УДК 364.8.34

Дрововозов В.И., к.т.н,
Хемраев А.К.

АНАЛИЗ ГРАДИЕНТА ДЛЯ НЕЙРОННЫХ СЕТЕЙ С ВЕЙВЛЕТ-РАЗЛОЖЕНИЕМ ЦЕЛЕВОГО ВЕКТОРА

Национальный авиационный университет

Представлен анализ градиента для некоторых случаев нейронных сетей с вейвлет и вейвлет-подобным разложением целевого вектора – типа нейронной сети, специализированного на распознавании речи и преобразовании сигнала, и позволяющего ускорить скорость и качество обучения по сравнению со стандартным перцептроном. Показывается, что в достаточно широких рамках нейронные сети с вейвлет-разложением целевого вектора эффективнее стандартного многослойного перцептрона.

Вступление

Искусственные нейронные сети (ИНС) с вейвлет-разложением целевого вектора (НСВЦ) созданы для задач распознавания речи и используются в этой области, хотя могут быть применены для широкого круга задач, подразумевающих преобразование одного сигнала (в значении – система отсчётов некоторой функции) в другой сигнал (другую систему отсчётов некоторой функции). Так же существует множество работ, обсуждающих и доказывающих высокую эффективность вейвлет-методов для фонемного распознавания [3], особенно следует упомянуть биологическую аналогию вейвлет-обработки звукового сигнала, описанную у Добеши [4]. Широкий спектр задач, решаемых ИНС обуславливает актуальность и целесообразность исследования.

Целью работы является анализ градиента нейронных сетей с вейвлет-разложением целевого вектора.

Пути решения

НСВЦ – специализированный тип нейронных сетей. Их практическая эффективность обусловлена тем, что рассматривая множество выходных сигналов обучающей выборки (далее целевых векторов) с помощью вейвлет-разложения находятся спектральные диапазоны, в которых локализован сигнал, и спектральные области, не значащие для решения, исключаются из области поиска [5]. Исключение производится с помощью проецирования значений нейросети на области спектральной локализации нейронной сети, причём сам процесс проецирования производится за счёт модуля обратного вейвлет-разложения выходного сигнала многослойного перцептрона автоматически.

Так как модуль обратного вейвлет-разложения реализован в нейронном базисе, это позволяет говорить о новой нейронной структуре на базе перцептрона – НСВЦ. Данная структура построена по принципу локализации решения, в соответствии с которым подбирается преобразование выходных векторов выборки, которое позволяет обнаружить и использовать диапазоны локализации данных векторов и выходные значения ИНС проецируются на данные диапазоны, что позволяет сузить область поиска решения, т.е уменьшить вероятность попадания в локальные максимумы, скорость и точность обучения.

Пусть модуль обратной вейвлет-декомпозиции осуществляет преобразование, обратное вейвлет-преобразованию, заданному двумя зеркльно-квадратурными *FIR* фильтрами G и H, определяемыми коффициентами h_i и g_i.

В этом случае вейвлет-разложение может быть реализовано последовательным применением свёрток $H(z_i)=\sum_{l=1}^{n_H} h_{l-2k} z_i$ и $G(z_i)=\sum_{l=1}^{n_G} g_{l-2k} z_i$.

Пусть дана обучающая выборка $(\widetilde{x}_i,\widetilde{y}_i)$, $i=1,2,\dots,i_{\max}$ где вектора $\widetilde{x}_i,\widetilde{y}_i$ равны $\widetilde{x}_i=(x_{i1},x_{i2},\dots,x_{in})$, $\widetilde{y}_i=(y_{i1},y_{i2},\dots,y_{im})$, и вейвлет преобразование F, преобразующее вектор $\widetilde{y}_i$ в набор коэффициентов $F_j^{(l)}(\widetilde{y}_i)$: $F_j^{(l)}(\widetilde{y}_i)=F(j,l,\widetilde{y}_i)$, $j=1,2,\dots,j_{\max}$, $l=1,2,\dots,l_{\max}$, где l – номер уровня вейвлет-разложения, а j – номер коэффициента на данном уровне вейвлет-разложения. Компонент $\tilde{y}_i$ обучающей пары $\tilde{x}_i, \tilde{y}_i$ будем называть целевым компонентом или целевым вектором. Тогда:

1. Целевые компоненты пар-примеров обучающей выборки подвергаются вейвлет-разложению.

2. На каждом уровне l вейвлет разложения по всей выборке выбираются минимальные и максимальные значения

$$S_l = \max_{i,j} F_j^{(l)}(\tilde{y}_i), \tag{1}$$

$$I_l = \min_{i,j} F_j^{(l)}(\tilde{y}_i). \tag{2}$$

3. Те же самые операции проводятся над контрольной выборкой. Результатом являются контрольные значения $\hat{S}_l$ и $\hat{I}_l$.

4. Если контрольный максимум не превосходит максимум обучающей выборки более чем на заданную константу точности $\varepsilon, \hat{S}_l < S_l + \varepsilon$, а минимум обучающей выборки не превосходит контрольный минимум более чем ту же ε, $\hat{I}_l > I_l + \varepsilon$, то можно говорить о корректности выбранных минимальных и максимальных значений для данной задачи.

5. Диапазонами частотной локализации будут области $[I_l; S_l]$ для каждого уровня l вейвлет-разложения.

Данный алгоритм дан здесь в виде инициализации ИНС, но практически возможно не инициализировать величины I_l, S_l, а корректировать их в процессе предъявления обучающих примеров.

НСВЦ состоит из двух модулей – многослойного перцептрона и модуля обратной вейвлет-проекции. Значения многослойного перцептрона принципиально ограничены. Процессом проецирования выходных значений перцептрона, соответствующих l-му уровню масштаба вейвлет-разложения на область частотной локализации будет сдвиг и масштабирование области значений нейронной сети на область частотной локализации $[I_l; S_l]$. Обозначим верхнее значение, принимаемое входами ИНС будет S, а нижнее I.

Проецирование значений перцептрона o_j^l на диапазоны частотной локализации решения осуществляется следующим образом:

$$\hat{o}_j^l = \left(o_j^l - I\right)\frac{S_l - I_l}{S - I} + I_l. \tag{3}$$

Анализ влияния масштабирования на градиент. Введём следующие обозначения:

E – ошибка слоя.

$E(w)$ – целевая функция.

d_j – желаемый выходной сигнал j-го нейрона слоя.

y_j – выходное значение, $y \in (0,1)$.

x_i – входное значение.

w_{ij} – вес, связывающий i-й вход с j-м выходом.

u_i^t – взвешенная сумма i-го нейрона, t-го слоя.

$$u_i^t = \sum_{k=0}^{N} w_{ki}^t x_k^t .$$

Согласно метода распространения ошибки, имеют место следующие формулы:

$$E(w) = \tfrac{1}{2}\sum_{i=1}^{n_N}(y_i - d_i)^2 \tag{4}$$

$$E_i^t = \sum_{l=1}^{n_t} E_l^{t+1} w_{il}^{t+1}, \tag{5}$$

$$E_i^N = y_i - d_i. \tag{6}$$

$$\Delta w_{ij}^t = -\eta\left(E_j^t \cdot \frac{dy_i^t}{du_i^t} \cdot x_i^t\right) \tag{7}$$

Где η – скорость обучения.

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_i}\frac{\partial u_j}{\partial w_{ij}}\frac{dy_j}{du_j} = \frac{\partial E}{\partial y_i} x_j \frac{dy_j}{du_j} \tag{8}$$

Пусть $D(x)$ – функция, которая позволяет найти производную активационной функции f по её значению

Т.е. если $y_j = f(u_j)$,

то $D(y_j) = f'(u_j)$,

$$\Rightarrow \frac{dy_j}{du_j} = D(y_j) \tag{9}$$

Рассмотрим случай биполярной функции.

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_i} x_i D(y_j) \qquad (10)$$

Рассмотрим влияние масштабирования на компоненты градиента. Пусть выход y_j во всех ситуациях не превосходит $1/k_{\max}$.

С учётом того, что нейронная сеть с униполярной функцией активации выходного слоя выдаёт значения в области (0,1) возможно промасштабировать выход, увеличив его в $k \le k_{\max}$ раз. Для этого достаточно промасштабировать соответствующий целевой вектор. Для этого введём замену переменной $y' = ky$

Тогда, для масштабированного выхода получаем

$$\frac{\partial E'}{\partial w_{ij}} = k(y_i - d_i) x_i D(ky_j) \qquad (11)$$

$$D(ky_j) = 1 - k^2 y_j^2 \qquad 12)$$

Тогда для биполярной функции:

$$\frac{\partial E'}{\partial w_{ij}} = k \frac{\partial E}{\partial w_{ij}} \frac{ky_j(1-ky_j)}{y_j(1-y_j)} = k^2 \frac{\partial E}{\partial w_{ij}} \frac{(1-ky_j)}{(1-y_j)} \qquad (13)$$

Введём величину M как отношение

$$M = \frac{\partial E'}{\partial w_{ij}} \Big/ \frac{\partial E}{\partial w_{ij}}. \qquad (14)$$

Тогда

$$M(y) = k\frac{1-k^2y^2}{1-y^2}. \qquad (15)$$

Этот множитель отображает участок биполярной (логистической) сигмоиды на отрезке $(0;1/k]$ в полноценную сигмоиду на отрезке $(0;1)$. При этом:

$$\frac{\partial E'}{\partial w_{ij}} = \frac{\partial E}{\partial w_{ij}} M(y_j). \qquad (16)$$

Исследуем свойства множителя M.

$$M(y) = k\frac{1-k^2 y_j^2}{1-y_j^2} = k^3 + \frac{k-k^3}{1-y^2} \qquad (17)$$

$$[M(y)]' = \left[k^3 + \frac{k-k^3}{1-y^2}\right]' = (k-k^3)\frac{2y}{(1-y^2)^2}. \qquad (18)$$

С учётом того, что k>1, множитель $(k-k^3)$ всегда отрицательный. Очевидно, что M имеет единственный экстремум – максимум $\max\{M(y)\} = k$ в точке y=0.

Таким образом очевидно, что в случае биполярной функции её градиент в k раз больше при y=0, а далее монотонно убывает при $y \to +1$ и $y \to -1$. Выбор коэффициента k будет производится на основании этой особенности.

Для дальнейших рассуждений необходимо вычислить отрезок, на котором $M(y)$ больше единицы, т.е. найти то множество y, на котором градиент $\frac{\partial E'}{\partial w_{ij}} > \frac{\partial E}{\partial w_{ij}}$.

Исходя из монотонного убывания M(y) на отрицательной и положительной полуосях для нахождения этого множества достаточно найти такие y, при которых M(y)=1.

$$M(y) = 1 \Leftrightarrow$$

$$k^3 + \frac{k-k^3}{1-y^2} = 1 \Leftrightarrow y^2 = 1 - \frac{k-k^3}{1-k^3} \Rightarrow$$

$$y^2 = \frac{1}{k^2+k+1} \Rightarrow y = \pm\frac{1}{\sqrt{k^2+k+1}}. \qquad (19)$$

Итак, очевидно, что $M(y) > 1$ при $y \in \left(-\frac{1}{\sqrt{k^2+k+1}}; \frac{1}{\sqrt{k^2+k+1}}\right)$.

Обозначим y_{eq} такую величину, что $M(\pm y_{eq}) = 1$.

$$M(y) > 1 \Leftrightarrow y \in (-y_{eq}, +y_{eq}). \qquad (20)$$

Подберём коэффициент k масштабирования таким образом, чтобы

$$ky_{\max} = k_{\max} y_{eq}, \qquad (21)$$

где $y_{\max}$ – максимальное значение y, причём $y_{\max} = 1/k_{\max}$.

При этом условии будет верно $M(y) > 1$ или, что равносильно

$$\frac{\partial E'}{\partial w_{ij}} \ge \frac{\partial E}{\partial w_{ij}}, \qquad (22)$$

для всех значений ограниченной величины y.

Так как:

$$y_{eq} = \frac{1}{\sqrt{k_{\max}{}^2 + k_{\max} + 1}}, \quad (23)$$

$$ky_{\max} = k_{\max} y_{eq} \Leftrightarrow$$

$$k\frac{1}{k_{\max}} = \frac{k_{\max}}{\sqrt{k_{\max}{}^2 + k_{\max} + 1}}, \quad (24)$$

что возможно если:

$$k = \frac{(k_{\max})^2}{\sqrt{k_{\max}{}^2 + k_{\max} + 1}}. \quad (25)$$

При данном коэффициенте масштабирования градиент целевой функции больше на всей области значений нейронной сети, что ускоряет обучение ИНС.

Сравним градиенты исходной и масштабированной функции.

Градиент целевой функции $gradE$ базовой нейронной сети:

$$gradE = \sqrt{\sum_{i,j}\left(\frac{\partial E}{\partial w_{ij}}\right)^2}. \quad (26)$$

Градиент целевой функции НСВЦ:

$$gradE' = \sqrt{\sum_{i,j}\left(\frac{\partial E'}{\partial w_{ij}}\right)^2}. \quad (27)$$

Пусть $y'_j = k_j y_j$. Тогда:

$$\frac{\partial E'}{\partial w_{ij}} = M(k_j, y_j)\frac{\partial E}{\partial w_{ij}}, \quad (28)$$

При этом из соотношения 25 следует что для любых k_j, y_j, удовлетворяющих условию задачи, $M(k_j, y_j) > 1$. (29)

$$gradE' = \sqrt{\sum_{i,j}\left(M(k_j, y_j)\frac{\partial E}{\partial w_{ij}}\right)^2} = \quad (30)$$

$$= \sqrt{\sum_{i,j} M^2(k_j, y_j)\left(\frac{\partial E}{\partial w_{ij}}\right)^2} > \sqrt{\sum_{i,j}\left(\frac{\partial E}{\partial w_{ij}}\right)^2} = gradE$$

Выводы

Показано, что при нахождении областей частотной локализации и проецировании значений нейронной сети на область локализации решения, в достаточно широких рамках можно добиться увеличения градиента (по сравнению с той же ИНС без частотной локализации и модуля обратной вейвлет-проекции), а следовательно, скорости сходимости нейронной сети.

Список литературы

1. Tebelskis, J. Speech Recognition using Neural Networks: PhD thesis, Doctor of Philosophy in Computer Science/ Joe Tebelskis; School of Computer Science, Carnegie Mellon University.– Pittsburgh, Pennsylvania. – 1995.– 179 p.
2. Handbook of neural network signal processing/ Edited by Yu Hen Hu, Jenq-Neng Hwang. – Boca Raton; London; New York, Washington D.C.: CRC press. – 2001. – 384 p.
3. Ф.Г. Бойков Применение вейвлет-анализа в задачах автоматического распознавания речи: Дис. кандидата физико-математических наук: 05.13.18 / Фёдор Геннадьевич Бойков. – М., 2003. – 111 с.
4. Добеши И. Десять лекций по вейвлетам.– Ижевск: НИЦ «Регулярная и хаотическая динамика», – 2001. – 464 с.
5. Астапов К.А. Применение вейвлет-преобразования для сокращения области значения искусственных нейронных сетей на примере задачи распознавания речи // Электронный научно-инновационный журнал «Инженерный вестник Дона», Ростов-на-Дону: – 2009. – № 1. – С. 24-29.
6. Червяков Н.И., Астапов К.А. Использование вейвлетов для улучшения параметров нейронных сетей в задачах распознавания речи. // Инфокоммуникационные технологии – №°4 (2008) – Самара: Издательство ПГУТИ, – 2008. – С. 9-12.