Understanding Traffic Cruising Causation

Via Parking Data Enhancement

Mirza Jasarevic

This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfilment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

The authors declare that they are the sole authors of this thesis and that they have not used any sources other than those listed in the bibliography and identified as references. They further declare that they have not submitted this thesis at any other institution to obtain a degree.

Contact Information:

Author(s): Mirza Jasarevic

E-mail: mijf12@student.bth.se

University advisor:

Prof. Veselka Boeva

Department of Computer Science

Faculty of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

Abstract

Background. Some computer scientists have recently pointed out that it may be more effective for the computer science community to focus more on data preparation for performance improvements, rather than exclusively comparing modeling techniques. Testing how useful this shift in focus is, this paper chooses a particular data extraction technique to examine the differences in data model performance.

Objectives. Five recent (2016-2020) studies concerning modeling parking congestion have used a rationalized approach to feature extraction rather than a measured approach. Their main focus was to select modeling techniques to find the best performance. Instead, this study picks a feature common to them all and attempts to improve it. It is then compared to the performance of the feature when it retains the state it had in the related studies. Weights are applied to the selected features, and altered, rather than using several modeling techniques. Specifically in the case of time series parking data, as the opportunity appeared in that sector. Apart from this, the reusability of the data is also gauged.

Methods. An experimental case study is designed in three parts. The first tests the importance of weighted sum configurations relative to drivers' expectations. The second analyzes how much data can be recycled from the real data, and whether spatial or temporal comparisons are better for data synthesis of parking data. The third part compares the performance of the best configuration against the default configuration using k -means clustering algorithm and dynamic time warping distance.

Results. The experimental results show performance improvements on all levels, and increasing improvement as the sample sizes grow, up to 9% average improvement per category, 6.2% for the entire city. The popularity of a parking lot turned out to be as important as occupancy rates(50% importance each), while volatility was obstructive. A few months were recyclable, and a few small parking lots could replace each other's datasets. Temporal aspects turned out to be better for parking data simulations than spatial aspects.

Conclusions. The results support the data scientists' belief that quality- and quantity improvements of data are more important than creating more, new types of models. The score can be used as a better metric for parking congestion rates, for both drivers and managers. It can be employed in the public sphere under the condition that higher quality, richer data are provided.

Keywords: Machine learning, Unsupervised learning, Time series, Parking, Dynamic time warping

Acknowledgments

First and foremost I have to thank professor Veselka Boeva, my supervisor through this trialing moment, for providing me with invaluable advice and a sense of direction when I was lost. Her experience and competence were indispensable, as she was always astute and able to spot a dead-end and mistake from a mile away, sparing me a lot of grief.

Secondly, I owe a lot of gratitude to the municipality of Karlskrona, especially Per-Olav Gramstad, Fredrik Sjölin, Mostafa Hassan, Mats Hellman, and Johan Lindh who helped facilitate, organize and coordinate the necessary data, as well as meetings and discussions to help materialize the plan and the study itself. Despite their busy schedules, they made sure to reply quickly to all inquiries, and they made the administrative processes run very smoothly, which is highly commendable.

I also want to thank my relatives deserve great, big thanks as well for their enthusiasm and support. Though they could not help me directly, they tried their utmost to handle errands on my behalf so that I could spend more time studying.

Contents

Abstract	i
Acknowledgments	iii
1 Introduction	2
1.1 Problem statement	2
1.2 Scope and Research Questions	2
1.3 Summary of Research Questions	3
1.4 Aim and Objectives	3
1.5 Outline	4
2 Related Work	5
2.1 Search Criteria	5
2.1.1 Databases	5
2.1.2 Search terms	5
2.1.3 Selection criteria	5
2.2 Literature Review	6
2.3 Related Work	6
2.3.1 Summary and gaps	7
3 Method	8
3.1 Setup	8
3.1.1 Why an experiment?	8
3.2 Materials, methods and preprocessing	8
3.2.1 Data	8
3.2.2 K-means clustering algorithm	9
3.2.3 Dynamic time warping	9
3.2.4 Silhouette Analysis	9
3.2.5 Data analysis, preparation and enhancement	10
3.2.6 Incorporation of driver and expert knowledge as validation sets	12
3.2.7 Variables	13
3.2.8 Data limitations	13
3.2.9 Implementation	13
4 Results and Analysis	15
4.1 Results addressing RQ1	15
4.2 Results addressing RQ2	16
4.3 Results addressing RQ3	21

5	Discussion	24
5.1	Validity Threats	25
5.1.1	External Threats	25
5.1.2	Internal Threats	25
5.1.3	Conclusion Validity	25
6	Conclusions and Future Work	26
6.1	Future Work	26
A	Supplemental Information	30
A.1	Truth table	30

List of Figures

4.1	Overview of combined K-S and E-S heatmaps.	17
4.2	Identities of the 146 substitutable months	17
18		
4.4	Arrival distributions: Entire city vs. town square only.	19
4.5	Kurtosis-Skewness ratios: Red Zone	19
4.6	Kurtosis-Skewness ratios: Green Zone	20
4.7	Kurtosis-Skewness ratios: Months	20
4.8	Sums and averages of silhouette analysis	23
4.9	Sums and averages of silhouette analysis	23

List of Tables

3.1	Example data	8
3.2	A truth table built through incorporating driver and expert knowledge	12
4.1	Results of all combinations	15
4.2	Averages of 10 machines	15
4.3	Accuracy rate for factor intensities	16
4.4	Measurements of machine <i>BA 8</i> for 2018	20
4.5	Congestion score silhouettes vs. occupancy-rate silhouettes	21
4.6	Compared performance of clustering groups	22
4.7	Comparison of machine silhouette analysis results	22

Glossary

- **Decision Support System:** Abbreviated "*DSS*"; this is an often multifaceted program displaying high-level business information, such as graphs, from collected data for domain experts [1].
- **Machine learning:** This is a subset of artificial intelligence programming and statistical techniques used to imitate the human act of learning [2].
- **Unsupervised learning:** A subset of machine learning focused on pattern recognition [3].
- **K-means:** A subset of unsupervised learning to divide a set of data into groups, or clusters rather, according to the attributes of the entries [4].
- **Dynamic Time Warping:** Abbreviated "*DTW*"; a way of measuring and comparing two time series of data to decide how similar they are. This is used in unison with k-means [5].
- **Granularity:** This denotes the level of detail or size of the data units being processed.
- **Cruising Risk/Congestion Rates:** These terms are used interchangeably to denote the likelihood of a driver having to leave a selected parking lot after entering it, and cruise for parking. They are interchangeable in this context because the more a parking is congested, the higher the risk is that the driver must leave the parking to look elsewhere.
- **Parking lots/Machines:** These terms are also used interchangeably to denote the place where drivers park. Each parking lot has a payment machine where drivers pay after parking, then the machine registers the stay. Because the machines identify where the rows of data came from, they represent their parking lots, therefore they are used as the identity of parking lots and denote parking lots.
- **Kurtosis:** A measure for the sharpness of the distribution.
- **Skewness:** A measure for the slant and width of the distribution.
- **Empirical Distribution Function:** Abbreviated "*EDF*"; The distribution function of the sample. Each one has a complementary Empirical Characteristic Function (ECF).
- **Kolmogorov-Smirnov Test:** A statistical test to compare two EDFs.
- **Epps-Singleton Test:** Another statistical test to compare two ECFs.

In late 2019, the municipality of Karlskrona presented an opportunity to explore their parking data. Their stated goal was to explore the data and to find a way to use it for parking congestion management, to guide drivers to less congested parking lots. The reason for this was increased reports of drivers unable to find parking spaces. They expressed interest in finding out how machine learning could help to mitigate the issue.

1.1 Problem statement

On the outset, most machine learning projects going forward from that moment in time at BTH were head-on comparisons of different machine learning models to find the best one suited for the problem. But as Andrew Ng pointed out in one of his articles for DeepLearningAI; there may be more value in terms of efficiency to improve the data instead, to take a data-centric approach rather than a model-centric approach [6]. Some of the cited examples were projects where defects in steel production were detected through object detection algorithms. The baseline performance, when using an up-to-date model, would be 76.2% accuracy for instance, and improving the model itself would result in 0% improvement. Whereas improving the data resulted in a 16.9% improvement. For solar panel defect detection, the baseline was 75.68%, improvement of the model resulted in 0.04% better accuracy, while improvement of the data resulted in 3.06% better accuracy. A third example was a surface inspection problem solved through the same object detection approach, the baseline accuracy was 85.05%, model improvement gave 0% better accuracy, but data improvement gave 0.4% better accuracy. On all counts, the data improvement outperformed the model improvement by big margins. This may also be the case with parking data.

1.2 Scope and Research Questions

There are many readily available studies comparing techniques and models themselves already, some of which we will see in the section *Related Work*, making another one potentially redundant. This motivates the paper's work to be focused on studying how the parking data improvement would benefit for better understanding of traffic cruising causation. The improvements are based on the summarized opinions of several veteran drivers and GIS experts from the municipality. These opinions are

condensed into a truth table to represent their consensus, where 125 different parking situations are presented in numbers, and the respondents write whether they would attempt to park under those circumstances or look elsewhere. The parking situations are described by combining three factors (occupancy, volatility, and popularity) with varying degrees of importance. The interest here forms the first research question (**RQ1**); *based on domain knowledge, what weights should be assigned to the factors?*

The created truth table is the basis for a multi-attribute utility function used to automatize classifying of 3416400 different pre-processed data points, to validate and establish ground truth. This matches the optimal configuration before the later performance comparison through clustering and silhouette analysis. The silhouette analysis acts as a representative for how clear and confident the prognosis of the algorithm are. The performance is compared to the performance of a version mimicking that of related studies. The reason for this approach is two-fold: (i) the number of prognoses required to give useful results takes too much time to do manually; (ii) an element of human judgment must be introduced to get results tailored to human preferences per human judgments and consumption. This forms the third research question (**RQ3**); *do the factors with the correct domain knowledge configuration make the cluster silhouette rating better?*

In addition to the above, a section is dedicated to the analysis of the parking data, attempting to explore the potential for data recycling and data synthesis where data are missing. This leads to the second research question (**RQ2**); *can results from one parking lot be applied to a parking lot missing data for the same time period?*

1.3 Summary of Research Questions

To summarize the research questions, they are:

- **RQ1:** *Based on domain knowledge, what weights should be assigned to the factors?*
- **RQ2:** *Can results from one parking lot be applied to a parking lot missing data for the same time period?*
- **RQ3:** *Do the factors with the correct domain knowledge configuration make the cluster silhouette rating better?*

1.4 Aim and Objectives

To summarize the context, the goals of this study are:

- To use a data-centric approach, comparing a data composition of previous works, to a more nuanced data composition.
- To find out which of the three factors used are important when modeling parking cruising causation.

- To find what, if any, aspects of the parking data are more useful for simulation and recycling of the data.
- To confirm or reject the suspicion that a data-centric approach improves the accuracy of the results.

1.5 Outline

The next section will examine some similar works and their approaches, comparing them to this study.

After that, the entirety of the experiment and its setup will be described, everything from the data partitioning to the procedure.

Once that is finished, the results will be analyzed in a section of its own, followed by a discussion where the verdicts and validity threats are summarized and ended with the conclusions and suggestions for future works.

2.1 Search Criteria

2.1.1 Databases

In order to find related studies, the following databases were used:

- IEEExplore
- Summon
- Springer
- arXiv
- Researchgate
- DiVA
- Google Scholar

2.1.2 Search terms

Combinations of the following terms were used while searching through the databases:

- "Parking"
- "Cruising"
- "Congestion"
- "Dynamic Time Warp"
- "Clustering"
- "Occupancy"

2.1.3 Selection criteria

Studies were considered qualified and sufficiently related if they contained one or more attributes of each point in the list of inclusion criteria below:

- Written in English or Swedish
- Entails an approach with machine learning or simulation

- Involves data about arrival times, occupancy rates, or cruising times
- Was published in the past five years; i.e. 2016 or later
- The study design was an experiment

A systematic search was primarily employed to find related studies. In cases where there were very few, relevant results in a given database, snowball sampling was used.

Boolean AND-combinations of the phrases were used occasionally to narrow large quantities of results. OR-combinations were used in cases where the used database returned fewer results.

Initially, the criterion for date (2016 and onward) was used until more than 15 related studies were found. Then the most recent studies were used.

2.2 Literature Review

The idea of multi-attribute utility to model preferences dates back to 1969-1976 [7], [8] when psychologist Ward Edwards and economist Howard Raiffa explored the field of decision analysis to develop ways of formulating a person's preferences in a succinct, mathematical way [9], [10]. Their findings would eventually find their way into the larger clade of decision theory, and later into a relatively new branch of machine learning and data mining called "*preference learning*". It would be introduced through workshops as early as 2004 through to 2009, before the notion of a new branch became entrenched into popular discourse, according to Fürnkranz and Hüllermeier in their book "*Preference Learning*" [11]. They, along with Sammut and Webb [12] also state as late as 2010-2011, when their books were published, that preference learning was still a new, emerging branch. Since then, there have been several innovative methods tested to obtain some semblance of personal preferences as generated by AI such as different types of rankings and new, innovative utility functions [13]. Truth tables have been readily used in the wider field of decision theory and artificial intelligence for decades already, it is not foreign to the fields per se [14], [15], [16]. It appears to be a relatively new idea in the context of machine learning, however. Something similar has been implemented for the creation of optimal activation functions for Classification and Regression Trees in a study from 2017 by Tenace and Calimera. But rather than using the truth table to select the best configuration, they made supervised machine learning algorithms attempt to infer the boolean utility function from the inherent logic of the truth table. The model is in turn used to make classifications [17]. Leandro (2016) conducted a similar experiment in 2016 [18], [19], so there is precedent for this approach.

2.3 Related Work

There is no cruising data available to use as indicators of "goodness of fit" otherwise. This is reflected in the related works from which the choice of techniques is drawn, in the context of similar data.

Sayarshad et al. (2020) constructed several simulations of cars cruising for parking space, using factors like occupancy rates, turning them into a "social welfare score", which showed the desirability of different parking lots at different times during the day. It then proposed dynamic pricing schemes based on the outcomes [20].

Fiez et al. (2017) also tried a similar approach, making queuing network simulations, predicting cruising percentages and parking congestion costs relative to parking prices. They found that using probabilities of a block face being full is a more useful metric than using occupancy rates [21]. Some drivers may feel that a particular spot is so important, that it is worth a high price anyways.

Sonntag et al. (2020) acknowledge this problematic nature of occupancy rate representation and opt for testing different measurements to restructure the data with a deep-learning approach. This was done with a focus on different spatial granularity levels. They attempted to find a better way to predict occupancy than using parking meter information, which they also state has not been investigated much before their study. Their results varied wildly from 2-100% accuracy depending on the granularity level [22].

Nilsson et al. (2020) used a deep-learning approach with long short-term memory and gradient boosting regression trees to predict occupancy. They also tested which exogenous features were useful for the predictions, there-among occupancy rates evaluated in different granularity levels. The study also focused on how to measure epistemic uncertainty of the predictions, using the Monte Carlo Dropout method [23].

Zhao (2020) compared the performance of a multilayer perceptron to that of agglomerated k -means clustering when using unlabeled, parking meter data. He concluded that sensor data should be gathered, and from different locations to improve the performance. Among his chosen metrics was occupancy rates [24].

2.3.1 Summary and gaps

The studies suggest strategies (unimplemented in a production environment) to lessen the cruising times and improve the data; composition of social welfare scores, bare data validations, probabilistic uncertainty calculations, different data sources, and restructuring data. Most have been done after pre-processing rather than during, and most of the studies have measured performances against different models, instead of a definitive number. Results are often given such as reduction of walking distances, error measurements, counts of cars, or improvements relative to other models and other metrics. Others have wildly ranging percentages with different techniques and granularities, this makes fair comparisons difficult. So replicating a condition common to all of them (occupancy) can be compared to the improvement later.

3.1 Setup

3.1.1 Why an experiment?

This type of issue concerns the exploration of data and testing hypotheses. Thereby warranting the manipulation, measurement, and transformation of the data points themselves. The setting involves a high level of control where treatments can be applied to subjects systematically. A lot of results can be generated, which helps in making more realistic conclusions [25], [26]. To isolate variables and examine their effects more closely, a higher level of control is needed during testing. At the same time, the in-depth analysis dedicates a significant amount of time and effort to particular aspects of the phenomenon of parking behavior in a given area, therefore the choice was to make a mixture between an experiment and a case study.

3.2 Materials, methods and preprocessing

3.2.1 Data

We begin by examining the data provided by the municipality. An excerpt of two rows of the unprocessed data entries (set vertically, to fit the width of the document) is given in Table 3.1.

Parking_No	1198780	1198789
Car_No	174340	174341
SubscriberMobileCar_No	178665	178666
StartDateTime	2018-01-03 12:15:58.000	2018-01-03 12:16:51.000
EndDateTime	2018-01-03 12:56:58.000	2018-01-03 15:48:51.000
ParkType	CALE	CALE
Customer_No	1	1
Zone_No	29787108	29787108
PaymentType	CALE	CALE
PaymentStatus	ACTIVE	ACTIVE
RenewedFromParking_No	0	0
Active	0	0
CreateDateTime	2018-01-03 12:16:16	2018-01-03 12:17:35
StoppedDateTime	NULL	NULL
OrgType	GROUP	GROUP
Name	BA 8 Röd zon	BA 8 Röd zon

Table 3.1: Example data

The only features relevant are "*StartDateTime*", "*EndDateTime*" and "*Name*". These were chosen for further pre-processing. The total number of rows is 151325 and features data between 2017-12-31 and 2020-02-15. 40 parking machines are included, of which 15 are from the red zone ("*Röd Zon*", the busiest parkings), and 25 from the green zone ("*Grön Zon*", the less busy parkings).

All of them are located on the island of Trossö. The machines do not have the same amount of data, all of the machines in the green zone have approximately one whole month of data, while 10 machines in the red zone have at least one entire year of data.

The data between 2019-01-01 to 2020-01-01 of 10 machines in the red zone are used for large-scale comparisons in all research questions. Additionally, for RQ1 and RQ2 in a smaller inter-machine comparison, data from a red-zone machine between 2018-01-01 to 2019-01-01 are used. For RQ2 and RQ3, one month of data from all 40 available machines are used.

3.2.2 K-means clustering algorithm

The simple function of k -means is to presuppose some centroids in a swath of data points, then to find the nearest centroid for each data point and assign it as belonging to that centroid. Then all points in each cluster are summed up to find their average, to get a new centroid for the cluster. Then the process is repeated. This is done until the method *converges*, i.e. when no more data points "switch sides" when the recalculation is done [27], [4]. Because, as we saw in the Related Works, there is no clear way to go about this problem, and there are many measurements for the results, this study will test how well the data are categorized for the user's convenience. Therefore k -means was chosen. The schedule for this study was not sufficient to test more algorithms, but comparing them was not the point of the study to begin with.

3.2.3 Dynamic time warping

The previously described mechanism can be used for entire arrays of points as well, turning them into a signature, i.e. the centroid array describing its cluster. This is called dynamic time warping [28], [5], and this is what k -means time series clustering is all about, which will be used to differentiate the types of parking congestion days in this study. Dynamic time warping is simply a way to compare a series of points to another series of points, and to then decide if they belong in the same cluster or not. But there are different types of dynamic time warping. In this study, "*soft*" dynamic time warping will be used. The difference between this one and "*hard*" DTW is that soft DTW will tolerate more minor differences between compared time series, counting them as similar [29]. Whereas hard DTW is pickier, so to speak.

3.2.4 Silhouette Analysis

This is done by clustering a dataset and then measuring the dissimilarities in distance of each data point within their respective clusters relative to the other points in the same cluster. The average of these dissimilarities is then calculated as the number $a(i)$. Then the dissimilarity of the data points versus the data points of other clusters

is calculated. When this has been done systematically, the result with the smallest dissimilarities is chosen to be represented as the number $b(i)$ [30]. This formula can be summarized as follows.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3.1)$$

3.2.5 Data analysis, preparation and enhancement

For each of the chosen machines, data are analyzed and pre-processed. Initially, all of the data belonging to the machine is extracted. A time interval is then selected, for which a vacant matrix is created, matching the number of hours (row length) and days (number of rows) of the interval. Then a linear filter [31] is applied to the processed data. It goes step by step through each selected row, and increments by one each of the slots in the matrix corresponding to each hour in the interval, creating a count of cars per hour.

By the same process, another matrix is created, where the number of minutes the respective cars stayed in each hour, is counted for the entire parking lot.

The two entities are represented by B_i and T_i , respectively. Their formulas are given below in Eq. 3.2.

$$B_i = \sum_{k=-\infty}^{+\infty} C_{i,k} X_k \ \& \ T_i = \sum_{k=-\infty}^{+\infty} C_{i,k} X_k \quad (3.2)$$

where $C_{i,k}$ is the filter, i.e. the action of inserting values into each corresponding slot in the matrix based on the current time point i in the time series interval, and X_k is the chosen time series interval. The two matrices containing the two types of entities calculated for each time point of the studied period (see Eq. 3.2) are denoted by B and T , respectively.

Since the data are unlabeled and temporal, clustering and dynamic time warping techniques are chosen for the task at hand. The clusters resulting from it represent low, medium, or high amounts of parking congestion rates. The parking congestion rates are in turn constructed from three factors: *occupancy*, *volatility* and *popularity*.

It should be noted that the formulas for these filters are generic formulas used for most types of both time-variant- and time-invariant filters. Though the one used here is not the same, it is inspired specifically by Jan Grandell's version who derived his from P.J. Brockwell's and R.A. Davis' version [32][33]. The difference is that these filters are time-variant, and the input signal $C_{i,k}$ is time-independent in the same sense as X_k , but the output signal results in a discrete, finite count of cars or minutes for each hour in the chosen interval. Therefore the resulting series is not called Y , nor are the iterations for C points of time t , but indices i . Furthermore the capital C is to denote *count* specifically. All other formulas used in this study are entirely custom and specifically produced for this study.

Occupancy timeline

The first factor, denoted by $\mathcal{O}$, is simply the set of ratios between the number of occupied spaces in a parking lot per hour and the parking capacity (number of spaces

in total) for the whole studied period. It is represented by the formula in Eq. 3.3, where C is the parking capacity and B is the above-introduced set of counts of cars per hour for the whole timeline.

$$\mathcal{O} = \left\{ \frac{i}{C} \mid i \in B \right\} \quad (3.3)$$

Volatility timeline

Volatility is the amount of time cars stay on average in a parking lot. It is denoted by $\mathcal{V}$ and to be constructed, each element i of the timeline T is divided by 60 minutes and the number of cars present in the corresponding hour to get the percentage of minutes out of the full hour that the cars there stay on average.

$$\mathcal{V} = \left\{ \frac{T_i}{(60 \times B_i)} \mid i \in T \right\} \quad (3.4)$$

Popularity timeline

Lastly, the popularity factor, denoted by $\mathcal{P}$, represents the arrival-to-departure ratio in a parking lot relative to the number of free spaces available. To construct it, each element i of the timeline B is deduced from the number of spaces in the parking to find the number of free spaces left in that hour. In addition, the difference between the previous hour's count of cars and the current hour's count of cars is calculated, then deduced from the capacity as well, to get the change in free spaces. The number of free spaces is added to it, and then the total is divided by the capacity to get a percentage of competition, or popularity of the parking lot in each hour.

$$\mathcal{P} = \left\{ \begin{array}{l} \text{If } B_{i-1} < B_i \text{ then } \frac{2(C - B_i) + B_{i-1}}{C} \\ \text{else } 0 \end{array} \mid i \in B \right\} \quad (3.5)$$

Aggregated score matrix

The above defined factors are then combined through the weighted averaging into a *parking congestion rate*, or *score*, i.e. different importance can be assigned to the factors before being integrated together. The parking congestion score at time point i , denoted by $\mathcal{S}_i$, is defined by Eq. 3.6 given below.

$$\mathcal{S}_i = \mathcal{O}_i \times w_1 + \mathcal{V}_i \times w_2 + \mathcal{P}_i \times w_3 \quad (3.6)$$

Where $w_j \in [0, 1]$, $j \in \{1, 2, 3\}$ and $\sum_{j=1}^3 w_j = 1$, are relative weights that define the importance of the factors. In addition, $\mathcal{O}_i$, $\mathcal{V}_i$ and $\mathcal{P}_i$ are the factors' scores at time point i . Note that the congestion score is restrained to constitute a total between 0 and 1 at all times, i.e. $\mathcal{S}_i \in [0, 1]$.

The parking congestion scores calculated for the whole timeline are ordered into a matrix of 24 hours per row, for as many rows as there are days in the studied time interval. For this, time intervals have always been chosen in even, full amounts of days, i.e. the year, month, and day were changed, but the hour, minute, and second

were always set to 00:00:00, to ensure dimensional symmetry. The $m \times 24$ dimension matrix below, denoted by $\mathcal{S}$, presents parking congestion daily profiles for m days. Each profile contains 24 congestion scores, one per day.

$$\mathcal{S} = \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,24} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,24} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m,1} & s_{m,2} & \cdots & s_{m,24} \end{pmatrix} \quad (3.7)$$

3.2.6 Incorporation of driver and expert knowledge as validation sets

Only drivers themselves can judge whether a rate has been true to reality or not. Hence their input is needed [34]. This has been gauged by submitting a form to 10 participants, long-time drivers and GIS experts from the municipality, to give their judgments of what is a tolerable and what is an intolerable situation in the context of deciding whether to enter a parking lot and try to park there (the forms were anonymized, and a non-disclosure agreement was signed for this). The decisions where they agreed have been defined as are, and where they have differed, the majority's opinion was chosen. The form itself is the truth table, containing 125 combinations of the various factors (in stepped percentages of severity, 0, 0.25, 0.50, 0.75, and 1.00), and the margins to the right are left empty to be filled in. The final truth table (condensed) looks the way it is shown in Fig. 3.2.

O	V	P	VERDICT
0	0	0	YES
0	0	0.25	YES
0	0	0.5	YES
0	0	0.75	YES
0	0	1	YES
0	0.25	0	YES
0	0.25	0.25	YES
...	...	...	...
...	...	...	...
...	...	...	...
1	0.75	0.75	NO
1	0.75	1	NO
1	1	0	NO
1	1	0.25	NO
1	1	0.5	NO
1	1	0.75	NO
1	1	1	NO

Table 3.2: A truth table built through incorporating driver and expert knowledge

Using the truth table, the boundaries for the multi-attribute utility function are defined through an estimate resembling a sum-of-products operation. The rows of the table where the first "NO" appears, up until the row where the last "YES" appears, are analyzed. We can then infer the approximate boundaries for the function. This creates the basis for evaluating the different configurations of weights to arrive at the best one by the opinions of the experts. Ordinarily, this operation could be performed by a neural network, training until its weights reach an optimum. However, it requires feedback to the network, but there is no labeled training data for this purpose.

Instead, we can infer, by way of mathematical and logical operations, what the weights *ought* to be, relied on the knowledge we have of the world. This utility function is used to judge 39 configurations of 8760 data points (total number of hours in a year of data) each, validating the results by showing their respective accuracy. The multi-attribute utility function firstly evaluates all of the data and gives its verdicts for whether it would or would not try to park in each and every single hour in the chosen parking lot. These represent labels, and then the different configurations' attempts are compared to the labels, giving us a count of accuracy for each configuration.

3.2.7 Variables

For RQ1, the independent variables are the weights to be varied. The dependent variables are the resulting matrices of parking congestion rates. For RQ2 the independent variables are different months, different machines, and the number of months used in the calibrations. The dependent variables are similarity statistics and dimensions of distributions(skewness, kurtosis, standard deviation, mean, variance, and variation). For RQ3 the independent variables are the inputs of different machines with the correct weight configuration chosen from RQ1, to produce the dependent variables of optimal cluster amounts and cluster silhouette scores from silhouette analysis.

3.2.8 Data limitations

There are problems in the data which make the conclusions applicable to a narrower population. The local data are asymmetrical, i.e. some machines have much more data than others. Some machines are missing from the data, so the studied area is not represented entirely. The capacities of the parking lots are not available, proportion has to be counted from the highest occupancy in the data for each parking lot. The proportion may therefore not be the actual size. The data starts and stops being registered on different days for each machine.

Other existing problems are also mentioned in the related works. For example, there is no data about drivers who entered the parking lot but left when they could not find an open slot. There are no measures for current cruising times. Cars that parked during "free hours" and areas without machines are not represented in the data. It is also unknown if the cars remained in their paid spot until the end of the paid term.

3.2.9 Implementation

The following software tools have been used in the study:

- Atom 1.56.0
- Matplotlib 3.3.4
- Numpy 1.20.1
- Pandas 1.2.2

- Powershell 5.1.18362.1474
- Python 3.8.5
- Scikit-learn 0.24.1
- Scipy 1.6.0
- Tslearn 0.5.0.5
- Xlrd 2.0.1
- Seaborn 0.11.1

Chapter 4

Results and Analysis

4.1 Results addressing RQ1

In Table 4.1 we can see the results of one machine on one year of data, as judged by the utility function. In Table 4.2 the average results of 10 machines on one year of data are presented. We can notice that the singular machine has the configuration $[0.5, 0.1, 0.4]$ as its best configuration with 100% accuracy. In comparison, the averages of 10 machines show that both 100% weight assigned to occupancy rates $([1, 0, 0])$, and 100% weight assigned to popularity rates $([0, 0, 1])$ produce the same 99.9% accuracy. This is a conflicting result.

Index	O	V	P	Accuracy
1	0.0	0.0	1.0	0.999
...	...	...	...	...
5	0.1	0.4	0.5	0.998
6	0.1	0.5	0.4	0.987
7	0.1	0.6	0.3	0.931
8	0.1	0.7	0.2	0.725
9	0.1	0.8	0.1	0.289
10	0.0	1.0	0.0	0.241
11	0.2	0.1	0.7	0.999
...	...	...	...	...
15	0.2	0.5	0.3	0.988
16	0.2	0.6	0.2	0.933
17	0.2	0.7	0.1	0.724
18	0.3	0.1	0.6	0.999
...	...	...	...	...
21	0.3	0.4	0.3	0.998
22	0.3	0.5	0.2	0.988
23	0.3	0.6	0.1	0.936
24	0.4	0.1	0.5	0.999
...	...	...	...	...
28	0.4	0.5	0.1	0.988
29	0.5	0.1	0.4	1.000
30	0.5	0.2	0.3	0.999
...	...	...	...	...
39	1.0	0.0	0.0	0.999

Table 4.1: Results of all combinations

Index	O	V	P	Accuracy
1	0.0	0.0	1.0	0.999
2	0.1	0.1	0.8	0.998
...	...	...	...	...
6	0.1	0.5	0.4	0.995
7	0.1	0.6	0.3	0.981
8	0.1	0.7	0.2	0.925
9	0.1	0.8	0.1	0.383
10	0.0	1.0	0.0	0.333
11	0.2	0.1	0.7	0.998
...	...	...	...	...
16	0.2	0.6	0.2	0.981
17	0.2	0.7	0.1	0.925
...	...	...	...	...
21	0.3	0.4	0.3	0.997
22	0.3	0.5	0.2	0.995
23	0.3	0.6	0.1	0.982
24	0.4	0.1	0.5	0.998
...	...	...	...	...
27	0.4	0.4	0.2	0.997
28	0.4	0.5	0.1	0.995
...	...	...	...	...
31	0.5	0.3	0.2	0.998
32	0.5	0.4	0.1	0.997
...	...	...	...	...
39	1.0	0.0	0.0	0.999

Table 4.2: Averages of 10 machines

The performances are broken into the following groups to conduct further analysis:

- **Highest occupancy weights (HOW)**: Rows with the occupancy column weights above 0.5.
- **Lowest occupancy weights (LOW)**: Rows with the occupancy column weights 0.5 or below.

- **Highest volatility weights (HVW)**: Rows with the volatility column weights above 0.5.
- **Lowest volatility weights (LVW)**: Rows with the volatility column weights 0.5 or below.
- **Highest popularity weights (HPW)**: Rows with the popularity column weights above 0.5.
- **Lowest popularity weights (LPW)**: Rows with the popularity column weights 0.5 or below.

	Average Accuracy Per Group	
	1 Machine	10 Machines
HOW	0.999	0.999
LOW	0.872	0.912
HVW	0.495	0.642
LVW	0.999	0.999
HPW	0.999	0.999
LPW	0.872	0.912

Table 4.3: Accuracy rate for factor intensities

The analyzed, average accuracies of the groups are listed in Table 4.3. One may notice that volatility diminished the accuracy in both tests, as demonstrated by its average group accuracy (HVW). This has motivated the choice to conduct a singular test with the configuration $[0.5, 0, 0.5]$ to study the effect of removing volatility altogether. This results in a 100% accuracy on all 10 machines, outperforming the 99.9%, and explains the discrepancy. Though this seems like a very marginal improvement, it produces quite different results as can be seen in the investigation of RQ3 when it is compared to $[1.0, 0, 0]$. The best configuration is to assign 50% weight to occupancy, and 50% to popularity. This configuration is used in the experiments studying RQ2.

4.2 Results addressing RQ2

Some machines are missing months of data. Finding substitutes for those months in other machines would be useful. We need to be certain that the machine being imputed to has driver tendencies like the one being imputed from. We can measure the similarities of the arrival-departure EDFs (Empirical Distribution Functions) between the machines, months or otherwise, using two non-parametric, two-sample, distribution-free statistical tests: Kolmogorov-Smirnov (K-S) [35], and Epps-Singleton (E-S) [36]. The difference between them is that the E-S test tests the Fourier transform, the Empirical Characteristic Function, of the same distribution, and is considered a more powerful test of the same statistic. Since the same one-to-one comparisons are done to produce the same statistic, and Levy's inversion formula implies [37] that each ECF corresponds directly to an EDF, they are treated as comparable, and the same goodness-of-fit measures in this study.

When two EDFs are compared using one of these tests, a p -value is obtained based on the confidence interval (95% in this study's case) used for the test. This value is what is plotted in the heatmap in Fig. 4.1. The darker the color is, the higher the p -value, the more similar the compared pairs are. In one set of these tests, one month of data all machines have, have been compared to each other. In the other set, all months of one machine have been compared to each other. In the first studied case, there are three pairs of similar machines, all of which are parking lots with low capacities between 7-16 slots. While comparison of months yielded no similarities. Another test conducted was a larger test of 12 months from 10 machines each from the red zone, comparing them all to each other (Fig. 4.1), for a total of 14400 comparisons. We find 146 (86 K-S and 60 E-S) pairs of similar months from the total 120 months compared. This represents 1% of all pairs compared. 54.8% of these are in the months of March-July (Fig. 4.2).

Figure 4.1: Overview of combined K-S and E-S heatmaps.

Figure 4.2: Identities of the 146 substitutable months

The above-reported results mean that there is potential for data to be recycled from machine to machine, this is the answer to RQ2. It should be noted that the

radial heatmap would have contained many more matching pairs if the confidence interval had been lowered from 95% to, for instance, 90%. The reason for this threshold was to ensure that the drivers would get as much certainty as possible.

Another point of interest is the potential of missing data to be simulated based on existing data. To examine this, it is necessary to study the data's distributions, stochasticity, and predictability. Imagine several distributions of the same proportions (shape) but different scales, like the ones plotted in Fig. 4.3. They are so similar that if the smaller is scaled up, or the larger down, they may be turned into each other approximately. Let us suppose for instance that the kurtosis (height) and skewness (width and position on the x-axis) of these graphs are (9,3) for the largest, (6,2) for a middle one, and (3,1) for the smallest. Dividing the kurtosis by the skewness for all three gives us the same ratio, namely 3. If we plot this result on a graph, it will be a straight line. In another case when the distributions have the same kurtosis, but growing skewness, plotting their ratio would result in a multiplicative inverse function. In both cases, we would find a pattern or mathematical function, which provides predictability. Based on this we could scale distributions' kurtosis and skewness numbers of one machine to the size of other machines that we are interested in simulating. This can be used to synthesize missing data in a particular week, month, etc. The reason why we use kurtosis and skewness instead of mean and variance is that the latter two are plain measurements of the scale, but not necessarily the entire shape. Variance does not show how the distribution is slanted for example, but skewness does. Mean does not indicate the thickness of the tails or peakedness of the distribution, but kurtosis does.

Figure 4.3: "Illustration of the skewness and kurtosis", *Application of localization factor for the detection of tin oxidation with AFM*, taken from Attila Bonyár (2015)¹

¹https://www.researchgate.net/publication/298415862_Application_of_localization_factor_for_the_detection_of_tin_oxidation_with_AFM

Figure 4.4: Arrival distributions: Entire city vs. town square only.

In Fig. 4.4 we can see that one machine has high kurtosis and high skewness, but that the whole dataset for the same year has a low kurtosis and lower skewness. The assumption was then made that the data will tend towards the city's shape as the number of samples grows, and plotting the ratio is expected to result in a gradually descending line. Calculating and plotting this kurtosis-skewness ratio for each machine in the red zone (Fig. 4.5), the green zone (Fig. 4.6) and for each month of one machine (Fig. 4.7) reveals that spatial relations are very different. Each location has a rather unique tendency. However, the months of one machine in the red zone (see Table 4.4) are very similar, indicating that a good starting point for the simulation of data would be able to be the temporal dimension rather than the spatial one. It is worth noting that Fisher-Pearson skewness was used, as a Gaussian distribution cannot be assumed [38].

Figure 4.5: Kurtosis-Skewness ratios: Red Zone

Figure 4.6: Kurtosis-Skewness ratios: Green Zone

Figure 4.7: Kurtosis-Skewness ratios: Months

Month	Sample	Skewness	Kurtosis	Std Dev	Variance	Variation	Mean
Jan	746	25.425	657.75	0.194	0.038	1.227	0.158
Feb	1317	19.846	449.893	0.257	0.066	1.149	0.224
Mar	3256	21.396	532.864	0.262	0.069	0.647	0.405
Apr	666	24.01	599.2	0.125	0.016	0.805	0.155
May	864	18.924	440.747	0.194	0.038	1.012	0.192
Jun	1582	11.858	228.175	0.188	0.035	0.755	0.249
Jul	782	21.764	538.051	0.184	0.034	1.087	0.17
Aug	701	0.857	2.867	0.036	0.001	0.23	0.155
Sep	929	24.754	645.049	0.199	0.04	1.093	0.182
Oct	3433	18.107	414.942	0.315	0.099	0.744	0.423
Nov	3832	15.171	331.634	0.161	0.026	0.382	0.421
Dec	3064	-0.058	11.573	0.146	0.021	0.413	0.353

Table 4.4: Measurements of machine *BA 8* for 2018

4.3 Results addressing RQ3

The calculated silhouette scores and their resulting, optimal number of clusters are displayed in Table 4.5.

In order to conduct the analysis, data in Table 4.5 are split into eight groups:

- **Red:** All machines from the red zone.
- **Green:** All machines from the green zone.
- **Small:** All machines with a sample size up to 200.
- **Large:** All machines with a sample size above 200.
- **Few SA clusters (FSA):** All machines with 2 or 3 clusters in column *SA Result*.
- **Many SA clusters (MSA):** All machines with 4 or more clusters in column *SA Result*.
- **Few OSA clusters (FOSA):** All machines with 2 or 3 clusters in column *Occ.SA Result*.
- **Many OSA clusters (MOSA):** All machines with 4 or more clusters in column *Occ.SA Result*.

Machine	Zone	Samples	SA Result	SA Shape	Occ. SA Result	Occ. SA Shape
100	Green	160	2	0.708	2	0.727
101	Green	294	2	0.793	2	0.842
102	Green	485	3	0.628	3	0.572
103	Green	122	8	0.868	2	0.795
...	...	...	...	...	...	...
13	Green	50	2	0.715	2	0.677
14	Red	428	2	0.793	2	0.811
16	Red	304	2	0.737	3	0.740
18	Red	647	2	0.488	2	0.600
...	...	...	...	...	...	...
22	Red	262	2	0.651	2	0.600
23	Red	197	2	0.475	3	0.546
24	Red	365	2	0.788	2	0.506
25	Red	448	6	0.499	2	0.489
...	...	...	...	...	...	...
29	Red	578	2	0.594	2	0.459
30	Green	302	2	0.682	2	0.684
31	Green	70	3	0.546	2	0.603
34	Green	73	2	0.692	2	0.657
...	...	...	...	...	...	...
41	Green	49	4	0.761	3	0.680
42	Green	115	4	0.616	2	0.627
44	Green	51	4	0.536	2	0.491
49	Green	116	2	0.866	2	0.887
...	...	...	...	...	...	...
64	Green	124	2	0.670	2	0.664
66	Green	197	2	0.739	2	0.680
68	Green	206	2	0.730	2	0.714
69	Green	128	2	0.547	2	0.608
8	Red	2251	3	0.565	2	0.564

Table 4.5: Congestion score silhouettes vs. occupancy-rate silhouettes

Since the distribution cannot be assumed, the statistical test used for analysis is non-parametric. Namely, a Mann-Whitney U-Test is used to find improvements

between the groups [39]. We can see that there is a rising rate of improvement as the sample sizes grow. MSA, Large and Red are all larger versions (more samples) of their comparative counterparts, and they all follow the trend of increased, relative improvement compared to their counterparts. The Z-score from the Mann-Whitney U-Test validate them as significant (see Table 4.6). Note that the test is two-tailed and the p-values farther from the middle (0.5) indicate more support for the hypothesis that the data enhancement improved the result more in general.

	Z-score	p-value	Mean-D	Median-D	RI
Green	0.116	0.546	0.002	0.012	1-2%
Red	0.696	0.769	0.031	0.022	2-3%
Small	0.153	0.561	0.002	0.014	0-1%
Large	0.767	0.788	0.025	0.022	2-2.5%
FSA v FOSA	0.236	0.593	0.007	0.003	0.5%
MSA v MOSA	0.601	0.809	0.092	0.081	8-9%

Table 4.6: Compared performance of clustering groups

Note: Mean-D, Median-D and RI stand for Mean Difference, Median Difference and Relative Improvement, respectively.

When instead an entire year's worth of data for 10 red-zone machines is compared (Table 4.7), we can see confirmation that improvement is a general trend even in a large-scale comparison. However, no trend in strong correlation between sample size increase and percentage of improvement could be proven from the sample sizes and results of the months and machines in this measurement. The underlying factor is therefore as of yet inconclusive. Figures 4.8 and 4.9 show the summaries in the form of graphs. The negative values in Table 4.7 represent the advantage of the occupancy-only version over the congestion score-version.

Machine	BA-18	BA-20	BA-22	BA-23	BA-24	BA-26	BA-27	BA-28	BA-29	BA-08	SUM	AVG
Jan	0.7%	0.8%	2.2%	1.1%	-0.1%	2.3%	1.9%	-1.3%	1.5%	-7.9%	1.0%	-0.3%
Feb	-2.7%	-5.7%	1.5%	1.3%	2.6%	-4.3%	1.3%	6.4%	-4.9%	1.3%	-3.1%	0.1%
Mar	8.4%	0.4%	2.6%	3.4%	-0.2%	1.0%	-6.0%	-0.9%	-0.7%	-1.2%	6.9%	0.7%
Apr	1.5%	-0.7%	6.4%	4.2%	0.2%	-0.8%	3.8%	6.5%	-1.3%	-0.9%	18.9%	1.9%
May	0.7%	0.8%	0.5%	3.3%	-0.2%	0.6%	1.4%	0.7%	0.9%	0.4%	9.0%	0.9%
Jun	-3.3%	-1.2%	0.4%	2.2%	-0.3%	4.9%	0.8%	0.2%	3.9%	1.2%	8.8%	0.9%
Jul	-2.9%	0.1%	1.8%	3.9%	-1.0%	0.2%	-0.7%	-1.9%	-0.1%	0.1%	-0.7%	-0.1%
Aug	0.7%	-0.2%	8.8%	-1.7%	-1.4%	0.2%	-1.2%	-1.8%	1.0%	0.3%	4.7%	0.5%
Sep	-1.0%	-1.7%	0.8%	1.2%	1.9%	0.2%	5.0%	4.1%	1.6%	-2.7%	9.4%	0.9%
Oct	0.8%	1.3%	2.6%	2.4%	-0.2%	15.00%	0.5%	1.7%	1.6%	-0.6%	25.0%	2.5%
Nov	7.1%	-1.0%	12.6%	-1.5%	0.3%	0.3%	6.1%	5.8%	20.9%	0.6%	51.1%	5.1%
Dec	1.6%	-3.2%	1.7%	0.4%	-6.7%	0.0%	-0.9%	2.1%	5.6%	-1.6%	-0.9%	-0.1%
SUM	11.7%	-10.3%	41.8%	20.4%	-5.3%	19.5%	12.0%	21.6%	30.0%	-11.1%		
AVG	1.0%	-0.9%	3.5%	1.7%	-0.4%	1.6%	1.0%	1.8%	2.5%	-0.9%		

Table 4.7: Comparison of machine silhouette analysis results

Silhouette analysis of the entire city's data has resulted in a rating of 80.4% when using the score, and 65.5% average for all machines. Using occupancy only, it is 74.2%, and 64.2%, respectively. This is an increase of 6.2% in performance for the

Figure 4.8: Sums and averages of silhouette analysis

Figure 4.9: Sums and averages of silhouette analysis

score, for the entire city, and a 1.3% increase in the machines' average. This further confirms that the improvement persists.

It appears that there is a trend in performance improvement when using the data-centric approach. This supports Andrew Ng’s conclusion.

The answer to RQ1 suggests that it would be a good idea for organizations like the municipality to set sensors by the entrances of parking lots to measure the arrivals and departures of cars because popularity was given almost the same amount of importance as the occupancy rates under the assumption that the expert gave the correct opinion. It also suggests that researchers should pay attention to this phenomenon as well in their future studies, and give half of the weight to popularity, half to occupancy, then use the product as a representation of the parking situation at any given parking lot. It is also worth noting that in none of the tests was volatility useful, so that attribute is formally rejected.

RQ2 showed that this type of data is very specific, it is difficult to transplant data from one parking lot to represent another, and the sufficiently equivalent parking lots were very small. It showed that there was no uniformity nor pattern between most of the parking lots given the nature of the particular data provided. This also emphasizes the importance of consistent measures of all populations involved and the importance of having many measurements. If nothing else, a rule of thumb could be inferred from this part of the experiment that 200 samples or less per month give bad results and that over 2000 samples per month can be considered good. In a month of 30 days, this would correspond to around 67 cars arriving per day to the parking lot. Therefore this is the final verdict to the municipality’s initial request or question; the data are insufficient to create a robust model for the time being, as the plan was to span most of the city, where most is defined as reaching farther out than the town square.

The results of RQ3 suggest that, at least in the context of parking or traffic, the performance of clustering techniques may improve if we recompose the preprocessed data using the experience of an expert. Creating more features from one set of features, and recomposing them into a new measure makes a difference.

Creating a complete model out of the given data was not possible, at least not for the intended scope of representing the entirety of Trossö. Out of 57 requested parking lots to be included as representative of all of Trossö, only 40 had any data, and of those 40 only 10 up to one year worth of data (around 300 samples per month, which

is very low). Out of the remaining 10, only three had consistent measurements for entire years, and all of those three were near one another, namely in the town square. Therefore this study turned the attention to how the data can be improved in various aspects to start progressing towards the possibility of a model. This is even more essential and important than the model itself, as prominent computer scientists such as Andrew Ng urge the computer science community to focus on the improvement of data, stating in a recent article "*When a system isn't performing well, many teams instinctually try to improve the Code. But for many practical applications, it's more effective instead to focus on improving the Data.*" (DeepLearning.ai, Issue 84, March 2021) [6].

One report even suggests that up to 87% of related projects never go into production because of the problems with data they face [40]. A recent survey conducted by Anaconda suggested that data scientists still spend at least 60% of their time cleaning and organizing data [41]. This stresses the importance of data quality, and good preparation.

Overall this score can be used as an indication of parking congestion once enough data are gathered, until then there is no guarantee as to how it will be received by the general populace, whether it will be successful or not. The score can be used both to guide drivers and to advise managers, combined into a larger DSS with features such as GPS or map services. Furthermore, since problems like this one are difficult to quantify, and there are no measures of the "real dependent variable" which we attempt to replicate, the truth table and the quantification of the experience of experts are recommended in further endeavors. Exploring this may lead to the solution of the entire problem, potentially opening up a new way to produce data with less effort.

5.1 Validity Threats

5.1.1 External Threats

As mentioned in the section *Data limitations*, there is a question of whether the results can be generalized to parking populations across the entire city. Some machines are missing in the registry, unpaid parkings are not registered, and the data for the available machines is not complete in all cases.

5.1.2 Internal Threats

A potential internal threat is that the respondent drivers who filled in the truth table may have been influenced by their particular habits and which routes they choose in their everyday drives.

5.1.3 Conclusion Validity

The conclusions will encapsulate only the populations of machines available in the data, and only in periods when abnormal events occur, such as COVID 19.

Chapter 6

Conclusions and Future Work

The study has shown a general increase in performance and silhouette evaluation of the model (answering RQ3) when the data are improved through domain knowledge. Improvements ranged from 0-9%. This is a considerable improvement when we take into account how small the difference is in the configurations' accuracies. It appears that an occupancy-only approach is acceptable, and that popularity-only is just as accurate. But a 50-50 split between them is slightly better, answering RQ1. In addition, this improves performance more between temporal aspects (months) rather than spatial aspects (locations/machines). The study also reveals that temporal aspects are more useful for parking data simulation than spatial aspects. Furthermore, it has been discovered that some months have the potential to replace missing months in the data of other machines, and that entire datasets can only be substitutes between a few small parking lots. This answers RQ2. The conclusions are (as of yet) verifiably applicable to the context of parking exclusively. The data has too many limitations compared to the data of the related works, preventing broader generalizations. For future endeavors, the following can be considered:

6.1 Future Work

Imputation of parking data: Making a simulation to impute more data from the available data. If reliable results are achieved, the study can be revisited, and the persistence of results when there are over 10000 samples can be further studied. RQ2 hints at what could be used for imputation. A mixture of DTW, clustering, and linear regression could advance this exploration further.

Minute-level differentiation: It is also of interest to study the performance when minute-level detail is produced during the pre-processing and used in the evaluation, clustering, and comparison of distributions.

Testing more thresholds: The extension of the truth table to incorporate more detailed experts' opinions may make a new optimum emerge. A neural network could mimic an experts' opinion. This may be useful in the evaluation of parking congestion rates, and identifying and involving more factors. It would be interesting to study which type of neural network mimics the truth table the best. Success with this could help circumvent the scarcity of experts and their time.

Bibliography

- [1] M. S. Scott-Morton and P. G. Keen, “Decision support systems: an organizational perspective,” *Reading: Addison-Wesley*, 1978.
- [2] A. M. Turing, “Computing machinery and intelligence,” in *Parsing the turing test*, pp. 23–65, Springer, 2009.
- [3] G. E. Hinton, T. J. Sejnowski, *et al.*, *Unsupervised learning: foundations of neural computation*. MIT press, 1999.
- [4] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *proc. 5th berkeley symp. math. stat. prob...* 1967, pp. 281–297. [1] h. robbins and s. monro,” *A stochastic approximation method*,” *Ann. Math. Stat.*, vol. 22, pp. 400–407, 1951.
- [5] M. Müller, “Dynamic time warping,” *Information retrieval for music and motion*, pp. 69–84, 2007.
- [6] A. Ng, “Issue 84,” *DeepLearning.AI*, Mar 2021.
- [7] M. Abdellaoui and C. Gonzales, “Multiattribute utility theory,” in *Decision-making Process*, 2010.
- [8] D. L. Olson, “Decision aids for selection problems,” *Journal of the Operational Research Society*, vol. 48, no. 5, pp. 541–542, 1997.
- [9] H. Raiffa, “Preferences for multi-attributed alternatives,” tech. rep., RAND CORP SANTA MONICA CALIF, 1969.
- [10] W. Edwards, “How to use multiattribute utility measurement for social decisionmaking,” *IEEE transactions on systems, man, and cybernetics*, vol. 7, no. 5, pp. 326–340, 1977.
- [11] J. Fürnkranz and E. Hüllermeier, “Preference learning and ranking by pairwise comparison,” in *Preference learning*, pp. 65–82, Springer, 2010.
- [12] C. Sammut and G. I. Webb, *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.
- [13] C. Benzmüller and H. Stuckenschmidt, *KI 2019: Advances in Artificial Intelligence: 42nd German Conference on AI, Kassel, Germany, September 23–26, 2019, Proceedings*, vol. 11793. Springer Nature, 2019.
- [14] C. C. Aggarwal, “Propositional logic,” in *Artificial Intelligence*, pp. 105–135, Springer, 2021.
- [15] B. Coppin, *Artificial intelligence illuminated*. Jones & Bartlett Learning, 2004.
- [16] M. T. Jones, *Artificial Intelligence: A Systems Approach: A Systems Approach*. Jones & Bartlett Learning, 2008.

- [17] V. Tenace and A. Calimera, “Activation-kernel extraction through machine learning,” in *2017 New Generation of CAS (NGCAS)*, pp. 5–8, IEEE, 2017.
- [18] C. Leandro, “Reverse engineering and symbolic knowledge extraction on Łukasiewicz fuzzy logics using linear neural networks,” *arXiv preprint arXiv:1604.02774*, 2016.
- [19] C. Leandro, “Symbolic knowledge extraction using Łukasiewicz logics,” *arXiv preprint arXiv:1604.03099*, 2016.
- [20] H. R. Sayarshad, S. Sattar, and H. O. Gao, “A scalable non-myopic atomic game for a smart parking mechanism,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 140, p. 101974, 2020.
- [21] T. Fiez and L. Ratliff, “Data-driven spatio-temporal analysis of curbside parking demand: A case-study in seattle,” *arXiv preprint arXiv:1712.01263*, 2017.
- [22] J. Sonntag and L. Schmidt-Thieme, “A machine learning approach to infer on-street parking occupancy based on parking meter transactions,” in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, IEEE, 2020.
- [23] M. Nilsson and S. von Corswant, “How certain are you of getting a parking space?: A deep learning approach to parking availability prediction,” 2020.
- [24] Y. Zhao, *Parking Availability Prediction based on Machine Learning Approaches: A Case Study in the Short North Area*. PhD thesis, The Ohio State University, 2020.
- [25] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.
- [26] P. Runeson and M. Höst, “Guidelines for conducting and reporting case study research in software engineering,” *Empirical software engineering*, vol. 14, no. 2, pp. 131–164, 2009.
- [27] R. Xu and D. Wunsch, *Clustering*, vol. 10. John Wiley & Sons, 2008.
- [28] M. Vichi, P. Monari, S. Mignani, and A. Montanari, *New developments in classification and data analysis: proceedings of the meeting of the Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society, University of Bologna, September 22-24, 2003*. Springer Science & Business Media, 2006.
- [29] M. Cuturi and M. Blondel, “Soft-dtw: a differentiable loss function for time-series,” in *International Conference on Machine Learning*, pp. 894–903, PMLR, 2017.
- [30] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.
- [31] L. H. Koopmans, *The spectral analysis of time series*. Elsevier, 1995.
- [32] J. Grandell, “Formulas and survey time series analysis,” KTH Royal Institute of Technology, 2005.
- [33] P. J. Brockwell, P. J. Brockwell, R. A. Davis, and R. A. Davis, *Introduction to*

- time series and forecasting*. Springer, 2016.
- [34] Y. Bi, S. Kapoor, and R. Bhatia, *Intelligent systems and applications*. Springer, 2016.
 - [35] F. J. Massey Jr, “The kolmogorov-smirnov test for goodness of fit,” *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
 - [36] S. J. Goerg and J. Kaiser, “Nonparametric testing of distributions—the epps–singleton two-sample test using the empirical characteristic function,” *The Stata Journal*, vol. 9, no. 3, pp. 454–465, 2009.
 - [37] H. Fischer, *A history of the central limit theorem: from classical to modern probability theory*. Springer Science & Business Media, 2010.
 - [38] A. R. Jones, *Probability, statistics and other frightening stuff*, vol. 2. Routledge, 2018.
 - [39] T. W. MacFarland and J. M. Yates, “Mann–whitney u test,” in *Introduction to nonparametric statistics for the biological sciences using R*, pp. 103–132, Springer, 2016.
 - [40] V. Staff, “Why do 87% of data science projects never make it into production?,” *VentureBeat*, Jul 2019.
 - [41] “Anaconda state of data science report 2020,” *Anaconda*, Jul 2020.

Appendix A

Supplemental Information

A.1 Truth table

O	V	P	VERDICT
0	0	0	YES
0	0	0.25	YES
0	0	0.5	YES
0	0	0.75	YES
0	0	1	YES
0	0.25	0	YES
0	0.25	0.25	YES
0	0.25	0.5	YES
0	0.25	0.75	YES
0	0.25	1	YES
0	0.5	0	YES
0	0.5	0.25	YES
0	0.5	0.5	YES
0	0.5	0.75	YES
0	0.5	1	YES
0	0.75	0	YES
0	0.75	0.25	YES
0	0.75	0.5	YES
0	0.75	0.75	YES
0	0.75	1	YES
0	1	0	YES
0	1	0.25	YES
0	1	0.5	YES
0	1	0.75	YES
0	1	1	YES
0.25	0	0	YES
0.25	0	0.25	YES
0.25	0	0.5	YES
0.25	0	0.75	YES
0.25	0	1	YES
0.25	0.25	0	YES
0.25	0.25	0.25	YES
0.25	0.25	0.5	YES
0.25	0.25	0.75	YES
0.25	0.25	1	YES
0.25	0.5	0	YES
0.25	0.5	0.25	YES
0.25	0.5	0.5	YES
0.25	0.5	0.75	YES
0.25	0.5	1	YES
0.25	0.75	0	YES
0.25	0.75	0.25	YES
0.25	0.75	0.5	YES
0.25	0.75	0.75	YES
0.25	0.75	1	YES
0.25	1	0	YES
0.25	1	0.25	YES
0.25	1	0.5	YES
0.25	1	0.75	YES
0.25	1	1	YES
0.5	0	0	YES
0.5	0	0.25	YES
0.5	0	0.5	YES
0.5	0	0.75	YES
0.5	0	1	YES
0.5	0.25	0	YES
0.5	0.25	0.25	YES
0.5	0.25	0.5	YES
0.5	0.25	0.75	YES
0.5	0.25	1	YES
0.5	0.5	0	YES
0.5	0.5	0.25	YES
0.5	0.5	0.5	YES
0.5	0.5	0.75	YES
0.5	0.5	1	YES
0.5	0.75	0	YES
0.5	0.75	0.25	YES
0.5	0.75	0.5	YES
0.5	0.75	0.75	YES
0.5	0.75	1	YES
0.5	1	0	YES
0.5	1	0.25	YES
0.5	1	0.5	YES
0.5	1	0.75	YES
0.5	1	1	YES
0.5	0.5	0.5	YES

O	V	P	VERDICT
0.5	0.5	0.75	YES
0.5	0.5	1	YES
0.5	0.75	0	YES
0.5	0.75	0.25	YES
0.5	0.75	0.5	YES
0.5	0.75	0.75	YES
0.5	0.75	1	YES
0.5	1	0	YES
0.5	1	0.25	YES
0.5	1	0.5	YES
0.5	1	0.75	YES
0.5	1	1	YES
0.5	1	1	NO
0.75	0	0	YES
0.75	0	0.25	YES
0.75	0	0.5	YES
0.75	0	0.75	YES
0.75	0	1	YES
0.75	0.25	0	YES
0.75	0.25	0.25	YES
0.75	0.25	0.5	YES
0.75	0.25	0.75	YES
0.75	0.25	1	YES
0.75	0.5	0	YES
0.75	0.5	0.25	YES
0.75	0.5	0.5	NO
0.75	0.5	0.75	NO
0.75	0.5	1	NO
0.75	0.75	0	YES
0.75	0.75	0.25	YES
0.75	0.75	0.5	NO
0.75	0.75	0.75	NO
0.75	0.75	1	NO
0.75	1	0	YES
0.75	1	0.25	YES
0.75	1	0.5	NO
0.75	1	0.75	NO
0.75	1	1	NO
1	0	0	NO
1	0	0.25	NO
1	0	0.5	NO
1	0	0.75	NO
1	0	1	NO
1	0.25	0	NO
1	0.25	0.25	NO
1	0.25	0.5	NO
1	0.25	0.75	NO
1	0.25	1	NO
1	0.5	0	NO
1	0.5	0.25	NO
1	0.5	0.5	NO
1	0.5	0.75	NO
1	0.5	1	NO
1	0.75	0	NO
1	0.75	0.25	NO
1	0.75	0.5	NO
1	0.75	0.75	NO
1	0.75	1	NO
1	1	0	NO
1	1	0.25	NO
1	1	0.5	NO
1	1	0.75	NO
1	1	1	NO
-	-	-	-