

О построении признакового пространства в задаче обучения¹

Б.М. Гавриков, Н.В. Пестрякова

Аннотация. Рассмотрена проблема оценивания состояния здоровья человека на основе результатов анализа периферической крови. Исследования проведены по четырем градациям поражения систем организма отдельно для мужчин и женщин. Используется статистический метод классификации, базирующийся на полиномиально-регрессионном подходе и имеющий вероятностные оценки. В разработанных и обученных версиях классификатора реализованы полные наборы комбинаций мономов до пятого порядка.

Ключевые слова: состояние здоровья человека, периферическая кровь, классификация, полиномиальная регрессия.

Введение

Ключевой проблемой в медицине является качество диагностики, поскольку от правильной постановки диагноза зависит адекватность дальнейшего лечения. Основным подходом к оптимизации медицинской диагностики является поиск некоторого универсального правила, по которому принимается то или иное решение. Для получения такого правила нужно обработать огромное количество данных медицинского обследования множества людей, результатом которого стал бы правильно поставленный диагноз. Отсюда вытекает необходимость развития методов предварительной диагностики, основанных на технологиях распознавания образов, эффективных при решении задач такого рода.

Одним из основных средств, используемых врачами при первичной диагностике, является анализ периферической крови (из пальца), состоящий из ряда показателей. Их набор определяется типом автоматического анализатора и включает 15 – 20, а иногда и большее число наименований. Известные гематологи, Кассир-

ский И.А., Воробьев А.И., Бергану Ш., Vinatier I., Nayachy Y. и другие, отмечали, что любое заболевание организма проявляется в виде изменения показателей крови.

Для здорового человеческого организма параметры крови варьируются в некоторых известных диапазонах [5]. Значительные отклонения от нормы могут быть характерными проявлениями определенных заболеваний, выявить которые не составляет большого труда. Но во многих случаях ситуация неоднозначная, и врачу необходима помощь в принятии решения о состоянии систем организма (СО).

1. Общая постановка задачи

Предложен способ, позволяющий оценивать состояния здоровья человека (СЗЧ) по четырем градациям поражения организма на основании лабораторного анализа периферической крови. Используется разработанный ранее авторами статистический метод классификации, базирующийся на полиномиально-регрессионном подходе и имеющий вероятностные оценки.

¹ Работа выполнена при финансовой поддержке РФФИ (грант №16-07-00742 а).

Для мужчин и женщин строятся и используются различные классификаторы, поскольку диапазоны вариации показателей крови людей существенно зависят от пола. Кроме того, гинекологическая система имеется только у женщин.

По каждой СО проводится самостоятельное исследование СЗЧ посредством отдельного классификатора. При его построении используется обучающая выборка, соответствующая рассматриваемой СО.

СЗЧ включает четыре градации – от практически здорового состояния до максимальной степени поражения организма (онкологические заболевания):

- 1 класс – здоровые;
- 2 класс – начальные отклонения состояния здоровья;
- 3 класс – выраженное отклонение состояния здоровья;
- 4 класс – тяжелое заболевание.

При обучении используются выборки, полученные в результате детального обследования пациентов большой группой специалистов из различных областей медицины [1]. Для конкретной СО из рассматриваемого перечня к каждой из четырех возможных градаций СЗЧ относится список заболеваний, соответствующих этой СО [5].

База показателей крови практически здоровых людей одинакова для всех СО, но при этом существенно различается для мужчин и женщин.

Рассматривается задача предварительной диагностики, при которой выявляется не конкретное заболевание, а состояние каждой из СО пациента. В случае, если будет отмечено отклонение от нормы, необходимо дальнейшее обследование соответствующими узкими специалистами.

Предложенная методика может применяться также для наблюдения за динамикой состояния СО пациентов, проходящих сложное лечение, например, онкобольных. Используемые высокотоксичные химические препараты, а также лучевая терапия, убивая онкоклетки, одновременно наносят существенный вред организму в целом. Врачу-онкологу необходимо контролировать состояние всех органов при проведении подобной терапии и в случае необходимости вносить определенные коррективы в протокол лечения. Требуется соблюдение некоторого разумного балан-

са между вызванными лечением негативными проявлениями и принесенной им пользой.

Основанием для использования при решении описанной задачи подхода, основанного на полиномиальной регрессии [2-5], является то, что данный метод хорошо зарекомендовал себя при распознавании столь сложных объектов, как печатные и рукопечатные символы. Этот статистический метод является точным, быстрым, устойчивым к шумам, генерирует монотонные (надежные) оценки, имеющие вероятностную природу.

При построении дифференцированных по полу обучающих выборок для каждой из СО используются измеренные на автоматизированном анализаторе крови стабильные показатели. Они перечислены ниже, причем приведены их общепринятые обозначения и размерность:

- RBC [L^{-1}] – эритроциты,
- HGB [$g L^{-1}$] – гемоглобин,
- PLT [L^{-1}] – тромбоциты,
- WBC [L^{-1}] – лейкоциты,
- LIMPH [L^{-1}], [%] – лимфоциты,
- GRAN [L^{-1}], [%] – гранулоциты
(GRAN = NEUT + EOS + BASO).

Ввиду объективных трудностей составления обучающих выборок принцип соразмерности не выполняется в необходимой степени. А именно, для ряда СО некоторые из градаций СЗЧ недостаточно заполнены или вообще отсутствуют. По каждому из полов градация, соответствующая здоровым людям (класс «1»), наиболее обширная и ввиду универсальности используется для всех СО. Следующим по заполненности является класс «4». Наборы выборок показателей крови для классов «2», «3», «4» уникальны для каждой из СО. Предъявляемое к выборкам условие случайности, несомненно, выполняется для параметров крови. Это обусловлено большим разнообразием человеческих организмов.

2. Математическая постановка задачи

Для конкретной СО по имеющемуся анализу периферической крови определим, какой из четырех градаций СЗЧ он соответствует. Перечень градаций СЗЧ представляет собой множе-

ство с $K = 4$ элементами. Для представления наборов показателей крови вводится вектор $\mathbf{v} \in \mathbf{R}^N$, i -я компонента которого суть отнормированная величина i -го показателя крови, лежащая на отрезке $[0,1]$, причем $N = 8$.

Пространство первичных признаков определяется параметрами крови, которые измеряются несопоставимыми величинами, а поэтому принципиально различаются и по наименованию, и по порядкам величин, и по диапазону вариации. Число независимых компонент полиномиального вектора соответствует количеству используемых при построении классификатора параметров крови и равно восьми в данном исследовании. Их набор определяется типом автоматического анализатора крови и включает существенно большее число наименований. Однако нет необходимости использовать какие-либо параметры крови в дополнение к этим восьми, поскольку они представляют собой их комбинации и не являются независимыми.

Нормирование на отрезок $[0,1]$ проводим следующим образом. Пусть по рассматриваемой СО имеется обучающая выборка, включающая все градации СЗЧ. Для каждого i -го показателя крови находим минимальное и максимальное значение v_i^{\min}, v_i^{\max} , где $i = 1, \dots, N$.

$$v_i^{\min} = \min_j \{v_i^j\}, j=1, \dots, Q$$

$$v_i^{\max} = \max_j \{v_i^j\}, j=1, \dots, Q$$

где Q – объем выборки по данной СО.

Затем выполняем следующее преобразование:

$$v_i \rightarrow (v_i - v_i^{\min}) / (v_i^{\max} - v_i^{\min}).$$

Отождествляем k -й элемент множества градаций СЗЧ с базисным вектором $\mathbf{e}_k = (0 \dots 1 \dots 0)$ (здесь 1 находится на k -м месте, причем $1 \leq k \leq K$) из \mathbf{R}^K . Обозначаем $Y = \{\mathbf{e}_1, \dots, \mathbf{e}_K\}$.

Пусть можно найти $p_k(\mathbf{v})$ – вероятность того, что набор (отнормированных) показателей крови соответствует k -му элементу СЗЧ, где $1 \leq k \leq K$. Искомый элемент СЗЧ будет иметь порядковый номер r , где

$$p_r(\mathbf{v}) = \max_k \{p_k(\mathbf{v})\}, 1 \leq k \leq K. \quad (1)$$

Приближенные значения компонент $(p_1(\mathbf{v}), \dots, p_K(\mathbf{v}))$ представляются в виде многочленов от координат $\mathbf{v} = (v_1, \dots, v_N)$:

$$p_k(\mathbf{v}) \cong c_0^{(k)} + \sum_{i=1}^N c_i^{(k)} v_i + \sum_{i,j=1}^N c_{i,j}^{(k)} v_i v_j + \dots, \quad (2)$$

$$1 \leq k \leq K.$$

Суммы в правых частях равенств (2) конечные и определяются выбором базисных мономов. А именно, если

$$\mathbf{x}(\mathbf{v}) = (1, v_1, \dots, v_N, \dots)^T -$$

конечный вектор размерности L из приведенных в (2) базисных мономов, упорядоченных некоторым образом и определяющих соответствующее признаковое пространство, тогда (2) можно записать следующим образом:

$$\mathbf{p}(\mathbf{v}) = (p_1(\mathbf{v}), \dots, p_K(\mathbf{v}))^T \cong A^T \mathbf{x}(\mathbf{v}), \quad (3)$$

где A – матрица размера $L \times K$, столбцами которой являются векторы $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(K)}$. Каждый такой вектор составлен из коэффициентов при мономах соответствующей строки (2) (с совпадающим верхним индексом), упорядоченных так же, как в векторе $\mathbf{x}(\mathbf{v})$. Следовательно, приближенный поиск вектора вероятностей $\mathbf{p}(\mathbf{v})$ сводится к нахождению матрицы A .

Значение A вычисляется приближенно в процессе обучения с использованием содержащихся в некоторой базе данных множеств пар векторов $[\mathbf{v}^{(1)}, \mathbf{y}^{(1)}], \dots, [\mathbf{v}^{(j)}, \mathbf{y}^{(j)}]$. Здесь первый элемент $\mathbf{v}^{(j)}$ – набор параметров крови, соответствующий элементу СЗЧ с каким-либо номером k ($1 \leq k \leq K$), а второй элемент – его базисный вектор $\mathbf{y}^{(j)} = (0 \dots 1 \dots 0)$, где 1 стоит на k -м месте, $1 \leq j \leq J$:

$$A \cong \left(\frac{1}{J} \sum_{j=1}^J \mathbf{x}^{(j)} (\mathbf{x}^{(j)})^T \right)^{-1} \left(\frac{1}{J} \sum_{j=1}^J \mathbf{x}^{(j)} (\mathbf{y}^{(j)})^T \right). \quad (4)$$

При получении правой части (4) используется следующая рекуррентная процедура, где A_0 и G_0 заданы:

$$A_j = A_{j-1} - \alpha_j G_j \mathbf{x}^{(j)} [A_{j-1}^T \mathbf{x}^{(j)} - \mathbf{y}^{(j)}]^T, \quad \alpha_j = 1 / (m J / 2), \quad m \geq 1$$

$$G_j = \frac{1}{1 - \alpha_j} \left[G_{j-1} - \alpha_j \frac{G_{j-1} \mathbf{x}^{(j)} (\mathbf{x}^{(j)})^T G_{j-1}}{1 + \alpha_j ((\mathbf{x}^{(j)})^T G_{j-1} \mathbf{x}^{(j)} - 1)} \right], \quad (5)$$

$$1 \leq j \leq J$$

$$G_j \cong D^{-1}, \quad D = \text{diag} (E\{x_1^2\}, E\{x_2^2\}, \dots, E\{x_L^2\}).$$

Здесь x_1, x_2, \dots, x_L – компоненты вектора $\mathbf{x}(\mathbf{v})$, m – экспериментально подбираемый множитель. Получаемые оценки могут выходить за рамки отрезка $[0,1]$ из-за того, что используе-

мый метод является приближенным. Отрицательные значения обнуляли, а превышающие 1 делали равными 1.

Табл. 1. Зависимость точности классификации от структуры полинома для основных систем организма

Пищеварительная система												
м/ж- объем базы	длина полинома											
	33			61			69			77		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-109							13	11,9%	2	12	11,0%	2
ж-77	12	15,6%	2	4	5,2%	2	4	5,2%	2	4	5,2%	2
м/ж- объем базы	длина полинома											
	85			165			495			1287		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-109	13	11,9%	2	4	3,7%	2	2	1,8%	4	2	1,8%	8
ж-77	1	0,8%	2	0	0%	2						
Органы дыхания												
м/ж- объем базы	длина полинома											
	33			61			69			77		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-76							4	5,3%	2	1	1,3%	2
ж-62							0	0%	2			
м/ж- объем базы	длина полинома											
	85			165			495			1287		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-76	1	1,3%	2	0	0%	4						
Опорно-двигательный аппарат												
м/ж- объем базы	длина полинома											
	33			61			69			77		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-76				1	1,3%	2				0	0%	2
ж-72				1	1,4%	2	1	1,4%	2	1	1,4%	2
м/ж- объем базы	длина полинома											
	85			165			495			1287		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-76												
ж-72	1	1,4%	2	0	0%	4						
Урологическая система												
м/ж- объем базы	длина полинома											
	33			61			69			77		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-110										18	16,4%	2
ж-119							6	5,0%	2	4	3,4%	2
м/ж- объем базы	длина полинома											
	85			165			495			1287		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-110	9	8,2%	2	3	2,7%	2	1	0,9%	6	3	2,7%	18
ж-119				3	2,5%	4	1	0,8%	16	1	0,8%	60

Табл. 1. Зависимость точности классификации от структуры полинома для основных систем организма (продолжение таблицы)

Эндокринная система												
м/ж- объем базы	длина полинома											
	33			61			69			77		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-121	29	24,0%	2	22	18,2%	4	21	17,4%	4	24	19,8%	3
ж-80	6	7,5%	2	2	2,5%	2	2	2,5%	2	2	2,5%	2
м/ж- объем базы	длина полинома											
	85			165			495			1287		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
м-121				9	7,4%	6	6	5,0%	20	6	5,0%	60
ж-80	2	2,5%	2	0	0%	4						
Гинекологическая система												
м/ж- объем базы	длина полинома											
	33			61			69			77		
	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m	mis	mis-%	m
ж-56							0	0%	2			

Рассматривались различные модификации вектора $x(v)$. Указанной в Табл. 1 длине полинома соответствует следующая его структура.

1). Длина полинома 33.

$$x=(1, \{v_i\}, \{v_i^2\}, \{v_i^3\}, \{v_i^4\}, 1 \leq i \leq 8). \quad (6)$$

Имеются мономы степенного вида первого, второго, третьего и четвертого порядка. Перекрестные произведения отсутствуют.

2). Длина полинома 61.

$$x=(1, \{v_i\}, \{v_i^3\}, \{v_i^4\}, \{v_i v_j\}), 1 \leq i \leq 8, i \leq j \leq 8. \quad (7)$$

Имеются мономы степенного вида первого, второго, третьего и четвертого порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

3). Длина полинома 69.

$$x=(1, \{v_i\}, \{v_i^3\}, \{v_i^4\}, \{v_i^5\}, \{v_i v_j\}), 1 \leq i \leq 8, i \leq j \leq 8. \quad (8)$$

Имеются мономы степенного вида первого, второго, третьего, четвертого и пятого порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

4). Длина полинома 77.

$$x=(1, \{v_i\}, \{v_i^3\}, \{v_i^4\}, \{v_i^5\}, \{v_i^6\}, \{v_i v_j\}), 1 \leq i \leq 8, i \leq j \leq 8. \quad (9)$$

Имеются мономы степенного вида первого, второго, третьего, четвертого, пятого и шестого

порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

5). Длина полинома 85.

$$x=(1, \{v_i\}, \{v_i^3\}, \{v_i^4\}, \{v_i^5\}, \{v_i^6\}, \{v_i^7\}, \{v_i v_j\}), 1 \leq i \leq 8, i \leq j \leq 8. \quad (10)$$

Имеются мономы степенного вида первого, второго, третьего, четвертого, пятого, шестого и седьмого порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

6). Длина полинома 165.

$$x=(1, \{v_i\}, \{v_i v_j\}, \{v_i v_j v_k\}), 1 \leq i \leq 8, i \leq j \leq 8, j \leq k \leq 8. \quad (11)$$

Имеются мономы первого, второго и третьего порядка. Перекрестные произведения используются для мономов второго и третьего порядка.

7). Длина полинома 495.

$$x=(1, \{v_i\}, \{v_i v_j\}, \{v_i v_j v_k\}, \{v_i v_j v_k v_l\}), 1 \leq i \leq 8, i \leq j \leq 8, j \leq k \leq 8, k \leq l \leq 8. \quad (12)$$

Имеются мономы первого, второго, третьего и четвертого порядка. Перекрестные произведения используются для мономов второго, третьего и четвертого порядка.

8). Длина полинома 1287.

$$x=(1, \{v_i\}, \{v_i v_j\}, \{v_i v_j v_k\}, \{v_i v_j v_k v_l\}, \{v_i v_j v_k v_l v_m\}), 1 \leq i \leq 8, i \leq j \leq 8, j \leq k \leq 8, k \leq l \leq 8, l \leq m \leq 8. \quad (13)$$

Имеются мономы первого, второго, третьего, четвертого и пятого порядка. Перекрестные произведения используются для мономов второго, третьего, четвертого и пятого порядка.

В (6 - 13) выражения в фигурных скобках соответствуют цепочкам элементов вектора, вычисляемым по всем показателям крови из имеющегося набора.

3. Технология вычислений и полученные результаты

Используемые базы показателей крови обладает значительной вариабельностью и небольшим объемом. Поскольку исходные обучающие множества имели неравноценные по объему подмножества для различных градаций заболевания организма, то при получении обучающей последовательности использова-

лась лишь часть объема данных по классу «1» (здоровых людей), сопоставимая с количеством элементов класса «4» (наивысшая степень заболевания), далее добавлялись элементы классов «2» и «3», а затем проводилось перемешивание по всему объему обучающего множества.

Этот прием, конечно, не позволил осуществить приблизительное выравнивание числа элементов для всех четырех классов по каждой СО. Однако вычислительная практика показала, что используемого количества элементов класса «1» достаточно для безошибочного распознавания всех элементов этого класса, в том числе не включенных в процесс обучения.

В Табл. 2 приведены данные для обоих полов о том, как именно количественно распределены по классам элементы обучающих последовательностей различных СО.

Табл. 2. Распределение объема базы по классам основных систем организма

Пищеварительная система					
<i>Распределение объема базы по классам</i>					
<i>м/ж</i>	«1»	«2»	«3»	«4»	«1-2-3-4»
<i>м</i>	33	17	26	33	109
<i>ж</i>	24	9	20	24	77
Органы дыхания					
<i>Распределение объема базы по классам</i>					
<i>м/ж</i>	«1»	«2»	«3»	«4»	«1-2-3-4»
<i>м</i>	32	11	12	21	76
<i>ж</i>	24	4	10	24	62
Опорно-двигательный аппарат					
<i>Распределение объема базы по классам</i>					
<i>м/ж</i>	«1»	«2»	«3»	«4»	«1-2-3-4»
<i>м</i>	33	3	7	33	76
<i>ж</i>	33	0	6	33	72
Урологическая система					
<i>Распределение объема базы по классам</i>					
<i>м/ж</i>	«1»	«2»	«3»	«4»	«1-2-3-4»
<i>м</i>	33	18	26	33	110
<i>ж</i>	42	0	42	35	119
Эндокринная система					
<i>Распределение объема базы по классам</i>					
<i>м/ж</i>	«1»	«2»	«3»	«4»	«1-2-3-4»
<i>м</i>	34	26	27	34	121
<i>ж</i>	35	10	0	35	80
Гинекологическая система					
<i>Распределение объема базы по классам</i>					
<i>м/ж</i>	«1»	«2»	«3»	«4»	«1-2-3-4»
<i>ж</i>	30	6	5	15	56

В СО «Эндокринная система» (женщины) нет элементов класса «3». Для женщин в базах СО «Опорно-двигательный аппарат», «Урологическая система» отсутствуют элементы класса «2». Однако классификаторы обучаются и обеспечивают высокую точность при распознавании обучающего множества. В наборах альтернатив распознавания эти классы отсутствуют.

Наиболее проблемными в отношении точности классификации являются для мужчин пищеварительная, урологическая и эндокринная системы. Они характеризуются высокой заполненностью классов «2» и «3».

С целью повышения точности распознавания проводилось многократное обучение на одной и той же базе с контролем точности распознавания, поскольку при неограниченном увеличении числа таких итераций точность сначала стабилизируется на некотором минимальном значении, а затем может начинать нарастать. В ряде случаев процесс имеет вид периодических колебаний с одним и тем же значением минимума количества ошибочных распознаваний, но с частично изменяющимся набором неправильно распознанных элементов, а для одинаковых элементов – с различным перечнем альтернатив.

Как показали многократные численные эксперименты, наиболее успешным оказался следующий способ обучения на используемых базах показателей крови. Полученные из них обучающие последовательности повторялись десять раз. Один цикл обучения проводился на такой десятикратной последовательности. Это позволило ускорить данную процедуру без потери информативности в отношении особенностей процесса обучения.

Из приведенных в Табл. 1 результатов видно, что использование перекрестных произведений для мономов второго, третьего, четвертого и пятого порядка позволило значительно увеличить точность классификации.

Заключение

Для разработанного авторами статистического метода распознавания на основе полиномиальной регрессии реализовано приложение в качестве классификатора СЗЧ по показателям периферической крови из пальца для различных СО.

Данный метод по самой своей природе подходит для решения задач медицинской диагностики, поскольку он именно посредством статистической обработки данных, относящихся к большому количеству людей, позволяет найти некоторый принцип, по которому принимается то или иное решение. Статистический характер метода помогает справиться с проблемой «человеческого фактора», если в обучающем множестве показатели крови случайно оказались не в том классе (например, по ошибке медика-специалиста или технического работника). Разумеется, речь идет об относительно небольшом числе таких случаев.

Устойчивость метода к шумам позволит, например, правильно оценить состояние СО в ситуации с не очень точным анализатором крови.

Вероятностный характер оценки означает, что проверяемый набор показателей крови будет отнесен к каждому из четырех классов с некоторой вероятностью. Решением будет считаться класс с максимальной оценкой. Однако часто может представлять интерес не только нахождение класса с максимальной оценкой, но и сопоставление ее с другими оценками.

Сравнительно небольшой объем и возможная неполнота обучающих множеств, несомненно, являются существенным недостатком и могут привести к уменьшению уровня достоверности полученных результатов и, как следствие, к понижению точности классификации в реальной медицинской практике. Однако это обстоятельство несколько не умаляет значимость исследований, поскольку речь идет о создании метода, в принципе пригодного для классификации таких объектов, как наборы показателей крови. Кроме того, в целях «чистоты эксперимента» рассматривается распознавание элементов самих обучающих множеств.

Имеется очевидный порядок расположения градаций СЗЧ от здорового до максимальной степени поражения организма. Однако при решении поставленной задачи классификации в полученном перечне альтернатив он может нарушаться. Увеличение объема и полноты обучающих множеств в дальнейшем позволит решить и эту проблему.

Используемые в работе базы показателей крови обладают значительной вариабельностью

и небольшим объемом. Были разработаны и реализованы новые подходы в организации обучающей последовательности.

Достигнутая точность распознавания на обучающих множествах для различных рассматриваемых СО находится в диапазоне 95 – 100 %. Эти результаты соответствуют данным, полученным при помощи алгебраического подхода Журавлева [1].

Литература

1. Количественная оценка гомеостатической активности здоровых и больных людей / Ставицкий Р.В. [и др.]. // М.: ГАРТ. 2013. 131 с.

Пестрякова Надежда Владимировна. Ведущий научный сотрудник ФИЦ ИУ РАН. Окончила МФТИ в 1983 году. Доктор технических наук. Количество печатных работ: более 90, в том числе 1 монография. Область научных интересов: вычислительная математика и физика, распознавание образов. E-mail: pestryakova@isa.ru

Гавриков Борис Михайлович. Аспирант ФИЦ ИУ РАН. Окончил МИФИ в 2014 году. Количество печатных работ: 12, в том числе 3 монографии. Область научных интересов: вычислительная математика, распознавание образов, медицинская физика. E-mail: bmgavrikov@gmail.com

About building the feature space in the problem of learning

B.M. Gavrikov, N. V. Pestryakova

The problem of state estimation of human health based on the results of the analysis of peripheral blood. Studies were conducted on four grades of lesions of the body systems separately for men and women. Used statistical classification method based on polynomial regression approach, and with probabilistic estimation. In developed and trained version of the classifier implemented complete sets of combinations of monomials up to the fifth degree.

Key words: state of human health, peripheral blood, classification, polynomial regression

References

1. Stavitskii, R.V. [et al.]. 2013. Kolichestvennaya otsenka gomeostaticheskoy aktivnosti zdorovih i bol'nih lyudey. [Quantitative estimation of homeostatic activity in healthy and sick people]. Moscow, GART. 131p.
2. Gavrikov, M.B., and N.V. Pestryakova 2004. Keldysh Institute Preprints. № 22. Metod polinomial'noy regressii v zadachah raspoznavaniya pechatnih i rukopechatnih simvolov. [Polynomial regression method in pattern recognition of printed and handprinted characters]. 12p.
3. Gavrikov, M. B. [et al.]. Automation and Remote Control. –2006,. On a Pattern Recognition Method Based on Polynomial Regression // 67(2). p. 278-292. DOI: 10.1134/S000511790602007X.
4. Gavrikov, B.M., M.B.Gavrikov, and N. V.Pestryakova. 2015. ITVS. Statisticheskii analiz harakteristik metoda raspoznavaniya na obuchayushem mnojestve. [A statistical analysis of the recognition method when recognizing a training set]. 2:50-67.
5. Gavrikov, B.M. [et al.]. 2016. Trudi ISA RAN. Ob odnom statisticheskom metode otsenivaniya sostoyaniya zdorov'ya che-loveka. [One statistical method of evaluating the state of human health]. 66(2):54-59.

Pestryakova N.V. laboratory of artificial intelligence methods ISA FRC CSC RAS, E-mail: pestryakova@isa.ru

Gavrikov B. M. post-graduate courses ISA FRC CSC RAS, E-mail: bmgavrikov@gmail.com