

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

ЗАЙКО ТЕТЯНА АНАТОЛІВНА

УДК 004.93

**МЕТОДИ ВИДОБУВАННЯ АСОЦІАТИВНИХ ПРАВИЛ
В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ ДІАГНОСТУВАННЯ**

05.13.23 – системи та засоби штучного інтелекту

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків – 2014

Дисертацією є рукопис.

Робота виконана у Запорізькому національному технічному університеті Міністерства освіти і науки України.

Науковий керівник – доктор технічних наук, доцент,
Субботін Сергій Олександрович,
Запорізький національний технічний університет,
професор кафедри програмних засобів.

Офіційні опоненти: доктор технічних наук, професор,
Кучеренко Євген Іванович,
Харківський національний університет
радіоелектроніки, професор кафедри
штучного інтелекту;

кандидат технічних наук, доцент,
Леонов Сергій Юрійович,
Національний технічний університет
"Харківський політехнічний інститут",
професор кафедри обчислювальної техніки
та програмування.

Захист відбудеться “_____” _____ 2014 р. о 15⁰⁰ годині на засіданні спеціалізованої вченої ради 64.052.01 у Харківському національному університеті радіоелектроніки за адресою: 61166, м. Харків, проспект Леніна, 14.

З дисертацією можна ознайомитися у бібліотеці Харківського національного університету радіоелектроніки за адресою: 61166, м. Харків, проспект Леніна, 14.

Автореферат розісланий “_____” _____ 2014 р.

Учений секретар
спеціалізованої вченої ради

О. А. Винокурова

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Розробка інтелектуальних систем діагностування пов'язана, як правило, з необхідністю обробки та аналізу великих масивів даних, що характеризують досліджувані об'єкти і процеси. Для виявлення нових закономірностей у наявних даних доцільним є використання методів видобування асоціативних правил, які дозволяють синтезувати бази правил, що мають високу інтерпретовність.

Значний внесок у розвиток методів і моделей обчислювального інтелекту як базису для побудови інтелектуальних систем діагностування внесли Л. Заде, Е. Мамдані, М. Сугено, Є. В. Бодяньський, Ю. П. Зайченко, Є. І. Кучеренко, В. І. Литвиненко, О.Г. Руденко та ін., у створення та дослідження методів видобування асоціативних правил – Дж. Адамо, І. Кох, Ч. Жанг, В. Верікос та ін.

Проте відомі методи видобування асоціативних правил, як правило, призначені для обробки бінарних даних. Більшість реальних завдань діагностування пов'язані з необхідністю обробки не тільки якісної, але і кількісної інформації, що обумовлює необхідність видобування чисельних асоціативних правил, які містять інформацію не тільки про наявність деякого набору подій, а й про їх числові характеристики. Крім того, методи пошуку асоціативних правил не дозволяють вирішувати завдання аналізу даних, що виникають при побудові інтелектуальних систем діагностування та пов'язані з факторним аналізом, скороченням розмірності навчальних вибірок і синтезом моделей на основі великих масивів неструктурованих даних.

Тому актуальним є удосконалення існуючих і створення нових методів видобування асоціативних правил для підвищення рівнів узагальнення й інтерпретовності синтезованих діагностичних моделей на основі інтелектуальних обчислень.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконана відповідно до тематичних планів Міністерства освіти і науки України та Запорізького національного технічного університету (ЗНТУ) згідно з державною науково-технічною програмою “Нові вітчизняні інтелектуальні комп'ютерні засоби”, у рамках трьох науково-дослідних робіт (НДР):

– держбюджетної НДР ЗНТУ “Інформаційні технології автоматизації розпізнавання образів і прийняття рішень для діагностування в умовах невизначеності на основі гібридних нечіткологічних, нейромережових і мультиагентних методів обчислювального інтелекту” (№ держ. реєстрації 0109U007673);

– держбюджетної НДР ЗНТУ “Інтелектуальні інформаційні технології автоматизації проектування, моделювання, управління та діагностування виробничих процесів і систем” (№ держ. реєстрації 0112U005350);

– госпдоговірної НДР ТОВ «Дніпро-Трейд» “Інтелектуальні інформаційні технології автоматизації діагностування” (№ держ. реєстрації 0110U006873).

В процесі роботи над НДР авторка дисертації брала участь як виконавець, дослідила та розробила нові методи синтезу асоціативних правил, а також розв'язала практичні завдання розпізнавання та медичного діагностування.

Мета і задачі дослідження. Метою дисертаційної роботи є розробка методів видобування асоціативних правил для підвищення рівнів узагальнення й інтерпретовності синтезованих діагностичних моделей.

Для досягнення поставленої мети у роботі вирішувалися такі задачі:

- аналіз процесу і методів побудови асоціативних правил в інтелектуальних системах;
- розробка методу видобування чисельних асоціативних правил;
- створення методу факторного аналізу на основі асоціативних правил;
- розробка методу скорочення розмірності навчальної вибірки на основі асоціативних правил;
- створення методу синтезу нейро-нечітких мереж (ННМ) з використанням асоціативних правил;
- розробка програмного забезпечення для синтезу діагностичних моделей на основі асоціативних правил за допомогою запропонованих методів;
- експериментальне дослідження розроблених методів і програмних засобів на основі вирішення практичних завдань.

Об'єкт дослідження – процес побудови асоціативних правил.

Предмет дослідження – методи видобування асоціативних правил в інтелектуальних системах.

Методи дослідження. Для вирішення завдань, поставлених у роботі, використовувалися методи: теорії нечіткої логіки – для визначення функцій належності в процесі синтезу асоціативних правил; теорії асоціативних правил – як базис для розробки нових методів вирішення задач інтелектуального аналізу даних; математичної статистики – для розробки системи критеріїв оцінювання та аналізу ефективності розроблених методів синтезу асоціативних правил.

Наукова новизна дисертаційної роботи полягає у тому, що:

- вперше запропоновано метод видобування чисельних асоціативних правил, який передбачає фазифікацію заданої бази транзакцій та автоматичне розбиття діапазонів значень ознак на інтервали, враховує індивідуальну значущість ознак, використовує критерії для оцінювання непрямих асоціацій, що знижує ступінь участі користувача в процесі пошуку асоціативних правил і дозволяє виявляти не тільки набори, що часто зустрічаються, але й набори, що рідко виникають у вихідній базі даних, проте є цікавими та дозволяють виявляти нові знання про досліджувані об'єкти або процеси;

- вперше запропоновано метод синтезу нейро-нечітких мереж, що заснований на застосуванні витягнутого з заданої транзакційної бази даних набору асоціативних правил, використовуваних для визначення структури нейро-нечіткої моделі, а також для обчислення значень параметрів функцій належності та вагових коефіцієнтів, що дозволяє на основі неструктурованих наборів даних синтезувати прості й інтерпретовні нейро-нечіткі мережі;

- удосконалено метод факторного аналізу на основі асоціативних правил, в якому, на відміну від відомих методів, визначаються еквівалентності ознак для формування факторних груп виходячи з частоти їх спільного попадання в асоціативні правила синтезованої бази правил, що дозволяє оцінювати тісноту зв'язку між різними ознаками, виключати з подальшого розгляду надлишкові ознаки та виконувати факторний аналіз в транзакційних базах даних;

- дістав подальшого розвитку метод скорочення розмірності навчальної вибірки на основі асоціативних правил, який, на відміну від відомих методів, пе-

редбачає виконання редукції екземплярів, ознак і надлишкових термів, для оцінювання інформативності ознак використовує інформацію про витягнуті асоціативні правила та дозволяє формувати розбиття простору ознак з меншою кількістю екземплярів у порівнянні з вихідною вибіркою, що в свою чергу дозволяє синтезувати більш прості та зручні для сприйняття діагностичні моделі.

Практичне значення результатів роботи полягає у тому, що:

- розроблено програмне забезпечення, яке дозволяє видобувати асоціативні правила на основі заданих транзакційних баз даних, а також вирішувати завдання аналізу даних шляхом застосування методів виявлення асоціативних правил;

- проведено експерименти з вирішення практичних завдань на основі запропонованих методів, результати яких дозволили надати рекомендації щодо застосування розроблених методів пошуку асоціативних правил на практиці;

- наукові положення, висновки і рекомендації, викладені в дисертаційній роботі, використані при підготовці курсів "Інтелектуальні системи", "Основи обчислювального інтелекту", "Системи штучного інтелекту" на кафедрі програмних засобів ЗНТУ (акт впровадження від 18.04.14);

- розроблені програмні засоби побудови асоціативних правил впроваджені у ТОВ "Дніпро-Трейд" (акт впровадження від 11.02.14) та дитячій лікарні № 1 м. Запоріжжя (акт впровадження від 17.12.13).

Особистий внесок здобувача полягає у тому, що наукові положення, висновки та рекомендації, які складають суть дисертаційної роботи, були сформульовані, розроблені та досліджені ним самостійно. Основні результати дисертації опубліковано в [1–17]. У публікаціях, написаних у співавторстві, здобувачу належать: [1] – реалізація методів вирішення задач аналізу даних на основі асоціативних правил; [2] – дослідження методів пошуку асоціативних правил для побудови інтелектуальних систем; [3] – реалізація методу побудови нейро-нечітких моделей на основі асоціативних правил; [4] – запропоновано використання асоціативних правил для вирішення завдання діагностування нейро-артритичних аномалій; [6] – реалізація методу видобування чисельних асоціативних правил з урахуванням значущості ознак; [7] – метод скорочення розмірності навчальної вибірки на основі асоціативних правил; [8] – обґрунтування доцільності застосування асоціативних правил для вирішення задач інтелектуального аналізу даних; [9] – реалізація методу факторного аналізу на основі асоціативних правил; [10] – обґрунтування доцільності використання асоціативних правил для побудови діагностичних моделей; [11] – реалізація методів видобування асоціативних правил для інтелектуального аналізу даних; [12] – метод синтезу ННМ ANFIS; [13] – метод синтезу чисельних асоціативних правил; [15] – метод факторного аналізу з виділенням асоціативних правил; [16] – програмна реалізація методів пошуку асоціативних правил; [17] – обґрунтування доцільності застосування асоціативних правил для синтезу ННМ.

Апробація результатів дисертації. Основні положення та результати дисертаційної роботи доповідалися та обговорювалися на міжнародних конференціях і семінарах: "Моделювання нерівноважних систем" (Красноярськ, 2011, 2013), "Радіоелектроніка та молодь в XXI сторіччі" (Харків, 2012, 2013), "Сис-

темний аналіз. Інформатика. Управління" (Запоріжжя, 2012), "Нейроінформатика, її застосування та аналіз даних" (Красноярськ, 2012, 2013), "Тиждень науки" (Запоріжжя, 2012), "Робототехніка та штучний інтелект" (Железногорськ, 2013), "Інформаційні процеси і технології" (Севастополь, 2013).

Публікації. Основні положення та результати дисертації опубліковано у 17 друкованих працях, з них 1 наукова монографія, 8 статей, з яких 7 статей у наукових фахових виданнях України з технічних наук (з яких 3 статті у виданнях, що входять до міжнародних наукометричних баз), 1 стаття у провідному іноземному науковому фаховому виданні, 7 публікацій у збірниках праць наукових конференцій та семінарів, а також 1 патент України.

Структура та обсяг роботи. Дисертаційна робота складається з переліку скорочень, вступу, чотирьох розділів, висновків, списку використаних джерел із 142 найменувань на 16 сторінках та одного додатку на 7 сторінках. Робота містить 18 рисунків на 18 сторінках та 6 таблиць на 7 сторінках. Загальний обсяг дисертації становить 180 сторінок, з них 143 сторінки основного тексту.

ОСНОВНИЙ ЗМІСТ РОБОТИ

Вступ містить обґрунтування актуальності теми дисертації, зв'язок роботи із науковими програмами, планами та темами, формулювання мети та основних завдань дослідження, перелік основних методів вирішення сформульованих завдань, наукову новизну та практичне значення одержаних результатів роботи, характеристику особистого внеску автора у роботи, виконані у співавторстві, відомості про апробацію та практичну реалізацію результатів роботи.

Перший розділ дисертації присвячено аналізу процесу і методів видобування асоціативних правил (АП) в інтелектуальних системах діагностування.

Проаналізовано процес видобування АП з заданої бази транзакцій $D = \{T_1, T_2, \dots, T_{N_D}\}$, де кожний елемент T_j , $j = 1, 2, \dots, N_D$ містить інформацію про деякі взаємозалежні події, $N_D = |D|$ – кількість елементів (транзакцій) у наборі даних D , $T_j = (tid_j, item_j)$, tid_j – ідентифікатор j -ї транзакції T_j , $item_j = \{t_{1j}, t_{2j}, \dots, t_{N_{item_j}j}\} \subseteq I$ – список елементів, що входять у транзакцію T_j , t_{ij} – i -й елемент списку $item_j$, $i = 1, 2, \dots, N_{item_j}$, $N_{item_j} = |item_j|$ – кількість елементів множини $item_j$, $I = \{\tau_1, \tau_2, \dots, \tau_{N_I}\}$ – множина можливих змінних (ознак), які можуть входити в список елементів $item_j$ кожної транзакції T_j , $j = 1, 2, \dots, N_D$, набору даних D , τ_a – a -й елемент множини I , $a = 1, 2, \dots, N_I$, $N_I = |I|$ – кількість елементів множини I .

Показано, що для обробки великих масивів неструктурованих даних доцільно використовувати АП, які дозволяють синтезувати бази правил, зручні для подальшого сприйняття та аналізу експертами в прикладних областях. Відзначено, що асоціативним правилом називається імплікація $X \rightarrow Y$, в якій набори X та Y : $X \rightarrow Y: X \subset I, Y \subset I, X \cap Y = \emptyset$.

Проаналізовано процес видобування АП. Досліджено основні методи побудови АП. Запропоновано класифікацію методів пошуку АП. Виявлено основні переваги і недоліки методів видобування АП. Показано, що більшість методів побудови АП висувають високі вимоги до обсягів пам'яті, пов'язані зі значними розмірами аналізованих баз даних. Крім того, як правило, процес побудови АП є досить тривалою процедурою. Іншими проблемами при видобуванні АП є необхідність вибору порогових значень підтримки та достовірності, а також необхідність забезпечення пошуку цікавих і нетривіальних правил.

Відзначено, що відомі методи видобування АП не дозволяють вирішувати завдання скорочення розмірності даних, факторного аналізу, синтезу моделей на основі великих масивів неструктурованих даних, які виникають при вирішенні реальних практичних завдань побудови інтелектуальних систем діагностування.

Проведений у роботі аналіз показав, що потреба усунення виявлених недоліків обумовлює необхідність розробки нових та модифікації існуючих методів видобування АП та їх застосування до розв'язання задач аналізу даних.

Основні результати розділу опубліковані у роботах [1, 2, 8, 10, 11].

У **другому розділі** розроблено *метод видобування чисельних асоціативних правил*, у якому пропонується використовувати підхід на основі теорії нечітких множин, що дозволяє розбивати вихідні ознаки на нечіткі інтервали та працювати з кожною ознакою, а не з окремими інтервалами її розбиття, а також при пошуку АП використовуються розраховані оцінки індивідуальної інформативності ознак, що дозволяє враховувати їхню значущість у вихідній базі даних. Пропонований метод містить такі етапи.

На початковому етапі виконується фаззифікація бази транзакцій D , тобто приведення всіх її чисельних значень до нечіткого вигляду: $D \rightarrow FuzzyD$. Таке перетворення дозволить виділити нечіткі терми кожної ознаки для можливості виконання подальшого видобування АП. У якості функцій належності доцільно використовувати такі функції, які дозволяють обмежувати інтервал значень ознак: трапецієподібну, П-подібну, трикутну функцію. Для визначення значень параметрів функцій належності виконується розбиття кожної чисельної a -ї ознаки $\tau_a \in I$ на деяку кількість інтервалів $N_{\text{інт.}a}$ з наступним визначенням границь отриманих інтервалів: $\Delta_{ak} = [l_{ak}; r_{ak})$, де l_{ak} та r_{ak} – відповідно, ліва й права границі k -го інтервалу Δ_{ak} a -ї ознаки τ_a .

Як правило, ознаки, що описують досліджувані об'єкти або процеси, мають різну інформативність, тому з метою видобування цікавих АП, що адекватно описують досліджувані залежності, пропонується враховувати індивідуальну значущість ознак. Оскільки вихідний параметр у транзакційних базах даних, як правило, не заданий, пропонується оцінювати індивідуальну значущість ознак за допомогою параметрів, що характеризують границі областей групування екземплярів (транзакцій) у просторі ознак.

Отже, для визначення індивідуальної значущості ознак пропонується виконувати кластерний аналіз, у результаті якого виділяти групи (кластери) компактно розташованих транзакцій у просторі ознак $\tau_a \in I$, I – множина усіх ознак, що входять у базу D . При цьому ознаки попередньо нормуються з метою

приведення значень усіх ознак до одного діапазону, що усуне вплив величини граничних значень ознаки на її індивідуальну значущість.

У результаті кластеризації виділяється $N_{\text{кл.}}$ кластерів. Для визначення значущості кожного елемента $\tau_a \in I$ будемо оцінювати його вплив для віднесення транзакції до кожного з кластерів. Очевидно, що чим менше ширина діапазону зміни значень a -ї ознаки в множині транзакцій кластера K_b ($b = 1, 2, \dots, N_{\text{кл.}}$), тим більше її значущість у даному кластері. Ширину діапазону будемо оцінювати як середньоквадратичне відхилення (1):

$$\sigma_{ab} = \sqrt{\sum_{g=1}^{N_{\text{тр.}b}} (\overline{\tau_{ab}} - \tau_{abg})^2}, \quad (1)$$

де $\overline{\tau_{ab}}$ – середнє значення a -ї ознаки в b -му кластері; τ_{abg} – g -те значення a -ї ознаки в b -му кластері; $N_{\text{тр.}ab}$ – кількість транзакцій в b -му кластері.

Ознаці з мінімальним значенням величини σ_{ab} будемо присвоювати максимальне значення рангу $Rg_{ab} = |I|$ в b -му кластері, наступній за зростанням значення σ_{ab} ознаці присвоїмо ранг $Rg_{ab} = |I| - 1$ і т.д. У випадку, якщо ознаки мають однакове значення σ_{ab} , їм присвоюються однакові значення Rg_{ab} . Ознаки, що рідко зустрічаються, із середнім значенням у групі $\overline{\tau_{ab}}$, нижчим мінімально припустимого ($\overline{\tau_{ab}} < \overline{\tau_{\min}}$), вважаються неінформативними в даному кластері, внаслідок чого їм присвоюється нульове значення рангу: $Rg_{ab} = 0$. Потім для кожної a -ї ознаки τ_a складаються значення рангів по всіх кластерах (2):

$$Rg_a = \sum_{b=1}^{N_{\text{кл.}}} Rg_{ab}. \quad (2)$$

Значущість (вага) w_a ознаки τ_a може визначається в такий спосіб:

– як відношення рангу Rg_a до суми рангів усіх ознак (3):

$$w_a = \left(\sum_{A=1}^{|I|} Rg_A \right)^{-1} Rg_a; \quad (3)$$

– як відношення рангу Rg_a до максимального значення рангів (4):

$$w_a = \frac{Rg_a}{\max_{A=1,2,\dots,|I|} Rg_A}. \quad (4)$$

Крім запропонованого вище для обчислення індивідуальної інформативності ознак можна використовувати підхід, що враховує межі інтервалів розбиття ознак у кластерах. У даному методі пропонується сортувати масив значень кожної ознаки τ_a за зростанням. Ліва l_{ak} й права r_{ak} межі k -го інтервалу Δ_{ak} a -ї ознаки τ_a обираються таким чином, щоб екземпляри (транзакції) зі зна-

ченням ознаки $\tau_a \in \Delta_{ak} = [l_{ak}; r_{ak})$ відносилися до одного кластеру K_b , а екземпляри із сусідніх інтервалів – до інших кластерів $K_c \neq K_b$.

Як міру інформативності a -ї ознаки в транзакційній базі даних D доцільно використовувати кількість інтервалів $N_{\text{інт.}a}$, на які розбивається діапазон її значень $\Delta_a = [\tau_{a\min}; \tau_{a\max}]$: чим менше буде кількість таких інтервалів, тим більше інформативність ознаки.

Тому значущість ознаки τ_a запропоновано обчислювати як:

– відношення мінімальної кількості інтервалів серед усіх ознак до величини $N_{\text{інт.}a}$ a -ї ознаки за формулою (5):

$$w_a = \frac{\min_{A=1,2,\dots,|I|} N_{\text{інт.}A}}{N_{\text{інт.}a}}; \quad (5)$$

– нормоване значення величини $N_{\text{інт.}a}$ за формулою (6):

$$w_a = 1 - \frac{N_{\text{інт.}a} - \min_{A=1,2,\dots,|I|} N_{\text{інт.}A}}{\max_{A=1,2,\dots,|I|} N_{\text{інт.}A} - \min_{A=1,2,\dots,|I|} N_{\text{інт.}A}} = \frac{\max_{A=1,2,\dots,|I|} N_{\text{інт.}A} - N_{\text{інт.}a}}{\max_{A=1,2,\dots,|I|} N_{\text{інт.}A} - \min_{A=1,2,\dots,|I|} N_{\text{інт.}A}}. \quad (6)$$

Важливим етапом є визначення граничних значень підтримки наборів елементів, яке в запропонованому методі відбувається з використанням інформації про індивідуальну значущість ознак, розраховану раніше. У розробленому методі видобування чисельних АП підтримку транзакції T_j будемо розраховуватися як перетинання функцій належності ознак, що входять у транзакцію T_j .

Підтримка набору X визначається як сума підтримок усіх транзакцій, які містять цю множину. Крім того, за рахунок використання критеріїв для оцінювання непрямих асоціацій передбачається можливість видобування наборів, які не часто зустрічаються, однак є цікавими та дозволяють виявляти нові знання про досліджувані об'єкти або процеси. При побудові бази АП у процесі їх видобування використовуються значення індивідуальної інформативності ознак, розраховані раніше, що дозволяє враховувати значущість кожного атрибута при пошуку правил. При генерації нових наборів-кандидатів у процесі синтезу АП ураховується властивість антимонотонності підтримки, застосування якої дозволяє суттєво скоротити простір пошуку.

Запропонований метод передбачає фазифікацію заданої бази транзакцій і автоматичне розбиття діапазонів значень ознак на інтервали, враховує індивідуальну значущість ознак, використовує критерії для оцінювання непрямих асоціацій, що знижує ступінь участі користувача в процесі пошуку асоціативних правил, зменшує ймовірність виявлення правил, які некоректно описують досліджувані об'єкти й процеси, а також дозволяє витягати набори, що не тільки часто зустрічаються, але й рідко виникаючі цікаві асоціативні правила. Використання апріорної інформації про значущість ознак у розробленому методі дозволяє скоротити простір пошуку та час видобування правил, зменшити кількість витягнутих правил, і, відповідно, підвищити рівні узагальнення й інтерпретовності синтезованої бази асоціативних правил.

Основні результати розділу опубліковані в [1, 5, 6, 13, 14].

У **третьому розділі** запропоновано методи факторного аналізу, скорочення розмірності навчальної вибірки та синтезу ННМ на основі АП.

В *методі факторного аналізу на основі асоціативних правил* виконується видобування АП, що дозволяє здійснювати оцінювання еквівалентності термів ознак, виключити з подальшого розгляду надлишкові ознаки, скоротивши тим самим простір пошуку й зменшивши час факторного аналізу.

На початковому етапі задається транзакційна база даних D , яка може містити як чисельні, так і бінарні або якісні ознаки.

Потім із заданої бази транзакцій D видобуваються АП, використовуючи відомі методи пошуку таких правил $D \rightarrow \text{БП}$, у результаті чого виконується узагальнення даних, і, відповідно, виключення з подальшого розгляду надлишкових ознак, а також деяких термів ненадлишкових ознак. Це дозволяє скоротити простір пошуку та час виконання факторного аналізу. Далі виконується спрощення синтезованої бази асоціативних правил БП, поєднуючи за можливостю деякі правила. Після цього на основі побудованої бази правил БП виділяються терми ознак $\tau_a \in I$. Для цього аналізується кожне АП з бази БП ($\text{АП}_l \in \text{БП}$), у результаті чого формуються масиви термів кожної з ознак $\tau_a \in I$: $\Delta\tau_a = \{\Delta\tau_{a1}, \Delta\tau_{a2}, \dots, \Delta\tau_{aN_{\Delta\tau_a}}\}$, де $\Delta\tau_{ak} \in [\Delta\tau_{ak \min}; \Delta\tau_{ak \max}]$ – k -й терм (інтервал) a -ї ознаки; $\Delta\tau_{ak \min}$ і $\Delta\tau_{ak \max}$ – мінімальне й максимальне значення в k -му термі a -ї ознаки, відповідно; $N_{\Delta\tau_a}$ – кількість термів a -ї ознаки.

Далі визначається еквівалентність термів ознак. Будемо вважати, що терми тим еквівалентніше, чим вище ймовірність (частота) того, що екземпляри (асоціативні правила), які потрапили в один терм $\Delta\tau_{ak}$ першої ознаки $\tau_a \in I$, потраплять в інший терм $\Delta\tau_{bm}$ другої ознаки $\tau_b \in I$. Тому для визначення еквівалентності термів ознак будемо розраховувати частоту попадання асоціативних правил у терми різних ознак за формулою (7):

$$\alpha_M = \frac{1}{N_{\text{БП}}} \sum_{l=1}^{N_{\text{БП}}} \beta_{Ml}, \quad (7)$$

де $M = \langle a, b, k, m \rangle$ – кортеж, який визначає взаємозв'язок k -го $\Delta\tau_{ak}$ терму a -ї ознаки $\tau_a \in I$ й m -го терму $\Delta\tau_{bm}$ b -ї ознаки $\tau_b \in I$; $N_{\text{БП}}$ – кількість правил у базі БП; β_{Ml} – величина, що визначає наявність зв'язку між термами ознак кортежу M в l -му правилі $\text{АП}_l \in \text{БП}$ синтезованої бази правил БП.

Після визначення еквівалентності термів α_M визначається еквівалентність ознак. Для оцінювання еквівалентності a -ї та b -ї ознак визначається величина γ_{ab} за формулою (8):

$$\gamma_{ab} = \frac{1}{N_{\Delta\tau_b} N_{\Delta\tau_a}} \sum_{m=1}^{N_{\Delta\tau_b}} \sum_{k=1}^{N_{\Delta\tau_a}} \alpha_{abkm}. \quad (8)$$

Після цього здійснюється формування груп $\Psi = \{\Psi_1, \Psi_2, \dots, \Psi_{N_\Psi}\}$ близьких ознак. Для цього використовується значення еквівалентностей γ_{ab} між a -ю та

b -ю ознаками. Крім того, пропонується визначати групи Ψ за допомогою еволюційного підходу. Для цього множина факторних груп $\Psi = \{\Psi_1, \Psi_2, \dots, \Psi_{N_\Psi}\}$ подається у вигляді хромосом $H: \Psi \rightarrow H$. Гени h_{ji} хромосоми H_j відповідають i -му елементу множини $\Psi \rightarrow H_j$. Таким чином, кожна j -та хромосома t -ї популяції $H_j^{(t)}$ буде відповідати j -му розв'язку на t -й ітерації еволюційного пошуку, що представляє собою j -ту множину факторних груп (9):

$$H_j^{(t)} \rightarrow \Psi_j^{(t)} = \left\{ \Psi_{1j}^{(t)}, \Psi_{2j}^{(t)}, \dots, \Psi_{N_{\Psi_j}j}^{(t)} \right\}. \quad (9)$$

Гени h_{ji} хромосом H_j являють собою вектори цілих чисел, що відповідають номерам ознак $\tau_a \in I$ із множини I . Для оцінювання хромосом H_j пропонується використовувати цільову функцію γ_{H_j} (10), що враховує еквівалентність γ ознак τ_a у кожному з генів h_{ji} (факторній групі):

$$\gamma_{H_j} = \left(1 + \sum_{i=1}^{N_{H_j}} \gamma_{h_{ji}} \right)^{-1} N_{H_j} \rightarrow \min, \quad (10)$$

де $\gamma_{h_{ji}}$ – середнє значення еквівалентності ознак в i -му гені h_{ji} j -ї хромосоми H_j (11):

$$\gamma_{h_{ji}} = \frac{2}{N_{h_{ji}} (N_{h_{ji}} - 1)} \sum_{a=1}^{N_{h_{ji}}} \sum_{\substack{b=a+1 \\ a,b: \tau_a, \tau_b \in h_{ji}}}^{N_{h_{ji}}} \gamma_{ab}, \quad (11)$$

де $N_{h_{ji}}$ – розмір i -го гену j -ї хромосоми.

Формування нової множини розв'язків при еволюційному пошуку відбувається за допомогою еволюційних операторів відбору, схрещування та мутації.

Еволюційна оптимізація триває доти, поки не буде досягнута максимально припустима кількість ітерацій N_{it} або не знайдений розв'язок $H_j \rightarrow \Psi_j$ із прийнятним значенням цільової функції, що не перевищує мінімально припустиме значення $\gamma_{H_j} \leq \gamma_{H \min}$.

Запропонований метод на основі еволюційного підходу передбачає видобування правил із заданих баз транзакцій, що дозволяє здійснити оцінювання еквівалентності ознак та узагальнення даних, і, відповідно, виключення з подальшого розгляду надлишкових ознак, що дозволяє скоротити простір пошуку та час виконання факторного аналізу. У розробленому методі визначення еквівалентності ознак для формування факторних груп виконується виходячи із частоти їх спільного попадання в асоціативні правила синтезованої бази правил, що дозволяє оцінювати тісноту зв'язку між різними ознаками (якісними, кількісними), не висувати вимог до вхідних даних і виконувати факторний аналіз у транзакційних базах даних. Використання еволюційного підходу для пошуку

груп якісно близьких ознак дозволяє більш детально в порівнянні з жадібною стратегією досліджувати простір пошуку, а також формувати групи близьких ознак, що характеризуються більш прийнятними оцінками еквівалентності.

Метод скорочення розмірності навчальної вибірки на основі асоціативних правил для редукції даних виявляє АП, інформація про цікавість яких використовується для оцінювання ступеню впливу ознак на вихідний параметр, а також взаємозв'язки деяких значень ознак між собою.

На початковому етапі для заданої вибірки D виконується редукція її екземплярів. Для цього дискретизуються значення ознак (діапазон значень $\Delta_a = [\tau_{a \min}; \tau_{a \max}]$ кожної ознаки τ_a розбивається на $N_{\text{інт.}a}$ інтервалів). Величина $N_{\text{інт.}a}$ може задаватися користувачем як параметр методу та бути єдиною для всіх ознак вибірки D . Крім того, кількість $N_{\text{інт.}a}$ інтервалів дискретизації a -ї ознаки τ_a може бути визначена шляхом кластеризації вибірки D і проектування границь кластерів на координатні осі в просторі ознак.

Після дискретизації виконується перетворення $D \rightarrow D'_1$, у результаті якого значення вихідних ознак τ_a замінюються номерами інтервалів значень ознак, виділених у процесі дискретизації. Отримані в результаті перетворення $D \rightarrow D'_1$ екземпляри T'_j та T'_k з однаковими значеннями ознак τ'_{aj} та τ'_{ak} , $a = 1, 2, \dots, N_I$ вважаються еквівалентними й надлишковими.

Після виконання етапу редукції екземплярів виявляються неінформативні ознаки з подальшим їх виключенням із вибірки. Для редукції ознак τ_a з вибірки D'_1 будемо витягати правила $\text{АП}_l \in \text{БП}$, оцінювати їх цікавість та цікавість кожного терму ознак, на основі чого будемо робити висновок про інформативність кожної ознаки. Як оцінки цікавості правил можна використовувати різні критерії, зокрема: $I_{\text{АП}_l} = \text{supp}(X_l \rightarrow Y_l) + \text{supp}(\bar{X}_l \rightarrow \bar{Y}_l)$, $I_{\text{АП}_l} = (\text{conf}(\bar{X}_l \rightarrow Y_l))^{-1} \text{conf}(X_l \rightarrow Y_l)$, де $\text{supp}(A)$ – підтримка множини A , обчислена як відношення кількості елементів T_j , що містять A , до загальної кількості екземплярів N_D у наборі даних D , $\text{conf}(A)$ – вірогідність множини A , що розраховується як відношення підтримки імплікації $A (X \rightarrow Y)$ до підтримки її лівої частини X .

Використовуючи інформацію про цікавості $I_{\text{АП}_l}$ витягнутих асоціативних правил, оцінюються цікавості термів $\Delta\tau_{ak}$, $k = 1, 2, \dots, N_{\text{інт.}a}$ кожної ознаки τ_a , $a = 1, 2, \dots, N_I$. Цікавість термів $\Delta\tau_{ak}$ пропонується визначати за формулою (12):

$$I_{\Delta\tau_{ak}} = \frac{1}{N_{\Delta\tau_{ak}}} \sum_{\substack{l: \text{АП}_l \in \text{БП}, \\ \Delta\tau_{ak} \in \text{АП}_l}} I_{\text{АП}_l}, \quad (12)$$

де $N_{\Delta\tau_{ak}}$ – кількість асоціативних правил $\text{АП}_l \in \text{БП}$, що містять терм $\Delta\tau_{ak}$.

Інформативність I_a ознак τ_a будемо оцінювати, виходячи з оцінок цікавостей термів, що входять у відповідну ознаку: $I_a = \frac{1}{N_{\text{інт.}a}} \sum_{k=1}^{N_{\text{інт.}a}} I_{\Delta\tau_{ak}}$.

Ознаки τ_a з низькими значеннями інформативності I_a виключаються з вибірки D'_1 . У результаті видалення неінформативних ознак і надлишкових екземплярів виконується перетворення $D'_1 \rightarrow D'_2$ й скорочення розмірності навчальної вибірки.

З метою виконання етапу скорочення надлишкових термів з вибірки D'_2 витягаються асоціативні правила та виявляються взаємозв'язки між різними інтервалами $\Delta\tau_{ak}$ й $\Delta\tau_{bm}$ ознак. З транзакцій (екземплярів) T'_{2j} вибірки D'_2 виключаються терми $\Delta\tau_{ak} \in X_l$ при наявності в цих ж транзакціях термів $\Delta\tau_{bm} \in Y_l$, що входять у консеквенти Y_l правил АП₁ бази БП₂. Шляхом виключення надлишкових термів з вибірки D'_2 виконується перетворення $D'_2 \rightarrow D'_3$ та формування вибірки D'_3 скороченої розмірності. У такий спосіб отримане розбиття простору ознак D'_3 містить суттєво меншу кількість елементів $\Delta\tau_{ak}$ у порівнянні з вихідною вибіркою D , характеризується більш високими узагальнюючими властивостями й дозволяє понизити структурну та параметричну складність синтезованих діагностичних моделей.

Метод синтезу нейро-нечітких мереж на основі асоціативних правил передбачає застосування набору асоціативних правил, витягнутих з заданої транзакційної бази даних D , для визначення структури ННМ, а також для обчислення значень параметрів функцій належності та вагових коефіцієнтів ННМ. Пропонований метод синтезу ННМ складається з таких етапів.

На першому етапі відбувається витягання асоціативних правил вигляду $X \rightarrow Y$ з заданого набору даних D . При цьому заданою є також вихідна змінна $Y \subset I$, модель залежності якої від набору вхідних параметрів $X = \{X_1, X_2, \dots, X_{N_X}\} \subset I$ необхідно побудувати.

На другому етапі виконується побудова бази правил вигляду «Якщо X , то Y ». На даному етапі використовуються побудовані раніше асоціативні правила. У випадку, якщо вихідний параметр Y є якісним (при розв'язанні завдань класифікації), у базу правил об'єднуються $N_{\text{поз.}}$ позитивних та $N_{\text{нег.}}$ негативних правил, знайдених на попередньому етапі. Таким чином, синтезується база, що містить $N_{\text{поз.}} + N_{\text{нег.}}$ правил. У випадку, якщо вихідний параметр Y і вхідні ознаки X приймають кількісні значення з відповідних діапазонів значень, база правил будується на основі чисельних АП. У результаті синтезується база з $N_{\text{пр.}}$ правил.

На третьому етапі визначаються параметри першого шару ННМ ANFIS, що реалізує систему нечіткого виведення Такагі-Сугено у вигляді п'ятишарової нейромережі прямого поширення. Нейрони першого шару відповідають нечітким термам вхідних змінних з функціями належності $\mu_{ac}^{(1)}$, $a = 1, 2, \dots, N_X$, $c = 1, 2, \dots, N_{\text{розб}a}$, де $\mu_{ac}^{(1)}$ – функція належності a -ї ознаки c -му терму; N_X – кількість вхідних параметрів X у базі правил, синтезованої на попередньому етапі; $N_{\text{розб}a}$ – кількість інтервалів розбиття a -ї ознаки. Як функцію належності пропонується використовувати функцію Гауса: $\mu_{ac} = \exp(-(X_a - M_{Xac})^2 / (2\sigma_{Xac}^2))$, де M_{Xac} – математичне сподівання a -ї ознаки X_a в c -му термі; σ_{Xac} – середньоквадратичне відхилення значень ознаки X_a в

c -му термі від математичного сподівання M_{Xac} – величина, що визначає ширину функції μ_{ac} . Визначимо значення параметрів M_{Xac} та σ_{Xac} таким чином, щоб для кожної a -ї ознаки ширина функції належності μ_{ac} для c -го терма визначалася виходячи з меж діапазонів значень ознаки X_a на кожному з інтервалів розбиття її значень, знайдених раніше, з (13) та (14):

$$M_{Xac} = \frac{1}{2}(X_{a \min c} + X_{a \max c}); \quad (13)$$

$$\sigma_{Xac} = \frac{X_{a \max c} - M_{Xac}}{2} = \frac{1}{4}(X_{a \max c} - X_{a \min c}). \quad (14)$$

Розрахунок параметрів функцій належності μ_{ac} за формулами (13) і (14) забезпечує перетинання функцій належності сусідніх інтервалів, що дозволяє ввести нечіткість при визначенні належності вхідних змінних термам. Значення параметрів M_{Xac} і σ_{Xac} розраховуються за формулами (13) і (14) для всіх інтервалів діапазону розбиття a -ї ознаки X_a : $c=1, 2, \dots, N_{\text{розб}a}$.

Етап 4. Синтез прихованих шарів нейро-нечіткої мережі. Другий шар містить $N_{\text{пр}}$ нейронів, на входи яких надходять виходи нейронів першого шару, що утворюють антецеденти правил. Кожний b -й нейрон, $b=1, 2, \dots, N_{\text{пр}}$ цього шару визначає ступінь виконання відповідного правила, використовуючи одну

з формул: $\mu_b^{(2)} = \min_{\substack{a=1, 2, \dots, N_X \\ c=1, 2, \dots, N_{\text{розб}a}}} (w_{ac}^{(2,b)} \mu_{ac}^{(1)})$ або $\mu_b^{(2)} = \prod_{a=1}^{N_X} \prod_{c=1}^{N_{\text{розб}a}} w_{ac}^{(2,b)} \mu_{ac}^{(1)}$, де $w_{ac}^{(2,b)}$ – ваго-

ві коефіцієнти, що визначають наявність зв'язків між нейронами першого шару й відповідними нейронами другого шару, що й відображають наявність c -го терма a -ї ознаки в умові b -го правила: $w_{ac}^{(2,b)} = 1$, якщо c -й терм a -ї ознаки входить в умову b -го правила, а якщо ні, то: $w_{ac}^{(2,b)} = 0$.

Нейрони третього шару знаходять нормалізовані ступені виконання правил. Четвертий шар містить нейрони, на виходах яких обчислюються лінійні комбінації вхідних ознак X_a з урахуванням ступенів виконання правил (визначаються консеквенти правил):

$y_b = \mu_b^{(3)} \sum_{a=1}^{N_X} w_a^{(4,b)} X_a$, $b=1, 2, \dots, N_{\text{пр}}$. Кожний b -й

нейрон четвертого шару з'єднаний з усіма входами мережі X_a , $a=1, 2, \dots, N_X$, а також з b -м нейроном третього шару. Єдиний нейрон п'ятого шару мережі призначений для обчислення загального виходу мережі.

Розроблений метод заснований на застосуванні витягнутого із заданої транзакційної бази даних набору асоціативних правил, використовуваних для визначення структури ННМ, а також для обчислення значень параметрів функцій належності та вагових коефіцієнтів ННМ, що дозволяє будувати ННМ на основі неструктурованих наборів даних, у яких кожний екземпляр подається транзакцією, що описує конкретну послідовність деяких взаємозалежних подій.

Застосування методів видобування асоціативних правил для синтезу ННМ дозволяє суттєво скорочувати масиви даних, зменшуючи в такий спосіб

число екземплярів (транзакцій) і ознак, які описують досліджувані об'єкти та процеси, що, у свою чергу, спрощує процес подальшої побудови моделей (зменшується час синтезу та кількість використовуваних апаратних ресурсів комп'ютера), а також дозволяє синтезувати більш прості моделі, зручні для подальшого застосування на практиці.

Основні результати розділу опубліковані в [1, 3, 7, 9, 11, 12, 15, 17].

У **четвертому розділі** описано експериментальне дослідження розроблених методів аналізу даних на основі асоціативних правил. З метою дослідження розроблених методів та їх порівняння з аналогами запропоновано систему критеріїв оцінювання асоціативних правил і методів їх видобування.

Розроблено автоматизовану систему пошуку асоціативних правил (свідчення про реєстрацію авторського права на комп'ютерну програму № 52412 від 03.12.2013 р.), що дозволяє видобувати асоціативні правила на основі заданих транзакційних баз даних, а також вирішувати завдання аналізу даних шляхом застосування різних методів виявлення асоціативних правил.

Виконано чисельні експерименти з вирішення практичних завдань автоматизації технічного та медичного діагностування, а також по дослідженню властивостей і характеристик методів видобування асоціативних правил.

Розв'язано задачу автоматизації медичного діагностування нейро-артритичних аномалій. Виявлено найбільш інформативні фактори, що дозволяють діагностувати нейро-артритичні аномалії на ранніх стадіях. Виявлені фактори і залежності дозволять своєчасно вживати необхідні дії для запобігання небажаних переходів від латентної форми до відкритої форми захворювання.

Вирішено актуальну задачу синтезу діагностичної моделі для контролю якості кондитерської продукції на основі асоціативних правил. Виявлено набір асоціативних правил, що дозволяє прогнозувати якість виробів в залежності від характеристик сировини і параметрів технологічного процесу. Оцінювання характеристик сировини шляхом застосування витягнутих правил дозволяє перевіряти якість сировини, одержуваної від окремих постачальників. Виконано відбір інформативних ознак і побудовано нейро-нечітку діагностичну модель якості кондитерської продукції, що узагальнює інформацію, подану у вигляді бази транзакцій і містить відомості про характеристики використовуваної сировини, її параметри, а також параметри обладнання та технологічних процесів.

Результати експериментів з дослідження різних методів видобування чисельних АП наведено в табл. 1, де $N_{АП}$ – кількість відібраних правил; $S_{сер.}$ – середнє значення підтримки в побудованій базі правил; $C_{сер.}$ – середнє значення достовірності витягнутих АП; t – час пошуку АП.

Таблиця 1 – Характеристики методів видобування АП

Метод	Критерій порівняння			
	$N_{АП}$, шт.	$S_{сер.}$, %	$C_{сер.}$, %	t , сек.
FARM	129	5,6	81,9	19,1
FWARM	93	4,1	89,1	15,7
МУЗП	82	3,3	91,3	12,3

Результати експериментів показали, що запропонований метод видобування чисельних АП з урахуванням значущості ознак (МУЗП) виявляв менше правил порівняно з методами FARM і FARM, що пояснюється використанням апріорної інформації про значущість ознак, це дозволяло не розглядати деякі набори (з низькими оцінками індивідуальної значущості) як такі, що часто зустрічаються і, відповідно, не тільки скорочувало час пошуку, але і зменшувало кількість видобутих правил, що в свою чергу, підвищувало рівні узагальнення та інтерпретовності синтезованої бази асоціативних правил.

Виявлено, що використання запропонованого математичного забезпечення дозволяє зменшити складність синтезованих діагностичних моделей в середньому на 67 % і підвищити їх рівень узагальнення в середньому на 53 % порівняно з нейро-нечіткими діагностичними моделями, побудованими шляхом відображення множини екземплярів на правила.

Основні результати розділу опубліковані в [1, 4, 8, 11, 16].

Висновки містять основні наукові та практичні результати дисертаційної роботи.

Додаток містить копії документів про впровадження та належність результатів роботи.

ОСНОВНІ РЕЗУЛЬТАТИ ТА ВИСНОВКИ

У дисертаційній роботі вирішено актуальну наукову задачу удосконалення існуючих і створення нових методів видобування асоціативних правил для підвищення рівнів узагальнення й інтерпретовності синтезованих діагностичних моделей. Було отримано такі результати.

1. З метою вирішення завдань, що виникають при розробці інтелектуальних систем діагностування, виконано аналіз процесу видобування асоціативних правил, що дозволяють виявляти нові закономірності, встановлювати і виявляти нові знання. Досліджено основні методи побудови асоціативних правил, виявлені їх переваги та недоліки. У результаті проведеного аналізу встановлено, що більшість методів видобування асоціативних правил висувають високі вимоги до обсягів пам'яті, пов'язані зі значними розмірами аналізованих баз даних. Крім того, як правило, процес побудови асоціативних правил є досить тривалою процедурою. Показано, що проаналізовані методи не дозволяють вирішувати завдання скорочення розмірності даних, факторного аналізу, синтезу моделей на основі великих масивів неструктурованих даних та ін., що виникають при вирішенні реальних практичних завдань побудови інтелектуальних систем діагностування. Проведений аналіз показав необхідність розробки нових і модифікації існуючих методів видобування асоціативних правил.

2. Запропоновано метод синтезу чисельних асоціативних правил, який передбачає фазифікацію заданої бази транзакцій і автоматичне розбиття діапазонів значень ознак на інтервали, враховує індивідуальну значущість ознак, використовує критерії для оцінювання непрямих асоціацій, що знижує ступінь участі користувача у процесі пошуку асоціативних правил, зменшує ймовірність видобування правил, які некоректно описують досліджувані об'єкти та

процеси, а також дозволяє виявляти не тільки набори, що часто зустрічаються, але й набори, що рідко виникають у вихідній базі даних, проте є цікавими та дозволяють виявляти нові знання про досліджувані об'єкти або процеси. Використання апріорної інформації про значущість ознак в розробленому методі дозволяє скоротити простір пошуку і час видобування правил, зменшити кількість витягнутих правил, і, відповідно, підвищити рівні узагальнення й інтерпретовності синтезованої бази асоціативних правил.

3. Розроблено метод синтезу нейро-нечітких мереж, який заснований на застосуванні видобутого з заданої транзакційної бази даних набору асоціативних правил, використовуваних для визначення структури ННМ, а також для обчислення значень параметрів функцій належності та вагових коефіцієнтів ННМ, що дозволяє будувати ННМ на підставі неструктурованих наборів даних, в яких кожен екземпляр подається транзакцією, яка описує конкретну послідовність деяких взаємопов'язаних подій. Застосування методів видобування асоціативних правил для синтезу ННМ дозволяє істотно скорочувати масиви даних, зменшуючи таким чином кількість екземплярів (транзакцій) і ознак, що описують досліджувані об'єкти і процеси, що, у свою чергу, спрощує в подальшому процес побудови моделей (зменшується час синтезу і кількість використовуваних апаратних ресурсів комп'ютера), а також дозволяє синтезувати більш прості моделі, зручні для подальшого застосування на практиці. Запропонований метод використовує побудований набір АП для виділення нечітких термів при синтезі ННМ без необхідності участі людини в цьому процесі, що дозволяє підвищити рівень автоматизації інформаційної технології, пов'язаної з необхідністю побудови нейро-нечітких моделей. На відміну від градієнтних методів, в запропонованому методі синтезу ННМ на основі асоціативних правил не потрібно обчислення значень частинних похідних цільової функції для визначення значень вагових коефіцієнтів нейронів і параметрів функцій належності.

4. Удосконалено метод факторного аналізу на основі АП, який передбачає видобування АП із заданих баз транзакцій, в результаті чого виконується узагальнення даних, і, відповідно, виключення з подальшого розгляду надлишкових ознак, що дозволяє скоротити простір пошуку і час виконання факторного аналізу. У запропонованому методі визначення еквівалентності ознак для формування факторних груп виконується виходячи з частоти їх спільного попадання в асоціативні правила синтезованої бази правил, що дозволяє оцінювати тісноту зв'язку між різними ознаками (якісними, кількісними), не висувати вимог до вхідних даних і виконувати факторний аналіз в транзакційних базах даних.

5. Розвинено метод скорочення розмірності навчальної вибірки на основі АП, який передбачає виконання етапів редукції екземплярів, ознак і надлишкових термів, для оцінювання інформативності ознак використовує інформацію про витягнуті АП та дозволяє формувати розбиття простору ознак з меншою кількістю екземплярів у порівнянні з вихідною вибіркою, що, у свою чергу, дозволяє синтезувати більш прості та зручні для сприйняття діагностичні моделі.

6. Розроблено програмне забезпечення для пошуку асоціативних правил, що дозволяє видобувати асоціативні правила на основі заданих транзакційних баз даних, а також вирішувати завдання аналізу даних і синтезу діагностичних моделей шляхом застосування різних методів виявлення асоціативних правил.

7. Запропоновані методи складають математичну основу розробленого програмного забезпечення, істотно (до 53%) підвищують рівень узагальнення синтезованих діагностичних моделей, зменшують їх складність (до 67 %), підвищуючи таким чином їх інтерпретовність, що підтверджує доцільність застосування розроблених методів на практиці при побудові діагностичних моделей в інтелектуальних системах.

Виконано чисельні експерименти з вирішення практичних завдань автоматизації технічного та медичного діагностування, а також з дослідження властивостей і характеристик методів побудови асоціативних правил. Виявлено, що запропоновані методи дозволяють підвищити рівні узагальнення та інтерпретовності діагностичних моделей в порівнянні з існуючими методами.

Отримані в роботі теоретичні та практичні результати впроваджено у виробництво на ТОВ "Дніпро-Трейд" (акт впровадження від 11.02.14) та дитячій лікарні № 1 м. Запоріжжя (акт впровадження від 17.12.13) , а також у навчальний процес ЗНТУ (акт впровадження від 18.04.14).

СПИСОК ПУБЛІКАЦІЙ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Олейник А. А. Синтез диагностических и распознающих моделей на основе гибридных нейро-нечётких технологий вычислительного интеллекта : монография / А. А. Олейник, Т. А. Зайко, С. А. Субботин ; под ред. С. А. Субботина. – Харьков : ООО “Компания Смит”, 2014. – 284 с.

2. Зайко Т. А. Анализ методов построения ассоциативных правил в интеллектуальных системах / Т. А. Зайко, А. А. Олейник, С. А. Субботин // Комп'ютерно-інтегровані технології: освіта, наука, виробництво. – 2012. – № 9. – С. 34–42.

3. Зайко Т. А. Побудова нейро-нечітких моделей на основі неструктурованих даних / Т. А. Зайко, А. О. Олійник, С. О. Субботін // Штучний інтелект. – 2012. – № 4. – С. 546–556.

4. Диагностирование нейро-артритических аномалий на основе ассоциативных правил / Т. А. Зайко, А. А. Олейник, Н. В. Жихарева, С. А. Субботин // Бионика интеллекта. – 2012. – № 2 (79). – С. 53–57.

5. Зайко Т. А. Побудова баз чисельних асоціативних правил / Т. А. Зайко // Комп'ютерно-інтегровані технології: освіта, наука, виробництво. – 2013. – № 13. – С. 21–27.

6. Зайко Т. А. Извлечение численных ассоциативных правил с учетом значимости признаков / Т. А. Зайко, А. А. Олейник, С. А. Субботин // Східно-Європейський журнал передових технологій. – 2013. – № 5/4 (65). – С. 28–34. (Входить до міжнародних наукометричних баз Index Copernicus, EBSCO, РИНЦ).

7. Зайко Т. А. Скорочення розмірності навчальної вибірки на основі асоціативних правил / Т. А. Зайко, А. О. Олійник, С. О. Субботін // Радіоелектроніка. Інформатика. Управління. – 2014. – № 1. – С. 106–111. (Входить до міжнародних наукометричних баз INSPEC, EBSCO, Index Copernicus, РИНЦ).

8. Зайко Т. А. Ассоциативные правила в интеллектуальном анализе данных / Т. А. Зайко, А. А. Олейник, С. А. Субботин // Вісник Національного технічного університету “Харківський політехнічний інститут” : зб. наук. праць. – Харків : НТУ “ХПІ”. – 2013. – № 39 (1012). – С. 82–96. (Входить до міжнародних наукометричних баз INSPEC та EBSCO).

9. Олейник А. А. Факторный анализ транзакционных баз данных / А. А. Олейник, Т. А. Зайко, С. А. Субботин // Автоматика и вычислительная техника. – 2014. – № 2. – С. 34–47. (Входить до міжнародних наукометричних баз SCOPUS, INSPEC та ін.).

10. Зайко Т. А. Методы извлечения ассоциативных правил для построения диагностических моделей / Т. А. Зайко, С. А. Субботин // Моделирование неравновесных систем : 14 Всероссийский семинар, 7–9 октября 2011 г. : материалы / под ред. В. В. Слабко. – Красноярск: Сиб. федер. ун-т, 2011. – С. 89–94.

11. Зайко Т. А. Интеллектуальный анализ данных на основе ассоциативных правил / Т. А. Зайко // Радіоелектроніка і молодь в ХХІ сторіччі : 16-ий міжнародний молодіжний форум, 17–19 квітня 2012 р. : матеріали форуму. – Харків : ХНУРЕ, 2012. – Т. 6. – С. 21–22.

12. Зайко Т. А. Метод построения нейро-нечетких сетей ANFIS / Т. А. Зайко, А. А. Олейник, С. А. Субботин // Нейроинформатика, её приложения и анализ данных : XX-й Всероссийский семинар, Красноярск, 28–30 сентября 2012 г. : материалы семинара. – Красноярск : Институт вычислительного моделирования СО РАН, 2012. – С. 47–51.

13. Зайко Т. А. Метод формирования систем численных ассоциативных правил / Т. А. Зайко, А. А. Олейник, С. А. Субботин // Нейроинформатика, её приложения и анализ данных : XXI-й Всероссийский семинар, Красноярск, 27–29 сентября 2013 г. : материалы семинара. – Красноярск : Институт вычислительного моделирования СО РАН, 2013. – С. 79–83.

14. Зайко Т. А. Извлечение знаний на основе численных ассоциативных правил / Т. А. Зайко // Радіоелектроніка і молодь в ХХІ сторіччі : 17-ий міжнародний молодіжний форум, 22–24 квітня 2013 р. : матеріали форуму. – Харків : ХНУРЕ, 2013. – Т. 6. – С. 5–6.

15. Зайко Т. А. Группировка признаков с выделением ассоциативных правил / Т. А. Зайко, А. А. Олейник, С. А. Субботин // Робототехника и искусственный интеллект : V Всероссийская научно-техническая конференция с международным участием, Железногорск, 9 ноября 2013 г. : материалы конференции. – Железногорск : Железногорский филиал СФУ, 2013. – С. 106–109.

16. Зайко Т. А. Программное обеспечение поиска ассоциативных правил на основе транзакционных баз данных / Т. А. Зайко, А. А. Олейник, С. А. Субботин // Моделирование неравновесных систем: XVI-й Всероссийский семинар, Красноярск, 4–6 октября 2013 г. : материалы семинара. – Красноярск : Институт вычислительного моделирования Сибирского отделения Российской академии наук, 2013. – С. 42–47.

17. Пат. 88634 Україна, МПК G06N 3/00. Спосіб побудови нейро-нечітких мереж на основі асоціативних правил / Т. А. Зайко, А. О. Олійник, С. О. Субботін, (Україна); заявник Запорізький національний технічний університет. – № u2013 12005 ; заявл. 14.10.13 ; опубл. 25.03.14, Бюл. № 6. – 6 с.

АНОТАЦІЯ

Зайко Т. А. Методи видобування асоціативних правил в інтелектуальних системах діагностування. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту. – Харківський національний університет радіоелектроніки Міністерства освіти і науки України, Харків, 2014.

Мета дисертації – розробка методів видобування асоціативних правил для підвищення рівнів узагальнення й інтерпретовності синтезованих діагностичних моделей.

Об’єкт дослідження – процес побудови асоціативних правил. Предмет дослідження – методи видобування асоціативних правил в інтелектуальних системах.

Розроблено нові методи і програмні засоби, які дозволяють видобувати асоціативні правила для підвищення рівнів узагальнення й інтерпретовності синтезованих діагностичних моделей. Запропоновано метод синтезу чисельних асоціативних правил, який враховує індивідуальну значущість ознак, використовує критерії для оцінювання непрямих асоціацій. Розроблено метод факторного аналізу на основі асоціативних правил, який передбачає видобування правил із заданих баз транзакцій, в результаті чого виконується узагальнення даних, і, відповідно, виключення з подальшого розгляду надлишкових ознак, що дозволяє скоротити простір пошуку і час виконання факторного аналізу. Запропоновано метод скорочення розмірності навчальної вибірки на основі асоціативних правил. Розроблено метод синтезу нейро-нечітких мереж на основі асоціативних правил. Виконано чисельні експерименти з вирішення практичних завдань автоматизації технічного та медичного діагностування, а також з дослідження властивостей і характеристик методів побудови асоціативних правил. Отримані результати доведені до практичної реалізації та впроваджені на підприємствах і в організаціях.

Ключові слова: асоціативне правило, база транзакцій, діагностування, інтерпретовність, нейро-нечітка мережа.

АННОТАЦИЯ

Зайко Т. А. Методы извлечения ассоциативных правил в интеллектуальных системах диагностирования. – На правах рукописи.

Диссертация на соискание учёной степени кандидата технических наук по специальности 05.13.23 – системы и средства искусственного интеллекта. – Харьковский национальный университет радиоэлектроники Министерства образования и науки Украины, Харьков, 2014.

Цель диссертации – разработка методов извлечения ассоциативных правил для повышения уровней обобщения и интерпретабельности синтезированных диагностических моделей.

Объект исследования – процесс построения ассоциативных правил.

Предмет исследования – методы извлечения ассоциативных правил в интеллектуальных системах.

Выполнен анализ процесса извлечения ассоциативных правил, позволяющих выявлять новые закономерности, устанавливать и выявлять новые знания. Установлено, что большинство методов построения ассоциативных правил выдвигают высокие требования к объемам памяти, связанные со значительными размерами анализируемых баз данных. Кроме того, как правило, процесс построения ассоциативных правил является достаточно длительной процедурой.

Предложен метод синтеза численных ассоциативных правил, который предполагает фазификацию заданной базы транзакций и автоматическое разбиение диапазонов значений признаков на интервалы, учитывает индивидуальную значимость признаков, использует критерии для оценивания косвенных ассоциаций, что снижает степень участия пользователя в процессе поиска ассоциативных правил, уменьшает вероятность извлечения правил, некорректно описывающих исследуемые объекты и процессы, а также позволяет извлекать не только часто встречающиеся наборы, но и редко возникающие интересные ассоциативные правила.

Разработан метод факторного анализа на основе ассоциативных правил, который предполагает извлечение правил из заданных баз транзакций, в результате чего выполняется обобщение данных, и, соответственно, исключение из дальнейшего рассмотрения избыточных признаков, что позволяет сократить пространство поиска и время выполнения факторного анализа.

Предложен метод сокращения размерности обучающей выборки на основе ассоциативных правил, который предполагает выполнение этапов редукции экземпляров, признаков и избыточных термов, для оценивания информативности признаков использует информацию об извлеченных ассоциативных правилах и позволяет формировать разбиение пространства признаков с меньшим числом экземпляров по сравнению с исходной выборкой, что, в свою очередь, позволяет синтезировать более простые и удобные для восприятия диагностические модели.

Разработан метод синтеза нейро-нечетких сетей (ННС), который основан на применении извлеченного из заданной транзакционной базы данных набора ассоциативных правил, используемых для определения структуры ННС, а также для вычисления значений параметров функций принадлежности и весовых коэффициентов ННС, что позволяет строить ННС на основании неструктурированных наборов данных, в которых каждый экземпляр представляется транзакцией, описывающей конкретную последовательность некоторых взаимосвязанных событий. Применение методов извлечения ассоциативных правил для синтеза ННС позволяет существенно сокращать массивы данных, уменьшая таким образом число экземпляров (транзакций) и признаков, описывающих исследуемые объекты и процессы, что, в свою очередь, упрощает в дальнейшем процесс построения моделей, а также позволяет синтезировать более простые модели, удобные для дальнейшего применения на практике.

Разработано программное обеспечение для поиска ассоциативных правил, позволяющее извлекать ассоциативные правила на основе заданных транзакционных баз данных, а также решать задачи анализа данных и синтеза диагностических моделей путем применения различных методов выявления ассоциативных правил.

Предложенные методы составляют математическую основу разработанного программного обеспечения, существенно (до 53%) повышают уровень обобщения синтезированных диагностических моделей, уменьшают их сложность (до 67%), повышая таким образом их интерпретабельность, что подтверждает целесообразность применения разработанных методов на практике при построении диагностических моделей в интеллектуальных системах.

Выполнены численные эксперименты по решению практических задач автоматизации технического и медицинского диагностирования, а также по исследованию свойств и характеристик методов построения ассоциативных правил. Выявлено, что предложенные методы позволяют повысить уровни обобщения и интерпретабельности диагностических моделей по сравнению с существующими методами.

Полученные в работе теоретические и практические результаты внедрены в производство, а также в учебный процесс, что подтверждается актами внедрения.

Ключевые слова: ассоциативное правило, база транзакций, диагностирование, интерпретабельность, нейро-нечеткая сеть.

ABSTRACT

Zaiko T. A. Methods of association rule extraction in intelligent diagnosis systems. – Manuscript.

The thesis for a candidate degree in technical sciences on the speciality 05.13.23 – systems and means of artificial intelligence. – Kharkiv National University of Radio Electronics of Ministry of Education and Science of Ukraine, Kharkiv, 2014.

The purpose of the thesis – the development of methods for association rules extraction to improve the generalization and interpretability levels of synthesized diagnostic models.

The object of research – the process of association rule extraction. The subject of research – the methods of association rule extraction in the intelligent systems.

New methods and software that allow to mine the association rules to improve the generalization and interpretability levels of synthesized diagnostic models are developed. The method of numerical association rules is proposed, which taking into account the individual importance of features, using the criteria for evaluation of indirect associations. The method of factor analysis based on the association rules, which involves extracting rules from a given database of transaction, resulting in the synthesis of data is performed, and therefore, the exclusion from further consideration redundant features, which can reduce the search space and time performance of factor analysis is developed. The method of the dimension reduction of training samples based on the association rules is proposed. The method for the neuro-fuzzy networks synthesis based on the association rules is developed. Numerical experiments for solving of practical problems of automation of technical and medical diagnostics and study of the properties and characteristics of methods of association rules mining are provided. The obtained results are enhanced to the practical realization and introduced in enterprises and organizations.

Key words: association rule, base of transactions, diagnostics, interpretability, neuro-fuzzy network.

Підписано до друку 00.00.2014 р. Зам. № 0000.
Формат 60x90 ¹/₁₆. Тираж 100 прим. Обсяг 0,9 д.а.

69063, Запоріжжя, вул. Жуковського, 64,
Запорізький національний технічний університет, Друкарня.