

ВІДГУК

офіційного опонента – доктора фізико-математичних наук, професора, професора кафедри дослідження операцій факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка Самойленка Ігоря Валерійовича на кваліфікаційне дослідження «Оцінки параметрів авторегресійних моделей» Книгніцької Тетяни Василівни, яка здобуває науковий ступінь доктора філософії у галузі знань 11 «Математика та статистика» за спеціальністю 113 «Прикладна математика».

Актуальність теми виконаної роботи.

Актуальність роботи зумовлена необхідністю активно досліджувати та застосовувати елементи штучного інтелекту у різноманітних сферах людської діяльності. Це призводить до появи цілого спектру нових задач прикладної математики, зокрема проблем дослідження часових рядів, встановленню подібності між ними та визначенню оптимальної кількості кластерів для даних, що представлені у вигляді графів. Саме ці аспекти є об'єктом дисертаційного дослідження Книгніцької Т.В.

Зауважимо, що оскільки збереження різноманітних даних у формі часових рядів є одним з найбільш поширених підходів, кластеризація даних, які представлені часовими рядами або графами, є важливим методом який дозволяє спрощувати аналіз великих масивів даних, та знаходить застосування в різних сферах, від фізики до біології та медицини. Зважаючи на збільшення кількості різноманітних досліджень, спрощення збереження та доступу до великих об'ємів даних через інформаційні мережі, зацікавленість науковців у обробці все більших масивів даних, розробка нових методів які б дозволили спростити, укрупнити та узагальнити дані є одним з основних викликів для сучасної прикладної науки.

Саме тому тематика дисертаційного дослідження Книгніцької Т.В. є, безумовно, актуальною та важливою.

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

Обраний напрям дослідження відповідає програмі наукової тематики кафедри прикладної математики та інформаційних технологій Чернівецького національного університету імені Юрія Федьковича "Математичне моделювання і числово-аналітичні методи дослідження динамічних та інформаційних процесів".

Обґрунтованість і достовірність наукових положень, висновків і рекомендацій забезпечена публікацією відповідних результатів у рейтингових рецензованих міжнародних виданнях. Три статті, що були опубліковані у періодичних наукових виданнях за темою дисертації повністю охоплюють проблематику та основні положення наукової роботи. Апробація основних результатів дисертації відбулася під час представлення доповідей на 8 наукових конференціях. Розгорнутий та повний огляд наукової літератури з тематики представленого дослідження, аналіз та застосування сучасних підходів до розв'язання задач дослідження свідчать про те, що дисертантка повністю володіє необхідною методологією наукового дослідження.

Наукове і практичне значення роботи.

Серед нових та важливих наукових результатів зазначимо наступне:

1. В роботі запропоновано нову метрику для знаходження міри подібності в просторі стаціонарних часових рядів, представлених моделями ARMA(p, q).

2. Зроблено порівняння запропонованої метрики та класичних метрик в просторі часових рядів та показано, що отримана метрика є більш стійкою до викидів і дає більш точні результати для часових рядів з великою кількістю вимірювань.

3. Запропоновано новий метод визначення оптимальної кількості кластерів при розгляді задач кластеризації об'єктів, що задаються неструктурованими даними (графами, часовими рядами тощо) на основі спектрального аналізу стохастичної матриці даних.

4. За допомогою симуляції методом Монте-Карло, показано, що запропонований метод підбору кількості кластерів дає кращі результати для визначення оптимальної кількості кластерів у порівнянні з рядом класичних методів.

5. Встановлено, що розроблений алгоритм знаходження оптимальної кількості кластерів є менш чутливим до наявності кластерів різного розміру.

Результати, отримані у дисертаційному дослідженні, можуть бути використані при кластеризації різних типів даних, зокрема в дисертаційній роботі наведено далеко не вичерпні можливі варіанти застосування у галузях медицини, соціології, економіки, тощо. Усі ці приклади підкреслюють важливість кластеризації даних у великій кількості різноманітних галузей, де вона допомагає у аналізі і використанні складноструктурованих великих наборів даних для прийняття рішень, підвищення ефективності та досягнення більшого розуміння ключових питань.

Оцінка змісту дисертації, її завершеність. Кваліфікаційна робота Кнігніцької Т.В. має загальний обсяг 150 сторінок машинописного тексту та структурно складається з анотації (українською та англійською мовами), списку опублікованих праць автора, переліку умовних позначень, вступу, трьох розділів, списку використаних джерел (122 позиції) та додатку (список публікацій здобувачки за темою дисертації).

У вступі обґрунтовано актуальність теми дослідження, сформульовано мету, завдання, предмет, об'єкт та методи дослідження, вказано наукову новизну, практичне значення отриманих результатів, зв'язок роботи з науковими дослідженнями та особистий внесок здобувачки, а також наведено дані про те, де доповідались, обговорювались та були опубліковані основні результати дисертації.

У першому розділі здійснено огляд наукової літератури, присвяченої дослідженню часових рядів, зокрема, визначенню метрик подібності між

часовими рядами та підходи до кластеризації даних, які представлені у вигляді неструктурованих типів даних. Детально проаналізовано хронологію розвитку наукових підходів до задач кластеризації, класифікації, зменшення розмірності часових рядів. Перший пункт розділу 1 відображає загальний огляд розвитку наукових досліджень при дослідженні часових рядів та існуючі метрики для встановлення подібності між часовими рядами. У другому пункті наведено методи дослідження структурних стрибків у часових рядах. У третьому пункті зроблено огляд наукових досліджень, які стосуються неперервних часових рядів. Вибір оптимальної кількості кластерів при поділі даних на групи представлено у пункті четвертому.

У другому розділі запропоновано визначати подібність або відстань між часовими рядами за допомогою моделей часових рядів. Запропонований алгоритм для встановлення подібності двох наборів даних використовує параметри моделі, а не самі вимірювання. У якості моделей часових рядів розглянуто стаціонарні *ARMA* моделі. Отриманий алгоритм порівнюється з уже існуючими метриками знаходження відстаней у випадку збільшення вимірювань часового ряду та у випадку зростання кількості викидів у вхідному часовому ряді. Отриманий алгоритм має меншу обчислювальну складність, ніж алгоритми Евкліда, *DTW* та *ERP*. Запропоновану відстань можна використовувати для кластеризації сильно зашумлених даних.

Наукову новизну висновків, зроблених на основі отриманих у другому розділі результатів, розкривають такі положення:

- Описано алгоритм для знаходження відстані між часовими рядами на основі моделей часових рядів. Отримана відстань є більш стійкою до викидів у часових рядах. У випадку збільшення кількості викидів запропонований у дисертаційному дослідженні алгоритм дає кращі результати (відносна похибка зростає логарифмічно), ніж аналогічні

алгоритми (Евклідова відстань, ERP, DTW) для знаходження відстані між часовими рядами (відносна похибка зростає лінійно).

- Запропонований метод знаходження відстані між вимірюваннями часового ряду дає кращі результати для великих часових рядів, коли кількість вимірювань $T > 1000$. До того ж обчислювальна складність отриманого алгоритму є меншою за обчислювальну складність уже існуючих алгоритмів.

У третьому розділі розглянуто проблему кластеризації на графах на основі власних значень стохастичної матриці графа. Доведено, що власні значення стохастичної матриці для великих графів ($N > 100$) поділяються на три групи, одна із яких є визначальною для числа кластерів у графі. Використовуючи теорію випадкових матриць, вдалося показати, що асимптотичний розподіл підгрупи дійсних частин власних значень стохастичної матриці графу описується напівколовим розподілом Вігнера. Використання стохастичних матриць дало змогу точно локалізувати власні значення, що відповідають за кількість кластерів, чого не вдавалося зробити для матриць суміжності. Основні припущення моделі пов'язані з властивостями дискретних ланцюгів Маркова, що дозволяє розширити область застосування отриманих результатів на більш широкий клас об'єктів. Теоретичні результати перевірені на кластеризації часових рядів, що описують вартості $N = 470$ акцій S&P500 в період з 2013 до 2018 року.

Наукову новизну висновків, зроблених на основі отриманих у третьому розділі результатів, розкривають такі положення:

- У роботі запропоновано новий метод визначення оптимальної кількості кластерів k_{opt} при кластеризації об'єктів, що задаються неструктурованими даними (графами та часовими рядами) на основі спектрального аналізу стохастичної матриці даного графу.

- Використовуючи метод Монте-Карло, вдалося показати, що запропонований метод дає кращі результати для визначення оптимальної кількості кластерів k_{opt} у порівнянні із деякими класичними методами.
- Оскільки запропонований алгоритм є спектральним, то його складність збігається зі складністю знаходження власних значень для стохастичної матриці P .
- Описаний алгоритм не є чутливим до кластерів різного розміру, тобто співвідношення між розмірами кластерів практично не впливають на точність алгоритму.
- Теоретичні результати роботи перевірено на реальних даних $N = 470$ акцій $S \& P500$, розглянутих в період з 2013 до 2018 року. Результати оцінки оптимального значення k_{opt} збіглися із відповідними оцінками для даних компаній в інший період часу.

Дисертаційні положення та побажання щодо вдосконалення змісту дисертації. Загалом позитивно оцінюючи отримані результати дослідження, їх наукову новизну та практичну значущість, водночас, вважаю за доцільне звернути увагу на деякі недоліки:

- 1) Робота містить суттєву кількість одруківок, невідформатованих елементів тексту та невдалих, як на мене, словосполучень, зокрема на с. 18 «рішення групування про об'єктів» на с. 82 «нуьовим математичним сподіванням»; на с. 57 текст на полях сторінки, на с. 98 посилання на літературу не відкомпільовано в тексті, на с. 112 текст англійською мовою; на с. 78 «процес приймає вигляд» варто було б замінити на «процес набуває вигляду», на с. 82 «до цих пір» варто було б замінити на «до цього моменту», на с. 118 «за ймовірністю має місце збіжність» варто замінити на «має місце збіжність за ймовірністю»; тощо.

- 2) У роботі використовуються поняття, які не були означені і, хоча, вони є загальноживаними, це ускладнює читання та розуміння тексту, особливо для вузькопрофільних спеціалістів. Наприклад, на с. 74 згадано «білий шум» без означення цього процесу, на с. 77 стверджується, що «розв'язок характеристичного рівняння буде стійким» без пояснення в якому сенсі розуміти стійкість.
- 3) На с. 78 стверджується що значення білого шуму не корелюють між собою, що не дуже зрозуміло читачу без пояснення природи цього процесу.
- 4) На с. 83 згадано «процес броунівського руху» у контексті випадкового блукання. Не зрозуміло чи має на увазі авторка вінерівський процес, чи якусь іншу модель броунівського руху, як от, наприклад, модель Каца?
- 5) На мою думку, згаданих вище зауважень можна було б уникнути за рахунок зменшення оглядової частини, яка здається занадто деталізованою. Натомість, в основній частині варто було б більш детально описати деякі поняття та властивості, які використовуються без належних пояснень.

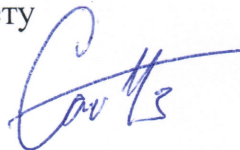
Відсутність порушень академічної доброчесності. Кваліфікаційне дослідження є самостійною науковою працею автора. Висновки, рекомендації та пропозиції, що характеризують наукову новизну кваліфікаційного дослідження, одержані автором особисто. При використанні праць інших вчених для аргументації актуальних положень дослідження обов'язково вказано посилання на відповідні праці.

Загальний висновок. Кваліфікаційна наукова робота Книгніцької Тетяни Василівни «Оцінки параметрів авто регресійних моделей» за актуальністю, науковою новизною, загальним переліком отриманих результатів, а також їх взаємозв'язком та повнотою їх викладу в журнальних публікаціях та апробацією 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 авторка кваліфікаційної наукової роботи Книгніцька Тетяна Василівна заслуговує присудження їй наукового ступеня доктора філософії з галузі знань 11 Математика та статистика за спеціальністю 113 Прикладна математика.

Офіційний опонент:

доктор фізико-математичних наук,
професор кафедри дослідження операцій
факультету комп'ютерних наук та кібернетики
Київського національного університету
імені Тараса Шевченка

 Ігор САМОЙЛЕНКО

СІЛСЬКО-ГОСПОДАРСЬКИЙ
УНІВЕРСИТЕТ ІМЕНІ
ТАРАСА ШЕВЧЕНКА
4.3.11.2023

