УДК 004.412

НЕЛІНІЙНІ РЕГРЕСІЙНІ РІВНЯННЯ ДЛЯ ОЦІНЮВАННЯ ТРИВАЛОСТІ РОЗРОБКИ JAVA-ЗАСТОСУНКІВ ДЛЯ ПЛАТФОРМИ MIDRANGE

Приходько С.Б., д.т.н., проф.[1]; Фаріонова Т.А., к.т.н., доц.[2]
[1, 2] Національний університет кораблебудування імені адмірала Макарова
[1,2] Україна, Миколаїв
[1] ORCID 0000-0002-2325-018X; [2] ORCID 0000-0003-3384-4712

Анотація. *Проведено оцінку існуючої моделі ISBSG для оцінювання часу розробки програмних проєктів в залежності від трудомісткості їх розробки для платформи midrange. Побудовано нелінійні регресійні рівняння тривалості розробки Java-застосунків для платформи midrange з урахуванням типів ПЗ. Також порівняно значення коефіцієнту детермінації R^2, середньої величини відносної похибки MMRE та рівня прогнозування PRED(0,25) для трьох побудованих нелінійних регресійних рівнянь.*
Ключові слова: *трудомісткість; тривалість розробки програмного забезпечення; нелінійна регресія; Java-застосунок; платформа midrange*

Вступна частина. Midrange – клас комп'ютерних систем, який знаходиться між мейнфреймами і мікрокомп'ютерами. Це системи середнього рівня, в першу чергу високопродуктивні мережеві сервери та інші типи серверів, які можуть обробляти великомасштабну обробку багатьох бізнес-додатків. Вони не такі потужні, як мейнфрейми, однак їх покупка, експлуатація та обслуговування дешевше, ніж у мейнфреймів, і, таким чином, вони задовольняють обчислювальні потреби багатьох організацій. Midrange стали популярними як потужні мережеві сервери, які допомагають керувати великими веб-сайтами в Інтернеті, корпоративними інтрамережами, тощо. Сьогодні системи середнього рівня включають в себе сервери, що використовуються на промислових підприємствах і виробничих підприємствах, грають важливу роль в автоматизованому виробництві (CAM). Вони також можуть приймати форму потужних технічних робочих станцій для автоматизованого проектування (САПР) та інших обчислювальних і графічних додатків. Системи середнього рівня також використовуються в якості зовнішніх серверів для допомоги мейнфрейм-комп'ютерам в телекомунікаційній обробці та управлінні мережею [1].

Згідно з PMBOK та стандарту ISO/IEC 12207 – Software life cycle processes – процес оцінювання тривалості робіт з розробки програмного забезпечення (ПЗ) є невід'ємною складовою інженерії вимірювання ПЗ [2]. За статистикою біля 30% програмних проєктів мають труднощі з реалізацією (не вкладаються в бюджет, виходять з порушенням часових обмежень, або неналежної якості). Програмні проєкти для платформи midrange характеризуються своїм значним розміром, великою кількістю компонентів та спеціальними вимогами до надійності. Тому достовірне оцінювання тривалості розробки програмних проектів для цієї платформи є актуальним завданням.

За статистичними даними компанії ISBSG (International Software Benchmarking Standards Group) (https://www.isbsg.org/about-isbsg), частка ПЗ для платформи midrange, основною мовою розробки якого є Java, складає приблизно 25%.

Метою роботи є підвищення достовірності оцінювання тривалості розробки Java-застосунків для платформи midrange за рахунок побудови нелінійних регресійних рівнянь**.**

Основна частина. Найбільш поширеними моделями, для оцінювання тривалості програмних проектів є COCOMO [3] і ISBSG [4]: нелінійні регресійні рівняння для оцінювання тривалості програмних проектів в залежності від трудомісткості. У зв'язку з тим, що емпіричні дані тривалості програмних проектів та трудомісткості програмних проектів мають закон розподілу, який відрізняється від гаусівського [3-5], для побудови таких моделей виконувалась нормалізація емпіричних даних з тривалості та трудомісткості розробки програмних проектів. Для побудови моделі ISBSG для платформи midrange, було використано десятковий логарифм для нормалізації емпіричних даних тривалості програмних проектів та емпіричних даних трудомісткості програмних проектів [4]:

$$D = 0{,}548\, E^{0{,}360}. \tag{1}$$

Тут D –тривалість програмних проектів з розробки ПЗ (місяці), E – трудомісткість розробки програмних проектів для платформи midrange (люд.-год.). Однак, слід зазначити, що модель (1) була розроблена за даними проєктів 1989-1996 рр.

Для оцінки моделі (1) було використано дані з 529 програмних проєктів з репозитарію ISBSG та отримані такі результати її якості: R^2=0,253; MMRE=0,708; PRED(0,25)=0,301. Аналізуючи результати, можна зробити висновок, що модель (1) має невисоку якість.

Щоб вдосконалити модель ISBSG для оцінювання тривалості розробки програмних проектів для midrange (1) як додаткову ознаку впливу визначимо мову розробки ПЗ. А саме побудуємо нелінійне регресійне рівняння для оцінювання тривалості розробки Java-застосунків платформи midrange. В якості даних було використано 129 Java проєктів з репозитарію ISBSG.

Емпіричні дані тривалості програмних проектів та трудомісткості мають закон розподілу, який відрізняється від гаусівського [4, 5]. Тому для побудови нелінійного регресійного рівняння авторами використані нормалізовані значення тривалості програмних проектів та нормалізовані значення трудомісткості. Нормалізацію випадкової величини виконуємо за перетворенням у вигляді десяткового логарифму. Це дозволить побудувати рівняння лінійної регресії для нормалізованих значень

$$\hat{z}_D(z_E) = b_0 + b_1 z_E \tag{2}$$

де z_D, z_E – нормалізовані випадкові величини D та E; b_0 і b_1 – коефіцієнти рівняння лінійної регресії.

Після перевірки даних на викиди, коефіцієнти нелінійного регресійного рівняння для оцінювання тривалості розробки Java-застосунків платформи midrange мають наступні значення: b_0 =-0,0778, b_1=0,273. Побудоване рівняння нелінійної регресії тривалості програмних проектів для midrange має наступні характеристики: R^2=0,3849; MMRE=0,2676; PRED(0,25)=0,57. Таким чином, ми отримали кращі оцінки за параметрами R^2, MMRE та PRED(0,25).

В наборі даних Java-застосунків платформи midrange, що використовувалися в розрахунках, були присутні як нові проєкти (ND), так й такі, що вдосконалювалися (E). Тому авторами було запропоновано покращити оцінку за рахунок побудови окремо нелінійного рівняння регресії як для (ND), так й для (E). В табл. 1 наведені результати розрахунків.

З табл. 1 можна зробити висновок, що тип розробки (ND або E) має вплив на якість оцінки. Однак, необхідно проводити подальше дослідження факторів впливу з метою побудови покращеною нелінійної регресійної моделі для оцінювання тривалості розробки програмних проектів для midrange.

Таблиця 1 – Порівняльні результати розрахунків Java-застосунків платформи midrange

Показники моделі	New Development (ND), n=32	Enhancement (E), n=97
b_0	-0,74	-0,005
b_1	0,418	0,252
R^2	0,5916	0,3251
MMRE	0,3564	0,2979
PRED(0,25)	0,4	0,5974

Як видно з табл. 1, значення наведених параметрів кращі для нелінійного регресійного рівняння для проєктів, що вдосконалюються. Однак прийнятні значення MMRE та PRED(0,25) (не більше 0,25 та не менше 0,75 відповідно) для нелінійної регресії з використанням нормалізуючого перетворення у вигляді десяткового логарифму не досягнуті, що свідчить про необхідність застосування більш складних перетворень, наприклад, нормалізуючого перетворення Джонсона.

Висновки. Проведено оцінку існуючої моделі ISBSG для оцінювання часу розробки програмних проєктів в залежності від трудомісткості їх розробки для платформи midrange. Побудовано нелінійні регресійні рівняння тривалості розробки Java-застосунків для платформи midrange з урахуванням типів ПЗ. Порівняно значення оцінок R^2, MMRE та PRED(0,25) для трьох побудованих нелінійних регресійних рівнянь. Перспектива подальших досліджень полягає у побудові нелінійних регресійних моделей із врахуванням впливу додаткових чинників, таких як сфера використання програмного продукту (банківська справа, виробництво, телекомунікації тощо), функціональний розмір тощо. Слід також дослідити використання нелінійних моделей, включення інших змінних у процес моделювання та використання інших видів нормалізуючих перетворень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Компьютер среднего уровня – Midrange computer. (2021). Retrieved from: https://star-wiki.ru/wiki/Midrange_computer.
[2] Лавріщева, К.М. (2008). *Програмна інженерія*. Київ: Академперіодика
[3] Boehm, B. W. (1981). *Software engineering economics*. Englewood Cliffs, NJ : Prentice Hall.
[4] Oligny, S., Bourque, P., Abran, A, Fournie, B. (2000). Exploring the relation between effort and duration in software engineering projects. *In proc. of the World Computer Congress*, 175-178.
[5] Приходько С.Б., Пухалевич А.В. (2012). Розробка нелінійної регресійної моделі тривалості програмних проектів на основі нормалізуючого перетворення Джонсона. *Радіоелектронні і комп'ютерні системи, 4 (56)*, 90-93.

Prykhodko S.B., Farionova T.A.

Nonlinear regression equations for estimating the duration of Java-applications development for the midrange platform

Abstract. *An evaluation of the existing ISBSG model for estimating the time of software project development depending on the effort of their development for the midrange platform. Nonlinear regression equations of duration of development of Java-applications for the midrange platform taking into account software types are constructed. The values of the coefficient of determination R^2, the average value of the relative error MMRE and the prediction level PRED (0.25) for the three constructed nonlinear regression equations were also compared.*

Keywords: *effort; duration of software development; nonlinear regression; Java application; midrange platform.*