Prace Naukowe Instytutu Cybernetyki Technicznej
Nr 9 Politechniki Wrocławskiej Nr 9

Studia i Materiały Nr 8 1973

Włodzimierz GREBLICKI

Rekurencyjny, asymptotycznie optymalny algorytm uczenia rozpoznawania

Podano asymptotycznie optymalny algorytm uczenia rozpoznawania, w którym nieznane gęstości prawdopodobieństwa, według których pojawiają się obiekty z poszczególnych klas, szacuje się przy zastosowaniu nieparametrycznych, rekursywnych metod estymacji. Prowadzi to do otrzymania rekurencyjnego algorytmu uczenia.

1. WSTĘP

W uczeniu rozpoznawania, na podstawie ciągu uczącego można aproksymować nieznane rozkłady prawdopodobieństwa w poszczególnych klasach i na tej podstawie tworzyć odpowiednie reguły decyzyjne [1,2,3]. Jeśli gęstości estymuje się w sposób nieparametryczny, to można otrzymać w ten sposób algorytmy zbieżne do optymalnego algorytmu Bayesa [1]. W poniższej pracy podano nowy sposób uczenia rozpoznawania, w którym wykorzystuje się tzw. rekursywne oszacowania gęstości prawdopodobieństwa [4] i wykazano jego asymptotycznie optymalne własności.

2. PRZEDSTAWIENIE PROBLEMU

Obiekt należący do jednej z M klas rozpoznaje się na podstawie p pomiarów, które tworzą wektor x. Prawdopodobieństwo pojawienia się obiektu z klasy i wynosi p_i, a gęstość prawdopodobieństwa w tej klasie $f_i(x)$ · $L(i,j)$ jest stratą poniesioną przez zaliczenie do klasy i obiektu należącego do klasy j. Ryzyko, dla reguły rozpoznawania $\psi(x)$ jest równe

$$R[\psi(x)] \overset{def}{=} \sum_{j=1}^{M} p_j \int L(\psi(x), j) f_j(x) dx. \tag{1}$$

Jak wiadomo, optymalna reguła Bayesa $\psi^*(x)$ zalicza obiekt o pomiarze x do klasy i, dla której ryzyko warunkowe

$$\sum_{j=1}^{M} L(i,j)\, p_j f_j(x) \tag{2}$$

jest najmniejsze.

Rozpoznawanie prowadzi się jednak w warunkach, gdy ani prawdopodobieństwa p_i ani rozkłady $f_i(x)$ nie są znane. Jedyną informacją jest ciąg uczący

$$(x_1, j_1), \ldots, (x_n, j_n),$$

gdzie j_k jest numerem klasy, do której należy obiekt o pomiarze x_k. Zarówno kolejne pomiary, jak i numery klas są stochastycznie niezależne. Na podstawie tego ciągu, nieznane prawdopodobieństwa p_i szacuje się ilorazami

$$p_{in} \overset{\text{def}}{=} \frac{n_i}{n}, \tag{3}$$

gdzie n_i jest liczbą obiektów z klasy i, natomiast nieznane gęstości - za pomocą tzw. oszacowań rekursywnych [4]. Jeśli przez $f_{in}(x)$ oznaczymy taki estymator dla klasy i, to reguła decyzyjna ustalona na podstawie ciągu uczącego zalicza obiekt x do klasy i, dla której wyrażenie

$$\sum_{j=1}^{M} L(i,j)\, p_{jn}\, f_{jn}(x) \tag{4}$$

osiąga minimum. Wykażemy, że otrzymany w ten sposób algorytm uczenia rozpoznawania jest asymptotycznie optymalny.

3. REKURSYWNA ESTYMACJA GĘSTOŚCI PRAWDOPODOBIEŃSTWA

W pracy [4] podano rekursywny sposób oszacowania nieznanej gęstości prawdopodobieństwa. Udowodnimy teraz, że estymator ten jest zgodny. Załóżmy w tym celu, że $Y, Y_1, Y_2, \ldots, Y_n$ jest ciągiem niezależnych p-wymiarowych zmiennych losowych o gęstości $f(y)$. Szacuje się ją wyrażeniem

$$f_n(y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^p}\, K\left(\frac{y - Y_i}{h_i}\right) \tag{5}$$

gdzie $K(y)$ jest odpowiednio wybraną funkcją, a h_i pewnym ciągiem liczbowym. Wykażemy teraz, że przy prawidłowym wyborze funkcji $K(y)$ i ciągu h_n powyższy estymator jest zgodny.

Twierdzenie

Jeśli

$$\int_y K(y)dy = 1, \int_y |K(y)|dy < \infty, \int_y K^2(y)dy < \infty, \quad \lim_{\|y\|\to\infty} \|y\|^p |K(y)| = 0 \tag{6}$$

oraz

$$h_n > 0, \lim_{n\to\infty} h_n = 0, \lim_{n\to\infty} nh_n^p = \infty, \tag{7}$$

to

$$f_n(y) \to f(y) \tag{8}$$

według prawdopodobieństwa, gdy $n\to\infty$, w punktach ciągłości funkcji $f(y)$.

D o w ó d

Załóżmy teraz, że pierwsze z założeń (6) nie jest spełnione lecz

$$\int_y K(y)\, dy < \infty .$$

Zauważmy, że

$$Ef_n(y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^p}\int_x K\left(\frac{y-x}{h_i}\right) f(x)dx = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^p}\int_x K\left(\frac{x}{h_i}\right) f(y-x)dx$$

i oznaczmy

$$g(y) \stackrel{def}{=} f(y)\int_x K(x)dx = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^p}\int_x K\left(\frac{x}{h_i}\right) f(y)dx.$$

Stąd

$$Ef_n(y) - g(y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^p}\int_x \left[f(y-x) - f(y)\right] K\left(\frac{x}{h_i}\right)dx.$$

Niech $\delta > 0$. Podzielmy teraz przestrzeń całkowania względem zmiennej x na dwa obszary, w których odpowiednio $\|x\| \leqslant \delta$ i $\|x\| > \delta$. Zatem

$$|Ef_n(y) - g(y)| \leqslant \frac{1}{n} \sum_{i=1}^{n} \Big[\max_{\|x\|\leqslant\delta} |f(y-x) - f(y)| \int_X |K(x)|\,dx +$$

$$+ \int_X \frac{|f(y-x)|}{\|x\|^p} \frac{\|x\|^p}{h_i^p} K\left(\frac{x}{h_i}\right)dx + |f(y)| \int_{\|x\|>\delta} \frac{1}{h_i^p} K\left(\frac{x}{h_i}\right)dx \Big] \leqslant$$

$$\leqslant \max_{\|x\|\leqslant\delta} |f(y-x) - f(y)| \int_X |K(x)|\,dx + \frac{1}{\delta^p} \frac{1}{n} \sum_{i=1}^{n} \sup_{\|x\| > \frac{\delta}{h_i}} \|x\|^p |K(x)| +$$

$$+ f(y) \frac{1}{n} \sum_{i=1}^{n} \int_{\|x\| \frac{\delta}{hi}} |K(x)|\,dx.$$

Zauważymy teraz, że dla dowolnego $\varepsilon > 0$ można wybrać $\delta > 0$ tak, aby spełniona została nierówność

$$\max_{\|x\|\leqslant\delta} |f(y-x) - f(y)| \int_X |K(x)|\,dx < \frac{\varepsilon}{3}.$$

Z założeń (6) (z wyjątkiem pierwszego z nich) wynika ponadto, że istnieje H_1 takie, że

$$\frac{1}{\delta^p} \sup_{\|x\|>\frac{\delta}{h_i}} \|x\|^p |K(x)| < \frac{\varepsilon}{3}$$

dla $h_i < H_1$ oraz H_2 takie, że

$$f(y) \int_{\|x\|>\frac{\delta}{h_i}} |K(x)|\,dx < \frac{\varepsilon}{3}$$

dla $h_i < H_2$. Stąd i z założeń (7) wynika, że istnieje N takie, że $h_n < \min(H_1, H_2)$ dla $n > N$ oraz, że

$$\left| E\, f_n(y) - g(y) \right| < \frac{1}{n} \sum_{i=1}^{N} \frac{\varepsilon}{3} + \bar{K} + \frac{\varepsilon}{3} + \frac{1}{n} \sum_{i=N+1}^{n} \left[\frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \right] \leqslant$$

$$\leqslant \frac{1}{n} \sum_{i=1}^{N} \left(\frac{2}{3} \varepsilon + \bar{K} + \varepsilon \right)$$

dla $n > N$, przy czym $\bar{K} = \sup \|x\|^{p} |K(x)|$. Weźmy teraz

$$n > \max \left[N, \frac{2}{3} N + \frac{1}{\varepsilon} N \bar{K} \right].$$

Zatem

$$\left| Ef_n(y) - g(y) \right| < 2\varepsilon$$

Wynika stąd, że

$$\lim_{n \to \infty} \left| Ef_n(y) - f(y) \int_X K(x)dx \right| = 0. \qquad (9)$$

W szczególności, jeśli przyjąć $K(y)$ w ten sposób, aby

$$\int_Y K(y)dy = 1,$$

to

$$\lim_{n \to \infty} \left| Ef_n(y) - f(y) \right| = 0. \qquad (10)$$

Dla wykazania zgodności estymatora wystarczy jeszcze udowodnić, że

$$\lim_{n \to \infty} \operatorname{var} \left[f_n(y) \right] = 0.$$

Zauważmy w tym celu, że

$$\operatorname{var}\left[f_n(y)\right] = \frac{1}{n}\sum_{i=1}^{n}\operatorname{var}\left[\frac{1}{h_i^P}K\left(\frac{y-Y_i}{h_i}\right)\right] \leqslant \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^{2P}}EK^2\left(\frac{y-Y_i}{h_i}\right) \leqslant$$

$$\leqslant \frac{1}{n\,\min^P(h_1,\ldots,h_n)}\sum_{i=1}^{n}\frac{1}{h_i^P}EK^2\left(\frac{y-Y_i}{h_i}\right).$$

Ze zbieżności (9) oraz założeń (6) i (7) wynika, że

$$\lim_{n\to\infty} n\,\min^P(h_1,\ldots,h_n)\operatorname{var}\left[f_n(y)\right] = f(y)\int_y K^2(y)\,dy,$$

co kończy dowód.

4. ALGORYTM UCZENIA

Stosowanie powyższego, zgodnego estymatora gęstości prawdopodobieństwa zapewnia, że reguła decyzyjna dąży, przy wzroście długości ciągu uczącego, do reguły Bayesa, a w przypadkach, gdy gęstości $f_i(x)$ są prawie wszędzie ciągłe, ryzyko - do ryzyka Bayesa [1]. Obiekt zalicza się do klasy i, dla której wyrażenie

$$\Phi_n(x,i) \overset{\mathrm{def}}{=} \sum_{j=1}^{n} L(i,j)\sum_{k=1}^{n_j}\frac{1}{h_K^P}K\left(\frac{x-x_{jk}}{h_K}\right)$$

jest najmniejsze, gdzie x_{jk} jest k-ym kolejnym obiektem z klasy j w ciągu uczącym. Powyższy algorytm uczenia można zapisać w postaci wzoru rekurencyjnego

$$\Phi_1(x,i) = L\,(i,j=j_1)\,\frac{1}{h_1^P}K\left(\frac{x-x_1}{h_1}\right),$$

$$\Phi_{n+1}(x,i) = \Phi_n(x,i) + L(i,j=j_{n+1})\,\frac{1}{h^P_{(n+1)j}}K\left(\frac{x-x_{n+1}}{h_{(n+1)j}}\right), n = 1, 2, \ldots .$$

Dla różnych funkcji $K(y)$ otrzymuje się różne wersje algorytmu. Jako przykłady można podać:

$$K(y) = \prod_{i=1}^{P}\left(\frac{1}{\sqrt{2\pi}}\right)^{P} e^{-\frac{|y^{(i)}|^2}{2}}, \quad K(y) = \prod_{i=1}^{P}\left(\frac{1}{2}\right)^{P} e^{-|y^{(i)}|}$$

Podany w pracy algorytm uczenia rozpoznawania wykorzystuje nieparametryczne, zgodne sposoby szacowania nieznanych gęstości. Dzięki rekurencyjnemu ich charakterowi procesy obliczeniowe dają przedstawić się w postaci prostego wzoru rekurencyjnego.

LITERATURA

[1] G r e b l i c k i W., Nieparametryczna estymacja w uczeniu rozpoznawania, Podstawy sterowania, nr 3, 1972.

[2] G r e b l i c k i W., O pewnej metodzie uczenia w nieparametrycznym zadaniu rozpoznawania obrazów, Prace Naukowe Instytutu Cybernetyki Technicznej Politechniki Wrocławskiej nr 1, Studia i Materiały, nr1, 1972.

[3] B u b n i c k i Z., Least interval pattern recognition and its application in control systems, Materiały IV Kongresu IFAC, Warszawa 1969.

[4] W o l v e r t o n C. T., W a g n e r T. J., Recursive estimates of probability densities, IEEE Trans. SSC, nr 3, 1969.

Praca wpłynęła do Redakcji 27 IX 1972.
Po poprawkach 16 VI 1973.

RECURSIVE ASYMPTOTICALLY OPTIMAL LEARNING ALGORITHM TO RECOGNIZE PATTERNS

In the paper a recurrent and asymptotically optimal learning algorithm is given. Probability densities of patterns to be classified are estimated by recursive methods. This approach yields a recurrent learning algorithm.

РЕКУРРЕНТНЫЙ АСИМПТОТИЧЕСКИ ОПТИМАЛЬНЫЙ АЛГОРИТМ ОБУЧЕНИЯ РАСПОЗНАВАНИЮ

В работе представлен асимптотически оптимальный алгоритм обучения распознаванию. Неизвестные функции распределения оцениваются при помощи непараметрических рекуррентных оценок. Этим методом получен рекуррентный алгоритм распознавания.