

ВІДГУК

офіційного опонента на дисертаційну роботу

Павлишенка Богдана Михайловича

«Методи інтелектуального аналізу консолідованих даних для підтримки прийняття рішень», подану на здобуття наукового ступеня доктора технічних наук

за спеціальністю 05.13.23 – системи та засоби штучного інтелекту

Актуальність теми. Дисертаційну роботу Павлишенка Богдана Михайловича присвячено вирішенню важливої науково-практичної проблеми, пов'язаної із застосуванням сучасних методів інтелектуального аналізу даних (ІАД) до побудови математичних моделей, оцінювання прогнозів та прийняття рішень з метою їх подальшого використання у системах підтримки прийняття рішень для підвищення якості прогнозів та управлінських рішень.

Підвищення вимог до якості розв'язання задач управління на різних рівнях діяльності пов'язано з принциповими недоліками традиційних методів вирішення цих задач, які недостатньо враховують складний системний характер реальних задач аналізу даних, прогнозування розвитку процесів та управління в сучасних умовах. Крім того, нові і суттєво важкі проблеми виникають у зв'язку з неповною даних, їх спотворенням, а також недостатньою адекватністю моделей внаслідок існуючих невизначеностей статистичного, структурного і параметричного характеру. Відсутня обґрунтована теоретична і практична бази для побудови методології системного аналізу складних процесів взаємодії об'єктів управління та інформаційного середовища, математичних моделей, необхідних для коректного оцінювання і прогнозування стану досліджуваних об'єктів, а також розв'язання задач поглибленого аналізу даних, моделювання досліджуваних процесів, їх прогнозування і високоякісного управління в умовах наявних невизначеностей.



Важливим моментом аналізу даних є їх консолідація, під якою розуміють коректне об'єднання масивів даних, що надходять з різних джерел для вирішення певної проблеми, наприклад, моделювання і прогнозування розвитку вибраних процесів. Вирішення цієї проблеми потребує обробки значних об'ємів статистичних даних, які характеризуються, на сьогодні: неповнотою, нечіткістю, спотворенням, неоднозначністю інформації, яка може проявлятися як ризик в певному розумінні. Так, невизначеність інформації, яка зберігається в базах даних, впливає на неможливість отримання коректної статистичної вибірки і застосування кількісного методу оцінювання відповідного економічного ризику. Неточність інформації призводить до побудови хибних моделей та обчислення некоректних оцінок фінансового ризику.

В цілому, несвоєчасність отримання достовірної інформації може спричинити некоректне оцінювання або навіть неусвідомлення катастрофічності ситуації і, відповідно, до запізненого реагування та застосування механізмів та засобів зниження ризиків різних типів, у тому числі фінансово-економічних. Загалом, невизначеність у обробці даних – це фактор негативного впливу на процеси моделювання, прогнозування та прийняття рішень, який може призвести до значного погіршення остаточних результатів або і до зовсім некоректних висновків. Все сказане визначає та підкреслює актуальність тематики дисертаційного дослідження.

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

Наукові положення, результати і висновки дисертації є у достатній мірі обґрунтованими. Вони ґрунтуються на ретельному аналізі великого об'єму інформації із сучасних науково-технічних джерел з досліджуваної проблеми; коректному визначенні мети і постановках задач дослідження; бездоганному використанні відомих і розробці нових методів розв'язання задач поглибленого аналізу даних, моделювання і прогнозування вибраних процесів, які характеризуються значною невизначеністю; критичному порівнянні відомих і

отриманих результатів у напрямі розробки сучасних методів поглибленого аналізу даних і їх практичного використання, а також чіткому формулюванні висновків за результатами теоретичних і практичних досліджень, виконаних автором дисертаційної роботи.

Структура, задачі та короткий зміст роботи

Робота складається із вступу, шести розділів, висновків, списку використаних джерел з 361-го джерела та додатків загальним обсягом 407 сторінок друкованого тексту, з яких 314 сторінок основного тексту.

У вступі обґрунтовано актуальність теми роботи, визначено науково-прикладну проблему, відображено зв'язок роботи з науковими програмами, темами, сформульовано мету і основні задачі дослідження, охарактеризовано наукові і практичні результати. Наведено дані про особистий внесок здобувача, апробацію та публікації результатів, показані структура і обсяг дисертації.

У першому розділі виконано аналіз відомих публікацій з проблеми аналізу даних, математичного моделювання на основі даних з використанням інтелектуального аналізу даних. Показано, що ймовірнісні моделі байєсівського типу дають можливість побудувати моделі у формі розподілів ймовірностей, байєсівських мереж і т. ін. Виконаний аналіз вказав на необхідність виконання подальших досліджень методів комбінування моделей різних типів, створення ансамблів моделей з метою подальшого ефективного використання моделей кожного типу.

У другому розділі розглянуто сучасні методи виявлення та формування необхідних для моделювання ознак. Розглянуто комбінування моделей різних типів у прогнозний ансамбль на основі стекінгового підходу, що забезпечує використання прогнозних значень, отриманих на валідаційній вибірці, у якості прогнозних ознак для моделей другого рівня.

Третій розділ присвячено аналізу концепції семантичних та тематичних лексикографічних полів з точки зору їх можливого використання в алгоритмах інтелектуального аналізу даних у формі текстових масивів. Як впливає з отриманих результатів дослідження, семантичні і тематичні поля мають

значний класифікаційний потенціал, що надає можливість використовувати кількісні ознаки з тематичних полів у задачах інтелектуального аналізу текстових даних. На основі теорії нечітких множин створено модель нечіткого семантичного поля лексемного складу текстових масивів.

Четвертий розділ присвячено подальшому розвитку методів машинного навчання в аналітиці текстових даних з використанням концепції семантичного поля. З метою дослідження кластерних та класифікаційних методів машинного навчання розглянуто текстові вибірки різних типів, зокрема, авторські тексти англomовної художньої літератури, повідомлення з груп новин та короткі повідомлення з Інтернету. Запропонований автором метод кластеризації текстових документів у просторі семантичних і тематичних полів надає можливість отримувати новий структурний поділ документів за семантичними ознаками у просторі суттєво меншої розмірності ніж простір, утворений лексемним складом текстової вибірки.

У **п'ятому розділі** розглянуто можливості застосування теорії частих множин та асоціативних правил в аналітиці текстових повідомлень соціальних мереж. Це дало можливість тематичне семантичне поле, яке у подальшому аналізі використано для асоціативних правил. Показано, що на основі текстових повідомлень з мережі Інтернет можна побудувати аналітичні моделі, які відображають вплив значущих подій, пов'язаних з аналізованими бізнес-процесами, на динаміку відповідних фінансових часових рядів, наприклад, ціни акцій вибраних для аналізу компаній на фондовому ринку.

Шостий розділ присвячено використанню теорії аналізу формальних концептів, на основі якої створено модель семантичного контексту, яка відображає структурну семантичну організацію текстових масивів. Запропоновано модель ґратки семантичних концептів для виконання аналізу твітів, згрупованих за користувачами. Формальний об'єм такого семантичного концепту об'єднує користувачів, які часто вживають у своїх повідомленнях деякі ключові слова, що утворюють множину формального змісту цього концепту.

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

Список використаних джерел є достатньо повним, він охоплює вітчизняні та зарубіжні публікації і містить 361 найменування.

У додатках описано результати аналізу врахування стохастичних патернів у прогностичній аналітиці часових рядів. Проаналізовано взаємний вплив наявності товарів в аналітиці продажів. Також виконано порівняльний аналіз впливу економічних криз на фінансовий ринок.

Достовірність результатів досліджень

Достовірність сформульованих в роботі положень, висновків і результатів забезпечується коректною постановкою проблеми математичного моделювання, прогнозування динаміки вибраних процесів в умовах невизначеності на базі комплексного використання запропонованого автором математичного апарату обробки інформації стосовно стану досліджуваних об'єктів і методів розв'язання поставлених задач в умовах наявності структурної, статистичної і параметричної невизначеності вихідних даних. Достовірність отриманих результатів також підтверджується узгодженням теоретичних і експериментальних результатів досліджень, а також успішним застосуванням цих результатів на практиці і в навчальному процесі кафедри.

Нові наукові результати, отримані автором дисертаційного дослідження не викликають сумнівів; до них необхідно віднести, у першу чергу, такі:

– метод оптимізації політики прогнозування часових рядів з використанням стекінгового об'єднання та відбору різнотипних моделей на основі регресійних моделей і байєсівської регресії, що забезпечує підвищення точності прогнозування та формування оптимального прогнозного ансамблю моделей;

– метод виявлення технічних відмов, який забезпечує за рахунок об'єднання байєсівської, лінійної та машинно-навчальної логістичної регресій

підвищення точності та достовірності результатів і надає можливість побудувати ефективні диверсифіковані процеси підтримки прийняття рішень;

– метод виявлення додаткових аналітичних ознак на основі лексемних поєднань у семантичних структурах текстових масивів, який розширює інформаційну основу для підтримки прийняття рішень в аналітиці консолідованих даних за рахунок використання теорії та асоціативних правил.

Значимість отриманих результатів для науки і практичного використання

Сформульовані автором теоретичні положення і розроблені методи мають узагальнений стандартизований характер і практично є базою для прийняття рішень у задачах математичного моделювання складних процесів на основі статистичних/експериментальних даних, стратегічному прогнозуванні і плануванні та управлінні діяльністю підприємств в різних галузях господарства України. Розроблені методи прогнозування і математичні моделі доведені до програмної реалізації і використовуються у компанії SoftServe Inc. при розробці програмного забезпечення для розв'язання задач аналізу даних, а також впроваджені у навчальні курси Львівського національного університету імені Івана Франка.

Повнота викладу результатів досліджень в опублікованих роботах

Основний зміст дисертації викладено в 52 наукових працях, з них: 30 статей у наукових фахових виданнях України і 22 публікації у матеріалах конференцій. Серед публікацій 7 статей опубліковано у виданнях, включених у міжнародну наукометричну базу Scopus, а також 5 статей опубліковано у матеріалах конференцій, які реферуються у базі Scopus.

Автореферат ідентичний за змістом основним положенням дисертації, він з достатньою повнотою відображає основні наукові і практичні результати роботи, написаний грамотно з використанням сучасної наукової термінології.

Наукові положення, результати і висновки, які виносилися на захист кандидатської дисертації, не виносяться на захист докторської дисертації.

Зауваження по роботі

1. При побудові комплексних прогнозуючих моделей бажано вказати конкретно, які нові структури моделей були отримані, наприклад, які нові нелінійні елементи були введені у моделі нелінійних процесів; які типи розподілів ймовірностей використані для опису випадкових збурень досліджуваних процесів; які методи оцінювання параметрів були використані для оцінювання ускладнених структур.
2. В алгоритмах консолідації даних у формі часових рядів доцільно застосувати методи калмановської або байєсівської фільтрації для покращення статистичних характеристик вибраних для аналізу процесів. Це дасть можливість покращити результати моделювання і прогнозування цих процесів, у першу чергу це стосується нелінійних нестационарних фінансових процесів.
3. Оскільки дослідження має явно виражений системний характер, то вважаю за доцільне сформулювати новий принцип системного аналізу, який безпосередньо відноситься до тематики дослідження, а також до інших напрямів досліджень, пов'язаних з використанням мережі Інтернет, тобто цей принцип може звучати так: *активне цілеспрямоване використання засобів інтелектуального аналізу даних і системи Інтернет у процесі аналізу функціонування складних об'єктів, прогнозування їх стану та управління ними надає можливість мінімізувати вплив невизначеностей різного характеру на точність результатів аналізу та управлінських рішень, що приймаються на основі виконаного аналізу.* На мою думку, такий принцип цілком має право на життя.

4. При виконанні аналізу взаємного впливу часових рядів у дослідженнях характеристик фінансових процесів доцільно використати таку ймовірно-статистичну характеристику як *взаємна інформація* – ймовірнісний критерій, який характеризує ступінь зв'язаності часових рядів між собою.
5. Вважаю за доцільне розглянути критеріальну базу (статистичні критерії адекватності побудованих математичних моделей, якості оцінок прогнозів, кластеризації даних) у окремому розділі і надати докладне пояснення стосовно можливостей їх коректного практичного використання.
6. Всі типи невизначеностей, що згадуються у роботі, доцільно звести у таблицю і зазначити використані методи боротьби з ними.
7. У роботі зустрічаються незначні помилки стилістичного характеру. Наприклад, слово «напрямок» необхідно вживати тільки у геометричному контексті, а в інших випадках – «напряму»; замість незрозумілого багатозначного слова «сет» доцільно використовувати слово «вибірка».

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

Загальні висновки по роботі

Вважаю, що дисертаційна робота Павлишенка Богдана Михайловича «Методи інтелектуального аналізу консолідованих даних для підтримки прийняття рішень» виконана на високому сучасному науковому рівні, містить достатню наукову новизну і за своїм змістом повністю відповідає паспорту спеціальності 05.13.23 – системи та засоби штучного інтелекту.

Дисертація є завершеною науково-дослідною роботою, яка вирішує важливу актуальну проблему наукового обґрунтування і розробки комплексу математичних моделей і методів аналізу консолідованих даних, прогнозування та підтримки прийняття рішень на основі консолідованих даних в умовах наявності невизначеностей різних типів і природи. Результати дослідження

містять необхідну наукову новизну, знайшли практичне застосування і мають значні перспективи для подальшого розвитку. Дисертація відповідає вимогам п. п. 9, 10, 12 «Порядку присудження наукових ступенів», затверджених постановою Кабінету міністрів України № 567 від 24.07.2013 року щодо докторських дисертацій, а здобувач, **Павлишенко Богдан Михайлович**, заслуговує на присудження наукового ступеня доктора технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту.

Офіційний опонент:

професор кафедри математичних методів системного аналізу

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

доктор технічних наук, професор

 Петро Бідюк

