

УДК 004.4:006.91

**ВИКОРИСТАННЯ МЕТРОЛОГІЧНИХ ПІДХОДІВ ДЛЯ
ВЕРИФІКАЦІЇ ЕКСПЕРТНИХ ОЦІНОК ТРУДОМІСТКОСТІ
РОЗРОБКИ ПЗ**

Заболотний Є.О.

e-mail: yevhen.zabolotnyi@nure.ua

Науковий керівник – к.т.н., доц. Запорожець О. В.

Харківський національний університет радіоелектроніки, каф. ІВТ
м. Харків, Україна

This study presents a metrological framework for verifying expert software development effort estimates to reduce subjectivity in project planning. By treating requirement decomposition (feature lists) as an indirect measurement of software complexity, the authors implement an ensemble of hybrid neural networks acting as parallel measurement channels. Experimental validation showed an MMRE of 77.67% and RMSPE of 92.78%. Calculation of Type A standard uncertainty (56.11%) based on ensemble dispersion enables the use of statistical control and outlier detection to identify metrologically unreliable estimates. This approach transitions effort estimation from intuitive guessing to a verifiable statistical process with predictable confidence intervals.

У сучасній практиці управління програмними проектами процес оцінювання трудомісткості розробки програмного забезпечення (ПЗ) часто базується на експертних методах, які характеризуються високим ступенем суб'єктивізму та залежністю від досвіду експерта [1]. З точки зору метрології, процедуру декомпозиції вимог у формі фічелістів (англ. feature list – структурований список задач) із подальшим присвоєнням часових еквівалентів (наприклад, в людино-години) можна розглядати як процес непрямого вимірювання латентного параметра, що відповідає величині складності програмного забезпечення. Такий підхід дозволяє застосувати до аналізу оцінок класичний метрологічний інструментарій, спрямований на визначення точності, збіжності та надійності отриманих результатів.

Запропонована модель верифікації базується на побудові ансамблю нейронних мереж, де кожен вузол навчається на різних підмножинах історичних даних (використовується стратегія випадкового виключення 20% даних для валідації кожного члена ансамблю). Подібні моделі часто розглядаються в дослідженнях і показують стабільні результати [2, 3]. Для обробки текстових описів у фічелістах застосовуються методи NLP (Natural Language Processing) виключно з метою семантичної верифікації та перетворення природної мови у векторне представлення. Це дозволяє моделі порівнювати нові задачі з еталонними значеннями з навчальної вибірки.

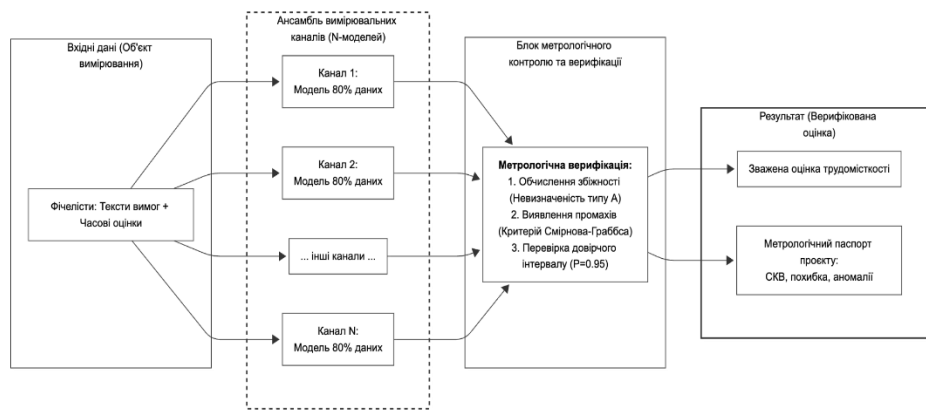


Рисунок 1 – Схема верифікації експертних оцінок

Використання ансамблю, де кожен вузол (вимірювальний канал) навчається на випадковій підмножині історичних даних, дозволяє отримати статистичний ряд прогнозів для кожної оцінюваної одиниці трудомісткості. При оцінці кожної задачі i -та модель дає свою відповідь, далі складається оцінка проєкта Y_i . Фінальний результат Y розраховується як:

$$Y = \frac{\sum w_i y_i}{\sum w_i}, \quad (1)$$

де вага w_i обернено пропорційна похибці моделі ($w_i = 1/\sigma_i^2$).

Метрологічний аспект дослідження полягає у впровадженні процедури статистичного контролю результатів оцінювання. Використання ансамблю моделей дозволяє розрахувати показник невизначеності за типом А, що базується на статистичному аналізі розбіжностей між прогнозами окремих моделей. Якщо середньоквадратичне відхилення всередині ансамблю перевищує встановлений поріг, результат вважається метрологічно недостовірним через недостатню збіжність вимірювальних каналів. Додатково пропонується застосування статистичних критеріїв виявлення аномалій для автоматичної ідентифікації промахів, що дозволяє нівелювати вплив аномальних відхилень усередині ансамблю.

Для реалізації моделі верифікації розроблено гібридну архітектуру штучної нейронної мережі (ANN), що поєднує аналіз семантики та метаданих. Текстовий блок базується на енкодері all-MiniLM-L6-v2 (розмірність вектора 384) із застосуванням маскованого усереднення (masked mean pooling), оскільки ця модель демонструє високу швидкість обробки [4]. Отриманий семантичний вектор об'єднується з вектором ознак метаданих (176 параметрів, включаючи складність стеку та коефіцієнти впевненості) у загальний вектор розмірністю 560 одиниць. Регресійний блок складається з повнозв'язного шару (128 нейронів) з активацією ReLU та Dropout (0.1), що завершується 6 паралельними вихідними вузлами для прогнозування трудомісткості за категоріями (BE, FE, Mobile, Data, Design, DevOps) із невід'ємним обмеженням результатів.

Експериментальне дослідження проведено на вибірці з 46 фічелістів. Для зовнішньої валідації ансамблю було зарезервовано 9 об'єктів, а 37 використано для формування 10 вимірювальних каналів (моделей ансамблю) за схемою випадкового розподілу (30 для навчання та 7 для внутрішньої перевірки збіжності). Оцінка точності прогнозів проводилася за допомогою відносних показників похибки, що зумовлено масштабно залежною варіативністю значень трудомісткості.

Для валідаційної вибірки значення середньої відносної похибки модуля (MMRE) склали 77.67%, а середньоквадратична відносна похибка – 92.78%. Середня стандартна невизначеність за типом А (u_A) склала 56.11%, і була розрахована як розкид прогнозів усередині ансамблю відносно дійсного значення:

$$u_A = \frac{1}{n} \sum \sqrt{\frac{1}{m(m-1)} \sum (y_{ij} - \bar{y}_i)^2}, \quad (2)$$

де n – це кількість фічелістів; m – кількість вузлів ансамблю; y_{ij} – прогноз окремого i -того вузла для фічеліста j ; а $\bar{y}_i$ – це їх середнє значення.

Даний підхід може бути використаний як для покращення моделі оцінювання, так і для верифікації експертних оцінок, для яких мікроеваріації усередині ансамблю можуть бути використані як індикатор метрологічної достовірності.

Цей підхід забезпечує об'єктивне підґрунтя для верифікації експертних оцінок, дозволяючи перейти від інтуїтивного підтвердження результатів до їх метрологічної перевірки. Розрахована сумарна похибка ансамблю виступає кількісним показником якості моделі, що дає змогу прогнозувати довірчі інтервали для оцінок розробки ПЗ. Даний підхід може бути інтегрований у процеси розробки ПЗ для мінімізації ризиків помилкового планування.

Список використаних джерел:

1. Jørgensen M. A Review of Studies on Expert Estimation of Software Development Effort // *Journal of Systems and Software*. 2004. Vol. 70, no. 1-2. P. 37–60.
2. Pai D. R., McFall K. S., Subramanian G. H. Software Effort Estimation Using Neural Network Ensemble // *Journal of Computer Information Systems*. 2013. Vol. 53, no. 4. P. 49–58. DOI: 10.1080/08874417.2013.11645650.
3. Idri A., Hosni M., Abran A. Systematic literature review of ensemble effort estimation // *Journal of Systems and Software*. 2016. Vol. 118. P. 151–175. DOI: 10.1016/j.jss.2016.05.016.
4. Pavlyshenko B., Stasiuk M. Semantic Similarity Analysis Using Transformer-Based Sentence Embeddings // *Electronics and Information Technologies*. 2025. Issue 30. P. 43–58. DOI: 10.30970/eli.30.4.