

A PROBLEM OF APPLYING THE DECISION TREE TO HYPOTID DHYROISEASE EXAMINATIONS

Shing-Hwa Lu^a Ding-An Chiang^b Nan-Ching Huang^c

a Associate Professor, Department of Urology, School of Medicine, National Yang-Ming University; Deputy Superintendent, Zhong Xiao Branch, and Chief, Department of Urology, Taipei City Hospita
E-mail: shlu7@yahoo.com.tw

b Professor, Department of CSIE, Tamkang University E-mail: 694190124@s94.tku.edu.tw

c Instructor, National Taipei College of Nursing E-mail: glassstore@www2.tku.edu.tw

Abstract

When a decision tree is applied to a medical examination, it is important that a medical domain expert should review the decision tree carefully because the tree may have irrelevant conditions, which demands extra information to be supplied. For medical examinations, extra information needed means extra examinations needed to a patient, and extra examinations cause more expense and more burdens to the patient and society. Therefore, to save medical resources and reduce patients' burden, we have to deal with irrelevant conditions in the decision tree. In this paper, we use hypothyroid disease examinations as an example to demonstrate this point of the view.

Keywords: Decision tree, the irrelevant values problem, hypothyroid disease, medical resources, medical examination.

1 INTRODUCTION

The decision tree is one of the key data mining techniques and has been applied to medical applications [1,2,3]. However, since the decision tree creates a branch for each value of that appearing in the training data without considering whether the value is relevant to the classification. Consequently, the resultant rules of the tree may have irrelevant conditions [4, 12]. For medical examinations, it causes more expense and burdens to the patient and society. Moreover, saving medical resources

is a very important issue for the Taiwan's government. Therefore, when the decision tree is applied to medical applications, we have to deal with irrelevant conditions in the tree.

Until now, a number of different approaches have been proposed to solve this problem. For example, Fayyad has proposed two algorithms. GID3 and GID3*, to solve this problem of the decision tree constructed using ID3 [4, 5,6, 7]. These two algorithms over-whelm the irrelevant values problem at the step before attribute selection. However, the problem of these algorithms is that some branches in the GID3 tree or the GID3* tree may be longer than that in the ID3 tree. Quinlan uses a pessimistic estimate of the accuracy of rules to generalize each rule by removing not only irrelevant conditions but also conditions that do not seem helpful for discriminating the nominated class from other classes [8,9,12].

We have provided another view to solve the irrelevant values problem of the decision tree without the problem of GID3 and GID3* tree. We eliminate irrelevant values in the process of converting the decision tree to the rules for classification according to information on the decision tree [17]. The original algorithm can only deal with discrete values; however, for medical examinations, some attributes always contain continuous values. Therefore, the original algorithm is revised to solve the irrelevant values problem of the decision tree with

continuous values in this paper. Consequently, the modified algorithm can deal with not only discrete values but also continuous values.

2 BASIC ALGORITHM

Let $\mathbf{A} = \{A_1, \dots, A_n\}$ be a set of attributes and $\mathbf{C} = \{C_1, \dots, C_s\}$ be a set of classes. To represent a decision tree by a set of branches, the branch Br of the decision tree can be represented as the form $Br[A_1] \wedge \dots \wedge Br[A_n] \rightarrow C_k$, where $Br[A_i]$ is a set of branch values out of an attribute A_i in the branch Br , $i = 1 \dots n$ and $1 \leq k \leq s$. For easy explanation the irrelevant values problem for medical examinations, we consider only that attribute values are discrete values in this section. That is, $Br[A_i]$ is reduced to a branch value out of an attribute A_i in the branch Br .

According to the semantics of irrelevant values, a value, when $Br[A_j]$ is an irrelevance of a rule; this value can be deleted or replaced by any value from the same domain value without affecting the correctness of the rule. Therefore, we can trivially do combinatorial explosion in the number of comparisons to all the branches to identify irrelevant values of a branch. However, the process of identifying irrelevant values by this way is very time-consuming. Therefore, to enable users to focus on only relevant conditions of the rules, we provided the following theorems to solve the irrelevant values problem for a complex decision tree whose attributes contain discrete values only. These theorems are proven in [17], readers are suggested to refer to this paper, if further explanation is needed.

Theorem 1. Let $Br[A_1] \wedge \dots \wedge Br[A_j] \wedge Br[A_{j'}] \wedge \dots \wedge Br[A_{n1}] \rightarrow C_{k1}$ and $Br'[A_1] \wedge \dots \wedge Br'[A_j] \wedge \dots \wedge Br'[A_{n2}] \rightarrow C_{k2}$ be two branches through a non-leaf node P in the tree, where the branching attribute with respect to P is A_j . Let $\mathbf{A} = \{A_{j'}, \dots, A_{n1}\}$ and $\mathbf{A}_1$ be the same attributes in these two branches, where $\mathbf{A}_1 \subseteq \mathbf{A}$. Then, Br is in conflict with Br' with respect to $\mathbf{A}$ if and only if $Br[\mathbf{A}_1] = Br'[\mathbf{A}_1]$ and $C_{k1} \neq C_{k2}$.

Theorem 2. Let $Br[A_1] \wedge \dots \wedge Br[A_{j-1}] \wedge Br[A_j] \wedge Br[A_{j+1}] \wedge \dots \wedge Br[A_n] \rightarrow C_k$ be a branch through a non-leaf node P in a decision tree, and the branching attribute with respect to P be A_j . For all branches through P of the decision tree, if Br is not conflict with these branches with respect to attributes $A_{j+1} \dots A_n$, then $Br[A_j]$ is an irrelevant value in Br .

Theorem 3. Let Br be a branch through a non-leaf node P of the decision tree. When the branch value $Br[P]$ has been identified by theorem 1, all other branches through P are useless for the following process to identify the irrelevant values of Br .

To make the process of removing all irrelevant values of a branch more efficient, we can recursively apply the above theorems until the node P is the root of the decision tree.

3 AN ALGORITHM TO IDENTIFY THE IRRELEVANT VALUES TO CONTINUOUS VALUES

For medical examinations and many other applications, some attributes always contain continuous values. To deal with continuous values, the theorem 1 is revised into theorem 4. Consequently, the modified algorithm can deal with not only discrete values but also continuous values.

Theorem 4. Let $Br[A_1] \wedge \dots \wedge Br[A_j] \wedge Br[A_{j'}] \wedge \dots \wedge Br[A_{n1}] \rightarrow C_{k1}$ and $Br'[A_1] \wedge \dots \wedge Br'[A_j] \wedge \dots \wedge Br'[A_{n2}] \rightarrow C_{k2}$ be two branches through a non-leaf node P in the tree, where the branching attribute with respect to P is A_j . Let $\mathbf{A} = \{A_{j'}, \dots, A_{n1}\}$, $\mathbf{A}_1$ be the same attributes in these two branches and $\mathbf{a}_1$ be a branch's value of $Br[\mathbf{A}_1]$, where $\mathbf{A}_1 \subseteq \mathbf{A}$. Then, Br is in conflict with Br' with respect to $\mathbf{A}$ if and only if

- (1) when $\mathbf{A}_1 \neq \emptyset$, $\exists \mathbf{a}_1, \mathbf{a}_1 \in Br[\mathbf{A}_1], \mathbf{a}_1 \in Br'[\mathbf{A}_1]$ and $C_{k1} \neq C_{k2}$, or
- (2) when $\mathbf{A}_1 = \emptyset$, $C_{k1} \neq C_{k2}$.

According to theorem 4, when $C_{k1} = C_{k2}$, branches, $Br \rightarrow C_{k1}$ and $Br' \rightarrow C_{k2}$, are never in conflict with each other;

therefore, to identify all the irrelevant values of a branch Br , we need only to consider those branches, Br' , whose leaves are different from C_{kl} . Moreover, to make the process of removing all irrelevant values of a branch more efficient, we can recursively apply the above theorems until the node P is the root of the tree. The corresponding algorithm is shown as Fig. 2.

Input: A decision tree

Output: A set of rules without irrelevant conditions;

Let $Br = \{Br_1, \dots, Br_m\}$; /* the branches of the decision tree */

For each branch Br' in Br **Do**

{Let $Br' = Br'[A_1] \wedge \dots \wedge Br'[A_k] \rightarrow C_i$

For $j = k$ downto 1 **Do**

{Apply theorem 2, 3 and 4 to check whether $Br'[A_j]$ is an irrelevant value;

If $Br'[A_j]$ is an irrelevant value **Then** remove $Br'[A_j]$ from Br' ;}
 Represented Br' by a rule;}

Fig. 2. Converting a decision tree to a set of rules without irrelevant conditions

4 CASE STUDY: USING HYPOTHYROID DISEASE EXAMINATIONS

In this section, we discuss the problem of applying the decision tree to *hypothyroid disease* examinations. To economized the medical expenditures and alleviated the torture of the patient in the medical investigation, we have to avoid unnecessary examinations to the patient. However, when we follow the rules that generate by the tree to do medical examinations, some unnecessary examinations may be performed. We address this point of view by the following example.

Example. Let us consider the decision tree, as shown in fig 2. The original rules and the resultant rules without irrelevant conditions are shown in Table 1. Comparing these two sets of rules, we could find out that our rules are more concise and comprehensible than the original rules. The irrelevant conditions of the original rules are represented by boldfaces in Table 1.

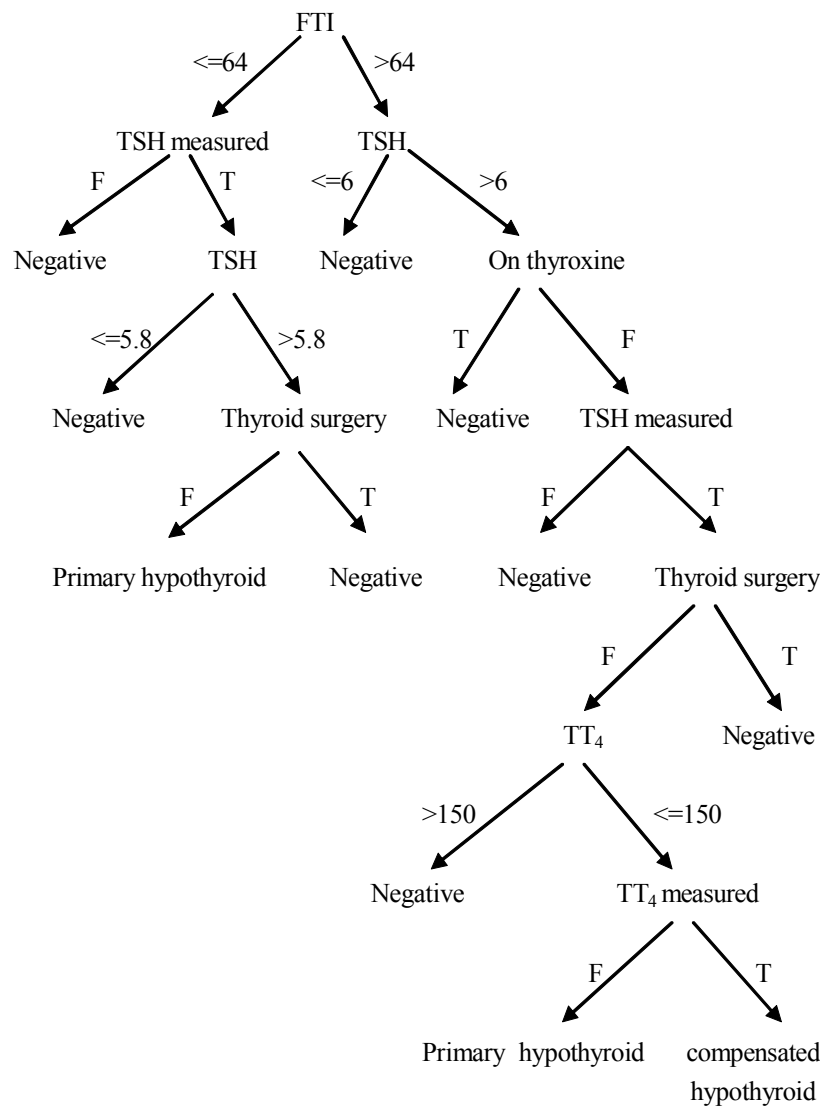


Fig 2. Decision tree for thyroid examinations

Comparing between the rules with and without irrelevant conditions in the table 1 with respect to the decision tree of fig. 2, some characters can be found. Firstly, the decision tree model recommends the doctors to do many investigations for the patient which might not be required. For example, in our rules, the rules of 1,2,4 show that it is much simpler for physicians to make diagnosis from the investigation results. In our algorithm, a history with negative TSH measured result or thyroid

surgery, or TSH less than 5.8 represents the negative result of the diagnosis. That means if a patient with a history of thyroid surgery or thyroxin medication, his diagnosis is always negative and does not need to have the other investigation such as TT4 or FT4. It is same that if a patient with a measurement of TSH below 5.8, he also doesn't need to do the upper testing either. Only when the TSH is above 5.8, he may need to do FTI, or TT4 to make a correct diagnosis.

Table 1. The rules convert from the decision tree

The original rules without considering irrelevant values problem	The resultant rules without irrelevant conditions
$FTI \leq 64 \wedge THS \text{ measured} = f \rightarrow \text{negative}$	(1) $THS \text{ measured} = f \rightarrow \text{negative}$
$FTI \leq 64 \wedge THS \text{ measured} = t \wedge TSH \leq 5.8 \rightarrow \text{negative}$	(2) $TSH \leq 5.8 \rightarrow \text{negative}$
$FTI \leq 64 \wedge THS \text{ measured} = t \wedge TSH > 5.8 \wedge \text{thyroid surgery} = f \rightarrow \text{primary hypothyroid}$	(3) $FTI \leq 64 \wedge THS \text{ measured} = t \wedge TSH > 5.8 \wedge \text{thyroid surgery} = f \rightarrow \text{primary hypothyroid}$
$FTI \leq 64 \wedge THS \text{ measured} = t \wedge TSH > 5.8 \wedge \text{thyroid surgery} = t \rightarrow \text{negative}$	(4) $\text{thyroid surgery} = t \rightarrow \text{negative}$
$FTI > 64 \wedge TSH \leq 6 \rightarrow \text{negative}$	(5) $FTI > 64 \wedge TSH \leq 6 \rightarrow \text{negative}$
$FTI > 64 \wedge TSH > 6 \wedge \text{on thyroxine} = t \rightarrow \text{negative}$	(6) $FTI > 64 \wedge \text{on thyroxine} = t \rightarrow \text{negative}$
$FTI > 64 \wedge TSH > 6 \wedge \text{on thyroxine} = f \wedge THS \text{ measured} = f \rightarrow \text{negative}$	The resultant rule is equal to (1)
$FTI > 64 \wedge TSH > 6 \wedge \text{on thyroxine} = f \wedge THS \text{ measured} = t \wedge \text{thyroid surgery} \rightarrow \text{negative}$	The resultant rule is equal to (4)
$FTI > 64 \wedge TSH > 6 \wedge \text{on thyroxine} = f \wedge THS \text{ measured} = t \wedge \text{thyroid surgery} = f \wedge TT_4 > 150 \rightarrow \text{negative}$	(7) $FTI > 64 \wedge TT_4 > 150 \rightarrow \text{negative}$
$FTI > 64 \wedge TSH > 6 \wedge \text{on thyroxine} = f \wedge THS \text{ measured} = t \wedge \text{thyroid surgery} = f \wedge TT_4 \leq 150 \wedge TT_4 \text{ measured} = f \rightarrow \text{primary hypothyroid}$	(8) $TSH > 6 \wedge THS \text{ measured} = t \wedge \text{thyroid surgery} = f \wedge TT_4 \leq 150 \wedge TT_4 \text{ measured} = f \rightarrow \text{primary hypothyroid}$
$FTI > 64 \wedge TSH > 6 \wedge \text{on thyroxine} = f \wedge THS \text{ measured} = t \wedge \text{thyroid surgery} = f \wedge TT_4 \leq 150 \wedge TT_4 \text{ measured} = t \rightarrow \text{compensated hypothyroid}$	(9) $FTI > 64 \wedge TSH > 6 \wedge THS \text{ measured} = t \wedge \text{thyroid surgery} = f \wedge TT_4 \leq 150 \wedge TT_4 \text{ measured} = t \rightarrow \text{compensated hypothyroid}$

Secondly, the resultant rules show that the better way for a clinician to make a correct diagnosis without doing every investigation available. Those patients without a thyroid surgery history and thyroxine medication need to have the test of FTI, TT₄ and “TSH measured” further to make the diagnosis. Therefore, we can optimize lab investigations for diagnosing a possible hypothyroid patient by the resultant rules. Doing less investigation is both good for patients and our society. For each patient, she might not spend time and money to do some lab examination which is not needed for diagnosis. For our society, it is a waste if we spend our fiscal expenditure for any medical procedure that is not quiet needed. The doctors’ time is limited, the patients are suffering from the disease both physically and mentally, the ever-increasing of medical expenditure is a huge burden to our

government. So it is a great benefit for all these three parts if we can find way to cut down the medical expenditure, such as the lab investigations for the diagnosis.

5 CONCLUSIONS

Medical problem is a field where “decision tree” algorithm is usually applied. As the cost of healthcare system is rising again and again, it is urgent for government to do anything to inhibit the fast increasing the medical expenditure while maintaining the quality. Using the decision tree model could be helpful for us to retrieve the relevant medical tests needed in clinical cases under different condition. However, the new rules we adopted here and testified in a hypothyroid model

showed more efficient and hence more useful to exclude the irrelevant conditions. Therefore, the new rules could be a good technique for physicians to find the easiest and

fastest way to make a diagnosis and thus cut down the indispensable expenditures spending in lab investigations.

REFERENCES

- [1] BOHANEK, M. — RAJKOVIC, V.: Expert system for the decision making. *Sistemica*, Vol. 1, 1990, No. 1, pp.145-157.
- [2] BREIMAN, L. — FRIEDMAN, J. H. — OLSHEN, R. A. — STONE, C. J.: *Classification and regression trees*. Wadsworth 1984.
- [3] CAI, Y. — GERCONI, N. — HAN, J.: An attribute-oriented approach for learning classification rules from relational databases, *Proc. of Conf. on Data Engineering*, Los Angeles 1990, pp.281-288.
- [4] CHENG, J. — FAYYAD, U. M. — IRANI, K. B. — QIAN, Z.: Improved decision trees: a generalized version of ID3. *Proc. of the Fifth Int. Conf. on Machine Learning*, 1988, pp. 100-108.
- [5] CHI, Z. — YAN, H.: ID3-derived fuzzy rules and optimized defuzzification for handwritten numeral recognition. *IEEE Tran. on Fuzzy Systems*, Vol. 4, 1996, No. 1, 24-31.
- [6] FAYYAD, U. M. — IRANI, K. B.: A machine learning algorithm (GID3*) for automated medical knowledge acquisition improvements and extensions. General Motors Research Report CS-634, Warren, MI: GM research labs 1991.
- [7] FAYYAD U M.: Branching on attribute values in decision tree generalization. *Proc. Twelfth National Conference on Artificial Intelligence AAAI-94*, Seattle, Washington 1994, pp. 104-110.
- [8] KAMBER, M — WINSTONE, L. — GONG, W — CHENG, S. — HAN, J.: Generalization and decision tree induction: Efficient, classification in data mining. *Proc. Seventh Int. Workshop on Research Issues in Data Engineering* 1997. pp. 11-121.
- [9] MAHER P. E. — GLAIR, D. ST.: Uncertain reasoning in an ID3 machine learning framework *Proc. of the Second IEEE Int. Conf. Fuzzy Syst.* 1993, pp. 7-12.
- [10] MANGASARIAN, O. L. — WOLBERG, W. H.: Cancer diagnosis via linear programming. *SIAM News*, Vol. 23, 1990, No. 25.
- [11] KUO Wen-Jia, CHANG Ruey-Feng, CHEN Dar-Ren, CHENG CHUN LEE, Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. *Breast cancer research and treatment*. 2001, vol. 66, no1, pp. 51-57.
- [12] Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine*, 2005, 34; 2: 113-127.
- [13] Neumann A, Holstein J, Lepage E, et al. Measuring performance in health care: case-mix adjustment by boosted decision trees. *Artificial intelligence in medicine*, 2004; 32: 97-113.
- [14] Jerez-Aragones J, Gomez-Ruiz J, Ramos-Jimenez G, et al. A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artificial intelligence in medicine*, 2003; 27: 45-63.
- [15] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial intelligence in medicine*, 2001; 23: 89-109.
- [16] Ou MH, West G, Lazarescu M, et al. Dynamic knowledge validation and verification for CBR teledermatology system. *Artificial intelligence in medicine*, In press.
- [17] Ding-An Chiang, Wei Chen, Yi-Fan Wang, The Irrelevant Values Problem In The ID3 Tree *Computers and Artificial Intelligence*, Vol, 19. 2000. 169-182