

기온 분석을 통한조류 인플루엔자 바이러스 감염 발생 예측 모델

두건

성균관대학교 소프트웨어대학

A model for predicting the avian flu virus infection through temperature analysis

Chien Tu

Sungkyunkwan University

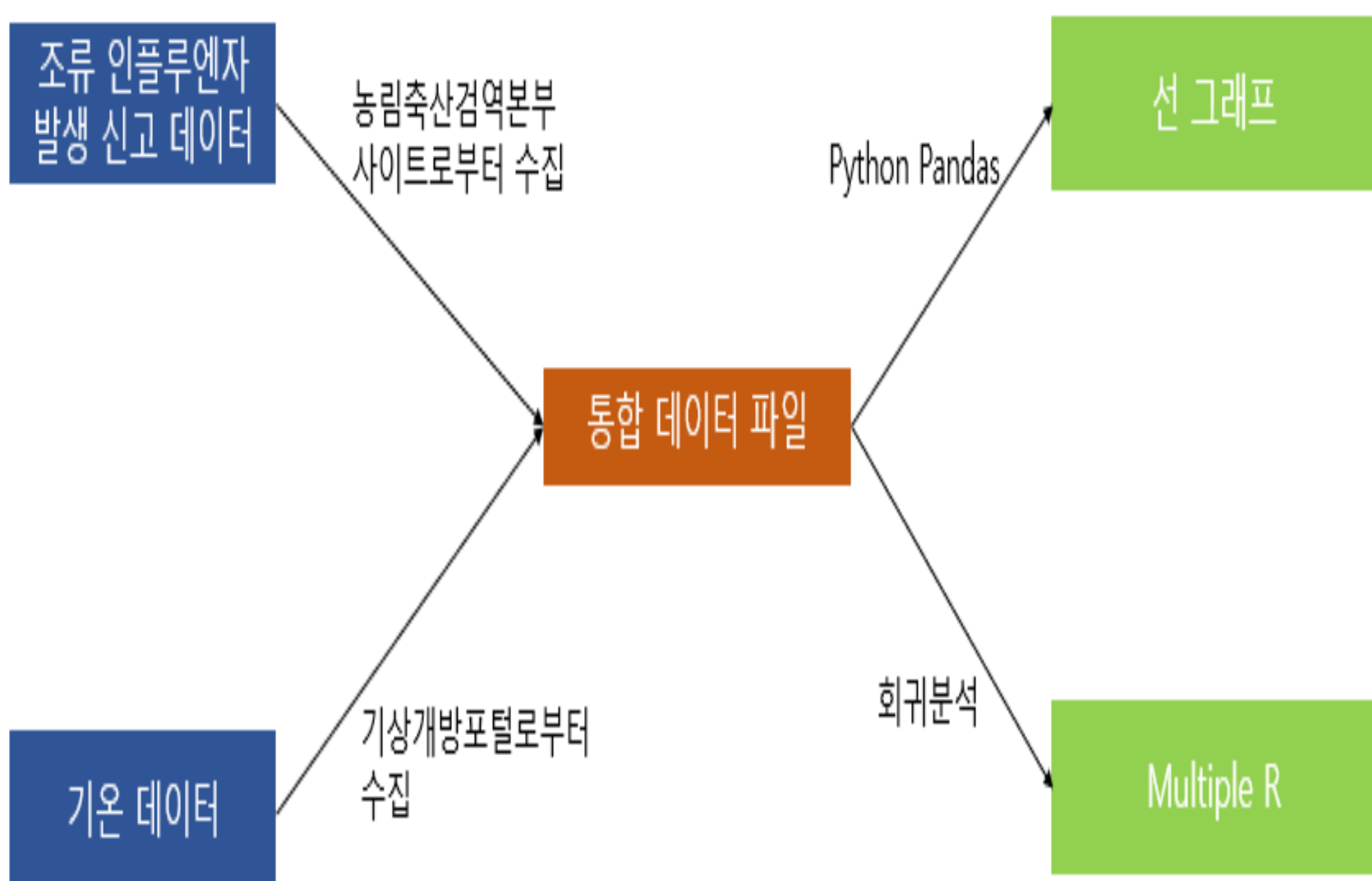
지도 교수: 김응모 교수님

연구실명: 데이터베이스 연구실

요약

본 논문에서는 조류 인플루엔자 발생 지역의 기온에 따른 조류 인플루엔자 바이러스 발생의 상관관계를 밝혀내고자 한다. 조류 인플루엔자 바이러스 발생에 영향을 끼치는 외부 환경 요인으로서는 기온, 습도, 강수량 등을 들 수 있으나, 본 논문에서는 수집 데이터의 한계로 인해 기온과의 관계만 분석한다. 본 논문에서는 조류 인플루엔자 발생 지역과 해당 지역의 기온 데이터의 연관성을 수치화하는 회귀분석 기법을 이용해 조류 인플루엔자와 외부 기온의 상관관계를 파악, 이를 통해 (1) 조류 인플루엔자 바이러스와 기온의 상관관계 측정, (2) 측정을 통해 향후 연구 혹은 활용방안으로써 활용할 수 있을 것이다.

시스템 구성



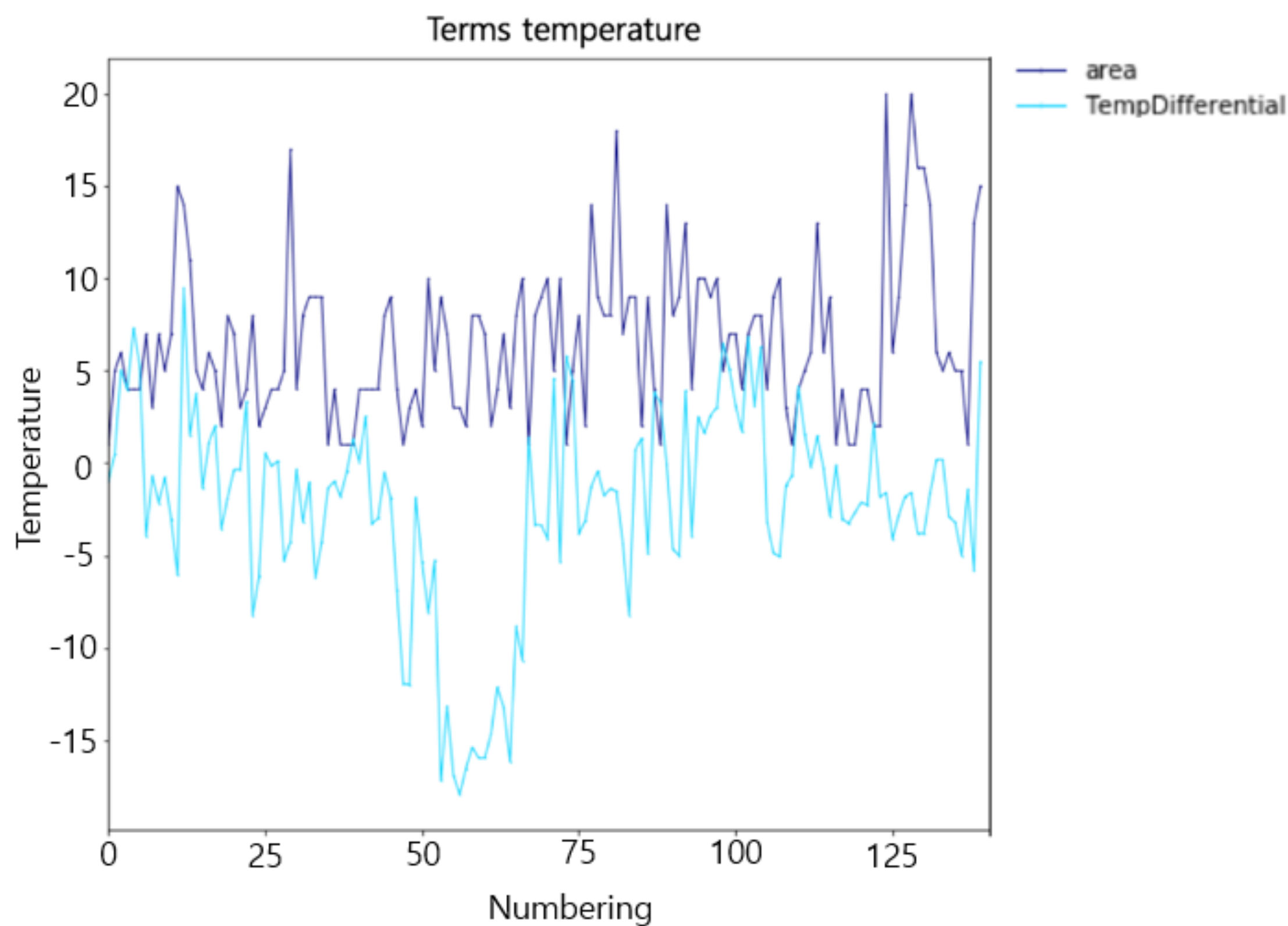
조류 인플루엔자 발생 신고 데이터	농림축산검역본부로부터 2014년 4월 ~ 2017년 6월까지 일 단위의 조류 인플루엔자 발생 신고 데이터를 수집하였다.
기온 데이터	기상개방포털로부터 조류 인플루엔자 발생 신고 데이터와 상응하는 날짜, 지역의 기온 데이터를, 그리고 당시 전국 평균 기온 데이터를 수집하였다.
데이터 통합 파일	조류 인플루엔자 발생 신고 데이터로부터 신고 지역 및 신고 순서를, 기온 데이터로부터 해당 지역 당시 기온과 당시 전국 평균 기온 값을 하나의 데이터 파일로 통합 저장하였다.
데이터 분석을 위한 tool	통합된 데이터 파일을 분석하기 위해 선 그래프와 회귀분석 기법을 활용하였다. 선 그래프의 경우, Anaconda의 Jupyter를 통해 Safari 브라우저에서 Python 2.7.14 언어를 활용, Pandas 라이브러리를 이용해 그래프를 작성하였다. 회귀분석의 경우, MS 엑셀 파일의 분석도구 기능을 활용하였다.

데이터 분석 방법

1. 데이터 분석은 시각적 분석을 위한 선 그래프, 그리고 객관적 수치를 얻기 위한 회귀분석 기법을 활용하였다.
2. 선 그래프는 Python 언어의 Pandas 라이브러리를 이용하였다. 사전에 작성한 데이터 파일(csv 파일)을 read하여 plot 함수로 그래프를 작성하였다.
3. 회귀분석은, 데이터 파일을 MS 엑셀 파일의 데이터 분석 기능을 통하여 Multiple R 값을 도출하였다.

결과

- 선 그래프를 통해 시각적 데이터를 얻었으나 연관성을 명확히 알 수 없었다.
- 조류 인플루엔자와 기온 간의 연관성을 객관적 수치인 회귀분석 기법의 Multiple R 값으로 알아내고자 하였다. Multiple R 값이 0.445 즉 44.5%로 계산되었고 두 요소의 연관성이 있음을 결론으로 도출하였다.



결론

본 논문에서는 조류 인플루엔자 발생 신고 데이터와 기온과의 상관관계를 분석하였다. 발생 신고 지역 데이터와 지역 기온 데이터, 그리고 전국 평균 기온 데이터를 정리하여 데이터 파일에 통합하였고, 회귀분석을 통해 데이터 간의 연관성을 multiple R, 즉 44.5%라는 객관적인 수치로 증명하여 결론을 도출하였다.
향후, 교차검증과 같이 더욱 다양한 기법, 그리고 습도 등의 다양한 주변 환경 데이터를 기반으로 분석하여 나은 결론을 도출할 수 있을 것으로 기대된다. 이를 이용해 본 논문의 결론을 보강할 계획이다.