

Спеціалізованій вченій раді ДФ 76.051.044
Чернівецького національного університету
імені Юрія Федьковича,
(58002, м. Чернівці, вул. Коцюбинського 2)

ВІДГУК

**офіційного опонента – доктора технічних наук, професора,
завідувача кафедри інженерії програмного забезпечення та кібербезпеки
Державного торговельно-економічного університету**

Криворучко Олени Володимирівни

**на дисертаційне дослідження Кириченко Оксани Леонідівни
на тему «Дослідження статистичних характеристик складних мереж
методами інтелектуального аналізу даних», поданої на здобуття наукового
ступеня доктора філософії за спеціальністю 121 – Інженерія програмного
забезпечення галузі знань 12 – Інформаційні технології**

1. Актуальність теми дисертаційної роботи

Сьогоднішній етап розвитку інформаційних технологій характеризується, безумовно, тим, що людство оперує величезними обсягами різноманітної інформації, для чого найчастіше використовують методи інтелектуального аналізу даних. Більше того, інформаційний простір, де ми живемо і працюємо та обробляємо потрібну нам інформацію, часто являє собою складну мережу, яка, внаслідок своєї складності, також потребує досліджень. Розуміння статистичних, кластерних та інших характеристик комплексних мереж сприяє їх кращому використанню у різних галузях людської діяльності. Велике значення при цьому має адекватна сегментація накопичених обсягів даних, оскільки аналіз повнорозмірних даних ще дуже ускладнений. Кластеризація дозволяє спростити такий аналіз, дослідити структуру мережі в цілому та її частин зокрема.

Не дивлячись на те, що процес інтелектуального аналізу складних мереж достатньо забезпечений математичним, алгоритмічним та програмним забезпеченням, існує ряд проблем, що потребує вирішення. Зокрема, це вибір оптимальної кількості кластерів, яка зазвичай невідома перед проведенням кластеризації, глибоке дослідження різних підмереж однакового суспільного

призначення, технології збирання та збереження та аналізу інформації про складні мережі тощо.

Дисертаційна робота Оксани Леонідівни Кириченко є вдалою спробою вирішити зазначені проблеми, і, зважаючи на наведене вище, вибір тематики дослідження та поставлені перед дисертанткою завдання є, безумовно, актуальними.

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

Дисертаційну роботу Кириченко О.Л. виконано на кафедрі програмного забезпечення комп'ютерних систем Чернівецького національного університету імені Юрія Федьковича. Кафедра проваджує діяльність за акредитованою аспірантською програмою «Інженерія програмного забезпечення», яку закінчила дисертантка, та науково-дослідною темою «Дослідження, моделювання та розробка програмного забезпечення складних динамічних систем» (номер Державної реєстрації 0121U109232). Тематика дисертації повністю відповідає зазначеній освітньо-науковій програмі III рівня вищої освіти та кафедральній тематиці науково-дослідних робіт, а рецензована дисертаційна робота виконана в рамках цієї тематики.

3. Аналіз змісту дисертації, обґрунтованість наукових положень, висновків та рекомендацій

Дисертація є завершеною науково-дослідною роботою, що складається з анотації (українською та англійською мовами), змісту, переліку умовних скорочень, чотирьох розділів, висновків, списку використаних джерел та додатків.

У *вступі* обґрунтовано актуальність теми дисертації, визначено об'єкт та предмет досліджень, сформульовано мету та задачі, наукова новизна дослідження та практична цінність отриманих результатів, подано відомості про апробацію роботи та наведено її структуру.

Перший розділ роботи присвячено аналізу теорії та існуючих моделей складних мереж; продемонстровано їхню роль у інформаційному просторі сьогодення. Дисертантка розглянула моделі Ердоша-Рені, Уатса-Строгаца та Барабаші-Альберта, навела їхні переваги та недоліки. Розглянуто найважливіші приклади складних мереж: соціальні, інформаційні, технологічні та біологічні. Наведено також основні методи та засоби інтелектуального аналізу даних з

особливим наголосом на процесі кластеризації при дослідженні складних мереж.

Проведення такого аналізу дозволило автору виокремити ряд нерозв'язаних задач та зробити висновок про необхідність проведення подальших досліджень у цій галузі інтелектуального аналізу даних.

Матеріал розділу висвітлено повно, проведено ґрунтовний аналіз предметної галузі, ретельно та всебічно обґрунтовано актуальність та завдання дисертаційного дослідження.

У другому розділі дисертантка розглянула засади концепції кроулінгу як засобу збирання інформації у складних мережах та зроблено детальний огляд існуючого програмного забезпечення для збирання інформації, що дозволило сформулювати вимоги до програмного забезпечення власної розробки, визначити його архітектуру та структуру, обрати оптимальні засоби розробки.

У третьому розділі увага дисертантки зосереджена на статистичному та кластерному аналізі веб-простору з використанням розробленої інформаційної технології. Авторкою проаналізовано освітянські сегменти веб-простору України, Польщі та Ізраїлю. На основі проведених досліджень встановлено статистичну подібність характеристик усіх сегментів, однак український освітянський сегмент явно ненасичений вузлами освітянських закладів, що й продемонстрував кластерний аналіз. Український освітянський сегмент становить найменш розвинену структуру, кількість вузлів у кластерах найменша з усіх досліджених сегментів.

Релевантність отриманих результатів підтверджується тим, що збільшення глибини зондування, проведене авторкою, не призвело до якісних змін статистичної картини характеристик, а вплинуло лише на кількісні показники.

Четвертий розділ дисертації присвячено розробці нового методу оцінки оптимальної кількості кластерів у складних мережах на основі випадкових матриць.

У попередніх розділах дисертації авторкою зроблено висновок, що існуючі сьогодні традиційні методи кластеризації вимагають у користувача попередньо визначити потрібну кількість кластерів у наборі даних, а оптимальна ця кількість чи ні, – вирішується додатковими дещо штучними методами, наприклад, методом “ліктя”. Такий підхід не дуже вдалий, оскільки у задачах, які вирішуються у рецензованій дисертації (та й у багатьох інших),

кількість кластерів завчасно невідома. Тому було поставлено завдання розробки власного методу оцінки оптимальної кількості кластерів, позбавленого таких недоліків. Розробку виконано на основі теорії випадкових матриць. Авторкою доведено ряд лем та теорем, які дозволили використати теорію стохастичних випадкових матриць до аналізу структури інтернет-мереж. Проведено моделювання та порівняння з класичними результатами, що продемонструвало адекватність, а в деяких випадках, і вищу точність розробленого методу оцінки оптимальної кількості кластерів.

Висновки дисертаційної роботи сформульовано чітко, вони повністю охоплюють отримані результати і повністю задовольняють вимоги, що ставляться до результатів дисертації на здобуття наукового ступеня доктора філософії.

Список використаних джерел повно охоплює предметну галузь і вказує на аналіз значної кількості літературних джерел, як українських, так і закордонних авторів, проведений автором.

Додатки містять діаграму класів розробленого кроулера та її опис; лістинг основних частин коду програми; акти впровадження результатів дисертаційної роботи; список публікацій за темою дисертації.

4. Наукова новизна одержаних результатів

До найсуттєвіших та нових наукових результатів дисертаційної роботи О.Л. Кириченко, на мою думку, можна віднести таке:

1. Авторкою вперше:

- з використанням сучасних методів розробки програмного забезпечення створено кроулер, який, на відміну від відомих, містить аналітичний модуль інтелектуальної обробки інформації, підтримує контейнеризацію, роботу в багатопотоковому режимі, легко масштабується і може бути використаний для аналізу не лише веб-, а й споріднених складних мереж;

- розроблено інформаційну технологію збирання та інтелектуального аналізу даних у веб-мережах, яка також може бути використана як основа для аналізу споріднених складних мереж;

- з використанням отриманих при кроулінгу результатів проведено порівняльний аналіз статистичних характеристик та кластерної структури освітянських сегментів України (edu.ua), Польщі (edu.pl) та Ізраїлю (ac.il).

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

Вважаю, що отримані дисертанткою наукові результати, є вагомим внеском у методи та засоби дослідження комплексних мереж за допомогою інтелектуального аналізу даних.

5. Достовірність отриманих результатів та висновків

Достовірність результатів, отриманих в дисертаційній роботі О.Л. Кириченко, забезпечено коректною постановкою задачі, мети та завдань дослідження, послідовним їх розв'язанням та аргументованим вибором методів та програмних засобів розробки.

Достовірність наукових положень, висновків та рекомендацій підтверджуються сучасними методами досліджень, які відповідають поставленій задачі, глибоким аналізом об'єкта та предмета дослідження за допомогою адекватних методів.

Обґрунтованість результатів дисертаційного дослідження обумовлена глибоким порівняльним аналізом отриманих результатів з результатами інших науковців, відомими теоретичними та практичними даними.

Достовірність результатів, висновків та наукових положень також підтверджується публікаціями основних результатів дисертаційної роботи, а також успішною апробацією на всеукраїнських та міжнародних науково-практичних конференціях.

6. Практична цінність одержаних результатів

Практична цінність наукових результатів, одержаних у дисертаційній роботі О.Л. Кириченко, полягає у практичному застосуванні теоретичних положень, методів та технологій, що підтверджується актами про впровадження результатів дисертації. Зокрема, компанія «Квант Азимут» використовує архітектуру аналітичного модуля для розробки власного програмного забезпечення; компанія Qlicks B.V. – для проведення поділу клієнтів на різні категорії, що потім використовується у маркетингових цілях.

Варто зауважити, що результати дисертаційного дослідження використовувалися у Чернівецькому національному університеті імені Юрія Федьковича при виконанні держбюджетного проєкту «Дослідження, моделювання та розробка програмного забезпечення складних динамічних систем» (номер Державної реєстрації 0121U109232).

7. Оформлення дисертації, дотримання вимог академічної доброчесності, повнота викладу наукових результатів у публікаціях

Дисертація має повний обсяг 229 сторінок друкованого тексту, причому основна частина викладена на 161 сторінці. Список використаних джерел складає 252 найменування.

Дисертаційна робота має логічну структуру, висновки та рекомендації відповідають отриманим у розділах результатам. Оформлення дисертації задовольняє усі вимоги до такого роду кваліфікаційних наукових праць.

Результати перевірки дисертаційної роботи на наявність академічного плагіату (8,24% схожості з джерелами з Інтернету) свідчать про високу індивідуальність роботи. Авторський стиль простежується по всьому тексту дисертації. Відсутні запозичення і використання результатів інших авторів без посилання на їхні джерела. Це підтверджує, що дисертаційна робота О.Л. Кириченко відповідає нормам академічної доброчесності.

Основні положення дисертації та найважливіші її результати опубліковано у науковій періодиці та апробовані на наукових конференціях. Зокрема, дисертанткою опубліковано 1 статтю у науковому виданні, яке індексується у наукометричній базі SCOPUS; 3 статті – у виданнях, включених до переліку наукових фахових видань України; 15 робіт – у збірниках матеріалів наукових конференцій; 3 статті – інші публікації, що додатково відображають наукові результати дисертації. Отже, вимоги щодо кількості та якості наукових публікацій авторкою виконано.

8. Недоліки та зауваження до дисертаційної роботи

Однак, дисертаційна робота Кириченко О.Л. не позбавлена деяких недоліків. До них можна віднести наступне:

1. На жаль, розроблений дисертанткою метод визначення оптимальної кількості кластерів у наборі даних апробовано лише на модельних прикладах. Доцільно було би застосувати його до реальних даних, які зібрані авторкою у реальному веб-просторі.

2. При збиранні даних у процесі кроулінгу не зазначається тип веб-сторінок (статичні чи динамічні) і як аналізуються динамічні сторінки.

3. Дисертаційна робота добре стилістично і грамотно оформлена, хоча і містить незначну кількість технічних неточностей.

Однак, зазначені недоліки не є принциповими, не впливають на загальну позитивну оцінку дисертаційної роботи, не змінюють її наукової новизни та практичної цінності.

9. Загальні висновки

Подана до захисту дисертаційна робота є завершеною науковою працею, яка містить нові, добре обґрунтовані результати. Дисертація розв'язує актуальну науково-прикладну задачу у галузі інтелектуальної обробки даних у складних мережах.

Тема, зміст та результати дисертації повністю відповідають спеціальності 121 – Інженерія програмного забезпечення.

Зважаючи на актуальність теми, обґрунтованість наукових положень, висновків, поданих у дисертації, їхню новизну та практичну цінність, достатню кількість наукових публікацій, а також відсутність порушень академічної доброчесності, вважаю, що дисертаційна робота «Дослідження статистичних характеристик складних мереж методами інтелектуального аналізу даних» цілком 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 №44 від 12 січня 2022 року зі змінами, внесеними згідно з Постановою Кабінету Міністрів України №341 від 21 березня 2022 року, а також «Вимогам до оформлення дисертації», затверджених Наказом Міністерства освіти і науки України №40 від 12 січня 2017 року, а її авторка, Кириченко Оксана Леонідівна, заслуговує присудження їй наукового ступеня доктора філософії з галузі знань 12 – Інформаційні технології за спеціальністю 121 – Інженерія програмного забезпечення.

Офіційний опонент – доктор технічних наук,
професор, завідувач кафедри інженерії
програмного забезпечення та кібербезпеки
Державного торговельно-економічного університету

Олена КРИВОРУЧКО

Людмила Олександрівна Криворучко
Генеральний секретар



29.01.2024