

Constrained Market Share Maximization by Signal-Guided Optimization

Bo Hui¹, Yuchen Fang^{2*}, Tian Xia¹, Sarp Aykent¹, Wei-Shinn Ku¹

¹ Auburn University

² Beijing University of Posts and Telecommunications

¹{bohui,tianxia,sarp,weishinn}@auburn.edu, ²fangyuchen@bupt.edu.cn

Abstract

With the rapid development of the airline industry, maximizing the market share with a constrained budget is an urgent econometric problem for an airline. We investigate the problem by adjusting flight frequencies on different flight routes. Owing to the large search space of solutions and the difficulty of predicting the market, this problem is in general daunting to solve. This paper proposes a novel two-stage optimization method to address the challenges. On the higher level, we use a signal to guide the optimization process toward a constrained satisfying solution. On the lower level, we consider the consecutive itineraries in real scenarios and model the unseen correlations between routes in itineraries for market share prediction. In theory, we prove the convergence of our optimization approach. In the experiment, we empirically verify the superiority of both our prediction model and optimization approach over existing works with large-scale real-world data. Our code has been released at: <https://github.com/codingAndBS/AirlineMarket>.

Introduction

Since the first aircraft flight in 1903, air travel has emerged as a crucial means of transportation and the airline industry has been developing rapidly. There are over 41,700 airports all over the world. The number of flights performed globally by the airline industry reached 38.9 million in 2019. While there are considerable airlines in the market and finite market demands, maximizing the market share with a constrained budget by adjusting flight frequencies on different flight routes is an econometric problem for an airline (Li et al. 2021), where the decision maker sets a budget constraint to avoid excess of expenditure over income. Intuitively, shifting the flight frequencies and the cost on some routes to other routes will change the market share of an airline. This research problem is especially challenging since it not only requires constrained combinatorial optimization methods to search for the optimal solution but also involves market share prediction.

The first challenge lies in the large search space of solutions. To maximize its market share, an airline needs to decide the frequency of flights on each route. The constrained budget implies a solution set including all possible combinations of flight frequencies on different flight routes and

their corresponding costs are under the settled budget. Considering that there are many routes to be covered and the flight frequency on each route is an ordinal number, the optimization problem is daunting to solve. For instance, if there are n routes and the maximum frequency is f^{max} , we need to find an optimal solution from $(f^{max} + 1)^n$ options under constraint. Enumerating and evaluating all solutions on such a complicated network is prohibitive. Note that traditional combinatorial optimization algorithms (Papadimitriou and Steiglitz 1998) cannot be used to solve this problem, since the market share is nondeterministic, and two routes are usually not independent from each other. A heuristic-based algorithm, called GroupGreedy, is introduced in (An et al. 2016) to run on a small subset of routes. Unfortunately, the scalability of this work is far from satisfactory. Another work (Li et al. 2021) adopts the gradient ascent strategy for the solution where cost overruns the predefined budget. However, this work sets the optimization as a maximizing problem and solves the problem in only one space, where the constrained maximization is actually a max-min problem and involves two spaces. Therefore, the solution in (Li et al. 2021) may be trapped in a stationary point and the global optimal solution is not guaranteed.

Second, plenty of factors (e.g., average ticket price, flight frequency and on-time performance) will influence the air market. Intuitively, if the ticket price of an air carrier on a route decreases, the market share on the route will increase. Likewise, if the flight frequency increases, the market share will decrease. Furthermore, the market share on a flight route involves not only one airline but all airlines with flights on the route, which makes the market share prediction more complicated. The butterfly effect in economics refers to the compounding impact of small changes and states that it is nearly impossible to make accurate predictions for the future and identify the precise cause of a specific change in the air market. Therefore, designing an accurate prediction model is demanding.

In the last few years, spatial dependency (Yu, Yin, and Zhu 2018; Hui et al. 2021b,a; Fang et al. 2021) has drawn attention in volume prediction in the transportation system. To capture the spatial and unseen correlations between any two flight routes, we use attention mechanism (Vaswani et al. 2017) to model the correlations among routes for market share prediction. Specifically, we transform the airport-wise directed graph into a line graph where the nodes are

*These authors contributed equally.

flight routes in the airport network. Then attention layers are stacked to learn the correlations, and the learned correlations are used to aggregate representations of other routes. Note that we consider not only the correlations between two adjacent routes but also the unseen dependency between any two non-adjacent routes. With our design, we are able to reveal the potential relationships between any two routes and model the complicated economic effects.

To address the challenge of a large searching space of solutions under constraint, we formulate the optimization process as a min-max problem. In the problem of market share maximization, the task is to maximize total market share on all routes while minimizing the cost of the solution under an inequality constraint regarding the budget. Constrained optimization is usually solved using the Lagrange relaxation technique (Bertsekas 1997) where the optimization is converted into an equivalent unconstrained problem. Then a penalty term is added to make infeasible solutions sub-optimal. Inspired by actor-critic algorithms (Schulman et al. 2015; Mnih et al. 2016) in deep reinforcement learning, we consider the distribution of frequencies on different routes as the actor and use TD-learning critic to estimate the penalized market share. We propose an optimization algorithm to alternatively update the actor and Lagrange multiplier with the gradient which is estimated using the log-likelihood trick (Williams 1992). We theoretically prove that our algorithm converges almost to a constrained solution and demonstrates fast convergence. In the experiment, we verify the superiority of our optimization algorithm over existing works on real-world datasets.

The contributions of this paper can be summarized as:

- We propose a constrained optimization algorithm to incorporate the constraint as a penalty signal into the market share maximization, where the signal guides the optimization towards a constrained solution.
- We theoretically prove that our algorithm converges almost to a constrained solution.
- We utilize the connectivity information to pre-train an accurate market share prediction model.
- We empirically verify the superiority of our optimization algorithm over existing works and the effectiveness of our market share prediction model.

Preliminary

Airport Network

On the airport network, airports are connected by many flight routes. Different airlines schedule different flight frequencies on these routes under different conditions to maximize profit. Mathematically, the airport network for airline k is defined as: $\mathcal{G}_k = (\mathcal{V}, \mathcal{R})$, where \mathcal{V} and \mathcal{R} indicate the set of airports and flight routes. We use f_k to represent the flight frequencies of airline k on all routes, *i.e.*, $f_{k,r}$ is the frequency of airline k on route r . Besides the flight frequency, many other features are associated with routes and airlines, such as ticket price, occupancy rate, and delay ratio. We use $X_k \in \mathbb{R}^{|\mathcal{R}| \times F}$ to represent the features of airline k , where $|\mathcal{R}|$ and F denote the number of routes and the dimension of features, respectively.

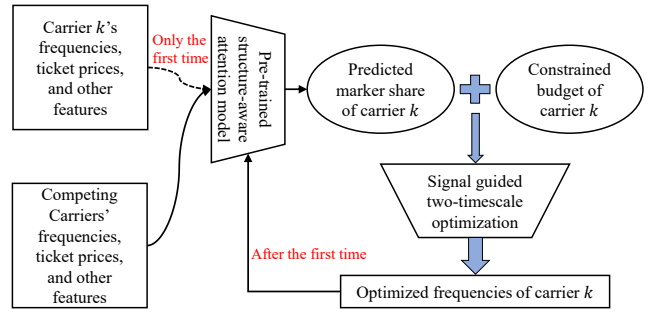


Figure 1: Framework Overview

Problem Definition

Given \mathcal{G}_k and X_k for airline k , the target is to maximize the overall market share by adjusting flight frequencies f_k on all flight routes under the constrained budget. Formally, the problem can be defined as:

$$\begin{aligned} \max_{f_r^{max} \geq f_{r,k} \geq 0} \quad & \sum_{r \in \mathcal{R}} d_r \times m_{r,k} \\ \text{s.t.} \quad & \sum_{r \in \mathcal{R}} c_{r,k} \times f_{r,k} \leq b_k \end{aligned} \quad (1)$$

where f_r^{max} and b_k are the observed maximum frequency on flight route r and the budget of airline k . Due to the limited number of tracks and finite boarding time, the frequency is constrained by a maximum number on the route. Moreover, to avoid excess offset by the decision maker to avoid excess expenditure over income, the overall cost should not be larger than the budget. Therefore, we consider the problem as a constrained optimization problem under two constraints: $f_r^{max} \geq f_{r,k} \geq 0$ and $c_{r,k} \times f_{r,k} \leq b_k$, where $f_{r,k}$ is the flight frequency to be optimized on route r for airline k . We use d_r , $c_{r,k}$, and $m_{r,k}$ to denote the passenger demand for Route r , the cost per flight on Route r for airline k , and the market share of airline k on route r . The market share $m_{r,k}$ is estimated by a pre-trained prediction model.

Architecture

Figure 1 shows the framework of our method. We first train a neural network model to predict the market share for airline k with the features and flight frequencies. Given an input regarding flight frequencies on all flight routes for airline k , we can estimate the market share with the pre-trained model. Then the key step is to find the optimal flight frequencies which can maximize the market share for airline k under a constrained budget. We use a signal-guided optimization algorithm to search for a constrained solution.

Methodology

Parametrized Frequencies

In this paper, we consider parametrized frequencies $f_k = \{f_{r,k} | r \in \mathcal{R}\}$ as learnable parameters in a pre-trained neural network model to maximize the market share of airline k through the updating of gradient, which is different from existing black-box search methods. Specifically, we use $\Theta(\cdot)$

to denote the pre-trained neural network. The market share can be estimated as:

$$m_{r,k} = \Theta(f_{r,k}, \cdot) \quad (2)$$

In the following sections, we will describe how to maximize $m_{r,k}$ and how to learn the pre-trained model $\Theta(\cdot)$.

Signal-Guided Two-Timescale Optimization

In the market share maximization problem, frequencies are integer values. We relax flight frequencies to real numbers during the maximization and round down the real numbers to integers at the end of the maximization process, which is similar to other continuous relaxation techniques (Pardalos, Prokopyev, and Busygin 2006). In the literature, the Lagrange multiplier has been adopted to maximize concave functions (or some special non-concave functions) with constraints (Boyd et al. 2011). However, the Lagrange multiplier is not suited for our work because of the high nonlinearity of our objective function in market share prediction based on a neural network. Therefore, we incorporate the budget constraint as a penalty signal into the market share maximization function to guide flight frequencies toward a constrained solution. We derive the objective Lagrange function as:

$$\mathcal{L}(\lambda, f_k) = o(f_k) - \lambda c(f_k) \quad (3)$$

where λ , $o(\cdot)$, and $c(\cdot)$ denote the Lagrange multiplier, market share maximization function and budget constraint, respectively. More specifically, $o(\cdot)$ and $c(\cdot)$ are defined as:

$$\begin{aligned} o(f_k) &= \sum_{r \in \mathcal{R}} d_r \times m_{r,k} \\ c(f_k) &= \sum_{r \in \mathcal{R}} c_{r,k} \times f_{r,k} - b_k \end{aligned} \quad (4)$$

Previous methods manually set the Lagrange multiplier to find the optimal solution to the original constrained problem because their objective functions are simple. In contrast, our objective function is more intricate since our market share prediction function is a complex neural networks model. Therefore, we convert the original constrained problem into an equivalent unconstrained problem and utilize a two-timescale method to solve the unconstrained problem, which can be formulated as:

$$\begin{aligned} \min_{\lambda} \max_{f_{r,k}} \mathcal{L}(\lambda, f_k) &= \min_{\lambda} \max_{f_{r,k}} (o(f_k) - \lambda c(f_k)) \\ \text{s.t. } f_r^{\max} &\geq f_{r,k} \geq 0 \text{ and } \lambda \geq 0 \end{aligned} \quad (5)$$

where L is the Lagrange function and λ is the Lagrange multiplier (a coefficient of the penalty signal). Notice that, as λ increases, the solution to Eq. (5) converges to that of Eq. (1). In our two-timescale optimization method, the flight frequencies are optimized by solving Eq. (5) on the faster timescale and λ is increased until the constraint is satisfied on the slower timescale. The target in the two-timescale method is to find a saddle point $(\lambda^*, f_k^*(\lambda^*))$, which is a feasible solution.

Algorithm 1: Signal-guided two-timescale Optimization

Input: demand d_k on each route, budget b_k for airline k , cost c_k for each route of airline k , and pre-trained prediction model $\Theta(\cdot)$

Output: flight frequencies $f_k^{(i+1)}$

- 1: Initialize: $f_k^{(0)} = f_{k, \text{freq}}$
 - 2: $\lambda^{(0)} = 0$
 - 3: **while** not convergence **do**
 - 4: $m_k^{(i)} = \Theta(f_{r,k}, \cdot)$
 - 5: $o(f_k) = \sum_{r \in \mathcal{R}} d_r \times m_{r,k}^{(i)}$;
 - 6: $c(f_k) = \sum_{r \in \mathcal{R}} c_{r,k} \times f_{r,k}^{(i)} - b_k$;
 - 7: $f_k^{(i+1)} = \Gamma_f(f_k^{(i)} + \eta_2^{(i)} \nabla_f \log(o(f_k^{(i)}) - \lambda^{(i)} c(f_k^{(i)})))$;
 - 8: $\lambda^{(i+1)} \leftarrow \Gamma_\lambda(\lambda^{(i)} + \eta_1^{(i)} c(f_k^{(i)}))$
 - 9: Return $f_k^{(i+1)}$
-

Estimating the Gradient

When we utilize the two-timescale optimization method to maximize the market share, the Lagrange multiplier and frequencies are updated as follows:

$$\begin{aligned} \lambda^{(i+1)} &= \Gamma_\lambda(\lambda^{(i)} - \eta_1^{(i)} \nabla_\lambda \mathcal{L}(\lambda^{(i)}, f_k^{(i)})) \\ f_k^{(i+1)} &= \Gamma_f(f_k^{(i)} + \eta_2^{(i)} \nabla_f \mathcal{L}(\lambda^{(i)}, f_k^{(i)})) \end{aligned} \quad (6)$$

where Γ_λ and Γ_f are projection operators. To keep the iterate frequencies stable, we project the frequency solution onto a compact and convex set and project the Lagrange multiplier into the range $[0, \infty]$. The value of step-size $\eta_1^{(i)}$ and $\eta_2^{(i)}$ ensure that the frequency update is performed on a faster timescale than that of the coefficient λ of the penalty signal. The gradients $\nabla_\lambda \mathcal{L}$ and $\nabla_f \mathcal{L}$ are derived from Eq. (5), where the formulation for $\nabla_f \mathcal{L}$ is obtained using the log-likelihood trick. Formally, two gradients can be calculated by:

$$\begin{aligned} \nabla_\lambda \mathcal{L}(\lambda, f_k) &= -c(f_k) \\ \nabla_f \mathcal{L}(\lambda, f_k) &= \nabla_f \log(o(f_k) - \lambda c(f_k)) \end{aligned} \quad (7)$$

Algorithm 1 illustrates the details of our two-timescale optimization process.

Assumption 1.

$$\begin{aligned} \sum_{i=0}^{\infty} \eta_1^{(i)} &= \sum_{i=0}^{\infty} \eta_2^{(i)} = \infty, \\ \sum_{i=0}^{\infty} (\eta_1^{(i)^2} + \eta_2^{(i)^2}) &< \infty, \\ \text{and } \frac{\eta_1^{(i)}}{\eta_2^{(i)}} &\rightarrow 0. \end{aligned} \quad (8)$$

Theorem 1. The iterates $(\lambda^{(n)}, f_k^{(n)})$ converge to a fixed point almost surely under Assumption 1 as well as the standard stability assumption for the iterates and bounded noise (Borkar 2009).

Proof. To prove our optimization algorithm will converge to a saddle point, we prove the convergence of f_k and λ respectively:

- **Convergence of f_k :** The f_k -recursion tracks an ODE in the asymptotic limit for any λ on the slowest timescale because f_k is stable according to the projection. Specifically, the value of λ is a constant in the frequency updating due to the timescale separation, thus the following ODE governs the evolution of f_k :

$$\dot{f}_k^t = \Gamma_f(\nabla_f \mathcal{L}(\lambda, f_k^t)) \quad (9)$$

where Γ_f is a projection operator which ensures that the evolution of the ODE stays within the compact and convex set. Therefore, the rule of f_k in Eq. (6) can be seen as a discretization of the ODE in Eq. (9). Finally, the convergence of f_k can be concluded by the standard stochastic approximation arguments from (Borkar 2009).

- **Convergence of λ :** We first show the convergence of λ and then show that the whole process converges to a local saddle point of $\mathcal{L}(\lambda, f_k)$. As with f_k , the following ODE governs the evolution of λ :

$$\dot{\lambda}^t = \Gamma_\lambda(\nabla_\lambda \mathcal{L}(\lambda^t, f_k(\lambda^t))) \quad (10)$$

where $f_k(\lambda^t)$ is the limiting point of the f_k -recursion corresponding to λ^t . As shown in Chapter 6 of (Borkar 2009), $(\lambda^{(n)}, f_k^{(n)})$ converges to the internally chain transitive invariant sets of the ODE in Eq. (19). According to Theorem 2 of (Borkar 2009), $\lambda^{(n)} \rightarrow \lambda^*$ and $f_k^{(n)} \rightarrow f_k^*(\lambda^*)$, which completes the proof.

Lemma 1. The fixed point λ^* and $f_k^*(\lambda^*)$ in Theorem 1 is a feasible solution.

Proof. The first order methods such as gradient descent converge almost surely to a local minimum (Panageas, Piliouras, and Wang 2019) and the local minima frequencies satisfy the budget constraints. Hence for unbounded Lagrange multiplier λ , the process converges to a fixed point $(\lambda^*, f_k^*(\lambda^*))$ is a feasible solution.

Pre-trained Market Share Prediction Model

We first train a neural network model to accurately forecast the market share of airlines with the given features and flight frequencies. Since both the flight frequencies and the other features (e.g., ticket price, on-time performance) have influences on the market share, we utilize a fully connected layer to extract key information from features, which can be formed as:

$$h_{r,k}^0 = W^0(f_{r,k} \oplus X_{r,k}) + b^0 \quad (11)$$

where $W^0 \in \mathbb{R}^{F \times d}$ and $b^0 \in \mathbb{R}^d$ are the learnable weights and bias of the fully connected layer. We concatenate the flight frequency $f_{r,k}$ with other features as the final input feature $f_{r,k} \oplus X_{r,k}$.

We also consider the hidden correlations between flight routes. First, the connected structure of routes in the airport network is essential for predicting market share. Intuitively, if a route is adjacent to many other routes, this route can

be a part of the itinerary for many passengers from other routes. Moreover, due to the complicity of markets, there are many unrevealed relations between any two routes. For example, suppose a travel itinerary covers a list of sequential cities (Los Angeles, New York, Philadelphia, and Miami). In that case, people may choose to drive from New York to Philadelphia since the distance between these two cities is short. In this case, though the flight route from Los Angeles to New York and that from Philadelphia to Miami are not adjacent, there is a flow between these two routes. Therefore, we elaborately design a structure-aware attention layer to auxiliary forecast market share by simultaneously aggregating feature and structure-based spatial information from other routes. Our structure-aware attention layer is a stack of multiple basic units, where the l^{th} unit can be formulated as:

$$\tilde{h}_{r,k}^l = \sum_{i \in \mathcal{R}} \alpha_{r,i,k}^l (W^{l_v} h_{i,k}^{l-1}) \quad (12)$$

where $W^{l_v} \in \mathbb{R}^{d \times d}$ is the learnable parameter. $\alpha_{r,i,k}^l$ denotes the correlation between route r and route i for airline k in the airport network, which is learned from the latent representation and the connection degree between two routes:

$$\alpha_{r,i,k}^l = \frac{\exp((W^{l_q} h_{r,k}^{l-1})^T (W^{l_k} h_{i,k}^{l-1}))}{\sum_{j \in \mathcal{R}} \exp((W^{l_q} h_{r,k}^{l-1})^T (W^{l_k} h_{j,k}^{l-1}) / \sqrt{d})} + \beta_{r,i,k}. \quad (13)$$

We use $W^{l_q}, W^{l_k} \in \mathbb{R}^{d \times d}$ to denote the learnable parameters. In Eq. (13), the first item shows semantic similarities between the two routes. With the second term, we are able to leverage structure-based spatial correlations. Specifically, $\beta_{r,i,k}^l$ denotes the prior structure knowledge, *i.e.*, the connection degree between route r and route i in the airport network. We transform the original airport-wise directed graph \mathcal{G}_k into the corresponding line graph $\bar{\mathcal{G}}_k$, where the nodes $\bar{\mathcal{V}}$ of $\bar{\mathcal{G}}_k$ are the ordered edges in \mathcal{R} , *i.e.*, routes in the airport network \mathcal{G}_k are nodes in the line graph $\bar{\mathcal{G}}_k$. Then, an unweighted adjacency matrix $\mathcal{A}_k \in \mathbb{R}^{|\mathcal{R}| \times |\mathcal{R}|}$ is derived from the line graph. The weight is defined according to the adjacency between routes, *i.e.*, $\mathcal{A}_{k,r,i} = 1$ if the end airport of route r is the beginning airport of route i and 0 otherwise. To mine the connection degree among routes, we calculate the shortest path length between routes and build the connection degree matrix $\mathcal{B}_k \in \mathbb{R}^{|\mathcal{R}| \times |\mathcal{R}|}$ using the Gaussian kernel. The connection degree matrix can be formulated as:

$$\beta_{r,i,k} = \begin{cases} \exp(-\frac{\text{spd}(r,i,\bar{\mathcal{G}}_k)^2}{\sigma^2}) & \text{if route } r \text{ and } i \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

where $\text{spd}(\cdot)$ denotes the calculation of the shortest path length between two routes with the given indices over the line graph \mathcal{A}_k . Here, σ denotes the standard deviation of all pair-wise shortest path lengths.

We further add a non-linear feed-forward layer to enhance the representation power:

$$h_{r,k}^l = W^{l_2} \sigma(W^{l_1} \tilde{h}_{r,k}^l + b^{l_1}) b^{l_2} \quad (15)$$

where σ denotes the rectified linear unit and can eliminate the negative value.

To estimate the market share, we utilize a linear projection to predict the number of passengers of airline k on route r and further use a rectified linear unit to restrict illegal negative predicted values. Finally, we use the ratio of passengers that take airplanes of airline k to all passengers on route r as the market share of airline k on route r :

$$\hat{m}_{r,k} = \frac{\sigma(W^e h_{r,k}^L + b^e)}{\sum_i \sigma(W^e h_{r,i}^L + b^e)} \quad (16)$$

where $W^e \in \mathbb{R}^d$ and $b^e \in \mathbb{R}$ are the learnable weights and bias of the linear projection. Note that h_r^L is the output of L th stacked attention layer as shown in Eq. (15).

We use $\Theta(f_{r,k}, \cdot)$ to represent our prediction model and train our market share prediction model by adopting the \mathcal{L}_1 loss:

$$\mathcal{L}_1(\Theta) = \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} |m_{r,k} - \hat{m}_{r,k}| \quad (17)$$

where Θ indicates all learnable parameters.

Experiment

Dataset

We conduct extensive experiments on the airline origin and destination survey (DB1B) dataset¹ released by the US Department of Transportation’s Bureau of Transportation Statistics (BTS), which records the flight route information in the US and releases 10% of tickets sold in the US every month of the year for research purposes. Specifically, we collect the historical flight information from the DB1B dataset for the past ten years and select 1,000 flight routes with the most significant number of passengers in history to conduct experiments. We choose the top 3 airlines in the US domestic air markets to illustrate the prediction and optimization of the market share. From the prediction perspective, we aim to learn a neural network-based model by leveraging data of the top 3 airlines in the past ten years except for the last month’s data. The data of the last month is used for testing. Besides the flight frequencies of each airline, we utilize five other features that have a high correlation with the forecasting of market share: the ticket price, delay ratio, cancel ratio, divert ratio, and seat availability, *i.e.*, $X_{r,k}$ is a 5-dimensional vector. From the perspective of optimization, we utilize the pre-trained model and the input features (along with flight frequencies) in the last month to maximize the overall market share of each airline in the last month under the corresponding budget constraint.

Baselines

In this subsection, we introduce baselines for our optimization algorithm and the pre-trained market share prediction model.

To demonstrate the superiority of our signal-guided optimization method, we compare it with four baselines:

- **Brute Force:** In this method, we enumerate all feasible solutions under the observed maximum flight frequency and budget constraints.

- **Normal Greedy:** This method compares the marginal improvement in the market share of all strategies and selects the strategy that brings the biggest improvement. The complexity of this method is $O(\frac{b_k \times |\mathcal{R}|}{\sum_{r \in \mathcal{R}} c_{r,k} / |\mathcal{R}|})$.
- **Group Greedy:** This method randomly splits flight routes into multiple groups and then utilizes the normal greedy on these groups. The complexity of it is $O(\frac{b_k \times N_g}{\sum_g c_{g,k} / N_g})$, where N_g is the number of groups.
- **AGA (Li et al. 2021):** This is a gradient-based white-box optimization algorithm, which adds a regularization term to the original Lagrange function and solves the new function by a heuristic search method. However, it depends on the regularization term to derive the optimal form of the Lagrange multiplier and violates the actual condition.

We compare our proposed market share prediction method with several representative baselines:

- **XGBoost:** XGBoost is the widely used boosting-based standard regression algorithm.
- **RandomForest:** RandomForest is a well-known ensemble-based standard regression algorithm.
- **Model1 (Suzuki 2000a):** This is a famous multi-logit regression model of predicting flight market share, which considers multiple flight factors.
- **Model2 (Wei and Hansen 2005):** This is another multi-logit regression model of predicting flight market share, which further utilizes aircraft size and seat availability.
- **Model3 (Li et al. 2021):** This is a fully connected neural network model and further utilizes some network features such as degree centrality and page rank.

Configuration

In this paper, we test the group number of the group greedy in [5,10,20,40] for the top 1,000 routes and choose 20 in our paper under the trade-off of run-time and optimized performance. Moreover, the group number is 2 for the top 10 routes. We train our market share prediction model with the learning rate of 1e-3, epochs of 500, and the Adam optimizer for updating weights. The number of structure-aware attention layers in our model is 4 and the dimension of the hidden states in our model is 64. To prevent over-fitting for the data of last month, we use the 10-fold cross-validation method to choose the best hyper-parameter, which means we have nine 12-month folds and one 11-month fold. Following the standard cross-validation, we randomly select nine folds to train our model and utilize the remaining fold as the validation set to tune hyper-parameters and repeat this process nine times. The learnable parameters in our model are initialized by the Xavier initializer.

Metrics To evaluate the effectiveness and efficiency of our proposed optimization method and baselines, we set the optimized passengers in all flight routes of the airline and the run-time of optimization as metrics. The larger number of passengers and the smaller run-time indicate better performance. We adopt widely used metrics correlation coefficient

¹<https://www.transtats.bts.gov/DataIndex.asp>

Methods	Airline1 (\uparrow)	Airline2 (\uparrow)	Airline3 (\uparrow)
Ground Truth	25851	19553	19543
Brute Force	N/A	N/A	N/A
Normal Greedy	31016	20577	22890
Group Greedy	25936	20251	21474
AGA	27025	20609	20658
Ours	30885	22292	25727

Table 1: Optimization results for the top-10 routes.

Methods	Airline1 (\downarrow)	Airline2 (\downarrow)	Airline3 (\downarrow)
Brute Force	Timeout	Timeout	Timeout
Normal Greedy	47	34	42
Group Greedy	20	19	19
AGA	14	15	15
Ours	18	18	18

Table 2: Runtime in seconds for the top-10 routes.

Methods	Airline1 (\uparrow)	Airline2 (\uparrow)	Airline3 (\uparrow)
Ground Truth	501919	566868	448126
Brute Force	N/A	N/A	N/A
Normal Greedy	N/A	N/A	N/A
Group Greedy	929345	978315	746235
AGA	896992	961166	636136
Ours	1089126	1074594	873283

Table 3: Optimization results for the top-1,000 routes.

Methods	Airline1 (\downarrow)	Airline2 (\downarrow)	Airline3 (\downarrow)
Brute Force	Timeout	Timeout	Timeout
Normal Greedy	Timeout	Timeout	Timeout
Group Greedy	282	237	194
AGA	27	26	26
Ours	37	36	37

Table 4: Runtime in seconds for the top-1,000 routes.

(CC), coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) to evaluate the market share prediction. Note that (1) the larger value of CC and R^2 and (2) the smaller value of MAE and RMSE indicate better performance.

Constrained Optimization

We first compare our signal-guided optimization method with baselines on the top 10 routes. Specifically, we show the optimization results of market share for the top 3 airlines in Table 1. Note that the computational cost of brute-force search is highly burdensome, and the running time is prohibitively long. We mark its optimization result with "N/A". For all 3 airlines, our optimization outperforms another gradient-based baseline (AGA) by 2,750 on average. Our optimization method achieves the largest market share except for Airline1 where we observe a slight difference be-

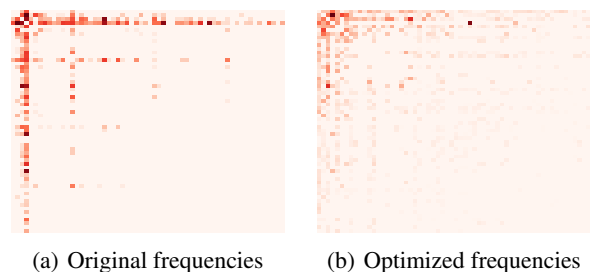


Figure 2: Visualizing the optimized frequencies.

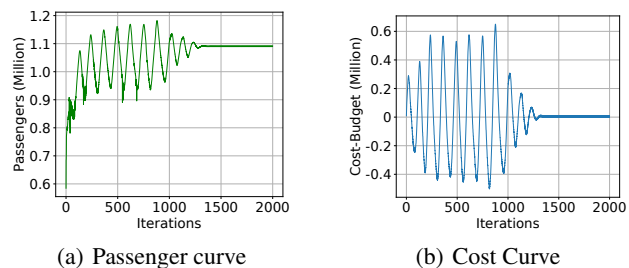


Figure 3: Visualizing convergence curves.

tween Normal Greedy and our method. Another indication of the effectiveness of the optimization methods is the runtime. As shown in Table 2, the run-time of our method is only around 50% of that with Normal Greedy.

We also investigate the effectiveness of our optimization method on top 1,000 routes in Table 3 and Table 4. As the number of routes increases, the size of the searching space will increase exponentially. In particular, optimization with Normal Greedy will result in a timeout on the top 1,000 routes. We find that our method consistently outperforms the baselines for all 3 airlines in terms of optimized market share. In addition, the maximum market share obtained by our method almost doubles the ground truth. It leaves plenty of room for improvement of strategies in these airlines.

Visualization We visualize the optimization process of flight frequency for top 1 airline with a heat-map. Figure 2 shows the difference between the original frequencies and the optimized frequencies. Note that each pixel denotes the frequency on a flight route between two airports. The darkness of the pixel represents the value of frequencies. The darker the color, the larger the frequency. We can observe that only a few pixels are darker than others. This means that only a few routes have larger frequencies under the budget constraint. However, after optimization, the darker pixels are distributed more evenly. By transferring the cost from some "hot" flight routes to other routes, the airline can increase its market share with a constrained budget.

Convergence Lastly, we verify the convergence of our optimization algorithm. As depicted in Figure 3, the total number of passengers who travel with the flights of the top 1 airline converges at about 1.1 million with our optimization algorithm. Note that it is the same airline in Figure 2. There-

Methods	10 routes				1,000 routes			
	CC (\uparrow)	R ² (\uparrow)	MAE (\downarrow)	RMSE (\downarrow)	CC (\uparrow)	R ² (\uparrow)	MAE (\downarrow)	RMSE (\downarrow)
XGBoost	0.9583	0.8644	0.0534	0.0758	0.9204	0.7769	0.0677	0.1358
RandomForest	0.9578	0.9124	0.0466	0.0643	0.9224	0.8743	0.0537	0.1332
Model1 (Suzuki 2000a)	0.9625	0.9231	0.0410	0.0579	0.9231	0.8757	0.0523	0.1309
Model2 (Wei and Hansen 2005)	0.9507	0.9086	0.0478	0.0665	0.9246	0.8763	0.0510	0.1294
Model3 (Li et al. 2021)	<u>0.9809</u>	<u>0.9608</u>	<u>0.0268</u>	<u>0.0430</u>	<u>0.9443</u>	<u>0.8864</u>	<u>0.0463</u>	<u>0.1220</u>
Model4 (Ours)	0.9881	0.9748	0.0220	0.0345	0.9613	0.9107	0.0360	0.1133
Model4 (No Att)	0.9814	0.9637	0.0255	0.0393	0.9463	0.8896	0.0446	0.1201
Model4 (No Prior)	0.9831	0.9667	0.0247	0.0386	0.9503	0.9023	0.0374	0.1148

Table 5: Market share prediction performance comparison. Bold: Best; underline: Second best.

fore, it further verifies the rationale of optimized frequencies in Figure 2(b). One advantage of our optimization method is that we adopt the Lagrange relaxation technique. We also visualize the difference between the cost and budget of the top 1 airline. It shows that the cost can be larger than the budget in the optimization process. However, we are able to confirm that the cost of our optimized frequencies is under budget constraints. It empirically demonstrates Theorem 1.

Market Share Prediction

Table 5 shows the performance comparison of our prediction model with the representative baseline models. It is clear that our method outperforms these baselines in terms of all performance metrics on both top 10 routes and top 1,000 routes. Specifically, our method can reduce about 21% RMSE and 18% MAE of the second performing baseline (Model3) on the top 10 routes. We also note that, when the number of routes is 1,000, the prediction performance will drop. This is because there are more complicated scenarios in 1,000 routes. However, our prediction accuracy is consistently higher than all baselines.

Ablation study: We also conduct an ablation study to validate the effectiveness of each component in model design. We first remove the attention coefficient (marked as "No Att" in the table). This variant only considers the direct connection (Equation (14)) between two airports. Without the attention coefficient in the prediction model, the prediction performance will drop in terms of all metrics. It verifies that our attention component can benefit the prediction accuracy. Our second variant (marked as "No Prior") removes $\beta_{r,i,k}$ in Equation (13). Likewise, removing this prior knowledge regarding connectivity will lead to performance degradation. It further verifies that the structure of the airport network plays a vital role in market share prediction.

Related Works

Constrained frequency optimization is NP-hard. In practice, NP-hard problems are usually solved using heuristics. Recently, deep learning has been introduced as a learning algorithm for NP-hard problems. For example, convolutional networks and deep reinforcement learning have been applied to master the famous Go (Silver et al. 2016, 2017). In terms of constrained optimization, greedy algorithms are most applicable to a large variety of learning settings (Lo-

catello et al. 2017). The stochastic gradient framework is also used to solve convex optimization problems with an infinite number of inclusion constraints that need to be satisfied (Fercoq et al. 2019). However, in the case where the inputs are integers, it is not clear how to adapt the updating steps. Also, optimistic gradient descent for constrained saddle-point optimization requires additional assumptions such as the uniqueness of the optimal solution (Wei et al. 2021). In (Lin, Ma, and Yang 2018), the finite-sum constrained convex optimization is studied and a new affine-minorized method is proposed to guarantee a feasible optimal solution path. Unfortunately, these combinatorial optimization algorithms cannot be used to solve the problem in this paper. This is because the input values in a solution are fixed and independent from each other in their setting.

A prior step of market share maximization is to predict market share. An early work (Suzuki 2000b) shows that there is a relationship between on-time performance and airline market share. Another work (Wei and Hansen 2005) studies the effect of other features including aircraft size, seat availability, price, and frequency. To forecast market share, (An et al. 2016) proposes an ensemble forecasting architecture considering two new types of features: features derived from clusters of similar routes and features based on equilibrium pricing. A later work (An et al. 2017) further considers frequency constraints and long-term profits. Li et al.'s work (Li et al. 2021) is the most germane to our paper. It also builds a neural network model to predict the market share. However, it ignores the hidden relation between two routes and fails to achieve satisfying prediction accuracy. Worse still, it uses an optimization algorithm based on the gradient to seek optimal flight frequency. This strategy may get trapped at a stationary point in the landscape.

Conclusion

In this paper, we propose an optimization method to maximize an airline's market share under constraints with a pre-trained market share prediction model. We incorporate the budget constraint as a penalty signal into the market share maximization function to guide flight frequencies towards a constrained solution. In the experiment, we empirically verify the superiority of both our signal-guided optimization method and the pre-trained market share prediction model.

References

- An, B.; Chen, H.; Park, N.; and Subrahmanian, V. S. 2016. MAP: Frequency-Based Maximization of Airline Profits based on an Ensemble Forecasting Approach. In Krishnapuram, B.; Shah, M.; Smola, A. J.; Aggarwal, C. C.; Shen, D.; and Rastogi, R., eds., *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, 421–430. ACM.
- An, B.; Chen, H.; Park, N.; and Subrahmanian, V. S. 2017. Data-Driven Frequency-Based Airline Profit Maximization. *ACM Trans. Intell. Syst. Technol.*, 8(4): 61:1–61:28.
- Bertsekas, D. P. 1997. Nonlinear programming. *Journal of the Operational Research Society*, 48(3): 334–334.
- Borkar, V. S. 2009. *Stochastic approximation: a dynamical systems viewpoint*, volume 48. Springer.
- Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J.; et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1): 1–122.
- Fang, Y.; Qin, Y.; Luo, H.; Zhao, F.; Zeng, L.; Hui, B.; and Wang, C. 2021. CDGNet: A Cross-Time Dynamic Graph-based Deep Learning Model for Traffic Forecasting. *CoRR*, abs/2111.02736.
- Fercoq, O.; Alacaoglu, A.; Necoara, I.; and Cevher, V. 2019. Almost surely constrained convex optimization. In Chaudhuri, K.; and Salakhutdinov, R., eds., *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, 1910–1919. PMLR.
- Hui, B.; Yan, D.; Chen, H.; and Ku, W. 2021a. Trajectory WaveNet: A Trajectory-Based Model for Traffic Forecasting. In Bailey, J.; Miettinen, P.; Koh, Y. S.; Tao, D.; and Wu, X., eds., *IEEE International Conference on Data Mining, ICDM 2021, Auckland, New Zealand, December 7-10, 2021*, 1114–1119. IEEE.
- Hui, B.; Yan, D.; Chen, H.; and Ku, W. 2021b. TrajNet: A Trajectory-Based Deep Learning Model for Traffic Prediction. In Zhu, F.; Ooi, B. C.; and Miao, C., eds., *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021*, 716–724. ACM.
- Li, D.; Liu, J.; Jeon, J.; Hong, S.; Le, T.; Lee, D.; and Park, N. 2021. Large-Scale Data-Driven Airline Market Influence Maximization. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 914–924.
- Lin, Q.; Ma, R.; and Yang, T. 2018. Level-Set Methods for Finite-Sum Constrained Convex Optimization. In Dy, J. G.; and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, 3118–3127. PMLR.
- Locatello, F.; Tschannen, M.; Rätsch, G.; and Jaggi, M. 2017. Greedy Algorithms for Cone Constrained Optimization with Convergence Guarantees. In Guyon, I.; von Luxburg, U.; Bengio, S.; Wallach, H. M.; Fergus, R.; Vishwanathan, S. V. N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 773–784.
- Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Lillicrap, T. P.; Harley, T.; Silver, D.; and Kavukcuoglu, K. 2016. Asynchronous Methods for Deep Reinforcement Learning. In Balcan, M.; and Weinberger, K. Q., eds., *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, 1928–1937. JMLR.org.
- Panageas, I.; Piliouras, G.; and Wang, X. 2019. First-order methods almost always avoid saddle points: The case of vanishing step-sizes. *Advances in Neural Information Processing Systems*, 32.
- Papadimitriou, C. H.; and Steiglitz, K. 1998. *Combinatorial optimization: algorithms and complexity*. Courier Corporation.
- Pardalos, P. M.; Prokopyev, O. A.; and Busygin, S. 2006. Continuous approaches for solving discrete optimization problems. In *Handbook on modelling for discrete optimization*, 39–60. Springer.
- Schulman, J.; Levine, S.; Abbeel, P.; Jordan, M. I.; and Moritz, P. 2015. Trust Region Policy Optimization. In Bach, F. R.; and Blei, D. M., eds., *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, 1889–1897. JMLR.org.
- Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T. P.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; and Hassabis, D. 2016. Mastering the game of Go with deep neural networks and tree search. *Nat.*, 529(7587): 484–489.
- Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; Chen, Y.; Lillicrap, T. P.; Hui, F.; Sifre, L.; van den Driessche, G.; Graepel, T.; and Hassabis, D. 2017. Mastering the game of Go without human knowledge. *Nat.*, 550(7676): 354–359.
- Suzuki, Y. 2000a. The relationship between on-time performance and airline market share: a new approach. *Transportation Research Part E: Logistics and Transportation Review*, 36(2): 139–154.
- Suzuki, Y. 2000b. The relationship between on-time performance and airline market share: a new approach. *Transportation Research Part E-logistics and Transportation Review*, 36: 139–154.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In Guyon, I.; von Luxburg, U.;

Bengio, S.; Wallach, H. M.; Fergus, R.; Vishwanathan, S. V. N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 5998–6008.

Wei, C.; Lee, C.; Zhang, M.; and Luo, H. 2021. Linear Last-iterate Convergence in Constrained Saddle-point Optimization. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Wei, W.; and Hansen, M. 2005. Impact of aircraft size and seat availability on airlines' demand and market share in duopoly markets. *Transportation Research Part E: Logistics and Transportation Review*, 41(4): 315–327.

Williams, R. J. 1992. Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Mach. Learn.*, 8: 229–256.

Yu, B.; Yin, H.; and Zhu, Z. 2018. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In Lang, J., ed., *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, 3634–3640. ijcai.org.