УДК 621.396

Ірина Робертівна Мальцева
Юрій Олександрович Процюк
Юлія Олександрівна Черниш
Максим Георгійович Тищенко

ПРИНЦИПИ ВИЗНАЧЕННЯ НАДІЙНОСТІ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ЧЕРЕЗ ВИКОРИСТАННЯ ТОЧНОСТНОЇ ТЕОРІЇ

Постановка проблеми

Потенційним джерелом небезпеки і ризику в складних системах є їх інформаційні помилки. У цій статті разом з класичним підходом побудови теорії надійності програмного забезпечення запропонований підхід, який ґрунтується на дослідженні точності його операторів.

Введення

"Обчислювальні машини призначені загалом для запису чисел, для дій над ними і видачі результату в числовій формі" [1]. Вони були створені на основі успіхів математики і в інтересах рішення завдань математики. Їх переваги - швидкодія і висока точність обчислень - зіграли величезну роль в рішенні ряду важливих практичних завдань найрізноманітнішого характеру.

Нині практично немає такої галузі науки і техніки, де б не використовувалися ці досконалі технічні засоби. Бурхливий розвиток техніки зв'язку і теорії інформації багато в чому зумовлений прогресом в області обчислювальної техніки. Зараз засоби обчислювальної техніки призначені не лише для виробництва складних і якнайточніших розрахунків і управління різними об'єктами, але і для представлення інформації в нечисловому, кількісному, смисловому виді. Проте, в якому б виді інформація не була представлена обчислювальними засобами, вона завжди у своїй основі являється числовою, цифровою.

Подальша цивілізація суспільства вимагає рішення усе більш складних завдань. Це неминуче призводить до лавиноподібного зростання складності алгоритмів рішення цих завдань, програмного забезпечення і самих обчислювальних засобів. Складність алгоритмів і програм рішення завдань - неминуче джерело помилок обчислень. Крім того, наші первинні виміри будь-яких величин, даних, які вводяться в обчислювальні засоби, містять помилки. Наприклад, константи π і e в комп'ютері, на якому набраний цей текст, представлені після коми усього лише 14 цифровими знаками. В результаті цього, ми вимушені звертати увагу на точність обчислень і разом з надійністю техніки займатися надійністю обчислень.

Сучасна теорія і практика надійності програмного забезпечення (точніше, надійності функціонування програмно-керованих обчислювальних засобів або просто обчислювальних засобів, дія яких не можлива без програм) ґрунтується на ймовірнісному підході, що використовує статистичні дані про помилки (початкових даних, самих програм, техніки, людини-оператора). Вона використовує великий ряд моделей і методів дослідження надійності програмного забезпечення (програм) [2,3]. Проте можливості такого підходу дуже обмежені із-за мізерності статистичних даних про помилки програм. І, слід зазначити, що цей статистичний дефіцит помилок у міру удосконалення елементної бази, техніки і якості програм, зростанням автоматизації, усе більш позначатиметься. Але це не означає, що з помилками програм буде покінчено.

Будь-яке обчислення пов'язане з певною точністю. Чим складніше обчислення, тим менше його точність. Складне обчислення складається з простих, малих обчислень. Складна програма складається з великої кількості елементарних операторів. Конкретна реалізація траєкторії обчислювань має певну точність, зумовлену складом, точністю операторів і точністю округлення результату обчислення кожного оператора.

Мета статті

Ми повинні звернути увагу дослідників на точностний аспект складних обчислень, показати елементарним аналізом, як змінюється точність обчислень залежно від складності програми, визначивши ймовірність виникнення помилки в простій, гіпотетичній програмі, і поставити завдання дослідження на точність усіх операторів, з яких складаються складні програми.

Виклад основного матеріалу

Найпростіший лінійний оператор.

Розглянемо властивості оператора:

$$Y = K \quad X. \tag{1}$$

Для простоти дослідження покладемо, що величини X Y детерміновані і довільні, а константа K – випадкова величина з щільністю розподілу

$$f_K(k), \quad a \le k \le b. \tag{2}$$

Знайдемо щільність розподілу тепер уже випадкової величини Y вона дорівнює:

$$f_y(y) = \frac{1}{x} \cdot f_k\left(\frac{y}{x}\right), \ xa \le y \le xb. \tag{3}$$

Припустимо, що точність величини K визначена або задана як:

$$K \le |\Delta_k| \text{ або } -\Delta_k \le K \le \Delta_k. \quad (4)$$

Тоді в обчислення вноситься помилка з приводу випадковості величини K. Ймовірність виникнення такої помилки дорівнює:

$$P_k(\Delta_k) = 1 - \int_{\nu_k - \Delta_k}^{\nu_k + \Delta_k} f_k(u)du, \quad (5)$$

де ν_k – математичне очікування величини K. Нехай, наприклад:

$$f_k(k) = \begin{cases} \frac{1}{b-a}, & a \le k \le b, \\ 0, & k \langle a, \ k \rangle b. \end{cases} \quad (6)$$

З (1) знайдемо:

$$P_k(\Delta_k) = 1 - \frac{2\Delta_k}{b-a}. \quad (6)$$

При $\Delta_k = 0$ ймовірність помилки дорівнює 1, а при $\Delta_л = \frac{b-a}{2}$ -0, що і слід було очікувати.

Поступаючи так само, знайдемо вірогідність появи помилки у вихідній величині якщо межі точності її задані величиною $|\Delta_y|$:

$$P_y(\Delta_y) = 1 - \int_{\nu_k \cdot x - \Delta_y}^{\nu_k \cdot x + \Delta_y} \frac{1}{x} f\left(\frac{u}{x}\right)du. \quad (7)$$

Зокрема, при рівномірному розподілі величини K (5) знайдемо:

$$P_y(\Delta_y) = 1 - \frac{2\Delta_y}{x(b-a)}. \quad (8)$$

Надалі припускаємо, що $\Delta_y = \Delta_k$, тобто точність вихідної величини дорівнює точності представлення величини K. Порівняємо(6) і (8), вводячи їх відношення, при рівномірній щільності:

$$G(\Delta_k) = \frac{P_y(\Delta_k)}{P_k(\Delta_k)} = (1 - \frac{2\Delta_k}{x(b-a)})/(1 - \frac{2\Delta_k}{(b-a)}). \quad (9)$$

З(9) витікає, що тільки при $x = 1$ ймовірності обох помилок однакові. Якщо, $x \rangle 1$, то ймовірність помилки на виході оператора більша, ніж ймовірність помилки константи K і ця перевага збільшуватиметься тим більше, чим буде більше діапазон межі точності Δ_k. Зі збільшенням x міра цього зростання буде більша. Звідси слідує висновок: за обумовлених умов простий лінійний оператор з випадковим коефіцієнтом передачі збільшує ймовірність помилки у вихідній величині. Це справедливо і відносно впливу на Y величини X, коли вона стає випадковою. Проте, спільний вплив похибок K, X на Y тут розглядати не будемо.

Якщо $x \langle 1$, відношення $G(\Delta_k)$ зі збільшенням Δ_k зменшується, беручи свій початок по колишньому з x при $\Delta_k = 0$.

Розглянемо інформаційну властивість цього оператора. Ентропія випадкової величини K буде рівна - $H_k = \ln(b-a)$, а ентропія випадкової величини Y – $H_y = \ln(x(b-a))$. Введемо і знайдемо передатну функцію оператора Y по ентропії:

$$W_y = \frac{H_y}{H_k} = 1 + \frac{\ln x}{\ln(b-a)}. \quad (10)$$

З формули(10) виходить, що невизначеність значень вихідною змінною Y по відношенню до невизначеності K зростає при $x \rangle 1$ і зменшується при $x \langle 1$. Якщо визначити приріст кількості інформації на виході оператора по відношенню до входу, то матимемо [4]:

$$I_y = H_k - H_y = -\ln J\left|\frac{dy}{dk}\right| = -\ln x, \quad (11)$$

де $J\left|\frac{dy}{dk}\right|$ – якобіан перетворення. З формули(11) виходить: при $x \rangle 1$ приріст кількості інформації негативний, а при $x \langle 1$ – позитивний, що узгоджується з(10).

У більшості літературних джерел стверджується, що кількість інформації не може бути негативна. Проте, це не узгоджується з рядом джерел, наприклад [5], у якому під кількістю інформації розуміється

$$I = \ln \frac{P_A}{P_B}, \quad (12)$$

де P_A, P_B – ймовірність настання події після досвіду і до досвіду(апостеріорна і апріорна ймовірності). І якщо апостеріорна ймовірність менша за апріорну, то кількість інформації негативна. Слід зауважити, що в [6] визначення(12) розуміється як не кількість інформації, а її цінність.

Ми дотримуватимемося визначення(12) і в якості виправдання цього факту розглянемо два " полярні" розподіли ймовірності : δ – дельта-функцію і рівномірну щільність, визначену на інтервалі $[0, \infty)$.Тоді ентропія і, відповідно, кількість інформації можуть набувати значень в інтервалі $(-\infty, \infty)$. І якщо під кількістю інформації розуміти міру знятої невизначеності, то можна стверджувати, що кількість інформації для безперервних розподілів, може бути і негативна.

Знайдемо величину ентропії на вході і виході оператора Y:

$$\begin{aligned} H_k &= -(\alpha \ln \alpha + (1-\alpha)\ln(1-\alpha)), \\ H_y &= -(\frac{\alpha}{x}\ln\frac{\alpha}{x} + (1-\frac{\alpha}{x})\ln(1-\frac{\alpha}{x})), \end{aligned} \quad (13)$$

де $\alpha = \frac{2\Delta_k}{b-a}$. Вважаємо, що в нашому випадку це досить мала величина. Скористаємося формулами(10) і (11), підставляючи в них співвідношення(13). Отримаємо при $x \rangle 1$ $W_y \langle 1$ і тим менше, чим більше x. В цьому випадку $I_y \rangle 0$, і тим більше, чим більше величини x. При $x \langle 1$

$W_y \rangle 1$ і тим більше, чим менше x. А величина $I_y \langle 0$ і тим менше, чим менше x. На наш погляд, це і відповідає реальній дійсності. У першому випадку, коли $x \rangle 1$, оператор знижує невизначеність результату, а в другому випадку, коли $x \langle 1$ – збільшує її. Погрішність в K при великому значенні x менш значима для Y, напроти, при малому значенні вона стає значимішою для Y. і може стати навіть сумірною з Y. Тут усюди під x розуміється значення змінної X оператора Y, як аргумент функції.

Неважко уявити і той факт, що якщо маємо два, що послідовно діють оператора

$$Y_2 = K_2 Y_1, \quad Y_1 = K_1 X, \qquad (14)$$

то при випадкових величинах K_1, K_2 описані вище точностний і інформаційний ефекти посилюватимуться.

Найпростіша складна програма

Реальна програма досить складного обчислювального процесу може бути представлена у вигляді графа Г(v,d,P,R), де v - безліч вершин (операторів), d - безліч дуг (переходів між операторами), P - безліч ймовірностей реалізації операторів, R - безліч ймовірностей переходів між вершинами. При певних значеннях початкових даних на входах графа, значеннях констант операторів, може існувати на графі одна і тільки одна траєкторія дотримання обчислювального процесу. У кожній вершині графа відбувається вибір наступної вершини, до якої здійснюватиметься перехід процесу. Цей вибір визначається виконанням або невиконанням предикативної умови, властивої тільки цій вершині. Увесь процес обчислень відбувається в часі.

Провести аналіз точностних і інформаційних властивостей для складних програм, що описуються подібними графами, не є можливим. Тому розглянемо одну просту складну гіпотетичну програму. Така програма міститиме n однотипних лінійних операторів, що розглянуті в п.2, послідовно реалізовуються від початку процесу до його закінчення. Вважаємо, що результат обчислень буде отриманий тільки тоді, коли все n операторів виду $Y = K \cdot X$ один за іншим будуть реалізовані. При цьому результат обчислень попереднього оператора є значенням вхідної змінної для наступного оператора.

У усіх операторів коефіцієнти K_i, $i = 1,2...n$, випадкові, підпорядковані одному і тому ж розподілу. Але при реалізації кожного елементарного оператора вибір значення константи K_i робиться наново, незалежно від вибору значень її попередників. Таким чином, наше завдання аналізу властивості програми зводиться до підсумовування помилок усіх операторів, що становлять програму. У реальних програмах кожен елементарний оператор має свої константи по точності, що характеризуються своїми розподілами

Проте, в подібному виді рішення задачі аналізу буде математично дуже громіздким. Нам же необхідно отримати загальні, принципові висновки про властивість складної програми.

Згідно [6] "ймовірність того, що при складанні великого числа наближених чисел, закруглених за звичайним правилом, помилки не накопичуються, а, навпроти, компенсуються", досить велика.

Але ми підійдемо до підсумовування випадкових величин знову з тієї ж позиції, яка передбачає вивчення щільності ймовірності суми випадкових величин, визначення ймовірності помилки при підсумовуванні, якщо межі точності доданків і результату задані. Як і раніше вважаємо, що вони однакові і рівні Δ.

Підсумовуючи n випадкових величин з щільністю вірогідності(2), в перетворенні Лапласа отримуємо:

$$f_y^*(sx), \qquad (15)$$

$$\left[f_y^*(sx)\right]^n. \qquad (16)$$

З(16) знайдемо математичне очікування, середньоквадратичне відхилення і коефіцієнт варіації вихідного значення програми :

$$\nu_n = nx\alpha\nu_k, \quad \sigma_n = \sqrt{n}x\sigma_k, \quad \eta_n = \frac{1}{\sqrt{n}}\eta_k, \qquad (17)$$

де ν_k, σ_k, η_k – математичне очікування, середньоквадратичне відхилення і коефіцієнт варіації константи K окремого оператора.

Для апроксимації щільності ймовірності результату реалізації програми скористаємося нормальним розподілом з параметрами(17).

Знайдемо ймовірність помилки програми залежно від її складності:

$$P_y(n) = 1 - \frac{1}{\sqrt{2\pi n}x\sigma_k}\int_{nx-\Delta}^{nx+\Delta} e^{-\frac{(u-nx\nu_k)^2}{2nx^2\sigma^2{}_k}}du. \qquad (18)$$

Для більшої наочності розглянемо простий чисельний, досить грубий приклад: $x = 5, \nu_k = 3, \sigma_k = 0{,}5, \Delta = 50.$ Отримаємо при значеннях $n = 10,20,30,40$:

$$P_y(10) = 7.827 \cdot 10^{-4}; P_y(20) = 0.500;$$
$$P_y(30) = 0.966; P_y(40) = 0.999.$$

Незважаючи на грубість прикладу, він досить наочно пояснює суть питання, що вивчається, показуючи як швидко росте ймовірність помилки програми залежно від її складності.

Рішення завдань в комп'ютері здійснюється, здавалося б, з дуже високою точністю $(10^{-15}, 10^{-30})$, але і це не змінює висловленого твердження про зростання ймовірності отримання помилкового результату від складності програм. Крім того, твердження посилюється і тим, що змінні завдань наводяться також приблизно, наближеними є округлення обчислень, реалізації умовних переходів, які також призводять до накопичення величини помилки результату.

Що стосується інформаційних властивостей програм, то вони зберігаються в основному такими ж, як і у розглянутого елементарного оператора.

Про побудову точностної теорії надійності програм

Нині основними чинниками, що визначають ненадійність програмного забезпечення, вважаються людський чинник, помилки даних, програмні помилки, відмови техніки [7].

Використовуються аналітичні, статистичні методи оцінювання і методи моделювання. Безперечно це дуже важливо і дає певні позитивні результати для підвищення надійності програмного забезпечення. Але якщо ми хочемо захистити від помилок програмного забезпечення дорогі об'єкти цивільного і військового призначення, необхідно приділяти більше уваги аспекту точності обчислень із-за невизначеності величин їх компонентів.

Класична теорія надійності технічних систем використала метод розчленовування складних об'єктів на елементи розрахунку. При цьому надійність елементів розраховувалася, в основному, за допомогою довідників, в яких наводилися значення інтенсивності відмов елементів. Потім оцінювалася надійність об'єкту і приймалися заходи по її підвищенню.

На наш погляд, для забезпечення надійності функціонування складних програмних комплексів необхідно йти по подібному шляху.

1. Визначати основні прості оператори, з яких складаються програми.
2. Відповідальніше і дуже ретельно вивчати точність представлення різних величин операторів.
3. Досліджувати точностні характеристики операторів при отриманих розподілах величин.
4. Інтенсивніше розвивати моделі дослідження точностних властивостей елементарних операторів.
5. Складати довідники за результатами дослідження операторів, в яких приводити інженерні рекомендації по практичному оцінюванню точності операторів.
6. Розробляти точностні моделі складних програм на основі інтегрування елементарних операторів. Для цього можна використати розроблені графоаналітичні методи і методи моделювання.
7. При побудові складних програм з окремих операторів необхідно виконувати дуже ретельне узгодження виходів операторів з входами подальших операторів, а саме, областей їх завдання. Невиконання цього є джерелом додаткових помилок і втрат інформації[8,9].

Зміст п.7 добре нагадує про те, як важливо погоджувати у багатокаскадному електронному підсилювачі один каскад по виходу з входом наступного каскаду. Невиконання цієї вимоги в граничних областях каскадів призводить до непрацездатності підсилювача в здавалося б заздалегідь передбаченій області зміни параметрів підсилювача. У програмах цей ефект менш очевидний, тому не завжди враховується при їх налагодженні.

Висновки

Здійснення запропонованого підходу до оцінювання і підвищення надійності функціонування програмного забезпечення вимагає істотних витрат різного виду. Але, на наш погляд, його здійснення потрібне. Це зумовлено кінцевою точністю обчислень в сучасних і перспективних комп'ютерах і іншими чинниками, з яких найбільш вагомим є лавиноподібне зростання складності перспективних програм. Іншого шляху немає, поки не будуть створені комп'ютери з надмірністю, що дозволяє організувати оперативний контроль і ухвалення рішень в умовах впливу перешкод різного виду, що адаптуються по точності в процесі обчислень. Повернення до класичної теорії помилок і подальший її розвиток дуже потрібні в інтересах інформатики і інформатизації суспільства. Такий шлях, на наш погляд, може стати джерелом нових конструктивних рішень як в елементній базі, архітектурі обчислювальних засобів, так і в програмному забезпеченні. Він також може стимулювати розвиток нових методів чисельної математики.

У цій статті не наводяться методи дослідження точності реалізації різних математичних операторів і приклади, що ілюструють значення вірогідності виникнення помилок при реалізації операторів.

Література

1. Вінер Н. Кібернетика, або зв'язок в тварині і машині/ Вінер Н // Пер. з Англ.-м.: Сов. радіо – 1958. **2. Тейєр Т.** Надійність програмного забезпечення./ Тейєр Т., Ліпов М., Нельсон Э. // Пер. з Англ.-м.: Світ, 1981. **3. Липаєв В.В.** Надійність програмних засобів / В.В.Липаєв М.: Сінтег, 1998. **4. Тарасенко Ф.П.** Введення в курс теорії інформації. / Ф.П. Тарасенко // Томск, 1963. **5. Красовський А.А.** Основи автоматики і технічної кібернетики. / А.А. Красовський, Г.С.Поспєлов //М.- Л. : Енергоатоміздат, 1962. **6. Крилов А.Н.** Лекції про наближені обчислення / А.Н. Крилов // М.: ГИТТЛ, – 1954. **7. Карповський Є.Я.** Надійність програмної продукції / Є.Я.Карповський, С.А. Чижов // -Київ : Техніка. – 1990. **8. Смагін В.А.** Метод ентропії дослідження інформаційних мережевих структур. / В.А. Смагін // На сайті Sir35.narod.ru - 2002. **9. Смагін В.А.** Про одне джерело втрат інформації. /В.А. Смагін // На сайті Sir35.narod.ru - 2002.

Потенциальным источником угрозы и риска в сложных системах являются их информационные ошибки. В этой статье вместе с классическим подходом построения теории надежности программного обеспечения предложен подход, который обоснован на исследовании точности его операторов.

Ключевые слова: программное обеспечение, надежность, сложность программ, ошибка, эффект накопления, точность вычисления, распределение вероятности, моделирование, прогноз.

The potential source of and risk in the systems are their informative errors. This article deals with both as a classic approach of construction theory of software reliability so as the new offered approach, which is ed on operators' accuracy research.

Key words: software, reliability, complication of programs, error, accumulation effect, accuracy of calculation, allocation of probability, modeling, prognosis .