

ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

доктора технічних наук, професора, професора кафедри кібербезпеки та інтелектуальних інформаційних технологій факультету радіоелектроніки, комп'ютерних систем та інфокомунікацій Національного аерокосмічного університету «Харківський авіаційний інститут»

Морозової Ольги Ігорівни

на дисертаційну роботу **Курінька Дмитра Дмитровича**

на тему: **«МОДЕЛІ ТА МЕТОДИ МАШИННОГО НАВЧАННЯ ДЛЯ
ВИЯВЛЕННЯ ТА УСУНЕННЯ АНТИПАТЕРНІВ В ПРОГРАМНИХ
КОМПОНЕНТАХ»**,

представлену на здобуття наукового ступеня доктора філософії за спеціальністю
122 – Комп'ютерні науки, галузь знань 12 – Інформаційні технології

1. Актуальність теми дисертації

Інтелектуальний аналіз програмного коду є одним із динамічно зростаючих напрямів комп'ютерних наук, що перебуває на перетині теорії графів, методів машинного навчання та формальних моделей подання структурованих даних. Зростання обсягів і складності програмних систем ставить перед дослідниками принципово нові задачі автоматичного розпізнавання патернів у структурованих і напівструктурованих даних, зокрема, у вихідному коді програм.

Антипатерни проєктування являють собою клас стійких структурних аномалій у програмному коді, виявлення яких є нетривіальною задачею класифікації на графових даних.

Складність полягає в тому, що антипатерни не зводяться до одного типу сигналів: їхні ознаки мають структурну, семантичну та еволюційну природу, проявляються на різних рівнях абстракції, від окремого методу до архітектурного компонента, і не завжди піддаються формальному опису через порогові значення метрик або детерміновані правила. Існуючі моделі, як правило, обробляють лише

один тип ознак, що призводить до суттєвих втрат у точності розпізнавання та неможливості узагальнення на нові типи аномалій.

Із розвитком графових нейронних мереж, методів машинного навчання на структурованих даних та підходів до оцінювання невизначеності у задачах класифікації з'явилися теоретичні передумови для побудови якісно нових моделей аналізу коду. Однак синтез цих підходів у єдиній обчислювальній архітектурі, здатній інтегрувати різнорідні ознаки, підтримувати багаторівневий контекст і забезпечувати калібровані прогнози в умовах неповноти даних, досі залишається відкритою науковою задачею.

Актуальність дисертації визначається необхідністю розроблення теоретично обґрунтованих моделей і методів, що дозволяють розв'язати задачу виявлення антипатернів як задачу мультимодальної класифікації на гетерогенних графових структурах із явним урахуванням невизначеності. Вирішення цієї задачі розширює методологічний апарат комп'ютерних наук у частині аналізу складних структурованих даних і сприяє розвитку досліджень у галузі інтелектуального аналізу програмного коду.

2. Ступінь обґрунтованості наукових положень, висновків та рекомендацій

У роботі послідовно розглянуто взаємопов'язані задачі – автоматичне виявлення антипатернів у програмних компонентах, рекомендацію та планування рефакторингових перетворень, а також оцінювання їх фактичної ефективності в умовах безперервної розробки. Для вирішення цих задач здобувач застосовує:

- багаторівневу модель виявлення антипатернів на основі гібридного графа властивостей коду, що об'єднує синтаксичні дерева, графи потоку керування та залежностей з 4-канальними ознаками вузлів (структурні, семантичні ембеддинги, метричні, еволюційні);

- метод рекомендації рефакторингів через багатоцільову оптимізацію з явним моделюванням невизначеності та механізмом відмови від рекомендацій у low-confidence режимі;

- метод композиції «мінімально інвазивних» послідовностей рефакторингів на основі причинно-наслідкового аналізу репозиторію;

- модель оцінювання ефективності рефакторингів на базі статистичного контролю процесів з атрибуцією впливу та керованим рівнем хибних тривог.

Здобувач здійснює постановку задач на формальному рівні, вводить необхідні позначення, описує алгоритмічні структури та обґрунтовує вибір архітектурних рішень.

Висновки підкріплені результатами експериментів на реальних Java-проектах з відкритим кодом з порівнянням із базовими підходами та абляційними варіантами, що підтверджує належний ступінь обґрунтованості наукових положень і рекомендацій.

3. Достовірність отриманих результатів

Достовірність результатів забезпечується:

- використанням реальних відкритих програмних проєктів у фіксованих релізних зрізах з контрольованим середовищем;

- оцінюванням якості за усталеними метриками – Accuracy, F1-macro, NDCG@5, U@5 для задачі рекомендації; Technical Debt Index (TDI), Maintainability Index (MI), Cognitive Complexity (CC), Clone density – для наскрізної оцінки; ймовірність детекції, частка хибних тривог, час до сигналу – для SPC-моделі;

- наявністю порівняння із базовими й сучасними підходами, а також проведенням абляційного аналізу для виявлення внеску окремих компонентів; узгодженістю прогнозованого та фактичного ефектів ($R^2 \approx 0,91$ при розбіжностях у межах $\pm 1,6\%$), що свідчить про коректність запропонованих рішень;

- наявністю впроваджень та актів про використання результатів досліджень у практичній діяльності підприємств і в навчальному процесі.

У сукупності це дозволяє вважати отримані результати достовірними.

4. Наукова новизна

У результаті проведених досліджень здобувачем отримано та теоретично обґрунтовано ряд наукових результатів, що вирізняються новизною і виносяться на захист. Серед них особливої уваги заслуговують такі:

- *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 доставлення;

- *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.

Перелічені результати є новими для спеціальності 122 «Комп'ютерні науки» та вносять вагомий внесок у розвиток методів інтелектуального аналізу структурованих програмних даних, розширюючи наявний науковий доробок у відповідній предметній галузі.

5. Практичне значення

Практичне значення роботи полягає в можливості використання розроблених моделей і методів:

- в інструментах статичного аналізу та підтримки рефакторингу промислових об'єктно-орієнтованих програмних систем, зокрема великомасштабних Java-проектів;
- у системах безперервної інтеграції та доставлення для автоматизованого контролю внутрішньої якості коду при виконанні pull request і між релізами;
- у підсистемах підтримки прийняття рішень спеціалістів з комп'ютерних наук, що надають пріоритизовані, безпечні та обґрунтовані рекомендації щодо рефакторингу;
- у навчальному процесі при підготовці фахівців з комп'ютерних наук та штучного інтелекту.

Наявність актів впровадження у діяльність підприємств та використання результатів у науково-освітньому середовищі підтверджує реальну прикладну цінність дисертаційної роботи.

6. Повнота викладу результатів у публікаціях та дотримання академічної доброчесності

Основні результати дисертаційної роботи висвітлені в 11 публікаціях, серед яких є 2 статті у фаховому виданні, що індексується міжнародною наукометричною базою Scopus, а також доповіді на 5 міжнародних науково-практичних конференціях.

Публікації відображають основні положення дисертаційної роботи; чітко простежується логічний зв'язок між опублікованими результатами та структурою дисертаційної роботи. Кожен пункт наукової новизни підкріплений щонайменше однією статтею у науковому виданні. Ознак порушення академічної доброчесності не виявлено.

7. Характеристика змісту дисертації

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел і додатків. Структура роботи є логічною та відповідає заявленій темі.

У *вступі* аргументовано актуальність, визначено мету, завдання, об'єкт і предмет дослідження, сформульовано наукову новизну та практичне значення роботи, наведено відомості про апробацію та публікації.

Перший розділ містить системний аналітичний огляд проблематики виявлення та усунення антипатернів у програмних компонентах. Розглянуто класичні підходи на основі метрик і правил, методи машинного навчання, а також сучасні графові та LLM-орієнтовані рішення. Окреслено обмеження існуючих підходів і обґрунтовано потребу в нових моделях, що слугує підґрунтям для постановки задач дослідження.

У *другому розділі* запропоновано гібридне графове подання коду у вигляді графа властивостей коду з чотириканальними ознаками вузлів та розроблено багаторівневу нейромережеву модель виявлення антипатернів. Описано математичний апарат, архітектуру моделі, механізм оцінювання невизначеності та конвеєр застосування в умовах безперервної інтеграції та доставлення.

Третій розділ присвячено методам рекомендації, планування та процесного оцінювання рефакторингів. Запропоновано метод багатоцільової оптимізації рефакторингів дублікатів коду з відбором Парето-оптимальних рішень та фільтром низької впевненості, метод композиції мінімально інвазивних послідовностей рефакторингів на основі причинно-наслідкового аналізу репозиторію, а також модель оцінювання ефективності рефакторингів на основі статистичного контролю процесів з атрибуцією впливу.

У *четвертому розділі* подано опис інтегрованого інструментального засобу для наскрізного керування антипатернами: від побудови гібридного подання коду до оцінювання ефекту рефакторингів. Наведено архітектуру засобу, структуру програмних компонентів та результати експериментального дослідження на семи

реальних програмних проєктах з відкритим кодом із порівнянням із базовими підходами.

У висновках узагальнено результати роботи, показано досягнення поставленої мети та виконання всіх задач дослідження.

Загалом, зміст дисертаційної роботи є послідовним, логічно побудованим і достатньо повно відображає проведені дослідження.

8. Зауваження та побажання

Наведені нижче зауваження мають характер зауважень і побажань та не знижують загальної позитивної оцінки роботи:

1) в першому розділі наведено широкий огляд літератури, однак класифікація підходів подекуди виглядає змішаною: в одному переліку сусідять застарілі метричні методи, класичні підходи машинного навчання та сучасні LLM-орієнтовані рішення без чіткої вказівки на загальну лінію еволюції. Було б доцільно жорсткіше структурувати огляд, наприклад, за ознаками «структурні / семантичні / еволюційні підходи» або «правила / машинне навчання / глибинне навчання», що дозволило б чіткіше позиціонувати власні рішення автора на тлі наявного стану справ;

2) хоча здобувач використовує широко відомі відкриті Java-проєкти, у тексті лише стисло вказуються їхні параметри. Для повнішого розуміння умов і відтворюваності експериментів доцільно було б подати у вигляді зведеної таблиці ключові характеристики кожного проєкту (кількість класів/методів, розмір кодової бази, час існування репозиторію, кількість розробників, розподіл антипатернів тощо);

3) у роботі показано загальний ефект від використання гібридного графового подання коду та ієрархічної графової нейронної мережі, однак бракує розгорнутої абляційної оцінки окремих компонентів моделі (4-канальні ознаки вузлів, еволюційні ознаки, механізм ієрархічної уваги). Наявність такої оцінки дозволила б точніше обґрунтувати архітектурні рішення та продемонструвати, за рахунок яких саме механізмів досягається перевага над базовими підходами;

4) при аналізі результатів значно більше уваги приділено середнім показникам якості. Разом з тим лише побіжно розглядаються ситуації, у яких якість знижується – наприклад, для проєктів з нестандартними угодами про кодування, рідкісними або нетиповими антипатернами, а також для дуже малих або великих кодових баз. Більш ґрунтовний аналіз типових помилок та причин їх виникнення був би корисним для практичного застосування запропонованих методів;

5) в описі інструментального засобу переважно розглядається його автономна робота. Для практичних впроваджень важливо розуміти, як саме пропонується pipeline інтегрується у конкретні CI/CD-платформи, які технічні вимоги ставляться до середовища виконання та як вирішуються питання масштабованості для великих монорепозіторіїв;

6) запропонована модель оцінювання ефективності рефакторинґів на основі статистичного контролю процесів є концептуально цінною, однак питання налаштування гіперпараметрів (довжина вікна базового режиму, порогові значення CUSUM/EWMA, мінімально необхідний час спостереження після інтервенції) розкрито недостатньо. Було б корисно навести практичні рекомендації щодо вибору цих параметрів або описати процедуру їх автоматичного налаштування;

7) у тексті подекуди зустрічаються неточності у нумерації формул і рисунків, не завжди послідовно оформлено посилання на джерела.

Наведені зауваження не впливають на коректність наукових результатів і обґрунтованість висновків дисертаційної роботи.

9. Загальний висновок

Дисертаційна робота Курінька Дмитра Дмитровича на тему «Моделі та методи машинного навчання для виявлення та усунення антипатернів в програмних компонентах» є завершеною науково-кваліфікаційною працею, в якій вирішено актуальну наукову задачу підвищення ефективності автоматизованого аналізу якості програмного коду. Зміст дисертаційної роботи відповідає спеціальності 122 –

«Комп'ютерні науки», галузі знань 12 – «Інформаційні технології», а сама робота – вимогам чинного Порядку присудження ступеня доктора філософії в Україні.

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 опонент:

доктор технічних наук, професор,
професор кафедри кібербезпеки та
інтелектуальних інформаційних
технологій факультету
радіоелектроніки, комп'ютерних систем
та інфокомунікацій Національного
аерокосмічного університету
«Харківський авіаційний інститут»

Ольга МОРОЗОВА

(підпис)

Підпис д.т.н., професора Ольги МОРОЗОВОЇ засвідчую

Учений секретар

Національного аерокосмічного
університету

«Харківський авіаційний інститут»

Тетяна БОНДАРЄВА

(підпис)