Offer Acceptance in the Freight Industry

Keagan Benson [1] Benson Duong [1] Radu Manea [1] Nima Yazdani [1] Matthew Stirling [2]

[1]University of California- San Diego [2]Flock Freight

Links

- Radu Manea - rmanea@ucsd.edu
- Benson Duong - beduong@ucsd.edu
- Keagan Benson - ktbenson@ucsd.edu
- Nima Yazdani - niyazdan@ucsd.edu

- *Website*
- *Github*

.

Background And Introduction

Flock Freight Information and Problem Set Up:

- **Flock Freight** (FF) deals with different carriers and shipments, acting as a marketplace for freight carriers to place bids on shipping orders needed to be fulfilled, which is otherwise known as a **freight broker**.

1. A Company X needs to ship an Order to a given place by a given time, but lack their own delivery vehicles.
2. Carriers are entities that offer delivery services.
3. FF is the intermediary between 1 and 2 above: One-by-one, FF processes an incoming stream of **N** offers from different carriers to deliver the order.
   - Only one offer will have the cheapest delivery rate: **rate***
   - FF cannot accidentally reject this offer **(there's no going back when FF rejects an offer!)**

Oftentimes, the decision to accept or reject a carrier's offer is very difficult. When trying to maximize margin, we must be wary of which offer we accept because there may be a future offer that is better than the one we choose.

Data Description

We use two data sources for our model. The records in the **Orders Table** attribute to a certain order. The records in the **Offers Table** attribute to a certain offer on an order.

Table 1. Orders Data Dictionary

Column Name	Description
REFERENCE_NUMBER	Unique ID for the order
ORDER_DATETIME_PST	Date and time of order in Pacific standard time
PICKUP_DEADLINE_PST	Date and time order must be picked up from origination in Pacific standard time
DELIVERY_TIME_CONSTRAINT	Type of delivery time scheduling constraint
ORIGIN_3DIGIT_ZIP	The first three digits of the origination location ZIP Code
DESTINATION_3DIGIT_ZIP	The first three digits of the destination location ZIP Code
APPROXIMATE_DRIVING_ROUTE_MILEAGE	Approximate number of driving miles from origination to destination
PALLETIZED_LINEAR_FEET	The length and weight of the shipment converted to the percent amount of the truck filled
FD_ENABLED	Customer paid for upgraded service with delivery deadline and no transfer of truck (hub and spoke system not allowed)
EXCLUSIVE_USE_REQUESTED	Customer paid for shipment to be delivered on its own truck (cannot be pooled)
HAZARDOUS	The shipment is hazardous material (cannot be pooled)
REFER_ALLOWED	The shipment can go on a refrigeration truck
STRAIGHT_TRUCK_ALLOWED	The shipment can go on a straight truck
LOAD_BAR_COUNT	The number of load bars required by the load
LOAD_TO_RIDE_REQUESTED	Delivery service without hub stops
ESTIMATED_COST_AT_ORDER	Flock Freight's estimated cost to fulfill the order (estimated at the time of order)
TRANSPORT_MODE	The type of shipment (FTL, LTL, PTL)

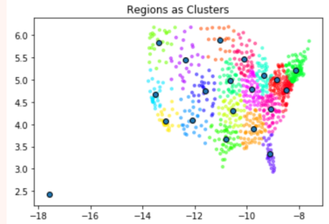
Table 2. Offers Data Dictionary

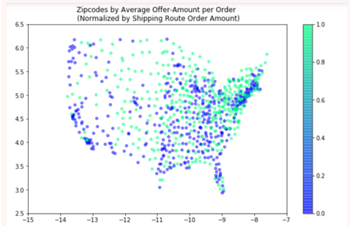
Column Name	Description
CARRIER_ID	Unique ID for the carrier providing the offer
REFERENCE_NUMBER	Set of order reference numbers the offer would deliver (more than one for a pool)
CREATED_ON_HQ	Date and time offer was submitted in Pacific standard time
RATE_USD	Amount carrier will be paid if offer is accepted
OFFER_TYPE	"pool" for two orders pooled together, or "quote" for one order
SELF_SERVE	Boolean field designating carrier made offer through the app without representative intervention
IS_OFFER_APPROVED	Boolean field designating if Flock Freight approved carrier's offer (carrier must still confirm contract)
AUTOMATICALLY_APPROVED	Boolean field designating if Flock Freight approval was done without representative intervention
MANUALLY_APPROVED	Boolean field designating if Flock Freight approval was done with representative intervention
WAS_EVER_UNCOVERED	Boolean field designating if agreed contract to deliver load was ever broken (e.g. carrier truck broke down)
COVERING_OFFER	Boolean field designating Flock Freight and carrier agreed contract together to deliver load
LOAD_DELIVERED_FROM_OFFER	Boolean field designating this offer was the offer to deliver load
RECOMMENDED_LOAD	Boolean field designating the load (set of order references numbers) was sent to the carrier as a recommended load

Methods Overview

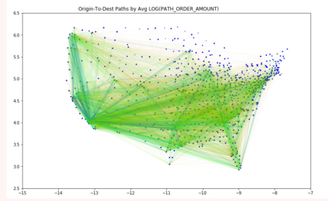
Geographic EDA: Geopandas was used to preprocess the given data, and extract potentially useful geospatial information for the models to be used.

1. * Zip3 Centroid Scatterplots: Point Maps of the Zipcodes show that metropolitan areas can be grouped with K-meansClustering (See "Regions as Clusters"), which will simplify one-hot-encoding the zipcodes of an order's destination and origin.
   * The zipcodes scatterplots are also shaded differently according to some quantitative variable that's useful for the model: The right map colors zipcodes by offer amounts. Overlaying the right map on top of the left can reveal which "regions" have more offers per order.

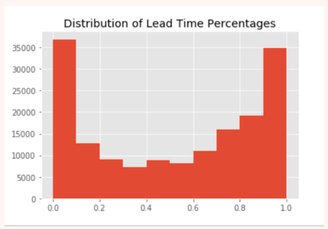




2. * Shipping Route Plots: Since each order needs a start and endpoint, it can be mapped as edges between nodes. This can potentially show supply lines and hubs, or regions with heavy activity. Quantitative variables can be shown at the same time (like Rates), showing if shipping routes tend to be cheaper than others.



3. The "Lead Time Percentage" histogram shows the lead time percentage or "patience". It's the duration between the chosen offer and order's initiation, divided by duration between the order's dead-line and initiation. Bimodal, it shows high amounts of "overly-early" acceptances, and "last-second" acceptances. This will be a very relevant issue and obstacle for the project, since overly early acceptances create coverage bias, especially for offer amount prediction.



Models

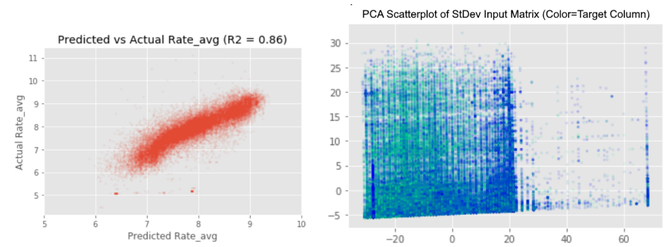
Our model comprises of 3 sub-models.

- **Offer Amount Prediction Model**: For a given order, predict how many offers ***N*** it will receive beforehand. A high ***N*** allows gambling with "over-rejecting" until a cheap offer comes along. * Linear Regression is used. The coverage bias from "overly early" acceptances mean that samples must be weighted by lead time.
- **Rate Avg & StDev Prediction Model**: For a given order, predict the avg rate $rate_\mu$ of its expected offers; When the offers $Offer^i$ come in, classify each as "cheap" if < predicted $rate_\mu$. Furthermore: predict the standard deviation $rate_\sigma$, so an offer can be classified as "very cheap" if $rate^i < rate_\mu - rate_\sigma$. Tolerance can be adjusted post-training with λ: $rate^i < rate_\mu - \lambda\,(rate_\sigma)$.

.

Summary of Findings

- The Rate Average Model used linear regression, with an R2 of 85% (or 90% if only with FTL). The most influential features were Distance, Truck Shipment Volume (Palletized Linear Feet), and most of the one-hot-encoded region columns.
- The Standard Deviation Model needed to be resimplified as ordinal classification with Random Forest (classes=2). The ROC AUC score is 60-67%. Its PCA Scatterplot, colored by the target column, is shown.
- The offer amount is a very useful feature for $rate_\sigma$, and so the offer amount model's success will help this model too. The Offer-Amount-per-order's distribution is severely discrete (50% of ***N's*** are <= 3, and 75% <= 5), explaining why this $rate_\sigma$ model had to be ordinal and not regression-based. But when normalized by the shipping route's order amount, the ROC AUC of the binary classification goes from 60% to 78%.



Quarter 1 Model Set Up

Our resulting offer acceptance method consists of a baseline secretary method that utilizes a trained linear regression model that estimates the number of offers expected and a prior that uses a linear regression model that predicts the estimated cost. The expected number of offers is modeled using the following features: the boolean features of the Orders Dataset, 'APPROXIMATE DRIVING ROUTE MILEAGE', 'PALLETIZED LINEAR FEET', 'SECONDS BETWEEN ORDER AND DEADLINE', 'LOAD BAR COUNT', and 'ESTIMATED COST AT ORDER' while the estimated cost is trained using two features: 'SECONDS BETWEEN ORDER AND DEADLINE' and 'ESTIMATED COST AT ORDER'. The seconds between order and deadline is a new column generated by taking the difference between the 'PICKUP DEADLINE PST' and the 'ORDER DATETIME PST' column.

Our prior works by comparing the estimated cost and the rate given for the first n/e offers and if any of those offers have a rate less than the estimated cost to ship then we automatically accept that offer instead of moving on with the secretary method. We evaluated our offer acceptance method by comparing the average rate of the offers we chose in the test dataset, to the average rate of the historically accepted offers and found that the historically accepted rates were around 150 cheaper. The distribution of our accepted offers and the historically accepted offers can be seen below.