УДК 004.052:004.738.2 (045)

Толстикова Е.В., к.т.н.

ОПТИМАЛЬНАЯ НАСТРОЙК ПАРАМЕТРОВ И СТРУКТУРЫ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Национальный авиационный университет

Рассмотрена задача оптимальной настройки параметров искусственной нейронной сети произвольной структуры. Показано, что для широкого класса искусственных нейронных сетей можно применять модифицированные градиентные методы поиска оптимальных параметров для работы сети в условиях наличия ошибок измерения и внешних помех.

Введение

Нейронные сети (НС) и, в частности, искусственные НС (ИНС) представляются универсальным средством эффективного построения моделей практически любых нелинейных структур [1]. С их помощью можно решать задачи распознавания образов, оптимизации, идентификации, управления динамическими объектами. Они строятся по принципам нейронных структур живой природы [2, 3].

Искусственные нейронные сети состоят из элементов, функциональные возможности которых аналогичны большинству элементарных функций биологического нейрона и демонстрируют большое число свойств, присущих мозгу. Например, они обучаются на основе опыта, обобщают предыдущие прецеденты на новые случаи и извлекают существенные свойства из поступающей информации, содержащей излишние данные. Однако, несмотря на такое функциональное сходство, даже самый оптимистичный их защитник не предположит, что в скором будущем искусственные нейронные сети будут дублировать функции человеческого мозга. Реальный «интеллект», демонстрируемый самыми сложными нейронными сетями, находится ниже уровня интеллекта дождевого червя, и энтузиазм должен быть умерен в соответствии с современными реалиями [2].

Постановка задачи

Известные применения ИНС – это распознавание образов и классификация признаков объектов, принятие решений и управление, кластеризация – разбиение множества входных сигналов на классы, аппроксимация и прогнозирование, сжатие данных и ассоциативная память и др.

Перечисленные задачи весьма близки друг к другу как по принципиальным особенностям, так и по методам решения. Основные этапы решения этих задач следующие:

– сбор данных для обучения;
– подготовка и нормализация данных;
– выбор топологии сети;
– экспериментальный подбор характеристик сети;
– экспериментальный подбор параметров обучения;
– собственно обучение;
– проверка адекватности обучения;
– корректировка параметров, окончательное обучение;
– вербализация (словесное описание) сети с целью дальнейшего использования.

Как видно из перечня этапов, процесс решения перечисленных задач с помощью ИНС основан на итерационном подборе некоторых параметров, доставляющих экстремум функции качества, которая, как правило, не обладает свойствами непрерывности и гладкости. Поэтому существенным недостатком такого подхода является не только невозможность обеспечения каких бы то ни было гарантий оптимальности применяемых методов и алгоритмов, но даже отсутствие предпосылок получения хотя бы асимптотических оценок сходимости к оптимальному решению.

В данной работе сделана попытка применения регулярных статистических методов для построения ИНС с оптимальной настройкой параметров для некоторых классов задач, решаемых с помощью ИНС.

Структуры ИНС могут быть самыми разными, однако основными элементами сети являются формальные аналоги нейрона – адалина Уидроу или персептрон Розенблатта [4, 5] – нелинейные элементы с симметричной или несимметричной относительно оси ординат амплитудной характеристикой активационной функции. Структурно эти элементы идентичны, и единственное отличие заключается в характере нелинейности характеристики активационной функции.

Для преодоления априорной неопределенности [6] и синтеза ИНС необходимо, прежде всего, определить условия оптимальности. В многочисленной литературе по нейронным сетям [2-5 и др.] данная проблема даже не затрагивается. Единственным исключением является работа [7], в которой рассмотрены условия оптимальности и синтезированы алгоритмы настройки нейронной сети, т.е. решена задача статистического синтеза ИНС как информационной системы.

Приведем классификацию нейронных сетей по их ключевым показателям.

1. Классификация по типу входной информации: аналоговые НС (используют информацию в форме действительных чисел)или цифровые НС (оперируют с информацией, представленной в двоичном виде).

2. Классификация по характеру обучения:

– обучение с учителем – выходное пространство решений нейронной сети известно априори;

– обучение без учителя – нейронная сеть с самоорганизацией, которая формирует выходное пространство решений только на основе входных воздействий;

– обучение с подкреплением – система назначения штрафов и поощрений в зависимости от результата.

3. Классификация по характеру настройки – сети с фиксированными или динамическими связями.

4. Классификация по времени передачи сигнала – синхронные или асинхронные сети.

5. Классификация по характеру связей: НС прямого распространения, рекуррентные НС.

Элементы (адалина или персептрон), образующие нейронную сеть, вообще говоря, могут быть соединены самым произвольным образом, образуя практически неограниченное число возможных структур. Как известно, наиболее широко распространенными являются слоистые структуры, в которых сигналы передаются от слоя с меньшим номером к ближайшему слою с большим номером, т.е. слои в такой структуре соединены последовательно. Входом первого слоя является внешнее воздействие, а выход последнего слоя является выходом всей НС. Далее для определенности будем рассматривать персептрон в качестве базового элемента НС.

Пусть число слоев НС равняется K. Число персептронов в k-м слое, $k=\overline{1,K}$, полагаем равным M_k. Размерность вектора входных сигналов $\mathbf{X}_{k-1}(m)$для k-го слоя будет равна M_{k-1}, а размерность вектора входных сигналов $\mathbf{X}_k(m)$, соответственно, M_k. Соотношение между $\mathbf{X}_{k-1}(m)$ и $\mathbf{X}_k(m)$ имеет вид:

$$\mathbf{X}_k(m)=\Phi\left[\mathbf{W}_k^T\mathbf{X}_{k-1}(m)\right], \tag{1}$$

где $\mathbf{W}_k=\left[\mathbf{w}_{k1} \quad \mathbf{w}_{k2} \quad \cdots \quad \mathbf{w}_{k,Mk}\right]$ – матрица размерностью $\left(M_{k-1}\times M_k\right)$ весовых коэффициентов; T – символ транспонирования. Столбцами матрицы $\mathbf{W}_k$ являются векторы $\mathbf{w}_{km}$ весовых коэффициентов m- го персептрона из k-го слоя. Размерность вектора $\mathbf{w}_{km}$ равна M_{k-1}. Оператор Φ есть оператор нелинейного преобразования вектора $\mathbf{W}_k^T\mathbf{X}_{k-1}(m)$ в соответствии с выбранной (симметричной или несимметричной) активационной функцией.

Рассмотрим дискретную динамическую K-слойную НС, построенную в соответствии с алгоритмом (1). Структурная схема k-го слоя сети изображена на рис. 1.

Весовые коэффициенты k-го слоя НС определяются вектором весовых коэффициентов

$$\mathbf{w}_k=\left[w_{k1},w_{k2},\cdots,w_{k,Mk}\right],\quad k=\overline{1,K}.$$

Применим соотношение (1) для рекуррентного вычисления выходного сигнала всей НС:

$$\mathbf{X}_K(m)=\Phi\left[\mathbf{W}_K^T\Phi\left[\mathbf{W}_{K-1}^T\ldots\Phi\left[\mathbf{X}_0(m)\right]\ldots\right]\right]. \tag{2}$$

Здесь $\mathbf{X}_0(m)$ – вектор входных сигналов сети, имеющий размерность M_0. Уравнением (2), по существу, определяется передаточная функция НС.

Рассмотрим теперь условия оптимальности нейронной сети как модели сложной нелинейной системы, работающей в условиях внутренних и внешних помех.

Оптимальный алгоритм настройки параметров нейронной сети

В реальных нелинейных системах присутствуют как внутренние ошибки измерения, так и внешние шумы и помехи. В этом случае связь между входным сигнальным вектором $\mathbf{X}_0(m)$ и выходным вектором $\mathbf{X}_K(m)$ описывается уравнением:

$$\mathbf{X}_K(m) = \Xi\left[\mathbf{W}_K^T \Xi\left[\mathbf{W}_{K-1}^T \ldots \Xi\left[\mathbf{X}_0(m)\right]\ldots\right]\right] + \mathbf{v}(m) + \zeta(m), \tag{3}$$

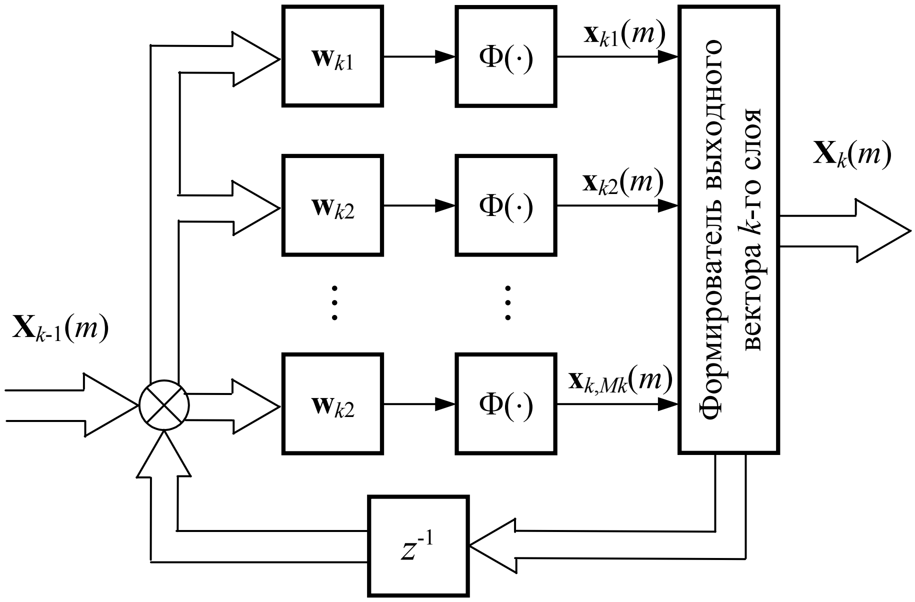

Рис. 1. Структурная схема k -го слоя сети

где $\mathbf{v}(m)$ и $\zeta(m)$ – векторы ошибок измерений и внешних шумов соответственно. Вектор $\mathbf{v}(m)$ имеет размерность M_K, а вектор $\zeta(m)$ – размерность M_0. Логично предположить, что ошибки измерения и внешние шумы представляют собой последовательности взаимно некоррелированных случайных величин с нулевыми математическими ожиданиями. Обозначим плотность распределения вектора $\mathbf{v}(m)$ через $p_\nu(x)$, а плотность распределения вектора $\zeta(m)$ – через $p_\zeta(x)$. Без потери общности можно сделать допущение о принадлежности распределений $p_\nu(x)$ и $p_\zeta(x)$ к гауссовскому семейству.

Допустим, что существует одна и только одна матрица весовых коэффициентов НС $\widetilde{\mathrm{W}}_k = \left[\widetilde{\mathrm{w}}_{k1}\ \widetilde{\mathrm{w}}_{k2} \ldots \widetilde{\mathrm{w}}_{k,Mk}\right]$, такая, что оператор нелинейного преобразования $\Xi\left[\mathbf{X}_0(m)\right]$ однозначно воспроизводит выходной вектор $\mathbf{X}_K(m)$:

$$\Xi\left[\mathbf{X}_0(m)\right] = \Phi\left[\widetilde{\mathrm{W}}_K^T \Phi\left[\widetilde{\mathrm{W}}_{K-1}^T \ldots \Phi\left[\mathrm{X}_0(m)\right]\ldots\right]\right] \tag{4}$$

При наличии априорной информации о семействе распределений ошибок измерения и внешних помех можно построить оптимальный алгоритм настройки весовых коэффициентов НС, при которых средние потери будут минимальными:

$$G(\mathbf{w}) = \mathrm{E}\left\{C\left[\varepsilon\left(\mathbf{X}_0(m), \mathbf{w}\right)\right]\right\} \underset{\mathbf{w}}{\rightarrow} \min, \tag{5}$$

где $C(\varepsilon)$ – функция потерь;

$$\varepsilon\left(\mathbf{X}_0(m), \mathbf{w}\right) = \mathbf{X}_K(m) - \Xi\left[\mathbf{W}_K^T \Xi\left[\mathbf{W}_{K-1}^T \ldots \Xi\left[\mathbf{W}_1^T \mathbf{X}_0(m)\right]\ldots\right]\right] \tag{6}$$

– вектор невязки размерности M_k.

В качестве функции потерь, которую интерпретируем как меру неопределенности относительно параметра, используем информационную функцию потерь вида [8]:

$$C(\lambda, s) = -\ln p(s \mid \lambda) \tag{7}$$

где $p(s \mid \lambda)$ – условная плотность вероятности параметра s, если принято решение с оценкой λ. Если, как в рассматриваемом случае, условная плотность вероятности описывается гауссовской кривой или другой четной функцией относительно некоторой фиксированной точки s^*, функция вида (7) является симметричной функцией разности $|\lambda - s|$. Отметим, что гауссовская плотность вероятности имеет максимальную энтропию на всей прямой $[-\infty, \infty]$ в классе распределений с фиксированными математическим ожиданием и дисперсией. В этом смысле она представляет наименее благоприятную плотность распределения из данного класса. Условие существования оптимального решения $w = \tilde{w}$ имеет вид:

$$\begin{aligned}\nabla G(\mathbf{w}) &= \mathrm{E}\left\{\nabla C\left[\varepsilon\left(\mathbf{X}_0(m), \mathbf{w}\right)\right]\right\} = \\ &= \mathrm{E}\left\{\nabla \Xi^T\left[\mathbf{W}_K^T \Xi\left[\mathbf{W}_{K-1}^T \ldots \Xi\left[\mathbf{W}_1^T \mathbf{X}_0(m)\right]\ldots\right]\right]\nabla C\left[\varepsilon\left(\mathbf{X}_0(m), \mathbf{w}\right)\right]\right\} = 0.\end{aligned} \quad (8)$$

Из условия (8) следует, что для определения оптимального решения $\tilde{w}$ необходимо вычислять либо градиент функционала $\nabla C\left[\varepsilon\left(\mathbf{X}_0(m), \mathbf{w}\right)\right]$, либо градиент выходного сигнального вектора $\nabla \Xi^T(\cdot)$ и градиент функционала по вектору невязки $\varepsilon\left(\mathbf{X}_0(m), \mathbf{w}\right)$.

Вычисление $\nabla C\left[\varepsilon\left(\mathbf{X}_0(m), \mathbf{w}\right)\right]$ можно выполнять методом обратного пересчета или так называемого обратного распространения ошибки [9]. Суть метода заключается в том, что общий градиент функционала по вектору весовых коэффициентов НС рассматривается как последовательность частных градиентов по матрице $\mathbf{W}_k$ весовых коэффициентов слоев от K-го до первого. Результаты расчетов градиента по матрице весов текущего слоя используются в качестве исходных данных для вычисления градиента по матрице весов предыдущего слоя. По существу, задача оптимальной настройки НС сводится к задаче типа динамического программирования.

Выводы

В работе рассмотрена задача настройки НС как задача идентификации нелинейной системы. Благодаря введению информационной функции потерь можно получить асимптотически оптимальные алгоритмы настройки, обладающие хорошей сходимостью и устойчивостью. Априорная информация об ошибках измерения и внешних шумах дает возможность однозначно задать функцию потерь и добиться максимально возможной скорости сходимости. При этом появляется возможность замены градиента средних потерь на псевдоградиент и достаточно простой модификации алгоритмов настройки, например, для дискретных нелинейных систем.

Список литературы

1. Нейрокомпьютеры и их применение: Нейронные сети: история развития теории. Кн. 5: Учеб. пособие для вузов. / Под общей ред. А.И. Галушкина, Я.З. Цыпкина. – М.: ИПРЖР, 2001. – 840 с.

2. Уоссермен Ф. Нейрокомпьютерная техника, М., Мир, 1992. – 184 с.

3. Хайкин С. Нейронные сети: Полный курс. 2-е изд., испр. – М.: Издательский дом Вильямс, 2008, – 1103 с.

4. Калан Р. Основные концепции нейронных сетей: Пер. с англ. – М.: Издательский дом "Вильямс", 2001. – 287 с.

5. Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2002. – 344 с.

6. Репин В.Г., Тартаковский Г.П. Статистический синтез при априорной неопределенности и адаптация информационных систем. – М.: Сов. радио, 1977. – 432 с.

7. Цыпкин Я. З. Информационная теория идентификации. – М.: Наука. Физматлит, 1995. – 336 с.

8. Куликов Е.И., Трифонов А.П. Оценка параметров сигнала на фоне помех. – М.: Сов. радио, 1978. – 296 с.

9. Галушкин А.И. Синтез многослойных систем распознавания образов [Текст]: монография / А.И. Галушкин. – М.: Энергия, 1974. – 368 с.