

УДК 621.391; 519.21; 519.245

© А. В. Меркушева, Г. Ф. Малыхина

МЕТОДОЛОГИЯ СЕТЕЙ С СИММЕТРИЧНЫМИ ФУНКЦИЯМИ ПРЕОБРАЗОВАНИЯ НЕЙРОНОВ

Представлены основы методологии сетей с симметричными функциями преобразования у нейронов (СФПН), которые находят применение в задачах аппроксимации, распознавания образов, идентификации систем (объектов), создания контроллеров, снижения уровня зашумленности сигналов в информационно-измерительных системах. Особенность структуры таких нейронных сетей (НС) состоит в локализации элементов скрытого слоя в многомерном векторном пространстве (размерность которого идентична размерности входной информации) и в наличии СФПН, зависящей от (метрической) нормы разности векторов локализации элементов скрытого слоя и входного вектор-сигнала. Даны элементы прикладной теории распознавания образов на сети с СФПН; критерий обучения НС на основе функционала, регуляризованного по методу А.Н. Тихонова; общий вид функции аппроксимации и интерполирования, полученный на основе этого критерия с использованием схемы Грина для обратной задачи при преобразовании линейным дифференциальным оператором; метод выбора параметра регуляризации.

ВВЕДЕНИЕ

При аппроксимации функций, определенных на многомерной области, при классификации, решении задач идентификации и создания контроллеров используются нейронные сети с функциями преобразования (ФП) элементов скрытого слоя, которые образуют базис для аппроксимирующей функции и для разделяющей поверхности (например, гиперплоскости) при решении задачи классификации. Это — сети с (сферически) симметричными функциями преобразования нейронов (СФПН). В методологии анализа таких сетей имеются существенные особенности, поэтому целесообразно рассмотреть элементы их теории¹⁾.

Структура сети с СФПН включает слой входных узлов (с числом, равным размерности входных векторов), скрытый слой с СФПН и выходной слой с линейными ФП. Нейроны с СФП характеризуются (многомерным) вектором своего положения и зависимостью реакции на входное воздействие в виде функции (ФП), зависящей от расстояния (нормы разности) между входным вектором и вектором положения нейрона. Увеличение числа нейронов в скрытом слое повышает точность аппроксимации и делает линейно разделимыми (гиперплоскостью) классифицируемые объекты (векторы), которые были линейно неразделимыми при меньшем числе базисных функций — узлов сети с СФПН. Так, если при классификации (типа дихотомии) образов m_0 -мерных векторов \mathbf{x} высокой размерности их разделение на два класса

может быть реализовано только с помощью некоторой поверхности, то после нелинейного преобразования $\mathbf{x} \rightarrow \boldsymbol{\varphi}(\mathbf{x})$, $\boldsymbol{\varphi}(\mathbf{x}) = \{\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots, \varphi_m(\mathbf{x})\}$, $m > m_0$ в пространство более высокой размерности эти образы могут стать линейно разделимыми (гиперплоскостью²⁾. По такому принципу осуществляется классификация на нейронной сети (НС) с СФПН, где в качестве ФПН используются $\varphi_1(\mathbf{x})$, $\varphi_2(\mathbf{x}), \dots, \varphi_m(\mathbf{x})$.

При дихотомии на классы $\{\mathbf{X}_1, \mathbf{X}_2\}$ набор (множество) образов \mathbf{X} считается $\boldsymbol{\varphi}$ -разделимым, если для некоторого вектора \mathbf{w} выполняются условия:

$$\begin{aligned} \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) &> 0 \text{ при } \mathbf{x} \in \mathbf{X}_1; \\ \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) &< 0 \text{ при } \mathbf{x} \in \mathbf{X}_2. \end{aligned} \quad (1)$$

Соотношение $\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) = 0$ описывает разделяющую гиперплоскость в m -пространстве признаков $\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots, \varphi_m(\mathbf{x})$ входного вектора (образа) при $\forall \mathbf{x}$.

Принцип построения сети с СФПН связан с методом интерполяции по Пауэллу (Powell) [2], в котором по заданным N точкам (по векторам $\mathbf{x}_i \in \mathbf{R}_{m_0}$, $i = 1, 2, \dots, N$) находится функция F , удовлетворяющая условиям $F(\mathbf{x}_i) = d_i$, $i = 1, 2, \dots, N$ и имеющая вид:

$$F(\mathbf{x}) = \sum_{k=1}^N w_k \varphi(\|\mathbf{x} - \mathbf{x}_k\|), \quad (2)$$

¹⁾ Приложения сетей с СФПН предполагается проанализировать в следующей статье.

²⁾ Возможность линейной разделимости многомерных образов после их нелинейного преобразования в пространство признаков большей размерности установлено Ковером (Cover) [1].

где d_i — заданные в узлах значения интерполируемой переменной³⁾; $\varphi(\|\mathbf{x} - \mathbf{x}_k\|)$ — набор значений одного из видов нелинейных СФПН⁴⁾, размещенных в точках $\{\mathbf{x}_k\}_{k=1,2,\dots,N}$.

При векторно-матричном обозначении $\mathbf{d} = [d_1, d_2, \dots, d_N]^T$; $\mathbf{w} = [w_1, w_2, \dots, w_N]^T$; $\Phi = [\varphi_{ik}]_{i,k=1,2,\dots,N}$; $\varphi_{ik} = \varphi(\|\mathbf{x}_i - \mathbf{x}_k\|)$ условие интерполирования $F(\mathbf{x}_i) = d_i$ ($i = 1, \dots, N$) принимает форму соотношения

$$\Phi \mathbf{w} = \mathbf{d}. \quad (3)$$

Матрица Φ является несингулярной⁵⁾ [3, 4], и поэтому веса (компоненты вектора \mathbf{w}) в представлении интерполирующей функции могут быть определены по (4):

$$\mathbf{w} = \Phi^{-1} \mathbf{d}. \quad (4)$$

В нейронных сетях используются СФПН в виде гауссиана $\varphi(r) = \exp(-r^2 / 2\sigma^2)$, где $r \in R$, и менее часто в виде: $\varphi(r) = 1/(r^2 + c^2)^{1/2}$ (σ и c — константы).

1. СВОЙСТВА СЕТЕЙ С СФПН

Анализ свойств сетей с СФПН без ограничения общности выводов можно выполнить на задаче аппроксимации и интерполирования⁶⁾.

Нахождение интерполирующей функции F относится к классу обратных задач. Устойчивость процедуры получения F обеспечивается, если прямая и обратная задачи являются хорошо определенными (обусловленными — по терминологии А.Н. Тихонова). Для этого необходимо, чтобы функция f , которую представляет интерполирующая функция F (получаемая в результате обучения НС) обладала свойствами существования

³⁾ В (2) и далее $\|\dots\|$ — евклидова норма:

$$\|\mathbf{x} - \mathbf{x}_j\| = [\sum_{j=1}^m (x_j - (\mathbf{x}_j)_j)^2]^{1/2} \frac{\delta y}{\delta x}.$$

⁴⁾ Имея в виду математическую форму представления интерполирующей функции F в виде выражения (2), значения СФПН $\varphi(\|\mathbf{x} - \mathbf{x}_k\|)$ называют также радиальными базисными функциями, поэтому в американских работах используется термин RBF-сети.

⁵⁾ У рассматриваемой матрицы Φ (по условию несингулярности) $\det \Phi \neq 0$, и она имеет обратную Φ^{-1} .

⁶⁾ Полезно отметить, что процедура обучения сети точной аппроксимации иногда может давать плохие показатели при тестировании (результаты работы НС на данных, не участвовавших в обучении и имеющих статистические характеристики, одинаковые с обучающей выборкой). Это связано с "переобученностью" сети, возникающей при большом количестве данных, их значительном рассеянии относительно интерполирующей поверхности, использовании числа СФПН, равного числу обучающих векторов, и при слишком длительном обучении.

(для $\forall \mathbf{x} \exists y, y = f(\mathbf{x})$),

единственности

$$(\mathbf{x} \neq \mathbf{t} \Rightarrow f(\mathbf{x}) \neq f(\mathbf{t}))$$

и непрерывности

$$(\|\mathbf{x} - \mathbf{t}\| < \delta \Rightarrow |f(\mathbf{x}) - f(\mathbf{t})| < \varepsilon(\delta) \text{ при малых } \delta \text{ и } \varepsilon).$$

Нарушение этих условий может быть связано с рядом причин: входной набор $\{\mathbf{x}_i; d_i\}_{i=1,2,\dots,N}$ может быть недостаточно точным для восстановления единственной интерполирующей поверхности; при высокой зашумленности входного набора обучающих данных возможно получение выхода НС вне диапазона, предусмотренного для y , т. е. нарушение непрерывности.

При этом, хотя явления, порождающие данные (речь, изображение, сигнал радара), дают хорошо определенные задачи, тем не менее обучение НС, связанное с восстановлением функции f (в виде F , интерпретируемой как поверхность Γ в многомерном пространстве значений вектор-сигналов), часто бывает плохо обусловленной задачей.

Определение синаптических весов НС (компонент вектора \mathbf{w}) основано на управлении и контроле степени обученности НС по критерию, который является функционалом ошибки [5, 6]. Согласно методу А.Н. Тихонова [7–9], этот критерий в виде функционала ошибки $E_{\text{ош.}}$ (отражающего расстояние между желаемыми и действительными значениями на выходе НС)

$$E_{\text{ош.}}(F) = \frac{1}{2} \sum_{i=1}^N |d_i - F(\mathbf{x}_i)|^2 \quad (5)$$

расширяется путем введения дополнительного "регуляризирующего" функционала

$$E_{\text{пер.}}(F) = \frac{1}{2} \|DF\|^2, \quad (6)$$

где D — линейный дифференциальный оператор, который включает априорную информацию о степени гладкости функции F . Функционал, который минимизируется, имеет вид

$$E(F) = E_{\text{ош.}} + \lambda \cdot E_{\text{пер.}}, \quad (7)$$

а положительное число λ определяет баланс относительной значимости требований минимизации ошибки обучения НС и гладкости аппроксимирующей функции F .

В связи с тем что для получения F осуществляется минимизация по функции векторного аргумента ($F(\mathbf{x})$), условие $dE(F) = 0$, необходимое для минимума, выражается в форме дифференциала Фреше ($D\Phi$) [10–12]:

$$dE(F, h) = \left[\frac{d}{d\beta} E(F + \beta h) \right]_{\beta=0} = 0, \quad (8)$$

где h — функция от вектора \mathbf{x} ($h(\mathbf{x})$) так же, как и $F(\mathbf{x})$, $\mathbf{x} \in \mathbf{R}_{m_0}$. Вычисление компонент этого дифференциала, основанное на использовании соотношения (7) и общего правила получения ДФ⁷⁾, приводит к выражениям (9):

$$dE_{\text{ош.}}(F, h) = dE_{\text{ош.}}(F, h) + \lambda \cdot dE_{\text{пер.}}(F, h) = 0, \\ dE_{\text{ош.}}(F, h) = - \left(h, \sum_{i=1}^N (d_i - F(\mathbf{x}_i)) \cdot \delta_{\mathbf{x}_i} \right)_H, \quad (9)$$

$$dE_{\text{пер.}}(F, h) = \int_{\mathbf{R}_{x_i}} DF \cdot Dh \cdot d\mathbf{x} = (Dh, DF)_H,$$

где $(\cdot, \cdot)_H$ — символ произведения в гильбертовом пространстве (интеграл произведения функций по области их определения) [13, 14]; $\delta_{\mathbf{x}_i}$ — дельта-функция $\delta_{\mathbf{x}_i}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_i)$.

Согласно правилу переноса дифференциального оператора D с одного множителя на другой в форме сопряженного оператора $\tilde{D} : \int_{\mathbf{R}_{m_0}} DF \cdot Dh \cdot d\mathbf{x} = \int_{\mathbf{R}_{m_0}} h(\mathbf{x}) \tilde{D} DF(\mathbf{x}) d\mathbf{x}$, последнее соотношение в (9) приобретает вид

⁷⁾ Для получения ДФ компонент общего (регуляризованного) функционала (7) служат преобразования:

$$dE_{\text{ош.}}(F, h) = \left[\frac{d}{d\beta} E_{\text{ош.}}(F + \beta \cdot h) \right]_{\beta=0} = \\ = \left[\frac{1}{2} \times \frac{d}{d\beta} \sum_{i=1}^N (d_i - F(\mathbf{x}_i) - \beta \cdot h(\mathbf{x}_i))^2 \right]_{\beta=0} = \\ = - \left[\sum_{i=1}^N (d_i - F(\mathbf{x}_i) - \beta \cdot h(\mathbf{x}_i)) \cdot h(\mathbf{x}_i) \right]_{\beta=0} = \\ = - \sum_{i=1}^N (d_i - F(\mathbf{x}_i)) \cdot h(\mathbf{x}_i) = - \left(h, \sum_{i=1}^N (d_i - F(\mathbf{x}_i)) \cdot \delta_{\mathbf{x}_i} \right)_H;$$

$$dE_{\text{пер.}}(F, h) = \left[\frac{d}{d\beta} E_{\text{пер.}}(F + \beta h) \right]_{\beta=0} = \\ = \left[\frac{1}{2} \cdot \frac{d}{d\beta} \int_{\mathbf{R}_{x_i}} D[F + \beta h]^2 d\mathbf{x} \right]_{\beta=0} = \\ = \frac{2}{2} \left[\frac{d}{d\beta} \int_{\mathbf{R}_{x_i}} D[F + \beta h] Dh \cdot d\mathbf{x} \right]_{\beta=0} = \\ = \int_{\mathbf{R}_{x_i}} DF \cdot Dh \cdot d\mathbf{x} = (Dh, DF)_H.$$

Отметим, что наряду с термином ДФ встречается эквивалентное по содержанию понятие дифференциала Га-то, например в [12].

$$dE_{\text{пер.}}(F, h) = \int_{\mathbf{R}_{m_0}} h(\mathbf{x}) D \tilde{D} F(\mathbf{x}) d\mathbf{x} = (h, \tilde{D} DF)_H.$$

С учетом этого ДФ от общего (регуляризованного) функционала (7) преобразуется к форме

$$dE(F, h) = \left(h, \left[\tilde{D} DF - \frac{1}{\lambda} \sum_{i=1}^N (d_i - F) \cdot \delta_{\mathbf{x}_i} \right] \right)_H, \quad (10)$$

и поскольку (10) справедливо для любого вида функции $h(\mathbf{x})$, то условие минимизации представляется в виде соотношения (11):

$$\tilde{D} DF - \frac{1}{\lambda} \sum_{i=1}^N (d_i - F) \cdot \delta(\mathbf{x} - \mathbf{x}_i) = 0. \quad (11)$$

Функция F , удовлетворяющая (11), зависит от λ (ниже это отмечено индексом — $F_\lambda(\mathbf{x})$) и находится с помощью метода Грина⁸⁾. Для этого необходимо:

- Изменить форму уравнения (11), выделив явным образом линейный дифференциальный оператор (L), и представить это уравнение в виде ([15], см. также⁸⁾)

$$L F_\lambda(\mathbf{x}) = \varphi(\mathbf{x}), \quad \text{где} \\ L \equiv \tilde{D} D; \quad (12)$$

$$\varphi(\mathbf{x}) = \frac{1}{\lambda} \sum_{i=1}^N [d_i - F(\mathbf{x}_i)] \delta(\mathbf{x} - \mathbf{x}_i) = 0.$$

- Получить функцию $F_\lambda(\mathbf{x})$ как интегральное выражение $F_\lambda(\mathbf{x}) = \int_{\mathbf{R}_{m_0}} G(\mathbf{x}, \xi) \varphi(\xi) d\xi$, в котором

$G(\mathbf{x}, \xi)$ — функция Грина для оператора $L \equiv \tilde{D} D$ и в соответствии с выражением (12)

⁸⁾ Методом Грина решаются уравнения (в частных производных) общего вида $LF(\mathbf{x}) = \varphi(\mathbf{x})$, $\mathbf{x} \in \mathbf{R}_{m_0}$, где L — линейный дифференциальный оператор; $\varphi(\mathbf{x})$ — произвольная функция. Решение может быть представлено в виде $F(\mathbf{x}) = \int_{\mathbf{R}_{m_0}} G(\mathbf{x}, \xi) \varphi(\xi) d\xi$, где $G(\mathbf{x}, \xi)$ является

решением квазиоднородного уравнения (с \mathbf{x} в качестве переменной и ξ в качестве параметра) $LG(\mathbf{x}, \xi) = \delta(\mathbf{x} - \xi)$, где $\delta(\mathbf{x} - \xi)$ — δ -функция. При этом δ -функцию (Дирака) называют точечным источником возбуждения; по форме — это предельно узкая плотность распределения, сосредоточенная в точке ξ , так что

$$\int_{\mathbf{R}_{m_0}} \varphi(\mathbf{x}) \delta(\mathbf{x} - \xi) d\mathbf{x} = \varphi(\xi), \quad \int_{\mathbf{R}_{m_0}} \varphi(\xi) \delta(\mathbf{x} - \xi) d\xi = \varphi(\mathbf{x}).$$

Более полные сведения о методах, связанных с использованием функций Грина, содержатся в [16–18].

$\varphi(\xi) = \frac{1}{\lambda} \sum_{i=1}^N [d_i - F(\mathbf{x}_i)] \delta(\xi - \mathbf{x}_i)$. Полное выражение для $F_\lambda(\mathbf{x})$ имеет вид:

$$\begin{aligned} F_\lambda(\mathbf{x}) &= \\ &= \int_{\mathbf{R}^m} G(\mathbf{x}, \xi) \left\{ \frac{1}{\lambda} \sum_{i=1}^N [d_i - F(\mathbf{x}_i)] \cdot \delta(\xi - \mathbf{x}_i) \right\} d\xi = \\ &= \frac{1}{\lambda} \sum_{i=1}^N [d_i - F(\mathbf{x}_i)] \cdot G(\mathbf{x} - \mathbf{x}_i). \end{aligned} \quad (13)$$

Решение задачи оценки аппроксимирующей функции $F_\lambda(\mathbf{x})$ (по критерию регуляризованного функционала) в форме разложения по функциям Грина имеет коэффициенты w_i , зависящие линейно от величины ошибки на выходе НС. Поэтому $F_\lambda(\mathbf{x})$ может быть представлено соотношением (14):

$$\begin{aligned} F_\lambda(\mathbf{x}) &= \sum_{i=1}^N w_i G(\mathbf{x}, \mathbf{x}_i), \\ w_i &= \frac{1}{\lambda} [d_i - F(\mathbf{x}_i)], \quad i = 1, 2, \dots, N. \end{aligned} \quad (14)$$

Более компактное представление получается после введения векторно-матричных обозначений

$$\begin{aligned} \mathbf{F}_\lambda &= [F_\lambda(\mathbf{x}_1), F_\lambda(\mathbf{x}_2), \dots, F_\lambda(\mathbf{x}_N)]^T; \\ \mathbf{d} &= [d_1, d_2, \dots, d_N]^T; \\ \mathbf{G} &= \begin{bmatrix} G(\mathbf{x}_1, \mathbf{x}_1) & \dots & G(\mathbf{x}_1, \mathbf{x}_N) \\ \dots & \dots & \dots \\ G(\mathbf{x}_N, \mathbf{x}_1) & \dots & G(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}; \end{aligned} \quad (15)$$

$$\mathbf{w} = [w_1, w_2, \dots, w_N]^T,$$

с помощью которых соотношения (14) переходят в эквивалентную форму (16):

$$\mathbf{F}_\lambda = \mathbf{G} \cdot \mathbf{w}; \quad \mathbf{w} = (1/\lambda)(\mathbf{d} - \mathbf{F}_\lambda). \quad (16)$$

При этом вектор коэффициентов разложения и вектор значений функции $F_\lambda(\mathbf{x})$ относятся к центрам функций Грина (и одновременно к центрам локализации нейронов в сети), а компоненты матрицы \mathbf{G} симметричны ($G(\mathbf{x}_i, \mathbf{x}_j) = G(\mathbf{x}_j, \mathbf{x}_i)$), поскольку они относятся к решениям уравнения с самосопряженным оператором L ($\tilde{L} = L$).

Сама матрица Грина \mathbf{G} симметрична, т. е.

$$\mathbf{G}^T = \mathbf{G}.$$

Исключение \mathbf{F}_λ из соотношений (16) ведет к системе (17) линейных уравнений (в векторно-матричной форме), которая служит для определения коэффициентов разложения — компонент вектора $\mathbf{w} = [w_1, w_2, \dots, w_N]^T$:

$$(\mathbf{G} + \lambda \cdot \mathbf{I}) \mathbf{w} = \mathbf{d}. \quad (17)$$

При обратимости матрицы $\mathbf{G} + \lambda \cdot \mathbf{I}$, которая, согласно [3], достигается при указанных выше видах ФПН и облегчается выбором величины λ , система линейных уравнений (17) имеет решение, определяющее коэффициенты разложения:

$$\mathbf{w} = (\mathbf{G} + \lambda \cdot \mathbf{I})^{-1} \mathbf{d}. \quad (18)$$

Таким образом, если в регуляризуемом функционале (6) критерия обучения выбран дифференциальный оператор D , то по соотношению (18) может быть определен вектор веса \mathbf{w} , соответствующий желаемому отклику НС и подходящей величине параметра регуляризации λ .

Вид функций Грина (ФГ) при выбранных ее центрах $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ определяется типом оператора D , т. е. фактически априорными предположениями относительно отображения вход—выход НС. При D , инвариантном к переносу, ФГ зависят только от разности своих аргументов $G(\mathbf{x}, \mathbf{x}_i) = G(\mathbf{x} - \mathbf{x}_i)$; при инвариантности к вращению и переносу — ФГ зависят от нормы разности тех же векторов $G(\mathbf{x}, \mathbf{x}_i) = G(\|\mathbf{x} - \mathbf{x}_i\|)$ и являются "зависящими от радиуса" базисными функциями. Решение задачи с регуляризованным критерием обучения в этом случае имеет вид

$$F_\lambda(\mathbf{x}) = \sum_{i=1}^N w_i G(\|\mathbf{x} - \mathbf{x}_i\|), \quad (19)$$

который за счет использования регуляризации по методу А.Н. Тихонова существенно отличается от полученного ранее более простого выражения (2).

2. ОСНОВНАЯ ФОРМА СФПН И НЕЙРОННЫЕ СЕТИ, ОСНОВАННЫЕ НА МЕТОДЕ РЕГУЛЯРИЗАЦИИ

Наиболее практически значимыми и применимыми являются сети с СФПН на основе функций Грина, у которых порождающий их оператор инвариантен к вращению и переносу и которые представляют многомерный гауссиан:

$$G(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{1}{2\sigma_i^2} \|\mathbf{x} - \mathbf{x}_i\|^2\right), \quad (20)$$

где \mathbf{x}_i, σ_i — центр и полуширина гауссиана и соответствующей СФПН в скрытом слое НС. Оператор L ($L \equiv \tilde{D}D$), который связан с дифференци-

альным оператором D , используется в регуляризующей компоненте общего функционала (критерия обучения НС) и порождает функцию Грина в форме (20), имеет следующий вид [19]:

$$L = \tilde{D}D; \quad (21)$$

$$L = \sum_{n=0}^{\infty} (-1)^n \alpha_n \nabla^{2n}, \quad \text{при } \alpha_n = \frac{\sigma_i^{2n}}{n!2^n},$$

где ∇^{2n} — n -кратная итерация многомерного аналога оператора Лапласа ∇^2 :

$$\nabla^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \dots + \frac{\partial^2}{\partial x_{m_0}^2}, \quad (22)$$

$$\nabla^{2n} f(\mathbf{x}) = (\nabla^2 (\nabla^2 \dots (\nabla^2 f))).$$

Преобразования, выполненные Поггио (Poggio) [19], дают явную форму для D (оператора регуляризации) и ему сопряженного \tilde{D} :

$$D = \sum_n a_n^{1/2} \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2} + \dots + \frac{\partial}{\partial x_{m_0}} \right)^n =$$

$$= \sum_{a+b+\dots+k=n} a_n^{1/2} \frac{\partial^n}{(\partial x_1)^a (\partial x_2)^b \dots (\partial x_{m_0})^k}; \quad (23)$$

$$\tilde{D} = \sum_n (-1)^n a_n^{1/2} \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2} + \dots + \frac{\partial}{\partial x_{m_0}} \right)^n =$$

$$= \sum_{a+b+\dots+k=n} (-1)^n a_n^{1/2} \frac{\partial^n}{(\partial x_1)^a (\partial x_2)^b \dots (\partial x_{m_0})^k}; \quad (24)$$

Таким образом, решение задачи аппроксимации средствами сети с СФПН при применении метода А.Н. Тихонова может быть получено на основе оператора регуляризации, включающего все частные производные (а точнее все необходимые — исходя из априорных представлений о требуемой степени гладкости F_λ).

Использование ФГ в форме гауссиана (20), рассматриваемого как решение операторного уравнения $LG(\mathbf{x}, \xi) = \delta(\mathbf{x} - \xi)$ (см. примечание ⁸), в котором оператор L определен выражением (21), при замене ξ на \mathbf{x}_i приводит к соотношению (25):

$$\sum_n (-1)^n \frac{\sigma_i^{2n}}{n!2^n} \nabla^{2n} \exp\left(\frac{1}{2\sigma_i^2} \|\mathbf{x} - \mathbf{x}_i\|^2\right) =$$

$$= \delta(\mathbf{x} - \mathbf{x}_i). \quad (25)$$

Этому уравнению удовлетворяет ФГ в форме гауссиана; эта же ФГ является СФПН анализируе-

мой структуры сети; и регуляризованное нейросетевое решение определяется выражением

$$F_\lambda(\mathbf{x}) = \sum_{i=1}^N w_i \exp\left(\frac{1}{2\sigma_i^2} \|\mathbf{x} - \mathbf{x}_i\|^2\right), \quad (26)$$

в котором веса w_i выражаются вторым соотношением в (14)⁹.

Практическая реализация представления регуляризованной функции аппроксимации F_λ в форме взвешенной суммы ФГ $G(\mathbf{x}, \mathbf{x}_i) = G(\mathbf{x} - \mathbf{x}_i)$, центрированных в точках \mathbf{x}_i ($i = 1, \dots, N$), предполагает определенную структуру НС (рис. 1). Как отмечено в начале статьи (Введение), простой вариант структуры сети с СФПН состоит из трех слоев: слоя входных узлов (общего для большинства структур НС) с числом узлов m_0 , равным размерности исследуемой задачи (т. е. размерности входных векторов); скрытого слоя, который при использовании метода регуляризации состоит из нейронов с СФП, эквивалентной функции Грина для выбранного оператора регуляризации D , и состоит из N нейронов, где N — размер обучающей выборки $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$; одного линейного выходного нейрона, полностью связанного со всеми нейронами скрытого слоя весами w_i ($i = 1, \dots, N$). Веса выходного слоя являются неизвестными коэффициентами разложения (26) и, согласно (18), определяются значениями ФГ $G(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1, 2, \dots, N$ и параметра регуляризации λ . При решении задач распознавания образов число нейронов выходного слоя соответствует количеству дифференцируемых классов (групп со специфичным набором признаков).

СФПН (функции возбуждения нейронов скрытого слоя НС) в форме ФГ центрированы в различных точках обучающей выборки, так что выходом i -го скрытого нейрона ($\forall i, i = 1, 2, \dots, N$) служит $G(\mathbf{x}, \mathbf{x}_i)$.

3. РАСШИРЕННАЯ КОНЦЕПЦИЯ СЕТЕЙ С СПФН

Кроме достоинств сетей с СФПН и регуляризацией функционала (являющегося критерием качества обучения НС), таких как:

- универсальность возможностей, обеспечивающая получение желаемого качества аппроксимации любой непрерывной функции многих переменных и распознавания многопараметрических объектов (образов);

⁹ Следует отметить, что использование различных значений полуширины гауссианов (параметров σ_i) в структурах сетей с СФПН с регуляризацией в реальных приложениях пока не находит применения.

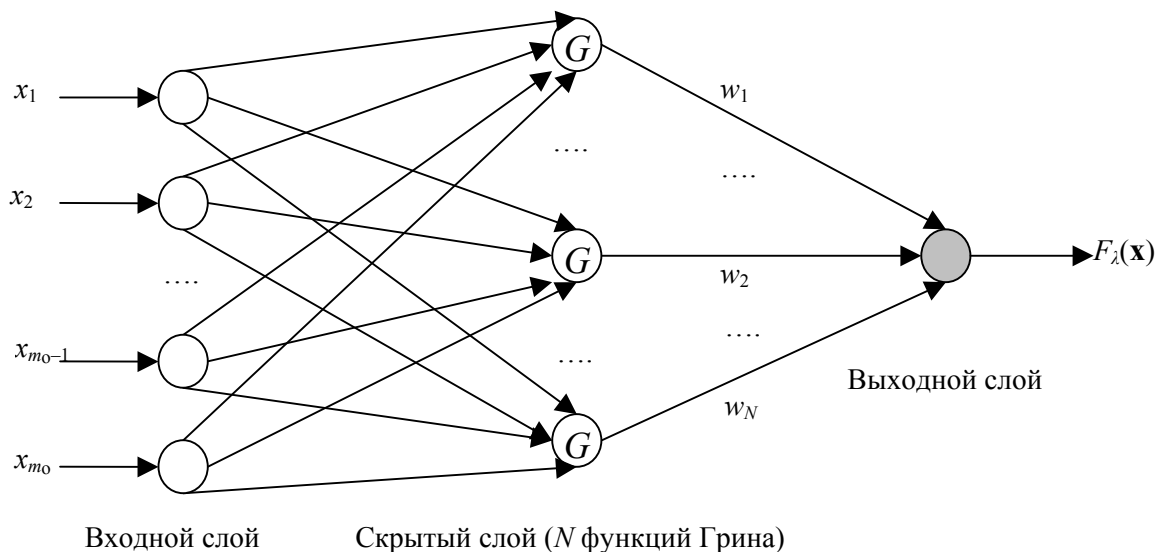


Рис. 1. Нейронная сеть с СФПН и регуляризацией функционала критерия обучения НС

- линейность схемы аппроксимации и распознавания образов относительно неизвестных коэффициентов, позволяющая выбрать их наилучшие оценки;

- оптимальность решения при применении сети с регуляризацией в смысле минимизации функционала, который служит мерой близости показаний сети от истинной величины, определяемой обучающими данными, — имеются и некоторые трудности. Так, соответствие числа функций Грина и количества обучающих данных ($\mathbf{x}_i, i=1,2,\dots,N$) при больших N делает избыточным объем вычислений; например, для определения коэффициентов разложения $F_\lambda(\mathbf{x})$ по (19) требуется обращение матрицы \mathbf{G} порядка $N \times N$. Кроме того, для матриц высокого порядка более вероятно плохая обусловленность (большое относительное различие минимального и максимального собственных значений матрицы).

В связи с этим получила развитие расширенная концепция сетей с СПФН, которая отражает тенденцию контролируемого снижения сложности сети и получения приближенной формы регуляризованного решения. Основу метода упрощения структуры НС составляет поиск субоптимального решения (достаточно близкого к решению в форме (19)) в пространстве меньшей размерности¹⁰⁾.

Приближенное решение $F^*(\mathbf{x})$ представляется

¹⁰⁾ В прикладной математике и механике описанные далее процедуры называют методом Галеркина.

в виде разложения по сокращенному набору базисных функций (БФ): $F^*(\mathbf{x}) = \sum_{i=1}^{m_1} w_i \varphi_i(\mathbf{x})$, $\{\varphi_i(\mathbf{x}) \quad i=1,2,\dots,m_1\}$, причем число БФ меньше, чем количество обучающих данных $m_1 < N$. В качестве БФ берутся СФПН $\{\varphi_i(\|\mathbf{x} - \mathbf{t}_i\|) \quad i=1,2,\dots,m_1\}$, т. к. только такой вид БФ при $m_1 \rightarrow N$ и $\mathbf{t}_i = \mathbf{x}_i \quad (i=1,2,\dots,N)$ обеспечивает плавный переход приближенного решения $F^*(\mathbf{x})$ в полное корректное решение, определяемое выражением (19).

Использование БФ типа $\{\varphi_i(\|\mathbf{x} - \mathbf{t}_i\|)\}$ в виде ФГ дает новую форму аппроксимации

$$F^*(\mathbf{x}) = \sum_{i=1}^{m_1} w_i G(\mathbf{x}, \mathbf{t}_i) = \sum_{i=1}^{m_1} w_i G(\|\mathbf{x} - \mathbf{t}_i\|). \quad (27)$$

Структура последующих преобразований аналогична основной схеме получения решения по методу регуляризации. Сокращенный набор коэффициентов разложения $\{w_i \quad (i=1,2,\dots,m_1)\}$ определяется на основе минимизации регуляризованного функционала $E(F^*)$, который сформирован с усеченным разложением аппроксимирующей функции $F^*(\mathbf{x})$

$$E(F^*) = \sum_{i=1}^N \left[d_i - \sum_{j=1}^{m_1} w_j G(\|\mathbf{x}_i - \mathbf{t}_j\|) \right]^2 + \lambda \|DF^*\|^2. \quad (28)$$

Использование векторно-матричных обозначений, сходных с (15), позволяет представить (28) в виде

$$E(F^*) = \| \mathbf{d} - \mathbf{G}\mathbf{w} \|^2 + \lambda \| DF^* \|^2, \quad (29)$$

где

$$\mathbf{d} = [d_1, d_2, \dots, d_N]; \quad \mathbf{w} = [w_1, w_2, \dots, w_{m_1}];$$

$$\mathbf{G}_{N \times m_1} = \begin{bmatrix} G(\mathbf{x}_1, \mathbf{t}_1) & G(\mathbf{x}_1, \mathbf{t}_2) & \dots & G(\mathbf{x}_1, \mathbf{t}_{m_1}) \\ G(\mathbf{x}_2, \mathbf{t}_1) & G(\mathbf{x}_2, \mathbf{t}_2) & \dots & G(\mathbf{x}_2, \mathbf{t}_{m_1}) \\ \dots & \dots & \dots & \dots \\ G(\mathbf{x}_N, \mathbf{t}_1) & G(\mathbf{x}_N, \mathbf{t}_2) & \dots & G(\mathbf{x}_N, \mathbf{t}_{m_1}) \end{bmatrix}.$$

Регуляризирующая компонента функционала (29) преобразуется к выражению (30)¹¹⁾:

$$\lambda \| DF^* \|^2 = \lambda (DF^*, DF^*)_H = \lambda \mathbf{w}^T \mathbf{G}_0 \mathbf{w}, \quad (30)$$

где

$$\mathbf{G}_0_{m_1 \times m_1} = \begin{bmatrix} G(\mathbf{x}_1, \mathbf{t}_1) & G(\mathbf{x}_1, \mathbf{t}_2) & \dots & G(\mathbf{x}_1, \mathbf{t}_{m_1}) \\ G(\mathbf{x}_2, \mathbf{t}_1) & G(\mathbf{x}_2, \mathbf{t}_2) & \dots & G(\mathbf{x}_2, \mathbf{t}_{m_1}) \\ \dots & \dots & \dots & \dots \\ G(\mathbf{x}_{m_1}, \mathbf{t}_1) & G(\mathbf{x}_{m_1}, \mathbf{t}_2) & \dots & G(\mathbf{x}_{m_1}, \mathbf{t}_{m_1}) \end{bmatrix};$$

$$\begin{aligned} \| DF^* \|^2 &= (DF^*, DF^*)_H = \\ &= \left[\sum_{i=1}^{m_1} w_i G(\mathbf{x}, \mathbf{t}_i), \tilde{D}D \sum_{i=1}^{m_1} w_i G(\mathbf{x}, \mathbf{t}_i) \right]_H = \\ &= \left[\sum_{i=1}^{m_1} w_i G(\mathbf{x}, \mathbf{t}_i), \sum_{j=1}^{m_1} w_j \delta(\mathbf{x} - \mathbf{t}_j) \right]_H = \\ &= \sum_{i=1}^{m_1} \sum_{j=1}^{m_1} w_i w_j G(\mathbf{t}_j, \mathbf{t}_i) = \mathbf{w}^T \mathbf{G}_0 \mathbf{w}. \end{aligned}$$

Напомним, что индексами под символом матрицы обычно указывается ее размерность, как в (29) и (30).

Минимизация функционала (29) (критерия качества обучения сети с СФПН) относительно вектора коэффициентов осуществляется так же, как выше в разделе 2, и приводит к соотношению

$$(\mathbf{G}^T \mathbf{G} + \lambda \mathbf{G}_0) \mathbf{w} = \mathbf{G}^T \mathbf{d}. \quad (31)$$

При этом, если параметр регуляризации λ приближается к нулю, вектор веса \mathbf{w} сходится к решению, определяемому псевдообратной матрицей

$\mathbf{G}^+ : \lambda \rightarrow 0 \Rightarrow \mathbf{w} = \mathbf{G}^+ \mathbf{d}$, где $\mathbf{G}^+ = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$. В этом случае вектор синаптических весов НС \mathbf{w} является решением (с минимальной нормой) задачи подгонки методом наименьших квадратов при переопределенном ($N > m_1$) количестве данных.

В случае, когда отдельные параметры исследуемого объекта (компоненты входного вектора \mathbf{x}) имеют в решаемой задаче различную значимость, используется взвешенная норма $\| \mathbf{x} \|_C$, которая определяется специально подобранной матрицей \mathbf{C} с размерностью $(m_0 \times m_0)$, соответствующей размерности m_0 входного вектора:

$$\| \mathbf{x} \|_C = (\mathbf{C} \mathbf{x})^T (\mathbf{C} \mathbf{x}) = \mathbf{x}^T \mathbf{C}^T \mathbf{C} \mathbf{x}.$$

При использовании такой взвешенной нормы аппроксимация регуляризованного решения (27) представляется в виде

$$F^*(\mathbf{x}) = \sum_{j=1}^{m_1} w_j G(\| \mathbf{x} - \mathbf{t}_j \|_C).$$

Применение взвешивающей нормы в СФПН в виде гауссиана с центром \mathbf{t}_i приводит к соотношению

$$\begin{aligned} G(\| \mathbf{x} - \mathbf{t}_i \|_C) &= \exp \left[-(\mathbf{x} - \mathbf{t}_i)^T \mathbf{C}^T \mathbf{C} (\mathbf{x} - \mathbf{t}_i) \right] = \\ &= \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{t}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{t}_i) \right], \end{aligned} \quad (32)$$

где обратная матрица $\boldsymbol{\Sigma}^{-1}$ определена выражением $\frac{1}{2} \boldsymbol{\Sigma}^{-1} = \mathbf{C}^T \mathbf{C}$.

Правая часть в (32) представляет многомерное распределение Гаусса с вектором среднего \mathbf{t}_i и с ковариационной матрицей $\boldsymbol{\Sigma}$. Поэтому соотношение (32) является обобщением выражения (20), которое относится к основной форме сети с СФПН, построенной по методу регуляризации функционала ошибки обучения.

Решение задачи аппроксимации, представленное в форме $F^*(\mathbf{x}) = \sum_{i=1}^{m_1} w_i \varphi_i(\mathbf{x})$ (с укороченным разложением и процедурой определения синаптических весов w_i на основе функционала (28)), дает основу для построения структуры сети с СФПН, показанной на рис. 2.

Структура сети по расширенной концепции (СРК) сходна с общим типом сети с регуляризацией функционала (ОТСР), но имеет существенные различия с ней.

- Число нейронов в скрытом слое СРК равно m_1 и обычно значительно меньше числа N образцов, доступных для обучения НС, тогда как число нейронов в скрытом слое ОТСР равно в точности N .

¹¹⁾ Преобразования, ведущие к (30) учитывают представление $F^*(\mathbf{x})$ в виде (27), правило введения сопряженного оператора \tilde{D} в гильбертовом произведении функций $(DF^*, DF^*)_H = (F^*, \tilde{D}DF^*)_H$, и то, что ФГ $G(\mathbf{x}, \mathbf{t}_j)$ является решением уравнения $\tilde{D}DG(\mathbf{x}, \mathbf{t}_j) = \delta(\mathbf{x} - \mathbf{t}_j)$.

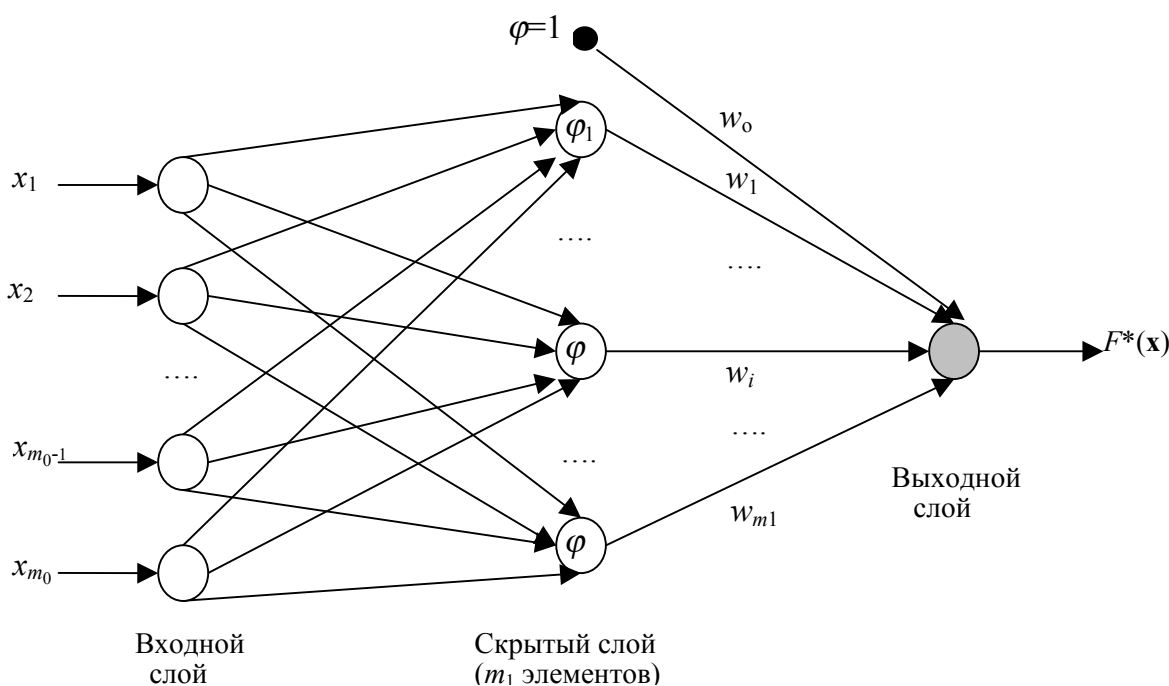


Рис. 2. Сеть с укороченным разложением $F^*(\mathbf{x})$ по СФПН (на основе расширенной концепции сетей с регуляризацией). Дополнительный элемент скрытого слоя, не зависящий от данных, обеспечивает смещение w_0 для нейрона выходного слоя

• У СРК (рис. 2) веса выходного слоя (w_1, \dots, w_{m1}), положение центров СПФН и взвешивающая матрица являются неизвестными и входят в состав параметров, значения которых определяются в процессе обучения сети. Напротив, у ОТСР функции СФПН известны и определяются функциями Грина. Центры этих ФГ находятся в точках обучающей выборки и единственно неизвестными параметрами у ОТСР (которые определяются при обучении) являются линейные веса выходного слоя.

4. ТОЧНОСТЬ ФУНКЦИОНИРОВАНИЯ СЕТИ ПРИ МЕТОДЕ РЕГУЛЯРИЗАЦИИ И ОЦЕНКА ПАРАМЕТРА λ

В процедурах, связанных с методом регуляризации сети с СФПН, с анализом среднеквадратичной ошибки НС и ее контролем, существенную роль играет параметр λ . Подход к оценке λ удобно показать, используя в качестве модели задачу нелинейной регрессии, в которой выход НС y_i на входной вектор-сигнал \mathbf{x}_i определяется соотношением

$$y_i = f(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, 2, \dots, N, \quad (33)$$

где $f(\mathbf{x}_i)$ — гладкая зависимость; ε_i — отсчеты белого шума с нулевым средним и дисперсией σ^2 , т. е. $E[\varepsilon_i] = 0 \forall i$; $E[\varepsilon_i \varepsilon_k] = \sigma^2 \delta_{ik}$ (здесь E — символ оператора математического ожидания, δ_{ik} — символ Кронекера), а задача состоит в восстановлении функции $f(\mathbf{x}_i)$ из заданного класса на основе использования обучающих данных $\{(\mathbf{x}_i, y_i)\}_{i=1,2,\dots,N}$.

Если $F_\lambda(\mathbf{x}_i)$ — это оценка $f(\mathbf{x}_i)$ по методу регуляризации для некоторой величины параметра λ , то $F_\lambda(\mathbf{x}_i)$ минимизирует функционал, составленный для задачи нелинейной регрессии в виде: $E(F) = (1/2) \sum_{i=1}^N [y_i - F(\mathbf{x}_i)]^2 + (\lambda/2) \|DF(\mathbf{x}_i)\|^2$. При этом выбор подходящего значения λ должен уравнивать относительную значимость двух требований к функционированию обученной НС:

• Обеспечение гладкости $F(\mathbf{x})$, предусмотренной заданием вида дифференциального оператора D .

• Обеспечение желательной меньшей суммарной квадратичной ошибки $\sum_{i=1}^N [y_i - F(\mathbf{x}_i)]^2$, источником которой служит неполнота соответствия данных и модели регрессионной зависимости с предусмотренной гладкостью.

Представление $F_\lambda(\mathbf{x}_k)$ на обучающем векторе \mathbf{x}_k (через набор наблюдаемых значений $\{y_i\}_{i=1,2,\dots,N}$) в виде $F_\lambda(\mathbf{x}_k) = \sum_{i=1}^N a_{ik}(\lambda)y_i$ при использовании векторно-матричных обозначений принимает форму соотношения (34):

$$\mathbf{F}_\lambda = \mathbf{A}(\lambda)\mathbf{y}, \quad (34)$$

где

$$\mathbf{F}_\lambda = [F_\lambda(\mathbf{x}_1), F_\lambda(\mathbf{x}_2), \dots, F_\lambda(\mathbf{x}_N)]^T; \quad \mathbf{y} = [y_1, y_2, \dots, y_N]^T;$$

$$\mathbf{A}(\lambda) = \begin{bmatrix} a_{11} & \dots & a_{1N} \\ \dots & \dots & \dots \\ a_{M1} & \dots & a_{MN} \end{bmatrix}.$$

Аналогичным образом при использовании (в качестве обозначения) вектора отсчетов функции $\mathbf{f} = [f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_N)]^T$ модели регрессии в точках обучающей выборки может быть представлена среднеквадратичная ошибка $R(\lambda)$, определенная как $R(\lambda) = (1/N) \sum_{i=1}^N [f(\mathbf{x}_i) - F_\lambda(\mathbf{x}_i)]^2$, в виде

$$R(\lambda) = \frac{1}{N} \|\mathbf{f} - \mathbf{F}_\lambda\|^2 = \frac{1}{N} \|\mathbf{f} - \mathbf{A}(\lambda)\mathbf{y}\|^2. \quad (35)$$

Таким же способом получается эквивалентное векторное представление выражения (33):

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\varepsilon}, \quad \text{где вектор } \boldsymbol{\varepsilon} \equiv [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N]^T.$$

Проделанные преобразования позволяют придать среднеквадратичной ошибке $R(\lambda)$ форму, удобную для получения для нее определенных характеристик:

$$\begin{aligned} R(\lambda) &= (1/N) \|(\mathbf{I} - \mathbf{A}(\lambda))\mathbf{f} - \mathbf{A}(\lambda)\boldsymbol{\varepsilon}\|^2 = \\ &= (1/N) \|(\mathbf{I} - \mathbf{A}(\lambda))\mathbf{f}\|^2 - \\ &\quad - (2/N) \boldsymbol{\varepsilon}^T \mathbf{A}(\lambda) (\mathbf{I} - \mathbf{A}(\lambda)) \mathbf{f} + \\ &\quad + (1/N) \|\mathbf{A}(\lambda)\boldsymbol{\varepsilon}\|^2, \end{aligned} \quad (36)$$

\mathbf{I} — матрица тождественного преобразования.

При этом путем применения к (36) оператора математического ожидания (E) получается оценка

средней величины $R(\lambda)^{12)}$:

$$\begin{aligned} E[R(\lambda)] &= \\ &= (1/N) \|\mathbf{I} - \mathbf{A}(\lambda)\mathbf{f}\|^2 + (\sigma^2/N) \text{tr}\{\mathbf{A}^2(\lambda)\}. \end{aligned} \quad (37)$$

В связи с тем, что для оценки $E[R(\lambda)]$ по (37) требуется знание регрессионной функции $f(\mathbf{x})$ (которую требуется восстановить средствами НС), Кравеном и Вахбой (Craven, Wahba) [20] введено определение следующей аппроксимации $\hat{R}(\lambda)$:

$$\begin{aligned} \hat{R}(\lambda) &= (1/N) \|\mathbf{I} - \mathbf{A}(\lambda)\mathbf{y}\|^2 + \\ &+ (\sigma^2/N) \text{tr}[\mathbf{A}^2(\lambda)] - (\sigma^2/N) \text{tr}[(\mathbf{I} - \mathbf{A}(\lambda))^2]. \end{aligned} \quad (38)$$

$\hat{R}(\lambda)$ является несмещенной оценкой; $E\{\hat{R}(\lambda)\} = E\{R(\lambda)\}$; минимизация $\hat{R}(\lambda)$ дает хорошую оценку параметра регуляризации λ .

Для практики применения сетей с СПФН по методу регуляризации важным является предложенный в [20] метод перекрестной проверки-подтверждения (ППП). На основе этого метода осуществляется процедура оценки неизвестной дисперсии шума σ^2 , которая необходима для определения $\hat{R}(\lambda)$ с помощью выражения (38). Метод ППП имеет две разновидности — простую и расширенную.

• В простой форме метода при минимизации функционала $E(F) = \sum_{i=1}^N [y_i - F(\mathbf{x}_i)]^2 + (\lambda/2) \|DF(\mathbf{x}_i)\|^2$ определяется $F_\lambda^{[k]}$, соответствующая $\min_F E^{[k]}$, где $E^{[k]} = \sum_{i=1, i \neq k}^N [y_i - F(\mathbf{x}_i)]^2 + (\lambda/2) \|DF(\mathbf{x}_i)\|^2$, т. е. для функционала $E^{[k]}$, в котором исключена обучающая пара (\mathbf{x}_k, y_k) . Способность $F_\lambda^{[k]}$ предсказывать все точки данных оценивается как средняя ошибка их предсказания

$$V_0(\lambda) = (1/N) \sum_{k=1}^N [y_k - F_\lambda^{[k]}(\mathbf{x}_k)]^2. \quad (39)$$

Это выражение зависит только от данных. Процедуры получения функции V_0 и способ оцен-

¹²⁾ Необходимые для этого преобразования основаны на том, что квадрат нормы вектора равен скалярному произведению $\|\mathbf{f}\|^2 = (\mathbf{f}, \mathbf{f})$; норма $\|(\mathbf{I} - \mathbf{A}(\lambda))\mathbf{f}\|^2$ не является случайной и оператор E на нее не влияет. $E\{\boldsymbol{\varepsilon}^T \mathbf{A}(\lambda) (\mathbf{I} - \mathbf{A}(\lambda)) \mathbf{f}\} = 0$, т. к. $E\{\boldsymbol{\varepsilon}^T\} = 0$. По правилам матричной алгебры $E\{\|\mathbf{A}(\lambda)\boldsymbol{\varepsilon}\|^2\} = E\{\text{tr}[\mathbf{A}^T(\lambda)\mathbf{A}(\lambda)\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T]\} = \sigma^2 \text{tr}[\mathbf{A}^T(\lambda)\mathbf{A}(\lambda)] = \sigma^2 \text{tr}[\mathbf{A}^2(\lambda)]$; поскольку $\text{tr}[\mathbf{BC}] = \text{tr}[\mathbf{CB}]$, $\text{tr}[\mathbf{A}^T(\lambda)\mathbf{A}(\lambda)] = \text{tr}[\mathbf{A}^2(\lambda)]$.

ки λ , основанный на минимизации V_0 , составляют обычную форму метода ППП.

• Расширенная форма метода для получения $V_0(\lambda)$ использует три положения.

1) Если точку данных y_k заменить ее предсказателем $F_\lambda^{[k]}(\mathbf{x}_k)$, то минимизация регуляризованного функционала $E(F) = (1/2) \times \sum_{i=1}^N [y_i - F(\mathbf{x}_i)]^2 + (\lambda/2) \|DF(\mathbf{x}_i)\|^2$ с использованием данных $\{y_1, y_2, \dots, y_{k-1}, y_k, y_{k+1}, \dots, y_N\}$ в качестве решения дает $F_\lambda^{[k]}(\mathbf{x}_k)$.

2) Для каждого входного вектора \mathbf{x}_k функция $F_\lambda^{[k]}(\mathbf{x})$, минимизирующая функционал $F_\lambda^{[k]}(\mathbf{x})$, линейно зависит от y_k .

3) Из положений 1) и 2) следует справедливость соотношения $F_\lambda^{[k]}(\mathbf{x}_k) = F_\lambda(\mathbf{x}_k) + (F_\lambda^{[k]}(\mathbf{x}_k) - y_k) \frac{\partial F_\lambda(\mathbf{x}_k)}{\partial y_k}$, а учитывая, что

$$\frac{\partial F_\lambda(\mathbf{x}_k)}{\partial y_k} = a_{kk} \quad (\text{поскольку } F_\lambda(\mathbf{x}_k) = \sum_{i=1}^N a_{ik}(\lambda)y_i),$$

это соотношение можно решить относительно $F_\lambda^{[k]}(\mathbf{x}_k)$ и представить эту функцию в виде (40):

$$F_\lambda^{[k]}(\mathbf{x}_k) = \frac{F_\lambda(\mathbf{x}_k) - a_{kk}(\lambda)y_k}{1 - a_{kk}(\lambda)} = \frac{F_\lambda(\mathbf{x}_k) - y_k}{1 - a_{kk}(\lambda)} + y_k. \quad (40)$$

Подстановка (40) в (39) дает новую форму $V_0(\lambda)$:

$$V_0(\lambda) = (1/N) \sum_{k=1}^N \left\{ \frac{F_\lambda(\mathbf{x}_k) - y_k}{1 - a_{kk}(\lambda)} \right\}^2. \quad (41)$$

Чтобы уменьшить влияние различия множителей $(1 - a_{kk}(\lambda))^{-1}$ на взвешивание ошибок аппроксимации $(F_\lambda(\mathbf{x}_k) - y_k)$ в выражении (41) для $V_0(\lambda)$, расширенный метод ППП использует процедуру вращения координат. При этом модифицированная функция $V(\lambda)$ отличается от обычной формы (41) введением весовых коэффициентов ω_k :

$$V(\lambda) = (1/N) \sum_{k=1}^N \omega_k \left\{ \frac{F_\lambda(\mathbf{x}_k) - y_k}{1 - a_{kk}(\lambda)} \right\}^2, \quad (42)$$

$$\omega_k = \left\{ \frac{1 - a_{kk}(\lambda)}{(1/N) \text{tr}[\mathbf{I} - \mathbf{A}(\lambda)]} \right\}^2.$$

Таким образом, после преобразования числителя в (42) с использованием соотношения

$F_\lambda(\mathbf{x}_k) = \sum_{i=1}^N a_{ik}(\lambda)y_i$ расширенный метод ППП предлагает функцию $V(\lambda)$ (минимизация которой определяет величину параметра регуляризации λ) в виде

$$V(\lambda) = \frac{(1/N) \|\mathbf{I} - \mathbf{A}(\lambda)\mathbf{y}\|^2}{\{(1/N) \text{tr}[\mathbf{I} - \mathbf{A}(\lambda)]\}^2}. \quad (43)$$

Согласно [20], при больших N среднеквадратичная ошибка $R(\lambda)$ при значении λ , полученном по расширенному методу ППП, приближается к минимально возможному значению.

ЗАКЛЮЧЕНИЕ

Представлены элементы теории и аналитические процедуры, составляющие методологию сетей с (сферически) симметричной функцией преобразования нейронов (СФПН). Особенность структуры таких нейронных сетей (НС) состоит в использовании двух принципов:

- локализации элементов скрытого слоя (нейронов) в многомерном пространстве, размерность которого соответствует размерности входной информации (размерности аргумента аппроксимируемой функции, распознаваемых образов, объектов с отслеживаемой или управляемой по модели траекторией);

- зависимости СФПН от (нормы) разности входного вектор-сигнала и вектора локализации, фиксирующего каждый нейрон скрытого слоя НС в пространстве параметров анализируемой информации.

Специфика методологии сетей с СФПН определяется некоторой сложностью аналитических преобразований, составляющих ее основу и относящихся к (нетрадиционным для прикладных методов обработки сигналов в ИИС) разделам функционального анализа, математической физики и теории решения некорректно поставленных (как правило, обратных) задач.

Краткая сводка процедур и преобразований, связанных с основами методологии включает:

- Введение нелинейных преобразований пространства параметров входного вектора НС, улучшающих разделимость объектов при распознавании образов — получение линейной разделимости на классы с помощью (гипер-)плоскостей вместо сложных гиперповерхностей. Этот принцип имеет приложение также при преобразовании нелинейной фильтрации в более простую задачу линейной фильтрации.

- Анализ задачи интерполирования с использованием сети со сферически СФПН и особенности алгоритма обучения, связанные с переобученностью НС.

- Элементы процедуры регуляризации по методу А.Н. Тихонова для плохо обусловленной задачи.

- Схему применения метода регуляризации при получении корректного решения задач для сетей с СФПН. Анализируемый подход состоит в расширении критерия (функционала) качества обучения НС путем введения аддитивной компоненты в виде регуляризирующего функционала (РФ). РФ представляет квадрат нормы преобразованной дифференциальной оператором (ДО) аппроксимируемой функции. При этом ДО и сам РФ отражает априорные представления о гладкости восстанавливаемой функциональной зависимости, а специально вводимый множитель при РФ (называемый параметром регуляризации) определяет относительную значимость показателя ошибок НС и требования к гладкости решения — функции, определяемой сетью после обучения.

- Аналитические преобразования, связанные с получением решения для НС с изотропной формой СФПН в виде гауссиана, включающие минимизацию критерия общего вида (с РФ) и использующие процедуры для дифференциала Фреше, определение функций Грина (ФГ) и представление аппроксимируемой функции в виде разложения по ФГ.

- Метод получения решений для сетей с изотропными СФПН в виде многомерных гауссианов, показанный как в основной, так и в расширенной формах, последняя из которых использует НС с существенно меньшим числом нейронов.

- Рассмотрение подхода к определению величины параметра регуляризации, использующего критерий среднеквадратичной ошибки НС, и метода перекрестной проверки-подтверждения (ППП). Метод PPP описан в простом и расширенном вариантах.

Основные идеи, методы и аналитические процедуры, описанные выше, разработаны Ковером, Микчели, Квоком и Йонгом, А.Н. Тихоновым, Ченом и Чангом, Хэйкиным, Поггио, Кравеном и Вахбой [1, 3, 5, 7, 11, 15, 19, 20].

СПИСОК ЛИТЕРАТУРЫ

1. Cover T.M. Geometrical and statistical properties of systems of linear inequalities to pattern recognitions // IEEE Transactions on Electronic Computers. 1965. V. E14. P. 326–334.
2. Химмельблау Д. Прикладное нелинейное программирование. М.: Мир, 1975. 534 с.
3. Micchelli C.A. Interpolation of scattered data: distance matrices and conditionally positive-definite functions // Constructive Approximation. 1986. V. 2. P. 11–22.
4. Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. М.: Наука, 1984. 318 с.
5. Kwok T.K., Young D.-Y. Objective functions for training new hidden units in constructive neural networks // IEEE Transactions on Neural Networks. 1997. V. 8, N 8. P. 1131–1148.
6. Малыхина Г.Ф., Меркушева А.В. Элементы статистической концепции обучения нейронной сети и прогнозирование точности ее функционирования // Научное приборостроение. 2005. Т. 15, № 1. С. 29–45.
7. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1979. 280 с.
8. Kirsch A. Introduction to the mathematical theory of inverse problems. N. Y.: Springer-Verlag, 1996. 420 p.
9. Lagaris I.E., Likas A. Artificial neural networks for solving ordinary and partial differential equations // IEEE Transactions on Neural Networks. 1998. V. 9, N 5. P. 897–1000.
10. De Figueiredo R.J.P. Nonlinear feedback control systems. N. Y.: Academic Press, 1993. 480 p.
11. Chen S., Chang E.S. Regularized orthogonal least square algorithm for constructing RBF neural networks // International Journal of Control. 1996. V. 64, N 5. P. 829–937.
12. Бронштейн И.Н., Семендяев К.А. Справочник по математике. М.: Наука, 1986. 544 с.
13. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М.: Наука, 1966. 120 с.
14. Треногин В.А. Функциональный анализ. М.: Наука, 1980. 350 с.
15. Haykin S. Neural networks. Upper Saddle River, New Jersey: Prentice Hall, 1994. 680 p.
16. Джеффрис Т., Свирлс Б. Методы математической физики. М.: Мир, 1970. Вып. 1–3. 860 с.
17. Курант Р. Уравнения с частными производными. М.: Мир, 1964. 396 с.
18. Камке Э. Справочник по теории уравнений в частных производных. М.: Наука, 1976. 484 с.
19. Poggio T., Girosi F. Networks for approximation and learning // Proceedings of IEEE. 1990. V. 78. P. 1481–1497.
20. Craven P., Wahba G. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of cross-validation // Numerische Mathematik. 1979. V. 31. P. 377–403.

Санкт-Петербург

Материал поступил в редакцию 7.12.2005.

METHODOLOGY OF NETWORKS WITH SYMMETRICAL TRANSFORMATION FUNCTIONS FOR NEURONS

A. V. Merkusheva, G. F. Malychina

Saint-Petersburg

The basic methodology of networks with symmetrical transformation functions for neurons (STFN) is presented. Such neural networks (NN) find application in the problems of approximation, pattern recognition, systems (objects) identification, controller designing, noise level lowering for signals in information-measurement systems. The characteristic property of NN structure is the localization of hidden-layer elements in the multidimensional vector space (whose dimension is identical to the input information dimension) and the presence of STFN depending on the (metric) norm of the difference between the hidden-layer element localization vectors and the input signal-vector. The paper presents the applied theory elements for pattern recognition using a network with STFN; NN learning criteria on the basis of a functional regularized with the aid of A.N. Tikhonov's method; a general form of approximation and interpolation functions (obtained on the basis of those criteria) using Green's scheme for the inverse problem generated by transformation with a differential operator; a method for regularization parameter selection.