Bachelorarbeit

Comparison of supervised learning models in arrival time estimation

an der Technischen Universität
Carolo-Wilhelmina zu Braunschweig
Carl-Friedrich-Gauß-Fakultät
Institut für Wirtschaftsinformatik, Abteilung Decision Support.

Eingereicht von:	Emre Gezer
Matrikelnummer:	4901507
Studiengang:	Wirtschaftsinformatik
Referent:	Prof. Dr. Dirk Christian Mattfeld
Betreuer:	M. Sc. Florentin Hildebrandt
Eingereicht:	Braunschweig, den 10.02.2021

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Braunschweig, 10.02.2021

Contents

List of Tables vi

List of Figures vii

List of Abbreviations viii

1 Introduction 1

2 Literature Review 3
- 2.1 Most related work . . . 3
- 2.2 Arrival Time Estimation . . . 3
- 2.3 Supervised learning . . . 4

3 Model 5

4 Solution Approach 6
- 4.1 Motivation . . . 6
- 4.2 Algorithms . . . 6
  - 4.2.1 GBDT . . . 7
  - 4.2.2 Random Forest . . . 7
  - 4.2.3 Neural Network . . . 7
- 4.3 Feature Engineering . . . 7
- 4.4 Evaluation . . . 7

5 Computational Study 8
- 5.1 Experimental Design . . . 8
- 5.2 Parametrization of Methods . . . 8

6 Analysis 9

7 Conclusion 10

References 11

List of Tables

List of Figures

List of Abbreviations

ML	Machine Learning
ETA	Estimation of arrival time

1 Introduction

Nobody likes waiting - be it when you order your next book on Amazon for class or the pizzas at UberEATS for your birthday. In fact, when doing research on this topic, psychologists found out that increased waiting times generally have a significant negative impact on customer satisfaction and loyalty [1,2]. If we take a glance at the online food delivery market, there are more than 700 million people globally that used food delivery services in 2017 with twice as many users being expected in 2024 [3]. According to forecasts for the same time frame, the userbase of eCommerce platforms in general will grow from 2.480 billion to roughly 4.6 billion users [4]. Thus, waiting becomes a large scaled economic issue. What one might not expect is that the customers' own perceptions regarding their waiting time negatively affect the perceived service quality stronger than actual waiting times do [5,6]. Given these points, we can conclude that a key challenge lies in the communication of accurate arrival time estimations to the customers, rather than solely shipping the product fast.

Current research tackles this task commonly by means of machine learning techniques. Exemplary, while Hildebrandt and Ulmer used Gradient Boosting Decision Trees in their offline approach to map state features directly to expected arrival times in food delivery [7], Zhu et al. implemented a Multilayer Perceptron to predict accurate Order Fulfillment Cycle Times [8], whereas Ulmer and Thomas used approximation to estimate mean arrival times for values of aggregated states in service routing [9]. Other than Khiari et al. who compared several boosting and linear regression algorithms [10] (without incorporating neural networks in their comparison and spatial features in their underlying model which is what we will do), hardly any kind of machine learning model comparison for arrival time estimation purposes in vehicle routing settings can be found. Therefore, we intend to conduct further research in this area.

This paper examines the forecast quality and performance of different supervised machine learning (ML) techniques in the context of arrival time estimation problems (ETA). To accomplish this task, the algorithms will be trained on historical meal delivery data collected within Iowa City.

This paper is organized as follows. Firstly, we will review and discuss literature on machine learning in general and its usage in the field of arrival time estimation. We will then proceed with our model, where we describe the ETA problem generally and define the model with respect to the underlying problem of our experimental design. Several steps are taken in the solution approach that follows upon our model definition. In the beginning, we present the used ML techniques and define them generally for the sake of understanding the way each of them functions. Concretely, our dataset will undergo a transformation process with the goal in mind of having a reduced feature space with the most impactful features left in the end. We will then advance to the computational study, where we present its experimental design and discuss its results. Finally, we will create variations of our experimental design in order to analyze each algorithm's robustness and impact in these different settings.

2 Literature Review

In this section, we present related literature. Since we're estimating arrival times by means of supervised learning algorithms, our problem is associated to following research areas: arrival time estimation and machine learning in DVR settings.

2.1 Most related work

This work is mostly inspired by Hildebrandt and Ulmer (2020), who contributed a supervised offline and another supervised online-offline approach to predict arrival times in the Restaurant Meal Delivery problem setting, a dynamic pick up and delivery problem with uncertainty in travel times, processing times and order bundling presented in **?**. In their offline approach, Hildebrandt and Ulmer (2020) map state features to expected arrival times by means of a gradient boosting decision tree (GBDT) model. In their online-offline approach, they calculate the running average in an online simulation by training a supervised deep neural network that approximates the decisions a given routing policy would make for a given state. Because we operate in the identical problem setting and share a common goal in estimating arrival times fast and accurately via supervised learning, this paper will incorporate the GBDT model presented in Hildebrandt and Ulmer (2020) in its empirical comparison.

2.2 Arrival Time Estimation

Li, Tian, and Leung (2010)

Zhang, Lam, and Chen (2013) formulate a SVRPTW with stochastic travel and service times that seeks to minimize vehicle employing costs, travel times and deviances w.r.t expected arrival times. They predict expected arrival times by calculating the sum of expected service and travel times which are then used to calculate the minimum on-time probability which every vehicle must fulfill for every visited customer.

Ulmer and Thomas (2018) estimate mean arrival times for a vehicle routing problem with stochastic requests (VRPSR) using value function approximation. The RMDP is more complex wrt. to its fleet size since the RMDP is in contrast a multi vehicle problem and its logistical context since the RMDP is a pick-up and delivery problem whereas the VRPSR only incorporates the delivery component.

Chun-Hsin Wu, Jan-Ming Ho, and Lee (2004) apply support vector regression (SVR) for travel time prediction for single origin-destination routes in the context of intelligent transportation systems. Their experimental procedure contains of selecting a relatively less biased subset of the provided traffic data, and comparing the SVR to two analytical baseline methods for travel time prediction as benchmarks. They rely solely on temporal features.

Chen, Liu, Xia, and Chien (2004) who proposed an artifical neural network.

Wang, Fu, and Ye (2018) propose an offline trained Wide Deep Recurrent Neural Network (WDRNN) model to predict vehicle travel times for single origin-destination routes by aggregating spatial, temporal, traffic, personalized and augmented features. In their offline comparison, they compared their approach to competing classical machine learning methods and route-based ETA. Two indications of their results are interesting in our viewpoint: First, all machine learning methods included in their experiment outperformed route-based ETA. Secondly their deep learning approach outperformed the competing classical machine learning methods. They also employ the WDRNN online for real-time service.

+++ Wird wahrscheinlich nicht genommen +++: Salari, Liu, and Shen (2020) propose a regression tree based and a quantile regression forests model to forecast delivery time distributions in a B2C E-Commerce context. While they use temporal (i.e. „order hour“), processing (i.e. „Number of waiting orders“), and other order-related features (i.e. „Total quantity“), they do not incorporate spatial features and do not define their underlying problem model explicitly. In contrast to other arrival time prediction approaches, Salari et al. (2020) attempt to predict arrival time distributions rather than continuous values.

2.3 Supervised learning

3 Model

Vehicle veV Node peP, deD actions a (consist of type, time t, len(action)) Routes 0 –> consists of actions

State (time, planned route, updated route)

4 Solution Approach

This chapter presents the solution approach. Section 4.1 motivates our problem. Section 4.2 discusses the selected features. Section 4.3 explains the algorithms considered in the comparison in detail. Section 4.4 defines the process by which we evaluate the algorithms in order to assess the quality of each algorithm.

4.1 Motivation

We motivate this paper by answering the following questions: Why is machine learning useful and widely-common for esitmation tasks? Which technical problems are encountered when using other approaches? Which machine learning algorithms are appropriate for our problem? Why bother comparing them?

4.2 Algorithms

This section gives a short introduction to the conceptual framework of supervised learning and then examines the algorithms included in our comparison. For further research, the reader is refered to Bishop (2006) and Wolf (2020).

Supervised learning models receive a finite sequence $S = \{(x, y)|x \in X, y \in Y\}$ of length $N \in \mathbb{N}$ as inputs where $x_{i \in N} \in X$ is an **observation** and $y_i \in Y$ its corresponding **target**. They aim for the approximation of an **hypothesis** $h : X \to Y$ that describes the underlying patterns of X accurately w.r.t. a **loss function** $L : Y \times Y \to \mathbb{R}$ and predicts a target of a given observation. Loss functions represent the deviance of $h(x_i)$ for a given x_i and its correspondent target y_i by which we can measure how accurate the predictions of a model are. An optimal approximation function h_{opt} is obtained by finding the minimal loss. h is then used to predict any $y_i \in Y$ for a corresponding $x_i \in X$.
Predictions can generally be made for either a **regression** or **classification** problem, which differ in the nature of their target variables. While regression is used to predict continuous targets Y, classification finds its use in the prediction of discrete Y. In machine learning terms, the arrival time estimation problems classifies as a regression task since we are estimating continuous time points. Although various supervised learning models use different mathematical procedures to predict targets, all of them follow the principle of induction, meaning that general rules are inductively inferred from the observations. This is also called **generalization**. In order for a model to generalize the data properly, two things have to be avoided: **Underfitting**

and **Overfitting**.

4.2.1 GBDT

4.2.2 Random Forest

- Example - Formal Definition

4.2.3 Neural Network

- Example - Formal Definition

4.3 Feature Engineering

- Raw Data - Manual feature selection - Feature learning via deep autoencoder

4.4 Evaluation

- Use MSE (and maybe more criteria) etc. to derive infos about accuracy - Hyperparameter sensitivity analysis (Random Search) for robustness - Variations in data set for robustness –> Noise einführen –> Anzahl der Trainingsdaten –> Andere Features - Time for performance

5 Computational Study

5.1 Experimental Design

5.2 Parametrization of Methods

- Different subsets of samples w.r.t size

6 Analysis

- Variations in experimental design
- Introduce noise to data - Change model assumptions (Changes in fleet, customer, restaurant variables, general assumptions (?)) - Hyperparameter sensitivity - Discuss Interpretability

7 Conclusion

References

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics).* Berlin, Heidelberg: Springer-Verlag.
6

Chen, M., Liu, X., Xia, J., & Chien, S. I. (2004). A dynamic bus-arrival time prediction model based on apc data. *Computer-Aided Civil and Infrastructure Engineering*, *19*(5), 364-376. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8667.2004.00363.x` doi: https://doi.org/10.1111/j.1467-8667.2004.00363.x
4

Chun-Hsin Wu, Jan-Ming Ho, & Lee, D. T. (2004). Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems*, *5*(4), 276-281. doi: 10.1109/TITS.2004.837813
3

Hildebrandt, F., & Ulmer, M. (2020, 08). *Supervised learning for arrival time estimations in restaurant meal delivery.*
3

Li, X., Tian, P., & Leung, S. C. (2010). Vehicle routing problems with time windows and stochastic travel and service times: Models and algorithm. *International Journal of Production Economics*, *125*(1), 137 - 145. Retrieved from `http://www.sciencedirect.com/science/article/pii/S092552731000023X` doi: https://doi.org/10.1016/j.ijpe.2010.01.013
3

Salari, N., Liu, S., & Shen, Z.-J. M. (2020, 09 28). Real-time delivery time forecasting and promising in online retailing: When will your package arrive? Retrieved from `http://dx.doi.org/10.2139/ssrn.3701337`
4

Ulmer, M., & Thomas, B. (2018, 03). Enough waiting for the cable guy - estimating arrival times for service vehicle routing. *Transportation Science.*
3

Wang, Z., Fu, K., & Ye, J. (2018). Learning to estimate the travel time. New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3219819.3219900`
4

Wolf, M. M. (2020). *Mathematical foundations of supervised learning.* Retrieved from `https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA4801_2020S/ML_notes_main.pdf`
6

Zhang, J., Lam, W., & Chen, B. Y. (2013, 05). A stochastic vehicle routing problem with travel time uncertainty: Trade-off between cost and customer service. *Networks and Spatial Economics*, *In press*. doi: 10.1007/s11067-013-9190-x 3