

НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ МЕТРИКИ RFC ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА JAVA

Приходько С.Б., д.т.н., професор¹; Позняков Р.О.²

^{1,2} Національний університет кораблебудування імені адмірала Макарова

^{1,2} Україна, Миколаїв

¹sergiy.prykhodko@nuos.edu.ua; ²211264@nuos.edu.ua

Анотація. В роботі на основі перетворення у вигляді десяткового логарифму побудована двохфакторна нелінійна регресійна модель для оцінювання метрики RFC застосунків з відкритим кодом на Java в залежності від метрик CBO та WMC.

Ключові слова: оцінювання, метрика RFC; застосунок з відкритим кодом; Java; нелінійна регресійна модель.

Вступна частина. Для визначення якості програмного забезпечення (ПЗ) запропоновано використовувати різні метрики ПЗ, включаючи RFC (response for a class), CBO (coupling between object classes) та WMC (weighted methods per class) на рівні програми [1]. Однак відомі лише точкові оцінки їх середніх для виявлення якості програмних систем [1]. А інтервальних оцінок для показників ПЗ на рівні програми для визначення її якості немає. Також відомо [1], що метрика RFC залежить від метрик CBO та WMC. У [1] автори побудували рівняння лінійної регресії для оцінювання середнього RFC. Водночас не завжди є теоретичне обґрунтування використання моделей лінійної регресії, наприклад, коли розподіл залишків не є гаусовим [2]. У цьому випадку необхідно застосовувати нелінійні регресійні моделі [3].

Метою дослідження є побудова нелінійної регресійної моделі для оцінювання метрики RFC застосунків з відкритим кодом на Java в залежності від метрик CBO та WMC.

Основна частина. Розглянемо побудову нелінійної регресійної моделі для оцінювання метрики RFC застосунків з відкритим кодом на Java в залежності від метрик CBO та WMC. Для цього зазначені три метрики з 49 застосунків з відкритим кодом на Java, які розміщені на сайті GitHub (<https://github.com>), були отримані за допомогою інструменту СК (<https://github.com/mauricioaniche/ck>). Далі ми перевірили тривимірні дані на нормальність за допомогою тесту Мардіа (Mardia's test), який 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 десятковим логарифмом даних.

Використовуючи зібрані дані із зазначених вище метрик застосунків з відкритим кодом на Java спочатку була побудована лінійна регресійна модель $Y = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \varepsilon$, де $\hat{b}_0$, $\hat{b}_1$ та $\hat{b}_2$ є оцінками параметрів. Змінна ε , яка описує залишки, 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, яка була побудована на основі десяткового логарифму має вигляд $Y = 10^{\varepsilon + \hat{b}_0} X_1^{\hat{b}_1} X_2^{\hat{b}_2}$, де ε – гаусівська випадкова величина з нульовим математичним сподіванням та середньоквадратичним відхиленням, оцінка якого дорівнює

0,06538, $\hat{b}_0$, $\hat{b}_1$ та $\hat{b}_2$ є оцінками параметрів, які мають відповідно такі значення: -0,054776, 0,672260 та 0,441541. Оцінки параметрів були отримані за допомогою методу найменших квадратів. Ця модель була побудована за методом [3] із врахуванням наявності викидів за даними з 46 застосунків, що залишилися, і може застосовуватися для значень метрик, для яких квадрат відстані Махаланобісу їх десяткових логарифмів менше або дорівнює 12,84.

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 0,8718, 0,1262 і 0,9783. Це свідчить про добру якість отриманої моделі. За побудованою нелінійною регресійною моделлю виконано оцінювання метрики RFC для трьох популярних застосунків з відкритим кодом на Java (FreeMind, jEdit та TuxGuitar), метрики яких аналізуються в [4] на рівні застосунку. Для FreeMind, jEdit та TuxGuitar значення оцінок метрики RFC дорівнюють 8,316, 7,813 та 10,199 відповідно. А актуальні значення RFC за [4] є такими 13,54, 10,62 та 17,78 відповідно. Отриманий результат вказує на те, що лише для jEdit відносна похибка (MRE) менше за 30%, а саме – 26,4%. Для застосунків FreeMind та TuxGuitar значення MRE складає 38,6% та 42,6% відповідно. Такий результат можна пояснити тим, що значення квадрату відстані Махаланобісу для десяткових логарифмів метрик застосунків FreeMind, jEdit та TuxGuitar більші за 12,84, а саме дорівнюють 17,11, 13,59 та 16,16 відповідно. Це вказує і на необхідність у подальшому використання більш складних перетворень, наприклад, тривимірних перетворень Джонсона або Бокса-Кокса.

Висновки. В роботі на основі десяткового логарифму була побудована нелінійна регресійна модель для оцінювання метрики RFC застосунків з відкритим кодом на Java в залежності від метрик CBO та WMC, яка має задовільну якість відповідного оцінювання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Laing, V., & Coleman, C. (2001). *Principal Components of Orthogonal Object-Oriented Metrics*. White Paper Analyzing Results of NASA Object-Oriented Data (Task 323-08-14). Greenbelt, Maryland (US): NASA Goddard Space Flight Center, the Software Assurance Technology Center (SATC).
- [2] Bates, D.M., & Watts, D.G. (1988). *Nonlinear Regression Analysis and Its Applications* (2nd ed.). New York: John Wiley & Sons. <http://dx.doi.org/10.1002/9780470316757>
- [3] 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
- [4] Molnar, A.J., Neamtu, A., & Motogna, S. (2020). Evaluation of software product quality metrics. In E. Damiani, G. Spanoudakis, & L. Maciaszek (eds). *Evaluation of Novel Approaches to Software Engineering : 14th International Conference (ENASE 2019), Communications in Computer and Information Science: Vol. 1172* (pp. 163-187). Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-40223-5_8

Prykhodko S.B., Poznyakov R.O.

A nonlinear regression model for estimating the RFC metric of open-source applications in Java

Abstract. In work, a nonlinear regression model with two predictors for estimating the RFC metric of open-source applications in Java depending on CBO and WMC metrics has been built.

Keywords: estimation; RFC metric; open-source app; nonlinear regression model.