

ЛИНЕЙНЫЕ РЕШАЮЩИЕ ФУНКЦИИ, БЛИЗКИЕ К ОПТИМАЛЬНЫМ

Н.Г. Загоруйко

Известно, что правила принятия решений и аппаратура, их реализующая, основанные на использовании линейных решающих функций (гиперплоскостей) просты. Однако, можно показать, что эти правила можно подвергнуть дальнейшему упрощению, либо совсем не увеличивая риска ошибок, либо увеличивая его на незначительную контролируемую величину.

Будем рассматривать пример распознавания K образов при следующих предположениях. Априори известно, что распределения реализаций образов в многомерном выборочном пространстве подчиняются нормальному закону с математическим ожиданием μ_i , $1 \leq i \leq K$. Для всех образов матрицы ковариаций единичны, так что поверхности равной плотности вероятности представляют собой сферы одинакового радиуса. Априорные вероятности q_i появления всех образов одинаковы: $q_i = \frac{1}{K}$, $1 \leq i \leq K$.

Выборочное пространство с помощью гиперплоскостей разбивается на K областей, в каждой из которых вероятность появления одного образа больше, чем вероятность появления любого другого. Такое разбиение пространства оптимально в том смысле, что оно обеспечивает минимум ожидаемого риска ошибок. Под риском ошибок будем понимать следующее. Вероятность ошибочного отнесения реализации z -го образа к i -тому образу есть

$$R(j/i) = \int_{y_j} P_i(x) dx,$$

где $P_i(x)$ - закон распределения i -го образа в n -мерном выборочном пространстве $x = (x_1, x_2, \dots, x_n)$; а y_j - область j -го образа в этом пространстве, ограниченная с помощью решающих гиперплоскостей. При априорных вероятностях появления 1-го и 2-го образов $q_1 = q_2 = \frac{1}{2}$ средние потери для двух образов выразятся так:

$$R = \frac{1}{2} [R(2/1) + R(1/2)].$$

Если число образов $k > 2$, то общая вероятность ошибки распознавания равна

$$R_{\text{оп.}} = \frac{1}{k} \sum_{j=1}^k R(j/i) = \frac{1}{k} \sum_{i=1}^k R_i,$$

где R_i - общая вероятность ошибки распознавания i -го образа.

Отделять все образы друг от друга попарно обычно нет необходимости, так что количество оптимальных гиперплоскостей θ , как правило, меньше C_k^2 . [1,2]. На рис.1 θ гиперплоскостей показаны жирными линиями. Пунктиром показаны $C_k^2 - \theta$ гиперплоскостей, которые были бы необходимы для разделения всех образов попарно.

Чем меньшим количеством гиперплоскостей выборочное пространство делится на k областей, тем более экономичной может быть выбрана схема распознающего автомата. Однако уменьшение количества гиперплоскостей до числа, меньшего θ , приведет к увеличению ожидаемого риска потерь.

Отсюда вытекает постановка задачи, рассматриваемой в данной работе: найти способ уменьшения количества гиперплоскостей до числа N , меньшего θ , который бы при любом N обеспечивал минимум ожидаемого риска ошибок. В отличие от θ оптимальных решающих функций, N выбранных таким способом гиперплоскостей образуют "линейные решающие функции, близкие к оптимальным."

Возможности уменьшения $R_{\text{оп}}$ после сокращения числа гиперплоскостей до $N < \theta$ заложены в следующем: гиперплоскость (u_j), на которую дополнительно возлагается функция одной из исключенных гиперплоскостей (u_{ei}), может быть повернута и (или) смещена в такое положение, при котором она выполняла бы свою роль двух гиперплоскостей оптимальным, в смысле общих потерь, образом. В отличие от исходных "оптимальных" гиперплоскостей будем называть эти новые гиперплоскости "близкими к оп-

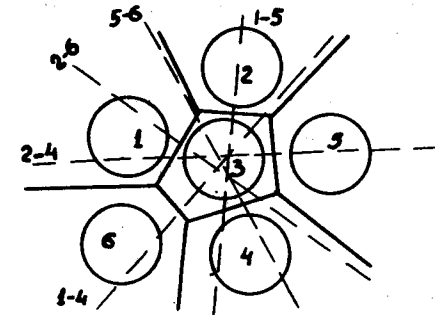


Рис. 1. Оптимальные линейные решающие функции.

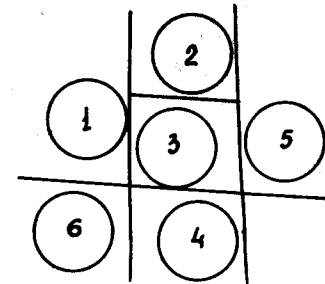


Рис. 2. Линейные решающие функции, близкие к оптимальным.

тимальным", а правила принятия решения, использующие эти гиперплоскости, - "решающими функциями, близкими к оптимальным". Задача синтеза последних гораздо более сложная, чем задача простого сокращения числа гиперплоскостей без изменения их положения. Но так как она решается лишь при разработке распознающего устройства, то идти на большие, но разовые затраты машинного времени в большинстве случаев целесообразно, имея в виду, что ее решение может существенно упростить схему устройства. Так на рис. 2 показан тот же случай, что и на рис. 1, но разделенный гиперплоскостями, близкими к оптимальным. Оказалось, что без значительного увеличения потерь, число гиперплоскостей можно свести к 4.

Трудность задачи состоит в том, что ее решение, если не использовать какого-нибудь аналитического или эвристического метода, достигается полным перебором всех возможных замен θ исходных гиперплоскостей близкими к оптимальным, причем положение этих гиперплоскостей также определяется перебором всех ее возможных положений в n -мерном пространстве. Ясно, что таким путем задачу решить невозможно.

Задача замены оптимальных плоскостей U_{ij} и $U_{\ell z}$ одной плоскостью $U_{i\ell-jz}$ сводится к задаче поиска линейной решающей функции, наилучшим образом отделяющей образы i и ℓ от образов j и z . Ясно, что при этом в общем случае не может идти речь о нормальном распределении суммы реализаций i и ℓ (так же, как и j и z) образов. Линейное разделение двух множеств точек, произвольно распределенных в пространстве, может быть осуществлено двумя путями.

Первый состоит в том, что, следуя Себастьяну [3], ищется полиномиальное приближение ($F^r(v)$) оптимальной (нелинейной) решающей функции ($U(v)$). Если при этом ограничиться полиномом первой степени, то решение будет представлять собой наилучшую линейную аппроксимацию $U(v)$.

Второй путь состоит в использовании метода Андерсона-Бахадера [4], развитого в работе [2].

Если координаты математического ожидания суммы реализаций i и ℓ образов заданы вектором μ_1 , а суммы реализаций j и z образов - μ_2 , то уравнение гиперплоскости, отделяющей две эти суммы друг от друга, можно найти в виде:

$$b^T x + \rho = 0, \quad (1)$$

где ρ - постоянный член, x - n -мерный вектор, а b^T - транс-

понированный вектор b

$$b = [z\sigma_1 + (1-z)\sigma_2]^{-1} \cdot (\mu_2 - \mu_1). \quad (2)$$

Здесь σ_1 и σ_2 - матрицы ковариаций суммарного распределения i и ℓ и j и z образов, соответственно, а z может быть найдено из решения уравнения (3):

$$b^T [z^2\sigma_1 - (1-z)^2\sigma_2] b = 0. \quad (3)$$

Значение ρ определяется формулой (4):

$$\rho = - \frac{(b^T\sigma_2 b)^{1/2} b^T \mu_1 + (b^T\sigma_1 b)^{1/2} b^T \mu_2}{(b^T\sigma_2 b)^{1/2} + (b^T\sigma_1 b)^{1/2}}. \quad (4)$$

Если одно из отделяемых множеств составлено из S образов, появляющихся с вероятностью q_i ($i=1, \dots, S$), то, зная математические ожидания μ_i и матрицы ковариаций σ_i этих образов, можно найти μ_Σ и σ_Σ суммарного множества.

Действительно, легко показать, что

$$\mu_\Sigma = \sum_{i=1}^S q_i \cdot \mu_i,$$

$$D_\Sigma = \sum_{i=1}^S q_i (D_i + h_i^2).$$

Здесь D_i и D_Σ - дисперсии i -го образа и суммарного множества образов, соответственно, а $h_i = \mu_\Sigma - \mu_i$.

Суммарная ковариация множества по двум переменным x и y $K_{xy\Sigma}$ через ковариации образов K_{xyi} , входящих во множество, и через расстояния h_{xi} и h_{yi} между математическим ожиданием множества и отдельных образов выражается так:

$$K_{xy\Sigma} = \sum_{i=1}^S q_i (K_{xyi} + h_{xi} \cdot h_{yi}).$$

В рассматриваемом нами простом случае $q_i = q_j = \frac{1}{S}$; $K_{xyi} = 0$; $D_i = 1$, так что

$$\mu_\Sigma = \frac{\sum_{i=1}^S \mu_i}{S};$$

$$D_\Sigma = \frac{1 + \sum_{i=1}^S h_i^2}{S};$$

$$K_{xy\Sigma} = \frac{\sum_{i=1}^S h_{xi} \cdot h_{yi}}{S};$$

Итак, определение "решающими функциями, близкими к опти-

малыми" можно осуществить с помощью сравнительно простых вычислений. Однако при большом числе образов K число возможных сочетаний пар оптимальных плоскостей $C_{(C_k^2)}^{(2)}$ может оказаться настолько большим, что перебор и сравнение эффективности всех плоскостей, близких к оптимальным, будет занимать неприемлемо большое машинное время.

Упростить задачу можно следующим способом. Условимся считать ошибки δ распознавания любого образа пренебрежимо малыми. Отсюда следует, что плоскость, проходящая на расстоянии ρ_δ от математического ожидания i -го образа, будет достаточно надежно отделять область i -го образа от всех образов, удаленных от данной плоскости с противоположной стороны на расстоянии $\rho > \rho_\delta$, если ρ_δ удовлетворяет условию

$$(1-\delta) > \frac{2}{\sqrt{\pi}} \int_0^{\rho_\delta} e^{-t^2} dt.$$

Составим таблицу оптимальных плоскостей, отделяющих каждый образ от каждого (таких плоскостей C_k^2). В таблице укажем евклидовы расстояния (E_{ij}) между математическими ожиданиями образов, оптимально разделяемых этими плоскостями. Отметим те плоскости U_{ij} , для которых $E_{ij} < 2\rho_\delta$. Эти плоскости следует использовать без изменения их положения, так как ошибки даже при оптимальном их размещении уже превышают δ . Затем следует найти расстояния от отмеченных плоскостей до математических ожиданий всех остальных образов. Если окажется, что плоскость U_{ij} проходит между ℓ -тым и z -тым образом, причем $\rho_{ij-\ell} > \rho_\delta$ и $\rho_{ij-z} > \rho_\delta$, то это будет означать, что функции плоскости $U_{z\ell}$ можно переложить на U_{ij} . Отметим все сочетания пар образов, разделенных "попутно" плоскостями U_{ij} и эти сочетания исключим из дальнейшего рассмотрения. Затем берем плоскость U_{nm} , которая делит пару образов n и m , причем E_{nm} минимально из всех оставшихся неотмеченными пар образов. Ищем пары образов из числа неотмеченных, между которыми проходит U_{nm} . Если U_{nm} проходит только между образами, расстояние от которых до $U_{nm} > \rho_\delta$, то такие пары образов отмечаются, а плоскость U_{nm} остается в своем исходном положении.

Если U_{nm} проходит между парой образов p и q , причем $|\rho_p| + |\rho_q| > 2\rho_\delta$, но $\rho_p < \rho_\delta$ или $\rho_q < \rho_\delta$, то можно попытаться найти плоскость U_1 , близкую к оптимальной, заменяющую U_{nm} и U_{pq} . Для этого объединяются в два множества образы n и p и образы m и q , определяются математичес-

кие ожидания и матрицы ковариации этих множеств и ищется близкая к оптимальной гиперплоскость U_1 , их разделяющая. В случае успеха (т.е. при $\rho_p > \rho_\delta$ и $\rho_q > \rho_\delta$) образы p и q отмечаются и ищется следующая пара образов, между которыми проходит плоскость U_1 . Если такая пара (например, v и w) имеется, то создаются множества образов n, p, v и m, q, w и делается попытка найти плоскость U_2 , удаленную от каждого образа этих множеств на расстояние $\rho > \rho_\delta$.

После того, как все такие возможности исчерпаны, аналогичный процесс начинается с рассмотрения очередной пары образов, не отмеченной на предыдущих шагах и минимально удаленных друг от друга.

Эта процедура продолжается до тех пор, пока не останутся неотмеченных (т.е. не разделенных удовлетворительно) пар образов. В итоге получаем минимальный набор решающих плоскостей, разделяющих выборочное пространство на K зон с заданной достоверностью.

Обсудим следующие вопросы: при каких условиях целесообразен поиск решающих функций, близких к оптимальным? Нельзя ли найти более простое решение?

Известно, что вместо оптимальных гиперплоскостей можно использовать сравнение евклидовых расстояний от точки - реализации - до математических ожиданий образов. При этом, если не вводить пороговых ограничений, результаты распознавания совершенно одинаковы [3]. В то же время сложность алгоритма и аппаратуры вычисления и сравнения K евклидовых расстояний не на много сложнее алгоритма и аппаратуры, использующей K гиперплоскостей [5]. Значит, использовать гиперплоскости целесообразно лишь в том случае, когда их число не больше числа распознаваемых образов.

Как часто можно перейти к $N < K$ плоскостям, близким к оптимальным?

Интуитивно ясно, что при случайном взаимном расположении образов в выборочном пространстве высокая надежность распознавания с помощью оптимальных гиперплоскостей достигается тогда, когда образы удалены друг от друга на достаточно большие расстояния. Большая плотность упаковки пространства образами ассоциируется с меньшей надежностью распознавания. Следовательно, если надежность распознавания (P_0) при использовании оптимальных решающих функций высока, то велики возможности для смещения и поворота гиперплоскостей. В этом случае вероятность

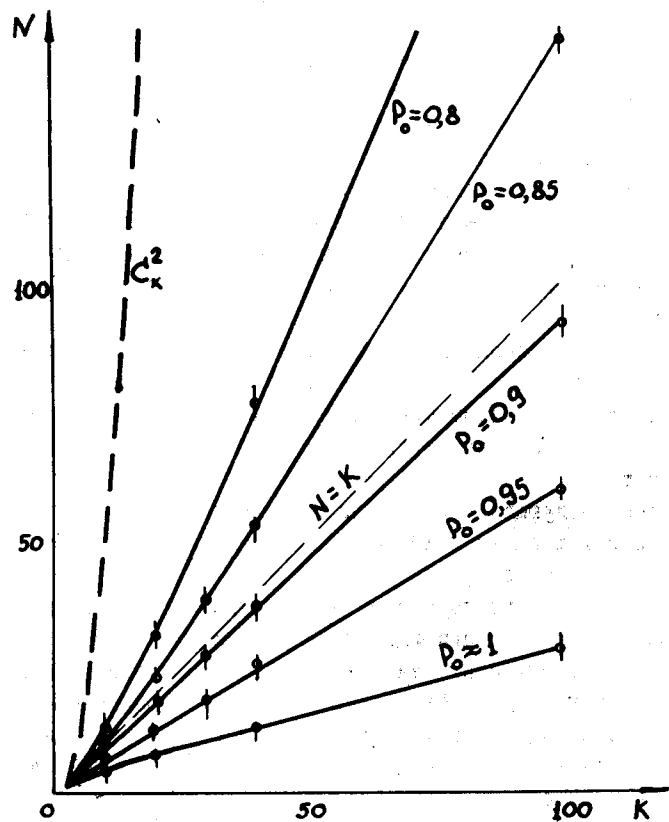


Рис. 3. Зависимость числа гиперплоскостей (N), близких к оптимальным, от числа образцов (K).

существенного выигрыша в числе гиперплоскостей при переходе от оптимальных к близким к оптимальным также возрастает.

Необходимо определить величину исходной надежности распознавания P_0 , при которой можно ожидать, что число N гиперплоскостей не будет превышать числа образцов.

Определение P_0 было проделано путем моделирования.

В выборочном пространстве n переменных с помощью датчика случайных чисел задавались координаты математических ожиданий K образцов. Дисперсии образцов так согласовались с размером пространства, что потери при оптимальном разделении образцов не превышали $\alpha = 1 - \rho_0$. Затем описанным здесь методом оптимальные плоскости заменялись близкими к оптимальным. Отношение (R) числа этих плоскостей (N) к числу образцов (K) фиксировалось. Эксперимент проводился при $n = 2$ (образы на плоскости), $K = 4, 6, 8, 10, 15, 20, 30, 40, 100$, для $P_0 = 0,8; 0,85; 0,9; 0,95; 0,999$. Для каждого такого варианта сочетания K и P_0 было проведено несколько (до 10 опытов), отличающихся друг от друга взаимным (случайным) расположением образцов друг относительно друга. Всего было промоделировано около 300 вариантов. Результаты эксперимента приведены на рис. 3. $R < 1$ имеет место при значениях P_0 в диапазоне от $\approx 0,9$ до 1. Отсюда следует, что использование решающих функций, близких к оптимальным, более целесообразно по сравнению с использованием евклидовых расстояний до математических ожиданий образцов в тех случаях, когда надежность распознавания при оптимальных решающих функциях равна или больше 90%.

Автор благодарен профессору А.А.Боровкову за ценные советы и дискуссии.

ЛИТЕРАТУРА

1. Welch P.D. and Wimpres R.S. Two Multivariate Statistical Computer Programs and their Application to the Vowel Recognition Problem. The Journal of the Acoustical Society of America 1961, vol.33, № 4, pp.426-434.
2. Загоруйко Н.Г., Лбов Г.С. Алгоритм упрощения линейных решающих функций. Вопросы радиоэлектроники, серия общетехническая, 1965, № 1.
3. Sebestyen G.S. Decision - Making Processes in Pattern Recognition, Macmillan Company, N 4, Lond., 1962.

4. Anderson T.W., Bahadur R.R. Classification into two multivariate normal distribution with different covariance matrices. The annals of Mathematical Statistics, June 1962, vol.33, N 2, pp.420-431.
5. Ковалевский В.А. Корреляционный метод распознавания изображений. Ж. вычисл. матем. и матем. физ., 1962, т. 2, № 4, с. 684-694.