

РЕГРЕСІЙНА МОДЕЛЬ ОЦІНЮВАННЯ РОЗМІРУ JAVA-ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ ДЛЯ РОБОТИ З БАЗАМИ ДАНИХ

Л.О. Латанська, к.ф.-м.н., доцент¹, Л.Б. Прокопович, магістрант²

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

^{1,2} Україна, м. Миколаїв

¹ llatanskaya@gmail.com, ² kalka.root@gmail.com

Анотація. Розглянута проблема побудови регресійної моделі оцінювання розміру Java-застосунків з відкритим кодом для роботи з базами даних. Побудована нелінійна регресійна модель за допомогою нормалізуючого перетворення на основі натурального логарифму. Проведено порівняння отриманих результатів з лінійної моделлю.

Ключові слова: регресійна модель; нормалізуючі перетворення; Java-застосунки; розмір коду програми.

Розмір є вагомим показником програмного забезпечення. Інформація, яка отримана при оцінці розміру програмного проекту, використовується при плануванні, контролі та аналізі діяльності ІТ-підприємств; при аудиті якості програмного забезпечення. Питанням оцінки розміру програмного забезпечення присвячено багато досліджень, наприклад [1-5]. Зараз існує велика кількість методів оцінки розміру програмного забезпечення, серед яких можна згадати наступні: експертні оцінки, регресійні моделі, нейронні мережі, комбіновані методи [6]. Незважаючи на те, що оцінки розміру програмного продукту недостатньо для повного розуміння складності проекту, існує взаємозв'язок між розміром програмного продукту та вартістю його розробки.

Приймаючи до уваги вище вказане, питання оцінювання розміру Java-застосунків з відкритим кодом та побудова відповідної регресійної моделі є актуальними.

Мета роботи: підвищення достовірності оцінювання розміру Java-застосунків з відкритим кодом для роботи з базами даних.

Для побудови регресійної моделі оцінювання розміру Java-застосунків з відкритим кодом для роботи з базами даних була сформована вибірка із 50 проектів, яка містить: кількість строк коду програмного забезпечення Y , середню кількість викликів методів X_1 , кількість класів та інтерфейсів X_2 , середню кількість методів X_3 . Вибірка формувалася за допомогою програми SourceMonitor та з використанням інформації веб-сервісу для хостингу ІТ-проектів GitHub.

Встановлено, що отримані емпіричні дані не підпорядковуються закону нормального розподілу. Це може призвести до помилок у розрахунках та негативно впливає на достовірність отриманих результатів, у нашому випадку, на оцінку розміру програмного забезпечення. Щоб уникнути цієї проблеми, перед тим як будувати математичну модель, потрібно нормалізувати вихідні дані. Для виконання нормалізуючих перетворень можна використовувати десятковий та натуральний логарифми, перетворення Бокса-Кокса, перетворення Джонсона та інші. В якості нормалізуючого перетворення було обрано натуральний логарифм.

Після усунення викидів для нормалізованих даних була побудована лінійна регресійна модель, загальний вигляд якої може бути представлений у вигляді рівняння (1):

$$Z_y = b_0 + b_1 Z_{x1} + b_2 Z_{x2} + b_3 Z_{x3} + \varepsilon, \quad (1)$$

де b_0, b_1, b_2, b_3 – коефіцієнти лінійної регресії;

ε – випадкова величина, яка розподілена за нормальним законом.

Застосувавши метод найменших квадратів, отримали наступні коефіцієнти лінійної регресії: $b_0=0,95595$; $b_1= 0,59190$; $b_2= 1,09993$; $b_3= 0,44306$. Проте p -значення $b_0= 0,17271$ більше 0,05. Це означає, що параметр b_0 не є значимим. Тому для нормалізованих даних було

побудовано нове лінійне регресійне рівняння без b_0 з такими коефіцієнтами: $b_1 = 0,55204$; $b_2 = 1,22996$; $b_3 = 0,55385$. Для отримання нелінійної регресійної моделі було виконано зворотне перетворення.

З метою порівняння була побудована лінійна регресійна модель в припущенні про нормальність вихідних даних без виконання нормалізації (2):

$$Y = -109531,2479 + 1830,1856X_1 + 92,7494X_2 + 9109,4542X_3 + \varepsilon. \quad (2)$$

За даними табл. 1 можна побачити, що нелінійна модель більш якісна.

Таблиця 1

Порівняння лінійної та нелінійної моделей

Модель	Коефіцієнт детермінації, R^2	Рівень передбачення, PRED(25%)
Лінійна	0,866	0,54
Нелінійна	0,923	0,78

Висновки. В результаті проведеного дослідження була побудована нелінійна регресійна модель оцінювання розміру Java-застосунків з відкритим кодом для роботи з базами даних. Перспективи подальших досліджень полягають в розробці програмного забезпечення для автоматизації процесу розрахунків за запропонованою моделлю.

REFERENCES

- [1] Этингоф, Е. В. (2013). Оценка затрат на информационные системы. *Отраслевая экономика*. Взято з <http://uecs.ru/marketing/item/2594-2013-12-05-12-00-33>.
- [2] Ваганова, Е. В., Земцов, А. А., Миньков С. Л. (2016). Оценка стоимости разработки программного продукта: обзор. *Проблемы учета и финансов*, 1(21), 58-62.
- [3] Заровский, А. П., Богуш, В. А. (2012). Использование имитационного моделирования для оценки сроков разработки программного обеспечения. *Экономика и управление*, 1, 95-101.
- [4] Василенко, Ю. Н. (2005). *Алгоритмические методы оценки программного обеспечения*, Мат. конф. "Инженерия программного обеспечения". Київ: НАУ.
- [5] Голованова, М. А., Надин, Е. В. (2014). Оценка трудоемкости работ на ранних стадиях создания программного обеспечения. *Системы обработки информации*, 8 (124), 151-156.
- [6] Титов, А. И. (2016). Выбор метрики размера проекта в модели оценки трудоемкости разработки программ. *Интеллектуальные технологии на транспорте*, 1, 31-38.

Liudmyla Latanska, Leonid Prokopovich

Open source Java applications size regression model for working with databases

Text annotation. The problem of building a regression model of the size of an open source Java application for working with databases is considered. A nonlinear model is constructed using a normalizing transformation based on the natural logarithm. The results are compared with a linear model.

Key words: regression model; normalizing transformation; Java applications; program code size.

Людмила Латанская, Леонид Прокопович

Регрессионная модель оценивания размера JAVA-приложений с открытым кодом для работы с базами данных

Текст аннотации. Рассмотрена проблема построения регрессионной модели размера Java-приложения с открытым кодом для работы с базами данных. Построена нелинейная модель с помощью нормализующего преобразования на основе натурального логарифма. Проведено сравнение полученных результатов с линейной моделью.

Ключевые слова: регрессионная модель; нормализующее преобразование; Java-приложение; размер кода программы.