АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
АВТОМЕТРИЯ
№ 5 1971

УДК 621.391.193

А. Г. СЕНИН
(Новосибирск)

К АНАЛИЗУ КРИТЕРИЕВ ОПТИМАЛЬНОСТИ, ИСПОЛЬЗУЕМЫХ ПРИ ОБУЧЕНИИ РАСПОЗНАВАНИЮ ОБРАЗОВ

Рекуррентные алгоритмы обучения находят широкое применение при решении различных задач оптимизации автоматических устройств, в частности распознавания образов.

Если разделяющая поверхность $\hat{f}(\bar{x})$ аппроксимируется в виде суммы

$$\hat{f}(\bar{x}) \sum_{j=1}^{N} c_j \varphi_j(\bar{x}), \tag{1}$$

то коэффициенты c_j, формируемые регулярным алгоритмом стохастической аппроксимации, на каждом шаге обучения определяются из условия [1]

$$c_j[n] = c_j[n-1] + \gamma[n] F'(y[n] - \sum_{i=1}^{N} c_i[n-1]\varphi_i(\bar{x}[n]))\varphi_j(\bar{x}[n]). \tag{2}$$

При соответствующем выборе шага $\gamma[n]$ такие алгоритмы обеспечивают сходимость искомых коэффициентов к оптимальным, минимизирующим функционал ошибки

$$I = \overline{F\Big(y - \sum_{j=1}^{N} c_j \varphi_j(\bar{x})\Big)}, \tag{3}$$

где y — реакция оператора, равная, например, величине

$$y = \begin{cases} 1; & \bar{x} \in A; \\ -1; & \bar{x} \in B. \end{cases} \tag{4}$$

Очевидно, если нам удалось отыскать такую поверхность, то использование решающего правила

$$\hat{f}(\bar{x}) \begin{cases} \geqslant 0; & \bar{x} \in A, \\ < 0; & \bar{x} \in B \end{cases} \tag{5}$$

позволяет классифицировать реализации образа $\bar{x}$, принадлежность которых неизвестна.

В зависимости от вида выпуклой функции $F(\cdot)$ оптимальные коэффициенты в (3) могут быть различными, поэтому несомненный интерес представляют вопросы анализа выбранного критерия, выявления характера разделяющей поверхности, сформированной на основе минимизации последнего, и сравнения решающего правила (5) с байесовым. Ограничимся рассмотрением лишь наиболее простых критериев, а именно:

$$I = \overline{\left(y - \sum_{j=1}^{N} c_j \varphi_j(\bar{x})\right)^2}, \tag{6}$$

а также

$$I = \overline{\left| y - \sum_{j=1}^{N} c_j \varphi_j(\bar{x}) \right|}. \tag{7}$$

Соответствующие алгоритмы обучения (2) в этом случае оказываются элементарными и могут быть реализованы в сравнительно простых аналоговых устройствах. Далее введем критерий, непосредственно минимизирующий вероятность ошибки распознавания.

Будем минимизировать функционал

$$I = \overline{(y - f(\bar{x}))^2}, \tag{8}$$

где $f(\bar{x})$ определена в произвольном классе функций. Тогда искомая поверхность $f(\bar{x})$ оказывается равной [2]

$$f(\bar{x}) = \frac{(2\nu - 1)\, P_B P(\bar{x}/B) - (2\mu - 1)\, P_A P(\bar{x}/A)}{P_B P(\bar{x}/B) + P_A P(\bar{x}/A)}, \tag{9}$$

где μ, ν — вероятности правильной классификации оператором обучающей последовательности из классов A и B; P_A, P_B — априорные вероятности; $P(\bar{x}/A)$, $P(\bar{x}/B)$ — многомерные условные распределения признаков соответствующих классов.

Если оператор ошибается в равной степени в оценке ситуаций, т. е. $\mu=\nu$, то решающее правило (5) и разделяющая поверхность

$$f(\bar{x}) = 0 \tag{10}$$

обеспечивают минимум вероятности ошибок в распознавании.

Практически используют разделяющую поверхность (1) в виде суммы некоторых детерминированных функций с коэффициентами, настраиваемыми при обучении. Можно показать, что минимум функционала (6) достигается в этом случае на поверхности $\hat{f}(\bar{x})$, аппроксимирующей оптимальную поверхность по критерию среднего квадрата ошибки, т. е. $\hat{f}(\bar{x})$ минимизирует также функционал

$$\overline{\varepsilon^2} = \int_{-\infty}^{\infty} (f(\bar{x}) - \hat{f}(\bar{x}))^2 P(\bar{x})\, d\bar{x}, \tag{11}$$

где $P(\bar{x})$ — безусловное распределение признаков.

Рассмотрим теперь критерий вида

$$\overline{|y - f(\bar{x})|} = \int\int_{-\infty}^{\infty} |y - f(\bar{x})|\, P(\bar{x}, y)\, d\bar{x}\, dy =$$

$$= \int_{-\infty}^{\infty} [(P_A P(\bar{x}/A) \mu + P_B P(\bar{x}/B) (1-\nu)) | 1 - f(\bar{x}) | +$$

$$+ (P_A P(\bar{x}/A) (1-\mu) + P_B P(\bar{x}/B) \nu) | 1 + f(\bar{x}) |] \, dx. \qquad (12)$$

Разобьем пространство интегрирования на две области G_1, для которой

$$P_A P(\bar{x}/A) \mu + P_B P(\bar{x}/B) (1-\nu) > P_A P(\bar{x}/A) (1-\mu) + P_B P(\bar{x}/B) \nu, \qquad (13)$$

и G_2, для которой

$$P_A P(\bar{x}/A) \mu + P_B P(\bar{x}/B) (1-\nu) < P_A P(\bar{x}/A) (1-\mu) + P_B P(\bar{x}/B) \nu. \qquad (14)$$

Очевидно, минимум функционала (12) имеет место, если в области G_1 $\bar{f}(x) = -1$, в области G_2 $\bar{f}(x) = 1$. Таким условиям удовлетворяет функция

$$f(\bar{x}) = \operatorname{sign} k \, [(2\nu - 1) P_B P(\bar{x}/B) - (2\mu - 1) P_A P(\bar{x}/A)], \qquad (15)$$

где k — произвольный положительный множитель.

Если оператор при обучении ошибается в равной степени, т. е. $\mu = \nu$, тогда

$$f(\bar{x}) = \operatorname{sign} k \, [P_B P(\bar{x}/B) - P_A P(\bar{x}/A)] \qquad (16)$$

обеспечивает минимум ошибок распознавания.

Таким образом, возвращаясь к формуле (16), можно сделать заключение о том, что по критерию модуля ошибки в классе произвольных функций, так же как и при квадратичном критерии, формируется оптимальная разделяющая поверхность, минимизирующая вероятность ошибки распознавания.

Следует заметить, что если оптимальная разделяющая поверхность

$$f(\bar{x}) = P_B P(\bar{x}/B) - \bar{P}_A P(\bar{x}/A) = 0 \qquad (17)$$

и представима рядом (1), то при квадратичном критерии и в этом случае не гарантируется формирование такой поверхности. Для обеспечения оптимального распознавания необходимо удовлетворить более жесткому требованию представимости в виде ряда (1) функции (9). В этом смысле обучение распознаванию по критерию модуля ошибки более предпочтительно, поскольку для формирования оптимальной поверхности при этом достаточно выполнения первого условия.

Однако в общем случае эти критерии, а также и другие известные не обеспечивают после обучения минимума ошибок распознавания. Так, например, для двух равномерных распределений (рис. 1)

$$P_1(x) = \begin{cases} 2; & -0{,}25 < x < 0{,}25; \\ 0; & -0{,}25 > x; \; x > 0{,}25; \end{cases} \qquad (18)$$

$$P_2(x) = \begin{cases} 1; & -1{,}25 < x < -0{,}25; \\ 0; & -1{,}25 > x; \; x > -0{,}25 \end{cases} \qquad (19)$$

и линейной разделяющей функции

$$\hat{f}(x) = x - c = 0. \qquad (20)$$

оптимальное по вероятности ошибок значение «порога» $c^* = -0{,}25$, при квадратичном критерии

$$c_{\text{кв}} = \frac{m_1 + m_2}{2} = -0{,}38 \tag{21}$$

и модульном

$$-0{,}25 > c_{\text{м}} > -0{,}75. \tag{22}$$

Известные алгоритмы обучения, минимизирующие математическое ожидание несовпадения знаков, приводят к усложнению вычислительных операций, учитывающих ограничения типа неравенств. Но и в этом случае не гарантируется формирование оптимальных коэффициентов, минимизирующих вероятность неправильных решений.

Действительно, минимум функционала

$$I = \overline{F\Big(y \sum_{j=1}^{N} c_j \varphi_j(\bar{x}) - \alpha\Big)} \tag{23}$$

при ограничении

$$\alpha \geqslant 0 \tag{24}$$

имеет место, если при совпадении знаков принять

$$\alpha = \Big| \sum_{j=1}^{N} c_j \varphi_j(\bar{x}) \Big|; \tag{25}$$

в противном случае

$$\alpha = 0. \tag{26}$$

В силу того, что $F(\cdot)$ является выпуклой функцией, функционал (23) определяется не только вероятностью неправильных ответов, но также и весом $\Big| \sum_{j=1}^{N} c_j \varphi_j(\bar{x}) \Big|$, с которым каждая ошибка участвует в осреднении.

Минимум ошибок будет обеспечен, если при обучении минимизируется такой критерий, для которого цены правильных ответов приняты, например, равными нулю, как в (23), цены неправильных ответов — некоторой другой положительной величине, не зависящей от абсолютного значения рассогласования между сигналами y и $\hat{f}(\bar{x})$. Очевидно, таким условиям удовлетворяет критерий вида

$$I = \overline{F\Big(y - \operatorname{sign} \sum_{j=1}^{N} c_j \varphi_j(\bar{x})\Big)}, \tag{27}$$

экстремум которого определяется лишь качеством распознавания.

Нетрудно убедиться, что для произвольной функции $f(\bar{x})$ функционал

$$I = \overline{F(y - \operatorname{sign} f(\bar{x}))} \tag{28}$$

принимает экстремальное значение, если

$$f(\bar{x}) = k\,(P_B P(\bar{x}/B) - P_A P(\bar{x}/A)), \tag{29}$$

что соответствует байесовому правилу.

К сожалению, минимизировать в процессе обучения критерий (27) невозможно, так как производная его по искомым параметрам везде равна нулю, за исключением тех точек, для которых

$$\sum_{j=1}^{N} c_j \varphi_j(\bar{x}) = 0. \tag{30}$$

Однако возможно сколь угодно точно в процессе обучения приблизиться к оптимальной поверхности, если воспользоваться критерием вида

$$I = \Big(y - \alpha \operatorname{sign} \sum_{j=1}^{N} c_j \varphi_j(\bar{x}) - \sum_{j=1}^{N} c_j \varphi_j(\bar{x})\Big);\ 0 \leqslant \alpha < 1. \tag{31}$$

На рис. 2 представлена зависимость величины функционала

$$F = \Big(y - \alpha \operatorname{sign} \sum_{j=1}^{N} c_j \varphi_j(\bar{x}) - \sum_{j=1}^{N} c_j \varphi_j(\bar{x})\Big)^2 \tag{32}$$

от значения

$$z = y \sum_{j=1}^{N} c_j \varphi_j(\bar{x}) \tag{33}$$

при разных коэффициентах α. Очевидно, при значениях α, близких к единице, коэффициенты разделяющей поверхности практически будут совпадать с оптимальными по критерию минимума ошибок распозна-

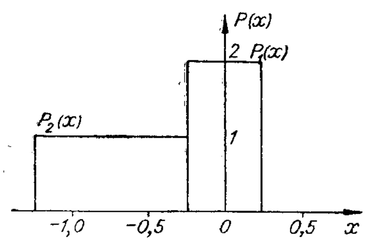

Рис. 1.

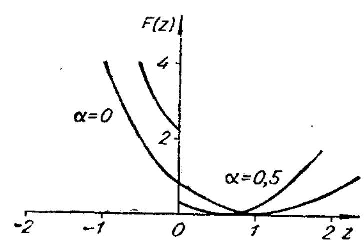

Рис. 2.

вания, причем вид выпуклой функции будет влиять лишь на скорость сходимости алгоритма, не меняя, по существу, результат обучения. Если $\alpha \geqslant 1$, то коэффициенты c_j, сохраняя оптимальное положение разделяющей поверхности в пространстве признаков, стремятся к нулю, что приводит к дополнительным погрешностям в подстройке коэффициентов, обусловленным дискретным характером вычислительных операций или ограниченными потенциальными возможностями реальной аппаратуры. Поэтому достаточным может оказаться значение $\alpha = 0{,}8 \div 0{,}9$.

Искомые коэффициенты в этом случае определяются рекуррентным выражением вида

$$c_j[n] = c_j[n-1] + \gamma[n] F'\Big(y[n] - \alpha \operatorname{sign} \sum_{i=1}^{N} c_i[n-1] \times$$

$$\times \varphi_i(\bar{x}[n]) - \sum_{i=1}^{N} c_i[n-1] \varphi_i(\bar{x}[n])\Big) \varphi_j(\bar{x}[n]). \tag{34}$$

На рис. 3 представлены значения оптимальных коэффициентов для ранее рассмотренного примера с той лишь разницей, что в аргумент функционала введена пороговая функция. При этом искомые величины определяются решением соответствующих уравнений:

$$c_{\text{кв}} + \alpha\left(\int\limits_{-\infty}^{c_{\text{кв}}} [P_1(x) + P_2(x)]\,dx + P_1(c_{\text{кв}}) - P_2(c_{\text{кв}}) - 1\right) = \frac{m_1 + m_2}{2}; \tag{35}$$

$$\alpha\left(P_1(c_{\text{м}}) - P_2(c_{\text{м}})\right) + \int\limits_{-\infty}^{c_{\text{м}}+1-\alpha} P_1(x)\,dx - \int\limits_{c_{\text{м}}-1+\alpha}^{\infty} P_2(x)\,dx = 0. \tag{36}$$

Введение пороговой функции в критерий оптимальности (31) позволяет при обучении существенно упростить вычислительные операции по сравнению с известными алгоритмами совпадения знаков и вместе с тем сформировать разделяющую поверхность, минимизирующую ошибки распознавания.

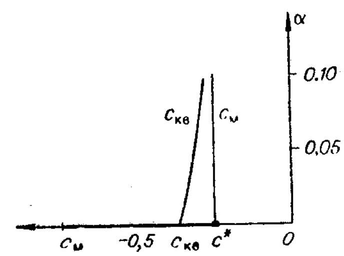

Рис. 3.

ЛИТЕРАТУРА

1. Я. З. Цыпкин. Адаптация и обучение в автоматических системах. М.. «Наука», 1968.
2. Б. Д. Борисов, М. И. Могильницкий, З. И. Нестерова, А. Г. Сенин. О принципах построения аппаратуры для распознавания шумовых сигналов.— Тезисы докладов IV Всесоюзного совещания по автоматическому управлению (технической кибернетике). Тбилиси, 1968.

Поступила в редакцию
3 ноября 1970 г.