O. V. Seraya, Cand. Sc. (Eng), Ass. Prof.; T. I. Katkova, Cand. Sc. (Pedagogic), Ass. Prof.

NEURAL NETWORK PRODUCTIONAL SYSTEM FOR STATE DIAGNOSTICS

Expert system with neural network mechanism of inference is suggested. Procedure has been developed for predicting the state of the object diagnosed by such mechanism.

Key words: *expert system, productional rules, neural network structures, inference mechanism, objects state prediction.*

Introduction

Modern information technologies for object state diagnostics make use of the specific systems of artificial intellect – expert systems (ES). Such systems are organized so that the results of measuring a set of object controlled parameters $x_1, x_2, ..., x_n$ can be used for its state diagnostics.

Operation principles and structure of the diagnostic expert system depend, to a great extent, on the type of the inference mechanism (IM). In practice two essentially different approaches to IM construction are used: productional and Bayesian. Productional approach, that has found wider application, is based on the system of the so called productional rules constructed in the following way [1]:

If x_1 is A_1, x_2 is $A_2, ..., x_n$ is A_n, then with the probability $p(A_1, A_2, ..., A_n)$ the object is in the state $H(A_1, A_2, ..., A_n)$.

Production systems are convenient for practical application and simple in realization. However, they have an essential disadvantage: they are constructed so that that there exists a dependence of the diagnostic system efficiency on the dimensionality of the controlled parameter. Diagnostic system must be complete, i. e. for each possible variant of the controlled parameters set a corresponding productional rule must exist. This means: if each parameter can adopt one of m possible values, then the total number of productional rules will be $N = m^n$ and will increase with the growth of m and n.

Bayesian inference mechanism [2] practically excludes the dimensionality problem. However, practical capabilities of such systems are limited due to the necessity of statistical independence of the controlled parameters. The efforts to avoid this problem were made in [3]. Here the dependent parameters are joined in groups. Then, parameters, included in one group, are processed using productional rules. The obtained results are already practically independent and used in Bayesian technology. Such ES with combined IM makes it possible to solve diagnostics problems for high dimensionality of the controlled parameters set and their possible correlation. At the same time it has common principle disadvantage of all ES – discrete character of the controlled parameters. Both in productional and in Bayesian systems each parameter is either a symptom with Boolean character or manifestation, or if it is a continuous parameter, its possible values range must be divided into subranges, i.e. discretized. This factor causes a number of problems in practical development of ES. First, if a number of subranges is large, the choice of a rational number of subranges must result from the non-trivial compromise between IM complexity and state evaluation accuracy. Second, the boundaries of subranges are difficult for theoretical explanation. The very existence of boundaries can alone lead to the unnatural situation when different diagnoses correspond to two parameter sets that are similar as to their numerical values.

All these problems can be excluded if IM is constructionally adapted to processing of the parameters that are discrete by their nature. Such IM can be implemented with the application of

artificial neural networks (ANN). Let's formulate the problem of the development of ES system with neural network inference mechanism designed for the evaluation and prediction of the diagnosed object state.

Problem statement

As it is known, any ANN performs mapping of the points from multidimensional observation space X with n dimensionality onto the points of the multidimensional solution space Y that in general case has another dimensionality H. Here, correct mapping of the points from X onto Y is provided by a specially arranged procedure of the network training. Assume that a series of measurements of the object controlled parameters is performed. As a result, sets $X_1 = (x_{11}, x_{12},..., x_{1j},..., x_{1n})$, $X_2 = (x_{21}, x_{22},..., x_{2j},..., x_{2n}),\ldots,\quad X_l = (x_{l1}, x_{l2},..., x_{lj},..., x_{ln}),\ldots,\quad X_L = (x_{L1}, x_{L2},..., x_{Lj},..., x_{Ln})$ are obtained. In the course of network training these sets are presented to Q experts that for each X_l, $l = 1,2,...,L$ set assign distribution of probabilities $P_{lq} = (p_{lq1}, p_{lq2},..., p_{lqh},..., p_{lqH})$, $l = 1,2,...,L$, $q = 1,2,...,Q$ of $h = 1,2,...,H$ of object states. Elementary statistical processing of the expert evaluation results matches each measurement set X_l with the distribution of average values of state (range) probabilities $\hat{P}_l$ and a set of probability dispersions $\hat{\sigma}_l^2$. The obtained data are used for ANN training. After this, for each new vector of controlled parameter measurements neural network forms corresponding vectors $P(X)$ and $\sigma^2(X)$. The question about the possibility of ANN application for object state prediction remains to be a problematic one. Let's consider two alternative procedures of this problem solution.

Main results

A. Microapproach. As trained ANN for each observation vector of controlled parameters determines corresponding distribution of diagnoses probabilities, the solution of the object state prediction problem can be obtained by the observation vector prediction.

Sets $X_1, X_2,..., X_n$ are used for the observation matrix formation:

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

The columns of this matrix correspond to the readings of each controlled parameter for the time points $t_1, t_2,..., t_L$. Let's introduce the model of parameters evolution in time:

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

Calculation of (c_{ij}) parameters of the model (1) is performed by the least-squares method using matrix

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

and vectors

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

Now the set of vectors of regression equations parameter estimates is found by minimization of functionals:

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

Optimal (in terms of least-squares method) vectors of regression coefficients are found by the formulas:

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

Substitution of (2) into (1) gives a set of analytical descriptions of controlled parameters in time, which provides the possibility to calculate their values at t_{np} moment of the prediction:

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

Trained ANN will match the obtained set of parameter values

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

with the distribution of the object state probabilities at the moment of prediction.

B. Macroapproach. By means of entering observation sets $x_1, x_2,..., x_n$ sequentially to ANN input, the corresponding distributions of diagnoses probabilities will be obtained:

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

The law of probability variations for each range, e. g. for h-range, $h = 1,2,...,H$, is described by the corresponding time function $P_h(t)$, that can be represented by expansion into series in terms of a certain set of basic functions in accordance with the model:

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

where $\varphi_0(t), \varphi_1(t), ..., \varphi_m(t)$ is a set of basic functions.

It should be noted that standard independent evaluation of the parameters of the model (4) for each diagnosis separately is impossible as in this case normalization condition will not be taken into account: the sum of diagnoses probabilities at any time point must be equal to one.

In this connection let's consider another approach.

Simple technology for the application of the set of probability distributions (3), corresponding to the observation moments $t_1, t_2,..., t_L$, for the calculation of distributions at the prediction moment consists in the following.

We approximate each of the distributions (3) of the continuous curve (3) representing the density of diagnosis probabilities distribution and depending on q parameters. Now for each of the sets $\left(\theta_1(t_1), \theta_1(t_2),..., \theta_1(t_L)\right)$, $\left(\theta_2(t_1), \theta_2(t_2),..., \theta_2(t_L)\right)$,…, $\left(\theta_q(t_1), \theta_q(t_2),..., \theta_q(t_L)\right)$ we build analytical

continuation and calculate the parameter set $\theta_1(t_{np}),\theta_2(t_{np}),...,\theta_q(t_{np})$, that defines unambiguously the desired distribution of the diagnosis probabilities at t_{np} moment of prediction. Unfortunately, this elementary procedure can turn out to be extremely labor-consuming because the character of each distribution (3) could be so complicated that would require an unacceptably large number of q parameters for its adequate approximation or otherwise would be inadmissibly rough.

The alternative is to use for the description of distributions (3) the laws of empirical distribution, their principal feature being monotonous non-decreasing. For the evaluation of empiric distribution laws the functional can be used

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

where $y(t)$ provides the empirical law $g(t)$ adaptation to real data. In [4] it is shown that quite acceptable (in terms of quality) approximation of real empirical distribution laws can be provided by the description of $y(t)$ in the form of quadratic polynomial. General prediction scheme will be as follows. A distribution set $P_1(t_1),P_2(t_2),...,P_L(t_L)$ is transformed into the empiric integral laws:

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

where

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

Then each of the obtained integral laws is approximated independently using functional (5), presented in the form of

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

Approximation quality index has the form of

$$W_l=\sum_{h=1}^{H}\big(R_{lh}-g_l(h)\big)^2.$$

Minimization (7) is performed by least-squares method. As a result, a number of sets $\{(a_{01},a_{11},a_{21}),(a_{02},a_{12},a_{22}),...,(a_{0L},a_{1L},a_{2L})\}$ are obtained. These sets are used for receiving polynomial functions $a_0(t),a_1(t),a_2(t)$ that provide the possibility to describe approximation (6) for any time point.

Let's consider this procedure in detail. For the description of functions $a_0(t),a_1(t),a_2(t)$ we use the system of functions $\psi_1(t_l),\psi_2(t_l),...,\psi_d(t_l)$, orthogonal on the set of equidistant points $t_1,t_2,...,t_L$ so, that

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

We describe $a_0(t),a_1(t),a_2(t)$ in the following way:

$$\begin{aligned}a_0(t)&=\sum_{i=0}^{d}b_{0i}\psi_i(t),\\a_1(t)&=\sum_{i=0}^{d}b_{1i}\psi_i(t),\\a_2(t)&=\sum_{i=0}^{d}b_{2i}\psi_i(t).\end{aligned} \tag{9}$$

Using the results of approximation (6) – (7), we introduce values sets of these functions on the set of arguments $t_1,t_2,...,t_L$:

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

Now we introduce matrix

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

and vectors

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

Vectors B_0, B_1, B_2 to be found are estimated by minimization of functionals:

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

Here

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

As due to (8), $\left(\psi^T \psi\right)^{-1} = \frac{1}{L} I$, where I is unit matrix, then

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

By substitution of (10) into (9), the values of a_{0l}, a_{1l}, a_{2l} for any t are calculated. Now, using (6), we calculate the empirical law of the diagnoses probability distribution for t_{np} moment of the prediction.

Thus, we have solved the problem of prediction of the system state probability vector.

Conclusions

In this paper method is suggested for using ES with neural network inference mechanism for predicting the state of the object being diagnosed. Two alternative approaches to the solution of this problem are considered. One of them is based on the prediction of the controlled parameters behavior. The other implements the suggested procedure of the direct prediction of failure state probabilities distribution.

REFERENCES

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. – 251с.

Seraya Oksana – Cand. Sc. (Eng), Ass. Prof., the department of economic cybernetics and marketing management. Phone: (8057)-707-66-28.

National Technical University “Kharkiv Polytechnic Institute”.

Katkova Tatyana – Cand. Sc. (Pedagogic), Ass. Prof., head of the department of mathematics and mathematical methods. Phone: (06153)-71971.

Berdiansk University of Management and Business.