

ВІДГУК

офіційного опонента

на дисертаційну роботу

Кириченко Оксани Леонідівни

на тему «Дослідження статистичних характеристик складних мереж методами інтелектуального аналізу даних», представлену на здобуття ступеня доктора філософії в галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення

Актуальність теми дисертації

Сучасний інформаційний простір є прикладом складної мережі з певними особливостями та зв'язками. Дослідження статистичних особливостей сприяє кращому розумінню організації та використання мереж. Статистичною основою досліджень є адекватна сегментація складних мереж на організовані за різними критеріями структури меншої розмірності з метою спрощення аналізу великих обсягів даних, дослідження структури мережі та її кластерів.

Існуючі математичний, алгоритмічний та програмний апарати інтелектуального аналізу складних мереж не дають відповіді на запитання щодо оптимального вибору кластерів складної мережі, вивчення статистичних особливостей підмереж, методик збирання та аналізу інформації про такі мережі, тощо.

Все це зумовлює актуальність теми дисертаційної роботи Кириченко Оксани Леонідівни, присвяченої розв'язанню важливої та актуальної задачі проведення статистичного та кластерного аналізу окремих зон веб-простору на основі розроблених програмного забезпечення та підходів для визначення оптимальної кількості кластерів мереж.

Зв'язок роботи з науковими програмами, планами, темами.

Дисертаційна робота виконана на кафедрі програмного забезпечення комп'ютерних систем Чернівецького національного університету імені Юрія Федьковича. Її зміст відповідає тематиці науково-дослідних робіт: «Математичне та програмне забезпечення обчислювальних систем»

(Державний реєстраційний номер 011U007046) і «Дослідження, моделювання та розробка програмного забезпечення складних динамічних систем» (Державний реєстраційний номер 0121U109232).

Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни

Наукова новизна результатів дисертаційного дослідження полягає в наступному:

1. Уперше розроблено кроулер, який базується на сучасних методах розробки програмного забезпечення та, на відміну від існуючих, містить аналітичний модуль інтелектуальної обробки інформації, підтримує контейнеризацію, роботу в багатопотоковому режимі, легко масштабується, і може бути використаний для аналізу споріднених мереж;

2. Уперше розроблено інформаційну технологію збирання та інтелектуального аналізу даних у складних мережах, яка уможливорює дослідження статистичних характеристик складних мереж, збирання статистичної інформації у веб-просторі та її обробки методами інтелектуального аналізу даних та може в подальшому використовуватися для практичного дослідження складних мереж;

3. Набула подальшого розвитку теорія дослідження графових структур, зокрема, сформульовано і доведено твердження, які дозволяють оцінювати розподіл власних значень випадкових матриць та переносити класичні результати на матриці із слабо корельованими елементами; на основі розширення границь теорії дослідження графових структур розроблено алгоритм перевірки належності елементів до одного кластеру, що надає можливість визначити її оптимальну кластерну структуру.

Наукові положення, висновки і рекомендації дисертаційної роботи Кириченко О.Л. достатньо обґрунтовані коректним використанням математичного апарату, підкріплені успішною реалізацією, ефективним практичним впровадженням результатів дисертаційних досліджень, яке продемонструвало збігання теоретичних досліджень з реальними результатами.

Обґрунтованість наукових положень та висновків, сформульованих у дисертаційній роботі, є достатньою і базується на детальному аналізі джерел за даною проблемою, чіткій постановці задач дослідження, використанні сучасних методів дослідження, правильним застосуванням математичного апарату при теоретичному розгляді наукових положень дисертації, а також проявляється у якісному та аргументованому формулюванні висновків.

Достовірність та обґрунтованість запропонованих методів і засобів підтверджується результатами експериментальних досліджень та коректним застосуванням методів, які були використані під час виконання роботи.

Наукові положення, висновки та рекомендації, сформульовані в дисертації, логічно випливають із результатів, отриманих за допомогою чітких викладок. Тому можна стверджувати, що висновки та практичні рішення, отримані у роботі, коректні, достатньо обґрунтовані й можуть бути рекомендовані до використання.

Отже, в дисертаційній роботі поставлене наукове завдання виконано повністю, здобувач повною мірою оволодів методологією наукової діяльності.

Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності

За своїм змістом дисертаційна робота здобувача Кириченко О.Л. повністю відповідає Стандарту вищої освіти зі спеціальності 121 Інженерія програмного забезпечення та напрямкам досліджень відповідно до освітньої програми «Інженерія програмного забезпечення» Чернівецького національного університету ім. Ю. Федьковича. Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувача у науковий напрям інженерії програмного забезпечення.

Розглянувши звіт подібності за результатами перевірки дисертаційної роботи на текстові співпадіння, можна зробити висновок, що дисертаційна робота Кириченко Оксани Леонідівни є результатом самостійних досліджень здобувачки і не містить елементів фальсифікації, компіляції, фабрикації, плагіату та запозичень. Використані результати і тексти інших авторів мають належні посилання на відповідне джерело.

Практичне значення одержаних результатів

Розроблені у дисертаційній роботі кроулер, інформаційна технологія та метод визначення оптимальної кількості кластерів можуть в подальшому використовуватися для практичного дослідження складних мереж.

Запропоновані підходи до архітектури аналітичного модуля використовуються у компанії «Квант Азимут» для розробки власного програмного забезпечення та компанії «Qlicks B.V.» – для проведення сегментації клієнтів на різні категорії, які потім ефективно використовуються для персоналізованих маркетингових кампаній і стратегій та передбачення поведінки клієнтів на основі аналізу покупок, історії пошуку або профілей в соціальних мережах.

Мова та стиль викладення результатів

Дисертаційна робота написана українською мовою.

Дисертація написана логічно, доступно, на високому технічному рівні з використанням сучасної термінології. Матеріали дисертаційної роботи викладено послідовно, доступно для розуміння і сприйняття. Стиль мовлення задовольняє вимоги до текстів науково-технічного змісту. Текст дисертації в достатній мірі проілюстрований таблицями та рисунками, а також фрагментами програмного коду. Здобувачка використовує загальноприйнятту термінологію. Дисертаційна робота має логічну структуру.

Дисертація складається зі вступу, 4 розділів, висновків, списку використаних джерел та додатків. Загальний обсяг дисертації 229 сторінок, в т. ч. 182 сторінки основного тексту.

У вступі обґрунтовано актуальність теми дисертації, зазначено мету і задачі дослідження, сформульовано наукову новизну отриманих результатів, а також показано зв'язок роботи з науковими програмами підрозділу, в якому виконувались дослідження.

У першому розділі дисертації проаналізовано складні мережі та їх роль в інформаційному просторі, теорію складних мереж, моделі та приклади складних мереж, методи та засоби інтелектуального аналізу даних та

кластеризації у дослідженні складних мереж. Обґрунтовано доцільність проведення статистичного та кластерного аналізу окремих зон веб-простору.

У другому розділі розроблено програмне забезпечення для збору даних, зокрема, описано концепцію кроулінгу як засобу збирання інформації, визначено та проаналізовано вимоги до кроулера, розроблено архітектуру та структуру програмного забезпечення, обрано засоби розроблення програмного забезпечення та описано результати роботи розробленого програмного забезпечення.

У третьому розділі виконано статистичний та кластерний аналіз веб-простору методами інтелектуального аналізу даних, зокрема, проаналізовано статистичні характеристики інформації у веб-просторі, визначено особливості застосування методів кластерного аналізу до сегментів веб-простору, розроблено інформаційну технологію для збору статистичної інформації та проведення кластерного аналізу у веб-просторі та проаналізовано отримані результати.

У четвертому розділі виконано оцінку оптимальної кількості кластерів складних мереж на основі випадкових матриць, зокрема, проаналізовано основні відомості та твердження, застосування теорії випадкових матриць, виконано аналіз кластерної структури інтернет-мереж на основі випадкових матриць, виконано моделювання та порівняння з класичними результатами.

У висновках охарактеризовані та систематизовані наукові та практичні результати, отримані в ході дисертаційного дослідження. Основні висновки і рекомендації логічно витікають із результатів, які наведено у розділах роботи.

Дисертаційна робота оформлена відповідно до вимог наказу МОН України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

Оприлюднення результатів дисертаційної роботи

Наукові результати дисертації висвітлені у 22 наукових публікаціях здобувача, серед яких: 1 стаття в журналі, що індексується в наукометричній базі Scopus, 5 статей у наукових виданнях, включених на дату опублікування до переліку наукових фахових видань України, 1 розділ монографії міжнародного

видання та 15 робіт у збірниках матеріалів міжнародних та всеукраїнських наукових конференцій.

Також результати дисертації були апробовані на 8 міжнародних та всеукраїнських наукових фахових конференціях, семінарах, літніх школах.

Сумарно опубліковані праці віддзеркалюють повноту викладу результатів дисертаційної роботи. Науковий рівень публікацій – високий. У всіх публікаціях здобувачем дотримано принципів академічної доброчесності.

Таким чином, наукові результати, описані в дисертаційній роботі, повністю висвітлені у наукових публікаціях здобувача.

Недоліки та зауваження до дисертаційної роботи

Незважаючи на загальне позитивне враження від дисертаційної роботи, варто відзначити деякі недоліки та зауваження до дисертаційної роботи:

1. На мою думку, некоректно сформульовано мету дисертаційного дослідження, оскільки «проведення статистичного та кластерного аналізу окремих зон веб-простору...» є засобом досягнення мети, а не, власне, метою;

2. У вступі до дисертаційного дослідження відсутній опис практичної значущості отриманих під час дисертаційного дослідження результатів, такий опис є лише наприкінці анотації (стор. 6). Але при формулюванні практичного значення одержаних результатів у анотації здобувачка лише констатує можливість подальшого використання розроблених нею кроулера, інформаційної технології та методу, проте не зазначає ефект та кількісне значення ефекту, який забезпечили розроблені нею методи та засоби;

3. У підрозділі 2.2 дисертантка виконує визначення та аналіз вимог до кроулера, проте при визначенні нефункціональних вимог (стор. 70-72) вона оминає своєю увагою такі важливі нефункціональні вимоги для будь-якого програмного забезпечення, як якість, надійність, безпека, зручність використання (usability), тощо;

4. Оскільки здобувачка розробила інформаційну технологію для збирання статистичної інформації та проведення кластерного аналізу у веб-просторі (підрозділ 3.3), то, на мою думку, доцільно було б у дисертаційній роботі представити структуру формування цієї інформаційної технології, визначити

принципи проектування та функціонування інформаційної технології, описати етапи запропонованої інформаційної технології, представити узагальнену та деталізовану схеми розробленої інформаційної технології;

5. Здобувачка регламентує, що «статистичною основою досліджень служить адекватна сегментація складних мереж...», однією із розв'язаних в дисертаційній роботі задач вона зазначає «дослідити адекватність запропонованого методу визначення оптимальної кількості кластерів за допомогою методів моделювання», натомість висновок про адекватність розроблених підходів зібраних статистичних даних приймається на основі отримання практично однакових результатів двох методів визначення оптимальної кількості кластерів у наборі даних: метода “ліктя” та k-core decomposition (стор. 139). На мою думку, доцільно було б більше уваги приділити визначенню саме адекватності як сегментації складних мереж, так і пропонованого методу визначення оптимальної кількості кластерів;

6. Враховуючи, що в роботі використовуються методи кластеризації та класифікації, на мою думку, доцільно було б виконати розрахунок метрик кластеризації (Cluster Homogeneity, Cluster Completeness, Rag Bag, Size vs Quantity, тощо) та метрик класифікації (Accuracy, Precision, Recall, F1, Specificity, площа під Precision-Recall-кривою, площа під ROC-кривою, тощо), які б дали можливість зробити висновок про якість розроблених та використаних методів кластеризації та класифікації;

7. Неприпустимим, як на мене, для галузі інформаційних технологій є посилання на літературні джерела 15-річної і більше давнини (джерела 1, 2, 5-8, 11-15, 18, 19, 22, 25, 27-32, 36, 37, 40, 41, 48-53, 55, 59-67, 70, 71, 77, 82, 90, 92, 105, 117, 119, 121, 123, 129, 133, 134-138, 163, 164, 166, 167, 170, 171, 172, 181, 186-194, 197-202, 213-215, 224, 225, 228, 230, 237, 245, 248). Доцільніше було б замінити посилання на ці застарілі джерела посиланнями на актуальні англійські статті, які містять дійсно сучасні дослідження в галузі статистичних характеристик складних мереж.

Вважаю, що висловлені зауваження не є визначальними і не применшують загальної наукової новизни і практичної значущості результатів дисертаційного дослідження та не впливають на позитивну оцінку дисертаційної роботи.

Висновок про дисертаційну роботу

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Кириченко Оксани Леонідівни на тему «Дослідження статистичних характеристик складних мереж методами інтелектуального аналізу даних» виконана на високому науковому рівні, не порушує принципів академічної доброчесності та є закінченим науковим дослідженням, сукупність теоретичних та практичних результатів якого розв'язує наукове завдання, що має істотне значення для інформаційних технологій. Дисертаційна робота за актуальністю, практичною цінністю та науковою новизною повністю відповідає вимогам чинного законодавства України, що передбачені в п. 6 – 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 січня 2022 р. № 44.

Здобувач Кириченко Оксана Леонідівна заслуговує на присудження ступеня доктора філософії в галузі знань Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення.

Офіційний опонент:

Завідувачка кафедри комп'ютерної інженерії та
інформаційних систем

Хмельницького національного університету,
докторка технічних наук, професорка

 Тетяна ГОВОРУЩЕНКО

«Підпис Говорущенко Т.О. засвідчую»:

Проректор з НР ХНУ

01.02.2024



 Олег СИНЮК