1. The training process of improved random forest based on AFSA

Step 1 Parameters of the AFSA and RF algorithms are initialized. These include the number of iterations, the number of artificial fish, the location of the fish, the visual field of the artificial fish, the maximum step length, the crowding degree factor, the maximum number of behavior attempts, the size of the random forest scale, and the attribute subset size.

Step 2 Assuming that the original training set is (X,Y),

$(x_1,y_1),(x_2,y_2),\cdots(x_N,y_N)\in(X,Y)$, the number of the listed sets is N . Bootstrap sampling method randomly extracts N samples (x_i,y_i) as training samples$\left(X^*,Y^*\right)$.

Step 3 Input training sample$\left(X^*,Y^*\right)$, select k feature subsets when tree splits to train the decision tree.

Step 4 Loop the two steps above until the decision tree forest with preset number $nTree$ is generated, meanwhile, the initial establishment of the random forest classifier is completed.

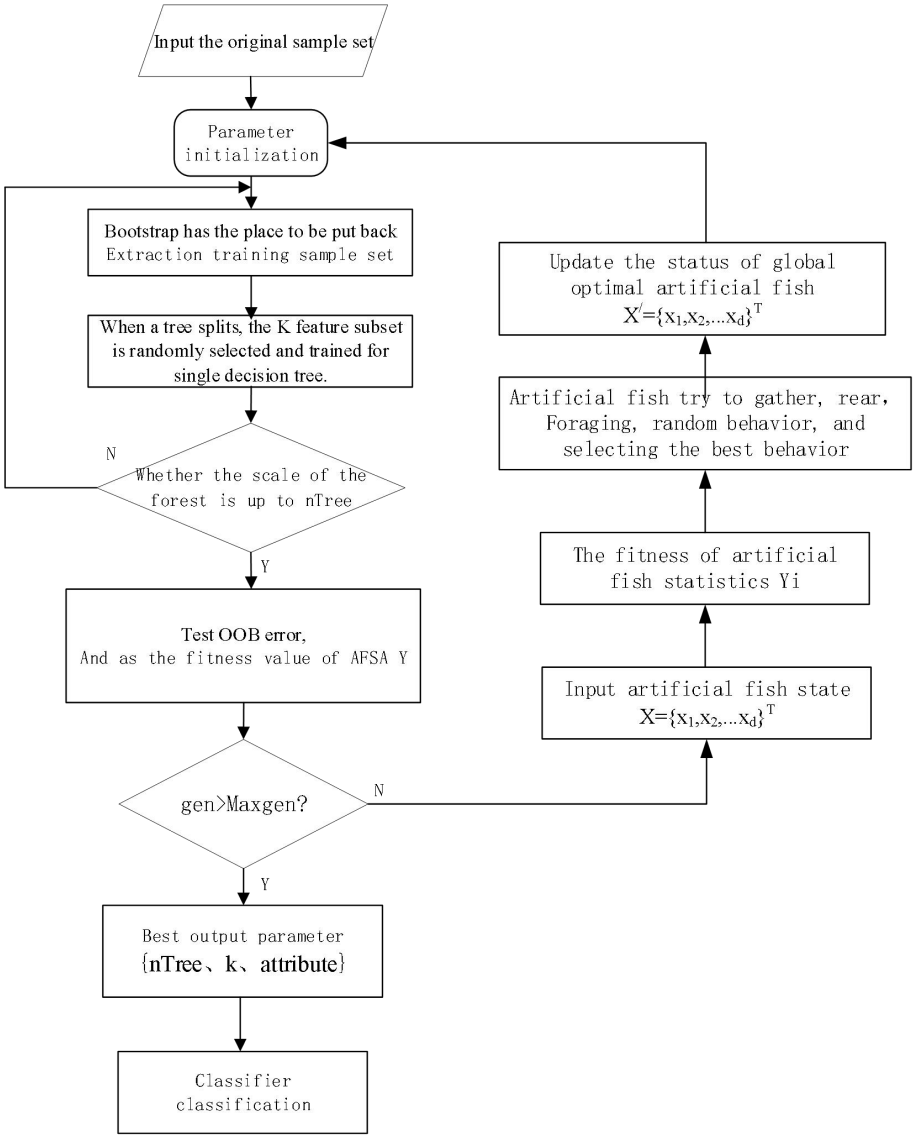
Step 5 The OOB sample is used as a test sample to test the classifier. Then the OOB error is obtained and used as the fitness value Y in the AFSA algorithm.

Step 6 Judge whether $gen > Maxgen$ is achieved, if is, output the RF classifier parameter $nTree$, k , $\left\{Attribute_i \middle| i=1,2,\cdots,M\right\}$ and enter the Step 8; otherwise, we need to set the RF classifier parameter $X=\{nTree,k,Attribute\}$ as the initial state of the artificial fish to forage, cluster, tailgate by the AFSA, and update the global optimal artificial fish state $X^{'}=\{nTree,k,Attribute\}$.

Step 7 Use the optimized parameters $X^{'}=\{nTree,k,Attribute\}$ to reset the RF classifier and enter the Step 2.

Step 8 Finish the training of the classifier and keep OOB error as the correct rate of classification λ 。

2. Flow chart of cervical epithelial cell recognition algorithm based on improved random forest model is as follows.


Input the original sample set
Parameter initialization
Bootstrap has the place to be put back Extraction training sample set
When a tree splits, the K feature subset is randomly selected and trained for single decision tree.
Whether the scale of the forest is up to nTree
N
Y
Test OOB error, And as the fitness value of AFSA Y
gen>Maxgen?
N
Y
Best output parameter {nTree、k、attribute}
Classifier classification
Update the status of global optimal artificial fish X'={x1,x2,...xd}T
Artificial fish try to gather, rear, Foraging, random behavior, and selecting the best behavior
The fitness of artificial fish statistics Yi
Input artificial fish state X={x1,x2,...xd}T