Learning Value Functions for Same-Day Delivery Problems in the Tardiness Regime*

Nikolaus Frohner and Günther R. Raidl

Institute of Logic and Computation, TU Wien, Vienna, Austria
{nfrohner|raidl}@ac.tuwien.ac.at

Same-day delivery problems [6] are concerned with delivering orders placed dynamically by customers on the same day. Delivery route planning and driver dispatching have to be performed online within short time. We consider a problem variant [3] originating from a real-world online super market in Vienna, where orders have to be delivered within either one or two hours and where delivery deadlines are soft, while every order has to be delivered. The optimization goal is to minimize and balance the tardiness and the travel times.

Our current solution approach makes use of an Adaptive Large Neighborhood Search (ALNS) [5] with a surrogate objective function [2]. It replaces a computationally intense online scenario sampling/consensus function approach [1] by an offline training phase, where we learn to estimate the quality of routes which start in the future to account for the dynamic and stochastic aspect of the problem. We observed that different surrogate models are superior both to myopic optimization and the sampling approach on artificial instances and also in real-world use as reported by practitioners. An explanation for the behavior of the surrogate functions is that they favor routes which have larger slacks and are more widespread over the delivery area. Routes might look inefficient in a static context and would not be constructed considering only the myopic costs, but are likely to be improved over time when new orders arrive. An extreme example is a route containing exactly one far away but delayable order.

This approach is targeted at the so-called zero-tardiness regime, where we have sufficient capacity to satisfy the demand to create routes with (almost) no tardiness. There, we replaced the myopic travel distance with the aforementioned surrogate function. When the demand exceeds the capacity, we move to the tardiness regime, where the goal of the optimization is to reduce and balance the unavoidable tardiness evenly among the customers. We observe that in this case the performance decreases, since the travel times are only indirectly respected. The approach focuses then only on the tardiness in the static context facilitated by creating shorter routes with as many available vehicles as possible. Due to the waste of driver resources, these inefficient routes might then result into even more tardiness for later orders—a self-amplifying downward trend. This can be addressed by willingly accepting additional tardiness. In Figure 1, we see on an exemplary real-world instance the positive impact of the route efficiency when allowing extra tardiness of a couple of minutes.

* This project is partially funded by the Doctoral Program "Vienna Graduate School on Computational Optimization", Austrian Science Foundation (FWF) Project No. W1260-N35.

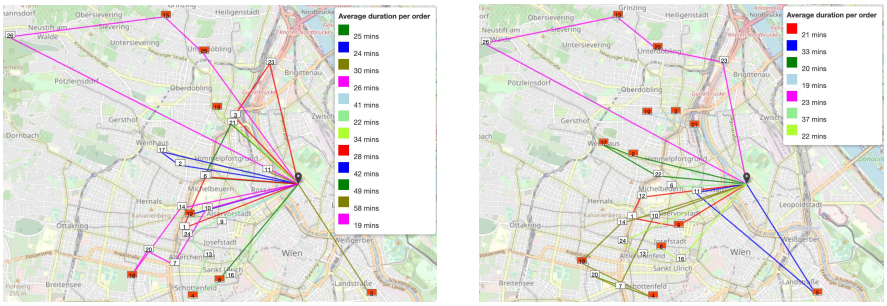

Fig. 1. Planned routes in a high load situation of a real-world instance in Vienna. Left: Myopic minimization of the tardiness, where many, partially inefficient, routes are planned. Right: Fewer, more efficient routes are planned when we allow extra 5 minutes of tardiness per order, naturally resulting into more tardy orders (in orange).

In our current work, we formalize our problem as Markov Decision Process to model explicitly the immediate reward R for an action (i.e., starting a route with a certain tardiness) and the value V of the resulting post-decision state (i.e., how much remaining tardiness do we expect). Following [4], a neural network is employed to create a value approximation $\hat{V}$ using temporal difference learning in an offline training phase. Both reward and value are then used in the objective function of the online optimization by the ALNS. Relevant features for the network need to be investigated—the predicted route costs, the driver capacity, remaining expected orders, and mean order delivery times are expected to be relevant. The intuitive goal is to learn making clever sacrifices regarding the current reward leading to less tardiness over the whole day.

References

1. Bent, R.W., Van Hentenryck, P.: Scenario-based planning for partially dynamic vehicle routing with stochastic customers. Operations Research **52**(6), 977–987 (2004)
2. Bracher, A., Frohner, N., Raidl, G.R.: Learning surrogate functions for the short-horizon planning in same-day delivery problems. In: Stuckey, P.J. (ed.) 17th International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR'21). LNCS, vol. 12735, pp. 283–298. Springer, Vienna, Austria (2021)
3. Frohner, N., Raidl, G.R.: A double-horizon approach to a purely dynamic and stochastic vehicle routing problem with delivery deadlines and shift flexibility. In: Causmaecker, P.D., et al. (eds.) Proceedings of the 13th International Conference on the Practice and Theory of Automated Timetabling - PATAT 2021: Volume I. Bruges, Belgium (2020)
4. Joe, W., Lau, H.C.: Deep reinforcement learning approach to solve dynamic vehicle routing problem with stochastic customers. In: Proceedings of the International Conference on Automated Planning and Scheduling. vol. 30, pp. 394–402 (2020)
5. Ropke, S., Pisinger, D.: An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. Transportation Science **40**(4), 455–472 (2006)
6. Voccia, S.A., Campbell, A.M., Thomas, B.W.: The same-day delivery problem for online purchases. Transportation Science **53**(1), 167–184 (2019)