Секция
«Математическая теория интеллектуальных систем»

КЛАССИФИКАЦИЯ ПСИХИЧЕСКИХ ЗАБОЛЕВАНИЙ НА ОСНОВЕ ДЕНДРОГРАММ ЭЭГ ГОЛОВНОГО МОЗГА И ИХ ХАРАКТЕРИСТИК

В. С. Анашин (Москва), Л. Б. Тяпаев, В. В. Давыдов (Саратов)

Проблема постановки корректного диагноза заболевания пациента всегда является актуальной. В данной работе рассматривается задача классификации пациентов по типу психического заболевания на основе анализа дендрограмм ЭЭГ головного мозга. В работе [1] предложен метод анализа дендрограмм, построенных по ЭЭГ для пациентов из нескольких групп, предварительно разделенных на кластеры по вынесенному врачебному вердикту. Авторы работы [1] вычисляют p-адический квантовый потенциал — параметр, по которому удаётся разделить пациентов на группы по ментальным заболеваниям: болезнь альцгеймера, шизофрения, депрессия и умеренное когнитивное расстройство (УКР). В данном исследовании было решено пойти иным путём, а именно, провести анализ дендрограмм, на которые можно смотреть как на префиксные коды.

В данной работе используются реальные данные, которые были использованы в статье [1]. Они представляют собой набор файлов формата hdf5 с записями ЭЭГ пациентов с различными психическими расстройствами и одной контрольной группой здоровых людей:

№	Файл	Пациенты	Заболевание
1	alz_c1_new.mat	43	болезнь альцгеймера
2	dep_c1_new.mat	28	депрессия
3	mci_c1_new.mat	28	УКР
4	schiz_c1_new.mat	42	шизофрения
5	controls_c1_new.mat	96	контрольная группа

Всего имеется 237 пациентов, представленных в обезличенной форме. Основными данными являются записи ЭЭГ и частота их дискретизации. ЭЭГ представлена в виде двумерного массива — (единичные считывания сигнала $\times$ 19 значений электрических потенциалов электродов в *мВ*). Есть и иные данные, например, пол и возраст.

Перечислим шаги, необходимые для построения дендрограммы.

1. Извлечение данных ЭЭГ из файлов формата HDF5.
2. Применение двух фильтров к ЭЭГ. Первый, для исключения из него помех частотой 50 Герц, вызванных колебаниями в электрической сети. Второй, для исключения частот ниже 1 Герц. В отличие от статьи [1] в данной работе используется вся запись;
3. Установка величины временного окна — одна секунда. Переменная времени t лежит в интервале $[1, n/W]$, где $n = \lfloor \frac{\text{число единичных считываний сигнала}}{\text{частота дискретизации}} \rfloor$, а W — размер временного окна. Единичное считывание это вектор из 19 значений, по 1 на электрод.
4. Нормализация значений электрических потенциалов электродов внутри каждого окна. Для расчета применяется формула $\widehat{ep}_{elec,t} = \frac{|ep_{elec,t}|}{max(|ep_{elec,t}|)}$, $\widehat{ep}_{elec,t} \in [0, 1]$, где $\widehat{ep}_{elec,t}$ — нормализованная во временном окне запись ЭЭГ для соответствующего электрода, $|ep_{elec,t}|$ — абсолютные значения записи ЭЭГ во временном окне.
5. свёртка ЭЭГ данных во временном окне с шагом в 0.01. Каждую секунду записи сворачиваем в вектор из 100 элементов, учитывая частоту дискретизации, получаем новый набор данных $h_{elec,t}$.

6.1. вычисление расстояния Хеллингера попарно, по формуле $H(x, y) = \frac{1}{\sqrt{2}} \left(\sqrt{\sum_{i=1}^{k} (\sqrt{x_i} - \sqrt{y_i})^2} \right)$, где $k = 100$, $x = h_{elec,t}$ — вектор из k элементов для электрода в единицу времени t, $y = h_{elec,t}$ — аналогично, но для другого электрода. В данной работе расстояния вычисляются для всего вектора, для всех t и k.

6.2. Построение дендрограммы на основе расстояний (рис. 1).

7. Сопоставление каждому электроду бесконечной последовательности из 0/1 (целого 2-адического числа [2]), префиксом которого является бинарный код ветви.

Дендрограмма представляет собой бинарное кодовое дерево, а значит, каждому электроду, его висячей вершине, можно сопоставить двоичный код (рис. 1, справа). Кодовое дерево формируемое дендрограммой является насыщенным, а значит для него выполняется равенство Макмиллана $\sum_{i=1}^{r} 2^{-l_i} = 1$ [3], где r — число кодовых слов, l_i — длина кодового слова с индексом i. Данное свойство позволяет рассматривать бинарные коды электродов, как полную группу

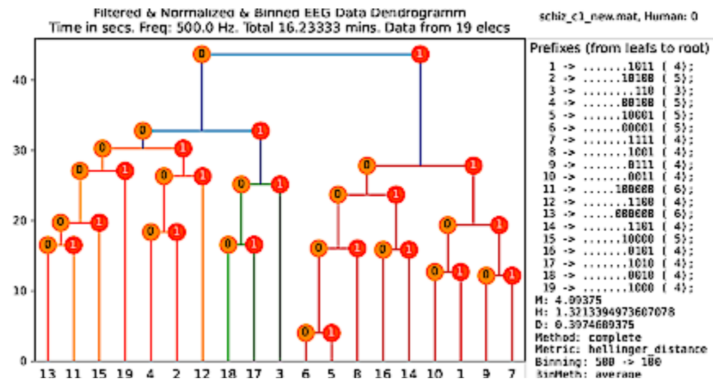

Рис. 1: Дендрограмма для человека с шизофренией

событий длин кодов. Следовательно, можно посчитать математическое ожидание, энтропию и дисперсию кодов. Средние, минимальные и максимальные значения упомянутых параметров приведены ниже.

Файл	Параметр	Мат. ожидание	Энтропия	Дисперсия
alz	Среднее	3.5741733	1.8168757	1.6849957
	Максимум	4.0937500	2.2526204	4.0898438
	Минимум	2.4453125	1.1292237	0.3974609
controls	Среднее	3.5953878	1.7979498	1.6737500
	Максимум	4.1875000	2.2995481	4.3085938
	Минимум	2.3896484	0.6962123	0.1523438
dep	Среднее	3.6573661	1.7742883	1.4166489
	Максимум	4.1875000	2.2995481	3.9940796
	Минимум	2.8359375	0.6962123	0.1523438
mci	Среднее	3.6290458	1.8588847	1.4953717
	Максимум	4.0937500	2.2653195	4.0082397
	Минимум	2.4140625	1.1862781	0.3974609
shiz	Среднее	3.6320219	1.8641116	1.5150029
	Максимум	4.0937500	2.2653195	4.1875000
	Минимум	2.7968750	1.2741306	0.3974609

Анализ вычисленных параметров префиксных кодов выявил следующую особенность: средние значения математического ожидания, энтропии и дисперсии для группы дендрограмм пациентов, принадлежащих одному ментальному кластеру, являются уникальными.

Стоит отметить, что предложенный метод анализа ЭЭГ может быть использован лишь как вспомогательное средство, указывающее возможный тип заболевания. Верный диагноз, посредством дополнительных исследований, может поставить только врач-специалист.

Список литературы

1. Shor O., Glik A., Yaniv-Rosenfeld A., Valevski A., Weizman A., Khrennikov A., et al. EEG p-adic quantum potential accurately identifies depression, schizophrenia and cognitive decline // PLoS ONE. — 2021. — 16(8): e0255529.

2. Anashin V., Khrennikov A. Applied Algebraic Dynamics — Berlin.: Walter de Gruyter, 2009.

3. Иорданский М. А., Смышляева О. В. Кодирование комбинаторных объектов: учебное пособие. — Н.Новгород: Мининский университет, 2017.

DOI: 10.20948/dms-2022-64

ЗАДАЧА К-КОНЕЧНОПОРОЖДЕННОСТИ ПРЕДПОЛНЫХ КЛАССОВ ЛИНЕЙНЫХ АВТОМАТОВ

В. А. Бирюкова (Москва)

В данной работе будет рассмотрена проблема K-конечнопорожденности предполных классов линейных автоматов, функционирующих над конечным полем, состоящим из двух элементов.

Пусть $V = (A, Q, B, \varphi, \psi, q_0)$ — инициальный абстрактный конечный автомат [1]. V — линейный автомат, если $n, s \in \mathbb{N}$, $A = E_k^n$, $Q = E_k^s$, $B = E_k$, где E_k — поле Галуа, состоящее из k элементов, φ и ψ — линейные операторы k-значной логики. Множество линейных автоматов над полем E_2 обозначим L_2. Мы будем рассматривать K- и A- замыкания [1] множеств автоматов из L_2. В теореме А.А. Часовских [2] доказано: f — линейный автомат тогда и только тогда, когда $f(x_1, \ldots, x_n) = \sum\limits_{i=1}^{n} \mu_i x_i + \mu_0$, где $n \in \mathbb{N}$, $\mu_j \in E_2^{'}(\xi)\ \forall\ j = \overline{0, n}$. Переменную x_i назовем непосредственной переменной линейного автомата f, если $\mu_i(0) = 1$.