Shared Q&A

INFORMS RAS · 2020 Problem-Solving Competition

Dated: May 25, 2020

- **What does the specified speed on segments (or Wyes, sidings and spurs) mean? Do we assume that all trains travel in the given speed? Or are they the maximum speed in the segments?**
  The specified speeds on the sidings, wyes, and spurs denote the speeds that the train should be going when entering these branches, while the speeds on the mainlines are the speeds on each of the segments if the train was bypassing the siding, wye, etc., altogether. In the network for the Toy Problem Dataset, there are speeds for Track #1, Track #2, and then for the sidings, wyes, spurs, etc.

- **Can trains travel slower than the listed speed on tracks or segments? Or must trains travel at each listed speed and arrive early at some intermediate station?**
  That is a very good question. Participants are to assume trains are all traveling at the permissible speeds for the track segment and not less than that. As such, it is possible that a train will arrive early at the intermediate station if there is no meet and pass delay in between stations. In other words, please assume the train is traveling at the speeds given for the segments in the network and the changes in speeds are instantaneous for simplicity. In the real-world, trains do proceed at slower than posted speeds for various reasons (e.g., maintenance in the area, or just cautionary movements).

- **It seems that Wye is used for trains to change direction. But why does a train change direction in practice?**
  You are correct; a wye can be used to change the direction of a train, but it is also used to move trains traveling in either direction to another branch line or subdivision. Also, it is possible that a specific rail car needs to be oriented a specific way. As such, wyes allow a set of cars to be re-oriented if they're used in conjunction with storage tracks. In practice, a wye can be found at one of two places: an industrial location, such as a power plant, or at a junction where another mainline enters the network. In either of these cases, when the train has offloaded its cargo or a train is entering the network from the branch line, it may need to proceed in either direction (e.g., east or west). The wyes allow this movement. I hope this wasn't too confusing; it's a bit difficult to explain without visualizations, but please let me know if anything was unclear.

- **What will the format of all the inputs be? When will these information be available?**
  The format of the validation dataset will be identical to that format of the Toy Problem dataset. We are expecting to release the validation dataset on May 29, assuming no delay from our railroad partner who's graciously providing the real-world data for us to use.

- **What is the final output that is expected from participants? How is the output evaluated?**
  The final output will be a final report that provides the following, which will be evaluated by a team of judges from industry and academia:
  - Feasibility of the proposed solution, it must satisfy all the given constraints.
  - The quality of the solution in terms of its objective function value.
  - The tractability of the solution approach.
  - The implementation quality of the approach.
  - The practical usability/reproducibility of the solution approach.
  - Computational time of the proposed solution approach.

   - The generalizability of the solution approach.
   - The quality of the paper describing the solution approach. How clear is the explanation? Is it possible to reproduce the approach just by reading the paper?

- **Shall we derive a generic prediction model that will be evaluated in completely new networks? Or shall we derive a prediction model based on a given network; and later our model is tested against different train schedules in the same network.**
  The goal of the competition is to develop a generic prediction model that can be used independent of the network. Although, you will be given a validation dataset with which to compare how your model performs against real-world data. This will allow you to tweak your model as needed. Ideally though, the solution should be flexible enough that it *could* be applied to other networks.

- **What will be the evaluation criteria? We assume that we will predict actual arrival time of trains with a given schedule in a given network; and the criteria is the accuracy of prediction. Is our understanding correct?**
  Above, I've listed the criteria on which the approach will be judged. I will be working with the team of judges over the next couple months to derive an evaluation matrix for them to follow, but it will be based on the criteria above. In general, while the accuracy is important, the creativity in the approach and the considerations made to achieve the results are just as important as the predictions. So, it has to be looked at as a whole, rather than individual parts.

- **In the problem description, it writes that "The accuracy of the ETA will be determined using the historical data by comparing the model-generated ETAs with actual arrival times." Will you use average historical arrival time for evaluation or one-time historical arrival time?**
  Given that we incorporated random variables into this problem, I'd expect a range of values will be acceptable, but we will have on-time historical arrival times from an operating railroad for a number of trains. I wish I could expand upon this more, but we have not received the the final validation set just yet, so I'd expect to have more of an answer once we've seen what is provided in the railroad data.

- **Do you plan to provide a large number of (historical) samples of train operations in a certain network, so that some data-driven machine learning algorithm can be trained?**
  The validation dataset will provided a large amount of samples within a certain network. Unfortunately, it will likely be one dataset over a small length of time, but it is expected to be large enough to be split, if needed, to derive a data-driven approach using machine learning.