Continuous Planning of a Fleet of Shuttle Vans as Support for Dynamic Pricing

Filip Stavrov
stavrovf@gmail.com
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia

Luka Stopar
luka.stopar@ijs.si
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia

ABSTRACT

This paper solves the problem of estimating the number and type of required resources for pickup and delivery of passengers at some time in the future. By combining optimization and sampling methods, as well as making plans based on several statistical samples, we estimate the real values for the required resources and show how the sample values converge towards the real values. Our approach combines machine-learning based demand predictions, for the number of passengers, and a route optimization engine that assigns the passengers into shared shuttle vehicles. In order to validate our method we create a baseline data that is representative of the real values. We test our approach using this baseline data, and we obtain statistically significant results.

KEYWORDS

statistical samples, demand predictions, route optimization engine, sampling techniques, optimization technique

1 INTRODUCTION

The effective allocation of resources is a critical topic in the mobility industry. Anticipating the number and type of resources required can significantly enhance a company's ability to plan accurately for the future. Our work addresses this challenge by focusing on how to estimate the number and type of vehicles needed for passenger pickup and delivery at a future time. The input to our problem consists of machine learning-based demand predictions, which provide estimates of the number of passengers across various routes offered by the company. These predictions are provided daily and further broken down into hourly estimates for each day.

*Both authors contributed equally to this research.

Information Society 2024, 7–11 October 2024, Ljubljana, Slovenia

https://doi.org/10.70314/is.2024.sikdd.27

Once we receive these predictions, our goal is to simulate reservations based on this data. For instance, if the predictions indicate that 12 passengers will travel from Ljubljana to Koper on October 20, 2024, we would simulate reservations using sampling techniques. One particular example is creating four separate bookings—one for five passengers, one for three, and two for two passengers each. We will introduce the sampling techniques used in this process in greater detail later on.

After generating these reservations, the next step is to input them into the Route Optimization Engine to generate a plan for that day. This plan will specify the number of vehicles required and the specific reservations each vehicle will serve.

The main hypotheses that our approach explores and experimentally tests are the following:

- H1: We can accurately estimate the number of required resources using optimization methods based on predicted passenger numbers.
- H2: Monte Carlo sampling of historical distributions can effectively model uncertainty in demand predictions, leading to stable resource estimations.
- H3: Creating plans based on several sample values will converge towards the actual number of required resources.

On the other hand, the key assumptions and limitations that underline our research are:

- Prediction Accuracy: We assume that the predictions effectively estimate the number of future passengers.
- Passenger Distribution: We assume that the number of passengers follows a Poisson distribution and that the distributions on different routes are independent.
- Independence: We assume that the passenger distribution and the window type distributions are independent to each other.
- Concept Drift: We assume there is no concept drift in the data, meaning the underlying data patterns do not change over time.

2 RELATED WORK

The problem of resource allocation in the mobility industry, particularly in the context of vehicle routing and passenger demand prediction, has been extensively studied. Traditional methods for vehicle routing often rely on static models that assume known and deterministic demand. However, recent advances in machine learning and optimization have enabled more dynamic approaches that can account for uncertainty and variability in demand. [3][4] For instance, predictive analytics has been employed to forecast passenger demand using historical data, which can then be fed into optimization algorithms to determine the optimal allocation of vehicles. Monte Carlo simulation is another technique commonly used to model uncertainty in demand predictions, providing a probabilistic framework for decision-making under uncertainty. [2] Moreover, dynamic vehicle routing approaches, have demonstrated the benefits of real-time adjustments to routing plans based on updated demand information. [1] The integration of these methodologies into a continuous planning framework is relatively novel and addresses the limitations of static planning approaches, particularly in highly variable and uncertain environments. [1][5]

3 METHODOLOGY

Our methodology begins with demand predictions for the number of passengers, and the ultimate goal is to determine the number and type of vehicles required, as well as the reservations each vehicle will serve. The figure below provides a detailed overview of this process.

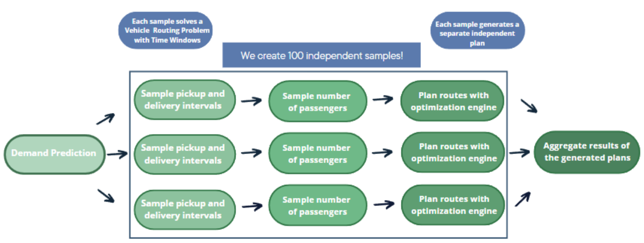

Figure 1. Methodology

Starting with the demand predictions, we apply sampling techniques to simulate reservation data. Specifically, we take the predicted number of passengers for different routes at various times and generate reservations through sampling. This reservation data follows a specific format, including fields such as ID, start location, end location, pickup time, and more. Key attributes include the number of passengers per reservation and the window type, which reflects travel preferences. For instance, some passengers may prefer a private vehicle (VIP), while others are open to sharing the ride. Additionally, the window interval is crucial—it can be a specific time or a more flexible period, affecting both the service pricing and overall experience. These factors will be incorporated into the dynamic pricing model later on.

The process begins with demand predictions and culminates in the generation of reservation data. Critical steps include sampling the number of passengers per reservation, the window type, and the window length. Sampling is done from probabilistic distributions derived from historical data, with the distributions illustrated below.

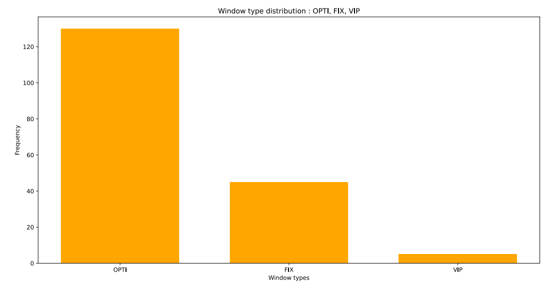

Figure 2. Window type distribution

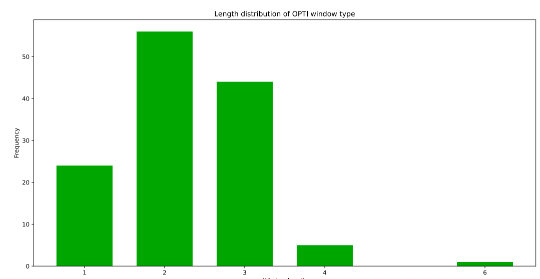

Figure 3. Window length distribution

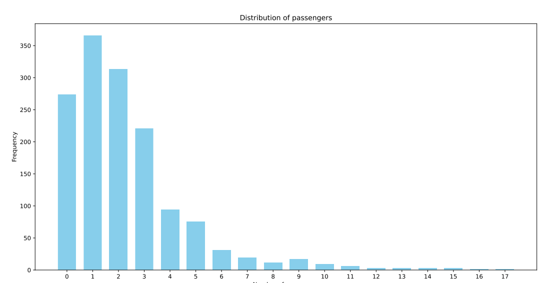

Figure 4. Number of passengers distribution

Please note that from a single demand prediction input file, we generate 100 independent samples of reservation data. This approach introduces uncertainty through probabilistic sampling. Each independent sample is then submitted as a separate job to the Route Optimization Engine, where it solves a vehicle routing problem with time constraints. The output for each job is a plan corresponding to the reservation data. Our final objective is to aggregate these results and analyze the insights they provide.

4 RESULTS

After solving all 100 jobs, we obtained 100 independent plans and began analyzing the results. As shown in the figure below, the distribution of the number of passengers yielded a mean value of 325.87 with a standard deviation of 16.85. For the number of vehicles, the mean was 38.01 with a standard deviation of 3.06. It's notable that the passenger data exhibits significantly more variance compared to the vehicle data. This is expected, as passengers are grouped into visits, and visits are then allocated to vehicles, resulting in less variation in the vehicle count.

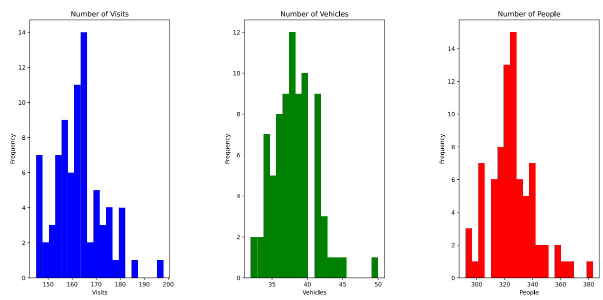

Figure 5. Sampled data: visits, vehicles and passengers distributions

To further validate our approach, we created a baseline using the same data from which the demand predictions were generated. We generated 100 samples from this baseline and submitted them as independent jobs. Upon completion, we compared the baseline results with those of our sampled data. The mean number of vehicles from the baseline was 37.81 with a standard deviation of 3.01, which closely aligns with the values from our sampled data. You can observe the comparison on the figure below.

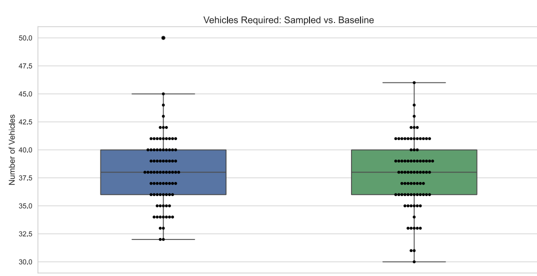

Figure 6. Comparison of required vehicles between sampled and baseline data

We also analyzed the error distribution for the number of vehicles between the baseline and sampled data, finding a mean absolute error of 3.16. This suggests that the difference between the two sets is minor, considering the sampling of data, and it is indicating a good alignment. Additionally, the average number of vehicles in both the sampled and baseline data is quite similar. While the mean absolute error reflects some variability in the sampled values, this is acceptable given the overall similarity to the global mean, and the sampling of values. Thus, despite the variance, the sampled values converge towards the actual values. This error distribution is displayed on the figure below.

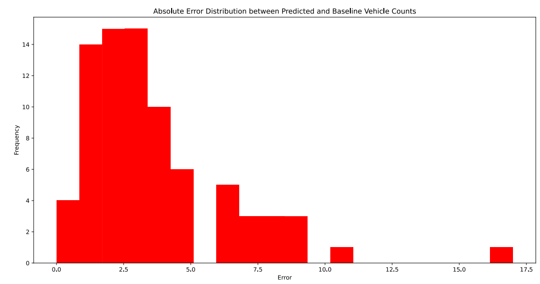

Figure 7. Required vehicles - error distribution

To statistically test whether the sampled and baseline data have the same mean number of vehicles, we conducted a Welch's t-test. The results showed a test statistic of 0.59, a p-value of 0.55, and a 95% confidence interval ranging from -0.64 to 1.23. Given the p-value, we fail to reject the null hypothesis, meaning there is no statistically significant difference between the sampled and baseline vehicle counts. Additionally, the range of the mean difference of vehicles between the sampled and the baseline data, which is from - 0.64 to 1.23, falls within our practical significance threshold of up to 2 vehicles, further supporting the similarity between the two datasets. This indicates that we can effectively estimate the number of required resources by applying optimization techniques on top of the demand prediction values.

We also analyzed the mean number of vehicles and observed that this value converges toward the actual values as the number of samples increases. This is shown on the figure below.

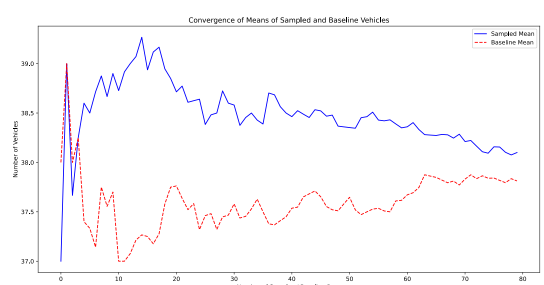

Figure 8. Convergence of means of sampled vehicles

Finally, after obtaining both the number of passengers and the number of vehicles, we decided to fit a linear regression to explore whether we could simplify the process and avoid the detailed approach previously described. As illustrated in the figure below, the regression line serves as a reasonable estimator for the number of vehicles based on the number of passengers. However, this model struggles to capture the non-linear relationships influenced by various optimization types, window lengths, and travel modes, resulting in considerable variance around the regression line. While it is generally true that a higher number of passengers correlates with an increased number of vehicles, this relationship can be misleading. Different travel types can accommodate more passengers per vehicle, which can disrupt the linear relationship, especially in cases where these travel types dominate. Consequently, although the linear regression provides a solid approximation, it overlooks essential non-linear factors that are critical to our analysis. Our approach, which integrates these factors, demonstrates greater robustness and effectiveness. The linear regression line and the data correlation are presented in the figure below.

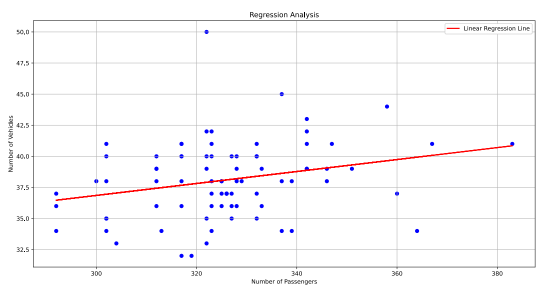

Figure 9. Regression Analysis

5 CONCLUSION

In conclusion, our findings demonstrate that we can effectively estimate the number of required resources by employing optimization methods based on predicted passenger numbers. As the number of samples increases, the sampled values consistently converge toward the actual resource requirements, reinforcing the reliability of our approach. Alternative methods, such as linear regression, fail to adequately address the non-linear complexities inherent in resource allocation, such as varying optimization types and window lengths. Our method, which incorporates these factors, proves to be a far more accurate and effective solution for resource estimation in the mobility industry.

ACKNOWLEDGMENTS

Our research is part of a broader, multi-partner initiative called CONDUCTOR. The primary objective of this project is to design, integrate, and demonstrate advanced, high-level traffic and fleet management systems. These systems aim to optimize the transport of passengers and goods efficiently on a global scale, ensuring seamless multimodality and interoperability. The CONDUCTOR project is co-funded by the European Union's Horizon Europe research and innovation programme under the Grant Agreement No 101077049.

REFERENCES

[1] Berbeglia, G., Cordeau, J. F., & Laporte, G. (2010). Dynamic pickup and delivery problems. Transportation Research Part B: Methodological, 44(5), 667-684. https://doi.org/10.1016/j.trb.2009.10.004

[2] Ulmer, M. W., Thomas, B. W., & Mattfeld, D. C. (2018). Preemptive depot returns for same-day delivery under uncertain customer availability. European Journal of Operational Research, 269(2), 356-371. https://doi.org/10.1016/j.ejor.2017.08.008

[3] Bertsimas, D., & Sim, M. (2004). The Price of Robustness. Operations Research, 52(1), 35-53. https://doi.org/10.1287/opre.1030.0065

[4] Ghiani, G., Guerriero, F., Laporte, G., & Musmanno, R. (2003). Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies. European Journal of Operational Research, 151(1), 1-11. https://www.sciencedirect.com/science/article/abs/pii/S0377221702009153

[5] Psaraftis, H. N., Wen, M., & Kontovas, C. A. (2016). Dynamic vehicle routing problems: Three decades and counting. Networks, 67(1), 3-31. https://doi.org/10.1002/net.21628