Predicting the Availability of Parking Spaces with Publicly Available Data

Christoph Pflügler¹, Thomas Köhn¹, Maximilian Schreieck¹, Manuel Wiesche¹ and Helmut Krcmar¹

Abstract: Searching for parking spaces on the street causes a significant part of the urban traffic and results in extra costs for the drivers in terms of time and fuel consumption. Existing approaches to predict the availability of parking spaces have significant drawbacks as they are either expensive or rely on the users’ information. This article deals with the prediction of the parking situation based on publicly available data that can be accessed cost-efficiently. Suitable categories of data are identified based on a literature review. Subsequently, a prototypical system that employs a neural network is implemented. The relevance of the different categories of data is evaluated based on 2,779 real world records. The results show that weekday, time of the day, location, and temperature have a significant impact on the prediction; whereas events, traffic, vacation time and rainfall are only of secondary importance. This article categorizes existing solutions to support finding parking spaces and shows that publicly available information can provide a good starting point for the prediction of the availability of parking spaces.

Keywords: Smart Parking, Public Data, Smart City, Parking Prediction, Parking, Neuronal Network

1 Introduction

The volume of traffic in urban regions is continuously increasing [MAZ15] [Ba05]. Additionally, to increasing traffic congestion and longer travel times, health risks for the residents rise [ZB13], [Ch15], [Ra14]. A significant portion of the traffic in urban environments is caused by searching for parking spaces [Sh06]. According to Giuffrè et al. [GST12] the search for parking space is responsible for up to 40% of the total traffic within cities. Apart from the increase of the own consumption of fuel, the search for parking spaces will also negatively impact other road users’ consumption. Parking seekers often drive slowly and cause the traffic flow behind them to slow down [In14]. According to Le Fauconnier and Gantelet [LG06] on average it takes 12 minutes to find a free parking space.

In the literature and on the market there are a number of applications and concepts that aim to simplify the search for a free parking space. These applications are usually based either on user data or data from sensors in parking spaces. Therefore, they are either dependent on data entered by the users or they have high setup costs due to the installation of sensors in the parking spaces. A system based on publicly available data that predicts...
the availability of parking by means of an intelligent algorithm would not have the mentioned disadvantages. Up to now, it has not yet been investigated which data is suitable for the prediction of parking spaces within a city.

Based on a literature review, weather, time, traffic, events and locations are identified as possibly relevant categories of data. It is important that the data is publicly available to ensure its independence and sustainability. Subsequently, the data is integrated in a prototypical solution based on a neural network to predict the availability of free parking spaces within a certain area. The data is evaluated in terms of its contribution to the improvement of the predictions. The results show that weekday, time of the day, location, and temperature have a significant impact on the prediction, whereas events, traffic, vacation time and rainfall are only of secondary importance. The training data was gathered in the city of Munich, Germany, where the parking situation is a significant transport problem [CM16].

The Design Science Research (DSR) approach was selected for this article; therefore, the paper is organized according to the framework of Gregor and Hevner [GH13]. The first section provides an overview of existing solutions for the identification of available parking spaces. In the next section, we present a literature review on the categories of data that can influence the parking situation and are publicly accessible. Consecutively, the methodology and the exemplary implementation are described. Subsequently, an exemplary data set is used to evaluate the prototypical implementation. The article concludes with a discussion of the results and a summary of the contribution.

2 Existing solutions to support the search for parking spaces

For the identification of existing solutions in addition to the literature, the market has been systematically studied with a focus on existing concepts and applications. The search was concentrated on systems that support the search for parking spaces on the street. Other kinds of parking such as parking garages or closed parking were not considered. The identified concepts and applications can be divided into three categories: systems based on user data, systems based on parking data and systems based on publicly available data.

2.1 Systems based on user data

A variety of systems work on the basis of data that is provided by users (crowd sourcing). Two types of applications can be provided: Free parking spaces explicitly reported by the users or implicitly recognized by an application installed on the smartphone of the users. The explicitly reported kind is directly activated by the user. The implicit recognition is carried out by the application using the data of the user but without his participation [NEM13]. In both types, users can view the reported free parking spaces and reserve them via a website or a mobile application (app). Applications of this type are widely used and successful in the market. Examples of established applications based on explicit user data...
are: ParkoPedia [PA15c], ParkMünchen [PM15], Parkonaut [PA15b] and ParkTAG [PA15d]. Systems based on the implicit recognition of users’ data are not yet widely used. A system based on implicit user data is Pocket Parker of Nandugudi et al. [NNC14]. This system measures the variations of the speed of the user and identifies the occupied parking space. Similarly, the system Park Sense [NEM13] measures how fast wireless networks in the vicinity of the user change to derive the users’ speed. This data is used to identify when the user has left a parking space. Other example of a system with implicit recognition is the cooperation between the car manufacturer BMW and INRIX which evaluates user data to generate parking information [In15].

Systems based on user data are generally the most widely used and have the advantage of low construction and operation costs [NEM13]. Disadvantages of this type of systems are that a large amount of users is needed to generate sufficient data to provide adequate parking information [CSR12], [Gr15]. This effectively represents a paradox, since such applications are used by users only if sufficient parking information is displayed, but the parking information can be generated only after the users provide the data [DRH11]. Moreover, the quality of the parking information is not ensured. Since the parking spaces are not monitored in real time, it is not guaranteed that these are still available when other users reach them and want to park [Ma10].

2.2 Systems based on parking data

Systems in this category are based on parking spaces equipped with sensors. The sensors detect whether a parking space is currently occupied and reports it to a central database. Various companies offer this kind of solutions [SI15], [GE15], [SP15]. Concrete concepts are also presented by Sujith et al. [Su14], McNeal [Mc13] and Seong-Eun et al. [Se08]. The advantage of this type of systems is that the parking information is very precise and available in real time [NEM13]. The main disadvantage that can hinder the deployment of these technologies is shown by Nandugudi et al. [NNC14]: the equipment necessary for one single parking lot costs around US$ 2,500. There are also additional costs for cabling, interfaces and communication systems. The SFPark project in San Francisco endowed a total of 8,622 parking sensors and costed a total of about US$ 18 million [Mc13].

2.3 Systems based on publicly available data

The two previously presented types of systems bring significant disadvantages. Either they have very high initial and operational costs (systems based on parking data) or there is a strong dependence on explicit or implicit user data (systems based on users’ data). A system that analyzes publicly available data to generate predictions for the present, would be cost-effective and not as dependent on a user base. The focus on publicly available data also ensures a higher independence. A system that is already used in practice and is not described in the literature, is the system ParkNav. ParkNav combines data of different categories and, using an algorithm, predicts probability values for the availability of
parking spaces [Gr15], [PA15a]. However, it has not been studied in literature which data categories are useful for such a system.

3 Relevant categories of publicly available data

For the development of a system based on publicly available data, it is necessary to identify possible influencing factors to find a parking space. To this end, a literature review was conducted with the aim to identify categories of data that influence the parking situation and are publicly available. The following categories were identified and are summarized in Table 1.

According to Yang et al. [YLW03] weather information is of central importance, since weather significantly affects the current traffic behavior and the traffic flow’s intensity. Greengard [Gr15] agrees with this argument and specifically considers rainfall and temperature as important influencing factors for parking prediction.

According to Zehe et al. [Ze07], the time of the day is of great importance. Their results show that holidays, weekdays and time of the day lead to different parking situations. Greengard [Gr15] has a similar conclusion, his work argues that the traffic volume is different during holidays and vacation time as more parking spaces will be searched, making it more difficult to find a parking space. In general, the time can be divided into five data sets, which can impact the traffic volume: time of the day, day, month, bank holidays and vacation time.

Another factor that may influence the availability of parking spaces are events [KMP12], [YLW03]. Events, such as concerts or matches, cause a significant increase in the volume of traffic, consequently more parking spaces will be searched and the search for parking spaces will be more difficult.

Shin and Jun [SJ14] and Yang et al. [YLW03] argue that traffic information is an important factor for the availability of parking spaces. A generally higher traffic volume makes it harder to find a parking space, because more parking spaces will be searched.

Another important factor influencing the car parking is identified by Mathur et al. [Ma10]. They describe that parking availability often depends on the location. The chance to find a parking space generally is higher in certain locations than in others. Giuffrè et al. [GST12] use the example of Milan where the parking situation differs strongly from one area of the city to another. To integrate the impact of the location, it is helpful to divide the urban area into equal rectangular cells. Mathur et al. [Ma10] divided the city of San Francisco into cells of 175x190 meters, because this size ensures that most of the time only one street segment lies within a cell.
### Table 1: Overview of the categories of data, the data and possible data sources

<table>
<thead>
<tr>
<th>Database category</th>
<th>Data</th>
<th>Data source example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Time of the day</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weekday</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Month</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bank holidays</td>
<td>Coordinated Universal Time (UCT) Server</td>
</tr>
<tr>
<td></td>
<td>Vacation time</td>
<td></td>
</tr>
<tr>
<td>Weather</td>
<td>Temperature</td>
<td>Deutscher Wetterdienst [DW15]</td>
</tr>
<tr>
<td></td>
<td>Rainfall</td>
<td>openweatherAPI [OW15]</td>
</tr>
<tr>
<td>Traffic</td>
<td>Traffic situation</td>
<td>Bing Maps [MI15]</td>
</tr>
<tr>
<td>Event</td>
<td>Event locations</td>
<td>eventful [Ev15]</td>
</tr>
<tr>
<td>Users’ request for parking</td>
<td>Location of users’ request</td>
<td>GPS coordinates of the app’s user</td>
</tr>
</tbody>
</table>

#### 4 Implementation and evaluation of a prototypical system for the prediction of the parking situation

To develop the prototypical implementation and the integration of the different data sets, we followed the DSR approach [GH13], [He04]. The procedure is based on the DSR process proposed by Peffers et al. [Pe07]. It starts with the identification of existing approaches as well as solutions and their disadvantages (chapter 2). The implementation is based on the insights from this analysis: we chose a system that uses publicly available data to predict the availability of parking spaces and identified relevant data sources (chapter 3). Based on this, the exemplary implementation by means of a neural network is described in this chapter followed by an evaluation of the system. The system and its components are shown in Figure 1, the components will be described in detail below. For the evaluation of the data categories, training data was recorded during the period from July to September 2015 in the downtown area of Munich with a specially created web app. In total, 2,779 records for parking availability in the downtown area were collected. In about 23% of the cases at least one parking space was available within a cell. The web app has been designed for being used while going for a run within the city or walking the dog. Therefore it had a very simple interface. The data collection has been conducted by the authors and by voluntarily participating students.
4.1 Data preparation

The previously discussed categories of data were integrated as input components of the system (Figure 1). These were designed as autonomous and independent components, because categories of data can be easily included or excluded in the prediction. These components gather the information from the appropriate data sources and make sure that the resulting data is processed appropriately for the subsequent neural network. Some datasets change less frequently than others and it is not necessary to re-query them, therefore these components use the temporary memory.

To process the location data, the urban area of Munich (area within the “Mittlere Ring”) was divided into equal rectangular cells according to Mathur et al. [Ma10] as shown in Figure 2. The grid has a size of 74,429m x 111,193m, as these scales relate to 0.001 degrees in latitude and longitude. Using this cell size ensures that usually only one street segment lies in one cell. Latitude and longitude coordinates were used for segmentation, since they facilitate the identification of relevant cells for a given position.
Several reasons support the usage of a neural network as a prediction algorithm. First, neural networks are particularly suitable for predicting events where little or nothing is known about the underlying relationships and features of the events, but enough training data or observation values are available. The information needed can be derived from the training data [ZEH14], [YLW03]. Second, neural networks have the advantage of continuous learning. The training dataset can be easily complemented by other data, so that a continuous improvement of the prediction is possible. Third, neural networks offer the possibility to consider non-linear relationships and interaction effects through the integration of hidden layers [ZEH98], [Ze07].

The number of input neurons of the neural network corresponds to the number of the data from the data sources. It has to be noted that the use of the continuous influences of certain data might not be useful and the input values should be considered as individual neurons. For example, the input value “weekday” should not be used as a continuous value from 1 to 7, since this would mean that the last weekdays (5, 6 or 7) are associated to a higher weight than the first ones (1, 2 and 3). Instead, each weekday should be integrated as a single binary neuron. This procedure was used for all the data for which continuity is not meaningful. This is the case for the following neurons: Weekday (7 neurons); Month (12 neurons); Time (24 neurons); Location (1,323 neurons, one for each cell area); Temperature (12 neurons, by establishing a scale of $\leq -20$ °C to $\geq 40$ °C and dividing the
values in steps of 5 °C). The remaining data sources are represented by one single neuron each, in which the continuous influence can be taken (traffic intensity, rainfall, i.e.) or the input value is only binary (event, holiday, vacation time, i.e.). Overall, we consider 1383 input neurons. As output neuron a single neuron is used, which represents the predicted value. The amount of neurons and intermediate layers depends of the complexity of the decision [ZEH98], [He08]. There are no rules how a neural network exactly should look, but it is recommended to try out several attempts and choose the one with the lowest error rate [ZEH98], [Sc97]. As a guideline, it can be assumed that no more than two intermediate levels are needed and that the amount of neurons should be between the number of input and output neurons [He08].

4.3 Training data set

The feedback function presented in the architectural model (Figure 1), makes it possible to enhance the training set of the neural network continuously with new training values. The feedback function also ensures that the user can enter information about the correctness of the prediction after receiving a prediction on the parking situation. By providing this information, the prediction can steadily increase its accuracy.

For the collection of an initial training dataset a separate web application has been implemented. Users open the app and start a tour. Every 5 seconds, the current position and the current time are recorded. If they see a free parking space, they press a button. If a user passes through a cell without reporting a free parking space, it is declared as occupied, otherwise as free. The web application was designed for users that go for a run or walk their dog within the city.

4.4 Probability database

To ensure the fastest possible response time of the system, the predicted values are not separated from each request, but are calculated in time intervals and stored in the probability database. The division in cells of the urban area mentioned above is applied here, for the downtown of Munich, resulting in 1,323 cells. The cells are displayed in the database and a probability of availability is assigned to each cell, which states how likely it is to find a free parking space in this area. As the prediction algorithm continuously calculates values for the cells, it is ensured that the database always contains up-to-date values. The calculation of values for all 1,323 cells takes about 20 minutes on modern hardware.

4.5 Evaluation of the data categories

Regarding to the construction of the neural network, the best results were found with two intermediate layers and 25 and eight neurons on each of them. The network has been tested with different activation and training functions. The best results were achieved with the
sigmoid activation function and the “Resilient Backpropagation” training method. In the 2,779 records collected on training, the neuronal network reaches an average square error of 0.16321. Regarding to the training iterations the prediction model was configured so that it stops automatically right after 100 iterations.

For the significance analysis of single neurons or single data sources, the change of Mean Square Error (MSE) can be used, which determine the significance of individual or groups of neurons by removing them. The results (Table 2) indicate that the weekday, location, temperature and time of the day improve significantly the prediction whereas events like traffic, holidays and rainfall have only a secondary relevance.

<table>
<thead>
<tr>
<th>Data</th>
<th>Error value</th>
<th>Difference to optimal</th>
<th>Normalized (ordered)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekday removed</td>
<td>0.17686</td>
<td>-0.01365</td>
<td>186.3225</td>
</tr>
<tr>
<td>Location removed</td>
<td>0.17554</td>
<td>-0.01233</td>
<td>152.0289</td>
</tr>
<tr>
<td>Temperature removed</td>
<td>0.17252</td>
<td>-0.00931</td>
<td>86.6761</td>
</tr>
<tr>
<td>Time of the day removed</td>
<td>0.16972</td>
<td>-0.00651</td>
<td>42.3801</td>
</tr>
<tr>
<td>Month removed</td>
<td>0.16777</td>
<td>-0.00456</td>
<td>20.7936</td>
</tr>
<tr>
<td>Vacation time removed</td>
<td>0.16523</td>
<td>-0.00202</td>
<td>4.0804</td>
</tr>
<tr>
<td>Rainfall removed</td>
<td>0.16481</td>
<td>-0.0016</td>
<td>2.56</td>
</tr>
<tr>
<td>Events removed</td>
<td>0.16188</td>
<td>0.00133</td>
<td>1.7689</td>
</tr>
<tr>
<td>Holidays removed</td>
<td>0.16397</td>
<td>-0.00076</td>
<td>0.5776</td>
</tr>
<tr>
<td>Traffic state removed</td>
<td>0.16369</td>
<td>-0.00048</td>
<td>0.2304</td>
</tr>
</tbody>
</table>

Table 2: Results of the significance analysis of neurons

### 4.6 Discussion

First, the systems based on publicly available data will be discussed. These have the advantage that they are less expensive than systems based on parking data. They are not
dependent on a large user base either. They still need real data to train the neural network, however far fewer data is required in comparison with the systems based on user’s data. Furthermore, the training data can also be derived from the past. By means of the presented feedback function, the system ensures that the training dataset is continually expanding. Thereby, new pattern in the parking availability are automatically recognized.

The evaluation of the data shows that the time, the location and the weather have the greatest impact. Interestingly, these data are relatively easy to obtain. The time - with its dimensions of week, time, month, holiday season, holiday - can be generated from the current system time. The location is also easy to obtain. It corresponds to the cell in which a user searches for a parking space. The weather, especially the temperature, affects the parking situation significantly and is also easily accessible. Various data providers such as the German Weather Service [DW15] provide this information free of charge.

Unlike the data discussed above, the prediction accuracy is only slightly improved by data of traffic and events. This information is more difficult to access than the data described above and it could not contribute significantly to the quality of to the prediction of the parking situation. Similarly, additional metadata about the events could not be retrieved from the events database in a structured way. For instance, the event’s audience size could not be obtained.

Overall, the evaluation indicates that the prediction of the parking situation can be conducted with publicly available information. For each data category, it should be evaluated whether the quality increase of the prediction is worth the effort to integrate the source.

The results of the presented evaluation are limited insofar that despite the active diversification of the collected training data, some data are primarily homogeneous. In this regards, time data can be distributed over several time ranges, though the data collection took place only in the months of July, August and September 2015. However, due to this period, weather data is largely consistent, and does not include low temperatures. For further improvement of the accuracy of the prediction, it is necessary to gather additional training data over a longer period of time, at least of the whole year, and recording days with warm and cold temperatures, as well as peak and off-peak hours. Another limitation is that the model presented has been evaluated only in one city, therefore city-based effects cannot be identified and excluded.

Future research should focus on these limitations. The authors plan to evaluate the categories of data with larger data sets extended to several years and enhanced by explicit user data. Furthermore, future evaluation could compare the different systems used to support the search for parking spaces as this has not yet been done in literature.

This work extends existing research by a classification of systems that support the search for parking spaces. It also shows that there are few approaches based on publicly already available data despite their advantages over other systems. Furthermore, the evaluation of
individual data sources shows that simple data is a good starting point for the prediction of the parking situation.

4.7 Conclusion

This paper has been motivated by the effort needed to search for parking spaces on the roadside nowadays in urban environments. First, existing approaches have been researched and categorized based on the data sources they use: data from parking sensors, from users of the system or from publicly available sources. For the systems based on publicly available data, it is not clear which kind of data should be included, therefore the relevant data categories were identified. Subsequently, a prototypical system was implemented by means of a neural network. This served as a basis for the evaluation of the different categories of data. The results suggest that easily available data – i.e. location, time, weather – have the biggest influence for the accuracy of the prediction. This contribution therefore serves as a basis for future systems that predict the parking situation more efficiently.

Acknowledgements

We thank the German Federal Ministry for Economic Affairs and Energy for funding this research as part of the project 01MD15001D (ExCELL).

5 References


