МЕТОДЫ ОПРЕДЕЛЕНИЯ ОБЪЕМА ОБУЧАЮЩЕЙ ВЫБОРКИ ДЛЯ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ

А.С. Михайлов, М.С. Нехорошкина
Кафедра автоматики и микропроцессорной техники, кафедра информационных технологий
ФГБОУ ВПО «Костромской государственный технологический университет»
Кострома, Российская Федерация
E-mail: amt@kstu.edu.ru, vt@kstu.edu.ru

В докладе представлены новые методы определения объема обучающей выборки, достаточного для успешного обучения искусственной нейронной сети, выполняющей функции универсального аппроксиматора.

Введение

Несмотря на большое количество публикаций, свидетельствующих об успешном опыте применения искусственных нейронных сетей (ИНС) для решения различных задач экономики, медицины и техники, в настоящее время практически неисследованной остается проблема определения объема обучающей выборки для ИНС. Эта проблема выражается в необходимости соответствия обучающей выборки двум противоречивым требованиям:

а) обучающая выборка должна быть репрезентативной и иметь количество примеров, достаточное для качественного обучения ИНС;
б) для сокращения временных затрат на предварительную подготовку данных и процедуру обучения объем обучающей выборки должен быть минимально возможным.

На сегодняшний день известны несколько эвристических правил для приближенного определения необходимого количества примеров в обучающей выборке, например, измерение Вапника-Червоненкиса (Yapnik-Chervonenkis dimension, VC-измерение) [1, 2] или правило Видроу (Widrow's rule of thumb) [3]. Однако, часто оказывается, что между действительно достаточным объемом обучающей выборки и этими оценками существует большой разрыв [4].

В настоящей работе предлагаются два новых метода определения объема обучающей выборки — метод половинного деления, по своей сути аналогичный одноименному методу вычислительной математики, и метод, основанный на анализе частотного спектра функции, которую аппроксимирует нейронная сеть, и вычислении частот Найквиста. При изложении предлагаемых методов принимаются три следующих обстоятельства:

а) задача, решаемая с помощью ИНС, может быть сведена к задаче аппроксимации;
б) известен вид зависимости, которую должна аппроксимировать ИНС (без ущерба для общности ограничимся рассмотрением функции двух переменных $y = f(x_1, x_2)$, т. е. примем, что ИНС имеет 2 входа x_1 и x_2, а также один выход y);
в) известны область определения, т. е. $x_1 \in (x_{1_{min}}; x_{1_{max}})$ и $x_2 \in (x_{2_{min}}; x_{2_{max}})$, а также область допустимых значений функции, которую должна аппроксимировать ИНС, т. е. $y \in (y_{min}; y_{max})$.

I. Метод половинного деления

Данный метод определения объема обучающей выборки демонстрирует хорошие результаты при использовании радиальных базисных сетей с нулевой ошибкой [5, 6].

Суть метода заключается в следующем:

а) на координатной плоскости, называемой картой обучения (рис. 1), по осям абсцисс и ординат в соответствующем масштабе отмечаются диапазоны изменения аргументов функции, которую должна аппроксимировать ИНС, в нашем случае $x_1 \in (x_{1_{min}}; x_{1_{max}})$ и $x_2 \in (x_{2_{min}}; x_{2_{max}})$;
б) указанные в предыдущем пункте диапазоны делятся пополам, и на карту обучения добавляются точки с координатами $(x_{1_0}; x_{2_0})$, $(x_{1_0}; x_{2_{min}})$, $(x_{1_0}; x_{2_{max}})$, $(x_{1_{min}}; x_{2_0})$, $(x_{1_{max}}; x_{2_0})$, где $x_{1_0} = \frac{x_{1_{min}} + x_{1_{max}}}{2}$ и $x_{2_0} = \frac{x_{2_{min}} + x_{2_{max}}}{2}$;
в) в обучающую выборку включаются следующие сочетания аргументов функции: $(x_{1_{min}}; x_{2_{min}})$, $(x_{1_{min}}; x_{2_{max}})$, $(x_{1_0}; x_{2_0})$, $(x_{1_{max}}; x_{2_{min}})$, $(x_{1_{max}}; x_{2_{max}})$, $(x_{1_0}; x_{2_{min}})$, $(x_{1_0}; x_{2_{max}})$, $(x_{1_{min}}; x_{2_0})$, $(x_{1_{max}}; x_{2_0})$, а также соответствующие этим сочетаниям значения функции y;
г) проводится обучение ИНС и проверка ее работы на тестовом множестве;
д) в случае неудовлетворительного качества обучения ИНС начинается следующая итерация, диапазоны $(x_{1_{min}}; x_{1_0})$ и $(x_{1_0}; x_{1_{max}})$, а также $(x_{2_{min}}; x_{2_0})$ и $(x_{2_0}; x_{2_{max}})$ делятся пополам, и на карте обучения отмечаются новые точки;
е) в обучающую выборку добавляются соответствующие данные, и проводится повторное обучение ИНС.

При необходимости осуществляется еще одна или несколько итераций.

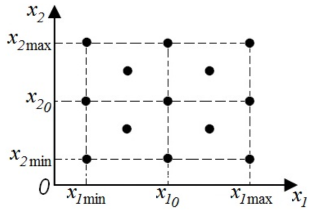

Рис. 1 – Пример карты обучения для метода половинного деления

Очевидно, что данный метод определения объема обучающей выборки является итерационным, и в этом заключается его главный недостаток. Достоинством метода является его прозрачность и интуитивно понятный алгоритм формирования обучающей выборки. Необходимо также отметить, что данный метод может применяться не одновременно к обоим диапазонам изменения аргументов функции, а поочередно. В этом случае будет иметь место иной характер распределения точек на карте обучения при сохранении общей идеи метода.

II. Частотный метод

На сегодняшний день получены удовлетворительные результаты при использовании данного метода определения объема обучающей выборки для ИНС с прямым распространением сигнала и обратным распространением ошибки.

Суть метода заключается в следующем:

а) для получения частотного спектра функции $y = f(x_1, x_2)$ проводится двумерное, а в общем случае — многомерное, преобразование Фурье (непрерывное либо дискретное), результатом которого является некоторая зависимость $S = f(\Omega_{x_1}, \Omega_{x_2})$ (рис. 2) [7];

б) из графика функции $S = f(\Omega_{x_1}, \Omega_{x_2})$ для обоих аргументов x_1 и x_2 определяются так называемые частоты Найквиста $\Omega_{H_{x_1}}$ и $\Omega_{H_{x_2}}$, при которых высшими гармониками можно пренебречь. За частоты Найквиста принимаются такие значения Ω_{x_1} и Ω_{x_2}, при которых значение функции $S = f(\Omega_{x_1}, \Omega_{x_2})$ не превышает 10 % от ее максимального значения S_{max} ;

в) далее определяются периоды дискретизации для обоих аргументов $\Delta x_{1_{max}}$ и $\Delta x_{2_{max}}$, соответствующие частотам Найквиста $\Omega_{H_{x_1}}$ и $\Omega_{H_{x_2}}$ [8];

г) с учетом полученных значений $\Delta x_{1_{max}}$ и $\Delta x_{2_{max}}$ на карте обучения проводится «разбивка» диапазонов изменения аргументов функции $y = f(x_1, x_2)$, в нашем случае $x_1 \in (x_{1_{min}}; x_{1_{max}})$ и $x_2 \in (x_{2_{min}}; x_{2_{max}})$;

д) в обучающую выборку добавляются данные в соответствии с картой обучения (рис. 3), проводится обучение ИНС и проверка ее работы на тестовом множестве.

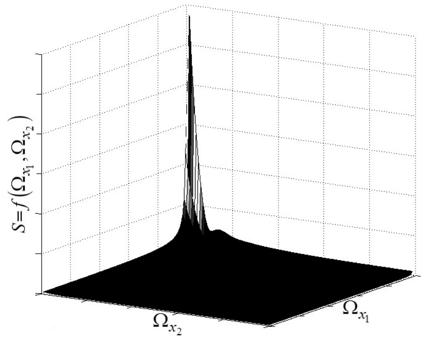

Рис. 2 – Пример графика функции $S = f(\Omega_{x_1}, \Omega_{x_2})$

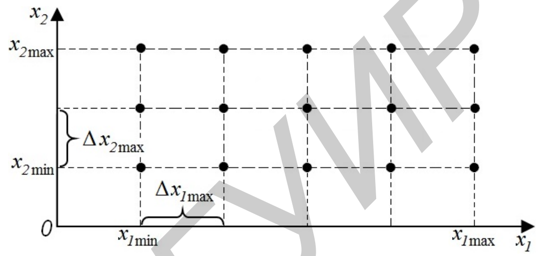

Рис. 3 – Пример карты обучения для частотного метода

Применение частотного метода позволяет снизить трудоемкость и длительность процесса обучения, а также исключить его итерационный характер. Вместе с тем, могут возникнуть сложности, связанные с невозможностью визуализации и, как следствие, невозможностью полноценного анализа n -мерных графиков при $n > 3$.

1. Vapnik, V. N. On the uniform convergence of relative frequencies of events to their probabilities / V. N. Vapnik, A. Ya. Chervonenkis // Theoretical Probability and Its Aplications. – 1971. – Vol. 17. – P. 264–280.
2. Koiran, P. Neural networks with quadratic VC-dimension / P. Koiran, E. D. Sontag // Advances in Neural Information Processing Systems. – 1996. – Vol. 8. – P. 197–203.
3. Widrow, B. Adaptive Signal Processing / B. Widrow, S. D. Steams // Englewood Cliffs. – NJ: Prentice-Hall. – 1985.
4. Гулаков, К. В. Применение методов нейросетевого моделирования для уменьшения объемов экспериментальных работ при разработке сварочных материалов / К. В. Гулаков // Вестник Брянского государственного технического университета. – 2011. – № 3(31). – С. 111–117.
5. Михайлов, А. С. Динамический регулятор состояния с нейросетевой настройкой для нестационарного объекта управления / А. С. Михайлов, Б. А. Староверов // Вестник ИГЭУ. – 2014. – № 3. – С. 53–59.
6. Михайлов, А. С. Синтез динамического регулятора состояния с контуром нейросетевой адаптации / А. С. Михайлов, Б. А. Староверов // Труды III Всероссийской научной конференции молодых ученых с международным участием «Теория и практика системного анализа». – Т. I. – Рыбинск: РГАТУ имени П. А. Соловьева. – 2014. – С. 64–73.
7. Сергиенко, А. Б. MATLAB и преобразование Фурье / А. Б. Сергиенко // Exponenta Pro. Математика в приложениях. – 2003. – № 3(3). – С. 84–86.
8. Даджион, Д., Мерерсо Р. Цифровая обработка многомерных сигналов: Пер. с. англ. – М.: Мир. – 1988. – 448 с.