Setup time prediction using machine learning algorithms : A real-world case study

Manuel Iori[1], Marco Lippi[1], Alberto Locatelli[1], Marco Locatelli[2]

[1] Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Italy

{manuel.iori,marco.lippi,alberto.locatelli}@unimore.it

[2] Department of Engineering and Architecture, University of Parma, Italy

marco.locatelli@unipr.it

Mots-clés : *Uncertain machine-dependent and job sequence-dependent setup times, Setup time prediction, Machine learning algorithms.*

1 Abstract

According to Allahverdi and Soroush [3], *setup time* (ST) can be defined as the time required to prepare the necessary resource (e.g., machines, people) to perform a task (e.g., a job, an operation). Setup operations may include, for instance, switching tools, cleaning up, changing material, adjusting tools, etc. As all these operations are often strongly time-consuming and may take a considerable part of the entire production time, the reduction of STs plays a crucial role in scheduling. The deep connection between scheduling activities and STs was analyzed in detail by Allahverdi [1], who provided a comprehensive survey on scheduling problems involving STs.

Although, in many real-world manufacturing environments, STs are influenced by various and random factors (e.g., crew experience, breakdowns of a tool or a machine, lack of personnel, complex and non-fixed procedures, etc.), in the extant literature, the vast majority of scheduling studies consider STs as a given input taking a deterministic value, calculated by means of simplistic average-based methods that are not capable to catch the complexity of reality. As already highlighted by Kim and Bobrowski [7], assuming stochastic STs as fixed and constant values may lead to develop inefficient schedules. This is also confirmed by practitioners, who often view the lack of uncertainty and dynamic elements in the modeling of scheduling processes as the major source of gap between scheduling theory and practice (see Sabuncuoglu and Goren [8]). Only a few of studies, recently reviewed by Allahverdi [2], took into account the issues related to the STs uncertainty. These studies mainly focused on robust optimization approaches (modeling STs as a probability distribution, or a fuzzy number, or a random variable within some interval), but did not provide any methods to estimate these times.

In this work, we seek to fill this gap of the literature by provide models that can be used to provide good estimations of the STs starting from a set of empirical data. In detail, we use *machine learning* (ML) regression algorithms to predict the STs, and we apply them to a real-world scheduling application arising in the color printing industry, where a finite set of jobs must be sequentially performed by a heterogeneous set of parallel flexographic printer machines. Specifically, we deal with *uncertain machine-dependent and job sequence-dependent setup times* (UMDJSTs) with an additional issue : the UMDJST between two jobs not only depends on the two jobs and the involved machine, but also on all jobs previously scheduled on that specific machine, owing to tool configurations (see, e.g., Soares and Carvalho [9]). Indeed, jobs have different tooling requirements and it is often beneficial to leave a certain tool unused in a machine magazine only to use it again a few jobs later (see Iori et al. [4, 5, 6]).

In the addressed real-world industrial application, UMDJSTs are strongly time-consuming (around 65% of the total production time) and subject to significant uncertainties, because depending on many factors : characteristics of the specific machine, status of the machine (which directly depends on the jobs already processed), characteristics of the current job, operators' experience, etc. It turns out that the problem of predicting such times is not trivial and may require a large computational effort (see Iori et al. [6]). To be able to solve practical industrial instances, Iori et al. [6] introduced a heuristic evaluation method, which expresses UMDJSTs in terms of some specific and pre-fixed features (i.e., characteristics of jobs and machines). This analytic approach suffers from intrinsic limitations and cannot account for the whole complexity of the problem (see [6]).

To improve the accuracy of UMDJSTs prediction, we aim at exploiting the knowledge extraction capability of the ML algorithms to be exploratory both in selecting the significant features and expressing the UMDJSTs in terms of these features. Using a real-world industrial database, we train three different ML models : linear regression, random forests, and gradient boosting machines. For each model, we take into account a wide set of features and several possible inter-dependencies amongst them, in order to identify a parsimonious but comprehensive subset of these features. We compare the three models with the heuristic evaluation method introduced by Iori et al. [6] on a real-world industrial test set. The experimental results demonstrate that the gradient boosting machine approach obtains the best performance overall, immediately followed by random forests. For both models, the mean squared error on the predicted UMDJSTs is less than half of that of the heuristic evaluation method, proving their effectiveness in modeling the application. The results are of interest because the ML models can be easily adapted to deal with ST evaluation in many other scheduling problems.

Références

[1] A. Allahverdi. The third comprehensive survey on scheduling problems with setup times/costs. *European Journal of Operational Research*, 246(2) :345–378, 2015.

[2] A Allahverdi. A survey of scheduling problems with uncertain interval/bounded processing/setup times. *Journal of Project Management*, 7 :255–264, 2022.

[3] A. Allahverdi and H.M. Soroush. The significance of reducing setup times/setup costs. *European Journal of Operational Research*, 187(3) :978 – 984, 2008.

[4] M. Iori, A. Locatelli, and M. Locatelli. Scheduling of parallel print machines with sequence-dependent setup costs : A real-world case study. In *IFIP International Conference on Advances in Production Management Systems*, pages 637–645. Springer, 2021.

[5] M. Iori, A. Locatelli, M. Locatelli, and J. J. Salazar-González. Tool switching problems in the context of overlay printing with multiple colours. In *Combinatorial Optimization*, pages 260–271. Springer, 2022.

[6] M. Iori, A. Locatelli, and M. Locatelli. A GRASP for a real-world scheduling problem with unrelated parallel print machines and sequence-dependent setup times. *International Journal of Production Research*, 2022 (forthcoming).

[7] S.C. Kim and P.M. Bobrowski. Scheduling jobs with uncertain setup times and sequence dependency. *Omega*, (4) :437 – 447, 1997.

[8] I. Sabuncuoglu and S. Goren. Hedging production schedules against uncertainty in manufacturing environment with a review of robustness and stability research. *International Journal of Computer Integrated Manufacturing*, 22(2) :138 – 157, 2009.

[9] L.C.R. Soares and M.A.M. Carvalho. Biased random-key genetic algorithm for scheduling identical parallel machines with tooling constraints. *European Journal of Operational Research*, 285(3) :955 – 964, 2020.