The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Acquisition of Rules for Selecting Suppliers of Raw Materials in Distributed Production Systems by means of Reinforcement Learning

Hisaaki Yamaba[1], Kayoko Takatsuka[1], and Shigeyuki Tomita[1]

[1] University of Miyazaki, Miyazaki, Miyazaki 889-2192, Japan
(Tel: 81-985-58-7425, Fax: 81-985-58-7425)
(yamaba@cs.miyazaki-u.ac.jp)

Abstract: In these days, many production systems are consist of several factories. Such factories are dispersed in wide area and form "production networks". In such networks, each factory produces intermediate materials for other factories. In order to operate production networks efficiently, some rational and sound operational strategy is needed for realizing cooperative operation. In the previous work, "Behavior Model" of scheduling activities in decentralized production networks was developed and the validity was confirmed. Also, an attempt was made to obtain proper scheduling rules by means of reinforcement learning. Concretely, Profit Sharing was adopted in order to obtain rules for selecting suppliers of intermediate materials under the proposed operational model. In this work, improvement of the representation of states used in the rule learning was attempted. A series of experiments was carried out in order to examine the performances of the rules obtained under the new representations.

Keywords: Production System, Supply Chain, Reinforcement Learning

1 INTRODUCTION

In these days, many production systems are composed of several factories. Such factories are dispersed in wide area and form "production networks". In such networks each factory produces intermediate materials for other factories. In such networks, decision-making in each factory is currently carried out independently of other factories, and this causes low performance of the whole system. In order to operate production networks efficiently, some rational and sound operational strategy is needed for realizing cooperative operation among such networks. In order to manage such a production network soundly, negotiations among factories are very important. In the previous work, "Behavior Model" of scheduling activities in decentralized production networks was developed (Yamaba [1]). Also, an attempt was made to obtain proper scheduling rules by means of reinforcement learning (Yamaba [2]). Concretely, Profit Sharing was adopted in order to obtain rules for selecting suppliers of intermediate materials under the proposed operational model.

In this work, improvement of the representation of states used in the rule learning was attempted. A series of experiments was carried out in order to examine the performances of the rules obtained under the new representations.

2 CHARACTERISTICS OF TARGET PRODUCTION SYSTEM NETWORK

The target production system network in this work is composed of a single business department and several factories. Locations of them are widely dispersed.

Each order from customers is completed through multi-stage manufacturing processes. For example, plastic film manufacturing companies with an upper process of raw plastic film creation and several lower processes of chemical treatment of plastic films are assumed. Each stage of manufacturing of an order is allowed to be performed at different factories. Some factories have facilities only to create raw plastic film, some factories have facilities to perform chemical treatment of plastic film, and others have both of the facilities.

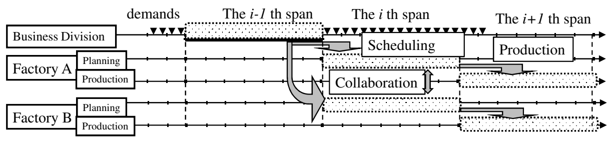

Fig. 1. An example of a timetable of the target production system

Orders from customers are concentrated to the single business department. The business department selects a factory for manufacturing processes of each order periodically (e.g., every one month). Each factory makes out production schedules of given orders together with procurement plans of intermediate materials at the same time. "Procurement plans" mean selection of factories for each intermediate material.

An example of a timetable of the production system is shown in Fig. 1. In general, schedules of orders given in the i 1 th span are made in the i th span and bases produce them in the $i+1$ th span.

In such a production network, if "Factory A" makes an order of its intermediate material needed to produce an order given in i 1 th span to "Factory B" in the i the span, "Factory A" cannot finish making out its schedule until it will be determined that "Factory B" will accept the order or not. Since "Factory B" makes out its own schedule for orders given in i 1 th span in the same span, "Factory B" cannot reply whether it can accept the order or not until the schedule is fixed. In general, "Factory B" also makes orders of its intermediate materials to other factories. This causes longer time until the whole scheduling process is finished. In particular, if "Factory B" also made an order to "Factory A", the scheduling process never finishes because of a deadlock. So it is indispensable to realize a scheduling algorithm which is equipped collaboration mechanism among factories.

3 BEHAVIOR MODEL OF DECENTRALIZED SCHEDULING

In order to realize the sound scheduling method, the concept of "behavior model" was introduced (Yamaba [1]).

Receiving orders of intermediate materials from other factories, each factory create several schedule candidates. However, it could be happened that no schedule candidate can satisfy all of the limitations of production (due date) of the given requirements in case that much workload which exceeds the productive ability of the factory is given. In this work, it is assumed that factories are allowed to reject some requirements on such occasions. But factories have to reply to the orderer that the requirements are rejected within the designated time span.

Receiving a reply of rejection ("reject messages"), production bases select another factory and make an order to the selected factory again. It is assumed that there is a limitation for the number of attempts of re-selection of factories. When the number of the re-selecton for the N 1th process exceeds the limitation, the orderer factory abandon the request of N th process given to the factory itself. Such factories reply the "reject message" to the orderer factory of the N th process.

4 LEARNING OF SUPPLIER SELECTION RULES

4.1 Supplier Selection Rules

There are several factors affecting the performance of schedules of production networks discussed here. In Yamaba [2], an acquisition method of rules for selecting proper factories for intermediate materials was proposed. Reinforcement learning was adopted in the method. Profit Sharing is one of the most promising methods of reinforcement learning. The method was used in order to obtain operational rules of engineering systems (Arai[3], Yamaba [4], Yamaba [5]).

The target production systems are plastic films factories. It is assumed that there are three processes in the production

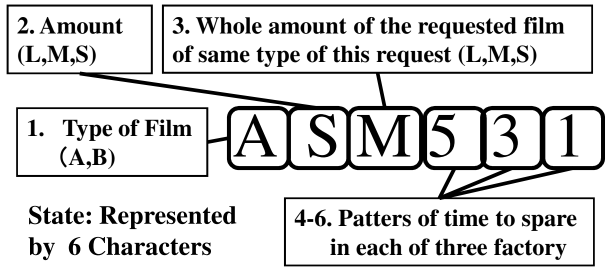

Fig. 2. Representation of states of an intermediate material requirement

system: the film creation and two stages of chemical treatment of film processes (Fig. **??**.). Also, there are 3 types of factories in the target production system.

Type 1: factories which has facilities for film creation processes only.

Type 2: factories which has facilities for chemical treatment processes only.

Type 3: factories which has all facilities.

It is assumed that there is one factory of Type 1 (factory 1), two factories of Type 2 (factory 2 and 5) and two factories of Type 3 (factory 3 and 4).

The target production systems deal with two types of films α and β. Each type of films is produced by fixed type of machines: machine type A and B. And in this work, it is assumed that requests of intermediate materials are limited to be required from Type 2 to Type 1 and 3. As for the factories of Type 3, intermediate materials required for the final products are processed the factory itself.

4.2 States of a production system

In this work, proper selection of factory for each of intermediate material requirement according to the conditions of the whole production network and the feature of the requirements is intended. Basically, next two points are concerned in representation of the states.

1. Vacant time in the current schedule (This means a time to spare for new requirements.)
2. Whole amount of requests

Concretely, a state of each intermediate material requirement is described as a string composed of six characters (See Fig. 2.).

1 The first character is one of "A" or "B". It represents the type of machine for film creation.

Table 1. The concrete values of L, M, S

	2	3 (α)	3 (β)	4-6
L	25	90	70	40
M	15 – 25	50 - 90	40 - 70	20 - 40
S	< 15	< 50	< 40	< 20

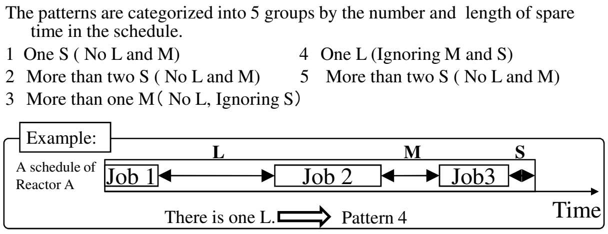

Fig. 3. Representation of vacant time in schedules

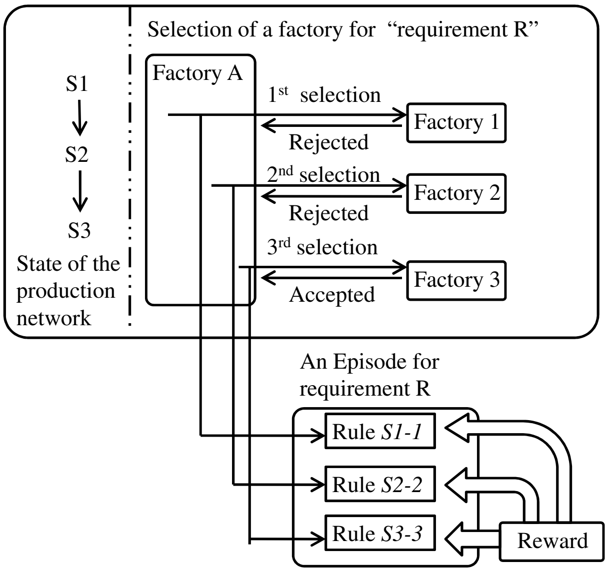

Fig. 4. The process of learning of rules

2 The second character is one of "S", "M" or "L". It represents the amount of the requested film. The first column in TABLE 1 shows the correspondence of an concrete amount and the three characters.

3 The third character is one of "S", "M" or "L". It represents the whole amount of the requested film which is the same type with the requests. Besides, requirements which are accepted already by some factory are ignored. The second and the third column in TABLE 1 shows the correspondence of an amount and the three characters.

4 – 6 Each of the 4 th , 5 th and 6th characters is one of "1", "2", "3", "4" or "5". They represent the pattern of vacant time in the factory 1, 3, and 4.

The 5 patterns used in from 4th to 6th characters are categorized by the number and the length of vacant time in the schedule of the corresponding factory (Fig. 3.). The length of vacant time is represented by "S", "M" or "L". The 4th column in TABLE 1 shows the correspondence of amounts and the three characters. Patterns are categorized by the number of each type of vacant time:

Pattern 1 There is only one S. (No M ,L).

Pattern 2 There are more than two S. (No M, L).

Pattern 3 There are more than one M. (No L).

Pattern 4 There is one L.

Pattern 5 There are more than two L.

4.3 Process of Learning Rules

There are three actions (selection of factory 1, 3 or 4) for each state. A combination of each of the three actions and a state form "rules". The module of reinforcement learning observes the target production system and obtains data about current condition of the system. Next, the reinforcement learning module identifies the state of each intermediate material requirement. Then, the module selects a rule from three rules of the identified state.

Episodes of Profit Sharing start when requirement of intermediate materials are generated (See Fig.4.). A rule is selected from rules which are corresponding to the state of the requirement and the production network at that time. The selected rules are added to the episode of the requirement. The request of intermediate material is sent to the factory which is indicated by the selected rule. If the selected factory accepts the requests, a reward is given to the rules in the episode. On the other hand, in case that the request is not accepted, another factory is selected and the request is sent to the new factory. Since some requirements may be accepted, the state of each intermediate will be changed at each "re-scheduling".

5 IMPROVEMENT OF STATES REPRESENTATION

5.1 Introduction of new representation

First, preliminary experiments were carried out in order to examine the performance of the original representation. From the results of the experiments, several problems below were found.

- There are 2250 states under the proposed "6 characters"

Table 2. New ranges of spare times

Set No.	original	1	New 2
L	> 40	> 3000	> 5000
M	20 40	1500 3000	3000 3000
S	< 20	< 1500	< 3000

Table 3. Appearance of states

Set No.	original	1	2
Appeared states	307	1301	1349

representation, but the number of states which appeared frequently was almost 300 in the experiments.

- Appearance of states with Type 3 (including M size spare time) was very rare.
- 95% of appearedd states were type 4 or 5 (including L size spare time).

These results show that the ranges of a spare time used in the previous work did not match time length of production jobs. Concretely, the ranges of L and M have to be expanded.

So, new several candidates of ranges to represent spare time of machines were introduced. Table 2 shows two examples of the range sets used in the experiments below.

5.2 Experiments

A series of experiments was carried out in order to confirm that factory selection rules obtained under the new spare time ranges had ability to separate states of production networks properly keeping a performance of the rules. Since the model production network are same with the one used in the previous work (Yamaba [2]), the details is omitted here.

The numbers of appeared states are shown in Table 3. This result shows that the new ranges seems separate states of a production network more adequet then original one.

Table 4 shows the number of whole orders and rejected orders in the simulation experiments operated using the supplier selection rules obtained under the each spare time ranges. The column of "random" is the result of the experiment under the condition that supplier selection was carried out at random. Percentages of a rejected order ratio to the one when random selection was used are also shown in Table 4. The percentages show that the obtained rules can select proper factories.

6 CONCLUSION

An attempt was made to obtain operation rules by reinforcement learning in order to manage production network effectively.

The new candidates of ranges were proposed and it was confirmed that the new ranges separated the sates of the target production networks more adequate keeping the performance of the obtained supplier selection rules.

Table 4. Comparison of performance

Set No.	original	1	2	random
Whole orders	170098	171556	172070	171234
Rejected orders	22216	22907	23327	26192
Rejected ratio	13.06	13.35	13.56	15.30
Percentage to random	85.39	87.29	88.63	100

REFERENCES

[1] Yamaba H and Tomita S (2006), Network-based Support System for Decentralized Scheduling of Distributed Production Systems through Man-Machine Collaboration. Proceedings of 2006 IEEE International Conference on Indeutirial Informatics:1285-1290

[2] Yamaba H, Matsumoto S and Tomita S (2008), An Attempt to Obtain Scheduling Rules of Network-based Support System for Decentralized Scheduling of Distributed Production Systems. Proceedings of 2008 IEEE International Conference on Indeutirial Informatics:506-511

[3] Arai S, Miyazaki K and Kobayashi S (1998), Cranes Control Using Multi-Agent Reinforcement Learning. International Conference on Intelligent Autonomous System(5):335-342

[4] Yamaba H, Yoshioka H and Tomita S (2004), Acquisition of AGV Control Rules Using Profit Sharing Method and Evaluation of the Rules. Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies(2):405-411

[5] Yamaba H and Tomita S (2010), A Design Support System for Discrete Production Systems with AGVs – Object-Oriented Simulator Generator and Acquisition of AGV Control Rules Using Reinforcement Learning– . Kagaku Kogaku Ronbunshu 36:127-135