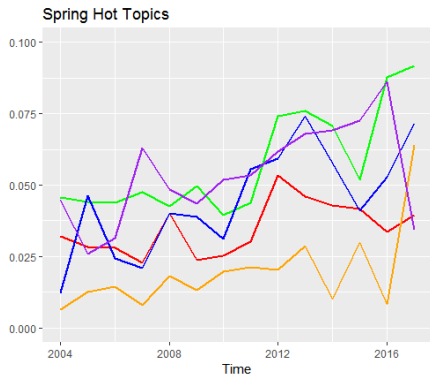
봄				여름			
No	regression coefficient	p-value	Hot/Cold	No	regression coefficient	p-value	Hot/Cold
T01	0.000106	0.971	-	T01	0.006981	0.0613	-
T02	-0.000706	0.593	-	T02	-0.001031	0.631	-
T03	0.004946	0.0483 *	Hot	T03	0.0042815	0.000207 ***	Hot
T04	-0.0004819	0.714	-	T04	0.0028024	0.00613 **	Hot
T05	0.00334	0.99	-	T05	-0.001725	0.342	-
T06	-0.006456	0.0529	-	T06	-0.001083	0.426	-
T07	-0.001409	0.597	-	T07	0.0001617	0.818	-
T08	-0.006183	0.119	-	T08	0.0008003	0.336	-
T09	0.005636	0.166	-	T09	-0.001032	0.682	-
T10	-0.002733	0.255	-	T10	0.005718	0.0104 *	Hot
T11	-0.004290	0.0735	-	T11	0.0004295	0.556	-
T12	0.008534	0.0292 *	Hot	T12	0.002459	0.119	-
T13	0.005436	0.143	-	T13	-0.001060	0.476	-
T14	-0.0001539	0.921	-	T14	0.0029598	0.0111 *	Hot
T15	0.014260	0.000874 ***	Hot	T15	0.0004845	0.764	-
T16	0.006434	0.11	-	T16	-0.0005796	0.552	-
T17	0.004890	0.145	-	T17	-0.0002289	0.902	-
T18	-0.009205	0.039 *	Cold	T18	-0.003477	0.0988	-
T19	-0.001665	0.599	-	T19	0.004307	0.0423 *	Hot
T20	-0.001113	0.501	-	T20	-0.002461	0.123	-
T21	0.002389	0.178	-	T21	0.0006878	0.817	-
T22	0.004080	0.346	-	T22	0.002903	0.153	-
T23	-0.004132	0.172	-	T23	-0.0008837	0.316	-
T24	-0.008454	0.291	-	T24	-0.002052	0.0967	-
T25	-0.007019	0.263	-	T25	0.0002976	0.727	-
T26	-0.005321	0.0648	-	T26	-0.0007517	0.566	-
T27	0.013846	0.00204 **	Hot	T27	0.004227	0.089	-
T28	-0.002340	0.286	-	T28	0.002179	0.0731	-
T29	-0.001192	0.565	-	T29	-0.0013056	0.197	-
T30	-0.001178	0.445	-	T30	0.003403	0.236	-
T31	-0.002885	0.0751	-	T31	-0.003540	0.0178 *	Cold
T32	0.010047	0.0288 *	Hot	T32	-0.003836	0.0382 *	Cold
T33	-0.004969	0.0405 *	Cold	T33	0.004537	0.00351 **	Hot
T34	-0.010669	0.0505	-	T34	-0.00000451	0.997	-
T35	0.001918	0.344	-	T35	-0.007814	0.121	-
				T36	-0.002377	0.35	-
				T37	-0.002744	0.0284 *	Cold
				T38	0.001690	0.34	-
				T39	0.0008039	0.554	-
				T40	-0.0008918	0.254	-
				T41	-0.0007899	0.524	-
				T42	0.0004471	0.876	-
				T43	-0.003949	0.105	-
				T44	-0.002055	0.0888	-
				T45	0.001138	0.729	-
				T46	-0.001710	0.233	-
				T47	-0.006318	0.164	-

〈표 7〉 가을, 겨울 시계열회귀분석 결과

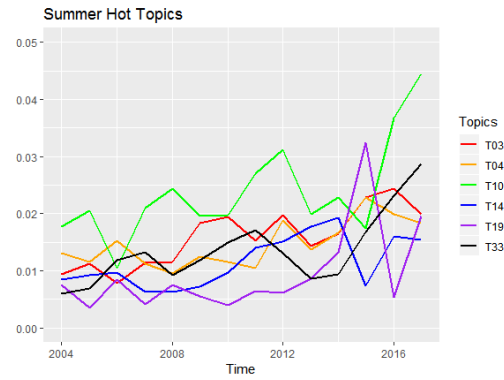
가을				겨울			
No	regression coefficient	p-value	Hot/Cold	No	regression coefficient	p-value	Hot/Cold
T01	-0.0004738	0.806	-	T01	-0.0001701	0.958	-
T02	-0.003588	0.0207 *	Cold	T02	-0.002247	0.167	-
T03	-0.005446	0.385	-	T03	-0.005944	0.244	-
T04	0.001626	0.38	-	T04	-0.006888	0.00862 **	Cold
T05	0.011333	0.0739	-	T05	0.001626	0.682	-
T06	-0.002991	0.0559	-	T06	0.0001419	0.978	-
T07	-0.002418	0.628	-	T07	-0.005906	0.0905	-
T08	-0.005604	0.0836	-	T08	-0.002022	0.37	-
T09	-0.001064	0.48	-	T09	-0.0017568	0.0791	-
T10	0.001728	0.645	-	T10	0.018801	0.0147 *	Hot
T11	0.002397	0.155	-	T11	-0.002700	0.12	-
T12	-0.015420	0.0152 *	Cold	T12	0.005888	0.0197 *	Hot
T13	-0.009393	0.15	-	T13	-0.002613	0.213	-
T14	0.002921	0.355	-	T14	0.009086	0.138	-
T15	0.010490	0.00064 ***	Hot	T15	0.0009355	0.448	-
T16	0.011969	0.00753 **	Hot	T16	0.004098	0.218	-
T17	-0.006973	0.0379 *	Cold	T17	0.005919	0.126	-
T18	-0.003037	0.181	-	T18	-0.006459	0.124	-
T19	-0.005492	0.387	-	T19	-0.001384	0.488	-
T20	-0.003656	0.332	-	T20	0.013681	0.000471 ***	Hot
T21	-0.002942	0.0665	-	T21	-0.007675	0.325	-
T22	0.0012800	0.207	-	T22	0.0003998	0.805	-
T23	0.009398	0.0000901 ***	Hot	T23	-0.001094	0.752	-
T24	-0.002396	0.246	-	T24	0.004049	0.11	-
T25	-0.001706	0.438	-	T25	-0.003238	0.156	-
T26	-0.002468	0.183	-	T26	0.002999	0.37	-
T27	-0.0006053	0.829	-	T27	-0.013317	0.0000389 ***	Cold
T28	-0.003419	0.188	-	T28	-0.005849	0.0858	-
T29	-0.0004056	0.786	-	T29	-0.001985	0.554	-
T30	-0.002163	0.536	-	T30	-0.0038125	0.000939 ***	Cold
T31	0.010037	0.0339 *	Hot	T31	0.004148	0.118	-
T32	0.0005463	0.39	-	T32	-0.005681	0.0832	-
T33	0.010465	0.0766	-	T33	0.008916	0.208	-
T34	-0.0008683	0.788	-	T34	-0.0004812	0.821	-
T35	0.013312	0.174	-	T35	0.0005334	0.799	-
T36	-0.004974	0.234	-				

삼척(T33)으로 나타났다. 〈그림 9〉는 연도별 여름의 Cold-Topic을 나타낸다. Cold-Topic은 해수욕장-민원(T31), 행정·정책-정책도모(T32), 해수욕장-쓰레기문제(T37)로 나타났다. 〈그림 10〉은 연도별 여름의

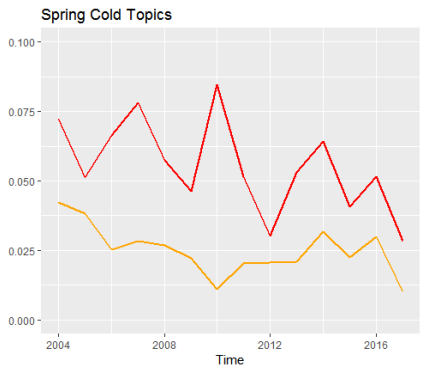
Neutral-Topic을 나타내며 Hot-Topic과 Cold-Topic을 제외한 나머지 토픽으로 나타났다. 〈그림 11〉은 연도별 가을의 Hot-Topic을 나타낸다. Hot-Topic은 관광지-강릉(T15), 대회-뉘시(T16), 축제·행사-



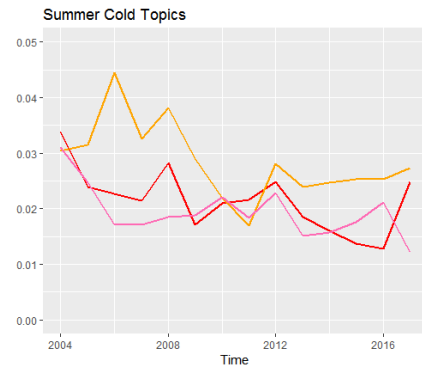
〈그림 5〉 연도별 봄 Hot-Topic



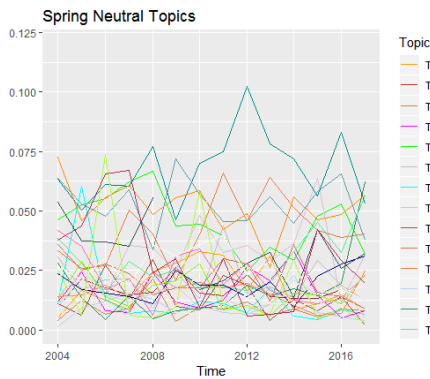
〈그림 8〉 연도별 여름 Hot-Topic



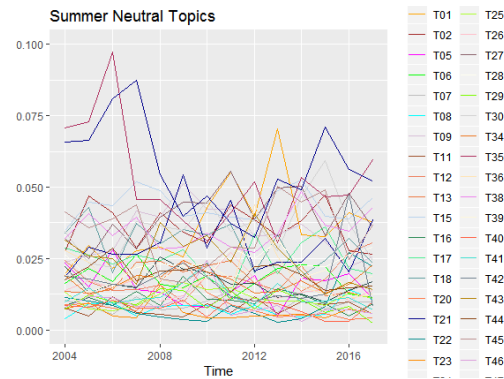
〈그림 6〉 연도별 봄 Cold-Topic



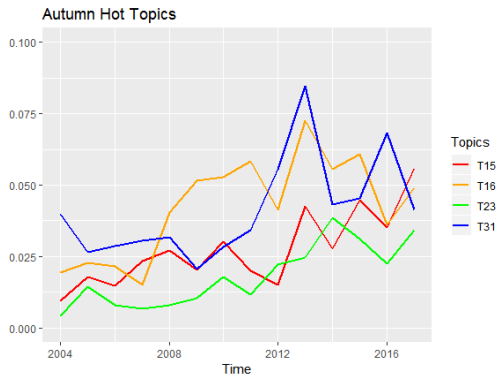
〈그림 9〉 연도별 여름 Cold-Topic



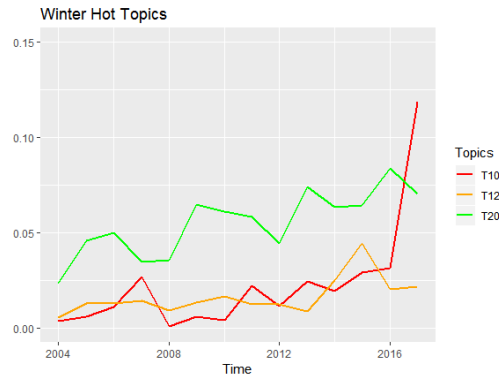
〈그림 7〉 연도별 봄 Neutral-Topic



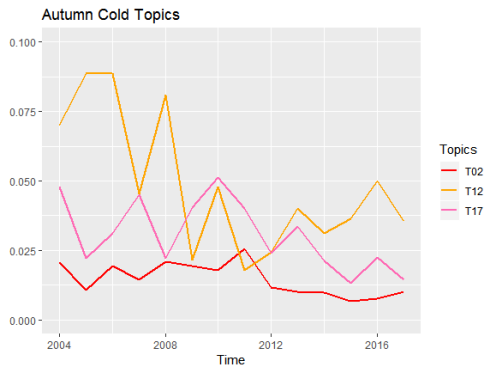
〈그림 10〉 연도별 여름 Neutral-Topic



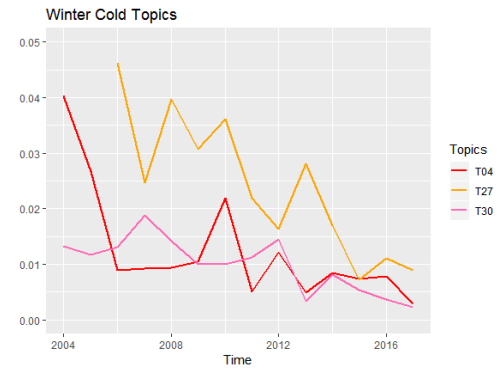
〈그림 11〉 연도별 가을 Hot-Topic



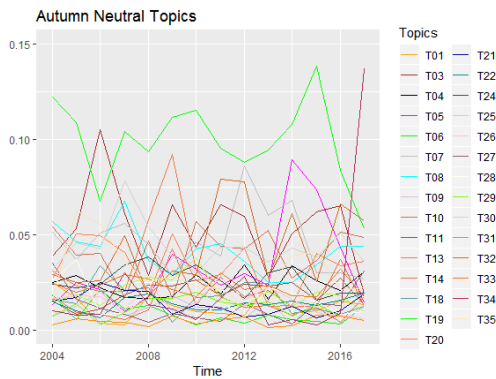
〈그림 14〉 연도별 겨울 Hot-Topic



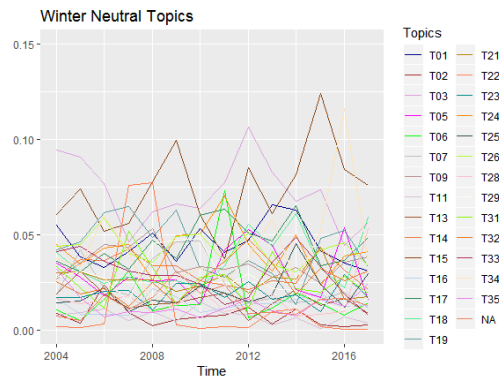
〈그림 12〉 연도별 가을 Cold-Topic



〈그림 15〉 연도별 겨울 Cold-Topic



〈그림 13〉 연도별 가을 Neutral-Topic



〈그림 16〉 연도별 겨울 Neutral-Topic

히 대처해야함을 의미한다.

## 5. 결론

본 연구에서는 2004년부터 2017년까지 총 14년간의 강원도민일보와 강원일보의 기사 17,931건을 대상으로 토픽모델링과 시계열회귀분석 방법을 활용하여 해변과 관련한 계절별 지역현안을 분석하였다.

연구를 통해 도출된 결과는 다음과 같다. 첫째, 오픈소스 프로그램인 R을 활용하여 해수욕장과 해변을 키워드로 하는 17,931건의 기사를 웹크롤링으로 수집하였다. 그리고 수집된 기사 중 본문을 추출하고 작성 월을 기준으로 3~5월은 봄, 6~8월은 여름, 9~11월은 가을, 12~2월은 겨울로 나누어 계절을 구분하고 계절별 토픽모델링과 시계열회귀분석을 진행하였다. 둘째, 토픽모델링을 통해 봄 35개, 여름 47개, 가을 36개, 겨울 35개의 계절별 토픽을 도출하였다. 계절별 공통적으로 도출된 주제는 해수욕장, 축제·행사, 사건사고 및 환경문제, 관광지, 개발·분양, 행정·정책, 날씨로 나타났다. 셋째, 시계열회귀분석 결과 봄에는 35개의 토픽 중 5개의 상승 토픽과 2개의 하락 토픽을 도출하였다. 여름에는 47개의 토픽 중 6개의 상승 토픽과 3개의 하락 토픽을 도출하였다. 가을에는 36개의 토픽 중 4개의 상승 토픽과 3개의 하락 토픽을 도출하였다. 마지막으로 겨울에는 35개의 토픽 중 3개의 상승 토픽과 3개의 하락 토픽을 도출하였다. 각 계절별로 상승 토픽과 하락 토픽에 해당하지 않는 토픽은 중립 토픽으로 구분하였다.

계절별로 세분화하여 진행한 토픽모델링과 시계열회귀분석 결과는 해변과 같이 사계절 용도가 다른 경우에 지역현안을 분석하는데 있어 더욱 유용한 결과를 도출하고 이에 따른 세부적인 진단이 가능하다고 판단된다. 각 지역의 자치단체에서는 계절별로 도출된 Cold-Topic과 Neutral-Topic보다는 Hot-Topic에 대한 대응 및 개선방안을 강구한다면 지역 현안을 보다 원활하게 해결할 수 있을 것으로 판단된다. 또한 설문조사와 의견수렴 등의 질적 현안분석 방법에 비

해 웹크롤링을 이용한 토픽모델링 분석 방법은 경제적·시간적 절약을 기대할 수 있고 좀 더 객관적인 결과를 도출할 수 있다고 판단된다.

본 연구에서는 지역신문기사 자료를 활용하여 계절별 토픽모델링과 시계열회귀분석을 진행하였는데 강원도민일보와 강원일보 2개의 언론사를 활용한 점은 분석 자료의 다양성이 부족하다는 한계점을 가진다. 또한 본 연구에서 활용된 잠재디리클레할당(LDA)은 토픽 수 선정에 따라 분석 결과가 크게 달라진다는 한계가 있다.

향후 연구 진행시 중앙지 혹은 강원도지역의 여타 지역 언론기사, 타 지역의 언론기사와 같은 보다 다양한 지역신문기사 자료를 활용하고 데이터에 따라 적절한 토픽 수가 선정되는 디리클레 프로세스(Dirichlet Process) 혹은 계층적 디리클레 프로세스(Hierarchical Dirichlet Process)를 토픽모델링에 적용한다면 더욱 다양하고 의미 있는 결과를 얻을 수 있을 것으로 판단된다. 또한 토픽모델링 분석 이후 지역 간의 연계 및 주제 간의 연계 등이 가능한 사회연결망분석(Social Network Analysis, SNA)을 활용한다면 좀 더 다양한 분석 결과의 시각적 표현이 가능할 것으로 판단된다.

## 주

- 1) 본 연구에서는 우도 값의  $\text{harmoni mean}$ 이 최대가 되는 지점을 토픽 수로 선정하기 위하여 반복횟수를 1,000회로 설정하였음.

## 참고문헌

- 김창식 등, 2017, 토픽모델링과 시계열회귀분석을 활용한 정보시스템분야 연구동향 분석, 『한국디지털콘텐츠학회 논문지』, 18(6), pp.1143-1150.
- 김태경 등, 2016, 토픽 모델링을 이용한 핀테크 기술 동향 분석, 『한국산학기술학회 논문지』, 17(11), pp.670-681.
- 박상인, 2016, 토픽모델링을 이용한 방송미디어 관련 소셜 미디어 콘텐츠 분석, 『한국지능정보시스템학회 학술대회논문집』, pp.22-22.
- 박자현 등, 2013, 토픽모델링을 활용한 국내 문헌정보학 연구동향 분석, 『정보관리학회지』, 30(1), pp.7-32.

- 우현지 등, 2017, 토픽 모델링을 이용한 트위터 데이터의 공간 분포 패턴 분석, 『한국지역지리학회지』, 23(2), pp.376-387.
- 윤인주 등, 2016, 우리나라 해수욕장 현황과 발전 방향에 관한 연구, 『해양정책연구』, 31(1), pp.115-148.
- 윤혜진 등, 2017, 지역관광 연구트렌드 분석: 텍스트마이닝 활용, 『Paper presented at the pp』, pp.275-282.
- 이수상, 2016, 독후감 텍스트의 토픽모델링 적용에 관한 탐색적 연구, 『한국도서관·정보학회지』, 47(4), pp.1-18.
- 환경부, 2011, 국립공원 내 해수욕장 해변으로 명칭 변경, 보도자료.
- Blei, D.M., et al, 2003, Latent dirichlet allocation, 『Journal of Machine Learning Research』, 3(Jan), pp.993-1022.
- Chandelier, M., 2018, Content analysis of newspaper coverage of wolf recolonization in France using structural topic modeling, 『Biological Conservation』, 220, pp.254-261.
- Cheng, X., et al, 2018, Topic modelling of ecology, environment and poverty nexus: An integrated framework, 『Agriculture, Ecosystems & Environment』, 267, pp.1-14.
- Li, x., et al, 2018, Exploring coherent topics by topic modeling with term weighting, 『Information Processing & Management』, 54, pp.1345-1358.
- Ponweiser, M., 2012, Latent Dirichlet allocation in R (Diploma Thesis), 『Vienna University of Economics and Business, Vienna』.
- Yang, T., et al, 2011, Topic modeling on historical newspapers, Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, 『Social Sciences, and Humanities』, pp.96-104.

계재신청 2018.09.25.

심사일자 2018.09.26.

계재확정 2018.12.05.

주저자: 유무상, 교신저자: 손철