Models and methods of decision-making and probabilistic methods in information processing as an applied tool for data evaluation

S G Zbrishchak\textsuperscript{1} and L S Zvyagin\textsuperscript{1}

\textsuperscript{1} Department of System Analysis in Economics, Financial University under the Government of the Russian Federation, 49 Leningradsky Prospekt, 125993, Moscow, Russia

E-mail: sv1417@yandex.ru, sdmif@yandex.ru

Abstract: Decision-making is a process that requires a lot of effort, sufficient knowledge, and making a choice by analyzing various solutions to many problems. The set of problems is called in another way the set of alternatives \( x \). When solving the problem of decision-making in organizational models, in fact, we make a choice, giving preference to one or another option, based on characteristics and certain criteria. But first, you need to correctly form the initial alternatives, based solely on the conditions and limitations of external systems that affect the problem being solved. If the decision-maker (hereinafter referred to as the LPR) decides to ignore the requirements of external systems, the formed alternatives will not be complete enough, and the choice will be limited. As a result, it will not be possible to find the best solution among the presented ones. In General, it is accepted to consider the decision-making process as a system that consists of a certain set of typical stages and their actions that interact with each other. The number and stages of actions may vary, because the conditions and types of tasks are different. In turn, among all existing methods of data processing and analysis, mathematical methods of information processing are distinguished, which include statistical sequential analysis, the basis of which is probabilistic methods, which serve as one of the most effective analysis tools.

1 Introduction

Probabilistic methods belong to the field of General scientific and practical activity, they are represented by a wide variety of methods used for data analysis in various spheres of public life. Among them, the least squares method, exponential smoothing method, probabilistic modeling method, and adaptive probabilistic forecasting methods are highlighted, which will be discussed in this article. This is due not only to the fact that the scope of their use is quite extensive, but also to the fact that probabilistic methods have their own specifics in processing information, which is based on making a certain probability of an event, based on the source data directly related to the model and method of data collection.

The source information is the input element of the decision-making system (DSS), and the set of valid decisions is the output element of the DSS. There are also informal elements of the SPR: forming alternatives, evaluating them, and selecting optimal options. This article will discuss the various levels of models to build organizational system, namely the model objectives, model design, model of research process and design of organizational systems (hereinafter – OS), model skills, the technological model. Next, special attention will be paid to decision-making models classified by two criteria: by the number of goals and by the scope of application. It will also cover methods for finding solutions, such as the full search strategy, the implicit method, heuristic methods, the method of analogy, simplification, aggregation, and so on.
2 Model levels for building an organizational system

An organizational system is an artificial object (hereinafter referred to as an IO) that is created by people based on their goal, having developed a plan in advance. The organizational system is based on a constructive model.

The process of modeling constructive activity (hereinafter-CD) is quite specific and time-consuming. Because requirements, capabilities, and knowledge affect the modeling process, it is subject to constant changes. And in order to choose a rational version of the model of constructive activity, it is necessary to be based on schematization, formalization and automation of building a series of models.

You can think of a CD for creating an organizational system as interconnected models:

\[ M_{KD} = \{ M_C, M_o, M_L(M_{PR} \cup M_{USL}), M_T, M_Y \} \]

where \( M_{KD} \) – is a model of constructive activity.

Next, let's look at each model separately.

1) Goal model (\( M_g \)). In this model, there are typically only knowledge about a certain subject area. There are four different ways to describe this knowledge:

\* Morphological description – this refers to the description of objects, relationships, and structures;
\* Functional – describes parameters, processes, and quality;
\* Natural – the basis of laws, rules, relationships, principles of construction;
\* Informational – includes a description of hypotheses, restrictions, and conditions.

In this model, a certain theory is formed, and various physical and constructive invariants of components and connections are considered as axioms. Here it is impossible to formulate and calculate the number of such axioms in advance, since new invariants-axioms can arise in the course of activity. Then, at the stage of constructive activity, the list of axioms becomes more or less complete.

2) Model of the design object (\( M_o \)). In this model, there are typically only knowledge about a certain subject area. There are different ways to describe this knowledge:

\* Morphological description – this refers to the description of objects, relationships, and structures;
\* Functional - describes parameters, processes, and quality;
\* Natural – the basis of laws, rules, relationships, principles of construction;
\* Informational – includes a description of hypotheses, restrictions, and conditions.

In this model, a certain theory is formed, and various physical and constructive invariants of components and connections are considered as axioms. Here it is impossible to formulate and calculate the number of such axioms in advance, since new invariants-axioms can arise in the course of activity. Then, at the stage of constructive activity, the list of axioms becomes more or less complete.

3) Model of the process of research and design of the OS (\( M_{pr} \)). The model under consideration contains knowledge about the means of OS research and construction. These research tools are methods, procedures, and operations related to the transformation, construction, and analysis of models.

4) the skill Model describes the process of applying OS research and design tools. This model contains certain actions for the use of research tools, which are supplemented by certain conditions (meta-knowledge). This model includes logical action diagrams that help you build and transform models of the first, second, and third layers. The skill model (logical) can be defined as follows: \( M_L(M_{PR} \cup M_{USL}) \)

5) In the technology model (\( M_T \)) contains functions for detecting and correcting discrepancies that may occur between the selected model layers, as well as the models used and the external world. For different stages of the life cycle of KD models, different classes of functions are needed to detect discrepancies and resolve them. There are three classes of such functions:

\( C_I \) – the misalignment of the verbal structures appear as the result of the decomposition of knowledge;
\( C_{II} \) – misalignment between the three different States of the OS: objective, predictable and desirable. Objective refers to the actual state of the operating system acquired during the operation stage. The predicted state provides a model for predicting the functioning of the organizational system. If there is a mismatch between the objective and predicted state, you need to review the OS models.
\( C_{III} \) – mismatch between different States of CD process models as objects of research and construction. Next, we can consider the life cycle of IO as the final product of activity (see figure 1).

![Figure 1. The life cycle of IO.](image-url)
First, you need to create a need to solve the problem. You can call this a stage of awareness that there is a problem situation. Next, you need to directly create a goal to resolve the problem situation. At the third stage, a plan is built to achieve the goal, and it is approved. And then the implementation of the plan is already organized and monitored. This life cycle can be considered both in space and in time.

3 Methods for solving multi-criteria selection problems

In multi-criteria problems, LPR is given alternatives that need to be evaluated using various criteria and find the best alternative. But in multi-criteria problems, there is a difficulty associated with the ambiguity of the choice. To resolve this problem, use the following groups of methods:

Group I: methods related to the first group are designed to reduce the number of criteria by introducing additional assumptions that relate to ranking criteria and comparing alternatives.

This group includes the following methods:
- convolutions;
- main criterion;
- threshold criterion;
- distances

Group II: this group contains methods aimed at reducing the number of alternatives in the original set by excluding deliberately bad alternatives.

For example, the second group includes the following methods:
- method for constructing the Pareto set;
- guaranteed result method;
- the method is based on the principle of Nash (equilibrium)

Let's look at each group in more detail. Let's start with the first one.

With the convolution method, we replace the original criteria (let's call them local) with one General criterion. In another way, this operation is called aggregation of private criteria. This method is useful if it is possible to arrange local criteria in descending order of importance.

There are three convolution models: additive, multiplicative, and Maximin.

The essence of the main criterion method is clear from the name. The main criterion is selected from the set of initial criteria, which exceeds the other criteria in importance.

And the last method from the first group, as already mentioned before, is the state method, which consists in introducing distance in space. This is done in order to find a reference (ideal) solution that corresponds to the absolute maximum point in the criteria space.

Next, we will consider methods aimed at reducing the number of alternatives, excluding bad ones. One of these methods is the construction of the Pareto set. This method is used for a large number of alternatives that have conflicting ratings for different criteria.

In this case, if we used methods from the first group, the results would be unreliable. Therefore, in order to reduce the number of alternatives, they build a set of paretos. An alternative will belong to a set if it is not worse than other criteria and is better than at least one. Thus, the number of alternatives is reduced, and the final choice is left to the LPR.

The advantage of analyzing the Pareto set is that the alternatives do not lend themselves to pairwise comparison, because one alternative will be better for one criterion and another for another, and this is the complexity that the Pareto set solves. Using this model, you can find a compromise between conflicting requirements.

Only after the Pareto set is constructed, the best solution from the remaining options is determined, and then the methods from the first group are applied.

Multi-criteria tasks are even more difficult to solve if the system interacts with the environment, since the solution will depend on uncontrolled factors. These factors increase the ambiguity of choosing the best solution. There are two strategies for choosing the best option that allow you to make informed decisions.

The guaranteed result method is used in an unpredictable or hostile environment. Since we do not know the behavior of the environment, it is not possible to determine the best solution. However, in this case, you can define a guaranteed solution that is valid for any state of the environment.
In conclusion, we will consider an approach based on the principle of equilibrium. This method is particularly used in game theory, since the Nash principle reduces the number of alternatives when it comes to a collective decision that is made by all participants, but each proceeds from their own interests.

4 Methods for finding solutions

People are faced with situations when the solution to the problem is not available or is not unambiguous, for example, when the solution is difficult to identify explicitly, methods are used to find solutions. Let's look at each of the methods:

The full search strategy is used when there is not enough information about the task, or when the set of alternatives is relatively small (up to $10^3$ elements – if the account is performed manually, and up to $10^9$ - if on a computer).

Implicit iteration contains several methods, such as the minimum cost method, branch and boundary method, simplex method, dynamic programming, $(\alpha - \beta)$ method, and others. In each of these methods, at each step, the search is performed not for the entire task space, but only for a fragment of it.

Heuristic methods model qualitative and situational ways of solving problems. Step-by-step procedures are created in which a certain number of steps provide a satisfactory solution to the problem, since possible options are reduced during the search for solutions. Heuristic methods are usually used to solve poorly structured, poorly formalized problems that cannot be described using a numerical model. In addition, such tasks have such characteristics as inaccuracy, incompleteness, ambiguity and ambiguity of information. It is also advisable to use this method in extreme or unknown situations.

Heuristic search:
* Includes system analysis of the issue;
* Identifies external and internal constraints that affect the result;
* Analyzes how possible it is to get the result by simple means;
* Identifies features, limitations, and "bottlenecks" that require additional resources to be used, or Vice versa – reducing and;
* Simulates tasks and possible situations to get the best solution.

The heuristic approach uses General approaches applied by a person when solving problems, when generating solutions, comparing them and choosing the optimal solution.

Analogy method.

The simplification method is used if you can't find solutions using the analogy method. In this case, some conditions and restrictions are dropped, thereby increasing the" symmetry " of the problem.

The aggregation method includes the characteristics of the previous method, applying a higher-level conceptual apparatus. This approach is used to solve the so-called ill-posed problems.

Methods for finding solutions can be divided into 3 groups.

Group I: state search strategy.

The initial information is given in the form of various situations, describing the state of the system. It is necessary to find such a way, which will lead from the initial state to one of the target States.

This group includes the following methods:
- Search method "in width" and "in depth»;
- $(\alpha - \beta)$ method;
- The method of the shortest path;
- The method of branches and borders;
- Gradient methods (minimum cost, dynamic programming, vector optimization)

Group II: task search strategy.

Information is given in the form of a problem $s$ and the set of elements of its solution $s_{ij}$, where $j$ – is the number of levels of the solution; $i$ – is the number of elements on the $j$-th level.

In this case, the original problem must be reduced to simpler ones, so that the final result is elementary problems.

This group includes the following methods:
- key operators;
- General problem solver, etc.

Group III: methods that use logical inference
The information describes the state of a formal system that includes the alphabet, axioms, and inference rules. In this case, it is necessary to check whether the end state $s_k$ or the initial state $s_0$ can be reached.

This group includes the following methods:
- deductive method;
- production method, etc.

At the moment, there are various modifications to the methods of finding solutions to improve their effectiveness, as well as targeted search strategies of a General nature that simulate the process of reasoning people.

5 Models of decision-making

The decision-making model implies an evaluation procedure that helps you make the right choice between options. But making decisions can be difficult, because there are many conflicting criteria for choosing.

It is possible to classify decision-making models by various criteria.

1. by the number of goals (ways to describe an object):
   * single-purpose models (Profit-Cost model; Efficiency-Cost model)
   * multi-purpose models (multidimensional utility functions and a priori models for comparing variants)
2. depending on the problem situation (scope), there are models:
   * compromises'
   * optimization
   * diagnostic and so on

Let's look at each of the models in more detail.

Single-purpose models. In this type of model, each alternative is evaluated using a single criterion, which is why this type of model is sometimes referred to as single-criteria models in the literature. As mentioned above, there are basically two models used in single-purpose models: "Profit-Cost" and "Efficiency – Cost".

The Profit-Cost model uses one economic criterion – the cost coefficient (let's call it "C"). This coefficient expresses the following relations:

$$c(x) = \sum_{j=1}^{n} a_j(x) - \sum_{k=1}^{m} b_k(x),$$

where $\sum_{j=1}^{n} a_j(x)$ – this is the amount that takes into account the total profit for the option under consideration for each element of positive impact;

$\sum_{k=1}^{m} b_k(x)$ - this is the amount that takes into account the total profit for the option under consideration for each element of negative impact.

Then the cost coefficient will be equal to the difference between profit and cost. Sometimes the cost factor is defined as the ratio of profit to cost, i.e.:

$$c(x) = \frac{\sum_{j=1}^{n} a_j(x)}{\sum_{k=1}^{m} b_k(x)}$$

After determining the cost coefficients for each of the solution alternatives, choose the highest—this will be the best solution.

The Model Of "Efficiency – Cost". This model compares the degree of achievement of goals and costs.

The model can be presented in the following form:

$$I(x) = \frac{(a(x) - a_0)}{b(x)},$$

where $I(x)$ – this is the cost effectiveness index for the solution option;

$a(x)$ – result after implementing the option $(x)$

$a_0$ – degree of achievement of the goal (result) before the implementation of the option $(x)$

$b(x)$ – costs of implementing the option $(x)$

After making calculations for each of the alternatives, the option with the highest efficiency index is selected.

You can ensure the accuracy and reliability of these two calculations by taking into account as many components of profit and cost (the first model) and effect and cost (the second model) as possible. The models considered are simplified and they are applied to various types of selection tasks. Moreover, the "Profit-Cost" and "Efficiency-Cost" models are used to find a compromise if competing factors influence the decision.

Example 1.
Let’s say there is a factory that produces certain products. It is necessary to find the optimal amount of costs for product control. Let the measure of effectiveness be the quality (accuracy) of control. Then, if the control accuracy is high, the direct costs will increase. If the quality of control is reduced, indirect costs will increase. The best option is to combine the two options and get the best one.

Next, we need to consider multi-criteria models.

The compromise model. This model describes how to evaluate substitutions in tools for complex systems that contain interdependent subsystems. There are two types of models:

Models that describe trade-offs between mutually substituting systems

Models describing trade-offs between mutually complementary systems. In this case, one of them will either strengthen or weaken the other.

Optimization model. These types of models include the following:
* differential calculus;
* Lagrange’s method;
* linear programming methods;
* target programming;
* dynamic programming;
* quadratic and nonlinear programming;
* etc.

Diagnostic models. Under this model, we understand systematic Troubleshooting if the normal operation of the system is disrupted. This model uses the method of pattern recognition, classification, and taxonomy.

In multi-purpose models, each alternative is evaluated by a variety of different criteria. They also include: multidimensional utility functions, multidimensional scaling models, and the method of analyzing hierarchies.

They mainly use additive and multiplicative multidimensional utility functions. A utility function (let’s denote it U) is a scalar function that establishes an order relation on a set of variants. In an additive model, the utility function looks like this:

\[ U_a(x) = \sum_{i=1}^{n} p_i \bar{U}_i(x) \]

where \( U_a(x) \) – is the utility function option \( (x) \); \( p_i \) – is the weight of criterion; \( \bar{U}_i(x) \) – assessment of the usefulness of the option \( (x) \) on the criterion \( i \).

In a multiplicative model, the utility function looks like this:

\[ U_m(x) = \prod_{i=1}^{n} p_i \bar{U}_i(x) \]

Usually, in order to get an estimate, use the expert path. However, sometimes estimates can also be set analytically using a suitable approximating function.

6 Characteristics of probabilistic methods of information processing

Any type of information processing has certain specifics and a number of disadvantages. Starting from the definition of the conceptual side of probabilistic methods of information processing, it is worth noting that the probability is a certain ratio of the studied options to the total number of outcomes. This dependence can be represented by the following formula:

\[ P(A) = \frac{M}{n} \]

where \( P(A) \) – is the probabilities being investigated,
\( M \) – it is a characteristic of favorable outcomes,
\( n \) – is the total number of outcomes of the situation under study.

In addition to the above definition, there is another, statistical, which is based on the eponymous definition of probability, according to which the probability is the relative frequency of occurrence of an event when the conditions of the experiment are repeated many times.
7 Bayesian approach

Almost all statistical problems have a common property: before a certain set of information is obtained, a number of probabilistic models are considered as potentially applicable to the studied situation. One of the ways to "revise" the relative acceptability of probabilistic models is the Bayesian approach, on the basis of which the well-known Bayes theorem arose. Its formula, which makes it possible to evaluate events that exist objectively and are based on everyday experience, plays an important role in modern mathematical statistics and the theory of probabilistic methods of information processing.

The idea of the Bayesian approach is to move from a priori knowledge (or more precisely, ignorance) to a posteriori (corrected) with the observed phenomena in mind. The Bayesian approach allows you to get the most accurate results with time savings, track the dynamics of the model, get a variety of representations of the model under study, clearly and clearly reflect the results of the study, and classical probability theory and mathematical statistics help justify and simplify the application of these principles.

The Bayes formula looks like this:

\[ P(x|y) = \frac{P(y|x) \cdot P(x)}{P(y)} \]

Bayes methods have been applied, for example, to sorting resumes of new applicants for various job categories based on their similarity to already classified or grouped resumes. Another interesting use of them is to sort the news feeds that are most interesting to a person—this is done by selecting new articles based on previous ones that the person has rated as suitable and relevant. The next step is to sort spam and other inappropriate emails based on Bayesian similarity to previously classified messages.

The Bayesian approach is also effective and useful when probabilistic data is often not available, which makes the use of many traditional approaches incorrect.

In addition, in probability theory, there are laws of distribution of random variables, which are also considered probabilistic methods of information processing. Among these methods, the following distributions are distinguished: normal distribution, uniform distribution, exponential distribution, and Poisson distribution.

The Bayes formula is well known from the course in probability theory and mathematical statistics and allows you to recalculate the conditional probability (or density) of \( P(A|B) \) through \( P(B|A) \).

Currently, Bayesian methods have become quite widespread and are actively used in a variety of fields of knowledge. However, unfortunately, not many people have an idea of what it is and why it is necessary. One of the reasons is the lack of a large amount of literature in Russian. So here I will try to set out their principles as simply as I can, starting with the basics (I apologize if this seems too simple to some).

The Bayesian approach helps to optimize the solution of many problems in modern system analysis and system modeling. This explains the interest in this topic and the relevance of the work. We can conclude that the Bayesian approach helps us solve many economic problems and build economic models in the most optimal way. It is worth noting that such decisions are fairly accurate, which makes them acceptable when justifying a particular economic decision. However, it is worth noting that the construction of economic models and system analysis based on the Bayesian approach are not without disadvantages. Among them, we can highlight the complexity of interpreting the results obtained, the obligation to comply with certain conditions that may not be observed in the real economic activity of subjects, moreover, this approach is labor-intensive and resource-intensive. Nevertheless, in the current state of digital and it technologies, the Bayesian approach has prospects for development and improvement, which will allow it to be more widely applied in the creation of economic systems and system analysis.

8 Conclusion

Summing up, it is worth noting once again that the modeling process is quite time-consuming. It is very difficult to choose the optimal solution among many alternatives. First, you need to determine the correct model for which you will be able to find the best solution. There are a huge variety of decision-making models. According to the methods of describing an object, there are single-purpose models (they include the "Profit-Cost" model and the "Efficiency-Cost" model) and multi-purpose models (these are multidimensional utility functions and a priori models for comparing options). Also, depending on the problem situation, there are compromise models, optimization, diagnostic, etc. each model has been considered in More detail in this paper.
There are also methods for solving multi-criteria selection problems, which can be divided into two groups. The first group includes the convolution method, the main criterion, thresholds, and distance. The methods contained in this group reduce the number of criteria by introducing additional assumptions. The second group is also aimed at reducing alternatives, but excluding obviously bad alternatives. This can be achieved using the Pareto set construction method, the guaranteed result method, and the equilibrium method.

The scientific literature contains many models of decision-making. In order to study and describe each of them in detail, it will take a large amount of time. This paper presents only a few of them with a brief description.

Thus, the methods considered are only a certain part of all existing probabilistic methods of information processing. All these methods, described and not described, are currently widely used in all spheres of life, from the economic activities of economic entities to the use of probabilistic methods of information processing for control for medical purposes. However, as described in the paper, each described method has both advantages and disadvantages, which, for example, as in the case of adaptive forecasting methods, led to the emergence of a new direction – Adaptive Econometrics.

In addition, probabilistic methods of information processing are used to analyze markets for raw materials and goods in cases of large batches of products, to analyze the operation of information systems and check the logic and regularity of their operation, to reduce software errors of various technical devices, as well as to conduct social and psychological research.

It is worth noting that the methods of probabilistic processing of information can also be used during the audit. The methods defined for this area identify those areas of the data set that are supposed to contain deviations from the norm based on probability, which will help significantly speed up the audit time, but at the same time, due to the huge amounts of data being calculated, the values of some probabilities may be approximate and not accurate. However, with the development of computer technology, these calculations will be able to be more accurate, which will significantly reduce the cost of conducting an audit and improve its quality.

References
[1] Aksenov K A 2018 Modeling and decision-making in organizational and technical systems (Moscow: FLINT) 104 (in Russian)
[2] Arsenyev Yu N, Shelobaev S I and Davydova T Yu 2003 Decision-Making. Integrated intelligent systems (Moscow: UNIGI-DANA) (in Russian)
[3] Berstein L S, Karelin V P and Tselyov A N 1999 Models and methods of decision-making in integrated intelligent systems (Rostov n/A: RSU Publishing house) (in Russian)
[4] Demidova L A, Kirakovsky V V and Pylkin A N 2015 Decision-making in conditions of uncertainty (Moscow: Hotline-Telecom) 283 (in Russian)
[5] Zavgorodny V N 2014 Modeling of decision-making processes in complex organizational and technical systems Software products and systems 1 (105) (in Russian)
[6] Karelin V P 1995 Theory and means of decision-making support in organizational and technological systems (Taganrog: TRTH) (in Russian)
[7] Karelin V P 2009 Methods and means of information and analytical support for decision-making in organizational systems Vestnik Tiuie 2 (10) (in Russian)
[8] Kutergin V A 2009 Constructive processes and artificial objects Electronic scientific publication: "Sustainable innovation development: design and management" 2 (in Russian)
[9] Kutergin V A 2012 Goal and decision-making in organizational systems Electronic scientific publication: "Sustainable innovation development: design and management" 8 (3) (16) (in Russian)
[10] Rizvanov D A and Yusupova N I 2017 Fundamentals of decision support in resource management in complex systems using intelligent technologies Modern science-intensive technologies 1 69-73: http://www.top-technologies.ru/ru/article/view?id=36558 (in Russian)
[11] Rosenberg I N and Starostina T A 2006 Solution of placement problems with fuzzy data using geoinformation systems (M. Nauchny Mir) 208 (in Russian)