

11
Применение энтропии перестановок при анализе хаотических, шумовых и хаотических зашумленных рядов

© С.А. Макаркин, А.В. Стародубов, Ю.А. Калинин

Саратовский государственный университет им. Н.Г. Чернышевского,
410012 Саратов, Россия
e-mail: MakarkinStanislav@mail.ru, StarodubovAV@gmail.com

(Поступило в Редакцию 30 августа 2016 г.)

Рассмотрен один из методов анализа хаотических, зашумленных и хаотических зашумленных рядов — метод анализа энтропии перестановок. С помощью обзора некоторых особенностей информационной энтропии (или энтропии Шеннона) введено понятие энтропии перестановок, описан алгоритм ее расчета и отмечены преимущества такого подхода для анализа временных рядов, а также показано применение данного метода для анализа различных модельных систем и экспериментальных данных.

DOI: 10.21883/JTF.2017.11.45135.2022

Введение

Различие хаотических и шумовых сигналов в настоящее время является одной из важных задач, стоящих перед физиками. Современные медицинские исследования, построенные на анализе данных работы головного мозга и сердца, показали, что анализ характера вариативности сердечного ритма (случайный или хаотический) и характера электроэнцефалограммы головного мозга может выявить приступ эпилепсии или сердечный приступ до их наступления [1]. Сложность этой проблемы заключается в том, что зачастую временные ряды с хаотической (детерминированной) и шумовой (случайной) природой, полученные при наблюдении за естественными системами, практически неразличимы при их сравнении с помощью анализа автокорреляционной функции, ляпуновского показателя, энтропии или фрактальной размерности — т.е. методов анализа типовых орбит предположительно эргодических систем [2,3]. Таких методов существует множество [4–7], но они обладают недостатком — они хорошо подходят для анализа систем с низкой размерностью, при добавлении сколь-нибудь значительного шума они перестают работать точно. Поэтому были разработаны некоторые алгоритмы, которые обходят вышеуказанные проблемы.

В настоящей работе рассмотрен следующий метод — анализ энтропии перестановок, разработанный Vandt и Rompe [8], который находит широкое применение при анализе временных рядов различной природы. Энтропия перестановок применяется при изучении электроэнцефалограмм (ЭЭГ), полученных при анализе работы головного мозга. Veisi и др. [9] иллюстрировал способность энтропии перестановки для классификации нормальной и эпилептической ЭЭГ. Результаты классификации, выполненные при помощи дискриминантного анализа, указали, что методы, использующие энтропию перестановки, могут отличить нормальные и эпилептические сигналы ЭЭГ с точностью, большей чем 97% для

чистого и большей чем 85% для сильно зашумленного сигнала ЭЭГ.

Эпилептические припадки проявляются ограниченными последовательностями симптомов с ограниченной изменчивостью. Schindler и др. [10] предугадали, что эти стереотипические последовательности могут подразумевать, что у иктальной нейронной динамики могут быть детерминированные характеристики, и что они будут, по-видимому, увеличены в иктальных областях мозга.

Также энтропия перестановок может применяться при анализе сердечной активности [11]. Сердечные болезни часто связываются с отклонениями в изменчивости сердечного ритма и в последовательностях характерных интервалов между биениями сердца (Beat to beat interval, BBI). Различие между физиологическими и патологическими образцами BBI представляет собой очень важный диагностический инструмент. Успешная классификация временного ряда BBI сильно зависит от доступности существенных особенностей. Энтропия перестановки, как было показано в [12], значительно улучшила способность отличить изменчивость сердечного ритма при различных физиологических и патологических состояниях. Порядковая статистика паттерна, как доказано, была более эффективной, чем установленные индикаторы изменчивости сердечного ритма при сравнении пациентов, страдающих от застойной сердечной недостаточности, и здоровых пациентов.

В последнее время часто обсуждается применение хаотических сигналов в радиолокации и телекоммуникации. Например, в работе [13] описана возможность применения хаотических дискретных последовательностей для кодировки информации. Как упоминалось ранее, схожесть хаотических сигналов с шумовыми делает затруднительным их распознавание обычными методами, однако остается возможность обнаружения полезной информации с помощью анализа энтропии перестановок.

Разумеется, анализ временных рядов с помощью энтропии перестановок можно применить к самым разным

по своей природе системам, а не только к вышеперечисленным. Например, в работах [14,15] показан анализ энтропии перестановок временных рядов различных финансовых индикаторов.

Информационная энтропия в анализе временных рядов и ее особенности

При рассмотрении энтропии перестановок вспомним определение энтропии Шеннона или информационной энтропии [16]. Информационная энтропия — мера неопределенности, связанной со случайной величиной; определяет количество информации, содержащейся в сообщении (обычно в битах или битах на символ) или

$$H(X) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i), \tag{1}$$

где X — некоторая случайная величина, $p(x_i)$ — вероятность i -го значения, $X(P(X = x_i) = p(x_i))$, n — количество возможных состояний X . Если $H = 0$, то наше знание о процессе будет наибольшим, если же $p(x_i) = 1/n$ для всех i , то информационная энтропия будет наибольшей, а неопределенность процесса — максимальной.

Однако применение классической информационной энтропии при анализе временных рядов имеет следующие недостатки: пренебрежение временными отношениями между значениями временного ряда и использованием знаний об анализируемой системе.

Основным недостатком при анализе временных рядов является пренебрежение временными отношениями между значениями временного ряда, так, что структура и возможные временные закономерности, присутствующие в процессе, не учитываются.

К примеру, возьмем последовательности [1,1,0,0] и [0,1,0,1]. Энтропия Шеннона (1) для них будет одинаковой, хотя очевидно, что это разные последовательности. Для преодоления этого недостатка необходимо, чтобы расчетная формула учитывала не только вероятность получения какого-либо значения, но и информацию о предыдущих значениях.

Следующий недостаток — использование классической энтропией некоторых заранее известных свойств системы (например, информацию об алфавите какого-либо языка при анализе текста) делает затруднительной работу с системами типа „черный ящик“.

Также классические методы вычисления энтропии слабо подходят для исследования сильно нелинейных хаотических режимов.

С учетом вышесказанного Vandt и Pompe [8] разработали свой достаточно простой и мощный метод, который учитывает вышеописанные недостатки.

Описание метода расчета энтропии перестановок

Рассмотрим временной ряд $X = \{x_t : t = 0, 1, 2, 3 \dots N\}$. Для каждого момента времени введем вектор из d значений, где d — вложенная размерность: $\{x_t, x_{t+1}, \dots, x_{t+(d-2)}, x_{t+(d-1)}\}$. Для этого вектора определим перестановки вида (паттерна) $\pi = (r_0 r_1 \dots r_{d-1})$ или $\pi = (\llcorner 0 \gg - \llcorner 1 \gg \dots - \llcorner d-1 \gg)$, для которых выполняется условие

$$x_{s+r_0} \leq x_{s+r_1} \leq \dots \leq x_{s+r_{d-1}}. \tag{2}$$

Поясним на примере. Рассмотрим последовательность чисел $X_1 = \{2, 3, 6, 10, 7\}$. Для $d = 2$ отсчитываем следующие векторы — для начального времени $s = 1$ имеем (2,3); для $s = 2$ имеем (3,6) и так далее, всего четыре пары. Паттерну перестановки $\llcorner 0 \gg - \llcorner 1 \gg$ будут соответствовать векторы (2,3); (3,6); (6,10), так как выполняется условие $x_s \leq x_{s+1}$.

Паттерну перестановки $\llcorner 1 \gg - \llcorner 0 \gg$ будет соответствовать пара, выполняющая условие $x_s \geq x_{s+1}$. Это условие выполняется для последней пары чисел — (10,7). В итоге имеется три пары чисел с паттерном $\llcorner 0 \gg - \llcorner 1 \gg$ и одна пара чисел с паттерном $\llcorner 1 \gg - \llcorner 0 \gg$.

Рассчитаем теперь информационную энтропию для паттернов перестановки (здесь, как очевидно, возможны только два состояния паттернов — $\llcorner 0 \gg - \llcorner 1 \gg$ и $\llcorner 1 \gg - \llcorner 0 \gg$) с заданной вложенной размерностью

$$\begin{aligned} H(d = 2) &= -\sum_{i=1}^2 p_i(\pi) \log_2 p_i(\pi) \\ &= -\left(\frac{3}{4}\right) \log_2 \left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) \log_2 \left(\frac{1}{4}\right) \approx 0.81. \end{aligned}$$

Теперь проведем аналогичные действия для $d = 3$. Для данного d будет существовать три вектора — (2,3,6); (3,6,10); (6,10,7). Первые два будут подходить под паттерн $\llcorner 0 \gg - \llcorner 1 \gg - \llcorner 2 \gg$, для которого $x_s \leq x_{s+1} \leq x_{s+2}$. Третий вектор будет соответствовать паттерну $\llcorner 0 \gg - \llcorner 2 \gg - \llcorner 1 \gg$, удовлетворяющему условию $x_s \leq x_{s+2} \leq x_{s+1}$, и информационная энтропия для данных паттернов будет

$$\begin{aligned} H(d = 3) &= -\sum_{i=1}^2 p_i(\pi) \log_2 p_i(\pi) \\ &= -\left(\frac{2}{3}\right) \log_2 \left(\frac{2}{3}\right) - \left(\frac{1}{3}\right) \log_2 \left(\frac{1}{3}\right) \approx 0.92. \end{aligned}$$

Общей формулой энтропии перестановок для произвольного d будет

$$H(d) = -\sum_{i=1}^d p_i(\pi) \log_2 p_i(\pi). \tag{3}$$

Значения энтропии перестановок будет лежать в следующем промежутке: $H(d) \in [0, \log_2 d!]$.

Нижняя граница соответствует временному ряду, в котором все значения расположены в порядке убывания (или возрастания), верхняя — ряду, в котором реализуются все возможные для данного d паттерны, при этом они все будут иметь одинаковую вероятность обнаружения.

Поведение энтропии перестановок в случае анализа хаотического и хаотического зашумленного сигналов

В абсолютно случайных (шумовых) процессах появление того или иного паттерна перестановки должно происходить с одинаковой по отношению к другим вероятностью. Из этого следует, что при достаточно долгом наблюдении за процессом все возможные паттерны перестановки рано или поздно появятся. В хаотических процессах присутствует хоть и сложная, но все же упорядоченность. Поэтому какие-то паттерны перестановки будут иметь большую вероятность появления, а некоторые могут не появиться. Это обуславливает большее значение энтропии перестановок при большем зашумлении хаотического процесса, чем значение энтропии перестановок во временном ряду с меньшим зашумлением.

Из вышесказанного становится ясно, что для различения систем хаотических, хаотических зашумленных и шумовых достаточно проанализировать количество всех паттернов перестановки N , которые можно наблюдать во временном ряду с длительностью T и вложенной размерностью (она же порядок энтропии перестановок) d и не рассчитывать собственно энтропию перестановок. При одинаковом d хаотическому ряду будет присуще более быстрое насыщение зависимости $N(T)$, чем хаотическому зашумленному ряду. При этом максимальное N будет выше у более зашумленных систем.

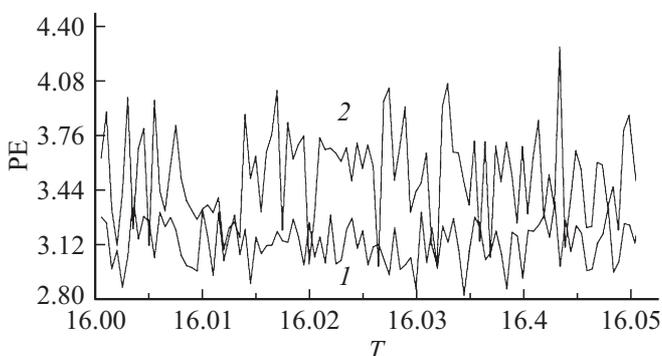


Рис. 1. Увеличенный фрагмент графика энтропии перестановок PE для системы (5) для зашумленных рядов с разными амплитудами центрированного гауссова шума, a : 1 — 0, 2 — 0.1.

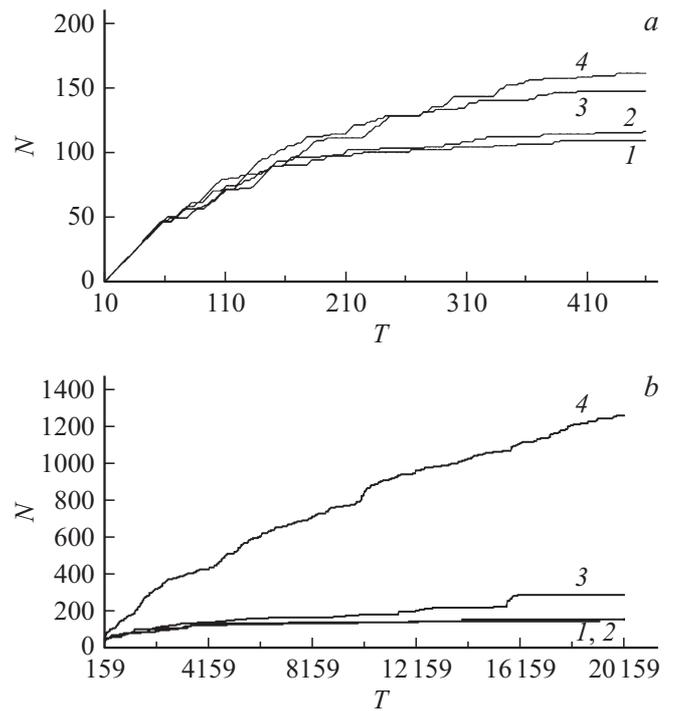


Рис. 2. Графики зависимости количества наблюдаемых паттернов N от длины временного зашумленного ряда T для разных амплитуд центрированного гауссова шума. a — система (4), b — система (5); 1 — $a = 0$, 2 — $a = 0.00025$, 3 — $a = 0.001$, 4 — $a = 0.004$.

Для выявления характерных зависимостей рассмотрим два типа систем — с дискретным и непрерывным временем. Система с дискретным временем, а именно логистическое отображение, является одной из наиболее изученных и наглядных, при этом удобна для проведения численных расчетов. Логистическое отображение будет иметь вид

$$x_{t+1} = Lx_t(1 - x_t) + R(a). \quad (4)$$

С параметром $3.5 \leq L \leq 4$ (т.е. взят параметр, соответствующий хаотической области). $R(a)$ — функция, описывающая гауссов центрированный шум с минимальным значением $R_{\min} = -a$ и максимальным $R_{\max} = a$.

Однако, являясь хорошими модельными системами, системы с дискретным временем не всегда достаточно точно описывают реальные процессы, так как не содержат информации о системе между отсчетами дискретного времени, а также дискретные отображения, полученные из систем с потоковым временем, необратимы — восстановить систему в начальном виде из соответствующего ей дискретного отображения (к примеру, сечения Пуанкаре) невозможно. Поэтому для большего соответствия реальному процессу необходимо рассмотреть соответствующую ему систему с непрерывным временем. В настоящей работе рассматривается система Лоренца, которая описывает процесс конвекции в жидкости и является одной из самых изученных [17].

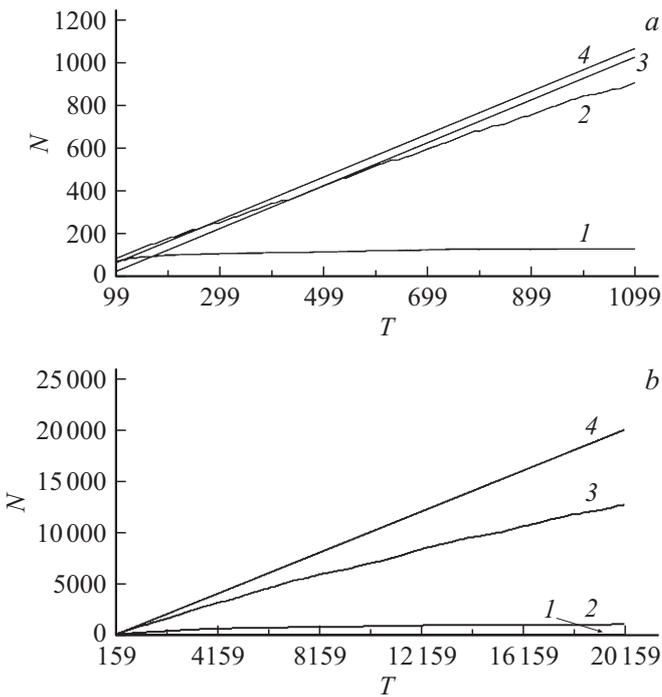


Рис. 3. Графики зависимости количества наблюдаемых паттернов N от длины временного не зашумленного ряда T для разных вложенных размерностей d . a — система (4), b — система (5); 1 — $d = 10$, 2 — $d = 20$, 3 — $d = 40$, 4 — $d = 80$.

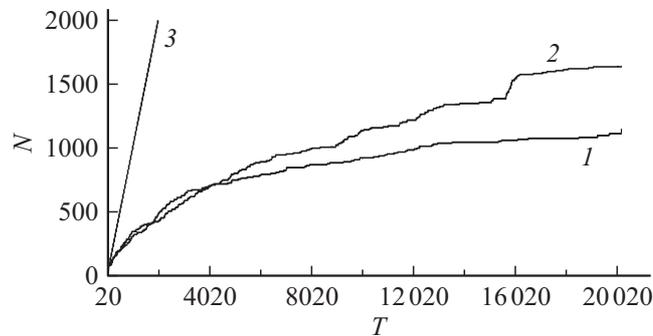


Рис. 4. Графики зависимости количества наблюдаемых паттернов N от длины временного ряда T . 1 — реализация без шума, 2 — реализация с амплитудой шума $a = 0.01$, 3 — шумовая реализация.

Система Лоренца имеет вид

$$\begin{cases} \frac{dx}{dt} = -s(x - y) + R_1(a), \\ \frac{dy}{dt} = -xz + rx - y + R_2(a), \\ \frac{dz}{dt} = xy - bz + R_3(a). \end{cases} \quad (5)$$

Здесь $s = 16$ (или $16 \leq s \leq 16.5$ в случае анализа зависимости РЕ от параметра), $r = 45.92$, $b = 4$ — параметры; a , $R_1(a)$, $R_2(a)$, $R_3(a)$ — аналогичны $R(a)$ и a из (1). Для решения этой системы использовался численный метод Рунге-Кутты четвертого порядка с шагом $h = 0.01$.

Как и предполагалось, кривая зависимости РЕ от параметра системы для системы (4) для последовательности с ббльшей амплитудой шума лежит выше, чем кривая для меньших значений амплитуды шума [8]. Для системы Лоренца будет прослеживаться аналогичная зависимость (рис. 1).

Теперь определим, какое время наблюдения можно считать достаточно длинным. При слишком малом времени наблюдения между сигналами с разной зашумленностью не будет разницы. Другими словами, все возможные паттерны не успеют появиться как в системах с сильным шумом, так и в системах со слабым шумом. Слишком большое время наблюдения не внесет значительных изменений в метод расчета и сравнения РЕ для разных систем, но, очевидно, сильно увеличит время вычислений.

Важным также является определение величины вложенной размерности d . Естественно, оптимальная величина d будет такой, что количество всех встречающихся типов паттернов в системе будет $d!$, но заранее определить это не представляется возможным. Для систем со сложной динамикой не имеет смысла брать малые d , так как при них учитывается слишком короткая предыстория временного ряда, что приведет к малому разнообразию паттернов, которые будут все время повторяться во временном ряду.

Определение оптимальных d и T зависит от каждой конкретной системы. Для системы (4) оптимальным временем для рядов с указанными амплитудами шума и $d = 10$ будет $T = 450$, так как количество паттернов перестановки при больших T уже не увеличивается (рис. 2, a). В системе (5) (рис. 2, b) оптимальную величину T для разных величин амплитуды шума подобрать сложнее, так как увеличивается время переходного процесса по сравнению с системой (4) и насыщение кривой $N(T)$ для $a = 0$ и $a = 0.001$ происходит на сравнительно малой $T \approx 5159$, для $a = 0.01$ оптимальное $T \approx 16159$, для $a = 0.1$ оптимальное $T \gg 20000$.

Выбор d взаимосвязан с выбором T , так как с малым T не все паттерны перестановки успеют проявиться, как ранее упоминалось. Взяв оптимальное T

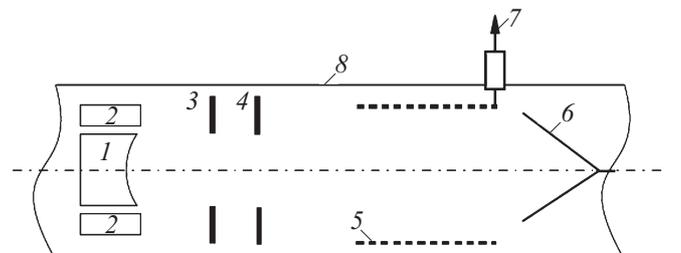


Рис. 5. Принципиальная схема отпаянного лабораторного макета низковольтного генератора: 1 — катод, 2 — управляющий электрод, 3 — первый анод, 4 — второй анод, 5 — отрезок спиральной замедляющей системы, 6 — коллектор, 7 — коаксиальный широкополосный СВЧ вывод, 8 — труба дрейфа.

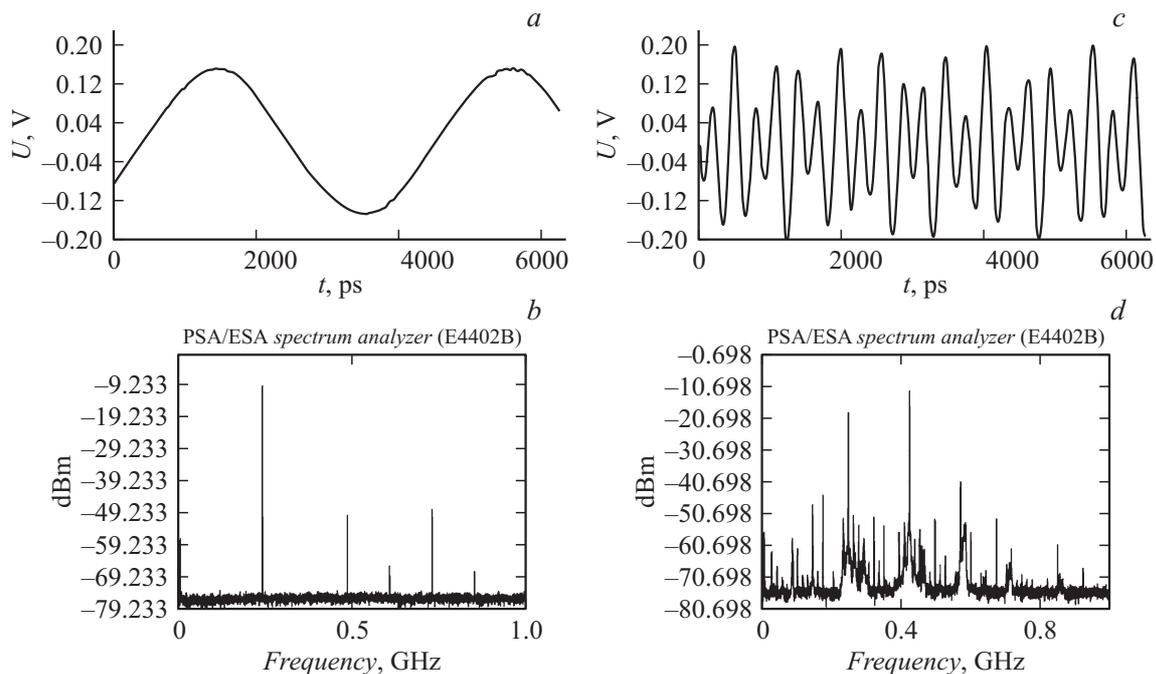


Рис. 6. *a* — временная развертка ряда «1», *b* — спектр временной развертки ряда «1», *c* — временная развертка «2», *d* — спектр временной развертки «2».

для какого-либо d , с увеличением d надо увеличивать и T , чтобы пронаблюдать появление насыщения $N(T)$. В работе [7] для применения метода к дискретным системам советуется брать $6 \leq d \leq 12$.

Продолжая анализ влияния вложенной размерности d , рассмотрим количество встречаемых паттернов в системах (4) (рис. 3, *a*) и (5) (рис. 3, *b*) для разных d при отсутствии шума. На графике видно, что для $d > 40$ зависимость N от T становится близка к линейной вида $N = KT$, где $K \approx 1$, что соответствует максимальной скорости прироста N (т.е. с каждой итерацией появляется новый паттерн перестановки), при больших d других видов зависимости появляться не будет. Такая кривая будет достигать насыщения только при очень больших временах наблюдения ($T \gg 20000$). Однако, как было показано выше, четко различать зашумленные сигналы для данных условий можно и при меньших d .

Теперь продемонстрируем поведение N для шумового временного ряда и сравним с хаотическим. Возьмем функцию $R(a)$ в системе (4) такую, что ее значения будут лежать в тех же пределах, что и значения хаотического и хаотического зашумленного ряда этой системы.

Рассмотрим график $N(T)$ для этих реализаций (рис. 4). Как и ожидалось, для шумовой реализации характерно очень быстрое нарастание количества паттернов в зависимости от длительности временной реализации.

Анализ экспериментальных данных

Теперь применим метод анализа количества паттернов перестановки к данным, полученным экспериментально.

Эксперимент проводился с макетом низковольтного генератора виркаторного типа, схема которого приведена на рис. 5.

Макет представляет собой электровакуумный прибор с выходной мощностью до 1 W в диапазоне генерируемых частот до 3 GHz. Для формирования внутренней распределенной обратной связи в подобных приборах используется дополнительное торможение электронного потока за счет подачи тормозящего напряжения U_{inh} на коллектор ($U_{inh} < U_0$, где U_0 — ускоряющее напряжение). Таким образом, в пространстве дрейфа создается электрическое поле, которое тормозит электронный пучок, инжектируемый с катода. Питание экспериментального образца генератора осуществлялось в непрерывном режиме. В качестве основных управляющих параметров были выбраны величина ускоряющего напряжения U_0 и напряжение на коллекторе U_{inh} . Наблюдение и анализ динамики физических процессов генерации СВЧ излучения в макете низковольтного виркатора осуществлялись по спектрам мощности и временным реализациям генерируемых сигналов. Для регистрации спектров мощности сигнала использовался цифровой анализатор спектра Agilent Technologies ESA-E Series Spectrum Analyzer E4402B (9.1 kHz–3.0 GHz). Регистрация временных реализаций выходных СВЧ сигналов осуществлялась с использованием цифрового запоминающего осциллографа реального времени Agilent Technologies Infiniium DSO 81004B Series oscilloscopes. Эксперимент заключался в получении сигналов различного типа на выходе прибора. Полученные сигналы оцифровываются во временные ряды, к которым и будет применяться описанный в

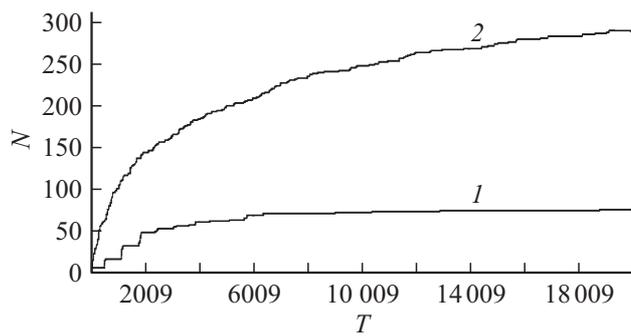


Рис. 7. Графики зависимости количества наблюдаемых паттернов N от длины временного ряда T . 1 — реализация «1», 2 — реализация «2».

настоящей работе метод. Из всех рядов выделим два характерных, которые обозначим «1», «2».

Рассмотрим ряд «1». На временной развертке (рис. 6, *a*) и спектре (рис. 6, *b*) видно, что ряд «1» представляет собой зашумленный периодический режим.

Перейдем к рассмотрению ряда «2». Анализ временной развертки (рис. 6, *c*) и вида спектра (рис. 6, *d*) показывает, что ряд «2» представляет собой шумоподобную широкополосную временную реализацию. В работе [18] было показано, что наблюдаемые режимы широкополосной шумоподобной генерации представляют собой режимы генерации принципиально хаотических сигналов. Указанный вывод сделан на основе анализа сценариев перехода к режимам широкополосной шумоподобной генерации. Таким образом, выбранная для анализа экспериментальная реализация является хаотической.

Теперь построим для выбранных рядов данных зависимость количества прироста паттернов от длительности наблюдения для $d = 6$ (рис. 7). Как и в случае с временными рядами модельных систем (4) и (5), зависимость $N(T)$ для систем с более сложной динамикой (более высокий уровень шума, хаотичность) лежит выше, чем для систем с более простой. Также из рис. 7 отчетливо видно, что зависимость $N(T)$ для ряда «2» выходит на насыщение, что говорит о детерминированности наблюдаемого процесса, другими словами, исследуемая временная реализация является хаотической, что подтверждается выводами в работе [18].

Заключение

В настоящей работе описан алгоритм расчета энтропии перестановок, который позволяет анализировать и различать временные ряды с хаотической и шумовой динамикой, а также предложено и обосновано применение анализа динамики прироста количества паттернов перестановки для аналогичных целей. Также было показано преимущество используемого метода по сравнению с анализом энтропии Шеннона. Приведены результаты

применения используемого метода к временным хаотическим и хаотическим зашумленным рядам модельных систем с дискретным (логистическое отображение) и непрерывным (система Лоренца) временем, а также данным, полученным в ходе экспериментальных исследований систем вакуумной СВЧ электроники. Также приведено определение оптимальных параметров в алгоритмах расчета.

В дальнейшем предполагается рассмотреть применение описанного метода для детектирования полезной информации в различных сигналах, использующих кодирование информации с помощью хаоса. Также предполагается дальнейшее проведение лабораторных экспериментов как с описанным в настоящей работе лабораторным макетом, так с другими подобными системами.

Работа выполнена при поддержке гранта РФФИ (грант № 16-02-00238-а), а также Министерства образования и науки РФ в рамках проектной части государственного задания (проект № 3.859.2017/4.6).

Список литературы

- [1] *Kantz H. et al. (eds.). Nonlinear Analysis of Physiological Data. Berlin: Springer, 1996.*
- [2] *Eckmann J.-P. Ruelle D. // Rev. Mod. Phys. 1985. Vol. 57. P. 617.*
- [3] *Pesin Ya.B. Dimension theory in dynamical systems. University of Chicago Press, 1998.*
- [4] *Grassberger P., Procaccia I. // Phys. Rev. 1983. Vol. 28A. P. 2591.*
- [5] *Ding M., Grebogi C., Ott E., Sauer T., Yorke J. A. // Phys. Rev. Lett. 1993. Vol. 70. P. 3872.*
- [6] *Abarbanel H.D.I. Analysis of Observed Chaotic Data. NY.: Springer, 1996.*
- [7] *Bofetta G., Cencini M., Falcioni M., Vulpiani A. // Phys. Rep. 2002. Vol. 356. P. 367–474.*
- [8] *Bandt C., Pompe B. // Phys. Rev. Lett. 2002. Vol. 88. P. 174102:1–174102:4.*
- [9] *Veisi I., Pariz N., Karimpour A. // In Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, Boston, MA, USA, 2007. P. 200–203.*
- [10] *Schindler K., Gast H., Stieglitz L., Stibal A., Hauf M., Wiest R., Mariani L., Rummel C. // Epilepsia. 2011. Vol. 52. P. 1771–1780.*
- [11] *Bian C., Qin C., Ma Q.D.Y., Shen, Q. // Phys. Rev. E. 2012. Vol. 85. P. 021906:1–021906:7.*
- [12] *Parlitz U., Berg S., Luther S., Schirdewan A., Kurths J., Wessel N. // Comput. Biol. Med. 2012. Vol. 42. P. 319–327.*
- [13] *Kolesov V.V., Potapov A.A. // Electromagnetic Phenomena. 2005. Vol. 5. N 2(15). P. 89–104.*
- [14] *Zanin M. // Chaos. 2008. Vol. 18. P. 013119.*
- [15] *Zunino L., Zanin M., Tabak B.M., Perez D.G., Rosso O.A. // Phys. A. 2009. Vol. 388. P. 2854–2864.*
- [16] *Shannon C.E. // Bell Syst. Tech. J. 1948. Vol. 27. P. 379–423.*
- [17] *Кузнецов С.П. Динамический хаос. Серия „Современная теория физики колебаний и волн“. М.: Физматлит, 2001. С. 43–63.*
- [18] *Калинин Ю.А., Стародубов А.В., Кузнецов Н.Н. // ЖТФ. 2013. Т. 83. Вып. 6. С. 151–154.*