

A Methodology for Assisting Customers Service Selection in Freight Transportation by Data Analytics

Aysel Biyik,¹ Bruno Albert Neumann-Saavedra,¹ Dirk Christian Mattfeld¹

Abstract: Over the past several decades, logistics activities have been sourced by freight logistics providers (FLPs). However, by the development of digital technologies, digital freight platforms (DFPs) has started to substitute FLPs. Due to the complexity of the freight transportation domain, it is challenging to provide customer-oriented delivery solutions. In such a complex process, which involves many actors, it is necessary to develop a methodology to assist customer service selection. This methodology should be able to represent the multi-modal network adequately with respect to customer preferences, learn from customer's past choices via data and generate accurate and quick solutions to satisfy them. Therefore, this paper proposes a methodology that uses data analytics, predictive analytics, to estimate the cost in multi-modal freight transportation concerning customer preferences. Given some customer preferences, the proposed methodology is able to match a variety of services as closely as possible.

Keywords: digital transformation; digital freight platforms; multi-modal freight transportation; customer expectations; data analytics; predictive analytics

1 Introduction

In recent years, digital platform companies have emerged as a key player in the logistics market to fulfil customer expectations more efficiently. In the field of freight transportation, digital freight platforms (DFPs) play the role of intermediate between sender, carriers, and customers. DFPs aim to substitute human interactions with an automated process for offering convenient multi-modal transport services to stakeholders. Indeed, the acceptance of such DFPs is rapidly growing to such an extent that they have been starting to replace traditional freight logistics providers (FLPs). For instance, digital platforms such as Cargomatic [1] enable cost-efficient, real-time, and on-demand arrangements of transports that cut into the domain of logistics services. Reports indicate that FLPs focusing on standardized services such as transportation and warehousing are likely to lose significant market share to new transport technologies and customized service solutions, see for example [2]. For a more detailed discussion about the impact of digital platforms on FLPs, the reader is referred to [3].

One major challenge in adopting DFPs is to meet customer expectations of service by automatic procedures rather than human expertise in such a complex process, which involves

¹ Technische Universität Braunschweig, Decision Support Group, Mühlenpfordtstraße 23, 38106 Braunschweig
a.biyik@tu-braunschweig.de

many actors that have a different perspective of service evaluation. Traditional service providers have advantages utilizing long-time experience in industry and customer. However, as the nature of the service is complex, providing proper service to customers with a quick response is challenging with the human experience. One reason for this is that there are several criteria a customer may consider selecting transport services. These criteria include cost, CO_2 emission, transport mode, flexibility, and time reliability of these services [2]. Every customer weights these criteria differently according to his or her preferences. To address this challenge, the field of operations research presents a rich literature on multi-objective optimization models on multi-modal transport to reach optimal solutions with respect to customer preferences [4]; [5]; [6]. Nevertheless, finding optimal solutions for real-world sized instances of such models typically involves large computational times and is not unbiased, as the weighting of these criteria is different for each customer. Thus, it is challenging that a DFPs can provide customers with a quick response to their requests by using optimization models. Indeed, providing a quick response is essential for the success of DFPs, as customers do not intend to spend so much time to get suitable results from the digital platforms. These situations encourage researchers to benefit from other methods like data analytics in such a complex domain. Therefore, it is proposed a methodology that can learn from customer's past choices to generate quick solutions of acceptable accuracy. Thus, by embedding proposed methodology to the DFPs, provides a reliable estimation of the quality of the offered service.

This study proposes a methodology by using data analytics to predict a solution that satisfies customer preferences most. The aim is to quickly generate a solution as well as an accurate estimation of their actual quality. We rely on a mathematical model to generate instances. Given the individual preferences of a customer, we can generate the optimal solution based on an instance. Then, we build a predictive model by machine learning aiming at assessing relationships among criteria. Thus, from a pool of solutions of such an instance, a predictive model determines routing options with respect to customer preferences without the need for conducting an optimization process. We test our methodology on a case study based on portions of the transportation network in Shanghai-Singapore. We consider six criteria a customer may concern when selecting a service: cost, delivery time, distance, CO_2 emission, mode of transportation, number of transshipment. Statistical techniques have been applied to depict patterns, correlations, and other insights among parameters. The tendency of the relation of parameters has led us to implement Multiple Linear Regression (MLR). A cross-validation method has been applied to prevent overfitting. A model can make accurate predictions on unseen data; we observe that it can generalize from the training set to the test set without any help from a human. Besides, computational experiments show that our methodology reduces the long computational time given a large set of criteria to consider when determining routing decisions. The proposed methodology enables to match a variety of services as closely as possible with respect to customer preferences.

The rest of this paper is organized as follows. Section 2 outlines the proposed approach and methodology. Section 3 provides a brief overview of the case study to evaluate the

experimental results of the proposed approach on portions of the transportation network in Shanghai-Singapore.

2 Proposed Methodology

The overall approach is conducted by the predictive analytic and mathematical model. Predictive analytics covers a variety of statistical techniques from data mining, predictive modelling, and machine learning that analyse current and historical data to make predictions about the future. Predictive analytics does not only assist in creating practically useful models, but they also play an important role alongside explanatory modelling in theory building and theory testing [9]. When building the predictive model required dataset is generated by using the mathematical model. The proposed approach is given in Figure 1. It consists of the following layers systematically.

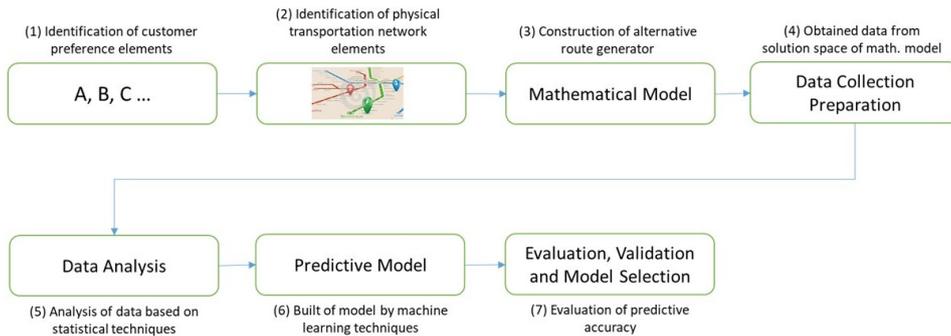


Fig. 1: The proposed methodology.

A. Identification of Customer Preference Elements This step is the foundation of the mathematical model as well as the predictive model. Customer expectations or preferences should be selected and represented accurately. In this study, it is defined based on previous literature.

B. Identification of Physical Transportation Network Elements This step comprises the use of data derived from GPS (global positioning system), sensor detection of the traffic status and incidents, among others.

C. Mathematical Model (Service Graph) In this step, we design the optimization model aiming at determining a routing solution which satisfies customer expectations at most. We use the collected data to generate real-world based instances of the mathematical model as well as alternative route solutions according to customer requests.

D. Data Collection and Preparation This dataset is constructed by taking samples from the routing solution space of the model. The initial number of variables is usually large to

capture new sources of information and new relationships. The explanation for each variable is based on combining theory, domain knowledge, and exploratory analysis.

E. Data Analysis In this step, the data is aggregated to reduce their dimension and to handle outliers. Due to the often large number of predictors, reducing the dimension can help reduce sampling variance (even at the cost of increasing bias), and in turn increase predictive accuracy [7]. Statistical techniques are used to analyse how the dependent variable changes with the variation of any of the independent variables keeping other variables fixed. Correlation analysis is used to measure the strength and direction of the association among variables. In addition, visualization provides the means for learning about different measurements of quality as well as associations derived from predictive modelling.

F. Predictive Model We design a predictive model to estimate the quality of a solution that potentially answers customer preferences at most. Machine-learning algorithms based on supervised learning are applied for the predictive model. Given data in the form of examples with labels, we can feed a learning algorithm that pairs these example-labels one by one. Thus, we allow the algorithm to predict the label for each example, giving it feedback as to whether it predicted the right answer or not. Over time, the algorithm will learn to approximate the exact nature of the relationship between examples and their labels. When fully trained, the supervised learning algorithm will be able to observe a new, never-before-seen example and predict a good label for it.

G. Evaluation, Validation and Model Selection This part is applied to certain methods for evaluating the predictive power of a model by such as MSE (the mean square error), R^2 (measure of goodness-of-fit of the model or accuracy of the model). Predictive power is related to an empirical model's ability to predict new observations accurately [7]. For evaluating, it can be applied the method such as portioning dataset or crossvalidation. Over-fitting is a major focus in predictive analytics [8]; [9]. Assessing over-fitting is achieved by comparing the performance of the training and holdout sets. Then, the model selection step is applied, is aimed at finding the right level of model complexity that balances bias and variance, in order to achieve high predictive accuracy.

3 Case Study and Results

The proposed approach is evaluated in a case study. The transportation network presented in [10] is considered. The network includes four cities (Shanghai, Wuxi, Singapore, Malaysia), 16 ports and 4 transportation modes. A product originates from different cities and has different destinations. Each city has four airports, railway stations, seaports and warehouses. There are in total 50 direct routes connecting different ports. Each route has a specific transportation mode, transportation cost, delivery time and distance. Each transportation mode has a specific capacity, speed and CO_2 emission level.

Breadth-first search algorithm was used to generate routes. Then, 50 different alternative route samples have been obtained to correspond to the request of the customer. In the data

analysis step, it was found that costs and CO_2 emissions are directly proportional to the origin-destination distance. It was also observed that CO_2 emissions do not increase with respect to the number of transshipment. If environmental efficiency is of utmost importance for a customer, he or she will usually prefer the ship option, even if it involves large delivery times.

The predictive model is constructed by using MLR analysis. MLR is a suitable and efficient method for our continuous dataset due to the tendency of the relation of parameters. According to the ANOVA test, the effect of the number of transshipment has not a significant contribution to the model, and it can be removed from the model in order to decrease the complexity of the model. A common measure is the coefficient of determination R^2 , which measures the fraction of the total variation in the dependent variable that is captured by the model [11]. The higher value, the better our model fits the data. R^2 is obtained 90.90% and adjusted R^2 is obtained 90.09%. By omitting the statistically insignificant parameter from the model, R^2 and adjusted R^2 is obtained 90.65% and 90.04%, respectively. Moreover, the normality test has been done and it was found that the residuals of the model were normally distributed, and the variance was homogeneous. It means that the test results are usually reliable when the sample is large enough.

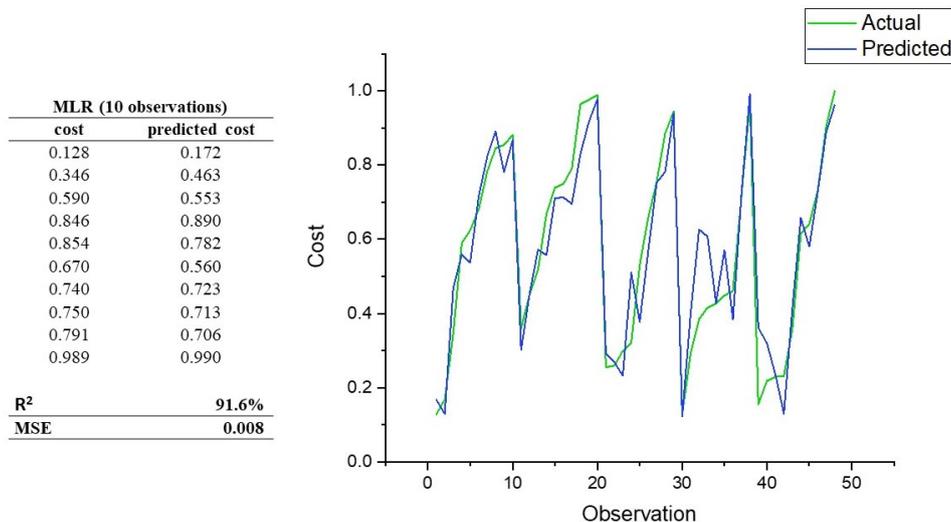


Fig. 2: The performance of MLR; 10 observations for actual and predicted cost value, and comparison of the actual versus predicted cost in 50 observations.

To prevent overfitting, cross-validation method is applied, the data are randomly partitioned into 10 mutually exclusive subsets and the algorithm is run 10 times, with each run on a different set of 9 subsets joined as a training set and with testing done on the remaining subset. The 10 runs thus produce 10 different parameter sets for the algorithm, and the prediction performances of these runs can be compared to each other. After cross-validation, the model performance is that R^2 is 91.6% and MSE is 0.008. Actual and predicted cost for 10 samples

is given in Figure 2. The graphical display, to compare the actual versus the predicted cost, is given in Figure 2. The predictive model is constructed to estimate transportation cost in multi-modal freight transportation network without long computational time. To apply these techniques through the case study, Minitab, Python, and Rapid-Miner were used [12]; [13]. The case study results motivate further research. For the next step, at the modelling, other well-known machine-learning algorithms such as Neural Network and Support Vector Machine will be implemented to compare their performance, and to choose the best. Then, the number of parameters in the dataset will be extended regard to different customer preferences such as service quality attributes.

Bibliography

- [1] Cargomatic Homepage, <https://www.cargomatic.com/lncs>. Last accessed 4 Oct 2019
- [2] Handfield, R. and Straube, F. and Pfohl, H. and Wieland, A.: Embracing global logistics complexity to drive market advantage. DVV Media Group GmbH, BVL International ,(2013)
- [3] Hofmann, E., Osterwalder, F.: Third-party logistics providers in the digital age: towards a new competitive arena?. *Logistics* **1**(2), 9 (2017)
- [4] Bektas, T., Crainic, TG.: A brief overview of intermodal transportation. *Logistics Engineering Handbook*. In G. D. Taylor edn. Taylor and Francis Group, Boca Raton, FL, USA (2008)
- [5] Caris, A., Macharis, C., Janssens, GK.: Planning problems in intermodal freight transport: accomplishments and prospects. *Transportation Planning and Technology* **31**(3), 277–302 (2008)
- [6] Crainic, TG., Kim, K.: Intermodal transportation. *Handbooks in operations research and management science* **14**, 467–537 (2007)
- [7] Shmueli, G., Koppius, O.R.: Predictive analytics in information systems research. *MIS quarterly* , 553–572 (2011)
- [8] Stone, M.: Cross-validators choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)* **36**(2), 111–133 (1974)
- [9] Hastie, T., Tibshirani, R., Friedman, J., Franklin, J.: The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer* **27**(2), 83–85 (2005)
- [10] Multimodal Transportation Network, <https://github.com/hzjken/multimodal-transportation-optimization>. Last accessed 1 Feb 2019
- [11] VanderPlas, J.: *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media, Inc. (2016)
- [12] Minitab, <https://www.minitab.com/en-us/>. Last accessed 1 Dec 2019
- [13] RapidMiner, <https://rapidminer.com/>. Last accessed 1 Dec 2019