

A NEW WAY FOR FINDING THE OPTIMAL DECISION TREE USING A MODIFIED ITERATION OF ID3 ALGORITHM

Daniela Qendraj (Halidini)¹, Etleva Halidini²

¹Department of Mathematics, Faculty of Information Technology, University
“Aleksandër Moisiu” Durrës, Durrës, Albania, Email: daniela_qendraj@hotmail.com

²Highschool “Besnik Sykja”, Tirana, Albania, Email: etlevahalidini@hotmail.com

Abstract

The aim of this study is to find the optimal binary decision tree using the distribution obtained by transforming our variables into binary variables. We know that the decision tree represents a distribution and every path between attributes is a conditional probability. Choosing binary tree makes simply finding the optimal way because the number of leaves and the number of nodes are the same. We will prove the existence of a sequence of steps for finding the optimal binary decision tree from a given tree that represents a given distribution, by a modified iteration of ID3 algorithm. Many works are related to decision trees, including their growing, selection, splitting, complexity. There are many algorithms in decision trees that use the top-down induction selecting best attribute for every split, such as ID3, CART, C4.5, C5 . Appart them we want to use decision trees by implementing to edge and vertices numeric characteristics, using the probability theory. The first node represent the tree stump, which is the starting point of the logical structure. From the stump node we obtain edges that represent the alternatives of the first level. The vertices of these edges show the alternatives of the first level. From each of them we obtain edge for the next level, and so on. This structure is formed by obtaining from each vertice the same number of edges that have the next alternatives of the level. If we have to construct a binary tree for all decisions of the structure, the way of finding the optimal path in the tree is simply, because from each vertice we obtain two alternatives edge or none alternative. As we see the new version of ID3 has only 4 leaves compared to that with 7 leaves, with identical prediction on the test of data set. This implies a new way for constructing decision trees that are equivalent, and finding an optimal decision tree is easily if we use the conditional probability in binary trees.

Keywords: *decision tree, P-distribution, ID3 algorithm*