Flatland Challenge

G. Berselli, R. De Matteo, M. M. L. Pulici

July 19, 2021

Contents

1 Introduction

1.1 Flatland Challenge

The aim of the challenge [1], is to achieve efficient management of railway traffic. In particular, this problem is tackled in a simple grid world environment, in order to simulate and experiment different scenarios.

In more detail, the goal consists in making a number of trains arrive at their destinations minimizing the travel time. Even though for simple environments the train can follow explicit plans, as complexity increases the problem of mapping all possible states becomes intractable, for this reason a class of algorithms known as Reinforcement Learning (Section 1.2) can be exploited.

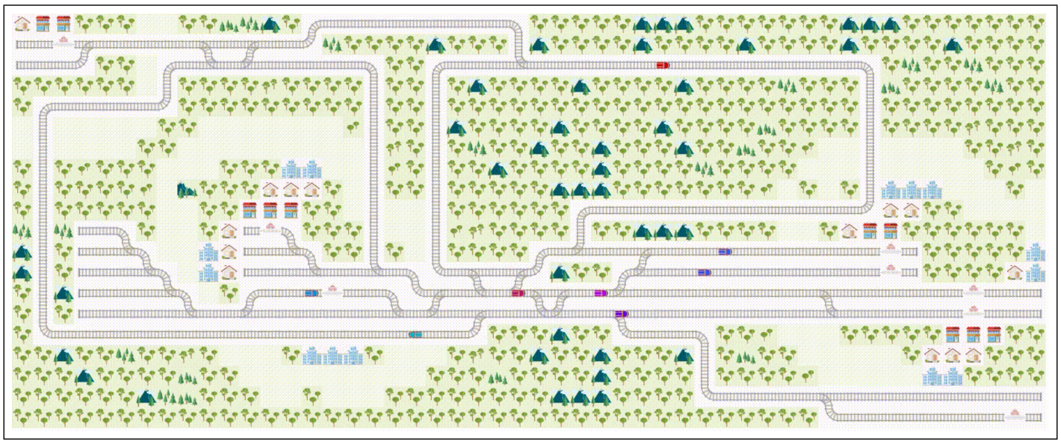

Figure 1: A possible Flatland instance

1.1.1 The environment

The Flatland environment [2] consists in a discrete time simulation, meaning that each action is performed with a constant time step. At each step, an agent for each simulated train chooses an action. An agent is defined as "an entity that can move within the grid and must solve tasks" [2]. More precisely, each agent can choose between two actions: waiting or moving in a direction. Each agent has an individual starting position and its goal is to reach its target destination. Of course, two agents can not occupy the same cell at the same time.

Each cell in the Flatland grid can take the form of any of 8 tile types, as shown in Fig. 2. More configurations can be obtained by rotating and mirroring the 8 basic tiles.

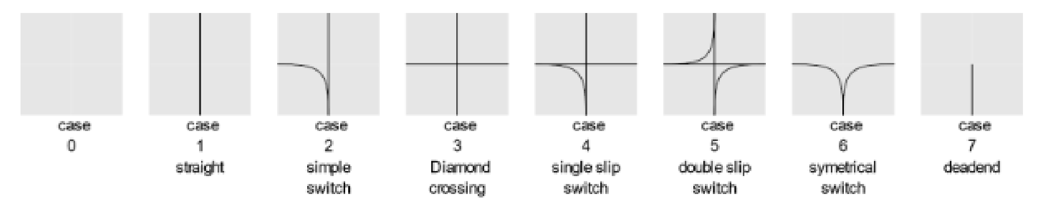

Figure 2: The 8 cell types

When the tile is of the straight type, the agent can only choose to continue moving or to stop. In presence of a simple or a double switch, the train is forced to decide in which of the offered direction to move. When arriving at a dead end, an agent can only stop or go backward.

1.2 Reinforcement Learning

The term Reinforcement Learning [3] refers to the class of agents which rely on feedbacks, or rewards, produced by the agents themselves when interacting with the environment. The way Reinforcement Learning works is to use observed rewards to learn the best possible policy for the given environment. In other words, the agent has no prior knowledge of the environment and must learn how to behave based only on posterior trial-and-error feedbacks.

There are three basic agent designs for Reinforcement Learning:

- Utility-based agent, which learns a utility function on states and uses it to select actions;
- Q-learning, which learns an action-utility function giving the expected utility of an action given a state;
- Reflex agent, which learns a policy mapping directly from states to actions.

In addition to these designs, Reinforcement Learning can be either passive, meaning that the policy is fixed and the task is to learn the utilities of states, or active, referring to agents which must also learn how to act.

1.2.1 Q-learning

For the Flatland Challenge, Q-learning is used. This class of algorithms learn an action-utility representation instead of simply learning the utilities. The value of an action-state tuple is typically indicated as $Q\left(s,a\right)$. There is a direct correlation between Q-values and state utilities, as expressed by the following equation:

$$U\left(s\right) = \max_a Q\left(s,a\right). \tag{1.1}$$

The main difference between Q-functions and basic utility information si that no model of the form $P\left(s'|s,a\right)$ is needed, either for learning or for action selection. This characteristic makes Q-learning a model-free method.

At equilibrium, when the Q-values are correct, the following equation must hold:

$$Q\left(s,a\right) = R\left(s\right) + \gamma \sum_{s'} P\left(s'|s,a\right) \max_{a'} Q\left(s',a'\right). \tag{1.2}$$

References

[1] AIcrowd. (2021). Flatland, [Online]. Available: `https://www.aicrowd.com/challenges/flatland`.

[2] ——, (2020). Flatland challenge, [Online]. Available: `https://www.aicrowd.com/challenges/flatland-challenge`.

[3] S. J. Russell and P. Norvig, “Reinforcement learning”, in *Artificial Intelligence, A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Pearson Education, 2009, ch. 21, pp. 830–859.