

УДК 004.032.26.001.63

Снитюк В.Е., Шарапов В.М.

Черкасский государственный технологический университет, Черкассы, Украина

ЭВОЛЮЦИОННО-ПАРАМЕТРИЧЕСКАЯ ОПТИМИЗАЦИЯ RBF-СЕТИ

Радиально-базисные нейронные сети (RBF) являются достаточно хорошими аппроксиматорами в области изменения данных, на которых происходил процесс их обучения. Вне этой области, в общем случае, поведение сети непредсказуемо, но может быть оптимизировано через настройку ее параметров. В статье предложен метод эволюционного поиска оптимальных окон RBF-сети и рассмотрены различные варианты его реализации.

Введение

Задача аппроксимации сложных нелинейных зависимостей неизбежно возникает при проектировании сложных систем, идентификации законов их функционирования, извлечении знаний из баз данных. Базируясь на классической теореме Вейерштрасса о приближении непрерывной функции полиномом, одним из направлений решения этой задачи является построение временных рядов и полиномов Колмогорова-Габора, Чебышева и других. Представителем этого направления является метод группового учета аргументов (МГУА), композиция которого с методом предельных упрощений (МПУ) рассматривается в [1]. Авторы указывают на то, что большой проблемой их применения есть правильное соотношение сложности модели с объемом обучающей выборки. Кроме того, принцип свободы выбора, реализуемый при определении параметров метода, приводит к значительной неопределенности, что влечет его адекватное применение только опытными специалистами. Временные ряды [2], выбор их модели [3], как и ее усовершенствование [4], ограничивают исследователя одним экзогенным фактором модели.

Другое направление связано с синтетическим подходом к получению аппроксимационной зависимости [5]. Типичным его представителем являются искусственные нейронные сети (ИНС). Базируясь на том, что ИНС с непрерывными функциями активации является универсальным аппроксиматором и теореме Колмогорова о возможности получения любой непрерывной функции n переменных с помощью операций сложения, умножения и суперпозиции из непрерывных функций *одного* переменного [6], их активно используют при решении задач восстановления зависимостей [7], пропущенных данных, классификации, кластеризации, ассоциативного поиска и многочисленных других. ИНС не ограничены требованиями непрерывности, дифференцируемости функций, но синергетический эффект их обучения требует решения задачи структурной и параметрической оптимизации.

RBF-сети являются оптимальным аппроксиматором в области изменения данных, на которых происходит их обучение. Вне этой области функционирование сети дает самые неожиданные результаты. Мы предлагаем использовать для увеличения предсказуемости поведения сети генетический алгоритм (ГА). Эта идея не является новой [6,7,8,9,10,11], однако подавляющая часть работ содержит лишь теоретические предположения, в них отсутствует формализация задачи и

экспериментальные результаты и, кроме того, они ориентированы на ИНС, которые функционируют на базе алгоритма обратного распространения ошибки или подобных. Известно, что такой алгоритм обладает медленной скоростью сходимости и низкой точностью. Алгоритм функционирования RBF-сети является точным, прямым и, оптимизируя ее параметры, можно получить эффективный механизм восстановления аппроксимационных зависимостей.

Постановка задачи

Пусть $\vec{X} = (X_1, X_2, \dots, X_n)$ - вектор входных факторов, $\vec{Y} = (Y_1, Y_2, \dots, Y_m)$ - вектор выходных характеристик, $A = \{x_{i1,j1}, y_{i2,j2}\}_{i1+i2=1}^{n+m}{}_{j1+j2=1}^k$ - матрица обучающих образов, k - их количество. Матрицу A разделим на две части: матрицу A_1 и матрицу A_2 . О том, как это рационально сделать, можно прочесть, например, в [12]. Количество строк в A_1 равно k_1 , в A_2 - k_2 , причем $k_1 + k_2 = k$ и $k_2 < k_1$. Необходимо решить следующую задачу:

$$\frac{1}{mk_2} \sum_{j=1}^m \sum_{i=1}^{k_2} (y_{ij} - d_{ij})^2 \rightarrow \min, \quad (1)$$

где d_{ij} - результат, полученный обученной сетью на j -м выходе, при поданном i -м образе из контрольной матрицы A_2 . Поскольку обучение RBF-сети (рис.1) заключается в вычислении матрицы весовых коэффициентов $W^r = F^{-1}Q$, где F - матрица значений выходов скрытого слоя, $Q = (y_{ij})_{i=1}^{k_1}{}_{j=1}^m$ - значения из матрицы A_1 , а также то, что функционирование сети зависит от вектора $\vec{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_{k_1})$, получим уточненную задачу:

$$\frac{1}{mk_2} \sum_{j=1}^m \sum_{i=1}^{k_2} (y_{ij} - d_{ij}(W^T, \vec{\sigma}))^2 \rightarrow \min. \quad (2)$$

Решению задачи (2) предшествует определение области Ω , в которой изменяются значения вектора $\vec{\sigma}$. Это необходимо для адекватного использования генетического алгоритма, поскольку задача оптимизации (2) может быть адекватно решена только в компактном пространстве. При этом, заметим, что чем меньшим оно будет, тем более эффективным будет процесс решения и точнее результат.

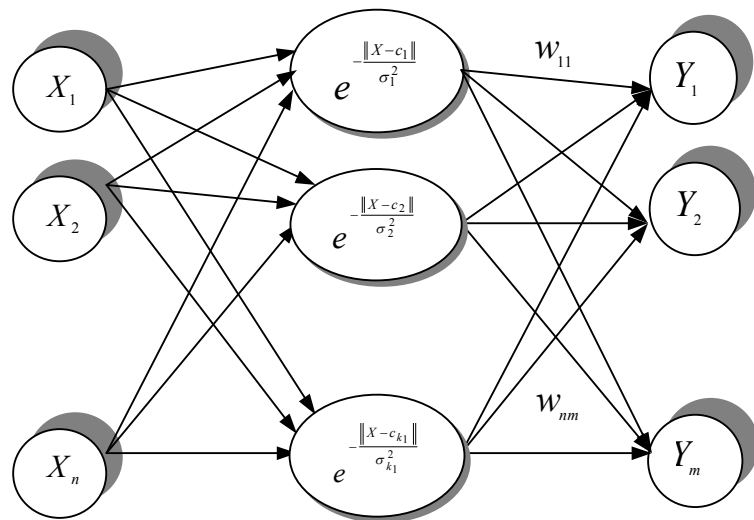


Рис.1. Структура RBF-сети

Определение области изменения “ширин окон” активационных функций

Норма, которая присутствует в активационной функции, вычисляется следующим образом:

$$\|X_l - c_k\| = \sum_{i=1}^n (x_{li} - c_{ki})^2. \quad (3)$$

Предположим, что исходные данные нормированы и их значения находятся на отрезке $[0;1]$. Адекватное функционирование сети возможно и в противном случае, но является более трудоемким при определении параметров и требует отдельного рассмотрения. “Центры окон” активационных функций c_k совпадают с обучающими образами из матрицы A_1 . Тогда норма (3) будет не меньшей нуля и

не превышающей n . Значения функции $e^{-\frac{\|x-c\|}{\sigma^2}}$, соответственно, находятся в отрезке $\left[0; e^{-\frac{n}{\sigma^2}}\right]$. Учитывая то, что при $x \rightarrow \infty$ $e^{-x^2} \rightarrow 0$, а также опасность паралича сети, необходимо требовать выполнения неравенства

$$0,1 < e^{-\frac{n}{\sigma^2}} \leq 1. \quad (4)$$

Тогда получим $\ln 0,1 < -\frac{n}{\sigma^2} \leq 0$, или $-\ln 0,1 > \frac{n}{\sigma^2} \geq 0$. Отсюда

$-\frac{\ln 0,1}{n} > \frac{1}{\sigma^2} \geq 0$, или $\sigma^2 > -\frac{n}{\ln 0,1}$. Считая, без ограничения общности, что

$\sigma > 0$, получим $\sigma > \sqrt{-\frac{n}{\ln 0,1}}$. Исключая возможность “перескакивания” сетью

через оптимум, будем считать, что $e^{-\frac{n}{\sigma^2}} < 0,9$. Тогда, окончательно получим оценку для области изменения значений σ :

$$\sigma \in \left(\sqrt{-\frac{n}{\ln 0,1}}; \sqrt{-\frac{n}{\ln 0,9}} \right). \quad (5)$$

Например, при количестве обучающих образов равном 10, получим такую оценку: $\sigma \in (2,08; 9,74)$.

Эволюционная технология оптимизации функционирования RBF-сети

Оптимизируя работу RBF-сети, мы использовали следующие процедуры: кроссовер однотоочный (ОТ) и двухтоочный (ДТ); выбор родителей пропорциональный (ПР), методом панмиксии (ПА), селективный (СЕ), методом инбридинга (ИН) и аутбридинга (АУ); отбор потомков обычный (ОБ), элитный (ЭЛ), с вытеснением (ВТ).

Композиционный алгоритм работы ГА вместе с RBF-сетью требует предварительной подготовки данных. Простейший ее этап заключается в нормировании исходных данных. Если в качестве формулы преобразования используется

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}},$$

то в данные для нормирования необходимо включать и элементы матриц A_1 и A_2 , также элементы данных, содержащихся в векторе \vec{X} , для которых необходимо рассчитывать прогнозные значения \vec{Y} . Это необходимо делать для того, чтобы и прогнозные данные попадали в единичный гиперкуб.

Размерность вектора $\vec{\sigma}$, как и количество нейронов скрытого слоя, совпадает с количеством тренировочных образов k_1 . Каждая его компонента принадлежит области (5) и применения ГА заключается в получении таких значений элементов $\vec{\sigma}$, при которых получаем решение (2). В качестве функции приспособленности (fitness function) будем использовать среднеквадратическую ошибку на контрольных образах (элементах матрицы A_2), обученной на элементах матрицы A_1 RBF-сети.

Заметим, что на одном шаге ГА кроссовер можно проводить как для одного элемента вектора $\vec{\sigma}$, так и для всех. Алгоритм пропорционального выбора [13] заключается в том, что родители выбираются пропорционально значениям их функции приспособленности; панмиксия не предполагает никаких ограничений на выбор; селективный выбор позволяет включать в пару только особей, приспособленность которых не ниже средней по популяции, при равных шансах всех; в результате инбридинга первая особь пары является случайной, а второй выбирается максимально близкая к ней по расстоянию Хэмминга; в результате аутбридинга выбираются максимально далекие по тому же критерию особи. Обычный выбор потомка и включение его в новую популяцию осуществляется с вероятностью 0,5 из двух; элитный отбор предполагает включение в новую популяцию лучших родителей и потомков, а при отборе с вытеснением в новой популяции будут только различные особи.

Алгоритм оптимизации работы RBF-сети с помощью ГА имеет следующий вид:

Шаг 1. Нормируем начальные данные.

Шаг 2. Формируем обучающий и контрольный массив.

Шаг 3. Определяем границы изменения оптимизируемых параметров.

Шаг 4. Задаем параметры генетического алгоритма.

Шаг 5. Вычисляем количество элементов в генеральной совокупности (зависит от предполагаемой точности результата *eps*).

Шаг 6. Определяем количество бинарных элементов в генотипе.

Шаг 7. Формируем случайным образом начальную популяцию в двоичном, вещественном и целом эквиваленте.

Шаг 8. Положим $S_Fit = 0$ и будем интерпретировать значение этой переменной как среднюю приспособленность на предыдущем шаге.

Шаг 9. Для всех элементов начальной популяции $\vec{\sigma}$ по очереди обучаем RBF-сеть и находим среднюю квадратическую ошибку на контрольных образах $Fit(i)$.

Шаг 10. Находим значение средней приспособленности S_Pop . Если $|S_Fit - S_Pop| < eps$, то переход на шаг 20.

Шаг 11. Находим максимальное Max_Fit и минимальное Min_Fit значения приспособленности в популяции. Если $|Max_Fit - Min_Fit| < eps$, то переход на шаг 20.

Шаг 12. Нормируем значения $Fit(i)$ и формируем соответствующий массив вероятностей.

Замечание. Шаги 13-16 выполняются k_1 раз.

Шаг 13. Выбор родителей (по одной из схем ПР, ПА, СЕ, ИН, АУ).

Шаг 14. Осуществление кроссовера (по одной из схем ОТ, ДТ).

Шаг 15. Осуществление мутации с заданной вероятностью.

Шаг 16. Формирование множества кандидатов в новую популяцию.

Шаг 17. Отбор особей в новую популяцию по одной из схем ОБ, ЭЛ, ВТ.

Шаг 18. Положим $S_Fit = S_Pop$.

Шаг 19. Переход на шаг 9.

Шаг 20. Вычисляем прогнозные значения для полученных массивов весовых коэффициентов W и параметров $\vec{\sigma}$.

Шаг 21. Конец.

Анализ результатов

Этот алгоритм был реализован как составная часть программной системы идентификации и прогнозирования “Alternative S”. Проведенное тестирование дало возможность сделать определенные выводы. В качестве функции, которую необходимо аппроксимировать, была взята функция $y = x^2$. Обучающие образы приведены в табл.1.

Таблица 1.

Исходные данные для аппроксимации

x	1	3	5	7	9	11	13	15	17	19
y	1	9	25	49	82	120	169	223	290	361

Точки прогнозирования $x_1 = 4$, $x_2 = 20$. Отношение количества тренировочных образов к контрольным было взято как 4:1, т.е. в контрольную матрицу попали два образа, в тренировочную - восемь. Несмотря на то, что в исходных данных присутствовали шумовые эффекты, сеть достаточно эффективно их обошла. Так, средние квадратические ошибки на контрольных образах составляли порядка $10^{-8} - 10^{-6}$. Результаты работы программной системы представлены в таблице 2.

Анализ результатов показывает, что среднее количество итераций составило 13,5. Прогнозные значения: для $x = 4$, $y = 16,043064$; для $x = 20$, $y = 399,962027$. Абсолютные ошибки составили $\varepsilon_1 = 0,043064$ и $\varepsilon_2 = 0,037973$, что свидетельствует о высоком качестве прогнозирования. Тестирование проводилось при 14 представителях в популяции и заканчивалось, когда абсолютное значение разности средней приспособленности в настоящей и предыдущей популяции или абсолютное значение минимальной и максимальной приспособленности в популяции было меньше 10^{-8} . Вероятность мутации составляла 0,02. Не-

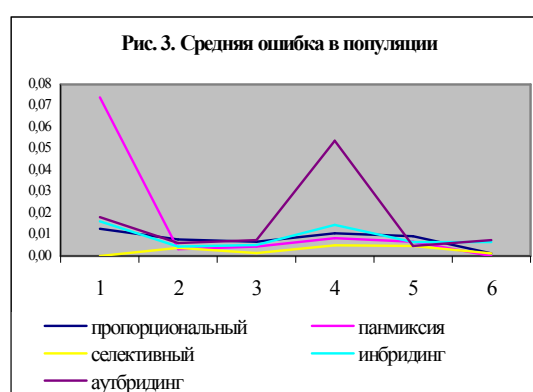
большое среднее количество итераций свидетельствует о том, что сеть достаточно хорошо отслеживает закономерности в данных. Наибольшее количество итераций соответствует варианту, когда в роли механизма формирования новой популяции применяется отбор с вытеснением. Этому же варианту соответствует большинство случаев, когда прогнозирование осуществляется с некоторым положительным смещением.

Таблица 2

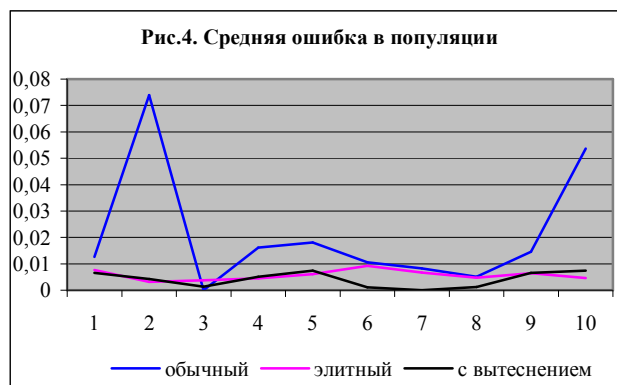
Результаты оптимизации RBF-сети и прогноза

Кроссовер	Выбор родителей	Выбор потомков	Прогноз для 4	Прогноз для 20	Кол-во итераций	Относит. ошибка, %
ОТ	ПР	ОБ	15,999997	399,94949	4	0,01264625
ОТ	ПР	ЭЛ	16,000139	399,97297	6	0,00762625
ОТ	ПР	ВТ	15,999996	399,97372	6	0,00659500
ОТ	ПН	ОБ	16,000180	399,70870	3	0,07395000
ОТ	ПН	ЭЛ	15,999997	399,98721	8	0,00321625
ОТ	ПН	ВТ	16,000008	399,98298	45	0,00430500
ОТ	СЕ	ОБ	16,999970	399,96334	4	6,2589
ОТ	СЕ	ЭЛ	16,000013	399,98551	4	0,00370375
ОТ	СЕ	ВТ	16,000014	399,99513	10	0,00130500
ОТ	ИН	ОБ	15,999990	399,93582	5	0,01610750
ОТ	ИН	ЭЛ	15,999993	399,98240	6	0,00444375
ОТ	ИН	ВТ	16,000010	399,97989	60	0,00509000
ОТ	АУ	ОБ	15,999862	399,93102	6	0,01810750
ОТ	АУ	ЭЛ	16,000011	399,97599	5	0,00607125
ОТ	АУ	ВТ	15,999983	399,97079	9	0,00740875
ДТ	ПР	ОБ	16,000043	399,95899	4	0,01052125
ДТ	ПР	ЭЛ	16,000025	399,96369	4	0,00923375
ДТ	ПР	ВТ	15,999987	399,99578	13	0,00113625
ДТ	ПН	ОБ	15,999989	399,96722	3	0,00826375
ДТ	ПН	ЭЛ	15,999993	399,97364	12	0,00663375
ДТ	ПН	ВТ	16,291981	400,06517	21	1,8411
ДТ	СЕ	ОБ	15,999875	399,98256	3	0,00514125
ДТ	СЕ	ЭЛ	16,000000	399,98130	5	0,00467500
ДТ	СЕ	ВТ	15,999990	400,00477	13	0,00125500
ДТ	ИН	ОБ	15,999813	399,94657	5	0,01452625
ДТ	ИН	ЭЛ	16,000028	399,97497	20	0,00643250
ДТ	ИН	ВТ	16,000021	399,97430	50	0,00655625
ДТ	АУ	ОБ	15,999955	399,78644	6	0,05367125
ДТ	АУ	ЭЛ	15,999987	400,01832	5	0,00466125
ДТ	АУ	ВТ	16,000064	399,97213	61	0,00736750

Анализ данных позволил установить ряд любопытных фактов. Так, средняя ошибка в популяции (рис.2) не зависит от выбора варианта кроссовера. Для одно-



точечного кроссовера наибольшая ошибка соответствует селективному варианту выбора родителей и общему принципу формирования новой популяции. Для двухточечного кроссовера – это пропорциональный выбор родителей и отбор в новую популяцию с вытеснением. График зависимости средней ошибки в популяции от варианта выбора родителей (рис.3) указывает на то, что наилучшим вариантом, дающим наибольшую скорость сходимости, является селективный метод, самым медленным является аутбридинг. При формировании новой популяции наиболее эффективным оказывается выбор с вытеснением, незначительно ему уступает элитный метод выбора потомков (рис.4). Таким образом, оптимальной комбинацией является композиция селективного метода выбора родителей с формированием новой популяции методом вытеснения. Результаты табл. 2 это подтверждают.



Прогнозирование было проведено и для других значений (табл. 3). Величины абсолютных ошибок имеют тенденцию к росту как квадратичная зависимость, однако, их величина при прогнозировании в области, которая значительно удалена от области обучения RBF-сети, свидетельствует о точности и эффективности метода.

Заключение

Если прогнозные значения попадают на границу гиперкуба, а остальные значения сосредоточены в некоторой узкой его области, то точность прогноза будет убывать по мере удаления этой области от соответствующей границы гиперкуба. Эффективность алгоритма будет еще выше, если для бинарного кодирования использовать не простое двоичное представление числа, а его код Грея, поскольку он обладает свойством “непрерывности”, т.е. если числа в десятичной записи отличаются на единицу, то и в записи их кодов Грея будут различными значениями только в одном разряде.

Композиция прямого метода получения весовых коэффициентов RBF-сети и итерационного генетического алгоритма дает возможность получения оптимальных значений параметров сети, что, в свою очередь, позволяет решить задачу идентификации целевой функции и осуществить прогнозирование с минимальными ошибками. В отличие от других ИНС, RBF-сеть в композиции с ГА требует значительно меньших вычислительных ресурсов. И одним из главных преимуществ использования RBF-сети вместе с ГА при установлении аппроксимационных зависимостей является возможность идентификации вектор-функций.

Таблица 3

X	Y	Y [^]	Y-Y [^]
4	16	16,00207	0,002074
20	400	400,0003	0,00027
25	625	624,9972	0,0028
30	900	899,9784	0,02163
35	1225	1224,889	0,1106
40	1600	1599,574	0,4263
45	2025	2023,664	1,3362
50	2500	2496,422	3,5784
55	3025	3016,524	8,4756

Литература

1. Васильев В.И., Шевченко А.И. Комбинированный алгоритм оптимальной сложности // Искусственный интеллект. – 2002. - №3. – С. 504-509.
2. Бриллинджер Д. Временные ряды. Обработка данных и теория. – М.: Мир, 1980. – 535с.
3. Бидюк П.И., Зворыгина Т.Ф. Структурный анализ методик построения регрессионных моделей по временным рядам наблюдений // УсиМ. – 2003. – №2. – С.93-99.
4. Крислов В.А., Побережник С.М. Аппроксимация сложных зависимостей структурно-гибкими полиномиальными и гармоническими рядами // УсиМ. – 2003. - №2. – С. 80-86.
5. Ляшенко І.М., Снитюк О.І. Синтетична оптимізація структури виробництва на прикладі агропромислового комплексу Черкаської області // Економічна кібернетика. – 2002. - №3-4. – С. 67-73.
6. Шарапов В.М., Снитюк В.Е. Биокрибернетический метод определения оптимума целевой функции в условиях неопределенности // Искусственный интеллект. – 2002. - №4. – С. 123-128.
7. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного. Докл. АН СССР, 1957. Т. 114, №. 5. С. 953-956.
8. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. /Вороновский Г.К., Махотило К.В., Петрашев С.Н., Сергеев С.Н. – Харьков: Основа. - 1997. – 112с.
9. Ротштейн А.П., Митюшкин Ю.И. Извлечение нечетких баз знаний из экспериментальных данных с помощью генетических алгоритмов // Кибернетика и системный анализ. – 2001. - №4. – С.45-53.
10. Курейчик В.В., Эволюционные методы принятия решений с синергетическими и гомеостатическими принципами управления // Таганрог: Перспективные информационные технологии и интеллектуальные системы. – 2002. - №1. С. 6-10.
11. Божич В.И., Лебедев О.Б., Шницер Ю.Л. Разработка генетического алгоритма обучения нейронных сетей // Таганрог: Перспективные информационные технологии и интеллектуальные системы. – 2002. - №1. С. 21-24.
12. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. – К.: Техніка, 1975. - 312 с.
13. Батищев Д.И., Исаев С.А. Оптимизация многоэкстремальных функций с помощью генетических алгоритмов / Мужвуз. сборник, ВГТУ, Воронеж, 1997.