Власов Л.В., Малыхина Г.Ф., Тархов Д.А.
Vlasov L.V., Malyhina G.F, Tarhov D.A.

НЕЙРОННЫЙ ЭМУЛЯТОР “ESSENCE”

NEURAL EMULATOR “ESSENCE”

Санкт-Петербургский государственный технический университет
St. Petersburg State Technical University

Neural emulator “Essence” is suggested for identification of dynamic inlinear systems in their parameters measurement problems.

Развитие и совершенствование средств вычислительной техники и их широкое применение в информационно-измерительных системах обусловило процесс алгоритмизации измерений. Разработка и реализация современных алгоритмов обработки измерительной информации позволяет не только уменьшить погрешность традиционных методов измерения, но и на основе косвенных и совокупных методов выполнять измерение величин, которые ранее не могли быть измерены. В этом плане перспективным является построение процедур измерения, использующих модель объекта измерения как нелинейной динамической системы со стационарными или нестационарными параметрами и привлечение аппарата темпоральных нейронных сетей.

Несмотря на существование различных технологий выделения знаний из данных, к которым относятся, в частности, деревья решений, генетические алгоритмы, эволюционное программирование, групповой учёт аргументов, наиболее отработанной и зрелой является нейросетевая технология. Она реализована как в программном виде, так и аппаратно для многочисленных приложений, таких, как косвенные измерения, прогнозирование значений сигналов, техническая и медицинская диагностика. При этом под нейронной сетью понимают модель системы, состоящую из однородных элементов - нейронов, содержащих сумматоры и активационные функции, линейного или нелинейного вида, функционирование которой зависит от коэффициентов, величины которых, определяются в результате обучения.

Соотношение между входными $\mathbf{x}_i$ и выходными $\mathbf{y}_i$ параметрами динамической системы может быть представлено системой дифференциальных уравнений. Применение нейронной сети при измерении в нелинейных динамических системах включает этап обучения нейронной сети в процессе калибровки средства измерения, связанной с подстройкой коэффициентов $\mathbf{w}$ таким образом, чтобы в среднем за время обучения ошибка выхода сети была минимальной:

$$E\{\varepsilon(\mathbf{w})\} = E\left\{\frac{1}{2}\sum_{i=1}^{N}\left(d_i - F(\mathbf{x_i}, \mathbf{w})\right)^2\right\} \qquad (1)$$

где E{.}символ усреднения, $\mathbf{x}_i = \left[\mathbf{x}_{i,t}, \mathbf{x}_{i,t-1}, \ldots, \mathbf{x}_{i,t-p}\right]$ вектор входных параметров, который включает текущие $\mathbf{x}_{i,t}$ и предыдущие $\mathbf{x}_{i,t-n}$, задержанные на n отсчетов значения параметров, d_i - известный выходной сигнал динамической системы, $F(\mathbf{x_i}, \mathbf{w})$ - функция преобразования входного сигнала динамической системой и этап измерения, связанный с вычислением выходного сигнала нейронной сети по известным входным сигналам и коэффициентам нейронной сети:

$$y_i = F(\mathbf{x_i}, \mathbf{w}) \qquad (2)$$

Нейронная сеть обладает свойствами адаптивной системы, поскольку является самоорганизующейся системой, не требующей тщательной предварительной разработки. Нейронная сеть является нелинейной системой с изменяющимися во времени параметрами, способной самонастраиваться в процессе измерений при изменениях в динамической системе.

@2002, электронная версия подготовлена ЗАО АВТЭКС Санкт-Петербург, http://www.autex.spb.ru

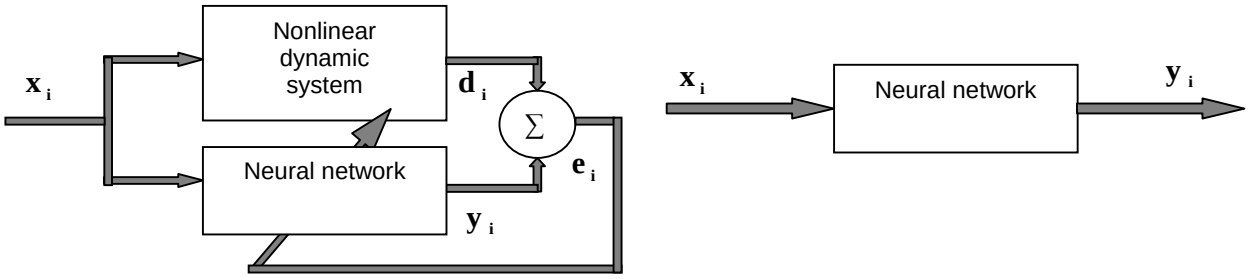


Fig.1 Dynamic system parameter measurement - a) learning, б) measurement

Процесс идентификации объекта, связанный с выбором класса моделей, является трудоемким и плохо формализуемым этапом разработки алгоритма измерения. Для автоматизации этого этапа при использовании нейросетевых моделей объекта измерения авторами предлагается нейронный эмулятор “Essence”. “Essence” представляет собой пакет программ, выполненных на языке Java, который позволяет достаточно легко выполнять сетевые соединения для обмена данными в распределенной измерительно-вычислительной системе и открывать сеансы с базами данных, содержащими измерительную информацию, предназначенную для обучения нейронной сети. Распределенные измерительные системы должны давать возможность извлекать измерительную информацию из источников разного рода: локальных и распределённых баз данных, текстовых и иных файлов, информационных систем различного рода, включая Интернет.

Пакет программ имеет современный интерфейс, выполненный с использованием javaBeans. Пакет позволяет автоматизировать этапы предварительной обработки информации, выбора архитектуры нейронной сети и требуемого размера нейронной сети, определения характера нелинейности и выбора подходящего алгоритма обучения нейронной сети.

Предварительная обработка данных направлена на то, чтобы наиболее полно извлечь содержащуюся в них информацию. Нормализация входных параметров включает вычитание среднего или начального значения, устранение корреляции входных параметров на основе метода главных компонент, приведение дисперсии всех параметров к одному и тому же значению. Эти методы существенно повысят скорость обучения нейронной сети, уменьшат время обучения. Желаемые значения выходных параметров нейронной сети должны быть приведены к интервалу, который определяется видом активационных функций нейронов. Например, при использовании в качестве активационной функции гиперболического тангенса $\varphi(v) = b \cdot \tanh(a \cdot v)$, где a=1.7159, b=0.66667, желаемый отклик сети должен находиться в диапазоне [-1,1]. На этапе предварительной обработки информации “Essense” предоставляет возможность выбора типа нормализации параметров сети, применение метода кластеризации для входных параметров сети и использование метода главных компонент для устранения избыточности, связанной с коррелированностью параметров. От выбора подходящего метода нормализации параметров сети во многом зависит успешность обучения. В пакете предложены три варианта нормализации входных параметров, которые, по мнению авторов, позволяют повысить информативность всех входных параметров для обучения и последующего использования.

В результате анализа главных компонент входных параметров часто оказывается возможным использовать в дальнейшем не все, а только часть главных компонент, которые определяют близкую к единице, например 0.98, долю энергии входных сигналов. Этот метод позволяет уменьшить размер сети за счет уменьшения числа входов. Новые входные параметры, число которых сокращается за счет исключения корреляции, являются линейно независимыми комбинациями исходных параметров. При этом уменьшается число оцениваемых синаптических коэффициентов, более эффективно используются обучающие данные, быстрее проходит процесс обучения.

Другим важным методом предварительной обработки является кластеризация. Применение кластеризации векторов входных параметров важно для правильной идентификации объектов измерения с нестационарными, изменяющимися во времени, характеристиками. При этом отдельные классы входных векторов представляют локально стационарные состояния системы. Для каждого локально стационарного состояния может быть определена отдельная нейронная сеть с параметрами, зависящими от класса состояний.

В настоящее время авторы работают над включением в блок предварительной обработки разложения по адаптивному вайвлет-базису, что актуально для входных сигналов, имеющих определённую сложную структуру.

В пакете автоматизирован этап выбора архитектуры нейронной сети. “Essence” позволяет с помощью специального окна задавать число слоев и число нейронов в каждом слое нейронной сети. Методика выбора архитектуры сети предполагает наличие двух выборок данных, полученных в результате наблюдения объекта измерения: обучающей и тестирующей. Одновременно в процессе обучения нейронной сети выполняется ее проверка по тестирующей выборке. Если уменьшение ошибки сети, вычисленной по обучающей выборке, связано с возрастанием ошибки, вычисленной по тестовой выборке, то сеть является переобученной или слишком сложной для данного объекта.

Важным параметром сети является вид нелинейности активационных функций нейронов. В пакете предложено использовать восемь видов активационных функций, среди которых сигмоидная, пороговая функция, функция синуса, арктангенса, гауссинан и другие известные функции активации нейронов.

Практическая реализация нейросетевых алгоритмов измерения во многом определяется эффективностью вычислительного процесса, связанного с обучением сети. Поэтому важным этапом является выбор метода обучения нейронной сети. Обучение по типу «образец за образцом», когда изменения весов сети происходит каждый раз по предъявлении нового вектора входных параметров, эффективно при больших множествах обучающих данных. Получение измерительных данных недостаточно, то наиболее полно извлечь из них информацию позволяют методы обучения, основанные на эпохах, когда коррекция весов сети происходит каждый раз после анализа всех обучающих векторов данных.

Для различных исходных данных и различных сетевых архитектур наиболее эффективным может оказаться тот или иной метод обучения. “Essence” позволяет выполнить сопоставление двенадцати методов обучения нейронной сети, заданной архитектуры, и на основе этого сопоставления сделать окончательных выбор метода обучения. Среди предлагаемых методов обучения метод случайного поиска, статистического градиента, Руммельхарта, Бройдена-Флетчера, сопряженных градиентов и другие известные методы минимизации ошибки сети в процессе обучения.

В качестве оценки ошибки на выходе нейронной сети, возникающей в процессе обучения и тестирования, предложено использовать традиционную оценку (1) для среднего значения суммы квадратов отклонений выхода сети от обучающих данных, среднее значение суммы модулей:

$$E\{\varepsilon(\mathbf{w})\} = E\left\{\sum_{=1}^{N} \left| d_i - F(\mathbf{x_i}, \mathbf{w}) \right|\right\} \qquad (3)$$

и максимальное значение отклонений:

$$e = \max_{i=1\ldots N} \left| d_i - F(\mathbf{x_i}, \mathbf{w}) \right| \qquad (4)$$

Таким образом, разработанный пакет программ “Essence” позволяет автоматизировать процесс идентификации моделей объектов при измерении параметров нелинейных динамических систем со стационарными и нестационарными параметрами.

В настоящее время авторы доклада занимаются разработкой технологии создания гетерогенных распределённых нейронных сетей. Под гетерогенными сетями при этом понимают сети, обладающие следующими свойствами:

- сети обрабатывают различный вид входной и выходной информации;
- имеют сложную архитектуру, образованную из нескольких сетей, и сетей-экспертов:
- допускают многообразные формы участия человека в принятии решений.

Первым шагом в создании такой системы является Essence 1.1 – версия функционирующая на отдельном компьютере, структура и принципы построения которой предполагают возможность перехода к распределенному варианту.

В настоящее время ведётся работа над версией, в которой отдельные части системы будут адаптировать свой размер и структуру, а не только коэффициенты во время непрерывного функционирования или в промежутках между поступлением входных данных во время измерений. В качестве элементов сети предполагается брать не только простейшие стандартные нейроны, но и более сложные элементы, в том числе и с динамическим поведением, реализующие известные модели подсистем исследуемого объекта.