УДК 621.396

А.О. Левченко, к.т.н., с.н.с., доц.
Академія сухопутних військ ім. гетьмана Петра Сагайдачного

ШЛЯХИ РЕАЛІЗАЦІЇ БАЗОВОГО АЛГОРИТМУ МЕТОДУ МАКСИМУМУ КОМПАКТНОСТІ ЯК ГЕНЕРАТОРА СТРУКТУРИ ПРИКЛАДНИХ ПРОГРАМ СТАТИСТИЧНОЇ ОБРОБКИ ДАНИХ

У статті розглянуто шляхи реалізації базового алгоритму методу максимуму компактності як генератора структури прикладних програм статистичної обробки даних. Наведено результати аналізу трансформації загального випадку до випадків, опублікованих раніше в роботах інших авторів. Побудовано інформаційну модель методу розрахунку каппа-критерію алгоритмів методу максимуму компактності для випадку впровадження принципу самоорганізації моделей у програмних продуктах, що реалізовують представлення чисел у 16-річній системі обчислення з плаваючою комою.

Вступ. Актуальність роботи, її зв’язок з науковими проблемами. Статистичне моделювання отримує широке розповсюдження як ефективний метод вирішення низки практичних та теоретичних завдань. Особливостями реалізації методів статистичного моделювання є необхідність багатократного отримання числових значень, які характеризують динаміку процесу, що досліджується, шляхом розв’язування рівнянь, які призначені для аналітичного опису процесу для різних умов. У цьому випадку оцінка похибки підсумків моделювання має імовірнісний характер.

Значення ускладнення аналітичних моделей [1] під час розрахунків призводить до погіршення точності обчислень. Проблема поліпшення точності обчислень під час статистичної обробки даних вимагає врахувань похибки вихідних даних, що призводить до необхідності вирішення некоректних завдань [2]. Алгоритми розв’язку некоректних задач реалізують чисельні методи математики [3], які, в свою чергу, є основою алгоритмів та методів синтезу прикладних програмних продуктів статистичної обробки даних [4, 5, 6].

Досвід останніх років вказує на один з можливих перспективних виходів зі становища, що полягає в комбінованому застосуванні чисельних та аналітичних методів [1, 7, 8].

Одним з методів моделювання, що реалізує зазначений підхід, є метод максимуму компактності (ММК). Таким чином, пошук шляхів вирішення проблеми забезпечення точності та достовірності статистичного моделювання за рахунок розвитку алгоритмів ММК є актуальним науковим завданням.

Аналіз попередніх досліджень, постановка завдань. Першою спробою вирішення завдання створення комбінованих алгоритмів моделювання для побудови нефізичних моделей складних систем [9] можна вважати низку робіт [10, 11], в яких вперше реалізовано та обґрунтовано застосування чисельних та аналітичних методів для побудови регресійних моделей випадкових процесів.

Істотним обмеженням на використання методу групового урахування аргументів (МГУА) під час створення комерційних програмних продуктів стало використання декількох критеріїв відбору остаточної моделі [12]. Саме це вимагало постійного втручання оператора в процес відбору моделі [13] та виключало можливість створення повністю автоматичних систем (автономних програм).

Введення каппа-критерію еквівалентності випадкових моделей [14, 15] та розробка ММК створило передумови для розроблення повністю автоматизованих систем підтримки прийняття рішень (СППР) з прогнозуванням та відповідних програмних продуктів.

Однак можливість порушення передумов можливості застосування ймовірнісно-статистичних методів була визнана лише на початку 80-х рр. минулого сторіччя у зв’язку з різким зростанням вимог до точності під час розробки сучасних інформаційних технологій у фундаментальних дослідженнях. До цього часу в розробці науково-методичного забезпечення розв’язання початкових задач математичної статистики було досягнуто беззаперечного прогресу. Варто згадати схему крос-перевірки.

З усвідомленням цього факту вже в 1990–1992 рр. отримано ряд результатів, що дозволили розробити рекомендації врахування статистичної неоднорідності даних і одним з яких стало застосування схеми перехресного іспиту. Найбільш відомими методами, що використовують схему перехресного іспиту та отримали поширення в прикладних дослідженнях в Україні, є МГУА і ММК моделей. Однак роботи з розвитку ММК та дослідження властивостей його алгоритмів припинено наприкінці 1990 р. у зв’язку з відсутністю фінансування та обмеженнями, що були викликані недостатнім розвитком засобів обчислювальної техніки.

Однак у [16] вперше запропоновано реалізацію алгоритмів ММК мовою програмування Бейсік.

Визначення частини невирішених завдань. Завдання статті, умови дослідження. Завданням статті є визначення структурних моделей алгоритмічного забезпечення програмних продуктів та інформаційних технологій СППР з прогнозуванням технічного стану складних об'єктів, відмова або несправний стан яких може призвести до катастрофічних наслідків.

Обмеженнями на реалізацію розроблених алгоритмів визначено такі вимоги:

– програмні продукти, що реалізують розроблені методи та алгоритми, створюються мовами програмування високого рівня з представленням чисел у 16-річній системі обчислень з плаваючою комою;

– забезпеченість гарантованого результату прогнозу в межах визначеної похибки на період прогнозної ретроспекції, що визначається відповідним інтервалом врахованої передумови;

– моделі алгоритмічного забезпечення відповідних практичних результатів [10] повинні передбачати процедури, які реалізують методи та способи виключення виродження задачі ідентифікації параметричної моделі прогнозу в класі степеневих рядів;

– для забезпечення точності аналітичних моделей, які будуються в результаті роботи відповідних процедур моделювання, використовується ідея самоорганізації моделей у схемі їх перехресного відбору.

Викладення основного матеріалу. Виникає завдання визначення полінома оптимальної складності й необхідної кількості вимірювальної інформації для його ідентифікації.

Розв'язання цих задач, як правило, пов'язане зі статистичною обробкою даних, отриманих у результаті проведених вимірювань за допомогою формальних розрахованих схем, серед яких найбільш часто використовується метод найменших квадратів (МНК), а оптимізація складності інтерполяційного полінома проводиться за результатами порівняння суми квадратів залишкових відхилень. Проте оцінки, які при цьому отримують, забезпечуються відносно складними обчисленнями, крім того, МНК оцінки чутливі до неоднорідностей у вибірці, а іноді й до помилок округлення.

Якщо як мірою статистичної стійкості оцінюваних статистичних характеристик аналізованих даних вдається скористатися розподілом похибок відповідним чином організованого інтерполяційного процесу, то в такій ситуації обчислювальну схему обробки можна спростити. При цьому показником якості апроксимації випадкових функцій $f(\Theta, Z, x)$ служить розподіл помилок апроксимації, якому відповідає деяка функція цього розподілу, що характеризує його компактність, – функція компактності.

Основою будь-яких алгоритмів чи процедур, що використовують для побудови моделей ММК, є побудова екстраполяційного функціонала в розрахунковій схемі перехресного відбору моделей. Побудова екстраполяційного функціонала передбачає виконання операцій за такими етапами:

– розподіл вихідної вибірки даних на (m+n) підгруп, m груп даних формують пробну вибірку, n – контрольну вибірку з (m+1) варіантом перебору, кількість груп даних в одному варіанті перебору формують велику контрольну вибірку, або, за [18, 19], величину контрольного "вікна", цей крок формує базу вихідних даних, яка буде поступово збільшуватися за рахунок збільшення кількості елементів;

– побудова на кожній пробній вибірці (m+1) варіанта m – параметричної моделі. Цей крок формує базу даних моделі, яка може змінюватися;

– екстраполяція отриманих моделей на контрольну вибірку (контрольні вибірки в разі реалізації процедур змінного контрольного вікна), формування бази даних екстраполяцій;

– об'єднання екстраполяцій в екстраполяційний функціонал, відхили екстраполяційного функціонала характеризують невиключену систематичну похибку (НСП) моделі, а розподіл вибірки відносно екстраполяційного функціонала, або функція компактності, характеризує випадкову складову систематичної похибки (ВССП); числові значення екстраполяційного функціонала виступають головним критерієм відбору моделі;

– структурна ідентифікація моделі за критерієм ММК: максимуму каппа-статистики [20] з урахуванням НСП, похибок округлення, похибок, що викликані системою представлення чисел;

– остаточна параметрична ідентифікація структурованої моделі.

У [17] зазначено, що якщо виділяти окрему процедуру формування пробних та контрольних вибірок, можна передбачити варіацію самих класів моделей. Такий підхід до формування структури інформаційної технології реалізації СППР з прогнозуванням дозволить використовувати одні програмні модулі як для побудови моделей дрейфу в класі степеневих рядів, так і моделей розподілу. Під час побудови моделей розподілу та варіації їх класів шляхом вибору класу моделей, що найбільш підходить, при гіпотезі про одиничність компонента в моделі [21] усувається проблема стійкості статистичних оцінок при невеликих відхиленнях від нормального [4].

В загальному випадку сенс проблеми полягає в різному значенні оцінок МНК навіть для симетричного гауссівського розподілу зі збільшенням ваги "хвостів" розподілу. Відносна ефективність оцінки середньоарифметичного значення незалежних випробувань, що використовується як статистичний аналог математичного очікування, наведена в [4].

Ще одна проблема, яка виникає при вирішенні прикладних завдань, може бути розв'язана шляхом варіації самих методів та їх алгоритмів параметричної ідентифікації (МНК, МНМ тощо). Наприклад, при припущенні про нормальний вигляд розподілу та одиничність компонентів моделі знаходження значень параметрів моделі методом найменших модулів (МНМ) чи МНК просто неможливе.

Така можлива інваріантність структури алгоритму реалізації ММК до складових розрахункових модулів (процедур) до кінцевого результату забезпечує широкий клас задач, що розв'язуються. Це в [4] дозволило розглядати ММК як генератор прикладних алгоритмів.

Зазначене вище дозволяє стверджувати, що базовим алгоритмом, який безпосередньо формує спосіб реалізації ММК в прикладних програмах та інформаційних технологіях статистичної обробки даних, є алгоритм реалізації рекурсивної процедури формування пробних та контрольних вибірок.

Припустимо, що на першому кроці наближення модель процесу являє собою функціональну залежність процесу, який моделюється від однієї змінної *x*:

$$y = f(\overline{\Theta}_m, x),$$

де $\overline{\Theta}_m$ – m -вимірний вектор параметрів моделі.

Вихідна вибірка являє собою множину дійсних чисел:

$$Y_x = \{y_i(x_i); i = 1...N\}. \tag{1}$$

Множину (1) можна представити двома одновимірними масивами (векторами):

$$Y = \{y_i; i = 1...N\}, \quad X = \{x_i; i = 1...N\}.$$

До процедури формування пробної $Y_{пр}; X_{пр.}$ та контрольної $Y_{контр.}; X_{контр.}$ вибірок висуваються такі вимоги за аналогією з [17, 18, 19]:

$$Y_{пр.} \cup Y_{контр.} = Y; \quad Y_{пр.} \cap Y_{контр.} = \varnothing$$

$$Y_{контр.} = \{y_{kj}; j = 1...N_k\};$$

$$Y_{пр.} = \{y_{nL}; L = 1...N_L\};$$

$$N_k + N_L = N$$

відповідно до X. Таким чином, слід встановити значення впорядкованих за індексами масивів $Y_{пр.}; X_{пр.}; Y_{контр.}; X_{контр.}$, що не викликає великих труднощів, якщо, наприклад, встановлені значення масиву $S = \{S(j); j = \overline{1}, \overline{N}\}$ індексів елементів $[S(j)+1] \in \{1,2,...N\}$ вихідної вибірки, що складають контрольну. Тоді для пробної вибірки необхідно встановити значення масиву, аналогічного S, що вже пов'язано з суттєвими труднощами алгоритмічного характеру, особливо якщо маємо j, при яких $|S(j) - S(j+1)| > 1$.

Уникнути вказаних труднощів дозволяє введення індикаторної функції пробної вибірки:

$$I_L = \{e(j); L = 1,..., N\}, \quad e(j) \in \{1,0\},$$

нулі якої відповідають індексам контрольної та індикаторної функцій контрольної вибірки:

$$I_k = \{eL(i); K = 1,..., \overline{N}\}, \quad eL(i) = 1 - e(L),$$

визначених на множині індексів $1,2,...,N$. У цьому випадку формування масивів пробної та контрольної вибірок буде забезпечуватися через встановлення значень відповідних індексів за сумами:

$$j = \sum_{l=1}^{i} eI(l), \quad h = \sum_{l=1}^{i} e(l).$$

У [17] для забезпечення можливості програмної реалізації алгоритму пропонується введення додаткової процедури з використанням штучно створених елементів вибірки. Однак з використанням бази даних елементів масивів вихідної статистики ця необхідність відпадає.

Процедура формування пробних і контрольних вибірок у рамках розрахункової схеми ММК потребує генерації індикаторних функцій $I_L; I_k$; використання програмного продукту Microsoft Office Excel пропонується для відповідного способу реалізації алгоритму мовами високого рівня в операційній системі Microsoft Windows.

Програмний модуль розрахунку можливих варіантів структур реалізовується за алгоритмом, наведеним на рисунку 1.

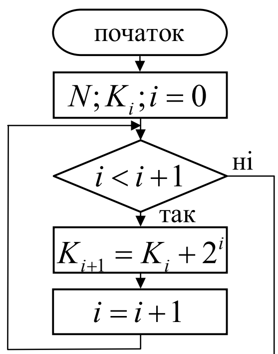

Рис. 1. Алгоритм програмного модуля розрахунку можливих варіантів структур моделей у класі степеневих рядів

Загальна кількість варіантів структур визначається за співвідношенням:

$$K_{i+1} = K_i + 2^i; i = 0...N,$$

де K_i – кількість варіантів структур на попередньому кроці алгоритму;

N – максимальний степінь апроксимаційного полінома (визначається кількістю членів вибірки).

Параметри N; K_i та i реалізуються як цілі числа довжиною 4 байта.

Якщо передбачати варіацію кількості членів у вибірках у межах від $N-1$ до 1, можна отримати схему випробувань базового алгоритму ММК при ідентифікації моделі в класі степеневих рядів на вихідній вибірці.

Розширення реалізацій ММК проводиться шляхом застосування різноманітних процедур формування пробної та контрольної вибірок. Наприклад, застосування процедури формування $m+1$ рівних підвибірок та їх почергове використання як контрольної з формуванням відповідного екстраполяційного функціонала призводить до методу, що пропонується в [22]. Це підтверджує коректність процедур, що пропонуються, і дозволяє говорити про виявлення загальних закономірностей у різних моделях інформаційного процесу, які полягають у можливості управління процесом шляхом обмеження розміру контрольних вибірок та їх кількості.

Перейдемо тепер до розгляду випадку ідентифікації моделі функціональної залежності параметра φ від двох змінних x_1, x_2:

$$y = \varphi(\Theta_{m_1}, \Theta_{m_2}, x_1, x_2)$$

на множині даних

$$Y = \{y_{ij}; i = \overline{1}, \overline{N}; j = \overline{1}, \overline{N}\};$$

$$X_1 = \{x_{1i}; i = \overline{1}, \overline{N}\};$$

$$X_2 = \{x_{2j}; j = \overline{1}, \overline{N}\}.$$

Загальний вигляд функціональної структури ММК-генератора наведено на рисунку 2.

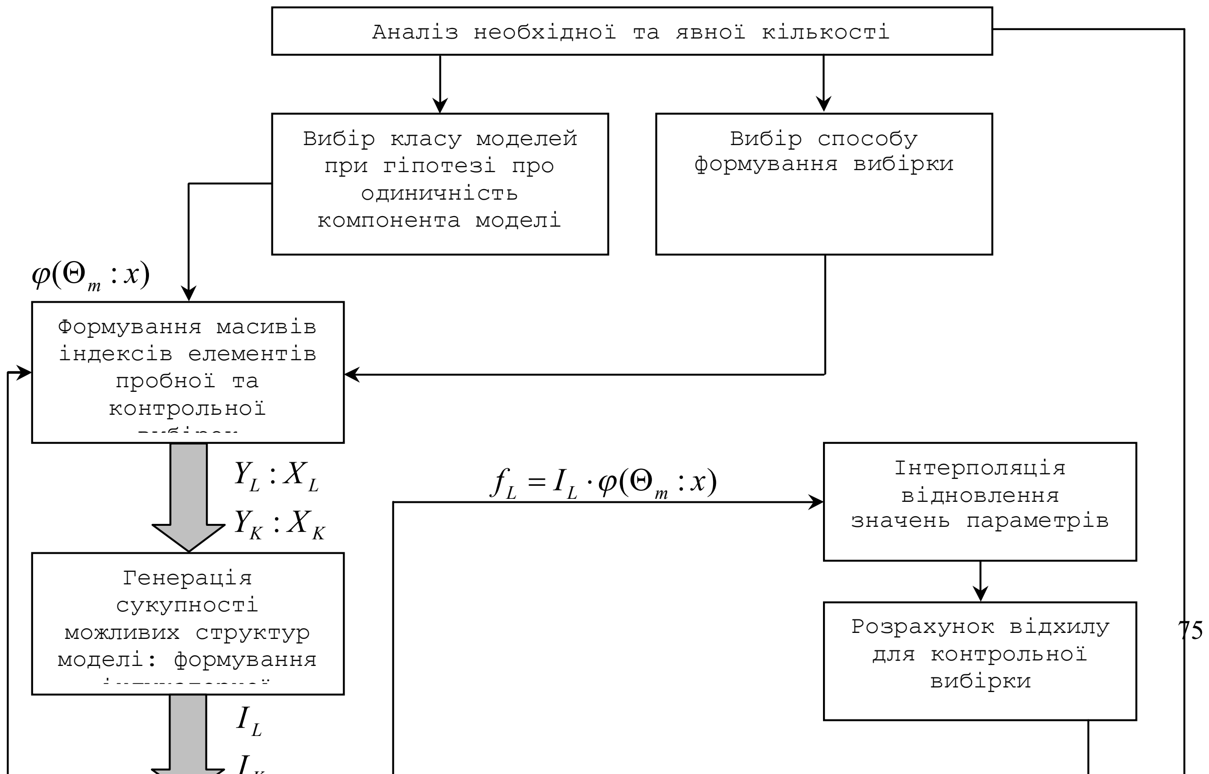

Рис. 2. Інформаційна модель методу розрахунку каппа-критерію алгоритмів ММК для випадку впровадження принципу самоорганізації моделей

Очевидно, що при фіксації в φ для однієї зі змінних даний випадок зводиться до попереднього, і для того, щоб скористатися алгоритмом, що пропонується на рисунку 1, необхідно ввести двовимірну індикаторну функцію пробної

$$I_{пр.} = \left\{ e(f, K_f); f = 1{,}2; K_f = \overline{1, N_f} \right\}, \quad e(f, K_f) \in \{1;0\}$$

і контрольної

$$I_{контр.} = \left\{ e1(f, K_f); f = 1{,}2; K_f = \overline{1, N_f} \right\}, \quad e1(f, K_f) = 1 - e(f)$$

вибірок. Тоді перетворення масивів Y, X_1, X_2, що відповідають перетворенню (1), буде відбуватися через встановлення значень індексів:

$$jf_{пр.} = \sum_{l=1}^{K_f} e(f, l), \qquad jf_{контр.} = \sum_{l=1}^{K_f} e1(f, l),$$

де будь-який елемент з нульовим індексом є резервним.

Це доповнює процедуру формування структури кількістю циклів за кількістю змінних, яка в принципі може бути скільки завгодно великою.

Висновки. Введення індикаторної функції пробної та контрольної вибірок до розрахункової схеми алгоритмів ММК дозволяє:

1) реалізувати базовий алгоритм ММК-генератора на достатньо простих процедурах;

2) забезпечити рекурсивність базового алгоритму за кількістю змінних, що досліджуються, що дасть змогу розширити логіку ММК на статистичний аналіз багатофакторних моделей;

3) застосовувати стандартні пакети прикладних програм відновлення функціональних залежностей на дослідних даних.

Напрямки подальших досліджень. Напрямками подальших робіт є постановка та проведення розрахункових експериментів з визначення стійкості алгоритмів методу максимуму компактності до визначення структури поліноміальної моделі та їх завадостійкості до зашумлених статистичних даних.

ЛІТЕРАТУРА:

1. *Левин С.Ф.* Комбинированный метод статистического моделирования / *С.Ф. Левин.* – М. : АН СССР, 1978. – 75 с.
2. *Бахвалов Н.С.* Численные методы / *Н.С. Бахвалов, Н.П. Житков, Г.М. Кобельков.* – М. : Наука, 1992. – 164 с.

3. Машинные методы математических вычислений / *Дж. Форсайт* и др. – М. : Мир, 1980. – 335 с.
4. *Айвазян С.А.* Прикладная статистика: основы моделирования и первичная обработка данных / *С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин.* – М. : Финансы и статистика, 1983. – 471 с.
5. *Айвазян С.А.* Прикладная статистика: исследование зависимостей / *С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин.* – М. : Финансы и статистика, 1985. – 487 с.
6. Прикладная статистика: классификация и снижение размерности / *С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин.* – М. : Финансы и статистика, 1989. – 607 с.
7. *Лагутин С.А.* Статистические модели и методы в измерительных задачах [Електронний ресурс] / *С.А. Лагутин, М.В. Петухов.* – Режим доступу : www.pti.ac.ru\other\\conf\rus
8. Відповідність характеристик екстраполяційного функціонала похибкам моделей фізичних процесів : зб. наук. пр. з матеріалами I міжнар. наук.-практ. конф. “Безпека та захист інформації в інформаційних та телекомунікаційних системах”. – Харків : Харківський нац. економ. ун-т, 2008. – № 7. – С. 70–71.
9. *Ивахненко А.Г.* Особенности применения метода группового учета аргументов в задачах прогнозирования случайных процессов / *А.Г. Ивахненко, В.С. Степашко* // Автоматика. – 1986. – № 5. – С. 3–14.
10. *Івахненко О.Г.* Поліноміальна і логічна теорія динамічних систем. Частина II / *О.Г. Івахненко, Ю.В. Копа, Ву Суан Минь* // Автоматика. – 1970. – № 4. – С. 17–23.
11. *Івахненко О.Г.* Поліноміальна і логічна теорія динамічних систем. Частина I / *О.Г. Івахненко, Ю.В. Копа, Ву Суан Минь* // Автоматика. – 1970. – № 3. – С. 26–36.
12. *Ивахненко А.Г.* Объективная компьютерная кластеризация / *А.Г. Ивахненко, Н.А. Петухова, Н.А. Ивахненко* // Автоматика. – 1986. – № 3. – С. 3–11.
13. *Левченко А.А.* Идентификация модели параметра потока отказов при многорежимном содержании радиотехнических средств / *А.А. Левченко, О.И. Кравчук* // Вісник Черкаського державного технологічного університету. – 2006. – № 2. – С. 23–26.
14. *Левин С.Ф.* Верификация экспертных и использующих экспертные оценки систем, ориентированных на вероятностно-статистические методы в программах обеспечения эксплуатации аэрокосмической техники / *С.Ф. Левин* // Проблема разработки и внедрения экспертных систем. – М. : ВНИИМС, 1989. – С. 144–145.
15. *Левин С.Ф.* Основы теории контроля / *С.Ф. Левин.* – М. : МО СССР, 1983. – 51 с.
16. *Левин С.Ф.* Статистический анализ систем обеспечения эксплуатации технических объектов / *С.Ф. Левин.* – М. : Изд. АН СССР, 1989. – 432 с.
17. *Блинов А.П.* Научно-методическое обеспечение гарантированности решения метрологических задач вероятностно-статистическими методами / *А.П. Блинов, С.Ф. Левин* // Измерительная техника. – 1988. – № 12. – С. 8–12.
18. *Левченко А.О.* Процедура виявлення структурних змін моделей дрейфу нормованого параметра ОВТ / *А.О. Левченко* // Військово-технічний збірник АСВ. – 2009. – № 1(1). – С. 69–73.
19. *Левин С.Ф.* Гарантированность программ обеспечения эксплуатации техники / *С.Ф. Левин.* – К. : Знание, 1989. – 23 с.
20. А. с. 32413, МКИ В 25 J 15/00. Спосіб апроксимації щільностей розподілу ймовірностей випадкових процесів багатомодового вигляду / *А.О. Левченко, В.А. Багінський.* – № 32413 ; заявл. 17.03.10 ; опубл. 06.05.2010, Бюл. № 05.
21. *Левченко А.А.* Теоретические вопросы моделирования и оценки качества систем обеспечения эксплуатации сложных технических комплексов / *А.О. Левченко* // Системи озброєння та військової техніки. – 2007. – № 3(11). – С. 119–124.
22. *Блинов А.П.* Построение градуировочных характеристик средств измерений методом максимума компактности / *А.П. Блинов* // Измерительная техника. – 1987. – № 7. – С. 15–17.

ЛЕВЧЕНКО Андрій Олександрович – кандидат технічних наук, старший науковий співробітник, доцент, заступник начальника наукового центру Сухопутних військ Академії сухопутних військ ім. гетьмана Петра Сагайдачного.

Наукові інтереси:
- моделювання складних систем;
- перевірка достовірності отриманих моделей.

Тел.: (068)505–05–55; (067)776–07–53.

E-mail: AndreyLevchenko2007@yandex.ru

Подано 08.07.2010

Левченко А.А. Шляхи реалізації базового алгоритму методу максимуму компактності як генератора структури прикладних програм статистичної обробки даних
Левченко А.А. Пути реализации базового алгоритма метода максимума компактности как генератора структуры прикладных программ статистической обработки данных
Levchenko A.A. Ways of realisation of base algorithm of a method of a maximum of compactness as generator of structure of applied programs of statistical data processing

УДК 621.396

Пути реализации базового алгоритма метода максимума компактности как генератора структуры прикладных программ статистической обработки данных / А.А. Левченко

В статье рассмотрены пути реализации базового алгоритма метода максимума компактности как генератора структуры прикладных программ статистической обработки данных. Приведены результаты анализа трансформации общего случая к случаям опубликованным раньше в работах других авторов. Построена информационная модель метода расчета каппа-критерия алгоритмов ММК для случая внедрения принципа самоорганизации моделей в программных продуктах, которые реализовывают представление чисел в шестнадцатиричной системе исчисления с плавающей запятой.

УДК 621.396

Ways of realisation of base algorithm of a method of a maximum of compactness as generator of structure of applied programs of statistical data processing / A.A. Levchenko

In article ways of realisation of base algorithm of a method of a maximum of compactness as generator of structure of applied programs of statistical data processing are considered. Results of the assaying of transformation of a blanket case are led to cases published earlier in works of other authors. The information sample piece of a method of calculation of kappa-criterion of algorithms method of a maximum of compactness for a case of introduction of a principle of self-organising of sample pieces in software products, which realisation representation of numbers in hexadecimal system of calculation from a floating comma is constructed.