# Аппроксимация непрерывных функций через синтез нейронных сетей с минимальной конфигурацией

Н. А. Игнатьев
*Национальный университет Узбекистана, Ташкент*
e-mail: n_ignatev@rambler.ru

Рассматривается решение задачи численной аппроксимации непрерывных функций в пространстве количественных и разнотипных признаков через синтез нейронных сетей с минимальной конфигурацией.

*Ключевые слова*: нейронные сети, численная аппроксимация, разнотипные признаки.

## Введение

Аппроксимация непрерывных функций от многих переменных с помощью искусственных нейронных сетей (НС) с любой наперед заданной точностью является базовой задачей для многих прикладных исследований. Методы обучения НС решению этой задачи еще далеки от совершенства. В частности, это проявляется в отсутствии строго формализованных процедур адаптации конфигурации нейронной сети под сложность восстанавливаемых функций. Как правило, выбор конфигурации производится эвристическим путем и определяется интуицией и опытом исследователя.

Предлагаемый структурный и алгоритмический синтез однослойных НС основан на оптимизации через построение минимального покрытия обучающей выборки объектами-эталонами в пространстве количественных и разнотипных признаков. Решается задача локально-оптимального покрытия обучающей выборки объектами-эталонами при заданной величине максимального уклонения $\varepsilon$ табличного значения функции от аппроксимируемого.

Аппроксимация функций в окрестностях объектов-эталонов (локальных областях) осуществляется с помощью радиально-базисных функций активации. В отличие от известных методов аппроксимации (например, [1, 2]) базовые элементы сети относительно равномерно (в смысле задаваемой точности) распределяются по локальным областям признакового пространства. Рассматриваются вопросы о способности обобщения НС решаемой задачи в зависимости от точности аппроксимации непрерывной случайной функции и объема обучающей выборки.

## 1. Аппроксимация функций в пространстве количественных признаков

Считается что табличное представление (по строкам) функции задается с помощью $n+1$ количественных признаков. Один из признаков строки объявляется целевым, остальные — зависимыми от него. Множество значений целевого признака обозначим через $Y$, зависимых — через $X_1, ..., X_n$ и будем считать, что между ними существует некоторая явно не заданная функциональная зависимость $y = f(x_1, ..., x_n)$. Далее по тексту при употреблении термина объект подразумевается, что он описывается на множестве зависимых признаков. Требуется синтезировать НС через аппроксимацию вектора значений функции $Y = \{y_i\}_1^m$, каждому элементу которого соответствует объект из множества $E_0 = \{S_1, S_2, ..., S_m\}$.

Аппроксимационные свойства НС для восстановления функциональной зависимости реализуются с помощью радиально-базисных функций так называемыми $RBF$-сетями. Описание множества радиально-базисных функций и их использование в $RBF$-сетях можно найти в [1]. Однослойные НС, использующие эти функции, в данной работе рассматриваются как альтернатива многослойным НС, которые в силу многоэкстремальности решаемой задачи не всегда могут выдать глобальный минимум ошибки. В данной работе при описании процесса выбора локально-оптимального покрытия $E_0$ объектами-эталонами для аппроксимации табличных функций используется радиально-базисная функция $\exp(\alpha z)$.

Пусть объект $S^j \in E_0(S^j = (x_{j1}, ..., x_{jn}))$ является одним из эталонов выборки и значения весов его признаков [3, 4], вычисляются как $w_{jt} = x_{jt}$ и $w_{j0} = -\frac{1}{2}\sum_{t=1}^{n} w_{jt}^2$. Для синтеза НС, вычисляющей значения целевого признака, необходимо:

а) задать значение $\varepsilon$ — максимальной величины уклонения $\mid y - \overline{y} \mid \leq \varepsilon$ значения целевого признака $\overline{y}$, восстанавливаемого с помощью радиально-базисной функции, от табличного $y$;

б) построить локально-оптимальное покрытие $\Pi(\varepsilon)$ множества $E_0$ объектами-эталонами с учетом значения $\varepsilon$.

Значение взвешенной суммы по объекту-эталону $S^j \in E_0$ для произвольного допустимого объекта $S = (a_1, ..., a_n)$ вычисляется как

$$\varphi(S, S^j) = \sum_{i=1}^{n} w_{ji}a_i + w_{j0} \tag{1}$$

и используется для определения по $\max_j \varphi(S, S^j)$ локальной области пространства для аппроксимации функции по объекту (точке) $S$.

По каждому объекту $S_r \in E_0$ из уравнения

$$y_r - \mathrm{sign}(y_r)\exp(\alpha_r\varphi(S_r, S_r)) = 0$$

вычисляется значение параметра $\alpha_r$ радиально-базисной функции $\exp(\alpha z)$.

Для каждого $S_i \in E_0$ строится упорядоченная по мере убывания значений $\varphi(S_j, S_i)$ по (1) последовательность

$$S_{i_1}, S_{i_2}, ..., S_{i_m}, S_i = S_{i_1}. \tag{2}$$

Из (2) выделяется подпоследовательность $\{S_{i_1}, S_{i_2}, ..., S_{i_k}\}, 1 \leq k < m$, для каждого $S_{i_j}, j = \overline{1,k}$, которой выполняется

$$| \; y_i - \operatorname{sign}(y_i) \exp(\alpha_{i_j} \varphi(S_{i_j}, S_i)) \; | \leq \varepsilon$$

$(| \; y_i - \operatorname{sign}(y_i) \exp(\alpha_{i_{k+1}} \varphi(S_{i_{k+1}}, S_i)) \; | > \varepsilon)$, и объект $S_i$ включается в состав множества $T_{i_j}$.

Процесс построения минимального покрытия $\Pi(\varepsilon)$ обучающей выборки производится следующим образом.

1. Выбирается максимальное по числу объектов множество $T_i$. Объект $S_i \in E_0$ включается в состав объектов покрытия $\Pi(\varepsilon)$. По каждому $S_j \in T_i$ производится коррекция состава множеств $\{T_i\}$ как

$$T_u = \begin{cases} \emptyset, & u = j, \\ T_u \backslash S_j, & u \neq j. \end{cases}$$

2. Если $\bigcup\limits_{i=1}^{m} T_i = \emptyset$, то процесс формирования $\Pi(\varepsilon)$ заканчивается, в противном случае происходит переход на 1.

Описанная технология синтеза НС позволяет предотвращать паралич сети, который имеет место при использовании градиентных методов обучения. $RBF$-сети являются универсальными аппроксиматорами и при необременительных ограничениях могут быть использованы для аппроксимации любой непрерывной функции.

Вопрос о соотношении точности (задаваемой значением $\varepsilon$) восстанавливаемой функции, длины обучающей выборки и количества зависимых параметров в [5] исследовался через равномерную сходимость частот появления событий $J_{m\varepsilon} = \{S_i \mid (y_i - \overline{y_i})^2 > \varepsilon\}$ к их вероятностям. Пусть $\Pi_{mn}(\varepsilon)$ — покрытие объектами-эталонами множества $E_0$ в $n$-мерном признаковом пространстве при аппроксимации функций из $Y$ с максимальной величиной уклонения $\varepsilon$. Способность НС к обобщению должна проявляться при выполнении минимум двух условий:

1) $\lim\limits_{m \to \infty} | \; \Pi_{mn}(\varepsilon) \; | = C_n$;

2) $\lim\limits_{m \to \infty} \dfrac{| \; \Pi_{mn}(\varepsilon) \; |}{m} = 0$.

Первое условие обеспечивает ограниченность (конечность) числа объектов-эталонов покрытия $E_0$. Значение $C_n = \text{const}$ при фиксированной размерности пространства и состава его признаков. Смысл второго условия заключается в том, что число точек, которое может обобщить каждый объект-эталон для аппроксимации значений функции, стремится к бесконечности.

## 2. Аппроксимация функций в разнотипном признаковом пространстве

Считается, что множество значений целевого признака $Y$ измеряется по количественной шкале, $r$ признаков из числа зависимых $X_1, ..., X_n$ измеряются по количественным шкалам, $n - r$ — по номинальным. Обозначим множество количественных признаков через $I$, номинальных — через $J$.

Для решения проблемы разномасштабности измерений и согласования синаптических весов количественных и номинальных признаков используется дробно-линейное отображение значений количественных признаков в интервал [0,1]. По каждому признаку $x_t \in J$ с числом градаций $l_t$ производится разбиение значений целевого признака $\{y_i\}_1^m$ на $l_t$ непересекающихся классов $K_1, K_2, ..., K_{l_t}$ для вычисления весового коэффициента $w_t$. Упорядоченное множество значений $\{y_i\}_1^m$ разбивается на $l_t$ интервалов $(c_{2k-1}, c_{2k}]$, $c_{2k-1} < c_{2k}, k = \overline{1, l_t}$. Критерий для определения границ интервалов $(c_{2k-1}, c_{2k}]$ основывается на проверке гипотезы (утверждения) о том, что каждый интервал содержит значения признака только одного класса.

Пусть $u_i^p$ — множество значений целевого признака класса $K_i$ в интервале $(c_{2p-1}, c_{2p}]$, $A = (a_0, ..., a_{l_t}), a_0 = 0, a_{l_t} = m, a_p$ — порядковый номер элемента упорядоченной по возрастанию последовательности $r_{j_1}, ..., r_{j_m}$ значений целевого признака, определяющий правую границу интервала $c_{2p} = r_{a_p}$.

Критерий

$$\left( \frac{\sum\limits_{p=1}^{l_t} \sum\limits_{i=1}^{l_t} u_i^p (u_i^p - 1)}{\sum\limits_{i=1}^{l_t} \mid K_i \mid (\mid K_i \mid -1)} \right) \left( \frac{\sum\limits_{p=1}^{l_t} \sum\limits_{i=1}^{l_t} u_i^p (m - \mid K_i \mid - \sum\limits_{j=1}^{l_t} u_j^p + u_i^p)}{\sum\limits_{i=1}^{l_t} \mid K_i \mid (m - \mid K_i \mid)} \right) \to \max_{\{A\}} \qquad (3)$$

позволяет вычислять оптимальное значение синаптического веса $w_t$ при классификации $\{y_i\}_1^m$ по $l_t$ градациям признака $x_t \in J$.

Значение взвешенной суммы синаптических весов объекта $S = (b_1, ..., b_n)$ по объекту-эталону $S^d = (x_{d1}, ..., x_{dn})$ вычисляется как

$$\varphi(S^d, S) = \sum_{x_u \in I} w_{du} b_u + \sum_{x_u \in J, x_{du} = b_u} w_u^2 + w_{d0}, \qquad (4)$$

где $\{w_{d0}, w_{d1}, ..., w_{dn}\}$ — веса нейронов сети, $w_{d0} = -\frac{1}{2} \left( \sum\limits_{x_u \in I} w_{du}^2 + \sum\limits_{x_u \in J} w_u^2 \right)$.

При заданной величине максимального уклонения $\varepsilon$ табличных значений функции от аппроксимируемых объекты минимального покрытия $\Pi(\varepsilon)$ отбираются по тому же самому принципу, что и для количественных признаков. Локальная область для вычисления значений функции по $S$ определяется через объект-победитель $S^d \in \Pi(\varepsilon)$ по (4), для которого $\varphi(S^d, S) = \max\limits_{S^v \in \Pi(\varepsilon)} \varphi(S^v, S)$.

Проблема выбора признакового пространства при аппроксимации функций заключается в поиске критериев для выделения информативных наборов признаков и в данной работе специально не рассматривается.

## 3. Вычислительный эксперимент

Для эксперимента был взят модельный пример из 40 объектов с данными об аномальной сетевой активности, описываемых множеством из восьми разнотипных признаков:

1) характер воздействия;
2) цель воздействия;
3) условие начала воздействия;
4) наличие обратной связи с атакуемым объектом;
5) расположение субъекта атаки;

Результаты эксперимента

| Максимальное уклонение $\varepsilon$ | Число объектов покрытия | Среднее уклонение |
|---|---|---|
| 0.1 | 23 | 0.0437 |
| 0.15 | 18 | 0.0576 |
| 0.2 | 9 | 0.0807 |

6) объем передаваемой информации;

7) вероятность нанесения ущерба;

8) сумма ожидаемого ущерба.

Максимальное число градаций номинальных признаков в описании объектов не превышало 3. Точность аппроксимации значений целевого признака "Вероятность нанесения ущерба" семью зависимыми признаками определялась по объектам обучения $E_0 \backslash \Pi(\varepsilon)$ как величина среднего абсолютного уклонения вычисленного значения функции от табличного. Для вычисления значения функции использовалась радиально-базисная функция $\exp(\alpha_r \varphi(S^r, S))$. Влияние выбора величины максимального уклонения $\varepsilon$ на результаты аппроксимации целевого признака представлены в таблице.

## Список литературы

[1] Терехов В.А., Ефимов Д.В., Тюкин И.Ю. Нейросетевые системы управления М.: Высшая школа, 2002.

[2] Буцев А.В., Первозванский А.А. Локальная аппроксимация на искусственных нейросетях // Автом. и телемеханика. 1995. № 9. С. 127–136.

[3] Игнатьев Н.А., Мадрахимов Ш.Ф. О некоторых способах повышения прозрачности нейронных сетей // Вычисл. технологии. 2003. Т. 8, № 6. С. 31–37.

[4] Игнатьев Н.А. О синтезе факторов в искусственных нейронных сетях // Вычисл. технологии. 2005. Т. 10, № 3. С. 32–38.

[5] Вапник В.Н. Восстановление зависимостей по эпирическим данным. М.: Наука, 1979.

*Поступила в редакцию 9 ноября 2006 г.,*
*в переработанном виде — 28 июля 2008 г.*