



Е. С. Кирик

аспирант Института
вычислительного моделирования
СО РАН

научный руководитель
д. т. н., проф. А. В. Медведев

Ekaterina S. Kirik (Agarova)

УДК 62-506.1

О НЕПАРАМЕТРИЧЕСКОМ ПОДХОДЕ К ВОССТАНОВЛЕНИЮ РАЗДЕЛЯЮЩЕЙ ПОВЕРХНОСТИ В ЯВНОМ ВИДЕ В ЗАДАЧЕ РАСПОЗНАВАНИЯ ОБРАЗОВ

Copyright © 2000, Е. С. Кирик

В работе рассматривается вопрос о непараметрическом восстановлении разделяющей поверхности в явном виде в задаче распознавания образов. В отличие от предложенных ранее, данный способ позволяет получить оптимальную оценку в смысле минимума функции риска. Идея алгоритма состоит в выделении из начальной обучающей выборки элементов, лежащих непосредственно вблизи разделяющей поверхности, и последующем восстановлении последней по этим элементам, для чего используются методы непараметрического регрессионного анализа ■

Nonparametric algorithm of the separating surface restoration in a direct form in the pattern recognition problem

This work deals with problem of non parametrical dividing surface restoration in a direct form in probability statement of the pattern recognition problem. In difference to known suggested approach allows get an optimal asymptotically convergent estimation. The idea consists in extracting from the initial learning samples elements which are laying just near separating surfaces, and then in direct restoration of latter on these elements using the methods of nonparametric regression analysis ■

Введение

В общей постановке задача распознавания образов состоит в отнесении предъявляемого объекта к одному из классов. Классы характеризуются тем, что принадлежащие им объекты обладают некоторым сходством. Если удастся представить набор существенных для классификации признаков в виде вектора параметров $x = (x^1, \dots, x^k)$, то M классам отвечает набор M областей в k -мерном пространстве параметров.

Ясно, что если границы областей заданы в аналитическом виде, то для любого опознаваемого объекта легко определить его класс. Например, для двух смежных областей можно ввести *решающую функцию* $f(x)$, которая положительна в одной области, отрицательна в другой и обращается в ноль на *разделяющей поверхности*. Тогда знак $f(x)$ позволяет определить, к какому классу принадлежит объект, описываемый вектором x .

Таким образом, выделение разделяющих поверхности по существу решает задачу распознавания. Кроме того, в некоторых случаях вид этих поверхностей сам по себе представляет интерес. Это относится прежде всего к динамическим задачам, когда в каждый момент времени необходимо знать положение объекта по отношению к границе класса. Например, можно рассматривать задачу об удержании объекта в заданном классе (подмножестве параметрического пространства). В этом случае явное уравнение для разделяющей поверхности дает информацию, необходимую для управления объектом.

На практике границы областей, соответствующих классам, как правило неизвестны. Взамен имеется множество объектов (набор векторов x_i , $i = \overline{1, n}$ — т. н. *обучающая выборка*), для которых заранее известно, например, из эксперимента, к какому именно классу принадлежит каждый из них. В этом случае можно поставить задачу о приближенном восстановлении разделяющих поверхностей по расположению элементов обучающей выборки в параметрическом пространстве.

Можно выделить три группы методов решения данной задачи. При наличии дополнительной информации о вероятностных характеристиках классов применяется *байесовский подход*. Согласно этому подходу, решающая функция строится исходя из условия минимума т. н. функции риска [2]. Если имеется некоторая информация о виде разделяющей поверхности (например, что она может быть описана некоторой функцией с неизвестными значениями параметров), то можно использовать *параметрический подход*, где стоит вопрос о нахождении оптимальных параметров. Третий, *непараметрический подход* не требует наличия какой-либо информации, кроме самой обучающей выборки, и поэтому является наиболее универсальным. Однако в отличие от первых двух методов, непараметрический подход не позволяет записать уравнение разделяющей поверхности в явном виде, т. е. разрешить уравнение $f(x) = 0$ относительно одной из компонент вектора x .

Предлагаемый в данной статье метод позволяет снять последнее ограничение.

Задачи о явном непараметрическом восстановлении разделяющей поверхности уже рассматривались ранее [1]. Однако предложенный в [1] алгоритм не позволяет получить единственную (как это имеет место на самом деле в вероятностной постановке задачи) оценку разделяющей поверхности, которой соответствует минимальное количество ошибок классификации. В этом смысле последняя не является оптимальной. Кроме того, алгоритм достаточно сложен в вычислительном плане, особенно для задач большой размерности.

1. Постановка задачи. Пути решения

Прежде всего заметим, что достаточно уметь решать поставленную задачу для двух смежных классов, так как любую многоальтернативную задачу ($M > 2$) можно свести к решению последовательности двувальтернативных методом дихотомии. На каждом шаге процедуры выделяется один из классов, а все остальные рассматриваются в совокупности. Если опознаваемый объект не принадлежит выделенному классу, то последний исключается из рассмотрения и процедура продолжается с оставшимися классами.

Учитывая сказанное, далее всюду будем рассматривать задачу для двух классов.

Пусть дана обучающая выборка $V = \{\mathbf{x}_i, i = \overline{1, n}\}$: $\mathbf{x}_i = (x_i^1, \dots, x_i^k)$ — вектор признаков, причем $V = V_1 \cup V_2$, $V_1 = \{\mathbf{x}_i, i = \overline{1, n_1}\}$ — элементы, принадлежащие X_1 , $V_2 = \{\mathbf{x}_i, i = \overline{1, n_2}\}$ — элементы, принадлежащие X_2 , $n = n_1 + n_2$. Предполагается, что данные центрированы и нормированы. Задача состоит в восстановлении истинной разделяющей поверхности в явном виде по выборке V . Например, явный вид разделяющей поверхности можно представить как функцию $x_n^1 = x_n^1(x^2, \dots, x^k)$, тогда поставленная задача сводится к нахождению непараметрической оценки $x^1 - x_n^1 = x_n^1(x^2, \dots, x^k)$.

Решение этой задачи содержит два этапа. Первый заключается в выделении элементов обучающей выборки, лежащих вблизи к разделяющей поверхности (подвыборки), второй — это непосредственно восстановление разделяющей поверхности в явном виде по найденным элементам.

Рассмотрим последовательно решение каждой из этих подзадач. Сначала дадим обзор наиболее известных методов построения решающей функции.

1.1. Методы построения решающей функции

Как уже упоминалось, существуют различные методы решения задачи распознавания образов. Для того чтобы использовать тот или иной метод распознавания образов, необходимо располагать определенным количеством априорной информации.

1.1.1. Байесовские алгоритмы

Если известны вероятностные характеристики классов, то успешно применяются методы теории статистических решений. Здесь получены оптимальные байесовские алгоритмы распознавания образов, обеспечивающие минимум функции риска [2]

$$R = \int_{X_1} F_{11} P_1 p_1(\mathbf{x}) d\mathbf{x} + \int_{X_1} F_{21} P_2 p_2(\mathbf{x}) d\mathbf{x} + \int_{X_2} F_{12} P_1 p_1(\mathbf{x}) d\mathbf{x} + \int_{X_2} F_{22} P_2 p_2(\mathbf{x}) d\mathbf{x}, \quad (1.1)$$

где F_{km} — элементы т. н. платежной матрицы, характеризующие потери, возникающие при отнесении объекта \mathbf{x} из класса k к классу m ($k, m = \overline{1, 2}$), P_k — априорная вероятность появления нового объекта в классе k ($k = \overline{1, 2}$), $p_k(\mathbf{x})$ — условная плотность распределения элементов в классе k , X_1, X_2 — классы, на которые разбита исследуемая область X .

Минимуму функции риска R удовлетворяет байесовская решающая функция [2]

$$f_{12}(\mathbf{x}) = [F_{11} - F_{12}]P_1 p_1(\mathbf{x}) + [F_{21} - F_{22}]P_2 p_2(\mathbf{x}) = 0. \quad (1.2)$$

Уравнение (1.2) определяет, причем единственным образом, поверхность, разделяющую области X_1 и X_2 . Знак решающей функции $f_{12}(\mathbf{x})$ позволяет различать эти области. Поэтому оптимальное байесовское решающее правило можно сформулировать следующим образом:

$$\mathbf{x} \in X_1, \text{ если } f_{12}(\mathbf{x}) \leq 0; \quad \mathbf{x} \in X_2, \text{ если } f_{12}(\mathbf{x}) > 0. \quad (1.3)$$

При необходимости уравнение (1.2) может быть разрешено относительно одной из компонент вектора \mathbf{x} , например,

$$x^1 = f_{sm}^1(x^2, \dots, x^k),$$

и тогда разделяющая поверхность будет найдена в явном виде.

При условии, что вероятностные характеристики известны, данный подход позволяет точно находить уравнение разделяющей поверхности как неявно, так и в явном виде.

Когда вероятностные характеристики классов неизвестны, что наиболее характерно для практических задач, для построения решающей функции используют либо параметрический подход (если вид истинной разделяющей поверхности предполагается известным), либо непараметрический (где, кроме обучающей выборки, не требуется никакой дополнительной информации).

1.1.2. Параметрический подход

Параметрический подход заключается в том, что для оценивания решающей функции выбирается класс функций $\psi_i(\mathbf{x})$, $i = \overline{1, N}$ [2]. Вид искомой зависимости задается с точностью до набора параметров \mathbf{a} :

$$\tilde{f}_n(\mathbf{x}, \mathbf{a}) = \sum_{i=1}^N a_i \psi_i(\mathbf{x}). \tag{1.4}$$

Здесь $\psi_i(\mathbf{x})$, $i = \overline{1, N}$ — некоторая система базисных функций. Параметры a_i , $i = \overline{1, N}$ вычисляются с использованием обучающей выборки. Значения параметров каждого элемента выборки подставляются в оценку (1.4) при фиксированных значениях a_i , $i = \overline{1, N}$, и с помощью байесовского решающего правила определяется класс элемента, затем значение класса сравнивается с известным (истинным), подсчитывается суммарное количество ошибок классификации. Оптимальным считается тот набор параметров a_i , $i = \overline{1, N}$, который соответствует минимальному количеству ошибок.

Как и в предыдущем случае, при необходимости можно выписать уравнение разделяющей поверхности в явном виде, разрешив (1.4) относительно одного из признаков, например, $x^1 = \tilde{f}_n^1(x^2, \dots, x^k)$.

Основной проблемой параметрического подхода является выбор базисных функций $\psi_i(\mathbf{x})$, $i = \overline{1, N}$, правильность которого непосредственно влияет на качество оценки (1.4). Этот выбор может быть сделан только на основе априорной информации о возможном виде разделяющей поверхности, которая нередко является недоступной. Существуют также проблемы вычислительного плана, если число базисных функций достаточно велико, особенно это касается задач большой размерности. Если все же размерность не велика, и достаточно данных, чтобы задать степень полинома и ограничить область поиска значений a_i , $i = \overline{1, N}$, то довольно быстро можно получить хорошие оценки.

1.1.3. Непараметрический подход

В рамках непараметрического подхода в выражении для оптимальной байесовской решающей функции условные плотности $p_k(\mathbf{x})$, $k = \overline{1, 2}$ заменяются их асимптотически сходящимися [1] непараметрическими оценками. Большой популярностью пользуются т. н. оценки плотности ядерного типа, введенные Розенблатом и Парзенем [5],

$$p_{n_i}(\mathbf{x}) = \frac{1}{n_i} \sum_{m=1}^{n_i} \frac{1}{C_{n_i}} \prod_{j=1}^k \Phi\left(\frac{x^j - x_m^j}{C_{n_i}}\right), \tag{1.5}$$

$p_{n_i}(\mathbf{x})$, $i = \overline{1, 2}$ — оценка плотности распределения элементов в i -м классе.

Здесь $\Phi(\cdot)$ — ядро — финитная колоколообразная интегрируемая с квадратом функция, удовлетворяющая условиям [1]

$$0 < \Phi(z) < \infty \quad \forall z \in \Omega(z), \quad \frac{1}{C_n} \int_{\Omega(x)} \Phi\left(\frac{x - x_i}{C_n}\right) dx = 1, \quad \lim_{n \rightarrow \infty} \frac{1}{C_n} \Phi\left(\frac{x - x_i}{C_n}\right) = \delta(x - x_i), \tag{1.6}$$

C_n — т. н. *параметры размытости*, определяющие размер носителя и “дельтаобразность” ядра $\Phi(\cdot)$.

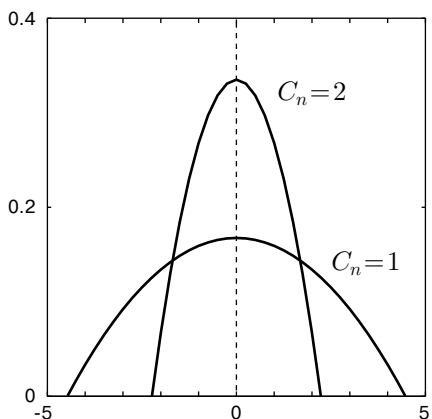


Рис. 1

В приложениях используются колоколообразные функции $\Phi(\cdot)$ различных видов [3, 4, 9], например,

$$\Phi(z) = \begin{cases} 0.335 - 0.067(z)^2, & |z| \leq \sqrt{5}, \\ 0, & |z| > \sqrt{5}; \end{cases} \tag{1.7}$$

$$\Phi(z) = \begin{cases} 0.05968(e^z + e^{-z}) - 0.154293z^2 + 0.2311, & |z| \leq 2.6, \\ 0, & |z| > 2.6; \end{cases}$$

$$\Phi(z) = \begin{cases} 0.5, & |z| \leq 1, \\ 0, & |z| > 1. \end{cases}$$

На рисунке 1 представлена функция $\Phi(x/C_n)/C_n$ (1.7), построенная для двух значений параметра размытости: $C_n = 1$, $C_n = 2$.

Параметры размытости C_{n_i} , $i = 1, 2$ удовлетворяют следующим условиям:

$$C_{n_i} > 0, \quad \lim_{n_i \rightarrow \infty} C_{n_i} = 0, \quad \lim_{n_i \rightarrow \infty} n_i (C_{n_i})^k = \infty. \quad (1.8)$$

Строго говоря, если рассматривается k -мерная задача, то параметры размытости C_{n_i} являются векторами соответствующей размерности: $C_{n_i} = (C_{n_i}^1, \dots, C_{n_i}^k)$. Предварительное центрирование и нормирование данных позволяет уйти от этой сложности и считать C_{n_i} скалярами. Последнее в дальнейшем и подразумевается.

Положив в (1.2), что элементы платежной матрицы F_{11}, F_{22} равны нулю, заменив условные плотности их непараметрическими оценками, а вероятности P_1 и P_2 на n_1/n и n_2/n соответственно, получим сходящуюся непараметрическую оценку следующего вида:

$$\varphi_n(\mathbf{x}) = \frac{F_{21}}{n_2} \sum_{i=1}^{n_2} \frac{1}{(C_{n_2})^k} \prod_{j=1}^k \Phi\left(\frac{x^j - x_i^j}{C_{n_2}}\right) - \frac{F_{12}}{n_1} \sum_{i=1}^{n_1} \frac{1}{(C_{n_1})^k} \prod_{j=1}^k \Phi\left(\frac{x^j - x_i^j}{C_{n_1}}\right), \quad (1.9)$$

где $\mathbf{x}_i = (x_i^1, \dots, x_i^k)$ — элементы обучающей выборки $V = \{\mathbf{x}_i\}$, $i = \overline{1, n}$, причем $V = V_1 \cup V_2$, $V_1 = \{\mathbf{x}_i\}$, $i = \overline{1, n_1}$ — элементы, принадлежащие X_1 , $V_2 = \{\mathbf{x}_i\}$, $i = \overline{1, n_2}$ — элементы, принадлежащие X_2 .

Если опознаваемый объект лежит далеко от разделяющей поверхности, глубоко внутри одного из классов класса, то одно из слагаемых (1.2) равно нулю, так как в окрестности этого объекта встречаются представители только этого класса. Вблизи разделяющей поверхности встречаются элементы обоих классов, и оба слагаемых (1.2) отличны от нуля.

Знак оценки $\varphi_n(\mathbf{x})$ позволяет различать рассматриваемые области (классы), решающее правило формулируется по аналогии с байесовским [2]:

$$\mathbf{x} \in X_1, \quad \text{если } \varphi_n(\mathbf{x}) \leq 0, \quad \mathbf{x} \in X_2, \quad \text{если } \varphi_n(\mathbf{x}) > 0. \quad (1.10)$$

В оценке (1.2) неизвестными являются параметры размытости C_{n_i} . Они находятся в ходе скользящего экзамена, который заключается в следующем. При фиксированных C_{n_i} , $i = 1, 2$ из обучающей выборки по очереди исключается каждый элемент, и на оставшихся определяется его класс при помощи алгоритма (1.2), (1.4), затем результат сравнивается с действительным значением класса этого элемента, и подсчитывается суммарное количество допущенных ошибок для заданных значений параметров размытости. Эта процедура проводится для различных значений C_{n_i} , $i = 1, 2$. Оптимальной считается пара, для которой допускается минимальное количество ошибок.

Непараметрический подход хорошо зарекомендовал себя в решении задач распознавания образов, так как для его применения требуется наименьшее количество априорной информации. Напомним, что в отличие от байесовской и параметрической решающих функций, непараметрическую оценку (1.2) нельзя разрешить относительно какой-нибудь из компонент \mathbf{x} . В результате оказывается невозможным получить явное уравнение для разделяющей поверхности.

1.2. Идея построения алгоритма

Поверхность, разделяющая классы, представляет собой некоторую неизвестную зависимость в пространстве параметров объектов $x^1 = x^1(x^2, \dots, x^k)$. Будем предполагать, что эта зависимость однозначна, т. е. каждому набору x^2, \dots, x^k соответствует только одно значение x^1 . Задача о восстановлении разделяющей поверхности в явном виде сводится к нахождению непараметрической оценки $x^1 = x_n^1(x^2, \dots, x^k)$.

Для восстановления однозначной зависимости $a(\mathbf{b})$ используют непараметрическую оценку регрессии (условного математического ожидания) — $\widehat{M}\{a|\mathbf{b}\}$. Для этого необходимо обладать набором независимых статистических наблюдений искомой функции — обучающей выборкой $\{a_i, \mathbf{b}_i\}$, $i = \overline{1, n}$ (допускается, что наблюдения имеют помехи).

Заменим в формуле условного математического ожидания $M\{a|b\} = \int_{\Omega(a)} a p(a, b) / p(b) da$ плотности вероятности их непараметрическими оценками типа (1.5). Используя свойство ядра $\Phi(\cdot)$

$$\frac{1}{C_n} \int_{\Omega(a)} a \Phi\left(\frac{a - a_i}{C_n}\right) da = a_i, \quad i = \overline{1, n},$$

получим сходящуюся непараметрическую оценку регрессии [1]

$$a_n(x) = \frac{\sum_{i=1}^n a_i \prod_{j=1}^k \Phi\left(\frac{b^j - b_i^j}{C_n}\right)}{\sum_{i=1}^n \prod_{j=1}^k \Phi\left(\frac{b^j - b_i^j}{C_n}\right)}. \quad (1.11)$$

Параметр размытости C_n определяется из условия минимума квадратичного критерия оптимальности в ходе скользящего экзамена на обучающей выборке:

$$\omega^2(C_n) = \frac{1}{n} \sum_{i=1}^n (a_i - a_n(b_i))^2 \rightarrow \min_{C_n}. \tag{1.12}$$

Подходить к восстановлению x^1 напрямую как к оцениванию $M\{x^1|(x^2, \dots, x^k)\}$ нельзя. Чтобы воспользоваться оценкой (1.11), нужны измерения (пусть даже сделанные с ошибкой) искомой функции. В данном случае таких измерений нет. Имеется лишь обучающая выборка V , элементы которой разбросаны по всей исследуемой области X . Информация о том, какие из элементов лежат непосредственно вблизи разделяющей поверхности, отсутствует. Если не принимать во внимание эту особенность обучающей выборки, то зависимость, восстановленная с помощью алгоритма (1.11), (1.12), будет представлять собой усреднение по всей V .

Будем называть совокупность элементов, лежащих в окрестности разделяющей поверхности, *подвыборкой* $V^0 = \{x_i\}$, $i = \overline{1, n^0}$ выборки V , $V^0 \subset V$. Если решена задача выделения V^0 , то восстановление поверхности сводится к оцениванию следующего условного математического ожидания: $M\{x^1|(x^2, \dots, x^k), x \in V^0\}$ ([6], [7]).

Возникает вопрос, как выделить из всего облака точек выборки V часть V^0 , необходимую для решения задачи. Источником информации о расположении элементов выборки в пространстве параметров могут стать непараметрические оценки решающей функции.

Преобразуем оценку (1.2) следующим образом. Умножим слагаемые в (1.2) на $n_1(C_{n_1})^k > 0$, $n_2(C_{n_2})^k > 0$ соответственно и разделим на $F_{21} \sum_{i=1}^{n_2} \prod_{j=1}^k \Phi\left(\frac{x^j - x_i^j}{C_{n_2}}\right) + F_{12} \sum_{i=1}^{n_1} \prod_{j=1}^k \Phi\left(\frac{x^j - x_i^j}{C_{n_1}}\right) > 0$. Получим оценку следующего вида:

$$\tilde{\varphi}_n(x) = \frac{F_{21} \sum_{i=1}^{n_2} \prod_{j=1}^k \Phi\left(\frac{x^j - x_i^j}{C_{n_2}}\right) - F_{12} \sum_{i=1}^{n_1} \prod_{j=1}^k \Phi\left(\frac{x^j - x_i^j}{C_{n_1}}\right)}{F_{21} \sum_{i=1}^{n_2} \prod_{j=1}^k \Phi\left(\frac{x^j - x_i^j}{C_{n_2}}\right) + F_{12} \sum_{i=1}^{n_1} \prod_{j=1}^k \Phi\left(\frac{x^j - x_i^j}{C_{n_1}}\right)}. \tag{1.13}$$

Если элемент находится внутри класса, то эта зависимость имеет тот же знак, что и (1.2). Поэтому в смысле определения класса, к которому принадлежит x , нет существенной разницы, какую из оценок использовать: (1.2) или (1.13), но с точки зрения построения искомого алгоритма предпочтительной оказывается оценка решающей функции (1.13).

Дело в том, что в отличие от оценки (1.2), которая может принимать значения от $-n_1(C_{n_1})^k \max(p_{n_1}(x))$ до $n_2(C_{n_2})^k \max(p_{n_2}(x))$ (где $p_{n_1}(x)$ и $p_{n_2}(x)$ — непараметрические оценки плотности первого и второго классов соответственно), оценка (1.13) нормирована, и область ее значений постоянна и не зависит от выборки (интервал $[-1, 1]$). Зная, чему равно значение $\tilde{\varphi}_n(x)$, о каждом элементе x можно сделать вывод о его расположении относительно разделяющей поверхности на основе следующих рассуждений.

Поскольку $\Phi(\cdot)$ — финитная функция, то равенство $|\tilde{\varphi}_n(x)| = 1$ означает отсутствие “поблизости” представителей другого класса, (слагаемые числителя и знаменателя в (1.13), соответствующие другому классу, равны нулю) и, соответственно, периферийное положение элемента x относительно разделяющей поверхности. Если $\tilde{\varphi}_n(x) \in (-1, 1)$, то колоколообразная функция $\Phi(\cdot)$ покрывает элементы, принадлежащие как первому классу так и второму. В этом случае оба слагаемых числителя и знаменателя в (1.13) отличны от нуля. Следовательно, экзаменуемый элемент выборки x лежит в окрестности разделяющей поверхности.

Равенство $\tilde{\varphi}_n(x) = 0$ можно трактовать двояко. Оно может означать как то, что в силу финитности $\Phi(\cdot)$ в окрестности x , определяемой оптимальным параметром размытости, не оказалось ни одного элемента из обучающей выборки, так и то, что элемент x лежит на разделяющей поверхности. Поэтому можно считать элементами выборки V^0 те $x \in V$, для которых выполняется условие $\tilde{\varphi}_n(x) \in (-1, 0) \cup (0, 1)$.

Таким образом, задача восстановления разделяющей поверхности в явном виде состоит из двух этапов. Во-первых, путем скользящего экзамена на начальной обучающей выборке решается задача распознавания образов (1.13), (1.4). В результате каждому элементу из выборки ставится в соответствие значение оценки решающей функции (1.13), вычисленной на этом шаге, то есть формируется выборка $V^+ = \{x_i, \tilde{\varphi}_n(x_i)\}$, $i = \overline{1, n}$, $V^+ = V_1^+ \cup V_2^+$. На втором этапе с использованием полученной информации строится непараметрическая оценка условного математического ожидания $M\{x^1|(x^2, \dots, x^k), x \in V^0\}$, которая является оценкой разделяющей поверхности в задаче распознавания образов.

1.3. Другие методы выделения подвыборки

Вопрос о выделении точек, лежащих в непосредственной близости от разделяющей поверхности изучается достаточно давно. Впервые эта задача была поставлена в связи с проблемой большого объема обучающей выборки. При больших n задача распознавания образов становилась “неподъемной” в вычислительном плане (особенно для больших размерностей). Возникла необходимость в выделении наиболее информативных с точки зрения качественной классификации элементов выборки. Лежащие вблизи разделяющей поверхности элементы как раз таковыми и являются. Вероятно, первая работа по этой тематике была опубликована Хартом в 1968 году [8].

В работе [8] было предложено использовать правило m ближайших соседей, которое заключается в том, что плотности в (1.2) заменяются оценками типа

$$p_{n_i}(\mathbf{x}) = \frac{m-1}{n_i W(\mathbf{x}, m, n)}, \quad i = \overline{1, 2}, \quad (1.14)$$

где $W(\mathbf{x}, m, n)$ — k -мерный шар с центром в \mathbf{x} , содержащий m ближайших к \mathbf{x} точек из i -го класса. радиус шара равен расстоянию до максимально удаленного элемента из m ближайших. В качестве меры “близости” точек обычно используется евклидово расстояние. Если все m “соседей” тестируемого элемента оказываются из одного класса, к тому же классу причисляется и рассматриваемый элемент.

Этот способ оценивания плотности является так же непараметрическим. Он близок по основной идее к методу построения оценок типа Розенבלата — Парзена [9, 10]. В последнем также выбираются несколько точек ближайших к x с весом, зависящим от формы ядра. Если взять прямоугольное ядро (с постоянным весом), то обе оценки совпадут.

Качество получаемой выборки V^0 в первом и во втором случае с точки зрения последующего явного восстановления разделяющей поверхности определяется по оптимальной величине байесовского риска. Если последняя мала, что означает слабое перекрытие классов на границе, то алгоритмы работают так, что к V^0 причисляются элементы, лежащие непосредственно вблизи границы классов, а точки, лежащие в глубине классов, в подвыборку не попадают. Если наоборот — оптимальный байесовский риск высок, то V^0 представляет собой более обширное облако элементов, так как в V^0 попадают не только те элементы V , что лежат непосредственно вблизи восстанавливаемой поверхности, но и достаточно удаленные от нее. Понятно, что восстановленная по V^0 разделяющая поверхность в этом случае даст более грубое приближение истинной зависимости.

Вообще говоря, нет необходимости вычислять значения байесовского риска при различных значениях параметров C_{n_i} , $i = 1, 2$ или m соответственно, чтобы определить качество V^0 в том и другом случае. Косвенным показателем оптимального значения последнего является минимальное количество ошибок (в процентном соотношении), допущенное при скользящем экзамене на обучающей выборке.

Предлагаемый подход к формированию V^0 позволяет найти выход в подобной ситуации. Например, можно считать элемент x лежащим на периферии, если $\tilde{\varphi}_n(\mathbf{x}) \in [-1, -\varepsilon] \cup [0] \cup [\varepsilon, 1]$, и наоборот — вблизи разделяющей поверхности, если $\tilde{\varphi}_n(\mathbf{x}) \in (-\varepsilon, 0) \cup (0, \varepsilon)$, где $0 < |\varepsilon| \leq 1$.

Смысл этих ограничений состоит в том, что согласно байесовской решающей функции (1.2), модифицированной оценкой которой является (1.13), и байесовскому решающему правилу, чем ближе расположен элемент выборки к разделяющей поверхности, тем меньше по модулю значение решающей функции. При этом, как уже отмечалось в предыдущем пункте, равенство нулю непараметрической оценки $\tilde{\varphi}_n(\mathbf{x})$ решающей функции (1.2) не несет в себе однозначную информацию, и поэтому такие элементы выборки не рассматриваются.

Величина ε зависит от w — минимального количества ошибок (в процентном соотношении), допущенных при скользящем экзамене и является убывающей функцией от w .

В [8] вопрос об улучшении качества выборки V^0 не затрагивается. Применить описанную выше процедуру к выборке V^0 , полученной с помощью метода m ближайших соседей, не представляется возможным. Так как никакими преобразованиями нельзя добиться того, чтобы все значения оценок решающих функций попадали в интервал $[-1, 1]$, то и применить описанную выше процедуру к выборке V^0 , полученной с помощью метода m ближайших соседей, не представляется возможным.

Позднее в работе [11] была предпринята попытка модифицировать алгоритм Харта. Смысл сводился к сокращению объема V^0 , однако не в смысле выделения элементов, сконцентрированных непосредственно вблизи разделяющей поверхности из числа уже вошедших в V^0 . Поэтому и этот способ выделения последней нельзя использовать для решения задачи.

В работах [12–16] тоже излагаются методы выделения V^0 из начальной обучающей выборки, в основе которых также лежит правило m ближайших соседей. Однако вопрос о явном восстановлении разделяющей поверхности на основе выборки V^0 не обсуждается ни в одной из работ.

Другой принцип формирования V^0 описан в [1]. В этой работе предлагается считать элементами V^0 те, что удовлетворяют условию $|(\varphi_n(\mathbf{x}) - 0)/C_n^\varphi| \leq a$, где $\varphi_n(\mathbf{x})$ — оценка (1.4). Смысл этого условия также состоит в том, что чем ближе расположен элемент выборки к разделяющей поверхности, тем меньше по модулю значение решающей функции. С использованием оценки регрессии восстанавливается x^1 в s точках исследуемого пространства

$$\hat{x}_n^1(x^2, \dots, x^k) = \frac{\sum_{i=1}^n x_i^1 \prod_{j=2}^k \Phi\left(\frac{x^j - x_i^j}{C_n^\varphi}\right) I_1\left(\frac{\varphi_n(\mathbf{x}_i)}{C_n^\varphi}\right)}{\sum_{i=1}^n \prod_{j=2}^k \Phi\left(\frac{x^j - x_i^j}{C_n^\varphi}\right) I_1\left(\frac{\varphi_n(\mathbf{x}_i)}{C_n^\varphi}\right)}. \tag{1.15}$$

Здесь I_1 — индикаторная функция, селектирующая элементы выборки

$$I_1\left(\frac{\varphi_n(\mathbf{x})}{C_n^\varphi}\right) = \begin{cases} 1, & |\varphi_n(\mathbf{x})/C_n^\varphi| \leq a, \tilde{\varphi}_n(\mathbf{x}) \neq 0; \\ 0, & |\varphi_n(\mathbf{x})/C_n^\varphi| > a, \tilde{\varphi}_n(\mathbf{x}) \neq 0 \text{ или } \varphi_n(\mathbf{x}) = 0. \end{cases} \tag{1.16}$$

Таким образом формируется выборка $V_s = \{\hat{x}_n^1(x_i^2, \dots, x_i^k), (x_i^2, \dots, x_i^k)\}$, и затем по этой выборке оценка разделяющей поверхности параметризуется, то есть ищется в виде полинома с некоторыми коэффициентами.

В этом подходе можно обозначить как минимум две проблемы. Во-первых, нет критерия выбора C_n^φ и C_n , хотя они оба влияют на качество восстанавливаемой оценки x^1 , следовательно, последняя не является единственной и не оптимальна. Другая сложность — вычислительного плана — возникает при нахождении коэффициентов полинома.

2. Восстановление разделяющей поверхности в явном виде

Прежде чем приступить непосредственно к восстановлению разделяющей поверхности, необходимо решить задачу распознавания образов (1.13), (1.4). Предположим, что решение этой задачи получено. В результате каждому элементу $\mathbf{x}_i, i = \overline{1, n}$ из V ставится в соответствие оценка $\tilde{\varphi}_n(\mathbf{x}_i), i = \overline{1, n}$.

Для того чтобы воспользоваться регрессионным алгоритмом для явного восстановления разделяющей поверхности, то есть искать ее как оценку условного математического ожидания $M\{x^1 | (x^2, \dots, x^k), \mathbf{x} \in V^0\}$, необходимо выделить подвыборку V^0 . Согласно пункту 1.2, к V^0 причисляются элементы, удовлетворяющие условию $0 < |\tilde{\varphi}_n(\mathbf{x})| < 1$. Но, как уже указывалось, такое ограничение обеспечивает хорошее (в смысле последующего явного восстановления разделяющей поверхности) качество выборки V^0 только в случае пересечения классов непосредственно вблизи разделяющей поверхности. Если же наблюдается более глубокое проникновение классов друг в друга, качество получаемой оценки ухудшается. Поэтому следует придерживаться следующих правил при формировании выборки V^0 .

Элемент выборки \mathbf{x} лежит на периферии, если $\tilde{\varphi}_n(\mathbf{x}) \in [-1, \varepsilon] \cup [0, \varepsilon] \cup [\varepsilon, 1]$, и наоборот — вблизи разделяющей поверхности, если $\tilde{\varphi}_n(\mathbf{x}) \in (-\varepsilon, 0) \cup (0, \varepsilon)$, где $0 < |\varepsilon| \leq 1$.

Тогда непараметрическая оценка $M\{x^1 | (x^2, \dots, x^k), \mathbf{x} \in V^0\}$ примет вид

$$\hat{x}_n^1(x^2, \dots, x^k) = \frac{\sum_{i=1}^n x_i^1 \prod_{j=2}^k \Phi\left(\frac{x^j - x_i^j}{C_n^\varphi}\right) I(\tilde{\varphi}_n(\mathbf{x}_i))}{\sum_{i=1}^n \prod_{j=2}^k \Phi\left(\frac{x^j - x_i^j}{C_n^\varphi}\right) I(\tilde{\varphi}_n(\mathbf{x}_i))}, \tag{2.1}$$

где $I(\cdot)$ — индикаторная функция, роль которой заключается в дифференциации элементов выборки,

$$I(\tilde{\varphi}_n(\mathbf{x})) = \begin{cases} 1, & \tilde{\varphi}_n(\mathbf{x}) \in (-\varepsilon, 0) \cup (0, \varepsilon); \\ 0, & \tilde{\varphi}_n(\mathbf{x}) \in [-1, -\varepsilon] \cup [0, \varepsilon] \cup [\varepsilon, 1]. \end{cases} \tag{2.2}$$

Параметр размытости C_n настраивается в ходе скользящего экзамена на подвыборке V^0 , так как только эти элементы дают вклад в оценку (2.1). По этой же причине в квадратичный критерий оптимальности типа (1.12), который используется для нахождения оптимального параметра размытости в непараметрической оценке регрессии, так же следует ввести функцию $I(\cdot)$.

Оценку (2.1) можно оптимизировать также и по ε . Таким образом, критерий для нахождения оптимальных значений параметров C_n и ε

$$\omega^2(C_n) = \frac{1}{n^0} \sum_{i=1}^n (x_i^1 - \hat{x}_n^1(x_i^2, \dots, x_i^k))^2 I(\tilde{\varphi}_n(\mathbf{x}_i)) \rightarrow \min_{C_n, \varepsilon}, \tag{2.3}$$

где $n^0 = \sum_{i=1}^n \text{sign } I(\tilde{\varphi}_n(\mathbf{x}_i))$.

Эквивалентным (2.3) является критерий нахождения оптимальных значений параметров C_n и ε в ходе скользящего экзамена на обучающей выборке V^0 при определении класса их принадлежности. Поскольку

оценка (2.1) представляет разделяющую поверхность в явном виде, байесовский критерий изменится следующим образом:

$$\begin{aligned} \mathbf{x} \in X_1, & \text{ если } x^1(x_i^2, \dots, x_i^k) - \hat{x}_n^1(x^2, \dots, x^k) \leq 0; \\ \mathbf{x} \in X_2, & \text{ если } x^1(x_i^2, \dots, x_i^k) - \hat{x}_n^1(x^2, \dots, x^k) > 0. \end{aligned} \quad (2.4)$$

При оптимальных значениях параметров C_n и ε оценка разделяющей поверхности (2.1) удовлетворяет минимуму функции среднего риска (1.1).

Если выборочные значения \mathbf{x}_i поступают последовательно, и в каждый момент времени необходимо оценивать $M\{x^1|(x^2, \dots, x^k), \mathbf{x} \in V^0\}$ в некоторых фиксированных точках, то алгоритму можно придать рекуррентный вид, позволяющий корректировать оценку $M\{x^1|(x^2, \dots, x^k), \mathbf{x} \in V^0\}$ по мере поступления информации. Рекуррентная оценка выглядит следующим образом:

$$\begin{aligned} \hat{x}_n^1(x^2, \dots, x^k) &= \frac{\sum_{i=1}^{n-1} x_i^1 \prod_{j=2}^k \Phi\left(\frac{x^j - x_i^j}{C_{n-1}}\right) I(\tilde{\varphi}_{n-1}(\mathbf{x}_i)) + x_n^1 \prod_{j=2}^k \Phi\left(\frac{x^j - x_n^j}{C_{n-1}}\right) I(\tilde{\varphi}_{n-1}(\mathbf{x}_n))}{\sum_{i=1}^{n-1} \prod_{j=2}^k \Phi\left(\frac{x^j - x_i^j}{C_{n-1}}\right) I(\tilde{\varphi}_{n-1}(\mathbf{x}_i)) + \prod_{j=2}^k \Phi\left(\frac{x^j - x_n^j}{C_{n-1}}\right) I(\tilde{\varphi}_{n-1}(\mathbf{x}_n))} = \\ &= \frac{A_{n-1} + x_n^1 \prod_{j=2}^k \Phi\left(\frac{x^j - x_n^j}{C_{n-1}}\right) I(\tilde{\varphi}_{n-1}(\mathbf{x}_n))}{B_{n-1} + \prod_{j=2}^k \Phi\left(\frac{x^j - x_n^j}{C_{n-1}}\right) I(\tilde{\varphi}_{n-1}(\mathbf{x}_n))} = \frac{A_n}{B_n}, \quad (2.5) \end{aligned}$$

где $A_0 = B_0 = 0$.

За начальное приближение принимается оценка, полученная на исходной обучающей выборке, параметр размытости C_n и величина ε выбираются из условия минимума критерия (2.3) или (2.4). Далее, при появлении новых выборочных значений, по формуле (2.5) производится коррекция оценки \hat{x}_n^1 . Причем сначала вычисляется значение оценки (1.13), поскольку коррекция нужна, только если значения индикаторной функции (2.2) в новых точках не равны нулю. Поэтому, если требуется значение оценки $M\{x^1|(x^2, \dots, x^k), \mathbf{x} \in V^0\}$ в ряде фиксированных точек \mathbf{x} области возможных значений, то расчеты во всех точках можно выполнять параллельно по экономичной в вычислительном отношении рекуррентной формуле (2.5) и избавиться от необходимости хранить всю обучающую выборку.

Асимптотическая сходимость оценки (2.1) не вызывает сомнений в силу сходимости непараметрической оценки регрессии [1], это подтверждают и численные эксперименты (см. следующий параграф).

Таким образом, предложенный метод позволяет находить оптимальную асимптотически сходящуюся непараметрическую оценку разделяющей поверхности $x^1(x^2, \dots, x^k)$.

3. Численное исследование алгоритма

Приведем результаты численного исследования, иллюстрирующие эффективность алгоритма, основанного на использовании оценки (2.1).

В качестве тестового примера, с целью максимальной наглядности, рассматривалась двумерная двальтернативная задача ($k = 2, M = 2$) — $V = \{x_i\}, i = \overline{1, n}$. Исследуемое пространство X представляет собой единичный квадрат, а в качестве разделяющей поверхности взята парабола, делящая площадь квадрата примерно пополам. Пространство X заполнялось обучающей выборкой, сгенерированной датчиком случайных чисел.

В качестве колоколообразной финитной функции использовалась усеченная парабола (1.7).

Допустимая область значений параметра размытости C_n в оценках решающей функции (1.13) и разделяющей поверхности (2.1) определялась с учетом размеров исследуемой области X и вида функции $\Phi(\cdot)$. Максимальное значение параметров должно быть таково, что колокол с центром в любой точке области X должен покрывать все элементы выборки.

Алгоритм тестировался на обучающих выборках различного объема, при этом последовательное увеличение объема выборки производилось путем добавления новых элементов к уже имеющимся. Качество оценок разделяющей поверхности, полученных для $n = 100, n = 200, n = 400$, можно оценить по их графической реализации (рис. 2).

Точками обозначены элементы обучающей выборки V , кружками — элементы подвыборки V^0 .

На соответствующих графиках представлены зависимости, описывающие изменение значения оценки $\omega^2(C_n)$ в зависимости от параметра размытости C_n . Хорошо прослеживается свойство асимптотической сходимости оценки (2.1) при увеличении объема обучающей выборки.

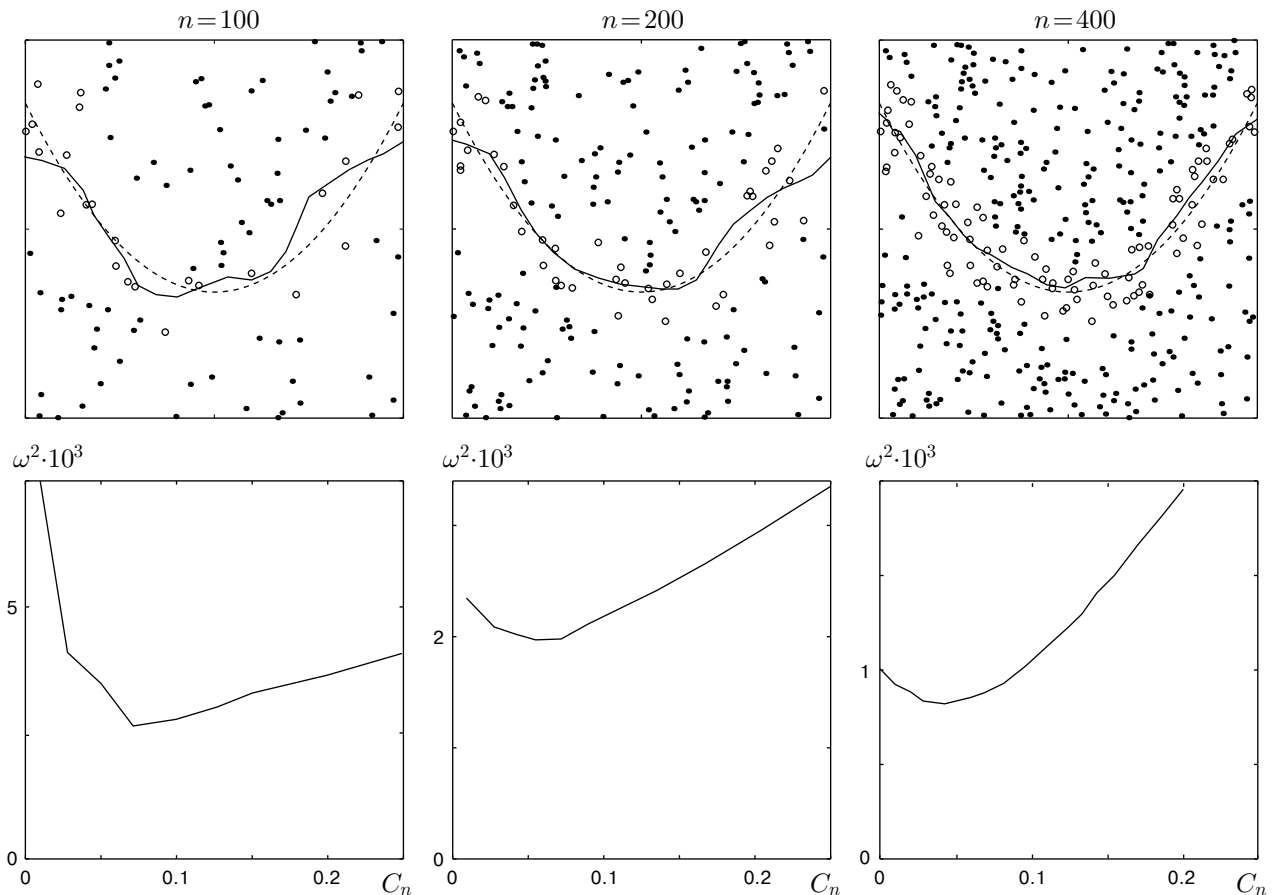


Рис. 2

4. Направления дальнейших исследований

Рассмотренная задача возникла в рамках построения оценки решающей функции в вероятностной постановке задачи распознавания образов при условии, что разделяющая поверхность изменяется со времени. В этом случае алгоритмы, не предполагающие такой особенности РП, работают плохо и возникает большое количество ошибок классификации. В этих условиях проблема явного восстановления разделяющей поверхности является ключевой.

Отметим, что оценка (1.6) имеет один существенный недостаток. Она пригодна лишь при условии, что функция, описывающая разделяющую поверхность, является однозначной. В то же время для задач распознавания в большинстве случаев наблюдается многозначность последней. Поэтому дальнейшие исследования связаны с нахождением непараметрического регрессионного алгоритма, восстанавливающего линию регрессии многозначного вида, после чего построение соответствующей оценки разделяющей поверхности в задаче распознавания образов не составит труда.

В настоящее время уже существуют некоторые представления о том, как можно восстановить многозначную функцию [17]. Но, к сожалению, в алгоритме накладываются определенные условия на обучающую выборку, выполнение которых невозможно, если речь идет о выборке V^0 . Поэтому в настоящий момент рассматриваются другие пути решения этой задачи.

Планируется использование представленного в статье подхода в системе автоматического управления сжиганием угля в котлоагрегате на Красноярской ГРЭС с целью контролирования и изучения протекания технологического процесса.

Литература

1. *Медведев А. В.* Непараметрические системы адаптации. Новосибирск: Наука, 1983. 174 с.
2. *Цыпкин Я. З.* Основы теории обучающихся систем. М.: Наука, 1970. 251 с.
3. *Епанечников В. А.* Непараметрическая оценка многомерной плотности вероятности // Теория вероятностей и ее приложения. 1968. Т. XIV. С. 156–161.

4. *Тарасенко Ф. П.* Непараметрическая статистика. Томск: Изд-во Томского ун-та, 1976. 292 с.
5. *Parzen E.* On estimation of probability density function and mode // *Ann. Math. Stat.* 1962. V. 33. P. 1065–1076.
6. *Кирик Е. С.* Алгоритм восстановления разделяющей поверхности в задаче распознавания образов // *Материалы 3-го сибирского конгресса ИНПРИМ, посвященного памяти С. Л. Соболева.* Новосибирск: 1998. С. 95–96.
7. *Кирик Е. С., Панышин А. Б.* Непараметрические алгоритмы в информационных технологиях технической диагностики // *Вестник НИИ СУВПТ.* 1999. Вып. 2. С. 83–96.
8. *Hart P. E.* The condensed nearest neighbor rule // *IEEE Trans. on Information Theory.* 1968. V. IT-14. № 3. P. 515–516.
9. *Рубан А. И.* Методы анализа данных: Учебное пособие. Ч. 1. Красноярск: КГТУ, 1994. 220 с.
10. *Ripley B. D.* Pattern recognition and neural networks. Cambridge University Press, 1996. 405 p.
11. *Gates G. W.* The reduced Nearest Neighbor rule // *IEEE Trans. on Information Theory.* 1972. V. 18, P. 431–433.
12. *Fukunaga K., Mantock J. M.* Nonparametric data reduction // *IEEE Trans. on Pattern Analysis and Machine Intelligence.* 1984. V. 6. P. 115–116.
13. *Ritter G. L., Woodroff H. B., Lowry S. R., Isenhour T. L.* An algorithm for a selective nearest neighbor decision rule // *IEEE Trans. on Information Theory.* 1975. V. 21. P. 655–669.
14. *Swonger C. W.* Sample set condensation for a condensed nearest neighbor decision rule for pattern recognition // *Frontiers of Pattern Recognition*, ed. S. Watanabe. 1972. P. 511–519.
15. *Tomek I.* Two modification of CNN // *IEEE Trans. on Systems, Man and Cybernetics.* 1976. V. 6. P. 769–772.
16. *Ullmann S. R.* Automatic selection of reference data for use in a nearest neighbor method of pattern classification // *IEEE Trans. on Information Theory.* 1974. V. 20. P. 541–543.
17. *Кирик Е. С.* О восстановлении многозначных зависимостей // *Материалы 7 Всероссийского семинара “Нейроинформатика и ее приложения”.* Красноярск: ИВМ СО РАН, 1999. С. 62–63.

■ Работа поддержана Красноярским краевым фондом науки. Грант № 1М0033.

■ Статья поступила в редакцию 12 сентября 2000 года; в окончательном варианте — 19 ноября 2000 года.