<날씨 정보 시각화>

20181288 윤성호

날씨 정보를 확인해야 하는 상황은 아주 많다. 그러나, 날씨 정보 중 얻고 싶은 정보가 정리되어 있는 경우도 있지만 아닌 경우도 있다. 따라서, 이번 프로젝트에서는 정리되어 있는 정보를 가져와 시각화 하는 것은 물론, 정리되어 있지 않은 정보를 시각화해 의미있는 날씨분석자료를 얻는 것을 목표로 한다.

날씨 정보를 받아온 후 처리하기 위해, Weather 클래스를 하나 생성한 후, 그 아래 매서드 형태로 원하는 기능을 구현하는 방식으로 프로그램을 작성하려고 한다. 또한 날씨 정보는 서울을 기준으로 받아오려고 한다.

<날씨 데이터 받아오기>

날씨 데이터는 “기상청 날씨누리” 사이트에서 제공하는 날씨 정보를 웹 크롤링으로 긁어온 후, 정제하여 사용하고자 한다. 이를 위해 requests 라이브러리와, beautifulSoup4 라이브러리를 이용하려고 한다. 또한, 뽑아온 데이터에서 필요한 부분만을 골라내기 위해서 정규식을 이용할 계획이다.

<Weather 클래스의 매서드 : 오늘의 날씨 출력> 끝!

“기상청 날씨누리” 에서 현재시간부터 24시간이후까지의 날씨 데이터를 받아온 후, numpy 라이브러리를 이용해 기온, 체감온도, 강수확률 등의 정보를 저장한 후 matplotlib 라이브러리를 이용하여 그래프를 그려서 날씨의 추이를 알아보려 한다. 그래프로 시각화해 나타낸다면 변화를 한눈에 알아볼 수 있을 것 같아 구현하게 되었다.

<Weather 클래스의 매서드 : 미래의 날씨 출력> 끝!

“기상청 날씨누리” 에서는 오늘뿐만 아니라 내일, 모레, 글피 날씨의 예측정보까지 제공하고 있다. 입력으로 내일, 모레, 글피 중 하나를 받으면, 해당 날의 기온, 체감온도, 강수확률 등의 정보를 저장한 후 matplotlib 라이브러리를 이용하여 그래프를 그려서 추이를 알아보고자 한다.

<Weather 클래스의 매서드 : 과거 날씨 출력> 끝!

년월일을 입력하면 해당 년원일에 해당하는 날씨정보를 받아와서 평균 기온, 최고 기온, 최저 기온, 강수량 등의 정보를 화면에 출력해준다. 이 때, 유효하지 않은 날짜가 입력된다면 예외처리를 통해 에러 메세지를 출력할 수 있도록 구현하려고 한다.

<Weather 클래스의 매서드 : 과거의 오늘 날씨 출력> 끝!

“기상청 날씨누리” 에는 1960년부터 현재까지의 날씨 정보를 제공하고 있다. 82년정도의 데이터가 쌓여있는 것이다. n월 m일을 입력하면, 1960년부터 2021이나 2022년의 모든 n년 m월 날씨 데이터를 받아와서, 같은 날짜의 날씨 정보가 어떻게 변화하는 지 matplotlib을 통해 시각화 해보고자 한다. 유효하지 않은 날짜 입력이 들어오면 예외처리를 통해 에러 메세지를 출력하게 한다.

<Weather 클래스의 매서드 : 연평균기온 추이(numpy\_version)> 끝!

“기상청 날씨누리” 에서 1960년부터 현재까지의 날씨 정보를 모두 받아온다. 이를 numpy의 ndarray를 활용하여 처리한다. 각 년도의 평균기온을 받아온 데이터를 실제 연산하여 구하고, matplotlib을 이용해 시각화한다. 또한, 실행시간을 측정해본다. 매년 365일이 있는데 약 82년동안 쌓여있으므로 365\*82 = 29930 개의 기온에 대해 평균을 계산해야한다. 이 매서드를 작동시키는데 드는 실행시간도 확인해본다.

---------------------------------------- proposal 기능 추가 -------------------------------------------

<Weather 클래스의 매서드 : 최고기온, 최저기온> 끝!

입력으로 두 날짜를 입력한다. 두 날짜 사이의 평균기온들을 탐색하면서 주어진 기간동안의 날씨 데이터를 matplotlib을 이용해 시각화하고, 최고기온이 측정된 날짜로, 최저기온이 측정된 날짜를 출력해준다.

<Weather 클래스의 매서드 : 내일 날씨 예측을 통한 기온별 옷차림 추천>

날씨를 예측하는 일은 굉장히 어려운 일이다. 초기조건의 약간의 변화만으로도 엄청난 결과의 차이가 발생할 수 있기 때문이다. 그러나 결과값이 유사할 가능성을 높이기 위해서는 초기조건의 차이가 유사해야한다는 것은 알 수 있다. 정확한 예측은 쉽지 않지만, 초기조건의 유사도를 분석하면 결과값도 어느정도 예측은 가능하다는 의미이다. 이번 매서드에서는 내일의 날짜에 해당하는 과거 년도의 날씨자료를 분석하여, 내일의 날씨를 예측한다. 그리고 예측한 기온 및 강수확률을 바탕으로 입을만한 옷차림을 추천해준다. 즉, 오늘을 기준으로 과거 6개월전의 날씨 데이터를 받아온 후, 과거의 오늘 날짜로부터 6개월 전의 날짜를 모두 불러온다. 기상청에는 82년간의 데이터가 쌓여있기 때문에 81개의 과거 데이터를 얻을 수 있다. 이러한 과거 데이터와 현재 데이터간의 유사도를 분석하여 제일 유사도가 높은 3개의 날씨를 선택하고 가중치를 고려하여 날씨의 평균기온을 구하여 내일의 기온을 출력해준다. 유사도 측정 방법은 세가지로 진행 할 예정이다.

1. Euclidean Similarity

81개의 과거 데이터와 1개의 올해 데이터를 비교하는데, 각 날짜의 평균기온의 차이를 모두 합한 것을 유사도로 정의하고, 유사도가 가장 큰 3개의 년도에서의 내일 날짜의 평균기온을 유사도를 가중치로 두어 평균을 구해 기온을 예측한다.

1. Cosine similarity

81개의 과거 데이터와 1개의 올해 데이터를 비교하는데, 1차원 벡터의 관점에서 각도의 차이의 코사인의 절댓값으로 유사도를 정의한다. 이렇게 정의된 유사도를 계산하여 유사도가 가장 큰 3개의 년도에서의 내일 날짜의 평균기온을 유사도 가중치를 두어 평균을 구하여 기온을 예측한다.

1. 내가 정의한 similarity

날씨 데이터는 조금의 차이도 결과적으로 큰 차이를 만들어낸다. 이러한 날씨의 특성을 고려하며 차이에 더 가중치를 두어서 유사도를 계산하고자 한다. 이를 위해 평균기온의 차이의 제곱을 모두 합한 것을 유사도로 정의한다. 이렇게 정의하면 Euclidean similarity에서 계산한 유사도 탑 3는 변화가 없지만, 가중치에 변화가 생겨서 위의 두개의 유사도보다 더 날씨의 특성에 부합한 예측이라고 생각한다.