

НЕЛІНІЙНІ РЕГРЕСІЙНІ МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ

Голубєв С.С.¹, Зелінський В.В.², Плигачов О.О.³, Пухалевич А.В. к.т.н.⁴,
Юрченко М.В.⁵

¹⁻⁵ Національний університет кораблебудування імені адмірала Макарова

¹⁻⁵ Україна, Миколаїв

⁴ andrii.pukhalevych@nuos.edu.ua

Анотація. Моделі, довірчі інтервали та інтервали передбачення нелінійних регресій для оцінювання розміру програмних застосунків з відкритим кодом побудовані на основі нормалізуючого перетворення у вигляді десяткового логарифму. Проведено порівняння побудованих моделей з відповідними лінійними регресійними моделями.

Ключові слова: оцінювання розміру програмного застосунку; нелінійна регресійна модель; довірчий інтервал; інтервал передбачення; нормалізуюче перетворення.

Задача оцінювання розміру програмних застосунків на ранній стадії програмного проекту є важливою, оскільки інформація, отримана при оцінюванні їх розміру, використовується для прогнозування трудомісткості по розробці застосунків, включаючи програмні застосунки з відкритим кодом.

У роботах [1-3] запропоновано рівняння лінійної регресії для оцінювання розміру програмного забезпечення (ПЗ), що написано деякими мовами програмування, включаючи C++, Java, Php, Visual Basic. Запропоновані рівняння побудовано за допомогою множинного лінійного регресійного аналізу на основі метрик з діаграми класів. Проте існують чотири основні припущення, які виправдовують використання моделей лінійної регресії, одним з яких є нормальність розподілу помилок. Але це припущення справедливе лише в окремих випадках. Це призводить до необхідності використання нелінійних регресійних моделей, в тому числі для оцінювання розміру ПЗ застосунків з відкритим кодом.

Нормалізуючі перетворення часто є добрим способом побудови моделей, рівнянь, довірчих інтервалів та інтервалів передбачення нелінійних регресій. Зокрема відомі моделі СОСОМО та СОСОМО II побудовані саме із використанням нормалізуючого перетворення у вигляді десяткового логарифму. А враховуючи на те, що розподіл розміру ПЗ не є гаусівським, потрібно використовувати саме нелінійні регресійні моделі, а їх побудову вести на основі нормалізуючих перетворень [4].

В роботі для оцінювання розміру програмних застосунків з відкритим кодом для різних мов програмування (C++, C#, Java) використовується трьохфакторна нелінійна регресійна модель, що побудована на основі нормалізуючого перетворення у вигляді десяткового логарифму:

$$Y = 10^{\varepsilon_x + b_0} X_1^{b_1} X_2^{b_2} X_3^{b_3}, \quad (1)$$

де ε_x – гаусівська випадкова величина, $\varepsilon_x \sim N(0, \sigma_x)$; X_1 – кількість класів; X_2 – кількість методів; X_3 – кількість атрибутів; b_0, b_1, b_2, b_3 – параметри моделі, які мають різні значення для відповідної мови програмування (C++, C# або Java).

Також було побудовано трьохфакторні лінійні регресійні моделі для відповідних мов програмування (C++, C# та Java).

Для порівняння моделі (1) та лінійної регресійної моделі були визначені значення множинного коефіцієнту детермінації R^2 , середньої величини відносної помилки MMRE та відсотка прогнозування PRED(0,25). Значення R^2 , MMRE та PRED (0,25) виявилися кращими для моделі (1) у порівнянні з лінійною регресійною моделлю для всіх мов програмування (C++, C# та Java). Також було побудовано довірчі інтервали та інтервали передбачення трьохфакторних лінійної і нелінійної регресій. Ширини довірчих інтервалів та інтервалів передбачення виявилися меншими для нелінійної регресії (1) у порівнянні з лінійною для всіх мов програмування, що розглядалися.

Висновки.

Удосконалено трьохфакторну нелінійну регресійну модель нелінійної регресії для оцінювання розміру програмних застосунків з відкритим кодом для різних мов програмування (C++, C#, Java) в залежності від метрик діаграми класів на основі нормалізуючого перетворення у вигляді десятичного логарифму, що дозволяє підвищити достовірність оцінювання залежної змінної нелінійної регресії у порівнянні з лінійною регресійною моделлю. Модель, що побудована, в порівнянні з лінійною регресійною моделлю, має більші значення множинного коефіцієнту детермінації і відсотку передбачень та менші середні величини відносної похибки, ширин довірчого інтервалу і інтервалу передбачення нелінійної регресії.

REFERENCES

- [1] Kaczmarek, J., & Kucharski, M. (2004). Size and effort estimation for applications written in Java. *Information and Software Technology*, 46, (9), 589-601. DOI: 10.1016/j.infsof.2003.11.001
- [2] Tan, H.B.K., Zhao, Y., & Zhang, H. (2006). *Estimating LOC for information systems from their conceptual data models*. Proceedings of the 28th International Conference on Software Engineering (ICSE '06), Shanghai, China, May 20-28, 2006.
- [3] Kiewkanya, M., & Surak, S. (2016). *Constructing C++ software size estimation model from class diagram*. Computer Science and Software Engineering: 13th International Joint Conference, Khon Kaen, Thailand, July 13-15, 2016. DOI: 10.1109/JCSSE.2016.7748880
- [4] Prykhodko, N.V., & Prykhodko, S.B. (2018). Constructing the non-linear regression models on the basis of multivariate normalizing transformations. *Electronic modeling*, 40, (6), 101–110. DOI: 10.15407/emodel.40.06.101

Golubev S.S., Zelinsky V.V., Pligachev O.O., Pukhalevych A.V., Yurchenko M.V.

Non-linear regression models to estimate the size of open source applications

Annotation. The models, confidence and prediction intervals of non-linear regressions for estimating the size of open source applications are constructed on the basis of the normalizing transformation in the decimal logarithm form. Comparison of the constructed models with the linear regression models is performed.

Keywords: application size estimation; non-linear regression model; confidence interval; prediction interval; normalizing transformation.

Голубев С.С., Зелинский В.В., Плигачев А.А., Пухалевич А.В., Юрченко М.В.

Нелинейные регрессионные модели для оценки размера программных приложений с открытым кодом

Аннотация. Модели, доверительные интервалы и интервалы прогнозирования нелинейных регрессий для оценки размера программных приложений с открытым кодом построены на основе нормализующего преобразования в виде десятичного логарифма. Проведено сравнение построенных моделей с соответствующими линейными регрессионными моделями.

Ключевые слова: оценка размера программного приложения, нелинейная регрессионная модель, доверительный интервал, интервал прогнозирования, нормализующее преобразование.