

ГЕНЕТИЧЕСКИЙ АЛГОРИТМ ДИАГНОСТИКИ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ

А. В. Парадинец

Генетические алгоритмы – семейство вычислительных алгоритмов, основанных на эволюционных процессах. Используя процессы, похожие на естественный отбор в природе, эти алгоритмы кодируют потенциальное решение поставленной задачи определенной структурой данных и применяют операции рекомбинирования. Генетические алгоритмы часто рассматриваются как оптимизаторы функций, однако

круг задач, решаемых с помощью генетических алгоритмов, достаточно широк.

В генетическом алгоритме переменные, отвечающие за конкретный параметр, обычно представляются битовыми строками (рис. 1). Это означает, что переменные дискретизируются определенным образом, и область дискретизации представляется некоторой степенью 2. Далее эти параметры объединяются в одну строку – хромосому, которая и является основной структурой данных в генетическом алгоритме.

$x1_0 \dots x1_{k1}$	$x2_0 \dots x2_{k2}$	$x3_0 \dots x3_{k3}$		$xn_0 \dots xn_{kn}$
----------------------	----------------------	----------------------	-------	----------------------

Рис. 1. Вид битовой строки-хромосомы, кодирующей набор параметров некоторой задачи

Применение генетического алгоритма начинается с образования популяции хромосом (как правило, случайных). Далее происходит оценка этих структур и распределение возможности «воспроизводства» в соответствии с приспособленностью каждой хромосомы. Таким образом, каждая хромосома получает некоторую величину в зависимости от приспособленности решения, представляемую ей, являющуюся мерой шанса воспроизводства в следующем поколении.

Выполнение генетического алгоритма рассматривается как процесс, состоящий из двух стадий. Процесс начинается с текущей популяции. К текущей популяции применяется отбор (на основании значений фитнес-функции), после чего формируется промежуточная популяция.

Существует несколько методов отбора. В одном из них можно рассматривать популяцию, каждой особи которой сопоставлен сектор «рулетки» с площадью, пропорциональной значению фитнес-функции (далее – фитнес-значение). Повторяющимся раскручиванием рулетки производится отбор особей в промежуточное поколение. После того как был произведен отбор и создание промежуточного поколения закончено, применяется рекомбинирование (скрещивание). Этот процесс может быть рассмотрен как создание следующего поколения из промежуточного. Кроссовер (скрещивание) применяется к случайно выбранным парам с вероятностью p_c . (Популяция уже должна быть достаточно хорошо перемешана после процесса отбора). Выбирается пара из двух строк. С вероятностью p_c эти строки рекомбинируются, и полученные две новые строки попадают в последующее поколение.

Теперь перейдем к обработке данных электрокардиограмм. Существуют несколько методов анализа медицинских кривых. В первом ис-

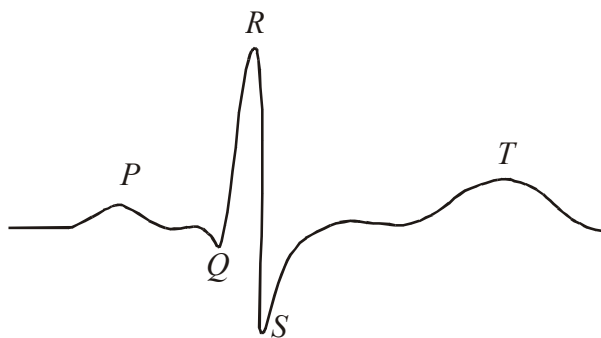


Рис. 2. Вид периода кардиограммы

следователи пошли по пути использования медицинских признаков, которые были выработаны на основе многолетнего опыта и врачебной интуиции. При этом на кривой отыскиваются те или иные характерные особенности (зубцы, изгибы, провалы) и производятся их замеры. Численные результаты этих замеров и некоторых со-

отношений между ними, а также описания их особенностей рассматриваются как признаки соответствующей медицинской кривой. Так, на ЭКГ, типичный вид которой представлен на рис. 2, выделяются зубцы P , Q , R , S и T , затем описываются особенности их вида, замеряются некоторые характерные размеры и т. д.

В основе второго направления лежит идея разложения исходных данных по той или иной системе базисных функций. Речь идет о приближенном представлении непрерывного процесса $f(t)$ в виде некоторой линейной комбинации

$$f(t) = \sum_{i=1}^m c_i \varphi_i(t) \quad (1)$$

базисных функций $\varphi_1(t)$, $\varphi_2(t)$, ... Коэффициенты c_1 , c_2 , ..., c_m при этом рассматриваются как код непрерывной кривой $f(t)$, позволяющий приближенно восстановить кодируемую временную зависимость $f(t)$. Представляя координаты разложения в n -мерном пространстве, в большинстве случаев можно построить плоскость, ограничивающую различные множества кардиограмм с определенными заболеваниями.

Можно анализировать кривые электрокардиограмм по принципу «похожести». Пусть имеется выделенный комплекс PQRS (рис. 3). Различные заболевания видоизменяют форму этого комплекса (отсутствуют некоторые зубцы, изменяются отношения расстояний и так далее). Необходимо составить большую базу таких комплексов для различных заболеваний. С использованием этой базы и исходной кардиограммы можно найти такой экземпляр комплекса PQRS в базе, который будет наилучшим образом похож на исходный. Для этого требуется найти коэффициенты подобия k_x , k_y и координаты смещения d_x , d_y .

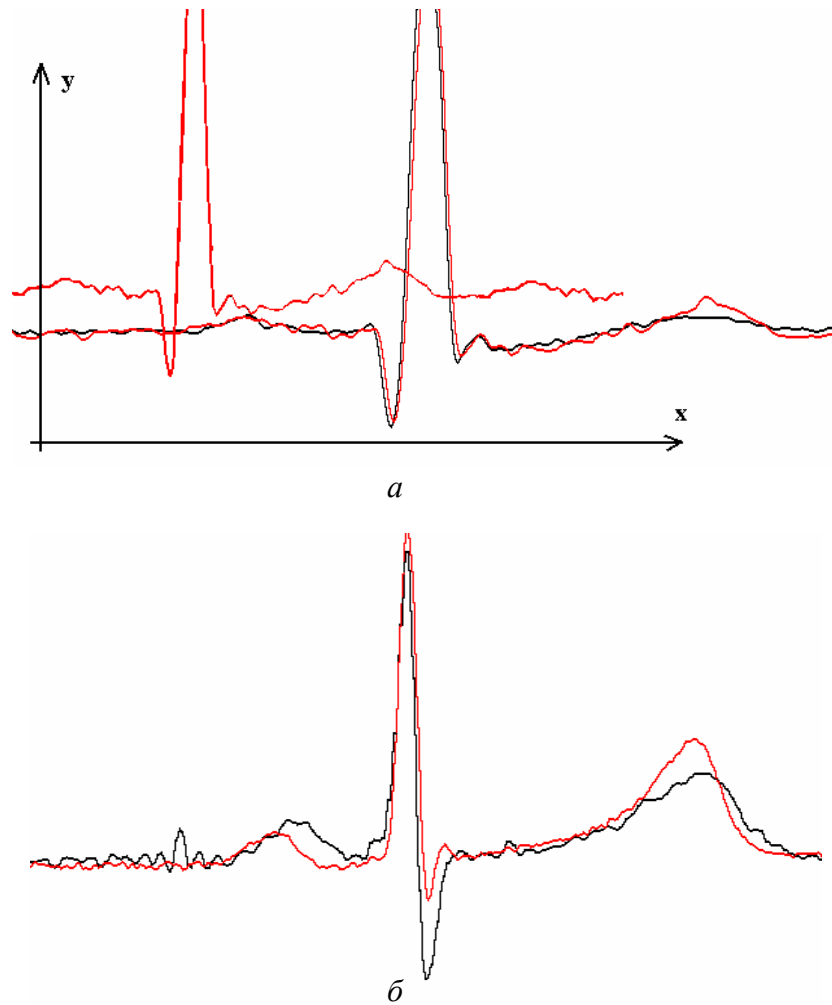


Рис. 3

a – сопоставление двух фрагментов кривой ЭКГ;
б – результат работы генетического алгоритма – найденный эталон к исходному комплексу QRS реальной кардиограммы

Генетический алгоритм может быть использован для быстрого поиска этих коэффициентов. Таким образом, имеется задача оптимизации, трудно решаемая градиентными методами (так как на рис. 3*a* можно видеть, что смещение кривой вдоль координаты x не несет никакой градиентной информации, когда заметно увеличение значения целевой функции при уменьшении коэффициента k_y). Целевая функция будет выглядеть как сумма квадратов разностей между точками исходного образца и точками эталонного образца из базы данных, для которого введены коэффициенты пропорциональности по координатам и координаты начального смещения

$$g(k_x, k_y, d_x, d_y) = \sum_{x=0}^{len} \left(F(x) - k_y f(k_x x + d_x) - d_y \right)^2. \quad (2)$$

Результатом работы генетического алгоритма будут являться 4 параметра k_x, k_y, d_x, d_y , при которых эталон наилучшим образом подходит к исходному образцу. Теперь, находя наилучший с точки зрения похожести эталон и принимая минимальный коэффициент похожести (целевая функция), для того чтобы отбросить варианты заболеваний, не существующих в базе данных, можно поставить предварительный диагноз.

На рис. 3б можно видеть результаты работы алгоритма, нашедшего оптимально похожий эталон из 80 возможных для исходной кривой.

Однако на практике, диагностика сердечно-сосудистых заболеваний не является такой простой задачей. Для диагностики используется семейство кардиографических кривых – отведений. Ряд заболеваний может отражаться только в некоторых из них. Некоторые заболевания могут влечь за собой незначительные изменения формы ЭКГ. При этом необходимо искать похожие записи уже не в одной кривой, а параллельно в нескольких. Дополнительно часто используется и другая физиологическая информация. Также данный способ не позволяет диагностировать все изменения ритма (аритмии).

Литература

1. *Darrell Whitley* A Genetic Algorithm Tutorial / Computer Science Department, Colorado State University. 1991.
2. *Неймарк Ю. И.* Распознавание образов и медицинская диагностика /. М.: Наука, 1972.
3. *Зюзенков М. В.* Основы практической электрокардиографии / Мн.: Высшая школа, 1992 г.