## Author's Information

Alexander Makarenko – National Technical University of Ukraine "KPI", Institute for Applied System Analysis Department of Mathematical Methods of System Analysis
37, Pobedy Avenue, 03056, Kiev-56, Ukraine, makalex@i.com.ua, makalex@mmsa.ntu-kpi.kiev.ua

# TOWARDS THE PROBLEMS OF AN EVALUATION OF DATA UNCERTAINTY IN DECISION SUPPORT SYSTEMS

Victor Krissilov, Daria Shabadash

*Abstract: The question of forming aim-oriented description of an object domain of decision support process is outlined. Two main problems of an estimation and evaluation of data and knowledge uncertainty in decision support systems – straight and reverse, are formulated. Three conditions being the formalized criteria of aim-oriented constructing of input, internal and output spaces of some decision support system are proposed. Definitions of appeared and hidden data uncertainties on some measuring scale are given.*

*Keywords: decision support systems, straight and reverse problems of data uncertainty, three conditions of aim-oriented object domain constructing, appeared and hidden uncertainties.*

*ACM Classification Keywords: H.4.2 Information Systems Applications: Decision support Systems*

## Introduction

One of the most actual questions of decision making theory – is the question of forming aim-oriented description of an object domain, namely, description of input, internal and output spaces of decision support systems (DSS). Practically, any input data has uncertainty, sources of which can be: inaccuracy of measuring and inaccuracy of rounding-up, scale restrictions, impossibility of measuring or definition of values with needed precision, hidden semantic uncertainty of qualitative data, etc [1, 2]. In addition, uncertainty in DSS may be caused by methods, used for obtaining, storage and processing of knowledge. A great deal of uncertainty to the decision making process brings the subjective factor that appears when the person making a decision (PMD) formulates the set of alternatives decisions and the set of descriptive criteria for them.

Main known approaches to the evaluation of uncertainty in DSS are methods of the probability theory [3, 4] and methods of fuzzy logic [2, 5]. The first are used in that case, when the extensive statistical information about the decision making process is accessible. The second are applied for description of system behavior, when it is too expensive or practically impossible to construct precise mathematical models. However, frequently in real DSS there is a necessity of the composite approach for estimation and aim-oriented handling of input and output space uncertainty.

The given paper is devoted to the problems of an estimation and evaluation of data and knowledge uncertainty in DSS.

## Straight and Reverse Problems of Data Uncertainty in DSS

We will consider some DSS in the way of a "black box" (fig. 1).

On fig. 1. are represented:

$X = \{x_1, x_2, ..., x_n\}$ - the set of input parameters (dimensions);

$Y = \{y_1, y_2, ..., y_m\}$ - the set of output parameters (dimensions);

$Q = \{q_1, q_2, ..., q_l\}$ - the set of internal (intermediate) states (dimensions).

The representation form of results, to be exact – uncertainty that exists in them, we shall designate it $N_3$, essentially influences on a constructional usage of them in a particular problem of decision making, and it is characterized by the working conditions of DMS as a whole. Uncertainty of results $N_3$, is conditioned by uncertainty of input data ($N_1$) and uncertainty of system ($N_2$) (fig. 1.) [1].

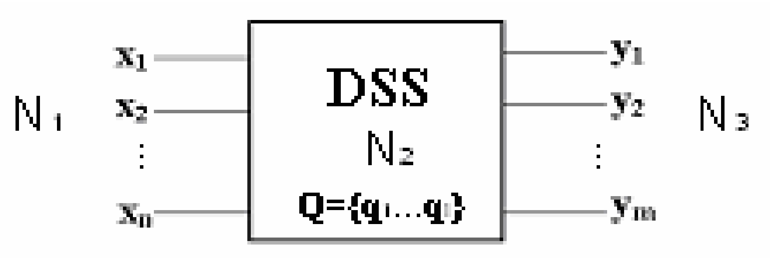


Fig. 1. Uncertainty in DSS

Within the frameworks of such approach, let's formulate two main problems of estimation and evaluation of data and knowledge uncertainty in DMS – straight and reverse.

The straight problem consists of determination of result's limit accessible uncertainty $N_3$, on the base of known uncertainty of input data $N_1$ and uncertainty of system functional $N_2$. Then making a comparison of received $N_3$ with the value of a result's limit acceptable uncertainty $N_{3max}$, that is determined by PMD, on the base of solving tasks aim. This problem arises when, on the base of already available data, for example, stored in some data warehouse [6, 7] and had some level of uncertainty, it is necessary to construct the definite rules for decision making.

The reverse problem consists of aim-oriented forming of internal and input dimensions so, that it can provide an uncertainty of output dimension $N_3$ not bigger than top limit acceptable uncertainty $N_{3max}$. This problem arises at solving tasks of pattern recognition, cluster analysis, constructing of object domain of some DSS [6, 7].

Solving two main problems of estimation and evaluation of data and knowledge uncertainty in DMS makes possible to formulate three main conditions, being the formalized criteria of aim-oriented constructing of input, internal and output spaces of some DSS.

1. Condition of insufficient detailing (an excessive generality) of space:

$$N_3 > N_{3max}$$

2. Condition of redundant detailing of space:

$$N_3 < N_{3min}.$$

3. Condition of constructive usage of space:

$$N_{3\,max} \geq N_3 \geq N_{3\,min} .$$

Where:

$N_3$ – uncertainty of result, calculated on the base of input data uncertainty ( $N_1$ ) and uncertainty of system ( $N_2$ )

$N_{3\,max}$ , $N_{3\,min}$ – respectively, top limit acceptable and low limit sufficient uncertainties, determined from the aim of decision support task

Surely, essential requirement is that - $N_{3\,min} <= N_{3\,max}$ .

## Concepts of Appeared and Hidden Data Uncertainties

In practice, usually, process of formation of DSS's input and output spaces has iterative character. At the same time, each iteration represents conversion between various types of scales, or transition to more or less detailed scale of the same type. So, the straight problem formulated above is, from this point of view, the process of sequential granulation. The reverse problem represents the process of sequential decomposition. Traditionally values $N_1$ , $N_2$ and $N_3$ characterize uncertainty of DSS on some final iteration [3, 4]. Hence, the big influence on the solving problem has type of the scale, which is used for display of input and output spaces. Depending on a required precision, measuring scales of various types are used: nominative, order, interval, relative and absolute [7].

Let's consider more in detail representing of some data on different scales.

First of all, in an explicit form, there is some set of values on a scale, the amount and form of which depends on the type of selected measuring scale. Up to the moment of measurement (observation), there is uncertainty of what value on a scale will be selected as a result of measurement. This uncertainty can be semantically compared to the entropy of the initial alphabet, known in information theory [3, 4]. Thus, the uncertainty of the measuring scale values set, described above, we shall name *the appeared uncertainty*, and designate as $H_{np}$ . Usually, during the characterizing of some measurement uncertainty only this uncertainty is taken into consideration.

However, on the other hand, data on a measuring scale are represented with some finite precision. It means, that each value on the scale hides in itself whole "cloud" of the real values. At that, distinguishing these values is impossible because of resolution limitation of measuring devices or inexpedience of this for the given task. Thus, some value on a scale represents analogue of concept of the granule, offered by L. A. Zadeh [2]. Therefore, takes place the uncertainty of the data, which is "hidden" in values of a measuring scale. We shall name it as hidden uncertainty, and designate as $H_{ск}$ .

Let's choose the scale of absolute type and consider the limiting case, when only two values are located on it (for example, «0» and «1»). In this case, appeared uncertainty of the scale is minimal, as the possible quantity of values on it – is minimal. Hidden uncertainty, in this case, on the contrary – is maximal, as in two values, that lies on the given scale, all variety of possible values of entrance data is contained. When increase in scale detailing, obviously, the quantity of values on the scale increases and the number of the "not distinguished" values decreases. Hence, appeared uncertainty of the scale increases, and hidden - decreases. At use of all possible values on the absolute scale, hidden uncertainty - is minimal and is defined only by inaccuracy of the received data. Appeared uncertainty, at the same time, - achieves its maximum.

As there is unique transformation from strong to weaker scales, the changing of appeared and hidden uncertainties values, described above, is valid for other types of scales - nominative, order, interval, and relative. Definitions of appeared and hidden uncertainties are given independently of measuring scales types.

## Conclusion

Choice of measuring scale type determines the form of data representation in DSS data domain. Then, the ratio of hidden and appeared uncertainties can characterize conversion between various measuring scales, on each iteration of forming DSS's input, internal and output spaces.

So, considering the straight problem from the point of view of appeared and hidden uncertainties, formulated above, we shall receive the following. At the known uncertainty of input data ($N_1$) and uncertainty of the system ($N_2$) the process of solving the straight problem represents the process of sequential granulation of input scales values up to obtaining the result with the uncertainty $N_{3min} \le N_3 \le N_{3max}$. Thus, it is expedient to estimate changing of appeared and hidden uncertainties on each iteration of this transformation, in order to check up conditions (1), (2), (3).

Similarly, at solving of the reverse problem, the basic carried out operation is – decomposition. At the known uncertainty of results $N_3$, it is expedient to characterize process of sequential decomposition from the result scale to the input space scales by changing of appeared and hidden uncertainty values on each iteration.

Real tasks often are the composition of these processes, i.e. demands iterative execution of both: granulation and decomposition. And exactly the analysis of appeared and hidden uncertainties changes on each iteration makes all process of solving straight and reverse problems aim-oriented. Hence, on the basis of the introduced concepts of appeared and hidden uncertainty, it becomes possible to characterize and manage the processes of decomposition and granulation at formation input and output spaces of DSS.

The further studies should be directed to the development of formalized methods of the quantitative evaluation of data and knowledge uncertainty, supplying a choice and/or developing of adequate means for decision making process.

## Bibliography


[1] В.А. Крисилов, Д.В. Шабадаш. Неопределённость данных и знаний в системах поддержки принятия решений. In: Искусственный интеллект 1'2003. Наука і освіта, Донецк, 2003.

[2] L. A. Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. In: Fuzzy Sets Syst., Vol. 90, No. 2, 1997.

[3] C. E. Shannon. Information Theory. In: Encyclopedia Britannica. IL, 14th Edition, Vol. 12, pp. 246B-249, Chicago, 1968.

[4] A.N. Kolmogorov, A.N. Shiryayev. Selected Works of A.N. Kolmogorov: Information Theory and the Theory of Algorithms (Mathematics and Its Applications). In: Kluwer Academic Pub, 1993.

[5] R. R. Yager. On the measure of fuzziness and negation. II. Littices. In: Information and Control, 44, 236 – 260, 1980.

[6] Peter Jackson. Introduction to expert systems. In: West Group, Rochester, NY, 2001.

[7] Nikolay G. Zagoruiko. Applied methods of data and knowledge analysis. In: Novosibirsk, 1999.


## Authors' Information


**Victor A. Krissilov** – ph. d., head of chair "System Programming", computer departament, Odessa national polytechnic university. Shevchenko Av.1, Odessa 65044, Ukraine, e-mail: victorK@405.com.ua

**Daria V. Shabadash** – post-graduate student, chair "System Programming", ONPU, e-mail: dara_sh@mail.ru