

Вычисление корреляционной размерности для физиологических временных рядов

М.И. Шпитонков¹

¹ Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

Аннотация. Работа посвящена вычислению корреляционной размерности электрокардиограмм (ЭКГ) пациентов с болезнью Паркинсона, болезнью Альцгеймера и здоровых пациентов. Было показано, что у здоровых пациентов корреляционная размерность не выходит на насыщение, тогда как у больных пациентов она принимает вполне определенные значения. Таким образом, появляется возможность диагностики типа заболевания на основе данных о корреляционной размерности электрокардиограммы.

Ключевые слова: корреляционная размерность, метод автокорреляционной функции, система Лоренца, оптимальная временная задержка.

DOI: 10.14357/20790279200209

Введение

В последнее время получили распространение методы анализа физиологических временных рядов, основанные на применении формального аппарата теории динамических систем и, в частности, детерминированного хаоса. В шестидесятых годах прошлого века в динамической системе, при изучении турбулентности было обнаружено очень сложное движение, которое воспринималось как хаотическое. Для характеристики таких движений впервые ввели понятие динамического хаоса. Затем был введен новый математический образ динамического хаоса – странный аттрактор. Определение «странный» подчеркивает два основных свойства аттрактора. Во-первых, необычность его геометрической структуры. Размерность странного аттрактора является дробной (фрактальной). Во-вторых, странный аттрактор – это притягивающая область для траекторий из окрестных областей. При этом все траектории внутри странного аттрактора динамически неустойчивы, что выражается в сильной экспоненциальной расходимости близких в начальный момент траекторий. Главная идея применения методов хаотической динамики к анализу временных рядов состоит в следующем. Оказывается основная структура хаотической системы, содержащая в себе всю информацию о системе, а именно, ее аттрактор (подмножество фазового пространства, притягивающее траектории в пределе бесконечного времени),

может быть восстановлена через измерение только одной наблюдаемой этой динамической системы, фиксированной как временной ряд. Согласно методу Грасбергера и Прокаччи [1], процедура реконструкции фазового пространства и восстановления хаотического аттрактора системы при динамическом анализе временного ряда, сводится к построению так называемого лагового или восстановленного пространства с помощью метода задержки. Векторы в новом пространстве, пространстве вложения, сформированы из значений временного ряда скалярных измерений с временным запаздыванием. Или, другими словами, реальный аттрактор динамической системы и «аттрактор», восстановленный в лаговом пространстве по временному ряду (псевдоаттрактор), являются топологически эквивалентными и обладают одинаковыми обобщенными фрактальными размерностями, ляпуновскими показателями и другими численными характеристиками. Одной из величин, вычисляемых при помощи данного формального аппарата, является корреляционная размерность восстановленного аттрактора. При помощи этой величины можно судить, например, насколько синхронно протекают процессы в головном мозге, или насколько вариативна его деятельность [2,3]. В основе вычисления корреляционной размерности лежит теорема Такенса [4], пользуясь которой по динамике одного временного ряда можно судить об эволюции всей системы, порождающей его, по-

строив псевдоаттрактор, обладающий такими же метрическими характеристиками, как аттрактор системы в фазовом пространстве. В том случае, если анализируемый временной ряд является реализацией случайного процесса, то восстановленный псевдоаттрактор представляет собой бесструктурное облако точек, которое при последовательном бесконечном увеличении размерности вложения лагового пространства, подобно газу, заполняет весь предоставленный ему объем. Однако для хаотических данных корреляционная размерность будет, в конечном счете, насыщаться при ее истинном значении. Для случайных данных такого насыщения не наблюдается, и корреляционная размерность растет монотонно. Вычисление корреляционной размерности сопряжено с рядом трудностей. Во-первых, математические теоремы и разработанные на их основе алгоритмы имеют теоретическое обоснование лишь для бесконечных и не зашумленных хаотических процессов [5]. Во-вторых, вычислительные процедуры, применяемые разными исследователями, могут несколько различаться между собой, и на данный момент не существует единого мнения относительно выбора параметров для вычислений [6,7,8]. Это ведет к трудностям при сравнении результатов, полученных разными исследователями. В итоге обработка больших объемов экспериментальных данных достаточно трудоёмка.

1. Понятие корреляционной размерности

Для динамических систем принятым представлением развития процесса во времени является построение «портрета» в фазовом пространстве (то есть пространстве, координатами которого являются переменные состояния). В пространстве с числом независимых переменных большим или равным трем в нелинейной динамической системе может существовать странный аттрактор – притягивающее множество в фазовом пространстве, в котором расположены хаотические траектории [5].

Размерностью вложения m называется наименьшая целая размерность пространства, содержащего весь аттрактор. Она соответствует количеству независимых переменных, однозначно определяющее установившееся движение динамической системы.

Множество, соответствующее странному аттрактору, фрактально. Фрактальное множество (самоподобный объект) – характеризуется дробной фрактальной размерностью (точнее, целым спектром различно определяемых размерностей, совпадающих для регулярных фракталов, но различающихся для природных систем [5]).

Важной количественной характеристикой аттрактора, несущей информацию о степени сложности поведения динамической системы, является корреляционная размерность D_c [3,5]. Алгоритм расчета D_c основан на вычислении корреляционного интеграла, в качестве которого выступает функция $C(\delta)$, для каждого δ равная нормированному числу пар точек рассматриваемого объекта, расстояние

$$C(\delta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=i+1}^{n-1} \theta(\delta - |y_i - y_j|)$$

которыми не превосходит δ :

где $\theta(x)$ – функция Хевисайда, $|y_i - y_j|$ – расстояние между образами i -й и j -й точек временного ряда в восстановленном пространстве.

Величина суммы зависит от δ , причем, если

эта зависимость имеет степенной вид $C(\delta) \sim \delta^{D_c}$, то исследуемое множество фрактально, а величина D_c – его корреляционная размерность.

2. Алгоритм вычисления корреляционной размерности

Для практического вычисления корреляционной размерности на графике $\ln(C(f)) = f(\ln(\delta))$ выделяют область линейной зависимости (области скейлинга) и функция аппроксимируется прямой линией методом наименьших квадратов. Тогда тангенс угла наклона графика является размерностью D_c .

Для известной динамической системы $\hat{m}$ и D_c легко определить – ведь известны все компоненты вектора $x(t) = \{x_1(t), \dots, x_m(t)\}$, описывающего поведение системы в фазовом пространстве (так, для системы Лоренца $D_c = 2,06$, $m = 3$). Однако при изучении природных систем, в том числе нейрофизиологических, обычно приходится иметь дело с сигналом, который выглядит достаточно сложно и кажется похожим на случайный [6]. Для природных объектов измерение всех компонент, характеризующих систему, невозможно – хотя бы потому, что они не все известны. Однако согласно теореме Такенса [4] можно восстановить некоторые свойства аттрактора (например, m и D_c) по временной последовательности одной из составляющих вектора $x(t)$ [4].

Методика основана на построении псевдоаттрактора, где в качестве компонент вектора служит сама измеренная последовательность, но взятая с некоторой временной задержкой:

$$xp(t) = \{x(t), x(t+\tau), x(t+2\tau), \dots, x(t+(m-1)\tau)\} \quad (1)$$

Поскольку заранее размерность вложения m неизвестна, то процедура сводится к следующему:

1. Последовательно увеличивают размерность фазового пространства и добавляют компоненты псевдовектора $xp(t)$.
2. При каждом $m=2,3,\dots$ вычисляют корреляционную размерность D_c и строят зависимость $D_c(m)$. Сначала при добавлении новых компонент псевдовектора корреляционная размерность растет. Это значит, что мы еще не достигли нужного количества измерений, и, соответственно, нужной сложности, степень которой характеризует D_c .
3. Начиная с некоторой размерности m пространства, корреляционная размерность D_c достигает насыщения и перестает изменяться. Значение m , при котором это происходит, является оценкой минимальной размерности вложения, а значение D_c – оценкой корреляционной размерности аттрактора.

Для выбора временной задержки τ используются следующие подходы:

1. Метод автокорреляционной функции 2. Метод взаимной информации

Эти два метода основаны на предположении, что временная задержка является оптимальной, если координаты реконструированного аттрактора являются максимально независимыми притом, что сама задержка τ имеет как можно меньшее значение.

Метод автокорреляционной функции:

Рассматривается автокорреляционная функция

$$R(\tau) = \frac{1}{N_1} \sum_{t=1}^{N_1} y(t)y(t+\tau), \quad \text{где} \quad N_1 = N - \tau,$$

$$y(t) = x(t) - M(x) \text{ – центрированная версия ряда}$$

Для каждого конкретного значения τ функция $R(\tau)$ возвращает коэффициент корреляции между исходным временным рядом и его версией, полученной при помощи временной задержки на τ шагов.

Оптимальная временная задержка обычно выбирается в соответствии с первым нулевым (либо близким к нулю) значением автокорреляционной функции. Данный метод является достаточно простым и не требует больших вычислений. Но поскольку вместо свойства независимости используется неэквивалентное ему свойство некоррелированности, то полученное значение задержки τ не всегда является оптимальным.

Метод взаимной информации:

Пусть $(a,b) \in R^1$ – минимальный интервал, содержащий все значения временного ряда. Разобьем данный интервал на L равных частей (количество интервалов разбиения обычно выбирают по

известной формуле Старка [4] $L = ([\log_2 N] + 1)$. Обозначим событие «значение $x(t)$ принадлежит i -му интервалу» через A_i , а событие «значение $x(t+\tau)$ принадлежит j -му интервалу» через B_j . Тогда функция взаимной информации определяется соотношением:

$$I(\tau) = \sum_{i=1}^L \sum_{j=1}^L P(A_i B_j) \log_2 \frac{P(A_i B_j)}{P(A_i)P(B_j)} \quad (2)$$

где P – вероятность соответствующего события.

Функция взаимной информации является, с одной стороны, более точной мерой независимости, чем автокорреляционная функция, а с другой – более сложной в вычислительном смысле. В данном методе задержка выбирается в соответствии с первым минимумом функции $I(\tau)$.

Для ускорения расчетов можно использовать модифицированную версию формулы:

$$I_1 = - \sum_{i=1}^L \sum_{j=1}^L [P(A_i B_j) - P(A_i)P(B_j)]^2 \quad (3)$$

Функция $I_1(\tau)$ ведет себя так же как $I(\tau)$, то есть обе эти функции возрастают на одних и тех же промежутках, а также достигают экстремумов в одних и тех же точках. Время вычислений же сокращается на 10-15%.

Как следует из определения размерности вложения, она соответствует числу независимых переменных, описывающих систему. Таким образом, восстанавливая размерность вложения, мы получаем информацию о сложности системы. Из этого следует также возможность отличить динамическую систему со сложным поведением (но характеризующуюся конечным m), и случайный (стохастический) шум, который описывается (теоретически) бесконечно большим числом независимых переменных. Для полностью случайной системы увеличение m на единицу приводит к увеличению D_c также примерно на 1, то есть $D_c \sim m$.

Отметим, что данный метод предъявляет большие требования к длине ряда, если исследуемая система описывается достаточно большим числом управляющих переменных. Существует оценка зависимости минимальной необходимой длины ряда $N_{\min}$ от корреляционной размерности:

$$N_{\min} = 10^{2+0,4D_c}.$$

Для примера, подобная методика применялась для анализа сигнала ЭЭГ и позволила выявить интересные закономерности для здоровых и больных эпилепсией детей [6]. В рамках данной работы

мы получили подобные результаты для больных болезнями Альцгеймера и Паркинсона и здоровыми пациентами.

3. Расчет корреляционной размерности для системы Лоренца

При реализации описанного выше алгоритма необходимо удостовериться в его корректности и провести начальное тестирование подсчета корреляционной размерности. Отличным примером подобного тестирования может служить система Лоренца, для которой, как известно, $D_c = 2,06$, $m = 3$.

Система Лоренца – система трех нелинейных дифференциальных уравнений первого порядка:

$$\begin{aligned} \dot{x} &= P_r(y-x) \\ \dot{y} &= -y + rx - xz \\ \dot{z} &= xy - bz \end{aligned} \quad (4)$$

Решения этой системы уравнений в широкой области параметров являются нерегулярными функциями времени и по многим своим характеристикам неотличимы от случайных функций. Система Лоренца служит моделью описания тепловой конвекции в горизонтальном слое жидкости, по

догреваемой снизу (P_r – число Прандтля, $r = \frac{Ra}{Ra_c}$ –

приведенное число Рэлея, b определяется выбором моды в Фурье разложении поля скорости и температуры).

С помощью вышеописанного метода для системы Лоренца было получено значение корреляционной размерности $D_c = 2,1$, что достаточно хорошо согласуется с теорией. Это позволяет переходить к расчетам на реальных данных.

4. Расчет корреляционной размерности для эмпирических данных

Для расчета корреляционной размерности были исследованы ЭКГ пациентов с болезнью Паркинсона (65 человек), болезнью Альцгеймера (53 человека) и здоровые пациенты (48 человек). Размер ЭКГ был около 2000 точек для каждого пациента.

Большой проблемой при работе с реальными данными было то, что значение смещения τ , определенное как с помощью метода корреляционной функции, так и с помощью метода взаимной информации, составляло порядка 20% от длины всего ряда.

Построив зависимость корреляционной размерности от размерности вложенного аттрактора, мы можем найти, когда график выходит на насы-

щение и таким образом получить итоговое значение корреляционной размерности.

На Рис.1 показана на одном графике зависимость корреляционной размерности от размерности вложенного аттрактора для разных типов данных.



Рис. 1. Зависимость корреляционной размерности от размерности вложенного аттрактора.

В нашем случае не наблюдается насыщения зависимости корреляционной размерности D_c от размерности вложенного аттрактора для здоровых людей, т.е. можно сказать, что ЭКГ здорового человека ведет себя как случайный процесс. Для больных пациентов наблюдается совсем другая ситуация. Отчетливо видно насыщение зависимости корреляционной размерности от размерности вложенного аттрактора. При этом корреляционная размерность ЭКГ больных болезнью Паркинсона $D_c \approx 3,3$, больных болезнью Альцгеймера – $D_c \approx 4$.

Заключение

В представленной работе проанализировано применение методов стохастической динамики к анализу физиологических временных рядов. Так с использованием корреляционной размерности были исследованы ЭКГ здоровых пациентов и пациентов с болезнью Паркинсона и Альцгеймера. Было показано, что у здоровых пациентов корреляционная размерность не выходит на насыщение, тогда как у больных пациентов (болезнь Паркинсона и болезнь Альцгеймера) она принимает вполне определенные значения. Таким образом, появляется дополнительная возможность диагностики типа заболевания на основе данных о корреляционной размерности временного ряда $R-R$ интервалов электрокардиограммы.

Литература

1. Grassberger P., Procaccia I. Measuring the strangeness of strange attractors. // *Physica D*. №1, v.9, 1983. P. 189-208.
2. Семенова Н.Ю., Захаров В.С. Анализ корреляционной размерности данных ЭЭГ при эпилепсии у детей. // *Нелинейный мир*, №3, т.8, 2010, С. 180-188.

3. Головкин В.А. Нейросетевые методы обработки хаотических процессов. // В книге «Лекции по нейроинформатике». – М.: МИФИ, 2005. С. 43-88.
4. Takens F. On the numerical determination of the dimension of an attractor. In: Dynamical systems and bifurcations (Eds. B.L.J. Braaksma, H.W. Broer and F. Takens). Lect. Notes in Math. 1125, Springer, Heidelberg. 1985.
5. Малинецкий Г.Г., Потанов А.Б. Современные проблемы нелинейной динамики. Москва. УРСС.2002, 358 с.
6. Меклер А.А. Применение аппарата нелинейного анализа динамических систем для обработки сигналов ЭЭГ. // Вестник новых медицинских технологий, №1, т. XIV, 2007, С. 73-76.
7. Койчубеков Б.К., Сорокина М.А., Пашев В.И. Особенности нелинейной динамики ЭЭГ в различных возрастных группах. International Journal of Experimental Education, 2013, № 4, С. 68–72.
8. Jelles B., Scheltens Ph., van der Flier W.M., Jonkman E.J., Lopes da Silva F.H., Stam C.J. Global dynamical analysis of the EEG in Alzheimer's disease: Frequency-specific changes of functional interactions. J. Clinical Neurophysiology, 2008, v. 119, p. 837–841.

Шпитонков Михаил Иванович. Вычислительный центр им. А.А.Дородницына ФИЦ ИУ РАН. Москва. Старший научный сотрудник. Кандидат физико-математических наук, доцент. Количество печатных работ: 95. Область научных интересов: математические модели в биологии, медицине, эпидемиологии, физиологии, экологии, экономике. E-mail: mixash@bk.ru

The calculation of correlation dimension to physiological time series

M.I. Shpitionkov¹

¹ Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

Abstract. This work is devoted to calculating the correlation dimension of electrocardiograms (ECG) in patients with Parkinson's disease, Alzheimer's disease, and healthy patients. It was shown that in healthy patients, the correlation dimension does not reach saturation, whereas in sick patients it takes quite certain values. Thus, it is possible to diagnose the type of disease based on data on the correlation dimension of the electrocardiogram.

Keywords: correlation dimension, autocorrelation function method, Lorentz system, optimal time delay.

DOI: 10.14357/20790279200209

References

1. Grassberger P., Procaccia I. Measuring the strangeness of strange attractors. // Physica D. №1, v.9, 1983. P. 189-208.
2. Semenova N.Yu., Zhakharov V.S. 2010. Analiz korrelyatsionnoj razmernosti dannykh EEG pri epilepsii u detej. [Analysis of the correlation dimension of EEG data in children with epilepsy]. Nelinejnyj mir. [Non-linear world]. №3, v.8, p. 180-188.
3. Golovko V.A. Nejrosetevye metody obrabotki khaoticheskikh protsessov. [Neural network methods for processing chaotic processes]. V knige «Lekcii po nejroinformatike». [In the book "Lectures on Neuroinformatics"]. Moscow. MIFI, 2005. p. 43-88.
4. Takens F. On the numerical determination of the dimension of an attractor. In: Dynamical systems and bifurcations (Eds. B.L.J. Braaksma, H.W. Broer and F. Takens). Lect. Notes in Math. 1125, Springer, Heidelberg. 1985.
5. Malinetsky G.G., Potapov A.B. Sovremennye problemy nelinejnoj dinamiki [Modern problems of nonlinear dynamic]. Moscow. URSS.2002, 358 p.
6. Mekler A.A. 2007. Primenenie apparata nelinejnogo analiza dinamicheskikh sistem dlia obrabotki signalov EEG. [Application of the device for nonlinear analysis of dynamic systems for processing EEG signals]. Vestnik novykh meditsynskih tekhnologij. [Bulletin of new medical technologies]. №1, т. XIV, p. 73-76.
7. Kojchubekov B.K., Sorokina M.A., Pashev V.I. 2013. Osobennosti nelinejnoj dinamiki EEG v razlichnykh vozrastnykh gruppakh. [The peculiarities of nonlinear dynamics of EEG in different age groups]. International Journal of Experimental Education, № 4, p. 68–72.
8. Jelles B., Scheltens Ph., van der Flier W.M., Jonkman E.J., Lopes da Silva F.H., Stam C.J. Global dynamical analysis of the EEG in Alzheimer's disease: Frequency-specific changes of functional interactions. J. Clinical Neurophysiology, 2008, v. 119, p. 837–841.

Shpitionkov Mikhail Ivanovitch. Ph.D, Dorodnitsyn computing center Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, 44/2 Vavilova str., Moscow, 119333, Russia. E-mail: mixash@bk.ru.