

Проблемы формирования входного вектора для нейросетевых систем прогнозирования электропотребления

Р. Р.Беглов, К. Н.Воронов, Б.А.Гнатюк

В статье представлены результаты определения минимального набора значимых факторов, из которых формируется входной вектор размерностью 8 и 12 для обучения нейронных сетей. Определен синтезированный входной вектор, который даёт возможность прогнозировать почасовое потребление электроэнергии независимо от вида графиков потребления, нагрузочных диаграмм, часовых поясов и погодных условий.

Наиболее универсальными являются методы прогнозирования электроэнергии, в которых используются искусственные нейронные сети. Для создания таких систем необходимо определить состав входного вектора для обучения нейросетей. Рассмотрено формирование входного вектора на основе минимальный набор значимых факторов. Проведены экспериментальные исследования прогнозирования при использовании входного вектора разной размерности. Определен состав входного вектора, который даёт возможность прогнозировать почасовое потребление электроэнергии независимо от вида графиков потребления, и погодных условий.

Задача планирования потребления электроэнергии всегда была важна для правильного расчета процессов ее приобретения и расходования, а сейчас стала весьма актуальной в связи с преобразованиями в энергетике. Поскольку электроэнергия ключевым продуктом для экономики и составляет весомую долю в расходах практически всех ее отраслей. Переход к оптовому рынку электроэнергии и мощности – ФОРЭМ, повысил актуальность прогнозирования потребления, но также вызывал проблемы, связанные с точностью ужесточив требования по скорости и надежности прогнозов. Для этих целей наиболее востребованы прогнозы почасового потребления электроэнергии на ближайшие сутки.

Наиболее перспективными из большого разнообразия применяемых методов прогнозирования являются те, в которых используются искусственные нейронные сети (ИНС) [1]. Однако разработка таких систем прогнозирования связана с необходимостью решать слабо формализуемых задач: определение состава входного вектора для обучения ИНС, выбор типа ИНС, расчет необходимого объема статических данных – обучающей выборки.

В ходе проведенного исследования решались первоочередная задача – определение минимального количества факторов, необходимых для прогнозирования, и формирование на этой основе входного вектора.

Для эксперимента были выбраны следующие переменные: статические переменные – день недели (понедельник, вторник, ... воскресенье), месяц (январь, февраль, ... декабрь), час (1,2, ... ,24); динамические переменные – почасовое потребление электроэнергии.

Прогнозирование, рассмотренное в [2], предполагает что динамические переменные должны идти со сдвигом во времени, таким образом, чтобы отражать динамику изменения потребления электроэнергии в предшествующие дни. При почасовом прогнозировании электроэнергии на сутки вперед, учитывая, что график потребления электроэнергии имеет выраженный недельный цикл, возможный сдвиг во времени должен варьироваться от одних до восьми суток. Результаты проведенных экспериментальных исследований представлены в Таблице 1. Эксперименты проводились на нейронной сети с прямым распределением и задержкой по времени [3].

Из анализа полученных данных следует, что чем больше «откат» во времени, тем меньше погрешность прогнозирования, однако при этом увеличивается размерность входного вектора обучающих переменных и, как следствие, в квадратичной зависимости увеличивается время обучения нейронных сетей. В результате экспериментов установлен, что оптимальным является сдвиг на пять суток назад. Так как при меньшем сдвиге существенно снижается точность прогнозирования, а при большем временном откате или сдвиге вектор становится слишком большим, и возникают проблемы с обучением нейронных сетей.

Таблица 1

Длительность сдвига (сутки)	Количество и % попадания в интервалы ошибок			
	0-3%	3-5%	5-10%	>10%
1	341 47,36 %	175 24,3 %	182 25,27 %	22 3,055 %
3	335 49,3 %	181 25,14 %	190 26,39 %	14 1,94 %
5	368 51,11 %	178 24,72 %	156 21,6%	18 2,5%
7	378 52,5 %	181 25,14 %	147 20,41 %	14 1,94 %
8	396 55 %	184 25,55 %	135 18,75 %	5 0,69 %

Таким образом, получается компактный входной вектор, состоящий из 8 переменных:

$$\{X_{N-1} \ X_{N-2} \ X_{N-3} \ X_{N-4} \ X_{N-5} \ T \ D \ M\},$$

где $X_{N-1} \ X_{N-2} \ X_{N-3} \ X_{N-4} \ X_{N-5}$ – данные часового потребления электроэнергии со сдвигом на пять дней;

$T [1, 2, 3 \ 4, \dots 24]$ – переменная «час»;

$D [1, 2, 3 \ 4 \ 5 \ 6 \ 7]$ – переменная «день недели» (1 – понедельник; 2 – вторник, ..., 7 – воскресенье);

$M [1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12]; \}$ - переменная «месяц» (1 - январь; 2 – февраль, ..., 12 – декабрь)

При этом не требуется операция масштабирования переменных. Подчеркнём, что среди входных переменных нет показателей погоды, характеристики графиков нагрузки (пики, полу-пики, провалы). Отказ от таких переменных сделан для придания системе прогнозирования универсальности и независимости от часовых и климатических поясов.

Также экспериментальным путем было установлено (см. Таблицу 2), что для повышения точность прогнозирования необходимо в обучающем векторе продублировать переменную «час» таким образом, чтобы количество дублирований совпадало с количеством динамических переменных почасового потребления электроэнергии. В этом случае получаем входной вектор размерностью 12.

Таблица 2

Количество переменных входного вектора	Количество и % попадания в интервалы ошибок			
	0-3%	3-5%	5-10%	>10%
8	74 44,04	49 29,16	40 23,8	5 2,97
12	87 51,78	47 27,98	34 20,24	0 0

Была проверен следующий набор переменных: статические переменные – день недели (понедельник, вторник, ... воскресенье), месяц (январь, февраль, ... декабрь), час (1,2, ... ,24); динамические переменные – почасовое потребление электроэнергии со сдвигом на трое суток, таким образом, чтобы месяц и день недели исторических данных соответствовал дню и месяцу прогнозируемых данных. Также данные почасового потребления прошлого года формируются таким образом, чтобы месяц и день недели исторических данных соответствовал дню и месяцу прогнозируемых данных.

В результате получается входной вектор размерностью 9:

X_N – данные почасового потребления электроэнергии предшествующего года (час, месяц и день недели соответствуют часу, дню и месяцу прогнозируемых данных);

$X_{N-1}\ X_{N-2}\ X_{N-3}$; – данные часового потребления электроэнергии текущего года (со сдвигом на три дня)

$T [1,2,3, \dots, 24]; [1,2,3, \dots, 24]; [1,2,3, \dots, 24]$ – переменные «час»;

$D [1,2,3,4,5,6,7]$ – переменная «день недели»;

$M [1,2,3,4,5,6,7,8,9,10,11,12]$ – переменная «месяц».

Исследования показали, что такой вектор при использовании нейронной сети с прямым распределением и задержкой по времени дает погрешность прогнозирования, занимающую меньше чем при 8 размерном векторе, но больше чем при 12 размерном векторе.

В ходе практических испытаний разработанной нейросетевой системы прогнозирования выяснилось, что отдельной проблемой является получение прогнозов в предпраздничные и праздничные дни. Эти дни не совпадают по

дням недели с текущими выходными днями, на данных потребления электроэнергии которых обучались нейронные сети. Для повышения точности прогнозирования в эти дни необходимо модифицировать входной вектор. Так как в праздничные дни потребление электроэнергии варьируется и отличается от будней, было принято решение ввести дополнительную тринадцатую переменную, где час праздничного дня обозначается как 1, а час будничного дня как 0. Результаты экспериментов с модифицированным входным вектором представлены в Таблице 3.

Таблица 3

Название нейронной сети	Количество и % попадания в интервалы ошибок			
	0-3%	3-5%	5-10%	>10%
Двухслойная Нейросеть Элмана	555 77.08	119 16.52	41 5.69	5 0.69
Нейросеть с прямым распределением и задержкой по времени	547 75.97	125 17.36	41 5.69	7 0.97
Нейросеть с каскадным распределением	534 74.16	121 16.8	54 7.5	11 1.52

Прогнозирование электроэнергии для 2015 года стало лучше в праздничные дни, но хуже в предпраздничные дни. Так же стоит отметить, что прогноз потребления в новогодние праздники значительно меньше чем реальное потребление. Из чего следует, что каждый час с 1 по 9 января необходимо отмечать в векторе определения отдельной цифрой. Поэтому было принято при обучении обозначать каждый час с 1 по 9 января дня как 3, час праздничного дня – как 1, а час будничного дня – как 0.

Однако, эксперименты показали, что прогнозирование электроэнергии стало хуже в новогодние праздники. Следовательно, каждый час с 1 по 9 января не нужно отмечать как праздничный день при прогнозировании, но при обучении необходимо отмечать праздничным днем. Получим, что тринадцатая переменная должна при обучении с 1 по 9 января принимать значение 1, в праздничный день принимать также значение 1, в будничные и обычные воскресные дни – значение 0. При прогнозировании эта переменная меняется по-другому: с 1 по 9 принимает значение 0, в часы праздничного дня принимает значение 1, в часы будничного и обычного воскресного дня – опять значение 0. Результаты экспериментов представлены в Таблице 4.

Таблица 4

Название нейронной сети	Количество и % попадания в интервалы ошибок			
	0-3%	3-5%	5-10%	>10%
Нейросеть с каскадным распределением	543 75.41	127 17.63	45 6.25	5 0.69
Двухслойная нейросеть Элмана	565 78.47	121 16.8	34 4.72	0 0
Нейросеть с прямым распределением и	563 78.19	125 17.36	32 4.44	0 0

задержкой по времени				
----------------------	--	--	--	--

Отличие входного вектора при обучении и при прогнозировании, как правило, не применяется. В данном случае такой прием применен, чтобы улучшить способность нейросети к обобщению, т.е. чтобы нейросеть не переучивалась. Такая модификация дала следующий эффект. Прогнозирование электроэнергии улучшилось и в праздничные дни, но не изменилось в предпраздничные дни.

Что бы уменьшить погрешность прогнозирования в предпраздничные, дни модифицируем вектор, обозначив часы предпраздничного дня значением переменной 2, часы праздничного дня – значением переменной 1, а часы будничного дня – значением переменной 0. Экспериментальная проверка показала (Таблица 5), что точность прогнозирования в праздничные и предпраздничные дни заметно увеличилась, но снизилась общая эффективность прогнозирования. Увеличилось количество неудовлетворительных и удовлетворительных погрешностей и уменьшилось количество хороших и отличных.

Таблица 5

Название нейронной сети	Количество и % попадания в интервалы ошибок			
	0-3%	3-5%	5-10%	>10%
Двухслойная нейросеть Элмана	506	144	61	9
	70.27	20	8.47	1.25
Нейросеть Элмана	573	95	46	6
	79.58	13.19	6.38	0.83
Нейросеть с прямым распределением и задержкой по времени	561	118	36	5
	77.91	16.38	5	0.69

Из этого следует, что для увеличения общей эффективности прогнозирования с учетом данных предыдущего опыта, необходимо не отмечать предпраздничный день при прогнозировании, но отмечать их при обучении нейронных сетей. Окончательно получается, что значение тринадцатой переменной модифицированного вектор необходимо устанавливать при обучении для предпраздничного дня – равным 2, для праздничного дня – равным 1, а для будничного дня – равным 0. При процедуре прогнозирования необходимо тринадцатую переменную установить для предпраздничного дня равной 0, для праздничного – равной 1, для будничного дня – равной 0.

Экспериментальная проверка, результаты которой приложены в Таблице 6, показала, что данная модификация входного вектора наиболее результативна.

Таблица 6

Название нейронной сети	Количество и % попадания в интервалы ошибок			
	0-3%	3-5%	5-10%	>10%
Двухслойная нейросеть Элмана	519	142	55	4
	72.08	19.72	7.63	0.55

Нейросеть Элмана	584 81.11	95 13.19	40 5.55	1 0.13
Нейросеть с прямым распределением и задержкой по времени	571 79..3	119 16..52	30 4.16	0 0

ВЫВОДЫ

Определен минимальный набор значимых факторов, из которых формируется входной вектор размерностью 8 и 12 для обучения нейронных сетей. Модификация этого вектора путем добавления тринадцатой составляющей, позволило существенно повысить точность прогнозирования в предпраздничные и праздничные дни. Особенностью этой составляющей является то, что она изменяется по-разному при процессе обучения и процессе прогнозирования. Синтезированный входной вектор даёт возможность прогнозировать почасовое потребление электроэнергии независимо от вида графиков потребления, нагрузочных диаграмм, часовых поясов и погодных условий. Это делает возможным создание универсального инструмента прогнозирования электропотребления.

Список литературы

- 1.Шумилова Г.П., Готман Н.Э., Старцева Т.Б. Прогнозирование электрических нагрузок при оперативном управлении электроэнергетическими системами на основе нейросетевых структур. Сыктывкар: КНЦ УрО РАН, 2008. 76 С.
- 2.Осовский С. Нейронные сети для обработки информации / Пер. с польского И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с.
3. Хайкин С., Нейронные сети: полный курс, 2-е издание.: Пер. с англ. - М.: Издательский дом «Вильямс», 2006. - 1104с