Model design as a new socio-cyber-physical systems creating mechanism

E.V. Yurkevich and L. N. Kryukova

V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

Email: Lidkryukova@yandex.ru

Abstract. The purpose of this work is to consider the methodology of designing intellectual models of cyber-physical social systems that take into account the cybernetic properties of the hardware, software and the human factor specifics of the system developers and users. To ensure customizable solutions, model design process is based on systems analysis. This methodology can be implemented by optimizing the objective tree, which determines the development vector of the model being designed based on their importance; the information tree, which determines the structure of the knowledge base optimized by the cost of resources used; the task tree, which allows optimizing the development of the cyber-physical social systems based on the results of analyzing the criteria of the two trees. Intellectual properties of the model being designed are determined by automating the management of its development vector in accordance with the dynamics of the utility of its specific goals. Due to practical limitations on the amount of resources available, the task is to optimize the plan for creating a cyber-physical social system that maximizes the solution value with limited cost. The choice between more precise algorithms that are not applicable given fuzzy limitations caused by human factor allows solving the problems faster than when using deterministic or heuristic methods.

Introduction

The establishment of post-industrial economy greatly increases the role of information in ensuring the efficiency of developing radically new technological structures [1]. Introducing elements of artificial intelligence into the process of developing such solutions is becoming particularly relevant.

The traditional approaches to creating technical solutions are mainly based on the principle of imperative data formats provided by the developers. Today the focus is on making sure that the technology solutions meet the requirements of each and every customer. In this regard, it is important to provide a prompt assessment of the correlation between the results offered by the developers and the expectations of the customers. When creating such assessment, real-time accounting of human factor must become an important link in the feedback chain, ensuring dynamic adjustment of the properties of the new systems.

Today practically every technology system can be considered as a human-computer one. Therefore, the problem is to produce an algorithmic presentation combining the possibility of formally approaching engineering tasks using computer hardware and expert assessments. Reviewing the mechanisms of developing solutions that can maximize the design efficiency of the systems being set up is what determines the relevance of addressing the issue by introducing elements of hybrid intellect.

Hybrid intellect is understood as a combination of artificial intelligence and interpretation of decisions made by specialists. In this work hybrid intellect will be implemented by creating cyber-
physical social systems (CPSS) [2], which are seen as a development in designing embedded systems [3]. The peculiarity of CPSS systems lies in the fact that they compute the dynamic properties of the object being designed in the context of restrictions present in the form of qualitative evaluation of the dynamic impact of destabilizing factors.

This peculiarity determines the problem statement in the transition from the traditional modeling of a development item to designing its intellectual model. This approach to design is based on the concept of a model developed in accordance with the object design assignment. The concept implementation is structured in the form of an objective tree in a system of functional blocks model, built on the basis of an analysis of the demand for the system. The combination of models can be built using a knowledge base required to develop the system. In the next stages in the model design, the authors are to declare the set of objectives addressed by the system and to assess the reliability and relevance of the model.

Mechanisms for creating a socio-cyber-physical system intelligent model

From the methodological point of view, CPSS model design shall be based on the systems analysis ideology [4]. In this work, that ideology will be produced by optimizing the structure of three trees: the objective tree "A", which determines the development vector of the model; the information tree "B", which determines the properties of the knowledge base used for working with the model; and the task tree "C", which presents the ways to meet the objectives from the set "A" via abilities determined by the set "B".

To create the tree "A", we will proceed from the most generic formulation of the objective provided by the customer. This original objective is decomposed into sub-objectives, sub-sub-objectives, etc. The process ends when the formulation of objective in each element in the tree represents the requirement for a specific action. Thus, we arrive at subsets $A \ni \{a_1, a_2, \ldots, a_i, \ldots, a_n\}$, $\forall a_i \ni \{a_{i1}, a_{i2}, \ldots, a_{iy}, \ldots, a_{in}\}$, $\forall a_y \ni \{a_{y1}, a_{y2}, \ldots, a_{yk}, \ldots, a_{yn}\}$, etc.

The next objective is to assign a utility score to each element in the tree A. We consider the implementation of tree A as a Markov chain with yield (MCY). In this case, meeting the target of each element in the tree can be presented as an event characterized with a certain effect (yield). In terms of this work, the evaluation of said yield will be understood as the utility of the respective event occurring.

To assess the effect on reliability from the transitions between one subset of sub-tasks being addressed and another, the following aggregated reliability indicators are suggested:

- the average time to hitting the given subset of sub-tasks;
- the average number of transitions from $i^{th}$ level to $j^{th}$ level before hitting the given subset;
- the average number of transitions in the interval of time $(0, T)$ before hitting the given subset;
- the probability of transition from the subset of sub-tasks $\{a_i\}$ to subset of sub-tasks $\{a_j\}$ in the interval $(0, T)$;
- the probability of the task being addressed at the level of $\{a_i\}$ at the moment $T$.

The expected yield for reliability indicator is to be determined using the following equation:

$$E = M\{f(x(t))\},$$

where $M\{*\}$ is the expected utility symbol; $x(t)$ is the process of transition to a lower level in the subset of tasks being addressed; $f(\cdot)$ is the integral-type function declared at the process trajectories and determining the form of reliability and efficiency indicator, i.e. yield. Let’s denote the yield as $\omega_{ij}$, if at the moment $t$ the process state $x(t) = x(i)$ (the tasks are solved at the $i^{th}$ level), or $\omega_{ij}$, if $t$ is the moment of transition from the $i^{th}$ level to $j^{th}$ level ($i \neq j$). The basic expression for MCY now looks as follows:
\[
\frac{dM_i(t)}{dt} = \sum_{j=1}^{n} \lambda_{ij} M_j(t) + \omega_i + \sum_{j \neq i} \lambda_{ij} \omega_j, \quad M_i(t) = 0.
\]

Given the integral represented by functional \( \Phi(*) \), we suggest recording the reliability indicators via Stieltjes integral as follows:

\[
E = M\left[ \int_0^T d\Phi(x(t)) \right],
\]

where \( d\Phi(x(t)) \) is the differential of integrating function.

In this work we assume that the approximated solution is based on the assumption that achieving the goal \( A \) to solve the tasks requires some random time \( \tau \), with the cumulative distribution \( F(\tau) \), with the universal mean \( T \). In this case, the indicator estimates will look as averaged assessments for each time period \( \tau \in [0, \infty] \)

\[
M(\bar{T}) = \int_0^\infty M(\tau) dF(\tau).
\]

The function \( F(\tau) \) is chosen as Gamma distribution \( G(k, \beta) \) with universal mean \( m = k / \beta = T \).

From here we can conclude that introducing random time distributed according to the law \( G(k, \beta) \) results in a Markov model, resulting in a \( k \)-fold expansion of the space of tasks being addressed.

This produces a set of alternative solutions, from which one must be chosen. To support this choice, \( OLAP \) (On Line Analytical Processing) technology for comprehensive multidimensional data analysis is provided. It is based on \( FASMI \) (Fast Analysis of Shared Multidimensional Information), which uses the fuzzy logic mechanism and possibility theory to perform multidimensional data analysis.

We create the proposed mechanism in accordance with the following requirements:

- the need to provide analysis results in a given time, even at the cost of less detailed analysis;
- the possibility to perform logical and statistical analysis of the properties of each block of tasks in dynamics, as needed to achieve the sub-goals being part of tree \( A \);
- multiple-user access to data with support for the respective locking mechanisms and access authentication options;
- multi-dimensional conceptual representation of the data, including support for multiple hierarchies;
- ability to address the information regardless of its size and storage location.

The unique feature of this mechanism is the use of Markov chain model with yield, which allows optimizing the solutions by the criterion of maximum probability of solving the sub-tasks from tree \( A \) without failures.

Since there is no universal measure in the existing practice that would make physical sense and allow measuring the effect of achieving the goals, we introduce an artificial measure. It will be determined via the utility of alternative (the utility of yield from the respective event \( a \)). Let us define it as the technological solution value ratio \( (\omega) \). Numerically, \( \omega_i \) is determined as the difference between the probability of reaching the goal of system creating in presence of element \( a_i \) and the probability of reaching the same goal in absence thereof. If calculating the said probability is not possible, utility is estimated by expert opinion.

When the utility is determined using expert opinion, the following key phases are suggested:

- ordering the multiple outcomes of meeting the objective via the elements of the \( A \) multitude by preference \( (a_1 \rightarrow a_2 \rightarrow \ldots \rightarrow a_n) \);
identifying the utility of each outcome \( F(a_i) \), validating the resulting estimates for non-contradiction by comparing the preference scores of different outcomes;

- eliminating contradictions in estimates by adjusting or reordering the outcomes, or utility scores, or both.

Approximation of the utility function is often used when addressing practical tasks. The technology of such approximation consists of the following. When reviewing the outcomes in addressing a task via a specific element of the multitude \( A \), characteristic points are located, for example, corresponding to the extremities of the utility function. The unknown values in-between are filled with some known dependency. The approximation type is chosen based on the information available or qualitative considerations of the outcome utility. In practice, multiple-stage and other complex approximations of the utility function can be used. The simplest forms of approximation are single-stage, cosine and triangular representations of the utility functions [5].

Important provision is that the scale of utility scores must be the same for the entire set of \( \{ a_1, a_2, \ldots, a_n \} \). Utility scores shall not change in subsequent stages of the model design.

To conclude the work with tree “\( A \)” , the branches of maximum value are determined using the bottom-to-top dynamic programming mechanism.

In the next step of CPSS model design, a knowledge base is built based on the tree “\( B \)”, which repeats the architecture of “\( A \)”. We use “\( B \)” to assess the possibility of using the resources available and required. Since these resources are heterogeneous, we can introduce the cost \( (w_i) \) as a common equivalent of the value of resources used in reaching goal \( A \).

Let us assume that creating a knowledge base with the help of tree “\( B \)” requires prioritizing the resources in terms of priority, based on the values of multitude \( \{w_i\} \). In this regard, the mechanism for creating tree “\( B \)” can be considered as a dialog-based technology that considers human factor during the interaction between specialists in CPSS.

In essence, this form of model design acts as a special form of information support organization, which contains an intellectual product based on the cognitive activities of the knowledge base developers. For example, efficiency of software and hardware suites is affected by interactive and multimedia capabilities motivating the developers to use the information resources more efficiently [4].

Technologically, the use of information resources in model design implies structuring the information contained in feedback provided to its developer. To maintain this structuring, it is suggested to use intellectual data analysis (IDA) methods. The most common ones are: Knowledge Discovery, Intelligent Analysis Data.

In our case, intellectual data analysis is seen as the possibility of using the information processing algorithms to identify and assess the trends, patterns, correlations and prospects in the development of the object being modeled. Using intellectual analysis mechanisms in model design allows increasing the volume of information processed and handling more complex tasks; increasing the demands to the analysis results; minimizing the time restrictions on the situation analysis and model development; improving the interface elements that made them accessible for use by specialists.

When creating tree “\( B \)”, text mining is considered as a private case of data mining, i.e., a non-trivial process for identifying new, potentially useful, correct and interpretable patterns in the data. The specific of objectives in \( A \) requires processing a lot of information, not just as text and numbers, but also in the form of speech or video footage. One of the central problems in this case is clustering, as a process of grouping information based on linguistic and mathematical methods.

Algorithmic presentation of the links between content elements allows creating semantic networks, which helps the developers to identify the descriptors and use them for navigation. Visualization is a key element in visual presentation of unstructured textual information, as a method of revealing its semantic content and implementing a navigation mechanism used in the analysis of its classes.

An important property of intellectual data mining tools is the reliability of their interface to the model designer. Several blocks can be used when developing a multi-stage information processing model, and each one must work in its own applied area (areas) and use its own language. It is
advisable to expand the use of the following formats in improving the solution development process: business games, role play, discussions. All these methods speed up the acquisition of new experiences.

Tree “C” must facilitate the formation of a multitude of tasks maximizing the efficient usage of available resources. With the help of tree “C”, the tasks used to achieve the goals set by tree “A” can be seen as elements in the system with their cost ($w_i$), derived by analyzing the tree “B”, and their value ($\omega_i$) derived from analysis of the tree “A”.

Limited resource availability is a constraint in almost every situation. In this case, the design of a CPSS model is reduced to writing an algorithm for addressing the tasks with maximum value while maintaining the restriction on the total costs of such solution.

In formal terms, this algorithm can be classified as a “knapsack problem”. Suppose reaching each objective in $a_i$ ($i = 1, 2, \ldots, n$) requires solving the respective task. This means there are $n$ tasks. For each $i^{th}$ task, we know the value of its solution $\omega_i > 0$ and the cost of resources required $w_i > 0$. The restriction on the total cost of addressing the full range of tasks is defined as the available amount of resources $W$. A series of tasks needs to be developed, described by the multitude $\{\omega_i, i = 1, 2, \ldots, n\}$, meeting the condition for maximizing the value of technological solution ($\omega$), given the limitations

$$\sum_{i=1}^{n} w_i x_i \leq W \quad \text{and} \quad x_i \in \{0, 1\},$$

where $x_i$ is the Kronecker’s delta, which defines acceptance or rejection of $i^{th}$ task, $\forall i$.

The proposed formulation allows generalizations depending on the conditions pitched against the total resources available and the specifics of the quantity of the tasks. For example, no more than one resource value can be used for addressing each task, or generally no more than a given value of resources for each task; another common type of restrictions allows using a variable number of resource values.

Due to a large volume of data being used, multiple-choice tasks are common, i.e., the tasks are divided into groups and only one task must be chosen from each group; also, several groups of resources can be available, where each group has a fixed size, and each task can be delegated to any group. The user has to divide the tasks into subsets so that the total use of any resource does not exceed the maximum available amount of that resource, while maximizing the total value of the solutions. Intellectual nature of that model is determined by preparing its development vector in accordance with the utility of reaching the objectives in each element of the tree “A”.

In practical applications, one has to choose between exact algorithms, which work fast but are not applicable given “greater freedom” in restrictions, and approximations that work fast but do not guarantee optimal solution to the task, while offering a “good enough” solution faster than any other deterministic or heuristic methods. For example, the proximity to optimal solution is determined by the fitness function. This can take the form of a total value of the solutions, provided the total value does not exceed the quantity of resources available. To ensure reliability of CPSS operation, group decision-making is often used. We suggest coordinating the two factors by alternating between two phases of the negotiation process: individual decision-making phase and decision coordination phase.

The presence of an intellectual component allows introducing elements of forecasting into the description of parameter dynamics of the system being modeled. Traditional modeling methods work well when the process is stationary, i.e., its properties change slowly over time. The suggested model design mechanisms allow introducing adaptivity in the CPSS.

**Conclusion**

In order to create methodological support of innovations in the design of CPSS intellectual models, which take into account both cybernetic properties of the software and hardware and the human factor of their developers and users, we suggest considering a series of aspects in using systemic approach to create complex man-machine systems. The results of testing conducted include: a design mechanism and the actual model of a complex system featuring components of expert interaction at all stages of
its lifecycle, including the mechanism to set up adequate space for interaction between the experts, and methodological provisions to develop a knowledge base for creating functional modules.

In this regard, we suggest introducing hybrid intellect as a feedback mechanism to adjust the dynamic properties of the systems being designed, which is a tool to adapt them in real time to the influence of human factor. Hybrid intellect is understood as a combination of artificial intelligence and interpretation of decisions made by specialists. This approach allows minimizing the errors in making decisions on stabilizing the operation of the system being considered.

As a vector criterion of the value of a technological solution, we use an assessment of the modules’ operational functional readiness to perform the technological functions assigned in normal mode, under fixed influence of external factors over a fixed period. On the other hand, as a solution to one of the optimization tasks in this class, we suggest using the methodology of considering the dynamics of aggregate value of these solutions, with the costs not to exceed the fixed limit, using the resources required. In practical consideration, the choice is often between exact algorithms, which are not applicable given “greater freedom” in restrictions, and approximations that work fast but do not guarantee optimal solution to the task.

Using the proposed models can offer flexibility in creating a CPSS that allows optimizing the development by the criteria of maximum value of its elements. Hybrid intellect mechanisms allow diagnosing changes in the CPSS at an early stage in the development of its reaction to external factors. As a result, we can provide a warning signal on time and prevent the failure of the entire system or its individual elements.

References
[1] Yurkevich E V, Krivopalov D M and Kryukov L N 2019 Algorithmic features of reliability design of onboard spacecraft systems South siberian scientific bulletin 2 18–27
[2] Yurkevich E V, Kalygina A D and Kryukova L N 2018 Cyber-social systems as a tool for optimizing strategic management at an industrial enterprise Proc. of the 11-th Int. Conf. "Management of Large-Scale Systems Development" vol 1 (Moscow: IPU RAN) pp 236-8
[3] Yurkevich, E V, Politikin A G, Promyslov V G, Stepanovskaya I A and Kryukova L N 2018 Virtualizing the Process Engineering for Managing Spacecraft Resilience Against Electrophysical Effects (Moscow: IPU RAN) p 168
[4] Yurkevich E V 2007 Introduction to the Theory of Information Systems. 2nd revision, updated (Moscow: IDT Group LLC) p 272
[5] Burkov V N and Novikov D A 1996 Introduction to the Theory of Active Systems (Moscow: IPU RAS) p 124