JAIST Repository
https://dspace.jaist.ac.jp/

Title	フローショップ・スケジューリング問題への強化学習の適用
Author(s)	田中, 雄介
Citation	
Issue Date	1999-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1320
Rights	
Description	Supervisor:平石 邦彦, 情報科学研究科, 修士

Japan Advanced Institute of Science and Technology

An Application of Reinforcement Learning to Flow Shop Scheduling Problems

Yusuke Tanaka

School of Information Science,
Japan Advanced Institute of Science and Technology

August 13, 1999

Keywords: flow shop scheduling, reinforcement learning, artificial neural network, genetic algorithm.

The feasibility of applying reinforcement learning (RL) to manufacturing scheduling problems is studied. Flow shop scheduling problems, which objective is to minimize the maximum completion time (denotated as $F||C_{max}$ problems), are dealt with in this study in order to examine the feasibility.

A precedent study reports that the application of RL to a specific practical problem, minimizing the maximum completion time in job shop scheduling, realized the better feasible solution and the faster computation comparing to the former solution methods. One of the remarkable points of the study is that, in the process of formulation into a RL problem, the search space of the problem is squeezed into the smaller one by utilizing the heuristics for solving the problem.

Concerning $F||C_{max}$ problem with two machines, the optimal solution is given by Johnson's algorithm, whereas more than two machines, it is hard to obtain the optimal except for the full search. The reasons this study deals with $F||C_{max}$ problems, the special case of a job shop problem, are that the optimal rule can be utilized for the RL formulation, and that the effectiveness of the formulation can be evaluated, for the problem hard to obtain the optimal solution except for the full search.

In this paper, the main objectives are as follows:

- To propose formulations of the scheduling problems as RL problems, where any optimal solution cannot be obtained except for the full search.
- To examine the effectiveness of the above formulations , through their implementation, and to evaluate the feasibility to apply RL.

Copyright © 1999 by Yusuke Tanaka

First, the RL formulations are proposed by utilizing Johnson's algorithm, while considering the application into the three-machine problem. RL consists of five elements, which are an environment, an agent, a policy, a reward, and a value function. The environment of this problem has machines and jobs. The agent makes the decision on the job processing order in the waiting queue in front of the first machine.

There are two ways of formulating the decision making on the job processing order into a sequential decision making of RL. The one is to improve sequentially the schedule, the job processing order of all jobs in the queue. However, the other way is adopted, to choose a job to be processed sequentially until all jobs are processed.

A policy decides an agent's action corresponding to the state it senses. The policy to maximize the accumulated reward is an agent's target and is acquired by a learning of an agent. The quality of an agent's learning is greatly affected by the formulation of a state and an action. An action is to choose a job to be processed in the waiting queue of the first machine. It is better to select some candidate jobs among the queue by utilizing the heuristics to solve a problem if known, in order to acquire scheduling rules and to obtain the improved solutions. Three types of candidate job sets are formulated as follows.

a1) A job set which exploits the regularity of a job processing sequence obtained by Johnson's algorithm and which limits the number of candidate jobs

a2) A job set which partially exploits the regularity of a job processing sequence obtained by Johnson's algorithm

a3) A job set which contains all jobs in the queue

A state is formulated into the following seven features. The processes of the first and the second machine are represented as the former process and the processes of the second and the third machine as the latter process.

s1) The number of jobs remained in the queue

s2) The number of jobs with the longer processing time of the former process in the queue

s3) The ratio of the former machine workload remained in the queue

s4) The ratio of the latter machine workload remained in the queue

s5) The Boolean value whether the job with the minimum processing time in the former process is now waiting in the queue

s6) The Boolean value whether the job with the minimum processing time in the latter process is now waiting in the queue

s7) The utilization of the latter machines

The objective of the problem leads a reward to be given as the maximum completion time at the terminal state. $TD(\lambda)$ method is employed to update a value function, which is implemented into an artificial neural network as an approximated function. In addition, an RL agent, which learns the superior solution obtained by a genetic algorithm (GA), is proposed in order to decrease the number of training tasks of the RL agent.

Solving a two-machine problem preliminarily, which optimal solution is obtained by Johnson's algorithm, the formulation of the state **s2** and the action **a1** realized the same solution as Johnson's. Besides, several combinations of state and action formulations were experimented, one of which realized the maximum completion time 7 % greater than the optimal solution. Even when the formulation was not perfect as shown, an agent learned and obtained the improved schedules. This implies that an agent has potential to learn effectively even in the problem domain where it is hard to obtain any optimal solution except for the full search.

Solving a three-machine problem, an agent performed equivalent to that for two-machine problem. The result, however, was not superior to that obtained by Johnson's-based algorithm, which applied Johnson's algorithm after the three-machine problem was converted to a two-machine problem.

An RL agent with GA obtained as a stable solution as an agent without GA in less numbers of training tasks.

This study is concluded as follows:

- The several formulations of flow shop scheduling problems into RL problems have been proposed.
- Both in a two-machine problem where the optimal solution is given by Johnson's algorithm and in a three-machine problem where the optimal solution is not given except for the full search, the formulation mentioned above has been observed effective by the results of their implementation.

The future study topics are to continue brushing up RL formulation, and to comparing RL to other probabilistic search algorithms. It is worth studying a learning agent capable even in the more complex and more extensive environments, in terms of a class of scheduling problems, from flow shop to job shop, and in terms of an objectives, from minimizing the maximum completion time to other objectives.