Second fORged-by-Machines Contest (2020)

This is the second iteration of the "fORged-by-Machines" contest for doctoral students. It is sponsored by Amazon Web Services (AWS), and supported by the USC Data-Driven Decisions Lab (which is sponsored by grants from NSF, AFOSR, and ONR). The winning prize will be $1,000, "runner-up" will get $500, and the honorable mention will get $250. Last year, this contest had 17 participants and they only had a week to submit. This year, we are giving the contestants more time: a full three weeks, but at most two students per team. Three teams will be chosen to present their solution on Nov 8, 2020 at the INFORMS Annual Meeting. The Prize ceremony will be held on Monday Nov.9, 2020 at the INFORMS Computing Society Business Meeting.

Problem Statement

Large cities (e.g., Los Angeles) provide tremendous opportunities for online suppliers due to increasing demand for package deliveries. A large e-commerce firm is engaged in a modeling effort to improve the efficiency of their fulfillment network serving such a large city. Providing an efficient network design, at scale, is a challenge and requires tools from both predictive and prescriptive analytics. Such models usually require inputs representing quantities to be shipped (O-D flows) from each origin to destination, and these inputs are usually estimated using predictive models. Because these flows vary significantly over time, an important consideration is to accommodate O-D flows which are uncertain. Not only do these flows vary on a daily basis, patterns change seasonally as well. As a result, predicting variability and accommodating it in prescriptive models is essential to providing high-quality network designs.

The data for the planning instance under consideration involves 4 origin nodes, 2 transshipment nodes and 7 destination nodes (warehouses). The plan requires the allocation of an appropriate number of trucks on the lanes from origin nodes to warehouses, and the assignment of customer locations to warehouses. Since customer locations are too numerous to enumerate, it suffices to assign customer ZIP codes to destination nodes, while also allocating the necessary number of trucks on each arc. The plan must be static, therefore cannot be altered day-by-day, based on demand. Once a plan is accepted, the firm uses it over the next planning period. All packages will be shipped from an origin node to a destination node, either through a transshipment node or directly. The packages will be delivered from the destination node to customer locations via a different model which is excluded from this exercise. The capacity of each truck is 1000 packages. The cost for operating a truck is approximately, $100 + $2 x miles. Given the latitude and longitude of two points: (lat1, lon1), (lat2, lon2), the delivery distance can be estimated by: $(69.5\,|\mathrm{lat}_1 - \mathrm{lat}_2| + 57.3|\mathrm{lon}_1\text{-}\mathrm{lon}_2|)$ miles. The locations of the nodes can be found in "node_location.csv" and illustrated in "node_location.html". You can assume that any network link that you deem necessary is available for use, free-of-charge. We have the daily delivery data from origin nodes to customer location with quantity of packages, in "DeliveryData" folder. This delivery data varies day-by-day and we provide the data for N=365 days. All data is available at https://core.isrd.isi.edu/chaise/record/#1/Core:File/RID=WCRY

Evaluation

To evaluate the plan, we require the contestant (groups of no more than two doctoral students) to submit the following files: "TruckAssignment.csv" and "CustomerAssignment.csv". "TruckAssignment.csv" includes the allocation for trucks on 50 lanes (origin node->destination node, origin node->transshipment node, transshipment node->destination node), and "CustomerAssignment.csv" should contain the assignment for 118 ZIP code areas to the destination nodes.

We will evaluate solutions based on three metrics as listed below:

1. Cost per package (based only on allocated trucks between facilities and "delivered" packages).
2. Distance per package (from origin nodes all the way to the customer locations).
3. Total missed package count
4. Equitable loads are among different warehouses

The evaluation will be based on solving a multi-commodity flow optimization model with the objective of minimizing distance and missed packages, which will be subject to the arc capacities defined by the truck allocations recommended by the contestant. The penalty for a missed package is $1000. Finally, note that the distance between destination nodes and customer locations will be inferred directly from the provided zip to destination node assignments, and no cost is associated with the final deliveries. All evaluations will take place over a set of test instances, involving potentially different customer locations and package demands. These locations and demands will be generated using the same underlying uncertainty model as the one which created the demand data.

Registration Deadline: **8:59 pm (Pacific Daylight Time), October 2, 2020.** In order to participate, please register at https://forms.gle/6BJ22p54qv9BFs7s6

Submission Deadline: 8:59 pm (Pacific Daylight Time), October 16, 2020. Send your entry by e-mail to sen@datadrivendecisions.org Every submission must be accompanied by a letter from a faculty member who must be an INFORMS member, and must certify that the students participating in the contest are Ph.D. students at their university.