

РАДИОТЕХНИКА, СИСТЕМЫ ТЕЛЕКОММУНИКАЦИЙ, АНТЕННЫ И УСТРОЙСТВА СВЧ

УДК: 621.396.6

С.В. Зимина

ФЛУКТУАЦИИ ВЕСОВЫХ КОЭФФИЦИЕНТОВ В ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ, НАСТРАИВАЮЩЕЙСЯ ПО АЛГОРИТМУ LMS С КВАДРАТИЧНЫМ ОГРАНИЧЕНИЕМ

Нижегородский государственный университет им. Н.И. Лобачевского

Приведены результаты статистического анализа с учетом флуктуаций настраиваемых весовых коэффициентов искусственной нейронной сети, настраиваемой по алгоритму минимизации среднего квадрата ошибки с квадратичным ограничением на усиление полезного сигнала. Методами теории возмущений по коэффициенту адаптации алгоритма, который будем считать малым, в первом (так называемом «борновском») приближении найдено выражение корреляционной функции выходного сигнала. Показано, что учет флуктуаций весовых коэффициентов приводит к искажениям выходного сигнала искусственной нейронной сети. Полученные результаты позволяют выбирать величину коэффициента адаптации алгоритма настройки искусственной нейронной сети, приводящей к оптимальному соотношению между скоростью настройки сети и точностью настройки.

Ключевые слова: искусственная нейронная сеть, флуктуации весовых коэффициентов, алгоритм минимизации среднего квадрата ошибки с квадратичным ограничением.

Введение

Флуктуации весовых коэффициентов в адаптивных системах определяют скорость и точность настройки таких систем. Одним из вариантов адаптивных систем являются искусственные нейронные сети (ИНС) [1-2]. Флуктуации настраиваемых весовых коэффициентов, возникающие вследствие статистической связи весовых коэффициентов и входного сигнала ИНС, искажают ее статистические характеристики и, в частности, выходной сигнал искусственной нейронной сети. В работах [3-6] был проведен статистический анализ функционирования ИНС, настраиваемых по дискретному градиентному, быстрому рекуррентному алгоритмам и алгоритму Хэбба, с учетом флуктуаций весовых коэффициентов. Было показано, что флуктуации приводят к искажениям выделяемого искусственной нейронной сетью полезного сигнала, причем в искусственной нейронной сети, настраиваемой по алгоритму Хэбба, величина искажений увеличивается по мере увеличения числа слоев сети.

В последнее время внимание исследователей снова начинает привлекать классический алгоритм минимизации среднего квадрата ошибки (алгоритм МСКО или алгоритм LMS), который одним из первых был предложен еще Уидроу. Однако применение алгоритма LMS встретило трудности, выражающиеся в неконтролируемом подавлении полезного сигнала в случае, когда помеха коррелирована с сигналом. По этой причине в настоящее время интерес вызывают в первую очередь различные современные модификации LMS алгоритма, не содержащие этого недостатка, и в частности, алгоритм LMS с квадратичным ограничением на усиление полезного сигнала [7]. Представляет интерес исследование влияния флуктуаций весовых коэффициентов на статистические характеристики искусственной нейронной сети,

настраивающейся по этому алгоритму. В данной работе представлен статистический анализ корреляционной функции выходного сигнала произвольного слоя искусственной нейронной сети, настраивающейся по алгоритму минимизации среднего квадрата ошибки с квадратичным ограничением, полученной с учётом флуктуаций весовых коэффициентов.

1. Постановка задачи

Рассмотрим работу искусственной нейронной сети, настраивающейся по LMS алгоритму с квадратичным ограничением. Настройка вектора весовых коэффициентов $\vec{W}_i^\alpha(k+1)$ i -ого искусственного нейрона слоя α такой ИНС в момент времени $k+1$ описывается уравнением следующего вида:

$$\vec{W}_i^\alpha(k+1) = \vec{W}_i^\alpha(k) - \mu \cdot [\vec{Z}^{\alpha-1}(k) \cdot F[(\vec{Z}^{\alpha-1}(k))^H \vec{W}_i^\alpha(k)] + \frac{|F[(\vec{Z}^{\alpha-1}(k))^H \vec{W}_i^\alpha(k)]|^2}{|F[\vec{S}^H \vec{W}_i^\alpha(k)]|^2} \vec{S} \cdot F[\vec{S}^H \vec{W}_i^\alpha(k)]] \quad (1)$$

Здесь $\vec{Z}^{\alpha-1}(k)$ - вектор выходного сигнала слоя $\alpha-1$ искусственной нейронной сети в момент времени k ; μ - коэффициент адаптации LMS алгоритма с квадратичным ограничением; $\vec{S} = [S_1, S_2, \dots, S_N]^T$ - вектор полезного сигнала, H - операция эрмитовского сопряжения.

В выражении (1) $\vec{Z}^{\alpha-1}(k)$ - входной сигнал i -го нейрона слоя α , являющийся также вектором выходного сигнала слоя $\alpha-1$. Выходной сигнал i -го нейрона слоя α $Z_i^\alpha(k)$ может быть записан в виде:

$$Z_i^\alpha(k) = F \left[\sum_{q=1}^N Z_q^{\alpha-1}(k) W_{iq}^\alpha(k) \right] = F[y_{ai}(k)] = \sum_{l=1}^{N_l} a_l \cdot y_{ai}^l(k) = \vec{A}^T \vec{Y}_{ai}(k),$$

где $F[\]$ - нелинейная активационная функция рассматриваемого искусственного нейрона, a_l - коэффициенты разложения нелинейности F в ряд Вольтерра [8], $y_{ai}(k)$ - выходной сигнал линейной части i -го нейрона слоя α . $\vec{A} = [a_1 \ a_2 \ \dots \ a_{N_l}]^T$ - вектор коэффициентов разложения функции $F[\]$ в ряд Вольтерра, а j -й коэффициент разложения может быть представлен формулой $a_j = \frac{1}{j!} F^{(j)}[y=0]$. Таким образом, первый коэффициент разложения в ряд

Вольтерра пропорционален первой производной нелинейной активационной функции, находящейся на выходе рассматриваемого искусственного нейрона.

$\vec{Y}_{ai}(k) = [y_{ai}(k) \ y_{ai}^2(k) \ \dots \ y_{ai}^{N_l}(k)]^T$ - вектор степеней выходного сигнала линейной части искусственного нейрона i , принадлежащего слою α .

В статье предполагается, что все искусственные нейроны в описываемой искусственной нейронной сети имеют одну и ту же нелинейную функцию $F[\]$ на выходе. Поэтому при обобщении результатов от одного искусственного нейрона на слой, а потом и на произвольное число слоев коэффициенты разложения нелинейной функции в ряд Вольтерра будут одни и те же.

Алгоритм настройки (1) описывает в большей степени своего рода каскадное соединение искусственных нейронов, чем единую искусственную нейронную сеть, поскольку весовые коэффициенты каждого отдельного нейрона зависят только от собственных значений в предыдущий момент времени и не связаны непосредственными математическими отношениями с весовыми коэффициентами других искусственных нейронов. Тем не менее, косвенное

влияние весовых коэффициентов различных искусственных нейронов друг на друга существует и осуществляется через выходные сигналы данных элементов сети, что также видно из формулы (1).

Будем предполагать, что рассматриваемая искусственная нейронная сеть является узкополосной. Корреляционную матрицу входных сигналов узкополосной ИНС любого искусственного нейрона входного слоя данной сети можно представить в виде:

$$\mathbf{R}_{xx}(k, k+n) \equiv \langle \vec{X}^*(k) \vec{X}^T(k+n) \rangle = \mathbf{R}_{xx} r^{|n|}, \quad (2)$$

где $\vec{X}(k)$ - вектор полезного сигнала в момент времени k ; * и T - операции комплексного сопряжения и транспонирования; r - коэффициент корреляции между отсчётами входного сигнала; \mathbf{R}_{xx} - пространственная часть корреляционной матрицы входных сигналов.

Аналогичную форму в виде произведения пространственной и временной частей имеют также корреляционные матрицы входных сигналов искусственных нейронов любого слоя искусственной нейронной сети, различается только вид пространственной части \mathbf{R}_{xx} . Необходимо отметить, что данная корреляционная матрица имеет различный вид в зависимости от номера слоя искусственной нейронной сети. Так, для всех скрытых слоев в частном случае одинаковой помеховой обстановки, подаваемой на все нейроны входного слоя, который рассматривается в данной работе, корреляционная матрица будет иметь в качестве всех своих элементов единицы во всех скрытых слоях ИНС. Во входном слое эта матрица будет равна корреляционной матрице входных сигналов искусственной нейронной сети.

Будем ограничиваться для простоты и в силу существенно нелинейного вида представленного алгоритма (1) первым слагаемым при разложении нелинейных функций на выходе искусственных нейронов в ряд Вольтерра, т.е. только первой производной активационной функции. Добавление при разложении в ряд слагаемых более высокого порядка усложняет анализ, не внося новой информации в выводы.

Будем искать статистические характеристики описанной нейронной сети с учётом флуктуаций настраиваемых весовых коэффициентов в предположении, что между весовым вектором и вектором входных сигналов существует негауссовская статистическая зависимость. Это наиболее общий случай такой зависимости между данными векторами. С этой целью необходимо на первом этапе анализа найти статистические характеристики отдельного нейрона с учётом флуктуаций весового вектора, а затем обобщить полученные результаты на искусственную нейронную сеть и получить соответствующие рекуррентные формулы.

2. Корреляционная функция выходного сигнала искусственного нейрона, настраиваемого по алгоритму LMS с квадратичным ограничением

Методами теории возмущений по малому параметру μ в первом, так называемом «борновском» приближении, было получено выражение для корреляционной функции выходного сигнала искусственной нейронной сети, настраиваемой по LMS алгоритму с квадратичным ограничением, с учётом флуктуаций весовых коэффициентов:

$$K_z(m, m+n) = A_1 + B_1 + C_1 + D_1. \quad (3)$$

Здесь коэффициенты корреляционной функции (3) могут быть представлены в виде:

$$A_1 = a_1^2 \vec{W}_{CT}^H \mathbf{R}_{xx} \vec{W}_{CT} \cdot r^{|n|}, \quad (4)$$

$$B_1 = \mu^2 a_1^4 \frac{M^{-1}}{1-r^2} r^{|n|} Sp(\mathbf{R}_{xx}) \vec{W}_{CT}^H \mathbf{R}_{SS}^H \mathbf{R}_{xx} \vec{W}_{CT} + \mu^2 a_1^4 \frac{M^{-1} r^2}{1-r^2} r^{|n|} Sp(\mathbf{R}_{xx}) \vec{W}_{CT}^H \mathbf{R}_{xx} \mathbf{R}_{SS} \vec{W}_{CT}, \quad (5)$$

$$C_1 = \mu^2 a_1^4 r \frac{1+r-r^n}{(1-r)^2} r^{|n|} \vec{W}_{CT}^H \mathbf{R}_{XX}^3 \vec{W}_{CT} + \mu^2 a_1^4 \frac{r}{(1-r)^2(1+r)} r^{|n|} Sp(\mathbf{R}_{XX}) \vec{W}_{CT}^H \mathbf{R}_{XX}^2 \vec{W}_{CT} +$$

$$+ \mu^2 a_1^4 r^{|n|} \left\{ \frac{2}{(1-r^2)^2} + \frac{1-r^{2n+2}}{1-r^2} \right\} Sp(\mathbf{R}_{XX}^2) \vec{W}_{CT}^H \mathbf{R}_{XX} \vec{W}_{CT} \quad (6)$$

$$D_1 = \mu^2 a_1^4 \frac{r^{|n|+3}}{1-r^2} \left\{ \frac{1-r^n}{1-r^2} + \frac{1}{1-r} \right\} Sp(\mathbf{R}_{XX}^2) \vec{W}_{CT}^H \mathbf{R}_{XX}^2 \vec{W}_{CT}. \quad (7)$$

\vec{W}_{CT} - постоянный стационарный весовой вектор произвольного искусственного нейрона первого слоя искусственной нейронной сети, a_1 - первый коэффициент разложения нелинейной функции $F[]$ в ряд Вольтерра, M^{-1} - постоянный коэффициент, возникающий при усреднении LMS алгоритма с квадратичным ограничением.

Из выражений (4)–(7) видно, что влияние флуктуаций весовых коэффициентов проявляется в появлении в выражении корреляционной функции слагаемых, пропорциональных коэффициенту адаптации μ алгоритма настройки ИНС, которые искажают вид полезного сигнала, выделяемого искусственной нейронной сетью. Можно также видеть, что величина влияния флуктуаций имеет второй порядок малости по коэффициенту адаптации μ , т.е. достаточно малая.

Из формул (4)–(7) также можно видеть, что представленные слагаемые корреляционной функции выходного сигнала зависят также от величины первого коэффициента разложения в ряд Вольтерра нелинейной функции, которая присутствует на выходе искусственного нейрона. В частном случае наиболее распространенного вида логистической нелинейной функции или сигмоида $f(x) = \frac{1}{1+e^{-px}}$ [9] первый коэффициент разложения равен $a_1 = p/4$, где величина p определяет крутизну сигмоида. При $p = 0$ сигмоид вырождается в горизонтальную линию на уровне 0.5, при увеличении p сигмоид приближается к виду функции единичного скачка [9]. Для практических задач вполне достаточно брать величину $p = 0.5 \div 1$, что приводит к тому, что $a_1^4 = 1/4^4 = 1/256 = 0.0039$. Иными словами, «флуктуационные» слагаемые будут малы не только в силу малости коэффициента адаптации, а также и по причине малости коэффициента a_1^4 .

3. Корреляционная функция выходного сигнала искусственного нейрона, принадлежащего произвольному слою ИНС

Обобщая данный результат последовательно на второй, третий и т.д. слои можно получить выражение для корреляционной функции выходного сигнала i -ого искусственного нейрона, принадлежащего слою α искусственной нейронной сети, настраиваемой по LMS алгоритму с квадратичным ограничением, с учётом флуктуаций весового вектора:

$$K_\alpha(m, m+n) = \sum_{i=1}^4 A_i K_{\alpha-1}^i(m, m+n) = \sum_{i=1}^4 A_{\alpha i} \left(\sum_{j=1}^4 A_{\alpha-1j} K_{\alpha-2}^j(m, m+n) \right)^i =$$

$$= \sum_{i=1}^4 A_{\alpha i} \left(\sum_{j=1}^4 A_{\alpha-1j} \left(\sum_{k=1}^4 A_{\alpha-2k} K_{\alpha-3}^k(m, m+n) \right)^j \right)^i = \quad (8)$$

$$= \sum_{i_1=1}^4 A_{\alpha i_1} \left(\sum_{i_2=1}^4 A_{\alpha-i_2} \left(\sum_{i_3=1}^4 A_{\alpha-2i_3} \cdots \left(\sum_{i_\alpha=1}^4 A_{1i_\alpha} K_1^{i_\alpha}(m, m+n) \right)^{i_{\alpha-1}} \cdots \right)^{i_2} \right)^{i_1}$$

Здесь коэффициенты $A_{\alpha i}$, возникшие вследствие учета флуктуаций весовых коэффициентов, имеют вид:

$$A_{\alpha 1} = a_1^2 (\vec{W}_{CTi}^\alpha)^H \mathbf{B} (\vec{W}_{CTi}^\alpha) r^{|n|}; \quad (9)$$

$$A_{\alpha 2} = \mu^2 a_1^4 Sp(\mathbf{B}) \frac{1}{1-r^2} \times \left\{ M^{-1} (\vec{W}_{CTi}^\alpha)^H \mathbf{R}_{SS}^H \mathbf{B} (\vec{W}_{CTi}^\alpha) + M^{-1} r^2 (\vec{W}_{CTi}^\alpha)^H \mathbf{B} \mathbf{R}_{SS} (\vec{W}_{CTi}^\alpha) + \frac{r}{1-r} (\vec{W}_{CTi}^\alpha)^H \mathbf{B} (\vec{W}_{CTi}^\alpha) \right\} r^{|n|}; \quad (10)$$

$$A_{\alpha 3} = \mu^2 a_1^4 \times \left\{ \frac{r(1+r-r^n)}{(1-r)^2} + \frac{2+(1-r^2)(1-r^{2n+2})}{(1-r^2)^2} Sp(\mathbf{B}^2) \right\} (\vec{W}_{CTi}^\alpha)^H \mathbf{B} (\vec{W}_{CTi}^\alpha) r^{|n|}; \quad (11)$$

$$A_{\alpha 4} = \mu^2 a_1^4 \frac{r^3(2+r-r^{|n|})}{(1-r^2)^2} Sp(\mathbf{B}^2) (\vec{W}_{CTi}^\alpha)^H \mathbf{B}^2 (\vec{W}_{CTi}^\alpha) r^{|n|}. \quad (12)$$

Здесь \vec{W}_{CTi}^α - постоянный стационарный весовой вектор искусственного нейрона i , принадлежащего слою α искусственной нейронной сети.

Матрица \mathbf{B} зависит от того, коэффициенты какого слоя вычисляются. Для входного слоя искусственной нейронной сети эта матрица равна корреляционной матрице входных сигналов, для последующих слоев, в силу постановки задачи, матрица \mathbf{B} имеет все элементы, равные единице.

Из выражений (8) – (12) следует, что влияние флуктуаций весовых коэффициентов проявляет себя в том, что корреляционная функция выходного сигнала предыдущего слоя входит в выражение для корреляционной функции последующего слоя как полином четвертой степени, коэффициентами в котором служат обусловленные флуктуациями слагаемые второго порядка малости по коэффициенту адаптации μ . От слоя к слою ИНС это явление повторяется, и результирующий вид корреляционной функции выходного сигнала произвольного слоя ИНС представляет из себя набор вложенных одна в другую функций, так что запись результата через все слои искусственной нейронной сети содержит наибольшее число вложенных друг в друга функций. Если говорить о максимальной степени, в которой войдет корреляционная функция выходного сигнала первого слоя ИНС в формулу корреляционной функции произвольного слоя сети, то она пропорциональна $4^{\alpha-1}$.

Иными словами, если учесть, что коэффициенты в каждом полиноме в корреляционной функции содержат коэффициент адаптации во второй степени, который полагается малым, то можно ожидать, что искажения выделяемого ИНС полезного сигнала вследствие флуктуаций, хоть и сложны по форме, но вклад их будет уменьшаться по мере увеличения числа слоев сети, т.к. степень при коэффициенте адаптации во «флуктуационных» слагаемых также будет расти.

Для сравнения приведем выражение для корреляционной функции выходного сигнала i -го нейрона, принадлежащего слою α искусственной нейронной сети, полученное без учета флуктуаций весовых коэффициентов:

$$\begin{aligned} K_\alpha(m, m+n) &= A_{\alpha 1} K_{\alpha-1}(m, m+n) = A_{\alpha 1} A_{\alpha-1} K_{\alpha-2}(m, m+n) = \\ &= A_{\alpha 1} A_{\alpha-1} A_{\alpha-2} K_{\alpha-3}(m, m+n) = A_{\alpha 1} A_{\alpha-1} A_{\alpha-2} \cdots A_{11} \cdot 1 = \\ &= \prod_{j=1}^{\alpha} A_{j1} = a_1^{2\alpha} r^{\alpha|n|} \prod_{j=1}^{\alpha} (\vec{W}_{CTi}^j)^H \mathbf{B}_j (\vec{W}_{CTi}^j) \end{aligned} \quad (13)$$

Здесь

$$\mathbf{B}_j = \begin{cases} \mathbf{R}_{xx}, & j = 1 \\ \mathbf{E}, & j = 2, 3, \dots \end{cases} \quad (14)$$

Верхние индексы в обозначениях весовых коэффициентов являются номером слоя, а не степенью.

Из сравнения выражений (8) и (13) видно, что формула выходной мощности, полученная без учета флуктуаций весовых коэффициентов, существенно упрощается и содержит произведение выходных мощностей искусственных нейронов всех слоев ИНС. Из формулы (13) также можно видеть, что без учета флуктуаций корреляционная функция выходного сигнала каждого последующего слоя ИНС перестает быть полиномом, составленным по степеням корреляционной функции выходного сигнала предыдущего слоя искусственной нейронной сети, и связана с ней линейной зависимостью. Кроме того, формула (14) не содержит «флуктуационных» слагаемых, пропорциональных коэффициенту адаптации, а зависит только от первого коэффициента разложения нелинейной функции F на выходе искусственного нейрона в ряд Вольтерра.

Заключение

Таким образом, статистический анализ работы искусственной нейронной сети, настраиваемой по LMS алгоритму с квадратичным ограничением на усиление полезного сигнала, с учетом флуктуаций весовых коэффициентов показал, что они приводят к искажению выделяемого искусственной нейронной сетью сигнала. По мере роста числа слоев искусственной нейронной сети происходит накопление искажений, вносимых флуктуациями в выделяемый сетью полезный сигнал, однако величина данных искажений сдерживается тем, что увеличивается также и степень коэффициента адаптации, которому пропорциональны «флуктуационные» слагаемые, а коэффициент адаптации обычно бывает значительно меньше единицы. При практическом использовании такой искусственной нейронной сети необходимо добиваться оптимального соотношения между скоростью настройки сети (скорость настройки увеличивается при увеличении коэффициента адаптации) и точностью настройки ИНС (минимальным уровнем искажений, вносимых флуктуациями весовых коэффициентов в выделяемый сетью сигнал, что достигается при уменьшении коэффициента адаптации). Выбор зависит от задач конкретного использования искусственной нейронной сети.

Библиографический список

1. Haykin S. Neural Networks: A comprehensive Foundation. - New York: Macmillan College Publishing Company, 1994.
2. **Галушкин, А.И.** Теория нейронных сетей / А.И. Галушкин. – М.: ИПРЖР, 2000. – 416 с.
3. **Зимина, С.В.** Статистические характеристики искусственной нейронной сети с дискретным градиентным алгоритмом настройки с учётом флуктуаций весовых коэффициентов // Нейрокомпьютеры: разработка и применение, 2006. № 10. С. 9–15.
4. **Зимина, С.В.** Влияние флуктуаций весовых коэффициентов на статистические характеристики искусственной нейронной сети с алгоритмом рекуррентного обращения выборочной оценки корреляционной матрицы входных сигналов // Нейрокомпьютеры: разработка и применение, 2007. № 5. С. 3–7.
5. **Литвинов, О.С.** Статистический анализ флуктуаций весовых коэффициентов искусственной нейронной сети, настраиваемой по алгоритму Хэбба / О.С. Литвинов, С.В. Зимина // Нейрокомпьютеры: разработка и применение, 2009. №3. С. 33–43.
6. **Зимина, С.В.** Флуктуации весовых коэффициентов в искусственной нейронной сети с алгоритмом Хэбба // Нейрокомпьютеры: разработка и применение, 2013. N4. С. 3–8.
7. **Орешкин, Б.Н.** Алгоритм LMS с квадратичным ограничением / Б.Н. Орешкин, П.А. Бакулев // Антенны, 2007. N9. С. 29–34.

8. **Пупков, К.А.** Функциональные ряды в теории нелинейных систем / К.А. Пупков, В.И. Капалин, Ющенко А.С. – М.: Наука, 1976.
9. **Круглов, В.В.** Искусственные нейронные сети. Теория и практика / В.В. Круглов, В.В. Борисов. – М.: Горячая линия – Телеком, 2002. – 382 с.

*Дата поступления
в редакцию 23.04.2015*

S.V. Zimina

WEIGHT VECTOR JITTER IN THE NEURAL NETWORK WITH LMS ALGORITHM TUNING WITH SQUARE CONSTRAINT

Lobachevsky State University of Nizhni Novgorod

The article presents the results of statistical analysis taking into account weight vector jitter of an neural network tuning the algorithm to minimize the average squared error with a quadratic constraint on the gain of the desired signal. Methods of perturbation theory in the first (so-called "born") at a rate adaptation algorithm, which we assume to be small, the approximation is found the expression of the correlation function of the output signal. It is shown that taking into account fluctuations of weight coefficients leads to distortions in the output signal of an neural network. The obtained results allow to choose the value of the coefficient adaptation algorithm configuration of an neural network, resulting in optimal ratio between the speed of the network configuration and tuning accuracy.

Key words: neural network, fluctuations of weight coefficients, the algorithm to minimize the average squared error with a quadratic constraint.