A Deep Learning Approach to Predict Parking Availability in a Real-World Deployment

Anonymous Author*, Anonymous Author*‡, Anonymous Author†‡, Anonymous Author†‡
*Anonymous Place
†Anonymous Place
‡Anonymous Place
Anonymous Emails

Abstract—Smart parking technologies allow both users to be timely informed about the availability of parking lots and corporations to offer services leveraging that information. Because of these useful and likely profitable features, their adoption is increasingly gaining traction. In this paper we describe and analyze a full-stack, industrial-grade, smart parking deployment. Innovative deep-learning forecasting models (based on CNN-LSTM), designed for predicting parking lots availability up to 8 hours ahead are described and tested on two different data sets. Experimental results show that the best performing model is capable of predicting the number of available parking lots over the next 8 hours (in a $\pm 3\%$ range) with 80% accuracy. Given such forecasts, we also articulate ideas about future services and applications that can be built on top of them.

1 Introduction

According to [1], 50+ millions of new vehicles are sold every year. The increasing number of circulating vehicles is likely to increase the average number of vehicles roaming around in search for parking lots which in turn negatively affect traffic flows and consequent pollution levels. Drivers, in fact, rely on on-road perceptions, empirical tries, and past-experience instead of being informed on the number and location of available lots. In 2015, the search of free parking lots was already responsible for up to 40% of the traffic within major cities in the US [1]. Smart parking technologies are being developed particularly for addressing these challenges.

Most smart parking applications involve collecting the state of parking lots, processing it, and distribute useful information to final users. Nevertheless, these applications, often relying on a blend of different software, hardware and communication devices, are usually assembled aiming at different trade-offs. For example, a number of different line of development have been recently proposed (a complete review can be found in [2], [3], [4]), spanning from crowd-sourcing [5], to sensing parking lots using cameras [6], [7], [8], and predicting the availability of parking lots in the future [9], [10], [11], [12].

Services integrating predictive capabilities, which are one of the key topics of this work, allow drivers to organize their trips either beforehand or during their trips. They not only allow drivers to more informed decision-making processes, but can eventually represent the foundation for more complex services such as dynamic pricing (based on the future states of the parking system) or remote reservation that might even improve the ability of large heterogeneous smart cities to self-regulate themselves in an autonomous fashion.

The contribution of this work is twofold. Firstly, we describe a full-stack deployment of a smart parking system installed in the city center of an Italian mid-sized city. The system is based on 70 parking sensors transmitting data via a LoRaWAN channel to an enterprise IoT platform. Both the sensors and the IoT platform have been provided by Bosch GmbH, while LoRaWAN connectivity has been provided by the A2A Group (a large Italian multi-utility company). We also highlight important aspects related to operational issues emerged during the collection of an original data set 10 weeks long.

Secondly, we apply deep neural networks to the collected data for the sake of forecasting parking lots availability. Experimental results show that our approach is capable of predicting the availability of parking lots 8 hours ahead with an accuracy around 80%. Similar results have also been obtained on an alternative data set collected in Birmingham. To the best of our knowledge, the proposed data collection architecture and prediction methodology uses a combination of smart city concepts, IoT devices, and machine learning techniques still overlooked in literature.

The remainder of this paper is organized as follows. The architecture dedicated to data collection and processing (including practical issues that arose during the deployment), from parking sensors until a REST API exposing the results, is described in Section 2. In Section 3 we statistically characterize the data set that has been collected. In Section 4, we present two prediction approaches for parking lot availability. The first one, based on statistics and used as a baseline, is compared with a more recent one based on convolutional neural networks. In Section 5 we also discuss novel services that could be implemented on top of these predictions and their impact on both parking management and the city as a whole. Finally, related works are discussed

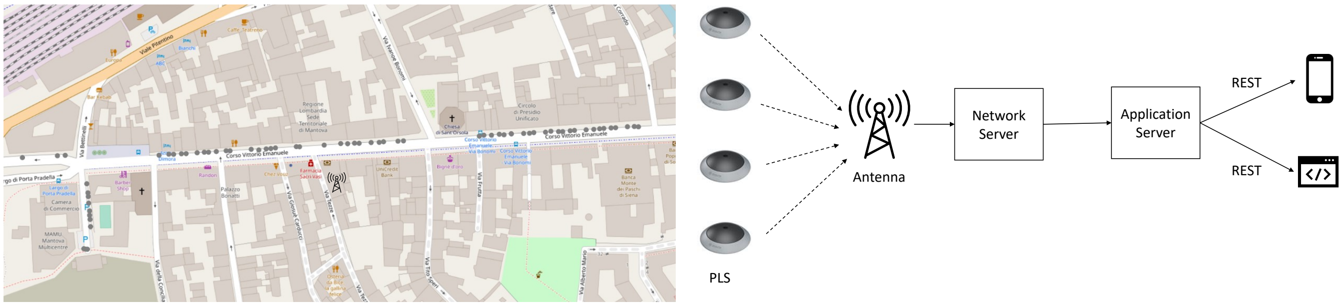


Fig. 1: The deployment area at the center Mantova. 70 sensors (represented as grey dots) and 1 LoRaWAN gate have been deployed (left). The data collection infrastructure used is also represented. Sensors send MQTT packets over LoRaWAN to a dedicated network server (right).

in Section 6. Section 7 concludes the paper.

2 Deployment

2.1 Data Collection Infrastructure

The infrastructure for data collection has been deployed on the main street of Mantova, a mid-sized Italian city. It leverages parking lot sensors and a network gateway deployed on the edge, communicating with a network server routing the messages to cloud-based application servers processing the data and exposing results via REST interfaces as shown in Figure 1.

Concerning the sensor layer, 70 Bosch parking lot sensors have been installed in a straight line in a crowded street of the center. These surface-mount sensors, lasting 5 years with a single charge, use an adaptive model-based detection algorithm developed with examples of real parking events. It has been measured from the vendor that 96% of parking state changes are correctly detected (tested with 2,000 sensors and 46 different car types in real parking environments). These sensors, based on a magnetometer, are also capable of self-learning the needed calibration during the first five parking events and of reporting parking state changes within 35 seconds. They send data encoded in MQTT packets over a low energy LoRaWAN channel (EU868 standard) to a single gate covering an approximate radius of 5 km which has been installed on a telecommunication tower provided by Cellnext, located near the center of the city.

MQTT packets are then routed to an instance of the Pt-NetSuite network server. It is a Linux-based, modular server designed with security and scalability in mind that can be regarded as the backbone of the whole LoRaWAN network. It receives packets from dedicated gates (for this project only one has been installed) and pushes data, without caching for avoiding latency and storage issues, to registered applications via a publish-subscribe service. These communication devices have been provided by the A2A Group.

Once received on our application server, MQTT packets are stored in a MySQL database. The internal structure of MQTT packets sent by Bosch parking lot sensors is composed of 12 different fields describing the status of the parking lot, the sensor itself, and the communication link (see Table 1). Because of most of the fields are not useful for our study on prediction, we stored them in a more compact form (see Table 2). The key fields comprise: a timestamp *(time)*, an id for identifying a specific sensor *(dev_eui)*, a sequence number for recognizing duplicate messages *(seqno)*, an identifier for recognizing system messages from status changes *(msg_type)*, and finally a value describing the most recent status of the sensor *(state)*.

The internal models for the studied prediction algorithms, detailed in Section 4, are trained with these simplified packets. More specifically, we use an online training technique which re-trains the model each month for both improving prediction accuracy and reacting to changes due to unexpected events or seasonal changes.

Whenever a user asks for a prediction using the REST API it is either generated on-request or fetched from a cache. The REST API, written in Flask, implements HTTP endpoints exposing data about both current and predicted conditions of the parking area. Being a web-service, the API can be either queried directly from a client or used indirectly for integrating predictions within other information systems, mobile apps, or vendor-specific services (e.g., Bosch IoT suite). For example, as detailed in Section 5, we developed an Android application capable of guiding users to parking areas likely to have available places at the time of their arrival (under the assumption it will happen within 8 hours).

2.2 Operational issues

During the study two problems materialized. They did not impact the validity of the study and have only an anecdotal value, however, because of their nature, we think their description could add value for the reader, as they emphasize the key role of corporate operations in large IoT deployments [13].

Firstly, after about 6 weeks sensors started to detach from the pavement of the road (glue failure). Some of them have been removed by people while other losses are linked with unusual atmospheric or social events such as the city Carnival floats which passed through the street. As a consequence, whenever sensor where detached from the ground, the number of messages has decreased thus lowering the amplitude of the daily occupancy changes. As shown in Figure 2 the daily difference between maximum and minimum occupancy decreased from 15% in January to about 5% in March. This issue has been carefully discussed with the sensors provider and outlined the key relevance of

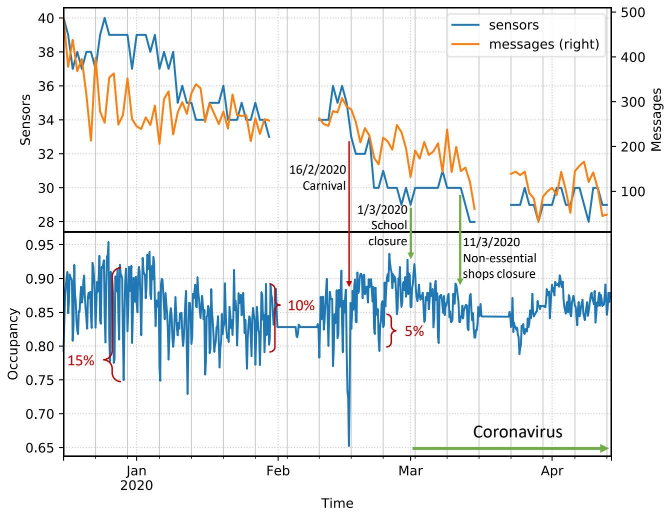


Fig. 2: Comparison of sensors activity (top: number of sensors in blue, number of messages in orange) and parking occupancy (bottom). It is visible the activity decreasing because of the losing of the sensors and the COVID-19 outbreak.

properly executing corporate operations for achieving large-scale deployments.

Secondly, at the beginning of March 2020 (immediately after the end of this study) a general lockdown due to the COVID-19 outbreak has been imposed in Italy. People movements have been restricted and nearly all commercial activities except supermarkets and pharmacies closed. Figure 2 clearly shows a long decline of the occupancy rate from 95% on the 1st of March to 75% on the 1st of April coupled with a reduction of the daily oscillations between daytime and night time. Once lockdown measures have been lifted, the missing sensors have been restored and the occupancy returned to the pre-COVID-19 levels.

3 Data Set Characterization

TABLE 1: An example of MQTT message received from Bosch parking lot sensors.

topic	/unimore/apps/1/devices/fcd6bd000019427f/uplink/2
payload	01
seqno	195
adr	false
channel	1
freq	868.3
modBW	125
rssi	-123
seqno	195
sf	10
snr	-4.7
time	1581266616757

Although the data collection infrastructure is still operational at the time of writing, for this study we focused on a data set comprising 70 days, from 2019-12-16 to 2020-02-23, during which the parking sensors sent on average 500 messages per day (35,000 in total), 90% of which representing status changes *(type = update)* while the remaining 10% representing maintenance messages *(type = heartbeat)*.

TABLE 2: Filtered messages used for storage and analysis purposes.

time	2020-02-09 17:43:37	2020-02-09 17:53:59
dev_eui	fcd6bd000019427f	fcd6bd000019427f
seqno	195	196
msg_type	heartbeat	update
state	1	0

The behavior in time of the collected data has been initially characterized. The area of the city in which parking lots are located greatly influences their daily usage patterns. As an example, a parking space located in a residential area is likely to be fully occupied during the night, while for another one, located close to a commercial area, the opposite is likely true.

In our case, being the sensors installed in a mostly commercial area, the occupancy follows a sine-like wave with more lots occupied during the day and less during the night, as showed in Figure 3-left. However, the observed wave is not pronounced in that the average load lies around 90% and the minimum and maximum loads around 75% and 95%, respectively. Furthermore, according with the fact that urban parking lots change occupant more frequently during daytime, sensor messages have been received at higher frequency between 7 AM and 7 PM. The distribution of daily normalized frequencies is represented in Figure 3-right.

Figure 4 adds further details regarding the behavior in time. In particular, instead of reporting the occupancy rate over time as in Figure 3-left, the daily distribution is reported. The dark area represents the second and the third data quartiles. The whiskers extend to the minimum and maximum values excluding outliers (which are instead represented as dots). Given that the dark area is frequently narrower than 5 lots or 7%, the chart suggests that the system behaves in a reasonably predictable manner. In fact, the narrower the dark area, the more constant is the behaviour week over week.

Figure 4-right, in addition, shows the daily trend of available lots. The chart represent the average number of available lots for a given couple hour/day, for the 10 weeks of the study. Weekends are clearly differentiated from working days. In particular, as a confirmation of the commercial use of the area, the number of available lots is significantly higher during weekend mornings.

4 Lot Availability Prediction

Three different algorithms for predicting the number of available parking lots have been studied. The first two, based on statistics, only capable of learning the average behaviour of the parking area. The third one, based on Convolutional Neural Network (CNN), Long Short-Term Memory cells (LSTM) and Fully Connected layers (FC) has been developed for learning more complex behaviours. The collected dataset has been divided in training and testing sets comprising, respectively, 80% and 20% of the data (28,000 and 7,000 messages).

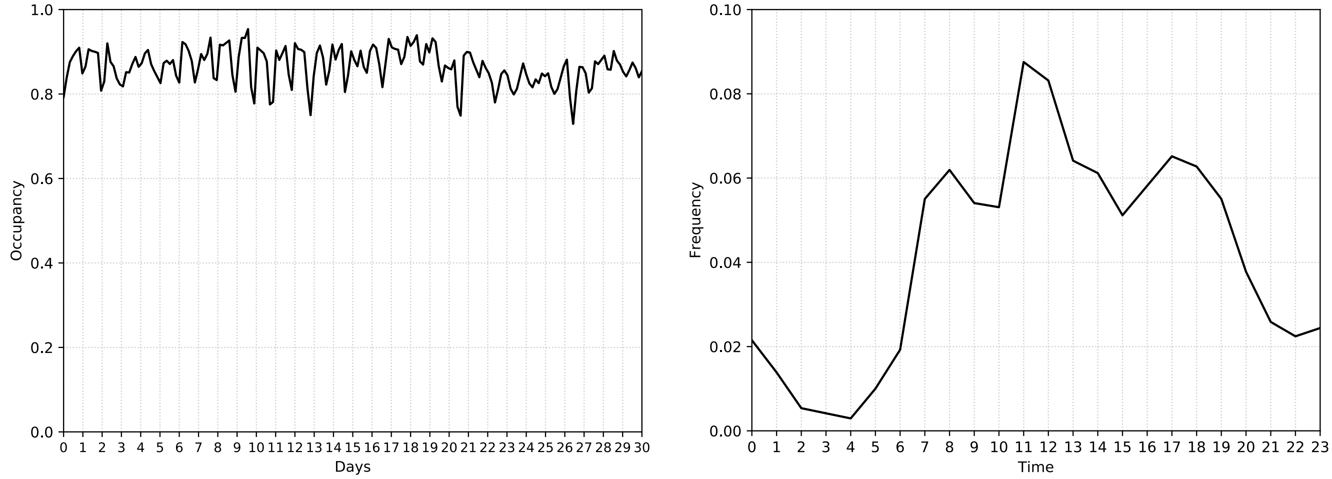


Fig. 3: Parking occupancy rate during the first month. Day 0 corresponds to 2019-12-16. The occupancy is relatively consistent over time varying between 95% and 75% (left). Normalized frequencies of the messages received (right).

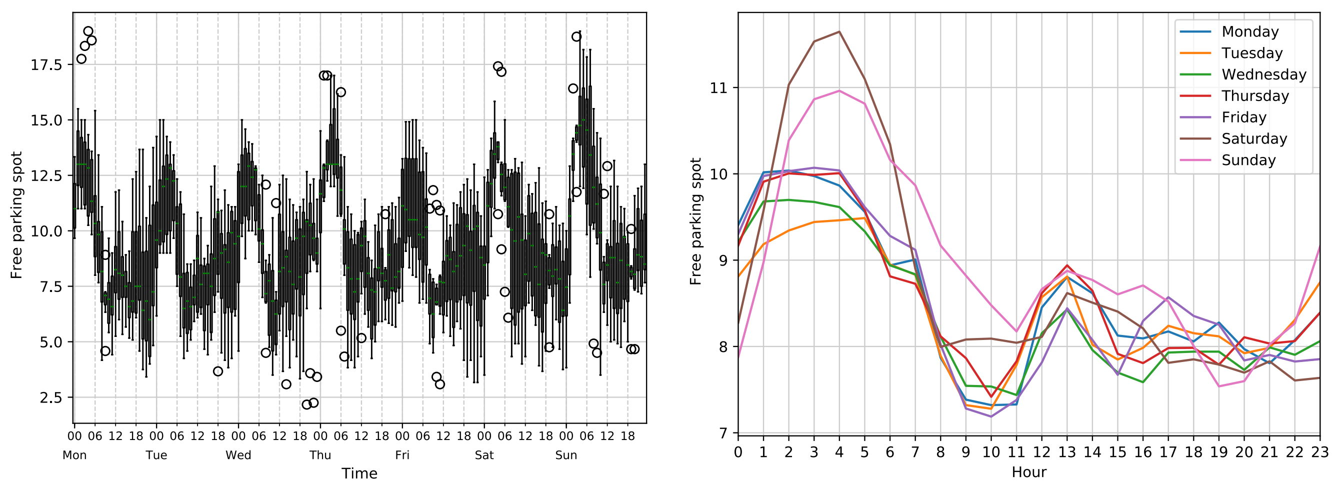


Fig. 4: Parking lots availability distributed over a week: the darker area indicates the second and third data quartiles (left). Parking lots availability represented for specific days of the week. Each line represents a specific day. Given the commercial usage of the area, there is a visible difference between working days and weekend days during the morning (right)

4.1 Statistical Prediction

We started with developing a simple prediction model based on the statistical median. For each day of the week and each hour of the day, the model records the median of all the observed samples. On the one hand, this represent a clear limitation in that the model is unable to learn behaviors distant from the median. On the other hand, despite its simplicity, this approach can provide predictions for arbitrary moments in the future with minimal computation.

Figure 5 shows the results. The median value computed over the 8 weeks used as training set is represented as blue line, while tolerable errors of ± 2 ($\pm 3\%$) and ± 3 ($\pm 4\%$) are represented as the gray and light gray areas, respectively. The number of available parking lots observed in the testing set is depicted as a colored dot (i.e., different colors have been used for each week). The accuracy is calculated as the ratio between the number of points inside the gray areas and the total number of points. Given the reasonably predictable behaviour of the parking area, this approach reaches 67% and 81% of accuracy when considering acceptable an error of $\pm 3\%$ and $\pm 4\%$ respectively.

For the sake of comparing results with other works, an alternative prediction model based on Fourier series has been implemented as well [14]. This model fits a Fourier series to the input data, in this case the entire training set. Then, the computed coefficients of the Fourier series are used to predict the testing set data. Using 75 harmonics on the Mantova data set we obtained an accuracy of 34% when accepting an error of $\pm 3\%$ and 47% when accepting an error of $\pm 4\%$.

These results highlight the circadian patterns of the parking area under study and confirm the actual usability of the approach in real-world deployments where computational resources might be constrained. However, it is relevant to note that this approach is unable to cope with cases in which the number of available lots diverges, for various reasons,

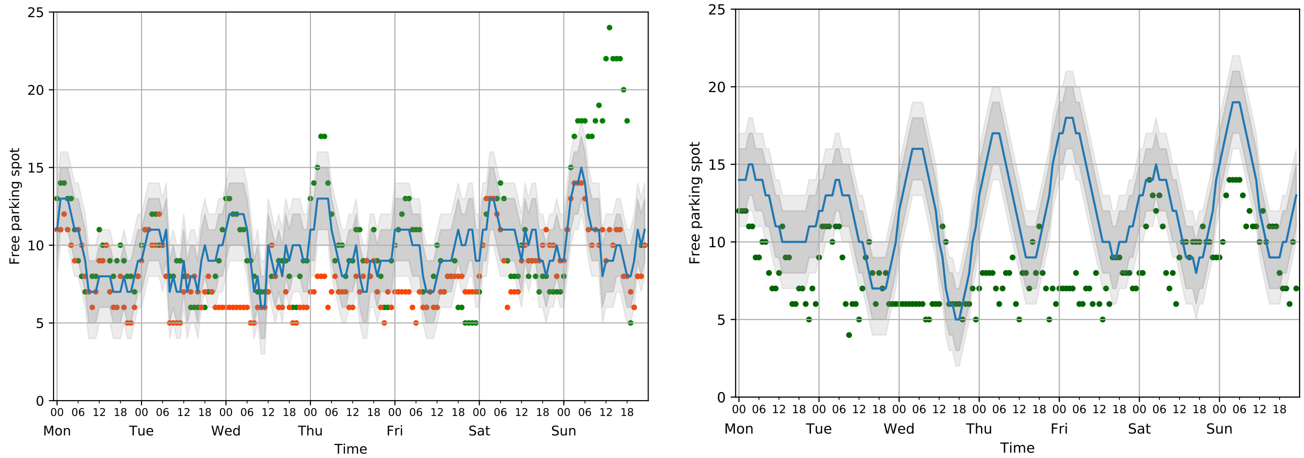


Fig. 5: Predictions computed using both the median behaviour (left) and Fourier series (right). The number of available parking lots during computed from the training set (8 weeks) is represented as a blue line. The gray and light gray areas, instead, represent a band of acceptable error of $\pm 3\%$ and $\pm 4\%$, respectively. Different colors represent different weeks of the testing set.

TABLE 3: Comparison between deep learning architectures the median and Fourier approaches (used as a baseline) on data collected both in Mantova and Birmingham.

	Mantova	Birmingham
MEDIAN ($\pm 3\%$)	0.67	0.71
MEDIAN ($\pm 4\%$)	0.81	0.85
FOURIER ($\pm 3\%$)	0.34	0.32
FOURIER ($\pm 4\%$)	0.47	0.37
CNN+FC ($\pm 3\%$) (1-hour)	0.58	0.85
CNN+FC ($\pm 3\%$) (8-hours)	0.51	0.85
CNN+LSTM+FC ($\pm 3\%$) (1-hour)	0.83	0.84
CNN+LSTM+FC ($\pm 3\%$) (8-hours)	0.76	0.80
CNN+FC ($\pm 4\%$) (1-hour)	0.80	0.91
CNN+FC ($\pm 4\%$) (8-hours)	0.77	0.86
CNN+LSTM+FC ($\pm 4\%$) (1-hour)	0.96	0.86
CNN+LSTM+FC ($\pm 4\%$) (8-hours)	0.93	0.83

from the usual behaviour. These cases are represented by the green dots in the upper right portion of Figure 5-left. Because of this limitation, and because of their ability of learning complex patterns, we also tested the use of deep neural networks on this problem.

4.2 Deep Networks Prediction

Deep networks instead of modelling the behaviour in terms of simple statistical constructs are capable of capturing more complex relations eventually leading to more accurate predictions. This comes with a price: *(a)* from a computational viewpoint training the network is largely more expensive; *(b)* the network we developed cannot predict arbitrary points in time but, instead, requires the last 24 hourly samples of the availability rate for predicting the next 8 hours.

The architecture we have chosen is based on CNN and LSTM layers. On the one hand, CNN layers are useful for both their feature extraction capabilities and their robustness to shifts and distortions. As an example, these layers can identify localized drops or spikes in parking availability due to lunch-breaks. Moreover, because the network uses a small subset of the entire data set it can generalize predictions likely allowing higher accuracy in case of outliers. On the other hand, LSTM layers are often used with time series because of their capacity of learning temporal dependencies from data. This allows to learn longer temporal trends which otherwise would be lost. For example, LSTM allows to effectively encode week-long patterns storing the difference between working days and weekends. A proper combination of both CNN and LSTM layers can take the best from both aspects and improve the overall performance of the network. A set of about 300 different configurations composed by dense, convolutional, and recurrent layers in different orders and with different hyperparameters has been tried before choosing one specific architecture. As shown in Figure 6, it is composed of 1 1D CNN layer, 1 LSTM layer and 3 Fully Connected (FC) layers. The CNN layer is composed by 60 filters (*kernel size = 6*), the LSTM layer is composed of 1,000 cells, and the 3 FC layers are composed respectively of 300, 200 and 100 neurons. There is, lastly, a final FC layer composed of 8 neurons for feature extraction (1 neuron represents 1 hour in the future). For avoiding overfitting, *early stopping* and *dropout = 0.08* have been used. The number of parameters has been reduced to less than 5 millions, mostly located within the LSTM layer.

Figure 7 shows the approach acting on 1 week of the testing set. The dashed blue line represents the predicted number of available lots and the cyan areas represent a band of acceptable error of $\pm 3\%$ and $\pm 4\%$, respectively. The orange line represents the ground truth. In Figure 7-left, each point of the blue dotted line represent the value of the prediction made one hour earlier. In Figure 7-right, instead, there are 8 colored dotted lines each one representing a prediction made n hours before, with $0 < n \leq 8$.

We tested the architecture for predicting both 1 hour and 8 hours in the future, with and without the LSTM layer embedded. Table 3 represents the accuracy obtained with these 4 configurations on two different data sets. It also

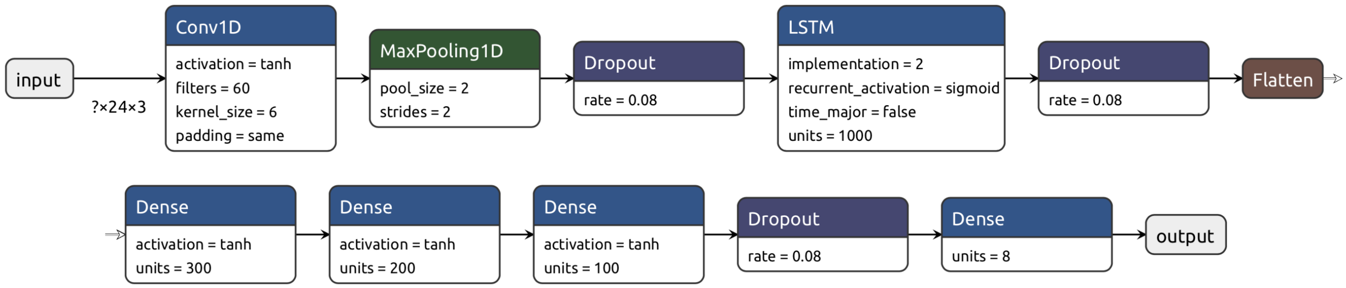


Fig. 6: The architecture used for parking lot prediction. Each layer is represented with its own parameters. The Flatten layer formats the output of the LSTM layer in a shape compatible with the dense layers. The last dense layer is used for extracting the output features.

contains the accuracy achieved with both the median and Fourier approaches used as a baseline. Accuracy is defined as the number of predictions falling within an acceptable error of both $\pm3\%$ and $\pm4\%$ around the ground truth.

Experimental results showed that the precision of the prediction decreases as the distance in time increases. If fact, small changes integrate over time making the final state of the system more uncertain. In general, shifting from 1 to 8 hours in the future reduces the accuracy of the prediction around 5-10%. Moreover, given that future states of the system are not only tied to individual previous states but also to their history over time, the LSTM layer can be beneficial. We observed, in fact, that the LSTM layer implies a 10-40% improvement in prediction accuracy, and that this improvement is more significant when accepting a smaller error ($\pm3\%$). The higher the accuracy required, the more beneficial the introduction of the LSTM layer appears. As an example, the 8 hours prediction without the LSTM layer reaches 51% of accuracy (acceptable error = $\pm3\%$) while the 1 hour prediction with the LSTM layer reaches 83%. These two results translates to 77% and 96% respectively when accepting an error of $\pm4\%$.

4.3 Birmingham Parking Data Set

For the sake of testing the generality of the learning architecture, we also tested it on an well known public data set. We have chosen the *Parking in Birmingham* data set published by Birmingham City Council and licensed under the Open Government License v3.0. It comprises data from several car collected from 2016-10-04 to 2016-12-19 and updated every 30 minutes from 8 AM to 5 PM. Each record of the data set is composed of the following fields: a timestamp, an id for the parking area, the number of occupied parking lots, and the total number of parking lots.

This data set, however, has some known flaws that has been addressed using the approach described in [14]. More specifically: *(a)* the occupancy rate has been normalized on the car park's capacity; *(b)* values above 100% or below 0% have been discarded; *(c)* values associated with invalid timestamps have been discarded; *(d)* days with a low variance in occupancy rate (below 5%) have been discarded. The data set contains data regarding multiple parking area. Among all, we have selected the *Others-CCCPS202*. It comprises 2937 lots and its behaviour is described with 1312 records, each one representing half an hour. Because of the significant differences between the area studied in Mantova and this one (number of lots, missing data, deviation of the daily occupancy, etc.) the network has been re-trained. Figure 8 shows the predictions on 1 week of the testing set using the same graphical formalism used in Figure 7. Given increased size of the parking area, the two cyan areas representing bands of $\pm3\%$ and $\pm4\%$ here correspond to ±88 and ±117 lots.

Table 3 summarizes experimental results and confirms previous observations in Mantova. These results also confirm a reduction in accuracy around 5% when increasing the temporal horizon from 1 to 8 hours ahead. A comparative analysis of the two data sets also suggests that the relevance of the LSTM layer is not universal within the parking availability prediction problem but depends on other factors such as the area of the city in which the parking area is located. As an example, in this data set an improvement in accuracy has been observed when removing the LSTM instead of introducing it. It seems that the LSTM layer is mostly beneficial only when the CNN+FC does not performs well. In fact, the higher the accuracy obtained with the CNN+FC configuration, the smaller the improvement provided by LSTM layer. These differences are likely rooted in differences between the two parking areas. For example, the one in Mantova is open 24/7 while the one in Birmingham only during the day. Furthermore, the daily oscillations look much more repetitive and smooth in Birmingham as shown in Figures 7 and 8.

As a concluding remark, our approach showed a significant generalization capability. In fact, the configuration CNN+LSTM+FC performs uniformly (about 80% accuracy) with both data sets. The only drawback lies in that, when applied to different scenarios, the network must be re-trained. Transfer learning approaches can be applied to overcome this limitation [15].

5 Forecasting-enabled Services

A REST API has been developed for the sake of enabling a range of services and applications. The API exposes via

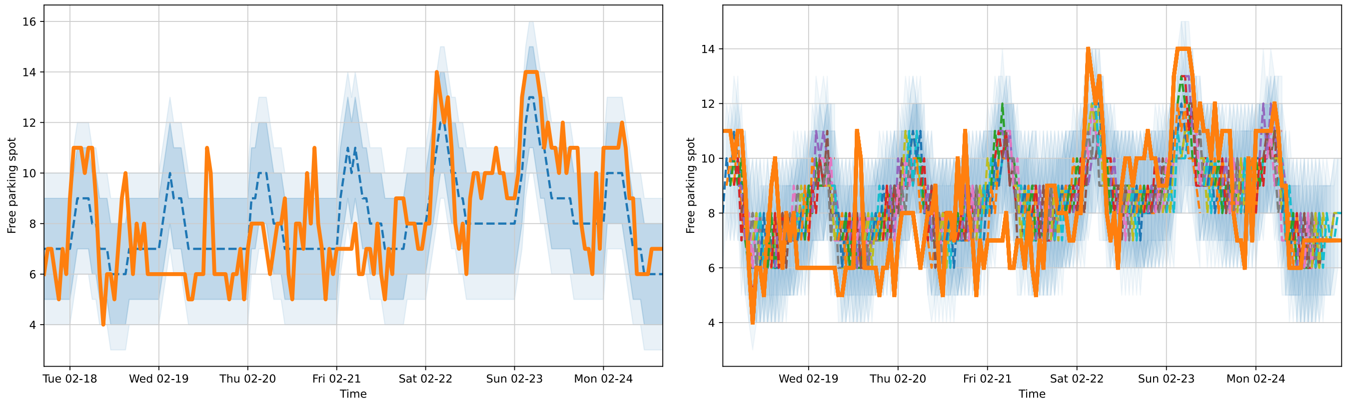


Fig. 7: 1-hour-ahead and 8-hours-ahead NN prediction (Mantova data set). Occupancy at the top, free lots at the bottom. The orange lines are the real values, the dashed lines are the predicted values, the cyan areas are the errors of the prediction (± 2 lots).

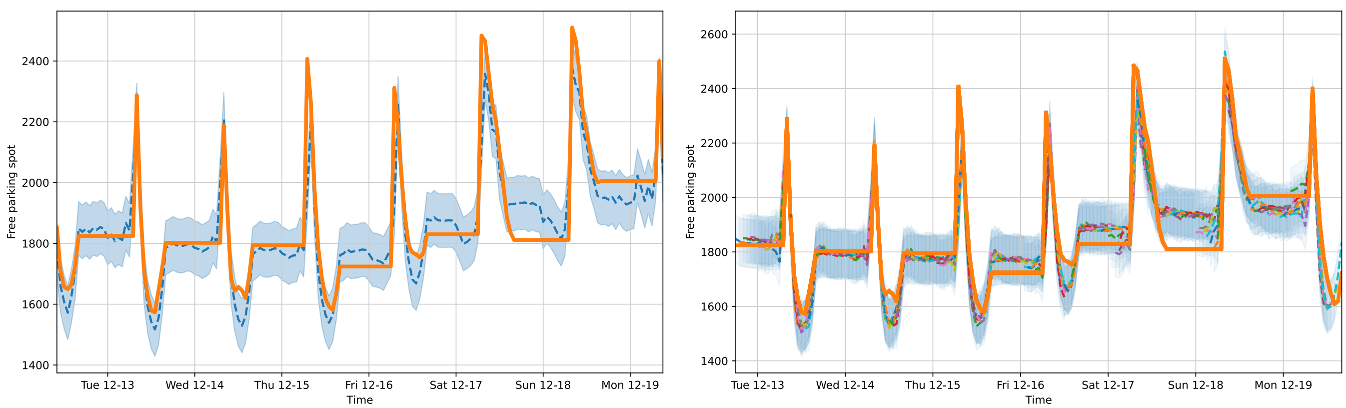


Fig. 8: 1-hour-ahead and 8-hours-ahead NN prediction (Birmingham data set). Occupancy at the top, free lots at the bottom. The orange lines are the real values, the dashed lines are the predicted values, the cyan areas are the errors of the prediction (± 88 lots).

HTTP the past, present, and predicted availability rates of a specific smart parking area (in our case, the one in Mantova).

As a first case study, predictions have been integrated within information signs as shown in Figure 9-left. The number of available parking lots does not represent the current availability, but the predicted one considering the distance between the sign and the parking area. This kind of approach might alleviate one of the key issues associated with information signs and parking guidance. In fact, results from a recent survey [16] shows that 17% of the drivers do not trust such systems believing that the information presented does not reflect the actual situation at the time of arrival.

As a second case study, we developed a mobile application enabling the search for a parking lot using both our API (for sensored parking), and the OpenStreetMaps API (for normal parking). Figure 9-right shows a screenshot of the application. Our application server response contains information about the parking, the current and the predicted number of free lots (within 8 hours). The red column, represents the predicted availability rate at the time of arrival. In case of numbered parking lots, the application can also reserve a specific lot and provide navigation to reach it.

In addition to applications and services targeting citizens, analyzing parking data might benefit local authorities and policymakers as well. For example, precise and timely information about the availability of parking lots can be used to dynamically set the price of parking. On the basis of such information, it might be possible to propose novel pricing models for parking lots which take into account both predicted traffic flows and parking demand. These models, other than likely improve the profitability of parking provides, can also lean toward traffic reduction (e.g., discouraging travel to crowded areas because parking becomes expensive) and, more in general, to more sustainable urban behaviors. Our system, even in its current form, naturally fulfills this need.

6 Related Work

Smart parking is a vast field comprising a number of challenges in different domains. Thus, this section is focused on the areas strictly related to the topics of this paper. Please refer to [2], [3] for a more complete survey concerning all the aspects spanning from sensors, to cloud infrastructures, until machine learning techniques.

(a) (b)

Fig. 9: Information signs integrated with the developed REST API *(left)*. Screenshot of the mobile application showing both normal parking spaces (red points) and smart parking spaces (green icons). When selected, smart parking spaces are shown with the predicted availability for 8 hours ahead. The red column relates to the expected time of arrival *(right)*.

Concerning predictive capabilities, the first works appeared slightly after 2010. They mostly used the Poisson distribution for describing arrival times of cars and Markov chains for predicting the number of available parking lots [17], [18]. However, those assumptions revealed too simplistic in that parking areas are frequently influenced by other factors such as time, day of the week, natural or social events. In addition, in 2016 the Weibull distribution proved to be more accurate than the Poisson distribution for describing arrivals at parking spaces using a realistic data set collected in Santander. [11]. In [9], authors used real-time parking data from the San Francisco area for developing a model capable of predicting the available parking lots for both on-street and off-street parking at the driver's expected arrival time. The model uses both temporal and spatial correlations for estimating parking lots availability. The same data have been also used for comparing clustering techniques [19] aiming at reducing the storage required by prediction systems without impacting accuracy. In [10], [9], regressions, support vector machines, and neural networks have been also compared for predicting parking lots availability. Stolfi et. al [14], tested several forecasting strategies such as polynomial adjustment, fourier series, k-means clustering. All the alternatives tested showed similar forecasting accuracy. For comparison purposes, we implemented the one based on Fourier series and used it a baseline. Results obtained with our CNN+LSTM+FC network are significantly more accurate on the data sets collected both in Mantova and Birmingham (see Table 3). Regarding deep neural networks, they initially have been applied on this problem using feed-forward networks without impressive results. Nevertheless, the increase in available data due to the spread of IoT systems and the introduction of back-propagation significantly improved their accuracy [11]. Camero et. al. [12], presented a deep learning architecture based on recurrent networks for forecasting parking lots availability. The authors also propose to apply the concept of NeuroEvolution for optimizing deep learning architectures in processing parking data. The work of Camero et. al., however, focus mainly on prediction while neglects technological aspects (e.g., IoT) that have been used for further improvements in this paper.

Concerning, instead, the design and architecture of a smart parking system, early works evaluated different sensors (i.e., acoustics, light, magnetic) and collected data for processing on a single server [20]. In [21], a comprehensive analysis of the key aspects for designing a smart parking system has been proposed. The work mainly focuses on identifying and deploying the most suitable sensors for detecting state changes of a lot. Other authors such as [22] used RFID and IR sensors for detecting both the presence of a car and its details. Data were made available to drivers over cloud services. A similar technique, based on RFID sensors, has been tested by Pawowicz et al. [23] for improving traffic flow in the context of a smart city. Another attempt to optimize the usage of parking space in a smart city has been described in [24]. The authors introduced an automated valet for parking based on robotic arms and using Q-Learning as a learning method. Recently, artificial vision has been increasingly introduced in the field. Amato et. al. [7] proposed a parking system using cameras and deep learning for determining the availability of parking lots. Similar proposals such as PKLot and CNRPark-EXT can be also found in [25] and [6]. Furthermore, Mago et. al. [26] proposed another system for detecting available parking lots and assigning arriving vehicles to them. All these these proposals, however, neglect predictive aspects.

7 Conclusion

In this paper we proposed a comprehensive approach suitable for the smart parking domain in which we combine a set of IoT and machine learning technologies still overlooked in literature. The system makes use of 70 parking lot sensors sending data to a dedicated application server via a low-power, low-bandwidth LoraWAN network covering a radius of 5 km. Firstly, this experience showed how the LoRaWAN technology is well suited for IoT applications requiring long running times without recharging or maintenance. It also showed the relevance of corporate operations for creating long lasting, large-scale deployments. Secondly, using this architecture we collected a 10 weeks data set comprising 35,000 data points in the center of Mantova, a mid-sized Italian city. Using this data set and another one publicly available collected in Birmingham, we assessed how deep neural networks perform in predicting the availability of parking lots in the future (within a 8 hours time horizon). Experiments showed that a well constructed deep learning architecture comprising both CNN and LSTM layers provides clear advantages in terms of accuracy over previous approaches. However, the benefits derived from the use of a LSTM layer did not prove universal for the problem but depending on the specific parking area.

Acknowledgment

Anonymous acknowledgments

References

[1] Y. Zhang and A. Haghani, "A gradient boosting method to improve travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 308–324, 2015.

[2] T. Lin, H. Rivano, and F. L. Mouel, "A survey of smart parking solutions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, 2017.

[3] F. Al-Turjman and A. Malekloo, "Smart parking in iot-enabled cities: A survey," *Sustainable Cities and Society*, vol. 49, p. 101608, 2019.

[4] F. Zantalis, G. Koulouras, S. Karabetsos, and D. Kandris, "A review of machine learning and iot in smart transportation," *Future Internet*, vol. 11, no. 4, p. 94, 2019.

[5] C. Zhu, A. Mehrabi, Y. Xiao, and Y. Wen, "Crowdparking: Crowdsourcing based parking navigation in autonomous driving era," in *2019 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, 2019, pp. 1401–1405.

[6] G. Amato, F. Carrara, F. Falchi, C. Gennaro, and C. Vairo, "Car parking occupancy detection using smart camera networks and deep learning," in *2016 IEEE Symposium on Computers and Communication (ISCC)*. IEEE, 2016, pp. 1212–1217.

[7] G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Meghini, and C. Vairo, "Deep learning for decentralized parking lot occupancy detection," *Expert Systems with Applications*, vol. 72, pp. 327–334, 2017.

[8] R. Ke, Y. Zhuang, Z. Pu, and Y. Wang, "A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on iot devices," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.

[9] T. Rajabioun and P. A. Ioannou, "On-street and off-street parking availability prediction using multivariate spatiotemporal models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2913–2924, 2015.

[10] Y. Zheng, S. Rajasegarar, and C. Leckie, "Parking availability prediction for sensor-enabled car parks in smart cities," in *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*. IEEE, 2015, pp. 1–6.

[11] E. I. Vlahogianni, K. Kepaptsoglou, V. Tsetsos, and M. G. Karlaftis, "A real-time parking prediction system for smart cities," *Journal of Intelligent Transportation Systems*, vol. 20, no. 2, pp. 192–204, 2016.

[12] A. Camero, J. Toutouh, D. H. Stolfi, and E. Alba, "Evolutionary deep learning for car park occupancy prediction in smart cities," in *International Conference on Learning and Intelligent Optimization*. Springer, 2018, pp. 386–401.

[13] Yu Huang, Yi-Chao Chen, Chuang-Wen You, Dian-Xuan Wu, Yi-Ling Chen, Kai-Lung Hua, and Jane Yung-Jen Hsu, "Toward an easy deployable outdoor parking system — lessons from long-term deployment," in *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2017, pp. 227–236.

[14] D. H. Stolfi, E. Alba, and X. Yao, "Predicting car park occupancy rates in smart cities," in *International Conference on Smart Cities*. Springer, 2017, pp. 107–117.

[15] K. Weiss, T. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 9, 2016.

[16] Z. Liu, W. Deng, and D. Pan, "Driver response to parking guidance and information systems in nanjing," in *International Conference of Chinese Transportation Professionals (ICCTP)*, Nanjing, China, 2011.

[17] A. Klappenecker, H. Lee, and J. L. Welch, "Finding available parking spaces made easy," *Ad Hoc Networks*, vol. 12, pp. 243–249, 2014.

[18] S. Pullola, P. K. Atrey, and A. El Saddik, "Towards an intelligent gps-based vehicle navigation system for finding street parking lots," in *2007 IEEE International Conference on Signal Processing and Communications*. IEEE, 2007, pp. 1251–1254.

[19] F. Richter, S. Di Martino, and D. C. Mattfeld, "Temporal and spatial clustering for a parking prediction service," in *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*. IEEE, 2014, pp. 278–282.

[20] R. Kumar, N. K. Chilamkurti, and B. Soh, "A comparative study of different sensors for smart car park management," in *the 2007 International Conference on Intelligent Pervasive Computing (IPC 2007)*. IEEE, 2007, pp. 499–502.

[21] M. Bachani, U. M. Qureshi, and F. K. Shaikh, "Performance analysis of proximity and light sensors for smart parking," *Procedia Computer Science*, vol. 83, pp. 385–392, 2016.

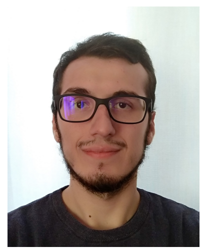
[22] B. K. Gandhi and M. K. Rao, "A prototype for iot based car parking management system for smart cities," *Indian Journal of Science and Technology*, vol. 9, no. 17, pp. 1–6, 2016.

[23] B. Pawłowicz, M. Salach, and B. Trybus, "Infrastructure of rfid-based smart city traffic control system," in *Conference on Automation*. Springer, 2019, pp. 186–198.

[24] N. Shoeibi and N. Shoeibi, "Future of smart parking: Automated valet parking using deep q-learning," in *International Symposium on Distributed Computing and Artificial Intelligence*. Springer, 2019, pp. 177–182.

[25] P. R. De Almeida, L. S. Oliveira, A. S. Britto Jr, E. J. Silva Jr, and A. L. Koerich, "Pklot–a robust dataset for parking lot classification," *Expert Systems with Applications*, vol. 42, no. 11, pp. 4937–4949, 2015.

[26] N. Mago and S. Kumar, "A machine learning technique for detecting outdoor parking," *International Journal of Engineering & Technology*, vol. 7, no. 2.30, pp. 39–43, 2018.



Manuele Barraco is Master student in Computer Science at the University of Modena and Reggio Emilia ....

Nicola Bicocchi is associate professor in Computer Science at the University of Modena and Reggio Emilia.....

Marco Mamei is full professor in Computer Science at the University of Modena and Reggio Emilia, since 2014. He received the PhD in Computer Science from the same University in 2004. His work focuses on data mining applied to mobile phone data and Internet of Things applications. In these areas we published more than 100 papers, 8 patents, and won several best paper awards.

Franco Zambonelli is full professor in Computer Science at the University of Modena and Reggio Emilia.....