Machine Learning Method Within the Context of a Socially Aware Solution for Vehicle Routing Problems

Robert Maurer
Department of Business and Management
University of Applied Sciences Brandenburg
Brandenburg, Germany
E-Mail: maurer@th-brandenburg.de

Andreas Johannsen
Department of Business and Management
University of Applied Sciences Brandenburg
Brandenburg, Germany
E-Mail: johannse@th-brandenburg.de

***Abstract*—The market for courier, express and parcel services has seen an immense increase in sales and relevance in times of the pandemic. Not only has the volume of shipments increased, but also the demand for social Vehicle Routing Problems (VRP) solution procedures based on modern IT solutions supporting the dispatching or routing process. This article provides an answer to social responsible and sustainable logistics services and a conceptual prototype for the practical implementation of a Machine Learning method to solve Vehicle Routing Problems (VRP) in the context of sustainable "last mile" logistics. Aspects of combinatorial optimization algorithms in the form of an ant algorithm were used to support the applied Machine Learning (ML) system. The prototype is based on the "Reinforcement Learning" system and uses "REINFORCE with baseline" as the algorithm for updating a parameterized policy. A benchmark analysis provides a comparison between the prototype and Google-OR, as a representative for combinatorial optimization algorithms, applied in two examples. The results show that Google-OR prevails over the prototype in terms of solution quality, but the prototype convinces in runtime and automatism. In addition, the applied Machine Learning context results only in minor advantages for small to medium sized logistic domains, as they do not generate enough data. Hence, using Machine Learning methods for Vehicle Routing problems is recommended for a larger stop volume in urban areas. Furthermore, the prototype represents an alternative solution to outsourcing to third party providers and provides an approach to gain a competitive advantage for solving Vehicle Routing Problems.**

Keywords-Vehicle Routing Problem; Machine Learning; Reinforcement Learning; last mile logistics

I. INTRODUCTION

A. Motivation and goals

In 2020, the amount of shipments in Germany comprised 4,05 billion package, express and courier shipments. The amount of shipments tends to increase within the upcoming years whereby the majority of shipments is serving the B2B sector [1]. In addition to the increased amount of package shipments and the resulting routing problems, the scientific interest in Vehicle Routing Problems (VRP) based on modern IT solutions has also increased. A search led by the keyword "Vehicle Routing" in the IEEE Explore database resulted in more than 6500 hits within the last ten years. Nevertheless, there is just a limited amount of scientific work in the field of Machine Learning (ML)-methods to solve VRP. A search through Google Scholar resulted in 15 publications within the last five years that used the terms "Vehicle Routing" and Machine Learning" in their titles. In addition, the current research findings are predominantly in a theoretical environment with a focus on mathematical models and assumptions, such as [2] and [3], but without proper practical relevance. Furthermore, social factors such as comfort of driving (e.g., through weather conditions), route preferences, road conditions or interactions between the drivers are barely considered. Additionally, there is only limited literature considering comprehensive solutions for sustainability related challenges [4]. The main motivation of this work is to develop a practical concept, in form of a prototype, to solve VRP. Furthermore, the paper shows that ML-methods - in a productive working environment - can be a long-term alternative to outsourcing to third party suppliers and can be a potential tool to increase employee loyalty. This is shown by a comparison in terms of effectiveness under selected evaluation criteria between ML-methods and classical optimizing algorithms to solve VRP.

B. Thematical Introduction

In the subject area of the zero-emissions „Last-Mile"-metropolitan logistics, the VRP describes a combinatorial optimization problem that addresses the following basic question: "What is the optimal set of routes for a specific vehicle fleet to deliver goods to a specific amount of customers?" [5]. VRP was first discussed by [6] in their scientific work „The Truck Dispatching Problem" where the problem context was the delivery of fuel, which was solved by using algorithmic ways. To solve VRP, ML- methods are especially suitable because their decision making process is based on algorithms and experience. The experience arises from specific subscription structures occurring in certain subject domains and the resulting known VRP instances. In this work, the specific use case of the VRP is within the context of the "Last- Mile" metropolitan logistics. This is a modern form of the urban logistics, which contains the last step of a package delivery. Especially a zero-emissions approach was pursued. That is realized through the usage of cargo bikes and micro-depots, which justifies the term "Green VRP" [7]. The aspect of „Last Mile" metropolitan logistics is relevant for solving combinatorial optimization

 ISBN: 978-1-61208-981-2

problems in the way that it can be assumed that there is a small distance between the routes and the delivery tours are limited in capacity through the usage of cargo bikes. Looking at rural areas, these factors are becoming more important because the distances between the stops are longer and the amount of stops is less, compared to urban areas. Therefore, the planning of efficient routes, as well as the driving conditions, are not only a decisive factor for the high logistic costs of zero emission delivery, but also necessary in order to keep up with the alternative of Diesel- or Petrol- based transporters.

C. Challenges and Competitive Position

There are challenges in the later implementation of the prototype regarding the competition between classical optimization algorithms and the used ML-method to solve VRP. Especially concerning performance, the ML- method of the prototype is going to be very demanding in the beginning, because it includes more process steps such as training the model. The prototype is competitive, if it generates short routes based on the length of the tour, provides the results within a few seconds, generally reacts to unknown VRP instances, performs equally well on VRP instances and does not need a manual intervention. A further competitive advantage could be gained by the prototype though the consideration of social framework conditions.

II. METHODOLOGY

A. Selection of a Machine Learning Method

In the context of this article, the ML-model uses a Reinforcement Learning (RL)-system to solve VRP of the "Last Mile" metropolitan logistics. In this paragraph, we explain why we chose reinforcement learning in our work. The RL differs significantly from the alternative supervised- and unsupervised learnings because it uses a different approach for the construction of a learning system. The learning model describes an agent or a decision maker in the RL that observes an environment, executes actions on it and independently learns the dynamics from an unknown starting point [8]. After each action, the agent can receive two different kinds of rewards: an immediate or a delayed reward. An immediate reward is applied for actions of the agent that allow an immediate assessment, which are also used in the developed prototype. For instance, the crossing of a red traffic light can be assessed immediately because a negative behavior is directly identifiable and does not only become identifiable through subsequent behavior. A delayed response can be thought of as an action in the game of chess where the reward is measured to the follow up reaction, which was also used in the scientific work of [9]. For this article, RL was chosen in order to find solutions to complex optimization problems without prior human knowledge to reduce development and dispositioning effort.

B. Training and Result Data

Through the cooperation with a Germany based company, a solid foundation of anonymized customer were used for training and testing the model. Due to the missing availability of data, certain framework conditions such as route preferences or impact of the weather could not be directly considered in the learning process. The data objects have the following properties: longitude, latitude, stop weight, stop volume and an anonymized identifier. The data objects are specific to a certain date and a micro depot.

C. Prototyping

For the foundation of the research and based on the practical motivation, a qualitative-constructivist approach in the form of the prototyping was chosen. The final version of the prototype is supposed to be a service that trains, stores and applies ML-models. The ML-model is sensitive towards the sender, the micro depot, the supplier, the weekday, the vehicle weight, the vehicle volume and the ML-method. The ML-model is instance-based saved. The ML-model should only be trained by stops from one day to one week in order to avoid that subscribed customers change the strategy of the agent through repetitive appearance.

D. Comparative Analysis

In the context of this article, a comparative analysis between the developed prototype and the VRP-solver of Google-OR was performed. According to [10] the VRP-solver is based on heuristic algorithms that are categorized as „First Solution Strategy" and are optionally extendable through „Local-Search"- strategies. Since the competitor's solution is not considering any social framework conditions regarding the driver, the comparison is mainly focused on the technical factors. The results of this comparison are supposed to evaluate the prototype, reveal opportunities for improvement and consider the potential for the implementation of the RL-method for solving the VRP. The analysis is structured in two parts. In the first part of the comparative analysis, the main focus lies on an estimation of the effectiveness of the developed prototype including the predefined basic functionalities of the chosen RL-algorithm "REINFORCE with baseline" to solve VRP. The estimation of the effectiveness is based on the focus of the trainings and testing time, as well as further comparison criteria such as distance and time effort. The Haversine formula proves the distance effort. The time effort is calculated through the deposited vehicle speed, the distance and a predefined stop dwell time of five minutes. For a larger experiment, further comparison criteria could be profitability, service quality, consistence, external (especially social factors) and further [11]. It is looked at a stop volume of 430 stops spread over four weeks in May 2021, whereby the ML-model is trained with 200 iterations after each day. Since the training is progressive, each training of each week builds on the experience of the previous one. The second part of the analysis focused on the scope for improvement of the ML-method and the developed adjustment impulses, which are derived from the previous first part. The goal is it to converge as close as possible to the calculated distance and time effort compared to the competitors in order to convince with a better runtime. In the third part, the same problem context is looked at for a more significant comparison. The training was analogous to the first part with adjusted 115

iterations per day. The amount of iterations was reduced to minimize the possibility of "overfitting" of the ML model compared to the training data.

III. RESEARCH RESULTS

A. Modelling and Conception of the Machine Learning Model

The Machine Learning model is fundamentally based on the Markov Decision Process (MDP) in conjunction with an Ant Colony Optimization (ACO)-algorithm, which is a combinatorial optimization method, whose foundation was laid by [12]. The ACO-procedure was used in the training mode for the prototype within the first instance of the process to provide the RL-method in the following step of the process with a premonition of the transition probability distribution in the subsequent process step and to tune already trained ML-models for the considered problem context. Comparatively, a similar sub-procedure was performed by [13]. They considered an ACO-model supported by ML as the starting point. The ML-model is defined by MDP. In Markov-models, the subsequent states and the RL-reward only depend on the current state and the chosen action of the agent [17]. The adapted MDP in the developed prototype was modeled in accordance to [15]. As an ML-algorithm the prototype uses the "REINFORCE"-algorithm or Monte-Carlo Policy-Gradient, in order to find an optimal policy π^*. For the development of the prototype and the application of the "REINFORCE"-algorithm, the book "Reinforcement Learning- An Introduction" from [16] was used for guidance. The Monte-Carlo-methods are characterized by the fact that there is no holistic knowledge about the environment needed to find an optimal policy. This is because systems are learning from the interaction with the environment [16]. A modified form the "REINFORCE" with baseline"-algorithm was constructed according to [16]. A baseline can be a random variable or, as used in the context of this prototype, a „state-value"-function, which reflects valuations of possible rewards of the entire condition space [16].

B. Development of the Prototype

The backend of the prototype was developed in the programming language python and the web-frontend, which was necessary for better debugging, in react. Consciously, no existing ML-frameworks, such as Keras, Tensorflow or PyTorch, were used. This decision was made because, in case complex ML-frameworks were used, there would constantly be a risk of becoming dependent on the respective framework support. Furthermore, the usage of self-developed complex ML-methods lead to better gains of understanding and experience, than the usage of ML-frameworks. However, there is a stronger tendency to learn the framework rather than the ML-procedure based on it. In addition, special challenges occurred during the development of the prototype. Above all the thematic of the local minimum, in which the prototype often was stuck in the early stages of the development. The progression of cumulated rewards, which are understood as total length in km of all formed tours within one episode were monitored over 2000 episodes. In this case, the agent considers a problem context of 19 stops, which are distributed in the city, and detects an adequate solution towards the end, but not the best one even though it already detected it in the first 250 iterations. This was caused by the learning factor being set too low and by the fact, that the exploration factor did not influence the agent enough for an extended exploration of the environment. For a better exploration, a dynamic epsilon exploration was used [16]. In the use case of the prototype, the fundamental goal was finding the global minimum in a specific problem context in order to reach the shortest possible total distance. Furthermore, it proved to be difficult to implement the "Baseline" update exactly according to [16]. Partial updates of specific parameters were too "heavy" and caused a noise in policy rewards. In order to solve this problem, two areas in the concept were adjusted. On the one hand, the advantage calculation was replaced by a „Simple every-visit Monte Carlo"[16]. On the other hand, a direct update of the parameterized policy was prohibited and instead handled through an adjustable increase and decrease factor. Moreover, after the first part of the comparative analysis, certain adaptation impulses regarding "local-search" and "bin-packing" strategies were implemented in the prototype. In the case of "local-search" strategies, the agent's action-selection execution is matched to ensure that the made decision does not exceed an adjustable threshold against the lowest possible reward. A simple „First Fit Decreasing"-approach, realized the "bin-packing"-strategy, which enables an optimal distribution of the capacity requirements to the maximum capacity of the vehicle.

IV. INTERPRETATION

A. Evaluation of the Results

In the first part of the comparative analysis and the used basis implementation, it was shown that the ML-procedure performed worse than the competitor in the chosen aspects of evaluation did. Enforcement, based on the runtime is possible, but the tours should not require significantly higher distances for that. The payment in the area of the "last-mile"-logistics is usually per finished stop. Because of this, the driver does not want to be slowed down by badly optimized routes. The prototype with the basic implementation of „REINFORCE with Baseline" partially does not recognize nearby stops and chooses unnecessary long distances in the first step of the analysis. It deviates +6.94 km and +14.13 min on average from the competitor's solution. The results in the second step of the analysis show an increase in the aspects of evaluation, compared to the first version. The average deviations between prototype solution and OR-strategy are +6,51 km +11,01 min. Also in the second part of the analysis, regarding runtime, the prototype is significantly faster in solving VRP-instances compared to the competitor, because he can execute the experience-based decisions. This is still too expansive for a productive environment in relation to the number of stops and should be optimized in terms of competitiveness. With regard to the named challenges and the resulting indirect requirements to the competitiveness of

 ISBN: 978-1-61208-981-2

the prototype in point one, the following resulting points can be derived, which the prototype is capable to fulfil or not.

- After finishing the training process, the prototype was able to detect a "good" up to an "excellent" solution.
- The prototype can detect solutions without manual interference.
- In the testing modus, the results of the prototype are provided within milliseconds.
- The prototype performs optimally on mainly known VRP-instances. If the major part of the VRP-instances is not known to the prototype, the prototype will not get close to finding an optimal solution.
- Because the prototype reacts badly to unknown VRP-instances, it is not possible that the prototype performs equally on all VRP-instances.

B. Comparison to Combinatory Optimization Algorithm

The comparative analysis has shown that the ML-procedure in the current implementation does not provide the same level of solution quality VRP-instances compared to optimization algorithms. Furthermore, it was shown that the chosen adjustment impulses for the basic implementation of "REINFORCE with baseline" did not provide a significant improvement. The adjustment impulses attained that the prototype was able to reach a better solution with less iterations. However, the improvement of the solution differs only slightly from the basic implementation. Regardless, the comparative studies of RL-methods, with different evaluation criteria and the same runtime, show rarely an optimal solution compared to combinatorial optimization algorithms [17]. This is justified by the fact that the fundamental goal of the RL-method is both, to avoid bad solutions as well as to achieve an average solution. Therefore, it also defines the goal of the prototype. In contrast to the prototype, the Google-OR-Solver considers the holistic structure of the problem and reaches an asymptotically optimal solution with enough runtime and processing power. Besides Google-OR, there are other alternatives for solving combinatorial optimizing problems, such as Concorde TSP Solver or the services of openrouteservices.

C. The Relevance of the „Last-Mile"-Logistics

With regard to the relevance of the prototype for the „last-Mile"-logistics and in consultation with the cooperating company, which is supporting a network of the "last-mile"-metropolitan logistics, the following results can be derived. The prototype is not efficient enough for low stop volume with less than 1200 stops per day. One of the reason for that is the fact that the manual dispatch effort and the usage of combinatorial optimizing algorithms, for less than 400 stops within a small "last-mile" area, are significantly less than the training effort of the prototype.

Starting from 400 stops, the Google-OR-Solver shows signs of weakness with a runtime of 10s. This was also shown by the work of [2], where the developed RL-method lead partially to even better results, for a high stop volume, compared to the OR tools. Considering the retro perspective and the feedback of the company, the decision of the RL for the "last-mile" was reconsidered and ideas regarding other ML-systems were elaborated. One of the favorite solutions is the implementation of supervised-learning to learn the dispatch mode of the supplier-dispatchers so that the ML-component is supportive and does not automatically solve VRP-instance. However, looking at solving Green Vehicle Routing Problems as a whole, e.g. [4] just partially agrees, because applying a multi-dimensional approach is suggested.

V. CONCLUSION

This article, provides a concept for the basic implementation of a reinforcement learning (RL)-method, in the form of a "REINFORCE with baseline", in combination with an Ant Colony Optimization (ACO)-algorithm to solve Vehicle Routing Problems (VRP). Regarding rural areas, the prototype is not suitable due to the necessary high amount of stops for the training. However, due to the large amount of stops, the developed prototype based on the enhanced implementation of "REINFORCE with baseline" can be, especially for urban areas, an alternative to third party providers such as Google-OR. In addition, impulses for the adaptive solving of VRP by using different optimization mechanism within the prototype are provided. It is also shown that the social factor is barely considered in the ML-context for solving VRP. Therefore, there is a potential for the prototype to gain a competitive advantage.

Looking ahead to further research landscapes of Machine Learning (ML) and VRP, the research results have shown that an enhanced implementation of an RL-method can achieve a good, up to an excellent result, regarding the investigated evaluation criteria compared to classic optimization methods. For future improvement of the prototype, a reduction of the complexity of the adjustable parameter could be considered in order to avoid possible "overfitting". Additionally, a redesigned prototype by using a different proximal policy according to [3] could lead to a relevant increase in performance. An extension of the prototype regarding other RL-procedures would be possible and important to validate the results. An exploration of further RL-procedures would be important as well, because in the field of RL the slightest changes in parameters or the used procedure result in strong deviations in the results. However, the human factor, such as personal preferences of the delivery personnel, should be incorporated in the optimization model and algorithm. This, for example, could consider preferences of dealing with environmental conditions (e.g., construction areas or traffic jams) on the proposed routes.

REFERENCES

[1] Bundesverband Paket und Expresslogistik e. V., "CEP Study 2021 - Analysis of the market in Germany: A study commissioned by the German Parcel and Express Logistics Association" Transl. Germany KE-CONSULT Kurte&Esser GbR. [Online]. Available from: https://www.biek.de/download.html?getfile=2897. 2021 [retrieved: Jun.2022.

[2] M. Nazari, A. Oroojlooy, M.Takac and L. V. Snyder, “Reinforcement Learning for Solvoing Vehicle Routing Problem”, arXiv:1802.04240, 2018

[3] C. C. Hsu, C. Mendler-Dünner, M. Hardt, Revisiting Design Choices in Proximal Policy Optimization, arXiv:2009.10897, 2020

[4] Moghdani et al., “The Green Vehicle Routing Problem: A Systematic Literature Review” Journal of Cleaner Procustion 279, 2021

[5] K. Braekers, K. Ramaekers, and I. Nieuwenhuyse, „The Vehicle Routing Problem: State of the Art Classification and Review“, Computers & Industrial Engineering, Vol. 99, pp.1-3. 2015

[6] G. B. Dantzig and J. H. Ramser, „The Truck Dispatching Problem“, Management Science, Vol. 6, No. 1, pp.80–91 [Online]. Available from: http://www.jstor.org/stable/2627477, 1959 [retrieved: Jun..2022

[7] R. Richter, M. Söding, and G. Christmann, “Logistics and mobility in the city of tomorrow. An expert study on last mile, sharing concepts and urban production” Transl. Germany, 2020

[8] S. S. Richard and G. B. Andrew, “Reinforcement Learning: An Introduction”, A Bradford Book, 2018

[9] D. Silver et al., „Mastering the game of Go without human knowledge“, Nature, Vol. 550, pp. 354-355, 2017

[10] Google (2021) Routing Options [Online]. Available from: https://developers.google.com/optimization/routing/routing_options#local_search_options (retrieved: Jun .2022).

[11] T. Vidal, G. Laporte and P. Matl: „A concise guide to existing and emerging vehicle routing problem variants“, European Journal of Operational Research, Vol. 286, pp. 2-3, 2019.

[12] H. S. Chang, W. J. Gutjahr, J. Yang and S. Park: „An ant system approach to Markov decision processes“, Proceedings of the 2004 American Control Conference, Vol. 4, pp. 3820-3825, 2004

[13] Y. Sun et al., “Boosting Ant Colony Optimization via Solution Prediction and Machine Learning” [Online]. Available from: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwihnfXOx474AhUvSfEDHd--B-cQFnoECAgQAQ&url=https%3A%2F%2Farxiv.org%2Fpdf%2F2008.04213&usg=AOvVaw3YbcGVlSr4VpkCz12nJhCM. [retrieved: Jun-2022]

[14] U. Lorenz, „Reinforcement Learning“, Springer Berlin Heidelberg, 2020

[15] M. L. Puterman: Markov Decision Processes: Discrete Stochastic Dynamic Programming, 2014 [Online], Available from: https://books.google.de/books?id=VvBjBAAAQBAJ. [retrieved: Jun. 2022]

[16] R. S. Sutton and A. G. Barto: Reinforcement Learning: An Introduction. 2018 [Online], The MIT Press. Available from: http://incompleteideas.net/book/the-book-2nd.html. [retrieved Jun. 2022]

[17] N. Mazyavkina, S. Sviridov, S. Ivanov and E. Burnaev: „Reinforcement Learning for Combinatorial Optimization: A Survey“, CoRR, abs/2003.03600, 2020.

 ISBN: 978-1-61208-981-2