УДК 004.78:519.72 **doi: 10.32620/reks.2019.2.12**

С. Ю. ГАВРИЛЕНКО

Національний технічний університет «ХПІ», Україна

МЕТОДИКА ВІДБОРУ СИСТЕМИ ПОКАЗНИКІВ ДЛЯ ІДЕНТИФІКАЦІЇ СТАНУ КОМП'ЮТЕРНОЇ СИСТЕМИ

Процеси проектування, розробки, тестування, супроводу та управління сучасними комп'ютерними системами (КС) вимагають вирішення завдань оптимізації в умовах отримання багатьох характеристик в режимі реального часу. При цьому однією з пріоритетних задач є задача відбору цих характеристик або показників. ***Предметом*** *вивчення в статті є методи оцінки інформативності показників функціонування комп'ютерної системи.* ***Метою*** *є розробка методики відбору системи показників для ідентифікації стану комп'ютерних систем за умови нечітких вихідних даних. Для автоматизації процесу вибору найбільш впливових показників необхідно залучення додаткової інформації, щодо впливовості окремих показників та ефективних математичних рішень оброки нечітких даних, що зможе підвищити достовірність результатів ідентифікації стану КС. Це обумовлює* ***актуальність*** *задачі розробки методики відбору системи показників для ідентифікації стану КС.* ***Результати.*** *Проведений аналіз показав, що завдання ідентифікації розпадається на дві підзадачі: вибір сукупності контрольованих показників об'єкта та рішення відповідно завдання ідентифікації. Розглянуто вирішення задачі вибору підмножини інформативних параметрів із множини параметрів, які оцінюються з використання «відстані» Кульбака. Відзначено, що ця технологія базується на припущенні, що щільності параметрів системи, які контролюються, відомі та можуть бути статистично оцінені для кожного із станів системи. В реальних умовах малої вибірки вихідних даних, ця гіпотеза не може бути коректно обґрунтована. В цих умовах природно використовувати інші можливості для опису невизначеності вихідних даних, не використовуючи при цьому математичний апарат теорії ймовірності. Виходячи із цього, сформовані вимоги до критерію інформативності. Показано, що цим вимогам задовольняє математичний апарат нечіткої математики, на базі якого сформована відповідна методика.* ***Висновки.*** *Розроблено методику відбору системи показників для ідентифікації стану комп'ютерної системи. Основною відмінністю цієї методики є введення нового критерію оцінки інформативності за умови нечітких вихідних даних, що підвищує точність ідентифікації стану КС.*

Ключові слова: *комп'ютерна система; ідентифікація стану; оцінки інформативності; «відстань» Кульбака.*

Вступ

Сьогодні розвиток інформаційної сфери та рівень інформаційної безпеки визначає економічну і політичну роль держав на міжнародній арені і вимагає пріоритетного розгляду.

Характерною рисою сучасних держав є активне впровадження комп'ютерних технологій в усі сфери життєдіяльності суспільства.

Процеси проектування, розробки, тестування, супроводу та управління сучасними КС вимагають вирішення завдань оптимізації в умовах отримання багатьох характеристик в режимі реального часу. При цьому однією з пріоритетних задач є задача відбору цих характеристик або показників.

Особливо важливою ця задача стає в процесі ідентифікації стану КС [1-3].

Методологія, що передбачає такій вибір може будуватися на теорії прийняття рішень [4]. При цьому вибір найбільш раціонального показника з множини дуже часто здійснюється експертом без використання теоретично обґрунтованих методологій та без врахування нечіткої природи вхідних даних.

Для автоматизації процесу вибору найбільш впливових показників необхідно залучення додаткової інформації, щодо впливовості окремих показників та ефективних математичних рішень оброки даних, що зможе підвищити достовірність результатів ідентифікації стану КС.

Крім того задача може ускладнюватися, наприклад, коли кількість спостережень є малою або гіпотеза про вид їх ймовірнісного розподілу є недостовірною або вхідні дані є нечіткими числами.

Це обумовлює актуальність задачі розробки методики відбору системи показників для ідентифікації стану КС в реальних умовах малої вибірки вихідних даних.

Предметом дослідження є методи оцінки інформативності показників функціонування комп'ютерних систем для ідентифікації їх стану.

Метою статті є розробка методики відбору сис-

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

теми показників для ідентифікації стану КС за умови нечітких вихідних даних.

1. Аналіз існуючих підходів

Аналіз літератури показав, що в основі сучасних моделей ідентифікації стану об'єкту лежать роботи К.Ф. Гаусса, в яких вперше була сформована і реалізована ідея побудови моделі об'єктів на основі спостереження їх поведінки. Природним розвитком цієї ідеї було формулювання задачі ідентифікації стану об'єктів відповідно до множини параметрів, які контролюються. Ця задача, з урахуванням невизначеності вихідних даних, придбала характер важливої складової вирішення проблеми оцінки ефективності та управління в техніці, економіці, соціології ті ін. [5]. Достатньо загальна математична модель задачі ідентифікації стану в теоретико-імовірнісному трактуванні описується наступним чином [6].

Нехай система може бути в одному із множин m станів: $\left(H_1, H_2, ..., H_m\right)$. Є сукупність параметрів $x = \left(x_1, x_2, ..., x_p\right)$, які контролюються. Значення цих параметрів вимірюються та надалі використовуються для ідентифікації. Крім того, відома матриця розподілу умовних щільностей розподілу випадкових значень параметрів, які контролюються, для можливих станів системи $-f\left(x_i / H_k\right)$, $i = 1, 2, ..., p$, $k = 1, 2, ..., m$. Технологія ідентифікації забезпечує обробку цього вихідного статистичного матеріалу з метою розрахунку розподілу ймовірностей станів об'єкта. Таким чином завдання ідентифікації розпадається на дві підзадачі:

– вибір сукупності контрольованих показників об'єкта;

– рішення відповідно завдання ідентифікації.

2. Використання «відстані» Кульбака

Якщо кількість параметрів, які контролюються, велика і їх контроль потребує великих ресурсних (матеріальних, часових і т. д.) затрат, природним є бажання зменшити кількість параметрів, залишив найбільш інформативні. Задача вибору підмножини інформативних параметрів із множини параметрів, які контролюються, може бути вирішена різним чином [7, 8]. Одним із традиційних підходів для цього випадку є використання «відстані» Кульбака, яка знаходиться наступним чином [9].

Для спрощення, будемо вважати, що множина можливих станів редукована до двох: основна H_0 та альтернативна H_1. Тоді корисність (інформативність) конкретного параметра x оцінюється «відстанню» між розподілами $f\left(x / H_0\right)$ та $f\left(x / H_1\right)$, яке обчислюється за формулою:

$$J_1\left(f\left(x / H_0\right), f\left(x / H_1\right)\right) = = \int_{-\infty}^{\infty} f\left(x / H_0\right) \ln \frac{f\left(x / H_0\right)}{f\left(x / H_1\right)} dx. \quad (1)$$

Чим більше ця відстань, тим вище інформативність параметру x.

Розглянемо декілька прикладів. Нехай $f\left(x / H_0\right)$ та $f\left(x / H_1\right)$ гаусовий розподіл вигляду:

$$f\left(x / H_0\right) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\left\{-\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2}\right\},$$

$$f\left(x / H_1\right) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left\{-\frac{\left(x-\mu_1\right)^2}{2\sigma_1^2}\right\}.$$

Тоді маємо:

$$J_1\left(f\left(x / H_0\right), f\left(x / H_1\right)\right) = = \int_{-\infty}^{\infty} \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2}} \ln\left[\frac{\sigma_1}{\sigma_0} \frac{e^{-\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2}}}{e^{-\frac{\left(x-\mu_1\right)^2}{2\sigma_1^2}}}\right] dx =$$

$$= \int_{-\infty}^{\infty} \left(\begin{array}{c} \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\left\{-\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2}\right\} \times \\ \times \ln\left\{\frac{\sigma_1}{\sigma_0} \exp\left\{-\left[\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2} - \frac{\left(x-\mu_1\right)^2}{2\sigma_1^2}\right]\right\}\right\} \end{array}\right) dx.$$

Якщо $\sigma_0 = \sigma_1 = \sigma$, то

$$J\left(f\left(x / H_0\right), f\left(x / H_1\right)\right) = = \int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2}\right\} \cdot \left[\frac{\left(x-\mu_1\right)^2}{2\sigma_1^2} - \frac{\left(x-\mu_0\right)^2}{2\sigma_1^2}\right] dx =$$

$$= \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2}\right\} \cdot \left[x^2 - 2x\mu_1 + \mu_1^2 - x^2 + 2x\mu_0 - \mu_0^2\right] dx =$$

$$= \frac{2\left(\mu_0 - \mu_1\right)}{2\sigma^2} \int_{-\infty}^{\infty} x \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2}\right\} dx +$$

$$+\frac{\mu_1^2+\mu_0^2}{2\sigma^2}\cdot\int\limits_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2}\right\}dx =$$

$$=\frac{\left(\mu_0-\mu_1\right)\mu_0}{\sigma^2}+\frac{\mu_1^2-\mu_0^2}{2\sigma^2}=$$

$$=\frac{1}{2\sigma^2}\left[2\mu_0-2\mu_0\mu_1+\mu_1^2-\mu_0^2\right]= \tag{2}$$

$$=\frac{1}{2\sigma^2}\left(\mu_1^2-2\mu_0\mu_1+\mu_0^2\right)=\frac{\left(\mu_0-\mu_1\right)^2}{2\sigma^2}.$$

Нехай зараз щільність розподілу випадкових значень параметрів *x*, які контролюються, для різних станів системи (H_0 та H_1) є експоненціальними, тобто:

$$f\left(x/H_0\right)=\lambda_0 e^{-\lambda_0 x},\ f\left(x/H_1\right)=\lambda_1 e^{-\lambda_1 x}.$$

Тоді:

$$J\Big(f\left(x/H_0\right),f\left(x/H_1\right)\Big)=$$

$$=\int\limits_0^{\infty}\lambda_0 e^{-\lambda_0 x}\ln\left[\frac{\lambda_0}{\lambda_1}e^{-\left(\lambda_0-\lambda_1\right)x}\right]dx=$$

$$=\int\limits_0^{\infty}\lambda_0 e^{-\lambda_0 x}\left[\ln\frac{\lambda_0}{\lambda_1}-\left(\lambda_0-\lambda_1\right)x\right]dx= \tag{3}$$

$$=\ln\frac{\lambda_0}{\lambda_1}-\left(\lambda_0-\lambda_1\right)\int\limits_0^{\infty}x\lambda_0 e^{-\lambda_0 x}dx=$$

$$=\ln\frac{\lambda_0}{\lambda_1}-\frac{\left(\lambda_0-\lambda_1\right)}{\lambda_0}=\ln\frac{\lambda_0}{\lambda_1}+\frac{\lambda_1}{\lambda_0}-1.$$

Із структури співвідношення (1), яка визначає значення критерію, видно, що його числове значення дорівнює нулю, якщо щільності розподілу $f\left(x/H_0\right)$ та $f\left(x/H_1\right)$ співпадають.

З другого боку, із аналізу результатів розрахунку критерію в конкретних задачах ясно, що значення критерію збільшується, зі збільшенням різниці між цими щільностями, і воно не обмежено зверху.

Необхідно відмітити наступну важливу обставину, яка суттєво обмежує використання технології Кульбака, для оцінки інформативності параметрів, які контролюються. Ця технологія основана на допущенні, що щільності параметрів системи, які контролюються, відомі та можуть бути статистично оцінені для кожного із станів системи. В реальних умовах малої вибірки вихідних даних, ця гіпотеза не може бути коректо обґрунтована. В цих умовах природно використовувати інші можливості для опису невизначеності вихідних даних, не використовуючи при цьому математичний апарат теорії ймовірності.

Проаналізуємо можливості розрахунку аналога міри Кульбака для випадку, коли реальна невизначеність відносно значення параметрів, які контролюються задається в термінах нечіткої математики. Нехай шляхом аналізу експериментальних даних отримано функції приналежності контрольованих параметрів *x* для якої-небудь пари можливих станів системи H_i та H_k. Виконаємо розрахунок для окремого випадку, коли функції приналежності є гаусовими.

Введемо:

$$\mu\left(x/H_i\right)=\exp\left\{-\frac{\left(x-\mu_i\right)^2}{2\sigma_i^2}\right\},$$

$$\tilde{\mu}\left(x/H_i\right)=\frac{M\left(x/H_i\right)}{\int\limits_{-\infty}^{\infty}M\left(x/H_i\right)dx};$$

$$\mu\left(x/H_k\right)=\exp\left\{-\frac{\left(x-\mu_k\right)^2}{2\sigma_k^2}\right\},$$

$$\tilde{\mu}\left(x/H_k\right)=\frac{M\left(x/H_k\right)}{\int\limits_{-\infty}^{\infty}M\left(x/H_k\right)dx}.$$

Тоді:

$$J\Big(\mu\left(x/H_i\right),\mu\left(x/H_k\right)\Big)=\int\limits_{-\infty}^{\infty}\tilde{\mu}\left(x/H_i\right)\ln\frac{\tilde{\mu}\left(x/H_i\right)}{\tilde{\mu}\left(x/H_k\right)}dx=$$

$$=\int\limits_{-\infty}^{\infty}\frac{1}{\sigma_i\sqrt{2\pi}}\exp\left\{-\frac{\left(x-\mu_i\right)^2}{2\sigma_i^2}\right\}\times$$

$$\times\ln\left[\frac{\sigma_k}{\sigma_i}\cdot\exp\left\{-\frac{\left(x-\mu_i\right)^2}{2\sigma_i^2}\right\}\Bigg/\exp\left\{-\frac{\left(x-\mu_k\right)^2}{2\sigma_k^2}\right\}\right]dx=$$

$$=\int\limits_{-\infty}^{\infty}\frac{1}{\sigma_i\sqrt{2\pi}}\exp\left\{-\frac{\left(x-\mu_i\right)^2}{2\sigma_i^2}\right\}\times$$

$$\times\ln\left[\frac{\sigma_k}{\sigma_i}\exp\left\{-\frac{\left(x-\mu_i\right)^2}{2\sigma_i^2}+\frac{\left(x-\mu_k\right)^2}{2\sigma_k^2}\right\}\right]dx.$$

Нехай $\sigma_i=\sigma_k=\sigma$. Тоді:

$$J=\frac{1}{\sigma\sqrt{2\pi}}\times$$

$$\times\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{\left(x-\mu_i\right)^2}{2\sigma^2}\right\}\left[\frac{\left(x-\mu_k\right)^2-\left(x-\mu_i\right)^2}{2\sigma^2}\right]dx=$$

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

$$
\begin{aligned}
&= \frac{1}{2\sigma^3\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-\mu_i)^2}{2\sigma^2}\right\} \times \\
&\times \left[x^2 - 2x\mu_k + \mu_k^2 - x^2 + 2x\mu_i - \mu_i^2\right] dx = \\
&= \frac{\mu_i - \mu_k}{\sigma^3\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left\{-\frac{(x-\mu_i)^2}{2\sigma^2}\right\} dx + \\
&+ \frac{\mu_k^2 - \mu_i^2}{2\sigma^3\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-\mu_i)^2}{2\sigma^2}\right\} dx = \\
&= \frac{\mu_i - \mu_k}{\sigma^2} \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_i)^2}{2\sigma^2}\right\} dx \\
&+ \frac{\mu_k^2 - \mu_i^2}{2\sigma^2} = \frac{1}{\sigma^2}\left[(\mu_i - \mu_k)\mu_i + \frac{\mu_k^2 - \mu_i^2}{2}\right] = \\
&= \frac{1}{2\sigma^2}\left[2\mu_i^2 - 2\mu_i\mu_k + \mu_k^2 - \mu_i^2\right] = \frac{1}{2\sigma^2}(\mu_i - \mu_k)^2 .
\end{aligned} \tag{4}
$$

Отримане співвідношення (4), природно, співпадає із (2).

Аналіз результатів конкретних рішень дозволяє зробити наступні висновки.

1. Значення міри Кульбака дорівнює нулю у випадку, коли умовні щільності розподілу (або функції приналежності) співпадають, і ця міра може приймати довільно велике, не обмежено зверху значення в іншому випадку.

2. Діапазони можливих числових значень критерію залежать від типів щільності розподілу (функцій приналежності).

3. Критерій Кульбака є асиметричними, тобто його числове значення може залежати від того, який стан вибрано основним і який є альтернативним, тобто від правила входження щільності в розрахункову формулу. Дійсно, якщо щільності розподілу (або функції приналежності) експоненціальні, то отримане при інвертуванні H_0 та H_1 співвідношення не співпадає з (3):

$$J = \ln\frac{\lambda_1}{\lambda_0} + \frac{\lambda_0}{\lambda_1} - 1 .$$

Зазначені обставини ініціюють розробку інших методів оцінки інформативності контрольованих показників.

3. Методика відбору системи показників

Відзначимо тепер, що рішення цього завдання не вичерпує проблем, пов'язаних з побудовою системи ідентифікації стану об'єкта.

В зв'язку з цим сформуємо вимоги до критерію інформативності.

1. Критерій повинен дорівнювати нулю, якщо функції приналежності нечітких значень параметрів, які контролюються, для різних станів системи співпадають.

2. Числове значення критерію в випадку, якщо функції приналежності параметра для різних станів різні, повинні належати кінцевому інтервалу, який не залежить від типу функції приналежності, наприклад інтервалу [0,1]. При цьому, при збільшення різниці в функціях приналежності, числове значення критерію повинно монотонно зростати.

3. Числове значення критерію не повинно залежати від правила входження функцій приналежності в формулу для його розрахунку.

Цим вимогам задовольняє наступне співвідношення, яке для пари функцій приналежності $\mu(x/H_i)$, $\mu(x/H_k)$ має вигляд:

$$
\begin{aligned}
&J\left(\mu(x/H_i), \mu(x/H_k)\right) = \\
&= 1 - \int_{-\infty}^{\infty} \left(\tilde{\mu}(x/H_i)\tilde{\mu}(x/H_k)\right)^{1/2} dx,
\end{aligned} \tag{5}
$$

де, як і раніше:

$$\tilde{\mu}(x/H_i) = \frac{\mu(x/H_i)}{\int_{-\infty}^{\infty} \mu(x/H_i)},$$

$$\tilde{\mu}(x/H_k) = \frac{\mu(x/H_k)}{\int_{-\infty}^{\infty} \mu(x/H_k)}.$$

Розрахуємо значення критерія (5) для гаусових функцій приналежності. При цьому:

$$
\begin{aligned}
&J\left(\mu(x/H_i), \mu(x/H_k)\right) = \\
&= 1 - \int_{-\infty}^{\infty} \left(\begin{matrix} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right\} \frac{1}{\sigma_k\sqrt{2\pi}} \times \\ \times \exp\left\{-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right\} \end{matrix} \right)^{1/2} dx = \\
&= 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_i\sigma_k}} e^{-\frac{1}{2}\left[\frac{(x-\mu_i)^2}{2\sigma_i^2} + \frac{(x-\mu_k)^2}{2\sigma_k^2}\right]} dx = \\
&= 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_i\sigma_k}} e^{-\frac{1}{4\sigma_k^2\sigma_k^2}\left[\begin{matrix} x^2\sigma_i^2 + x^2\sigma_k^2 - 2x\mu_i\sigma_k^2 - \\ -2x\mu_k\sigma_i^2 + \mu_i^2\sigma_k^2 + \mu_k^2\sigma_i^2 \end{matrix}\right]} dx =
\end{aligned}
$$

$$=1-\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_i\sigma_k}}e^{-\frac{\sigma_i^2+\sigma_k^2}{4\sigma_k^2\sigma_k^2}\left[x^2-2x\frac{\mu_i\sigma_k^2+\mu_k\sigma_i^2}{\sigma_i^2+\sigma_k^2}+\frac{\mu_i^2\sigma_k^2+\mu_k^2\sigma_i^2}{\sigma_i^2+\sigma_k^2}\right]}dx= \tag{7}$$

$$=1-\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_i\sigma_k}}e^{-\frac{\sigma_i^2+\sigma_k^2}{4\sigma_k^2\sigma_k^2}\left[\left(x^2-\frac{\mu_i\sigma_k^2+\mu_k\sigma_i^2}{\sigma_i^2+\sigma_k^2}\right)^2-\left(\frac{\mu_i\sigma_k^2+\mu_k\sigma_i^2}{\sigma_i^2+\sigma_k^2}\right)^2+\frac{\mu_i^2\sigma_k^2+\mu_k^2\sigma_i^2}{\sigma_i^2+\sigma_k^2}\right]}dx=$$

$$=1-\int_{-\infty}^{\infty}\sqrt{\frac{2\sigma_i\sigma_k}{\sigma_i^2+\sigma_k^2}}\frac{1}{\sqrt{2\pi}\cdot\sigma_i\sigma_k\Big/\left(\sigma_i^2+\sigma_k^2\right)^{1/2}}\times$$

$$\times e^{-\left(x^2-\frac{\mu_i\sigma_k^2+\mu_k\sigma_i^2}{\sigma_i^2+\sigma_k^2}\right)^2\Big/\left(2\frac{2\sigma_i\sigma_k}{\sigma_i^2+\sigma_k^2}\right)}dx\left(e^{-A}dx\right),$$

де

$$A=\frac{1}{4\left(\sigma_i^2+\sigma_k^2\right)\sigma_i^2\sigma_k^2}\times$$

$$\times\left[\begin{matrix}\mu_i^2\sigma_k^2\sigma_i^2+\mu_i^2\sigma_k^4+\mu_k^2\sigma_i^4+\mu_k^2\sigma_i^2\sigma_k^2-\\-\mu_i^2\sigma_k^4-2\mu_i\mu_k\sigma_k^4\sigma_i^2-\mu_k^2\sigma_i^4\end{matrix}\right]=\frac{\left(\mu_i-\mu_k\right)^2}{4\left(\sigma_i^2+\sigma_k^2\right)}.$$

Підставляючи A в (7), отримуємо:

$$J\left(\mu\left(x/H_i\right),\mu\left(x/H_k\right)\right)=$$

$$=1-\sqrt{\frac{2\sigma_i\sigma_k}{\sigma_i^2+\sigma_i^2}}e^{-\left(\mu_i-\mu_k\right)^2\Big/\left(4\left(\sigma_i^2+\sigma_k^2\right)\right)}. \tag{8}$$

Легко бачити, що числові значення міри (8) завжди належать інтервалу [0,1]. Дійсно,

$$0<e^{-\left(\mu_i-\mu_k\right)^2\Big/\left(4\left(\sigma_i^2+\sigma_k^2\right)\right)}\le 1, \tag{9}$$

причому рівність справа досягається тільки в тому випадку, коли $\mu_i=\mu_k$. Крім того, так як:

$$\frac{\sigma_i}{\sigma_k}+\frac{\sigma_k}{\sigma_i}-2=\frac{\sigma_i^2+\sigma_k^2-2\sigma_i\sigma_k}{\sigma_i\sigma_k}=\frac{\left(\sigma_i-\sigma_k\right)^2}{\sigma_i\sigma_k}\ge 0,$$

то
$$\frac{\sigma_i}{\sigma_k}+\frac{\sigma_k}{\sigma_i}=\frac{\sigma_i^2+\sigma_k^2}{\sigma_i\sigma_k}\ge 2,$$

звідки
$$\frac{\sigma_i\sigma_k}{\sigma_i^2+\sigma_k^2}\le\frac{1}{2},$$

та
$$\sqrt{\frac{2\sigma_i\sigma_k}{\sigma_i^2+\sigma_k^2}}\le 1, \tag{10}$$

причому рівність досягається тільки, коли $\sigma_i=\sigma_k$. Таким чином, із (5) – (10) випливає необхідне:

$$0\le J\left(\mu\left(x/H_i\right),\mu\left(x/H_k\right)\right)<1.$$

Таким чином, міра (5) дорівнює нулю, якщо $\mu\left(x/H_i\right)=\mu\left(x/H_k\right)$ та наближується до одиниці, при збільшені різниці між цими функціями приналежності. Крім того, ця міра симетрична, тобто її числове значення не залежить від правила входження функції приналежності в розрахункове співвідношення.

Отримаємо співвідношення для розрахунку критерію (5) в випадку, коли функції приналежності $\mu\left(x/H_i\right),\mu\left(x/H_k\right)$ – експоненціальні.

При цьому маємо:

$$J\left(\mu\left(x/H_i\right),\mu\left(x/H_k\right)\right)=$$

$$=1-\int_0^{\infty}\left(\tilde{\mu}\left(x/H_i\right)\tilde{\mu}\left(x/H_k\right)\right)^{1/2}dx=$$

$$=1-\int_0^{\infty}\left(\lambda_i e^{-\lambda_i x}\cdot\lambda_k e^{-\lambda_k x}\right)^{1/2}dx=$$

$$=1-\int_0^{\infty}\left(\lambda_i\lambda_i\right)^{1/2}e^{-x\frac{\lambda_i+\lambda_k}{2}}dx= \tag{11}$$

$$=1-\frac{2\left(\lambda_i\lambda_i\right)^{1/2}}{\lambda_i+\lambda_k}\int_0^{\infty}\frac{\lambda_i+\lambda_k}{2}e^{-x\frac{\lambda_i+\lambda_k}{2}}dx=$$

$$=1-\frac{2\left(\lambda_i\lambda_i\right)^{1/2}}{\lambda_i+\lambda_k}=1-\frac{2}{\sqrt{\frac{\lambda_i}{\lambda_k}}+\sqrt{\frac{\lambda_k}{\lambda_i}}}.$$

Так як середнє арифметичне двох чисел не менше, чим їх середнє геометричне, то

$$\frac{1}{2}\left(\sqrt{\frac{\lambda_i}{\lambda_k}}+\sqrt{\frac{\lambda_k}{\lambda_i}}\right)\ge\left(\sqrt{\frac{\lambda_i}{\lambda_k}}\sqrt{\frac{\lambda_k}{\lambda_i}}\right)^{1/2}=1.$$

Тоді
$$\sqrt{\frac{\lambda_i}{\lambda_k}}+\sqrt{\frac{\lambda_k}{\lambda_i}}\ge 2.$$

Звідси випливає, що значення критерію (11) лежить в інтервалі [0, 1]. Такий висновок становиться ще біль прозорим, якщо зробити елементарні перетворення другого доданку в (11):

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

$$1-\frac{2}{\sqrt{\frac{\lambda_i}{\lambda_k}}+\sqrt{\frac{\lambda_k}{\lambda_i}}}=1-\frac{2\sqrt{\frac{\lambda_k}{\lambda_i}}}{\frac{\lambda_k}{\lambda_i}+1}=1-\frac{2\lambda_k\sqrt{\frac{\lambda_i}{\lambda_k}}}{\lambda_i+\lambda_k}=$$

$$=1-\frac{2\sqrt{\lambda_i+\lambda_i\lambda_k}}{\lambda_i+\lambda_k}=1-\frac{\sqrt{\lambda_i\lambda_k+\lambda_k}}{\frac{\lambda_i+\lambda_k}{2}}.$$

Таким чином, другий доданок в (11) є відношення середнього геометричного двох чисел λ_i та λ_k до їх середнього арифметичного, яке не є більшим одиниці. При цьому рівність нулю в (11) має місце тільки, якщо $\lambda_i=\lambda_k$ і це значення асимптотично прямує до 1 по мірі збільшення розбіжності між $\mu(x/H_i)$ та $\mu(x/H_k)$.

Таким чином, критерій (5) задовольняє висунутим вимогам.

Розглянемо ще один простий критерій інформативності. Введемо наступний критерій інформативності нечітких параметрів, який базується на оцінці площі ділянок, які виникають при перетині $\mu(x/H_i)$ та $\mu(x/H_k)$. Нехай множина можливих старів описується набором $H=(H_1,H_2,...,H_m)$. Для нечіткого параметра *x* введемо множину умовних функцій приналежності

$$\mu=\left(\mu(x/H_1),\mu(x/H_2),...,\mu(x/H_m)\right),$$

де $\mu(x/H_i)$ функція приналежності нечіткого значення параметра x для стану $H_i, i=1,2,...,m$.

Тепер для пари станів H_i, H_k введемо:

$$\mu_{ik}^{(C)}(x)=\min\{\mu(x/H_i),\mu(x/H_k)\} \qquad (12)$$

$$\mu_{ik}^{(D)}(x)=\max\{\mu(x/H_i),\mu(x/H_k)\}.$$

Співвідношення (12) проілюстровані на рис. 1 заштрихованими областями.

Знайдемо площу під кривими $\mu_{ik}^{(C)}(x)$ та $\mu_{ik}^{(D)}(x)$:

$$S\left(\mu_{ik}^{(C)}(x)\right)=\int_{-\infty}^{x^*}\mu(x/H_k)dx+\int_{x^*}^{\infty}\mu(x/H_i)dx,$$

$$S\left(\mu_{ik}^{(D)}(x)\right)=\int_{-\infty}^{x^*}\mu(x/H_i)dx+\int_{x^*}^{\infty}\mu(x/H_k)dx.$$

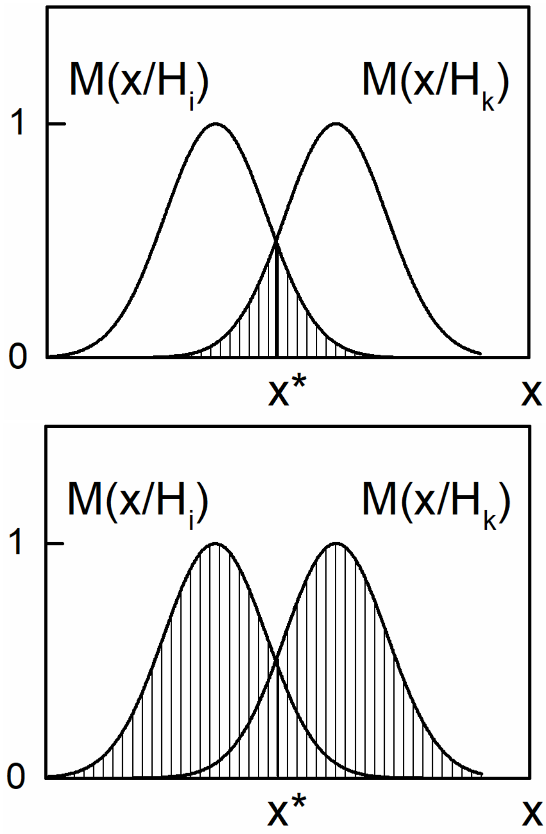

Рис. 1. Критерій інформативності, який базується на значенні площі

Визначимо правило розрахунку x^*. Для параметра x задамо функції приналежності (L-R)-типу, які відповідають станам H_i та H_k:

$$\mu_i(x)=\begin{cases}L\left(\frac{a_i-x}{\alpha_i}\right), & x\le a_i,\\ R\left(\frac{x-a_i}{\beta_i}\right), & x>a_i;\end{cases}$$

$$\mu_k(x)=\begin{cases}L\left(\frac{a_k-x}{\alpha_k}\right), & x\le a_k,\\ R\left(\frac{x-a_k}{\beta_k}\right), & x>a_k.\end{cases}$$

Знайдемо точку перетину функцій приналежності $\mu(x/H_i)$ та $\mu(x/H_k)$, вирішуючи рівняння:

$$R\left(\frac{x-a_i}{\beta_i}\right)=L\left(\frac{a_k-x}{\alpha_k}\right), \quad a_i<a_k.$$

Тоді міра інформативності параметра x для різних станах H_i та H_k визначається відношенням:

$$\xi_{ik} = 1 - \frac{S\left(\mu_{ik}^{(C)}(x)\right)}{S\left(\mu_{ik}^{(D)}(x)\right)} \in [0,1].$$

Із цього відношення слідує, що запропонований критерій задовольняє пред'явленим вимогам: він дорівнює нулю, якщо функції приналежності показників для різних станів об’єкту співпадають, та асимптотично наближається до одиниці по мірі їх віддалення один від одного.

Наведемо приклад. Нехай задано функції приналежності:

$$\mu_i(x) = \begin{cases} \exp\left\{-\left(\dfrac{\mu_i - x}{\alpha_i}\right)^2\right\}, & x \le \mu_i, \\ \exp\left\{-\left(\dfrac{x - \mu_i}{\beta_i}\right)\right\}, & x > \mu_i; \end{cases}$$

$$\mu_k(x) = \begin{cases} \exp\left\{-\left(\dfrac{\mu_k - x}{\alpha_k}\right)^2\right\}, & x \le \mu_k, \\ \exp\left\{-\left(\dfrac{x - \mu_k}{\beta_k}\right)\right\}, & x > \mu_k. \end{cases}$$

Рівняння для розрахунку x^* має вигляд:

$$\exp\left\{-\left(\frac{x - \mu_i}{\beta_i}\right)^2\right\} = \exp\left\{-\left(\frac{\mu_k - x}{\alpha_k}\right)^2\right\}.$$

звідки

$$\frac{x - \mu_i}{\beta_i} = \frac{\mu_k - x}{\alpha_k},$$

$$\alpha_k x - \alpha_k \mu_i = \beta_i \mu_k - \beta_i x,$$

$$x(\alpha_k + \beta_i) = \beta_i \mu_k + \alpha_k \mu_i,$$

$$x^* = \frac{\beta_i \mu_k + \alpha_k \mu_i}{\alpha_k + \beta_i}.$$

4. Дослідження розробленої методики відбору системи показників

Проведемо дослідження розробленої методики відбору системи показників на прикладі розробленої в роботі [10] системи ідентифікації КС. Загальну систему ідентифікації стану КС наведено на рис. 2. Синтезовані показники нормального функціонування КС, а саме: BDS-тестування, показник Херста, вихідні дані контрольних карт є нечіткими величинами. Інформативність цих показників оцінюється за допомогою розробленого методу відбору і надалі використовується для ідентифікації стану КС.

Крім того ідентифікація стану КС на основі дискримінантного та кластерного аналізу може бути використана як для чітких так і для нечітких вихідних даних, що також потребує оцінки їх інформативності [11].

Результати досліджень можливостей використання запропонованої методики оцінки інформативності показників наведено у табл. 1.

Як видно з таблиці, запропонована методика, разом з відомою методикою відбору інформативних показників функціонування комп’ютерної системи, дозволила проводити обробку та відбір показників на повній множини вихідних даних (з урахуванням нечітких даних), що підвищує точність ідентифікації стану КС.

Висновки

Таким чином, розроблено методику відбору системи показників для ідентифікації стану комп’ютерної системи. Основною відмінністю цієї методики є введення нового критерію оцінки інформативності за умови нечітких вихідних даних. Це дозволило оцінити показники інформативності КС та ідентифікувати стан системи в критичних умовах, коли показники, які контролюються, не можуть бути статистично оцінені для кожного із станів системи, тобто мають ознаки нечіткості.

Працездатність цієї методики підтверджується багаточисельними розрахунками за допомогою математичних пакетів MathCad та Mathlab.

У подальшому цю методику можливо використовувати у сукупності з іншими методами ідентифікації стану КС в загальній нечіткій експертній системі ідентифікації стану КС.

Література

1. *Семенов, С. Г. Защита данных в компьютеризированных управляющих системах [Текст] : моногр. / С. Г. Семенов, В. В. Давыдов, С. Ю. Гавриленко. – «LAP LAMBERT ACADEMIC PUBLISHING» Германия, 2014. – 236 с.*

2. *Шелухин, О. И. Обнаружение вторжений в компьютерные сети [Текст] / О. И. Шелухин, Д. Ж. Сакалема, А. С. Филинова. – М. : Гарячая линия-Телеком, 2013. – 220 с.*

3. *Agrawal, S. Survey on Anomaly Detection using Data Mining Techniques [Text] / S. Agrawal // Procedia Computer Science. – 2015. – Vol. 60. – P. 708 –713.*

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

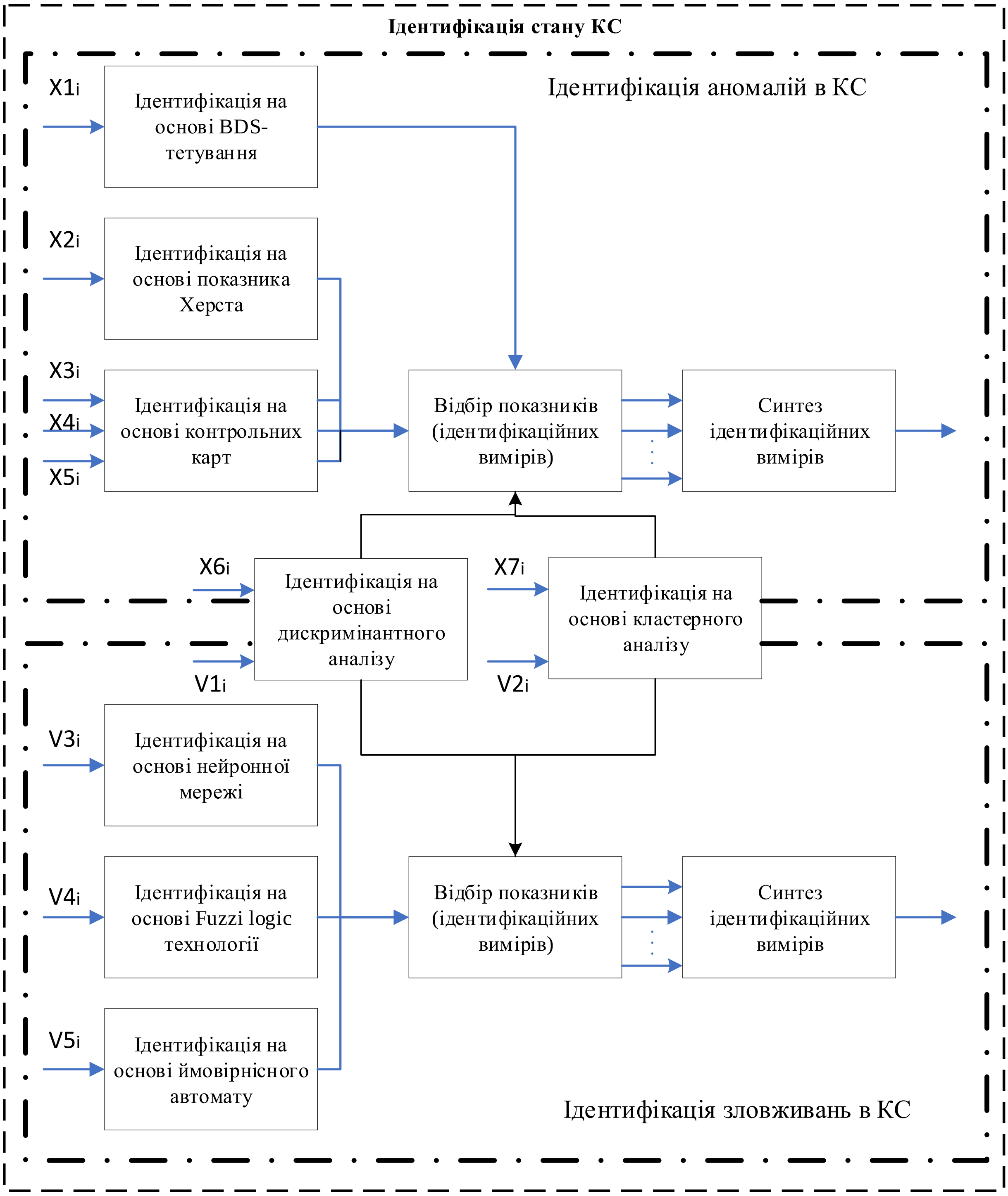

Рис. 1. Загальна система ідентифікації стану КС

Таблиця 1

Результати досліджень можливостей використання запропонованої методики

	Показники BDS-тестування	Показники Херста	Вихідні дані контрольних карт	Вихідні дані дискримінантного та кластерного аналізу	Вихідні дані ймовірнісних автоматів	Вихідні дані нейронної мережі	Вихідні дані Fuzzi Logic технології
Запропонована методика відбору показників	+	+	+	+	+	+	+
Звісна методика на основі міри Кульбака	–	–	–	+/–	+	+	+

4. Боровиков, В. П. Искуство анализа данных [Текст] / В. П. Боровиков. – Питер, 2005. – 432 с.

5. Ципкін, Я. З. Информаційна теорія ідентифікації [Текст] / Я. З. Ципкін. – М. : Наука. Фізматліт, 1995. – 336 с.

6. Chandola, V. Anomaly detection for discrete sequences: A survey [Text] / V. Chandola, A. Banerjee, V. Kumar // IEEE Transactions on Knowledge and Data Engineering. – 2012. – Vol. 24, No. 5. – P. 823–839.

7. Дилигенская, А. Н. Идентификация объектов управления [Текст] : учеб. пособ. / А. Н. Дилигенская. – Самара : Самар. гос. техн. ун-т., 2009. – 136 с.

8. Русаков, К. Д. О задаче выбора признаков наблюдаемого состояния сложного динамического объекта в условиях различного качества измерительной информации [Текст] / К. Д. Русаков, С. Ш. Хиль // Нейрокомпьютеры и их применение: 15-я Всероссийская научн. конф.: тез. докл. – М. : ФГБОУ ВО МГППУ, 2017. – С. 246-248.

9. Kullback, S. Information Theory and Statistics [Text] / S. Kullback. – New York : Wiley, 1959. – 395 p.

10. Гавриленко, С. Ю. Синтез ідентифікаційних вимірів в комп'ютерній системі критичного призначення [Текст] / С. Ю. Гавриленко // Сучасний стан наукових досліджень та технологій в промисловості. – 2019. – № 2 (8). – С. 36-43. – DOI: 10.30837/2522-9818.2019.8.036.

11.Identification of the state of an object under conditions of fuzzy input data [Text] / S. Semenov, O. Sira , S. Gavrylenko, N. Kuchuk // Eastern-European Journal of Enterprise Technologies.– 2019. –Vol. 1, No. 4 (97). – P. 22-29. DOI: 10.15587/1729-4061.2019.157085.

References

1. Semenov, S. G., Davydov, V. V., Gavrilenko, S. Yu. *Zashhita dannykh v komp`yuterizirovannykh upravlyayushhikh sistemakh* [Data protection in computerized control systems]. Germany, LAP LAMBERT ACADEMIC PUBLISHING Publ., 2014. 236 p.

2. Shelukhin, O. I., Sakalema, D. Zh., Filinova, A. S. *Obnaruzhenie vtorzhenii v komp'yuternye seti* [Detection of intrusions into computer networks]. Moscow, Goryachaya liniya-Telekom Publ., 2013. 220 p.

3. Agrawal, S. Survey on Anomaly Detection using Data Mining Techniques. *Procedia Computer Science*, 2015, vol. 60, pp. 708-713.

4. Borovikov, V. P. *Iskusstvo analiza dannykh* [Art of data analysis]. St. Petersburg, Piter Publ., 2005. 432 p.

5. Tsypkin, Ya. Z. *Informatsionnaya teoriya identifikatsii* [Information theory of identification]. Moscow, Nauka, Fizmatlit Publ., 1995. 336 p.

6. Chandola, V., Banerjee, A., Kumar, V. Anomaly detection for discrete sequences: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2012, vol. 24, no. 5, pp. 823-839.

7. Diligenskaya, A. N. *Identifikatsiya ob"ektov upravleniya* [Identification of control objects]. Samara, Samar. gos. tekhn. un-t Publ., 2009. 136 p.

8. Rusakov, K. D., Khil', S. Sh. O zadache vybora priznakov nablyudaemogo sostoyaniya slozhnogo dinamicheskogo ob"ekta v usloviyakh razlichnogo kachestva izmeritel'noi informatsii [On the problem of choosing the signs of the observed state of a complex dynamic object under conditions of different quality measurement information]. *Neirokomp'yutery i ikh primenenie: 15-ya Vserossiiskaya nauchn. konf.* [Neurocomputers and their applications: the 15th All-Russian Scientific. konf.]. Moscow, FGBOU VO MGPPU, 2017, pp. 246-248.

9. Kullback, S. Information Theory and Statistics. New York, Wiley Publ., 1959. 395 p.

10. Gavrylenko, S. Yu. Syntez identyfikatsiinykh vymiriv v kompiuternii systemi krytychnoho pryznachennia [Synthesis of identification measurements in the computer system of critical use]. *Suchasnyi stan naukovykh doslidzhen ta tekhnolohii v promyslovosti* [Current state of scientific researches and technologies in industry], 2019, no. 2 (8), pp. 36-43. DOI: 10.30837/2522-9818.2019.8.036.

11. Semenov, S., Sira, O., Gavrylenko, S., Kuchuk, N. Identification of the state of an object under conditions of fuzzy input data. *Eastern-European Journal of Enterprise Technologies*, 2019, vol. 1, no 4 (97), pp. 22-29. DOI: 10.15587/1729-4061.2019.15708.

Надійшла до редакції 14.05.2019, розглянута на редколегії 12.06.2019

МЕТОДИКА ОТБОРА СИСТЕМЫ ПОКАЗАТЕЛЕЙ ДЛЯ ИДЕНТИФИКАЦИИ СОСТОЯНИЯ КОМПЬЮТЕРНОЙ СИСТЕМЫ

С. Ю. Гавриленко

Процессы проектирования, разработки, тестирования, сопровождения и управления современными компьютерными системами (КС) требуют решения задач оптимизации. Особенностью таких процессов является получение многих характеристик в режиме реального времени. При этом одной из приоритетных задач является задача отбора этих характеристик или показателей. **Предметом изучения** в статье являются методы оценки информативности показателей функционирования компьютерных систем. **Целью статьи** является разработка методики отбора системы показателей для идентификации состояния компьютерных систем в условиях нечетких исходных данных. Для автоматизации процесса выбора наиболее влиятельных показателей необходимо привлечение дополнительной информации относительно влияния отдельных пока-

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

зателей, а также эффективных математических решений обработки нечетких данных, что сможет повысить достоверность результатов идентификации состояния КС. Это обусловливает **актуальность** задачи разработки методики отбора системы показателей для идентификации состояния КС. **Результаты.** Проведенный анализ показал, что задача идентификации распадается на две подзадачи. Первая – выбор совокупности контролируемых показателей объекта, вторая – решение соответственно задачи идентификации. Рассмотрено решение задачи выбора подмножества информативных параметров из множества контролируемых параметров с использованием «расстоянии» Кульбака. Отмечено, что эта технология базируется на предположении, что плотности контролируемых параметров известны и могут быть статистически оценены для каждого из состояний системы. В реальных условиях малой выборки исходных данных эта гипотеза не может быть корректно обоснована. В этих условиях естественно использовать другие возможности для описания неопределенности исходных данных. При этом желательно не использовать математический аппарат теории вероятности. Исходя из этого, сформированы требования к критерию информативности. Показано, что этим требованиям удовлетворяет математический аппарат нечеткой математики, на базе данного аппарата сформирована соответствующая методика. **Выводы.** Разработана методика отбора системы показателей для идентификации состояния компьютерной системы. Основным отличием этой методики является введение нового критерия оценки информативности при условии нечетких исходных данных, что повышает точность идентификации состояния КС.

Ключевые слова: компьютерная система; идентификация состояния; оценки информативности; «расстояние» Кульбака.

THE INDICATORS' SYSTEM SELECTION METHOD FOR THE COMPUTER SYSTEM STATUS IDENTIFICATION

S. Yu. Gavrylenko

The processes of design, development, testing, maintenance, and management of modern computer systems (CS) require the solution of optimization tasks. The peculiarity of such processes is to obtain many characteristics in real time. At the same time, one of the priority tasks is the task of selecting these characteristics or indicators.

The subject of the article study is method for evaluating the informativeness of the indicators of the functioning of computer systems. The purpose of the article is to develop a methodology for selecting a system of indicators for identifying the state of the CS under the condition of fuzzy output data. To automate the process of choosing the most influential indicators, it is necessary to involve additional information. Information is needed on the impact of individual indicators and effective mathematical decisions on false data cessation. This information can increase the reliability of the results of the identification of the state of the CS. This determines the relevance of the task of developing a methodology for selecting a system of indicators for identifying the state of the CS. Results. The analysis showed that the identification problem falls into two subtasks. The first is the choice of a set of monitored indicators of an object. The second is the solution for the identification task accordingly. The solution to the problem of choosing a subset of informative parameters from a set of parameters controlled by using the "distances" of Kulbak is considered. This technology requires the following assumption. The densities of system parameters that are monitored are known. They can be statistically evaluated for each state of the system. In real conditions of a small sample of output data, this hypothesis cannot be properly justified. In these circumstances, it is natural to use other possibilities to describe the uncertainty of the output data. At the same time, it is not desirable to use the mathematical apparatus of probability theory. Proceeding from this, the requirements for the informative criterion are formed. It is shown that these requirements are met by a mathematical apparatus of fuzzy mathematics. On the basis of this apparatus, an appropriate method is formed. Conclusions. The method of selection of the system of indicators for the identification of the state of the computer system is developed. This allowed us to expand the range of incoming processed metrics in the state identification system on the condition of fuzzy input data and improve the accuracy of identifying the state of the CS.

Keywords: computer system; state identification; estimation of informality; "distance" of Kulbak.

Гавриленко Світлана Юріївна – кандидат технічних наук, доцент, професор кафедри «Обчислювальна техніка та програмування», Національний технічний університет «Харківський політехнічний інститут», Харків, Україна.

Gavrylenko Svitlana – PhD (Technical), Associate Professor, Professor of the Department of "Computer Science and Programming", National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; e-mail: gavrilenko08@gmail.com, ORCID Author ID: 0000-0002-6919-0055.