PLASMA – Approximate Dynamic Programming finally cracks the locomotive optimization problem

Warren B. Powell, Belgacem Bouzaïene-Ayari, Coleman Lawrence, Clark Cheng, Sourav Das, Ricardo Fiorillo

For decades, the freight railroads in the United States have tried to use the power of mathematical programming to optimize the flows of locomotives over their networks. The problem was always that classical mathematical programming formulations could not handle the complex operational details that characterize freight railroads in the U.S. and Canada. Consist formation, foreign power, communications equipment, flush toilets and shop routing are just some of the critical details that need to be handled if a model is going to accurately capture locomotive productivity.

In addition, there is the problem of uncertainty in transit times, yard delays, equipment failures, train schedules and tonnages.

Given both the complexity and uncertainty, some have turned to rule-based simulation models, creating open debates within at least one railroad between whether optimization or simulation is the best approach.

Approximate dynamic programming can be viewed as an “optimizing simulator.” It combines the intelligence of optimization with the flexibility of simulation. ADP solves sequences of small optimization problems, while still stepping through time just as a simulation model would. However, ADP requires that we run the simulation repeatedly, and uses the principle of feedback learning to learn the value of locomotives in the future. The result is an intelligent simulator that can respond in a realistic way to changes in fleet size, new train schedules and new operating policies.

PLASMA is currently running at Norfolk Southern for strategic planning (such as fleet size and mix), or for short-term operational forecasting (say a week into the future). It can also be adapted to be used for real-time locomotive assignments, but this version has not been implemented.

Features

It is useful to start by highlighting some of the features of PLASMA:

- Each locomotive is modeled individually, making it possible to capture both horsepower and tractive effort, ownership, special equipment, shop status, and consist breakup costs.
- Arrivals and departures of trains and locomotives are modeled down to the minute.
- PLASMA will delay trains as necessary, prioritizing trains by importance.
- PLASMA can add locomotives to a train to reposition the power, or it can schedule light engine moves (moving locomotives without a train).
- PLASMA will optimize consists while working to get locomotives to their shop appointments on time. PLASMA will simultaneously balance the demands on each shop to minimize shop congestion.
- Foreign power is routed toward interchange points, although PLASMA can decide to hold on to foreign power if needed.
- When used as a strategic planning system, PLASMA is used to decide the fleet size and mix for a planned schedule.

PLASMA can be run in deterministic mode (which is how Norfolk Southern uses the system at the moment), or stochastic mode, where it explicitly models uncertainty in transit times and yard delays. It can also capture uncertainty in train schedules and tonnages. When uncertainty is introduced, PLASMA adopts a more conservative style, holding onto power to be prepared for contingencies.

For more on PLASMA, see

http://www.castlelab.princeton.edu/plasma.html
How it works
ADP breaks the problem of optimizing locomotives over time into a series of assignments of locomotives to a single train at a time. A single subproblem is depicted in figure 1, where we show the assignment of individual locomotives to trains.

The value of locomotives in the future are captured by nonlinear value function approximations (VFAs) which capture the value of different types of locomotives in the future. VFAs are estimated internally within the model through repeated simulations. The VFAs represent the magic of approximate dynamic programming, and this is how a single large problem is broken into a series of smaller problems. The value functions are learned by simulating our way through time, solving one assignment problem at a time.

Usage at Norfolk Southern
PLASMA is the model within a larger locomotive planning system called LARS developed by Norfolk Southern.

The LARS Fleet Size model has proven to be a valuable tool to Norfolk Southern’s fleet management process since 2008. With new locomotive order-to-delivery cycles generally totaling a year or more, determining accurate future locomotive fleet requirements is critical.

To accomplish this, a marketing forecast (carloads on a lane basis) for a future period (typically one to five years out) is converted to a train plan with the assistance of train and block models. Separately, a known historical period is used as a basis for comparison. After varying the fleet-sizes on the input and estimating train delay for each run, a comparison plot can be made to determine differentials at any specified service level.

A sample of this comparison is shown in figure 2 where October 2015 is the forecasted period using October 2010 as the base. The red circles represent the historical service level for 2010 in terms of delay due to the lack of available locomotive power. The ability of the model to replicate historical performance metrics was an important plateau that gave management the confidence to use the system for projecting future locomotive requirements. The difference, represented by the horizontal bar, is the change in units required in 2015 to retain 2010 service levels. In addition to providing these point estimates of locomotive demand, the LARS fleet size model has been applied to develop statistically-based ranges of future locomotive requirements based upon forecasted operating plans and observed variations in other operational parameters (e.g., train speed and crew start).

For the past four years, LARS’ fleet-size estimates have been a critical decision driver for NS’ annual locomotive purchases which totaled 172 units from 2008 to 2011.