ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

Сборник трудов

Института математики СО АН СССР

1967 г. Выпуск 28

ПРЕОБРАЗОВАНИЕ СТАТИСТИЧЕСКИХ ОПИСАНИЙ КЛАССОВ, ВЫПОЛНЯЕМОЕ В ПЕРСЕПТРОНЕ

В.И. Лебедев

Рассматривается преобразование описаний классов, разделяемых с помощью персептрона. Первичное описание, реализуемое в пространстве ретины персептрона, с помощью системы связей $S \to A$ преобразуется ко вторичному описанию, реализуемому в пространстве ассоциативного слоя.

Показано, что преобразование носит неоптимальный характер и сопряжено со значительными потерями полезной информации.

Каждой реализации на ретине персептрона соответствует своя совокупность возбужденных рецепторов, определяющая некоторую точку x пространства представления информации $\mathcal{X}$, реализованного ретиной. При некоторых очевидных предположениях о структуре ретины пространство $\mathcal{X}$ можно считать евклидовым и, в зависимости от типа используемых рецепторов, может быть непрерывным, дискретным или ортонормированным. Размерность $\mathcal{X}$ будем считать равной числу рецепторов N_s.

Легко видеть, что координаты любой точки $x \in \mathcal{X}$, соответствующей некоторой реализации на ретине персептрона, будут определяться не только геометрической формой образа, но и его площадью и положением на ретине. В результате этого в пространстве $\mathscr{X}$ оказывается возможной высокая коррелированность векторов точек, соответствующих различным, в том числе и подлежащим разделению, реализациям на ретине. То есть описание образов, реализованное в пространстве ретины, оказывается неоптимальным с точки зрения решения задачи распознавания.

Первичное описание класса на ретине персептрона подлежит далее преобразованию с помощью системы связей между ретиной и ассоциативным слоем. Если обозначить через $\mathscr{N}_S$ число рецепторов ретины, через $\mathscr{N}_A$ – число A – элементов и через ℓ –число связей каждого из A – элементов с ретиной, то рассматриваемое преобразование может быть выражено математически в виде матрицы порядка $\mathscr{N}_A \times \mathscr{N}_S$, строки которой для обычного персептрона представляет собой случайные векторы с ℓ отличными от нуля координатами в пространстве $\mathscr{X}$.

Подобная матрица описывает вырожденное линейное преобразование с отбрасыванием $(\mathscr{N}_S - \ell)$ координат $\mathscr{X}$. Рассмотрим потери полезной информации в процессе такого преобразования.

Преобразования имеет смысл слияния нескольких точек $x \in \mathscr{X}$ в одну точку $y \in \mathscr{Y}$, где $\mathscr{Y}$ – $\mathscr{N}_A$ – мерное пространство вторичных описаний персептрона, которое может быть определено аналогично пространству $\mathscr{X}$ для совокупности входов A– элементов.

Слияние выполняется для совокупности точек $x \in \mathscr{X}$, определяемых при помощи ℓ рецепторов данного A-элемента и имеющих каждая не более ℓ отличных от нуля координат. Так, в случае двоичной ретины и беспорогового A-элемента число сливаемых точек x будет:

$$m = \sum_{i=1}^{\ell} C_\ell^i .$$

Сливаемые точки представляют собой в этом случае всевозможные линейные комбинации [I] из ℓ случайным образом выбранных базисных точек ортонормированного пространства $\mathscr{X}$.

Смысл любого преобразования пространства представления информации в задаче распознавания образов можно видеть в преобразовании статистических описаний разделяемых классов с целью уменьшения возможной неоднозначности решения. В нашем случае статистическими описаниями классов являются апостериорные распределения вероятностей в пространстве представления информации, и преобразование должно иметь целью уменьшение числа точек пространства, имеющих отличные от нуля вероятности реализации как первого $p(x/K_I)$, так и второго $p(x/K_{II})$ (в случае дихотомии) разделяемых классов.

Указанная задача обычно решается одним из двух возможных способов: посредством "разнесения" множеств точек разделяемых классов (например, в смысле увеличения среднеквадратического расстояния между точками множеств, [2]) при сохранении значений $p(x/K_I)$ и $p(x/K_{II})$ в каждой точке или посредством преобразования формы апостериорных распределений (за счет подчеркивания информативных и устранения избыточных характеристик) при сохранении постоянным определенного каким-либо образом расстояния между множествами точек классов.

Рассмотрим с этой точки зрения преобразование первичного пространства представления информации в персептроне.

Увеличение "разнесенности" классов или улучшение геометрических характеристик представления информации в случае использования преобразования типа слияния точек требует отбора для слияния точек $x \in \mathscr{X}$, имеющих отличные от нуля вероятности реализации одного и того же из разделяемых классов. При этом слияние должно выполняться отдельно для совокупности точек, реализуемых каждым из классов.

Рассматриваемое преобразование оперирует со случайно взятыми точками пространства $\mathscr{X}$, выбор которых никоим образом не связан с оценкой соответствующих значений $p(x/K_I)$ и $p(x/K_{II})$. Такое преобразование, очевидно, не способно улучшить геометрию представления информации.

Для оценки изменения информационных характеристик представления проследим преобразование распределений $p(x/K_I)$ и $p(x/K_{II})$.

Условную вероятность $p(A_i/K_j)$ возбуждения элемента A_i образом класса K_j можно записать в виде:

$$p(A_i/K_j) = \sum_{\mu} p(x/K_j) = \sum_{\mu} p(\ldots, S_k, \ldots, S_\ell, \ldots, S_m, \ldots /K_j), \qquad k, \ell, m = 1, 2, \ldots, \mathscr{N}_S, \tag{I}$$

где $\mathscr{M}$ – указанное выше подмножество точек $x \in \mathscr{X}$, сливаемых с помощью данного A_i в одну точку $y \in \mathscr{Y}; p(\ldots, S_k, \ldots, S_m, \ldots /K_j)$–

условная вероятность одновременного возбуждения реализаций класса k_j совокупности рецепторов $\ldots, k, \ldots, \ell, \ldots, m, \ldots$, определяющей точку $x \in \mathcal{M}$.

По аналогии можно записать:

$$p(y/k_j) = p(\ldots, A_p, \ldots, A_q, \ldots, A_s, \ldots /k_j), \; p, q, s = 1, 2, \ldots, N_A. \quad (2)$$

Как видно из (I) и (2), зависимость между $p(x/k_j)$ и $p(y/k_j)$ достаточно сложна. Будем поэтому рассматривать соотношение между $p(x/k_j)$ и $p(A_i/k_j)$, считая последнюю вероятность одной из составляющих $p(y/k_j)$.

Как это следует из (I), одно и то же значение $p(A_i/k_j)$ реализуется в персептроне (по завершении его обучения и подбора системы весов) при возбуждении реализацией класса k_j любой точки x на совокупности точек $\mathcal{M}$, сливаемых с помощью данного А-элемента. Но $p(A_i/k)$ есть элемент вторичного статистического описания j-того класса, и в качестве такового используется далее в персептроне для принятия решения о классовой принадлежности. Оценка $p(A_i/k_j)$, получаемая по формуле (I), не несет информации о распределении вероятностей $p(A_i/k_j)$ на множестве сливаемых точек $x \in \mathcal{M}$, однако именно эта информация как деталь первичного статистического описания представляет для нас интерес с точки зрения решения задачи распознавания. Одно и то же значение $p(A_i/k_j)$, реализуемое в персептроне, может соответствовать целому ряду распределений вероятностей $p(x/k_j)$ на подмножестве $\mathcal{M}$, то есть может соответствовать целому ряду различных (в том числе и подлежащих разделению) входных воздействий. Определяемая в этом смысле информативность оценки $p(A_i/k_j)$ повышается только в частном случае равенства $p(x/k_j)$ для всех $x \in \mathcal{M}$, не представляющем практического интереса.

Таким образом, преобразование апостериорных распределений по классам, выполняемое в обычном персептроне, не только не улучшает информационных характеристик этих статистических описаний и не облегчает тем самым решения задачи классификации, но и сопряжено с потерями полезной информации, тем большими, чем сложнее оказывается функциональная форма преобразуемых распределений.

Картина механизма рассматриваемого преобразования может быть дополнена изучением статистических закономерностей сигнала на входе А-элемента.

Сигнал y_i на входе i-того А-элемента записывается в следующем виде:

$$y_i = \sum_{k=1}^{\ell} x_k,$$

где x_k - сигнал с k-того рецептора из общего числа ℓ подключенных к элементу A_i. Совокупность $x_1, x_2, \ldots, x_\ell$ сигналов, снимаемых с рецепторов данного А-элемента при подаче на ретину некоторой реализации, можно считать совокупностью ℓ случайных величин. Эти величины не будут независимы в общем случае, так как дельта-коррелированность образа на ретине по пространству означала бы вырождение этого образа в точку.

Предположим, что нам известны несколько первых моментов или семиинвариантов распределений $p(x_k/k_j), k = 1, 2, \ldots, \ell, j = const$. Тогда задача нахождения распределения $p(y_i/k_j)$ или, что то же самое, распределения $p(A_i/k_j)$ сводится к известной задаче теории вероятности: найти распределение суммы при конечном, но достаточно большом числе слагаемых при условии, что для распределений слагаемых известны по крайней мере несколько первых моментов или семиинвариантов.

Поставленная задача может быть разбита на два этапа: необходимо доказать сам факт объективного существования распределения $p(y_i/k_j)$ и в случае существования $p(y_i/k_j)$ необходимо определить его тип.

Доказательство факта объективного существования $p(y_i/k_j)$ может быть проведено с использованием одной из форм закона больших чисел теории вероятности, а именно - теоремы Маркова (см., например, [3]), согласно которой для того, чтобы для последовательности $x_1, x_2, \ldots, x_\ell$ как угодно зависимых случайных величин при любом $\xi > 0$ выполнялось соотношение

$$\lim_{\ell \to \infty} P\left\{ \left| \frac{1}{\ell} \sum_{k=1}^{\ell} x_k - \frac{1}{\ell} \sum_{k=1}^{\ell} \langle x_k \rangle \right| < \xi \right\} = 1, \quad (3)$$

необходимо и достаточно, чтобы

$$\lim_{\ell\to\infty} < \frac{\left[\sum_{K=1}^{\ell}(x_K - \langle x_K\rangle)\right]^2}{\ell^2 + \left[\sum_{K=1}^{\ell}(x_K - \langle x_K\rangle)\right]^2} > = 0. \qquad (4)$$

Условие (4), как обычно, требует, чтобы дисперсии величин x_K были конечны и не слишком разнились между собой, а совокупность x_K удовлетворяла бы (4) при достаточном большом ℓ.

Строгое доказательство выполнения условия (4) в каждом отдельном случае требует экспериментального изучения распределений $p(x_K/K_j)$, $K=1,2,\ldots,\ell$. Однако практика показывает, что при не слишком малой разрешающей способности рецепторов ретины, при соответствующем выборе конфигурации совокупности рецепторов данного А-элемента, а также при использовании $\ell \geqslant 20 \div 50$ условие (4) практически всегда выполняется и имеет место факт сходимости распределения $p(y_i/K_j)$.

Более или менее строгое определение функциональной формы $p(y_i/K_j)$ потребовало бы привлечения математического аппарата центральной асимптотической проблемы (см., например, [4]) и вряд ли необходимо в настоящей работе. Поэтому ограничимся здесь указанием, что при использовании $\ell \geqslant 50 \div 100$ и при соблюдении оговоренных выше условий форма распределения $p(y_i/K_j)$ оказывается весьма близкой к нормальной для широкого класса входных воздействий.

Таким образом, если считать $p(x_K/K_j)$ деталью первичного статистического описания, а $p(y_K/K_j)$ деталью вторичного статистического описания, то можно утверждать, что в то время как информативность $p(x_i/K_j)$ определяется его формой, которая отражает специфику входных воздействий и может быть сколь угодно сложной, информативность $p(y_i/K_j)$ можно оценивать при помощи всего лишь двух первых моментов распределения, которые полностью его определяют. Этот факт еще раз подтверждает значительность потерь полезной информации в процессе преобразования первичного пространства представления информации в персептроне. Случайный характер этого преобразования, выполняемого при посредстве операций суммирования и сравнения, не обеспечивает избирательности к полезной информации, и тот небольшой её объем, которым персептрон располагает в первичном пространстве представления, в процессе преобразования уменьшается как абсолютно, так и относительно.

Учет вышеизложенных особенностей механизма преобразования пространства представления информации в персептроне может оказаться полезным при поиске методов оптимизации структуры и логики устройства.

ЛИТЕРАТУРА

1. Б.А. Розенфельд. Многомерные пространства. Изд. "Наука", 1966 г.
2. Г. Себастиан. Процессы принятия решений при распознавании образов. Изд. "Техника", Киев, 1965 г.
3. Б.В. Гнеденко. Курс теории вероятностей. Физматгиз, Москва, 1965 г.
4. М. Лоэв. Теория вероятностей. Пер. с англ. под ред. Ю.В. Прохорова, И.И.Л., Москва, 1962 г.

Поступила в редакцию
I/X-1967г.