

A Reinforcement Learning Based Approach for Welding Sequence Optimization

Jesus Romero-Hdz, Baidya Saha, Gengis Toledo-Ramirez
and Ismael Lopez-Juarez

Abstract We develop and implement a Q-learning based Reinforcement Learning (RL) algorithm for Welding Sequence Optimization (WSO) where structural deformation is used to compute reward function. We utilize a thermomechanical Finite Element Analysis (FEA) method to predict deformation. We run welding simulation experiment using well-known Simufact[®] software on a typical widely used mounting bracket which contains eight welding beads. RL based welding optimization technique allows the reduction of structural deformation up to $\sim 66\%$. RL based approach substantially speeds up the computational time over exhaustive search.

Keywords Reinforcement learning (RL) · Welding sequence optimization
Structural deformation · Finite element analysis (FEA) · Simufact software

1 Introduction

Welding is the most common fabrication process typically used for joining metals [1]. It is widely used in various industries such as automotive, shipbuilding, aerospace, construction, gas and oil trucking, nuclear, pressure vessels, and heavy and earth-moving equipment [2, 3]. Structural deformation of welded structures is a natural outcome of internal stresses produced while welding due to intrinsic nonuniform heating and cooling of the joint. Nevertheless, welding deformation plays a negative role in the process having high impacts in several ways, such as

J. Romero-Hdz (✉) · G. Toledo-Ramirez
Centro de Ingeniería Y Desarrollo Industrial (CIDESI), Santiago de Querétaro, Mexico
e-mail: jaromero@cidesi.edu.mx

B. Saha
Centro de Investigación En Matemáticas (CIMAT), Guanajuato, Mexico
e-mail: baidya.saha@cimat.mx

I. Lopez-Juarez
Centro de Investigación Y de Estudios Avanzados (CINVESTAV),
Mexico City, Mexico

constraints in the design phase, reworks, quality cost, and overall capital expenditure. Selection of welding sequence could substantially reduce structural deformation. The common industrial practice is to select a sequence by experience using a simplified design of experiments which does not guarantee an optimal sequence [4]. Welding deformation can be numerically computed through Finite Element Analysis (FEA) using thermomechanical models. FEA can provide reasonable solutions for various welding conditions and geometric configurations. However, it can be computationally very expensive and time consuming.

The optimal welding sequence can be achieved by using a full factorial design procedure. The total number of welding configurations (N) is counted by, $N = n^r \times r!$, where n and r are the number of welding directions and seams (beads) respectively. This is an NP -hard problem and these possible configurations grow exponentially with the number of welding beads. As an example, the mounting bracket used in this study has eight weld seams that can be welded in two directions; hence the number of welding configurations for exhaustive search is 10,321,920. In real-life application, a complex weldment like an aero-engine assembly might have between 52 and 64 weld segments [5]. Therefore, full factorial design is practically infeasible even for simulation experiment using FEA.

Here, we develop and implement a Q-Learning based RL algorithm [6] for WSO. The technical contributions of this paper are as follows. First, a deformation based reinforcement learning significantly reduces the computational complexity over extensive search. In this experiment, we achieve the optimal solution through RL after executing 2 iterations. Second, we exploit a novel reward function for RL consisting of the inverse of the maximum structural deformation for WSO. Third, we compare our RL algorithm with both single objective [7] and multi-objective [8] Genetic Algorithm (GA) and we demonstrate that RL finds a pseudo optimal solution which is much faster than GA, in both cases.

We conduct simulation experiments for Gas Metal Arc Welding (GMAW) process through the simulation software Simufact[®]. The scope of this study is limited to GMAW process. The average execution time for each simulation (welding configuration) is 30 min using a workstation with two Intel Xeon[®] @2.40 GHz processors, 48 GB of RAM and 4 GB of dedicated video memory. We conduct our experiment on a mounting bracket which is usually used in telescopic jib [9] and automotive industries [10, 11].

The organization of this paper is as follows. Section 2 presents literature review. Section 3 demonstrates reinforcement learning based welding sequence optimization method. Section 4 illustrates experimental results and discussions. In Sect. 5 conclusions and future directions of this work are presented. Relevant references are listed at the end of the paper.

2 Literature Review

We organize the literature review into two parts. In the first part we present a brief literature review on WSO using artificial intelligence techniques and then we briefly introduce the Q-learning based RL algorithm related to WSO.

2.1 Welding Sequence Optimization

Several optimization methods are available in literature for welding sequence optimization. Among them, GA is one of the most popular methods available in the literature for welding sequence optimization. Mohammed et al. [12] presented GA based WSO where the distortion data computed by FEA has been used as a fitness function. Kadivar et al. [13] also presented GA based solution for WSO where only the distortion is used in the objective function and they completely neglect the effect of the welding sequence on the maximum residual stress. Damsbo and Ruboff [14] incorporated domain specific knowledge into a hybrid GA for welding sequence optimization. They minimized the robot path length to minimize the operation time but neglected the effect of welding sequence on deformation and residual stress. Islam et al. [15] coupled FEA approach with GA where the maximum structural deformation was used in the fitness function and other design variables such as welding direction and upper and lower bounds of welding process parameters were taken into effect in the model. Warmefjord et al. [16] discussed several alternative approaches to GA in spot welding sequence selection method where they suggested general simple guidelines, minimize variation in each step, sensitivity, and relative sensitivity. Kim et al. [17] proposed two types of heuristic algorithms called construction algorithm and an improvement algorithm where heuristics for the traveling salesman problem are tailored to the welding sequence optimization. However, they did not consider the inherent heat-caused deformation with the aim of minimizing the time required to process the task.

Romero-Hdz et al. [18] presented a literature overview of the artificial intelligence techniques used in WSO. The AI techniques reported include GA, Graph Search, Artificial Neural Networks (ANN) and Particle Swarm Optimization (PSO). Other popular methods are also described such as Joint Rigidity Method, Surrogate Models and the use of generalized guidelines. Some of the limitations of these studies are the lack of experimental validation in real components and ignoring some factors like residual stress and temperature which are important factors for the resulting welded quality.

Okumoto [19] presented an implementation of the Q-learning based RL algorithm to optimize the welding route of an automatic machine. The machine uses a simple truck system and it moves only in one direction until a force is detected. The fitness function in this work is based on the time, because, this type of machines are moved from one joint to another manually by the welder. As the amount of welding

seams increase, the total number of possible execution combinations grows exponentially. So, a bad decision can increase the labor hours that impacts on the cost and lead time. They used ϵ -greedy based selection method. This method stochastically adopts a lesser reward to avoid the local optima.

In this paper, we proposed a Q-Learning based Reinforcement learning algorithm where we use maximum structural deformation as the reward function that are described in the next section.

2.2 Reinforcement Learning

RL is a branch of machine learning which has been extensively used in different fields such as gaming [20], neuroscience [21], psychology [22], economics [23], engineering communications [24], engineering power systems [25], and robotics [26]. Some of the algorithms are inspired by stochastic dynamic programming like Q-learning algorithm which is the base of the proposed algorithm in this paper. RL techniques learn directly from empirical experiences of the environment. RL can be subdivided into two fundamental problems: learning and planning. While learning, the agent interacts with the real environment, by executing actions and observing their consequences. On the other hand, while planning the agent can interact with a model of the environment: by simulating actions and observing their consequences.

Figure 1 shows the basic framework of the RL algorithm. An RL task requires decisions to be made over many time steps. We assume that an agent exists in a world or environment, E , which can be summarized by a set of states, S . First, the agent receives observations from the environment E , then the agent solves the exploration-exploitation dilemma, whether to explore and get new information or

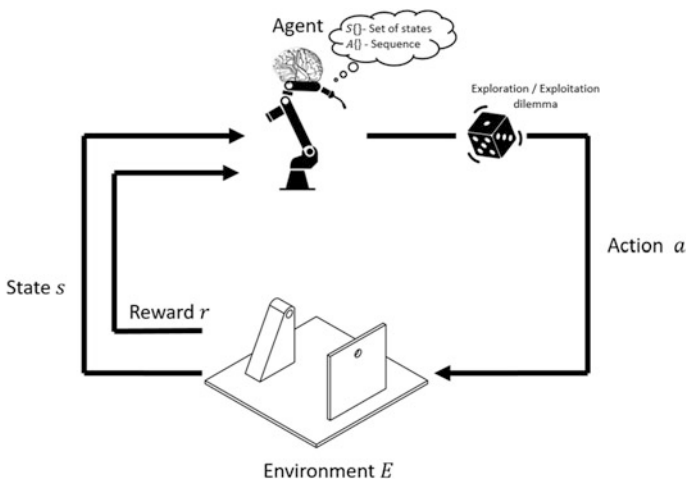


Fig. 1 Basic framework of RL

act with the information it knows and trust. Then selects an action a , and then it receives an immediate reward r and then go to next state. The agent again acts repeatedly so as to receive maximum reward from an initial state to a goal state. It is an optimization technique that can manage with moderate dynamic change in the environment through learning by repeating and evaluating the action.

In this paper, we applied Q-learning algorithm, one of the algorithms of RL for WSO. Q-learning algorithm estimates the value function $Q(s, a)$ which is obtained by the agent's repeating of the action by trial and error, for the environment. $Q(s, a)$ expresses the expectation of a gain when the agent takes the most suitable action after having chosen an action in a state s . The most suitable action is defined as the action a in state s , for which the value of $Q(s, a)$ becomes greatest among all the actions permissible in state s . The Q value is updated by the following Eq. (1) [19]:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \{ r(s, a) + \gamma \max_{a' \in A(s)} Q(s', a') - Q(s, a) \} \quad (1)$$

where $Q(s, a)$ is the value of action a in state s , $Q(s', a')$ is the value of action a' at state s' after transition, α is the learning rate ($0 < \alpha < 1$), and γ is the discount rate ($0 < \gamma < 1$). A number of selection methods have been proposed to solve the exploration-exploitation dilemma and choose one action executed among the many possible actions that exist. We used ε -greedy method in this study. This method stochastically adopts a lesser reward to avoid the local minima. The ε -greedy selects an action a in state s for which the value of $Q(s, a)$ is maximum at probability $(1-\varepsilon)$, $0 < \varepsilon < 1$, as illustrated in the equation:

$$p(s, a) = \begin{cases} 1 - \varepsilon, & \text{when } Q(s, a) \text{ is maximum} \\ \varepsilon, & \text{elsewhere} \end{cases} \quad (2)$$

3 Reinforcement Learning Based Welding Sequence Optimization

Figure 2 shows the flowchart and pseudo-code of the Q-learning based RL method on WSO respectively. We solve the exploration-exploitation dilemma by generating a random number between 0 and 1 and if it is less than or equal to 0.2 (the value of ε is taken as 0.2), exploration is executed, on the other hand, exploitation will be performed. For exploration, we choose the second best weld seam and for exploitation we choose the weld seam with a particular welding direction that gives the minimum of the maximum structural deformation. In this WSO experiment, the agent is considered as the robot or human, the actions of the agent are the weld seams that can be placed into the workpiece along with the direction of the welding, the state is defined as the set of actions already executed. The reward is defined as the inverse of the maximum structural deformation. The most suitable action is defined as the welding weld seam along a particular direction that provides minimum of the maximum structural deformation (Fig. 3).

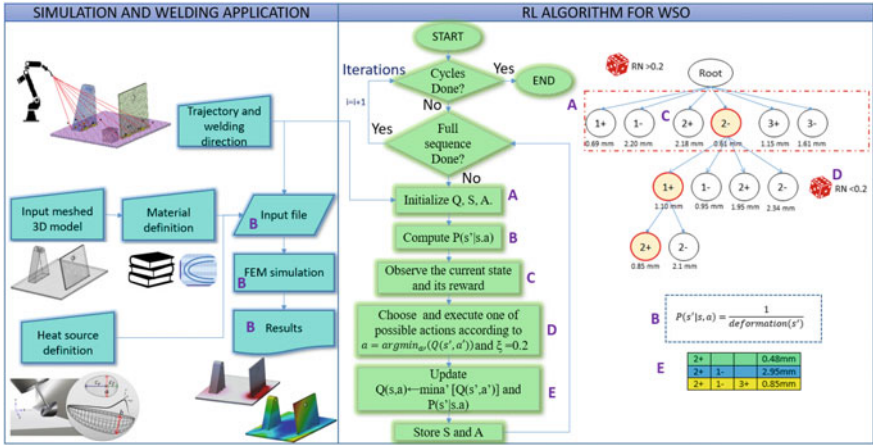


Fig. 2 Flowchart of Q-learning based reinforcement learning method on WSO

```

Initialize  $Q(s, a) = 0; S = \{\}, A = \{\}, s = \{\}, A' = \{1+, 1-, 2+, 2-, \dots, 8+, 8-\};$ 
Initialize a state-transition model  $P(s'|s, a) \quad P(s'|s, a) = 1/$ 
  deformation( $s'$ );
Do until learning end /*finish all the cycles or iterations*/
Do until we reach a goal /*all the seams are welded */
Observe a current state  $s$ ,
Choose an action  $a$  from the set consisting of available welding
  seams along
    all possible directions which are not welded yet
  which provides
    minimum of the maximum structural deformation;
 $a = \text{argmin}_{a' \in A'} [Q(s', a')]$  with  $\epsilon$ -greedy policy;  $[Q(s', a') =$ 
  minimum of the maximum structural deformation,  $\epsilon = 0.2]$ 
Execute action  $a$ ;
Update Q-value with  $Q(s, a) \leftarrow \min_{a'} [Q(s', a')]$  and remove  $a$  and -
   $a$  from  $A'$ ;
Update state-transition model  $P(s'|s, a);$ 
Store  $S = \{s\}$  and  $A = \{A, a\};$ 
End
End
  
```

S, A: states and actions consisting of stochastic shortest paths.

Fig. 3 Pseudo-code of Q-learning based RL method on WSO

4 Experimental Results and Discussions

This section is organized as follows. First we describe the study case. Second, we mention the values of the parameters used in this study. Third, we illustrate the results of the FEA for the best and worst sequence found by the proposed RL method. Fourth, we demonstrate the effects of welding sequence on WSO. Finally, we show a comparative study among single objective GA [7], multi-objective GA [8] and Q-learning based RL method.

4.1 Study Case

Figure 4 illustrates some sample geometries of the mounting brackets available in the market. These geometries are typically used in heavy equipment, vehicles, and ships. Figure 5 demonstrates a mounting bracket which we chose as a study case in our experiment as well as the engineering drawing with all specifications.



Fig. 4 Mounting brackets available in the market as an example of welded parts

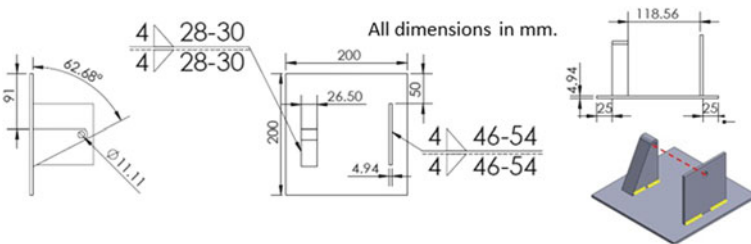


Fig. 5 Engineering drawing of the mounting bracket with 8 weld beads

4.2 Parameters Used

Table 1 illustrates values of the parameters of the Q-learning based RL algorithm used in the simulation experiment. We conduct the RL algorithm up to two iterations to find the pseudo optimal solution. Moreover, Table 2 shows the welding simulation and real experiment parameters.

4.3 Discussions About the FEA Results

The best sequence found was $[-5, -8, +6, -7, -1, +3, +2, -4]$ with maximum structural deformation 0.93 mm. On the other hand, the worst sequence found is $[-3, +4, -7, +6, +5, -1, +8, -2]$ and the corresponding maximum structural deformation is 2.76 mm as shown in Fig. 6. We conduct the RL experiment for two iterations. We select ϵ -greedy algorithm strategy where we choose the value of ϵ as 0.2. This indicates that the RL allows exploration 20% time and exploitation 80% time of the actions chosen. Since the value of our reward function at any stage in terms of maximum structural deformation cannot be computed by summing the value of the reward function of the previous stages, the value of the learning rate and discount factor are inapplicable in our application.

Table 1 RL parameters used in this experiment

Parameter	Value
Epochs (number of iterations)	2
Exploration rate	0.2
Exploitation rate	0.8
Learning rate	–
Discount factor	–
Selection method	ϵ -greedy method

Table 2 Welding simulation and real experiment parameters

Parameter	Value
Material	ASTM A36
Robot speed	50 cm/min
Current	234 A
Voltage	20.5 V

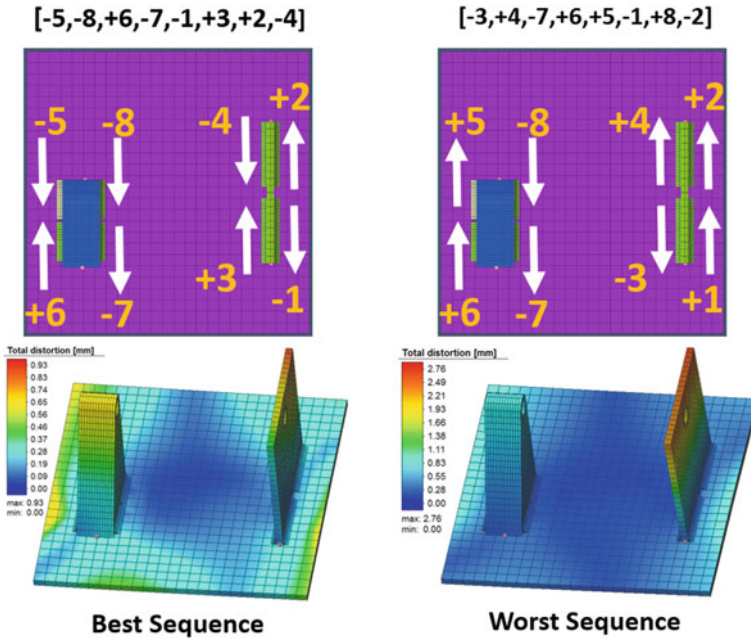


Fig. 6 Comparison between best (left column) and worst (right column) sequence and their deformation patterns

4.4 Effects of Welding Sequence on Welding Process Optimization

Normalized frequency of the structural deformation and effective stress values on the mounting bracket used in this study are shown in Fig. 7. If we consider the deformation value of the worst sequence as 100%, RL algorithm reduces about 66% maximum structural deformation over worst sequence (maximum structural deformation of best and worst sequence are 0.93 and 2.76 mm respectively). These results clearly demonstrate that welding sequence has significant effect on welding deformation. However, welding sequence has less effect on effective stress. These results are consistent with the results reported in the literature [7, 8].

4.5 Comparative Analysis: Reinforcement Learning Versus Genetic Algorithm

We compare the proposed RL method with the GA methods widely used in the WSO. In order to due this comparison, we have used the same parameters for GA reported in our previous work [7]. The structural deformation of the mounting

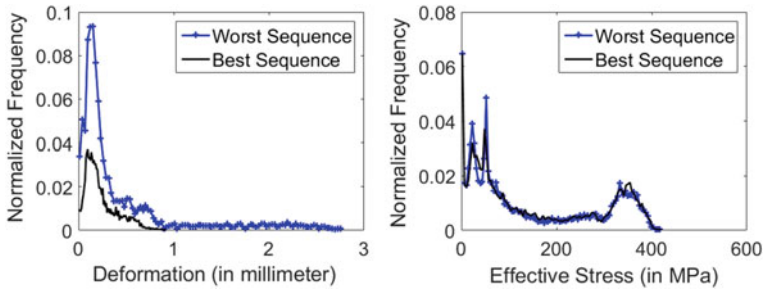


Fig. 7 Normal distribution of deformation and effective stress for best and worst sequence

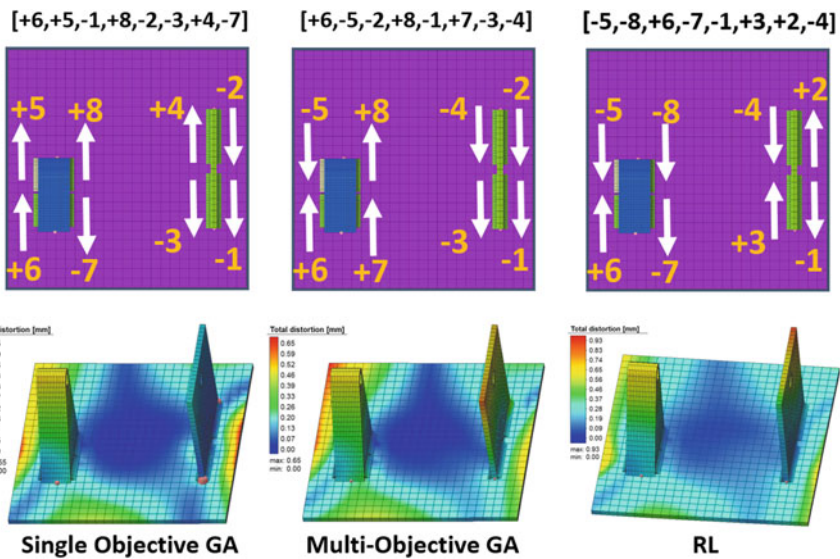


Fig. 8 From top to bottom: single objective GA, Multi-objective GA and RL algorithm

bracket for the best sequence found by the single objective, multi-objective and RL are demonstrated in Fig. 8. Table 3 illustrates the comparative analysis among single objective [7] and multi-objective GA [8] and RL method. Table 3 shows that though RL yields a bit more structural deformation, however, RL converges much faster than GA. RL finds a pseudo optimal solution in only two iterations whereas single objective and multi-objective GA require 115 and 81 simulations respectively. Since the average time required for each welding simulation experiment needs 30 min, RL method for our study case takes 30 h whereas single objective and multi-objective GA method take 57.5 and 40.5 h respectively to converge the algorithm. Figure 9 shows the search space (tree) for the best sequence explored by the RL algorithm.

Table 3 Comparative analysis among single objective genetic algorithm, multi-objective genetic algorithm and reinforcement learning

Method	Best sequence	Maximum structural deformation (mm)	# of sequences executed	Time (h)
Single objective GA	[+6, +5, -1, +8, -2, -3, +4, -7]	0.55	115	57.5
Multi-objective GA	[+ 6, -5, -2, +8, -1, +7, -3, -4]	0.66	81	40.5
Proposed RL	[-5, -8, +6, -7, -1, +3, +2, -4]	0.93	2	30

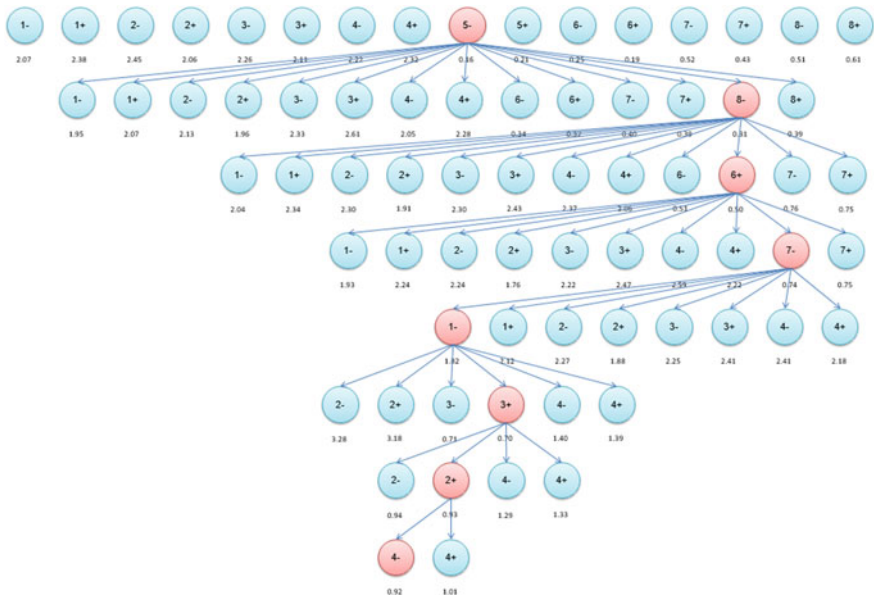


Fig. 9 Search space for the best sequence explored by Q-learning based RL algorithm

5 Conclusions and Future Works

Structural deformation plays an important role to measure the quality of the welded structures. Optimization of the welding sequence reduces the deformation of the welded structures. In this paper, the maximum structural deformation is exploited as the reward function of a proposed Q-learning based RL algorithm for WSO. RL is used to reduce significantly the search space of the exhaustive search. Structural deformation is computed using FEA. We conduct a simulation experiment on a mounting bracket which is typically used in vehicles and other applications. We

compare our RL based welding sequence optimization method with widely used single objective and multi-objective GA. Proposed RL based WSO method significantly reduces the search space over GA and thus RL finds the pseudo optimal welding sequence much faster than GA by slightly compromising the welding quality. Welding quality could be enhanced by incrementing the number of iterations of the RL method.

This work opens up different avenues for WSO research. In the near future, we would like to develop a more robust multivariate reward function including structural deformation, residual stress, temperature, and robot path time for welding sequence optimization. Information of the deformation, residual stress and temperature after welding each seam in a sequence needs to be investigated for achieving better reduction of welding deformation and residual stress.

Acknowledgements The authors gratefully acknowledge the support provided by CONACYT (The National Council of Science and Technology) and CIDESI (Center for Engineering and Industrial Development) as well as their personnel that helped to realize this work and the Basic Science Project (254801) supported by CONACYT, Mexico.

References

1. Goldak JA, Akhlaghi M (2005) Computational welding mechanics. <http://www.worldcat.org/isbn/9780387232874>. Accessed 21 May 2016
2. Masubuchi K (1980) Analysis of welded structures. Pergamon Press Ltd., Oxford
3. Islam M, Buijk A, Rais-Rohani M et al (2014) Simulation-based numerical optimization of arc welding process for reduced distortion in welded structures. *Finite Elem Anal Des* 84: 54–64
4. Kumar DA, Biswas P, Mandal NR et al (2011) A study on the effect of welding sequence in fabrication of large stiffened plate panels. *J Mar Sci Appl* 10:429–436
5. Jackson K, Darlington R (2011) Advanced engineering methods for assessing welding distortion in aero-engine assemblies. *IOP Conf Ser Mater Sci Eng* 26:12018
6. Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. MIT press, Cambridge
7. Romero-Hdz J et al (2016) An elitism based genetic algorithm for welding sequence optimization to reduce deformation. *Res Comput Sci* 121:17–36
8. Romero-Hdz J et al (2016) Deformation and residual stress based multi-objective Genetic Algorithm (GA) for welding sequence optimization. In: *Proceedings of the Mexican international conference on artificial intelligence, Cancun, Mexico, vol 1*, pp 233–243
9. Derlukiewicz D, Przybyłek G (2008) Chosen aspects of FEM strength analysis of telescopic jib mounted on mobile platform. *Autom Constr* 17:278–283
10. Subbiah S, Singh OP, Mohan SK et al (2011) Effect of muffler mounting bracket designs on durability. *Eng Fail Anal* 18:1094–1107
11. Romeo ŞI et al (2015) Study of the dynamic behavior of a car body for mounting the rear axle. *Proc Eur Automot Congr* 25:782
12. Mohammed MB, Sun W, Hyde TH (2012) Welding sequence optimization of plasma arc for welded thin structures. *WIT Trans Built Environ* 125:231–242
13. Kadivar MH, Jafarpur K, Baradaran GH (2000) Optimizing welding sequence with genetic algorithm. *Comput Mech* 26:514–519
14. Damsbo M, Ruboff PT (1998) An evolutionary algorithm for welding task sequence ordering. *Proc Artif Intell Symbolic Comput* 1476:120–131

15. Islam M, Buijk A, Rais-Rohani M, Motoyama K (2014) Simulation-based numerical optimization of arc welding process for reduced distortion in welded structures. *Finite Elem Anal Des* 84:54–64
16. Warnefjord K, Soderberg R, Lindkvist L (2010) Strategies for optimization of spot welding sequence with respect to geometrical variation in sheet metal assemblies. *Design and Manufacturing* 3: 569–57
17. Kim HJ, Kim YD, Lee DH (2004) Scheduling for an arc-welding robot considering heat-caused distortion. *J Oper Res Soc* 56(1):39–50
18. Romero-Hdz J et al (2016) Welding sequence optimization using artificial intelligence techniques: an overview. *Int J Comput Sci Eng* 3(11):90–95
19. Okumoto Y (2008) Optimization of welding route by automatic machine using reinforcement learning method. *J Ship Prod* 24:135–138
20. Tesauro G (1994) TD-gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comput* 6:215–219
21. Schultz W, Dayan P, Montague P (1997) A neural substrate of prediction and reward. *Science* 16:1936–1947
22. Sutton R (1990) Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In: *Proceedings of 7th international conference on machine learning*, vol 1, pp 216–224
23. Choi J, Laibson D, Madrian B et al (2007) Reinforcement learning and savings behavior. Yale Technical Report ICF Working Paper, 09–01
24. Singh S, Bertsekas D (1997) Reinforcement learning for dynamic channel allocation in cellular telephone systems. *Adv Neural Inf Process Syst* 9:974–982
25. Ernst D, Glavic M, Geurts P et al (2005) Approximate value iteration in the reinforcement learning context-application to electrical power system control. *Int J Emerg Electr Power Syst* 3(1):1066
26. Abbeel P, Coates A, Quigley M et al (2007) An application of reinforcement learning to aerobatic helicopter flight. *Adv Neural Inf Process Syst* 19:1–8