Algorithm 1. Generate new random forest!

Input: training set D; attribute set A

Output: multiple expression classification decision trees.

1: Count=0; number=0;

2: Create the root node node;

3 : If all samples in D belong to the same category C, then,

4: Mark node as class C leaf node, return,

5: end if

6: If A=ϕ

, OR the sample values on A are the same, then.

7 :   Mark node as a leaf node and its category as the class with the largest number of samples, return

8: end if

9 : For each attribute, information gain rate is calculated by Equation (2).

10: Select the optimal partition attribute from A, and assume that the test attribute A \* has

  the highest information gain rate during the experiment.

11: Find the segmentation point of the attribute;

12: A new leaf node is separated from node a\*;

13: If the sample subset corresponding to this leaf node is empty, then this leaf node is

    divided to generate a new leaf node, which is marked as the expression with the highest number.

14: Else,

15   continue to split this leaf node;

16: end if;

17: One decision tree is created.

18: make the test sample into the established tree and calculate the recognition rate,

19: if accuracy<0.6, count=count,

20: else

21:   count=count+1;

22: end if

23: if count <M,

24:   repeat step(2)-step(22)

25: else

26:   count =M,

27: break.

28: end if

29: Set the threshold value δ

30: If random <δ

31:   Select the optimal decision tree from all the currently established decision trees

    as the alternative decision tree. number=number+1;

32: else

33:   The decision tree is randomly selected from all the currently established decision trees

    as an alternative decision tree. number=number+1;

34: if number<m,

35   repeat step(29)-step (33)

36 : if number=m,

37:   break

38: end if

39: All the selected decision trees are combined to form a random forest

40 : The test samples are put into the random forest, and the classification results of each decision tree are collected.

   The results with the most votes will be used as the prediction classification of the current sample.