

ВІДГУК
офіційного опонента

на дисертаційну роботу Бабія Андрія Степановича
«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 роботи. Дисертаційна робота Бабія Андрія Степановича «Моделі, методи та інтелектуальна інформаційна технологія аналізу неоднорідних послідовностей» присвячена розв'язанню актуальної науково-практичної задачі розробки моделей, методів та інтелектуальної інформаційної технології аналізу неоднорідних послідовностей даних для підвищення ефективності оцінювання поточного стану предметних областей в інформаційно-аналітичних системах.

На сьогоднішній день існує нагальна потреба в удосконаленні моделей і методів, які б дозволяли аналізувати інформацію, яка отримана з різних джерел. Сучасні напрямки досліджень охоплюють не лише проблематику обробки великих масивів однорідних даних із значною кількістю вимірів, але і такі аспекти, як комбінування даних, отриманих із різнорідною природою походження за умов наявності лише коротких наборів даних. Розмір вибірки також може бути обмеженим із різних причин, пов'язаних із особливостями предметних областей.

Більшість з існуючих сучасних інструментальних засобів інтелектуального аналізу даних характеризуються широким набором функцій для обробки структурованих масивів даних, однак інформаційні технології для вирішення задачі аналізу неоднорідних послідовностей потребують вдосконалення. Вважаю, що запропоновані у дисертаційній роботі Бабія А.С. засоби аналізу неоднорідних послідовностей дадуть можливість розширити функціональність існуючих інформаційно-аналітичних систем за рахунок інтелектуального аналізу даних.

Вважаю також, що сформульовані завдання у зазначеній постановці раніше не ставилися. Важливість їх розв'язку та можливість поширеного використання отриманих результатів визначають **актуальність теми дисертації.**

Структура та зміст дисертаційної роботи. Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та одного додатку.

У вступі обґрунтовано актуальність теми дисертації, сформульовано мету та задачі дослідження, показано наукову новизну і практичну цінність отриманих результатів.



В першому розділі проведено детальний аналіз предметної області та розглянуті актуальні публікації за напрямком аналізу даних, отриманих з різних джерел в умовах коротких вибірок даних. На підставі проведеного аналізу автор виявив ряд недоліків в існуючих моделях і методах аналізу неоднорідних послідовностей, в тому числі, окреслив обмеження застосування.

Другий розділ присвячено розробці моделей неоднорідних послідовностей. Для досліджень даних, які отримуються нерівномірно за часом, пропонується для узагальнення використовувати статистичні показники. Обробку кількісних показників з метою виявлення взаємозв'язків запропоновано виконувати існуючими методами кореляційного аналізу та множинної регресії. Для обробки даних, представлених у вигляді нечітких значень, запропоновано використовувати нечітку лінійну регресійну модель. Модифіковано тренд-сезонну модель динамічного ряду даних, а саме запропоновано використання нечітких розбиттів і F-компонент для подання трендової складової у вигляді усереднених значень. Це дозволяє уникнути втрати крайових значень для коротких вибірок даних.

Третій розділ присвячено розробці методів аналізу неоднорідних послідовностей. В роботі пропонується застосувати підхід на основі методу попарних порівнянь для непрямої побудови функцій належності. Цей підхід застосовується для врахування оцінок експертів стосовно рівнів динамічного ряду. Далі динамічний ряд із значеннями, поданими у вигляді нечітких множин, моделюється за допомогою застосування нечіткої регресійної моделі. Значущі чинники визначаються за допомогою запропонованого методу, який здійснює підбір коефіцієнтів за критерієм рівнозначності кутів відхилення між вектором похибки і векторами змінних, із наступним відбором підмножини значущих чинників із застосуванням порогового значення. Отримав розвиток метод фільтрації компонент тренд-сезонної моделі неоднорідних послідовностей, за рахунок ітеративного застосування F-перетворення для виділення трендової складової.

Четвертий розділ присвячено розробці інтелектуальної інформаційної технології аналізу неоднорідних послідовностей. На основі запропонованих моделей і методів аналізу даних здобувачем були запропоновані такі етапи інтелектуальної інформаційної технології, які включають в себе етап збору початкових даних, формування динамічних рядів і попередньої обробки та вилучення аномальних рівнів даних, вибір способу побудови моделі і проведення моделювання. Розглянуті способи застосування запропонованої інформаційної технології в рамках підходів KDD, CRISP-DM або SEMMA. Практична реалізація викладених етапів інформаційної технології здійснена у вигляді програмного модуля. В якості прикладу даних для експериментального підтвердження використовувалися відомості про кількість скоєних правопорушень та дані оцінювання носового дихання.

У висновках стисло відображені основні результати, отримані в роботі.
У додатку наведені акти впровадження дисертаційної роботи.

Ступінь обґрунтованості та достовірності наукових положень, отриманих в дисертаційній роботі. Достовірність результатів, одержаних автором, базується на:

- коректному використанні сучасного математичному апарату, відомих моделей та методів із областей теорії нечітких множин, регресійного аналізу і математичної статистики;
- узгодженості теоретичних і практичних результатів;
- результатах впровадження розроблених моделей та методів.

Наукова новизна результатів роботи полягає у постановці і розв'язанні нової актуальної науково-практичної задачі побудови інтелектуальної інформаційної технології аналізу неоднорідних послідовностей даних для оцінювання поточного стану предметних областей. Серед важливих результатів можна відзначити наступні.

1. Вперше запропонований у роботі метод визначення значущих чинників нечіткої регресійної моделі неоднорідних послідовностей даних, містить етапи підбору коефіцієнтів за критерієм рівнозначності кутів відхилення між вектором похибки і векторами змінних та відбору підмножини значущих чинників з коефіцієнтами, що перевищують порогове значення та дозволяє запобігти перенаванчання нечіткої лінійної регресії та отримати підмножину значущих чинників за скінченну кількість ітерацій.

2. Отримала подальший розвиток тренд-сезонна модель неоднорідних послідовностей, в якій трендова складова подається у вигляді інтерпольованих усереднених значень із врахуванням функції належності, що дозволяє застосовувати дану модель для коротких вибірок без втрати крайових значень і тим самим підвищити ефективність моделювання стану предметних областей в інформаційно-аналітичних системах.

3. Отримав подальший розвиток метод фільтрації компонент неоднорідних часових послідовностей, в якому для виявлення тренду початкова послідовність ітеративно розбивається на скінченну кількість нечітких розділів, для кожного з яких розраховується усереднене значення із врахуванням функції належності, що дозволяє підвищити ефективність оцінювання зміни стану предметної області за рахунок фільтрації коливань різних періодів та виділення трендової складової.

Оформлення дисертаційної роботи. Структура та обсяг дисертаційної роботи відповідають вимогам щодо тексту дисертацій на здобуття наукового ступеня кандидата технічних наук. Дисертацію написано з використанням сучасної бібліографії та наукової термінології. Задачі дисертаційного дослідження вирішуються послідовно на всіх етапах дослідження. Автореферат відповідає змісту дисертаційної роботи.

Висновок щодо відповідності дисертації паспорту 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.

Публікації за темою дисертації. Основні результати викладені в 7 статтях у наукових фахових виданнях України з технічних наук, 1 статті – за кордоном у виданні, що входить до міжнародної наукометричної бази Scopus. Також опубліковано 8 тез у матеріалах міжнародних конференцій (з них 3 включено до наукометричної бази Scopus). Результати роботи достатньо апробовані на міжнародних наукових та науково-практичних конференціях.

Порівняння змісту дисертаційної роботи і опублікованих праць дисертанта показує, що наукові положення дисертації достатньо повно відображено в публікаціях автора в відкритому друку. Апробація матеріалів на конференціях достатня.

Практичне значення результатів дисертаційної роботи. Запропонована інтелектуальна інформаційна технологія надає можливість при побудові моделі предметної області додатково враховувати відомості подані у вигляді нечітких даних, та здійснювати відбір значущих чинників і моделювати періодичну компоненту на основі адитивної декомпозиційної моделі, що сприяє проведенню аналізу в умовах коротких вибірок даних.

Розроблену інформаційну технологію було апробовано на реальних даних та впроваджено у вигляді програмного засобу у діяльність ТОВ «Ендейвер», м.Полтава та в діяльність Головного управління національної поліції Харківської області і в навчальний процес кафедри програмної інженерії ХНУРЕ, що підтверджується відповідними актами впровадження, наведеними в додатках до роботи.

Робота не вільна від недоліків та зауважень.

Зауваження.

1. В першому розділі роботи дисертантом проведено аналіз існуючих моделей і методів обробки неоднорідних послідовностей даних, але не розглянуті методи узагальнення групових експертних оцінок, що, на наш погляд, не є обов'язковим, але було б доцільним.

2. В підрозділі 2.1 під час розгляду нерегулярної складової динамічного ряду після вилучення систематичних компонент відсутнє обґрунтування властивостей випадкової компоненти.

3. В дисертаційній роботі, в підрозділі 2.1, в якому досліджуються показники, що використовуються для аналізу неоднорідних послідовностей, слід було б розглянути також і інші критерії репрезентативності вибірок даних.

4. В другому розділі дисертації здобувачем запропоновані моделі для аналізу неоднорідних послідовностей. Проте доцільно було б також розглянути базові онтологічні поняття предметної області.

5. В підрозділі 3.1, присвяченому розробці підходу для узгодження експертних оцінок, доцільно було б розглянути проблеми виявлення неякісного підбору експертів та шляхи покращення групових експертних оцінок.

6. Підрозділ 4.1 присвячено розробці інтелектуальної інформаційної технології аналізу неоднорідних послідовностей, але під час опису етапу вибору моделі користувачем не викладено, яким саме чином відбувається генерування альтернатив.

7. В четвертому розділі, під час викладення результатів експериментальних досліджень, здобувачем не розкрито умови збору даних для проведення аналізу неоднорідних послідовностей.

8. У тексті дисертації і автореферату присутні похибки стилістичного характеру. Так, наприклад, на стор. 5 автореферату двічі повторюється слово «подається» і т. і.

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

Висновки. В цілому, дисертаційна робота є завершеною науковою працею, в якій отримані нові наукові і практичні результати, що є суттєвим для вирішення задач оцінювання поточного стану предметних областей в інформаційно-аналітичних системах.

Наукова новизна не викликає сумнівів, а зазначені зауваження не впливають істотно на загальну високу оцінку роботи. Результати роботи достатньо повно представлені у 16 наукових публікаціях, пройшли необхідну апробацію.

Зміст автореферату повною мірою відображає основні положення

дисертації. Роботу написано зрозумілою мовою, що є важливим для розповсюдження і широкого застосування її результатів.

Вважаю, що дисертаційна робота відповідає вимогам пунктів 9, 11, 12 «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 наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології.

Офіційний опонент:

доктор технічних наук, професор,
завідувач кафедри інтелектуальних
комп'ютерних систем Національного
технічного університету «Харківський
політехнічний інститут»

Н.В. Шаронова

Підпис опонента професора Н.В. Шаронової
засвідчую:

Учений секретар НТУ «ХПІ», д.т.н.



О.Ю. Заковоротний