

머신러닝 모델을 활용한 모기 활동량 측정

이세훈*, 김기태*, 김영호^o, 허유진*

^o인하공업전문대학 컴퓨터시스템과,

*인하공업전문대학 컴퓨터시스템과

e-mail: {seihoon, kkt}@inhatc.ac.kr*, kymin555@naver.com^o, ujin7339@naver.com*

Measurement of Mosquito Activity using Machine Learning Model

Se-Hoon Lee*, Ki-Tae Kim*, Yeong-Ho Kim^o, Yu-Jin Hur*

^oDept. of Computer Systems & Engineering, INHA TECHNICAL COLLEGE,

*Dept. of Computer Systems & Engineering, INHA TECHNICAL COLLEGE

● 요약 ●

본 논문에서는 모기 활동 수치를 측정하기 위한 효율적인 머신러닝 모델을 제안한다. 수집된 데이터의 분석을 통해 효율적인 모델을 선정한다. 또한 데이터셋의 상관관계를 분석하고 데이터 가중치에 따라 모기의 활동에 영향을 주는 환경이 무엇인지를 분석한다. 본 논문에서는 모델을 이용한 앱 개발하여 실질적으로 모델을 활용한 예시를 보이고 실생활에서의 해당 모델을 도입하였을 때 가져올 일상의 긍정적 효과를 보인다.

키워드: 모기 활동량 측정(Measurement of Mosquito Activity), 머신러닝 모델(Machine Learning Model), 랜덤 포레스트(Random Forest), 웹앱(Web App)

I. Introduction

최근 국내 말라리아 환자가 급증하고 있고, 여름철에는 모기로 인해 생활적으로 불편함을 겪고 있다.

서울특별시 감염병관리과에서는 모기에 대처할 수 있게 모기예보제를 시행하고 있고, 모기 예보 단계에 따른 시민 행동 요령을 제공하고 있다. 그러나, 서울특별시의 모기예보제의 경우 현재 모기 예보를 서울특별시 홈페이지의 업로드된 포스터의 QR코드를 통해 접속하여 확인할 수 있지만 현재는 금일의 수치를 계산하여 기록할 뿐, 미래의 모기 활동 정도를 예보해주지 않는다.

본 논문에서는 직접 데이터를 수집하여 내일의 모기 활동을 예측하는 모기예보제를 만들고 쉽게 확인하여 모기를 예방할 수 있는 서비스 모델을 제시하고자 한다. 머신러닝을 위한 데이터셋은 서울특별시에서 수집하였고 일일 최저 기온, 평균 기온, 최고 기온과 일일 강수량을 기록하고 모기 수치도 함께 기록하는 것을 확인할 수 있다. 해당 데이터를 머신러닝으로 활용하기 위해 데이터 전처리를 하였다.

Table 1. Feature_importances

	Date	Rain	Mean_Temp	Min_Temp	Max_Temp
중요도	0.227	0.005	0.03	0.714	0.021

Random Forest 모델에서 최저기온 -> 날짜 -> 평균기온 -> 최고기온 -> 강수량 순서로 중요도를 나타 낸다. 이를 통해 최저기온에 따라 모기 출현이 결정된다는 것을 확인할 수 있다.

III. Design of the System

라즈베리파이와 ESP32 보드를 활용하여 데이터 수집 시스템을 구성하였다. ESP32 보드는 온습도 센서를 탑재하고 온도 값을 측정하여 전송할 수 있는 기능을 갖추고 있다.

II. Analysis of Environmental Factors

여러 가지 모델 중 Random Forest의 feature_importances 함수를 사용하여 학습시킨 모델의 데이터 셋의 데이터 특성 중요도는 표 2와 같이 나타난다.



Fig. 1. Data Collection Logic

ESP32 보드는 주기적으로 온도 값을 측정하고, 플라스크 서버에 GET 방식으로 전송한다. 측정된 온도 값은 HTTP 요청을 통해 Flask 서버로 전달된다.

Flask 서버는 온도 값을 받으면 즉시 Beautiful Soup을 사용하여 기상청에서 서울의 현재 강수량을 크롤링한다. 이후, 수집된 온도 값과 강수량 값을 리즈베리파이 데이터베이스에 저장하기 위해 요청을 보낸다. 리즈베리파이의 MySQL을 활용하여 데이터베이스를 구축하였다. 수집된 온도 값과 강수량 값은 MySQL에 저장된다. MySQL은 뷰를 활용하여 일일 뷰를 생성하였다. 이를 통해 각 일자별로 최저 기온, 평균 기온, 최고 기온 및 평균 강수량과 같은 정보를 볼 수 있도록 하였다.

Random Forest, LightGBM, XGBoost, GradientBoosting 모델을 비교해본 결과는 표2와 같다.

Table 2. Score comparison by model

	Train Score	Test Score
Random Forest	0.933	0.897
XGBoost	0.997	0.872
LightGBM	0.959	0.849
Gradient Boosting	0.994	0.863

LightGBM, XGBoost, Gradient Boosting 은 훈련 세트의 점수는 높지만, 테스트 세트의 점수와 차이가 크게 나는 과소 적합 현상을 확인할 수 있다. 그러나, Random Forest의 경우 훈련 세트와 테스트 세트의 점수 차가 크지 않고 데이터를 분석해보았을 때, 과소 적합과 과대 적합 현상이 일어나지 않아 활용하기 적합한 모델이라고 판단된다.

최고온도, 최저온도, 평균온도가 높아질수록 모기의 군집도가 모여 있는 것을 확인할 수 있다. 이를 통해 우리나라의 날씨 특성상 여름철에 최고온도, 최저온도, 평균온도가 높아 모기의 군집도가 모여 있을 것이라 추론할 수 있다. 속성별 결과값 영향정도는 Mosquito_Indicator, rain_mm, mean_T, min_T, max_T의 순서이다. 이에 따라 환경 변수의 영향력은 중요도에 따라 측정이 되며 가장 중요한 속성별 중요도는 Mosquito_Indicator이고, 다음으로 중요한 값은 rain_mm인 것을 확인할 수 있다.

단계를 보여주는 서비스를 제공하고 있다. 또한 단계와 지수를 확인한 후에 행동 수칙 및 방법을 통하여 행동 수칙 화면을 통해 모기 단계별 행동 수칙을 알려준다.

IV. Conclusion

본 논문에서 기존의 서울 모기 예보제를 기반으로 IoT, 날씨 API와 머신러닝 모델을 활용하여 모기 군집도를 파악할 수 있는 서비스를 개발하였다. 이 서비스를 통해 여름철에 모기의 발생률이 높아짐에 따라 생기는 문제들을 방지하고 공중 보건에 기여하는데 집중하고자 하였다.

하지만 현재 데이터는 수도권 모기 출현 값에 맞춰진 데이터로 다른 지역에 적용하기에 한계가 있다. 이러한 한계를 극복하기 위해 지역마다 모기 출현 정도를 측정하고 시간에 맞는 온습도 강수량을 측정한 데이터셋을 준비한 다음, 머신러닝 모델을 활용하여 모기를 예방할 수 있을 것이다.

REFERENCES

- [1] Seoul, <https://news.seoul.go.kr/welfare/mosquito>
- [2] Scikit-learn, <https://scikit-learn.org/stable/>
- [3] Korea Weather, <https://data.kma.go.kr/climate/>

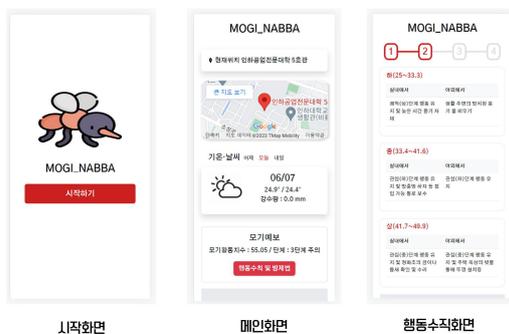


Fig. 2. Mosquito activity measurement service

생성된 모델을 기반으로 작성된 모기 나뉘 서비스는 어제, 오늘, 내일에 따른 모기 예보를 알려주며 해당 날짜의 모기 활동 지수와