

НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПЛАГІНІВ ДЛЯ WORDPRESS

Чичерін В.О.¹; Пухалевич А.В., к.т.н.²

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

¹ vovaempire@gmail.com; ² andrii.pukhalevych@nuos.edu.ua

Анотація. Зібрано ООП метрики 128 плагінів для Wordpress на PHP. Побудовано трифакторну нелінійну регресійну модель для оцінювання розміру плагінів для системи управління вмістом Wordpress. Перевірено характеристики побудованої моделі.

Ключові слова: оцінювання розміру програмного забезпечення; Wordpress плагіни; нелінійна регресійна модель.

Вступна частина. При розробці програмного забезпечення, в тому числі плагінів для Wordpress – найпоширенішої системи управління вмістом веб-сайтів на PHP, дуже важливо мати оцінки розміру коду, так як вони потім використовуються для передбачення трудомісткості створення програмного забезпечення.

Для оцінювання розміру коду використовуються лінійні та нелінійні регресійні моделі [1] в залежності від трьох метрик: загальна кількість класів, загальна кількість зв'язків та середня кількість атрибутів на клас. Розмір плагінів, що створені для Wordpress, може також залежати від інших метрик та мати інші параметри [2]. Тому це потребує побудови математичних моделей для оцінювання розміру плагінів, що застосовуються для такої системи як Wordpress.

Метою дослідження є побудова нелінійної регресійної моделі для оцінювання розміру плагінів для системи управління вмістом Wordpress.

Основна частина. Так як плагіни для Wordpress написані мовою програмування PHP, ООП метрики 128 плагінів, розміщених на сайті Wordpress (<https://wordpress.org/plugins/>), були отримані за допомогою інструменту PhpMetrics (<https://phpmetrics.org/>). Було взято такі метрики: розмір плагіну Y у тисячах рядків коду; кількість класів X_1 ; середня кількість методів на клас X_2 ; глибина дерева спадкування X_3 . Вибір цих метрик пов'язаний з тим, що їх можна одержати з діаграми класів, а також тому, що між ними немає мультиколінеарності.

Отримані багатовимірні дані було перевірено на нормальність закону розподілу. Виявилось, що розподіл даних не являється гаусівським, і тому необхідно провести нормалізацію даних. Для цього було використано десятковий логарифм.

Після нормалізації дані було перевірено на наявність викидів за допомогою квадрату відстані Махаланобіса [3] та побудовано лінійну регресійну 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 являються оцінками параметрів, а ε - це випадкова помилка, яка повинна мати нормальний розподіл.

Було отримано наступні значення оцінок параметрів лінійної регресії: $\hat{b}_0 = 1,3663$; $\hat{b}_1 = 0,9989$; $\hat{b}_2 = 1,0229$; та $\hat{b}_3 = 0,0252$. Нормальність розподілу випадкової помилки ε була перевірена за критерієм Пірсона.

На наступному етапі було побудовано нелінійну регресійну модель. Побудована нелінійна регресійна модель була перевірена за множинним коефіцієнтом детермінації R^2 , середньою величиною відносної помилки MMRE і PRED(0,25) – відсотком прогнозованих результатів, для яких величини відносної помилки MRE менші за 0,25. Допустимі значення R^2 , MMRE і PRED(0,25) складають не менше 0,7, не більше 0,25 і не менше 0,75 відповідно. Для побудованої моделі вони складають $R^2 = 0,92$; MMRE = 0,22; PRED(0,25) = 0,75. Це свідчить про середню якість отриманої моделі.

Висновки. В роботі, з використанням десяткового логарифму, було побудовано нелінійну регресійну модель з трьома факторами, для оцінювання розміру Wordpress плагінів. Побудована модель має достатні значення R^2 , MMRE і PRED, що свідчить про високу якість отриманої моделі. В подальшому планується створення програмного забезпечення, що реалізує побудовану модель для оцінювання розміру Wordpress плагінів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Tan, H.B.K., Zhao, Y., Zhang, H. (October, 2009). Conceptual data model-based software size estimation for information systems. *Transactions on Software Engineering and Methodology*, vol. 19, issue 2, article No. 4.
- [2] 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.
- [3] Prykhodko, S., Prykhodko, N., Makarova, L., & Pugachenko, K. (2017). Detecting outliers in multivariate non-Gaussian data on the basis of normalizing transformations. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*. doi:10.1109/ukrcon.2017.8100366

Chycherin V.O., Pukhalevych A.V.

THE NONLINEAR REGRESSION MODEL FOR ESTIMATING WORDPRESS PLUGINS SIZE

Abstract. OOP metrics for 128 Wordpress plugins are collected. A three-factor nonlinear regression model for estimating plugin size for the Wordpress content management system is built. The characteristics of the built model are checked.

Keywords: software size estimation; Wordpress plugins; nonlinear regression model.