

### Заклучение

Проводя БОС-тренинг в несколько сеансов (количество сеансов индивидуально) можно выработать такие психофизиологические качества водителей как скорость и точность сенсомоторных реакций, внимание, мышление, память, способность быстро воспринимать сложившуюся ситуацию, умение оценивать временные интервалы. Стабилометрическую платформу D-1 можно использовать для выработки пороговых критериев профессионально важных психофизиологических качеств водителей транспортных средств.

## STABILOMETRIC METHODS OF EVALUATION AND DEVELOPMENT OF PROFESSIONALLY IMPORTANT PSYCHOPHYSIOLOGICAL QUALITIES OF DRIVERS

N.V. SHCHERBINA

### Abstract

It is proposed to use biofeedback training to develop professionally important psychophysiological qualities of the drivers of vehicles.

*Keywords:* stabilometry, biofeedback, training.

### Список литературы

1. Савченко В.В., Свистун М.С., Сикорский В.В. // Автомобильная промышленность. 2008. № 1. С. 32–34.
2. Савченко В.В. // Системный анализ и управление в биомедицинских системах. 2006. Т. 5, № 1. С. 187–191.
3. Романов А.Н. Автотранспортная психология. М., 2002.
4. Дубовский В.А. // Медицинская техника. 2011. № 2. С. 14–17.
5. Дубовский В.А. Устройство для тренировки двигательных функций человека / Патент РБ № 8397.
6. Щербина Н.В., Ковалевич О.В., Савченко В.В. и др. // Сб. матер. II Междунар. науч.-практ. конференции «BIG DATA and Advanced Analytics. Использование BIG DATA для оптимизации бизнеса и информационных технологий». Минск, 2016. С. 256–263.
7. Дубовский В.А. // Докл. БГУИР. 2016. № 4. С. 67–72.

УДК 51.761.615

## ПРЕОБРАЗОВАНИЕ КЛАСТЕРОВ КЛИНИЧЕСКИХ ДАННЫХ НА ЭТАПЕ ПЕРВИЧНОЙ ДИАГНОСТИКИ

А.И. МИТЮХИН, Е.Н. МАЙСЕНЯ

*Институт информационных технологий  
Белорусского государственного университета информатики и радиоэлектроники  
ул. Козлова, 28, Минск, 220037, Беларусь*

*Минский городской клинический онкологический диспансер,  
пр. Независимости, 64, Минск, 220013, Беларусь*

*Поступила в редакцию 10 ноября 2016*

Рассмотрен метод кластеризации биомедицинских данных, позволяющий повысить эффективность первичной диагностики. Процесс диагностики реализуется на основе методологии распознавания диагностических признаков. Упрощение процесса распознавания выполняется посредством преобразования пространства исходных кластеров. Приведен вычислительный алгоритм кластеризации и классификации признаков.

*Ключевые слова:* диагностика, клинические данные, распознавание, информативные признаки, ковариация, корреляция, пространство кластеров.

### Введение

Получение определенного диагностического решения основывается на анализе диагностических данных. По мере увеличения количества клинических данных принятие достоверного диагностического решения врачом усложняется и может потребовать значительных временных затрат. Медико-биологическое исследование, основанное на специальном математическом описании и анализе, совместно с цифровыми технологиями становится все более востребованным в современной клинической практике, способствует эффективному установлению диагноза или подтверждению его. При этом обеспечивается более высокая степень достоверности результата.

В работе рассматривается метод, описывающий получение первичного диагностического решения с использованием аппарата корреляционного анализа данных. Ввиду того, что ответственность за правильность диагноза остается на враче, проводящем обследование, положительный результат первичной (начальной) диагностики указывает врачу на необходимость проведения более глубоких и детальных обследований, получения другой дополнительной клинически важной информации. Эффективное проведение первичной диагностики способствует более раннему выявлению патологических процессов, выбору адекватного лечения заболевания.

### Теоретические принципы

Диагностическое решение с прогнозируемой точностью можно получить с использованием методологии распознавания образов. Процесс распознавания включает в себя такие этапы как: 1) кластеризация; 2) выделение наиболее информативных признаков; 3) процедура обучения; 4) классификация с помощью заранее накопленных эталонных данных; 5) анализ текущей информации пациентов. Рассмотрим составляющие процесса распознавания.

В практических задачах диагностические данные удобно представлять в виде вектора  $\mathbf{g}_k = (g_{1k} \ g_{2k} \ \dots \ g_{Nk})$  признаков (образа). Здесь  $N$  обозначает число признаков,  $k$  – это номер класса  $\omega_k$  на множестве  $\{\omega_1, \omega_2, \dots, \omega_w\}$  классов. Числовые значения координат (или другие представления данных) вектора-образа соответствуют диагностическим признакам пациента. Предполагается, что все векторы пространственно совмещены в момент регистрации признаков пациентов. Набор этих векторов запишем в виде матрицы  $\mathbf{G} = (\mathbf{g}_1 \ \dots \ \mathbf{g}_M)^T$  размером  $N \times M$ . Если рассматривать процедуру распознавания с учителем, необходимо иметь набор из  $w$  априорно заданных матриц признаков. Структуры матриц определяют размер и форму кластеров.

Отбор наиболее информативных признаков связан с процессом минимизации признакового пространства через преобразование кластеров. В результате такого преобразования максимизируются расстояния между классами и минимизируются внутриклассовые расстояния. Степень уменьшения размерности (размеров кластеров) зависит от уровня статистической зависимости данных. Количественная оценка статистической зависимости выполняется расчетом ковариационных и корреляционных связей [1]. Ковариационная матрица  $\text{cov}(\mathbf{g})$  для признаков  $\mathbf{G}$  определяется выражением

$$\text{cov}(\mathbf{g}) = E\{(\mathbf{g} - \mathbf{m}_g)(\mathbf{g} - \mathbf{m}_g)^T\}, \quad (1)$$

где  $E$  – оператор математического ожидания,  $\mathbf{m}_g = E\{\mathbf{G}\} = (m_1, \dots, m_N)^T$  – средний вектор.

Матрица описывает внутриклассовые и межклассовые диагностические признаки через значение  $\sigma_{ii}^2$  – дисперсии  $i$ -ой координаты случайного вектора  $\mathbf{g}$  и значение  $\sigma_{ij}^2$  – ковариации  $i$ -ой и  $j$ -ой координаты вектора  $\mathbf{g}$ . Коррелированность данных определяется коэффициентом

$$c_{ij} = \frac{\sigma_{ij}^2}{\sqrt{\sigma_{ii}^2 \sigma_{jj}^2}}. \tag{2}$$

В обследуемых случаях реальные диагностические параметры характеризуются изменчивостью от одного пациента к другому. Коэффициенты корреляции (2) могут принимать значения в диапазоне  $c_{ij} \in [-1; 1]$ . Характерные же признаки имеют значения коэффициентов корреляции  $c_{ij} \rightarrow 1$ , что свидетельствует о наличии межкомпонентных связей в векторах матрицы  $\mathbf{G}$ . Диагностическая информация становится в целом избыточной. Упрощение принятия диагностического решения требует устранения или уменьшения избыточности анализируемых данных, что эквивалентно операции понижения размерности диагностического пространства. Подобные рассуждения справедливы как нормальных, так и для аномальных данных обследуемых пациентов.

В общем подходе устранение межкомпонентных связей состоит в обработке коррелированных данных посредством линейного обратимого преобразования [2] вида

$$\hat{\mathbf{g}} = \mathbf{A} \mathbf{g}, \tag{3}$$

где ортогональная матрица  $\mathbf{A}$  – ядро преобразования. Результат преобразования определяется свойствами ядром  $\mathbf{A}$ . Полная декорреляции исходных данных с целью устранения избыточности достигается, если использовать преобразование, которое учитывает утверждение теоремы [3].

*Теорема.* Ковариационная матрица  $\text{cov}(\hat{\mathbf{g}})$  в области преобразований вектора  $\mathbf{g}$  будет иметь диагональный вид, если матрицу  $\mathbf{A}$  составить из транспонированных собственных векторов ковариационной матрицы  $\text{cov}(\mathbf{g})$ . Таким образом, для ковариационной матрицы  $\text{cov}(\mathbf{g})$  в области исходных данных существует матрица  $\mathbf{A}$  размером  $n \times n$ , составленная из собственных векторов матрицы  $\text{cov}(\mathbf{g})$ . Преобразование коррелированных данных с помощью ядра, составленного из собственных векторов, можно записать как

$$\hat{\mathbf{g}}_k = \sum_{i=1}^N g_{ik} \mathbf{A}_i, k = 1, 2, \dots, M, \tag{4}$$

где  $k$  обозначает номер коэффициента преобразования исходного вектора  $\mathbf{g}_k = (g_{1k} \ g_{2k} \ \dots \ g_{Nk})$ ,  $\mathbf{A}_i$  – вектор-столбец матрицы  $\mathbf{A}$ .

Для минимизация признакового пространства путем преобразования  $n$ -мерного пространства в пространство меньшей размерности вместо использования всех собственных векторов матрицы  $\mathbf{A}$  применяется ядро  $\mathbf{A}_l$  меньшей размерностью,  $(n \times l), l \leq n$ . Матрица  $\mathbf{A}_l$  состоит из  $l$  собственных векторов, которым отвечают  $l$  наибольших собственных чисел  $\lambda$  матрицы  $\mathbf{A}$ . Наиболее информативные признаки коэффициентов  $\hat{\mathbf{g}}_k$  соответствуют дисперсиям  $\sigma_{\hat{\mathbf{g}}}^2 = \lambda$  коэффициентов преобразования (4).

Эффективность описания пространства признаков определяется выбором сохраняемого множества коэффициентов  $\hat{\mathbf{g}}_k$ . Критерием при выборе множества коэффициентов является сохранение  $l$  коэффициентов с наибольшими дисперсиями  $\max \sigma_{\hat{\mathbf{g}}}^2$  матрицы  $\text{cov}(\hat{\mathbf{g}})$ . Вычисление собственных векторов и, соответствующих им собственных чисел, матрицы  $\text{cov}(\mathbf{g})$ , осуществляется путем решения характеристического полинома ковариационной матрицы (можно найти в [4]). Далее рассмотрим пример автоматической классификации нормальных и аномальных клинических признаков.

### Вычислительные этапы

Не теряя общего представления о  $n$ - мерных векторах диагностических признаков  $\omega$  классов, ограничимся решением задачи классификации 3-D векторов образов принадлежащих к

двум классам:  $\omega_1$  – нормальные образы и  $\omega_2$  – аномальные образы. Обучающие множества векторов признаков записаны в виде матриц  $G_1$  и  $G_2$ :

$$G_1 = \begin{pmatrix} 4 & 3 & 4 & 3 & 5 & 3 \\ 1 & 2 & 1 & 1 & 1 & 2 \\ -1 & -2 & 0 & -1 & 0 & -2 \end{pmatrix}; G_2 = \begin{pmatrix} 6 & 6 & 6 & 7 & 7 & 6 \\ 3 & 3 & 4 & 3 & 4 & 4 \\ 1 & 2 & 3 & 2 & 2 & 3 \end{pmatrix}.$$

На вход классификатора подается вектор признаков  $g = (g_1 \ g_2 \ g_3)^T$  с неизвестной классификацией. Необходимо определить его классификацию.

*Первый этап. Выбор исходных признаков*

1. Ковариационные матрицы, соответствующие классам  $\omega_1$  и  $\omega_2$ , определяются из выражения (1).

$$\text{cov}(G_1) = \begin{pmatrix} 0,6667 & -0,2667 & 0,6000 \\ -0,2667 & 0,2667 & -0,4000 \\ 0,6000 & -0,4000 & -0,8000 \end{pmatrix}, \text{cov}(G_2) = \begin{pmatrix} 0,2667 & 0 & -0,0667 \\ 0 & 0,3000 & 0,3000 \\ -0,0667 & 0,3000 & 0,5667 \end{pmatrix}.$$

Как видно, значительный уровень коэффициентов ковариации  $\sigma^2_{ij}$  компонент  $g_1, g_2, g_3$  на всем множестве векторов позволяет реализовать процесс декорреляции признаков выполнением преобразования (4). Ядро этого преобразования строится на основе знания статистических характеристик (ковариационных матриц) признаков разных классов. Тогда общая ковариационная матрица, описывающая внутриклассовые и межклассовые диагностические признаки, вычисляется как

$$\text{cov}(G) = p_1 \text{cov}(G_1) + p_2 \text{cov}(G_2), \tag{5}$$

где  $p_1 = p_2 = 0,5$  – априорное значение вероятности признаков классов.

$$\text{cov}(G) = \begin{pmatrix} 0,9333 & -0,2667 & 0,5333 \\ -0,2667 & 0,5667 & -0,1000 \\ 0,5333 & -0,1000 & 1,3667 \end{pmatrix}.$$

2. Решение характеристического полинома, соответствующего матрице (5), приводит к нахождению следующих собственных значений:  $\lambda_1 = 0,3834; \lambda_2 = 0,7121$  и  $\lambda_3 = 1,7712$ . Основная информация о классификации диагностических признаков содержится в третьем и втором собственном числе. Тогда практически без потери диагностической информации можно обеспечить эффективное представление данных в области преобразований (4). Матрица собственных векторов, соответствующих  $\lambda_1, \lambda_2$  и  $\lambda_3$ , имеет вид

$$A = \begin{pmatrix} 0,6108 & -0,5497 & 0,5698 \\ 0,7495 & 0,6333 & 0,1925 \\ -0,2551 & 0,5447 & -0,7989 \end{pmatrix}. \tag{6}$$

*Выбор признаков в области преобразований*

На этом этапе рассматривается процедура минимизации пространства кластеров. Для этого строится ядро преобразования  $A_l$ , состоящее лишь из двух собственных векторов, а вектор  $\hat{g}_k$  будет содержать 2 ненулевые координаты в области преобразований. Применяя усечение матрицы (6) и используя выражение  $\hat{g}_l = A_l g$ , получим следующие матрицы обучающих признаков в области преобразований для множеств  $G_1$  и  $G_2$ :

$$\hat{g}_{1l} = \begin{pmatrix} 3,8238 & 3,9001 & 3,0743 & 3,6313 & 4,3808 & 3,9001 \\ -1,2746 & -1,2737 & -1,0195 & -0,4757 & -0,7308 & -1,2737 \end{pmatrix}; \tag{7}$$

$$\hat{g}_{2l} = \begin{pmatrix} 6,2044 & 6,0119 & 6,4527 & 6,7614 & 7,2022 & 6,4527 \\ 0,9024 & 1,7013 & 3,0449 & 1,4462 & 2,7898 & 3,0449 \end{pmatrix}. \tag{8}$$

Таким образом, устранение избыточности за счет минимизацию внутриклассовой энтропии признаков позволило преобразовать, уменьшить пространственный размер кластеров.

*Второй этап. Классификация.* Решение задачи классификации сводится к определению оптимальных разделяющих границ кластеров или в разработке процедур отнесения признаков к различным классам образов. Пусть в качестве обучающих множеств классификатора выступают вектора признаков  $\hat{\mathbf{g}}_{lk} = (\hat{g}_{1k} \ \hat{g}_{2k} \ \dots \ \hat{g}_{lk})^T$ , где  $l$  обозначает число признаков,  $k$  – номер класса на множестве  $\{\omega_1, \omega_2, \dots, \omega_w\}$ . К классу  $\omega_1$  и  $\omega_2$  относятся векторы-столбцы матриц (7) и (8) соответственно. Если образы  $\{\hat{\mathbf{g}}_k\}$  располагаются в двумерном пространстве признаков и кластеры не пересекаются, разделяющая граница определяется как [5]

$$\|\bar{\mathbf{g}}_1\|^2 - \|\bar{\mathbf{g}}_2\|^2 = 2(\bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_2)^T \hat{\mathbf{g}}, \quad (9)$$

где  $\bar{\mathbf{g}}_1$ ,  $\bar{\mathbf{g}}_2$  средние значения векторов образов классов  $\omega_1$  и  $\omega_2$  соответственно.

Процесс обучения классификатора распознаванию образов происходит после получения функции, описывающей разделение кластеров. Подставляя в выражение (9) значения координат векторов  $\bar{\mathbf{g}}_1$ ,  $\bar{\mathbf{g}}_2$ ,  $\hat{\mathbf{g}}$ , получаем линейную разделяющую (дискриминантную) функцию  $C(\hat{\mathbf{g}})$  классификатора вида  $C(\hat{\mathbf{g}}) = 2,7291 \hat{g}_1 + 3,1629 \hat{g}_2 - 15,8679$ .

Коэффициенты дискриминантной функции задают параметры и порог классификатора. Автоматическая классификация признака  $\hat{\mathbf{g}}_k$  выполняется на основе решающего правила классификатора. Принятия решения о классе образа с неизвестной классификацией реализуется с помощью контрольного множества  $\{\tilde{\mathbf{g}}_k\}$  из испытательных образов, точная классификация которых известна. Вычисления значений  $C(\hat{\mathbf{g}})$  для испытательного множества  $\{\tilde{\mathbf{g}}_k\}$  привели к следующим решающим правилам работы автоматического классификатора: если  $C(\hat{\mathbf{g}}) < 0$  то  $\mathbf{g} \in \omega_1$ ; если  $C(\hat{\mathbf{g}}) > 0$ , то  $\mathbf{g} \in \omega_2$ .

### Заключение

Преобразование кластеров клинических данных на основе вычисления распределения дисперсий диагностических признаков позволяет уменьшить размерность анализируемых данных, упростить решающее правило работы классификатора практически без потери диагностической информации. Предлагаемый подход может оказаться особенно эффективным на этапе первичной диагностики заболеваний и для текущего контроля состояния тяжелобольных пациентов.

## TRANSFORMATION OF CLINICAL DATA CLUSTERS AT THE STAGE OF INITIAL DIAGNOSIS

A.I. MITSUKHN, E.N. MAISENIA

### Abstract

A clustering method of biomedical data, which allows to increase the efficiency of the primary diagnosis is considered. Diagnostic process is realized on the basis of diagnostic signs recognition methodology. Simplification of the recognition process is performed by converting the space of initial clusters. It is shown a computational algorithm of clustering and classification characteristics.

*Keywords:* diagnosis, clinical data, recognition, informative signs, covariance, correlation, cluster space.

### Список литературы

1. Митюхин А.И. Цифровая обработка речи и анализ изображений. Минск, 2016.
2. Mitsiukhin A.I. // Proceedings 58 International Scientific Colloquium. DE, TU Ilmenau, 2014. ID 2066.
3. Стренг Г. Линейная алгебра и ее применения. М., 1980.
4. Мэтьюз Дж., Финк К. Численные методы. Использование МАТЛАБ. М., 2001.
5. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М., 2005.

УДК 620.3; 6:531-022.532; 612.821; 612.821.8; 681.3.07; 621.383

## МИКРОФЛЮИДНАЯ ЯЧЕЙКА ДЛЯ МАТРИЦ ОПТИЧЕСКИХ РЕЗОНАТОРНЫХ СЕНСОРНЫХ СИСТЕМ

А.В. САЕЧНИКОВ<sup>1,2</sup>, Э.А. ЧЕРНЯВСКАЯ<sup>1</sup>, В.А. САЕЧНИКОВ<sup>1</sup>, А. ОСТЕНДОРФ<sup>2</sup>

<sup>1</sup>Белорусский государственный университет  
пр. Независимости, 4, Минск, 220030, Республика Беларусь.

<sup>2</sup>Ruhr Universität Bochum  
Universitätsstraße, 150, Bochum, 44801, Germany

Поступила в редакцию 11 ноября 2016

Разработана микрофлюидная ячейка, содержащая матрицу распределенных сферических оптических сенсоров. Моделирование динамики потока жидкости показало, что ячейка обеспечивает высокую степень гомогенности скорости движения потока исследуемого биомедицинского раствора. Проведены тестовые эксперименты по детектированию изменений биологической среды внутри ячейки.

*Ключевые слова:* моды шепчущих галерей, микрофлюидная ячейка, оптический резонанс, сенсор, вычислительная гидродинамика.

### Введение

Последнее время оптические сенсорные системы, основанные на резонансе мод шепчущих галерей (WGM), получили большое распространение для детектирования биологических соединений вплоть до отдельных молекул и генов [1]. Для выполнения анализа многокомпонентных соединений была разработана матрица чувствительных оптических сенсоров [2] с применением аддитивных технологий. В дальнейшем была предложена соответствующая методика индивидуализации сенсоров для обеспечения выборочного взаимодействия с компонентами сложных соединений [3].

Известно, что область чувствительности WGM сенсора определяется границами распространения световой волны внутри кольцевого резонатора [4]. Таким образом, сенсор чувствителен только к тем биологическим компонентам, которые находятся вблизи данной области. Поэтому должна быть обеспечена контролируемая водная среда вокруг чувствительной ячейки матрицы. Это позволит проводить количественный анализ/сравнение спектральных откликов от различных сенсоров матрицы.

Простое размещение водного биологического раствора на матрице резонаторов может привести к следующим неоднозначностям последующего анализа данных:

- трудно обеспечить хорошую обтекаемость раствора вокруг сенсора, находящегося внутри матричной структуры;
- временная неопределенность концентрации биологического агента из-за испарения водной основы;