Wireless Personal Communications
https://doi.org/10.1007/s11277-020-07181-w

Q-Learning-Based Fuzzy Logic for Multi-objective Routing Algorithm in Flying Ad Hoc Networks

Qin Yang[1] · Sung-Jeen Jang[1] · Sang-Jo Yoo[1]

Abstract
Flying ad hoc networks (FANETs) that consist of multiple unmanned aerial vehicles (UAVs) have developed owing to the rapid technological evolution of electronics, sensors, and communication technologies. In this paper, we propose a multi-objective routing algorithm for FANETs. In addition to the basic transmission performance in the construction of the routing path, the network impact according to the mobility of the UAV nodes and the energy state of each node should be considered because of the characteristics of the FANET, and the overall efficiency and safety of the network should be satisfied. We therefore propose the use of Q-learning-based fuzzy logic for the FANET routing protocol. The proposed algorithm facilitates the selection of the routing paths to be processed in terms of link and overall path performances. The optimal routing path to the destination is determined by each UAV using a fuzzy system with link- and path-level parameters. The link-level parameters include the transmission rate, energy state, and flight status between neighbor UAVs, while the path-level parameters include the hop count and successful packet delivery time. The path-level parameters are dynamically updated by the reinforcement learning method. In the simulation results, we compared the proposal with the conventional fuzzy logic and Q-value-based ad hoc on-demand distance vector. The results show that the proposed method can maintain low hop count and energy consumption and prolong the network lifetime.

Keywords Routing algorithm · Flying ad hoc networks · Multi-objective · Fuzzy logic · Q-learning

1 Introduction

In the wake of the rapid technological development of electronics, sensors, and communication technologies, it has become possible to produce unmanned aerial vehicle (UAV) systems that can fly autonomously and can be operated remotely without the requirement for carrying human personnel. UAVs are now used in a wide range of

✉ Sang-Jo Yoo
sjyoo@inha.ac.kr

[1] Department of Information and Communication Engineering, Inha University, 253 YongHyun-dong, Nam-gu, Incheon, Korea

Published online: 21 January 2020

applications from package delivery to data acquisition over the Internet-of-Things (IoT) sensor networks [1]. UAVs form wireless networks that can employ UAVs in order to extend the range of communication, maximize network data communication capability by using vehicles as relay nodes, collect data from a wide area network in remote or harsh environments and aid node localization in a mobile network. If all UAVs are directly connected to an infrastructure, such as a ground base or a satellite, then this infrastructure based communication architecture restricts the capabilities of the multi-UAV systems. Ad hoc networking between UAVs can solve the problems arising from a fully infrastructure based UAV networks. This created a new paradigm, referred to as the flying ad-hoc networks (FANETS) where each flying UAV can act as a router and they establish fully distributed wireless networks.

FANET can be considered as a special form of mobile ad hoc network (MANET) and vehicle ad hoc network (VANET). Even though the FANET and these networks have common characteristics, it has several unique design challenges [2]. In the FANET, the degree of mobility of each node is much higher than that in the VANET and MANET. A UAV has a speed of 30–460 km/h [3]. In other words, the locations of UAVs change from moment to moment, and the links among UAVs are established and destroyed intermittently. Owing to the higher degree of mobility, the FANET topology also changes more frequently than that of the MANET and VANET topology. The topology of UAV network changes quickly, which results in a link variation problem and additionally, frequent topology changes also increases the latency, packet loss, and control signaling overhead. In multi-UAV applications, flight plan changes, fast and sharp UAV movements, and various UAV formations directly affect the mobility of the multi-UAV systems. The initial FANET studies and experiments were designed using the existing VANET or MANET routing protocols. Unlike VANET routing, in which movement direction and path of vehicles are generally limited and usually predictable, in FANET frequent topology changes make it be difficult to use VANET routing protocols directly. Also conventional MANET protocols may not appropriately adapt to the frequent changes of UAV topology in real time. In most of MANET routing protocols utilize proactive or reactive routing path finding algorithms. Because of the high mobility of the FANET nodes, maintaining a routing table before actual data transmission, as in proactive methods, is not optimal. Repetitive path finding before each packet delivery, as in reactive routing, can also be exhaustive. However, owing to the various environmental non-standardized conditions and network characteristics of FANETs, there are many factors (e.g. routing, network formation, energy resource management, flight squadron configuration, and etc.) that some computational intelligence techniques can contribute, which includes swarm intelligence [4], machine learning, and fuzzy logic (FL).

Fuzzy logic was introduced by Zadeh and is a mathematical discipline used for expressing human reasoning in terms of rigorous mathematical notation [5]. Human beings can make decisions even in the presence of imprecise or incomplete knowledge. Fuzzy logic facilitates the approximation of human reasoning, begins with a set of user-supplied human language, and builds the rules. The fuzzy systems convert these rules into their mathematical equivalents. Additional benefits of fuzzy logic include its simplicity and its flexibility, which aid in solving problems comprising imprecise and incomplete data, and it can be used to model nonlinear functions of arbitrary complexity. The potential of the fuzzy logic has been fully explored in many fields including signal processing, speech recognition, aerospace, robotics, embedded controllers, networking, business, and marketing [6]. Moreover, the use of the fuzzy logic in wireless ad hoc networks has been shown to be a promising technique as it facilitates the combination and evaluation of diverse parameters

in an efficient manner. In ad hoc networks, fuzzy logic has been used in localization, clustering, cluster head election, routing, data aggregation, and security [7].

Another potential technique used in wireless ad hoc network routing problems and that has received extensive attention is machine learning. Machine learning allows ad hoc networks to learn from previous experience, conduct optimal routing actions to save energy and prolong lifetime, and also adapt to dynamic environments. The quality of service requirements can be met in routing problems using simple computational methods and classifiers. Barbancho et al. [8] introduced sensor intelligence routing (SIR) while using self-organizing-map-based (SOM-based) unsupervised learning for detecting optimal routing paths. The SIR introduces a slight modification in Dijkstra's algorithm to form the network backbone and shortest paths from a base station to every node in the network. During the route learning, the second layer neurons compete with each other to reserve high weights in the learning chain. Evidently, the learning phase is a highly computational process owing to the neural network generation task. Some previous researches in wireless networks have treated the problems of ad hoc routing issues as a Markov decision process. The subfield of machine learning that deals with sequential control problems is called reinforcement learning. Reinforcement learning is used quite frequently in routing problems.

In this paper, we propose a routing algorithm that combines fuzzy logic and reinforcement learning algorithms in FANETs. Multiple UAVs are connected to determine the routing path while considering the transmission rate (TR), residual energy, energy drain rate, hop count (HC), and successful packet delivery time (SPDT). A fuzzy system is used to derive reliable links between two UAV nodes, and Q-learning supports the fuzzy system by providing a reward on the path. The link-related parameters are defined using the transmission rate (TR), energy state (ES), and flight status (FS). The ES takes into consideration the residual energy and energy drain rate. The FS takes into consideration the flying velocity and moving direction of neighbor UAV nodes. However, path-related parameters (the HC and SPDT) are computed by the Q-learning from the previous selected path. During continuous packet delivery, the fuzzy system and Q-learning can ensure good communication quality and energy savings for the FANET.

The remainder of this paper is organized as follows. Related works are presented in Sect. 2. Section 3 presents the network model and considered input parameters along with the proposed system architecture and operation process. The parameter design methods and proposed Q-learning-based fuzzy logic routing algorithm are described in Sect. 4. The performance evaluation with the simulation results are presented in Sect. 5. The conclusions of this study are presented in Sect. 6.

2 Related Work

Fuzzy logic provides a simple method for arriving at a definite conclusion based on vague, ambiguous, imprecise, noisy, or missing input information. This aids in managing the performance parameters of the hop count, transmitted packet, and residual energy for the ad hoc networks and converts the complex behavior of the network using a simple rule-based system [9]. A routing protocol called adaptive fuzzy multiple attribute decision routing (AFMADR) was proposed for the VANETs [10]. Four attributes including the distance, direction, road density, and location are used to characterize the candidate vehicles, which form inputs to the fuzzy mapping systems. A proposed adaptive weight algorithm is used to calculate the weights of the attributes to enhance the scalability and

robustness of the AFMADR scheme. A routing protocol named fuzzy-based energy efficient multicast routing (FBEEMR) was proposed for the ad hoc network in [11]. The basic idea of the FBEEMR is to select the best path that reduces the energy consumption of the ad hoc nodes based on fuzzy logic. This protocol is mainly used to extend the lifetime of the ad hoc network with respect to energy-efficient multicast routing by calculating the route lifetime values for each route. In another study on the VANET [12], the use of a novel routing information called machine-learning-assisted route selection (MARS) system was proposed for estimating the necessary information for the routing protocols. In MARS, the road information is maintained in roadside units with the help of machine learning. They used the K-means algorithm for classifying and predicting the moves of vehicles and then selected some suitable routing paths with a better transmission capacity to transmit packets.

Routing schemes comprising the use of Q-learning, which is a reinforcement learning technique, have been proposed. Sun et al. [13] demonstrated the use of the Q-learning algorithm for enhancing the multicast routing protocol in wireless ad hoc networks. The Q-MAP multicast routing algorithm is designed as two phases, which comprise the join query forward and join reply backward. The roles of their design are the establishment of a routing path and resource reservation in a distributed manner. Arroyo-Valles et al. [14] introduced Q-probabilistic routing (Q-PR), which is an enhanced geographic routing algorithm for wireless ad hoc networks that learns from previous routing decisions. Depending on the importance of messages, the expected delivery rate, and the power constraints, Q-PR determines the optimal routes using the reinforcement learning and Bayesian decision model. This algorithm discovers the next hop during the message routing time. The Bayesian method is used to handle the decision of transmitting the packets to the set of candidate neighbor nodes while taking into consideration the data importance, node profiles, expected transmission, and reception energy. Dong et al. [15] proposed the use of reinforcement-learning-based geographic routing (RLGR), which takes into consideration the energy efficiency, delay, and routing failure to improve the network performance in sensor networks.

Research has also been conducted to combine the advantages of fuzzy logic with reinforcement learning, which operates adaptively to the surrounding environment. The protocol, called a portable fuzzy constraint Q-learning protocol, based on ad hoc on-demand distance vector (PFQ–AODV) [16] comprised the use of fuzzy logic to evaluate whether a wireless link is good by considering multiple metrics, which are the available bandwidth, link quality, and relative vehicle movement. Based on an evaluation of each wireless link, the proposed Q-learning protocol learns the best route using the route request messages and hello messages. The dynamic-fuzzy-energy-state-based AODV (DFES-AODV) routing protocol was presented for MANET [17]. The system inputs are the residual battery level and energy drain rate of the mobile nodes. In [18, 19] for wireless sensor network fuzzy based routing protocols were also proposed, in which they considered multi types of awareness for sensor network data delivery. In [20] UAV network formation and packet delivery path are determined using particle swarm optimization.

Most of FANET routing algorithms with intelligence only consider short range (or short time scale) or long range (or long time scale) parameters. In fuzzy logic, they generally take into count the neighbor node's link quality in terms of movement and energy consumption so that only short range or short time scale decision is made. In FANET to deliver the acquired data to the destination node entire path delay and reliability also should be carefully considered. In other hand, Q-learning based FANET reactive routing protocols try to find an optimal path by iterative updating Q-values so that they require convergence time and may not adapt quickly to frequent link quality changes.

In this paper, we take into account both of link and path related quality parameters. In the proposed system, each node observes all neighbor link quality in terms of transmission rate, energy conditions and flight movement pattern. The proposed Q-learning rule evaluates the path cost from source to destination in terms of hop count distance and successful delivery time. Due to the different packet waiting time at each UAV node queue and packet retransmission trials at any links on the path by packet losses, the smaller hop count distance does not guarantee the shorter successful packet delivery time. These path-related parameters are measured whenever data packet is delivered to the destination and the Q-values are updated iteratively. In contrast to other algorithms that combine Q-learning and fuzzy logic, the designed routing path finding algorithm improves the network performance in two aspects. One aspect is the link-related parameters considering the performance of a single link, and the other aspect is the path-related parameters for the whole routing path. The path-related parameters are managed through Q-learning to improve the system's routing paths while learning. In designing the Q-learning, the intention was to find the path efficiently via a unique reward grant method that propagates from the destination node. The proposed fuzzy logic connects link parameters and Q-learning path parameters in the combined fuzzy system to determine the next UAV node to delivery data packets by considering not only short range and short time scale but also long range and long time scale network and environment.

3 Network Model and System Architecture

3.1 Network Model and Input Parameters

In this paper, it is assumed that N UAV nodes are randomly deployed in the geographical area, as shown in Fig. 1, and each UAV node has a different initial velocity and moving direction. The data in the source node may be transmitted to the destination node by the specific purpose of the FANET. Figure 1 shows an example geographical topology with

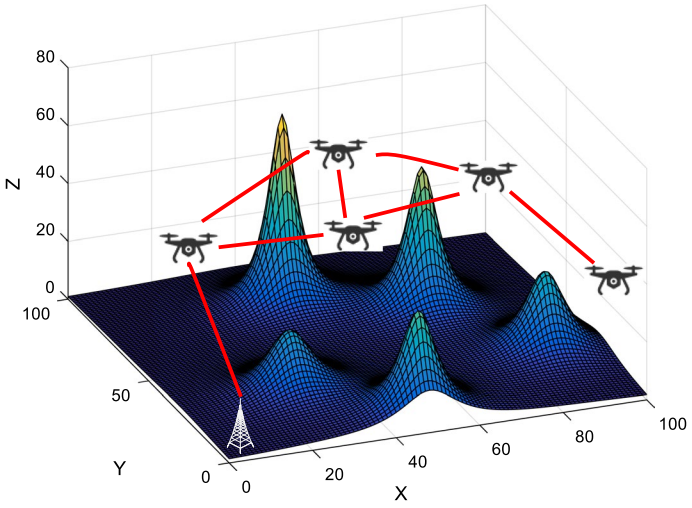

Fig. 1 A routing path scenario in the geographical topology with multiple UAV nodes

multiple UAV nodes, and the entire field is divided into small unit areas. The continuous azimuth and velocity variation provide more practical information for the UAV flying environment regarding the flight computation, and the wireless communication supports UAVs in connecting to each other to build a routing path from a base station to the predetermined sink node.

Our proposed system first finds a path via which the data packet reaches the destination node through a good link in a short-sighted manner and then finds a path that considers not only a good link between two nodes but also a good path in a macroscopic manner through a repetitive routine based on the found path. In this manner, two types of input parameters are used for the system to determine the routing path. The first type comprises link-related inputs, which include the TR, ES, and FS. The second type of inputs comprise path-related parameters that include the HC and SPDT. The definition and characteristic of each parameter are explained as:

3.1.1 Transmission Rate

When determining the network throughput, the data transmission rate (TR) represents the data transmission efficiency and is essential for individual links between the two nodes as well as the overall network performance.

3.1.2 Energy State

It is quite important for the system to consider the battery energy of the nodes to select paths in terms of the whole network. The path selection that does not take into account the battery energy is likely to deplete the energy of a specific node and divide the network [7, 11, 17]. Furthermore, if a node with a high residual energy is used too often to deliver traffic, then it can deplete the energy of the node at a fast rate and shorten the life of the network [18]. Therefore, we consider each node's residual energy after the packet delivery and the energy drain rate in a certain period to present the ES.

3.1.3 Flight Status

In FANETs, the node velocity and moving direction mainly determine the link connection quality. If one node moves to a sink node at a high speed, the possibility of successfully delivering the packet increases [3]. Furthermore, if the moving direction of the current node is similar to that of the neighboring node, the possibility of the two nodes being included within the communication range in the future increases. Therefore, the similarity of the moving direction of two neighbor nodes is also a crucial necessary consideration. The FS takes into consideration the above to indicate whether the link reliability for the neighbor is good according to the flight information.

3.1.4 Hop Count and Successful Packet Delivery Time

The link-related parameters focus only on the link reliability between the two nodes but do not take into consideration the network efficiency for the entire routing path to the sink. We introduce the path-related parameters of the HC and SPDT to evaluate the overall path from a source node to a destination node. The HC indicates how many nodes are required to deliver packets to the destination. Hence, a path with a low HC

can quickly deliver the data and causes a low load on the network [11, 20]. The SPDT records the transmission time from the source to the destination for successfully delivering a data packet. In this paper, the proposed reinforcement learning gives the HC reward and SPDT reward to the visited nodes passing through to the sink. For the next iteration, there will then be a high chance of selecting a neighbor node with a low HC to the sink and a short SPDT.

3.2 System Architecture

The proposed system structure is presented in Fig. 2. The overall structure consists of an environmental parameter part that recognizes the surroundings, a fuzzy logic part that synthesizes and evaluates the input parameters, an action part that selects a path according to the preference from the result of the input parameters, and a Q-learning part that performs the evaluation based on the entire path. The input parameters obtained from the given environment are the TR, ES, and FS, which are link-related parameters described in the previous section, and the path-related parameters include the HC and SPDT. The system evaluates the neighboring UAV nodes starting from the source node until it discovers the destination node within the fixed iteration number, and only link-related parameters are used as inputs of the fuzzy system to select the next link. After finding the destination node, the system inputs path-related parameters to the fuzzy system. The fuzzy system combines individual input parameters for selecting a routing path to the destination based on the evaluation. Finally, the Q-learning assigns the Q-values about the HC and SPDT to the nodes for each link that constitutes the selected path from the source to the destination. The path-related Q-values (Q^h and Q^t for HC and SPDT, respectively) assigned to each node is used to evaluate each link by inputting it into a fuzzy system with the link-related parameters in the following assessment.

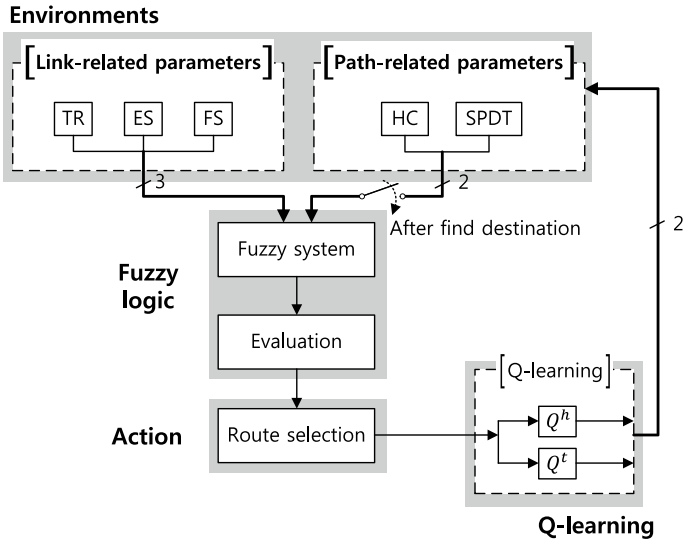

Fig. 2 Schematic of the proposed algorithm

Here, the Q-learning considers the parameters obtained through each selected path in the entire routing path as an observed environment. Each selected UAV node is set to a state, and the previous node (state) determines an action as the next node (state) to deliver the data packet.

Figure 3 describes how the system works based on Fig. 2. The data packets attempt to find a way to reach the destination node (D) starting from the source node (S), and the small circles on the link represent the UAV nodes. First, it starts from the source node, as in Fig. 3a, and evaluates the preferences about the peripheral nodes in the fuzzy system based on the link-related parameters and then select one of the neighboring nodes. This link-based evaluation is performed repeatedly until the destination node is found

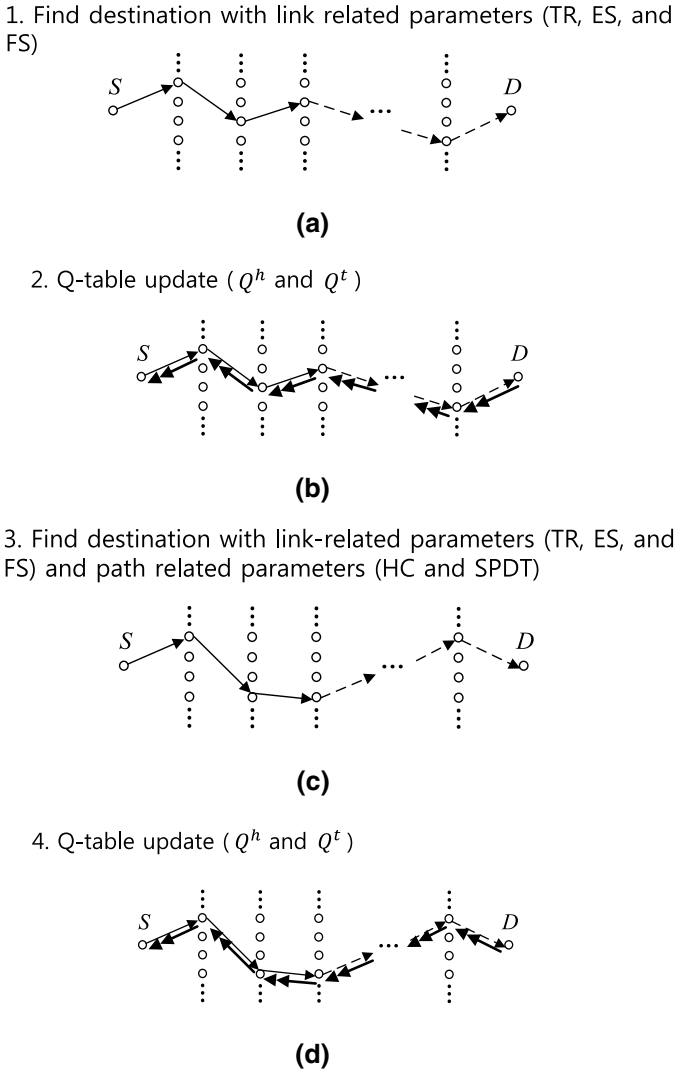

Fig. 3 Operating process of the proposed system

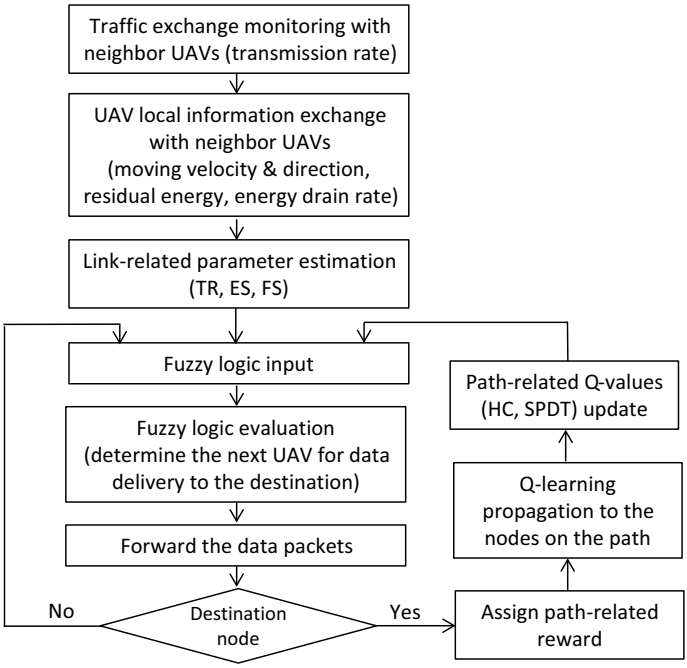

Fig. 4 Overall procedure of the proposed mechanism at each UAV node

within a certain number of iterations of the algorithm. After determining the destination node using the first process of Fig. 3a, each Q-value for SPDT and the HC is assigned to each node from the destination node to the source node direction for the corresponding path, as shown in Fig. 3b. After allocating the Q-values for the path, the destination node is again found using link-related parameters and path-related parameters, as shown in Fig. 3c. The routing path found in the third process may be different from the routing path found in the first and second processes. After finding a path from Fig. 3c to the destination node, each Q-value for the HC and SPDT is assigned to each node as shown in Fig. 3d; the third process is then returned to and repeatedly performed, as shown in Fig. 3c, d.

Therefore, the overall process of the proposed system can be explained as follows. First, it finds a path to reach the destination node based on an individual link performance from the source node. However, this routing path is not considered in terms of the performance of the network as only the individual link performance is considered. Therefore, we evaluate the path-related performance of the found routing path, and the performance on the network wide is considered when nodes find the link for the next time.

The general reinforcement learning method can be used to determine the state of the system using the environmental parameter and the appropriate operation is then performed. The reward for the corresponding action is obtained and this process is repeated. However, the proposed algorithm allows two aspects of link and path performance to be considered when searching for the routing paths in the network and allows the input parameters to be integrated into meaningful values using fuzzy logic rather than them being discretely used. In addition, Q-learning provides a basis for the evaluation of each link in terms of the overall path. In the following section, we describe the

method for generating each input parameter and assigning the Q-value and the method for evaluating the fuzzy logic link preference using input parameters and Q-values.

Figure 4 shows the overall procedure of the proposed mechanism that is performed at each UAV node.

4 Q-Learning-Based Fuzzy Logic Routing Algorithm

4.1 Link-Related Parameter Design

In this section, we design the link-related parameters that are input directly into the fuzzy system. According to Fig. 2, the fuzzy variables include the ES, FS, TR, HC, and SPDT. Among them, the TR is used as the original data transmission characteristic without any modification. The ES for the link from the current node i to the neighbor node j is defined as $ES_{i,j}$.

$$ES_{i,j} = \beta_1 \times \frac{RE_j}{IE} - \beta_2 \times EDR_j \tag{1}$$

where RE_j is the residual energy at node j, IE is the initial energy that is assumed as the fixed value for all the nodes, EDR_j presents the energy drain rate at node j, and β_1 and β_2 are the scale factors for the equation. Therefore, by subtracting the energy consumption rate from the current remained energy ratio, the lower the value obtained, the shorter will be the life expectancy of the node.

Figure 5 shows an example topology to explain the FS. Node i, which moves at an angle of θ_i, attempts to transmit data. Node j, which is a neighboring node of node i, moves at an angle of θ_j and a speed of V_j. Node k is also a peripheral node of node i and is similarly described. The projection of the velocity vector on the direction of the destination node d from j or k can then be represented by P_j^d and P_k^d for each node j and k. For node i to deliver packets quickly and stably to the destination node, it is better to select a node that moves in a direction similar to node i and quickly approaches to the destination node among the surrounding nodes. Therefore, in Fig. 5, it is desired that node j be selected rather than node k, and the FS for the link from current node i to neighbor node j is described as follows:

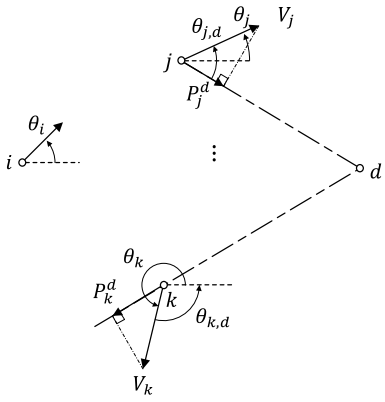

Fig. 5 Topology scenario of FS calculation

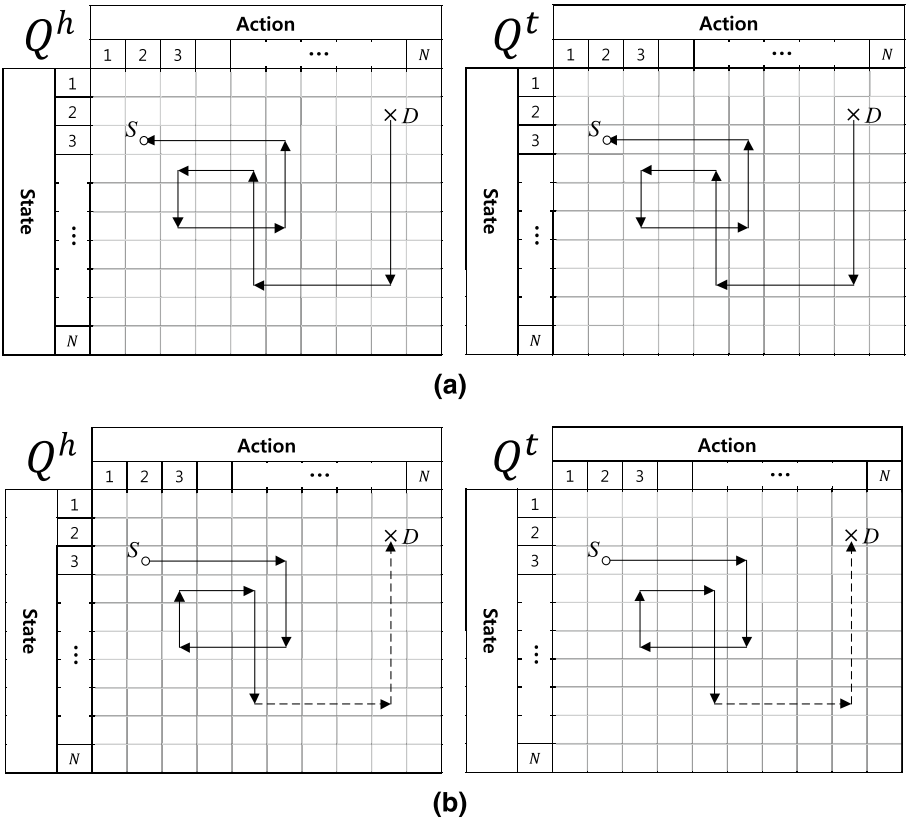

Fig. 6 Operation and learning process of the Q-table

$$FS_{i,j} = \cos\left(\theta_i - \theta_j\right) P_j^d \tag{2}$$

where $\cos\left(\theta_i - \theta_j\right)$ represents the similarity of the flight direction of the transmission node with that of the adjacent node. The velocity projection vector P_j^d of node j on the destination node d is expressed using the Eq. (3).

$$P_j^d = V_j \cos\theta_{j,d} \tag{3}$$

where V_j is the current velocity of node j, and $\cos\theta_{j,d}$ is the angle between the moving vector of node j and the vector from node j to the destination node d. $\cos\theta_{j,d}$ is expressed as follows:

$$\cos\theta_{j,d} = \frac{J \cdot D}{\|J\|\|D\|} \tag{4}$$

where J and D are the coordinates of nodes j and d, respectively, and $\|\cdot\|$ represents 2-norm.

4.2 Path-Related Parameter Design with Q-Learning

In this section, the path-related parameters that are inputted to the fuzzy system after reaching the destination node using Q-learning are discussed. The path-related parameters, HC and SPDT, are learned through each Q-learning and are allocated to each node as a Q-value. As explained in Figs. 2 and 3, the update of the Q-table can be expressed as shown in Fig. 6. The left and right figures of Fig. 6a, b represent the Q-table ($\mathcal{Q}^h$ and $\mathcal{Q}^t$) for learning the Q-values related to HC and SPDT, respectively. The row and column of the table are defined as the state and action, the state represents a node for transmitting a data packet, and the action represents a node for receiving a data packet. Figure 6a shows the process of updating the Q-value of the nodes in the corresponding routing path after finding the destination node using the link-related parameter, and Fig. 6b shows the process of visiting the destination node using the updated Q-value along with the link-related parameter. The dotted line of Fig. 6b indicates that it may find a different path than Fig. 6a, in which actual routing path decision is not only based on the Q-values but also other link related parameters. The proposed fuzzy logic in Fig. 2 combines link and path level parameters.

Q-learning is an off-policy technique for which no information is provided in advance and is a type of reinforcement learning that performs learning based on the Markov decision rule [21]. A system called an agent can learn about the environment experienced by executing an action and thus execute the dynamically appropriate operation according to the intended purpose. The system performs a specific action in a state to which it belongs and obtains a corresponding reward. Subsequently, the reward is combined with the Q-value previously calculated in the corresponding state, and the expected Q-value is obtained in a new state after the action is performed. The combined Q-value replaces the Q-value for the state and action. This general Q-learning process can be represented by Eq. (5).

$$\mathcal{Q}(s_t, a_t) = (1-\alpha)\mathcal{Q}(s_t, a_t) + \alpha\left\{ r_t + \gamma \max_{a_{t+1}} \mathcal{Q}(s_{t+1}, a_{t+1}) \right\} \tag{5}$$

where s_t and a_t represent the state and action of the system at time t, and r_t is a reward obtained by the action a_t. α is the learning rate, and γ is the discount factor. The first term represents the previous Q-value in the corresponding state and operation, and the third term is the maximum (expected) reward that can be obtained by action a_{t+1} among the possible actions when the state is changed to s_{t+1}. Therefore, α is the mixing ratio of $\mathcal{Q}(s_t, a_t)$, which is the previously learned information, r_t is obtained by performing the current action, and $\max_{a_{t+1}} \mathcal{Q}(s_{t+1}, a_{t+1})$ is expected to be obtained in the future. Moreover, γ determines the extent to which the expected reward will be reflected in the Q-value, and the synthesized result replaces the previous Q-value.

It is likely that this general Q-learning process will be applied as shown in Fig. 6b when applied to the proposed method. However, if there is a node that leads to a place that is far from the destination while searching for a path from the source node to the destination node, a lengthy route and a considerable amount of time is required to reach the destination node. This is disadvantageous in terms of the required time and the HC performance for the overall routing path. Also, even a path is short and fast, some links of the path may suffer bad link quality. Therefore, the proposed system finds a destination node using a fuzzy system with the link-related parameters as inputs, and then spread the Q-value of the HC and SPDT from the destination node in

the reverse direction to the previous nodes, as shown in Fig. 6a. We then consider both performances of the link and the path side in determining the routing path by inputting the Q-values and link-related parameters allocated in the nodes to the fuzzy system.

The proposed method is started by assuming that the previous node of the destination UAV node is a node m for allocating the Q-value from the destination node to the source node. There are two types of Q-values, which are the HC and SPDT. All the Q-values are initially set to arbitrary values. First, in relation to the Q-learning for the HC, the update of the Q-value for the node m just before the end node is given as follows:

$$\mathcal{Q}^h\left(m, a_m\right) = \left(1-\alpha^h\right)\mathcal{Q}^h\left(m, a_m\right)' + \alpha^h R^h \tag{6}$$

where $\mathcal{Q}^h\left(m, a_m\right)'$ represents the Q-value before the node m executes the action a_m, and R^h is a predefined terminal reward given to the last destination UAV node. α^h is a scale factor that controls the influence of the previous Q-value and terminal reward.

After determining the rewards of node m and R^h of the terminal, the Q-value is updated in the order of the source node direction, which is given as follows:

$$\mathcal{Q}^h\left(m', a_{m'}\right) = \left(1-\alpha^h\right)\mathcal{Q}^h\left(m', a_{m'}\right)' + \alpha^h\left\{\gamma \max_{a^*} \mathcal{Q}^h(m, a^*)\right\} \tag{7}$$

where m' is a node just before the node m, and $\mathcal{Q}^h\left(m', a_{m'}\right)'$ is the previous Q-value before the node m' executes the action $a_{m'}$. $\max_{a^*} \mathcal{Q}(m, a^*)$ is a maximum Q-value among the nodes (actions) that the node m(the next state of the current state m') can select. This Q-value update is repeated until it reaches to the source node.

With respect to the Q-learning of the SPDT, the Q-value for the node m just before the destination node is calculated as follows:

$$\mathcal{Q}^t\left(m, a_m\right) = \left(1-\alpha^t\right)\mathcal{Q}^t\left(m, a_m\right)' + \alpha^t\left(R^t - t_{m,d}\right) \tag{8}$$

where $\mathcal{Q}^t\left(m, a_m\right)'$ represents the Q-value before the node m executes the action a_m and R^t is a predefined terminal reward given to the last destination UAV node for SPDT parameter. α^t is a scale for SPDT. The overall framework is similar to that in Eq. (6). $t_{m,d}$ is the time taken for node m to successfully deliver packets to the destination node d.

After updating the Q-value of the SPDT for node m and destination node d, the influence should be propagated in the source node direction, and it is represented as shown in Eq. (9).

$$\mathcal{Q}^t\left(m', a_{m'}\right) = \left(1-\alpha^t\right)\mathcal{Q}^t\left(m', a_{m'}\right)' + \alpha^t\left\{\gamma \max_{a^*} \mathcal{Q}^t(m, a^*) - t_{m',m}\right\} \tag{9}$$

where the parameters are the same as those in Eq. (7), except that $t_{m',m}$ is the time required to successfully deliver the packet from node m' to m.

Therefore, in the two Q-learning algorithm of the HC and SPDT, the rewards R^h and R^t given to the final terminal are propagated to node m using Eqs. (6) and (8), and the previous nodes on the routing path are affected in turn through Eqs. (7) and (9), respectively. The updated path-related Q-value is input to the fuzzy system with the link-related parameters to aid in evaluating the priority of the node when determining the path from the source node to the destination node, and the following section describes the fuzzy system in detail.

4.3 Fuzzy Controller Implementation

This section describes how the fuzzy system initially uses link-related parameters to identify a routing path from the source node to the destination node, and how the link- and path-related parameters are used in the fuzzy system after updating the Q-values of the HC and SPDT. The fuzzy logic is based on the fuzzy set introduced by Zadeh to quantitatively express ambiguousness such as that of a natural language. A fuzzy set is a set of members that indicate how much the input parameters belong to each classification and mark the input without distinguishing whether the object of interest belongs to a classification by binary logic. A fuzzy logic system can be considered to be an expert system that encompasses a set of linguistic fuzzy rules. The fuzzy rules follow this general pattern: If premise(s) Then conclusion(s). In a fuzzy rule, the premises and conclusions correspond to the fuzzy input and output sets respectively. The fuzzy inputs in this paper are link- and path-related parameters, and the output is the selection priority among the peripheral nodes for receiving data packets.

The fuzzy system structure of the proposed method is shown in Fig. 7. Fuzzy control is performed through fuzzification, fuzzy inference, and defuzzification. Fuzzification is the process of changing the exact input (crisp) values measured in the system into each membership function using the fuzzy rules. The fuzzy inference infers the fuzzy result with the fuzzy rules and membership values obtained by the fuzzification. Finally, the defuzzification converts the fuzzy output into an equivalent crisp value for use in the system.

First, the values of the parameters (crisp values) described in Sects. 4.1 and 4.2 are input into the fuzzy module and converted to each membership function value. Figure 8 represents the membership functions of the link-related parameters (TR, ES, and FS), path-related input parameters (HC and SPDT), and output. Figure 8a presents the membership function of the TR, which presents the median level from 0 to 0.8 and high level from 0.6 to the top value of 1. Figure 8b shows the membership function of the ES, which denotes a low level from 0 to 6 and high level from 4 to 10. It is assumed that the flying velocity of the UAV is between 5 and 15 m/s, and the membership function of the FS is given a bad level from −15 to 5 and good level from −5 to 15, as shown in Fig. 8c. Figure 8d presents the membership function of the HC, which is designed from 1 to 50 per trail. They are separated as small and large statuses before the value of 10 and after the value of 5. Figure 8e represents the membership function of the SPDT on one trail from 0.01 to 0.5, which is divided as a short and long status before the value of 0.05 and after the value of

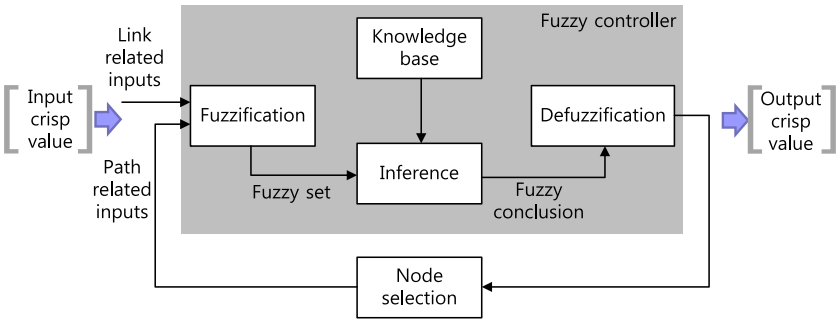

Fig. 7 Proposed fuzzy system structure

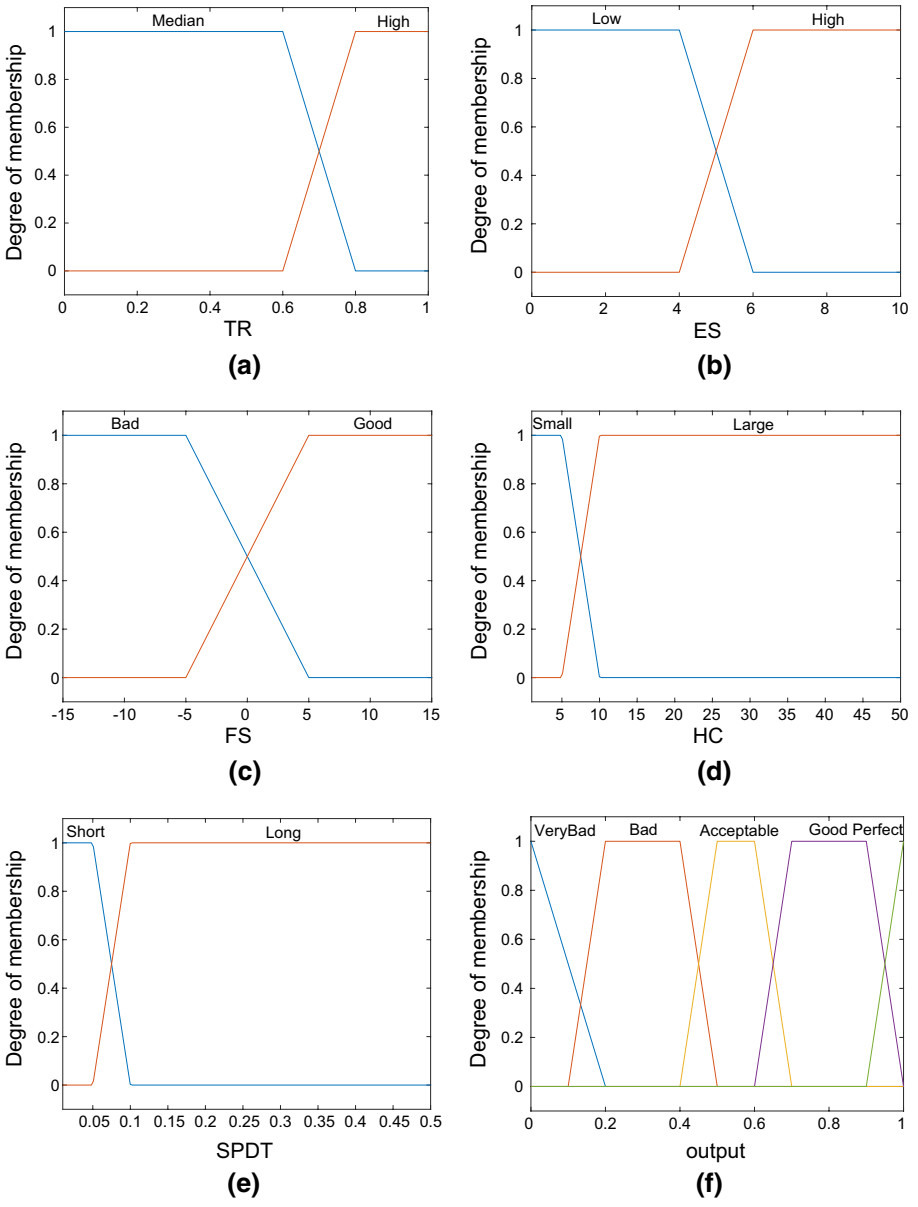

Fig. 8 Membership functions: **a** TR; **b** ES; **c** FS; **d** HC; **e** SPDT; **f** priority index

0.1, respectively. Meanwhile, Fig. 8d is used to combine these inputs to calculate the crisp output value in the defuzzification module of Fig. 7.

When the crisp values of each parameter are converted to a fuzzy set using the membership functions, an IF–THEN rule is required to determine the corresponding results. These rules are supplied in the knowledge base module, and Tables 1 and 2 represent the fuzzy rule of the proposed fuzzy system. Table 1 is a rule used to determine a routing path

Table 1 IF–THEN rules for the fuzzy system for initial route selection

No.	TR	ES	FS	Output
1	M	L	B	O_VB
2	H	L	B	O_B
3	M	H	B	O_B
4	H	H	B	O_ACT
5	M	L	G	O_B
6	H	L	G	O_ACT
7	M	H	G	O_G
8	H	H	G	O_P

M, Median; H, High; L, Low; B, Bad; G, Good; O_VB, Very Bad; O_B, Bad; O_ACT, Acceptable; O_G, Good; O_P, Perfect

through link-related parameters until the destination is initially found, and Table 2 is a rule for determining a routing path using updated path-related Q-values. As can be observed in the tables, human logic is involved in the design. For example, IF TR is high, ES is high, and FS is good, THEN the output of this link reliability is perfect. In this manner, the inference engine obtains fuzzy conclusions using the IF–THEN rule. In addition, the fuzzy conclusions are input to the defuzzification module and converted into a crisp value for the system to use as an output, and in the proposed system, this value is used as a priority index to select one of the peripheral nodes. Therefore, this operates as an action from the Q-learning point of view.

Figure 9 presents an example of a computational process for an operation process in which the fuzzy system initially uses link parameters to find a route to a destination node. We use the Mamdani method, each row represents each rule of Table 1, and three columns represent each input parameter. For example, in the case of TR = 0.7, ES = 5, and FS = 2, a fuzzy value corresponding to each crisp input value is found, and the value for the output of the IF–THEN rule corresponding to each row is assigned—the plots of the fourth column represent this. The assigned value in the output is allocated as a minimum value among the input membership function values of each input parameter by the AND fuzzy rule. The graph of the right lower end is obtained as the final output. General centroid method is used to obtain the crisp output value of the final graph in the defuzzification module.

5 Simulation Results and Discussion

In this section, the performance of our proposed algorithm is evaluated in terms of the HC, ES, FS, and TR for different topologies and node numbers. Comparisons are performed among our proposed Q-learning based fuzzy system, original fuzzy system, and AODV-based Q-learning routing protocols. Table 3 presents the simulation parameters.

Figure 10 represents the average HC trend for 100 different topologies during 100 trails. One hundred different topologies provide a more generous environment of packet transmission for verifying the reliability and feasibility of the Q-learning in our algorithm. Overall, it can be observed that the average HC decreases as the iteration increases. However, the HCs increase briefly at 15 and 55 trains, which is due to the learning process of Q-learning for the changes of topologies. After the learning, the average HC is again reduced rapidly.

Table 2 IF–THEN rules for the fuzzy system after destination arrival

No.	TR	ES	FS	HC	SPDT	Output
1	M	L	B	LG	LN	O_VB
2	H	L	B	LG	LN	O_B
3	M	H	B	LG	LN	O_B
4	H	H	B	LG	LN	O_B
5	M	L	G	LG	LN	O_B
6	H	L	G	LG	LN	O_B
7	M	H	G	LG	LN	O_ACT
8	H	H	G	LG	LN	O_ACT
9	M	L	B	SM	LN	O_B
10	H	L	B	SM	LN	O_B
11	M	H	B	SM	LN	O_B
12	H	H	B	SM	LN	O_B
13	M	L	G	SM	LN	O_B
14	H	L	G	SM	LN	O_B
15	M	H	G	SM	LN	O_ACT
16	H	H	G	SM	LN	O_G
17	M	L	B	LG	SH	O_B
18	H	L	B	LG	SH	O_B
19	M	H	B	LG	SH	O_B
20	H	H	B	LG	SH	O_B
21	M	L	G	LG	SH	O_B
22	H	L	G	LG	SH	O_B
23	M	H	G	LG	SH	O_ACT
24	H	H	G	LG	SH	O_G
25	M	L	B	SM	SH	O_B
26	H	L	B	SM	SH	O_B
27	M	H	B	SM	SH	O_B
28	H	H	B	SM	SH	O_B
29	M	L	G	SM	SH	O_B
30	H	L	G	SM	SH	O_B
31	M	H	G	SM	SH	O_G
32	H	H	G	SM	SH	O_P

M, Median; H, High; L, Low; B, Bad; G, Good; LG, Large; SM, Small; LN, Long; SH, Short; O_VB, Very Bad; O_B, Bad; O_ACT, Acceptable; O_G, Good; O_P, Perfect

The HC has a high value of over 20 up to 20 trains, but decreases sharply by half after 30 trains, increases slightly, and again shows a tremendous decrease. Therefore, it can be observed that the topology can have a lower HC on using the proposed Q-learning, and even if the topology changes during the data transmission, a good routing path with a low HC based on the Q-value is found.

Figure 11 shows 100 different topologies that are used to obtain the average FS during 100 successful packet transmission trails. It can be observed from the figure that the FS values have a relatively stable trend without negative values. This means that the topology

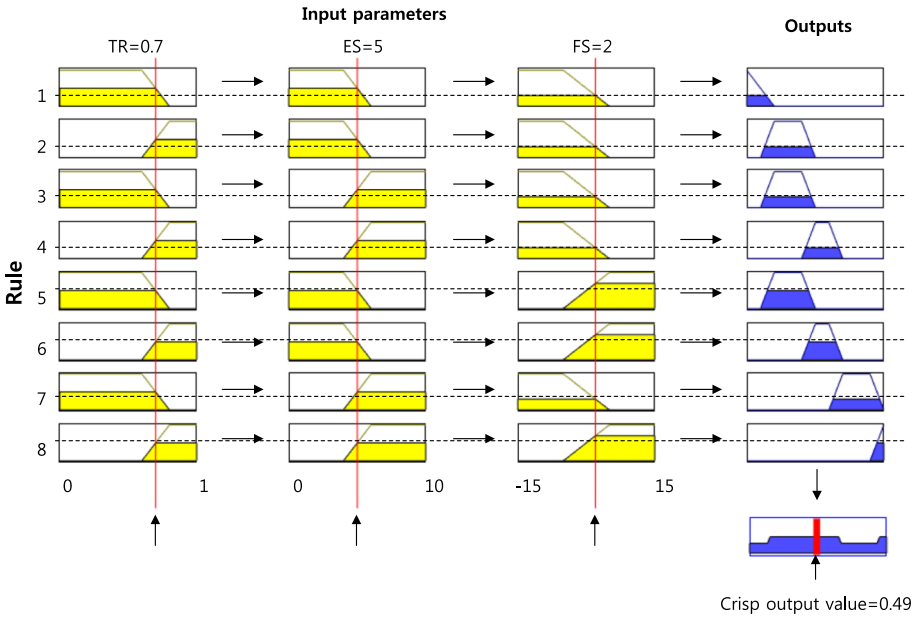

Fig. 9 The fuzzy process example

Table 3 Simulation parameters

Group	Parameter values
Network environments	Topology area: 100 m × 100 m Number of nodes (N): 15 Source node coordination: [1, 1] Destination node coordination: [90, 90]
Node characteristics	Initial energy (IE): 10 Communication range: 45m Energy drain rate range per second: 0–0.05 Energy consumption rage for one transmission 0–0.01 Velocity range: 5−15 m/s TR range: 0−1(1 = 100%) UAV flying angle: − 180° to 180° SPDT per trail: 0.01 s
Fuzzy variable and Q-learning parameters	Scale factors β_1 and β_2: 0.5 Learning rate γ: 0.7 Q-learning factors α^d and α^l: 0.5

change will not affect the FS values significantly on using our proposed method. This FS can thus be used to ensure that the moving trend of the UAV nodes in our selected routing path is positive and stable.

In the following experiment, we compare our proposed method with the original fuzzy system and the Q-value-based AODV. Figure 12 presents the average TR of the nodes on the path according to the successful packet transmission trails. As can be observed, within the first 50 trails, the fuzzy system and Q-value-based fuzzy system cause the TR to remain

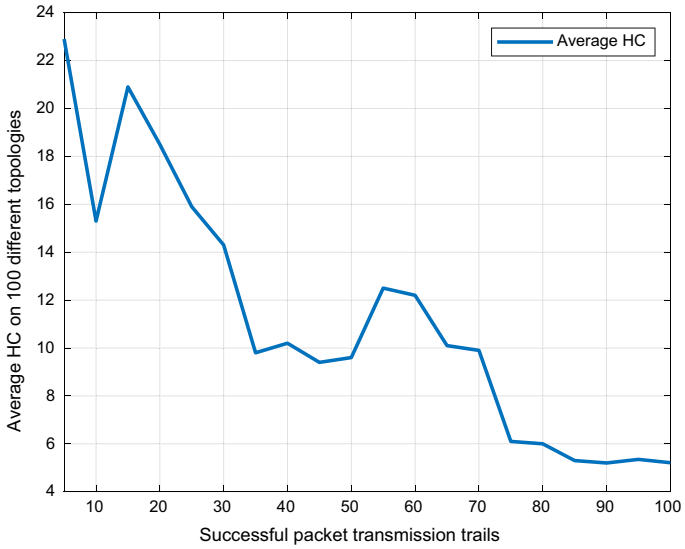

Fig. 10 Average HC for 100 different topologies during 100 trails

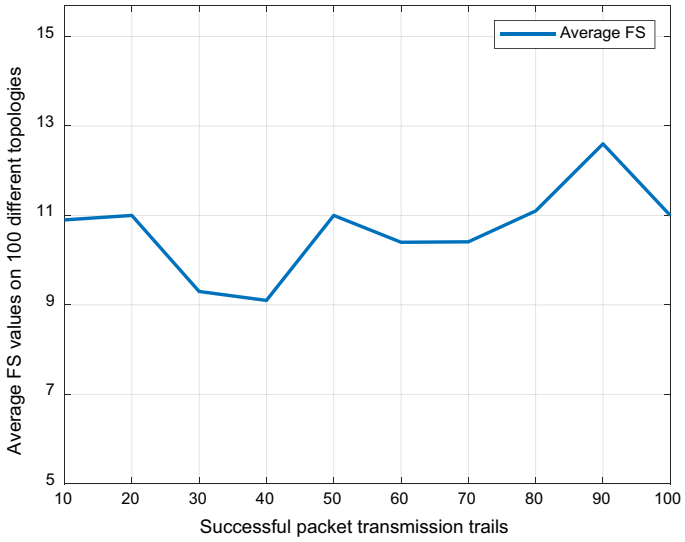

Fig. 11 Average FS for 100 different topologies during 100 trails

over 5 while the TR of the Q-value-based AODV is less than 5. Within 100 trails, the three methods are found to be similar but their results differ after 100 trails. The TR of the fuzzy system decreases significantly at 150 trails and increase sharply at 200 trials owing to the node disconnection and reconnection, while the TR of the Q-value-based fuzzy system algorithm increases significantly at 150 trails and remains relatively constant at 200 trails while considering the effect of the Q-value. The TR of the Q-value-based AODV

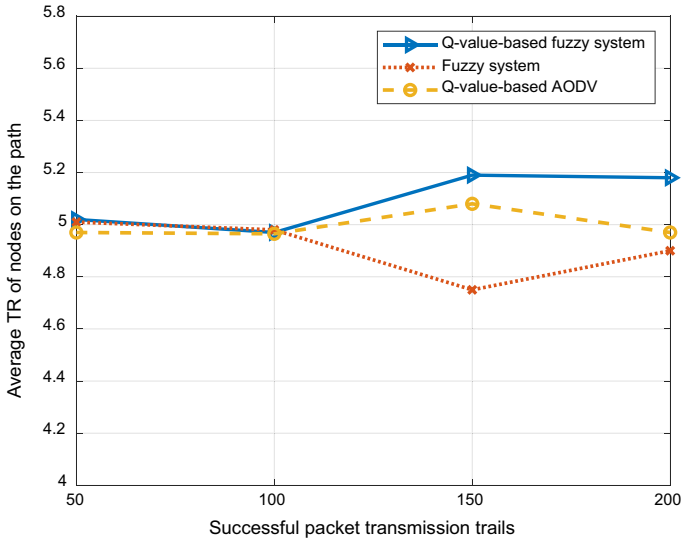

Fig. 12 Average TR of nodes on the path under 200 trails

is also increased but not by much as compared to our proposed method, and decreases at 200 trails. Hence, our proposed method can be used to achieve a better capacity during the change in topology during UAV node flying.

Considering that the number of nodes may significantly influence the topology and network performance, the number of nodes is changed over the range $\{10, 15, 30\}$, and other parameters are maintained at a constant, as shown in Table 3. The simulation results are displayed in Figs. 13, 14 and 15. We evaluate the performance of the average HC and average remained energy of nodes on the path. If the number of nodes in the network is 10, the routing path is likely to be cut off between the source node and the destination node owing to the transmission coverage limit and the disconnection problem caused by the topology change. In Fig. 13a, the average HC of the fuzzy system is higher than those of the other

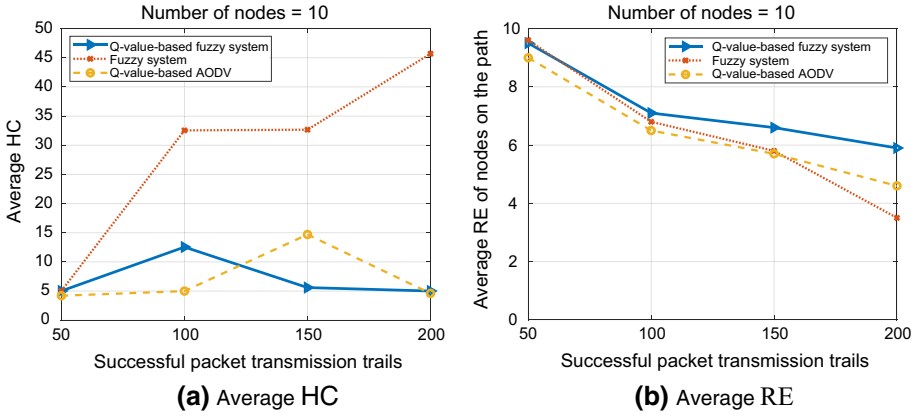

Fig. 13 Comparison of the results when the number of nodes equals 10 under 200 trails

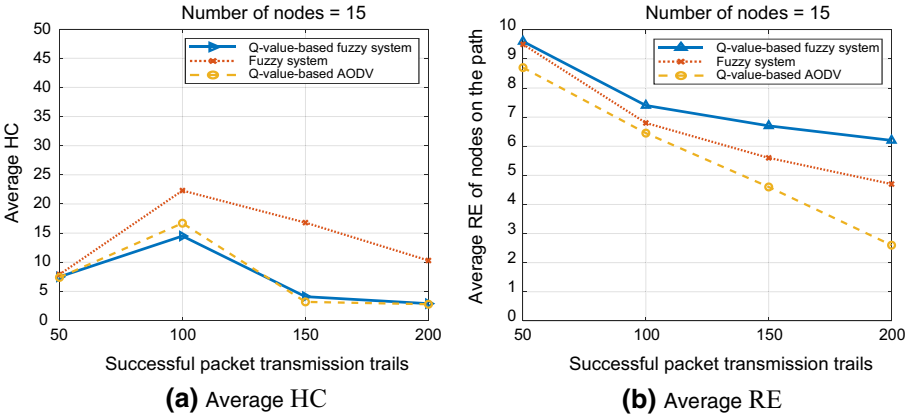

Fig. 14 Comparison results when the number of nodes equals 15 under 200 trails

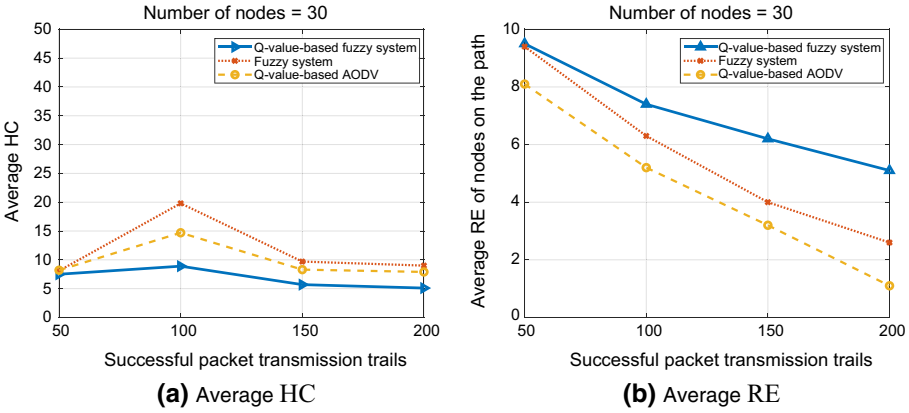

Fig. 15 Comparison results when the number of nodes equals 30 under 200 trails

two methods. At the same time, the energy remaining on using the fuzzy system is less than that in the case of the others', as shown in Fig. 13b. When we increase the number of nodes to 15, as compared with Fig. 13a, Fig. 14a shows a lower HC of the fuzzy system but that is still higher than those of the other two methods. However, in Fig. 14b, the average remaining energy of the Q-value-based AODV decreased to a greater extent than that in the case of the fuzzy system owing to the broadcasting of messages to all the neighbors, and we can observe the same trend in Fig. 15b—the average remaining energy of the proposed method decreases to a greater extent when the number of nodes increases. In addition, in Figs. 13a and 14a, the HC could be greater than the number of nodes owing to the mobility of the topology.

In summary, it can be clearly confirmed the merits of the proposed method in the average HC because it takes into consideration the entire path and finds a short routing path by using Q-learning. In the case of a large number of nodes, the overall HC decreases and the performance difference between the proposed and conventional method is small. In contrast, when the number of available nodes in the network is small, the performance

enhancement increased. In terms of the average residual energy, the proposed method is not outstanding when the number of nodes is small because the nodes consuming power could be overlapped. However, when the number of nodes is large, the performance of the proposed method is evidently excellent. In such situations, irrespective of the average HC or average remaining energy, our proposed method has apparent advantages over the other aforementioned two methods.

6 Conclusions

In this paper, we have proposed a Q-learning based fuzzy logic to implement a multi-objective routing algorithm in flying ad-hoc networks. To capture the short-term and long-term UAV network characteristics, we have defined link-related parameters and path-related parameters, Transmission rate, energy state and flight status are considered for evaluating link quality of all neighbor UAV nodes. The proposed Q-learning mechanism evaluates the routing path used in terms of hop count and successful packet delivery time. The Q-values for two path related parameters are dynamically updated whenever a packet is delivered to the destination. The proposed fuzzy logic effectively combines the short time/range scale link-related parameters and the long time/range scale path-related parameters. Using the defined fuzzy membership functions and fuzzy rules, when a UAV node receives a data packet it determines the best next UAV node to deliver the packet to the destination. Therefore, the proposed Q-learning based fuzzy logic is able to adapt to the frequent UAV network topology changes and achieve the multi-objectives required in FANET.

We set 100 different topologies to verify the effectiveness of the proposed Q-learning-based fuzzy system. Performance comparisons were conducted with the original fuzzy system and Q-value-based AODV. For many different initial network topologies and frequent topology changes in a simulation, the proposed routing algorithm provided higher average transmission rate and shorter hop count than those of the compared methods. The proposed method also had larger average residual energy so that it can prolong the network lifetime.

Acknowledgements This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT under Grant 2017R1A2B4003512.

References

1. Yang, Q., & Yoo, S.-J. (2018). Optimal UAV path planning: Sensing data acquisition over IoT sensor networks using multi-objective bio-inspired algorithms. *IEEE Access, 6,* 13671–13684.
2. Bekmezcia, İ., Sahingoza, O. K., & Temelb, Ş. (2013). Flying ad-hoc networks (FANETs): A survey. *Ad Hoc Networks, 11,* 1254–1270.
3. Li, X., & Yan, J. (2017). LEPR: Link stability estimation-based preemptive routing protocol for flying ad hoc networks. In *Proceedings of IEEE symposium on computers and communications (ISCC)* (pp. 1–6).
4. Dengiz, O., Konak, A., & Smith, A. E. (2011). Connectivity management in mobile ad hoc networks using particle swarm optimization. *Ad Hoc Networks, 9,* 1312–1326.
5. Zadeh, L. A. (1988). Fuzzy logic. *Computer, 21,* 83–93.
6. Amiri-Doomari, S., Mirjalily, G., & Abouei, J. (2018). Joint load balanced stable routing and communication segment assignment in mobile cognitive radio ad-hoc networks. *International Journal of Communication System, 31,* 1–24.

7. Marwaha, S, Srinivasan, D., Tham, C. K., & Vasilakos, A. (2004). Evolutionary fuzzy multi-objective routing for wireless mobile ad hoc networks. In *Proceedings of the congress on evolutionary computation* (pp. 1964–1971).
8. Doherty, P., Haslum, P., Heintz, F., et al. (2007). A new QoS routing algorithm based on self-organizing maps for wireless sensor networks. *Telecommunication Systems, 36,* 73–83.
9. Maksimović, M., Vujović, V., & Milošević, V. (2014). Fuzzy logic and wireless sensor networks—A survey. *Journal of Intelligent & Fuzzy Systems, 27*(2), 877–890.
10. Li, G., Ma, M., Liu, C., & Shu, Y. (2017). Adaptive fuzzy multiple attribute decision routing in VANETs. *International Journal of Communication Systems, 30*(4), 1–20.
11. Das, S. K., Tripathi, S., & Burnwal, A. P. (2015). Fuzzy based energy efficient multicast routing for ad-hoc network. In *Proceedings of the third international conference on computer, communication, control and information technology (C3IT)* (pp. 1–5).
12. Lai, W. K., Lin, M.-T., & Yang, Y.-H. (2015). A machine learning system for routing decision-making in urban vehicular ad hoc networks. *International Journal of Distributed Sensor Networks, 2015,* 1–13.
13. Sun, R., Tatsumi, S., & Zhao, G. (2002). Q-MAP: A novel multicast routing method in wireless ad hoc networks with multi-agent reinforcement learning. In: *Proceedings of 2002 IEEE region 10 conference on computers, communications, control and power engineering* (pp. 667–670).
14. Arroyo-Valles, R., Alaiz-Rodriguez, R., Guerrero-Curieses, A., & Cid-Sueiro, J. (2007). Q-probabilistic routing in wireless sensor networks. In *Proceedings of 3rd international conference on intelligent sensors, sensor networks and information* (pp. 1–6).
15. Dong, S., Agrawal, P., & Sivalingam, K. (2007). Reinforcement learning based geographic routing protocol for UWB wireless sensor network. In *Proceedings of IEEE global telecommunication conference* (pp. 652–656).
16. Wu, C., Ohzahata, S., & Kato, T. (2013). Flexible, portable, and practicable solution for routing in VANETs: A fuzzy constraint Q-learning approach. *IEEE Transactions on Vehicular Technology, 62,* 4251–4263.
17. Chettibi, S., & Chikhi, S. (2016). Dynamic fuzzy logic and reinforcement learning for adaptive energy efficient routing in mobile ad-hoc networks. *Applied Soft Computing, 38,* 321–328.
18. Al-Kiyumi, R. M., Foh, C. H., Vural, S., Chatzimisios, P., & Tafazolli, R. (2018). Fuzzy logic-based routing algorithm for lifetime enhancement in heterogeneous wireless sensor networks. *IEEE Transactions on Green Communications and Networking, 2,* 517–532.
19. Saleh, A. I., Abo-Al-Ez, K. M., & Abdullah, A. A. (2017). "A multi-aware query driven (MAQD) routing protocol for mobile wireless sensor networks based on neuro-fuzzy inference. *Journal of Network and Computer Applications, 88,* 72–98.
20. Na, H. J., & Yoo, S.-N. (2019). PSO-based dynamic UAV positioning algorithm for sensing information acquisition in wireless sensor networks. *IEEE ACCESS, 7,* 77499–77513.
21. Alsheikh, M. A., Lin, S., Niyato, D., & Tan, H.-P. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials, 16,* 1996–2018.

Qin Yang received the B.E degree in communication and information engineering from Chongqing University of Posts and Telecommunications in 2016. She is currently working toward the M.S degree in information and communication engineering with Multimedia Network Laboratory, Inha University, South Korea. Her research interests include wireless sensor network, FANET, machine learning.

Sung-Jeen Jang received the B.S degree in electrical engineering from Inha University Incheon, Korea, 2007. He received his M.S. and Ph.D. degrees in Graduate School of Information Technology and Telecommunication, Inha University, in 2009 and 2019, respectively. Having been awarded his Ph.D. degree in Feb. 2019, he continued his work in Inha University as a postdoc. His current research interests include cognitive radio network protocols and machine learning applied wireless communications.

Sang-Jo Yoo received the B.S. degree in electronic communication engineering from Hanyang University, Seoul, South Korea, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, in 1990 and 2000, respectively. From 1990 to 2001, he was a Member of Technical Staff with the Korea Telecom Research and Development Group, where he was involved in communication protocol conformance testing and network design fields. From 1994 to 1995 and from 2007 to 2008, he was a Guest Researcher with the National Institute Standards and Technology, USA. Since 2001, he has been with Inha University, where he is currently a Professor with the Information and Communication Engineering Department. His current research interests include cognitive radio network protocols, adhoc wireless network, MAC and routing protocol design, wireless network QoS, and wireless sensor networks.