Efficient prediction of spatio-temporal events on the example of the availability of vehicles rented per minute

Bartlomiej Balcerzak[1], Radoslaw Nielek[1,2], and Jerzy Pawel Nowacki[1]

[1] Polish-Japanese Academy of Information Technology, Warsaw, Poland
{b.balcerzak,nielek,nowacki}@pja.edu.pl
[2] Vooom Sp. z o.o., ul. Jana Czeczota 9, 02-607 Warsaw, Poland

Abstract. This article shows a solution to the problem of predicting the availability of vehicles rented per minute in a city. A grid-based spatial model with use of LSTM network augmented with Time Distribution Layer was developed and tested against actual vehicle availability dataset. The dataset was also made publicly available for researchers as a part of this study. The predictive model developed in the study is used in a multi-modal trip planner.

Keywords: carsharing · spatio-temporal events · machine learning · prediction · city

1 Introduction

Nowadays, the need for mobility in cities is increasingly met by free-float vehicles sharing systems. Typically, such systems allow vehicles to be rented and returned in any, location within the city area. The specificity of the system is the reason why both from the perspective of service providers and customers, there is a need for efficient prediction of vehicle demand and availability in terms of time and space.

Prediction of spatio-temporal events in urban space is a difficult task due to the number of dimensions that need to be taken into account and high level of non-linearity. The availability of vehicles depends on the day of the week, time of day and location. Next to zones with low and high demand, we also have areas where cars cannot be parked such as closed areas (e.g. military base), parks, ponds, rivers, or pedestrian zones. Moreover, European cities, as opposed to American ones, usually have irregular road networks, what makes the task even more complex. Due to the fact that the prediction results are intended for use in a commercial route planning tool[3], the model, in addition to high precision, must be characterized by high level of adaptability, robustness, and moderate demand for computing power.

[3] The prediction module based on results described in this paper is part of the commercial MaaS solution offered by Vooom Inc. available at https://planner.app.vooom.pl/

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_8

1.1 Contribution

In this article, we present attempts to build an efficient prediction model of the cars availability with the aim to use it in an intelligent multi-modal trip planner. Planning a transfer in intra-city trips using shared free-floating vehicles requires confirmation of the availability of the means of transport in a specific place in the perspective of several dozen minutes (usually in a horizon up to 60 minutes).

The model we propose is new contribution to the field, since it mostly focuses on the supply side of the car sharing market. Thus, it is based of the trends in availability of cars in a specific city. What is more, by being based on the supply side of things, the model does not require user information for profiling and efficient work. Finally, the model we propose offers means for predicting car availability in long time spans, exceeding the standard time frames for booking cars for short term rent. Creating a city spanning model offers also a challenge for creating efficient architectures for predicting the dynamics of objects sparsely spread within the city grid.

2 Related Work

Prediction of spatio-temporal events in the context of urban transportation has a variety of practical applications, and, therefore, it has been intensively researched in recent years. Li et al. have studied the problem of the availability of free parking lots in NYC [9]. They tested four different spatial units, such as: POIs (Places of Interest), streets, census tracts and 1km grid. It turns out that the best results can be achieved with a combination of 1km grid and random forest algorithm. Robustness of this approach were also confirmed by Goa et al. in a work focused on legality of on-street parking [6], but other methods, such as Graph-Convolutional Neural Networks [15] or Long Short Memory Model (LSTM) [11], were also tested with good results.

Studies done by Schmoeller et al. [10] and Wagner et al. [12] revealed that free-floating carsharing usage concentrates on relatively few hot spots, POIs, such as restaurants and stores, can explain part of spatial variation in car-sharing activities, and the fact that strong imbalance spatial imbalance exists between vehicles demand and supply. As shown by Willing et al. [14], spatial and temporal dimensions are interdependent.

In order to reduce spatio-temporal supply and demand mismatch, various relocation strategies were studied [13]. The majority of studies were focused on solving relocation problem in one-way, station-based systems (see [8] for literature review) or algorithmic aspects of multidimensional optimization problem with constrains (e.g. with use of evolutionary algorithm [7] or operational research [4]). Noteworthy exceptions are works done by Ai et al. [1] (convolutional long short-term memory network applied to prediction of dockless bikes distribution), Bao et al. [2] (hybrid deep learning neural networks used to predict bike availability within 15, 20 and 30 minutes), Formentin et al. [5] (prediction of the distance to the nearest available vehicle), and Dario et al. [3] (an extensive study on the impact of the type of learning algorithm, time horizon and the specificity

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_8

of the environment on the accuracy of prediction of vehicle availability in a given area).

3 Model and problem definition

Having the presented research in mind, we decided to propose a solution that mainly focuses on the supply side of the car sharing market. Based on the literature review, we decided to use an LSTM model augmented with a Time Distributed Layer with information about car location taken from the last 60 minutes used as input. This architecture has proven to be the most efficient of those we tested in our work. The location is a 100m x 100m square. The total space used in our study (the city of Warsaw and it's direct surroundings) are represented as a matrix of 40 thousands such squares, arranged in 200 by 200 squares shape (thus, covering a shape of 400 squared kilometers). This granularity allows the model to predict car availability at a close proximity of the user. The selection of a 100m x 100m cell (henceforth defined as a location) was chosen based on usability of the model. 100m distance is within a walking distance of a user searching to find available vehicles, and thus our model requires this level of precision.

As a baseline model, we decided to use a simple model which assumes that all cars available at a given time in a given location would remain available at that location. We wanted to test the model's performance at a 15,30,45 and 60 minute mark.

The model is trained in real time, with each training step occurring on a minute by minute basis. During each training step, the current map of Warsaw (herein designated as t_0 is treated as a target, with 10 past maps serving as inputs. For example, in the case of the 15 minute iteration of the model, these 10 inputs are randomly selected from a time span between t_{-15} and t_{-15-60}. This additional randomisation step allows us to minimize the risk of overfitting. The model weights are reset every 24 hours at 4am, since, based on observations of the model behavior, doing so prevents the occurrence of concept drift.

4 Dataset

In our work we used a dataset spanning a period of 30 days between November 1st and November 30th.The data was collected through the use of APIs made available by car rental providers operating in Warsaw, in real time and stored in an SQL database. Due to occasional issues with data availability (instances when API failed to provide data at a given time), the dataset was pruned in order to remove timestamps with a significantly lower number of missing values(these were datastamps that yielded a number of available cars lower than the mean daily value by two standard deviations).

The dataset represents a minute-by-minute availability of cars for rent within the boundaries of the city of Warsaw during daytime hours (between 6am and 8pm). Performance comparison within night time (from 1am to 6am) produced a similar pattern to day time hours.Table 1 below represents the basic information

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_8

about our dataset, while Figure 1 represents an example of a single observation from the dataset.

Property	Value
Number of observations	25200 (over 50 million records; 4.5 GB raw data)
Size of the observation	200x200 binary matrix with each cell representing a 100x100 meter square.
Coordinates delimiting the analyzed area	20.9° - 21.1°, 52.14°-52.35°
Time period represented in the data	11.01.2020 - 11.30.2020
Hour time span represented in the data	6am - 8pm
Average number of locations with available cars	646.58
% of locations where no cars were observed	64.1%

Table 1. Basic characteristics of the dataset

The data collected represents a span of 30 days, thus giving us an opportunity to view how well does the prediction model work during a prolonged period. In addition, the average number of occupied locations at a single moment (around 646), when compared to the size of the grid (40,000 locations) indicates, a very sparse matrix, which provides an additional level of challenge for the classifier. We made the dataset available online[4].

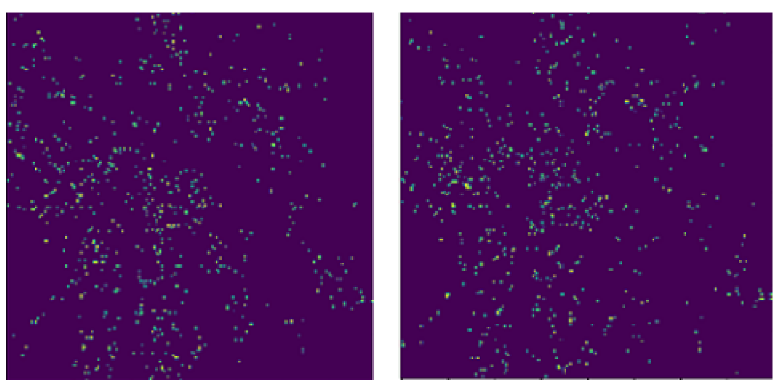

Fig. 1. A heat-map representation of the locations of available cars on December 3rd 2020. The maps were taken for 8am (left), 11am (right).

[4] http://nielek.com/datasets/ICCS2021.html

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_8

4.1 Performance measures

From a business perspective, there are two instances of incorrect prediction which lead to the most critical outcomes. Firstly, a case when the classifier incorrectly predicts the car to be available at a given location in the future. In such a case the travel planner leaves the user at a place with possibly no available transport, thus increasing user frustration.

Secondly, the classifier may predict a location to have no available vehicles, while such vehicles are present at the location. That may lead to sub optimal routes, thus increasing travel time.

Therefore, our selected measures of classifier performance ought to take these scenarios in to account. With that in mind, we decided to focus on two main measures of classifier performance, i.e. positive class precision and positive class recall.

In order to provide a more general view of the model's performance, we will primarily use F1-score as method for evaluating the feasibility of the proposed model. Since the output of the model is a 200x200, the precision and recall can be calculated on a minute-by-minute basis, thus providing information about the number of correctly predicted locations at a given time.

5 Results

Firstly, we decided to compare results for the entirety of the 30 day period. These aggregated results are presented in Table 2. While the performance measured with the F-1 score shows no improvement over the baseline in the 15 minute time frame, the model outperforms the base line in the more distant time frames. One can notice gradual descent of the F-1 score value for the baseline, compared with the more consistent results of the model. What is more, when it comes to the precision score, the model outperforms the baseline in all selected time frames. In the case of the recall, the base line proves superior only in the 15 minute mark. These results may be partially explained by the very nature of car sharing. Most providers allow for the booking of a car to stay active for 15 minutes, thus reducing the mobility of the cars within this time frame. This feature also makes the 15 minutes time frame less crucial for the performance of the model, while further emphasizing the importance of prediction in the long term, where the user has no means of booking a car for that long.

Time Frame	Precision/Recall/F1-Score	Baseline Precision/Recall/F1-Score
15 minutes	0.9/ 0.83/0.87	0.86/ 0.86/0.86
30 minutes	0.84/0.85 /0.84	0.79/0.8 /0.79
45 minutes	0.81/0.87 /0.83	0.76/ 0.76/0.76
60 minutes	0.77/ 0.87/0.81	0.72/ 0.72/0.72

Table 2. Comparison of the averaged performance measures for the entire month of November. Prediction threshold of 80%

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_8

In addition a logistic regression model was also used, the results were below the baseline level. In order to further review the robustness of our model, we decided to analyze it's performance during different times of the day, and different days of the week. Results are presented in Figure 2. The model seems to be giving the best performance during morning hours and in the early afternoon, with a significant drop around noon. A similar trend can be observed when comparing performance by weekdays, with higher scores being observed on weekends.

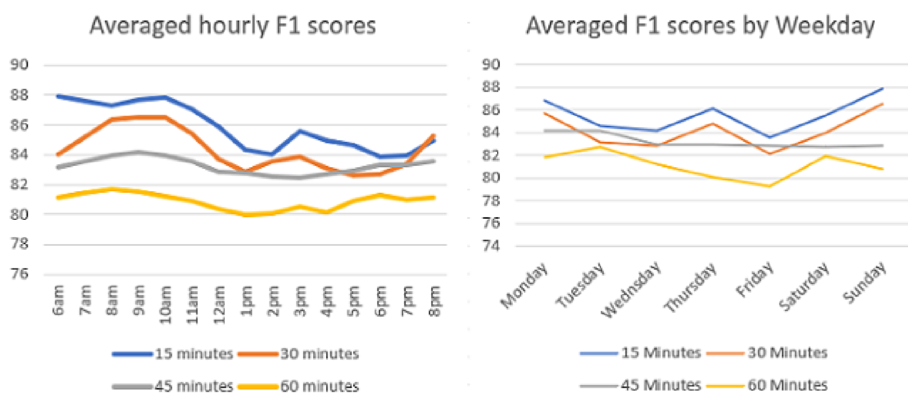

Fig. 2. Averaged F1 scores by hour (left) and by weekday (right)

6 Conclusion

In this article we wanted to propose model allowing for the prediction of the availability of cars for rent in Warsaw. The model was able to predict the availability of cars within the time spans ranging from 15 to 60 minutes. The results we obtained, have shown that our model outperforms the baseline when it comes to the precision of the prediction. This improvement in performance was mostly observed when considering time frames such as 45 and 60 minutes marks. The results have also shown that the prediction model exceeds the 80% threshold for the F1 score, regardless of the time of day and time of week; although higher performance was noted in morning hours and during the weekends. This effect, however, was weaker when predicting longer time spans. The model proposed in this article proved to be an effective method of predicting car for rent availability in short time spans. In our future work, we will aim to extend the model to cover other shared means of transportation, as well as test to the validity of the model in the case of other cities.

Acknowledgments. Project was partially financed by EU European Regional Developement Fund within the Inteligent Developement Program. Project realised within the Narodowe Centrum Badań i Rozwoju Szybka Ścieżka program.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_8

References

1. Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., Ju, Y.: A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. Neural Computing and Applications **31**(5), 1665–1677 (2019)
2. Bao, J., Yu, H., Wu, J.: Short-term ffbs demand prediction with multi-source data in a hybrid deep learning framework. IET Intelligent Transport Systems **13**(9), 1340–1347 (2019)
3. Daraio, E., Cagliero, L., Chiusano, S., Garza, P., Giordano, D.: Predicting car availability in free floating car sharing systems: Leveraging machine learning in challenging contexts. Electronics **9**(8), 1322 (2020)
4. Folkestad, C.A., Hansen, N., Fagerholt, K., Andersson, H., Pantuso, G.: Optimal charging and repositioning of electric vehicles in a free-floating carsharing system. Computers & Operations Research **113**, 104771 (2020)
5. Formentin, S., Bianchessi, A.G., Savaresi, S.M.: On the prediction of future vehicle locations in free-floating car sharing systems. In: 2015 ieee intelligent vehicles symposium (iv). pp. 1006–1011. IEEE (2015)
6. Gao, S., Li, M., Liang, Y., Marks, J., Kang, Y., Li, M.: Predicting the spatiotemporal legality of on-street parking using open data and machine learning. Annals of GIS **25**(4), 299–312 (2019)
7. Herbawi, W., Knoll, M., Kaiser, M., Gruel, W.: An evolutionary algorithm for the vehicle relocation problem in free floating carsharing. In: 2016 IEEE Congress on Evolutionary Computation (CEC). pp. 2873–2879. IEEE (2016)
8. Illgen, S., Höck, M.: Literature review of the vehicle relocation problem in one-way car sharing networks. Transportation Research Part B: Methodological **120**, 193–204 (2019)
9. Li, M., Gao, S., Liang, Y., Marks, J., Kang, Y., Li, M.: A data-driven approach to understanding and predicting the spatiotemporal availability of street parking. In: Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 536–539 (2019)
10. Schmöller, S., Weikl, S., Müller, J., Bogenberger, K.: Empirical analysis of free-floating carsharing usage: The munich and berlin case. Transportation Research Part C: Emerging Technologies **56**, 34–51 (2015)
11. Shao, W., Zhang, Y., Guo, B., Qin, K., Chan, J., Salim, F.D.: Parking availability prediction with long short term memory model. In: International Conference on Green, Pervasive, and Cloud Computing. pp. 124–137. Springer (2018)
12. Wagner, S., Brandt, T., Neumann, D.: In free float: Developing business analytics support for carsharing providers. Omega **59**, 4–14 (2016)
13. Weikl, S., Bogenberger, K.: Relocation strategies and algorithms for free-floating car sharing systems. IEEE Intelligent Transportation Systems Magazine **5**(4), 100–111 (2013)
14. Willing, C., Klemmer, K., Brandt, T., Neumann, D.: Moving in time and space–location intelligence for carsharing decision support. Decision Support Systems **99**, 75–85 (2017)
15. Yang, S., Ma, W., Pi, X., Qian, S.: A deep learning approach to real-time parking occupancy prediction in transportation networks incorporating multiple spatio-temporal data sources. Transportation Research Part C: Emerging Technologies **107**, 248–265 (2019)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_8