Lean Sma

How to cost-effectively collect high-quality data.

By Christopher Dance

On-street parking sensors are working in cities from Los Angeles to Moscow. The benefits of data from such sensors are undeniable. As guidance apps move from cellphones 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 additional 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.

rt Parking

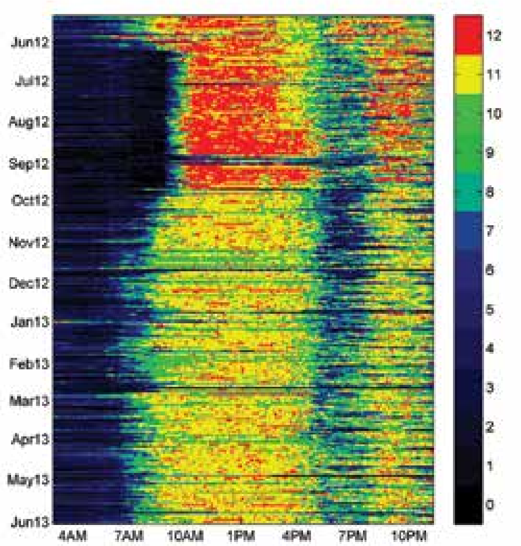

FIGURE 1: Occupancy time series for 701 S. Olive St., Los Angeles. Colors: blue (<70 percent occupied), green (70-90 percent occupied), red (100 percent occupied). Rows correspond to weekdays and columns to minutes. During the period shown, parking was initially free after 6 p.m., but the price became $5 per hour when operating hours were extended to 8 p.m. There was also a price increase from $4 per hour to $5 per hour during the day. The corresponding decrease in occupancy is very clear.

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

- 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 because payment rates vary highly within a city due to varying levels of placard use and abuse, for example, the utility of payment data as a substitute for sensor data is very variable. Furthermore, we might sample in both space and time, for instance by moving 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.

Finally, we discuss the key questions of what high-quality data is and the accuracy of sampling methods to illustrate the tremendous potential of lean smart parking.

Where Should Sensors Go?

There are 12,870 different ways to put eight sensors in 16 stalls. So if we want to do spatial sampling, which arrangement should we choose? Shouldn't we have more sensors on longer block faces?

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 of more than 50 percent.

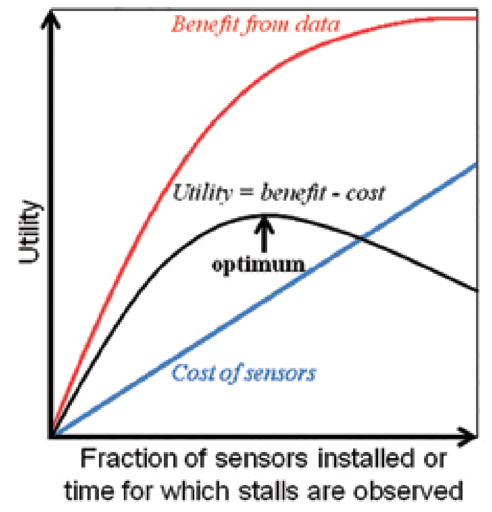

FIGURE 2: The trade-off between the benefits and costs of data collection.

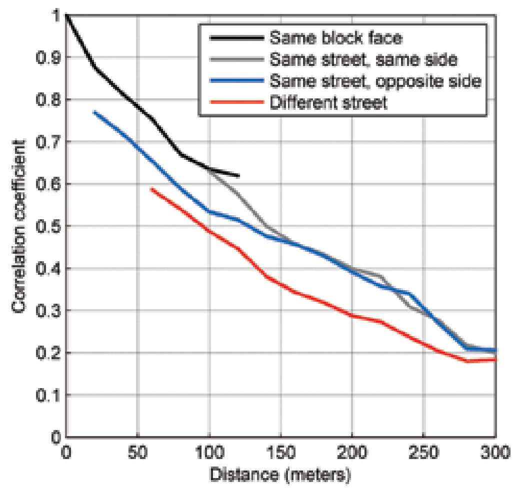

FIGURE 3: The spatial correlation between occupancy for pairs of stalls at different distances and with different relationships.

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 3).

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 that minimizes this error.

Where Should 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. If the block face is not observed, the error increases because people's behavior changes unpredictably.

This situation is analogous to the tracking of multiple military targets (block faces) with a limited number of radars (mobile units) 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?

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 month to month.

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

Measuring the Quality of Sampling Methods

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

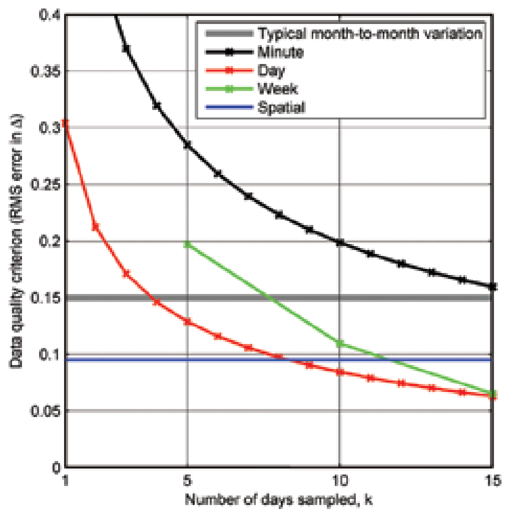

FIGURE 4: Comparison of the quality of sampling methods. We estimate a criterion defined as the fraction of time that a block face is more than 90 percent occupied minus the fraction of time that it is less than 70 percent occupied; thus lies between -1 and 1. The sampling methods are then compared in terms of the root-mean-squared (RMS) error.

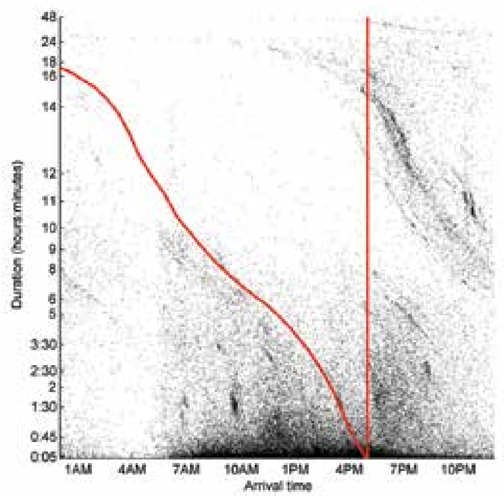

FIGURE 5: Parking events for a block face in Los Angeles. Each dot corresponds to a parking event and the region between the red lines corresponds to the set of vehicles present at 5 p.m. The duration axis has been nonlinearly scaled so that equal increments on this axis correspond to equal contributions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

Results

Figure 4 compares the data quality from the following sampling methods:

- Minute: observe one minute on distinct days.
- Day: observe all minutes of distinct days.
- Week: observe days from whole weeks.
- Spatial: permanently observe 50 percent 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 percent 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 percent sensor coverage due to month-to-month variations.

Regarding minute/day/week observations, there is big variation in demand between 11 a.m. and 4 p.m. on any given day, thus many minute observations are required to do as well as 50 percent 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. Therefore, five single-day-observations from different weeks provide better estimates than weeklong observations.

The results presented here made no use of payment data. 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.

Conclusion

The best choice of lean smart parking technology depends on the city. For instance, in 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.

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 of more than 50 percent, while ensuring only a small loss in the utility of the data relative to a 100 percent 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!

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.

CHRISTOPHER DANCE is research fellow at Xerox Research Centre Europe. He can be reached at dance@xerox.com or +33.47661.5050.