

Распознавание биосигналов с хаотическими свойствами по оценкам условной энтропии

Л. А. Манило

Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
lmanilo@yandex.ru

Аннотация. Рассмотрены теоретические вопросы и результаты экспериментов по распознаванию биосигналов методами, основанными на оценках условной энтропии. Показано, что скорректированные энтропийные показатели могут эффективно использоваться для описания регулярных и хаотических составляющих анализируемых процессов, а также для решения задач медицинской диагностики.

Ключевые слова: анализ биомедицинских сигналов; хаотические компоненты; условная энтропия; распознавание аритмий

I. ВВЕДЕНИЕ

Исследование сигналов, генерируемых живыми системами, представляет особый интерес, поскольку развивает новые подходы к анализу сложных систем, отличающихся нелинейной динамикой поведения, а также способствует решению ряда прикладных задач диагностики состояний живого организма. Наличие математических моделей, описывающих поведение сложной системы, позволяет эффективно строить траектории ее движения в фазовом пространстве для любых начальных условий и даже оценивать во времени вероятность приближения такой системы к равновесному состоянию [1]. Однако в реальных задачах, особенно при анализе живых систем, часто трудно построить математическую модель, учитывающую все особенности развития исследуемого состояния организма. Поэтому приходится оценивать закономерные и случайные составляющие траектории поведения по экспериментальным данным. В то же время задача обнаружения регулярных и хаотических составляющих, комплексно описывающих анализируемый процесс, особенно актуальна при анализе биологических сигналов. Такой анализ может быть направлен на обнаружение хаоса в самых разнообразных электрофизиологических данных, которые являются результатом взаимодействия большого числа переменных, каждая из которых имеет свой статистический закон распределения. Это могут быть задачи исследования электрической активности головного мозга в различных стадиях сна и при бодрствовании, обнаружение ЭЭГ-патологий, классификация нарушений ритма сердца по ЭКГ [2, 3].

Современная теория хаоса дает нам ряд показателей (размерность аттрактора, показатель Ляпунова, энтропия Колмогорова), которые описывают свойства системы и теоретически позволяют реконструировать уравнения, порождающие анализируемый процесс. Однако на практике применение их к задачам анализа биосигналов ограничивается трудностью аппроксимации этих показателей по выборкам данных конечной длины [1].

Другой путь анализа хаотических процессов основан на теории информации, предложенной Э. Шенноном [4]. Теория информации дает нам количественную меру информации, содержащейся в последовательности символов из данного алфавита, и она может быть применима к любым дискретным последовательностям. По Шеннону информация служит мерой «неожиданности» обнаружения каждой новой буквы в анализируемой последовательности символов. В качестве меры «неожиданности» появления новых последовательностей, несогласованных с предшествующими, может быть использована условная энтропия [5], а изменение ее во времени, т.е. при удлинении цепочки символов, представляющей анализируемый процесс, может служить эффективной оценкой хаотичности поведения системы.

II. ВЫЧИСЛЕНИЕ ОЦЕНОК УСЛОВНОЙ ЭНТРОПИИ

Вычисление условной энтропии предполагает выполнение следующей последовательности шагов. Вначале для заданной выборки отсчетов $\{x(i)\}, i = 1, 2, \dots, N$ формируется ряд L -мерных последовательностей

$$\{X_L(i) = x(i), x(i+1), \dots, x(i+L-1)\},$$

$$i = 1, 2, \dots, (N-L+1).$$

Затем посредством операции квантования и последующего преобразования из числовой в символьную форму это множество может быть представлено в виде множества последовательностей состояний

$$\{Z_L(i)\}, i = 1, 2, \dots, (N-L+1).$$

При этом каждый элемент последовательности $Z_L(i) = z(i), z(i+1), \dots, z(i+L-1)$ кодируется символами из произвольного алфавита $A = \{a_p\}, p = 1, \dots, \xi$.

В результате такого преобразования получаем упорядоченные по i последовательности событий. Они связаны с интервальной оценкой каждого отсчета $x(i)$ в диапазоне (x_{\min}, x_{\max}) и интерпретируются в дальнейшем как соответствующие цепочки символов длины L . Изменяя параметр L , можно формировать последовательности событий разной длины.

Условная энтропия определяется в виде:

$$E(L|L-1) = - \sum_{L-1} p_{L-1} \sum_{L|L-1} p_{L|L-1} \log p_{L|L-1},$$

где $p_{L|L-1}$ задает вероятность появления конкретного символа в последней L -ой ячейке цепочки длины L при условии выделения всех одинаковых цепочек укороченной длины $L-1$.

В информационном плане это выражение в среднем оценивает потерю информации, связанную с появлением L -го символа при условии получения всей подцепочки предшествующих символов. Иными словами величина $E(L|L-1)$ определяет дополнительную информацию, необходимую для предсказания L -го события в любой из последовательностей $\{Z_L(i)\}, i = 1, 2, \dots, (N-L+1)$, если известна вся последовательность предыдущих событий. Условная энтропия стремится к нулю с увеличением L для процессов с выраженной регулярной составляющей и принимает постоянное положительное значение для процессов, генерируемых системами с хаотическими свойствами. Последнее означает снижение уровня прогнозируемости поведения динамической системы при наличии шума в исследуемом процессе.

Величина $E(L|L-1)$ может быть вычислена как приращение, получаемое безусловной энтропией $E(L)$ при переходе от последовательности событий длины $(L-1)$ к длине L :

$$E(L|L-1) = E(L) - E(L-1) = - \sum_L p_L \log p_L + \sum_{L-1} p_{L-1} \log p_{L-1}.$$

Однако при анализе ограниченной выборки отсчетов замена распределения условной вероятности $p_{L|L-1}$ значениями частот попадания $x(i)$ в соответствующие интервалы приводит к значительному смещению оценки $E(L|L-1)$. В результате величина $E(L|L-1)$ с ростом L

будет стремиться к нулю независимо от вида анализируемого процесса.

Преодолеть эту трудность можно путем коррекции оценки условной энтропии:

$$\tilde{E}_1(L|L-1) = E(L|L-1) \left(1 + \frac{N_{L-1}^{(1)}}{(N_L - N_{L-1}^{(1)})} \right), \quad (1)$$

или

$$\tilde{E}_2(L|L-1) = E(L|L-1) + E(1) \cdot \frac{N_{L-1}^{(1)}}{N_L}, \quad (2)$$

где N_L – число анализируемых цепочек символов длины L , $N_{L-1}^{(1)}$ – число лишь однажды встретившихся цепочек символов длины $L-1$.

В выражении (1) «неизвестным» цепочкам придается среднее приращение энтропии, оцениваемое по $(N_L - N_{L-1}^{(1)})$ числу информативных последовательностей событий.

В формуле (2) редким событиям присваиваются веса, равные $E(1)$, т.е. соответствующие энтропии одиночных событий, благодаря чему «неизвестные» цепочки рассматриваются как случайные, а не как регулярные последовательности. Это позволяет избежать ложного представления о не подтвержденных регулярных изменениях в исследуемых процессах, что связано с ограниченным величиной N размером выборки данных.

Введение весов, задаваемых величиной $E(1)$, было обосновано в работе [5], однако придавались они всем одиночным событиям, обнаруженным на L -ом шаге, что приводило к значительному смещению скорректированной энтропии. В наших работах в разряд совершенно случайных событий переводятся лишь те цепочки, которые были представлены лишь одним вектором в ячейках $(L-1)$ – мерного пространства состояний.

Анализ зависимостей рассматриваемых оценок условной энтропии на модельных сигналах позволил предложить следующие способы выбора информативных параметров. Во-первых, это могут быть точечные оценки скорректированной кривой $\tilde{E}_1(L|L-1)$ или $\tilde{E}_2(L|L-1)$ при значениях $L = 2, 3, 4$, т.е. в той области, где вклад одиночных цепочек незначителен. Это позволяет в дальнейшем использовать их в качестве дискриминантных признаков при решении классификационных задач. Во-вторых, в качестве меры нерегулярности процесса может быть использована нижняя граница кривой $\tilde{E}_2(L|L-1)$, которая принимает

малое значение для регулярных процессов и большое значение в случае наличия в сигнале выраженных хаотических изменений. В работе [5] с этой целью обосновано использование величины максимального отклонения скорректированной функции от $E(1)$, поэтому в настоящей работе также введена оценка $ME = E(1) - \min_{L=2, \dots, 10} \{ \tilde{E}_2(L|L-1) \}$.

III. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Рассмотренные оценки были использованы при анализе нарушений ритма сердца, в частности, в задаче распознавания мерцательной аритмии по последовательности кардиоциклов. На рис. 1 приведены примеры ритмограмм для трех классов сердечного ритма: нормального ритма (a), мерцательной аритмии (b) и частой желудочковой экстрасистолии (c).

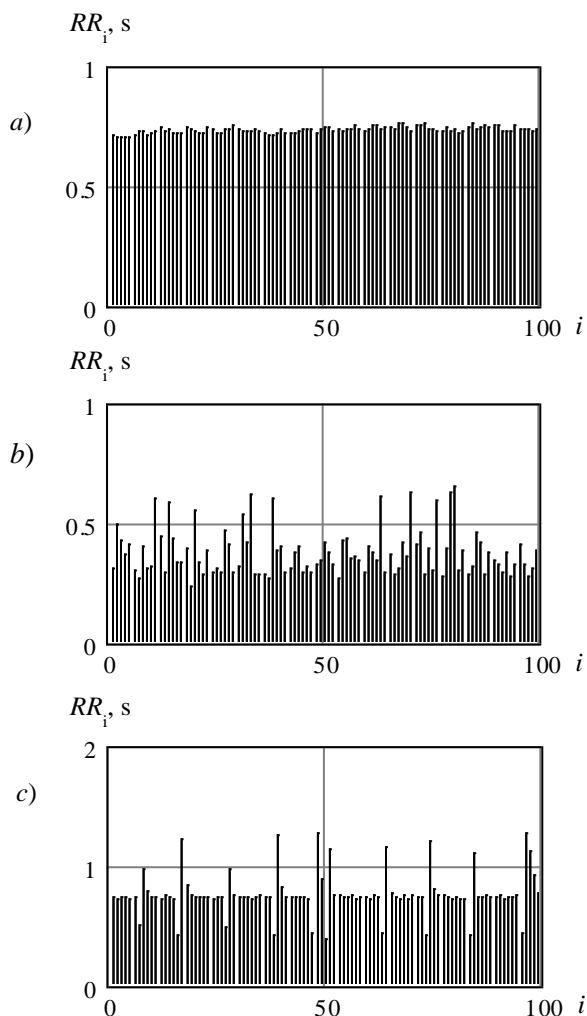


Рис. 1. Примеры ритмограмм для разных видов сердечного ритма: нормального ритма (a), мерцательной аритмии (b) и частой желудочковой экстрасистолии (c)

Полученные для этих реализаций зависимости $E(L|L-1)$, $\tilde{E}_1(L|L-1)$, $\tilde{E}_2(L|L-1)$; $L=1, \dots, 10$ в графическом виде представлены на рис. 2. Длина анализируемых фрагментов составляла $N=300$ отсчетов ритмограммы.

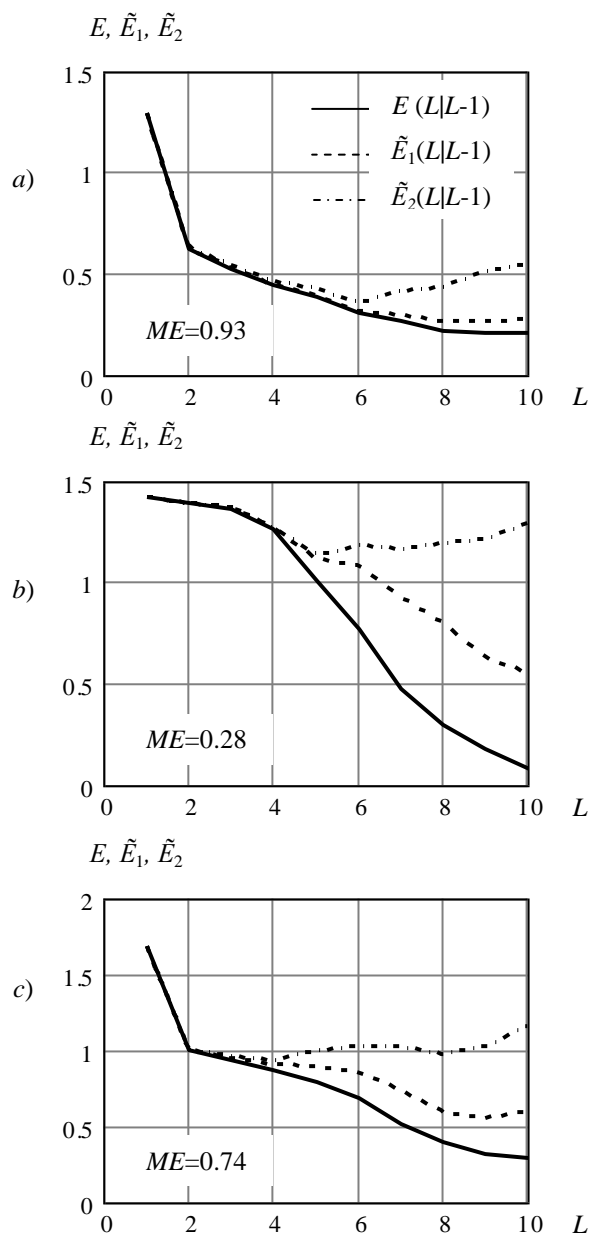


Рис. 2. Зависимости оценок условной энтропии $E(L|L-1)$, $\tilde{E}_1(L|L-1)$, $\tilde{E}_2(L|L-1)$ от длины цепочек L для трех видов ритмограмм: нормального ритма (a), мерцательной аритмии (b) и частой желудочковой экстрасистолии (c)

Как видно из рисунков, нормальный ритм имеющий волновую структуру ритмограммы, характеризуется показателем ME значительно превышающим его значение при мерцательной аритмии ($ME = 0.93$; рис. 2a).

В случае частой экстрасистолии, для которой характерны кратковременные регулярные изменения в последовательности RR-интервалов, наблюдается существенное изменение функции $\tilde{E}_2(L|L-1)$, и значение показателя ME также велико ($ME = 0.74$; рис. 2с). И только при мерцательной аритмии, отличающейся нерегулярным распределением длительностей кардиоциклов, показатель ME значительно меньше ($ME = 0.28$; рис. 2б).

Эксперименты на реальных данных позволили установить, что функция $\tilde{E}_2(L|L-1)$ и показатель ME могут эффективно использоваться для оценки степени регулярности изменений, наблюдаемых во временных рядах конечной длины, а также могут найти применение в задаче распознавания нарушений ритма со сложной динамикой распределения кардиоинтервалов. В процессе исследования было обработано более 150 записей электрокардиосигналов, полученных из электрокардиографической базы MIT-BIH [6], для которых были построены ритмограммы и вычислен относительный минимум уточненной энтропии $\tilde{E}_2(L|L-1)$.

Анализ результатов эксперимента показал, что параметр ME имеет устойчивые интервальные оценки при уровне значимости $\alpha = 0.05$ для разных видов рассматриваемых ритмограмм, причем значение этого параметра при мерцательной аритмии значительно меньше. Так для нормального ритма это значение равно (0.85 ± 0.09), при частой желудочковой экстрасистолии оно несколько меньше (0.74 ± 0.10) и резко уменьшается при появлении приступов мерцательной аритмии (0.17 ± 0.11).

В то же время распознавание хаотических фрагментов ритмограммы на фоне альтернативных групп ритма, отличающихся выраженными регулярными изменениями, целесообразно проводить в пространстве следующего набора признаков: $E(L|L-1)$, $L = 2, 3, 4$; ME .

Построение решающих функций осуществлялось с применением линейного дискриминантного анализа на базе критерия Фишера [7]. С целью минимизации ошибок классификации была введена коррекция разделяющей функции [8, 9]. Она заключалась в нахождении дополнительного вектора, оптимизирующего положение границ разделяющих областей. Результаты экспериментов показали, что переход от анализа параметра ME к

описанию сигналов предложенным набором признаков позволил в 1,5 раза повысить эффективность классификации. При распознавании фрагментов мерцательной аритмии средняя ошибка классификации не превышала значения $\varepsilon = 1\%$.

IV. ЗАКЛЮЧЕНИЕ

Полученные теоретические и экспериментальные результаты позволяют предложить данный подход для идентификации процессов конечной длины с разной выраженностью детерминированных и хаотических компонентов. В частности, показатели $E(L|L-1)$, $L = 2, 3, 4$; ME могут эффективно использоваться при распознавании фрагментов мерцательной аритмии в ходе слежения за ЭКГ больного, что является чрезвычайно важной задачей кардиомониторного наблюдения.

СПИСОК ЛИТЕРАТУРЫ

- [1] Шустер Г. Детерминированный хаос: Введение / Пер. с англ. М.: Мир, 1988. 240 с.
- [2] Рангайян Р.М. Анализ биомедицинских сигналов. Практический подход / Пер. с англ. под ред. А.П. Немирко. М.: ФИЗМАТЛИТ, 2007. 440 с.
- [3] Nonlinear biomedical signal processing / Edited by Metin Akay. Vol. 2, Dynamic analysis and modelling. New York: IEEE. 2001. 341 p.
- [4] Shannon C. E. A mathematical theory of communication In: Bell System Technical Journal. 1948. Vol. 27, № 3, pp. 379–423; Vol. 27, № 4, pp. 623–656.
- [5] Porta A., Baselli G., Liberati D., Montano N., Cogliati C., Gnecchi-Ruscone T., Malliani A., Cerutti S. Measuring regularity by means of a corrected conditional entropy in sympathetic outflow // Biological Cybernetics. 1998, 78, pp. 71–78.
- [6] MIT-BIH Arrhythmia Database. Available from MIT-BIH database distribution // Massachusetts institute of technology. URL: <http://www.physio-net.org/physiobank/database/mitdb>; <http://ecg.mit.edu>.
- [7] Дуда З., Харт П. Распознавание образов и анализ сцен / Пер. с англ. М.: Мир, 1976. 511 с.
- [8] Манило Л.А. Упорядочение спектральных признаков по эмпирическим оценкам межгруппового расстояния в задачах классификации биосигналов // Изв. вузов России. Радиоэлектроника. 2006. Вып.3, с. 20–29.
- [9] Nemirko A.P. Transformation of feature space based on Fisher's linear discriminant. Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. 2016. Vol. 26, No 2. pp. 257–261.