УДК 519.685.3

Н.Ф. КИРИЧЕНКО, Ю.Г. КРИВОНОС, Н.П. ЛЕПЕХА

ОПТИМИЗАЦИЯ СИНТЕЗА ГИПЕРПЛОСКОСТНЫХ КЛАСТЕРОВ И НЕЙРОФУНКЦИОНАЛЬНЫХ ПРЕОБРАЗОВАНИЙ В СИСТЕМАХ КЛАССИФИКАЦИИ СИГНАЛОВ

Ключевые слова: *кластеризация, синтез систем классификации, распознавание образов, псевдообратные и проекционные матрицы*.

ВВЕДЕНИЕ

Статья является логическим продолжением исследований задач кластеризации и классификации сигналов, изложенных в работах [1, 2]. Для исследования этих задач, как и ранее, используются известные [3–5] и ранее полученные авторами свойства псевдообратных и проекционных матриц [6–8]. Цель данной работы состоит в демонстрации эффективности использования этого математического аппарата при решении важных задач цифровой обработки информации. А именно: для задач кластеризации множества точек в пространстве признаков получен оптимизационный алгоритм синтеза гиперплоскостных кластеров, предлагается алгоритм решения задачи линейной отделимости конечного множества точек, определен критерий линейной полосной разделимости точек в пространстве признаков на два класса, определена оптимальная по толщине нелинейная полоса разделимости точек на два класса, разработан алгоритм определения оптимального нелинейного преобразования координаты вектора признаков в заданных классах функций и ее индекса. С точки зрения авторов, предлагаемые математические средства имеют перспективу существенно улучшить решение прикладных задач распознавания речевых сигналов, человеческих лиц, рукописных символов и текстов, ультразвуковой информации, экономических ситуаций, других сигналов и событий.

1. СИНТЕЗ ГИПЕРПЛОСКОСТНЫХ КЛАСТЕРОВ СРЕДСТВАМИ ОПТИМИЗАЦИИ

1.1. Краткий обзор средств кластеризации. При большом разнообразии различных методов кластеризации (краткий обзор приведен в [9]) остановимся на нескольких приемлемых для использования в нашей задаче методах кластеризации в пространстве признаков.

Довольно часто используется алгоритм кластеризации K-средних (K-means method) [10], в котором формируется некоторая целевая функция $\Phi(\cdot)$, выражающая качество текущего разбиения множества точек $x(j)$, $j=\overline{1,n}$, на K кластеров с центрами в точках c_i, $i=\overline{1,K}$, K — задано. В начальный момент центры кластеров выбираются произвольно, далее для каждой точки множества итеративно определяем его принадлежность к одному из K кластеров и вычисляем новые значения $\Phi(\cdot)$ для центров кластеров, стремясь при этом минимизировать функцию $\Phi(\cdot)$. В литературе можно найти различные его модификации [10].

Одним из классических считается метод сегментации информации, основанный на разбиении и слиянии кластеров (split and merge methods) [11]. В начальный момент единственным регионом является все множество точек $x(j)$, $j=\overline{1,n}$, и на каждой итерации метода производится расщепление каждого региона R_i, для которого не выполняется условие $\mathrm{P}(R_i)$ принадлежности точек к i-му региону, на некоторое количество подрегионов. Далее сливаются те смежные регионы R_i и R_k, для объединения которых $\mathrm{P}(R_i \cup R_k)$ это условие выполняется. Процесс заканчивается, если существующие регионы невозможно ни расщепить, ни объединить. В качестве логического предиката $P(\cdot)$ можно использовать, например, критерий K-means метода, либо строить различные меры подобия [10].

Среди множества других публикаций по этой проблеме следует отметить фундаментальную работу [12], в которой дается детальный анализ и обзор различных методов, в том числе класс методов последовательного формирования кластеризирующих подпространств (The Subspaces Methods of Classification), которые в определенном смысле идейно близки к предлагаемому ниже авторами методу кластеризации.

1.2. Постановка задачи синтеза гиперплоскостных кластеров средствами оптимизации и алгоритм ее решения. Рассмотрим задачу синтеза гиперплоскостных кластеров, используя ранее полученные результаты по псевдообратным и проекционным матрицам, а также свойства сингулярного разложения матриц [4, 5].

Для множества точек в пространстве признаков $x(j) \in R^m$, $j=\overline{1,n}$, рассмотрим задачу разделения этого множества на два кластера (подмножества) таким образом, чтобы точки первого кластера были расположены ближе к некоторой гиперплоскости, порождающей свой класс, а точки второго кластера, в свою очередь, ближе к иной гиперплоскости и соответственно порождающей этот класс. Особенностью постановки рассматриваемой задачи является то обстоятельство, что заранее ни сами эти подмножества, ни порождающие гиперплоскости неизвестны.

Для формулирования этой задачи в математическом смысле и построения алгоритма ее решения введем в рассмотрение расстояние $\rho(x(j), L(A,b))$ от точки $x(j) \in R^m$ до некоторой гиперплоскости

$$L(A,b) = \{x : x = A^+ b + Z(A)w, \forall w \in R^m\}, \tag{1}$$

определяемой как множество решений (псевдорешений) системы алгебраических уравнений

$$Ax = b, A \in R^{s \times m}, b \in R^s, \tag{2}$$

где A и b — соответственно матричный и векторный параметры рассматриваемой гиперплоскости, а $Z(A) = I_m - A^+ A$. Из определения псевдообратной матрицы A^+

$$\forall b \in R^s, \ A^+ b = \arg \min_{\operatorname{Arg} \min\limits_{x \in R^m} \|Ax-b\|^2} \|x\|^2$$

расстояние $\rho(x(j), L(A,b))$ от точки $x(j)$ до гиперплоскости $L(A,b)$ определяется согласно соотношению

$$\rho^2(x(j), L(A,b)) = \| A^+ (b - Ax(j))\|^2, \tag{3}$$

где $\|x\|^2 = x^{\mathrm{T}} x$.

Тогда справедливо утверждение следующей леммы.

Лемма 1. Для множества точек $x(j)$, $j=\overline{1,n}$, сумма квадратов их расстояний до гиперплоскости $L(A,b)$

$$\rho^2(\{x : x(j), j=\overline{1,n}\}, L(A,b)) = \sum_{j=1}^{n} \rho^2(x(j), L(A,b)) \tag{4}$$

принимает значение

$$\rho^2(\{x : x(j), j=\overline{1,n}\}, L(A,b)) = \sum_{j=1}^{n} (b - Ax(j))^{\mathrm{T}} R(A^{\mathrm{T}})(b - Ax(j)) =$$
$$= \operatorname{tr} R(A^{\mathrm{T}}) \sum_{j=1}^{n} (b - Ax(j))(b - Ax(j))^{\mathrm{T}}, \tag{5}$$

где $R(A^{\mathrm{T}}) = (A^+)^{\mathrm{T}} A^+$ — взвешенная проекционная матрица [7].

Используя свойства псевдообратных матриц, несложно осуществить доказательство леммы об оптимальном выборе вектора b.

Лемма 2. При заданных значениях $A, x(j), j=\overline{1,n}$, имеет место соотношение

$$b_{\text{opt}} = A\hat{x} = \arg\min_{b \in R^s} \rho^2(\{x\colon x(j),\, j=\overline{1,n}\}, L(A,b)),\ \hat{x} = \frac{1}{n}\sum_{j=1}^{n} x(j). \tag{6}$$

Доказательство. На основании соотношения (5) и в силу необходимого условия оптимальности

$$\sum_{j=1}^{n} (A^+)^{\mathrm{T}} A^+ (b - Ax(j)) = 0 \tag{7}$$

для искомого оптимального значения вектора b получаем систему уравнений

$$(A^+)^{\mathrm{T}} A^+ b = (A^+)^{\mathrm{T}} \left(\frac{1}{n} \sum_{j=1}^{n} x(j) \right), \tag{8}$$

которая является совместной и существование решения которой вытекает из очевидного соотношения

$$\left\| Z((A^+)^{\mathrm{T}} A^+)\, (A^+)^{\mathrm{T}} \sum_{j=1}^{n} x(j) \right\|^2 = 0.$$

Поскольку $((A^+)^{\mathrm{T}} A^+)^+ = AA^{\mathrm{T}}$, то

$$b_{\text{opt}}(A) = AA^{\mathrm{T}} (A^+)^{\mathrm{T}} \left(\frac{1}{n} \sum_{j=1}^{n} x(j) \right) = A \left(\frac{1}{n} \sum_{j=1}^{n} x(j) \right). \tag{9}$$

Теорема. Расстояние $\rho(\{x : x(j),\, j=\overline{1,n}\}, L(A,b))$ множества точек $x(j),\ j=\overline{1,n}$, до гиперплоскости $L(A,b)$ при оптимальном выборе вектора b согласно формуле (6) имеет значение

$$\rho(\{x\colon x(j),\, j=\overline{1,n}\}, L(A, b_{\text{opt}}(A))) = (\operatorname{tr} A^+ A \cdot \widetilde{X}\widetilde{X}^{\mathrm{T}})^{1/2}, \tag{10}$$

где $\widetilde{X} = (\widetilde{x}(1) \vdots \ldots \vdots \widetilde{x}(n)),\ \widetilde{x}(j) = x(j) - \hat{x},\ j=\overline{1,n}$.

Доказательство. Так как имеют место соотношения

$$\rho^2(\{x : x(j),\, j=\overline{1,n}\}, L(A, b_{\text{opt}})) =$$

$$= \sum_{j=1}^{n} \left(\frac{1}{n} \sum_{i=1}^{n} Ax(i) - Ax(j) \right)^{\mathrm{T}} (A^+)^{\mathrm{T}} A^+ \left(\frac{1}{n} \sum_{i=1}^{n} Ax(i) - Ax(j) \right) =$$

$$= \sum_{j=1}^{n} (\hat{x} - x(j))^{\mathrm{T}} A^{\mathrm{T}} (A^+)^{\mathrm{T}} A^+ A (\hat{x} - x(j)) = \sum_{j=1}^{n} (\hat{x} - x(j))^{\mathrm{T}} A^+ A (\hat{x} - x(j)) =$$

$$= \sum_{j=1}^{n} \widetilde{x}^{\mathrm{T}}(j) A^+ A \widetilde{x}(j) = \operatorname{tr} A^+ A \cdot \sum_{j=1}^{n} \widetilde{x}(j) \widetilde{x}^{\mathrm{T}}(j) = \operatorname{tr} A^+ A \widetilde{X} \widetilde{X}^{\mathrm{T}},$$

то справедливость теоремы не вызывает сомнения.

Средство оптимального выбора матрицы A получим как следствие из утверждения предыдущей теоремы. А именно, используя известное сингулярное разложение для матрицы $\widetilde{X}$

$$\widetilde{X} = \sum_{i=1}^{r} u_i v_i^{\mathrm{T}} \lambda_i,\ u_i^{\mathrm{T}} u_j = \delta_{ij},\ v_i^{\mathrm{T}} v_j = \delta_{ij},\ \lambda_1 \geq \ldots \geq \lambda_r > 0,\ r = \operatorname{rank} \widetilde{X},$$

для следующего представления расстояния множества точек $x(j),\ j=\overline{1,n}$, до $L(A, b_{\text{opt}})$

$$\rho(\{x: x(j),\ j=\overline{1,n}\}, L(A, b_{\text{opt}}(A))) = \left(\operatorname{tr} A^{+} A \cdot \sum_{j=1}^{r} u_j u_j^{\mathrm{T}}(j) \lambda_j^2)\right)^{1/2},$$

получаем средство оптимального выбора матрицы A.

Следствие. Оптимальный выбор матрицы $A \in R^{s\times m}$ определяется как решение оптимизационной задачи

$$A_{\text{opt}} = \arg \min_{AA^{\mathrm{T}}=I_s,\ A\in R^{s\times m}} \rho^2(\{x: x(j),\ j=\overline{1,n}\}, L(A, b_{\text{opt}}(A))) = \begin{pmatrix} u_{m-s+1}^{\mathrm{T}} \\ \vdots \\ u_m^{\mathrm{T}} \end{pmatrix}, \quad (11)$$

при этом $\operatorname{tr} A_{\text{opt}}^{+} A_{\text{opt}} \widetilde{X} \widetilde{X}^{\mathrm{T}} = \sum_{j=m-s+1}^{r} \lambda_j^2$, $(u_1, \ldots, u_m)^{\mathrm{T}} (u_1, \ldots, u_m) = I_m$.

На основании приведенных выше лемм, теоремы и следствия алгоритм синтеза гиперплоскостных кластеров представляется аналогично алгоритму K-средних:

1) создать начальное разбиение исходного множества $\Omega_x = \{x: x(j),\ j=\overline{1,n}\}$ на два подмножества:

$$\Omega_x = \Omega_x(1) \cup \Omega_x(2),\ \Omega_x(1) \cap \Omega_x(2) = \varnothing;$$

2) вычислить оптимальные $A_{\text{opt}}(1)$, $b_{\text{opt}}(1)$ для $\Omega_x(1)$ и $A_{\text{opt}}(2)$, $b_{\text{opt}}(2)$ — для $\Omega_x(2)$;

3) обновить $\Omega_x(1)$ и $\Omega_x(2)$ по расстоянию $(b_{\text{opt}} - A_{\text{opt}} x(j))^{\mathrm{T}} R(A_{\text{opt}}^{\mathrm{T}}) \times$ $\times (b_{\text{opt}} - A_{\text{opt}} x(j))$, сформировать новые подмножества $\Omega_x(1)$, $\Omega_x(2)$ из элементов множества Ω_x согласно условиям

$$\| A_1^{+}(b_1 - A_1 x(j)) \|^2 \le \| A_2^{+}(b_2 - A_2 x(j)) \|^2,$$

$$\| A_2^{+}(b_2 - A_2 x(j)) \|^2 < \| A_1^{+}(b_1 - A_1 x(j)) \|^2;$$

4) перейти на выполнение операции 2 с новыми подмножествами $\Omega_x(1), \Omega_x(2)$;

5) прекратить выполнение операций при не обновлении в новом цикле значений A_1, b_1 и A_2 b_2.

2. СИНТЕЗ КЛАССИФИКАТОРОВ СРЕДСТВАМИ ЛИНЕЙНЫХ И НЕЛИНЕЙНЫХ ОПТИМАЛЬНЫХ ПРЕОБРАЗОВАНИЙ ПРОСТРАНСТВА ПРИЗНАКОВ

Для рассматриваемых задач распознавания принадлежности исследуемых объектов к тому или иному классу будем предполагать известным, к какому из классов относятся векторы обучающей выборки $x(j)$, $j=\overline{1,n}$, в пространстве признаков. Тогда целесообразно рассмотреть последовательно такие постановки задач:

— определить необходимые и достаточные условия линейной отделимости конечного множества точек в многомерном пространстве;

— определить необходимые и достаточные условия линейной полосной разделимости точек $x(j)$, $j=\overline{1,n}$, на два класса;

— определить оптимальную по толщине нелинейную полосу разделимости точек $x(j)$, $j=\overline{1,n}$, на два класса;

— упорядочить координаты $x_i(j)$, $i=\overline{1,m-1}$, $j=\overline{1,n}$, $x_m(j)\equiv 1$, согласно их влиянию или на размеры (толщину) линейной оптимальной полосы, или на критерий линейной полосной разделимости;

— сформулировать правило замены слабо влияющих (неинформативных) координат вектора признаков на более информативные из множества конкурирующих признаков;

— для выбранных согласно прикладной целесообразности классов нелинейных преобразований предложить средства определения оптимального как самого нелинейного преобразования из выбранного класса некоторой координаты $x_i(j)$, так и индекса i этой координаты (примеры таких классов приводятся ниже);

— сгенерировать суперпозицию оптимальных преобразований координат вектора признаков для достижения условий оптимальной полосной разделимости точек в новом пространстве признаков.

2.1. Условия и оптимизация полосной разделимости образов в пространстве признаков. Сформулируем сначала условия линейной отделимости точек $x(j)$, $j=\overline{1,n}$, от начала координат в пространстве R^m. Используя операции псевдообращения матриц, эту задачу будем понимать и исследовать в смысле существования решения системы относительно вектора $a \in R^m$

$$x^{\mathrm{T}}(j)a = y_j, \;\; y_j \geq \Delta, \;\; j=\overline{1,n}, \tag{12}$$

при некоторых $\Delta > 0$ и значениях y_j.

Тогда при фиксированном $\Delta > 0$ необходимое и достаточное условие существования решения этой задачи будет иметь вид

$$\min_{y \in D(\Delta)} y^{\mathrm{T}} Z(X) y = y_*^{\mathrm{T}}(\Delta) Z(X) y_*(\Delta) = 0, \tag{13}$$

где

$$X = (x(1) \vdots \ldots \vdots x(n)) = \begin{pmatrix} x_{(1)}^{\mathrm{T}} \\ \ldots \\ x_{(m)}^{\mathrm{T}} \end{pmatrix}, \;\; Z(X) = I_n - X^+ X,$$

$$D(\Delta) = \{y \colon y = (y_1, \ldots, y_n)^{\mathrm{T}}, \; y_j \geq \Delta, \; j=\overline{1,n}\}. \tag{14}$$

При этом искомый вектор a принимает значение

$$a(\Delta) = (X^{\mathrm{T}})^+ \, y_*(\Delta), \tag{15}$$

а толщина полосы δ, отделяющей множества точек $x(j)$, $j=\overline{1,n}$, от начала координат, равна величине

$$\mathrm{y}_* = \delta = \frac{\Delta}{(\mathrm{y}_*^{\mathrm{T}}(\Delta) \mathrm{R}(\mathrm{X}) \mathrm{y}_*(\Delta))^{1/2}}, \tag{16}$$

где $R(X) = X^+ (X^T)^+$.

Поскольку $y_*(k\Delta) = k\, y_*(\Delta)$, то без ограничения общности можно положить $\Delta = 1$, $y_*(1) = y_*$. Следовательно, максимальная толщина полосы достигается при значениях

$$y_{\mathrm{opt}} = \arg\min_{y \in D} y^{\mathrm{T}} R(X) y, \;\; a_{\mathrm{opt}} = (X^{\mathrm{T}})^+ \, y_{\mathrm{opt}}, \tag{17}$$

где $D = \{y \colon y^{\mathrm{T}} Z(X) y = y_*^{\mathrm{T}} Z(X) y_* = 0, e_j^{\mathrm{T}} y \geq 1, j=\overline{1,n}\}$. Используя сингулярное разложение матриц (SVD)

$$X = \sum_{j=1}^{r} u_j v_j^{\mathrm{T}} \lambda_j, \;\; Z(X) = I_n - \sum_{j=1}^{r} v_j v_j^{\mathrm{T}},$$

$$\mathrm{u}_{\mathrm{i}}^{\mathrm{T}} \mathrm{u}_{\mathrm{j}} = \delta_{\mathrm{ij}}, \;\; v_{\mathrm{i}}^{\mathrm{T}} v_{\mathrm{j}} = \delta_{\mathrm{ij}}, \;\; \mathrm{XX}^{\mathrm{T}} \mathrm{u}_{\mathrm{j}} = \lambda_{\mathrm{j}}^2 \mathrm{u}_{\mathrm{j}}, \;\; \mathrm{X}^{\mathrm{T}} \mathrm{X} v_{\mathrm{j}} = \lambda_{\mathrm{j}}^2 v_{\mathrm{j}},$$

$$\lambda_1^2 \geq \ldots \geq \lambda_{\mathrm{r}}^2, \;\; \overline{\mathrm{i}, \mathrm{j} = 1, \mathrm{n}},$$

и учитывая, что $y^{\mathrm{T}} Z(X) y = 0$ для $y \in D$, можно записать следующие соотношения:

$$y = \sum_{i=1}^{r} \alpha_i v_i, \;\; \sum_{i=1}^{r} \alpha_i e_j^{\mathrm{T}} v_i \geq 1 \;\; \forall j = \overline{1,n}, \;\; y^{\mathrm{T}} R(X) y = \sum_{j=1}^{r} \alpha_j^2 \lambda_j^{-2}. \tag{18}$$

Таким образом, задача поиска оптимальной линейной отделимости точек $x(j)$, $j=\overline{1,n}$, от начала координат сводится к решению задачи оптимизации квадратичной функции на выпуклом множестве:

$$\alpha_{\text{opt}} = \arg \min_{\alpha=\{\alpha:\, e_j^{\text{T}}(v_1 \vdots \ldots \vdots v_r)\alpha \geq 1,\; j=\overline{1,n}\}} \alpha^{\text{T}} \operatorname{diag}(\lambda_1^{-2}, \ldots, \lambda_r^{-2})\alpha,$$

$$y_{\text{opt}} = (v_1 \vdots \ldots \vdots v_r)\alpha_{\text{opt}}, \tag{19}$$

$$a_{\text{opt}} = \sum_{j=1}^{r} u_j v_j^{\text{T}} \lambda_j^{-1} \cdot (v_1 \vdots \ldots \vdots v_r)\alpha_{\text{opt}} = (\lambda_1^{-1} u_1 \vdots \ldots \vdots \lambda_r^{-1} u_r)\alpha_{\text{opt}},$$

$$\delta_{\text{opt}} = (\alpha_{\text{opt}}^{\text{T}} \operatorname{diag}(\lambda_1^{-2}, \ldots, \lambda_r^{-2})\alpha_{\text{opt}})^{-1/2}.$$

Используя приведенные выше условия линейной отделимости точек в R^m для задачи линейной полосной разделимости двух классов точек в пространстве признаков, получим как условия существования решения такой задачи, так и средство получения самого решения. Пусть известно, что для последовательности точек $x(j)$ в пространстве признаков R^m

$$x(j) \in R^m, \; x(j) = \begin{pmatrix} x_1(j) \\ \vdots \\ x_m(j) \end{pmatrix}, \; x_m(j) = 1, \; j = \overline{1,n},$$

точки $x(i_k)$, $k=\overline{1,n_1}$, принадлежат первому классу, а $x(j_s)$, $s=\overline{1,n_2}$ — второму. Тогда линейную полосную разделимость этих классов будем понимать в смысле существования такого вектора $a \in R^m$, для которого

$$a^{\text{T}} x(i_k) \geq 1, \; k=\overline{1,n_1},$$
$$a^{\text{T}} x(i_s) \leq -1, \; s=\overline{1,n_2}. \tag{20}$$

Тогда условие линейной полосной разделимости приобретает вид

$$\min_{y \in D} y^{\text{T}} Z(X) y = 0, \; D = \{y:\, e_{i_k}^{\text{T}} y \geq 1, e_{j_s}^{\text{T}} y \leq -1, k=\overline{1,n_1}, s=\overline{1,n_2}\}, \tag{21}$$

а значение вектора a определяется оптимально в смысле максимизации толщины разделяющей полосы

$$y_{\text{opt}} = \arg \min_{y \in D_1} y^{\text{T}} R(X) y, \tag{22}$$

где $D_1 = \{y:\, y^{\text{T}} Z(X) y = 0\} \cap D$,

$$a_{\text{opt}} = (X^{\text{T}})^{+} y_{\text{opt}}. \tag{23}$$

Применяя представление SVD для матрицы X, эти операции можно интерпретировать в более удобной для использования форме:

$$\alpha_{\text{opt}} =$$

$$= \arg \min_{\alpha \in \{\alpha:\, e_{i_k}^{\text{T}}(v_1 \vdots \ldots \vdots v_t)\alpha \geq 1,\, e_{i_s}^{\text{T}}(v_1 \vdots \ldots \vdots v_t)\alpha \leq -1,\, k=\overline{1,n_1},\, s=\overline{1,n_2}\}} \alpha^{\text{T}} \operatorname{diag}(\lambda_1^{-2}, \ldots, \lambda_r^{-2})\alpha, \tag{24}$$

$$y_{\text{opt}} = (v_1 \vdots \ldots \vdots v_r)\alpha_{\text{opt}}, \tag{25}$$

$$a_{\text{opt}} = (u_1 \vdots \ldots \vdots u_r)\operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_r^{-1})\alpha_{\text{opt}}. \tag{26}$$

2.2. Оптимальное нелинейное преобразование компонент вектора признаков. Весьма часто условие линейной полосной разделимости (21) в выбранном пространстве признаков не выполняется, т.е.

$$\min_{y \in D} y^{\text{T}} Z(X) y > 0. \tag{27}$$

В этом случае можно использовать следующие средства возможного выхода из создавшегося положения. Прежде всего попытаться улучшить выбор информативных признаков. Второе средство состоит в использовании каскадного полосного линейного разделения двух классов [1] и, как нам представляется, весьма перспективно синтезировать оптимально суперпозицию нелинейных преобразований в пространстве признаков. Этот синтез предлагается осуществлять поэтапно следующим образом.

Рассмотрим сначала задачу синтеза нелинейного преобразователя при известных входных и выходных данных соответственно: $X = (x(1) \vdots \ldots \vdots x(n))$ и $y = (y_1, \ldots, y_n)$. Попытаемся путем поиска нелинейного преобразования некоторых компонент вектора $x_{(i)}$, $i = \overline{1, m-1}$, изменить их таким образом, чтобы уменьшить невязку $y^{\mathrm{T}} Z(X) y$, и определим среди них оптимальную компоненту с индексом i_0 и соответствующее ей оптимальное преобразование. Для этого выполним указанную последовательность действий:

1) найти $a = (X^{\mathrm{T}})^{+} y$;

2) вычислить

$$a(i) = (X_{(i)}^{\mathrm{T}})^{+} y, \ X_{(i)} = \begin{pmatrix} x_{(1)}^{\mathrm{T}} \\ \vdots \\ x_{(i-1)}^{\mathrm{T}} \\ x_{(i+1)}^{\mathrm{T}} \\ \vdots \\ x_{(m)}^{\mathrm{T}} \end{pmatrix}, \text{ где } i = \overline{1, m-1}; \tag{28}$$

3) образовать вектор отклонений

$$\Delta y_i = (\Delta y_i(1), \ldots, \Delta y_i(n))^{\mathrm{T}}, \ \Delta y_i(j) = y(j) - \hat{y}_i(j), \ j = \overline{1, n}, \tag{29}$$

где $\hat{y}_i = (\hat{y}_i(1), \ldots, \hat{y}_i(n))^{\mathrm{T}} = Z(X_{(i)}) y = (I_n - X_{(i)}^{+} X_{(i)}) y = y - X_{(i)}^{\mathrm{T}} a(i)$;

4) заменить удаленную компоненту $x_i(j)$ нелинейной функцией вида $\varphi(a^{\mathrm{T}} x(j), \alpha(i))$, неизвестные параметры $\alpha(i)$ определяются из условия минимума отклонения

$$\sum_{j=1}^{n} \left\| \varphi(a^{\mathrm{T}} x(j), \alpha(i)) - \Delta y_i(j) \right\|^2 . \tag{30}$$

В качестве примера нелинейного преобразования можно рассмотреть сплайны третьего порядка

$$\varphi(a^{\mathrm{T}} x(j), \alpha(i)) = (1 \quad z(j) \quad z^2(j) \quad z^3(j)) \alpha(i), \ z(j) = a^{\mathrm{T}} x(j), \ j = \overline{1, n},$$

$$\alpha(\mathrm{i}) = \left(\sum_{\mathrm{j}=1}^{\mathrm{n}} \begin{pmatrix} 1 \\ \mathrm{z(j)} \\ \mathrm{z}^2(\mathrm{j}) \\ \mathrm{z}^3(\mathrm{j}) \end{pmatrix} (1 \quad \mathrm{z(j)} \quad \mathrm{z}^2(\mathrm{j}) \quad \mathrm{z}^3(\mathrm{j})) \right)^{+} \sum_{\mathrm{j}=1}^{\mathrm{n}} \begin{pmatrix} 1 \\ \mathrm{z(j)} \\ \mathrm{z}^2(\mathrm{j}) \\ \mathrm{z}^3(\mathrm{j}) \end{pmatrix} \cdot \Delta \mathrm{y}_{\mathrm{i}}(\mathrm{j}); \tag{31}$$

5) найти компоненту с индексом $i_0 \in \{i : \overline{1, m-1}\}$, при которой достигается минимум величины $y^{\mathrm{T}} Z(X_{\varphi, i}) y$, где

$$X_{\varphi, i} = (x_{(1)} \ldots x_{(i-1)} \ \varphi_i \ x_{(i+1)} \ldots x_{(m)}), \ \varphi_i (\varphi(a^{\mathrm{T}} x(1), \alpha(i)) \ldots \varphi(a^{\mathrm{T}} x(n), \alpha(i)))^{\mathrm{T}} .$$

Таким образом, на первом этапе получили наилучшее преобразование оптимально выбранной компоненты $x_{(i_0)}$.

Затем для новых входных данных повторяем описанный выше процесс нелинейного преобразования, реализуя, таким образом, суперпозицию нелинейных функциональных преобразований в целях минимизации невязки.

В задаче синтеза нелинейной системы распознавания образов задача усложняется в связи с тем, что выходные данные (вектор y) наперед неизвестны, т.е. если $\min\limits_{y \in D} y^{\mathrm{T}} Z(X) y > 0$, то в этом случае, как правило, такой линейной разделяющей полосы не существует. Возникает вопрос, какие компоненты $x_{(i_0)}$ требуется последовательно нелинейно преобразовать, чтобы разделяющая полоса в новом полученном пространстве входных данных существовала?

Для этого воспользуемся приведенным выше алгоритмом, в котором выходные данные составляют вектор y_*. Этот вектор находим, решая задачу оптимизации квадратичной функции при линейных ограничениях

$$y_* = \operatorname{Arg} \min_{y \in D} y^{\mathrm{T}} Z(X) y, \tag{32}$$

используя одну из модификаций градиентного метода, например описанную в [1].

Далее выполняем все приведенные выше операции пп.1–5, полагая $y = y_*$. Для новых векторов признаков

$$\begin{pmatrix} x_1(j) \\ \vdots \\ x_{i_0-1}(j) \\ \varphi(a^{\mathrm{T}} x(j), \alpha(i_0)) \\ x_{i_0+1}(j) \\ \vdots \\ x_m(j) \end{pmatrix}, \quad j = \overline{1, n}, \tag{33}$$

определяем новое значение $y \in D$, при котором достигается минимум выражения $y^{\mathrm{T}} Z(X_{\varphi, i_0}) y$ и соответствующее новое значение

$$a = (X^{\mathrm{T}}_{\varphi, i_0}) y. \tag{34}$$

Если это минимальное значение $y^{\mathrm{T}} Z(X_{\varphi, i_0}) y$ равняется нулю, т.е. выполняется условие линейной полосной разделимости в новом пространстве признаков, то необходимо найти

$$a_{\mathrm{opt}} = (X^{\mathrm{T}}_{\varphi, i_0})^{+} y_{\mathrm{opt}}, \quad y = \arg \min_{y \in D_1} y^{\mathrm{T}} R((X^{\mathrm{T}}_{\varphi, i_0})^{+}) y, \tag{35}$$

где $D_1 = \{y \colon y^{\mathrm{T}} Z(X_{\varphi, i_0}) y = 0\} \cap D$. И, таким образом, построить оптимальную линейную разделяющую операцию $a^{\mathrm{T}}_{\mathrm{opt}} x(i_k) \geq 1$, $a^{\mathrm{T}}_{\mathrm{opt}} x(j_s) \leq -1$, $k = \overline{1, n_1}$, $s = \overline{1, n_2}$, для новых векторов признаков, которые получаются после нелинейного преобразования компонент.

3. ОРГАНИЗАЦИЯ СИСТЕМЫ РАСПОЗНАВАНИЯ ОБРАЗОВ

Рассмотрим структурно организацию системы распознавания образов. Более детальные исследования по задаче организации системы и ее реализации планируется развить в последующей работе авторов.

1. На основе идеальных представителей классов конкретной рассматриваемой предметной области создаем базу данных и средствами синтеза кластеров строим дихотомное дерево классификации образов. На этом этапе решаем задачу выбора характерных признаков классификации и их количество (размерность пространства признаков).

2. На основе уже не идеальных, а реальных представителей классов обучающей выборки для каждой вершины дерева синтезировать оптимальный полосно-разделяющий классификатор согласно информации об объектах, содержащейся в обучающей последовательности.

3. В итоге настроенная таким образом система распознавания может уточняться путем оптимизации всех параметров в ранее организованной структуре, т.е. при этом осуществляется реализация обучения синтезированной системы распознавания образов.

ЗАКЛЮЧЕНИЕ

В настоящей статье осуществляется развитие средств псевдообращения матриц применительно к исследованию задач кластеризации информации и распознаванию образов. Приведенные методы нелинейного преобразования компонент вектора признаков позволяют расширить сферу использования аппарата псевдообращения матриц на нелинейные задачи распознавания образов. В статье предлагаются детальные схемы построения таких алгоритмов, работоспособность которых проверена авторами на решении ряда конкретных задач, связанных с распознаванием ультразвуковых сигналов, рукописных символов. В настоящее время ведется работа по их применению для распознавания содержащейся информации в речевых сигналах.

СПИСОК ЛИТЕРАТУРЫ

1. Кириченко Н.Ф., Кривонос Ю.Г., Лепеха Н.П. Синтез систем нейрофункциональных преобразователей в решении задач классификации // Кибернетика и системный анализ. — 2007. — № 3. — С. 47–57.
2. Кириченко Н.Ф., Донченко В.С., Сербаев Д.П. Нелинейные рекурсивные регрессионные преобразователи: динамические системы и оптимизация // Там же. — 2005. — № 3. — С. 58–68
3. Алберт А. Регрессия, псевдоинверсия, рекуррентное оценивание / Пер. с англ. — М.: Наука, 1977. — 305 с.
4. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений / Пер. с англ. — М.: Мир, 1980. — 279 с.
5. Беклемишев Д.В. Дополнительные главы линейной алгебры. — М.: Наука, 1983. — 336 с.
6. Кириченко Н.Ф. Аналитическое представление возмущений псевдообратных матриц // Кибернетика и системный анализ. — 1997. — № 2. — С. 98–107.
7. Кириченко Н.Ф., Лепеха Н.П. Применение псевдообратных и проекционных матриц к исследованию задач управления, наблюдения и идентификации // Там же. — 2002. — № 4. — С. 107–124.
8. Кириченко Н.Ф., Крак Ю.В., Полищук А.А. Псевдообратные и проекционные матрицы в задачах синтеза функциональных преобразователей // Там же. — 2004. — № 3. — С. 116–129.
9. Верченко А.П., Кириченко Н.Ф., Лепеха Н.П. Рекуррентные средства кластеризации в применении к задачам сегментации изображений // Проблемы управления и информатики. — 2005. — № 2. — С. 62–71
10. Форсайт Дж., Понс М. Компьютерное зрение. Современный подход / Пер. с англ. — М.: Вильямс, 2004. — 928 с.
11. Lucchese L., Mitra S. Color image segmentation: A state-of-the-art survey // Proc. of the Indian National Sci. Academ. (INSA-A). — 2001. — **67**, N 2. — P. 207–221.
12. Kohonen T. Self-organizing maps. — Berlin: Springer, 2001. — 501 p.

Поступила 03.04.2008