Extending prediction of parking occupancy to untracked city areas using city background information

6th March 2017

1 Abstract

Several smart cities around the world have begun monitoring parking areas in order to predict free spots and help drivers that are looking for parking. The current results are indeed promising, however this approach is limited by the high cost of sensors that need to be installed throughout the city in order to achieve an accurate prediction rate. This work investigates the extension of forecasted parking information from areas equipped with sensors to areas that are missing them. To this end, similarity values between city neighborhoods will be computed based on background data, e.g. from geographic information systems. Using the derived similarity values, the adaptation of occupancy rates from monitored- to unmonitored parking areas will be analysed.

2 Technical details

2.1 Method

1. **Cluster all blocks into compact areas**

   To start with, k-means to group individual block faces (see figure 1) into a convenient number of clusters. Only the geographic position and (Euclidean/Manhattan) distance are needed for clustering. Monitored and unmonitored areas will be clustered separately.

2. **Calculate pairwise similarity between unmonitored- and monitored clusters**

   Per cluster, vectors with the following dimensions will be built using GIS attributes:

   - count of office buildings × their capacity/size
   - count of cinemas/theaters/concert halls × their capacity/size
   - count of restaurants/fast food × their capacity/size
   - (other types of buildings that create parking demand)

Figure 1: Blocks in a city, with each block displaying four faces for each street[1]

Cosine Similarity will be used further. In general, between vectors A and B with n dimensions, it is defined as:

$$\text{similarity}(A, B) = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Its value lies generally between -1 and 1, but in our case it is limited to [0..1], since all dimensions have positive magnitudes. 0 corresponds to uncorrelated vectors, while 1 indicates that the vectors are the same.

3. **Build prediction models for monitored clusters**

   For training ML models, the following SFpark attributes will be used:

   - parking price - varies with zones and block faces
   - traffic density - determined by sensors installed on the street surface, it indicates how much from a street section is occupied by cars
   - fuel price - daily rate, valid for the whole city
   - precipitation - the amount of rain or snow that fell per day, valid for the entire city
   - street closing information - due to street parades, construction work, etc.
   - parking demand up - signals the block faces that have a higher parking demand at a certain point, due to concerts or other extraordinary events
   - construction site - signaled for a block face that it affects
   - parking place count - the number of operable parking spots
   - occupancy - the rate of occupied spots divided by the parking place count

   The target variable for the model is *occupancy*.

   The prediction models will be computed using state-of-art methods, like neural networks, autoregression models and support vector regression[2][3].

   In figure 2 there is a snippet of training data.

BLOCK_FACE_ID	TIME_POINT	PRICE $/H	TRAFFIC %	FUEL PRICE $	RAIN inch/day	S_CLOSED	P_DEMAND	C_SITE	PP_COUNT	OCCUPANCY %
10202	01.04.2011 12:00	2	42.7	1.76	1.299	no	no	yes	47	100.00
10302	01.04.2011 12:00	1.5	23.5	1.76	1.299	no	no	no	32	67.44
10403	01.04.2011 12:00	2	12.2	1.76	1.299	no	no	no	21	84.19
10702	01.04.2011 12:00	2	28.5	1.76	1.299	no	no	no	54	66.67
32608	01.04.2011 12:00	1.5	23.1	1.76	1.299	no	yes	no	56	21.79
32800	01.04.2011 12:00	1.5	62.1	1.76	1.299	yes	no	no	14	55.78
10202	08.04.2011 12:00	2	71	1.84	0.197	no	no	yes	47	95.00
10302	08.04.2011 12:00	1.5	64.5	1.84	0.197	no	yes	no	32	83.33
10402	08.04.2011 12:00	3	23	1.84	0.197	yes	no	no	7	71.43
10403	08.04.2011 12:00	2	75	1.84	0.197	yes	no	no	21	40.24
10702	08.04.2011 12:00	2	13	1.84	0.197	no	yes	no	54	66.67
32608	08.04.2011 12:00	1.5	8	1.84	0.197	no	no	yes	56	2.61
32800	08.04.2011 12:00	1.5	19	1.84	0.197	no	no	yes	14	18.40

Figure 2: Training data snippet indicating SFpark attributes

4. **Apply prediction models to unmonitored clusters, factoring in the respective similarity value**

   For every unmonitored cluster C_u:

   (a) the monitored cluster C_m will be selected that has the maximum similarity between the two clusters

   $$S = max\{similarity(C_u, C_m)\}, \forall C_m \in \mathcal{C}$$

   (b) C_m's model will be applied on the SFpark attributes of C_u, resulting in the prediction

   $$P = M_{C_m}(attr)$$

   (c) the final result will be a value interval

   $$\text{occupancy}(C_u, \text{attr}) = [max\{0, P-(1-S)\cdot 100\}, min\{100, P+(1-S)\cdot 100\}]$$

   (d) in case the resulting interval is too large, the previous steps can be repeated for the next most similar cluster C'_m. By calculating the respective interval and intersecting all of them, one may arrive at a more precise result.

2.2 SFpark Data

The data from the SFpark project in San Francisco offers the richest such datasets to date[4]. Parking occupancy data has been collected there for over two years between April 2011 - July 2013 in more than 400 blocks, resulting in over 1 million data records.

The project was conducted with the main purpose of leveling off the parking occupancy in the city. In seven pilot areas of the city parking prices were adapted to the level of occupancy. Therefore drivers received incentives to park in areas less occupied by paying a smaller parking fee. Another two control areas were used to verify the effectiveness of the pricing measures. Parking data for all nine areas exist (see figure 3).

2.3 City background information

Metadata that indirectly indicates parking demand is highly interesting to analyse. Do areas with a high number of restaurants display a similar parking occupancy

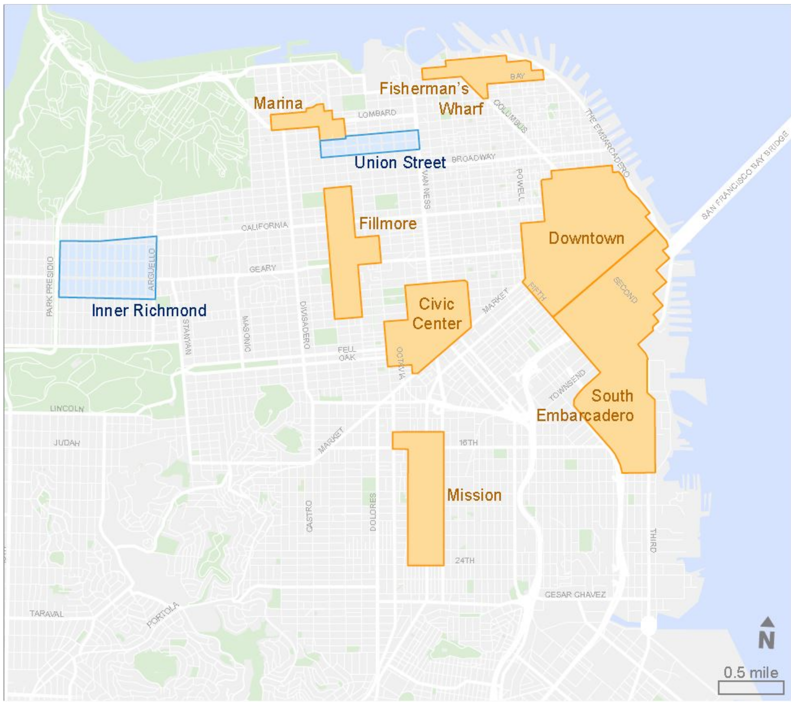

Figure 3: SFpark pilot and control areas

graph? Can parking behavior for quarters with a high sales revenue be correctly guessed?

Some of this metadata can be found in geographic information systems. OpenStreetMap has a relatively large number of Points of Interest (POIs), which for the northern part of San Francisco are displayed as nodes in figure 4. POIs store types of buildings and sometimes other related information that can be used to create a "profile" of an area.

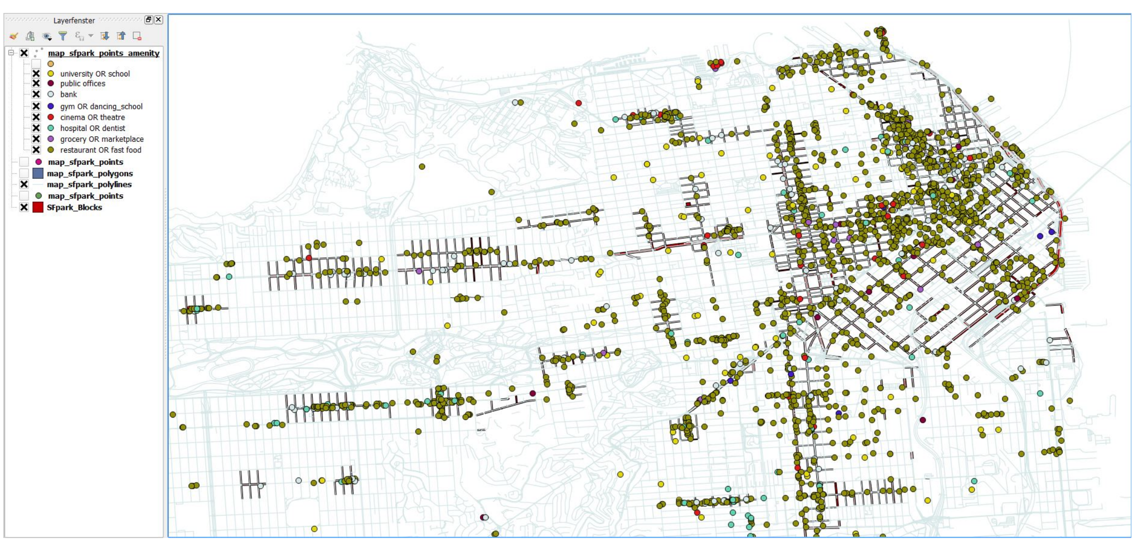

Figure 4: Screenshot from QGIS showing the dimensions of metadata available in OSM

Additionally, SFpark provides sales revenue information for its original nine areas, concerning food product and general retail. Even some of the information that the ML model is trained with but has little added value there can be used to consolidate an area profile.

2.4 Evaluation

To test the accuracy of the inferred occupancy values, we estimate to leave out about 10% of the available SFpark data for testing (training will be performed with 90% of the data).

2.5 Originality of work

Up to now, the smart parking literature concentrates on capturing data and predicting parking information using static or mobile sensors, by covering the areas for which future rates are calculated[5]. An extrapolation approach as described here, to the best of my knowledge, has not been explored yet.

References

[1] *Japanese addresses: No street names. Block numbers.* `http://sivers.org/jadr`. Accessed: 2017-02-19.

[2] Xiao Chen. *Parking occupancy prediction and pattern analysis.* Tech. rep. Technical report, Stanford University, 2014. Machine Learning Final Projects.

[3] Tooraj Rajabioun and Petros A Ioannou. "On-street and off-street parking availability prediction using multivariate spatiotemporal models". In: *IEEE Transactions on Intelligent Transportation Systems* 16.5 (2015), pp. 2913–2924.

[4] *SFpark project.* `http://sfpark.org`. Accessed: 2017-03-05.

[5] Trista Shuenying Lin. "Smart parking: Network, infrastructure and urban service". PhD thesis. Lyon, INSA, 2015.