No 705

TALLINNA TEHNIKAÜLIKOOLI TOIMETISED
TRANSACTIONS OF TALLINN TECHNICAL UNIVERSITY

UDC 681.3

R. Kuusik

APPLICATION OF THEORY OF MONOTONIC SYSTEMS FOR DECISION TREES GENERATION

Abstract

In this paper a new way for generation of the decision trees immediately from initial data matrix X(N,M) is proposed. The use of the features extracted in conjunction with the theory of monotonic systems is described. The corresponding algorithms are estimated.

1. Formulation of the problem

Formation and use of the so-called decision trees (DT) deserves deep attention. The problem involves the following.

Let us assume that X(H,M) is a final data matrix, where H is the number of objects and M is the number of quantities Fj in X, in which each element Xij can have value from interval Cj = 1,...,Ej.

The decision tree <T,Fx> is defined as a rooted tree T with a root Fx so that all subtrees Ti of the tree T have exactly the same successors as they had in the tree <T,Fx>.

The best decision tree is called a tree <T,Fx>, to which the greatest value corresponds according to the previously determined criterion of "goodness".

2. The present situation

To solve the above-mentioned problem the algorithms for formation of a decision tree, the criterion of "goodness" and an algorithm of extraction the best DT are proposed in [1].

It is supposed that from data matrix X(H,M) by the GUHA method [2] elementary conjunctions Ki, l = 1,...,L,

were previously generated which are the nodes of a decision tree, and the corresponding output results Dg, g=1,...,G, i.e. elementary implications Kl --> Dg are determined. For each Dg its own best decision tree is generated. Value V= =A/B, where A is the number of objects in X(H,M), for which Kl--> Dg is true on X, B is the number of objects in X, containing Kl, serves as a criterion of "goodness" of a decision tree.

It is assumed that the best decision tree is the one to which the greatest value of criterion V corresponds.

General algorithm for generation of the best DT, according to [1], is the following:

1) form a tree for Fx according to Dt from {Kl};

2) determine the existence of one whole branch from a root (Cj)Fx to a leaf (i.e. that it contains all quantities j=1,...,M in the branch as nodes) on $\langle T,Fx \rangle$;

3) calculate the measure of a DT goodness. If Vx > > Vmax, then Vmax:= Vx and elementary conjunctions, corresponding to nodes $j \in \langle T,Fx \rangle$, j=1,...,R, $R \leq M$, being excluded from {Kl};

4) if there are no more trees, then the best DT is found, or else determine the next root quantity and pass to 1.

For Dt formation a table is constructed which measurements

$$P = Ex * M1 * \sum_{j=1}^{M1} Ej,$$ where M1 - the number of input quantities, $M \leq M1$.

If, for example, M1=10 and Ex=Ej=3, then P=900 cells.

E x a m p l e 1

Let {Kl} be a set of elementary conjunctions, generated by the GUHA method from X, l=1,...,11. Let F1, F2, F3 be input quantities and F4 an output quantity. For each Kl it is determined, what result Dg=F4 it comes to (see example in [1]).

Let us introduce elementary implications Kl--> Dg by table X(11,4):

Kl \ Pj	F1	F2	F3	F4
1	1			1
2	3			2
3		2		1
4	2	1		1
5	2	2		2
6	2	3		2
7	2		2	2
8		2	1	1
9	2	1	1	2
10	2	1	2	1
11	3	2	1	2

For example, to the eleventh row of X corresponds the elementary implication (3)F1 & (2)F2 & (1)F1 --> (2)F4.

If you neglect the resulting quantity F4, then at the existence of elementary conjunction K12= (2)F1 from the root F1 by the data of the table X, a DT is generated

F1 --> (1)F1
--> (2)F1 --> (2)F2
--> (1)F2 --> (1)F3
--> (2)F3
--> (3)F2
--> (3)F1

The node (3)F1 is a list because an elementary conjunction in {Kl} equal to (3)F1 & (2)F2 does not exist.

3. A new solution

Solution of the problem suggested in[1] has some faults.

First, the set of elementary conjunction {Kl}, generated from X by the GUHA method and the corresponding elementary implications are the initial for DT generation algorithm, not the initial data matrix X(H,M).

The application of the GUAH method is a polynominal process. GUHA does not guarantee coverage of elements of the

initial data matrix X with generated by it elementary implications K1⟶Dg because of existence of

a) subjective restrictions. The researcher determines previously the maximum length of the elementary conjunctions. It is assumed that the short conjunctions are "more interesting" than long ones,

b) algorithmical restrictions. They do not have quick algorithms for determination of all elementary conjunctions in X.

Secondly, the algorithm suggested for generation of the best DT is quite work-consuming:

1) the usage of the GUHA method,

2) constructing a table for each Fx,

3) determination of existence of the whole branch from a root Fj to a leaf (do there exist all quantities j=1,... ...,M as nodes). If it is so, we have to determine whether this branch is the best one.

Further we suggest a new approach for realization of the described problem using the theory of monotonic systems. The algorithms suggested can be used for generation of all DT on the initial data matrix X(H,M), or the best DT on X, or the corresponding DT also on the matrix of elementary implications, generated by the GUHA method or by its analogues (see example 1).

3.1. Description of the algorithm

A decision tree (DT) is defined in [1] as follows:

1) in one tree with a root Fx all values Fx: 1,...,Ex are the nodes of the root;

2) the sequence of quantities in DT for all branches with a root Fx is the same.

Using the criterion given in [1] , the best Dt is the one which has the greatest number of leaves.

In [3] the combinatorial algorithm for generating the best DT by the definition above directly from the initial data matrix is suggested. In [3] it is shown that the best DT is the tree which has the greatest coverage of elements of the initial data matrix, i.e. the best DT is the tree with the greatest number of leaves.

Below we will introduce a new solution of the problem, specify a DT, describe the algorithms and criteria of "goodness" for generating the best DT.

In [1] it is suggested that from the examined data matrix by the GUHA method Dtree nodes are previously generated. From these nodes various DT are formed and on the basis of the criterion, suggested by the authors, the best one is defined. From this point of view the suggested criterion satisfies the goals.

In reality it is more complicated. For example, during the expert interrogations a certain part of the experts may give similar answers. It means that the choice of criteria of the best DT may be put some other way too, taking as a principle also frequency of suggestions of certain answers or its combinations. For example, if we have two decision branches

1) K9= (2)F1 & (2)F2 & (1)F3 given by different experts 10 times and 2) K10= (2)F1 & (2)F2 & (2)F3 accordingly 6 times, then at the equal criterion value by [1] they can be equal, but K9 is preferable by the expert evaluation.

Concluding the above described discussion, we will suggest a new criterion and a new algorithm generation of the best DT direct from the initial data matrix X(H,M). According to this, to generate from X one node of the tree less than 2HM operations are needed.

Further, under the best DT we shall assume a whole branch of DT (from root to leaf), to which the greatest criterion value T (will be determined further) corresponds.

The algorithm for DT formation suggested below is based upon the theory of monotonic systems [4,5]. To create a monotonic system on X the so-called frequency conversion described in [6] is used.

Let X(H,M) be a final matrix of nominal data, where each element Xij may have a value from the interval Cj=1,2,. ...,Ej.

In algorithm A we use the following denotations:

X^l - the subset of X, $X^l \subset X$, $X^{l+1} \subset X^l$, $X^0 = X(N,M)$, max $l \leq M$;

A^l - the table of frequencies of the subset X^l, $A^{l+1} \subset A^l$.

ALGORITHM A

A1. Find frequencies for every quantity $j=1,\ldots,M$ values $Cj=1,\ldots Ej$ from X^0 (they form table A^0) and find the greatest of them: MAX:=max|Cj|. F:=j; Y:=Cj; l:=0; t:=0; Kt:=(Y)F.

A2. Find from X objects which contain a Kt. They form data matrix X^{l+1}.

A3. Find frequencies for every quantity $j=1,\ldots,M$ values $Cj=1,\ldots,Ej$ from X^{l+1} (we form table A^{l+1}). t:=t+1. The values of quantities, whose frequencies are equal to the value of MAX, form an elementary conjunction $Kt=\&_u(Yu)Fu$ with a frequency MAX.

A4. l:=l+1. If all quantities are described in Kt, then DO $A^{l-1}:=A^{l-1}-A^l$, l:=l-1. If l=0 then go to A5. END. Go to A2.

A5. The end of the algorithm

The frequency of the elementary conjunction is a monotonic function of weight. Elementary conjunctions Kt, separated by the algorithm A, are kernels according to the theory of monotonic systems [4]. The corresponding theorems are proved in [3].

Let us represent the elements of the set {K1} (see example 1) in the form of a data matrix and assuming that zero corresponds to "is not determined", we obtain the data matrix X(11,4):

i \ Fj	1	2	3	4
1	1	0	0	1
2	3	0	0	2
3	0	2	0	1
4	2	1	0	1
5	2	2	0	2
6	2	3	0	2
7	2	0	2	2
8	0	2	1	1
9	2	1	1	2
10	2	1	2	1
11	3	2	1	2

Table of frequencies for quantities in X is as follows:

Fj / Cj	F1	F2	F3	F4
0	2	3	6	0
1	1	3	3	6
2	6	4	2	5
3	2	1	0	0

E x a m p l e 2

Using X as the initial for the algorithm A and assuming that

a) zero value is not considered in the analysis,

b) the quantity F4 is not considered in formation of DT,

c) sequence of quantities in the DT is not defined previously, then in answer to the algorithm A usage, we acquire the following DT:

```
          6          3          1
1) --> (2)F1 --> (1)F2 --> (1)F3    (=10)
                           1
                      --> (2)F3    (=10)
                     2     1
                   --> (2)F3 --> (1)F2    (=9)
                     1
                   --> (1)F3 &   (1)F2    (=8)
                     1
                   --> (2)F2             (=7)
                     1
                   --> (3)F2             (=7)
          4          2          1
2) --> (2)F2 --> (1)F3 --> (3)F1    (=7)
                    1
                  --> (2)F1         (=5)
                    1
                  --> (3)F1 &   (1)F3    (=6)
```

```
      3                    1
3) --> (1)F2  &  (2)F1 --> (1)F3        (=7)
                           1
                       --> (2)F3        (=7)
      3          2       1
4) --> (1)F3 --> (2)F2 --> (3)F1        (=6)
                 1
              --> (1)F2  &  (2)F1       (=5)
      2                    1
5) --> (2)F3  &  (2)F1 --> (1)F2        (=5)
      2          1
6) --> (3)F1 --> (2)F2  &  (1)F3        (=4)
      1
7) --> (3)F2  &  (2)F1                  (=2)
   1
8) --> (1)F1                            (=1)
```

In the represented trees the numbers above the pointers indicate frequency of a conjunction corresponding to the node j (it is an elementary conjunction which contains all antecedent to the node j elements ()Fx of the tree). For example, in the DT No. 1 |(2)F1| =6, |(2)F1&(1)F2|=3, etc.

We suggested the measure $T = \sum_{j=1}^{M} Sj$, where Sj is weight of the node j, as a criterion of the DT branch "goodness".

The combination of values of quantities (see example 2, the fourth branch of DT No. 1) may serve as a node. According to [1] it means that there is no branch from the root node ()Fx which contains all quantities as the nodes of DT. On the basis of our suggestions and the criterion of the DT "goodness", quantities, which belong to the node j, are not separated in relations to the node j-1, i.e. we do not prefer any of them.

If in the node j there are several elements (as a combination of values several quantities), the weight of the node is calculated as $Sj = Qj \cdot Zj$, where Qj is the number of elements ()Fx in the node j and Zj is an elementary conjunction frequency in X corresponding to the node j.

The corresponding values of the criterion T are given in the brackets through the label "=" at the end of each branch of the DT. The measurement T of each node of a DT is calculated immediately in generation of DT.

Let us assume that the initial value of the measurement T=0 and after generation of the first branch Tmax:=T1. Then for each following branch b we may immediately say whether it is better or not: if Tb > Tmax then Tmax=Tb.

Forming a DT with the help of the algorithm A, formation of a new branch does not begin from the root Fx, but from the antecedent node b-1. If there are no more paths from this branch, then it is from b-2, etc. For each node the corresponding measurements equal to a sum of weights of the previous nodes are recorded. If in the formation of a new branch from a node b a value U (so-called forecast) $U=T_{b-1} + (M-b) \cdot Sb < Tmax$ (where T_{b-1} is the measurement of the previous node to b, M-b is the number of this branch quantities not yet described), then we may finish formation of all branches from this node, because the measure T will always be less than Tmax. This is determined by monotonous of the weight function.

If for the succedent root of DT its forecast U < Tmax, then we finish work of the algorithm A, because we cannot find a new branch on X whose measure T > Tmax.

Therefore in our example we have to form only three completed branches:

```
  6          3          1
--> (2)F1 --> (1)F2 --> (1)F3      (=10)
                      1
                   --> (2)F3      (=10)
           2          1
           --> (2)F3  --> (1)F2   (= 9)
```

U=6+2·2=10 <=Tmax=10

For

```
  6          1
--> (2)F1  --> (1)F2 &  (1)F3 --> ... and we go to
U=3·6=18>Tmax  U=6+2·1 < Tmax      form a new DT
                                   from root (2)F2.
```

For the next branch

4 2

2) --> (2)F2 --> (1)F3... we finish the formation of branches from the node (2)F2 and go to a new root.

U=3*4 > Tmax U=4+2*2 < Tmax

3

3) --> (1)F2 & (2)F1 --> ...

U=3*3=9 < Tmax

Since already for the root of the tree 3) the forecast U < Tmax, the best DT is found and there is no reason to proceed.

Now let us concentrate our attention upon a peculiarity of measure T suggested by us. You may have noticed that its value can change if the sequence of quantities in the branch changes. It means that the algorithm A determines also the sequence of steps F1,...,Fm in making decisions, i.e. what is the best sequence of leading to the required result. The best sequence is the one to which the greatest value of a function T corresponds.

Taking into account also the resulting quantity F4 in choosing the best DT, the shape of the generated DT strongly changes, because depending on the value F4, we obtain two best DT, one for the data matrix X1 and the other for X2:

1) X1(5,3) and 2) X2(6,3):

	1	2	3	4
1:	1	0	0	1
2:	0	2	0	1
3:	2	1	0	1
4:	0	2	1	1
5:	2	1	2	1

	1	2	3	4
1:	3	0	0	2
2:	2	2	0	2
3:	2	3	0	2
4:	2	0	2	2
5:	2	1	1	2
6:	3	2	1	2

4. Conclusion

In conclusion, we will present the final algorithm of extracting the best DT.

ALGORITHM B

B1. T=0, j=1, Y=1 (contents of Y is No. of a branch =j).

B2. By the algorithm A we generate one node j of Dt. Let us compute the forecast U=Tj-1 + (M-j)*Sj.

If U < Tmax, then if j=1 (the node is a root), then go to B3, otherwise j=j-1 and go to B2.

If the node j is a leaf, then Tmax=Tj, Y=j, j=j+1 and go to B2.

B3. The end of algorithm.

This algorithm ensures a fairly rapid process of generation of the best DT from the initial data matrix X(H,M), because there is no need for preliminary formation and tracing of all DT. By the suggested criterion T we can preliminary to value "goodness" of DT, i.e. if it can serve as the best DT, or not. If the value of the forecast for a root node is not greater than Tmax, then according to the suggested algorithm, the best DT is found.

If we take into account that the search for the combination, corresponding to the node of DT, differs from the combinatorial algorithms and needs no more than 2HM operations, then the generation process of the best tree is very fast.

The algorithm suggested enables us to generate also K of the best DT, or even the worst one. Adding to the algorithm B the requirement that in a generated tree the sequence of quantities related to a root should be equal for all branches of the DT, then it is possible to generate also DT according to definition [1].

The algorithms described above are realized on PC IBM/AT.

References

1. Renc Z., Setikova L. Decision trees: a contribution to automatic interpretation of GUHA results // Int. J. Man-Machine Studies. 1985. No 22. P. 193-207.

2. Hajek P., Havranek T. Mechanizing hypothesis formation. Berlin-Hagelberg: Springer-Verlag, 1978.

3. Kuusik R. Generator hypotheses for qualitative data // Trans. of Tallinn Tech. Univ. 1987. No. 645. P. 141-148.

4. Mullat I. Extremal monotonic systems // Automation and Remote Control. 1976. No 5. P. 130-139; No 8. P. 169--178.

5. Vyhandu L., Müllat I. Monotonic system in scene analysis. Symposium "Mathemathical processing methods for data analysis and processing of cartographical data". Tallinn, 1979. P. 63-66.

6. Vyhandu L. Fast methods for data analysis and processing// Trans. of Tallinn Tech. Univ. 1986. No 614. P. 15-23.

R. Kuusik

Monotoonsete süsteemide teooria rakendusest otsusepuude formeerimisel

Kokkuvõte

Käesolevas artiklis esitatakse uus lähenemine nn. otsusepuude formeerimiseks, neist parima leidmiseks vahetult nominaalsete algandmete maatriksist X(N,M) ilma kõiki otsusepuid läbimata. Kirjeldatud lähenemine põhineb monotoonsete süsteemide teoorial. Esitatakse vastavad algoritmid.