Business Analytics for Intermodal Capacity Management

1. Introduction

This paper reports the theory and practice of business analytics in the intermodal industry. It grows out of a prior collaboration with Hub Group, a major Intermodal Marketing Company (hereafter IMC). The IMC operates an intermodal network. It sells freight services to customers, provides containers for holding cargo, and coordinates local pickup and long haul shipment. A typical transaction goes as follows. A customer (shipper) who wants to move a full container load, e.g., from Los Angeles (LA) to Boston, contacts the IMC. Within a few hours, the customer service representative (CSR) of LA must decide whether the load is profitable enough to accept. Once the order is accepted, an empty container is delivered to the shipper for loading. Then the loaded container is picked up via truck, shipped to the destination via train, and delivered to the receiver via truck. After the container has been unloaded and returned, it becomes available for future use in Boston, or other locations if the container is repositioned out of Boston.

A major constraint on the service capacity is the number of containers available in a location. To manage it, the industry has adopted a hierarchical mechanism. It is a decentralized system with two authority levels. At the tactical level, central management (CM) controls monthly container supply over the network: it carries out container repositioning based on mid-term forecasts. At the execution level, each CSR manages daily demand for a single designated location: given container supply, he handles load acceptance in real time, based on the short-term forecasts.

For effective capacity management, the IMC must address three challenges. (i) Heterogenous delivery leadtime: Customers typically have long-term contracts in place with the IMC, and premium customers pay higher rates for prioritized access to container capacity. Such differentiation can improve profitability, but it also demands more efficient operations. (ii) Environmental uncertainty: For each location, daily demands differ substantially in profit margin and leadtime requirements; they exhibit strong regional and cyclical day-of-week patterns, ranging from hundred per day at one location to just a few per week at another. Through load acceptance and shipment, volatile demand from order origins translates into volatile supply at order destinations. Effective capacity management, therefore, must account for random demand and supply, which are co-evolving over time. (iii) Network interdependence: The IMC operates a complex network of 24 locations (CSRs); they are tightly connected by demand, acceptance policy, and resulting traffic. The CSR’s current supply is determined by the past acceptance elsewhere; his current acceptance affects future supply for the destination CSRs. This strong interdependence further complicates capacity management.

In the past, the IMC had employed the hierarchical mechanism for container repositioning, and the first-come-first-served (FCFS) policy for load acceptance. This simple arrangement has produced two problems.

---

1 The delivery leadtime is the maximal time allowed from the order requested until the order must be filled.
The first is chronic network imbalance. It is driven by demand disparity and the FCFS acceptance policy. In the U.S., large imports of Asian goods drive dominant eastbound traffic, leaving the west coast in constant container shortage. This situation is further exacerbated by the FCFS policy, which relies on the one-way economics, accepting a load as long as its one-way revenue exceeds the service cost. This implies that repositioning empty containers would inflict a net loss on CSRs, which discourages them from voluntarily repositioning excess containers to the deficit locations. As a result, demand disparity perpetuates into a chronic supply imbalance. To control it, the IMC must periodically reposition empty containers, a costly undertaking.

The second problem is operational inefficiencies. First, under the FCFS policy, CSRs make no use of short-term forecasts, exercising no active control over the traffic. As a result, future profitable loads are often dislodged by the less profitable ones accepted earlier, and the service and relationships with premium customers suffer. Second, current load acceptance determines future container supply. So load acceptance and container repositioning should be managed jointly. Yet the IMC takes a separate approach, which cripples the CSR’s ability to respond quickly to changing local conditions. Third, CM indeed acts on mid-term forecasts, repositioning containers only infrequently, not every day. This delay is costly: a shortage faced by a CSR can only be resolved by CM a week or more later; until then, she must turn away additional orders, no matter how profitable they are.

Effective management, therefore, must leverage local short-term forecasts, deploy responsive policies, and integrate a network perspective; it must coordinate CM and CSR activities, across locations and over time. These principles guide our solution. In this paper, we focus on how to tackle the three challenges—heterogenous leadtime, environmental uncertainty, and network interdependence—in a practical way. In particular, we seek to address three questions: (1) Given the decentralized hierarchies and dispersed information, how should the IMC reduce chronic network imbalance? (2) How should the IMC improve operational efficiency, in light of heterogenous leadtime and evolving supply? (3) How does our new approach fare with existing systems?

Our study makes three contributions. The first is a novel scarcity pricing scheme. It is a decomposition method with the rolling horizon execution. The scheme has two components: the network flow model for CM, and the load acceptance model for the CSRs. In particular, the acceptance model is framed as a Markov decision process, fully embracing environmental uncertainty. For each CSR, we compute the scarcity price, a sufficient statistic that distills all the relevant information other CSRs need to know. We then refine each CSR’s objective function with the scarcity price vector, which provides current scarcity condition of the network.

Our scheme addresses the main deficiencies of the old practice. Through scarcity prices, it internalizes the externalities that each CSR’s action imposes on the future and the network, thereby calibrating the
timing and magnitude of adaptive responses. By refining the CSR’s objectives, it motivates the CSRs to reposition empty containers, which would not happen under one-way economics. Through these actions, the scheme also extracts and propagates the short-term forecasts that influence the CSR’s decisions in the first place, thereby addressing the challenge of network interdependence. As a result, the scheme is able to create dynamic balancing process to reduce the chronic network imbalance.

Our second contribution is a refined load acceptance policy. The model accounts for two key characteristics of intermodal operations: heterogenous leadtime and evolving supply. To solve this high-dimensional problem, we propose a decoupling approach, which reduces the problem into a sequence of one-dimensional convex programs. We then characterize the nested-threshold structure of the optimal policy. The thresholds are driven by both demand and supply risks; they translate leadtime heterogeneity into fulfillment flexibility which helps improve efficiency. Using an explicit form, we pin down three driving forces of the optimal policy. We show each force has its own effective range, and it pays to improve the forecast accuracy.

Our third contribution is algorithmic testing and validation with real-world data. Gorman (2010) implemented a heuristic inventory targeting system at the IMC. We improve on it with the scarcity pricing scheme and a refined acceptance policy. We further derive analytical insights on the heterogeneous leadtime and random supply—both are original in the literature. We then put the insights into algorithmic implementation: using structural properties, we develop a linear-time algorithm for computing the policy, and a Monte Carlo algorithm for computing scarcity prices. Our experiments reveal that the improvement over Gorman (2010) can be substantial, especially when either network imbalance or supply risk is high.

To our knowledge, this is one of the first few studies that model, analyze, and implement business analytics in the intermodal industry. With certain modifications, our general framework and solution ideas can also apply to other settings, e.g., truckload shipping and car rental.

References