

Рецензія
на дисертаційну роботу
Сінька Дмитра Павловича
на тему «Комп'ютерна модель прогнозування стану розщеплення
кластеру розподіленої інформаційної системи з мікросервісною
архітектурою»,
представлену на здобуття ступеня доктора філософії в галузі знань
12 – Інформаційні технології за спеціальністю 122 – Комп'ютерні науки

Актуальність теми дисертації

Запропонована до захисту дисертаційна робота присвячена актуальному науковому дослідженню розвитку розподілених інформаційних систем, що обслуговують об'єкти енергетики та інші критично важливі галузі. Розвиток таких систем супроводжується зростанням складності мережевої взаємодії між компонентами. Керування такими системами в реальному часі не терпить розходження станів, а наслідки неузгодженості вимірюються не транзакціями, а стабільністю надання послуг для мільйонів споживачів.

Стан розщеплення кластера є фундаментальним наслідком розділення мережі й не може бути усунений у принципі, що впливає з відомих теоретичних результатів щодо неможливості одночасного забезпечення узгодженості, доступності та стійкості до мережевих розділень. Дослідження надійності розподілених сховищ даних показують, що мережеві розділення трапляються в промислових середовищах значно частіше, ніж прийнято вважати, а багато систем, формально спроектованих стійкими до них, насправді втрачають узгодженість за певних сценаріїв. Наявні інженерні рішення зосереджені на реагуванні на вже наявний поділ, тоді як кількісне оцінювання близькості конфігурації до критичного стану залишається малодослідженим.

Саме прогнозування цієї величини дає змогу завчасно виявляти вразливі конфігурації та переходити від реактивних до превентивних механізмів управління ризиками. З огляду на це, тема дисертаційної роботи Сінька Д. П. є актуальною.

Підтвердженням актуальності теми дисертації може бути те, що робота виконувалася відповідно до плану наукових досліджень Інституту проблем моделювання в енергетиці ім. Г. Є. Пухова НАН України, зокрема в рамках НДР «Розвиток наукових засад алгебраїчної теорії сильного штучного інтелекту стосовно кібернетичної безпеки об'єктів критичної інфраструктури в галузі енергетики» (№ ДР 0123U100913, 2023–2027 рр.) та НДР «Розвиток розподіленої енергетики в умовах ринку електричної енергії України з використанням технологій та систем цифровізації. Розділ 1. Організаційні та

математичні моделі взаємодії учасників децентралізованого ринку електроенергії» (№ ДР 0125U000237, 2025–2026 рр.).

Оцінка обґрунтованості наукових результатів дисертації, їх новизни

У дисертаційній роботі відповідно до мети та задач дослідження вирішено наукове завдання розроблення комп'ютерної моделі прогнозування відстані до розщеплення кластера розподіленої інформаційної системи з мікросервісною архітектурою на основі регресійних моделей машинного навчання та графово-інваріантного представлення стану кластера.

Обґрунтованість отриманих наукових положень підтверджується коректністю математичних постановок задач, застосуванням апробованих методів теорії графів та машинного навчання для їх розв'язання. Важливим етапом обґрунтування прикладних результатів є підтвердження властивостей інваріантності представлення процесу пошуку рішень автоматизованими тестами, а також відтворюваність наведених експериментальних результатів, що забезпечується фіксацією джерел випадковості та версій використаного програмного середовища.

Наукова новизна дисертаційної роботи полягає в такому:

- Вперше формалізовано відстань до стану розщеплення кластера як неперервну міру структурної крихкості та поставлено задачу її регресійного прогнозування, що, на відміну від відомої бінарної класифікації факту розщеплення, дозволяє зберігати дуже важливу для прийняття рішень щодо управління ризиками інформацію про ступінь близькості до критичного стану.
- Вперше запропоновано подвійно інваріантне представлення стану кластера, інваріантне водночас до перестановки вузлів і до перейменування типів.
- Вперше розроблено комп'ютерну модель прогнозування відстані до розщеплення, яка, на відміну від відомих, поєднує точний алгоритм обчислення цільової величини, конвеєр формування інваріантних ознак, регресійну модель CatBoost та механізми калібрування.
- Удосконалено спосіб представлення стану кластера завдяки поєднанню канонізованого матричного опису з графово-інваріантними структурними ознаками.
- Дістали подальшого розвитку методи калібрування та оцінювання невизначеності прогнозу, зокрема ізотонічне калібрування та конформні інтервали, застосовані до задачі регресії структурної крихкості.

Теоретичне значення роботи

Теоретичне значення роботи полягає у формалізації відстані до розщеплення кластера розподіленої інформаційної системи як неперервної міри структурної крихкості та в обґрунтуванні представлення стану кластера,

інваріантного до перестановки вузлів і перейменування типів. Такий підхід до вирішення задачі підвищення стійкості розподіленої інформаційної системи до розпаду кластера, що призводить до ризиків консистентності даних у системах, дозволяє перейти від констатації факту відмови до її кількісного передбачення. Він закладає основу для подальших досліджень у сфері оцінювання стійкості розподілених систем. Особливої уваги заслуговує також те, що автор не обмежився емпіричним добором ознак відмови, а пов'язав кожен з них із цільовою величиною через її теоретичні обмеження або релаксації, завдяки чому розроблена комп'ютерна модель прогнозування відстані до розщеплення кластера розподіленої інформаційної системи виконує справжню роботу наближених обчислень до результату.

Практична цінність результатів роботи

Практична цінність полягає у можливості застосування розробленого програмного комплексу для прогнозування відстані до стану розщеплення кластера та надання неперервної оцінки ризику з інтервалом надійності у зручній для користувача формі через вебсервіс. Комплекс інтегрується в системи моніторингу розподілених систем без суттєвих змін їх архітектури. Можливість оцінювання рівня ризику створює передумови для виконання етапу обробки ризиків і сприяє своєчасному прийняттю керуючих рішень, зокрема: коригуванню політик реплікації; обмеженню кількості операцій запису; ініціації контрольованих процедур відновлення запису. Вихідний код програмного комплексу доступний у відкритому репозиторії, що забезпечує можливість незалежної його перевірки та застосування задля подальшого вдосконалення отриманих теоретичних результатів.

Оцінка змісту дисертації, її завершеності та дотримання принципів академічної доброчесності

Дисертація Сінька Д. П. є завершеною кваліфікаційною науковою працею, яка за своїм змістом відповідає Стандарту вищої освіти зі спеціальності 122 «Комп'ютерні науки» та напрямам досліджень її освітньої програми. Дисертація свідчить про наявність особистого внеску автора в означених напрямках.

Дисертацію викладено на 204 сторінках; вона складається із вступу, чотирьох розділів, висновків, списку використаних джерел та додатка. Обсяг основного тексту дисертації становить 158 сторінок друкованого тексту. Робота містить 19 таблиць. Список використаних джерел містить 119 найменувань.

У вступі обґрунтовано актуальність теми дослідження, сформульовано його мету, поставлено наукові завдання, визначено предмет, об'єкт та наведено використані методи дослідження, викладено основні наукові

результати роботи, які характеризують її наукову новизну та практичне значення.

У першому розділі проведено досить докладний аналіз відкритих джерел щодо побудови архітектур розподілених інформаційних систем, природи стану розщеплення кластера розподіленої інформаційної системи та методів машинного навчання, придатних для його прогнозування. Розглянуто еволюцію архітектурних підходів, моделі реплікації, класичні алгоритми консенсусу та координаційні сервіси. Показано, що розщеплення кластера є фундаментальним наслідком розділення мережі, а наявні підходи здебільшого трактують його як бінарну подію, факт настання якої легко детектується. На підставі аналізу виявлено науково обґрунтовану прогалину та сформульовано мету, об'єкт, предмет і завдання дослідження.

У другому розділі побудовано математичну модель кластера у вигляді орієнтованого графа з типізованими вузлами, де ефективним вважається лише взаємне ребро, наявне в обох напрямках. Стан розщеплення кластера означено через поняття функціональних островів, тобто зв'язних компонент, що містять усі присутні типи. Уведено відстань до розщеплення кластера як неперервну міру структурної крихкості та обґрунтовано практично значущий діапазон цільової досліджуваної величини. Ключовим результатом розділу є інваріантне представлення стану кластера фіксованої розмірності, що досягається канонічним впорядкуванням вузлів на основі кольорового уточнення та канонічним маркуванням типів за структурним рангом. Запропоновано набір із 17 графово-інваріантних структурних ознак та виконано порівняльний аналіз методів машинного навчання, за результатами якого обґрунтовано вибір ансамблів дерев рішень.

У третьому розділі описано програмну реалізацію комп'ютерної моделі системи мовою Python у вигляді модульного комплексу з єдиною точкою формування ознак. Розроблено точний алгоритм обчислення відстані до розщеплення кластера, що використовує оригінальний інкрементний алгоритм як метод послідовного проведення обчислень на кожному кроці. Замість повного перерахунку стану системи з нуля алгоритм спирається на попередній результат, що дає економію часу зі зростанням кількості видалень, а також обмежує множину кандидатів ребрами функціонального острова. Коректність застосування запропонованого алгоритму підтверджено повним звіренням з еталонним перебором усіх підмножин ребер. Окрему увагу приділено машинному навчанню та калібруванню моделей, зокрема виявленню й усуненню виродження моделі за функцією втрат Хубера, застосуванню ізотонічного калібрування та конформних інтервалів, а також інтеграції навчених моделей у вебсервіс.

У четвертому розділі наведено результати експериментального дослідження запропонованих засобів моделювання. Центральним результатом дослідження є порівняння представлень за фіксованої моделі, що кількісно підтверджує визначальну роль ознакового опису. Виконано порівняння базових моделей за якістю та обчислювальною вартістю. Обґрунтовано вибір функції втрат, оцінено ефект калібрування та емпіричне покриття конформних інтервалів. Наведено додаткові абляції щодо обсягу даних, вагування прикладів і внеску канонізації. Дослідження цього розділу надало можливість авторові сформулювати обмеження запропонованих засобів комп'ютерного моделювання та визначити напрями подальших досліджень.

Дисертаційна робота оформлена відповідно до чинних вимог до оформлення дисертацій. У роботі відсутні порушення академічної доброчесності. У ній викладено результати особистих досліджень автора, а використані результати з інших джерел мають відповідні посилання. Окремо слід відзначити, що здобувач прозоро задекларував використання інструментів штучного інтелекту для пошуку літератури та стилістичної перевірки тексту із зазначенням особистої відповідальності за зміст роботи, що відповідає сучасним підходам до академічної доброчесності.

Оприлюднення результатів дисертаційної роботи

Наукові результати дисертації висвітлено у 12 наукових публікаціях здобувача, серед яких статті у наукових виданнях, включених на дату опублікування до переліку наукових фахових видань України категорій «А» та «Б», статті у періодичних наукових виданнях, проіндексованих у базах даних Scopus, а також тези доповідей на всеукраїнських і міжнародних науково-технічних конференціях.

Наукові публікації здобувача свідчать про належний науковий рівень проведених досліджень. Для кожної праці, опублікованої у співавторстві, здобувачем чітко визначено особистий внесок. Аналіз цих внесків показує, що здобувач брав безпосередню участь у постановці задачі дослідження, запропонував поняття відстані до розщеплення як оцінки ризику розпаду кластера на острови, обґрунтував підхід до зменшення розмірності під час моделювання складних кластерних систем, а також виконав практичну реалізацію проактивної системи виявлення розщеплення кластера. Таким чином, особистий внесок здобувача є вагомим, відповідає основному змісту дисертаційної роботи і підтверджує самостійність отриманих наукових результатів.

Недоліки та зауваження до дисертаційної роботи

Незважаючи на належний науковий рівень та актуальність проведеного дослідження, можна виокремити кілька аспектів, які потребують додаткової уваги або могли б бути оформлені більш вдало.

1. Формула (2.7) фактично є арифметичною рівністю, що визначає підсумкову розмірність вектора ознак. Її природніше було б навести безпосередньо в тексті абзацу без окремої нумерації, залишивши нумерацію для змістовних співвідношень.

2. Перелік умовних скорочень містить дев'ять позицій, серед яких SVR, GNN та ГНМ. Проте до нього не внесено низку скорочень, що активно вживаються в тексті, зокрема MAE, RMSE та назви конкретних реалізацій градієнтного бустингу. Ця обставина ускладнює читання роботи.

3. Ефект ізотонічного калібрування добре проілюстровано таблицею 4.6, у якій наведено середні прогнози за класами істинної відстані. Порівняння з класичною калібрувальною кривою додало б цьому важливому результату наочності.

4. У таблицях 4.2 та 4.3, що містять порівняння представлень і базових моделей, доцільно було б виділити найкращі значення метрик напівжирним накресленням — це полегшило б читачеві сприйняття підсумку порівняння.

5. У тексті подекуди зустрічаються друкарські неточності, зокрема «діверсифікація» замість «диверсифікація».

6. У роботі наведено єдиний рисунок, позначений як «Рис. 2.1 – Послідовність дій системи прогнозування СРК». За наявності одного рисунка на всю дисертацію подвійна нумерація за розділами втрачає сенс.

7. У тексті паралельно вживаються форма «вузол» і калька «нода». Доцільно уніфікувати ці форми в усьому тексті, щоб матеріал легше читався.

8. На відміну від таблиць, які мають розгорнуті назви, лістинги 3.1 і 3.2 позначено лише номером, без найменування. Оскільки лістингів у роботі два, доцільно або назвати їх за зразком таблиць, або, зважаючи на невелику кількість, оформити наскрізною нумерацією.

9. Висновки до розділів 1 і 4 оформлено маркованим списком, а до розділів 2 і 3 — нумерованим. Це варто було б уніфікувати.

Наведені зауваження не є визначальними, мають рекомендаційний характер, стосуються оформлення роботи та не впливають на позитивну оцінку її наукової новизни й практичної значущості.

Висновок про дисертаційну роботу

Вважаю, що дисертаційна робота здобувача Сінька Дмитра Павловича «Комп'ютерна модель прогнозування стану розщеплення кластеру розподіленої інформаційної системи з мікросервісною архітектурою» є завершеним, самостійно виконаним науковим дослідженням, у якому

вирішено важливе наукове завдання відповідно до мети дослідження, теоретичні та прикладні результати якого розв'язують комплекс задач прогнозування стану розщеплення кластера розподіленої інформаційної системи з мікросервісною архітектурою, що має істотне значення для галузі інформаційних технологій у напрямі підвищення стійкості розподілених інформаційних систем до розпаду кластерів. Робота відзначається чіткою логічною структурою, матеріал викладено послідовно, висновки до кожного розділу органічно впливають з його змісту.

За актуальністю, науковою новизною, теоретичною та практичною цінністю отриманих результатів дисертація повністю 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.

Здобувач Сінько Дмитро Павлович заслуговує на присудження ступеня доктора філософії в галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп'ютерні науки».

Рецензент:

головний науковий співробітник відділу
математичного та економетричного
моделювання Інституту проблем
моделювання в енергетиці

ім. Г.Є. Пухова

НАН України, д.т.н., с.н.с.



Зелімхан БОРУКАЄВ