Predicting employee absenteeism to generate robust rosters

Pieter Smet[1], Martina Doneda[2,3], Giuliana Carello[2], Ettore Lanzarone[4], and Greet Vanden Berghe[1]

[1] KU Leuven, Department of Computer Science, CODeS, Gent, Belgium
[2] Politecnico di Milano, Department of Electronics, Information and Bioengineering, Milan, Italy
[3] National Research Council, Institute for Applied Mathematics and Information Technologies, Milan, Italy
[4] University of Bergamo, Department of Management, Information and Production Engineering, Dalmine (BG), Italy

Abstract. Employee absences often lead to disruptions in rosters, necessitating last-minute changes to employee schedules. A common strategy to minimize the adverse effects of these changes is to assign employees to on-call duties, thereby increasing the robustness in the rosters. This study explores the effectiveness of a data-driven robust rostering approach, using predictions of employee absences to schedule an appropriate number of on-call duties. Numerical experiments demonstrate how the accuracy of absence predictions significantly impacts the robustness of the resulting rosters. We introduce a methodology to assess the conditions under which a data-driven robust rostering approach can outperform simple, non-data-driven rostering strategies.

Keywords: Personnel rostering, Employee absenteeism, Robustness, Proactive rostering, Machine learning.

1 Introduction

Employee absenteeism is the term used to describe when an employee is not present at work during their scheduled hours. Intertwined factors such as health issues, difficulties in achieving work-life balance and instances of workplace harassment can all contribute to employee absenteeism. Whatever the root cause, absenteeism typically has several negative effects on organizations: reduced productivity, additional costs from overtime or from hiring and training replacement employees, low team morale and job disengagement [2].

There have been several studies on how to make personnel rosters more robust with respect to disruptions caused by employee absenteeism. The most common approach is to include *buffers* in the roster that manage unexpected absences through the use of surplus resources. Capacity buffers involve assigning more employees than required to a shift [3]. Meanwhile, reserve shift buffers are created by assigning a subset of employees to special on-call duties which can be converted into working shifts to cover for absences [4].

Approaches employing buffers typically have one or more parameters to set buffer size, thereby affecting the degree of robustness of the generated roster. These parameters are usually set by a human expert or based on results from extensive empirical studies. In our work, we investigate how a Machine Learning (ML) model for predicting employee absenteeism can help determine a suitable number of reserve shifts. More specifically, we analyze under which conditions an ML-informed robust rostering approach can outperform non-data-driven approaches that schedule a fixed number of reserve shifts on each day.

2 Problem definition

The considered personnel rostering problem is based on a general problem definition [1]. The goal is to find an assignment of shifts to employees subject to various personal and organizational constraints. Table 1 provides an overview of the problem's hard and soft constraints. The roster of the preceding scheduling period is taken into account to correctly evaluate the constraints at the beginning of the current scheduling period. Robustness is ensured by including a number of reserve shifts in the roster on each day of the scheduling period. The objective function is a weighted sum of the scheduled employee wage costs (regular, overtime and on-call), the wages of interim personnel needed to cover any understaffing and a penalty term for assigning fewer reserve shifts than required.

Hard constraints
At most one shift assignment per day per employee
Skill requirements
Forbidden shift succession (e.g. no early after late shift)
Minimum number of days worked per employee
Maximum number of consecutive working days
Maximum number of consecutive night shift assignments
Shift and day off requests
Soft constraints
Minimum staffing requirements for each day, shift and skill
Maximum number of days worked per employee
Number of reserve shifts in the roster

Table 1: Hard and soft constraints in the problem.

3 ML-informed robust rostering

The problem described in Section 2 is modeled as an integer programming problem and solved using Gurobi 10.0.3. The number of reserve shifts required on each day is determined by an ML model. In contrast to other studies, we do not actually train an ML model. Instead, we propose a way of *simulating* the predictions a model would make at a given prediction performance level, i.e., given the ground truth and a performance characterization of the ML model, our methodology derives what predictions the model would make. These predictions are used to determine the number of reserve shifts that must be included in the roster. Predicting whether or not an employee will be absent on a given day is a binary classification problem. We use the True Positive Rate α and True Negative Rate β to characterize the prediction performance of the ML model. Figure 1 shows a confusion matrix, used to compare ground truth and model predictions for binary classification problems. Given a confusion matrix, we can compute $\alpha = TP/(TP+FN)$ and $\beta = TN/(TN + FP)$.

		Ground truth	
		Positive	**Negative**
Predicted	**Positive**	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

Fig. 1: Confusion matrix.

The probability that an employee is absent on a given day, derived from historical data, is denoted by ρ. Note that we do not consider employee-specific absence probabilities, but instead use the average over all employees. The correct prediction of an absence by the ML model depends, with probability α, on whether the realization will be correctly classified as a True Positive. With probability $1 - \alpha$, a true absence will result in a False Negative. Similarly, if the employee is not absent, this will be considered a potential False Positive with probability $1 - \beta$. Any potential False Positive will become a proper False Positive with probability ρ, so that the overall number of absences will be reasonable even when β is very small. Each time the prediction results in a True Positive or a False Positive, the number of reserve shifts required is increased by one. In case of True Negatives or False Negatives, the number of reserve shifts is unaffected.

The robustness of the generated roster is measured by the expected re-rostering cost. This cost is computed by running several simulations in which employees become absent and the roster is repaired using an exact re-rostering method. The re-rostering costs obtained for different values of α and β are recorded and compared against those obtained by a non-data-driven baseline approach. More specifically, we compare against an approach from the literature that has no knowledge about what will happen and instead schedules a fixed number of reserve shifts on each day [4]. The results of numerical experiments enable us to identify for which levels of sensitivity and specificity better solutions are generated. Detailed results will be presented at the conference.

Acknowledgements This research was supported by KU Leuven C24E/23/012 “Human-centred decision support based on new theory for personnel rostering”. Editorial consultation provided by Luke Connolly (KU Leuven).

References

1. Ceschia, S., Dang, N., De Causmaecker, P., Haspeslagh, S., Schaerf, A.: The second International Nurse Rostering Competition. Annals of Operations Research **274**(1-2), 171–186 (2019)
2. Hudson, C.K., Shen, W.: Understaffing: An under-researched phenomenon. Organizational Psychology Review **5**(3), 244–263 (2015)
3. Ingels, J., Maenhout, B.: Optimised buffer allocation to construct stable personnel shift rosters. Omega **82**, 102–117 (2019)
4. Wickert, T.I., Smet, P., Vanden Berghe, G.: Quantifying and enforcing robustness in staff rostering. Journal of Scheduling **24**(3), 347–366 (2021)