Recognition and Analysis of Complex Structured Objects

Tatiana Kosovskaya

kosovtm@gmail.com

St.Petersburg State University

Prelimenary words

While investigation properties of complex structured objects, which are a set of elements that have some specified properties and are in specified relationships, a convenient tool is to describe them using predicate calculus formulas.

If we use the predicate formulas, the relations between the elements are set by atomic predicate formulas in which the order of arguments is strictly defined.

It is even more difficult to write down such a property of a group of elements in which the number of elements for different groups of the same type is different.

If the relation does not depend on the order of the arguments (for example, it is commutative like the relation "to be brothers"), then it becomes necessary to write down atomic formulas with all possible orders of arguments.

A group of elements "family"can include from two up to a sufficiently large number of elements. It is necessary to introduce many predicates with different number of arguments which define the same relation between different number of arguments.

Main definitions

Definition 1. *A complex structured object (CSO) is defined as an object* $\omega = \{\omega_1, \ldots, \omega_n\}$*, whose elements have given properties and are in specified relationships defined by predicates* $p_1, \ldots, p_k$*.*

Definition 2. *A description of the complex structured object* ω *is an elementary conjunction* $S(\omega)$ *of atomic formulas with predicates* $p_1, \ldots, p_k$*, which contains the maximum number of literals and is true for* ω*.*

Let the set of all CSOs Ω be divided into disjoint subsets $\Omega_1, \dots, \Omega_M$.

Definition 3. *A description $S(\Omega_j)$ of the set Ω_j is a disjunction of elementary conjunctions with variables for arguments*

$$S(\Omega_j) = A_j^1(\overline{x}_j^1) \vee \cdots \vee A_j^{k_j}(\overline{x}_j^{k_j}),$$

which is true for every object from Ω_j and only for them.

Definition 4. *Two elementary conjunctions of atomic predicate formulas* $F1(a_1, \dots, a_n)$ *and* $F2(b_1, \dots, b_n)$ *are called isomorphic*

$$F1(a_1, \dots, a_n) \sim F2(b_1, \dots, b_n),$$

if there is such an elementary conjunction $R(x_1, \dots, x_n)$ *and substitutions of arguments* $a_{i_1}, \dots, a_{i_n}$ *and* $b_{j_1}, \dots, b_{j_n}$ *of formulas* $F1(a_1, \dots, a_n)$ *and* $F2(b_1, \dots, b_n)$ *respectively, instead of all occurrences of variables* $x_1, \dots, x_n$ *of the formula* $R(x_1, \dots, x_n)$*, that the results of these substitutions* $R(a_{i_1}, \dots, a_{i_n})$ *and* $R(b_{j_1}, \dots, b_{j_n})$ *coincide up to the order of literals with the formulas* $F1(a_1, \dots, a_n)$ *and* $F2(b_1, \dots, b_n)$*, respectively.*
The resulting substitutions $\lambda 1 = \{x_1 : a_{i_1}, \dots,\ x_n : a_{i_n}\}$ *and* $\lambda 2 = \{x_1 : b_{j_1}, \dots,\ x_n : b_{j_n}\}$ *are called unifiers of formulas* $F1(a_1, \dots, a_n)$ *and* $F2(b_1, \dots, b_n)$ *with the formula* $R(x_1, \dots, x_n)$ *respectively.*

Definition 5. *An elementary conjunction that does not contain constants is called a common subformula of two elementary conjunctions if it is isomorphic to some subformulas of each the initial ones.*

Definition 6. *An elementary conjunction that does not contain constants is called a maximal common subformula if it is their common subformula with the largest number of literals.*

If the description of the class of objects has the form

$$S(\Omega_j) = A_j^1(\overline{x}_j^1) \vee \cdots \vee A_j^{k_j}(\overline{x}_j^{k_j}),$$

then the recognition of a CSO with the description $S(\omega)$ consists in the checking the logical sequence

$$S(\omega) \Rightarrow \exists \overline{x}_{\neq} A(\overline{x}), \qquad 1$$

where $A(\overline{x})$ is a disjunctive term of the classes descriptions.

The problem (1) is an NP-complete problem.

Computational complexity of such a checking is exponential under the length of the right-hand part.

Multi-level description of classes

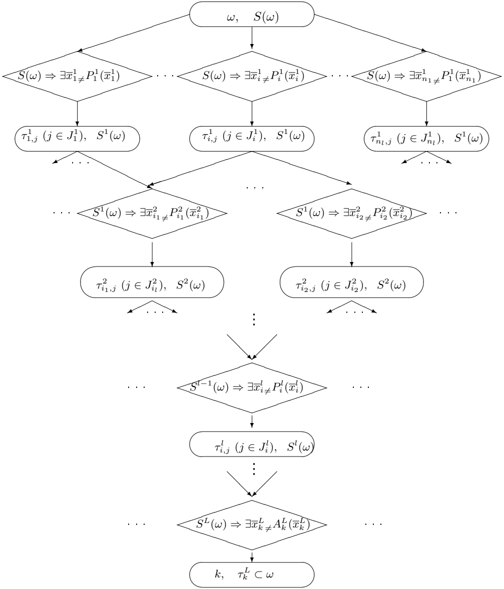
Construction of a multi-level description of classes consists in pairwise extraction of maximal common subformulas from elementary conjunctions $A_j^l(\overline{x}_j^l)$ and $A_j^r(\overline{x}_j^r)$.

This procedure is repeated with the obtained elementary conjunctions until the newly received ones do not have common subformulas. The process of recognition starts from checking of (1) for the shortest of the received.

Computational complexity remains exponential but with essentially lower exponent.

Level description of classes

$$\begin{cases} & A^L_{k,j}(\overline{x}^L_{k,j}) & (j = 1, \ldots, m_k) \\ p^1_1(y^1_1) & \Leftrightarrow & P^1_1(\overline{y}^1_1) \\ & \vdots & \\ p^1_{n_1}(y^1_{n_1}) & \Leftrightarrow & P^1_{n_1}(\overline{y}^1_{n_1}) \\ & \vdots & \\ p^l_i(y^l_i) & \Leftrightarrow & P^l_i(\overline{y}^l_i) \\ & \vdots & \\ p^L_{n_L}(y^L_{n_L}) & \Leftrightarrow & P^L_{n_L}(\overline{y}^L_{n_L}) \end{cases} .$$


ω, S(ω)
S(ω) ⇒ ∃x̄¹₁≠ P¹₁(x̄¹₁)
S(ω) ⇒ ∃x̄¹ᵢ≠ P¹ᵢ(x̄¹ᵢ)
S(ω) ⇒ ∃x̄¹ₙ₁≠ P¹₁(x̄¹ₙ₁)
τ¹₁,ⱼ (j ∈ J¹₁), S¹(ω)
τ¹ᵢ,ⱼ (j ∈ J¹ᵢ), S¹(ω)
τ¹ₙₗ,ⱼ (j ∈ J¹ₙₗ), S¹(ω)
S¹(ω) ⇒ ∃x̄²ᵢ₁≠ P²ᵢ₁(x̄²ᵢ₁)
S¹(ω) ⇒ ∃x̄²ᵢ₂≠ P²ᵢ₂(x̄²ᵢ₂)
τ²ᵢ₁,ⱼ (j ∈ J²ᵢₗ), S²(ω)
τ²ᵢ₂,ⱼ (j ∈ J²ᵢ₂), S²(ω)
Sˡ⁻¹(ω) ⇒ ∃x̄ˡᵢ≠ Pˡᵢ(x̄ˡᵢ)
τˡᵢ,ⱼ (j ∈ Jˡᵢ), Sˡ(ω)
Sᴸ(ω) ⇒ ∃x̄ᴸₖ≠ Aᴸₖ(x̄ᴸₖ)
k, τᴸₖ ⊂ ω

Disadvantages of using predicate calculus formulas

The use of predicate calculus formulas has such disadvantages that

– in cases that the relations are commutative (for example, «to be friends»), you have to explicitly specify both $p(x, y)$ and $p(y, x)$, for n-ary relation it is needed to write down $n!$ atomic formulas;

– if the relation can be multi-place without specifying the number of arguments (for example, «be members of the same family»), then you need to enter several predicates like $p^j(x_1, \ldots, x_j)$ for each number of arguments. Question: what is the maximum number of such arguments?

The possibility of using hypergraphs to eliminate these disadvantages is suggested below.

Main definitions related to the use of special type hypergraphs

Definition 7. *A set $\omega = \{\omega_1, \ldots, \omega_n\}$, on which the properties of these objects and the relationships between them are dtfined, is called a complex object (CO).*

For every element of a complex object ω it is known what properties are fulfilled and in what relations these elements are.

Relations may be commutative and without fixed arity.

Definition 8. *A hypergraph of relations is such a hypergraph that its vertices are divided into two parts:*
– one part (predicate vertices) contains the names of the properties of objects and the relations between them, the in-degree of these nodes are 0;
– the other one (objective vertices) contains the names of the elements of an object.

The edges are defined as follows:
– for a property p the edge has the form

$$p \to \omega_i \ ,$$

– for a noncommutative k-ary relation

$$p \to (\omega_{i_1} \to \cdots \to \omega_{i_k}) \ ,$$

– for a commutative relation of arbitrary dimension

$$p \to \{\omega_{i_1}, \cdots, \omega_{i_k}\} \ ,$$

– for a noncommutative k-ary relation, which can have every element of a set specific to a position

$$p \to \{\{X_1\}, \cdots, \{X_k\}\} \ .$$

Definition 9. *Two hypergraphs of relations G_1 and G_2 are called isomorphic if there exists such a hypergraph of relations H with variables as vertex names and such substitutions λ_1 and λ_2 of the names of the object elements in G_1 and G_2 instead of the variables of the graph H, the results of applying these substitutions are graphs G_1 and G_2.*
Such substitutions will be called the unifiers of the hypergraphs G_1 and G_2 with the hypergraph H.

Definition 10. *A description of CO ω is a hypergraph of relations, the names of properties and relations defined on ω are in one part of which, and the names of all objects are in the other part.*

Let the set of all objects Ω be divided into disjoint subsets $\Omega_1, \dots, \Omega_M$.

Definition 11. *The description $S(\Omega_j)$ of the set Ω_j is such a collection of hypergraphs of relations with variables for the names of the vertices in its second part, that for each object of the class this collection contains a hypergraph isomorphic to the subgraph defining the description of the object.*

Example of descriptions of an object and a class of objects

Let an object be a family of Johnson. It contains 5 members: father John, mother Mary, two sons Peter and Basil, daughter Katherine.
Two properties and one relation are defined as
$m(x)$ – "x is a man
$w(x)$ – "x is a woman
$p(\{X\},\{Y\})$ – "the element of the set $\{X\}$ is the parent of the element of the set {Y}".

Description of the Johnson's family can be represented by a hypergraph, in one part of which there are vertices m, w, p, and in the other part vertices with the names of family members J, M, P, B, K.
The edges have the form $m \to J$, $m \to P$, $m \to B$,
$w \to M$, $w \to K$,
$p \to (\{J, M\}, \{P, B, K\})$.
The last edge means that every element of the set $\{J, M\}$ is the parent of every element of the set $\{P, B, K\}$.

Description of the object class "families with brothers"corresponds to the fact that the formula

$$\exists x x_1 x_2 (p(x, \{x_1, x_2\}) \& m(x_1) \& m(x_2)$$

follows from the description of a particular family.

This formula (without quantifiers of existence) corresponds to a hypergraph with variables x, x_1, x_2 as vertex names and with edges

$$m \to x_1, \quad m \to x_2, \quad p \to (x, \{x_1, x_2\}).$$

It is obvious that it is a subgraph of a hypergraph defining a description of the Johnson family with the following unifiers (values of variables): $x = J$, $x_1 = P$, $x_2 = B$.

About further investigations

Our conference is devoted to mathematics which is assistened with computer.
This presentation is about what mathematical instruments may help us to solve some problems connected with Artificial Intelligence.
The main problem arising with this approach is to develop algorithms for checking hypergraphs of relations for isomorphism and for extracting maximal common (up to the names of arguments) sub-hypergraph of relations for two ones.

THANK YOU for attention

Tatiana Kosovskaya
St.Petersburg State University
kosovtm@gmail.com