Interpretation Method in Automated Reasoning

Aleksandar Perović, Nedeljko Stefanović, Dejan Ilić, Aleksandar Jovanović

GIS (Group For Intelligent Systems), Faculty of Mathematics, Belgrade

Abstract: Generally in mathematics and its applications, some problems are becoming much easier if we can express them in, should we say, suitable framework. For instance, S^3 can be treated as model of Euclidian space, so we could use apparatus of analysis, algebra and model theory for solving particular problem in Euclidian geometry. On the other hand, some analytical, or algebraical problems became much easier if we can find their geometrical interpretation. Logical background of this method can be found in two different approaches: interpretation method, i.e. formal representation of one first order theory in another (in general case one formal system in another), and category theory, where we exploit equivalence and existence of adjoint functors between certain categories. Our particular interest is application of these methods in automated reasoning.

1 Introduction

Various problems in mathematics and its application can be described and treated within different mathematical theories and concepts. Some of them are made to solve particular problems, which for centuries endured and remain unsolved. Marvellous example of this kind is Galois theory, arguably one of the uttermost achievements of men kind at all.

Formal background of interpretation method is due to Gödel, thou method itself was informally used earlier (for instance to establish equiconsistence between the Euclidian and hyperbolic geometry). This method was introduced in his monumental paper on formal undecidability of *Principia Mathematica* and related

systems. Gödel succeeded to recursively[1] interpret logical notions such as “to be a formula”, “to be a sentence”, “to be a proof”, “to be consistent” etc. within formal arithmetic[2] PA (Peano arithmetic), and to prove that if PA is consistent, then this consistency cannot be shown in PA. Strictly speaking, Gödel constructed a sentence Con(PA) which asserts that theory PA is consistent and then showed the following:

- If PA is consistent, then Con(PA) is not provable from PA;

- If PA is ω-consistent, then ¬Con(PA) is not provable from PA;

- PA is essentially incomplete, i.e. there is no recursive, consistent and complete extension of PA.

Namely, interpretation method is formal representation of one first order theory into another. Thou much of interesting theories (at least for mathematicians) are undecidable, there are decidable fragments which can be used for automated theorem proving/ automated reasoning. Variety of algorithms and methods are proved to be useful here (such as syntax analysis and processing, tableau, resolution, quantifier elimination etc.).

With the respect to the category theory, our interest also lies in the proof theory and its applications. Namely, we can refer to formulae (or sets of formulae) as objects and to inference proofs as morphisms between them. In this way we can construct a category related to some formal system, and then try to establish categorical equivalence (if not equivalence, then adjunction) between some other categories, such as Cartesian closed categories, λ-calculi categories, certain types of topological categories etc. For instance, some problems in graph theory can be solved by shifting to another category, such as λ-calculi, some topological categories etc.

---

[1] Gödel's *recursive functions* are the first formal system of computability , which later inspired Alan Turing to develop the *Turing machines* and Alonso Church to develop the λ- calculus.

[2] Formal arithmetic PA is a first order theory in language $L_{PA}=\{+,\cdot,',0,1\}$ with following axioms:

- #x(x'≠0)
- #x#y(x'=y'Â x=y)
- #x(x+0=x)
- #x#y(x+y'=(x+y)')
- #x(x· 0=0)
- #x#y (x· y'=(x· y)+x)
- Let k(x,…) be an arbitrary formula in language L_{PA}. Then universal closure of the sentence

$$k(0,\ldots) \wedge \#x(k(x,\ldots)\hat{A}\ k(x',\ldots))\hat{A}\ \#xk(x,\ldots))$$

is an axiom of PA.

Some of research projects in the *Group for Intelligent Systems* are aimed on the quantifier elimination method and its applications. More details of this work are presented in [2], [4], [8] and [9]. Obviously we were curious if we could find a way to expand developed procedures in the broader class of theories and applications. One way to do this is to use interpretation method. That could also help to find more applications of decidable fragments of some undecidable theories.

2 Definability

Let L be a first order language (i.e. L may contain some constant, function and relation symbols, but L can also be an empty set) and let T be a theory of language L (i.e. set of some sentences, where sentence is a first order formula which every variable is bounded with some quantifier).

A property is definable in theory T if it can be uniquely expressed (in T) by a first order formula of language L. Now we will give a strict definition for definability of constant, function and relation symbols:

- Each formula $k(x)$ of language L such that

$$T \vdash \exists_1 x k(x)$$

can be used for definition of new constant symbol c in the following way:

$$x=c \Leftrightarrow_{def} k(x).$$

- Each formula $k(x_1, ..., x_n, y)$ of language L such that

$$T \vdash \#x_1, \dots, x_n \exists_1 y\ k(x_1, \dots, x_n, y)$$

can be used for definition of new n-ary function symbol F in the following way:

$$y=F(x_1,\dots, x_n) \Leftrightarrow_{def} k(x_1,\dots, x_n, y).$$

- Each formula $k(x_1,\dots,x_n)$ of language L can be used for definition of new n-ary relation symbol R in the following way:

$$R(x_1,\dots,x_n) \Leftrightarrow_{def} k(x_1,\dots, x_n).$$

Example 1. Let us show that each natural number n is definable in PA. To see that, let $k_n(x)$ be a formula

$$x=\underline{n},$$

where $\underline{1}$ =0', $\underline{2}$ =0'' etc. Now for every pair of natural numbers m and n we have that

$$PA \vdash \underline{m}=\underline{n} \quad \textit{if and only if} \quad m=n,$$

so each formula $k_n(x)$ has the unique witness $\underline{n}$. Thus we can (in PA) define each natural number n with $\underline{n}$. □

If T is a theory of fields of characteristic zero, then it is easy to see that each rational number is definable in T. Now let us state another example.

Example 2. Let L={<} be a language of linear ordering and let T be a set of all sentences of language L which are true in the structure (O, <) where

$$\mathbf{N} = \{1,2,3,\ldots\},$$

is a set of all natural numbers and < is usual ordering of O. Then, every natural number is definable in T. To see that, for each natural number n let $k_n(x)$ be a formula

$$\exists x_1,\ldots,x_n(\ddot{a}_{i\neq j}x_i\neq x_j \wedge x_1<x \wedge \ldots \wedge x_n<x \wedge \#y(y<x \hat{A}\ y=x_1 \vee \ldots \vee y=x_n)).$$

Note that formula $k_n(x)$ asserts that x has exactly n predecessors. Thus,

$$T \vdash \exists_1 x\ k_n(x),$$

since the only witness in (O,<) for $k_n(x)$ is the natural number n+1. To verify that 1 is definable, note that 1 is the only witness for the formula

$$\#y(x \leq y)$$

in the structure (O,<). □

As an interesting contrast to the previous example we state the theory T* which contains all sentences of the language L of linear ordering which are true in the structure ([,<), where

$$[= \{0,1,-1,2,-2,\ldots\},$$

is the set of all integers and < is usual ordering of integers. It can be shown that none of the integers is definable in T*. To gain definability, it is sufficient to add one constant symbol to the language L.

3 Interpretation

Let L and L* be first order languages, T be a theory of language L and T* be a

theory in language L*. We say that language L is *interpretable* in a theory T* if following conditions hold:

- There is an unary predicate U definable in T* such that

$$T \vdash \exists x\, U(x).$$

- For each constant symbol c of language L there is a constant symbol c* definable in T* such that

$$T^* \vdash U(c^*).$$

- For each n-ary function symbol F of the language L there is an n-ary function symbol F* definable in T* such that

$$T^* \vdash \#x_1,\ldots,x_n(U(x_1) \wedge \ldots \wedge U(x_n) \hat{A}\ U(F(x_1,\ldots,x_n)).$$

- For each n-ary relation symbol R of language L there is an n-ary relation symbol R* definable in T*.

In order to obtain the fundamental interpretation theorem, the following is needed:

for a given formula κ in a language L we are going to define a formula κ^* in a language L* by induction on the complexity of formula κ as follows:

- $(t_1=t_2)^*$ is the formula $t_1^*=t_2^*$, where t_i are terms of language L;

- $(R(x_1,\ldots,x_n))^*$ is the formula $R^*(x_1,\ldots,x_n)$, where R is n-ary relation symbol of language L;

- $(\neg\kappa)^*$ is the formula $\neg\kappa^*$;

- $(\kappa \wedge \psi)^*$ is the formula $\kappa^* \wedge \psi^*$;

- $(\exists x\kappa(x,\ldots))^*$ is the formula $\exists x(U(x) \wedge \kappa^*(x,\ldots))$.

We say that theory T is *interpretable* in T* if there is an interpretation of the language L in theory T* such that for every nonlogical axiom κ of T we have $T^* \vdash \kappa^*$.

Theorem. Let L and L* be a first order languages and let T and T* be a theories of languages L and L* respectively, such that T is interpretable in T*. Then, for each formula κ of language L, if $T \vdash \kappa$ then $T^* \vdash \kappa^*$. □

With next example we will illustrate the use of quantifier elimination in the interpretation method.

Example 3. Let T be a recursive theory. Suitable theory T* should satisfy the following conditions:

- There is a recursive interpretation of axioms of T into T* and there is a recursive interpretation of atomic formulae of T* into T;

- T* admits the quantifier elimination and we have an effective procedure for it;

- There is a recursive procedure for validity of quantifier free formulas in the theory T*.

So, we have the following algorithm:

input: formula κ of the language L

output: YES, if $T \vdash \kappa$; NO, otherwise

step 1: Find an interpretation κ^* in language L* of a given formula κ.

step 2: Find a quantifier free formula ψ of the Language L* such that

$T^* \vdash \kappa^* \leftrightarrow \psi$.

step 3: Check the validity for ψ. If ψ is valid in T*, then the output is YES, otherwise it is NO.

Of course, the interpretation theorem guaranties the correctness of the algorithm stated above.

Example 4. Here we are briefly going to discuss an interpretation of monadic calculus in ZFC theory (ZFC states for Zermelo-Frankel set theory together with the axiom of choice). The embedding of the monadic calculus in the set theory is quite natural: for instance, silogism Bocardo

Some M are not P

Every M rm is S

Some S are not P

can be expressed in ZFC as

$$M \setminus P \neq 0 \wedge M \acute{Y} S \hat{A} \; S \setminus P \neq 0,$$

where 0 is abbreviation for empty set.

To be more precise, we will define a *monadic formula* in ZFC by induction on complexity as follows:

- Atomic formula is monadic formula;

- Boolean combination κ of monadic formulae is monadic formula if for no variables x, y and z, $x \in y$ and $y \in z$ are subformulas of κ;

- If $\kappa(x,\ldots)$ is monadic formula and there is no variable y such that $y \in x$ is a

subformula of κ, then $\exists x \kappa(x,\ldots)$ and $\#x\kappa(x,\ldots)$ are also monadic formulae.

It is obvious that the set of all monadic formulae in ZFC is recursive. If we combine this with effective procedure of quantifier elimination for the monadic calculus developed in GIS (see [2]), we obtain a theorem prover for monadic formulae in ZFC. □

Example 5. Here we are going to put some light on the fact that theory ZFC cannot be interpreted in formal arithmetics PA (assuming that both of this theories are consistent). If we define ordinals and ordinal arithmetics in the usual way (see [6]), then we have that

$$ZFC \vdash (\omega, +, \tilde{a}, S, 0, \{0\})^{1} \models PA.$$

where ω is the least infinite ordinal, $S(x)=x \cup \{x\}$ and + and ãare respectively ordinal addition and multiplication. Thus

$$ZFC \vdash Con(PA).$$

On the other hand, by the first Gödel incompleteness theorem we have that

Con(PA) is not provable from PA, so by interpretation theorem we can conclude that ZFC cannot be interpreted in PA. □

With next two examples we are going to illustrate the general interpretation method.

Example 6. Let PR be a first order predicate calculus and let PC be a propositional calculus. We will give the sketch of a proof for consistency of PR.

Namely, we are going to specify an interpretation of PR in PC in the following way:

- Atomic formulae are interpreted as propositional letters;

- Interpretation of Boolean combination of formulae is the same Boolean combination of interpretations of formulae;

- Interpretation of the formula $\#xk$ is interpretation of k.

In this way we can also interpret PR-proofs in PC-proofs. Now any proof of contradiction in PR we can translate to the proof of contradiction in PC. Since propositional calculus is consistent, we can conclude that the first order predicate calculus is also consistent. □

Example 7. Decidability of monadic calculus and modal S5 calculus were proved very early (Skolem, Tua). Decidability for theory of Boolean algebras was proved by Tarski in 40's. However interpretation based equivalences of these theories were studied in detail by Žarko Mijajlović in his MS thesis in 1973. Mijajlović introduced effective interpretations (in linear time) between modal S5 calculus, monadic calculus without equality and universal sentences in theory of Boolean algebras. These theories were implemented in GIS several times in alternative design.

Here we are going to briefly describe the interpretation of monadic calculus without equality in modal S5 calculus. Modal S5 calculus is propositional calculus with two additional unary operators: L (L(p) we interpret as "p is necessary") and

M (M(p) we interpret as “p is possible”). Beside propositional axioms, we have three additional axioms:

- L(p) Â p (axiom of necessity);

- L(p Â q) Â (L(p) Â L(q));

- M(p) Â L(M(p)).

Rules of detachment are modus ponens k, $\frac{\varphi, \varphi \to \psi}{\psi}$ and rule of necessity $\frac{\varphi}{L(\varphi)}$.

The following fact allows us to construct desirable interpretation: *Each monadic formula* k *with monadic predicates* $P_1,\ldots, P_n$ *is equivalent (in monadic calculus) to Boolean combination* z($r_1,\ldots, r_n$)*, where formula* r_i *is some of predicates* P_j *or* r_i *has a form*

$$\%x\ (Q_1(x)\ \wedge \ldots \wedge Q_n(x)),$$

where $Q_i \in \{P_1,\ldots, P_n\}$. *In particular, reduced monadic formulas are formulas of type* z($r_1,\ldots, r_n$).

Now we will define the interpretation of reduced formula as follows:

- Interpretation of P(x) is propositional letter p;

- Interpretation of %x($P_1(x)\ \wedge \ldots \wedge\ P_n(x)$) is M($p_1 \wedge\ \ldots \wedge\ p_n$);

- Interpretation of Boolean combination is the same Boolean combination of adequate interpretations.

It can be shown that reduced monadic formula k is theorem in monadic calculus if and only if its interpretation is theorem in modal S5 calculus.

4 Categories

For definition of the notion of category and related features we refer the reader to [7]. Let us state some basic examples of categories.

- Category of sets: objects are sets, morphisms are functions and composition of morphisms is composition of functions;

- Category of topological spaces: objects are topological spaces, morphisms are continuous functions and composition of morphisms is composition of functions;

- Category of Abelian groups: objects are Abelian groups, morphisms are group homomorphisms and composition of morphisms is composition of functions;

- Category induced with monoid (M,◦): the only object is ◦ morphisms are elements of M and composition of morphisms is ◦

- Category induced with certain formal theory: objects are formulae, morphisms are inference proofs and composition of morphisms is concatenation of proofs.

Note that some categories could be a proper classes. However, without loss of generality we could restrain our argumentation to fragments which are sets. So, we will assume that all categories are small (i.e. they are not proper classes).

There are various applications of category theory. The essence of the method lies in transition from one category in another via adequate functor. Usually we need category equivalence, or in more general case, adjunction between categories. With following example we want to illustrate application of the category theory in the proof theory.

Example 8. As before, let ω be the least infinite ordinal (for strict definition see [6]). We will define a category O as follows:

- Objects are all finite ordinals (i.e. $O_O = \omega$);

- Morphisms are relations between finite ordinals. Thus, for arbitrary finite ordinals m and n the set of all morphisms from m to n is $P(m \times n)$;

- Composition of morphisms is composition of relations.

Now let X be an infinite set (for our purpose X will be a set of all propositional letters) such that $T \notin X$ (¾ stands for logical constant "true"). To simplify things, we will consider only conjunctive fragment of the intuitionistic calculus (with ¾ T). So, set of formulae is the minimal superset of $X \cup \{T\}$ which is closed under the $\wedge$ (i.e. under the conjunction). We have the following inference rules:

- t_A: A ⊢ ¾ (tautology)

- 1_A: A ⊢ A (identity)

- p_1: $A \wedge B$ ⊢ A (first projection)

- p_2: $A \wedge B$ ⊢ B (second projection)

- If $f : A \to C$ (i.e. f is an inference proof from formula A to formula C) and $g : B \to D$ then (p_1,p_2): $A \wedge B \vdash C \wedge D$.

As additional rules we have the following equalities among inference proofs:

- h(gf)=(hg)f (here gf is concatenation of proofs f and g);

- Let $f : A \to B$. Then $f1_A$=f and $1_B f$=f;

- $p_1(f,g)=f$, $p_2(f,g)=g$;

- $(p_1h, p_2h)=h$;

- $(f,g)h=(fh, gh)$;

- Let $f : A \rightarrow T$. Then $f = t_A$.

In this way we actually defined the category $L_{\wedge T}$ ¾(X). It can be shown that the functor F: $L_{\wedge T}$ ¾(X) → NO defined with

- $F(x) = F(T) = 1$, $x \in X$

- $F(A \wedge B)=|F(A) \times F(B)|$

- Let p_1: $A \wedge B$ ¼→ A. Then $F(p_1)$:F(A)©F(B)· F(A) is defined with $F(p_1)(x,y)=x$;

- Let p_2: $A \wedge B \rightarrow$ ¼B. Then $F(p_1)$:F(A)©F(B)· F(B) is defined with $F(p_1)(x,y)=y$;

- $F(1_A)=id_{F(A)}$, $F(t_A)=id_1$

- F((f,g))=(F(f), F(g)), where (F(f),F(g))(x,y)=(F(f)(x),F(g)(y))

has the following property: morphisms f and g in the category L.¾(X) are equal if and only if corresponding morphisms F(f) and F(g) are equal in the category O. Note that question of equality between morphisms in O is quite easy: we have to check equality of two subsets of m x n. □

References

[1]C. C. Chang, H.J. Keisler: *Model theory*, third edition, North-Holland 1990

[2]F. Marić, M. Marić, Ž. Mijajlović, A. Jovanović: *Theorem provers based on the quantifier elimination method*, Proc. XLVII ETRAN conference, Herceg-Novi, June 8-13, 2003

[3]Ž. Mijajlović: *Decidable theories* (in Serbian), Computer science num. 1, Belgrade 1991

[4]Ž. Mijajlović, A. Jovanović: *Provers in algebra based on the quantifier elimination*, Proc. of algebraic conference, Niš 1996

[5]J. Shoenfield: *Mathematical logic*, Addison-Wesley 1967

[6]T. Jech: *Set theory*, second edition, Springer-Verlag 1997[7]J. Lambek, P. J.

Scott: *Introduction to higher order categorical logic*, Cambridge University Press 1986

[8]A. Jovanović, Ž. Mijajlović: *Logic and databases*, Proc. of the Sixth International Symposium on Computer Science at the University of Dubrovnik, 1984

[9]A. Jovanović: *Group for Intelligent Systems - Problems and Results*, Intelektualnie sistemy, Lomonossov Un and RAN, 6, 2002