Manko G.I., Chistokletov E.P.

USE OF INFORMATION UNCERTAINTY IN IDENTIFICATION TASKS

Ukrainian State University of Chemical Technology, Dnipro, Ukraine

Informational methods for analyzing and managing systems under uncertainty are studied. The expediency of use of information uncertainty in tasks of identification of control objects and synthesis of regulatory systems is justified. For the numerical evaluation of information uncertainty, an amount of disinformation is used as Bongard's negative useful information. Such information uncertainty can serve as a criterion for the adequacy of a mathematical model of a control object. Information assessment of the modeling accuracy is applicable to any method of identification and allows a researcher to compare the accuracy of models that differ from each other by the method of obtaining (analytical or experimental), nature (deterministic or stochastic models), specific implementation (physical or mathematical models). In the course of parametric identification, an optimization problem of finding a minimum of information uncertainty in the parameter space of a mathematical model can be solved. The information uncertainty criterion provides verification of the statistical hypothesis about the adequacy of a particular model. If the criterion value exceeds a certain critical value, the adequacy hypothesis must be rejected. To calculate the critical values of the information adequacy criterion, a statistical experiment was performed. Using the Monte Carlo method, the probability distribution of the information criterion was investigated. A sufficiently smooth empirical criterion distribution function was constructed. The distribution of the information criterion has a pronounced asymmetry and a small positive kurtosis. It is revealed that this distribution is best approximated by the Generalized Extreme Value Distribution law. The critical value can be defined as a quantile of the level of 0.01 or 0.05 for this distribution.

Keywords: adequacy, computer modeling, distribution law, identification, information criterion, Monte Carlo method, control object, statistical hypothesis, uncertainty.

Statement of the problem

As known, the synthesis of control systems begins with the process of obtaining a mathematical model of a controllable object. Meanwhile, developers are encountering the problem of incomplete correspondence of models with an identified object. Thus, during system synthesis, the models remain uncertain.

The sources of model uncertainty are:
− incomplete knowledge of the physical and chemical processes occurring in the object;
− assumptions and simplifications, taken during a model formation;
− noises, interferences and measurement errors taking place in experimental researches of objects;
− parameters changes in the object over time.

Another problem is the quality assessment of the obtained model. To solve this problem, additional researches are carried out, listed in the report [1].

Verification is the process of determining that a model implementation accurately represents the developer's conceptual description of the model and its solution.

Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of a model.

Both verification and validation are processes that accumulate evidence of a model's correctness or accuracy for a specific scenario; thus, they cannot prove that a model is correct and accurate for all possible scenarios, but, rather, it can provide evidence that a model is sufficiently accurate for its intended use.

Adequacy is the decision that the model fidelity is sufficient for its intended use.

Fidelity is the difference between simulation and experimental outcomes.

Analysis of recent research and publications

Conceptual model validity, model verification, operational validity, and data validity were discussed by Robert G. Sargent [2]. Four different approaches to determining model validity are described, a graphical paradigm that relates verification and validation to the model development process is presented, and various validation techniques are defined.

In the article [3], the analytical expressions are obtained to determine the dynamic characteristics of controllable objects by the parameters of closed systems transient function. However, the automation of such process is extremely difficult.

In the work [4], the peculiarities of the controllable object identification by measures of Toolbox PID Tuner in the MATLAB environment are considered. There is stated that automatic identification does not always give satisfactory results.

A.N. Trunov [5] gives an overview of the works devoted to the general patterns of selection criteria search for settling the adequacy. There is an indication of a single expression of adequacy assessment absence, which makes possible taking into account several factors of influence and suggesting expressions for their calculation. It is proposed to use a criterion that is inversely proportional to the standard deviation of a model from experimental values.

The dispersion methods for adequacy assessment are widely known. E.P. Chernogorov [6] states that the system model allows a researcher to model some general population. The experiment provides a separate implementation of this general population. The model also provides an implementation of a general population, reproducing the experimental conditions. The adequacy assessment lies in checking statistical hypotheses about the belonging of these realizations to the same general population.

It should be noted that the dispersion criteria of the adequacy, like Fisher criterion [7], are applicable only in the case of a normal distribution of model variables, which doesn't always take place in practice.

Objective

The purpose of this work is to substantiate the criterion of information uncertainty application to assess the adequacy of models obtained in the course of identifying controllable objects.

Informational uncertainty

The philosophical category of "uncertainty" is in the focus of many scientists' attention. The authors of the article [8] consider uncertainty as a concept expressing a certain phenomenon associated with a subject-object relation. This makes it possible to distinguish four aspects of the study of the uncertainty concept. Uncertainty is a concept that can display: the properties of an object, the conditions of interaction between an object and a subject, the features of subjective perception, the features of interaction of new knowledge with current knowledge. So uncertainty is a category that reflects the absence of any (material or ideal) orderliness.

For control systems that are the subject of our research, in work [9], questions of quantitative estimation of orderliness and related concepts of complexity, organization, and entropy are considered. According to [9], ranking can be viewed as a characteristic of order — the absence of a difference in relation to something taken as the standard of order.

We consider a multidimensional parallelepiped in the parameter space of a system used in interval systems theory as such standard. We will consider such n-dimensional parallelepiped as a target area, a vector $\dot{X}_{et}$, directed from the origin of coordinates to the target area as a standard of order, and a system in such a state as ranked. The difference measure of the system parameters real vector $\dot{X} = \{x_1, x_2, ..., x_n\}$ from order standard X_{et} characterizes the disorder $\overline{Y}$.

Uncertainty takes different meanings in different situations and at different times. If m situations are possible, the probabilities of which are $p_1, p_2, ..., p_m$, then the ensemble gives a complete characteristic of disorder at each moment of time

$$\begin{pmatrix} \overline{Y}_1, \overline{Y}_2, ..., \overline{Y}_m \\ p_1, p_2, ..., p_m \end{pmatrix}. \tag{1}$$

A generalized ranking characteristic is the organization of the system, which, like ranking, is more convenient to estimate through the opposite value — disorder:

$$\overline{O} = \bigcup_{\beta}^{l} \alpha_{\beta} \bigcup_{i}^{d} s_i \bigcup_{j}^{m} p_j f(\overline{Y}), \tag{2}$$

where $\bigcup$ is the symbol of generalization of the disorder characteristics, respectively, for m situations, d elements of a system and l time intervals; f is a function by which weighting of the disorder is

performed by the existence factor of its occurring in relation to a certain functioning of the system indicator; p_j, s_i, α_β are the weights of j-th situation, i-th element and β-th time interval, respectively.

Taking formula (2) as a basis, we can derive a set of relations that can be used as an estimate of uncertainty. To simplify, let's take d=1 and l=1. We perform the operation of generalization by m situations with weighting by probabilities included in the ensemble (1):

$$\overline{O} = \sum_{j=1}^{m} p_j \log \Pi_{\overline{y}j}, \qquad (3)$$

where $\Pi_{\overline{y}j}$ is some numerical characteristic of differences $\dot{X}$ from X_{et}, called the disorder parameter.

Taking a j-th situation disorder probability $\Pi_{\overline{y}j} = p_j$ as a parameter, we obtain Shannon's informational entropy formula:

$$\overline{O} = H = -\sum_j p_j \log p_j. \qquad (4)$$

Mathematical model of a system under investigation due to inevitable inaccuracies gives distribution Q={q_j} instead of real distribution P={p_j}. Taking $\Pi_{\overline{y}j} = q_j$, we obtain Bongard's uncertainty formula [10]:

$$\overline{O} = N(p/q) = -\sum_j p_j \log q_j. \qquad (5)$$

According to Bongard, useful information stored in the distribution Q={q_j}, regarding the problem with a probability of response P={p_j}, is:

$$I_n = H - N(p/q) = \sum_j p_j \log \frac{q_j}{p_j}. \qquad (6)$$

Unlike the amount of information according to Shannon's classical theory, useful information can have both positive and negative meaning. Due to inaccuracies of a system's mathematical description, we should talk about the misinformation:

$$D = N(p/q) - H = \sum_j p_j \log \frac{p_j}{q_j}. \qquad (7)$$

Thus, the uncertainty estimate brought by the model is disorganization according to formula (3), in which $\Pi_{\overline{y}j} = p_j / q_j$. We will call the amount of misinformation (7) the information uncertainty.

Obviously, the largest value of misinformation occurs when H=0. Dividing the information uncertainty (7) by its maximum possible value, we obtain the relative information uncertainty:

$$\eta = \frac{D}{\sup D} = \frac{\sum_j p_j \log \dfrac{p_j}{q_j}}{-\sum_j p_j \log q_j} = 1 - \frac{\sum_j p_j \log p_j}{\sum_j p_j \log q_j}. \qquad (8)$$

Ways to use informational uncertainty in identification tasks

Information uncertainty in the form (7) can be used as a criterion for the adequacy of mathematical models. Accurate assessment of modeling accuracy is possible with any method that makes it possible to compare models with different characteristics (analytical or experimental), character (deterministic or stochastic models), implementation of realizations (physical or mathematical models). A mathematical model cannot be defined as minimal uncertainty.

Relative information uncertainty in the form (8) can be used as a criterion for the adequacy of models obtained by other methods. Moreover, there is both the possibility of comparing several models and choosing the most adequate, taking the least amount of misinformation, as well as examining the adequacy of a particular model's statistical hypothesis.

Here is the way to determine whether the model is adequate. Let us assume that there is some object, a state of which is characterized by the variable y (output value of the object). As a result of experimental researches, which used the method of a passive experiment, for instance, a statistical sampling of the object's values was obtained. The estimate of probabilities p_j is a ratio of a sample's number of v_j implementations, falling within a j-th separation interval of a range of y value, to the sample size n.

Let there also be a mathematical model of such an object, for which model value of the output variable $\hat{y}$ can be calculated, and by methods of information theory or mathematical statistics, a distribution Q={q_j} can be found.

We hypothesize that a mathematical model is adequate to an object. To examine the hypothesis, we need to calculate a criterion of information uncertainty value for model in the form (7), for which

we use statistics

$$S = \frac{1}{n} \sum_j v_j \log \frac{v_j}{nq_j}. \tag{9}$$

The resulting criterion value is compared with its critical value. If the calculated value is greater than critical, the adequacy hypothesis is rejected.

To calculate critical values, it is necessary to know the adequacy criterion distribution law. The distribution law can be determined by the statistical computer simulation method, described in [11].

The calculating power of modern computers allows us to solve many problems using a statistical experiment. Regulatory document P 50.1.037−2002 [12] recommends: if the probability distribution law is used to describe the sample $F(x,\theta)$ and estimates of its parameters are found, and to test a complex hypothesis H_0: $F(x) \in \{F(x,\theta), \theta \in Q\}$ a researcher doesn't know the statistical distribution of the corresponding criterion, it is rational to use the methodology of statistical laws computer analysis, which is well proven in modeling distributions of criteria statistics.

To do so, by the law $F(x,\theta)$ there should be modeled N statistical samples of the same volume n as a sample, for which an examination of the hypothesis H_0: $F(x) \in \{F(x,\theta), \theta \in Q\}$ is needed. Next, for each of the N samples, it is necessary to calculate the same parameters of the law estimates, and then S statistics of the studied criterion value. As a result, statistics values sampling $S_1, S_2, ..., S_N$ will be obtained with the distribution law $G(Sn^{\frac{1}{2}}H_0)$ for examined hypothesis H_0. For this sample, with a sufficiently large N, a quite smooth empirical distribution function of $G_N(Sn^{\frac{1}{2}}H_0)$ can be built, which can be directly used to conclude whether the hypothesis H_0 can be accepted. Simulated distribution deviations from the theoretical one at N=2000 usually have an order of H≈±0.015. If necessary, it is possible to build an approximate analytical model, approximating $G_N(Sn^{\frac{1}{2}}H_0)$, and to make afterward a decision about the examined hypothesis, based on this model.

The information criterion of the probability distribution values research was carried out by the Monte Carlo method. A sample of random v_j values was generated, for which the value of the S statistics was determined, and this was repeated 3000 times. This is a number greater than N=2000, recommended in [12]. For a sample of criterion values obtained this way, the distribution of criterion values histogram was built. The critical value was defined as 0.01-quantile.

The information criterion distribution has a pronounced asymmetry and a small positive kurtosis. Lognormal Distribution, Gamma Distribution, Weibull Distribution, Generalized Extreme Value Distribution can give such a picture. Compliance with all these distributions was verified using the information criterion. The calculation results showed that the Generalized Extreme Value Distribution provides the best approximation. However, as shown by additional researches, in the small values of information uncertainty criterion region, these distributions density curves differ from the normal distribution curve a little bit, and the quantile level of 0.05 for these distributions is almost equal. This gives grounds to state that at a trust level of 90%, it is permissible to use value tables of the normal distribution function to determine the critical values of the information adequacy criterion.

Conclusions

The use of the informative uncertainty criteria in the tasks of identification has been proposed in this paper. As a criterion for informational uncertainty, it was proposed to use the amount of misinformation brought by the mathematical model. The formulas for calculating such a criterion were given. The ways of using the criterion in the course of solving identification problems are shown.

REFERENCES

1. *Concepts* of Model Verification and Validation / Thacker B.H., Doebling S.W., Hemez F.M. et al. − Los Alamos: LANL, 2004. − 41 p.

2. *Sargent R.G.* Verification and validation of simulation models // WSC: Proceedings of the Winter Simulation Conference. − Syracuse, NY 13244, U.S.A. − 2011. − P.183-198.

3. *Стрижнев А.Г., Марков А.В., Русакович А.Н.* Идентификация объекта управления по переходной характеристике замкнутой системы // Доклады БГУИР. − 2012. − № 5(67). − C.65-72.

4. *Бильфельд Н.В.* Идентификация объектов управления в Toolbox PID Tuner // Новый университет. Серия: Технические науки. − 2016. − № 6-7. − C.52-53.

5. *Трунов О.М.* Критерій адекватності як оцінка ефективності процесу побудови моделі // Восточно-Европейский журнал передовых технологий. − 2015. − № 1/4(73). − C.36-41.

6. *Черногоров Е.П.* Некоторые вопросы оценки адекватности модели // Наука ЮУрГУ: материалы 66-й научн. конф. − Челябинск: Издат. центр ЮУрГУ, 2014. − C.162-168.

7. *Fang X., Wang L.* Feature selection based on fisher criterion and sequential forward selection for intrusion detection // Revista de la Facultad de Ingenierna U.C.V. — 2017. — Vol.32. — No 1. — P.498-503.

8. *Дорожкин А.М., Соколова О.И.* Понятие «неопределенность» в современной науке и философии // Вестник Вятского государственного гуманитарного университета. — 2015. — С.5-12.

9. *Электрические системы. Кибернетика электрических систем* : учеб. пос. для электроэнерг. вузов / Под ред. В.А. Веникова. — М.: Высш. школа, 1974. — 328 с.

10. *Бонгард М.М.* Проблемы узнавания. — М.: Наука, 1967. — 320 с.

11. *Манко Г.И., Довгополая А.А.* Об использовании информационного критерия проверки гипотез о законе распределения ошибок измерения // Вопросы химии и химической технологии. — 2013. — № 1. — С.181-184.

12. *Р 50.1.037-2002. Рекомендации по стандартизации. Правила проверки согласия опытного распределения с теоретическим: методические рекомендации.* — Новосибирск: Изд-во НГТУ, 2002. — 153 с.

ВИКОРИСТАННЯ ІНФОРМАЦІЙНОЇ НЕВИЗНАЧЕНОСТІ В ЗАДАЧАХ ІДЕНТИФІКАЦІЇ

Манко Г.І., Чистоклетов Є.П.

Вивчаються інформаційні методи аналізу та управління системами в умовах невизначеності. Обґрунтовано доцільність використання інформаційної невизначеності в задачах ідентифікації об'єктів управління і синтезу числового оцінювання інформаційної невизначеності використовується кількість дезінформації як негативної корисної інформації Бонгарда. Така інформаційна невизначеність може слугувати як критерій адекватності математичної моделі об'єкта управління. Інформаційна оцінювання точності моделювання може бути застосована при будь-якому методі ідентифікації і дозволяє порівнювати точність моделей, що відрізняються одна від одної методом отримання (аналітичним або експериментальним), характером (детерміновані або стохастичні моделі), конкретною реалізацією (фізичні або математичні моделі). В ході параметричної ідентифікації може вирішуватися оптимізаційна задача пошуку мінімуму інформаційної невизначеності в просторі параметрів математичної моделі. Критерій інформаційної невизначеності забезпечує перевірку статистичної гіпотези про адекватність конкретної моделі. Якщо значення критерію перевищує деяке критичне значення, гіпотеза про адекватність повинна бути відкинута. Для розрахунку критичних значень інформаційного критерію адекватності був здійснений статистичний експеримент. З використанням методу Монте-Карло було досліджено розподіл ймовірностей інформаційного критерію. Була побудована досить гладка емпірична функція розподілу критерію. Розподіл інформаційного критерію має виражену асиметрію і невеликий позитивний ексцес. Виявлено, що цей розподіл найкращим чином апроксимується законом Generalized Extreme Value Distribution. Критичне значення може бути визначено як квантиль рівня 0,01 або 0,05 цього розподілу.

Ключові слова: адекватність, комп'ютерне моделювання, закон розподілу, ідентифікація, інформаційний критерій, метод Монте-Карло,

контрольований об'єкт, статистична гіпотеза, невизначеність.

ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННОЙ НЕОПРЕДЕЛЕННОСТИ В ЗАДАЧАХ ИДЕНТИФИКАЦИИ

Манко Г.И., Чистоклетов Е.П.

Изучаются информационные методы анализа и управления системами в условиях неопределенности. Обоснована целесообразность использования информационной неопределенности в задачах идентификации объектов управления и синтеза систем регулирования. Для числовой оценки информационной неопределенности используется количество дезинформации как отрицательной полезной информации Бонгарда. Такая информационная неопределенность может служить в качестве критерия адекватности математической модели объекта управления. Информационная оценка точности моделирования применима при любом методе идентификации и позволяет сравнивать точность моделей, отличающихся друг от друга методом получения (аналитическим или экспериментальным), характером (детерминированные или стохастические модели), конкретной реализацией (физические или математические модели). В ходе параметрической идентификации может решаться оптимизационная задача поиска минимума информационной неопределенности в пространстве параметров математической модели. Критерий информационной неопределенности обеспечивает проверку статистической гипотезы об адекватности конкретной модели. Если значение критерия превышает некоторое критическое значение, гипотеза об адекватности должна быть отвергнута. Для расчета критических значений информационного критерия адекватности был проведен статистический эксперимент. С использованием метода Монте-Карло было исследовано распределение вероятностей информационного критерия. Была построена достаточно гладкая эмпирическая функция распределения критерия. Распределение информационного критерия имеет выраженную асимметрию и небольшой положительный эксцесс. Выявлено, что это распределение наилучшим образом аппроксимируется законом Generalized Extreme Value Distribution. Критическое значение может быть определено как квантиль уровня 0,01 или 0,05 этого распределения.

Ключевые слова: адекватность, компьютерное моделирование, закон распределения, идентификация, информационный критерий, метод Монте-Карло, контролируемый объект, статистическая гипотеза, неопределенность.

Manko G.I., Chistokletov E.P.

USE OF INFORMATION UNCERTAINTY IN IDENTIFICATION TASKS

Manko G.I., Chistokletov E.P.

Ukrainian State University of Chemical Technology, Dnipro, Ukraine

Informational methods for analyzing and managing systems under uncertainty are studied. The expediency of use of information uncertainty in tasks of identification of control objects and synthesis of regulatory systems is justified. For the numerical evaluation of information uncertainty, an amount of disinformation is used as Bongard's negative useful information. Such information uncertainty can serve as a criterion for the adequacy of a mathematical model of a control object. Information assessment of the modeling accuracy is applicable to any method of identification and allows a researcher to compare the accuracy of models that differ from each other by the method of obtaining (analytical or experimental), nature (deterministic or stochastic models), specific implementation (physical or mathematical models). In the course of parametric identification, an optimization problem of finding a minimum of information uncertainty in the parameter space of a mathematical model can be solved. The information uncertainty criterion provides verification of the statistical hypothesis about the adequacy of a particular model. If the criterion value exceeds a certain critical value, the adequacy hypothesis must be rejected. To calculate the critical values of the information adequacy criterion, a statistical experiment was performed. Using the Monte Carlo method, the probability distribution of the information criterion was investigated. A sufficiently smooth empirical criterion distribution function was constructed. The distribution of the information criterion has a pronounced asymmetry and a small positive kurtosis. It is revealed that this distribution is best approximated by the Generalized Extreme Value Distribution law. The critical value can be defined as a quantile of the level of 0.01 or 0.05 for this distribution.

Keywords: adequacy, computer modeling, distribution law, identification, information criterion, Monte Carlo method, control object, statistical hypothesis, uncertainty.

REFERENCES

1. Thacker B.H., Doebling S.W., Hemez F.M. Concepts of Model Verification and Validation. Los Alamos: Los Alamos National Lab., 2004. 41 p.

2. Sargent R.G. Verification and validation of simulation models. WSC: Proceedings of the Winter Simulation Conference. Syracuse, NY 13244, U.S.A., 2011, pp. 183-198.

3. Strizhnev A.G., Markov A.V., Rusakovich A.N. *Identifikaciya ob#ekta upravleniya po perekhodnoj harakteristike zamknutoj sistemy* [Control object identification by a closed system transition characteristic]. *Doklady BGUIR* [Reports of BSUIRE], 2012, no. 5(67), pp. 65-72. (in Russian).

4. Bil'fel'd N.V. *Identifikaciya ob#ektov upravleniya v Toolbox PID Tuner* [Identification of control objects in Toolbox PID Tuner]. *Novyj universitet: serija «Tekhnicheskie nauki»* [New University: «Technical Engineering» series], 2016, no. 6-7. pp. 52-53. (in Russian).

5. Trunov O.M. *Kriterij adekvatnosti yak ocinka efektivnosti procesu pobudovi modeli* [The adequacy criteria as the process of model building efficiency assessment]. *Vostochno-Evropejskij zhurnal peredovyh tekhnologij* [East European Journal of Advanced Technologies], 2015, no. 1/4(73). pp. 36-41. (in Ukrainian).

6. Chernogorov E.P. *Nekotorye voprosy ocenki adekvatnosti modeli* [Some questions of the model adequacy assessment]. *Nauka YUUrGU: materialy 66-j nauchnaja konferencija* [Science of SUSU: materials of the 66th scientific conference]. Chelyabinsk: publishing center YUUrGU, 2014, pp. 162-168. (in Russian).

7. Fang X., Wang L. Feature selection based on fisher criterion and sequential forward selection for intrusion detection. Revista de la Facultad de Ingenierнa U.C.V., 2017, vol. 32, no. 1, pp. 498-503.

8. Dorozhkin A.M., Sokolova O.I. *Ponjatie «neopredelennost'» v sovremennoj nauke i filosofii* [The concept of «uncertainty» in modern science and philosophy]. *Vestnik Vjatskogo gosudarstvennogo gumanitarnogo universiteta* [Bulletin of the Vyatka State Humanitarian University], 2015, pp 5-12. (in Russian).

9. Venikova V.A. *Elektricheskie sistemy* [Electrical systems]. *Kibernetika elektricheskih sistem: ucheb. pos. dlya elektroenerg. vuzov* [Cybernetics of electrical systems: stud guide for electrical energy universities]. Moscow, 1974, 328 p. (in Russian).

10. Bongard M.M. *Problemy uznavaniya* [Problems of Recognition]. Moscow: Science, 1967, 320 p. (in Russian).

11. Manko G.I., Dovgopolaya A.A. *Ob ispol'zovanii informacionnogo kriteriya proverki gipotez o zakone raspredeleniya oshibok izmereniya* [About using the information criteria for testing the hypotheses of measurement error distribution law]. Questions of chemistry and chemical technology, 2013, no. 1, pp. 181-184. (in Russian).

12. R 50.1.037-2002. *Rekomendacii po standartizacii* [Recommendations for standardization]. *Pravila proverki soglasiya opytnogo raspredeleniya s teoreticheskim: metodicheskie rekomendacii* [Rules for checking the accordance of an experimental and theoretical distribution: methodological guidelines] Novosibirsk: NGTU publ., 2002, 153 p. (in Russian).