

УДК 681.3.01:621.311

© 2009 г. **Е.С. Анушина**

(Санкт-Петербургский государственный электротехнический университет)

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНЫХ ВЫЧИСЛИТЕЛЬНЫХ ТЕХНОЛОГИЙ

Рассмотрен синтез оптимальных систем управления с использованием. В статье рассматриваются проблемы прогнозирования в технических системах на примере прогнозирования электропотребления. Предложена система прогнозирования на основе нейронечетких систем и вейвлет-теории, что позволяет повысить точность краткосрочного прогноза.

Ключевые слова: прогнозирование электропотребления, временной ряд, вейвлет-теория, нейронечеткие модели.

Введение

Задача прогнозирования является на сегодняшний день одной из наиболее актуальных как для технических, так и для социально-экономических систем, решение ее является условием их эффективного и надежного функционирования.

Для обеспечения наиболее качественного прогноза необходимо иметь адекватную математическую модель объекта. В случае отсутствия такой модели самый распространенный подход – анализ временных рядов, которые отражают поведение системы в прошлом и дают возможность спрогнозировать ее будущее поведение.

Задача прогнозирования формулируется следующим образом: зная предыдущие l значений временного ряда $x(t)$, предсказать последующее значение $x(t + d)$

$$x(t + d) = \Psi(x(t), \dots, x(t - l), \eta_1, \eta_2, \dots, \eta_q),$$

где d – шаг прогнозирования; q – количество независимых переменных η ; $\Psi(\cdot)$ – прогнозирующая функция [1].

За последнее время большие успехи в прогнозировании временных рядов были получены с использованием интеллектуальных средств обработки данных [2]. К ним, в частности относятся искусственные нейронные сети, объединение которых с нечеткой логикой и генетическими алгоритмами позволяет получать новые эффективные решения задачи прогнозирования.

Известно, что задача обучения нейронной и нейронечеткой сетей (сети на основе алгоритма Такаги-Сугено [3]) заключается в минимизации ошибки обобщения, т.е. достижении минимальной погрешности при решении прикладной за-

дачи, когда фактические исходные данные отличаются от данных обучающей выборки. Уменьшение ошибки обучения в нейронной сети обеспечивается в результате увеличения числа нейронов в скрытых слоях, в нейронечеткой сети – путем увеличения числа нечетких правил.

Согласно результату, приведенному в работе [4], для формирования произвольной непрерывной функции с помощью нейронной сети необходимо организовать сеть с двумя скрытыми слоями.

Для увеличения числа нечетких правил в нейронечетких сетях следует увеличить число функций принадлежности входных переменных при обучении с использованием гибридного алгоритма обучения, либо уменьшить радиус кластеров при обучении на основе метода адаптивного алгоритма самоорганизации [3].

Как правило, уменьшение ошибки обучения не всегда приводит к уменьшению ошибки обобщения, и при определении параметров сетей требуется оценка погрешности и обучения, и обобщения. Например, при уменьшении радиуса кластеров возрастает число кластеров и нечетких правил. В результате аппроксимация данных обучающей выборки становится более точной, но это достигается увеличением ошибок обобщения и прогноза в приведенных исследованиях.

Необходимое число нейронов в слоях, требуемое число функций принадлежности и оптимальное значение радиуса кластеров определяются экспериментально.

В предлагаемом исследовании для прогнозирования электропотребления использован нейронечеткий подход.

Нечеткий подход прогнозирования. В нечетких системах результат нечеткого условного логического вывода на основе алгоритма Мамдани определяется по

выражению [3]:
$$B' = A' \circ \bigcup_{i=1}^m R_i, \quad \text{где} \quad R_i = A_i \rightarrow B_i, \quad A' = A'_1 \times A'_2 \times \dots \times A'_n$$

$A_i = A_{1i} \times A_{2i} \times \dots \times A_{ni}$; $i = \overline{1, m}$, m – число нечетких правил; A_{ji} , A' – нечеткие множества, соответствующие входным переменным; $j = \overline{1, n}$, n – число входов системы (необходимое число предыдущих значений нагрузки); B_i , B' – нечеткие множества, соответствующие выходной переменной (значение нагрузки $x(t + d)$); R_i – нечеткая логическая операция импликации i -го правила.

Результат прогнозирования [3]:
$$x_k = \frac{\sum_{i=1}^m y_i \left[\prod_{j=1}^n \mu_{A_{ji}}(x_j) \right]}{\sum_{i=1}^m \left[\prod_{j=1}^n \mu_{A_{ji}}(x_j) \right]}, \quad \text{где } y_i \text{ – центр не}$$

четкого множества B_i i -го нечеткого правила; $\mu_{A_{ji}}(x_j)$ – функция принадлежности A_{ji} нечеткого множества x_j – входной переменной.

Для построения прогнозирующей модели необходимо задать нечеткие правила, число и тип функций принадлежности нечетких множеств, способы реализации нечетких логических операций “и”/“или” и импликации.

Нейронечеткий подход прогнозирования. При использовании алгоритма Такаги – Сугено [4] $B_i = f(x_1, x_2, \dots, x_n)$, как правило, $B_i = b_{0i} + b_{1i}x_1 + \dots + b_{ni}x_n$.

$$\text{Результат прогнозирования [3]: } x_k = \frac{\sum_{i=1}^m \left[\prod_{j=1}^n \mu_{Aji}(x_j) \right] (b_{0i} + b_{1i}x_1 + \dots + b_{ni}x_n)}{\sum_{i=1}^m \left[\prod_{j=1}^n \mu_{Aji}(x_j) \right]}.$$

Параметры нечеткой системы на основе алгоритма Такаги-Сугено можно определить путем обучения нейронечеткой сети [3].

Традиционные методы прогнозирования сложных нестационарных временных рядов состоят в разложении последовательности на составляющие, прогнозировании каждой составляющей в отдельности и объединении полученных прогнозов. В общем случае структура временного ряда представляет собой набор следующих компонент: систематическая составляющая, показывающая в среднем характер изменения прогнозируемого явления во времени и называемая трендом; колебания около тренда с возможно изменяющейся частотой; сезонные колебания со значительно большей частотой, чем колебания около тренда, и случайная составляющая. Идентификация этих компонент является особой задачей, и разработано множество способов выделения тренда и сезонной компоненты [5]. Однако в любом случае неизбежна потеря информации, а значит, и ухудшение точности прогнозирования.

В статье предлагается использовать вейвлет-преобразование для выделения разночастотных компонент временного ряда.

Вейвлет-преобразование «Избыточный Хаар»

Выбор способа вейвлет-преобразования для разложения временного ряда на составляющие зависит от особенностей решаемой задачи. В данном случае для задачи прогнозирования дискретных временных рядов был использован так называемый алгоритм избыточного дискретного вейвлет-преобразования (ИДВП). Для реализации ИДВП используется алгоритм «Избыточный Хаар» (или алгоритм «a trous» – «с дырами») [6].

В алгоритме «a trous» производится последовательность сверток сигнала $c = \{c_i\}$ с фильтром $h(l)$

$$c_{i+1}(k) = \sum_{l=-\infty}^{+\infty} h(l)c_i(k + 2^i l).$$

Возрастающая дистанция $2^i l$ между отсчетами объясняет, почему этот алгоритм называется «a trous». В алгоритме «Избыточный Хаар» для свертки ис-

пользуется простой фильтр $h = (0,5;0,5)$:

$$c_j(k) = \frac{1}{2}(c_{j-1}(k) + c_{j-1}(k - 2^j)).$$

Далее, из последовательностей сглаженных сигналов берется разность между соседними сигналами для получения вейвлет-коэффициентов

$$w_j(k) = c_j(k) - c_{j-1}(k).$$

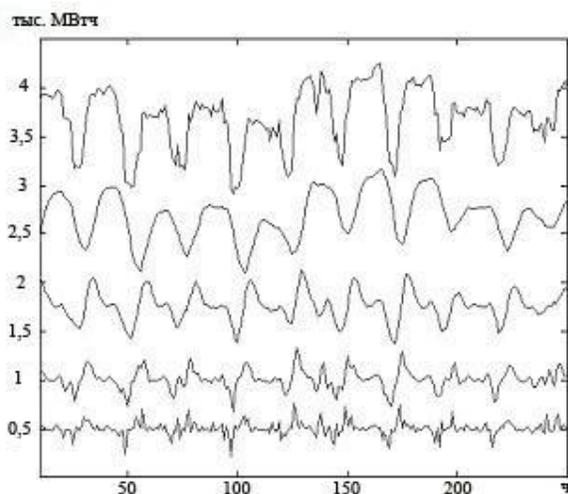


Рис. 1.

временной ряд уровня аппроксимации (медленная динамика) и детализации: 3 (среднетемповая динамика), 2 и 1 (быстрые динамики).

В результате исходный временной ряд можно представить как

$$c_0(t) = x(t) = c_p(t) + \sum_{i=1}^p w_i(t),$$

где p – количество уровней декомпозиции (глубина вейвлет-разложения, или глубина разложения), которые используются для приближения временного ряда; $c_p(t)$ – аппроксимация на уровне p ; $w_i(t)$ – коэффициенты детализации на уровнях $i = 1 \dots p$, соответственно.

Разложение временного ряда с глубиной $p = 3$ показано на рис. 1, где сверху вниз – исходный временной ряд,

Пример прогнозирования

В качестве примера рассматривается временной ряд, описывающий колебания потребления электрической нагрузки для первой ценовой зоны федерального оптового рынка электроэнергии и мощности за рабочие дни февраля 2004 г. Ретроспективные данные за последние девять часов раскладываются по алгоритму «Избыточный Хаар» с глубиной детализации, равной трем. В результате временной ряд разбивается на четыре составляющие (аппроксимирующую и три детализирующие), обладающие различной динамикой и в сумме дающие исходный временной ряд (рис. 1). Прогнозирование осуществляется для каждого уровня разложения в отдельности. Для этого сформирован комитет из четырех нейронечетких сетей (ННС), построенных на основе алгоритма нечеткого логического вывода Такаги-Сугено первого порядка. Далее прогнозы, выполненные каждой из четырех ННС, суммируются и, таким образом, получается итоговое прогнозируемое значение электрической нагрузки [7].

Каждая нечеткая модель имеет некоторое число входов по количеству ретроспективных данных, используемых для прогноза, определенное эмпирически, и один выход; прогнозирование осуществляется на час вперед.

Кроме того, в каждую модель вводится еще один дополнительный вход, идентифицирующий период, на котором производится прогнозирование. Для это-

го временной ряд визуально разбивается на участки возрастания, убывания и относительного постоянства динамики. Каждый такой период обозначается цифрой, которая поступает на вход модели. Такой подход позволяет учесть значимый фактор – номер часа и повысить точность прогнозирования [7].

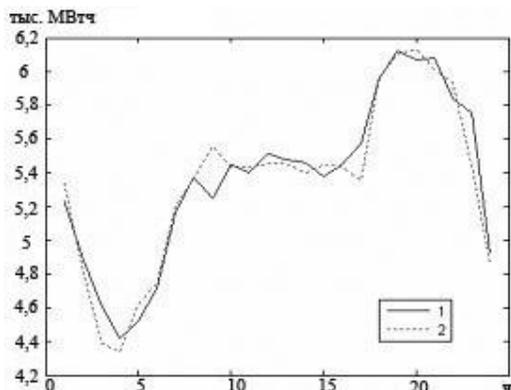


Рис. 2.

где кривая 1 – реальное значение; кривая 2 – прогноз). Точность оценивается как абсолютная средняя ошибка (mean absolute percentage error – MAPE).

Сравнительный анализ построенной системы прогнозирования осуществлялся с регрессионными моделями, а также с системой прогнозирования, построенной на нейронечеткой модели, но без предварительного вейвлет-преобразования.

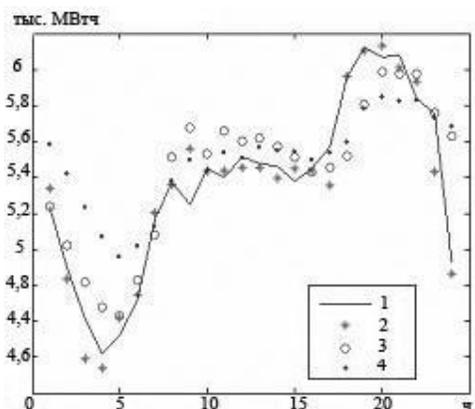


Рис. 3.

неточности входной информации, полученной на основе телеметрических измерений, т.е. оценка работоспособности модели при искажении входной информации. Для этого входная информация системы подвергалась искажениям посредством ограниченного шума.

Данные, полученные в результате прогнозирования, сравнивались с реальными значениями. Результаты исследований приведены в таблице. Как видно из анализа таблицы, сильнее всего искажение данных оказывает влияние на ошибку детализирующих уровней, а особенно – на ошибку уровней детализации 1 и 2, отвечающих быстрой динамике временного ряда.

При искажении входных данных до 4% модель все еще обеспечивает относительно адекватный прогноз.

Результаты исследований показали, что прогнозирование компонент временного ряда с различной динамикой и последующее их суммирование повышают точность производимого прогноза. Средняя точность прогнозирования составляет: для уровня аппроксимации – 0,566%, для детализации на уровне 3 – 3,541%, для детализации на уровне 2 – 4,135 %, для детализации на уровне 1 – 6,269%. Суммарная точность прогнозирования равна 1,68% (рис. 2,

Результаты сравнения приведены на рис. 3, где кривая 1 – фактические данные почасовой нагрузки рабочего дня февраля, кривая 2 – прогноз ННС с использованием вейвлет-преобразования, кривая 3 – прогноз ННС и кривая 4 – прогноз авторегрессионной модели скользящего среднего АРСС(20,10).

Точность прогнозирования для ННС с использованием вейвлет-преобразования – 1,68%; для ННС – 2,1 %; АРСС – 2,8% [8].

Кроме того, были проведены исследования на грубость данного подхода к возможной

Искажение шумом, %	MAPE, % аппроксимация	MAPE, % детализация 3	MAPE, % детализация 2	MAPE, % детализация 1	MAPE, % прогноз
0	0,566	2,541	4,135	6,269	1,867
1,385	0,700	5,222	6,456	10,74	2,690
2,771	0,930	7,239	11,591	13,66	3,959
4,156	1,178	9,810	16,241	24,115	5,228
5,542	1,393	10,439	31,106	54,092	12,361

Затем происходит резкое ухудшение прогнозных качеств модели, когда вейвлет-преобразование с глубиной разложения 3 уже не в состоянии отфильтровать высокочастотную составляющую на уровне аппроксимации и детализации 3.

На рис. 4 приведен результат прогнозирования при искажении входной информации на 4% (кривая 1 – реальные значения, кривая 2 – итоговый прогноз системы). Точность прогнозирования составила 5,1%.

На рис. 5 приведены результаты сравнения разработанной и регрессионной модели относительно устойчивости к возможным неточностям входной информации. Кривая 1 – фактические данные, кривая 2 – прогноз ННС с использованием вейвлет-преобразования, кривая 3 – прогноз ННС и кривая 4 – прогноз АРСС. При искажении исходных данных на 3% точность прогнозирования составила для ННС с использованием вейвлет-преобразования 4%; для ННС – 4,5%, АРСС – 6,3%.

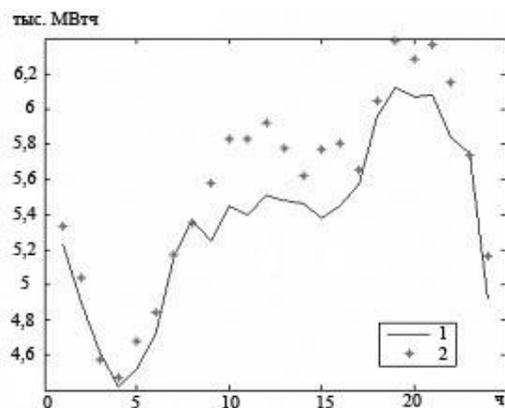


Рис. 4.

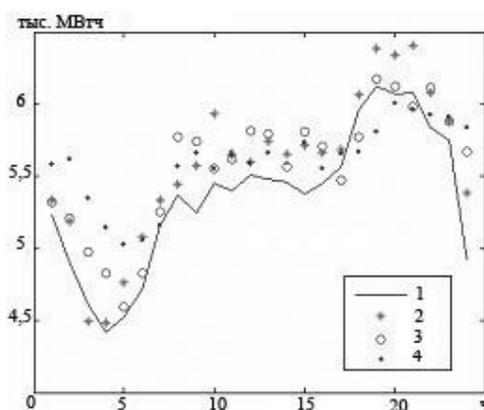


Рис. 5.

Таким образом, можно сделать вывод, что предложенная система прогнозирования характеризуется меньшей погрешностью прогноза по сравнению со статистической моделью и с ННС без использования вейвлет-преобразования. Как было сказано выше, это объясняется тем, что отфильтрованные на высоких уровнях детализации высокочастотные искажения имеют меньший вес (см. рис. 1) и ошибка прогнозирования для этих временных рядов вносит наименьший вклад в общую ошибку прогноза.

Кроме того, повышение точности прогнозирования с использованием вейвлет-преобразования объясняется тем, что нет необходимости в подборе такой архитектуры нейронечеткой сети, которая учитывала бы динамику изменения одновременно быстроменяющейся и медленноменяющейся компонент временного

ряда. Точный учет изменения детальных компонент временного ряда (быстроменяющаяся компонента) дает плохую способность к обобщению сети, что приводит к значительным ошибкам прогнозирования, как и при недостаточном обучении сети (игнорирование быстроменяющейся компоненты). При использовании вейвлет-разложения временного ряда эту проблему удастся частично ослабить.

Заключение

Основной трудностью в задаче прогнозирования сложных временных рядов является одновременное отслеживание разночастотных компонент временного ряда с целью повысить качество прогноза. Для решения этой проблемы разработана система прогнозирования, объединяющая теорию нейронных сетей, нечетких множеств и вейвлет-теорию. Система включает четыре нейронечеткие сети различной структуры, построенные на основе алгоритма нечеткого логического вывода Такаги-Сугено. Прогноз осуществляется на час вперед по данным из предыстории (данные по нагрузке на прошедшие девять часов), а также по номеру часа, для которого осуществляется прогноз. Точность прогнозирования составила 1,68 %. Использование вейвлет-разложения позволило повысить точность прогнозирования в 1,3 раза по сравнению с прогнозом, выполненным с нейронечеткой моделью, и в 1,7 раза по сравнению с прогнозом модели АРСС. Исследование данной системы относительно работы с неточными входными данными показало, что при искажении входных данных до 4% точность прогнозирования сохраняется в пределах 5%.

ЛИТЕРАТУРА

1. *Цукерман Е.В.* Прогнозирование временных рядов. – Казань: Магариф, 1997.
2. *Минаев Ю.Н., Филимонова О.Ю., Бенамеур Л.* Методы и алгоритмы решения задач идентификации и прогнозирования в условиях неопределенности в нейросетевом логическом базисе. – М.: Горячая линия – Телеком, 2003.
3. *Осовский С.* Нейронные сети для обработки информации /пер. с польского И.Д. Рудинского – М.: Финансы и статистика, 2004.
4. *Hecht-Nielsen R.* Kolmogorov's mapping neural network existence theorem // Int. Conf.NN. IEEE Press, 1987. – Vol. 3. – P.11–13.
5. *Попов Л.А.* Анализ временных рядов и прогнозирование. – М.: Изд-во Российской экономической академии, 2005.
6. *Murtagh F., Starck J.-L., Renaud O.* On Neuro-Wavelet Modeling // Decision Support Systems Journal. – 2004. – № 37 (March). – P. 475–484.
7. *Поляхов Н.Д., Приходько И.А., Анушина Е.С.* Прогнозирование электропотребления на основе нейронечеткого подхода с использованием вейвлет-теории // Тр. междунар. конф. по мягким вычислениям и измерениям SCM-2005. – Saint-Petersburg. – 2005. – Vol. 1. – P.247–250.
8. *Анушина Е.С.* Использование интеллектуальных вычислительных технологий для решения задач прогнозирования // Естественные и технические науки. – 2006. – №5. – С.184–186.

Статья представлена к публикации членом редколлегии В.В. Путовым.

E-mail:

Анушина Е.С. – anushinaes@mail.ru