

УДК 519.7:007.52

Е.В. Бодянский, А.П. Чапланов, Е.Б. Чапанова

Харьковский национальный университет радиоэлектроники

РАДИАЛЬНО-БАЗИСНАЯ НЕЙРОННАЯ СЕТЬ С ПОЛИНОМИАЛЬНЫМИ ФУНКЦИЯМИ АКТИВАЦИИ

Рассмотрен синтез алгоритма обучения радиально-базисной нейронной сети с полиномиальными функциями активации 4-й степени для задачи прогнозирования временных рядов.

алгоритм обучения, радиально-базисная нейронная сеть, функция активации

Введение

В настоящее время для анализа временных последовательностей все чаще используются методы на основе аппарата искусственных нейронных сетей.

Большое внимание к нейронным сетям обусловлено их сходством с биологическими системами, обладающими высокими способностями к идентификации и адаптации. Параллельный характер сетей позволяет существенно повысить скорость выполнения вычислений. Благодаря наличию нелинейных передаточных функций, нейросеть способна аппроксимировать существенно нелинейные функции с любой степенью точности. Тот факт, что параметры сети настраиваются лишь на основе данных о входе и выходе, дает возможность использования сетей как адаптивных или обучающихся систем в условиях неопределенности. Благодаря этим качествам, искусственные нейронные сети показали высокие результаты в задачах распознавания образов, в проблемах оптимизации, идентификации и прогнозирования временных последовательностей. В частности, для распознавания образов, в сложных административных системах, в банках и на биржах для прогнозирования курсов валют, где модели носят нелинейный характер и требуют решения задач идентификации в реальном времени [1].

Актуальность исследований в этом направлении подтверждается массой возможных сфер применения нейросетей. Это автоматизация процессов распознавания визуальных, акустических и искусственно синтезированных образов, адаптивное управление [2], обработка сигналов, организация ассоциативной памяти, задачи оптимизации и линейной алгебры большой размерности, задачи обучения, робототехнические задачи, задачи прогнозирования, диагностики и многие другие приложения [3]. Применение искусственных нейронных сетей является эффективным инструментом решения актуальных проблем в технике, экономике, биологии и медицине.

В число этих проблем входит ряд задач, решение которых осложняется именно нелинейностью, нелокальностью, дискретностью и нестационарно-

стью постановки, а именно: задачи прогнозирования временных последовательностей различной природы в условиях существенной неопределенности относительно структуры и параметров исследуемого процесса в реальном времени.

Одним из эффективных средств решения вышеперечисленных задач является применение радиально-базисных искусственных нейронных сетей из-за их универсальных свойств по аппроксимации нелинейных сигналов и преимуществ перед многослойным перцептроном. На основе данной архитектуры нейронной сети можно решать достаточно большой круг задач. Для обучения сети можно применить градиентные методы настройки параметров. Однако аналитическая форма градиента существенно зависит от выбора радиально-базисных функций. В классической радиально-базисной нейронной сети (РБФ-сети) используются колоколообразные функции (гауссианы), однако можно также использовать и радиально-базисные функции, как, например, полиномиальные функции 4-го порядка.

Постановка задачи. Основной целью проведения данного исследования является разработка архитектуры РБФ-сети с полиномиальными функциями активации 4-го порядка (ПРБФ) и алгоритмов её обучения (настройки) для прогнозирования временных последовательностей.

Полиномиальная РБФ

Как отмечалось выше, наиболее широкое распространение получила функция активации - гауссиан, расширить возможности которой можно, используя многомерный гауссиан

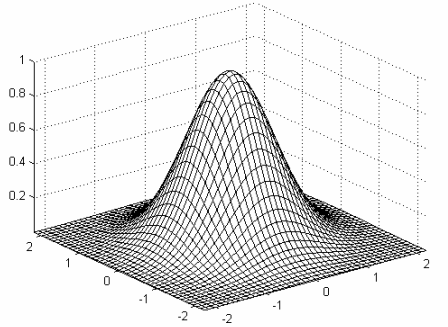
$$\begin{aligned} \varphi(x) &= \Phi(\|x - c\|, \Sigma^{-1}) = \exp(-(x - c)^T \Sigma^{-1} (x - c)) = \\ &= \exp(-\|x - c\|_{\Sigma^{-1}}^2), \end{aligned} \quad (1)$$

где ковариационная матрица Σ^{-1} определяет форму, размер и ориентацию так называемого рецепторного поля радиально-базисной функции. Виды РБФ и соответствующего рецепторного поля представлены на рис. 1 и 2. При $\Sigma = \sigma^2 I$ (I - $(n \times n)$ единичная матрица) рецепторное поле представляет гиперсферу с цен-

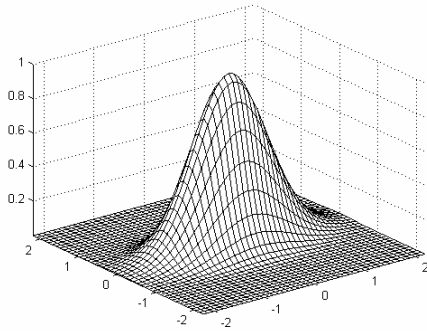
тром c и радиусом σ ; при $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ – это гиперэллипсоид, чьи оси совпадают с осями входного пространства и имеют длину $2\sigma_i$ по i -й оси, и, наконец, при Σ – недиагональной положительно определенной матрице

$$\Sigma = Q^T \Lambda Q, \quad (2)$$

где матрица собственных значений Λ определяет форму и размер рецепторного поля, а ортогональная матрица вращения Q – его ориентацию [3].

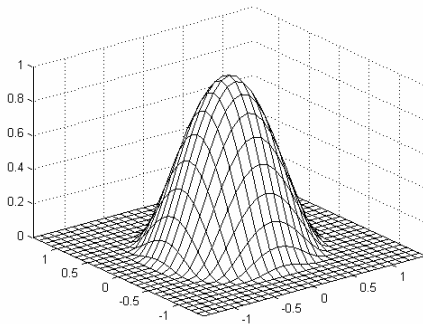


$$\Sigma^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad c = (0,0)^T$$



$$\Sigma^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}, \quad c = (0,0)^T$$

Рис. 1. Многомерный гауссиан



$$y = \begin{cases} \left(1 - \|x - C_i\|_{\Sigma_i^{-1}}^2\right)^2, & \|x - C_i\|_{\Sigma_i^{-1}}^2 < 1; \\ 0, & \|x - C_i\|_{\Sigma_i^{-1}}^2 \geq 1. \end{cases}$$

Рис. 2. Полиномиальная РБФ 4-го порядка

Как следует из вышеизложенного, РБФ сеть характеризуют три типа параметров:

- линейные весовые параметры выходного слоя w_{ij} (входят в описание сети линейно);
- центры c_j – нелинейные параметры скрытого слоя (входят в описание сети нелинейно);
- отклонения (радиусы базисных функций) σ_{ij} – нелинейные параметры скрытого слоя.

Обучение сети, состоящее в определении этих параметров, может сводиться к одному из следующих вариантов.

1. Задаются центры и отклонения, а вычисляются только веса выходного слоя.

2. Определяются путем самообучения центры и отклонения, а для коррекции весов выходного слоя используется парадигма обучения с учителем.

3. Определяются все параметры сети с помощью парадигмы обучения с учителем.

Итак выход РБФ-сети вычисляется в соответствии со следующим выражением:

$$\begin{aligned} \mathfrak{f}(k) &= w_0 + \sum_{i=1}^h w_i \Phi_i(\|x(k) - C_i\|_{\Sigma_i^{-1}}^2) = \\ &= \sum_{i=0}^h w_i \varphi_i(x(k)) = w^T \varphi(x(k)); \\ \varphi_i(x(k)) &= \Phi_i(\|x(k) - C_i\|_{\Sigma_i^{-1}}^2); \end{aligned} \quad (3)$$

$$\varphi_0(x(k)) \equiv 1, i = 1, \dots, h;$$

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

$$w = (w_0, w_1, \dots, w_h)^T;$$

$$\varphi(x(k)) = (1, \varphi_1(x(k)), \dots, \varphi_h(x(k)))^T,$$

где $\mathfrak{f}(k)$ – выход сети в текущее время k , $x(k)$ – n -мерный входной вектор, h – число модулей (базисных функций) в скрытом слое, $\Phi_i(\bullet)$ – i -ая базисная функция, C_i – n -мерный прототипный вектор (центр) i -й базисной функции, Σ_i^{-1} – матрица ($n \times n$), которая определяет область влияния i -ой базисной функции, $\varphi_i(\bullet)$ – i -ая базисная функция, w – вектор настраиваемых синаптических весов, $\varphi(\bullet)$ – нелинейный регрессионный вектор.

Наиболее широко используемые РБФ – многомерные гауссианы, в общем виде определенные как

$$\Phi_i(\|x - C_i\|_{\Sigma_i^{-1}}^2) = \exp\left(-\frac{1}{2} \|x - C_i\|_{\Sigma_i^{-1}}^2\right), \quad (4)$$

где Σ_i – в общем случае недиагональная, симметрическая и положительно определенная матрица.

Предлагается использовать полиномиальную радиально-базисную функцию 4-го порядка (ПРБФ), выход которой ненулевой только внутри гиперэллипсоидной области (рецепторного поля), т.е.

$$\Phi_i(\|x - C_i\|_{\Sigma_i^{-1}}^2) = \begin{cases} \left(1 - \|x - C_i\|_{\Sigma_i^{-1}}^2\right)^2, & \|x - C_i\|_{\Sigma_i^{-1}}^2 < 1; \\ 0, & \|x - C_i\|_{\Sigma_i^{-1}}^2 \geq 1. \end{cases} \quad (5)$$

ПРБФ – локализована, требует меньше вычислительных усилий по сравнению с гауссовыми РБФ и позволяет получить эффективную упрощенную процедуру обучения. Цель состоит в выборе на каждой итерации k параметров $w(k), C_i(k), \Sigma_i^{-1}(k)$, которые минимизировали бы функцию ошибок

$$E(k) = \frac{1}{2} (d(k) - \hat{y}(k))^2 = \frac{1}{2} e(k)^2, \quad (6)$$

где $d(k)$ – целевой выход, $e(k)$ – ошибка модели в текущее время k .

Ниже представлена процедура обучения градиентного спуска:

$$\begin{cases} w(k) = w(k-1) - \eta_w \nabla_w E(k); \\ C_i(k) = C_i(k-1) - \eta_{C_i} \nabla_{C_i} E(k), \quad i = \overline{1, h}; \\ \Sigma_i^{-1}(k) = \Sigma_i^{-1}(k-1) - \eta_{\Sigma_i^{-1}} \partial E(k) / \partial \Sigma_i^{-1}, \quad i = \overline{1, h}, \end{cases} \quad (7)$$

где $\nabla_w E, \nabla_{C_i} E$ – градиенты функции ошибок по отношению к соответствующим векторам параметров; $\partial E / \partial \Sigma_i^{-1}$ – производная функции ошибок по отношению к матрице Σ_i^{-1} ; $\eta_w, \eta_{C_i}, \eta_{\Sigma_i^{-1}}$ – коэффициенты шага обучения.

Для базисных функций (5), мы можем записать производную $\Phi'_i(\bullet)$ в форме:

$$\Phi'_i(\|x(k) - C_i\|_{\Sigma_i^{-1}}^2) = \begin{cases} -2 \cdot \left(1 - \|x(k) - C_i\|_{\Sigma_i^{-1}}^2\right), & \|x(k) - C_i\|_{\Sigma_i^{-1}}^2 < 1; \\ 0, & \|x(k) - C_i\|_{\Sigma_i^{-1}}^2 \geq 1. \end{cases} \quad (8)$$

Таким образом, производная $\Phi'_i(\bullet)$ ПРБФ равна $-2 \cdot \left(1 - \|x(k) - C_i\|_{\Sigma_i^{-1}}^2\right)$, когда входной сигнал попадает в рецепторное поле и 0 в противном случае. Далее можно записать выражения для градиента функции ошибок по отношению к центрам i -й ПРБФ:

$$\nabla_{C_i} E(k) = \begin{cases} -4e(k)w_i \left(1 - \|x(k) - C_i\|_{\Sigma_i^{-1}}^2\right) \times \\ \times \Sigma_i^{-1} (x(k) - C_i), & \|x(k) - C_i\|_{\Sigma_i^{-1}}^2 < 1; \\ 0, & \|x(k) - C_i\|_{\Sigma_i^{-1}}^2 \geq 1 \end{cases} \quad (9)$$

и производную по отношению к матрице Σ_i^{-1} :

$$\partial E(k) / \partial \Sigma_i^{-1} = \begin{cases} 2e(k)w_i \left(1 - \|x(k) - C_i\|_{\Sigma_i^{-1}}^2\right) \times \\ \times (x(k) - C_i)(x(k) - C_i)^T, & \|x(k) - C_i\|_{\Sigma_i^{-1}}^2 < 1; \\ 0, & \|x(k) - C_i\|_{\Sigma_i^{-1}}^2 \geq 1. \end{cases} \quad (10)$$

Используя выражения (7), (9), (10), получим в результате набор обучающих процедур для всех параметров ПРБФ сети:

$$\begin{cases} w(k) = w(k-1) + \Delta w(k), \\ C_i(k) = C_i(k-1) + \Delta C_i(k), \quad i = 1, \dots, h, \\ \Sigma_i^{-1}(k) = \Sigma_i^{-1}(k-1) + \Delta \Sigma_i^{-1}(k), \quad i = 1, \dots, h, \end{cases} \quad (11)$$

где обновления $\Delta w(k), \Delta C_i(k)$, и $\Delta \Sigma_i^{-1}(k)$ вычисляются в виде:

$$\Delta w(k) = \eta_w e(k) \varphi(x(k)), \quad (12)$$

$$\Delta C_i(k) = \begin{cases} 4\eta_{C_i} e(k) w_i (k-1) \cdot \Sigma_i^{-1}(k-1) (x(k) - C_i(k-1)) \times \\ \times (1 - \|x(k) - C_i(k-1)\|_{\Sigma_i^{-1}(k-1)}^2), \\ \text{если } \|x(k) - C_i(k-1)\|_{\Sigma_i^{-1}(k-1)}^2 < 1; \\ 0, \text{ если } \|x(k) - C_i(k-1)\|_{\Sigma_i^{-1}(k-1)}^2 \geq 1, \end{cases} \quad (13)$$

$$\Delta \Sigma_i^{-1}(k) = \begin{cases} -2\eta_{\Sigma_i^{-1}} e(k) w_i (k-1) \cdot \left(1 - \|x(k) - C_i(k-1)\|_{\Sigma_i^{-1}(k-1)}^2\right) \times \\ \times (x(k) - C_i(k-1))(x(k) - C_i(k-1))^T, \\ \text{если } \|x(k) - C_i(k-1)\|_{\Sigma_i^{-1}(k-1)}^2 < 1; \\ 0, \text{ если } \|x(k) - C_i(k-1)\|_{\Sigma_i^{-1}(k-1)}^2 \geq 1. \end{cases} \quad (14)$$

Как можно увидеть из уравнений (11) – (14) параметры i -й ПРБФ обновляются только когда входной сигнал ненулевой. Это случается, когда выход находится внутри соответствующего рецепторного поля.

Таким образом, в каждый текущий момент времени только ближайшая к текущему входу ПРБФ обновляется. Так обучающие процедуры (11)-(14) напоминают правило конкурентного обучения самоорганизующихся карт, а предложенная нейронная сеть, являясь развитием радиально-базисных структур, позволяет решать задачу прогнозирования временных последовательностей в условиях структурной и параметрической неопределенности [7].

Реализация метода

Работоспособность предложенного подхода продемонстрирована на примере реконструкции в реальном времени хаотического временного ряда, сгенерированного известным дифференциальным уравнением Мэки-Гласса:

$$\dot{x}(t) = \frac{0.2 x(t - \tau)}{1 + x^{10}(t - \tau)} - 0.1 x(t). \quad (15)$$

Значения временного ряда (15) в каждой целой точке были получены с помощью метода Рунге-Кутты четвертого порядка.

Временной шаг метода был равен 0.1, начальные условия $x(0) = 1.2$, задержка $\tau = 17$, и $x(t)$ были получены для $t = 0, \dots, 51000$. Значения $x(t-18), x(t-12), x(t-6)$ и $x(t)$ использовались в качестве входов ($n=4$) для прогнозирования значений $x(t+6)$.

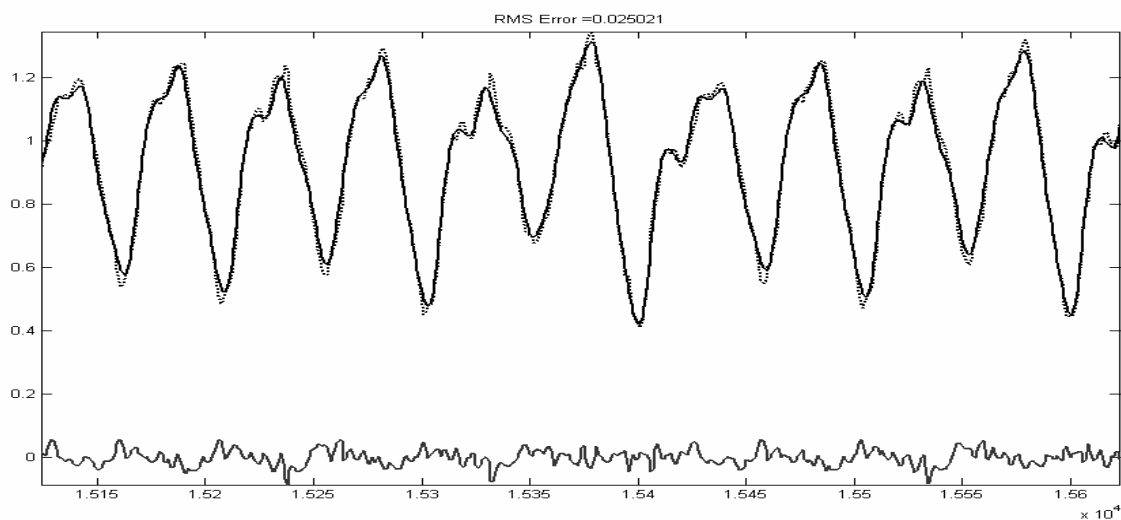


Рис. 3. Прогноз временного ряда Мэки-Гласса

Базисные функции были созданы с помощью метода кластеризации [8] 500 точек данных для $t = 118, \dots, 617$. Метод кластеризации обнаружил 10 кластеров, параметры которых были использованы для инициализации 10 узлов скрытого слоя ($h = 10$). Начальные рецепторные поля выбирались как гиперсферы радиуса $3\sigma = 0.4689$, где параметр σ был найден методом кластеризации. Начальные веса выходного слоя были равны нулю.

С помощью предложенного метода осуществлялась настройка параметров скрытого слоя нейронной сети. Для сравнения также использовался рекуррентный метод наименьших квадратов (РМНК) для настройки весов выходного слоя.

Во всех экспериментах выполнялось обучение сети РБФ с одинаковыми начальными параметрами для $t = 118, \dots, 50117$. Затем использовалась обученная сеть для прогнозирования точек данных из проверочной выборки для $t = 50118, \dots, 50617$. Чтобы оценить точность прогнозирования, вычислялась среднеквадратичная ошибка на проверочных данных ($RMSE_{СНК}$). Результаты показаны в табл. 1.

Таблица 1
Сравнение алгоритмов по точности аппроксимации

№	Используемая сеть	Метод обучения	$RMSE_{СНК}$
1	Классическая РБФ-сеть	РМНК	0.02733
2	ПРБФ	(11)	0.02502

Прогноз данных из проверочной выборки показан на рис. 3. Сплошная линия представляет реальный хаотический процесс, а пунктирная линия обозначает прогноз. Ошибка показана на том же рисунке под графиками реального процесса и прогноза и очень мала.

Выводы

Задача прогнозирования и идентификации временных рядов имеет значительный интерес с точки

зрения практического использования. Так, обнаружение разладок, изменения параметров временных сигналов может быть полезно в отраслях с автоматизированным производством, где большей частью за производство отвечают машины. В том числе, данный подход может быть применим на рынке ценных бумаг, где слежение за характеристиками временных процессов может оказаться существенным для финансистов и аналитиков.

Список литературы

1. Haykin S. *Neural Networks. A Comprehensive Foundation*. – Upper Saddle River, N. J.: Prentice Hall, Inc., 1999. – 842 p.
2. Tsoukalas L.H., Uhrig R.E. *Fuzzy and Neural Approaches in Engineering*. – New York: A Wiley-Interscience Publication, 1997. – 600 p.
3. Rojas I., Gonzalez J., Canas A., Diaz A. F., Rojas F. J., Rodriguez M. *Short-term prediction of chaotic time series by using RBF network with regression weights // Int. J. of Neural Systems*. – 2000. – 10. – P. 353–364.
4. Руденко О.Г., Бодянский Е.В. *Основы теории искусственных нейронных сетей*. – X.: Телетех, 2002. – 317 с.
5. Бодянский Е.В., Руденко О.Г. *Искусственные нейронные сети: архитектуры, обучение, применения*. – X.: ТЕЛТЕХ, 2004. – 369 с.
6. Цыпкин Я.З. *Основы теории обучающихся систем*. – М.: Наука, 1970. – 252 с.
7. Бодянский Е.В., Кучеренко Е.И., Чапманов А.П. *Диагностика и прогнозирование временных рядов с помощью многослойной радиально-базисной нейронной сети // Труды VIII Всероссийской конф. «Нейрокомпьютеры и их применение» НКП-2002*. – М.: Институт проблем управления им В.А. Трапезникова РАН, 2002. – С. 209–213.
8. Chiu, S. *Fuzzy model identification based on cluster estimation // J. of Intelligent & Fuzzy Systems*. – 1994. – 2. – P. 267–278.

Поступила в редколлегию 27.02.2007

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

