

目标分类和假设检验

L.Meier

(北约萨克兰特反潜战研究中心)

摘要 本文提出了能说明所有目标分类问题的一种简单的目标分类问题，并给出了它的解。讨论了由于不知道目标性质而对于分类所产生的各种问题及其解决办法。最后把这些概念推广到动态的情况。

一、引言

图 1 是目标分类问题的方框图。譬如在一个声纳系统中，测量过程可以由声纳信号的传

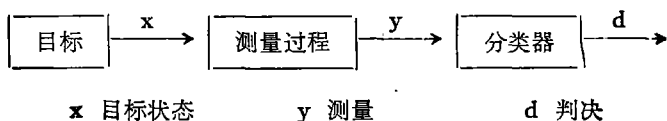


图 1 分类问题的方框图

播、接收、检测和测量的提取所组成。与目标分类密切相关的是目标检测。在检测中，人们希望把目标从噪声中分离出来，而在分类问题中，人们则要把真目标既同噪声又同假目标分开。在一个实际系统中，在分类之前可以有也可以没有检测。譬如，在声纳系统中，最佳分类器可以对各种未经加工的声纳回波发生作用—使用检测器和测量提取器只是使问题较为易于处理。

假设检验的统计理论已被证明在检测的理论和实践中是非常有用的^[1]。在本文中，把这种应用推广以包括目标分类问题。分类区别于检测的一个方面是，在分类问题中可以包含多于两个的假设。在假设检验的文献中大多数是关于两个假设的工作，但是多假设问题也并没有被忽视^[2]。在本文中，把这些结果应用于分类问题。更重要的是，在目标分类问题中，对于某些假设，其参数可能不是很好地确定的一特别是某些类的假目标的统计参数可能实际上是不知道的。本文提出称为势函数的统计概念作为解决由于未知统计参数所引起的困难的一种方法。最后给出一个例子来说明所提出的某些概念。

* 原文题目：“Target Classification and Hypothesis Testing.”

二、分类问题

在这一节中定义一种简单的分类问题，给出它的解，并讨论由于目标性质的不确定性所引起的各种困难。对于更为一般的分类问题，也不难给以形式上的表述，但是这里没有时间和篇幅去作讨论。幸而一般来说，对于这里给出的那些结果进行推广是相当明显的。

1. 问题的表述

图 1 是分类问题的方框图。目前将只考虑静态的情况，即所有测量都同时作出并是此刻目标状态的函数。关于这些测量存在三种假设，即它们是下列三种情况之一的结果：光是噪声(H_0)，假目标(H_1)，或真目标(H_2)。分类器必须确定 H_2 是否为真。

存在三种错误概率：漏分类概率(当 H_2 为真时说它不为真)；光有噪声时的虚警概率(当 H_0 为真时说 H_2 为真)；和假目标时的虚警概率(当 H_1 为真时说 H_2 为真)。可以以两种方式把它们结合入性能准则：在纽曼—皮尔逊方法中，在两个概率受到限制的情况下使第三个概率为最小；在贝叶斯方法中，使这三个概率的一个线性组合为最小。

2. 问题的解

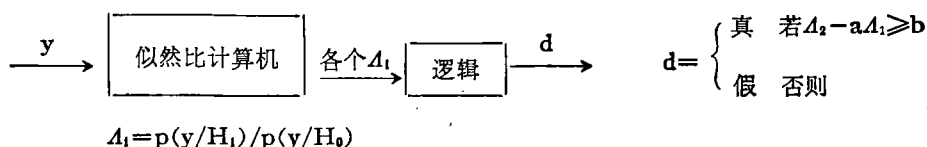


图 2 最佳分类器的方框图

图 2 是对两种性能准则都为最佳的分类器的方框图。参数 a 和 b 必须选得使所用的准则为最佳。对这个结果的证明既是标准的^[2]又是简单的，所以在此不再重复。为了计算似然性 $p(y/H_i)$, $i=0, 1, 2$ ，如果把目标的模型取成概率密度 $p(x/\bar{H}_i)$, $i=1, 2$ 的形式，并把测量过程的模型取成概率密度 $p(y/H_0)$ 及 $p(y/x, \bar{H}_0)$ (其中 $\bar{H}_0$ 的意思是非 H_0) 的形式是方便的，于是 $p(y/H_i)$, $i=1, 2$ ，由下式给出：

$$p(y/H_i) = \int_{\mathbf{x}} p(y/\mathbf{x}, \bar{H}_0) p(\mathbf{x}/H_i) d\mathbf{x} \quad (2.1)$$

3. 未知目标性质

一般来说， $p(y/H_0)$ 和 $p(y/\mathbf{x}, \bar{H}_0)$ 或多或少是由测量过程的物理特性确定的。另一方面，要确定 $p(\mathbf{x}/H_1)$ 可不是那么容易。对于真目标（譬如潜艇）和某些类型的假目标（譬如模拟器）， $p(\mathbf{x}/H_1)$ 的形式不太难确定，可以以某种置信度来选择合适的参数。但是未知的假目标（譬如鱼群）的 $p(\mathbf{x}/H_1)$ 的形式又是什么呢？一种合理的假定是，这种假目标的 $p(\mathbf{x}/H_1)$ 的形式与 $p(\mathbf{x}/H_2)$ 相类似（否则它们将不会同真目标那么容易混淆），即 $p(\mathbf{x}/H_1) = p(\mathbf{x}/\alpha_1)$, $i=1, 2$ ，其中 α 为一适当的参数矢量。选择统计参数值 α_i 是一个复杂得多的任务。

零假设比替换假设更好确定的情况在统计学中并不是不常见的。在这种情况下，势函数

非常有用^[3]。在我们的问题中, 势曲线就是当给定 $\mathbf{x}=\xi$ 时判决为真目标存在的概率与 ξ 的关系曲线。对于一个给定的分类器, 这样一条曲线只是 $p(\mathbf{y}/\mathbf{x}, \bar{H}_0)$ 的函数, 因此比作为系统性能的量度的三种错误概率的值更为可靠。用一个例子能使势曲线的利用得到更好的说明。

三、例 子

为了说明这些概念, 我们来考虑这样一种情况: 声纳发射一个长的连续波 (LCW) 脉冲信号来探测和识别一条潜艇。

1. 说明

在用 LCW 信号时, 距离变化率 $\dot{r}$ 和标称回波强度 (ES)_{标称} 是可以测得的。(前者由回波的频率偏移得到, 后者从测得的 S/N 比除以标称的零目标强度 S/N 比得到。) 包含在这些测量中的目标状态变量为: 目标速度 v 和平均目标强度 (TS)_{平均}。作下列假定:

- (1) 距离变化率的测量误差可以忽略。
- (2) 所有的目标舷角都是等可能性的。
- (3) 测得的回波强度是指数分布, 在给定目标强度时与舷角无关。
- (4) 目标速度在 v_{MIN} 与 v_{MAX} 之间均匀分布。
- (5) 平均目标强度在 (TS)_{平均 MIN} 与 (TS)_{平均 MAX} 之间均匀分布。
- (6) 检测阈为 $|\dot{r}| > 1.5$ 及 (TS)_{标称} > 1 。
- (7) S/N 足够高, 以致只有目标被检测到。

假定(2)意味着 $\dot{r}$ 可以用 $|\dot{r}|$ 来代替。对此例子, 显然有

$$\mathbf{x} = \begin{bmatrix} v \\ (\text{TS})_{\text{平均}} \end{bmatrix}; \quad \mathbf{y} = \begin{bmatrix} |\dot{r}| \\ (\text{ES})_{\text{标称}} \end{bmatrix}; \quad \alpha = \begin{bmatrix} v_{\text{MIN}} \\ v_{\text{MAX}} \\ (\text{TS})_{\text{平均 MIN}} \\ (\text{TS})_{\text{平均 MAX}} \end{bmatrix}$$

进而, 假定(1)到(3)确定 $p(\mathbf{y}/\mathbf{x}, \bar{H}_0)$, 假定(4)和(5)确定 $p(\mathbf{x}/\alpha)$ 。

2. 解

因为假定(7), $p(\mathbf{y}/H_0) = 0$, 因此图 1 的判决规则可以由下式代替:

$$\lambda'_{\text{FT}} = l_n p(\mathbf{y}/H_1) / p(\mathbf{y}/H_2) \leq \lambda_{\text{TH}} \rightarrow \text{潜艇存在} \quad (3.2)$$

似然性 $p(\mathbf{y}/H_i)$, $i=1, 2$ 的精确计算是相当复杂的, 所以作了下列简化假定: $|\dot{r}|$ 是正态分布, (ES)_{标称} 为指数分布。给定了这些假设, 就有

$$\lambda'_{\text{FT}}(|\dot{r}|, \text{ES}_{\text{标称}}) = C_r (|\dot{r}| - |\mathbf{r}|)^2 - C_{\text{ES}} (\text{ES})_{\text{标称}} \quad (3.3)$$

3. 模拟结果

本文作了这个例子的蒙特—卡罗模拟。假定已知 $v_{\text{MIN}}^{\text{SUB}} = 6$, $v_{\text{MAX}}^{\text{SUB}} = 10$, $(\text{TS})_{\text{平均 MIN}}^{\text{SUB}} = 12.6$, $(\text{TS})_{\text{平均 MAX}}^{\text{SUB}} = 200$ 。对于假目标, 这些相应的值则是不知道的, 因此假定假目标速度的变化范围

尽可能地大，而假目标强度小于潜艇目标强度： $v_{MIN}^{FT}=0, v_{MAX}^{FT}=16, TS_{MIN}^{FT}=0.3, TS_{MAX}^{FT}=63$ 。最后，用下列似然参数： $|i|=7, C_i=0.02, C_{TS}=0.03$ 。注意，当分类器是根据简化来设计的时候，是用实际分布来产生分类器输入的。结果汇总在图3和4上。图3是熟知的ROC： $P_C=0.866$ 是最高的分类概率，它是在所有检测到的回波都被识别为潜艇时得到的。画在图4上的 $\lambda_{TH}=0.326$ 的势函数仅仅依赖于测量过程的模型和分类器的设计，而与目标模型无关。另一方面，ROC除了其他因素以外却是依赖于目标模型的，因此不太可靠。

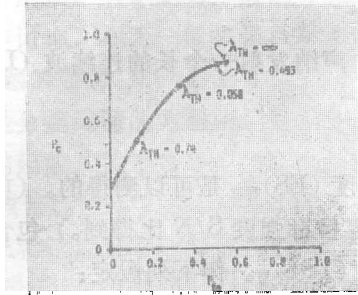


图 3 对于例子的ROC

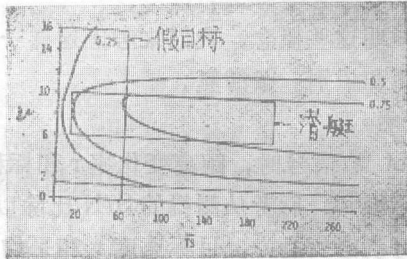


图 4 对于 $\lambda_{TH}=0.326$ 的 $d=真$ 的概率的等值线

现在来研究在分类器设计中如何应用象在图4中所给出的那种势函数的问题。在这个例子中，分类器有四个参数： $C_i, |\bar{r}|, C_{TS}$ 和 λ_{TH} 。下列要求应该说是合理的：不管潜艇处于什么状态，把它分类为潜艇的概率为，譬如说，0.75。这样，对这四个参数进行调整，直到在图4中围绕标明真目标的矩形的等值线 $P(d=真)=0.75$ 为止。可能有很多组参数都能使这点成立，所以应当调节这些参数以使得 $P(d=真)=0.75$ 的等值线尽可能地与真目标的边界靠紧。在分类器参数选择方面依然可能有一定程度的任意性，在实际操作中来修正这些参数既是必要的也是可能的。另一方面，要找到任意一些参数，使得对于识别潜艇的概率的限制得到满足也许是不可能的，或者更普遍地讲，不满足这个限制的那些参数组不能消除许多假目标。如果是这样的话，必须得出结论：这个信号对于分类来说是不充分的，应当试用另一个信号。

四、推 广

上面假定所有的测量都是在同一时刻作出的。但往往各个测量可能是在一系列时刻作出的(譬如，如果发出一串声纳信号的话)，这就要求采用目标的动态模型。由描述模型的动态行为的微分方程可以得到状态作为，测量时间的函数的差分方程。这里依然采用方程(1.1)*，但是由状态方程和描述测量过程的那些方程来进行计算可能是非常困难的。另外，还存在着是否需要跟踪检测作为检测过程或者分类过程的一部分的问题。对于一给定的跟踪，可以利用对贝叶斯估计的那些方程作的小小的推广以递归方式进行计算^[4]。一般来说，即便是这些递归方程也是无望解出的。但是，在许多情况下，可以用推广的卡尔曼滤波器对它们作近似^[5]。

在动态的情况中，把关于目标模型的所有不确定性分解成原始的目标状态的不确定性是方便的。在这种情况下，势曲线变成了分类概率与原始状态之值的关系曲线。

*此处似有误，因文中未出现过方程(1.1)。——译注

五、结 论

在目标分类中利用假设检验的统计理论看起来象将其应用于目标检测一样富有成效。但是,假目标统计参数知识的相对缺乏使得分类问题远不是那么简单明了。这个困难通过考虑势函数的办法来克服,这个函数给出把一目标识别为真目标的概率作为目标状态(动态情况的原始状态)的函数。将来还有许多工作要做。现在只提几种可能性:需要处理一些更实际的例子;对于分类器和目标参数的改变的敏感程度尚需要研究(初步的迹象表明,对分类器参数的敏感程度是低的);并且应该探索在工作过程中自适应地使分类器参数最佳的途径。

(陶笃纯译)

参 考 资 料

- [1] Van Trees, H. L., "Detection, Estimation, and Modulation Theory", pt 1. New York, Wiley, 1968: pp239-418.
- [2] *ibid*: pp46-52.
- [3] Kendall, M. G. and Stuart, A., "The Advanced Theory of Statistics, Vol. 2: Inference and Relationship". New York, Hafner, 1967: pp179-184.
- [4] Ho, Y. C. and Lee, R. C. K., "A Bayesian approach to problems in stochastic estimation and control." IEEE AC-9, 1966: 333-339.
- [5] Cox, H. "On the estimation of state variables and parameters for noisy dynamic systems." IEEE AC-9, 1964: 5-12.