Expert Systems with Applications 38 (2011) 7609–7615

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Solving flexible flow-shop problem with a hybrid genetic algorithm and data mining: A fuzzy approach

H. Khademi Zare, M.B. Fakhrzad *

Department of Industrial Engineering, Yazd University, Yazd, Iran

ARTICLE INFO

Keywords:
Flexible flow-shop scheduling
Genetic algorithm
Data mining
Attribute-driven deduction
Fuzzy sets

ABSTRACT

In this paper, an efficient algorithm is presented to solve flexible flow-shop problems using fuzzy approach. The goal is to minimize the total job tardiness. We assume parallel machines with different operation times. In this algorithm, parameters like "due date" and "operation time" follow a triangular fuzzy number. We used data mining technique as a facilitator to help in finding a better solution in such combined optimization problems. Therefore, using a combination of genetic algorithm and an attribute-deductive tool such as data mining, a near optimal solution can be achieved. According to the structure of the presented algorithm, all of the feasible solutions for the flexible flow-shop problem are considered as a database. Via data mining and attribute-driven deduction algorithm, hidden relationships among reserved solutions in the database are extracted. Then, genetic algorithm can use them to seek an optimum solution. Since there are inherited properties in the solutions provided by genetic algorithm, future generation should have the same behavioral models more than preliminary ones. Data mining can significantly improve the performance of the genetic algorithm through analysis of near-optimal scheduling programs and exploration of hidden relationships among pre-reached solutions.

1. Introduction

Flow-shop with parallel machines (FSPM) problem is well-known to flexible flow-shop problem (Hoogeveen, Lenestra, & Veltman, 1996; Janiak & Lichtenstein, 2001; Tian, Jian, & Zhang, 1999). This problem includes a number of products with the same production steps. There are parallel machines in each production stage (work-station). All products have should pass step 1 to step "n" (Artiba & Elmaghrabhy, 1997; Shiau, Cheng, & Huang, 2008).

Each job is processed just by one machine in each step. Also, each machine is working on one job each time and equally each job is under processing just using one machine each time. Machine center can be filled by uniform or unrelated machines.

In this paper, we suppose that three types of machines in each center. These machines are equal in operation type and different in operation speed.

In recent two decades, many researchers focus on flexible flow-shop problem which is a NP-Hard problem. Most researchers considered maximum finish time as a goal function. As well as mathematical modeling, deterministic, heuristic and even meta-heuristic methods are used (Chen, 1995; Gupta, 1988; Haouari & Hallah, 1997). Others preferably applied deterministic and heuristic methods to solve small-size problems (Botta-Genoulaz, 2000; Braha & Loo, 1999; Kyparisis & Koulamas, 2001). The main parameter used in these papers is to minimize tardiness time for all jobs. Customer's orders represent input jobs to the system. For each job, the setup and process times are determined by process unit and due date is established by customers. Botta-Genoulaz (2000) presented six heuristic algorithms to minimize the maximum tardiness in a flexible flow-shop problem with different due dates. Lin and Liao (2003) developed a near-optimal heuristic algorithm to minimize maximum tardiness in a 2-stage flexible flow-shop problem. Bertel and Billaut (2004) proposed a heuristic algorithm for a scheduling problem in a flow-shop environment with regards to minimizing the numbers of operations with balanced tardiness. They have used a combination of integer-linear programming and genetic algorithm for the mentioned problem. Vob and Witt (2007) studied on a flexible flow-shop which is an actual scheduling problem including 16 processing steps to minimize balanced tardiness time. They have developed a mathematic model on the basis of scheduling problem with limitation of resources. Then they have solved the problem heuristically. Besides pointed out algorithm used to solve different types of problems, a number of other algorithms can be found. For example, there are lots of algorithms most of which use neighborhood search, tabu search and simulated annealing. Some of them also use branch and bound algorithm (Chen, Pan, & Lin, 2008; Garey & Johnson, 1979; Karimi & Zandieh, 2009).

The above techniques have significantly success to solve combinational optimization problem. The flexibility of these techniques

* Corresponding author.
E-mail address: mfakhrzad@yazduni.ac.ir (M.B. Fakhrzad).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.12.101

especially genetic algorithm has extended their application scope (Deborah & Freitas, 2004; Roshanaei, Seyyed Esfehani, & Zandieh, in press). In this paper, we have used a different analytical glass to the flexible series scheduling problem to improve the efficiency of the solution. We have also used data mining to study the properties of the solutions to find their hidden relationships. Data mining is an effective tool to manage huge databases. In the presented methodology, each feasible scheduling solution is considered as a database member. After clustering the data, their properties are recorded. Then, logical relationships among the operation's properties and their sequence have been developed through data mining deduction algorithm. The extracted rules help genetic algorithm to boost its speed to the optimal point. In each stage of scheduling, there are three types of machines with low, medium and high operation speed. A triangular fuzzy number is assigned for the time of operations. The structure of the remaining part of the paper is as follows: In Section 2, we have studied the flexible series scheduling problem. In Section 3, the application of data mining in the flexible series scheduling is studied. The genetic algorithm is applied to the problem in Section 4. The proposed methodology is discussed in Section 5. In Section 6, we analyzed the efficiency of the methodology and finally, we illustrate the findings in Section 7.

2. Flexible series scheduling problem

A fuzzy flexible flow-shop problem is considered as a difficult problem in production management and combinational optimization. In this problem, a set of jobs having same sequence are existed. There are n jobs which totally should pass m production steps. Each job in the whole process uses a unique machine. Each step includes a number of machines with different operation speed: low, medium or high (Behnamian, Fatemi Ghomi, & Zandieh, 2009). The time of operations for each job would be different on the basis of the assigned machine. The operation time follows a triangular fuzzy number. The objective of such problem is to sequence jobs on the existing machines so that the sum of tardiness time would be minimized. In general, the start time and operation time for each job can be different. In this paper, we suppose the starting times for all the jobs are equal.

Since the flexible flow-shop problem is an NP-Hard problem, the number of flexible solutions for sequencing jobs increase on the basis of exponential function.

For a flexible flow-shop machine including 6 jobs and 1 machine, the total number of feasible and infeasible solutions to sequence jobs are equal to 720 (i.e. 6!). Suppose the number of machines is 6, then, the answer is equal to 518400 sequencing routes. To encounter to this problem, a simple search mechanism can hardly find the optimal solution. Different search algorithms are developed to solve such problems but, they still cannot guarantee to reach the optimal solution.

It should be noted that since we can consider different idle times between two adjacent machines hence there are unlimited quantity of feasible scheduling programs regarding fuzzy approach. It is so clear that after sequencing the machines, the mentioned idle time between two adjacent machines isn't so effective. As a result, there is no need to search within all feasible solutions. Thus, a set of feasible times is defined and the solution space can be decreased. Now, the optimal solution will be in this space reasonably.

3. Flexible flow-shop problem and data mining

Today, human has to use new tools to capture the knowledge via analyze gathering information intelligently. Knowledge perception and mining data bases are some kinds of Artificial Intelligence tools which help us to analyze huge volume of information. At the other hand, knowledge perception includes process of evaluation and interpretation of models to capture what we call is knowledge. As well, data mining is a typical application of deduction algorithms which is used to identify behavioral models in data or trends in a specified context. Most of data mining algorithms are derived from machine-learning, trend recognition and statistical techniques. These algorithms include clustering, grouping and graphical models. Primary objective are described in these algorithms. Before mining data, the appropriate algorithm should be selected and data partially organized. In this paper, property-driven deduction algorithm is selected as a data mining tool. The first step in studying flexible flow-shop problem using property-driven deduction algorithm with fuzzy approach is to prepare data.

Firstly, the effective parameters for sequencing the jobs should be recognized. Then, concepts and levels of the hierarchical structure are presented. Also, defined concepts are applied on higher levels. The whole indicated procedure above must be done by an expert whose knowledge and intuition is sufficient to the problem space (Chen, Hong, & Vincent, 2009; Lin, Shi-Jinn Horng, Yuan-Hsin Chen, Rong-Jian Chen, & I-Hong Kuo, 2009). In this paper, we identified five effective parameters which can affect job sequencing. They are:

i. **Machine grade:** jth job would be done by machine with grade m in kth step. For each operation, there are three types of machines including low, medium and high speed. Note that lower grade machine equals to higher speed one.
ii. **Operation time:** The operation time needed for the jth job in step k is determined by machine of grade m.
iii. **Remaining time:** It is equal to the sum of processing time for the remaining j jobs after passing k steps
iv. **Load on machine:** Sum of the total processing time should be considered on machine without any specific sequencing.
v. **Delivery time:** Time needed for delivering jth job to the customer.

4. Genetic algorithm and FFS problem

Genetic algorithm is a statistical method for the optimization and search problems. The algorithm is a member of evolutionary calculations in the field of Artificial Intelligence. Its unique property makes the algorithm so effective that we cannot assume it as a simple stochastic search method. In fact, the preliminary idea of genetic algorithm comes from Darvin's evolution theory (Casillas, Francisco, & Martínez-López, 2009). The algorithm is based upon natural behavior of human genes. To solve optimization problems, parameters like search space, coding type and method, shapes of chromosomes and their lengths, population size in each generation, fitness function and the most important ones including mutation and cross-over types must be considered. The search space refers to the space we are looking for optimal solution in it.

The shape and type of chromosomes in the genetic algorithm should be in such a way that the algorithm can solve the problem with predefined accuracy. It also must search the space well. The population size in each generation relates to the problem type. The fitness function shows the ideality level of the genetic algorithm. The chance of each parent chromosome to be selected to create child chromosomes is due to its fitness function. Most of the times, the best chromosome is transferred to the next generation. The mutation and cross-over operators have significant role on performance of genetic algorithm. The "pmx" and "6x" are two famous cross-over operators (Garey & Johnson, 1979). Besides,