

An Analysis of Effects of Emergency Fine Dust Reduction Measures and National Petition Using Regression Analysis and Text Mining

Annie Kim[†] · So-Hee Jeong[†] · Hyun-Bin Choi[†] · Hyon Hee Kim^{††}

ABSTRACT

Recently, the Seoul government implemented 'Free Public Transportation' policy and 'Citizen Participatory Alternative-Day-No-Driving' system as 'Emergency Fine Dust Reduction Measures'. In this paper, after identifying the effectiveness of the two traffic policies, suggestions for direction of future fine dust policy were made. The effect of traffic on the fine dust was analyzed by regression analysis and the responses to the two traffic policies and petitions were analyzed using text mining. Our experimental results show that the responses to the policy were mostly negative, and the influence of the domestic factors was considerable unlike expectation of citizens. Moreover, the result made us possible to know people's specific needs on fine dust reduction policy. Finally, based on the result, the suggestions for fine dust reduction policy direction were provided.

Keywords : Fine Dust, Transportation, Regression, Text Mining, Topic Modeling

회귀분석과 텍스트마이닝을 활용한 미세먼지 비상저감조치의 실효성 및 국민청원 분석

김애니[†] · 정소희[†] · 최현빈[†] · 김현희^{††}

요약

최근 서울시에서는 '미세먼지 비상저감조치'로 '대중교통 무료' 정책과 '시민 참여형 차량 2부제'를 시행하였다. 본 논문에서는 두 교통정책에 대한 실효성을 파악한 뒤, 향후 미세먼지 정책의 방향을 제시하였다. 교통이 미세먼지에 미치는 영향은 회귀분석으로, 두 정책에 대한 시민들의 반응과 향후 정책에 대한 시민들의 의견은 텍스트마이닝 기법을 통해 알아보았다. 분석 결과, 정책에 대한 시민들의 의견은 대부분 부정적이었고 국외 요인이 미세먼지의 주된 원인이라는 시민들의 생각과 달리 국내 요인의 영향도 상당하였다. 또 국민청원을 통해 시민들이 원하는 구체적인 정책의 내용을 알 수 있었다. 위 결과를 토대로 향후 미세먼지 정책이 나아갈 방향을 제시하였다.

키워드 : 미세먼지, 교통, 회귀분석, 텍스트마이닝, 토픽모델링

1. 서론

최근 미세먼지 문제가 심각해지면서 이로 인해 오염된 대기 질의 개선 및 대기오염 방지를 통해 국민들의 건강을 보호하고자 '미세먼지 비상저감조치'가 발령되었다. 서울시는 '미세먼지 비상저감조치'의 일환으로, 2018년 1월 15, 17, 18일 3일간 '대중교통 무료' 정책을 시행하였다. 하지만 위 정책의

목표인 '미세먼지에 대한 시민 관심 제고', '차량 2부제 법제화'를 상당 부분 달성했다고 판단하여 이를 폐지하였다[1]. 이후 서울시는 2월 27일에 '시민 주도형 미세먼지 비상저감조치 8대 대책'의 하나로 '시민 참여형 차량 2부제'(이하 '차량 2부제')를 발표하였다[2]. 위 미세먼지 관련 정책들은 시민 생활과 맞닿아 있고 시민들의 참여를 적극적으로 유도하기 때문에 본 정책에 대한 관심을 더욱 불러일으켰다.

본 논문은 회귀분석[3]과 텍스트마이닝[4, 5]을 활용해 위 두 교통정책에 대한 시민들의 반응을 알아보고 향후 미세먼지 정책이 나아가야 할 방향을 제시하고자 한다.

먼저, 텍스트마이닝을 통해 위 두 교통정책에 대한 시민들의 반응을 분석해 보았다. 정책 명을 검색어로 하여 5개 주요 언론사의 기사를 정책별로 선정하고 선정한 기사들의 '좋아요', '화나요' 수와 댓글을 수집하였다. 정책별로 댓글에 텍스트마이닝 기법을 적용하여 주요 단어를 추출한 다음 워드 클

* 이 논문은 2016년도 동덕여자대학교 학술연구비 지원에 의하여 수행된 것임.

** 이 논문은 2018년도 한국정보처리학회 춘계학술발표대회에서 '미세먼지 감축을 위해 회귀분석과 텍스트 마이닝을 활용한 교통 정책에 대한 반응 분석'의 제목으로 발표된 논문을 확장한 것임.

† 준회원: 동덕여자대학교 정보통계학과 학사과정

†† 정회원: 동덕여자대학교 정보통계학과 부교수

Manuscript Received: July 6, 2018

First Revision: August 2, 2018

Accepted: August 3, 2018

* Corresponding Author: Hyon Hee Kim(heekim@dongduk.ac.kr)

라우드로 결과를 시각화하였다. 그 결과, 미세먼지 정책에 대한 시민들의 관심과 부정적인 의견이 크게 증가함을 알 수 있었다. 또한 두 정책에 대한 워드 클라우드에서 가장 많이 언급된 ‘중국’이라는 단어를 통해 미세먼지에 미치는 국내 요인의 영향은 미미하고 미세먼지의 주된 원인은 국외에 있다는 시민들의 생각을 알 수 있었다.

시민들이 간과하고 있는 국내 요인의 영향을 확인하기 위해, 위 정책과 관련된 교통을 중심으로 회귀분석을 실시하여 국내 산업과 미세먼지와의 관계를 파악하였다. 미세먼지농도와의 관련성을 높이면 대기오염물질 배출량을 고려해야 한다는 Shin et al.의 연구[6] 결과를 반영하여 배출원 분류 체계를 기반으로 산업을 재분류하고 미세먼지와의 관계를 알아 보았다. 회귀분석의 기법으로는 단순선형회귀와 전진 선택, 후진 제거, 그리고 단계적 변수 선택을 이용하였다. 그 결과, 교통수단이라고 해서 반드시 미세먼지와 양의 관계를 갖는 것은 아니며 도로를 이용하는 교통수단만이 미세먼지에 악영향을 미친다는 사실을 알 수 있었다.

마지막으로 국민들이 원하는 정책의 방향을 파악하기 위해 ‘미세먼지’를 검색어로 하여 국민청원을 추출하고 이를 분석하였다. 추출한 청원에 토픽 모델링을 적용하여 10개의 토픽과 토픽별로 30개의 단어를 추출하고 각 토픽의 주요 단어가 갖는 유사성을 바탕으로 토픽을 크게 네 가지 그룹으로 나누었다. 그룹명은 ‘Petitioner’, ‘Other Countries’, ‘Transportation’, ‘Children’이며 ‘Petitioner’는 현재 미세먼지 정책에 대한 청원인의 일반적인 견해, ‘Other Countries’는 미세먼지를 유발하는 국외 요인의 해결 촉구, ‘Transportation’은 교통수단과 관련한 정책 제안, ‘Children’은 아이들의 건강과 관련한 정책 제안에 대한 내용으로 구성되어 있다. 이를 통해 시민들이 원하는 정책의 구체적인 내용을 알 수 있었다.

본 연구의 공헌은 다양한 분석 결과를 통합하여 교통을 중심으로 ‘미세먼지 비상저감조치’에 대한 국민들의 반응을 분석하고 향후 정책의 방향을 제시한 것이다. 본 연구 결과에 따르면, 향후 정책은 시민들에게 국내 요인의 영향이 상당함을 알리고 보다 적극적인 정책 참여를 유도해야 한다. 또한, 시민들에게 비도요동 교통수단의 이용을 장려하고 도요동 교통수단은 현재보다 친환경적인 교통수단으로 변화해야 한다. 그 외에도 전 세계 국가들의 협력과 정부의 외교적인 노력이 필요하며 정책의 내용에 행동요령과 같이 쉽게 실천할 수 있는 내용이 반영되어야 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 살펴보고 3장에서는 두 정책에 대한 시민들의 반응과 국내 요인의 영향력을 분석한다. 4장에서는 국민청원 내용을 분석하고 마지막 5장에서는 결론 및 정책의 향후 방향을 제시한다.

2. 관련 연구

미세먼지에 미치는 요인은 크게 국내 요인과 국외 요인으로 나뉜다. 국내 요인에 관한 연구로는 수도권 지역에서 기상 조건들이 미세먼지에 미치는 영향 판단을 위해 3차원 대기

모델(TAPM: The Air Pollution Model)을 이용한 분석 연구가 진행되었다[7]. 분석 결과, 수도권 지역의 기상 조건은 안개 형성에 유리하여 미세먼지 고농도 발생이 쉽고 중국의 영향을 받지 않아도 미세먼지 농도가 증가할 수 있다는 사실이 알려졌다.

또한, 타이어/브레이크 마모로 발생하는 미세먼지의 물리화학적 특성 및 배출량 평가와 관련한 연구가 진행되었다[8]. 브레이크 마모에 의한 미세먼지는 차량 속도, 제동 빈도 및 제동력에 의해 발생 특성이 다르며 교차점, 신호등, 코너 근처에서 미세먼지농도가 높게 관찰된다. 또, 향후 타이어/브레이크 마모에 의한 미세먼지 관리 대책을 수립하기 위해서는 이에 대한 명확한 측정법이 필요하다고 보았다.

마지막으로 권역별 식생지수와 미세먼지 발생량의 상관관계를 분석한 연구가 진행되었다[9]. 식생지수는 식생조건을 대표하는 값으로 정규화된 값을 평균하여 사용하였다. 그 결과, 식생지수와 권역별 미세먼지 발생량은 음의 상관관을 보이고 내륙도시가 서해안, 내륙농촌, 동해안보다 식생의 영향을 가장 크게 받는 것으로 나타났다. 따라서 미세먼지 발생량은 식생지수가 클수록 낮아진다는 것을 알 수 있다.

국외 요인에 관한 연구로는 한국의 미세먼지 발생 원인을 한국과 중국의 화석연료 소비로 보고 미세먼지와 실질 GDP와의 관계를 파악한 연구가 있다[10]. 한국과 중국의 실질 GDP가 증가할수록 한국의 미세먼지농도는 현재보다 더 심각해짐을 확인할 수 있었고 미세먼지의 농도변화에 있어 기상 변수를 도입하지 못한 것을 한계점으로 제시하였다.

또한 [11]의 연구는 국내 요인과 중국 요인을 변수로 설정하고 지역별 패널 데이터를 통해 각 변수가 국내 초미세먼지 농도에 미치는 영향을 분석하였다. 국내 요인으로는 경유 소비량, 석탄화력발전 거래량, 화학, 시멘트 제조업 생산지수를, 중국 요인으로는 한국과 거리가 가까운 산둥성 지역의 초미세먼지농도, 월별 서풍계열 풍향일수 비율을 선정하였다. 분석 결과, 중국 산둥성 지역의 초미세먼지농도와 서풍 계열의 풍향이 국내 초미세먼지농도에 양의 영향을 미치는 것으로 확인되었다. 하지만 국내 요인인 경유 소비량과 석탄화력발전 거래량, 화학 산업 제조업 생산지수의 영향은 발견하지 못하였고 시멘트 제조업 생산지수는 초미세먼지농도에 양의 영향을 주는 것으로 나타났다.

위에서 본 기존 연구들 이외에도 미세먼지의 원인 규명에 관한 다양한 연구가 진행 중에 있다. 하지만 아직까지도 미세먼지의 정확한 원인은 밝혀지지 않은 상태이며 미세먼지 저감 정책이 시행되고 있는 최근까지도 미세먼지 고농도 현상이 빈번히 발생하고 있다.

3. 기존 정책의 실효성 분석

3.1 정책에 대한 시민들의 반응 분석

정책 명을 검색어로 하여 주요 언론사별로 기사를 선정하였다. 기사 선정 시, 네이버뉴스에 등록된 기사 중 의견의 편향을 방지하기 위해 댓글이 10개 이상인 기사만을 선정하였

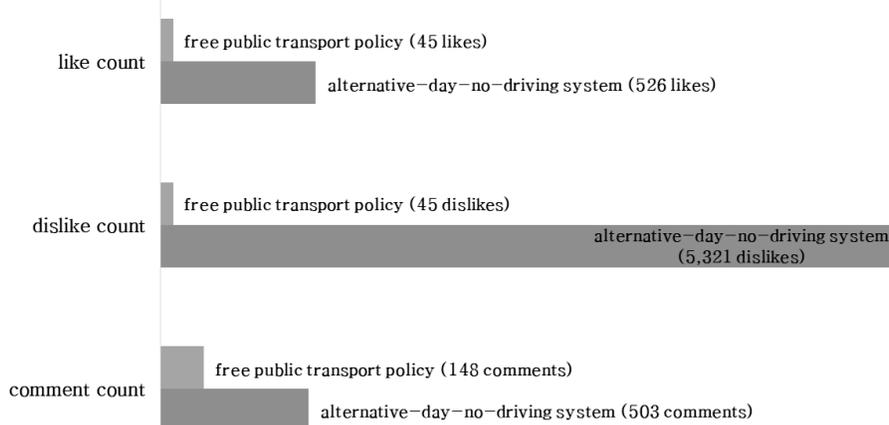


Fig. 1. Number of Likes/Dislikes and Comment for Each Policy

다. ‘대중교통 무료’ 정책에 대한 기사는 조선, 중앙, 동아일보와 한겨레, 경향신문에서, ‘차량 2부제’에 대한 기사는 조선, 중앙일보와 경향신문에서 선정하였다. 선정된 기사마다 ‘좋아요’, ‘화나요’ 수와 댓글을 수집하였다. 댓글은 R을 사용하여 크롤링하고 명사만 추출하였다. 전처리 단계로, 추출된 명사에서 정책 명과 ‘미세먼지’, 특수문자는 제거하였다. 또 명확한 분석 결과를 도출하기 위해 문맥상 의미가 유사한 단어들은 동일한 단어로 통일하였다[12].

위 두 정책에 대한 시민들의 반응을 알아보기 위해 먼저 각 정책에 대한 기사들의 ‘좋아요’, ‘화나요’ 수와 댓글 수를 비교해보았다. 그 결과는 Fig. 1과 같다.

‘좋아요’, ‘화나요’ 수 합을 변화를 보면, ‘대중교통 무료’ 정책에 관한 기사에서는 90개이지만 ‘차량 2부제’에 관한 기사에서는 5,847개로 크게 증가하였다. 마찬가지로 댓글 수도 148개에서 503개로 증가하였다. ‘차량 2부제’에 관한 기사가 ‘대중교통 무료’ 정책에 관한 기사보다 더 적음에도 불구하고 ‘좋아요’, ‘화나요’ 수와 댓글이 월등히 많다는 것은 ‘대중교통 무료’ 정책보다 ‘차량 2부제’에 대한 시민들의 관심이 더 높음을 의미한다.

다음, 각 정책에 대한 기사들의 ‘좋아요’, ‘화나요’의 비율을 알아보았다. ‘대중교통 무료’ 정책에 대한 시민들의 생각은 ‘좋아요’가 50%(45개), ‘화나요’가 50%(45개)로 그 차이가 없지만 ‘차량 2부제’에 대한 시민들의 생각은 ‘좋아요’가 9%(526개), ‘화나요’가 91%(5,321개)로 부정적인 반응이 대부분이었다. ‘화나요’의 비율이 많이 증가한 것으로 보아 현재 ‘미세먼지 비상저감조치’ 정책에 대한 시민들의 생각은 부정적인 것으로 나타났다.

마지막으로 R을 이용하여 정책별로 댓글에 텍스트마이닝을 실시하고 그 결과를 워드 클라우드로 나타내었다. 워드클라우드는 글에서 언급된 주요 단어를 시각화하는 방법으로 이를 통해 정책별로 많이 언급된 단어를 알 수 있다.

두 정책에 대한 워드 클라우드는 Fig. 2와 같다. ‘대중교통 무료’ 정책의 워드 클라우드를 살펴보면, 가장 많이 언급된 단어는 ‘세금’, ‘중국’, ‘서울만’이고 긍정의 의미를 가진 단어로는 ‘좋다’, ‘노력’, ‘서울만’ 등이 있다. ‘서울만’은 단어 자체로는



Fig. 2(A). Word Cloud of Free Public Transport Policy



Fig. 2(B). Word Cloud of Alternative-Day-No-Driving System

긍정과 부정을 판단하기 어렵지만 댓글의 문맥상 긍정의 의미를 내포한다. 부정적인 단어로는 ‘혈세’, ‘지옥철’이 있다. ‘차량 2부제’의 워드 클라우드로서 가장 많이 언급된 단어는 ‘시민’, ‘중국’, ‘차량’, ‘출퇴근’, ‘세금’이고 긍정의 의미를 가진 단어는 없다. 부정의 의미를 가진 단어로는 ‘한심’, ‘답답’, ‘헛짓’, ‘탁상행정’, ‘남비’, ‘불편’ 등이 있다.

두 정책의 워드 클라우드로 공통으로 등장하는 단어로는 ‘중국’, ‘세금’, ‘시민’이 있다. ‘중국’이라는 단어가 공통으로 등장한 이유는 미세먼지의 주된 원인이 국내가 아닌 중국에 있다는 시민들의 생각 때문인 것으로 예상된다. 반대로, 두 워드 클라우드로서 나타난 가장 큰 차이점은 ‘차량 2부제’에 관한 부정적인 단어들의 수가 ‘대중교통 무료’ 정책보다 월등히 많았다는 점이다. 이는 앞서 ‘화나요’의 비율이 많이 증가했다는 분석 결과와 같은 맥락으로 해석할 수 있다.

3.2 국내 요인의 영향력 분석

기존 정책의 실효성 분석에서 국내 요인의 영향력에 대한 시민들의 인식이 낮음을 보였다. 이를 통해 미세먼지에 대한 국내 요인의 영향력 분석이 추가적으로 필요함을 알 수 있었고 두 정책과 관련된 교통산업을 중심으로 회귀분석을 실시하여 배출원별 산업과 미세먼지와의 관련성을 파악하고자 한다. 자료 수집 당시, 산업별 사업체 수 자료[13]는 2006년부터 제공된다는 점과 산업별 사업체 수 자료와 미세먼지농도 자료[14]의 가장 최근 시점이 2016년이라는 점을 고려하여 자료의 시점을 2006년부터 2016년으로 통일하였다.

미세먼지농도(PM10, 단위: $\mu\text{g}/\text{m}^3$) 자료는 환경부 한국환경공단에서 제공한 2006년부터 2016년까지 서울, 경기, 인천 지역의 연평균자료이다. 대기오염물질 배출원별 사업체 자료는 대기오염물질 배출원이 산업별로 구분된 것을 고려하여, 먼저 통계청의 한국표준산업분류(Korean Standard Industrial Classification, 이하 KSIC)에 따라 산업별 사업체 수 자료를 수집하고 미세먼지농도와의 관련성을 높이기 위해 국립환경과학원의 대기오염물질 배출원 분류 체계[15, 16]로 재분류하였다.

KSIC에 따른 사업체 수 자료 수집 시, 자료의 일치성을 위해 앞서 수집한 미세먼지농도 자료와 시점과 지역을 동일하게 하여 수집하였다.

국립환경과학원의 대기오염물질 배출원 분류 체계 항목은 총 13개로 해당 항목과 구체적인 내용은 Table 1과 같다.

Table 1. Air Pollution Source Categories

Air pollutant	Sources
Combustion in energy industries	Emission from district heating and solid fuel transformation facilities
Non-industrial combustion plants	Emission from agriculture, forestry, aquaculture and residential facilities
Combustion in manufacturing industries	Emission from boiler, gas turbine and stationary engine
Production processes	Emission from production process facility
Storage and distribution of fuels	Emission from chemical fuel and gas supply
Solvent use	Emission from paint and ink applications, dry cleaning and other use of solvent
Road transport	Emission from cars on the road
Other mobile sources and machinery	Emission from railway, ship, aircraft and agricultural machinery
Waste treatment and disposal	Emission from waste incineration, liquid waste treatment and composting
Agriculture	Emission from fertilizer and enteric fermentation in agriculture
Other sources & sinks	Emission from vegetation, wetland or soil and fire
Fugitive dust	Fugitive dust from automotive or emission from non-ventilation facility
Biomass combustion	Emission from roast, agricultural incineration, wood stove or fireplace

위의 배출원 분류 체계를 기준으로, KSIC에 따른 산업들을 각 항목에 분류하였다. 이때, 배출원 분류 체계와 유사성이 떨어지는 산업, 관련 산업이 없는 배출원은 자료에서 제외하였다. 그 결과는 Table 2와 같다.

Table 2. Reclassification of KSIC industry

Air pollutant	Industry
Combustion in energy industry	Power generation industry
Production processes	All manufacturing industry except printing and recording
Solvent use	Printing and recording industry, Painting and construction industry, Services industry for building and industrial facility cleaning, Industrial and household laundry industry
Road transport	Bus industry for city and intercity, Taxi industry, Hearse operation industry, Freight industry, Package specialized freight industry
Other mobile sources and machinery	Railway industry, Metropolitan industry, Water transportation industry, Flight industry
Waste treatment and disposal	Sewage waste and human waste treatment industry, Waste disposal industry
Agriculture	Farm industry, Livestock industry, Farm and livestock combined industry, Forestry industry

자료 수집 및 처리의 마지막 단계로 미세먼지농도 자료와 대기오염물질 배출원별 사업체 수 자료를 합하여 최종 데이터셋을 만들었다.

미세먼지농도는 종속변수로 대기오염물질 배출원별 사업체 수는 독립변수로 설정하고, 독립변수는 값의 범위 차이를 고려하여 평균과 분산을 이용해 표준화하였다. 다음, 통계분석 프로그램인 R을 이용하여 미세먼지농도에 대한 단순선형 회귀식을 도출하였다. 해당 회귀식은 Equation (1)과 같다.

$$\hat{y} = -6.393c - 16 - 0.3n_1 + 1.52n_2 + 1.53n_3 + 0.69n_4 - 3.18n_5 - 1.88n_6 - 0.25n_7 \quad (1)$$

위 회귀식에서 \hat{y} 은 미세먼지농도, n_1 은 에너지연소 산업의 사업체 수, n_2 는 생산공정 산업의 사업체 수, n_3 는 유기용제사용 산업의 사업체 수, n_4 는 도로이동 산업의 사업체 수, n_5 는 비도로이동 산업의 사업체 수, n_6 은 폐기물처리 산업의 사업체 수, n_7 은 농업의 사업체 수를 의미한다.

도출된 회귀식의 설명력은 0.75이고 유의수준 0.1에서 비도로이동(n_5)과 폐기물처리(n_6) 산업 변수만이 유의한 것으로 나타났다.

단순선형회귀식으로는 유의한 변수가 적어 미세먼지농도와 산업 간의 관련성을 알아보기 어렵다고 판단하여 불필요한 변수를 제거하고 최적의 모형을 찾는 변수선택을 이용한 회귀분석을 실시하였다. 변수선택 방법으로 전진 선택법, 후진 제거법, 단계적 방법을 사용하였고 세 모형의 판단 기준으로는 AIC(Akaike Information Criterion) 값을 이용하였다.

AIC 값은 전진 선택법이 -30.8, 후진 제거법이 -33.66, 단계적 방법이 -33.66으로 가장 AIC 값이 작은 후진 제거법과 단계적 방법 중 전진 선택법과 후진 제거법의 보완적인 성격을 띠는 단계적 방법을 최종 모형으로 선택하였다. 단계적 방법을 통한 회귀식은 Equation (2)와 같다.

$$\hat{y} = -6.393e-16 - 0.3n_1 + 2.1n_3 + 0.95n_4 - 2.47n_5 - 2.31n_6 \quad (2)$$

단계적 방법을 통해 선택된 변수는 에너지연소(n_1), 유기용제사용(n_3), 도로이동(n_4), 비도로이동(n_5), 폐기물처리(n_6) 산업으로 총 5개의 변수이다. 회귀식의 설명력은 0.74이고 유의수준 0.1에서 에너지연소 산업을 제외한 모든 변수가 유의한 것으로 나타났다. 하지만 에너지연소 산업의 유의확률이 0.14로 유의수준과 큰 차이가 없기 때문에 비교적 유의한 편이라고 할 수 있다.

각 산업이 미세먼지농도에 미치는 영향력을 그래프로 표현한 것은 Fig. 3과 같다.

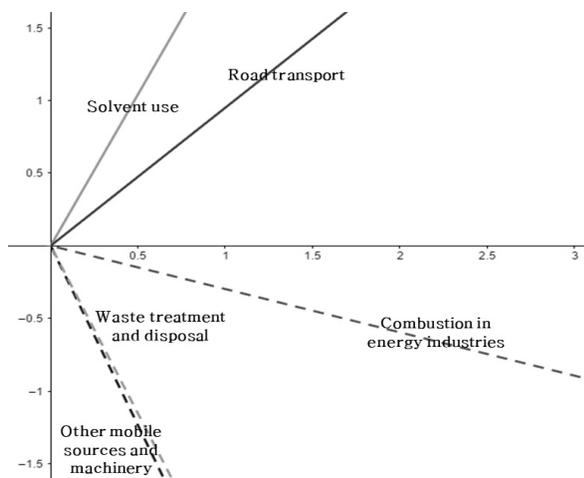


Fig. 3. Coefficients of Stepwise Regression Analysis

위 그래프에서 회귀계수가 양수인 것은 실선, 음수인 것은 점선으로 나타냈다. 미세먼지농도와 양의 관계를 보이는 산업은 유기용제사용(n_3), 도로이동(n_4) 산업이며 이 산업들의 사업체 수가 증가할수록 미세먼지농도가 증가함을 의미한다. 음의 관계를 보이는 산업은 에너지연소(n_1), 폐기물처리(n_6), 비도로이동(n_5) 산업이고 이는 사업체 수가 증가할수록 미세먼지농도가 감소함을 의미한다. 결과적으로 교통산업이라고 해서 반드시 미세먼지농도와 양의 관계를 갖는 것은 아니다. 도로의 이용 여부에 따라 미세먼지농도에 미치는 영향력의 방향이 달라진다.

미세먼지농도와 교통산업의 관계를 더 자세하게 알아보기 위해 도로이동(n_4)과 비도로이동(n_5) 산업을 추가분석 해보았다. 도로이동 산업에 해당하는 산업으로는 시내버스, 시외버스, 택시, 전세버스, 장의차량, 도로 화물, 소화물 전문 운송업이 있다. 이 7개의 산업을 독립변수로, 미세먼지 농도를 종

속변수로 하여 단순선형회귀분석을 실시했을 때, 시외버스, 택시, 전세버스, 도로 화물 운송업은 미세먼지농도와 양의 관계를, 시내버스, 장의차량, 소화물 전문 운송업은 음의 관계를 보였다. 시내버스 운송업이 미세먼지농도와 음의 관계를 가지는 이유는 환경부가 추진한 천연가스 버스 보급정책으로 인해 경유 버스가 천연가스 버스로 전환되면서 대기오염물질이 감소했기 때문인 것으로 보인다[17].

비도로이동 산업에 해당하는 철도, 도시철도, 수상, 항공 운송업을 독립변수로, 미세먼지농도를 종속변수로 하여 단순선형회귀분석을 실시하였다. 그 결과, 철도, 수상, 항공 운송업이 미세먼지 농도와 양의 관계를 보였고 도시철도 운송업은 음의 관계를 보였다. 미세먼지 농도와 양의 관계를 보이는 비도로이동 산업은 대기오염물질 배출량이 도시철도 운송업 보다는 많고 도로이동 산업보다는 적기 때문에 회귀계수의 부호가 양수인 것으로 판단된다. 도시철도 운송업은 오직 전력을 이용하여 움직이고 대기오염물질 배출량이 절대적으로 적어 미세먼지농도와 음의 관계를 갖는 것으로 보인다.

4. 국민청원 분석

기존 정책의 실효성 분석을 통해 두 정책을 바라보는 시민들의 시각이 부정적임을 알 수 있었다. 정책에 대한 부정적인 반응은 시민들의 참여를 저지시키므로 긍정적인 반응을 불러일으킬 수 있는 정책이 필요하다. 따라서 본 분석에서는 국민청원을 분석하여 시민들이 원하는 '미세먼지 비상저감조치'의 방향을 파악하고자 한다.

청와대 홈페이지의 '국민청원 및 제언' 카테고리에서 제목에 '미세먼지'가 포함되고 동의 수가 10 이상인 청원을 추출하였다[18]. 추출 시에는 청원의 내용만 Python을 사용해 추출하였다[19]. 추출한 데이터에서 '미세먼지'와 관련이 없고 불필요하다고 판단되는 단어와 길이가 2 이하인 단어는 제거하였다. 그 결과, 최종적으로 추출된 청원은 805개, 단어 수는 5,257개이다.

청원의 내용 분석을 위해 LDA(Latent Dirichlet Allocation) 알고리즘 기반 토픽 모델링을 사용하였다. 토픽 모델링의 한 기법인 LDA는 확률을 기반으로 토픽과 토픽에 분포하는 단어를 분석하여 대량의 문서 분석에 용이한 기법이다. 본 분석에서는 토픽은 10개로, 토픽의 구성 단어는 30개로 설정하고 토픽 간 관련성을 한눈에 알아보기 위해 공통된 단어들을 포함하는 토픽들을 그룹화하였다. 생성된 토픽의 그룹은 Fig. 4와 같다.

토픽 모델링 결과를 통해 10개의 토픽을 크게 네 그룹으로 나누었다. 그룹 1은 토픽 1과 8, 그룹 2는 토픽 2, 4, 5, 그룹 3은 6, 7, 9, 10, 그룹 4는 토픽 3이다. 그리고 각 그룹에 속하는 단어들의 유사성과 청원 내용을 바탕으로 주제를 도출하고 하나의 주제어를 선정하여 해당 그룹의 그룹명으로 사용하였다. 그룹 1에 속하는 단어로는 '경각심', '보여주기식', '최우선과제', '두려움' 등이 있고 이 단어들은 청원인의 일반적인 견해라고 판단하여 'Petitioner'라는 그룹명을 부여하였다.

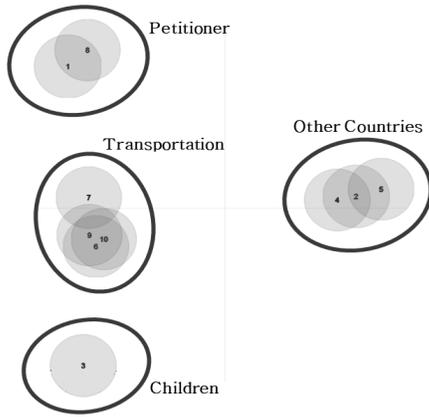


Fig. 4. Clusters of Topic

그룹 2에 속하는 단어는 ‘중국’, ‘외교적’, ‘편서풍’, ‘산둥반도’, ‘주변국’ 등으로 국외라는 공통점을 보여 ‘Other Countries’를 그룹명으로 부여하였다. 그룹 3에는 ‘경유차’, ‘화석연료’, ‘자가용’, ‘지하철’ 등 교통과 관련된 단어들 주를 이루고 있어 ‘Transportation’을 그룹명으로 하였다. 그룹 4는 ‘초등학교’, ‘중학교’, ‘고등학교’, ‘교육기관’, ‘선생님’ 등 아이들과 관련이 있다고 판단하여 ‘Children’을 그룹명으로 하였다. 아래 Table 3은 그룹명과 해당 그룹의 주요 단어, 특성을 정리한 것이다.

각 그룹의 특성을 요약하면 ‘Petitioner’에 속하는 청원들은 청원인의 일반적인 견해 혹은 현재 미세먼지 정책에 대한 반응이 대부분이었다. 또 미세먼지를 다른 문제보다 우선시하고 보여주기식이 아닌 근본적인 문제해결을 원하였다. 제시한 정책으로는 ‘국민 행동요령 마련’, ‘마스크 보급’ 등이 있

다. ‘Other Countries’에 속하는 청원들은 미세먼지의 국내 요인보다 국외 요인 해결을 원한다. 또 미세먼지의 주된 원인이 중국에 있다고 생각하고 외교를 통해 미세먼지 문제를 해결해 주기를 바란다. ‘Other Countries’에서 제시한 정책으로는 ‘중국 공장에 미세먼지 집진기 설치’, ‘중국 정부에 공장 내 공기청정기 설치 요구’, ‘중국 정부에 산둥반도로의 공장 이전 중단 요구’, ‘중국의 쓰레기 소각장 건설 예정지의 위치 변경’, ‘한국과 중국의 공동투자를 통한 태양열 발전 단지 건설’ 등이 있다. ‘Transportation’은 교통수단과 관련된 정책을 제시하여 궁극적으로 대기오염물질의 양을 줄이고자 한다. 제시한 정책으로는 ‘화석연료를 전기 및 친환경 연료로의 전환’, ‘출퇴근 시, 대중교통 사용의 활성화를 위해 대중교통 차량 수의 증대’, ‘노후차량의 교체를 위한 보조금 지원 및 차량 등급제 개선’, ‘차량 2부제를 자율에서 강제로 전환’, ‘LPG 차량으로의 개조 규제 완화’, ‘차량 내에 공기청정기 설치’, ‘자동차운행 제한’ 등이 있다. ‘Children’은 미세먼지 관련 교내 방침 마련에 관한 정책이 대부분이었다. 제시된 정책으로는 ‘환기 수단 마련’, ‘학생들에게 마스크 무상 지원’, ‘자율 등교제’, ‘미세먼지 관련 교육 시행’, ‘교내에 실내체육관 건립’ 등이 있다.

5. 결론

본 논문에서는 회귀분석과 텍스트마이닝을 이용해 기존 ‘미세먼지 비상저감조치’ 정책에 대한 실효성을 파악하고 국민청원을 분석하여 향후 정책의 방향성을 알아보았다.

‘좋아요’, ‘화나요’의 비율 변화와 워드 클라우드를 통해 현

Table 3. Properties of Each Topics in Topic Modeling

Topic	Key word	Property
Petitioner	Carcinogen, Minimum, Awareness, Early Death, Priority, National Emergency, Just to Show, Fear, Stress, Wind Power Generation, Mask	<ul style="list-style-type: none"> - General reactions to current fine dust reduction policy - People’s dissatisfactions over fine dust in their daily life and concerns about their health - Demands for fundamental solution by prioritizing the fine dust matter
Other Countries	China, Diplomacy, West Wind, Shandong Peninsula, Nation, National Level, Air Purification, Harmful Substances, Neighboring Countries, Actual Effectiveness	<ul style="list-style-type: none"> - Needs for foreign factor controlled rather than domestic one - Demands for policies through cooperating with various countries such as China
Transportation	Energy, Diesel, Fossil Fuel, Traffic, Effluent Gas, LNG, Car, Rating System, Subway, LPG, Thermoelectric Power Plant, Artificial Rainfall, Commute	<ul style="list-style-type: none"> - Proposals for new transportation policy - Demands for fossil fuel to be replaced by electricity, natural gas and hydrogen - Demands for rising the number of public transportation vehicles in order to increase the use of it - Demands for subsidy or improvement of vehicle rating system for replacing old vehicles
Children	Children, Elementary School, Middle School, High School, Kindergarten, Educational Institution, School Board, Teacher, Elementary School Student, Playground, Clean Air	<ul style="list-style-type: none"> - Demands for policy protecting vulnerable children from fine dust - Suggesting Kindergarten, Elementary, Middle and High School to be closed when fine dust is high - Demands for free mask and an education about fine dust for children

재 ‘미세먼지 비상저감조치’를 바라보는 시민들의 시각이 부정적임을 알 수 있었다. 또 ‘중국’이라는 단어를 통해 국내 요인보다 국외 요인을 미세먼지의 주된 원인으로 생각하고 있음을 알게 되었다. 시민들이 간과하고 있는 국내 요인의 영향을 파악하기 위해 회귀분석을 실시하였다. 그 결과를 보면, 도로를 이용하는 교통은 미세먼지에 양의 영향을 미치고 이용하지 않는 교통은 음의 영향을 미치는 것으로 나타났다.

토픽 모델링을 통해 미세먼지 관련 청원을 ‘Petitioner’, ‘Other Countries’, ‘Transportation’, ‘Children’ 총 네 가지의 주제로 나눌 수 있었다. ‘Petitioner’는 현재 미세먼지 정책에 대한 일반적인 견해가 대부분이었고 ‘Other Countries’는 미세먼지의 국외 요인의 해결을 촉구하였다. ‘Transportation’은 교통수단, ‘Children’은 아이들의 건강과 관련한 미세먼지 정책을 제시하였다.

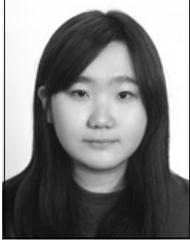
위 내용을 바탕으로 교통과 관련해 ‘미세먼지 비상저감조치’의 향후 방향을 살펴보면 다음과 같다. 국내 요인 해결에 중점을 둔 현재 정책은 국외 요인을 중시하는 시민들에게 부정적인 반응을 불러일으켜 활발한 정책 참여를 저지한다. 따라서 시민 참여를 활성화하기 위해 긍정적인 반응을 유발할 수 있는 정책으로 변화해야 하며 이를 위해 국외 요인뿐 아니라 국내 요인, 그중 도로이동 교통수단의 영향도 상당함을 알려야 한다. 또한, 대기오염물질의 양을 줄이기 위해 지하철과 같은 비도로이동 교통수단의 이용을 유도할 수 있는 정책을 펴야 한다. 이와 함께 대중교통뿐만 아니라 도로이동 교통수단이 좀 더 친환경적인 교통수단으로 변화하도록 정부 차원의 지원이 이루어진다면 미세먼지 감축에 큰 효과를 가져올 것이다. 또 시민들이 제시한 정책 및 의견을 향후 정책에 반영한다면 좀 더 다양하고 실효성 있는 정책이 될 것이다.

이 외에도 미세먼지는 한국에만 국한된 환경문제가 아니므로 전 세계 국가들의 협력이 반영된 정책이어야 한다. 이는 미세먼지에 미치는 국외 요인의 영향이 크다는 시민들의 생각을 반영한 정책이기 때문에 시민들의 더 적극적인 참여를 끌어낼 수 있다. 또 미세먼지에 취약한 아이들 혹은 노약자, 임산부에서 나아가 일반 시민까지 실천할 수 있는 행동요령을 수립한다면 적극적인 시민 참여와 시민 참여를 통한 미세먼지 감축을 달성할 수 있다.

References

- [1] Fine Dust Information Center of Seoul, “The ways of high-density fine dust(PM2.5) matter improvement in Seoul,” Press Release of Seoul(2018.2.27.), 2018.
- [2] Fine Dust Information Center of Seoul, “The evolution of ‘Emergency fine dust reduction measures’ to ‘8 Citizen-led measures,’” Press Release of Seoul(2018.2.27.), 2018.
- [3] A. Gelman and J. Hill, “Linear regression: before and after fitting the model,” in *Data Analysis Using Regression and Multilevel/Hierarchical Models*, 1st ed., Cambridge University Press, pp.53-73, 2006.
- [4] H. Choi, G. Park, S. Choi, and S. Kim, “Research on the big data collecting system for measuring of broadcast content influence,” *The Korean Institute of Broadcast and Media Engineers Conference*, pp.171-174, 2014.
- [5] Y. Cha, J. Lee, J. Choi, and H. Kim, “A Topic Modeling Approach to Marketing Strategies for Smartphone Companies,” *Knowledge Management Research*, Vol.16, No.4, pp.69-87, 2015.
- [6] M. Shin, C. Lee, H. Ha, C. Choe, and Y. Kim, “The Influence of Meteorological Factors on PM10 Concentration in Incheon,” *Journal of Korean Society for Atmospheric Environment*, Vol.23, No.3, pp.322-331, 2007.
- [7] I. Park, “What is the root cause of high-density fine dust in Seoul metropolitan region?,” *Journal of Korean Society for Atmospheric Environment*, Vol.32, No.3, pp.352-353, 2016.
- [8] S. Lee, “Characterization of Non-exhaust Particulate Matters (PM) Generated from Tire and Brake,” *Journal of the Korean Society of Automotive Engineers*, Vol.39, No.8, pp.34-38, 2017.
- [9] S. Kwon, M. Kim, G. Jung, S. Hong, S. Hong, and M. Chae, “Dust emission and vegetation indices of different regions,” *The Conference of Korean Society of Soil Science and Fertilizer*, pp.279-279, 2012.
- [10] K. Jang and J. Yeo, “The Effects of Korean and Chinese Economic Growth on Particulate Matter in Korea: Time Series Cointegration Analysis,” *Journal of Environmental Policy and Administration*, Vol.23, No.1, pp.97-117, 2015.
- [11] S. Park and H. Shin, “Analysis of the Factors Influencing PM 2.5 in Korea: Focusing on Seasonal Factors,” *Journal of Environmental Policy and Administration*, Vol.25, No.1, pp.227-248, 2017.
- [12] H. Cho, Y. Chung, J. Lee, and J. Lee, “Sentiment Analysis Using News Comments for Public Opinion Mining,” *KIIS Spring Conference*, Vol.23, No.1, pp.149-150, 2013.
- [13] Statistic Korea, “Number of businesses and employees by city, industry, and business division (‘06 ~),” 2017 [Internet], <http://kostat.go.kr>.
- [14] Air Korea “Domestic/International air quality,” [Internet], <http://www.airkorea.or.kr>, 2018.
- [15] National Institute of Environmental Research, “National Air Pollutants Emission Service,” 2018 [Internet], <http://airemissioner.go.kr>.
- [16] J. Kim, “Assessment and Estimation of Particulate Matter Formation Potential and Respiratory Effects from Air Emission Matters in Industrial Sectors and Cities/Regions,” *Journal of Korean Society of Environmental Engineers*, Vol.39, No.4, pp.220-228, 2017.
- [17] Y. Park, “The direction of the government’s policy to promote natural gas-powered buses,” Policy Report of the Ministry of Environment (2003.05.15.), 2003.
- [18] Cheongwadae, “Petition and Proposal,” 2018 [Internet], <http://www.president.go.kr>.

- [19] E. Vargiu and M. Urru, "Exploiting web scraping in a collaborative filtering-based approach to web advertising," *Artificial Intelligence Research*, Vol.2, No.1, pp.44-54, 2013.



김애니

<https://orcid.org/0000-0002-7269-1950>
e-mail : anniekim0714@gmail.com
2014년~현재 동덕여자대학교
정보통계학과 학사과정
관심분야: Text Mining &
Pattern Recognition



정소희

<https://orcid.org/0000-0002-0596-2220>
e-mail : sohee_e2@naver.com
2014년~현재 동덕여자대학교
정보통계학과 학사과정
관심분야: Biostatistics &
Pattern Recognition



최현빈

<https://orcid.org/0000-0003-2042-1657>
e-mail : chb4477@naver.com
2014년~현재 동덕여자대학교
정보통계학과 학사과정
관심분야: Experimental Design &
Regression Analysis



김현희

<https://orcid.org/0000-0002-7507-8342>
e-mail : heekim@dongduk.ac.kr
1996년 이화여자대학교 컴퓨터학과(학사)
1998년 이화여자대학교 컴퓨터학과(석사)
2005년 이화여자대학교 컴퓨터공학과
(공학박사)
2005년~2006년 LG전자 디지털미디어 연구소 선임연구원
2006년~현재 동덕여자대학교 정보통계학과 부교수
관심분야: Digital Pharmacovigilance, Machine Learning,
Ontology