SFpark dataset combined with taxi trajectories from Cabspotting

This dataset contains records of the measured parking availability in San Francisco, obtained from the public API of the SFpark project[1]. In addition, by combining the SFpark data with taxi trajectories from the Cabspotting project[2], we simulated parking crowd-sensing with taxis as probe vehicles. For different numbers N of taxis as probe vehicles, the parking availability data is contained in the files **sfpark_filtered_Ntaxis.csv.** The files comprise records for a period of six weeks, from 2013-06-13 until 2013-07-24. The parking availability information was provided per parking segment (also called block face), for a total of 420 parking segments after filtering implausible parking segments. The description of the parking segments is contained in the file **sfpark_filtered_segments.csv.**

If you use this dataset, please refer to our papers where the generation of this dataset is described:

- Bock, F., Attanasio, Y., Di Martino, S., 2017. Spatio-temporal road coverage of probe vehicles: A case study on crowd-sensing of parking availability with taxis. In: Societal Geo-innovation. Lecture Notes in Geoinformation and Cartography. Springer International Publishing, pp. 165-184.
- Bock, F., Di Martino, S., 2017. How many probe vehicles do we need to collect on-street parking information? In: International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). IEEE, pp. 538-543.

sfpark_filtered_Ntaxis.csv

Each line of the files corresponds to the parking situation of a parking segment in the SFpark pilot area at a specific timestamp. The parking segments of the SFpark pilot area were filtered for data plausibility. The crowd-sensing observations were computed as described in our papers. The columns correspond to the following content:

Original SFpark data:

- **timestamp:** Timestamp of the API record rounded to the closest minute
- **segmentid:** ID of the parking segment (also called block face by SFpark) as defined by SFpark
- **capacity:** current total number of parking spaces in the parking segment (capacity may vary over time due to parking restrictions)
- **occupied:** current number of occupied parking spaces in the parking segment

Simulated information for parking crowd-sensing:

- **observedK:** has value 1, if this time instant is observed by the taxis; K corresponds to the Kth repetition of the random sampling of taxis and of their lane choice
- **diffK:** difference between the last occupancy measurement of the taxis and the current number of occupied parking spaces; K like above

sfpark_filtered_segments.csv

Each line of the file corresponds to a parking segment in the SFpark pilot area. The parking segments of the SFpark pilot area were filtered for data plausibility as described in our papers. The columns correspond to the following content:

- **segmentid:** ID of the parking segment (also called block face by SFpark) as defined by SFpark
- **streetname:** Name of road with house number range
- **startx, starty, endx, endy:** WGS84 coordinates of start and end point of parking segment

[1] http://sfpark.org/

[2] https://crawdad.org/epfl/mobility/20090224/