

АЛГОРИТМЫ ФОРМИРОВАНИЯ ПРИЗНАКОВОГО ПРОСТРАНСТВА ПО КРИТЕРИЮ СОПРЯЖЕННОСТИ ВЕКТОРОВ ИЗМЕРЕНИЙ

В.А. Фурсов, В.А. Шустов
Институт систем обработки изображений РАН

Введение

Одной из центральных в общей проблеме распознавания образов является проблема формирования признакового пространства. К сожалению, пока отсутствуют полностью формализованные методы ее решения. Обычно в рамках этой общей проблемы рассматривают два типа задач: добавление к имеющемуся набору признаков новых, улучшающих качество распознавания и, наоборот, исключение из заданного избыточного множества признаков наименее информативных (редукция признакового пространства).

Пусть обобщенная (линейная по параметрам) разделяющая функция имеет вид:

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

где $\mathbf{a} = [w_0, \mathbf{w}]^T = [w_0, w_1, \dots, w_d]^T$ - $M \times 1$ - вектор искоемых параметров (весовых коэффициентов [4]), \mathbf{x} - вектор, компоненты которого суть признаки, а $\mathbf{y}(\mathbf{x})$ - $M \times 1$ - вектор обобщенных признаков, который, в частности, может быть задан в виде $\mathbf{y}(\mathbf{x}) = [1, \mathbf{x}]^T$. 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 $\mathbf{Y}_j, j = \overline{1, M}$.

В работах [1, 2] рассматривалась задача формирования признакового пространства при обучении распознаванию образов по малому числу наблюдений с использованием показателей сопряженности векторов признаков. В частности, была предложена мера взаимной ориентации векторов признаков, вычисляемая для каждого вектора $\mathbf{Y}_i, i = \overline{1, M}$ величина

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

где $\mathbf{T}_0(L-1)$ - $N \times (N-L+1)$ - матрица, составленная из $(N-L+1)$ собственных векторов, соответствующих нулевым собственным значениям матрицы $\mathbf{Y}(L-1) \mathbf{Y}^T(L-1)$, \mathbf{Y}_L - $N \times L$ - матрица, составленная из L векторов-столбцов матрицы \mathbf{Y} ($2 \leq L \leq M$), а матрица $\mathbf{Y}(L-1)$ формируется из матрицы \mathbf{Y}_L путем вычеркивания одного, в данном случае, i -го вектора-столбца \mathbf{Y}_i .

В работе [2] указывалось, что в общем случае для матрицы \mathbf{Y} из M столбцов возможно $J_M = \sum C_M^i, i = \overline{2, M}$ различных вариантов построения матриц \mathbf{Y}_L , для каждого из которых должно рассчитываться L показателей (2). Если вычисляются только M показателей (для исходной

$N \times M$ - матрицы \mathbf{Y}) и исключаются векторы (признаки), для которых этот показатель оказался равным нулю, то гарантируется, что среди сингулярных чисел матрицы \mathbf{Y} не окажется нулевых. Однако при этом возникает опасность, что будут исключены также и информативные признаки.

Например, если среди векторов признаков имеются два линейно-зависимых, показатель (2), вычисленный по матрице \mathbf{Y}_M , будет равен нулю для обоих, в случае трех компланарных векторов он окажется равным нулю для всех трех и так далее. Конечно, последовательный "просмотр" всех возможных из множества J_M вариантов подпространств позволит избежать исключения информативных признаков. Однако при этом объем вычислительной работы, особенно при большой размерности исходного признакового пространства (например, несколько сотен), будет непомерно большим даже для многопроцессорных систем.

В настоящей работе исследуются процедуры формирования признакового пространства, позволяющие, как правило, избежать потери информативных признаков, но при этом объем вычислительной работы остается примерно таким же, как и в случае вычисления M показателей (2) для исходной $N \times M$ - матрицы \mathbf{Y} .

1. Особенности задачи

В [3] рассматривался основанный на вычислении показателей (2) алгоритм, в котором использовалась матрица \mathbf{Y}_{M-1} , формируемая из матрицы \mathbf{Y} путем вычеркивания одного, в данном случае, i -го вектора-столбца \mathbf{Y}_i , и вычислялись M показателей. Далее полученные величины S_i сравнивались с заданным допустимым значением D . Если оказывалось, что $S_i < D$, то столбец \mathbf{Y}_i исключался. Величина D определялась экспериментально и не должна была быть слишком малой, так как малые значения S_i говорят о сильной мультиколлинеарности векторов матрицы \mathbf{Y} и, следовательно, о плохой обусловленности задачи оценивания.

В описанной процедуре при наличии, например, двух линейно-зависимых векторов признаков величина S_i для обоих столбцов матрицы будет равна нулю, независимо от того, какой вклад каждый из этих признаков вносит в эффективность распознавания. При этом по результатам сравнения S_i с порогом D будут исключены оба столбца и есть опасность избавиться от "хороших" признаков, хотя информативность одного из них или их суммы, возможно, достаточно велика.

В таблице 1 приведен простой пример матрицы \mathbf{Y} из четырех признаков и значения величин S_i

для каждого из них. В этом примере, строго следуя описанному выше правилу сравнения параметров сопряженности с допуском, следовало бы исключить все векторы (признаки), хотя среди них имеется четыре варианта признакового пространства, содержащих пары ортогональных векторов.

Таблица 1. Пример неопределенной ситуации

№ объекта	Признаки			
	Y_1	Y_2	Y_3	Y_4
1	0	2	0	4
2	0	4	0	8
3	1	0	3	0
4	2	0	6	0
S_i	0	0	0	0

Ясно, что если сначала последовательно вычислить все коэффициенты парной сопряженности и из каждой пары векторов (Y_1, Y_3) и (Y_2, Y_4), для которых показатель S_i с очевидностью равен нулю, исключить по одному вектору, то неопределенность снимается. Поскольку среди признаков может оказаться любое число линейно-зависимых, процедуру вычисления S_i мы обязаны выполнить для трех, четырех и так далее векторов во всех возможных сочетаниях. Ниже описываются другие, обладающие меньшей вычислительной сложностью, алгоритмы, позволяющие избежать исключения информативных признаков.

2. Процедуры формирования признакового пространства

Возможны несколько путей уменьшения вычислительной сложности процедуры отбора информативных признаков. Можно, например, исключать признаки последовательно по одному, каждый раз после исключения производя перерасчет параметров S_i для остающихся признаков. Можно изменить способ вычисления S_i таким образом, чтобы линейная зависимость между признаками учитывалась при вычислении меры сопряженности только один раз. Для этого необходимо вычислять сопряженность с подпространством, которое формируется несколько иначе. В частности, при вычислении S_i для i -го столбца исключаются все столбцы с первого по i -тый включительно. Для того чтобы подчеркнуть отличия в схеме формирования подпространства, вычисленный таким способом показатель сопряженности типа (2), будем обозначать V_i .

Ясно, что значения V_i уже не будут инвариантны к порядку следования признаков в обучающей матрице, однако такой способ позволяет избежать указанные ранее трудности. По сути, если S_i принять за величину, которая характеризует вклад вектора в сопряженность всей совокупности признаков обучающей матрицы, то величина V_i представляет собой некоторую “добавочную” сопряженность, вносимую одним признаком к ранее выбранной группе признаков.

Различные алгоритмы выбора признаков можно построить, комбинируя во всех возможных сочетаниях следующие три фактора:

- использование параметров S_i или V_i ,
- исключение “плохих” признаков из начальной группы (уменьшение) или добавление “хороших” признаков (увеличение),
- вычисление параметров сопряженности векторов один раз или пересчет этих показателей каждый раз при изменении количества признаков.

Для удобства их дальнейшего сопоставления эти комбинации сведены в таблицу 2 и обозначены цифрами 1-7. Далее алгоритмы будут называться по номерам, указанным в таблице.

Таблица 2. Классификация алгоритмов

№ п/п	Показатель	Изменение размерности	Пересчет показателя
1	S_i	уменьшение	да
2	S_i	уменьшение	нет
3	S_i	увеличение	да
4	S_i	увеличение	нет
5	V_i	уменьшение	да
6	V_i	уменьшение	нет
7	V_i	увеличение	все равно

3. Результаты экспериментов

Для оценки эффективности алгоритмов формирования признакового пространства установим зависимость относительного количества неправильно классифицируемых обучающих объектов от числа выбранных из начальной группы признаков для случая распознавания двух классов. Для оценки параметров разделяющей гиперплоскости используется метод наименьших квадратов [4].

Задача оценивания параметров классификатора решается следующим образом. Ищется разделяющий вектор \mathbf{a} , удовлетворяющий в смысле минимальной среднеквадратической ошибки уравнению

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

где \mathbf{b} - вектор, формируемый по заданным допускам, определяющим область решений [4]. Предполагается, что число строк N матрицы \mathbf{Y} не намного превышает число столбцов (признаков) M .

Последнее обстоятельство дает основание классифицировать задачу определения вектора \mathbf{a} как задачу оценивания по малому числу наблюдений. Это вовсе не означает, что число наблюдений исчисляется единицами. Речь идет о числе степеней свободы, то есть фактически о размерности нуль-пространства матрицы \mathbf{Y}^T . Например, при числе признаков 500, выборка из 700 обучающих объектов должна считаться малой.

Для сравнения указанных в таблице 2 алгоритмов проводились следующие эксперименты. Строились зависимости процента ошибочного распознавания цифр для всех типов алгоритмов для различных систем признаков. В качестве цифр использовались монохромные 256-цветные изображения раз-

личных шрифтов размера 16×12. Системы признаков представляли собой суммы интенсивностей пикселей не перекрывающихся областей изображения размера 4×4, 2×2 и 1×1. Количество обучающих примеров по каждой цифре равнялось 30. Таким образом, размеры обучающей матрицы были равны соответственно 60×13, 60×49 и 60×193. В последнем случае, при использовании в качестве системы признаков интенсивностей пикселей изображения, обучающая матрица получается недоопределенной, поэтому можно было использовать алгоритмы только 3, 4 и 7 типов.

Для сравнения эффективности алгоритмов с тем эффектом, которого можно достичь, формируя набор признаков, по которым происходит классификация наилучшим образом, строились графики зависимостей процента достижимого минимума и среднего путем перебора всех возможных наборов указанного количества признаков.

Проведенный анализ результатов показал следующее. Алгоритмы 3, 4 и 7, которые формируют признаковое пространство путем наращивания количества признаков, дают одинаковые результаты на различных обучающих матрицах.

На рисунке 1 представлены графики зависимости процента ошибок распознавания от числа используемых признаков, выбранных соответствующим методом для распознавания цифр 2 и 3. В качестве признаков использовалась сумма интенсивностей пикселей прямоугольных областей изображения размера 4×3. График иллюстрирует количество ошибок распознавания при использовании части признаков, отобранных различными способами: *a* – выбор лучших признаков перебором; *b* – средняя ошибка по всем комбинациям признаков; *c* – выбор признаков алгоритмом 1-го типа; *d* – алгоритмом 6-го типа; *e* – 2-го типа; *f* – 5-го типа.

Графики ошибок при использовании алгоритмов всех типов близки к графику среднего, который соответствует среднему числу ошибок при случайном выборе группы признаков. При этом алгоритмы, использующие перерасчет параметров дают более хорошую выборку, а алгоритмы с наращиванием признакового пространства более плохую, чем остальные. Наилучшие результаты по различным обучающим матрицам показал метод первого типа.

Заключение

С помощью описанных алгоритмов анализа признакового пространства можно упорядочить признаки по степени их важности для распознавания, используя только обучающую матрицу. Эксперименты подтверждают существование связи мер сопряженности с информативностью признаков

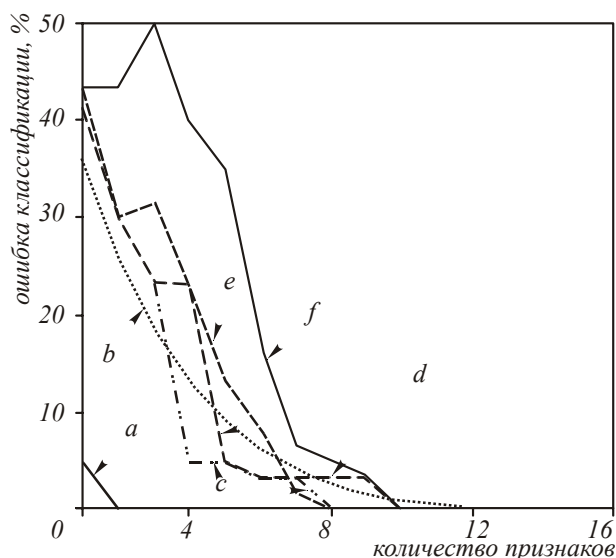


Рис. 1. Зависимость относительного числа ошибок распознавания от числа признаков.

При совместном использовании описанных алгоритмов формирования признакового пространства с оценкой параметров разделяющей гиперплоскости методом наименьших квадратов можно получить оценку сверху ошибки распознавания при использовании для классификации подпространства признакового пространства указанного размера. При этом такие оценки достигают нуля быстрее, нежели оценка по среднему и требуют для своего вычисления гораздо меньших вычислительных затрат.

Все алгоритмы гарантируют невырожденность информационной матрицы, формируемой для оценивания по методу наименьших квадратов.

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

Литература

1. Фурсов В.А. Метод проекций на нуль-пространство в проблеме распознавания образов по малому числу наблюдений // Тр. Всероссийской конференции “Математические методы распознавания образов” (ММРО-9). Москва, 15-19 ноября, 1999 г. С. 119-121.
2. Vladimir A. Fursov. Training in Pattern Recognition from a small number of observations using projections onto null-space // Prog 15th International Conference on Pattern recognition (ISPR) 2000. V. 2/ P. 789-792.
3. Фурсов В.А. Шустов В.А. Формирование признакового пространства по критерию сопряженности векторов измерений // Компьютерная оптика. № 20. С. 140-142.
4. Дуда Р., Харт П. Распознавание образов и анализ сцен. Пер с англ. М., Мир. 1976. 512 с.