

Некоторые приложения нового метода прогнозирования, основанного на анализе временных рядов*

А.В. Евсеева, асп., Е.Л. Никологорская, канд. хим. наук

Рассматривается новый алгоритм прогнозирования на основе анализа временных рядов, приводятся результаты применения разработанного метода для построения прогноза потребления электроэнергии, предлагается возможность использования методики для прогнозирования вероятности возникновения лесных пожаров.

Ключевые слова: прогноз, временной ряд, фильтр Винера, нейронная сеть, эволюционное моделирование, параллельный алгоритм, гибридный алгоритм, технология CUDA, потребление электроэнергии, лесной пожар.

SEVERAL SUPPLEMENTS FOR NEW FORECASTING METHOD BASED ON ANALYSIS OF TIME SERIES

A.V. Evseeva, Post Graduate Student, E.L. Nikolgorская, Candidate of Chemistry

The article describes a new forecasting method based on the time series analysis, gives the application results of the developed method for forecast building of electricity consumption, describes the possibility of the new method application to forest fire emergence probability.

Key words: forecasting, time series, Wiener filter, neural network, evolutionary modeling, parallel algorithm, hybrid algorithm, CUDA technology, electricity consumption, forest fire.

Расчет прогнозных значений на основе имеющейся информации является актуальной задачей для многих сфер человеческой деятельности. От точности полученного решения зависят как экономическая выгода компаний, работающих на современном рынке предоставления услуг, так и человеческие жизни, если речь идет о прогнозировании тех или иных факторов, определяющих опасность для жизни. Погрешность полученных предсказаний в десятые и сотые доли процента может привести к миллионным убыткам и десяткам потерянных человеческих жизней в случае возникновения чрезвычайной ситуации. Поэтому так важно бороться за каждый процент улучшения того или иного метода.

Математически задача прогнозирования имела следующую формулировку.

Предположим, что случайный вектор \vec{Z} размерности k доступен для наблюдения, а случайная величина X недоступна. Ставится задача угадать значение X по \vec{Z} . Любая функция φ , заданная на k -мерном пространстве, принимающая действительные значения и такая, что мы будем использовать $\varphi(\vec{Z})$ вместо X , называется предиктором X по \vec{Z} .

Другими словами, предиктор – это оценка X по \vec{Z} .

Таким образом, решить задачу прогнозирования значит определить вид предиктора, позволяющего определять значение случайной величины X по \vec{Z} с отклонением из допустимого интервала [4].

Несмотря на все многообразие существующих на сегодняшний день методов построения прогноза, каждый из рассматриваемых подходов в конечном итоге сводится к использованию и обработке временных рядов.

Временной ряд – это последовательность упорядоченных во времени числовых показателей,

характеризующих уровень состояния и изменения изучаемого явления.

В терминологии временных рядов задача построения прогноза имеет следующую формулировку.

Временной ряд $Z = z(t_1), z(t_2), z(t_3), \dots, z(t_N)$ обозначим $Z_t^N = z_1, z_2, z_3, \dots, z_N$. Набор последовательных значений $Z_t^M = z_t, z_{t+1}, z_{t+2}, \dots, z_{t+M-1}$, лежащих внутри временного ряда Z_t^N , назовем выборкой из этого ряда длины M с моментом начала отсчета t , $M \in [1, N-1]$, $t \in [1, N-M]$. Разность начала отсчетов выборок Z_t^M , Z_{t-k}^M назовем задержкой k , $k \in [1, t-1]$. Тогда задача прогнозирования на k тактов вперед заключается в построении выборки Z_t^M на основании значений элементов выборки с задержкой k Z_{t-k}^M .

Вначале в качестве методов прогнозирования были исследованы и реализованы три способа обработки временных рядов: фильтр Винера, нейронная сеть и эволюционное моделирование. Для каждого из указанных методов характерен свой подход к решению поставленной проблемы [1].

После рассмотрения трех базовых подходов к прогнозированию был предложен алгоритм построения гибридного метода на их основе.

При помощи определенного набора коэффициентов $\alpha_i \geq 0$, $\sum_i \alpha_i = 1$, $i = 1, 2, 3$, результаты ра-

боты трех базовых методов было предложено объединить в единый синтетический алгоритм.

Таким образом, принцип действия гибридного алгоритма следующий:

$$p(t) = \alpha_1 f(t) + \alpha_2 n(t) + \alpha_3 m(t),$$

$$\alpha_i \geq 0, \sum_i \alpha_i = 1, i = 1, 2, 3,$$

где $p(t)$ – спрогнозированное значение наблюдаемой величины, полученное с применением гибридного алгоритма; $f(t)$ – значение прогнозируемой величины, полученное с применением фильтра Винера; $n(t)$ – значение прогнозируемой величины, полученное с применением нейронной сети; $m(t)$ – значение прогнозируемой величины, полученное с применением эволюционного моделирования.

Одной из принципиальных и сложных задач является подбор искомого оптимального набора коэффициентов $\alpha_i, i = 1, 2, 3$, для получения наименьшей ошибки прогнозирования синтетического алгоритма.

Чтобы избежать полного перебора параметров, ставится линейная задача оптимизации для нахождения указанных коэффициентов:

$$r(t) - p(t) \rightarrow \min,$$

где $r(t)$ – реальное значение величины потребляемой электроэнергии; $p(t)$ – спрогнозированное значение наблюдаемой величины, полученное с применением гибридного алгоритма.

Задача оптимизации позволяет снизить количество итераций при нахождении необходимого набора коэффициентов $\alpha_i, i = 1, 2, 3$.

Для повышения точности при решении задачи оптимизации принято решение использовать метод наименьших квадратов.

Перейдем от непрерывного изменения времени к дискретному. Пусть в соответствии с гибридным алгоритмом получены прогнозные значения для моментов времени $t_1, t_2, t_3, \dots, t_k$: $p(t_1), p(t_2), \dots, p(t_k)$.

Разность между значением наблюдаемой величины, вычисленным в результате применения метода прогнозирования, и истинным ее значением называется невязкой. В данном случае невязка будет определяться выражением $(r(t) - p(t))$.

Вычислим значение невязки для каждого момента времени:

$$t_1: (r(t_1) - p(t_1));$$

$$t_2: (r(t_2) - p(t_2));$$

$$t_3: (r(t_3) - p(t_3));$$

.....

$$t_k: (r(t_k) - p(t_k)).$$

Введем функцию $F(\alpha_1, \alpha_2, \alpha_3)$, равную сумме квадратов невязок:

$$F(\alpha_1, \alpha_2, \alpha_3) = \sum_{i=0}^n (r(t_i) - p(t_i))^2,$$

где k – число проведенных испытаний по построению прогноза.

Теперь задача оптимизации примет вид

$$F(\alpha_1, \alpha_2, \alpha_3) \rightarrow \min.$$

Для решения поставленной задачи оптимизации предложено использовать метод градиентного спуска, а для избежания попадания в локальные минимумы использовать несколько поисковых точек.

Определив оптимальный набор коэффициентов $\alpha_i, i = 1, 2, 3$, в дальнейшем для построения прогноза необходимо придерживаться следующего алгоритма:

1. Определить момент времени t_i , на который надо выполнить прогноз.

2. Получить прогнозные значения $f(t_i), n(t_i)$ и $m(t_i)$, воспользовавшись базовыми методами: фильтрации Винера, нейронной сетью и эволюционным моделированием соответственно.

3. Вычислить прогнозируемое значение $p(t_i)$, полученное с помощью гибридного алгоритма, применив формулу (1).

Разработанный гибридный алгоритм показал улучшение точности построения прогноза по сравнению с методами, выбранными в качестве базовых [2]. Однако предлагаемый метод прогнозирования имеет достаточно высокую трудоемкость, поскольку в нем приходится неоднократно проводить вычисления трех базовых методов и метода градиентного спуска для нахождения оптимального значения коэффициентов. В связи с этим важным является вопрос ускорения вычислений. В [3] был представлен разработанный метод параллельной организации изложенного алгоритма с использованием системы CUDA.

Оценка полученных результатов проводилась с использованием следующих критериев:

1. Относительной ошибки прогноза

$$\varepsilon_t = \frac{|y_t - \tilde{y}_t|}{y_t} \cdot 100\%,$$

где y_t – фактическое значение показателя энергопотребления на момент времени t ; \tilde{y}_t – значение показателя в момент времени t , полученное с применением прогнозной модели.

2. Средней абсолютной ошибки в процентах

$$MAPE = \frac{1}{Q} \sum_{i=1}^Q \frac{|y_t - \tilde{y}_t|}{y_t} \cdot 100\%,$$

где Q – количество значений временного ряда, попавших в контрольный период; y_t и \tilde{y}_t – фактическое и спрогнозированное значения временного ряда в момент времени t соответственно.

Представленная модель была применена для получения почасового прогноза потребления электроэнергии на 24 часа вперед для г. Костромы и Костромской области и для г. Иванова.

1. *Кострома и Костромская область.* Значения ошибки прогнозирования с помощью гибридного алгоритма по показателю MAPE, максимальное и минимальное значения относительной ошибки для каждого месяца года даны в табл. 1.

Разработанный гибридный алгоритм показал высокую точность прогнозирования. Среднее значение абсолютной ошибки за месяц составило 2,139, что является высоким результатом для прогнозной модели данной предметной области. Минимальное отклонение за месяц наблюдается в декабре и январе, максимальное – в мае и июне. Среднее значение абсолютной ошибки в будни составило 1,998 %, в

выходные дни – 2,481 %. В будни максимальное значение показателя MAPE пришлось на август (2,434 %), минимальное – на апрель (1,564 %). Максимальное значение абсолютной ошибки наблюдалось в выходные дни мая (3,242 %).

2. Ивановская энергосбытовая компания.

По Ивановской энергосбытовой компании были исследованы данные за три месяца (январь, февраль и март) для ряда фидеров (607А, 609А, 612А) и для системы в целом.

Значения ошибки прогнозирования с помощью гибридного алгоритма по показателю MAPE, максимальное и минимальное значения относительной ошибки для указанных месяцев даны в табл. 2.

Анализ полученных данных (табл. 2) показывает, что полученный результат удовлетворяет заявленным требованиям нахождения средней абсолютной ошибки в пределах 5 %. При этом для энергосистемы в целом результаты намного лучше, чем показатели по отдельным фидерам, так как колебания, происходящие на отдельных фидерах, сглаживаются при объединении результатов. Но даже по отдельным фидерам данный показатель превышен лишь для фидера 607А в будни и выходные дни марта и для фидера 612А в выходные, что свидетельствует о высокой устойчивости метода.

Кроме того, данный подход может быть использован как для реализации системы прогнозирования энергопотребления, так и для временных рядов показателей, отличающихся наличием периодической со-

ставляющей в характере поведения, из других предметных областей, например, для прогнозирования вероятности возникновения лесных пожаров.

Под лесными пожарами понимается стихийное, неуправляемое распространение огня по лесным площадям. Отличие природных пожаров от других чрезвычайных катастроф природного характера заключается в их регулярности и большой территории распространения. Лесные пожары могут наблюдаться почти в любое время года. Основными причинами их возникновения является деятельность человека, грозовые разряды, самовозгорания торфяной крошки и сельскохозяйственные палы в условиях жаркой погоды или в так называемый пожароопасный сезон (период с момента таяния снегового покрова в лесу до появления полного зеленого покрова или наступления устойчивой дождливой осенней погоды).

Ежегодно по всему миру данное явление наносит экологический и экономический ущерб, а также уносит человеческие жизни. Для России эта тема особенно актуальна в связи с широкой распространенностью лесных угодий и их хозяйственной значимостью. В связи с вышеизложенным является очевидной необходимость создания и использования методов прогнозирования возникновения лесных пожаров, которые бы позволили пожарным службам взять развитие ситуации под свой контроль и принять своевременные меры по предупреждению катастрофы.

Таблица 1. Результаты прогнозирования энергопотребления для г. Костромы и Костромской области

Месяц	MAPE, %			Минимальное отклонение, %	Максимальное отклонение, %
	будни	выходные	будни и выходные		
январь	1,800	1,838	1,811	0,010	7,915
февраль	1,579	2,650	1,885	0,022	7,249
март	1,790	2,438	1,975	0,006	8,884
апрель	1,564	3,008	1,977	0,008	7,860
май	2,311	3,242	2,577	0,068	10,192
июнь	2,386	3,004	2,606	0,010	10,303
июль	2,373	2,561	2,427	0,030	8,011
август	2,434	2,634	2,471	0,024	7,824
сентябрь	2,057	2,575	2,205	0,076	9,956
октябрь	2,012	2,347	2,122	0,001	9,242
ноябрь	1,753	1,976	1,816	0,027	8,070
декабрь	1,921	1,499	1,800	0,039	6,782
среднее	1,998	2,481	2,139	0,027	8,530

Таблица 2. Результаты построения прогноза для г. Иваново

Месяц	Фидер	MAPE, %		Минимальное отклонение, %		Максимальное отклонение, %	
		будни	выходные	будни	выходные	будни	выходные
январь	607А	2,755	3,292	0,024	0,040	7,673	12,814
	609А	2,253	3,973	0,035	0,010	9,068	16,381
	612А	2,872	2,785	0,074	0,013	10,225	9,819
	система в целом	2,094	3,156	0,012	0,113	8,959	7,426
февраль	607А	3,068	4,688	0,001	0,208	10,326	11,937
	609А	3,654	3,083	0,069	0,268	11,056	9,651
	612А	2,981	6,109	0,018	1,079	18,911	11,963
	система в целом	2,459	2,824	0,029	0,278	7,516	11,996
март	607А	7,421	5,370	0,141	0,419	25,565	16,991
	609А	3,525	3,771	0,037	0,082	8,831	10,810
	612А	4,172	5,004	0,043	0,083	12,038	15,229
	система в целом	2,611	3,610	0,038	0,022	7,134	11,908

Для применения разработанной методики в решении указанной задачи наблюдаемый временной ряд будут составлять показатели пожарной опасности, рассчитываемые по формуле Нестерова [5], изменяющиеся с течением времени.

Список литературы

1. **Никологорская А.В., Сидоров С.Г.** Опыт прогнозирования энергопотребления в энергосетях Костромской области // Высокие технологии, исследования, промышленность. Т. 4: сб. тр. IX Междунар. науч.-практич. конф. «Исследование, разра-

ботка и применение высоких технологий в промышленности». – СПб.: Изд-во Политехн. ун-та, 2010. – С. 266–270.

2. **Евсеева А.В.** Прогнозирование энергопотребления при помощи гибридного алгоритма с применением технологии CUDA // Вестник ИГЭУ. – 2011. – Вып. 3. – С. 47–49.

3. **Никологорская А.В., Ясинский Ф.Н.** Построение гибридного метода прогнозирования энергопотребления с использованием суперкомпьютера на графических ускорителях // Вестник ИГЭУ. – 2010. – Вып. 4. – С. 89–91.

4. **Дронов С.В.** Многомерный статистический анализ: учеб. пособие. – Изд-во Алт. гос. ун-та, 2006.

5. **Нестеров В.Г.** Горимость леса и методы ее определения. – М.; Л.: Гослесбуиздат, 1949.

Работа выполнена при поддержке Министерства образования и науки РФ ГК № 13 G25.31.0077.

*Евсеева Анна Владимировна,
ФГБОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина»,
аспирант кафедры высокопроизводительных вычислительных систем,
телефон (4932) 26-98-29.*

*Никологорская Елена Леонидовна,
ФГБОУВПО «Ивановский государственный химико-технологический университет»,
доцент кафедры высшей математики,
телефон (4932)32-72-56.*