

УДК 621.311

## РАЗМЕР ОБУЧАЮЩЕЙ ВЫБОРКИ И ЕЕ ВЛИЯНИЕ НА АРХИТЕКТУРУ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ В ЭНЕРГОСИСТЕМЕ

© В.З. Манусов, И.С. Макаров, С.А. Дмитриев,  
С.Е. Кокин, С.А. Ерошенко

*Ключевые слова:* прогнозирование; искусственные нейронные сети; обучение.

Представлен подход, использующий теорию искусственных нейронных сетей для прогнозирования потребления электрической энергии, при условии минимизации ошибки прогнозирования на длительном интервале упреждения. Одновременное выполнение условий небольшого отклонения прогнозируемых значений от реальных значений и сохранение квадрата ошибки в заданных пределах на предопределенном интервале достигается за счет сбалансированного выбора архитектуры нейронной сети. Испытания метода проводились с использованием реальных данных.

### ВВЕДЕНИЕ

Реформирование энергетической отрасли повлекло введение новых принципов функционирования рынков *электроэнергии* (ЭЭ) и существенное изменение договоров энергоснабжения и купли продажи. Энергоснабжающие организации и потребители (субъекты розничного рынка электроэнергии) в сложившихся на данный момент условиях заинтересованы в достоверном планировании (прогнозировании) потребностей в ЭЭ и проведении эффективного анализа по ее использованию. Вследствие этого положения существенно возрастает ответственность в достоверном решении задач анализа и прогнозирования режимных параметров (в первую очередь, электрической нагрузки потребителей, перетоков мощности и т. д.) и характеристик (потерь электроэнергии, температуры наружного воздуха и т. д.) для каждой конкретной *энергоснабжающей организации* (ЭСО) [1].

*Технологии искусственного интеллекта* (ТИИ) – одна из возможностей повышения качества решения задач прогнозирования, среди которых наибольший интерес представляют *искусственные нейронные сети* (ИНС).

### 1. ПОСТАНОВКА ЗАДАЧИ

Решение задач прогнозирования, как правило, осуществляется с помощью регрессионных моделей, позволяющих связать известное суммарное потребление электроэнергии с шин питающей подстанции с потреблением электроэнергии исследуемым потребителем. Однако в настоящее время большое распространение получили системы, базирующиеся на принципах работы живых организмов, в частности, моделирующие их нервную деятельность. Такие системы получили название ИНС.

В литературе имеется большое число публикаций по использованию ИНС для прогнозирования. Эти публикации охватывают широкий спектр прогностиче-

ских применений, варьируя от финансово-экономических до физических явлений) [2].

Рассмотрим следующую задачу: крупный промышленный потребитель электрической энергии ( $P_{иссл}$ ), питающийся с шин подстанции (ПС), в целях обеспечения процесса планирования производства и поставок электрической энергии подает заявку на предполагаемый объем электрической энергии своему поставщику электрической энергии. Кроме того, с шин ПС питается жилой сектор и ряд мелкопромышленных потребителей, эквивалентированных как  $P_{ст}$ .

Учет суммарного расхода ЭЭ осуществляется вручную лишь раз в месяц. В этом случае при выполнении прогноза мы сталкиваемся с ситуацией, когда последний достоверно известный член временного ряда датируется последним числом предыдущего месяца. Поэтому при выполнении прогноза следует опираться на данные, поступающие онлайн, получаемые с шин питающей подстанции.

Целью работы является разработка комплекса оперативного прогнозирования на основе ИНС, отвечающего следующим требованиям:

- прогнозирование потребления ЭЭ на каждый час суток;
- выполнение прогноза на 2-е, 3-е и 4-е сутки вперед;
- минимизация ручных манипуляций при получении прогноза;
- удобная форма предоставления прогноза;
- небольшое отклонение прогноза от реальных значений.

### 2. АНАЛИЗ ЗАВИСИМОСТИ ОШИБКИ ОТ ДЛИНЫ ОБУЧАЮЩЕЙ ВЫБОРКИ

В данном случае прогнозирование сводится к задаче, в которой некоторому входному вектору  $X$ , составленному из последних известных данных по общему потреблению с шин ПС, посредством преобразования с

помощью ИНС ставится в соответствие некоторый выходной вектор  $Y$ , представляющий прогнозные значения потребления ЭЭ. Для этого нужно выбрать архитектуру ИНС и обучить ее, решая при этом две противоречивые задачи: уменьшение среднего квадрата ошибки и сохранение квадрата ошибки в заданных пределах на предопределенном интервале.

Обучить ИНС значит минимизировать следующую функцию [3]:

$$F(W, X) = \sum_{i=1}^N (Y_i - d_i)^2 \rightarrow \min, \quad (1)$$

где  $Y_i$  – выходной прогноз сети при подаче примера  $i$ ;  $d_i$  – соответствующее истинное значение потребления ЭЭ;  $N$  – общее количество примеров в множестве обучения. Стоит отметить, что обучение производится на «обучающей выборке», и значения  $d_i$  известны заранее.

Отклик сети  $Y$  получается в результате нелинейного преобразования вектора  $X$ :

$$Y = \varphi \left( \sum_k w_{0k} \varphi \left( \sum_j w_{kj} \varphi \left( \dots \varphi \left( \sum_i w_{li} x_i \right) \right) \right) \right), \quad (2)$$

где  $\varphi(\cdot)$  – функция активации общего вида;  $w_{0k}$  – синаптический вес связи между нейроном  $k$  последнего скрытого слоя и единственным выходным нейроном 0 и т. д. для всех остальных синаптических весов;  $x_i$  – элемент  $i$  входного вектора  $X$ .

Таким образом, значение  $Y$  есть функция от входного вектора  $X$  и вектора синаптических весов  $W$ .

Располагая первыми и вторыми производными целевой функции  $F(W, X)$ , в качестве ее квадратной модели можно взять сумму первых трех членов тейлоровского разложения  $F(W, X)$  в окрестности текущей точки  $W_k$ , т. е. воспользоваться соответствием:

$$F(W_k + \Delta W, X) \approx F(W_k, X) + g_k^{-T} \Delta W + \frac{1}{2} \Delta W^T H \Delta W, \quad (3)$$

где  $g_k$  – вектор локального градиента в точке  $W_k$ :

$$g = \left. \frac{\partial F(W, X)}{\partial W} \right|_{w=\bar{W}_k}; \quad (4)$$

$H_k$  – матрица Гессииана в точке  $W_k$ :

$$H_k = \left. \frac{\partial^2 F(W, X)}{\partial W^2} \right|_{w=\bar{W}_k}. \quad (5)$$

Минимум правой части выражения (3) (если он существует) достигается на векторе  $\Delta W_k^*$ , доставляющем минимум квадратичной форме:

$$\Phi(W) = g_k^{-T} \Delta W + \frac{1}{2} \Delta W^T H_k \Delta W. \quad (6)$$

Будучи стационарной точкой, этот вектор должен удовлетворять равенству:

$$H_k \Delta W_k^* = -g_k \quad (7)$$

или

$$\Delta W_k^* = -H_k^{-1} g_k. \quad (8)$$

В итоге, на итерации  $k$  имеем

$$\Delta W_k = W_{k-1} + \Delta W_k^* = W_{k-1} - H_k^{-1} g_k. \quad (9)$$

Итерации повторяются до тех пор, пока не будет достигнут некий наперед заданный критерий сходимости, например:

$$2 \cdot \frac{|F(W_k, X) - F(W_{k-1}, X)|}{F(W_k, X) + F(W_{k-1}, X)} \leq \varepsilon. \quad (10)$$

В ходе обучения для выбранной архитектуры сети вектор  $W$  примет значение  $W'_{opt}$ , близкое к оптимальному  $W_{opt}$ , в котором функция  $F(W, X)$  достигает глобального минимума.

Оптимальный результат  $W_0$  зависит от количества примеров в обучающей выборке. Действительно, компонент  $j$  вектора  $g_k$  равен:

$$g_{kj} = \sum_{i=1}^N \frac{\partial ((f(W, X_i) - d_i)^2)}{\partial w_j}, \quad (11)$$

а элемент  $(q, j)$  гессииана  $H_k$  равен:

$$H_{k_{qj}} = \sum_{i=1}^N \frac{\partial^2 ((f(W, X_i) - d_i)^2)}{\partial w_q \partial w_j}, \quad (12)$$

где  $N$  – общее число примеров в обучающей выборке.

На каждой из итераций вектор  $\Delta W_k^*$ , равный произведению матрицы, обратной гессииану, на вектор локального градиента также будет функцией от общего числа примеров.

Вектор  $W'_{opt}$  также будет зависеть от числа примеров, поскольку

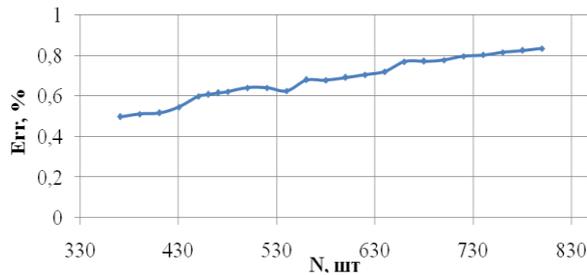


Рис. 1. Изменение средней ошибки на обучающей выборке в зависимости от длины обучающей выборки

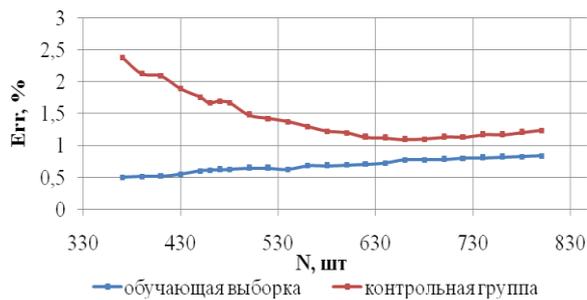


Рис. 2. Изменение ошибки в контрольной группе и ошибки на обучающей выборке в зависимости от длины обучающей выборки ( $M = 168$ )

$$W'_{opt}(N) = W + \sum_{k=1}^S \Delta W_k^*(N), \quad (13)$$

где  $W_0$  – начальное приближение;  $S$  – номер итерации, на которой выполняется критерий сходимости.

Исследуем зависимость значения целевой функции в точке  $W'_{opt}$  от длины обучающей выборки.

Ошибка на обучающей выборке рассчитывается следующим образом:

$$Err_{train} = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - d_i|}{d_i} \cdot 100\%, \quad (14)$$

где  $Y_i$  – выходной прогноз сети при подаче примера  $i$ ;  $d_i$  – соответствующее истинное значение потребления ЭЭ;  $N$  – общее количество примеров в множестве обучения.

Следует обратить внимание на то, что при расчете средней ошибки используется функция модуля, а не квадратичная функция, т. к. в большинстве практических приложений прогнозирования в энергетике более важным является модуль отклонения.

Графически зависимость средней ошибки на обучающей выборке от длины выборки при фиксированной архитектуре сети и одинаковом алгоритме обучения имеет следующий вид (рис. 1).

С ростом длины выборки можно наблюдать тенденцию к росту средней ошибки. Это связано с тем, что также как и в задаче описания случайного процесса, заданного во времени  $N$  точками, полиномом степени

$M$ , при увеличении числа  $P$ , равного  $M - N + 1$ , ошибка описания будет увеличиваться.

Действительно, известно, что для набора, состоящего из  $N$  точек  $(x_0, y_0), (x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)$ , существует интерполяционный полином Лагранжа степени  $N - 1$ , для которого  $L(x_i) = y_i$ . То есть такой полином абсолютно точно проходит через точки данного набора или, по аналогии с нашим случаем, для него  $Err_{train} = 0$ . Однако описать абсолютно точно полиномом степени  $M$  набор из  $N$  точек при условии, что  $M < N + 1$  и что ординаты точек выбираются с некоторой степенью случайности, можно только с очень малой вероятностью. И чем больше разность между количеством точек и степенью полинома, тем меньше эта вероятность и тем больше математическое ожидание среднего отклонения значения интерполяционного полинома от значений ординат точек заданного набора:

$$Err = \frac{1}{N} \sum_{i=1}^N |L(x_i) - y_i|. \quad (15)$$

Несмотря на то, что  $F(W)$  – это трансцендентная функция, а не полином, изложенные выше соображения о связи ошибки на обучающей выборке с разностью между количеством примеров и числом переменных (весов синапсов) применимы также и для нее.

Весь процесс обучения сети построен на минимизации ошибки на обучающей выборке, для того чтобы на дистанции упреждения ошибка была приблизительно мала. Однако уменьшить ошибку можно, сократив длину выборки (рис. 1). Зачастую решение о сокращении обучающей выборки выглядит очень привлекательным. Но при чрезмерном сокращении обучающей выборки разработчик будет сталкиваться с ситуацией, когда значение ошибки при прогнозировании будет сильно отличаться от ошибки на обучающей выборке.

Для исследования этого аспекта из обучающей выборки выделяется т. н. «контрольная группа». Элементы этой группы известны, но не участвуют в обучении, поэтому выделение контрольной группы позволяет получить более точное представление о средней ошибке прогнозирования, которая будет допускаться ИНС.

## ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

Построим зависимость средней ошибки в контрольной группе от длины обучающей выборки.

После процесса обучения сеть тестируется на неизвестных для нее примерах контрольной группы, после чего вычисляется ошибка:

$$Err_{control} = \frac{1}{C} \sum_{i=1}^C \frac{|Y_i - d_i|}{d_i} \cdot 100\%, \quad (16)$$

где  $Y_i$  – выходной прогноз сети при подаче примера  $i$ ;  $d_i$  – соответствующее истинное значение потребления

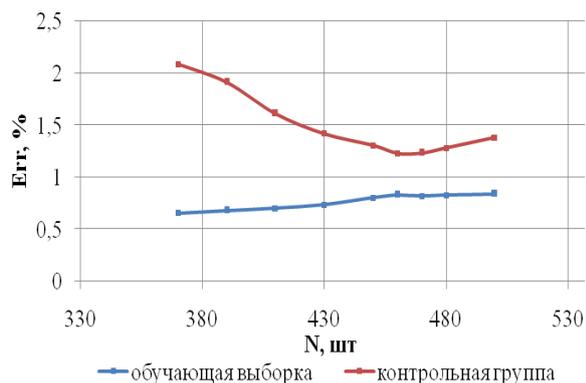


Рис. 3. Изменение ошибки в контрольной группе и ошибки на обучающей выборке в зависимости от длины обучающей выборки ( $M = 120$ )

ЭЭ;  $C$  – общее количество примеров в контрольной группе. В представленном случае  $C = 30$ .

Полученные результаты показывают, что минимум ошибки  $Err_{control\ min} = 1,088\%$  для сети с 168 синапсами и пороговыми значениями достигается при обучении на выборке длиной около 660 примеров.

На практике нередки ситуации, когда разработчик сталкивается с некоторым недостатком данных для обучения. Так, рассмотрим ситуацию, когда доступны лишь 500 обучающих примеров. Минимальная ошибка в контрольной группе сети с 168 синапсами и пороговыми значениями в этом случае будет равна  $Err_{control}|_{N=470} = 1,693\%$  ( $C = 30$ ,  $N = 500 - 30 = 470$ ) (рис. 2), что больше минимального значения почти на 0,6%. Исходя из вышеизложенных соображений, можно предположить, что ИНС с меньшим количеством настраиваемых параметров может показать меньшую минимальную ошибку в контрольной группе. За счет сокращения количества синапсов во входном слое облегчим предыдущую сеть до 120 настраиваемых параметров.

После тестирования получаем следующие результаты (рис. 3).

Минимум ошибки  $Err_{control\ min} = 1,23\%$  для сети с 120 синапсами и пороговыми значениями достигается при обучении на выборке длиной около 460 примеров. Минимуму ошибки в контрольной группе соответствует ошибка на обучающей выборке  $Err_{train}|_{N=460} = 0,83\%$ . При  $N = 460$  ошибка на обучающей выборке сети со 168 настраиваемыми параметрами равна 0,608%, т. е. меньше, чем у сети со 120 параметрами. Однако в

контрольной группе при  $N = 460$  большая сеть имеет ошибку  $Err_{control}|_{N=460} = 1,667\%$  против  $Err_{control\ min} = 1,23\%$  у меньшей сети, поэтому облегченный вариант сети стоит признать лучшим. Данные рассуждения показывают важность контроля ошибки не только на обучающей выборке, но и в контрольной группе.

## ЗАКЛЮЧЕНИЕ

При разработке ИНС необходимо производить проверку качества прогнозирования на примерах контрольной группы, в противном случае, разработчик рискует выбрать вариант сети, далекий от оптимального варианта.

При выборе архитектуры ИНС, количества нейронов и прилегающих синапсов необходимо отталкиваться от общего количества известных примеров, которые можно использовать для обучения.

Зависимость ошибки в контрольной группе от длины обучающей выборки при фиксированной архитектуре ИНС имеет ярко выраженный экстремум. Увеличение числа синаптических весов в ИНС ведет к смещению экстремума в сторону большего числа элементов в обучающей выборке.

## ЛИТЕРАТУРА

1. Томин Н.В. Анализ и прогнозирование режимных параметров и характеристик для субъектов розничного рынка электроэнергии на базе технологий искусственного интеллекта: автореф. дис. ... канд. тех. наук. Иркутск, 2007.
2. Weizhong Y. Toward automatic time-series forecasting using neural networks // IEEE Transaction on neural networks and learning systems. 2012. V. 23. № 7. P. 1028-1039.
3. Haykin S. Neural Networks: A Comprehensive Foundation. Prentice Hall, 1999. P. 842.

Поступила в редакцию 8 апреля 2013 г.

Manusov V.Z., Makarov I.S., Dmitriyev S.A., Kokin S.E., Eroshenko S.A. SAMPLE SIZE OF TRAINING AND ITS IMPACT ON ARCHITECTURE OF ARTIFICIAL NEURAL NETWORKS IN POWER SYSTEMS

An approach that uses the theory of artificial neural networks to predict the consumption of electric energy, while minimizing the prediction error on long range preemption is presented. The simultaneous fulfillment of the conditions of a small deviation of the predicted values from the actual values and the preservation of squared errors in the set limits at a predetermined interval is achieved through well-balanced select of architecture of the neural network. The tests were conducted using the method of real data.

*Key words:* forecasting; artificial neural network; training.