Learning Patterns in Train Position Data

Automatic Detection of Whether a Solution of the Train Unit Shunting Problem (TUSP) is a Week or a Weekend Day

Supervisor: Issa Hanou[1] Author: Ilia Tomov Responsible Professor: Mathijs de Weerdt[1]

[1]Delft University of Technology

Introduction

- When not in use, trains are parked at shunting yards (Figure 1).
  - Planning movements and tasks on these shunting yard is commonly known as the **Train Unit Shunting Problem (TUSP)** [1].
  - The TUSP is an NP-hard problem.
- Currently, most planning is done by humans.
  - Sensitive, time-consuming task that is prone to errors.
  - The utilization of shunting yards is growing ⇒ more difficult.
- Automated planning tools help human planners.
  - Take as input a timetable with the arrival and departure times of trains and a description of the structural layout of the shunting yard.
  - Output a feasible shunting plan.
- However, current automated planning models are not able to distinguish valuable patterns in the data, as opposed to human planners [2].
- Therefore, the aim of this paper is to study train data for the presence of patterns.
  - This data contains train positions for different timestamps, which can be combined to reconstruct the routes of individual trains.
  - Together, these routes construct a solution to the Train Unit Shunting Problem.

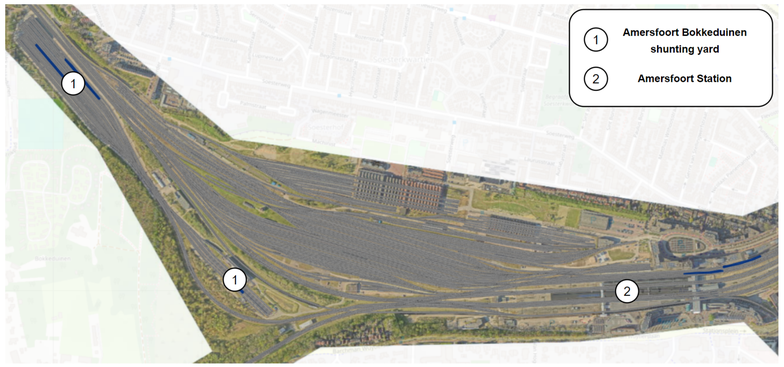

Figure 1. Railway hub in Amersfoort, NL.

Research Question

How to automatically detect whether a solution to the Train Unit Shunting Problem is a week or a weekend day?

- *What is the distribution of the number of parked trains in a shunting yard during a week/weekend day?*
- *What types of trains are parked most often during a week/weekend day?*
- *What tracks are most used for shunting during a week/weekend day?*
- *For each type of train, what tracks are most used for shunting this type during a week/weekend day?*

Methodology

- Data Preprocessing:
  - Train positions for different timestamps are combined to reconstruct the paths of individual trains.
  - Solution data structure: contains train routes and represents a solution of the TUSP.
- Pattern Extraction:
  - Due to the large amount of data, a small sample is used that consists of seven days for each day of the week.
  - Only one area and one shunting yard are considered.
  - The data is examined for the presence of four types of patterns that correspond to the research subquestions.
- Binary Classification of Solutions to the TUSP:
  - The main research question can be viewed as a binary classification task on TUSP solutions.
  - Several binary classification methods are employed.

Results

- Pattern Extraction:

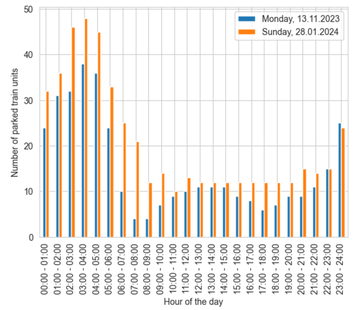

Figure 2. Distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard.

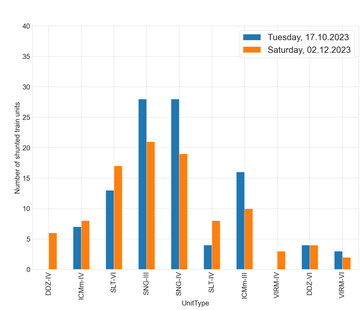

Figure 3. Number of parked trains per unit type in the Amersfoort Bokkeduinen shunting yard.

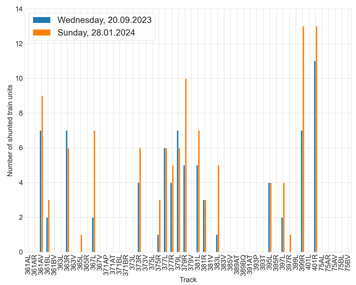

Figure 4. Number of parked trains per track in the Amersfoort Bokkeduinen shunting yard.

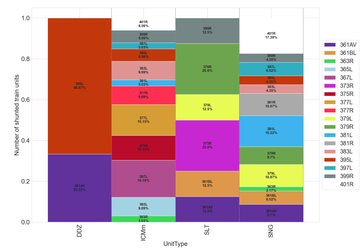

Figure 5. Distribution of the Number of Parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Thursday, 08.06.2023.

Results

- Binary Classification of Solutions to the TUSP:

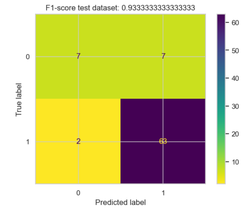

Figure 6. Confusion matrix and F1-score of the logistic regression model.

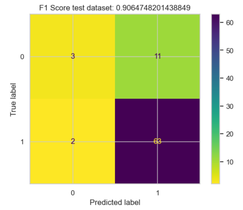

Figure 7. Confusion matrix and F1-score of the support vector machine model.

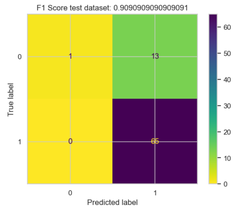

Figure 8. Confusion matrix and F1-score of the random forest model.

Discussion

- Pattern Extraction:
  - Patterns in the distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard were found, that differ between week and weekend days.
  - Patterns corresponding to the other research subquestions were not found.
  - Since only a small part of the data was used in the experiment, the presence or absence of patterns in the entire dataset cannot be concluded.
  - To prove or disprove the existence of patterns in the entire dataset, more extensive research is required.
- Binary Classification of Solutions to the TUSP:
  - The employed binary classification methods reach reasonable performance, however, misclassify a large portion of weekend solutions, probably due to class imbalance in the data.
  - The results proposed by the Logistic Regression model suggest the presence of inherent patterns in the data, that allow for the classification of week and weekend day solutions.

Conclusions & Future Work

- Train data has been studied for the presence of patterns.
- Binary classification has been performed on solutions to the TUSP.
- Results might differ for different areas and shunting yards of interest.
- The insights obtained throughout this research could be applied to improve the real automated planning process by designing new planning heuristics.
- Future work: examine the data in more detail, possibly augmenting it with further insights.

References

[1] R. Freling, R. Lentink, L. Kroon, and D. Huisman, "Shunting of passenger train units in a railway station," *Transportation Science*, vol. 39, no. 2, pp. 261–272, May 2005.

[2] L. van de Gevel, "How human knowledge can support algorithmic decision-making in the train unit shunting problem - an exemplary study," 2022.

I.D.Tomov@student.tudelft.nl