ISSN 2224-087X. Електроніка та інформаційні технології. 2019. Випуск 11. С. 73–80
Electronics and information technologies. 2019. Issue 11. P. C. 73–80

UDK 004.8: 004.89: 519.7 DOI: https://doi.org/10.30970/eli.11.7

GENERAL SCHEME FOR CONSTRUCTING THE MOST COMPLEX LOGICAL TREE OF CLASSIFICATION IN PATTERN RECOGNITION DISCRETE OBJECTS

I. Povhan

Department of software systems,
State High School "Uzhgorod National University",
st. Zankovetska №89B, 88000 Uzhgorod, Ukraine
igor.povkhan@uzhnu.edu.ua

The work raises an important question in the theory of recognition of discrete objects associated with the General scheme of construction of the most complex resulting logical classification tree. The General step-by-step scheme of construction of such tree is offered, and the estimation of complexity of the received graph-circuit models is given. Numerical estimates in the long term allow us to develop effective models of minimization schemes of logical classification trees. The result obtained is of fundamental importance in the problem of estimating the stability of the maximum logical classification tree with respect to the permutation of tiers and the effect of such a permutation on the overall complexity of the tree. The work is relevant for all methods of pattern recognition in which the resulting classification function can be represented as a logical tree.
Key words: tasks of pattern recognition, logical tree, graph-scheme of the model.

This work is a logical continuation of a cycle of works [1,2,3] dedicated to such an important issue associated with the logical trees classification, as a matter of minimizing, the study of stability on the reshuffling of tiers, evaluation of complexity of the largest trees. Here is explored the complexity of graph circuit patterns (logical classification trees) which are constructed in the learning process of the system of recognition (the logical tree classification actually represents the generated function recognition). For this estimated the complexity of the wood that is used in the scheme of distributed choice signs [4] for recognition of $n-$ dimensional discrete sets (objects).

Note that the main focus of the work will be paid exactly the upper estimate of the complexity of the logical tree. This is especially important due to the fact that these assessments reveal that memory must have real, autonomous system of recognition, which is constructed by the algorithms given in the works [5,6].

At the beginning, we will assume that fixed some system of signs $P_1, P_2, ..., P_n$, $(0 \le P_i \le k_i - 1)$, $(i-1,2,...,n)$, and a system of symbols $O_0, O_1, ..., O_{k-1}$. We also consider that the signs $P_1, P_2, ..., P_n$ - sorted (for example in the order in which they are recorded here).

Will consider all regular logical tree, i.e. in which the tops of the $i-$tier of the stands of signs P_i, $(1 \le i \le n)$, and the tops of the $(n+1)-$ tier are symbols $O_0, O_1, ..., O_{k-1}$ value functions $f_R(P_1, ..., P_n)$.

Let the given regular tree D . Through the label $|D|$, the number of all the different labels that are received as a result of the process of placing labels on the tree D . Obviously $|D|$, that is the number of nodes [1]. The problem lies in the fact that among all regular trees D to find this tree, for which the value $|D|$ will be the biggest. The maximum tree label $D_{\max}$.

Definition 1. The set of all vertices, standing in circles with numbers i , which $l \le i \le s$, $(1 \le l \le s \le n)$ will call stripe. Stripe, consisting of circles with numbers i , $l \le i \le s$, let through D_s^l . Through $|D_s^l|$ denote the number of all the different between the labels that are in stripe D_s^l . Prove the following.

Statement 1. Let in the $i-$ tier of regular wood D with different tops are different tags. Then the parts D_i^1 at different level are different.

This follows directly from the functions of the above process of placing tags. Indeed, if the $i-$tier with different tops are different tags, then all level $i-1$ tier will stand the different tags. In addition, all tags that are at the tops of the $(i-1)-$ tier, in this case, will be different from all the labels that stand at the mountain tops, the i , $i+1,...,n$, $(n+1)$ - tier.

Reflecting similar, you can see all the tags of the $(i-2)-$tier, which are at different level also vary among themselves and from all levels of labels $i, i+1, i+2,...,n, n+1$. From the $(i-2)-$tier, we can move to $i-3$, from $i-3$ to $i-4$, and so on.

So, all tags of the $(i-j)$ - tier $(1 \le j \le i-1)$, which are at different level, differ between themselves and all the labels that stand at the tops of the $i-j+1, i-j+2,....,n,(n+1)$ - tiers. And this in turn means that the statement is proved.

Note that the above, it is still the next statement.

Statement 2. Provided the *Statement 1* all tags that are in stripe D_{i-1} , different from all the labels are in stripe D_{n+1}^i .

Let us now directly to the construction of the tree $D_{\max}$. Recall that all vertices α in the $i-$tier of dependent traits $P_i, P_{i+1},...,P_n$. Thus, each vertex α of the $i-$ tier is a certain function $f_\alpha(P_i, P_{i+1},...,P_n)$.

Yes, $P_j \in \{0,1,2,...,k_j-1\}$, $(j = i, i+1,...,n)$ and all the functions f_α taking values in the set $\{O_0, O_1,...,O_{k-1}\}$, then the number of all functions of the appearance $f_\alpha(P_i, P_{i+1},...,P_n)$ will be equal $K^{K_i K_{i+1} ... K_n}$. The number of vertices of the $i-$tier is $K_1 K_2 ... K_{i-1}$.

It is obvious that the different peaks of the $i-$tier, then and only then can stand various functions f_α (this can be interpreted as what the various tops $i-$ tier stand different tags) when $K_1 K_2 ... K_{i-1} \le K^{K_i K_{i+1} ... K_n}$. If $K_1 K_2 ... K_{i-1} > K^{K_i K_{i+1} ... K_n}$ then in the $i-$tier of arbitrary trees D would though there will always be two different peaks with the same labels.

As the size $K_1, K_2,...,K_{i-1}$ and $K^{K_i K_{i+1} ... K_n}$ growth, according to grow i , then m , there is that:

$$\left.\begin{aligned} K_1K_2...K_{m-1} \le K^{K_mK_{m+1}...K_n} \\ K_1K_2...K_m > K^{K_{m+1}K_{m+2}...K_n} \end{aligned}\right\}. \tag{1}$$

Story number m call lower breaking.

Consider first the case when the first ratio (1) equality, that is $K_1K_2...K_{m-1} = K^{K_mK_{m+1}...K_n}$.

In this case, the number of all vertices of the $m-$ stage coincides with the number of all functions $f_\alpha(P_i, P_{i+1},...,P_n)$.

The tree $D_{\max}$ is constructed in the following way: each top of the $m-$ tier α is placed in line with some function $f_\alpha(P_i, P_{i+1},...,P_n)$, with different vertices α, β and $m-$ tier put in line with the various functions f_α and f_β. As a result, we get some tree D^*. Show $D^* = D_{\max}$, that the tree D^* take the maximum number of tags.

Let set up a tree D. It is necessary to show that $|D^*| \ge |D|$. Let us break trees D^* and D on stripes $(D^*)^1_{m-1}, (D^*)^m_{n+1}, D^1_{m-1}, D^m_{n+1}$. From the construction of tree D^* and affirmations (1), (2) it follows that none of the labels stripe $(D^*)^1_{m-1}$ does not coincide with any of the stripe $(D^*)^m_{n+1}$. Hence we obtain:

$$|D^*| = |(D^*)^1_{m-1}| + |(D^*)^m_{n+1}|. \tag{2}$$

For a tree D will have the following:

$$|D| = |D^1_{m-1}| + |D^m_{n+1}|. \tag{3}$$

According to №1 in different tops stripes $(D^*)^1_{m-1}$ are different tags. This in turn means that:

$$|(D^*)^1_{m-1}| \ge |D^1_{m-1}|. \tag{4}$$

Obviously, all the tags α at the tops of tiers $m, m+1,....,n+1$ is a function f_α that depend only on the signs $P_m, P_{m+1},...,P_n$.

So we actually performed the original task and built exactly complicated tree D^*.

Since all tags that are on the $m-$ floor of the tree D^*, will use all functions that depend $P_m, P_{m+1},...,P_n$, on:

$$|D^m_{n+1}| \le |(D^*)^m_{n+1}|. \tag{5}$$

(2), (3), (4), (5) get $|D^*| \ge |D|$ that for any tree D. This means that D^* there is a maximum.

Now we calculate the number of labels in the tree D^*. To build D^* directly follows that $|D^*| \ge |(D^*)^1_m|$.

So in stripe $(D^*)^1_m$ at different level are a variety of tags

$$(D^*)^1_m = I + K_1 + K_1K_2 + ... + K_1K_2 * ... * K_{m-1} = I + \sum_{j=2}^{m} K_1...K_{j-1}.$$

So this means:

$$\left|D^*\right| = I + \sum_{j=2}^{m} K_1 K_2 ... K_{j-1} . \quad (6)$$

If $K_1 = K_2 = ... = K_n = r, (r \geq 2)$, then the ratio (6) has the form:

$$\left|D^*\right| = \frac{r^m - 1}{r - 1} . \quad (7)$$

Moreover, the ratio $K_1 K_2 ... K_{m-1} = K^{K_m K_{m+1} ... K_n}$ in this case is:

$$r^{m-1} = K^{r^{n-(m-1)}} . \quad (8)$$

On the basis of the ratio of (8), consider the equality:

$$r^l = K^{r^{n-l}} . \quad (9)$$

We are interested in the case where this equation has a solution in integers l, and that is the solution.

Note that the equation (9) has no more than one solution in the integers l, because the left part (9) is strictly increasing and the right is strictly comes with growth l.

Consider only case $r = K$, since this case is mostly of interest to practice. Then the ratio (9) has the form: $K^l = K^{K^{n-l}}$. The last value is equivalent to the following:

$$l = K^{n-l} . \quad (10)$$

The next step we fix l in (10) and will change n. Show that the ratio (10) then and only then has a solution in integers l, when n has the form $n = K^\delta + \delta$, where $\delta -$ is a number.

Indeed, if $n = K^\delta + \delta$, then putting $l = K^\delta$ a get $K^{n-l} = K^{K^\delta + \delta - K^\delta} = K^\delta = l$.

So, in the case $n = K^\delta + \delta$ of we have: $m - 1 = l = K^\delta$.

Away and (7) are:

$$\left|D^*\right| = \frac{K^m - 1}{k - 1} \approx \frac{K^m}{k - 1} = \frac{K^{K^\delta + 1}}{k - 1} . \quad (11)$$

Note, if $K = 2$ the value (11) has the form:

$$\left|D^*\right| \approx 2^{2^\delta + 1} . \quad (12)$$

Also, numbers $\frac{K^{K^\delta + 1}}{k - 1}$, and $2^{2^\delta + 1}$ according to (11) and (12) are different from D^* not more than per unit.

This is easy to show that the number of all the vertices d of any regular tree D when is $K_1 = K_2 = ... = K_n = r = K$ equal to:

$$d = \frac{K^{n+1} - 1}{K - 1} . \quad (13)$$

Note that while $K = 2$ we have $d = 2^{n+1} - 1 \cong 2^{n+1}$.

Value:

$$E = \frac{d}{\left|D^*\right|} \quad (14)$$

describes generally the effect obtained during the transition from tree D to $\tilde{D}$. Substituting into (14) value (7) and (13) we get:

$$E = \frac{d}{\left|D^*\right|} = \frac{\frac{K^{n+1}-1}{K-1}}{\frac{K^m-1}{K-1}} = \frac{K^{n+1}-1}{K^m-1} \approx \frac{K^{n+1}}{K^m} = K^{n-m+1} = K^{\delta}.$$

The latter comes from the fact that $n = K^{\delta} + \delta$ та $m = K^{\delta} + 1$.

Since the value K^{δ} is greater than δ, then it can be put $n \approx K^{\delta}$.

Thus, when $K_1 = K_2 = ... = K_n = r = K$ and $n = K^{\delta} + \delta$, in the transition from D to $\tilde{D}$ the complexity of the tree D decreases almost n times. This number d and $\left|D^*\right|$ case, $K_1 = K_2 = ... = K_n = r = 2$ and $n = 2^{\delta} + \delta$, $\delta = 1,2,3,4$ (table 1).

Table 1.

Case logic tree $r = 2$, $\delta = 1,2,3,4$

δ	**1**	**2**	**3**	**4**
$n = 2^{\delta} + \delta$	3	6	11	20
d	15	127	4095	2097151
$\left|D^*\right|$	7	31	511	131071

We are left to show that if n you can not submit the form $n = K^{\delta} + \delta$, then the equation (10) has no solution in positive integers l.

Note first that for any integers of positive numbers K and n, there $(k, n \geq 2)$ is one thing δ that:

$$K^{\delta} \leq n < K^{\delta+1}. \tag{15}$$

Note that if $n = K^{\delta} + \delta$ then δ satisfies (15).

The latter follows from that $K \geq 2$ and $K^{\delta} > \delta$.

So, let δ satisfies condition (15). In this case the possible three cases:

1) $n - K^{\delta} = \delta$;
2) $n - K^{\delta} < \delta$; (16)
3) $n - K^{\delta} > \delta$.

In the case of (16.1), as has been shown above, the equation (10) has a solution in integers of positive numbers l.

Show that in cases (16.2) (16.3) equation (10) has a solution in integers numbers l.

Consider the case (16.2) numbers:

$$l = K^{\delta} + j - \delta,$$
$$l + 1 = K^{\delta} + j - \delta + 1, \tag{17}$$

Where $j = n - K^{\delta}$. In the case of (16.2) $j < \delta$ and $n = K^{\delta} + j$. Calculate the value K^{n-l} and $K^{n-(l+1)}$ data l and $l+1$:

$$K^{n-l} = K^{K^{\delta}+j-l} = K^{K^{\delta}+j-(K^{\delta}+j-\delta)} = K^{\delta},$$
$$K^{n-(l+1)} = K^{K^{\delta}+j-l-1} = K^{K^{\delta}+j-K^{\delta}-j+\delta-1} = K^{\delta-1}. \tag{18}$$

On the other hand:

$l+1=K^{\delta}+j-\delta+1\geq K^{\delta}-(\delta-1)=K*K^{\delta-1}-(\delta-1)=K^{\delta-1}+$

$+(K-1)*K^{\delta-1}-(\delta-1)\geq K^{\delta-1}+(K^{\delta-1}-(\delta-1))>K^{\delta-1}=K^{n-(l+1)}$.

The latter follows from that $K^{\delta-1}>\delta-1$.

So, we have in the case (16.2)

$l<K^{n-l}$,

$l+1>K^{n-(l+1)}$. (19)

From (19) and the fact that the left and right parts of the equation (10) while increasing l, respectively, is strictly increasing and decreasing, it follows that in the case of the equation (10.2) has no solution in integers numbers l.

Consider the case of (16.3). In this case, we have:

$$\left.\begin{aligned} l=K^{\delta}+j-(\delta+1) \\ l+1=K^{\delta}+j-\delta \end{aligned}\right\}. \quad (20)$$

Here $j=n-K^{\delta}$.

We calculate with these numbers K^{n-l} and $K^{n-(l+1)}$:

$K^{n-l}=K^{K^{\delta}+j-(K^{\delta}+j-(\delta+1))}=K^{\delta+1}$,

$K^{n-(l+1)}=K^{K^{\delta}+j-(K^{\delta}+j-\delta)}=K^{\delta}$. (21)

From (20) and (21), you can get:

$l=K^{\delta}+(j-(\delta+1))<K^{\delta}+j<K^{\delta}+(K-1)*K^{\delta}=K^{\delta+1}=K^{n-l}$,

$l+1=K^{\delta}+j-\delta>K^{\delta}=K^{n-(l+1)}$. (22)

Note that here we use the ratio $(K-1)*K^{\delta}>j$. The latter follows from that $K^{\delta}\leq n\leq K^{\delta+1}$.

With ratio (22) and with the fact that in equation (10) left and right parts with an increase n, in accordance with strictly increasing and decreasing, it follows that in the case of (16.3) equation (10) has no solution in integers l.

Thus from the above it follows that when the number K,δ,n of satisfying criteria (15), (16) and (21), then the number m of the layer break in the tree D^{*} is found by the formula:

$m=l+1=K^{\delta}+j-\delta+1$. (23)

Note that here $j=n-K^{\delta}$.

If the number K,δ,n of satisfying criteria (15) and (16.3), then the number m of the layer break is according to the formula:

$m=l+1=K^{\delta}+j-\delta$. (24)

Again, note that here $j=n-K^{\delta}$.

Result. So all the goals in the work have been achieved. The General scheme of construction of the maximum logical classification tree is proposed, and the complexity of the obtained graph-circuit models is estimated. The paper presents a numerical evaluation of the complexity of the logical tree of classification. The obtained result is of fundamental importance in the problems of minimization of logical classification trees. At the end, we note again that the result is relevant for all methods of pattern recognition in which the resulting classification function can be represented as a logical tree.

REFERENCES

1. *Vashchuk F. G.* The problem of estimating the complexity of the logic trees, recognition and general method optimization / F. G. Vashchuk, Y. A. Vasilenko, I. F. Povhan // Scientific and technical journal "European Journal of Enterprise Technologies". - 2011. – 6/4 (54). – P. 24–28.
2. *Vashchuk F. G.* Overall assessment of minimization of logical tree structures / F. G. Vashchuk, Y. A. Vasilenko, I. F. Povhan // Scientific and technical journal "European Journal of Enterprise Technologies". – 2012. – 1/4 (55). – P. 29–33.
3. *Povhan I. F.* Minimization of logic tree structures for the task of recognition images / I. F. Povhan, Yu. A. Vasilenko, E. Yu Vasylenko, M. I. Kovacs, O. D. Nickovic // Scientific and technical journal "European Journal of Enterprise Technologies". – 2004. – 3[9]. – P. 12–16.
4. *Povhan I. F.* Conceptual framework of recognition systems based on the method of extensive array of signs / Povhan I. F., Vasilenko Yu. A., Vasilenko E. Y. // Scientific and technical journal "European Journal of Enterprise Technologies". – 2004. – №7[1]. – P. 13–15.
5. *Povhan I. F.* Basis, the importance of discrete traits (analysis of some approaches) / Povhan I. F., Vashchuk F. G., Vasilenko Yu. A., Povhan L. S // East European journal of advanced technologies. – 2010. – No. 5/4 (47). – P. 71–75.
6. *Vasylenko Y. A.* Group and individual assessment of the importance of boolean arguments / Vasylenko Y. A., Vashchuk F. G., Pohan I. F., Polischuk F. G. // Bulletin of national technical University "Kharkiv Polytechnic Institute". – 2011. – №53.– P. 57–64.

ЗАГАЛЬНА СХЕМА ПОБУДОВИ САМОГО СКЛАДНОГО ЛОГІЧНОГО ДЕРЕВА КЛАСИФІКАЦІЇ В ЗАДАЧАХ РОЗПІЗНАВАННЯ ДИСКРЕТНИХ ОБ‘ЄКТІВ

І. Повхан

Кафедра програмного забезпечення систем,
ДВНЗ "Ужгородський національний університет",
вул. Заньковецкої №89Б, 88000 Ужгород, Україна
igor.povkhan@uzhnu.edu.ua

Робота піднімає важливе питання в теорії розпізнавання дискретних об‘єктів пов‘язане з загальною схемою побудови самого складного результуючого логічного дерева класифікації. Представлення результуючої схеми розпізнавання на основі логічного дерева, дозволяє забезпечити мінімальну форму системи розпізнавання та високу якість класифікації. В роботі пропонується загальна покрокова схема побудови такого логічного дерева, та дається оцінка складності отриманих граф-схемних моделей.

Побудовану схему розпізнавання у вигляді логічного дерева можна представити або в ДНФ, або в КНФ формі. Так дерево розпізнавання, яке являє собою певне правило класифікації, можна представити за допомогою відповідної логічної функції. Отже важливою проблемою при побудові схеми розпізнавання (відповідного логічного дерева) такого типу буде проблема синтезу логічної функції, яка еквівалентна даному дереву розпізнавання. Зі

збільшенням числа аргументів логічної функції (результуючого дерева) швидко зростає складність одного з етапів синтезу функції – етапу мінімізації. Виходом з даного положення може бути не знаходження її мінімальної форми, а представлення у вигляді декомпозиції функцій (відповідних дерев). Важливими перевагами методу дерева для мінімізації логічних функції є те, що з деревом досить просто працювати при великій кількості аргументів. Крім того, дуже просто автоматизувати процес мінімізації за рахунок написання простої комп‘ютерної програми.

Зауважимо, що в роботі також виведені числові оцінки в перспективі дозволяють розробити ефективні моделі схем мінімізації логічних дерев класифікації. Отриманий результат принципово важливий в задачі оцінки стійкості максимального логічного дерева класифікації щодо перестановки ярусів та впливи такої перестановки на загальну складність дерева. Дане питання має принциповий характер, тому що впливає на загальну складність отриманої схеми класифікації та обсяг пам‘яті необхідної для зберігання та роботи відповідної системи розпізнавання. Робота актуальна для всіх методів розпізнавання образів в яких отримана функція класифікації може бути представлена у вигляді логічного дерева.

Ключові слова: задачі розпізнавання образів, логічне дерево, граф-схемні моделі.

Стаття: надійшла до редакції 11.12.2018,
доопрацьована 22.01.2019,
прийнята до друку 24.01.2019.