Journal of Applied Sciences

ISSN 1812-5654

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 8 (21): 3924-3930, 2008
ISSN 1812-5654

Q-Learning Based Cooperative Multi-Agent System Applied to Coordination of Overcurrent Relays

J. Sadeh and M. Rahimiyan
Department of Electrical Engineering, Faculty of Engineering,
Ferdowsi University of Mashhad, Mashhad, Iran

Abstract: In this study, a new method for optimum coordination of overcurrent relays is proposed. In order to solve the over current relays coordination problem, a Q-learning based cooperative multi-agent system is designed and applied. In the proposed method, each relay based on the Q-learning algorithm can act as an autonomous adaptive intelligent agent to find the desirable setting according to each power system configuration. In addition, by injecting the cooperation concept in the reward function of the Q-learning algorithm, all relays cooperate each other to minimize the overall operating time of protection system, while, they satisfy the coordination constraints. The efficiency of the suggested technique is illustrated by applying the method for coordination of overcurrent relays in a typical power system.

Key words: Overcurrent relay coordination, reinforcement learning, cooperative multi-agent system, Q-learning algorithm

INTRODUCTION

Overcurrent relays have been used as an applicable tool for distribution and transmission line protection and, also for many years, their coordination problem has been studied by the researchers and engineers of power system. The objective function of the problem, which is the total operating time of the relays, is minimized to have a high speed protective system. In addition, some coordination constraints must be satisfied so that the protection system is kept reliable and selective. In the existing literature, several methods have been utilized to solve this problem. Since years, mathematical formulation-based classical optimization techniques such as linear programming (Urdaneta *et al*., 1996, 2000), non-linear programming (Sadeh, 2005; Keil and Iäger, 2008) and mixed-integer linear programming (Zeineldin *et al*., 2004) have been applied. The non-linear methods try to find the Time Multiplier Setting (TMS) and Current Setting (I_{set}) simultaneously. However, these approaches work on iterative search, which does not have enough ability to achieve the global optimal setting of the relays. To reduce the complexity of non-linear problem, at first, the problem is transformed to the linear programming by determining the I_{set} of relays using some heuristic techniques. Then, the TMS of relays are calculated by solving the obtained linear optimization problem. However, in order to solve the nonlinear problem, it would be beneficial to use the stochastic optimization approaches such as genetic and evolutionary algorithms (Razavi *et al*., 2008; So and Li, 2000), particle swarm optimization (Mansour *et al*., 2007; Zeineldin *et al*., 2006) which can overcome the drawbacks of traditional methods.

After restructuring the vertically electricity industry to competitive power market, the relay coordination has become more critical and important problem. In this new environment, establishing the competitive energy market and providing open access to transmission network, the efficient operation of transmission network encountered with some issues and challenges. This leads to increase the complexity and sensitivity of overcurrent relay coordination problem. Due to considerable change of power system configuration in the power market, the obtained optimal setting of overcurrent relays in a configuration may not be suitable for the new configurations. In order to have a selective, reliable and high speed protection system for different situations of power system, it is necessary to use an intelligent approach with more ability than the mentioned classical and stochastic optimization methods.

In this study, the overcurrent relay coordination problem is modeled as Q-learning based cooperative multi-agent system. Each relay tries to find the desirable setting in each situation of power system based on Q-Learning (QL) algorithm which is kind of model-free Reinforcement Learning (RL). In fact, each relay learns

Corresponding Author: Dr. Javad Sadeh, Department of Electrical Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, P.O. Box 91775-1111, Mashhad, Iran

from the past experiences which obtained by taking actions in the repeated games to minimize its operating time for all possible situations, while, it does not have any knowledge about actions of other relays and the occurrence of new power system configurations. As a result, the overcurrent relays can act as an autonomous adaptive intelligent agent in the Multi-Agent System (MAS). Moreover, by combining the cooperation concept in the MAS with the QL algorithm, the intelligent adaptive agents attempt cooperating each other to minimize the overall operating time whereas they satisfy their coordination constraints. Considering the ability of RL methods to solve the problem with finite discrete variables, the structural constraints of overcurrent relays are respected easily.

REINFORCEMENT LEARNING

The reinforcement learning problem is the learning problem of agent interacts with its environment so as to achieve its goal. In fact, RL is learning policy which means what to do or, in other words, how to map each situation to action to maximize the received long run reward. The trial and error and delayed reward are the most important characteristics of RL. In order to find the best policy, at first, the agent must obtain new experiences based on trial and error. In each state of environment, it evaluates how good the chosen action is considering the immediate reward and value of new state. The value of a state is the long term reward expected to be acquired over the future starting from that state (Sutton and Barto, 1998).

Assume that an agent interacts with its environment at a sequence of discrete time steps, t = 0, 1, 2, .., as shown in Fig. 1. Also, assume that $S = \{s_1, s_2,. . . , s_n\}$ is the finite set of possible states of the environment and $A = \{a_1, a_2,..., a_m\}$ is the finite set of admissible actions, which the agent can take. At each time step t, the agent senses the current state $st = s \in S$ of its environment and accordingly selects an action $a_t = a \in A$. As a result of its action, the state of environment changes to the new state $s_{t+1} = s' \in S$, with a transition probability Pss' (a) and the agent receives an immediate reward r_{t+1}.

However, the reinforcement learning is established on an important assumption that the interaction of agent and dynamic environment satisfies the Markov property and the reinforcement learning task is a Markov Decision Process (MDP). In this condition, the value of a state s under policy π, denoted by $V^\pi(s)$, as given in Eq. 1, is the expected return when starting in state s and afterward following policy π.

$$V^\pi(s) = E\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s\right\} \qquad (1)$$

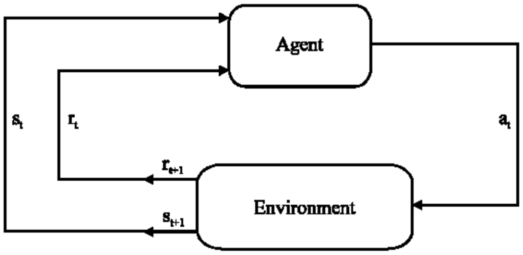

Fig. 1: The agent-environment interaction in RL

where, γ is discounted factor and r is the reward function. Considering the Bellman equation in dynamic programming, the above equation can be written as:

$$V^\pi(s) = \sum_a \pi(s,a) \sum_{s'} P^a_{ss'}\left[R^a_{ss'} + \gamma V^\pi(s')\right] \qquad (2)$$

In which π(s,a) is the probability of taking action a in state s and $R^a_{ss'}$ is the expected value of reward r_{t+1}. Now, the optimal value and the best action in state s are calculated by taking the maximum value of Eq. 2 over action space A(s) and building Bellman optimality equation as follows:

$$V^*(s) = \max_{a \in A(s)} \sum_{s'} P^a_{ss'}\left[R^a_{ss'} + \gamma V^*(s')\right] \qquad (3)$$

Here, it is explained that how the RL algorithm is used to find the desirable setting of relays in power system protection. In order to reach this objective and to solve relay coordination problem, we apply the cooperation concept in RL-based MAS.

RL-BASED COOPERATIVE MAS APPLIED TO RELAY COORDINATION

The relay coordination problem is modeled as a cooperative MAS in which each agent intelligently finds out the desirable setting using RL algorithm. As a result, the overcurrent relay coordination changes to the RL problem which can be solved using the Dynamic Programming (DP) or Temporal-Difference (TD) methods.

Modeling state of environment and agent's action: Each relay in the coordination problem is an intelligent agent which interacts with the power system as dynamic environment. The power system configuration is the state of environment. The environment can experience some states due to outage or entrance of generators and transmissions and also load variation. In this condition, it is necessary to adopt the relays setting to the different situations and update the policy of coordination for each

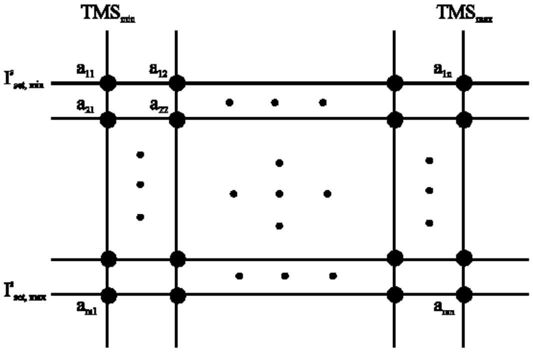

Fig. 2: The definition of action space

one. However, as described earlier, the RL is a capable method to improve the selected strategy in the previous state of dynamic environment to the new one by evaluating the new state. Therefore, using the RL method, an adaptive coordination of relays is possible which increases the reliability and security of power system protection.

In the RL-based MAS, each agent selects its setting i.e., TMS and I_{set} as action among action space. Considering relay structure, the range of TMS is limited between TMS_{min} and TMS_{max}. Also, the $I^s_{set,min}$ and $I^s_{set,max}$ of relays for each state s can be obtained considering the power flow and short circuit calculations. Then, the two dimensional action space is built for each relay using its discrete parameters as shown in Fig. 2. It is obvious that $I^s_{set,min}$ and $I^s_{set,max}$ of each relay can be different from other relays.

Defining agent's punishment function: Here, the cooperation concept of agents is used to solve the overcurrent relay coordination problem. Generally, based on the literature of overcurrent relay coordination, the setting of relays is adjusted to minimize their total operating time (Eq. 4), such that the physical constraints in Eq. 5, the assumed operation curve of relay and the coordination constraints of back up-primary pairs in Eq. 6 are satisfied. In order to apply the RL-based cooperative MAS in coordination problem, the reward function is replaced by the punishment function and the max operator in Eq. 3 replaced by min operator. The mathematical formulation of relays coordination problem is presented as follows:

$$\min \sum_{i=1}^{Nr} t_i \tag{4}$$

$$\begin{aligned} &\text{subject to:} \\ &TMS_{i,min} \le TMS_i \le TMS_{i,max} \\ &I_{set,i,min} \le I_{set,i} \le I_{set,i,max} \end{aligned} \tag{5}$$

$$\begin{aligned} &t_i^B(I_{max}^{B,fault_j}, TMS_i, I_{set,i}) - t_j^P(I_{max}^{P,fault_j}, TMS_j, I_{set,j}) \ge CTI \\ &(i,j) \in \text{all Backup} - \text{Primary pairs} \end{aligned} \tag{6}$$

where, Nr is the number of relays, $I_{max}^{B,fault_j}$ is the maximum current flows through the back up overcurrent relay i when the fault is happened close to and in front of primary relay j and $I_{max}^{P,fault_j}$ is the corresponding current through the primary relay. Thus, in order to minimize the total operating time of all relays using the MAS, the intelligent agents must give support and cooperate each other. Nevertheless, the MAS can be designed such that the agents compete or cooperate with each other. This strongly depends on how the punishment function of agents is defined.

To reach this goal, two additive components are assigned as the punishment function of each intelligent agent. First part is the total operating time of all relays the same as Eq. 4 and the second part is taken into account to respect the coordination constraint in Eq. 6. To keep the coordination of back up-primary pairs, a penalty function is defined as a second part of punishment function. Now, we call its primary and back up relays as primary and back up agents. Suppose the sets of P^i and B^i consist of all primary and back up agents of the intelligent agent ith, respectively. The penalty function of ith intelligent agent is sum of following two-argument functions given in Eq. 7 and 8. Each two-argument function is driven from the coordination constraint of the agent ith with one of the primary or back up agents.

$$Penalty_{i,j}^{P_i} = \begin{cases} \exp\left(-\beta\left(t_i^B - t_j^P - CTI\right)\right) - 1 & t_i^B - t_j^P - CTI < 0 \\ 0 & \text{else} \end{cases} \tag{7}$$

$$Penalty_{i,j}^{B_i} = \begin{cases} \exp\left(-\beta\left(t_j^B - t_i^P - CTI\right)\right) - 1 & t_j^B - t_i^P - CTI < 0 \\ 0 & \text{else} \end{cases} \tag{8}$$

In the Eq. 7 j belongs to P^i and in Eq. 8 j belongs to B^i. Using these two equations, the penalty function of ith intelligent agent as a second term of its punishment function is written as:

$$Penalty_i = \sum_{j \in P_i} penalty_{i,j}^{P_i} + \sum_{j \in B_i} penalty_{i,j}^{B_i} \tag{9}$$

Now, the punishment function of ith intelligent agent is constructed as follows:

$$Punishment_i = \sum_{k=1}^{Nr} t_k + penalty_i \tag{10}$$

As a result, each intelligent agent takes an action in each state of power system to minimize the cumulative punishment in long run. Considering the defined punishment function, each agent not only try to minimize its operating time, however, it selects an action so that minimizes the total operating time when satisfies its coordination constraints. Whereas, it does not have any knowledge about other intelligent agents' actions, the operating time of other agents is added to its operating time to guide the agents to be cooperative. Consequently, instead of solving the relays coordination problem using one agent with 2*Nr dimensional action space, using the RL-based MAS, this studied problem is modeled as a distributed system. Using the proposed model, the dimensional of action space is reduced and, therefore, the search process to find the optimal action can be performed efficiently. Though, the sub optimality issue arises in the MAS which is unavoidable.

To find the optimal setting of each agent in state s, the optimal value function must be computed. If the model of environment, means $P^a_{ss'}$ and $R^a_{ss'}$, is known, the DP can be a suitable technique to evaluate the value function. The $P^a_{ss'}$ can be calculated using the historical data, but the second component i.e., $R^a_{ss'}$ is not absolutely known for all agent. Because the punishment function of each agent not only depends on its action, but is affected by other agents' actions which are not known from the viewpoint of the agent. In this condition the QL algorithm (TD(0)) which is kind of reinforcement learning can be useful to solve the mentioned problem. The QL algorithm does not need any model of environment and develop the optimal action by estimating the value function.

OVERCURRENT RELAY COORDINATION BASED ON QL ALGORITHM

To find the optimal solution of Markov Decision Problems without a prior knowledge of the environment, the Watkins's QL algorithm, which is a kind of reinforcement learning algorithm (Sutton and Barto, 1998), can be used. The QL algorithm is a very effective model-free algorithm that is insensitive to exploration for learning from delayed reinforcement (Kaelbling *et al*., 1996). Thus, it may be a suitable method to model the overcurrent relay coordination problem as a QL-based cooperative MAS. Each agent is able to select the best action by cooperating with other agents when it does not have any knowledge about the model of power system environment and other agents' actions.

QL algorithm: In the QL algorithm, the value function for each admissible (s, a) is defined as a Q-value. Considering the QL algorithm, each agent attempts to discover the optimal policy $\pi^*(s)\in A$ to maximize the Q-value of each state over the long run or maximize the total reward that it receives in repeated games, using the Bellman optimality Eq. 11:

$$Q^*(s,a)=\sum_{s'} P^a_{ss'}\left[R^a_{ss'}+\gamma \max_{a'} Q^*(s',a')\right] \tag{11}$$

$$\pi^*(s)=\arg\max_{a}\,(Q^*(s,a)) \tag{12}$$

Without learning a model, however, the QL algorithm is able to determine the optimal policy for every admissible (s, a) by online estimating its Q-value using the zero-order temporal difference TD(0) method. To start the learning process, the Q-value lookup table is initialized randomly or using the agent's knowledge. In each iteration of game, after doing action a_t, the only available information for the QL algorithm is s_t, a_t, s_{t+1} and r_{t+1}. The updating rule for the state-action pair (s_t, a_t) is given by:

$$Q_{t+1}(s_t,a_t)=Q_t(s_t,a_t)+\alpha\Delta Q(s_t,a_t) \tag{13}$$

$$\Delta Q(s_t,a_t)=r_{t+1}+\gamma \max_{a'} Q(s_{t+1},a')-Q_t(s_t,a_t) \tag{14}$$

where, α is the agent's learning rate. α can be interpreted as how much the estimated Q-values are updated by new data. To apply the QL algorithm in coordination problem, the reward function must be substituted by the punishment function. Also, in Eq. 14, the min operator is used instead of max operator.

QL-based relay coordination: Here, it is presented that how the QL- based cooperative MAS is used to solve the relay coordination problem. Interactions of all agents with environment in the MAS are shown in Fig. 3.

The same as self-play problem, in MAS, all agents after sensing the state of environment and evaluating their situations take the desirable action simultaneously.

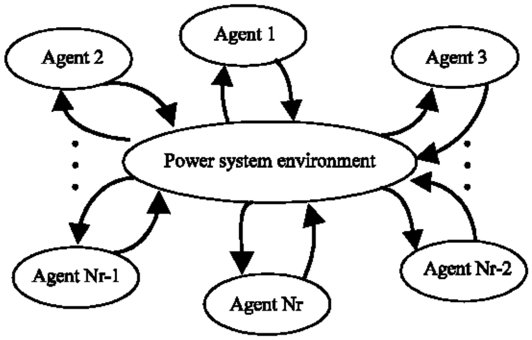

Fig. 3: The schematic diagram of cooperative MAS

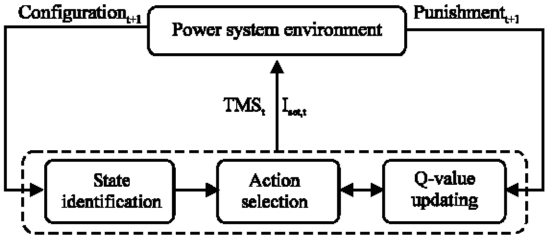

Fig. 4: The flowchart of QL-based overcurrent relay coordination

To find the desirable overcurrent relay coordination, each agent participates in a repeated game in which the behaviors of all agents are oriented based on QL algorithm. In Fig. 4, it is illustrated that how an intelligent agent interacts with the power system environment to adjust its setting.

According to this flowchart, each agent' learning process to determine its setting is done as follows:

- **State identification:** The state of environment for the current step is the power system configuration in the pervious step. It is obtained using the Monte Carlo simulation which is run based on the calculated occurrence probability of all scenarios of power system configuration.
- **Action selection:** After obtaining the current state, the agent uses its Q-value lookup table which saves the Q-values for each state-action pair. The action selection through the QL algorithm is done by choosing the action with minimum Q-value in the current state. To trade off between exploitation and exploration, the agent can utilize from the ε-greedy strategy. It means that the agent selects the action that has the minimum Q-value with high probability (1-ε) and an arbitrary action from all admissible actions with small probability ε, independent of the Q-values.
- **Q-value updating:** At the end of each step, the agent is notified of the new configuration (s_{t+1}) to take the next decision (a_{t+1}). Also, the power flow and short circuit calculations are performed in the current configuration (s_t). Then, regarding the setting of agents (a_t), the punishment function is calculated considering the mentioned cooperative philosophy using the Eq. 7-10 and then updates its Q-value according to Eq. 13 and 14 by using the min operator instead of max operator.

RESULTS AND DISCUSSION

Here, the QL-based cooperative MAS is applied to solve the coordination problem of overcurrent relays. In order to study the performance of the proposed method, a typical distribution system, consists of three overcurrent relays, is considered (Fig. 5). The generator and the parallel transformers are the Thevenin equivalent calculated from bus A. Considering the change of power system topology, the Thevenin equivalent can vary. As a result, it is obvious that the Short Circuit Capacity (SCC) can take different values. It leads to change of the short circuit current in the distribution system and, consequently, the re-set of relays setting is needed. For example, in this simulation, two scenarios are possible for Thevenin equivalent which is shown in Fig. 5 and 6.

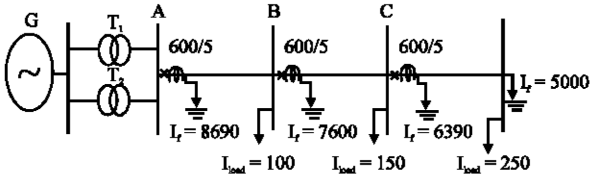

Fig. 5: The studied distribution system in scenario 1

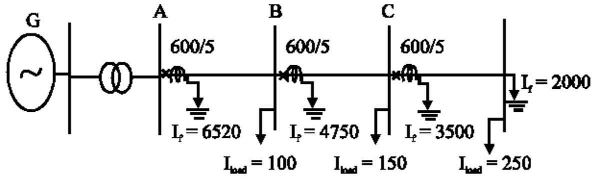

Fig. 6: The studied distribution system in scenario 2

The operation characteristic of the overcurrent relays are assumed to be as Eq. 15.

$$t(I,TMS,I_{set}) = 3\times\frac{TMS}{\log(\frac{I}{I_{set}})} \qquad (15)$$

where, t is the operation time of the relay and I is the current through the relay. Also, I_{set} and TMS can take the discrete value between 0.25 and 2.5 by stepping 0.125 and between 0.1 and 1 by stepping 0.025, respectively. Thus, the action space of each intelligent agent as an overcurrent relay consists of 19×37 settings. The required fault currents of coordination problem are presented in the Fig. 5 and 6 for both network topologies. The CTI is assumed to be 0.3 sec. Also, in each scenario, there are two primary-back up pairs i.e., (B, A) and (C, B).

It should be noted that the I_{set} of each relay must be greater than the load current and be smaller than the value of minimum current flows through it when fault occurs at far bus of its primary relays. As a result, some actions in the 19×37 actions may be inadmissible which are

Table 1: Settings of overcurrent relays using the proposed method

	Scenario 1		Scenario 2	
Relay	TMS	I_{set}	TMS	I_{set}
A	0.325	1.375	0.275	0.875
B	0.225	0.875	0.175	0.875
C	0.100	0.500	0.100	0.625

Table 2: Settings of overcurrent relays using nonlinear programming

	Scenario 1		Scenario 2	
Relay	TMS	I_{set}	TMS	I_{set}
A	0.2545	1.375	0.2630	0.875
B	0.1904	0.875	0.1673	0.875
C	0.1000	0.500	0.1000	0.625

neglected. The admissible actions, which satisfy the mentioned constraint, are identified. Therefore, the intelligent agent is only able to select the admissible actions. Now, the QL-based cooperative MAS is implemented based on the proposed flowchart of coordination problem in Fig. 4. The obtained setting of overcurrent relays are shown in Table 1.

The total operation time of relays according to the obtained settings in scenarios 1 and 2 are 1.7609 and 1.6122 sec, respectively.

It should be noted that in the studied system with two possible scenarios, the occurrence of state s_{t+1} is independent of the state s_t and a .$_t$ Thus, the value of parameter γ is chosen zero without loss of optimality. Moreover, in this situation, the punishment function in state s_t only depends on a_t. Consequently, the probability of occurrence of two scenarios can not affect the setting of relays.

At the end of this section, the results obtained using the proposed method are compared with those of a classical optimization approach, namely Sequential Quadratic Programming (SQP) method. Optimization toolbox of MATALB® is used to solve the nonlinear programming. The optimal values of time multiplier settings and the current settings of the relays are shown Table 2. The current settings of the relays in both methods are completely the same. But the time multiplier settings of the relays obtained using nonlinear programming are lower than those of obtained using the proposed method. However, it is very important to note that in the proposed method, the change of topology is taken into account adaptively and the settings are modified after changing the system topology automatically.

CONCLUSION

In this research, the reinforcement learning-based cooperative multi-agent system is constructed and applied in solving the overcurrent relay coordination problem for first time in the existing literature. As a result, combining the reinforcement learning method and cooperation concept in the multi-agent system, each overcurrent relay as an intelligent agent adjusts its setting such that it improves the overall operating time when respects the coordination constraints. Considering the ability of reinforcement learning to select the best strategy according to the state of the environment, the designed tool can adopt the relay coordination to the new power system configuration. Consequently, we could achieve an adaptive coordination which improves and reinforces the security and reliability of power system protection. Also, due to increasing uncertainty and considerable change of power system configuration in the competitive electricity market, the importance of using such method to provide a flexible operation of power system should be considered more than before.

REFERENCES

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

Keil, T. and J. Jäger, 2008. Advanced coordination method for overcurrent protection relays using nonstandard tripping characteristics. IEEE Trans. Power Deliv., 23: 52-57.

Mansour, M.M., S.F. Mekhame, N. El-Sherif and El-Kharbawe, 2007. A modified particle swarm optimizer for the coordination of directional overcurrent relays. IEEE Trans. Power Deliv., 22: 1400-1410.

Razavi, F., H.A. Abyaneh, M. Al-Dabbagh, R. Mohammadi and H. Torkman, 2008. A new comprehensive genetic algorithm method for optimal overcurrent relays coordination. Elect. Power Syst. Res. J., 78: 713-720.

Sadeh, J., 2005. Optimal coordination of overcurrent relays in an interconnected power system. 15th Power System Computation Conference (PSCC), August 22 to 26, Liege, Belgium, pp: 1-5.

So, C.W. and K.K. Li, 2000. Time coordination method for power system protection by evolutionary algorithm. IEEE Transa. Ind. Appl., 36: 1235-1240.

Sutton, R.S. and A.G. Barto, 1998. Reinforcement Learning: An Introduction. 1st Edn., MIT Press, Cambridge, MA, ISBN 0-262-19398-1.

Urdaneta, J., H. Restrepo, S. Marquez and J. Sanchez, 1996. Coordination of directional overcurrent relay timing using linear programming. IEEE Trans. Power Deliv., 11: 122-129.

Urdaneta, A.J., L.G. Perez, J.F. Gomez, B. Feijoo and M. Gonzalez, 2000. Presolve analysis and interior point solutions of the linear programming coordination problem of directional overcurrent relays. Elect. Power Energy Syst. J., 23: 819-825.

Zeineldin, H., E. El-Saadany and M. Salama, 2004. A novel problem formulation for directional overcurrent relay coordination. Large Engineering Systems Conference on Power Engineering LESCOPE-04, July 28-30, pp: 48-52.

Zeineldin, H.H., E.F. El-Saadany and M.M.A. Salama, 2006. Optimal coordination of overcurrent relays using a modified particle swarm optimization. Elect. Power Syst. Res. J., 76: 988-995.