2013 13th Iranian Conference on Fuzzy Systems (IFSC)

Gait Analysis of a Six-Legged Walking Robot using Fuzzy Reward Reinforcement Learning

Mohammadali Shahriari
School of Science and Engineering
Sharif University of Technology, Int. Campus
Tehran, Iran
Corresponding author, Shahriari@kish.sharif.edu

Amir A. Khayyat
School of Science and Engineering
Sharif University of Technology, Int. Campus
Kish, Iran
Khayyat@sharif.edu

***Abstract*—Free gait becomes necessary in walking robots when they come to walk over discontinuous terrain or face some difficulties in walking. A basic gait generation strategy is presented here using reinforcement learning and fuzzy reward approach. A six-legged (hexapod) robot is implemented using Q-learning algorithm. The learning ability of walking in a hexapod robot is explored considering only the ability of moving its legs and using a fuzzy rewarding system telling whether and how it is moving forward. Results show that the hexapod robot learns to walk using the presented approach properly.**

***Index Terms*—Fuzzy systems, gait analysis, hexapod, reinforcement learning.**

I. Introduction

Control of legged robots is difficult, requiring fairly heavy on-line computations to be performed in real time. Hence a machine-learning solution is needed [1]. One machine-learning method for legged robots, which has great potential, is Reinforcement Learning (RL). RL is a promising approach to achieve the control of complex robots in dynamic environments; Josep M. Porta developed a Robotic Oriented Reinforcement Learning [2]. This approach helps with the basic learning platform for walking. A simple RL approach is used to develop walking gaits for hexapod [3][4]. Matt R. Bunting has implemented Q-learning [5] (a form of RL) on a hexapod [6] and has shown the capability of this algorithm for walking control. One of the strengths of Q-learning is that it is able to compare the expected utility of the available actions without requiring a model of the environment. In RL techniques one of the challenges is how to reward the actions in different states. There are different approaches and researches for rewarding regarding different systems and purposes. Accuracy and computation time are two parameters which should considered in this content. A fuzzy system is used in this paper for giving the proper reward signal to the robot in walking learning problem [7]. In the first section the reinforcement learning structure and fuzzy reward for walking is discussed. The next section studies the learning algorithm of the robot and in the final section the results and discussion of the presented formulation is analyzed.

II. Reinforcement Learning for Hexapod Walking

RL techniques are interesting subjects in both control theory and cognitive sciences. In control theory, building a system that works completely perfect is quite difficult, and it is an exhaustive procedure when unexpected errors or disturbances affect the system. Building a system that learns how to accomplish a task on its own, there becomes no need to calculate and predict complex control algorithms. In cognitive sciences, the ability to learn is a core component of cognition. RL algorithm is one such simple learning algorithm. This section explores the ability of a robotic hexapod agent to learn how to walk, using only the ability to move its legs and tell whether if the robot is moving forward. Therefore, the hexapod may be seen as an analog for a biological subject lacking all but the basic instincts observed in infants and having no external support or parental figure to learn from.

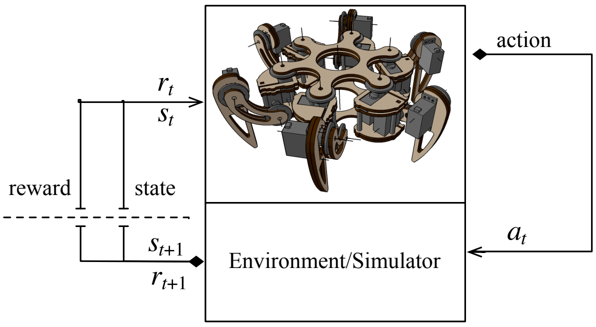


Fig. 1: The diagram of reinforcement learning procedure. Taken actions by agent lead to certain states and rewards. The goal is to find the best policy which takes actions that maximize total pay off in walking.

A. Reinforcement Learning Problem Architecture

In supervised learning the targets i.e. right answers are given to the agent as the training set. In some control problems it is difficult or not feasible to define the exact correct supervision. For example in hexapod gait analysis it is hard to define explicit supervision that the algorithm of learning is trying to mimic. In RL instead of telling the exact supervision only a reward function shows that whether is the agent doing well or not. Therefore it will be learning algorithms job to find best actions which lead to better rewards. There will be no need to define input/output sets. This kind of learning

978-1-4799-1228-5/13/$31.00 ©2013 IEEE.

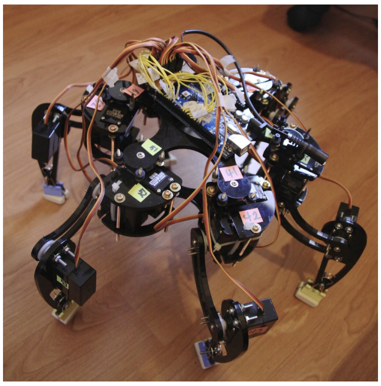


Fig. 2: Experimental 18 degrees of freedom SiWaReL prototype.

focuses on online performance and the interaction between exploration and exploitation. The trade of between these two has been one of the most challenges which are studied in recent years [8]. It is also shown that reinforcement learning has different successful applications in autonomous systems and legged robot locomotion [9].

Q-learning is a simple approach of RL which is chosen for walking learning in this paper. The robot explores the the state action domain and gets reward in taking actions. Future actions are taken based specified rewards in a way which maximizes the coming reward in the present and future states and actions. The schematic of actions, states and rewards can be seen in figure 1.

It is desired to learn the time sequencing of hexapod gait. The legs are moving in possible states with different actions. At the end of learning procedure, the optimal policy should be found satisfying the goal which is walking on a straight line minimizing tilt and other undesired translations.

B. State and Actions specification in SiWaReL

Learning problem for six-legged walking robot requires defining some discrete state and actions because continuous or huge state and action space means high computational costs. In this paper only three states are considered for each leg which means 729 states for the robot. The first state is when the leg is not on the ground and lifted. The other two states are when the leg is on front and back of the body [5]. 1 shows the specified states set of the hexapod robot.

$$S = \left\{ \sum_{i=1}^{6} s_{j_i} 3^{(6-i)} | j = 1, 2..729 \right\} \tag{1}$$

where s_{ji} stands for the jth state of ith leg and S is the set of all possible states. For the goal of learning to walk, actions are defined as going to new state i.e. leg position. Therefore each action can be specified by moving between the states. As in each state the robot can move to any other state the actions space have the same dimension of states.

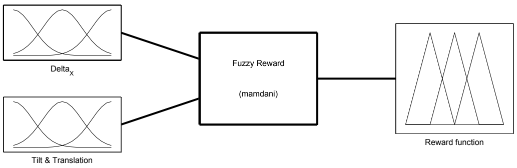


Fig. 3: The diagram of fuzzy reward.

$$(s, a) \quad \in \quad S \times A \tag{2}$$

where s, a and A denote state, action and set of actions respectively.

The learning procedure is a time consuming task. Also high number of iterations are required for the robot to explore possible state action set, update the rewards and find a desirable policy which results in the robot to walk. In RL the robot explores the states and actions with possible higher rewards. It is obvious that some part of this set would not explored at all or definitely would be with high punishment. Hence, decreasing the the state-action space by omitting the unreachable configurations or actions would help the computation speed of learning which is considerable.

C. Fuzzy Reward in learning to walk

One of The challenges in this content is to establish the interaction between the environment and the six-legged robot. A system that would tell the robot how good is its movement or its actions. A typical way is using a mathematical function which uses sensory data and tells how good or bad an action is in a certain state. [5]. Another approach which is implemented here is a fuzzy system that uses sensory data and tells the reward value. The design of fuzzy system is shown in figure 3 and discussed in details in [7]. The rules are defined as walking in straight line takes better reward as walking backward or tilting. Fig 4 shows the surface of fuzzy reward.

In RL for walking the reward signal is generated using a fuzzy system and comparison of the results is shown that fuzzy rewarding is accurate enough for learning to walk.

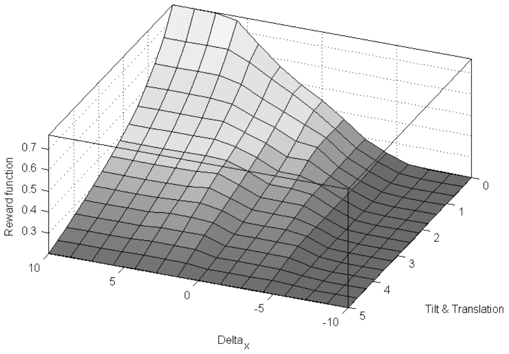


Fig. 4: The surface of fuzzy reward for walking.

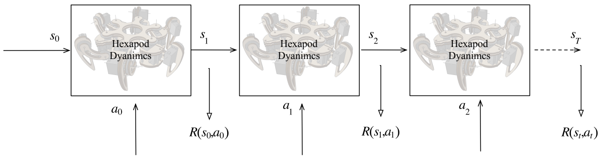


Fig. 5: Policy search for hexapod walking learning.

III. Learning algorithm and discussion

Although at the first step the aim was to implement the learning on the real built prototype of SiWaRel hexapod, the literature was shown that online learning requires a long time and in most of same projects, circuits or motor burning was so common. Therefore a dynamic model of the robot in simulation environment as the simulator is developed. The simulator task is to simulate the robot in dynamic environment as accurate as how the robot would act in real world. In different states different actions are taken and the reward is then calculated as it is illustrated in figure 5.

The robot takes actions based on ϵ-greedy action selection and goes to the next states. ϵ-greedy is a sub optimal action selection which as the literature shows it is promised to have a better performance due to the fact that a good learner does not always choose the optimal choice. The reward of the current state and action is calculated and stored. The reward values should be stored in a way to be used in action selection and finding the best optimal policy.

Q-learning is a kind of reinforcement learning in which no model of environment is needed. In Q-learning all the values are stored in a Q matrix. The simple Q-learning updates Q in every state and action by the immediate reward and the maximum Q value in the next state. As it is clear the Q matrix should have the dimension of $s \times a$.

$$Q(s,a) = R(s,a) + \lambda \underset{a}{\max} Q(s',a') \tag{3}$$

s', a' denotes the future action and state that would be taken. R and $\lambda < 1$ are the reward value and the discount factor respectively. It is shown in [5] that the Kalman filtered version of Q-learning has better performance for this purpose as Burton and Sutton shown in text [9].

$$Q(s,a) = \alpha[R(s,a) + \lambda \underset{a}{\max} Q(s',a')] + (1-\alpha)Q(s,a) \tag{4}$$

The schematic of learning to walk for hexapod robot is shown in figure 6. The learning starts from s_0 when all the legs are lifted. The state is looked up in the state/inverse-kinematic table for skipping kinematic analysis computation time and the joint values are sent into the hexapod dynamics simulator. An action i.e. a new state is taken by the action selection policy and again the new joint variables are looked up in the state/inverse-kinematic table and then are sent to the hexapod dynamics simulator. Then the simulation runs and the movement of the robot from the state to a new state is simulated. After finding out the simulator's result, the fuzzy reward is calculated. Using the reward, action, new and old states the Q is updated and this procedure is repeated for the new state until the learning completes.

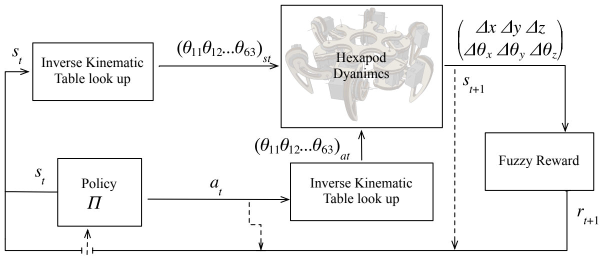

s_t
Inverse Kinematic Table look up
$(\theta_{11}\theta_{12}...\theta_{63})_{st}$
Hexapod Dyanimcs
$\begin{pmatrix} \Delta x\ \Delta y\ \Delta z \\ \Delta\theta_x\ \Delta\theta_y\ \Delta\theta_z \end{pmatrix}$
s_{t+1}
Fuzzy Reward
$(\theta_{11}\theta_{12}...\theta_{63})_{at}$
s_t
Policy
Π
a_t
Inverse Kinematic Table look up
r_{t+1}

Fig. 6: The schematic of reinforcement learning algorithm for hexapod robot in learning to walk.

As it is said before, action a_t is defined as the new state s_{t+1} but it should be mentioned that the opposite is not necessarily correct. It means that s_{t+1} is not always a_t. Assuming a time that a leg has became malfunctioned or is not working properly, in this situation a_t which is defined as new state is sent to the simulator and although the robot tries to go to new state, s_{t+1} is something different from that state which a_t tried to go. If failure has been caused in one or two leg it can learn how to walk with the other healthy legs.

IV. Results

Executing the iterative procedure of the learning algorithm leads to convergence of the Q values after about 150000 iterations. By increasing the greediness of action selection it is shown that Q matrix would be updated with spending less time on exploring non optimal places in state action space with the cost of acting more restricted. Figure 7 shows the updated $Q(s,a)$ values on the left and the explored state action space on the right.

Implementing the learning results on the hexapod dynamic model shows that the hexapod robot learned to walk well using the fuzzy system for rewarding in reinforcement learning. As it is shown in figure 8 the robot is walking in the right direction using the optimal policy.

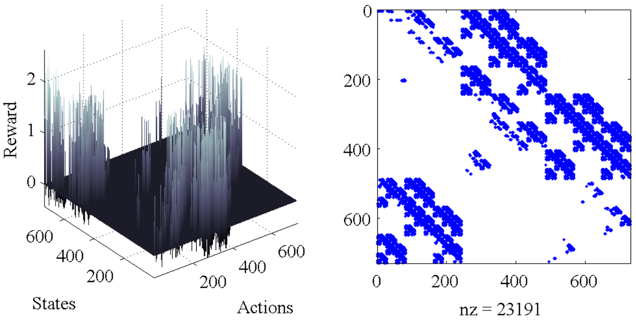
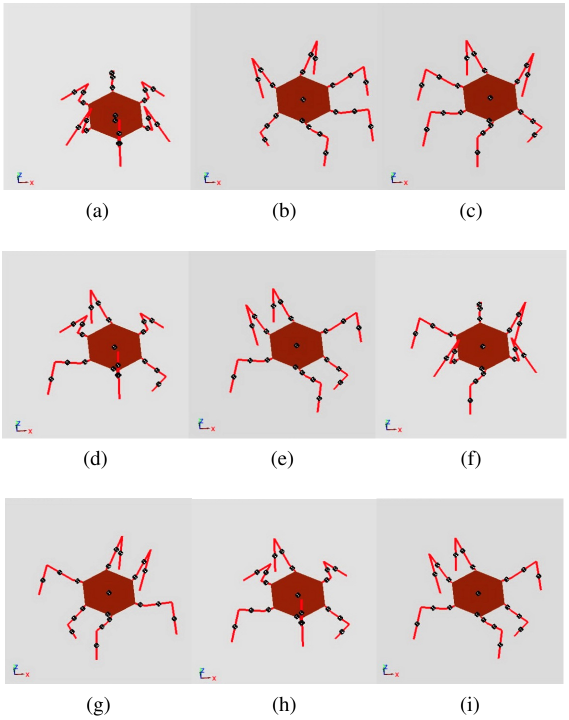


Fig. 7: Converged Q with ϵ-greedy action selection, $\epsilon = 0.2$.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 8: Evaluating the optimal policy on the dynamic model of hexapod robot. As it can be seen the robot is moving on the desired direction $(+x)$.

V. Conclusion

This paper studied the learning capability of an six-legged walking robot to walk using fuzzy reward reinforcement learning approach. As it is analyzed the hexapod robot has the ability of learning to walk with the presented approach. The simulation results have shown the efficiency of fuzzy rewarding system in gait analysis.

References

[1] A. Preumont, "An investigation of the kinematic control of a six-legged walking robot," *Mechatronics*, vol. 4, no. 8, pp. 821–829, 1994.

[2] R. Vidoni and A. Gasparetto, "Efficient force distribution and leg posture for a bio-inspired spider robot," *Robotics and Autonomous Systems*, vol. 59, no. 2, pp. 142–150, 2011.

[3] D. E. Koditschek, R. J. Full, and M. Buehler, "Mechanical aspects of legged locomotion control," *Arthropod Structure & Development*, vol. 33, no. 3, pp. 251–272, 2004.

[4] J. Estremera, J. Cobano, and P. Gonzalez de Santos, "Continuous free-crab gaits for hexapod robots on a natural terrain with forbidden zones: An application to humanitarian demining," *Robotics and Autonomous Systems*, vol. 58, no. 5, pp. 700–711, 2010.

[5] M. R. Bunting and J. Rogers, "Q-learning hexapod (may 2009)."

[6] A. Roennau, T. Kerscher, and R. Dillmann, "Design and kinematics of a biologically-inspired leg for a six-legged walking machine," in *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*. IEEE, 2010, pp. 626–631.

[7] M. Shahriari, K. G. Osguie, and A. A. A. Khayyat, "Modular framework kinematic and fuzzy reward reinforcement learning analysis of a radially symmetric six-legged robot," *Life Science Journal*, vol. 10, no. 8s, 2013.

[8] S. B. Thrun, "Efficient exploration in reinforcement learning," 1992.

[9] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. Cambridge Univ Press, 1998, vol. 1, no. 1.