

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN MÔN HỌC
TOÁN ỨNG DỤNG VÀ THỐNG KÊ
ĐỀ TÀI: Bài toán khí hậu
Bài toán 1: Mô tả dữ liệu về lượng mưa

Giảng viên lý thuyết: PGS.TS Nguyễn Đình Thúc

Lớp: 20TN

Thành viên thực hiện:

- 20120131 – Nguyễn Văn Lộc
- 20120536 – Võ Trọng Nghĩa
- 20120572 – Nguyễn Kiều Minh Tâm

THÀNH PHỐ HỒ CHÍ MINH, THÁNG 4 NĂM 2022

Lời nói đầu

Hiện nay, biến đổi khí hậu đang là một chủ đề nhận được sự quan tâm trên toàn thế giới. Nhiều tổ chức liên chính phủ như Ủy ban Liên chính phủ về Biến đổi Khí hậu (IPCC) được lập ra nhằm tìm kiếm những giải pháp cho tình trạng này. Trong đó, vấn đề về lượng mưa nổi lên như một trong những vấn đề bức thiết nhất. Nhận thấy được tầm quan trọng này, nhóm chúng em quyết định làm bài toán về mô tả dữ liệu lượng mưa, nhằm cung cấp một cái nhìn tổng quan về phân bố lượng mưa tại một trạm khí tượng.

Thành phố Hồ Chí Minh, tháng 4 năm 2022

Mục lục

1	Đặt vấn đề	3
1.1	Xác định và hình thức hóa mục tiêu của bài toán	3
1.2	Dạng bài toán	3
1.3	Đối tượng được chọn cho bài toán	3
1.4	Phạm vi, mức độ, quy mô của bài toán	3
2	Thu thập và xử lý dữ liệu	3
2.1	Thu thập dữ liệu	3
2.2	Xử lý dữ liệu	3
2.2.1	Trích xuất dữ liệu	3
2.2.2	Xử lý phần dữ liệu bị khuyết	3
3	Phân tích, đánh giá và kết luận	4
3.1	Kết quả xử lý	4
3.2	Mô tả từ dữ liệu	5

Danh sách hình vẽ

1	Biểu đồ đường	4
2	Biểu đồ phân tán	5
3	Giá trị mode	5
4	Giá trị median	5
5	Giá trị độ lệch chuẩn	5

Danh sách bảng

1 Đặt vấn đề

1.1 Xác định và hình thức hóa mục tiêu của bài toán

Bài toán: Phân tích dữ liệu về lượng mưa trung bình năm từ năm 1886 đến năm 2021 ở trạm Concordia Muni AP, Kansas, miền Tây Hoa Kỳ.

Hình thức hóa mục tiêu của bài toán: vẽ biểu đồ histogram, xác định các giá trị thống kê mẫu của dữ liệu rồi đưa ra nhận xét.

1.2 Dạng bài toán

Bài toán này thuộc dạng bài toán **Mô tả** (nêu ra các đặc trưng của dữ liệu).

1.3 Đối tượng được chọn cho bài toán

Dữ liệu về lượng mưa trung bình năm của trạm Concordia Muni AP.

Cột PRCP (cột BU) của tập tin **USW00013984.csv** trong thư mục **gsoy-latest**, lấy từ [dataset về Global Summary of the Year của NOAA](#).

1.4 Phạm vi, mức độ, quy mô của bài toán

Theo không gian: dữ liệu được xử lý trong bài toán đại diện cho bang Kansas của Hoa Kỳ.

Theo thời gian: dữ liệu được thu thập trong vòng 136 năm, từ năm 1886 đến năm 2021.

2 Thu thập và xử lý dữ liệu

2.1 Thu thập dữ liệu

Dữ liệu xử lý trong bài toán này được thu thập từ dữ liệu của Cơ quan Quản lý Khí quyển và Đại dương Quốc gia Hoa Kỳ (NOAA), phần dữ liệu tổng hợp theo năm (Global Summary of the Year).

2.2 Xử lý dữ liệu

2.2.1 Trích xuất dữ liệu

Dữ liệu lượng mưa từ tập tin **USW00013984.csv** được trích xuất thành tập dữ liệu **data.csv** nhờ các hàm của module **pandas**.

2.2.2 Xử lý phần dữ liệu bị khuyết

Dữ liệu từ tập tin gốc bị khuyết khoảng thời gian 14 năm từ năm 1949 đến năm 1962, do đó, nhóm chúng em đã tìm cách "điền" vào những ô bị khuyết này. Lượng mưa bị khuyết ở năm thứ i được tính bằng trung bình cộng của lượng mưa 2 năm trước đó,

tức là trung bình cộng của năm thứ $i - 1$ và năm thứ $i - 2$.
 Đoạn code sau là hàm xử lý đoạn dữ liệu bị khuyết:

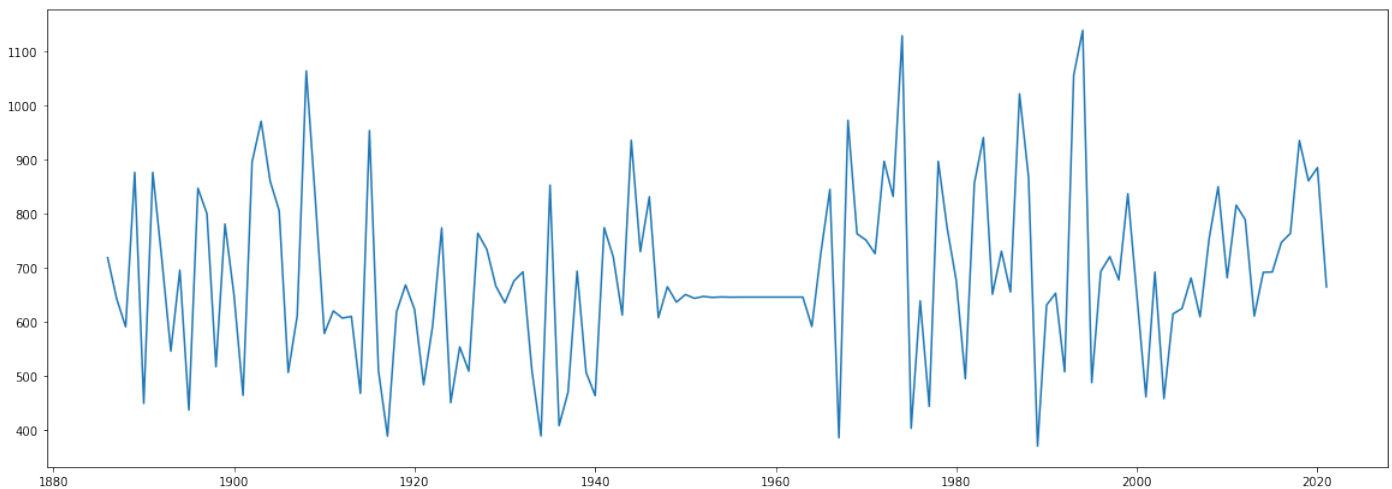
```
def fill_missing(data_list, begin, end):
    for i in range(begin, end + 1, 1):
        sum = 0
        for j in range(i - 2, i, 1):
            sum += data_list[j]
        data_list[i] = round(sum / 2, 1)
```

3 Phân tích, đánh giá và kết luận

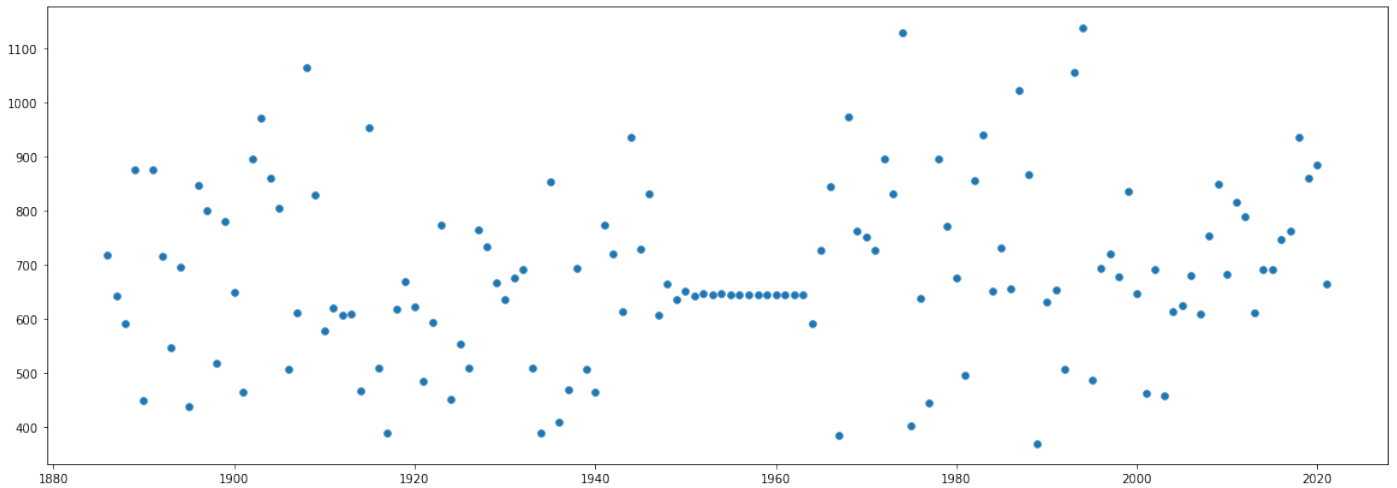
Sau khi được tiền xử lý, dữ liệu được trực quan hóa (visualized) bằng các hàm hỗ trợ của module `matplotlib`. Bên cạnh đó, các số liệu thống kê liên quan đến dữ liệu được tính bằng các hàm của module `numpy`.

3.1 Kết quả xử lý

Dưới đây là hai biểu đồ thể hiện tập dữ liệu trên: biểu đồ đường và biểu đồ phân tán.



Hình 1: Biểu đồ đường



Hình 2: Biểu đồ phân tán

Sau đây là kết quả tính toán mode và median của dữ liệu.

Mode: [645.2]

Hình 3: Giá trị mode

Mean: 684.3316176470589

Hình 4: Giá trị median

Standard deviation: 160.43275212961458

Hình 5: Giá trị độ lệch chuẩn

3.2 Mô tả từ dữ liệu

Ngoại trừ khoảng thời gian số liệu bị khuyết (số liệu của 14 năm, từ năm 1949 đến 1962), đã được tiền xử lý như trình bày trong phần trên, nhìn chung giá trị lượng mưa trung bình năm ở trạm Concordia Muni AP, Kansas, miền Tây Hoa Kỳ có sự biến động liên tục qua từng năm với những khoảng biến động tương ứng tương đối lớn (trên 300 mm) trong trung bình 2 lần trong 20 năm liên tiếp.

Lượng mưa tại trạm này được tính trung bình trong khoảng 684.3 mm. Con số là xấp xỉ so với lượng mưa trung bình năm 2017 của Hoa Kỳ là 715mm (số liệu từ <https://en.>

wikipedia.org/wiki/List_of_countries_by_average_annual_precipitation). Độ lệch chuẩn của dữ liệu lớn, dữ liệu phân bố không đều, có những năm lượng mưa đạt hơn 1100 mm, nhưng có những năm chỉ đạt ít hơn 400mm. Lượng mưa trung bình mỗi năm ở trạm này trong khoảng thời gian 1886 đến 2021 là lớn hơn 350 mm, chứng tỏ vùng đất này không bị hạn hán triền miên.

Có thể thấy, mặc dù trạm này không bị hạn hán triền miên, nhưng lượng mưa phân bố không đều, có thể gây ảnh hưởng nghiêm trọng đến các hoạt động sản xuất, sinh hoạt của người dân nơi đây. Chính phủ sở tại cần đưa ra giải pháp nhằm giải quyết vấn đề này.

Lời cảm ơn

Trong quá trình thực hiện đồ án này, chúng em đã nhận được những bài giảng tận tâm, chi tiết về các phương pháp của Thầy Nguyễn Đình Thúc. Bên cạnh đó là những sự hướng dẫn của các Thầy, Cô trợ giảng cho môn học Toán ứng dụng và thống kê. Chúng em xin cảm ơn các Thầy, Cô.

Ngoài ra, chúng em cũng xin cảm ơn các bạn trong lớp đã đưa ra những chỉ dẫn, góp ý trong quá trình chúng em thực hiện đồ án.

Thành phố Hồ Chí Minh, tháng 4 năm 2022