

Data- and Rule-Based Integrated Mechanism for Job Shop Scheduling

Yanhong Wang*, Dandan Ji

Department of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China.

* Corresponding author. Email: wangyh@sut.edu.cn

Manuscript submitted July 10, 2014; accepted January 17, 2015.

doi: 10.17706/ijcce.2015.4.3.180-186

Abstract: In this paper, a new data-based mechanism has been proposed to solve job shop scheduling problems (JSPs). The proposed method is an intelligent-algorithm-based approach implanted with data- and rule-based scheduling algorithm. The proposed data-based framework is based on decision tree algorithm and rule mining technique to solve the scheduling in job shop with abundant information and massive unforeseen events. A data- and rule- based scheduling system framework for JSP was developed, and then, the knowledge discovery and rule-based adaptive scheduling algorithm were developed. The simulation result of a simplified scheduling problem is presented as a preliminary validation to show the feasibility of the proposed integrated scheduling mechanism. The state-of-the art research on the key technologies of data-based scheduling is introduced together with some related research results of information based or rule based ones. The results show that data- and rule-based scheduling offers better solutions.

Key words: Data-based, integrated scheduling mechanism, job shop, rule-based.

1. Introduction

Scheduling is a decision making process which deals with the allocation of limited resources to tasks over time, which plays an important role in achieving the goal of system optimization. The job shop scheduling problems (JSPs) is one of the most particular hardest combinatorial optimization problems that have been proven to be NP-hard.

JSPs presents a simplified model of many real production scheduling problems [1] in manufacturing practice, therefore it has attracted the attention of many researchers due to its wide applicability and inherent difficulty. However, in today's business environment, products need to be manufactured with huge product variety and large volume of small batches to achieving productivity and profitability, which make the actual manufacturing processes becoming increasingly complex. The uncertain and unforeseeable dynamic manufacturing environment presents new challenges for the existing JSPs research results.

In such a complex manufacturing environments described above, vast amounts of data are collected in information system or database management systems from all involved areas in manufacturing management, using barcodes, RFID, sensors, vision systems etc. The collected manufacturing data contains valuable information that could be integrated within the manufacturing system to improve decision making and enhance productivity (Elovici and Braha 2003). Recently, data-based scheduling has emerged as a new way to solving JSPs in such a manufacturing environment supporting by on- and off-line data accumulated

and stored in data base. It has drawn great attentions from both academia and industry to the application of data-based methods to the scheduling problems in complex manufacturing systems [2].

Data-based scheduling method has a distinct advantage in solving complex JSPs, however how to extract useful information and knowledge for scheduling from a large number of data is still an intractable problems. In this paper, we propose our preliminary research results in data-based JSP with a data-based and rule-based integrated scheduling mechanism is discussed. The remainder of this paper is organized as follows. Section 2 reviews some recent works related to the data-based scheduling. Then a data-based scheduling framework is developed and explained in Section 3. Details of our novel knowledge discovery approach as well as the rule-based scheduling algorithm are given in Section 4 and Section 5. A simple example is discussed in Section 6 to illustrate the efficient of the proposed method and this is followed by conclusions in Section 7.

2. Literature Review

JSPs is a complex combinatorial optimization problem full of numerous constraints as well as uncertain or unpredictable changed condition factors especially in today's challenge manufacturing environment. Data-based scheduling gives a feasible pathway to deal with the above troublesome in JSPs. Recently, several data-based scheduling models or algorithms have been proposed, which integrated with data processing algorithm, simulation technique or other technical measures [3]. Since there are necessary relation between data and knowledge as well as rule, those knowledge or rule based scheduling approach would give references or lessons for data-based scheduling. We now review some of the literature pertaining to the field of data-based scheduling as well as the knowledge or rule based scheduling.

2.1. Data-Based Scheduling

Data accumulated in the production process for scheduling can be classified into on- and off-line data, which includes information related with part, device, as well as other relevant aspects. Choudhary *et al.* [4] clearly indicated that it is necessary to adapt effective data processing methods, due to the concealment character of data, and data mining approach might become an important tool to obtain knowledge from data.

To solve the data-based scheduling, Li *et al.* [2] proposed a data-based scheduling framework composed of data layer, model layer, and scheduling layer. Meanwhile, it indicated that integrating data-based scheduling with conventional scheduling would bridge the gap between research and application. For dynamic manufacturing system scheduling, Shahzad *et al.* [5] proposed a data mining based approach framework integrated the data mining optimization and simulation, for the dynamic job shop scheduling problems. The framework applied a tabu-search algorithm to find optimal scheduling solutions, and then these solutions are approximated with a decision-tree based learning algorithm.

Compared with traditional scheduling methods, data-based scheduling has the characteristics of complexity. Consequently, the combination of data-based scheduling with existing scheduling methods and algorithm would be a realistic way.

2.2. Knowledge or Rule Based Scheduling

Knowledge discovery is a process to obtain potential, novel, and useful knowledge from data, by using some or certain kinds of data processing techniques. The process of knowledge discovery can be regarded as a process of data processing that might be a specific step for data-based scheduling.

Panwalkar and Iskander [6] presented a summary of over 100 priority dispatching rules, and classified them into simple priority rules, heuristic scheduling rules and other rules. Each rule had been description in detail, with its features. Koonce and Tsai [7] used data mining methods to discover useful patterns from

data for JSPs. The training data were generated by the genetic algorithm firstly, and then the attribute-oriented induction method was used to obtain useful patterns of knowledge.

The motivation of research on data-based scheduling is the limits of traditional scheduling facing to complex manufacturing systems [2]. It is naturally a new idea to solve complex scheduling problems, however, how to extract useful knowledge from related on- and off-line data to improve the operational performance of the JSPs is still a novel and unsolved issues. The aim of this paper is therefore to present a trying to use knowledge discovery and rule-based decision tree for data processing and usage in JSPs.

3. System Framework

A new generation of methods and tools are required to assist humans in intelligently managing modern manufacturing system. Taking into account features on shop floor related to scheduling, a data-based scheduling framework for job shop is proposed as shown in Fig. 1.

In the data-based jobshop scheduling, data processing is one of the basic and critical sections. This framework gives a platform for extracting knowledge or rules by processing on- and off-line data, which would be used for the decision making in scheduling. In this framework, data processing model as well as the followed rule-based algorithm would be a selective use. If there are no obvious changes emerging in manufacturing environment, the JSPs can be optimally solved with existing scheduling methods, and the “data-based scheduling” will not be triggered. On the other hand, on-line data as well as related off-line data would be “processing” to digging out rules and to make rational decisions under changed environment.

In the proposed framework, there are several key points need to be interpreted to show the train of our thought. The following offered several explanations.

3.1. Data Sources

Data are the basis and premise of the realization of data-based scheduling. These data may include both from different information systems in the enterprises, such as ERP, MES, ACP, SCADA, and from production scheduling process. The large amounts of data can be classified into two categories: off-line data and on-line data. The former includes related information of product, order, equipment, and other key element in shop. The latter is reflection of the real-time status of production scenes and their environment, including state information of equipment, WIP status, and so forth.

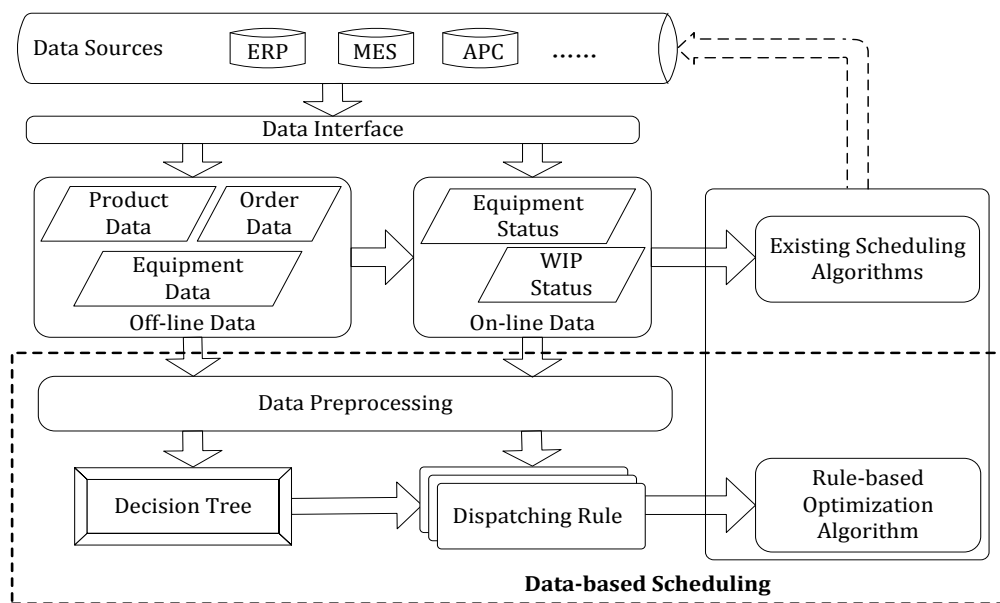


Fig. 1. Data-based scheduling framework.

3.2. Data Preprocessing

The obtained data from manufacturing processes might be generally incomplete, inconsistent, and with noisy, due to the lag of updating or inaccurate transmission. Therefore, these data should be pre-processed, for the purpose of data-purification and noise reduction. Since these data would be used as training data to discover scheduling rules, aggregation and attribute selection need to be further carried out to ensure the completeness and reliability of data pre-processing.

- 1) Data cleaning is designed to eliminate data noise and handle missing data, where Data Smoothing Technique is used to get solutions.
- 2) Data aggregation is about to merge two or more objects into a single object. Techniques adopted might be support vector machines or fuzzy K-means methods.
- 3) Data feature subset selection is characteristic selection, which eliminating some irrelevant and redundant attributes. Commonly, technologies been used are computational intelligence methods.
- 4) Discretion will transfer continuous attributes into discrete attributes to facilitate certain algorithms using discrete value. Decision tree, fuzzy aggregation and other methods are might be used.

3.3. Decision Tree

After the data have been pre-processed, data mining algorithms are used to obtain effective knowledge rules. Decision tree is one of the common used data mining algorithms [4], which has relatively small calculation and high intelligibility. It could derive rules from the training or sample data for scheduling. The branch nodes of the tree are made up of attribute values of training data, while the root node is made up of the most informative one among the training data. Further scheduling rules as well as knowledge will be acquired via constructing trees and pruning trees. The proposed framework also provided a flexible switching mechanism for fitting the actual shop floor environment. It usually runs in the regular mode, which making scheduling decision directly by these derived rules. However, if there are any non-routine incidents happened on the floor, the data-based scheduling modeler will be awoken to making a occasional decision using the rule-base intelligent scheduling algorithm as well as online data. Besides, the results of scheduling will be record as resources for acquiring new rules for future use under the similar occasion.

4. Process of Knowledge Discovery

Knowledge discovery is a process of discovering interesting and useful knowledge from the objective data set [8]. For data-based scheduling, knowledge discovery can help for making helpful knowledge to making further suitable scheduling decision. Main steps of the process are specialized now.

Step 1. Data selection. Determine the scope of objective data to be processed, and select corresponding data set related to the objective task.

Step 2. Data pre-processing. This step include several operations such as cleaning data noise, eliminating duplicate data, deriving and repairing missing data, completing the data type conversion and *et al.* The object is to ensure the completeness, reliability and operability of the data using processes.

Step 3. Data transformation. Data transformation will convert the data into a unified form which might suitable for mining. Operations include data summarizing or aggregating, standardization.

Step 4. Data mining. Suitable data mining algorithms will be selected to obtain useful patterns and make further efforts to achieve effective knowledge acquisition.

Step 5. Interpretation and knowledge presentation.

This process will run iteratively until a rational and meaningful pattern has been gained. If a meaningful pattern has been identified, visualization and knowledge representation technology will be adopted to present the obtained knowledge for use.

5. Rule-Based Adaptive Scheduling Algorithm

To solve the problem under the proposed data-based scheduling framework, an intelligent scheduling algorithm should be roused with the help of related rule from rule base.

In this paper, A rule-based ant colony algorithm is elaborated to make the optimal decision where scheduling rules, such as the longest processing time (LPT), which are derived from the decision tree algorithm as the suitable rule for this occasion, are applied to the ant colony algorithm for the optimal scheduling decision making. Details of the algorithm include the following steps.

Step 1. Initialization parameters, such as pheromone, number of ants, the influence of pheromone heuristic factor and pheromone self-heuristic factor.

Step 2. Create a generation of ants, and construct paths to search the feasible solution in this iteration.

Step 3. Optimize the local or global feasible solution, according to the dispatching rule in rule base

Step 4. Adjust and update the pheromones, meanwhile, update the influence of pheromone heuristic factor and pheromone self-heuristic factor dynamically based on the ant colony algorithm [9].

Step 5. Judging whether the result meets the stopping criterion. If the answer is yes, export the best solution and go to Step 6, otherwise return to Step 1.

Step 6. End the program.

6. Simulation and Results

A preliminary simulation research is given to verify the effective of the proposed approach. To demonstrate the viability of the developed method, a miniature JSPs with two-job and three-machine has been considered. The process planning characteristics of the instance are described in the Table 1. There are two jobs in the example, and each has three different operations to be processed according to a given sequence.

Table 1. Parameters of a Two-Job and Three-Machine Problem

Jobs	Operation sequence	Machine alterative	Processing time (s)
J_1	O_{11}	$M1$	10
	O_{12}	$M3$	18
	O_{13}	$M2$	10
J_2	O_{21}	$M1$	20
	O_{22}	$M2$	20
	O_{23}	$M3$	10

Fig. 2 gives the result which adopts the conventional ant colony optimization algorithm [9], and the rule-based adaptive ant colony algorithm is adopted in Fig. 3. The minimum make-span of the conventional ant colony algorithm is 60, and the proposed data- and rule-based algorithm is 58, which gives a shorter minimum make-span. It suggests that the dispatching rule has a great impact on the total processing time.

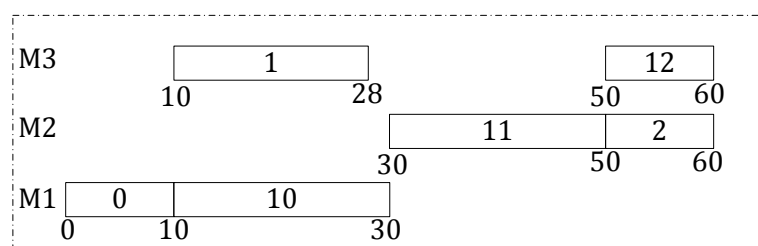


Fig. 2. Gantt chart of a conventional algorithm.

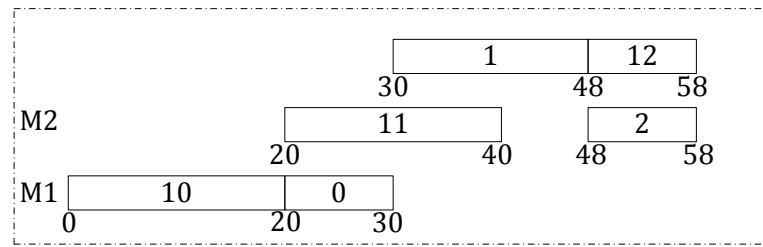


Fig. 3. Gantt chart of the proposed data-based algorithm.

7. Conclusion

In this paper, many attempts have been made to construct effective novel data- and rule-based scheduling algorithm with special decision tree implanted rule mining for solving flexible job shop scheduling problems. The data-based scheduling framework as well as accompanying algorithms is proposed, after a brief overview of scheduling strategy related to the data-based scheduling research is provided. The performance of the proposed algorithm has been compared with conventional algorithms, such as ant cloning scheduling. From the results of simulations, we can see, our approach, novel scheduling algorithm with data-based decision tree and rule mining uses ant cloning intelligent, take into account the current state of the scheduling environment, can make better scheduling decisions. However, the research and simulation is only preliminary work. Future work includes applying the proposed method to other scheduling problems with different constraints, especially for the scenario with machine disruption, job released delay, processing time prolonging or other similar environment.

Acknowledgment

This work is supported by Natural Science Foundation of Liaoning Province, China, under Grant 2013020022.

References

- [1] Ziaee, M. (2014). A heuristic algorithm for solving flexible job shop scheduling problem. *International Journal of Advanced Manufacturing Technology*, 71(1-4), 519-528.
- [2] Li, L., Sun, Z. J., Ni, J. C., & Qiao, F. (2013). Data-based scheduling framework and adaptive dispatching rule of complex manufacturing systems. *International Journal of Advanced Manufacturing Technology*, 66, 1891-1905.
- [3] Liu, M. (2009). A survey of data-based production scheduling methods. *Acta Automatica Sinica*, 35(6), 785-806.
- [4] Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2009). Data mining in manufacturing: a review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5), 501-521.
- [5] Shahzad, A., Mebarki, N. & Irccyn, I. (2010). Discovering dispatching rules for job shop scheduling problem through data mining. *Proceedings of the 8th International Conference of Modeling and Simulation*.
- [6] Panwalkar, S. S., & Iskander, W. (1977). A survey of scheduling rules. *Operations Research*, 25(1), 45-61.
- [7] Koonce, D. A., & Tsai, S. C. (2000). Using data mining to find patterns in genetic algorithm solutions to a job shop schedule. *Computers & Industrial Engineering*, 38(3), 361-374.
- [8] Kurgan, L. A., & Musilek, P. (2006). A survey of knowledge discovery and data mining process models. *Knowledge Engineering Review*, 21(1), 1-24.
- [9] Wang, W. X., Wang, Y. H., Yu H. X., & Zhang, C. Y. (2013). Dynamic-balance-adaptive ant colony

optimization algorithm for job-shop scheduling. *Proceedings of the Fifth Conference on Measuring Technology and Mechatronics Automation*.



Yanhong Wang received the PhD degree in control theory and control engineering in 2003 from Jilin University, Changchun, China. She is currently a professor in the Department of Information Science and Engineering, Shenyang University of Technology, Shenyang, China. Her current research interests include production scheduling, distributed information processing, and enterprise information system.



Dandan Ji is a graduate student in the Department of Information Science and Engineering, Shenyang University of Technology, Shenyang, China. Her research interests include data-based scheduling, data mining, and distributed information processing. She intends to expand research issues to big data and data-based information systems in the future.