АКУСТИЧЕСКИЙ ЖУРНАЛ, 2004, том 50, № 5, с. 590–602

УДК 519.22:534.87

ФУНКЦИЯ "ЛОЖНОГО ОБУЧЕНИЯ" КЛАССИФИКАТОРА АКУСТИЧЕСКИХ СИГНАЛОВ

© 2004 г. А. Н. Гончаров

Акустический институт им. Н.Н.Андреева РАН

117036 Москва, ул. Шверника, 4

E-mail: bvp@akin.ru

Поступила в редакцию 25.12.2003 г.

В работах по статистическому распознаванию образов, использующих обучение и экзамен, результат обучения зависит не только от эффективности признаков, но также и от соотношения мощности выбранной решающей функции, объема обучающей выборки и числа признаков. Расчет ошибок распознавания, связывающих эти основные величины для конкретного классификатора, обычно сложен, а вычисления приближенны и мало наглядны при оценке получаемых результатов в процессе исследований. Для оценки ошибок классификации предложено формировать отрезки случайного шума в аппарате распознавания вместо признаков реальных сигналов.

Целью данной работы является разработка простого, наглядного и эффективного способа экспериментальной оценки ожидаемых ошибок классификации используемого аппарата распознавания при обучении. Метод заключается в формировании выборки отрезков случайного шума, которая в режиме обучения вводится в используемый аппарат распознавания вместо признаков реальных сигналов. Части этой однородной выборки имитируют различные классы. Функция "ложного обучения" получается при последовательном увеличении для распознавания числа случайных признаков. Соответствующий рост вероятности распознавания искусственных классов при подобном "ложном обучении" зависит от объема "обучающей" выборки и мощности используемого решающего правила.

Основным результатом настоящей работы является предложенная функция "ложного обучения" для любого конкретного классификатора. Функцию получают для того же объема обучающей выборки, что и при распознавании реальных сигналов. Сравнивая эту функцию и экспериментальные вероятности по распознаванию сигналов при одинаковом числе использованных признаков, определяют меру надежности полученных результатов на реальных сигналах.

Применение простой функции "ложного обучения" полезно для характеристики достоверности любых экспериментальных результатов по статистическому распознаванию сигналов в акустике, сейсмоакустике, гидроакустике, при распознавании речи, в медицинской и технической диагностике, в радиолокации, а также других областях.

При создании новой распознающей системы для классификации сложных акустических сигналов или образов, состоящих из совокупности различных независимых компонент, порождаемых различными процессами в источниках [1–4], ведется поиск наиболее информативного комплекса признаков, образующих многомерное пространство классификации. Причем часто такой поиск проводится по ограниченному материалу обучения и экзамена. На этом этапе особенно важно оценивать достоверность результатов, поскольку вероятность распознавания существенно зависит от соотношения объема обучающей выборки и размерности признакового пространства у статистической распознающей системы. Примерами для распознавания акустических сигналов со сложной структурой могут быть в частности внутренние шумы пассажирских самолетов [1]. На основе анализа и классификации спектра этих шумов проводится диагностика состояния отдельных механизмов либо с целью снижения уровня их шумов, либо с целью диагностики их технического состояния. Другой пример. Сложный состав акустических шумов центробежных насосов со спектром, состоящим из неравномерной сплошной части и дискретных, линейчатых составляющих, осложняет задачу своевременного распознавания приближающегося момента возникновения опасной кавитации жидкости [2]. В этой задаче использование комплекса различных спектральных признаков в дополнение к наиболее информативной дискретной составляющей способно повысить вероятность диагностики кавитации. Аналогично, гидроакустические шумы морских объектов имеют в спектрах сигналов множество различных независимых составляю-

щих, и по спектральной структуре пришедших сигналов во многих задачах необходимо классифицировать их источники с возможно более высокой надежностью [3]. В другой актуальной проблеме распознавания звуков речи [4] для обеспечения надежности распознавания предъявляются весьма высокие требования как к информативности отдельных физических и формальных признаков, так и к формированию эффективного комплекса классификационных признаков с использованием оптимального решающего правила.

Во многих подобных проблемах объем исходных данных ограничен, в то время как области распознаваемых классов в многомерном пространстве признаков частично перекрываются и неудовлетворительно разделяются по малому числу признаков, а требования по надежности их распознавания предъявляются высокие. В таких случаях задачи решаются путем увеличения числа используемых признаков n и выбором более сложного и мощного решающего правила [5, 6]. Однако малый объем обучающей выборки L_{lr} в свою очередь жестко ограничивает как максимально допустимое число одновременно используемых признаков N (т.е. при статистическом распознавании – размерность признакового пространства), так и приемлемую сложность классификатора (число изменяемых при обучении параметров и мощность) [7]. При нарушении этих ограничений получаются недостоверные и завышенные результаты распознавания. В связи с этим важно иметь возможность оперативной оценки надежности получаемых результатов при поиске эффективного комплекса признаков. При этом в основном для случая нормально распределенных плотностей вероятностей признаков существуют различные методы расчета, связывающие основные взаимозависимые величины: вероятности ошибок классификации 1-го и 2-го рода – α и β; объемы обучающей и контрольной, экзаменационной выборок – L_{lr} и L_{ex}; число признаков или размерность пространства классификации N, а также расстояние Махаланобиса между средними значениями областей классов d в многомерном пространстве, которое для нормально распределенных признаков при равных ковариационных матрицах M в классах s_1 и s_2 имеет вид:

$$d^2 = (\mathbf{m}_2 - \mathbf{m}_1)^T M^{-1} (\mathbf{m}_2 - \mathbf{m}_1), \qquad (1)$$

где $\mathbf{m}_1$, $\mathbf{m}_2$ – вектора средних многомерных (n-мерных) областей классов s_1 и s_2;

M – общая ковариационная матрица;

T – здесь обычный знак транспонирования.

Физический смысл величины d – это расстояние между средними классов, нормированное на дисперсию их распределений.

В результате обучения распознающей системы по ограниченным выборкам классов при последовательном увеличении размерности признакового пространства обычно получают возрастающую кривую вероятности распознавания в функции числа признаков. При ограниченных выборках экспериментально полученные значения вероятности распознавания имеют тем меньшую статистическую обеспеченность, чем больше размерность решающего правила. Это связано с объективно возрастающими требованиями к объему статистик для получения многомерных распределений более высокой размерности в признаковом пространстве. Возникает задача оценки надежности, статистической достоверности результатов по многомерному распознаванию. Сопоставление экспериментальных результатов по вероятности распознавания реальных сигналов с аналитическими расчетами соответствующих вероятностей затруднено из-за сложности таких расчетов и их ориентации в основном на нормальные распределения признаков. Поэтому для ориентировочной, но простой и быстрой оценки надежности экспериментальных значений вероятности распознавания реальных сигналов предлагается получать для сопоставления другую кривую распознавания, также основанную на обучении. Для этой второй функции обучающие реализации получаются из отрезков однородного шума, имитирующих выборки с признаками классов, которые используются для обучения той же самой распознающей системы при аналогично возрастающей размерности многомерного пространства случайных признаков. Для таких шумовых выборок небольшого объема и при большой мощности решающего правила вероятность распознавания подобных условных классов довольно быстро увеличивается при увеличении размерности признакового пространства и даже может приближаться к значению 1.0 уже при числе случайных признаков порядка 10. Эту вторую функцию назовем "функцией ложного обучения". Объемы выборок для нее берутся такими же, как для реальных сигналов. Сравнивая с нею результаты обучения по распознаванию реальных сигналов, мы можем оперативно оценить возможность чисто случайного отбора реальных признаков в многомерный комплекс в очередном эксперименте и тем самым оценить надежность достигнутой вероятности распознавания на обучении по реальным сигналам при возрастающей размерности признакового пространства для используемого объема обучающих выборок в данной распознающей системе. Как уже было сказано, другой возможный способ оценки надежности экспериментальных результатов по вероятности распознавания состоит в сопоставлении с результатами аналитических расчетов соответствующей вероятности. Такой способ, хотя и возможен, но связан со сложностью применения расчетных методов.

Методы аналитического расчета ошибок классификации рассматриваются в основном в теоретических работах [7–12; 13]. В них приводятся общие расчетные зависимости ошибок распознавания α, $\beta(L_{tr}, L_{ex}, N)$ и даются характерные примеры для конкретных исходных данных, которые применимы в основном к нормальным и некоторым симметричным законам распределения совокупностей в классах. Кроме того, рассматриваются распределения, близкие к нормальным, а также лишь в некотором приближении учитывается влияние "хвостов" распределений [10]. При непараметрических методах статистического распознавания не рассматриваются ошибки, связанные с учетом этих "хвостов" из-за необходимости иметь значительно большую статистику для получения удовлетворительных оценок. Обычно при этом рассматриваются стандартные критерии принятия решений и типовые случаи как априорно известных распределений или известных только средних и ковариационных матриц, так и получаемых их оценок при непараметрических методах с неизвестными распределениями и параметрами [8–14].

Однако в известных нам теоретических работах по статистическим методам распознавания в многомерных пространствах фактически все же подразумевается не только нормальный или симметричный характер распределений в генеральной совокупности, но также и постоянство исходных распределений и их параметров, и их независимость от режимов функционирования источников, условий прохождения сигналов через среду и т.д. Но именно эти факторы в реальных задачах (неоднородная выборка) часто приводят к изменению как основных параметров распределений в классах (тренд средних, изменение дисперсий и другие), так и самих типов распределений (задаваемых или определяемых), приводя к многомодальности и к появлению вообще нетипичных распределений областей классов в признаковом пространстве. Все это затрудняет возможность использования предлагаемых методов аналитического расчета ошибок распознавания.

В случаях же нормальных распределений признаков и областей классов расчеты ошибок распознавания последовательно усложняются по мере возрастания неопределенности при задании исходных данных. Приведем типичные аналитические методы исследования ошибок классификации. При этом покажем, насколько расчетные методы оценки вероятности распознавания действительно опираются на нормальное распределение областей классов, а также то, что они сложны в применении на ответственном начальном этапе разработки распознающей системы и в процессе поиска наиболее информативного многомерного пространства признаков. В некоторых случаях мы будем особо отмечать отсутствие учета потерь по надежности получаемых расчетных оценок вероятностей распознавания. Рассмотрим далее кратко несколько таких характерных методов расчета в порядке их усложнения.

АНАЛИТИЧЕСКИЕ МЕТОДЫ РАСЧЕТА ОШИБОК РАСПОЗНАВАНИЯ

Для случая нормального распределения плотностей вероятностей областей двух равновероятных классов в пространстве n признаков с разными средними $\mathbf{m}_1$ и $\mathbf{m}_2$, равными ковариационными матрицами и с использованием байесовского классификатора в работе [8] получается общее выражение для оценки ошибок классификации. Затем для конкретного случая рассчитывается семейство кривых, связывающих вероятности ошибок распознавания p_e с определяемой по (1) величиной d: $p_e(d)$, и с параметром семейства – размерностью n признакового пространства $p_e(d; n)$ для n = 1; 3; 10; 20. Соответствующие графики приводятся в логарифмическом масштабе [8; рис. 5.3]. Там же получены оценки среднего и дисперсии величины p_e для случая неравных матриц ковариации и для разных объемов обучающей выборки: L = 12; 50; 100; 200 и 400. Для подобного же простого случая более наглядная зависимость $p_e(d)$ была получена в работе [9] при распознавании двух равновероятных классов s_1 и s_2, $p(s_1) = p(s_2) = 0.5$ с нормальными распределениями в одномерном признаковом пространстве при использовании байесовского классификатора. Этот рассчитанный простой пример только иллюстрирует полученные выражения для более общего случая распознавания нескольких классов $s_{i;j}$ с априорными вероятностями $p(s_i)$, $p(s_j)$, когда вероятность ошибки распознавания получается как сумма ошибок по классу s_i и по классу s_j:

$$P_e = P(s_i)\Phi\left(\left(\lambda - \frac{1}{2}d_{ij}\right)/\sqrt{d_{ij}}\right) + \\ + P(s_j)\left[1 - \Phi\left(\left(\lambda + \frac{1}{2}d_{ij}\right)/\sqrt{d_{ij}}\right)\right], \quad (2)$$

где: $\lambda = \ln[P(s_j)/P(s_i)]$ – логарифм порога $\theta = P(s_j)/P(s_i)$ при принятии решения о классе s;

$P(s)$ – априорные вероятности образов класса s;

Φ – табулированный интеграл вероятностей нормального распределения;

$d_{ij} = (\mathbf{m}_i - \mathbf{m}_j)^T M^{-1}(\mathbf{m}_i - \mathbf{m}_j)$ – расстояние Махаланобиса между классами s_i и s_j.

В случае равных априорных вероятностей классов $P(s_i) = P(s_j)$ порог равен 1, а величина λ

равна соответственно 0; тогда выражение (2) имеет вид [9]:

$$p_e = 1/2\Phi(-1/2\sqrt{d_{ij}}) + 1/2[1 - \Phi(1/2\sqrt{d_{ij}})] = \\ = \int_{(1/2)\sqrt{d_{ij}}}^{\infty} (1/\sqrt{2\pi})\exp(-y^2/2)dy. \quad (3)$$

В соответствии с выражением (3) для случая двух классов построен график $p_e(d)$ [9, рис. 4.4]. По графикам хорошо видно, что область основного спада ошибок классификации от максимальных до 5% и ниже соответствует области значений d от минимальных до значений 10–15, что уже соответствует мало пересекающимся, достаточно удаленным и хорошо различимым областям классов.

Несколько более обобщенные теоретические расчеты для аналогичных простых случаев приведены в работе [10], которая в значительной мере посвящена именно методам расчета ошибок распознавания. Рассматривая возможность повышения достоверности распознавания за счет увеличения объема выборки, в работе ошибки классификации α выражаются через табулированный интеграл вероятностей в функции от обобщенного параметра Ld:

$\alpha = 1 - \Phi[(c_0 + Ld/2)/\sqrt{Ld}]$, где: $c_0 = -\frac{8}{L}\ln(2\alpha_0) \le$

$\le d$ – условие, чтобы ошибки классификации не превышали заданного уровня α_0;

α_0 – верхняя граница Чернова для допустимых вероятностей ошибок α при классификации нормальных совокупностей.

Для частного случая распознавания двух классов с нормальными распределениями в n – мерном пространстве признаков с различными средними и с общей ковариационной матрицей M и порогом $c_0 = 0$, выражение для ошибок α и β имеет вид: $\alpha = \beta = 1 - \Phi(\sqrt{Ld}/2)$. Для этих условий полученная расчетным путем монотонно спадающая зависимость $\alpha(Ld)$ приводится в виде графика; по иллюстрации видно, что также как для d в выражении (3) ошибкам распознавания менее 5% соответствуют значения параметра кривой $Ld > 11$.

Мы постоянно обращаем внимание на иллюстративный графический материал в работах по следующим причинам. Обучающая выборка обычно используется при разработке классификатора, а экзаменационная, контрольная – при оценке его эффективности [9]. Мы поэтому считаем, что на первом этапе создания распознающей системы принципиально важна как простота расчета взаимной зависимости числа признаков, объема выборки и ошибок классификации, так и наглядность соответствующих оценочных критериев, упрощающая оперативную оценку текущих результатов исследований.

В несколько более сложных задачах, чем предыдущие, теоретический анализ связи ошибок распознавания с объемом выборки для заданных многомерных размерности N нормальных распределений классов с разными средними ($\mathbf{m}_1$ и $\mathbf{m}_2$) и ковариационными матрицами (M_1 и M_2) оказывается уже связан со сложными расчетами и реализуется с использованием метода статистического моделирования Монте-Карло [10, ч. 3, разд. 3.4]. Для этой задачи анализ проводился при дихотомической классификации с использованием решающего правила по методу максимального правдоподобия:

$$\lambda_L = \frac{1}{2}\sum_{l=1}^{L}[(\mathbf{x}_l - \mathbf{m}_1)^T M_1^{-1}(x_l - \mathbf{m}_1) - \\ - (\mathbf{x}_l - \mathbf{m}_2)^T M_2^{-1}(\mathbf{x}_l - \mathbf{m}_2)] + \\ + (L/2)\ln(\det M_1/\det M_2) \ge \theta, \quad (4)$$

где λ_L – логарифм отношения правдоподобия;

$\theta = \ln c$ – порог принятия решений о классе s_1 или s_2, определяемый в соответствии с выбранным критерием качества;

L – число реализаций в выборке, $l = 1, \ldots, L$;

$\mathbf{m}_1, \mathbf{m}_2$ – вектора средних для классов s_1 и s_2;

$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_l, \ldots \mathbf{x}_L)$ – значения N – мерного вектора $\mathbf{x}_l^{(n)}$ признаков $n = 1, \ldots, N$ в реализациях выборки с номерами l; каждая реализация выборки с номером l представлена набором из n конкретных действительных значений признаков $x_{n,l}$;

M_1, M_2 – ковариационные матрицы классов s_1 и s_2.

Для получения аналитических результатов и графических зависимостей α_0, $\hat{\alpha}(L_{lr}, L_{ex})$ и β_0, $\hat{\beta}(L_{lr}, L_{ex})$ в выражении (4) заменены параметры нормальных распределений их оценками $\hat{\mathbf{m}}_1$, $\hat{\mathbf{m}}_2$, $\hat{M}_1$, $\hat{M}_2$, так что полученные результаты сходятся к предельным α_0, β_0 лишь асимптотически при бесконечном увеличении объема обучающей выборки по обоим классам: $(L_{lr})_1, (L_{lr})_2 \longrightarrow \infty$. Конкретные расчетные графические материалы были получены в соответствии с (4) для следующих исходных данных: $\theta = 0$; $N = 3$; $(L_{lr})_1 = (L_{lr})_2$;

$$\mathbf{m}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}; \quad \mathbf{m}_2 = \begin{pmatrix} 1 \\ 0.5 \\ 0.5 \end{pmatrix}; \quad M_1 = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{pmatrix};$$

$$M_2 = \begin{pmatrix} 1.5 & 0.75 & 0.75 \\ 0.75 & 1.5 & 0 \\ 0.75 & 0 & 1.5 \end{pmatrix}.$$

Для этого случая области существенного замедления спада ошибок I-го и II-го рода (α и β) для трехмерного пространства признаков соответствуют объемам выборок $L \sim 15$–30, при которых вероятности ошибок соответственно имеют значения $\alpha, \beta \approx 0.05$–0.02, причем эти пределы α, β существенно зависят от соответствующих объемов выборок. В области значений L от 30 до 70 ошибки α и β асимптотически приближаются к нулю.

Определение функциональной зависимости вероятности ошибок распознавания от объемов выборок в случаях, когда типы одномерных функций распределения по каждому признаку неизвестны и поэтому используются непараметрические решающие правила, связано с более сложными расчетами, с использованием экспериментальных оценок законов распределения и проводится с использованием специально разработанных программ [10; разд. 5.3] даже в случае двухальтернативного одномерного непараметрического распознавания. При многомерном непараметрическом распознавании анализ и расчет вероятностей ошибок уже становится относительно сложным и громоздким.

При анализе задач статистической оптимизации признакового пространства классификатора, помимо использования методов декорреляции признаков и увеличения выборки рассматриваются также расчеты для методов обеспечения заданной вероятности ошибок классификации с помощью обоснованного увеличения числа признаков n с определением минимально необходимого их количества в функции обобщенного параметра ρ [10, гл. 5.4]:

$$\rho = n[\sum_{k=1}^{K}((L_{lr})_k + L_{ex})],\ n' \le n,$$ где: n – первоначальное число признаков; n' – число признаков после оптимизации; $(L_{lr})_k$ – объемы обучающих выборок по каждому из k классов s_k; L_{ex} – объем контрольной выборки.

На сложность расчетов при решении подобной задачи по оптимизации характеристик непараметрического обучения уже существенно влияет и степень перекрываемости областей классов, и неопределенность влияния "хвостов" распределений признаков. Отметим, что для получения достаточных статистик по "хвостам" необходима действительно представительная выборка из генеральной совокупности, а для оценки их влияния надо иметь данные о надежности распознавания. Далее используется численное интегрирование выражения для ошибок распознавания, а расстояние Махаланобиса заменяется его оценкой. Если для достижения заданного уровня ошибок распознавания по расчетам требуется объем выборок больше, чем имеется в распоряжении исследователей, задача переводится в другую плоскость – поиск возможности снижения ошибок с помощью увеличения числа признаков n. В работе [10], фактически посвященной этим проблемам, решение и этой задачи [гл. 5.4] опять сводится к применению нормальных распределений с принятием дополнительных ограничений: средние по обоим классам неизвестны, дисперсии одинаковы, общая матрица принимается диагональной, а упрощенное вследствие этого расстояние Махаланобиса d выражение (1) заменяется его оценкой снизу. Точность измерения d определяется практической разрешающей способностью системы Δd и принимается конкретно равной $\Delta d = 0.1$. При этих условиях за критерий оптимизации принимаются условия: $(2L_{lr} + L_{ex})N \longrightarrow \min$; $\alpha(L_{lr}, L_{ex}, N) \le \alpha_0$.

Уравнение для оптимизации вероятности ошибок по числу выборок и числу признаков решается при последовательных итерациях методом численного интегрирования. Получена зависимость $\alpha(n; L_{lr}, L_{ex})$ в функции от размерности признакового пространства. Семейство кривых приводится в виде графика [10, рис. 5.2]. В частности, проводилась также оптимизация для набора возрастающих максимально допустимых величин L от $(L_{lr})_{max} = 20$ и $(L_{ex})_{max} = 50$ до значений 400 и 200 соответственно. Эти величины выбраны, исходя из заданных d, α и N, а также исходя из объемов выборок жестко ограниченных временем допустимого обучения. Полученное оптимизированное число признаков N', необходимых для достижения заданных вероятностей ошибок α от 0.09 до 0.02, для названного набора возрастающих L последовательно снижалось соответственно от $N' = 7$ до $N' = 1$. При этом уже оптимизированное число необходимых объемов обучающей и экзаменационной выборок L'_{lr} и L'_{ex} также последовательно уменьшалось и составило для наибольших L'_{lr} вместо прежних 400 всего 120, а для L'_{ex} вместо 200 лишь 160. Семейства кривых функций $\alpha(N, L_{lr}, L_{ex})$ рассчитаны для $N = 1$–14 и $L_{lr} = 10$–200, $L_{ex} = 200$–400. Для больших выборок ($L > 100$ – 200) кривые быстро приближаются к уровню $\alpha \approx$ $\approx (0.01 - 0)$ при числе признаков $N \le 3$–5. В целом приведенные в качестве иллюстраций рассчитанные семейства кривых $\alpha(N)$ с параметрами L_{lr} и L_{ex} также, как и данные, приведенные выше, показывают, что и здесь используются фактически малопересекающиеся и хорошо разделимые области классов при нормальном распределении признаков в этих областях и достаточно больших объемах выборок. Однако если требуемый объем выборок превышает предел имеющихся, тогда для достижения заданной надежности распознавания просто используется метод увеличения числа признаков. Следует отметить, что это делается без оценки соответствующего снижения надежности распознавания.

ФУНКЦИЯ ЛОЖНОГО ОБУЧЕНИЯ

Таким образом, как видно из приведенного краткого рассмотрения некоторых характерных аналитических методов расчета зависимости вероятности ошибок статистического распознавания от размерности пространства признаков и объемов обучающих и контрольных выборок, все эти методы оказываются, как правило, достаточно трудоемкими и в основном относятся к нормальным распределениям областей классов и к традиционным вариантам решающих правил. В то же время при разработке новых распознающих систем именно на ответственном первоначальном их этапе, с одной стороны, определяется основной подход к синтезу структуры системы, но с другой – к этому времени еще не накоплено достаточных объемов выборок, не определен состав ни физических, ни формализованных признаков и не получены еще удовлетворительные оценки их распределений. Все это делает проблематичным использование таких расчетов для текущей оценки достоверности результатов поисковых экспериментальных исследований [5; 6; 7; 11; 12; 13: разделы 1.3, 2, 7, 8; 14]. Кроме того, на первом этапе создания распознающей системы обычно и само решающее правило, часто со значительными специфическими особенностями, либо выбирается, исходя из имеющихся возможностей, либо подбирается, исходя из предварительных данных о распределениях признаков в классах. Оптимизация же системы возможна на втором этапе ее разработки, на котором уточняется состав признаков, отобранных в наиболее эффективный их комплекс, проводится оптимальное их преобразование для сокращения размерности, а для полученного признакового пространства уже подбирается решающее правило, адекватное структуре границ классов в этом многомерном пространстве.

В связи с этим понятно, что особенно на первом этапе исследований важно иметь достаточно простой метод оперативной оценки максимально допустимого числа признаков, используемых в апробируемых на информативность многомерных их комплексах. Метод должен обеспечить упрощенную оценку надежности полученных результатов по вероятности распознавания в функции от числа признаков и от фактически имеющегося объема обучающей выборки из генеральной совокупности. В то же время такой метод должен учитывать мощность используемого конкретного решающего правила с учетом всех его практических особенностей, добавленных на основе эвристических подходов. Что касается мощности только самого решающего правила, то, в частности, дихотомическую мощность решающей функции $C_к$ (в отличие от мощности критерия) определяют как $C_к = 2(к + 1)$, где к – число регулируемых параметров. Величина $C_к$ для разделяющих границ в виде гиперплоскости равна $2(n + 1)$, где n – размерность признакового пространства; для гиперсферы – $2(n + 2)$; для поверхности 2-го порядка – $(n + 1)(n + 2)$; для полиномиальной поверхности порядка r – $2C^r_{n+r}$ (здесь C – число сочетаний) [9, разд. 2.5.3].

Помимо этого необходимо, чтобы такой метод оценки надежности результатов по распознаванию учитывал и существенные особенности процедур алгоритмического формирования и оценки информативности классификационных параметров при их анализе и отборе в рабочий комплекс признаков для получения эффективного многомерного признакового пространства. Сложность такого учета алгоритмических особенностей классификатора при расчетных методах оценки ошибок распознавания видна в частности из теоретической оценки связи необходимого объема выборки и заданной разрешающей способности измерения признаков.

Кроме зависимости вероятности классификации от факторов информационной неопределенности исходных данных, размерности пространства признаков и объемов обучающей и контрольной выборок, на определение вероятности ошибок распознавания и на оценку требуемых объемов выборки влияет также шаг $\Delta\theta_n$ при поиске по каждому признаку x_n оптимального порога θ_n, разделяющего области классов. Этот шаг вычисляется в соответствии: с распределениями признаков в классах, с разрешающей способностью $(\Delta x_n)_\varepsilon \approx (\Delta m_n)_\varepsilon$ измерения признаков x_n и расстояния между средними в классах в реальной системе, а также в соответствии с заданными стоимостями ошибок 1-го и 2-го рода по каждому из классов. Измерение нормированного по дисперсиям расстояния между средними значениями признаков по классам, то есть расстояние Махаланобиса d – не только оценивается с учетом разрешающей способности $(\Delta m_n)_\varepsilon$, но и в оптимальном случае не должно быть меньше нее, причем в простейшем варианте:

$$(\Delta d)_{\min} \geq (\Delta m)_\varepsilon/\sigma, \quad (\Delta d)_{\min} > 0. \qquad (5)$$

Когда в условии задачи распознавания повышается требование к точности измерения разности средних Δd, это фактически означает, что допустимое минимальное расстояние между средними в классах измеряется в соответствующих долях от величины дисперсии. Например принятие величины $\Delta d = 0.01$ означает, что далее может решаться задача с различающимися значениями средних на величину 0.01σ [10].

Математическое решение такой задачи оптимизации объемов обучающей и контрольной выборок в функции от требуемой достоверности распознавания (ошибки 1-го и 2-го рода) и в соот-

ветствии с задаваемой точностью измерения нормированной разности $(\Delta d)_{\min}$ между средними 2-х классов было проведено в работе [10, гл. 5.1] для случая одномерного параметрического распознавания. Вывод и решение всех уравнений для этого случая проведены при следующих условиях: число классов K равно 2; распределения областей классов – нормальные; средние по классам неизвестны, а дисперсии одинаковы; решение принимается по критерию максимального правдоподобия; вероятности ошибок распознавания α и β считаются равными и рассчитываются через интеграл вероятности. Используемое исходное выражение для вероятности ошибок в соответствии с условиями следующее:

$$\alpha = \beta = F(-m'/\sigma_2)F(m'/\sigma_1) + F(m'/\sigma_2)F(-m'/\sigma_1), \quad (6)$$

где $m' = (m_2 - m_1)/\sigma$;

$F(z)$ – интеграл вероятности:

$F(x) = [\Phi(x/\sqrt{2})/2] + \frac{1}{2}$, где для нормально распределенной x: $\Phi(x) = (2/\sqrt{\pi})\int_0^x \exp(-z^2)\,dz$.

После соответствующих подстановок это выражение используется для решения задачи оптимизации распознающей системы, т.е. для отыскания таких объемов L_{lr}^* и L_{ex}^*, которые удовлетворяют ограничениям и минимизируют принятый критерий ρ в соответствии с выражениями:

$$\rho = (2L_{lr} + L_{ex}) \longrightarrow \min;$$

$$\alpha = \beta = F(\Delta d/\sqrt{2/L_{lr}})F[-\Delta d/\sqrt{(2/L_{lr}) + (4/L_{ex})}] +$$

$$+ F(-\Delta d/\sqrt{2/L_{lr}})F[\Delta d/\sqrt{(2/L_{lr}) + (4/L_{ex})}] \le \alpha_0 = \beta_0.$$

Или, после преобразований, задача оптимизации имеет вид:

$$\Delta d^2 = (2L_{lr} + L_{ex}) \longrightarrow \min;$$

$$\Phi(\Delta d/\sqrt{L_{lr}/2})\Phi[-\Delta d/2\sqrt{(1/L_{lr}) + (2/L_{ex})}] \ge \quad (7)$$

$$\ge 1 - 2\alpha_0,$$

где: Φ – интеграл вероятности $\Phi(x) = 2/\sqrt{\pi}\int_0^\infty \exp(-z^2)dz$;

α_0 – заданное значение верхних границ вероятностей ошибок 1-го рода, $\alpha_0 = \beta_0$.

Решение этой задачи состоит в определении оптимальных значений L_{lr}^* и L_{ex}^*. Для каждого выбранного значения $\Delta d \equiv \Delta m'$ величина оптимизированных в соответствии с заданной величиной Δd объемов L_{lr}^* и L_{ex}^* находится при решении (6)–(7) стандартными методами целочисленного программирования, а искомый объем выборок выражается в обобщенном виде: $L_{lr}^*(\Delta d^2)$ и $L_{ex}^*(\Delta d^2)$ и получается в виде зависимости от заданной вероятности правильной классификации $(1 - \alpha_0)$.

Расчеты частного примера для этой задачи по условиям (7) показывают, что в случае, если требуемая точность измерения разности между средними по классам повышается (т.е. уменьшается Δd) в a раз (по отношению например к исходному значению 0.1 новая точность будет $0.1/a$), то тогда требуемый объем выборок по отношению к исходному объему L_{lr} возрастет в a^2 раз (т.е. $L_{lr}a^2$) по крайней мере для любых значений вероятностей правильного распознавания p выше 0.9 [10, гл. 5.1]. Поскольку обычно требуемые объемы выборок являются относительно большими числами, то такое их увеличение является весьма существенным и не всегда осуществимо.

Из анализа расчетных методов определения ошибок классификации понятно, почему на начальных стадиях исследований оценка достоверности экспериментальных результатов по многопараметрической классификации, как правило, проводится на интуитивном уровне. Это относится как к результатам по обучению, так и на экзамене. При этом, особенно при полученных вероятностях распознавания близких к 1.0, возникают излишне оптимистичные оценки результатов и их неадекватная физическая интерпретация из-за статистической ошибки обучения, связанной с избыточной размерностью признакового пространства, либо с завышенной мощностью решающего правила по отношению к недостаточному объему обучающей выборки. Кроме того, обычно реальные алгоритмы распознавания включают ряд эвристических операций, которые существенно влияют на ошибки классификации, но сложны для их учета в математических расчетах.

Известно, что экспериментально получаемая вероятность многопараметрической классификации весьма критично зависит от практического соотношения между объемом выборки и числом признаков. Так известно, что для 100%-ной классификации L реализаций (образов) при любой их группировке в L-мерном пространстве признаков достаточно использовать L линейных разделяющих границ [9; разд. 2.5.3]. Такая классификация возможна даже при полном перекрытии областей классов. При этом может и не происходить никакого обобщения, т.е. обучения классификации. Такое обучение назовем “ложным”.

Поэтому практика решения различных задач распознавания, в частности в акустике, доказывает необходимость наличия простой оценки надежности получаемых в реальных исследованиях вероятностей многопараметрической классификации. Это особенно важно, когда полученные вероятности близки к 1, так как такой результат может быть получен в значительной степени бла-

годаря завышенному числу признаков при ограниченном объеме выборки [5–7, 11, 12].

Для облегчения практической оценки статистических ошибок классификатора на обучении или на экзамене предлагается использовать статистическую функцию ложного обучения (ФЛО).

Функцией ложного обучения назовем такую зависимость вероятности p_r классификации обучающих выборок объемом L_r со случайными признаками $x_{l,n}$, $l = 1, \ldots, L_r$, набранными из однородного шума, которая получена в функции от возрастающего числа случайных признаков $n_r = 1, \ldots, N_r$ в многомерном решающем правиле при фиксированном объеме выборок. Таким образом, в случае распознавания, например 2-х классов A и B реальных акустических сигналов для обучения формируются выборки с объемами L^A и L^B, содержащие отдельные реализации с наборами признаков x_n реальных сигналов: $(x_{l,n})^{A;B}$, $l = 1, \ldots, L^{A;B}$; $n = 1, \ldots, N$. Для получения функции ложного обучения формируются шумовые обучающие выборки из случайного процесса, например с дискретно- равномерным распределением, и объемами L_r^A и L_r^B с отдельными реализациями, содержащими наборы случайных признаков: $(x_r)_{l,n}^{A;B}$, $l_r = 1, \ldots, L_r^{A;B}$; $n_r = 1, \ldots, N_r$ с соответствующими распределениями вероятностей.

Экспериментально ФЛО $p_r(n_r/L_r)$ получают в режиме обучения по разным частям единой однородной выборки шума, искусственно отождествляя эти части с разными классами. Теоретически расстояние Макаланобиса между классами s_i и s_j по всем признакам n_r для неограниченной однородной выборки из случайного шумового процесса равно нулю $d_{ij} \equiv 0$. Однако в малых выборках L_r^A и L_r^B его оценка $\hat{d}_{ij}$ для некоторых n_r может давать случайную возможность проведения ложной классификации при $\hat{d}_{ij}(x_r)_n > 0$, поскольку чем меньше выборка, тем больше дисперсия выборочных средних по классам и выше вероятность найти такие $(x_r)_n$, для которых разница между средними будет значимо больше нуля.

Таким образом, ФЛО получают, пропуская искусственно генерируемые реализации шума через тот же полный аппарат отбора признаков и классификации, с помощью которого были получены результаты распознавания реальных сигналов. По мере наращивания числа случайных признаков n_r, используемых в многомерном решающем правиле, искусственно разделенная на "классы" однородная шумовая выборка L_r начинает постепенно все "лучше" классифицироваться.

С увеличением числа признаков скорость нарастания вероятности такой "ложной" классификации $p_r(n_r)$ существенно зависит также и от мощности используемого решающего правила.

Приведем далее пример применения ФЛО при разработке одной распознающей системы. Для понимания примера приведем описание использованного алгоритма классификатора и множества реальных классификационных признаков сложного акустического сигнала.

Использованный алгоритм классификатора обладает сравнительно большой мощностью решающего правила и основан на предложенном в работах [15–17] методе (Федосеев В.А.; Брайловский В.Л; модифицированные варианты реализованы Черновым А.Н., Акустический институт им. акад. Н.Н. Андреева). Этот статистический алгоритм многопараметрического распознавания, используя байесовский подход, реализует линейное разделение классов в отбираемых наиболее информативных многомерных подпространствах признаков, оцениваемых по отношению правдоподобия.

Принципиально, структура алгоритма состоит из двух основных частей: предварительная обработка признаков по статистике выборки и принятие решений в многомерном признаковом пространстве. Исходные двумерные массивы конкретных значений признаков $x_{n,l}$ в реализациях обучающих выборок с объемами L_{lr}^A в классе A: $l^A = 1, \ldots, L_{lr}^A$ и L_{lr}^B в классе B: $l^B = 1, \ldots, L_{lr}^B$, сформированы отдельно для каждого класса $(x_{n,l})^A$, $(x_{n,l})^B$. В каждой реализации с номером l любого класса имеется N признаков, $n = 1, \ldots, N$; порядковый номер n каждого физического признака $x_{n,l}$ зафиксирован и одинаков во всех реализациях выборок классов A и B.

Сначала в процедуре предварительной обработки по обучающей выборке L_{lr} находятся либо оценки одномерных распределений признаков $x_{n,l}$, $n = 1, \ldots, N$ в классах, либо находятся только средние в классах и интервал между ними дробится на ряд отрезков. В использованном нами варианте алгоритма информативность признаков оценивается при проектировании распределений на координатные оси; проектирование на случайные направления здесь не рассматривается. Далее для пар классов по каждому признаку отыскиваются оптимальные пороги по какому-либо критерию принятия решений (идеального наблюдателя, Неймана-Пирсона или максимального правдоподобия). Из общего числа признаков отбирается состав наиболее информативных, количество которых с 5–10-кратным запасом превышает то, которое планируется использовать в окончательно сформированном многомерном решающем правиле. Значения этих признаков кодируются относительно своих порогов (возможны варианты: по двоичной, троичной, четверичной и т.д. систе-

мам) и в многомерном дискретном признаковом пространстве гиперкубов для каждого элемента квантованного признакового пространства находится своя оценка отношения правдоподобия λ. Опишем алгоритм подробнее.

Итак, пусть на входе системы имеются выборки N-мерных векторов признаков, представляющие распознаваемые классы (случай дихотомии) A и B:

$$X_l^A = (x_{l,1}^A, \ldots, x_{l,n}^A, \ldots, x_{l,N}^A), \quad l = 1, \ldots, \mathrm{L}^A;$$

$$n = 1, \ldots, N;$$

$$X_l^B = (x_{l,1}^B, \ldots, x_{l,n}^B, \ldots, x_{l,N}^B), \quad l = 1, \ldots, \mathrm{L}^B;$$

где L^A, L^B – число обучающих реализаций (образов) в классах A и B соответственно;

n – число признаков.

По выборкам L^A, L^B для каждого признака x_n вычисляется оценка минимальных, максимальных и средних значений по классам (либо по исходным значениям $x_{l,n}$, либо по робастной оценке):

$$\begin{gathered}(x_n^A)_{\min}, (x_n^B)_{\min}; (x_n^A)_{\max}, (x_n^B)_{\max}; \\ \langle x_n \rangle^A, \langle x_n \rangle^B; \quad n = 1, \ldots, N.\end{gathered} \tag{8}$$

В случае работы по средним в классах, расстояния между ними по каждому признаку

$|\langle x_n \rangle^A - \langle x_n \rangle^B|$ делится границами $\theta_\varepsilon^{(n)}$ на некоторое число E отрезков $\Delta_\varepsilon^{(n)}$:

$$\begin{gathered}\theta_\varepsilon^{(n)} = \inf_{A,B}(\langle x_n \rangle^A, \langle x_n \rangle^B) + \\ + (\left|\langle x_n \rangle^A - \langle x_n \rangle^B\right|/E)\varepsilon\Delta_\varepsilon^{(n)}, \text{ где } \varepsilon = 1, \ldots, E-1;\end{gathered} \tag{9}$$

$\Delta_\varepsilon^{(n)}(x_n)$ – шаг при дискретном изменении пробных границ классов;

$(E - 1)$ – число интервалов на отрезке $[\langle x_n \rangle^A, \langle x_n \rangle^B]$.

Величина E может варьироваться. Для каждой из границ как для пробного порога $\theta_\varepsilon^{(n)}$, разделяющего классы A и B, например по критерию идеального наблюдателя, далее по выборке определяется оценка суммарной ошибки распознавания p_e (если неизвестны априорные вероятности классов, то $P(A) = P(B) = 0.5$). При последовательном переборе значений $\theta_\varepsilon^{(n)}$ в выражении (9) величина ε окончательно выбирается такою, чтобы минимизировать суммарную ошибку $p_e = \alpha + \beta$, $\hat{p}_e^{(n)}(\theta_\varepsilon^{(n)}) = \hat{\alpha}^{(n)}(\theta_\varepsilon^{(n)}) + \hat{\beta}^{(n)}(\theta_\varepsilon^{(n)})$ разделения выборки на классы по каждому признаку x_n:

$$\hat{p}_e^{(n)} = \inf_\varepsilon[\hat{p}_e^{(n)}(\theta_\varepsilon^{(n)})], \quad \varepsilon = 1, \ldots, E-1.$$

Здесь $\hat{\alpha}$ и $\hat{\beta}$ – оценки ошибок распознавания соответственно 1-го и 2-го рода. Оценки вероятности ошибок $\hat{p}_e^{(n)}(\theta_\varepsilon^{(n)})$ или соответствующие оценки вероятности правильного распознавания $\hat{p}^{(n)}[\theta_\varepsilon^{(n)}] = 1 - p_e^{(n)}[\theta_\varepsilon^{(n)}]$ определяются в использованном варианте по критерию максимального правдоподобия Фишера с порогом, соответствующим минимуму средней ошибки:

$$\hat{p}^{(n)}[\theta_\varepsilon^{(n)}] = \sup_\varepsilon[(l_\varepsilon^A + l_\varepsilon^B)/(\mathrm{L}^A + \mathrm{L}^B)],$$

где l_ε^A, l_ε^B – здесь числа правильно распознанных по признаку x_n элементов из обучающих выборок L^A и L^B в классах A и B соответственно при использовании пробной границы $\theta_\varepsilon^{(n)}$ в качестве разделяющего классы порога.

Заметим здесь, что соотношение, с одной стороны, между разрешающими способностями при измерении: исходных признаков $(\Delta x_n)_\varepsilon$ расстояния между средними $(\Delta m_n)_\varepsilon$ (или в соответствии с (8) – между минимальными и максимальными значениями) и расстояния Махаланобиса $(\Delta d_{ij}^{(n)})_{\min}$, а с другой – величиной шага $\Delta\theta_n \equiv \Delta_\varepsilon^{(n)}(x_n)$ при поиске оптимального порога, может существенно влиять и на вероятность распознавания реальных сигналов, и на ФЛО. Как было показано в этом разделе выше, при фактической высокой или избыточной заданной точности измерения Δd по отношению к точности измерения признаков, то есть когда $(\Delta d)_{\min} < (\Delta m_n)_\varepsilon/\sigma$, могут резко возрастать требуемые объемы обучающей выборки (выражение (5) и далее). Но если при этом практически имеющиеся обучающие выборки малы, то тогда из-за недостаточности статистики будет существенно снижена надежность полученных, пусть даже высоких, вероятностей распознавания реальных сигналов в многомерном признаковом пространстве, а соответствующая ФЛО будет круче нарастать при увеличении размерности решения.

Далее в использованном для эксперимента алгоритме границы, соответствующие максимальным вероятностям распознавания по каждому признаку x_n, фиксируются в качестве порогов θ_n. Относительно этих порогов затем проводится двоичное кодирование (меньше порога – 0 и больше порога – 1) по всем признакам во всех реализациях выборки $x_{l,n}^A$ и $x_{l,n}^B$ и получаются двоично-кодированные признаки $y_{l,n}^A$ и $y_{l,n}^B$. Каждой двоичной области (0; 1) каждого признака ($y_{l,n} = 0$ или $y_{l,n} = 1$) соответствует свое соотношение чисел

попавших в нее реализаций из класса A и класса B и своя оценка вероятности распознавания в ней:

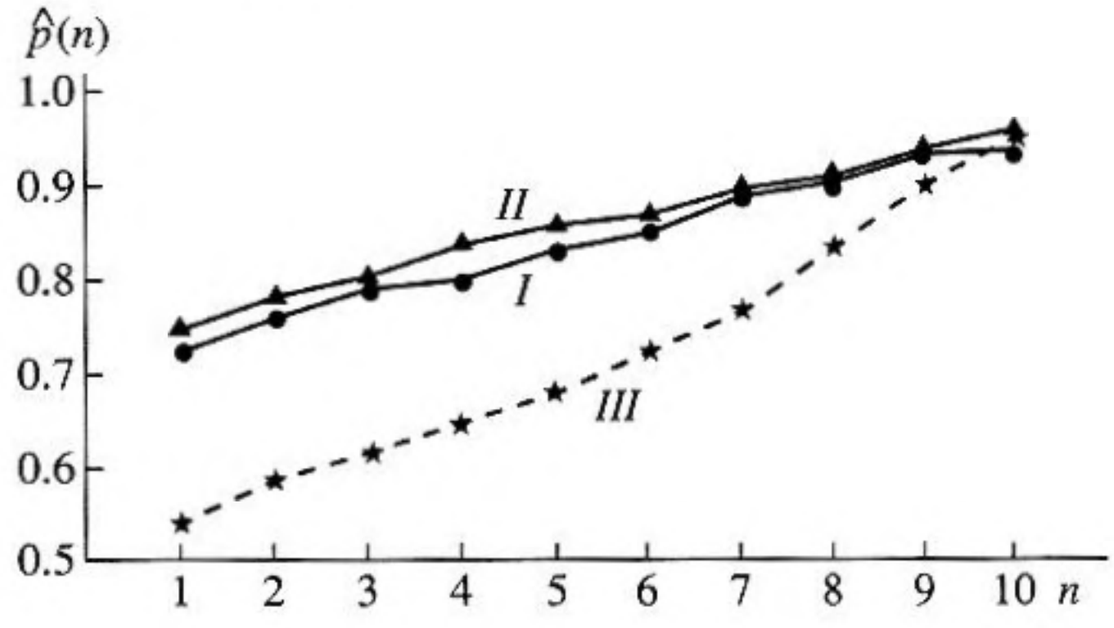
$$\hat{p}_n^A = l_n^A/\mathrm{L}^A; \quad \hat{p}_n^B = l_n^B/\mathrm{L}^B;$$

$$\hat{\lambda}_n = (l_n^A/\mathrm{L}^A)/(l_n^B/\mathrm{L}^B) \text{ или } \hat{\lambda}_n = (l_n^B/\mathrm{L}^B)/(l_n^A/\mathrm{L}^A),$$

где l_n^A, l_n^B – значения l_ε^A, l_ε^B для оптимальных значений $\theta_\varepsilon^{(n)}$. При оценке информативности признаков y_n для каждой области относительно порога (0;1) помимо общей вероятности правильного распознавания область идентифицируются по тому, к какому классу она в основном относится.

Кодированные двоичные признаки y_n ранжируются по информативности ($\hat{p}_n$) и для дальнейшей работы используются лишь те признаки, для которых информативность не меньше некоторого значения, так что после сокращения состава их число будет N', $N' < N$. На основе этих двоичных массивов отобранных признаков для всех реализаций обучающей выборки проводится новое кодирование различных пар признаков. В новом массиве каждой паре признаков в реализации соответствует двухразрядный двоичный код, каждый разряд которого соответствует значению (0; 1) одного из признаков y_n, y_{n+a} в этой реализации. Получается перекодированный массив с двухразрядными двоичными кодами для различных сочетаний (C_n^2) пар признаков y_n, y_{n+a}, который образует следующие подпространства в двумерном признаковом пространстве: 00; 01; 10; 11.

Оценка условных вероятностей распределения элементов выборки по этим подпространствам была описана еще в работах [15–17]. Здесь мы только кратко укажем, что в данном варианте алгоритма при полном переборе всех пар признаков y_n, y_{n+a} оценивается информативность получаемых подпространств по величине $\hat{p}_{n;\,n+\alpha}$ для каждой такой пары при фактическом учете зависимости между ними. Таким образом по обучающей выборке определяется наиболее информативная – "лучшая" пара. При этом для нее запоминается и соответствующий набор выборочных оценок классификационной информативности λ_n по каждому из четырех признаковых подпространств (00, 01, 10, 11), соответствующих статистическому материалу обучающей выборки. Далее в алгоритме к выбранной лучшей паре признаков подбирается в качестве дополнительного (без полного перебора всех возможных сочетаний из трех признаков) наилучший третий признак из числа ранее отобранных наиболее информативных. При этом используется трехзначная двоичная кодировка по трем признакам каждой реализации выборки, а метод оценки информативности соответствует предыдущим циклам. Для предлагаемой далее на рисунке иллюстрации применения ФЛО был использован вариант алгоритма, в котором максимальное число последовательно присоединяемых к лучшей паре наилучших дополнительных признаков равно восьми, так что конечное квантованное признаковое пространство для распознавания является десятимерным.



Зависимость вероятности распознавания от размерности пространства признаков. *I* – реальный сигнал, $L_A = 206$, $L_B = 214$; *II* – реальный сигнал, $L_A = 100$; $L_B = 100$; *III* – шумовая выборка, $L_r^A = L_r^B = 216$.

Для построения гиперплоскости, разделяющей классы в многомерном двоичном пространстве признаков, используются не все его элементарные области – гиперкубы, соответствующие различным многоразрядным кодовым наборам из 0 и 1, общим числом N'. Те из них, в которых соотношение реализаций классов λ оказалось относительно близко к 1.0

($\hat{\lambda}_n = (\hat{p}_n^A/\hat{p}_n^B) \approx 1 \pm g$, $g = 0.1$–0.3) не используются, поскольку в противном случае это может снизить вероятность распознавания всей системы [16, 17]. Область каждого гиперкуба идентифицируется с одним из классов A или B, а оценка вероятности правильного распознавания образов, попадающих в эту область, определяется приведенными выше величинами отношений: $\hat{p}_n^A$, $\hat{p}_n^B$, $\hat{\lambda}_n$.

Подпространства размерности 2–10, состоящие из конкретных наборов информативных элементарных областей – гиперкубов для каждого класса, разделяются гиперплоскостью в общем признаковом многомерном пространстве. Как результат синтеза решающего правила по обучающей выборке запоминаются все основные параметры сформированного общего признакового пространства, а также параметры подпространств каждого из классов и разделяющей их гиперплоскости. Эффективность этой распознающей системы оценивается по статистике контрольной, экзаменационной выборки.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА ПО ПРИМЕНЕНИЮ ФЛО

На рисунке приведена полученная экспериментально ФЛО $p_r(n_r/L_r)$ для алгоритма распознавания, изложенного выше. Используется предварительный отбор наиболее информативных признаков по критерию идеального наблюдателя и далее принимается решение в пространстве гиперкубов с разделяющей границей, реализуемой в виде логических булевых функций. (Этот эксперимент проводился с программным комплексом классификации Акустического института им. акад. Н.Н. Андреева, г. Москва). Как видно из рисунка (кривая III) при использовании 9–10 случайных признаков данное решающее правило способно с вероятностью 0.96 разделить на искусственные классы однородную ограниченного объема выборку модельного шума. Этому эксперименту соответствовали следующие данные:

– число реализаций (образов) в искусственно созданных двух классах равно $L_r^A = L_r^B = 216$ (общая выборка шума из 432 реализаций была разделена пополам);

– использован генератор шума с равномерной дискретной функцией плотности вероятности с округлением генерируемых в диапазоне 0–1 значений до точности $(\Delta x_n)_\varepsilon = (\Delta m_n)_\varepsilon = 0.001$; это – упрощенная модель распределения нормированных составляющих усредненного энергетического спектра сигнала, которые в действительности имеют бета – распределение плотности вероятности;

– в каждой реализации было сгенерировано по 64 случайных числа, имитирующих 64 исходных признака $N_r = 64$, из которых в дальнейшем по описанному выше алгоритму предварительно отбиралось по $N_r' = 40$ наилучших, т.е. дающих по отдельности более высокую вероятность разделения искусственных классов;

– при поиске оптимального порога по каждому отобранному случайному признаку разница между средними в классах разбивалась на 8 отрезков и анализировалось 7 пробных границ $\theta_\varepsilon^{(n)}$, $\varepsilon = 1, \ldots, 7$, как и при оценке информативности признаков реальных сигналов; при этом проведенное ранее округление генерируемых случайных чисел по точности $(\Delta x_n)_\varepsilon = (\Delta m_n)_\varepsilon$ до 0.001 не сопоставлялось с условием (5): $(\Delta d)_{\min} \geq (\Delta m)_\varepsilon/\sigma$;

– наиболее информативный случайный признак позволил решающему правилу распознать одну половину шумовой выборки от другой с вероятностью около $p_r \approx 0.53$;

– два наиболее информативных случайных признака, выявленные при полном переборе всех пар из 40, совместно (двумерное признаковое пространство и решающее правило) дали вероятность ложного распознавания $p_r \approx 0.58$;

– последовательное наращивание числа дополнительных наиболее информативных случайных признаков приводит к ускоряющемуся нарастанию вероятности ложной классификации, вплоть до значений $p_r \approx 0.96$ при совместном использовании 10 случайных признаков, т.е. по 10-мерному решающему правилу.

Приведенная на рисунке статистическая ФЛО (кривая III) соответствует только описанному выше решающему правилу и при фиксированном числе реализаций в классах (по 216). Понятно, что ход ФЛО будет меняться не только при изменении решающего правила или объемов выборки, но также при других функциях распределения вероятности модельного шума, при другом количестве исходных или отобранных признаков и при изменениях правила их отбора.

Для практического использования ФЛО необходимо сопоставить ее с экспериментальными результатами по классификации на обучении для реальных сигналов. Иллюстрация по применению ФЛО приведена на рисунке.

Кривые I и II на рисунке соответствуют экспериментально полученным результатам по классификации реальных акустических сигналов. Данные получены на начальном этапе исследований по поиску наиболее эффективного комплекса признаков для распознавания двух классов этих сложных сигналов в многомерном признаковом пространстве. Кривая III на том же рисунке – это ФЛО для той же предварительной обработки, для того же решающего правила и при тех же значениях параметров. Кривые I, II и III получены в режиме обучения классификации для описанного выше статистического алгоритма распознавания. Для формирования реальной обучающей выборки сначала по каждой реализации принятых акустических сигналов $x(t)$ получался текущий энергетический спектр $G(\omega)$:

$$G(\omega) = \lim_{T \to \infty} \frac{1}{2\pi} |S_T(\omega)|^2/T,$$

где $S_T(\omega)$ – текущий спектр, преобразование Фурье от выборки $x_T(t)$:

$$S_T(\omega) = \int_0^T x(t)\, e^{-j\omega t} dt = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t} dt,$$ здесь T – длина выборки реализации сигнала $x(t)$.

Спектры получались численно по дискретным отсчетам исходного сигнала с использованием алгоритма Кули-Тьюки для быстрого преобразования Фурье.

По специальному алгоритму проводился автоматический анализ спектра акустических сигналов и формировалось 64 разнородных спектральных признака, совместно представляющих минимально полную систему спектральных признаков сигнала. Одна часть признаков соответствует различным параметрам отдельных дискретных компонент, т.е. узкополостных выбросов спектра, не

входящих в гармонические звукоряды. Другая часть – различным параметрам групп дискретных компонент, кратных по частоте и входящих в звукоряды. Третья часть признаков состоит из параметров сплошной части спектров сигналов. Число классов – 2; они соответствуют различным источникам исследуемых акустических сигналов.

В соответствие с алгоритмом, изложенным выше, из общего числа этих признаков, равного 61 – 64, в программе предварительной обработки было отобрано по критерию максимального правдоподобия 40 наиболее информативных для данной выборки акустических сигналов. Перебором сочетаний по два признака из 40 определялась пара, дающая в двумерном пространстве минимум средней ошибки при классификации выборки. Из оставшихся 38 лучших отобранных признаков, начиная с наилучшего, выбирались наилучшие дополнительные к первой паре признаки; сначала присоединялся один, потом – два, три и так далее признаков для соответственно трех -, четырех – и так далее – мерных решающих правил. Как видно по рисунку, в связи с сильной коррелированностью признаков (декорреляция не проведена) наблюдается медленное нарастание вероятности классификации (кривые I и II) реальных сигналов 2-х классов. Кривая II на рисунке соответствует приблизительно половине общего числа реализаций (по 100 в каждом классе), кривая I получена по полному объему экспериментальных данных (216 и 208 реализаций в классах *A* и *B*).

Совместное размещение экспериментальной кривой вероятности классификации реальных акустических сигналов $p(n)$ по выборкам L_{lr} и кривой ФЛО $p_r(n_r)$ для шумовой выборки L_r на общем графике позволяет значительно объективнее интерпретировать результаты исследований. Для данного случая очевидно, что использовать состав формальных акустических признаков, вошедших в многомерный комплекс, можно не более, чем до совместного их числа, равного 5–6. При этом вероятность случайной классификации по ФЛО оказывается уже близкой к $p_r(n_r = 6) \sim$ ~ 0.7. При числе признаков $n = n_r > 5$–6 из сравнения с ФЛО становится очевидной малая надежность дальнейшего отбора реальных признаков сигнала в комплекс при данном решающем правиле и объеме выборки. Это связано с недопустимым повышением вероятности фактически случайного выбора последующих признаков реального сигнала и, следовательно, с ложным обучением. Полученные результаты по ФЛО были использованы при разработке многоальтернативного статистического классификатора акустических сигналов, в котором для проведения попарных дихотомий использовалось лишь по три признака.

Заметим, что и природа в созданных ею механизмах работы мозга избегает правил принятия решений повышенной размерности, хотя и не скупится на очень сложную, изощренную и многообразную обработку тех же самых, жизненно важных акустических сигналов. Например при обработке импульсных сигналов в слуховой системе дельфина распознавание важных особенностей принятых импульсных сигналов осуществляется последовательно не более чем по трем признакам, начиная с одного важнейшего [18]. Исследования также показали, что мозг человека для принятия решения представляет информацию преимущественно в двумерном виде и реже – в трехмерном.

В отличие от задания ФЛО в виде вероятностей для традиционных алгоритмов распознавания, для случая недетерминированных алгоритмов – нейронных сетей и персептронов – ФЛО может определяться, например, как число циклов Q работы алгоритма при адаптивном решении задачи. Так, для трехслойного персептрона с обратным распространением ошибки (из пакета Neural Networks Toolbox системы MATLAB) в качестве ФЛО была получена быстро убывающая функция числа циклов адаптивной подстройки структуры персептрона для обеспечения заданной вероятности распознавания $p = 0.9$. Алгоритм медленно различал две случайные выборки $L_r^A = L_r^B = 20$ из шума с дискретно-равномерным распределением при работе по одному-двум признакам $n_r = 1$–2 (использовано 300–500 циклов $Q(n_r = 1$–$2) = 300$–500) и значительно быстрее – по нескольким $n_r = 3, \ldots, 10$. В этом случае ФЛО имеет вид экспоненциально спадающей кривой $Q(n_r)$, имеющей при $n_r = 9$–10 значение 30–50 циклов. При отсутствии усреднения кривых разброс относительно аппроксимирующей экспоненты составляет 20–30%, так что ФЛО для персептрона имеет вид: $Q(n_r) \sim Q_0 \exp(-n_r a)[1 \pm (0.2 - 0.3)]$.

В заключение отметим, что простота получения ФЛО обеспечивает возможность ее широкого использования совместно с текущими результатами исследований по многопараметрическому распознаванию. Путем сравнения с ФЛО результатов по обучению распознаванию реальных сигналов для любого выбранного решающего правила можно приближенно, но оперативно оценить меру надежности полученных данных по многомерным комплексам признаков для исследуемого сигнала. Предложенная ФЛО позволяет также правильно строить план эксперимента и более объективно интерпретировать результаты многопараметрического распознавания реальных сигналов при разработке детерминированного статистического классификатора или нейросистемы.

Понятно, что при проведении некоторых модификаций алгоритма распознающей системы необходимо вновь получать ФЛО, адекватную новому

варианту. Такими модификациями являются: изменение алгоритма или параметров предварительной обработки и критериев оценки признаков; изменение решающего правила; изменение нормировки исходного пространства признаков; изменение шага при переборе пробных порогов; изменение числа первоначальных или лучших признаков для выбора из них наиболее информативных.

Применение простой и наглядной функции ложного обучения полезно для характеристики достоверности любых экспериментальных результатов по автоматическому статистическому распознаванию сигналов в акустике, в гидроакустике и сейсмоакустике, при распознавании дикторов по произношению, в медицинской диагностике, при диагностике неисправностей механизмов по шумам и вибрациям, а также при распознавании зрительных образов, в радиолокации и в других задачах.

Автор благодарит Л.Е. Казакевич за помощь в проведении экспериментов на ЭВМ и В.А. Баранова за помощь при расчете ФЛО для персептрона.

СПИСОК ЛИТЕРАТУРЫ

1. *Кузнецов В.М.* Проблемы снижения шума пассажирских самолетов // Акуст. журн. 2003. Т. 49. № 3. С. 293–317.
2. *Чудина М.* Шум как индикатор кавитации в центробежном насосе // Акуст. журн. 2003. Т. 49. № 4. С. 551–564.
3. *Кадашников С.К., Машошин А.И.* О статистическом описании канала распространения сигнала в применении к гидроакустической классификации // Акуст. журн. 1998. Т. 44. № 3. С. 462–469.
4. *Edward Chilton and Ioannis Paraskevas.* Audio fine classification using the statistical analysis of acoustic images // J. Acoust. Soc. Am. 2003. V. 113. № 4. Pt. 2. P. 2271. 145 th Meeting Acoust. Soc. Amer., Session 3pSP4, Signal Processing in Acoustical and Random Analysis.
5. *Raudys S.J. and Jain A.K.* Small Sample Size Effect in Statistical Pattern Recognition: Recommendations for Practitioners // IEEE Trans. Pattern Analysis and Machine Intelligence. 1991. V. 13. № 3. P. 252–264.
6. *Jain A.K. and Chandrasekaran B.* Dimentionality and Sample Size Consideration in Pattern Recognition Practice. Handbook of Statistics. P.R. Krishnaiah and L.N. Kanal. Amsterdam: North – Holland. 1982. Eds. V. 2. P. 835–855.
7. *Raudys S.J. and Pikelis V.* On dimensionality, Sample Size, Classification Error, and Complexity of Classification Algorithms in Pattern Recognition // IEEE Trans. Pattern Analysis and Machine Intelligence. 1980. V. 2. № 3. P. 243–251.
8. *Фукунага К.* Введение в статистическую теорию распознавания образов. М.: Гл. редакция физ-мат. литературы. 1979. С. 367.

Keinosuke Fukunaga. Introduction to Statistical Pattern Recognition. New York and London: Academic Press. 1972.

9. *Гонсалес Дж. Ту, Р.* Принципы распознавания образов. М.: Мир, 1978. С. 441.

Julius T. Tou, Rafael C. Gonzalez. Pattern Recognition Principles. London – Amsterdam – Don Mills, Ontario – Sydney – Tokyo. 1974.

10. *Фомин Я.А., Тарловский Г.Р.* Статистическая теория распознавания образов. М.: Радио и связь, 1986. С. 364 .
11. *Devroye L.* Automatic Pattern Recognition: A Study of the Probability of Error // IEEE Trans. Pattern Analysis and Machine Intelligence. 1988. V. 10. № 4. P. 530–543.
12. *Raudys S.J. and Duin R.P.W.* Expected Classification Error of the Fisher Linear Classifier with Pseudoinverse Covariance Matrix // Pattern Recognition Letters. 1998. V. 19. nos. 5–6. P. 385–392.
13. *Anil K. Jain, Fellow, IEEE, Robert P.W.* Duin and Jianchang Mao, Senior Member, IEEE. Statistical Pattern Recognition: // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000. V. 22. № 1. P. 4–37.
14. *Alberto Ruiz.* A Nonparametric Bound for the Bayes Error // Pattern Recognition. 1995. V. 28. № 6. P. 921–930.
15. *Брайловский В.Л., Лунц А.Л.* Формулировка задачи распознавания образов со многими параметрами и методы ее решения // Техническая кибернетика. М.: 1964. № 1. С. 20–31.
16. *Брайловский В.Л.* Алгоритм распознавания объектов со многими параметрами и его применение // Техническая кибернетика. М.: 1964. № 2. С. 30–39.
17. *Брайловский В.Л., Лунц А.Л.* Об оценке признаков в статистических решающих правилах // Техническая кибернетика. М.: 1967. № 3. С. 99–110.
18. *Дубровский Н.А., Зориков Т.В., Квижинадзе О.Ш., Куратишвили М.М.* Признаковое описание акустических импульсов при их различении слуховой системой афалины // Сенсорные системы. 1992. Т. 2. вып. 1. С. 37–48.

False Learning Function of an Acoustic Signal Classifier

A. N. Goncharov

Andreev Acoustics Institute, Russian Academy of Sciences, ul. Shvernika 4, Moscow, 117036 Russia
e-mail: bvp@akin.ru

Abstract—In works on statistical pattern recognition that use learning and examination, results of the learning depend not only on the feature efficiencies, but also on the proportion between the capacity of the decision rule, length of the learning sample, and number of features. It is usually difficult to calculate the recognition errors, which connect these basic quantities for a particular classifier, while the calculations are approximate and do not clearly characterize the results obtained in the process of the study. In evaluating classification errors, it is proposed to produce segments of random noise in the recognition engine, rather than the features of real signals.