УДК 681.3

МУЛЬТИСТАРТОВЫЙ СУБГРАДИЕНТНЫЙ МЕТОД ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ В ПРОСТРАНСТВЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Щербакова Г.Ю., Крылов В.Н.
Одесский национальный политехнический университет, г. Одесса
yuri4@mail.ru victor_krylov@inbox.ru

Abstract

Shcherbakova G. Yu., Krylov V. N. Multi-start sub gradient method for neural networks learning in the wavelet transformed domain*. Sub gradient iterated method for neural networks learning with back-propagation algorithm in the wavelet transformed domain is proposed. This method allows noise stability raising, error, local extreme and initial point search sensitiveness reducing.*

Введение

Искусственные нейронные сети широко применяются в задачах прогнозирования, аппроксимации, управления. Важной областью их применения являются кластеризация и классификация при распознавании образов [1].

При выборе метода обучения нейронной сети определяют критерий и (или) функционал качества, метод его оптимизации, алгоритм и режим обучения нейронной сети (последовательный или пакетный).

Поскольку в условиях помех при малом объеме обучающих выборок функционал качества обучения нейронной сети может иметь многоэкстремальную, зашумленную поверхность, выбор метода поиска оптимума во многом определяет достоверность классификации при распознавании.

Существующие методы оптимизации функционала качества при обучении нейронной сети можно разделить на две категории. К первой категории относятся методы глобальной оптимизации. Они используют перебор значений переменных, от которых зависит целевая функция. Ко второй категории относятся методы, позволяющие найти локальный оптимум. При использовании методов перебора с ростом размерности решаемой задачи значительно растет и время обучения. Методам локальной оптимизации, применяемым при обучении нейронных сетей, также присущ ряд недостатков. При стохастической оптимизации они могут требовать большого числа итераций. Регулярные итеративные методы оптимизации, наряду с высокой точностью, могут обладать низкой помехоустойчивостью (в зависимости от особенностей оценки градиента). Обе группы методов чувствительны к локальным экстремумам и к начальной точке поиска.

Задачу повышения устойчивости к воздействию помех у регулярных итеративных методов оптимизации решают, применяя субградиентные методы оптимизации. Однако этим методам присуща низкая точность. Предлагается повысить помехоустойчивость и снизить чувствительность к локальным экстремумам и начальной точке поиска, применяя для оптимизации функционала качества при обучении нейронной сети иерархический субградиентный итеративный метод оптимизации в пространстве вейвлет-преобразования, в дальнейшем это метод называется мультистартовым [2,3].

Целью данной работы является разработка метода обучения нейронной сети, обладающего высокой помехоустойчивостью и точностью при низкой чувствительности к локальным экстремумам.

Для достижения этой цели решены задачи:

-анализа современных методов оптимизации при обучении нейронных сетей;

-разработки мультистартового субградиентного метода обучения нейронных сетей в пространстве вейвлет-преобразования (ВП);

-проведены экспериментальные исследования для оценки повышения скорости сходимости, повышения помехоустойчивости, снижения чувствительности к локальным экстремумам этого метода при обучении многослойной нейронной сети с обратным распространением ошибки.

Методы оптимизации при обучении нейронных сетей

Для решения задачи оптимизации при обучении нейронных сетей могут использоваться следующие методы:

- локальной оптимизации (регулярные итеративные) с вычислением оценок градиента (частных производных первого порядка) (градиентный метод (наискорейшего спуска); метод сопряженных градиентов; одношаговые и многошаговые методы с поиском оптимума целевой функции в направлении антиградиента);

- локальной оптимизации с вычислением частных производных первого и второго порядка (метод Ньютона; метод Гаусса-Ньютона; методы оптимизации с разреженными матрицами Гессе; квазиньютоновские методы; метод Левенберга-Марквардта);

- стохастической оптимизации (используют поиск в случайном направлении; имитация отжига; метод Монте-Карло (численный метод статистических испытаний)):

- глобальной оптимизации (используют перебор значений переменных, от которых зависит целевая функция) [4].

Этим методам присущи различные недостатки.

Стохастические алгоритмы требуют большого числа шагов обучения. В алгоритмах глобальной оптимизации при отсутствии априорной информации о характере целевой функции экспоненциально растет сложность перебора с ростом размерности решаемой задачи [4]. Метод сопряженных градиентов очень чувствителен к точности вычислений. У методов, учитывающих направление антиградиента на нескольких шагах алгоритма, и методов, включающих вычисление матрицы Гессе, также растет количество дополнительных переменных, необходимых для организации вычислений. Это усложняет их использование для обучения нейронных сетей больших размерностей.

Поэтому для поиска оптимума функционала качества в многослойных сетях используются регулярные итеративные методы оптимизации с вычислением оценок градиента первого порядка [4]. Они обладают высокой точностью, но (в связи с особенностями оценки градиента) низкой помехоустойчивостью и высокой чувствительностью к локальным экстремумам и начальной точке поиска. Эти проблемы можно преодолеть с помощью субградиентных методов оптимизации с высокой помехоустойчивостью [5]. Но погрешность этих методов высока. Для уменьшения влияния помех в субградиентных методах применяют подход, использующий вейвлет преобразование [5,6].

Целевая функция при оптимизации, как правило, пространственно неоднородна. Адекватным аппаратом анализа таких функций является вейвлет преобразование [6,7].

Однако разработанный в работе [5] метод поиска оптимума с оценкой градиента на основе гиперболического вейвлет преобразования имеет высокую погрешность.

Для снижения влияния указанных выше недостатков в этой работе предлагается применить метод обучения нейронной сети на основе мультистартового субградиентного метода оптимизации в пространстве вейвлет преобразования.

Метод оптимизации при обучении нейронной сети

Итеративный алгоритм в пространстве гиперболического вейвлет преобразования (ГВП) имеет вид:

$$x[k+1] = x[k] + \gamma HWT(x[k]), \qquad (1)$$

где $HWT(x[k])$ – оценка субградиента в точке $x[k]$, γ -шаг.

В качестве функционала качества обучения используется /1/:

$$E(n) = \frac{1}{2}\sum_{j\in M} e_j^2(n), \qquad (2)$$

где M - множество нейронов выходного слоя; $e_j(n) = d_j(n) - y_j(n)$ - ошибка нейрона j для примера n; $d_j(n)$ - желаемый отклик нейрона j; $y_j(n)$ - сигнал на выходе нейрона j на итерации n.

На первом этапе оптимизации при обучении сети для оценки субградиента используется взвешенная сумма значений минимизируемой функции с вейвлет-функцией Хаара в окрестности, определяемой длиной носителя.

Такой поиск позволяет достичь окрестности глобального экстремума с погрешностью, определяемой асимметрией целевой функции.

Для снижения погрешности на втором этапе оптимизации для оценки субградиента используется взвешенная сумма с гиперболической функцией $\Psi(i) = \dfrac{1}{\alpha x}$, регуляризованной по лифтинговой схеме /8/. В качестве начальной точки используется результат предыдущего этапа.

Вычисляют гиперболическую функцию $\Psi(i)$ для $i = \overline{1,N}$, и минимизируемая функция $J[x]$ взвешивается с функцией $\Psi(i)$

$$HWT(x) = J[x] * \Psi(i), \qquad (3)$$

где $*$ – операция взвешенного суммирования.

Если найденная на этом этапе координата минимума отличается от результата предыдущего этапа не более чем на δ (заданное значение точности поиска координат оптимума), процесс поиска оптимума заканчивается. Если координата минимума отличается от результата предыдущего этапа более чем на δ, масштаб α увеличивается на 1, пока не выполнится условие окончания поиска оптимума. При таком поиске последовательно переходят от оптимизации с помощью вейвлета Хаара, способного обеспечить высокую помехоустойчивость, к оптимизации с помощью дифференциатора, способного дать максимальную точность (так как при $\alpha \to \infty$ взвешивающая функция $\frac{1}{\alpha x}$ стремится к дифференциатору).

Процедура обучения нейронной сети

Обучение методом обратного распространения ошибки с использованием мультистартового субградиентного метода оптимизации в пространстве вейвлет-преобразования рассматривается на примере последовательного обучения. Обучение предполагает два прохода по слоям сети: прямой проход и обратный проход. Исходные данные: начальные значения весовых коэффициентов c_{ij}; точность обучения ε; длина носителя L_1 и шаг вейвлет-функции Хаара и гиперболической функции dx_1.

1.Прямой проход

1.1. Для узлов всех слоев сети, предшествующих выходному, определяются весовые суммы I_j

$$I_j(n) = \sum_{k=1}^{N_K} c_{jk} O_k , \qquad (4)$$

где I_j - сигнал на входе активирующего элемента каждого узла слоя J для $j = 1,2,\ldots,N_j$; N_j - число узлов в слое J; N_R - число узлов в слое K, предшествующем слою J; c_{jk} - веса, модифицирующие выходные сигналы O_k узлов слоя K.

1.2. Определяются активаторы для этих узлов

$$h_j(I_j) = \frac{1}{1 + e^{-(\sum_{k=1}^{N_k} c_{jk} O_k)/\Theta_0}} , \qquad (5)$$

где $h_j(I_j)$ - сигмоидальная функция активации; параметр крутизны сигмоидальной функции $\Theta_0 = 1$.

1.3. Для всех весовых коэффициентов c_{jk} и смещений узлов выходного слоя:

- рассчитывают $2L_1$ значений активатора j-го узла $h_j(I_j)$ по (5); для одного из коэффициентов определяются значения с приращениями

$$c_{jk} = c_{jk} \pm l\Delta_{cjk} , \qquad (6)$$

где $l = 1,\ldots, L_1$; Δ_{cjk} - приращение;

- рассчитывают $2L_1$ значений целевой функции - ошибки обучения (для $2L_1$ значений c_{jk}).

- рассчитывают взвешенную сумму функции (2) с вейвлетом Хаара, и (при необходимости) с гиперболической функцией;

- с учетом полученной оценки субградиента корректируется c_{jk}.

2. Обратный проход

Синаптические веса сети настраиваются в соответствии с правилом коррекции ошибок [1]; от выхода к входу последовательно корректируются весовые коэффициенты предыдущих слоев согласно п. 1.3.

Повторяются расчеты по пунктам 1.1 - 2, пока не будет выполнено условие останова. Полученное значение используется как начальная точка для следующего этапа поиска.

Формируется новая оценка субградиента путем взвешивания с гиперболической функцией. Проводится итеративная процедура поиска. Сравнивают значения результатов поиска на текущем и предыдущем этапе. Если разность превышает пороговое значение, то увеличивается значение α и осуществляется переход к следующему этапу уточнения координат оптимума, в противном случае – останов.

Экспериментальные исследования

Разработанный метод оптимизации при обучении нейронной сети был проверен экспериментально.

Оценка увеличения скорости сходимости относительно метода градиентного спуска проводилась с помощью функционала качества в виде «оврага Розенброка».

Разработанный метод при обучении нейронной сети позволил достичь оптимума в 1,7 раза быстрее (по количеству итераций) по сравнению с методом градиентного спуска.

Оценка пониженной чувствительности разработанного метода оптимизации к локальным экстремумам и стартовой точке поиска при обучении нейронной сети проводилась с помощью функционала в виде функции Швефеля. Точка старта выбиралась случайным образом. Метод градиентного спуска отыскивал ближайший к стартовой точке минимум. Разработанный метод оптимизации при обучении нейронной сети обеспечивает вероятность обнаружения глобального минимума 0,8.

Оценка помехоустойчивости метода проводилась в два этапа с использованием функционала в виде функции Де Йонга 1

$$f(x) = x^2 \qquad (7)$$

На первом этапе принималось для формулы (7) $x \in (-5{,}12;\ 5{,}12)$, глобальный минимум $f(x) = 0, \quad x = 0$.

При отношении сигнал/шум по амплитуде до 1,05 (помеха распре–делена по нормальному закону с нулевым средним и среднеквадратическим отклонением от 0 до 40000, максимальное значение тестируемой функции $f(x) = 42000$) метод позволяет выйти в область глобального минимума.

На втором этапе оценки помехоустойчивости метода была произведена оценка среднего квадратического отклонения (СКО) для трех методов оптимизации – посредством взвешенного суммирования с вейвлетом Хаара, взвешенного суммирования с гиперболической функцией $\Psi(i)$ и разностного метода.

Оценено среднее квадратическое отклонение (СКО) для сверток с вейвлетом Хаара, с гиперболической функцией $\Psi(i)$ и разностного метода. СКО разностного метода $\sigma(x) = \sqrt{2} \cdot \sigma_i$ (σ_i - СКО оценки приращения функции $J_+(c,a)$ и $J_-(c,a)$). СКО для свертки с вейвлетом Хаара $\sigma_{NHaar}(x) = \frac{\sqrt{2}}{N} \cdot \sigma_i$, СКО для свертки с гиперболической функцией $\sigma_{N\Psi}(x) = \frac{\pi}{\sqrt{3N}} \cdot \sigma_i$ (σ_i - СКО на i шаге). Для оценки помехоустойчивости метода применялся показатель $K = \sigma^*(x) / \sigma(x)$. Здесь $\sigma^*(x)$ - СКО после свертки с вейвлетом Хаара либо гиперболической функцией.

При оценке помехоустойчивости метода на этом этапе выявлено: самая помехоустойчивая– оценка сверткой с вейвлетом Хаара ($\sigma_{NHaar}(x)/\sigma(x)$ при длине носителя 20 возросло более чем в 160 раз). Помехоустойчивость свертки с $\Psi(i)$ возросла до 12. Следовательно, оценка градиента с помощью вейвлет-функции Хаара повышает помехоустойчивость метода оптимизации при обучении нейронной сети.

Метод опробован при классификации реперных знаков в системе автоматизированного позиционирования фотошаблонов (ФШ) интегральных микросхем (ИС) [5, 10].

Современное производство качественных ИС предполагает применение систем автоматизированного оптического контроля (АОК) на ряде этапов, начиная с производства ФШ. Целью системы обработки изображений является выделение и анализ дефектов ФШ посредством сравнения с эталонным ФШ. Для этого должно быть обеспечено позиционирование (ориентация) изображения контролируемого ФШ, то есть определение и коррекция угла наклона, сдвигов и изменения масштаба относительно эталона.

Системы АОК состоят из оптико-электронного блока и системы обработки изображений.

Позиционирование ФШ в современных системах АОК выполняется в оптико-электронном блоке с помощью сложной прецизионной механики и дорогостоящей осветительной аппаратуры. Оптико-электронный блок включает оптическую систему, датчик изображения (видеокамеру), двухкоординатный стол, осветительную систему и блок обработки видеосигнала, осуществляющий усиление, квантование и нормирование видеосигнала.

Система обработки изображений представляет собой проблемно-ориентированный аппаратно-программный комплекс обработки и распознавания изображений. Снизить ресурсоемкость процесса АОК и требования к осветительно-фокусирующей аппаратуре могут позволить более совершенные методы обработки изображений. С этой целью авторами было предложено проводить обнаружение и распознавание реперных знаков (РЗ) при обработке изображений с помощью системы автоматизированного оптического позиционирования ФШ ИС [5, 10].

В этой системе после локализации РЗ на изображении фотошаблона и его контурной обработки выполняется идентификация РЗ. На этапе идентификации необходимо по рассчитанным признакам РЗ определить его вид, то есть провести классификацию.

Для этого был разработан метод классификации РЗ. Классификация РЗ при позиционировании проводилась с помощью многослойной нейронной сети с обратным распространением ошибки, обученной посредством предложенного мультистартового субградиентного метода оптимизации в пространстве вейвлет-преобразования.

Показано, что при применении процедуры классификации с использованием мультистартового субградиентного итерационного алгоритма в пространстве вейвлет-преобразования, средний риск уменьшился от 1,2 до 5 раз при изменении СКО в признаковом пространстве от 1 до 5 по сравнению с методом градиентного спуска.

Время обучения возрастало на порядок.

Подобные результаты получены и при сравнении разработанного метода с методами, основанными на других градиентных алгоритмах.

Выводы

Разработан метод обучения многослойной нейронной сети в последовательном режиме с обратным распространением ошибки, позволяющий повысить помехоустойчивость и точность при низкой чувствительности к локальным экстремумам.

Литература

1. Хайкин С. Нейронные сети: полный курс, 2-е издание: Пер. с англ. – М. Издательский дом «Вильямс», 2006. – 1104 с.
2. Крылов В.Н., Щербакова Г.Ю., Козина Ю.Ю., Волошин В.В. Субградиентный итеративный метод оптимизации в пространстве вейвлет-преобразования // Труды 9 междунар. науч.-практ. конф. "Современные информационные и электронные технологии". – Одесса. – 2008. – С. 62.
3. Крылов В.Н., Щербакова Г.Ю. Субградієнтний ітеративний метод оптимізації в просторі вейвлет-перетворення // Збірник наукових праць Військового інституту Київського національного університету ім. Т. Шевченка. К.:ВІКНУ – 2008. – Вип. № 12. – С.56 – 60.
4. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. – М.: Горячая линия – Телеком, 2002. – 382 с.
5. Крылов В.Н., Щербакова Г.Ю., Козина Ю.Ю., Волошин В.В. Помехоустойчивая классификация реперных знаков в пространстве гиперболического вейвлет-преобразования // Міжнародна наукова конференція «Інтелектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій ISDMIT'2007». - Том 3. – Євпаторія. – 2007. – С.153 – 155.
6.Полякова М.В. Исследование субградиентного поискового метода адаптации в пространстве вейвлет-преобразования // Тр. Одес. политехн. ун-та. — Одесса, 2007.— Вып. 1(27). — С. 207 — 213.
7.Полякова М.В., Крылов В.Н. Характеристика локальной регулярности функций с помощью обобщенных вейвлет-функций // Тр. Одес. политехн. ун-та. — Одесса, 2005.— Вып. 2(24). — С. 192 — 198.
8. Krylov V.N., Polyakova M.V. Contour images segmentation in space of wavelet transform with the use of lifting // *Optical-electronic informatively-power technologies*. – №2 (12), 2007, p. 48 - 58.
9.Крылов В.Н., Щербакова Г.Ю. Субградиентный итеративный метод обучения нейронных сетей в пространстве гиперболического вейвлет-преобразовния // Материалы Международной научно-технической конференции ИИ-2008. 22-27 сентября 2008. пос. Кацивели, АР Крым, Украина. – Том 2. – С.285 – 289.
10. Крылов В.Н., Щербакова Г.Ю., Козина Ю.Ю., Волошин В.В. Помехоустойчивая классификация реперных знаков в пространстве гиперболического вейвлет-преобразования // Системні технології. Регіональний міжвузівський збірник праць. – Вип. 6 (52). – Дніпропетровськ, 2007. – С. 125 – 130.

Поступила в редколлегию 13.03.2009