Reactive power planning in Distribution Systems using A Reinforcement learning method

Mehdi Ahrari Nouri, Amir Hesami and Alireza Seifi
Department of Electric Engineering, Shiraz University,
Sama Organization (Affiliated with Islamic Azad University) – Shiraz Branch, Shiraz, Iran
m_ahrarinouri@yahoo.com, sa_hesami@yahoo.com, seifi@shirazu.ac.ir

Abstract: This work presents a new reinforcement learning (RL) algorithm for capacitor allocation in distribution feeders. The problem formulation considers two distinct objectives related to total cost of power loss and total cost of capacitors including the purchase and installation costs. The formulation is a multi-objective and non-differentiable optimization problem. The proposed method of this article uses RL procedure for sizing and siting of capacitors in radial distribution feeders. The proposed method has been implemented in a software package and its effectiveness has been verified through a 9-bus radial distribution feeder and also a 34-bus radial distribution feeder for the sake of conclusions supports. A comparison has been made between the proposed method of this paper and similar methods in other research works to show its effectiveness for solving optimum capacitor planning problem.

Keywords: Reactive Power Planning, Reinforcement Learning, Radial Distribution Feeder

I. *Introduction*

Capacitors are widely installed in distribution systems for reactive power compensation to achieve power and energy loss reduction, reduce voltage regulation, system capacity release, power flow control, improving system stability and power factor correction. Capacitor planning must determine the optimal site and size of capacitors to be installed on the buses of a radial distribution system. Many approaches have been proposed to solve the capacitor planning problem. For instance, [1] formulated the problem as a mixed integer programming problem that incorporated power flows and voltage constraints. The problem was decomposed into a master - slave problem to determine the siting of the capacitors, as well as the types and the size of the capacitors placed on the system. Refs [2,3] proposed heuristic approaches to identify the sensitive nodes by the levels of effect on the system losses. Ref. [4] adopted an equivalent circuit of a lateral branch to simplify the distribution loss analysis, which obtained the capacitor operational strategies according to the reactive load duration curve and sensitivity index. Moreover, optimal capacitor planning based on the fuzzy algorithm was implemented to present the imprecise nature of its parameters or solutions in practical distribution systems [5-7]. Several investigations have recently applied artificial intelligence (AI) techniques to resolve the optimal capacitor planning problem due to the growing popularity of AI. Refs. [8,9] presented a solution methodology based on a simulated annealing (SA) technique. Ref. [10] applied the tabu search (TS) technique to determine the optimal capacitor planning in Chiang et al's [8] distribution system, and compared the results of the TS with the SA. In Refs.[11-12], genetic algorithms (GA) were implemented to obtain the optimal selection of capacitors, but the objective function only considered the capacitor cost and power losses without involving operation constraints.

The capacitor planning problem is formulated as a multiple objective problem. The formulation proposed here considers two distinct objectives related to total cost of power loss and total cost of capacitors and also considers load flow restrictions security and operational constraints like loading of feeders and voltage profile maximum reactive compensation.

The rest of this article is organized as follows: Section II describes a novel formulation of the capacitor planning problem. A solution algorithm based on the RL for the multi-objective problems is developed in section III. Section IV demonstrates the effectiveness of the solution algorithm on two distribution case study. Conclusions are finally made in section V.

II. *Mathematical Model of the Problem*

The objective function in the capacitor planning problem for radial distribution feeders is:

$$Minimize\{(K_P \times (\sum_{i=0}^{N} P^k_{loss(i,i+1)}) + (\sum_{k=1}^{N_C} (C^{Q_i}{}_{inst} + C^{Q_i}{}_{purc}))\} \quad (1)$$

Such that:

$$P_{gi} - P_{di} - V_i \sum_{j=1}^{N} V_j Y_{ij} \cos(\delta_i - \delta_j - \theta_{ij}) = 0 \quad (2)$$

$$Q_{gi} - Q_{di} - V_i \sum_{j=1}^{N} V_j Y_{ij} \sin(\delta_i - \delta_j - \theta_{ij}) = 0 \quad (3)$$

$$V_i^{\min} \le V_i \le V_i^{\max} \qquad i=1 \ldots N \quad (4)$$

$$P_{ij}^{\min} \le P_{ij} \le P_{ij}^{\max} \quad i=1...N \tag{5}$$

$$Q_C^{Total} \le Q_L^{Total} \tag{6}$$

Where:
K_p, Cost per power loss, $/kW/year
N, Total Number of buses in radial distribution network

$P_{loss(i,i+1)}$, Active power loss of (i,i+1) branch
N_C, Total number of possible capacitor sizes
$C^{Q_i}{}_{inst}$, The cost of installation of a capacitor bank of Q (Var) on bus i
$C^{Q_i}{}_{purc}$, The cost of purchasing of a capacitor bank of Q (Var) for bus i
P_{gi}, Q_{gi}, Active and reactive power generations at bus i.
P_{di}, Q_{di}, Active and reactive power load at bus i.
V, δ, System bus voltages magnitudes and phase angles.
Y_{ij}, θ_{ij}, Bus admittance matrix elements
Q_C^{Total}, Total connected Var by capacitor banks for radial distribution network
Q_L^{Total}, Total Var of connected loads in radial distribution network

This objective function considered here in (1), consists of two terms. The first term denotes the cost of power loss and the second term includes the total cost of capacitors that consist of the purchase and installation costs.
Regarding the constraints, (2) and (3) present the well-known load flow restrictions while security and operational constraints like voltage profile and loading of feeders have been formulated in inequality of (4) and (5).
As a general rule, for reactive-power compensation, the maximum capacitor size should not exceed the connected reactive load. This results having a limited number of available capacitor sizes for installing on the radial distribution network. This concept has been formulated by (6) in the set of constraints of introduced objective function.

III. *Reinforcement Learning*

Reinforcement learning is defined by Kaelbling, Littman and Moore (1996) as 'the problem faced by an agent that must learn behavior through trial-and-error interactions with a dynamic environment'. Mathematically, the reinforcement learning problem has been formalized as a Markov Decision Process (a process where the probability of the agent moving from one state to another, given its choice of action, is independent of the history of the system prior to reaching that state). The mathematics of Markov processes has been extensively studied, one significant result, Bellman (1957), showed that an algorithm based on dynamic programming can be shown to converge to an optimal policy if the Markov process is stationary (a stationary Markov process is one in which the state transition probabilities, given the agent's choice of action, do not change over time). In the standard reinforcement-learning model, an agent is connected to its environment via perception and action. On each step of interaction the agent receives as input, i, some indication of the current state, s, of the environment; the agent then chooses an action, a, to generate as output. The action changes the state of the environment. The value of this state transition is communicated to the agent through a scalar reinforcement signal [13].
Formally, a RL problem consists of:

- A discrete set of environment states, S;
- A discrete set of agent actions, A;
- A set of scalar reinforcement signals, R;
- Policy π which chooses the actions that has to be taken.
- Value Function, which maps each state to a measure of the expected discounted future reward that agent will receive by the following policy π.

In a RL problem the agent's goal is to maximize the reward it receives in the long run. In general, we seek to maximize the expected return, where the return R_t is defined as some specific function of the reward sequence. In the simplest case the return is the sum of rewards:

$$R_t = r_{t+1} + r_{t+2} + ... + r_T \tag{7}$$

That T is the final time step. We have this notion of final time step when the agent – environment interaction breaks naturally into subsequences called *Episodes*. We can use discount factor $\gamma, (0 \le \gamma \le 1)$ in the Equation (8), so we have:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{8}$$

Almost all reinforcement learning algorithms are based on estimating value functions which value functions are functions of states that estimates how good it is for the agent to be in a given state. We have the policy π which is mapping from each state $s \in S$, and action $a \in A$, to the probability
$p(s,a)$ of taking action a when in state s. The value of a state s under a policy π, denoted $V^{\pi}(s)$, is the expected return when starting in s and following π thereafter.

$$V^{\pi}(s) = E_{\pi}\{R_t \Big| s_t = s\} = E\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\} \tag{9}$$

A. *Q – Learning*

One of the most important breakthroughs in reinforcement learning was development of an off-policy temporal-difference (TD) control algorithm known as Q-learning

(Watkins, 1989). Its simplest form, one–step Q–Learning, is defined by:

$$Q(s_t,a_t)=Q(s_t,a_t)+\alpha[r_{t+1}+\gamma\max Q(s_{t+1},a_t)-Q(s_t,a_t)] \qquad (10)$$

It is important to note that the new value for $Q(s_t,a_t)$ memory is based both on the current value of $Q(s_t,a_t)$, and the values of immediate rewards obtained by next searches (r_{t+1}). So, the α parameter plays a critical role representing the amount of the updated Q-memory (10) and affects the number of iterations.

This is identical to Sarsa learning except that when considering the next state action transition, the action is chosen that will maximize the next Q-value. Q-learning is shown to converge to an optimal policy under the usual assumptions (Watkins and Dayan, 1992), and it remains the most popular reinforcement learning algorithm because no model of the environment is required, it is intuitive, easy to implement, and can be run interactively with updates made immediately, as and when states are visited [14].

IV. *Problem Formulation and Implementation*

For the purpose of our analysis, the Q-learning algorithm is the "agent", the state vectors are the number of buses available for capacitor placing, and the action vector are the discrete values of possible capacitors [15]. The algorithm proceeds as follows:

The agent observes the initial state (*s*) of the system, as obtained by the load flow solution, and chooses one action (*a*) from the action vector. A new load flow is executed. The agent observes the resulting state of the solution and the immediate reward expressing the degree of satisfaction of the operating limits of the constrained variables. A new action is selected next, leading to a new load flow solution and a new reward. Selection of new control actions is repeated until the voltage constraint is satisfied. The goal of the agent is to learn the optimal Q-function using the mappings of states to actions such that the long-term reward is maximized. The procedure is repeated for a large number of operating states covering the whole planning period. The agent finds the set of actions that leads to a greedy-optimal policy.

A. Rewards

Application of the Q-learning algorithm to reactive power planning problem is linked to the choice of an immediate reward (*r*), such that the iterative value of Q-function (10) is maximized for the whole planning period. The reward function is calculated by (11) as follow:

$$r=\frac{1}{K_P\times(\sum_{i=0}^{N}P_{loss(i,i+1)})}+\frac{1}{\sum_{k=1}^{N_C}(C^{Q_i}{}_{inst}+C^{Q_i}{}_{purc})} \qquad (11)$$

The first term in (11) is the reverse of power losses cost and the second term is the reverse of capacitors cost that are used. So as the agent maximizes the reward function it will follow our goal and minimizes the objective function (1).

B. Simulation Results

The proposed algorithm is applied to two distribution systems. For the two test feeders in case study, yearly loss cost is selected to be U.S. $ 168/kW [16], and the voltage limits are 0.9p.u and 1.1p.u. The sizes and cost of available capacitors are according to [16]. In all test cases it is assumed that all buses are available for placement of capacitors.

In addition to the above definitions, two parameters for implementing the Q-learning algorithm need to be chosen. Parameter γ is the control factor by which later rewards are discounted and it must be between 0 and 1. In our application, later rewards are not important because there is no interdependence among load flow solutions produced by Q-learning steps; therefore, the value of γ is chosen close to 0 ($\gamma = 0.005$).

The critical parameter α, expresses the amount of the updated Q-function. A large enough parameter (close to 1) allows fast convergence of the Q-learning algorithm, while a small value (close to 0) avoids instability of Q-learning. Since the Q-learning enforced in constrained load flow problem does not depend on previous Q-learning steps as stated above, this parameter works well close to 1. In our application, a value of 0.995 is selected.

C. Case Study 1: 9-bus system

The 9-bus radial distribution feeder of [16] is taken as the test feeder. The rated voltage is 23 kV. The total reactive load of the system is 4186kVar that leads to 27 practical combinations of mentioned standard capacitor banks so the action vector has 28 members (including member zero) and state vector has 9 members.

Applying the load flow program on this feeder before compensation, the cost function and the total power losses are U.S. $ 131675 and 783.8 kW, respectively. The maximum and minimum bus voltage magnitudes were 0.9929p.u and 0.8375p.u respectively, where the voltage of the substation (bus number 0) is assumed to be 1p.u.

Table 1 shown the results of capacitor planning in [16].The methods 1-5 and the exact solutions described in [16] and method 6 is our purposed RL method compared to them. The reward of RL method in 9 bus feeder for 20000 iterations is shown in Fig 1.

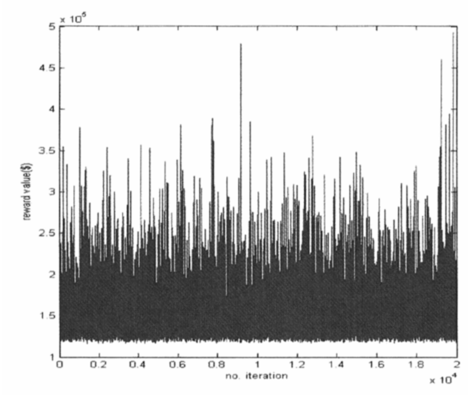

Fig.1. The reward of RL method in 9-bus test feeder

Table 1 Results for all methods applied to the 9-bus feeder including original data in [16] and RL Method

Bus No.	No Q_C (KVAr) Placed	Qc(KVAr) Using Method 1	Qc(KVAr) Using Method 2	Qc(KVAr) Using Method 3	Qc(KVAr) Using Method 4	Qc(KVAr) Using Method 5	Exact Solution	Qc Using RL Method 6
1								450
2				3300			3600	4050
3			1050	3900	3300	2850		300
4		2100	1050		1800	2100	4050	3750
5		2500	1950	1200	1050	1050	1650	
6								600
7								600
8							600	
9		900	900	900	900	900		450
Real Loss(KW)	783.8	707	705	689	692	691.6	686	680.18
$ Cost	131675	119736	119420	117330	117571	117479	117095	116275
Min V (p.u.)	0.8375	0.9000	0.9029	0.9006	0.90004	0.9000	0.9003	0.9004
Max V (p.u.)	0.9929	1.0000	1.0000	1.006	1.0012	1.001	1.007	1.0068

D. Case Study 2: 34-bus system

A radial distribution network with 34 load points is used to simulate the proposed RL Method. The data of this test system has been taken from [16].The system voltage is 11 kV. Before compensation, the cost is U.S. $ 37212, this is based on the previously defined cost function, the active and reactive losses are 221.5 kW and 65.04 KVAr, respectively, and the voltage limits in per unit are 0.9417 and 1.0. Considering total connected reactive load of 2873.5 KVAr of this system, 19 capacitor bank combination can be used.

The result of RL method is illustrated in table 2, while the comparison between results of the methods 1-5 of [16] and methods 6 are presented in Figs. 2&3.

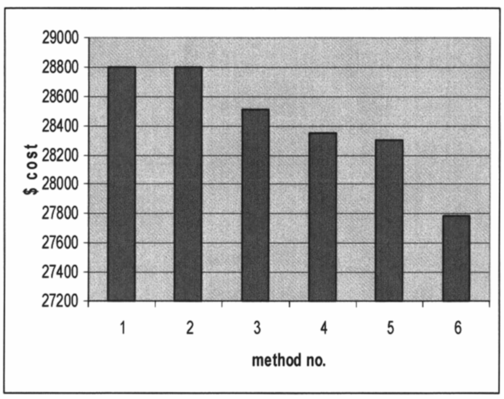

Fig.2. Cost function for the 6 methods applied to 34-bus test feeder

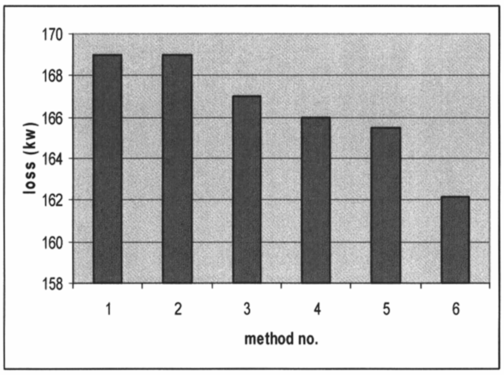

Fig.3. Active power losses for the 6 methods applied to 34-bus test feeder.

V. *Conclusions*

This article presents a new combined optimization method for optimum capacitor planning problem. The proposed method uses reinforcement learning method for siting and sizing of capacitors in distribution system.

The method developed herein is tested on 9-bus distribution system and the results have been compared to similar research works. The comparison shows the effectiveness of proposed method in case of investment and improving the performance of the distribution network. In addition, to verify the practical feasibility and performance of the proposed algorithm for practical cases, a 34-bus test system has been used.

VI. References

[1] M.E. Baran, F.F. Wu, optimal capacitor placement in distribution systems, IEEE Trans. Power Delivery, pp 725-734, 1989

[2] T.S. Abdel-Salam, A.Y. Chikhani, R. Hackam, A new technique for loss reducing using compensating capacitors applied to distribution systems with varying load condition, IEEE Trans. Power Delivery ,pp 819-827, 1994

[3] M Chis, M.M.A Salama, S.Jayaran, capacitor placement in distribution systems using heuristic search strategies, IEE Proc. Generation Transmission & Distribution, pp 225-230, 1997

[4] MY. Cho, YW. Chen, Fixed/switched type shunt capacitor planning of distribution systems by considering customer load patterns and simplified feeder model, IEE Proc Generation Transmission & Distribution, pp 533-540,1997

[5] HN. Ng, M.M.A. Salama, Fuzzy optimal capacitor sizing and placement, Proceedings of the Canadian conference on electrical and computer engineering, pp 684-687, 1995

[6] C.T. Su, C.C. Tsai, A new fuzzy-reasoning approach to optimum capacitor allocation for primary distribution systems, Proceedings of the IEEE international conference on industrial technology, December, pp 237-241, 1996

[7] K.H. Abdul-Rahman, S.M. Shahidepour, A fuzzy-based optimal reactive power control, IEEE Trans. On Power Systems, pp 662-670, 1993

[8] H.D. Chiang, J.C. Wang, O. Cocking, Optimal capacitor placements in distribution systems: part 1: a new formulation and the overall problem, IEEE Trans. On Power Delivery, pp 634-642, 1990

[9] H.D. Chiang, J.C. Wang, O. Cocking, Optimal capacitor placements in distribution systems: part 2: solution algorithms and numerical results, IEEE Trans on Power Delivery, pp 643-649, 1990

[10] H.T. Yang, Y.C. Huang, C.L. Huang, Solution to capacitor placement problem in radial distribution system using tabu search method, Proceedings of the international conference on energy management and power delivery, pp 388-393, 1995

[11] V. Ajjarapu, Z. Albanna, Application of genetic based algorithms to optimal capacitor placement, Proceedings of the first international forum on applications of neural networks to power systems, pp 251-255, 1991

[12] S. Sundharajan, A. Pahwa, optimal selection of capacitors for radial distribution systems using a genetic algorithm, IEEE Trans. On Power Systems, pp 499-507, 1994

[13] L.P. Kaelbling, M.L. Littman, A.W. Moore, and Reinforcement Learning: A Survey, Journal of Artificial Intelligence Research 4, pp 237-285, 1996.

[14] R. Sutton, A.G. Barto, Reinforcement Learning: An introduction, MIT press, 1998

[15] J.G. Vlachogiannis, N.D. Hatziargyriou, " Reinforcement Learning for Reactive Power Control", IEEE Trans. Power Systems, Vol. 19, No. 3, AUGUST 2004

[16] S.F. Mekhamer, S.A. Soliman, M.A. Moustafa, M.E. El-Hawary, Application of Fuzzy Logic for Reactive-Power Compensation of Radial Distribution Feeders, IEEE Trans. on Power Systems, Vol.18, No.1,February, 2003.