Int. J. Production Economics 173 (2016) 153–160

Contents lists available at ScienceDirect

Int. J. Production Economics

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

Scheduling controllable processing time jobs with position-dependent workloads

Daniel Oron

The University of Sydney Business School, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 9 July 2014
Accepted 21 December 2015
Available online 31 December 2015

Keywords:
Makespan
Total flowtime
Resource consumption function
Workload

ABSTRACT

In various real life applications job processing times are controllable through the allocation of a limited resource. To date research has been conducted under the assumption that the relationship between the amount of resource allocated to a job and its processing time is independent of the number of tasks processed previously. However, there exist many manufacturing and service systems where workers and machines acquire, develop and refine skills through the repetition of identical or similar operations. In this paper we consider a scheduling model where job processing times are a convex function of the amount of resource they are allocated. In addition, we assume that the parameters of this function are position-dependent, i.e., vary with the job's position in the sequence. This assumption reflects general processes of learning or aging, or a combination of both. We first focus on a single machine setting and the makespan and total flowtime criteria. We show that the combined problem of finding an optimal job sequence and an optimal resource allocation can be solved in $O(n^3)$ time. We show that our algorithm can be used to address a bicriteria objective comprising of a linear combination of makespan and the total flowtime criteria on a single machine. We then extend the results to a parallel machine setting for the total flowtime criteria.

1. Introduction

The majority of deterministic scheduling models assume that job processing times remain unchanged throughout the production process. However, in various real life applications, job processing times vary either over time or as a function of the job position in the sequence. The latter model is referred to as the learning or the aging effect, depending on whether job processing times become shorter or longer when scheduled later in the sequence. The motivation for learning and aging models is driven by the existence of an improvement (or deterioration) process which is inherent in various production systems. The learning process is often the result of continuous repetition of similar operations. Both workers and machines are able to refine their skills through reflection and analysis of previous tasks. Moreover, machine processing rates are often determined by exogenous considerations such as the amount of energy they are allocated, their surrounding environment (temperature, humidity, etc.) and the operator assigned to oversee the production process.

In a recent paper, Fang and Lin (2013) study a parallel machine scheduling problem to minimize an objective function which comprises of tardiness penalties and power consumption costs. Their model is motivated by a setting whereby CPU speed is determined by the magnitude of electricity it is allocated. The scheduler's task is to determine the optimal allocation of jobs to machines and the optimal frequency for each machine-to-job pairing. During this process, CPU speeds are continuously adjusted in order to meet energy consumption requirements. Mansouri et al. (2016) introduce the concept of green production in the context of scheduling, studying a two-machine flowshop problem where the objective is to determine the best trade-off between the completion time of a set of jobs and the energy invested in the process. They discuss the scarcity and likely future shortage in a variety of resources, including key materials and energy. The authors develop heuristics to find the optimal scheduling policy for a given restriction on energy consumption. One of the main drivers for this new stream of research is increasing regulatory restrictions on emissions and the introduction of a carbon tax. More than ever, decision makers are now faced with the additional challenge of determining how to allocate energy, and how to best operate under varying allocations of energy.

Motivated by Fang and Lin (2013) and by Mansouri et al. (2016), we consider a setting whereby the scheduler is required to complete a set of jobs subject to limited resources. We assume that the energy allocation to machines is determined at a higher level in the decision making process and is updated at the completion time of each task. Hence, each position in the processing sequence

E-mail address: daniel.oron@sydney.edu.au

http://dx.doi.org/10.1016/j.ijpe.2015.12.014
0925-5273/

is allocated a certain amount of energy, resulting in a position-dependent processing time for each job. Combined with the assumption that there is a general learning or aging process which occurs, job processing times are a general function of their position in the processing sequence. We also assume that there exists a limited resource which is controlled by the scheduler and also affects job processing times. Thus, the scheduler controls one dimension of the processing time by the allocation of one resource to the various jobs; the second factor determining job processing times is controlled through the allocation of jobs to a position in the processing sequence.

Although the learning effect was identified in various fields and applied in industry as early as the first half of the twentieth century (see Wright, 1936), the first paper to consider the learning effect in the context of scheduling was Biskup (1999). The author assumes that the processing time of a job decreases as a function of its position in the sequence according to the following relationship:

$$p_{jr} = p_j r^a \tag{1}$$

where p_{jr} is the processing time of job j when executed in the r-th position in the sequence. p_j is the basic processing time of job j and a is a parameter whose value is less than zero. The parameter a is often referred to as the learning index. The relationship expressed in Eq. (1) reflects a decreasing rate of learning, which is consistent with most real life systems. Biskup's model (Biskup, 1999) was adopted by numerous researchers, including a paper by Mosheiov (2001a) who shows that various regular objectives are minimized when jobs are scheduled according to non-decreasing order of their basic processing times. Mosheiov (2001b) also extends these results to a parallel machine setting. Mosheiov and Sidney (2003) identify the shortcomings of this model; first, they claim that it constraints all jobs to use the same learning rate, or index (the parameter a). Furthermore, it does not allow different types of learning curves. To overcome these limitations, the authors present a general model where each job follows a different learning curve. Actually, the model they develop is not restricted to learning and can cater for any position-dependent processing times. They show that the makespan, the total flowtime and a common due-date assignment problem on a single machine can be solved in $O(n^3)$ time through the formulation of a corresponding linear assignment problem. The authors extend their results to a uniform parallel machine setting for the total flowtime criterion, showing that the problem is solved in polynomial time for a fixed number of machines. In a later paper, Mosheiov and Sidney (2005) present a polynomial time algorithm to minimize the number of tardy jobs for general position-dependent processing times and a common due date. Gawiejnowicz (2008) provides complexity results and solution algorithms for various time-dependent scheduling models.

Another common assumption in classical scheduling is that job processing times are constant parameters and determined exogenously. The scheduler's task focuses on allocating and sequencing the jobs on the different machines. In reality, there exist many applications where job processing times are controlled by the allocation of a non-renewable resource, such as sub-contracting, additional budget, fuel, electricity, fertilizer, etc. Thus, by allocating additional resources to a job its processing time can be shortened. One of the earliest papers to consider controllable processing times was Vickson (1980) who points out that although controllable processing time systems are widespread, they have never been addressed in the context of scheduling. Vickson's paper (Vickson, 1980) has since motivated numerous researchers to consider controllable inputs, including processing times, due-dates, and job release times, under various machine environments. Janiak (1989) describes a problem in steel mills where ingots are processed in batches, first preheated and then hot-rolled in a blooming mill. Both the time required at the preheating stage and that of the hot-rolling process are inversely proportional to the intensity of the gas flow. Kaspi and Shabtay (2003) describe an interesting application in a machine tooling environment. The job processing times depend on the feed rate and spindle speed set at each stage of the production process. A complete state of the art survey on scheduling with controllable processing times by Shabtay and Steiner (2007) describes a wide range of applications with controllable job processing times in various scheduling environments. The survey includes over 100 references, a clear indication of the importance and popularity of this class of problems.

There are two main models which attempt to model the relationship between the amount of resource allocated to a job and its processing time. The first model is linear and defined as follows:

$$p_{ij}(u_{ij}) = \overline{p}_{ij} - a_{ij}u_{ij}, \quad 0 \le u_{ij} \le \overline{u}_{ij} \le \overline{p}_{ij}/a_{ij} \tag{2}$$

where $\overline{p}_{ij}$ is the *non-compressed processing time* for job j on machine i. The variable $\overline{u}_{ij}$ denotes the *upper bound* on the amount of resource which can be allocated to job j, and a_{ij} is its positive *compression rate*. This model has been extensively studied by various researchers (see Shabtay, 2004; Kaspi and Shabtay, 2003; Lee and Lei, 2001; Ng et al., 2005; Shakhlevich and Strusevich, 2006). Note that the linear function does not reflect the *law of diminishing marginal returns* which exists in various manufacturing and production processes. In addition, it allows for scenarios where jobs are not allocated any resources and can still be processed in reasonable time. Due to the fact that the linear relationship does not always accurately describe the effect of resource allocation on processing times, the *convex resource consumption function* is proposed as an alternative:

$$p_{ij}(u_{ij}) = \left(\frac{w_{ij}}{u_{ij}}\right)^k \tag{3}$$

where u_{ij} is the amount of resource allocated to job j on machine i, w_{ij} is a positive parameter which represents the *workload* of job j on machine i, and k is a positive constant (See Alidaee and Ahmadian, 1993; Janiak and Kovalyov, 1996; Shabtay and Kaspi, 2004). Monma et al. (1990) make the observation that in very large scale integration (VLSI) circuit design, the area of the silicon surface is proportional to the time spent on an individual job. In this case the resource is the silicon slice and the value of the constant k is $\frac{1}{2}$. They also identify various industrial processes where the value of k is equal to 1.

In this paper, we study a scheduling model with a convex resource allocation function (as given in Eq. (3)) and a continuous non-renewable resource. By continuous we mean that any amount of available resource can be allocated to any given job. We also assume that the production system experiences a process of learning and that a job's workload depends on its position in the sequence. Thus, the scheduler's task is to simultaneously sequence the jobs while allocating the resource so as to minimize a given criterion. Much to our surprise, research papers which consider the joint existence of these two widespread phenomena in machine scheduling have only appeared in recent years. Koulamas et al. (2010) and Wang et al. (2010) were the first to consider time- and resource-dependent processing times in the context of scheduling. Koulamas et al. (2010) study several single machine scheduling problems with controllable processing times and learning effect or deterioration. Wang et al. (2010) study a similar problem setting and show that minimizing a cost function which comprises of the makespan, the total completion time and the total absolute difference in waiting times can be solved in polynomial time. More recent papers which assume resource allocation under the assumption of linear deterioration include Wang and Wang (2013), Wang and