УДК 519

ГИБРИДНЫЙ МЕТОД РАСПОЗНАВАНИЯ НА ОСНОВЕ ОПТИМАЛЬНЫХ ТУПИКОВЫХ НЕЧЕТКИХ ТЕСТОВ И РЕКУРРЕНТНОГО МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

И.В. Котельников, Ю.И. Неймарк, Л.Г. Телкина

Научно-исследовательский институт прикладной математики и кибернетики,
Нижегородский государственный университет им. Н.И. Лобачевского,
ул. Ульянова, 10, г. Нижний Новгород, Россия, 603005
e-mail: *neymark@pmk.unn.runnet.ru*

Abstract

The hybrid method of recognition based on a combination of logic and statistical approaches to the data analysis and focused on solving recognition problems for multidimensional dynamically changing objects and problems with any size, including small, of learning sample is offered. Method is based on adaptive formation of the new description sufficient for construction of a reliable decision rule on the submitted sample of the data. The new description is formed by means of the clusters or syndromes construction with use of the optimal irreducible fuzzy tests and their coding with help of the least squares method.

Постановка задачи

В настоящее время существует обширное множество различных методов распознавания, что, с одной стороны, затрудняет их практическое использование и ограничивает возможности исследования по созданию принципиально новых методов, а, с другой стороны, порождает интересные как теоретические, так и практические задачи их исследования, анализа и применения. *Одним из таких современных направлений исследований в распознавании образов является разработка гибридных алгоритмов, сочетающих в себе идеи и возможности нескольких известных методов распознавания.* В качестве примера можно сослаться на работы В.В. Рязанова с соавторами по склеиванию нейросетевого и комбинаторного подхода для решения задач распознавания с большими обучающими выборками [1]. Результаты исследований в этом направлении излагаются и в предлагаемой вашему вниманию статье.

Ставится задача разработки гибридного метода, ориентированного на *решение обучающего класса задач распознавания при любых, в том числе и небольших объемах обучающей выборки, включая и актуальные задачи распознавания многомерных динамически изменяющихся объектов,* и основанного на формировании нового описания, достаточного для построения достоверного решающего правила на представленной выборке данных. Для решения поставленной задачи предлагается использовать два дополняющих друг друга подхода к анализу данных, развиваемых авторами в последние годы, а именно: логический, учитывающий конкретные особенности всех анализируемых множеств, и статистический, опирающийся на среднестатистическое оценивание характеристик обучающего множества.

Используемый метод анализа данных основаны на применении *уникального математического аппарата, разработанного или расширенного и модифицированного авторами* и хорошо зарекомендовавшего себя в решении множества проблем, возникающих в задачах распознавания. Логический анализ выполняется с использованием аппарата *оптимальных тупиковых нечетких тестов* (ОТНТ) [2], [3], а статистический анализ проводится путем постановки и решения проблем с помощью *универсальной рекуррентной формы метода наименьших квадратов* (МНК) [4].

Процесс решения задачи распознавания с помощью предлагаемого гибридного метода складывается из трех этапов:

1) анализ обучающих данных и возможностей сокращения исходного описания;
2) формирование нового описания;
3) решение задач распознавания на новом описании.

В данной работе акцент сделан на решении важных первых двух этапов с использованием тестового подхода на первом этапе решения задачи распознавания и рекуррентного МНК - на втором этапе ее решения. После этого на третьем этапе могут быть построены различные решающие правила, в том числе и обладающие адаптивными возможностями правила на базе ОТНТ [2] и рекуррентного (МНК) [4], с последующим принятием коллективного решения [4].

1. Метод оптимальных тупиковых нечетких тестов и синдромов в процедуре формирования новых информативных признаков описания объектов

Тупиковые нечеткие тесты (ТНТ) являются естественным расширением понятия двоичных тестов [5] на случай, когда признаки объектов могут принимать не только двоичные, из множества $\{0,1\}$, но любые значения из их непрерывного ряда на вещественной оси. За это важное для практического применения расширение пришлось заплатить получением таких значений мощности полных множеств ТНТ, которые практически не реализуемы для подавляющего числа реальных задач распознавания. Выход из положения был найден выделением оптимального подмножества ТНТ (ОТНТ) малой мощности [2], [3], соответствующего важному условию максимально доступного на выборке различимости объектов разных классов, и разработкой алгоритма их построения, минуя недостижимое построение полного множества тестов.

В n-мерном пространстве признаков тесту соответствует k-мерный прямоугольный параллелепипед P^k , $k \leqslant n$, с замечательным свойством содержать в своем внутреннем объеме, исключая поверхность, точки объектов какого-то одного и только этого класса.

Под синдромом понимается сочетание из $k,\ k \leqslant n$, признаков описания объектов с известными границами их изменчивости

$$\{c_{i_\nu} \leqslant x_{i_\nu} \leqslant d_{i_\nu}\ ,\ \nu = \overline{1,n},\ i_\nu \in \{1,2,\ldots,n\}\},$$

которым удовлетворяет группа объектов какого-то одного и только этого класса. В n-мерном пространстве признаков синдрому соответствует k-мерный прямоугольный параллелепипед P^k, во внутренней области и на поверхности которого располагаются точки объектов какого-то одного и только этого класса, удовлетворяющего определению синдрома.

Из сравнения геометрических интерпретаций теста и синдрома следует их практическая адекватность. Благодаря этому построив тест и определив множество объектов в его внутренней области, для этого множества объектов очень легко построить синдром. Для этого необходимо лишь определить интервалы, минимальным образом покрывающие выделенную тестом группу объектов по каждому из k тестовых признаков.

Оптимальным тестом соответствуют синдромы (ОС). Выделенные ОТНТ и ОС обладают следующими оптимальными характеристиками и свойствами [2], [3]:

- Минимально допустимое на выборке число признаков их описания, что обусловлено построением ОТНТ только минимальной длины.
- Максимально допустимая на выборке различимость объектов различных классов. Это - основное свойство, по которому выделяется подмножество ОТНТ.
- Многомерный характер отношений сходства-различимости объектов выборки на минимально допустимом числе признаков против одномерного характера этих отношений по значениям функции расстояния между объектами по полному набору признаков. Это позволяет поднять на более высокий качественный уровень оценку сходства-различимости объектов анализируемой выборки.
- Максимально допустимая на выборке мощности множества объектов, принадлежащих отдельным синдромам, что, как следствие, приводит к минимуму априорной вероятности ошибки для объектов принадлежать классу объектов с рассматриваемым синдромом.
- Возможность для отдельных признаков объектов иметь неопределенные знания (пропуски). Методом ОТНТ и ОС успешно решались задачи, в которых ни один из объектов не имел описания полным набором признаков.
- Независимость получаемых синдромов от масштаба измерения признаков - важный фактор получения единственности решения задач анализа данных и учета полного влияния каждого из признаков на результаты анализа данных.
- Прозрачное для представления практически всех предметных областей представление результатов анализа данных через ОС в значениях исходных признаков объектов, или в терминах рассматриваемой предметной области.

При кодировании выборок многомерных объектов выделенное небольшое число ОС принимается за новые признаки описания объектов. Каждая n-мерная комбинация признаков(объект или состояние объекта в определенный момент времени) заменяется порядковым номером выделенного ОС, которому она принадлежит. С одной стороны, это уже в n раз сокращает описание анализируемой выборки объектов, что при больших n довольно существенно. С другой стороны, множество n-мерных состояний объектов в исходной выборке часто очень большой мощности заменяется небольшим фиксированным числом возможных состояний в виде выделенных ОС, общим для всех объектов выборки. Такая замена по существу тоже эквивалентна существенному сжатию описания объектов выборки. Номера ОС в описании объектов являются по существу их фиктивными именами, количественное значение различимости между которыми в общем случае не отражает реального значения различимости между объектами. Для метода ОТНТ и ОС это затруднение легко преодолевается заменой различимости между номерами реальными значениями различимости между синдромами. Иной путь формирования нового описания — численное кодирование кластеров или синдромов, которое может быть реализовано с помощью МНК.

Такое описание дает возможность использовать любые методы анализа данных и построения решающих правил. Этот путь предпочтителен при распознавании динамически изменяющихся объектов для построения адаптивного описания с помощью математического моделирования динамики изменения объекта и сведения динамической задачи к классической статической задаче распознавания с учителем [6].

2. Оптимальное кодирование, согласованное с МНК

Пусть все множество данных X, состоящее из N объектов, удалось описать с помощью K кластеров (синдромов), каждый из которых описывает некоторое подмножество объектов

$$X^s = \{x_j^s / j = 1, 2, \dots, N_s\},\ s = 1, 2, \dots, K,$$

причем $\sum_{s=1}^{K} N_s = N_0 \geqslant N$ (для кластеров имеет место равенство). Ставится задача присвоения каждому из X^s такого числового кода b_s, чтобы в некотором спрямляющем пространстве базисных функций $\varphi_1(x), \varphi_2(x), \dots, \varphi_{m-1}(x), \varphi_m(x)$ минимизировать квадратичный функционал

$$J(a, b) = \sum_{s=1}^{K} \sum_{j=1}^{N_s} \left(\sum_{i=1}^{m} a_i \varphi_i(x_j^s) - b_s \right)^2$$

по $\mathbf{a} = (a_1, a_2, \dots, a_m)$ и $\mathbf{b} = (b_1, b_2, \dots, b_K)$ при условии нормировки $\sum_{\nu=1, \nu<\mu}^{K-1} (b_\nu - b_\mu)^2 = C$, выражающем максимизацию различимости кодов кодируемых кластеров (синдромов). Поставленная задача оптимизации геометрически интерпретируется как поиск в пространстве базисных функций некоторого направления максимальной различимости кодируемых кластеров (синдромов), задаваемого искомым вектором $\mathbf{a}$. При выбранном направлении коды кластеров (синдромов) определяются из условия

$$\frac{\partial J}{\partial b_s} = 0, \text{ т.е. } b_s = \frac{1}{N_s} \sum_{j=1}^{N_s} \sum_{i=1}^{m} a_i \varphi_i(x_j^s), \ s = 1, 2, \ldots, K,$$

а это означает, что коды кластеров (синдромов) представляют собой усредненные значения проекций объектов из этих кластеров (синдромов) на выбранное направление.

Такое одновременное кодирование, представленное в работе [7], не всегда достаточно для решения задачи распознавания. Для более полного отражения особенностей взаимного расположения кластеров (синдромов) может потребоваться их многомерное кодирование, необходимость которого определяется решением задачи распознавания, а именно: если при одномерном кодировании не удаляется решить задачу с требуемой степенью точности, описание кластеров (синдромов) уточняется путем введения двумерного кодирования с последующей коррекцией решающего правила с помощью адаптивных алгоритмов распознавания и т.д.

При многомерном кодировании кластеров (синдромов) после введения направлений $\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^{L-1}$ новое направление $\mathbf{a}^L$ и, соответственно, новые значения $\mathbf{b}^L$ ищутся из условия минимизации функционала

$$J(\mathbf{a}^L, \mathbf{b}^L) = \sum_{s=1}^{K} \sum_{j=1}^{N_s} \left(\sum_{i=1}^{m} a_i^L \varphi_i(x_j^s) - b_s^L \right)^2$$

при условиях

$$\sum_{\nu=1, \ \nu<\mu}^{K-1} (b_\nu - b_\mu)^2 = C \ \& \ a^k (a^L)^T = \sum_{i=1}^{m} a_i^k a_i^L = 0, k = 1, 2, \ldots, L-1,$$

включающихся в себя уже упоминавшееся выше условие нормировки для кодовых значений и дополнительные $L-1$ линейных ограничений, представляющих собой условия ортогональности выбираемого нового направления к уже введенным $L-1$ направлениям.

Если ввести матрицы
$F_s = \|\varphi_i(x_j^s)\|$ размерности $N_s \times m$ (матрицы значений функций $\varphi_1(x), \ldots, \varphi_m(x)$ на векторах из множеств $X^s, \ s = 1, \ldots, K$),
$F = \|F_1 \ \cdots \ F_K\|^T$ размерности $N_0 \times m$,
$G_{L-1} = \|a^1 \ \cdots \ a^{L-1}\|^T$ размерности $L-1 \times m$
и вектора $b_s = \|b_s^L \ \cdots \ b_s^L\|^T$ размерности $N_s \times 1$
и $b = \|b_1 \ \cdots \ b_K\|^T$ размерности $N_0 \times 1$,
то задача условной оптимизации (задача 1) может быть записана в виде:

$$\min_{a^L, \ b^L} J(a^L, b^L) = \min_{a^L, \ b^L} (Fa^L - b)^T (Fa^L - b) \qquad (1)$$

при условиях

$$G_{L-1}a^L = 0 \qquad \& \qquad \sum_{\nu=1,\nu<\mu}^{K-1} (b_{\nu-b_\mu})^2 = C.$$

Докажем утверждение, позволяющее найти решение задачи 1.

Предложение 1. Вектор оптимальных кодовых значений b^L есть решение однородной системы линейных уравнений - ого порядка $(Q - \lambda R)b^L = 0$, где $\lambda = \lambda_{min}$ — минимальный по величине множитель Лагранжа, обращающий в ноль определитель этой системы, а Q и R — квадратные симметричные неотрицательно определенные матрицы размерности $K \times K$ с элементами:

$$q_{ij} = \begin{cases} sum(-F_i\Gamma F_j^T), & \text{при } i \neq j \\ sum(-F_i\Gamma F_i^T) + N_i, & \text{при } i = j \end{cases} \text{ и } r_{ij} = \begin{cases} k-1, & \text{при } i = j \\ -1, & \text{при } i \neq j \end{cases},$$

где через $sum(-F_i\Gamma F_j^T)$ обозначается сумма элементов матрицы $(-F_i\Gamma F_j^T)$, а $\Gamma = F^TF - P^{-1}G_{L-1}^T(G_{L-1}P^{-1}G_{L-1}^T)^{-1}G_{L-1}P^{-1}$. Направление максимальной различимости кодов $a^L = \Gamma F^T b$.

Доказательство. При заданном b решение задачи оптимизации функционала

$$\min_{a^L} J(a^L) = \min_{a^L} (Fa^L - b)^T(Fa^L - b) \tag{2}$$

при условии $\quad G_{L-1}a^L = 0$
может быть легко найдено и записано в виде $a^L(b) = \Gamma F^T b$, где $\Gamma = P^{-1} - D$, а $P = F^TF$ и $D = P^{-1}G_{L-1}^T(G_{L-1}P^{-1}G_{L-1}^T)^{-1}G_{L-1}P^{-1}$ [4] (условия существования решения — линейная независимость базисных функций на множестве X и введенный $L-1$ направлений). Если представить в $J(a^L, b^L)$ полученное выражение для $a^L(b)$, то от задачи 1 через решение задачи 2 перейдем к задаче 3:

$$\min_{b^L} \bar{J}(b^L) = \min_{b^L} b^T(F\Gamma F^T - E)^T(F\Gamma F^T - E)b = \min_{b^L} b^T H b \tag{3}$$

$$\text{при условии} \qquad \sum_{\nu=1,\nu<\mu}^{K-1} (b_{\nu-b_\mu})^2 = C,$$

где E — единичная матрица размерности $N_0 \times N_0$.

Отметив, что $P = F^TF = \sum_{i=1}^K F_i^TF_i$ и $(F_i\Gamma F_j^T)^T = F_j\Gamma F_i^T$, т.к. Γ — матрица симметричная, рассмотрим матрицу H, состоящую из $K \times K$ блоков H_{ij} вида:

$$H_{ij} = \sum_{\nu \neq i,j} (F_i\Gamma F_\nu^T)(F_\nu\Gamma F_j^T) + (F_i\Gamma F_i^T - E_i)F_i\Gamma F_j^T +$$

$$+F_i\Gamma F_j^T(F_j\Gamma F_j^T - E_j) = \sum_{\nu=1}^{K} F_i\Gamma F_\nu^T F_\nu \Gamma F_j^T 2F_i\Gamma F_j^T =$$

$$= F_i\Gamma P\Gamma F_j^T - 2F_i\Gamma F_j^T = -F_i\Gamma F_j^T$$

при $i \neq j$, т.к. легко показать, учитывая выражения для матрицы Γ через матрицы P и G_{L-1}, что $\Gamma P \Gamma = \Gamma$. Аналогично при $i = j$

$$H_{ii} = \sum_{\nu \neq i} (F_i \Gamma F_\nu^T)(F_\nu \Gamma F_i^T) + (F_i \Gamma F_i^T - E_i)^2 =$$

$$= \sum_{\nu=1}^{K} F_i \Gamma F_\nu^T F_\nu \Gamma F_j^T - 2F_i \Gamma F_i^T + E_i = -F_i \Gamma F_i^T + E_i.$$

Таким образом, функционал

$$\bar{J}(b) = b^T H b = \sum_{i,j=1}^{K} b_i^T H_{ij} b_j = \sum_{i,j=1}^{K} b_i^L b_j^L \sum_{\mu=1}^{N_i} \sum_{\nu=1}^{N_j} h_{\mu\nu}^{ij},$$

где $h_{\mu\nu}^{ij}$ — элемент матрицы H_{ij} размерности $N_i \times N_j$. Поэтому, если обозначить через q_{ij} сумму всех элементов матрицы H_{ij}, то

$$\bar{J}(b) = \sum_{i,j=1}^{K} b_i^L q_{ij} b_j^L = (b^L)^T Q b^L = I(b^L),$$

где b^L — вектор искомых кодовых значений, а Q — матрица размерности $K \times K$. И т.к. $H_{ij}^T = H_{ji}$, то $q_{ij} = q_{ji}$, т.е. Q — квадратная симметричная полуположительно определенная матрица с элементами

$$q_{ij} = \left\{ \begin{array}{rl} sum(-F_i \Gamma F_j^T), & \text{при } i \neq j \\ sum(-F_i \Gamma F_i^T) + N_i, & \text{при } i = j \end{array} \right. .$$

Если условие нормировки в задаче 3 представить в виде
$(b^L)^T R b^L = C (R$ — матрица с элементами $r_{ij})$, то задачу 3 можно записать в виде:

$$\min_{b^L} I(b^L) = \min_{b^L} (b^L)^T Q b^L$$

при условии

$$(b^L)^T R b^L = C, \quad \text{где} \quad r_{ij} = \left\{ \begin{array}{rl} k-1, & \text{при } i = j \\ -1, & \text{при } i \neq j \end{array} \right. .$$

Решение последней задачи условной оптимизации с помощью метода множителей Лагранжа приводит к однородной системе линейных уравнений $(Q - \lambda R) b^L = 0$, при решении которой в качестве множителя Лагранжа λ выбирается наименьший положительный корень, обращающий в ноль определитель этой однородной системы, т.к.

$$I = (b^L)^T Q b^L = \lambda (b^L)^T R b^L = \lambda C.$$

Из однородности системы также следует, что значение кодов $b_1^L, b_2^L, \ldots, b_K^L$ изменяются пропорционально $\sqrt{C}$ при изменении величины константы в условии нормировки, а, следовательно, также изменяются и координаты вектора a^L, определяющего искомое направление, т.е. выбор направления определяется выбором условия нормировки, но не зависит от выбора константы C в этом условии, которая, без ограничения общности, может быть принята за 1.

Заметим,что все решения задачи по оптимальному кодированию кластеров (синдромов) может быть проведено с использованием универсальной рекуррентной формы МНК, включая поиск минимального по величине корня $\lambda_{\min}$, при котором определитель однородной системы обращается в ноль. Последняя задача может быть легко сведена к задаче поиска минимального собственного значения некоторой преобразованной матрицы, которая, в свою очередь, легко решается с помощью МНК [4].

Заключение

Предложенный гибридный метод распознавания сочетает в себе логический анализ выборки, учитывающий конкретные особенности всех данных, и статистическое кодирование тех закономерностей, что выявлены с помощью логического анализа. Особенностью предлагаемого подхода является рекуррентное увеличение размерности нового описания, т.е. уточнение его до тех пор, пока оно не окажется достаточным для построения решающего правила распознавания требуемой эффективности. *Такой подход является новым и позволяет добиваться нужного качества распознавания на минимальном описании.*

Предложенный метод может быть использован для решения ряда актуальных задач распознавания: задач с малыми выборками данных, распознавание динамически изменяющихся объектов [7]*, оценивание состояния динамического объекта, слежение за динамически изменяющимся объектом* [8]*, а также в задачах автоматического интеллектуального анализа данных (data mining).*

Работа выполнена при финансовой поддержке РФФИ (проект 02-01-00274) и ФЦП «Интеграция» (госконтракт №Б-0039/2102).

Список литературы

1. *Ryazanov V.V., Senko O.V., Aslanyan L., Saakyan Kh., Mingo L.Ph.* On Certain Hybrid Classifikation Models Based on Neural Network and Logical Approaches // Pattern Recognition and Image Analysis, Vol.11, 2001. Pp.80-82.
2. *Kotel'nikov I.V.* A Syndrome Recognstion Method Based on Optimal Irreducible Fuzzy Tests. // Pattern Recognition and Image Analasis, v.11., no.3, 2001. Pp. 553-559.
3. *Kotel'nikov I.V.* Cluster Analisis of Multimensional Objects Based on Optimal Irreducible Fuzzy Tests and Syndromes.// Pattern Recognition and Images Analisis, v.14, No.3, 2004, Pp. 361-369.
4. *Неймарк Ю.И., Теклина Л.Г.* Новые технологии применения метода наименьших квадратов. - Нижний Новгород. Изд. Нижнегорского госуниверситета, 2003. 196 с.
5. *Чегис И.А., Яблонский С.В.* Логические способы контроля работы электрических схем.// Труды математического института АН СССР им. В.А.Стеклова. М.: Наука, 1958. Т. 51, С. 269-360.
6. *Неймарк Ю.И., Теклина Л.Г.* Решение задачи распознавания многомерных временных рядов на основе редукции размерности // Сб. Математические методы распознавания образов. Доклады X Всероссийской конференции. М., 2001. С. 96-99.

7. *Неймарк Ю.И., Теклина Л.Г.* Роль кодирования образов при распознавании // Доклады РАН. Т363, №6, 1998. С.751-752.
8. *Kotel'nikov I.V., Неймарк Ю.И., Теклина Л.Г.* The analisis of the dynamic objects changes by the pattern recognition methods // Pattern Recognitions and Image Analisis: New Information Technologies. Confirence Proceedings. V.I. Pp.66-69.