

УДК 681.518.2:519.816

Обучаемые искусственные нейронные сети в построении формальных математических моделей систем при априорной неопределенности данных

Е. М. Угрюмова

Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Украина

Предложены методология и компьютерная система поддержки принятия решений для формирования формальных математических моделей на основе статистических моделей систем. Разработан метод оценивания параметров статистических моделей систем на базе обучаемых искусственных нейронных сетей на основе метода стохастической аппроксимации путем применения регуляризирующего последовательного (адаптивного) алгоритма синтеза решений с отложенной коррекцией.

Ключевые слова: компьютерная система поддержки принятия решений, обучаемые искусственные нейронные сети, метод стохастической аппроксимации.

Запропоновані методологія і комп'ютерна система підтримки прийняття рішень для формування формальних математичних моделей на основі статистичних моделей систем. Розроблений метод оцінювання параметрів статистичних моделей систем на базі штучних нейронних мереж, що навчаються, на основі методу стохастичної апроксимації шляхом вживання послідовного (адаптивного) алгоритму, що регуляризує, синтезу рішень з відкладеною корекцією.

Ключевые слова: комп'ютерна система підтримки прийняття рішень, штучні нейронні мережі, що навчаються, метод стохастичної апроксимації.

Methodology and decisions support computer system for forming of formal mathematical models on the basis of statistical models of the systems are offered. The method of parameters evaluation of the systems statistical models on the foundation of the learning artificial neural networks on the basis of method of stochastic approximation by application of regularization sequential (adaptive) algorithm of synthesis of solutions with the lazy correction is developed.

Key words: decisions support computer system, learning artificial neural networks, method of stochastic approximation.

1. Постановка проблемы и её актуальность

Объективная и субъективная неопределенности являются атрибутами процессов принятия решений, которые основываются, как правило, на использовании исходных, либо формальных математических моделях рассматриваемых физических процессов, на всех этапах жизненного цикла технических систем (ТС). Вследствие «кажущейся» простоты, широкое применение в компьютерных системах поддержки принятия решений (КСППР) в автоматизированных системах, предназначенных для решения широкого круга задач в разных областях практической деятельности, получили формальные математические модели (ФММ) на основе статистических моделей. Вытекающие из современных тенденций и особенностей развития ТС требования к качеству и срокам их создания в условиях экономии сырьевых и ограничения финансово-производственных ресурсов, обуславливают актуальность разработки методов оценивания параметров статистических

моделей ТС, обеспечивающих при априорной неопределенности входных данных стабильность (робастность) и информативность параметров ФММ.

При структуризации задач формирования ФММ ТС на основе дискретных данных об аналогах возникает ряд неопределенностей:

– концептуальная неопределенность – неопределенность выбора типа ФММ, используемого в дальнейшем для определения ее структуры и параметров;

– методологическая неопределенность – неопределенность выбора постановки задачи синтеза оптимальных параметров ФММ, сводящейся, как правило, к многокритериальной задаче параметрической оптимизации (МЗПО), метода и алгоритма ее численного решения.

Процесс раскрытия этих неопределенностей сводится к разработке методологии решения задачи структурно-параметрической оптимизации, включающей методы формализации представления векторной функции векторной переменной и построения решений МЗПО. Следует отметить, что задачи формирования ФММ на основе дискретных данных об аналогах (аппроксимации данных) в общем случае относятся к числу некорректных задач. Таким образом, возникает необходимость разработки устойчивых к возмущениям входных данных и погрешностям вычислений математических методов для их решения.

К настоящему времени опубликовано множество работ, посвященных описанию теории, методов обучения, практике применения обучаемых искусственных нейронных сетей (ИНС) в различных областях науки и техники [1-18]. Среди приведенных публикаций следует выделить работы [11, 12, 15-17], в которых достаточно подробно описаны не только общие положения теории ИНС, но и недостатки существующих методов обучения и проблемы, связанные с применением известных типов ИНС для решения практических задач. Следует отметить, в большинстве работ, посвященных решению задач формирования ФММ на базе обучаемых ИНС, отсутствует анализ устойчивости решений к возмущениям входных данных и погрешностям вычислений.

Данная работа посвящена исследованию некорректных задач восстановления многомерных статистических зависимостей на основе эмпирических данных, построению нейросетевых моделей (на примерах однонаправленной многослойной и радиально-базисной обучаемых ИНС) с помощью интеллектуальных информационных технологий. Представлены примеры практического использования построенных таким образом ФММ для решения задач анализа параметров ТС на базе реализующей представленную методологию КСППР «Concept_Pro®» [19].

2. Общая постановка задачи

Задана векторная функция набором обучающих пар $\left(\vec{y}^{(0)}, \vec{d} \right)_p$, $p = 1 \dots P$, где

$\vec{y}^{(0)}, \vec{d}$ – вектора входа, размерности H_0 , и выхода, размерности H_{K+1} , соответственно. Необходимо аппроксимировать данную выборку. Результатом решения задачи должен являться некий математический механизм, в результате работы которого можно было бы получить любое значение векторной функции,

представленной данной обучающей выборкой, по заданному вектору входа, в диапазоне, ограниченном исходными данными.

3. Метод обучения однонаправленных многослойных искусственных нейронных сетей (ОМС)

Однонаправленная многослойная обучаемая искусственная нейронная сеть (ОМС), используемая для аппроксимации данных, представляет собой параллельно распределенный процессор, который обладает способностью сохранения опытного знания и обработки информации между локальными процессорными элементами – нейроэлементами (нейронами), связанными между собой посредством специальных соединений – синаптических связей.

ОМС включает в себя три типа слоев нейронов:

- слой входных данных, которые известны из задачи;
- слои промежуточных данных, которые принимают значения соответствующих данных от предыдущих нейронов, формируют данные и передают их последующим слоям;
- слой выходных данных, которые требовалось получить в процессе обучения.

Запоминаемая информация, касающаяся данных, распределяется по сети в виде весовых параметров этих соединений, а развитие возможностей ОМС осуществляется путем обучения ОМС.

Исходными данными для аппроксимации данных с помощью ОМС являются:

- входные параметры и управляющие переменные опытных образцов (аналогов) $\{Y_{ph}^{(0)}\}$;

- выходные параметры $\{d_{pi}\}$.

Обычно все исходные данные предварительно приводятся к безразмерному виду. В нашем случае использовались прямое преобразование:

$$f^\circ = \frac{2(f - \langle f \rangle)}{(f_{\max} - f_{\min})}, \text{ где } \langle f \rangle = (f_{\max} + f_{\min})/2, f^\circ \in [-1,1],$$

обратное:

$$f = [(f_{\max} - f_{\min})f^\circ + (f_{\max} + f_{\min})]/2.$$

Алгоритм обучения ОМС с обратным распространением ошибки, заключается в нахождении значения ошибки между фактическим и желаемым выходными данными сети. Уменьшение значения ошибки можно добиться, модифицируя параметры сети. Процесс повторяется до достижения сетью способности выполнять желаемый тип преобразования «вход – выход». В результате обучения ОМС находятся веса связей слоев ОМС, а по входным параметрам и управляющим переменным разрабатываемого объекта – его выходные параметры.

Простейшая ОМС с одним скрытым слоем ($K = 1$) изображена на рис. 1. Здесь $\{Y_{ph}^{(0)}\}$ – множество входных данных, $\{Y_i^{(k)}\}$ – множество выходных данных k -ого слоя; k – номер слоя, $k = 1 \dots (K+1)$; $p=1 \dots P$, P – число аналогов;

$\{w_{ij}^{(k)}\}$ – множество весов k -ого слоя; i – элемент k -ого слоя; j – элемент $(k-1)$ -ого слоя. На рис. 1 введены следующие обозначения: H_0 – количество входов сети; H_1 – количество нейронов скрытого слоя; H_2 – количество выходов сети.

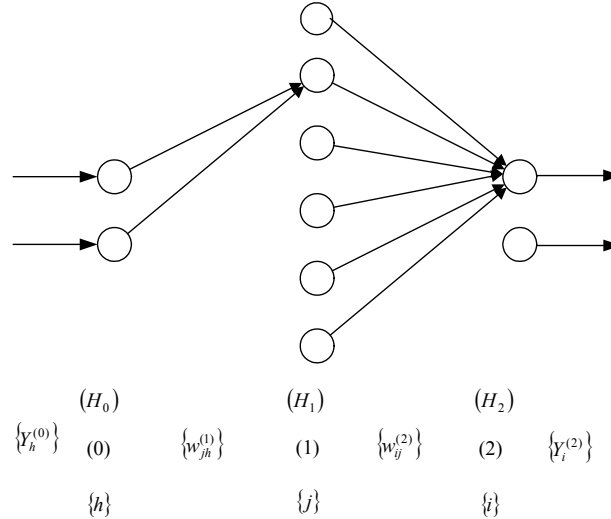


Рис. 1. Структура однонаправленной многослойной сети.

Для обеспечения стабильности (робастности) и информативности параметров статистических формальных математических моделей систем (процессов) на основе обучаемых искусственных нейронных сетей (ИНС) при априорной неопределенности входных данных, а также достаточной с практической точки зрения точности аппроксимации данных целесообразно в качестве методов обучения ИНС использовать стабильные (робастные) статистические методы оценки их параметров [1 – 4].

Пусть M – вектор случайных чисел размерности m (параметры модели), D – вектор случайных чисел размерности H_{K+1} (данные измерений).

Апостериорная плотность распределения вероятностей $\rho(M/D)$ по правилу Байеса имеет вид:

$$\rho(M/D) = \frac{\rho(D/M)\rho(M)}{\rho(D)}. \quad (3.1)$$

Если имеется P измерений вектора D , то можно принять гипотезу:

$$\rho_p(M/D_p) = \frac{\rho_p(D_p/M)\rho_{p-1}(M/D_{p-1})}{\rho(D_p)}, \quad p=1\dots P,$$

где $\rho_0(M/D_0) \equiv \rho(M)$.

Тогда апостериорная плотность распределения вероятностей вектора M после P измерений равна:

$$\rho_P(M/D_P) = \frac{\rho(D_P/M)\rho(M)}{\rho(D_P)},$$

где $\rho(D_P/M) = \prod_{p=1}^P \rho_p(D_p/M)$, $\rho(M) = \prod_{n=1}^m \rho_n(M_n)$, $\rho(D_P) = \prod_{p=1}^P \rho_p(D_p)$ –

константа, $m = \sum_{k=1}^{K+1} (H_{k-1} + 1)H_k$ – для ОМС ($m = 2H_0H_1 + (H_1 + 1)H_2$ – для РБС), $m \leq PH_{K+1}$.

Будем выбирать вектор \hat{M} согласно принципу максимума апостериорной плотности распределения вероятностей:

$$\hat{M}_{t+1} = \arg \sup_{M \in D_M} \rho_P(M/D_P). \quad (3.2)$$

Если принять, что $\rho(D_p/M) = Normal[\Delta_p/0, \sigma_D^2]$, $\rho_n(M_n) = Normal[\Delta_n/0, \sigma_M^2]$, где $Normal[f/0, \sigma_f^2]$ – нормальный закон распределения плотности вероятностей случайной величины f с математическим ожиданием “0” и дисперсией σ_f^2 , $\Delta_p = Y_p^{(K+1)}(\bar{Y}^{(0)}) - d_p(\bar{Y}^{(0)})$, $\Delta_n = M_{n,t+1} - M_{n,t}$, то задача (3.2) будет эквивалентна задаче

$$\hat{M}_{t+1} = \arg \inf_{M \in D_M} \left\{ \frac{1}{4PH_{K+1}} \sum_{p=1}^P \Delta_p^2(M) + \frac{\beta_t}{4m} \sum_{n=1}^m \Delta_n^2 \right\}, \quad (3.3)$$

где $\beta_t = \frac{\sigma_D^2}{\sigma_M^2}$, σ_D^2 , σ_M^2 – дисперсии случайных чисел, $t = 1 \dots T$ – номер эпохи обучения. Второе слагаемое в (3.3) обеспечивает выполнение принципа Уидроу – принципа минимального возмущения.

Определим меру информативности системы как ее среднюю энтропию:

$$S = \int s(M)F(dM), \quad (3.4)$$

где $s(M) = -\rho(M) \log_2 \rho(M)$, $F(dM)$ – априорная вероятностная мера параметров модели M . Если принять, что $\rho(M_{t+1}) \sim \exp[-\beta_t D_{KL}(M_{t+1}, M_t)]$, где D_{KL} – относительная энтропия Кульбака-Лейблера:

$$D_{KL}(M_{t+1}, M_t) = \rho(M_{t+1}) \log_2 \left[\frac{\rho(M_{t+1})}{\rho(M_t)} \right], \quad (3.5)$$

то вместо (3.3) получим:

$$\hat{M}_{t+1} = \arg \inf_{M \in D_M} \left\{ \frac{1}{4PH_{K+1}} \sum_{p=1}^P \Delta_p^2(M) + \beta_t D_{KL}(M_{t+1}, M_t) \right\}. \quad (3.6)$$

Будем использовать оценки:

$$D_{KL}(M_{t+1}, M_t) = \frac{1}{\sum_{k=1}^{K+1} (H_k + 1)} \sum_{k=1}^{K+1} \sum_{i=0}^{H_k} D_{KL,i}(M_{i,t+1}^{(k)}, M_{i,t}^{(k)}), \quad (3.7)$$

а плотность распределения вероятностей участия $i^{(k)}$ – нейрона в формировании величины $Y_i^{(k)}$ (см. формулу (3.5)) зададим распределением Коши:

$$\rho_i^{(k)}(M_i^{(k)}) \sim \frac{1}{1 + 1/\sum_{n=0}^{H_{k-1}} [w_{in}^{(k)}]^2}. \quad (3.8)$$

Следует отметить, что применение стабилизирующего оператора $D_{KL}(M_{t+1}, M_t)$ в (3.6) с использованием выражений (3.5, 3.7, 3.8) является одним из способов обеспечения гладкости аппроксимирующих дискретные данные поверхностей $Y_i^{(K+1)}(\bar{Y}^{(0)})$.

В дальнейшем в качестве функции выбора использовалась функция вида:

$$E = \frac{1}{8PH_{K+1}} \sum_{p=1}^P \gamma^{P-p} \sum_{i=1}^{H_{K+1}} f_{fit}[\Delta_{pi}^2(M)] + \frac{1}{2} \beta_t D_{KL}(M_{t+1}, M_t), \quad (3.9)$$

$$(\gamma = 0,95 \dots 0,99; \beta_t = 10^{-7}).$$

Решение – аппроксимирующую функцию вида $Y_i^{(K+1)}(\bar{Y}^{(0)})$ – будем искать методом стохастической аппроксимации. Коррекцию весов связей будем осуществлять по следующей формуле (представлен рекуррентный алгоритм обучения, соответствующий методу стохастической аппроксимации, обеспечивающий сходимость $w_{ij}^{(k)}(t) \xrightarrow{t \rightarrow \infty} \hat{w}_{ij}^{(k)}$ с вероятностью $P = 1$):

$$w_{ij}^{(k)}(t+1) = w_{ij}^{(k)}(t) + \mu(t) \{ \eta_{ij}^{(k)}(t) r_{ij}^{(k)}(t) + \alpha_{ij}^{(k)}(t) [w_{ij}^{(k)}(t) - w_{ij}^{(k)}(t-1)] \} + \tilde{w}_{ij}^{(k)}(t+1), \quad (3.10)$$

где $\mu(t) = \frac{\mu(0)}{[1 + (1 + \frac{t}{H_0 + H_1 + 1})^{-\lambda}]^t}$ – коэффициент обучения (определяемый в

соответствии с условиями Дворецкого: $\mu(t) > 0$, $\sum_{t=1}^{\infty} \mu(t) = \infty$, $\sum_{t=1}^{\infty} \mu^2(t) < \infty$);

$\lambda < 1$ (здесь принималось $\mu(0) = 1$, $\lambda = 0,5$); $t = 1 \dots P \cdot T$, T – количество эпох обучения;

$$r_{ij}^{(k)}(t) = S_{ij}^{(k)}(t) + \frac{(S_i^{(k)}(t))^T (S_i^{(k)}(t) - S_i^{(k)}(t-1))}{(S_i^{(k)}(t-1))^T S_i^{(k)}(t-1)} r_{ij}^{(k)}(t-1) \quad - \text{ проекции}$$

вектора направления поиска, определяемые в соответствии с методом сопряженных градиентов (Полака – Рибьера), $r_{ij}^{(k)}(0) = 0$;

$$S_{ij}^{(k)}(t) = -\frac{\partial E}{\partial w_{ij}^{(k)}} = -\gamma^{P-p} f'_{fit, \Delta_i} \delta_j^{(k)} Y_j^{(k-1)} - \frac{\beta w_{ij}^{(k)}}{\ln 2 \sum_{k=1}^{K+1} (H_k + 1)} \cdot \frac{(1 + \sum_{n=0}^{H_{k-1}} w_{in}^2)}{(1 + \sum_{n=0}^{H_{k-1}} w_{in}^2)^2} \left\{ 1 + \ln \left[\frac{\rho_i^{(k)}(M_{i,t}^{(k)})}{\rho_i^{(k)}(M_{i,t-1}^{(k)})} \right] \right\} \quad (3.11)$$

– проекции градиента;

$$\tilde{w}_{ij}^{(k)}(t+1) = \nu(t) w_{ij}^{(k)}(t) \Delta w^\circ (2\xi_{ij} - 1) \quad - \quad \text{«аддитивный шум»} \quad (\text{здесь}$$

$$\nu(t) = \mu(t) \sqrt{\frac{2}{\ln(2+t)}}; \quad \text{принималось} \quad \Delta w^\circ = 0,2; \quad \xi_{ij} \in [0,1] \quad - \quad \text{нормально}$$

распределенные случайные величины, имеющие нулевое значение среднего и единичную дисперсию). «Аддитивный шум» подключался к обучению, согласно представленной выше формуле, в двух случаях:

– в конце g эпох обучения (так называемый алгоритм глобальной «встряски» весов), если выполнялось условие:

$$\Delta E_g^\circ < \varepsilon^g, \quad \text{где} \quad \Delta E_g^\circ = \prod_{\tau=1}^g \Delta E_\tau^\circ, \quad \Delta E_\tau^\circ = \text{abs}(E_\tau^\circ - 1), \quad E_\tau^\circ = \sqrt{E_\tau / E_{\tau-1}}, \quad \tau = 1 \dots T;$$

– если $\text{abs}(\Delta w_{ij}^{(k)}(t)) < (1 - \varepsilon) \text{abs}(\Delta w_{ij}^{(k)}(t-1))$ (так называемый алгоритм локальной «встряски» весов). Принималось $\varepsilon = 0,03$.

В рассматриваемом случае использовалась функция вида:

$$f_{fit}(\Delta_{pi}^2) = 1 - \exp((-L_{fit} / 4) \Delta_{pi}^2), \quad L_{fit} \geq 4.$$

При обучении использовался регуляризирующий алгоритм, реализующий прерывания в итерационном процессе в случаях накопления ошибок вычислений (верхний индекс T в представленных ниже формулах означает операцию транспонирования вектора в строку):

$$\text{if } \frac{(S_i^{(k)}(t))^T (S_i^{(k)}(t) - S_i^{(k)}(t-1))}{(S_i^{(k)}(t-1))^T S_i^{(k)}(t-1)} \geq r_{\max} \quad \text{then } r_{ij}^{(k)}(t-1) = 0;$$

$$\text{if } (S_i^{(k)}(t-1))^T S_i^{(k)}(t) \geq r_{\min} (S_i^{(k)}(t))^T S_i^{(k)}(t) \quad \text{then } r_{ij}^{(k)}(t-1) = 0$$

(принималось: $r_{\max} = 5, r_{\min} = 0.2$);

$$\delta_j^{(k)} = \varphi_j^{(k)} \sum_{i=1}^{H_{k+1}} \delta_i^{(k+1)} w_{ij}^{(k+1)}; \quad \delta_i^{(K+1)} = -\Delta_i \varphi_i^{(K+1)};$$

$$\varphi_i^{(k)} = b(1 - f^2(Y_i^{(k)}));$$

$$Y_i^{(k)} = f(w_{io}^{(k)} + \sum_{j=1}^{H_{k-1}} w_{ij}^{(k)} Y_j^{(k-1)});$$

$k = 1 \dots (K+1); i = 1 \dots H_k; j = 1 \dots H_{k-1}; H_k$ – число элементов в k -ом слое;

$f(s) = \frac{e^{bs} - e^{-bs}}{e^{bs} + e^{-bs}}$ – передаточная функция; $\varphi(s)$ – производная передаточной функции.

Коэффициенты обучения и момента находились по следующим зависимостям:

$$\eta_{ij}^{(k)}(0) = \eta_{\max}; \quad \tilde{\alpha}_{ij}^{(k)}(t) = \frac{S_{ij}^{(k)}(t)}{S_{ij}^{(k)}(t-1) - S_{ij}^{(k)}(t)};$$

if $E_{\tau}^{\circ} > K_w$ and $S_{ij}^{(k)}(t)S_{ij}^{(k)}(t-1) < 0$ then $\eta_{ij}^{(k)}(t) = \rho_d \eta_{ij}^{(k)}(t-1)$
else $\eta_{ij}^{(k)}(t) = \rho_i \eta_{ij}^{(k)}(t-1); \alpha_{ij}^{(k)}(t) = \tilde{\alpha}_{mid};$
if $\eta_{ij}^{(k)}(t) < \eta_{\min}$ then $\eta_{ij}^{(k)}(t) = \eta_{\min}.$

Реализация условия: $E_{\tau+1} > E_{\tau}, E_{\tau} < E_{\tau-1}$ – представлена в форме:

if $E_{\tau}^{\circ} \leq K_w$ and $S_{ij}^{(k)}(t)(w_{ij}^{(k)}(t) - w_{ij}^{(k)}(t-1))\tilde{\alpha}_{ij}^{(k)}(t) < 0$ then $\alpha_{ij}^{(k)}(t) = \tilde{\alpha}_{\max}$
else $\alpha_{ij}^{(k)}(t) = \tilde{\alpha}_{ij}^{(k)}(t);$
if $\alpha_{ij}^{(k)}(t) > \tilde{\alpha}_{\max}$ then $\alpha_{ij}^{(k)}(t) = \tilde{\alpha}_{\max}.$

Принималось:

$$\rho_d = 0.7, \rho_i = 1.05, \eta_{\min} = 0.01, \eta_{\max} = 0.4 \dots 0.8, \tilde{\alpha}_{mid} = 0.01, \tilde{\alpha}_{\max} \leq 1.75, K_w = 1.04.$$

При переходе к новой эпохе обучения порядок предъявления обучающих пар $(\bar{Y}_p^{(0)}, \bar{d}_p)$, $p = 1 \dots P$, в рекуррентном алгоритме обновлялся (использовался генератор случайных чисел на интервале $[1 \dots P]$).

Инициализация начальных значений весов осуществлялась в диапазоне указанных ниже значений с помощью генератора случайных чисел:

- для скрытого слоя: $\pm \sqrt[H_0]{H_1}$;
- для выходного слоя: $\pm 0,5$.

Порядок операций для оценки выходных параметров включает следующие шаги.

1. Находим веса связей слоев ОМС.

2. Задаваясь входными параметрами и управляющими переменными разрабатываемого объекта, находим его выходные параметры.

Для проверки значимости (качества) предсказания данных ОМС могут использоваться следующие – усредненные, исходя из результатов 10 независимых запусков – величины: энергия среднеквадратичной ошибки:

$$E_{av} = \frac{1}{2P} \sum_{p=1}^P \sum_{i=1}^{H_{K+1}} \Delta_{pi}^2(M); \quad \text{средние относительные погрешности: } \delta_i^{\circ} = \frac{1}{P} \sum_{p=1}^P \left| \frac{\Delta_{pi}}{d_{pi}} \right|;$$

остаточные дисперсии: $S_{ocm,i}^2 = \frac{\sum_{p=1}^P \Delta_{pi}^2}{P-m-1}$; коэффициент множественной

корреляции: $R_i = \sqrt{1 - \frac{S_{ocm,i}^2}{S_{di}^2}}$, где S_{di}^2 – дисперсии средних $d_i(\bar{Y}^{(0)})$.

Далее сравнивались дисперсии средних S_{di}^2 с остаточными дисперсиями $S_{ocm,i}^2$ с помощью F_i – критерия Фишера: $F_i = \frac{R_i^2}{1-R_i^2} \cdot \frac{P-m-1}{m}$ с числом степеней свободы $\mathcal{G}_1 = m$ и $\mathcal{G}_2 = P-m-1$.

Считается, что ОМС аппроксимирует результаты опытов, если F_i достигают или превышают свои критические значения $F_{crt}(\phi, \mathcal{G}_1, \mathcal{G}_2)$ при выбранном уровне значимости (обычно принимают $\phi = 0.95$).

4. Метод обучения радиально-базисных искусственных нейронных сетей (РБС)

Рассмотрим алгоритм обучения радиально-базисных обучаемых искусственных нейронных сетей (РБС), предназначенных для интерполяции данных. В случае многомерной интерполяции с помощью РБС принимается, что количество нейронов скрытого слоя равно количеству тренировочных шаблонов. РБС для интерполяции данных применяется только в том случае, если количество обучающих пар $(\bar{Y}_p^{(0)}, \bar{d}_p)$, $p=1\dots P$, невелико, иначе (при большом количестве тренировочных шаблонов, а, соответственно, таком же количестве нейронов скрытого слоя) проблема обучения нейронной сети с математической точки зрения становится трудно разрешимой.

Структура простейшей РБС с одним скрытым ($K=1$) слоем аналогична представленной на рис. 1. Введем следующие обозначения:

$\bar{Y}^{(k)} = [Y_1^{(k)}, \dots, Y_{H_k}^{(k)}]^T$, $k = 0, 1, 2$ – вектор входных данных k -ого слоя;

$\bar{c}_j = [c_{j1}, c_{j2}, \dots, c_{jH_0}]^T$, $j = 1 \dots N_1$ – вектор координат центров активационной функции для нейронов скрытого слоя;

$\bar{\sigma}_j = [\sigma_{j1}, \sigma_{j2}, \dots, \sigma_{jH_0}]^T$, $j = 1 \dots N_1$ – вектор, задающий ширину окна активационной функции j -ого нейрона скрытого слоя;

$\varphi_j(\bar{Y}_p^{(0)}, \bar{c}_j, \bar{\sigma}_j) = \exp \left[-\frac{1}{2} \sum_{h=1}^{H_0} \frac{(Y_{ph}^{(0)} - c_{jh})^2}{\sigma_{jh}^2} \right] \equiv \varphi_{pj}$ – радиально-базисная

активационная функция нейрона скрытого слоя;

w_{ij} – вес связи между i -м нейроном выходного слоя и j -м нейроном скрытого слоя (здесь, согласно обозначениям, принятым на рис. 1, имеется в виду, что $w_{jh}^{(1)} = e_{jh} = 1$, $w_{ij}^{(2)} \equiv w_{ij}$).

Выберем размер скрытого слоя H_1 равным количеству тренировочных шаблонов P .

Разместим центры активационных функций нейронов скрытого слоя в точках пространства входных сигналов сети, которые входят в набор тренировочных шаблонов сети: $\bar{c}_j = \bar{Y}_j^{(0)}$, $j = 1 \dots P$.

Определим веса нейронов выходного слоя сети $\bar{w}_i = [w_{i1}, w_{i2}, \dots, w_{iH_1}]^T$, $i = 1 \dots H_2$, $j = 1 \dots P$. Для этого предъявим сети весь набор тренировочных шаблонов $\{\bar{d}_p\}$. Выход j -го нейрона выходного слоя для p -го шаблона будет равен:

$$\begin{aligned} Y_i^{(2)} &= w_{i1}\varphi(\bar{Y}_p^{(0)}, \bar{c}_1) + w_{i2}\varphi(\bar{Y}_p^{(0)}, \bar{c}_2) + \dots + w_{iP}\varphi(\bar{Y}_p^{(0)}, \bar{c}_P) = \\ &= w_{i1}\varphi(\bar{Y}_p^{(0)}, \bar{Y}_1^{(0)}) + w_{i2}\varphi(\bar{Y}_p^{(0)}, \bar{Y}_2^{(0)}) + \dots + w_{iH_1}\varphi(\bar{Y}_p^{(0)}, \bar{Y}_P^{(0)}) = d_{pi} \end{aligned} \quad (4.1)$$

Расписав это уравнение для всех выходов сети и всех шаблонов, получим следующее уравнение в матричной форме:

$$\Phi W^T = D,$$

$$\text{где } \Phi = \begin{bmatrix} \varphi_{11} & \dots & \varphi_{1P} \\ \varphi_{21} & \dots & \varphi_{2P} \\ \dots & \dots & \dots \\ \varphi_{P1} & \dots & \varphi_{PP} \end{bmatrix} \text{ – интерполяционная матрица, } \varphi_{pj} = f(\bar{Y}_p^{(0)}, \bar{Y}_j^{(0)});$$

$$W = \begin{bmatrix} w_{11} & \dots & w_{1P} \\ \dots & \dots & \dots \\ w_{H_21} & \dots & w_{H_2P} \end{bmatrix} \text{ – матрица линейных параметров РБС (весов}$$

выходного слоя);

$$D = \begin{bmatrix} d_{11} & \dots & d_{1H_2} \\ \dots & \dots & \dots \\ d_{P1} & \dots & d_{PH_2} \end{bmatrix} \text{ – матрица выходных шаблонов.}$$

Решение $W^T = \Phi^{-1}D$ даст нам искомые значения выходных синаптических весов, обеспечивающих прохождение интерполяционной поверхности через тренировочные шаблоны в пространстве выходных сигналов сети. Фактически необходимо решить систему уравнений отдельно для каждого выхода сети, вычислив весовые коэффициенты от нейронов скрытого слоя к данному нейрону выходного слоя. Получаем систему для i -го нейрона выходного слоя:

$$\begin{cases} w_{i1}\varphi(\bar{Y}_1^{(0)}, \bar{Y}_1^{(0)}) + w_{i2}\varphi(\bar{Y}_1^{(0)}, \bar{Y}_2^{(0)}) + \dots + w_{iP}\varphi(\bar{Y}_1^{(0)}, \bar{Y}_P^{(0)}) = d_{1i} \\ w_{i1}\varphi(\bar{Y}_2^{(0)}, \bar{Y}_1^{(0)}) + w_{i2}\varphi(\bar{Y}_2^{(0)}, \bar{Y}_2^{(0)}) + \dots + w_{iP}\varphi(\bar{Y}_2^{(0)}, \bar{Y}_P^{(0)}) = d_{2i} \\ \dots \\ w_{i1}\varphi(\bar{Y}_P^{(0)}, \bar{Y}_1^{(0)}) + w_{i2}\varphi(\bar{Y}_P^{(0)}, \bar{Y}_2^{(0)}) + \dots + w_{iP}\varphi(\bar{Y}_P^{(0)}, \bar{Y}_P^{(0)}) = d_{Pi} \end{cases}, \quad (4.2)$$

или в матричном виде:

$$[\varphi_{pj}] \begin{cases} w_{i1} \\ w_{i2} \\ \dots \\ w_{iP} \end{cases} = \begin{cases} d_{1i} \\ d_{2i} \\ \dots \\ d_{Pi} \end{cases},$$

где $\bar{Y}_p^{(0)}$ – вектор из набора тренировочных шаблонов.

Решение этой системы находилось методом квадратного корня [20]. Таким образом, находим матрицу весовых коэффициентов, если центры и ширина окон активационных функций уже заданы, в результате получаем обученную нейронную сеть.

Рассмотрим алгоритм обучения радиально-базисных обучаемых искусственных нейронных сетей, предназначенных для аппроксимации данных. Для обучения радиальной нейронной сети в случае, когда количество обучающих пар значительно превышает количество нейронов в скрытом слое $P \gg H_1$, применим гибридный алгоритм. В гибридном алгоритме процесс обучения разделяется на два этапа:

- подбор линейных параметров сети (весов выходного слоя) на основе использования метода псевдоинверсии;
- адаптация нелинейных параметров активационных функций (центров \bar{c}_j и ширины $\bar{\sigma}_j$ этих функций).

Оба этапа тесно переплетаются. При наличии P обучающих пар $(\bar{Y}_p^{(0)}, \bar{d}_p)$, $p=1..P$, и фиксации конкретных значений центров и ширины активационных функций (в первый момент это будут начальные значения) получаем систему уравнений ($P \gg H_1$):

$$\Phi \bar{w}_i = \bar{d}_i, \quad i=1..H_2,$$

где $\Phi = [\varphi_{pj}]$, $p=1..P, j=0..H_1; \varphi_{p0}=1$, $\bar{w}_i = [w_{i0}, w_{i1}, \dots, w_{iH_1}]^T$, $\bar{d}_i = [d_{1i}, d_{2i}, \dots, d_{Pi}]^T$.

Вектор весов \bar{w}_i может быть определен за один шаг псевдоинверсией матрицы Φ , т.е. $\bar{w}_i = \Phi^+ \bar{d}_i$, где Φ^+ – псевдообратная к Φ матрица. Практически псевдоинверсия рассчитывается с использованием разложения по собственным значениям (SVD – декомпозиция), в соответствии с которым $\Phi = USV^T$.

Матрицы U и V ортогональны и имеют размерности $(P \times P)$ и $((H_1 + 1) \times (H_1 + 1))$, соответственно, тогда как S – это псевдодиагональная

матрица с размерностью $(P \times (H_1 + 1))$. Пусть исходная матрица Φ имеет ранг r . Тогда количество столбцов ортогональных матриц U и V может быть уменьшено до r . Соответственно, получаем зависимость: $\Phi \cong U_r S_r V_r^T$, где $S_r = \text{diag}[s_1, s_2, \dots, s_r]$, $s_1 \geq s_2 \geq \dots \geq s_r > 0$.

Вектор весов РБС, подвергающейся обучению, определяется формулой –

$$\vec{w}_i = \Phi^+ \vec{d}_i,$$

где $\Phi^+ = V_r S_r^{-1} U_r^T$, $S_r^{-1} = \text{diag}[\frac{1}{s_1}, \frac{1}{s_2}, \dots, \frac{1}{s_r}]$.

Выходные веса РБС подбираются за один шаг простым перемножением соответствующих матриц (если $r > (H_1 + 1)$, то принималось – $r = H_1 + 1$). Для вычисления SVD-декомпозиции использовался алгоритм, предложенный в работе Голуба и Кахана [21].

Далее (второй этап), при зафиксированных значениях выходных весов возбуждающие сигналы пропускаются по сети до выходного слоя, что позволяет рассчитать величину погрешности для последовательности векторов $\{\vec{Y}_p^{(0)}\}$. После этого происходит возврат к скрытому слою (обратное распространение). По величине погрешности $\|\vec{Y}^{(2)} - \vec{d}\|_{L_2}$ определяется вектор градиента функции выбора относительно конкретных центров \vec{c}_j и ширины $\vec{\sigma}_j$.

При существовании P обучающих пар функцию выбора можно представить в виде (3.9).

Начальные значения матрицы координат центров $C = \{c_{jh}\}$ активационных функций для нейронов скрытого слоя можно задать в первом приближении случайным образом. Выберем $c_{h,\min} = \min_p(Y_{ph}^{(0)})$, $c_{h,\max} = \max_p(Y_{ph}^{(0)})$, а далее зададим c_{jh} случайным образом на основе равномерного распределения $c_{jh} \in [c_{h,\min}, c_{h,\max}]$, где $j = 1 \dots H_1, h = 1 \dots H_0$ (если предварительно все исходные данные приведены к безразмерному виду, то $c_{jh} \in [-1, 1]$). Для уточнения значений $C = \{c_{jh}\}$ применим алгоритм K -средних ($K = H_1$) [11], что позволяет сократить длительность обучения радиальной нейронной сети.

Когда будут известны значения центров радиальных базисных функций, необходимо определить значения элементов ковариационной матрицы $\sigma = \{\sigma_{jh}\}$.

В качестве начального приближения выбиралось: $\sigma_{jh} = \frac{\rho}{\sqrt{2H_1}}$, где ρ – максимальное расстояние между центрами \vec{c}_j , $j = 1 \dots H_1, h = 1 \dots H_0$;

$$\rho = \max_j \max_k \left(\sqrt{\sum_{h=1}^{H_0} (c_{jh} - c_{kh})^2} \right), \quad k = 1 \dots H_1 \quad (\text{если предварительно все исходные}$$

данные приведены к безразмерному виду, то $\rho = 2\sqrt{H_0}, \sigma_{jh} = \sqrt{\frac{2H_0}{H_1}}$).

Для уточнения значений $\sigma = \{\sigma_{jh}\}$ применялся также алгоритм формирования «области охвата» радиально-базисными функциями с учетом K «соседей»: $\sigma_{jh}^2 = \Sigma_j = \frac{1}{K} \sum_{k=1}^K \sum_{h=1}^{H_0} (c_{jh} - c_{kh})^2$, $k = 1 \dots K$, что позволяло сократить длительность обучения РБС. Принималось $K \in [3, 5]$.

Уточнение элементов ковариационной матрицы $\sigma = \{\sigma_{jh}\}$ и координат центров $C = \{c_{jh}\}$ завершает очередной цикл обучения. Коррекция элементов ковариационной матрицы осуществляется по следующей формуле (представлен рекуррентный алгоритм обучения, соответствующий методу стохастической аппроксимации, обеспечивающий сходимость $\sigma_{jh}(t) \xrightarrow{t \rightarrow \infty} \hat{\sigma}_{jh}$ с вероятностью $P = 1$):

$$\sigma_{jh}(t+1) = \sigma_{jh}(t) + \mu(t) \{ \eta_{jh}(t) r_{jh}(t) + \alpha_{jh}(t) [\sigma_{jh}(t) - \sigma_{jh}(t-1)] \} + \tilde{\sigma}_{jh}(t+1), \quad (4.3)$$

где проекции вектора направления поиска $r_{jh}(t)$, определялись в соответствии с методом сопряженных градиентов (Полака – Рибьера), а проекции градиента функции выбора находились по формулам:

$$S_{jh}(t) = - \sum_{p=1}^P \gamma^{P-p} \sum_{i=1}^{H_2} \left\{ f'_{fit, \Delta_i} \Delta_{pi} w_{ij} \exp \left[- \frac{1}{2} \sum_{h=1}^{H_0} \frac{(Y_{ph}^{(0)} - c_{jh})^2}{\sigma_{jh}^2} \right] \frac{(Y_{ph}^{(0)} - c_{jh})^2}{\sigma_{jh}^3} \right\} -$$

$$- \frac{\beta \sigma_{jh}}{\ln 2(H_1 + 1)} \cdot \frac{(1 + \sum_{n=0}^{H_0} \sigma_{jn}^2)}{\left(1 + \sum_{n=0}^{H_0} \sigma_{jn}^2\right)^2} \left\{ 1 + \ln \left[\frac{\rho_j(M_{j,t})}{\rho_j(M_{j,t-1})} \right] \right\}. \quad (4.4)$$

Аналогичным образом осуществлялось уточнение элементов матрицы координат центров $C = \{c_{jh}\}$. Проекция градиента функции выбора в этом случае находились по формулам:

$$S_{jh}(t) = - \sum_{p=1}^P \gamma^{P-p} \sum_{i=1}^{H_2} \left\{ f'_{fit, \Delta_i} \Delta_{pi} w_{ij} \exp \left[- \frac{1}{2} \sum_{h=1}^{H_0} \frac{(Y_{ph}^{(0)} - c_{jh})^2}{\sigma_{jh}^2} \right] \frac{(Y_{ph}^{(0)} - c_{jh})}{\sigma_{jh}^2} \right\} -$$

$$-\frac{\beta c_{jh}}{\ln 2(H_1 + 1)} \cdot \frac{(1 + \sum_{n=0}^{H_0} c_{jn}^2)}{(1 + \sum_{n=0}^{H_0} c_{jn}^2)^2} \left\{ 1 + \ln \left[\frac{\rho_j(M_{j,t})}{\rho_j(M_{j,t-1})} \right] \right\}. \quad (4.5)$$

В завершении осуществлялось восстановление аналитического представления искомых функций $Y_{pi}^{(2)} = w_{io} + \sum_{j=1}^{H_1} w_{ij} \varphi_j(\bar{Y}_p^{(0)}, \bar{c}_j, \bar{\sigma}_j)$, $i = 1 \dots H_2$, вначале в безразмерной форме, а затем с использованием обратного преобразования – в размерной форме.

Многочисленное повторение обоих этапов ведёт к полному и быстрому обучению сети, особенно когда начальные значения параметров радиально-базисных функций близки к оптимальным значениям.

На практике выделенные этапы в разной степени влияют на адаптацию параметров. Как правило, быстрее функционирует алгоритм SVD (он за один шаг находит локальный минимум функции). Для выравнивания этой диспропорции одно уточнение линейных параметров сопровождается обычно несколькими циклами адаптации нелинейных параметров.

5. Сравнительный анализ ОМС и РБС применительно к задачам аппроксимации данных

В качестве примера была взята выборка из $P = 398$ значений соответствующих параметров аэродинамических характеристик осевого многоступенчатого компрессора (МОК) авиационного двигателя. В качестве входных данных для обучаемых искусственных нейронных сетей (ИНС) задавались значения расхода воздуха (G_A) и количество оборотов в минуту ротора (n) МОК, в качестве выходных данных – степень сжатия (π_k^*) и коэффициент полезного действия (η_k^*). Для решения задачи аппроксимации данных использовались ОМС и РБС с разнообразной структурой. Эти же сети были построены и обучались с помощью математического пакета MATLAB 7.0.1.

Введем следующие обозначения для выделения метода обучения ОМС и РБС: СГ1 – обучение методом сопряженных градиентов ОМС с одним скрытым слоем, СГ2 – обучение методом сопряженных градиентов ОМС с двумя скрытыми слоями, ЛМ1 – обучение методом Левенберга-Марквардта ОМС с одним скрытым слоем, ЛМ1Р – обучение ОМС с одним скрытым слоем методом ЛМ1 на основе байесовской регуляризации, ЛМ2 – обучение методом Левенберга-Марквардта ОМС с двумя скрытыми слоями, СГ1 (ХАИ) – обучение предложенным методом ОМС с одним скрытым слоем, РБС (ХАИ) – обучение предложенным методом РБС.

На рис. 2 представлены результаты аппроксимации характеристик МОК, полученные с помощью ОМС методом СГ1 (ХАИ) – зависимости $[G_A, \pi_E^*]$

(рис.2, а) и $[G_A, \eta_K^*]$ (рис. 2, б) при различных значениях оборотов в минуту ротора (n) МОК.

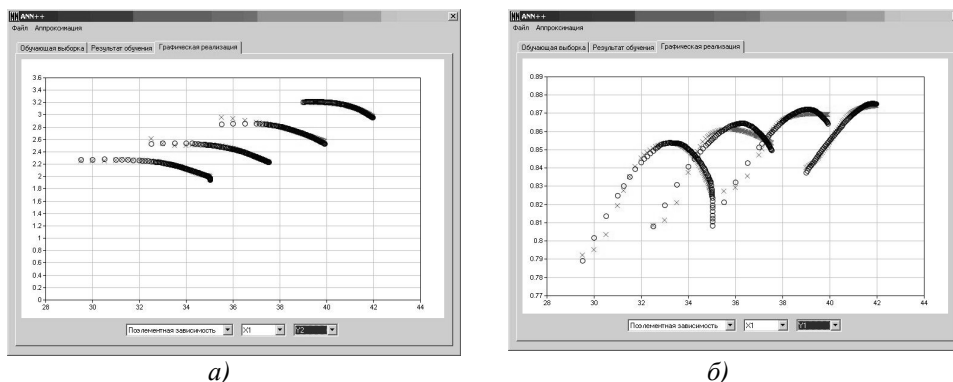


Рис. 2. Результаты аппроксимации характеристик МОК с помощью ОМС:
 ○ – исходные данные, × – значения аппроксимирующих функций

Для проверки значимости (качества) предсказания данных ОМС и РБС вычислялись средние относительные погрешности по физическим параметрам δ_i^0 и энергия среднеквадратичной ошибки E_{av} . Результаты обучения ИНС для рассматриваемой выборки представлены в таблице.

Общий анализ данных показывает, что ОМС с двумя скрытыми слоями обучается лучше, чем с одним скрытым слоем. ЛМ1 и ЛМ2 имеют эффективную реализацию в системе MATLAB 7.0.1, однако эффективность этих методов ухудшается в случае подключения байесовской регуляризации (см. ЛМ1Р в сравнении с ЛМ1). Таким образом, разработанные методы СГ1 (ХАИ) и РБС(ХАИ) обучения ИНС на основе предложенного метода стохастической аппроксимации с регуляризацией обеспечивают при априорной неопределенности входных данных стабильность (робастность) и информативность параметров статистических формальных математических моделей систем (процессов), а также достаточную с практической точки зрения точность аппроксимации данных.

Таблица. Результаты обучения ИНС ($P = 398, H_0 = 2, T=300$)

Метод	δ_π^0	δ_η^0	E_{av}	Параметры
СГ1	0.0762	0.0121	0.0536	2-20-2
СГ2	0.0263	6.7e-3	0.0816	2-20-10-2
ЛМ1	4.1e-4	2.3e-4	6.3e-5	2-20-2
ЛМ1Р	5.0e-4	3.5e-4	0.0122	2-20-2
ЛМ2	1.6e-4	6.1e-5	2.7e-6	2-20-10-2
СГ1(ХАИ)	2.7e-3	1.7e-3	5.1e-5	2-20-2
РБС	3.9e-4	4.0e-4	2.5e-4	2-398-2
РБС(ХАИ)	7.5e-4	7.4e-4	6.5e-6	2-20-2

6. Выводы по результатам и направления дальнейших исследований

Предложены методология и компьютерная система поддержки принятия решений (КСППР) для формирования многомерных формальных математических моделей (ФММ) на основе статистических моделей систем. Разработан метод оценивания параметров статистических моделей систем на базе обучаемых искусственных нейронных сетей (ИНС) на основе метода стохастической аппроксимации путем применения регуляризирующего последовательного (адаптивного) алгоритма синтеза решений с отложенной коррекцией, в котором, в отличие от существующих, реализовано адаптивное управление вычислениями в соответствии с принципом минимального возмущения, а в качестве функций выбора рационального решения используются модифицированные вероятностные критерии. Разработанные методы СГ1 (ХАИ) и РБС (ХАИ) обучения ИНС на основе предложенного метода стохастической аппроксимации с регуляризацией обеспечивают при априорной неопределенности входных данных стабильность (робастность) и информативность параметров статистических математических моделей систем (процессов), а также достаточную с практической точки зрения точность аппроксимации данных.

ЛИТЕРАТУРА

1. Горбань А.Н. Обучение нейронных сетей. – М.: СП Параграф, 1990. – 125 с.
2. Уоссермен Ф. Нейрокомпьютерная техника: теории и практика. – М.: Мир, 1992. – 127 с.
3. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности / Г.К. Вороновский, К.В.Махотило, С.Н. Петрашев, С.А. Сергеев. – Харьков: ОСНОВА, 1997. – 112 с.
4. Горбань А.Н., Дунин-Барковский В.Л., Кирдин А.Н. Нейроинформатика. – Новосибирск: Наука, СО РАН, 1998. – 296 с.
5. Заенцев И.В. Нейронные сети: основные модели: Учебное пособие по курсу «Нейронные сети». – Воронеж: Воронежский государственный университет, 1999. – 76 с.
6. Нейроні мережі в системах автоматизації / В.І. Архангельський, І.М. Богаєнко, Г.Г. Грабовський, М.О. Рюмшин. – К.: Техніка, 1999. – 364 с.
7. Галушкин А.И. Теория нейронных сетей. Кн. 1: Учеб. пособие для вузов. – М.: ИПРЖР, 2000. – 416 с.
8. Каллан Р. Основные концепции нейронных сетей. – М. – Санкт-Петербург – Киев: Издательский дом «Вильямс», 2001. – 287 с.
9. Дюк В., Самойленко А. Data mining: учебный курс. – СПб: Питер, 2001. – 368 с.
10. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. – М.: Горячая линия – Телеком, 2001. – 382 с.
11. Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2002. – 344 с.
12. Медведев В.С., Потемкин В.Г. Нейронные сети. MATLAB 6 / Под общ. ред. В.Г. Потемкина» – М.: ДИАЛОГ – МИФИ, 2002. – 496 с. (Пакеты прикладных программ: Кн. 4)

13. Назаров А.В., Лоскутов А. И. Нейросетевые алгоритмы прогнозирования и оптимизации систем. – СПб: Наука и техника, 2003. – 384 с.
14. Рутковская Д, Пилинский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. – М.: Горячая линия-Телеком, 2004. – 452 с.
15. Бодянский Е.В., Рученко О.Г. Искусственные нейронные сети: архитектуры, обучение, применения. – Харьков: ТЕЛТЕХ, 2004. – 372 с.
16. Руденко О.Г., Бодянский Е.В. Искусственные нейронные сети. Учебное пособие. – Харьков: ООО «Компания СМИТ», 2005. – 408 с.
17. Хайкин С. Нейронные сети: полный курс. – М: Издательский дом «Вильямс», 2006. –1104 с.
18. Системы искусственного интеллекта. Практический курс: учебное пособие / В.А. Чулюков, И.Ф. Астахова, А.С. Потапов и др.; под ред. И.Ф.Астаховой. – М.: БИНОМ. Лаборатория знаний, 2008. – 292 с.
19. Комп'ютерна програма „Комп'ютерна інтерактивна система підтримки прийняття рішень при формуванні вигляду елементів складних технічних систем «Concept_Pro»”: Свідчення про реєстрацію авторського права на твір №28975 / Е.М. Угрюмова, О.А. Трончук В.С. Афанасьєвська (Україна). – Дата реєстрації 29.05.2009.
20. Самарский А.А., Гулин А.В. Численные методы. – М.: Наука. Гл. ред. физ-мат. лит., 1989. – 432 с.
21. Голуб Дж., Ван Лоун Ч. Матричные вычисления. – М.: Мир, 1999. – 534 с.
22. Соломахова Т.С., Беляновский Е.С. Применение аэродинамических характеристик круговых решеток тонких профилей при проектировании радиальных вентиляторов // В кн.: Промышленная аэродинамика. – М.: Машиностроение, 1986. – Вып. 1/33. – С. 63 – 70.