

МЕТОД ПОБУДОВИ НЕЙРОМОДЕЛЕЙ ІЗ ДІЙСНИМ ВИХОДОМ ЗА ПРЕЦЕДЕНТАМИ

Вирішено завдання автоматизації синтезу моделей кількісних залежностей за прецедентами. Запропоновано метод побудови моделей кількісних залежностей за прецедентами на основі нейронних мереж, що враховують топологічні та статистичні властивості даних. Синтезована за допомогою запропонованого методу нейронна мережа має блочну архітектуру: перший блок визначає належність розпізнаваного екземпляра до кластера у просторі ознак, другий – реалізує обчислення за частковими регресійними моделями для кластерів, третій здійснює вибір часткової моделі відповідно до належності розпізнаваного екземпляра до певного кластера. Використання запропонованого методу дозволяє отримати інтерпретабельну нейромодель з мінімальною складністю та високими узагальнюючими здатностями. Проведено експериментальне дослідження розробленого методу при вирішенні практичних завдань.

Ключові слова: нейронна мережа, ознака, вибірка, машинне навчання, діагностування.

Вступ. На практиці в задачах технічного та біомедичного діагностування часто виникає потреба в отриманні не стільки якісного (категоріального) рішення, скільки кількісної оцінки вихідного параметра моделі, на основі якої користувач на власний розсуд, або шляхом використання відомих методик та моделей зможе приймати рішення. При цьому, як правило, відсутні аналітичні моделі та експертні знання. Це вимагає створення та використання моделей із дійсним виходом за прецедентами.

Постановка завдання. Нехай задана навчальна вибірка екземплярів $x = \{x^s_j\}$, де x^s_j – значення j -ої ознаки s -го екземпляра вибірки, $j = 1, 2, \dots, N$, $s = 1, 2, \dots, S$, N – кількість вхідних ознак, S – кількість екземплярів у навчальній вибірці. Кожному екземпляру навчальної вибірки зіставлено значення вихідного дійсного параметра y^s .

Тоді залежність y від x можна записати у вигляді: $y = f(w, x) + Err$, де $f(w, x)$ – деяка функція, вид якої задається методом апроксимації або користувачем, w – множина параметрів, що дозволяють налаштувати функцію (навчити) на вирішення певної апроксимаційної задачі, Err – деяка помилка, що виникає завдяки неповній відповідності виду апроксимуючої функції виду реально існуючої залежності та похибкам у визначенні значень параметрів апроксимуючої функції, наприклад середньоквадратична помилка, яка визначається формулою (1):

$$Err = \frac{1}{2} \sum_{s=1}^S (f(x^s) - y^s)^2. \quad (1)$$

Отже, завдання побудови моделі залежності $y = f(w, x)$ за набором прецедентів полягає в ідентифікації такої структури f та набору значень параметрів w , за яких буде досягнуто прийнятний рівень помилки Err .

Аналіз літератури. Відомі методи побудови кількісних моделей залежностей за прецедентами можна класифікувати на параметричні та непараметричні [1].

Оскільки параметричні методи вимагають глибокого знання діагностованого об'єкта та залежностей між його характеристиками, використання цих методів є неможливим у застосуваннях, де відсутні або недостатні експертні знання та фізичні моделі.

Серед непараметричних методів, які знайшли широке застосування на практиці, можна виділити методи і моделі: регресійного аналізу [1, 2], штучні нейронні [3, 4] та нейро-нечіткі мережі [5, 6].

Методи регресійного аналізу [1, 2] можна безпосередньо використовувати для знаходження параметрів апроксимуючої функції, проте лише за умови, що відомий її вигляд, а це вимагає знань про об'єкт діагностування. При використанні одновимірних лінійних регресійних моделей для апроксимації багатовимірних нелінійних залежностей, як правило, не вдається забезпечити бажану точність прогнозування, але параметри апроксимуючої функції розраховуються за один прохід у неітеративному режимі. Використання нелінійної регресії може призвести до надмірного ускладнення моделі та вимагатиме великих витрат часу, а також додатково висуває питання вибору виду функції регресії. З іншого боку, у певних областях у просторі ознак значення прогнозованого параметра може визначатися, в основному, якоюсь однією ознакою, що є найбільш сильно корельованою з прогнозованим параметром у цій області.

Нейронні мережі [3, 4], зокрема мережі прямого поширення сигналу, є ефективним засобом для вирішення завдань побудови моделей за прецедентами, що мають високі узагальнювальні властивості. Проте відомі методи побудови нейромереж вимагають визначення користувачем структури мережі

(кількості шарів, кількості нейронів у кожному шарі, дискримінантних й активаційних функцій нейронів), є високо ітеративними та призводять до отримання неінтерпретабельних моделей.

Нейро-нечіткі мережі [5, 6] є різновидом нейронних мереж, що будуються за правилами системи нечіткого виведення. Ці мережі більш орієнтовані на вирішення завдань класифікації або моделювання лінійної регресії та не забезпечують високого узагальнення даних. Крім того, нейро-нечіткі мережі вимагають визначення користувачем функцій належності до нечітких термів та правил виведення, витяг яких за прецедентами є тривалим і неоднозначним.

Узагальнюючи сказане вище, можна заключити, що модель із реальним виходом за прецедентами доцільно будувати, поєднуючи нейронні мережі та часткові регресійні моделі, отримуючи тим самим кускову регресію.

Метою статті є створення методу, який дозволяв би будувати моделі із дійсним виходом за точковими спостереженнями, та був би вільний від розглянутих недоліків.

Метод побудови нейромоделей із дійсним виходом за прецедентами. Простір ознак є неоднорідним за заповненням даними, а також поведінкою вихідної ознаки. Тому доцільно виділити однорідні області у просторі ознак та побудувати часткову регресійну модель для кожної області. При цьому для забезпечення високого рівня узагальнення даних, підвищення швидкості побудови та емуляції моделі часткові моделі слід будувати з адаптацією складності поліному: спочатку обирати індивідуально найбільш інформативну в локальній області ознаку і будувати за нею лінійну одновимірну локальну регресійну модель, якщо вона не забезпечить локально прийнятну точність, то будували багатовимірну лінійну регресійну модель, а якщо і вона буде недостатньо точною, то будувати локально нелінійну регресійну модель, наприклад на основі нейроелемента. Метод побудови нейромоделей із дійсним виходом за прецедентами, що реалізує наведені ідеї, полягає у виконанні таких етапів.

1. Ініціалізація: задати навчальну вибірку екземплярів x та набір зіставлених ним значень прогнозованого параметра y .

2. Кластеризація: на основі екземплярів навчальної вибірки сформувати кластери з центрами $C^q = \{C_j^q\}$, $q = 1, 2, \dots, Q$, де Q – кількість сформованих кластерів. Для виділення кластерів пропонується використовувати метод формування кластерів [7]. Після чого розділити екземпляри навчальної вибірки на групи, що відповідають кожному кластеру за мінімумом відстані від екземпляра до центра кластера C^q у просторі ознак. Визначити S^q – кількість екземплярів вихідної вибірки, що належать до q -го кластера. У результаті ми отримаємо множину екземплярів кожного кластера та зіставлений ним набір значень вихідної ознаки.

3. Оцінювання інформативності ознак: для екземплярів кожної q -ої групи, $q = 1, 2, \dots, Q$, знайти локальні значення коефіцієнтів кореляції ζ х ознак та значень параметра y для $j = 1, 2, \dots, N$ за формулою (2):

$$r^q(x_j, y) = \frac{\sum_{s=1}^S \left(x_j^s - \frac{1}{S^q} \sum_{s=1}^S x_j^s \right) \left(y^s - \frac{1}{S^q} \sum_{s=1}^S y^s \right)}{\sqrt{\sum_{s=1}^S \left(x_j^s - \frac{1}{S^q} \sum_{s=1}^S x_j^s \right)^2 \sum_{s=1}^S \left(y^s - \frac{1}{S^q} \sum_{s=1}^S y^s \right)^2}}, x^s \in C^q, \quad (2)$$

де N – кількість ознак.

Після чого для кожної групи екземплярів знайти номер ознаки, модуль коефіцієнта кореляції якої з прогнозованим параметром є локально найбільшим за формулою (3):

$$g^q = \arg \max_j |r^q(x_j, y)|, q = 1, 2, \dots, Q; j = 1, 2, \dots, N. \quad (3)$$

4. Побудова локальних моделей. Спочатку побудувати локальну одновимірну лінійну регресію: для кожної групи екземплярів оцінити значення параметрів рівняння лінійної регресії для прогнозованого параметра за ознакою g^q за формулою (4):

$$y_g^s = b_{q0} x_{g^q}^s + b_{q1} x_{g^q}^s \in C^q, \quad (4)$$

де b_{q0} та b_{q1} – значення коефіцієнтів, що можуть бути знайдені за допомогою формул [2]. Після чого оцінити помилку отриманої моделі Err_q , яка визначається подібно до Err , але для екземплярів тільки q -го кластера. Для кожної q -ої області, для якої $Err_q/S^q > Err/S$, побудувати лінійну багатовимірну локальну регресійну модель виду (5):

$$y^s = \sum_{j=1}^N b_{qj} x_j^s, x_j^s \in C^q, \quad (5)$$

де $\{b_{qj}\}$ – коефіцієнти, що знаходяться за допомогою формул [1]. Після чого оцінити помилку отриманої моделі Err_q . Для кожної q -ої області, для якої $Err_q/S^q > Err/S$, побудувати багатовимірну нелінійну локальну регресійну модель виду (6):

$$y^s = \frac{1}{1 + \exp\left(-\sum_{j=1}^N \beta_{qj} x_j^s\right)}, x_j^s \in C^q, \quad (6)$$

коефіцієнти якої $\{\beta_{qj}\}$ знайти за допомогою методу Уїдрю–Хоффа [3].

5. Синтез структури та налаштування параметрів нейромережі: для сформованого набору кластерів та локальних моделей побудувати нейромережу прямого поширення сигналу із блочною архітектурою, структура якої зображена на рисунку 1.

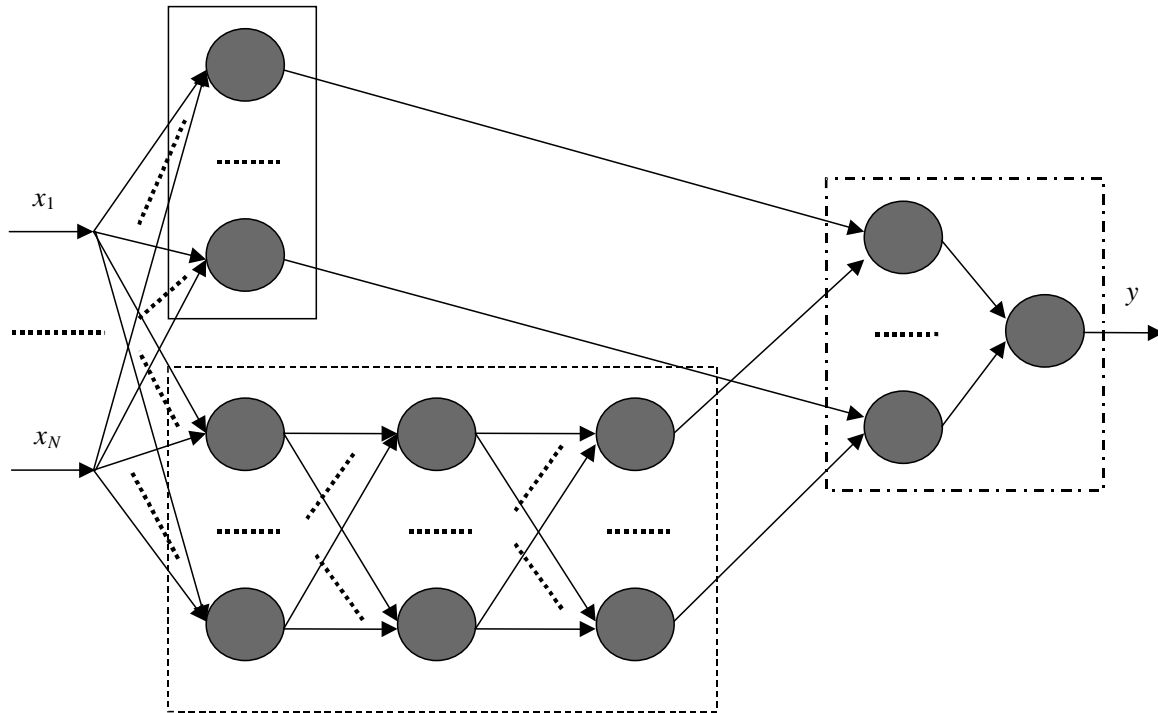


Рис. 1. Схема структури нейромережі

На входи мережі надходять значення ознак розпізнаваного екземпляра x^s . Перший блок нейронів (обведений суцільною лінією на рисунку) реалізує обчислення для часткових регресійних моделей. Другий блок нейронів (обведений штриховою лінією на рисунку) визначає належність розпізнаваного екземпляра до найближчого кластера. Третій блок нейронів (обведений штрих-пунктирною лінією на рисунку) здійснює селекцію часткової моделі для відповідного кластера. На виході мережа видає оцінку значення вихідного параметра моделі.

Введемо позначення для формального опису структури нейромережі: $\phi^{(\mu,i)}$ – дискримінантна функція i -го нейрона μ -го шару, $\psi^{(\mu,i)}$ – активаційна функція i -го нейрона μ -го шару, $w_j^{(\mu,i)}$ – ваговий коефіцієнт j -го входу i -го нейрону μ -го шару.

Перший блок утворює усього один шар нейронів, які використовують як дискримінантні функції зважені суми (7):

$$\phi^{(1,i)}(x, w) = \sum_{j=1}^N w_j x_j + w_0, \quad i = 1, 2, \dots, Q, \quad (7)$$

де Q – кількість сформованих кластерів, а як функції активації для кластерів одновимірної та багатовимірної лінійних регресій – лінійні функції (8):

$$\psi^{(1,i)}(\alpha) = \alpha, \quad i = 1, 2, \dots, Q, \quad (8)$$

для кластерів нелінійної багатовимірної регресії – логістичні сигмоїдні функції (9):

$$\psi^{(1,i)}(\alpha) = \frac{1}{1 + e^{-\alpha}}, \quad i = 1 \dots Q. \quad (9)$$

Другий блок на першому шарі містить нейрони, що визначають близькість розпізнаваного екземпляра до центрів кластерів та використовують як дискримінантні функції відстані (10):

$$\phi^{(1,i)}(x, w) = \sum_{j=1}^N (x_j - w_j)^2, \quad i = Q + 1, Q + 2, \dots, 2Q; \quad (10)$$

а як функції активації – лінійні функції (11):

$$\psi^{(1,i)}(\alpha) = \alpha, \quad i = 1, 2, \dots, Q. \quad (11)$$

Інші шари другого блоку визначають кластер, до центра якого від розпізнаваного екземпляра відстань є найменшою у просторі вхідних ознак. Дискримінантні й активаційні функції цих нейронів визначаються за формулами (12)–(16):

$$\phi^{(1,i)}(x, w) = \sum_{j=1}^N (x_j - w_j)^2, \quad i = Q + 1, Q + 2, \dots, 2Q; \quad (12)$$

$$\phi^{(2,i)}(x, w) = \sum_{j=1}^2 w_j x_j + w_0, \quad i = 1, 2, \dots, Q^2; \quad (13)$$

$$\phi^{(3,i)}(x, w) = \sum_{j=1}^Q w_j x_j + w_0, \quad i = 1, 2, \dots, Q; \quad (14)$$

$$\psi^{(1,i)}(\alpha) = \alpha, \quad i = Q + 1, Q + 2, \dots, 2Q; \quad (15)$$

$$\psi^{(\mu,i)}(\alpha) = \begin{cases} 0, & \alpha \leq 0, \\ 1, & \alpha > 0; \end{cases} \quad \mu = 2: i = 1, 2, \dots, Q^2; \quad \mu = 3: i = 1, 2, \dots, Q. \quad (16)$$

Дискримінантні й активаційні функції нейронів третього блоку визначаються за формулами (17)–(19):

$$\phi^{(4,i)}(x, w) = \prod_{j=1}^2 w_j x_j, \quad i = 1, 2, \dots, Q; \quad (17)$$

$$\phi^{(5,i)}(x, w) = \sum_{j=1}^Q w_j x_j + w_0, \quad i = 1; \quad (18)$$

$$\psi^{(\mu,i)}(\alpha) = \alpha, \quad i=1,2,\dots,Q, \quad \mu = 4, 5. \quad (19)$$

Після синтезу структури розрахуємо параметри нейромережі: для другого та третього блоків вагові коефіцієнти пропонується розраховувати за формулою (20):

$$w_j^{(\mu,i)} = \begin{cases} 0, & \mu = 5, i = 1, j = 0; \\ 1, & \mu = 5, i = 1, j > 0; \\ 1, & \mu = 4, \forall i, j > 0; \\ 1 - Q, & \mu = 3, i = 1, j = 0; \\ 1, & \mu = 3, i = 1, j > 0; \\ 0, & \mu = 2, \forall i, j = 0; \\ 1, & \mu = 2, \forall i, j = 2; \\ -1, & \mu = 2, \forall i, j = 1; \\ C_j^q, & \mu = 1, i = Q + 1, Q + 2, \dots, 2Q, j = 1, 2, \dots, N. \end{cases} \quad (20)$$

Для першого блоку для одновимірної лінійної регресії вагові коефіцієнти пропонується визначати за формулою (21):

$$w_j^{(\mu,i)} = \begin{cases} b_{q0}, & \mu = 1, i = q, j = g^q; \\ b_{q1}, & \mu = 1, i = q, j = 0; \\ 0, & \mu = 1, i = q, 0 < j \neq g^q. \end{cases} \quad (21)$$

Для першого блоку для багатовимірної лінійної та нелінійної регресій вагові коефіцієнти пропонується встановлювати за формулою (22):

$$w_j^{(\mu,i)} = \begin{cases} \beta_{ij}, & \mu = 1, i = 1, \dots, Q, j = 1, \dots, N; \\ 0, & \mu = 1, i = 1, \dots, Q, j = 0. \end{cases} \quad (22)$$

Експерименти та результати. Для перевірки практичної придатності та дослідження пропонуваного методу була створена його програмна реалізація. За допомогою розробленої програми експериментально досліджено пропонуванний метод шляхом вирішення практичних завдань технічного діагностування [8]. Зокрема для задачі прогнозування коефіцієнта поверхневого зміцнення лопаток газотурбінних авіадвигунів [8] навчальна вибірка даних складалася з 57 екземплярів, кожен з яких характеризувався набором 16 вхідних ознак: x_1 – твердість матеріалу, x_2 – середній контактний тиск у зоні деформування, x_3 – піввісь еліпса торкання в зоні пружного контакту, x_4 – подача при вигладжуванні, x_5 – межа міцності, x_6 – межа текучості матеріалу, x_7 – показник деформаційного зміцнення, x_8 – теоретичний коефіцієнт концентрації напруг від слідів обробки, x_9 – параметр вихідної шорсткості деталі, x_{10} – сила вигладжування, x_{11} – радіус сфери алмазного інструмента, x_{12} – параметр шорсткості після вигладжування деталі, x_{13} – теоретичний коефіцієнт концентрації напруг натурної деталі (зразка), x_{14} – діаметр зразка в небезпечному перерізі, x_{15} – радіус скруглення галтеля або надрізу, x_{16} – відносний градієнт першої головної напруги. Вихідною ознакою є коефіцієнт зміцнення лопатки після алмазного вигладжування. Для заданої вибірки, фрагмент якої наведено у таблиці 1, за допомогою розробленого математичного забезпечення побудовано нейромодел, фрагмент результатів роботи якої наведено у стовпці f таблиці (значення наведені з округленням).

Фрагмент вхідних даних та результатів моделювання

s	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀	x ₁₁	x ₁₂	x ₁₃	x ₁₄	x ₁₅	x ₁₆	y	f
1	2350	3390	0,14	0,08	900	650	0,116	1,45	1,1	100	3	0,25	1	10	10	0,43	1,16	1,159
6	3300	5630	0,219	0,085	1150	1000	0,126	0,35	0,85	400	3	0,12	1	10	10	0,43	1,39	1,389
10	2700	5570	0,22	0,08	950	600	0,103	0,35	0,85	400	3	0,6	1	10	10	0,43	1,51	1,509
12	3300	5110	0,199	0,06	1150	1000	1,126	1,35	0,85	300	3	0,1	1	10	10	0,43	1,38	1,379
20	2300	3900	0,126	0,08	950	850	0,106	1,25	0,23	110	3	0,14	1	7,5	10	0,50	1,27	1,271
21	3100	5500	0,202	0,08	960	850	0,131	1,45	1,1	300	2,5	0,19	1	10	10	0,43	1,45	1,449
22	2850	4860	0,196	0,08	1000	700	0,103	1,35	0,85	250	2,5	0,12	1	10	10	0,43	1,50	1,501
25	2600	3970	0,125	0,08	1000	800	0,126	1,3	0,7	110	3	0,07	1	7,5	10	0,5	1,44	1,439
27	3300	5630	0,219	0,08	1150	1000	0,126	1,45	1,15	400	3	0,14	1,15	10	6	0,575	1,29	1,290
30	2350	3140	0,40	0,08	950	850	0,106	1,25	0,63	200	2,5	0,40	1,45	60	10	0,30	1,57	1,570
32	2550	5330	0,224	0,08	960	770	0,115	1,25	0,65	400	3	0,11	1,0	10	10	0,43	1,47	1,469
33	3300	4810	0,176	0,085	1150	1000	0,126	1,8	2,8	200	2,5	1,0	1,0	10	10	0,43	1,43	1,429
38	2700	4780	0,177	0,08	950	600	0,103	1,8	2,8	200	2,5	0,35	1,0	10	10	0,43	1,49	1,491
44	2350	4620	0,18	0,08	900	650	0,116	1,8	2,8	200	2,5	0,35	1,0	10	10	0,43	1,66	1,659
50	2300	3900	0,126	0,08	950	850	0,106	1,7	2,5	110	3,0	0,6	1,0	7,5	10	0,5	1,20	1,200
52	2600	3970	0,125	0,1	1000	800	0,126	1,65	2,4	110	3	0,30	1,0	7,5	10	0,5	1,13	1,131
54	3300	3480	0,139	0,08	1180	1000	0,126	1,65	2,4	100	3	1,2	2,52	10	0,25	8,2	1,83	1,829

Як видно з таблиці 1, запропонований метод дозволяє отримувати моделі кількісних залежностей з прийнятною точністю. Порівняно з раніше отриманими регресійними моделями [8], побудована модель забезпечує більшу точність, а порівняно з нейромоделями на основі багатошарових нейромереж прямого поширення [8] є більш інтерпретабельною.

Експерименти показали, що розроблене математичне забезпечення є працездатним та дозволяє вирішувати завдання побудови моделей кількісних залежностей за прецедентами.

Висновки. У роботі вирішено актуальне завдання автоматизації синтезу моделей кількісних залежностей за прецедентами для вирішення задач технічного діагностування.

Наукова новизна отриманих результатів полягає у тому, що вперше запропоновано метод побудови моделей кількісних залежностей за прецедентами на основі нейронних мереж, що враховують топологічні та статистичні властивості даних. Синтезована за допомогою запропонованого методу нейронна мережа має блочну архітектуру: перший блок визначає належність розпізнаваного екземпляра до кластера у просторі ознак, другий блок реалізує обчислення за частковими регресійними моделями для кластерів, третій блок здійснює вибір часткової моделі відповідно до належності розпізнаваного екземпляра до певного кластера. Використання запропонованого методу дозволяє отримати інтерпретабельну нейромоделю з мінімальною складністю та високими узагальнювальними здатностями.

Практична значущість отриманих результатів полягає у тому, що розроблено програмне забезпечення, яке реалізує запропонований метод побудови моделей кількісних залежностей, проведено експериментальне дослідження розробленого методу при вирішенні практичних завдань технічного діагностування. Встановлено, що нейромоделю, яка отримується у результаті синтезу й навчання на основі запропонованого методу, є інтерпретабельною та за потреби легко перетворюється у дерево правил "якщо-то", що дозволяє використовувати її для видобування знань з даних. Точність класифікації та швидкість навчання НМ, сформованої на основі запропонованого алгоритму, є достатніми для багатьох прикладних задач діагностування та прогнозування.

Подальші дослідження можуть бути спрямовані на розробку процедур автоматизації визначення кількості розбиттів вибірки (кластерів), оптимізації (спрощення структури) отриманої моделі, перетворення отриманої нейромережі у нейро-нечітку мережу, а також на дослідження властивостей розробленого математичного забезпечення у широкому колі практичних застосувань різної природи, розмірності та складності.

Роботу виконано у межах держбюджетної науково-дослідної теми Запорізького національного технічного університету "Інтелектуальні інформаційні технології автоматизації проектування, моделювання, керування та діагностування виробничих процесів і систем" (номер держ. реєстрації 0112U005350).

Список використаної літератури:

1. Дрейнер Н. Прикладной регрессионный анализ : в 2-х книгах / Н.Дрейнер, Г.Смит. – Кн. 1. – М. : Финансы и статистика, 1986. – 366 с.

2. *Гринчишин Я.Т.* Алгоритмы и программы на бейсике / *Я.Т. Гринчишин, В.И. Ефимов, А.Н. Ломакович.* – М. : Просвещение, 1988. – 160 с.
3. *Руденко О.Г.* Штучні нейронні мережі / *О.Г. Руденко, С.В. Бодяньський.* – Харків : Компанія СМІТ, 2006. – 404 с.
4. *Хайкин С.* Нейронные сети : полный курс / *С.Хайкин.* – СПб. : Вильямс, 2005. – 1104 с.
5. *Митюшкин Ю.И.* Soft Computing: идентификация закономерностей нечёткими базами знаний / *Ю.И. Митюшкин, Б.И. Мокин, А.П. Ротштейн.* – Винница : Универсум-Винница, 2002. – 145 с.
6. *Субботин С.О.* Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень : навч. посібник / *С.О. Субботин.* – Запоріжжя : ЗНТУ, 2008. – 341 с.
7. *Субботин С.А.* Алгоритм планирования он-лайнного эксперимента в нейросетевой диагностике / *С.А. Субботин* ; под ред. *А.Н. Горбаня* ; отв. за выпуск *Г.М. Цибульский.* // Нейроинформатика и ее приложения : IX Всерос. семинар (5–7 окт. 2001 г., г. Красноярск). – Красноярск : КГТУ, 2001. – С. 180–181.
8. *Прогрессивные технологии моделирования, оптимизации и интеллектуальной автоматизации этапов жизненного цикла авиационных двигателей : монография / А.В. Богуслаев, Ал.А. Олейник, Ан.А. Олейник и др.* ; под ред. *Д.В. Павленко, С.А. Субботина.* – Запорожье : ОАО "Мотор Сич", 2009. – 468 с.

СУББОТІН Сергій Олександрович – кандидат технічних наук, доцент, професор кафедри програмних засобів Запорізького національного технічного університету.

Науковці інтереси:

– нейронні та нейро-нечіткі мережі;

– розпізнавання образів;

– технічне діагностування.

Тел.: (067)394–11–80.

Email: subbotin@zntu.edu.ua

Стаття надійшла до редакції 10.10.2013