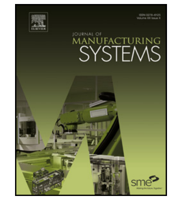
Journal of Manufacturing Systems 38 (2016) 114–133

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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



Technical Paper

A metamodel-based Monte Carlo simulation approach for responsive production planning of manufacturing systems

Minqi Li [a], Feng Yang [a,*], Reha Uzsoy [b], Jie Xu [c]

[a] Industrial and Management Systems Engineering Department, West Virginia University, Morgantown, WV 26505, USA
[b] Edward P. Fitts Department of Industrial & Systems Engineering, North Carolina State University, Raleigh, NC 27695, USA
[c] Department of Systems Engineering & Operations Research, George Mason University, Fairfax, VA 22030, USA

ARTICLE INFO

Article history:
Received 29 March 2015
Received in revised form 8 October 2015
Accepted 15 November 2015
Available online 5 January 2016

Keywords:
Production planning
Metamodeling
Non-stationary time series
Monte Carlo simulation
Multi-objective optimization

ABSTRACT

Production planning is concerned with finding a release plan of jobs into a manufacturing system so that its actual outputs over time match the customer demand with the least cost. For a given release plan, the system outputs, work in process inventory (WIP) levels and job completions, are non-stationary bivariate time series that interact with time series representing customer demand, resulting in the fulfillment/non-fulfillment of demand and the holding cost of both WIP and finished-goods inventory. The relationship between a release plan and its resulting performance metrics (typically, mean/variance of the total cost and the fill rate) has proven difficult to quantify. This work develops a metamodel-based Monte Carlo simulation (MCS) method to accurately capture the dynamic, stochastic behavior of a manufacturing system, and to allow real-time evaluation of a release plan's performance metrics. This evaluation capability is then embedded in a multi-objective optimization framework to search for near-optimal release plans. The proposed method has been applied to a scaled-down semiconductor fabrication system to demonstrate the quality of the metamodel-based MCS evaluation and the results of plan optimization.
© 2015 The Society of Manufacturing Engineers. Published by Elsevier Ltd. All rights reserved.

1. Introduction

This work is concerned with production planning in manufacturing, which can be defined as the problem of finding a release schedule of jobs into the manufacturing system so that the realized outputs over time satisfy predetermined requirements as close as possible [1]. The planning horizon of production activities usually ranges from several months to two years, and the frequency of planning/replanning is weekly or monthly [2].

Typically, the planning horizon is divided into a number of discrete periods, and the decision variables represent the quantities of work of different types released into the system in each period. The performance metrics to be optimized usually include (i) the total cost (or profit), which may consist of the holding cost for finished goods (FG) and work in process (WIP) inventories, production costs and backordering costs and (ii) the fill rate, defined as the percentage of immediately satisfied demand.

Optimizing the performance metrics with respect to (w.r.t.) the release plan is challenging because it is notoriously difficult to quantify the relationships between the performance metrics and the input decisions. A manufacturing system is subject to inherent uncertainties such as probabilistic processing times, machine failures, etc., leading to complicated input–output relationships as discussed in Section 3. We shall focus on three principal time series: $A(t)$, the number of jobs released for processing during period t, $Q(t)$, the number of jobs (i.e., WIP) in the system at the start of period t, and $D(t)$ the number of completed jobs departing from the system during period t. Fig. 1 illustrates the input–output process of a manufacturing system. The release process $A(t)$, which will usually vary over time, is determined by the decision variables (the release plan). Given $A(t)$, $Q(t)$ and $D(t)$ are non-stationary time series describing the system's outputs over time, whose evolution also depends on the initial status of the system. The realized performance metrics depend on $Q(t)$, $D(t)$, and the customer demand $\mathfrak{D}(t)$; the WIP holding cost incurred is determined by $Q(t)$; and the FG holding cost and the fill rate by the departure process $D(t)$ and the customer demand $\mathfrak{D}(t)$. In practice, demand is generally a non-stationary time series, and is specified through forecasting efforts exogenous to production planning.

Despite extensive research, it remains a challenge to adequately quantify the relationship between the performance metrics and the release schedule due to the complex interactions between the non-stationary time series $A(t)$, $Q(t)$, $D(t)$, and $\mathfrak{D}(t)$. To address this difficulty, this paper develops a metamodel-based Monte Carlo

* Corresponding author. Tel.: +1 3042939477.
E-mail address: Feng.Yang@mail.wvu.edu (F. Yang).

http://dx.doi.org/10.1016/j.jmsy.2015.11.004
0278-6125/© 2015 The Society of Manufacturing Engineers. Published by Elsevier Ltd. All rights reserved.

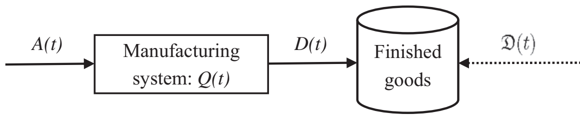


Fig. 1. The input–output process of a manufacturing system.

simulation (MCS) method with the following features. First, for a given release plan, it enables the evaluation of the probabilistic measures of system performance including expectations (e.g., the mean cost), variances (e.g., the variance of the cost) and probabilities of interest (e.g., the fill rate). Second, it is able to accommodate a wide variety of demand patterns. Third, it allows rapid evaluation of a candidate plan in terms of its performance metrics, and permits timely plan optimization.

The remainder of this paper is organized as follows. Section 2 provides a review of the related literature. Section 3 gives an overview of the metamodel-based MCS method for responsive production planning. Section 4 details the input–output metamodeling of a manufacturing system, and the metamodel-based MCS is discussed in Section 5. Section 6 formulates the multi-objective optimization problem for production planning, and presents the optimization scheme that uses the MCS method to quickly evaluate each candidate plan. In Section 7, the plan optimization approach is applied on a scaled-down semiconductor fabrication system. A brief summary is given in Section 8.

2. Literature review

Production planning problems have been addressed by several different streams of research, each emphasizing certain aspects of the problem, and thus different mathematical models. A strong case can be made [3] that none of these formulations addresses the problem faced in industry in its full complexity and generality. Hence in practice the production planning function will often combine several different mathematical models. In addition, it is possible to formulate the production planning problem in different ways depending on the time frame covered and the level of information aggregation involved [4,5]. In this paper we shall focus on a limited formulation, that of how to release work into a production system over time in order to match its output with demand in some optimal or near-optimal manner.

A central issue in matching production output to demand arises from the presence of substantial cycle times in most production systems. The cycle time is defined as the time elapsing between a unit of work (a job) being released into a production system and its emergence from the system as a finished product. In a practical production system the cycle time of any given job is a random variable whose distribution potentially depends on all uncertainties arising in the production environment, such as behavior of human decision makers, process times, machine failures, and so on. Queueing models [6,7], simulation models (e.g., [8,9]) and empirical observation are all in agreement that the distribution of the cycle time will also depend on the average utilization of the production resources. This creates a central difficulty for production planning systems: in order to match output to demand, they need to consider cycle times; but the distribution of the cycle time is determined to a considerable degree by the resource utilization, which, in turn, is determined by the release decisions made by the planning system. Hence cycle times are endogenous to the release decisions made by the planning system. However, cycle times are determined in practice by complex interactions between several complex stochastic processes evolving over time, such as the pattern of releases into the system, customer demand, machine failures, and the arrival of jobs at machines within the system over time. The difficulty of a comprehensive analytical treatment of these interactions in their entirety constitutes the central difficulty faced by production planning.

Mathematical programming models have focused on the allocation of limited resource capacity among different products over time. Much of this work assumes all inputs are deterministic, leading to formulations as linear or mixed integer programs [10–14]. Most of these models, as well as the widely used Material Requirements Planning (MRP) approach [15–17] treat cycle times as an exogenous parameter independent of resource utilization. These models ignore the stochastic nature of the problem, requiring enhancements to their solutions to be useful in practice, and also ignore the relation between release decisions and cycle times. However, deterministic mathematical programming models have been used extensively in industry as the basis for successful planning systems [18,19]. These deterministic models have been extended in several ways to incorporate uncertainty in both production and demand. Several authors [20–23] have proposed scenario-based stochastic programming models. The main difficulty with this approach is the extremely rapid growth in the size of the models as the number of decision epochs and random variables increases. A variety of robust optimization approaches have been proposed, in which one seeks a production plan that will provide a satisfactory solution over a restricted set of uncertain outcomes [24,25]. Yet another approach has been the use of chance constraints, where constraints may be violated with a specified probability [26,27]. These techniques tend to be computationally less demanding than stochastic programming, but also represent uncertainty and its consequences in different ways. Aouam and Uzsoy [28,29] compare a number of these models in the context of a very simple single-stage production-inventory system under stochastic demand, and find that they need to be parameterized with care to yield desirable results.

Another extension of mathematical programming models has been in the direction of explicitly representing the dependence between cycle times and planning decisions, making cycle times endogenous to the planning models [30,31]. One such approach, which is closely related to this work, is the use of nonlinear clearing functions that represent the expected output of a production resource during a planning period as a function of some measure of its planned workload during that period. The planned workload, is usually computed from a set of state variables such as the total amount of work available to the resource in the period or the average work in process (WIP) level during the period. The clearing function is usually assumed to be, and can in many cases be shown to be, a concave nondecreasing function of the state variables which admits of piecewise linearization, yielding tractable optimization models. Simple clearing functions can be derived from classical steady state queueing models (e.g., [32,33]) or transient queueing models [14,34,35]. In general, the form of the clearing function will change from period to period based on the values of the state variables used to estimate it. However most models using clearing functions have assumed a time-stationary function representing the expected performance of the system over an appropriate range of operating conditions. Models using clearing functions have shown considerable promise in extensive computational experiments [33,36–38]. The work in this paper can be viewed as a generalization of the clearing functions employed to date, extending the number and nature of the state variables considered and explicitly modeling the evolution of the clearing function over time based on the evolution of the underlying state variables.

While these approaches may still be sufficient in many cases to provide important insights, detailed discrete-event simulation (DES) models appear to be the only methodology that permits the detailed modeling of complex stochastic systems and their interactions that typify production systems. However, DES comes with

Download English Version:

https://daneshyari.com/en/article/1697383

Download Persian Version:

https://daneshyari.com/article/1697383

Daneshyari.com