УДК 519.216

И.П. Атаманюк, канд. техн. наук, доц.,
Николаев. нац. аграр. ун-т,
Т.В. Олейникова, врач-кардиолог, БРИТ ОКЦ,
г. Николаев

АВТОМАТИЗАЦИЯ ПРОЦЕССА ДИАГНОСТИКИ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ НА ОСНОВЕ АППАРАТА КАНОНИЧЕСКИХ РАЗЛОЖЕНИЙ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

І.П. Атаманюк, Т.В. Олейникова. **Автоматизація процесу діагностики серцево-судинних захворювань на основі апарату канонічних розкладів випадкових послідовностей.** Метою роботи є розробка інформаційної технології для автоматизації аналізу результатів кардіографії. В основу алгоритму розпізнавання покладено канонічний розклад випадкової послідовності, що описує зміну кардіограми. Отримане рішення задачі класифікації кардіограм по критерію максимуму щільності розподілу суттєво простіше відомого вирішального правила за рахунок переходу від багатовимірної щільності розподілу до добутку одномірних щільностей. Результати обчислювального експерименту вказують на високу достовірність ідентифікації серцево-судинних захворювань на основі запропонованої інформаційної технології.

Ключові слова: випадкова послідовність, канонічний розклад, алгоритм розпізнавання.

И.П. Атаманюк, Т.В. Олейникова. **Автоматизация процесса диагностики сердечно-сосудистых заболеваний на основе аппарата канонических разложений случайных последовательностей.** Целью работы является разработка информационной технологии для автоматизации анализа результатов кардиографии. В основу алгоритма распознавания положено каноническое разложение случайной последовательности, описывающей изменение кардиограмм. Полученное решение задачи классификации кардиограмм по критерию максимума плотности распределения существенно проще известного решающего правила за счет перехода от многомерной плотности распределения к произведению одномерных плотностей. Результаты численного эксперимента указывают на высокую достоверность идентификации сердечно-сосудистых заболеваний на основе предложенной информационной технологии.

Ключевые слова: случайная последовательность, каноническое разложение, алгоритм распознавания.

I.P. Atamanyuk, T.V. Oleynikova. **Automation of process of diagnostics of cardiovascular diseases on the basis of vehicle of canonical decompositions of casual sequences**. The aim of work is an information technology for automation of analysis of results of cardiography. In basis of algorithm of recognition canonical decomposition is canonical decomposition of casual sequence, describing the change of cardiograms. Got decision of task of classification of cardiograms on the criterion of a maximum of closeness of distribution substantially simpler of the known decision rule due to a transition from the multidimensional closeness of distribution to increase of unidimensional closenesses. The results of numeral experiment are specified on high authenticity by authentications of cardiovascular diseases on the basis of the offered information technology.

Keywords: casual sequence, canonical decomposition, algorithm of recognition.

В настоящее время сердечно-сосудистые заболевания прочно занимают первое место среди наиболее распространенных и опасных болезней современности. По данным Всемирной организации здравоохранения показатель смертности из-за сердечных болезней в Украине достигает 64 %, в США ежегодно болезнь сердца поражает более 800 тыс. людей. В настоящее время резко увеличилось число сердечных заболеваний у трудоспособного населения (нередко

ISSN 2076-2429 (print)
ISSN 2223-3814 (on line)

возраст больного инфарктом миокарда не превышает 23…25 лет).

Поскольку заболевания сердца принадлежат к болезням, протекание и результаты лечения которых, непосредственно зависят от своевременного выявления и устранения патологических отклонений, достоверная диагностика является важнейшей и первоочередной задачей в проблеме сердечно-сосудистых заболеваний. Автоматизация процессов диагностики является одним из основных направлений в кардиологии для повышения надежности и достоверности идентификации заболевания. На сегодняшний день разработано большое количество подходов к решению данной задачи, однако каждый из них имеет свои недостатки и ограничения, поэтому необходимость разработки эффективных методов и информационных технологий медицинской диагностики не утратила своей актуальности.

Одним из самых распространенных методов диагностики и распознавания сердечно-сосудистых заболеваний является электрокардиография (ЭКГ) — метод графической регистрации характеристик электрического поля сердца и их изменений в процессе сердечных сокращений. Сигнал ЭКГ характеризуется набором зубцов, по временным и амплитудным параметрам которых ставится диагноз. Учитывая, что изменение параметров ЭКГ носит случайный характер, математическим содержанием диагностики сердечных заболеваний является задача классификации реализации случайной последовательности (каждому классу соответствует некоторое заболевание или отсутствие болезни). С целью повышения достоверности диагностики сердечно-сосудистых заболеваний необходимо разработать на основе теории случайных последовательностей информационную технологию распознавания электрокардиограмм с полным учетом их стохастических свойств.

Объектом исследования является случайная последовательность $\{X\}=\{X(1),X(2),...,X(12)\}$ с двенадцатью элементами, каждый из которых соответствует некоторому наиболее информативному параметру электрокардиограммы: $X(1)$ — ширина зубца P; $X(2)$ — высота зубца P; $X(3)$ — интервал P-Q ; $X(4)$ — высота зубца Q; $X(5)$ — интервал QRS; $X(6)$ — высота первого зубца R; $X(7)$ — высота второго зубца R; $X(8)$ — высота зубца S; $X(9)$ — интервал Q-T; $X(10)$ — высота зубца T; $X(11)$ — длительность первого цикла кардиограммы; $X(12)$ — длительность второго цикла кардиограммы. В результате проведения ЭКГ получена некоторая последовательность значений $x(i),\ i=\overline{1,12}$, о которой априорно известно, что она порождена одной из случайных последовательностей $X^{(j)}(i),\ i=\overline{1,12},\ j=\overline{1,J}$ ($J-1$ заболеваний и нормальное состояние). Требуется определить к какой именно из этих последовательностей (к какому из J классов) относится данная реализация. Сформулированная таким образом задача распознавания полностью сводится к стандартной байесовской постановке, однако при использовании критерия Байеса могут быть не распознаны маловероятные (и поэтому особенно опасные) заболевания. В этой связи для решения задачи медицинской диагностики наиболее приемлемым является критерий максимума правдоподобия, согласно которому при наблюдении реализации $\vec{x}=\{x(1),x(2),...,x(12)\}$ принимается та гипотеза, которая удовлетворяет условию:

$$j^{*}=\arg\max_{j}\left\{f_{12}\left(\vec{x}/j\right)\right\}, \tag{1}$$

где $f_{12}\left(\vec{x}/j\right),\ j=\overline{1,J}$ — условная плотность распределения признаков $\vec{x}$ при условии, что реализация принадлежит данному классу.

Задача распознавания реализации случайной последовательности сводится к определению принадлежности реализации $\vec{x}$ к одному из J заданных распределений $f_{12}\left(\vec{x}/j\right),\ j=\overline{1,J}$.

Таким образом, следующим этапом является оценка неизвестных плотностей $f_{12}(\vec{x}/j),\ j=\overline{1,J}$, что в свою очередь, учитывая большое количество результатов наблюдения $x(i),\ i=\overline{1,12}$, является достаточно сложной и трудоемкой процедурой. Данная задача в рамках линейных связей существенно упрощается [1] при переходе от последовательности $x(i),\ i=\overline{1,12}$ к анализу набора некоррелированных значений $v_i,\ i=\overline{1,I}$, которые определяются из канонической модели представления случайной последовательности [2]:

$$X(i)=\sum_{\nu=1}^{i}V_\nu\varphi_\nu(i),\ i=\overline{1,12}, \tag{2}$$

$$V_i=X(i)-\sum_{\nu=1}^{i-1}V_\nu\varphi_\nu(i),\ i=\overline{1,12}, \tag{3}$$

$$\varphi_\nu(i)=\frac{1}{D_\nu}\left\{M\left[X(\nu)X(i)\right]-\sum_{j=1}^{\nu-1}D_j\varphi_j(\nu)\varphi_j(i)\right\},\ \nu=\overline{1,I},\ i=\overline{\nu,I}. \tag{4}$$

$$D_i=M\left[X^2(i)\right]-\sum_{\nu=1}^{i-1}D_\nu\varphi_\nu^2(i),\ i=\overline{1,12}; \tag{5}$$

где $\varphi_\nu(i),\ \nu,i=\overline{1,I}$ — неслучайная координатная функция: $\varphi_\nu(\nu)=1,\ \varphi_\nu(i)=0$, если $\nu>i$.

В этом случае замена $\vec{x}$ на вектор $\vec{v}$ с учетом $f_I(\vec{v}/j)=\prod_{i=1}^{12}f_1(v_i/j),\ j=\overline{1,J}$ позволяет записать решающее правило в следующем виде

$$j^*=\arg\max_j\left\{\prod_{i=1}^{12}f_1(v_i/j),\ j=\overline{1,J}\right\}. \tag{6}$$

Задача распознавания, таким образом, сводится к последовательной аппроксимации двенадцати одномерных плотностей распределения. Стохастический алгоритм диагностики существенно упрощается, однако переход от вектора $\vec{x}$ к вектору $\vec{v}$ возможен при условии, что случайные последовательности $\{X(i)/j\},\ i=\overline{1,12},\ j=\overline{1,J}$ обладают только линейными связями. Снятие ограничения о нормальном распределении случайных последовательностей $X^{(j)}(i),\ i=\overline{1,12},\ j=\overline{1,J}$ возможно в результате использования соответствующего нелинейного канонического разложения [3…5]:

$$V_i^{(\lambda)}=X^\lambda(i)-\sum_{\nu=1}^{i-1}\sum_{j=1}^{N}V_\nu^{(j)}\beta_{\lambda\nu}^{(j)}(i)-\sum_{j=1}^{\lambda-1}V_i^{(j)}\beta_{\lambda i}^{(j)}(i),\ i=\overline{1,12}; \tag{7}$$

$$D_\lambda(i)=M\left[X^{2\lambda}(i)\right]-\sum_{\mu=1}^{i-1}\sum_{j=1}^{N}D_j(\mu)\left\{\beta_{\lambda\mu}^{(j)}(i)\right\}^2-\sum_{j=1}^{\lambda-1}D_j(i)\left\{\beta_{\lambda i}^{(j)}(i)\right\}^2,\ i=\overline{1,12} \tag{8}$$

$$\beta_{h\nu}^{(\lambda)}(i)=\frac{1}{D_\lambda(\nu)}\left(M\left[X^\lambda(\nu)X^h(i)\right]-\sum_{\mu=1}^{\nu-1}\sum_{j=1}^{N}D_j(\mu)\beta_{\lambda\mu}^{(j)}(\nu)\beta_{h\mu}^{(j)}(i)-\right.$$
$$\left.-\sum_{j=1}^{\lambda-1}D_j(\nu)\beta_{\lambda\nu}^{(j)}(\nu)\beta_{h\nu}^{(j)}(i)\right),\ \lambda=\overline{1,N},\ \nu=\overline{1,i}. \tag{9}$$

ISSN 2076-2429 (print)
ISSN 2223-3814 (on line)

Учитывая различные свойства $\left\{X(i)/j\right\},\ i=\overline{1,12},\ j=\overline{1,J}$ параметры канонического разложения (7)…(9) являются уникальными для исследуемых последовательностей. Преимущество использования разложения (7)…(9) заключается в том, что из некоррелированности $V_i^{(N)},\ i=\overline{1,I}$ следует их независимость, так как все стохастические связи более низких порядков из данных коэффициентов удалены. Таким образом, также как и в предыдущем случае, перевод задачи распознавания из 12-мерного пространства признаков $\left\{X(1),...,X(12)\right\}$ в пространство признаков $\left\{V_1^{(N)},...,V_{12}^{(N)}\right\}$ такой же размерности упрощает процедуру оценки плотностей распределения $f\left(v_1^{(N)},...,v_{12}^{(N)}/j\right)=\prod_{i=1}^{12} f\left(v_i^{(N)}/j\right),\ j=\overline{1,J}$, которая сводится к аппроксимации двенадцати одномерных плотностей распределения. Решающее правило принимает вид

$$j^*=\arg\max_j\{\prod_{i=1}^{12} f_1(v_i^{(N)}/j),\ j=\overline{1,J}\}. \tag{10}$$

Отсутствие предположений о виде плотности распределения случайных величин $\left\{V_1^{(N)},...,V_{12}^{(N)}\right\}$ приводит к необходимости использования для их описания непараметрических методов. Наиболее простым и эффективным подходом в данных условиях является использование непараметрических оценок парзеновского типа [7]:

$$f_L\left(v_i^{(N)}\right)=\frac{1}{dL}\sum_{l=1}^{L} g\left(u_l\right), \tag{11}$$

где $u_l=d^{-1}\left(v_i^{(N)}-v_{i,l}^{(N)}\right)$;

$v_{i,l}^{(N)},\ l=\overline{1,L}$ — реализации случайной величины $V_i^{(N)}$;

$g\left(u_l\right)$ — некоторая весовая функция (ядро);

d — константа (коэффициент размытости).

Выбор в качестве функции ядра $g(u)$ равномерной плотности распределения позволяет записать выражение для оценки плотности распределения $V_i^{(N)}$ в следующем виде:

$$f_L\left(v_i^{(N)}\right)=\frac{1}{dL}\sum_{l=1}^{L} g_l\left(v_i^{(N)}\right),$$

где

$$g_l\left(v_i^{(N)}\right)=\begin{cases}0,5, & v_{i,l}^{(N)}-d\le v_i^{(N)}\le v_{i,l}^{(N)}+d,\\ 0, & \left|v_i^{(N)}-v_{i,l}^{(N)}\right|>d,\end{cases}\quad l=\overline{1,L};$$

$$d=0,5\sup_l\left|v_{i,l}^{(N)}-v_{i,l-1}^{(N)}\right|,\ v_{i,l}^{(N)}>v_{i,l-1}^{(N)},\ l=\overline{2,L}.$$

Схема функционирования системы диагностики сердечно-сосудистых заболеваний, на основе разработанной информационной технологии, представлена на рис. 1.

Алгоритм апробирован на трех классах случайных последовательностей: $\left\{X(i)/1\right\},\ i=\overline{1,12}$ — “здоровое сердце”, $\left\{X(i)/2\right\},\ i=\overline{1,12}$ — “гипертрофия миокарда”, $\left\{X(i)/3\right\},\ i=\overline{1,12}$ — “тяжелая аритмия”. Проверка статистической гипотезы о независимости случайных коэффициентов канонического разложения (7) на основе критерия χ^2 показала справедливость гипотезы при $N=3$ для всех трех последовательностей с вероятностью не меньше $P_Д = 0{,}98$. Таким образом, разложение (7) с соответствующим набором координатных функций $\beta_{h\nu}^{(\lambda)}(i)$, $h,\lambda = \overline{1,3}$, $\nu, i = \overline{1,12}$ модифицируется в адекватную модель исследуемой случайной последовательности $\{X(i)/j\}$, $i = \overline{1,12}$, $j = \overline{1,3}$. Например, в таблице представлены значения $\beta_{1\nu}^{(1)}(i)$, $\nu, i = \overline{1,12}$ для $\{X(i)/3\}$, $i = \overline{1,12}$. Распознавание выполнялось для 150 различных кардиограмм, относящихся к трем исследуемым случайным последовательностям. В результате эксперимента все кардиограммы были правильно классифицированы.

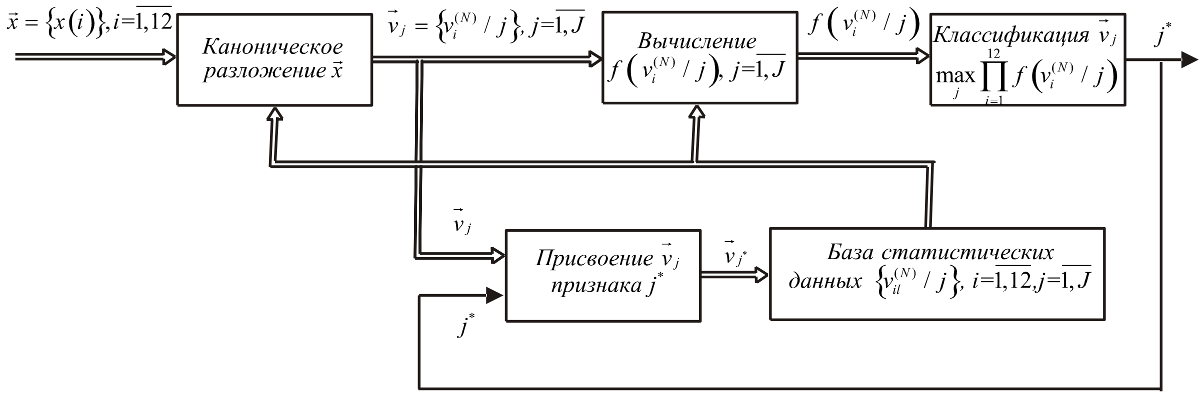

Рис. 1. Схема функционирования системы диагностики сердечно-сосудистых заболеваний

Значения координатной функции $\beta_{1\nu}^{(1)}(i)$ *для случайной последовательности* $\{X(i)/3\}$, $i = \overline{1,12}$

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0,14	1,46	0,12	0,92	1,49	0,06	0,22	3,06	-0,36	7,11	5,66
2	0	1	6,50	0,34	3,81	5,70	0,37	1,56	12,59	-2,41	28,72	22,10
3	0	0	1	0,08	0,63	1,01	0,04	0,15	2,13	-0,24	4,91	3,92
4	0	0	0	1	4,22	9,07	0,72	1,12	14,16	-0,18	33,40	25,39
5	0	0	0	0	1	1,22	0,20	0,29	3,02	-0,25	6,42	5,46
6	0	0	0	0	0	1	0,08	0,15	1,61	-0,14	3,79	2,79
7	0	0	0	0	0	0	1	0,24	2,70	-0,53	6,16	5,05
8	0	0	0	0	0	0	0	1	5,69	-0,50	11,17	9,52
9	0	0	0	0	0	0	0	0	1	-0,07	2,12	1,82
10	0	0	0	0	0	0	0	0	0	1	-6,08	-4,15
11	0	0	0	0	0	0	0	0	0	0	1	0,83
12	0	0	0	0	0	0	0	0	0	0	0	1

Таким образом, в работе предложена автоматизированная система диагностики сердечно-сосудистых заболеваний. В основу системы положено каноническое разложение случайной последовательности изменения кардиограмм, что позволило сформировать решающее правило максимума совместной плотности распределения в виде произведения одномерных плотностей распределения параметров кардиограммы. Каноническое разложение не накладывает никаких существенных ограничений (линейность, стационарность, марковость и т.д.) на класс исследуемых случайных последовательностей, что дает возможность максимально учесть стохастические характеристики последовательностей, относящихся к различным сердечно-сосудистым заболеваниям. Результаты численного эксперимента указывают на высокую

ISSN 2076-2429 (print)
ISSN 2223-3814 (on line)

достоверность диагностики сердечно-сосудистых заболеваний на основе предложенной информационной технологии.

Литература

1. Кудрицкий, В.Д. Фильтрация, экстраполяция и распознавание реализаций случайных функций / В.Д. Кудрицкий. — К.: ФАДА, ЛТД, 2001. — 176 с.
2. Пугачев, В.С. Теория случайных функций и ее применение / В.С. Пугачев. — М.: Физматгиз, 1962. — 720 с.
3. Atamanyuk, I.P. The algorithm of optimal polynomial extrapolation of random processes / I.P. Atamanyuk, V.Y. Kondratenko, O.V. Kozlov, Y.P. Kondratenko // Lecture Notes in Business Information Processing, 2012. — pp. 78 — 87.
4. Атаманюк, И.П. Алгоритм оптимальной нелинейной экстраполяции реализации случайного процесса с фильтрацией погрешностей измерений / И.П. Атаманюк, Ю.П. Кондратенко // Электронное моделирование. — 2012. — Т. 34, № 4. — С. 23 — 40.
5. Атаманюк, И.П., Кондратенко Ю.П. Синтез оптимальных линейных стохастических систем управления на базе аппарата канонических разложений случайных последовательностей / И.П. Атаманюк, Ю.П. Кондратенко // Управляющие системы и машины. — 2012. — №1. — С. 8 — 12.

References

1. Kudritskiy, V.D. Fil'tracija, ekstrapoljacija i raspoznavanie realizacij sluchajnyh funkcij [Filtering, extrapolation and recognition of the random function] / V.D. Kudritskiy. — Kiev, 2001. — 176 p.
2. Pugachev V.S. Teoriya sluchainikh funktsiy i eio primenenie [The theory of random functions and its application] / V.S. Pugachev. — Moscow, 1962. — 720 p.
3. Atamanyuk, I.P. The algorithm of optimal polynomial extrapolation of random processes / I.P. Atamanyuk, V.Y. Kondratenko, O.V. Kozlov, Y.P. Kondratenko // Lecture Notes in Business Information Processing, 2012. — pp. 78 — 87.
4. Atamanyuk, I.P. Algoritm optimalnoy nelineynoy extrapolatsii realizaciy sluchainogo procesa s philtraciyey pogreshnostey izmereniy [Optimal nonlinear extrapolation algorithm realization of a random process with filtering of measurement error] / I.P. Atamanyuk, Y.P. Kondratenko // Elektronnoye modelirovaniye [Electronic simulation]. — 2012. — # 4. — Vol. 34, — pp. 23 — 40.
5. Atamanyuk, I.P. Sintez optimalnih lineynih stohasticheskih sistem upravleniya na baze apparata kanonicheskih razlozheniy sluchainikh posledovalnostey [Synthesis of linear stochastic optimal control systems based on the apparatus of canonical expansions of random sequences] / I.P. Atamanyuk, Y.P. Kondratenko // Upravljajushhie sistemy i mashiny [Control systems and machines]. — 2012. — # 1. — pp. 8 — 12.

Рецензент д-р техн. наук, проф. Одес. нац. политехн. ун-та Дрозд А.В.

Поступила в редакцию 18 декабря 2012 г.