

СОКРАЩЕНИЕ ПРОСТРАНСТВА ПРИЗНАКОВ ПО КРИТЕРИЮ СОПРЯЖЕННОСТИ С НУЛЬ-ПРОСТРАНСТВОМ

В.А.Шустов

Институт систем обработки изображений РАН
Самарский государственный аэрокосмический университет

Аннотация

Часто при формировании признакового пространства намеренно задают избыточное число признаков, а затем, исключают малоинформативные. Этот прием известен как задача редукции признакового пространства [1]. При этом возникает проблема выбора показателя информативности и порогового значения выбранного показателя.

В работах [2, 3] исследовалась возможность использования в качестве критерия информативности признака показателя сопряженности соответствующего этому признаку вектора с нуль-пространством транспонированной матрицы, составленной из оставшихся векторов-признаков. Эта методика предполагает задание некоторого порога для показателя сопряженности и исключение признаков, у которых он меньше заданной. Однако вопрос обоснованного выбора порога для получения оптимального набора признаков из исходного множества остается открытым.

В настоящей работе предлагается и исследуется способ задания количественного значения порога на основе вычисления критерия сопряженности составляющей признакового пространства, которая вводится для учета смещения разделяющей гиперплоскости.

Исходная система признаков

Рассматривается задача распознавания рукописных цифр, написанных разным почерком. Различные изображения цифр представляют собой растр 12×16 черно-белого полутонового 256-цветного изображения. Предполагается, что цифры распознаются попарно (случай распознавания двух классов).

Система признаков формируется следующим образом. Поле цифры разбивается на квадраты размера 2×2 пиксела. Суммируется величина яркости, приходящаяся на пиксели внутри каждого из этих квадратов. Полученные значения после нормирования используются в качестве признака изображения.

На рис. 1 приводится пример преобразования изображения цифры «3». Слева приведено исходное изображение, справа – преобразованное изображение. Значения интенсивности отсчетов преобразованного изображения используются в качестве признаков. Такое преобразование равносильно масштабированию размеров изображения с уменьшением в 2 раза и распознаванию уменьшенных изображений. В рассматриваемом случае количество признаков равно 48.

Для выбранной системы признаков существенное значение имеют размеры и число квадратов на поле цифры. При разбиении на большое число квадратов малых размеров может оказаться, что часть из них на поле области, ограничивающей всю совокупность цифр, окажутся малоинформативными. Ясно, что на этапе обучения исключение соответствующих признаков может быть произведено без особого ущерба для точности классификации. При этом за счет сокращения размерности признакового пространства уменьшается её вычислительная сложность. На эффективность процедуры исключения признаков это не должно оказывать влияние, так как никаких ограничений на характер признакового про-

странства не налагается, за исключением того, что количество обучающих примеров должно быть больше количества признаков (переопределенность обучающей матрицы). Исключение таких признаков должно происходить автоматически в процессе формирования признакового пространства.

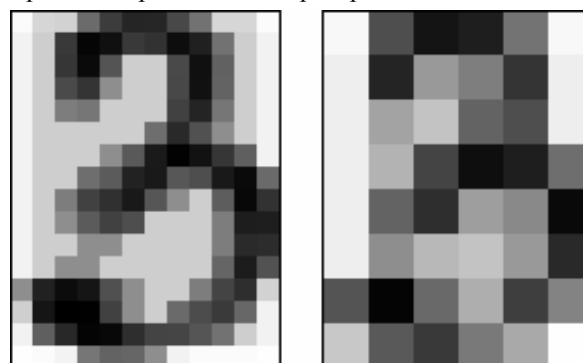


Рис. 1. Формирование признаков по изображениям цифр (слева – исходное изображение; справа – преобразованное для распознавания)

Метод исключения признаков

Рассмотрим случай распознавания двух классов. Пусть обобщенная линейная разделяющая функция имеет вид

$$g(X) = a^T y(X) = w_0 + \sum w_i x_i, \quad i = \overline{1, m}, \quad (1)$$

где $\mathbf{a} = [w_0, \mathbf{w}]^T = [w_0, w_1, \dots, w_m]$ – $M \times 1$ - вектор искомых параметров (весовых коэффициентов [4]), \mathbf{X} – вектор, компоненты которого являются признаками, а $y(\mathbf{X})$ – $M \times 1$ - вектор обобщенных признаков ($M = m + 1$), который, в частности, может быть задан в виде $y(\mathbf{X}) = [1, \mathbf{X}]^T$. Для N обучающих объектов ($N > M$) имеем $N \times M$ -матрицу \mathbf{Y} , строки которой $\mathbf{y}_i^T, i = \overline{1, N}$,

а столбцы (векторы признаков) далее обозначаются как \mathbf{Y}_j , $j = \overline{1, M}$.

Для принятия решения об отборе признаков для классификации используется критерий сопряженности, вычисляемый для каждого вектора \mathbf{Y}_i , определяемый как [1]:

$$S_i = \left(\mathbf{Y}_i^T \mathbf{T}_0 \mathbf{T}_0^T \mathbf{Y}_i \right)^{1/2} / \left(\mathbf{Y}_i \mathbf{Y}_i^T \right)^{1/2}, \quad (2)$$

где $\mathbf{T}_0 - N \times (N - M + 1)$ - матрица, составленная из $N - M + 1$ собственных векторов, соответствующих нулевым собственным значениям матрицы $\mathbf{Y}(M - 1)\mathbf{Y}^T(M - 1)$, а матрица $\mathbf{Y}(M - 1)$ формируется из матрицы \mathbf{Y} путем вычеркивания одного, в данном случае, i -го вектора-столбца \mathbf{Y}_i .

В работах [2, 3] показано, что величина (2) связана с информативностью признаков при распознавании двух классов и исключение признаков, для которых она принимает малые значения, не ухудшает качества классификации при оценивании параметров разделяющей функции (1), удовлетворяющих в смысле минимальной среднеквадратической ошибки уравнению

$$\mathbf{Y}\mathbf{a} = \mathbf{b}, \quad (3)$$

где \mathbf{b} - вектор, формируемый по заданным допускам, определяющим область решений [4].

Решение о принадлежности к классу осуществляется по знаку функции $g(\mathbf{X})$, т.е. её значение сравнивается с нулем. Решающая граница между классами принимает вид

$$g(\mathbf{X}) = 0. \quad (4)$$

Уравнение (4) представляет собой уравнение гиперплоскости в пространстве признаков. Если перенести w_0 из разделяющей функции (1) в правую часть уравнения (4), то принцип, по которому происходит классификация, не изменится, и уравнение решающей границы будет иметь вид

$$g(\mathbf{X}) = \text{const}. \quad (5)$$

При этом размер вектора обобщенных признаков уменьшится на единицу, что равносильно удалению из обучающей матрицы \mathbf{Y} одного признака.

Можно предположить, что значение «1», добавляемое к набору наблюдаемых векторов-признаков, служит только для упрощения нахождения значения const решающей границы (5) (значения параметра w_0 при записи вида (4)) и не несет в себе никакой дополнительной информации с точки зрения описания расположения различных классов в признаковом пространстве. В тоже время этот вектор-столбец наряду с остальными входит в обучающую матрицу и участвует в вычислении величин (2).

Очевидно, что при выявлении в обучающей матрице вектора-столбца, все компоненты которого являются одинаковыми, он должен быть исключен, как не несущий информации о различии классов. Так, при распознавании цифр «2» и «3» непосредственно по их изображениям в виде прямоугольных областей интуитивно понятно, что области в правом

и левом верхних углах изображений можно исключить из состава признаков без ущерба для качества распознавания. В самом деле, если фон у этих изображений будет одинаковым, то значения интенсивностей отсчетов в указанных областях изображений будут представлять собой константу, что приведет к появлению в обучающей матрице столбца с одинаковыми значениями.

При существовании нескольких таких признаков, независимо от их конкретных значений, обучающая матрица окажется вырожденной ввиду наличия линейных зависимостей между столбцами. Тот факт, что фактическое увеличение количества признаков при добавлении «1» происходит без измерений дополнительных характеристик объекта, говорит о том, что добавленный признак не может нести дополнительной информации, а только изменяет форму записи решающей границы.

Исходя из этого предлагается использовать значение величины (2) вектора признаков «1» в качестве порога $S_{\text{доп}}$ для выявления группы наименее информативных признаков в обучающей матрице.

Принимая во внимание результаты, полученные в [3], процедура оценивания строится следующим образом. Для каждого вектора \mathbf{Y}_j , $j = \overline{1, M}$ вычисляется величина S_j . В матрице \mathbf{Y} находится столбец с минимальным значением S_j . Если таким столбцом оказался столбец «1», то исключение признаков прекращается, иначе столбец \mathbf{Y}_j исключается из матрицы \mathbf{Y} , и величины S_j для оставшихся столбцов вычисляются заново.

Преимущества используемого подхода к решению задачи выбора системы признаков особенно проявляются в случае распознавания с использованием малого числа наблюдений, что может быть вызвано различными условиями (невозможность получить достаточно большой объем наблюдений или нестационарность распределений образов в признаковом пространстве, требующая частой настройки классификатора).

При использовании критерия сопряженности в качестве критерия информативности добавление в обучающую матрицу \mathbf{Y} столбца с константой приведет к исключению в первую очередь тех признаков, значения которых в разных наблюдениях изменяются незначительно. Следовательно, в первую очередь будут исключены менее изменчивые признаки.

Далее приводятся результаты экспериментов, подтверждающие возможность снижения размерности указанной системы признаков по показателю сопряженности (2) составляющей признакового пространства, вводимой для учета смещения разделяющей гиперплоскости.

Результаты экспериментов

Для проверки работоспособности предлагаемого подхода проведены следующие эксперименты. Для каждой пары цифр формировалась обучающая матрица и выполняется процедура исключения признаков с оцениванием параметров разделяющей

функции (1). Количество образцов различных цифр было различным – от 30 до 150. На рис. 2 изображены графики процента правильно распознаваемых цифр в зависимости от количества оставленных признаков. Вертикальными линиями обозначены точки, в которых заканчивается исключение признаков, у которых величина (2) меньше, чем у столбца констант. Приводятся графики при классификации пар цифр 0-1 и 0-6. Цифры 1 и 6 соответственно наиболее непохожи и похожи на цифру 0 по сравнению с остальными. Видно, что при выполнении описанной процедуры исключения качество классификации начинает ухудшаться при исключении числа признаков, большего, чем количество признаков, исключенных на основании сравнения с признаком-константой.

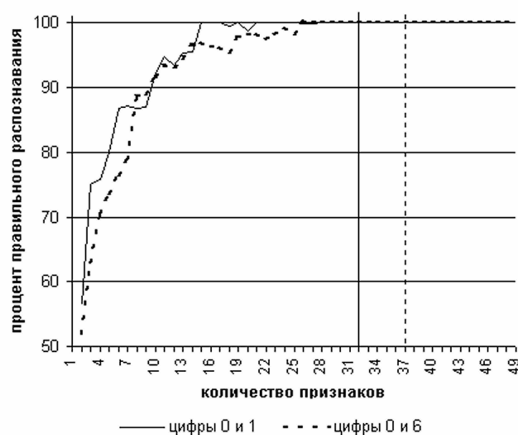


Рис. 2. Зависимость качества классификации от количества признаков

Для других обучающих матриц получены аналогичные результаты. Это может говорить о том, что величину S_j признака-константы нельзя использовать в качестве порога для выбора пространства признаков минимальной размерности, достаточного для линейной разделимости классов в нем. Однако эту

величину можно взять в качестве критерия избыточности признакового пространства и использовать описанную процедуру для его сокращения без ущерба для качества классификации.

Заключение

Предлагаемая процедура анализа признакового пространства на основе показателя сопряженности векторов с нуль-пространством и использование в качестве допуски критерия сопряженности признака с постоянным значением, позволяет осуществлять обоснованное сокращение числа признаков без ущерба для качества классификации. Для применения предлагаемого метода не требуется использование априорных вероятностных характеристик обучающей выборки или требований к её статистической достоверности в виде её большого объема.

При необходимости возможно также решать задачи обоснованного наращивания признакового пространства.

Работа выполнена при поддержке РФФИ, грант 99-01-00079.

Литература

1. Фу К. Последовательные методы в распознавании образов и обучении машин. Пер с англ. М.: Наука, 1971, 256с.
2. Фурсов В.А. Шустов В.А. Формирование признакового пространства по критерию сопряженности векторов измерений // Компьютерная оптика, № 20. С. 140-142.
3. Фурсов В.А. Шустов В.А. Алгоритмы формирования признакового пространства по критерию сопряженности векторов измерений// Компьютерная оптика, № 21. С. 176-178.
4. Дуда Р., Харт П. Распознавание образов и анализ сцен // Пер с англ. М.: Мир, 1976, 512 с.

Feature space reduction by the criterion of contingency with null-space

V.A. Shustov^{1,2}

¹Image Processing Systems Institute of RAS,

²Samara State Aerospace University

Abstract

Often, when forming a feature space, an excessive number of features is defined deliberately, and then, uninformative features are excluded. This technique is known as the feature space reduction [1]. It raises a problem of choosing an informativeness indicator and a threshold value of the selected indicator.

The works [2, 3] investigated the possibility that an indicator of contingency of a vector corresponding to a particular feature with null-space of the transposed matrix, composed of the remaining feature vectors, can be used as a criterion of informativeness of this feature. This technique involves setting a certain threshold for the contingency indicator and excluding the features if their contingency is less than specified. However, the issue of a reasonable choice of a threshold for obtaining an optimal set of features from the original set remains open.

This paper proposes and investigates a method for assigning a quantitative value to the threshold based on calculating the criterion of contingency of a feature space component, which is introduced to allow for the displacement of the separating hyperplane.

Keywords: null-space, feature space, feature space reduction, indicator of contingency of a vector, transposed matrix, hyperplane.

Citation: Shustov VA. Feature space reduction by the criterion of contingency with null-space. *Computer Optics* 2002; 23: 66-68.

References

- [1] Fu KS. Sequential methods in pattern recognition and machine learning. New York, London: Academic Press; 1968.
- [2] Fursov VA, Shustov VA. Formation of feature space by the criterion of conjugacy of measurement vectors [In Russian]. *Computer Optics* 2000; 20: 140-142.
- [3] Fursov VA, Shustov VA. Algorithms of feature space formation by the criterion of conjugacy of measurement vectors [In Russian]. *Computer Optics* 2001; 21: 176-178.
- [4] Duda RO, Hart PE. Pattern classification and scene analysis. New York: John Wiley and Sons; 1973.