УДК 004.415.2:519.237.5

МАТЕМАТИЧНА МОДЕЛЬ ОЦІНЮВАННЯ СКЛАДНОСТІ ОБ'ЄКТНО-ОРІЄНТОВАНОГО ПРОЄКТУВАННЯ .NET-ЗАСТОСУНКІВ ЗА СЕМАНТИКОЮ КЛАСІВ

Руденко К.Р.[1]; Латанська Л.О., к. ф.-м. н., доц.[2]
[1, 2] Національний університет кораблебудування імені адмірала Макарова
[1, 2] Україна, Миколаїв
[1] dreamtroperx@gmail.com; [2] liudmyla.latanska@nuos.edu.ua

Анотація. *Запропоновано математичну модель та алгоритм оцінювання складності об'єктно-орієнтованого проєктування .NET-застосунків за семантикою класів (WMC, RFC, LCOM), що поєднує перевірку нормальності за критеріями Мардія, логарифмічну нормалізацію і Z-Score стандартизацію, відсікання викидів, множинну лінійну регресію та кластеризацію K-Means для реалізації мовою C#.*

Ключові слова: *об'єктно-орієнтоване проєктування; семантика класів; нормалізація метрик; еліпсоїд прогнозування; множинна лінійна регресія; кластеризація K-Means.*

Вступна частина. Оцінювання складності об'єктно-орієнтованого проєктування (ООП) .NET-застосунків мовою C# є актуальною задачею забезпечення якості ПЗ. Семантику класів описують метрики Ш. Чидамбера та К. Кемерера [1]: WMC, RFC і LCOM. Через негаусовий розподіл метрик і їх кореляцію [2-6] порогові оцінки є некоректними.

Метою дослідження є розроблення математичної моделі й алгоритму оцінювання складності ООП .NET-застосунків за семантикою класів на основі нормалізації, стандартизації, еліпсоїда прогнозування, множинної регресії [2-5] та K-Means [7].

Основна частина. У вибірці з N .NET-застосунків i-й проєкт ($i = 1,\dots,N$) описується вектором середніх метрик WMC, RFC, LCOM $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3})^T$. На першому кроці при $\alpha = 0{,}005$ перевіряється багатовимірна нормальність за критеріями Мардія $\beta_{1,k}$ і $\beta_{2,k}$ [5, 6] при $k = 3$ ($\bar{\mathbf{X}}$, $\mathbf{S}_N$ – середнє і коваріаційна матриця):

$$\beta_{1,k} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[(\mathbf{X}_i - \bar{\mathbf{X}})^T\mathbf{S}_N^{-1}(\mathbf{X}_j - \bar{\mathbf{X}})\right]^3,\ \beta_{2,k} = \frac{1}{N}\sum_{i=1}^{N}[(\mathbf{X}_i - \bar{\mathbf{X}})^T\mathbf{S}_N^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})]^2. \quad (1)$$

За відхилення розподілу від Гаусового закону [5, 6] на другому кроці виконується нормалізація метрик шляхом натурального логарифмування (перетворення Бокса-Кокса при $\lambda_j = 0$ [2, 5]) та їх просторова Z-Score стандартизація до єдиної безрозмірної шкали:

$$Y_{ij} = \ln(X_{ij}), \qquad Z_{ij} = \frac{Y_{ij} - \bar{Y}_j}{\sigma_j}, \qquad \bar{Y}_j = \frac{1}{N}\sum_{i=1}^{N} Y_{ij}, \qquad \sigma_j = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Y_{ij} - \bar{Y}_j)^2}.$$

На третьому кроці у нормалізованих, стандартизованих даних $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})^T$ виявляються викиди за квадратом відстані Махаланобіса d_i^2 та еліпсоїдом прогнозування [4, 5]:

$$d_i^2 = (\mathbf{Z}_i - \bar{\mathbf{Z}})^T\mathbf{S}_N^{-1}(\mathbf{Z}_i - \bar{\mathbf{Z}}) \le \chi^2_{k,\alpha},\ \ \mathbf{S}_N = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T.$$

Точки з $d_i^2 > \chi^2_{k,\alpha}$ (квантиль χ^2 при $k = 3$, $\alpha = 0{,}005$) відсікаються як викиди з повторною перевіркою за критеріями Мардія (1) [5].

На четвертому (основному) кроці будується система трьох рівнянь множинної лінійної регресії, використовуючи метод найменших квадратів, у якій кожна нормалізована,

стандартизована метрика Z_Y (WMC, RFC, LCOM) прогнозується за двома іншими предикторами Z_1, Z_2 [3, 5]:

$$\hat{Z}_{Y_i} = \hat{b}_0 + \hat{b}_1 Z_{1i} + \hat{b}_2 Z_{2i}, \quad (2)$$

де $\hat{b}_0, \hat{b}_1, \hat{b}_2$ – оцінки параметрів регресії, точність якої оцінюється за R^2, MMRE і PRED(0,25) [5]. За системою (2) обчислюються оцінки $\hat{Z}_{WMC}, \hat{Z}_{RFC}, \hat{Z}_{LCOM}$: якщо метрики проєкту не перевищують оцінок (2), складність ООП є невисокою (оптимальною), а за перевищення 1, 2 або 3 рівнянь визначається локальне ускладнення, структурна диспропорція чи висока складність ООП [5]. На п'ятому кроці виконується PCA-кластеризація 3D K-Means [7] для зіставлення регресійних оцінок із кластерами масштабу.

Висновки. Запропоновано математичну модель і алгоритм оцінювання складності ООП .NET-застосунків за метриками WMC, RFC, LCOM на основі критеріїв Мардія, логарифмічної нормалізації, Z-Score стандартизації, відсікання викидів, множинної лінійної регресії та K-Means для реалізації на C# (Roslyn API).

Використання інструментів ШІ. Модель Gemini 4 Argon застосовувалася мінімально для редагування окремих речень і виправлення дрібних багів; модель, алгоритм і висновки розроблено авторами самостійно.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Chidamber, S. R., & Kemerer, C. F. (1994). A metrics suite for object oriented design. *IEEE Transactions on Software Engineering, 20*(6), 476-493. https://doi.org/10.1109/32.295895

[2] Prykhodko, S. B., & Prykhodko, N. V. (2020). Mathematical modeling of non-Gaussian dependent random variables by nonlinear regression models based on the multivariate normalizing transformations. *Mathematical Modeling and Simulation of Systems, 8*(3), 166-174. https://doi.org/10.1007/978-3-030-58124-4_16

[3] Prykhodko, S. (2024). Evaluating quality of software systems by the confidence and prediction intervals of regressions for RFC, CBO and WMC metrics. *WSEAS Transactions on Systems, 23,* 322-330. https://doi.org/10.37394/23202.2024.23.36

[4] Prykhodko, S., Makarova, L., Latanska, L., & Bryzghalov, M. (2025). Mathematical model for detecting outliers in the two-dimensional data of software metrics RFC and CBO from applications in Java. *CEUR Workshop Proceedings, 4048*, 509-519.

[5] Prykhodko, A. S. (2026). *Matematychni modeli dlia obrobky informatsii z metryk Web zastosunkiv, shcho rozrobliaiutsia z vykorystanniam PHP-freimvorkiv* (PhD dissertation). Odesa I. I. Mechnikov National University, Odesa. (in Ukrainian).

[6] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519-530. https://doi.org/10.1093/biomet/57.3.519

[7] Ding, C., & He, X. (2004). K-means clustering via principal component analysis. In *Proceedings of the 21st International Conference on Machine Learning (ICML 2004)* (pp. 29-36). https://doi.org/10.1145/1015330.1015408

Rudenko K.R., Latanska L.O.

Mathematical model for assessing object-oriented design complexity of .NET applications by class semantics

Abstract. *A mathematical model and algorithm are proposed for assessing object-oriented design complexity of .NET applications by class semantics (WMC, RFC, LCOM), combining Mardia's normality test, logarithmic normalization and Z-Score standardization, outlier filtering, multiple linear regression, and K-Means clustering for implementation in C#.*

Keywords: *object-oriented design; class semantics; metric normalization; prediction ellipsoid; multiple linear regression; K-Means clustering.*