

## ВІДГУК

офіційного опонента на дисертаційну роботу **Бабія Андрія Степановича** на тему **«Моделі, методи та інтелектуальна інформаційна технологія аналізу неоднорідних послідовностей»**, подану на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології.

**Актуальність теми виконаних досліджень.** Дисертаційна робота Бабія Андрія Степановича є науково прикладним дослідженням, метою якого є розробка моделей, методів та інтелектуальної інформаційної технології аналізу неоднорідних послідовностей даних для підвищення точності та ефективності оцінювання характеристик об'єктів для різних предметних областей в інформаційно-аналітичних системах.

Збільшення об'ємів та неоднорідність даних, які доступні для обробки із застосуванням інформаційно-аналітичних систем ставить підвищені вимоги до їх проектування. Можливості інформаційних технологій дозволяють не лише збільшувати кількість вимірювань і зберігати великі масиви однорідних даних, але і значно розширяти типи та кількість джерел, з яких надходить інформація. У тому числі поширені випадки, коли обробляються набори коротких вибірок даних, які поєднують в собі як дані отримані за допомогою натурального експерименту, так і дані експертних опитувань.

Дослідження проблеми комбінування даних, отриманих із різнорідних джерел, є актуальним напрямком наукових досліджень.

Сучасні реалізації інформаційно-аналітичних систем та інструментальні засоби інтелектуального аналізу даних також надають можливості для обробки інформації, отриманої із різних джерел, зокрема в межах стохастичного та інтервального підходів. Разом з тим методи та засоби обробки даних, які представлено у вигляді нечітких значень потребують додаткових наукових досліджень.

Тому тема та поставлені задачі дисертаційного дослідження, а саме розробка методів та удосконалення інформаційної технології аналізу різнорідних даних, частину з яких представлено у вигляді кількісних значень, отриманих в результаті натурального експерименту, а друга частина має «суб'єктивну» та нечітку природу, є **актуальними**.

Актуальність задачі, важливість та перспективність отриманих результатів підтверджується також тим, що робота виконувалася відповідно до наукового напрямку кафедри програмної інженерії Харківського національного університету радіоелектроніки «Теорія методи і моделі управління життєвим циклом інтелектуальних інформаційних середовищ регіональних соціо-економічних об'єктів» (номер державної реєстрації № ДР 0115U002430) в рамках наукового напрямку "Технології та засоби розробки програмних продуктів і систем", визначеного Кабінетом Міністрів України в переліку пріоритетних тематичних напрямів наукових досліджень і науково-технічних розробок на період до 2015 року.



**Оцінка змісту дисертації, її завершеності та оформлення.** Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел, додатку і за структурою та обсягом відповідає вимогам щодо кандидатських дисертацій з технічних наук.

У вступі обґрунтовано актуальність обраної теми дисертації, визначені мета і задачі дослідження, сформульовано об'єкт і предмет, а також методи дослідження, викладено наукову новизну та практичне значення отриманих результатів, надано інформацію про особистий внесок автора в роботи, виконані у співавторстві, наданий перелік публікацій та відомості про апробацію за темою роботи.

У першому розділі проаналізовано існуючі моделі, методи та інформаційні технології обробки великих масивів даних з різномірною природою походження, викладені в наукових публікаціях. Досліджено процес аналізу даних, представлених у вигляді неоднорідних послідовностей та розглянуті переваги і недоліки основних моделей і методів, за допомогою яких здійснюють комбінування даних. На основі існуючих публікацій за темою дослідження обґрунтована потреба в розробленні нових та розвитку існуючих моделей і методів аналізу неоднорідних послідовностей.

*Разом з тим, доцільно було б доповнити цей розділ аналізом методів структурної ідентифікації регресійних та інтервальних моделей (при побудові останніх використовують прості схеми перевірки значущості чинників в регресії), який необхідний для обґрунтування методів визначення значущих чинників моделі. Варто також було б більшу увагу звернути на особливості існуючих інтелектуальних інформаційних технологій аналізу неоднорідних даних.*

Другий розділ присвячений визначенню та обґрунтуванню показників, які пропонується застосовувати для аналізу неоднорідних послідовностей та моделей для їх представлення. У розділі розглянуто неоднорідні послідовності, які формуються із повідомлень про появу події чи результатів запитів з інформаційних систем обліку, що отримуються нерівномірно у часі. Для їх аналізу обґрунтовано застосування статистичних показників. Пропонується застосовувати спосіб розрахунку коефіцієнту кореляції із врахуванням наявної різниці в дисперсіях та математичних сподіваннях різних груп в середині однієї вибірки із застереженнями, в тому числі у випадку негомогенності кореляційного зв'язку і застосування в умовах «парадоксу об'єднання» двох (різномірних) груп даних (парадокс Сімпсона). Отримала подальший розвиток тренд-сезонна модель неоднорідних послідовностей, а саме для запобігання втраті крайових значень для коротких рядів даних, запропоновано подання трендової складової у вигляді інтерпольованих усереднених значень із врахуванням функції належності, яка асоційована із кожним нечітким розділом.

*У загальному, розділ викладено логічно, проте, доцільно було б частину цього розділу долучити до оглядової частини дисертації, зокрема, п.2.1. Потрібно також було б надати більш старанне обґрунтування критеріїв ефективності оцінювання та моделювання характеристик об'єктів для різних*

*предметних областей, зокрема, для відбору найбільш значущих чинників нечіткої регресії. Слід також було б окреслити недоліки використаного в роботі методу для обчислення нечітких коефіцієнтів лінійної регресії і навести схеми визначення порогового рівня коефіцієнту чіткості  $h$ .*

**Третій розділ** присвячено розробці методів аналізу неоднорідних послідовностей. Розглянуто підхід для узгодження експертних оцінок на основі попарних порівнянь, розроблені метод визначення значимих чинників нечіткої регресійної моделі та метод фільтрації компонент неоднорідних послідовностей. Для подання динамічного ряду із врахуванням експертних оцінок показників динамічного ряду пропонується використовувати підхід на основі непрямого методу побудови функцій належності. Для визначення значущих чинників нечіткої регресійної моделі неоднорідних послідовностей даних запропоновано метод, за допомогою якого здійснюється підбір коефіцієнтів за критерієм рівнозначності кутів відхилення між вектором похибки і векторами змінних. Підмножина значущих чинників обирається з усієї множини факторних змінних із застосуванням порогового значення. Для фільтрації компонент тренд-сезонної моделі неоднорідних послідовностей пропонується модифікація методу виділення трендової складової, яка полягає в розрахунку усереднених значень із врахуванням функції належності для кожного із нечітких розділів, на які розбивається динамічний ряд.

*Цей розділ є центральним в дисертаційній роботі. Проте його матеріал містить ряд недоліків. Зокрема, недостатньо обґрунтовано використання критерію Маллоуза для визначення значущості чинників (вхідних змінних) нечіткої регресійної моделі. Не ясно, чому автор не використав прості процедури інтервального аналізу для визначення значущості коефіцієнтів нечіткої регресії, адже, як відомо, інтервальній моделі є узагальненим випадком - нечітких. Складною процедурою в розглянутому методі згладжування трендової складової моделі за допомогою  $F$ -перетворення є визначення нечітких розділів послідовності. Доцільно було б обґрунтувати інтервали розбиття.*

**Четвертий розділ** присвячено розробці інтелектуальної інформаційної технології аналізу неоднорідних послідовностей. У цьому розділі запропоновано та обґрунтовано етапи інтелектуальної інформаційної технології, які практично реалізовані у вигляді модуля обробки неоднорідних послідовностей даних із використанням моделей і методів розроблених в дисертаційній роботі. Розроблену інтелектуальну інформаційну технологію запропоновано використовувати в рамках одного із поширених підходів до аналізу даних: KDD, CRISP-DM або SEMMA. Пропонується отримувати дані неоднорідних послідовностей з різних типів сховищ даних та файлових сховищ, які зберігають структуровані відомості. Практична реалізація виконана за допомогою мови програмування Python. Також у цьому розділі наведено результати аналізу неоднорідних послідовностей на прикладах відомостей про кількість скоєних правопорушень та риноманометричних даних оцінювання носового дихання. Результати застосування удосконаленої технології

засвідчують підвищення ефективності оцінювання характеристик об'єктів різних предметних областей.

*Матеріали цього розділу в основному підтверджують достовірність теоретичних результатів, викладених у попередніх розділах.*

У висновках стисло відображені основні результати, отримані в роботі.

У додатках наведені акти про впровадження результатів дисертаційної роботи.

### **Наукова новизна одержаних результатів.**

**Вперше** запропоновано та обґрунтовано метод визначення значущих чинників нечіткої регресійної моделі на основі неоднорідних послідовностей даних, який, на відміну від існуючих, містить ітераційну процедуру налаштування коефіцієнтів за критерієм рівнозначності кутів відхилення між вектором похибки і векторами змінних та редукції незначущих чинників за критерієм Маллоуза, що дозволяє оптимізувати структуру нечіткої лінійної регресії та її прогностичні властивості.

**Удосконалено** метод фільтрації компонент неоднорідних часових послідовностей, в якому, на відміну від існуючих, для виявлення тренду початкова послідовність ітеративно розбивається на скінчену кількість нечітких розділів, для кожного з яких розраховується середнє зважене значення із врахуванням функції належності, яка асоційована із нечітким розділенням, а також застосовується фільтрація коливань різних періодів, що дозволяє підвищити ефективність оцінювання характеристик об'єктів заданої предметної області.

### **Отримали подальший розвиток:**

- тренд-сезонна модель неоднорідних послідовностей, в якій, на відміну від існуючих моделей, трендова складова подається у вигляді інтерпольованих середньо-зважених значень із врахуванням функції належності, яка асоційована із кожним нечітким розділенням, що дозволяє застосовувати дану модель для коротких вибірок без втрати крайових значень і тим самим підвищити ефективність моделювання характеристик об'єктів різних предметних областей в інформаційно-аналітичних системах;

- інтелектуальна інформаційна технологія аналізу неоднорідних послідовностей даних, яка, на відміну від існуючих, містить методи та засоби визначення значущих чинників нечіткої регресійної моделі та фільтрації компонент часових рядів даних, що дозволяє підвищити ефективність аналізу неоднорідних послідовностей.

**Значущість отриманих результатів для науки і практичного використання** полягає у розвитку моделей, методів та інтелектуальної інформаційної технології аналізу неоднорідних послідовностей даних у випадку представлення цих даних випадковими, нечіткими та інтервальними послідовностями. Практичне використання розроблених теоретичних положень та створеної інформаційної технології забезпечує підвищення точності та

ефективності оцінювання і моделювання характеристик об'єктів для різних предметних областей в інформаційно-аналітичних системах.

**Ступінь обґрунтованості наукових положень, висновків і рекомендацій.** Основні наукові результати дисертаційного дослідження в цілому обґрунтовані теоретично та підтверджені впровадженням розроблених методів для розв'язування прикладних задач.

Обґрунтованість отриманих наукових положень, висновків і рекомендацій забезпечується коректним використанням перевірених теоретичних та практичних засобів наукового дослідження, зокрема: методів теорії нечітких множин, математичної статистики, регресійного та нечіткого-регресійного аналізу, аналізу динамічних рядів.

Отримані теоретичні результати частково зіставлені з відомими в літературних джерелах, а результати моделювання з даними експерименту та експертних опитувань.

У цілому наукові положення і висновки, сформульовані в дисертаційній роботі, є обґрунтованими теоретично та підтверджені практичним впровадженням.

**Достовірність результатів досліджень.** Достовірність результатів дисертаційного дослідження забезпечено теоретичним обґрунтуванням зроблених припущень, практичним впровадженням результатів досліджень та результатами їх аналізу. Для підтвердження достовірності отриманих показників ефективності запропонованих методів та інформаційної технології розглянуто ряд прикладів та експериментальних досліджень. Зокрема, побудовано адекватні тренд-сезонну модель рівнів злочинності та модель показників носового дихання. Достовірність результатів роботи також підтверджено на багаточисленних науково-практичних конференціях.

**Практичне значення роботи.** Результати дисертаційного дослідження дають змогу підвищити ефективність аналізу даних, представлених у вигляді неоднорідних послідовностей. Практична цінність роботи полягає у створенні та застосуванні ряду алгоритмів та програмних засобів:

- для визначення значущих чинників нечіткої регресійної моделі із етапом підбору коефіцієнтів за критерієм рівнозначності кутів відхилення між вектором похибки і векторами змінних та відбору підмножини значущих чинників з коефіцієнтами, що перевищують порогове значення;

- для фільтрації компонент неоднорідних часових послідовностей за допомогою модифікованого методу на основі F-перетворення;

- модуля аналізу неоднорідних даних для оцінювання характеристик об'єктів, який реалізує етапи запропонованої інформаційної технології.

Отримані практичні результати використано в ході виконання практичної частини науково-дослідних робіт за темою «Теорія, методи і моделі управління

життєвим циклом інтелектуальних інформаційних середовищ регіональних соціо-економічних об'єктів» (№ ДР 0115U002430).

На основі результатів дисертаційних досліджень розроблений програмний засіб для аналізу неоднорідних послідовностей, який впроваджено:

- в діяльність ТОВ «Ендейвер», що дозволило зменшити рівень похибки оцінювання характеристик виконаної роботи бригад;

- в діяльність управління інформаційної підтримки та координації поліції 102 ГУНП в Харківській області по аналізу динамічних рядів показників злочинності.

Результати, отримані у ході дисертаційних досліджень, було впроваджено в організацію навчального процесу на кафедрі програмної інженерії Харківського національного університету радіоелектроніки при проведенні лекційних занять та лабораторних робіт з курсу «Теорія ймовірностей і математична статистика» та «Емпіричні методи програмної інженерії»

Усі впровадження результатів роботи підтверджено відповідними актами.

### **Повнота викладення результатів в опублікованих матеріалах.**

Основні наукові результати дисертації відображені в 16-ти наукових працях (серед них 3 – одноосібних), у тому числі 7 статей у наукових фахових виданнях України з технічних наук і 1 стаття – за кордоном у виданні, що входить до міжнародної наукометричної бази Scopus, 8 тез у матеріалах міжнародних конференцій (з них 3 включено до наукометричної бази Scopus.

В опублікованих працях викладено в повному обсязі основні положення дисертаційної роботи, які винесено на захист. Особистий внесок здобувача в сумісних публікаціях є підтвердженим. Рівень та кількість публікацій, рівень апробації відповідають вимогам, що ставляться до кандидатських дисертацій в Україні.

**Автореферат** відображає суть основних наукових положень, практичну значущість та висновки. Дисертаційна робота та автореферат оформлені у відповідності до вимог, що ставляться до кандидатських дисертацій в Україні, хоча по тексту дисертації зустрічаються опечатки та стилістичні неточності

### **Недоліки та зауваження до дисертаційної роботи**

1. У першому розділі відсутній аналіз методів структурної ідентифікації регресійних моделей, який необхідний для обґрунтування методів визначення значущих чинників моделі. Відсутній також аналіз методів побудови регресії у випадку інтервальних даних, які відзначаються достатньо простою схемою перевірки значущості чинників в регресії і можуть бути застосовані для нечітких регресійних моделей.

2. У другому розділі не достатньо обґрунтовано критерії ефективності оцінювання та моделювання характеристик об'єктів для різних предметних областей. Вважаю також, що частину цього розділу можна було б винести в оглядову частину дисертації, наприклад п.2.1.

3. Використаний в роботі метод Х. Танаки для обчислення нечітких коефіцієнтів нечіткої лінійної регресії на основі розв'язування задачі лінійного програмування має суттєвий недолік, оскільки вимагає емпіричного задання порогового рівня коефіцієнта чіткості  $h$ .

4. Недостатньо обґрунтовано використання коефіцієнта Маллоуза для порогового критерію визначення значущості чинників (вхідних змінних) нечіткої регресійної моделі, який, як відомо, включає складові для оцінки точності та структурної складності регресії для випадку нормально розподілених похибок у вихідних змінних. Взамін, автором за значення вихідної змінної використано центри нечітких величин з невідомими розподілами. Виникає запитання достовірності такого узагальнення? До того ж такий підхід неможливо застосувати у випадку нечітких вхідних змінних.

5. Недоліком розглянутого методу згладжування трендової складової моделі за допомогою F-перетворення є визначення нечітких розділів послідовності з регулярним розбиттям. Таке розбиття часто спотворює згладжуваний тренд. Доцільно було б обґрунтувати інтервали розбиття.

6. У третьому розділі п.3.1 розглянуто метод визначення коефіцієнтів нечіткості рівнів динамічного ряду, який ґрунтується на методі Сааті. Разом з тим, запропонований метод не використано для апроксимації трикутних функцій належності, а також для обґрунтування розділення динамічного ряду на нерегулярні нечіткі розбиття в задачі фільтрації компонент неоднорідних послідовностей.

7. У першому прикладі п. 4.3 (медичне застосування) в побудованій нечіткій регресійній моделі другий нечіткий коефіцієнт при змінній  $X_5$  представлений трикутною функцією належності з центром 0,001 та шириною 0,01. Тобто інтервал розмитості  $[0,001-0,01; 0,001+0,01]$  з достатньою ступінню належності включає «0». Чи не означає це, що чинник позначений  $X_5$  в цій регресії є незначущим?

8. В тексті дисертації і автореферату є окремі невідповідності та стилістичні неточності, а також опечатки. Наприклад, на ст. 39 написано: «викривленнях внесених лінеаризацією»; на ст. 93 написано: «чіткі фактори, обрані за допомогою обрані за допомогою запропонованого методу»; на ст. 108 написано: «Як правило, об'єм таких файлів порівняно невеликий, але досить розповсюджений».

Зазначені недоліки є не принциповими і суттєво не знижують загального позитивного враження від поданої роботи.

#### **Відповідність дисертаційної роботи обраній спеціальності.**

Дисертаційна робота відповідає формулі спеціальності та виконана у відповідності до наступних пунктів, зазначених в паспорті спеціальності 05.13.06 – інформаційні технології:

2) розроблення інформаційних технологій для аналізу та синтезу структурних, інформаційних і функціональних моделей об'єктів і процесів, що автоматизуються;

3) розроблення моделей і методів автоматизації виконання функцій і завдань виробничого й організаційного управління у звичайних і багаторівневих структурах на основі створення та використання нових інформаційних технологій;

10) моделювання предметних галузей інформаційних систем (аналітичне, імітаційне, інфологічне, об'єктно-орієнтоване, тощо) на підґрунті створення і застосування відповідних інформаційних технологій.

### Загальні висновки.

Оцінюючи роботу в цілому, вважаю, що дисертація Бабія А.С. на тему «Моделі, методи та інтелектуальна інформаційна технологія аналізу неоднорідних послідовностей» є завершеною науковою працею, у якій отримані нові наукові і практичні результати, що вирішують важливу науково-технічну задачу створення моделей, методів та інтелектуальної інформаційної технології аналізу неоднорідних послідовностей даних для підвищення точності та ефективності оцінювання і моделювання характеристик об'єктів для різних предметних областей в інформаційно-аналітичних системах.

Дисертаційна робота здобувача за своїм рівнем, обсягом і якістю досліджень відповідає вимогам пунктів 9, 11, 12-14 «Порядку присудження наукових ступенів», затвердженого постановою Кабінету Міністрів України від 24.07.2013 № 567 щодо кандидатських дисертацій, а її автор Бабій Андрій Степанович заслуговує на присудження йому наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології.

### ОФІЦІЙНИЙ ОПОНЕНТ:

Декан факультету комп'ютерних  
інформаційних технологій  
Тернопільського національного  
економічного університету,  
д.т.н., професор



*Дивак М. П.*

Дивак М. П.

Забірюю:

Зав. загальним відділом

*Дивак М. П.*

*Гарашова Г. В.*