УДК 004.056

Экспериментальное исследование метода прогнозирования, основанного на универсальных кодах

П.А. Приставка

В статье предлагается и экспериментально исследуется метод прогнозирования, основанный на универсальных кодах. На примерах прогнозирования числа солнечных пятен и солнечного излучения показывается, что указанный метод обладает высокой точностью прогноза.

Ключевые слова: прогнозирование, универсальные коды, временны́е ряды, солнечная активность.

1. Введение

На сегодняшний день задача прогнозирования является актуальной для множества областей человеческой деятельности. К их числу можно отнести, например, сейсмологию, финансовую аналитику, геологию, медицину, изучение физических явлений и т.д.

Следует отметить, что на практике очень часто априорные сведения о распределении вероятностей исходов прогнозируемого процесса отсутствуют и это затрудняет решение поставленной задачи. В таком случае можно воспользоваться оценками указанных величин. Одним из основных подходов для получения таких оценок является использование статистических методов, построенных с помощью анализа взаимосвязи последовательных исходов реализации процесса и выявления закономерностей. В конце 1980-х был предложен способ, использующий взаимосвязь задачи прогнозирования временны́х рядов и универсальных кодов и описывающий применение последних для построения методов данного типа, см. [1, 2]. Фактически стало ясно, что универсальные коды представляют собой средство, позволяющее обнаруживать скрытые периодичности и зависимости в данных. Данная статья посвящена исследованию результатов практического применения указанного подхода для решения задач прогнозирования, т.к на сегодняшний день подобные работы ([3, 4]) остаются немногочисленными и, фактически, данный вопрос всё ещё остаётся малоизученным.

Целью данной работы является построение, реализация и экспериментальная оценка метода, предложенного в [5], в основе которого лежит универсальный код R, описанный в [1]. Выбор именно этого кода объясняется тем, что он асимптотически оптимален (однако важно отметить, что его практическая точность была неизвестна).

В качестве объекта исследования были взяты показатели числа солнечных пятен и солнечного излучения, и рассматривалась задача, когда по N известным значениям соответствующего временно́го ряда необходимо спрогнозировать следующее, (N+1)-ое, значение процесса. При этом, для того чтобы иметь возможность сравнить результаты, полученные с помощью рассматриваемого метода, с соответствующими реально зафиксированными величинами процесса, как период основания, так и период упреждения прогноза были взяты из на самом деле уже прошедшего временно́го интервала. Следует отметить, что задача прогнозирования солнечной активности (СА) представляет большой практический и теоретический

интерес ввиду её влияния на многие процессы, идущие на Земле. Так, например, в настоящее время широко исследуются и выявляются статистические связи погоды и климата с СА [6]. При этом, согласно [7], недостаточность знаний о механизмах, определяющих поведение СА, оставляют актуальным вопрос точности соответствующих прогнозов.

Полученные в ходе работы экспериментальные результаты показывают, что методы прогнозирования, основанные на применении универсальных кодов, обладают достаточно высокой точностью.

2. Описание метода и разработанного алгоритма

2.1. Описание метода

Здесь будет описан метод прогнозирования из [5], который был реализован и исследовался в рамках данной статьи, а также приведена необходимая теоретическая информация.

Сначала кратко сформулируем задачу прогнозирования, которая рассматривалась в данной работе. Пусть есть некоторый стационарный и эргодический источник, который порождает последовательности $x_1x_2...$ элементов из некоторого множества A, которое может быть как конечным множеством, так и некоторым непрерывным интервалом. Предполагается, что распределение вероятностей этого источника $P(x_1 = a_{i_1}, x_2 = a_{i_2}, ..., x_t = a_{i_t})$ (или плотность $p(x_1, x_2, ..., x_t)$) неизвестно. Пусть источник порождает сообщение вида $x_1...x_{t-1}x_t$, $x_i \in \mathrm{A}$ для всех i, и следующий элемент требуется спрогнозировать.

Теперь перейдём к описанию метода прогнозирования.

По определению, код U называется универсальным, если для любого стационарного и эргодического источника P верны следующие равенства:

$$\lim_{t\to\infty} |U(x_1...x_t)|/t = H(P)$$

с вероятностью 1 и

$$\lim_{t\to\infty} |E(U(x_1...x_t)|/t = H(P) \ ,$$

где $H(P)$ – энтропия, а $E(f)$ – среднее значение f.

Теорема 1. *Пусть U – универсальный код и*

$$\mu_U(u) = 2^{-|U(u)|} \Big/ \sum_{\upsilon\in \mathrm{A}^{|u|}} 2^{-|U(\upsilon)|} .$$

Тогда для любого стационарного и эргодического источника P верны следующие равенства:

$$\lim_{t\to\infty} \frac{1}{t}(-\log P(x_1...x_t) - (-\log \mu_U(x_1...x_t))) = 0 ,$$

с вероятностью 1, и

$$\lim_{t\to\infty} \frac{1}{t} \sum_{u\in \mathrm{A}^t} P(u) \log(P(u)/\mu_U(u)) = 0$$

Доказательство теоремы приводится в [5].

Таким образом, можно заметить, что, в некотором смысле, мера μ_U является непараметрической оценкой для неизвестной меры P.

Обозначим через A^t и A^* множество всех слов длины t на A и множество всех конечных слов на алфавите A соответственно ($\mathrm{A}^* = \bigcup_{i=1}^{\infty} \mathrm{A}^i$). Через $M_\infty(\mathrm{A})$ мы обозначим множество всех стационарных и эргодических источников, которые порождают символы из A и пусть $M_0(\mathrm{A}) \subset M_\infty(\mathrm{A})$ будет множеством всех случайных источников. Пусть $M_m(\mathrm{A}) \subset M_\infty(\mathrm{A})$ будет классом всех марковских источников порядка (или связности) не больше m, $m \geq 0$. Пусть $M^*(\mathrm{A}) = \bigcup_{i=0}^{\infty} M_m(\mathrm{A})$ будет множеством всех источников конечной связности.

Мера Кричевского для $M_m(\mathrm{A})$ определяется следующим образом:

$$K_m(x_1 \ldots x_t) = \begin{cases} \dfrac{1}{|\mathrm{A}|^t}, & \text{если } t \leq m, \\ \dfrac{1}{|\mathrm{A}|^m} \prod_{\upsilon \in \mathrm{A}^m} \dfrac{\prod_{\alpha \in \mathrm{A}} (\Gamma(\nu_x(\upsilon a) + 1/2)/\Gamma(1/2))}{(\Gamma(\bar{\nu}_x(\upsilon) + |\mathrm{A}|/2)/\Gamma(|\mathrm{A}|/2))}, & \text{если } t > m, \end{cases}$$

где $\nu_x(\upsilon)$ – число последовательностей υ, встречающихся в x, $\bar{\nu}_x(\upsilon) = \sum_{\alpha \in \mathrm{A}} \nu_x(\upsilon\alpha)$, $x = x_1 \ldots x_t$, а $\Gamma()$ – это гамма-функция.

Пусть $\omega = \omega_1, \omega_2, \ldots$ – это распределение вероятностей, каждый член которого находится по формуле:

$$\omega_i = 1/\log(i+1) - 1/\log(i+2), \tag{1}$$

где i – целое из множества $1, 2, \ldots$.

Введём определение меры R, которая является оценкой вероятностей для класса всех стационарных и эргодических источников на конечном алфавите:

$$R(x_1 \ldots x_t) = \sum_{i=0}^{\infty} \omega_{i+1} K_i(x_1 \ldots x_t).$$

(В дальнейшем в данной работе в качестве ω будет использоваться распределение (1), хотя результаты будут справедливы и для любого распределения ненулевых вероятностей).

Важно заметить, что мера R основывается на универсальном коде из [1] и также может применяться при решении задач прогнозирования. Если говорить более точно, то теорема 1 является истинной, если заменить μ_U на R.

Рассмотрим прогнозирование для различных типов последовательностей, порождаемых источником.

1) *Источник порождает символы из конечного алфавита*

Как упоминалось выше, мера R является оценкой вероятностей для класса всех стационарных и эргодических источников на конечном алфавите. Поэтому в данном случае метод прогнозирования выглядит достаточно просто. Пусть $x_1 \ldots x_t$ – некоторая известная последовательность, порождённая источником. Тогда для любого элемента $a \in \mathrm{A}$ можно вычислить вероятность того, что он будет следующим символом на выходе источника (т.е. условную вероятность этого символа), воспользовавшись формулой

$$R(a \mid x_1 \ldots x_t) = \frac{R(x_1 \ldots x_t a)}{R(x_1 \ldots x_t)}.$$

2*) Источник порождает значения из непрерывного интервала*

Пусть (Ω, F, P) – вероятностное пространство и $X_1, X_2, \ldots$ – стохастический процесс, каждое X_t которого принимает значения из стандартного борелевого пространства. Предположим, что совместное распределение P_n для $(X_1, X_2, \ldots, X_n)$ имеет функцию плотности вероятности $p(x_1 x_2 \ldots x_n)$ по отношению к сигма-конечной мере M_n. В качестве M_n может выступать мера Лебега, счетная мера и т.д. Пусть Π_n, $n \geq 1$ будет возрастающей последовательностью конечных разбиений Ω, которые асимптотически генерирует борелевское поле на *F*, и пусть $x^{[k]}$ обозначает элемент Π_k, который содержит точку *x*. Для целых *s* и *n* следующим образом определим приближение величины плотности:

$$p^s(x_1, \ldots, x_n) = P(x_1^{[s]}, \ldots, x_n^{[s]}) \Big/ M_n(x_1^{[s]}, \ldots, x_n^{[s]})$$

Пусть *U* – это универсальный код, который определён для любого конечного алфавита. Определим соответствующую плотность как

$$r_U(x_1 \ldots x_t) = \sum_{i=1}^{\infty} \omega_i R(x_1^{[i]} \ldots x_t^{[i]}) \Big/ M_t(x_1^{[i]} \ldots x_t^{[i]}). \quad (2)$$

(Здесь предполагается, что $U(x_1^{[i]} \ldots x_t^{[i]})$ определён для алфавита, который содержит буквы $|\Pi_i|$).

В [4] показывается, что плотность $r_U(x_1 \ldots x_t)$ может выступать в качестве оценки неизвестной плотности $p(x_1, \ldots, x_t)$, а условная плотность

$$r_U(a \mid x_1 \ldots x_m) = \frac{r_U(x_1 \ldots x_m a)}{r_U(x_1 \ldots x_m)} \quad (3)$$

является подходящей оценкой для условной вероятности $p(a \mid x_1 \ldots x_m)$.

При применении данного метода на практике может возникнуть вопрос о том, последовательность какой длины нужно взять в качестве входных данных. Казалось бы, что лучший результат будет достигаться при использовании максимально возможной длины известных членов временно́го ряда. Однако в действительности это не всегда так. Например, в силу того, что данные из начала временно́го ряда могут соответствовать нетипичному для этого процесса поведению, учёт данных из этого периода негативно скажется на точности полученного результата. Итак, если ответ на данный вопрос неочевиден, то в качестве вероятностной оценки можно использовать величину $\tilde{r}_U$, определение которой приводится ниже.

Пусть $x_1 \ldots x_t$ – некоторая последовательность источника. Рассмотрим множество различных натуральных чисел N_i и множество выборок $x_{t-N_i+1} \ldots x_t$, $1 \leq N \leq t, 1 \leq i \leq t$. Иными словами, любая выборка данного множества, по сути, состоит из N_i последних членов исходной последовательности. Каждую отдельно взятую выборку данного множества назовём «окном», а соответствующее N_i – его размером. Пусть *i=k*, тогда

$$\tilde{r}_U = \sum_{i=1}^{k} \tilde{\omega}_i r_U(x_{t-N_i+1} \ldots x_t),$$

где $\tilde{\omega} = \tilde{\omega}_1, \tilde{\omega}_2, \ldots, \tilde{\omega}_k$ – распределение ненулевых вероятностей, $\sum_{i=1}^{k} \tilde{\omega}_i = 1$. Аналогично величине r_U, $\tilde{r}_U$ может использоваться для оценки условной вероятности $p(a \mid x_1 \ldots x_m)$.

Как видно из определения, при расчёте данной величины учитываются последовательности («окна») различной длины, и соответствующие r_U «склеиваются» с некоторыми вероятностями. Таким образом, вычисления на основе $\tilde{r}_U$ должны автоматически обеспечивать оптимальные, с точки зрения простоты использования метода и точности прогноза, результаты.

2.2. Реализация алгоритма

Рассмотрим теперь некоторые аспекты практической реализации исследуемого метода. Итак, пусть есть некоторый источник, который порождает значения из непрерывного отрезка $A;B$, и известна последовательность членов временно́го ряда $x_1...x_t$ длины t. Требуется спрогнозировать x_{t+1}. Для простоты рассмотрим вычисления на основе r_U . Действия при использовании $\tilde{r}_U$ аналогичны.

Шаг 1. Вычислим на основе (2) $r_U(x_1...x_t)$. Опишем этот этап немного подробнее. Разобьём интервал [A; B] на 2 равные части и преобразуем $x_1...x_t$ в последовательность символов, каждый из которых равен номеру участка разбиения, который содержит соответствующую точку x_i. (Т.е. если номера участков были 0 и 1, то в результате получим некоторую последовательность, состоящую из этих символов). Далее, вычислим первое слагаемое суммы из формулы (2). В качестве M_n возьмём произведение длин всех участков разбиения, содержащих x_i, а меру R будем вычислять для последовательности, полученной на этом шаге на основе $x_1...x_t$. После этого снова разделим каждый из имеющихся участков (их будет два) на 2 равные части (теперь их станет четыре) и для этого нового разбиения выполним аналогичные действия с целью нахождения второго слагаемого из правой части (2). Будем продолжать данный процесс до тех пор, пока не получим разбиение, для которого все различные значения из временны́х рядов (в том числе и используемые на следующем шаге) не будут принадлежать различным участкам. Сложив все имеющиеся слагаемые, получим величину $r_U(x_1...x_t)$ из (2).

Заметим, что на данном шаге можно использовать любой другой алгоритм получения возрастающей последовательности конечных разбиений интервала $A;B$.

Шаг 2. Рассмотрим множество a_i , которое состоит из величин: $A, A+h, A+2h, ..., B$, где h – некоторый достаточно маленький шаг. В данной работе использовался шаг h=0.01. Подставляя по очереди каждый элемент этого множества в конец последовательности $x_1...x_t$, вычислим соответствующие величины $r_U(x_1...x_t a)$, где a – элемент из множества a_i . При вычислениях будем использовать те же способ и разбиение, что и на шаге 1.

Шаг 3. Теперь по формуле (3) для всех величин из множества a_i вычислим оценки соответствующих условных вероятностей и на их основе определим прогнозное значение. В данной работе прогнозным значением считалась та величина, чья оценка условной вероятности была наибольшей среди всех остальных, но можно использовать и любой другой подходящий способ.

3. Экспериментальные результаты

В данном разделе приводятся результаты экспериментальной оценки метода, описанного в предыдущей секции. В качестве объектов исследования были выбраны временны́е ряды, состоящие из следующих показателей: среднемесячное и сглаженное среднемесячное число

солнечных пятен, абсолютные ежедневное и ежемесячное значение солнечного излучения. Данные, использованные в экспериментах, можно найти на сайте National Geophysical Data Center (NGDC) в разделе «Space Weather & Solar Events» [8]. Все исследуемые процессы принимали значения из некоторого непрерывного интервала. Проводимые в рамках данного раздела вычисления можно разделить на две независимые части. В первой рассматривалась задача прогнозирования значений процессов на один шаг вперед, а во второй – сравнение между собой точности краткосрочных прогнозов для сглаженного среднемесячного числа солнечных пятен: опубликованного на сайте NGDC и полученного помощью метода на основе кода *R*. Более подробное описание проводимых в каждой части вычислений, а также результаты этих экспериментов приводятся ниже, в пунктах 3.1 и 3.2 соответственно.

3.1. Прогнозирование значений временны́х рядов

Суть каждого проводимого на данном этапе эксперимента заключалась в следующем. Имея в распоряжении только данные об N последовательных значениях временно́го ряда, требовалось спрогнозировать значение его (N+1)-го члена. Для того чтобы иметь возможность сравнить результаты, полученные с помощью рассматриваемого метода, с соответствующими реально зафиксированными величинами процесса, как период основания, так и период упреждения прогноза были взяты из на самом деле уже прошедшего временно́го интервала. В ходе вычислений по каждому выбранному процессу для анализа зависимости точности прогноза от длины последовательности известных значений использовались временны́е ряды различной продолжительности. Кроме того, с целью изучения возможности оптимизации вычислений, дополнительно проводились расчёты с использованием «окна». (Подробное описание данного метода см. в предыдущем разделе). Следует отметить, что в качестве размеров «окон» для каждого процесса были взяты все длины, используемые для данного временного ряда при вычислениях без «окна», а значения вероятностей при «склеивании» были одинаковыми. Для каждой длины временно́го ряда определённого процесса, а также при соответствующих вычислениях с использованием «окна», проводилось 25 экспериментов над независимыми наборами данных. Результаты экспериментальных вычислений содержатся в табл. 1, приведённой ниже.

Первая колонка таблицы содержит наименование исследуемого временно́го ряда, вторая – рассматриваемый диапазон принимаемых им значений. Остальные колонки содержат значения средней абсолютной погрешности, полученные либо при вычислениях с применением «окна», либо при использовании соответствующей длины временно́го ряда. Надпись «н/д» («нет данных») в ячейке означает, что вычисление не проводилось по причине отсутствия требующихся данных. Так, например, в строке № 3 таблицы содержится информация о том, что для временно́го ряда с данными об абсолютном ежедневном солнечном излучении средняя абсолютная погрешность при прогнозировании на основе 4000 известных значений равна 1.45. Диапазон значений, которые принимают члены этого временно́го ряда, равен [50; 300]. А на основании строки № 1 таблицы можно сказать, что для временно́го ряда с данными о среднемесячном числе солнечных пятен средняя абсолютная погрешность при подсчёте с использованием «окна», размеры которого составляли 500, 700, 1000, 1200, 2000 и 3000, составила 23.97.

Таблица 1. Результаты экспериментальных вычислений первой части

Временной ряд	Диапазон значений процесса	Исп. «окно»	Длина						
			500	700	1000	1200	2000	3000	4000
Среднемесячное число солнечных пятен	[0; 256]	23.97	6.54	2.56	9.58	15.85	21.7	19.63	н/д
Сглаженное среднемесячное число солнечных пятен	[0; 210]	2.34	1.5	1.1	1.99	0.77	3.36	2.56	н/д
Абсолютное ежедневное солнечное излучение	[50; 300]	1.78	1.17	1.17	2.71	5.52	8.35	1.72	1.45
Абсолютное ежемесячное солнечное излучение	[580; 2540]	45.29	211.29	45.88	н/д	н/д	н/д	н/д	н/д

3.2 Сравнение краткосрочных прогнозов

Согласно информации, предоставленной на сайте NGDC, в настоящее время существует специализированная программа прогнозирования (далее – программа NGDC), которая формирует «предварительный взгляд» на текущий солнечный цикл, используя улучшенный метод МакНиша – Линкольна (McNish-Lincoln). В соответствии с предоставленной на сервере информацией, программа NGDC формирует прогноз для сглаженного среднемесячного числа солнечных пятен, и также использует для своих вычислений известные значения членов временно́го ряда для данного процесса.

На этом этапе оценки экспериментальных результатов будет произведено сравнение точности краткосрочных прогнозов: полученного с помощью программы NGDC и составленного на основе исследуемого метода. Из тех же, что и в пункте 3.1, соображений, здесь рассматривался прогноз на уже прошедший период.

В качестве прогноза, сформированного программой NGDC, был взят файл sunspot.predict от 06.08.2008, в котором для каждого месяца текущего солнечного цикла указано соответствующее прогнозное значение и его доверительный интервал. Здесь следует уточнить, что в качестве величины для сравнения будет использоваться непосредственно предоставленное прогнозное значение без учёта доверительного интервала. При этом известно, что на момент создания этого прогноза о рассматриваемом процессе имелись в наличии и могли использоваться данные только по февраль 2008 года включительно. Следовательно, на момент проведения вычислений этой работы в июне 2010 года имелось 21 уже известное значение исследуемого временно́го ряда (за каждый месяц с марта 2008 по ноябрь 2009). Вычисления исследуемым методом без использования «окна» проводились по следующей схеме. Рассматривался такой временной ряд некоторой фиксированной в рамках эксперимента длины, что его последний член содержал данные за февраль 2008-го года. На основании содержимого данного ряда строился прогноз на один шаг вперед, т.е. март 2008-го года. После этого, на следующей итерации, производилось смещение границ временно́го ряда с извест-

ными значениями на один элемент вправо, при этом последний элемент (соответствующий реальному значению за март 2008) заменялся на полученное на предыдущем шаге прогнозное значение. На основании такого преобразованного ряда снова строился прогноз на один шаг вперед – апрель 2008-го года. И так далее. В общей сложности для каждой рассматриваемой длины временно́го ряда было проведено 21 вычисление. Последнее прогнозное значение было сформировано для ноября 2009 года. Вычисления с использованием «окна» проводились подобным образом, с той лишь разницей, что здесь на каждой итерации рассматривался не один, а несколько временны́х рядов различной длины. Аналогично предыдущему пункту в качестве размеров «окон» взяты все длины, используемые при подсчётах без «окна», а значения вероятностей, используемые при «склеивании», были одинаковыми. Результаты экспериментальных вычислений для исследуемого метода приводится в табл. 2. Первая строчка содержит данные о величине средней абсолютной погрешности при использовании исследуемого метода и соответствующих входных данных. Вторая – среднюю абсолютную погрешность, вычисленную на основе прогноза программы NGDC. Из таблицы ясно, напри́мер, что при прогнозировании исследуемым методом на основании имеющегося временно́го ряда длины 1000 средняя абсолютная погрешность за 21 эксперимент составила 1.87. Для программы NGDC данная величина вычислялась один раз на основе единственного прогнозного файла.

Таблица 2. Результат прогнозирования временно́го ряда
для сглаженных среднемесячных значений солнечных пятен

R	Длина			Исп. «окно»
	500	1000	2000	
	1.75	1.87	2.23	1.42
NGDC	16.39			

Для наглядности полученные в данной части экспериментальной оценки результаты представлены на рис. 1 ниже. На данном графике по оси N отложено сглаженное среднемесячное число солнечных пятен, а по оси t – условные единицы времени, каждая из которых соответствует некоторому месяцу солнечного цикла. Область графика, где $t \le 0$, является периодом основания прогноза, а область, где $1 \le t \le 21$ – периодом упреждения прогноза. Сплошная жирная линия отображает реальные значения временно́го ряда, пунктирные линии из коротких и длинных штрихов – прогнозные значения, полученные программой NGDC и вычисленные на основе кода R с использованием «окна» соответственно.

При использовании программы NGDC сопоставимая средняя абсолютная погрешность наблюдалась лишь на первых четырёх прогнозных значениях и была равной 1.95. При проведении 10 экспериментов она выросла до 5.51, а по завершении всех 21 вычислений данной серии экспериментов указанная величина составила 16.39, при этом наблюдалась ярко выраженная тенденция к росту ошибки прогноза с увеличением его порядкового номера.

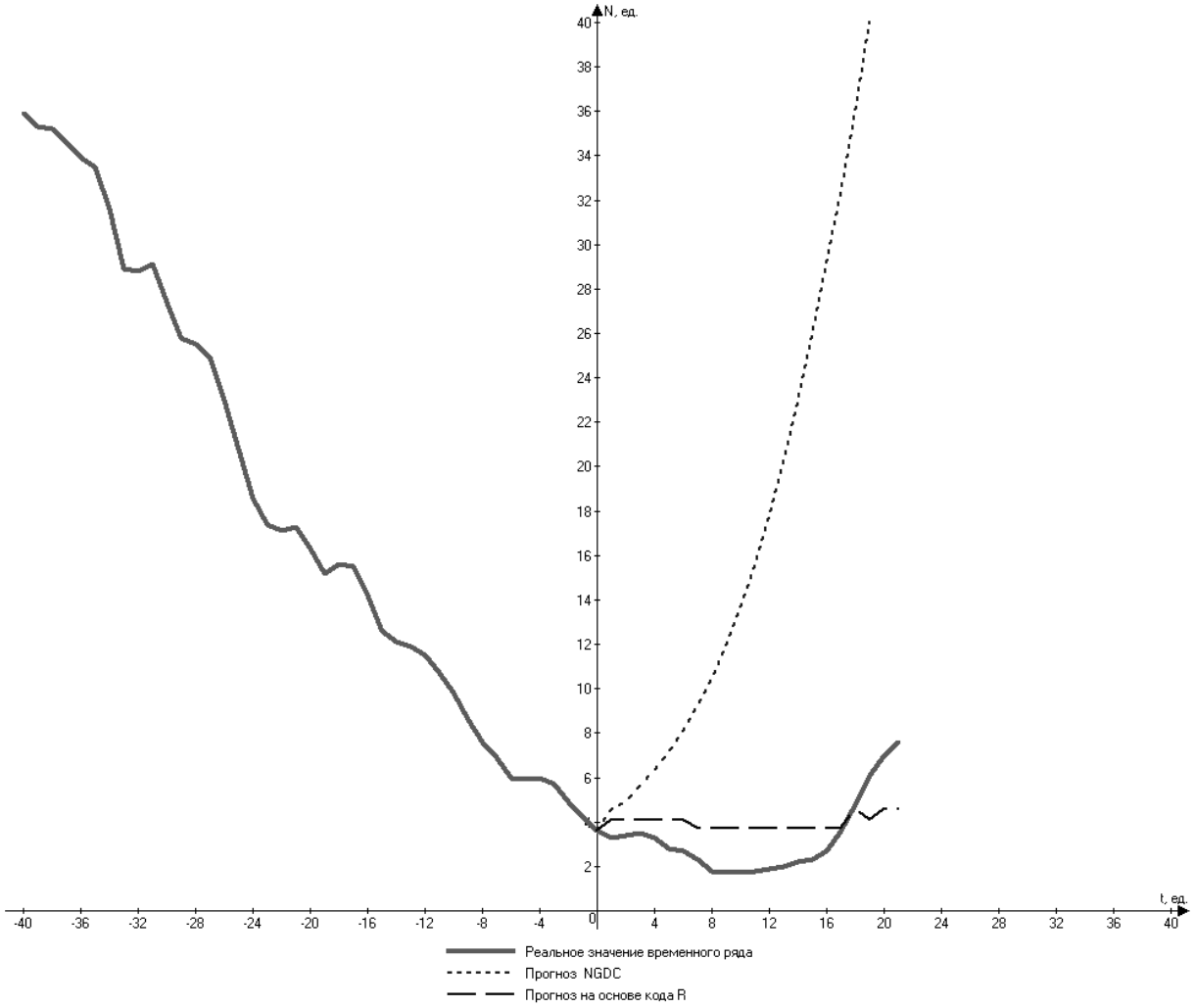

Рис. 1. Результаты сравнения краткосрочных прогнозов

Таким образом, на основании графика можно сказать, что в данном случае при построении краткосрочного прогноза исследуемый метод на основе универсального кода *R* оказался более эффективным. Данные на сервере NGDC представлены так, что провести другие эксперименты по сравнению точности невозможно.

Безусловно, следует отметить, что сравнение указанных методов на основе одного файла прогнозов может быть недостаточным для окончательных выводов, однако, в любом случае размер абсолютной и относительной (мы видим, что она составляет около 1 %) погрешностей, полученных в данном испытании при применении исследуемого метода, подтверждают его право на существование.

4. Заключение

В данной статье были рассмотрены реализация и экспериментальная оценка метода прогнозирования, основанного на универсальном коде *R*. Анализ итогов работы показал достаточно высокую точность полученных результатов. Более того, при сравнении краткосрочных прогнозов, составленных с помощью программы NGDC и указанного метода, было выявлено значительное преимущество последнего. Таким образом, можно сделать вывод о том, что универсальные коды являются эффективным инструментом при построения методов прогнозирования для решения практических задач.

Литература

1. B. Ya. Ryabko, Twice-universal coding // Problems of Information Transmission. 1984. V.20, № 3. P. 173–177.
2. B. Ya. Ryabko, Prediction of random sequences and universal coding // Problems of Information Transmission. 1988. V. 24, №. 2. P. 87–96.
3. A. Gruzin, B. Ryabko. Practical Application of Universal Codes to Time Series Forecasting // 2009 XII International Symposium on Problems of Redundancy in Information and Control Systems, Proceedings. P. 10 - 15.
4. B. Ryabko, V. Monarev. Experimental investigation of forecasting methods based on data compression algorithms // Problems of Information Transmission. 2005. V. 41, № 1. P. 65-69.
5. B. Ryabko. Compression-based methods for nonparametric on-line prediction, regression, classification and density estimation of time series // Festschrift in Honor of Jorma Rissanen on the Occasion of his 75th Birthday. Tampere, 2008. P. 271 – 288.
6. Солнечная активность // Википедия: свободная электронная энциклопедия: на русском языке (Последнее изменение: 10:46 21.06.2010) [Электронный ресурс] URL: http://ru.wikipedia.org/wiki/Солнечная_активность (дата обращения: 25.06.2010)
7. Котов Ю. Д. Солнечный спутниковый проект «Коронас-Фотон» // Земля и Вселенная. 2009. № 3. С. 3–19.
8. Space Weather & Solar Events [Электронный ресурс] // National Geophysical Data Center: [сайт] URL: http://www.ngdc.noaa.gov/stp/spaceweather.html (дата обращения: 04.06.2010)

Статья поступила в редакцию 8.11.2010

Приставка Павел Анатольевич
Аспирант кафедры прикладной математики и кибернетики СибГУТИ, e-mail: ppa@ngs.ru

Experimental investigation of forecasting method based on universal codes

P. A. Pristavka

In the article a forecasting method based on universal codes is proposed and experimentally investigated. It is shown on the examples of sunspot number and solar flux prediction that the method possesses high forecasting precision.

Keywords: forecasting, universal codes, time series, solar activity.