An Iterative Approach to Optimizing an NHL Schedule

Kyle A. Mana
Kyle.a.mana@gmail.com

1 – Problem

The NHL schedule introduces an interesting optimal routing problem; with over 25 destinations per season, variable numbers of visits to each location, and constraints imposed due to maximum road trip lengths, schedule variability to maintain fan interest, conflicting interests between teams, and many more such considerations.

This brief paper offers a possible approach to optimizing the schedule for a single NHL team with regards to distance traveled. The model has been constructed in pyomo and the code can be found at publicly on github.

2 – Model Composition

Prior to mathematically constructing the model, certain parameters needed to be constructed and schedule complexities such as variable visits to each city need to be managed.

2a – Calculating Distance Between Arena's

Before applying any optimization logic, the data needs to be constructed for an optimal routing problem. Each team's schedule (away games/home games) and the lat-long of each team's arena is easily accessible.

The main parameter that needs to be constructed is the distance from one arena to another. To easily construct the distance between two arenas we can calculate the haversine distance for each lat-long pairing between cities. This will underpin out optimization model with the assumption that the distance will be traveled "as the crow fly's", or in the shortest distance from A to B. Obviously this assumption is not entirely accurate, as there will be variable distance traveled from airport to arena, however it provides the level of precision needed for all intents and purposes.

The haversine distance for a lat-long pairing can be calculated as:

2b – Managing Model Constraints

A majority of the schedule and routing constraints can be managed within the linear program itself. However, the presence of multiple routes (limited by a certain number of destinations per trip) and the variable number of visits per destination greatly increases the level of computational complexity and stretches runtimes (particularly when using glpk solver).

A common approach to managing the computation complexity of a “greenfield” analysis is to leverage a secondary classification model; placing departure locations (in the context of distribution networks) in areas of high location density or eliminating particular A to B location pairs (referred to moving forward as “arcs”) due to geographical location. This initial model can greatly reduce the computational complexity of considering every possible location, route, etc. and the inherent explosion of parameters and decision variables.

An issue with segmenting location sets in this way for the problem being considered is the variable number of games in each location. Some locations are visited as many as four times per season, while others are only visited only once. Doing initial segmentation to limit the number of considerable arcs from A to B creates risks of discontinuity in the number of games for a subset and suboptimal routes where a trip is being made to visit only a single team.

To manage the variable number of games in each location, an iterative optimization approach can be leveraged to optimize the NHL schedule for each individual team. To do so, the optimization problem is set up for an individual road trip with several assumptions such as:

1. The road trip can only visit four teams
2. The road trip must play only one game per city visited
3. The road trip must be circular (leave and return to the home city)

Within this construction a parameter is nested with the number of games remaining in each destination. From this point forward, the optimization problem can be iteratively solved with the number of games in each location being decremented by the number of visits within the optimization iteration. The optimization problem is then considered to be completed when there are no remaining games in the schedule.

With the high-level model framework laid out, we can now review the mathematical formulation (continued on page 3).

Sets:

- h = home location
- L = possible game locations

Parameters:

- d_{ij} = distance from location i to location j
- n_i = number of games remaining at location i
- m = number of games per road trip

Variables:

- x_{ij} = whether or not to make the trip from location i to location j

While:

$$\sum_{i \in L} n_i > 0:$$

Solve:

$$\text{Min} \sum_{j \in L} \sum_{i \in L} x_{ij} d_{ij}$$

Subject to:

Must make trip out of home arena

$$\sum_{i \in L} x_{hi} = 1$$

Must make trip back to home arena

$$\sum_{i \in L} x_{ih} = 1$$

Must both arrive at and depart from of every arena on a trip

$$\sum_{j \in L} x_{ij} - \sum_{j \in L} x_{ji} = 0 \text{ for all } i \in L$$

The trip cannot consist of simply going back and forth between to arenas (circular trip)

$$\sum_{j \in L} X_{ij} - (X_{it} + x_{ti}) \geq 0 \text{ for all } i \in L, t \in L$$

Total games in the trip must not exceed the maximum limit per trip

$$\sum_{j \in L} \sum_{i \in L} x_{ij} = m$$

You cannot visit a location unless there is a game to play

$$\sum_{i \in L} X_{ij} - n_i \leq 0 \text{ for all } j \in L$$

Deprecate:

$$n_i -= \sum_{j \in L} x_{ji} \text{ for all } i \in L$$

Assumptions:

- Each road trip must be four games or less
- The trip must keep moving after every game (i.e. cannot play in location *y* several times after landing there)

Vancouver Canucks Optimization Results