

РЕЦЕНЗІЯ

кандидата технічних наук

Немченка Вадима Вячеславовича

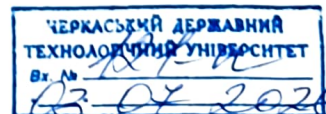
на дисертаційну роботу ОСТАПЮКА Володимира Вікторовича
«Агентні моделі в інформаційній технології інтелектуального моніторингу»,
подану на здобуття ступеня доктора філософії
з галузі знань 12 – «Інформаційні технології» за спеціальністю
121 – «Інженерія програмного забезпечення»

1. Актуальність теми дисертаційної роботи

Сучасний етап розвитку розподілених інформаційних систем, мікросервісних архітектур та об'єктів критичної інформаційної інфраструктури характеризується стрімким зростанням складності та динамічності середовищ їх функціонування. У процесі експлуатації таких систем неминуче змінюються статистичні властивості вхідних потоків даних та режими роботи об'єктів спостереження – явище, відоме як «дрейф концепцій» (concept drift). За цих умов традиційні статичні засоби моніторингу втрачають адекватність, що може призводити до пропуску критичних збоїв та зниження надійності систем загалом.

Окрему складність становить те, що в потоках телеметрії практично неможливо миттєво отримати істинні мітки (ground truth) для нових даних. Це робить класичні методи контролю помилок малоефективними та зумовлює потребу у механізмах виявлення дрейфу без учителя (unsupervised drift detection), здатних працювати в режимі реального часу. Тому особливої ваги набуває створення моніторингових програмних агентів, які спроможні автономно синтезувати й адаптувати власні прогностичні моделі без втручання людини.

Додаткову гостроту проблематиці надають сучасні виклики інженерії машинного навчання (MLOps): обмежений рівень автоматизації побудови складних ансамблів, переважно імперативний характер налаштування взаємодії алгоритмів та



обмеження наявних форматів серіалізації щодо опису багатокрокових конвеєрів. Зазначені чинники перешкоджають надійному перенесенню навчених моделей у середовище експлуатації із гарантуванням машинної відтворюваності обчислень. Актуальність теми посилюється й прикладним контекстом кризового моніторингу складних інформаційних процесів в умовах сьогодення.

З огляду на викладене, дисертаційна робота В.В. Остапюка, спрямована на вдосконалення методів багатoshарового синтезу агентних моделей та розроблення інструментів їх відтворюваного виконання на основі декларативного підходу й подання обчислень у вигляді спрямованого ациклічного графа (DAG), є своєчасною та актуальною як у науковому, так і в практичному вимірі.

2. Наукова новизна одержаних результатів

У дисертаційній роботі розв'язано актуальне науково-прикладне завдання підвищення ефективності функціонування моніторингового програмного агента шляхом розробки методів побудови та засобів програмної реалізації багатoshарових агентних моделей. Наукова новизна одержаних результатів полягає в наступному:

- *вперше запропоновано метод багатoshарового синтезу агентних моделей, який ґрунтується на формуванні масиву вхідних даних, синтезі моделі та додаванні сигналу з її виходу до вхідного масиву для повторного синтезу і який, на відміну від існуючих, використовує попередню кластеризацію вхідних даних та побудову спеціалізованих моделей для кожного кластера. Це дозволяє підвищити однорідність точок спостереження у кластерах і, як наслідок, точність та адекватність результатів моделювання;*

- *удосконалено метод синтезу багатoshарових моделей та їх програмної реалізації за рахунок залучення різнотипних (гетерогенних) алгоритмів машинного навчання для формування структурних шарів, що дає змогу компенсувати алгоритмічні обмеження окремих класів моделей, підвищити повноту опису об'єкта та покращити результати моделювання;*

– *набув подальшого розвитку* метод багатошарового синтезу агентних моделей завдяки формалізації процесів конструювання алгоритмів синтезу на основі спрямованого ациклічного графа (DAG), що забезпечує багатоетапність удосконалення структури моделі, обробку проміжних сигналів на виході шарів та об'єднання моделей, побудованих у різних програмних середовищах, із гарантуванням детермінізму обчислень.

3. Практичне значення та цінність одержаних результатів

Практична цінність роботи полягає у створенні цілісної інформаційної технології проектування програмного забезпечення моніторингових програмних агентів. Розроблений метод підвищення однорідності даних та удосконалений метод рециркуляції дозволяють агенту автономно розпізнавати різні режими функціонування інформаційної системи й адаптувати під них власну прогностичну логіку завдяки збагаченню простору ознак.

Важливим прикладним здобутком є застосування декларативного підходу та подання процесів конструювання у вигляді DAG, що вирішує критичну інженерну проблему MLOps – надійне перенесення навчених складних ансамблів (із вбудованими етапами кластеризації та рециркуляції) у середовище експлуатації зі збереженням повної машинної відтворюваності обчислень та формуванням самодостатніх відтворюваних пакетів результатів. Експериментально підтверджено зниження середньоквадратичної похибки прогнозування (RMSE) на величину до 27,9%, а застосування удосконаленого методу рециркуляції забезпечило додаткове зниження похибки на 1,9...15,9 % порівняно з класичними статичними ансамблями.

Прикладну спрямованість результатів підтверджує їх використання у виконанні науково-дослідних робіт кафедри програмного забезпечення автоматизованих систем ЧДТУ – «Програмні агенти в інформаційній технології кризового інтелектуального моніторингу» (№ 0121U113866) та «Інтелектуальний моніторинг кібербезпеки корпоративної мережі програмними агентами» (№ 0125U001347). Одержані рішення мають потенціал для впровадження в

системах технічної діагностики, моніторингу мікросервісних архітектур та автоматизованого контролю складних інформаційних процесів.

4. Достовірність та обґрунтованість одержаних результатів

Наукові положення, висновки та рекомендації дисертаційної роботи є обґрунтованими. Достовірність забезпечується коректним застосуванням методів системного аналізу, машинного навчання, кластеризації (K-Means, гаусові суміші GMM, ієрархічна кластеризація), методів оптимізації гіперпараметрів (алгоритм Tree-structured Parzen Estimator, TPE) та граф-орієнтованого моделювання на основі DAG.

Теоретичні положення підкріплено масштабними експериментальними дослідженнями, проведеними на трьох структурно відмінних відкритих наборах табличних даних (зокрема прогнозування вартості поїздок, оцінювання вартості ювелірних виробів та аналіз демографічно-поведінкових показників). Якість моделей оцінено за загальноприйнятими метриками регресійного аналізу (RMSE, MAE, R^2) із застосуванням методології перехресної перевірки, а збіг результатів повторних запусків складної архітектури підтверджує машинний детермінізм запропонованого підходу. Достовірність результатів додатково засвідчена їх апробацією у виданні, що індексується базою Scopus, та у фахових наукових виданнях України.

5. Структура, оцінка змісту та завершеність дисертації

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел і додатків. Загальний обсяг становить 149 сторінок, з яких основний текст – 119 сторінок; робота містить 8 таблиць, 17 рисунків, 4 додатки обсягом 18 сторінок, а список використаних джерел налічує 100 найменувань.

У вступі обґрунтовано актуальність теми, сформульовано мету, завдання, об'єкт і предмет дослідження, розкрито наукову новизну та практичне значення

одержаних результатів, наведено відомості щодо зв'язку роботи з науковими темами, особистого внеску здобувача, апробації та публікацій.

Перший розділ присвячено аналізу проблематики динамічних змін у середовищах інтелектуального моніторингу складних інформаційних систем. Здійснено теоретичний аналіз методів прогнозування часових рядів, обґрунтовано доцільність застосування алгоритмів машинного навчання для задач табличної регресії та критично проаналізовано проблеми MLOps щодо відтворюваності експериментів, форм подання архітектур і обмежень наявних форматів серіалізації моделей. На цій основі сформульовано науково-прикладну проблему та завдання дослідження.

Другий розділ містить теоретичні основи та математичну формалізацію розроблених методів. Описано метод підвищення однорідності вхідних даних шляхом кластеризації простору ознак і збагачення його прогнозами локальних моделей, а також удосконалений метод багат шарової рециркуляції з використанням гетерогенних пулів алгоритмів і пошаровим тестуванням структури ансамблю.

Третій розділ присвячено експериментальній перевірці запропонованих рішень на трьох структурно відмінних відкритих наборах даних. Для оптимізації гіперпараметрів застосовано алгоритм TPE, а оцінювання виконано за метриками RMSE, MAE та R^2 . Наведені порівняльні таблиці й графіки переконливо демонструють, що комплексне застосування методів кластеризації та гетерогенної рециркуляції дозволяє суттєво знизити похибку прогнозування порівняно як з ізольованими базовими алгоритмами, так і з класичними статичними ансамблями.

Четвертий розділ відображає інженерно-програмну складову дослідження. Запропоновано компонентну об'єктно-орієнтовану архітектуру моніторингового агента та декларативний опис його конфігурації, а ключовим здобутком є подання багатоетапних процесів конструювання алгоритмів синтезу моделей у вигляді спрямованого ациклічного графа (DAG) із формуванням відтворюваного пакета результатів (bundle). Таке рішення усуває проблему втрати логіки взаємодії

алгоритмів під час збереження моделей і гарантує детермінізм та машинну відтворюваність обчислень.

У висновках узагальнено основні наукові та практичні результати, констатовано виконання всіх поставлених завдань і підтверджено досягнення мети дослідження. Загалом дисертаційна робота характеризується цілісністю, логічною завершеністю та достатнім рівнем обґрунтованості одержаних результатів.

6. Повнота викладення наукових положень у публікаціях

Основні наукові результати, положення та висновки дисертації повною мірою висвітлено у 14 наукових працях автора. Серед них 4 статті у фахових наукових виданнях України, 1 стаття у періодичному науковому виданні, проіндексованому в базі даних Scopus, та 9 тез доповідей на міжнародних і всеукраїнських науково-практичних конференціях. Аналіз змісту публікацій свідчить про їх відповідність тематиці дисертації, достатній рівень апробації результатів дослідження та їх наукову значущість. Опубліковані праці адекватно й повно відображають зміст дисертаційної роботи.

7. Відсутність порушення академічної доброчесності

Дисертаційну роботу виконано самостійно. За результатами аналізу ознак порушення принципів академічної доброчесності не виявлено: у тексті відсутні ознаки академічного плагіату, фабрикації чи фальсифікації результатів. Використані ідеї, твердження та результати інших авторів супроводжуються належними посиланнями на першоджерела, а внесок здобувача у працях, опублікованих у співавторстві, чітко розмежований і коректно зазначений. Робота відповідає принципам академічної доброчесності.

8. Зауваження та недоліки дисертації щодо її оформлення і змісту

Незважаючи на високий науковий рівень та практичну значущість роботи, можна відзначити окремі зауваження:

1. У роботі зустрічаються стилістичні та орфографічні помилки, недоліки в оформленні таблиць (окремі таблиці розбиті на дві сторінки без відповідних позначень), малюнків, відступів заголовків (спостерігається різне оформлення заголовків одного і того ж рівня), наявні пусті фрагменти на сторінках в різних місцях роботи. Також наявні таблиці, де на нову сторінку переноситься лише один рядок, що погіршує загальне сприйняття даних таких таблиць. Проте, зазначені помилки оформлення, не впливають на загальне сприйняття суті дослідження.

2. Позитивно оцінюючи застосування спрямованого ациклічного графа (DAG) для забезпечення детермінізму обчислень, варто зауважити, що питання споживання оперативної пам'яті під час опрацювання «глибоких» шарів рециркуляції висвітлено лише фрагментарно. Оскільки вузли графа зберігають масиви проміжних ознак, доцільно було б навести хоча б орієнтовні оцінки використання пам'яті залежно від кількості шарів ансамблю.

3. Автор аргументовано обґрунтовує об'єктно-орієнтовану архітектуру системи та формування стандартизованого пакета результатів. Водночас механізми інтеграції розробленого агента із наявними зовнішніми платформами збереження й версіонування моделей описано доволі стисло. Деталізація інтерфейсів автоматичного експорту результатів до сторонніх систем розширила б межі практичного застосування розробки.

4. У роботі здійснено ґрунтовний порівняльний аналіз різних методів кластеризації. Проте дискусійним видається підхід до визначення оптимальної кількості кластерів, яку в наведених експериментах підібрано емпірично. Для повністю автономного моніторингового агента доцільно було б передбачити внутрішній алгоритмічний механізм автоматичного визначення цієї кількості на основі метрик якості кластеризації.

5. Вибір трьох різнопланових відкритих наборів даних забезпечив надійну загальну валідацію методу багатошарової рециркуляції. Разом із тим, з огляду на орієнтованість роботи на моніторингові агенти інформаційних систем, серед тестових наборів бракує масивів суто інфраструктурної телеметрії (зокрема

системних журналів подій чи метрик навантаження серверів). Апробація саме на таких даних додатково підтвердила б ефективність розробки у її безпосередньому цільовому домені.

6. Експериментально доведено, що багатошарова рециркуляція знижує загальну похибку прогнозування. Водночас нарощування глибини ансамблю (до семи шарів) суттєво підвищує обчислювальну складність та ризик перенавчання моделі. Запровадження вбудованого механізму ранньої зупинки побудови нових шарів на основі контролю похибки валідації зробило б архітектуру агента більш ресурсоефективною.

Зазначені недоліки не мають принципового характеру, не применшують наукової новизни та практичної цінності отриманих результатів і не впливають на загальну позитивну оцінку дисертаційної роботи.

9. Загальні висновки та відповідність вимогам до ступеня доктора філософії

Дисертаційна робота Остапюка Володимира Вікторовича «Агентні моделі в інформаційній технології інтелектуального моніторингу» є завершеною, самостійно виконаною науково-дослідною працею, у якій розв'язано актуальне науково-прикладне завдання підвищення ефективності функціонування моніторингових програмних агентів шляхом удосконалення методів синтезу та програмної реалізації багатошарових адаптивних моделей.

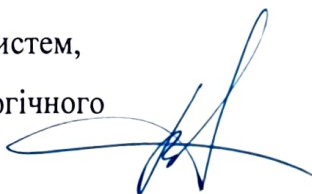
Одержані результати відзначаються науковою новизною, достатнім рівнем теоретичного обґрунтування та практичною спрямованістю. Запропоновані моделі, методи й алгоритмічні рішення мають прикладне значення та можуть бути використані при створенні й удосконаленні сучасних інформаційних технологій інтелектуального моніторингу складних розподілених систем.

Вважаю, що представлена до захисту дисертаційна робота за своїм змістом, ступенем актуальності, рівнем обґрунтованості наукових положень, новизною, повнотою викладу в наукових публікаціях та практичною значущістю відповідає

вимогам «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 січня 2022 р. № 44 (зі змінами), а її автор, ОСТАПЮК Володимир Вікторович, заслуговує на присудження наукового ступеня доктора філософії з галузі знань 12 «Інформаційні технології» за спеціальністю 121 «Інженерія програмного забезпечення».

Рецензент:

доцент кафедри програмного
забезпечення автоматизованих систем,
Черкаського державного технологічного
університету, к.т.н.



Вадим НЕМЧЕНКО

Підпис к.т.н. (PhD), доцента Вадима НЕМЧЕНКА засвідчую:

Учений секретар,
Черкаського державного
технологічного університету,
к.т.н. , доцент



Ірина МИРОНЕЦЬ