электронное научно-техническое издание
НАУКА и ОБРАЗОВАНИЕ
Эл № ФС 77 - 30569. Государственная регистрация №0421100025. ISSN 1994-0408

Методика селекции признаков классификации в задачах распознавания образов сложных пространственных объектов

77-30569/316296

01, январь 2012

Гулевич С. П., Шевченко Р. А., Прядкин С. П., Веселов Ю. Г.

УДК 778.35:629.7

МГТУ им. Н.Э. Баумана
Радиотехнический институт имени академика А.Л. Минца
ВУНЦ ВВС «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина»
bauman@bmstu.ru
info@rti-mints.ru

ОСОБЕННОСТИ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ ОБРАЗОВ ДЛЯ СЛОЖНЫХ ПРОСТРАНСТВЕННЫХ ОБЪЕКТОВ

Анализ дифракционных изображений, полученных в результате рассеивания широкополосных радиолокационных сигналов на объектах со сложной трехмерной поверхностью, во многом зависит от качества обработки информации. На формирование радиолокационного дифракционного изображения сказываются ряд случайных факторов. Случайные факторы можно в зависимости от причин возникновения разделить на группы. Первая группа факторов определяется самим объектом - характером отражения и рассеивания сигнала от его поверхности. Ввиду структурной сложности дифракционного изображения объекта, многовариантностью ракурсов, внешнего оборудования и средств маскировки эта группа имеет детерминированную и случайную составляющие. Вторая группа связана с влиянием среды - рассеивание электромагнитного излучения в атмосфере и зависит от метеоусловий, замирания радиолокационного сигнала в атмосфере на участке “РЛ - объект”. Третья группа случайных факторов связана с приемной антенной. Проходя приемные тракты входных усилителей происходит оцифровка дифракционного изображения - квантование по мощности сигнала и квантованию по времени, обусловленные спектральной чувствительностью антенны и характеристиками АЦП входных каскадов приемного тракта. К этой же группе относятся помехи: естественные и искусственные (крыши домов, трубопроводы, линии электропередач, кабельные каналы и другие). К шумовому воздействию можно также отнести различные ложные объекты.

Успешность выполнения задачи анализа дифракционного изображения можно характеризовать показателями эффективности применяемыми к системам технического зрения. В большинстве систем технического зрения в качестве показателей используются вероятность правильного распознавания (классификации) или достоверность оцениваемых классификатором признаков РЛ изображения. Очевидно, что показатели эффективности зависят от дальности, помех, метеоусловий и целого ряда случайных факторов.

В статье для определенности будем рассматривать систему технического зрения с обучением по эталонным сигналам, в этой связи будем рассматривать ее работу на 2 этапах:

- проектирование и первичное обучение,
- эксплуатация и дообучение с целью распознавания новых объектов.

Рассмотрим некоторые принципы построения систем технического зрения, которые закладываются на первом этапе и определяют потенциальные возможности по распознаванию дифракционных изображений объектов. Общая схема решения задачи распознавания приведена на рисунке 1.

Рисунок 1 – Схема решения задачи распознавания

Изображение объекта, сформированное в приемном тракте РЛС и оцифрованное АЦП, будем обозначать ***Z***. По-существу, ***Z*** представляет собой двумерную матрицу размером N·M (размер в пикселях). Последовательно считывая строки матрицы ***Z,*** преобразуем её в вектор $\vec{Z}$ размерности M·N. Вектор $\vec{Z}$ будем называть *образом* ***Z***. Будем считать, что образ объекта может принадлежать к классу ω_j множества классов $\mathbf{\Omega}=\{\omega_j, j=1\ldots K\}$. Множество Ω счетно. Каждый класс определяется конкретной моделью объекта. Алгоритм соотнесения образа одному из классов ω_j будем называть *классификатором*. Если образу объекта $\vec{Z}$ в L-мерном метрическом пространстве признаков соответствует вектор $\vec{X}$. Из рисунка 1 видно, что выбор информационных признаков происходит перед первичным обучением

системы технического зрения, и после обучения оказывает определяющее значение на эффективность процесса распознавания дифракционного изображения.

В качестве признаков, как правило, выбирают следующие признаки изображения:

- геометрические (отрезки прямых, дуги – обобщенный метод Хафа, детекторы углов – методы Харриса и Томаши, контуры и другие особенности изображения);
- спектральные (спектральные характеристики на основе Фурье или вейвлет преобразования);
- структурные (на основе морфологического анализа и других методов);
- энергетические;
- статистические (на основе анализа распределений и статистик) и другие.

Без ограничения общности могут быть использованы любые признаки и их комбинации для обучения системы технического зрения (СТЗ). Будем предполагать, что при фиксированном выборе векторного пространства признаков классификатор обеспечивает оптимальное решение по распознаванию объекта.

Обработка признаков в СТЗ наиболее часто ведется тремя способами: отбором наиболее информативных признаков (отбором подпространств вектора признаков), образованием отношений отдельных признаков (отношений отдельных компонент вектора признаков) и образованием линейных комбинаций отдельных признаков. В теории распознавания используется термин “существенная размерность”, обозначающий минимальное число измерений (выделяемых признаков), необходимое для достаточно точной идентификации распознаваемых объектов. Поэтому при разработке новых и совершенствовании существующих СТЗ важно отобрать минимальное число признаков, обеспечивающих заданные показатели эффективности работы СТЗ без усложнения её конструкции, тем самым, повысить надежность работы и понизить стоимость её производства и эксплуатации.

Постановка задачи исследования

Выбор наиболее информативных признаков, описывающих множество классов объектов $\Omega=\{\omega_j,\ j=1\ldots K\}$, является ключевым моментом в задаче распознавания сложных по форме, спектру и другими признакам объектов, находящихся на “пестром” фоне. Задача ставится следующим образом: разработать критерии выбора информационных признаков образов, обеспечивающих решение задачи распознавания образов из K классов с максимальным уровнем вероятности распознавания $P_{\text{расп}\ \omega j}$ объекта. Классы ω_j и ω_i являются не пересекающимися: если $\vec{X} \in \omega_j$ и $\vec{X} \in \omega_i$, то следует $\vec{X} \in \emptyset$.

Допущения:

1 в кадре присутствует информация только об одном нормализованном образе;

2 векторы признаков $\vec{X}$ для всех классов имеют одинаковый закон распределения.

КРИТЕРИЙ РАЗДЕЛИМОСТИ КЛАССОВ С ТОЧКИ ЗРЕНИЯ ТЕОРИИ ИНФОРМАЦИИ

Для оценки информативности признаков воспользуемся положениями теории информации - используем понятия *объем информации* $I(\vec{Z})$ и *энтропия* $H(\vec{Z})$. Ральф Хартли в работе [1] предложил *объем информации* $I(\vec{Z})$, содержащийся в образе и связанный с реализацией события $\vec{Z} = \vec{z}_k$ с вероятностью p_k, определять как логарифмическую функцию:

$$I(\vec{z}_k) = \log\left(\frac{1}{p_k}\right) = -\log p_k, \qquad (1)$$

где основание логарифма, как правило, равно 2 *(*единица информации - *bit)* или ***e*** *(*единица информации - *nat).* Для определенности будем использовать последнюю. Ввиду того, что разрешение приемника ОЭС определено – *M·N*, тогда множество реализаций образов объекта можно рассматривать как случайное событие $\vec{Z} = \vec{z}_k$. Это множество счетно и ограничивается условием: $0 \le I(\vec{z}_k) \le M \cdot N$.

Здесь предполагаем, что минимальный объем информации соответствует достоверному событию ($p_k = 1$), а максимальный соответствует событию с наибольшей неопределенностью - реализация образа в виде 1 точки $p_k = 1/NM$).

Среднее значение объема информации $I(\vec{z}_k)$ при всех возможных реализациях образа называют *энтропией* $H(\vec{Z})$:

$$H(\vec{Z}) = E[I(\vec{z}_k)] = \sum_{k=1}^{MN} p_k \cdot I(\vec{z}_k) = -\sum_{k=1}^{MN} p_k \cdot \log p_k. \qquad (2)$$

По своей сути энтропия определяется как мера априорной неопределенности информации о системе (образе), учитывающая как вероятность наступления события в системе, так и число степеней свободы системы (сомножитель $\log p_k$).

Считая, что *0·log0=0* для энтропии $H(\vec{Z})$ получим диапазон:

$$0 \le H(\vec{Z}) \le \log(M \cdot N).$$

Под условной энтропией $H(\vec{X}|\vec{Y})$ будем понимать меру оставшейся неопределенности относительно $\vec{X}$ после того, как было получено наблюдение $\vec{Y}$. Величину $H(\vec{X},\vec{Y})$ определим как совместную энтропию:

$$H(\vec{X},\vec{Y}) = -\sum_{s=1}^{R}\sum_{k=1}^{L} p(x_k, y_s) \cdot \log p(x_k, y_s), \qquad (3)$$

где $p(x_k, y_s)$ - функция совместной вероятности случайных векторов $\vec{X},\vec{Y}$.

Так как энтропия $H(\vec{X})$ представляет неопределенность относительно входа системы перед наблюдением выхода системы, а условная энтропия $H(\vec{X}|\vec{Y})$ - ту же неопределенность после наблюдения выхода, то разность $H(\vec{X}) - H(\vec{X}|\vec{Y})$ будет определять ту часть неопределенности, которая была разрешена наблюдением выхода системы. Эта величина называется *взаимной информацией* между случайными векторами $\vec{X}, \vec{Y}$. Обозначив эту величину как $I(\vec{X},\vec{Y})$, можно записать[1]:

$$I(\vec{X},\vec{Y}) = H(\vec{X}) - H(\vec{X}|\vec{Y}) = \sum_{s=1}^{R}\sum_{k=1}^{L} p(x_k, y_s) \cdot \log\left(\frac{p(x_k, y_s)}{p(x_k) \cdot p(y_s)}\right), \quad (4)$$

т.е. $I(\vec{X},\vec{Y})$- величина разрешенной данной реализацией образа неопределенности.

Прежде чем приступить к решению задачи селекции минимального набора признаков решим 2 подзадачи.

1. Какой из двух наборов признаков F_X и G_X с распределениями $f_X(x)$, $g_X(x)$ несет больший объем информации об образе?

2. Как изменится взаимная информация $I(\vec{X},\vec{Y})$ между входом $\vec{X}$ и выходом $\vec{Y}$ классификатора при появлении гауссова шума в измерении признаков $\vec{X}$ образа $\vec{Z}$?

Рассмотрим первую подзадачу.

В работе [2] был сформулирован ***принцип максимальной энтропии***: *“Если выводы основываются на неполной информации, она должна выбираться из распределения вероятности, максимизирующего энтропию при заданных ограничениях на распределение”*. В [3] доказано, что принцип максимальной энтропии корректен и существует только одно распределение, обеспечивающее максимум энтропии, которое можно выбрать с помощью "аксиом согласованности":

1 уникальность (результат должен быть уникальным);

2 инвариантность (выбор системы координат не должен влиять на результат);

3 независимость системы (не должно иметь значения, будет ли независимая информация о независимых системах браться в расчет как в терминах различных плотностей вероятности раздельно, так и вместе, в терминах совместной плотности вероятности);

4 независимость подмножеств (не должно иметь значения, будет ли независимое множество состояний системы рассматриваться в терминах условной плотности вероятности или полной плотности вероятности системы).

Для случая многомерного *гауссового* распределения $f_X(\vec{X})$ центрированного вектора признаков $\vec{X}$ известно

$$f_X(\vec{X}) = \frac{1}{(2\pi)^{\frac{L}{2}}(\det(\boldsymbol{\Sigma}))^{1/2}} \cdot \exp\left(-\frac{1}{2}\,\vec{X}^T \boldsymbol{\Sigma}^{-1}\vec{X}\right). \quad (5)$$

Тогда доопределим энтропию для непрерывного вектора признаков $\vec{X}$ следующим выражением [7]:

$$H(\vec{X}) = \int f_X(\vec{X}) \log(f_X(\vec{X}))d\vec{X}\,. \qquad (6)$$

Последняя величина получается из (2) как предельный переход по количеству состояний для вектора признаков $\vec{\mathrm{X}}$ и в литературе [6] носит название *дифференциальной энтропии*. Поскольку на практике, как правило, используется квантованная оценка признаков и выборочные оценки вектора математического ожидания и матрицы ковариации, то нас будет интересовать только вывод некоторых выражений для оценки энтропии с переходом к дискретному представлению. Подставляя выражение (5) в интеграл (6), для многомерного гауссового распределения $\mathrm{f_X}(\vec{\mathrm{X}})$ получено выражение для энтропии $H(\vec{X})$[4]:

$$H(\vec{X}) = \frac{1}{2}[L + L \cdot \log(2\pi) + \log|\det(\boldsymbol{\Sigma})|], \qquad (7)$$

где L - размерность вектора признаков $\vec{X}$, $\boldsymbol{\Sigma}$ - матрица ковариации случайного вектора признаков $\vec{X}$, $\det(\boldsymbol{\Sigma})$- определитель матрицы $\boldsymbol{\Sigma}$.

Из выражения (7) вытекает 3 вывода.

1 Если вектор признаков $\vec{X}$ описывается гауссовой статистикой, то энтропия информации, содержащейся в $\vec{X}$ имеет наибольшее значение среди прочих векторов признаков $\vec{X}^I$, описываемых иной статистикой с таким же вектором средних значений. При этом выполняется неравенство: $H(\vec{X}) \geq H(\vec{X}^I)$, причем равенство достигается только при совпадении распределений $\vec{X}$ и $\vec{X}^I$.

2 Энтропия $H(\vec{X})$ гауссового случайного вектора признаков $\vec{X}$ однозначно определяется его ковариационной матрицей $\boldsymbol{\Sigma}$ и не зависит от среднего значения вектора признаков $\vec{X}$.

3 Понижение размерности вектора признаков $\vec{X}$ (например, с L до L-1) вызывает скачкообразное уменьшение энтропии, поэтому при выборе информативности признаков следует учитывать разбиение классов образов на подклассы. Это означает, например, что если у объекта изменяется ракурс и при этом какой-то из компонентов вектора признаков пропадает или появляется, то с этого ракурса объекта следует выделить новый подкласс данного объекта.

В этой связи может быть сформулирован ***критерий 1- разделение множества признаков сложных объектов на подклассы***: скачкообразное изменение энтропии вектора признаков $\vec{X}$ образа, принадлежащего некоторому классу, является необходимым и достаточным условием выделения его подкласса.

Рассмотрим вторую подзадачу.

Для этого рассмотрим линейный классификатор сигналов из вектора признаков (метод главных компонент, метод линейной нейронной сети). Пусть выход этого классификатора с учетом шума выражается соотношением:

$$Y = \sum_{i=1}^{L} \alpha_i \cdot x_i + \aleph, \qquad (8)$$

где α_i - *i*-й весовой коэффициент; $\aleph$ - гауссов шум с дисперсией ${\sigma_\aleph}^2$ с нулевым средним, оставшийся в признаках после их выделения из образа. Будем также предполагать, что компоненты случайного вектора признаков $\vec{X} = (x_1, x_2, \dots, x_L)^T$ также имеют гауссово распределение, тогда выход Y также будет иметь гауссово распределение, как взвешенная сумма гауссовых случайных переменных.

Учитывая свойство взаимности информации из (4) можно записать:

$$I(Y, \vec{X}) = H(Y) - H(Y|\vec{X}). \qquad (9)$$

Несложно заметить из (9), что функция плотности вероятности переменной Y для входного вектора $\vec{X}$ равна функции плотности вероятности суммы константы и гауссовой случайной переменной, следовательно, условная энтропия $H(Y|\vec{X})$ является "информацией", которую выход Y линейного классификатора (8) накапливает о шуме обработки сигнала $\aleph$, а не о самом векторе полезного сигнала $\vec{X}$. Исходя из этого, будем считать

$$H(Y|\vec{X}) = H(\aleph).$$

Тогда (9) перепишем в следующем виде:

$$I(Y, \vec{X}) = H(Y) - H(\aleph). \qquad (10)$$

Из выражения (7) для $H(Y)$ и $H(\aleph)$ получим:

$$H(Y) = \frac{1}{2}[1 + \log(2\pi\sigma_Y^2)], \qquad H(\aleph) = \frac{1}{2}[1 + L \cdot \log(2\pi{\sigma_\aleph}^2)].$$

После подстановки в формулу (6) и упрощения получим:

$$I(Y, \vec{X}) = \frac{1}{2}\log(2\pi\frac{\sigma_Y^2}{{\sigma_\aleph}^2}). \qquad (11)$$

Из выражения (11) вытекает 2 вывода.

1 Частное $\frac{\sigma_Y^2}{{\sigma_\aleph}^2}$ можно рассматривать как отношение сигнал/шум. Предполагая, что дисперсия ${\sigma_\aleph}^2$ является константой, взаимная информация $I(Y, \vec{X})$ достигает максимума при увеличении дисперсии σ_Y^2 выхода линейного классификатора.

2 Рост дисперсии шума $\sigma_{\text{ш}}{}^2$ снижает взаимную информацию между входным и выходным сигналами классификатора.

Теперь вернемся к решению основной задачи по разработке критериев выбора признаков, обеспечивающих решение задачи распознавания с максимальным уровнем вероятности распознавания $P_{\text{расп}}$ объекта.

Поскольку на основании зависимости (7) энтропия $H(\vec{X})$ для наиболее информативного набора признаков $\vec{X}$ образа определяется его ковариационной матрицей $\mathbf{\Sigma}$, то учитывая ее симметричный характер, выполним разложение Карунена-Лоева[5]:

$$\mathbf{\Sigma} = \mathbf{\Psi}^{\mathbf{T}} \cdot \mathbf{\Lambda} \cdot \mathbf{\Psi}, \tag{12}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{\text{L}} \end{bmatrix},$$

где $\mathbf{\Lambda}$ - диагональная матрица состоящая из собственных значений λ_{i}, а $\mathbf{\Psi}$ - матрица составленная из столбцов – собственных векторов, соответствующих λ_{i}. Заметим, что ранг матрицы $\mathbf{\Sigma}$ т.е. $L^I = rank(\mathbf{\Sigma}) \leq \text{L}$ определяет количество линейно независимых векторов признаков, отвечающих за классификацию $\vec{X}$. Выполняя процедуру нормализации для собственных значений и собственных векторов $\mathbf{\Sigma}$, получим:

$$\mu_i = \frac{\lambda_i}{tr(\mathbf{\Sigma})}.$$

Учитывая тот факт, что $\mathbf{\Psi}$ соответствует ортогональному преобразованию т.е. $\mathbf{\Psi}^{\mathbf{T}} \cdot \mathbf{\Psi} = 1$, для $\det(\mathbf{\Sigma})$ получим:

$$\det(\mathbf{\Sigma}) = \prod_{\text{i}=1}^{\text{L}^{\text{I}}} \mu_i. \tag{13}$$

Подставляя выражение (13) в (7), получим:

$$H(\vec{X}) = \frac{1}{2} \sum_{i=1}^{\text{L}^{\text{I}}} [1 + \log(2\pi) + \log \mu_i]. \tag{14}$$

В работе [5] показано, что максимальная погрешность разложения (12) определяется выражением:

$$\varepsilon^2 = E\{\|\Delta X(L)/X\|^2\} \leq \mu_{L^I}. \tag{15}$$

В (15) предполагается, что все собственные значения и их собственные вектора упорядочены в порядке убывания, так как значимость каждого компонента вектора признаков $\vec{X}$ определяется его собственным значением:

$$\mu_1 > \mu_2 > \cdots > \mu_{L^I}. \quad (16)$$

Рассмотрим ортогональное преобразование исходного вектора признаков $\vec{X}$образа:

$$\vec{Y} = \boldsymbol{\Psi^T} \cdot \vec{X},\ \boldsymbol{\Psi} \cdot \boldsymbol{\Psi^T} \cdot \vec{X} = \boldsymbol{\Psi} \cdot \vec{Y},\quad \vec{X} = \boldsymbol{\Psi} \cdot \vec{Y}, \quad (17)$$

т.е. матрица $\boldsymbol{\Psi}$ состоит из L^I линейно независимых векторов-столбцов ($|\boldsymbol{\Psi}| \neq \mathbf{0}$) и эти столбцы ортонормированны:

$$\Psi_i^T \Psi_j = \begin{cases} 1, & i = j \\ 0, & \mathrm{i} \neq \mathrm{j} \end{cases}.$$

Тогда для матрицы ковариации $\boldsymbol{\Sigma_X}$ получим:

$$\boldsymbol{\Sigma_X} \equiv \mathrm{E}\left[(\vec{\mathrm{X}} - \mathrm{E}\{\vec{\mathrm{X}}\}) \cdot (\vec{\mathrm{X}} - \mathrm{E}\{\vec{\mathrm{X}}\})^{\mathrm{T}}\right] = \mathrm{E}\left[(\boldsymbol{\Psi} \cdot \vec{Y} - \mathrm{E}\{\boldsymbol{\Psi} \cdot \vec{Y}\}) \cdot (\boldsymbol{\Psi} \cdot \vec{Y} - \mathrm{E}\{\boldsymbol{\Psi} \cdot \vec{Y}\})^{\mathrm{T}}\right] =$$

$$= \boldsymbol{\Psi} \cdot E\left[(\vec{Y} - E\{\vec{Y}\}) \cdot (\vec{Y} - E\{\vec{Y}\})^T\right] \cdot \boldsymbol{\Psi^T} = \boldsymbol{\Psi} \cdot \boldsymbol{\Sigma_Y} \cdot \boldsymbol{\Psi^T}. \quad (18)$$

На основании выражения (18) и с учетом (12) получим:

$$\det(\boldsymbol{\Sigma_X}) = \det(\boldsymbol{\Psi\Psi^T}) \cdot \det(\boldsymbol{\Sigma_Y}) = \det(\boldsymbol{\Sigma_Y}). \quad (19)$$

Можно сделать вывод о том, что ортогональное преобразование пространства признаков не меняет энтропии (7) для вектора $\vec{Y}$. Из свойств ортогонального преобразования вытекает, что оно не изменяет расстояний в пространстве признаков и эквивалентно повороту относительно центра рассеивания. Рассеивание, в нашем случае, описывается матрицей ковариации вектора признаков $\boldsymbol{\Sigma_X}$. Для многомерного гауссового распределения $\mathrm{f_X}(\vec{\mathrm{X}})$ из (5) видно, что поверхность постоянной плотности вероятности вектора признаков $\vec{\mathrm{X}}$ является поверхность второго порядка. Учитывая симметричный характер гауссового распределения относительно вектора математического ожидания $\vec{M}_X = \mathrm{E}\{\vec{X}\} = \left(\mathrm{m_{x1}}, \mathrm{m_{x2}}, \cdots, \mathrm{m_{xL^I}}\right)^{\mathrm{T}}$, установлено [7], что эта поверхность - L-мерный эллипсоид рассеивания. Поскольку в результате выполнения ортогонального преобразования матрица ковариации вектора признаков $\boldsymbol{\Sigma_Y}$ приведена к диагональному виду (12), то это означает, что полуоси эллипсоида указывают направления максимального рассеивания в L-мерном пространстве признаков.

Смысл ортогонального преобразования $\vec{Y} = \boldsymbol{\Psi^T} \cdot \vec{X}$ заключается в повороте системы координат относительно центра рассеивания $\vec{M}_X$ до совмещения координатных осей $\overrightarrow{OY}$ с главными полуосями рассеивания в L-мерном пространстве признаков. Так как каждый ω_j-й класс образов имеет свой эллипсоид рассеивания B_j, то задача выбора набора признаков, обеспечивающих решение задачи определения вероятности распознавания $P_{\text{расп}\,\omega_j}$ сводится к рассмотрению геометрической задаче с K эллипсоидами постоянной плотности вероятности.

ГЕОМЕТРИЧЕСКИЙ ПОДХОД К ЗАДАЧЕ РАЗДЕЛИМОСТИ КЛАССОВ

Для определенности будем рассматривать эллипсоиды постоянной плотности вероятности с полуосями равными $\sigma_{y_i}^j = 1/\sqrt{\lambda_i}$, где λ_i - собственные значения выборочной ковариационной матрицы (12). При этом правильность выбора вектора признаков $\vec{X}$ будет интерпретироваться как отсутствие пересечений эллипсоидов рассеивания различных классов. В этой связи может быть сформулирован ***критерий 2-разделимости классов***: множества признаков для классов разделимы тогда и только тогда, когда выполняются условия теоремы 1.

Теорема 1[8]. *Пусть* $\vec{X}^T \boldsymbol{A_1} \vec{X} = 1$ *и* $\vec{X}^T \boldsymbol{A_2} \vec{X} = 1$ *– квадрики в* R^L*, причем первая является эллипсоидом. Квадрики не пересекаются тогда и только тогда, когда матрица* $\boldsymbol{A_1} - \boldsymbol{A_2}$ *является знакоопределенной.*

Последняя теорема может быть использована для матриц $\boldsymbol{A_1} = \boldsymbol{\Sigma_1}^{-1}$ и $\boldsymbol{A_2} = \boldsymbol{\Sigma_2}^{-1}$. Применяя критерий Сильвестра[9] для матрицы $\boldsymbol{\Delta_{12}} \equiv \boldsymbol{\Sigma_1}^{-1} - \boldsymbol{\Sigma_2}^{-1}$, необходимо проверить, что все главные миноры матрицы $\boldsymbol{\Delta_{12}}$ положительны, либо все главные миноры отрицательны.

Непересечение эллипсоидов рассеивания различных классов является необходимым условием при выборе эффективного набора признаков классификатора образов. В случае зашумления вектора признаков среднеквадратические отклонения ошибок (СКО) компонент вектора признаков, а, следовательно, и главные оси эллипсоидов рассевания будут расти, что, в конце концов, приведет к их пересечению и возникновению состояния неопределенности. Для контроля выполнения необходимого критерия выбора признаков и оценки устойчивости классификатора к шумам можно использовать расстояние между близкими классами (эллипсоидами рассеивания). Для этого воспользуемся следующей теоремой.

Теорема 2[8]. *Если выполняется условие теоремы 1, то квадрат расстояния между эллипсоидом* $\vec{X}^T \boldsymbol{A_1} \vec{X} = 1$ *и квадрикой* $\vec{X}^T \boldsymbol{A_2} \vec{X} = 1$ *совпадает с минимальным положительным корнем полинома*

$$F(z) = D_\rho(\boldsymbol{det}[\rho \boldsymbol{A_1} + (z-\rho)\boldsymbol{A_2} - \rho(z-\rho)\boldsymbol{A_1}\boldsymbol{A_2}]),$$

в предположении, что этот корень не является кратным. Здесь D_ρ*— дискриминант полинома рассматриваемого относительно переменной* ρ.

Таким образом, для каждой пары близких эллипсоидов рассеивания с матрицами $\boldsymbol{A_1} = \boldsymbol{\Sigma_1}^{-1}$ и $\boldsymbol{A_2} = \boldsymbol{\Sigma_2}^{-1}$ необходимо составить матрицу размерности $L \cdot L$: $\boldsymbol{D}(z,\rho) = \rho \boldsymbol{A_1} + (z-\rho)\boldsymbol{A_2} - \rho(z-\rho)\boldsymbol{A_1}\boldsymbol{A_2}$. Далее, разложив детерминант $\boldsymbol{det}[\boldsymbol{D}(z,\rho)]$ в виде характеристического многочлена по степеням ρ, получим полином порядка $n=L+2$: $d(z) \equiv \boldsymbol{det}[\boldsymbol{D}(z,\rho)] = d_0 + d_1\rho^1 + \cdots + d_n\rho^n$, где $d_i = d_i(z)$. Отметим, что дискриминант полинома равен нулю тогда и только тогда, когда полином имеет кратные корни. Поскольку машинное определение выборочной оценки ковариационных матриц выполняется с точностью до погрешностей округления, то событие получения кратных корней практически является маловероятным. Выражение для дискриминанта имеет вид:

$$F(z) = ((-1)^{n(n-1)/2}/d_n) \cdot R\left(d, d'_\rho\right),$$

где $R\left(d, d'_\rho\right)$ - результант полинома $d(z)$ и его производной $d'_\rho(z)$

$$d'_\rho(z) = d_1 + 2d_2\rho^1 + \cdots + nd_n\rho^{n-1}.$$

Результант $R\left(d, d'_\rho\right)$ равен определителю следующей *(2n-1)·(2n-1)* матрицы:

d_n	d_{n-1}	d_{n-2}	.	.	.	d_0	0	.	.	0
0	d_n	d_{n-1}	d_{n-2}	.	.	.	d_0	0	.	0
.	.	.	.	.	.	.	.	.	.	.
0	0	0	0	0	d_n	d_{n-1}	d_{n-2}	.	.	d_0
nd_n	$(n-1)d_{n-1}$	$(n-2)d_{n-2}$	.	.	d_1	0	0	.	.	0
0	nd_n	$(n-1)d_{n-1}$	$(n-2)d_{n-2}$	.	.	d_1	0	0	.	0
0	0	nd_n	$(n-1)d_{n-1}$	$(n-2)d_{n-2}$	.	.	d_1	0	.	0
.	.	.	.	.	.	.	.	.	.	.
0	0	0	0	0	0	nd_{n-1}	$(n-1)d_{n-2}$	$(n-2)d_{n-2}$	.	d_1

Тогда, выбирая минимальный положительный корень - $\tilde{z}$ полинома $d(z)$, получим евклидово расстояние между двумя эллипсоидами в L-мерном пространстве признаков:

$$D_{12} = \sqrt{\tilde{z}}\ . \qquad (20)$$

Отметим, что указанная процедура оценки расстояний может быть рассчитана на этапе проектирования классификатора. Выполняя пропорциональное расширение (*k*-коэффициент пропорциональности) главных полуосей эллипсоида рассеивания

$$\sigma_{yi}^{*} \to k \cdot \sigma_{yi}, \qquad k > 1$$

получим таблицу соответствий $k \to k\sigma_{yi} \to D_{js}$ расстояний между классами *j* и *s*. Максимальное значение *k* при котором $D_{js} > 0$, определит запас устойчивости выбранного набора L признаков к воздействию шума.

При некотором упрощении задачи может быть получена оценка вероятности распознавания. Выполняя процедуру нормализации и ранжирования (16) собственных значений для выборочной ковариационной матрицы каждого класса, можно оставить только три значимых компоненты вектора признака:

$$\mu_1 > \mu_2 > \mu_3.$$

В работе [7] найдено аналитическое выражение для вероятности попадания случайного трехмерного вектора признаков с гауссовой плотностью вероятности в эллипсоид рассеивания:

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2} = k^2.$$

Последнее выражение было получено для центрированного вектора признаков. В более общем случае - рассматриваем несколько эллипсоидов в одной системе координат, $\vec{M}_X = (m_x, m_y, m_z)^{\mathrm{T}} \neq 0$ и уравнение эллипсоида рассеивания выглядит так:

$$\frac{(x-m_x)^2}{\sigma_x^2} + \frac{(y-m_y)^2}{\sigma_y^2} + \frac{(z-m_z)^2}{\sigma_z^2} = k^2.$$

Здесь, как и при составлении таблицы расстояний между классами, полуоси эллипсоида (*a,b,c*) пропорциональны СКО компонент вектора признаков:

$$a = k \cdot \sigma_x,\ b = k \cdot \sigma_y,\ c = k \cdot \sigma_z. \qquad (21)$$

В соответствии с выражением (5) для постоянной плотности вероятности вектора признаков при L=3, потребуем для степени экспоненты:

$$D_{\mathrm{M}}^2 \equiv (\vec{\mathrm{X}} - \vec{M}_X)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\vec{\mathrm{X}} - \vec{M}_X) = k^2, \qquad (22)$$

где D_{M}^2 - квадрат расстояния Махаланобиса (предложено индийским математиком в 1936 году и носит его имя), являющегося обобщением меры расстояния в многомерном

пространстве признаков с учетом характера рассеивания вектора признаков разных классов. Подставляя в (22) выражение для $\boldsymbol{\Sigma^{-1}} = \boldsymbol{\Psi^{T}} \cdot \boldsymbol{\Lambda^{-1}} \cdot \boldsymbol{\Psi}$ (следует из (12), свойства ортогонального преобразования $\boldsymbol{\Psi^{T}} = \boldsymbol{\Psi^{-1}}$ и $\boldsymbol{\Sigma^{-1}} \cdot \boldsymbol{\Sigma} = \mathbf{E}$), получим:

$$(\vec{X} - \vec{M}_X)^{T} \boldsymbol{\Psi^{T}} \cdot \boldsymbol{\Lambda^{-1}} \cdot \boldsymbol{\Psi} (\vec{X} - \vec{M}_X) = k^2.$$

Матрица $\boldsymbol{\Psi} = (\vec{\gamma_1}, \cdots, \vec{\gamma_L})$ состоит из столбцов – собственных векторов соответствующих собственным значениям, например, $\vec{\gamma_1} \rightarrow \lambda_1$. В последнем выражении матрица $\boldsymbol{\Lambda^{-1}}$ на основании (12) определяется так:

$$\boldsymbol{\Lambda^{-1}} = \begin{bmatrix} 1/\lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\lambda_L \end{bmatrix}.$$

Выполняя замену переменных $\vec{X'} \equiv \boldsymbol{\Psi} (\vec{X} - \vec{M}_X)$, получим:

$$D_M^2 = \vec{X'}^{T} \cdot \boldsymbol{\Lambda^{-1}} \cdot \vec{X'} = \frac{(x' - m_x)^2}{\sigma_x^2} + \frac{(y' - m_y)^2}{\sigma_y^2} + \frac{(z' - m_z)^2}{\sigma_z^2} = k^2.$$

Из последнего выражения следует:

$$\lambda_1 \equiv \sigma_x^2 k^2, \ \lambda_2 \equiv \sigma_y^2 k^2, \ \lambda_3 \equiv \sigma_z^2 k^2. \qquad (23)$$

Далее получим выражение для вероятности попадания вектора признаков $\vec{X}=(x,y,z)^T$ в эллипсоид B_j:

$$P(\vec{X} \subset B_j) = \frac{1}{(2\pi)^{3/2} \sigma_x \sigma_y \sigma_z} \iiint_{B_j} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2}\right)} dxdydz.$$

Переходя к сферическим координатам и интегрируя по частям, получим:

$$P(\vec{X} \subset B_j) = 2 \cdot \Phi(k) - 1 - \sqrt{\frac{2}{\pi}} \cdot k \cdot e^{-\frac{k^2}{2}}, \qquad (24)$$

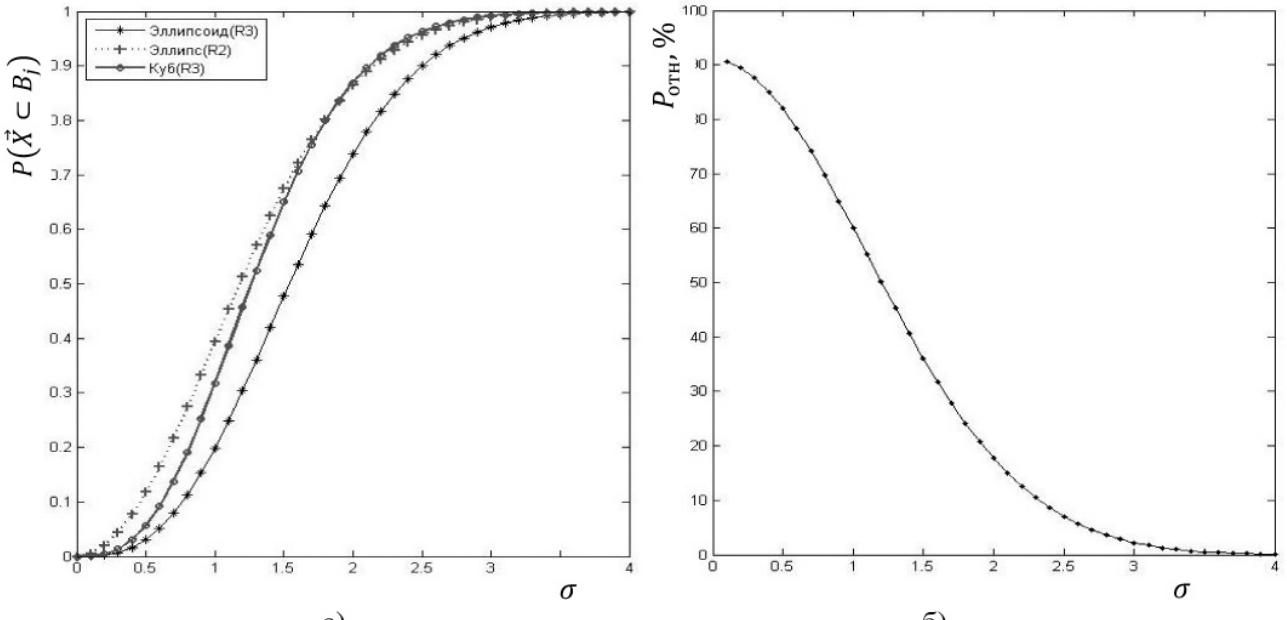
где $\Phi(k) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{k} e^{-\frac{t^2}{2}} dt$ - функция Лапласа[7].

В [6] показано, что уровень значимости признаков, как правило, резко падает с увеличением ранжированного номера собственного значения выборочной матрицы ковариации и на основании (15) относительная ошибка представления вектора признаков для размерности L>3 составит единицы процента. Реализация вектора признаков $\vec{X}=(x,y,z)^T$ позволяет выра-

зить отклонение от центра эллипсоида рассеивания $\vec{M}_X = (m_x, m_y, m_z)^T$ через расстояние, выраженное в среднеквадратических отклонениях, аналогично (21):

$$x - m_x = k \cdot \sigma_x, \qquad y - m_y = k \cdot \sigma_y, \qquad z - m_z = k \cdot \sigma_z.$$

Поскольку значения СКО были получены на этапе первичного обучения, то для текущего вектора признака $\vec{X}=(x,y,z)^T$ найдем $k = k_j$ для заданного класса с вектором $\vec{M}_X$ матрицей $\mathbf{\Sigma_j}$. Подставляя k_j в (24) получим вероятность попадания в эллипсоид $P(\vec{X} \subset B_j)$. Обратно, задавая вероятность $P(\vec{X} \subset B_j)$, получим размер эллипсоида k_j, т.е. расстояние Махаланобиса D_{M_j}. Так для $P(\vec{X} \subset B_j) = 0.997$ вычислено *k=3,9*, что превышает оценку для трехмерного нормального распределения *k=3,4* (обобщенное «правило трех сигм» на пространство R^3). Это факт свидетельствует о том, что не правомерно использовать аппроксимацию многомерного эллипсоида соответствующей размерности параллелепипедом (как это делается в большинстве расчетов), так как в этом случае получается завышенная оценка вероятности, а доверительные интервалы имеют заниженную оценку. Для эллипсоида рассеивания доверительный интервал составил 3,9 СКО (на 15 % больше).



а) б)

а) - вероятность попадания вектора признаков в эллипсоид, куб и эллипс,
б) - отношение вероятностей попадания вектора признаков в куб и эллипсоид

Рисунок 2 – Зависимость вероятности рассеивания признаков от СКО

При использовании более низкого уровня доверительной вероятности, как видно из рисунка 2.а) доверительный интервал уменьшается и при малых значениях СКО (рисунок 2.б) такое несоответствие весьма существенно.

Обобщая на случай размерности нормально распределенного вектора признаков с $L>3$ и учитывая недиагональность матрицы ковариации, можно построить доверительную область с помощью обобщения распределения Стьюдента на случай $L>1$. В теории многомерного статистического анализа[10] для построения доверительной области оценки неизвестного математического ожидания вектора признаков $\vec{M}_X = (m_1, m_2, \cdots, m_L)^{\mathrm{T}}$ используется статистика Хотеллинга ($\boldsymbol{T^2}$).

Для выборки объемом n нормально распределенного вектора признаков $\vec{X}$ размерностью L и уровнем значимости $\alpha = 1 - q$ (q - доверительная вероятность) величина

$$T_\alpha^2 = n \cdot (\overline{\boldsymbol{X}} - \vec{M}_X)^T \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{X}} - \vec{M}_X) \qquad (25)$$

определяется через $F_{\nu_1,\nu_2}(\alpha)$ распределение Фишера-Снедекора[10] со степенями свободы $\nu_1=L$ и $\nu_2=n-L$ в соответствии с выражением:

$$T_\alpha^2 = \left[\frac{L(n-1)}{n-L}\right] \cdot F_{\nu_1,\nu_2}(\alpha). \qquad (26)$$

Здесь $\overline{\boldsymbol{X}}$ – вектор среднего по выборке, $\boldsymbol{\Sigma}$ – выборочная оценка матрицы ковариации. Поскольку из (25) величина $k^2 = T_\alpha^2/n$ определяет в соответствии с (22) размер эллипсоида рассеивания вектора математического ожидания, то получаем доверительную области с уровнем значимости α.

На рисунке 2 изображена зависимость распределения Фишера-Снедекора от объема выборки n для уровня значимости $\alpha=0,01$. Так, например, для размерности вектора признаков 6 и объема выборки 31 получим числа степеней свободы $\nu_1=6$ и $\nu_2=25$, тогда из 1 графика получим $F_{6,25}(0,01) = 3,94$. В соответствии с (26) получим:

$$k^2 = \frac{T^2}{n} = \left[\frac{6 \cdot 30}{25 \cdot 31}\right] \cdot 3,94 = 0,92.$$

В соответствии с (25) величина k^2 определяет размер эллипсоида рассеивания вектора математического ожидания - центра эллипсоида. Приводя эллипсоид рассеивания к каноническому виду (путем ортогонального преобразования осей координат), получим вместо (25) выражение:

$$\frac{\varepsilon_i^2}{\sigma_i^2} \le k^2, \qquad \varepsilon_i \le k \cdot \sigma_i, \qquad (i = 1 \cdots L),$$

где $\vec{\varepsilon} = \vec{X} - \vec{M}_X$, а $\sigma_i^2 (i = 1 \cdots L)$- квадраты полуосей эллипсоида рассеивания (диагональные элементы матрицы $\mathbf{\Sigma}$- оценки дисперсий компонент вектора признаков по обучающей выборке).

Таким образом, может быть сформулирован ***критерий 3-объема обучающей выборки классификатора***: уровень значимости α, оценки дисперсий компонент вектора признаков- $\sigma_i^2 (i = 1 \cdots L)$ и размерность пространства признаков L определяют объем обучающей классификатор выборки - n_j, для каждого класса $\omega_j, j = 1 \cdots K$ в соответствии с выражениями (25), (26).

Как правило, в реальных классификаторах выбирают максимальный объем обучающей выборки для всех классов, т.е. $N = sup\{ n_j \mid \omega_j, j = 1 \cdots K \}$.

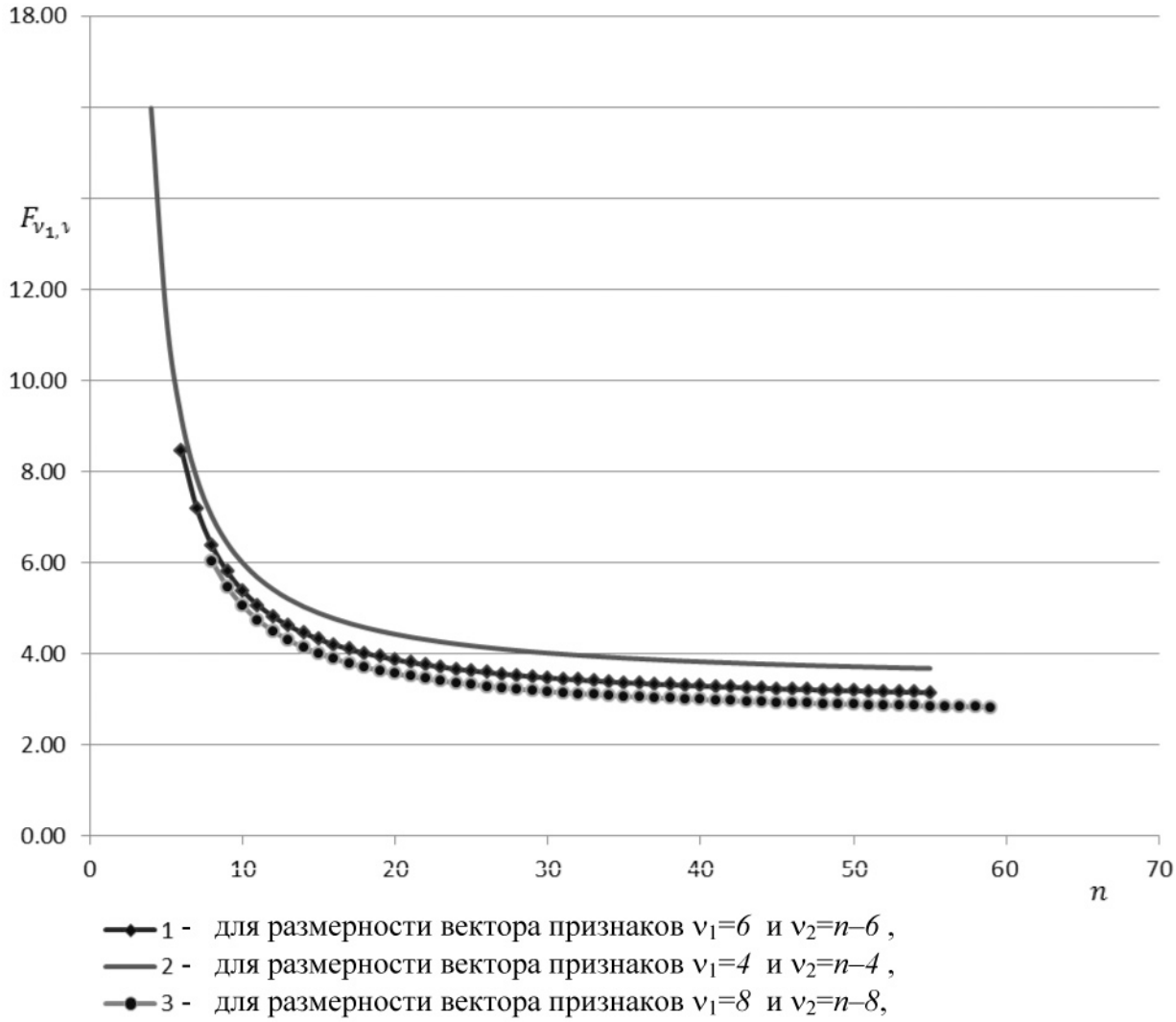


Рисунок 3 – Зависимость распределение $F_{\nu_1,\nu_2}(\alpha)$ от объема выборки *n*

Найдем теперь вероятность попадания вектора признаков $\vec{X} = (x_1, x_2, \cdots, x_L)^T$ в многомерный эллипсоид B_j, определяемый для класса ω_j квадратом расстояния Махаланобиса $D^2_{\mathrm{M}j}$ в соответствии с (22). На основании (5) получим:

$$P(\vec{X} \subset B_j) = \frac{1}{(2\pi)^{L/2}(\det(\boldsymbol{\Sigma}))^{1/2}} \iint_{B_j} \cdots \int e^{-\frac{1}{2} D^2_{\mathrm{M}j}(\vec{X})} dx_1 dx_2 \cdots dx_L. \qquad (27)$$

Поскольку точность определения центра эллипса определяется доверительной вероятностью $q_j=1-\alpha_j$, тогда условная вероятность того, что реализация вектора признаков $\vec{X}$ принадлежит классу ω_j определяется выражением:

$$\tilde{P}(\vec{X} \subset B_j^*) = \frac{q_j \cdot P(\vec{X} \subset B_j)}{\sum_{s=1}^{K} q_s \cdot P(\vec{X} \subset B_s)}. \qquad (28)$$

Из постановки задачи следует, что совместная вероятность независимых событий определяется произведением их вероятностей. Тогда вероятность распознавания класса ω_j определяется принадлежностью $\vec{X}$ к эллипсоиду B_j^* ($\tilde{P}(\vec{X} \subset B_j^*)$) и непринадлежностью к другим эллипсоидам ($1 - \tilde{P}(\vec{X} \subset B_s^*)$):

$$P_{\text{расп}\omega_j} = \tilde{P}(\vec{X} \subset B_j^*) \cdot \prod_{\omega_s, s=1\cdots K, s\neq j} \left(1 - \tilde{P}(\vec{X} \subset B_s^*)\right). \qquad (29)$$

Подчеркнем тот факт, что выражение (29) позволяет оценивать максимальную вероятность распознавания объекта на множестве классов Ω, так как на этапе разработки (при обучении) системы учитываются СКО, обусловленные лишь ошибками квантования (оцифровки) сигнала изображения. В действительности СКО компонент вектора признаков всегда будет больше на реальном сигнале. При увеличении отношения сигнал/шум в соответствии с (11) взаимная информацию между входным и выходным сигналами классификатора снижается. Это эквивалентно расширению эллипсоида рассеивания, то есть росту D_{M_j} и соответствующему повышению уровня значимости α (понижения вероятности распознавания для каждого класса).

Выявим некоторые закономерности относительно количества классов –*K* при фиксированной размерности пространства признаков *L*. Из выражения (29), видно, что увеличение

мощности множества распознаваемых образов (увеличение K) приводит к снижению вероятности распознавания, так как:

$$P_{\text{распω}_j}(K) = \tilde{P}\left(\vec{X} \subset B_j^*\right) \cdot \prod_{\omega_s, s=1\cdots K, s\neq j} \left(1 - \tilde{P}\left(\vec{X} \subset B_s^*\right)\right).$$

$$P_{\text{распω}_j}(K+1) = P_{\text{распω}_j}(K) \cdot \left(1 - \tilde{P}\left(\vec{X} \subset B_{K+1}^*\right)\right). \quad (30)$$

Поскольку $\gamma_s \equiv 1 - \tilde{P}\left(\vec{X} \subset B_s^*\right) \leq 1,\ s = 1\cdots K$, то получаем:

$$P_{\text{распω}_j}(K+1) \leq P_{\text{распω}_j}(K) \cdot \gamma_{K+1}. \quad (31)$$

На рисунке 3 приведена относительное снижение вероятности распознавания $P_{\text{отн}}(\Delta K)$ для случая, когда все $\gamma_s = 0{,}99$:

$$P_{\text{отн}}(\Delta K) \equiv \frac{P_{\text{распω}_j}(K+\Delta K)}{P_{\text{распω}_j}(K)} = \gamma^{\Delta K}$$

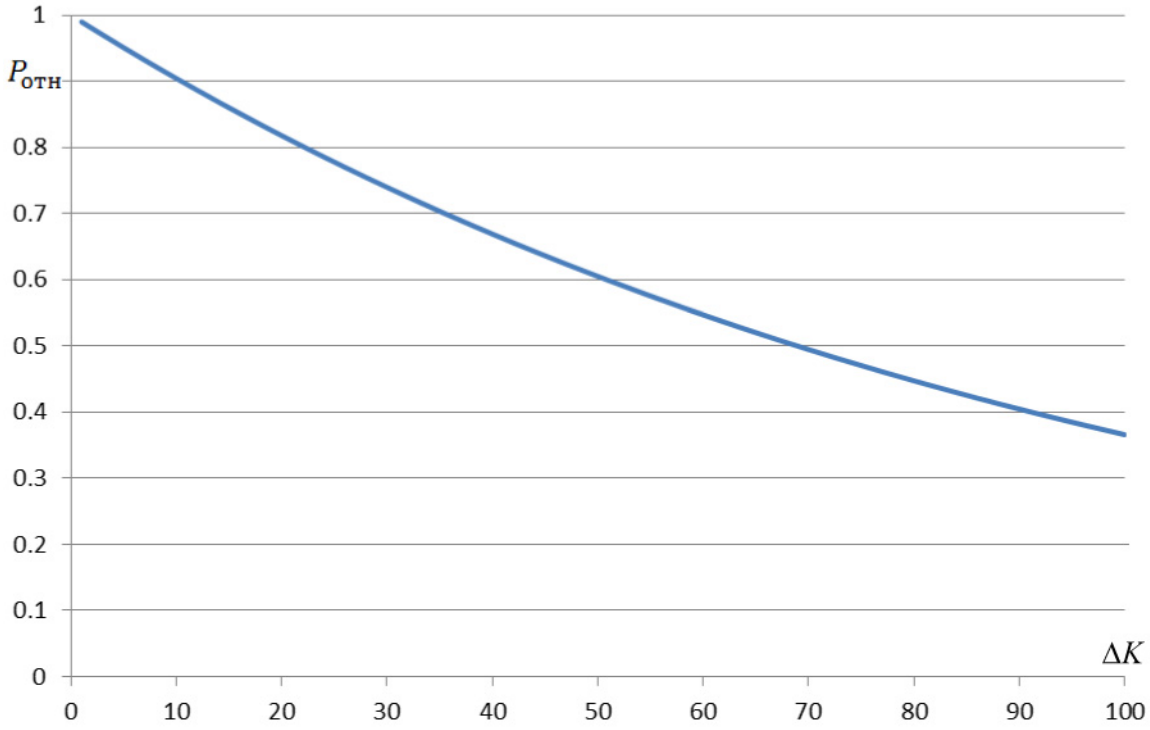


Рисунок 4 – Зависимость $P_{\text{отн}}$ от увеличения числа классов - ΔK

Из рисунка видно, что уже для K=23 классов $P_{\text{отн}} \cong 0{,}8$, поэтому целесообразно для систем технического зрения с большим количеством распознаваемых объектов либо разбивать классификаторы на группы, реализующих одновременную обработку входных образов,

либо увеличить вероятности $\tilde{P}\left(\vec{X} \subset B_s^*\right)$ для всех классов. Например, увеличение $\gamma_s = 0{,}995$ приводит к увеличению классов до K=45 при уровне $P_{\text{отн}} \cong 0{,}8$. Увеличение $\tilde{P}\left(\vec{X} \subset B_s^*\right)$ может быть выполнено, в том числе, за счет повышения информативности вектора признаков, то есть увеличения его размерности –*L*.

Критерий 4-мощности множества классов: увеличение количества распознаваемых объектов классификатором приводит к снижению вероятности распознавания в соответствии с выражением (31).

В заключении сформулируем еще один критерий по выбору размерности набора признаков L_{min}, обеспечивающих решение задачи распознавания с соответствующим ему максимальным уровнем вероятности распознавания.

С одной стороны в соответствии с методом величина $L_{min} = L^I$ определяется максимальным рангом матриц ковариаций $\boldsymbol{\Sigma_j}, j = 1 \ldots K$ то есть

$$L^I = \max_{j=1\ldots K} rank\left(\boldsymbol{\Sigma_j}\right). \qquad (32)$$

Заметим, что выражение (31) скорее является верхней границей для выбора размерности пространства признаков. Как правило, назначая уровень значимости признаков удается существенно понизить размерность этого пространства, например, используя упорядоченный набор собственных значений разложения Карунена-Лоева для ковариационной матрицы класса. Именно в этой связи был введен термин “существенная размерность” и на основании (15), (16) получим значение L^{II}.

С другой стороны, минимальное возможное число признаков описывается степенями свободы физической системы. Например, движение твердого тела в пространстве описывается 6 фазовыми координатами (3- линейных, 3- угловых). Считая, что вдоль каждой фазовой координаты необходимо иметь хотя бы 1 координату вектора признаков, получим 6 компонент. Если движения объекта можно считать плоским, то получим 5 компонент (одно уравнение связи координат движения объекта, например y=0). Обобщая, приходим к выводу: если на движение объекта наложено *m* уравнений связи, то получим *6-m* компонент вектора признаков. Тогда получим $L^0 = 6 - m$ степеней свободы системы, значение L^0 определяет “истинную размерность” системы.

Тогда для размерности пространства признаков можно записать:

$$L^o \le L_{min} \le \{L^I, L^{II}\}. \qquad (33)$$

В этой связи может быть сформулирован ***критерий 5-существенной размерности пространства признаков***: размерностью пространства признаков классификатора определяется выражением (33).

В заключении сформулируем выводы по работе.

1 Рассмотренные выше критерии качества информационных признаков могут быть положены в основу построения классификаторов на самых общих положениях статистической теории оценивания и теории информации и не используют конкретных методов и схем построения классификаторов.

2 Изложенные в статье критерии составляют основу научно-обоснованной методики селекции информативных признаков классификаторов.

3 Критерии отбора информационных признаков сложных пространственных объектов и сигналов могут быть применены для построения оптимального, в том числе, байесовского классификатора.

СПИСОК ЛИТЕРАТУРЫ

1. Hartley, R.V.L. Transmission of Information. - Bell System Technical Journal, July 1928, pp.535–563.
2. Шеннон К. Работы по теории информации.: Пер. с англ. - М.: ИЛ, 1963. 624 с.
3. Shore J.E. and R.W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy: IEEE Transactions on Information Theory, 1980, vol. IT-26, p. 26-37.
4. Jaynes E.T. On the rationale of maximum-entropy methods: Proceedings of the IEEE, 1982, vol. 70, p. 939-952.
5. Фукунага К. Введение в статистическую теорию распознавания образов.: Пер. с англ. - М.: Наука, 1979. - 366 с.
6. Хайкин С. Нейронные сети: полный курс, 2-е издание.: Пер. с англ. - М.: Издательский дом "Вильямс", 2006. -1104 с.
7. Вентцель Е.С. Теория вероятностей. Учебн. для вузов- 5-е изд. стер.-М.: Высш. шк., 1998.- 576 с.
8. Утешев А.Ю., Яшина М.В. Нахождение расстояния от эллипсоида до плоскости и квадрики в Rn.// Доклады АН. 2008. Т.419, №4, с.471-474
9. Гантмахер Ф.Р. Теория матриц. - М.: Наука, 1966. - 576 с.
10. Андерсон Т.В. Введение в многомерный статистический анализ.: Пер. с англ. - М.: Физматгиз, 1963. – 500 с.

electronic scientific and technical periodical
SCIENCE and EDUCATION
EL № FS 77 - 30569. №0421100025. ISSN 1994-0408

Method of selection of classification characteristics in pattern recognition problems of complex spatial objects

77-30569/316296

01, January 2012

Gulevitch S., P., Shevchenko R.A., Pryadkin S.P., Veselov Yu., G.

Bauman Moscow State Technical University
Radio Engineering Institute named after Academician AL Mintz
VUNTS Air Force "Air Force Academy named after professor NE Zhukovskii and Yu Gagarin"
bauman@bmstu.ru
info@rti-mints.ru

The methodology of selecting of classifiers' features was described in this article; these classifiers were used in processing of diffraction image of complex spatial objects. The substantiation of the separation of the object class into subclasses in terms of information theory, as well as reasonable selection of object classes with the usage of ellipsoidal estimation were given. The criteria of qualitative selection of classifier's feature vector, answering questions about the size of teaching selection, the cardinality of the set of classes, the essential dimension of feature vector, were developed. Evidence-based approach to the selection of informative features of classifier of signals, based on the concept of entropy of information theory and ellipsoidal estimation, were proposed.

Publications with keywords: pattern recognition, selection of characteristics, systems of machine vision, information theory, ellipsoidal estimation
Publications with words: pattern recognition, selection of characteristics, systems of machine vision, information theory, ellipsoidal estimation

Reference

1.Hartley, R.V.L., Transmission of Information, Bell System Technical Journal July (1928) 535–563.

2. Shennon K., Work on information theory, Moscow, IL, 1963, 624 p.

3. Shore J.E. and R.W. Johnson, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, IEEE Transactions on Information Theory IT-26 (1980) 26-37.

4. Jaynes E.T., On the rationale of maximum-entropy methods, Proceedings of the IEEE 70 (9) (1982) 939-952.

5. Fukunaga K., Introduction to the statistical theory of pattern recognition, Moscow, Nauka, 1979, 366 p.

6. Khaikin S., Neural networks: a comprehensive course, Moscow, Izdatel'skii dom "Vil'iamc", 2006, 1104 p.

7. Venttsel' E.S., Probability theory, Moscow, Vyssh. shk., 1998, 576 p.

8. Uteshev A.Iu., Iashina M.V., Finding the distance from the ellipsoid to the plane and quadrics in Rn, Doklady AN 419 (4) (2008) 471-474.

9. Gantmakher F.R., Theory of Matrices, Moscow, Nauka, 1966, 576 p.

10. Anderson T.V., An introduction to multivariate statistical analysis, Moscow, Fizmatgiz, 1963, 500 p.