УДК 519.711

Ф.Г. Гаращенко, Г.И. Кудин

ЗАДАЧИ КЛАССИФИКАЦИИ СРЕДСТВАМИ ПСЕВДООБРАЩЕНИЯ: ПРЕОБРАЗОВАНИЕ ПРОСТРАНСТВА ПРИЗНАКОВ, ВЛИЯНИЕ ВОЗМУЩЕНИЙ ИСХОДНОЙ ИНФОРМАЦИИ

Введение

На протяжении последних десяти лет в Институте кибернетики имени В.М. Глушкова НАНУ (ИК НАНУ), а также на кафедре моделирования сложных систем факультета компьютерных наук и кибернетики Киевского национального университета имени Тараса Шевченко разрабатывались алгоритмы классификации и кластеризации информации средствами псевдообращения матриц, которые базировались на фундаментальных результатах профессора Н.Ф. Кириченко [1–9].

При решении многих прикладных задач цифровой обработки информации используются критерии выбора алгоритма решения задачи на этапе построения классификатора — внутренние критерии, а также эффективность работы классификатора на множестве данных контрольной выборки — внешние критерии. Эти вопросы исследовались в публикациях А.Г. Ивахненко [10]. Относительно алгоритмов классификации и кластеризации средствами псевдообращения матриц такие исследования не проводились, что в известной мере объясняется использованием алгоритмов при решении отдельных прикладных задач, которые не нуждались в углубленном анализе качественных характеристик систематического использования выбранной технологии. В прикладных задачах классификации предполагается обработка экспериментальной информации, для которой желательно иметь определенные дополнительные сведения: для случайных данных — это вероятностные параметры, для детерминированных — границы возмущений. Поэтому зависимость погрешности полученных решений задач от возмущений исходных данных — важная составляющая научных исследований. Например, в работе [1] математически строго исследована точность решений системы линейных алгебраических уравнений при наличии погрешностей в исходных элементах матрицы, правых частей системы.

В настоящей работе представлены результаты исследований решения задачи гиперплоскостной классификации информации средствами псевдообращения матриц. С использованием критериальных функций качества процесса распознавания получены количественные оценки влияния погрешностей экспериментов на условия разрешимости задачи классификации и ширину полосы разделимости классов. Используются результаты теории возмущения псевдообратных и проекционных матриц.

1. Необходимые сведения из теории псевдообратных и проекционных матриц

Приняты следующие способы представления матрицы $A \in R^{m\times n}$:

$$A = (a(1) \vdots \ldots \vdots a(n)) \equiv (a_{(1)} \vdots \ldots \vdots a_{(m)})^{\mathrm{T}} \equiv (a_{ij})_{\substack{i=\overline{1,m} \\ j=\overline{1,n}}} \in R^{m\times n},$$

где
$a(j) = (\alpha_{1j},..., \alpha_{ij},..., \alpha_{mj})^{\mathrm{T}} \in R^m,\ a_{(i)} = (\alpha_{i1},..., \alpha_{ij},..., \alpha_{in}) \in R^n, \quad i = \overline{1, m}, \quad j = \overline{1, n}.$

Оптимизационная форма определения псевдообратной матрицы. Для матрицы $A \in R^{m\times n}$ псевдообратная по Пенроузу матрица $A^+ \in R^{m\times n}$ определяется соотношением

$$\forall b \in R^m \ \ A^+ b = \underset{x \in \Omega_A(b)}{\arg\min} \left\| x \right\|^2,$$

где

$$\Omega_A(b) = \mathrm{Arg} \min_{x \in R^n} \left\| Ax - b \right\|^2.$$

Сингулярное представление псевдообратной матрицы. Используем сингулярное представление матриц, т.е. представление в виде

$$A = \sum_{i=1}^{r} y_i x_i^{\mathrm{T}} \lambda_i, \ \ r = \mathrm{rank}\, A,$$

где y_j, x_j — ортонормированные системы собственных векторов матриц AA^{T}, $A^{\mathrm{T}}A$ соответственно, а λ_j^2 — общие их собственные значения:

$$AA^{\mathrm{T}} y_j = \lambda_j^2 y_j, \ \ A^{\mathrm{T}} A x_j = \lambda_j^2 x_j,$$

$$y_i^{\mathrm{T}} y_j = \delta_{ij}, \ \ x_i^{\mathrm{T}} x_j = \delta_{ij}, \ \ i, j = \overline{1, r}, \ \ \lambda_1^2 \ge \ldots \ge \lambda_r^2,$$

δ_{ij} — символ Кронекера.

Псевдообратную матрицу A^+ можно записать в виде сингулярного представления

$$A^+ = \sum_{i=1} x_j y_j^{\mathrm{T}} \lambda_j^{-1}.$$

Основные матрицы, связанные с псевдообратной матрицей. При применении псевдообращения к задачам классификации важными являются матрицы, которые определяются и вычисляются с использованием матриц A и A^+:

1) проекционная матрица $P(A) = A^+ A \equiv \sum_{j=1}^{r} x_i x_i^{\mathrm{T}}$ — ортогональный проектор на подпространство $L_{A^{\mathrm{T}}}$, порожденное векторами-строками матрицы A, т.е. на подпространство значений A^{T};

2) проекционная матрица $Z(A) = I_n - P(A)$ — ортогональный проектор на подпространство, ортогональное подпространству $L_{A^{\mathrm{T}}}$;

3) матрица $R(A) = A^+ (A^+)^{\mathrm{T}} \equiv \sum_{j=1}^{r} x_j x_j^{\mathrm{T}} \lambda_j^{-2}$.

2. Зависимость псевдообратных матриц при добавлении произвольных вектор-строк к исходной матрице

Если предположить, что расширение матрицы A происходит добавлением новой строки $a^{\mathrm{T}} \in R^n$ после $(i-1)$-й строки $(i = \overline{2, m+1})$, т.е. образуется матрица

$$A_{i,a} = (a_{(1)} \vdots \ldots \vdots a_{(i-1)} \vdots a \vdots a_{(i)} \vdots \ldots \vdots a_{(m)})^{\mathrm{T}} \in R^{(m+1)\times n},$$

то при известной псевдообратной матрице $A^+ \in R^{n\times m}$ для рекуррентного вычисления псевдообратной матрицы $A_{i,a}^+ \in R^{n\times(m+1)}$ применяют прямые формулы Гревиля–Кириченко.

Прямые формулы Гревиля–Кириченко для псевдообратных матриц [6]. Если неизвестную псевдообратную матрицу $A_{i,a}^{+}$ представить в виде

$$A_{i,a}^{+} = (p(1) \vdots \ldots \vdots p(i-1) \vdots p(i) \vdots p(i+1) \vdots \ldots \vdots p(m+1)) \in R^{n\times(m+1)},$$

где

$$P_i = (p(1) \vdots \ldots \vdots p(i-1) \vdots p(i+1) \vdots \ldots \vdots p(m+1)),$$

т.е. считать, что матрица $A_{i,a}^{+}$ может быть получена из матрицы $P_i \in R^{n\times m}$ добавлением вектор-столбца $p(i)$ после $(i-1)$-го столбца, то для матрицы P_i имеет место формула

$$P_i = (1 - p(i)\, a^{\mathrm{T}})\, A^{+}.$$

При этом вектор $p(i)$ определяется в зависимости от свойств линейной зависимости вектора a от векторов подпространства вектор-строк матрицы A:

1) если вектор-строка a линейно независима от векторов подпространства вектор-строк матрицы A, т.е. $a^{\mathrm{T}} Z(A)\, a > 0$, то

$$p(i) = Z(A)\, a \| Z(A)\, a \|^{-2};$$

2) если вектор-строка a линейно зависит от векторов подпространства вектор-строк матрицы A, т.е. $a^{\mathrm{T}} Z(A)\, a = 0$, то

$$p(i) = R(A)\, a (1 + a^{\mathrm{T}} R(A)\, a)^{-1}.$$

Следствие. Операция псевдообращения коммутативна с транспонированием, поэтому прямые формулы Гревиля–Кириченко аналогично выписываются для варианта расширения матрицы столбцом.

Используя прямые формулы Гревиля–Кириченко для представления псевдообратной матрицы A_{ia}^{+}, можно получить формулы для вычисления проекционной матрицы $Z(A_{i,a})$, а также матрицы $R(A_{i,a})$:

а) при $a^{\mathrm{T}} Z(A)\, a > 0$ (вектор a линейно независим)

$$Z(A_{i,a}) = Z(A) - p(i)\, p^{\mathrm{T}}(i) \| p(i) \|^{-2} \in R^{n\times n},$$

$$R(A_{i,a}) = R(A) - (p(i)\, a^{\mathrm{T}} R(A) - R(A)\, a p^{\mathrm{T}}(i)) \| p(i) \|^{-2} +$$

$$+ p(i)\, p^{\mathrm{T}}(i)(1 + a^{\mathrm{T}} R(A)\, a)^{-1} \in R^{n\times n};$$

б) при $a^{\mathrm{T}} Z(A)\, a = 0$ (вектор a линейно зависим)

$$Z(A_{i,a}) = Z(A) \in R^{n\times n},$$

$$R(A_{i,a}) = R(A) + R(A)\, a a^{\mathrm{T}} R(A)(1 + a^{\mathrm{T}} R(A)\, a)^{-1} \in R^{n\times n}.$$

3. Обратные формулы Гревиля–Кириченко для псевдообратных матриц [6]

Если для матрицы $A(i, a) \in R^{(m+1)\times n}$ (после $(i-1)$-й строки $(i = \overline{2, m+1})$ матрицы A вставлена вектор-строка a^{T}), известна ее псевдообратная матрица $A^{+}(i, a) \in R^{n\times(m+1)}$, то для определения псевдообратной матрицы $A^{+} \in R^{n\times m}$ (строка a^{T} матрицы $A(i, a)$ удаляется) имеют место обратные формулы Гревиля–Кириченко:

1) для случая линейной зависимости удаляемого вектор-строки от векторов подпространства вектор-строк матрицы $A(i, a)$, а это определяется выполнением условия $a^{\mathrm{T}} p(i) < 1$, то псевдообратная матрица $A^{+} \in R^{n\times m}$ имеет вид

$$A^{+} = (I_n + p(i)\, a^{\mathrm{T}} (1 - a^{\mathrm{T}} p(i))^{-1})\, P_i,$$

при этом ранг псевдообратной матрицы не изменяется, т.е. $\operatorname{rank} A = \operatorname{rank} (A^{\mathrm{T}} \vdots\, a)^{\mathrm{T}}$;

2) для случая линейной независимости удаляемого вектор-строки a^{T} от векторов подпространства вектор-строк матрицы $A(i, a)$, а это определяется выполнением условия $a^{\mathrm{T}} p(i) = 1$, псевдообратная матрица $A^{+} \in R^{n\times m}$ определяется формулой

$$A^{+} = (I_n - p(i)\, p^{\mathrm{T}}(i) \| p(i) \|^{-2})\, P,$$

при этом ранг псевдообратной матрицы уменьшается: $\operatorname{rank} A = (A^{\mathrm{T}} \vdots\, a)^{\mathrm{T}} - 1$.

Используя обратные формулы Гревиля–Кириченко для представления псевдообратной матрицы $A^{+} \in R^{n\times m}$, можно получить полезные для применений формулы экономного вычисления возмущенных матриц $Z(A)$, $R(A)$:

1) для случая линейной зависимости удаляемого вектор-строки a^{T} от векторов подпространства вектор-строк матрицы $A(i, a)$, а это определяется выполнением условия $a^{\mathrm{T}} p(i) < 1$,

$$Z(A) = Z(A(i, a));$$

$$R(A) = R(A(i, a)) + p(i)\, p^{\mathrm{T}}(i) Z_{ii}^{-1} (A^{\mathrm{T}}(i, a));$$

2) для случая линейной независимости удаляемого вектор-строки a^{T} от векторов подпространства вектор-строк матрицы $A(i, a)$, а это определяется выполнением условия $a^{\mathrm{T}} p(i) = 1$,

$$Z(A) = Z(A(i, a)) + p(i)\, p^{\mathrm{T}}(i) \| p(i) \|^{-2},$$

$$R(A) = Z(p^{\mathrm{T}}(i))\, R(A(i, a))\, Z(p^{\mathrm{T}}(i)).$$

4. Оптимизация полосной разделимости образов в пространстве признаков [6]

Используя операции псевдообращения матриц, целесообразно сформулировать условия линейной отделимости точек $x(j)$, $j = \overline{1, n}$, от начала координат в пространстве R^m. Эта задача понимается в смысле существования решения системы $x^{\mathrm{T}}(j) a = y_j$, $y_j \geq \Delta$, $j = \overline{1, n}$, относительно вектора $a \in R^m$ при некоторых $\Delta > 0$ и значениях y_j. При фиксированном $\Delta > 0$ необходимое и достаточное условие существования решения этой задачи будет иметь вид

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

где

$$X = (x(1) \vdots \ldots \vdots\, x(n)) = (x_{(1)} \;\; \ldots \;\; x_{(m)})^{\mathrm{T}}, \;\; D(\Delta) = \{ y : y = (y_1, \ldots, y_n)^{\mathrm{T}}, \; y_j \geq \Delta, \; j = \overline{1, n} \}.$$

При этом искомый вектор a принимает значение $a(\Delta) = (X^{\mathrm{T}})^{+} y_*(\Delta)$, а толщина δ полосы, которая отделяет множество точек $x(j)$, $j = \overline{1, n}$, от начала координат, равна

$$y_* = \delta = \Delta(y_*^{\mathrm{T}}(\Delta) R(X) y_*(\Delta))^{-1/2}.$$

Поскольку $y_*(k\Delta) = ky_*(\Delta)$, то можно положить $\Delta = 1$, $y_*(1) = y_*$. Таким образом, максимальная толщина полосы достигается при значениях

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

где

$$D = \{ y : y^{\mathrm{T}} Z(X) y = y_*^{\mathrm{T}} Z(X) y_* = 0, e_j^{\mathrm{T}} y \geq 1, \; j = \overline{1, n} \}.$$

Используя сингулярное разложение матриц $X = \sum_{j=1}^{r} u_j v_j^{\mathrm{T}} \lambda_j, Z(X) = I_n - \sum_{j=1}^{r} v_j v_j^{\mathrm{T}}$, а также то, что $y^{\mathrm{T}} Z(X) y = 0$ для $y \in D$, можно установить следующие соотношения:

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

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

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

$$a_{\mathrm{opt}} = (\lambda_1^{-1} u_1 \vdots \ldots \vdots \lambda_r^{-1} u_r) \alpha_{\mathrm{opt}}, \;\; y_{\mathrm{opt}} = (v_1 \vdots \ldots \vdots v_r) \alpha_{\mathrm{opt}},$$

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

Условия линейной отделимости точек R^m используются в задаче линейной полосной отделимости двух классов точек в пространстве признаков. Пусть известно, что для последовательности точек $x(j)$ в пространстве признаков R^m

$$x(j) \in R^m, \;\; x(j) = (x_1^{\mathrm{T}}(j) \; \ldots \; x_m^{\mathrm{T}}(j))^{\mathrm{T}}, \;\; x_m(j) = 1, \;\; j = \overline{1, n},$$

точки $x(i_k), k = \overline{1, n_1}$, принадлежат первому классу, а $x(j_s), s = \overline{1, n_2}$, — второму. Линейная полосная отделимость этих классов понимается в смысле существование такого вектора $a \in R^m$, для которого

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

Тогда условие линейной полосной отделимости двух классов точек в пространстве признаков принимает следующий вид:

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

а значение вектора a определяется оптимально: $a_{\mathrm{opt}} = (X^{\mathrm{T}})^{+} y_{\mathrm{opt}}$, где

$$y_{\text{opt}} = \arg\min_{y \in D_1} y^{\mathrm{T}} R(X) y, \;\; D_1 = \{y : y^{\mathrm{T}} Z(X)\, y = 0\} \cap D.$$

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

$$\alpha_{\text{opt}} = \arg \min_{\alpha \in \{\alpha : e_{i_k}^{\mathrm{T}} (v_1 \vdots \ldots \vdots v_r)\alpha \geq 1,\, e_{js}^{\mathrm{T}} (v_1 \vdots \ldots \vdots v_r)\alpha \leq -1,\, k=\overline{1,n_1},\, s=\overline{1,n_2}\}} \alpha^{\mathrm{T}} \operatorname{diag}(\lambda_1^{-2}, \ldots, \lambda_r^{-2})\, \alpha,$$

$$y_{\text{opt}} = (v_1 \vdots \ldots \vdots v_r)\, \alpha_{\text{opt}}, \;\; a_{\text{opt}} = (u_1 \vdots \ldots \vdots u_r) \operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_r^{-1})\, \alpha_{\text{opt}}.$$

Толщина разделяющей полосы в задаче оптимальной линейной полосной отделимости точек определяется формулой

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

5. Технологии отбора информативных признаков и минимизация их числа

Эффективность процесса учебы и режима использования классификатора в большой мере зависит от наличия характеристик информации о рассматриваемых классах, неизбыточности пространства признаков, отсутствия возможных эффектов переобучения. Алгоритмы классификации предусматривают выполнение определенного числа вычислений больших массивов информации, поэтому естественно ожидать при решении задач существенных погрешностей алгоритма. Практика цифровой обработки информации подтверждает наличие погрешностей значений компонент векторов пространства признаков, что обуславливает потребность согласовать погрешность работы классификатора с погрешностями измерений. В рамках постановки задачи об эффективности работы классификатора можно обоснованно считать, что погрешности компонент векторов пространств признаков известны.

Учитывая вычислительную схему алгоритма гиперплоскостной классификации с использованием теории псевдообращения матриц, оптимальным преобразованием пространства векторов признаков и повторением алгоритма на векторах признаков модифицированного пространства можно улучшить работу классификатора.

Поочередное удаление, преобразование признаков [5]. Если гиперплоскостной классификатор построен, но возникает потребность его улучшить, то, используя критерий разделимости классов, из матриц информации о классах можно удалить те признаки, которые не увеличивают ширину полосы между классами. Для случая неразделимости классов с помощью обратных формул Гревиля–Кириченко можно удалить неинформативные признаки и достичь свойств разделимости классов.

Наименее информативная координата x_{i^*} определяется решением оптимизационной задачи

$$y_*^{\mathrm{T}} Z(X_{i^*})\, y_* = \min_{i=\underline{1,n}} y_*^{\mathrm{T}} Z(X_i)\, y_*,$$

$$X_i = (x_{(1)} \;\ldots\; x_{(i-1)} \;\; x_{(i+1)} \;\ldots\; x_{(m)})^{\mathrm{T}}.$$

Используя результаты теории возмущения псевдообратных и проекционных матриц, условие существования наименее информативной координаты можно интерпретировать таким образом. Если $q(i)$ — i-й вектор-столбец матрицы X^+ и для некоторого $i = i_*$ выполняется условие $x^{\mathrm{T}}(i_*)\, q(i_*) \neq 1$, то i_*-я компонента вектора признаков неинформативна. Если же $x^{\mathrm{T}}(i)\,q(i)=1,\ \forall i=\overline{1,m},$ то наименее информативная i_*-я компонента вектора признаков определяется из условия

$$\frac{|\,y_*^{\mathrm{T}}q(i_*)\,|^2}{\|q(i_*)\|^2}=\min_{i=\overline{1,m}}\frac{|\,y_*^{\mathrm{T}}q(i)\,|^2}{\|q(i)\|^2}.$$

Выбор новой компоненты x_0 для замены наименее информативной компоненты x_{i^*} вектора признаков $x=(x_1,\ldots,x_m)^{\mathrm{T}}$, т.е. рассмотрение нового вектора $(x_1,\ldots,x_{i^*-1},x_0,x_{i^*+1},\ldots,x_m)^{\mathrm{T}}$ вместо предыдущего x, осуществляется по условию оптимальности:

$$y_*^{\mathrm{T}}Z(X_{(i^*,x_{(0)})})\,y_*\to\min_{x_0\in P_x},$$

$$Z(X_{(i^*,x_{(0)})})=Z(X_{i_*})-\frac{Z(X_{i_*})\,x_{(0)}x_{(0)}^{\mathrm{T}}\,Z(X_{i_*})}{\|\,Z(X_{i_*})\,x_{(0)}\,\|^2},$$

$$Z(X_{i_*})=\begin{cases}Z(X)+\dfrac{q_{i_*}q_{i_*}^{\mathrm{T}}}{\|q_{i_*}\|^2}, & x_{(i_*)}^{\mathrm{T}}q_{i_*}=1,\\ Z(X), & x_{(i_*)}^{\mathrm{T}}q_{i_*}\neq 1.\end{cases}$$

Здесь $X_{(i^*,x_{(0)})}$ — матрица, полученная из матрицы X заменой неинформативной вектор-строки $x_{(i_*)}^{\mathrm{T}}$ новой вектор-строкой $x_{(0)}^{\mathrm{T}}=(x_0(1),\ldots,x_0(n)),$ составленной по оптимально выбранному признаку x_0. В работе [5] качество классификации сигналов описанными выше средствами предложено улучшить за счет нелинейной трансформации выбранного пространства признаков, а именно, если i_*-я координата пространства признаков x_{i_*} наиболее неинформативна, то осуществляется замена этой компоненты каждого входного вектора $x(j)=(x_1(j),\ldots,x_{i*}(j),\ldots,x_m(j))^{\mathrm{T}}$ значениями некоторой нелинейной функции $\psi_0(x_{i_0}(j))$ от компоненты x_{i_0}. Функция $\psi_0(\cdot)\in\Psi$, Ψ — заданный допустимый класс функций нелинейных трансформаций, а для определения индекса компоненты $i_0\in\{\overline{1,m}\}$ также выписываются условия оптимальности.

Несколько проще в вычислительном плане может быть замена наиболее неинформативной компоненты $x_{i_*}(j)$, $j=\overline{1,n}$, каждого входного вектора $x(j)=$ $=(x_1(j),\ldots,x_{i*}(j),\ldots,x_m(j))^{\mathrm{T}}$ новыми значениями, полученными с использованием кластеризации признаков, применением методов математического программирования.

Кластеризация признаков. Уменьшения размерности пространства признаков можно достичь нахождением групп подобных признаков, выделить в каждой по типичному представителю, т.е. в виде нового признака используется доминирующий носитель построенного кластера группы. При этом целесообразно кластеры групп формировать с информативных признаков и предварительным выбором функции расстояния между признаками в составе группы.

Отбор признаков методами математического программирования. Технология базируется на построении линейной комбинации определенной группы информативных признаков. В пространстве неопределенных коэффициентов такой линейной комбинации по критерию разделимости классов или/и величины ширины полосы разделимости классов нужно решить соответствующую задачу математического программирования.

6. Технологии построения классификаторов при наличии погрешностей экспериментальных данных

В алгоритме гиперплоскостной классификации информации средствами псевдообращения матриц предполагается, что для компонент векторов пространства признаков классов проведена процедура стандартизации признаков — приведение их к единому масштабу, поэтому существует возможность количественной оценки влияния погрешностей экспериментальных данных на результат построения классификатора. Теория псевдообращения матриц предусматривает преобразование элементов евклидового пространства, поэтому необходимые для последующего изложения расстояния соответствий считаются евклидовыми — это расстояния между векторами, расстояние вектора от класса, расстояния между классами.

Целесообразно считать, что два класса линейно разделимы, если расстояние между центрами классов без суммы их дисперсий больше ширины полученной полосы разделимости. Для повышения качества работы построенного классификатора можно учесть дополнительно требование, чтобы ширина полосы разделимости классов была больше некоторого порогового значения, определение которого учитывает информацию о погрешностях проведенных вычислений, а также других обстоятельств, связанных с постановкой задачи. Такой поход к построению классификатора, когда варианты расстояний между классами дополняются выбором расстояния между центрами классов с учетом величины дисперсий элементов каждого класса, хорошо согласовывается с принципом учета компактности классов.

Цифровая обработка в рамках распознавания информации кроме задачи классификации предполагает возможность решения задач кластеризации. Эти задачи принципиально отличаются: в классификации присутствует информация о составе и количестве классов, а для классической задачи кластеризации такая информация отсутствует. В то же время, по нашему мнению, при необходимости линейной дискриминации многих классов использование методов кластеризации информации может существенно сузить количество вариантов построения последовательности линейных двуместных классификаторов. Один из подходов сочетания методов классификации и кластеризации информации — использование идей шкалирования информации, которые можно реализовать построением кусочно-гиперплоскостного кластера на множестве центров классов [9].

В дальнейшем изложении предполагается, что матрицы информации о классах полноранговые, обусловленность матриц вычислительного процесса будет отслеживаться величиной их сингулярных чисел. Это предположение является существенным с учетом результатов фундаментальных исследований научных сотрудников ИК НАН Украины [1]: получена погрешность результатов вычислений решений систем линейных алгебраических уравнений при наличии погрешностей в исходных элементах матрицы системы, правых частей системы.

Исследование влияния погрешностей экспериментальных данных на решение задачи классификации может предоставить важные характеристики классификатора:

1) входной сигнал классификатора попадает в сферу определенного класса — классификатор это подтверждает;

2) входной сигнал классификатора попадает в сферу определенного класса, но при этом полоса разделимости сужается;

3) решение задачи гиперплоскостной классификации неустойчиво при возмущении определенной компоненты вектора признаков.

7. Определение смещений минимумов условия разделимости классов при возмущении определенной компоненты вектора признаков

Условие линейной полосной отделимости двух классов имеет следующий вид:

$$\min_{y\in\Omega_y} y^{\mathrm{T}} Z(X)\, y = y_*^{\mathrm{T}} Z(X)\, y_* = 0,$$

$$\Omega_y = \{y : y = (y(1),..., y(n))^{\mathrm{T}},\ y(i_k) \geq 1, k = \overline{1, n_1}, y(j_s) \leq -1, s = \overline{1, n_2}\},$$

где $X = (x_{(1)}\ \ldots\ x_{(i-1)}\, x_{(i)}\ \ x_{(i+1)}\ \ldots\ x_{(m)})^{\mathrm{T}}$.

Если предположить, что возмущается i-я компонента вектора признаков, то это значит, что изменяется i-я строка матрицы, т.е. нужно рассматривать новую образованную матрицу

$$X(i, a) = (x_{(1)}\ \ldots\ x_{(i-1)}\, a\ \ x_{(i+1)}\ \ldots\ x_{(m)})^{\mathrm{T}},$$

где $a = x_{(i)} + \delta_i$.

Возмущения компонент вектора признаков целесообразно считать малыми, но их значения достаточные для выхода компоненты за пределы соответствующей дисперсии.

Условия, которым должен удовлетворять вектор $a = x_{(i)} + \delta_i$, формулируются с использованием матрицы $X(i) = (x_{(1)}\ \ldots\ x_{(i-1)}\ \ x_{(i+1)}\ \ldots\ x_{(m)})^{\mathrm{T}}$:

1) вектор a линейно зависим от векторов подпространства вектор-строк матрицы $X(i)$, т.е. $a^{\mathrm{T}} Z(X(i))\, a = 0$, случай интересен тем, что возмущение не влияет на условие разделимости классов, но влияет на ширину полосы разделимости;

2) вектор a линейно независим от векторов подпространства вектор-строк матрицы $X(i)$, т.е. $a^{\mathrm{T}} Z(X(i))\, a > 0$, этот случай нуждается в последующей детализации.

Для случая, когда вектор a линейно независим от векторов подпространства вектор-строк матрицы $X(i)$, используя формулу для вычисления возмущенной проекционной матрицы

$$Z(X(i, a)) = Z(X(i)) - p(i)\, p^{\mathrm{T}}(i) \left\| Z(X(i))\, a \right\|^2 \in R^{n\times n},$$

где $p(i) = Z(X(i))\, a \left\| Z(X(i))\, a \right\|^{-2}$, а также первое приближение метода малого параметра (малой величиной считается $\|\delta\|$) для определения векторной поправки γ к экстремуму y_* невозмущенного условия разделимости классов, вычисления можно подать в такой последовательности.

1. Условие разделимости классов при возмущении компоненты вектора признаков:

ISSN 0572-2691

$$\min_{\gamma} (y_* + \gamma)^{\mathrm{T}} Z(X(i, a))(y_* + \gamma) =$$

$$= \min_{\gamma} (y_* + \gamma)^{\mathrm{T}} (Z(X(i)) - Z(X(i))(x_{(i)} + \delta_i)(x_{(i)} + \delta_i)^{\mathrm{T}} Z(X(i)) /$$

$$/ (x_{(i)} + \delta_i)^{\mathrm{T}} Z(X(i))(x_{(i)} + \delta_i))(y_* + \gamma),$$

$$y_*^{\mathrm{T}} Z(X(i)) y_* = 0, \quad y_* + \gamma = y,$$

$$\Omega_y = \{ y : y = (y(1), ..., y(n))^{\mathrm{T}},$$

$$y(i_k) \geq 1, \; k = \overline{1, n_1}, \; y(j_s) \leq -1, \; s = \overline{1, n_2} \}.$$

2. Условие разделимости классов при возмущении компоненты вектора признаков в первом приближении метода малого параметра:

$$\min_{\gamma} (y_* + \gamma)^{\mathrm{T}} (Z(X(i)) - \frac{Z(X(i)(x_{(i)} x_i^{\mathrm{T}} + (x_{(i)} \delta_i^{\mathrm{T}} + \delta_i x_i^{\mathrm{T}})) Z(X(i))}{x_{(i)}^{\mathrm{T}} Z(X(i) x_{(i)} + (\delta_i^{\mathrm{T}} Z(X(i) x_{(i)} + x_{(i)} Z(X(i) \delta_i)}) (y_* + \gamma) +$$

$$+ o(\|\delta\|) = \min_{\gamma} (\gamma^{\mathrm{T}} (Z(X) + C) \gamma + 2 y_*^{\mathrm{T}} C \gamma + y_*^{\mathrm{T}} C y_* + o(\|\delta\|)),$$

где использованы обозначения

$$C = (b_1 E_n - A_1) Z(X(i) / b,$$

$$b = x_{(i)}^{\mathrm{T}} Z(X(i) x_{(i)}, \quad A_1 = Z(X(i)(x_{(i)} \delta_i^{\mathrm{T}} + \delta_i x_i^{\mathrm{T}}),$$

$$b_1 = \delta_i^{\mathrm{T}} Z(X(i) x_{(i)} + x_{(i)} Z(X(i) \delta_i.$$

3. Необходимое условие минимума:

$$(Z(X) + C_1) \gamma + C_1 y_* + o(\|\delta\|) = 0,$$

$$C_1 = A b_1 - A_1 b) / b^2, \quad A = Z(X(i) x_{(i)} x_i^{\mathrm{T}} Z(X(i)).$$

4. Векторная поправка γ к экстремуму y_* невозмущенного условия разделимости классов в первом приближении вычисляется по формуле

$$\gamma = Z^{-1}(X)(A_1 - b_1 E_n) Z(X(i)) y_* / b + o(\|\delta\|).$$

8. Определение изменения ширины полосы разделимости классов при возмущении определенной компоненты вектора признаков

Предполагая, что возмущения исходной матрицы не выводят за пределы окрестности экстремума условия разделимости, поправка к ширине полосы разделимости определяется выражением

$$h(x_{(i)} + \delta_i) - h(x_{(i)}) = ((y_* + \gamma)^{\mathrm{T}} R(X(i, a))(y_* + \gamma))^{-1/2} -$$

$$- (y_*^{\mathrm{T}} R(X) y_*)^{-1/2} = -D_1 h^3(x_{(i)}) / 2 + o(\|\delta\|),$$

где $h(x_{(i)}) = (y_*^{\mathrm{T}} R(X) y_*)^{-1/2}$ — ширина полосы при невозмущенном значении признака,

$$D_1=\gamma^{\mathrm{T}}R(X)\,y_* + y_*^{\mathrm{T}}R(X)\,\gamma + y_*^{\mathrm{T}}Dy_* + o(\|\delta\|),$$

$$D=(-(B_1^{\mathrm{T}}+B_1)\,b^2+((B+B^{\mathrm{T}})\,b_1+B_1c+c_1A)\,b-2b_1Ac)\,/\,b^3,$$

$$B_1=Z(X(i)(x_{(i)}\delta_i^{\mathrm{T}}+\delta_i x_i^{\mathrm{T}})R(X(i)),$$

$$B=Z(X(i)\,x_{(i)}x_i^{\mathrm{T}}R(X(i)),$$

$$c=1+x_{(i)}^{\mathrm{T}}R(X(i)x_{(i)},\ \ c_1=\delta_i^{\mathrm{T}}R(X(i)x_{(i)}+x_{(i)}R(X(i)\delta_i.$$

Для первого случая, когда вектор δ_i линейно зависим от векторов подпространства вектор-строк матрицы $X(i)$, поправка к ширине полосы разделимости вычисляется согласно выражению

$$h(x_{(i)}+\delta_i)-h(x_{(i)})=-D_2h^3(x_{(i)})\,/\,2+o(\|\delta\|),$$

$$D_2 = y_*^{\mathrm{T}}Dy_* + o(\|\delta\|).$$

9. Аналитический эксперимент

Если предположить, что возмущающий вектор δ_i линейно зависим от векторов подпространства вектор-строк матрицы $X(i)$, то множество экстремумов условия разделимости классов с учетом величины возмущения останется без изменений ($\gamma = 0$), а поправка к ширине полосы разделимости определяется формулой

$$h(x_{(i)}+\delta_i)-h(x_{(i)})=-2(\delta_i^{\mathrm{T}}R(X(i)\,x_{(i)})\,\|\,x_i^{\mathrm{T}}Z(X(i)y_*\,\|/\|\,x_i^{\mathrm{T}}Z(X(i)\,\|^{-4},$$

что обуславливает при некоторых возмущениях как расширение полосы разделимости, так и ее сужение.

Заключение

Псевдообращение является эффективным методом исследования в прикладных задачах, конструктивного описания решений задач, расширяющим возможности оптимизации структуры математической модели. В настоящей публикации метод классификации информации, предложенный в работе [5], дополняется расчетом погрешности метода в зависимости от возмущений исходной информации, оптимизацией процесса классификации.

Ф.Г. Гаращенко, Г.І. Кудін

ЗАДАЧІ КЛАСИФІКАЦІЇ ЗАСОБАМИ ПСЕВДООБЕРНЕННЯ: ПЕРЕТВОРЕННЯ ПРОСТОРУ ОЗНАК, ВПЛИВ ЗБУРЕНЬ ПЕРВИННОЇ ІНФОРМАЦІЇ

Для задачі гіперплощинної класифікації інформації засобами псевдообернення матриць запропоновано технології оптимізації простору векторів ознак. Отримано оцінки впливу збурень компонент векторів простору ознак на умови роздільності класів та ширину смуги роздільності. Проведені дослідження використовують результати теорії збурення псевдообернених і проекційних матриць.

 ISSN 0572-2691

F.G. Garashchenko, G.I. Kudin

THE PROBLEMS OF CLASSIFICATION BY MEANS OF PSEUDOINVERSION: THE TRANSFORMATION OF THE FEATURE SPACE, INFLUENCE OF PERTURBATIONS OF INITIAL INFORMATION

For the task of hyper plane classification of information by means of pseudoinversion of matrices the technologies for space optimization of feature vectors are offered. The estimates of influence of disturbances of feature space vectors components on conditions of separation of classes and width of a strip of separation of classes. The conducted researches are based on the results of the theory of perturbation of pseudo inverse and projective matrices.

1. *Сергиенко И.В., Химич А.Н., Яковлев Н.Ф.* Методы получения достоверных решений систем линейных алгебраических уравнений // Кибернетика и системный анализ. — 2011. — № 1. — С. 68–80.
2. *Кириченко Н.Ф.* Аналитическое представление возмущений псевдообратных матриц // Там же. — 1997. — № 2. — С. 98–107.
3. *Гаращенко Ф.Г., Волошин О.Ф., Кириченко М.Ф. та ін.* Розвиток методів і технологій моделювання та оптимізації складних систем: — Київ : Сталь, 2009. — 668 с.
4. *Кириченко Н.Ф., Кривонос Ю.Г., Лепеха Н.П.* Оптимизация синтеза гиперплоскостных кластеров и нейрофункциональных преобразований в системах классификации сигналов // Кибернетика и системный анализ. — 2008. — № 6. — С. 107–124.
5. *Кириченко Н.Ф., Кривонос Ю.Г., Лепеха Н.П.* Синтез систем нейрофункциональных преобразователей в решении задач классификации // Там же. — 2007. — № 3. — С. 47–57.
6. *Кириченко Н.Ф., Кудин Г.И.* Анализ и синтез систем классификации сигналов средствами возмущений псевдообратных и проекционных операций // Там же. — 2009. — № 3. — С. 47–57.
7. *Кириченко Н.Ф., Кудин Г.И.* Вариации псевдообратных и проекционных матриц при возмущении элементов исходной матрицы // Проблемы управления и информатики. — 2009. — № 4. — С. 7–20.
8. *Кривонос Ю.Г., Кудин Г.И., Лепеха Н.П.* Оптимальный синтез систем классификации средствами преобразований пространства признаков // Материалы Междунар. науч.-техн. конф. «Искусственный интеллект», 20–25 сентября, 2010 г. — Кацивели, Украина. — 2010. — Т. 1. — С. 238–241.
9. *Кудін Г.І., Зінько Т.М.* Синтез кусково гіперплощинних кластерів для навчальних виборок векторів у просторі ознак // Вісник Київського університету. Сер. фіз.-мат. науки. — 2017. — № 2. — С. 233–237.
10. *Ивахненко А.Г., Юрачковский Ю.П.* Моделирование сложных систем по экспериментальным данным. — М. : Радио и связь, 1987. — 239 с.

Получено 14.12.2017