European Journal of Operational Research 246 (2015) 413–420

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

On heuristic solutions for the stochastic flowshop scheduling problem

Jose M. Framinan*, Paz Perez-Gonzalez

Industrial Management, School of Engineering, University of Seville, Camino de los Descubrimientos s/n, 41092 Seville, Spain

ARTICLE INFO

Article history:
Received 13 August 2014
Accepted 3 May 2015
Available online 8 May 2015

Keywords:
Scheduling
Flowshop
Stochastic
Makespan objective
Heuristics

ABSTRACT

We address the problem of scheduling jobs in a permutation flowshop when their processing times adopt a given distribution (stochastic flowshop scheduling problem) with the objective of minimization of the expected makespan. For this problem, optimal solutions exist only for very specific cases. Consequently, some heuristics have been proposed in the literature, all of them with similar performance. In our paper, we first focus on the critical issue of estimating the expected makespan of a sequence and found that, for instances with a medium/large variability (expressed as the coefficient of variation of the processing times of the jobs), the number of samples or simulation runs usually employed in the literature may not be sufficient to derive robust conclusions with respect to the performance of the different heuristics. We thus propose a procedure with a variable number of iterations that ensures that the percentage error in the estimation of the expected makespan is bounded with a very high probability. Using this procedure, we test the main heuristics proposed in the literature and find significant differences in their performance, in contrast with existing studies. We also find that the deterministic counterpart of the most efficient heuristic for the stochastic problem performs extremely well for most settings, which indicates that, in some cases, solving the deterministic version of the problem may produce competitive solutions for the stochastic counterpart.

1. Introduction

The flowshop scheduling problem with makespan objective (usually denoted as $Fm|prmu|C_{max}$) has been subject of research for more than 60 years, being one of the most comprehensively studied problems in Operations Research (see in this regard the reviews by Framinan, Gupta, & Leisten, 2004; Reza Hejazi & Saghafian, 2005 and Ruiz & Maroto, 2005). This decision problem consists of how to schedule jobs in a permutation flowshop in order to minimize the maximum completion time or makespan. A classical assumption is that the processing times of each job in each machine are considered different, but known in advance (deterministic). In contrast, our paper deals with the problem of scheduling n jobs in a permutation flowshop consisting of m machines where the processing times are not deterministic, but follow some known distribution. The objective considered is that of minimizing the expected makespan. This problem is considered to be more realistic that their deterministic counterpart, as it allows capturing part of the inherent variability present in many real-life manufacturing environments (see e.g. Hopp & Spearman, 2008). In the following, we will denote our problem as $Fm|prmu|E[C_{max}]$.

The $Fm|prmu|E[C_{max}]$ problem has been much less studied than its deterministic counterpart, and it is clearly much more complex. In fact, apart from a dominance rule obtained by Makino (1965) for the case of two jobs, no exact solution is available without assumptions on the distribution of the processing times. For $m = 2$ and exponential distribution of the processing times, Talwar (1967) conjectured an exact solution for the problem that was later proved to be optimal by Cunningham and Dutta (1973), and is currently known as Talwar's rule.

Despite these advances, for the rest of the cases, no optimal procedure has been found. For the two-machine case, three approximate solutions have been proposed by Baker and Trietsch (2011) based both in Talwar's rule and in Johnson's rule (Johnson, 1954) for the deterministic flowshop, all of them with similar (near optimal) performance. For the general m machine case, (Baker & Altheimer, 2012) suggest three heuristics based on adaptations of the CDS (Campbell, Dudek, & Smith, 1970) and NEH (Nawaz, Enscore, & Ham, 1983) heuristics, again with similar and near optimal performance. Although it might seem that, from these results, the problem $Fm|prmu|E[C_{max}]$ is already solved, some issues have to be discussed:

- First of all, the evaluation of sequences in a stochastic flowshop is far from being a trivial task. Since the objective is to obtain the expected makespan of a given sequence, $E[C_{max}]$ has to be estimated by running N simulations using the sequence as a solution, from

* Corresponding author. Tel.: +34 95 448 7327; fax: +34 95 448 7329.
E-mail address: framinan@us.es, framinan@gmail.com (J.M. Framinan).

http://dx.doi.org/10.1016/j.ejor.2015.05.006
0377-2217/

which a sample C^i_{max} ($i = 1, \ldots, N$) is obtained. Then, $E[C_{max}]$ is estimated by averaging the sample, i.e. $E[C_{max}] \approx \bar{C}_{max} = \frac{1}{N}\sum_{i=1}^{N} C^i_{max}$. Up to now, there is no standardized procedure to determine N, although the authors of related contributions use a large number in order to ensure the significance of the estimation. Thus, Baker and Altheimer (2012) use $N = 100{,}000$ whereas Gourgand, Grangeon, and Norre (2003) employ $N = 200{,}000$ regardless the size and characteristics of each instance, while Portougal and Trietsch (2006) set N to 10,000 for the 2-machine case. In addition, there is no mechanism to establish the statistical significance of the so-obtained $\bar{C}_{max}$ and, consequently, to assess the differences in the performance among different heuristics.
- Due to the computational complexity of the stochastic problem, the experiments in the literature have been limited to very small problem sizes (up to $n = 10$ and $m = 6$ in the most recent studies). This makes the conclusions obtained so far to be restricted to very small problem sizes.
- Finally, to the best of our knowledge, the necessity of heuristics specifically designed for the stochastic problem has not been yet determined. In other words, one may try to solve the stochastic flowshop scheduling problem by transforming it into its deterministic counterpart, i.e. by obtaining a flowshop with the same number of jobs and machines but with deterministic processing times equal e.g. to the means of those from the stochastic problem. Then, heuristics for the $Fm|prmu|C_{max}$ problem can be applied and a (possibly good) sequence for the deterministic problem can be obtained. If this sequence performs well when applied to the stochastic problem, then the need of specific stochastic heuristics can be questioned. However, such test has not been conducted so far. It is foreseeable that the so-obtained sequences perform worse that those specifically designed for the stochastic version, but maybe the differences in the quality of the results do not justify the much higher computation times required for the stochastic heuristics. Even if the deterministic heuristics are not valid for some cases, it would be interesting to quantify the degree of variability for which using them is still acceptable, as it is clear that a stochastic flowshop with low variability would resemble very much to a deterministic flowshop.

With these issues in mind, we first discuss and propose a procedure for estimating the expected makespan of a sequence in a stochastic flowshop, so the error in such estimation is bounded by a given percentage. In this way, the statistical significance of the results obtained by the different heuristics can be more clearly established. Interestingly, the results show that the sample sizes (N) obtained from our procedure proposed range from very small to very high values, thus supporting the conclusion that no predetermined value can be easily found regardless the variability of the instances and the error assumed in the estimation.

Next, we compare the main heuristics proposed in the literature as well as the expected makespan obtained from the application of purely deterministic procedures to the mean processing times of the stochastic problem. These heuristics are tested for problem sizes larger than those presented so far in the literature, so the conclusions from the results can be better supported. The results show that, in contrast to Baker and Altheimer (2012), there are significant differences in the performance of the heuristics, and that – perhaps not so surprisingly – the performance of the sequences obtained from purely deterministic methods in the stochastic flowshop do not differ greatly from that obtained from specific stochastic methods.

The remainder of the paper is organized as follows: First, the problem under consideration is formally described in Section 2, where the main contributions of the literature are discussed. Since our work is of computational nature, we devote Section 3 to discuss the key issue of the testbed in which the heuristics are to be compared, as we intend to capture different problem sizes and different degrees of variability of the flowshop. Next, we present in Section 4 the procedure to estimate the expected mean makespan of a given solution, and compare the number of iterations required with those employed in the literature. The comparison of the performance of different heuristics for the problem is done in Section 5, where the main results are also discussed. Finally, Section 6 present the conclusions and points out future research lines.

2. Background

A flowshop consists of n jobs that must be processed on m machines in the same order, where job i requires p_{ij} time units to be processed on machine j. The scheduling problem in flow shops is to find a sequence of jobs for each machine according to certain performance measure(s). Additionally, for many situations, it is assumed that the job sequences will be the same on every machine (permutation flowshops). Other hypotheses common in scheduling research include the simultaneous availability of all jobs and of all machines, deterministic processing times, etc. For a complete list of these assumptions, see e.g. Framinan, Leisten, and Ruiz-Usano (2005).

While the deterministic flowshop scheduling problem with makespan objective has been extensively studied (see the reviews mentioned above), the same cannot be said about its stochastic counterpart. For the two-jobs case and a general distribution of the processing times, a dominance rule is given by Makino (1965), but this result is extremely restrictive and with little applicability for most practical settings.

By making assumptions on the distribution of the processing times of the jobs, an important result is due to Talwar (1967). He conjectures that the expected makespan is minimized when the processing time of the jobs follows an exponential distribution by sequencing the jobs in non increasing order of $\frac{1}{\mu_{i1}} - \frac{1}{\mu_{i2}}$, where μ_{ij} is the mean processing times of job i on machine j. This order is proved to be optimal by Cunningham and Dutta (1973), and it is currently known as Talwar's rule. As an extension of this rule to other distributions, Kalczynski and Kamburowski (2006) heuristically adapt Talwar's rule for the Weibull distribution. For a general family of distributions, Portougal and Trietsch (2006) develop a heuristic named PSH which starts with the solution given by Johnson's rule (Johnson, 1954) for the deterministic flowshop, and applies an adjacent pairwise interchange (a reason why this heuristic is later renamed API by Baker and Trietsch, 2011). Finally, Baker and Trietsch (2011) test three different procedures for different families of distributions, i.e. Talwar's rule, Johnson's rule, and the API heuristic. They conclude that the (deterministic) Johnson's rule could be better unless the coefficient of variation of the jobs is very high. For such cases, Talwar's rule or API may perform better.

For the general flowshop problem with m machines, Baker and Altheimer (2012) propose different heuristics. The first heuristic is called CDS/Johnson and consists of obtaining a set of $m - 1$ 2-machine flowshop subproblems with the addition of the processing times of the jobs in the manner of the CDS heuristic by Campbell et al. (1970). More specifically, 2-machine flowshop subproblem k (with $k = 1, \ldots, m - 1$) is constructed by obtaining the processing times of job i in the first (second) machine of this subproblem as $A_i = \sum_{j=1}^{j=k} p_{ij}$ ($B_i = \sum_{j=k+1}^{j=m} p_{ij}$). Then, Johnson's procedure is applied to each of the resulting $m - 1$ subproblems, and $m - 1$ sequences are obtained. The estimation of the expected makespan of each of this sequences is obtained (in their case by running 100,000 simulations and taking the average makespan value), and the one yielding the lowest value is selected.

The second tested heuristic is the CDS/Talwar heuristic. In a similar manner to the previous one, a set of $m - 1$ 2-machine flowshop subproblems are obtained, and a sequence is obtained for each one by applying Talwar's rule. Out of these $m - 1$ sequences, the one with the minimum estimation of the expected makespan is selected.