*Journal of Industrial and Systems Engineering*
Vol. 14, Special issue: 18th International Industrial Engineering Conference
Winter (January) 2022, pp. 86-90

**17 November 2021**
***Iran Institute of Industrial Engineering***

# Data-driven optimization model: Digikala case study

**Sanaz Hamidi[1], Seyed Mohammad Taghi Fatemi Ghomi[1*]**

[1]*Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran*

*sanazh3@aut.ac.ir, fatemi@aut.ac.ir*

**Abstract**

Increasing software as a service (SaaS) requires the provision of more updated models for services, so trying to develop a model customized for the customer is important. We used the linear Knapsack problem model proposed by Mike Hewitt and Emma Frejinger in 2020. Then historical data of Digikala was applied and shown that how the model works on it.
**Keywords:** Optimization modeling, statistical learning, mixed integer linear programming, third-party logistics

## 1-Introduction

Third-party logistics (3PL) companies have now increased their products' sales and the use of online sales sites and software. Classic optimization models help with organizational planning systems. So, we try to use historical data and business rules. Software as a Service (SaaS) works by using applications over the Internet as a service. This market is large and growing. It is estimated to grow the SAAS market by $272.49 billion in 2021. The user accesses the software using an application or browser through the Internet. Centrally hosting the application has benefits for both the customer and the provider. In this paper, we use historical data and training data from the 3PL company to review the linear Knapsack problem model proposed by Mike Hewitt and Emma Frejinger (Hewitt & Frejinger, 2020). This company has been selected as a third-party logistics company that sells products through software. The performance of third-party logistics is such that the owner of the products is assigned to 3PL companies to carry out transportation, customs, warehousing, order fulfillment, distribution, etc. Furthermore, a third-party logistics provider is a company that offers its customers logistics services for part or all of their supply chain management functions.

Section 2 reviews the relevant literature, whereas section 3 introduces the framework for a general mathematical program and discusses the General Optimization Problem (GOP) and business rules utilized to assess the effectiveness of the framework and explain its application. Section 4 describes how the data sets were generated to validate the framework. Then, section 5 reports the effectiveness of the framework.

## 2-Literature review

Lombardi, Milano, & Bartolini propose a methodology to integrate constraints related to complex systems in an optimization model (Lombardi, Milano, & Bartolini, 2017).

---

*Corresponding author
ISSN: 1735-8272, Copyright c 2022 JISE. All rights reserved

They assume an accurate simulator and use it to generate data for supervised representation learning. They delineate the general setting including the embedding of linear regression in mixed-integer linear programs (MILP) in a model. Focus is on problems that can be effectively solved with MILP and rules that can be accurately expressed with linear constraints. Such as (Pawlak & Krawiec, 2017), this paper focuses on the automated modeling of business rules for MILP. The problem setting is related to constraint programming (CP), parametric Markov decision processes, and inverse optimization. The aim of CP is to define constraints based on solution examples (Kolb s., 2016).

There are about five logistics parties that are working in the market spread in the physical distribution industry. The first stage is that production and transport were done by one producer, then some of these activities are divided into businesses that terms of insourcing and outsourcing, which is given to 3rd party logistics providers. At the beginning of third-party logistics in the 1980s, some companies have set up to provide services functions both transport and storage. These companies basically lifted a certain burden from manufacturing or producing companies related to supply chain management. Accordingly, it has correlated with a number of companies and a number of trucking carriers, which are using third-party logistics services (KIM, Studies on Total Logistics Management in Physical Distribution Process, 2019). Mike Hewitt and Emma Frejinger focused on optimization problems that are MILP. Two inputs are used: (1) MILP model to solve, which referred to as GOP, and, (2) historical data regarding past decisions implemented. Mathematical representations of business rules that are not present in GOP are learned by the framework, and the historical data indicates impacted actual decision making. The problem of learning mathematical representations of business rules from historical data is treated as a statistical learning problem and it expands GOP with new limitations, which is called the adaptive optimization problem (AOP). So, there is an algorithmic framework that combines optimization and statistical learning for automating the adaptation of classical optimization models to specific operational contexts.

## 3-General framework

This section presents GOP that will be adapted and shows how to learn mathematical representations of business rules from historical data. This GOP is MILP widely used in operational planning. So, GOP is:

$$\text{maximize } \mathrm{r}^{\mathrm{t}} x$$

$$\text{s.t.}$$

$$A^t x \leq b^t \qquad x \in X \tag{1}$$

$$G^t x + H^t v \leq h^t \qquad v \in V \tag{2}$$

Some of the business rules' models are not represented in GOP but still be observed by an actionable plan. Such as some business rules are often modeled with auxiliary variables such as binary variables for precedence constraints in scheduling. Parameters are for a time period $t = 1, \dots\dots, T$ , and $y^t$ and $x^t$ are plans executed in each of those time periods and an optimal solution to GOP, respectively. The methodology is to yield an optimization problem that focuses on learning linear representations.

The function $y_i = f\left(x; \alpha^i, \beta^i\right)$ is an affine function and regression equation is defined:

$$y_i = \beta^i + \sum_{j=1}^{n} \alpha_j^i x_j \quad + \varepsilon^i \qquad i = 1, \dots., n \tag{3}$$

Where $\beta^i \in \mathbb{R}$ are intercepts, $\alpha_j^i \in \mathbb{R}$ are parameters and $\varepsilon^i$ are Gaussian noise and they are independent. Training yields the parameter estimates $\hat{\beta}$ and $\hat{\alpha}$. $\hat{\beta}$ and $\hat{\alpha}$ are used to create an AOP by adding to the GOP three sets of decision variables. The first, $y_i$ are the predictions of the executed plan given a solution to GOP. The second, $\delta i \geq 0$ , $i = 1, \dots, n$ measures the difference between those plans with respect to element $i$. Finally, the third, $\Delta$, measures the total difference between these plans.

$$\hat{\beta}^i + \sum_{j=1}^{n} \hat{\alpha}_j^i x_j \quad + \varepsilon^i \; = y_i \tag{4}$$

$$\delta_i \geq y_i \; - x_i \tag{5}$$

$$\delta_i \; \geq x_i - y_i \tag{6}$$

$$\Delta = \sum_{i=1}^{n} \delta_i \tag{7}$$

As mentioned, a third-party logistics (3PL) company is selected and the company's goal is to maximize revenue and the provider tries to determine which customers' products to load into a single container. The 3PL has the ability to formulate and solve a different type of linear knapsack problem, KP (t):

$$Z_{KP}^t = maximize \sum_{i=1}^{n} r_i^t x_i$$

$$s.t.$$

$$\sum_{i=1}^{n} a_i^t x_i \; \leq b^t$$

$$0 \leq x_i \leq 1$$

We assume that the revenue is $r_i$ , and size is $a_i$ , associated with customer $i$'s products, the capacity is $b$, and t refers to the day. These business rules may be satisfied in an executable plan. Table 1 provides representations of these rules.

**Table 1.** Business rules for linear Knapsack problem

| Rule | Logical/Mathematical constraint | Functional mapping |
|---|---|---|
| 1 | $x_1 \geq x_2$ | $f_1: \bar{y}_i = \begin{cases} \max(\bar{x}_1, \bar{x}_2) & i = 1 \\ \bar{x}_i & i = 2,3, \dots, n \end{cases}$ |
| 2 | $a_1^t x_1 + a_2^t x_2 \leq 0.4b^2$ | $f_2: \bar{y}_i = \begin{cases} \max\left(0, \frac{0.4b^t - a_2\bar{y}_2}{a_1}\right) & i = 1 \\ \min\left(1, \frac{0.4b^t}{a_2}\right) & i = 2 \\ \bar{x}_i & i = 3, \dots, n \end{cases}$ |

AOP version of KP(t) is called A-KP (c, t). To formulate A-KP (c, t), we fit the following regression equation for each $i = 1, \dots, n$ :

$$\beta^i + \sum_{j=1}^{n} \alpha_j^i \, x_j + \varepsilon^i = y_i$$

A-KP (c, t):

$$Z_{KP}^t = maximize \sum_{i=1}^{n} r_i^t x_i + c\Delta$$

$$s.t.$$

$$\sum_{i=1}^{n} a_i^t x_i \; \leq b^t$$

$$0 \leq x_i \leq 1$$

## 4-Data generation

We divide the data into two sets that they are training and test sets. We use a training set to learn the linear representation of a mapping and a test set is used to evaluate the effectiveness of learning mathematical objects. We use sales data for two days and two customers (two cities) as a training set. Table 2 presents a data set.

**Table 2.** Data set for the linear knapsack problem

| | | Day1 | Day 2 |
|---|---|---|---|
| Customer 1 | Size | 180 | 150 |
| | Revenue (Rials) | 23499300 | 75677500 |
| Customer 2 | Size | 125 | 160 |
| | Revenue (Rials) | 138314800 | 90459930 |

## 5-Results

This section validates the approach in the regular setting. Recall this setting consists of small instances of GOP and the application of the same single business rule each day. It is observed that for small values of c, the optimal solution to A-KP (c, t) nearly always satisfies the constraints $G^t x + H^t v \leq h^t$ . Finally, we see that with respect to satisfying the business rule, the linear mapping approach works well on the linear knapsack problem.

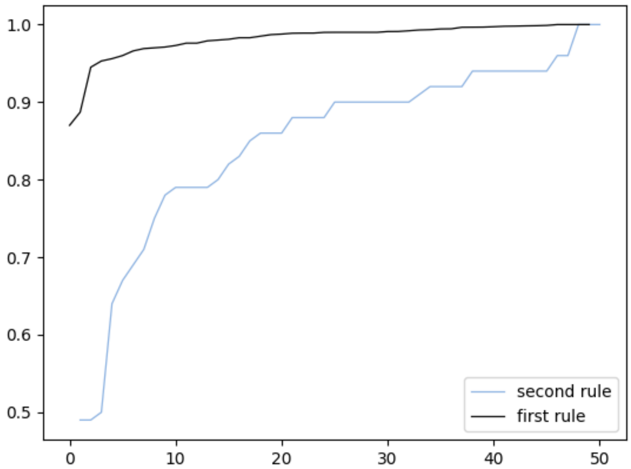


**Fig 1.** Linear knapsack problem: percentage of days in test data set where business rule is satisfied

Finally, we understand that this system helps product shippers plan and execute transportation moves, usually by third-party carriers, in order to deliver their products. We assessed the performance of the framework on a class of optimization problems. Our experiment showed that this problem has satisfied.

**Acknowledgment**

We acknowledge Digikala for providing sales data and assistance in preparing this paper.

## References

Hewitt, M., & Frejinger, E. (2020). Data-driven optimization model customization. *European Journal of Operational Research, 287*, 438-451.

KIM, J.-H. (2019). Studies on Total Logistics Management in Physical Distribution Process. *East Asian Journal of Business Economics, 7*(4), 15-26. doi:http://dx.doi.org/10.20498/eajbe.2019.7.4.15

Kolb, s. (2016). Learning Constraints and Optimization Criteria. *Technical Report . The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence* (pp. 403-409). Phoenix, Arizona, USA: AAAI Press.

Kolb, S., Teso, S., Passerini, A., & De Raedt. (2018). Learning SMT(LRA) Constraints using SMT Solvers. *In Proceedings of the twenty-seventh international joint conference on artificial intelligence* (pp. 2333-2340). Stockholm Sweden: AAAI Press. doi:https://doi.org/10.24963/ijcai.2018/323

Lombardi, M., Milano, M., & Bartolini, A. (2017). Empirical decision model learning. *Artificial Intelligence, 224*, 343-367. doi:https://doi.org/10.1016/j.artint.2016.01.005

Pawlak, T., & Krawiec, K. (2017). Automatic synthesis of constraints from examples using mixed integer linear programming. *European Journal of Operational Research, 261*(3), 1141-1157. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S037722171730156X