图 样 识 别

I. G. Liddell

（英国海军水下武器中心）

一、引 言

在很多领域内都出现图样识别问题，例如：医学（血球分类、癌细胞分类等等）；通讯（自动语言识别及字码—数码识别）；气象（天气图识别）；刑事科学（指纹识别）以及军事科学（摄影图片分析、雷达与声纳目标信号的自动识别，其他等等）。

尽管上述问题以及许多其他问题在近年来受到极大的注意，然而进展仍很缓慢。获得充分成功足以在实际环境中被采用的系统为数甚少。而在这种系统中其设计主要是靠设计者的经验和直观。某些情况下是由于图样本身具有明显差别使问题变得很简单才获得成功的。通常这种情况产生于图样的形式能够被设计者加以控制，例如银行支票的签字的识别。但是大部分问题对于设计者来说都是相当困难的。甚至某些比较简单的问题，例如打字字母的识别问题，也沒有得到很大的成功。字母的正确识别率很少超过90%，对大多数实际应用来说这是不能令人满意的，即使这样的水平也还是需要设计者动很多脑筋和试验才可能达到。在这种意义上，把图样识别看作为艺术比看作为科学或许更恰当些。

广泛地说，在图形识别中可遵循两种显然不同的途径：一个是解析方法，此时研究者试图使用统计决策理论、判别式分析等手段以使分类方案公式化；另一个是直观方法，由感知单元（perception）或其他许多模仿人类感知和学习过程的学习机或自适应机来具体化。这两种方法是分别独立发展的，然而就其结果而言可能差别不大。从长远观点来看，图形识别的前途要依赖于支撑其发展的理论框架（基础）的坚实性和适用性，而沿着目前已想到的习训机的路子不大可能获得满意的通用的识别装置。一个通用机必须具备十分广泛的自适应性。例如，在字母识别问题中，判断某一字是字母“O”时应与字母大小无关；而在识别不同管径的圆圈时则显然要求有不同的程序。另外的例子是，识别停机坪上的飞机，此时识别装置必然对图形的旋转不敏感；而反过来，在字母识别中如果不对图形的旋转加以约制的话几乎不可能有效地进行识别。从概念上来说，这种灵活性总是可能的，然而实际上常常是难以克服的，因而就目前来说致力于解决某一特殊识别问题的方法是适宜的。

1．问题的性质

未知图样可由一个N-维空间的矢量或测量数组来描述：

$$X_N=(x_1, x_2, \cdots, x_N)$$

图形识别问题就是要确定这一特定的样本是属于Q类不同图样的哪一类。解决这一问题的办法是估计每一类成员的概率分布函数，然后确定一个最佳分类准则——典型的是使总错误概率最小。

为解决这一问题作了大量工作，但是对于输入图样矢量本身却很少进行系统的研究。矢量X_N的选取是根据直观，并受到各种不同的实际约制，诸如实时操作，或使用有限分辨能力的特定的接收器等。综合所有这些因素，将导致包含于图样矢量中的鉴别信息少于包含于源图样中的鉴别信息。慎重选择图样矢量空间的重要性不能不特别强调，因为在任意选择矢量空间时所损失的鉴别潜力，不管识别过程作得如何精细也是无法补救的。在许多場合中，人类自己的巨大的图样识别能力可以用来挑选哪些图样矢量可以不至引起较高的信息损失。所有这类挑选尽管有用，但均属于定性的，并且不能满足一个较为有理论基础的方法在选择测量空间时的需要，也不能满足对不同的选择之间的相对信息损失进行比较的需要。

为了在下面的讨论中明确起见，对于选取适当的图样矢量空间的问题将使用术语“图样描述”（Pattern Description）（见图1）。而术语“图样识别”或“分类”将用于其后的鉴别阶段，在此阶段中与每一特定矢量相关联的测量被用来指定此图样隶属于Q类中的一类。总的过程叫作“图样分析”。在图1中，图样描述阶段又被分为两步。第一步形成一个图样矢量，它（理想地）包含所有源图样的信息。然后此矢量经受一个降维过程或预处理以适应实际的处理时间要求和存贮要求等，由此产生一个修正图样矢量作为分类器的输入。

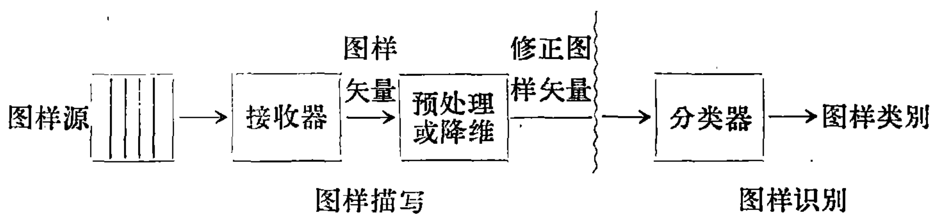

图1　图样分析器框图

二、图样识别（分类）

在本节，我们来考虑图1中所示的最后一步的操作。假定一个合适的修正图样矢量Y_M已经确定：

$$Y_M=(y_1,\ y_2,\ \cdots\cdots,\ y_M)=F(x_1,\ x_2,\ \cdots\cdots,\ x_N)$$

其中$(x_1,\ x_2,\ \cdots\cdots,\ x_N)$是修正前的图样矢量$X_N$。

下面考虑的问题是确定Y_M属于两个可能类别中的一类。按照Bayes判决准则，如果满足下述条件，则指定图样属于第1类：

$$\frac{P_1(Y_M)}{P_2(Y_M)}\geqslant\frac{C_2}{C_1}\cdot\frac{P_2}{P_1} \tag{1}$$

其中$P_1(Y_M)$是Y_M属于第1类时的概率分布，$P_2(Y_M)$是Y_M属于第2类时的概率分布，P_1是观察到第1类成员的先验概率，P_2是观察到第2类成员的先验概率，C_1是对第1类成员错误识别时的代价，C_2是对第2类成员错误识别时的代价，并且假定正确识别时代价为零。

在许多情况下，P_1和P_2都不能精确已知。此外，错误识别的代价事实上不好选定或者它本身就是变化不定的。在雷达和声纳中就是这种情况，因此通常是将复合量C_2P_2/C_1P_1

看作为一个主观判决阈。

如果 $P_1(Y_M)$ 和 $P_2(Y_M)$ 可以表为M-维多变量高斯分布[1]，它们取如下形式：

$$P(Y_M)=\frac{1}{(2\pi)^{M/2}|K|^{1/2}}\exp\left[-\frac{1}{2}(Y_M-\overline{Y}_M)^T K^{-1}(Y_M-\overline{Y}_M)\right] \tag{2}$$

其中$\overline{Y}_M$是分量平均值矢量，K是协方差矩阵。

取对数得到：

$$\log P(Y_M)=-\frac{1}{2}(Y_M-\overline{Y}_M)^T K^{-1}(Y_M-\overline{Y}_M)+\text{常数} \tag{3}$$

Bayes判决表面是由超二次型函数确定的曲面：

$$\log\frac{P_1(Y_M)}{P_2(Y_M)}=\log\frac{C_2P_2}{C_1P_1} \tag{4}$$

即：

$$\text{常数}+\frac{1}{2}(Y_{M2}-\overline{Y}_{M2})^T K_2^{-1}(Y_{M2}-\overline{Y}_{M2})-\frac{1}{2}(Y_{M1}-\overline{Y}_{M1})^T K_1^{-1}(Y_{M1}-\overline{Y}_{M1})$$

$$=\log\frac{C_2P_2}{C_1P_1}$$

推广到Q类情况，则当满足下式时，将输入图样判定为属于第k类：

$$\sum_{\substack{i=1\\ i\neq k}}^{Q}P_iP_i(Y_M)C_{ki}<\sum_{\substack{i=1\\ i\neq j}}^{Q}P_iP_i(Y_M)C_{ji} \tag{5}$$

其中$j=1,\ \cdots,\ k-1,\ k+1,\cdots,Q$，而$C_{ki}$是将第$i$类成员错误判别为第$k$类成员时的代价。

在许多实际情况下，概率分布函数的形式是未知的。Sebestyen及其他人[1,2]使用了判别式分析技术(鉴别量分析技术)，在对于各个图样类具有不等协方差矩阵的一般情况下能导致与Bayes 判决理论方法相同的结果。对于等协方差矩阵的特殊情况，则有判决理论与 Fisher 线性鉴别[3]的相似等价性。Fisher使用了测量的线性函数：

$$Y=\sum_{i=1}^{M}a_iy_i \tag{6}$$

它使两类Y的平均值之差除以Y的方差（假定两类具有相同方差）这个量达到极大。

人们对于线性方法寄以很大兴趣，因为它具有计算上的明显的优越性。一般而言，即使各类互不重叠也沒有得到很成功的结果，除非各图样能在测量空间中很好的分离开并且图样矢量的维数很大于类别数。在理想情况下，在测量空间中各类是由单一的点来代表的，此时可由线性方法作到完全地相互分隔开，只要类别数比维数至少多 1 。换言之，任何单独一个类别或一组类别均可线性地与其余者相分隔开。当各类不再是一个点而开始围绕其均值点分散时（如图 2 所示），则到一定程度时（依赖于测量空间中类内集群和类间集群的可分性）用线性分割就变为不可能了（如图 2 (c)）。如果还有未被使用的测量维，换言之可以增加维数，

则线性分隔仍是可能的，只要图样在测量空间是处在“一般位置”上*。否则，必须使用非线性方法。

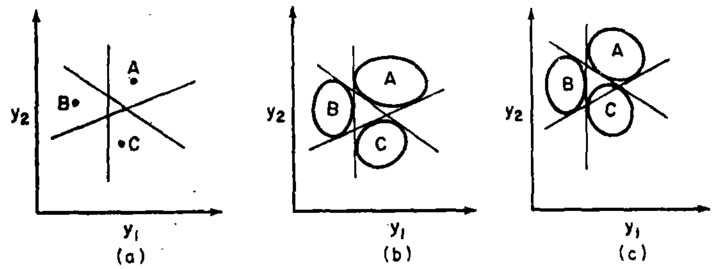

图 2　在m维（m＝2）空间中分隔m＋1类.（a)点类；（b)和(c)分散类

对于不可能进行线性分离的情况，甚至是各类不互相重叠的情况，建立了各种非线性方法。其中的一种方法是用“最紧邻分类器”（Nearest Neighbour Classifier），它将新输入的图样与每一类别的一组图样样板（已知的图样“典型”）进行检验。判决标准是看它与哪一个样板点最靠近就指定它是属于哪一类。换言之，图样矢量Y_M被识别为A类，如果对所有Q类满足：

$$\min_{\text{对所有的}u}\left[\sum_{i=1}^{M}(y_i-a_{ui})^2\right]<\min_{\text{对所有}v}\left[\sum_{i=1}^{M}(y_i-b_{vi})^2\right]$$

$$<\min_{\text{对所有}w}\left[\sum_{i=1}^{M}(y_i-C_{wi})^2\right]\cdots\cdots\text{等等} \tag{7}$$

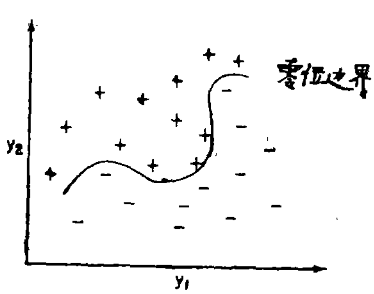

图 3　位函数法

其中y_i为新图样矢量在第i维的测量值，而a_{ui}，b_{vi}为各样板点相应的值。除了在存贮上明显地有问题以外，上述方法还有一个缺点，就是对于样板点集当中的所谓“流散”样点非常敏感。

克服这一困难可以用更为细致的方法，如位函数法[4]。图3中以二维问题简释了位函数的原理。图中，两类别的样点由测量空间中的正负电荷来代表。据此，可以划出空间中零位线。

三、图 样 描 写

讨论到此，已对图样分类技术的一小部分作了简单的介绍。然而，正如早已指出的那样，这种技术在实际环境中能否成功地使用还要依赖于描写原始图样的矢量所载荷信息量的多少。显然，如果根据随便选取的数据，其中主要鉴别信息均被弃置，则盲目地进行各种可能的分类过程是不会成功的。然而这种情况并不是没有的，结果是造成巨大浪费。尽管图样描写问题

*指不共线或共面。——译注

是十分困难的课题，但作者坚信，将来会针对图样描写的问题连同有关的数据的降维问题以及信息含量的评价问题作出巨大的努力。

可用来解决某些这类问题的一个可能的途径是所谓的“主分量分析”技术[5]，它是在多变量统计中用以降维的方法。图4中，以二维为例对此技术作了示意。图中的点子代表在y_1，y_2空间中的一组测量，由图4(a)可见，这些点子之间存在有十分理想的正相关性。因此，可能定义一个单一的新变量z_1，它完全可给出这些数据的特性。此变量称为第一主分量。在图4(b)中，点子稍有分散，不再严格位于一条直线上但仍然具有高度相关性。

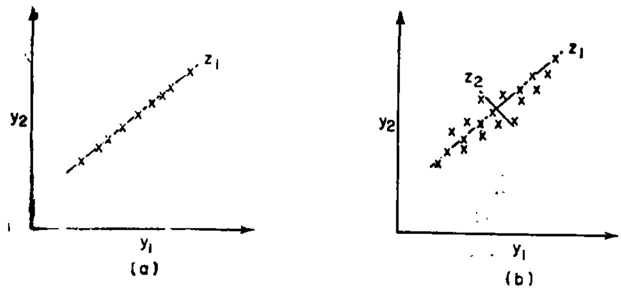

图4　主分量分析

第一主分量z_1，根据下式系数的求得而确定：

$$z_1=a_1y_1+a_2y_2 \tag{8}$$

求定系数的原则是使z_1方差最大。

第二主分量z_2，可由与z_1正交，即：

$$z_2=b_1y_1+b_2y_2$$

加上不相关条件$(z_1, z_2)=0$和z_2方差最大而确定。对多维问题的处理，可由上述方法直接推广。在所示的这种较简单的情况中，有可能第二主分量只具有很小的方差，以致可以只用第一主分量来近似描写数据，这样就将维数由2降为1。当涉及多维问题时，往往采用前五个左右的主分量就可获得较好的近似。

当然，这样大幅度的简化是十分吸引人的，但是对于实际的图样描写问题所涉及的大维数，必须为这种主分量分析付出巨大代价，即在分析过程中要求巨大的存贮量和运算量。主分量是根据考察测量数据组的方差—协方差矩阵并计算其本征矢量及本征值来确定的。前者确定原始测量的线性组合系数，由它给出主分量；而后者确定每一主分量对总的方差的贡献。在大多数实际的图样识别问题中包含很大的维数，在某些情况中甚至会超过10^6。处理这样大的矩阵不是轻而易举的，但如果能对维数作大量削减从而可能获得很大收益的话也还是值得的。克服此困难的一个简单办法当然是以牺牲分辨率来缩减原始图样的维数。这样作的风险也是熟知的，即降低分辨可能导致关键的鉴别信息的丢失。

为了在使用这种技术时能实现完整的图样分析，必须随后应用某种形式的判别式分析。其中的技术之一就是对来自所有各类图样的混合数据进行主分量分析，然后对包含了主要离散性的这些分量进行多变量判别式分析。

在图样描述中还存在另外一个问题，就是实际的图样类别的数目未知。在声纳显示屏上观察到的回波信号以及各种来源的噪声信号就是这种情况的最好例子。虽然，将数据分为“目标回波”及“背景事件”两种类别是方便的，但无疑的在后者中将包括许多未被定义的“子类别”(sub-classes)，因而存在这样的实际危险性，即选择一个代表性不够的背景子集为基础来进行图样分析。在很短的时间间隔和操作条件几乎不变的情况下，观察到的背景事件的数目也可能是非常之大的，对它们进行仔细考察所需的时间是非常可观的，并且可能是浪费的。

针对这一问题，可以使用集群分析(Cluster analysis)技术[6]，它将在某种最佳意义

下对测量空间进行分割。我们来考虑图5示出的一维和二维问题的简单情况。

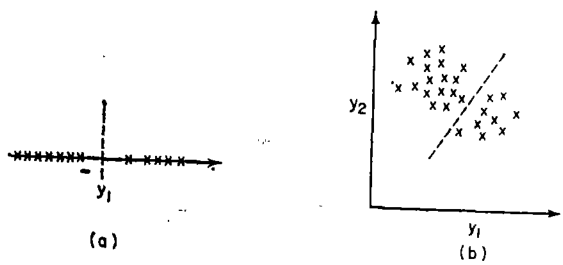

图5 在一维和二维空间中图样的集群

在一维的情况下，数据可设想分为两组。考虑所有可能的划分方法，并计算每一划分时相应的这两组的方差和（以每组的成员数为权重），从而找到使这个方差和最小的特定划分。在图中以虚线标出这种划分，它代表在使两组的重心距离最大意义下的最佳分割。图5(b)给出二维空间中类似的情况。然而当维数增加时，计算量迅速增加。此时，还需再借助于主分量分析法来实现降维。

如果发现在未知数据中存在着某种能彼此分开的群集现象，则可以采用所谓“链”集群法（“Chain” method of clustering）[7]。使用此技术时，要计算第一个样本和第二个样本的距离。如果此距离小于某一预先取定的最小距离时，则将第二样本作为第一集群的一部分。否则，第二样本形成第二集群的基础。此过程连续进行下去，则每一样本根据最小距离准则，不是加入某一存在的集群中去，就是作为新生集群第一个成员。

1. 附加信息

在许多实际应用中，包含在一个单个图样中的信息可能不足以获得十分低的误识率，即使图样矢量可能已包括了源图样的所有鉴别信息。在文字字母识别中就有这种情况，人能够几乎正确无误的识别是因为他能利用相邻图样之间存在的关系，即语义。在声纳和雷达中，也可以利用时域上的图样之间的关系，即跟踪，来改进对目标与非目标的分类置信度。

在图样描写上其他有关的问题是对于平移、标度（放大缩小）及旋转的不变性问题，这些已经在本文的引言中提到过。对于平移不变性的要求几乎在所有的应用中都是需要的；即不管图样是字母、数字也好，血球也好，雷达回波也好，指纹也好，都要求分类器保持其不变的性能而不论这些图样出现在观察場上的位置如何。在实际应用中，这一问题常靠“手控”中心化来解决（虽然自动化技术已开展），例如，重力中心或棱边注记。

当涉及到标度及旋转不变性时，将按照具体任务而有不同要求。象平移情况一样，手控或自动技术均可利用。螺线扫描技术和样板的手控旋转就是取得旋转不变性的手段的例子。标度不变性可借助于利用图样描述中对有关测量的标度比的关系来解决。

四、试　　验

在找寻一个图样分析的满意的方法的过程中，对大多数实际应用来说，进行大量的尝试常是必要的。因此找寻适当的试验过程是特别重要的，而且它也是图样分析的组成部分，后者往往需要大量的人力和时间。

可能最简单的办法是测量平均的图样误识百分率，并且在许多情况下（例如，自动数码识别器）这一技术是可行的。尽管它十分简单，然而它也有弊病，即一个单一的总括的因数

可能会使得对某一或某些特定类别的识别性能很差。特别是，当各个类别中图样成员数目不一以及误识代价不相同的情况下，使用这一办法可能导致严重错误。在这种情况下，一个可以采取的技术是“混淆矩阵”法（Confusion Matrix），它把各类的正确分类数与误识数均在一个矩阵中表示出来。每当属于第i类的一个图样被错误分到第j类时，则在矩阵中的第ij元素的数目x就增加 1 。这样，数目x_{ij}当$i=j$时（构成对角线元素）就代表各类的正确识别数目，而其余的，即$i\neq j$，均代表误识。

另外一个有用的方法是从通讯论中借用来的，即所谓接收机工作特性曲线(ROC)。在此方法中，由对A类正确识别概率相对于将其它各类误识为A类概率作曲线。这办法对“两类别”问题特别有用，即只存在两个类别或者只需将某一类同其余各类分开而其余各类之间无须再分的情况。前面提到的声纳回波信号的识别以及癌细胞的识别就是可以使用这一办法的最好例子。在这两种情况中，特别重要的类别的图样(目标、癌细胞)的数目较之构成余类的非目标及健康细胞的数目要少好几个数量级。在这两种情况中误识代价都是不同的。显然，将健康细胞误识为癌细胞要比反过来的损失小得多。类似的，目标漏报的代价要大于将背景事件误识为潜艇的代价。

利用ROC还有一点是非常可取的，就是它可以对不同的识别方法进行相互比较。随机分类器的性能是由过原点及（1，1）点的对角线表示的。优于随机分类器的性能则应由位于对角线之上的曲线来表示。对于两类以上的问题，则可绘出一系列ROC曲线，来表示对每一类正确识别概率与对余类误识概率之间的关系。然而，当类别数目较多时，想要得出总的结论将变得很困难。

为了使对错误率的估计获得足够高的置信度，显然应使试验图样具有代表性。只要有可能，就应当搞类别的随机抽样。即使如此，本文作者仍然认为，被某些研究人员所采用的以相同的“代表性”抽样进行设计然后再用于试验的办法是危险的。正如已提到的，在声纳信号数据中，想要获得一个有限的能有代表性的背景样本数据可能是十分困难的，因为由未确定的类别组将引起一些可能的次级结构。

另外一个问题是，图样描写阶段对于諸如平移、放大和旋转是敏感的，甚至是过于敏感的。对于不同的问题应对此加以不同的约制条件，因而给定的系统的性能将依赖于这些约制条件如何给法。

五、结　　论

前面列举了图样分析的某些基本原理。只着重于那些有一定理论基础的方法，远不是全面的。作者认为无论是对于总的问题还是对于已存在的理论的了解都还是很不够的。

我们特别强调了图样描写问题（包括降维问题）需要一个更具逻辑性的方法。普遍适用的解决办法看来在很长时间内都未必能得到，而对于某些特定问题或一组类似的问题，在方法上的定量化和理论化，只对目前是有益的。

关于通用性问题，作者认为，如许多研究者所寻求的那样，构造一种能够在图样分析上具有可与人相比的能力的通用型式的学习机还是为期遥远的。不同的图样识别任务具有不同的问题：某些任务要求不必管旋转；另外的某些任务可能只关心旋转问题；某些问题要求能有标度不变性的识别，等等。对于识别手写字母的识别装置很难企望它能与人竞争，除非它

能象人一样利用语义，即利用在相邻字母间的群相似性或利用对于所遇到的字母的某些先验知识。例如在确定邮寄地址时，工作人员靠其任务特点知道他会遇到某些类型的字母组（城市、国家，等等）。至于通用学习机在原理上不是不行，但这种机器要比目前试图构造的最精密的系统要复杂好几个量级。因此，通过对基本原理的深入了解而致力于建立一种能解决特定问题的系统，从而取得缓慢的更富于逻辑的进展似乎更为有利。

参 考 文 献

1. N.J.Nilsson. “Learning machines.” 1965.
2. G.S.Sebestyen. “Decision making processes in pattern recognition”. 1962.
3. R.A.Fisher. “Contributions to mathematical statistics.” 1952.
4. Arkadev & Braverman. “Teaching computers to recognise patterns”. 1967.
5. D.F.Morrison. “Multivariate statistical methods”. 1967.
6. R.R.Sokal & P.H.A.Sneath. “Principles of numerical taxonomy”. 1963.
7. R.E.Bonner. “On some clustering techniques.” IBM J.Res. Dev. 6, p.353, 1962.

討 論

M.G.Vezzosi指出了在图样识别的统计判决理论方法与多重假设检验之间的十分相似性，并问到是否可以从Neyman-Pearson观点出发而不用Bayes观点。

回答：后者只不过作为例子提出。可以有多种方法来解决问题，但本文的目的不在于搞最佳识别方法，而在于强调这样一个事实，即找到一个能够不至引起过高的信息损失的合适的对输入图样矢量的描写是当前的主要问题。

T.G.Birdsall指出，自动图样识别中的许多问题看来仍是由于对图样特性缺少经验而引起的，声纳目标问题尤其是如此。

回答：作者同意上述观点，但指出，无论是对图样信息含量的评价方面或是确定图样描写方面的技术发展都希望它能不依赖于人的感觉。虽然如此，在对任一特殊问题上目前的办法仍然必须利用人对图样的经验。

P.Stocklin指出，信号处理通常分为下列过程：开始是能量检测，随后是估计和分类。并提出类似的观点对图样识别是否也适用。

作者认为在大部分问题中上述途径基本上都是一样的。虽然在整个处理过程的早期阶段希望利用尽可能多的信息，然而过程本身的性质可能不允许利用全部信息。例如，对于声纳，在最广泛的意义下的形式（shape）应包含时域，即多脉冲发射信息。然而实际上，首先是使用一个单个脉冲，并确定其对一给定事件的似然比，然后再由相继的脉冲发射所获得的信息进行补充修正。

（尚尔昌译　黄曾旸校）