

ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

доктора технічних наук, професора, професора кафедри штучного інтелекту
навчально-наукового Інституту прикладного системного аналізу

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського».

Литвиненка Володимира Івановича

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

Попри значний прогрес у галузі статичного аналізу та інструментів рефакторингу, задача автоматичного виявлення антипатернів у промисловому програмному коді досі не має загальноприйнятого вирішення. Це пояснюється не браком уваги до проблеми, а її принциповою складністю: антипатерни є явищем мультимодальним – їхні прояви одночасно охоплюють синтаксичну структуру, семантику залежностей між компонентами та еволюційну динаміку кодової бази. Жоден із наявних підходів не інтегрує всі три виміри в єдиній моделі, що й зумовлює систематичні втрати в точності та узагальнювальній здатності існуючих рішень.

З наукової точки зору ця задача є нетривіальним випадком класифікації на гетерогенних графових структурах з неповними та зашумленими даними. Саме в цьому вимірі вона становить самостійний дослідницький інтерес для комп'ютерних наук – незалежно від прикладної мотивації. Розвиток методів навчання на графах, представлення коду у вигляді графів властивостей та підходів до оцінювання епістемічної невизначеності створив необхідні теоретичні

передумови для нового покоління моделей, проте їх синтез у цілісній архітектурі, придатній до застосування в реальних процесах розробки, залишається відкритою задачею.

Дисертаційна робота Курінька Дмитра Дмитровича вирішує саме цю задачу – і робить це комплексно, охоплюючи не лише виявлення антипатернів, а й рекомендацію рефакторингів та формалізоване оцінювання їх ефективності. Такий наскрізний підхід є рідкістю в існуючих дослідженнях і надає роботі додаткової наукової та практичної цінності.

Вважаю, що дисертаційна робота Курінька Дмитра Дмитровича є актуальною і своєчасною.

2 ЗВ'ЯЗОК РОБОТИ З НАУКОВИМИ ПРОГРАМАМИ, ПЛАНАМИ, ТЕМАМИ

Дисертацію виконано згідно тематичних планів НДР Національного університету «Одеська політехніка» за період 2024 – 2025 рр. в рамках держбюджетної теми № 237-177 – «Програмні системи машинного навчання та їх застосування в галузі інтелектуального аналізу даних».

3 ОЦІНКА ОБҐРУНТОВАНOSTІ НАУКОВИХ ПОЛОЖЕНЬ, ВИСНОВКІВ І РЕКОМЕНДАЦІЙ

При розв'язанні поставлених у дисертації задач здобувач використав широкий спектр сучасного математичного та алгоритмічного апарату: гетерогенні графові нейронні мережі з механізмом ієрархічної уваги, гібридне графове подання коду у вигляді графа властивостей, що об'єднує синтаксичні дерева, графи потоку керування та залежностей, методи багатоцільової оптимізації на основі принципу Парето, а також статистичний контроль процесів із використанням критеріїв кумулятивних сум та експоненційно зваженого ковзного середнього.

Обґрунтованість наукових положень підтверджується:

- чіткою постановкою задач і формальним описом розроблених моделей та методів;
- логічно побудованою системою ознак, що інтегрує структурні, семантичні, метричні та еволюційні характеристики коду в єдиному просторі;
- проведенням серії експериментів на реальних програмних проєктах з відкритим кодом у фіксованих релізних зрізах;
- порівнянням із базовими та сучасними підходами за загальноприйнятими кількісними метриками, а також проведенням абляційного аналізу окремих компонентів.

Висновки та рекомендації дисертації логічно впливають із проведеного теоретичного аналізу, постановки задач, запропонованих моделей і результатів експериментальних досліджень, а отже можуть вважатися достатньо обґрунтованими.

4 ДОСТОВІРНІСТЬ НАУКОВИХ ПОЛОЖЕНЬ, РЕКОМЕНДАЦІЙ ТА ВИСНОВКІВ

Підставою для позитивної оцінки достовірності результатів дисертації слугує насамперед методологічна коректність побудови експериментального дослідження. Здобувач використовує реальні програмні проєкти з відкритим кодом у фіксованих релізних зрізах із суворо контрольованим середовищем виконання, що виключає вплив зовнішніх чинників на отримані показники та забезпечує принципову відтворюваність результатів.

Достовірність додатково підкріплюється тим, що оцінювання якості проводилось за загальноприйнятими в науковій спільноті метриками для задач класифікації, ранжування та статистичного контролю процесів, а порівняння здійснювалось із актуальними підходами, що представляють сучасний стан досліджень у відповідних підзадачах. Покращення показників спостерігається стабільно на різних наборах даних, а не є артефактом окремо взятого

сприятливого випадку, що свідчить про узагальнювальну здатність запропонованих рішень.

Нарешті, факт впровадження результатів у діяльність НВП «Каре» та у навчальний та науковий-дослідницький процес Національного університету «Одеська політехніка», підтверджений відповідними актами, є незалежним свідченням практичної обґрунтованості отриманих висновків.

5 НАУКОВА НОВИЗНА РЕЗУЛЬТАТІВ, ВИКЛАДЕНИХ У ДИСЕРТАЦІЇ

Дисертаційна робота присвячена розв'язанню актуальної наукової задачі автоматизованого виявлення та усунення антипатернів у програмних компонентах шляхом розроблення моделей і методів машинного навчання, що інтегрують структурні, семантичні та еволюційні ознаки коду в єдиному гібридному графовому поданні з явним урахуванням невизначеності прогнозів.

За результатами проведених досліджень здобувачем отримано нові науково обґрунтовані результати, що не були предметом попереднього захисту і становлять самостійний внесок у розвиток методів інтелектуального аналізу структурованих програмних даних, а саме:

1) *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;

2) *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;

3) *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;

4) *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.

Отримані результати в сукупності є новими для спеціальності 122 «Комп'ютерні науки» та розширюють наявний науковий доробок у галузі інтелектуального аналізу структурованих програмних даних, формуючи цілісну методологічну основу для автоматизованого виявлення антипатернів, планування рефакторингів і доказового оцінювання їх ефективності в умовах реальних процесів розробки програмного забезпечення.

6 ПРАКТИЧНЕ ЗНАЧЕННЯ ОДЕРЖАНИХ РЕЗУЛЬТАТІВ

Практичне значення роботи полягає в тому, що розроблені моделі та методи підвищують точність автоматичного виявлення антипатернів у промислових програмних проєктах; забезпечують формування обґрунтованих, пріоритизованих і безпечних рекомендацій щодо рефакторингу безпосередньо у процесі безперервної розробки; дозволяють кількісно оцінювати фактичний ефект від проведених рефакторингових перетворень із контрольованим рівнем хибних сигналів.

Розроблений інструментальний засіб впроваджено у діяльність НВП «Каре» та використано в освітній і науково-дослідницькій діяльності Національного університету «Одеська політехніка», що підтверджено відповідними актами і свідчить про реальну прикладну цінність отриманих результатів.

7 ПОВНОТА ВИКЛАДУ РЕЗУЛЬТАТІВ У ПУБЛІКАЦІЯХ ТА ДОТРИМАННЯ АКАДЕМІЧНОЇ ДОБРОЧЕСНОСТІ

Основні результати дисертації опубліковано в 11 наукових працях, серед яких:

- 5 статей у фахових виданнях України з технічних наук;
- з них 2 статті у фаховому виданні, що індексується в наукометричній базі Scopus;
- 6 публікацій у матеріалах міжнародних наукових і науково-практичних конференцій.

Тематика та зміст зазначених праць корелюють із основними науковими положеннями дисертації, що забезпечує належний рівень апробації результатів.

Ознак порушення академічної доброчесності (плагіат, фальсифікація результатів тощо) в дисертаційній роботі не виявлено.

8 АНАЛІЗ ФОРМИ ТА ЗМІСТУ ДИСЕРТАЦІЙНОЇ РОБОТИ

Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел і додатків. Логіка побудови роботи відображає послідовний перехід від теоретичного обґрунтування до розроблення моделей, їх практичної реалізації та експериментальної верифікації.

Вступна частина окреслює наукову проблему, визначає мету, об'єкт і предмет дослідження, формулює наукову новизну та практичне значення отриманих результатів, а також містить відомості про апробацію роботи.

Перший розділ виконує функцію теоретичного фундаменту дослідження. На основі критичного аналізу наявних підходів – від традиційних метричних методів до сучасних рішень на базі глибинного навчання – здобувач обґрунтовує обмеження існуючих інструментів і формулює вимоги до нової моделі, що й визначає напрями подальшої роботи.

Другий і третій розділи становлять ядро наукового внеску дисертації. У другому розділі запропоновано концептуально новий підхід до подання програмного коду у вигляді гібридного графа властивостей та розроблено багаторівневу нейромережеву модель класифікації антипатернів із вбудованим механізмом оцінювання невизначеності. Третій розділ розвиває цей фундамент у напрямку прикладного застосування: представлено метод рекомендації рефакторингів через багатоцільову оптимізацію, метод побудови мінімально інвазивних послідовностей перетворень та модель статистичного контролю для оцінювання фактичного ефекту від проведених змін. Разом ці два розділи формують наскрізний методологічний ланцюжок – від виявлення проблеми до підтвердження результату її усунення.

Четвертий розділ забезпечує інтеграцію розроблених компонентів у єдиний інструментальний засіб та подає результати його комплексної експериментальної перевірки на реальних програмних проєктах. Отримані дані переконливо демонструють ефективність запропонованих рішень порівняно з базовими підходами.

Висновки стисло і точно підсумовують досягнуті результати та підтверджують повне виконання задач, сформульованих у вступі.

Загалом дисертація справляє враження цілісної та концептуально завершеної наукової праці, в якій кожен розділ органічно пов'язаний із загальною логікою дослідження.

9 ЗАУВАЖЕННЯ ДО ДИСЕРТАЦІЙНОЇ РОБОТИ

Наведені нижче зауваження не зменшують загальної позитивної оцінки дисертаційної роботи, проте, на думку опонента, заслуговують на увагу здобувача:

1) Експериментальне дослідження проводилось на проєктах, написаних мовами Java, Kotlin та Scala, що забезпечує достатньо репрезентативну основу для верифікації запропонованих моделей у межах екосистеми віртуальної машини

Java. Водночас у подальших дослідженнях було б цікаво розширити корпус за рахунок проєктів на мовах з принципово іншою парадигмою, зокрема, Python або C#, що дозволило б додатково підтвердити узагальнювальну здатність підходу за межами однієї мовної екосистеми;

2) У другому розділі описано механізм роботи в режимі низької впевненості, який передбачає відмову від рекомендації при перевищенні порогу невизначеності. Разом з тим процедура вибору конкретного значення цього порогу подана описово, без формального обґрунтування. Більш прозорий виклад, наприклад, через оптимізацію за кривою «точність-повнота» або фіксацію цільового рівня покриття, підвищив би відтворюваність запропонованого методу;

3) Базові підходи, з якими порівнюється запропонована система, охоплюють широкий спектр актуальних рішень і забезпечують коректний контекст для оцінювання отриманих результатів. У подальших дослідженнях було б доцільно доповнити це порівняння інструментами на основі великих мовних моделей, спеціалізованих для аналізу коду, що дозволило б ще повніше окреслити позиціонування запропонованого підходу в актуальному технологічному ландшафті;

4) У третьому розділі метод причинно-наслідкового виведення з історії репозиторію описано на концептуальному рівні, однак питання чутливості методу до якості та повноти історії комітів залишається нерозкритим. Зокрема, не розглянуто поведінку методу на проєктах із короткою або нерегулярною історією змін, де причинно-наслідкові зв'язки між рефакторингами можуть бути ненадійно ідентифіковані;

5) Запропонована модель передбачає інкрементальне оновлення графового подання коду, що є важливою перевагою для практичного застосування. Разом з тим у роботі відсутній аналіз накопичення помилок при тривалому інкрементальному оновленні без повної перебудови графа. Для повноти дослідження доцільно було б розглянути умови, за яких накопичені неточності стають критичними і потребують повного перерахунку;

6) У четвертому розділі end-to-end оцінка системи проводилась у фіксованих релізних зрізах, що є методологічно коректним, однак не відображає динаміки роботи системи в умовах реального безперервного розвитку проєкту. Моделювання поведінки системи в умовах потокової обробки нових комітів у реальному часі становить перспективний напрям для подальших досліджень;

7) Окремі формулювання у тексті дисертації потребували б уточнення з погляду термінологічної узгодженості, зокрема, паралельне вживання термінів «антипатерн» та «запах коду» у деяких підрозділах без чіткого розмежування цих понять може вносити неоднозначність у сприйняття наукових результатів.

Зазначені зауваження мають переважно дискусійний або редакційний характер і не впливають на наукову цінність дисертаційної роботи.

10 ЗАГАЛЬНІ ВИСНОВКИ ТА ОЦІНКА ДИСЕРТАЦІЇ

Дисертаційна робота Курінька Дмитра Дмитровича є завершеною кваліфікаційною науковою працею, в якій вирішено актуальну наукову задачу автоматизованого виявлення та усунення антипатернів у програмних компонентах шляхом розроблення нових моделей і методів машинного навчання на основі гібридного графового подання коду, багатоцільової оптимізації рефакторингів та статистичного контролю їх ефективності.

Робота відзначається актуальністю тематики, науковою новизною отриманих результатів, достатнім рівнем теоретичної обґрунтованості та переконливою експериментальною базою, а практична цінність підтверджена реальним впровадженням розроблених рішень.

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 штучного
інтелекту
навчально-наукового Інституту
прикладного системного аналізу
Національного технічного університету
України «Київський політехнічний
інститут імені Ігоря Сікорського

Володимир ЛИТВИНЕНКО