КОМПЬЮТЕРНАЯ МАТЕМАТИКА

Теория и методы оптимизации

УДК 519.685.3

А.С. ГАВРИЛЕНКО

РАЗРАБОТКА АЛГОРИТМА СИНТЕЗА СИСТЕМ ДЛЯ РЕШЕНИЯ ЗАДАЧ КЛАССИФИКАЦИИ СИГНАЛОВ

Представлено усовершенствование линейных алгоритмов синтеза систем распознавания принадлежности сигналов к классу путем уточнения информативности выбранных компонент вектора признаков и при необходимости замены соответствующих компонент нелинейными трансформациями.

Введение. Задачам проектирования систем классификации сигналов уделяется достаточно много внимания и в этой области достигнуты значительные результаты [1–3]. Однако, поскольку эта задача в общем случае не имеет универсального решения и эффективно решается с учетом специфики каждой конкретной предметной области, данная проблема сохраняет свою актуальность и простор для исследований и усовершенствований.

В данной работе для линейных алгоритмов синтеза систем распознавания принадлежности сигналов к тому или иному классу [4] предлагается усовершенствование путем уточнения вектора признаков с учетом информативности выбранных компонент в пространстве признаков и при необходимости их замены нелинейными трансформациями соответствующих компонент. Для этого используются методы, разработанные на основе теории возмущения псевдообратных и проекционных матриц, а также обобщенные полиномы, описанные в [5]. Рассматриваемые операции допускают использование их в суперпозиции, а также, что представляется особенно важным в прикладном отношении, в форме каскадной дихотомной классификации точек в пространстве признаков.

Постановка задачи

Пусть, $x(j)\in R^m$, $j=\overline{1,N}$ полученное в результате эксперимента обучающее множество $\{x(1),\ldots,x(N)\}$ векторов $x=\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$ характерных признаков рассматриваемых сигналов.

В дальнейшем обучающее множество будем представлять в виде матрицы

$$X = \left(x(1),\ldots,x(N)\right) = \begin{pmatrix} x_{(1)}^T \\ \vdots \\ x_{(m)}^T \end{pmatrix}.$$

Рассмотрим, не нарушая общности, задачу определения принадлежности сигнала из представленного множества к одному из двух классов, которые в обучающей выборке представлены множествами Ω_1 и Ω_2:

$$x(i_1),\ldots,x(i_{N_1}) \in \Omega_1,\ x(j_1),\ldots,x(j_{N_2}) \in \Omega_2,$$

$$x(i_l),\, l = \overline{1,N_1}, \quad x(j_k),\, k = \overline{1,N_2}, \quad N_1 + N_2 = N.$$

С целью формирования над компонентами x неоднородных линейных операций будем предполагать $x_m = 1$.

В линейной задаче синтеза системы классификации дискриминантная функция имеет вид $y = a^T x$, для которой необходимо найти такой вектор $a_\Delta \in R^m$, чтобы

$$\begin{cases} a_\Delta^T x(i_s) \ge \Delta & s = \overline{1,N}_1 \\ a_\Delta^T x(j_k) \le -\Delta & k = \overline{1,N}_2 \end{cases}, \tag{1}$$

либо найти такие векторы $a_\Delta \in R^m$ и $y \in R^N$, для которых

$$a_\Delta^T x(j) = y(j),\ \ j = \overline{1,N} \quad \text{и} \quad \begin{cases} y(i_s) \ge \Delta & s = \overline{1,N_1} \\ y(j_k) \le -\Delta & k = \overline{1,N}_2 \end{cases} \tag{2}$$

при некотором заданном $\Delta > 0$.

Теоретическая основа алгоритма

Решение линейной задачи. Обозначим область значений y, удовлетворяющих условию

$$y(j_1) = a^T x(j_1) \ge \Delta \ \ \forall \ \ x(j_1) \in \Omega_x^{(1)}$$
$$\text{и } \ y(j_2) = a^T x(j_2) \le -\Delta \ \ \forall \ \ x(j_2) \in \Omega_x^{(2)}.$$

как $\Omega_y(\Delta)$.

С целью унификации условий для дискриминантной функции $y(j) \ge \Delta,\ j = 1,\ldots,n$ сделаем замену знака для векторов $x(j_2)$, принадлежащих $\Omega_x^{(2)}$. Тогда задачу можно переформулировать следующим образом: определить векторы $a \in R^m$ и $y \in \Omega_y(\Delta)$, удовлетворяющие условиям

$$a^T x(j) = y(j),\ \forall\ \ j = \overline{1, N},$$

или

$$X^T a = y\,,\ y = \begin{pmatrix} y(1) \\ \vdots \\ y(N) \end{pmatrix}. \tag{3}$$

Исходя из необходимого и достаточного условия разрешимости данной задачи [6, 7], а именно:

$$y_\Delta^T Z(X) y_\Delta = \min_{y|y(j)\geq\Delta,\ j=1,\ldots,n} y^T Z(X) y = 0\,, \tag{4}$$

где $Z(X) = I - X^+ X$ – матрица проектирования на ортогональное дополнение к линейной оболочке, натянутой на вектор-строки матрицы X в R^n, значение y_Δ будем находить при некотором фиксированном Δ путем минимизации квадратичной функции с ограничениями

$$\min_{y|y(j)\geq\Delta,\ j=1,\ldots,n} y^T Z(X) y\ .$$

Тогда искомое значение вектора a можно найти, решая систему уравнений

$$X^T a_\Delta = y_\Delta\,,$$

а именно:

$$a_\Delta = X^{T^+} y_\Delta.$$

Необходимым и достаточным условием существования полученного решения линейной задачи классификации предполагалось выполнение условия (4) ([4], [6]). Если в процессе решения задачи условие (4) не выполняется, а именно $y_\Delta^T Z(X) y_\Delta > 0$, то линейная аппроксимация разделяющей поверхности гиперплоскостью невозможна. В этом случае целесообразно либо рассматривать задачу уточнения выбора пространства признаков, либо задачу поиска нелинейной аппроксимации разделяющей поверхности путем замены переменных такой суперпозицией нелинейных преобразований в рассматриваемом пространстве признаков, для которой выполняются условия (1).

Условия определения информативности компонент. Более удачный выбор пространства признаков при сохранении его размерности можно осуществить, используя результаты [4, 5], полученные средствами теории возмущения псевдообратных и проекционных матриц. А именно, если $q(i)$ – i-ый вектор-столбец матрицы X^+ и для некоторого $i = i_*$ выполняется условие

$$x^T(i_*) q(i_*) \neq 1, \tag{5}$$

то i_*-ая компонента вектора признаков x является неинформативной. Если же

$$x^T(i)q(i)=1,\ \forall i=\overline{1,m}, \tag{6}$$

то наименее информативная i_*-ая компонента вектора признаков x определяется из условия

$$\frac{\left|{y_*}^T q(i_*)\right|^2}{\left\|q(i_*)\right\|^2}=\min_{i=\overline{1,m}}\frac{\left|{y_*}^T q(i)\right|^2}{\left\|q(i)\right\|^2}. \tag{7}$$

Выбор новой компоненты x_0 в векторе признаков x и замена наименее информативной компоненты x_{i^*} этой компонентой x_0, т. е. рассмотрение нового вектора признаков $\left(x_1,\ \dots, x_{i^*-1},\ x_0,\ x_{i^*+1}, \dots, x_m\right)^T$ вместо старого $x=$ $=\left(x_1,\ \dots,x_m\right)^T$, осуществляется согласно следующему условию оптимальности:

$${y_*}^T Z(X_{(i^*,x_{(0)})})y_*\to\min_{x_0\in P_x}, \tag{8}$$

$$Z(X_{(i^*,x_{(0)})})=Z(X_{i_*})-\frac{Z(X_{i_*})x_{(0)}x_{(0)}^T Z(X_{i_*})}{\left\|Z(X_{i_*})x_{(0)}\right\|^2}, \tag{9}$$

$$Z(X_{i_*})=\begin{cases}Z(X)+\dfrac{q_{i_*}q_{i_*}^T}{\left\|q_{i_*}\right\|^2}, & x_{(i_*)}^T q_{i_*}=1\\ Z(X), & x_{(i_*)}^T q_{i_*}\neq 1\end{cases}. \tag{10}$$

Качество классификации сигналов вышеописанными средствами можно улучшить за счет нелинейной трансформации выбранного пространства признаков. Если согласно (7) или (5) – (7) i_*-ая координата пространства признаков x_{i_*} является наименее информативной, тогда осуществляется нелинейная трансформация этой компоненты $x_{i_*}(j)$ $j=\overline{1,n}$ каждого входного вектора $x(j)=\left(x_1(j),\ \dots,x_{i_*}(j),\dots,x_m(j)\right)^T$ и замены ее на значения некоторой нелинейной функции $\psi_0(x_{i_0}(j))$ от компоненты x_{i_0}.

Следует ожидать [5], что нелинейные преобразования над сигналом $\hat{y}_1(j)$ приведут к улучшению значения y_*

Синтез нелинейной системы классификации

На основании вышеизложенных идей синтеза функциональных преобразований приведем описание средств формирования конкретных систем классификации сигналов.

Организация класса базисных функций для построения классификатора.

Класс базисных функций $y = \varphi(x; a)$, из которого функции будут использоваться для построения покомпонентных наименее информативных компонент нелинейных преобразований выбранного пространства признаков, включает в себя наиболее часто используемые в практике аппроксимации функции, такие как $y = ax + b$, $y = ax^2 + bx + c$, полином n-й степени $y = P_n(x; a)$ и т. п.

Представляет интерес класс нелинейных преобразований $y = \varphi(x; a)$, построенных на базе классов функций часто используемых при построении эмпирических формул, которые позволяют некоторым преобразованием x и y сводить задачу определения коэффициентов a к решению простых линейных систем. Примеры таких классов функций, используемые в рассматриваемой системе классификации, и соответствующие преобразования приведены в таблице.

В ряде случаев полезно использовать эрмитовы кубические сплайны $S_3(x)$ [8, 9] для нелинейного преобразования s-й компоненты.

ТАБЛИЦА

Базисные функции	Преобразование координат	Преобразованная функция $Y = AX + B$	Преобразование коэффициентов
$y = \dfrac{1}{ax+b}$	$X = x;\ Y = 1/y$	$Y = ax + b$	$a = A;\ b = B$
$y = \dfrac{a}{x} + b$	$X = 1/x;\ Y = y$	$Y = aX + b$	$a = A;\ b = B$
$y = \dfrac{x}{ax+b}$	$X = x;\ Y = x/y$	$Y = ax + b$	$a = A;\ b = B$
$y = \dfrac{1}{ax^2+bx+c}$	$X = x;\ Y = 1/y$	$Y = ax^2 + bx + c$	$a = A;\ b = B;\ c = C$
$y = \dfrac{x}{ax^2+bx+c}$	$X = x;\ Y = x/y$	$Y = ax^2 + bx + c$	$a = A;\ b = B;\ c = C$
$y = a + \dfrac{b}{x} + \dfrac{c}{x^2}$	$X = 1/x;\ Y = y$	$y = a + bX + cX^2$	$a = A;\ b = B;\ c = C$
$y = ax^b$	$X = \ln x;\ Y = \ln y$	$Y = \ln a + b \cdot \ln x$	$a = e^A;\ b = B$
$y = ab^x$	$X = x;\ Y = \ln y$	$Y = \ln a + x \cdot \ln b$	$a = e^A;\ b = e^B$
$y = ae^{bx}$	$X = x;\ Y = \ln y$	$Y = \ln a + b \cdot x$	$a = e^A;\ b = B$
$y = ae^{-bx^2}$	$X = x^2;\ Y = \ln y$	$Y = \ln a - b \cdot x^2$	$a = e^A;\ b = B$
$y = ax^b e^{cx}$	$X = \ln x;\ Y = \ln y$	$Y = \ln a + b \cdot \ln x + cx$	$a = e^A;\ b = B; c = C$

Естественно, класс базисных функций должен быть открытым для расширения новыми базисными функциями.

Построение аппроксимации дискриминантных функций $\Phi(X)$ как функции многих переменных возможно с помощью операций сложения и суперпозиции представленных классов функций одного переменного. А именно:

а) в виде обобщенного полинома $y = a_0 + \sum_{i=1}^{m} a_i \varphi_i(x)$, где $\varphi_i(x)$ – базисные функции;

б) в виде суперпозиции $y = \Phi\left(\varphi(x)\right)$;

в) $y = a_0 + \sum_{i=1}^{m} a_i \Phi_i(\varphi(x))$;

г) $y = \Phi\left(a_0 + \sum_{i=1}^{m} a_i \varphi_i(x)\right)$.

Схема алгоритма. В описании алгоритма будут приняты следующие обозначения. Обучающая выборка – массив точек $x(j_1) \in \Omega_x(1)$, $j_1 = \overline{1, N_1}$ $x(j_2) \in \Omega_x(2)$, $j_2 = \overline{1, N_2}$, представляется соответственно матрицами $X_e^{(1)}$ и $X_e^{(2)}$, дополненными первым вектор-столбцом, заполненным единицами. Экспериментальные данные будут представлены матрицами вида

$$X_e = \begin{pmatrix} X_e^{(1)} \\ X_e^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & x_1^{(1)}(1) & \dots & x_m^{(1)}(1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(1)}(N_1) & \dots & x_m^{(1)}(N_1) \\ 1 & x_1^{(2)}(1) & \dots & x_m^{(2)}(1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(2)}(N_2) & \dots & x_m^{(2)}(N_2) \end{pmatrix},$$

$$X = \begin{pmatrix} 1 & x_1^{(1)}(1) & \dots & x_m^{(1)}(1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(1)}(N_1) & \dots & x_m^{(1)}(N_1) \\ -1 & -x_1^{(2)}(1) & \dots & -x_m^{(2)}(1) \\ \vdots & \vdots & \vdots & \vdots \\ -1 & -x_1^{(2)}(N_2) & \dots & -x_m^{(2)}(N_2) \end{pmatrix}.$$

Тогда схему алгоритма можно представить следующим образом.

1. Зададим некоторое начальное значение $d > 0$ и сформируем первое приближение $y = (\Delta_1, \dots, \Delta_N)$, $\Delta_j = d$, $j = \overline{1, N}$.

2. Решая систему $N \times m + 1$ уравнений $y = Xa$, получаем вектор коэффициентов $\hat{a} = X^+ y$.

3. Вычисляем:

а) вектор модельных значений $y_M = X_e \hat{a}$;

б) дискриминантную функцию для точек, условно отнесенных к первой и второй группам:

$$\begin{cases} y_e^{(1)} = X_e^{(1)} \hat{a}, \\ y_e^{(2)} = X_e^{(2)} \hat{a}; \end{cases}$$

в) невязку $y_M - y$ и среднеквадратичное отклонение $\left\| y_M - y \right\|^2 = y^T Z(X) y$.

4. В случае, если выполняется условие $\begin{cases} y_e^{(1)} \geq \Delta \\ y_e^{(2)} \leq -\Delta \end{cases}$, осуществляется выход из алгоритма.

5. Иначе проводим уточнение значения вектора допусков $y^T = (\Delta_1, ..., \Delta_N)$, $\Delta_j = d$, $j = \overline{1, N}$, минимизируя квадратичную форму $y^* = \arg\min\limits_{y \in D} y^T Z(X) y$ градиентным методом с ограничением $y \in D$, где $D = \left\{ y : d \leq \Delta_j \leq d', \ \ j = \overline{1, N} \right\}$ (положим $d' = 10^6$).

6. Зная y^*, находим уточненное значение вектора коэффициентов линейной регрессии $\hat{a}_* = X^+ y^*$.

7. В случае, если выполняется условие, что проекция y_M^* на ортогональное дополнение к объединенному массиву экспериментальных данных X :

$$y_M^{*\,T} Z(X) y_M^* < \varepsilon,$$

(положим точность приближения $\varepsilon = 10^{-6}$), где $y_M^* = X\hat{a}_*$ – уточненный вектор модельных значений, осуществляется выход из алгоритма.

8. Иначе проводим уточнение выбора пространства признаков $X = (x_1, \dots, x_m)$.

Определяем наименее информативную координату x_s, используя условия (5) или (7).

9. Осуществляем замену выбранной наименее информативной координаты

$$x_s = \begin{pmatrix} x_s(1) \\ \vdots \\ x_s(N_1) \\ -\tilde{x}_s(1) \\ \vdots \\ -\tilde{x}_s(N_2) \end{pmatrix},$$

нелинейной трансформацией $\psi(x_s)$ такой, что $y_*^T Z(X_{s,\psi}) y_* = = \min\limits_{\varphi \in \Phi} y_*^T Z(X_{s,\varphi}) y_*$, где

$$X_{s,\varphi} = \begin{pmatrix} x_1 \\ \vdots \\ \varphi(x_s) \\ \vdots \\ x_m \end{pmatrix},$$

причем выбор ψ из класса функций $\Phi(x)$, $x \in R^1$ осуществляется с нормализацией по диапазону значений компоненты x_s. Следует учесть, что нелинейная трансформация осуществляется над экспериментальными значениями X_e.

10. В случае, если выполняется условие $y_*^{\psi T} Z(X_{S,\psi}) y_*^{\psi} < \varepsilon$, где $y_*^{\psi} = \arg\min\limits_{y \in D} y^T Z(X_{S,\psi}) y$, осуществляется выход из алгоритма.

11. Иначе в матрице экспериментальных данных X_e проводится замена значений x_S на $\psi(x_S)$ с последующей заменой знака аналогично шагу 1, после чего повторяются шаги 8–9.

12. Иначе проводится замена значений x_S на $\sum\limits_{i \in I} c_i x_i$, где $I = \{1, \ldots, s-1, s+1, \ldots, m+1\}$, $c = (c_i), i \in I$ такое, что $y_*^T Z(X_{s,c}) y_* < \varepsilon$,

где $X_{s,c} = \begin{pmatrix} x_1 \\ \vdots \\ \sum\limits_{i \in I} c_i x_i \\ \vdots \\ x_m \end{pmatrix}$, с последующей заменой знака аналогично шагу 1,

после чего повторяются шаги 8–9.

13. Иначе проводится замена значений x_S на $\psi(\sum_{i\in I} c_i x_i)$, где $\psi \in \Phi$,

$c = (c_i)$, $i \in I$ такое, что $y_*^T Z(X_{s,\psi,c}) y_* < \varepsilon$, где $X_{s,\psi,c} = \begin{pmatrix} x_1 \\ \vdots \\ \psi(\sum_{i\in I} c_i x_i) \\ \vdots \\ x_m \end{pmatrix}$,

после чего повторяются шаги 8–9.

Заключение. Представленный в работе усовершенствованный алгоритм синтеза систем распознавания полезен в случаях, когда нет возможности найти решение задачи классификации, оставаясь в рамках линейной модели. Уточнение информативности признаков позволяет исключить из результирующей системы избыточную информацию и осуществить нелинейную трансформацию с минимизацией результирующей невязки. Рассмотренные алгоритмы синтеза нелинейных преобразований с автоматическим поиском на заданных классах функциональных преобразователей обеспечивают расширенные возможности и гибкий инструментарий для решения самых разнообразных задач классификации. Приведенные в [5] результаты численных экспериментов позволяют утверждать, что данный подход перспективен и может быть с успехом использован на практике.

А.С. Гавриленко

РОЗРОБКА АЛГОРИТМА СИНТЕЗУ СИСТЕМ ДЛЯ РОЗВ'ЯЗКУ
ЗАДАЧ КЛАСИФІКАЦІЇ СИГНАЛІВ

Представлено удосконалення лінійних алгоритмів синтезу систем розпізнавання приналежності сигналів до класу шляхом уточнення інформативності обраних компонент вектора ознак та за необхідності заміни відповідних компонент нелінійними трансформаціями.

A.S. Gavrylenko

DEVELOPMENT OF SYSTEM SYNTHESIS ALGORYTHM FOR SIGNAL CLASSIFICATION
PROBLEM SOLVING

An improvement of linear synthesis algorithms of the systems of recognition of belonging signals to a certain class by refinement of self-descriptiveness of selected components of attribute vector and in the presence of necessity of replacement of the proper components by nonlinear transformations is presented.

1. *Carmen Lai, David M.J. Tax, Robert P.W. Duin, Elżbieta Pękalska, Pavel Paclík.* On combining one-class classifiers for image database retrieval // Lecture Notes In Computer Science; Vol. 2364. Proceedings of the Third International Workshop on Multiple Classifier Systems. – 2002. – P. 212 – 221.
2. *Форсайт П.* Компьютерное зрение. Современный подход. Пер. с англ. – М.: Изд-во «Вильямс», 2004. – 928 с.
3. *Gonzalez R.C., Woods R.E., Eddins S.L.* Digital Image Processing Using MATLAB(R), 2nd Edition Gatesmark Publishing, 2009.
4. *Кириченко Н.Ф., Крак Ю.В., Полищук А.А.* Псевдообратные и проекционные матрицы в задачах синтеза функциональных преобразователей // Кибернетика и системный анализ. – 2004. – № 3. – С. 116–129.
5. *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.
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 // Кибернетика и системный анализ. – 2002. – № 4. – С. 107–124.
8. *Роджерс Д., Адамс Дж.* Математические основы машинной графики. – М.: Мир, 2001. – 604 с.
9. *Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л.* Методы сплайн-функций. – М.: Наука, 1980. – 350 с.

Получено 14.12.2009

Об авторе:

Гавриленко Анастасия Сергеевна,
младший научный сотрудник, аспирантка
Института кибернетики имени В.М. Глушкова НАН Украины.
e-mail ngav@dept115@icyb.kiev.ua