

УДОСКОНАЛЕННЯ МАТЕМАТИЧНОЇ МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ МУЛЬТИМЕДІЙНИХ ANDROID ДОДАТКІВ, НАПИСАНИХ МОВОЮ JAVA

Рябков С.І., ст. гр. 6151м, Приходько С.Б., доктор технічних наук
Національний університет кораблебудування ім. адм. Макарова
Україна, Миколаїв
141255@nuos.edu.ua, sergiy.prykhodko@nuos.edu.ua

Анотація. Під час дослідження було розроблено лінійну регресійну модель для оцінки розміру мультимедійних Android додатків, написаних мовою Java. Отримана модель, в порівнянні з іншими моделями, має більший коефіцієнт множинної детермінації, менше значення величини середньої похибки та більше величину відсотка прогнозування.

Ключові слова: прогнозування розміру програмного забезпечення, Java Android додатки, лінійна регресійна модель.

Вступна частина. Проблема оцінювання розміру програмного забезпечення (ПЗ) на ранніх етапах розробки є важливою [2, 3], оскільки результати, отримані при оцінюванні розміру ПЗ, використовуються для прогнозування трудомісткості розробки ПЗ за допомогою таких моделей як СОСОМО II, СОСОМО 2000 та СОСОМО 81. Це призводить до необхідності розробки відповідних моделей для оцінювання розміру ПЗ, включаючи мультимедійні Android додатки, що написані мовою Java. Множинна лінійна регресія – це статистична методика, яка використовує кілька пояснювальних змінних (незалежні змінні) для прогнозування результату змінної відповіді (залежна змінна), а також є розширенням простої лінійної регресійної моделі. Завдяки своїй гнучкості моделювання множинної регресії стало важливою методикою статистичного аналізу в багатьох сферах. Найголовнішою метою множинної лінійної регресії є моделювання зв'язку між пояснювальними змінними та змінними відповіді. Множинна лінійна регресія належить до алгоритмічної групи методик. Таким чином, для того щоб оцінити програмний проект, для цієї моделі потрібні дані минулих проектів. На даний час існують рішення для оцінки інформаційних систем, написаних мовою Java [1]. У даній роботі лінійної регресії на основі метрик концептуальної моделі даних у вигляді діаграми класів для оцінювання інформаційних систем, проте Android додатки мають власну специфіку розробки та, приймаючи до уваги новизну сфери програмування, обраних метрик може бути недостатньо для точної оцінки, що призводить до необхідності створення власної моделі.

Основна частина. Рівняння множинної регресії може бути представлено виразом:

$$Y = b_0 + \sum (b_n X_n) + \varepsilon$$

де b_0 - вільний член, який визначає значення Y , в разі, коли всі пояснюючі змінні X_n рівні 0, b_n - коефіцієнти змінних, а ε - залишкове значення або похибка. Для побудови моделі лінійної множинної регресії оцінки розміру мультимедійних Android додатків було зібрано дані з 35 програмних проектів, дані представлені у Таблиці 1, Де Y - кількість строк коду, $X1$ - кількість абстрактних методів, $X2$ - кількість публічних методів, $X3$ - кількість анонімних класів, $X4$ - кількість підкласів.

Для оцінки параметрів рівняння множинної регресії скористаємося методом найменших квадратів. Для початку визначимо вектор оцінки коефіцієнтів регресії. Відповідно до методу найменших квадратів, вектор s знаходиться з вектору:

$$S = (X^T X)^{-1} (X^T Y)$$

Виконавши обчислення отримаємо наступні коефіцієнти: $b_0 = 0$, $b_1 = 76.727$, $b_2 = 41.957$, $b_3 = 74.86$. Рівняння регресії приймає вигляд:

$$Y = 0 + 76.727X_1 + 41.957X_2 + 39.355X_3 + 74.86X_4$$

Таблиця 1 – Дані по Android проектам

[illegible]

Необхідно провести статистичний аналіз отриманого рівняння регресії: пошуку стандартної похибки та дисперсії параметрів моделі, перевірки значимості коефіцієнтів, коефіцієнту детермінації (R^2), значення величини середньої похибки ($MMRE$) та величини відсотка прогнозування ($PRED(0.25)$)

Оцінка дисперсії знаходиться за формулою:

$$S_e^2 = (Y - Y(X))^T (Y - Y(X)) = 121636468.183$$

Наступним кроком знайдемо незміщену оцінку дисперсії:

$$S^2 = \frac{1}{n-m} * S_e^2 = \frac{1}{35-4} * 121636468.183 = 3923757.036$$

Отримавши незміщену оцінку дисперсії можемо знайти значення стандартної похибки:

$$S = \sqrt{S^2} = \sqrt{3923757.036} = 1980.848$$

Для того, щоб в подальшому провести перевірку на значимість коефіцієнтів регресії потрібно знайти дисперсію параметрів моделі, а отже необхідно побудувати оцінку коваріантної матриці:

$$k = S^2 * (X^T X)^{-1}$$

Коваріантна матриця представлена у Таблиці 2.

Таблиця 2 – Коваріантна матриця $k = S^2 * (X^T X)^{-1}$

366786.422	-2775.029	113.512	-1194.768	-1788.871
-2775.022	361.418	-41.268	42.902	-30.834
113.542	-41.241	18.5	-15.85	-4.636
-1194.748	42.867	-15.838	28.656	-7.033
-1788.867	-30.815	-4.622	-7.014	64.994

Дисперсії параметрів моделі відповідають співвідношенням $S_i^2 = K_{ij}$, тобто це елементи, що знаходяться на головній діагоналі:

$$S_{b1} = \sqrt{361.418} = 19.011, S_{b2} = \sqrt{18.5} = 4.301$$

$$S_{b3} = \sqrt{28.656} = 5.353, S_{b4} = \sqrt{64.994} = 8.062$$

Для перевірки значимості параметрів множинної регресії використовують t-статистику (або t-критерій Стюдента). Критичне значення для даної моделі має значення:

$$t_{крит} = (31; 0.25) = 2.0395$$

Якщо відношення дисперсії параметру до його величини більше за критичне, то можна робити висновок про значимість параметру:

$$t_i = \frac{b_i}{S_{bi}}$$

$$t_1 = \frac{76.727}{19.011} = 4.036 > 2.0395$$

$$t_2 = \frac{41.957}{4.301} = 9.755 > 2.0395$$

$$t_3 = \frac{39.355}{5.353} = 7.352 > 2.0395$$

$$t_4 = \frac{76.86}{8.062} = 9.286 > 2.0395$$

З отриманих результатів можна зробити висновок, що статистична значимість підтверджується для усіх коефіцієнтів.

Тісноту впливу факторів на результат оцінює індекс множинної кореляції. На відміну від парного коефіцієнта кореляції, який може приймати негативні значення, він приймає значення від 0 до 1. При значенні R близькому до 1, рівняння регресії краще описує фактичні дані і фактори сильніше впливають на результат. При значенні R близькому до 0 рівняння регресії погано описує фактичні дані і фактори мають слабкий вплив на результат. В отриманому рівнянні регресії константа дорівнює нулю, а отже коефіцієнт множинної кореляції знаходиться за формулою:

$$R = \sqrt{1 - \frac{S_e^2}{\sum Y_i^2}} = \sqrt{1 - \frac{121636468.183}{12045298828}} = 0.994$$

Отримавши коефіцієнт множинної кореляції знайдемо коефіцієнт детермінації:

$$R^2 = 0.994^2 = 0.98$$

За результатами можна зробити висновок, що зв'язок між Y та факторами X сильний.

Величина середньої відносної похибки ($MMRE$) та відсотка прогнозування ($PRED(0.25)$) приймаються як стандарт оцінки результатів прогнозування регресійних моделей та рівнянь. Допустимі значення $MMRE$ та $PRED(0.25)$ не більше 0,25 та не менше 0,75 відповідно. В Таблиці 1 наведено отримані значення MRE , знайдемо значення $MMRE$ та $PRED(0.25)$:

$$MMRE = \frac{\sum MRE_i}{i} = \frac{4.961}{35} = 0.1417$$

$$PRED(0.25) = \frac{35 - 4}{35} = 0.886$$

З результатів можна зробити висновок, що отримана регресійна модель є достатньо точною

Висновки. У ході даної роботи було побудовано лінійну регресійну модель з наступними характеристиками $R^2 = 0.98$, $MMRE = 0.1417$, $PRED(0.25) = 0.886$. Модель для оцінки інформаційних систем, написаних мовою Java, приведена у роботі [1] має наступні характеристики $R^2 = 0.9737$, $MMRE = 0.15$, $PRED(0.25) = 0.81$. Порівнюючи отримані результати можна зробити висновок, що отримана модель має більший коефіцієнт множинної детермінації, менше значення величини середньої похибки та більшу величину відсотка прогнозування, а отже більше підходить для оцінювання розміру мультимедійних Android додатків.

REFERENCES

- [1] Tan H. B. K. Estimating LOC for information systems from their conceptual data models / H. B. K. Tan, Y. Zhao, H. Zhang // Software Engineering : the 28th International Conference (ICSE '06), Shanghai, China, May 20–28, 2006 : proceedings. – P. 321–330. DOI: 10.1145/1134285.1134331
- [2] Boehm, B. W. et al. Software Cost Estimation with COCOMO II. Prentice Hall, 2000.
- [3] Canfora, G., Cerulo, L., and Troiano, L. An experience of fuzzy linear regression applied to effort estimation. In Proc. 16th International Conference on Software Engineering & Knowledge Engineering, 2004, 57-61.

Improvement of mathematical model for estimating the size of Java based Android multimedia applications

Annotation. During the research, a linear regression model was developed to estimate the size of Java based multimedia Android applications.. The resulting model, in comparison with other models, has has a larger multiple coefficient of determination, a smaller value of the mean magnitude of relative error, a larger value of percentage of prediction.

Keywords: software size estimation, Java Android applications, linear regression model

Рябков С.И., Приходько С.Б.

Усовершенствование математической модели для оценки размера мультимедийных Android приложений, написанных на языке Java

Аннотация. В ходе исследования была разработана линейная регрессионная модель для оценки размера мультимедийных Android приложений, написанных на языке Java. Полученная модель, по сравнению с другими моделями, имеет больший коэффициент множественной детерминации, меньшее значение величины средней ошибки и большую величину процента прогнозирования.

Ключевые слова: прогнозирование размера программного обеспечения, Java Android приложения, линейная регрессионная модель.