

РЕЦЕНЗІЯ

офіційного рецензента, кандидата технічних наук, доцента,
завідувачки кафедри прикладної математики та інформаційних технологій
Національного університету «Одеська політехніка»
Волкової Наталії Павлівни
на дисертаційну роботу здобувача Курінька Дмитра Дмитровича
«Моделі та методи машинного навчання для виявлення та усунення антипатернів
в програмних компонентах»,
подану до захисту в разову спеціалізовану вчену раду
Національного університету «Одеська політехніка»
Міністерства освіти і науки України
на здобуття наукового ступеня доктора філософії
за спеціальністю 122 – Комп'ютерні науки
галузь знань 12 – Інформаційні технології

1. Загальна характеристика та актуальність теми

Дисертаційна робота Курінька Д.Д. спрямована на вирішення актуальної проблеми сучасної програмної інженерії – накопичення технічного боргу внаслідок антипатернів проєктування в об'єктно-орієнтованих системах. Поширення методологій безперервної інтеграції та доставлення (CI/CD) суттєво загострює цю проблему, оскільки деградація якості коду безпосередньо впливає на стабільність збірок, збільшує час виведення нових версій та підвищує ризики регресій.

Наявні засоби статичного аналізу та виявлення «запахів» коду, незважаючи на їх широке промислове застосування, мають принципові обмеження: вони оперують локальними структурними сигналами і не беруть до уваги семантику, міжмодульні залежності та еволюційну динаміку репозиторію. Внаслідок цього значна частина рекомендацій виявляється хибною або нерелевантною у конкретному контексті проєкту, що знижує довіру розробників до автоматизованих інструментів і гальмує їх впровадження в практичні робочі процеси.

Запропонований у дисертації підхід, що базується на багаторівневому гібридному графовому поданні коду з інтеграцією структурних, семантичних та еволюційних ознак, відповідає сучасним тенденціям розвитку методів машинного навчання для аналізу програмних систем. Дисертаційне дослідження є актуальним і відповідає спеціальності 122 «Комп'ютерні науки».

2. Структура й зміст дисертації

Дисертація побудована у вигляді чотирьох взаємопов'язаних розділів, яким передуює розгорнутий вступ, а завершує роботу – узагальнення отриманих результатів, список використаних джерел та додатки.

У вступі чітко окреслено проблемне поле дослідження: обґрунтовано невідповідність між потребами промислової розробки та можливостями наявних

інструментів, сформульовано мету, об'єкт, предмет і завдання роботи, визначено наукову новизну та практичне значення одержаних результатів.

Перший розділ виконує роль аналітичного підґрунтя дослідження: у ньому систематизовано наявні класифікації антипатернів за структурними, семантичними та еволюційними проявами, критично розглянуто традиційні методи, засновані на метриках і правилах, а також сучасні підходи, такі як графові нейромережі та LLM-орієнтовані рішення, виявлено їхні принципові обмеження й обрано напрям подальшого дослідження.

У другому розділі розроблено багаторівневу модель виявлення антипатернів на базі Code Property Graph (CPG). Гетерогенний граф об'єднує синтаксичні дерева, графи потоку керування та залежностей і доповнюється 4-канальними ознаками вузлів. Ієрархія «метод – компонент – проєкт» із механізмом передачі повідомлень і ієрархічною увагою забезпечує виявлення як типових, так і нетипових антипатернів із явним моделюванням невизначеності (open-set/low-confidence режим).

Третій розділ розвиває результати попереднього й зосереджується на методах підтримки прийняття рішень щодо рефакторингу: багатоцільовий метод рекомендації усунення дублікатів коду з Парето-оптимізацією та оцінкою невизначеності; метод планування «мінімально інвазивних» послідовностей змін із причинно-наслідковим виведенням з історії репозиторію; модель статистичного контролю процесів (CUSUM/EWMA) для доказового оцінювання ефекту інтервенцій.

Четвертий розділ демонструє системну завершеність роботи: описано наскрізний інструментальний засіб, що об'єднує всі запропоновані компоненти в єдиний CI/CD-сумісний конвеєр, і наведено результати його валідації на семи великомасштабних Java-проєктах.

Висновки системно узагальнюють здобуті результати і засвідчують виконання всіх поставлених завдань.

3. Науковий рівень, новизна та теоретичне значення

Здобувач обґрунтовано виділяє низку результатів, які мають наукову новизну. Серед них:

- *вперше* запропоновано багаторівневу модель виявлення антипатернів на основі гібридного графа, що уніфікує синтаксичні дерева, графи потоку керування і графи залежностей у граф властивостей коду, з додатковими семантичними та еволюційними ознаками. Запропонована модель дозволяє виявляти типові і нетипові антипатерни, працювати з багатомовними програмними проєктами і забезпечувати сумісність із процесами безперервної інтеграції та доставки;

- *удосконалено* метод рекомендації рефакторингів через багатоцільову оптимізацію з оцінкою невизначеності, спеціалізований на усуненні дублікатів коду. Використання методу забезпечує безпечне, прозоре автоматизоване усунення дублікатів коду із прогнозованим впливом на процес;

- *вперше* запропоновано метод композиції «мінімально інвазивних» послідовностей рефакторингів на основі причинно-наслідкового виведення з

історії репозиторію. Використання методу дозволяє максимізувати очікувану користь за умов обмежень інтерфейсних змін і бюджету безперервної інтеграції та доставлення;

– *удосконалено* модель оцінювання ефективності рефакторингів на основі статистичного контролю процесів із атрибуцією впливу. Використання моделі забезпечує доказові рішення «покращення, погіршення, змін не виявлено» з заданим рівнем довіри, зменшує хибні тривоги та надає керований зворотний зв'язок для безперервної інтеграції та доставлення.

4. Практичне значення роботи

Прикладна значущість роботи підтверджується кількома взаємопов'язаними аспектами:

– запропонована система здатна виявляти антипатерни різних типів у багатомовних кодових базах промислового масштабу, що є практично нездійсненним для більшості наявних монолітних аналізаторів;

– механізм Парето-оптимізації з фільтром невизначеності суттєво зменшує кількість небезпечних або нерелевантних рекомендацій, що безпосередньо знижує ризик дестабілізації продуктивних збірок;

– SPC-оцінювання на основі CUSUM/EWMA надає командам розробників формалізований і статистично обґрунтований зворотний зв'язок щодо реального ефекту проведених змін, що особливо цінно в умовах тривалих або паралельних рефакторингових кампаній;

– результати дисертації впроваджено у виробничу діяльність науково-виробничого підприємства «КАРЕ», а також у навчальний і науково-дослідний процес Національного університету «Одеська політехніка».

Експериментальне підтвердження практичної ефективності отримано на семи широко відомих Java-проектах відкритого коду: середнє зниження індексу технічного боргу склало 16,4%, приріст індексу підтримуваності – +5,66, зниження когнітивної складності – 12,1%, зменшення щільності клонів – 32,0% (при узгодженості прогнозів $R^2 \approx 0,91$).

5. Повнота публікацій та апробація результатів

Здобувач оприлюднив результати дослідження через 11 наукових праць. До публікацій у виданнях міжнародних наукометричних баз належать: дві статті, що індексуються в Scopus (Herald of Advanced Information Technology), та три – у наукових фахових виданнях України з технічних наук. Матеріали дослідження було представлено на шести міжнародних наукових конференціях.

Тематична послідовність публікацій дозволяє простежити логіку формування наукового результату – від огляду і постановки задачі до розроблення конкретних моделей і методів та їх інтеграції в єдину систему. Обсяг і якісний рівень публікацій є достатніми для присудження наукового ступеня доктора філософії.

6. Зауваження та побажання

Наведені нижче зауваження не знижують загальної позитивної оцінки роботи, мають переважно дискусійний та рекомендаційний характер:

1) у роботі стисло висвітлено питання отримання еталонних міток антипатернів – через ручну експертну анотацію, автоматичні детектори чи їх поєднання. Коротке роз'яснення цієї процедури та досягнутого рівня узгодженості між експертами було б корисним для повнішого розуміння умов відтворюваності результатів;

2) у другому розділі вузлам CPG призначаються семантичні ознаки, проте вибір конкретної попередньо навченої моделі (CodeBERT, GraphCodeBERT тощо) окремо не обґрунтовано. Надання короткого коментаря щодо чутливості результатів до заміни ембеддингу дозволив би краще оцінити стійкість запропонованого підходу;

3) у реальній розробці нерідко трапляються ситуації, коли декілька рефакторингів виконуються майже одночасно. Стисле обговорення стратегії поведінки SPC-моделі в таких сценаріях із паралельними інтервенціями доповнило б практичне уявлення про застосування методу;

4) у кодових базах різні типи антипатернів трапляються з різною частотою, і деякі з них є об'єктивно рідкісними. Пояснення того, чи враховувався цей дисбаланс під час навчання, наприклад шляхом застосування зважених функцій втрат або повторної вибірки, надало б додаткову впевненість у коректності отриманих метрик для малочисельних класів;

5) поза межами стандартних відкритих проєктів залишається відкритим питання поведінки системи на автогенерованому або надмірно шаблонному коді, де структура CPG може помітно відрізнятися від навчальних зразків. Це питання доцільно розглянути як перспективний напрям подальших досліджень;

6) оскільки модель навчається на фіксованому корпусі проєктів, залишається відкритим питання щодо стабільності її характеристик у разі зміни мовних стандартів і стилів розробки. Стисле обговорення можливих стратегій адаптації або донавчання в умовах змінного розподілу даних збагатило б практичну частину роботи;

7) у зведених результатах наведено переважно агреговані показники за корпусом проєктів без деталізації міжпроєктного розкиду значень. Наведення такої деталізації або її коментування сприяло б кращому розумінню того, в яких конкретних умовах ефект запропонованих методів є найбільш вираженим.

Зазначені зауваження не впливають на загальний висновок щодо наукового рівня дисертації та можуть бути враховані автором у подальшій роботі.

7. Загальний висновок про відповідність роботи встановленим вимогам

Дисертаційна робота Курінька Дмитра Дмитровича на тему «Моделі та методи машинного навчання для виявлення та усунення антипатернів в програмних компонентах» є завершеним самостійним науковим дослідженням, яке вирішує важливу прикладну задачу автоматизованого виявлення та безпечного усунення антипатернів у програмних компонентах. Запропоновані моделі та методи вирізняються комплексністю підходу: поєднання структурного,

семантичного та еволюційного аналізу коду, явного моделювання невизначеності й статистично обґрунтованого оцінювання ефекту змін формує цілісне методологічне рішення, якого бракує в наявних інструментах.

Враховуючи викладене, вважаю, що дисертаційна робота Курінька Дмитра Дмитровича на тему «Моделі та методи машинного навчання для виявлення та усунення антипатернів в програмних компонентах» відповідає вимогам Порядку підготовки здобувачів вищої освіти ступеня доктора філософії та доктора наук у ЗВО (наукових установах), затвердженого Постановою Кабінету Міністрів України від 23 березня 2016 року № 261 (зі змінами) та вимогам п.п. 5 – 9 Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії, затвердженого постановою Кабінету Міністрів України від 12.01.2022 №44 (зі змінами).

Автор дисертаційної роботи – Курінько Дмитро Дмитрович – заслуговує на присудження наукового ступеня доктора філософії за спеціальністю 122 – Комп'ютерні науки, галузь знань 12 – Інформаційні технології.

Офіційний рецензент:

Завідувачка кафедри прикладної
математики та інформаційних
технологій Національного
університету «Одеська
Політехніка», кандидат технічних
наук, доцент

Наталія ВОЛКОВА