

Artificial Intelligence in Knowledge-based Technologies and Systems

Viktor Krasnoproshin^{1*}, Vladimir Obratsov¹, Vladimir Rjazanov², Herman Vissia³

¹Department of Information Management Systems, Belarusian State University, Republic of Belarus

²Dorodnicyn Computing Centre of the Russian Academy of Sciences, Russia

³Byelex B.V., the Netherlands

Copyright©2016 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract A modification of the paradigm of Artificial Intelligence (AI) is proposed in the paper. The modification is based on the assumption that there are algorithms which are inductive by construction, but can be mathematically proved. The content of traditional artificial intelligence concepts (knowledge, form of presentation, knowledge base, etc.) is determined within the proposed paradigm. The modification ensures unification of many concepts in the field of artificial intelligence.

Keywords Artificial Intelligence, Model Modification, Inductive Inference, Justification Problem, Technology for Knowledge Systems Processing

1. Introduction

Currently, the prevailing trend in the development of mankind is a steady increase in the volume of information. People have learned to cope with the information and have a lot of tools and technology to work with it. Thus, the particular significance is not only data acquisition, storage and transfer but the usage as well. Perhaps, the first question is knowledge acquisition from the ever increasing body of data. Undoubtedly, useless data do not exist. But if all the information, being accumulated in various information systems, is stored, then the resources may simply be exhausted very soon.

Knowledge extraction from data is the most difficult problem in artificial intelligence. However, there are all the necessary prerequisites for its solution.

What are the prerequisites under consideration? First of all the ones that relate to the concept of artificial intelligence, corresponding computer technologies and their applicability for solving applied problems. Decision-making support systems, executive information systems as well as systems for storing and obtaining data are left somewhere aside. But actually the above-enumerated systems are parts of a wide

range of AI systems. We can state that they are particular cases of AI systems.

Let's consider the state of the art in AI. The key notion of AI is knowledge which is an object, a means and an aim of modelling. However, there are still questions concerning the notion of knowledge, its processing and the evaluation of the results.

What are the reasons? In our opinion, the following ones are of fundamental importance:

1) knowledge is not specified on a formal level (unlike information and data). Knowledge should become an object of mathematical formalization.

2) the level of computer technologies. It is well known that any technology allows building user-friendly systems under certain conditions. Such conditions can be: solution of problems of information representation and handling, the necessary level of information independence (separation) from algorithms, etc. Such conditions should exist for knowledge.

From the above, we can make the following conclusion. All the necessary notions exist, but they are either inaccurately defined or their systematization is not quite correct. Without modification of AI paradigm it is hardly probable to expect fundamental changes in solving problems of knowledge acquisition, handling, etc.

2. Brief Description of the Existing Paradigm

Let's consider the existing system of notions in AI. We can single out three levels of their construction: *theoretical*, *technological* and *applied*.

Theoretical level. The starting point in all theoretical constructions is a definition of knowledge. By now the following definitions may be thought of as the best known and logically complete:

- definitions based on informal aspects of the notion, knowledge is interpreted data [1];

- definitions based on ways of representing the source information. The most widely used are frames, semantic networks and “axiomatic” means (with the use of languages of propositional calculus, predicate calculus, modal logics, etc.) [2,3]. The main postulate may be formulated in the following manner: knowledge is all that can be obtained on the basis of the above-mentioned representations with the help of relevant mechanisms of inference.

Next is the notion of knowledge representation. The choice of representation mechanisms logically follows from the choice of the concept of knowledge. For the first conception, by convention, use is made of representations in the form of formal grammars [2]. Means peculiar to the second conception are practically all enumerated above.

Furthermore, the notion of algorithms (mechanisms) of knowledge manipulation is defined. The purpose of manipulation is to solve one of the following interconnected problems: obtaining new objects (knowledge) or inclusion of certain objects (information of which is not given beforehand) into a knowledge system existing a priori. The particular realization of algorithms is determined by the choice of previous notions, the purpose and specialization of the language used for formalization. As for the content, all algorithms realize a certain inference. As for the form, they can be divided into two types: deductive and inductive ones. If it is considered that deduction is not only a property of an algorithm but a characteristic of objects (information) as well, then almost all algorithms may be called inductive ones. Hence it is natural to classify algorithms according to the language of formalization. From this standpoint we can determine the following types of algorithms: resolution (most often parametrized) [2,3,4], inductive inference [5], fuzzy inference (including probability algorithms [2,5]), etc.

Technological level. The notions of this level (and of the applied one as well) are not quite settled. Hence, we will give some explanations first. The technological level comprises a set of methods and software that are intended for the support of design, realization and functioning of systems based on knowledge. These methods and software require that the objects being processed satisfy certain specifications and the logic of their processing be standardized. If we now disengage ourselves from specifications and standards, then the basic notions of this level will be [2,3]:

- knowledge base;
- interpretation of objects;
- interface;
- knowledge extraction.

The notions almost completely determine the meaning and content of the technology for working with knowledge. It can be enlarged taking into account the specialization of means. The means can be universal (all kinds of programming systems for AI), or specialized ones (expert systems, decision-making support systems, etc.). The following notions can be singled out:

- language for knowledge description;

- representation scheme;
- allowable operations on knowledge;
- explanation, etc.

We leave the notions without additional comments, because their meanings greatly depend on the content of notions selected on the theoretical level. Besides, now it is more important to determine how these notions form a system.

Applied level. This level is formalized in the least extent. Nevertheless, the following notions can be mentioned:

- problem;
- qualitative and quantitative estimates characterizing solvability of a particular problem and the result of its solution.

Why are these problems assigned to the applied level? It is obvious that certain distinctions always exist between an applied problem and its formal model. A theoretical construction provides only the necessary prerequisites allowing to estimate the quality of problem solution. In reality such an estimate depends on many conditions and can be obtained from an experiment. Although the way of conducting experiments and obtaining estimates is determined on the theoretical level.

Now let's determine the essence of the existing AI paradigm. We will describe a way of constructing a system on the basis of the notions introduced above as theses. The theses in a greater extent will have an applied nature. Any theory has sense if at least one problem can be solved with its help. Hence theoretical, technological or any other systematization is a direct consequence of an applied one. A system should describe interaction and interconditionality of notions.

Statement 1. The prime notion is a *problem* that is formulated on the informal level. For each problem (in case it relates to AI) it is possible to choose the corresponding *knowledge representation model*. A particular way of representation depends on the aims of problem solution, the necessity to estimate the result and other conditions.

Statement 2. A certain set of manipulation algorithms is associated with each knowledge representation model. The choice of a particular algorithm is stipulated by a number of restrictions. The major restriction is that an algorithm should not be reduced to a simple exhaustion under all allowable conditions imposed on the problem.

Statement 3. The choice made at the previous stages stipulates a possibility of choice on the technological level. Realization and particular content of notions of this level greatly depend on the choice of programming environment or system.

Statement 4. It is necessary to carry out an estimate of quality of problem solution. If the estimate is unsatisfactory, a correction is possible.

It is obvious that notions themselves do not depend on applied problems. But their ability to form systems is completely determined by a possibility to solve all problems that can be related to the field of AI. In this sense any

paradigm is a closed one (notions are more than enough), but it is incomplete. This results in drawbacks of the existing system. Let's enumerate the major ones:

1) impossibility to say a priori something about the choice of the way of knowledge representation for the problem being solved. This choice is subjective and greatly depends on the person who solves the problem;

2) the necessity to use (in the majority of problems) algorithms that are not justified. From the finite body of the defined knowledge justified inference on new objects can be obtained only in very special cases. The cases, when the body of the defined knowledge is infinite (may be built potentially), are very special in the sense of information. But then there is a possibility to use justified algorithms;

3) lack of a technology as, for example, used in databases. Hence design and construction of a particular system is basically intuitive rather than technological.

The list can be easily enlarged. But even the enumerated drawbacks are sufficient to say that the existing paradigm calls for a modification.

3. Brief Description of a Possible Modification

Let's consider the essence of the proposed modification. To do this, we will discuss knowledge manipulation algorithms in more detail. There is a close interdependence between the notion of an algorithm and the notion of knowledge. But algorithms have a property that depends not only on constituents of the notion of knowledge, but on qualitative and quantitative characteristics of the problem as well. This property is called justification.

Notes. Let's explain what is meant by justification [2]. To do this, we consider one of "axiomatic" systems, that is based on the language of propositional calculus. Its essence is as follows: a language L is chosen with the help of which the system is built. The system consists of axioms A and inference rules R . For such a system, and only for it, it is shown that the system is in a unique fashion (within the accuracy of language L) characterized by the following properties:

- completeness (each object, written with the help of language L , should belong to a truth class, i.e. identically true, identically false or neutral);
- consistency (not a single object and its logic negation can simultaneously belong to the same truth class);
- closure (each object, written with the help of language L after applying R , is always entered in a truth class).

The result of the above-mentioned properties is the existence of a universal algorithm that guarantees entering of any object above language L in a truth class (may be by a countable number of steps). This algorithm is called resolution (let's denote it by R_0 for brevity). Thus, the main point of justification of resolution R_0 in relation to propositional calculus lies in the fact that the result is

guaranteed.

Now let's assume that there is an algorithm R_1 that has the following properties:

- it is inductive by construction (i.e. it works with a finite body of input information but allows to construct inference in relation to an infinite set);
- it is justified in a sense mentioned above;
- it works with universal knowledge representation. In this case it is assumed that a universal representation can be obtained from any existing one without loss of information.

It is easy to see that in this case there is a possibility to avoid at least two drawbacks mentioned above (below we will show how it is possible to get rid of the remaining one). The point is that justification is always associated with properties of algorithm R_1 , i.e. completeness, consistency and closure. These properties in turn are a firm basis for further realization.

Now we may present the modified paradigm. Let's do it again with the help of statements.

Statement 1'. The prime notion is a problem that is formulated on the informal level. For each problem (in case it relates to AI) a way of transition to a universal knowledge representation model is chosen.

Statement 2'. The choice of an algorithm is absent. For *knowledge manipulation* use is made of algorithm R_1 .

Statement 3'. The interrelation of notions of the technological level is provided with the help of a system that can be called knowledge-based management system (KBMS). The system contains means allowing to carry out design, realization and functioning of knowledge.

Statement 4'. It is necessary to carry out an estimate of quality of problem solution. If the estimate is unsatisfactory, a correction is possible. In this case only correction of knowledge (information) makes sense.

Let's briefly describe the main notions necessary to formulate the modified paradigm. All missing details one can find in [6,7,8].

Let's start with the *justification scheme of algorithm R_1* . For this purpose we will introduce several classes of problems. Let's denote by X an arbitrary set of objects. Then:

- Z_1 are problems in which information about X is given in the following manner. A certain finite subset X_0 is known as well as rules R with the help of which the whole set X can be constructed from X_0 . Algorithm R_0 is also known by means of which for any given object $x \in X$ it is possible to determine whether x can be deduced from X_0 with the help of R , cannot be deduced or can be deduced by convention;
- Z_2 are problems in which information about X is given like in Z_1 , but rules R are not known. To determine deducibility use is made of the same algorithm R_0 ;
- Z_3 are problems in which information about X is also given through a certain finite subset X_0 . But to determine deducibility use is made of algorithm R_1 which differs from R_0 .

Now let's assume that R_0 is justified for problem Z_1 . Besides, there is a formal representation of objects, any set X_0 in all problems Z_1 - Z_3 can be reduced to this representation without restrictions essential for the solution. Then the proof of justification of algorithm R_1 reduces to the proof of the following assertions.

Assertion 1 (local justification)

For any problem Z_2 there is such statement Z_3 that for any object $x \in X_0$ solutions in Z_2 by algorithm R_0 and solutions in the corresponding statement Z_3 by algorithm R_1 coincide, and for all remaining $x \in X$ are majorized from the point of view of the extent of belonging.

Assertion 2 (global justification)

For any problem Z_1 there is such statement Z_3 that for any object $x \in X$ solutions in Z_1 by algorithm R_0 and solutions in the corresponding statement Z_3 by algorithm R_1 coincide.

It is easy to see that algorithm R_1 is as much justified as R_0 . Note that classes of problems Z_1 - Z_3 are not empty. For example, a well-known problem of propositional calculus belongs to Z_1 . Most of problems related to AI are formulated and solved in statement Z_2 . A well-known problem of pattern recognition can be assigned to class Z_3 .

The description of algorithm R_1 and its properties one can find in [8]. A schematic description of R_1 is presented in this paper.

The notions that will be introduced while describing the scheme are intuitively clear and additional explanation, in our opinion, is not needed. Let's denote by X_0 a finite subset of formal objects from X . Let's assume that for X_0 the solution to the problem is known a priori, i.e. the set is precedent for X . The algorithm R_1 can be characterized by the following features:

- correctness: on objects $x \in X_0$ results of the algorithm operation should coincide with results of a simple sorting out;
- object monotony, understood in the sense of the implication:
monotony of the amount of information in object $x \in X \Rightarrow$ monotony in quantitative and qualitative estimations of properties correlated with x as a result of reasoning.

We can say more precisely about this feature if we introduce a notion of sub-object and relate monotony with a numerical estimation. Such estimation, for example, can be interpreted as a degree of belonging of object x to the set X . Realization of this feature is accomplished with the use of the principle of generalization known in the theory of fuzzy sets;

- decomposition of objects $x \in X$:
each object $x \in X$ may be represented as Cartesian product of sub-objects with respect to objects from X_0 . The estimation that an object $x \in X$ receives as a result of reasoning is a superposition of estimations of sub-objects. It is also monotonous.

The basis for decomposition is a super-object, i.e. a notion

dual of the sub-object. This feature determines the relationship of retrievals X_0 and X , and also permissibility of reasoning.

Any algorithm having the above-mentioned features is an inductive inference. It is an easy matter (in the constructive sense) to build a particular algorithm R_1 . In the general case the number of such algorithms is infinite because parametrization is possible. But even in the result of parametrization we obtain an algorithm R_1 which is justified.

Let's consider the notion of knowledge [7]. Above we have described two conceptions of such definition. It is easy to enumerate their drawbacks. Hence we proceed from the fact that when constructing a different conception of knowledge it is necessary to try to avoid the above-mentioned drawbacks. In so doing, a new conception as a special case should contain at least the ones mentioned above.

The notion of knowledge can be defined on the basis of the following quite evident points:

- knowledge can be defined on two levels, i.e. informal and formal. The relation between the levels exists and is realized with the help of coding (from the informal level to the formal one) and interpretation (from the formal level to the informal one). As a result it is possible to determine a sequence (in terms of categories) for obtaining and processing knowledge;
- knowledge, as an object, is a part of the notion of information. On the informal level objects that are described by these notions coincide. On the formal level information falls into objects that are described by the notions of knowledge and data. These are different notions and they describe different objects. As a result of interpretation both sets of objects are represented as knowledge on the informal level;
- the difference between data and knowledge on the formal level can be described in terms of connections. In this sense data are defined as objects that are completely characterized by connections of one of the following types (or their combination): absence of connection, connections of order, connections of type (or multiplicity). In terms of allowable operations we can say that all these types of connections are described by relational algebra. Data are a set closed under operations of relational algebra. Knowledge, in its turn, is defined as objects that are characterized by connections of succession (inheritance) and deducibility. The language, within which it is possible to unite these two sets of objects, is the language of algebraic systems [9].

It is easy to see that in this case the notion of knowledge representation should satisfy the following conditions:

- to provide support for algorithm R_1 and operations used to describe the connection of succession;
- to allow without distortions to pass from representations existing on the formal level to the desired one;

- to have a possibility to describe connections typical for the notion of data.

In [6,7] one can find description of the representation that satisfy all the enumerated conditions. On the whole, if we leave out details, the representation is quite equivalent to the one used in object-oriented programming. Hence the representation is called an object representation.

One component of such universal representation is a set of elements characterized by the generality of structure and the way of construction in the following sense:

$$\text{object: } \Pi_1 \times \dots \times \Pi_n \rightarrow D_1 \times \dots \times D_n$$

where Π_i are signs, D_i is a set of possible values of a sign ($i=1, \dots, n$). Another component is a certain characteristic of connection or relation between objects. On the formal level relation is a subset of Cartesian powers of the set $\{\text{object}\}^m$. Thus, any object can be identified with a pair $\langle \text{object}, \text{relation} \rangle$, or in the general case with the Cartesian product $\{\langle \text{object}, \text{relation} \rangle\}^m$. The formal sense also has the representation in terms of $\{\langle \text{object}, \text{connection} \rangle\}^m$. As this takes place, all allowable connections, that appear in the definition of knowledge, can be described in terms of the following representations:

$$F: \{\langle \text{object}, \text{connection} \rangle\}^m \rightarrow \{\langle \text{object}, \text{connection} \rangle\}^k$$

where $m, k \in \mathbb{N}$.

Now let's describe briefly the construction of the corresponding notions for the technological level. The key notion here is a *knowledge base* that must satisfy the following conditions:

- to inherit properties of previous technologies (bases and data structures);
- to ensure support for the whole spectrum of connections mentioned above;
- to provide separation of knowledge (data) from means of manipulation;
- to ensure a design process that in case of knowledge can be based on consistency.

By adding interpretation support for knowledge (obtained on the formal level), interface and other features of developed systems to the system having the enumerated properties, we obtain a *knowledge-based management system* (KBMS).

4. Conclusions

Let's discuss the essence of the proposed modification of AI paradigm. On the whole, everything is reduced to the use of justified inductive algorithms. Of course, this restricts the choice to some extent, but as a result we receive much more. First, we have a distinct definition of the notion of knowledge. Second, the result is guaranteed. Third, we have separation of knowledge from the means of its manipulation. The last two factors result in a possibility to have a technology for constructing systems based on knowledge.

The justification of algorithms exhibits not only

theoretical meaning. In practice, the fact that the result is guaranteed leads to the following. Let's assume that we have managed to obtain an a priori estimate characterizing the relative body of the available knowledge. This, for example, can be the capacity of information or entropy (E. Shannon). Let's denote such estimate by α . It is evident that any quality estimate α' characterizing results of the algorithmic processing of new knowledge cannot be higher than α . The justification of the algorithm guarantees that $\alpha' = \alpha$. Of course, this does not take off pathological heuristic nature [10] of practical problems that is the consequence of incompleteness of a priori knowledge.

The introduced paradigm somewhat change the essence of problem solution as well. Following R. Hamming we can say that the aim of problem solution is not numbers but understanding. Such understanding should lead to the construction of a model justified according to all canons of mathematical rigour. The proposed paradigm is a model allowing to advance towards understanding.

And finally. It is clear that the proposed modification is not the only one. Other variants are possible. But to understand what these variants can be it is necessary to have at least one. We hope that the proposed modification of AI paradigm will serve this purpose as well.

The practical significance of the proposed modification is proven by applications, e.g.: "ORTHO-EXPERT" computer system [7,8], which is intended for the diagnosis and treatment of diseases in the field of orthopedics; "Adviser" system that is aimed at supporting decision-making in Internet (<http://www.heeftmijnkindautisme.nl/lkh/>).

Acknowledgements

We are very grateful to Belarusian and Russian Foundations for Fundamental Research for support of this investigation.

REFERENCES

- [1] W.O. Quine Word and object. Cambridge, M.I.T. Press, Mass., 1960.
- [2] A. Thayse and P. Gribomont etc. Approche logique de l'intelligence artificielle. 1 De la logique classique a la programmation logique. Bordas, Paris, 1988.
- [3] J.-L. Lauriere Intelligence artificielle. resolution de problemes par l'homme et la machine. Eyrolles, Paris, 1987.
- [4] J. Barwise (Ed) Handbook of Mathematical Logic, North-Holland, Amsterdam, 1977.
- [5] B.N. Pyatnitsyn Towards a Problem of induction-deduction relations. "Methods of Logical Analysis". Nauka, Moscow, 1977.
- [6] V. Krasnoprosin, V. Obraztsov The Problem of algorithms

- choosing in pattern recognition, *Pattern Recognition and Image Analysis*, 1996, Vol 6. No. 2. p.p. 188-199.
- [7] V. Krasnoproshin, V. Obraztsov, H. Vissia Knowledge as an object of mathematical formalization. *Proc. of Intern. Conf. On Modelling and Simulation*, 1999, - p.p. 207-219.
- [8] V. Krasnoproshin, V. Obraztsov Problem of Solvability and Choice of Algorithms for Decision Making by Precedence, *Journal on Pattern Recognition and Image Analysis*. 2006. Vol. 16. no 2.- p.p.155-169
- [9] A.I. Maltsev Algebraic Systems. Moscow, Nauka, 1970
- [10] Yu.I. Zhuravlev etc. Algorithms for Algebraic and Logical Correction and Their Applications, *Journal on Pattern Recognition and Image Analysis*. 2010. Vol. 22. No 2.- P.155-169.