Research on Routing Algorithm of UAV AD Hoc Network Based on Decision Tree

Jian-qin LU* and Jian-guo NAN
Air Force Engineering University, Xi’an Shaanxi, China
*Corresponding author

Keywords: UAV AD Hoc, Routing algorithm, Decision tree, Delivery success rate.

Abstract. The efficient routing algorithm is the basis for ensuring the performance of the UAV AD hoc. Most of the current routing algorithms only consider a certain feature of the node, and often have network instability and large packet loss rate. Aiming at this problem, this paper proposes a routing algorithm based on decision tree, collects and analyzes various attribute characteristics of nodes, forms a decision-making scheme, and selects the transmission path with the highest comprehensive evaluation. Experiments show that the algorithm improves the delivery success rate of the packet.

Introduction

Wireless AD hoc network is a temporary network multi-hop system composed of mobile nodes with wireless signal transceiver. In AD hoc network, each node has equal status and can quickly network anytime and anywhere without relying on fixed infrastructure. Therefore, it is very suitable for uav cluster network. In unmanned aerial vehicle ad-hoc network (UAANET), in addition to the relay equipment such as satellite and ground station communication as a node, the node are unmanned aerial vehicles, because of the unmanned aerial vehicles communication distance is limited, there is no guarantee that each can be direct communication between two nodes, often a communication process may require several relay to complete, so it is particularly important to research an efficient routing algorithm.

In recent years, routing algorithms have developed rapidly, which often improve the efficiency of other aspects at the expense of the interests of one aspect. Clustering routing algorithm is a kind of relatively mature routing algorithm, it through electing cluster head nodes in the network, the communication is divided into communications within the cluster and cluster communication, to a certain extent, reduce the communication overhead, but the energy consumption of cluster heads, often require frequent replacement of cluster heads, destabilize the communication network. GPSR routing algorithm is also a popular routing algorithm at present. When carrying out peripheral forwarding, it is blind to some extent. It is greatly affected by node state.

The routing algorithm based on decision tree is studied in the text, using the ideas of machine learning to the next hop routing, in the process of path choice, give full consideration to the node properties, avoid the node is not stable, relatively fast transmission and reduced the partial node problems affect the entire network randomness.

Introduction of Decision Tree

Decision Tree is a common machine learning method. It first summarizes the feature attributes in the training data set. After clearing the criteria, the attributes are sorted by the attribute classification ability of the sample set to find a pair. The high-purity classification model of the training set, on which the corresponding decision-making scheme is designed according to different classification results, and finally the model is directly applied to the new task set.

Generally, a decision tree has a root node and a number of internal nodes and leaf nodes. The root node and the internal node represent a division of attributes. This division generally selects an optimal feature attribute and classifies the training set. Split into subsets so that the subset has the best
classification effect under current conditions. The leaf nodes represent a class or class distribution. Each node contains a sample set that is divided into child nodes according to the result of the attribute test, and the final sample is imported into all leaf nodes. Each path from the root node to the leaf node corresponds to a classification rule. The purpose of decision tree learning is to generate a decision tree with no strong examples, and the classification rules in the decision tree are based on the training set, so it is very practical.

**Decision Tree Algorithm Selection**

Decision tree is a top-down, divide-and-conquer classification process. In the process of decision tree growth, the selection of test attributes and the division of sample sets are particularly important. Different decision tree algorithms do not choose the criteria for classification.

**ID3 algorithm:** Calculate the information gain of all attributes of the node from the root node. The larger the information gain, the more information the test attribute provides for the classification. After the classification of the test attribute, the uncertainty of the sample is smaller. Therefore, the attribute with the largest information gain is selected as the node branching criterion, and the same test is performed on the child nodes until no attributes can be subdivided or all the attribute information gains are small, and finally the decision tree is obtained. The ID3 algorithm is the most classical decision tree algorithm. It tends to select attributes with more values, but this attribute is not necessarily optimal. If the lower limit of information gain is unreasonable, the attribute partition will be too fine.

**C4.5 algorithm:** The C4.5 algorithm overcomes the problem of attributes that tend to take more values in ID3. On the basis of information gain, the information gain rate is established. The information gain rate is the ratio of information gain to segmentation information. As a test attribute, an attribute with a large information gain rate can effectively reduce the preference for attributes with a large number of values. The disadvantage of this algorithm is that it chooses the attribute with the smallest entropy as the optimal attribute. When an attribute has only one attribute value, the information gain rate becomes meaningless, so the generated decision tree is still a multi-fork tree.

**CART algorithm:** In order to generate a more concise binary tree, the CART algorithm uses a two-point recursive segmentation technique. The test attribute divides the sample into two, so that the node of the decision tree has at most two branches, which improves the efficiency of generating the decision tree. The CART algorithm is different from the first two algorithms in terms of information entropy. It introduces the Gini coefficient. The Gini coefficient represents the impureness of the model. The smaller the Gini coefficient, the lower the purity and the better the characteristics. Therefore, the CART algorithm calculates the Gini coefficient of each attribute to be the smallest attribute of the Gini coefficient at the decision tree node as the partitioning attribute of the node. Compared with the entropy calculation, the Gini coefficient does not need to be logarithmically calculated and is more efficient. After the above comparison, this paper chooses CART algorithm as the basis of design routing algorithm.

**Evaluation Parameter**

When performing tasks in a drone cluster, UAANET usually includes the following elements: drone nodes, ground control stations, satellite stations, and unmanned unit network links. The ground control station is mainly used to assign tasks and control unmanned maneuvers to the drones. The satellite stations are mainly used to provide geographic location information to the drones, and the unmanned unit network links are established between the drones. In UAANET, whether the pre-defined task and path planning algorithm in the UAV is artificially controlled by the ground station, the UAV's self-organizing network has a certain mapping relationship with the algorithm or human thinking. Therefore, we can use this relationship to study a routing algorithm to provide better protection for communication between drones.

Definition 1: The sample nodes are represented by $S_i$ (i = 1, 2...n), and the set of characteristic of the UAV node is described by the set $A$ as
\[ A = (a_1, a_2, \ldots, a_m) \]  

**Definition 2:** In order to find the attribute that can best divide the sample set, for each attribute in the feature attribute set of the UAV node, calculate the GINI coefficient of each possible division on each attribute, and find the division with the smallest GINI coefficient. The value is used as the optimal partitioning value of the attribute, and then the GINI coefficient of all the attributes on the optimal dividing value is compared, and the attribute with the smallest GINI value is used as the final test attribute. The GINI coefficient is defined as:

\[
Gini(p) = 1 - \sum_{k=1}^{m} p_k^2
\]  

\[ p_k \] is the proportion of samples satisfying the attribute \( a_m \) in the total sample \( S \).

**Definition 3:** In the binary recursive segmentation technique, each time a certain attribute value \( a \) in the attribute set \( A \) is taken, the sample set is divided into two sub-sample sets, so that each node of the decision tree has two branches. When the attribute value \( a \) is divided into \( S1 \) and \( S2 \), the Gini coefficient is defined as:

\[
Gini(S, A) = \frac{S1}{S} Gini(S1) + \frac{S2}{S} Gini(S2)
\]

**Attribute Division**

Feature attribute division criteria: Select features that have strong classification ability for training data. If the classification result of a feature is not different from the result of the random classification, the feature is said to have no classification ability. Based on the actual application of UAANET, the selected attributes are as follows:

1. **Close to the center.** Near-centrality is used to describe the shortest path length from each node to other nodes. The near-centrality is mainly measured by the reciprocal of the sum of the distances from a node to all other nodes. Close to the center of the network, selecting a node that is close to the centrality as a relay node can effectively reduce network overhead and packet transmission delay.

2. **Intermediary intermediate.** Refers to the number of times a node acts as a relay node during packet transmission between the other two nodes. The better the intermediateity of the intermediary, the higher the number of times a node acts as an “intermediary”, and the greater its effect on the data communication of nodes in the entire network.

3. **Extensiveness.** Each node in the network needs data transmission, and a communication link needs to be established between each other. The more links established by a node, the higher its popularity. Extensiveness can measure the coverage of a node to the entire network. This paper uses the number of links established to measure the breadth of network nodes.

4. **Energy margin.** For UAVs, energy is always an important factor that restricts the efficiency of its missions. Nodes with large energy margins can take on more tasks in the network, and better maintain the network while reducing the routing overhead of other nodes. When the energy of the UAV node is lacking, it will reduce the overhead of all aspects and inevitably affect the quality of data transmission. This paper defines the relative residual energy index as the measure of the energy surplus of the drone.

5. **Speed similarity.** Speed similarity describes the degree of speed similarity between a node and a neighbor. The greater the speed similarity, the smaller the relative mobility of the node \( m \) relative to the neighboring neighbor nodes. The more favorable the link establishment and data transmission, the smaller the speed similarity node can be used as the relay node to improve the data packet delivery.

6. **Congestion degree:** When the source node transmits a data packet to the destination node, if a node on the transmission path is overloaded or the cache queue is full, the node cannot receive the new data packet, resulting in packet loss. Therefore we define the degree of node congestion to describe the occupancy of the node data cache.
Decision Tree Based Routing Algorithm

UAANET Decision Tree Formation Based on CART

First, according to the concept of the feature attribute, the data used for training or executing the task is collected and classified as a training set. Then the data set is calculated according to the defined GINI coefficient evaluation standard, the optimal division value of each attribute is calculated, and the attributes are sorted according to the order of GINI coefficients from large to small, and finally formed the decision tree according to CART algorithm in machine learning. After the decision tree is formed through the training set, the importance of the six attributes is determined and different decision-making schemes are formed. When the drone cluster performs the same task in a similar environment, a pre-formed decision-making scheme can be used. The flow chart is as follows:

![Algorithm formation flow chart](image)

Figure 1. Algorithm formation flow chart.

Analysis of Routing Algorithm Rules Based on CART

A. Choice of decision-making plan

Collect tasks and environmental information, and find out whether there is similar situation information in the decision category library. If there is, directly select the corresponding decision plan in the decision tree to enter the routing node selection stage; if not, enter the decision tree formation stage: collect the network system situation information. The state at the first N time is $T_n$ ($n = 0,$
1...N), classifying and calculating the attributes, forming a decision tree, inputting the system state $T_{N+1}$ at the time of N+1 to make a decision test. If it satisfies the delay, etc indicator threshold, program formation.

B. Forward node selection
After the formation of the decision tree, the gini coefficients of the six attributes are $G_i (i = 1, 2... 6)$, the same as entropy. The gini coefficient can also reflect the amount of information provided by the target. The smaller the gini coefficient, the higher the purity of the property. The larger the gini coefficient, the worse the data aggregation, the more weight should be added. According to the gini coefficient, the weight is

$$M = \omega_1 \varphi_{in} + \omega_2 \varphi_{old} + \omega_3 \varphi_{mc} + \omega_4 \varphi_{med} + \omega_5 \varphi_{ned} + \omega_6 \varphi_{ed}$$  \hspace{1cm} (4)

After quantifying and standardizing each attribute, the weight $\omega_i$ is added to obtain the decision factor $M$ of the node. In the data packet forwarding phase, in order to improve the message delivery success rate, When selecting a path, select the neighbor node with the highest decision factor as the next forwarding node.

C. Decision feedback
When the UAV ad hoc network system completes a data packet forwarding, it evaluates the forwarding effect. If the destination node successfully receives the data packet of the source node, it increases the trustworthiness of the decision scheme in the background. Failure to receive a packet from the source node reduces the trustworthiness in the context. The formula for calculating the trustworthiness $P$ is as follows:

$$P = \frac{N_{suc}}{N_{suc} + N_{fail}}$$  \hspace{1cm} (5)

$N_{suc}$ indicates that the destination node successfully receives the number of forwarded packets, and $N_{fail}$ indicates the number of failed destination packets received by the destination node. When the trustworthiness of the decision-making scheme is lower than the specified value, indicating that the decision-making scheme is not applicable to this background, the surrounding situation is re-evaluated, the training set is collected, and a new decision tree and decision-making scheme are formed.

Simulation
In this paper, NS2 is used for simulation. In order to verify the effectiveness of the proposed algorithm, the LEACH algorithm in the current popular clustering algorithm and the GPSR algorithm based on geographic location are simulated in the same environment. The simulation environment is set to be in the range of 1000*1000km$^2$, and the number of drones is 20 steps, from 40~140. The simulation results of the three algorithms are as follows:

![Figure 2. Delivery success rate simulation diagram.](image-url)
It can be seen from the figure that the decision tree-based routing algorithm studied in this paper has obvious advantages in route delivery success rate because it is based on the speed similarity and the impact of both aspects of node load compared to the other two algorithms. Nodes that are too fast can easily lead to link instability and affect the effect of packet transmission. Nodes with large load will often discard new data information when the cache is full, which will also adversely affect the delivery rate. As the number of nodes increases, the delivery success rate increases continuously. When the node density further increases, the delivery success rate decreases slightly due to inter-node interference. The simulation results show that the transmission path selected by the algorithm is more stable and relatively less affected by the nodes, which is beneficial to improve the delivery success rate of data packets.

![Figure 3. Routing overhead simulation diagram.](image)

The figure shows the relationship between routing overhead and nodes. As the number of nodes increases, the network size increases gradually, and the number of control packets used to transmit data packets increases. The cost of nodes in the network increases for maintaining neighbors. Compared with the GPSR algorithm, the LEACH algorithm and algorithm in this paper have more complex structural relationships. When selecting a path, more node information is needed, resulting in a larger routing overhead.

**Conclusion**

The experimental results show that the routing algorithm based on decision tree has great advantages in improving the delivery success rate of data packets. This is because the algorithm in this paper comprehensively considers six attributes in the process of finding the path, and after training in the previous training set, The resulting weights are more suitable for actual use and also have better adaptability. However, this algorithm is not ideal in reducing routing overhead. The next step should be based on how to minimize routing overhead.

**References**

[1] Floriana Esposito, Donato Malerba, Giovanni Semeraro. A further study of pruning methods in decision tree induction[J]. In AI & Statistics, 1995; (250):211~218.

[2] Gupta P, McKewon N. Algorithms for packet classification[J]. IEEE Network, 2001, 15(2):24-32.

[3] Breslow L A, Aha D W. Simplifying Decision Tree: A survey. http://www.aic.ntl.navy.mil/~aha/, 2000.

[4] C S Wallace, J D Patrick. Coding decision trees[J]. Machine Learning, 1993;11(1):7~22.

[5] WU Dapeng, Zhang Puning, Wang Ruyan. Connection status aware cost efficient message transmission mechanism in opportunistic networks[J]. Journal on Communication, 2013, 34(3):44-52.
[6] Liu Huiyuan, Dong Chunyang, Huang Qiong. Multiple copies of VANETs based on decision-tree for opportunistic routing protocol[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2016, 28(6):769-776.

[7] Yin Ru, Men Changqian, Wang Wenjian, Liu Shuze. Model Decision Tree: An Accelerated Algorithm of Decision Tree[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(7):643-652.

[8] Li Zhenqiang, Zhang Shengliang, Ma Yan, Zhao Xiaoyu. Multiple Decision Tree Algorithm for Packet Classification[J]. Journal of Electronics & Information Technology, 2008, 30(4):975-978.