

## Оптимизация разрядности аппаратных средств при обеспечении требуемой точности работы нейронных сетей

С.Н. Данилин, М.В. Макаров, С.А. Щаников

*Муромский институт Владимирского государственного университета  
602264, Муром, Владимирской области, ул. Орловская д.23,  
тел. (492-34) 2-06-98, факс (49234) 2-28-85,  
E-mail: [nauka-murom@yandex.ru](mailto:nauka-murom@yandex.ru)*

В процессе аппаратной реализации нейросетевых алгоритмов обработки информации на базе программируемых логических интегральных схем, элементная база накладывает ограничения по разрядности, быстродействию, числу входов и выходов, которые оказывают существенное влияние на основные технические показатели устройств [1].

Разрядность входной информации и весовых коэффициентов оказывает определяющее влияние на оптимальное использование ресурсов элементной базы при обеспечении наилучших технических показателей нейронных сетей, а именно: точности, надежности, стоимости [2].

В связи с вышеизложенным, на математических моделях специализированных вычислительных устройств с обученными нейронными сетями, было проведено исследование зависимости точности их работы от разрядности входной информации и весовых коэффициентов. Рассмотрим в качестве примера вычисление в нейросетевом логическом базисе следующих функций:

$$y = \sqrt{x}, y = \frac{1}{x}, \begin{cases} X = R \cdot \sin \varphi, \\ Y = R \cdot \cos \varphi. \end{cases}$$

В процессе исследований синтезировались двухслойные нейронные сети. В качестве инструментального средства исследований использовался пакет прикладных программ Neural Network Toolbox версии 4, входящий в состав Matlab версии 7.

Разрядность входной информации определялась разрядностью АЦП и равнялась 12, а затем, в процессе эксперимента, уменьшалась на 1, 2, ..., 11 единиц путем округления.

Округление результатов сложения существенно не влияет на конечную точность, так как оно производится в сторону ближайшего целого числа с точностью до половины младшего разряда, при этом положительные и отрицательные округления равновероятны и при большом количестве сложений они взаимно компенсируют друг друга. Округление результатов вычисления функции активации также не будет оказывать существенного влияния, так как она чаще всего реализуется таблично.

Разрядность входной информации и весовых коэффициентов формировалась логическим способом, путем наложения на операнд разрядной сетки в диапазоне от 0 до  $2N$  (где  $N$  изменяется от 12 до 1), что эквивалентно изменению разрядности представления операндов в том же диапазоне.

В качестве показателя точности работы нейронной сети выбрана относительная погрешность преобразования информации. Необходимо отметить, что нейронная сеть обучалась без учета

ограничений на разрядность представления входных данных и весовых коэффициентов, но, тем не менее, после наложения ограничений типа округлений показала хорошие результаты по точности работы.

Результаты математического моделирования показывают, что уменьшение разрядности весовых коэффициентов с 12 до 3 и разрядности входной информации с 12 до 7 для нейронной сети, реализующей функцию  $y = \sqrt{x}$  снижают точность работы нейронной сети на 15 – 20 % от исходного уровня.

Для функции  $y = \frac{1}{x}$  высокий уровень точности работы нейронной сети (0,98 от исходного значения) сохраняется при уменьшении разрядности входной информации с 12 до 10 и разрядности весовых коэффициентов с 12 до 1.

Для функции  $\begin{cases} X = R \cdot \sin \varphi, \\ Y = R \cdot \cos \varphi. \end{cases}$  уровень точности работы нейронной сети сохраняется до уровня 0,85 от исходного значения при уменьшении разрядности входной информации с 12 до 4 и разрядности весовых коэффициентов с 12 до 1.

Проведенные исследования показали, что предложенный метод является эффективным при нахождении оптимальной разрядности аппаратных средств, реализующих нейросетевые алгоритмы, и может быть рекомендован для применения, как в теоретических исследованиях, так и в инженерной практике.

### Литература

1. Галушкин А.И. Теория нейронных сетей. – М.: ИПРЖР, 2000.-416 с.
2. Данилин С.Н. Экспериментальное исследование надежности обученных нейронных сетей прямого распространения. Журнал "Нейрокомпьютеры: разработка, применение" – Москва: Изд. «Радиотехника», №3, 2006г., 78 с.- с.63-70.