

2.7 A recursive method.

In case of complex dependence between characteristics for decision tree construction, one can apply more complex methods than a method of sequential branching.

The general scheme of a method is similar to the scheme which was used in a method of sequential branching with the difference, that instead of operation of division, more complex operation of growth is used. Let us consider the basic steps of this method.

- 1) To carry out an operation of growth for the root node.
- 2) To check the degree of promise of branching for each of the new child nodes.

If a node becomes a leaf, then we give to it the appropriate decision.

- 3) To carry out an operation of growth for each child node;

After that we repeat steps 2 and 3, until there will be no more perspective nodes for branching, or maximal possible complexity of a tree will not be achieved.

Let us describe the operation of growth. In this operation, the next criterion of quality of the tree is used:

$$Q = N_{err}/N + \alpha M/N \quad (\text{for PR problem}) \text{ or}$$

$$Q = d/d_0 + \alpha M/N \quad (\text{for RA problem}),$$

where α is some given parameter.

The operation consists of the following steps.

- 1) To fix the next characteristic X_j ($j=1, \dots, n$).
- 2) Initial branching (initial tree construction). One can understand initial branching as the dividing of the node on maximal possible (determined by sample) number of new nodes (figure 14). In case of the quantitative characteristic, this number is equal to the number of subintervals of splitting (i.e. the new node is matched to each subinterval), in case of the qualitative or ordered characteristic one value of the characteristic is matched to each node.

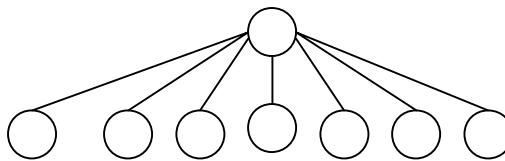


Fig. 14

- 3) To check up degree of promise of branching for all child nodes of the initial tree.
- 4) To carry out an operation of growth for perspective nodes of initial tree recursively (figure 15). The maximal recursion depth is limited to the given threshold R .

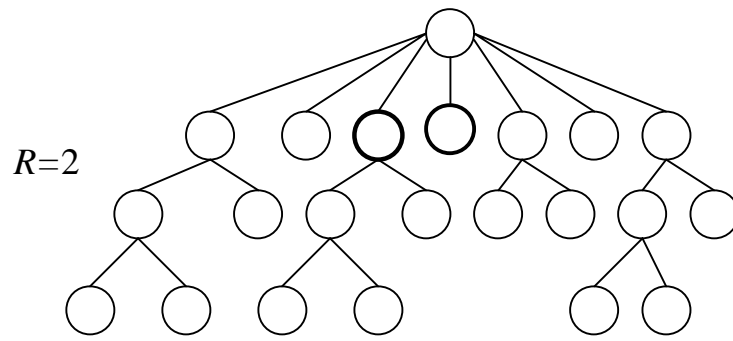


Fig. 15

- 5) To calculate the criterion of quality of the received tree.
- 6) To sort out all pairs of nodes of the initial tree allowable for aggregation.
- 7) To carry out for each pair the recursively operation of growth of the incorporated node and to calculate the quality of the received variant of a tree.
- 8) To remember the variant of a tree with the best criterion of quality (figure 16).

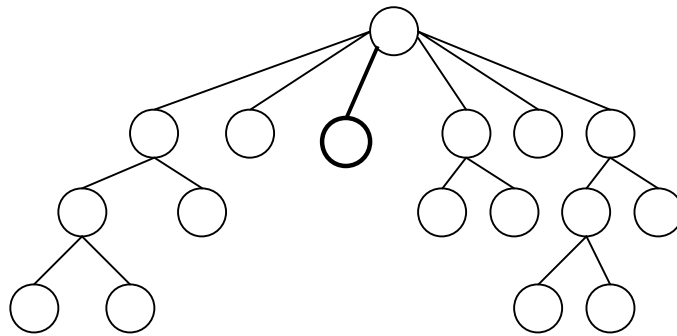


Figure 16

- 9) To repeat steps 6-8 for new pairs of incorporated nodes of initial tree. Every time to save the best variant. To repeat steps until all nodes will be united in one.
- 10) To pass to the next characteristic and to repeat steps 1-9 until all characteristics will be considered.

The most time-consuming in this algorithm is the fourth step. At each recursive reference to the operation of growth, steps 1-11 will be carried out. In other words, the optimum sub tree is forming for the appropriate subset of observations and this operation may call itself some times. By increasing the parameter R , it is possible to increase depth of a recursive nesting which allows finding more complex dependencies between characteristics (but in this case, time and a required memory is increased).

One more feature of this method is that the number of the branches leaving each node beforehand is not fixed and their optimum number is searched for.

The parameter α can be chosen equal to 1. By increasing the given parameter, the tree will become simpler (i.e. contain smaller number of leaves) and by reducing the tree will become more complex.