

УДК 621.391

И.К. ВАСИЛЬЕВА, Е.А. ПАНКРАТОВА

*Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Украина***ВЛИЯНИЕ СТЕПЕНИ КОРРЕЛЯЦИИ ПРИЗНАКОВ НА РЕЗУЛЬТАТЫ
РАСПОЗНАВАНИЯ ОБЪЕКТОВ ПО ДАННЫМ МОДЕЛИРОВАНИЯ
ДВУМЕРНЫХ НОРМАЛЬНЫХ СОВОКУПНОСТЕЙ**

Выполнено исследование зависимости статистической оценки вероятности правильной классификации от значения коэффициента корреляции между компонентами двумерных признаков объектов. Контрольные выборки признаков двух классов объектов были получены путем моделирования двумерных коррелированных случайных величин, распределенных по нормальному закону с различными параметрами. Для распознавания классов была выбрана непоследовательная одноступенчатая процедура принятия решения по критерию максимального правдоподобия. Эмпирическая вероятность правильного распознавания класса по исследуемому признаку интерпретировалась как мера информативности данного признака. Показано, что использование сильно коррелированных признаков может существенно повысить достоверность классификации слабо различимых объектов.

Ключевые слова: *распознавание, корреляция, информативность, различимость классов, максимальное правдоподобие, вероятность правильного решения.*

Введение

В основе методов распознавания образов лежит описание классов объектов через определенные значения значимых признаков. Каждому объекту ставится в соответствие матрица $r \times n$ признаков X , по которой происходит его распознавание (x_{ij} – j -е измерение i -го признака; $j = 1, 2, \dots, n$, $i = 1, 2, \dots, r$). Таким образом, X можно представить как некоторое пространство размерности $N = r \times n$ с определённой в этом пространстве метрикой. Любой объект (реализация) представляется в виде точки (вектора) в этом пространстве. Проекция этой точки на i -ю ось координат соответствует значению i -го признака. Очевидно, что количество признаков, необходимое для достоверного распознавания, зависит от разделяющих качеств выбранных признаков. Наилучшими являются те признаки, у которых разность условных по классу математических ожиданий m (МО) велика по сравнению со среднеквадратическими отклонениями σ (СКО). Для количественной оценки разделяющих качеств (информативности) признаков используются критерии, основанные на метриках элементов из разных классов либо на оценках вероятностей правильного распознавания [1 – 3]. Функции критерия, основанные на понятии «расстояния», можно получить из матриц рассеяния, используемых в дискриминантном анализе [1, 4, 5] и характеризующих разброс внутри и между классами.

Размерность признакового пространства обычно стремятся сделать как можно меньше, поскольку

при этом сокращается количество требуемых измерений, упрощаются вычисления, формирующие и реализующие решающие правила, повышается статистическая устойчивость результатов распознавания. Вместе с тем, уменьшение N ведёт к росту ошибок классификации. Поэтому формирование признакового пространства является компромиссной задачей, которую можно разделить на две части: формирование исходного признакового пространства и минимизация размерности этого пространства. В части, касающейся минимизации размерности, существуют формальные методы и алгоритмы, позволяющие количественно оценить информативность каждого из исследуемых признаков и затем сформировать новый набор признаков, образующих уточненное признаковое пространство с меньшей размерностью. Большинство этих методов предполагает функциональное отображение N -мерного вектора измерений в пространство меньшей размерности с учетом дисперсий либо корреляций между признаками [1, 4 – 8]. При этом процедура понижения размерности понимается как задача удаления или объединения (группировки) признаков с сильной корреляцией. Однако, как показано в [2], наличие корреляции между признаками еще не означает их слабой информативности; более того, использование сильно коррелированных признаков в алгоритмах распознавания при определенных условиях может существенно повысить достоверность принимаемых решений. Поэтому целью данной работы являлась экспериментальная оценка информативно-

сти коррелированных и некоррелированных признаков и исследование влияния степени взаимной корреляции компонент признаков на оценку вероятности правильного распознавания классов объектов.

1. Моделирование признаков

В качестве объектов исследования рассматривались два класса (a_1 и a_2), для эталонных описаний которых были приняты двумерные нормальные плотности распределения N_2 вида:

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

где \bar{m} – вектор МО для каждого класса объектов;

\mathbf{R} – корреляционная матрица (КМ).

Признаки классов отличались параметрами распределения (1). Для моделирования двух некоррелированных компонент вектора \bar{x} использовался алгоритм, основанный на центральной предельной теореме. Имитация отсчетов признака k -го класса \bar{x}_k^* осуществлялась путем преобразования \bar{x} [9]:

$$\bar{x}_k^* = \Phi \Lambda^{1/2} \bar{x} + \bar{m}_k^*,$$

где Φ – матрица собственных векторов \mathbf{R} ;

$\Lambda = \text{diag}(\lambda_i)$ – диагональная матрица собственных значений \mathbf{R} .

Исследовались следующие случаи:

1. Пара признаков \bar{x}^1 и \bar{x}^2 , характеризующихся одинаковым расстоянием между МО классов, при этом ориентация векторов $\bar{m}_1^i - \bar{m}_2^i$ не совпадает с направлениями собственных векторов \mathbf{R} (рис. 1, а):

$$f(\bar{x}^1 | a_1) = N_2(\bar{m}_1^1 = (0; 0); \bar{\sigma}_1^1 = (1; 1,5); r = \text{var}),$$

$$f(\bar{x}^1 | a_2) = N_2(\bar{m}_2^1 = (1; 0); \bar{\sigma}_2^1 = (1; 2); r = \text{var}),$$

$$f(\bar{x}^2 | a_1) = N_2(\bar{m}_1^2 = (0; 0); \bar{\sigma}_1^2 = (2,5; 1,5); r = \text{var}),$$

$$f(\bar{x}^2 | a_2) = N_2(\bar{m}_2^2 = (1; 0); \bar{\sigma}_2^2 = (2; 1,5); r = \text{var}).$$

2. Признаки \bar{x}^3 и \bar{x}^4 , векторы $\bar{m}_1^i - \bar{m}_2^i$ ($i = 3, 4$) которых имеют заметную составляющую вдоль собственных векторов, соответствующих наибольшему собственным числам КМ для a_1 и a_2 (рис. 1, б):

$$f(\bar{x}^3 | a_1) = N_2(\bar{m}_1^3 = (0; 0); \bar{\sigma}_1^3 = (3; 1,5); r = \text{var}),$$

$$f(\bar{x}^3 | a_2) = N_2(\bar{m}_2^3 = (1; 1); \bar{\sigma}_2^3 = (1; 2,5); r = \text{var}),$$

$$f(\bar{x}^4 | a_1) = N_2(\bar{m}_1^4 = (0; 0); \bar{\sigma}_1^4 = (2; 2); r = \text{var}),$$

$$f(\bar{x}^4 | a_2) = N_2(\bar{m}_2^4 = (1; 1); \bar{\sigma}_2^4 = (2; 2); r = \text{var}).$$

3. Пара признаков \bar{x}^5 и \bar{x}^6 с различной ориентацией векторов $\bar{m}_1^i - \bar{m}_2^i$ относительно направлений

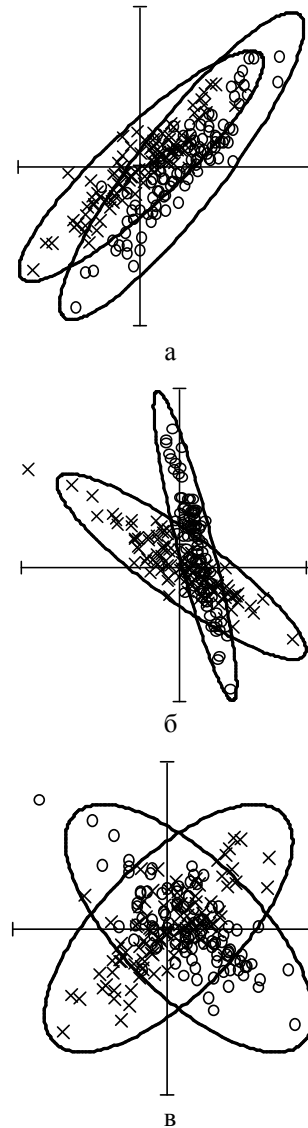


Рис. 1. Эллипсы рассеяния, соответствующие условным по классам плотностям распределения вероятности исследуемого двумерного признака, и результаты моделирования (выборочные данные, принадлежащие классу a_1 , помечены «x», a_2 – «o»):

а – признак \bar{x}^1 при $r | a_1 = r | a_2 = 0,9$;

б – признак \bar{x}^3 при $r | a_1 = r | a_2 = -0,9$;

в – признак \bar{x}^5 при $r | a_1 = 0,7, r | a_2 = -0,7$

собственных векторов \mathbf{R} , причем коэффициенты корреляции r между компонентами признаков классов a_1 и a_2 отличаются по знаку (рис. 1, в):

$$f(\bar{x}^5 | a_1) = N_2(\bar{m}_1^5 = (0; 0); \bar{\sigma}_1^5 = (2; 2); r = \text{var}),$$

$$f(\bar{x}^5 | a_2) = N_2(\bar{m}_2^5 = (1; 0); \bar{\sigma}_2^5 = (2; 2); r = -\text{var}),$$

$$f(\bar{x}^6 | a_1) = f(\bar{x}^5 | a_1),$$

$$f(\bar{x}^6 | a_2) = N_2(\bar{m}_2^6 = (1; 1); \bar{\sigma}_2^6 = (2; 2); r = -\text{var}).$$

4. Признаки \bar{x}^6 , \bar{x}^7 и \bar{x}^8 , которые отличаются друг от друга величиной разности МО классов:

$$\bar{\sigma}_k^6 = \bar{\sigma}_k^7 = \bar{\sigma}_k^8, \quad \bar{m}_1^6 = \bar{m}_1^7 = \bar{m}_1^8 = (0; 0),$$

$$\bar{m}_2^6 = (1; 1), \quad \bar{m}_2^7 = (2; 2), \quad \bar{m}_2^8 = (4; 4).$$

2. Результаты распознавания

Распознавание классов объектов проводилось по классической одноступенчатой процедуре по критерию максимального правдоподобия:

$$a = \begin{cases} a_1, & \text{если } L_N \geq 1; \\ a_2, & \text{если } L_N < 1, \end{cases} \quad (2)$$

где L_N – отношение правдоподобия:

$$L_N = f_N(\bar{x}|a_1)/f_N(\bar{x}|a_2),$$

где $f_N(\bar{x}|a_k)$ – совместная N -мерная плотность вероятности выборочных значений \bar{x} при условии их принадлежности к k -му классу.

Критерий (2) не требует знания априорных вероятностей классов $P(a_k)$ и функции потерь, позволяет оценивать достоверность решений и легко обобщается на случай многих классов.

Эмпирические вероятности правильного распознавания k -го класса P_{11} и P_{22} оценивались по отношению количества правильных решений к общему числу результатов распознавания контрольной выборки. Решения принимались по одному отсчету; объем испытаний $M = 500$. Оценка вероятности ошибки при условии равных $P(a_k)$ находилась как

$$P_{\text{ош}} = 1 - 0,5(P_{11} + P_{22}). \quad (3)$$

Графики зависимостей $P_{\text{ош}}$ (3) от величины коэффициента корреляции γ , $\gamma \in [-1; 1]$ показаны на рис. 2.

Заключение

Результаты оценки достоверности классификации объектов по двумерным нормальным совокупностям показывают, что наличие корреляционной зависимости между компонентами признаков не ухудшает их информативности. Увеличение степени взаимной корреляции компонент приводит к лучшим результатам распознавания, по крайней мере, в тех случаях, когда разность между МО классов составляет величины порядка $(0,3 \dots 3) \times \max\{\sqrt{\lambda_i}\}$, где λ_i – собственные числа корреляционных матриц признаков. Если вектор $\bar{m}_1 - \bar{m}_2$ имеет большую составляющую вдоль собственных векторов КМ, соответствующих наибольшему собственным числам, то существенное снижение вероятности ошибочного распознавания достигается только при отрицательных значениях коэффициентов корреляции признаков. Если ориентация вектора $\bar{m}_1 - \bar{m}_2$ не совпадает с направлениями собственных векторов \mathbf{R} , то с увеличением абсолютной величины коэффициента корреляции вероятность ошибки снижается на 2...4

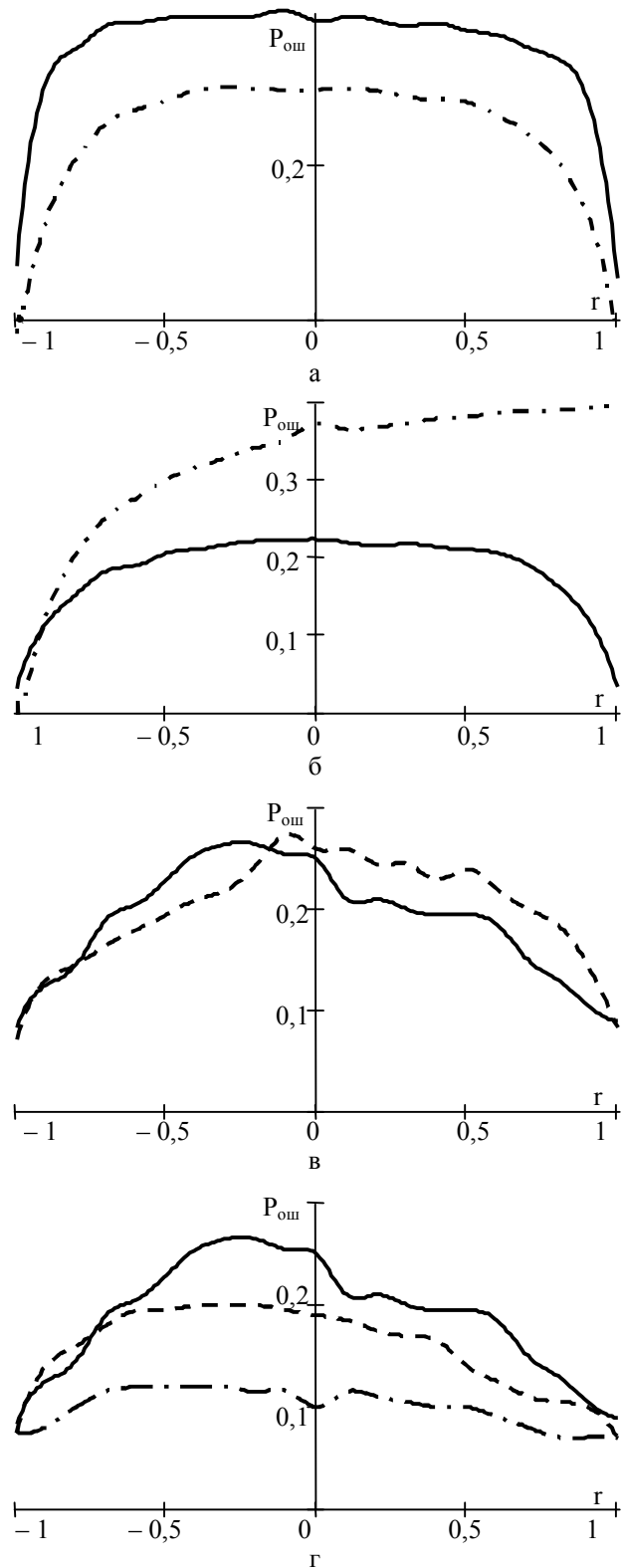


Рис. 2. Зависимости вероятности $P_{\text{ош}}$ ошибочного распознавания от значения коэффициента взаимной корреляции γ компонент двумерных признаков:
 а – \bar{x}^1 (штрихпунктирная линия) и \bar{x}^2 (сплошная);
 б – \bar{x}^3 (сплошная линия) и \bar{x}^4 (штрихпунктирная);
 в – \bar{x}^5 (штриховая линия) и \bar{x}^6 (сплошная);
 г – \bar{x}^6 (сплошная линия), \bar{x}^7 (штриховая)
 и \bar{x}^8 (штрихпунктирная линия)

порядка. Таким образом, использование сильно коррелированных признаков ($r \geq 0,9$) для классификации слабо различимых объектов может повысить достоверность принятого решения.

Литература

1. Дуда Р. Распознавание образов и анализ сцен: пер. с англ. / Р. Дуда, П. Харт. – М.: Мир, 1976. – 507 с.
2. Васильева И.К. Об информативности коррелированных признаков объектов распознавания / И.К. Васильева, А.В. Попов // *Радіоелектронні і комп'ютерні системи*. – 2008. – № 3 (30). – С. 56-61.
3. Jie Yu. Distance learning for similarity estimation / Yu Jie, J. Amores, N. Sebe, P. Radeva, Qi Tian // *IEEE Trans PAMI*. – 2008. – Vol. 30. – P. 451-462.
4. Hamsici O.C. Bayes optimality in linear discriminant analysis / O.C. Hamsici, A.M. Martinez //

IEEE Trans PAMI. – 2008. – Vol. 30. – P. 647-657.

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

6. Tae-Kyun Kim. Discriminative learning and recognition of image set classes using canonical correlations / Kim Tae-Kyun, J. Kittler, R. Cipolla // *IEEE Trans PAMI*. – 2007. – Vol. 29. – P. 1005-1018.

7. 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.

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

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

Поступила в редакцию 3.02.2009

Рецензент: д-р техн. наук, проф., проф. каф. 502 Г.Я. Красовский, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина.

ВПЛИВ СТУПЕНЯ КОРЕЛЯЦІЇ ОЗНАК НА РЕЗУЛЬТАТИ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ ЗА ДАНИМИ МОДЕЛЮВАННЯ ДВОВИМІРНИХ НОРМАЛЬНИХ СУКУПНОСТЕЙ

І.К. Васильєва, Є.О. Панкратова

Виконано дослідження залежності статистичної оцінки імовірності правильної класифікації від значення коефіцієнта кореляції між компонентами двовимірних ознак об'єктів. Контрольні вибірки ознак двох класів об'єктів були отримані шляхом моделювання двовимірних корельованих випадкових величин, розподілених за нормальним законом із різними параметрами. Для розпізнавання класів була обрана непослідовна одноступінчата процедура ухвалення рішення за критерієм максимальної правдоподібності. Емпірична імовірність правильного розпізнавання класу за досліджуваною ознакою інтерпретувалася як міра інформативності даної ознаки. Показано, що використання сильно корельованих ознак може істотно підвищити вірогідність класифікації об'єктів, що слабо розрізняються.

Ключові слова: розпізнавання, кореляція, інформативність, розрізнення класів, максимальна правдоподібність, імовірність правильного рішення.

EFFECT OF SIGNATURES' CORRELATION POWER ON RESULTS OF OBJECTS' RECOGNITION ACCORDING TO THE DATA OF MODELING OF TWO-DIMENSIONAL NORMAL SAMPLES

I.K. Vasilyeva, E.A. Pankratova

The research of a statistical estimation of exact classification probability relation from value of a correlation coefficient between components of two-dimensional signatures of objects was made. The control samples of signatures two classes of objects were obtained by modeling of the two-dimensional correlated random variables, distributed under the normal law with different parameters. For classes identification the nonsequential single-stage decision-making procedure by criterion of maximum likelihood was decided. The empirical probability of exact class identification over investigated signature was interpreted as a measure of self-descriptiveness of the given sign. It is shown, usage of the hardly correlated signatures can essentially increase veracity of weakly distinguishable objects' classification.

Key words: recognition, correlation, self-descriptiveness, distinguishability of classes, maximum likelihood, probability of sound decision.

Васильєва Ирина Карловна – канд. техн. наук, доцент кафедри производства радиоэлектронных систем летательных аппаратов, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина.

Панкратова Евгения Александровна – студентка кафедри производства радиоэлектронных систем летательных аппаратов факультета радиоэлектронных систем летательных аппаратов, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина.