

Макарова Л. М.

Національний університет кораблебудування імені адмірала Макарова

Камінський С. С.

Національний університет кораблебудування імені адмірала Макарова

Бриггало М. В.

Національний університет кораблебудування імені адмірала Макарова

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ 2D-ЗАСТОСУНКІВ НА ОСНОВІ ГРАФІЧНИХ РУШІЇВ

Розглянуто проблему оцінювання розміру програмних застосунків, що використовують 2D графічні рушії. Це питання є актуальним, оскільки достовірне прогнозування розміру програмного коду впливає на ефективність розробки та оцінювання необхідних ресурсів. Однак існуючі підходи до оцінювання розміру програмних продуктів часто не враховують специфічні особливості програмування для 2D графічних рушіїв, що призводить до неточностей у розрахунках. Виникає необхідність у створенні нових моделей, які б дозволяли більш точно передбачати розмір коду, необхідний для реалізації застосунків цього типу.

З метою розв'язання цієї проблеми запропоновано математичну модель, що враховує ключові метрики програмного забезпечення, такі, як кількість класів у програмному коді, кількість методів у класах та глибину дерева наслідування. Для підвищення достовірності оцінювання застосовано нормалізуюче перетворення у вигляді десяткового логарифму. Також у моделі використано методи регресійного аналізу, які дозволяють будувати достовірні прогнози розміру застосунків залежно від вхідних параметрів. Аналіз результатів показав, що запропонована модель дозволяє отримати достовірну оцінку розміру програмного забезпечення, що дає змогу оптимізувати процес розробки, покращити планування ресурсів і зменшити ризики перевищення бюджету проекту.

На основі запропонованої математичної моделі розроблено програмне забезпечення для оцінювання розміру застосунків. Проведено тестування програмного забезпечення, яке підтвердило достовірність прогнозування метрики KLOC (кількість рядків коду) у порівнянні з існуючими моделями для мов програмування Java, C# та PHP.

Результати дослідження можуть бути корисними для розробників та менеджерів проєктів, які працюють із застосунками на основі 2D графічних рушіїв. Вони сприяють підвищенню ефективності розробки, дозволяючи обґрунтовано передбачати обсяг програмного коду та необхідних ресурсів, що є важливим фактором у процесі створення високоякісного програмного забезпечення.

Ключові слова: оцінювання розміру програмного забезпечення, 2D графічні рушії, метрики програмного забезпечення, нелінійна регресійна модель, нормалізуюче перетворення.

Постановка проблеми. Зі зростанням використання 2D графічних рушіїв у розробці ПЗ, оцінювання розміру таких програмних застосунків стає важливою частиною процесу управління розробкою. Особливості архітектури програм, що використовують 2D графічні рушії, створюють додаткові складнощі в оцінюванні їхнього розміру в порівнянні зі стандартними моделями [1].

Традиційні метрики, такі як, кількість тисяч рядків коду (KLOC), кількість класів (NC), кількість методів на клас (NM) та глибина дерева наслідування (DIT), є ключовими показниками для оцінювання розміру ПЗ [2]. Однак стандартні моделі, розроблені для інших платформ та мов програмування, не завжди можуть ефективно

враховувати специфіку 2D графічних рушіїв, що може призводити до неточностей у прогнозуванні.

Особливо актуальним це питання стає при розробці проєктів з використанням 2D графічних рушіїв, де традиційні моделі оцінювання, такі як СОСОМО II [3], показують свою обмеженість. Специфіка таких проєктів включає в себе складні системи рендерингу, управління ресурсами та оптимізацію пам'яті, системи обробки користувачького вводу та компоненти фізичного моделювання. Саме тому дану проблему можна вважати актуальною.

Аналіз останніх досліджень і публікацій. Оцінювання розміру ПЗ є важливим етапом у процесі розробки, оскільки дозволяє визначити обсяг

роботи, необхідні ресурси та час для реалізації проекту. Існуючі моделі оцінювання розміру ПЗ можна поділити на кілька категорій: лінійні регресійні моделі, нелінійні регресійні моделі, моделі на основі машинного навчання та експертні системи. Кожна з моделей має свої переваги та обмеження, що впливають на точність та ефективність оцінювання розміру програмного продукту.

Лінійні регресійні моделі, хоча і є простими для використання та інтерпретації, часто не можуть точно описати складні взаємозв'язки між метриками ПЗ. Дослідження показують, що зв'язки між показниками розміру програмного продукту, такими як кількість рядків коду та кількість класів і методів, мають нелінійний характер [4–6]. Існують моделі для оцінювання розміру ПЗ, написаного, зокрема, на Java, PHP та C#, які ґрунтуються на припущеннях структури і організації коду, що потребує адаптації методів оцінювання розміру ПЗ під конкретні умови розробки. Розглянемо більш детально деякі з цих моделей.

Модель оцінювання розміру веб-додатків на Java базується на кількості класів як незалежній змінній. Використано два підходи до нормалізації: одновимірне перетворення Джонсона та десятикове логарифмічне перетворення. Аналіз показав, що точність моделі потребує покращення, зокрема використання багатовимірного перетворення Джонсона [4].

Для C#-додатків побудована багатofакторна модель з урахуванням кількості класів та кількості методів на клас. Модель показала високу точність, проте не враховує глибину дерева наслідування, що може впливати на архітектуру застосунків [5].

Модель для веб-застосунків для фреймворку Symfony, яка включає три фактори: кількість класів, середню кількість методів на клас і глибину дерева наслідування, краще відповідає задачам аналізу архітектури застосунків [6]. Але її використання для оцінювання розміру ПЗ з використанням 2D графічних рушіїв є обмеженим, оскільки модель зосереджується на архітектурних характеристиках, а не на безпосередніх показниках обсягу коду, що знижує її достовірність у прогнозуванні зусиль на розробку або масштабування проекту.

Постановка завдання. Метою роботи є підвищення достовірності оцінювання розміру програмних застосунків, що використовують 2D графічні рушії, за рахунок удосконалення відповідної математичної моделі, що забезпечить більш високу достовірність прогнозування розміру на ранніх етапах розробки.

Для досягнення поставленої мети потрібно вирішити наступні задачі:

- дослідити наявні методи оцінювання розміру ПЗ;
- визначити ключові метрики для оцінювання розміру застосунків, що створюються за допомогою 2D графічних рушіїв;
- розробити математичну модель для оцінювання розміру програмних проектів, що створюються за допомогою 2D графічних рушіїв;
- реалізувати програмне забезпечення для оцінювання розміру на основі розробленої моделі.

Виклад основного матеріалу. У процесі оцінювання розміру програмного забезпечення, зокрема при розробці моделей для прогнозування KLOC, перевірка багатовимірної нормальності даних є важливим етапом. Багато статистичних методів, включно з регресією, базуються на припущенні нормальності розподілу. Одним із основних підходів для перевірки багатовимірної нормальності є тест Мардіа (Mardia's Test), який оцінює асиметрію та ексцес багатовимірного розподілу через статистичні показники $\beta_{1,p}$ та $\beta_{2,p}$ [7, 8].

Коефіцієнт асиметрії $\beta_{1,p}$ обчислює симетрію багатовимірного розподілу, враховуючи кореляцію між змінними [8]:

$$\beta_{1,p} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \{(X_i - \bar{X})^T S^{-1} (X_j - \bar{X})\}^3, \quad (1)$$

де N – кількість спостережень, X_i – вектор i -го спостереження, $\bar{X}$ – середній вектор, S^{-1} – обернена коваріаційна матриця.

Про негаусівський розподіл даних свідчить значення багатовимірного ексцесу $\beta_{2,p}$, оцінка якого визначається як [8]:

$$\beta_{2,p} = \frac{1}{N} \sum_{i=1}^N \{(X_i - \bar{X})^T S_N^{-1} (X_i - \bar{X})\}^2. \quad (2)$$

Якщо значення $\beta_{1,p}$ і $\beta_{2,p}$ вказують на відхилення від нормальності, може знадобитися застосування методів, що не залежать від нормальності розподілу, або проведення трансформацій даних для наближення їх до нормального вигляду.

Для пошуку викидів у багатовимірних даних можна використати квадрат відстані Махаланобіса, який зважає на взаємозв'язки між змінними [9]:

$$d_i^2 = (X_i - \bar{X})^T S_N^{-1} (X_i - \bar{X}), \quad (3)$$

де $\bar{X}$ – вектор вибірових середніх, S_N^{-1} – обернена коваріаційна матриця.

Дані для побудови нелінійної регресійної моделі для оцінювання розміру програмних застосунків з 2D графічними рушіями були зібрані зі 108 проєктів, які знаходяться у відкритому доступі в репозиторії GitHub і проаналізовані за допомогою інструменту SourceMeter [10]. Метрики NC, NM та DIT використовувалися для оцінювання розміру застосунку в тисячах рядків коду KLOC. Фрагмент даних наведено в таблиці 1.

Після перевірки даних на нормальність за допомогою критеріїв Мардіа, оцінки асиметрії $\beta_{1,p}$ згідно (1) та ексцесу $\beta_{2,p}$ згідно (2), було встановлено, що дані не відповідають нормальному розподілу. Результати перевірки: $\beta_{1,p} = 139,97$, що перевищує критичне значення 40,00, $\beta_{2,p} = 179,22$, що також значно перевищує критичне значення 28,58. Це свідчить про суттєві відхилення від нормальності у розподілі даних.

Далі було виконано пошук викидів у багатовимірних негаусівських даних з використанням квадрату відстані Махаланобіса (3) та нормалізацією через десятковий логарифм відповідно до [11]. Квантіль розподілу Пірсона для рівня значущості 0,005 та 4 ступенів вільності склав 14,86. Оскільки є такі дані, для яких квадрат відстані Махаланобіса більший за квантіль розподілу Пірсона, такий спостережуваний набір даних слід вважати викидом та вилучити з вихідної вибірки [12].

Після видалення викидів залишилась вибірка із 103 застосунків. Максимальне значення ква-

драту відстані Махаланобіса дорівнює 12,91 та не перевищує квантіль розподілу Пірсона, тому можна зробити висновок, що передобробка даних завершена.

Далі дані були розподілені на навчальну та тестову вибірки. Навчальна вибірка становила 60 % від загального обсягу (59 застосунків) і використовувалася для побудови моделі. Тестова вибірка, що становила 40 % (44 застосунки), застосовувалася для перевірки якості моделі.

Розподіл вибірок здійснено з урахуванням збереження мінімальних і максимальних значень метрик, що забезпечило репрезентативність даних та достовірність оцінювання моделі.

Оскільки емпіричні дані не мають нормального розподілу, для прогнозування розміру побудовано нелінійну регресійну модель з використанням нормалізації десятковим логарифмом, аналогічно роботі [13].

Після обчислень були отримані наступні значення коефіцієнтів рівняння регресії: $\beta_0 = -1,7566$, $\beta_1 = 0,5425$, $\beta_2 = 0,5816$, $\beta_3 = -0,1212$.

Таким чином, остаточно нелінійна регресійна модель для оцінювання розміру програмних застосунків, що використовують 2D графічні рушії, має вигляд:

$$\hat{Y} = 10^{-1,7566} \cdot X_1^{0,5425} \cdot X_2^{0,5816} \cdot X_3^{-0,1212}.$$

Розрахунок параметрів якості нелінійної регресійної моделі здійснюється для оцінювання

Таблиця 1

Фрагмент даних для оцінювання розміру застосунків, що використовують 2D графічні рушії

№	URL	KLOC	NC	NM	DIT
1	https://github.com/azurite-engine/Azurite	15,57	165	2023	76
2	https://github.com/rafael-esper/JLud2D	57,512	318	3959	109
3	https://github.com/dillanspencer/Quad-Engine	2,319	24	284	2
4	https://github.com/tombefieux/2DPhysicsEngine	1,18	12	160	17
5	https://github.com/j-ackayao/KY-Engine	2,327	22	740	21
6	https://github.com/Flafila2/Remote2D-Engine	14,413	137	2032	137
7	https://github.com/fastjengine/FastJ	24,886	194	3304	185
8	https://github.com/benblaszczak/BobEngine	10,136	104	2626	186
9	https://github.com/mlabouardy/2dgamesdk-engine	0,174	6	30	2
10	https://github.com/b3dgs/lionengine	129,428	1323	16249	1544
11	https://github.com/gurkenlabs/litiengine	71,476	604	13553	486
12	https://github.com/vanZeben/2D-Game-Engine	1,376	24	200	14
13	https://github.com/nicolasgramlich/AndEngine	54,229	636	16773	1098
14	https://github.com/jbox2d/jbox2d	52,36	349	6808	242
15	https://github.com/AlmasB/FXGL	47,328	586	7097	349
16	https://github.com/sriharshachilakapati/SilenceEngine	30,133	286	3243	153
17	https://github.com/codepath/android_simple_game_engine	0,54	11	157	12
18	https://github.com/swordmaster2k/rpgwizard	24,967	290	3272	151
19	https://github.com/kennycason/java_games	7,675	119	1843	204
20	https://github.com/SypherEngine/SypherEngine	4,755	56	616	33

її достовірності та адекватності. Для перевірки якості моделі використаємо такі параметри: коефіцієнт детермінації R^2 , скоригований коефіцієнт детермінації $R^2_{Adjusted}$, середню абсолютну відносну похибку ($MMRE$) та $PRED(0,25)$ [13].

Результати оцінювання для навчальної та тестової вибірок:

– Навчальна вибірка: $R^2 = 0,9302$, $R^2_{Adjusted} = 0,9264$, $MMRE = 0,2438$, $PRED(0,25) = 0,5254$.

– Тестова вибірка: $R^2 = 0,8459$, $R^2_{Adjusted} = 0,8338$, $MMRE = 0,2630$, $PRED(0,25) = 0,5476$.

Побудована нелінійна регресійна модель показує прийнятні результати прогнозування на обох вибірках. Значення R^2 , більші за 0,8, свідчать про ефективне пояснення варіації змінної KLOC. Низькі значення $MMRE$ свідчать про достатню надійність побудованої моделі. Водночас, отриманий рівень показника $PRED(0,25)$ вказує на необхідність подальшого вдосконалення моделі, зокрема, шляхом застосування інших нормалізуючих перетворень при оцінюванні розміру програмних застосунків, створених із використанням 2D графічних рушіїв.

Висновки. У роботі удосконалено математичну модель для оцінювання розміру програмних застосунків, що працюють на 2D графічних рушійх. Модель використовує метрики кількості

класів NC, кількість методів у класі NM та глибину дерева наслідування DIT для прогнозування розміру у тисячах рядків коду KLOC. Було проведено збір та аналіз даних із 108 проєктів, які знаходяться у відкритому доступі в репозиторії GitHub та використовують 2D графічні рушії. Після ретельної обробки даних та видалення викидів, було сформовано вибірку із 103 застосунків, яка послужила основою для побудови моделі.

Загалом, побудована нелінійна регресійна модель дозволяє адекватно оцінювати розмір програмних продуктів, які працюють на 2D графічних рушійх, що є важливим для підвищення достовірності оцінювання обсягу відповідного ПЗ і для покращення планування його розробки. Результати, отримані на тестовій вибірці, підтвердили ефективність моделі, що проявилася у значній достовірності прогнозних оцінок.

У межах дослідження було реалізовано програмний засіб для автоматизації оцінювання розміру застосунків, створених на основі 2D графічних рушіїв, який здійснює аналіз даних та побудову регресійної моделі з урахуванням визначених метрик. Зазначений інструмент може бути корисним у практичній діяльності розробників відповідного програмного забезпечення.

Список літератури:

1. Liu J. Design and development of 2D game “Adventurous spirit”. *Proceedings of the 2023 International Conference on Machine Learning and Automation*. P. 102–109. DOI: 10.54254/2755-2721/37/20230482
2. Software Metrics. URL: <https://www.imagix.com/links/software-metrics.html> (дата звернення: 12.01.25).
3. Pratama R. Y., Sarno R., Sholiq. Optimizing COCOMO II Parameters Using Artificial Bee Colony Method. *11th International Conference on Information & Communication Technology and System (ICTS)*. October 2017. P. 125–129. DOI: 10.1109/ICTS.2017.8265657
4. Макарова Л. М., Приходько Н. В., Кудін О. О. Побудова нелінійної регресійної моделі для оцінювання розміру веб-додатків, реалізованих мовою Java. *Вісник ХНТУ*. 2019. № 2(69). С. 145–153.
5. Петриченко С. А., Пухалевич А. В. Багатофакторна нелінійна регресійна модель для оцінювання розміру вебзастосунків, що створюються з використанням мови програмування C#. *Матеріали IV Всеукраїнської науково-практичної інтернет-конференції «Інформаційні технології: моделі, алгоритми, системи (ITMAS – 2023)»* (30–31 жовтня 2023 р., м. Миколаїв). Миколаїв : НУК імені адмірала Макарова, 2023. С. 40–42.
6. Латанська Л. О., Кольцов А. В. Нелінійна регресійна модель для оцінювання кількості строк коду веб-застосунків, що створюються з використанням РНР фреймворку Symfony. *Матеріали III Всеукраїнської науково-практичної інтернет-конференції «Інформаційні технології: моделі, алгоритми, системи (ITMAS – 2022)»* (26–28 жовтня 2022 р., м. Миколаїв). Миколаїв : НУК імені адмірала Макарова, 2022. С. 10–11.
7. Multivariate Normality Testing (Mardia). URL: <https://real-statistics.com/multivariate-statistics/multivariate-normal-distribution/multivariate-normality-testing/> (дата звернення: 14.01.25).
8. Mardia K. V. Measures of multivariate skewness and kurtosis with applications. *Biometrika*. 1970. Vol. 57. P. 519–530. DOI: <https://doi.org/10.1093/biomet/57.3.519>
9. Відстань Махаланобіса. URL: https://askom.gitbooks.io/-/content/23_rasstoyanie_mahalanobisa.html (дата звернення: 16.01.25).
10. SourceMeter. URL: <https://sourcemeter.com> (дата звернення: 12.12.24).

11. Prykhodko S., Prykhodko N., Makarova L., Pukhalevych A. Outlier Detection in Non-Linear Regression Analysis Based on the Normalizing Transformations. *Proceedings of the 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, IEEE. (Lviv-Slavske, 2020). P. 407–410.
12. Камінський С.С., Бризгалов М.В., Макарова Л.М. Математична модель для оцінювання розміру програмних застосунків, що використовують 2D графічні рушії. *Матеріали V Всеукраїнської науково-практичної інтернет конференції «Інформаційні технології: моделі, алгоритми, системи (ITMAS-2024)»*. (30–31 жовтня 2024 р., м. Миколаїв). Миколаїв, НУК імені адмірала Макарова, 2024. С. 69–70.
13. Prykhodko S. B., Prykhodko N. V., Koltsov A. V. A Nonlinear Regression Model for Early LOC Estimation of Open-Source Kotlin-Based Applications. *Radio Electronics Computer Science Control*. April 2024. DOI: 10.15588/1607-3274-2024-1-8

Makarova L. M., Kaminsky S. S., Bryzgalov M. V. MATHEMATICAL MODELING FOR SIZE ESTIMATION OF 2D APPLICATIONS BASED ON GRAPHICS ENGINES

The aim of this work is to address the problem of estimating the size of software applications that utilize 2D graphics engines. This issue is relevant, since reliable prediction of the size of program code affects the efficiency of development and estimation of the required resources. However, existing approaches to estimating the size of software products often do not take into account the specific features of programming for 2D graphics engines, which leads to inaccuracies in calculations. There is a need to create new models that would allow more accurate prediction of the code size required to implement applications of this type.

To solve this problem, we propose a mathematical model that takes into account key software metrics, such as the number of classes in the program code, the number of methods in the classes, and the depth of the inheritance tree. To increase the reliability of the estimation, a normalizing transformation in the form of a decimal logarithm is applied. The model also uses regression analysis methods that allow making reliable predictions of application size depending on the input parameters. The analysis of the results showed that the proposed model allows obtaining a reliable estimate of the software size, which makes it possible to optimize the development process, improve resource planning and reduce the risks of exceeding the project budget.

Based on the proposed mathematical model, software for estimating the size of applications was developed. The software was tested, which confirmed the reliability of predicting the KLOC metric (number of lines of code) in comparison with existing models for Java, C#, and PHP programming languages.

The results of the study can be useful for developers and project managers working with applications based on 2D graphics engines. They help to increase development efficiency by allowing to reasonably predict the amount of program code and resources required, which is an important factor in the process of creating high-quality software.

Key words: software size estimation, 2D graphics engines, software metrics, nonlinear regression model, normalizing transformation.