A Surrogate-based Optimization for a Mega-Hub location problem in South Korea using Artificial Neural Network

Junghyun Kim¹, Kisun Song¹, Changyun Chung², and Dimitri Mavris¹

¹Aerospace Systems Design Laboratory, Georgia Institute of Technology, Atlanta, Georgia, United States
²Network Optimization Team, CJ Logistics, Seoul, South Korea

Abstract

In recent era, the urbanization and the growth of E-Commerce have led to a significant increase in parcel delivery service network complexity. Especially, South Korea is an exemplary market where next-day parcel delivery service is so dominant that various companies are committed to improve the current service networks. Although many companies in South Korea strive to minimize the amount of time for delivery, it is seemingly inevitable to confront network traffic saturation issues occurring in various fields such as limitation of hub capacity. In response to these concerns, the companies have started to consider opening a new mega hub in order to ensure timely delivery service. In this paper, we propose a surrogate-based optimization methodology employing state-of-the-art surrogate models such as Artificial Neural Network for a new mega hub location problem in a complex next-day parcel delivery service network in South Korea. This approach is to utilize the synergism of the collaboration between supervised learning techniques and advanced engineering design methods such as Design of Experiment and Monte-Carlo Simulation. Finally, we found the optimum location of new mega hub in South Korea where transportation costs were reduced by approximately 14% compared to the current hub network system operated by CJ Logistics.

Keywords: Logistics, Transportation network optimization, Hub location, Surrogate model, Artificial Neural Network

1. Introduction

In recent years, the emergence of ‘super-connected’ society due to the advancement of communication technology has significantly encouraged the E-Commerce companies to aggressively expand their parcel delivery service networks. Especially, South Korea is an exemplary market where it is so predominant that various companies and institutions are committed to improve the current service networks as much as possible. For example, CJ Logistics, which is one of the best parcel service providers in South Korea, operates 12 hubs and 270 subs with 18,000 vehicles for accurate and timely delivery of maximum 5.28 million boxes in a day. [1] To that end, however, it is seemingly inevitable to confront network traffic saturation issues occurring in various fields such as hub capacity, limitation of transportation system, non-optimal logistics, and etc. To add insult injury, the annual report in 2017 [2] projects parcel supply in South Korea continues to grow whilst average price for a box tends to be decreased as shown in Figure 1.

![Figure 1: Year-of-year rate of change of parcel supply and average price in South Korea](image_url)
In response to these concerns, many companies in South Korea have started to consider opening a new mega hub in order to ensure timely delivery service. This effort naturally caused the following research question ‘Where should the mega hub be located to minimize transportation costs?’

In fact, the first paper in term of hub location problems was addressed by Goldman in 1969. [3] After Goldman, O’Kelly presented the first recognized mathematical formulation for a hub problem by studying airline passenger networks in 1987. [4] Since then, the problem of hub location had attracted many researchers. For example, an important practical motivation for hub location research was the advent of express delivery firms, most notably Federal Express (FedEX). In spite of the prevalence of postal and freight transportation networks, the advent of FedEX in 1973 marked a turning point in the recognition of the value provided by hub networks. [5]

However, it was well known that the exact solution for these problems is mathematically limited because all of the classical hub location problems are Nondeterministic Polynomial time (NP) problem. [6] For this reason, stochastic optimization and heuristic methods have been recently received increasing attention among researchers. For example, Sim T, Lowe T, and Thomas B [7] provided a single assignment hub covering model where the arc travel time is normally distributed with a given mean and standard deviation. Also, Yang T-H [8] addressed uncertainty in demand via a stochastic two-stage model with multiple demand scenarios. Contreras I, Cordeau J-F, and Laporte G [9] considered a more straightforward variant of the cost minimizing problem with stochastic demands and transportation costs.

These have certainly cast light on strategies to improve next-day parcel delivery network. However, it is yet speculative at best and even hard to conclude that the direct or stochastic optimization is always the best leverage regardless of the network complexity that exponentially escalates as the nodes or segments increases. In this paper, we propose a surrogate-based network optimization methodology employing state-of-the-art surrogate models such as Artificial Neural Networks (ANN) of a new mega-hub location problem in a complex ‘next-day’ parcel delivery service network in South Korea. This approach is to utilize the synergism of the collaboration between supervised machine learning techniques and advanced engineering design methods such as Design of Experiment (DoE) and Monte-Carlo Simulation (MCS).

2. Methodology

2.1 Problem definition

A hub is generally defined as a special facility that serves as switching, transshipment, and sorting points in many-to-many distribution systems. The hub location problem is typically concerned with locating hub facilities in order to route the traffic between origin-destination pairs. [9] Instead of connecting origin-destination pairs directly, namely Point-to-Point distribution system, CJ logistics has mainly used Hub-and-Spoke distribution system where hub facilities concentrate flows to take advantage of economic benefit. For this reason, we assumed that there is a given network with a number of nodes in South Korea where origins, destinations, and potential hub locations are specified. For this case study, the following data were given from CJ Logistics: 1) Sub location and supply information, 2) Hub location and capacity information, 3) Real distance and cost information for each route, and 4) Truck carrying capacity information. For security concerns, the supply distribution information shown in Figure 2 was normalized.

Figure 2: Sub/Hub locations and Parcel supply distribution at the busiest day
2.2 Design of Experiment (DoE)

A DoE is defined as a procedure that selects a set of samples in the design space to maximize the amount of information with a limited set of experiments. [10] For this case study, Central Composite Design (CCD), one of the most commonly used structured-based DoE, was used to capture corner points of design space. In addition, Latin Hypercube Sampling (LHS) was used to capture inner points of design space. The sample points are shown in Figure 3.

![Design space with 200 potential hub locations](image)

Figure 3: Design space with 200 potential hub locations

2.3 Modeling and Simulation

The Transportation Problem (TP) is a special type of Linear Programming (LP) where the objective is to minimize the cost related to distribution of supply from origins to destinations. The TP formulated in this paper is as following:

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\
\text{subject to:} & \quad x_{ij} \geq 0 \quad (i = 1,2,\cdots,m; j = 1,2,\cdots,n) \\
& \quad \text{supply constraint: } \sum_{j=1}^{n} x_{ij} = s_i \\
& \quad \text{demand constraint: } \sum_{i=1}^{m} x_{ij} = d_i
\end{align*}
\]

To implement and solve the TP model, we used a Python open source package called as PuLP. [11] Then, we slightly modified the TP algorithm to reduce the gap between simulation and the reality. For instance, since parcel supplies are always larger than demands in delivery service in South Korea, a dummy variable was introduced to change the problem from unbalanced to balanced Transportation Problem. In addition, in order to consider impossible routes in the reality, the Big-M method was implemented into the TP algorithm as shown in Figure 4. Lastly, a variety of constraints considering the real operation in parcel delivery service at CJ logistics were implemented to the algorithm. The modified TP algorithm was then run for 200 sample points to generate results for fitting a surrogate model.

![The schematic of the TP algorithm used in this paper](image)

Figure 4: The schematic of the TP algorithm used in this paper
2.4 Surrogate modeling

In general, a surrogate model is defined as a multi-variate regression technique developed to model the response of a complex system. [12] For this case study, the surrogate modeling technique was used to predict capability of the original model in the design space whilst reducing computational time. In particular, Artificial Neural Network (ANN) based surrogate model, which is inspired by the structure of the human brain and is constructed by complex connections between the neurons, was used for this study. The diagram of ANN based surrogate model used in this study is shown in Figure 5. For the ANN model, we specified two hidden layers and eight neurons for each layer. For the hidden layer structures, the scaled version of the sigmoid function was used.

![Figure 5: Diagram of the ANN model](image)

In order to evaluate the accuracy of the surrogate model, a goodness of fit was performed with $R^2$, Actual by Predicted plot, and Residual by Predicted plot. Furthermore, K-Fold method was implemented in the ANN model in order to address over-fitting issues of the model. To be more specific, $R^2$ results, which describe how well the model predictions adhere to reality, were calculated for both training and validation datasets. ($R^2 = 0.99$) Second, Actual by Predicted plot was generated to determine how the regressed equation is sufficiently modeling the behavior of the supplied data. As shown left in Figure 6, it was observed that the data points were evenly scattered along the perfect line. In addition, Residual by Predicted plot shown right-top in Figure 6 shows the absolute errors associated with the assumed model for each predicted value were quite close to actual values. Finally, errors of the surrogate model were calculated by examining the distribution of the error. As a result, it was found that the absolute maximum error for the model was approximately 0.2% as shown right-bottom in Figure 6.

![Figure 6: Surrogate model validation results](image)

2.5 Monte-Carlo Simulation (MCS)

In general, MCS is defined as an algorithm that conduct random sampling repeatably to obtain output results. The MCS is usually used to understand the impact of risk and uncertainty; however, we used the MCS for mapping purpose of input and output in this study to see the trend of resulting outcomes. In order to determine lower and upper limits of
the design variables, we considered the reality to determine the design space and used uniform sampling distributions with respect to the min/max values for the MCS. Using the generated ANN based surrogate model, the MCS was performed with 100,000 points. The ANN based surrogate model enabled very rapid MCS runs of large number of samples. For this case study, it approximately took five seconds to populate 100,000 samples.

3. Results

After investigating the results from MCS, it was observed that: 1) the cost decreases as latitude increases. 2) the cost has a minimum point with respect to longitude. These trends are described in Figure 7. Based on the trends, we found the optimum location of the new mega hub in South Korea where transportation costs were reduced by approximately 14% compared to the current hub network system operated by CJ Logistics.

![Figure 7: Monte-Carlo Simulation results for the transportation costs w.r.t. longitude and latitude](image)

Although the computational results showed that the optimum location reduced transportation costs by approximately 14% compared to the current hub network system, it is still early to conclude that we found the optimum location. This is because it would be not possible to build the new mega-hub at the optimum location in reality. For instance, there would be already many people living on the location. For this reason, we analyzed and investigated a few design points close to the minimum. As a result, it was observed that there is a possible range to be an optimum, namely near-optimum solutions, where the cost difference is relatively small compared with the minimum point. For security concerns, we decided to not show the exact location; however, the notional description for the possible range and near-optimums is shown in Figure 8.

![Figure 8: Notional description of near-optimum solution for a mega-hub problem in South Korea](image)
4. Conclusion

In this paper, we proposed a surrogate-based optimization methodology with the synergism of the collaboration between supervised learning techniques and advanced engineering design methods to find an optimal location for a new mega-hub facility in South Korea. As the first step, the design space was specified by considering the real operations of CJ Logistics. Based on the design space, Design of Experiment was established with Central Composite Design and Latin Hypercube Sampling to capture not only corner but also inner points of the design space. For the Design of Experiment, 200 points were allocated as a potential mega hub location. For the sample points, the modified Transportation Problem algorithm was run to generate results for fitting a surrogate model. In particular, Artificial Neural Network was used as a surrogate model technique. To evaluate the surrogate model, goodness of fit was performed with R^2, Actual by Predicted, Residual by Predicted, and K-Fold method for the validation purpose. Using the generated surrogate model, Monte-Carlo Simulations were performed with 100,000 points to propagate the metric of interest. After investigating the results from Monte-Carlo Simulations, it was observed that: 1) the cost decreases as latitude increases. 2) the cost has a minimum point with respect to longitude. Based on the trends, we found the optimum location of the new mega hub in South Korea where transportation costs were reduced by approximately 14% compared to the current hub network system operated by CJ Logistics. Since it is difficult to conclude that the company builds the mega-hub at the optimum location in reality, a few design points close to the minimum were explored. As a result, it was observed that there is a possible range to be an optimum, namely near-optimum solutions, where the cost difference is relatively small compared with the minimum point.

Acknowledgements

We would like to thank Youngsoo Kim (Vice president, Consulting division, CJ Logistics) for his feedback on this study. We also would like to thank Byungdo Lee, Wansik Kim, and Junghoon Kim (Network Optimization Team, CJ Logistics) for their supports on data-gathering process.

References