

МАТЕМАТИЧНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНИХ ЗАСТОСУНКІВ, ЩО ВИКОРИСТОВУЮТЬ 2D ГРАФІЧНІ РУШІЇ

Камінський С.С.¹; Бризгалов М.В.²; Макарова Л.М., к.т.н., доц.³

^{1,2,3} Національний університет кораблебудування імені адмірала Макарова

^{1,2,3} Україна, Миколаїв

¹191181@nuos.edu.ua; ²2001bruzga2001@gmail.com; ³lidia.makarova@nuos.edu.ua

Анотація. У даній роботі було побудовано нелінійну регресійну модель для оцінювання метрики KLOC програмних застосунків, які працюють на 2D графічних рушіях, використовуючи нормалізацію за допомогою десяткового логарифму. Модель враховує три ключових показники: метрики NCL, NM і DIT.

Ключові слова: оцінювання розміру програмного забезпечення; нелінійна регресійна модель; нормалізація; 2D графічні рушії; KLOC; NCL; NM; DIT.

Вступна частина. Зі зростанням використання 2D графічних рушіїв у розробці ПЗ, оцінювання розміру таких застосунків є важливою частиною процесу управління розробкою ПЗ [1]. Особливості архітектури таких застосунків створюють труднощі в оцінюванні їхнього розміру стандартними засобами. Стандартні підходи, розроблені для інших програмних платформ та мов програмування, не враховують цих особливостей, що може призводити до неточностей у прогнозах. Метрика KLOC (кількість тисяч рядків коду) залежить від метрик NCL (кількість класів), NM (кількість методів) та DIT (глибина дерева успадкування) [2]. Це підкреслює важливість розробки математичної моделі для оцінювання розміру програмних застосунків, особливо метрики KLOC, яка є ключовою для аналізу розміру проектів. У такому контексті доцільно використовувати нелінійні регресійні моделі, побудовані за допомогою нормалізації десятковим логарифмом.

Метою дослідження є побудова нелінійної регресійної моделі для оцінювання метрики KLOC у застосунках, що працюють на 2D графічних рушіях, з урахуванням метрик NCL, NM та DIT і застосуванням нормалізації за допомогою десяткового логарифму.

Основна частина. У цій роботі побудовано нелінійну регресійну модель для прогнозування середнього значення KLOC на основі таких метрик: NCL, NM та DIT. Для побудови моделі були зібрані дані зі 121 програмного застосунку, що використовують 2D графічні рушії, з відкритим кодом, розміщених на GitHub. Метрики KLOC, NCL, NM та DIT були отримані за допомогою спеціалізованого інструменту аналізу коду SourceMeter [3]. Спочатку була проведена перевірка всіх значень на наявність викидів за допомогою методу, заснованого на обчисленні квадрата відстані Махаланобіса для нормалізованих даних. Після видалення викидів залишилась вибірка із 111 застосунків. Для нормалізації даних використовувався десятковий логарифм, що дозволило нормалізувати емпіричні дані та покращити якість подальшого аналізу, спираючись на аналогічні методики [4]. Нормалізовані дані були розбиті на 2 вибірки: 60% (66 застосунків) для побудови моделі та 40% (45 застосунків) для її тестування. При розбитті враховувалось, щоб у вибірці для побудови моделі були присутні мінімальні та максимальні значення метрик, а у тестовій вибірці – значення, близькі до граничних. На основі даних із 66 застосунків була побудована нелінійна регресійна модель виду:

$$Y = 10^{b_0} * X_1^{b_1} * X_2^{b_2} * X_3^{b_3},$$

де Y – розмір програмного застосунку $KLOC$, X_1 – кількість класів NCL , X_2 – кількість методів NM , X_3 – глибина дерева успадкування DIT , b_0 , b_1 , b_2 , b_3 – оцінки параметрів моделі, які мають відповідно такі значення: 1,0919, 0,4375, 0,7196, -0,1547. Оцінки параметрів були отримані за допомогою методу найменших квадратів. Остаточна модель має вигляд:

$$Y = 10^{1,0919} * X_1^{0,4375} * X_2^{0,7196} * X_3^{-0,1547}.$$

Значення R^2 , $MMRE$ та $PRED(0,25)$ для побудованої моделі дорівнюють відповідно: 0,92, 0,25, 0,51. Це свідчить про задовільну якість отриманої моделі. Отримана модель дозволяє оцінювати розмір програмних застосунків, що використовують 2D графічні рушії, на основі метрик NCL , NM , DIT . Варто зазначити, що негативне значення коефіцієнта b_3 для метрики DIT може вказувати на те, що збільшення глибини успадкування класів призводить до зменшення загального розміру програми, можливо, за рахунок більш ефективного повторного використання коду. Для практичного застосування цієї моделі була розроблена програма, яка дозволяє автоматизувати процес оцінювання розміру програмних застосунків, що використовують 2D графічні рушії, на основі метрик NCL , NM та DIT . Ця програма може бути використана розробниками та менеджерами проектів для швидкого отримання оцінок розміру майбутніх програмних застосунків, що використовують 2D графічні рушії.

Висновки. В роботі була побудована нелінійна регресійна модель для оцінювання метрики $KLOC$ програмних застосунків, що використовують 2D графічні рушії, в залежності від метрик NCL , NM та DIT , на основі нормалізації за допомогою десятичного логарифму. Побудована модель демонструє задовільну якість оцінювання за показниками R^2 та $MMRE$, але показник $PRED(0,25)$ не досягає нормативного значення, покращення якості якого можна досягнути за рахунок використання інших нормалізуючих перетворень. Модель може бути використана для прогнозування розміру проектів, що базуються на 2D графічних рушіях.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Institute of Data. (13 November, 2023). *Understanding Size Estimation in Software Engineering*. <https://www.institutedata.com/blog/size-estimation-in-software-engineering/>
- [2] GeeksforGeek. (13 May, 2024). *What is KLOC in Software Engineering?*. <https://www.geeksforgeeks.org/what-is-kloc-in-software-engineering/>
- [3] SourceMeter. (n.d.). <https://sourcemeter.com/>
- [4] Приходько, С.Б., Приходько, Н.В., & Смикодуб, Т.Г. (2020). Чотирьохфакторна нелінійна регресійна модель для оцінювання розміру Java-застосунків з відкритим кодом. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки*, 2 (70), 157-162. <https://doi.org/10.32838/2663-5941/2020.2-1/25>

Kaminskyi S.S, Bryzgalov M.V., Makarova L.M.

A mathematical model for estimating the size of software applications using 2d graphics engines.

Abstract. In this paper, a nonlinear regression model was built to estimate the KLOC metric of software applications running on 2D graphics engines using decimal logarithm normalization. The model takes into account three key indicators: NCL, NM, and DIT metrics.

Keywords: software size estimation; nonlinear regression model; normalization; 2D graphics engines; KLOC; NCL; NM; DIT.