Е.В. Бодянский, Е.В. Горшков, В.В. Колодяжный

ОБУЧЕНИЕ НЕЙРО-ФАЗЗИ СЕТИ КОЛМОГОРОВА НА ОСНОВЕ ОБРАТНОГО РАСПРОСТРАНЕНИЯ ОШИБОК И АЛГОРИТМОВ ОПТИМИЗАЦИИ ВТОРОГО ПОРЯДКА

Введение

Установление универсальных аппроксимирующих свойств искусственных нейронных сетей (ИНС) в значительной мере базировалось на фундаментальной теореме А. Н. Колмогорова о суперпозиции функций (ТСК) [1], утверждающей, что для всех $d \geq 2$ существуют непрерывные функции $\varphi_{li} : [0,1] \to \mathbb{R}$, $l = 1, 2, \ldots, 2d+1$; $i = 1, 2, \ldots, d$, такие, что выполняется следующее условие: для всех непрерывных функций $f : [0,1]^d \to \mathbb{R}$ существуют непрерывные функции $g_l : [0,1] \to \mathbb{R}$, такие, что

$$y = f(x_1, x_2, \ldots, x_d) = \sum_{l=1}^{2d+1} g_l \left(\sum_{i=1}^{d} \varphi_{li}(x_i) \right), \tag{1}$$

при этом функции $g_l(\cdot)$ определяются видом $f(\cdot)$, а функции $\varphi_{li}(\cdot)$ от $f(\cdot)$ не зависят.

Р. Хехт-Нильсеном [2] было показано, что ТСК может быть реализована с помощью нейросетевых технологий и была разработана ИНС с прямой передачей информации с единственным скрытым слоем, который конструировался независимо от вида функции $f(\cdot)$.

В. Куркова [3] предложила в конструкции (1) фиксированное число внешних активационных функций $g_l(\cdot)$ равное $2d+1$ заменить переменным n, а саму проблему представления функций $f(\cdot)$ заменить задачей нахождения ее аппроксимации $\hat{f}(\cdot)$, что гораздо удобнее при решении большинства реальных задач обработки информации. Ею же была введена соответствующая ИНС прямого распространения с сигмоидальными активационными функциями, обеспечивающая любую наперед заданную точность аппроксимации.

Д. А. Шпрехер [4] разработал численный алгоритм вычисления внутренних $\varphi_{li}(\cdot)$ и внешних $g_l(\cdot)$ одномерных активационных функций в суперпозиции Колмогорова специального вида, при этом основным предположением, используемым при вычислении этих функций, является существование $2d+1$ семейств попарно непересекающихся d-мерных кубов, объединение которых покрывает гиперкуб I^d наперед заданным образом.

Б. Игельник и Н. Парикх [5] предложили конструкцию, названную ими сплайн-сетью Колмогорова (Kolmogorov's spline network) и основанную на технике аппроксимации с помощью кубических сплайнов. Данная сеть обладает расширенными функциональными возможностями и лучшей адаптируемостью по сравнению с классическими ИНС.

А. Лопес-Гомес, Ш. Йошида, К. Хирота [6] на основе фаззификации ТСК [7] разработали нечеткую функционально связанную нейронную сеть (fuzzy functional link network), в которой внутренние функции $\varphi_{li}(\cdot)$ были заменены функциями принадлежности, используемыми в системах нечеткого вывода (fuzzy inference systems — FIS).

Вместе с тем следует отметить, что сети, основанные на ТСК, несмотря на свою эффективность и математическую элегантность не получили должного распространения для решения реальных задач, что объясняется, прежде всего, отсутствием численно простых и быстродействующих алгоритмов обучения.

В [8] описана гибридная структура, названная нейро-фаззи сетью Колмогорова (Neuro-Fuzzy Kolmogorov's Network — NFKN), сочетающая в себе колмогоровскую схему суперпозиции функций одной переменной, нейронную архитектуру прямого распространения и систему нечеткого вывода типа Такаги-Сугено, и обучаемая градиентной численно простой процедурой обратного распространения ошибок, однако обладающая, в то же время, всеми недостатками, присущими системам, параметры которых настраиваются с помощью алгоритмов оптимизации первого порядка. В данной работе предложен новый алгоритм обучения нейро-фаззи сети Колмогорова на основе обратного распространения ошибок и алгоритмов оптимизации второго порядка.

Архитектура NFKN

Введем в рассмотрение двухслойную архитектуру, приведенную на рис. 1 и реализующую отображение вида

$$\hat{f}(x_1, x_2, \ldots, x_d) = \sum_{l=1}^{n} f_l^{[2]}(o_l^{[1]}) = \sum_{l=1}^{n} f_l^{[2]}\left(\sum_{i=1}^{d} f_i^{[1,l]}(x_i)\right), \tag{2}$$

где $n \geq 2d+1$ — число сумматоров скрытого слоя, d — число входов нулевого слоя, $f_l^{[2]}(\cdot)$ ($l = 1, 2, \ldots, n$), $f_i^{[1,l]}(\cdot)$ ($i = 1, 2, \ldots, d$) — нелинейные преобразователи выходного и скрытого слоев, соответствующие функциям $g_l(\cdot)$ и $\varphi_{li}(\cdot)$ в выражении (1).

В качестве «строительных элементов» данной сети предлагается использовать нелинейные синапсы (nonlinear synapses — NS) и нео-фаззи нейроны (neo-fuzzy neurons — NFN), введенные и использованные в [9,10] и отличающиеся эффективными аппроксимирующими свойствами, простотой обучения и реализации.

На рис. 2 приведена схема нелинейного синапса $\mathsf{NS}_i^{[1,l]}$ скрытого слоя, осуществляющая преобразование

$$f_i^{[1,l]}(x_i) = \sum_{h=1}^{m_1} \mu_{ih}^{[1]}(x_i) w_{ih}^{[1,l]}, \tag{3}$$

где m_1 — число функций принадлежности $\mu_{ih}^{[1]}(\cdot)$ в нелинейном синапсе, $w_{ih}^{[1,l]}$ — настраиваемые синаптические веса, общее число которых в скрытом слое составляет $d \cdot m_1 \cdot n$.

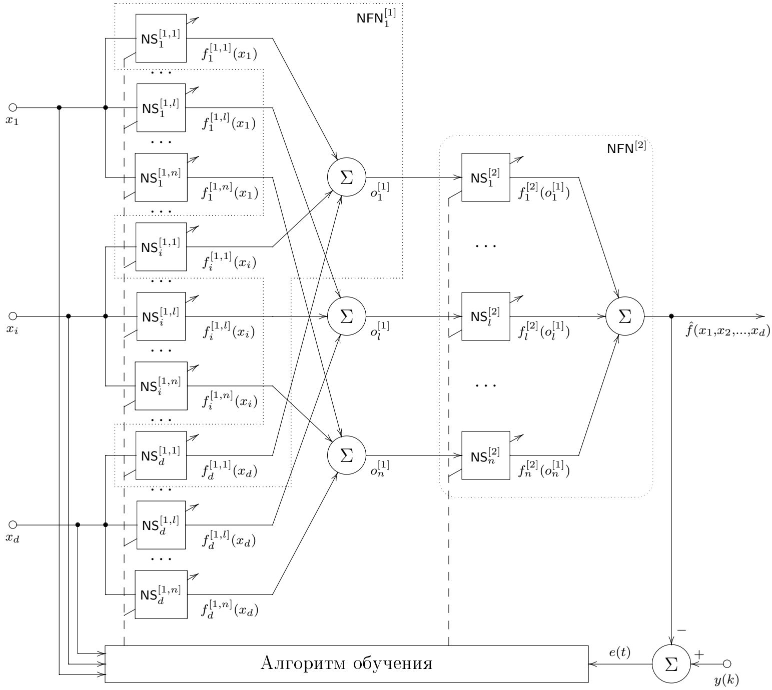

Рис. 1 – Архитектура нейро-фаззи сети Колмогорова

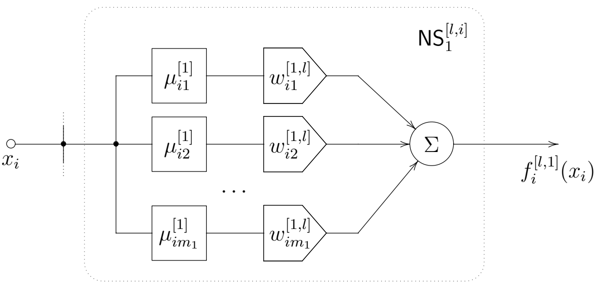

Рис. 2 – Нелинейный синапс скрытого слоя

На рис. 3 приведена схема нелинейного синапса $\mathsf{NS}_l^{[2]}$ выходного слоя, осуществляющего преобразование

$$f_l^{[2]}(o_l^{[1]}) = \sum_{j=1}^{m_2} \mu_{lj}^{[2]}(o_l^{[1]}) w_{lj}^{[2]}, \tag{4}$$

где m_2 — число функций принадлежности $\mu_{lj}^{[2]}(\cdot)$ в нелинейном синапсе, $o_l^{[1]}$ — выходные сигналы скрытого слоя, $w_{lj}^{[2]}$ — настраиваемые синаптические веса, общее число которых в выходном слое составляет $m_2 \cdot n$.

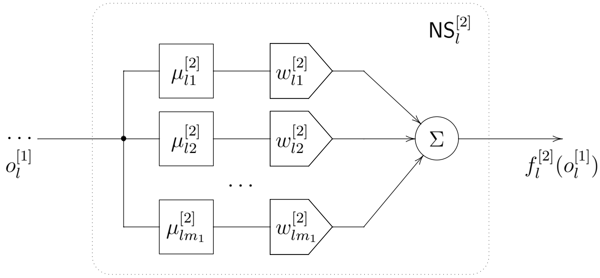

Рис. 3 – Нелинейный синапс выходного слоя

Несложно видеть, что нелинейный синапс NS есть не что иное как система нечеткого вывода типа Такаги-Сугено (TS-FIS) с одним входом и постоянным консеквентом. Как известно [11], TS-FIS является универсальным аппроксиматором и при соответствующем выборе синаптических весов и числе функций принадлежности может с наперед заданной точностью аппроксимировать любую ограниченную функцию одного переменного. Именно это обстоятельство делает целесообразным их применение в составе NFKN в качестве аналогов функций $g_l(\cdot)$ и $\varphi_{li}(\cdot)$.

Как видно из схемы на рис. 1, выходные сигналы скрытого слоя могут быть записаны в форме

$$o_l^{[1]} = \sum_{i=1}^{d} f_i^{[1,l]}(x_i) = \sum_{i=1}^{d} \sum_{h=1}^{m_1} \mu_{ih}^{[1]}(x_i) w_{ih}^{[1,l]}, \tag{5}$$

а выходного —

$$\hat{f}(x_1, x_2, \ldots, x_d) = \sum_{l=1}^{n} \sum_{j=1}^{m_2} \mu_{lj}^{[2]}(o_l^{[1]}) w_{lj}^{[2]}, \tag{6}$$

что, как несложно заметить, соответствует преобразованиям, реализуемым нео-фаззи нейронами, число которых в скрытом слое составляет n, а в выходном 1. На рис. 1 пунктирной рамкой обведены нео-фаззи нейроны $\mathsf{NFN}_1^{[1]}$ и $\mathsf{NFN}^{[2]}$.

Окончательно отображение, реализуемое с помощью NFKN, может

 ISSN 1562-9945

быть записано в виде

$$\hat{f}(x_1, x_2, \ldots, x_d) = \sum_{l=1}^{n} f_l^{[2]} \left(\sum_{i=1}^{d} f_i^{[1,l]}(x_i) \right) = \sum_{l=1}^{n} \sum_{j=1}^{m_2} \mu_{lj}^{[2]} \left(\sum_{i=1}^{d} \sum_{h=1}^{m_1} \mu_{ih}^{[1]} w_{ih}^{[1,l]} \right) w_{lj}^{[2]}, \tag{7}$$

что является гибридом двухслойной ИНС с прямой передачей информации и двухуровневой системы нечетких правил, реализующей двухуровневый подход (multiresolution approach) [12], что позволяет согласно теореме Яма-Нгуена-Крейновича обеспечить сколь угодно высокую точность аппроксимации любой многомерной непрерывной функции, заданной на ограниченном множестве.

Обучение NFKN

Для нахождения $d \cdot m_1 \cdot n$ неизвестных синаптических весов скрытого слоя $w_{ih}^{[1,l]}$ и $m_2 \cdot n$ весов $w_{lj}^{[2]}$ выходного используем стандартный критерий обучения, принятый в теории ИНС, вида

$$E^k = \frac{1}{2} \sum_{k=1}^{N} e^2(k) = \frac{1}{2} \sum_{k=1}^{N} (y(k) - \hat{f}(x_1(k), x_2(k), \ldots, x_d(k)))^2, \tag{8}$$

где $k = 1, 2, \ldots, N$ — номер наблюдения в обучающей выборке, N — ее объем, а $y(k)$ — обучающий сигнал.

Вводя в рассмотрение $(m_2 n \times 1)$-векторы синаптических весов выходного слоя и соответствующих их сигналов с выхода скрытого слоя

$$\begin{aligned}
&w^{[2]} = (w_{11}^{[2]}, w_{12}^{[2]}, \ldots, w_{1m_2}^{[2]}, \ldots, w_{l1}^{[2]}, w_{l2}^{[2]}, \ldots, w_{lm_2}^{[2]}, \ldots, w_{n1}^{[2]}, \ldots, w_{nm_2}^{[2]})^T, \\
&\mu^{[2]}(o^{[1]}(k)) = (\mu_{11}^{[2]}(o_1^{[1]}(k)), \mu_{12}^{[2]}(o_1^{[1]}(k)), \ldots, \mu_{1m_2}^{[2]}(o_1^{[1]}(k)), \ldots \\
&\ldots, \mu_{l1}^{[2]}(o_l^{[1]}(k)), \mu_{l2}^{[2]}(o_l^{[1]}(k)), \ldots, \mu_{lm_2}^{[2]}(o_l^{[1]}(k)), \ldots \\
&\ldots, \mu_{n1}^{[2]}(o_n^{[1]}(k)), \ldots, \mu_{nm_2}^{[2]}(o_n^{[1]}(k)))^T,
\end{aligned}$$

критерий (8) можно переписать в виде

$$E^k = \frac{1}{2} \sum_{k=1}^{N} (y(k) - w^{[2]T} \mu^{[2]}(o^{[1]}(k)))^2, \tag{9}$$

откуда видно, что для настройки выходного слоя может быть использован стандартный метод наименьших квадратов

$$w^{[2]}(N) = \left(\sum_{k=1}^{N} \mu^{[2]}(o^{[1]}(k)) \mu^{[2]T}(o^{[1]}(k)) \right)^{-1} \left(\sum_{k=1}^{N} \mu^{[2]}(o^{[1]}(k)) y(k) \right), \tag{10}$$

либо его рекуррентная форма

$$\begin{cases}
w^{[2]}(k+1) = w^{[2]}(k) + \dfrac{P(k)(y(k+1) - w^{[2]T}(k)\mu^{[2]}(o^{[1]}(k+1)))\mu^{[2]}(o^{[1]}(k+1))}{1 + \mu^{[2]T}(o^{[1]}(k+1))P(k)\mu^{[2]}(o^{[1]}(k+1))}, \\
P(k+1) = P(k) - \dfrac{P(k)\mu^{[2]}(o^{[1]}(k+1))\mu^{[2]T}(o^{[1]}(k+1))P(k)}{1 + \mu^{[2]T}(o^{[1]}(k+1))P(k)\mu^{[2]}(o^{[1]}(k+1))}.
\end{cases} \tag{11}$$

Настройка весов скрытого слоя вызывает бóльшие затруднения, поскольку выходной сигнал сети зависит от них нелинейно. В простейшем случае, используя идею обратного распространения ошибок, можно применить обычную градиентную процедуру локальной оптимизации

$$\begin{aligned} w_{ih}^{[1,l]}(k+1) = w_{ih}^{[1,l]}(k) - \eta(k)\frac{\partial \frac{1}{2}e^2(k)}{\partial w_{ih}^{[1,l]}} = \\ = w_{ih}^{[1,l]}(k) - \eta(k)\frac{\partial \frac{1}{2}e^2(k)}{\partial \hat{y}(k)}\frac{\partial \hat{y}(k)}{\partial o_l^{[1]}(k)}\frac{\partial o_l^{[1]}(k)}{\partial w_{ih}^{[1,l]}} = \\ = w_{ih}^{[1,l]}(k) + \eta(k)e(k)\mu_{ih}^{[1]}(x_i(k))\sum_{j=1}^{m_2} w_{lj}^{[2]}(k)\frac{\partial \mu_{lj}^{[2]}(o_l^{[1]}(k))}{\partial o_l^{[1]}}, \end{aligned} \quad (12)$$

где $\eta(k)$ — параметр шага, определяющий скорость сходимости процесса обучения и с нахождением которого связаны существенные проблемы.

Вводя далее в рассмотрение $(dm_1 n \times 1)$-вектор весов скрытого слоя

$$\begin{aligned} w^{[1]} = (w_{11}^{[1,1]}, w_{12}^{[1,1]}, \ldots, w_{1h}^{[1,1]}, \ldots, w_{1m_1}^{[1,1]}, w_{21}^{[1,1]}, w_{22}^{[1,1]}, \ldots, w_{2m_1}^{[1,1]}, \ldots \\ \ldots, w_{d1}^{[1,1]}, \ldots, w_{dm_1}^{[1,1]}, w_{11}^{[1,2]}, \ldots, w_{dm_1}^{[1,2]}, \ldots, w_{ih}^{[1,l]}, \ldots w_{dm_1}^{[1,n]})^T, \end{aligned}$$

$(dm_1 \times 1)$-вектор функций принадлежности скрытого слоя

$$\begin{aligned} \mu^{[1]} = (\mu_{11}^{[1]}(x_1), \mu_{12}^{[1]}(x_1), \ldots, \mu_{1m_1}^{[1]}(x_1), \mu_{21}^{[1]}(x_2), \mu_{22}^{[1]}(x_2), \ldots, \mu_{2m_1}^{[1]}(x_2), \ldots \\ \ldots, \mu_{ih}^{[1]}(x_i), \ldots, \mu_{d1}^{[1]}(x_d), \ldots, \mu_{dm_1}^{[1]}(x_d))^T, \end{aligned}$$

вспомогательный $(dm_1 n \times 1)$-вектор $\mu^{[1]} \otimes I_n$ (здесь $\otimes$ — символ тензорного произведения, I_n — $(n \times 1)$-вектор, образованный единицами), $(n \times 1)$-вектор-градиент выхода сети по выходам скрытого слоя

$$\nabla_{o^{[1]}}\ \hat{y} = \left(\frac{\partial \hat{y}(k)}{\partial o_1^{[1]}}, \ldots, \frac{\partial \hat{y}(k)}{\partial o_l^{[1]}}, \ldots, \frac{\partial \hat{y}(k)}{\partial o_n^{[1]}}\right)^T$$

(здесь $\frac{\partial \hat{y}(k)}{\partial o_l^{[1]}} = \sum_{j=1}^{m_2} w_{lj}^{[2]}(k)\frac{\partial \mu_{lj}^{[2]}(o_l^{[1]}(k))}{\partial o_l^{[1]}}$), вспомогательный $(dm_1 n \times 1)$-вектор $\nabla_{o^{[1]}}\hat{y} \otimes I_{m_1 d}$ и, наконец, $(dm_1 n \times 1)$ обобщенный вектор входов $\varphi = (\mu^{[1]} \otimes I_n) \odot (\nabla_{o^{[1]}}\hat{y} \otimes I_{m,d})$ (здесь $\odot$ — символ прямого произведения), можно переписать процедуру обучения (12) в векторной форме

$$\begin{aligned} w^{[1]}(k+1) = w^{[1]}(k) - \eta(k)\nabla_{w^{[1]}}\frac{1}{2}e^2(k) = \\ = w^{[1]}(k) + \eta(k)e(k)\nabla_{w^{[1]}}\hat{y}(k) = w^{[1]}(k) + \eta(k)e(k)\varphi(k). \end{aligned} \quad (13)$$

С позиций качества процесса обучения конструкции (12) и (13) ничем не отличаются, однако последняя может быть принята за основу при синтезе более эффективной процедуры на основе алгоритма Левенберга-Марквардта.

Вводя $(dm_1 n \times N)$-матрицу $\Phi(k) = (\varphi(1), \varphi(2), \ldots, \varphi(N))$, $(N \times 1)$-вектор ошибок обучения $\vec{e}(k) = (e(1), e(2), \ldots, e(N))^T$ и $(dm_1 n \times dm_1 n)$ единичную

 ISSN 1562-9945

матрицу E, несложно записать алгоритм обучения синаптических весов скрытого слоя в виде

$$w^{[1]}(\tau+1) = w^{[1]}(\tau) + (\Phi(k)\Phi^T(k) + \rho E)^{-1}\Phi(k)\vec{e}(k), \tag{14}$$

где ρ — неотрицательный регуляризующий параметр, $\tau = 0, 1, 2\ldots$ — индекс ускоренного машинного времени.

Обучение NFKN с помощью предложенного алгоритма осуществляется по эпохам. Вначале на основе обучающей выборки и при достаточно произвольных начальных значениях $w^{[1]}(0)$, $w^{[2]}(0)$ с помощью выражения (10) (или (11)) вычисляются значения весов выходного слоя и на их основе с помощью (14) пересчитываются веса скрытого слоя. Веса $w^{[1]}(1)$, $w^{[2]}(1)$ далее с помощью той же выборки и тех же процедур вновь уточняются, после чего $w^{[1]}(2)$, $w^{[2]}(2)$ запускаются на следующую эпоху обучения. После того, как веса застабилизируются на некотором уровне, процесс обучения может считаться завершенным.

Выводы

В работе предложен пакетный алгоритм обучения нейро-фаззи сети Колмогорова на основе обратного распространения ошибок и алгоритмов оптимизации второго порядка. Целесообразно использование предложенного алгоритма для решения задач классификации, прогнозирования временных рядов, эмуляции, нейроуправления и других приложений, связанных с интеллектуальным анализом данных, при наличии выборки обучающих наблюдений.

литература

1. Колмогоров А. Н. О представлении непрерывных функций нескольких перемен- ных в виде суперпозиции непрерывных функций одного переменного и сложения // ДАН СССР. — 1957. — №5. — С. 953–956.
2. Hecht-Nielsen R. Kolmogorov’s mapping neural network existence theorem // Proc. IEEE Int. Conf. on Neural Networks. — San Diego, CA, 1987. — 3. — P. 11–14.
3. Kurkova V. Kolmogorov’s theorem and multilayer neural networks Neural Networks. — 1992. — 5. — P. 501–506.
4. Sprecher D. A. A numerical implementation of Kolmogorov’s superpositions II // Neural Networks. — 1997. — N10. — P. 447–457.
5. Igelnik B., Parikh N. Kolmogorov’s spline network // IEEE Trans. on Neural Networks. — 2003. — N14. — P. 725–733.
6. Lopez-Gomez A., Yoshida S., Hirota K. Fuzzy functional link network and its application to the representation of the extended Kolmogorov theorem // Int. J. of Fuzzy Systems. — 2002. — N4. — P. 690–695.
7. Lopez-Gomez A., Hirota K. Fuzzif ication of the Kolmogorov theorem // J. of Advanced Computational Intelligence. — 2001. — N5. — P. 99–109.

8. Бодянский Е. В., Горшков Е. В., Колодяжный В. В. Нейро-фаззи-сеть Колмого- рова с полиномиальными функциями принадлежности и градиентный алгоритм ее обучения // Адаптивні системи автоматичного управління. — 2004. — №7(27). — С. 102–109.

9. Miki T., Yamakawa T. Analog implementation of neo-fuzzy neuron and its on-board learning // Mastorakis N. E. (ed.): Computational Intelligence and Applications. — WSES Press, Piraeus, 1999. — P. 144–149.

10. Yamakawa T., Uchino E., Miki T., Kusanagi H. A neo fuzzy neuron and its applications to system identification and prediction of the system behavior // Proc. 2nd Int. Conf. on Fuzzy Logic and Neural Networks «IIZUKA-92». — Iizuka, Japan, 1992. — P. 477–483.

11. Kosko B. Fuzzy systems as universal approximators // Proc. 1st IEEE Int. Conf. on Fuzzy Systems. — San Diego, CA, 1992. — P. 1153–1162.

12. Yam Y., Nguyen H. T., Kreinovich V. Multi-resolution techniques in the rules-based intelligent control systems: a universal approximation result // Proc. 14th IEEE Int. Symp. on Intelligent Control/Intelligent Systems and Semiotics ISIC/ISAS-99. — Cambridge, Massachusetts, September 15–17, 1999. — 3. — P. 213–218.

 ISSN 1562-9945