

미세먼지 감축을 위해 회귀분석과 텍스트 마이닝을 활용한 교통 정책에 대한 반응 분석

김애니, 정소희, 최현빈, 김현희
동덕여자대학교 정보통계학과

e-mail : (20141146, 20141180, 20141188, heekim)@dongduk.ac.kr

Analysis of response to transportation policy for particulate matter reduction using regression analysis and text mining

Annie Kim, So Hee Jeong, Hyun Bin Choi, Hyon Hee Kim
Dept. of Statistics and Information Science, Dongduk Women's University

요 약

최근 서울시에서는 ‘미세먼지 비상저감조치’로 ‘대중교통 무료’ 정책을 시행하고 후속 조치로 ‘시민 참여형 차량 2부제’를 제안하였다. 본 논문에서는 먼저, 위 두 교통 정책의 실효성을 파악하기 위해 ‘교통’을 중심으로 각 산업이 미세먼지에 미치는 영향을 알아보고, 위 정책들에 대한 시민들의 반응을 분석한다. 각 산업이 미세먼지에 미치는 영향은 회귀분석으로, 두 정책에 대한 시민들의 반응은 텍스트 마이닝 기법을 통해 알아보았다. 그 결과, 교통수단의 도로 이용 여부에 따라 미세먼지에 미치는 영향력의 정도와 방향이 다를 수 있었고 정책에 대한 관심과 부정적인 의견이 크게 증가함을 알 수 있었다. 또 국외 요인에 대한 해결책도 필요로 함을 알 수 있었다. 마지막으로 위 결과를 토대로 향후 미세먼지 문제와 관련된 정책이 나아갈 방향을 제시한다.

1. 서론

최근 미세먼지 문제가 심각해지면서 이에 대한 대책으로 ‘미세먼지 비상저감조치¹⁾’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. 이후 서울시는 2월 27일에 ‘시민 주도형 미세먼지 비상저감조치 8대 대책’의 하나로 ‘시민 참여형 차량 2부제’(이하 ‘차량 2부제’)를 발표하였다.

본 논문의 목적은 미세먼지에 영향을 미치는 산업을 알아보고, 이들 산업 중에서도 특히 교통정책과 관련이 있는 도로 이동 산업이 미세먼지에 미치는 영향을 파악하였다. 또한 서울시에서 제시한 교통 정책에 대한 시민들의 반응을 분석하였다.

먼저, ‘교통’을 중심으로 각 산업이 미세먼지에 미치는 영향을 알아보기 위해 미세먼지농도와 대기오염물질 배출원별 사업체 수 자료를 수집하여 회귀분석[1]을 실시하였다. 분석 결과, 교통수단이라고 해서 반드시 미세먼지와 양의 관계를 갖는 것은 아니며 도로를 이용하는 교통수단만이 미세먼지에 악영향을 미친다

는 사실을 알 수 있었다.

다음으로 ‘대중교통 무료’ 정책과 ‘차량 2부제’에 대한 시민들의 반응을 분석하기 위해, 정책 명을 검색어로 하여 5개 주요 언론사의 기사를 정책별로 수집하고 수집한 기사들의 ‘좋아요’, ‘화나요’, 댓글 수와 댓글을 추출하였다. 정책별로 댓글에 텍스트 마이닝 기법[2]을 적용하여 주요 단어를 추출한 다음 워드 클라우드를 이용하여 결과를 시각화하였다. 그 결과, 미세먼지 정책에 대한 시민들의 관심과 부정적인 의견이 크게 증가함을 알 수 있었다. 또 ‘중국’이라는 키워드를 통해 시민들이 국외 요인에 대한 해결책도 필요로 함을 알게 되었다.

본 연구의 공헌은 정형 및 비정형 자료 분석 결과를 통합하여 ‘미세먼지 비상저감조치’에 대한 향후 방향을 제시한 것이다. 이를 통해, 향후 정책은 시민들에게 비도로이동 교통수단의 이용을 장려하고 긍정적인 의견을 끌어 낼 수 있는 정책이어야 한다. 또 미세먼지에 미치는 부정적인 국내·외 요인을 모두 통제하기 위해서는 전 세계 국가들의 협력이 필요하다.

본 논문의 구성은 다음과 같다. 2장에서는 미세먼지와 관련된 연구에 대해, 3장에서는 각 산업이 미세먼지에 미치는 영향에 대해 알아본다. 4장에서는 두 정책에 대한 시민들의 반응을 분석하고 마지막 5장에서는 결론 및 정책의 향후 방향을 제시한다.

1) 미세먼지로 인해 오염된 대기 질의 개선 및 대기오염 방지를 통해 국민들의 건강을 보호하는 조치

2. 관련 연구

기존 미세먼지에 미치는 영향에 관한 연구는 다음과 같다. Shin, MoonKhee et al.의 연구[3]는 기상인자(풍향, 풍속, 상대습도, 일기 유형, 해륙풍 유무)가 미세먼지농도에 미치는 영향을 살펴보았다. 이 연구 결과에 의하면, 향후 연구에서는 미세먼지농도와 관련성을 높이기 위해 대기오염 배출량을 고려해야 한다고 말한다. 따라서 본 논문에서는 이를 반영하여 산업을 재분류하고 미세먼지와 산업 간의 관계를 알아 보았다.

3. 산업이 미세먼지에 미치는 영향 분석

3.1. 자료 수집 및 처리

미세먼지농도(PM10, 단위: $\mu\text{g}/\text{m}^3$) 자료는 환경부 한국환경공단에서 제공한 2006년부터 2016년까지 17개의 시도별 연평균 자료이다[4]. 자료의 일관성을 위해 세종특별시는 충청남도에 포함하였다. 충청남도의 미세먼지농도는 세종특별시와 충청남도에 있는 미세먼지 관측소들의 미세먼지농도를 산술평균 낸 것이다.

대기오염물질 배출원별 사업체 자료는 대기오염물질 배출원이 산업별로 구분된 것을 고려하여, 먼저 산업별로 사업체 수를 수집하였다. 이를 위해 통계청의 한국표준산업분류(Korean Standard Industrial Classification, 이하 KSIC)에 따라 사업체 수 자료[5]를 수집하고 미세먼지농도와 관련성을 높이기 위해 국립환경과학원의 대기오염물질 배출원 분류 체계[6][7]로 재분류하였다.

KSIC에 따른 사업체 수 자료 수집 시, 앞서 수집한 미세먼지농도 자료와의 일치성을 위해 2006년부터 2016년까지 17개의 시도별로 자료를 수집하고 세종특별시의 자료는 충청남도에 합하였다.

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

<표 1> 대기오염물질 배출원 분류 체계

항목	내용
에너지산업 연소	난방시설, 석유정제설비 산업 등의 배출
비산업 연소	주거시설, 농업·임업·양식업 등의 배출
제조업 연소	제조업의 보일러, 가스 터빈 및 엔진, 용광로 등의 배출
생산공정	생산 공정 산업에서의 배출
에너지수송 및 저장	화석연료, 휘발유 공급 등으로 인한 배출
유기용제 사용	페인트, 잉크, 세탁소 용매산업 등의 배출
도로이동오염원	도로에서 주행하는 자동차로 인한 배출
비도로이동오염원	철도, 선박 운항 및 항공기, 건설장비, 농기계 등의 배출
폐기물처리	폐기물소각, 폐수처리, 퇴비화 등의 폐기물 처리로 인한 배출

농업	가축의 분뇨관리 및 비료사용으로 인한 배출
기타 먼 오염원	식생에 의한 오염물질 배출, 습지나 토양에서의 오염물질 배출, 산불 및 화재
비산업연소	자동차 제비산먼지와 사업장 또는 공정상에서 일정한 배출구 없는 배출
생물성 연소	고기 및 생선구이, 노천소각, 농업 잔재물 소각, 목재 난로 및 보일러, 아궁이, 숯가마 등에서의 배출

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 <표 2>와 같다.

<표 2> KSIC의 산업 재분류

배출원	해당 산업
에너지산업 연소	발전업
생산공정	인쇄 및 기록 매체 복제업을 제외한 모든 제조업
유기용제 사용	인쇄 및 기록 매체 복제업, 도장·도배 및 내장 공사업, 건물 산업설비 청소 및 방제 서비스업, 산업용 세탁업, 가정용 세탁업
도로이동오염원	시내버스 운송업, 시외버스 운송업, 택시 운송업, 전세버스 운송업, 장의차량 운송업, 도로 화물 운송업, 소화물 진문 운송업
비도로이동오염원	철도운송업, 도시철도 운송업, 수상 운송업, 항공 운송업
폐기물처리	하수·폐수 및 분뇨 처리업, 폐기물 처리업
농업	작물 재배업, 축산업, 작물 재배 및 축산 복합농업, 영림업

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

3.2. 분석 및 결과

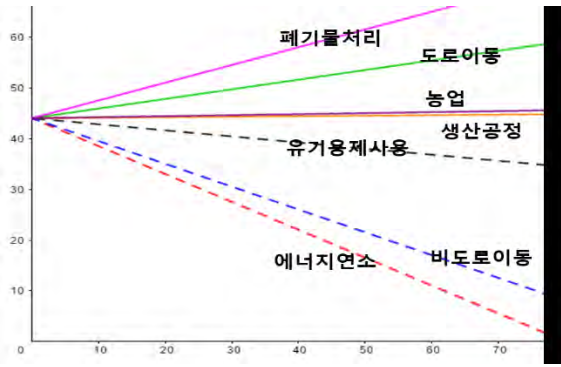
본 분석에서 사용한 자료는 총 176개로 지역별로 11개씩이다. 최종 데이터셋의 미세먼지농도는 종속변수로, 대기오염물질 배출원별 사업체 수는 독립변수로 설정하였다. 통계분석 프로그램인 R을 이용하여 미세먼지농도에 대한 회귀식을 지역별로 도출하였다.

먼저, 미세먼지농도에 미치는 전체 산업의 영향을 알아보기 위해 도출된 여러 회귀식 중 설명력이 높고 유의한 변수가 가장 많은 충청남도의 회귀식을 선택하였다. 해당 회귀식은 다음과 같다.

$$\hat{y} = 44.04 - 0.55n_1 + 0.01n_2 - 0.12n_3 + 0.19n_4 - 0.45n_5 + 0.35n_6 + 0.02n_7$$

(\hat{y} = 미세먼지농도, n_1 = 에너지연소 산업의 사업체 수, n_2 = 생산공정 산업의 사업체 수, n_3 = 유기용제사용 산업의 사업체 수, n_4 = 도로이동 산업의 사업체 수, n_5 = 비도로이동 산업의 사업체 수, n_6 = 폐기물처리 산업의 사업체 수, n_7 = 농업의 사업체 수)

위 회귀식의 설명력은 0.98 이고 유의수준 5%에서 $n_1 \sim n_7$ 변수 모두 유의한 것으로 나타났다. 각 산업의 사업체 수들이 미세먼지농도에 미치는 영향력을 그래프로 표현한 것은 (그림 1)과 같다.

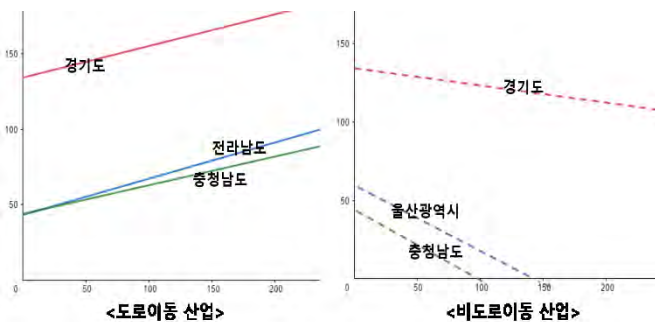


(그림 1) 충청남도의 회귀계수 그래프

위 그래프에서 회귀계수가 양수인 것은 실선, 음수인 것은 점선으로 나타냈다. 그 결과, 미세먼지농도와 양의 관계를 보이는 산업은 생산공정, 도로이동, 폐기물처리, 농업이며 음의 관계를 보이는 산업은 에너지연소, 유기용제사용, 비도로이동 산업이다.

미세먼지농도와 양의 관계를 갖는 산업들은 공통으로 배출물질이 대기와 접촉하는 정도가 크다고 판단된다. 반대로 미세먼지농도와 음의 관계를 갖는 산업 중 유기용제사용, 비도로이동 산업은 배출물질이 대기와 접촉하는 정도가 작은 것으로 보인다. 에너지연소 산업은 추가분석 한 결과, 이 산업에 속하는 화력, 수력, 원자력, 기타 발전업 중 수력의 영향력이 절대적으로 크고 영향력의 방향이 음이었다. 이 때문에 에너지연소 산업이 미세먼지와 음의 관계를 갖는 것으로 예측된다.

다음, 미세먼지농도에 미치는 교통산업의 영향력을 알아보기 위해 이와 관련된 도로이동과 비도로이동 산업의 회귀계수를 지역별로 살펴보았다. 그 결과는 (그림 2)와 같다.



(그림 2) 교통산업의 회귀계수 그래프

도로이동 산업이 유의하게 나타난 지역은 경기도, 전라남도, 충청남도이다. 각각의 회귀계수는 0.21, 0.24, 0.19 이고 미세먼지와 양의 관계를 갖는다. 이는 도로이동 산업의 사업체 수가 증가할수록 미세먼지농도가 증가함을 의미한다. 비도로이동 산업이 유의하게 나타난 지역은 경기도, 울산광역시, 충청남도이며 각각의 회귀계수는 -0.11, -0.42, -0.45 로 미세먼지와 음의 관계를 갖는다. 이는 비도로이동 산업의 사업체 수가 증가할 때, 미세먼지농도는 오히려

감소함을 의미한다. 결과적으로, 교통산업이라고 해서 반드시 미세먼지농도와 양의 관계를 갖는 것은 아니다. 도로의 이용 여부에 따라 미세먼지농도에 미치는 영향력의 방향이 달라진다.

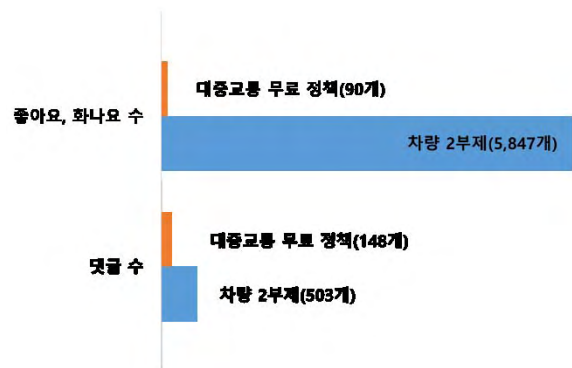
4. 정책에 대한 시민들의 반응 분석

4.1. 자료 수집 및 처리

먼저, 정책 명을 검색어로 사용하여 주요 언론사별로 기사를 선정하였다. ‘대중교통 무료’ 정책에 대한 기사는 조선, 중앙, 동아일보와 한겨레, 경향신문에서, ‘차량 2부제’에 대한 기사는 조선, 중앙일보와 경향신문에서 선정하였다. 선정한 기사마다 ‘좋아요’, ‘화나요’ 수와 댓글을 수집하였다. 댓글은 R 을 사용하여 크롤링하고 명사만 추출하였다. 다음, 전처리 단계로 추출된 명사에서 정책 명과 ‘미세먼지’, 특수문자는 제거하였다. 또 명확한 분석 결과를 도출하기 위해 문맥상 의미가 유사한 단어들은 같은 단어로 통일하였다.

4.2. 분석 및 결과

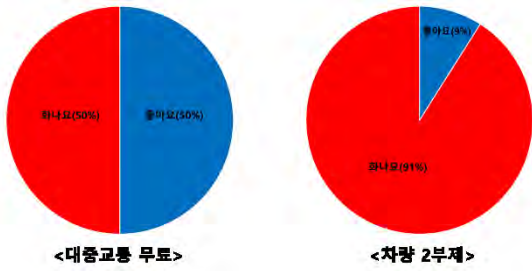
위 두 정책에 대한 시민들의 반응을 알아보기 위해 첫 번째로 각 정책에 대한 기사들의 ‘좋아요’, ‘화나요’ 수와 댓글 수를 비교해보았다. 그 결과는 (그림 3)과 같다.



(그림 3) 정책별 ‘좋아요’, ‘화나요’ 수와 댓글 수

위 그림에서 ‘좋아요’, ‘화나요’ 수의 변화를 보면, ‘대중교통 무료’ 정책에 관한 기사에서는 90 개이지만 ‘차량 2부제’에 관한 기사에서는 5,847 개로 많이 증가하였다. 마찬가지로 댓글 수도 148 개에서 503 개로 증가하였다. ‘차량 2부제’에 관한 기사가 ‘대중교통 무료’ 정책에 관한 기사보다 더 적음에도 불구하고 ‘좋아요’, ‘화나요’ 수와 댓글 수가 월등히 많았다. 이는 ‘대중교통 무료’ 정책보다 ‘차량 2부제’에 대한 시민들의 관심도가 더 높다는 것을 보여준다.

다음으로 각 정책에 대한 기사들의 ‘좋아요’, ‘화나요’의 비율을 알아보았다.



(그림 4) 정책별 ‘좋아요’, ‘화나요’ 비율

(그림 4)에 의하면, ‘대중교통 무료’ 정책에 대한 시민들의 생각은 ‘좋아요’가 50%, ‘화나요’가 50%로 차이가 없지만 ‘차량 2 부제’에 대한 시민들의 생각은 ‘좋아요’가 9%, ‘화나요’가 91%로 부정적인 반응이 대부분이었다. ‘화나요’가 많이 증가한 것으로 보아 결과적으로 현재 ‘미세먼지 비상저감조치’ 정책에 대한 시민들의 생각은 부정적인 것으로 나타났다.

마지막으로 R 을 이용하여 정책별로 댓글에 대한 텍스트 마이닝을 실시하였다. 텍스트 마이닝의 결과를 바탕으로 워드 클라우드를 생성하였다. 워드 클라우드는 글에서 언급된 핵심 단어를 시각화하는 방법으로 여기서는 정책별로 많이 언급된 단어를 살펴본다. 두 정책에 대한 워드 클라우드는 (그림 5)와 같다.



(그림 5) 정책별 워드 클라우드

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

두 정책의 워드 클라우드에 공통으로 등장하는 단어로는 ‘중국’, ‘세금’, ‘시민’이 있다. ‘중국’이라는 단어가 공통으로 등장한 이유는 미세먼지가 중국의 영향이라는 시민들의 생각 때문으로 예상된다. 반대로, 두 워드 클라우드에서 나타난 가장 큰 차이점은 ‘차량 2 부제’에 관한 부정적인 단어들의 수가 ‘대중교통 무료’

정책보다 월등히 많았다는 점이다. 이는 앞선 ‘화나요’의 비율이 많이 증가했다는 분석 결과와 같은 맥락으로 해석이 가능하다.

5. 결론 및 정책의 향후 방향

본 논문에서는 여러 산업 중 교통산업을 중심으로 ‘교통’이 미세먼지에 미치는 영향력을 알아보고 교통과 관련된 두 ‘미세먼지 비상저감조치’에 대한 시민들의 반응을 분석하였다. 회귀분석을 통해, 도로 이용 여부에 따라 도로를 이용하는 교통은 미세먼지에 양의 영향을 미치고, 이용하지 않는 교통은 음의 영향을 미치는 것으로 나타났다.

다음, ‘좋아요’, ‘화나요’의 비율 변화와 워드 클라우드를 통해 현재 ‘미세먼지 비상저감조치’를 바라보는 시민들의 시각이 부정적임을 알 수 있었다. 또 시민들이 ‘중국이 미세먼지에 미치는 영향이 크다’는 생각을 하는 것도 알게 되었다.

위 내용을 바탕으로 ‘미세먼지 비상저감조치’의 향후 방향을 살펴보면 다음과 같다. 먼저, 시민들에게 국외의 영향뿐만 아니라 국내 도로이동 교통수단의 영향도 상당함을 알리고 지하철과 같은 비도로이동 교통수단의 이용을 유도할 수 있는 정책을 펴야 한다. 다음으로, 정책에 대한 시민들의 부정적인 반응은 시민들의 정책 참여를 저지시키므로 긍정적인 반응을 불러일으킬 수 있는 정책이 요구된다. 그래야 시민들의 정책 참여를 이끌 수 있고 나아가 시민 참여를 통한 미세먼지 감축이라는 목적을 달성할 수 있다. 마지막으로 미세먼지는 동아시아에만 국한된 환경문제는 아니므로 중국뿐만 아니라 전 세계 국가들의 협력이 반영된 정책이어야 한다. 이는 미세먼지가 중국의 영향이라는 시민들의 생각을 반영한 정책이기 때문에 시민들의 더 적극적인 참여를 끌어낼 수 있다.

참고문헌

- [1] 서민구, 「R 을 이용한 데이터 처리&분석 실무」, 길벗, 2014, p.264-302
- [2] 최홍규 외 공저, “방송콘텐츠 영향력 도출을 위한 빅데이터 분석체계에 관한 연구”, 한국방송 미디어 공학회 학술발표대회 논문집, 2014, p.171-174
- [3] 신문기 외 공저, “기상인자가 미세먼지 농도에 미치는 영향”, 한국대기환경학회지, 2007, p.322-331
- [4] Air Korea, “국내/국외 대기질”, www.airkorea.or.kr, 2018
- [5] 통계청, “시도·산업·사업체구분별 사업체수, 종사자수('06~)”, http://kostat.go.kr, 2017
- [6] 국립환경과학원, “대기오염물질 배출원 대분류 체계”, http://airemiss.nier.go.kr, 2018
- [7] 김준범, “국내 산업 및 시도별 대기오염물질 배출량자료를 이용한 미세먼지 형성 가능성 및 인체 호흡기 영향 평가추정”, 대한환경공학회지, 2017, 39(4), p.220-228