УДК 62-506.1

ОБ ОТЫСКАНИИ ГЛОБАЛЬНОГО МАКСИМУМА В ОДНОЙ ЗАДАЧЕ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ

ЛИТИНСКИЙ Л. Б.

(Троицк)

Для специального функционала решается задача автоматической классификации в случае, когда отыскивается оптимальное разбиение объектов на два класса. Объектами являются векторы. Заданный на множестве их разбиений функционал многоэкстремален; отыскание его глобального максимума связывается со свойствами спектра матрицы взаимных корреляций исходных векторов.

1. Постановка задачи. Вспомогательная задача

Пусть $\{X^i\}$, $i=1,\ldots,m$ — совокупность m векторов произвольной размерности. Рассмотрим множество всех возможных разбиений этой совокупности на две непересекающиеся группы: если $I=(i_1, i_2, \ldots, i_{m_1})$ — список векторов, объединенных в одну группу, то список векторов из второй группы обозначим через $\bar I=(1, 2, \ldots, m)\setminus I$. На множестве разбиений определим функционал

$$F(I,\bar I)\equiv F(I)=\Big\|\sum_{i\in I}X^i\Big\|^2/|I|+\Big\|\sum_{j\in\bar I}X^j\Big\|^2/|\bar I|, \tag{1}$$

где $|I|$ — длина списка I, а $\|X\|$ — декартова длина вектора X. Задача состоит в отыскании глобального максимума функционала (1). Таким образом, имеем задачу автоматической классификации на два класса с функционалом качества типа средневзвешенной по классам дисперсии; этот функционал использовался, в частности, в [1].

Сформулируем задачу несколько иначе. Пусть $M=((X^i, X^j))_{i,j=1}^m$ — действующая в пространстве R^m матрица взаимных корреляций исходной совокупности векторов $\{X^i\}$. Каждому списку I сопоставим в R^m нормированный вектор $e(I)$ с координатами $e^j(I)$:

$$e^j(I)=0, \text{ если } j\notin I, \text{ и } e^j(I)=1/\sqrt{|I|}, \text{ если } j\in I. \tag{2}$$

Дополнительному к I списку $\bar I$ отвечает при этом вектор $e(\bar I)$:

$$e^j(\bar I)=0, \text{ если } j\in I, \text{ и } e^j(\bar I)=1/\sqrt{|\bar I|}, \text{ если } j\notin I. \tag{2'}$$

Легко видеть, что $\Big\|\sum_{i\in I}X^i\Big\|^2\Big/|I|=(Me(I),e(I))$, и тогда

$$F(I)=(Me(I),e(I))+(Me(\bar I),e(\bar I)),$$

что, учитывая ортогональность $e(I)$ к $e(\bar I)$, есть $Sp_{\pi(I)}M$ — след матрицы M на плоскости $\pi(I)$, проходящей через пару векторов $e(I)$, $e(\bar I)$.

Итак, исходная задача сформулирована в виде:

$$Sp_{\pi(I)}M\to\max, \tag{3}$$

где $\pi(I)$ — плоскость, натянутая на векторы (2), (2′).

Если бы в задаче (3), (2), (2′) варьируемая плоскость не задавалась условиями (2), (2′), а была произвольной, то решением задачи была бы плоскость π, проходящая через два наибольших собственных вектора f_1 и f_2 матрицы M (см. [2]):

$$Mf_i=\lambda_i f_i,\ \lambda_1\geqslant\lambda_2\geqslant\ldots\geqslant\lambda_m.$$

В том частном случае, когда плоскость π принадлежит множеству плоскостей $\pi(I)$, величина максимума в (3) равна $\lambda_1+\lambda_2$. В общем же случае плоскость π не принадлежит к числу плоскостей $\pi(I)$, соответственно и величина максимума в (3) меньше $\lambda_1+\lambda_2$. Естественно, однако, предположить: чем ближе варьируемая плоскость $\pi(I)$ к π, тем для нее значение $\mathrm{Sp}_{\pi(I)}M$ ближе к $\lambda_1+\lambda_2$.

Ниже будет определена мера близости плоскостей π_h, $\pi_g - \Delta(\pi_h, \pi_g)$; сейчас только подчеркнем, что от задачи (3), (2) (2′) мы переходим к задаче разыскания среди плоскостей $\pi(I)$ плоскости, ближайшей к π:

(4) $$\Delta(\pi, \pi(I))\to\max.$$

Разница между задачами (3), (2), (2′) и (4), (2), (2′) будет обсуждена позже.

2. Решение вспомогательной задачи

Определим удобную для наших целей меру близости между плоскостями: если $\pi_h(\pi_g)$ — плоскость, проходящая через пару ортонормированных векторов h_1 и h_2 (g_1 и g_2), то

(5) $$\Delta(\pi_h, \pi_g)=\begin{vmatrix} (h_1, g_1) & (h_1, g_2) \\ (h_2, g_1) & (h_2, g_2)\end{vmatrix}.$$

Замечание. Знак при Δ для наших целей неважен; в этом смысле (4) следует понимать как максимизацию $|\Delta(\pi, \pi(I))|$, хотя удобнее работать с характеристикой $\Delta(\pi, \pi(I))$ и отыскивать ее наибольшее и наименьшее значения. $\Delta(\pi_h, \pi_g)$ удовлетворяет требованиям, предъявляющимся к мерам близости: $\Delta(x, y)=\Delta(y, x)$, $0\leqslant|\Delta(x, y)|<1$, $\Delta(x, x)=1$. Однако истинной причиной использования именно этой меры близости является специфическая структура множества плоскостей $\pi(I)$.

Так как вектор $\sqrt{m}e_0=(1, 1, \ldots, 1)$ принадлежит каждой $\pi(I)$ (для каждого I: $\sqrt{|I|}e(I)+\sqrt{|\bar{I}|}e(\bar{I})=\sqrt{m}e_0$), удобно характеризовать $\pi(I)$ не парой векторов $e(I)$, $e(\bar{I})$, а вектором e_0 и ортогональным к нему в плоскости $\pi(I)$ вектором $t(I)$. Из 2^{m-1} различных векторов $t(I)$ выберем, и назовем их базисными, $m-1$ векторов, списки I которых состоят из одного индекса: $I=(i)$. Так как всего таких списков m, можно ограничиться например, $m-1$ первыми векторами.

Итак, базисными являются векторы $t(I)=t(i)\equiv t_i$, $i=1, 2, \ldots, m-1$. Величины близости соответствующих плоскостей к π обозначим Δ_i:

(6) $$\Delta_i\equiv\Delta(i)=\begin{vmatrix} (f_1, e_0) & (f_1, t_i) \\ (f_2, e_0) & (f_2, t_i)\end{vmatrix}.$$

Легко проверить, что для произвольного списка I (не содержащего индекса m) справедливо $t(I)=\sqrt{\dfrac{m-1}{|I|\cdot|\bar{I}|}}\sum\limits_{i\in I}t_i$. Отсюда следует, что

$$\Delta(I)=\sqrt{\frac{m-1}{|I|\cdot|\bar{I}|}}\sum_{i\in I}\Delta_i,$$

а тогда максимизировать $|\Delta(I)|$ довольно просто.

В самом деле, без ограничения общности можно считать, что

(7) $$\Delta_1\geqslant\Delta_2\geqslant\ldots 0\ldots\geqslant\Delta_{m-1},$$

а величина Δ_m, отбрасываемая при образовании базисных,— наименьшая. Тогда, как можно показать, для определения $\Delta(I^*)=\max|\Delta(I)|$ достаточно определить наибольшую из величин

$$(8)\qquad S_k=\sqrt{\frac{m-1}{k(m-k)}}\left|\sum_{i=1}^{k}\Delta_i\right|,\qquad k=1,2,\ldots,m-1.$$

При этом оказывается, что максимум, равный $\Delta(I^*)$, не зависит от того, какая из величин Δ_i отбрасывалась при образовании базисных.

В заключение сформулируем алгоритм решения задачи (4), (2), (2′):

а) отыскиваем наибольшие собственные векторы f_1, f_2 матрицы M; по формуле (6) вычисляем Δ_i и, упорядочив их по убыванию, образуем ряд (7);

б) вычисляем $m-1$ величин (8); список I^*, отвечающий наибольшей из этих величин, задает разбиение $(I^*, \bar{I}^*)$, оптимальное в смысле критерия (4).

3. Решение исходной задачи. Экспериментальная проверка алгоритма

Основные различия между критериями (3) и (4) связаны с тем, что в последнем не учитывается влияние собственных векторов матрицы M, отвечающих собственным значениям $\lambda_3, \lambda_4, \ldots$ Если же среди них имеются собственные значения, сравнимые по величине с λ_1 и λ_2, то влияние соответствующих собственных векторов может оказаться решающим при определении того, для какой из одинаково близких к π плоскостей $\pi(I)$ значение критерия (3) больше. Типична, по-видимому, следующая ситуация: несколько плоскостей приблизительно одинаково близки к π, и именно среди них находится плоскость, доставляющая критерию (3) глобальный максимум. Важно подчеркнуть, что для плоскости, доставляющей глобальный максимум функционалу (3), значение критерия (4) принципиально не может быть очень мало.

Таким образом, для решения задачи (3), (2), (2′) пункт б) алгоритма должен быть модифицирован:

в) вычисляем $m-1$ величин (8) и определяем те несколько наибольших из них, которые заметно превосходят остальные. Среди соответствующих списков I будет и тот, который доставляет глобальный максимум функционалу (3).

Экспериментальная проверка алгоритма а), в) проводилась на числовом материале из [3], где хорошие сверхпроводящие свойства проводников определенной кристаллической структуры прогнозировались с использованием методики [4]. Не вдаваясь здесь в физические подробности, укажем, что из исходного числового материала нами были взяты шесть матриц «объекты — параметры» с числом объектов, равным 60, и числом векторов-параметров $m=5$, 6, 8, 10. Для каждого из этих материалов полным перебором определялась группировка векторов, доставляющая глобальный максимум функционалу (3), а с другой стороны — проводились вычисления, диктуемые пунктами а), в) алгоритма. В трех из этих шести экспериментов группировка, доставляющая функционалу (3) глобальный максимум, была и оптимальной в смысле критерия (4), в двух экспериментах она была на втором месте по оптимальности, а в одном случае — на третьем.

Еще одной проверкой алгоритма явилась следующая модельная задача: две группы единичных векторов расположены на плоскости так, что внутри групп векторы повернуты друг относительно друга на один и тот же угол φ, а угол θ, отделяющий одну группу от другой, больше φ; число векторов в обеих группах одинаково.

Прямыми вычислениями можно показать, что в этом случае глобальный максимум функционала (1) достигается на естественном — погрупповом — разбиении. Этот материал характеризуется выполнением равенств $\lambda_3=\lambda_4=\ldots=0$, так что можно обойтись алгоритмом а), б). И действительно, вычисления по формулам (6)–(8) показывают, что группировка, доставляющая здесь глобальный максимум функционалу (3), обязательно будет и оптимальной в смысле критерия (4).

В заключение отметим, что весь подход естественным образом обобщается на произвольное число классов, однако работы в этом направлении еще нельзя считать законченными.

ЛИТЕРАТУРА

1. *Айзерман М. А., Браверман Э. М., Розоноэр Л. И.* Метод потенциальных функций в теории обучения машин. М.: Наука, 1970.
2. *Глазман И. М., Любич Ю. И.* Конечномерный линейный анализ. М.: Наука, 1969.
3. *Литинский Л. Б., Попова С. В.* Качественное прогнозирование $T_к$ для широкого класса двойных систем. Депонировано в ВИНИТИ, № 2404-79.
4. *Браверман Э. М.* Методы экстремальной группировки параметров и задача выделения существенных факторов.– АиТ, 1970, № 1, с. 123–132.

Поступила в редакцию
27.IX.1983

ON FINDING THE GLOBAL MAXIMUM IN ONE AUTOMATIC CLASSIFICATION PROBLEM

LITINSKIY L. B.

For a special kind of functional the problem of automatic classification is solved in the case where the objects have to be divided into two groups in an optimal way, the objects being vectors. The functional specified on the set of their decompositions has many extrema; its global maximum is found on the knowledge of the spectrum properties of the matrix of mutual initial scalars vectors.