TODO Title
TODO SubTitle

1 Robo Cup Logistic League Motivation

The Robo Cup Logistics League (RCLL) is designed to represent a common scenario in Industry 4.0. There one needs to assemble products based on dynamic custom orders in a real-time environment. In turn the environment is driven by a consistent stream of orders and individual delivery windows. The challenge here consists in scheduling the assembly process of orders in a feasible manner, to abide the delivery window of as many orders are possible.

2 Reinforcement Learning approach

We present a solution using Reinforcement Learning (RL) to tackle the problem of deciding which assembly step the robots need to process next. A straightforward Deep Q-Learning approach is applied, with a flexible reward design.

2.1 Environment

In the first steps we apply a more simplified environment, which for starters only provides an order schedule with delivery windows and randomly. The initial goal is thus for the system to learn which orders are most feasible to complete or award most intermediate rewards.
We build the environment so that we can have a precision of seconds, while not specifically modelling each second as a discrete time step. The idea is to merge the discrete time into the actual state as a numerical feature, which is incremented in the expected passed time for a processing step. Initially the elapsed time for the intermediate steps is drawn from Gaussian distributions, which can be later extended with application of the distance matrix and real-world data.
The environment utilizing most of the complexity behind the task

2.2 States

The state space consists of orders, up to 3 products in the pipeline, distance matrix of the machines, availability of machines and current time. While the distance matrix is constant throughout a game, the other parameters are simulating a dynamic real-time environment.
Specifically we can represent an order as a vector of $\{\mathcal{B}, \mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{C}, n, c, d_{start}, d_{end}\}$, where

- $\mathcal{B}$ represents a set of base types
- $\mathcal{R}$ is one of up to 3 ring types
- $\mathcal{C}$ is the cap type
- n is the requested amount
- c indicates whether the order is competitive
- d_{start} is the start of the delivery window
- d_{end} is the end of the delivery window

A intermediate product will look like $\{\mathcal{B}, \mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3\}$, which corresponds to the physical object one of the 3 robots can transport. We can also omit the cap as we will deliver straight after.
The distance matrix and

While there is still room for changes the current dimensions are

2.3 Actions

2.4 Rewards

References