Predicting nurse rosters with machine learning techniques

Shayekh Hassan[1[0000-0002-6196-0318]], Nadia Cissen[1[0000-0002-6792-8925]] and Leendert Kok[1]

[1] ORTEC B.V., Houtsingel 5 2719 EA Zoetermeer, The Netherlands
shayekh.hassan@ortec.com

Abstract. Optimizing nurse rosters is a challenge in practice. A large number of labor rules, many individual preferences, and fuzzy objectives make it hard and cumbersome to create the right optimization model with all the relevant data. Since there are a lot of data patterns in nurse rosters, we tried a different approach using machine learning. We implemented supervised machine learning techniques to predict nurse rosters for a medical center by training our models on past rosters. The medical center uses as a rule of thumb that a roster is good if at least 80% of the roster is executed as planned. In our computational experiments, we found the best results with ensemble learning with an accuracy of over 90%. We consider this a remarkable result, given that the machine learning models have zero explicit knowledge of labor rules, preferences, roster objectives, occupancy requirements, and availabilities.

Keywords: Nurse rostering, Supervised learning, Roster prediction

1 Introduction

ORTEC Workforce Scheduling (OWS) is a leading employee rostering solution for various industries. Traditionally, OWS creates and optimizes rosters based on hard and soft constraints [1]. However, it has been observed that many users of OWS, especially in the healthcare industry, make significant changes to the optimized rosters or create rosters even entire manually [2]. The main reason for this is that there are typically many tacit roster preferences and criteria of planners and nurses that are too cumbersome for the users to include in OWS as constraints [2]. Next to that, there are many data patterns in healthcare rosters, such as working several night shifts in a row, having an entire weekend on or off, certain colleagues typically working together, etc. [2]. These two observations raised the question: would it be possible to predict rosters by learning from the past ones?

2 Model and solution approach

For this research, we obtained a dataset from a large medical center in the Netherlands. We choose the neurology department to create the supervised learning models, and we use two other departments, cardiology and IC Nurses, to verify the robustness of the models. We choose these departments because:

- They have a high number of nurses (Neurology: 186, Cardiology: 357, and IC Nurses: 258).
- The cardiology department is similar to the neurology department in terms of planning difficulty.
- According to planners, the IC nurses department is the most difficult one to plan.

We use data from 2019 and 2020, because further in the past, nurses and shifts were quite different from now, and therefore can contribute very little in predicting current rosters. We excluded data from 2021, as those rosters were not finalized yet at the time of data availability.

We categorized the shifts in the data by their start time into five different categories: day off, early, day, late, and night shift. The motivation for categorization is that the number of unique shifts for a department is very high, with only minor differences in start and end time. For example, the Neurology department has 49 shifts between 2019 and 2020. If it is possible to correctly predict the shift category a nurse will work, we will be able to further narrow down the unique shift the nurse works based on other deterministic methods (e.g., matching required and available skills) or further investigation using machine learning techniques. Therefore, the scope of this study has been limited to predicting nurse rosters based on the above-mentioned five categories of shifts only.

We choose random forest as the first supervised learning method for this study, primarily because of its ability to perform without much feature selection [3] and handle discrete data well [4]. We train the model by taking the roster of a certain day as the output and the preceding days as input. For input, we experimented with different horizons: 7 days, 30 days, 91 days, 183 days, and 366 days.

For predicting the whole roster of a month, we start with predicting the first day of that month using the preceding days as input. For the second and subsequent days of the month, we include the preceding predicted days in the input and iteratively construct this way for the entire month.

3 Computational experiments

We split the data such that we could use the roster of December 2020 as the test set and the preceding 12 months of data as the training set. Since this is a time-series data, we applied cross-validation on a rolling basis in 5 steps [5]. We found that the random forest model predicts the rosters the best when the whole preceding year is used as input. For this input (366 days), the weighted F_1 scores for the three departments are 0.8505, 0.8904, and 0.6598 (See Figure 1).

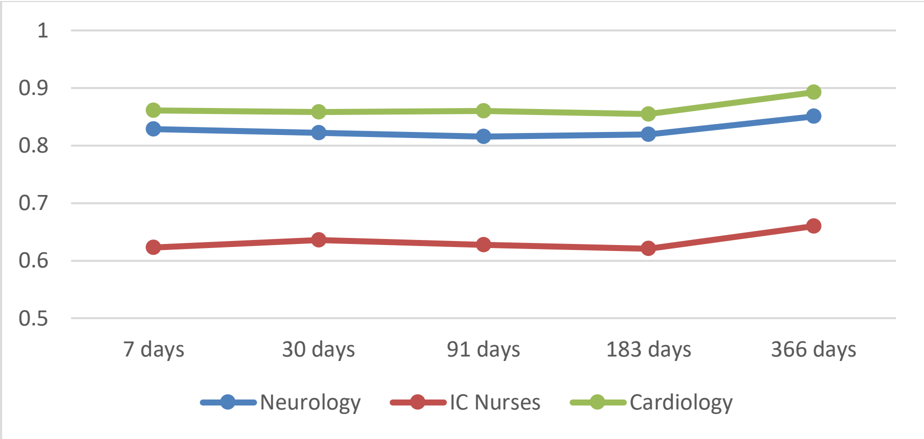

Fig. 1. Weighted average F_1 score of random forest model

We also developed gradient boosting and K-nearest neighbor (KNN) methods to compare the results. We used similar data processing methods as with the random forest model, and we used the same input and output structure. These models provided similar results: the highest weighted average F_1 scores are achieved when the input is 366 days, and the performance is the lowest for the department of IC nurses. Gradient boosting performed marginally better than the other two.

We also developed an ensemble learning model, combining the random forest, gradient boosting and KNN with maximum voting [6]. The results indicate that ensemble learning has a significantly higher weighted average F_1 score compared to the other models. For input of 366 days, the weighted F_1 scores for the three departments are 0.924 0.7432, and 0.9304.

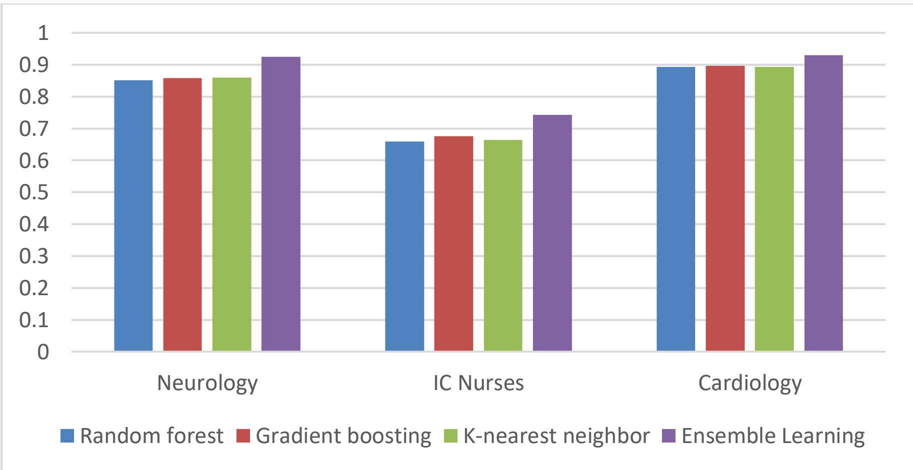

Fig. 2. Weighted average F_1 score of all models with 366 days as input

4 Conclusions

To conclude, supervised machine learning models, especially when multiple non-parametric models are combined, can predict nurse rosters to a high degree of accuracy based on past rosters of a year. Computational experiments indicate that accuracy and F_1 score of over 90% could be reached for certain departments. According to the planners, if the final roster is at least 80% similar to the planned roster, it is already considered good in practice. There is even more to win by including more information in the predictions, for example, labor rules, already scheduled holidays, etc. Further research will be conducted on applying these models to other customer data, improving these models' performances, and creating new models to predict more details (i.e., the exact shift a nurse will work).

5 References

1. Van Draat, L. F., Post, G., Veldman, B., & Winkelhuijzen, W: Harmonious personnel scheduling. Medium Econometrische Toepassingen, 14, 4-7 (2006).
2. Weersel, S. Preference learning: What defines an optimal shift schedule?. MS thesis. University of Twente, Twente (2019).
3. Fawagreh, K., Gaber, M.M., Elyan, E.: Random forests: From early developments to recent advancements. Systems Science & Control Engineering. 2, 602–609 (2014).
4. Janitza, S., Tutz, G., Boulesteix, A.-L.: Random Forest for ordinal responses: Prediction and variable selection. Computational Statistics & Data Analysis. 96, 57–73 (2016).
5. Hyndman, R.J., Athanasopoulos, G.: 5.10 Time series cross-validation. In: Forecasting: Principles and practice. Otexts, Lexington, Ky (2021).
6. Rokach, L.: Pattern classification using ensemble methods. World Scientific, Hackensack, New Jersey (2010).