COMPLEXITY AND LEVEL LOGICAL DESCRIPTION OF CLASSES FOR PATTERN RECOGNITION PROBLEMS

Kosovskaya T.M.

St.Petersburg Institute of Informatics and Automatization

Abstract

Various pattern recognition problems may be reduced to the proof of deducibility of propositional and predicate formulas from the set of atomic formulas. Here in the frameworks of logic-axiomatical approach to the pattern recognition problem the guaranteed time effectiveness of the solving algorithms is proved by means of consideration of several variants of setting of a problem including the problem of analysis of a compound object.
By the author a level description of classes was suggested. Such a description allows to decrease upper bounds of step number of algorithms solving pattern recognition problems. In this paper conditions for such a decreasing are proved.

1. Main definitions

In [1] such an approach to the pattern recognition problem was suggested.

Let Ω be a set of recognizable objects ω. Such an object may be regarded as a set of its elements $\omega = \{\omega_1, \ldots, \omega_t\}$. There exists a partition of the set Ω on K classes, i.e. $\Omega = \bigcup_{k=1}^{K} \Omega_k$. $p_1, \ldots, p_n$ are predicates defined on elements of ω.

The description of the class Ω_k may be represented by a formula A_k which has the form of disjunction of simple conjunctions of atomic formulas or its negations.

The description $S(\omega)$ of an object ω is the set of all true formulas in the form $p_i(\overline{\tau})$, where $\overline{\tau}$ is a string of elements of ω.

Identification problem consists in the checking whether the object ω or its part belongs to the class Ω_k. This problem may be solved by the proof of of deducibility of the formula

$$\exists \overline{y}(y \subseteq \omega \ \& \ A_k(\overline{y}))$$

from $S(\omega)$.

Classification problem consists in the finding of all such numbers k that ω belongs to the class Ω_k. This problem may be solved by the proof of of deducibility of the formula

$$\bigvee_{k=1}^{M} A_k(\overline{\omega})$$

from $S(\omega)$ for some ordering $\overline{\omega}$ of the set ω.

Problem of analysis of a compound object consists in the finding of all such numbers k and all parts $\overline{\tau}$ from ω that τ belongs to the class Ω_k. This problem may be solved by the constructive proof of deducibility of the formula

$$\bigvee_{k=1}^{M} \exists \overline{y}(y \subseteq \omega \;\&\; A_k(\overline{y}))$$

from $S(\omega)$.

2. Complexity of recognition problems with global characteristic of objects

Let predicates $p_1, \dots, p_n$ describe the whole object but not its parts. I such a case these predicates may be regarded as propositional (or boolean) variables. Then the description of the class Ω_k is a formula in the Disjunctive Normal Form (DNF) with propositional variables $p_1, \dots, p_n$. Identification problem is a partial case of classification problem and problem of analysis of a compound object can not be regarded.

Recognition problems with global characteristic of objects may be solved by polynomial algorithms. More precisely the following theorems are valid.

Theorem 1. *For every positive integer n there exists a Turing machine which for every string of logical constants $(\alpha_1, \dots, \alpha_n)$ and a DNF formula A_k with the length N_k checks whether this formula is true on this string of logical constants. The number of steps of such a Turing machine is bounded by $n + N_k$.*

Theorem 2. *The number of resolution rule applications (with the use of linear strategy) for deducibility of a DNF formula A_k from the set of formulas $(p_1^{\alpha_1}, \dots, p_n^{\alpha_n})$ is not greater then the number of propositional variables occurrences in the formula A_k.*

Theorem 3. *The number of sequential calculus rule applications for deducibility of a sequent $p_1^{\alpha_1} \dots p_n^{\alpha_n} \vdash A_k$ where A_k is a DNF formula is not greater then the number of occurrences of logical connectives $\vee$ and $\&$ in the formula A_k.*

3. Complexity of recognition problems with local characteristic of objects

In the case that predicates $p_1, \dots, p_n$ describe the parts of an object recognition problems are seams to be NP-complete. If we decide such problems for all possible class descriptions it is true.

Theorem 4. *The problem of checking wether the formula $A(\overline{y})$ is satisfiable on the set $\omega = \{\omega_1, \dots, \omega_t\}$ is NP-complete.*

Corollary. *If $A_k(\overline{y})$ is a parameter of a problem then identification problem is NP-hard.*

Theorem 5. *For every DNF formula $A(\overline{y})$ there exists an algorithm which checks wether the formula $A(\overline{y})$ is satisfiable on the set $\omega = \{\omega_1, \dots, \omega_t\}$ and the number of steps of such an algorithm is $O(t^{|y|})$, where $|y|$ – the number of variables in $\overline{y}$.*

Corollary. *For every DNF for-*

mula $A_k(\overline{y})$ identification problem may be solved by a polynomial algorithm the exponent of which is the number of variables in $\overline{y}$.

Theorem 6. *The problem of checking wether the formula $A(\overline{y})$ is true on some permutation of the set $\omega = \{\omega_1, \ldots, \omega_t\}$ is NP-complete.*

Corollary. *If $A_k(\overline{y})$ is a parameter of a problem then classification problem is NP-hard.*

In the next theorem the number of variables in $\overline{y}$ is equal to the number of elements in ω.

Theorem 7. *For every DNF formula $A(\overline{y})$ there exists an algorithm which checks wether the formula $A(\overline{y})$ is satisfiable on some permutation of the set $\omega = \{\omega_1, \ldots, \omega_t\}$ and the number of steps of such an algorithm is $O(t!)$, where $|y|$ – the number of variables in $\overline{y}$.*

Corollary. *For every DNF formula $A_k(\overline{y})$ classification problem may be solved by a polynomial algorithm the exponent of which is the number of variables in $\overline{y}$.*

Theorem 8. *If $A_1(\overline{y}_1), \ldots, A_K(\overline{y}_K)$ is a parameter of a problem then the problem of analysis of compound object is NP-hard.*

Theorem 9. *For every collection of DNF formulas $A_1(\overline{y}_1), \ldots, A_K(\overline{y}_K)$ the problem of analysis of compound object may be solved by a polynomial algorithm the exponent of which is the number of variables in $|y_1| + \ldots + |y_K|$.*

4. Level logical description of classes

In [2] objects, the structure of which allows to extract some simple parts and to describe classes in the terms of these parts, are regarded. It may be done by selecting of "often" occurred subformulas $P_j^1(\overline{x})$ of formulas $A_1(\overline{y}_1), \ldots, A_K(\overline{y}_K)$. Besides the equivalence system in the form ${p_i}^1(\overline{x}) \leftrightarrow P_i^1(\overline{x})$ is written down. Here ${p_i}^1$ – new predicates that will be called first level predicates. $A_1^1(\overline{y}_1^1)$, $\ldots$, $A_K^1(\overline{y}_K^1)$ are received from $A_1(\overline{y}_1)$, $\ldots$, $A_K(\overline{y}_K)$ by means of substitution of subformulas $P_j^1(\overline{x})$ by the correspondent atomic formulas $p_j^1(\overline{x})$.

Such a procedure may be repeated with formulas $A_1^1(\overline{y}_1^1), \ldots, A_K^1(\overline{y}_K^1)$ and so on. At last we can receive $L+1$-level description of classes in the form

$$\left\{\begin{array}{ccc} & A_k^L(\overline{x}^L) & \\ p_1^1(\overline{y}_1^1) & \Leftrightarrow & P_1^1(\overline{y}_1^1) \\ & \vdots & \\ p_{n_1}^1(\overline{y}_{n_1}^1) & \Leftrightarrow & P_{n_1}^1(\overline{y}_{n_1}^1) \\ & \vdots & \\ p_i^l(\overline{y}_i^l) & \Leftrightarrow & P_i^l(\overline{y}_i^l) \\ & \vdots & \\ p_{n_L}^L(\overline{y}_{n_L}^L) & \Leftrightarrow & P_{n_L}^L(\overline{y}_{n_L}^L) \end{array}\right.$$

Below conditions of decreasing of upper bounds of steps of algorithms for described recognition problems will be done.

5. Conditions of decreasing of complexity of recognition problems with global characteristic of

objects within level logical description of classes

Theorem 10. *Let a – the number of occurrences of propositional variables in DNF formulas $A_1, \dots, A_K$, $P_1^1, \dots, P_{n_1}^1$ – subformulas of $A_1, \dots, A_K$. $\nu_1^1, \dots, \nu_{n_1}^1$ – the number of occurrences of propositional variables in $P_1^1, \dots, P_{n_1}^1$, N_j^1 – the number of occurrences of P_j^1 in $A_1, \dots, A_K$. a^1 – the number of occurrences of propositional variables in the number of occurrences of propositional variables in $A_1^1, \dots, A_K^1$.*

Then for fulfilment the equality $a^1 = d \cdot a$ it is necessary

$$\sum_{j=1}^{n_1} (\nu_j^1 - 1) \cdot N_j^1 \geq (1-d) \cdot a.$$

Theorem 11. *Let a – the number of occurrences of propositional variables in DNF formulas $A_1, \dots, A_K$, $P_1^1, \dots, P_{n_1}^1$ – subformulas of $A_1, \dots, A_K$. $\nu_1^1, \dots, \nu_{n_1}^1$ – the number of occurrences of propositional variables in $P_1^1, \dots, P_{n_1}^1$, N_j^1 – the number of occurrences of P_j^1 in $A_1, \dots, A_K$. a^1 – the number of occurrences of propositional variables in the number of occurrences of propositional variables in $A_1^1, \dots, A_K^1$. The number d is defined be equality $a^1 = d \cdot a$.*

Then for decreasing of the upper bound of he number of sequential calculus rule or resolution rule applications while using a 2-level description of classes it is sufficient

$$\sum_{j=1}^{n_1} \nu_j^1 \leq (1-d) \cdot a. \qquad (5)$$

6. Conditions of decreasing of complexity of recognition problems with local characteristic of objects within level logical description of classes

Theorem 12. *Let m – the number of objective variables in formulas $A_1(\overline{y}_1), \dots, A_K(\overline{y}_K)$, $P_1^1(\overline{x}), \dots, P_{n_1}^1(\overline{x})$ – subformulas of formulas $A_1(\overline{y}_1), \dots, A_K(\overline{y}_K)$, $m_1, \dots, m_{n_1}$ – – the number of objective variables in formulas in $P_1^1(\overline{x}), \dots, P_{n_1}^1(\overline{x})$. Predicates p_j^1 are defined by equivalences $p_j^1(\overline{x}_j^1) \leftrightarrow P_j^1(\overline{x}_j^1)$. Formulas $A_1^1, \dots, A_K^1$ are received from $A_1, \dots, A_K$ by means of substitution of $p_1^1, \dots, p_{n_1}^1$ instead of $P_1, \dots, P_{n_1}$.*

Then the checking wether formulas $A_1, \dots, A_K$ are satisfiable on the set $\omega = \{\omega_1, \dots, \omega_t\}$ is equivalent to the checking wether formulas $A_1^1, \dots, A_K^1$ and equivalences $p_j^1(\overline{x}_j^1) \leftrightarrow P_j^1(\overline{x}_j^1)$ are satisfiable on the same set.

For decreasing the number of steps of algorithm deciding the problem of analysis if compound object it is sufficient that

$$\sum_{j=1}^{n_1} t^{m_j^1} + {t_1}^{s_1+n_1} < t^m,$$

where s_1 – the number of variables in $A_1, \dots, A_K$ which do not appear in $P_1, \dots, P_{n_1}$, t_1 – the sum of number of parts of ω for which one of subformulas $P_1, \dots, P_{n_1}$ is valid and the number of elements of ω which are not included into these parts.

References

[1] T.M.Kossovskaya and A.V.Timopheev, "Logic-axiomatical method of recognition and its application", in *Proc. IFAC Symp. Artificial Intelligence*, Leningrad, USSR, October 4-6, 1983, No 9, Pergamon, Oxford, 1984, pp. 523-528.

[2] T.M.Kossovskaya, "Level description of classes for decision in pattern recognition problems", in *Proc. III International Conf. "Discrete models in control system theory"*, Moscow, Dialog-MSU, 1998, pp. 48-50. (in Russian)