

ВІДГУК

офіційного опонента на дисертаційну роботу
Терентьєва Олександра Миколайовича
**«МОДЕЛІ, МЕТОДИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ
ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ В
УМОВАХ НЕВИЗНАЧЕНОСТІ»**,
подану на здобуття наукового ступеня доктора технічних наук
за спеціальністю **05.13.06 – інформаційні технології**

Актуальність теми. Дисертаційна робота Терентьєва Олександра Миколайовича присвячена вирішенню важливої науково-практичної проблеми, що пов'язана із створенням єдиної науково обґрунтованої інформаційної технології, яка враховує широкий клас задач із розробки прогнозів високої якості, що дозволяють приймати ефективні управлінські рішення. Означений комплекс задач пов'язаний із таким сучасним напрямом з аналізу даних, як наука о даних (data science), що використовує загальні класичні та спеціалізовані методи та підходи інтелектуального аналізу (data-mining) як структурованих так і не структурованих даних. Будь-які процеси життєдіяльності людини (соціально-економічні, екологічні, технологічні, тощо) описуються потоками великих даних, що потребують відповідної спеціалізованої обробки та подальшої побудови аналітичних моделей, які в свою чергу є невід'ємним компонентом сучасних систем підтримки прийняття рішень, що, в поєднанні із засобами аналітичної обробки інформації, призводить до підвищення керованості систем та точності при прогнозуванні. Все це загалом свідчить про значну актуальність обраної тематики дослідження.

Однією із актуальних проблем будь-якого аналізу даних є проблема неповноти інформації, що виражається у наявності пропусків у вимірах, технічних проблемах при передачі інформації, від об'єкта до транзакційного сховища даних, у деяких випадках відсутність розуміння які саме фактори, та за якої послідовності впливають один на одного, та загалом на процес як єдину систему. Все це в комплексі призводить до появи статистично-ймовірносних та структурно-параметричних невизначеностей для боротьби із якими необхідно

розробляти нові та удосконалювати вже існуючі методи та методології системного аналізу для повноцінного всебічного аналізу та опису функціонування складних об'єктів та процесів. Неостанню роль в цьому процесі є розробка нових та коректне керування вже існуючими інформаційними середовищами загального та спеціалізованого призначення, базами знань та правил, антологіями, математичними моделями та підходами щодо їх оцінювання багатокритеріальними методами та підходами оцінювання.

Обраний напрям та тематика дисертаційної роботи пов'язані із задачами щодо аналізу та об'єднання великих масивів даних як із структурованих, так і не структурованих джерел, для опрацювання невизначеностей що виникають при моделюванні та прогнозуванні процесів, що містять нелінійні та нестационарні параметри та фактори впливу.

Автором дисертаційної роботи означене широке коло задач, що вимагають розв'язання, а саме опрацювання невизначеностей, побудова математичних моделей що враховують нестационарності і нелінійності в процесах різної природи, дослідження та використання ансамблевих та адаптивних моделей.

Літературний аналіз джерел, виконаний здобувачем, показав, що на сьогодні немає універсальної методики або готової інформаційної системи, в сфері аналізу та моделювання нелінійних нестационарних процесів, що дозволяє вирішувати всі зазначені проблеми.

Як показано здобувачем, значні перспективи для вирішення означеного кола проблем та задач, закладені саме в нових моделях і методах для прогнозування нелінійних нестационарних процесів, що використовують багатомодельні та багатокритеріальні підходи, на основі синтезу та інтеграції різного за типом та структурою інформації на засадах системного аналізу засобами інтелектуального аналізу даних.

Все зазначене вище визначає актуальність обраної тематики дослідження здобувачем та необхідність вирішення описаних проблем.

Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертаційній роботі

Представлені в дисертаційній роботі ідеї, теоретичні положення, робочі гіпотези, отримані результати та висновки представлені в належному зрозумілому вигляді. Вони представлені на основі виконаного значного огляду літературних джерел та матеріалів в електронному вигляді з мережі Інтернет, пов'язаних із напрямком дисертаційного дослідження; коректному визначені мети і формалізованому представлені задачі дослідження; високопрофесійному застосуванню сучасних методів та підходи інтелектуального аналізу даних, у тому числі, адаптивних, у поєднанні із статистичним та ймовірнісним моделюванням; інформаційних технологіях моделювання і прогнозування нелінійних нестационарних процесів, із урахуванням умов невизначеності, нестационарності та ризику на основі використання; представленні та порівнянні отриманих практичних та теоретичних результатів, різними методами аналізу та прогнозування даних, що було виконано здобувачем на високому сучасному рівні.

Структура, задачі та короткий зміст роботи

Робота складається із вступу, 7 розділів, списку використаних джерел та додатків. Загальний обсяг роботи складає 293 сторінки, з яких основного тексту 285 сторінок, 77 ілюстрацій, 58 таблиць, 12 додатків. Список використаних джерел налічує 274 найменування на 32 сторінках.

У **вступі** обґрунтовано актуальність теми дисертаційної роботи, описано проблему дослідження та її стан, надане формулювання мети та завдань дослідження, приведено опис науково-практичних результатів. Представлено дані щодо апробації та опублікування результатів за тематикою роботи, приведено список публікацій із виділенням особистого внеску здобувача.

У **першому розділі** зроблено аналіз проблем, що виникають при розробці аналітичного інструментарію, призначеного для створення сучасних

інформаційних технологій прийняття рішень, в основі яких лежить прогнозне моделювання нелінійних нестационарних процесів різної природи. Проведений аналіз показав, що наявні методи щодо розв'язання задач прогнозування розвитку нелінійних нестационарних процесів містять як переваги, так і недоліки. Саме тому виникає потреба у розробці нових та покращенні вже існуючих методів, алгоритмів, інформаційних технологій, призначених для автоматизації наведених процесів.

Приведено характеристику та класифікацію нелінійних нестационарних процесів. Окрема виділено процеси із наявністю нестационарності, інтегрованої складової, гетероскедастичності та коінтеграції. Показано, що на якість результатів прогнозування значно впливають повнота та достовірність вхідної інформації. Для вирішення проблеми заповнення пропусків даних, наведено та узагальнено класифікацію пропусків. Було проведено чисельні експерименти та запропоновано власну методику, що дозволяє опрацьовувати пропуски у часових рядах.

У **другому розділі** описано розроблену методику для попереднього опрацювання різнорідних даних, що надаються із різних інформаційних джерел. Ця методика базується на використанні статистичних методів аналізу, використанні різнотипної інформації, методах моделювання, прогнозування та прийнятті рішень на основі набору критеріїв.

Для виявлення зав'язків в даних, опрацювання екстремальних значень, трансформації та перетворення значень та фільтрації шумових складових, виявлення та усунення мультиколінеарності, наведено розроблену методику попереднього аналізу даних, що складається з п'яти етапів.

Для розв'язування задач прийняття рішень щодо управління розвитком нелінійних нестационарних процесів, з інтеграцією різнотипної інформації, запропоновано виконувати аналіз неструктурованої інформації, в тому числі розміщеної в мережі Інтернет, засобами SAS Textual Analytics. Використання засобів текстової аналітики, дає можливість найбільш повно сформулювати альтернативи та визначити цільову настанову.

У **третьому розділі** запропоновано новий метод на основі аналізу нелінійних нестационарних процесів за їх подібністю. Метод є робастним та непараметричним, і полягає у виявленні схожості статистичної поведінки процесів, що описуються часовими рядами даних. Для оцінювання ступеня близькості часових рядів, запропонована метрика обчислення міри подібності, що складається з чотирьох кроків – побудувати матрицю відстаней; налаштуванні обмежень на стиснення та розширення; побудова оптимального шляху; обчисленні значення функції вартості шляху.

Використання цього підходу дозволяє не тільки аналізувати але і враховувати невизначеності, що виникають у задачах прогнозування розвитку нелінійних нестационарних процесів. Метод дозволяє виявляти схожі шаблони поведінки, в рядах даних, що досліджуються, навіть у випадках наявності значної кількості пропусків та спотворених чи неповних даних.

У **четвертому розділі** розглядається сучасний аналітичний інструмент для побудови нелінійних нестационарних процесів у вигляді причино-наслідкових мереж Байеса. Відповідний математичний апарат дозволяє враховувати структурно-параметричні невизначеності та будувати набори сценаріїв щодо розвитку подій об'єкту або процесу аналізу.

Запропоновано методику знаходження параметрів байесівських мереж із прихованими вершинами, на основі алгоритму максимізації математичного очікування, що дозволяє обчислювати параметри мережі Байеса за умови неповної вхідної інформації і відомої топології мережі. Для побудови ймовірнісного висновку запропоновано алгоритм, що включає три кроки – адитивна декомпозиція мережі Байеса на окремі складові загальної мережі; вивід окремих підмножин вузлів основної моделі за L-S алгоритмом; обчислення спільного розподілу ймовірностей для кожної підмножини вузлів. Для оцінювання якості роботи пропонується використання квадратичної відстані Хеллінджера.

У **п'ятому розділі** описано системну методологію побудови моделей нелінійних нестационарних процесів, в основі якого лежить адаптивний підхід

до моделювання, в рамках якого комбінується сценарне моделювання, регресійний аналіз та ймовірнісно-статистичний математичний апарат у формі мереж Байєса. Для вирішення проблеми неповноти та недостовірності даних нелінійних нестационарних процесів запропоновано багатомодельний підхід, що включає використання теорії подібності процесів та ймовірнісно-статистичного моделювання. Якість прогнозів забезпечується автоматизованою процедурою вибору на основі інтегрального критерія якості, що складається від двох до п'яти окремих статистичних критеріїв якості.

У **шостому розділі** надано опис інформаційної технології моделювання і прогнозування нелінійних нестационарних процесів в умовах невизначеності, нестационарності та ризику. Вона включає використання сучасних методів інтелектуального аналізу даних та використовує методологію системного аналізу. Наведена аналітична платформа дозволяє інтегрувати як власні доробки фахівців у вигляді реалізованих модулів та бібліотек, так і вже існуючі рішення від поставників бізнес-аналітичних рішень. Запропонована архітектура системи підтримки прийняття рішень дозволяє будувати додатки користувача, з метою вирішення комплексу задач прогнозування нелінійних нестационарних процесів.

У **сьомому розділі** представлено декілька практичних прикладів застосування багатомодельного підходу щодо моделювання нелінійних нестационарних процесів, на основі запропонованої в роботі інформаційної технології у вигляді системи підтримки прийняття рішень.

Висновки за розділами і результати дисертації сформульовані з достатньою повнотою, чітко, зрозуміло і відповідно до змісту роботи.

Список використаних джерел є достатньо повним та включає як вітчизняні, так і закордонні публікації і складається з 274 найменування.

У **додатках** подані матеріали, що ілюструють практичні реалізації отриманих рішень, акти впровадження результатів дисертаційного дослідження, а також авторські свідоцтва на розроблені комп'ютерні програми.

Достовірність результатів досліджень

Обґрунтованість та достовірність наукових положень, висновків та рекомендацій, що відображені в дисертації, засновані на коректному використанні математичного апарату з таких сучасних напрямків науки про дані як статистичний аналіз, регресійний аналіз, теорія мереж Байєса, теорія подібності процесів, інтелектуальний аналіз даних та методи системного аналізу, а також багатьма результатами практичного використання для вирішення практичних задач, для яких відповідні результати обчислювальних експериментів наведені в роботі, в логічній, зрозумілій, прозорій формі, підтвержені відповідними актами впровадження в організаціях, а також у навчальному процесі на кафедрі математичних методів системного аналізу ПСА НТУУ «КПІ ім. Ігоря Сікорського».

Наукова новизна отриманих результатів

Наукова новизна одержаних автором результатів дослідження сформульована логічно, коректно і не викликає сумніву.

Вперше запропоновано і розроблено:

– розроблено новий метод опрацювання невизначеностей, який ґрунтується на застосуванні теорії подібності процесів, який відрізняється робастністю результатів аналізу, що забезпечує отримання оцінок прогнозів високої якості за наявності неповних або спотворених даних;

– запропоновано новий метод побудови регресійних та ймовірнісно-статистичних моделей у формі мереж Байєса, який відрізняється можливістю врахування нестационарності і нелінійності стосовно змінних, що забезпечує високу адекватність моделей і якість прогнозів процесів досліджуваного типу;

– розроблено метод моделювання нелінійних нестационарних процесів різної природи в умовах невизначеності, який відрізняється від відомих урахуванням різних типів невизначеностей, що підвищує адекватність моделей і якість оцінок прогнозів за лінійними та нелінійними моделями;

– побудовано та досліджено ансамблі моделей для формального опису

нелінійних нестационарних процесів, які відрізняються модифікованою комплексною структурою та високою адекватністю, що дозволяє підвищити якість оцінювання прогнозів розвитку досліджуваних процесів;

– запропоновано метод прогнозування, оснований на використанні адаптивного підходу до моделювання у поєднанні із статистичним та ймовірнісним моделюванням, що дає можливість урахувати структурно-параметричні невизначеності і забезпечує адекватний опис причинно-наслідкових зв'язків і можливих варіантів розвитку процесів різної природи під впливом груп внутрішніх та зовнішніх чинників;

– запропоновано нові моделі і методи створення інформаційних технологій розв'язування задач побудови математичних моделей для прогнозування нелінійних нестационарних процесів, які ґрунтуються на принципах багатомодельного та багатокритеріального підходів, інтеграції різнотипної інформації і засновані на системному використанні методів інтелектуального аналізу даних, ймовірнісно-статистичного моделювання, теорії подібності процесів, прогнозування і підтримки прийняття рішень, що підвищує обґрунтованість прийняття рішень в умовах наявності невизначеностей та ризиків різних типів;

– розроблено інформаційну технологію, в основу якої покладено поєднання принципів системного аналізу, методів обробки та оцінювання якості даних, прогнозного моделювання із використанням нових моделей та їх композицій, запропонованих критеріїв адекватності моделей, оцінок якості прогнозів, яка забезпечує високу якість проміжних та остаточних результатів дослідження нелінійних нестационарних процесів.

Набуло подальшого розвитку:

– метод аналізу інформації на основі засобів статистично-ймовірнісного моделювання, який дає змогу зменшити суб'єктивізм відбору найбільш значимих чинників в задачах прогнозування процесів різної природи, що створює передумови для прийняття обґрунтованих управлінських рішень;

– метод побудови моделей у формі байєсівської мережі для оцінювання

розвитку динаміки процесів різної природи, який відрізняється коректністю формального опису за байєсівським інформаційним критерієм, що забезпечує обчислення високоякісних ймовірнісних оцінок прогнозів розвитку досліджуваних процесів;

– системна методологія побудови адаптивних моделей процесів різної природи, яка відрізняється удосконаленням існуючих методів моделювання в умовах невизначеності і нестационарності даних, методом структурно-параметричної адаптації, яка забезпечує підвищення адекватності моделей та якості оцінок прогнозів.

Удосконалено:

– інформаційну технологію розв'язання задач прогнозування розвитку досліджуваних процесів нелінійних нестационарних процесів, яка створює підґрунтя для прийняття ефективних рішень;

– метод оцінювання параметрів математичних моделей, який відрізняється комплексним застосуванням теорії оцінювання та байєсівського підходу, що забезпечує подолання проблеми зміщеності оцінок;

– інформаційну технологію, призначену для реалізації у системах підтримки прийняття рішень на основі системного підходу, множини методів ідентифікації і врахування невизначеностей, регресійного та інтелектуального аналізу даних, яка забезпечує побудову адекватних моделей досліджуваних процесів і обчислення високоякісних оцінок прогнозів.

Значимість отриманих результатів для науки і практичного використання

Необхідно відзначити суттєве практичне значення одержаних здобувачем результатів дослідження, що виражено у можливості швидкого використання для вирішення практичних задач, як цілісної інформаційної технології так і у вигляді окремих модулів та алгоритмів, в рамках вже існуючих аналітичних рішень та платформ, що реалізують запропоновані в дисертації методи та підходи. Апробація яких була проведена за участі автора в діяльності як

державних та комерційних установ, на що є підтверджуючі довідки з Державної служби України з лікарських засобів та контролю за наркотиками; ТОВ «Картезіан-Європа» та Smart Arbitrage Technologies Limited. Результати використовуються у навчальному процесі Інституту прикладного системного аналізу Національного технічного університету «Київський політехнічний інститут імені Ігоря Сікорського»

Повнота викладу результатів досліджень в опублікованих роботах

Результати дослідження О.М. Терентьєва пройшли необхідну широку апробацію: а саме у вигляді виступів з доповідями як на міжнародних, так і на національних науково-практичних конференціях та семінарах у провідних наукових закладах за обраною тематикою дослідження.

Вивчення дисертації та праць здобувача дає можливість зробити висновок про те, що основні результати дисертаційного дослідження опубліковані у 42-х наукових працях, у тому числі 2 колективних монографіях, 18 – статтях у фахових виданнях, в тому числі 5 статей у закордонних фахових виданнях, з них 4 – у виданнях, що індексуються у наукометричній базі Scopus, 15 – у збірниках наукових праць, матеріалах і тезах міжнародних і національних конференцій. Також опубліковано 7 робіт, що додатково відображають наукові результати дисертації: 3 навчальні посібники, 4 авторських свідоцтва на комп'ютерну програму.

Автореферат написаний із використанням сучасної наукової термінології, з достатньою повнотою відображає основні наукові і практичні результати роботи та є ідентичним за змістом основним положенням дисертації. Винесені раніше на захист кандидатської дисертації наукові положення, результати і висновки, не виносяться на захист докторської дисертації.

Зауваження по роботі

1. Бажано розробити та реалізувати додаткову методологію, яка в залежності від предметної галузі використання запропонованої інформаційної

технології та типів даних надає різні варіанти послідовності проведення аналізу.

2. У роботі недостатньо приділено уваги проблемі наявності невизначеностей та заповнення пропусків у випадку просторових геоданих.

3. Вважаю за доцільне вказати, що, окрім математичного апарату мереж Байеса та регресійного аналізу, широке розповсюдження серед методів інтелектуального аналізу даних набув метод опорних векторів, з яким бажано було б виконати порівняння на обчислювальних експериментах.

4. У роботі недостатньо чітко показано, які саме типи перетворення даних необхідно використовувати для підвищення якості прогнозів за умов наявності невизначеностей.

5. В інформаційній технології, що запропонована автором недостатньо розкриті питання пов'язанні із використанням такого популярного та розповсюдженого математичного апарату нейронних мереж для розв'язання задач прогнозного моделювання.

6. Запропоновану загальну схему застосування багатомодельного підходу для моделювання нелінійних нестационарних процесів бажано розширити модулем накопичення та управління побудованими моделями.

Необхідно зазначити, що зроблені зауваження не впливають на загальну високу позитивну оцінку дисертаційної роботи.

Загальні висновки по роботі

Дисертаційна робота Терентьєва Олександра Миколайовича «Моделі, методи та інформаційні технології прогнозування нелінійних нестационарних процесів в умовах невизначеності» є завершеною науковою працею, в якій сформульовано нові науково обґрунтовані положення, принципи та методи, стосовно розробки методів аналізу нелінійних нестационарних процесів та математичних моделей і в енергетиці, охороні здоров'я, фінансовій, банківській та інших галузях, для прогнозування та підтримки прийняття рішень в умовах

наявності невизначеностей різних типів і природи. Отримано практичні результати, що мають значні перспективи для їх подальшого розвитку та практичного застосування, Робота повністю відповідає паспорту спеціальності 05.13.06 – інформаційні технології.

Дисертація **відповідає** вимогам пп. 9, 10, 12 «Порядку присудження наукових ступенів», затверджених постановою Кабінету міністрів України № 567 від 24.07.2013 р. (зі змінами, внесеними згідно з постановою КМУ № 567 від 27.07.2016) щодо докторських дисертацій, а здобувач, Терентьев Олександр Миколайович, **заслуговує** на присудження наукового ступеня доктора технічних наук за спеціальністю 05.13.06 – інформаційні технології.

Офіційний опонент,

доктор технічних наук, професор,
завідувач кафедри математичного
забезпечення комп'ютерних систем
Одеського національного університету
імені І.І. Мечникова



Є.В. Малахов

Підпис громад. Є.В. Малахова
посада зав. кафедри
ЗАВІРЯЮ:
Проректор ОНУ імені І. І. Мечникова
Запорожченко О. В.

