

基于分离度的图象特征提取与识别方法

连石柱

(中国科学院遥感应用研究所 北京 100101)

摘要 对于图象中不同类别(以其统计分布函数来表征)的可识别性,分离度的统计定义可做出定量的描述。根据分离度在一一对应映射变换下的不变性,应用 Karhunen-Loève 变换对两类分布提取识别特征,发现分离度只依赖于特征值最大和最小的两个特征向量。而且,分离度的大小依赖于特征值与某个定值的偏差。由此我们提出一个识别模型,使每次分类后的样本集的特征值都趋近于两个定值,从而得到最佳的识别效果。由该模型设计的迭代算法在理论上证明是收敛的,在实际应用中是有效的,是在图象识别方法上的有益的理论探讨。

关键词 特征提取,混合归一化,样本自相关阵,K-L 变换,特征值,特征向量,分离度,迭代,聚类

1 引言

图象识别包含着特征选择的过程。对于单种类别,特征选择是选择对样本表示有贡献的特征;对于多种类别,特征选择则是选择对类别可分离性有效的特征。对于类别间的可分离性,数学上用“分离度”进行定量描述。基于分离度的定义,又根据分离度在一一对应映射下不变的原理^[1],用 K-L 变换进行特征提取,从而提出了一个样本交换迭代的聚类识别模式。该方法的基本思想是,对于两种不同的类别先进行混合归一变换,使其自相关阵具有相同的特征向量集,两个类别对应的同一特征向量的特征值排序相反。在此基础上,用同一特征向量阵对两类样本进行 K-L 变换^[2],在新的坐标空间,即特征空间里进行分类,发现两类之间的分离度只依赖于特征值最大和最小的特征向量,而其他特征向量对分离度没有贡献。进一步的分析表明,分离度的大小依赖于上述两个特征值对某个定值的偏差(取绝对值),偏差越大分离度越大,由此导出可收敛的迭代聚类过程。依据序贯分析的思想,两类样本的识别问题可以推广到多类问题中。

2 类别可分离性的特征选择

为了方便,以两种类别的问题为例,现设有 N 个样本的集合:

$$\{\mathcal{X}\} = \{x_1, x_2, \dots, x_N\}$$

$\{\mathcal{X}\}$ 中的 $x_i (i=1, \dots, N)$ 是 n 维向量。 $\{\mathcal{X}\}$ 的两个互不相交的子集分别为:

$$\{\mathcal{X}^{(1)}\} = \{x_1^{(1)}, \dots, x_{N_1}^{(1)}\}, \text{ 含 } N_1 \text{ 个样本}$$

$$\{\mathcal{X}^{(2)}\} = \{x_1^{(2)}, \dots, x_{N_2}^{(2)}\}, \text{ 含 } N_2 \text{ 个样本}$$

我们把 $\{\mathcal{X}\}$ 中的任一样本属于 $\{\mathcal{X}^{(k)}\}, k=1, 2$ 的先验概率表示为 $P(\omega_k), k=1, 2$ 。显然,

$$N_1 + N_2 = N$$

$$P(\omega_k) = \frac{N_k}{N} \quad k = 1, 2$$

2.1 混合归一化

为了提取特征,需要做两类分布样本的混合归一化。这里引用定理^[3]:

定理:设矩阵 Q 是由两个对称矩阵 Q_1 和 Q_2 的线性组合给出的, $Q = a_1 Q_1 + a_2 Q_2$, 假定 a_1, a_2 都

是正数,若将诸特征向量对Q归一化,则Q1和Q2有相同的特征向量,它们特征值的排序为:

$$\lambda_1^{(1)} > \lambda_2^{(1)} > \dots > \lambda_n^{(1)} \quad \text{对于 } Q_1$$

$$\lambda_1^{(2)} > \lambda_2^{(2)} > \dots > \lambda_n^{(2)} \quad \text{对于 } Q_2$$

现在假设, S1, S2 分别是样本集 {X(1)}, {X(2)} 的自相关阵,而 S 是自相关阵 S1 和 S2 的两个分布的混合自相关阵,则有:

$$\begin{aligned} S &= E\{xx^T\} \\ &= P(\omega_1)E\{xx^T/\omega_1\} + P(\omega_2)E\{xx^T/\omega_2\} \\ &= P(\omega_1)S_1 + P(\omega_2)S_2 \end{aligned}$$

按着上述定理,我们能够使 S1 和 S2 对角化而有相同的特征向量集,具体做法如下:

假定两类的先验概率相等,即 P(ω1) = P(ω2) = 1/2 并设 Φ, Λ 分别是混合自相关阵 S 的特征向量阵和特征值阵,经过白化变换和同时对角化[3]可将

$$\frac{1}{2}S_1 + \frac{1}{2}S_2 = S \text{ 变为:}$$

$$\frac{1}{2}S'_1 + \frac{1}{2}S'_2 = I \quad (S'_j \text{ 是新坐标系下的 } S_j)$$

考虑到

$$S'_2 \phi_j^{(2)} = (2I - S'_1) \phi_j^{(2)} = \lambda_j^{(2)} \phi_j^{(2)} \quad (3-1)$$

这里, φj(1), λj(1) 是 S'1 的特征向量和特征值; φj(2), λj(2) 是 S'2 的特征向量和特征值, j = 1, ..., n

整理(3-1)式,得

$$S'_1 \phi_j^{(2)} = (2 - \lambda_j^{(2)}) \phi_j^{(2)}$$

由特征值、特征向量的唯一性(不计顺序),得

$$\begin{aligned} \phi_j^{(1)} &= \phi_j^{(2)} \\ \lambda_j^{(1)} &= 2 - \lambda_j^{(2)} \quad j = 1, \dots, n \end{aligned}$$

这说明,两个不同分布集的自相关阵具有相同的特征向量,其相应的特征值排序相反,第一个分布有最大特征值的特征向量,对第二个分布有最小的特征值,反之亦然。并且,由于先验概率相等,由以上推导,显然有:

$$2 \geq \lambda_1 \geq \dots \geq \lambda_n \geq 0 \quad \text{对于类别 } \omega_1$$

$$0 \leq 2 - \lambda_1 \leq \dots \leq 2 - \lambda_n \leq 2 \quad \text{对于类别 } \omega_2$$

进一步的推导表明,当假设两个分布的协方差阵也相等时,两个分布自相关阵的特征值可以算出(推导计算过程见附录1):

$$\lambda_1 = 1 + |M_0| |M| (1 - \cos\theta) \geq 1$$

$$\lambda_2 = \lambda_3 = \dots = \lambda_{n-1} = 1$$

$$\lambda_n = 1 - |M_0| |M| (1 + \cos\theta) \leq 1$$

其中

$$M_0 = \frac{1}{2}(M_1 + M_2) \quad (M_1 \text{ 是 } \{X^{(1)}\} \text{ 的平均向量})$$

$$M = \frac{1}{2}(M_2 - M_1) \quad (M_2 \text{ 是 } \{X^{(2)}\} \text{ 的平均向量})$$

θ 是向量 M0 与 M 之间的夹角。

$$\cos\theta = \frac{M_0^T M}{|M_0| \cdot |M|}, \quad |\cdot| \text{ 表示向量长度}$$

2.2 K-L 变换后的分离度

分离度(Divergence)在文献[3]上定义为两个分布的密度函数似然比的期望值之差,如下式:

$$\begin{aligned} Div &= E\left\{-\ln \frac{P(x/\omega_1)}{P(x/\omega_2)} / \omega_2\right\} \\ &= E\left\{-\ln \frac{P(x/\omega_1)}{P(x/\omega_2)} / \omega_1\right\} \\ &= \int_{\varphi} P(x/\omega_1) \ln \frac{P(x/\omega_1)}{P(x/\omega_2)} dx \\ &\quad - \int_{\varphi} P(x/\omega_2) \ln \frac{P(x/\omega_1)}{P(x/\omega_2)} dx \end{aligned}$$

作者已经证明(过程在此从略),当两个分布为正态时,它们的分离度可由下式计算:

$$\begin{aligned} Div &= \frac{1}{2}(M_1 - M_2)^T (\Sigma_1^{-1} - \Sigma_2^{-1})(M_1 - M_2) \\ &\quad + \frac{1}{2} \text{tr}(\Sigma_1^{-1} \Sigma_2 + \Sigma_2^{-1} \Sigma_1 - 2I) \quad (3-2) \end{aligned}$$

这里, M1, M2 分别为两类的平均向量; Σ1, Σ2 分别为两类的协方差阵; tr(·) 是求矩阵特征值之和的运算。

因为分离度在一一对应的映射变换下是不变的,经过推导,我们可给出经 K-L 变换后的计算公式

$$\begin{aligned} Div &= \frac{1}{2}(L_1 - L_2)^T (\Theta_1^{-1} + \Theta_2^{-1})(L_1 - L_2) \\ &\quad + \frac{1}{2} \text{tr}[(\Theta_1 - \Theta_2)(\Theta_1^{-1} - \Theta_2^{-1})] \end{aligned}$$

其中, Lk = BTDk, B 为 S'j 的特征向量阵, k = 1, 2

$$D_k = \begin{cases} \Lambda \\ I - \Lambda \end{cases}$$

$$\Theta_K = D_k / P(\omega_k) - L_k L_k^T$$

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$$

当协方差阵相等时,即, Θ1 = Θ2 = Θ, (3-2)式可变为:

$$\begin{aligned} Div &= (L_1 - L_2)^T \Theta^{-1} (L_1 - L_2) \\ &= \frac{1}{\theta_{11} \theta_{nn} - \theta_{1n}^2} [\theta_{nn} (l_1^{(2)} - l_1^{(1)})^2 + \theta_{11} (l_n^{(2)} - l_n^{(1)})^2 \\ &\quad - 2\theta_{1n} (l_1^{(2)} - l_1^{(1)}) (l_n^{(2)} - l_n^{(1)})] \quad (3-3) \end{aligned}$$

其中,

$$l_j^{(k)} = \phi_j^T D_k = \begin{cases} \neq 0 & \text{对 } j=1, n \\ = 0 & \text{对 } j=2, \dots, n-1 \end{cases} \quad k=1, 2$$

$$\theta_{ii} = 2\lambda_i - l_i^{(1)} \cdot l_i^{(1)} \quad (i=1, n)$$

$$\theta_{ij} = \theta_{ji} = -l_i^{(1)} l_j^{(1)} \quad (i, j=1, n)$$

由(3-3)式可以看出,经过 K-L 变换之后,原来计算分离度需 n 个变量,现在只需 $\lambda_1, \lambda_n, \Phi_1, \Phi_n$ 四个变量, $2 \geq \lambda_1 \geq 1, 1 \geq \lambda_n \geq 0$, 换言之, $\lambda_j, \Phi_j, j=2, \dots, n-1$ 对分离度没有贡献。对于协方差不等的情况,理论证明将会相当繁琐,本文未做探讨。

由上可知,为了使分离度大,应选择距离 2 或 0 尽可能近的特征值,这就是对类别可分离性特征的最佳选择。由此即可产生以下聚类判别模型。

3 样本交换迭代的聚类算法

由以上不难看出,对于分类,一个相对“好”的样本集应使每个特征都做到最有效,也即使两类的每个特征值都趋近于 2 或 0,于是我们确立了如下的判据:

$$c = \frac{1}{n} \sum_{j=1}^n (\lambda_j^{(1)} - 1)^2 + \frac{1}{n} \sum_{j=1}^n (\lambda_j^{(2)} - 1)^2 \quad (3-4)$$

将此式稍加整理,得:

$$c = \frac{2}{n} \left[\sum_{j=1}^n \lambda_j^{(1)2} - 2 \sum_{j=1}^n \lambda_j^{(1)} \right] + 2$$

但, $\sum_{j=1}^n \lambda_j^{(1)2} = t_r S_1^2 = \sum_{i=1}^n \sum_{j=1}^n S_{ij}^2$

$$\sum_{j=1}^n \lambda_j = t_r S_1 = \sum_{i=1}^n S_{ij}$$

其中, S_{ij} 表示矩阵 S_1 的第 i 行 j 列元素的值。

所以 $c = \frac{2}{n} \left[\sum_{i=1}^n \sum_{j=1}^n S_{ij}^2 - 2 \sum_{i=1}^n S_{ii} \right] + 2$

合理的分类应使 c 值增加,为了量度 c 的变化,作者推导出了 Δc 的计算公式(推导过程见附录 2):

$$\Delta c = \Delta c_k + \Delta c_l$$

其中

$$\Delta c_k = \frac{4}{nN_1} (y_k^T S_1 y_k - y_k^T y_k)$$

$$\Delta c_l = -\frac{4}{nN_1} (y_l^T S_1 y_l - y_l^T y_l)$$

这里, y_l 是划入 ω_1 类的样本, y_k 是划入 ω_2 类的样本, Δc 表示 y_l 与 y_k 交换引起的 c 值的变化。 N_1 是 ω_1 类样本的个数。推导出 Δc 公式的好处是:不必计

算出样本交换前 S_1 的特征值,不必算出交换后的自相关阵,利于编程。

到此,我们可以归纳出以下具体的计算步骤:

第一步,找到一个判据 c ,当分类正确性增加时,使 c 达到最大。第二步,将所有 N 个样本任意地划分为各有 N_1, N_2 个样本的两个子集 ($N_1 + N_2 = N$), 设子集中的样本分别属于 ω_1, ω_2 类。第三步,在两个类别间交换同样数量的样本,使 c 增加。第四步,重复第三步,直至 c 最大为止。该计算步骤的细化流程见图 1。

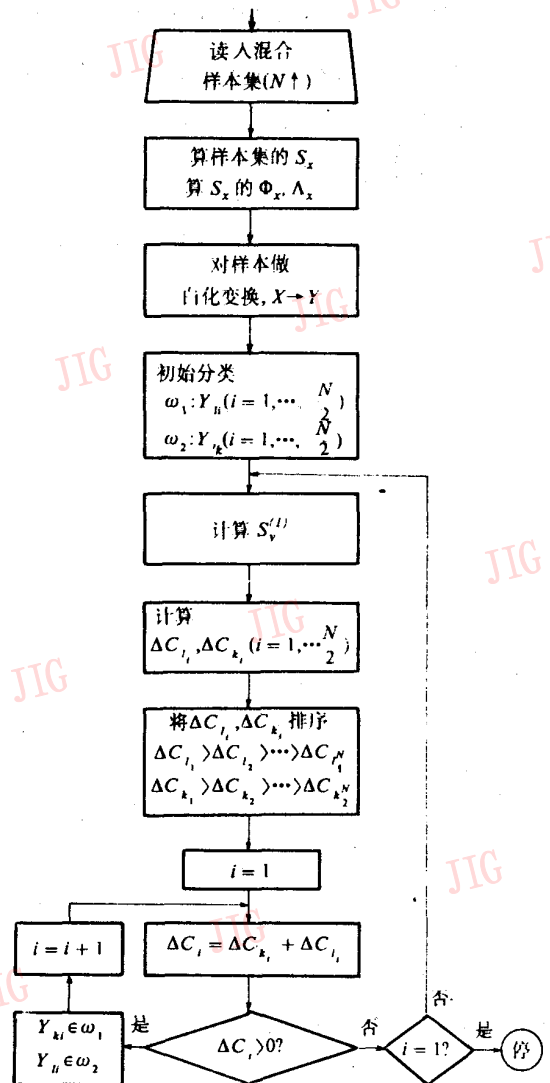


图 1 迭代聚类算法流程框图

Fig. 1 Flow chart of iterated cluster algorithm

现证明,以上算法收敛,即 c 终能达到最大值。

$$\text{由 } \Delta C_i = \Delta C_{k_i} + \Delta C_{l_i} \quad (i = 1, \dots, \frac{N}{2})$$

及 $\Delta C_{k_1} \geq \Delta C_{k_2} \geq \dots \geq \Delta C_{k_{\frac{N}{2}}}$

$$\Delta C_{l_1} \geq \Delta C_{l_2} \geq \dots \geq \Delta C_{l_{\frac{N}{2}}}$$

知: $\Delta C_1 \geq \Delta C_2 \geq \dots \geq \Delta C_{\frac{N}{2}}$

从分到第一类的样本集中任选 m 个样本

$$y_{l_{i_1}}, \dots, y_{l_{i_m}} \quad i_p \in I = \{i_1, \dots, i_m\}$$

$$P = 1, \dots, m$$

再从另一类样本集中任取 m 个样本

$$y_{k_{j_1}}, \dots, y_{k_{j_m}} \quad j_p \in J = \{j_1, \dots, j_m\}$$

$$p = 1, \dots, m$$

这 $2m$ 个样本交换,致使

$$\Delta c = \sum_{i \in J} \Delta C_{k_i} + \sum_{i \in I} \Delta C_{l_i} \leq m (\Delta C_{k_1} + \Delta C_{l_1})$$

由于最后停机时, $\Delta C_{k_1} + \Delta C_{l_1} \leq 0$

必有, $\Delta C \leq 0$

由于 m, I, J 都是任意的,到停机时,对任意两类之间的交换,必然有 $\Delta C \leq 0$,故此时 ΔC 必为最大值。

4 实验与讨论

在青藏高原湖泊动态遥感监测的课题中,为了弄清湖泊水域的变化,应用我们提出的识别模型比较精确地识别出了湖泊水域。并将本文提出的算法分类结果(图 2)与公认的精度较高的算法极大似然率分类方法作了对比(表 1)。

表 1 样本交换迭代聚类法与极大似然率法分类结果对比

Table 1. The classification result contrast between iterated algorithm of sample exchange and maximum likelihood method

分类方法	分类的湖泊	分类结果(水域象元数)
极大似然率	塞林湖	278816
本文方法	塞林湖	277797

表 1 的数据说明,样本交换迭代聚类方法达到了实用的精度。

当然,本方法尚属尝试和探索,有一些问题需进一步解决。如,在求解高维数矩阵的特征向量、特征值时,需要有快速算法,这是由来已久的大问题,也是大难题,非作者所能解决,好在硬件速度提高很快,可以逐渐弥补这个缺欠。另外,本文算法的程序也需进一步优化。对于两类识别到多类识别的推广,可采用序贯分析的方法。比如,对于一个多类别的混合群体,可以多次使用样本交换算法,依次分层,由

于算法的收敛性是严格的,类别的分出是可控的。



图 2 样本交换迭代聚类法分类结果

Fig. 2 The image of classification result by iterated algorithm of sample exchange

附录 1 协方差相等时特征值的计算

因为协方差阵

$$\Sigma^{(1)} = \Sigma^{(2)} = \Sigma$$

$$\frac{1}{2} S_1 = I - \frac{1}{2} (\Sigma_2 + M_2 M_2^T)$$

$$= I - \frac{1}{2} [(S_1 - M_1 M_1^T) + M_2 M_2^T]$$

$$\text{即 } S_1 = I - \frac{1}{2} (M_2 M_2^T - M_1 M_1^T)$$

设, M_1, M_2 分别是 ω_1 类, ω_2 类的平均向量,令

$$M_0 = \frac{1}{2} (M_2 + M_1), M = \frac{1}{2} (M_2 - M_1)$$

于是有:

$$S_1 = I - \frac{1}{2} [(M_0 + M)(M_0 + M)^T - (M_0 - M)(M_0 - M)^T]$$

从而,

$$S_1 \Phi^{(1)} = [I - (MM_0^T + M_0 M^T)] = \Phi^{(1)} \Lambda^{(1)}$$

整理此式可知 $\Phi^{(1)}$ 是 $(MM_0^T + M_0 M^T)$ 的特征向量阵,其特征值为阵 $I - \Lambda^{(1)}$,设其特征值为 $\mu_i, i = 1, \dots, n$, 则

$$1 - \lambda_i^{(1)} = \mu_i, \text{ 或 } \lambda_i^{(1)} = 1 - \mu_i$$

由于 $(MM_0^T + M_0 M^T)$ 的秩为 2,故其特征值应有 $n - 2$ 个为零,其余两个不为零,不妨设 $\mu_2 = \mu_3 = \dots = \mu_{n-1} = 0, \mu_1 \neq 0, \mu_n \neq 0$ 由于向量 M, M_0 不平行,夹角 $\theta \neq 0$,根据三角形余弦定理,可得:

$$\cos \theta = \frac{M_0^T M}{|M_0| |M|}$$

又根据矩阵求迹性质列出方程组:

$$\begin{cases} \mu_1 + \mu_n = 2 |M_0| |M| \cos \theta \\ \mu_1^2 + \mu_n^2 = 2(1 + \cos^2 \theta) |M_0|^2 |M|^2 \end{cases}$$

从而得到:

$$\mu_1 = |M_0| |M| (\cos \theta - 1)$$

$$\mu_n = |M_0| |M| (\cos \theta + 1)$$

附录 2 Δc 的计算

若将 ω_1 类的样本 Y_l 与 ω_2 类的样本 Y_k 交换, 则 ω_1 类自相关阵元素 S_{ij} 的增量由下式计算:

$$\begin{aligned} \Delta S_{ij} &= S'_{ij} - S_{ij} \\ &= \frac{1}{N_1} \left[\sum_{s \in J_1} y_s(t_i) y_s(t_j) - Y_l(t_i) y_l(t_j) + y_k(t_i) y_k(t_j) \right] \\ &\quad - \frac{1}{N_1} \sum_{s \in J_1} y_s(t_i) y_s(t_j) \end{aligned}$$

其中, $y_i(t_j)$ 表示第 i 个向量的第 j 个元素, J_1 表示属于 ω_1 类的向量标号集。由此可推出因交换引起的判据 c 的变化。设 c' 为交换后的判据值

$$\begin{aligned} \Delta c &= c' - c \\ &= \frac{2}{n} \left[\sum_{i=1}^n \sum_{j=1}^n S'_{ij} - 2 \sum_{i=1}^n S'_{ii} - \sum_{i=1}^n \sum_{j=1}^n S_{ij} + \sum_{i=1}^n S_{ii} \right] \\ &= \frac{2}{n} \left[\sum_{i=1}^n \sum_{j=1}^n (2S_{ij} \cdot \Delta S_{ij} + \Delta S_{ij}^2) - 2 \sum_{i=1}^n \Delta S_{ij} \right] \end{aligned}$$

当 N 相当大时, ΔS_{ij}^2 可忽略, 于是

$$\begin{aligned} \Delta c &\approx \frac{4}{n} \left[\sum_{i=1}^n \sum_{j=1}^n S_{ij} \Delta S_{ij} - \sum_{i=1}^n \Delta S_{ii} \right] \\ &= \frac{4}{nN_1} \left[\sum_{i=1}^n \sum_{j=1}^n S_{ij} Y_k(t_i) Y_k(t_j) - \sum_{i=1}^n Y_k^2(t_i) \right] \\ &\quad - \frac{4}{nN_1} \left[\sum_{i=1}^n \sum_{j=1}^n S_{ij} Y_l(t_i) Y_l(t_j) - \sum_{i=1}^n Y_l^2(t_i) \right] \\ &= \frac{4}{nN_1} [Y_k^T S_1 Y_k - Y_k^T Y_k] \\ &\quad - \frac{1}{nN_1} [Y_l^T S_1 Y_l - Y_l^T Y_l] \\ &= \Delta C_k + \Delta C_l \end{aligned}$$

参 考 文 献

- [1] T. Marill and DM Green "on the effectiveness of receptors in recognition System" *IEEE Trans. Information Theory* Vol 17 -9. Januavy 1963.
- [2] 程民德等. 图象识别导论. 上海: 上海科技出版社, 1982.
- [3] K. Fukunaga "Introduction to statistical pattern recognition", 1979.



连石柱, 中国科学院遥感应用研究所计算中心主任, 高级工程师。1968年毕业于南开大学数学系。1981年中国科学技术大学研究生院计算机图象处理专业毕业, 获理学硕士学位。1987—1988年, 公派到英国 BRISTOL 大学计算机系进修。长期以来主要从事计算方法、图象处理和软件研究工作。

A Method of Feature Selection and Classification Based on Divergence

Lian ShiZhu

(Institute of Remote sensing for Application, CAS, Beijing, 100101)

Abstract The separability of two pattern classes can be measured by the divergence for two Gaussian distributions. Because the divergence is invariant under the linear transformation we can extract "good" features for separating two patterns via Karhunen-Loève transformation. It is shown that the divergence is only dependent on two of n eigenvalues. One property of a "good" dichotomy is that each feature should be effective for classification. Thus, a criterion function is proposed.

This algorithm, which is called as sample exchange algorithm, is convergent and it is a reasonable unsupervised clustering method for classification.

Keywords Feature selection, Mixture normalization, Sample autocorrelation matrix, K-L expansion, Eigenvalue, Eigenvector, Divergence, Iteration, Clustering