# Trends in Applied Sciences Research

ISSN 1819-3579

Academic
Journals Inc.

www.academicjournals.com

Trends in Applied Sciences Research 7 (3): 210-220, 2012
ISSN 1819-3579 / DOI: 10.3923/tasr.2012.210.220


# A New Reinforcement Learning Optimization Method for Capacitor Allocation Considering Variable Load

Mehdi Ahrari Nouri and Ali Reza Seifi
School of Electrical and Computer Engineering, Shiraz University, Iran

*Corresponding Author: Mehdi Ahrari Nouri, School of Electrical and Computer Engineering, Shiraz University, Iran Tel: +98 711 2303081 Fax: +98 711 6287294*

## ABSTRACT

In distribution systems shunt capacitor banks are widely used for reactive power compensation, power and energy loss reduction and improving voltage profile. In this study by using Reinforcement Learning (RL) approach and heuristic strategies a method for reactive power optimization in distribution systems is presented. The approach is consist of determining values and locations of capacitor banks and also optimal position of tap in an Under Load Tap Changer (ULTC) transformer under voltage and current constraints for total load's curve duration. The optimization problem has to be solved in the way so that the load demand loss and systems energy loss are being minimized. By using double agent Q-Learning a new method for this problem is proposed and the results are compared to other similar researches.

**Key words:** Optimization, capacitor, ULTC, distribution system, Q-learning, variable load

## INTRODUCTION

One of the most common methods for loss reduction in medium voltage systems is reactive power compensation. Because of inductive loads in power systems there will be reactive power in the system as well as the active power. The primary sources for reactive power generation are power stations, so reactive current flow is from station to consumers. Therefore, according to this matter the reactive current has to pass through all the power system's section which will generate loss in the system and occupies bus's and equipment's capacity. The best location for reactive power compensation is the location close to consumers which is the distribution system.

By reactive power compensation we achieve other goals such as energy loss reduction, active power generation in the load's peak demand, system capacity release, power factor correction and improving voltage profile (Abdelaziz *et al*., 2011). One of the most common methods for reactive power compensation is using shunt capacitor banks in the distribution system. In this method, values and locations of the capacitors have to be determined so that the benefit from capacitor allocations would be maximized (Liang and Cheng, 2001; Augugliaro *et al*., 2004). The ULTC with switching capacitors are used for reactive power optimization so that the system operational constraints are satisfied and we use ULTC for the same reason in our proposed method. In many of papers published for reactive power optimization the load profile is consider constant (Azim Swarup, 2005; Bhattacharya and Goswami, 2009), but if the optimization process is done during the load's varying period, the results will be more usable in real applications so in this research the variable load is considered.

The solution technique for loss minimization can be classified into four categories: Analytical (Grainger and Civanler, 1985a-c), numerical programming (Wang *et al*., 1997a, b), heuristic (Abdel-Salam *et al*., 1994; Haque, 1999) and artificial intelligence based (Abdel-Salam *et al*., 1994; Mekhamer *et al*., 2003; Miu *et al*., 1997; Delfanti and Gianpietro, 2000; Adeyemo, 2011; Ahmed and Zamli, 2011; Dehini *et al*., 2012; Saxena *et al*., 2011; Otieno and Adeyemo, 2010; Sutha and Kamaraj, 2008; Yap *et al*., 2011). Reinforcement Learning (RL) recently has been used in different fields of electrical engineering and computer science applications. In reinforcement learning the agent learns by trial and error and exploring the dynamic environment, chooses the optimized action. Trial and error learning is based on animals learning psychological methods (Hagan and Krose, 1997).

In this study, a new optimized solution for the capacitor allocating problem in distribution systems by using RL approach and control of under load transformer ULTC for whole load curve is proposed.

## CAPACITOR ALLOCATION FORMULATION

The algorithm of finding the values and locations of capacitor banks is based on a goal function which is defined as a cost function (cost function means all the costs such as capacitor costs, energy production costs and etc.) then the values and locations of capacitor banks will be chosen to minimize the total cost for the entire load's period. For a constant load level the goal function is defined as:

$$\text{Cost function} = K_p \times P_{Loss} + \sum_{i=1}^{n_c} (C_{ins} + C_{K\,var} . Q_i) \tag{1}$$

And for several loads' levels the goal function is defined as:

$$\text{Cost function} = K_E . \sum_{i=1}^{n_t} P_{i,loss} . T_i + Kp \times P_{peak} + \sum_{i=1}^{n_c} (C_{ins} + C_{Kvar} . Q_i) \tag{2}$$

Where:

$K_p$ = Cost per power loss ($/kw/year)
$K_E$ = Cost of energy loss ($/kwh)
$P_{i, loss}$ = Power loss in the ith load level (Kw)
$T_i$ = The load level Duration of episode I (Hour)
$n_t$ = Number of load's levels during the studying period
$P_{peak}$ = The amount of peak station's generation (MW)
$n_c$ = Number of installed capacitors
$C_{kvar}$ = Cost for each KVAR capacitor bank ($/KVAR/year)
$C_{ins}$ = Capacitor allocation costs ($/KVAR/year)
$Q_i$ = The amount of capacitor bank installed on location I (KVAR)

The length of the study period is one year so we have:

$$\sum_{i=1}^{n_t} T_i = 8760 \tag{3}$$

The Under Load Tap Changer (ULTC) usually installed in a main transformer is employed to adjust the secondary voltage. ULTC is one the Voltage regulators that voltage changes according to the taps on the primary winding. In this study a 10% ULTC with 16 taps at the beginning of distribution feeder has been used during the study period.

In capacitor placement some constraints according to the power quality have to be satisfied. The first constraint is the load flow convergence. For satisfying this constraint the direct load flow is used (Teng, 2003). This algorithm is a classic method which can be use in radial and mesh distribution systems.

For controlling the consumer's power quality each bus's voltage has to be between a minimum and maximum allowed value. So in each capacitor allocation states this constraint should be satisfy by using Eq. 4:

$$V_{min}<V_i<V_{max} \quad I = 1, 2, \ldots, n \tag{4}$$

Where:

$V_{min}$ = Minimum allowed voltage
$V_{max}$ = Maximum allowed voltage
n = Number of busses
Vi = ith bus voltage value

According to the limitation of existing capacitor banks for installation, the maximum compensation constraint could be useful for achieving the optimized solution faster. For example the maximum compensation in each bus according to the Eq. 5 is equal to the total amount of system's reactive load. So during searching the optimized solution there is no need to search the states where compensation in those states is more than the total amount of the reactive load:

$$0<Q_i\leq Q_{max} \tag{5}$$

Where:

$Q_{max}$ = The maximum compensation (total amount of reactive loads of the system)
$Q_i$ = The amount of compensation installed on the ith node

## REINFORCEMENT LEARNING

Reinforcement learning is defined by Kaelbling *et al.* (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 (Sutton and Barto, 1998).

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 find (or learn) a policy π: S→A, mapping states to actions that maximizes the reward it receives in the long run:

$$\pi^* = \arg\max_{\pi} V^{\pi}(s), \forall s \tag{6}$$

where, $V^{\pi}(s)$ is called the value-function for policy π.

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 sϵS and action aϵA, to the probability P(s,a) of taking action a when the state is s:

$$V^{\pi}(s) = E_{\pi}\{R_t \Big| s_t = s\} = E\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \Big| s_t = s\} \tag{7}$$

where, $E_{\pi}\{\}$ denote the expected value given that the agent follows policy π. In general, we seek to maximize the expected return, where the return $R_t$ is defined as some specific function of the reward ($r_i$) sequences. In the simplest case the return is the sum of rewards:

$$R_t = r_{t+1} + r_{t+2} + r_{t+1} + \ldots + r_{TH} \tag{8}$$

where, t denote the time steps and TH is the final time step. We have this notion of final time step when the agent-environment interaction breaks naturally into subsequences called episodes. The additional concept that we need is that of discounting. According to this approach, the agent tries to select actions so that the sum of the discounted rewards it receives over the future is maximized.

We can use discount factor γ, ($0 \le \gamma \le 1$) in the Eq. 9, so we have:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \ldots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{9}$$

The discount rate determines the present value of future rewards: a reward received k time steps in the future, is worth only $\gamma^{k-1}$ times what it would be worth if it was received immediately.

**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).

Its simplest form, one-step Q-Learning, is defined by:

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

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 $\alpha$ parameter plays a critical role representing the amount of the updated Q-memory Eq. 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.

**Actoin-value method:** Unlike the supervised learning methods in RL the environment is explicitly on the trade-off between exploration and exploitation. The agent must learn which actions maximize reward function in the time, but also how to act to reach this maximization, looking for actions still not selected or regions not considered in a state space. The exploration and exploitation processes are usually mixed. Action-value methods are used to estimating the values of actions and for using the estimates to make action selection decisions. The simplest action selection rule is to select the action (or one of the actions) with highest estimated action value, that is, to select on play t one of the greedy actions, a*, for which:

$$Q_t(a^*) = \max_a Q_t(a) \tag{11}$$

$Q_t$ (a) Estimated value of action a at the t play.

This method always exploits current knowledge to maximize immediate reward; it spends no time at all sampling apparently inferior actions to see if they might really be better. A simple alternative is to behave greedily most of the time, but every once in a while, say with small probability $\varepsilon$; instead select an action at random, uniformly, independently of the action-value estimates. We call methods using this near-greedy action selection rule $\varepsilon$-greedy methods. An advantage of this method is that, in the limit as the number of plays increases, every action will be sampled an infinite number of times.

## THE PROPOSED METHOD

For solving the capacitor allocation problem using the reinforcement learning method states and actions, reward function and the optimized solution with a fast convergence are should be determined.

According to the power system's topology, the environment is defined as electric network and the standard capacitor banks values for the system (Mekhamer *et al.*, 2003) and the taps of ULTC are chosen as actions and the buses suitable for capacitor placement, are chosen as states. The policy is based on Q-learning algorithm and $\varepsilon$-greedy action value method is used for choosing the actions.

According to the novel method introduced for one load level capacitor placement (Nouri *et al.*, 2007), three methods have been proposed. In the first method the load's curve is divided into constant levels and after capacitor allocation process for each of them, among all the responses by using Eq. 2, the response with the minimum cost will be chosen as the optimized solution. The second method is similar to the first method except that after capacitor allocation for each load's

level and achieving sufficient results instead of testing all the results in the first step the most appropriate response will be chosen using the Eq. 1 for each load's level and then by using Eq. 2 the best solution will be determined among all the chosen solutions so by using a heuristic procedure a search space will be reduced. In both first and second methods the reward function is chosen by using Eq. 12 so the instant rewards have more effect on the optimal policy. The third method is similar to the second method but discount reward function has been used, where in each load's level the reward is constructed from total amount of power loss's costs, installed capacitor for that level and other level's power loss according to the discount factor (γ). In this way, the effect of each capacitor banks installed is considered for the other load levels during the load changing curve:

$$r_{\substack{i\\t+1}} = \frac{1}{cost(P_{loss})} + \frac{1}{cost(C_{total})} \tag{12}$$

$$r_{\substack{i\\t+1}} = \frac{1}{cost(Pi_{loss})} + \frac{1}{cost(Ci_{total})} + \sum_{\substack{k=1\\k\neq 1}}^{m} \gamma_k \times \frac{1}{cost(p_{\substack{k\\Loss}})} \tag{13}$$

$$\gamma_k = T_k \Big/ \sum_{k=1}^{m} T_k \tag{14}$$

Where:

$r_{\substack{i\\t+1}}$ = The ith load level's reward function

$P_{\substack{i\\Loss}}$ = Power loss in the ith load level (kW)

$C_{\substack{i\\total}}$ = The installed capacitor for the ith load level (KVAR)

m = The number of load's levels

$P_{\substack{k\\Loss}}$ = The power loss in the kth load level (kW)

$T_k$ = The kth load level's duration (Hour)

The algorithm of first method is as below

- Read input data (consist of number of system buses (n) and lines impedances) and number of load levels (m) and complex power of each of them.
- Put I = 1
- While I≤m

  3-1 j = n

  3-2 while j≥1

  3-2-1 state (s) = bus n

  3-2-2 choose greedy action a by using ε-greedy method

  3-2-3 perform load flow and calculate loss power

  3-2-4 Using Eq. 12 and calculate the reward of next episode

  3-2-5 Update Q-function using Eq. 10

  3-2-6 j = j-1

  3-3If the voltage constraint of Eq. 4 is guaranteed for all system buses go next step otherwise return step 3-1

  3-4 Save the results of chosen action, results and load flow

  3-5 I = I+1
- Calculate the cost function Eq. 2
- Print the result with the least cost

The values of two parameters $\gamma$ and $\alpha$ for implementing the Q-learning algorithm need to be chosen. Parameter $\gamma$, is the control factor by which later rewards are discounted and it must be between 0 and 1. In our application, in each load level later rewards are not important because there is no interdependence among load flow solutions, therefore, the value of $\gamma$ is initially set to be zero. The critical parameter $\alpha$ used in Eq. 10, expresses the amount of the updated Q-function, in other words the rate of learning. 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 will work well close to 1.

In our application, initially we set $\gamma = 0$ and $\alpha = 1$, but by these values the agent is so myopic and the effect of future actions will not take into account at all. Therefore by using a dynamic approach $\gamma$ and $\alpha$ are changing slightly between $0 \leq \gamma \leq 0.5$ and $0.5 \leq \alpha \leq 1$. By experiment the best value for $\alpha$ is 0.995 and for $\gamma$ is 0.005.

The value of $\varepsilon$ is chosen equal to 0.1. Small values for $\varepsilon$ prevents the agent from exploiting and choosing new actions and large values prolong the search time and the solution might not converge.

In order to consider the effect of ULTC in the power system after determining the value and position of capacitor banks by one of the methods described above, if the Eq. 4 during the whole load curve is not satisfied another Q-Learning policy is implemented so that the tap positions are new actions and the number of load levels are states. The same $\varepsilon$-greedy method is used for action selection. The optimal position of tap in the ULTC is determined so that all constraints are achieved and the tap movement in the ULTC is minimized too.

In order to considering the effect of load levels change during study period the load curve is discrete into several load levels.

## RESULTS

Two 9-buses and 33-buses systems have been studied in this section. In order to reviewing the energy loss role, the generation cost value is not taking account during the load's peak demand. The cost of generating one kilo watt hour energy ($K_E$) has been assumed equal to 1 \$/Kwh and loss's costs per year equal to $K_p = 168$ (\$/kw). Allowed voltage range is $V_{min} = 0.9$ p.u and $V_{max} = 1.1$ p.u.

Load's changing curve in one day is assumed as Fig. 1 (Taleski and Rajicic, 1996).

The load's changing curve is divided into different load levels and the three mentioned methods described above have been applied to them. In Fig. 2, load's changing curve is divided into 40 levels.

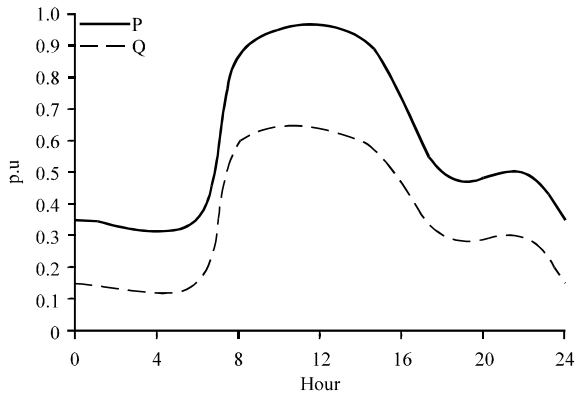


Fig. 1: Load's changing curve in one day (TALESKI, RAJICIC, 1996)

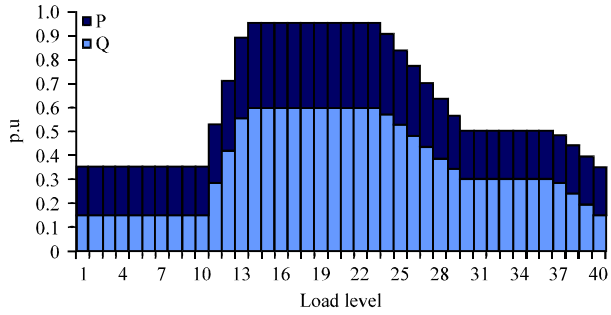


Fig. 2: Load's changing curve is divided into 40 levels

Table 1: Capacitor placement resulting in the 9 buses system for 40 load levels

| | (Kvar) Installed Capacitor | | |
|---|---|---|---|
| Bus number | Method 1 | Method 2 | Method 3 |
| 2 | 0 | 600 | 0 |
| 3 | 600 | 0 | 0 |
| 4 | 600 | 1200 | 1350 |
| 5 | 450 | 900 | 0 |
| 6 | 0 | 0 | 300 |
| 7 | 450 | 0 | 300 |
| Total cost ($) | 2,569,800 | 2,571,900 | 2,571,000 |
| ULTC tap position (p.u) | 1.0267 | 1.0267 | 1.0267 |
| Total cost decrease (%) | 4.4508 | 4.3713 | 4.4034 |

**9-buses system:** A 9-buses system has been used as the first case study (Chin and Lin, 1994). The systems nominal voltage is 23 Kv and system's total reactive load is 4186 Kvar. So, we can use 27 combinations of standards capacitor banks (Mekhamer *et al*., 2003) where there will be 9 states and 28 actions (the case of no capacitor that means no action is taken is added too). According to 40 levels division of the load's curve the energy loss's cost before capacitor allocation is equal to 2'689'500 $. In Table 1, the results of capacitor placement and comparison among three methods are shown.

It is clear the first method has the best cost decrease and the second position is belongs to method two. The position of ULTC tap shows that the voltage should by multiply by 1.0267 in order to Eq. 4 satisfied during load changing curve.

**33-buses systems:** The second case study is a 33-buses system according to Baran and Wu (1989). The systems nominal voltage is 12.66 Kv and system's total reactive load is 2300 Kvar. The energy loss's cost before capacitor allocation in spite of 40 load's levels is 605'720 $. The schematic of this system is illustrated in Fig. 3. In this system there is no need for using the ULTC and the tap changer's position at the beginning of the line is set to 1 per unit. The first method is steel has the best results so it is chosen as the optimal method (Table 2).

**Comparison:** The proposed method one is selected to compared with the method used for capacitor allocation by Galego *et al*. (2001) for continuous load curve. The results by Galego *et al*. (2001) for a 9-buses system with 3 load's levels (1000 h S = 1.1 kVA, 6760 h S = 0.6 KVA and 1000 h S = 0.3

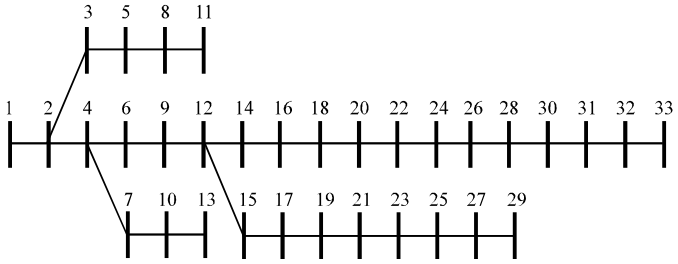


Fig. 3: The 33-buses system

Table 2: Capacitor placement resulting in the 33-buses system for 40 load levels

| Method | Bus number | 9 | 12 | 21 | 22 | 25 | Total cost ($) | Total cost decrease (%) |
|---|---|---|---|---|---|---|---|---|
| 1 | Install. Cap. (KVAR) | 150 | 150 | 150 | 150 | 150 | 520,030 | 14.1463 |
| 2 | Bus No. | 14 | 25 | - | - | - | 526,370 | 13.1005 |
| | Install. Cap. (KVAR) | 450 | 300 | - | - | - | | |
| 3 | Bus No. | 14 | 23 | - | - | - | 524,040 | 13.4845 |
| | Install. Cap. (KVAR) | 300 | 300 | - | - | - | | |

Table 3: The capacitor placement results in the 9-buses system for (GALEGO, MONTICELLI, ROMERO, 2001)

| Bus No. | (KVAR) Installed capacitor | Total cost decrease (%) |
|---|---|---|
| 2 | 300 | 8.15 |
| 5 | 300 | |
| 6 | 600 | |
| 7 | 1200 | |

Table 4: The results of method 1 for capacitor placement in the 9-buses system

| Bus No. | (KVAR) Installed Capacitor | Total cost decrease (%) | ULTC tap position (p.u) |
|---|---|---|---|
| 5 | 2400 | 9.95 | 1.0467 |
| 7 | 600 | | |
| 10 | 150 | | |

KVA) using heuristic methods, genetic algorithm, tabu search and simulate annealing are shown in Table 3.

The cost for the system before capacitor allocation is 329,039 $. We use method one that has the best answers for the comparison. The cost function is edited as Galego *et al.* (2001) and the results are mentioned in Table 4. As it is shown in this table the method one result is better than the results by Galego *et al.* (2001) and the system also does not have the voltage drop problem. However, Galego *et al.* (2001) divided load's curve into 3 levels whereas in the proposed method in this study the load's curve can be divided into several more levels.

## CONCLUSION

In the proposed method the multi agent reinforcement learning along with heuristic strategies is used for capacitor placement and reactive power compensation in the distribution power system. The method is quite simple compare to other complicated mathematical optimization artificial intelligence based methods. Fixed standard capacitor banks and under load tap changer transformer ULTC used and better results are achieved compare to previous methods.

In most of methods for optimizing the reactive power in distribution systems the load's profile is assumed as a constant level, that this assumption is the reason where results in theory are different to one done in practice, but by the proposed method the values and locations of the capacitor banks could be find over the entire distribution system's load curve and so more realistic results are achieving.

The designed algorithms were applied in standard power systems and comparison of results are shown the advantage of these methods to methods done by now.

## REFERENCES

Abdel-Salam, T.S., A.Y. Chikhani and R. Hackam, 1994. A new technique for loss reduction using compensating capacitors applied to distribution systems with varying load condition. IEEE Trans. Power Delivery, 9: 819-827.

Abdelaziz, S., B.K. Khadija, B. Chokri and E. Mohamed, 2011. Voltage regulation and dynamic performance of the tunisian power system with wind power penetration. Trends Applied Sci. Res., 6: 813-831.

Adeyemo, J.A., 2011. Reservoir operation using multi-objective evolutionary algorithms: A review. Asian J. Sci. Res., 4: 16-27.

Ahmed, B.S. and K.Z. Zamli, 2011. The development of a particle swarm based optimization strategy for pairwise testing. J. Artif. Intell., 4: 156-165.

Augugliaro, A., L. Dusonchet, S. Favuzza and E.R. Sanseverino, 2004. Voltage regulation and power losses minimization in automated distribution networks by an evolutionary multiobjective approach. IEEE Trans. Power Syst., 19: 1516-1527.

Azim, S., K.S. Swarup, 2005. Optimal capacitor allocation in radial distribution systems under APDRP. Proceeding of the IEEE Indices Conference, December 11-13, 2005, Chennai, India pp: 614-618.

Baran, M.E. and F.F. Wu, 1989. Network reconfiguration in distribution systems for loss reduction and load balancing. IEEE. Trans. Power Delivery, 4: 1401-1407.

Bellman, R.E., 1957. Dynamic Programming. Princetion University Press, Princeton, New Jersey, ISBN: 0486428095

Bhattacharya and Goswami, 2009. A new fuzzy based solution of the capacitor placement problem in radial distribution system. Exp. Syst. Appl., 36: 4207-4212.

Chin, H.C. and W.M. Lin, 1994. Capacitor placements for distribution systems with fuzzy algorithm. Proceedings IEEE Region Ann. Int. Conf. Theme Front. Comput. Technol., 2: 1025-1029.

Dehini, R., B. Ferdi and B. Bekkouche, 2012. Fuzzy logic controller optimization based on GA for harmonic mitigation. J. Artif. Intell., 5: 26-36.

Delfanti and Gianpietro, 2000. Optimal capacitor placement using deterministic and genetic algorithm. IEEE Trans. Power system, 15: 1041-1046.

Galego, R.A., A.J. Monticelli and R. Romero, 2001. Optimal capacitor placement in radial distribution networks. IEEE trans. Power Systems, 16: 4-4.

Grainger and Civanler, 1985a. Volt/Var control on distribution systems with lateral branches using shunt capacitors and voltage regulators Part II: The solution method, IEEE Trans. Power Apparatus System, 104: 3284-3290.

Grainger and Civanler, 1985b. Volt/Var control on distribution systems with lateral branches using shunt capacitors and voltage regulators part III: The numerical results. IEEE Trans. Power Apparatus and System, 104: 3291-3297.

Grainger, J.J. and S. Civanlar, 1985c. Volt/var control on distribution systems with lateral branches using shunt capacitors as voltage regulators-Part I, II and III. The numerical results. IEEE Trans. Power Apparatus Syst., 104: 3278-3297.

Hagan, T.S.H.G. and B.J.A. Krose, 1997. A short introduction to reinforcement learning. Proceedings of the 7th Belgian-Dutch Conference on Machine Learning, October 21, 1997, Tilburg University, The Netherlands, pp: 7-12.

Haque, M.H., 1999. Capacitor placement in radial distribution systems for loss reduction. IEE Proc. Gener. Transm. Distrib., 146: 501-505.

Kaelbling, L.P., M.L. Littman and A.W. Moore, 1996. Reinforcement learning: A survey. J. Artificial Intell. Res., 4: 237-285.

Liang, R.H. and C.K. Cheng, 2001. Dispatch of main transformer ULTC and capacitors in a distribution system. IEEE Trans. Power Delivery, 16: 625-630.

Mekhamer, S., S. Soliman, M. Moustafa and M. El-Hawary, 2003. Application of fuzzy logic for reactive-power compensation of radial distribution feeders. IEEE Trans. Power Syst., 18: 206-213.

Miu, K.N., H.D. Chiang and G. Darling, 1997. Capacitor placement, replacement and control in large-scale distribution systems by a GA-based two-stage algorithm. IEEE Trans. Power Syst., 12: 1160-1166.

Nouri, M.A., A. Hesami and A. Seifi, 2007. Reactive power planning in distribution systems using a reinforcement learning method. Proceedings of the International Conference on Intelligent and Advanced Systems, November 25-28, 2007, Kuala Lumpur, Malaysia, pp: 157-161.

Otieno, F.A.O. and J.A. Adeyemo, 2010. Strategies of differential evolution for optimum cropping pattern. Trends Applied Sci. Res., 5: 1-15.

Saxena, N.K., M.A. Khan and P.K.S. Pourush, 2011. GA analysis of parameters of magnetically biased microstrip rectangular patch antenna. J. Artif. Intell., 4: 197-206.

Sutha, S. and N. Kamaraj, 2008. Particle swarm optimization applications to static security enhancement using multi type facts devices. J. Artif. Intell., 1: 34-43.

Sutton, R.S. and A.G. Barto, 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA., USA., ISBN-13: 9780262193986, Pages: 322.

Taleski, R. and D. Rajicic, 1996. Energy summation method for energy loss computation in radial distribution networks. IEEE Trans. Power Syst., 11: 1104-1111.

Teng, J.H., 2003. A direct approach for distribution system load flow solutions. IEEE Trans. Power Delivery, 18: 882-887.

Wang, J.C., H.D. Chiang, K.N. Miu and G. Darling, 1997a. Capacitor placement and real time control in large-scale unbalanced distribution systems: Loss reduction formula, problem formulation, solution methodology and mathematical justification. IEEE Trans. Power Delivery, 12: 953-958.

Wang, J.C., H.D. Chiang, K.N. Miu and G. Darling, 1997b. Capacitor placement and real time control in large-scale unbalanced distribution systems: Numerical studies. IEEE Trans. Power Delivery, 12: 959-964.

Watkins, C.J.H.W. and P. Dayan, 1992. Tecknical note: Q-learning. Mach. Learn., 8: 279-292.

Yap, D.F.W., S.P. Koh, S.K. Tiong and S.K. Prajindra, 2011. Particle swarm based artificial immune system for multimodal function optimization and engineering application problem. Trends Applied Sci. Res., 65: 282-293.