КОМПЬЮТЕРНАЯ
МАТЕМАТИКА

Теория и методы оптимизации

УДК 519.685.3

А.С. ГАВРИЛЕНКО

АЛГОРИТМИ СИНТЕЗУ КЛАСИФІКАТОРІВ У ЗАДАЧАХ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ ЗАСОБАМИ ЛІНІЙНИХ І НЕЛІНІЙНИХ ОПТИМАЛЬНИХ ПЕРЕТВОРЕНЬ ПРОСТОРУ ОЗНАК

У роботі представлено алгоритм синтезу систем класифікації, розроблений на базі лінійних і нелінійних оптимальних перетворень простору ознак та реалізований із застосуванням методу активного набору для задачі квадратичного програмування та елементів методу побудови оптимального нелінійного перетворювача як узагальненого полінома на заданих класах функцій, а також показано результат роботи алгоритму на прикладі розв'язування задачі розпізнавання звукових сигналів.

Вступ. У задачах класифікації сигналів, незалежно від предметної області, в якій вони поставлені, широко застосовуються методи синтезу систем класифікації, засновані на автоматичному навчанні системи за допомогою деякої тренувальної вибірки сигналів. Стан проблеми та деякий огляд публікацій за даною темою висвітлено в роботах [1, 2]. У роботі [1] в розвиток поставлених задач попередньої статті [2] даються теоретичні обгрунтування можливостей і умови лінійної та лінійної полосної відокремлюваності множин точок за класами розпізнаваних об'єктів, а також представлені принципи побудови оптимальних алгоритмів розпізнавання засобами лінійних і нелінійних перетворень простору ознак. Ці результати отримані на основі апарату теорії збурення псевдообернених і проекційних матриць [3]. Питанням практичного використання такого типу алгоритмів і їх програмній реалізації на C++, а також апробації на конкретних задачах присвячені роботи [4, 5].

У даній роботі автор представляє розробку нових алгоритмів на базі втілення ідей лінійних і нелінійних оптимальних перетворень простору ознак, які розширюють можливості розроблених програмних продуктів синтезу класифікаторів, що і демонструється шляхом розв'язування нових практичних задач.

Для цього програмне забезпечення було доповнене реалізацією методу активного набору для задачі квадратичного програмування, до якої зводиться шляхом сингулярного перетворення матриці простору ознак задача оптимальної лінійної відокремлюваності точок навчаючої вибірки, а також реалізацією методу побудови оптимального нелінійного перетворювача як узагальненого полінома на заданих класах функцій.

Розглянуті операції допускають використання їх в суперпозиції, а також, що видається особливо важливим у прикладному відношенні, у формі каскадної дихотомної класифікації точок у просторі ознак. У роботах [1, 2] розглянута теоретична постановка як лінійної, так і нелінійної задачі розділення сигналів на класи і запропонований математичний апарат для її вирішення. В роботі [4] представлений розроблений алгоритм лінійного синтезу з подальшою нелінійною оптимізацією у разі неефективності базового методу.

Слід зауважити, що як методи, описані в цитованій літературі, так і ті, що розглядатимуться далі, базуються на побудові деякої дискримінантної гіперплощини або гіперповерхні, які являються роздільним бар'єром між кластерами навчальних виборок для окремих об'єктів. Проте можливо будувати алгоритми розпізнавання і на принципах побудови для кожного класу об'єктів деякого контейнера-канонічного тіла, який описується достатньо простим рівнянням для перевірки приналежності йому тієї чи іншої точки з вибірки, що підлягає розпізнаванню [5]. Розміри та орієнтація такого контейнера можуть легко варіюватися в процесі навчання алгоритму за умови, що контейнери, поставлені у відповідність різним класам об'єктів, не перетинаються, а сам алгоритм, побудований на такому принципі, суттєво спрощується.

Постановка задачі. Приведемо постановку задачі оптимального синтезу лінійних гіперплощинних кластерів та розглянемо в її межах задачу оптимальної лінійної полосної розділюваності двох множин точок, що відповідають характерним ознакам двох класів об'єктів. Для більшого числа класів об'єктів результати можуть бути узагальнені як на бінарні задачі багаторівневої класифікації, до чого ми ще повернемося.

Нехай $x(j) \in R^m$, $j = \overline{1...N}$ – отримана в результаті експерименту навчальна вибірка векторів характерних ознак сигналів, що підлягають класифікації. Навчальну вибірку далі будемо представляти у вигляді матриці

$X = \left(x(1) \dots x(N) \right) = \begin{pmatrix} x_{(1)}^T \\ \vdots \\ x_{(m)}^T \end{pmatrix}$ розмірністю $m \times N$ (тут покладено $x_m(j) = 1$),

а навчальні множини Ω_1 та Ω_2 у вигляді:

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

$$x(i_k),\, k = \overline{1, N_1}, \quad x(j_s),\, s = \overline{1, N_2}, \quad N_1 + N_2 = N,$$

тобто N_1 точок $x(i_k)$ належать першому класу, а N_2 точок $x(j_s)$ – другому класу.

Дискримінантна функція для лінійної задачі класифікації має вигляд $y = a^T x$. Тоді під лінійною полосною розділюваністю цих класів будемо розуміти факт існування такого вектора $a \in R^m$, для якого дискримінантна функція $y(j) = a^T x(j)$ на точках першого класу більша 1, а на точках другого класу – менша –1:

$$\begin{aligned} a^T x(i_k) \geq 1, \quad & k = \overline{1, N_1}, \\ a^T x(j_s) \leq -1, \quad & s = \overline{1, N_2}. \end{aligned} \tag{1}$$

Згідно з [1], умова існування лінійної полосної розділюваності множин точок двох об`єктів має вигляд

$$\min_{y \in D} y^T Z(X) y = 0, \text{ де } D = \left\{ y : e_{i_k}^T y \geq 1, e_{j_s}^T y \leq -1, k = \overline{1, N_1}, s = \overline{1, N_2} \right\}, \tag{2}$$

а $N \times N$ матриця $Z(X) = I_N - X^+ X$ – проекційний оператор на ортогональне доповнення до лінійної оболонки, натягнутої на вектор-стовпчики матриці X. Тобто, ця умова перевіряє, чи лежить вектор y в указаній лінійній оболонці. А $y^T Z(X) y = 0$ – це квадрат норми проекції вектора y на це ортогональне доповнення (або з іншої сторони це є квадрат норми нев'язки $\left\| y - a^T x \right\|^2$).

Враховуючи, що ширина роздільної смуги δ визначається як значення $\delta = \dfrac{\Delta}{(y_*^T(\Delta) R(X) y_*(\Delta))^{\frac{1}{2}}}$, (див. [1] і тут прийнято $\Delta = 1$) можна заключити, що максимальна ширина розділюючої смуги досягається при значеннях

$$y_{opt} = \arg\min_{y \in D_1} y^T R(X) y, \quad a_{opt} = \left(X^T \right)^+ y_{opt}, \tag{3}$$

де $D_1 = \left\{ y : y^T Z(X) y = 0 \right\} \cap D$, а $R(X) = X^+ (X^T)^+$ – проекційна матриця $N \times N$.

Дана задача є досить складною для практичного застосування, оскільки мінімізувати квадратичну функцію $y^T R(X) y$ необхідно на розв'язках задачі (2). Тому використовуючи сингулярний розклад матриць (SVD) і властивості цього розкладу для $m \times N$ матриці X:

$$\begin{aligned} & X = \sum_{j=1}^{r} u_j v_j^T \lambda_j, X^+ = \sum_{j=1}^{r} v_j u_j^T \lambda_j^{-1}, Z(X) = I_N - \sum_{j=1}^{r} v_j v_j^T, \ u_i^T u_j = \delta_{ij}, v_i^T v_j = \delta_{ij}, \\ & X X^T u_j = \lambda_j^2 u_j, X^T X v_j = \lambda_j^2 v_j, \lambda_1^2 \geq \ldots \geq \lambda_r^2 > 0, \ \lambda_{r+1} = \ldots = \lambda_N = 0, i, j = \overline{1, N}, \end{aligned} \tag{4}$$

та враховуючи, що $y^T Z(X) y = 0$ для $y \in D$, можна записати наступні співвідношення:

$$y = \sum_{i=1}^{r} \alpha_i v_i, \quad \sum_{i=1}^{r} \alpha_i e_j^T v_i \geq 1, \quad \forall j = \overline{1, N},$$

$$y^T R(X)y = y^T X^+ X^{+T} y = y^T \sum_{j=1}^{r} v_j v_j^T \lambda_j^{-2} y =$$

$$= \sum_{i=1}^{r} \alpha_i v_i^T \sum_{j=1}^{r} v_j v_j^T \lambda_j^{-2} \sum_{k=1}^{r} v_k \alpha_k =$$

$$= \sum_{j=1}^{r} \alpha_j^2 \lambda_j^{-2} == \alpha^T \mathrm{diag}(\lambda_1^{-2},\dots,\lambda_r^{-2})\alpha. \quad (5)$$

Тобто, використовуючи SVD-перетворення $X = U\Sigma V^T$ і перейшовши в простір власних векторів матриці X задачу пошуку (3) звели до розв'язку задачі оптимізації квадратичної функції (при $r = m$)

$$y^T R(X)y = (\alpha_1,\dots,\alpha_r)\begin{pmatrix} \lambda_1^{-2} & 0 & \dots & 0 \\ 0 & \lambda_2^{-2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_r^{-2} \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \dots \\ \alpha_r \end{pmatrix} = \sum_{i=1}^{r} \frac{\alpha_i^2}{\lambda_i^2},$$

а обмеження будуть мати вигляд

$$\sum_{i=1}^{r} \alpha_i v_i(k) \ge 1, \quad k = \overline{1, N_1},$$

$$\sum_{i=1}^{r} \alpha_i v_i(s) \le -1, \quad s = \overline{1, N_2},$$

де v_i – власні вектор-стовпчики розмірності N, що утворюють матрицю V у SVD-розкладі: $V = (v_1 \vdots \dots \vdots v_r)$.

Розв'язавши цю задачу квадратичного програмування, отримаємо

$$\alpha_{opt} = \arg \min_{\alpha \in \left\{\alpha : e_{ik}^T (v_1 \vdots \dots \vdots v_r)\alpha \ge 1, e_{js}^T (v_1 \vdots \dots \vdots v_r)\alpha \le -1, k=\overline{1,N_1}, s=\overline{1,N_2}\right\}} \alpha^T \mathrm{diag}\left(\lambda_1^{-2},\dots,\lambda_r^{-2}\right)\alpha, \quad (6)$$

Після чого можна обчислити

$$y_{opt} = (v_1 \vdots \dots \vdots v_r)\,\alpha_{opt}, \quad (7)$$

і коефіцієнти дискримінантної гіперплощини $y = a^T x$

$$a_{opt} = (u_1 \vdots \dots \vdots u_r)\,\mathrm{diag}\left(\lambda_1^{-1},\dots,\lambda_r^{-1}\right)\alpha_{opt}. \quad (8)$$

Розглянемо тепер покроковий алгоритм розв'язку задачі лінійної полосної роздільності двох множин точок з урахуванням вищевиведених співвідношень.

Схема алгоритму. В описі алгоритму будуть прийняті наступні позначення. Навчальна вибірка – масив точок $x(i_k) \in \Omega_x(1)$, $k = \overline{1,\dots,N_1}$ $x(j_s) \in \Omega_x(2)$, $s = \overline{1,\dots,N_2}$, представляється відповідно матрицями $X_e^{(1)}$ та $X_e^{(2)}$, доповненими першим вектор-стовпчиком, заповненим одиницями. Експериментальні дані будуть представлені матрицями вигляду

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

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

Тоді схему алгоритму можна представити наступним чином.

1. Задамо деяке початкове значення $d > 0$ та сформуємо перше наближення $y = (\Delta_1 \ldots \Delta_N),\, \Delta_j = d,\; j = \overline{1, \ldots, N}$.

2. Розв'язуючи систему $N \times (m+1)$ рівнянь $y = Xa$, отримаємо вектор коефіцієнтів $\hat{a} = X^{+}y$.

3. Знаходимо:

а) вектор модельних значень $y_M = X_e\hat{a}$;

б) дискримінантну функцію для точок, умовно віднесених до першої та другої груп:

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

в) нев'язку $y_M - y$ та середньоквадратичне відхилення $\|y_M - y\|^2 = y^T Z(X) y$.

4. У випадку, якщо виконується умова $\begin{cases} y_e^{(1)} \geq \Delta \\ y_e^{(2)} \leq -\Delta \end{cases}$, здійснюється вихід з алгоритму.

5. Інакше проводимо уточнення значення вектора допусків $y^T = (\Delta_1 \ldots \Delta_N)$, $\Delta_j = d,\; j = \overline{1, \ldots, N}$, мінімізуючи квадратичну форму

$y^* = \arg\min_{y \in D} y^T Z(X)y$ градієнтним методом з обмеженням $y \in D$, де $D = \left\{ y : d \le \Delta_j \le d', j = \overline{1, ..., N} \right\}$ (покладемо $d' = 10^6$).

6. Знаючи y^*, знаходимо уточнене значення вектора коефіцієнтів лінійної регресії $\hat{a}_* = X^+ y^*$.

7. У випадку, якщо виконується умова, що проекція y_M^* на ортогональне доповнення до об'єднаного масиву експериментальних даних X $y_M^{*\,T} Z(X) y_M^* < \varepsilon$, де ε – прийняте значення машинного нуля, $y_M^* = X\hat{a}_*$ – уточнений вектор модельних значень, роздільна площина вважається знайденою.

8. Далі, застосовуючи вищенаведені умови лінійної віддільності точок в R^m до поставленої задачі, представимо лінійну полосну роздільність двох класів навчальної вибірки у вигляді існування такого вектора $a \in R^m$, для якого

$$a^T x(i_k) \ge 1, \quad k = \overline{1, N_1},$$
$$a^T x(j_s) \le -1, \quad s = \overline{1, N_2}.$$

Тоді умова лінійної полосної роздільності приймає вигляд

$$\min_{y \in D} y^T Z(X) y = 0, \quad D = \left\{ y : e_{i_k}^T y \ge 1, e_{j_s}^T y \le -1, k = \overline{1, N_1}, s = \overline{1, N_2} \right\}.$$

9. Значення вектора коефіцієнтів a знаходимо через умову максимізації ширини роздільної смуги:

$$a_{opt} = \left(X^T\right)^+ y_{opt}, \text{ де } y_{opt} = \arg\min_{y \in D_1} y^T R(X) y, \quad D_1 = \left\{ y : y^T Z(X) y = 0 \right\} \cap D .$$

Можна побачити, що для обчислення оптимального значення a необхідно розв'язати задачу квадратичного програмування

$$\min_{\alpha \in \Delta} y^T R(X) y = \sum_{i=1}^{r} \frac{\alpha_i^2}{\lambda_i^2}$$

при обмеженнях Δ у вигляді лінійних нерівностей:

$$\sum_{i=1}^{r} \alpha_i v_i(k) \ge 1, \quad k = \overline{1, N_1},$$
$$\sum_{i=1}^{r} \alpha_i v_i(s) \le -1, \quad s = \overline{1, N_2}.$$

Приклад застосування. Наведений алгоритм був протестований у задачі автоматичного розпізнавання чисельників у комп'ютерній телефонії. Чисельники як звукові сигнали були представлені у форматі WAV (uncompressed PCM). Як вектор ознак обрано спектр вихідного сигналу, отриманий за допомогою композиції перетворення Фур'є та вікон Хеммінга:

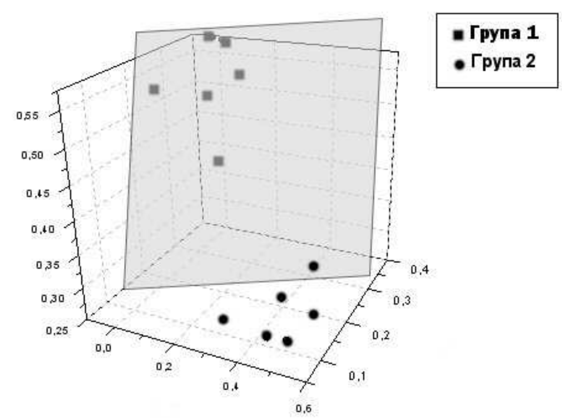
$$w(n) = 0.53836 - 0.46164 \cos\left(\frac{2\pi n}{N-1}\right).$$

Отримані множини векторів ознак навчальної вибірки, що складалася з чисельників «нуль» та «один», частково наведено в таблиці (для полегшення побудови графіка використовуються перші три координати вектора).

ТАБЛИЦЯ

«Нуль»	0,426707	0,115263	0,252032	0,725044
(Група 1)	0,478139	0,123449	0,250407	0,807893
	0,451432	0,138712	0,302439	0,723322
	…	…	…	…
«Один»	0,077387	0,31167	0,405589	–0,88242
(Група 2)	0,057001	0,35698	0,580578	–0,97261
	0,034388	0,34531	0,586168	–0,96633
	…	…	…	…

Для розв’язку задачі квадратичного програмування було розроблено алгоритм і програмну реалізацію на базі методів активного набору [6]. Після розв’язку задачі квадратичного програмування за формулами (6), (8) отримуємо коефіцієнти дискримінантної гіперплощини $y = a^T x$: $a_{opt} = (2{,}389 - 4{,}73 \;\; 0{,}98)$ (рисунок).



РИСУНОК

Заключення. Результат тестування показав ефективність даного методу обчислення оптимальної ширини роздільної смуги, а також зниження машинної ресурсоємності задачі за рахунок запропонованого перетворення. Надалі планується узагальнення алгоритму для каскадної дихотомної класифікації n множин точок в просторі векторів ознак, та подальше використання його в задачах розпізнавання мовних сигналів.

А.С. Гавриленко

АЛГОРИТМЫ СИНТЕЗА КЛАССИФИКАТОРОВ В ЗАДАЧАХ РАСПОЗНАВАНИЯ ОБЪЕКТОВ СРЕДСТВАМИ ЛИНЕЙНЫХ И НЕЛИНЕЙНЫХ ПРЕОБРАЗОВАНИЙ ПРОСТРАНСТВА ПРИЗНАКОВ

В работе представлен алгоритм синтеза систем классификации, разработанный на базе линейных и нелинейных оптимальных преобразований пространства признаков и реализованный с использованием метода активного набора для задачи квадратичного программирования и элементов метода построения оптимального нелинейного преобразователя как обобщенного полинома на заданных классах функций, а также показан результат работы алгоритма на примере решения задачи распознавания звуковых сигналов.

A.S. Gavrylenko

ALGORITHMS FOR CLASSIFICATION SYSTEMS SYNTHESIS IN OBJECT RECOGNITION PROBLEMS BASED ON LINEAR AND NON-LINEAR ATTRIBUTE SPACE TRANSFORMATIONS

An algorithm for signal classification systems synthesis, developed using principles of linear and non-linear transformations of attribute space and variation of active set method for quadratic programming, is introduced and described. An example of applying the algorithm to the problem of sound signal recognition is also presented.

1. *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. 50–58.
2. *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.
3. *Кириченко Н.Ф., Крак Ю.В.* Псевдообратные и проекционные матрицы в задачах синтеза функциональных преобразователей // Кибернетика и системный анализ. – 2004. – № 3. – С. 116–129.
4. *Гавриленко А.С.* Разработка алгоритма синтеза систем для решения задач классификации сигналов // Компьютерная математика. – 2010. – № 1. – С. 13–23.
5. *Кириченко Н.Ф., Куц Р., Лепеха Н.П.* Множества принадлежности в задачах классификации сигналов // Проблемы управления и информатики. – 2001. – № 5. – С. 71 –85.
6. *Гилл Ф., Мюррей У., Райт М.* Практическая оптимизация. Пер. с англ. – М.: Мир, 1985.– 509 с.

Отримано 15.11.2011

Про автора:

Гавриленко Анастасія Сергіївна,
молодший науковий співробітник Інституту кібернетики імені В.М. Глушкова НАН України.
e-mail: anastasiia.gavrylenko@gmail.com