Exploratory Data Analysis:

Q-learning for Automated Personnel Scheduling

Ben Platten
13408410

master information studies
Information Studies: Data Science
faculty of science
university of Amsterdam

fill in your date of defence in format yyyy-mm-dd

	1st Examiner	2nd Examiner
Title, Name	title and name of 1st Examiner	title and name of 2nd Examiner
Affiliation	affiliations of 1st Examiner	affiliations of 2nd Examiner
Email	email of 1st Examiner	email of 2nd Examiner

Q-learning for Automated Personnel Scheduling

Ben Platten
University of Amsterdam
Amsterdam, The Netherlands
ben.platten@student.uva.nl

KEYWORDS

Deep Q-learning, automated personnel scheduling

1 INTRODUCTION

This document describes an exploratory data analysis (EDA) performed as part of a Master's thesis investigating *to what extent can Deep Q-learning optimise personnel scheduling problems*.

Due to the difficulty of training Deep Reinforcement Learning algorithms in general, and for training them to optimise combinatorial optimisation problems specifically, this project will make use of simplified problems made from synthetic data. These problem formulations will be inspired by those used by [1] and will be described in the *Synthetic Data* section.

The *iPlan Data* section explores data from Randstad's iPlan personnel scheduling platform. Since this thesis is sponsored by Randstad Groep, we will attempt to create simplified problems that reflect the Randstad use case. The data is examine from the perspective of pools, shifts and employees and the analysis shows that some real scheduling scenarios may be simple enough to be optimised by an RL agent.

2 SYNTHETIC DATA

[1] use five different problem settings with increasing difficulty. The elements of a problem are; the required number of workers W, the number of tasks to be completed T and the events E which contain the "start time" and "location" features for the tasks. The base configuration is 3 Workers - 2 Tasks - 2 Events (3W-2T-2E). The most complicated configuration is 7W-6T-6E.

3 IPLAN DATA

iPlan is a Randstad proprietary tool for managing schedules. Each row of data is some combination of employee and shift; it is a timestamped record of whether the employee was contacted, assigned, or removed with regards to a particular shift. All employees are pre-filtered into pools based on skill set and physical location. All members of a pool are eligible for shifts assigned to that pool.

This EDA is based on 1 week of iPlan data from Monday 14th - Sunday 20th February 2022.
Number of shifts: 69,118
Number of employees: 43,562
Number of pools: 442

3.1 Pools

Pools are differentiated by the "pool_id" variable. 29 pools with over 500 shifts in one week (the highest had over 2000 shifts) were removed as outliers to improve the distribution plots. 1 additional pool with over 1500 employees was also removed.

Shifts per pool: min: 2, max: 498, mean: 103.6, median: 66.0
Employees per pool: min: 1, max: 1511, mean: 75.0, median: 42.0

Figure 1 shows 3 distribution plots for variables calculated at "pool_id" level. We can see that the number of shifts and the number of employees in a pool is not normally distributed: it is much more likely that the number of shifts and employees in a pool will be below the mean values of 104 and 75, respectively. The average number of shifts per day follows a similar distribution and has a mean of 15.

By looking at pools with an average of less than 7 shifts per day, we can see shift data of comparable complexity to that of [1]. There are 169 pools with an average of less than 7 shifts per day.

3.2 Shifts

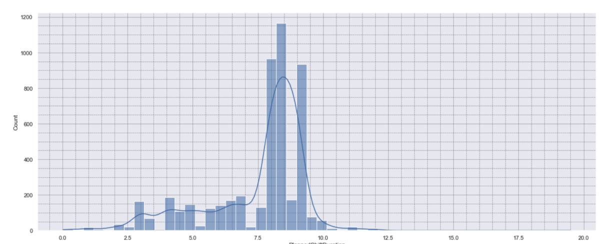

Figure 2: Distribution of shift duration

Figure 2 shows that the are usually around 7 hours in duration.

Distribution of shifts per day of the week:
Friday 13297
Monday 13661
Saturday 4174
Sunday 3790
Thursday 13990
Tuesday 13873
Wednesday 13721

3.3 Employees

Shifts per employee: min: 0, max: 95, mean: 3.4761038494362846, median: 4.0
Accepted shifts per employee: min: 0, max: 15, mean: 1.7, median: 0.0
Accepted shifts per employee: min: 0, max: 3, mean: 0.7, median: 1.0
Number of employees in more than 1 pool: 247 (0.6%)

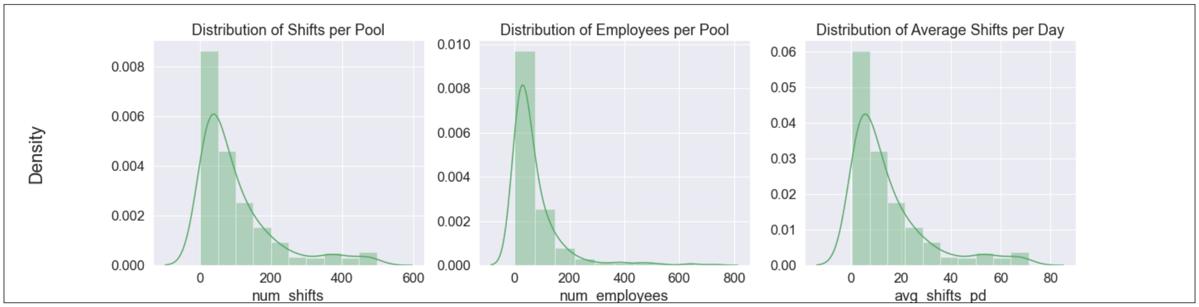

Figure 1: Distributions of Pool level features

REFERENCES

[1] AAAI Press Staff, Pater Patel Schneider, Sunil Issar, J Scott Penberthy, George Ferguson, Hans Guesgen, Francisco Cruz, and Marc Pujol-Gonzalez. [n.d.]. Can Reinforcement Learning solve the Human Allocation Problem? ([n. d.]), 9.

Appendix A TITLE OF YOUR APPENDIX

Put your appendix here.