

УДК 621.391.15

© 2016 г. А.С. Лысяк, Б.Я. Рябко

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ МЕТОДОВ СЖАТИЯ ДАННЫХ

Предложены эффективные (“быстрые” и не требующие большой памяти) алгоритмы для базирующихся на универсальном кодировании методов прогнозирования временных рядов, принимающих действительные значения. Ранее для таких методов было доказано только, что погрешность прогноза асимптотически минимальна, а вопросы сложности реализации вообще не рассматривались. Приведенные в статье экспериментальные результаты свидетельствуют о высокой точности предлагаемых методов.

§ 1. Введение

Прогнозирование временных рядов представляет большой практический интерес, так как позволяет решать задачи предсказания различных явлений в науке, технике и экономике. По-видимому, поэтому данная задача привлекает широкое внимание исследователей и в настоящее время существует много методов прогнозирования, среди которых наиболее известны прогнозирование на основе билинейной модели [1], авторегрессионный анализ различных типов [2–5], прогнозирование на основе методов Монте-Карло [6], методы на основе построения экспертных оценок (так называемые рекурсивные стратегии, описание которых можно найти в [7, 8]). Однако несмотря на существование многочисленных методов и подходов, задача повышения точности прогноза временных рядов не теряет своей актуальности.

В данной статье продолжается развитие методов прогнозирования, базирующихся на использовании универсальных кодов (или “методов сжатия данных”). Впервые возможность использования универсальных кодов для прогнозирования была описана в 1988 г. в работе [9], где рассматривались случайные процессы, порождающие значения из некоторого конечного множества (алфавита). В работе [10] было показано, как можно применять универсальные коды для прогнозирования временных рядов, принимающих значения из некоторого непрерывного интервала чисел, однако теоретического анализа предложенного метода в то время дано не было. Позже в [11] были предложены асимптотически наиболее точные методы прогноза временных рядов, принимающих значения из непрерывных интервалов, однако они не были практически реализованы и каких-либо сведений о (не асимптотической) точности прогноза и сложности не было.

В данной статье описываются алгоритмы, реализующие асимптотически оптимальные методы из [11] и экспериментально исследуется их точность на реальных временных рядах. Сравнение предлагаемого метода с известными показало, что его точность прогноза довольно высока.

§ 2. Универсальные коды и универсальные меры как основа для методов прогнозирования

2.1. Задача прогнозирования. Опишем сначала задачу прогнозирования. Пусть имеется некоторый стационарный и эргодический источник, порождающий последовательность элементов $x_1 x_2 \dots$ из некоторого множества A , которое назовем алфавитом. Алфавит может быть как конечным или счетным, так и представлять собой некоторый непрерывный отрезок на действительной прямой. Задача прогнозирования состоит в оценивании распределения вероятностей для случайной величины x_{t+1} по значениям x_1, x_2, \dots, x_t . В случае дискретного алфавита мы будем оценивать условные вероятности $P(x_{t+1} = a \in A | x_1 x_2 \dots x_t)$. Для непрерывного алфавита будем рассматривать только случай, когда существуют плотности условных вероятностей $p(x_{t+1} | x_1 x_2 \dots x_t)$ и задача прогнозирования сводится к построению оценок для них (все другие представляющие интерес характеристики, используемые при прогнозировании – среднее значение, дисперсия процесса и т.п. – легко вычисляются по плотности). В частности, мы (вслед за многими авторами) будем оценивать среднее значение процесса x_{t+1} (при заданных $x_1 \dots x_t$), вычисляя его по оценке плотности.

Остановимся кратко на связи между универсальными кодами и так называемыми универсальными мерами, с одной стороны, с задачей прогнозирования стационарных процессов. Пусть Ω – некоторое множество стационарных эргодических процессов, порождающих элементы из алфавита A . Предиктором будем называть функцию γ , определенную на всех словах $x_1 \dots x_{t+1}$, $t \geq 0$, которую для удобства будем обозначать аналогично условной вероятности через $\gamma(x_{t+1} | x_1 \dots x_t)$, и потребуем выполнения естественных условий $\sum_{a \in A} \gamma(a | x_1 \dots x_t) = 1$ и $\gamma(x_{t+1} | x_1 \dots x_t) \geq 0$ для всех $x_1 \dots x_{t+1} \in A^{t+1}$, $t \geq 0$.

Каждый предиктор естественным образом задает на множестве слов длины t меру $\gamma(x_1 \dots x_t)$:

$$\gamma(x_1 \dots x_t) = \gamma(x_1) \gamma(x_2 | x_1) \dots \gamma(x_t | x_1 \dots x_{t-1}). \quad (1)$$

При разных t эти меры естественно согласованы и задают распределение вероятностей на пространстве бесконечных последовательностей. Заметим, что справедливо и обратное: каждое распределение вероятностей π на бесконечных последовательностях в алфавите A задает предиктор

$$\pi(x_t | x_1 \dots x_{t-1}) = \pi(x_1 \dots x_t) / \pi(x_1 \dots x_{t-1}). \quad (2)$$

В качестве примера рассмотрим метод предсказания Лапласа. Он предложил оценивать условные вероятности $p(x_{t+1} = a \in A | x_1 x_2 \dots, x_t)$ по известным значениям $x_1 \dots x_t$ с помощью следующего предиктора L :

$$L(a | x_1 \dots x_t) = (\nu_{x_1 \dots x_t}(a) + 1) / (t + |A|),$$

где $\nu_{x_1 \dots x_t}(a)$ – количество вхождений буквы a в слово $x_1 \dots x_t$. Так, например, $L(0 | 01010) = 4/7$ при $A = \{0, 1\}$.

Естественным вопросом является оценка качества, или точности, предиктора. Одна из широко используемых в теории информации и математической статистике мер точности – расхождение по Кульбаку – Лейблеру (KL). В нашем случае расхождение между распределением вероятностей $P(x_{t+1} = a | x_1 \dots x_t)$ и предиктором $\gamma(a | x_1 \dots x_t)$ при заданном $x_1 \dots x_t$ дается равенством

$$\begin{aligned} \text{KL}(P, \gamma)_{x_1 \dots x_t} &= \\ &= \sum_{a \in A} P(x_{t+1} = a | x_1 \dots x_t) \log(\gamma(a | x_1 \dots x_t) / P(x_{t+1} = a | x_1 \dots x_t)) \end{aligned} \quad (3)$$

(здесь и далее $\log x \equiv \log_2 x$), а среднее значение (при заданной длине ряда t) – равенством

$$\text{KL}(P, \gamma)_t = \sum_{x_1 \dots x_t \in A^t} P(x_1 \dots x_t) \text{KL}(P, \gamma)_{x_1 \dots x_t}. \quad (4)$$

Заметим, что $\text{KL}(P, \gamma)_{x_1 \dots x_t}$ неотрицательна, и равна нулю только в случае, когда $\gamma(a | x_1 \dots x_t)$ и $P(x_{t+1} = a | x_1 \dots x_t)$ равны при всех a, x_1, \dots, x_t (см. [12]). В [9] показано, что для любого источника P , порождающего независимые и одинаково распределенные символы из алфавита A погрешность предиктора Лапласа удовлетворяет неравенству

$$\text{KL}(P, L)_t \leq \log e(|A| - 1)/(t + 1)$$

(здесь $e = 2,718 \dots$ – число Эйлера). Мы видим, что погрешность предиктора Лапласа стремится к нулю для любого источника, порождающего независимые и одинаково распределенные символы. К сожалению, не существует предиктора, обладающего таким свойством для любого стационарного эргодического источника (см. доказательство в [9]). Однако для таких источников существуют предикторы, для которых выполняется более слабое свойство: среднее по Чезаро погрешностей (4) стремится к нулю. Точнее, существует предиктор γ , такой что для любого стационарного эргодического источника ω , порождающего буквы из некоторого конечного алфавита A , справедливо равенство

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t \text{KL}(\omega, \gamma)_s = 0. \quad (5)$$

Предиктор γ называется универсальным для множества источников Ω , если равенство (5) выполняется для любого $\omega \in \Omega$. Величину $\frac{1}{t} \sum_{s=1}^t \text{KL}(\omega, \gamma)_s$ обозначим через $\overline{\text{KL}}(\omega, \gamma)_t$ и назовем средней погрешностью предиктора γ на источнике ω . Из этого определения и (3), (4) легко получить равенство

$$\overline{\text{KL}}(\omega, \gamma)_t = \sum_{x_1 \dots x_t \in A^t} \omega(x_1 \dots x_t) \log(\gamma(x_1 \dots x_t)/\omega(x_1 \dots x_t)), \quad (6)$$

где $\gamma(x_1 \dots x_t)$ определена в (1). Отсюда и из (5) становится понятен смысл универсальности предиктора: для любого источника $\omega \in \Omega$ величина предиктора $\gamma(x_1 \dots x_t)$ приближается к вероятности $\omega(x_1 \dots x_t)$.

2.2. Универсальные коды и меры. Понятие универсальной меры связывает задачи прогнозирования и универсального кодирования. Определяется она следующим образом: пусть дано некоторое множество стационарных и эргодических источников Ω . Мера μ называется универсальной, если для любого источника $P \in \Omega$ верно равенство

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u \in A^t} P(u) \log(P(u)/\mu(u)) = 0.$$

Данное равенство показывает, что в определенном смысле универсальная мера μ является непараметрической оценкой для неизвестного распределения вероятностей P . Из сравнения последнего равенства и (1)–(6) видно, что универсальная мера и предиктор тесно связаны и фактически совпадают.

Универсальные меры и предикторы связаны с универсальными кодами. Дадим описание этой взаимосвязи, так как она позволяет использовать для прогнозирова-

ния временных рядов так называемые архиваторы, т.е. программы, предназначенные для реального сжатия текстов. Здесь важно отметить, что современные архиваторы позволяют использовать для “сжатия” отклонения в частоте встречаемости различных букв и подслов, скрытые периодичности и целый ряд других закономерностей. Несомненно, это их свойство является ценным с практической точки зрения, и одна из задач этой работы – показать, как реальный архиватор может быть использован для построения предиктора, причем и в случае, когда значения случайного процесса – действительные числа.

Подробное определение неискажающего (“обратимого”) кода можно найти, например, в [12], а здесь кратко отметим, что код – это отображение слов длины t из алфавита A (т.е. A^t , $t \geq 1$) в множество различных слов в алфавите $\{0, 1\}$. Код U называется универсальным, если для любого стационарного эргодического источника P

$$\lim_{t \rightarrow \infty} E_P(|U(x_1 \dots x_t)|)/t = H(P),$$

где $E_P(f)$ – среднее значение f по мере P , а $H(P)$ – энтропия P по Шеннону, т.е.

$$H(P) = \lim_{t \rightarrow \infty} -t^{-1} \sum_{u \in A^t} P(u) \log P(u).$$

Заметим, что энтропия – асимптотически достижимая нижняя граница средней длины неискажающего кода, поэтому-то они и называются универсальными.

Следующее простое утверждение (см., например, [9]) говорит о том, что на базе любого универсального кода можно построить универсальную меру.

Предложение. Пусть U – универсальный код для некоторого множества источников Ω , порождающих буквы из алфавита A , и мера μ_U для каждого слова v в алфавите A задается равенством

$$\mu_U(v) = 2^{-|U(v)|} / \sum_{u \in A^{|v|}} 2^{-|U(u)|}. \quad (7)$$

Тогда μ_U – универсальная мера для Ω .

Для построения прогноза для реальных процессов необходимо выбрать конкретную универсальную меру (или сначала универсальный код, а потом по нему вычислять меру согласно (7)). Мы для прогнозирования будем использовать универсальную меру R (см. [13]). Выбор именно этой меры связан с тем, что она построена на основе универсального кода, избыточность которого асимптотически минимальна для классов бернуллиевских и марковских источников [13].

Для ее описания сначала приведем найденный в 1968 г. предиктор, для которого погрешность (6) асимптотически минимальна для множества всех бернуллиевских источников [14, 15]. Данный предиктор, позволяющий вычислить условные вероятности для следующего элемента ряда, определяется равенством

$$K_0(a | x_1 \dots x_t) = (\nu_{x_1 \dots x_t}(a) + 1/2)/(t + |A|/2), \quad (8)$$

где $\nu_{x_1 \dots x_t}(a)$ – число элементов a , встречающихся в слове $x_1 \dots x_t$. Интересно отметить, что для этого предиктора погрешность асимптотически в два раза ниже, чем аналогичная ошибка для предсказателя Лапласа.

На основе этого предиктора строится мера K_0 , которая, как впервые показано в [14], является универсальной для класса бернуллиевских источников:

$$K_0(x_1 \dots x_t) = \prod_{i=0}^{t-1} \frac{\nu_{x_1 \dots x_i}(x_{i+1}) + 1/2}{i + |A|/2}.$$

Например, $K_0(01010) = \frac{1}{1} \frac{1/2}{2} \frac{3/2}{3} \frac{3/2}{4} \frac{5/2}{5}$ при $A = \{0, 1\}$.

Для марковских источников аналогичная мера имеет следующий вид [15]:

$$K_m(x_1 \dots x_t) = \begin{cases} \frac{1}{|A|^t}, & t \leq m, \\ \frac{1}{|A|^m} \prod_{i=m}^{t-1} \frac{\nu_{x_1 \dots x_i}(x_{i+1-m} \dots x_{i+1}) + 1/2}{\nu_{x_1 \dots x_{i-1}}(x_{i+1-m} \dots x_i) + |A|/2}, & t > m, \end{cases} \quad (9)$$

где $\nu_x(\vartheta)$ – число последовательностей ϑ , встречающихся в x . Например, $K_1(01010) = \frac{1}{2} \frac{1/2}{1} \frac{1/2}{1} \frac{3/2}{2} \frac{3/2}{2}$ при $A = \{0, 1\}$. Данная мера является универсальной для множества марковских источников связности m (см. [15]).

Мера R , универсальная для множества всех стационарных и эргодических источников, определяется следующим образом:

$$R(x_1 \dots x_t) = \sum_{i=0}^{\infty} \omega_{i+1} K_i(x_1 \dots x_t), \quad (10)$$

где

$$\omega_i = 1/\log(i+1) - 1/\log(i+2). \quad (11)$$

2.3. Схема прогнозирования временных рядов, порождающих действительные значения. Прогнозирование временных рядов, порождающих значения из конечного алфавита, осуществляется по формуле (см. [9])

$$R(x_{t+1} = a | x_1 \dots x_t) = R(ax_1 \dots x_t)/R(x_1 \dots x_t).$$

В [11] показано, как этот метод можно перенести на случай рядов, порождающих действительные значения, и получить асимптотически оптимальный предиктор. Приведем необходимые сведения об этом методе.

Рассмотрим временной ряд, генерирующий последовательность x_t , каждый элемент которой принимает значения из интервала $[A, B]$. Пусть $\{\Pi_n\}$, $n \geq 1$, – возрастающая последовательность конечных разбиений интервала $[A, B]$, такая что максимальный размер подынтервала в разбиениях стремится к нулю (назовем этот процесс квантизацией). Определим также $x^{[k]}$ как элемент Π_k , содержащий точку x .

В дальнейшем будем рассматривать только процессы, для которых все многомерные плотности существуют. Обозначим через $p(x_1 x_2 \dots x_n)$ плотность вероятности процесса по мере Лебега L .

Определим теперь оценку плотности вероятностей r следующим образом:

$$r(x_1 \dots x_t) = \sum_{s=1}^{\infty} \omega_s \left(R(x_1^{[s]} \dots x_t^{[s]}) / L(x_1^{[s]} \dots x_t^{[s]}) \right). \quad (12)$$

Множители ω_s определяются формулой (11) и играют роль весовых коэффициентов для разбиений из Π_k . Как видно из (12), в процессе вычисления меры r происходит нормировка каждого слагаемого по мере Лебега L . Таким образом, мы соединяем между собой оценки плотности вероятностей для случая различных возрастающих разбиений, что избавляет нас от зависимости результатов прогноза от конкретного разбиения.

Как показано в [11], величина $r(x_1 \dots x_t)$ является оценкой неизвестной плотности вероятности $p(x_1 \dots x_t)$, а соответствующая условная плотность

$$r(a | x_1 \dots x_t) = r(x_1 \dots x_t a) / r(x_1 \dots x_t) \quad (13)$$

является оценкой плотности $p(a | x_1 \dots x_t)$, причем обе оценки в определенном смысле состоятельны [11].

§ 3. Описание алгоритма и оценка его трудоемкости

В [11] показано, что можно использовать для прогнозирования произвольные последовательности конечных разбиений, определяющих ступенчатые функции, оценивающие плотность. Экспериментальным путем мы выяснили, что разбиение на равные подынтервалы дает, как правило, наилучшую точность прогноза на реальных данных, поэтому в описываемых ниже экспериментах использовалось именно это разбиение. В качестве прогнозного значения x_{t+1} мы брали среднее значение, вычисляемое по оценке условной плотности (13).

Оценим трудоемкость описанного алгоритма прогнозного значения, понимая под трудоемкостью (или сложностью) число операций. Число операций при вычислении оценки плотности по n подынтервалам определяется сложностью вычисления меры R (см. (10)) в n -буквенном алфавите, которая, в свою очередь, зависит от сложности вычисления K_m (см. (9), (10)). Легко видеть, что сложность вычисления выражения (9) равна $O(tn^{t+1})$. Исходя из этого, получаем, что сложность вычисления прогнозного значения равна $O(t^3n^{t+2})$. Существенное уменьшение сложности вычислений было получено за счет того, что при большом числе подынтервалов n (например, $n > t$) многие частоты встречаемости подынтервалов (ν в (9)) совпадают, что позволяет использовать описанный в [16] метод группировки букв алфавита для уменьшения трудоемкости. В данном случае уменьшение сложности невозможно описать аналитически, так как эта величина, вообще говоря, зависит от значений, принимаемых рассматриваемым временным рядом, однако эксперименты показали, что время вычислений сокращалось в 3–5 раз при длине ряда от нескольких десятков до двух тысяч.

Итак, с одной стороны, количество подынтервалов разбиения n задает погрешность прогноза (которая, очевидно, не может быть меньше половины длины подынтервала). С другой стороны, сложность алгоритма существенно зависит от величины n . Экспериментально мы определили, что $n = \lfloor \log_2 t \rfloor + 5$ является хорошим компромиссом, так как большие значения n почти не влияют на точность прогноза, но существенно увеличивают время вычислений. Поэтому в наших экспериментах вместо (12) использовалась следующая формула:

$$r(x_1 \dots x_t) = \sum_{s=1}^{\lfloor \log_2 t \rfloor + 4} \omega_s R(x_1^{[s]} \dots x_t^{[s]}) / L(x_1^{[s]} \dots x_t^{[s]}) + \frac{1}{\lfloor \log_2 t \rfloor + 6} R(x_1^{\lfloor \log_2 t \rfloor + 5} \dots x_t^{\lfloor \log_2 t \rfloor + 5}) / L(x_1^{\lfloor \log_2 t \rfloor + 5} \dots x_t^{\lfloor \log_2 t \rfloor + 5}). \quad (14)$$

§ 4. Экспериментальные результаты прогнозирования

4.1. Описание методики экспериментальных исследований. Для конкретного ряда $x_1 \dots x_t$ по участку $x_1 \dots x_{t-10}$ вычислялось прогнозное значение для момента времени $t - 9$ (обозначим его через y_{t-9}), затем по $x_1 \dots x_{t-9}$ вычислялось прогнозное значение y_{t-8} для $t - 8$ и т.д.; по $x_1 \dots x_{t-1}$ вычислялось прогнозное значение y_t .

После этого вычислялась средняя ошибка прогноза $\left(\sum_{i=1}^{10} |x_i - y_i| \right) / 10$, которая и использовалась при сравнении методов.

В прогнозировании известны подходы, когда проводится предварительное преобразование данных (обычно фильтрация, сглаживание и т.п.), затем полученный ряд используется непосредственно для прогнозирования, после чего к прогнозным

Таблица 1

Прогнозирование цен на топливо в США. Метод R в режиме на 1 и на 20 шагов вперед

Величина разбиения	R-метод 1 шаг	R-метод 20 шагов
5	0,07260	0,44139
10	0,04658	0,57510
20	0,05202	0,29762
50	0,03915	0,12638



График цен на энергоносители (интервал – 1 неделя)

значениям применяется обратное преобразование. В рассматриваемых ниже примерах в некоторых случаях использовалось преобразование исходного ряда $x_1 \dots x_t$ к ряду $x_1^* \dots x_{t-1}^*$ по формуле $x_i^* = x_{i+1} - x_i$ (в прогнозировании такой подход обычно используется для удаления близких к линейным трендов).

В качестве примера рассмотрим задачу прогнозирования цен на топливо в США (см. [17]). В табл. 1 приведены результаты прогнозирования данного ряда с интервалом одна неделя в период с 01.01.2002 по 01.10.2013. Длина ряда равна 615 элементам, глубина вычислений бралась равной 5, а размер интервала ряда равен 0,832. Для того чтобы понять масштаб изменяемой величины, приводится значение Δ , равное максимальной разнице между двумя соседними элементами прогнозируемого ряда. График ряда приведен на рисунке. На нижней оси графика располагаются номера элементов временного ряда в порядке возрастания.

Прогнозирование для этого ряда проводилось для последних десяти точек (по всем предыдущим) и подсчитывалась погрешность прогноза. Среднее значение погрешности (по десяти точкам) равно 0,03915, тогда как разница между соседними значениями ряда достигает 1. Таким образом, погрешность прогноза меньше 4% от этой величины.

Для того чтобы сравнить точность предлагаемого метода с ранее известными, использовались данные по экономическим временным рядам США, по которым известны результаты прогнозирования методами международного института прогнозистов (International institute of forecasters (ИИФ)). Было взято четыре временных ряда из [17]: industry (индекс промышленного производства США), finance (1) и finance (2) (показатели финансовой активности США) и demographic (демографи-

Прогнозирование экономических временных рядов США. Методы R, Autobox, ForecastPro и PP-Autocast. Средняя ошибка прогноза

Временной ряд	Размер выборки	Δ	R-метод	Autobox	ForecastPro	PP-Autocast
Industry	144	6050	706,52	340,72	301,86	303,64
Finance (1)	144	1550	164,48	680,49	794,42	793,03
Finance (2)	132	118	21,07	76,12	71,98	41,40
Demographic	134	2642	53,46	122,08	152,71	286,19

ческие показатели США). Данные временные ряды брались в следующих временных периодах. Ряд Industry – в период с 01.1982 по 01.1994, ряд Finance (1) – с 01.1962 по 01.1974, ряд Finance (2) – с 01.1965 по 01.1976, ряд Demographic – с 01.1983 по 01.1994.

Эксперименты состояли из прогнозирования на один шаг вперед восемнадцати последних элементов представленных временных рядов. В качестве конкурентов методу *R* были выбраны следующие три наиболее известных метода, результаты по которым представлены на сайте ИФ: AutoBox, ForecastPro, PP-Autocast. Глубина вычислений для всех рассмотренных случаев бралась равной 3. Результаты представлены в табл. 2.

Из приведенных данных видно, что метод *R* в случае рядов Finance (1), Finance (2) и Demographic дает существенно меньшую ошибку прогноза по сравнению с другими известными методами. Как показали проведенные экспериментальные результаты, средняя ошибка прогноза у метода *R* примерно в два раза ниже, чем у известных методов.

§ 5. Заключение

Приведенные экспериментальные результаты показывают высокую точность метода на основе универсальной меры. Алгоритм, использующий метод группировки алфавита существенно уменьшает его трудоемкость, что позволяет использовать данный метод на обыкновенных компьютерах при длине ряда в несколько тысяч. Высокая точность предлагаемого метода прогноза подтверждается его сравнением с другими известными методами. Важно также отметить, что предложенный метод легко обобщается на случай прогнозирования многомерных временных рядов.

СПИСОК ЛИТЕРАТУРЫ

1. *Poskitt D.S., Tremayne A.R.* The Selection and Use of Linear and Bilinear Time Series Models // Int. J. Forecasting. 1986. V. 2. № 1. P. 101–114.
2. *Tong H.* Non-linear Time Series: A Dynamical System Approach. Oxford, UK: Clarendon Press, 1990.
3. *Tong H.* Threshold Models in Non-linear Time Series Analysis. Lect. Notes Statist. V. 21. Berlin: Springer, 1983.
4. *Tong H., Lim K.S.* Threshold Autoregression, Limit Cycles and Cyclical Data // J. Roy. Statist. Soc. Ser. B. 1980. V. 42. № 3. P. 245–292.
5. *Engle R.F.* Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation // Econometrica. 1982. V. 50. № 4. P. 987–1007.
6. *Bontempi G.* Local Learning Techniques for Modeling, Prediction and Control. PhD Thesis. IRIDIA, Université Libre de Bruxelles, Belgium, 1999.
7. *Zhang G., Patuwo B.E., Hu M.Y.* Forecasting with Artificial Neural Networks: The State of the Art // Int. J. Forecasting. 1998. V. 14. № 1. P. 35–62.

8. *Cheng H., Tan P.-N., Gao J., Scripps J.* Multistep-Ahead Time Series Prediction // Advances in Knowledge Discovery and Data Mining (Proc. 10th Pacific-Asia Conf. PAKDD'2006. Singapore. April 9–12, 2006). Lect. Notes Comp. Sci. V. 3918. Berlin: Springer, 2006. P. 765–774.
9. *Рябко Б.Я.* Прогноз случайных последовательностей и универсальное кодирование // Пробл. передачи инф. 1988. Т. 24. № 2. С. 3–14.
10. *Рябко Б.Я., Монарев В.А.* Экспериментальное исследование методов прогнозирования, базирующихся на алгоритмах сжатия данных // Пробл. передачи информ. 2005. Т. 31. № 1. С. 74–78.
11. *Ryabko B.* Compression-Based Methods for Nonparametric Prediction and Estimation of Some Characteristics of Time Series // IEEE Trans. Inform. Theory. 2009. V. 55. № 9. P. 4309–4315.
12. *Cover T.M., Thomas J.A.* Elements of Information Theory. Hoboken, NJ: Wiley, 2006.
13. *Рябко Б.Я.* Дважды универсальное кодирование // Пробл. передачи информ. 1984. Т. 20. № 3. С. 24–28.
14. *Кричевский Р.Е.* Связь между избыточностью кодирования и достоверностью сведений об источнике // Пробл. передачи информ. 1968. Т. 4. № 3. С. 48–57.
15. *Krichevsky R.* Universal Compression and Retrieval. Dordrecht: Kluwer Academic, 1993.
16. *Ryabko B.Y., Astola J., Gammerman A.* Adaptive Coding and Prediction of Sources with Large and Infinite Alphabets // IEEE Trans. Inform. Theory. 2008. V. 54. № 8. P. 3808–3813.
17. Gasoline and Diesel Fuel Update. Independent Statistics & Analysis. U.S. Energy Information Administration. <http://www.eia.gov/petroleum/gasdiesel/>.

Лысяк Александр Сергеевич
 Новосибирский государственный университет
 acsemt@gmail.com
Рябко Борис Яковлевич
 Новосибирский государственный университет
 Институт вычислительных технологий СО РАН
 boris@ryabko.net, rbya@yandex.ru

Поступила в редакцию
 19.03.2015
 После переработки
 19.12.2015