CSCDC Dataset

Overview

The CSCDC (CUIP Smart City Data Challenge) Dataset is a dataset compiled from a month of data generated by the CUIP Testbed. This dataset contains a full month of Video Events (events being objects which pass through the cameras' view points). In addition, the dataset includes 23 days of Air Quality Data (a week shy of the whole month of June) - this is where the challenge takes place. Your challenge is to generate the last week of air quality data as accurately as possible using the prior three weeks and the vehicle event data. How you design your model is completely up to you!

Video Event Format

Video Events includes data such as the event timestamp (in UTC), hit counts (how many times the object was found in this view), label (i.e. car, bus, person), event metadata (such as `pole_id` (`str`), `intersection` (`bool`), and `camera_id` (`str`)), and `locations`. Locations is a string representation of a JSON object that will need to be parsed via your programming language of choice. For example (in Python):

```
import json
.
.
.

for index, row in dataframe.iterrows():
  json_parsed = json.loads(row["locations"])
  # Proceed to iterate over locations, if necessary
```

Then you can refer to the stringified JSON object through your library and programming language of choice. The `locations` object is an array of objects, each of which contains a `timestamp` key (whose value is a UTC timestamp) and a `coords` key (whose value is an array of `float` values representing the `x1, y1, x2, y2` coordinates of the object at that given timestamp). The format looks something like this:

```
[{
  'coords': [737.4640244784002, 504.27940776028055,
967.5611213071068, 617.4017094599545],
  'timestamp': 1559364547782
}, {
  'coords': [795.9374239359277, 520.6466453487443,
1083.5997557290896, 662.3173861308863],
  'timestamp': 1559364547982
}, {
  'coords': [849.109040463825, 532.7213710812812,
1187.8635329313645, 699.1199888926651],
  'timestamp': 1559364548201
}]
```

This array is of **variable length**; some objects may have more objects within this array than others. The length of the array is the same as the `hit_count` value of that event.

H3 Air Quality Format

Air Quality Data includes data points such as `current_dewpoint_f`, `current_humidity`, `current_temp_f`, `lat`, `lon`, `timestamp`, `nicename` (the name of the sensor, this is similar to the `camera_id` of video events so this will help draw a correlation), and actual air quality values. An explanation of those values:

```
Pm1_0_atm: Channel A ATM PM1.0 particulate mass in ug/m3
Pm2_5_atm: Channel A ATM PM2.5 particulate mass in ug/m3
Pm10_0_atm: Channel A ATM PM10.0 particulate mass in ug/m3
Pm1_0_cf_1: Channel A CF=1 PM1.0 particulate mass in ug/m3
Pm2_5_cf_1: Channel A CF=1 PM2.5 particulate mass in ug/m3
Pm10_0_cf_1: Channel A CF=1 PM10.0 particulate mass in ug/m3
P_0_3_um: Channel A 0.3 micrometer particle counts per deciliter of
air
P_0_5_um: Channel A 0.5 micrometer particle counts per deciliter of
air
P_1_0_um: Channel A 1.0 micrometer particle counts per deciliter of
air
P_2_5_um: Channel A 2.5 micrometer particle counts per deciliter of
air
P_5_0_um: Channel A 5.0 micrometer particle counts per deciliter of
air
P_10_0_um: Channel A 10.0 micrometer particle counts per deciliter
of air
```

```
Pm1_0_atm_b: Channel B ATM PM1.0 particulate mass in ug/m3.
Pm2_5_atm_b: Channel B ATM PM2.5 particulate mass in ug/m3
Pm10_0_atm_b: Channel B ATM PM10.0 particulate mass in ug/m3
Pm1_0_cf_1_b: Channel B CF=1 PM1.0 particulate mass in ug/m3
Pm2_5_cf_1_b: Channel B CF=1 PM2.5 particulate mass in ug/m3
Pm10_0_cf_1_b: Channel B CF=1 PM10.0 particulate mass in ug/m3
P_0_3_um_b: Channel B 0.3 micrometer particle counts per deciliter
of air
P_0_5_um_b: Channel B 0.5 micrometer particle counts per deciliter
of air
P_1_0_um_b: Channel B 1.0 micrometer particle counts per deciliter
of air
P_2_5_um_b: Channel B 2.5 micrometer particle counts per deciliter
of air
P_5_0_um_b: Channel B 5.0 micrometer particle counts per deciliter
of air
P_10_0_um_b: Channel B 10.0 micrometer particle counts per
deciliter of air
```

H3 Directory Structure

In this folder, there are two main subdirectories. Below is a mapping of those subdirectories and how what they contain.

```
air_quality_csv/
  ├──all.csv
  ├──mlk-central/
  | ├──aq-mlk-central-2019-06-01.csv
  | ├──aq-mlk-central-2019-06-02.csv
  | ├──aq-mlk-central-2019-06-03.csv
  | .
  | .
  | .
  | └──aq-mlk-central-2019-06-30.csv
  ├──mlk-douglas/
  ├──mlk-georgia/
  ├──mlk-houston/
  ├──mlk-lindsay/
  ├──mlk-magnolia/
  └──mlk-peeples
video_event_csv/
  ├──all.csv
  ├──mlk-central-cam-2
  |   ├──mlk-central-cam-2_2019-06-01.csv
```

```
|   ├──mlk-central-cam-2_2019-06-02.csv
|   ├──mlk-central-cam-2_2019-06-03.csv
|   .
|   .
|   .
|   └──mlk-central-cam-2_2019-06-30.csv
├──mlk-douglas-cam-1
├──mlk-goergia-cam-3
├──mlk-houston-cam-1
├──mlk-lindsay-cam-2
├──mlk-magnolia-cam-2
├──mlk-peeples-cam-1
└──mlk-peeples-cam-3
```

The `all.csv` file in both `air_quality_csv` and `video_event_csv` contains *all* of the data of the subdirectories inside of the one file.Ï

About CUIP

The Center of Urban Informatics and Progress (CUIP) is a new nexus of research and innovation at the University of Tennessee at Chattanooga that will engage experts from across the statewide UT system in cross-disciplinary research focused on solutions to urban challenges. The vision for CUIP centers around research initiatives related to Energy, Mobility, Healthcare, Public Safety, Water, and Waste that will directly or indirectly benefit local citizens while developing methods and models for use around the world. CUIP is housed at The University of Tennessee at Chattanooga as an independent research entity, underscoring Chattanooga's emergence as an innovation hub.

With our particular approach to collaborative and informative research that will benefit the local community, CUIP's focus areas: **Energy, Mobility, Healthcare, Public Safety, Water,** & **Waste** are also designed in alignment with the goals of the Chattanooga Smart Community Collaborative.