УДК 681.518

О. В. Сєрая, к. т. н., доц.; Т. І. Каткова к. п. н, доц.

НЕЙРОМЕРЕЖЕВА ПРОДУКЦІЙНА ЕКСПЕРТНА СИСТЕМА ДІАГНОСТИКИ СТАНУ

Запропоновано експертну систему з нейромережним механізмом логічного виводу. Розроблено процедуру прогнозування стану об'єкту, який діагностується за таким механізмом.

Ключові слова: *експертна система, продукційні правила, нейромережеві структури, механізм логічного виводу, прогнозування стану об'єктів.*

Вступ

Сучасні інформаційні технології діагностики стану об'єктів ефективно використовують специфічні системи штучного інтелекту – експертні системи (ЕС). Такі системи організовані таким чином, щоб з використанням результатів вимірювання набору контрольованих параметрів $x_1, x_2, ..., x_n$ об'єкта діагностувати його стан. Принципи функціонування і структура діагностичних експертних систем суттєво залежать від типу механізму логічного виводу (МЛВ). На практиці використовуються два конструктивно різних підходи до побудови МЛВ: продукційний і байєсовий. Широко використовуваний продукційний підхід базується на системі так званих продукційних правил, побудованих таким чином [1]:

якщо $x_1 \in A_1,\ x_2 \in A_2, ..., x_n \in A_n$, то з ймовірністю $p(A_1, A_2, ..., A_n)$
об'єкт знаходиться в стані $H(A_1, A_2, ..., A_n)$.

Продукційні системи зручні для практичного застосування і прості в реалізації. Однак серйозний недолік полягає в конструктивній залежності ефективності системи діагностики від розмірності набору контрольованих параметрів. Річ у тім, що система діагностики повинна володіти повнотою, тобто необхідно, щоб для кожного можливого варіанту значень набору контрольованих параметрів існувало відповідне продукційне правило. Це означає, що якщо кожний параметр може прийняти одне m можливих значень, то загальне число продукційних правил дорівнюватиме $N = m^n$ і швидко зростає зі збільшенням m і n.

Байєсов механізм логічного виводу [2] практично знімає проблему розмірності. Однак практичні можливості таких систем обмежуються необхідністю статистичної незалежності контрольованих параметрів. Спроба обійти цю проблему була здійснена в [3]. При цьому залежні параметри об'єднуються в групи. Далі параметри, які входять в одну групу, обробляються з використанням продукційних правил. Отримані результати для різних груп вже практично незалежні і використовуються в байєсовій технології. Така ЕС з комбінованим МЛВ дозволяє розв'язувати задачі діагностики для великої розмірності набору контрольованих параметрів і їх можливої корельованості. Разом з цим, вона має загальний принциповий недолік будь-яких ЕС – дискретний характер контрольованих параметрів. І в продукційній, і в байєсовій системах кожний з параметрів – це або симптом з булевим характером прояву, або, якщо це неперервний параметр, діапазон можливих його значень потрібно розбити на піддіапазони, тобто дискретизувати. Ця обставина при практичній розробці ЕС приводить до ряду проблем. По-перше, якщо кількість піддіапазонів велика, вибір раціональної кількості піддіапазонів повинен бути результатом нетривіального компромісу між складністю МЛВ і точністю оцінювання стану в протилежному випадку. По-друге, межі піддіапазонів важко пояснити теоретично. По-третє, саме існування меж піддіапазонів може привести до неприродної ситуації, коли двом близьким за чисельними значеннями наборам параметрів будуть відповідати різні діапазони.

Всі ці проблеми можна усунути, якщо МЛВ буде конструктивно адаптований до обробки неперервних за своєю природою параметрів. Такий МЛВ може бути реалізований з використанням штучних нейронних мереж (ШНМ). Сформулюємо задачу розробки ЕС з нейромережевим механізмом логічного виводу, яка призначена для оцінки і прогнозування стану об'єкта, який діагностується.

Постановка задачі

Як відомо, будь-яка ШНМ здійснює відображення точок з багатомірного простору спостережень X розмірності n в точки багатомірного простору Y розв'язків іншої, в загальному випадку, розмірності H. За цих умов правильне відображення точок з X в Y забезпечується спеціальним чином організованою процедурою навчання мережі. Нехай проводиться серія вимірювань контрольованих параметрів об'єкта, в результаті яких отримаємо набори $X_1 = (x_{11}, x_{12}, ..., x_{1j}, ..., x_{1n})$, $X_2 = (x_{21}, x_{22}, ..., x_{2j}, ..., x_{2n}), \ldots,$ $X_l = (x_{l1}, x_{l2}, ..., x_{lj}, ..., x_{ln}), \ldots,$ $X_L = (x_{L1}, x_{L2}, ..., x_{Lj}, ..., x_{Ln})$. У процесі навчання мережі ці набори подаються Q експертам, які для кожного набору $X_l, l = 1,2,...,L$ задають розподіл ймовірностей $P_{lq} = (p_{lq1}, p_{lq2}, ..., p_{lqh}, ..., p_{lqH})$, $l = 1,2,...,L$, $q = 1,2,...,Q$, станів $h = 1,2,...,H$ об'єкта. Елементарна статистична обробка результатів експертного оцінювання ставить у відповідність кожному набору вимірювань X_l розподіл середніх значень ймовірностей станів (діапазонів) $\hat{P}_l$ і набір дисперсій ймовірностей $\hat{\sigma}_l^2$. Отримані дані використовуються для навчання ШНМ, після якого нейронна мережа для кожного нового вектора вимірювань контрольованих параметрів формує відповідні вектори $P(X)$ і $\sigma^2(X)$. Проблемним залишається питання про можливість використання ШНМ для прогнозування стану об'єкта. Розглянемо дві альтернативні методики розв'язання цієї задачі.

Основні результати

А. Мікропідхід. Оскільки навчена ШНМ для кожного вектора спостережень набору контрольованих параметрів визначає відповідний розподіл ймовірностей діагнозів, то розв'язання задачі прогнозування стану об'єкта може бути отримане шляхом прогнозування вектора спостережень.

Використаємо набори $X_1, X_2, ..., X_n$ для формування матриці спостережень:

$$X = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ x_{L1} & x_{L2} & \ldots & x_{Ln} \end{pmatrix}.$$

Стовпці цієї матриці відповідають відлікам кожного з контрольованих параметрів для моментів часу $t_1, t_2, ..., t_L$. Введемо модель еволюції параметрів у часі:

$$x_j(t) = \sum_{i=0}^{d} c_{ij} t^i, \quad j = 1,2,...,n. \tag{1}$$

Розрахунок параметрів (c_{ij}) моделі (1) проведемо за методом найменших квадратів, використовуючи матрицю

$$H = \begin{pmatrix} 1 & t_1 & \ldots & t_1^d \\ 1 & t_2 & \ldots & t_2^d \\ \ldots & \ldots & \ldots & \ldots \\ 1 & t_L & \ldots & t_L^d \end{pmatrix}$$

і вектори

$$c_1 = \begin{pmatrix} c_{01} \\ c_{11} \\ \dots \\ c_{d1} \end{pmatrix},\ c_2 = \begin{pmatrix} c_{02} \\ c_{12} \\ \dots \\ c_{d2} \end{pmatrix}, \dots,\ c_n = \begin{pmatrix} c_{0n} \\ c_{1n} \\ \dots \\ c_{dn} \end{pmatrix},\ X^{(1)} = \begin{pmatrix} x_{11} \\ x_{21} \\ \dots \\ c_{L1} \end{pmatrix},\ X^{(2)} = \begin{pmatrix} x_{12} \\ x_{22} \\ \dots \\ c_{L2} \end{pmatrix}, \dots,\ X^{(n)} = \begin{pmatrix} x_{1n} \\ x_{2n} \\ \dots \\ c_{Ln} \end{pmatrix}.$$

Тепер набір векторів оцінок параметрів рівнянь регресії (1) знайдемо, мінімізуючи функціонали:

$$J_1(c_1) = \left(Hc_1 - X^{(1)}\right)^T \left(Hc_1 - X^{(1)}\right),$$
$$J_2(c_2) = \left(Hc_2 - X^{(2)}\right)^T \left(Hc_2 - X^{(2)}\right),$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$J_n(c_n) = \left(Hc_n - X^{(n)}\right)^T \left(Hc_n - X^{(n)}\right).$$

При цьому оптимальні за критерієм найменших квадратів вектори регресійних коефіцієнтів отримаємо за формулами:

$$\hat{c}_j = \left(H^T H\right)^{-1} H^T X^{(j)},\ j = 1,2,...,n. \tag{2}$$

Підстанова (2) в (1) задає сукупність аналітичних описів поведінки контрольованих параметрів у часі, які забезпечують можливість розрахунку їхніх значень на момент t_{np} прогнозу:

$$x_j(t_{np}) = \sum_{i=0}^{d} \hat{c}_{ij} t_{np}^i,\ j = 1,2,...,n.$$

Отриманому при цьому набору значень параметрів

$$X(t_{np}) = \left(x_1(t_{np}), x_2(t_{np}),..., x_n(t_{np})\right)$$

навчання ШНМ поставить у відповідність розподіл ймовірностей станів об'єкта на момент пронозу.

Б. Макропідхід. Послідовно подаючи на вхід ШНМ набори спостережень $x_1, x_2,..., x_n$, отримаємо відповідні розподіли ймовірностей діагнозів:

$$P_1(t_1) = \left(p_{11}, p_{12},..., p_{1H}\right),$$
$$P_2(t_2) = \left(p_{21}, p_{22},..., p_{2H}\right), \tag{3}$$
$$\dots\dots\dots\dots\dots\dots\dots\dots..$$
$$P_L(t_L) = \left(p_{L1}, p_{L2},..., p_{LH}\right).$$

Закон зміни ймовірностей для кожного, наприклад, h-го, $h = 1,2,...,H$, з діапазонів описується відповідною функцією часу $P_h(t)$, яка може бути представлена розкладанням у ряд за деякою сукупністю базисних функцій у відповідністю з моделлю:

$$P_h(t) = \sum_{i=0}^{m} a_{ih}\varphi_i(t),\ \ h = 1,2,...,H, \tag{4}$$

де $\varphi_0(t), \varphi_1(t),..., \varphi_m(t)$ – набір базисних функцій.

Необхідно відзначити, що шаблонно незалежне оцінювання параметрів моделей (4) для кожного діагнозу окремо неможливе, тому що при цьому не враховується умова нормування: сума ймовірностей діагнозів на будь-яких момент часу повинна дорівнювати одиниці.

В зв'язку з цим розглянемо інший підхід.

Проста технологія використання набору розподілів ймовірностей (3), які відповідають моментам спостережень $t_1, t_2,..., t_L$, для розрахунку розподілу на момент прогнозу полягає в наступному.

Апроксимуємо кожний з розподілів (3) неперервною кривою $f\left(t_l, \theta_1(t_l), \theta_2(t_l),..., \theta_q(t_l)\right)$, яка має сенс щільності розподілу ймовірностей діагнозів і залежній від q параметрів. Тепер для кожного з наборів $\left(\theta_1(t_1),\theta_1(t_2),...,\theta_1(t_L)\right)$, $\left(\theta_2(t_1),\theta_2(t_2),...,\theta_2(t_L)\right),\ldots,$ $\left(\theta_q(t_1),\theta_q(t_2),...,\theta_q(t_L)\right)$ побудуємо аналітичне продовження і обчислимо сукупність параметрів $\theta_1(t_{np}),\theta_2(t_{np}),...,\theta_q(t_{np})$, яка однозначно задає шуканий розподіл ймовірностей діагнозів на момент t_{np} прогнозу. На жаль, ця елементарна процедура може вимагати великих затрат часу, тому що характер кожного з розподілів (3) може бути настільки складним, що вимагатиме для адекватної його апроксимації неприпустиму кількість параметрів q, інакше буде неприпустимо грубим.

Альтернатива полягає у використанні для опису розподілів (3) емпіричних законів розподілу, принципова властивість яких – монотонне неспадання. Для оцінки емпіричного закону розподілу можна використовувати функціонал

$$g(t)=\frac{1}{2}+\frac{1}{2}th\big(y(t)\big), \tag{5}$$

в якому $y(t)$ забезпечує адаптацію емпіричного закону $g(t)$ до реальних даних. У [4] показано, що достатньо добру за якістю апроксимацію реальних емпіричних законів розподілу забезпечує опис $y(t)$ у вигляді квадратичного полінома. Загальна схема прогнозування така. Сукупність розподілів $P_1(t_1),P_2(t_2),...,P_L(t_L)$ перетворюються в емпіричні інтегральні закони:

$$R_1=\left(R_{11},R_{12},...,1\right),R_2=\left(R_{21},R_{22},...,1\right),\ldots,R_L=\left(R_{L1},R_{L2},...,1\right),$$

де

$$R_{lh}=\sum_{h=1}^{H}P_{lh},\quad l=1,2,...,L.$$

Далі кожний з отриманих інтегральних законів незалежно апроксимується з використанням функціонала (5), який представлено у формі:

$$g_l(h)=\frac{1}{2}+\frac{1}{2}th\left(a_{0l}+a_{1l}h+a_{2l}h^2\right),\quad l=1,2,...,L,\ h=1,2,...,H\ . \tag{6}$$

Показник якості апроксимації має вигляд:

$$W_l=\sum_{h=1}^{H}\left(R_{lh}-g_l(h)\right)^2. \tag{7}$$

Мінімізація (7) здійснюється методом найменших квадратів. У результаті отримуємо сукупність наборів $\{(a_{01},a_{11},a_{21}),(a_{02},a_{12},a_{22}),...,(a_{0L},a_{1L},a_{2L})\}$. Ці набори використовуються для отримання поліноміальних функцій $a_0(t),a_1(t),a_2(t)$, які забезпечують можливість опису апроксимації (6) для будь-якого моменту часу.

Розглянемо цю методику детальніше. Для опису функцій $a_0(t),a_1(t),a_2(t)$ використаємо систему функцій $\psi_1(t_l),\psi_2(t_l),...,\psi_d(t_l)$, ортогональних на сукупності рівновіддалених точок $t_1,t_2,...,t_L$ так, щоб

$$\sum_{l=1}^{L}\psi_{k_1}(t_l)\psi_{k_2}(t_l)=\begin{cases}L,\ k_1=k_2,\\0,\ k_1\neq k_2.\end{cases} \tag{8}$$

Опишемо $a_0(t),a_1(t),a_2(t)$ таким чином:

$$a_0(t) = \sum_{i=0}^{d} b_{0i}\psi_i(t),$$
$$a_1(t) = \sum_{i=0}^{d} b_{1i}\psi_i(t), \qquad (9)$$
$$a_2(t) = \sum_{i=0}^{d} b_{2i}\psi_i(t).$$

Із використанням результатів апроксимації (6) – (7) введемо набори значень цих функцій на сукупності аргументів $t_1, t_2, ..., t_L$:

$$A_0 = \begin{pmatrix} a_{01} \\ a_{02} \\ ... \\ a_{0L} \end{pmatrix}, \ A_1 = \begin{pmatrix} a_{11} \\ a_{12} \\ ... \\ a_{1L} \end{pmatrix}, \ A_2 = \begin{pmatrix} a_{21} \\ a_{22} \\ ... \\ a_{2L} \end{pmatrix}.$$

Тепер уведемо матрицю

$$\psi = \begin{pmatrix} \psi_0(t_1) & \psi_1(t_1) & ... & \psi_d(t_1) \\ \psi_0(t_2) & \psi_1(t_2) & ... & \psi_d(t_2) \\ ... & ... & ... & ... \\ \psi_0(t_L) & \psi_1(t_L) & ... & \psi_d(t_L) \end{pmatrix}$$

і вектори

$$B_0^T = \begin{pmatrix} b_{00} & b_{01} & ... & b_{0d} \end{pmatrix}, \ B_1^T = \begin{pmatrix} b_{10} & b_{11} & ... & b_{1d} \end{pmatrix}, \ B_2^T = \begin{pmatrix} b_{20} & b_{21} & ... & b_{2d} \end{pmatrix}.$$

Шукані вектори B_0, B_1, B_2 знайдемо мінімізуючи функціонали:

$$M_k = (\psi B_k - A_k)^T (\psi B_k - A_k), \quad k = 1,2,3.$$

За цих умов

$$\hat{B}_k = \left(\psi^T \psi\right)^{-1} \psi^T A_k.$$

Оскільки, в силу (8) $\left(\psi^T \psi\right)^{-1} = \frac{1}{L} I$, де I – одинична матриця, то

$$\hat{B}_k = \frac{1}{L} \begin{pmatrix} \sum_{l=1}^{L} a_{kl}\psi_0(t_l) \\ \sum_{l=1}^{L} a_{kl}\psi_1(t_l) \\ ... \\ \sum_{l=1}^{L} a_{kl}\psi_d(t_l) \end{pmatrix}, \quad k = 1,2,3. \qquad (10)$$

Підставляючи (10) в (9), обчислимо значення параметрів a_{0l}, a_{1l}, a_{2l} для будь-якого t. Тепер, використовуючи (6), вирахуємо емпіричний закон розподілу ймовірностей діагнозів на момент t_{np} прогнозу.

Отже, задача прогнозування вектора ймовірностей станів системи розв'язана.

Висновки

У статті запропоновано методику використання ЕС з нейромережевим механізмом логічного виводу для прогнозування стану діагностуємого об'єкта. Розглянуто два альтернативних підходи до розв'язання задачі. Один з них обґрунтований на прогнозуванні поведінки контрольованих параметрів. Другий підхід реалізує запропоновану методику безпосереднього прогнозування розподілів ймовірностей станів відмови.

СПИСОК ЛІТЕРАТУРИ

1. Уотермен Д. Руководство по экспертным системам: Пер. с англ. – М.: МИР, 1989. – 388 с.
2. Нейлор К. Как построить свою экспертную систему: Пер с англ. – М.: Энергоиздат, 1991. – 286 с.
3. Миненкова З. Е. Комбинированный механизм логического вывода байесовой диагностической экспертной системы // Вестник ХПИ. – 2003. – № 6. – С. 69-74.
4. Серая О. В. Модели и информационные технологии оценки и прогнозирования состояния многомерных динамических объектов в условиях нечетких входных данных: Дис. канд. техн. наук: 05.13.06. – Х., 2001. – 251 с.

Сєрая Оксана Володимирівна – к. т. н., доцент кафедри економічної кібернетики і маркетингового менеджменту, тел. (8057)-707-66-28.

Національний технічний університет "Харківський політехнічний інститут".

Каткова Тетяна Ігорівна – к. п. н., доцент, завідувач кафедри математики і математичних методів, тел. (06153)-71971.

Бердянський університет менеджменту і бізнесу.