CS771 Assignment-2

Banoth Anand
220281
banotha22@iitk.ac.in

Bali Yaswanth Naidu
220279
baliyn22@iitk.ac.in

Nidhi Bajpai
230694
nidhib23@iitk.ac.in

Pankhuri Sachan
230734
pankhurisa23@iitk.ac.in

1 Solution

In this assignment, we are using the concept of Entropy Minimization and lookahead strategy to split the parent node into the corresponding children nodes. The Entropy Minimisation is basically the ID3 algorithm where we try to minimize the entropy of the children node compared to the parent node to maximize the Information Gain. Each split is done using Entropy Minimisation, where we had just split the words just to minimize the entropy of the children nodes.

When the node size is small instead of choosing the split that gives the maximum entropy reduction right now, a better strategy is to choose a split that will result in maximum entropy reduction two levels or three levels from now.

Entropy of a node,

$$H = -\sum_i p(i) * \log(p(i)) \tag{1}$$

where $p(i)$ is the probability of a class in the node.

Information Gain,

$$IG = H(Children) - H(Parent) \tag{2}$$

The root node is basically the complete collection of all the 5 biagrams obtained from each word. It splits the node into child nodes based on whether each tuple contains the query. The secret word is selected from the dictionary, then a query is chosen from the current children node and checked with the secret word. So the next split depends on the response i.e, if the word matches the answer. This goes on till the query word matches precisely with the secret word or the limit of the queries exceeds. Limit of queries is 5 in this case.

Training Time(s)	Dict Size	Accuracy	Avg Queries
0.0190417765994275	207392	0.8984104251338763	3.3359817689999547

References : ChatGPT

Preprint. Under review.