Software-Technological Complex for Adaptive Control of a Production Cycle of Robotic Manufacturing

Viktor Smorodin
Department of Mathematical Problems
of Control and Informatics
Francisk Skorina Gomel State University
Gomel, Belarus
Email: smorodin@gsu.by

Vladislav Prokhorenko
Department of Mathematical Problems
of Control and Informatics
Francisk Skorina Gomel State University
Gomel, Belarus
Email: snysc@mail.ru

***Abstract*—An approach is being proposed for constructing a new generation intellectual system based on OSTIS technology for decision making during realization of adaptive control procedures for technological cycle of robotic manufacturing based on the means of software-hardware coupling. At the basis of the decision making intellectual system lays the idea of using neural network controllers that solve the task of searching for optimal maintenance strategy for a technological cycle of robotic manufacturing. A formalization of such a system is being proposed based on OSTIS technology implementation.**

***Keywords*—technological production process, parameters of operation, probabilistic network graph, state indicators, adaptive control, neural network, LSTM, policy-gradient, reinforcement learning**

I. Introduction

The modern convergence direction of research in the sphere of intellectual systems development [1] requires creation of the corresponding software with elements of cognition based on semantically compatible artificial intelligence technologies. Such a direction must also include the creation of computer systems, that provide intellectualization of making analytical decisions, which is directly related to adapting control processes for complex technological systems (technological objects) in real time, creation of semantically compatible knowledge bases in the sphere of analysis of dynamic systems operation and optimization of complex technical systems operation based on them through the creation of open source software for intellectual decision making systems.

When constructing a software solution for solving complex control-related tasks it is important to have a universal interface that would provide semantic compatibility for its elements, allowing their interchangeability and independent development with simple integration. OSTIS technology can serve as a means of achieving this goal and providing a universal platform for connecting various separate problem solvers [2].

Technological systems that can be formalized as probabilistic network graph structures and mathematical models of semi-Markov processes are the object under study in this paper [3].

Adaptive control of a technological cycle is meant as the ability of a control system adequately react on external disturbances and standard control effects with changing the corresponding parameters of control during the system operation.

Optimal control in the scope of this paper is meant as a formalized by a neural network structure of an adaptive control of a technological cycle that is constructed in the base nodes of a probabilistic network graph structure or semi-Markov network model within the chosen quality criteria. The formalization of the control system and mathematical models of the object under study is based upon the authors' scientific research and development in the sphere of simulation modeling of complex technological systems [4].

In this paper a task of optimal maintenance strategy search is being considered for a technological cycle with implementation of reinforcement learning methods based on the selection of criteria chosen by user. An approach is proposed for solving such tasks based on neurocontroller usage that is trained sing policy gradient methods [5].

II. Formalized description of a technological cycle

A technological cycle is understood as a sequence of actions and operations, based on which the manufacturing of products is achieved. There are N technological nodes $(machines)$ in cycle M_i . During the execution of a cycle K operations O_j are being run sequentially. For each operation are given the execution time $t_{(O_i)}$, the set of nodes $\{M_{ijk}\}$, that operate in mode r_j (M_i) for the current operation. The set and content of such operations are defined by the corresponding technological production process.

During the execution of O_j operation an equipment failure of the i-th type node can occur, which demand pausing the cycle and performing maintenance and repair actions. The costs for repairing the i-th type node CM_i and costs for liquidating the consequences of it's failure during the cycle operation CMO_i are given.

When the i-th type node fails during the execution of the corresponding operation it is possible that an emergency may occur. The costs for repairing the i-th type node in case of emergency CE_i and the costs for liquidating the consequences of emergency CEO_i are given.

Before the cycle execution has started maintenance actions may be performed - one or more of the nodes may be checked and repaired.

The maintenance of all kinds is performed by the manufacturing facility personnel, that has a corresponding qualification. The available trained personnel is a limited resource and no more than L nodes can be repaired at the same time. In case a repair is necessary and the required personnel in unavailable, it is necessary to wait until one of the current repair operations ends. The repair times are not static and of probabilistic nature. The costs for cycle not operating (as a result of nodes failure, repair operations, maintenance, liquidation of emergencies) are also given - $CI(T)$ for the non operating period T.

It is assumed that the technological production cycle has integration of means of software-hardware coupling that allow transferring node observation data into the control system when cycle operates, as well as a system of processing recommendations for cycle maintenance that are produced by the control system.

III. Description of the simulation model that is used in this paper

For implementation of a simulation model in the given formalization the data is used:

- distributions for the duration of non-failure operation for the nodes of i-th type $F_{ir}(t_{wf})$ in mode $r(M_i)$;
- distributions for the restoration (repairs) for the nodes of i-th type after a failure $F_{if}(t_r)$;
- distributions for the liquidation of emergency for the nodes of i-th type $F_{ife}(t_{re})$;
- probabilities of emergency during a failure for the nodes of i-th type P_{ie}.

The simulation model operates during the given time period, it restarts the production cycle and, possibly, performs the maintenance actions before each start. The technological cycle control system is being used to make decisions regarding the necessity and contents of the maintenance procedure.

The data describing the current condition of the technological cycle nodes - duration of non-failure operation for all node M_i is passed inside the control system. Based on the control system recommendations the maintenance for the nodes is performed.

IV. Possible approaches to solving the task

When considering a task of such type the method for constructing an optimal strategy is not obvious which makes the usage of traditional supervised learning algorithms problematic. The complex structure and the nature of the possible solutions space in this task make it sensible to consider the reinforcement learning group of algorithms.

Analysis of the modern state of developments in the artificial intelligence field demonstrates that two most effective groups of reinforcement learning algorithms exist for solving complex control tasks:

- value-based - when controller is trained to estimate the future rewards for the actions it selects;
- policy-based - when controller is trained to predict distribution of actions that would lead to the choice of optimal action-selection policy.

In this paper a policy gradient neural network controller will be used for the task under consideration.

Implementation of the reinforcement learning methods implies construction of an environment in which the agent performs actions. The agent selects actions based on the current observations of the environment, and based on the actions performed and the possible changes in the state of the environment a reward is being calculated and may be observed by the agent.

In this paper the environment in which the agent operates is the control system of the technological production cycle that makes available of agent's observation of all nodes M_i non-failure operation duration. Based on the agent's action selection the decision making system forms requests for the maintenance of the technological cycle nodes. When the agent is being trained jointly with the simulation model reward calculation is also done.

V. Shaping the reward function

The reward shaping plays in important role, as it defines the agent's behavior that it learned during training. The choice of the reward function allows to select for optimization the criteria that user prioritizes.

The approach used in this paper includes into reward shaping such components as cycle non-failure operation time (R_{nop}), total sum of maintenance and emergency liquidation costs (R_{cost}), total number of nodes failures (R_f), including the ones that resulted in emergency (R_{fe}), total number of maintenance performed per cycle (R_{rep}). Each of the reward components is present in the equation with a weight coefficient α_i, which characterizes the importance of the component.

During the agent training the value of the reward function is calculated as following:

$$R = \alpha_1 R_{nop} + \alpha_2 R_{cost} + \alpha_3 R_f + \alpha_4 R_{fe} + \alpha_5 R_{rep}$$

VI. Policy gradient

In the policy-based methods instead of the approximation of a numeric function that estimates rewards that agent receives from the environment as a result of his actions, the policy function for action selection is being constructed directly, that connects environment states with agent's actions. The action selection policy is parameterized by the trained parameters of the model that is used to control agent.

Numeric function (reward function) in this case can be used to optimize policy regarding the trained parameters but is not used for action selection. Stochastic action selection policy gives the probability distribution for the possible actions. Such

policies are often used in the partially-observable environments when uncertainty exists.

It was shown that for some classes of tasks the policy based methods converge faster than value-based (Q-learning), and also are preferable when the action selection space is of large dimension [5]. The convergence towards at least the local quality maximum is guaranteed.

Policy π is parameterized by the trainable parameters θ.

$\pi_\theta(a|s) = P[a|s]$

This policy returns distribution of actions a when the observable state of the environment is s.

In order to find the values for trainable parameters an optimization problem must be solved for the quality estimation function $J(\theta)$.

$J(\theta) = E_{\pi\theta}(\sum \gamma r)$

Rules to update trainable parameters on the t step:

$\theta_{(t+1)} := \theta_t + \alpha \bigtriangledown J(\theta_t)$

According to the Policy Gradient Theorem[6]

$\bigtriangledown E_{\pi\theta}(r(\tau)) = E_{\pi\theta}(r(\tau) \bigtriangledown log\pi_{\theta(\tau)})$,

which can be transformed as

$\bigtriangledown E_{\pi\theta}(r(\tau)) = E_{\pi\theta}(r(\tau)(\sum_{t=1}^{T} \bigtriangledown log\pi_\theta(a_t|s_t)))$.

The REINFORCE algorithm that in meant to train the agent to perform actions according to the policy that results in maximization of the future rewards could be written like this [5]:

1. Initialize parameters θ

2. Generate an episode in which agent interacts with the environment with $\{S_i\}$, $\{A_i\}$, $\{R_i\}$ – sequences of length T of the observed environment

3. For each step t calculate discounted reward

$G_t := \sum_{k=t+1}^{T} \gamma^{k-t-1} R_k$

4. Update the parameters by a rule (perform a gradient ascent)

$\theta := \theta + \alpha\gamma^t G \bigtriangledown_\theta ln\pi(A_t|S_t, \theta)$

5. Repeat steps 2-4

VII. Neural network structure choice

For the agent control recurrent neural network based on multi=layer perceptron with LSTM block is being used. As we are working with policy gradient the network output must return the distribution of action probability, thus softmax is used. Network structure:

1) Dense x64 ReLU;
2) Dense x64 ReLU;
3) LSTM x32 ReLU;
4) Dense x6 Softmax.

VIII. Results of the training

One cycle execution in the normal condition takes 48 units of model time. One simulation lasts for 64*48 = 3072 units of model time.

On figure 1-5 the graphs show how various metrics change during the training that lasts 500 episodes.

Distribution of the most frequently produced by the system recommendations for maintenance (7) corresponds with the ones expected based on the chosen simulation parameters for

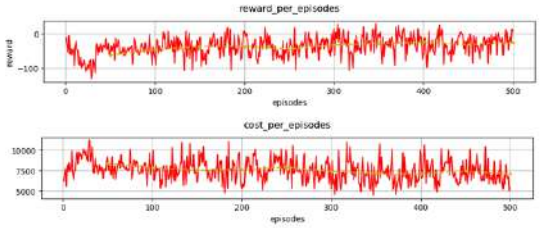

Figure 1. Total reward that agent receives during one simulation run. Total costs for executing the cycle during one simulation during training. A tendency to the increase of the reward and decrease of the costs can be observed during training.

Figure 2. Number of maintenance operations performed according to the system's recommendations.

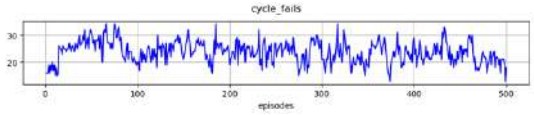

Figure 3. Number of failures during the simulations

Figure 4. The average time of normal operation of the cycle during simulation.

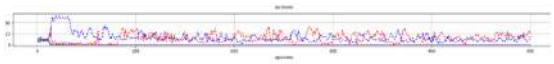

Figure 5. Distribution of recommendations for maintenance, that are most frequently generated by the system.

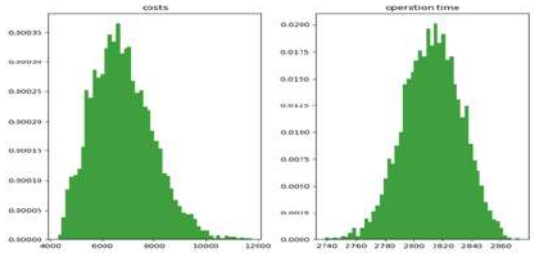

Figure 6. Histogram of distribution of costs and normal operation of the cycle during 5000 of test runs of the simulation

the distribution of the normal operations duration $F_{ir}(t_{wf})$ and probabilities of emergencies P_{ie} for nodes M_0, M_2, the ones for which the emergencies probabilities happens most frequently.

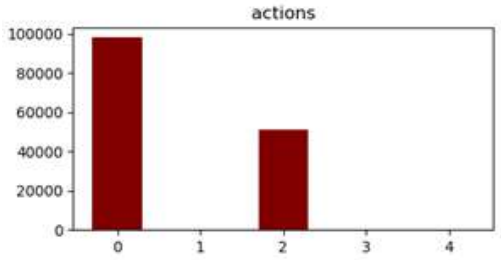

Figure 7. Distribution of the actions most frequently selected by the system

IX. Formalizing the controller as a problem solver of a decision making OSTIS system

In order to provide a possibility for the integration of the developed system concept with other intellectual systems a formalization of the proposed decision-making system based on OSTIS technology is proposed.

In the context of the OSTIS technology problem solvers are based on the multi-agent approach. According to this approach the problem-solver is implemented as a set of agents which are called sc-agents. These agents have shared memory and can exchange data through sc-texts. It is important to note that agents can be non-atomic, meaning that two or more sc-agents are operating to provide functionality for such an agent.

The problem solver for the task under consideration can be viewed as a decomposition of abstract non-atomic sc-agent.

abstract non-atomic sc-agent of cycle maintenance recommendation system

⇒ *decomposition of abstract sc-agent**:

{• *abstract sc-agent of interaction with the observation system*

• *abstract sc-agent of forming recommendations*

• *abstract sc-agent of forming maintenance requests*

}

1) abstract sc-agent of interaction with the observation system – performs extraction of observations from the means of hardware-software coupling in the technological production cycle, it initializes the operation of agent responsible for proposing recommendations.
2) abstract sc-agent of forming recommendations – based on the received observations initializes the operation of neurocontroller for receiving maintenance recommendations.
3) abstract sc-agent of forming maintenance requests – based on the data received from the agent of forming recommendation forms requests for maintenance for the corresponding means of hardware-software coupling.

X. Conclusion

In this paper an approach is proposed to constructing an intellectual system based on OSTIS technology for decision making when realizing the adaptive control procedures for the technological cycle.

The maintenance decision making system for the technological cycle is based on the neural network controller that is constructed using the methods of reinforcement learning for solving the task of optimal strategy search for the maintenance of the technological cycle.

A formalization of the decision making system based on the OSTIS technology is proposed, that allows integration into other intellectual systems when solving the task of technological production cycle control.

References

[1] V. Golenkov, N. Gulyakina, N. Grakova, I. Davydenko, V. Nikulenka, A. Eremeev, V. Tarasov. *From Training Intelligent Systems to Training Their Development Tools. Open Semantic Technologies for Intelligent Systems (OSTIS)*, Minsk, Belarussian State University of Informatics an Radioelectronics Publ., 2018, iss. 2, pp. 81–98.

[2] V. Golovko, A. Kroshchanka, V. Ivashenko, M. Kovalev, V. Taberko, D. Ivaniuk. *Principles of decision-making systems building based on the integration of neural networks and semantic models*, Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems], 2019, pp. 91-102.

[3] Smorodin, V.S., Maximey, I.V. textitMethods and means of simulation modeling of technological production processes, Gomel, F. Skorina State University, 2007.

[4] Maximey, I.V., Smorodin, V.S., Demidenko, O.M. *Development of simulation models of complex technical systems*, Gomel, F. Skorina State University, 2014. 298 p.

[5] Sutton, R. S., Barto, A. G. *Reinforcement Learning: An Introduction*, Cambridge, The MIT Press, 1998./

[6] Sutton, R. S., McAllester, D., Singh S., Mansour Y. *Policy Gradient Methods for Reinforcement Learning with Function Approximation, Advances in Neural Information Processing Systems 1*, NIPS 1999./

Программно-технологический инструментарий адаптивного управления технологическим циклом роботизированного производства

Смородин В.С., Прохоренко В.А.

Предлагается подход к построению интеллектуальной системы нового поколения в рамках технологии OSTIS для принятия управляющих решений при реализации процедур адаптивного управления технологическим циклом роботизированного производства на базе средств программного-аппаратного сопряжения.

В основе интеллектуальной системы принятия решений лежит идея применения нейросетевых контроллеров, решающих задачи поиска оптимальной стратегии обслуживания технологического цикла роботизированного производства. Предлагается формализация подобной системы в рамках применения технологии OSTIS.

Received 16.11.2022