Empirical Control System Development for Intelligent Mobile Robot Based on the Elements of the Reinforcement Machine Learning and Axiomatic Design Theory

Marko Mitić
PhD Student
University of Belgrade
Faculty of Mechanical Engineering

Zoran Miljković
Full Professor
University of Belgrade
Faculty of Mechanical Engineering

Bojan Babić
Full Professor
University of Belgrade
Faculty of Mechanical Engineering

This paper presents the authors' efforts to conceptual design of control system that can learn from its own experience. The ability of adaptive behaviour regarding the given task in real, unpredictable conditions is one of the main demands for every intelligent robotic system. To solve this problem, the authors suggest a learning approach that combines empirical control strategy, reinforcement learning and axiomatic design theory. The proposed concept uses best features of mentioned theoretical approaches to produce optimal action in the current state of the mobile robot. In this paper empirical control theory imparts the basis of conceptual solution for the navigation problem of mobile robot. Reinforcement learning enables the mechanisms that memorize and update environment responses, and combining with the empirical control theory determines best possible action according to the present circumstances. Axiomatic design theory accurately defines the problem and possible solution for the given task in terms of the elements defined by two previously mentioned approaches. Part of the proposed algorithm was implemented on the LEGO Mindstorms NXT mobile robot for the navigation task in an unknown manufacturing environment. Experimental results have shown good perspective for development of efficient and adaptable control system, which could lead to autonomous mobile robot behaviour.

Keywords: *learning mobile robot, empirical control theory, reinforcement learning, axiomatic design theory, mobile robot navigation.*

1. INTRODUCTION

One of the key objectives in modern robotics is to produce such a behaviour that is adaptive in real, stochastic conditions. In order to have productive, safe, and robust working robots, we need them to be able to cope with the dynamic nature of real environments: like humans or animals, robot should be able to adapt and learn from their own experiences instead of relying on predefined rules, models, or hardware controllers [1]. The same robot, running the same control program, can act differently considering real conditions in moment of a robot state transition. Hence, the reliability of such system cannot be satisfactory in terms of producing the best possible behaviour in a given moment. It is clear that adaptability is one of the main characteristics robot should possess.

The presented approach to conceptual design of control system inevitably leads to involving algorithm that includes active learning parameters. Those variables must store the environmental response of performed robot action, and also must indicate to control system what is the best possible action in the current robot state, with aspect to the real conditions and specified task. In that sense, several approaches can be distinguished: evolutionary computation, reinforcement learning (RL), empirical control (EC) theory [2], and others. These methods are well known and well established in various solutions for robot motion control problems.

Also, one of the main missions comprising mobile robot navigation task is properly and accurately defining a problem and a solution. From that point of view, it is necessary to design functional requirements and design parameters as elements of axiomatic design theory (AD) developed by Professor Suh of MIT [3]. Design, in Suh's terms, consists of a continuous interaction between the functional and the physical spaces. In that context, word functional may refer to the ability of a mobile robot learning based on the empirical data gathered from external sensors, i.e. main requirement in mobile robot navigation problem. The term physical may refer to the proposed empirical control algorithm, which is the main design parameter. One of the paper's aims is to present axiomatic design theory as a systematic tool for structure development of the mobile robot control system related to the navigation problem, and also a proper solution to that problem.

This paper sets up an original empirical control system based on the elements of reinforcement learning, particularly designed for solving the task of mobile robot navigation in unknown manufacturing environment.

Received: November 2010, Accepted: January 2011
Correspondence to: Marko Mitić, M.Sc.
Faculty of Mechanical Engineering,
Production Engineering Department,
Kraljice Marije 16, 11120 Belgrade 35, Serbia
E-mail: mmitic@mas.bg.ac.rs

Obtained results show good starting point for developing an autonomous behaviour of a mobile robot.

2. LEARNING METHODS

In this section will be explained basic concepts that are used in proposed empirical control system design. Empirical control theory and reinforcement learning theory will be clarified, with application to the proposed algorithm. Similarity and dissimilarity between these two approaches will also be pointed out.

2.1 Empirical control theory

Foundations of EC theory were first established by R.A. Brown [2]. Inspiration for this approach Brown got from observing real time natural systems and their interaction with the environment. He perceived that natural system or individual that behaves successfully appears to understand the requirements and information natural environment, which is direct consequence of the presence of natural intelligence. Much of the ability of natural systems comes about through practice and experience. The true value of obtained experience is demonstrated by comparing the systems' first attempt to execute a given task with a performance of the system after a large number of iterations. Such empirical systems have objectives which can be met by a system are designed to carry out just three steps presented in Table 1 [2,4,5].

Table 1. Three steps used by a self-learning system

No.	Description
1.	Produce certain behavior under certain conditions.
2.	Measure whether that behavior is carried out.
3.	Produce the behavior that has the highest probability of being carried out succesfully under those conditions.

On the basis of these three steps the empirical control algorithm for industrial robot learning has been developed [4,5]. Four simple steps defined in this algorithm [4,5] create the growth, evolution, i.e. successful development of this robot, which combined with artificial neural networks [6] represents successfully developed empirical control system.

Also, EC theory served as inspiration for designing a new hybrid control architecture for intelligent mobile robot navigation in a manufacturing environment [7]. So, from these examples one can realize the enormous potential that lies in described settings for control system design.

2.2 Introduction to reinforcement learning

In a RL paradigm [8], an agent interacts with the environment through a set of actions. The environment is modified in the sense of agent perception through external sensor and state in the next time step according to the selected action. Furthermore, at each step the agent receives an external award, as shown in Figure 1.

The objective of the RL agent is to maximize a numerical reward signal [9,10]. The main advantage of RL is that it does not need the model of the environment, i.e. the path planning algorithm for mobile robot navigation is not necessary.

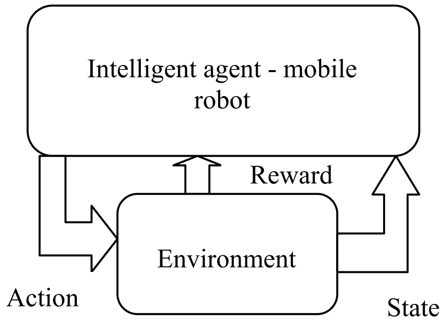

Figure 1. Basic model of a RL agent

Watkins introduced in 1989 the method of reinforcement learning called Q-learning. [11]. Q-learning algorithm attempts to learn a state-action value $Q\,(s,a)$, whose value is the maximum discounted reward that can be achieved by starting in state x, taking an action a, and following the optimal policy thereafter [12]. The action space is discrete and a separate $Q\,(s,a)$ value exists for each state-action pair.

In each time step the agent takes an action a from the state s, and the current state-action pair value estimate from a and s donated by $Q_t\,(s,a)$ is updated as follows [12]:

$$Q_{t+1}\left(s,a\right) \leftarrow Q_t\left(s,a\right) + \\ +\alpha\left[r_{t+1} + \gamma \max_a Q_t\left(s_{t+1},a\right) - Q_t\left(s_t,a\right)\right] \quad (1)$$

where: $t+1$ denotes the time constant in the next robot state; γ is the discount factor with value between 0 and 1; r_{t+1} is payoff that agent receives when action a is taken in state s; and parameter α is a learning rate.

Recommended values for scalars γ and α are ≥ 0.9 and ≤ 0.2, respectively.

Pseudo code for Q-learning algorithm is given in Table 2.

Table 2. Pseudo code of Q-learning algorithm

Initialize state-action function $Q\,(s,a)$
Present current state S_t
Calculate optimal action
Execute selected action (ε-greedy) a_t
Observe new state and reinforcement signal, S_{t+1} and r_{t+1}, respectively
Update state-action function as $Q_{t+1}\left(s,a\right) \leftarrow Q_t\left(s,a\right) + \alpha\left[r_{t+1} + \gamma \max_a Q_t\left(s_{t+1},a\right) - Q_t\left(s_t,a\right)\right]$
New state becomes current state

2.3 Similarity and diversity of presented learning methods

Although the presented methods have a lot in common, the main difference between them reflects onto the approach of learning state-action value function. While reinforcement learning modifies $Q_t\,(s,a)$ (in each iteration), the empirical theory tends to guide the agent to remember all of the previous transition states and applied actions. In that way, for the same sensory readings, the agent will select successful action according to the previously same obtained environment response. This should largely reduce the number of implied iterations of robot state-action transitions. At the same time, Q-learning represents great model for storing state transition probabilities, which can be employed in the novel control design approach. Thus, the authors propose the hybrid control system that contains the best features of both described methods.

3. PROBLEM STATEMENT

Described control system design methods will be partially implemented for the problem of the mobile robot navigation in unknown environment. The presented algorithm also includes a solution to an obstacle avoidance problem [13,14], although it has not yet been implemented mainly because of the large number of iterations needed. Iterations necessary for successful intelligent behaviour overcome tens of thousands, because the *tabula rasa* mobile robot [10] is not capable of faster learning. As mentioned, mobile robot should visit every possible state and produce every possible action in that state to have complete knowledge of optimal navigation path.

For this study the starting and goal point was chosen randomly. The robot was acting according to the defined actions choosing them randomly too. The values for $Q(s,a)$ were updated in accordance with the described Q-learning algorithm. Memorized sensor readings were stored in a matrix, and for each set of gathered empirical data $Q(s,a)$ value was assigned. Then, after a certain number of iterations, the two $Q(s,a)$ values were compared, and the one with higher probability was chosen. Obtained data was processed in the MATLAB software package [15], and then graphically presented as shown in the section below.

4. MAIN CONCEPT OF AXIOMATIC DESIGN

For the purpose of proper design and development of solution to the given task, the Axiomatic Design theory (AD) developed by Suh was adopted [3]. The described theory presents rigorous rules within solutions design for any given engineering problem. In this case, the main problem can be described as adaptive learning problem during navigation task performed by a mobile robot. That problem can be divided into two sub problems, that is, a problem of learning during specific task and a problem of using that learned experience to improve the existing behaviour.

According to [16,17] the design process in axiomatic design theory consists of five steps stated in the Table 3.

The design axioms present the basis for the concept of AD. The first axiom is known as the *Independence Axiom*, and the second one is known as the *Information Axiom*. Their description is given in the Table 4 [3].

As stated earlier, the engineering axiomatic design refers to mapping between functional and physical domain. These domains are defined with functional requirements (FR) and design parameters (DP), respectively [3,17,18]. In mathematical terms, the relationship between the FRs and DPs is expressed as:

Table 3. Several steps of design process defined by AD theory

Step No.	Description
1.	Establishment of designed goals to satisfy a given set of percieved needs.
2.	Conceptualization of designed solutions.
3.	Analysis of the proposed solution.
4.	Selection of the best design from among those proposed.
5.	Implementation.

Table 4. Axioms of AD and their short description

Axiom No.	Name and description
1.	*The Indipendence Axiom*. Maintain the independence of functional requirements.
2.	*The Information Axiom*. Minimizes the information concept.

$$\{\mathbf{FR}\} = |\mathbf{A}| \cdot \{\mathbf{DP}\}. \tag{2}$$

In given equation, $\{\mathbf{FR}\}$ denotes the functional requirement vector, $\{\mathbf{DP}\}$ denotes the design parameter vector, and $|\mathbf{A}|$ denotes the design matrix that characterizes the design process. The structure of the matrix $|\mathbf{A}|$, defines the type of design being considered. In order to satisfy the first axiom of AD, matrix $|\mathbf{A}|$ should be uncoupled or coupled design.

In *uncoupled design*, the $|\mathbf{A}|$ matrix is a diagonal matrix, whose shape indicates the independence of FR-DP pairs. So, logically, this type of design is most preferred. *Decoupled design* has the triangular design matrix $|\mathbf{A}|$. This indicates that FRs can be satisfied systematically from the first FR to the last one, by considering the first n DPs only. In the previous sentence n denotes total number of FRs. This type of design is most common in practice. If the design matrix has no specific shape, then those designs are called *coupled designs*. These designs are undesirable, and every system designer should try to avoid them.

4.1 Axiomatic design theory applied on the given control problem

The first stage in designing the solution for the stated problem is to define the main functional requirement, i.e. to define the functional requirements (FRs) of the system in the highest level of all FRs in the functional domain. In this step extreme care should be given to choosing the right functional requirement, since different FR can lead to completely different solutions. Since the main problem has been formulated, the functional requirement of top hierarchal level can be defined as follows:

- FR: The ability of mobile robot to learn (using Q-learning algorithm) based on the obtained empirical data from the environment.

Appropriate DP can be formulated as the proposed method of empirical control system of a mobile robot. So, the design parameter of the highest level is formulated as:

- DP: proposal of the empirical control system.

In the next steps the zigzagging process is applied. First, the main FR is decomposed into two levels, so that each level consists of several lower levels FRs. The same principle is applied on the main DP. In that manner, every DPs in the lower level corresponds to defined FRs in the same level. Also, the analysis of appropriate design matrix is given. Decoupled and uncoupled design matrices have been obtained, so the given result confirms the proposed solution to the problem. All of the experimental results regarding FRs and DPs at the lower levels with the corresponding design matrices are presented in Table 5.

As it is shown in the table, the decoupled designs are obtained in every case, except for the second FR in the third hierarchical level. In that case, the uncoupled design is obtained, which is the best possible outcome of the design process. The results show that the proposed control system developed for a given navigation task is based on good outlines, which give solid foundations for further research.

5. EXPERIMENTAL SETUP AND DISCUSSION

Every mentioned aspect of control system design is applied to get an operational system for a navigational problem task. The configuration selected for implementing the proposed algorithm is LEGO Mindstorms NXT configuration, as shown in Figures 2 and 3.

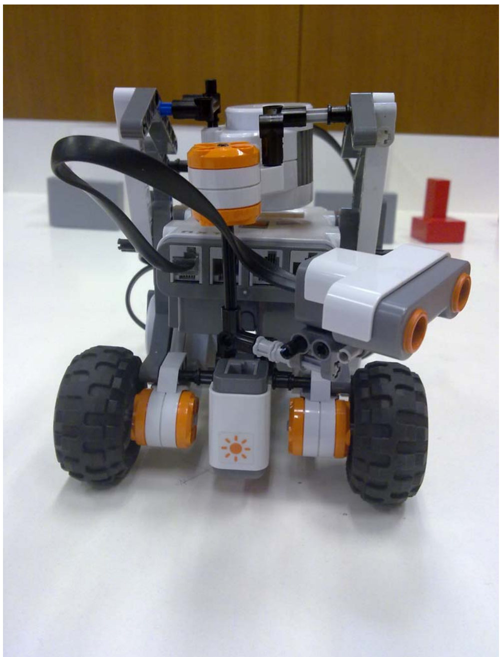

Figure 2. LEGO Mindstorms NXT configuration of a mobile robot in the laboratory model of environment – front view

As an external sensor for obtaining empirical data in the environment, the ultrasonic sensor available in the existing LEGO kit was selected. Its motor was placed above the configuration, so as to minimize size of the configuration and maximize robot's performance. The sensor range is defined by its manufacturer [19], and in case of testing the idea proposed in this paper it showed

Table 5. FRs and DPs at lower levels and their corresponding design matrix A

Hierarchical level	FRs (description)	DPs (description)	A (type of design)
II	FR_1 Execution of the given task	DP_1 Navigation task of mobile robot in a manufacturing environment	$\begin{bmatrix} X & & \\ X & X & \\ X & X & X \end{bmatrix}$ (decoupled design)
	FR_2 Ability for adaptive behaviour	DP_2 Collecting data from external sensors	
	FR_3 Memorizing successful actions and states, and also the level of confidence for all performed actions	DP_3 Application of Q-learning algorithm	
III	FR_{11} Defining possible actions	DP_{11} Set of three possible actions	$\begin{bmatrix} X & & \\ X & X & \\ X & X & X \end{bmatrix}$ (decoupled design)
	FR_{12} Defining possible states	DP_{12} Four possible states	
	FR_{13} Odometry model of mobile robot navigation	DP_{13} Data from the incremental encoders (mobile robot wheels)	
	FR_{21} Obstacle avoidance	DP_{21} Correct reading and processing of data received from external sensors	$\begin{bmatrix} X & & \\ & X & \\ & & X \end{bmatrix}$ (uncoupled design)
	FR_{22} Adaptability for various real time conditions	DP_{22} Comparison of real and calculated state response of a mobile robot	
	FR_{23} Partially adaptive behaviour to unexpected events in dynamic environment	DP_{23} Implementation of autonomous behaviour model into the future mobile robot decisions	
	FR_{31} Memorizing best action in the given moment	DP_{31} Reward endue depending of the selected action	$\begin{bmatrix} X & & \\ X & X & \\ X & X & X \end{bmatrix}$ (decoupled design)
	FR_{32} Determine level of confidence of the selected action	DP_{32} State-action value function $Q(s,a)$ estimation based on the Q-learning algorithm	
	FR_{33} Increasing/decreasing level of reliability for executed action	DP_{33} Comparison of selected action from the Q-table with empirical control theory	

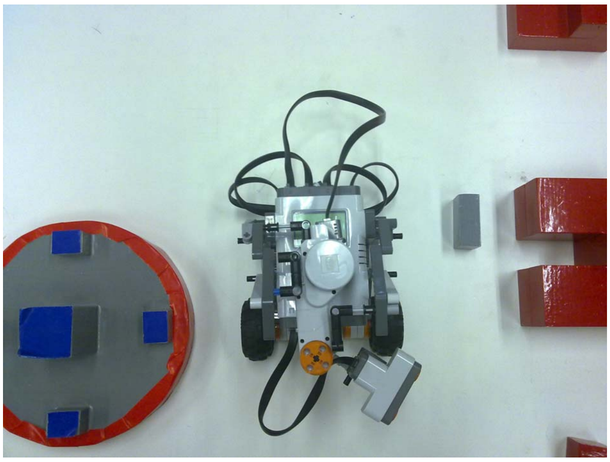

Figure 3. LEGO Mindstorms NXT configuration of a mobile robot in the laboratory model of environment – top view

satisfactory results. The ultrasonic sensor reads the distance in five measuring points, crossing the angle range of 180 degrees. Its initial position is on the left (position 0 in the Figure 4) viewing from the direction in which the robot is moving. It is activated at every 45 degrees, which in total gives five measurements in one state-action iteration. As mentioned, those values are then saved and are given an appropriate value. When the whole angle range of 180 degrees is visited, the sensor moves back to its initial position. In that backward motion the sensor is not active, i.e. it perceives readings in just one direction (clockwise). The robot is controlled by MATLAB package [15], using RWTH-toolbox [20]. In the Figure 4, the robot and sensor measurement process with measurement positions are symbolically represented.

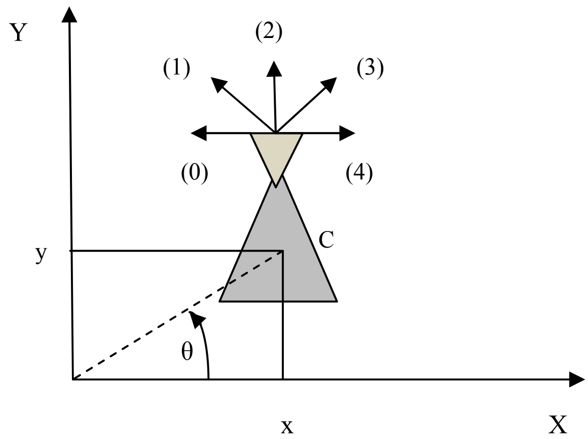

Figure 4. Mobile robot and sensor measurements in a global coordinate system

The set of possible actions is defined according to the Table 6.

Table 6. Defined actions and short description

Action No.	Name of action and description
1.	*Move forward.* Strateline forward movement of mobile robot cca 1 cm long.
2.	*Move left.* Mobile robot turns left at the angle of 45 degrees.
3.	*Move right.* Mobile robot turns right at the angle of 45 degrees.

The reward existing in the (1) is defined according to the measurement of an ultrasonic sensor. The numerical values for reward signals r are assigned as shown below:

$$r = \begin{cases} 0, \min J = \min U \\ 1, \min J < \min U \\ -1, \min U < \min J \\ 2, x_{\text{present}} = x_{\text{goal}} \end{cases} \quad (3)$$

where J denotes previous set of measurements, and U set of measurements in the current state. The variable x_{present} represents current state in which the mobile robot is, and x_{goal} denotes final state of a mobile robot. Clearly, the algorithm rewards robot movement away from the obstacle and reaching the final goal too, and punishments movement that lead closer to the static boundary. Odometry model was assigned for defining position and orientation of a mobile robot in global coordinates.

$$x' = \begin{bmatrix} x \\ y \\ \theta \end{bmatrix} + \begin{bmatrix} \dfrac{\Delta s_d + \Delta s_l}{2} \cdot \cos\left(\theta + \dfrac{\Delta s_d - \Delta s_l}{2b}\right) \\ \dfrac{\Delta s_d + \Delta s_l}{2} \cdot \sin\left(\theta + \dfrac{\Delta s_d - \Delta s_l}{2b}\right) \\ \dfrac{\Delta s_d - \Delta s_l}{2b} \end{bmatrix}. \quad (4)$$

In the (4) x' denotes the next state vector of a mobile robot. Constant b marks the wheelbase length, and Δs_d and Δs_l denotes the incremental path lengths in a mobile robot transition from one state to another. Clearly, x, y, and θ marks the current position and orientation of a robot.

For parameters α and γ constant values have been adapted accordingly to the best results obtained. In this case, these parameters have 0.1 and 0.99 values, respectively.

The environment state space in this setup is discretized, so that the Q-table has reasonably a large size. In that sense, the biggest problem of this approach is designing the state and action space in the way that matrix $Q\,(s,a)$ is not very computational expensive. Several solutions are proposed to reduce the $Q\,(s,a)$ size, from which the artificial neural network (ANN) [6] approach gave overall best results [12].

Given all stated in mind, a modification of obstacle avoidance Q-learning algorithm [21] is given. The improvement of this algorithm is reflected onto the employment of the empirical control theory in the whole process. The modified Q value is memorized not only in the present table, but also in the space reserved for the sensor measurement. In that sense, Q function will faster converge to its optimal value. Also, the new state will have additional source of information regarding earlier sensor measurement and prescribed Q values, in order to obtain and perform the best possible action in a given moment. With blue dashed line the novel empirically enhanced obstacle avoidance control system is denoted. This algorithm is presented in the Figure 5.

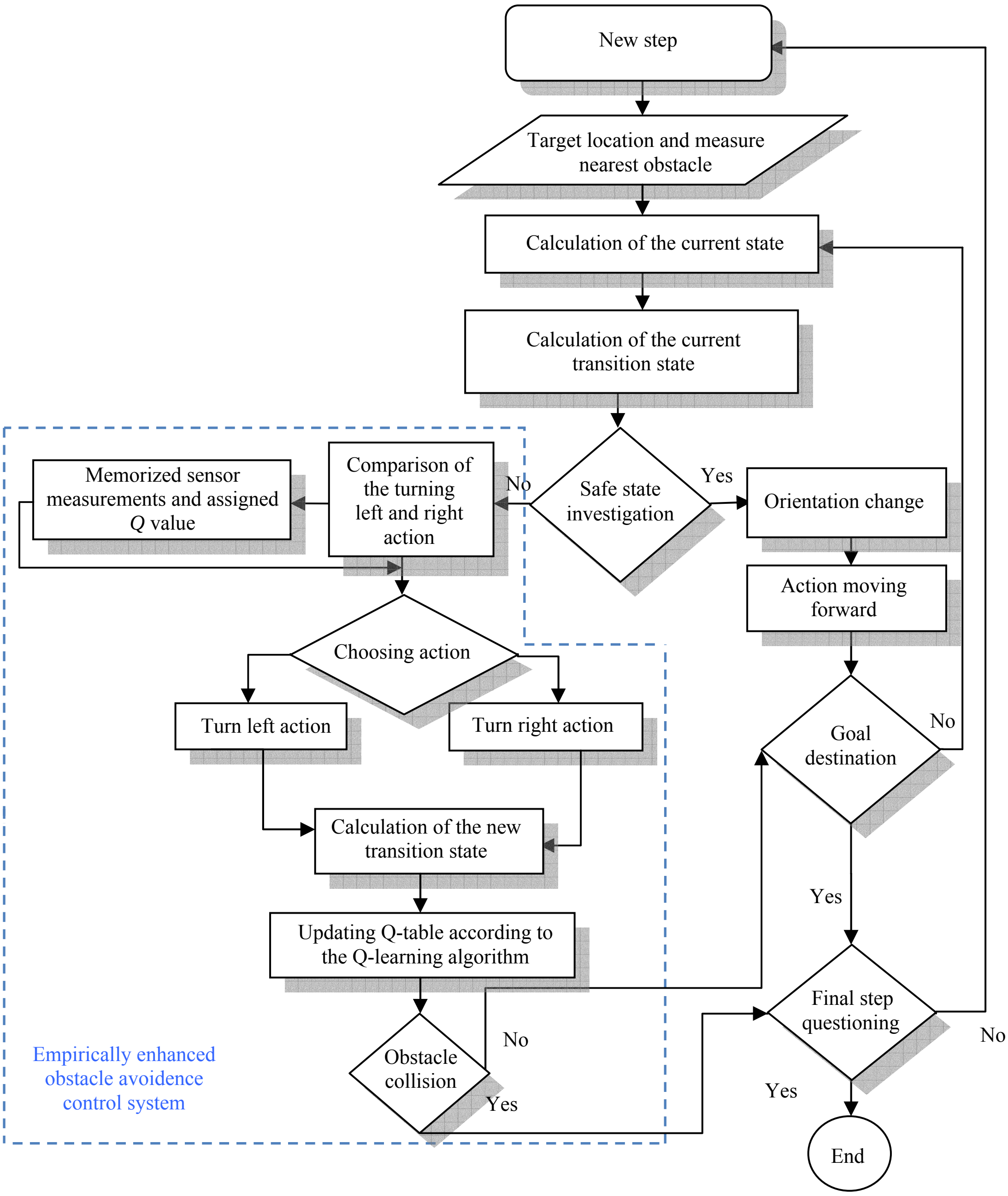

Figure 5. Empirically enhanced obstacle avoidance algorithm

5.1 Experimental results

Conducted experiment (without obstacle collision) within the laboratory model of manufacturing environment is based on presented algorithm (Fig. 5). Although *tabula rasa* mobile robot [10] needs enormous number of iterations to fully understand the environment, the results presented below showed that the Q function correctly updated its coefficients. Because the large number of iterations is conducted (for real time learning), they were divided into a number of steps in which the robot should reach the goal position. Iteration is ended in two possible cases: if a mobile robot reached the goal for a reasonable time or if the user defined time has been exceeded. The Q values were memorized (for each iteration), and shown in the graph. The obtained results (Figs. 6, 7 and 8) denote the number of steps to end iteration, and updated Q function for each iteration (episode) that has ended.

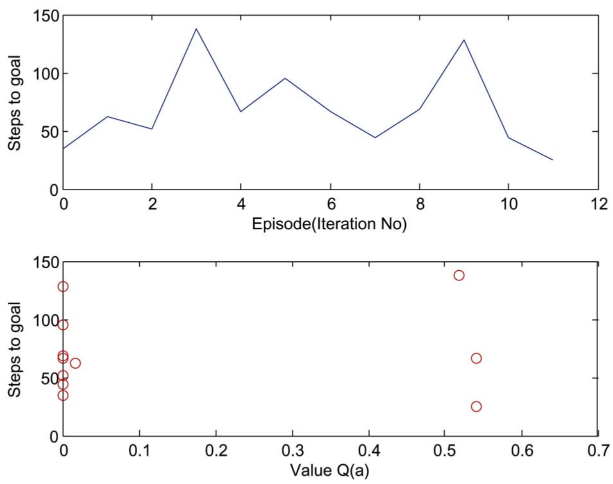

Figure 6. Experimental results at 11th iteration

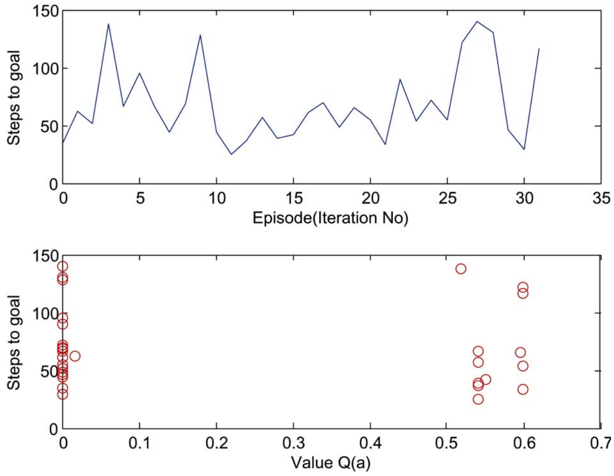

Figure 7. Experimental results at 31st iteration

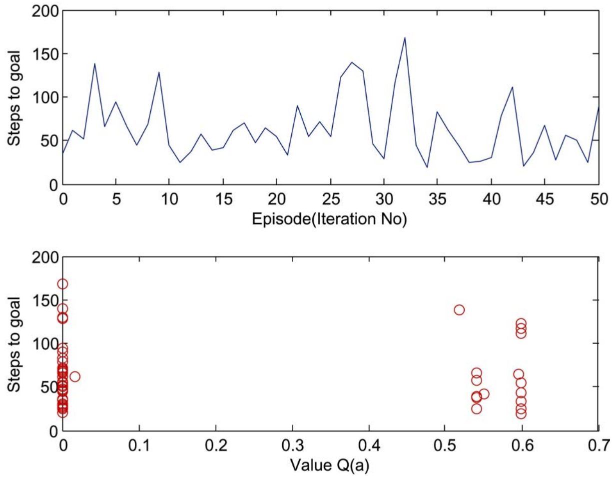

Figure 8. Experimental results at 50th iteration

6. CONCLUSION

In this paper a new approach for advance control system design is presented. The proposed approach is based on the empirical control theory, reinforcement learning, and the axiomatic design theory. The concept is verified for the control problem of mobile robot navigation in an unknown environment. For the algorithm evaluation LEGO Mindstorms NXT mobile robot was used, which was controlled with MATLAB software package.

The successful machine learning process of a mobile robot, as shown in the Figures 6, 7 and 8, was evident. Modifying the coefficients in the Q matrix, mobile robot was able to make difference between favourable actions in its current state. The Q values were adjusted in accordance with the described reinforcement learning algorithm. Also, a set of sensor measurements was memorized and for each of them the appropriate Q value was awarded. That value was used as a necessary advice for the decision of optimal action selection in the present robot state after 50^{th} iteration. With more iterations conducted, a mobile robot could perform autonomous behaviour as a solution for the navigation problem based on the shown experimental results and machine learning trend of the proposed control system.

For future research, as much needed development tool for Q function approximation, the artificial neural networks (ANN) arise above other solutions for the considered problem. Also, ANN can be very useful for speeding up the learning process, bearing in mind the computational expensiveness of the classical Q-learning algorithm. Particularly, implementation of the neural networks with a dynamical structure will be considered.

ACKNOWLEDGMENT

The authors would like to thank Ministry of Science and Technological Development of the Republic of Serbia for their financial support in the development of this research, which is a part of the project TR 35004.

The first author would additionally like to thank Ministry of Science and Technological Development of the Republic of Serbia for enabling him to do research work through the given PhD scholarship under the Contract No. 828 (2010).

REFERENCES

[1] Quintía, P., Iglesias, R., Rodríguez, M.A. and Regueiro, C.V.: Simultaneous learning of perception and action in mobile robots, Robotics and Autonomous Systems, Vol. 58, No. 12, pp. 1306-1315, 2010.

[2] Brown, R.A.: *Machines That Learn*, Oxford University Press, New York, 1994.

[3] Suh, N.P.: *The Principles of Design*, Oxford University Press, New York, 1990.

[4] Miljković, Z. and Babić, B.: Empirical control strategy for learning industrial robot, FME Transactions, Vol. 35, No. 1, pp. 1-8, 2007.

[5] Miljković, Z.: *Systems of Artificial Neural Networks in Production Technologies*, Faculty of Mechanical Engineering, University of Belgrade, Belgrade, 2003, (in Serbian).

[6] Miljković, Z. and Aleksendrić, D.: *Artificial Neural Networks – Solved Examples With Short Theory Introduction*, Faculty of Mechanical Engineering, University of Belgrade, Belgrade, 2009, (in Serbian).

[7] Vuković, N. and Miljković, Z.: New hybrid control architecture for intelligent mobile robot navigation in a manufacturing environment, FME Transactions, Vol. 37, No. 1, pp. 9-18, 2009.

[8] Kaelbling, L.P., Littman, M.L. and Moore, A.W.: Reinforcement learning: A survey, Journal of Artificial Intelligence Research, Vol. 4, pp. 237-285, 1996.

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

[10] Moreno, D.L., Regueiro, C.V., Iglesias, R. and Barro, S.: Using prior knowledge to improve reinforcement learning in mobile robotics, in: *Proceedings of the Conference Towards Autonomous Robotic Systems – TAROS 2004*, 06-08.09.2004, Colchester, UK.

[11] Watkins, C.: *Learning from Delayed Rewards*, PhD thesis, King's College, London, 1989.

[12] Huang, B.-Q., Cao, G.-Yi. and Guo, M.: Reinforcement learning neural network to the problem of autonomous mobile robot obstacle avoidance, in:

Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, 18-21.08.2005, Guangzhou, China, pp. 85-89.

[13] Lewis, F.L.: Neural network feedback control: Work at UTA's Automation and Robotics Research Institute, Journal of Intelligent & Robotic Systems, Vol. 48, No. 4, pp. 513-523, 2007.

[14] Carreras, M., Yuh, J., Batlle, J. and Ridao, P.: Application of SONQL for real-time learning of robot behaviours, Robotics and Autonomous Systems, Vol. 55, No. 8, pp. 628-642, 2007.

[15] http://www.mathworks.com.

[16] Babic, B.: Axiomatic design of flexible manufacturing systems International Journal of Production Research, Vol. 37, No. 5, pp. 1159-1173, 1999.

[17] Kim, S.-J., Suh, N.P. and Kim, S.-G.: Design of software system based on axiomatic design, CIRP Annals – Manufacturing Technology, Vol. 40, No. 1, pp. 165-170, 1991.

[18] Park, G.-J.: *Analytic Methods for Design Practice*, Springer-Verlag, London, 2007.

[19] http://mindstorms.lego.com.

[20] http://www.mindstorms.rwth-aachen.de.

[21] Jaradat, M.A.K., Al-Rousan, M. and Quadan, L.: Reinforcement based mobile robot navigation in dynamic environment, Robotics and Computer-Integrated Manufacturing, Vol. 27, No. 1, pp. 135-149, 2011.

ПРИЛОГ РАЗВОЈУ ЕМПИРИЈСКОГ УПРАВЉАЧКОГ СИСТЕМА МОБИЛНОГ РОБОТА БАЗИРАНОГ НА ЕЛЕМЕНТИМА МАШИНСКОГ УЧЕЊА ОЈАЧАВАЊЕМ И АКСИОМАТСКОЈ ТЕОРИЈИ ПРОЈЕКТОВАЊА

Марко Митић, Зоран Миљковић, Бојан Бабић

Овај рад представља истраживање аутора у домену концепцијског пројектовања управљачког система који може да учи на основу сопственог искуства. Способност адаптивног понашања при извршавању постављеног задатка у реалним, непредвидивим условима, један је од кључних задатака сваког интелигентног роботског система. У функцији решавања овог проблема, предлаже се приступ базиран на учењу, и то комбиновањем емпиријске управљачке стратегије, машинског учења ојачавањем и аксиоматске теорије пројектовања. Предложени концепт користи најбоље особине поменутих теоријских приступа у циљу остваривања оптималне одлуке мобилног робота за тренутно стање система. Емпиријска управљачка теорија се, у овом раду, *a priori* користи у утврђивању идејног решења за решавање проблема навигације мобилног робота. Учење ојачавањем реализује механизме који меморишу и ажурирају одговоре окружења, а у комбинацији са емпиријском управљачком теоријом одређује најбољу могућу одлуку у складу са тренутним околностима. Аксиоматска теорија пројектовања се користи при дефинисању управљачког проблема, као и при успостављању концепцијског решења за дати задатак, са аспекта примене поменутих приступа. Део предложеног алгоритма емпиријског управљања реализован је помоћу *LEGO Mindstorms NXT* мобилног робота, третирајући проблем навигације у непознатом окружењу. Остварени експериментални резултати наговештавају добру перспективу за реализацију ефикасног управљања базираног на искуству, чији даљи развој може да доведе до остварења аутономног понашања мобилног робота при избегавању препрека у технолошком окружењу, што је и очекивани научни циљ.