Decision Table Construction and Use [ArtH4992 / Sp 17]

Our approach to building a decision tree procedure to identify the sub-variants of the watermarks in Rembrandt's prints according to the taxonomy of Erik Hinterding is to select features that can be used in a sequence of yes-no questions leading us to the proper sub-variant. We have followed a "halving" strategy for selecting a sequence of questions that divides the remaining group of interest approximately in half in order to try to minimize the average number of questions that needs to be asked to end in the correct sub-variant. The multiple branches in the decision tree means that a different sequence of questions will be asked in distinguishing the different sub-variants.

In an attempt to convert the decision-tree construction procedure into a task susceptible to a mathematical algorithm that minimizes the average number of questions needed to isolate the sub-variants, we will use a decision table approach, in which all of the candidate questions about the presence or absence of certain observable features are answered for all sub-variants. For this list of all of the questions about watermark features to be used for a particular type, consider a table where the columns represent the questions and the rows the sub-variants. The entries in the table are the corresponding Yes/No answers. For example:

Questions	1	2	3	4	5
Subvariant A	Yes	No	No	Yes	No
Subvariant B	No	Yes	No	No	Yes
Subvariant C	No	Yes	No	Yes	No
Subvariant D	No	No	No	Yes	No
Subvariant E	No	Yes	Yes	Yes	No

Now each subvariant can be identified by the sequence of yes/no answers to the full set of questions. For example, subvariant A is defined by YNNYN, subvariant B by NYNNY, subvariant C by NYNYN, subvariant D by NNNYN, and subvariant E by NYYYN.

In order for the set of questions to lead to unique identification of each subvariant, the yes/no sequence associated with each subvariant must be different from all of the other identifying yes/no sequences. Our example table satifies this criterion.

To be assured that the sub-varaint is corectly identified, all questions need to be answerable with complete confidence. For proper progression through a decision tree, the absence or presence of a certain feature need only be unambiguous for the subset of possible sub-variants remaining at the current decision tree node. For example, assume that our journey through the decision tree has progressed to the point that we are left trying to distinguish between just two choices. A distinguishing feature for the remaining two choices could be that in a pair of elements within one choice one element is quite larger than the other and the reverse in the other choice. However, in the full class of possible sub-variants some watermarks may have the pair of elements of interest of approximately the same size. This question now becomes a problem in filling in all entries in the decision table. Thus, one challenge in building a decision table is to select features that are readily discernible in all possible sub-variants.

Task 1: Construct a decision table for your branch. If certain questions are difficult to answer for all sub-varaints within your type, replace those questions with others that are more confidently answered for all sub-variants but leave you with a complete decision tree distirnguishing all sub-variants.

It is possible to assemble a set of suitable questions without first building a decision tree by noticing features and posing questions one at a time and construction the decision table one column at a time. The addition of columns to the decision table continues until all of its rows are unique. To determine the sub-variant label of a watermark of interest from the decision table requires answering all of the questions for the watermark of interest. But, your original decision tree probably does not require answering this many questions in order to identify each particular subvariant.

To extract a decision tree from a decision table using the halving strategy, start with a question that has a column of Yes/No answers that divides the group approximately in half. For example, for the 5 questions in our sample table, question 2 divides the group of subvariants roughly in half with B, C, and E extracted for a Yes and A and D for a No. All other columns have only one Yes or one No. So, we can start our decision tree with Question 2. To separate A from D we can use question 1. To separate B and C from E, we can use question 3. To separate B from C, we can use question 5.

Thus, the sequence of questions resulting in subvariant A is Question 2 (answered No) followed by Question 1 (answered Yes). This decision tree path can be labeled with the sequence of questions 2-**1**, where the bold font indicates a Yes. Thus, the decision tree can be described by the sequence of questions producing each sub-variant

- A :: 2-**1**
- B :: **2**-3-**5**
- C :: **2**-3-5
- D :: 2-1
- E :: **2**-**3**

Thus, the average number of questions to be asked to reach a sub-varaint is (2+3+3+2+2)/5=2.4.

Note that we did not use Question 4. However, we clould use it instead of Question 5 to distinguish B from C. This would change the decision tree paths for B and C to **2**-3-4 and **2**-3-**4**, respectively.

This does not change the average number of questions that needs to be asked.

<u>Task 2</u>: From the decision table in Task 1, use the halving strategy to build a decision tree and deterrmine the average number of questions needed to reach a sub-variant.

Different questions may lead to different values for the average number of questions to be asked to reach an answer even when using the halving strategy with each subsequent question. For example, consider what happens if Question 2 was Yes for D but No for all other subvariants. The new decision table would be

Questions	1	2	3	4	5
Subvariant A	Yes	No	No	Yes	No
Subvariant B	No	No	No	No	Yes
Subvariant C	No	No	No	Yes	No
Subvariant D	No	Yes	No	Yes	No
Subvariant E	No	No	Yes	Yes	No

Now all questions divide the subvariants into one group of 4 with the same answer and the one remaining subvaraint with the opposite answer. So, it doen't matter which question we begin with in terms of attempting to divide the original group in half. One decision tree for this new table is

- A :: **1**
- B :: 1-2-3-4
- C :: 1-2-3-**4**
- D :: 1-**2**
- E :: 1-2-**3**

The average number of questions that needs to be asked to reach a sub-variant is now (1+4+4+2+3)/5=2.8, which is larger than that achievable with the previous decision table.

Task 3: Attempt to find (a) a different set of questions than you used in Tasks 1 and 2 that reduces the average number of questions needed when using the halving strategy and (b) a different set of questions than you used in Tasks 1 and 2 that increases the average number of questions needed when using the halving strategy.