

Mapping similarities in temporal parking occupancy behavior based on city-wide parking meter data

Fabian Bock^a, Karen Xia^b, and Monika Sester^a

^a *Institute of Cartography and Geoinformatics, Leibniz Universität Hannover, Germany; bock@ikg.uni-hannover.de, sester@ikg.uni-hannover.de*

^b *Columbia University, New York, US, k.xia@columbia.edu*

Abstract: The search for a parking space is a severe and stressful problem for drivers in many cities. The provision of maps with parking space occupancy information assists drivers in avoiding the most crowded roads at certain times. Since parking occupancy reveals a repetitive pattern per day and per week, typical parking occupancy patterns can be extracted from historical data.

In this paper, we analyze city-wide parking meter data from Hannover, Germany, for a full year. We describe an approach of clustering these parking meters to reduce the complexity of this parking occupancy information and to reveal areas with similar parking behavior. The parking occupancy at every parking meter is derived from a timestamp of ticket payment and the validity period of the parking tickets. The similarity of the parking meters is computed as the mean-squared deviation of the average daily patterns in parking occupancy at the parking meters. Based on this similarity measure, a hierarchical clustering is applied. The number of clusters is determined with the Davies-Bouldin Index and the Silhouette Index.

Results show that, after extensive data cleansing, the clustering leads to three clusters representing typical parking occupancy day patterns. Those clusters differ mainly in the hour of the maximum occupancy. In addition, the locations of parking meter clusters, computed only based on temporal similarity, also show clear spatial distinctions from other clusters.

Keywords: Data mining, Spatial analysis, Clustering, Parking search

1. Introduction

The shortage of parking spaces in many cities is a relevant problem in today's world. Drivers often need to circle around city blocks wasting time and gas until they find a free parking space (Shoup 2006). The provision of maps with typical parking space occupancy patterns has the potential to reduce the severity of this problem, as these maps would help drivers avoid the most crowded roads at certain times. Also, this information is important to policymakers since they can enact parking price changes based on these patterns. Studies show that dynamic parking price adaptations based on parking demand can have a positive impact on the parking situation (Millard-Ball et al. 2014).

Parking occupancy on a particular road typically demonstrates recurrent temporal patterns of the day and of the week, which also reveals similarities to several other roads within the dataset. Therefore, by clustering, temporal similarities among roads can be identified and only the typical parking patterns need to be stored or transmitted to navigation systems (Richter et al. 2014). Similarly, Vogel et al. (2011) found typical temporal patterns at specific bike-share stations through a clustering approach of bike sharing rental and return data. The prediction of parking availability can be performed with both model-based (Caliskan et al. 2007, Jossé et al.

2013) and data-driven (Rajabi-oun et al. 2015, Bock et al. 2016) methods.

In this paper, we study the clustering of daily patterns in parking occupancy derived from parking meter data. Therefore, we first apply some data cleansing to filter irregular parking meter usage (e.g., parking meter deactivated for some period due to construction). Then, we compute an estimate of the parking occupancy of every parking lane equipped with a parking meter based on the duration of validity of the parking tickets sold. This occupancy information is averaged over all weekdays for specific times of day. These day patterns are grouped with hierarchical clustering. The Davies-Bouldin Index and the Silhouette Index determine the optimal number of clusters. Finally, clustering results are evaluated and visualized in a map.

2. Dataset

In our study, we analyzed parking meter data from 176 meters with the same pricing regulation in Hannover, Germany over a one-year period from May 2012 through April 2013. The parking meters are located in the center of the city, which largely consists of shopping malls, shops, offices, and cultural sites (a map of the parking meter locations is shown in Figure 1). Every parking meter sells parking tickets for all parking spaces in the adjacent parking lane. For the chosen pricing regulation, the operating hours are from 9 AM until 7 PM every day

from Monday through Saturday with a fee of 75 euro cents per 30 minutes. Our dataset contains a record for every parking ticket, consisting of timestamp of payment, parking meter ID, duration of validity, and ticket price. In total, more than one million parking tickets were analyzed for this study.

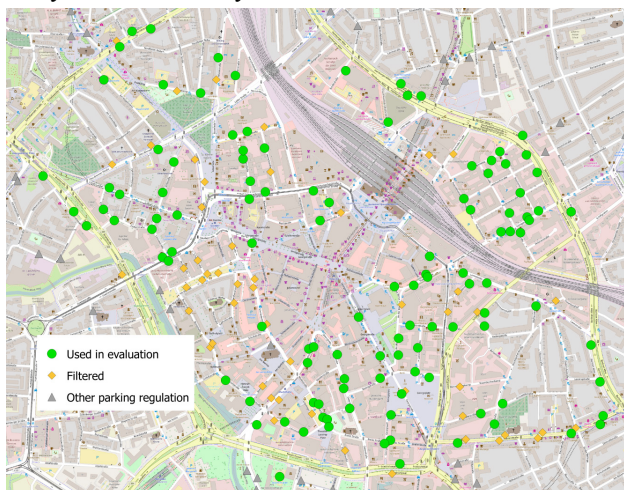


Fig. 1. Locations of parking meters in the city center of Hannover, Germany. Background map is taken from OpenStreetMap.

3. Methodology

To derive clusters of typical parking patterns, intensive data cleansing is applied by filtering parking meters with irregular usage patterns first (see Section 3.1). The occupancy of every parking lane is then calculated by accumulating all parking tickets valid at a specific time instance (see Section 3.2). Hierarchical clustering is computed based on a least-squares similarity measure of the average weekly patterns, limited to workdays, using the complete link-age method. The optimal cluster number is obtained by utilizing both the Davies-Bouldin and Silhouette Indices as cluster validation techniques to represent the intra- and inter-cluster relationships among parking meters within the dataset (see Section 3.3).

3.1 Data cleansing

Irregular parking ticket sales, which could be caused by temporal construction sites, parking meter malfunction, or changes in parking regulations, are cleansed from the dataset.

The following criteria are applied for filtering:

- Maximal period without a ticket sale:
Machines with at least one period of time in which no tickets were bought for at least 10 full days since the last ticket were excluded from further analysis. It is assumed that the most common reasons are nearby construction sites which lead to parking prohibition and removed parking meters.
- Minimum number of tickets sold:
Some parking meters show a very low total number of ticket sales in the aforementioned one-year period. While about 13,500 tickets

were sold on average per meter, we filtered out parking meters with less than 1,000 tickets.

- Drastic changes in ticket sales:
Some machines exhibit a drastic increase or decrease in tickets sold per day. ‘Drastic’ is defined as a change that is not repeated, does not form a part of a larger oscillating ticket pattern in the data. It also means that the change is sustained: that is, the average number of tickets changes to a new value that is consistent with the change (for at least a quarter of a year), and not the average that was previously associated with the data.

In addition, parking meters that allow free parking for residents are omitted since residents’ parking activities are not included in the dataset. In total, the number of parking meters is reduced from 176 to 117. The locations of filtered and remaining parking meters are illustrated in Figure 1.

3.2 Computation of average parking occupancy from parking tickets

The parking occupancy in the vicinity of every parking meter is computed based on the timestamp of ticket purchase and the duration of validity of every ticket. Assuming that drivers park at the parking space for as long as the ticket is valid, the count of valid tickets at every time instance gives an estimate of the parking occupancy. An example for this method is visualized in Figure 2.

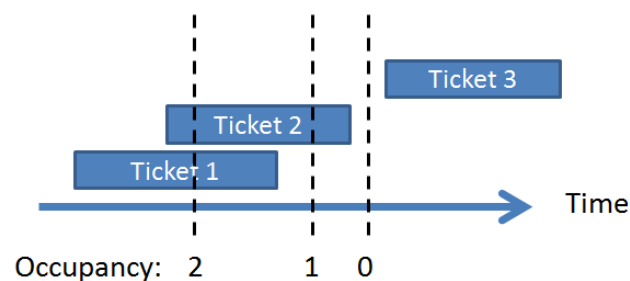


Fig. 2. Example for the calculation of occupancy from parking ticket data: the occupancy is derived from the number of tickets valid at this time instance.

The occupancy value is calculated for the entire one-year period in five-minute intervals. Since parking occupancy is a highly periodic phenomenon, we average over all weekdays (excluding bank holidays) according to the time of day. The result represents the typical daily parking occupancy pattern. To compare these patterns between parking lanes with different capacity, the average occupancy patterns of each parking meter are divided by their maximum average occupancy for normalization.

3.3 Hierarchical clustering

The goal of the clustering step is to group parking meters with similar parking patterns over a one-day period. The number of clusters is not known a priori. Thus, the

recursive splitting or merging of clusters in hierarchical clustering is well suited to find clustering results for all possible numbers of clusters. For the splitting case in particular, hierarchical clustering starts with one cluster containing all elements. Then, this cluster is split into two clusters based on a similarity measure and a linkage method. The similarity measure represents the pair-wise similarity between the cluster. The linkage method determines the best split containing similar elements in the same cluster or separating very dissimilar elements. In our case, the Euclidian distance and the ‘complete’ linkage method is used. ‘Complete’ means that the maximal distance of the members of the two clusters is used as the final distance. The splitting is repeated until all clusters contain only one element.

The optimal number of clusters can be determined by metrics describing the clustering quality. In our study, we use the Davies-Bouldin Index and the Silhouette Index. The Davies-Bouldin Index, which is essentially a ratio of intra- and inter-cluster distances, produces a relatively lower value to indicate better separated clusters and more tightly associated members within clusters. Values are strictly non-negative. On the other hand, for the Silhouette Index, whose value ranges from -1 to 1, the optimal value for a clustering result is 1. This index measures how similar an individual observation is to its own cluster compared to all other clusters.

4. Evaluation

As a result of data cleansing, the number of investigated parking meters is reduced from 176 to 117. Most of the removed meters revealed a long period without ticket sales (46 out of 59 meters). The results of the average daily occupancy pattern clustering and the evaluation of the number of clusters are shown in Section 4.1. These results of the clustering are visualized on a map and interpretations are given (see Section 4.2).

4.1 Clustering result

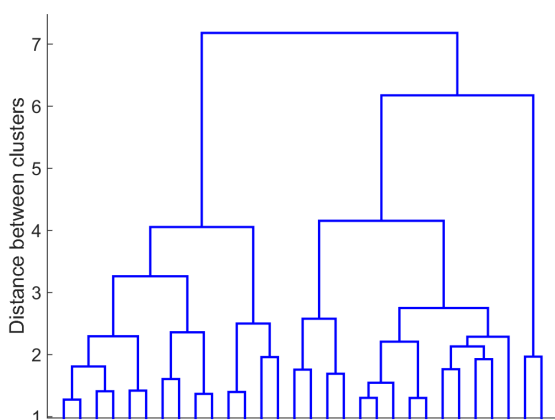


Fig. 3. Dendrogram of hierarchical clustering up to 30 clusters.

The result of the clustering can be visualized in a dendrogram where the hierarchical splitting of clusters and their similarities are shown in a tree diagram. Figure 3 shows the dendrogram for our clustering result up to 30 leaves. In our case, the branching is rather balanced, but

with a small cluster already separated as the third cluster. The optimal number of clusters is determined by the maximum of the Silhouette Index and the minimum of the Davis-Bouldin Index. Both indices indicate that three clusters lead to the best result (see Figure 4).

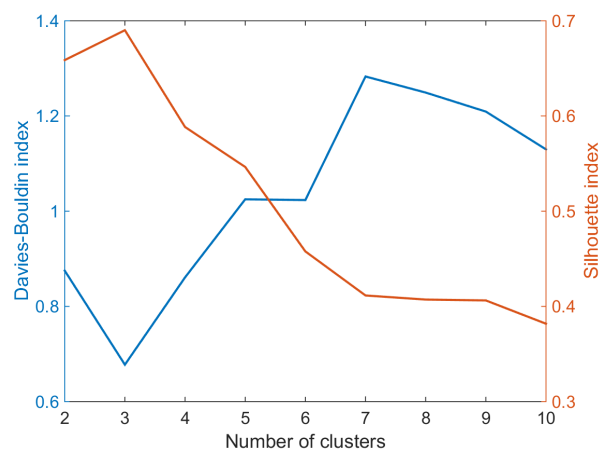


Fig. 4. Davies-Bouldin and Silhouette index for different numbers of clusters.

The average patterns of the cluster members for three clusters are visualized in Figure 5. They show clearly distinct day profiles: cluster 1 (three members) has a relatively small occupancy in the morning and afternoon, but a very high occupancy around 6 PM. Inversely, cluster 2 (37 members) has the highest occupancy in the morning followed by a continuous decrease. In cluster 3 (77 members), the occupancy remains on a similar level during the day and only drops before the end of parking meter operation hours. As an example, the occupancy patterns for all members of cluster 2 are shown in Figure 6.

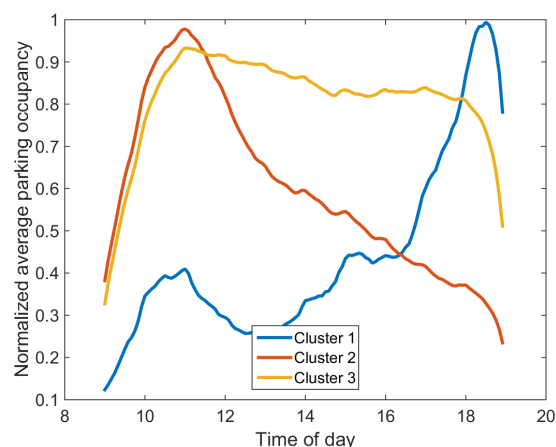


Fig. 5. Occupancy day pattern averaged for all members of the three clusters.

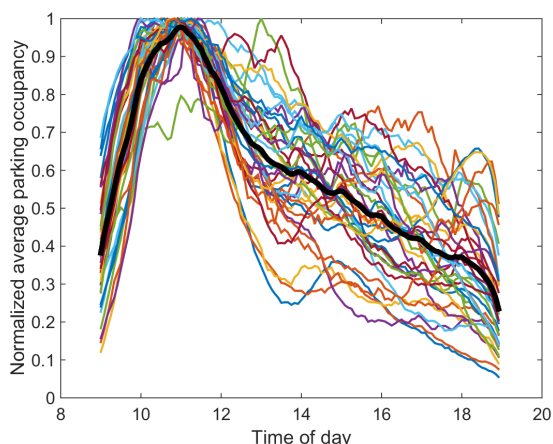


Fig. 6. Plot of day patterns of all members of cluster 2 and their average value in black.

4.2 Mapping of clustering result

The result of clustering the parking meters according to their occupancy patterns is also visualized in a map in Figure 7. The members of the clusters clearly reveal spatial proximity although spatial information was not included as a feature in the clustering process. The distribution of these clusters can be also interpreted by the existence of points of interest. All three members of cluster 1 (evening peak) are located in the same area around a cinema and several bars. Cluster 3 (constant occupancy) parking meters are mainly situated in the city center where there are many shops, cafes, and tourist attractions. Finally, cluster 2 (morning peak) meters surround the city center, mainly representing an area with many offices and public buildings such as courts and city administration. If the number of clusters is increased, the spatial proximity remains. As an example, the clustering result for six clusters is shown in Figure 8. While more clusters occur around the center, the city center is still covered by parking meters belonging to one cluster only. This result leads to the assumption that parking occupancy has a clear relation to points of interest and land use. The information about times and areas of high parking demand can be very helpful for drivers searching for a free parking space. However, since the parking occupancies are similar in their vicinity and thus fall into the same clusters, the benefit of parking guidance just based on the daily occupancy patterns of the clusters is limited.

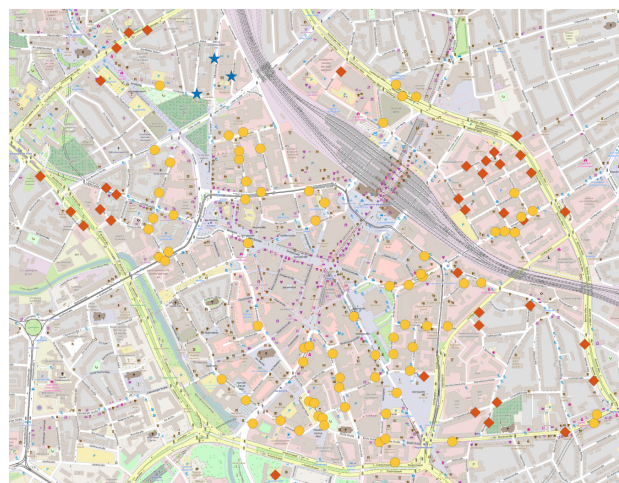


Fig. 7. Map of the parking meter cluster members for three clusters. The colors correspond to the colors in Figure 6: cluster 1 with evening peak (blue star), cluster 2 with morning peak (red diamond), and cluster 3 with constant occupancy (yellow circle). Back-ground map is taken from OpenStreetMap.

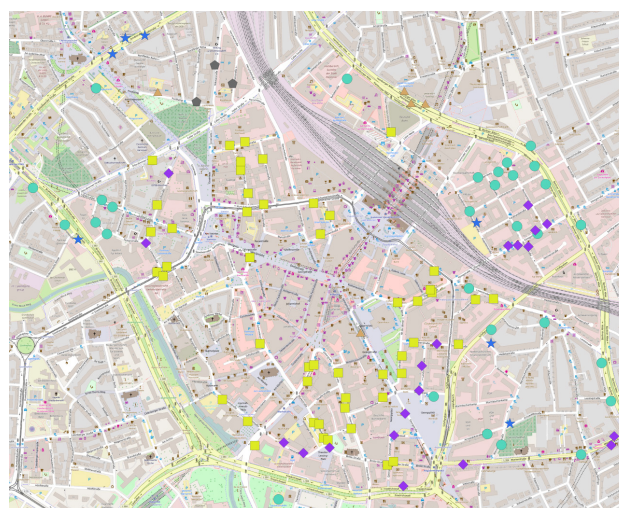


Fig. 8. Map of the parking meter cluster members for six clusters. Background map is taken from OpenStreetMap.

5. Conclusions

In conclusion, clustering of parking meter data leads to distinct temporal parking behavior for different roads, mainly differing in the peak hour of parking occupancy. Most interestingly, the locations of parking meter clusters, computed only based on temporal similarity, also show clear spatial region distinctions from other clusters. These results can be used to provide parking maps indicating parking areas with similar temporal parking usage at low data storage requirements. Future work will investigate the relationship of points of interest and specific spatio-temporal parking patterns.

6. Acknowledgements

This research has been supported by the German Research Foundation (DFG) through the Research Training Group SocialCars (GRK 1931). The focus of the SocialCars Research Training Group is on significantly improving the future of cities' road traffic, through

cooperative approaches. This support is gratefully acknowledged.

Acknowledgements of support for the project / paper / author are welcome.

7. References

- Bock, F., Martino, S. Di, & Sester, M. (2016). What Are the Potentialities of Crowdsourcing for Dynamic Maps of On-street Parking Spaces? *Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, 19–24. doi:10.1145/3003965.3003973
- Caliskan, M., Barthels, A., Scheuermann, B., & Mauve, M. (2007). Predicting Parking Lot Occupancy in Vehicular Ad Hoc Networks. *2007 IEEE 65th Vehicular Technology Conference - VTC2007-Spring*, 277–281. doi:10.1109/VETECS.2007.69
- Jossé, G., Schubert, M., & Kriegel, H.-P. (2013). Probabilistic parking queries using aging functions. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - SIGSPATIAL'13* (pp. 442–445). New York, New York, USA: ACM Press. doi:10.1145/2525314.2525458
- Millard-Ball, A., Weinberger, R. R., & Hampshire, R. C. (2014). Is the curb 80% full or 20% empty? Assessing the impacts of San Francisco's parking pricing experiment. *Transportation Research Part A: Policy and Practice*, 63, 76–92. doi:10.1016/j.tra.2014.02.016
- Rajabioun, T., & Ioannou, P. (2015). On-Street and Off-Street Parking Availability Prediction Using Multivariate Spatiotemporal Models. *IEEE Transactions on Intelligent Transportation Systems*, 16(5), 2913–2924. doi:10.1109/TITS.2015.2428705
- Richter, F., Di Martino, S., & Mattfeld, D. C. (2014). Temporal and Spatial Clustering for a Parking Prediction Service. *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, 278–282. doi:10.1109/ICTAI.2014.49
- Shoup, D. C. (2006). Cruising for parking. *Transport Policy*, 13(6), 479–486. doi:10.1016/j.tranpol.2006.05.005
- Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding bike-sharing systems using Data Mining: Exploring activity patterns. *Procedia - Social and Behavioral Sciences*, 20, 514–523. doi:10.1016/j.sbspro.2011.08.058