



DOI: <https://doi.org/10.15688/mpcm.jvolsu.2018.4.4>

УДК 004.89

ББК 55.6

АЛГОРИТМЫ ФОРМИРОВАНИЯ ДВУМЕРНЫХ ПРИЗНАКОВ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ МОЛОЧНЫХ ЖЕЛЕЗ МЕТОДАМИ КОМБИНИРОВАННОЙ РАДИОТЕРМОМЕТРИИ¹

Андрей Васильевич Зенович

Доцент кафедры фундаментальной информатики и оптимального управления,
Волгоградский государственный университет
fiou@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Николай Александрович Батурич

Студент кафедры фундаментальной информатики и оптимального управления,
Волгоградский государственный университет
fiou@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Денис Андреевич Медведев

Студент кафедры фундаментальной информатики и оптимального управления,
Волгоградский государственный университет
fiou@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Александр Юрьевич Петренко

Студент кафедры фундаментальной информатики и оптимального управления,
Волгоградский государственный университет
fiou@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Аннотация. Работа является частью группового проекта по созданию консультативной информационной системы, помогающей врачу диагностировать онкологические заболевания молочных желез методом комбинированной термометрии. Для использования в алгоритмах диагностики необходимо сформировать пространство характеристических и высокоинформативных признаков диагностики. В данной работе излагаются три алгоритма построения двумерных признаков и приводятся результаты вычислительных экспериментов, позволяющих оценить чувствительность и специфичность таких признаков. Комбинируя лучшие пары одномерных признаков, удается получить специфичность и чувствительность 70–80 %.

Ключевые слова: микроволновая радиотермометрия, интеллектуальный анализ данных, скрининг молочной железы, экспресс-диагностика злокачественных опухолей, маммология.

Введение

В данное время интенсивно развивается диагностика различных заболеваний на базе комбинированной (радио- и инфракрасной) термометрии [3]. Используемый для этих целей диагностический комплекс РТМ-01-РЭС позволяет проводить диагностику на основе измерения температур кожных покровов в ИК-диапазоне и внутренних температур на глубине до 5 см в РТМ-диапазоне. Детальное исследование температурных полей пациента позволяет врачу зафиксировать температурные аномалии, которые зачастую являются признаками серьезных, в том числе онкологических заболеваний. Создателями комплекса разработана методика диагностики заболеваний молочных желез при помощи измерения набора температур в РТМ- и ИК-диапазонах. Достоинствами данной методики является неинвазивный характер обследования, полная безвредность для пациентки и возможность обнаруживать температурные аномалии, порожденные опухолями малых размеров (несколько миллиметров).

Обследование пациентки начинается с измерения температур в опорных точках T_1 и T_2 , расположенных в центре грудной клетки, а также в 9 точках каждой молочной железы и в аксиллярной области (см. рис. 1). Измерение проводится в РТМ- и ИК-диапазонах, так что каждому пациенту соответствует набор из 44 температур. По данным температурам строится тепловая карта молочных желез пациентки, по которой специалист выявляет температурные аномалии, соответствующие предполагаемым опухолям.

Так как метод выявляет опухоли на ранней стадии и безвреден для организма пациентки, логично использовать его при профилактических медицинских осмотрах сотрудников, которые проводятся ежегодно в большинстве государственных учреждений и частных организаций. К сожалению, для постановки диагноза необходим специалист-маммолог высокой квалификации, знакомый с методом комбинированной радиотермометрии. Для массового применения метода необходимо создание электронной консультативной системы с элементами искусственного интеллекта, которая помогала бы врачу общей практики поставить предварительный диагноз. Такая система последние несколько лет разрабатывается в ВолГУ научным коллективом под руководством доктора физико-математических наук А.Г. Лосева (см., например, [2; 6]). На данный момент в системе реализованы три различных алгоритма диагностики, принятие решений

в системе производится методом взвешенного голосования, реализована возможность встраивания в систему других алгоритмов диагностики.

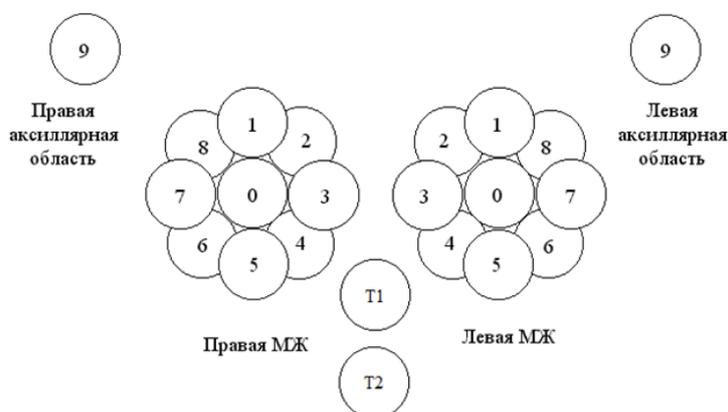


Рис. 1. Схема обследования молочной железы

Как было показано многочисленными вычислительными экспериментами, диагностика непосредственно по набору температур не позволяет достичь удовлетворительной чувствительности и специфичности алгоритма, поэтому для адекватной диагностики необходимо использовать высокоинформативные признаки.

1. Понятие о многомерных признаках

Высокоинформативные признаки строятся путем количественного моделирования качественных признаков заболевания, выявленных специалистами-маммологами [3]. К ним относятся повышенная разница температур сосков, повышенная величина термоасимметрии между одинаковыми областями правой и левой молочной железы, повышенный разброс температур между отдельными точками молочной железы, повышенная температура соска по сравнению со средней температурой молочной железы, повышенная разница кожной и глубинной температуры и некоторые другие.

Для адекватной диагностики представляется логичным изучать области изменения вектор-функций, каждая компонента которых моделирует один из указанных медицинских признаков. В простейшем случае можно изучать одну моделирующую функцию, вычисляя ее значения для каждого пациента из обучающей выборки и находя интервал, принадлежность которому дает высокую вероятность того, что пациент здоров.

Таким образом мы приходим к идее многомерного высокоинформативного признака заболевания.

Обозначим t_0-t_9 внутренние температуры (РТМ) левой молочной железы, $t_{10}-t_{19}$ — внешние температуры (ИК) левой молочной железы, $t_{20}-t_{29}$ — внутренние температуры правой молочной железы, $t_{30}-t_{39}$ — внешние температуры правой. Вторая цифра номера очередной температуры равна номеру области на рисунке 1, в которой измеряется данная температура. Пусть t_{40}, t_{41} — значение РТМ температур в опорных точках T_1 и T_2 , а t_{42}, t_{43} — значение ИК-температур в этих точках. Таким образом, j -му пациенту соответствует вектор из 44-х температур $(t_0^j, t_1^j, \dots, t_{43}^j)$. Пусть обучающая выборка состоит из n пациентов, первые k из которых являются здоровыми.

Пусть f^q — q -я исследуемая вектор-функция, рассматриваемая на множестве век-

торов $\Omega^q = \left\{ \left(t_0^j, t_1^j, \dots, t_{43}^j \right) \right\}$. Здесь $j = 1, 2, \dots, n$. Обозначим f_j^q значение q -й моделирующей вектор-функции на температурных данных j -го пациента.

Назовем высокоинформативным признаком тройку (f^q, V, X) , где X — «информативная» область множества значений функции f^q , $V = I(f^q, X)$ — информативность признака. Под информативностью понимается числовая функция, которая показывает, насколько хорошо указанный признак отделяет класс «больных» от класса «здоровых». При построении признаков использовались статистическая, эвристическая, энтропийная информативность, их различные комбинации, а также некоторые другие функции (см. [6]).

Будем говорить, что признак выполнен для j -го пациента, если f_j^q принадлежит X . Выполнение каждого признака повышает вероятность того, что пациент болен.

В одномерном случае, когда X является подмножеством R , а моделирующая функция f^q является обычной одномерной функцией, технология получения и применения соответствующих признаков детально разработана (см. [6]), реализован соответствующий модуль, который встроено в консультативную информационную систему. С геометрической точки зрения задача заключается в том, чтобы найти отрезок действительной оси, на котором расположены значения функции f^q подавляющего большинства «здоровых» пациентов из обучающей выборки и небольшого количества «больных». Для размерности 2 и выше задача становится нетривиальной.

Детальнее рассмотрим задачу построения двумерных высокоинформативных признаков. В этом случае значение вектор-функции f^q для каждого пациента представляется точкой координатной плоскости. Фактически можно предположить, что координатная плоскость изначально раскрашена в два цвета (синий соответствует значениям f^q для «здоровых» пациентов, красный — для «больных», причем логично предполагать, что множество синих точек образует выпуклую область V). Нам известен цвет конечного множества точек плоскости, соответствующих пациентам из обучающей выборки. Требуется найти алгоритм определения цвета каждой точки плоскости, при котором множество синих точек с хорошей точностью приближает область V .

В работе предложено три алгоритма решения данной задачи. Первый подход заключается в использовании кластеризации с помощью нейронных сетей. Второй делит точки обучающей выборки на кластеры с помощью известного алгоритма DBSCAN (Density-based spatial clustering of applications with noise) [1]. Третий подход заключается в построении семейства сужающихся эллипсов.

2. Моделирующие функции одномерных признаков, используемые для построения двумерных

На данный момент моделирующие функции одномерных признаков исследованы достаточно хорошо [6], получено около полутора тысяч моделирующих функций. Функции разделены на несколько групп. Помимо групп, моделирующих известные качественные медицинские признаки, получена группа функций, базирующихся на физико-математических моделях поведения температурных полей. Реализованы алгоритмы минимизации наборов одномерных признаков. В качестве моделирующих функций для построения двумерных признаков целесообразно взять небольшое количество функций, каждая из которых адекватно описывает всю группу, моделирующую соответствующий медицинский признак.

В качестве примера рассмотрим функции, моделирующие термоасимметрию. Повышенное значение разности температур между одноименными точками правой и левой молочных желез могут быть описаны функциями вида $|t_j - t_{(j+20)}|$, где $j = 0, \dots, 18$. Повышенное значение моделирующей функции при конкретном j сигнализирует о температурной аномалии в отдельной точке соответствующей молочной железе. Из этих функций сформируем две укрупненные моделирующие функции. ИК-термоасимметрию будет описывать $Pr5 = \max(|t_j - t_{j+20}|)$, где $j = 10, \dots, 18$, а РТМ-термоасимметрию будет описывать функция $Pr6 = \max(|t_j - t_{j+20}|)$, где $j = 0, \dots, 8$. Аналогичным образом были сформированы 6 укрупненных моделирующих функций, описывающих медицинские признаки (табл. 1).

Таблица 1

Моделирующие функции, используемые для построения двумерных признаков

| Имя функции | Признак |
|-------------|---|
| Pr1 | Повышенная (пониженная) ИК-температура соска |
| Pr2 | Повышенная (пониженная) РТМ-температура соска |
| Pr3 | Внутренний градиент |
| Pr4 | Разностный аналог второй производной |
| Pr5 | ИК-термоасимметрия |
| Pr6 | РТМ-термоасимметрия |

Выбирая пары из полученных шести функций, получаем 15 двумерных признаков. Обучающая выборка состояла из 73 «здоровых» и 56 «больных» пациентов. Эффективность проверялась на тестовой выборке из 36 «здоровых» и 27 «больных» пациентов. Для каждого двумерного признака вычислялась чувствительность *Sens* (доля верно диагностированных больных) и специфичность *Spec* (доля верно диагностированных здоровых). Мерой эффективности признака G считали среднее геометрическое чувствительности и специфичности.

3. Алгоритм кластеризации с помощью нейросетей

Рассматриваются два варианта алгоритма кластеризации. В первом варианте кластеризация проводится с помощью нейронной сети Кохонена с обучением без учителя [5]. Сеть состоит из слоя из N нейронов, каждый из которых соединен с входным вектором. Веса каждого i -го нейрона представляют собой вектор $W_i = [w_{i1}, w_{i2}]$, размерность которого совпадает с размерностью входного вектора $X_j = [x_{j1}, x_{j2}]$. Количество нейронов равно числу кластеров. В нашем случае производится деление на «больных» и «здоровых», поэтому $N = 2$.

Обозначим W^t — вектор весов на t -й итерации, W^{t+1} — вектор весов на следующей $(t + 1)$ -й итерации, $\eta(t) = \frac{A}{t+B}$ — функция скорости обучения. Легко видеть, что $\eta(t)$ уменьшается с увеличением t .

Алгоритм обучения нейронной сети Кохонена без учителя ([5]) состоит в повторении последовательности из трех описанных ниже шагов до тех пор, пока веса нейронов не стабилизируются.

- Шаг 1: Выбрать X_j из выборки.

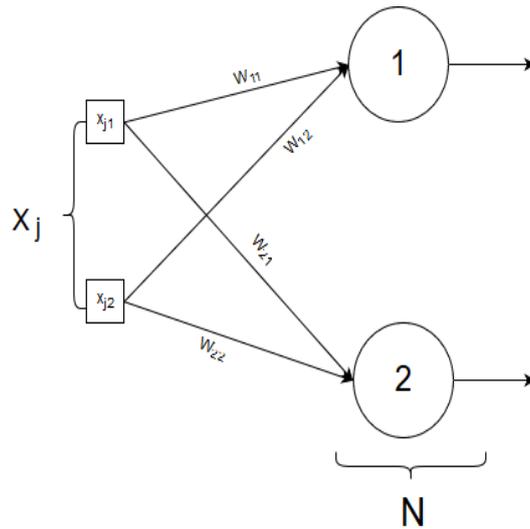


Рис. 2. Нейронная сеть Кохонена

- Шаг 2: Вычислить расстояние $d(X_j, W_i^t)$ между входным вектором X_j и вектором веса W_i^t каждого нейрона. Нейрон с наименьшим расстоянием до X_j назовем победителем, его номер обозначим за p .
- Шаг 3: Изменить веса нейронов по формуле: $W_i^{t+1} = W_i^t + K(i, p)\eta(t)(X_j - W_i^t)$.

В результате обучения векторы веса всех нейронов будут центрами различных кластеров. На рисунке 3 приведен пример работы данного алгоритма для одного из двумерных признаков. Крестиками обозначены «здоровые» пациенты, галочками — «больные». Два вида значков для «здоровых» и для «больных» соответствуют пациентам обучающей и тестовой выборки. Признак демонстрирует чувствительность и специфичность выше 80 %, что является хорошим показателем.

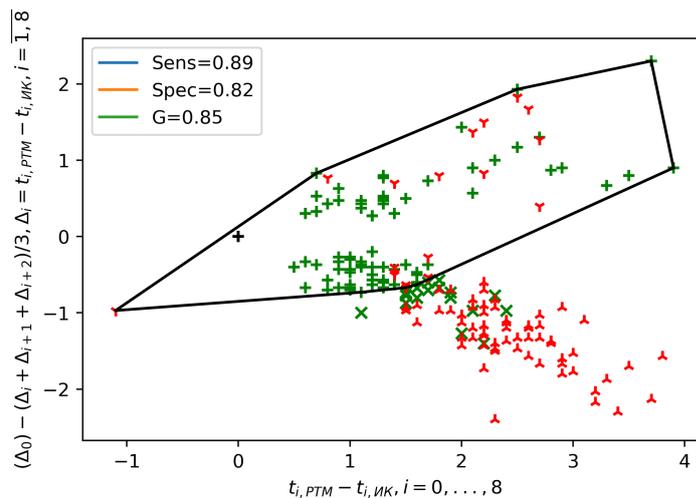


Рис. 3. Пример работы алгоритма обучения без учителя

В качестве второй версии алгоритма рассмотрим нейронную сеть Кохонена с обучением с учителем [5; 7]. Каждому нейрону добавим метку класса. Структура сети и

алгоритм обучения останутся такими же, как и в первом алгоритме, за исключением шага 3. На шаге 3 применяем следующие правила:

- если метка Y_j вектора X_j совпала с меткой нейрона, то нейрон движется в сторону входного вектора, $W_i^{t+1} = W_i^t + \eta(t)(X_j - W_i^t)$;
- иначе $W_i^{t+1} = W_i^t - \eta(t)(X_j - W_i^t)$.

В результате обучения векторы веса всех нейронов будут центрами скопления векторов соответствующих классов. На рисунке 4 показан пример работы второй версии алгоритма.

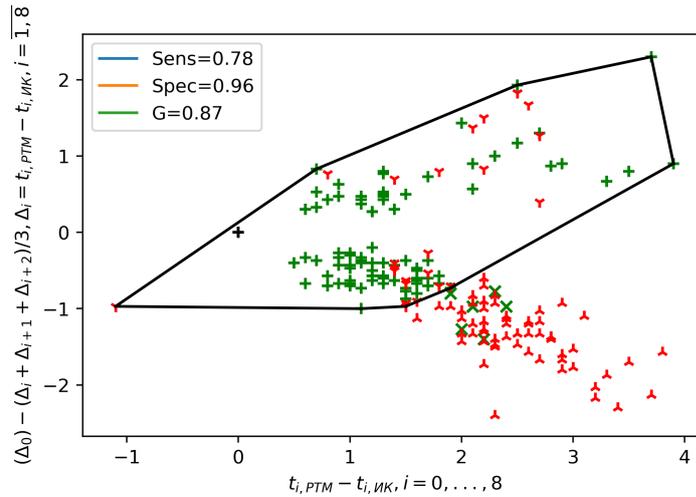


Рис. 4. Пример работы алгоритма обучения с учителем

4. Кластеризация DBSCAN

Алгоритм DBSCAN (см. [1]) решает задачу разбиения данных на кластеры произвольной формы. Идея алгоритма состоит в том, что плотность точек внутри кластера намного выше, чем в областях между кластерами. Обозначим MP — минимальное количество соседних точек, необходимое для создания кластера. Соседними точками для X являются точки, расстояние от которых до точки X меньше некоторого заранее заданного радиуса ξ .

Кластером A будем называть максимальный набор точек, в котором количество элементов больше MP и для каждой пары точек x, y из A существует цепочка $\{x, z_1, \dots, z_n, y\}$ такая, что расстояние между любыми двумя подряд идущими членами этой цепочки меньше ξ .

При данном подходе автоматически выявляются потенциальные выбросы. Ими становятся точки, не попавшие ни в один из кластеров, а, следовательно, находящиеся далеко от основных скоплений точек.

Алгоритм поиска кластеров состоит из следующих шагов:

- *Шаг 1:* Взять любую точку из выборки. Если она посещена или принадлежит какому-либо кластеру, повторить Шаг 1, иначе перейти на Шаг 2.

- *Шаг 2:* Отметить текущую точку как посещенную. Если количество соседей текущей точки меньше MP , вернуться на Шаг 1, иначе объявить эту точку началом нового кластера и перейти на Шаг 3.
- *Шаг 3:* Если нет не посещенных соседей для этой точки, перейти на Шаг 4, иначе отметить всех соседей как принадлежащих текущему кластеру. Взять не посещенную точку, принадлежащую к группе «соседние», отметить ее как посещенную и найти всех ее соседей. Если количество соседей не меньше, чем MP , то перейти на Шаг 3 с этими точками, иначе отметить их как принадлежащие текущему кластеру и вернуться на один уровень рекурсии назад.
- *Шаг 4:* Если в выборке остались не посещенные точки, перейти на Шаг 1, иначе отметить все точки вне кластеров как выбросы и закончить работу алгоритма.

После кластеризации пациент из тестовой выборки считался «больным», если он попадал в кластер с преобладающим числом «больных». В противном случае он считался «здоровым». На рисунке 5 показан пример работы алгоритма. Разными значками обозначены элементы различных кластеров.

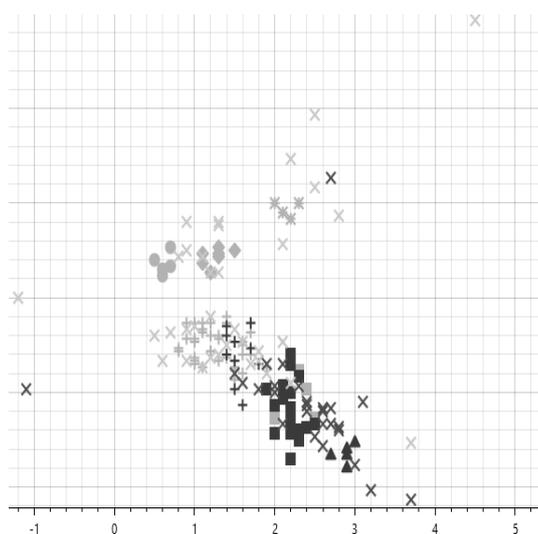


Рис. 5. Результат кластеризации методом DBSCAN

В таблице 2 приведены результаты проверки алгоритма на тестовой выборке.



Рис. 6. Изменение эффективности при уменьшении количества «здоровых» пациентов

Результаты работы алгоритма кластеризации методом DBSCAN

| Признаки | <i>Spec</i> | <i>Prec</i> | <i>G</i> | ξ | Кол-во кластеров | MP | Кол-во выбросов |
|-----------|-------------|-------------|----------|-------|------------------|----|-----------------|
| Pr1 и Pr3 | 0,68 | 0,79 | 0,73 | 0,29 | 6 | 3 | 50 |
| Pr1 и Pr4 | 0,51 | 0,75 | 0,62 | 0,59 | 3 | 3 | 11 |
| Pr2 и Pr1 | 0,86 | 0,71 | 0,79 | 0,37 | 6 | 3 | 19 |
| Pr2 и Pr3 | 0,51 | 0,82 | 0,65 | 0,54 | 2 | 4 | 10 |
| Pr2 и Pr4 | 0,68 | 0,93 | 0,79 | 0,38 | 6 | 3 | 17 |
| Pr3 и Pr4 | 0,81 | 0,86 | 0,83 | 0,24 | 6 | 3 | 28 |
| Pr5 и Pr1 | 0,81 | 0,32 | 0,51 | 1,15 | 3 | 20 | 20 |
| Pr5 и Pr2 | 0,81 | 0,39 | 0,56 | 0,81 | 2 | 13 | 16 |
| Pr5 и Pr3 | 0,70 | 0,93 | 0,81 | 0,32 | 5 | 3 | 34 |
| Pr5 и Pr4 | 0,95 | 0,46 | 0,66 | 0,37 | 8 | 3 | 25 |
| Pr6 и Pr1 | 0,89 | 0,46 | 0,64 | 0,65 | 3 | 3 | 13 |
| Pr6 и Pr2 | 0,57 | 0,93 | 0,73 | 0,31 | 7 | 3 | 35 |
| Pr6 и Pr3 | 0,89 | 0,57 | 0,71 | 0,50 | 3 | 3 | 16 |
| Pr6 и Pr4 | 0,41 | 1,00 | 0,64 | 0,64 | 2 | 12 | 17 |
| Pr6 и Pr5 | 0,68 | 0,71 | 0,69 | 0,21 | 9 | 3 | 52 |

5. Построение областей для двумерных признаков методом сужающихся эллипсов

Некоторые алгоритмы построения областей для двумерных признаков реализованы в [4]. Алгоритм, предложенный в данной работе, базируется на аналогичных идеях и является улучшением алгоритмов из [4].

Алгоритм начинается с построения основного эллипса, внутри которого лежат все «здоровые» пациенты из обучающей выборки. Далее необходимо сужать данный эллипс с целью достижения максимальной эффективности. Очевидная идея заключается в том, чтобы последовательно выбрасывать самые удаленные «здоровые» точки, проводить эллипс, включающий все оставшиеся, а потом варьировать (уменьшать) полуоси. На рисунке 6 приведен график зависимости эффективности от количества выброшенных точек. Монотонной зависимости не наблюдается, поэтому необходимо сохранять эллипс с максимальной эффективностью для каждого количества выброшенных точек, и из них выбирать наилучший.

Окончательная версия алгоритма состоит из следующих шагов:

- *Шаг 1:* Строим эллипс минимальной площади, внутри которого лежат точки всех «здоровых» пациентов из обучающей выборки, с помощью известного алгоритма Хачияна [8].
- *Шаг 2:* Считаем для каждой точки сумму расстояний от нее до всех остальных.
- *Шаг 3:* Исключаем точку с наибольшей суммой.
- *Шаг 4:* Перестраиваем эллипс по оставшимся точкам.
- *Шаг 5:* Сужаем построенный эллипс. Оставляя неизменным центр эллипса и направление его главных осей, варьировем его полуоси. Получаем двухпараметрическое семейство эллипсов, среди которых выбираем эллипс с наибольшей эффективностью. Сохраняем его.

Результаты работы алгоритма сужающихся эллипсов

| Названия пары признаков | <i>Sens</i> | <i>Spec</i> | <i>G</i> |
|-------------------------|-------------|-------------|----------|
| Pr6-Pr5 | 0,85 | 0,81 | 0,83 |
| Pr6-Pr2 | 0,78 | 0,78 | 0,78 |
| Pr6-Pr1 | 0,56 | 0,89 | 0,70 |
| Pr6-Pr3 | 0,85 | 0,69 | 0,77 |
| Pr6-Pr4 | 0,74 | 0,94 | 0,84 |
| Pr5-Pr2 | 0,59 | 0,89 | 0,73 |
| Pr5-Pr1 | 0,59 | 0,94 | 0,75 |
| Pr5-Pr3 | 0,44 | 0,92 | 0,64 |
| Pr5-Pr4 | 0,78 | 0,94 | 0,86 |
| Pr2-Pr1 | 0,85 | 0,83 | 0,84 |
| Pr2-Pr3 | 0,78 | 0,72 | 0,75 |
| Pr2-Pr4 | 0,81 | 0,83 | 0,82 |
| Pr1-Pr3 | 0,78 | 0,72 | 0,75 |
| Pr1-Pr4 | 0,52 | 0,89 | 0,68 |
| Pr3-Pr4 | 0,63 | 0,89 | 0,75 |

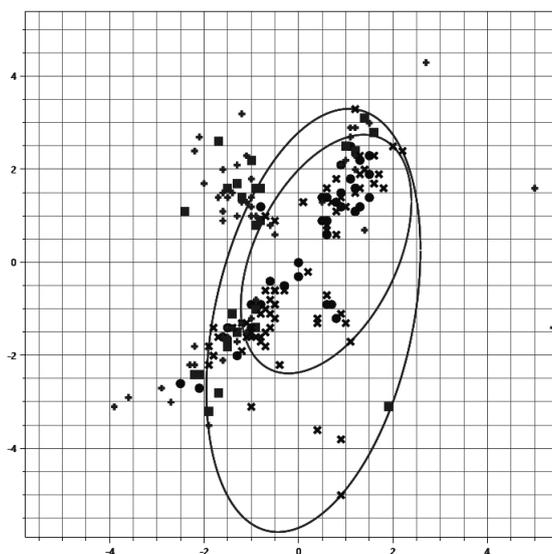


Рис. 7. Пример работы алгоритма

- *Шаг 6*: Если не достигнуто минимальное число точек, по которому можно построить эллипс, из оставшихся точек исключаем точку с наибольшей суммой расстояний и переходим к Шагу 4.

Из сохраненных на шаге 5 эллипсов выбираем эллипс с наибольшей эффективностью.

На рисунке 7 приведен пример работы алгоритма. Два типа крестиков соответствуют «здоровым» пациентам обучающей и тестовой выборки, квадраты и круги — «больным» пациентам обучающей и тестовой выборки. Большой эллипс охватывает всех «здоровых» пациентов из обучающей выборки, малый эллипс — область, на которой до-

стигается максимальная эффективность.

В таблице 3 приведены результаты работы данного алгоритма. Лучшие признаки дают эффективность порядка 80–85 %.

ПРИМЕЧАНИЕ

¹ Исследование выполнено при финансовой поддержке РФФИ и Администрации Волгоградской области, проект № 18-47-340004 р_а «Управление качеством классификации в диагностике рака молочной железы по многомерным данным микроволновой радиотермометрии».

СПИСОК ЛИТЕРАТУРЫ

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / Е. И. Большакова, Э. С. Клышинский, Д. В. Ландэ, А. А. Носков, О. В. Пескова, Е. В. Ягунова. — М. : МИЭМ, 2011. — 272 с.
2. Алгоритмы принятия решений в консультативной интеллектуальной системе диагностики молочных желез / А. В. Зенович, В. А. Глазунов, А. С. Опарин, Ф. Г. Примаченко // Вестник Волгоградского государственного университета. Серия 1, Математика. Физика. — 2016. — № 6 (37). — С. 141–154. — DOI: 10.15688/jvolsu1.2016.6.13.
3. Веснин, С. Г. Современная микроволновая радиотермометрия молочных желез / С. Г. Веснин, М. А. Каплан, Р. С. Авакян // Опухоли женской репродуктивной системы. — 2008. — № 3. — С. 28–33.
4. Интеллектуальный анализ многомерных термометрических данных в медицинской диагностике / А. Г. Лосев, А. В. Зенович, О. А. Бочкарев, В. В. Левшинский // Вестник Волгоградского государственного университета. Серия 1, Математика. Физика. — 2016. — № 5 (36). — С. 150–161. — DOI: 10.15688/jvolsu1.2016.5.13.
5. Кохонен, Т. Самоорганизующиеся карты / Т. Кохонен. — М. : БИНОМ. Лаборатория знаний, 2008. — 655 с.
6. Лосев, А. Г. Интеллектуальный анализ данных микроволновой радиотермометрии в диагностике рака молочной железы / А. Г. Лосев, В. В. Левшинский // Математическая физика и компьютерное моделирование. — 2017. — Т. 20, № 5. — С. 49–62. — DOI: 10.15688/мрсм.jvolsu.2017.5.6.
7. Ясницкий, Л. Н. Интеллектуальные системы / Л. Н. Ясницкий. — М. : Лаборатория знаний, 2016. — 221 с.
8. Khachiyani, L. G. Rounding of Polytopes in the Real Number Model of Computation / L. G. Khachiyani // Mathematics of Operations Research. — 1996. — Vol. 21, № 2. — P. 307–320.

REFERENCES

1. Bolshakova E.I., Klyshinskiy E.S., Lande D.V., Noskov A.A., Peskova O.V., Yagunova E.V. *Avtomaticheskaya obrabotka tekstov na estestvennom yazyke i kompyuternaya lingvistika* [Automatic Processing of Texts in Natural Language and Computational Linguistics]. Moscow, MIEM Publ., 2011. 272 p.
2. Zenovich A.V., Glazunov V.A., Oparin A.S., Primachenko F.G. *Algoritmy prinyatiya resheniy v konsultativnoy intellektualnoy sisteme diagnostiki molochnykh zhelez* [Algorithms of Decision-Making in Intelligent Advisory System for Diagnostics of the Mammary Glands]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 1, Matematika. Fizika* [Science Journal of Volgograd State University. Mathematics. Physics], 2016, no. 6 (37), pp. 141-154. DOI: 10.15688/jvolsu1.2016.6.13.

3. Vesnin S.G., Kaplan M.A., Avakyan R.S. Sovremennaya mikrovolnovaya radiotermometriya molochnykh zhelez [Modern Microwave Radiometry of the Mammary Glands]. *Opukholi zhenskoy reproduktivnoy sistemy*, 2008, no. 3, pp. 28-33.

4. Losev A.G., Zenovich A.V., Bochkarev O.A., Levshinskiy V.V. Intellektualnyy analiz mnogomernykh termometricheskikh dannykh v meditsinskoj diagnostike [Multidimensional Thermometric Data Mining in Medical Diagnostics]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 1, Matematika. Fizika* [Science Journal of Volgograd State University. Mathematics. Physics], 2016, no. 5 (36), pp. 150-161. DOI: 10.15688/jvolsu1.2016.5.13.

5. Kokhonen T. *Samoorganizuyushchiesya karty* [Self-Organizing Maps]. Moscow, BINOM. Laboratoriya znaniy Publ., 2008. 655 p.

6. Losev A.G., Levshinskiy V.V. Intellektualnyy analiz dannykh mikrovolnovoy radiotermometrii v diagnostike raka molochnoy zhelezy [Data Mining of Microwave Radiometry Data in the Diagnosis of Breast Cancer]. *Matematicheskaya fizika i kompyuternoe modelirovanie* [Mathematical Physics and Computer Simulation], 2017, vol. 20, no. 5, pp. 49-62. DOI: 10.15688/mpcm.jvolsu.2017.5.6.

7. Yasnitskiy L.N. *Intellektualnye sistemy* [Intellectual Systems]. Moscow, Laboratoriya znaniy Publ., 2016. 221 p.

8. Khachiyan L.G. Rounding of Polytopes in the Real Number Model of Computation. *Mathematics of Operations Research*, 1996, vol. 21, no. 2, pp. 307-320.

**ALGORITHMS FOR THE FORMATION
OF TWO-DIMENSIONAL CHARACTERISTIC AND INFORMATIVE SIGNS
OF DIAGNOSIS OF DISEASES OF THE MAMMARY GLANDS
BY THE METHODS OF COMBINED RADIO THERMOMETRY**

Andrey Vasilyevich Zenovich

Associate Professor, Department of Fundamental Computer Science
and the Optimal Control,
Volgograd State University
fiou@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Nikolay Aleksandrovich Baturin

Student, Department of Fundamental Computer Science
and the Optimal Control,
Volgograd State University
fiou@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Denis Andreevich Medvedev

Student, Department of Fundamental Computer Science
and the Optimal Control,
Volgograd State University
fiou@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Alexander Yuryevich Petrenko

Student, Department of Fundamental Computer Science
and the Optimal Control,
Volgograd State University
fiou@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Abstract. Current paper is a part of group project to create a consultative information system that helps doctors to diagnose breast cancer with combined thermometry. For use in diagnostic algorithms, it is necessary to form a space of characteristic and highly informative diagnostic features. This paper outlines three algorithms for constructing two-dimensional features and presents the results of computational experiments that allow one to estimate the sensitivity and specificity of such features. Combining the best pairs of one-dimensional signs, it is possible to obtain the specificity and sensitivity of 70–80 %.

Key words: microwave radiothermometry, data mining, breast screening, express diagnostics of malignant breast tumors, mammology.