



HAL
open science

A Classifier Approach for Probabilistic Forecasting in Transportation Planning

Ken Z M Arief, Justine Castan, Maria Restrepo Ruiz, Gauthier Soleilhac

► **To cite this version:**

Ken Z M Arief, Justine Castan, Maria Restrepo Ruiz, Gauthier Soleilhac. A Classifier Approach for Probabilistic Forecasting in Transportation Planning. ORBEL 2023, May 2023, Liège, Belgium. hal-04151814

HAL Id: hal-04151814

<https://hal.science/hal-04151814v1>

Submitted on 17 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Classifier Approach for Probabilistic Forecasting in Transportation Planning

Ken Z. M. Arief

IMT Atlantique, 4 rue Alfred Kastler, F-44307 Nantes, France
Universitas Gadjah Mada, Bulaksumur, I-55281 Yogyakarta, Indonesia
e-mail: ken.arief@imt-atlantique.net

Justine Castan,
Maria I. Restrepo-Ruiz

IMT Atlantique, 4 rue Alfred Kastler, F-44307 Nantes, France

Gauthier Soleilhac
CRC Services, Rueil-Malmaison, France.

Keywords: Forecasting, Transportation Planning, Stochastic Demand

1 Introduction and Problem Description

The freight transportation industry in France is becoming highly competitive, complex, and is currently facing many tensions. Environmental issues, energy prices, and the difficulty in recruiting drivers have accelerated the practice of pooling transport between several producers with common customers.

In the ultra-fresh product sector, producers are required to market products that must be consumed in a timely manner. Additionally, customers usually place their orders for next-day or even same-day delivery which makes for tight logistics flows. However, producers still need to book vehicles days before the date of dispatch in order to have sufficient capacity available and to let the carriers organize themselves.

The implementation of a shared ecosystem in this type of sector requires the coordination of transport planning when **customer demand is not known with certainty**. To address this issue, it becomes necessary to develop data-based decision-making tools for freight transportation under uncertainty. This problem has been addressed in [1]. The authors propose a two-stage stochastic model where the objective is to minimize the total expected cost of transporting all customers' orders and the transportation plans needs to be done one week in advance.

To model the decision making under uncertainty, the use of scenarios, where customers have varying values of demand, and its each own probability, is one of the most viable option. This decision making process also requires the **development of accurate forecasting models for demand planning**, since forecasting will produce the data to prepare and simulate the scenarios. Furthermore, poor quality predictions could lead to various problems such as shortages, over-stocks, delays, bullwhip effects, and lost customers [2]. Traditional point forecasting models do not capture the full range of possible outcomes (i.e. scenarios). Therefore, a method to provide probabilistic predictions of the set of possible outcomes might be adopted to build a stochastic transportation optimization model. Thus, probabilistic forecasting can be employed within this context.

Methods including time series or regression-based models and machine learning are able to generate probabilistic forecasts. But most of them either only provide the range of the possible outcomes or the probability of a certain forecast results, not both. To accommodate both variables, confidence intervals and probability distributions can be used. Conditional probability is proven to be effective in integrating these two to provide probabilistic insights [3]. Hence, we aim to develop an algorithm to forecast stochastic customers' demand through conditional probability as the end product. This algorithm ought to work by predicting the likely values and associated probabilities of the actual demand based on the known forecast. This will later be used to generate better scenarios for the stochastic optimization model in [1].

2 Methods and Findings

We perform Exploratory Data Analysis (EDA) on 7 daily time series data sets over a period of 10 months supplied by ultra-fresh companies. Each data set contains 2 variables: 1) actual historical demand, and 2) existing forecast of demand. The EDA shows differences in mean and standard deviation for each client and also highlights different time series patterns throughout the time. A feature selection process is also performed to take into account the time series characteristics (i.e. day of the week, month). Through seasonality analysis, the data reveals a 7-day pattern. Furthermore, from the feature extraction, the day of the week produces the highest importance factor, reaching around 80 percent of importance, while other features have negligible importance value. Thus, we split each data based on the day of the week and analyze the probability distribution within each group of days separately.

We use the prediction intervals approach and combine it with the probability distribution of both the actual and forecast data. Then, it is simplified into discrete histogram bins with their respective probabilities. The bins can be set to accommodate the desired number of scenarios for the optimization model. The algorithm is trained to predict the bin class of the actual demand based on the bin class of the forecast for each day of the week and for each client. To compare the results of the predictions, multi-class classification methods who inquire similar approach are also employed (Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbour, and Support Vector Machine). Metrics including accuracy and confusion matrix are computed to evaluate the performance of all models and they provide at least 60-percent accuracy for all the trained data sets. These preliminary results are deemed to be enough to generate the probabilistic scenarios, considering the limited data and forecasting algorithm.

Overall, modeling the probability of the demand is a crucial aspect in this step, which can provide valuable inputs for the optimization of transportation plans under stochastic demand in future works. This research is carried out in the framework of CRC Lab, a joint laboratory between IMT Atlantique and CRC Services company.

References

- [1] J. Castan, J. Medina, O. Péton, M-I. Restrepo. Un problème de tournée de véhicules avec demandes stochastiques et opportunités de cross-dock. *ROADEF 2023*. Feb 2023, Rennes, France.
- [2] J. T. Mentzer, M. B. Myers, T. P. Stank. *Handbook of Global Supply Chain Management*. 2007. CA: SAGE Publications, Inc.
- [3] H. Ba, S. Guo, Y. Zhong, S. He, X. Wu. Quantification of the forecast uncertainty using conditional probability and updating models. *Hydrology Research* 1 December 2019. 50 (6):1751–1771.