Reduced Rule Base in Fuzzy Rule Interpolation-based Q-learning

Dávid Vincze, Szilveszter Kovács

Department of Information Technology, University of Miskolc
Miskolc-Egyetemváros, H-3515 Miskolc, Hungary
{david.vincze,szkovacs}@iit.uni-miskolc.hu

Abstract: Fuzzy Q-learning, the fuzzy extension of the Reinforcement Learning (RL) is a well known topic in computational intelligence. It can be used to tackle control problems in unknown continuous environments without defining an exact method on how to solve it explicitly. In the RL concept the problem needed to be solved is hidden in the feedback of the environment, called reward or punishment (positive or negative reward). From these rewards the system can learn which action is considered to be the best choice in a given state. One of the most frequently applied RL method is the "Q-learning". The goal of the Q-learning method is to find an optimal policy for the system by building the state-action-value function. The state-action-value-function is a function of the expected return (a function of the cumulative reinforcements), related to a given state and a taken action following the optimal policy. The original Q-learning method was introduced for discrete states and actions. With the application of fuzzy reasoning the method can be adapted for continuous environment, called Fuzzy Q-learning (FQ-Learning). Traditional Fuzzy Q-learning embeds the 0-order Takagi-Sugeno fuzzy inference and hence inherits the requirement of the state-action-value-function representation to be a complete fuzzy rule base. An extension of the traditional fuzzy Q-learning method with the capability of handling sparse fuzzy rule bases is already introduced by the authors, which suggests a Fuzzy Rule Interpolation (FRI) method, namely the FIVE (Fuzzy rule Interpolation based on Vague Environment) technique to be the reasoning method applied with Q-learning (FRIQ-learning). The main goal of this paper is the introduction of a method which can construct the requested FRI fuzzy model in a reduced size. The suggested reduction is achieved by incremental creation of an intentionally sparse fuzzy rule base.

Keywords: reinforcement learning, fuzzy Q-learning, fuzzy rule interpolation, fuzzy rule base reduction

1 Introduction

Reinforcement learning methods can help in situations, where the task to be solved is hidden in the feedback of the environment, i.e. in the positive or negative rewards (the negative reward is often called a punishment) provided by the environment. The rewards are calculated by an algorithm created especially for expressing the task needed to be solved. Based on the rewards of the environment RL methods are approximating the value of each possible action in all the reachable states. Therefore RL methods can solve problems where a priory knowledge can be expressed in the form what is needed to be achieved, not in how to solve the problem directly. Reinforcement learning methods are a kind of trial-and-error style methods adapting to dynamic environment through incremental iterations. The primary ideas of reinforcement learning techniques (dynamical system state and the idea of "optimal return", or "value" function) are inherited from optimal control and dynamic programming [3]. A common goal of the reinforcement learning strategies is to gather an optimal policy by constructing the state-value- or action-value-function [19]. The state-value-function $V^{\pi}(s)$, is a function of the expected return (a function of the cumulative reinforcements), related to a given state $s \in S$ as a starting point, following a given policy π. These rewards, or punishments (reinforcements) are the expression of the desired final goal of the learning agent as a kind of evaluation following the previous action (in contrast to the instructive manner of error feedback based approximation techniques, for example the gradient descent optimisation). The optimal policy is basically the description of the agent behaviour, in the form of mapping between the agent states and the corresponding suitable actions. The action-value function $Q^{\pi}(s,a)$ is a function of the expected return, in case of taking action $a \in A_s$ in state s following policy π. In possession of the action-value-function, the optimal (greedy) policy, which always takes the optimal (the greatest estimated value) action in every state, can be constructed as [19]:

$$\pi(s) = \arg\max_{a \in A_s} Q^{\pi}(s,a) \tag{1}$$

For the estimation of the optimal policy, the action-value function $Q^{\pi}(s,a)$ should be approximated. Approximating the latter function is a complex task given that both the number of possible states and the number of the possible actions could be an extremely high value. Evaluating all the possibilities could take a considerable amount of computing resources and computational time, which is a significant drawback of reinforcement learning. However there are some cases where a distributed approach with continuous reward functions can reduce these resource needs [15]. Generally reinforcement learning methods can lead to results in practically acceptable time only in relatively small state and action spaces.

Adapting fuzzy models the discrete Q-learning can be extended to continuous state and action space, which in case of suitably chosen states can lead to the reduction of the size of the state-action space [12].

2 Q-learning and Fuzzy Q-learning

Q-learning is a reinforcement learning method which has the purpose of finding the fixed-point solution (Q) of the Bellman Equation [3] via iteration. In the case of discrete *Q-Learning* [23], the action-value-function is approximated by the following iteration:

$$Q_{i,u} \approx \widetilde{Q}_{i,u}^{k+1} = \widetilde{Q}_{i,u}^{k} + \Delta\widetilde{Q}_{i,u}^{k+1} = \widetilde{Q}_{i,u}^{k} + \alpha_{i,u}^{k} \cdot \left(g_{i,u,j} + \gamma \cdot \max_{v \in U} \widetilde{Q}_{j,v}^{k+1} - \widetilde{Q}_{i,u}^{k} \right) \quad (2)$$

$\forall i \in I, \forall u \in U$, where $\widetilde{Q}_{i,u}^{k+1}$ is the *k+1* iteration of the action-value taking the u^{th} action A_u in the i^{th} state S_i, S_j is the new (j^{th}) observed state, $g_{i,u,j}$ is the observed reward completing the $S_i \to S_j$ state-transition, γ is the discount factor and $\alpha_{i,u}^{k} \in [0,1]$ is the step size parameter (can vary during the iteration steps), *I* is the set of the discrete possible states and *U* is the set of the discrete possible actions. There are various existing solutions [1], [4], [5], [6] for applying this iteration to continuous environment by adopting fuzzy inference (called Fuzzy Q-Learning). Most commonly the simplest FQ-Learning method, the 0-order Takagi-Sugeno Fuzzy Inference model is adapted. Hereby in this paper the latter one is studied (a slightly modified, simplified version of the Fuzzy Q-Learning introduced in [1] and [6]). In this case for characterizing the value function *Q(s,a)* in continuous state-action space, the 0-order Takagi-Sugeno Fuzzy Inference System approximation $\widetilde{Q}(s,a)$ is adapted in the following way:

If *s* **is** S_i **And** *a* **is** A_u **Then** $\widetilde{Q}(s,a) = Q_{i,u}$, $i \in I, u \in U$, (3)

where S_i is the label of the i^{th} membership function of the *n* dimensional state space, A_u is the label of the u^{th} membership function of the one dimensional action space, $Q_{i,u}$ is the singleton conclusion and $\widetilde{Q}(s,a)$ is the approximated continuous state-action-value function. Having the approximated state-action-value function $\widetilde{Q}(s,a)$, the optimal policy can be constructed by function (1). Setting up the antecedent fuzzy partitions to be *Ruspini partitions*, the zero-order Takagi-Sugeno fuzzy inference forms the following approximation function:

$$\widetilde{Q}(s,a)=\sum_{i_1,i_2,\cdots,i_N,u}^{I_1,I_2,\ldots,I_N,U}\prod_{n=1}^{N}\mu_{i_n,n}(s_n)\cdot\mu_u(a)\cdot q_{i_1 i_2 \ldots i_N u} \tag{4}$$

where $\widetilde{Q}(s,a)$ is the approximated state-action-value function, $\mu_{i_n,n}(s_n)$ is the membership value of the i_n^{th} state antecedent fuzzy set at the n^{th} dimension of the N dimensional state antecedent universe at the state observation s_n, $\mu_u(a)$ is the membership value of the u^{th} action antecedent fuzzy set of the one dimensional action antecedent universe at the action selection a, $q_{i_1 i_2 \ldots i_N u}$ is the value of the singleton conclusion of the $i_1, i_2, \ldots, i_N, u$ -th fuzzy rule. Applying the approximation formula of the Q-learning (2) for adjusting the singleton conclusions in (4), leads to the following function:

$$\begin{aligned} q^{k+1}_{i_1 i_2 \ldots i_N u} &= q^{k}_{i_1 i_2 \ldots i_N u} + \prod_{n=1}^{N}\mu_{i_n,n}(s_n)\cdot\mu_u(a)\cdot\Delta\widetilde{Q}^{k+1}_{i,u} = \\ &= q^{k}_{i_1 i_2 \ldots i_N u} + \prod_{n=1}^{N}\mu_{i_n,n}(s_n)\cdot\mu_u(a)\cdot\alpha^{k}_{i,u}\cdot\left(g_{i,u,j}+\gamma\cdot\max_{v\in U}\widetilde{Q}^{k+1}_{j,v}-\widetilde{Q}^{k}_{i,u}\right) \end{aligned} \tag{5}$$

where $q^{k+1}_{i_1 i_2 \ldots i_N u}$ is the $k+1$ iteration of the singleton conclusion of the $i_1 i_2 \ldots i_N u$ th fuzzy rule taking action A_u in state S_i, S_j is the new observed state, $g_{i,u,j}$ is the observed reward completing the $S_i \rightarrow S_j$ state-transition, γ is the discount factor and $\alpha^{k}_{i,u} \in [0,1]$ is the step size parameter. The $\mu_{i_n,n}(s_n)\cdot\mu_u(a)$ is the partial derivative of the conclusion of the 0-order Takagi-Sugeno fuzzy inference $\widetilde{Q}(s,a)$ with respect to the fuzzy rule consequents $q_{u,i}$ according to (4). This partial derivative is required for the applied steepest-descent optimization method. The $\widetilde{Q}^{k+1}_{j,v}$ and $\widetilde{Q}^{k}_{i,u}$ action-values can be approximated by equation (4).

3 FRIQ-learning

The Fuzzy Rule Interpolation based Q-learning (FRIQ-learning) is an extension of the traditional fuzzy Q-learning method with the capability of handling sparse fuzzy rule bases. In the followings the FIVE FRI embedded FRIQ-learning (originally introduced in [22]) will be studied in more details.

3.1 The FRI method FIVE

Numerous FRI methods can be found in the literature. A comprehensive overview of the recent methods is presented in [2]. FIVE is one of the various existing techniques.

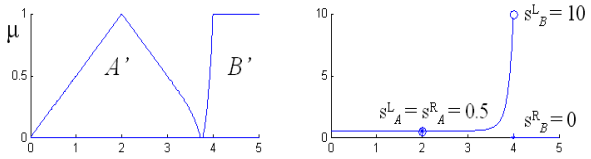

Figure 1

Approximate scaling function *s* generated by non-linear interpolation (on the right). On the left hand side the partition is shown as the approximate scaling function describes it (A', B').

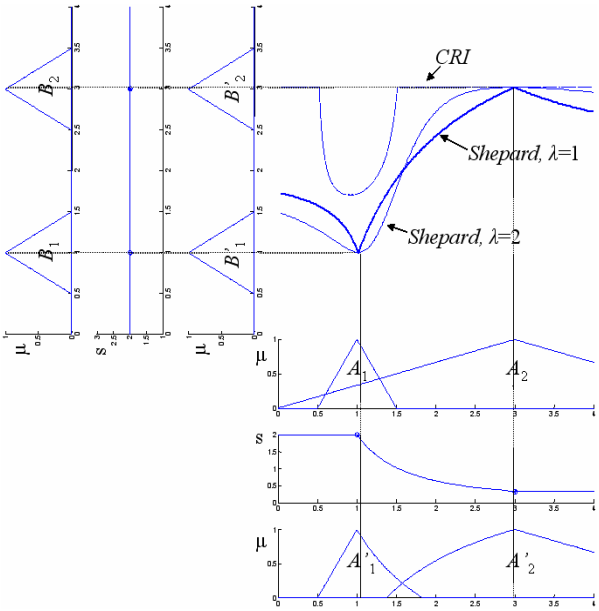

Figure 2

Interpolation of two fuzzy rules rules (R_i: $A_i \rightarrow B_i$), by the Shepard operator based FIVE, and for comparison the min-max CRI with COG defuzzification.

FIVE is an application oriented FRI method (introduced in [11], [9] and [13]), hence it is fast and serves crisp conclusions directly so there is no need for an

additional defuzzification step in the process. Also FIVE has been already proved to be capable of serving the requirements of practical applications [21].

The main idea of the FIVE is based on the fact that most of the control applications serves crisp observations and requires crisp conclusions from the controller. Adopting the idea of the vague environment (VE) [8], FIVE can handle the antecedent and consequent fuzzy partitions of the fuzzy rule base by scaling functions [8] and therefore turn the fuzzy interpolation to crisp interpolation. The idea of a VE is based on the similarity (in other words: indistinguishability) of the considered elements. In VE the fuzzy membership function $\mu_A(x)$ is indicating the level of similarity of x to a specific element a that is a representative or prototypical element of the fuzzy set $\mu_A(x)$, or, equivalently, as the degree to which x is indistinguishable from a [8]. Therefore the α-cuts of the fuzzy set $\mu_A(x)$ are the sets which contain the elements that are (1-α)-indistinguishable from a. Two values in a VE are ε-distinguishable if their distance is greater than ε. The distances in a VE are weighted distances. The weighting factor or function is called scaling function (factor) [8]. If a VE of a fuzzy partition (the scaling function or at least the approximate scaling function [11], [13]) exists, the member sets of the fuzzy partition can be characterized by points in that VE (see e.g. scaling function s on Figure 1). This way any crisp interpolation, extrapolation, or regression method can be adapted very simply for FRI [11], [13]. FIVE integrates the Shepard operator based interpolation (first introduced in [17]) method (see e.g. Figure 2.) because of its simple multidimensional applicability. Precalculating and caching of the consequent and antecedent sides of the vague environment is straightforward, speeding up the method considerably.

The source code of the FIVE FRI along with other FRI methods is freely available as a MATLAB FRI Toolbox [7]. These can be downloaded from [24] and [25] for free of charge.

3.2 FRIQ-learning based on FIVE

The introduction of FIVE FRI in FQ-learning allows the omission of fuzzy rules (i.e. action-state values in this case) from the rule base and also gains the potentiality of applying the proposed method in higher state dimensions with a reduced rule-base sized describing the action-state space. An example for effective rule base reduction by FRI FIVE is introduced in [18].

Substituting the 0-order Takagi-Sugeno fuzzy model of the FQ-learning with the FIVE FRI turns the FQ-learning to FRIQ-learning [22].

The FIVE FRI based fuzzy model in case of singleton rule consequents [10] can be expressed by the following formula:

$$\widetilde{Q}(s,a)=\begin{cases} q_{i_1 i_2 \ldots i_N u} & \text{if } \mathbf{x}=\mathbf{a}_k \text{ for some } k, \\ \left(\sum_{k=1}^{r} q_{i_1 i_2 \ldots i_N u}/\delta_{s,k}^{\lambda}\right)\Big/\left(\sum_{k=1}^{r} 1/\delta_{s,k}^{\lambda}\right) & \text{otherwise.} \end{cases} \tag{6}$$

where the fuzzy rules R_k have the form:

$$\textbf{If } x_1 = A_{k,1} \textbf{ And } x_2 = A_{k,2} \textbf{ And } \ldots \textbf{ And } x_m = A_{k,m} \textbf{ Then } y = c_k, \tag{7}$$

$\delta_{s,k}$ is the scaled distance:

$$\delta_{s,k} = \delta_s(\mathbf{a}_k, \mathbf{x}) = \left[\sum_{i=1}^{m}\left(\int_{a_{k,i}}^{x_i} s_{X_i}(x_i)\,dx_i\right)^2\right]^{1/2}, \tag{8}$$

and s_{X_i} is the i^{th} scaling function of the m dimensional antecedent universe, $\mathbf{x}$ is the m dimensional crisp observation and $\mathbf{a}_k$ are the cores of the m dimensional fuzzy rule antecedents $\boldsymbol{A}_k$.

The application of the FIVE FRI method with singleton rule consequents (6) to be the model of the state-action-value function results in the following:

$$\widetilde{Q}(s,a)=\begin{cases} q_{i_1 i_2 \ldots i_N u} & \text{if } \mathbf{x}=\mathbf{a}_k \text{ for some } k, \\ \sum_{I_1, I_2, \ldots, I_N, U} \prod_{n=1}^{N}\left(1/\delta_{s,k}^{\lambda}\right)\Big/\left(\sum_{k=1}^{r} 1/\delta_{s,k}^{\lambda}\right)\cdot q_{i_1 i_2 \ldots i_N u} & \text{otherwise} \end{cases} \tag{9}$$

where $\widetilde{Q}(s, a)$ is the approximated state-action-value function.

The partial derivative of the model consequent $\widetilde{Q}(s, a)$ with respect to the fuzzy rule consequents $q_{u,i}$, required for the applied fuzzy Q-learning method (5) in case of the FIVE FRI model from (9) can be expressed by the formula above (according to [14]):

$$\frac{\partial \widetilde{Q}(s,a)}{\partial q_{i_1 i_2 \ldots i_N u}} = \begin{cases} 1 & \text{if } x = a_k \text{ for some } k, \\ \left(1/\delta_{s,k}^{\lambda}\right)\Big/\left(\sum_{k=1}^{r} 1/\delta_{s,k}^{\lambda}\right) & \text{otherwise.} \end{cases} \tag{10}$$

where $q_{u,i}$ is the constant rule consequent of the k^{th} fuzzy rule, $\delta_{s,k}$ is the scaled distance in the vague environment of the observation, and the k^{th} fuzzy rule antecedent, λ is a parameter of Shepard interpolation (in case of the stable multidimensional extension of the Shepard interpolation it equals to the number of antecedents according to [20]), x is the actual observation, r is the number of the rules.

Replacing the partial derivative of the conclusion of the 0-order Takagi-Sugeno fuzzy inference (5) with the partial derivative of the conclusion of FIVE (10) with respect to the fuzzy rule consequents $q_{u,i}$ leads to the following equation for the Q-Learning action-value-function iteration:

$$\begin{aligned}
&\text{if } \mathbf{x}=\mathbf{a}_k \text{ for some } k: \\
&q_{i_1 i_2 \ldots i_N u}^{k+1} = q_{i_1 i_2 \ldots i_N u}^{k} + \Delta\widetilde{Q}_{i,u}^{k+1} = \\
&= q_{i_1 i_2 \ldots i_N u}^{k} + \alpha_{i,u}^{k} \cdot \left(g_{i,u,j} + \gamma \cdot \max_{v \in U} \widetilde{Q}_{j,v}^{k+1} - \widetilde{Q}_{i,u}^{k} \right) \\
&\text{otherwise} \\
&q_{i_1 i_2 \ldots i_N u}^{k+1} = q_{i_1 i_2 \ldots i_N u}^{k} + \prod_{n=1}^{N} \left(1/\delta_{s,k}^{\lambda}\right) \Big/ \left(\sum_{k=1}^{r} 1/\delta_{s,k}^{\lambda} \right) \cdot \Delta\widetilde{Q}_{i,u}^{k+1} = \\
&= q_{i_1 i_2 \ldots i_N u}^{k} + \prod_{n=1}^{N} \left(1/\delta_{s,k}^{\lambda}\right) \Big/ \left(\sum_{k=1}^{r} 1/\delta_{s,k}^{\lambda} \right) \cdot \alpha_{i,u}^{k} \cdot \left(g_{i,u,j} + \gamma \cdot \max_{v \in U} \widetilde{Q}_{j,v}^{k+1} - \widetilde{Q}_{i,u}^{k} \right)
\end{aligned} \tag{11}$$

where $q_{i_1 i_2 \ldots i_N u}^{k+1}$ is the $k+1$ iteration of the singleton conclusion of the $i_1 i_2 \ldots i_N u^{\text{th}}$ fuzzy rule taking action A_u in state S_i, S_j is the new observed state, $g_{i,u,j}$ is the observed reward completing the $S_i \rightarrow S_j$ state-transition, γ is the discount factor and $\alpha_{i,u}^{k} \in [0,1]$ is the step size parameter.

As in the previous chapter the $\widetilde{Q}_{j,v}^{k+1}$ and $\widetilde{Q}_{i,u}^{k}$ action-values can be approximated by equation (11). This way the FIVE FRI model is used for the approximation of the mentioned action-value function.

In multidimensional cases to slightly reduce the computational needs it is a good practice to omit updates on rules which have a distance (d_r) considered far away from the actual observation (for example a predefined limit $\varepsilon_d : d_r > \varepsilon_d$) in the state-action space.

4 Reducing Rule Base Size by Incremental Creation

For achieving the reduction of the fuzzy rule base size an incremental rule base creation is suggested. This method simply increases the number of the fuzzy rules by inserting new rules in the required positions (for an example see Figure 3.). Instead of building up a full rule base with the conclusions of the rules (q values) set to a default value, initially only a minimal sized rule base is created with 2^{N+1} fuzzy rules at the corners of the $N+1$ dimensional antecedent (state-action space) hypercube. Similarly like creating Ruspini partitions with two triangular shaped fuzzy sets in all the antecedent universes (see Figure 3/a). In cases when the

action-value function update (11) is high (e.g. greater than a preset limit ε_Q: $\Delta\widetilde{Q} > \varepsilon_Q$), and even the closest existing rule to the actual state is farther than a preset limit ε_s, then a new rule is inserted to the closest possible rule position (see Figure 3/a). The possible rule positions are gained by inserting a new state among the existing ones ($s_{k+1}=s_k$, $\forall k > i$, $s_{i+1}=\frac{s_i+s_{i+2}}{2}$, see e.g. on Figure 3/b.). In case if the update value is relatively low ($\Delta\widetilde{Q} \le \varepsilon_Q$), or the actual state-action point is in the vicinity of the already existing fuzzy rule, than the rule base remains unchanged. The next step is the value update done regarding to the FRIQ-Learning method according to the equation (11), as it was discussed earlier.

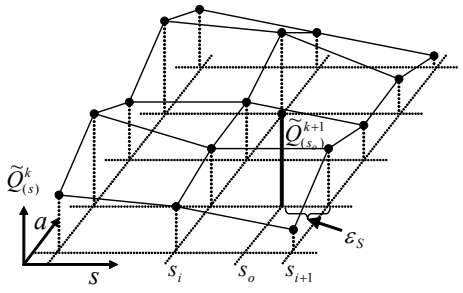

Figure 3/a

The next approximation of Q at s_o: $\widetilde{Q}^{k+1}_{(s_o)}$

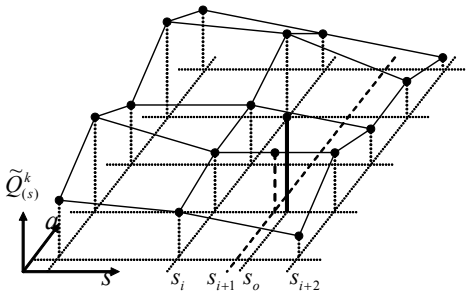

Figure 3/b

A new fuzzy rule is inserted at s_{i+1}.

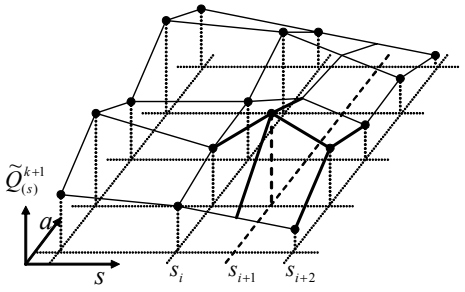

Figure 3/c

Next approximation, with a new rule inserted, and value updated according to (11)

This way the resulting action-value function will be modeled by a sparse rule base which contains only the fuzzy rules which seem to be most relevant in the model. Applying the FIVE FRI method, as stated earlier, allows the usage of sparse rule bases which could result in saving a considerable amount of computational resources and reduced state space.

Conclusions

By introducing the adaptation of the FIVE FRI method in Q-learning, continuous spaces can be applied instead of the originally discrete state-action spaces. Having continuous spaces can lead to better resolutions providing more precise description of the state-action pairs in Q-learning. The targeted reduced rule base size is achieved by incremental creation of an intentionally sparse fuzzy rule base. The fuzzy rule base is incrementally built up from a scratch and will contain only the rules which seem to be most relevant in the model. This way the real advantages of the proposed FIVE based FRIQ-learning method could be exploited: reducing the size of the fuzzy rule base has the benefits not only in decreasing the computing resource requirements, but having less rules (optimizable parameters), it also speeds up the convergence of the FRIQ-learning.

Acknowledgements

This research was partly supported by the Hungarian National Scientific Research Fund grant no: OTKA K77809 and the Intelligent Integrated Systems Japanese Hungarian Laboratory.

References

[1] Appl, M.: Model-based Reinforcement Learning in Continuous Environments. Ph.D. thesis, Technical University of München, München, Germany, dissertation.de, Verlag im Internet (2000)

[2] P. Baranyi, L. T. Kóczy, Gedeon, T. D., "A Generalized Concept for Fuzzy Rule Interpolation", IEEE Trans. on Fuzzy Systems, Vol. 12, No. 6, 2004, pp 820-837

[3] Bellman, R. E.: Dynamic Programming. Princeton University Press, Princeton, NJ (1957)

[4] Berenji, H.R.: Fuzzy Q-Learning for Generalization of Reinforcement Learning. Proc. of the 5th IEEE International Conference on Fuzzy Systems (1996) pp 2208-2214

[5] Bonarini, A.: Delayed Reinforcement, Fuzzy Q-Learning and Fuzzy Logic Controllers. In Herrera, F., Verdegay, J. L. (Eds.) Genetic Algorithms and Soft Computing, (Studies in Fuzziness, 8), Physica-Verlag, Berlin, D, (1996) pp 447-466

[6] Horiuchi, T., Fujino, A., Katai, O., Sawaragi, T.: Fuzzy Interpolation-Based Q-learning with Continuous States and Actions. Proc. of the 5th IEEE International Conference on Fuzzy Systems, Vol.1 (1996) pp 594-600

[7] Zs. Cs. Johanyák, D. Tikk, Sz. Kovács, K. W. Wong: Fuzzy Rule Interpolation Matlab Toolbox – FRI Toolbox, Proc. of the IEEE World Congress on Computational Intelligence (WCCI'06), 15th Int. Conf. on Fuzzy Systems (FUZZ-IEEE'06), July 16-21, Vancouver, BC, Canada, Omnipress. ISBN 0-7803-9489-5, 2006, pp. 1427-1433

[8] F. Klawonn, "Fuzzy Sets and Vague Environments", Fuzzy Sets and Systems, 66, 1994, pp. 207-221

[9] Sz. Kovács, and L.T. Kóczy, "Approximate Fuzzy Reasoning Based on Interpolation in the Vague Environment of the Fuzzy Rule base as a Practical Alternative of the Classical CRI", Proceedings of the 7th International Fuzzy Systems Association World Congress, Prague, Czech Republic, 1997, 144-149

[10] Kovács, Sz.: Extending the Fuzzy Rule Interpolation "FIVE" by Fuzzy Observation, Advances in Soft Computing, Computational Intelligence, Theory and Applications, Bernd Reusch (Ed.), Springer Germany, ISBN 3-540-34780-1, pp. 485-497, (2006)

[11] Sz. Kovács, "New Aspects of Interpolative Reasoning", Proceedings of the 6th. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Granada, Spain, 1996, pp. 477-482

[12] Sz. Kovács: SVD Reduction in Continuos Environment Reinforcement Learning, Lecture Notes in Computer Science, Vol. 2206, Computational Intelligence, Theory and Applications, Bernard Reusch (Ed.), Springer, ISBN 3-540-42732-5, pp.719-738, Germany, (2001)

[13] Sz. Kovács, and L.T. Kóczy, "The use of the concept of vague environment in approximate fuzzy reasoning", Fuzzy Set Theory and Applications, Tatra Mountains Mathematical Publications, Mathematical Institute Slovak

Academy of Sciences, Bratislava, Slovak Republic, vol.12, 1997, pp. 169-181

[14] Krizsán, Z., Kovács, Sz.: Gradient based parameter optimisation of FRI "FIVE", Proceedings of the 9th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics, Budapest, Hungary, November 6-8, ISBN 978-963-7154-82-9, pp. 531-538, (2008)

[15] José Antonio Martin H., Javier De Lope: A Distributed Reinforcement Learning Architecture for Multi-Link Robots. 4th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2007), 2007

[16] Rummery, G. A., Niranjan, M.: On-line Q-learning using connectionist systems. CUED/F-INFENG/TR 166, Cambridge University, UK. (1994)

[17] D. Shepard, "A two dimensional interpolation function for irregularly spaced data", Proc. 23rd ACM Internat. Conf., 1968, pp. 517-524

[18] Sz. Kovács: Interpolative Fuzzy Reasoning in Behaviour-based Control, Advances in Soft Computing, Vol. 2, Computational Intelligence, Theory and Applications, Bernd Reusch (Ed.), Springer, Germany, ISBN 3-540-22807-1, pp.159-170, (2005)

[19] Sutton, R. S., Barto, A. G.: Reinforcement Learning: An Introduction, MIT Press, Cambridge (1998)

[20] D. Tikk, I. Joó, L. T. Kóczy, P. Várlaki, B. Moser, and T. D. Gedeon (2002). Stability of interpolative fuzzy KH-controllers. Fuzzy Sets and Systems, (125) 1, 105-119

[21] Vincze, D., Kovács, Sz.: Using Fuzzy Rule Interpolation-based Automata for Controlling Navigation and Collision Avoidance Behaviour of a Robot, IEEE 6th International Conference on Computational Cybernetics, Stara Lesná, Slovakia, November 27-29, ISBN: 978-1-4244-2875-5, pp. 79-84, (2008)

[22] Vincze, D., Kovács, Sz.: Fuzzy Rule Interpolation-based Q-learning, SACI 2009, 5th International Symposium on Applied Computational Intelligence and Informatics, Timisoara, Romania, May 28-29, 2009, ISBN: 978-1-4244-4478-6, pp. 55-59, (2009)

[23] Watkins, C. J. C. H.: Learning from Delayed Rewards. Ph.D. thesis, Cambridge University, Cambridge, England (1989)

[24] The FRI Toolbox is available at: http://fri.gamf.hu/

[25] Some FRI applications are available at:
http://www.iit.uni-miskolc.hu/~szkovacs/
http://www.iit.uni-miskolc.hu/~vinczed/