

НЕЛІНІЙНЕ РЕГРЕСІЙНЕ РІВНЯННЯ ДЛЯ ОЦІНЮВАННЯ ТРИВАЛОСТІ РОЗРОБКИ COBOL – ЗАСТОСУНКІВ ДЛЯ ПЛАТФОРМИ MAINFRAME

Гусєва-Божаткіна В. А.

Національний університет кораблебудування імені адмірала Макарова

Україна, Миколаїв

guseva.bozhatkina@nuos.edu.ua; ORCID: 0000-0002-1117-3391

Анотація. Розроблено нелінійне регресійне рівняння для оцінювання тривалості розробки Cobol-застосунків для платформи Mainframe. Проведено оцінку існуючої моделі ISBSG для оцінювання часу розробки програмних застосунків в залежності від трудомісткості їх розробки для платформи Mainframe. Отримана модель у порівнянні з моделлю ISBSG, підвищує достовірність оцінювання тривалості розробки Cobol-застосунків для платформи Mainframe.

Ключові слова: програмний застосунок; тривалість розробки програмних застосунків; платформа Mainframe; регресійний аналіз; мова Cobol.

Вступна частина. Розробка та реінжиніринг програмних застосунків мовою програмування Cobol досі є більш актуальною темою для багатьох компаній, зокрема банків та фінустанов [1]. Варто підкреслити окремо, що мова програмування Cobol наразі не є досить поширеною і відповідного роду кваліфікації програмістів стає все менше. Поступово, у зв'язку зі збільшенням завдань від замовників, дана сфера діяльності переходить до невеликих аутсорсових організацій. Можна констатувати, що успіх програмних проєктів на початкових стадіях залежить від рішень, які приймаються проджект-менеджерами [2].

Актуальність проблеми отримання ефективної системи оцінювання тривалості розробки та реінжинірингу програмних застосунків в даний час є важливим завданням, що вимагає удосконалення існуючих методів, бо:

- за статистикою близько 30% програмних проєктів мають труднощі з реалізацією (не вкладаються в бюджет, виходять з порушенням часових обмежень, або неналежної якості);
- програмні проєкти для платформи Mainframe характеризуються своїм значним розміром, великою кількістю компонентів та спеціальними вимогами до надійності;
- частка програмного забезпечення для платформи Mainframe, основною мовою розробки якого є Cobol, складає приблизно 17,8%;
- існуючі моделі COCOMO та ISBSG не враховують особливості програмного продукту, такі як мова та засоби розробки, сфера використання. Для цих моделей немає побудованих рівнянь довірчих інтервалів та рівнянь інтервалів прогнозування тривалості.

Метою роботи є підвищення достовірності оцінювання тривалості розробки програмних застосунків мовою програмування Cobol за рахунок побудови нелінійної регресійної моделі.

Виходячи з поставленої мети та проблематики теми дослідження необхідно вирішити наступні завдання:

- провести аналіз існуючих методів та моделей оцінювання тривалості програмних застосунків;

- побудувати нелінійну регресійну модель тривалості розробки застосунків на Cobol для Mainframe в залежності від їх трудомісткості, шляхом побудови рівняння нелінійної регресії;
- на основі удосконалених моделей розробити програмне забезпечення для оцінювання тривалості розробки та Cobol-застосунків для платформи Mainframe.

У якості об'єкта дослідження було обрано процес оцінювання тривалості розробки Cobol-застосунків для платформи Mainframe.

Предметом дослідження є математична модель для оцінювання тривалості розробки Cobol-застосунків для платформи Mainframe та безпосередньо програма для їх реалізації.

Наукова новизна одержаних результатів полягає в отриманні подальшого розвитку моделі ISBSG для оцінювання тривалості розробки програмних застосунків для платформи Mainframe в залежності від трудомісткості їх розробки, за рахунок створення нелінійної регресійної моделі для Cobol-застосунків, яка побудована з урахуванням викидів. Це дозволяє підвищити достовірність оцінювання тривалості розробки Cobol- застосунків для платформи Mainframe.

Основна частина. Завдання оцінювання тривалості розробки програмних застосунків на ранній стадії є важливою задачею, оскільки ця інформація використовується для прогнозування часу розробки ПЗ за допомогою таких моделей як COCOMO II [3; 4]. Наразі існує певною мірою поширена проблема – не існує єдиного вимірювання, набору метрик та показників для оцінювання тривалості розробки ПЗ. Однак існують методи для оцінювання тривалості розробки ПЗ, деякі з них базуються на кількості строк коду, інші – обчислюють тривалість з функціональних, технічних або інших аспектів [1; 5]. Але більшість з них стають непридатними до використання через відсутність даних, ресурсів або експертних навичок в цій галузі.

Станом на сьогодні існує досить велика кількість методів для оцінювання тривалості розробки ПЗ. Нормалізуючі перетворення дуже часто є надійним способом побудови нелінійних регресійних моделей та рівнянь. Однак добре відомі методи їх побудови, що засновані на одновимірних нормалізуючих перетвореннях (таких як логарифмічне, Бокса-Кокса, Джонсона), не враховують кореляції між випадковими змінними у разі нормалізації негаусових даних. Це призводить до необхідності використання нормалізуючих перетворень, які враховують цю кореляцію, для побудови нелінійних регресійних моделей та рівнянь для оцінювання тривалості ПЗ інформаційних систем, в тому числі з відкритим кодом на Cobol [6].

Враховуючи рекомендації зі [7], загальний алгоритм процесу побудови нелінійної регресійної моделі (після процесу нормалізації емпіричних даних) має ітераційний характер та проходить в декілька етапів, що представлено у вигляді діаграми на рисунку 1.

Для побудови нелінійної регресійної моделі були використані дані за 124-ма програмними продуктами мовою програмування Cobol для платформи Mainframe з репозитарію International Software Benchmarking Standards Group (ISBSG) 2021 р. [5]. Вибір впав саме на цей репозитарій, через велику кількість програмних проєктів, що в ньому присутні, для яких вже визначена якість зібраних статистичних даних і даних щодо тривалості та трудомісткості розробки.

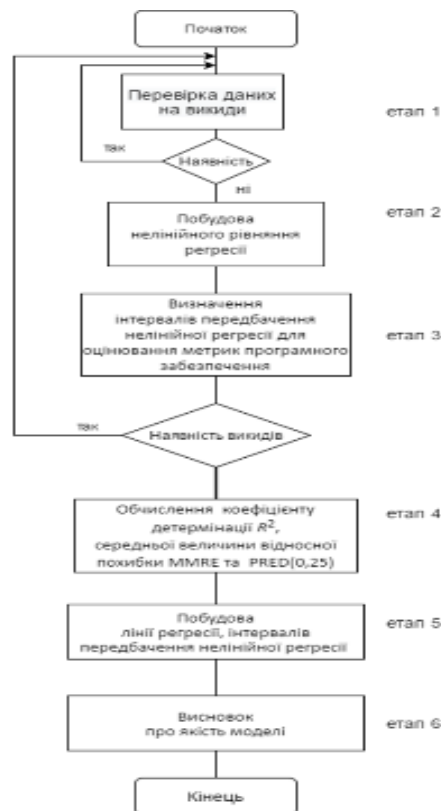


Рисунок 1 – Алгоритм побудови нелінійної регресійної моделі

На рисунку 2 зображено тривалість розробки Cobol-застосунків для платформи Mainframe в залежності від трудомісткості цих проектів за даними репозитарію ISBSG.

Проведено аналіз отриманих результатів розрахунків з моделлю ISBSG для платформи Mainframe. Показники побудованої нелінійної моделі регресії для оцінювання тривалості розробки Cobol-застосунків для платформи Mainframe наступні:

Якість побудованої моделі: $n = 124$; $R^2 = 0,3683$ $PRED(0,25) = 0,3952$; $MMRE = 0,5325$.

Якість моделі ISBSG: $n = 124$; $R^2 = 0,2962$; $PRED(0,25) = 0,4032$; $MMRE = 0,4846$.

Після відкидання викидів, зі 124 застосунків залишилось 97. Показник R^2 вищий ніж у моделі ISBSG, що говорить про те, що отримана модель більш достовірна.

Якість побудованої моделі: $n = 97$; $R^2 = 0,5743$; $PRED(0,25) = 0,4845$; $MMRE = 0,3002$.

Якість моделі ISBSG: $n = 97$; $R^2 = 0,5123$; $PRED(0,25) = 0,5155$; $MMRE = 0,2859$.

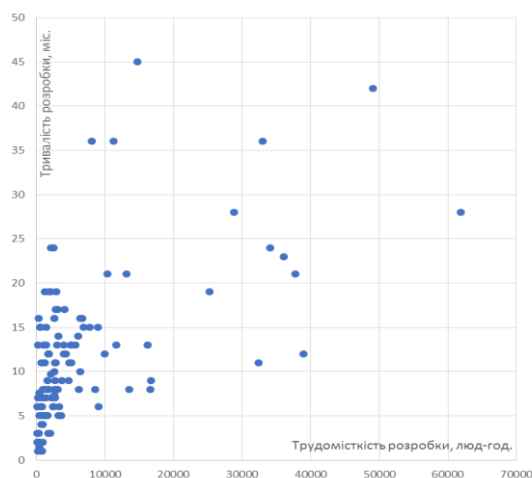


Рисунок 2 – Тривалість розробки Cobol-застосунків для платформи Mainframe в залежності від трудомісткості цих проектів за даними репозитарію ISBSG

Побудована лінійна модель регресії: $z_W = 0.312 - 0.086z_P + \varepsilon$.

Побудована нелінійна модель регресії: $P = 10^{\varepsilon - 0.086} \cdot W^{0.312}$.

Отримана модель регресії тривалості розробки застосунків, написаних мовою програмування Cobol для платформи Mainframe в залежності від трудомісткості розробки таких застосунків на основі десятичного логарифма має кращі характеристики R^2 , ніж ISBSG модель.

Проведений аналіз показав необхідність проведення подальших досліджень з метою побудови покращеної нелінійної регресійної моделі для оцінювання тривалості розробки застосунків, написаних мовою програмування Cobol для платформи Mainframe.

Висновки. Розроблена нелінійна регресійна модель тривалості розробки Cobol-застосунків для платформи Mainframe (з урахування типів ПЗ). Отримана модель у порівнянні з моделлю ISBSG, має кращі характеристики R^2 . Доцільно також розглянути можливість більш складних перетворень та дослідити вплив додаткових чинників, таких як: сфера використання програмного продукту (банківська справа, виробництво, телекомунікації тощо), функціональний розмір та ін.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Geet, J. V., Demeyer, S. (2009). Feature location in COBOL mainframe systems: an experience report. *Software Maintenance (ICSM 2009): Proc. of the 25th IEEE International Conference on*, Sept. 20–26. (2009). Edmonton, Alberta, Canada. DOI: 10.1109/ICSM.2009.5306312.
- [2] A guide to the project management body of knowledge. Fifth edition. Project Management Institute. (2013). 589 p.
- [3] Capers, J. (2013). Function points as a universal software metric. *ACM SIGSOFT Software Engineering Notes*. Vol. 38, No. 4. (P. 1–27). DOI: 10.1145/2492248.2492268.
- [4] International Software Benchmarking Standards Group (2016, March). ISBSG Data Demographics D&E Repository Release 2016 R1. <http://isbsg.org/wp-content/uploads/2016/04/ISBSG-DE-Demographics-2016-R1.pdf>.
- [5] Kissell, R. L. (2021). Nonlinear regression modeling. In book: *Algorithmic trading methods*. DOI: 10.1016/B978-0-12-815630-8.00008-9.
- [6] Sneed, H., Erdoes, K. (2013). Migrating AS400-COBOL to Java: A report from the field. *Software Maintenance and Reengineering (CSMR): Proc. of the 2013 17th European Conference*. DOI: 10.1109/CSMR.2013.32.
- [7] Prykhodko, S. B., Pukhalevich, A. V. (2014). Developing PC Software Project Duration Model based on Johnson transformation. *Proc. of the 12th Int. Conf. Modern Problems of Radio Engineering, Telecommunications and Computer Science TCSET'2014, Lviv – Slavske, Ukraine. Lviv Polytechnic National University*. P. 114–116.

Guseva-Bozhatkina V.A.

Nonlinear regression equation for estimating the development durations of Cobol – applications for the mainframe platform

Abstract. A nonlinear regression equation was developed to estimate the duration of development of Cobol applications for the Mainframe platform. An evaluation of the existing ISBSG model for estimating the development time of software applications depending on the complexity of their development for the Mainframe platform was carried out. The obtained model, in comparison with the ISBSG model, indicates the reliability of estimating the duration of development of Cobol applications for the Mainframe platform.

Keywords: Software application; duration of software application; Mainframe platform; regression analysis; Cobol language.