

НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНИХ СИСТЕМ З КОДОМ НА РНР

Приходько С.Б., д.т.н., професор¹, Сербулов Г.В.²

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

^{1,2} Україна, Миколаїв

¹sergiy.prykhodko@nuos.edu.ua; ²201250@nuos.edu.ua

***Анотація.** Удосконалено нелінійну регресійну модель для оцінювання розміру програмного забезпечення інформаційних систем з кодом на РНР. Ця модель має кращі значення показників R^2 , MMRE та PRED(0,25) у порівнянні з існуючими регресійними моделями, як лінійними, так і нелінійними.*

***Ключові слова:** оцінювання розміру програмного забезпечення; інформаційна система; нелінійна регресійна модель; РНР.*

Вступна частина. Задача оцінювання розміру програмного забезпечення (ПЗ) на ранній стадії його розробки є важливою, оскільки ця інформація використовується для прогнозування трудомісткості розробки ПЗ за допомогою такої відомої моделі як СОСОМО II (The Constructive Cost Model). Це призводить до необхідності розробки відповідних моделей та рівнянь для оцінювання розміру ПЗ, включаючи ПЗ інформаційних систем з кодом на РНР [1, 2].

У роботі [1] Hee Beng Kuan Tan, Yuan Zhao та Hongyu Zhang запропонували лінійні регресійні рівняння для оцінювання розміру (у тисячах строк коду) ПЗ інформаційних системи, що розробляються на базі певних мов програмування високого рівня, включаючи РНР. Запропоновані рівняння побудовано за допомогою множинного лінійного регресійного аналізу на основі метрик діаграми класів. Проте запропоноване в [1] лінійне регресійне рівняння для оцінювання розміру ПЗ інформаційних систем з відкритим кодом на РНР має незадовільні значення показників MMRE і PRED(0,25), які дорівнюють 0,4919 та 0,5313 відповідно. Тому в роботі [2] було побудовано нелінійне регресійне рівняння із застосуванням чотиривимірного перетворення Джонсона сім'ї S_B на основі тих же трьох метрик діаграми класів, що і в [1]: загальна кількість класів, загальна кількість зв'язків та середня кількість атрибутів на клас. Хоча запропоноване в [2] нелінійне регресійне рівняння для оцінювання розміру ПЗ інформаційних систем з відкритим кодом на РНР має добре значення MMRE – 0,2199, але значення PRED(0,25) складає 0,7188, що є меншим за 0,75. Тому виникає необхідність у подальшому удосконаленні нелінійного регресійного рівняння для оцінювання розміру ПЗ інформаційних систем з кодом на РНР.

Метою дослідження є удосконалення нелінійної регресійної моделі для оцінювання розміру ПЗ інформаційних систем з кодом на РНР.

Основна частина. Удосконалення нелінійної регресійної моделі для оцінювання розміру ПЗ інформаційних систем з кодом на РНР ми будемо за методом [3], що заснований на багатовимірних нормалізуючих перетвореннях та інтервалах передбачення. Суть цього методу [3] є такою. На першому етапі початкові дані перевіряються на наявність викидів і, якщо останні знайдено, то вони відкидаються. Для цього використовується критерій на основі квадрату відстані Махаланобіса для нормалізованих даних із 0,005 рівнем значущості. На другому етапі будується нелінійна регресійна модель із використанням відповідного

методу на основі нормалізуючих перетворень [4]. На третьому етапі визначаються границі інтервалу передбачення нелінійної регресії для рівня значущості, що дорівнює 0,05, за відповідним методом [4]. На останньому четвертому етапі перевіряють, чи є серед даних, за якими будувалася нелінійна регресійна модель такі, що виходять за границі інтервалу передбачення. Та, якщо відповідні дані знайдено, то вони відкидаються, і ми повторюємо знову всі етапи, починаючи з першого, для нових даних. Якщо таких викидів не було, то повторення етапів завершується, відповідна нелінійна регресійна модель побудована. Як і в [2-4], у якості багатовимірного нормалізуючого перетворення ми застосовуємо чотирьох-вимірне перетворення Джонсона для сімейства S_B .

Удосконалена нелінійна регресійна модель має таку ж саме структуру як і в [3, 4]. Ця 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 удосконаленої моделі дорівнюють відповідно 0,9885, 0,1725 і 0,7667, які є кращими за відповідні значення для лінійного та нелінійного регресійного рівнянь з [1] та [2]. Це свідчить про більш високу якість удосконаленої моделі у порівнянні з регресійними рівняннями з [1] та [2].

Висновки. На основі чотирьох-вимірного перетворення Джонсона сімейства S_B та видалення викидів у даних була удосконалена нелінійна регресійна модель з трьома предикторами для оцінювання розміру ПЗ інформаційних систем з кодом на PHP. Удосконалена модель має 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, S., Prykhodko, N., Makarova, L. (2019). Estimating the Software Size of Open-Source PHP-Based Systems Using Non-Linear Regression Analysis. *Proceedings of International Conference Advanced Computer Information Technologies (ACIT-2018). CEUR Workshop Proceedings, 2300*, 199-202.
- [3] Prykhodko, S., Prykhodko, N. (2021). Mathematical Modeling of Non-Gaussian Dependent Random Variables by Nonlinear Regression Models Based on the Multivariate Normalizing Transformations. In: Shkarlet S., Morozov A., Palagin A. (eds) *Mathematical Modeling and Simulation of Systems (MODS'2020)*. MODS 2020. Advances in Intelligent Systems and Computing, vol. 1265. (pp. 166-174). Springer, Cham. doi: https://doi.org/10.1007/978-3-030-58124-4_16
- [4] 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

Prykhodko S.B., Serbulov G.V.

A nonlinear regression model for estimating a software size of PHP-based information systems

Abstract. A nonlinear regression model for estimating a software size of PHP-based information systems have been improved. This model, in comparison with other regression models both linear and nonlinear, has better values of metrics named R^2 , MMRE and PRED(0.25).

Keywords: software size estimation; information system; nonlinear regression model; PHP.