

УДОСКОНАЛЕННЯ МАТЕМАТИЧНОЇ МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ КОМП'ЮТЕРНИХ ІГОР, ЩО НАПИСАНІ МОВОЮ JAVA

Дробишев В. В., ст. гр. 6151м, Приходько С. Б., д. т. н.

Національний університет кораблебудування ім. адм. Макарова

Україна, Миколаїв

141245@nuos.edu.ua, sergiy.prykhodko@nuos.edu.ua

Анотація. Під час дослідження було розроблено лінійну регресійну модель для оцінки розміру комп'ютерних ігор, 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: (оцінка розміру програмного забезпечення, модель лінійної регресії, комп'ютерні ігри, що написані мовою Java).

Вступна частина. Проблема оцінювання розміру програмного забезпечення (ПЗ) на ранніх етапах розробки є важливою[1], 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 [2, 3]. Це призводить до необхідності розробки відповідних моделей для оцінки розміру програмного забезпечення. 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. Для того щоби оцінити програмний проект, для цієї моделі потрібні дані минулих проектів. На даний час існують рішення для оцінки інформаційних систем, написаних мовою Java [3]. У даній роботі лінійної регресії на основі метрик концептуальної моделі даних у вигляді діаграми класів для оцінювання інформаційних систем, 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 у вигляді:

$$Y = f(\beta, X) + \varepsilon$$

де $X = (X_1, X_2, \dots, X_m)$ - вектор незалежних змінних; β - вектор параметрів (що підлягають визначенню); ε - випадкова похибка (відхилення); Y - залежна (яка пояснюється) змінна.

Для побудови моделі лінійної множинної регресії оцінки розміру комп'ютерних ігор, що написані мовою Java, було зібрано дані з 31 програмних проектів, дані представлені у Таблиці 1, Де Y – кількість строк коду, X_1 – кількість абстрактних методів, X_2 – кількість публічних методів, X_3 – кількість анонімних класів, X_4 – кількість підкласів.

Для оцінки параметрів рівняння множинної регресії скористаємося методом найменших квадратів [4]. 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 множення матриць отримаємо наступні коефіцієнти: $b_0 = 0$, $b_1 = 25.0035$, $b_2 = 44.0022$, $b_3 = 25.964$, $b_4 = 70.2348$. Рівняння регресії приймає вигляд:

$$Y = 0 + 25.0035X_1 + 44.0022X_2 + 25.963X_3 + 70.2348X_4$$

Таблиця 1 – Дані по проектам

№	Name	Y	X1	X2	X3	X4	Y(x)	$\varepsilon = Y - Y(x)$	MRE
1	15 Days	26891	96	332	272	56	28004,42	-1113,42342	0,041405
2	868-HACK	3958	8	62	19	15	4475,002	-517,002375	0,130622
3	911 Operator	25130	131	265	118	63	22424,59	2705,411941	0,107657
4	1000 Days of Syria	14231	59	119	50	42	10959,53	3271,469026	0,229883
5	Kaimkatbarnemshen	13648	51	135	71	36	11587,37	2060,627028	0,150984
6	Pavakvakenzyukigrakk	2189	9	12	44	8	2457,352	-268,352317	0,122591
7	Kazelasum	6292	20	37	123	16	6445,48	-153,480066	0,024393
8	Tona	21036	94	88	68	201	22105,27	-1069,27237	0,050831
9	Sookavyukcamob	5712	25	44	89	21	6487,381	-775,380877	0,135746
10	Lanabpektyurkadpakoz	12256	80	118	73	30	11194,96	1061,042978	0,086573
11	Noavnuzaleskanlomrem	24205	74	93	326	98	21289,74	2915,26146	0,12044
12	Syuanmevschakagkep	4754	78	32	52	17	5902,465	-1148,46471	0,241579
13	Afterfall: Reconquest	640	8	4	16	1	861,6957	-221,695716	0,3464
14	The Age of Decadence	6296	21	17	54	24	4360,802	1935,197547	0,307369
15	AGS Awards	25564	96	332	270	56	27952,5	-2388,49543	0,093432
16	War: Fleet Command	9513	42	47	92	59	9650,792	-137,792406	0,014485
17	Akane the Kunoichi	3166	40	28	40	21	4745,693	-1579,69324	0,498956
18	Albert & Otto	11945	11	175	163	59	16351,41	-4406,4079	0,368891
19	Passage	17452	217	144	68	83	19357,12	-1905,12185	0,109164
20	Cogmind	18739	164	185	77	64	18735,24	3,760303254	0,000201
21	It Just Ain't Your Story	10229	38	56	9	62	8002,491	2226,509102	0,217666
22	Bernband	18696	219	149	72	87	20011,94	-1315,93509	0,070386
23	Stronghold	7384	22	61	87	21	6968,01	415,9899637	0,056337
24	Beat Cop	2493	5	26	24	17	3086,202	-593,202319	0,237947
25	Aidos	2840	76	27	14	6	3873,232	-1033,232	0,363814
26	BatBat-Game	19375	86	200	36	80	17504,23	1870,769235	0,096556
27	Blackjack	31297	121	386	198	58	29224,76	2072,235863	0,066212
28	Cannon	17119	44	158	160	74	17404,12	-285,117008	0,016655
29	Catch_The_Drop	13215	48	106	161	38	12713,53	501,4721281	0,037947
30	Coldsquare	3800	10	24	10	13	2478,78	1321,219603	0,347689
31	Duelers	13336	29	230	32	52	15328,67	-1992,66516	0,14942
31	geobattle	3450	11	18	48	28	4279,925	-829,924651	0,240558

Необхідно провести статистичний аналіз отриманого рівняння регресії: пошуку стандартної похибки та дисперсії параметрів моделі, перевірці значимості коефіцієнтів, коефіцієнту детермінації (R^2), значення величини середньої похибки (MRE) та величини відсотка прогнозування ($PRED(0.25)$) [5, 6].

Знайдемо незміщену оцінку дисперсії:

$$S^2 = \frac{1}{n - m} * S_e^2 = 3045777.482$$

Отримавши незміщену оцінку дисперсії можемо знайти значення стандартної похибки:

$$S = \sqrt{S^2} = 1745.215$$

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

380611.62	-1009.364	-212.226	-973.275	-1979.565
-1009.363	58.087	-15.64	7.736	-33.491
-212.204	-15.636	21.663	-13.952	-0.582
-973.274	7.721	-13.935	28.151	-10.295
-1979.561	-33.471	-0.56	-10.283	108.631

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

$$S_{b1} = \sqrt{58.087} = 7.621, S_{b2} = \sqrt{21.663} = 4.654$$

$$S_{b3} = \sqrt{28.151} = 5.306, S_{b4} = \sqrt{108.631} = 10.423$$

Для перевірки значимості параметрів множинної регресії використовують t-статистику (або t-критерій Стьюдента) [7]. Критичне значення для даної моделі має значення:

$$T_{\text{крит}} = (27; 0025) = 2.052$$

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

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

$$t_1 = \frac{25.004}{7.621} = 3.281 > 2.052$$

$$t_2 = \frac{44.002}{4.654} = 9.454 > 2.052$$

$$t_3 = \frac{25.964}{5.306} = 4.894 > 2.052$$

$$t_4 = \frac{70.235}{10.423} = 6.739 > 2.052$$

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

Тісноту впливу факторів на результат оцінює індекс множинної кореляції [8]. 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:

$$R = \sqrt{1 - \frac{S_e^2}{\sum Y_i^2}} = 0.993402$$

Отримавши коефіцієнт множинної кореляції знайдемо коефіцієнт детермінації:

$$R^2 = 0.986847$$

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 [9]. 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 $PRED(0.25)$:

$$MMRE = \frac{\sum MRE_i}{i} = 0.15$$

$$PRED(0.25) = 0.82$$

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

Висновки. В результаті розрахунків було отримано рівняння множинної регресії: $Y = 0 + 25.0035X_1 + 44.0022X_2 + 25.964X_3 + 70.2348X_4$. Побудовано лінійну регресійну модель з наступними характеристиками $R^2 = 0.986847$, $MMRE = 0.15$, $PRED(0.25) = 0.82$. 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 комп'ютерних ігор, що написані мовою Java.

Статистична значимість рівняння перевірена за допомогою коефіцієнта детермінації і критерію Фішера. Встановлено, що в досліджуваній ситуації 95.82% загальної варіабельності Y пояснюється зміною факторів X_j . Встановлено також, що параметри моделі статистично значущі.

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, Barry W. (2000). *Software Cost Estimation With Cocomo II* – Prentice Hal, Hardcover,. –(544 p). – ISBN-10: 0137025769.

- [3] Приходько, С.Б. (2014). *Построение нелинейной регрессионной модели времени восстановления работоспособности устройств терминальной сети* [Текст] / С.Б. Приходько, Л.М. Макарова // *Проблемы информационных технологий*. №1 (015).-С97-102.-ISSN 1998-7005.
- [4] Prykhodko, S.B. (2017). *Constructing the non-linear regression equations on the basis of bivariate normalizing transformations* [Text] / S.B.Prykhodko, X.V. Prykhodko, L.M. Makarova, O.O. Kudin, T.G.Smykodub // *Вісник Херсонського національного технічного університету*. №3 (62). – С333-372. – Т.І. – ISSN 2078-4481
- [5] Приходько, С.Б. (2005). *Методичні вказівки до виконання лабораторних робіт з дисципліни “Обробка експериментальних даних на ЕОМ”* [Текст] / С.Б. Приходько. – Миколаїв:НУК.
- [6] Krishnaswamy K. N., Sivakumar A. I. and Mathirajan M. (2006). *Management Research Methodology: Integration of Principles, Methods and Techniques*. Dorling Kindersley (India) Pvt. Ltd.
- [7] Kvanli A. H., Pavur R. J. and Keeling K. B. (2006). *Concise Managerial Statistics*. Thomson Learning Inc.
- [8] Freund R. J. and Wilson W. J. (1998). *Regression Analysis: Statistical Modeling of a Response Variable*. Academic Press.
- [9] Ott R. L. and Longnecker M. (2010). *An Introduction to Statistical Methods and Data Analysis*. Cengage Learning Inc.

Drobyshev V. V., Prykhodko S. B.

Improving the mathematical model for estimating the size of Java-based computer games and developing a software for its implementation.

Annotation. During the research, a linear regression model was developed to estimate the size of Java based computer games. The resulting model, in comparison with other models, 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, linear regression model, Java-based computer games).

Дробышев В. В., Приходько С. Б.

Усовершенствование математической модели для оценивания размера компьютерных игр, написанных на языке Java и разработка программы для её реализации.

Аннотация. В ходе исследования была разработана линейная регрессионная модель для оценки размера компьютерных игр, 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: (оценка размера программного обеспечения, модель линейной регрессии, компьютерные игры, написанные на языке Java).