Automated Decision Making: Final Project

Luca Lumetti
244577@studenti.unimore.it

June 30, 2021

1 Introduction

In this project I've tried to face the Max-Mean Dispersion Problem using Tabu Search guided by deep reinforcement learning during the dispersion fase.
Given a complete graph G(V, E) where each edge has associated a distance or affinity d_{ij}, the Max Mean Dispersion Problem is the search of a subset of vertex $M \subset V, |M| >= 2$ which maximize the dispersion, calcuated as follow:

$$MeanDispersion(M) = \frac{\sum_{i<j;i,j\in M} d_{ij}}{|M|}$$

2 RLTS

In 2020, Nijimbere et al. proposed an approach based on the combination of reinforcement learning and tabu search, named RLTS. The main idea is to use Q-Learning to build an high-quality initial solution, then the initial solution is improved with a one-flip tabu search algorithm.

3 DQNTS

My idea is to let a network to learn an heuristich to build the initial solution, which can generalize to graphs of any size, then use one-flip tabu search to improve that solution as in RLTS.

3.1 Network Architecture

The network architecture is based on [1], the hyperparameters setting can be seen at www.github.com. The state2tens embedding is done with 5 features extracted by each node, which are:

- 1 if the node in in the solution, 0 otherwise
- the sum of all edges connected to the node
- the sum of all edges connected to the node and the solution nodes
- the sum of all edges connected to the node and the nodes not in the solution

To construct the initial solution The network has been trained over 10 different instances (MDPIx_35 and MDPIIx_35, $1 <= x <= 10$) for 5001 episodes.

3.2 Tabu Search

The tabu search implementation follows the one in RLTS paper, with the only difference in the parameter $\alpha = 100$ instead of $\alpha = 50000$. My implementation couldn't finish even a single iteration in the time limit imposed with the parameter α proposed in the paper. This made me think that my implementation is way slower than the one in RLTS, but still i left the same time contraints as the results were not that bad.

3.3 General Algorithm

The network architecture is used during the contruction of the inital solution: for all the nodes, the network estimate the reward, then all the values get interpolated in the range $[-1, +1]$ and all the nodes $>= 0$ are named as "good nodes". Among these "good nodes", a random amount is taken to construct the initial solution.

References

[1] Dieudonné Nijimbere, Songzheng Zhao, Xunhao Gu, Moses Olabhele Esangbedo, and Nyiribakwe Dominique. Tabu search guided by reinforcement learning for the max-mean dispersion problem. *Journal of Industrial & Management Optimization*, 2020.