

수도권 지역의 미세먼지(PM10)농도 예측 머신러닝 모델 구축

저자 (Authors)	류민지, 김진수 Minji Ryu, Jinsoo Kim
출처 (Source)	대한공간정보학회 학술대회 , 2020.6, 182-183 (2 pages) Proceedings of Korean Society for Geospatial Information Science , 2020.6, 182-183 (2 pages)
발행처 (Publisher)	대한공간정보학회 Korea Spatial Information Society
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09371412
APA Style	류민지, 김진수 (2020). 수도권 지역의 미세먼지(PM10)농도 예측 머신러닝 모델 구축. 대한공간정보학회 학술대회, 182-183.
이용정보 (Accessed)	부산도서관 210.103.83.*** 2021/09/24 13:53 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

수도권 지역의 미세먼지(PM10)농도 예측 머신러닝 모델 구축

류민지*, 김진수

Minji Ryu, Jinsoo Kim

부경대학교 지구환경시스템과학부 석사과정 (rmj12@naver.com)

부경대학교 공간정보시스템공학과 부교수 (jinsookim@pknu.ac.kr)

요약

본 연구는 수도권 지역 (서울, 경기, 인천) 의 정확한 미세먼지 농도 예측을 위해 머신 러닝 기반의 랜덤 포레스트 (Random Forest) 기법을 적용하여 모델 구축을 수행하고자 한다. 미세먼지 농도 예측 연구를 위한 자료로는 미세먼지 데이터와 18개의 미세먼지 농도 관련 인자를 사용하였다. 수집 자료를 기반으로 데이터베이스를 구축한 후 머신 러닝 기반의 모델을 구축하였다. 모델은 랜덤 포레스트 기법을 사용하여 구축하였으며 모델 구축 후 모델에 대한 성능 평가를 실시하였다. 결과적으로 구축된 모델은 성능에 대한 추가적인 연구가 필요한 것으로 판단했다.

1. 서론

2016년 기준 PM10 농도는 수도권 지역을 중심으로 지속적으로 감소하고 있으나 여전히 선진국에 비해 높은 수준이며, 특히 2013년부터 기상 및 국외 영향 등으로 다시 정체하는 추세를 보이고 있다 (환경부, 2016).

에어코리아 홈페이지에 제시된 정보에 의하면 미세먼지는 천식과 같은 호흡기계 질병을 악화시키고, 폐 기능의 저하를 초래한다. 미세먼지의 농도가 감소하지 않는 추세가 지속된다면 인간의 건강 피해가 늘어날 것이고 이에 대비한 미세먼지 농도를 예측하는 연구가 늘어나고 있다.

선행 연구를 통해 국내에서는 미세먼지의 농도를 예측하기 위해 통계적 기법이나 회귀 분석 등의 기법을 사용한 연구가 대다수인 반면, 인공지능 혹은 머신러닝을 이용한 연구는 다소 부족한 것으로 판단하였다.

그러므로 본 연구는 머신러닝을 기반으로 한 미세먼지 농도 예측 모델을 구축하고자 한다.

연구 지역은 수도권(서울, 경기, 인천) 지역이며, 데이터베이스 구축을 위한 입력 자료로는 미세먼지(PM10) 데이터와 18개의 미세먼지 관련 인자를 입력 자료로 수집하였다.

2. 본론

미세먼지(PM10) 자료는 에어코리아에서 제공하는 일반대기오염측정망 자료를 사용하였다. 수도권 지역 내 관측소 지점의 시간대별 관측 값을 일평균 값으로 환산하여 데이터를 구축하였다.

미세먼지 농도에 영향을 줄 수 있는 관련 인자는 1·2·3종 대기배출사업장으로부터의 거리, 4·5종 대기배출사업장으로부터의 거리, 인구 밀도, 건물 밀도, 기온, 강수량,

습도, 풍향, 풍속, 주거 지역으로부터의 거리, 상업 지역으로부터의 거리, 농업 지역으로부터의 거리, 공업 지역으로부터의 거리, 산림 지역으로부터의 거리, 도로로부터의 거리, 항만 지역으로부터의 거리, 나지 지역으로부터의 거리, 수역 지역으로부터의 거리 총 18가지로 기후 및 행정, 토지피복과 관련 된 인자를 선정하였다. 자료의 출처는 아래 표 1과 같다.

표 1 미세먼지 관련 인자

구분	출처
1, 2, 3종 대기배출사업장으로부터의 거리	SEMS, CAPSS
4, 5종 대기배출사업장으로부터의 거리	
인구밀도	통계청
건물밀도	행정안전부
기온	기상청 AWS
강수량	
습도	
풍향	
풍속	
주거지역으로부터의 거리	토지피복지도
상업지역으로부터의 거리	
농업지역으로부터의 거리	
공업지역으로부터의 거리	
산림지역으로부터의 거리	
도로로부터의 거리	
항만지역으로부터의 거리	
나지지역으로부터의 거리	
수역지역으로부터의 거리	

관련 인자들은 각 인자 별 중요도를 평가하여 모델 구축에 적용되었다.

입력 자료를 기반으로 데이터베이스를 구축한 후 모델 구축을 수행하였으며 모델은 랜덤 포레스트(Random Forest) 기법을 기반으로 제작하였다. 모델의 훈련 및 검증을 위해 입력 자료인 미세먼지(PM10) 데이터를 훈련 데이터 셋 80%, 검증 데이터 셋 20%의 비율로 나누어 모델 구축을 수행하였다. 또한 모델 구축에는 파라미터 최적화 과정도 적용하였다.

3. 결론

본 연구는 수도권 지역의 정확한 미세먼지 농도 예측을 위한 머신러닝 기반의 모델을 구축하고자 하였다. 데이터베이스 구축을 위해 미세먼지(PM10) 자료를 수집하고 관련 인자들을 선정하였다. 인자 별 중요도 평가 결과는 항만 지역으로부터의 거리가 가장 높은 기여도를 보였으며 도로로부터의 거리가 가장 낮은 기여도를 나타냈다. 훈련 데이터 셋과 검증 데이터 셋을 이용한 모델의 정확도 평가 결과에서 훈련 데이터 셋을 이용한 정확도는 0.95, 검증 데이터 셋을 이용한 정확도는 0.79였다. 결론적으로 본 연구에서 구축한 모델은 모델의 정확도 부분에 대한 개선이 필요하다고 판단되었다. 추후 연구를 통해 입력 자료 및 구축 데이터베이스를 보정한 후 이를 비교분석하여 높은 정확도의 모델을 구축하는 과정이 필요할 것이다.

참고문헌

1. 환경부, 2016, 2016 환경백서, p235-236