CONDUENT

Lean Smart Parking

How to Collect High-Quality Data Cost-Effectively

Lean Smart Parking

How to Collect High-Quality Data Cost-Effectively

On-street sensors are now installed in cities from Los Angeles to Moscow. The benefits of data from such sensors are undeniable. As guidance app's move from cell-phones to in-car systems, drivers will rely on sensor data to quickly find nearby spaces. Comparisons of sensor data with meter data can guide enforcement officers to parking violations. Finally, such data enables reliable decisions about prices and time limits, as well as retrospective evaluation of policy effectiveness (see Figure 1) .

Many other cities would love to use such sensors but rightly ask, "Do the costs of sensors outweigh the benefits of the data they collect?" Clearly, there is a trade-off here (see Figure 2) and the best choice usually does not involve installing a sensor in every stall.

In this article we focus on understanding this trade-off based on data from cities where we have 100% sensor coverage. We begin by describing two new methods which can reduce data-collection costs by more than 50% while still ensuring high-quality data:

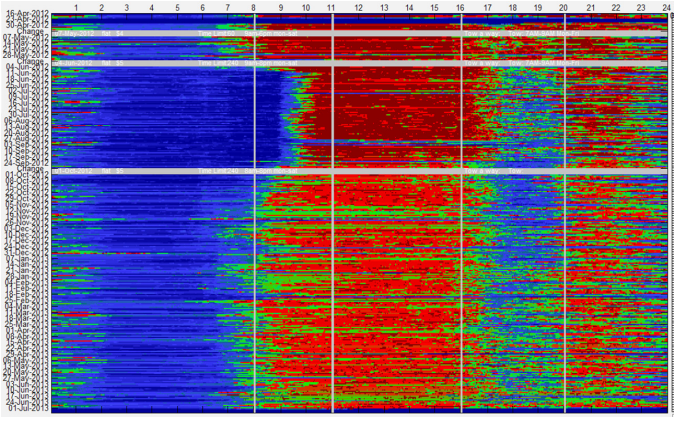

Figure 1: Occupancy time series for 701 S. Olive St., Los Angeles. Colors: blue (<70% occupied), green (70-90% occupied), red (> 90% occupied). Rows correspond to days and columns to minutes. The white box highlights a period where parking was initially free after 6 p.m., but the price became $5/hour when operating hours were extended to 8 p.m. The decrease in occupancy is very clear.

- Spatial sampling where one installs sensors in a fraction of the available spaces.
- Temporal sampling which uses mobile cameras coupled with computer vision algorithms (Bulan et al, 2013) to look at different streets at different times.

We refer to the use of these methods in guidance and policy decision making as lean smart parking.

Both of these methods can exploit payment data from meters to "fill in the gaps", although, since payment rates vary highly within a city, for instance due to varying levels of placard use and abuse, the utility of payment data as a substitute for sensor data is very variable. As shown in Figure 3, occupancy can be mostly paid (first column), mostly unpaid (second column), varying greatly with the time of day (third column) or varying greatly with the position in the block face (fourth column). We give a technical discussion of "Fusing payment and sensor data" in an inset later in this article.

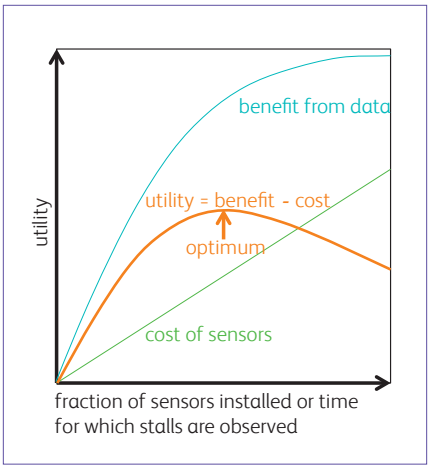

Figure 2: The trade-off between the benefits and costs of data collection.

Furthermore, sampling methods can combine sampling in both space and time. For instance we can move in-street sensors from one location to another, and we might use ultrasonic sensors rather than cameras. The considerations of this article also apply to such alternatives.

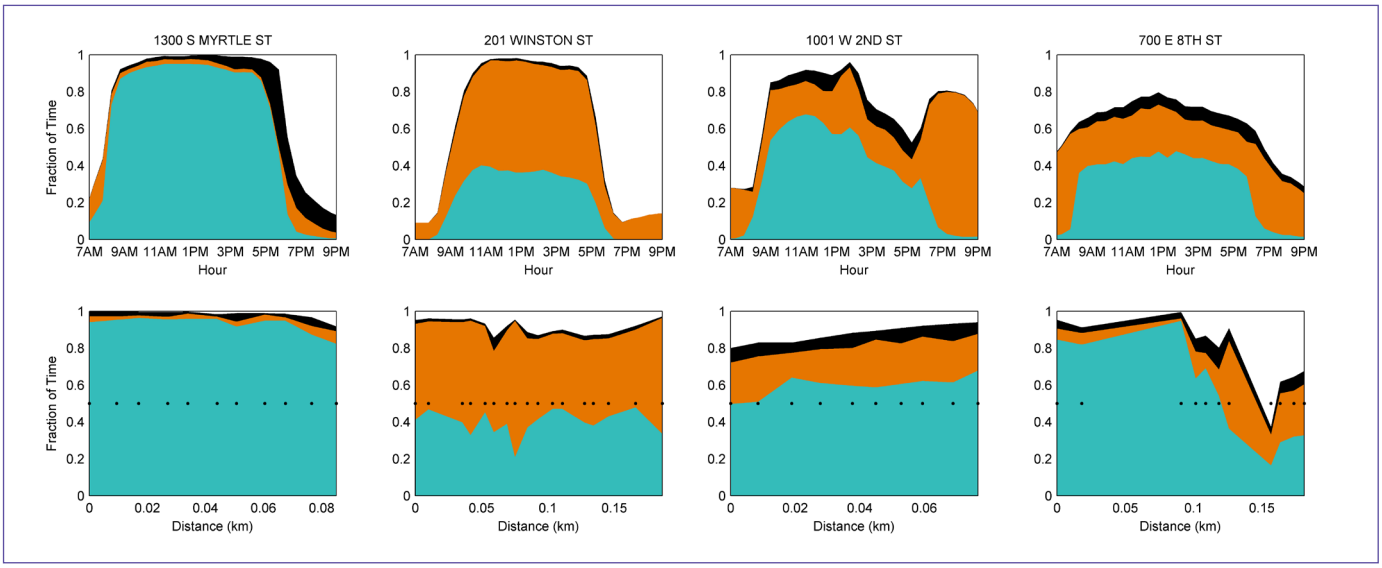

Figure 3: Typical example of the variation of occupancy and payment. Turquoise = occupied and paid, Orange = occupied and not paid, black = vacant and paid. The top row shows the variation with the time of day and the second row shows the variation with position in the block face, with each black dot representing the position

Where should we put the sensors?

There are 12870 different ways to put 8 sensors in 16 stalls. So if we want to do spatial sampling, which arrangement should we choose and shouldn't we have more sensors on longer block faces?

Fortunately, "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970) and the extent to which things are related in parking can be quantified in terms of the correlation between occupancy as a function of distance (see Figure 4). As one would expect, this Figure demonstrates that the correlation is higher for two points on the same block face than for two points on different block faces, even if the distance between the points is the same.

This spatial correlation can be used to calculate the error in predictions of occupancy for any given sensor arrangement using a method from oil and gas exploration called kriging. We can then search over all possible sensor arrangements and pick the arrangement which minimizes this error.

Where should the cameras look next?

Temporal sampling methods work by maintaining a belief about each block face's occupancy based on past observations. The error in this belief is reduced when a block face is observed. Otherwise, if the block face is not observed, the error increases because people's behaviour changes unpredictably.

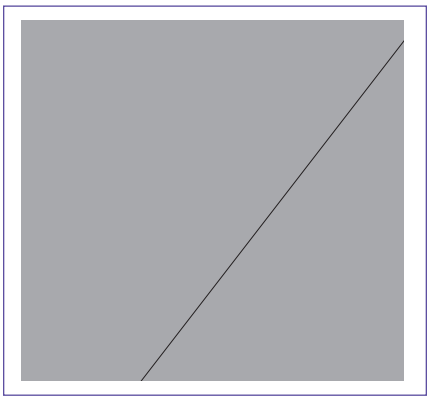

Figure 4: The spatial correlation between occupancy for pairs of stalls at different distances and with different relationships.

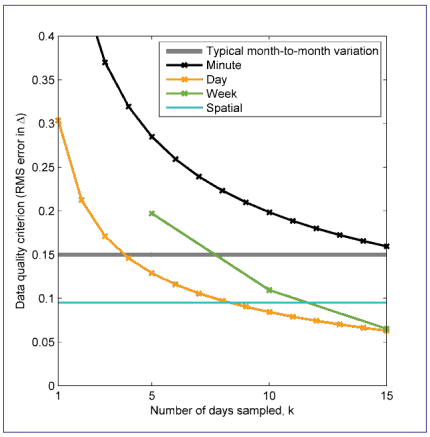

Figure 5: Comparison of the quality of sampling methods. We estimate a criterion Δ defined as the fraction of time that a block face is over 90% occupied minus the fraction of time that it is under 70% occupied; thus Δ lies between -1 and 1. The sampling methods are then compared in terms of the root-mean-squared (RMS) error in estimates of Δ.

This situation is analogous to the tracking of multiple military targets (block faces) with a limited number of radars (mobile units) so as to minimize the error in our belief about all the targets. It turns out that a near-optimal solution to this problem is given by the Whittle index policy (Whittle 1988).

This policy even allows us to use payment data from block faces that are not being observed. So, if payments indicate that the occupancy of a block face has changed

considerably since the last observation, our method will tend to choose to observe that face next. Similarly, we can ensure that places where payments are closely related to occupancy receive fewer observations.

What is "high-quality" data anyway?

Even decisions based on data from a sensor in every stall will sometimes be wrong, resulting in a loss of utility to drivers and policy makers. This is because such decisions are based on predictions of future occupancy in the presence of unpredictable changes from minute-to-minute and from month-to-month. For instance, one might reduce prices in response to low occupancy only to find that the next month is much busier due to an unpredicted demand spike.

Therefore, we describe data as "high-quality" if the loss of utility due to spatial or temporal sampling is less than the loss of utility due to inaccurate predictions given full data.

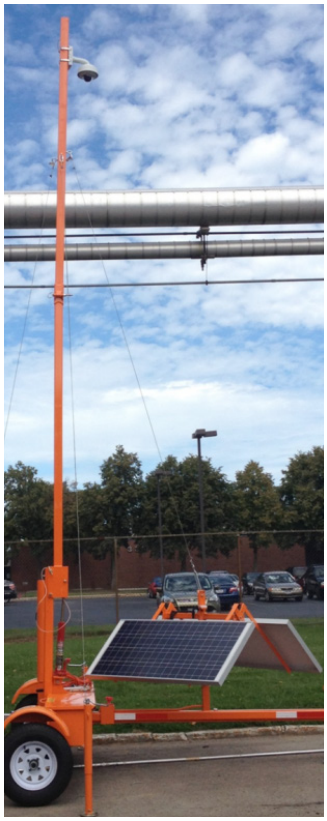

Photo of Mobile Camera System

Measuring the Quality of Sampling Methods

Data from cities where we have 100% sensor coverage is perfect for evaluating the effectiveness of sampling methods. After eliminating holidays and other atypical periods, we simply treat a fraction of the data as if it had been observed and use it to estimate the occupancy for the remaining fraction. Finally, we compare these estimates with the full data.

Results

Figure 5 compares the data quality from the following sampling methods:

1. Minute: observe 1 minute on k distinct days;

2. Day: observe all minutes of k distinct days;

3. Week: observe 5k days from k whole weeks;

4. Spatial: permanently observe 50% of the stalls.

It also shows the size of typical month-to-month variations as a guideline for "high-quality" data.

The figure shows that 50% sensor coverage (blue line) gives lower errors than the typical month-to-month variations. Thus the error due to partial sensor coverage will be less than the error faced by a system using 100% sensor coverage due to month-to-month variations.

Regarding minute/day/week observations, there is typically a big variation in demand over 11AM-4PM on any given day, thus many minute-observations are required to do as well as 50% sensor coverage. Also, there is correlation within a week, so if Monday is busy, the other days of the same week also tend to be busy. Thus, 5k single-day-observations from different weeks provide better estimates thank week observations.

The results presented here made no use of payment data. Thus they paint an appropriate picture of the performance of sampling methods for places where payments are not strongly related to occupancy, or a conservative picture for other places.

Fusing Payment and Sensor Data

The following table gives a simple-but-powerful model for a single block face at a single moment of time. By applying Bayes rule, it gives a predictor $\mathbb{P}(Z_o | Z_{so}, Z_{\bar{s}p})$ of the total Z_o occupancy given observations of the occupancy of stalls with-sensors Z_{so} and the number of stalls-without-sensors that are paid for $Z_{\bar{s}p}$.

Model	Motivation
$\boldsymbol{Z_o \sim}$ **beta-binomial**$(\boldsymbol{\theta_1, \theta_2, \theta_3})$ **truncated after** $\boldsymbol{Z}$ where $\mathbf{Z_o}$ = total occupancy $\mathbf{Z}$ = total number of stalls $\boldsymbol{A} \sim \boldsymbol{B}$ means "$\boldsymbol{A}$ is distributed as $\boldsymbol{B}$" $\boldsymbol{\theta_1, \ldots, \theta_5}$ are model parameters	We treat parking as a queuing system (the Engset model) in which we allow the variance of the occupancy to take different values for different block faces and times. The beta-binomial has 3 parameters (rather than 2) since we treat the maximum number of drivers who might wish to park as a free parameter.
$\boldsymbol{Z_{\bar{s}o} \mid Z_o \sim}$ **hypergeometric distribution** where $\mathbf{Z_{\bar{s}o}}$ = occupancy of stalls-without-sensors $\boldsymbol{A \mid B}$ means the random variable "$\boldsymbol{A}$ given $\boldsymbol{B}$"	This follows from an assumption that each stall is equally likely to be occupied given the total occupancy.
$\boldsymbol{Z_{\bar{s}p} \mid Z_{\bar{s}o} \sim}$ **binomial**$(\boldsymbol{\theta_4, Z_{\bar{s}o}})$ + **binomial**$(\boldsymbol{\theta_5, Z_{\bar{s}} - Z_{\bar{s}o}})$ where $\mathbf{Z_{\bar{s}p}}$ = number of paid stalls-without-sensors $\mathbf{Z_{\bar{s}}}$ = total number of stalls-without-sensors	This follows from an assumption that the payment state of each stall is independently and identically distributed given that it is occupied / vacant.

This model can be improved by removing the assumption of uniformity over the block face and by exploiting observations of arrival times (Dance et al, 2013). However this is non-trivial since the distribution of arrivals and durations varies a lot with time (see Figure 6).

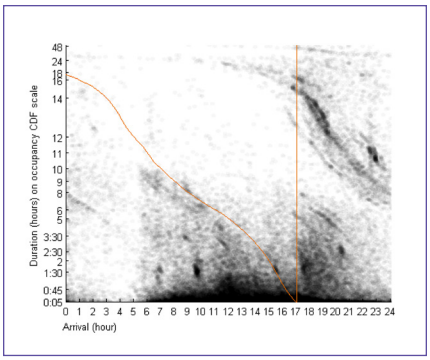

Figure 6: Parking events for a block face in Los Angeles. Each dot corresponds to a parking event and the region between the red lines corresponds to the set of vehicles present at 5 p.m. The duration axis has been nonlinearly scaled so that equal increments on this axis correspond to equal contributions to the total occupancy.

Conclusion

The best choice of lean smart parking technology depends on the city. For instance, for cities with long block faces, cameras can be more economical than in-street sensors because they observe multiple stalls at a time. Also, the extent to which payment data is a good substitute for sensor data varies considerably between and within cities. Therefore we advocate an iterative "lean start-up" approach to deploying sampling methods.

Although lean smart parking requires some work to determine the best solution for each city, our experiments clearly demonstrate that spatial and temporal sampling promise considerable savings, often over 50%, while ensuring only a small loss in the utility of the data relative to a 100% sensor installation. Several cities are currently deploying these first-of-a-kind technologies, so it is likely that lean smart parking is coming to a city near you soon.

Written by Chris Dance, Research Fellow at Conduent Research Centre Europe

Contact

transportation@conduent.com
1-877-414-2676
conduent.com/transportation

References

O. Bulan, R. Loce, W. Wu, Y. Wang, E. Bernal, and Z. Fan, Video-Based Real-Time On-Street Parking Occupancy Detection System.

Journal of Electronic Imaging, 22(4):041109, 2012.

C. Dance, S. Clinchant, and O. Zoeter, Approximate Inference for a Non-Homogeneous Poisson Model of On-Street Parking.

NIPS Workshop on Sustainability, Lake Tahoe, 2013.

W. Tobler, A computer Movie Simulating Urban Growth In The Detroit Region. Economic Geography, 46(2): 234-240, 1970.

P. Whittle. Restless Bandits: Activity Allocation In A Changing World. Journal of Applied Probability, 287-298, 1988.

CONDUENT