УДК 681.518

МОРОЗОВА Г.В., к.т.н. (УкрГАЖТ)

Построение формальных математических моделей систем с применением обучаемых искусственных нейронных сетей

Рассмотрено построение формальных математических моделей систем с применением обучаемых однонаправленных многослойных искусственных нейронных сетей. Проведены расчеты и получены результаты обучения искусственных нейронных сетей. На основе анализа результатов численных исследований устойчивости различных методов обучения однонаправленных многослойных искусственных нейронных сетей были выявлены их особенности.

Ключевые слова: *формальные математические модели, обучаемые однонаправленные многослойные искусственные нейронные сети.*

Актуальность

Вследствие «кажущейся» простоты, широкое применение в компьютерных системах поддержки принятия решений (СППР) в автоматизированных системах, предназначенных для решения широкого круга задач анализа, проектирования, диагностики технического состояния, синтеза и управления системами в разных областях: машиностроении, транспорте, энергетике, экологическом и экономическом мониторинге, медицине и других, получили формальные математические модели (ФММ) на основе статистических моделей.

В условиях априорной неопределенности входных данных влияние достоверности информации о состоянии систем, которую можно получить с помощью ФММ, на процесс принятия решений существенно возрастает.

Анализ последних достижений и публикаций

К настоящему времени опубликовано множество работ, посвященных описанию теории, методов обучения, практике применения обучаемых искусственных нейронных сетей (ИНС) в различных областях науки и техники. Среди приведенных публикаций следует выделить работы [1 - 6], 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 основе обучаемых ИНС, отсутствует анализ устойчивости решений к возмущениям входных данных и погрешностям вычислений.

В данной работе рассмотрено применение метода оценивания параметров статистических моделей в форме обучаемых однонаправленных многослойных искусственных нейронных сетей (ОМС) на основе метода стохастической аппроксимации, в котором реализовано адаптивное управление вычислениями в соответствии с принципом минимального возмущения, а модифицированные вероятностные критерии используются в качестве функций выбора рационального решения, что обеспечивает при наличии неопределенности входных данных стабильность и информативность параметров статистических моделей, а также достаточную с практической точки зрения точность аппроксимации данных.

Общая постановка задачи

Задана векторная функция набором обучающих пар $\left(\vec{\mathrm{Y}}^{(0)},\vec{\mathrm{d}}\right)_{\mathrm{p}}$, $p=1,...,P$; где $\vec{\mathrm{Y}}^{(0)},\vec{\mathrm{d}}$ – 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 $Y_p^{(K+1)}(\vec{Y}^{(0)})$, представленной данной обучающей выборкой, по заданному вектору входа, в диапазоне, ограниченном входными данными.

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 – синоптических 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.

Входными данными для аппроксимации данных с помощью ОМС являются: входные параметры и управляющие переменные опытных образцов (аналогов) $\left\{\mathrm{Y}_{\mathrm{ph}}^{(0)}\right\}$; выходные параметры $\left\{d_{pi}\right\}$.

Простейшая ОМС с одним скрытым слоем (K=1) изображена на рис. 1.

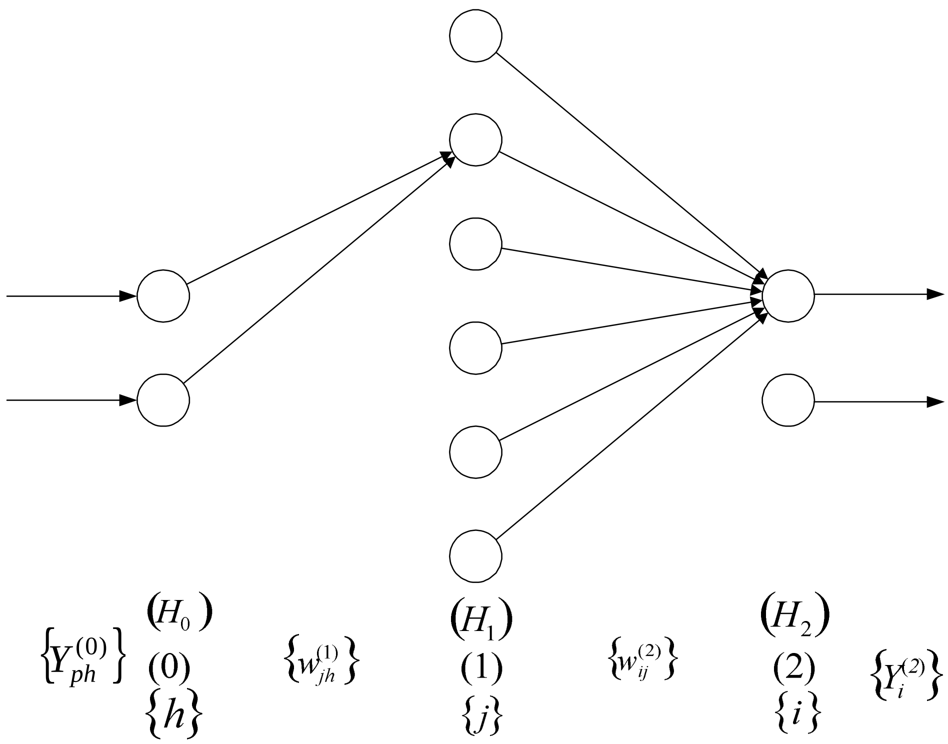

Рис. 1. Структура ОМС:

H_0 – количество входов сети; H_1 – количество нейронов скрытого слоя; H_2 – количество выходов сети; $\left\{\mathrm{Y}_{\mathrm{ph}}^{(0)}\right\}$ – множество входных данных; $\left\{\mathrm{Y}_{\mathrm{i}}^{(\mathrm{k})}\right\}$ – множество выходных данных k–ого слоя; k – номер слоя, k=1, …, (K+1); K – число скрытых слоев; p=1, …, P; P – число аналогов; $\left\{w_{ij}^{(k)}\right\}$ – множество весов k – ого слоя; i – элемент k – ого слоя; j – элемент (k – 1) – ого слоя.

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

$\mathrm{m} \le \mathrm{PH}_{\mathrm{K+1}}$.

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

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

Будем считать, что распределение плотности вероятностей безошибочного принятия гипотезы о достоверности найденных значений параметров модели M_{t+1}, когда гипотеза неверна, определяется законом $\rho(\mathrm{M}_{t+1}) \sim exp[-\beta_{t+1} D_{KL}(M_{t+1}, M_t)]$, где D_{KL} – относительная энтропия Кульбака-Лейблера.

В дальнейшем в качестве скалярной свертки функций выбора использовалась функция вида [7]

$$E = \frac{1}{2PH_{K+1}} \sum_{p=1}^{P} \gamma^{P-p} \sum_{i=1}^{H_{K+1}} f_{fit}[\Delta_{pi}^2(M)] + \frac{1}{2}\beta_{t+1} \nu D_{KL}(M_{t+1}, M_t) \ (\gamma = 0.95\ldots0.99; \beta_{t+1} = 10^{-7}). \quad (2)$$

Решение – аппроксимирующие функции вида $\mathrm{Y}_i^{(K+1)}(\vec{\mathrm{Y}}^{(0)})$ – будем искать методом стохастической аппроксимации. Коррекцию весов связей будем осуществлять по следующей формуле (представлен рекуррентный алгоритм обучения, соответствующий методу стохастической аппроксимации, обеспечивающий сходимость $w_{ij}^{(k)}(t) \xrightarrow[t\to\infty]{} \hat{w}_{ij}^{(k)}$ с вероятностью $\mathrm{P} = 1$):

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

где $\mu(t) = \dfrac{\mu(0)}{1+t}$ – коэффициент обучения; $t = 1\ldots\mathrm{P}\cdot\mathrm{T}$, Т – количество эпох обучения;

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

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

$$S_{ij}^{(k)}(t) = -\frac{\partial E}{\partial w_{ij}^{(k)}} = -\frac{\gamma^{P-p}}{PH_{K+1}} f'_{fit,\Delta_i^2} \delta_i^{(k)} Y_j^{(k-1)} -$$

$$- \frac{\beta_{t+1} \nu w_{ij}^{(k)}}{2\ln 2P \sum\limits_{k=1}^{K+1}(H_k+1)} \cdot \frac{(1+\sum\limits_{\substack{n=0\\(n\neq j)}}^{H_{k-1}} w_{in}^2)}{(1+\sum\limits_{n=0}^{H_{k-1}} w_{in}^2)^2} \left\{1 + \ln[\frac{\rho_i^{(k)}(M_{i,t}^{(k)})}{\rho_i^{(k)}(M_{i,t-1}^{(k)})}]\right\}$$

– проекции градиента;

$$\tilde{w}_{ij}^{(k)}(t+1) = \upsilon(t) w_{ij}^{(k)}(t) \Delta w^{\circ} (2\xi_{ij} - 1)$$

– «аддитивный шум» (здесь $\upsilon(t) = \mu(t)\sqrt{\dfrac{2}{ln(2+t)}}$; принималось $\Delta w^{\circ} = 0.2$; $\xi_{ij} \in [0,1]$ – нормально распределенные случайные величины, имеющие нулевое значение среднего и единичную дисперсию).

В рассматриваемом случае использовалась функция принадлежности вида:

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

При обучении использовался регуляризирующий алгоритм, реализующий прерывания в итерационном процессе в случаях накопления ошибок вычислений (верхний индекс Т в представленных ниже формулах означает операцию транспонирования вектора в строку) if $\dfrac{(S_i^{(k)}(t))^T(S_i^{(k)}(t)-S_i^{(k)}(t-1))}{(S_i^{(k)}(t-1))^T S_i^{(k)}(t-1)} \geq r_{\max}$ then $r_{ij}^{(k)}(t-1)=0$;

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

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

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

$$\varphi_i^{(k)}=b\left(1-f^2\left(s_i^{(k)}\right)\right);Y_i^{(k)}=f(s_i^{(k)});s_i^{(k)}=w_{io}^{(k)}+\sum_{j=1}^{H_{k-1}}w_{ij}^{(k)}Y_j^{(k-1)};$$

$f(s)=\dfrac{e^{bs}-e^{-bs}}{e^{bs}+e^{-bs}}$ – передаточная функция;

$\varphi(s)$ – производная передаточной функции;

$k = 1,\ldots,(K+1);\ i = 1,\ldots,H_k;\ j = 1,\ldots,H_{k-1};$

H_k – число элементов в k–ом слое;

$\eta_{ij}^{(k)}$, $\alpha_{ij}^{(k)}$ – коэффициенты обучения и момента, соответственно.

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: ± 0.5.

Применение ОМС для решения задач аппроксимации данных.

Для решения задачи аппроксимации данных использовались ОМС 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: 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(new)– обучение предложенным методом ОМС с одним скрытым слоем, МНК – обучение методом наименьших квадратов мультипликативной модели.

Для проверки значимости (качества) предсказания данных ОМС и РБС вычислялись средние относительные погрешности по физическим параметрам δ_i° и энергия среднеквадратичной ошибки E_{av}.

В качестве примера была взята выборка из Р=16 значений соответствующих параметров аэродинамических характеристик радиальных вентиляторов с загнутыми назад лопатками рабочего колеса [8].

В качестве входных данных для ИНС задавались значения: густоты решетки (τ), относительной вогнутости профиля ($\bar{f}$), безразмерной хорды профиля ($\bar{l}$), безразмерного диаметра входа решетки ($\bar{D}_1$).

В качестве выходных данных задавались: безразмерные параметры расхода ($y_1 \equiv \bar{c}_{2r}^* = \dfrac{\varphi^*}{4\bar{b}_2}$) и полного давления ($y_2 = 0.5\psi^* = \bar{c}_{2u}^*\cdot\eta_g$) для режима максимального КПД, где $\bar{c}_{2r}^*$ – коэффициент расхода, $\bar{c}_{2u}^*$ – относительная окружная составляющая абсолютной скорости на выходе, φ^* – коэффициент производительности, ψ^* – коэффициент полного давления, $\bar{b}_2$ – относительная ширина колеса на выходе, η_g – гидравлический КПД рабочего колеса.

Результаты обучения ИНС для второй выборки представлены в таблице 1.

Таблица 1

Результаты обучения ИНС (Р = 16, H_0 = 4, H_2 = 2, Т=300)

Метод	δ_{ϕ}^{0}	δ_{ψ}^{0}	E_{av}	Параметры
МНК	0.3509	0.0830	0.0205	–
СГ1	0.0020	0.0015	3.8e-5	4-10-2
СГ2	0.0023	0.0016	8.1e-5	4-10-5-2
ЛМ1	4.8e-7	2.8e-7	3.4e-12	4-10-2
ЛМ1Р	0.0241	0.0125	0.0743	4-10-2
ЛМ2	1.4e-6	2.6e-6	1.9e-10	4-10-5-2
СГ1(new)	0.0030	0.0053	5.4e-7	4-10-2

Заключение

Рассмотрено применение обучаемых однонаправленных многослойных искусственных нейронных сетей (ОМС) для построения формальных математических моделей систем.

На основе анализа результатов численных исследований устойчивости различных методов обучения ОМС были выявлены следующие особенности:

– процесс обучения ОМС быстро сходится;

– 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).

Литература

1. Осовский С. Нейронные сети для обработки информации / С.Осовский. – М.: Финансы и статистика, 2002. – 344 с.
2. Медведев В.С. Нейронные сети. MATLAB 6 / В.С. Медведев, В.Г. Потемкин; под общ. ред. В.Г. Потемкина. – М.: ДИАЛОГ–МИФИ, 2002. – 496 с. (Пакеты прикладных программ: Кн. 4)
3. Бодянский Е.В. Искусственные нейронные сети: архитектуры, обучение, применения / Е.В. Бодянский, О.Г. Руденко. – Харьков: ТЕЛЕТЕХ, 2004. – 372 с.
4. Зайченко Ю.П. Основи проектування інтелектуальних систем / Ю.П. Зайченко. – К. : Видавничий Дім «Слово», 2004. – 352 с.
5. Руденко О.Г. Искусственные нейронные сети. Учебное пособие / О.Г.Руденко, Е.В. Бодянский. – Харьков: ООО «Компания СМИТ», 2005. – 408с.
6. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – М.: Издательский дом «Вильямс», 2006. – 1104 с.
7. Угрюмова Е.М. Обучаемые искусственные нейронные сети в построении формальных математических моделей систем при априорной неопределенности данных / Е.М. Угрюмова // Вісник Харківського національного університету : зб. наук. пр. Сер. Математичне моделювання. Інформаційні технології. Автоматизовані системи управління. – 2010. – Випуск 13 (№890). – С. 237 – 253.
8. Соломахова Т.С. Применение аэродинамических характеристик круговых решеток тонких профилей при проектировании радиальных вентиляторов / Т.С. Соломахова, Е.С. Беляновский // В кн.: Промышленная аэродинамика. – М. : Машиностроение, 1986. – Вып. 1/33. – С. 63 – 70.

Морозова Г.В. ПОБУДОВА ФОРМАЛЬНИХ МАТЕМАТИЧНИХ МОДЕЛЕЙ СИСТЕМ З ВИКОРИСТАННЯМ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ, ЯКІ НАВЧАЮТЬ. Розглянуто побудову формальних математичних моделей систем із застосуванням односпрямованих багатошарових штучних нейронних мереж, які навчають. Проведено розрахунки і отримані результати навчання штучних нейронних мереж. На основі аналізу результатів чисельних досліджень стійкості різних методів навчання односпрямованих багатошарових штучних нейронних мереж були виявлені їх особливості.

Ключові слова: формальні математичні моделі, односпрямовані багатошарові штучні нейронні мережі, які навчають.

G. Morozova. THE CONSTRUCTION OF FORMAL MATHEMATICAL SYSTEM MODELS WITH THE APPLICATION OF TRAINABLE ARTIFICIAL NEURAL NETWORKS. The construction of formal mathematical system models with the application of trainable unidirectional multilayer artificial neural networks has been considered. The calculations have been conducted and the results of artificial neural network training have been obtained. The peculiarities of unidirectional multilayer artificial neural networks have been identified on the bases of the results of a large body of research on the stability of different training methods.

Key words: formal mathematical models, trainable unidirectional multilayer artificial neural networks.

Рецензент д.т.н., доцент Вамболь С.А. (Национальный университет гражданской защиты Украины)

Поступила 06.06.2014 г.