

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТРУДОМІСТКОСТІ РОЗРОБКИ ПРОГРАМНИХ ПРОЄКТІВ ЗА НЕЛІНІЙНОЮ РЕГРЕСІЙНОЮ МОДЕЛЛЮ

Приходько С.Б., д.т.н., професор¹; Пухалевич А.В., к.т.н.²; Машкін О.О.³
^{1,2,3} Національний університет кораблебудування імені адмірала Макарова

^{1,2,3} Україна, Миколаїв

¹sergiy.prykhodko@nuos.edu.ua; ²andrii.pukhalevych@nuos.edu.ua; ³201245@nuos.edu.ua

Анотація. В роботі виконано математичне моделювання трудомісткості розробки програмних проєктів за нелінійною регресійною моделлю в залежності від кількості функціональних точок.

Ключові слова: математичне моделювання; трудомісткість; програмний проєкт; нелінійна регресійна модель.

Вступна частина. У наші дні розробка програмного забезпечення (ПЗ) залишається бути складним процесом, який не завжди є успішним. Так, за дослідженнями Standish Group, які виконувалися з 2016 по 2020 р., лише 31% програмних проєктів були успішними, 50% завершилися лише з частковим успіхом, а 19% не були завершені взагалі [1], що у порівнянні з попередніми результатами, які були представлені у 2015 р. [2], практично не змінилися.

Однією із причин цього є не достовірне прогнозування трудомісткості розробки програмних проєктів в певних випадках. Тому достовірне прогнозування трудомісткості розробки програмних проєктів залишається важливою задачею.

На сьогоднішній день при прогнозуванні трудомісткості розробки програмних проєктів часто використовуються нелінійні регресійні рівняння трудомісткості їх виконання в залежності від певних метрик ПЗ, наприклад, таких як кількість строк коду, функціональних точок, сторінок проєктної документації, дефектів, тощо. Але зазначені рівняння, зокрема СОСОМО 81, СОСОМО II [3], не містять у своєму складі випадкових складових, що не дозволяє здійснювати моделювання трудомісткості розробки програмних проєктів як випадкової величини. Тому виникає потреба у побудові відповідних моделей для моделювання трудомісткості розробки програмних проєктів саме як випадкової величини.

Метою дослідження є побудова нелінійної регресійної моделі для моделювання трудомісткості розробки програмних проєктів саме як випадкової величини.

Основна частина. Розглянемо побудову нелінійної регресійної моделі для моделювання трудомісткості розробки програмних проєктів саме як випадкової величини на основі нормалізуючого перетворення у формі десятикового логарифму. Як і в [4], де було побудоване нелінійне регресійне рівняння за даними з [5], ми брали такі метрики з 133 програмних проєктів: трудомісткість Y у годинах; кількість функціональних точок X_1 . Далі ми перевірили двовимірні дані на нормальність за допомогою тесту Мардіа (Mardia's test), якій базується на багатовимірних асиметрії та ексцесі. Тест показав, що розподіл даних не є гаусівським. Тому для визначення наявності викидів у даних ми використовували тест на основі квадрату відстані Махаланобіса для нормалізованих за десятиковим логарифмом даних.

Використовуючи зазначені вище дані з 133 програмних проєктів спочатку була побудована лінійна регресійна модель $Y = \hat{b}_0 + \hat{b}_1 X_1 + \varepsilon$, де $\hat{b}_0$ та $\hat{b}_1$ є оцінками параметрів.

Причому змінна ε повинна бути випадковою величиною з розподілом Гаусу. Нульова гіпотеза про нормальність розподілу ε була перевірена за критерієм Пірсона. Ця гіпотеза була відкинута для рівня значимості 0,05. Те, що розподіл ε не є гаусівським вказує на відсутність теоретичного обґрунтування застосування лінійної регресійної моделі та на необхідність побудови нелінійної регресійної моделі.

Нелінійна регресійна модель має вигляд $Y = 10^{\varepsilon + \hat{b}_0} X_1^{\hat{b}_1}$, де ε – гаусівська випадкова величина з нульовим математичним сподіванням та середньоквадратичним відхиленням, оцінка якого дорівнює 0,1208, $\hat{b}_0$ та $\hat{b}_1$ є оцінками параметрів, які мають відповідно такі значення: -0,389258 та 0,876336. Оцінки параметрів були отримані за допомогою методу найменших квадратів. Ця модель була побудована за методом [6] із врахуванням наявності викидів за даними з 95 програмних проєктів, які залишилися після видалення викидів.

Створена нелінійна регресійна модель була перевірена за множинним коефіцієнтом детермінації R^2 , середньою величиною відносної помилки MMRE і відсотком прогнозованих результатів, для яких величини відносної помилки MRE менші за 0,25, PRED(0,25). Допустимі значення MMRE і PRED(0,25) складають не більше 0,25 і не менше 0,75 відповідно. Значення R^2 , MMRE і PRED(0,25) для створеної моделі дорівнюють відповідно 0,875, 0,230 і 0,621. Це свідчить про задовільну якість отриманої моделі.

Висновки. В роботі на основі десятичного логарифму була побудована нелінійна регресійна модель для моделювання трудомісткості розробки програмних проєктів саме як випадкової величини в залежності від кількості функціональних точок.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Portman, H. (2021, January 6). *Review Standish Group – CHAOS 2020: Beyond Infinity*. <https://hennyporrtman.wordpress.com/2021/01/06/review-standish-group-chaos-2020-beyond-infinity/>
- [2] Wojewoda, S., & Hastie, S. (2015, October 4). *Standish Group 2015 Chaos Report – Q&A with Jennifer Lynch*. Standish Group. <https://www.infoq.com/articles/standish-chaos-2015/>
- [3] Boehm, B.W., Abts, C., Brown, A.W., Chulani, S., Clark, B.K., Horowitz, E., Madachi, R., Reifer, D., & Steece, B. (2000). *Software Cost Estimation With Cocomo II*. Prentice Hall.
- [4] Prykhodko, S.B., Prykhodko, N.V., Makarova, L.M., Kudin, O.O., & Smykodub, T.G. (2017). Constructing the non-linear regression equations on the basis of bivariate normalizing transformations. *Вісник Херсонського національного технічного університету*. 3(62). 333-337.
- [5] Kitchenham, B., Pfleeger, S.L., McColl, B., & Eagan, S. (2002) An empirical study of maintenance and development estimation accuracy. *The Journal of Systems and Software*, 1(64). 57-77. [https://doi.org/10.1016/S0164-1212\(02\)00021-3](https://doi.org/10.1016/S0164-1212(02)00021-3)
- [6] Prykhodko, S., & Prykhodko, N. (2022). A modified technique for constructing nonlinear regression models based on the multivariate normalizing transformations. *Selected Papers of the VIII International Scientific Conference "Information Technology and Implementation" (IT&I-2021). Workshop Proceedings (IT&I-WS-2021), Kyiv, Ukraine, December 1-3, 2021. CEUR Workshop Proceedings*, 3179, 156-166. http://ceur-ws.org/Vol-3179/Paper_15.pdf

Prykhodko S.B., Pukhalevych A.V., Mashkin O.O.

Mathematical modeling of software project development effort by a nonlinear regression model

Abstract. In work mathematical modeling of software project development effort by a nonlinear regression model depending count of functional points has been performed.

Keywords: mathematical modeling; effort; software project; nonlinear regression model.