

РЕЦЕНЗІЯ

**доктора технічних наук, доцента,
доцента кафедри комп'ютерних наук
навчально-наукового інституту фізико-технічних та комп'ютерних наук
Чернівецького національного університету імені Юрія Федьковича**

УГРИНА Дмитра Ілліча

**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 дисертаційного дослідження

У наш час штучний інтелект уже допомагає в медицині, передбаченні стихійних лих, підборі контенту в соцмережах та створенні протезів. Машини, обладнані штучним інтелектом, можуть замінити людину у виконанні небезпечних та складних завдань, де необхідна велика сила, швидка увага, або обробка великої кількості схожої інформації. Розвиток інформаційних технологій має на меті створити засоби, що будуть полегшувати життя: удосконалювати виробництво, урізноманітнювати дозвілля, розвивати медицину, тощо. Усі проблеми розділяються на певні задачі, кожна з яких беруться вирішувати науковці за допомогою різноманітних засобів. За досить популярною у наш час назвою «штучний інтелект» стоять ґрунтовні математичні обчислення та статистичні підходи з використанням великих обсягів даних. Будь-який робот чи машина працює на основі висновків, які він отримує від обробки даних, які йому доступні. Тому методи обробки даних удосконалюються щодня.

Найчастіше, дані є неструктурованими – часовими рядами або графами. Кваліфікаційна наукова праця Книгніцької Тетяни Василівни присвячена розгляду неструктурованих типів даних, знаходженню подібності в даних та визначенню оптимальної кількості кластерів при розбитті вхідних даних на групи. Задача кластеризації даних представляє машинне навчання без учителя та є дуже актуальним напрямком дослідження у даний час. Кластеризацію застосовують при розсилці реклами, обробці зображень, визначенні атак на банківські операції, встановленні діагнозів у медицині тощо. Дисертація Книгніцької Т.В. описує ідеї щодо розв'язання двох основних задач кластеризації даних – визначення міри подібності між даними та встановлення оптимальної кількості кластерів для вхідних даних. Таким чином, кваліфікаційне дослідження Книгніцької Т.В., безперечно, є актуальним.

Зв'язок роботи з державними програмами, планами, темами

Кваліфікаційна наукова праця Книгніцької Т.В. виконана на кафедрі прикладної математики та інформаційних технологій Чернівецького

національного університету імені Юрія Федьковича у рамках науково-дослідної теми кафедри: «Математичне моделювання і числово-аналітичні методи дослідження динамічних та інформаційних процесів» (номер державної реєстрації: 0102U006591). Тему дисертації Кнігніцької Т.В. затверджено Науково-технічною радою Чернівецького національного університету імені Юрія Федьковича (протокол № 10 від 14 грудня 2016 року).

Ступінь обґрунтованості наукових положень, висновків, рекомендацій, сформульованих у дисертації

Наукові положення, висновки та результати дисертації обґрунтовані за допомогою загальноприйнятих методів аналізу часових рядів та теорії стохастичних випадкових матриць графу для розв'язання визначених завдань. У роботі ідеї та обчислення подані послідовно та логічно. Велика кількість посилань на наукову літературу дозволяє читачу легко здобути додаткову інформацію про використані математичні та статистичні підходи при розв'язуванні поставлених у дисертації задач. Кваліфікаційне дослідження Кнігніцької Т.В. включає приклади застосування та порівняння отриманих результатів з класичними підходами на реальних даних. Робота містить багато рисунків, які покращують розуміння та сприйняття отриманих висновків. Результати дисертаційного дослідження, застосовані до реальних даних, узгоджуються із результатами інших науковців, які використовували ті ж дані, але інші підходи до кластеризації даних. Кваліфікаційне дослідження було представлено на 8 наукових конференціях із публікацією тез чи матеріалів доповідей. Основні результати дослідження були опубліковані в трьох наукових статтях, дві з яких індексуються в базах даних Scopus та/або Web of Science. Додатково, результати дисертації відображено у чотирьох наукових статтях.

Структура дисертації

Дисертація складається з вступу, трьох розділів, висновків до кожного з розділів, загальних висновків, списку використаних джерел, який налічує 122 позиції, та списку публікацій здобувачки за темою дисертації (7 статей та 8 тез конференцій). Загальний обсяг роботи – 150 сторінок. Основні результати дисертації у повній мірі відображені у публікаціях. У першому розділі наведено огляд наукової літератури за темою дисертаційного дослідження. Алгоритм визначення міри подібності між часовими рядами описано у розділі другому. Третій розділ містить розроблений алгоритм для визначення оптимальної кількості кластерів, який ґрунтується на розгляді стохастичної матриці графу.

Наукова новизна

Результати і висновки дисертації, які становлять наукову новизну, включають наступні аспекти:

- Описано алгоритм для встановлення подібності між стаціонарними часовими рядами, які описуються моделями $ARMA(p, q)$. Отримана метрика використовує відстань між параметрами моделей $ARMA(p, q)$, а не між самими вимірюваннями часових рядів.
- Здійснено порівняння запропонованої метрики з класичними підходами (Евклідова відстань, DTW, ERP).
- Показано, що описана дисертанткою метрика є більш стійкою до викидів у часових рядах і дає більш точні результати для часових рядів з великою кількістю вимірювань.
- Встановлено, що складність алгоритму обчислення, запропонованого дисертанткою, при розгляді N часових рядів складає $O(T * N^2)$, в той же час аналогічна складність алгоритмів DTW, ERP становить $O(T^2 N^2)$. Таким чином за допомогою отриманої метрики отримуються більш стійкі до шумів кластери.
- Показано, що відносна похибка вимірювань для описаного алгоритму встановлення міри подібності між часовими рядами на основі моделей зростає логарифмічно. Відносна похибка вимірювань для класичних методів (Евклідова відстань, DTW, ERP) зростає лінійно. Крім того, при розгляді довгих часових рядів (коли кількість вимірювань часового ряду зростає, $T > 1000$) відносна похибка вимірювань для запропонованого методу спадає, а для класичних методів – залишається сталою величиною.
- Запропоновано новий метод визначення оптимальної кількості кластерів, який базується на проведенні спектрального аналізу стохастичної матриці графу. Власні значення стохастичної матриці графу поділяються на три групи, одна з яких і є визначальною для оптимальної кількості кластерів.
- За допомогою симуляції методом Монте-Карло, показано, що запропонований метод дає кращі результати для визначення оптимальної кількості кластерів у порівнянні з рядом класичних методів (Марковський алгоритм з двома типами параметрів та метод ліктя). Симуляція Монте-Карло використана для утворення багатовимірних даних – графу з фіксованою кількістю сукупностей (кластерів). Таким чином для порівняння запропонованого методу вибору оптимальної кількості кластерів з Марковським алгоритмом та методом ліктя дисертантка наперед володіла інформацією про оптимальну кількість кластерів.
- Встановлено, що розроблений алгоритм, який використовує стохастичну матрицю графу для визначення оптимальної кількості кластерів є менш чутливим до наявності кластерів різного розміру.

Практичне значення одержаних результатів

Дисертаційне дослідження Кнігніцької Т.В. має прикладний характер. Ідеї, описані в роботі, можуть бути застосовані для даних з будь-якої сфери життєдіяльності людини. Наприклад, у фінансовій сфері кластеризація часових рядів допомагає в розпізнанні закономірностей та трендів на фондових ринках. У медицині, цей підхід використовується для аналізу пацієнтських даних, таких як вимірювання параметрів здоров'я у часі, для виявлення паттернів та класифікації хвороб. В індустрії виробництва та технічному обслуговуванні, кластеризація часових рядів дозволяє виявляти аномалії та передбачати відмови обладнання, що сприяє плануванню обслуговування та уникненню аварій. У сфері електроенергетики, кластеризація часових рядів використовується для прогнозування витрат енергії та оптимізації режимів роботи енергосистем.

Кластеризація часових рядів також застосовується у транспорті для аналізу руху транспортних засобів та оптимізації трафіку. В сфері наукових досліджень – для класифікації та аналізу великих обсягів часових даних, зокрема в метеорології та астрономії. Загалом, кластеризація часових рядів стає невід'ємною частиною аналізу динамічних процесів у різних галузях, сприяючи розумінню та використанню складних взаємодій у часі.

Зауваження

1. Наукову новизну одержаних результатів слід формулювати відповідно класичного подання, а саме замість «Запропоновано», «Здійснено порівняння», «Встановлено», «Показано», «За допомогою» необхідно було «Вперше розроблено», «Отримано подальший розвиток», «Вдосконалено».

2. У першому розділі багато загальновідомого матеріалу, також було б добре представити матеріал використовуючи математичні моделі.

3. Пункт дисертації 2.1 треба було б перенести до першого розділу, так як має оглядовий характер.

4. У роботі не досліджено випадки при наявності великих викидів у часовому ряду для Евклідової відстані, відстані по моделях, DTW і ERP методів.

5. У різних представлених моделях одні й ті самі терміни використовуються різними позначеннями.

6. Для рисунків 6-8 не подано за яким принципом підбиралися параметри.

7. В дисертаційні роботі вказано, що проведені дослідження можуть використовуватись для різних галузей, зокрема медицини, рекламного бізнесу, економіки, соціальних та психологічних досліджень. Та в роботі представлені дослідження проведені на основі теоретичних результатів перевірки на кластеризації часових рядів, що описують вартості $N = 470$ акцій S&P500.

8. Є окремі неточності в оформленні роботи та зустрічаються деякі граматичні помилки.

Висловлені зауваження не применшують наукові досягнення дисертації Кнігніцької Т.В. та не впливають на висновок про належний рівень дослідження і його загальну позитивну оцінку.

Загальний висновок

Дисертаційна робота Кнігніцької Тетяни Василівни «Оцінки параметрів авторегресійних моделей», подана на здобуття наукового ступеня доктора філософії з галузі знань 11 Математика та статистика за спеціальністю 113 Прикладна математика за її актуальністю, науково-теоретичним рівнем, новизною розв'язання поставлених завдань, практичним значенням отриманих результатів, а також їх апробацією **цілком відповідає** пунктам 6, 7, 8, 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 №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 Кнігніцька Тетяна Василівна заслуговує на присудження їй наукового ступеня доктора філософії з галузі знань 11 Математика та статистика за спеціальністю 113 Прикладна математика.**

Рецензент

доктор технічних наук, доцент,
доцент кафедри комп'ютерних наук
навчально-наукового інституту
фізико-технічних та комп'ютерних наук
Чернівецького національного університету
імені Юрія Федьковича

Дмитро УГРИН

Підпис *Угрин Д.* засвідчую
Учений секретар Чернівецького національного
університету імені Юрія Федьковича
Лисовська Н. С.
" 27 листопада 2022 р.

