CS886 Final Project

Topic: Weather forecast in Hong Kong

ToDo:

* Concatenate dataset, choose time, select features and stations
* Objective of the project
* Start writing the paper
* Build a simple pipeline for the project (data preprocessing, model training, evaluation)

Objective:

* Predicting temperature next day?
* Predict temperature for selected stations in Hong Kong?

Challenge:

* Data Engineering and Transformation
  + No established dataset with relevant data for temperature forecast for each region
  + There are a lot APIs that provide different attributes in different formats (hourly, daily)
* Feature Engineering
  + Some important attributes are only measured in limited amount of stations
  + Most stations have a limited set of attributes

Data Preprocessing

* Select objective (Daily temperature prediction for 1/ multiple stations)
* Concatenate the dataset
* Choose relevant features, regions with enough data
  + Feature can share across nodes
  + If only a few regions miss the data, impute it with mean of other regions
* Encode the location attribute
  + Currently each station has similar lat and lon
  + Distinguish coastal areas, urban centers, and hilly regions

Coding:

* Choose region for prediction
* Connect API and download necessary datasets
* Concatenate different features according to date-time
* Convert the meteorological features, spatial features (location->nodes) and edges (with feature distance) into graph
* Visualise the graph
* Exploratory data analysis for graph features (?)

Paper Writing

* Structure (parts)
* Abstract
* Introduction

Research and literature review:

* Important features contribute to forecast
* Typical ML methods for temperature forecast, and methods used by HK Observatory (try to find baseline)
* How to encode spectral (location) and temporal (time, date) data
* Alternate method for forming edge (and features) in temperature forecast task
* EDA for graph features

Useful Dataset

|  |  |  |
| --- | --- | --- |
| Dataset | Link | Remarks |
| Daily maximum, mean and minimum temperatures | https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-temperature-info-hko |  |
| Daily total rainfall | https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-total-rainfall |  |
| Daily maximum and mean Hong Kong Heat Index | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-maximum-mean-heat-index> | * 1 region |
| Daily mean relative humidity | https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-mean-relative-humidity |  |
| Daily mean wind speed | https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-mean-wind-speed |  |
| Daily total bright sunshine (hours) | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-total-bright-sunshine> | * 1 region |
| Daily mean wet bulb temperature | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-mean-wet-bulb> | * 5 regions |
| Daily global solar radiation | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-global-solar-radiation> | * 2 regions |
| Daily prevailing wind direction | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-prevailing-wind-direction> | * 1 region |
| Daily mean pressure | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-mean-pressure> | * 12 regions |
| Daily mean dew point temperature | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-mean-dew-point> | * 1 region |
| Daily maximum and mean UV indices | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-maximum-mean-uv-index> | * 1 region |
| Daily mean amount of cloud | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-mean-amount-of-cloud> | * 1 region |
| Daily total evaporation | <https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-total-evaporation> | * 1 region |

Features:

- Hourly heights of astronomical tides

- Times and heights of astronomical high and low tides

- Times of sunrise, sun transit and sunset

- Times of moonrise, moon transit and moonset

- Gregorian-Lunar calendar conversion table

- Cloud-to-ground and cloud-to-cloud lightning count

- Latest 10-minute mean visibility

- Daily Mean Temperature

- Daily Maximum Temperature

- Daily Minimum Temperature

- Weather and Radiation Level Report

Good features but not included:

* Sunshine
* Humidity
* Precipitation
* Humidity
* Geospatial feature (location)
* Extreme event
* Cloud cover
* time