Solving stochastic machine scheduling problems by estimating the solution value within local search

Han Hoogeveen (Speaker) *
Marjan van den Akker, Guido Passage, and Jan Posthoorn †

1 Introduction

One of the standard assumptions in machine scheduling is that the processing times are deterministic. In many situations, this assumption does not cover reality, and therefore more and more researchers have started to work on stochastic scheduling problems.

In this presentation, I want to address the problem of how to find a good solution for problems with stochastic processing times for parallel machine scheduling problems with precedence relations. The standard approach for finding a not necessarily optimal solution in case of deterministic processing times is to apply some form of local search. In case of stochastic processing times, it is much harder to assess the quality of a move from the current solution to some neighbor. One possibility to compare the objective values is to run a discrete event simulation to get an estimate, as proposed by van den Akker et al. [1] for the stochastic job shop problem. The down-side of this approach is that a sufficient number of simulation runs are required to get a reliable estimate. To alleviate this computational burden, we have investigated algorithms to compute good estimates in less time than required by discrete event simulation, without reducing the quality of the obtained solutions. Below we will describe the algorithm that performed best.

2 Problem description

We consider the following scheduling problem. There are n jobs that have to be carried out by a set of m identical machines; processing job j takes time p_j. It is not allowed to start job j before its release date r_j. Furthermore, there are some *minimum delay precedence constraints*, which specify that the execution of job j

* `j.a.hoogeveen@uu.nl` Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

† `J.M.vandenAkker@uu.nl, g.j.p.n.passage@gmail.com, jiposthoorn@gmail.com`. Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

cannot start until a given amount q_{ij} of time has passed since the completion of job i. We assume that the values r_j and q_{ij} are deterministic, but the processing times are stochastic variables. Our initial objective function is to minimize the makespan (see also [5]), but we are currently analyzing the problem of minimizing total weighted tardiness (for which we add due dates and weights to the instance). We want to find a feasible *baseline schedule* with minimum objective function, that is, when executing the schedule, we are allowed to adjust the start times of the jobs by starting them as soon as possible, but we are not allowed to change the assignment of the jobs to the machines and the order in which the jobs are executed. As a result, to specify a schedule we need to determine for each machine which jobs it executes and in which order this is done; the start time S_j $(j = 1, \ldots, n)$, and hence its completion time C_j, are then determined by starting job j as soon as possible, which value becomes known during the execution of the schedule.

As our local search approach we use Iterated Local Search (see [2]) in combination with Variable Neighborhood Descent (see [3]). We estimate the value of the objective value by estimating the starting time S_j of each job j $(j = 1, \ldots, n)$; the start time S_j of job j is equal to the maximum of a number of stochastic variables, which are:

- the release date r_j;
- the values $C_i + q_{ij}$ for all predecessors i of j;
- the completion time of the machine predecessor of j.

Denoting these stochastic variables by $D_1, \ldots, D_l$, we find that $S_j = \max\{D_1, \ldots, D_l\}$. Next, we use [4], who describe how to find the expected value and variance of $X = \max\{D_1, D_2\}$, where D_1 and D_2 are two normally distributed stochastic variables with correlation coefficient ρ. Pretending that all these stochastic variables D_k are normally distributed, we estimate $E[S_j]$ and $Var[S_j]$, by iteratively computing $X_k = \max\{X_{k-1}, D_k\}$, where $X_1 = D_1$; again, we pretend that the maximum is normally distributed. We found that the quality of the approximation depends on the order in which we consider the variables D_k; it is best to first handle the variables D_k describing precedence constraints involving jobs scheduled on the same machine as job j starting with the machine predecessor of job j, then take the remaining precedence constraints into consideration, and end with r_j. For stochastic variables corresponding to jobs on different machines we use $\rho = 0$; for the other ones, the stochastic variables describing the part after the last common time-point are unrelated again.

Computational experiments reveal that local search with the above algorithm finds solutions that are just as good as local search with estimating the makespan using 300 simulation runs per iteration, but only requires a fraction of the computation time. This suggests that estimating the makespan works better and faster than simulation in local search for stochastic parallel machine scheduling problems. We are currently looking whether this approach can be applied to the problem of minimizing total weighted tardiness as well.

References

[1] J.M. van den Akker, K. van Blokland, and J.A. Hoogeveen (2013). Finding robust solutions for the stochastic job shop scheduling problem by including simulation in local search. In V. Bonifaci, C. Demetrescu, and A. Marchetti-Spaccamela (Eds.) In *Experimental Algorithms – SEA 2013*, Vol. 7933 of Lecture Notes in Computer Science, pp. 402–413. Springer Berlin Heidelberg

[2] H.R. Lourenço, O.C. Martin, and O. Stützle (2003). Iterated Local Search. In F. Glover and G.A. Kochenberger (Eds.) *Handbook of Metaheuristics.* Springer, Boston.

[3] N. Mladenović and P. Hansen (1997). Variable neighborhood search. *Computers and Operations Research 24*, pp. 1097–1100.

[4] S. Nadarajah and S. Kotz (2008). Exact distribution of the max/min of two Gaussian random variables. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 16*, pp. 210-212.

[5] G.J.P.N. Passage, J.M. van den Akker, and J.A. Hoogeveen (2019). A new, efficient approach to estimate starting times to improve the performance of local search for stochastic parallel machine scheduling. Submitted to *Computers and Industrial Engineering.*