

НЕЛІНІЙНІ РЕГРЕСІЙНІ МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ КІЛЬКОСТІ СТРОК КОДУ ВЕБ-ЗАСТОСУНКІВ, ЩО СТВОРЮЮТЬСЯ ЗА ДОПОМОГОЮ PHP-ФРЕЙМВОРКІВ

Приходько С.Б., д.т.н., професор¹, Приходько А.С.², Шутко І.С.³
^{1,2,3} Національний університет кораблебудування імені адмірала Макарова
^{1,2,3} Україна, Миколаїв

¹sergiy.prykhodko@nuos.edu.ua; ²whiterandrek@gmail.com; ³ishutko@gmail.com

***Анотація.** В роботі було побудовано нелінійні регресійні моделі для оцінювання кількості строк коду веб-застосунків, що створюються з використанням фреймворків CakePHP, Yii та Symfony.*

***Ключові слова:** оцінка кількості строк коду програмного забезпечення; веб-застосунок; нелінійна регресійна модель.*

Вступна частина. Оцінювання розміру веб-застосунків як і іншого програмного забезпечення (ПЗ) є важливою задачею на ранній стадії розробки програмного забезпечення, оскільки ця інформація використовується для прогнозування трудомісткості створення ПЗ за допомогою математичних моделей, наприклад, СОСОМО II. Зараз багато веб-застосунків створюються за допомогою різних фреймворків, які прискорюють розробку ПЗ. Найвідомішими з них є Symfony, Laravel, CakePHP, Yii.

На теперішній час відомі як лінійне регресійне рівняння [1], так і нелінійна регресійна модель [2] для оцінювання кількості строк коду інформаційних PHP-систем з відкритим кодом в залежності від трьох метрик концептуальної моделі даних у вигляді діаграми класів: загальна кількість класів, загальна кількість зв'язків та середня кількість атрибутів на клас. Але розмір веб-застосунків, що створюються з використанням різних фреймворків, може залежати від інших метрик [3]. Це потребує побудови відповідних моделей для оцінювання кількості строк коду веб-застосунків, що створюються з використанням PHP-фреймворків CakePHP, Yii та Symfony.

Метою дослідження є створення нелінійних регресійних моделей для оцінювання кількості строк коду веб-застосунків, що створюються з використанням фреймворків CakePHP, Yii та Symfony.

Основна частина. Розглянемо спочатку побудову нелінійної регресійної моделі для оцінювання кількості строк коду веб-застосунків, що створюються з використанням фреймворку CakePHP. Враховуючи на те, що CakePHP є PHP-фреймворком, метрики 38 веб-застосунків, які розміщені на сайті GitHub (<https://github.com>), були отримані за допомогою інструменту PhpMetrics (<https://phpmetrics.org/>). Ми брали такі метрики: розмір веб-застосунку Y у тисячах рядків коду; кількість класів X_1 ; середню кількість методів на клас X_2 ; глибину дерева спадкування X_3 . Ці метрики були обрані, тому що їх можна отримати з діаграми класів та між ними немає мультиколінеарності. Далі ми перевірили багатовимірні 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.

Використовуючи зібрані дані з із зазначених вище метрик веб-застосунків була побудована лінійна регресійна модель $Y = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \varepsilon$, де \hat{b}_0 , \hat{b}_1 , \hat{b}_2 та \hat{b}_3 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 в 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 $Y = 10^{\varepsilon + \hat{b}_0} X_1^{\hat{b}_1} X_2^{\hat{b}_2} X_3^{\hat{b}_3}$, де ε – гаусівська випадкова величина з нульовим математичним сподіванням та середньоквадратичним відхиленням, оцінка якого дорівнює 0,02976, \hat{b}_0 , \hat{b}_1 , \hat{b}_2 та \hat{b}_3 є оцінками параметрів, які мають відповідно такі значення: -1,8048, 0,9819, 1,1867 та 0,3198. Оцінки параметрів були отримані за допомогою методу найменших квадратів.

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 0,993, 0,107 і 0,929. Це свідчить про високу якість отриманої моделі.

Крім того були побудовані дві нелінійні регресійні моделі для оцінювання кількості строк коду веб-застосунків, що створюються з використанням фреймворків Yii та Symfony, які також мають добрі значення R^2 , MMRE і PRED(0,25).

Висновки. В роботі на основі десяткового логарифму було побудовано нелінійні регресійні моделі з трьома предикторами для оцінювання кількості строк коду веб-застосунків, що створюються за допомогою фреймворків CakePHP, Yii та Symfony. Створені моделі мають добру якість відповідного оцінювання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Tan, H.B.K., Zhao, Y., Zhang, H. (2006). Estimating LOC for information systems from their conceptual data models, *Software Engineering : the 28th International Conference (ICSE '06)* (pp. 321-330). Shanghai, China.
- [2] Prykhodko, N.V., Prykhodko, S.B. (2018). Constructing the non-linear regression models on the basis of multivariate normalizing transformations. *Electronic modeling*, 6 (40), 101-110. DOI: 10.15407/emodel.40.06.101
- [3] Приходько, С.Б., Приходько, Н.В., Фаріонова, Т.А., Ворона, М.В. (2020). Трьохфакторна нелінійна регресійна модель для оцінювання розміру Php-застосунків з відкритим кодом. *Науковий журнал «Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки»*, 31 (70), 124-131. DOI: 10.32838/2663-5941/2020.1-1/23

Prykhodko S.B., Prykhodko A.S., Shutko I.S.

Nonlinear regression models for estimating LOC of web applications created using PHP-frameworks

Abstract. In work three nonlinear regression models for estimating LOC of web applications created using the CakePHP, Yii, and Symfony frameworks have been created.

Keywords: LOC estimation; Web app; nonlinear regression model.