Methodology for the Synthesis of Information Technologies for Ignorance Modeling: the Key Concepts

Igor Kovalenko[1[0000-0003-2655-6667]], Yevhen Davydenko[2[0000-0002-0547-3689]], Alyona Shved[3[0000-0003-4372-7472]] and Kateryna Antipova[4[0000-0002-9012-5290]]

Petro Mohyla Black Sea National University, Mykolaiv, Ukraine
{ihor.kovalenko,davydenko,avshved}@chmnu.edu.ua

Abstract. The analysis of the most studied types of ignorance, such as imprecision, uncertainty, inconsistency, conflict, fuzziness, etc., with the aim of their identification, systematization and formalization was carried out. Methods of modeling various types of ignorance on the basis of modern theories have been considered: the fuzzy set theory, the theory of evidence, the theory of plausible and paradoxical reasoning, the rough set theory. In the framework of the study, the methodology for the synthesis of information decision support technologies for modeling different types of ignorance through the systematic application of artificial intelligence methods has been proposed. These concepts are based on a systematic approach to the identification of different types of ignorance, which creates the conditions for the correct selection and application of methods of analysis of the initial data. This, in turn, provides effective results when modeling relevant subject and problem areas of knowledge. An important task in this context is the reasonable choice of a mathematical apparatus capable of detecting, exploring and modeling various types of ignorance correctly.

Keywords: Information Technology, Decision-making, Expert Evidences, Ignorance.

1 Introduction

Information technologies (IT), which are a tool for implementing systems analysis methods, have been intensively developing in the last two decades within the framework of a scientific field called “knowledge engineering”. The basis of this scientific field is the results of development and research related to artificial intelligence (AI): knowledge representation and reasoning, and knowledge inference; artificial intelligence systems (expert systems, pattern recognition systems, decision support systems, etc.).

In the AI the analysis and management of various types of ignorance have a paramount importance, due to the creative nature of the tasks of creating intelligent technologies, which are always solved under conditions of inconsistency, incompleteness, inaccuracy, uncertainty of the source data, relations between them, processing operations (algorithms, processes solutions). The term “non-factors” is used in [1-3] to describe various types of ignorance.

Such types of ignorance (non-factors) as “fuzziness” and “inaccuracy”, were identified and studied in the framework of fuzzy mathematics, founded by Lotfi Zadeh [4]. Purposeful studies of non-factors began with research works of Narin`yani, which introduced the concept and defined the interpretation of non-factors [1, 2]. Works of Vagin, Ry`bina, Borisov, Val`kman, and etc., Burrus and Lesage reflect other approaches for the determination of non-factors [5-11].

Paper [5] identifies five basic types of ignorance (non-factors): contradiction, non-monotony, imprecision, uncertainty, fuzziness; [8] highlights: unknown, unreliability (falsity) and ambiguity.

At the same time, Smithson distinguishes two main types of ignorance: error and irrelevance [12]. Bonissone and Tong argue that there are three main types of ignorance: uncertainty, incompleteness and imprecision [13]. Bosc and Prade highlight four main types of ignorance that can penetrate information, namely: uncertainty, imprecision, vagueness, and inconsistency [14].

Thus, the study of them leads to the conclusion that despite a meaningful analysis of types of ignorance, they do not set out the principles of their unification and formalization.

But, it’s often modern methods of fuzzy mathematics, probabilistic-statistical inference, Bayesian and neural networks, genetic algorithms, etc., are used without proper analysis of the nature of the types of ignorance present. This fact can lead to inadequate models and conclusions.

The purpose of the article is to research the most studied types of ignorance and methods for their modeling, and to develop methodology for the synthesis of information technologies to support decision-making process under various types of ignorance.

2 Ignorance Handling Theories

Probability Theory deals with the chances of random events, while it is assumed that all events are well-defined concepts. In this case uncertainty is connected only with what chances each random event from the full group of such events can occur.

It should be pointed out that there are two main approaches for estimating the probabilities of events: objective probabilities based on the frequency method and experts, which are the sources of subjective probabilities. In the framework of probability theory for uncertainty modeling, analytical methods of probabilistic inference (probability trees, decision trees, Bayesian networks), methods of mathematical statistics, etc., can be used [15].

Fuzzy Set Theory is used to operate with fuzzy concepts that underlie the formation of sets of elements [4]. Elements are supposed to be well defined concepts. Uncertainty (fuzziness) here arises when trying to attribute elements to some classes (sets), since these classes (sets) are fuzzy, therefore, poorly defined.

In real conditions, there may also exist specific forms of non-factors (ignorance) that arise in the process of interaction between expert judgments. The forms of such interactions can have a different character – they can be consistent, compatible; can arbitrarily unite and intersect. To simulate these forms of interactions, the mathematical apparatus of the Dempster-Shafer Theory (DST, evidence) can be used. DST considers the frame of discernment (set of hypotheses) as a set of exclusive and exhaustive elements [16-18].

The Dezert-Smarandache Theory (DSmT) of plausible and paradoxical reasoning can be considered as a more in-depth version of the DST in the sense that it can operate with more complex forms of ignorance that can simultaneously be present in the formal system [19, 20]. The DSmT deals correctly with uncertain, highly conflicting and imprecise sources of evidence (set of hypotheses, expert data). Within the framework of this theory, elements of the frame of discernment (set of hypotheses) can reflect vague, imprecise concepts, and as a result can overlap each other. Thus, only the limitation of the mutual exhaustibility of elements of the frame of discernment is supported.

The methods proposed in the framework of Rough Set Theory (RST) allow to process large amounts of disordered data and obtain new knowledge based on the results of such processing [21-23]. This approach allows to correctly process the inaccurate expert information.

A number of approaches based on the integrated application of the mathematical apparatus of the considered theories to solve typical decision-making problems under complex forms of ignorance (e.g., fuzziness and uncertainty; height conflict, ambiguity and uncertainty; imprecision, incompleteness and uncertainty, etc.) have been proposed in [24-29].

3 Synthesis of Information Decision Support Technologies Under Various Types of Ignorance

The systematic methodology for the synthesis of IT for modeling non-factors allows to generate information decision support technologies on the basis of the generated set of rules and a set of parameters for the synthesis of IT, such as the type of data analysis task, data structure, identified types of ignorance (non-factors), or combinations thereof, etc.

The methodology for the synthesis of IT can formally be presented in the form of the following successive stages:

Stage 1. Determining the goals of analysis (evaluation).

There are two types of goals in system analysis: qualitative and quantitative. The form of obtaining the evaluation result depends on determining the purpose of the assessment.

Stage 2. Determining the composition and structure of the analysis task.

There are five basic data mining tasks: classification; 2) clustering; 3) association rule mining; 4) sequence data mining; 5) forecasting. In the practice of decision making, there are three main tasks: ordering of alternatives (ranking); distribution of alternatives to decision classes (clustering); choosing the best alternative.

The type of analysis task determines the type of data structuring procedure, in order to forming a final solution. So, for example, to solve the problem of choosing the best alternative, the ranking procedure of the initial set of alternatives can be used.

Stage 3. Determining the method of obtaining initial data (information).

The methods for obtaining information can conditionally be divided into next groups: empirical and theoretical. The most widespread are empirical methods of obtaining information, among which are the description, comparison, measurement, observation, experiment, analysis, etc. Examples of empirical data are research results, respondents' answers, experts' assessments, results of observations, measurements, etc. One of the most common empirical methods is the expert judgment. The obtained information can be both qualitative and quantitative, and represents estimates in one of four basic data measurement scales (ordinal, interval, ratio, nominal).

The choice of the method of obtaining the source information affects the structure of the source data.

Stage 4. Formation of a set of source data.

The procedure for generating a set of source data largely depends on the selected method of obtaining source information. At this stage, the source data takes the form of numbers, rankings, paired comparisons, intervals, etc., depending on the selected measurement scale. Thus, the structure of the source data is formed.

Stage 5. Identification and selection of methods for ignorance (non-factors) modeling.

The choice of modeling methods depends on the structure of the source data and the types of ignorance that influenced the process of extracting information and forming a set of source data, or contained in the received information (data set).

Stage 6. Synthesis of information decision support technologies under identified types of ignorance (non-factors).

The basis of the methodology for synthesis of IT for ignorance (non-factors) modeling is a model as follows:

$$SIT=\langle D, N, P, M, SGR, IP\rangle, \tag{1}$$

where $D = \{d_i \mid i = \overline{1,r}\}$ is a set of initial data; N is a ignorance identification procedure; P are the IT synthesis parameters (criteria); M is procedure for selection of the mathematical apparatus used to solve the problem of structuring of the initial information (data); SGR is rule system for information decision support technologies generation; IP are information processes.

The procedure for identifying the type of ignorance is a set of rules for identifying types of ignorance that analyzed in the system, based on a given set of criteria for their identification.

M may be represented by a method or group of methods that allow to correctly process data under identified type of ignorance, or a group thereof.

The IT generation rule is an algorithm for solving the stated analysis problem based on the mathematical apparatus used, taking into account the specifics of the source data. It is proposed to use the type of data structuring procedure, data structure, the method of obtaining initial information (data), form of result presenting, etc., as parameters (criteria) $\mathrm{K} = \{K_i \mid i = \overline{1, m}\}$ for IT generation rules construction. It can be used one K_i, or more criteria $\wedge K_i$, $i \le m$, for rules generation.

The IT generation rules can be represented as follows:

- one criteria is highlighted for the synthesis of IT:

$$PIT_j: \quad K_i \to IT_j. \tag{2}$$

- several criteria are highlighted for the synthesis of IT:

$$PIT_{j+1}: \quad \wedge K_i \to IT_{j+1}. \tag{3}$$

As an antecedent, one or a combination of criteria for the synthesis of IT is used, in the role of a consequent, the information technology generated, taking into account the formed criteria, is used.

IP is a set of algorithms for the implementation of information processes for obtaining, processing, exchanging, displaying data and knowledge generation.

The basic principles of the synthesis of IT invariant to the type of problem being solved and the method for identifying and presenting initial data (information).

Let us consider an example of the generation of IT for structuring of expert assessments under different types of non-factors based on the proposed concept (Fig. 1).

It was used one criterion (K_1 is a method of expert judgment as a method of obtaining initial information) for construction of IT generation rule.

The rule for IT generating will take the next form:

$$PIT: \quad K_1 \to IT.$$

Let us consider the structure and key steps of the information technology for structuring of expert assessments.

In general, the procedure for identifying and processing expert information consists of the following steps:

1. Identification of the goal of assessment. At this stage, the composition of the evaluation task, the type of procedure for structuring expert assessments, the form of presentation of the expected results are determined.
2. Development of a scenario for the examination. Within the framework of this stage, the technical issues of the examination and the method of presenting expert assessments (for example, numbers, rankings, partitions, etc.) are solved.
3. Expert group formation in accordance the level of competence for each expert.
4. Collection of expert information, its structuring.
5. Identification of types of ignorance (non-factors) that may occur in the information received.
6. Selection of a mathematical apparatus that allows to operate correctly with the identified types of ignorance.
7. Analysis of expert information based on the selected method.
8. Analysis of the results and synthesis of the final solution.

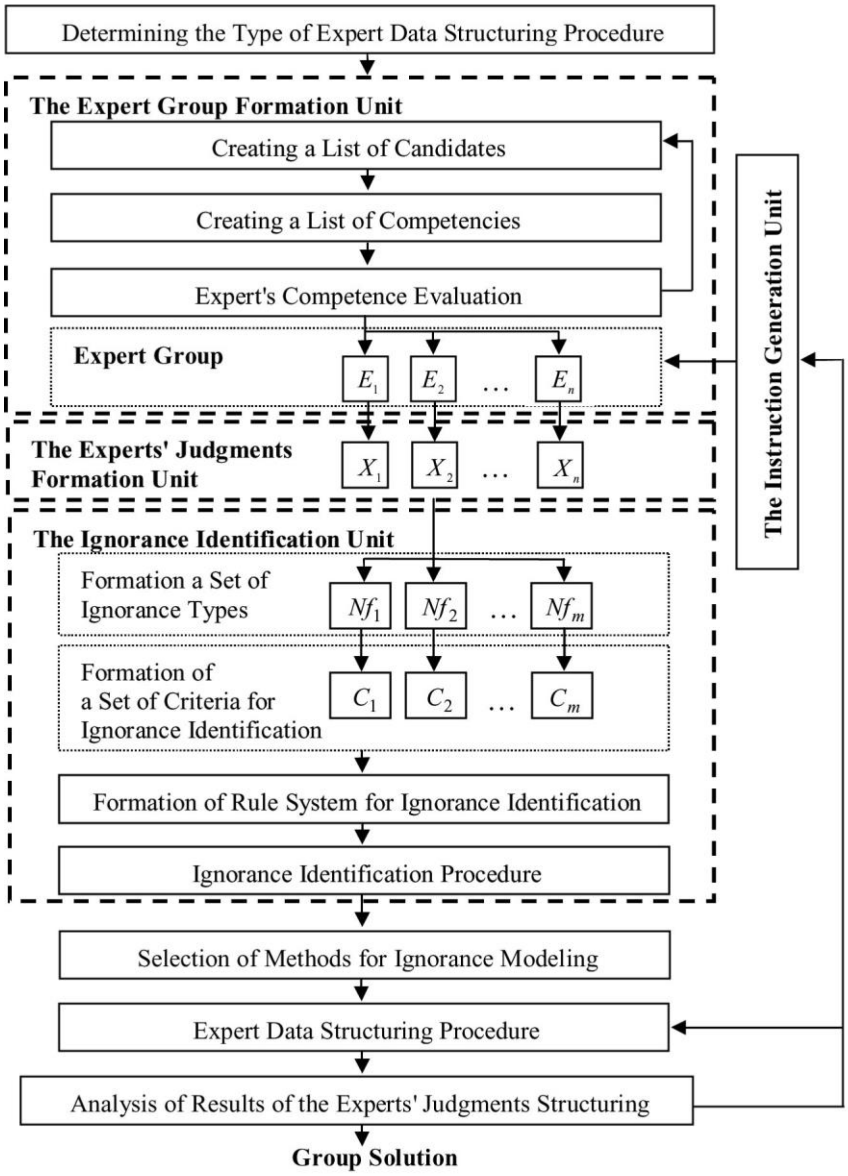

Fig. 1. – Structural scheme of IT for expert's data structuring under various types of ignorance

4 Conclusion

The complex analysis of the most studied types of ignorance (non-factors) was carried out. Methods of modeling various types of ignorance on the basis of modern theories have been considered: the fuzzy set theory, the theory of evidence, the theory of plausible and paradoxical reasoning, the rough set theory. The mathematical apparatus of the considered theories allows to operate correctly with various specific types of ignorance and their combinations. The performed analysis puts forward the conditions for a detailed analysis of non-factors, that ensures the correct choice of methods their modeling represented by the considered theories.

The methodology for the synthesis of information decision support technologies for ignorance modeling has been proposed. These concepts could be implemented as part of the tools of the automated expert support systems to ensure the choice of optimal solutions for the planning and implementation of projects for various purposes. Especially for solving ill-structured problems under imprecision, uncertainty, inconsistency and conflict.

Acknowledgment. This research was partially supported by the state research project "Development of information and communication decision support technologies for strategic decision-making with multiple criteria and uncertainty for military-civilian use" (research project no. 0117U007144, financed by the Government of Ukraine).

References

1. Narin`yani, A. S.: Netochnost` kak NE-faktor. Popy`tka doformal`nogo analiza [in Russian].Preprint RosNII II 2 (1994)
2. Narin`yani, A. S.: Nedoopredelenny`e modeli i operacii c nedoopredelenny`mi znacheniyami [in Russian]. Novosibirsk (1982)
3. Dushkin, R. V.: Metody` poluchiniya, predstavleniya i obrabotki znanij s NE-faktorami [in Russian]. Vebov i Knigin (2011)
4. Zadeh, L. A.: Outline of a new approach to the analysis of complex systems and decision processes, IEEE Transactions on Systems, Man and Cybernetics SMC-3(1), 28–44 (1973)
5. Vagin, V. N.: Znanie v intellectual`ny`x sistemax [in Russian]. Novosti iskusstvennogo intellekta 6, 8–18 (2002)
6. Dushkin, R. V., Ry`bina, G. V.: Ob odnom podxode k avtomatizirovannomu izvlecheniyu, predstavleniyu i obrabotke znanij s NE-faktorami [in Russian]. Izvestiya RAN. Teoriya i Sistemy` Upravleniya 5, 84–96 (1999)
7. Ry`bina, G. V.: Modeli, metody` i programmny`e sredstva dlya postroeniya integrirovanny`x e`kspertny`x system [in Russian]. Avtoreferat dissertacii na soiskanie uchenoj stepeni doktora texnicheskix nauk: 05.13.11 (2004)
8. Borisov, A. N., Krumberg, O. A., Fedorov, I. P.: Prinyatie resheniy` no osnove nechetkix modeley` [in Russian]. Zinatne, Riga (1990)
9. Val`kman, Yu. R. By`kov, V. S., Ry`xal`skij, A. Yu.: Modelirovanie NE-faktorov – osnova intellektualizacii komp`yuterny`x texnologiy` [in Russian]. Sy`stemni Doslidzhennya ta Informaciy`ni Texnologii 1, 39–61 (2007)
10. Burrus, N., Lesage, D.: Theory of evidence (DRAFT) (Technical report no. 0307–07/07/03). Laboratorie de Reserche et Developpment de l`Epita, Le Kremlin-Bicetre cedex (2003)
11. Bradley, R., Mareile, D.: Types of uncertainty. Erkenntnis, 79, 1225–1248 (2014)
12. Smithson, M.: Ignorance and uncertainty: emerging paradigms. Springer, New York (1989)

13. Bonissone, P. P., Tong, R. M.: Editorial: reasoning with uncertainty in expert systems. International Journal of Man-Machine Studies, 22, 241–250 (1985)
14. Bosc, P., Prade, H.: An introduction to the fuzzy set and possibility theory-based treatment of flexible queries and uncertain or imprecise databases. In: Motro, A., Smets, P. (eds) Uncertainty Management in Information Systems. Springer, Boston, 285–324 (1997)
15. Uzga–Rebrovs, O.: Modern concepts and applications of probability theory [in Russian]. RA Izdevnieciba, Rezekne (2004)
16. Shafer, G. A.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
17. Dempster, A. P.: A generalization of Bayesian inference. Journal of the Royal Statistical Society B(30), 205–247 (1968)
18. Liu, L., Yager, R. R.: Classic works of the Dempster-Shafer theory of belief functions: an introduction. Studies in Fuzziness and Soft Computing, 219, 1–34 (2008). doi: 10.1007/978-3-540-44792-4_1
19. Smarandache, F., Dezert, J.: Advances and applications of DSmT for information fusion. Collected works, 2. American Research Press, Rehoboth (2006)
20. Smarandache, F., Dezert J., Tacnet J.: Fusion of sources of evidence with different importances and reliabilities. In 13th International Conference on Information Fusion, pp. 1–8. IEEE, Edinburgh (2010). doi: 10.1109/icif.2010.5712071
21. Pawlak, Z.: Rough sets theoretical aspects of reasoning about data. Kluwer Academic Publishers, Boston; London (1991)
22. Uzga–Rebrovs, O.: Knowledge representing features in rough sets. In: 7th International Scientific and Practical Conference: Environment. Technology. Resources, vol. 2, pp. 169–175. RA Izdevnieciba, Rezekne (2009)
23. Skowron, A., Dutta, S.: Rough sets: past, present, and future. Natural Computing, 17, 4, 855–876 (2018). doi: 10.1007/s11047-018-9700-3
24. Kovalenko, I., Davydenko, Ye., Shved, A.: Development of the procedure for integrated application of scenario prediction methods. Eastern-European Journal of Enterprise Technologies, 2/4(98), 31–37 (2019). doi: 10.15587/1729-4061.2019.163871
25. Kovalenko, I., Davydenko, Ye., Shved, A.: Formation of consistent groups of expert evidences based on dissimilarity measures in evidence theory. In: 14th International conference on Computer sciences and Information technologies (CSIT 2019), pp. 113–116. IEEE Press, Lviv (2019)
26. Shved, A., Davydenko, Ye.: The analysis of uncertainty measures with various types of evidence. In: 1st International Conference on Data Stream Mining & Processing, pp. 61–64. IEEE Press, Lviv (2016). doi: 10.1109/dsmp.2016.7583508
27. Kovalenko, I., Davydenko, Ye., Shved, A.: Searching for Pareto-optimal solutions. In: Shakhovska, N., Medykovskyy, M. O. (eds) Advances in Intelligent Systems and Computing IV. Selected Papers from the International Conference on Computer Science and Information Technologies (CSIT 2019). Springer International Publishing (2020). doi: 10.1007/978-3-030-33695-0_10
28. Dezert, J., Tacnet, J.-M.: Evidential reasoning for multi-criteria analysis based on DSmT-AHP. Advances and Applications of DSmT for Information Fusion. Collected Works, 4, 95–100 (2015). doi: 10.5281/zenodo.22462
29. Beynon, M. J.: Reflections on DS/AHP: Lessons to Be Learnt. Lecture Notes in Computer Science, 95–104 (2014). doi: 10.1007/ 978-3-319-11191-9_11