УДК 621.391

И.К. ВАСИЛЬЕВА, А.В. ПОПОВ

Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Украина

ОБ ИНФОРМАТИВНОСТИ КОРРЕЛИРОВАННЫХ ПРИЗНАКОВ ОБЪЕКТОВ РАСПОЗНАВАНИЯ

Выполнено исследование влияния корреляционных связей между компонентами многомерных признаков объектов распознавания на различимость классов. Проведен сравнительный анализ ряда статистических критериев и информационных мер разделяющих качеств коррелированных признаков для случая двумерных нормальных совокупностей. Показано, что наличие корреляции между признаками однозначно не означает их слабой информативности; более того, использование сильно коррелированных признаков в алгоритмах распознавания при определенных условиях может существенно повысить достоверность принимаемых решений. Для понижения размерности вектора измерений предлагается исключать из первичного признакового пространства не максимально коррелированные признаки, а признаки с минимальным значением локальной дивергенции (различимости) классов.

Ключевые слова: *распознавание, корреляционная матрица, собственные значения, собственные вектора, информативность, дивергенция*

Введение

Первым этапом решения задачи распознавания объектов является выбор различительных признаков и определение способов их измерения. Очевидно, что количество признаков, необходимое для достоверного распознавания, зависит от разделяющих качеств (информативности) выбранных признаков. Выбор наиболее информативных признаков позволяет снизить размерность вектора измерений. Если оптимальный выбор признаков осуществляется вне связи с качеством алгоритма классификации, то он определяется максимизацией некоторой функции критерия $J\left(X_1,\ldots,X_g\right)$, обычно понимаемой как расстояние между классами в признаковом пространстве с координатами $X_1,\ldots,X_g$. В других случаях критерий выражает диаметр или объем области, занимаемый классом в признаковом пространстве, и новые признаки формируются путем минимизации критерия. В дискриминантном анализе принят критерий, сущность которого состоит в совместной минимизации внутриклассового разброса наблюдений и максимизации межклассового расстояния [1]. Другой подход связывает выбор признаков с качеством классификации: эффективность выбранных признаков непосредственно выражается в терминах вероятности правильного распознавания.

В известных процедурах, понижающих размерность признакового пространства, обычно предлагается последовательно исключать какой-либо признак из пары наиболее коррелированных [2, 3]. Эта рекомендация основана на доказанном факте, [4] что внутреннее расстояние множества $\{x_{ij}\}_{p\times n}$ в пространстве **X** минимально, если корреляционная матрица **R** p-мерного вектора $\vec{x}$ является диагональной, т.е. компоненты вектора некоррелированы. При этом понятие «внутреннее расстояние» означает максимальный диаметр области, занимаемый классом. Между тем известно, что эффект корреляции вызывает трансформацию совместной плотности распределения компонент случайного вектора (поворот и масштабирование осей эллипса рассеяния). В результате происходит сжатие поверхности многомерной плотности распределения вдоль одной из осей (снижение внутриклассового разброса по направлениям собственных векторов $\vec{\Phi}_i$ матрицы **R**, соответствующих таким собственным значениям, что $\left|\lambda_i\right| < 1$). Как следствие, классы могут становиться линейно разделимыми.

Таким образом, целью работы являлось исследование влияния корреляции признаков на различимость объектов классификации и разработка рекомендаций по уточнению признакового пространства.

1. Модели многомерных признаков

Признаки, несущие информацию об объектах распознавания, обычно имеют различную природу, как, например, амплитуда и фаза сигнала; следовательно, законы распределения одномерных совокупностей могут быть различны. В то же время наблюдаемые данные характеризуют один и тот же объект исследований; т.о., зачастую признаки могут быть зависимыми. В доступных литературных источниках аналитические выражения многомерных законов, включающих компоненты с различными распределениями, отсутствуют. Методология обработки данных при наличии корреляционных связей наиболее развита для многомерного нормального закона распределения и, как правило, в практических задачах распознавания эталонными описаниями объектов служат статистические модели многомерных нормальных совокупностей [4] вида:

$$f(\vec{x}) = (2\pi)^{-p/2} |\mathbf{R}|^{-1/2} \exp\left[-\frac{1}{2}(\vec{x}-\vec{m})^T \mathbf{R}^{-1}(\vec{x}-\vec{m})\right], \quad (1)$$

где $\vec{m}$ – вектор математического ожидания (МО) для класса объектов в пространстве $\mathbf{X}$.

Корреляционная матрица $\mathbf{R}$ – симметричная невырожденная матрица размером p×p, составленная из центральных моментов второго порядка, образованных составляющими случайного вектора:

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 & \cdots & r_{1p}\sigma_1\sigma_p \\ r_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & r_{2p}\sigma_2\sigma_p \\ \cdots & \cdots & \cdots & \cdots \\ r_{1p}\sigma_1\sigma_p & r_{2p}\sigma_2\sigma_p & \cdots & \sigma_p^2 \end{bmatrix}, \quad (2)$$

где σ_i – среднеквадратическое отклонение (СКО); r_{ij} – коэффициент корреляции между i-м и j-м компонентами случайного вектора.

Для некоррелированных компонент $r_{ij} = 0$ и матрица $\mathbf{R} = \mathrm{diag}(\sigma_i^2)$.

Кривые равной вероятности соответствуют тем значениям $\vec{x}$, при которых значение аргумента показательной функции остается постоянным, т.е.

$$(\vec{x}-\vec{m})^T \mathbf{R}^{-1}(\vec{x}-\vec{m}) = \chi^2 . \quad (3)$$

Уравнение (3) описывает эллипсоид с центром в $\vec{m}$.

Ограничимся двумерным случаем; это позволит упростить визуализацию результатов без потери общности рассуждений. Рассмотрим нормально распределенный объект $\mathbf{X}$ с единичной корреляционной матрицей – компоненты случайного вектора в этом случае независимы и одинаково распределены с единичной дисперсией (рис. 1, а).

Преобразование некоррелированного случайного вектора $\vec{x}$ в вектор $\vec{x}^*$ с заданными корреляционной матрицей $\mathbf{R}$ и вектором МО $\vec{m}^*$ осуществляется с помощью уравнения

$$\mathbf{X}^* = \left(\Lambda^{-1/2}\Phi^T\right)^{-1} \mathbf{X} = \Phi\Lambda^{1/2}\mathbf{X},$$

где Φ и $\Lambda = \mathrm{diag}(\lambda_i)$ – соответственно, матрицы собственных векторов и собственных значений $\mathbf{R}$.

Плотность распределения в пространстве $\mathbf{X}^*$ определяется уравнением

$$f(\vec{x}^*) = \sqrt[-p]{2\pi}|\Lambda|^{-1/2}\exp\left[-(\vec{x}^*-\vec{m}^*)^T \Lambda^{-1}(\vec{x}^*-\vec{m}^*)/2\right].$$

Т.о., кривыми равной вероятности являются эллипсы с центрами в точке $\vec{m}^*$ (рис. 1, б).

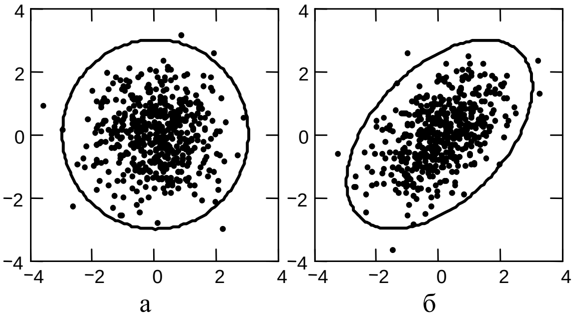


Рис. 1. Эллипсы рассеяния по уровню 0,9973 (3σ) и реализации двумерных нормальных величин с $\vec{m}^T = (0,\ 0)$ и $\vec{\sigma}^T = (1,\ 1)$: а – $r_{12} = 0$, б – $r_{12} = 0{,}5$

Направления главных осей эллипсов рассеяния совпадают с собственными векторами матрицы $\mathbf{R}$, а диаметры пропорциональны квадратному корню от соответствующих собственных чисел или СКО, поскольку $\sqrt{\lambda_k} = \sigma_k$.

При $r_{12} = \pm 1$ двумерное нормальное распределение вырождается в случайный вектор, расположенный на прямой $(x_1 - m_1)/\sigma_1 = \pm(x_2 - m_2)/\sigma_2$.

Вид эталонных описаний классов тесно связан с видом решающего правила. Процедура классификации состоит в сравнении измеренных значений признака $\vec{x}$ (непосредственно либо после функционального преобразования) с порогами – границами между классами. При этом для нормально распределенных совокупностей решающие правила, основанные как на функциях правдоподобия, так и на принципе расстояния, формулируются одинаково:

$$\vec{x} \in a_i \text{, если } d_j - d_i > c,$$

где $d_k = \left(\vec{x}^* - \vec{m}_k^*\right)^T \Lambda_k^{-1}\left(\vec{x}^* - \vec{m}_k^*\right)$ – среднеквадратическое расстояние для класса a_k; c – порог, определяемый принятым критерием стоимости (потерь) решения γ о классе объекта.

На рис. 2 изображены эллипсы рассеяния двумерных коррелированных и некоррелированных признаков, соответствующих двум классам объектов. В обоих случаях (для $r_{12} = 0$ и $r_{12} \neq 0$) были приняты следующие характеристики распределения: $\vec{\sigma}_1^T = \vec{\sigma}_2^T = (1,\ 1)$; $\vec{m}_1^T = (0,\ 2)$, $\vec{m}_2^T = (1,\ 2)$. Очевидно, что при таких условиях разделяющие качества коррелированных признаков будут выше; теоретически, для линейно зависимых признаков можно ожидать абсолютную достоверность распознавания классов, т.е.

$$\lim_{|r_{12}|\to 1} P(\gamma = a_k | a_k) = 1,$$

где $P\left(\gamma = a_k \middle| a_k\right)$ – вероятность правильного распознавания k-го класса.

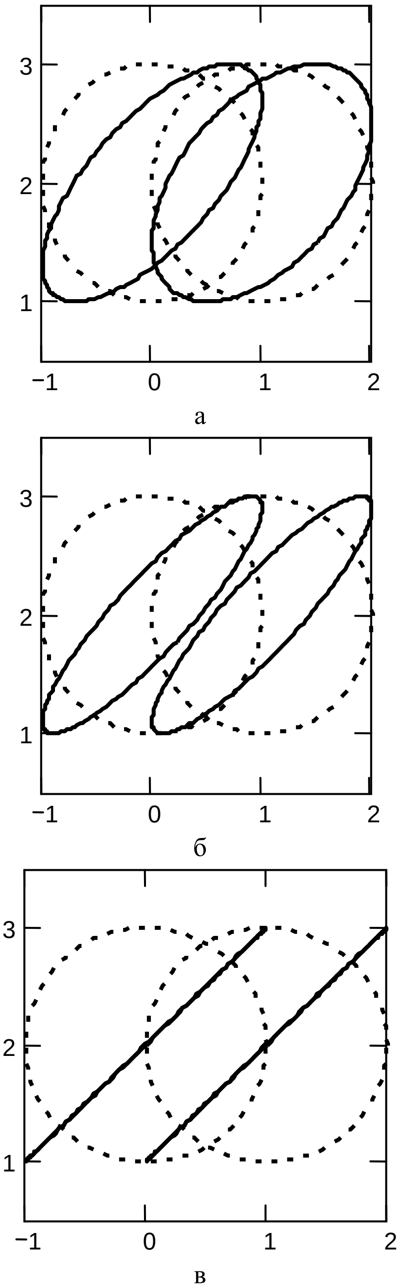


Рис. 2. Эллипсы рассеяния для классов a_1 и a_2 (пунктирная линия – некоррелированные признаки, сплошная – коррелированные признаки): а – $r_{12} \mid a_1 = 0{,}5$, $r_{12} \mid a_2 = 0{,}7$; б – $r_{12} \mid a_1 = r_{12} \mid a_2 = 0{,}9$; в – $r_{12} \mid a_1 = r_{12} \mid a_2 = 1$

2. Критерии информативности

Как отмечалось выше, в статистическом распознавании для сравнения информативности признаков, используются критерии, основанные на метриках элементов из разных классов либо на оценках вероятностей правильного распознавания.

Под «расстоянием внутри класса» понимают наибольшее расстояние между точками, принадлежащими одному классу:

$$\max\left(d\right)_{\{\{\vec{x},\vec{x}'\}\in X\}}, \quad d = \sqrt{\left(\vec{x}-\vec{x}'\right)^T\left(\vec{x}-\vec{x}'\right)} = \left\|\vec{x}-\vec{x}'\right\|.$$

Расстояние между двумя классами определяется наиболее удаленными точками из этих классов.

Функции критерия, основанные на понятии «расстояния», можно получить из матриц рассеяния, используемых в дискриминантном анализе [1].

Матрица разброса между K классами:

$$S_B = \sum_{i=1}^{K} n_i\left(\vec{m}_i - \vec{m}\right)^T\left(\vec{m}_i - \vec{m}\right),$$

где $\vec{m}_i = \sum_{\vec{x}\in X_i} \vec{x} \big/ n_i$ – средний вектор i-го класса, представленного выборкой из n_i отсчетов;

$\vec{m} = \sum_{i=1}^{K} n_i \vec{m}_i \big/ n$ – полный средний вектор.

Матрица разброса внутри класса S_W пропорциональна корреляционно-выборочной матрице:

$$S_W = \sum_{i=1}^{K} S_i,$$

где $S_i = \sum_{\vec{x}\in X_i}\left(\vec{x}-\vec{m}_i\right)^T\left(\vec{x}-\vec{m}_i\right)$.

Декоррелирующее преобразование диагонализирует **R** и тем самым минимизирует разброс внутри класса, т.е. доставляет $\min\left(\max\left(d\right)_{\{\{\vec{x},\vec{x}'\}\in X\}}\right)$. Но означает ли это, что декорреляция признаков однозначно приводит к лучшей различимости классов? Очевидно, нет, ведь, как показано выше (см. комментарий к рис. 2), наличие сильных корреляционных связей в ряде случаев обуславливает лучшую достоверность распознавания, понимаемую как вероятность правильных решений.

Оценка информативности признаков по отношению результатов распознавания объектов контрольной выборки в полном пространстве признаков к результатам распознавания, проводимого без учета исследуемого признака, зависит от решающего правила. В теории статистических решений все виды решающих правил основаны на формировании отношения правдоподобия L и его сравнения с определенным порогом, значение которого определяется выбранным критерием качества [4]:

$$L = \frac{f_p\left(\vec{x}\middle|a_1\right)}{f_p\left(\vec{x}\middle|a_2\right)} \geq c, \tag{4}$$

где $f_p\left(\vec{x}\middle|a_k\right)$ – совместная p-мерная плотность вероятности выборочных значений $\vec{x}$ при условии их принадлежности к классу a_k.

Теоретически, вероятности ошибочных решений количественно выражают объемы таких областей пересечения условных по классу плотностей распределения $f_p\left(\vec{x}\middle|a_k\right)$ (ограниченные плоскостью принятия решений), для которых решение в пользу класса a_k менее вероятно, чем альтернативная гипотеза. Трансформация совместных плотностей распределения вероятности, обусловленная корреляцией компонент векторного признака, несомненно, приводит к изменениям величин вероятностей ошибок; при этом возможно как уменьшение, так и увеличение достоверности распознавания.

Поскольку критерий вероятности ошибки связан с критерием принятия решения, то он больше характеризует саму систему распознавания, чем возможности сигналов и их параметров с точки зрения разделения выборочного пространства на классы объектов. Впрочем, для сравнения информативности различных признаков на этапе уточнения пространства классификатора можно использовать результаты тестовых распознаваний по какой-либо из традиционных процедур. К примеру, это может быть классическое распознавание по критерию максимального правдоподобия [4] с фиксированным объемом контрольных выборок; для обеспечения сопоставимости результатов распознавания объемы выборок должны быть одинаковы для всех исследуемых признаков.

Удобным критерием, который связан с отношением правдоподобия (4), но не предопределяет структуру классификатора, является мера различимости (дивергенции) классов a_k и a_j [5]:

$$J(k,j)=\int_{-\infty}^{\infty}\left(f\left(\vec{x}|a_k\right)-f\left(\vec{x}|a_j\right)\right)\ln\frac{P(a_k|\vec{x})}{P(a_j|\vec{x})}d\vec{x}, \qquad (5)$$

где $P(a_i|\vec{x})$ – апостериорная вероятность i-го класса.

Признаки, которым соответствует большее значение дивергенции, несут больше различающей информации. Однако (5) применимо лишь в случае двух классов объектов. В работе [6] получено выражение для меры различимости класса a_k на фоне остальных классов для случая, когда K > 2:

$$J(k)=\int_{\Omega}\left(f\left(\vec{x}|a_k\right)-f\left(\vec{x}|\tilde{a}_k\right)\right)\ln\frac{f\left(\vec{x}|a_k\right)}{f\left(\vec{x}|\tilde{a}_k\right)}d\vec{x}, \qquad (6)$$

где Ω – выборочное пространство; $f\left(\vec{x}|\tilde{a}_k\right)$ – смесь плотностей вероятностей для всех классов, за исключением a_k:

$$f\left(\vec{x}|\tilde{a}_k\right)=\frac{1}{1-P\left(a_k\right)}\sum_{i\neq k=1}^{K}P\left(a_i\right)f\left(\vec{x}|a_i\right),$$

где $P(a_k)$ – априорная вероятность класса a_k.

Признак является достаточно информативным для распознавания класса a_k, если $J(k)\geq\Theta_k$, где Θ_k – минимально достаточная дивергенция:

$$\Theta_k=\left[\frac{\left[1-P\left(a_k\right)\right]^2}{P\left(a_k\right)^2}-1\right]\ln\frac{\left[1-P\left(a_k\right)\right]^2}{P\left(a_k\right)^2}. \qquad (7)$$

Среднее значение J(k) представляет собой меру различимости множества классов на совокупности параметров $\{\vec{\xi}\}$ (или обобщенную дивергенцию)

$$J\left(\{\vec{\xi}\}\right)=\sum_{k=1}^{K}P\left(a_k\right)J(k) \qquad (8)$$

и может использоваться для сравнения информативности различных подмножеств совокупности параметров $\{\vec{\xi}\}$ [6, 7].

3. Влияние корреляции на различимость

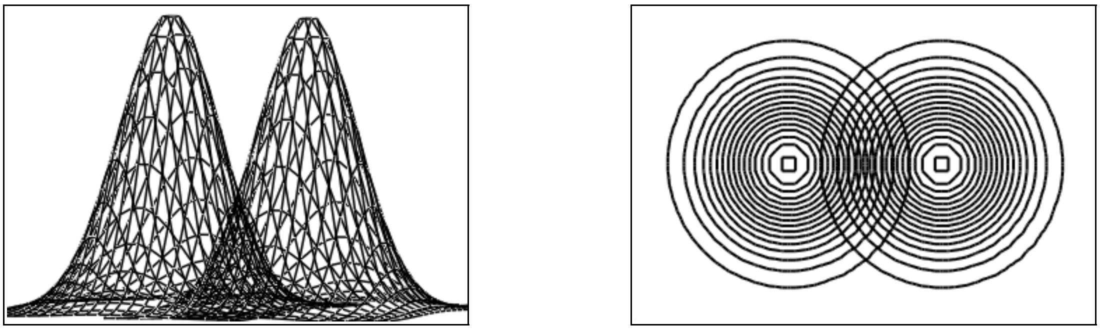
Для исследования влияния степени корреляции признаков на дивергенцию рассмотрим два класса объектов, эталонными описаниями которых являются двумерные нормальные распределения с единичными дисперсиями и МО $\vec{m}_1$, $\vec{m}_2$.

Будем считать, что $\vec{m}_1^T$ = (0, 0) – это условие обеспечивается всегда путем переноса начала отсчета системы координат в точку $\vec{m}_1\left(m_1^{(1)}, m_2^{(1)}\right)$; тогда $\vec{m}_2^T=\left(m_1^{(2)}-m_1^{(1)}, m_2^{(2)}-m_2^{(1)}\right)$, где $m_i^{(k)}$ – МО i-й компоненты векторного признака k-го класса. Примем для определенности $\vec{m}_2^T$ = (3, 0).

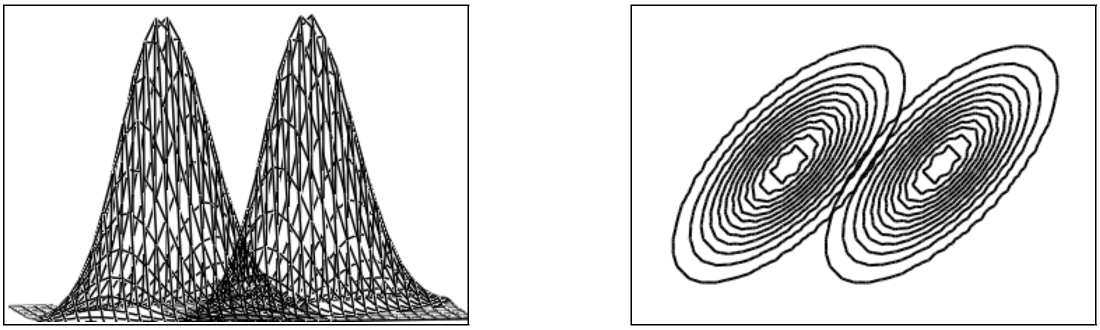
Если компоненты признаков некоррелированы (рис. 3, а), то дивергенция (5) J(1, 2) = 9,0.

Для компонент с коэффициентом корреляции r_{12} = 0,7 (рис. 3, б) в результате изменения расстояний внутри классов (а именно: уменьшения разброса вдоль направлений $\vec{\Phi}_2^{(k)}$ в $1\big/\sqrt{\lambda_2^{(k)}}\approx$ 1,83 раза) уменьшаются объемы областей пересечения условных по классам плотностей распределения, отвечающих за принятие неправильных решений. Следовательно, информативность коррелированных признаков должна быть выше. Об этом же свидетельствует и более высокое значение дивергенции J(1, 2) = 17,647.

Вместе с тем, эффект корреляции может оказывать и негативное влияние на различимость классов. Очевидно, что информативность коррелированных признаков будет ниже, чем информативность некоррелированных в тех случаях, когда коэффициенты корреляции $r_{12}\mid a_1$ и $r_{12}\mid a_2$ противоположны по знаку, а также, если вектор $\vec{m}_1-\vec{m}_2$ имеет значительную составляющую вдоль собственных векторов, соответствующих наибольшим собственным числам корреляционных матриц признаков классов a_1 и a_2. Так, на рис. 4 показаны плотности распределения и эллипсы рассеяния двух нормальных двумерных совокупностей с $\vec{m}_1^T$ = (0, 0), $\vec{m}_2^T$ = (2, 2) и единичными дисперсиями компонент векторов. В данном случае различимость классов по критерию дивергенции лучше, если признаки некоррелированы: при r_{12} = 0,7 J(1, 2) = 4,71; при r_{12} = 0 J(1, 2) = 8,0.

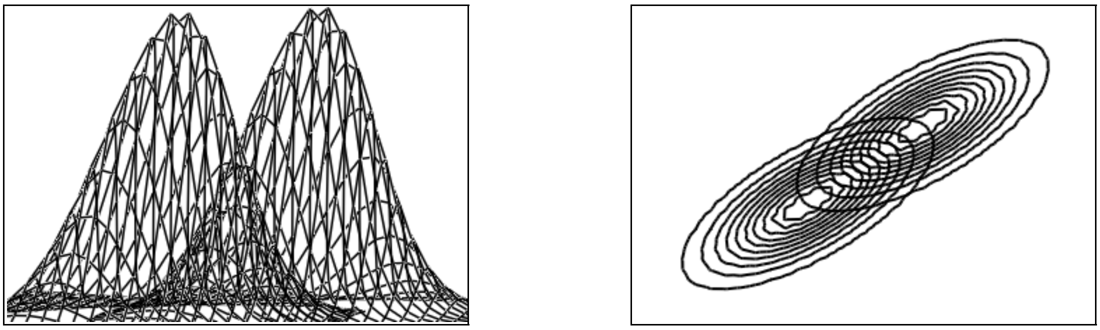


а

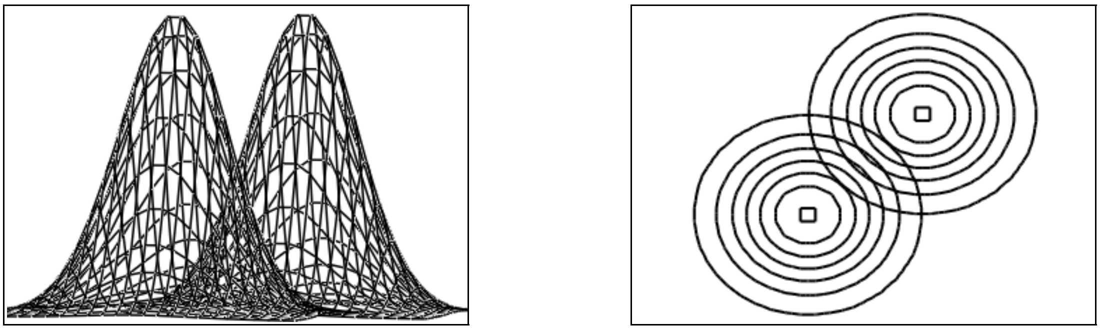


б

Рис. 3. Двумерные плотности нормального распределения и эллипсы рассеяния:
а – $r_{12} \mid a_1 = r_{12} \mid a_2 = 0$; б – $r_{12} \mid a_1 = r_{12} \mid a_2 = 0{,}7$



а



б

Рис. 4. Двумерные плотности нормального распределения и эллипсы рассеяния:
а – $r_{12} \mid a_1 = r_{12} \mid a_2 = 0{,}7$; б – $r_{12} \mid a_1 = r_{12} \mid a_2 = 0$

Заключение

На основании результатов анализа информативности двумерных признаков можно заключить, что наличие сильной корреляционной зависимости между парами элементов первичного признакового пространства не является само по себе достаточным основанием для исключения из алгоритма классификации какого-либо признака из пары наиболее коррелированных. Более обосновано, на наш взгляд, использовать в процедурах, понижающих размерность признакового пространства классификатора, критерии информативности, связанные с понятием достоверности распознавания, но не ограничивающие свободу выбора решающего правила, например, показатели дивергенции. Т.о., основные этапы базового алгоритма формирования уточненного признакового пространства будут состоять в выделении подмножеств признаков, построении их статистических моделей, оценке величин дивергенций классов по исследуемым признакам и исключении наименее информативных из признакового пространства.

Литература

1. Nenadic Z. Information discriminant analysis: Feature extraction with an information-theoretic objective / Z. Nenadic // IEEE Trans PAMI. – 2007. – Vol. 29. – P. 1394-1407.

2. Sanguinetti G. Dimensionality reduction of clustered data sets / G. Sanguinetti // IEEE Trans PAMI. – 2008. – Vol. 30. – P. 535-540.

3. Hua-Liang Wei. Feature subset selection and ranking for data dimensionality reduction / Hua-Liang Wei, S.A. Billings // IEEE Trans PAMI. – 2007. – Vol. 29. – P. 162-166.

4. Фукунага К. Ведение в статистическую теорию распознавания образов: пер. с англ. / К. Фукунага. – М.: Наука, 1979. – 367 с.

5. Кульбак С. Теория информации и статистика / С. Кульбак. – М.: Наука, 1968. – 302 с.

6. Попов А.В. Критерий информативности параметров сигналов для радиолокационного распознавания объектов / А.В. Попов // Авиационно-космическая техника и технология: сб. научн. тр. Нац. аэрокосм. ун-та им. Н.Е. Жуковского «ХАИ». – Х.: ХАИ, 1999. – Вып. 12. – С. 44-46.

7. Васильева И.К. Анализ информативности поляризационных признаков гидрометеорологических объектов / И.К. Васильева // Радіоелектронні і комп'ютерні системи. – 2004. – № 4 (8). – С. 126-130.

Поступила в редакцию 17.09.2008

Рецензент: д-р техн. наук, проф., директор центра Г.Я. Красовский, ГНПЦ «Природа», Харьков.

ПРО ІНФОРМАТИВНІСТЬ КОРЕЛЬОВАНИХ ОЗНАК ОБ'ЄКТІВ РОЗПІЗНАВАННЯ

І.К. Васильєва, А.В. Попов

Виконано дослідження впливу кореляційних зв'язків між компонентами багатовимірних ознак об'єктів розпізнавання на розрізненість класів. Проведено порівняльний аналіз ряду статистичних критеріїв і інформаційних мір класифікаційних якостей корельованих ознак для випадку двовимірних нормальних сукупностей. Показано, що наявність кореляції між ознаками однозначно не визначає їх слабку інформативність; більш того, використання сильно корельованих ознак в алгоритмах розпізнавання за певних умов може істотно підвищити вірогідність прийнятих рішень. Для зниження розмірності вектора вимірів пропонується виключати з первинного ознакового простору не максимально корельовані ознаки, а такі, що відповідають мінімальним значенням локальної дивергенції (розрізнення) класів.

Ключові слова: розпізнавання, кореляційна матриця, власні значення, власні вектора, інформативність, дивергенція.

ON SELF-DESCRIPTIVENESS OF CORRELATED SIGNATURES OF RECOGNITION OBJECTS

I.K. Vasilyeva, A.V. Popov

The research of correlation among components of recognition objects' multidimensional signatures effects on classes' discernibleness was performed. The comparative analysis of a number of statistical criterions and informational measures of correlated signatures separating properties for a case of bivariate normally distributed data was carry out. Is rotined, that if correlations among signatures are available doesn't signify their low self-descriptiveness; moreover, application of the hardly correlated signatures in classification algorithms under certain conditions can essential improve a veracity of the received solutions. In order to reducing of observations' vector length exclude from primary signs' space not the same signatures such as at the most correlated, but those local divergence (classes' discernibleness) are minimum value was proposed.

Key words: recognition, correlation matrix, eigenvalues, eigenvectors, self-descriptiveness, divergence.

Васильева Ирина Карловна – канд. техн. наук, доцент кафедры производства радиоэлектронных систем летательных аппаратов Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ», Харьков, Украина.

Попов Анатолий Владиславович – канд. техн. наук, доцент, доцент кафедры производства радиоэлектронных систем летательных аппаратов Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: off@xai.edu.ua.