

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДЛЯ ОЦІНЮВАННЯ ТРИВАЛОСТІ ПРОЄКТІВ З РОЗРОБКИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

УДК 005.8

ПРИХОДЬКО Сергій Борисович

д.т.н., професор, завідувач кафедри Програмного забезпечення автоматизованих систем,
Національний університет кораблебудування імені адмірала Макарова.

Наукові інтереси: математичне моделювання випадкових процесів в інформаційних технологіях.

E-mail: sergiy.prykhodko@nuos.edu.ua.

ПУХАЛЕВИЧ Андрій Володимирович

старший лаборант кафедри Програмного забезпечення автоматизованих систем,
Національний університет кораблебудування імені адмірала Макарова

Наукові інтереси: математичне моделювання випадкових процесів в інформаційних технологіях.

E-mail: andrii.pukhalevych@nuos.edu.ua, a.puhalevich@gmail.com.

ПОСТАНОВКА ПРОБЛЕМИ

На даний час в світі безперервно розробляється нове та модифікується існуюче програмне забезпечення для різних пристроїв та платформ. При розробці програмного забезпечення використовуються різні складні інформаційні технології переробки інформації для оцінювання тривалості програмних проєктів (таких як Comparative Estimating Tool [1] від International Software Benchmarking Standards Group (ISBSG) чи Microsoft Project [2]). Незважаючи на використання вказаних інформаційних технологій не перестає існувати проблема низької достовірності оцінювання тривалості проєктів з розробки програмного забезпечення [3]. Це пов'язано зі складністю сучасного програмного забезпечення та різноманітністю платформ, для яких воно створюється, через що 68% проєктів з розробки програмного забезпечення не виконуються в запланований термін [4]. В існуючих інфор-

маційних технологіях переробки інформації для оцінювання тривалості проєктів з розробки програмного забезпечення використовуються методи та моделі, які були побудовані без врахування реального розподілу емпіричних даних. Наприклад, метод PERT базується на припущенні про бета-розподіл, а моделі COCOMO та ISBSG – на припущенні про логнормальний розподіл емпіричних даних тривалості проєктів з розробки програмного забезпечення, хоча ці припущення не завжди виконуються [5,6]. Це призводить до зниження достовірності оцінювання тривалості проєктів з розробки програмного забезпечення. Крім того вказані інформаційні технології мають відносно високу вартість (Comparative Estimating Tool – 374,32 USD, Microsoft Project Standard 2016 – 589,99 USD). Тому наукова задача створення нових технологій переробки інформації для оцінювання тривалості проєктів з розробки програмного забезпечення з метою



підвищення достовірності оцінювання тривалості проектів з розробки програмного забезпечення є актуальною та має практичну цінність.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Зараз в існуючих інформаційних технологіях переробки інформації для оцінювання тривалості проектів з розробки програмного забезпечення використовуються параметричні моделі, які базуються на трудомісткості робіт. Серед цих моделей найбільш розповсюдженою є модель СОСОМО [6] – негаусівська нелінійна регресійна модель тривалості проектів з розробки програмного забезпечення в залежності від їх трудомісткості, яка була побудована для різних платформ: персональні комп'ютери (PC), платформи середнього розміру (MR), мейнфрейми (MF). В [3] для цих платформ на основі емпіричних даних International Software Benchmarking Standards Group (ISBSG) було побудовано більш нові регресійні моделі. При побудові вказаних регресійних моделей виконувалась нормалізація емпіричних даних тривалості робіт в програмних проектах та емпіричних даних трудомісткості проектів з розробки програмного забезпечення. Необхідність застосування нормалізуючих перетворень при побудові регресійних моделей виникає через те, що закон розподілу емпіричних даних тривалості робіт в програмних проектах та емпіричних даних трудомісткості цих проектів не є нормальним [5,6,7,8,9,10], і тому неможливо побудувати адекватні лінійні регресійні моделі.

При побудові негаусівських нелінійних регресійних моделей СОСОМО та ISBSG для нормалізації вказаних емпіричних даних було використано логарифмічне перетворення [3,6]. Логарифмічне перетворення не завжди дозволяє добре нормалізувати емпіричні данні, що приводить до необхідності використовувати інші норма-

лізуючі перетворення. Одним з таких перетворень є перетворення Джонсона [11].

В [7,8,9,10] на основі емпіричних даних ISBSG з використанням нормалізуючого перетворення Джонсона було побудовано нелінійні регресійні моделі та їх довірчі інтервали для платформ PC, MR, MF. Використання зазначених регресійних моделей та їх довірчих інтервалів дозволяє підвищити достовірність оцінювання тривалості проектів з розробки програмного забезпечення. Тому доцільним є створення інформаційної технології переробки інформації для оцінювання тривалості проектів з розробки програмного забезпечення із застосуванням регресійних моделей, побудованих з використанням нормалізуючого перетворення Джонсона.

Ціль роботи полягає у створенні інформаційної технології переробки інформації для оцінювання тривалості проектів з розробки програмного забезпечення із застосуванням негаусівських нелінійних регресійних моделей, побудованих з використанням нормалізуючого перетворення Джонсона.

ВИКЛАДЕННЯ ОСНОВНОГО МАТЕРІАЛУ

Термін "Інформаційна технологія" має декілька визначень, кожне з яких підкреслює певний його бік. Так, [12,13] наголошують на системі процесів, методів і способів переробки інформації, [14,15] акцентують увагу на сукупності моделей і методів, а також виробничих процесів і програмно-технічних засобів, що забезпечують обробку даних.

Інформаційна технологія, що пропонується, була створена на основі математичних моделей, наведених в [7,8,9,10] та автоматизованої системи (АС), приведеної далі в цій роботі.

Основною ціллю створення запропонованої інформаційної технології є підви-

щення достовірності оцінювання тривалості проектів з розробки програмного забезпечення.

Задачі, які вирішуються за допомогою пропонованої інформаційної технології:

- точкове та інтервальне оцінювання тривалості проектів з розробки програмного забезпечення;
- моделювання тривалості проектів з розробки програмного забезпечення;
- збір статистичних даних про завершені проекти з розробки програмного забезпечення;
- уточнення початкових параметрів моделей на основі зібраних статистичних даних.

Приведемо автоматизовану систему, на основі якої створюється пропонована інформаційна технологія. Автоматизована система (automated system), згідно [12,14], являє собою систему, що складається з персоналу і комплексу засобів автоматизації його діяльності, та реалізує інформаційну технологію виконання встановлених функцій. На практиці в середині АС виділяють кілька підсистем, необхідність такої дії зазвичай буває викликана організацій-

ними і фінансовими причинами. Функціональна підсистема – складова частина автоматизованої системи, що реалізує одну або кілька близьких функцій.

Носій функціональної ролі – користувач, безпосередньо зайнятий в автоматизованій діяльності та бере участь у виконанні певних функцій системи на певних етапах.

Всередині АС виділяють також компоненти – частини АС, з яких вона будується в об'єктивній реальності. Система фізично складається зі своїх компонентів, тому розподіл автоматизованої системи на компоненти носить найбільш об'єктивний характер. Компонентами АС є обладнання, програми, і в деякому роді, користувачі.

Кожен вид забезпечення об'єднує в собі компоненти або технічні рішення певного характеру. У [12,14] наводяться різні види забезпечення АС: організаційне, методичне, математичне, технічне, інформаційне, програмне, правове, лінгвістичне, ергономічне.

Архітектура автоматизованої системи переробки інформації для оцінювання тривалості проектів з розробки програмного забезпечення приведена на рисунку 1.



Рисунок 1. Архітектура автоматизованої системи переробки інформації для оцінювання тривалості проектів з розробки програмного забезпечення.

Користувачі представлені функціональними ролями в системі:

- ОПР (особа, приймаюча рішення) – або менеджер, погоджує та затверджує умови контракту з замовником, включаючи тривалість проектів з розробки програмного забезпечення та відповідальність розробника за зрив запланованих термінів виконання;
- аналітик – виконує аналітичне обґрунтування і формує пропозицію щодо виділення необхідних ресурсів проекту, включаючи тривалість проектів з розробки програмного забезпечення;
- експерт – фахівець з питань статистичної обробки даних; виконує побудову негаусівських нелінійних регресійних моделей тривалості робіт в програмних проектах в залежності від трудомісткості цих проектів.

Підсистема зберігання інформації складається з бази даних (БД) та системи управління базою даних (СУБД). Використовується для зберігання детальної інформації про виконані проекти з розробки програмного забезпечення.

Підсистема зберігання інформації призначена для:

- завантаження емпіричних даних (інформації про завершені проекти) з файлів в базу даних;
- ручного введення емпіричних даних (інформації про завершені проекти) в базу даних;
- перегляду та редагування емпіричних даних, збережених в базі даних.

Підсистема оцінювання призначена для:

- оцінювання тривалості проектів з розробки програмного забезпечення та визначення довірчого інтервалу отриманої оцінки;
- моделювання тривалості проектів з розробки програмного забезпечення;

Підсистема розрахунку параметрів моделей призначена для:

- попередньої обробки емпіричних даних тривалості проектів з розробки програмного забезпечення, а також трудомісткості цих проектів, з метою видалення даних, які значно відрізняються від основної вибірки;
- вибору нормалізуючого перетворення емпіричних даних;
- побудови негаусівських нелінійних регресійних моделей тривалості робіт в програмних проектах в залежності від трудомісткості цих проектів.

Розглянемо більш детально нелінійні регресійні моделі, які використовуються для оцінювання тривалості проектів з розробки програмного забезпечення. Нехай випадкова величина D – це емпіричні значення тривалості робіт в програмних проектах, а випадкова величина E – це емпіричні значення трудомісткості цих проектів. Розподіл випадкових величин D та E не є нормальним, асиметрія та ексцес розподілу величин D та E значно відрізняються від відповідних характеристик нормального розподілу [7,8,9,10]. Згідно [7,8,9] для оцінювання значення величини D в залежності від значення величини E можна застосовувати негаусівські нелінійні регресійні моделі тривалості проектів з розробки програмного забезпечення в залежності від їх трудомісткості:

$$\hat{D}(E) = \frac{\varphi_D + (\lambda_D + \varphi_D)c}{1 + c}, \quad (1)$$

де

$$c = \exp\left\{\frac{\hat{z}_D(z_E) - \gamma_D}{\eta_D}\right\};$$

$$\hat{z}_D(z_E) = b_0 + b_1 z_E; \quad z_E = \gamma_E + \eta_E \ln\left(\frac{E - \varphi_E}{\lambda_E + \varphi_E - E}\right)$$

; $\gamma_D, \eta_D, \varphi_D, \lambda_D, \gamma_E, \eta_E, \varphi_E, \lambda_E, b_0, b_1$ -

коефіцієнти, які залежать від вибраної платформи (наведені в [7,8,9]).

Довірчий інтервал $[\hat{D}(E)]$ для нелінійної регресії (1) задається як

$$[\hat{D}(E)] = \frac{\varphi_D + (\lambda_D + \varphi_D)c}{1 + c} \quad (2)$$

де $c = \exp\left\{\frac{[\hat{z}_D(z_E)] - \gamma_D}{\eta_D}\right\}$.

Для визначення за (2) довірчого інтервалу математичного сподівання оцінки тривалості проектів з розробки програмного забезпечення $[\hat{z}_D(z_E)]$ задається як

$$[\hat{z}_D(z_E)] = \hat{z}_D(z_E) \pm t_{\alpha/2, n-2} \cdot \sqrt{s_{z_D}^2 \left[\frac{1}{n} + \frac{(z_E - \bar{z}_E)^2}{S_{z_E}} \right]}$$

де n - кількість значень емпіричних даних, які використовувалися про побудові регресійної моделі для вибраної платформи (наведені в [7,8,10]);

$$f_B(D) = \frac{\eta_D \lambda_D}{\sqrt{2\pi}(D - \varphi_D)(\lambda_D + \varphi_D - D)} e^{-\frac{1}{2} \left[\gamma_D + \eta_D \ln\left(\frac{D - \varphi_D}{\lambda_D + \varphi_D - D}\right) \right]^2}, \quad \varphi_D < D < \varphi_D + \lambda_D$$

де $\gamma_D, \eta_D, \varphi_D, \lambda_D$ - коефіцієнти, які залежать від вибраної платформи (наведені в [7,8]), $\eta_D > 0; -\infty < \gamma_D < \infty; \lambda_D > 0; -\infty < \varphi_D < \infty$.

Алгоритм роботи автоматизованої системи залежить від вирішуваної задачі.

Задача 1. Точкове та інтервальне оцінювання тривалості проектів з розробки програмного забезпечення.

Аналітик на основі платформи проекту та його трудомісткості виконує точкове та інтервальне оцінювання тривалості проек-

$$s_{z_D}^2 = \frac{1}{n-2} \sum_{i=1}^n (z_{D_i} - \hat{z}_D(z_{E_i}))^2; \quad S_{z_E} = \sum_{i=1}^n (z_{E_i} - \bar{z}_E)^2;$$

$t_{\alpha/2, n-2}$ - квантіль t -розподілу Стюдента, визначається за таблицею верхніх $100\alpha\%$ -вих точок t -розподілу Стюдента за рівнем значимості $\alpha/2$ та кількістю ступенів вільності $n-2$.

Для визначення за (2) довірчого інтервалу прогнозування тривалості проектів з розробки програмного забезпечення $[\hat{z}_D(z_E)]$ задається як

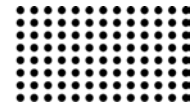
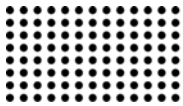
$$[\hat{z}_D(z_E)] = \hat{z}_D(z_E) \pm t_{\alpha/2, n-2} \cdot \sqrt{s_{z_D}^2 \left[1 + \frac{1}{n} + \frac{(z_E - \bar{z}_E)^2}{S_{z_E}} \right]}$$

Моделювання тривалості проектів з розробки програмного забезпечення виконано методом виключення (методом фон Неймана) згідно алгоритму наведеного в [17] та використовуючи функцію щільності ймовірності розподілу Джонсона сім'ї S_B . Вказана щільність ймовірності задається як

ту з розробки програмного забезпечення і надає ці дані ОПР для погодження та подальшого підписання контракту з замовником програмного забезпечення.

Задача 2. Моделювання тривалості проектів з розробки програмного забезпечення.

Аналітик, маючи значення точкової оцінки математичного сподівання тривалості проекту з розробки програмного забезпечення та довірчого інтервалу прогнозування, моделює можливі значення тривалості і надає ці дані ОПР для підтримки



прийняття рішень при складанні контракту з замовником програмного забезпечення.

Задача 3. Збір статистичних даних про завершені проекти з розробки програмного забезпечення.

Після завершення проекту аналітик вносить дані про прогнозовану та фактичну тривалість проекту в базу даних для подальшого аналізу та уточнення параметрів моделей.

Задача 4. Уточнення початкових параметрів моделей на основі зібраних статистичних даних про завершені проекти з розробки програмного забезпечення.

З певною періодичністю (наприклад, кожні шість місяців) експерт на основі нових статистичних даних про завершені проекти з розробки програмного забезпечення проводить попередню обробку вказаних даних, після чого визначає нові параметри нелінійної регресійної моделі тривалості проектів з розробки програм-

ного забезпечення в залежності від їх трудомісткості за алгоритмом, приведеним в [7,8,9]. Моделі з уточненими параметрами в подальшому використовуються при вирішенні задачі 1 та задачі 2.

ВИСНОВКИ

В даній роботі створено інформаційну технологію переробки інформації для оцінювання тривалості проектів з розробки програмного забезпечення із застосуванням негаусівських нелінійних регресійних моделей, побудованих з використанням нормалізуючого перетворення Джонсона. Застосування вказаних моделей враховує реальний розподіл емпіричних даних і дає можливість визначення довірчого інтервалу нелінійної регресії та інтервалу її прогнозування. Це дозволяє підвищити достовірність оцінювання тривалості проектів з розробки програмного забезпечення.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Comparative Estimating Tool [Elektronnyi resurs] // International Software Benchmarking Standards Group. – Rezhym dostupu : <http://isbsg.org/product/comparative-estimating-tool/>.
2. Microsoft Project 2016 Standard [Elektronnyi resurs] // Microsoft Corporation. – Rezhym dostupu : <https://products.office.com/uk-ua/Project/project-standard-desktop-software>.
3. Oligny, S. Exploring the relation between effort and duration in software engineering projects [Tekst] / S. Oligny, P. Bourque, A. Abran, B. Fournier // In proc. of the World Computer Congress, Aug. 2000. – pp. 175-178.
4. CHAOS Report [Elektronnyi resurs] // The Standish Group International, Inc. – 2009. – Rezhym dostupu : http://www.standishgroup.com/newsroom/chaos_2009.php.
5. Prykhodko, S. B. Porivniannia ymovirnisnykh modelei tryvalosti robit v prohramnykh proektakh [Tekst] / S. B. Prykhodko, A. V. Pukhalevych // Vostochno-Evropeiskiy zhurnal peredovykh tekhnolohiy. – Kharkov : Tekhnolohycheskyi tsentr, 2012. – № 1/13 (55). – S. 39-41. – ISSN 1729-3774.
6. Boehm, B. W. Software engineering economics [Tekst] / B. W. Boehm. – Englewood Cliffs, NJ : Prentice Hall, 1981. – 768 p.
7. Prykhodko, S. B. Rozrobka nelineinoy rehresii noi modeli tryvalosti prohramnykh proektiv na osnovi normalizuiuchoho peretvorennia Dzhonsona [Tekst] / S. B. Prykhodko, A. V. Pukhalevych // Radioelektronni i kompiuterni systemy. – Kharkiv : KhAI, 2012. – № 4 (56) – S. 90-93. – ISSN 1814-4225.
8. Prykhodko, S. B. Rozrobka nelineinykh rehresii nykh modelei tryvalosti prohramnykh proektiv na osnovi peretvorennia Dzhonsona [Tekst] / S. B. Prykhodko, A. V. Pukhalevych // Zbirnyk naukovykh prats NUK. – Mykolaiv : NUK, 2014. – № 2 (2014). – S. 76-80. – ISSN 2313-0415.
9. Prykhodko, S. B. Developing PC Software Project Duration Model based on Johnson transformation [Text] / S. B. Prykhodko, A. V. Pukhalevich // Proceedings of the 12th International Conference Modern Problems of Radio Engineering, Telecommunications and Computer Science TCSET2014, Lviv-Slavske, Ukraine. – Lviv : Polytechnic National University, 2014. – pp.114-116.
10. Prykhodko, S. B. Confidence interval estimation of PC software project duration regression based on Johnson transformation [Tekst] / S. B. Prykhodko, A. V. Pukhalevych // Radioelektronni i kompiuterni systemy. – Kharkiv : Kharkivskiy aviatsiinyi instytut, 2014. – № 2 (66). – S. 104-107. – ISSN 1814-4225.

11. Johnson, R. A. Applied Multivariate Statistical Analysis [Tekst] / R. A. Johnson, D. W. Wichern. – Pearson Prentice Hall, 2007. – 800 p.
12. DSTU 2226-93. Avtomatyzovani systemy. Terminy ta vyznachennia [Tekst]. – Nadano chynnosti 1994-07-01. – Kyiv : Derzhstandart Ukrainy, 1994. – 92 s.
13. DSTU 2481-94. Systemy obrobлення informatsii. Intelktualni informatsiini tekhnolohii. Terminy ta vyznachennia [Tekst]. – Nadano chynnosti 1995-01-01. – Kyiv : Derzhstandart Ukrainy, 1994. – 72 s.
14. HOST 34.003-90. Ynformatsyonnaia tekhnolohyia. Kompleks standartov na avtomatyzirovannye systemy. Avtomatyzirovannye systemy. Terminy i opredelenyia [Tekst]. – Vveden 1992-01-01. – Moskva : Standartynform, 2009. – 16 s.
15. HOST 7.0-99. Systema standartov po ynformatsyy, byblyotechnomu y yzdatelskomu delu. Ynformatsyonno-byblyotechnaia deiatel'nost, byblyohrafiya. Terminy i opredelenyia [Tekst]. – Vveden 2000-07-01. – Moskva : Yzdatel'stvo standartov, 1999. – 23 s.
16. Prykhodko, S. B. Intervalne otsiniuvannia statystychnykh momentiv nehausivskykh vypadkovykh velychyn na osnovi normalizuiuchykh peretvoren [Tekst] / S. B. Prykhodko // Naukovyi zhurnal "Matematychni modeliuvannia". – Dniprodzerzhynsk : DDTU, 2011. – № 1 (24). – С. 9-13.
17. Utkin, V.B. Informacionnye sistemy v jekonomike: ucheb. dlja stud. vyssh. ucheb. zavedenij [Tekst] / V.B. Utkin, K.V. Baldin. – Moskva : Izdatel'skij centr "Akademija", 2010. – 288 s.

Рецензент: д.т.н., проф. Коваленко І.І.
Чорноморський державний університет ім. Петра Могили