

- тво НТУУ "Київський політехнічний інститут", 1998, - с. 9-15.
10. Архипов А.Е., Архипова С.А. Идентификация аппроксимативных моделей методом варьирования данных. // Адаптивные системы автоматического управления. // Межвидом. научно-техн. зб. - Днепропетровск: Системні технології, 1998. - Вип. 1 (21). с. 81-86.
 11. Архипов А.Е., Архипова С.А. Селекция структуры линейной регрессии путем случайного варьирования исходных данных. // Системні технології. Дослідження динаміки і оптимізація параметрів технологічних процесів: Збір. наук. праць. - Вип. 2 - Дніпропетровськ: "Системні технології", 1998. - с. 3-9.
 12. Архипов А.Е. Выбор, оценка качества и методы повышения точности аппроксимативных моделей в задачах идентификации. - К.: УМК ВО, 1992. - 56 с.
 13. Архипов А.Е. Коррекция смещений коэффициентов регрессии, обусловленных ошибками в независимых переменных. Тези доповідей 1-ої Української конференції з автоматичного керування. АВТОМАТИКА-94 (Київ, 12-23 травня 1994 р.) - Київ, 1994, с.123.

Надійшла 02.03.2000

Після доробки 10.03.2000

УДК 681.513.6

ПРОГНОЗИРУЮЩАЯ НЕЙРОННАЯ СЕТЬ И АЛГОРИТМЫ ЕЕ ОБУЧЕНИЯ

Е. В. Бодянский, С. В. Попов

Предлагается подход к решению задач прогнозирования и раннего обнаружения изменения свойств нестационарных стохастических последовательностей, описываемых нелинейными разностными уравнениями авторегрессии - скользящего среднего (NARMA-модели). Предполагается, что прогнозируемая последовательность является нестационарной в широком смысле, причем изменения структуры, свидетельствующие о возникшей разладке, происходят в заранее неизвестные моменты времени. Предлагается архитектура рекуррентной искусственной нейронной сети и новые алгоритмы ее обучения, позволяющие с высокой точностью и быстродействием получать прогнозы контролируемого сигнала и обнаруживать моменты изменения его свойств.

Пропонується підхід до вирішення задачі прогнозування і раннього виявлення змін властивостей нестационарних стохастичних послідовностей, що описуються нелінійними різницевими рівняннями авторегресії - ковзного середнього (NARMA-моделі). Припускається, що прогнозована послідовність є нестационарною в широкому розумінні, при чому зміни структури, що свідчать про виникаюче розладнання, відбуваються у заздалегідь невідомі моменти часу. Запропоновано архітектуру штучної нейронної мережі і нові алгоритми її навчання, що дозволяють з високою точністю та швидкодією отримувати прогнози контрольованого сигналу та виявляти моменти зміни його властивостей.

An approach to prediction and early properties changes detection problems solution for nonstationary stochastic sequences that are described by nonlinear difference autoregression - moving average equations (NARMA-models) is proposed. The forecasted sequence is assumed to be nonstationary in a wide sense, and structure changes pointing to emerging faults are assumed to occur in moments, unknown a priori. Architecture of an artificial neural network and its learning algorithms are proposed that enable obtaining high quality predictions of the monitored signal and detection of its properties changes moments.

ВВЕДЕНИЕ

Задача раннего обнаружения разладок достаточно часто встречается на практике и, прежде всего, в технической диагностике механизмов и машин и тесно связана с проблемой адаптивного прогнозирования нестационарных стохастических последовательностей. Для решения этой проблемы предложено множество подходов, свя-

занных в значительной мере с методами математической статистики, теории случайных процессов, адаптивной идентификации, однако жесткие предположения о стохастических свойствах рядов и их структуре, как правило, линейной, ограничивают возможности традиционного адаптивного подхода.

Последние годы характеризуются всплеском исследований в области теории и практики искусственных нейронных сетей, в том числе и для решения задач прогнозирования [1-9]. Предложенные нейронные сети, обеспечивая высокое качество прогнозирования, тем не менее, не способны обнаруживать и сигнализировать об изменениях свойств контролируемого сигнала, что совершенно необходимо в задачах априорной диагностики в реальном времени.

В данной работе предложена архитектура прогнозирующей рекуррентной искусственной нейронной сети и алгоритмы ее обучения, позволяющие не только получать прогноз нелинейной стохастической последовательности, но и обнаруживать моменты изменения ее свойств по ходу процесса обучения и прогнозирования. Предлагаемая сеть обладает высокой скоростью обучения и обеспечивает оптимальную точность прогноза на принятом классе моделей.

1 АРХИТЕКТУРА ПРОГНОЗИРУЮЩЕЙ НЕЙРОННОЙ СЕТИ

Архитектура предлагаемой прогнозирующей искусственной нейронной сети приведена на рис. 1 и представляет собой сеть элементарных нейронов, отличающихся видом функций активации и алгоритмами обучения, являющимися в общем случае рекуррентными процедурами безусловной или условной оптимизации.

Прогнозируемая стохастическая последовательность $\{z(k)\}$, $k = 1, 2, 3, \dots$ предварительно обрабатывается входным нейроном С таким образом, что все ее значения (прошлые и будущие) принадлежат интервалу $[-1, +1]$.

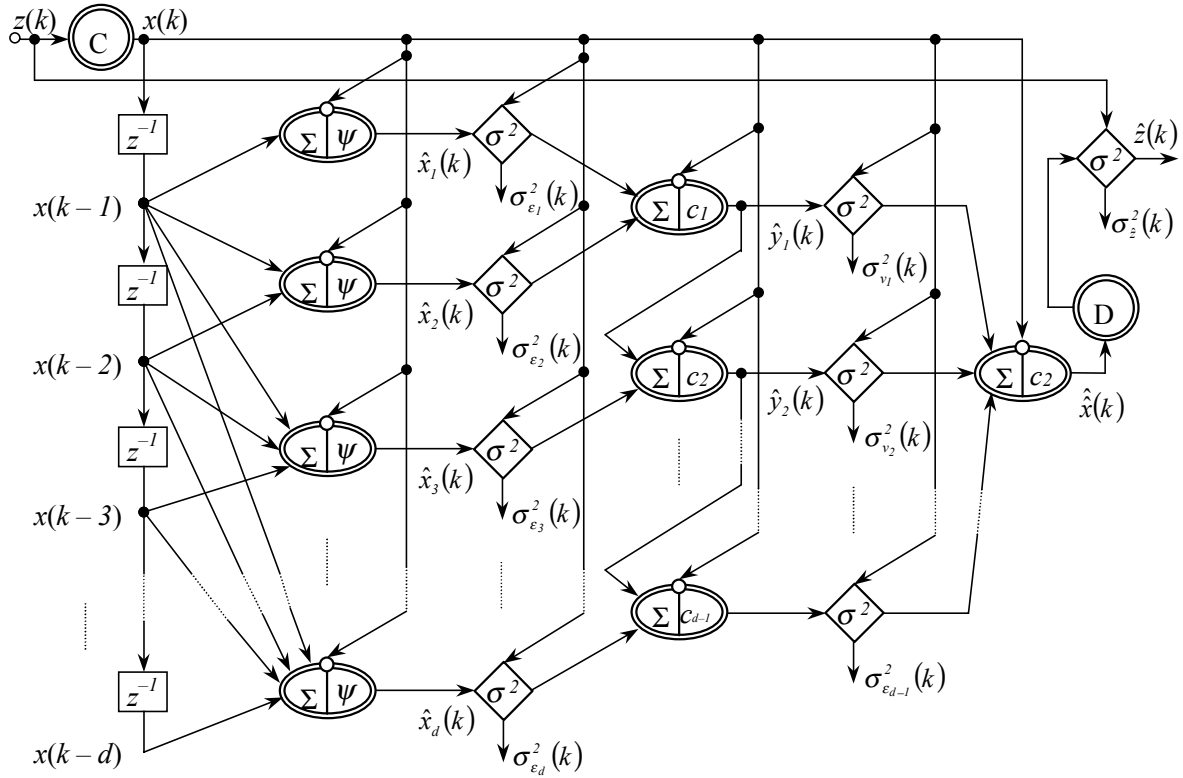


Рисунок 1 - Рекуррентная прогнозирующая нейронная сеть

Для этого в реальном времени вычисляются экстремальные значения последовательности $\{z(k)\}$.

$$\begin{cases} z^{\max}(k) = \max\{z^{\max}(k-1), z(k)\}, \\ z^{\min}(k) = \min\{z^{\max}(k-1), z(k)\}, \end{cases}$$

после чего сигнал кодируется так, что

$$x(k) = \frac{2z(k) - z^{\max}(k) - z^{\min}(k)}{z^{\max}(k) - z^{\min}(k)},$$

при этом всегда $-1 \leq x(k) \leq 1$.

Сигнал $\{x(k)\}$ далее подается на входной слой сети, образованный элементами чистого запаздывания z^{-1} : $z^{-1}x(k) = x(k-1)$, в результате чего на выходе этого слоя формируется набор задержанных значений $x(k-1)$, $x(k-2)$, ..., $x(k-d)$. При этом, чем больше значение d , тем более широкими возможностями обладает сеть.

Первый скрытый слой образован стандартными нейронами с нелинейной функцией активации, на входы которых подаются задержанные значения сигнала $x(k)$ и по цепи обратной связи каждого нейрона - задержанные значения прогноза $\hat{x}_j(k)$, $j = 1, 2, \dots, d$. Входы нейронов, обозначенные \circ , соответствуют входам обучающего сигнала, а ψ описывает функцию активации.

В результате обработки последовательности $\{x(k)\}$ нейронами первого скрытого слоя на их выходах появляются одношаговые прогнозы

$$\begin{cases} \hat{x}_1(k) = \psi(x(k-1), \hat{x}_1(k-1)) \\ \hat{x}_2(k) = \psi(x(k-1), x(k-2), \hat{x}_2(k-1), \hat{x}_2(k-2)) \\ \vdots \\ \hat{x}_d(k) = \psi(x(k-1), \dots, x(k-d), \hat{x}_d(k-1), \dots, \hat{x}_d(k-d)), \end{cases} \quad (1)$$

соответствующие нелинейному процессу авторегрессии - скользящего среднего (NARMA) порядка от 1 до d .

Второй скрытый слой образован однотипными элементарными нейронами, вычисляющими оценки точности прогнозов $\hat{x}_j(k)$ в виде дисперсий обновляющих последовательностей $\sigma^2_{\epsilon_j}(k)$.

Нейронами третьего скрытого слоя производится попарное объединение сигналов из второго слоя с целью получения оценок $y_j(k)$, $j = 1, 2, \dots, d-1$ таких, что

$$\begin{cases} y_j(k) = c_j(k)y_{j-1}(k) + (1-c_j(k))x_{j+1}(k), \\ \hat{y}_0(k) \equiv \hat{x}_1(k), \end{cases} \quad (4)$$

при этом весовые коэффициенты $c_j(k)$ характеризуют

оценку точности прогноза $y_j(k)$ по сравнению с $\hat{y}_{j-1}(k)$ и $x_{j+1}(k)$. Вектор весов $c(k) = (c_1(k), c_2(k), \dots, c_{d-1}(k))$ описывает качество прогнозов, получаемых в третьем слое, при этом оно, естественно, должно быть выше, чем в первом слое. Изменение соотношений между элементами этого вектора свидетельствует об изменении структуры модели сигнала, т.е. о возникшей разладке.

Четвертый слой полностью аналогичен второму и предназначен для оценки точности прогнозов $y_1(k)$,

$y_2(k), \dots, y_{d-1}(k)$, которые затем подаются на единственный нейрон пятого слоя, являющийся по структуре линейным ассоциатором и вычисляющий объединенный оптимальный прогноз $\hat{x}(k)$.

Нейрон, обозначенный D , предназначен для декодирования сигнала $\hat{x}(k)$ к исходной шкале

$$\hat{z}(k) = 0, 5(z^{\max}(k)(\hat{x}(k) + 1) - z^{\min}(k)(\hat{x}(k) - 1)). \quad (5)$$

И, наконец, последний нейрон сети вычисляет точность окончательного прогноза.

2 АЛГОРИТМЫ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

Выходной сигнал j -го нейрона первого скрытого слоя может быть представлен в виде

$$\begin{aligned} \hat{x}_j(k) = \psi \left(\sum_{i=1}^j w_{ji}(k)x(k-1) + \sum_{i=1}^j \hat{w}_{ji}(k)\hat{x}(k-1) + \right. \\ \left. + w_{j0}(k) \right) = \psi(w_j^T(k)X_j(k)) = \psi(u_j(k)), \end{aligned} \quad (6)$$

где $\psi(\cdot)$ - нелинейная функция активации, обычно сигмоидальная или гиперболического тангенса, а в нашем случае более простые в вычислительном отношении, но тем не менее удовлетворяющие всем необходимым условиям

$$\psi(u_i(k)) = \begin{cases} 1, 5u_i(k) - 0, 5u_i^3(k), \text{ или} \\ \sin \frac{\pi}{2} u_i(k), \text{ или} \\ \frac{2}{\pi} \arctan \gamma u_i(k), \end{cases} \quad (7)$$

$w_j(k) = (w_{j0}(k), w_{j1}(k), \dots, w_{jj}(k), \hat{w}_{j1}(k), \dots, \hat{w}_{jj}(k))^T -$
 $-(2j+1) \times 1$ - вектор настраиваемых синаптических весов, $X_j(k) = (1, x(k-1), \dots, x(k-j), \hat{x}_j(k-1), \dots, \hat{x}_j(k-j))^T$
 - вектор входов j -го нейрона, $u_j(k) = w_j^T(k)X_j(k)$, $j = 1,$

$2, \dots, d$, $k = 1, 2, 3, \dots$ - текущее дискретное время.

Вводя в рассмотрение ошибку прогнозирования j -го нейрона первого скрытого слоя

$$\epsilon_j(k) = x(k) - \hat{x}(k) = x(k) - \psi(u_j(k)), \quad (8)$$

запишем дельта-правило настройки синаптических весов

$$\begin{aligned} w_j(k+1) &= w_j(k) + \eta_j(k)\epsilon_j(k)\nabla_{w_j}\psi(u_j(k)) = \\ &= w_j(k) + \eta_j(k)\epsilon_j(k)G_j(k), \end{aligned} \quad (9)$$

где $\eta_j(k)$ - параметр шага поиска, принимаемый обычно постоянным, $\nabla_{w_j}\psi(u_j(k)) = G_j(k)$ - градиент функции активации по настраиваемым весам.

Сходимость алгоритма (9) обеспечивается в достаточно широком интервале варьирования параметра шага $\eta_j(k)$, однако естественным представляется выбор шага, обеспечивающего максимальную скорость. Можно показать [10], что в детерминированном случае максимальное быстроедействие обучения обеспечивается алгоритмом

$$w_j(k+1) = w_j(k) + \frac{x(k) - \hat{x}_j(k)}{\|G_j(k)\|^2} G_j(k) =$$

$$= \begin{cases} w_j k + \frac{xk - \hat{x}_j k}{1, 51 - u_j^2 k \|X_j k\|^2} X_j k, \text{ или} \\ w_j k + \frac{xk - \hat{x}_j k}{\frac{\pi}{2} \cos \frac{\pi}{2} u_j k \|X_j k\|^2} X_j k, \text{ или} \\ w_j k + \frac{xk - \hat{x}_j k}{2} X_j k. \end{cases} \quad (10)$$

Несложно видеть, что в линейном случае (10) совпадает с алгоритмом обучения Уидроу-Хопфа.

В случае, если контролируемый сигнал наблюдается на фоне помех, алгоритм обучения должен обладать как следящими, так и фильтрующими свойствами, для чего может быть использована процедура [11]

$$\begin{cases} w_j(k+1) = w_j(k) + r_j^{-1}(k)(x(k) - \hat{x}_j(k))G_j(k), \\ r_j(k) = \alpha r_j(k-1) + \|G_j(k)\|^2, 0 \leq \alpha \leq 1, r_j(0) = 0, \end{cases} \quad (11)$$

близкая по свойствам при $\alpha = 1$ к алгоритму стохастической аппроксимации Гудвина-Рэмеджа-Кэйнеса.

После каждого такта обучения нейронов первого скрытого слоя несложно получить многошаговый прогноз в виде

$$x_j(k+l) = \psi(w_j^T(k+l)X_j(k+l)), \quad (12)$$

где

$$\begin{cases} X_j(k+1) = (1, x(k), \dots, x(k-j+1), \hat{x}_j(k), \dots, \hat{x}_j(k-j+1))^T, \\ X_j(k+2) = (1, \hat{x}_j(k+1), x(k), \dots, x(k-j+2), \hat{x}_j(k+1), \dots, \\ \dots, \hat{x}_j(k-j+2))^T, \\ \vdots \\ X_j(k+j) = (1, \hat{x}_j(k+j-1), \dots, x(k), \hat{x}_j(k+j-1), \dots, \hat{x}_j(k))^T, \\ \vdots \\ X_j(k+l) = (1, \hat{x}_j(k+l-1), \dots, \hat{x}_j(k-l+1), \hat{x}_j(k+l-1), \dots, \\ \dots, \hat{x}_j(k-j+1))^T, \end{cases} \quad (13)$$

Во втором скрытом слое с помощью рекуррентных процедур

$$\sigma_{\varepsilon_j}^2(k) = \frac{k-1}{k} \sigma_{\varepsilon_j}^2(k-1) + \frac{1}{k} \varepsilon_j^2(k) \quad (14)$$

для стационарного и

$$\sigma_{\varepsilon_j}^2(k) = \alpha \sigma_{\varepsilon_j}^2(k-1) + (1-\alpha) \varepsilon_j^2(k), 0 \leq \alpha < 1 \quad (15)$$

для нестационарных сигналов вычисляется точность полученных прогнозов.

Попарное объединение прогнозов в третьем скрытом слое в виде (4) призвано обеспечить более высокое качество прогнозирования, при этом обучение нейронов этого слоя, являющихся двухходовыми линейными ассоциаторами, осуществляется с помощью процедуры [12]

$$\begin{cases} c_j(k+1) = \frac{s_j(k)}{s_j(k+1)} c_j(k) + \frac{\varepsilon_j(k+1)(\hat{y}_{j-1}(k+1) - \hat{x}_{j+1}(k+1))}{s_j(k+1)}, \\ s_j(k+1) = s_j(k) + (\hat{y}_{j-1}(k+1) - \hat{x}_{j+1}(k+1))^2, \\ c_j(0) = 0, 5. \end{cases} \quad (16)$$

В четвертом скрытом слое на основе ошибки $v_j(k) = x(k) - y_j(k)$ и формул типа (14), (15) вычисляются дисперсии $\sigma_{v_j}^2(k)$, при этом можно показать [12], что

$$\sigma_{v_j}^2(k) \leq \min\{\sigma_{v_{j-1}}^2(k), \sigma_{\varepsilon_{j+1}}^2(k)\}, \quad (17)$$

т.е. прогнозы, получаемые в третьем скрытом слое по точности не уступают лучшим прогнозам первого скрытого слоя.

Многошаговый прогноз аналогично (12) может быть записан в виде

$$y_j(k+1) = c_j(k+1)y_{j-1}(k+l) + (1-c_j(k+1))x_j(k+l). \quad (18)$$

В пятом скрытом слое сети, образованном $(d-1)$ -входовой адалиной, производится объединение прогнозов в виде

$$\hat{x}(k) = \sum_{j=1}^{d-1} c_j^0(k) \hat{y}_j(k) = \hat{y}^T(k) c^0(k), \quad (19)$$

где вектор весов $c^0(k) = (c_1^0(k), c_2^0(k), \dots, c_{d-1}^0(k))^T$ отве-

чает условию несмещенности, $\sum_{j=1}^{d-1} c_j^0(k) = E^T c^0(k) = 1$,

$y(k) = (y_1(k), y_2(k), \dots, y_{d-1}(k))^T$, $E = (1, 1, \dots, 1)^T$. Требование оптимальности прогноза $\hat{x}(k)$ приводит к алгоритму обучения вида [13]:

$$\begin{cases} P(k+1) = P(k) - \\ - (P(k) \hat{y}(k+1))(P(k) \hat{y}(k+1))^T \times \\ \times (1 + \hat{y}^T(k+1) P(k) \hat{y}(k+1))^{-1}, \\ c^*(k+1) = c^*(k) + \\ + P(k+1)(x(k+1) - \hat{y}^T(k+1) c^*(k)) \hat{y}(k+1), \\ c^0(k+1) = c^0(k) - \\ - P(k+1)(E^T P(k+1) E)^{-1} (E^T c^*(k+1) - 1) E, \\ c_j^* = (d-1)^{-1}, \end{cases} \quad (20)$$

где $c^*(k+1)$ - оценка стандартного рекуррентного метода наименьших квадратов.

Многошаговый прогноз этого слоя имеет вид

$$\hat{x}(k+1) = \hat{y}^T(k+l) c^0(k+1) \quad (21)$$

и после декодирования нейроном D может быть преобразован в прогноз исходной последовательности $z(k+l)$.

На выходе сети производится оценка точности объединенного прогноза в виде

$$\sigma_z^2(k) = \frac{k-1}{k} \sigma_z^2(k-1) + \frac{1}{k} (z(k) - \hat{z}(k))^2 \quad (22)$$

или

$$\sigma_z^2(k) = \alpha \sigma_z^2(k-1) + (1-\alpha) (z(k) - \hat{z}(k))^2, 0 \leq \alpha < 1, \quad (23)$$

при этом можно показать [13], что

$$\hat{\sigma}_z^2(k) \leq \min\{\min \sigma_{\varepsilon_j}^2(k), \min \sigma_{v_j}^2(k)\}, \forall j, \quad (24)$$

т.е. прогноз на выходе сети является действительно оптимальным на классе всех NARMA-моделей порядка от 1 до d .

ЗАКЛЮЧЕНИЕ

Предлагаемая искусственная нейронная сеть является специализированной вычислительной структурой, предназначенной для решения задач прогнозирования и раннего обнаружения изменения свойств стохастических по-

следовательностей, описываемых различными NARMA-уравнениями с неизвестными параметрами. В результате обучения сеть обеспечивает оптимальное на классе данных моделей прогнозирование, при этом, если вместо нейронов первого скрытого слоя использовать стандартные структуры прогнозирующих сетей типа MLP, RBFN, SOM и т.п., можно говорить об оптимальном на ансамбле нейросетей [14] прогнозе. С вычислительной точки зрения подобная метанейросеть не намного сложнее рассмотренной выше, поскольку программы, реализующие стандартные архитектуры, входят в состав многих коммерческих пакетов прикладных программ [15].

ПЕРЕЧЕНЬ ССЫЛОК

1. Wong F.S. Time series forecasting using backpropagation neural networks // Neurocomputing. - 1990/91. - 2. - P. 147-159.
2. Groot de C., Wuertz D. Analysis of univariate time series with connectionist nets: A case study of two classical examples // Neurocomputing. - 1991. - 3. - P. 177-192.
3. Connor J.T., Martin R.D., Atlas L.E. Recurrent neural networks and robust time series prediction // IEEE Trans. Neural Networks. - 1994. - 5. - N2. - P. 240-254.
4. Saxen H. Nonlinear time series analysis by neural networks. A case study // Int. J. Neural Systems. - 1996. - 7. - N2. - P.

5. 195-201.
5. Madhavan P.G. A new recurrent neural network learning algorithm for time series prediction // J. of Intelligent Systems. - 1997. - 7. - N1, 2. - P. 103-116.
6. Yu H.-Y., Bang S.-Ya. An improved time series prediction by applying the layer-by-layer learning method to FIR neural networks // Neural Networks. - 1997. - 10. - N9. - P. 1717-1729.
7. Nie J. Nonlinear time-series forecasting: A fuzzy-neural approach // Neurocomputing. - 1997. - 16. - P. 63-76.
8. Billings S.A., Hong X. Dual-orthogonal radial basis function networks for nonlinear time series prediction // Neural Networks. - 1998. - 11. - P. 479-493.
9. Conway A.J., Macpherson K.P., Brown J.C. Delayed time series prediction with neural networks // Neurocomputing. - 1998. - 18. - P. 81-89.
10. Бодянский Е.В. Адаптивные алгоритмы идентификации нелинейных объектов управления // АСУ и приборы автоматики. - 1987. - Вып. 81. - С. 43-46.
11. Бодянский Е.В. Обнаружение разладок в нелинейных стохастических последовательностях с помощью рекуррентных искусственных нейронных сетей // Проблемы бионики. - 1998. - Вып. 49. - С. 23-33.
12. Bodyanskiy Ye.V., Vorobyov S.A., Stephan A. Detection of NARMA-sequence order using recurrent artificial neural networks // Proc. of European Control Conference ECC'99. - Karlsruhe, Germany, 1999. - CD-ROM.
13. Бодянский Е.В. Автоматическое обнаружение разладок с помощью искусственной нейронной метасети // Проблемы бионики. - 1998. - Вып. 49. - С. 34-38.
14. Sharkey A.J.C. On combining artificial neural nets // Connection Science. - 1996. - 8. - N3, 4. - P. 299-313.
15. Braun H., Feulner J., Malaka R. Praktikum Neuronale Netze. Berlin: Springer-Verlag, 1996. - 242 P.

Надійшла 06.03.2000

Після доробки 10.03.2000

УДК 621.3.049.77.002:519.24

МЕТОД ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ГРАНИЧНОГО ВЫБОРОЧНОГО КОНТРОЛЯ

А. Ю. Долгов

Предлагается метод повышения эффективности прогнозирования (оценок) величины брака в принятых пластинах при граничном выборочном контроле качества кристаллов интегральных микросхем.

The method of increase of a forecasting effectiveness (ratings) of a spoilage size in the accepted plates at boundary selective quality surveillance of integrated microcircuits crystals is offered.

Граничный контроль [1] является основным методом контроля качества при производстве кристаллов интегральных микросхем. Основой метода является выборка малого объема (обычно $n=5$, иногда $n=10$), по которой следует судить о качестве кристаллов на пластине, содержащей от 400 до 4000 потенциально годных изделий. Такой малый объем выборки связан с количеством тестовых ячеек на пластине, которые имеют другую структуру, чем рабочие ячейки, и количество которых не может быть увеличено без ущерба для выхода годных. Но этот малый объем контрольной выборки исключает применение классических методов статистического контроля по количественному признаку [2], хотя сам по себе является достаточно эффективным. Так при реша-

ющем правиле "5 из 5" (т.е. при попадании всех пяти измеренных величин в норму по конструкторской документации) точность прогноза выхода годных, т.е. величина возможного брака на принятых пластинах, колеблется от 0 до 69 %, а при решающем правиле "3 из 5" - от 15 до 85% [1]. Такие точности не могут удовлетворить производство, однако метод остается по-прежнему востребованным, так как не имеется другого столь же простого в применении на практике, а для классических методов контроля, как уже упоминалось, не хватает объема выборки.

Классический метод прогнозирования состоит в том, что определяются квантили граничных значений по нормам конструкторской документации относительно среднего арифметического контрольной выборки, нормированные среднеквадратическим отклонением этой же выборки. По квантилям с помощью закона распределения Стюдента определяются вероятности, лежащие вне границ нормы, которые затем складываются в прогнозируемый брак. Анализ этого метода прогнозирования брака показал, что слишком большой разброс между минимальным и максимальным значениями брака при одних