

머신러닝 기반 기후 데이터를 활용한 모기 개체 수 예측

황세영, 차예빈, 차형빈, 고진광

순천대학교 컴퓨터공학과

Prediction of Mosquitoes using Climate Data based on Machine Learning

Se-Young Hwang, Ye-Bin Cha, Hyung-Bin Cha, JinGwang Koh

Dept. of Computer Engineering, Sunchon National University

요 약

최근 지구온난화에 따른 기온 및 강수량 증가 등으로 인해 모기 개체 수가 증가함에 따라 말라리아, 일본뇌염, 뎅기열 등 모기를 통해 전파되는 질병에 감염병의 위험률도 높아지고 있어 머신러닝기반 기후 데이터를 활용하여 모기 개체 수를 예측할 수 있는 모델을 제안하였다.

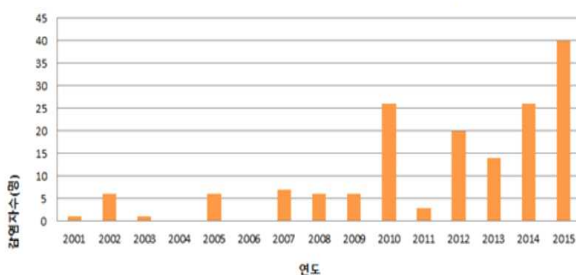
1. 서론

최근 기후변화로 인한 기온상승으로 인해 모기의 출현시기가 더 빨라지고 있으며, 서식지까지도 넓혀가고 있어 문제가 되고 있다. 기후변화가 모기에게 까지 영향을 미치는 것이다.

기후변화는 온도, 강수량의 기상요소 변화뿐만 아니라 모기의 성장과 발달에 영향을 줘 모기 감염병 위험을 증가시키기도 한다.

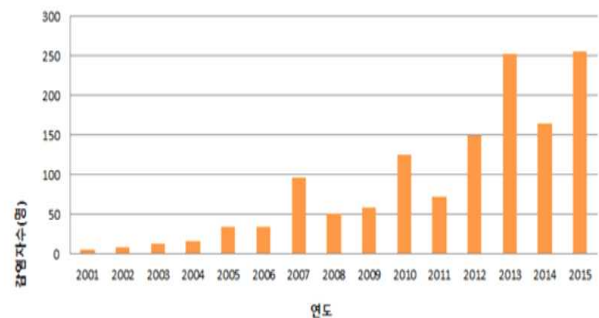
기후 변화로 인한 모기 감염병을 예방하기 위해 각 나라에서는 모기 활동을 예측하려는 노력을 펼치고 있다. 미국은 남동지역의 기후학자, 곤충학자, 공중 보건학자 전문가들이 공동연구를 수행해 모기 매개 질환의 위험을 예측하고 감시하는 시스템을 개발했고, 일본은 모기 채집 자료에 기상자료를 접목하여 모기 활동을 예측하는 모델을 개발하였다.

연도별 일본뇌염 감염자수(명)



(그림 1) 연도별 일본뇌염 감염자수




연도별 뎅기열 감염자수(명)



(그림 2) 연도별 뎅기열 감염자수

우리나라도 모기 감염병에 대한 선제적인 대응을 위해 기상자료와 지리정보를 활용한 모기 예측기술을 개발에 더욱 박차를 가해야 한다고 전문가들은 입을 모으고 있다.

국내 및 해외에서 감염될 수 있는 모기 매개 감염병에는 황열, 뎅기열, 웨스트나일열, 치쿤구니아열, 지카 바이러스, 말라리아, 일본뇌염 등이 존재하고, 이들은 발열, 오한, 두통, 구역, 빈혈 등을 일으킬 수 있다. 모기 매개 감염병은 심하면 사망에 다다를 수 있는데 황열은 5 ~ 50%, 뎅기열은 5%, 웨스트나일열은 신경계 감염을 일으킨 경우 약 10% 등 모기 매개 감염병에 의한 사망은 1년에 725,000명이라고 한다.

			
	지카바이러스	말라리아	일본뇌염
발생원	이집트숲모기, 아시아안 환충송모기	얼룩날개 모기류 암컷 모기	작은빨간집모기
증상	발열, 발진, 관절통, 눈 충혈, 근육통, 두통, 안구통 신생아 소두증	반복되는 오한, 두통, 구역, 발열, 빈맥, 빈호흡발한, 빈혈	고열(39~40도), 두통, 현기증, 구토, 복통, 지각 이상, 의식장애
잠복기	감염 모기에 물린 후 3~7일이 지나 증상 시작, 최대 잠복기는 2주	열대열, 삼일열 말라리아는 48시간	4~14일
예방법	백신 개발 중이기 때문에 지카 바이러스 발병국 여행 자제 (중남미, 동남아시아, 아프리카, 오세아니아 등)	중동, 중부아프리카, 중남미 등 말라리아 위험 국가로의 여행 시, 항말라리아제 복용	예방접종 필수, 야외 활동 시 모기에 물리지 않도록 주의

(그림 3) 모기가 물고 오는 대표적 질병들

위와 같은 자료를 기반에 두어 기상자료를 이용한 모기 개체수를 예측하는 머신러닝 모델을 만들고자 하였다.

2. 본론

Jupyter Notebook을 사용하여 머신러닝을 수행하였고, 회귀모델로 결과를 도출하였다. 여기서 머신러닝은 기존 데이터를 이용하여 아직 일어나지 않은 미지의 일을 예측하기 위해 만들어진 기법으로 데이터와 결과를 이용하여 특성과 패턴을 찾아내고 찾아낸 패턴을 이용하여 새로운 데이터에 대한 결과를 예측(추론)하는 것이다. 회귀모델이란 어떤 연속형 데이터 Y와 이 Y의 원인이 되는 X간의 관계를 추정하기 위해 만든 관계식이다.

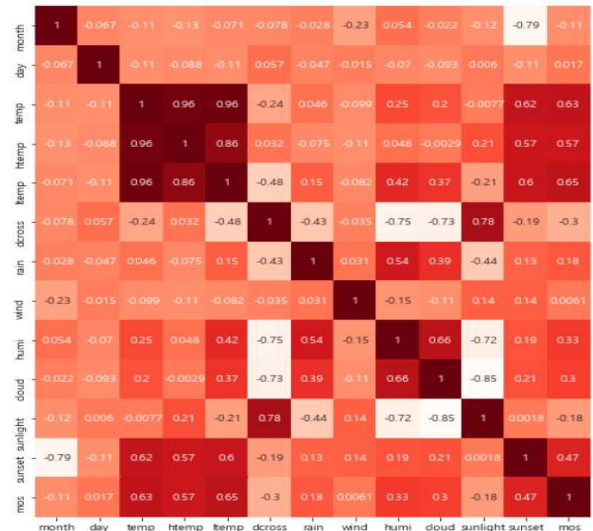
기상청에서 수집한 기상자료는 일별 평균기온(°C), 최고기온(°C), 최저기온(°C), 강수량(mm), 풍속(m/s), 습도(%), 전운량(1/10), 일조시간(h), 일몰시간(time)을 사용하였다.

date	month	day	temp	ltemp	htemp	dcross	rain	wind	humid	cloud	sunlight	sun
2017-04-17	4	17	14.7	18	11.2	6.8	5.5	2.2	67	8.2	0.7	1909
2017-04-18	4	18	11.7	17.4	9.4	8	7.5	2.5	77	5.2	2.7	1910
2017-04-19	4	19	12.6	18.4	7	11.4	0	2.7	43	2	12.3	1911
2017-04-20	4	20	12.9	18.2	9.9	8.3	0.1	2.4	45	7.4	2.6	1912
2017-04-21	4	21	13.1	19.3	7.6	11.7	0	1.9	57	3.1	8.5	1913
2017-04-22	4	22	14.8	22.5	8.5	14	0	2	48	0.5	12.4	1914
2017-04-23	4	23	14.9	21.7	8.9	12.8	0	2.6	40	0.1	12.6	1915
2017-04-24	4	24	15.1	21.3	8.3	13	0	2.5	39	4.8	12.1	1915
2017-04-25	4	25	16	20.4	12.9	7.5	0	2.8	57	7.9	4.7	1916

(그림 4) 수집된 기상자료

모기 채집량은 서울특별시 홈페이지에서 데이터를 수집하였으며, 모기 채집량은 DMS(Digital Mosquito monitoring System)을 사용하여 측정되었다.

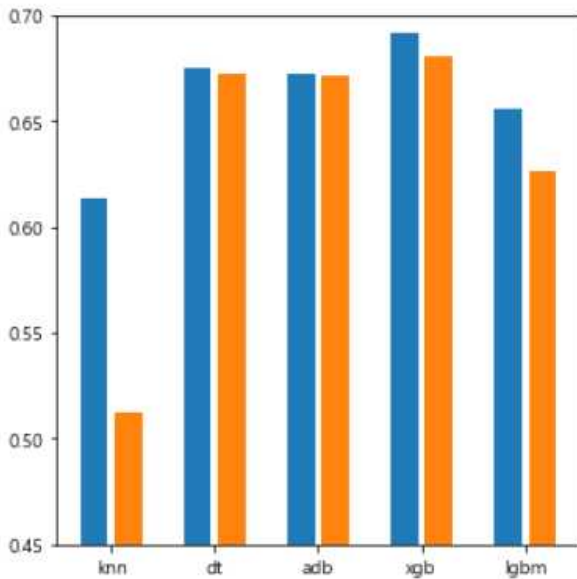
DMS란 이산화탄소를 유인제로 하여 암모기만 선택적으로 유인한 후 개체수를 자동으로 계측하는 장비로, 서울시 25개 구에 각각 2개씩 배치되어 50개소의 모기 채집량을 통합한 데이터가 제공되어 있다.



(그림 5) 연도별 일본뇌염 감염자수(명)

(그림 5)은 상관관계를 나타낸 표이다. 상관관계란 일정한 수치로 계산되어 두 대상이 서로 관련성이 있다고 추측되는 관계를 말한 것이다. 0에 가까울수록 대상 간의 관련성이 없음을 나타낸다.

(그림 5)을 보면 평균기온, 최고기온, 최저기온, 일몰시간이 연관이 높은 것을 볼 수 있다. 기온은 모기가 성충이 되는 시간을 변화시킨다. 기온이 높을수록 성충이 되는 기간이 단축되고, 이에 따라 모기의 개체수도 증가한다[1]. 특히 말라리아 환자가 서울, 경기, 인천, 강원에서 기온이 1°C 상승하면 발생 위험이 10.8%, 12.7%, 14.2%, 20.8% 증가하는 것으로 보고된 바 있다.



(그림 6) 모델별 예측결과표

(그림 6)은 위의 데이터를 사용하여 예측한 결과이다. 사용한 모델은 순서대로 KNN, Decision Tree, AdaBoost, XGBoost, LGBM이다. 그 외에도 LinearRegrassion, RandomForest 등을 사용하였으나 대표적인 모델 5개만 표로 나타내었다.

linear

not linear

모기 개체 수 예측하기

평균 온도 <input type="text" value="평균 온도를 입력하세요."/>	풍속 <input type="text" value="풍속을 입력하세요."/>
최고 온도 <input type="text" value="최고 온도를 입력하세요."/>	습도 <input type="text" value="습도를 입력하세요."/>
최소 온도 <input type="text" value="최대 온도를 입력하세요."/>	전운량 <input type="text" value="전운량을 입력하세요."/>
강수량 <input type="text" value="강수량을 입력하세요."/>	일조시간 <input type="text" value="일조시간을 입력하세요."/>
<input type="button" value="예측하기"/>	

(그림 7) 웹페이지를 이용한 결과 값 예측

이 중 가장 정확도가 높은 XGBoost를 사용하여 수집한 데이터를 입력 값으로 입력하게 되면 이에 대한 결과 값이 웹페이지로 반환되어 데이터에 맞는 모기 개체수를 예측할 수 있게 된다.

3. 결론

기상자료만을 가지고 모기 개체수를 예측하는 학습률을 70% 가까이 끌어올릴 수 있었고, 지리정보를 추가한다면 더욱 높은 학습을 끌어올릴 수 있을 것이라고 예측된다.

또한 자료에서는 강수량과 모기 개체수의 관계가 낮은 상관관계를 보이고 있는데, 이는 유충이 성충으로 되기까지의 5 ~ 14일의 시간이 필요하기 때문에 실제와 다른 값을 나타내어 낮은 상관관계를 나타낸 것으로 유추된다. 이를 활용한다면 더 높은 학습률을 끌어올릴 수 있을 것이다.

“본 연구는 과학기술정보통신부 및 정보통신기획평가원의 Grand ICT연구센터지원사업의 연구결과로 수행되었음”(IITP-2020-0-01489)

참고문헌

- [1] 채수미, 김동진, 윤석준, “기온과 지역특성이 말라리아 발생에 미치는 영향”, 보건사회연구 34(1) 436-455, 2014
- [2] 권형욱. “매개곤충 연구의 현재와 미래: 모기 연구의 융복합연구.” 한국응용곤충학회 학술대회는 문집 2016.10 (2016): 17-17.
- [3] 한창호, 윤용태, 이집호, 이정미, 최성민. “서울 지역에서 채집한 모기 분포 조사(2008~2011).” 대한 보건협회 보건종합학술대회 2011.- (2011): 99-99.