УДК 004.412:519.237.5

ОЦІНЮВАННЯ РОЗМІРУ ЗАСТОСУНКІВ МОВОЮ TYPESCRIPT

Трухов А.С. PhD (Комп'ютерні науки)[1]; Дімов В.С.[2]
[1,2] Національний університет кораблебудування імені адмірала Макарова
[1,2] Україна, Миколаїв
[1] artem.trukhov@nuos.edu.ua; [2] forij.main.mail@gmail.com

Анотація. *В роботі розглянуто проблему оцінювання розміру програмних застосунків, розроблених мовою TypeScript. Проаналізовано існуючі регресійні моделі та їх недоліки, зокрема ігнорування метрик структурної складності. На основі аналізу успішного застосування подібних метрик для Java та аналізу нових зібраних даних обґрунтовано актуальність розробки удосконаленої нелінійної моделі для TypeScript.*

Ключові слова: *оцінювання розміру ПЗ; typescript; нелінійна регресійна модель; метрики програмного забезпечення; коваріаційна матриця.*

Вступна частина. Важливість оцінки розміру проєкту на ранніх стадіях дозволяє зменшити ризики. Помилковий прогноз на ранніх етапах проєктування системи призводить до неправильного планування, що загрожує зривом термінів та виходом за межі бюджету на пізніших етапах.

Цільовою мовою програмування для дослідження було обрано TypeScript, оскільки він є не лише популярною, а й стратегічно важливою технологією. Як строго типізована надбудова над JavaScript, він вирішує ключові проблеми свого попередника, дозволяючи створювати більш надійний, масштабований та легкий у підтримці код. Актуальність мови підтверджується останніми даними: згідно з провідними опитуваннями, такими як Stack Overflow Survey [1], TypeScript посідає друге місце за популярністю серед мов програмування, поступаючись лише JavaScript, який він власне і доповнює. Проте, саме ці унікальні синтаксичні та структурні особливості мови роблять застосування універсальних моделей оцінювання, створених для інших технологій, недостатньо точним. Це створює нагальну потребу в розробці спеціалізованих інструментів.

Метою дослідження є огляд основних моделей для оцінювання розміру застосунків мовою TypeScript, аналіз їхніх обмежень та обґрунтування доцільності розробки нової, більш адекватної моделі на основі аналізу нових даних та метрик.

Основна частина Досліджуючи існуючі моделі [2] було виявлено, що переважна більшість має обмеження, що потребують вирішення. Їхня точність знижена через фундаментальне припущення про багатовимірний нормальний розподіл даних, що часто не виконується для реальних метрик програмного забезпечення. Ця проблема є системною, що підтверджено у роботах Приходько С.Б. [3], і вимагає застосування спеціалізованих методів - нормалізуючих перетворень.

Проте в останні роки було зроблено кілька регресійних моделей які враховують нормальний розподіл даних. Зокрема, в роботах Павлюк К.А. та Латанської Л.О. [4] було запропоновано нелінійну регресійну модель для проєктів на мові TypeScript у сфері криптовалютних операцій. Подібним чином, у праці Чебакова Д.В. [5] представлено удосконалену модель, що також базується на регресійному аналізі. Обидві моделі є значним кроком уперед, оскільки вони враховують недоліки попередніх робіт та орієнтовані

безпосередньо на TypeScript та використовують такі метрики, як кількість класів, та кількість атрибутів, у якості предикторів розміру.

Однак, навіть ці прогресивні моделі мають спільне, але ключове обмеження: вони ігнорують такий важливий аспект, як структурна складність застосунку, яка виражається у зв'язках між його компонентами. Кількість класів є важливою метрикою, але вона не відображає, наскільки тісно ці класи взаємопов'язані. Система з великою кількістю незалежних компонентів може бути простішою в розробці, ніж менша система з високою зв'язністю, де зміни в одному модулі викликають каскадні ефекти в інших. Саме ця зв'язність безпосередньо впливає на загальну складність і, як наслідок, на фінальний розмір проєкту.

Водночас дослідження моделей для архітектурно подібних мов, як-от Java [6], демонструють високу ефективність метрики «загальна кількість зв'язків» (total number of relationships). У роботі доведено, що включення цього параметра до нелінійної моделі дозволяє отримати значно точніші результати. На основі цього було сформовано гіпотезу, що введення аналогічної метрики структурної складності до моделі для TypeScript також дозволить суттєво підвищити її точність. Для перевірки цієї гіпотези було зібрано нові дані для низки TypeScript-проєктів. Проведений аналіз шляхом побудови коваріаційної матриці підтвердив, що запропонована метрика є статистично значущою.

В ході практичної перевірки гіпотези було виявлено проблему. Інструменти статичного аналізу, що використовувалися в розглянутих роботах [4-5], зокрема SonarQube, хоча й надають базові метрики, не мають функціоналу для обчислення загальної кількості залежностей між класами. Для вирішення цієї задачі було знайдено та адаптовано програмне рішення, що дозволяє генерувати UML-діаграми класів безпосередньо на основі вихідного коду TypeScript-проєктів. Використання цього інструментарію дало змогу автоматизувати збір розширеного набору метрик для вибірки з 49 відкритих TypeScript-проєктів. Були отримані наступні параметри: X1 – загальна кількість класів; X2 – середня кількість атрибутів на клас; X3 – загальна кількість залежностей між класами (включаючи наслідування, реалізацію інтерфейсів та асоціації); та Y – кількість рядків вихідного коду (SLOC).

Для емпіричного підтвердження значущості нововведеного параметра X3 на основі зібраних даних було побудовано коваріаційну матрицю. Аналіз матриці чітко продемонстрував високий ступінь кореляції між метрикою «загальна кількість залежностей» (X3) та фінальним розміром проєкту (Y), що підтверджує її високу прогностичну цінність та доцільність включення до удосконаленої регресійної моделі.

Висновки Проведений аналіз показав, що існуючі регресійні моделі для оцінювання розміру TypeScript-застосунків мають потенціал для вдосконалення. Було виявлено, що точність класичних підходів знижується через їхнє припущення про нормальний розподіл даних, яке не відповідає реальним метрикам програмного забезпечення. Крім того, навіть сучасні нелінійні моделі ігнорують такий важливий фактор, як структурна складність системи. Емпіричний аналіз нових даних, зібраних з 49 проєктів, підтвердив, що метрика «загальна кількість залежностей» є статистично значущою, демонструючи високу кореляцію з кінцевим розміром проєкту.

Таким чином, розробка нової нелінійної моделі, яка одночасно враховує негаусівський характер даних та включає показник структурної складності, є актуальною та обґрунтованою задачею. Це дозволить підвищити точність прогнозів і, як наслідок, покращити якість планування ресурсів та бюджету у сучасній веб-розробці.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] *Developer survey*. (2025). stackoverflow. https://survey.stackoverflow.co/2025/

[2] Tan, H. B. K., Zhao, Y., & Zhang, H. (2006). Estimating LOC for information systems from their conceptual data models. In *Software engineering: The 28th international conference (ICSE'06)* (с. 321–330). https://doi.org/10.1145/1134285.1134331

[3] Prykhodko, S., & Prykhodko, N. (2020). Mathematical modeling of non-gaussian dependent random variables by nonlinear regression models based on the multivariate normalizing transformations. In *Advances in intelligent systems and computing* (с. 166–174). Cham. https://doi.org/10.1007/978-3-030-58124-4_16

[4] Павлюк, К. А. (2022). *Нелінійна регресійна модель для оцінювання розміру програмних застосунків на TYPEScript* [Кваліфікаційна робота магістра, Національний університет кораблебудування імені адмірала Макарова]. https://eir.nuos.edu.ua/handle/123456789/8282

[5] Чебаков, Д. В. (2020). *Удосконалення регресійної моделі для оцінювання розміру програмних застосунків на TypeScript та розробка програми для її реалізації* [Кваліфікаційна робота магістра, Національний університет кораблебудування імені адмірала Макарова]. https://eir.nuos.edu.ua/handle/123456789/3707

[6] Prykhodko, N. V., & Prykhodko, S. B. (2018b). The non-linear regression model to estimate the software size of open source Java-based systems. *Radio Electronics, Computer Science, Control*, (3), 158–168. https://doi.org/10.15588/1607-3274-2018-3-17

Trukhov A.S. Ph.D.; Dimov V.S.

Evaluating the size of applications in the TypeScript language

Abstract*. This paper addresses the problem of estimating the size of software applications developed in the TypeScript language. Existing regression models and their drawbacks are analyzed, particularly their neglect of structural complexity metrics. Based on the analysis of the successful application of similar metrics for Java-based systems and the analysis of newly collected data, the relevance of developing an improved non-linear model for TypeScript is substantiated.*

Keywords*: software size estimation; typescript; non-linear regression model; software metrics; covariance matrix.*