PERGAMON

Computers & Industrial Engineering 38 (2000) 361–374

computers & industrial engineering

www.elsevier.com/locate/dsw

Using data mining to find patterns in genetic algorithm solutions to a job shop schedule

D.A. Koonce[a,*], S.-C. Tsai[b]

[a]*Industrial and Manufacturing Systems Engineering, Ohio University, Athens, OH 45701, USA*
[b]*Technical Specialist, SAS institute Taiwan Ltd., Taipei, Taiwan, ROC*

Accepted 25 July 2000

Abstract

This paper presents a novel use of data mining algorithms for the extraction of knowledge from a large set of job shop schedules. The purposes of this work is to apply data mining methodologies to explore the patterns in data generated by a genetic algorithm performing a scheduling operation and to develop a rule set scheduler which approximates the genetic algorithm's scheduler. Genetic algorithms are stochastic search algorithms based on the mechanics of genetics and natural selection. Because of genetic inheritance, the characteristics of the survivors after several generations should be similar. In using a genetic algorithm for job shop scheduling, the solution is an operational sequence for resource allocation. Among these optimal or near optimal solutions, similar relationships may exist between the characteristics of operations and sequential order. An attribute-oriented induction methodology was used to explore the relationship between an operations' sequence and its attributes and a set of rules has been developed. These rules can duplicate the genetic algorithm's performance on an identical problem and provide solutions that are generally superior to a simple dispatching rule for similar problems.

Keywords: Data mining; Job shop scheduling; Genetic algorithms

1. Introduction

In recent years, information growth has proceeded at an explosive rate. While database management systems (DBMS) provide us with basic tools for the efficient storage and look-up of large data sets, the capabilities for collecting and storing data have far outpaced our abilities to analyze, summarize and extract "knowledge" from this data. Traditional methods of data analysis were based mainly on humans

* Corresponding author. Tel.: +1-740-593-1544; fax: +1-740-593-0778.
E-mail address: koonce@ohio.edu (D.A. Koonce).

0360-8352/00/$ - see front matter
PII: S0360-8352(00)00050-4

dealing directly with data. Large volumes of data overwhelm the traditional manual methods of data analysis and make the task of analysis more difficult and less efficient. Also, traditional methods of data analysis, such as spreadsheets and ad hoc queries, can create informative reports from data, but cannot analyze the contents of those reports by focusing on important knowledge. These methods help only in data collecting and computing. They do not assist in improving the analysis task. Moreover, they overemphasize the statistical aspects of data while ignoring the domain knowledge of data. As a result, traditional analysis can fail to reveal the physical natures that the data implies.

Genetic algorithms (GAs) often provide fast solutions to traditional numeric problems. For example, a GA can generate schedules for a manufacturing job shop. However, GAs do not demonstrate repeatability or provide an explanation of how a solution is developed. Using data mining, this paper presents a method for inducing rules from the solutions of a GA, which describe its behavior. These rules have also been applied to similar job shop cases with success.

2. Background

To describe the process of mining solutions from a GA for rules, three background areas must be discussed: job shop scheduling, genetic algorithms and data mining.

2.1. Job shop scheduling

Job shop scheduling is an important and complex activity in manufacturing. A job shop model, typical in manufacturing, can be described as a set of jobs composed of sequences of operations that are processed on a set of machines. It is a decision making process which allocates limited resources over time to perform a set of jobs to meet objectives (Baker, 1974). The complexity of analysis for even small job shop problems has lead to significant research in heuristic rules. For example, a six job, one machine system has 6! or 720 possible schedules. Add two operations per job and a second machine and the number of schedules to search will be $6!^{*}6! = 518,400$; of which, half will not violate precedence constraints. The decision making process can be broken into two functions: resource allocation and operations sequencing. Many methodologies have been developed to solve the problem, such as dispatching rules, search algorithms approach, and artificial intelligence.

The dispatching rule approach uses information related to a specific job or operation to arrange the next job to be processed on a machine. Because they provide a straightforward method to solve a problem using characteristics of the jobs, operations and machines, the computation is simple and quick. But, each dispatching rule approach is limited to one specific objective or data element. Panwalker and Iskander (1977) classified dispatching rules into three categories-priority rules, heuristic scheduling rules, and others. Priority rules are based on information related to jobs, operations, or machine to determine the priority of resource allocation. Heuristic rules involve more complex considerations, such as machine load, alternative routing, alternative operation, etc. Blackstone, Phillips, and Hogg (1982) mentioned that Shortest Processing Time (SPT) was the best rule when the shop does not set due dates.

Search algorithms, such as simulated annealing and genetic algorithms, are another approach that provides the strategies to explore the solution space efficiently. However, sometimes their computation