Available online at www.sciencedirect.com

SCIENCE DIRECT®

European Journal of Operational Research 173 (2006) 827–838

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/ejor

Optimization and analysis aid via data-mining for simulated production systems

A.L. Huyet

Equipe de recherche en systèmes de production de l'IFMA, LIMOS UMR CNRS 6158, IFMA—Campus des Cézeaux—BP 265, F-63175 Aubière—France

Received 8 November 2004; accepted 21 July 2005
Available online 13 December 2005

Abstract

The optimization of a production system consists of determining a value for certain parameters which influence system performance. However, the majority of optimization methods deliver a solution without any form of explanation and this is no longer sufficient in a production context. Decision-makers would also like to have an analysis of their systems and especially of the high-performance behavior of those systems. In order to avoid the "black box" effect of many optimization methods, and to produce knowledge on system behavior (characterization of solutions that perform well, determination of critical parameters) and analyze efficient solutions, the author proposes a methodology which is based on the synergy between evolutionist optimization and an induction graph learning method. This approach is illustrated via the study of a simulated job-shop composed of five workstations, one entry station and one exit station; the numbers of machines at each station, the management method and the number of stock places in each station have to be optimized and analyzed.

Keywords: Decision analysis; Data-mining; Optimization; Simulation; Manufacturing

1. Introduction

Given the cost of new equipment investments, decision-makers want to acquire more knowledge about their production systems. In this context, knowledge that could help them to exploit or design their systems in an optimum way is in great demand. They need to know the different interactions with those systems in order to control them. These interactions or relations between different parameters affecting the system could influence the performance. For example, certain batch sizes can be efficient with one priority rule and disastrous with another. The performance measurements of those systems could be, for example,

---

E-mail address: huyet@ifma.fr

doi:10.1016/j.ejor.2005.07.026

makespan, magnitude of work-in-progress, customer satisfaction or financial impact. The studied parameters of the system and their associated performance constitute a solution.

In order to study the behavior of a system, we usually construct a model and then test it under different conditions. Production systems being often very complex, with interactive parameters, it is difficult to use an analytical approach to model them without formulating too restrictive a hypothesis. Thus, simulation is often used to study the different possible solutions. Nevertheless, simulation only enables the testing of modeled system behavior under certain conditions. It does not determine the critical parameters of the system regarding performance or the interactions between parameters which induce good performances.

In the literature, different methods such as data-analysis methods (sensitivity analysis (Kleijnen, 1992)), perturbations analysis (Heidergott, 1995), data-mining or learning methods (Michalski et al., 1984) could be used to determine important parameters or interactions between the studied parameters. Some of these approaches are also used on simulation results (Lereno et al., 2001; Pierreval and Ralambondrainy, 1990). Nevertheless, the obtained results are very dependent on the data set on which the analysis is made or the knowledge acquired. Moreover, since, in our context, we are especially interested in high-performance solutions, we will have to acquire knowledge or to provide analysis that is able to help decision-makers in their search for profits. Therefore, we will have to construct data sets that are automatically relevant to performance.

Our problematic, then, is to develop a method which at the same time produces some elements of understanding about system behavior and focuses these elements on the relevant points of system operation, i.e. those where it will be very efficient. The approach that we propose is based on a combination of a learning method and an optimization method. This synergistic effort allows the weaknesses of learning and optimization to be compensated for. The learning method will characterize the areas of the solution space that are pertinent to our case (i.e. those areas containing high-performance solutions). Optimization is then progressively guided towards those areas. Reciprocally, optimization automatically provides a data set for the learning method (Huyet and Paris, 2004). This data set contains more and more relevant data (solutions that represent high-performance settings for the system) as the optimization process progresses.

First, we will present our methodology, and then we will illustrate it through the study of a job-shop composed of five workstations, one entry station and one exit station. The number of machines in each station, the management method and the number of stock places in each station should be optimized and analyzed with regard to the performance.

2. Proposed methodology: Conjoint learning and optimization

2.1. General principle

One of the crucial points of learning methods is the construction of the data set from which we learn. Indeed, relevant knowledge is difficult to acquire due to the large search space. So an important improvement of the relevance of extracted knowledge regarding performance will be to construct a data set which is centered on what interests industrial managers: solutions representing high-performance system behavior.

The search for high-performance solutions is the aim of many optimization methods. Nevertheless, those methods generally work like "black boxes": they give solutions without any explanation of system behavior. It has to be noted that gradient methods (Spall, 2003) give some information about the impact of each decision variable on system performance, but this information could be difficult to analyze. Moreover, since we are dealing with non-numerical decision variables (priority rules, for example), gradient methods do not seem to be relevant to our purpose. However, in the context of manufacturing systems, we need to know which decision variables are important to build a "good" solution. In fact, in the industrial area, nobody will invest even in the best solution without any explanation. With the economic con-