Journal of Manufacturing Systems xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Multi-objective optimization of the textile manufacturing process using deep-Q-network based multi-agent reinforcement learning

Zhenglei He [a,*], Kim Phuc Tran [a], Sebastien Thomassey [a], Xianyi Zeng [a], Jie Xu [b,c], Changhai Yi [b,c]

[a] Univ. Lille, ENSAIT, GEMTEX–Laboratoire de Génie et Matériaux Textiles, F-59000 Lille, France
[b] Wuhan Textile University, 1st, Av Yangguang, 430200, Wuhan, China
[c] National Local Joint Engineering Laboratory for Advanced Textile Processing and Clean Production, 430200, Wuhan, China

ARTICLE INFO

Keywords:
Deep reinforcement learning
Deep Q-networks
Multi-objective
Optimization
Decision
Process
Textile Manufacturing

ABSTRACT

Multi-objective optimization, such as quality, productivity, and cost, of the textile manufacturing process is increasingly challenging because of the growing complexity involved in the development of textile industry in the upcoming big data era. It is hard for traditional methods to deal with high-dimension decision space in this issue, and prior experts' knowledge is required as well as human intervention. This paper proposed a novel framework that transformed the textile process optimization problem into a stochastic game, and introduced deep Q-networks algorithm instead of current methods to approach it in a multi-agent system. The developed multi-agent reinforcement learning system applied a utilitarian selection mechanism to maximize the sum of all agents' rewards (obeying the increasing ε-greedy policy) in each state, to avoid the interruption of multiple equilibria and achieve the correlated equilibrium optimal solutions of the textile process. The case study result reflects that the proposed MARL system can achieve the optimal solutions for the textile ozonation process, and it performs better than the traditional approaches.

1. Introduction

The textile manufacturing process adds value to fiber materials by converting the fibers into yarns, fabrics, and finished products [1]. Under the arousing global competition, textile companies have to face the challenges of cost reduction and performance improvement. There is a growing public concern on the environment which imposes bounds to the textile manufacturers on the exploitation of power, water and resources. The future development of textile manufacturing relies heavily on product customization and shortened manufacturing cycles since the distributors and consumers are increasingly looking for flexible capacity sensitive to demand variability. To deal with the high degree of variability in materials, processes and parameters, the manufacturers traditionally conduct trial and error, and lean on the expertise and experience [2]. There is a strong need to develop innovative methods to improve the textile manufacturing process.

Since textile manufacturing consists of a very long value chain of processes from raw materials to finished products (a brief example is provided in Fig. 1), the combinations of processes and parameters at different stages could be stochastic and immense when factors of the targeted performance vary in any respect [3–5]. And because of the number of factors such as increasing component (or product) complexity, it is difficult to obtain the optimal scenario of a textile manufacturing process. Meanwhile, the performance of the textile process is always governed by a few criteria and the quality of their significance with an overall objective is different [6]. Thus the optimization problems in this domain always take multiple objectives into account. It is very challenging for the simultaneous optimization of multiple targets in a textile production scheme from high dimensional space.

Scholars tended to employ mathematical programming methods and meta-heuristic algorithms to overwhelm textile manufacturing process optimization problems. Krishna et al. [7] utilized dynamic programming models to find the optimal maintenance policy of sewing machine and to decrease their costs in the textile industry. Majumdar [8] applied linear programming to maximize the overall profit of functional clothing production, and applied goal programming to optimize two conflict objectives, namely ultraviolet protective property and air permeability, of the functional clothing. Chakraborty and Diyaley [9] have

* Corresponding author.
E-mail address: zhenglei.he@ensait.fr (Z. He).

https://doi.org/10.1016/j.jmsy.2021.03.017
Received 30 November 2020; Received in revised form 19 March 2021; Accepted 19 March 2021

comparatively studied four evolutionary algorithms, i.e. artificial bee colony algorithm, ant colony optimization algorithm, particle swarm optimization algorithm (PSOA) and non-dominated sorting genetic algorithm-II (NSGA-II) for searching out the global optimal settings of ring and rotor spinning processes. However, in the background of Industry 4.0, the processes of textile manufacturing are expected to be more intelligent with quick reactivity to the market and adaptation to the big data environment. These classical methods either simplify the case by omitting certain non-essential details to achieve manageable equations based on scarification on the accuracy, or require prior experts' knowledge and human intervention. More importantly, they failed to work flexibly with the problems with respect to high-dimension searching space and continuously arriving data generated of the multi-input and multi-output variables.

This paper proposes a novel multi-objective optimization system with reinforcement learning (RL) and random forest (RF) in a multi-agent system, aiming to assist textile manufacturing firms to optimize the overall process performance and product quality as a whole. Specifically, it formulated the multi-objective optimization problems of the textile manufacturing process into a Markov game paradigm, and collaboratively applying multi-agent deep-Q-networks (DQN) reinforcement learning instead of current methods to address it. Due to the complicated nature of textile manufacturing process with multi-variables and flexibility, the scenario of a process can only be obtained through trial and error or ineligible classical methods. To deal with the future uncertainties, RF is applied to predict the unknown performance of a proposed textile process scenario. The performance of each proposed scenario will be predicted by RF models and got feedback to the agents, and this process is repeated in each time step until agents achieved their objectives. Furthermore, in cooperation with the forecasted performance of scenario, DQN is adopted to obtain the optimal scenario. There are a range of advantages employing the multi-agent DQN reinforcement learning to determine the optimal scenario of textile manufacturing process. First, the DQN is model–free. Instead of the metaheuristic methods requiring a predefined rule or prior knowledge, DQN discovers the optimal setting of process scenario by "learning" from direct interaction with the environment. Second, it expresses the high dimension variables by nonlinear function approximator, namely, deep neural network (DNN). Along with multi-agents improving the computation efficiency, it can reduce the complexity in the present optimization problem. Third, RL is adaptive. The textile manufacturers can autonomously acquire the optimal setting or parameters in an online fashion adapted to different scenarios, considering uncertainties and flexibilities of the materials, devices, designs, and operators.

To the best of our knowledge, this is the first paper that address the multi-objective optimization problem of the textile manufacturing process using DQN based multi-agent reinforcement learning (MARL) system. The main contributions of this paper are summarized below:

(1) Construction of a machine learning-based multi-objective optimization system for the textile manufacturing process.
(2) Formulation of textile manufacturing process optimization problem as a Markov decision process, and solve it by reinforcement learning.
(3) Transforming the multi-objective optimization problems of textile manufacturing into the game-theoretic model, and introducing multi-agent for searching the optimal process solutions.
(4) The application of DQN is extended to the multi-agent reinforcement learning system, which is more applicable and preferred to cope with the complicated realistic problem in the textile industry.

The rest of this paper is organized as follows: Section 2 consists of a comprehensive review of the existing research. Section 3 presents the problem formulation of textile manufacturing process multi-objective optimization and the mathematical representation of the problem in the system model. It is followed by the framework illustrated of the proposed MARL system in Section 4. And a case study of the system application for optimize an advanced textile finishing process is demonstrated in Section 5. Finally, conclusions and future works are discussed in Section 6.

2. Literature review

There have been a variety of works on the textile process multi-objective optimization from the last decades. Sette and Langenhove [10] simulated and optimized the fiber-to-yarn process to balance the conflicting targets of cost and yarn quality. Majumdar et al. [8] optimized the functional clothing in terms of ultraviolet protection factor and air permeability. Mukhopadhyay et al. [11] attempted to optimize the parametric combination of injected slub yarn to achieve the least abrasive damage on fabrics produced from it. Almetwally [12] optimized the weaving process performances of tensile strength, breaking extension, and air permeability of the cotton woven fabrics by searching optimal parameters of weft yarn count, weave structure, weft yarn density and twist factor. These works generally used the prior techniques that combine the multiple objectives into a single weighted cost function, the classical approaches such as weighted sum, goal programming, min-max, etc. which are not efficient as they cannot find the multiple solutions in a single run but times as many as the number of desired Pareto-optimal solutions. Pareto optimal solutions or non-dominated solutions are equally important in the search space that superior to all the other solutions when multiple objectives are considered simultaneously, and the curve formed by joining Pareto optimal solutions is the well-known Pareto optimal front [13].

Heuristic and meta-heuristic algorithms are also broadly investigated and applied in the textile manufacturing industry to approach the Pareto optimal solutions with regard to the multi-objective optimization [14], and evolutionary algorithms such as genetic algorithms (GA) and gene expression programming (GEP) were often the first choice. The study

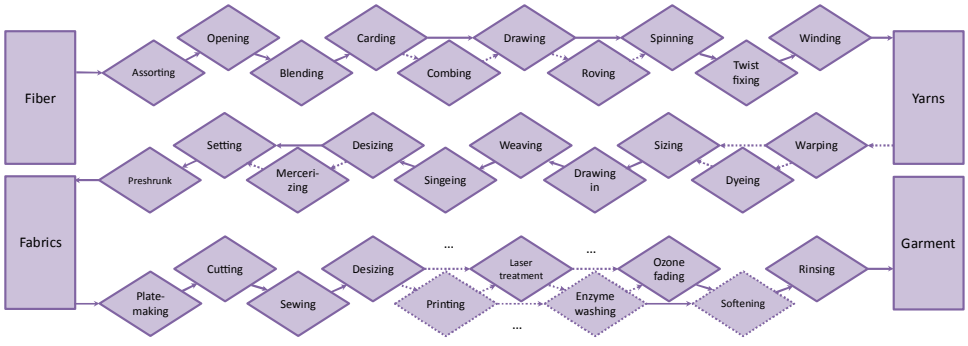

Fig. 1. A general illustration of the textile manufacturing processes from fiber to garment.

described in [15] scheduled the flow-shop of a fabric chemical finishing process aiming at minimal make-span and arresting time of machine simultaneously using multi-objective GA. The study of [16] optimized the electrospinning process performance in terms of fiber diameter and its distribution by searching for optimal solutions with regard to the processing parameters of solution concentration, applied voltage, spinning distance and volume flow rate. The electrospinning process parameters were mapped to the performances by the GEP model, and a multi-objective optimization method was proposed on the basis of GA to find the optimal average fiber diameter and its distribution. Wu and Chang [17] proposed a nonlinear integer programming framework on the basis of GA to globally optimized the textile dyeing manufacturing process. Ghosh et al. [18] optimized the yarn strength and the raw material cost of the cotton spinning process simultaneously with NSGA-II on the basis of two objective function models in terms of artificial neural networks and regression equation. Muralidharan et al. [19] described the combined use of NSGA-II with response surface methodology for the design and control of color fastness finish process to optimize five quality characteristics, i.e. shade variation to the standard, color fastness to washing, center to selvedge variation, color fastness to light and fabric residual shrinkage. Majumdar et al. [20] derived the Pareto optimal solutions using NSGA-II to obtain the effective knitting and yarn parameters to engineer knitted fabrics having optimal comfort properties and desired level of ultraviolet protection. Barzoki et al. [21] and Vadood et al. [22] employed this algorithm with artificial neural networks and Fuzzy logic respectively to optimize the properties of core-spun yarns in the rotor compact spinning process, where the investigated process parameters consist of the filament pre-tension, yarn count and type of sheath fibers, and the objectives were yarn tenacity, hairiness and abrasion resistance for the former but elongation and hairiness for the latter respectively. Apart from the GA frameworks, applications reported of other heuristic or meta-heuristic algorithms for multi-objective optimization in the textile domain also have been presented with synergetic immune clonal selection (SICS), artificial bee colony (ABC) algorithm, ant colony optimization (ACO), and particle swarm optimization (PSO) [9,23]. Meanwhile, simultaneous optimization using the desirability function [24], in addition to the heuristic or meta-heuristic algorithms, was very popular in the textile manufacturing process multi-objective optimization applications as well [25,26].

However, despite the above mentioned efforts on optimization of the textile manufacturing process, they still remain several significant limitations. First, research has taken little into account the high dimensional decision space and the increasing complexity in the textile processes optimization problem derived from growing factors of multi-inputs and multi-outputs in the manufacturing process. And commonly used heuristic methods like genetic algorithms are too time-consuming to be applied in the context of industrial practice when the number of involved variables becomes very large, along with large change intervals [27]. Second, it is expected that the textile manufacturing process reactive quickly to the market and adapt to the big data environment in the Industry 4.0 era. The previously developed system failed to illustrate the capacities of learning from the continuously arriving data to keep updated with the textile process development in this regard, thus it will be invalid when the textile process or applied scenarios vary in the future, and is unpractical to be implemented in the industry. Given this, it is desirable to develop innovative mechanisms for optimizing the textile manufacturing process.

Over the past few years, with the rapid evolution of artificial intelligence, more machine learning algorithms demonstrated increasingly versatile and powerful in the practical application of optimization issues in the industry. It is noticed that considerable research interest has been generated in adopting reinforcement learning (RL) algorithms in this regard [28–36]. RL is a machine learning approach using a well understood and mathematically grounded framework of Markov decision process (MDP) to get the agent acts to maximize its expected cumulative rewards via iterative interaction with an uncertain, unknown, and complex environment. It is model-free and does not rely on prior domain knowledge. The agent observes the environment in terms of state and selects an action at each step time. And according to the received numerical reward derived from the new state and the chosen action, agent will map the perceived environmental states to the probabilities of selecting any possible actions, to maximize the total amount of rewards over the long run. Studies have been reported to solve the optimization problems by using RL. For example, in [37], the pricing strategy optimization in the insurance industry was modeled as a sequential decision problem in terms of a MDP, and the revenue is optimized subject to the client retention by RL algorithm. Rana and Oliveira [32] used RL to model the optimal pricing of perishable interdependent products when demand is stochastic and its functional form unknown, and it is shown that RL can be used to price interdependent products. Similar application of RL for pricing optimization can also be found in references of [28,30,31]. The authors of [38] constructed a deep reinforcement learning (DRL) model to deal with the chemical reactions condition optimization problem, and claimed that it outperformed a state-of-the-art black-box optimization algorithm by using 71% fewer steps. Rocchetta et al. [39] and Kuznetsova et al. [40] have applied RL to address the energy management associated problems for optimizing the operation and maintenance of power grids equipped with prognostics and health management capabilities, and the planning of the battery scheduling, respectively. Mehdi et al. [41] employed the temporal difference based RL methods to reduce the dimension of data in feature selection has been reported. Jasmin et al. [42] have applied the RL to approach the economic dispatch problem. However, most of these previous RL studies focused only on single-objective problems.

It is also noticed that multi-objective optimization problem could be transformed into game theoretic models to be well solved [43,44], and recent developments of multi-agent system for optimizing multiple objectives on the basis of game theory have shown its extreme capability of dealing with functions having high dimensional space [45,46]. The multi agent reinforcement learning (MARL), on the other hand, has been proposed by many contributions for robotics distributed control, telecommunications, traffic light control, and dispatch optimization etc. [47–50]. However, traditional MARL algorithms generally can hardly handle the large-scale problem, the applicability of it was therefore very limited [51]. Yet the deep reinforcement learning (DRL) algorithms, which have been quickly developed in recent years, can make a difference. DQN is one of the DRL algorithms that utilize deep learning tools and strategies of experience replay [52] and fixed Q-target coping with the large-scale issues, has recently been well evaluated in many applications of DRL [33,53,54]. It is found that the high achievement of DRL has been illustrated in many applications in the MARL environment [55–57]. For instance, Wang et al. [58] optimized workflow scheduling with DQN based MARL algorithm. Mannion et al. [48] examined the application of MARL to a multi-objective dynamic dispatch optimization problem. Zhang et al. [59] formulated the large scale city traffic scenario to a MARL environment. However, although there have been several successful examples illustrating the effectiveness of DQN based MARL in optimization problems, very limited work has solved a complex production problem, especially in the textile manufacturing industry. Thus, to bridge the aforementioned research gap, a novel multi-objective optimization framework of the textile manufacturing process using DQN in the MARL environment is presented in this article.

3. Problem formulation

Considering the solution of a textile manufacturing process P is composed and determined by a set of parameter variables $\{v_1, v_2 \ldots v_n\}$, the impacts of these variables on the process performance could be varied a lot from n different respects with uncertainty, as the number of the processes and the related variables in the textile manufacturing industry is enormous and the influences of these variables on the targeted optimization performance are unclear. For example, the longer time was

taken of a textile process generally would lead to the increment of production cost, and a tiny enhance of temperature used in the textile production process could significantly arouse the power consumption, but sometimes the enhanced temperature may promote the process efficiency so that decrease the production cost eventually. Therefore, it is necessary to study the interrelated effects of process variables on process performance. From the engineering perspective, it is important to achieve a solution in the textile manufacturing process that can achieve good quality and avoid idle time, waste and pollutions at the same time. Models that incorporate the information of the process simulating the variation of multiple objective performances from the change of variable in the solutions are rather essential.

Suppose models exist that can map variables $v_1, v_2 \ldots v_n$ of the process solution P to its performance in accordance with m objectives, the performance of a specific solution could be simulated by:

$$f_i(P) = f_i(v_1, v_2 \ldots v_n) \quad for\ i = 1, \ldots m \tag{1}$$

When a decision-maker who wants to find a solution that satisfies m objectives of the process performances that the objectives are non-commensurable and no preference of the objectives related to each other is coming up with the decision-maker. The multi-objective problem could be defined as giving the n-dimensional variable vector $P = \{v_1, v_2 \ldots v_n\}$ in the solution space, finding a vector of p^* that optimizes a given set of m objective functions:

$$f(p^*) = \{f_1(p^*), f_2(p^*), \ldots, f_m(p^*)\} \tag{2}$$

The solution space is generally restricted by a series of constraints, when the domain of $v_j \in V_j$ for $j = 1, \ldots, n$ is known, and representing the m objectives by M, the objective of the problem is to find (3):

$$argmax_{v_j \in V_j}[f(v_1, v_2 \ldots v_n) \mid M] \text{ for } j = 1, \ldots, n \tag{3}$$

Eq. (3) aims at searching the optimal solution of variable settings, while there are always conflicting objectives that satisfying one single target but lead to unacceptable results to the others. A perfect multi-objective solution that simultaneously optimizes each objective function is almost impossible.

To this end, this paper proposes a self-adaptive DQN-based MARL framework where the m optimization objectives are formulated as m DQN agents that trained through a self-adaptive process constructed upon a Markov game.

4. Methodology

4.1. Multi-objective optimization of textile manufacturing process as Markov game

We begin by formulating the single objective textile process optimization problem as a Markov decision process (MDP) in terms of a tuple :{S, A, T, R}, where S is a set of environment states, A is a set of actions, T is the state transition probability function, R is a set of reward or losses. An agent in an MDP environment would learn how to take action from A by observing the environment with states from S, according to corresponding transition probability T and reward R achieved from the interaction. The Markov property indicates that the state transitions are only dependent on the current state and current action is taken, but independent of all prior states and actions [60]. While in the case of a multi-agent system, the joint actions are the result of multiple agents, the MDP is generalized to the stochastic Markov game of {S, A^1, …, A^m, T, R^1,…, R^m}, where S and T are similar to the MDP that are the finite set of environment states and the state transition probability function respectively in a Markov game, whereas differently, m is the number of agents, A^i for $i = 1, \ldots, m$ are the finite sets of actions available to the agent i, R^i for $i = 1, \ldots, m$ are the reward functions of the agent i.

As known that the solution of a textile manufacturing process is affected by a number of variables as $P\{v_1, v_2 \ldots v_n\}$, if the possible value of v_j is $h(v_j)$, the feasible values of the parameter in the process can define the environment space S from $\prod_{j=1}^{n} h(v_j), v_j \in V_j$ impacting the performance of the textile process with regard to the k objectives. These parameter variables are independent of each other and obey a Markov process that models the stochastic transitions from a state S_t at time step t to the next state S_{t+1}, where the environment state at time step t is:

$$St = [s_t^{v_1}, s_t^{v_2} \ldots] \in S \tag{4}$$

RL algorithm trains an agent to act optimally in a given multi-agent environment based on the observation of states and other agents as well as the feedback derived from the interactions, acquiring rewards and maximizing the accumulative future rewards over time from the interaction [60]. In our case, the agents learn in the interaction with the environment and other agents by taking action that can be conducted on the parameter variables $\in P\{v_1, v_2 \ldots v_n\}$ at time step t. Specifically, the action of an agent in a time step t of optimizing a textile manufacturing process in the Markov game, could be adjusting variable v_j to keep (0) or change to up (+) and down (-) with a specific unit u_j subjected to the constraint. As a result, there are 3^n actions in total in the joint action space A and, for simplicity, the action vector A_t at time step t could be:

$$A_t = \left[a_t^{v_1}, a_t^{v_2} \ldots a_t^{v_n}\right], \text{where } a_t^{v_j} \in \{-u_j, 0, +u_j\}, v_j \in V_j \text{ for } j = 1, \ldots, n \tag{5}$$

We define $A = \prod_{i \in m, s \in S} A^i(s)$ for the joint action from overall the agents, where $A^i(s)$ is agent i's finite set of pure actions at state s. It is also defined $A(s) \equiv \prod_{i \in I} A_i s$ and $A_{-i} = \prod_{j \neq i} A_j(s)$, therefore, $A(s) = A_{-i}(s) \times A_i(s)$; in order to distinguish player i, we define $a = (a_{-i}, a_I) \in A(s)$ with $a_i \in A_i(s)$ and $a_{-i} \in A_{-i}(s)$; It is also defined that $\mathscr{A} = \bigcup_{s \in S} \bigcup_{a \in A(s)} \{(s, a)\}$, the set of state-action pairs.

The m objectives of textile manufacturing process optimization are assigned to m agents in this Markov game. Agents are the main elements of this proposed system. Considering optimal management objectives and the formulated Markov game of textile manufacturing process optimization problem in a specific case, and the MARL architecture could be illustrated in in Fig. 2, where the optimization objectives are abstracted as RL agents. Given feedbacks from the RF models integrated with the Markov game environment with state-space formulated in Eq. (4) that consist of all the parameter variables of the simulated textile process, the agents are able to evaluate the values of its actions for adjusting the parameter variables with regard to the state (solution) and consequently improve its policy in the environment to optimize objectively gradually.

As known that apart from the benefits derived from the distributed nature of the multi-agent system such as parallel computation, the experience sharing from different agents also significantly improve the algorithms' performance. Therefore, it is assumed that each agent can observe each other's action and rewards. Then they select the joint

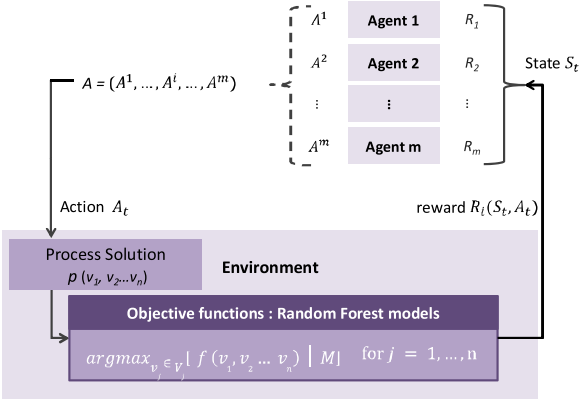

Fig. 2. The Markov game for textile manufacturing process multi-objective optimization in the proposed framework.

distribution (the combination of choices of all agents) which is determined by the actions selected of each agent $(A^1, \ldots, A^i, \ldots, A^m)$.

The state transition probabilities, as mentions that, are only dependent on the current state S_t and action A_t. It specifies how the reinforcement agents take action A_t at time step t to transit from S_t to next state S_{t+1} in terms of $T(S_{t+1} \mid S_t, A_t)$. For all $a_t^{v_j} \in \{-u_j, 0, +u_j\}, v_j \in V_j$, $T(S_{t+1}|S_t, A_t) > 0$ and $\sum_{S_{t+1} \in S} T(S_{t+1}|S_t, A_t) = 1$. The reward achieved by an agent in an environment is specifically related to its transition between states, which evaluates how good the transition agent conducts and facilitates the agent to converging faster to an optimal solution.

When the reinforcement agents perform a joint action A_t at time step t to divert the system from S_t to next state S_{t+1} with transition probability T, each agent would earn reward $R_i(S_t, A_t)$ from (3) of the objective functions. This procedure would be repeated at time t+1 again, and finally, converge agents' behaviors to a stationary policy. Random Forest (RF) is a predictive model composed of a weighted combination of multiple regression trees. It constructs each tree using a different bootstrap sample of the data, and different from the decision tree splitting each node using the best split among all variables, RF using the best among a subset of predictors randomly chosen at that node [61]. In general, combining multiple regression trees increases predictive performance. It accurately predicts by taking advantage of the interaction of variables and the evaluation of the significance of each variable [62]. According to a previous study [63], the random forest (RF) predictive model, constructed using Multivariate Random Forest (MRF) [64], is applied to simulate the textile process in this proposed framework and implement the objective functions (3) to earn the agents' rewards.

Stochastic games are neither fully cooperative nor fully competitive [47]. The performance of multi-objective optimization of our case in stochastic Markov game is determined by the agents' capability of gathering information about the other agents' behavior and the reward functions from the interaction to make a more informed decision thereafter. We consider each DQN agent observes all the other agents' actions and rewards and selects its own joint distribution action along with environment updates. The resulting textile process scenarios are generated through a self-learning and self-optimizing manner. The rewards mechanisms along with the interaction among agents perform a significant function in this respect, so that the proposed system, similar to the study of [58], employs a utilitarian selection mechanism $h = argmax_{A \in \Delta(A(S))} \sum_{i \in M} Q_i(s, a)$ that maximize the sum of all agents' rewards in each state to avoid the interruption of multiple equilibria. Convergence to equilibria is a basic stability requirement of MARL, and the Nash equilibrium is a well-known solution concept for the stochastic game that a joint strategy leading to a status of no agent is incentive to change its strategy. But a correlated equilibrium with increased generality instead of Nash equilibrium is taken into consideration in this issue as it allows agents' strategies to be interdependent. It is a joint distribution of actions from which none of the agents has any motivation to deviate unilaterally. Consequently, the solutions of the textile manufacturing process multi-objective optimization problem are correlated equilibria.

Formally, given a Markov game, a joint stationary policy π leads to a correlated equilibrium when:

$$\forall i \in M,\ s \in S \Big| \sum_{a \in A^{-i}(s)} \pi_s\, Q_i^{\pi}(s, a) \geq \sum_{a \in A^{-i}(s)} \pi_s\, Q_i^{\pi}(s, a') \tag{6}$$

where $A^{-i}(s)$ is the set of action vectors in state s excluding ones of agent i. The above inequality denotes that in state s, when it is recommended that agent i play a, it prefers to play a, because the expected utility of a is greater than or equal to the expected utility of a', for all a'.

4.2. Deep Q-networks reinforcement learning algorithm

Classical RL algorithms, such as the Q-learning and the SARSA (0/λ), are based on a memory-intensive tabular representation (i.e. Q-table) of the value or the instant reward, of taking an action a in a specific state s (the Q value of state-action pair, a.k.a Q(s, a)). These tabular algorithms impede the RL in realistic large-scale applications due to the huge amounts of states or actions involved. The tabular expression not only comes short of recording all of the Q(s, a) in these applications, but also shows poor generalization in the environment with uncertainty.

The deep neural networks (DNNs) is another widely applied machine learning technique coping with large-scale issues and has recently been innovatively combined with the RL to evolve toward deep reinforcement learning (DRL) algorithms. Deep-Q-network (DQN) is a DRL algorithm developed by Mnih et al. [54] in 2015 as the first artificial agent that is capable of learning policies directly from high-dimensional sensory inputs and agent-environment interactions. It is an RL algorithm proposed based on Q-learning, one of the most widely used model-free off-policy and value-based RL algorithms.

The Q-learning agent learns through estimating the sum of rewards r for each state S_t when a particular policy π is being performed. It uses a tabular representation of the $Q^{\pi}(S_t, A_t)$ value to assign the discounted future reward r of state-action pair at time step t in Q-table. The target of the agent is to maximize accumulated future rewards to reinforce good behavior and optimize the results. In the Q-learning algorithm, the maximum achievable $Q^{\pi}(S_t, A_t)$ obeys Bellman equation on the basis of an intuition: if the optimal value $Q^{\pi}(S_{t+1}, A_{t+1})$ of all feasible actions A_{t+1} on state S_{t+1} at the next time step is known, then the optimal strategy is to select the action A_{t+1} maximizing the expected value of $r + \gamma \cdot max_{A_{t+1}} Q^{\pi}(S_{t+1}, A_{t+1})$.

$$Q^{\pi}(S_t, A_t) = r + \gamma \cdot max_{A_{t+1}} Q^{\pi}(S_{t+1}, A_{t+1}) \tag{7}$$

According to the Bellman equation, the Q-value of the corresponding cell in Q-table is updated iteratively by:

$$Q^{\pi}(S_t, A_t) \longleftarrow Q^{\pi}(S_t, A_t) + \alpha[r + \gamma \cdot max_{A_{t+1}} Q^{\pi}(S_{t+1}, A_{t+1}) - Q^{\pi}(S_t, A_t)] \tag{8}$$

where S_t and A_t are the current state and action respectively, while S_{t+1} is the state achieved when executing A_{t+1} in the set of S and A in any given MDP tuples of{S, A, T, R}. $\alpha \in [0, 1]$ is the learning rate, which indicates how much the agent learned from new decision-making experience ($Q^{\pi}(S_{t+1}, A_{t+1})$) would override the old memory ($Q^{\pi}(S_t, A_t)$). r is the immediate reward, $\gamma \in [0, 1]$ is the discount factor determining the agent's horizon.

The agent takes action on a state in the environment and the environment interactively transmits the agent to a new state with a reward signal feedback. The basic principle of Q-learning algorithm essentially relies on a trial and error process, but different from humans and other animals who tackle the real-world complexity with a harmonious combination of RF and hierarchical sensory processing systems, the tabular representation of Q-learning is not efficient at presenting an environment from high-dimensional inputs to generalize past experience to new situations [54].

Q-table saves the Q value of every state coupled with all its feasible actions in a given environment, while the growing complexity in the problem nowadays indicates that the states and actions in an RL environment could be innumerable (such as Go game). In this regard, DQN applies DNNs instead of Q-table to approximate the optimal action-value function. The DNNs feed by the state for approximating the Q-value vector of all potential actions, for example, are trained and updated by the difference between Q-value derived from previous experience and the discounted reward obtained from the current state. While more importantly, to solve the instability of RL representing the Q value using nonlinear function approximator [65], DQN innovatively proposed two ideas termed experience replay [52] and fixed Q-target. As known that Q-learning is an off-policy RL, it can learn from the current as well as prior states. Experience replay of DQN is a biologically inspired mechanism that learns from randomly taken historical data for updating in each time step, which therefore would remove correlation in the

observation sequence and smooth over changes in the data distribution. Fixed Q-target performs a similar function, but differently, it reduces the correlations between the Q-value and the target by using an iterative update that adjusts the Q-value towards target values periodically.

Specifically, the DNNs approximate Q-value function in terms of $Q(s, a; \theta_i)$ with parameters θ_i which denotes weights of Q-networks at iteration i. The implementation of experience replay is to store the agent's experiences $e_t = (S_t, A_t, r_t, S_{t+1},)$ at each time step t in a dataset $D_t = \{e_1, \ldots e_t\}$. Q-learning updates were used during learning to samples of experience, $(S, A, r, S') \sim U(D)$, drawn uniformly at random from the pool of stored samples. The loss function of Q-networks update at iteration i is:

$$L_i(\theta_i) = \mathbb{E}_{(S, A, r, S') \sim U(D)}\left[\left(\left(r + \gamma \cdot \max_{A'} Q(S', A'; \theta_i^-) - Q(S, A; \theta_i)\right)^2\right]\right. \quad (9)$$

where θ_i^- are the network weights from some previous iteration. The targets here are dependent on the network weights; they are fixed before learning begins. More precisely, the parameters θ_i^- from the previous iteration is fixed as optimizing the i_{th} loss function $L_i(\theta_i)$ at each stage and are only updated with θ_i every F steps. To implement this mechanism, DQN uses two structurally identical but parametrically differential networks, one of it predicts $Q(S, A; \theta_i)$ using the new parameters θ_i, the rest one predicts $r + \gamma \cdot \max_{A'} Q(S', A'; \theta_i^-)$ using previous parameters θ_i^-. Every F steps, the Q network would be cloned to obtain a target network $\widehat{Q}$, and then $\widehat{Q}$ would be used to generate Q-learning target $r + \gamma \cdot \max_{A'} Q(S', A'; \theta_i^-)$ for the following F updates to network Q.

4.3. DQN based MARL for multi-objective optimization of textile manufacturing process

It is illustrated in Algorithm 1 of the pseudo-code of the DQN based MARL framework for multi-objective optimization of the textile manufacturing process. And correspondingly, a single episodic running of Algorithm 1 is graphically depicted in Fig. 3. It is shown that, on the basis of local updates of Q-values and policy at each state, the DQN agents interact with the environment (textile solution) and other agents iteratively to learn a correlated equilibrium strategy. The constructed random forest models (RF) approximate the objective performances of the textile process in the framework and feedback to the agents.

To achieve a correlated equilibrium, each DQN agent learns about the correlated equilibrium strategy π^t, where $\pi^{t+1} \in f(Q^{t+1}(s))$. Along with suitable reward mechanisms designed (in this system, the reward of an agent is given by the improvement of the objective performance from last state to current state), the convergence of the DQN-based algorithm in multi-agent settings can be guaranteed. For a more detailed description of the correlated equilibrium strategy, we refer the reader to the reference of [66].

Apart from the aforementioned parameters, it is also necessary to provide the expected process performance or optimization targets (P), the iterative steps for updating Q-value, and the size of experience dataset (D) in the system initialization. The detail of DQN applied follows its mechanism introduced in section 4.2. The given algorithm can work without episodes as the target of agents is to find the optimized solution with regard to the state in the environment satisfying multiple objective of the textile process. However, the lack of exploration of the agent in an environment may cause local optimum in a single running. So we initialize the first state randomly from each sub-state $s_t^{v_i}$ (where parameter variables $v_j \in V_j$), and introduce an episodic learning process to the agent for enlarging the exploration and preventing local optimum. An increasing ε -greedy policy is applied additionally to balance the exploration and exploitation of states at the learning period and optimizing period respectively.

As illustrated in Algorithm 2, increasing ε -greedy is employed with an increment given in each time step from 0 until it equals to ε_{max}. This helps the agents find the best actions in the present state to go to the next state with a possibility of ε that may also randomly choose an action with a possibility of $1 - \varepsilon$ to get a random next state. In this regard, the agents can explore the unexplored states without staying in the exploitation of already experienced states of Q-networks, and plentifully exploit them when the states are traversed enough.

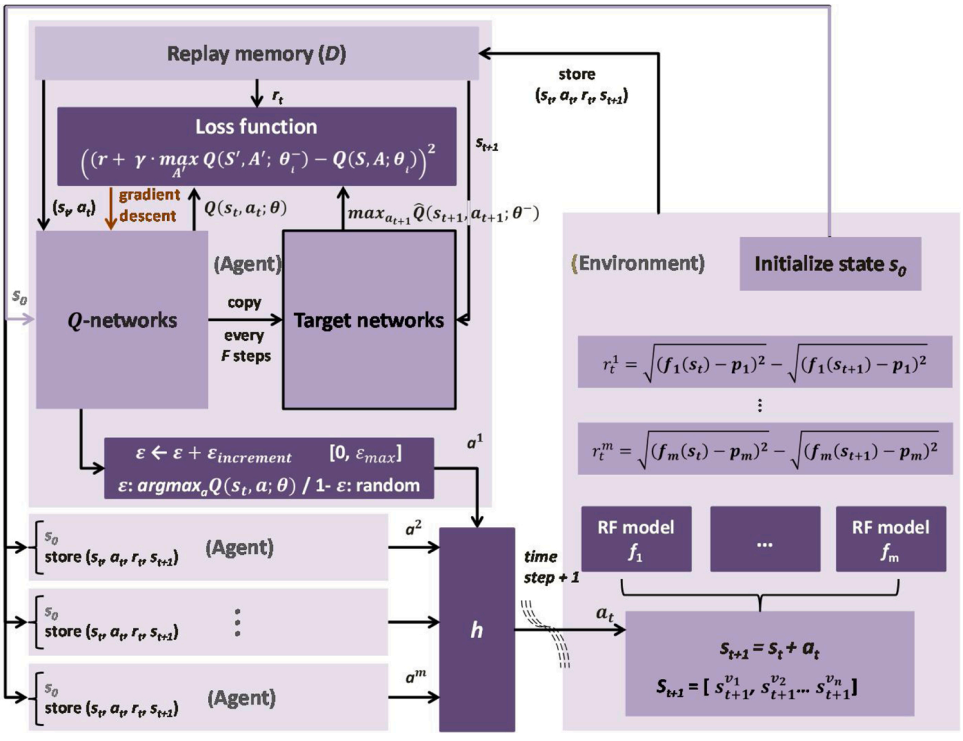

Fig. 3. Flowchart of the algorithm implementing the proposed DQN based multi-agent system for optimizing textile manufacturing process with multiple objectives.

Algorithm 1: DQN based MARL main body:

Input: game $\mathbb{\Gamma}$, RF models for simulating m objective performance $f(f_1 \dots f_m)$, selection mechanism h, expected performance of process P ($p_1, p_2 \dots p_m$,), number of episodes E , number of time steps N, learning rate α, discount factor γ , the step updating DQN F, replay memory size D;

Initialize function Q with random weights θ;

Initialize function $\hat{Q}$ with weights $\theta^{-} = \theta$;

Initialize state $s_0 = (v_1, v_2 \dots v_n)$

For episode =1, E **do**

For time step=1, N **do**

Choose an action randomly or $a_t \in h$ using increasing ε-greedy policy

Execute action a_t , observe next state s_{t+1}

Estimate $f_1(s_t) \dots f_m(s_t)$ and $f_1(s_{t+1}) \dots f_m(s_{t+1})$ to observe r_t ($r_t = \sqrt{(f_i(s_t) - p_i)^2} - \sqrt{(f_i(s_{t+1}) - p_i)^2}$)

Store transition (s_t, a_t, r_t, s_{t+1}) in D

Sample random minibatch of transitions (s_t, a_t, r_t, s_{t+1}) from D

Set $y_i = \begin{cases} r_j & \text{if terminates at step } j+1 \\ r_j + \gamma max_{a'} \hat{Q}(s_{j+1}, a';\ \theta^{-}) & \text{otherwise} \end{cases}$

Perform a gradient descent step on $\left(y_i - Q(s_j, a_j;\ \theta)\right)^2$ with regard to θ

Every F steps reset $\hat{Q} = Q$

$s_t \leftarrow s_{t+1}$

End For

End For

Algorithm 2: Increasing ε-greedy policy

Input: $\varepsilon_{increment}$, ε_{max}

$\varepsilon \leftarrow \varepsilon + \varepsilon_{increment}$ $(0 \le \varepsilon \le \varepsilon_{max})$;

If random(0,1) > ε

Randomly choose action a_t from action space

Else

$a_t = argmax_{A \in \Delta(A(S))} \sum_{i \in M} Q_i(s, a)$

End if

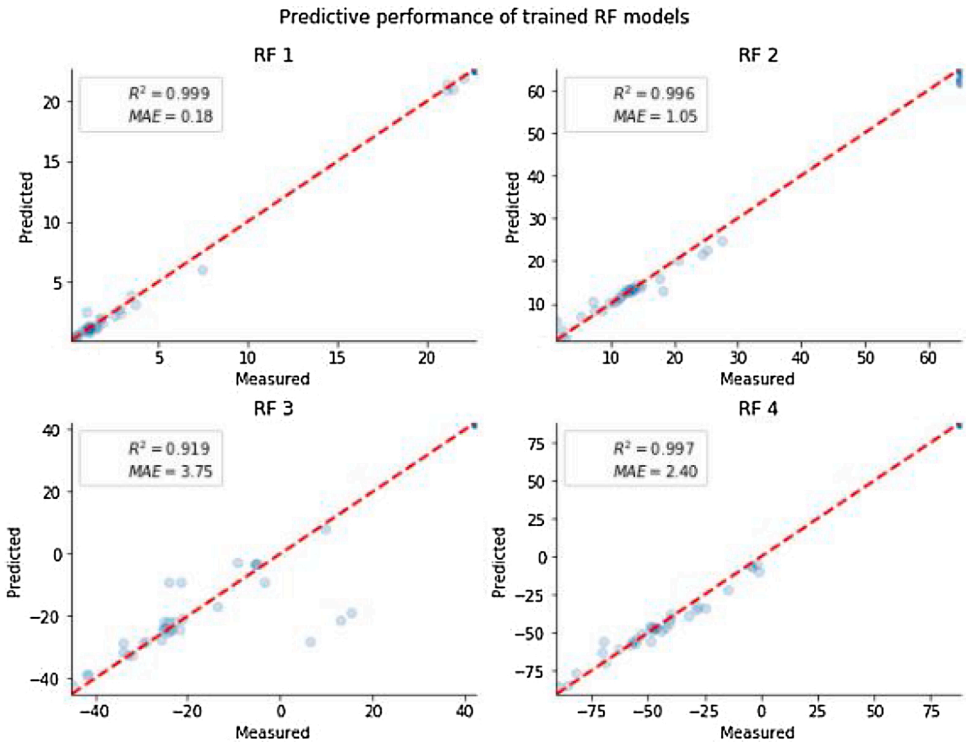

Fig. 4. Predictive performance of the RF models trained in the case study.

5. Case study

5.1. Experimental setup

Color fading is an essential finishing process for specific textile products such as denim to obtain a worn effect and vintage fashion style [67]. But this effect conventionally was achieved by chemical procedures which have an expensive cost, and highly consume water and power, resulting in heavy negative impacts on the environment. Instead, ozone treatment is an advanced finishing process employing ozone gas to achieve color faded effects on textile products without a water bath, so that save power and water, and causes less environmental issues. The interrelated influences of this process on its process performances have been investigated in our previous works [68–72], and according to the experience data with 129 samples we collected from these experimental studies, four random forests (RF) predictive models were constructed for simulating the 4 process performances of the color fading ozonation process. The present case study will attempt to solve the optimization problems of the color fading ozonation process with regard to the 4 process performance using the DQN based MARL system.

The RF models are inputted by four ozonation process parameters (water-content, temperature, *pH* and treating time) to predict four objective color faded performances in terms of color indexes known as *k/s*, *L**, *a**, and *b** of the treated fabrics with the accuracy of $R^2 = 0.999$, 0.996, 0.919 and 0.997 respectively. The predictive performance of the these RF models can be observed from Fig. 4. The *k/s* value indicates the color depth, while *L**, *a**, and *b** are the color indexes from a widely used international standard illustrating the color variation in three dimensions (lightness from 0 to 100, chromatic component from green to red and from blue to yellow from -120 to 120 respectively) [73].

Table 1
Dqn algorithm setting in textile ozonation process case study.

F	D	α	γ	$\varepsilon_{increment}$	a_{max}	E	N
5(>100)	2000	0.01	0.9	0.001	0.9	1	5000

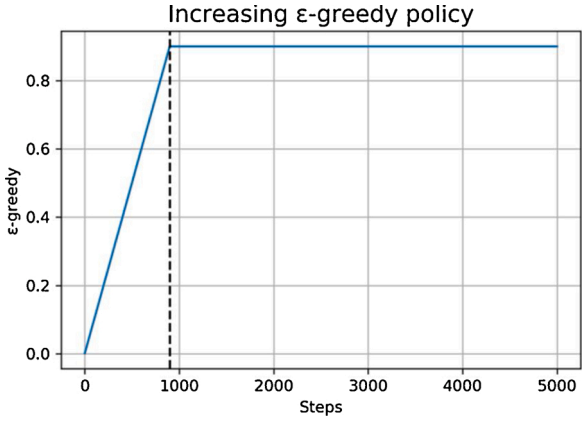

Fig. 5. Increasing ε -greedy policy for choosing action.

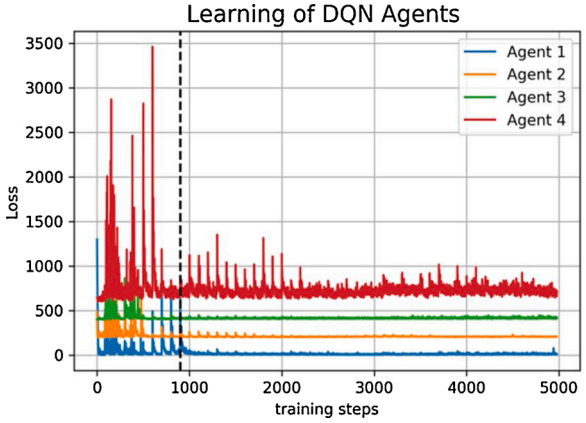

Fig. 6. The loss function of DQN for four agents in the Markov game.

Normally, the color of the final textile product in line with specific k/s, L^*, a^*, and b^* is within the acceptable tolerance of the consumer.

We optimize the color performance in terms of k/s, L^*, a^*, and b^* of the textile in ozonation process by finding a solution including proper parameter variables of water-content, temperature, pH and treating time that minimizes the difference between such specific process treated textile product and the targeted sample. Therefore, there are four agents in the stochastic Markov game, and the state space φ of it is composed by the solutions containing four parameters (water-content, temperature, pH and treating time) in terms of $S_t = [s_t^{v_1}, s_t^{v_2}, s_t^{v_3}, s_t^{v_4}]$. In a time step t, given the adjustable units of these parameter variables $u = 50, 10, 1, 1$ with regard to the constraint ranges of [0, 150], [0,100], [1, 14] and [1, 60] respectively, as the action of a single variable v_j could be kept (0) or changed up (+) / down (-) in the given range with specific unit u, so there are 3^4 =81 actions totally in the action space and the action vector every single agent at time step t is $A_t = [a_t^{v_1}, a_t^{v_2}, a_t^{v_3}, a_t^{v_4}]$, where $a_t^{v_1} \in \{-50, 0, +50\}$, $v_1 \in [0, 150]$; $a_t^{v_2} \in \{-10, 0, +10\}$, $v_2 \in [0, 100]$; $a_t^{v_3} \in \{-1, 0, +1\}$, $v_3 \in [1, 14]$; $a_t^{v_4} \in \{-1, 0, +1\}$, $v_4 \in [1, 60]$.

The transition probability is 1 for the states in the given range of state space above, but 0 for the states out of it. The reward r of an agent at time step t is expected to be in line with how close the agent gets to its target representing the related objective function. We set up the reward function as illustrated below to induce the agents to approach corresponding optimization objective results:

$$r_t = \sqrt{(f_i(s_t) - p_i)^2} - \sqrt{(f_i(s_{t+1}) - p_i)^2} for\ i = 1, \ldots m \tag{10}$$

As demonstrated the pseudo-code of DQN based MARL main body in Algorithm 1, the expected color performances of ozonation process treated samples (p_1, p_2, p_3, p_4, in terms of k/s, L^*, a^*, and b^*) are sampled by experts as 0.81, 15.76, -20.84, and -70.79 respectively to function the system in the present case study. Therefore, there are four agents in this case with respect to their corresponding optimization targets. In addition to the targets, the parameters of DQN agents such as step F for updating Q-networks and replay memory size D, as well as the learning rate α and the discount rate γ for updating loss function, etc., are listed in Table 1. In particular, the F step for updating DQN here denotes that after 100 steps, the Q-networks would be updated at every 5 steps.

The neural networks implemented by TensorFlow [74] are used in our case study to realize Q-networks, and specifically, the networks consist of two layers with 50 and 3^4 hidden nodes respectively, where the last layer corresponds to the actions. Due to the popularity of multi-objective particle swarm optimization (MOPSO) [75] and NSGA-II [76] in the application of textile process, they are considered as the baseline algorithms in this case study to show the effectiveness and efficiency of the proposed DQN-based MARL system for multi-objective optimization of the textile manufacturing process.

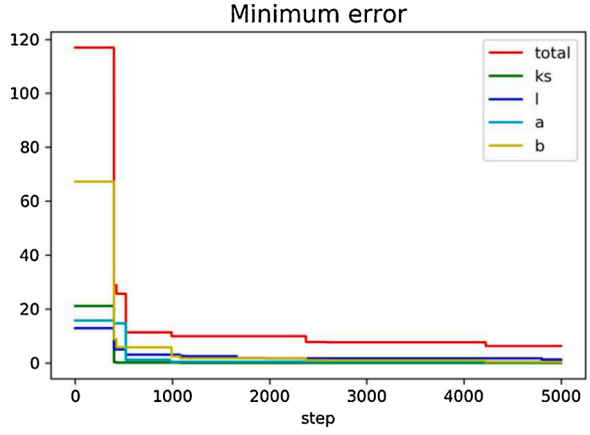

Fig. 7. The minimum error of DQN agents tuned and their sum value versus time steps.

Table 2
Comparison of baseline algorithms and the proposed framework of optimized result.

	Targets	MARL	MOPSO	NSGA- II
k/s	0.81	1.10	0.61	0.33
L^*	15.76	14.08	20.08	25.08
a^*	−20.84	−25.06	−37.06	−43.06
b^*	−70.79	−70.7	−78.7	−85.7
R^2	–	0.999	0.986	0.979
CPU time(s)	–	13.1	17.4	52.3

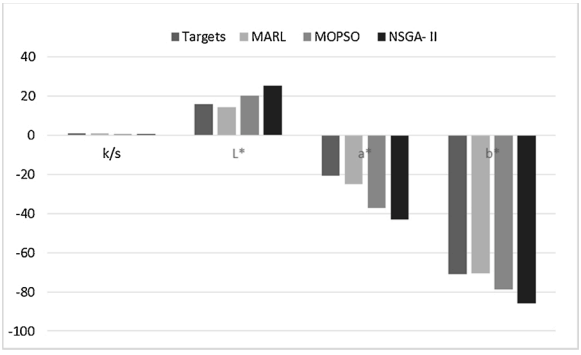

Fig. 8. Comparison of the proposed MARL framework with the baseline algorithms with regard to digital results.

Fig. 9. Comparison of the proposed MARL framework with the baseline algorithms with regard to simulated samples.

5.2. Results and discussion

In the case study, we trained four agents on the basis of the DQN algorithm in a Markov game to optimize an ozone textile process with multiple objectives. As shown in Fig. 5 the increasing ε -greedy policy was used for agents to balance the exploration and exploitation of states. It significantly affects the learning time and quality of learned policies of agents. In this policy, the exploration decays in the first 900 steps. As agents initially lack the information and policy for exploiting possible actions. But increasingly, they can follow its policy exploiting the available information by takes action selection mechanism h, rather than acting randomly. The effects of it could be clearly illustrated on the convergences of DQN agents given in Fig. 6 (for the illustration conveniences, 200, 400, and 600 units of loss are additional given to agent 2, agent 3, and agent 4 respectively). Where the losses of four deep Q-networks with respect to four agents were decreased dramatically after 900 steps agents exploring the environment, and were stable from the 900 steps on. It denotes that the deep Q-networks adapts successfully to the stochastic environment that the representation of Q-value in this deep Q-networks for agents is stable and accurate and the agents act deterministically after 900 steps when the ε -greedy increased to the maximum.

The agents targeted at optimizing the solution of a textile ozone

process to approach the fabric color performance of 0.81, 15.76, -20.84, and -70.79 in regard to k/s, L^*, a^*, and b^*. During the DQN agents interacted in the Markov game with 5000 steps, the minimum errors of each agent and the total sum of the minimum errors of four agents reflected by RF models are collected and displayed in Fig. 7. The convergence of minimum error verifies the effectiveness and efficiency of the designed reward function in our MARL system, and it seems that the solution with lower error can be obtained potentially along with growing time steps.

The comparison of the constructed framework with baseline approaches about optimized results is depicted in Table 2. And the comparison of the digital result are exhibited in Fig. 8. It is illustrated that the optimal scenario derived from the proposed MARL framework can achieve the desired color in the textile ozonation process, in terms of the colorimetric values of k/s, L^*, a^*, and b^*, which are very close to the optimization objectives. The simulated color performance of scenarios obtained from different algorithms are exhibited in Fig. 9. It can be observed directly and clearly about the color treating effects of the textile ozonation process in different system models. It implies that the proposed reinforcement learning (MARL) system proposed performed dominated the baseline methods of MOPSO and NSGA- II in our case study for optimizing the ozonation process solution efficiently and achieve the desired color on treated fabrics. The relatively shorter computation time and higher performance of MARL system could be attributed to that multiple gents can work in parallel mode, and they share experience in the process. While on the other hand, the meta-heuristic algorithms of MOPSO and NSGA-2 have been reported that may fail to work with smaller datasets [77] and take an impracticably long time in iteration [78]. And more importantly, though they are effective to deal with low dimension multi-objective optimization problems, the increased stress of selection from the growing dimension in the problem would decline the effects dramatically when the objectives are more than three.

6. Conclusions and future work

In this work, we proposed a multi-agent reinforcement learning (MARL) methodology to cope with the increasingly complicated multi-objective optimization problems in the textile manufacturing process. The multi-objective optimization of textile process solutions is modeled as a stochastic Markov game and multiple intelligent agents based on deep Q-networks (DQN) are developed to achieve the correlated equilibrium optimal solutions of the optimizing process. The stochastic Markov game is neither fully cooperative nor fully competitive so that the agents employ a utilitarian selection mechanism that maximizes the sum of all agents' rewards (obeying the increasing ε-greedy policy) in each state to avoid the interruption of multiple equilibria. The case study results reflect that the proposed MARL system is possible to achieve the optimal solutions for the textile ozonation process and enzyme washing process, and it performs better than the traditional approaches.

It is worth mentioning that this system can also be applied in practice with different objective functions such as energy optimization, material optimization, etc. However, the case studies in this paper are not eligible enough to reflect the power of it in the big data environment. As known that the practice and effectiveness of RF and DQN rely strongly on big data and computation power which is quite limited in the application of the textile industry nowadays. While along with the growing application of artificial intelligent techniques in the textile manufacturing industry, such concerns could be properly addressed in the industry 4.0 era when it is able to take full advantage of the Internet of Things (IoT) environment. The future works, thus, could try to investigate more to see the realistic and practical effects of this developed system in the real industrial implementation.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

This research was supported by the funds from National Key R&D Program of China (Project NO: 2019YFB1706300), and Scientific Research Project of Hubei Provincial Department of Education, China (Project NO: Q20191707).

The first author would like to express his gratitude to China Scholarship Council for supporting this study (CSC, Project NO. 201708420166).

References

[1] Uddin F. Introductory chapter: textile manufacturing processes. Rijeka: IntechOpen; 2019. https://doi.org/10.5772/intechopen.87968.
[2] Fan J, Hunter L. A worsted fabric expert system: part II: an artificial neural network model for predicting the properties of worsted fabrics. Text Res J 1998;68:763–71.
[3] Hasanbeigi A, Price L. A review of energy use and energy efficiency technologies for the textile industry. Renew Sustain Energy Rev 2012;16:3648–65. https://doi.org/10.1016/j.rser.2012.03.029.
[4] Kumar V, Koehl L, Zeng X, Ekwall D. Coded yarn based tag for tracking textile supply chain. J Manuf Syst 2017;42:124–39. https://doi.org/10.1016/j.jmsy.2016.11.008.
[5] Kumar V, Koehl L, Zeng X. A fully yarn integrated tag for tracking the international textile supply chain. J Manuf Syst 2016;40:76–86. https://doi.org/10.1016/j.jmsy.2016.06.007.
[6] Kaplan S. A multicriteria decision aid approach on navel selection problem for rotor spinning abstract. Text Res J 2006;76:896–904. https://doi.org/10.1177/0040517507069122.
[7] Krishna R, Guduru R, Shaik SH, Yaramala S. A dynamic optimization model for multi- objective maintenance of sewing machine. Int J Pure Appl Math 2018;118:33–43.
[8] Majumdar A, Singh SP, Ghosh A. Modelling, optimization and decision making techniques in designing of functional clothing. Indian J Fibre Text Res 2011;36:398–409.
[9] Chakraborty S, Diyaley S. Multi-objective optimization of yarn characteristics using evolutionary algorithms : a comparative study. J Inst Eng Ser E 2018;99:129–40. https://doi.org/10.1007/s40034-018-0121-8.
[10] Sette S, Van Langenhove L. Optimising the fiber-to-yarn production process: finding a blend of fiber qualities to create an optimal price/quality yarn. Autex Res J 2002;2:57–64.
[11] Mukhopadhyay A, Midha VK, Ray NC. Multi-objective optimization of parametric combination of injected slub yarn for producing knitted and woven fabrics with least abrasive damage. Res J Text Appar 2017;21:111–33.
[12] Almetwally AA. Multi-objective optimization of woven fabric parameters using Taguchi – grey relational analysis multi-objective optimization of woven fabric parameters using Taguchi – grey relational analysis. J Nat Fibers 2019;0:1–11. https://doi.org/10.1080/15440478.2019.1579156.
[13] Deb K. Multi-objective optimization using evolutionary algorithms, vol. 16. John Wiley & Sons; 2001.
[14] Talbi E-G. Metaheuristics: from design to implementation, vol. 74. John Wiley & Sons; 2009.
[15] Kordoghli B, Jmali M, Saadallah S, Liouene N. Multi-objective scheduling of flow-shop problems in finishing factories using genetic algorithms. J Text Apparel Technol Manage 2010;6:1–10.
[16] Nurwaha D, Wang X. Optimization of electrospinning process using intelligent control systems. 2013. p. 593–600. https://doi.org/10.3233/IFS-2012-0578. 24.
[17] Wu CC, Chang NB. Global strategy for optimizing textile dyeing manufacturing process via GA-based grey nonlinear integer programming. Comput Chem Eng 2003;27:833–54. https://doi.org/10.1016/S0098-1354(02)00270-3.
[18] Das S, Ghosh A, Majumdar A, Banerjee D. Yarn engineering using hybrid artificial neural network-genetic algorithm model. 2013. p. 1220–6. https://doi.org/10.1007/s12221-013-1220-2. 14.
[19] Jeyaraj KL, Muralidharan C, Senthilvelan T. Genetic algorithm based multi-objective optimization of process parameters in color fast finish process - a textile case study. J Text Apparel Technol Manage 2013;8:1–26.
[20] Majumdar A, Mal P, Ghosh A, Banerjee D. Multi-objective optimization of air permeability and thermal conductivity of multi-objective optimization of air permeability and thermal conductivity of knitted fabrics with desired ultraviolet protection. J Text Inst 2017;108:110–6. https://doi.org/10.1080/00405000.2016.1159270.
[21] Barzoki PK, Vadood M, Johari MS. Multi-objective optimization of rotorcraft compact spinning core-spun yarn properties. J Text Polym 2018;6:47–53.
[22] Vadood M, Barzoki PK, Johari MS. Multi objective optimization of rotorcraft compact spinning system using fuzzy-genetic model. J Text Inst 2017;108:2166–72. https://doi.org/10.1080/00405000.2017.1316178.

[23] Chen J, Ding Y, Jin Y, Hao K. A synergetic immune clonal selection algorithm based multi-objective optimization method for carbon fiber drawing process. Fibers Polym 2013;14:1722–30. https://doi.org/10.1007/s12221-013-1722-y.
[24] Derringer G, Suich R. Simultaneous optimization of several response variables. J Qual Technol 1980;12:214–9.
[25] Arain FA, Tanwari A, Hussain T, Malik ZA. Multiple response optimization of rotor yarn for strength, unevenness, hairiness and imperfections. Fibers Polym 2012;13:118–22. https://doi.org/10.1007/s12221-012-0118-8.
[26] Saggiomo M, Kemper M, Gloy Y-S, Gries T. Weaving machine as cyber-physical production system:multi-objective self-optimization of the weaving process. 2016 IEEE Int Conf Ind Technol 2016:2084–9. https://doi.org/10.1109/ICIT.2016.7475090.
[27] Chakraborty S, Agarwal S, Dandge SS. Analysis of cotton fibre properties: a data mining approach. J Inst Eng Ser E 2018;99:163–76. https://doi.org/10.1007/s40034-018-0125-4.
[28] Nanduri V, Das TK. A reinforcement learning model to assess market power under auction-based energy pricing. IEEE Trans Power Syst 2007;22:85–95. https://doi.org/10.1109/TPWRS.2006.888977.
[29] Krasheninnikova E, García J, Maestre R, Fernández F. Engineering applications of artificial intelligence reinforcement learning for pricing strategy optimization in the insurance. Eng Appl Artif Intell 2019;80:8–19. https://doi.org/10.1016/j.engappai.2019.01.010.
[30] Chizhov YA, Borisov AN. Markov decision process in the problem of dynamic pricing policy. 2011. p. 361–71. https://doi.org/10.3103/S0146411611060058.45.
[31] Lu R, Hong SH, Zhang X. A dynamic pricing demand response algorithm for smart grid: reinforcement learning approach. Appl Energy 2018;220:220–30. https://doi.org/10.1016/j.apenergy.2018.03.072.
[32] Rana R, Oliveira FS. Expert systems with applications dynamic pricing policies for interdependent perishable products or services using reinforcement learning. Expert Syst Appl 2015;42:426–36. https://doi.org/10.1016/j.eswa.2014.07.007.
[33] Waschneck B, Altenm T, Bauernhansl T, Knapp A, Kyek A. Optimization of global production scheduling with deep reinforcement learning. 51st CIRP Conf Manuf Syst 2018:1264–9.
[34] Wei Y, Kudenko D, Liu S, Pan L, Wu L, Meng XA. Reinforcement learning based workflow application scheduling approach in dynamic cloud environment. Collab. 2017. Springer International Publishing; 2018. p. 120–31. https://doi.org/10.1007/978-3-030-00916-8.
[35] Baykasoğlu A, Madenoğlu FS, Hamzadayı A. Greedy randomized adaptive search for dynamic flexible job-shop scheduling. J Manuf Syst 2020;56:425–51. https://doi.org/10.1016/j.jmsy.2020.06.005.
[36] Leng J, Jin C, Vogl A, Liu H. Deep reinforcement learning for a color-batching resequencing problem. J Manuf Syst 2020;56:175–87. https://doi.org/10.1016/j.jmsy.2020.06.001.
[37] Krasheninnikova E, García J, Maestre R, Fernández F. Reinforcement learning for pricing strategy optimization in the insurance. Eng Appl Artif Intell 2019;80:8–19. https://doi.org/10.1016/j.engappai.2019.01.010.
[38] Zhou Z, Li X, Zare RN. Optimizing chemical reactions with deep reinforcement learning. ACS Cent Sci 2017;3:1337–44. https://doi.org/10.1021/acscentsci.7b00492.
[39] Rocchetta R, Bellani L, Compare M, Zio E, Patelli E. A reinforcement learning framework for optimal operation and maintenance of power grids. Appl Energy 2019;241:291–301. https://doi.org/10.1016/j.apenergy.2019.03.027.
[40] Kuznetsova E, Li Y, Ruiz C, Zio E, Ault G, Bell K. Reinforcement learning for microgrid energy management. Energy 2013;59:133–46. https://doi.org/10.1016/j.energy.2013.05.060.
[41] Mehdi S, Fard H, Hamzeh A, Hashemi S. Using reinforcement learning to find an optimal set of features. Comput Math Appl 2013;66:1892–904. https://doi.org/10.1016/j.camwa.2013.06.031.
[42] Jasmin EA, College GE. Reinforcement learning solution for unit commitment problem through pursuit method. 2009 Int. Conf. Adv. Comput. Control. Telecommun. Technol. 2009:324–7. https://doi.org/10.1109/ACT.2009.87.
[43] Annamdas KK, Rao SS. Multi-objective optimization of engineering systems using game theory and particle swarm optimization. Eng Optim 2009;41:737–52.
[44] Jin M, Lei X, Du J. Evolutionary game theory in multi-objective optimization problem. Int J Comput Intell Syst 2010;3:74–87.
[45] Akopov AS, Hevencev MA. A multi-agent genetic algorithm for multi-objective optimization. 2013 IEEE Int. Conf. Syst. Man, Cybern., IEEE 2013:1391–5.
[46] Zhang H, Zeng Y, Jin X, Shu B, Zhou Y, Yang X. Simulating multi-objective land use optimization allocation using multi-agent system—a case study in Changsha, China. Ecol Modell 2016;320:334–47.
[47] Busoniu L, Cluj-napoca UT, Babuska R, De Schutter B. Multi-agent reinforcement learning : an overview. Innov. Multi-agent syst. Appl. Berlin Heidelberg: Springer; 2010. p. 183–221. https://doi.org/10.1007/978-3-642-14435-6.
[48] Mannion P, Mason K, Devlin S, Duggan J. Multi-objective dynamic dispatch optimisation using multi-agent reinforcement learning. Proc. 15th Int. Conf. Auton. Agents Multiagent Syst 2016:1–2 (AAMAS 2016).
[49] Khamis MA, Member S, Gomaa W. Enhanced multiagent multi-objective reinforcement learning for urban traffic light control. 11th Int. Conf. Mach. Learn. Appl. 2012 11th Int. Conf. Mach. Learn. Appl. Enhanc. 2012:586–91. https://doi.org/10.1109/ICMLA.2012.108.
[50] Kim YG, Lee S, Son J, Bae H, Do Chung B. Multi-agent system and reinforcement learning approach for distributed intelligence in a flexible smart manufacturing system. J Manuf Syst 2020;57:440–50. https://doi.org/10.1016/j.jmsy.2020.11.004.
[51] Hu J, Wellman MP. Multiagent reinforcement learning: theoretical framework and an algorithm, vol. 98. ICML; 1998. p. 242–50. Citeseer.
[52] Lin L-J. Reinforcement learning for robots using neural networks. Carnegie-Mellon Univ Pittsburgh PA School of Computer Science; 1993.
[53] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. ArXiv Prepr ArXiv13125602 2013.
[54] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature 2015;518:529–33.
[55] Gupta JK, Egorov M, Kochenderfer M. Cooperative multi-agent control using deep reinforcement learning. Int. Conf. Auton. Agents Multiagent Syst. Springer; 2017. p. 66–83.
[56] Tampuu A, Matiisen T, Kodelja D, Kuzovkin I, Korjus K, Aru J, et al. Multiagent cooperation and competition with deep reinforcement learning. PLoS One 2017;12:e0172395.
[57] Hernandez-Leal P, Kartal B, Taylor ME. Is multiagent deep reinforcement learning the answer or the question? A brief survey. Learning 2018;21:22.
[58] Wang Y, Liu H, Zheng W, Xia Y, Li Y, Chen P, et al. Multi-objective workflow scheduling with reinforcement learning. IEEE Access 2019;7:39974–82. https://doi.org/10.1109/ACCESS.2019.2902846.
[59] Zhang H, Ding Y, Zhang W, Feng S, Zhu Y, Yu Y, et al. CityFlow: a multi-agent reinforcement learning environment for large scale city traffic scenario. 2019 IW3C2 (International World Wide Web Conf. Committee) 2019:3620–4. https://doi.org/10.1145/3308558.3314139.
[60] Sutton RS, Barto AG. Introduction to reinforcement learning, vol. 135. Cambridge: MIT press; 1998.
[61] Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18–22. https://doi.org/10.1177/154405910408300516.
[62] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[63] He Z, K p Tran, Zeng X, Xu J, Yi C. Modeling color fading ozonation of reactive-dyed cotton using the extreme learning machine, support vector regression and random forest. Text Res J 2020;90:896–908. https://doi.org/10.1177/0040517519883059.
[64] Rahman R, Otridge J, Pal R. Gene expression IntegratedMRF : random forest-based framework for integrating prediction from different data types. 2017. p. 1407–10. https://doi.org/10.1093/bioinformatics/btw765. 33.
[65] Tsitsiklis JN, Van Roy B. An analysis of temporal-difference learning with function approximation. IEEE Trans Automat Contr 1997;42:674–90. https://doi.org/10.1109/9.580874.
[66] Greenwald A, Hall K, Zinkevich M. Correlated-Q learning. J Mach Learn Res 2007;1:1–30.
[67] Xu J, He Z, Li S, Ke W. Production cost optimization of enzyme washing for indigo dyed cotton denim by combining Kriging surrogate with differential evolution algorithm. Text Res J 2020;90:1860–71. https://doi.org/10.1177/0040517520904352.
[68] He Z, Li M, Zuo D, Yi C. Color fading of reactive-dyed cotton using UV-assisted ozonation. Ozone Sci Eng 2019;41:60–8. https://doi.org/10.1080/01919512.2018.1483817.
[69] He Z, Li M, Zuo D, Yi C. The effect of denim color fading ozonation on yarns. Ozone Sci Eng 2018;40:377–84. https://doi.org/10.1080/01919512.2018.1435259.
[70] He Z, Li M, Zuo D, Xu J, Yi C. Effects of color fading ozonation on the color yield of reactive-dyed cotton. Dyes Pigm 2019;164:417–27. https://doi.org/10.1016/j.dyepig.2019.01.006.
[71] He Z, Tran KP, Thomassey S, Zeng X, Yi C. A deep reinforcement learning based multi-criteria decision support system for optimizing textile chemical process. Comput Ind 2021;125:103373.
[72] Li M, He Z, Xu J. A comparative study of ozonation on aqueous reactive dyes and reactive-dyed cotton. Color Technol 2021:1–13. https://doi.org/10.1111/cote.12534.
[73] Cie C. In: Commission Internationale de l'eclairage Proceedings; 1931.
[74] Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, et al. TensorFlow : large-scale machine learning on heterogeneous distributed systems. 2015.
[75] Poli R, Kennedy J, Blackwell T. Particle swarm optimization. Icnn95-international Conf. Neural Networks 2002.
[76] Konak A, Coit DW, Smith AE. Multi-objective optimization using genetic algorithms: a tutorial. Reliab Eng Syst Saf 2006;91:992–1007.
[77] Giri C, Jain S, Zeng X, Bruniaux P. A detailed review of artificial intelligence applied in the fashion and apparel industry. IEEE Access 2019;7:95376–96. https://doi.org/10.1109/ACCESS.2019.2928979.
[78] Zimmerling C, Poppe C, Kärger L, Zimmerling C, Poppe C, Kärger L. Estimating optimum process parameters in textile draping of variable part geometries - a reinforcement learning approach. Procedia Manuf 2020;47:847–54. https://doi.org/10.1016/j.promfg.2020.04.263.