

Глава 6.

Погрешности в нейронных сетях

А.Н. Горбань, М.Ю. Сенашова

ВЦК СО РАН, КГУ

Рассматриваются нейронные сети слоистой структуры, состоящие из слоев стандартных нейронов. Изучаются ошибки, возникающие при технической реализации сетей, а также при шумах и повреждениях.

Определены максимально допустимые погрешности, возможные для сигналов и параметров каждого элемента сети, исходя из условия, что вектор выходных сигналов сети должен вычисляться с заданной точностью. Используются два типа оценок погрешности: гарантированные интервальные оценки и среднеквадратические оценки погрешностей.

Показано, что оценки допустимых погрешностей можно получить в ходе специального процесса “обратного распространения точности”. Он состоит в функционировании сети с той же системой связей, но от выходов к входам и с заменой элементов на двойственные. Эта двойственность принципиально отличается от той, которая используется в классическом методе вычисления градиентов оценки с помощью обратного распространения ошибок (back propagation of errors).

С помощью полученных результатов объясняется наблюдаемая высокая устойчивость нейронных сетей к шумам и разрушениям.

1. Введение

В настоящее время существуют различные технические реализации нейронных сетей, в том числе нейроимитаторы, то есть компьютерные модели нейронных сетей. Нейроимитаторы являются гибкими средствами для изучения сетей и работы с ними. С нейроимитаторами можно выполнять различные

операции – обучать, определять наиболее и наименее значимые связи, контрастировать, то есть удалять наименее значимые связи и т. д.

Существует подход, получающий все большее распространение, при котором сначала конструируется и обучается нейроимитатор, а затем создается техническая реализация полученной нейросети с уже вычисленными весами синапсов.

Нейроимитатор, работающий на универсальных цифровых ЭВМ, позволяет вычислять веса синапсов с большой точностью, которую трудно получить при других технических реализациях сети (в первую очередь – аналоговых) в силу ограниченной точности технических устройств. Поэтому возникает задача приведения весов синапсов к некоторому набору конкретных значений. Ее частный случай - задача бинаризации, то есть задача приведения весов синапсов к значениям 0 или 1 (связь либо есть, либо нет - без всяких весов синапсов).

При аналоговых реализациях, различных упрощениях архитектуры (в том числе - бинаризации) технически сложно получить результат работы сети той же точности, что и результат работы нейроимитатора [3-5]. Поэтому следует ограничиться некоторой точностью, с которой может работать сеть, то есть выбрать интервал, в котором могут изменяться значения вектора выходных сигналов сети.

Оценка погрешностей сигналов сети очень полезна при решении задачи упрощения нейронной сети. Зная допустимую погрешность выходного сигнала какого-либо элемента сети, мы можем заменять его более простыми, но менее точными элементами так, чтобы в итоге ошибка не превышала заданную.

Хорошо известно, что нейронные сети могут проявлять исключительную устойчивость к помехам и разрушениям. Иногда эти эффекты называют голографическими свойствами нейронных сетей, подразумевая, что полезные навыки распределены по сети примерно так же, как изображение - по голографической пластинке, и могут сохраняться при значительных разрушениях.

Как будет показано ниже, при прямом прохождении сигналов по достаточно большой сети погрешности гасятся: при больших погрешностях входных сигналов выходные сигналы сети могут иметь сравнительно малые погрешности. Это объясняет устойчивость нейронных сетей к шумам и повреждениям.

2. Два базовых подхода к оценкам погрешности

Рассмотрим два подхода к решению задачи вычисления погрешностей сигналов сети. При первом подходе (*гарантированные интервальные оценки*) вычисляются допустимые интервалы для погрешностей сигналов сети такие, что погрешность вектора выходных сигналов гарантированно (с вероятностью 1) не превышает заданную. При втором подходе (*среднеквадратические оценки погрешностей*) вычисляются среднеквадратические отклонения погрешностей сигналов. При этом часто используется предположение о том, что погрешности различных сигналов являются независимыми случайными величинами.

Существует принципиальное различие между этими двумя типами оценок. Гарантированные интервальные оценки исходят из рассмотрения наихудших возможных случаев, сколь бы малой не была их вероятность. Поэтому они, как правило, завышают реально имеющую место ошибку и слишком пессимистичны с практической точки зрения. Среднеквадратичные оценки, наоборот, стирают возможные большие отклонения и могут оказаться слишком оптимистичными.

Важное различие между двумя типами оценок демонстрируют следующие *формулы сложения*.

1. *Формула сложения для интервальных оценок*. Пусть для двух величин x , y определены гарантированные интервалы значений $x=x_0\pm\Delta_x$, $y=y_0\pm\Delta_y$. Тогда для их суммы имеем гарантированную оценку: $x+y= x_0+y_0 \pm(\Delta_x+\Delta_y)$, то есть $\Delta_{x+y}=\Delta_x+\Delta_y$.

2. *Формула сложения для среднеквадратичных отклонений*. Пусть для двух *независимых* величин x , y определены среднеквадратичные отклонения σ_x , σ_y . Тогда $\sigma_{x+y}=(\sigma_x^2+\sigma_y^2)^{1/2}$.

3. Структура сети

Предполагаем, что сеть имеет слоистую структуру. Это самоограничение позволит несколько сократить изложение, но не влияет на общность рассмотрения - исследование любой сети может быть формально сведено к изучению слоистых сетей.

Сеть слоистой структуры состоит из слоев стандартных нейронов, связанных между собой синапсами с весами, полученными при обучении. Причем сигналы передаются только в одном направлении, с предыдущего слоя на следующий. Под стандартным нейроном [1,2] понимается набор элементов, состоящий из адаптивного сумматора, нелинейного преобразователя и точки ветвления (рис.1). Точка ветвления – это элемент, посылающий выходной сигнал нелинейного преобразователя на вход нескольких стандартных нейронов следующего слоя.

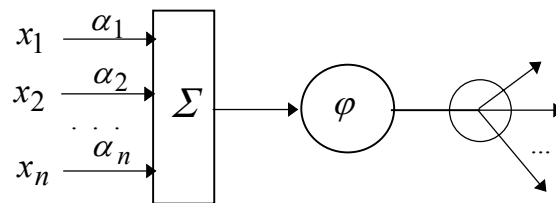


Рис.1. Стандартный нейрон.

Σ – адаптивный сумматор, φ – нелинейный преобразователь и точка ветвления.

Так как мы имеем дело с сетями слоистой структуры, состоящими из слоев стандартных нейронов, то выходные сигналы одного слоя являются входными сигналами другого слоя. В свою очередь, внутри самого стандартного нейрона выходной сигнал одного элемента (например, сумматора) является входным сигналом другого элемента (например, нелинейного преобразователя). Таким образом, можно проследить, начиная с выходных сигналов сети, от какого элемента сети пришел сигнал к данному элементу.

Стандартный нейрон является типичным участком любой нейронной сети. Поэтому достаточно выяснить, как вычисляются допустимые погрешности для элементов стандартного нейрона. В результате получим возможность вычислить допустимые погрешности для любого участка сети, двигаясь по сети от нейрона к нейрону.

4. Метод обратного распространения точности для гарантированных интервальных оценок

Пусть нам заданы допустимые погрешности вычислений для выходных сигналов сети. Для каждого элемента решим задачу: определить допустимые погрешности на входах элемента по заданным максимально допустимым погрешностям на его выходе. Если эту задачу решить для каждого элемента сети, то можно оценить допустимые погрешности для всех сигналов, проходящих через сеть, переходя по сети от элемента к элементу в обратном направлении (от выходов сети к ее входам). Этот процесс мы назовем ***обратным распространением точности***. В ходе него движение сигналов происходит от выходов ко входам, сигнал, проходящий по связи в обратном направлении, является допустимой погрешностью сигнала, проходящего по этой связи в прямом направлении.

Последним элементом стандартного нейрона является точка ветвления, поэтому начинаем рассмотрение метода обратного распространения точности именно с нее.

Точка ветвления имеет несколько выходов. Пусть для каждого ее выхода задана допустимая погрешность ε_i (i - номер выхода). Для того, чтобы удовлетворить всем этим ограничениям погрешности, необходимо и достаточно, чтобы входной сигнал точки ветвления имел погрешность $\varepsilon = \min\{\varepsilon_i\}_{i=1}^k$. Таким образом, при обратном распространении точности тока ветвления заменяется на двойственный элемент, выбирающий из поступающих сигналов ε_i (т.е. погрешностей) минимальный.

Следующим элементом стандартного нейрона является нелинейный преобразователь. Пусть входной сигнал нелинейного преобразователя равен A_0 , φ – его функция активации, $y = \varphi(A_0)$ – выходной сигнал и ε_1 – допустимая погрешность выходного сигнала. Вычислим максимальную погрешность ε входного сигнала нелинейного преобразователя, то есть найдем отрезок $[A_0 - \varepsilon, A_0 + \varepsilon]$ такой, что для любого $x \in [A_0 - \varepsilon, A_0 + \varepsilon]$ $\tilde{y} = \varphi(x)$ отличается от $y = \varphi(A_0)$ не более, чем на ε_1 : $|y - \tilde{y}| \leq \varepsilon_1$.

Ввиду непрерывности и дифференцируемости функции активации нелинейного преобразователя очевидно, что $\varepsilon \leq \varepsilon_1 / \max |\varphi'(x)|$, где $x \in [\varphi^{-1}(y - \varepsilon_1), \varphi^{-1}(y + \varepsilon_1)]$.

Пойдем традиционным путем, оценивая допустимую погрешность в линейном приближении: $\varphi(A_0 \pm \varepsilon) \approx \varphi(A_0) \pm \varphi'(A_0) \cdot \varepsilon$. По условию

$$\varepsilon_1 \geq |\varphi(A_0 \pm \varepsilon) - \varphi(A_0)| \approx |\varphi'(A_0) \cdot \varepsilon|.$$

Пользуясь этим неравенством, подберем ε следующим образом: $\varepsilon \leq \varepsilon_1 / |\varphi'(A_0)|$. В этом случае формула для вычисления допустимой погрешности более простая, но менее точная.

Получили погрешность, допустимую для входного сигнала нелинейного преобразователя, которая одновременно является допустимой погрешностью для выходного сигнала сумматора. Аналогично можем вычислить погрешность входного сигнала нелинейного преобразователя любого стандартного нейрона, если известна погрешность его выходного сигнала.

Двойственный к нелинейному преобразователю элемент – просто линейная связь! Ее вес равен $1/|\varphi'(A_0)|$ для линейного приближения в формуле ошибки или $1 / \max |\varphi'(x)|$ – в более общем случае (в последней формуле максимум берется по отрезку $x \in [\varphi^{-1}(y - \varepsilon_1), \varphi^{-1}(y + \varepsilon_1)]$ – так что линейность здесь уже кажущаяся).

Перейдем к следующему элементу стандартного нейрона – адаптивному сумматору с n синапсами, являющимися его входами. Адаптивный сумматор – это сумматор, в котором входные сигналы x_i суммируются с весами α_i .

Каждый вход x_i сумматора Σ_0 имеет некоторую погрешность ε_i , которая вносит свой вклад в допустимую погрешность выходного сигнала сумматора. Эти погрешности могут иметь различные величины в зависимости от того, какой способ распределения допустимой погрешности выходного сигнала по входам сумматора мы выберем. Погрешности по входам сумматора могут распределяться равномерно, пропорционально и приоритетно.

Рассмотрим сначала равномерное распределение. Для этого полагаем, что на всех входах погрешности равны между собой ($\varepsilon_i = \varepsilon, i = 1, \dots, n$). Пусть

$A_0 = \sum_{i=1}^n \alpha_i \cdot x_i$ – выходной сигнал сумматора без погрешностей. Тогда $\{A'_0\}$ –

множество выходных сигналов сумматора, получающихся, когда вектор входных сигналов сумматора пробегает вершины n -мерного куба с центром в точке

(x_1, x_2, \dots, x_n) и ребром длины 2ε : $\{A'_0\} = \sum_{i=1}^n \alpha_i \cdot (x_i \pm \varepsilon_i) = \sum_{i=1}^n \alpha_i \cdot x_i$

$+ \sum_{i=1}^n \alpha_i \cdot (\pm \varepsilon_i) = A_0 + \sum_{i=1}^n \alpha_i \cdot \varepsilon_i \cdot z_i$, где $z_i \in \{-1, 1\}$. Нам требуется, чтобы все

множество значений $\{A'_0\}$ попало в интервал $[A_0 - \varepsilon, A_0 + \varepsilon]$. Для этого

необходимо, чтобы $\max \left| \sum_{i=1}^n \alpha_i \cdot \varepsilon_i \cdot z_i \right| = \sum_{i=1}^n |\alpha_i| \varepsilon_i \leq \varepsilon$, где максимум берется по

всем z_i . Из этого неравенства и сделанного выше предположения о ε_i получаем требуемую оценку для равномерного распределения ε_i по входам сумматора:

$$\varepsilon_i \leq \varepsilon / \sum_{i=1}^n |\alpha_i|.$$

При пропорциональном распределении погрешностей допустимая погрешность выходного сигнала сумматора делится сначала на число входов, а затем для каждого входа делится на соответствующий вес синапса. То есть

погрешности распределяются пропорционально весам соответствующих синапсов. Формула расчета допустимой погрешности для каждого входа сумматора имеет вид: $\varepsilon_i = \varepsilon / (n \cdot \alpha_i)$, где ε – допустимая погрешность выходного сигнала сумматора, n – число входов сумматора, α_i – веса синапсов соответствующих входов сумматора.

При приоритетном распределении погрешностей сначала назначаются погрешности для тех входов, которые наиболее значимы по какому-либо признаку, а затем оставшуюся часть допустимой погрешности выходного сигнала сумматора распределяют между оставшимися входами равномерно или пропорционально.

Аналогично можно вычислить допустимые погрешности для входных сигналов сумматора любого стандартного нейрона, если известны погрешности для выходного сигнала сумматора.

Для адаптивного сумматора можно вычислять как допустимые погрешности входных сигналов сумматора, так и допустимые погрешности весов синапсов. Для вычисления допустимых погрешностей весов синапсов также можно использовать равномерное, пропорциональное и приоритетное распределение погрешностей. При равномерном распределении допустимые погрешности для весов синапсов вычисляются по формуле: $\varepsilon_i \leq \varepsilon / \sum_{i=1}^n |x_i|$, где x_i – входные сигналы сумматора.

При пропорциональном распределении допустимые погрешности для весов синапсов вычисляются по формуле: $\varepsilon_i \leq \varepsilon / (n \cdot x_i)$, где n – число входов сумматора, x_i – входные сигналы сумматора.

При приоритетном распределении сначала назначаются допустимые погрешности для тех весов синапсов, которые наиболее значимы по какому-либо признаку, а затем оставшуюся часть допустимой погрешности для выходного сигнала сумматора распределяют между оставшимися весами синапсов равномерно или пропорционально.

При обратном распространении точности имеет место специфическая двойственность - элементы сети заменяются на двойственные им. Однако, эта двойственность отличается от той, с которой мы встречаемся при изучении обратного распространения ошибки для вычисления градиентов функции оценки. Так, если в обычном обратном распространении двойственным элементом к точке ветвления является простой сумматор, то при обратном распространении точности вместо него, как было показано, появляется элемент, вычисляющий минимум приходящих на него сигналов. Нелинейный преобразователь при обратном распространении точности заменяется двойственным ему элементом, умножающим сигнал на число. Но если при обратном распространении ошибки множителем является значение градиента, то в нашем случае сигнал умножается на величину обратную производной от входного сигнала нелинейного преобразователя. Адаптивный сумматор также заменяется двойственным ему элементом. Этот элемент является своеобразной точкой ветвления. Но, в отличие от простой точки ветвления, он сначала преобразует приходящий к нему сигнал в соответствии с выбранным распределением погрешностей по входам адаптивного сумматора, а затем передает полученные сигналы дальше.