

НЕЛІНІЙНІ РЕГРЕСІЙНІ МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ ТРИВАЛОСТІ РОЗРОБКИ ПРОГРАМНИХ ЗАСТОСУНКІВ, ЩО СТВОРЮЮТЬСЯ МОВАМИ JAVA ТА VISUAL BASIC ДЛЯ ПК

Приходько С.Б., д.т.н., проф.¹; Пухалевич А.В., к.т.н.²; Халанчук І.П.²
^{1,2} Національний університет кораблебудування імені адмірала Макарова
^{1,2} Україна, Миколаїв
¹ sergiy.prykhodko@nuos.edu; ² andrii.pukhalevych@nuos.edu.ua

***Анотація.** Побудовано нелінійні регресійні моделі для оцінювання тривалості розробки в залежності від трудомісткості розробки програмних застосунків, що створюються мовами Java та Visual Basic для ПК.*

***Ключові слова:** нелінійна регресія; оцінювання тривалості; розробка застосунків.*

Вступна частина. Оцінювання тривалості розробки програмних застосунків є важливою задачею програмної інженерії, так як оцінки тривалості використовуються при заключенні договорів на розробку цих застосунків. Це вимагає математичних моделей, які дозволять виконати відповідне оцінювання.

До моделей, що використовуються для оцінювання тривалості розробки програмних застосунків, належать моделі ISBSG – нелінійні регресійні рівняння для оцінювання тривалості розробки програмного забезпечення в залежності від трудомісткості, які було побудовано для платформ PC (персональні комп'ютери – ПК), mid-range та mainframe [1]. Аналіз даних репозитарію ISBSG з релізу 2021 року [2] показав, що поділ моделей тільки по платформі вже є недостатнім, оскільки отримані моделі мають погані значення PRED та MMRE. Отже, виникає необхідність додавання інших ознак кластеризації даних. Однією з таких ознак, окрім платформи, природньо розглядати ще мову програмування.

Метою роботи є побудова нелінійних регресійних моделей для оцінювання тривалості розробки в залежності від трудомісткості розробки програмних застосунків, що створюються мовами Java та Visual Basic для ПК.

Основна частина. Для побудови нелінійних регресійних моделей для оцінювання тривалості розробки програмних застосунків, що створюються мовами Java та Visual Basic для ПК, використано метод побудови нелінійних регресійних моделей на основі нормалізуючих перетворень [3]. Цей метод полягає в тому, що емпіричні дані спочатку нормалізуються з застосуванням нормалізуючого перетворення, після чого будується лінійна регресійна модель за нормалізованими даними, а далі будується нелінійна регресійна модель з застосуванням перетворення, зворотного до нормалізуючого.

Щоб було можливо побудувати лінійну регресійну модель, нормалізовані дані мають мати нормальний розподіл, а відстань Махаланобіса для кожної пари значень має бути меншою чи дорівнювати критичному значенню. Розподіл залишків лінійної регресії повинен бути гаусівським, а нормалізовані значення емпіричних даних не повинні виходити за інтервал передбачення лінійної регресійної моделі. Тому з емпіричних даних потрібно видалити викиди, для чого виконуються наступні кроки:

1. Обчислюємо значення багатовимірної асиметрії нормалізованих емпіричних даних. Якщо це значення більше, ніж критичне значення, то з емпіричних даних видаляється пара значень з найбільшою відстанню Махаланобіса та знову переходимо до п. 1.

2. Обчислюємо значення багатовимірною ексцесу нормалізованих емпіричних даних. Якщо це значення більше, ніж критичне значення, то з емпіричних даних видаляється пара значень з найбільшою відстанню Махаланобіса та переходимо до п. 1.
3. За нормалізованими даними будуємо лінійну регресійну модель та перевіряємо закон розподілу залишків. Якщо розподіл не є гаусівським, то з емпіричних даних видаляємо пару значень з найбільшим за модулем залишку та переходимо до п. 1.
4. Якщо одне або більше нормалізованих значень даних виходить за інтервал передбачення лінійної регресії для нормалізованих даних, то ці значення видаляються та переходимо до п. 1.

Після видалення викидів за отриманою лінійною регресійною моделлю будується нелінійна регресійна модель із застосуванням перетворення, зворотного до нормалізуючого.

Емпіричними даними, що використовувалися при побудові нелінійних регресійних моделей для оцінювання тривалості розробки в залежності від трудомісткості розробки програмних застосунків, що створюються мовами Java та Visual Basic для ПК є дані репозитарію ISBSG з релізу 2021 року [2]. Нормалізація емпіричних даних виконувалася перетворенням у вигляді десяткового логарифму.

Для мови Java кількість емпіричних значень була 39, з них 7 значень було визначено як викиди. За значеннями, що залишилися було побудовано модель з наступними показниками якості: $R^2=0,9134$, $PRED(0,25)=0,9063$, $MMRE=0,1183$.

Для мови Visual Basic кількість емпіричних значень була 106, з них 27 значень було визначено як викиди. За значеннями, що залишилися було побудовано нелінійну регресійну модель з наступними показниками якості: $R^2=0,6487$, $PRED(0,25)=0,5063$, $MMRE=0,311$.

Висновки. В роботі було побудовано нелінійні регресійні моделі для оцінювання тривалості розробки програмних застосунків, що створюються мовами Java та Visual Basic для ПК. Показники якості побудованих моделей є кращими за відповідні показники нелінійної регресійної моделі, що створена для платформи ПК без врахування мови програмування. В подальшому планується розглянути можливість використання інших нормалізуючих перетворень при побудові нелінійних регресійних моделей для оцінювання тривалості розробки таких застосунків.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Oligny, S., Bourque, P., Abran, A. & Fournier, B. (2000). Exploring the relation between effort and duration in software engineering projects. *Proc. of the 16th IFIP World Computer Congress* (pp. 175-178). Aug, 2000, Beijing, China
- [2] *International Software Benchmarking Standards Group*. (2021). Retrieved from <https://www.isbsg.org/>.
- [3] Приходько, С.Б. (2012). Метод побудови нелінійних рівнянь регресії на основі нормалізуючих перетворень. *Тези доповідей міждерж. наук.-методич. конф. "Проблеми математичного моделювання"* (с. 31-33). 13-15 червня 2012, Дніпродзержинськ, Дніпродзержинськ : ДДТУ.

Prykhodko S.B., Pukhalevych A.V., Khalanchuk I.P.

Nonlinear regression models for the development duration estimation of software applications created in Java and Visual Basic for PC

Abstract. *Nonlinear regression models to estimate the development duration depending on the effort of developing software applications created in Java and Visual Basic for PC are built.*

Keywords: *nonlinear regression; duration estimation; software development.*