

УДК 519.711.3

О РЕАЛИЗАЦИИ РАСПОЗНАВАТЕЛЯ С МИНИМАЛЬНЫМ ЧИСЛОМ ВХОДОВ

О.В. GERMAN, Н.Л. БОБРОВА, А.Р. САМКО

Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 14 февраля 2011

Дан наиболее удобный, в практическом отношении, вариант в задачах распознавания с большим количеством входных признаков при взаимной независимости, где коэффициенты распознающей (дискриминаторной) функции могут быть получены через решение системы линейных неравенств, либо на основе методологии МГУА, либо обучением нейрораспознавателя. Для получения системы взаимно независимых входных признаков используется основная идея факторного анализа.

Ключевые слова: нейрораспознаватель, соотношение Хотеллинга, нормированное значение, соотношение Тэрстоуна, дискриминаторная функция, древовидная структура, метод обучения перцептрона, процедура обратного распространения ошибки.

Введение

Одна из основных задач реализации нейрораспределителя состоит в его обучении. При большом числе входов (измеряемом десятками) и их взаимозависимости обучение распознавателя становится крайне сложным и трудоемким процессом.

Наиболее удобный в практическом отношении вариант получаем при взаимной независимости входных признаков, так что коэффициенты распознающей (дискриминаторной) функции могут быть получены относительно легко через решение системы линейных неравенств [1], либо на основе методологии МГУА [2], либо обучением (нейро)распознавателя [3].

Для получения системы взаимно независимых входных признаков и значительного, в общем случае, сокращения их числа вполне можно использовать основную идею факторного анализа, определяемую соотношением Хотеллинга [4]:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m, \quad (1)$$

$$j = 1, \dots, n; m \ll n.$$

Здесь Z_j – нормированное значение j -го входного признака; F_1, \dots, F_m – вводимые в рассмотрение независимые факторы; a_{jk} ($k = \overline{1, m}$) – весовые коэффициенты факторов (нагрузки).

Наша задача в этой работе – показать, как, отправляясь от (1), получить (нейро)распознаватель с небольшим числом взаимно независимых входов и как адаптировать механизм распознавания в этом случае. Следует отметить важность этой задачи для практики. Во-первых, при небольшом числе входов обучение нейрораспознавателя становится относительно простым, например, на основе процедуры обратного распространения ошибки. Во-вторых, при независимости входов друг от друга принципиально можно добиться линейной распознающей функции или, во всяком случае, исключить в представлении функции комбинации, включающие попарные произведения входных признаков или их степеней.

Формализация

Рассматриваем входную таблицу (двумерную матрицу) входов $X=[x_{ij}]$, $j=\overline{1,n}$, $i=\overline{1,k}$ (n – число признаков; k – число объектов). Один дополнительный $(n+1)$ -й признак (который обозначим Y) играет роль диагностического (например, в простейшем случае трактуемый как «норма/отклонение», «пригоден/непригоден», «Да/Нет» и т.п.). Пронормируем значения x_{ij} , перейдя к

$$Z_{ij} = \frac{x_{ij} - \overline{x_j}}{\sigma_j}, \quad (2)$$

где $\overline{x_j}$ – среднее значение j -го входного признака; σ_j – его среднеквадратичное отклонение.

Пусть $Z=[Z_{ij}]$ – таблица нормированных входных признаков.

Замечание. Признак Y не нормируем; в таблице Z он не присутствует.

В отличие от (1) будем использовать разложение

$$Z = F \cdot A, \quad (3)$$

где

$$Z = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1n} \\ z_{21} & z_{22} & \dots & z_{2n} \\ \dots & \dots & \dots & \dots \\ z_{k1} & z_{k2} & \dots & z_{kn} \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kn} \end{bmatrix}, \quad F = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1n} \\ f_{21} & f_{22} & \dots & f_{2n} \\ \dots & \dots & \dots & \dots \\ f_{k1} & f_{k2} & \dots & f_{kn} \end{bmatrix}.$$

В описываемом здесь подходе к построению нейрораспознавателя Z заменяется на F , причем i -строке в Z соответствует i -строка в F и наоборот. При этом в соответствии с положениями факторного анализа число факторов m значительно меньше n , например, $m \approx (0,1 \div 0,25)n$. Матрица A является «фиксированной»: при обученном нейрораспознавателе эта матрица не изменяется и играет роль связки между Z и F , позволяя для каждого нового образца $\langle z_{i1}, z_{i2}, \dots, z_{in} \rangle$ легко находить соответствующий набор факторов $\langle f_{i1}, f_{i2}, \dots, f_{im} \rangle$. Далее имеем

$$\frac{1}{n} Z^T \cdot Z = \frac{1}{n} A^T F^T \cdot F \cdot A.$$

Произведение $\frac{1}{n} F^T \cdot F$ суть корреляционная матрица независимых факторов, а поскольку

факторы взаимно независимы, то $\frac{1}{n} F^T \cdot F = E$ (E – единичная диагональная матрица).

Аналогично $\frac{1}{n} Z^T \cdot Z$ дает нам корреляционную матрицу входных признаков – R_z .

Получаем известное соотношение Тэрстоуна

$$R_z = A^T \cdot A. \quad (4)$$

Отыскание матрицы A можно найти в [4]. В общем случае эта матрица не единственная, т.е. из (4) можно получить сколь угодно много решений. Для того, чтобы решение было единственным, применяют процедуру «вращения» пространства признаков. Более подробно об этом можно читать в [4].

Зная A , из (3) получим

$$Z \cdot A^T = F \cdot (A \cdot A^T), \quad (5)$$

$$Z \cdot A^T \cdot (A \cdot A^T)^{-1} = F.$$

Соотношение (5) по Z и A позволяет найти F . Нетрудно убедиться, что для каждого нового образца $\langle z_{i1}, z_{i2}, \dots, z_{im} \rangle$ из (5) находим вектор-строку $\langle f_{i1}, f_{i2}, \dots, f_{im} \rangle$ только по одной вектор-строке z_i .

Итак, мы достигли следующего. Вместо входной матрицы X мы имеем таблицу

$$(F, Y) = \begin{bmatrix} f_{11} f_{12} \dots f_{1m} & Y_1 \\ f_{21} f_{22} \dots f_{2m} & Y_2 \\ \dots & \dots \\ f_{k1} f_{k2} \dots f_{km} & Y_m \end{bmatrix}, \quad (6)$$

где значения диагностического признака Y определяются в общем случае существенно меньшим числом независимых факторов.

Таблица (6) используется как обучающая для определения дискриминаторной функции, которая в простейшем случае имеет линейный вид

$$\varphi(t_1, \dots, t_m) = \alpha_1 \cdot f_1 + \alpha_2 \cdot f_2 + \dots + \alpha_m \cdot f_m. \quad (7)$$

Следует также иметь в виду, что минимизация числа признаков в факторном анализе достигается не только тем, что отбрасываются зависимые признаки, но и тем, что оставляют лишь те признаки, которые в совокупности позволяют восстановить исходную таблицу наблюдений с приемлемой в статистическом смысле точностью.

Определение структуры распознавателя

Распознаватель с древовидной структурой [1] может быть представлен в общем виде, как показано на рис. 1.

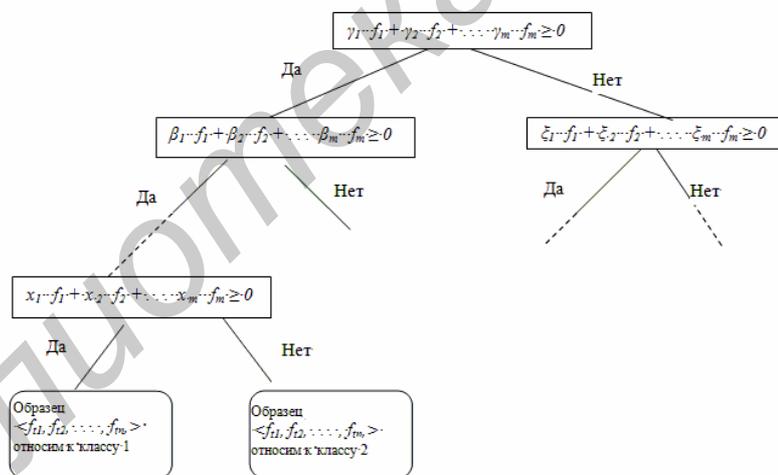


Рис. 1. Распознаватель с древовидной структурой

В узлах дерева находятся линейные неравенства вида $\sum_i \mu_i f_i \geq 0$, где μ_i – весовые коэффициенты при факторах f_i ($i = \overline{1, n}$). Вход в дерево осуществляется через его корневую вершину. Проверяется выполнение неравенства $\gamma_1 \cdot f_1 + \gamma_2 \cdot f_2 + \dots + \gamma_m \cdot f_m \geq 0$.

Если неравенство выполняется, то осуществляется переход по ребру «Да», показанный на рис. 1, если не выполняется, то переходим по ребру «Нет». В следующей вершине проверяется следующее неравенство и т.д., пока не доберемся к листовой вершине, например, $x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_m \cdot f_m \geq 0$. Теперь уже по выполнимости/невыполнимости этого неравенства на данном входном образце $\langle f_1, f_2, \dots, f_m \rangle$ получаем заключение о принадлежности входного объекта к классу 1 или классу 2. Проблема построения древовидного распознавания кратчайшей длины остается актуальной.

По таблице (6) строим систему неравенств с неравенствами вида

$$f_{i1} \cdot c_1 + f_{i2} \cdot c_2 + \dots + f_{im} \cdot c_m \geq 0,$$

если признак Y_i соответствует набору $\langle f_{i1}, f_{i2}, \dots, f_{im} \rangle$ класса 1, и

$$f_{i1} \cdot c_1 + f_{i2} \cdot c_2 + \dots + f_{im} \cdot c_m < 0,$$

если признак Y_i , ($i=1, \dots, k$) соответствует классу 2. Здесь C_i – неизвестные коэффициенты, которые необходимо найти. В [1, 5] описана стратегия решения системы линейных неравенств на основе устранения невязок (СУН). В этой стратегии все неравенства предварительно приводятся к виду « \geq » и последовательно устраняются невязки. Невязкой считается неравенство вида

$$f_{i1} \cdot d_1 + f_{i2} \cdot d_2 + \dots + f_{im} \cdot d_m \geq d_{m+1}, \quad (8)$$

где $d_{m+1} > 0$. Пусть, например, $f_{i1} > 0$. Тогда строится подстановка с новой переменной $d'_1 \geq 0$:

$$d_1 = \frac{d_{m+1}}{f_{i1}} - \frac{d_2 \cdot f_{i2}}{f_{i1}} - \dots - \frac{f_{im} d_m}{f_{i1}} + d'_1. \quad (9)$$

Подстановка (9) проводится в текущую систему неравенств. Получаем новую систему. Если невязок нет, то такая система выполняется для нулевых значений переменных. В противном случае находим очередную невязку, выражаем из нее переменную с положительным коэффициентом и т.д. Если в неравенстве-невязке все коэффициенты при переменных отрицательны, то при условии неотрицательности переменных (что обеспечивается стратегией СУН) делаем вывод о несовместимости системы. В описываемом подходе подобная невязка «исключается» из текущей системы неравенств и процесс ведется далее (с возможными последующими исключениями невязок из системы) до описанного решения.

Итак, после получения системы без невязок найдем коэффициенты c_1, c_2, \dots, c_m и получим неравенство для корневой вершины дерева:

$$c_1 \cdot f_1 + c_2 \cdot f_2 + \dots + c_m \cdot f_m \geq 0. \quad (10)$$

Если это неравенство выполняется на данном входе $\langle f'_1, f'_2, \dots, f'_m \rangle$, то вход относим к классу 1, в противном случае – к классу 2. Однако в силу того, что при выполнении СУН могли оказаться некоторые неравенства-невязки, следует проверить, для каких из этих неравенств принимается ложное решение: подставляем в (10) значения f_1, f_2, \dots, f_m соответствующей строки таблицы (6).

Таким образом, в общем случае следует искать новые разложения неравенства для второго уровня распознавателя на рис. 1,а, возможно и на более глубоких уровнях. На каждом очередном уровне число неравенств, участвующих в записи системы, последовательно сокращается. В идеале, разумеется, достаточно только одного единственного неравенства для работы распознавателя. В худшем случае – k (k – число строк в таблице 6). Все определяется тем, как «перемешаны» точки относительно класса 1 и 2 в n -мерном пространстве. Например, можно показать, что при перемешивании типа «локализованного выступа» (рис. 2) достаточно «в среднем случае» 4 неравенства. Наиболее плохой вариант полного перемешивания (рис. 3) потребовал в экспериментах около $0,3 \cdot k$ неравенств.

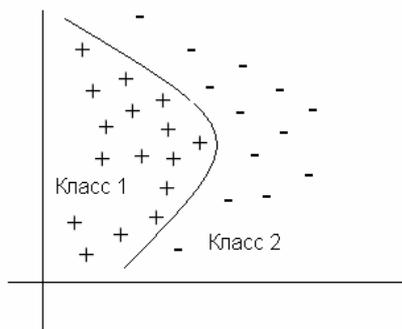


Рис. 2. Перемешивание типа «локализованного выступа»

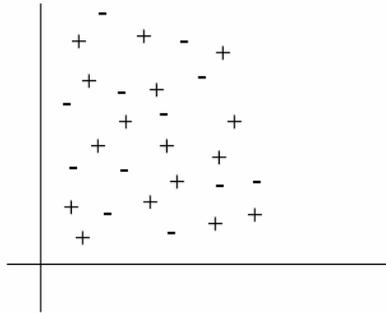


Рис. 3. Вариант полного перемешивания

Оценка сложности древовидного распознавателя

Имеется бесчисленное множество взаимного расположения двух множеств точек. Начнем рассмотрение со случая, представленного на рис. 4.

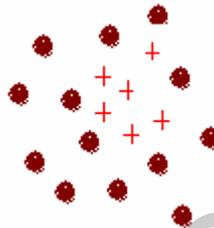


Рис. 4. Вариант расположения множеств

Здесь точки одного множества представлены кружками, а второго – крестиками. Особенность такого расположения в том, что точки обоих множеств расположены компактно, без перемежения. Вместе с тем, разделить такие множества точек с помощью линейной дискриминаторной функции нельзя. Такой функции просто нет, так как она в силу построения точек должна была быть неоднозначной. Обозначим через ρ – плотность точек рассматриваемого (внутреннего) множества, т.е. число точек, приходящееся на единицу площади (например, мм^2); N – число точек множества.

Сначала будем считать, что точки внутреннего множества можно заключить в круг, не содержащий точек внешнего множества (рис. 5,а).

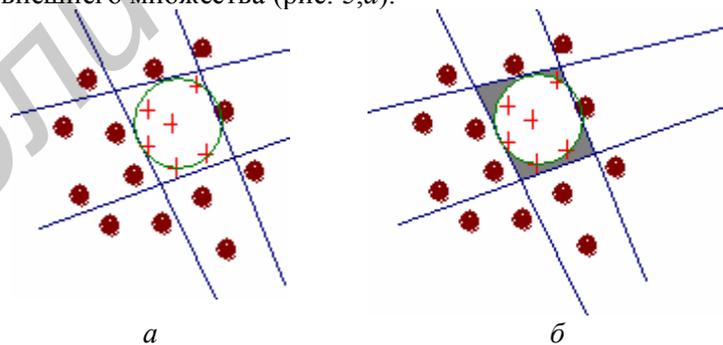


Рис. 5. Систематизация расположения множеств

Тогда число неравенств Z , требуемых для классификации, можно оценить исходя из числа сторон внешнего (минимального по числу сторон) многоугольника, описанного вокруг этого круга и такого, что в области, принадлежащие многограннику и не принадлежащие кругу (закрашены серым цветом на рис. 5,б), не попадает ни одна из точек внешнего множества.

Площадь внутреннего круга примерно равна

$$S_{in} = \frac{N}{\rho}. \tag{11}$$

Обозначив радиус окружности, ограничивающей круг, через r , получим

$$\pi \cdot r^2 = \frac{N}{\rho} \rightarrow r \approx \sqrt{\frac{N}{\pi \cdot \rho}} \quad (12)$$

Площадь правильного многоугольника с n сторонами, описанного вокруг окружности радиусом r , составляет

$$S_{MN} = \pi \cdot r^2 \cdot \operatorname{tg} \frac{\pi}{n}. \quad (13)$$

Теперь требуемое минимальное число неравенств можно оценить как $Z = n$ при условии

$$\rho \cdot (S_{MN} - S_{in}) < 1. \quad (14)$$

Имеем из (14)

$$\rho \cdot (\pi r^2 \cdot \operatorname{tg} \frac{\pi}{n} - \pi r^2) = \rho \cdot \pi r^2 (\operatorname{tg} \frac{\pi}{n} - 1) = \rho \cdot \pi \cdot \frac{N}{\pi \cdot \rho} (\operatorname{tg} \frac{\pi}{n} - 1) < 1.$$

Итак, получаем следующее результирующее соотношение:

$$N \cdot (\operatorname{tg} \frac{\pi}{n} - 1) < 1. \quad (15)$$

Это соотношение при больших N можно упростить до

$$\operatorname{tg} \frac{\pi}{n} < 1 \quad (16)$$

Это соотношение начинает выполняться при $n = 5$, а при $n = 4$ имеем пороговое значение. Замечательным является тот факт, что число неравенств не зависит ни от плотности, ни от числа точек при выполнении рассмотренных допущений. Поскольку эллипс является растяжением круга и описывающего его многоугольника, то соотношение (6) сохраняет свою силу, ибо площади при растяжении не изменяются. Итак, число неравенств для локализации вложенных множеств без перемешивания точек оценивается величиной в пределах десятка независимо ни от плотности, ни от числа точек. Рассмотренные случаи дают нам убедительные примеры тех ситуаций, когда преимущества классифицирующего дерева проявляются в полную силу.

Заключение

Итак, описываемый подход реализации включает следующие шаги. Определяем таблицу типа (6) для исходной таблицы X , причем число независимых факторов составляет порядка 0,1–0,25 от исходного числа признаков, что существенно важно в задачах распознавания с большим (десятки, сотни) числом признаков. Каждый новый входной экземпляр $\langle x_1, x_2, \dots, x_n \rangle$ с помощью (5) переводится в $\langle f_1, f_2, \dots, f_k \rangle$ и набор $\langle f_1, f_2, \dots, f_k \rangle$ подается на распознаватель. Размер матрицы A после обучения не растет и остается фиксированным. Построение дискриминаторной функции на F значительно проще и может строиться, например, методом обучения перцептрона на основе процедуры обратного распространения ошибки. Представленный в статье древовидный распознаватель дает отличные результаты при объеме области перемешивания двух множеств порядка 20–30% (от объема исходного множества).

ABOUT REALIZATION OF DISCRIMINATOR WITH MINIMUM NUMBER OF ENTRANCES

O.V. GERMAN, N.L. BOBROVA, A.R. SAMKO

Abstract

The most comfortable in a practical relation variant is given at mutual independence of entrance signs, where the coefficients of recognizing (by a discriminator) function can be got through the decision of set of linear inequalities, or on the basis of methodology of MGUA. For the receipt of the system of mutually independent entrance signs we use the basic idea of factor analysis.

Литература

1. *Герман О.В., Дорожжина Н.Н.* Теория информационных процессов и систем. Мн., 2007.
2. *Ивахненко А.Г., Юрачковский О.П.* Моделирование сложных систем по экспериментальным данным. М., 1987.
3. *Галушкин А.И.* Теория нейронных сетей. М., 2000.
4. *Иберла К.* Факторный анализ. М., 1980.
5. *Герман О.В., Дорожжина Н.Н.* // Вестник Ставропольского университета. 1999. В. 20. С. 85–99.