

ЭВОЛЮЦИОННЫЙ МЕТОД СТРУКТУРНО-ПАРАМЕТРИЧЕСКОГО СИНТЕЗА НЕЙРОННЫХ СЕТЕЙ

Введение

При решении задач аппроксимации зависимостей, а также их приложений для диагностики и автоматической классификации, прогнозирования и управления широкое распространение получили применение искусственных нейронных сетей [1]. Существенное влияние на эффективность использования построенной нейромодели на практике оказывает выбранная структура нейросети. Для структурного синтеза нейронных сетей используются методы последовательного добавления и последовательного удаления нейронов [2].

Весьма существенным недостатком существующих методов структурного синтеза является неоднозначность их результатов, вызванная двумя факторами, связанными с необходимостью обучения нейросети для оценивания ее структуры:

– случайный выбор начальных значений весовых коэффициентов нейромодели при параметрическом синтезе (обучении сети), что приводит к ситуациям, когда сетям с одинаковой архитектурой соответствуют разные значения критерия оценивания;

– получение различных результатов обучения, достигаемых при использовании разных методов настройки весовых коэффициентов, вследствие чего при выборе неоптимального для решаемой задачи метода параметрического синтеза оценивание структур нейромоделей в процессе структурного синтеза будет необъективным. Это, в свою очередь, приведет к неоптимальности найденной структуры нейронной сети.

Для устранения такой неоднозначности при оценивании хромосом в процессе структурного синтеза необходимо выполнять многократный (с применением различных начальных значений весовых коэффициентов и с использованием различных методов обучения) параметрический синтез нейросети, соответствующей оцениваемой структуре сети. Однако такой подход значительно увеличивает время, необходимое для выполнения структурного синтеза.

Другим, более эффективным, методом преодоления указанного недостатка является одновременный синтез структуры нейросетевой модели и поиск оптимальных значений параметров такой сети – структурно-параметрический синтез. Однако, как правило, существующие методы структурно-параметрического синтеза [2–4] подразумевают выбор структуры по заранее определенным правилам, которые при решении практических задач могут оказаться мало эффективными.

Для выполнения структурно-параметрического синтеза нейромодели эффективно могут использоваться методы эволюционной оптимизации [2, 5–7], в которых поиск оптимальной структуры сети и значений ее параметров является случайно направленным, что позволяет исследовать пространство поиска в различных областях. При таком подходе целесообразно каждую хромосому (решение) составить из нескольких частей, содержащих информацию о структуре сети и значениях ее весовых коэффициентов.

Однако правила вычисления значений генов новых хромосом в существующих операторах скрещивания и мутации [5–7], применяемых в традиционных методах эволюционной оптимизации, не учитывают то, какую именно информацию о нейросети (значения весовых коэффициентов, значения смещений, тип функции активации) содержат гены хромосомы, и являются одинаковыми для всех генов, что понижает эффективность эволюционного поиска для структурно-параметрического синтеза нейромодели.

Целью настоящей работы являлось создание эволюционного метода структурно-параметрического синтеза, позволяющего строить нейросетевые модели с незначительным количеством синаптических соединений, и свободных от указанных недостатков эволюционных операторов скрещивания и мутации для генерации новых хромосом при построении нейромодели.

1. Постановка задачи

Пусть задано максимально допустимое количество нейронов A в нейросети, используемой для аппроксимации зависимости по выборке исходных данных $\langle X, Y \rangle$, где $X = \{X_i\}$ – набор значений признаков, характеризующих рассматриваемый объект или процесс; $Y = \{y_p\}$ – массив значений выходного параметра в заданной выборке; $X_i = \{x_{ip}\}$ – i -ый признак в выборке, $i = 1, 2, \dots, L$; x_{ip} – значение i -го признака для p -го экземпляра выборки, $p = 1, 2, \dots, m$; y_p – значения прогнозируемого параметра для p -го экземпляра; L – общее количество признаков в исходном наборе; m – количество экземпляров выборки.

Тогда задача структурно-параметрического синтеза нейромодели заключается в поиске модели вида $HC = HC(S, W, B)$, для которой $\xi(HC, X, Y) \rightarrow \min$, где $S = S(L, A)$ – матрица, определяющая наличие синаптических связей между элементами сети (входными признаками, нейронами); $W = W(S)$ – матрица весовых коэффициентов, соответствующих связям, присутствующим в сети; $B = B(S)$ – вектор смещений нейронов сети; $\xi(HC, X, Y)$ – критерий, определяющий эффективность использования нейросетевой модели HC для аппроксимации зависимости между набором входных параметров X и соответствующем ему вектору значений выходного параметра Y , как правило, в качестве критерия оптимальности нейромодели используется среднеквадратическая ошибка:

$$\xi = \sum_{p=1}^m (y_p - y(HC, X_p))^2,$$

где X_p – набор значений признаков для p -го экземпляра; $y(\text{НС}, X_p)$ – значение выхода нейромодели НС, вычисленное для набора значений X_p .

2. Структурно-параметрический синтез нейросетевых моделей на основе методов эволюционной оптимизации

При структурно-параметрическом синтезе нейромоделей выбираются оптимальные значения весовых коэффициентов нейросети (входного слоя и скрытых слоев), смещений нейронов, а также функции активации нейронов. Поэтому предлагается решение (хромосому) составить из нескольких частей (рис. 1): в первой части хранить информацию о значениях весов нейронов входного слоя, во второй – значения весовых коэффициентов нейронов скрытых слоев, в третьей – значения смещений нейронов, в четвертой – функции активации для каждого нейрона сети.

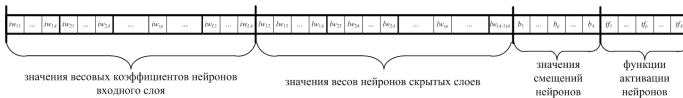


Рис. 1 – Рисунок 1 – Представление хромосомы при структурно-параметрическом синтезе нейросетей

На рис. 1 используются следующие обозначения: $iw_{\nu\rho}$ – значение весового коэффициента связи от ν -го признака к ρ -му нейрону; $lw_{\nu\rho}$ – значение весового коэффициента связи от ν -го нейрона к ρ -му нейрону; b_ρ – значение смещения ρ -го нейрона; tf_ρ – функция активации ρ -го нейрона.

Как видно из рис. 1, вторая часть хромосомы, содержащая информацию о значениях весовых коэффициентов нейронов скрытых слоев, кодируется с помощью способа, аналогичного параметрическому представлению хромосомы при структурном синтезе [2].

Размер хромосомы K определяется по формуле:

$$K = K_1 + K_2 + K_3 + K_4 = L \cdot A + \frac{A(A-1)}{2} + A + A = L \cdot A + \frac{A(A+3)}{2}, \quad (1)$$

где K_1, K_2, K_3, K_4 – количество генов первой, второй, третьей и четвертой частей хромосомы, соответственно.

С целью поиска нейромоделей с минимальным количеством синаптических соединений в предлагаемом методе структурно-параметрического синтеза разработаны специальные операторы скрещивания и мутации, стремящиеся уменьшить количество связей в сети.

Для повышения эффективности эволюционной оптимизации при структурно-параметрическом синтезе нейромоделей и сокращения времени поиска предлагается на этапе инициализации хромосом, соответствующих нейросетевым моделям, выбор начальных значений весовых коэффициентов и смещений осуществлять таким образом, чтобы равномерно распределить активную область определения функции активации каждого нейрона по пространству входных переменных, учитывая при этом индивидуальную информативность признаков.

Разработанный эволюционный метод структурно-параметрического синтеза нейросетевых моделей предлагается выполнять как следующую последовательность шагов.

Шаг 1. Сформировать начальное поколение хромосом, содержащих информацию о структуре сети и ее параметрах.

Шаг 2. Выполнить оценивание хромосом текущей популяции, вычленив значение фитнес-функции, учитывающее ошибку нейросети и сложность ее архитектуры (количество межнейронных связей, нейронов, слоев). Для расчета ошибки сети, соответствующей оцениваемой хромосоме H_j , целесообразно использовать формулу:

$$E(H_j) = \sum_{p=1}^m (y_p - y(H_j, X_p))^2, \quad (2)$$

где $y(H_j; X_p) = y(\text{НС}; X_p) = y_{Ap}$ – значение выхода нейромодели НС, построенной на основе хромосомы H_j , вычисленное для набора значений X_p :

$$y_{Ap} = tf_A \left(b_A + \sum_{c=1}^L iw_{cA} x_{cp} + \sum_{c=1}^{A-1} lw_{cA} y_{cp} \right), \quad (3)$$

где y_{Ap} – значение выхода A -го нейрона сети для p -го экземпляра; y_{cp} – значение выхода c -го нейрона сети для p -го экземпляра, рассчитывается по формуле, аналогичной для расчета значения y_{Ap} .

Шаг 3. Проверить критерии окончания поиска (достижение приемлемого значения ошибки синтезируемой модели, превышение максимально допустимого количества итераций, превышение допустимого времени функционирования метода). В случае удовлетворения хотя бы одного из таких критериев, выполнить переход к шагу 8.

Шаг 4. Выбрать наиболее приспособленные хромосомы для выполнения над ними эволюционных операторов скрещивания и мутации.

Шаг 5. Выполнить оператор скрещивания, уменьшающий количество синаптических соединений в сети.

Для частей хромосом, содержащих информацию о значениях весовых коэффициентах нейронов входного и скрытых слоев, значение i -го гена потомков предлагается определять по формулам:

$$h_{i\pi 1} = \begin{cases} 0, & \text{если } h_{i1} \cdot h_{i2} < 0; \\ kh_{i1} + (1 - k)h_{i2}, & \text{в противном случае,} \end{cases} \quad (4)$$

$$h_{i\pi 2} = \begin{cases} 0, & \text{если } h_{i1} \cdot h_{i2} > 0; \\ (1 - k)h_{i1} + kh_{i2}, & \text{в противном случае,} \end{cases}$$

где $h_{i\pi 1}$ и $h_{i\pi 2}$ – значения i -ых генов первого и второго потомков, соответственно; h_{i1} и h_{i2} – значения i -ых генов первого и второго родителей, соответственно; k – коэффициент, задаваемый пользователем, $k \in (0; 1)$.

Значения генов, соответствующих смещениям нейронов, определить по формулам: $h_{i\pi 1} = kh_{i1} + (1 - k)h_{i2}$ и $h_{i\pi 2} = kh_{i2} + (1 - k)h_{i1}$.

Значения генов, определяющих функцию активации нейрона, предлагается определять следующим образом:

$$\begin{aligned} h_{in1} &= \begin{cases} h_{i1}, & \text{если } h_{i1} = h_{i2} \text{ или } r > 0,5; \\ \text{rand}[\text{TF}], & \text{в противном случае,} \end{cases} \\ h_{in2} &= \begin{cases} h_{i2}, & \text{если } h_{i1} = h_{i2} \text{ или } r \leq 0,5; \\ \text{rand}[\text{TF}], & \text{в противном случае,} \end{cases} \end{aligned} \quad (5)$$

где r – случайно сгенерированное число в интервале (0; 1); $\text{rand}[\text{TF}]$ – случайно выбранный элемент множества TF , содержащего функции активации, используемые для построения нейросети.

Шаг 6. Выполнить оператор точечной мутации.

В случае, если для мутации выбран ген хромосомы, содержащий информацию о значениях весовых коэффициентах нейронов сети, тогда новое значение i -го гена предлагается вычислять по формуле:

$$h_{in} = \begin{cases} 0, & \text{если } |r| < 0.5h_i; \\ r, & \text{в противном случае,} \end{cases} \quad (6)$$

где h_i и h_{in} – значения i -го гена до и после мутации, соответственно; $r = \text{rand}[-h_i; h_i]$ – случайно сгенерированное число в интервале $[-h_i; h_i]$.

Если для мутации выбран ген хромосомы, соответствующий смещению нейрона, тогда значение i -го гена после мутации h_{in} определяется следующим образом: $h_{in} = \text{rand}[h_{i,\min}; h_{i,\max}]$, где $h_{i,\min}$ и $h_{i,\max}$ – минимальное и максимальное значения i -го гена в популяции.

В случае выбора для мутации гена хромосомы, содержащего информацию о функции активации нейрона, новое значение гена предлагается определять по формуле: $h_{in} = \text{rand}[\text{TF}]$.

Шаг 7. Создать новое поколение из полученных на предыдущем шаге хромосом-потомков и наиболее приспособленных хромосом текущего поколения. Выполнить переход к шагу 2.

Шаг 8. Останов.

Разработанный эволюционный метод структурно-параметрического синтеза нейросетей позволяет построить нейросетевую модель, используя при этом операторы скрещивания и мутации, уменьшающие количество синаптических соединений в сети.

3. Эксперименты и результаты по диагностике заболевания пиелонефритом

Методика экстренной помощи пациенту при заболеваниях пиелонефритом определяется формой заболевания. Поэтому во врачебной практике особо важное место занимают своевременная диагностика пиелонефрита и распознавание его формы [8].

Основным недостатком существующих методов дифференциальной диагностики пиелонефритов является необходимость определения большого количества лабораторных показателей, получение значений которых требует значительных затрат времени, является болезненным и дорогостоящим, что приводит к промедлению в оказании помощи больно-

му. Для автоматизации диагностики пиелонефритов целесообразно построить модели, позволяющие выявлять болезнь по данным несложного лабораторного исследования, например анализа крови. В настоящей работе для распознавания острого серозного пиелонефрита, острого гнойного пиелонефрита и обострения хронического пиелонефрита в качестве диагностических признаков используются следующие параметры общего анализа крови, предложенные в [9]: x_1 – эритроциты, Т/л; x_2 – гемоглобин, г/л; x_3 – лейкоциты, Г/л; x_4 – эозинофилы, %; x_5 – палочкоядерные нейтрофилы, %; x_6 – сегментоядерные нейтрофилы, %; x_7 – лимфоциты, %; x_8 – моноциты, %; x_9 – скорость оседания эритроцитов (СОЭ), мм/час.

Для построения диагностических моделей зависимостей различных форм пиелонефрита от параметров общего анализа крови использовался предложенный эволюционный метод структурно-параметрического синтеза нейросетевых моделей, который был программно реализован на языке пакета Matlab. Начальные значения параметров эволюционного поиска устанавливались следующими: количество особей в популяции $N = 100$, вероятность скрещивания $p_{скр} = 0.8$, максимальное количество итераций $T = 100$, количество элитных особей $N_e = 2$. Максимально допустимое количество нейронов в сети – 10.

В связи с вероятностным характером работы методов эволюционной оптимизации построение моделей выполнялось 100 раз для получения более объективных результатов.

В результате структурно-параметрического синтеза получили три нейросетевые модели зависимостей различных форм пиелонефритов от параметров общего анализа крови, среднеквадратические ошибки синтезированных моделей составили 0.035, 0.029 и 0.033 для диагностики острого серозного пиелонефрита, острого гнойного пиелонефрита и обострения хронического пиелонефрита, соответственно.

В табл. 1 приведена построенная нейромодель для диагностики острого серозного пиелонефрита, а на рис. 2 – графическое представление синтезированной нейронной сети.

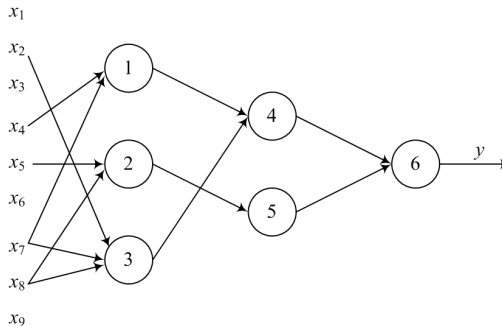


Рис. 2 – Графическое представление синтезированной нейронной сети

Построенная нейросетевая модель

Номер слоя, μ	Номер нейрона, ρ	Функция активации нейрона	Значение смещения	Связи (соединения)	
				узел, от которого идет соединение	значение весового коэффициента
1	1	тангенциальная сигмоидная	24,4132	признак $X_4 iw_{41}$	0,045477
				признак $X_7 iw_{71}$	0,041279
	2	тангенциальная сигмоидная	-88,3075	признак $X_5 iw_{52}$	0,041267
				признак $X_8 iw_{82}$	0,036844
	3	логистическая сигмоидная	28,8537	признак $X_{2i} iw_{23}$	0,016148
				признак $X_7 iw_{73}$	-0,031317
признак $X_8 iw_{83}$	-2,7601				
2	4	логистическая сигмоидная	59,1486	нейрон 1 lw_{14}	-0,021307
				нейрон 3 lw_{34}	-0,011642
	5	логистическая сигмоидная	-18,3753	нейрон 2 lw_{25}	-0,029741
3	6	пороговая	0,325	нейрон 4 lw_{46}	-0,026373
				нейрон 5 lw_{56}	-0,017855

Как видно, модель, построенная с помощью предложенного метода, не зависит от значений входных переменных x_1, x_6, x_9 . Сравнение достигнутых результатов с результатами, полученными в [9] показывает, что модели, синтезированные путем применения разработанного эволюционного метода структурно-параметрического синтеза нейромоделей, являются более простыми и интерпретабельными, поскольку имеют меньшее количество межнейронных связей, обеспечивая при этом приемлемую точность классификации.

Выводы

С целью построения простых и интерпретабельных нейромоделей разработан метод эволюционной оптимизации для структурно-параметрического синтеза нейронных сетей, что составляет *научную новизну* исследования. Предложенный метод позволяет выбирать оптимальную структуру нейронной сети, настраивая при этом значения весовых коэффициентов и смещений и не накладывая ограничений на вид функции активации нейронов, использует специальные операторы скрещивания и мутации, уменьшающие количество связей в сети, что позволяет синтезировать нейромодели с незначительным количеством синоптических соединений, обеспечивающих при этом высокую точность аппроксимации реальных зависимостей.

Практическая ценность результатов работы состоит в том, что разработано программное обеспечение, реализующее предложенный эволюционный метод структурно-параметрического синтеза нейромоделей, а также решена задача построения диагностических моделей зависимостей различных форм пиелонефрита от параметров общего анализа крови.

Работа выполнена в рамках госбюджетной темы “Научно-методические основы и математическое обеспечение для автоматизации и моделирования процессов управления и поддержки принятия решений на основе процедур распознавания и эволюционной оптимизации в нейросетевом и нечеткологическом базисах” (гос. регистрации 0106U008621).

Литература

1. Круглов В.В., Борисов В.В. Искусственные нейронные сети: Теория и практика. – М.: Горячая линия-Телеком, 2001. – 382 с.
2. Yao X. Evolving Artificial Neural Network // Proceedings of the IEEE. – 1999. – 9(87). – P. 1423–1447.
3. Интеллектуальные средства диагностики и прогнозирования надежности авиадвигателей: Монография / В.И. Дубровин, С.А. Субботин, А.В. Богуслаев, В.К. Яценко. – Запорожье: ОАО “Мотор-Сич”, 2003. – 279 с.
4. Каллан Р. Основные концепции нейронных сетей. – М.: Издательский дом “Вильямс”, 2001. – 287 с.
5. Haupt R., Haupt S. Practical Genetic Algorithms. – New Jersey: John Wiley & Sons, 2004. – 261 p.
6. Gen M., Cheng R. Genetic algorithms and engineering design. – New Jersey: John Wiley & Sons, 1997. – 352 p.
7. The Practical Handbook of Genetic Algorithms. Volume II. New Frontiers / Ed. L.D. Chambers. – Florida: CRC Press, 2000. – 421 p.
8. Тиктинский О.Л., Калинина С.Н. Пиелонефриты. – СПб.: МАЛО Медиапресс, 1996. – 238 с.
9. Дубровин В.И., Колесник Н.В., Субботин С.А. Нейросетевая дифференциальная диагностика острых пиелонефритов // Сборник трудов научно-технической конференции “Нейроинформатика–2003”, 29–31 января 2003 г. – М.: МИФИ, 2003. – С. 86–93.

Получено 02.04.2008