УДК 519.7:007.52

А.А. Беднарская, Е.А. Винокурова, И.П. Плисс

Харьковский национальный университет радиоэлектроники

АДАПТИВНЫЙ АЛГОРИТМ ОБУЧЕНИЯ ФАЗЗИ-ВЭЙВЛЕТ-НЕЙРОННОЙ СЕТИ

В статье предложена структура фаззи-вэйвлет-нейронной сети и ее алгоритм обучения для моделирования нелинейных систем в условиях априорной и текущей неопределенности. Алгоритм обучения оптимален по быстродействию и прост в реализации. В качестве активационных функций использованы различные виды семейств вэйвлетов. Имитационное моделирование разработанной структуры фаззи-вэйвлет-нейронной сети и ее алгоритма обучения подтверждают эффективность предложенного подхода.

вэйвлеты, фаззи-вэйвлет-нейронная сеть, алгоритм обучения, вычислительный интеллект

Введение

В настоящее время для обработки информации, содержащей сложные нелинейные закономерности и искаженной различного рода возмущениями, все более широкое распространение получают нейро-фаззи системы [1 – 4], реализующие технику мягких вычислений. Эти системы объединяют в себе лингвистическую интерпретируемость и аппроксимирующие свойства систем нечеткого вывода [5, 6] с обучаемостью и аппроксимирующими возможностями искусственных нейронных сетей [7, 8]. Это означает, что они с успехом могут быть использованы в задачах моделирования (эмуляции) нелинейных систем, описываемых уравнениями вида

$$y(k) = F(x(k)) + \xi(k),$$

где $y(k)$ – выходной сигнал системы в k-й момент дискретного времени $k = 0, 1, 2, \ldots$; $x(k)$ – $(n\times1)$-вектор входных сигналов; $F(\bullet)$ – произвольная функция в общем случае неизвестного вида; $\xi(k)$ – внешнее стохастическое возмущение с неизвестными характеристиками. Предполагается также, что функция $F(\bullet)$ задана на ортотопе

$$x_i(k) \in [x_i^{\min}, x_i^{\max}],\ i = 1, 2, \ldots, n,$$

где $x_i^{\min}$, $x_i^{\max}$ – известные нижняя и верхняя границы варьирования i-го входного воздействия.

Наряду с нейро-фаззи системами для обработки сигналов различной природы все чаще применяется аппарат вэйвлет-преобразования [9 – 11], обеспечивающий компактное локальное представление сигналов как в частотной, так и во временной области. На стыке теории искусственных нейронных сетей и вэйвлетов возникли вэйвлет-нейронные сети [12 – 18], для анализа нестационарных процессов с существенно нелинейной основой.

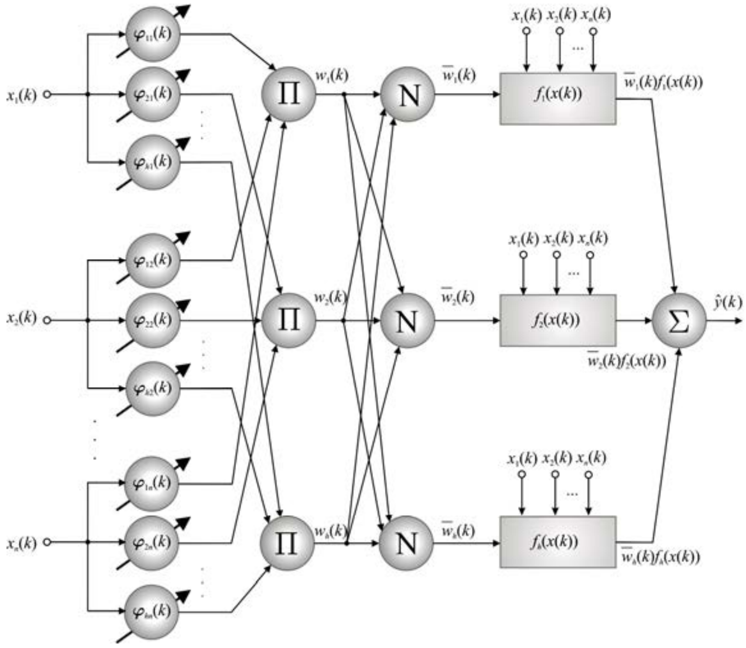

Рис. 1. Фаззи-вэйвлет-нейронная сеть

Естественным дальнейшим шагом является объединение прозрачности и интерпретируемости систем нечеткого вывода, мощных аппроксимирующих возможностей и обучаемости искусственных нейронных сетей и компактности и гибкости вэйвлет-преобразования в рамках гибридной системы вычислительного интеллекта, которую в дальнейшем будем именовать фаззи-вэйвлет-нейронной сетью.

Ключевым моментом, определяющим эффективность такого рода систем, является выбор алгоритма обучения, в основе которого чаще всего лежат градиентные процедуры минимизации принятого критерия. Сочетание градиентной оптимизации с обратным распространением ошибки приводит к низкой скорости процесса обучения гибридной системы [19] и необходимости использования обучающих выборок большого объема. В случае, когда обработка данных должна проводиться в реальном времени, моделируемая система нестационарна, а сигналы зашумлены помехами, традиционные градиентные алгоритмы обучения (не говоря уже о генетических) оказываются неэффективными.

Настоящая работа посвящена вопросам синтеза адаптивного алгоритма обучения фаззи-вэйвлет-нейронной сети, обладающего повышенной скоростью сходимости по сравнению с общепринятыми градиентными процедурами обратного распространения.

1. Архитектура фаззи-вэйвлет-нейронной сети

Введем в рассмотрение пятислойную архитектуру, приведенную на рис. 1 и совпадающую по сути с известной ANFIS [2], являющейся, в свою очередь, обучаемой системой нечеткого вывода Такаги-Сугено-Канга [20, 21].

Нулевой слой архитектуры является рецепторным и в текущий момент времени k на него подается входной сигнал в форме вектора

$$x(k) = (x_1(k), x_2(k), \ldots, x_n(k))^T$$

Первый скрытый слой в отличие от нейро-фаззи систем образован не традиционными неотрицательными функциями принадлежности, а набором из hn вэйвлетов (по h вэйвлетов на каждый вход) $\varphi_{ji}(x_i) = \varphi_{ji}(x_i, c_{ji}, \sigma_{ji})$ с 2hn настраиваемыми параметрами центра c_{ij}. и ширины σ_{ij}.

В качестве вэйвлетов в рассматриваемой фаззи-вэйвлет-нейронной сети могут использоваться различные аналитические вэйвлеты: вэйвлет Морле, вэйвлет «Мексиканская шляпа», Polywog вэйвлеты, Rasp вэйвлеты [12], генератор аналитических вэйвлетов [22], треугольный вэйвлет [23] и т.д.

Здесь можно отметить, что колебательный характер вэйвлет-функций не противоречит униполярности функций принадлежности, поскольку отрицательные значения $\varphi_{ji}(x_i)$ могут трактоваться в смысле малых уровней принадлежности или непринадлежности [24, 25].

Второй скрытый слой реализует операцию, аналогичную вычислению нечеткой Т-нормы

$$w_j(k) = \prod_{i=1}^{n} \varphi_{ji}(x_i(k)),\ j = 1, 2, \ldots, h,$$

после чего в третьем скрытом слое производится нормализация

$$\bar{w}_j(k) = w_j(k) \Big/ \sum_{j=1}^{h} w_j(k) = \prod_{i=1}^{n} \varphi_{ji}(x_i(k)) \Big/ \sum_{j=1}^{h} \prod_{i=1}^{n} \varphi_{ji}(x_i(k)),$$

обеспечивающая выполнение условия $\sum_{j=1}^{h} \bar{w}_j(k) = 1$.

Четвертый скрытый слой реализует операцию, аналогичную вычислению консеквента в системах нечеткого вывода, при этом в качестве функций $f_j(x(k))$ наиболее часто используют линейную форму

$$f_j(x(k)) = p_{j0} + \sum_{i=1}^{n} p_{ji} x_i(k).$$

В этом случае в четвертом слое вычисляются значения сигналов

$$\bar{w}_j(k)(p_{j0} + \sum_{i=1}^{n} p_{ji} x_i(k)) = \bar{w}_j(k) p_j^T \bar{x}(k),$$

где $\bar{x}(k) = (1, x^T(k))^T$, $p_j = (p_{j0}, p_{j1}, \ldots, p_{jn})^T$, а $h(n+1)$ параметров p_{ji}, $j = 1,2,\ldots,h$, $i = 0,1,2,\ldots,n$ подлежат определению.

И, наконец, в пятом выходном слое вычисляется выходной сигнал сети

$$\hat{y}(k) = \sum_{j=1}^{h} \bar{w}_j(k) f_j(x(k)) = \sum_{j=1}^{h} \frac{w_j(k)}{\sum_{j=1}^{h} w_j(k)} f_j(x(k)) =$$

$$= \sum_{j=1}^{h} \frac{\prod_{i=1}^{n} \varphi_{ji}(x_i(k), c_{ji}, \sigma_{ji})}{\sum_{j=1}^{h} \prod_{i=1}^{n} \varphi_{ji}(x_i(k), c_{ji}, \sigma_{ji})} f_j(x(k)),$$

который, вводя векторы переменных $f(x(k)) = (\bar{w}_1(k), \bar{w}_1(k)x_1(k), \ldots, \bar{w}_1(k)x_n(k), \bar{w}_2(k), \bar{w}_2(k)x_1(k), \ldots, \bar{w}_2(k)x_n(k), \ldots, \bar{w}_h(k), \bar{w}_h(k)x_1(k), \ldots, \bar{w}_h(k)x_n(k))^T$, $p = (p_{10}, p_{11}, \ldots, p_{1n}, p_{20}, p_{21}, \ldots, p_{2n}, p_{h0}, p_{h1}, \ldots, p_{hn})^T$ размерности $h(n+1)$, можно записать в компактной форме

$$\hat{y}(k) = p^T f(x(k)).$$

Настраиваемые параметры этой сети находятся только в первом и четвертом скрытом слоях. Это $2hn$ параметров вэйвлетов c_{ij} и σ_{ij} и $h(n+1)$ параметров линейных локальных регрессионных моделей p_{ij}. Именно они и должны быть определены в процессе обучения.

2. Обучение фаззи-вэйвлет-нейронной сети

Поскольку искомый вектор параметров p входит в описание сети линейно, для его уточнения может использоваться любой из алгоритмов, используемых в адаптивной идентификации [26], и, прежде всего, рекуррентный метод наименьших квадратов в форме

$$\begin{cases} p(k+1) = p(k) + \dfrac{P(k)(y(k) - p^T(k) f(x(k)))}{1 + f^T(x(k)) P(k) f(x(k))} f(x(k)), \\ P(k+1) = P(k) - \dfrac{P(k) f(x(k)) f^T(x(k)) P(k)}{1 + f^T(x(k)) P(k) f(x(k))} \end{cases}$$

(здесь $p^T(k) f(x(k)) = \hat{y}(k)$), являющийся по сути процедурой оптимизации второго порядка и обладающий сглаживающими свойствами; оптимальный по быстродействию одношаговый градиентный алгоритм Качмажа [27, 28], обладающий следящими свойствами

$$p(k+1) = p(k) + \frac{y(k) - p^T(k) f(x(k))}{f^T(x(k)) f(x(k))} f(x(k)),$$

или алгоритм Гудвина-Рэмеджа-Кэйнеса [29]

$$\begin{cases} p(k+1) = p(k) + r^{-1}(k)(y(k) - p^T f(x(k))) f(x(k)); \\ r(k+1) = r(k) + \| \varphi(x(k+1)) \|^2, \end{cases}$$

являющийся по сути процедурой стохастической аппроксимации).

Для настройки параметров первого скрытого слоя в ANFIS используется алгоритм обратного распространения ошибки, основанный на цепном правиле дифференцирования и градиентной оптимизации локального критерия

$$E(k) = \frac{1}{2} e^2(k) = \frac{1}{2}(y(k) - \hat{y}(k))^2.$$

В общем случае процедура обучения параметров этого слоя имеет вид

$$\begin{cases} c_{ji}(k+1) = c_{ji}(k) - \eta_c(k) \dfrac{\partial E(k)}{\partial c_{ji}(k)}; \\ \sigma_{ji}(k+1) = \sigma_{ji}(k) - \eta_\sigma(k) \dfrac{\partial E(k)}{\partial \sigma_{ji}(k)}, \end{cases}$$

а ее свойства полностью определены параметрами шага $\eta_c(k)$, $\eta_\sigma(k)$, выбираемыми из эмпирических соображений. Следует отметить, что, если параметры четвертого слоя могут быть настроены максимально быстро, то именно в первом слое это быстродействие теряется.

Повышение скорости сходимости может быть достигнуто путем использования более сложных, нежели градиентных, процедур типа алгоритма Хартли [30] или Марквардта [31], которые могут быть записаны в обобщенной форме [32]

$$\Phi(k+1) = \Phi(k) + \lambda (J(k) J^T(k) + \eta I)^{-1} J(k) e(k), \quad (1)$$

где $\Phi(k) = (c_{11}(k), \sigma_{11}^{-1}(k), c_{21}(k), \sigma_{21}^{-1}(k), \ldots, c_{ji}(k), \sigma_{ji}^{-1}(k), \ldots, c_{hn}(k), \sigma_{hn}^{-1}(k))^T$ – ($2hn \times 1$)-вектор настраиваемых параметров, причем для упрощения численной реализации в него входит не собственно параметр ширины σ_{ji}, а обратная ему величина σ_{ji}^{-1}; $J(k) - (2hn \times 1)$-вектор-градиент выходного сигнала $\hat{y}(k)$ по настраиваемым параметрам; $I - (2hn \times 2hn)$-единичная матрица; η – скалярный регуляризующий параметр; λ – положительный скалярный коэффициент усиления (демпфирования).

Для вычисления компонент вектора-градиента

$$J(k) = \left(\frac{\partial \hat{y}(k)}{\partial c_{11}}, \frac{\partial \hat{y}(k)}{\partial \sigma_{11}^{-1}}, \frac{\partial \hat{y}(k)}{\partial c_{21}}, \frac{\partial \hat{y}(k)}{\partial \sigma_{21}^{-1}}, \ldots, \frac{\partial \hat{y}(k)}{\partial c_{ji}}, \frac{\partial \hat{y}(k)}{\partial \sigma_{ji}^{-1}}, \ldots, \frac{\partial \hat{y}(k)}{\partial c_{hn}}, \frac{\partial \hat{y}(k)}{\partial \sigma_{hn}^{-1}} \right)^T$$

используется цепное правило, при этом

$$\begin{cases} \dfrac{\partial \hat{y}(k)}{\partial c_{ji}} = \dfrac{\partial \hat{y}(k)}{\partial \overline{w}_j} \cdot \dfrac{\partial \overline{w}_j}{\partial w_j} \cdot \dfrac{\partial w_j}{\partial \varphi_{ji}} \cdot \dfrac{\partial \varphi_{ji}}{\partial c_{ji}}; \\ \dfrac{\partial \hat{y}(k)}{\partial \sigma_{ji}^{-1}} = \dfrac{\partial \hat{y}(k)}{\partial \overline{w}_j} \cdot \dfrac{\partial \overline{w}_j}{\partial w_j} \cdot \dfrac{\partial w_j}{\partial \varphi_{ji}} \cdot \dfrac{\partial \varphi_{ji}}{\partial \sigma_{ji}^{-1}}. \end{cases}$$

С целью упрощения вычислительной сложности алгоритма обучения можно воспользоваться леммой обращения матриц в виде

$$(JJ^T + \eta I)^{-1} = \eta^{-1} I - \frac{\eta^{-1} IJJ^T \eta^{-1} I}{1 + J^T \eta^{-1} IJ},$$

с помощью которой несложно получить соотношение

$$\lambda (JJ^T + \eta I)^{-1} J = \lambda \left(J \middle/ \left(\eta + \|J\|^2 \right) \right),$$

подставляя которое в (1) получим алгоритм обучения параметров первого скрытого слоя в виде

$$\Phi(k+1) = \Phi(k) + \lambda \frac{J(k) e(k)}{\eta + \|J(k)\|^2}. \qquad (2)$$

Несложно видеть, что алгоритм (2) является нелинейной аддитивно-мультипликативной модификацией алгоритма Качмажа, а при $\lambda = 1, \eta = 0$ структурно с ним совпадает.

Заключение

Предложен численно простой оптимальный по быстродействию алгоритм обучения фаззи-вэйвлет-нейронной сети в задачах моделирования нелинейных систем. Экспериментальные исследования подтверждают его эффективность по сравнению с традиционными алгоритмами обучения гибридных систем вычислительного интеллекта.

Список литературы

1. Jang J.-S. R. Neuro-Fuzzy Modeling: Architectures, Analyses and Applications. – Berkeley, CA: University of California, 1992. – 155 p.

2. Jang J.-S. R. ANFIS: Adaptive network-based fuzzy inference systems // IEEE Tans. on Systems, Man, and Cybernetics. – 1993. – 23. – P. 665-685.

3. Jang J.-S. R., Sun C.-T. Neuro-fuzzy modeling and control // Proc. IEEE. – 1995. – 83 (3). – P. 378-406.

4. Jang J.-S. R., Sun C.-T., Muzutani E. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. – Upper Saddle River, N.J.: Prentice Hall, Inc., 1997. – 614 p.

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

6. Kreinovich V., Mouzouris Y. C., Nguen H. T. Fuzzy rule based modeling as a universal approximation tool // Fuzzy systems: Modeling and Control. – Boston: Kluwer Academic Publishers, 1998. – P. 135-195.

7. Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators // Neural Networks. – 1982. – 2. – P. 359-366.

8. Scarcelli F., Tsoi A.S. Universal approximation using feedforward neural networks: a survey of some existing methods and some new results // Neural Networks. – 1998. – 11. – P. 15-37.

9. Chui C. K. An Introduction to Wavelets. - New York: Academic Press. – 1992. – 264 p.

10. Daubechies I. Ten Lectures on Wavelets. – Philadelphia, PA: SIAM. – 1992. – 228 p.

11. Meyer Y. Wavelets: Algorithms and Applications. – Philadelphia, PA: SIAM. – 1993. – 133 p.

12. Lekutai G., VanLandingham H. F. Self-tuning control of nonlinear systems using neural network adaptive frame wavelets // Proc. IEEE Int. Conf. on Systems, Man and Cybernetics. Piscataway, N.J. – 2. – 1997. – P. 1017-1022.

13. Billings S. A., Wei H.-L. A new class of wavelet networks for nonlinear system identification // IEEE Trans. on Neural Networks. – 16 (4). – 2005. – P. 862-874.

14. Zhang Q. H., Benveniste A. Wavelet networks // IEEE Trans. on Neural Networks. – 3 (6). – 1992. – P. 889-898.

15. Oussar Y., Dreyfus G. Initialization by selection for wavelet network training // Neurocomputing. – 34. – 2000. –P. 131-143.

16. Zhang J., Walter G. G., Miao Y., Lee W. N. W. Wavelet neural networks for function learning // IEEE Trans. on Signal Process. – 43 (6). – 1995. – P. 1485-1497.

17. Zhang Q.H. Using wavelet network in nonparametric estimation // IEEE Trans. on Neural Networks. – 8 (2). – 1997. – P. 227-236.

18. Bodyanskiy Ye., Lamonova N., Pliss I., Vynokurova O. An adaptive learning algorithm for a wavelet neural network // Expert Systems. – Blackwell Synergy Ltd. – 22 (5). – 2005. – P. 235-240.

19. Avci E., Turkoglu I., Poyraz M. Intelligent target recognition based on wavelet adaptive network based fuzzy inference system // Lecture Notes on Computer Science. – 3522. – Berlin-Heidelberg: Springer-Verlag, 2005. – P. 594-603.

20. Takagi T., Sugeno M. Fuzzy identification of systems and its applications to modeling and control // IEEE Trans. on Systems, Man, and Cybernetics. – 1985. – 15. – P. 115-132.

21. Sugeno M., Kang G. T. Structure identification of fuzzy model // Fuzzy Sets and Systems. – 1998. – 28. – P. 15-33.

22. Винокурова Е. А., Ламонова Н. С., Плисс И. П. Генератор аналитических вэйвлетов // Проблемы бионики. – Вып. 60. – 2004. – С. 104-109.

23. Бодянский Е. В., Винокурова Е. А. Треугольные вэйвлеты переменной формы в вэйвлет-нейронных сетях // Сб. науч. тр. 13-й между. конф. «Автоматика 2006». – Винница: ВНТУ, 2006. – С. 393.

24. Mitaim S., Kosko B. What is the best shape for a fuzzy set in function approximation? // Proc. 5th IEEE Int. Conf on Fuzzy Systems "Fuzz-96". – 2. – 1996. – P. 1237-1213.

25. Mitaim S., Kosko B. Adaptive joint fuzzy sets for function approximation // Proc. 1997 Int. Conf. on Neural Networks "ICNN-97". – P. 537-542.

26. Льюнг Л. Идентификация систем. Теория для пользователя. – М.: Наука, ГРФМЛ, 1991. – 432 с.

27. Kaczmarz S. Angenaeherte Ausloesung von Systemen linearer Gleichungen // Bull. Int. Acad. Polon. Sci. – 1973. – Let. A. – S. 355-357.

28. Kaczmarz S. Approximate Solution of Systems of linear equation // Int. J. Control. – 1993. – 53. – P. 1269-1271.

29. Goodwin Y. C., Ramadge P. J., Caines P. E. A globally convergent adaptive predictor // Automatica. – 1981. – 17. – P. 135-140.

30. Hartley H. The modified Gauss-Newton method for the fitting of nonlinear regression functions of least squares // Technometrics. – 1961. – 3. – P. 269-280.

31. Marquardt D. W. An algorithm for least squares estimation of nonlinear parameters // SIAM J. Appl. Math. – 1963. – 11. – P. 431-441.

32. Бодянский Е. В. Адаптивные алгоритмы идентификации нелинейных объектов управления // АСУ и приборы автоматики. – 1987. – Вып. 81. – С. 43-46.

Поступила в редколлегию 27.12.2006

Рецензент: д-р техн. наук, проф. Е.В. Бодянский, Харьковский национальный университет радиоэлектроники, Харьков.