**ID3 Improvement**

The problem with ID3 decision tree is that partitioning by an attribute with many different values may result in a very low entropy due to overfitting. From my point of view, we should to avoid attributes with large numbers of possible values when creating decision trees. In general, there exists a problem of multi-value bias in the process of attribute selection [1], but the attribution that has more values is not always optimal; In addition, it is not easy to calculate information entropy by using algorithms, and its costs a lot of time; moreover the tree size is very hard to control difficult to control and the tree with a big size requires many long classification rules.

In order to solve the defect of multi-value bias in ID3, C4.5 selected gain ratio instead of the information gain equation[2], but this method have a lot of logarithmic operations, which may affect the whole performance. For my point of view, I want to try to add different weights for the information gain equation of every candidate attribution, and this way not only makes the selection of the optimal attribute more reasonable and with lower computational complexity., but also has a small effect on the running speed. Here is the steps of the algorithm[3]:

Input: the whole decision table

Output: a decision tree

1.Generate node

2.If the training sample belong to the same class. Then

3.The node is labeled as leaf node named LEAF; Return

4 End if

5 If A = f or the values in A are same in class Then

6 The node is labeled as leaf node; Return

7 Else

8 Minimum entropy is chosen as a heuristic strategy to select the optimal partition

attribute ai from A based on the simplified information gain;

9 For every value Vi in data do

10 Generate a branch for node; Dv is the sample subset that has value V

i from data;

11 If Dv is empty. Then

12 The node in a branch is labeled as a leaf node; Return

13 Else

14 For Every branch of data do

15 Determine split property by comparing the size of the coordination degree

and condition certainty degree.

16 If or there are no other condition attributes, define the branch as a leaf node.

17 ELSE define the branch as a non-leaf node, return Step 8.

18 End if

19 End for

20 End if

21 End for

22 End if

Reference

[1] Fayyad, U.M.; Irani, K.B. The Attribute Selection Problem in Decision Tree Generation. In Proceedings of

the National Conference on Artificial Intelligence, San Jose, CA, USA, 12–16 July 1992; pp. 104–110.

[2] Hssina, B., Merbouha, A., Ezzikouri, H. and Erritali, M. (2014). A comparative study of decision tree ID3 and C4.5. International Journal of Advanced Computer Science and Applications, 4(2), 13-19.

[3]]Wang, Y., Li, Y., Song, Y., Rong, X. and Zhang, S., 2017. Improvement of ID3 Algorithm Based on Simplified Information Entropy and Coordination Degree. *Algorithms*, *10*(4), p.124.