

# 模糊综合评判与数理统计知识结合的目标识别效果评估

李彦鹏, 黎湘, 王宏强, 庄钊文

(国防科技大学电子科学与工程学院空间电子技术研究所, 长沙 410073)

**[摘要]** 针对目标识别效果评估, 利用数理统计知识确立了评估指标及测度方法。随后, 使用模糊综合评判建立了一种开放的评估模型, 该模型可以给出客观的、量化的评估结果。最后, 结合实测数据完成了仿真。

**[关键词]** 数理统计; 模糊综合评判; 目标识别; 效果评估

**[中图分类号]** TN911.2; TP391 **[文献标识码]** A **[文章编号]** 1009-1742(2005)03-0064-05

## 1 前言

目标识别具有重大的理论意义和应用价值, 长期以来, 对目标识别效果评估的研究相当有限: a. 没有形成一种完整的、通用的评估理论; b. 成熟的、可直接应用于实际工作的评估方法非常少; c. 评估中没有将识别过程所处的条件有机地考虑进去。因此, 效果评估成为目标识别中最迫切需要解决的重要问题<sup>[1-8]</sup>。笔者将这一方向作为研究内容, 研究过程中首先找出目标识别效果评估工作思路, 而后结合恰当的数学工具建立评估模型, 最后进行仿真实验的技术路径。

目前, 模糊综合评判方法在复杂系统的多因素综合评判、模糊判决、天气预报、工农业生产中发挥了重要作用<sup>[9,10]</sup>。经考虑, 可以将这一工具引入到目标识别效果评估来解决问题。

## 2 目标识别效果评估整体思路

根据对识别系统的分析, 提出这样一种目标识别效果评估的整体思路: 在较为有力地刻画识别系统所处条件的基础上, 利用客观的、有效的评估模型得出定量的评估结果。评估模型必须具备较强的

通用性, 还要能够直接用于工程实践。主要步骤体现在:

1) 确立评估指标 根据目标识别中需要关注的内容, 从实验中可以获得的数据出发, 提炼出某些能够反映识别效果的量, 给出其准确的定义及计算方法。效果评估中就是通过考察这些评估指标来工作的。

2) 建立评估模型 生成评估指标后, 使用恰当的数学工具以及雷达信号处理、目标识别的知识建立合理的评估模型。

识别系统对某一种或几种训练目标的数据进行处理, 判决结果正确时记为 1, 称为正确识别; 不能做出判决时记为 0, 称为不能识别; 同时对各种错误判决状态都分别进行记录。

如果对非训练目标的数据进行处理, 则记录结果的方式会有不同。正确判决是指识别系统做出不是训练目标中的任何一种的判断, 也可以说是拒判, 此时记为 1, 称为正确鉴别; 错误判决是指识别系统将该目标判为训练目标中的某一种, 此时记为 0, 称为错误鉴别, 同时对各种错误鉴别状态都分别进行记录。

**[收稿日期]** 2004-04-22; **修回日期** 2004-09-05

**[基金项目]** 国家自然科学基金资助项目(69901005); 国防预研资助项目(41303040203); 国家杰出青年资助项目(600125102)

**[作者简介]** 李彦鹏(1972-), 男, 河北衡水市人, 博士, 国防科技大学电子科学与工程学院讲师

### 3 建立目标识别效果评估的评估指标

评估指标主要是依据人们对目标识别系统所关注的热点来选择的，从重点关心的问题出发，形成反映识别效果的能够量化的指标。综合评判中因素集元素来源于评估指标。

#### 3.1 识别率的含义及求取

传统上为了比较识别系统性能的优劣，一般是求识别率。计算方法是，经过多次识别实验后，正确识别的次数除以总测试次数就是该系统在该测试条件下对该目标的识别率，也就归结为该系统在这一条件下对该目标的识别能力。

这种方法有一个隐含的假设，即正确识别时样本值是1，不能识别时样本值是0。如果该系统识别率真值是 $p_0$ ，则第 $i$ 次实验结果为1的概率为 $p_0$ 。也就是说，识别系统单次识别的识别结果是0-1分布， $n$ 次实验正确识别的次数是 $B(n, p_0)$ 分布，其参数 $n$ 即为样本容量（总测试次数）， $p_0$ 是系统识别率真值（一般以总体均值代替）。

#### 3.2 识别率测试结果的提出及其分布

3.2.1 深入理解识别率 传统上的识别率，严格来说，是识别能力的测试结果，也可以说是识别率的测试结果。它是识别系统识别能力的一个估计值，将其不加分析的归结为系统的识别能力是有误差的。显然，识别率的测试结果是一个变量，在外界条件一定的情况下，多次测算识别率测试结果时它总是会在一个范围内摆动。摆动的范围、摆动的方式以及发生较大摆动的次数多少反映了一个系统内在的处理能力。

传统的求解识别率方法至少在以下几个方面对系统识别能力的认识存在模糊性：

1) 测试需要的样本容量 在求解识别率的过程中，不同的系统和不同的外界条件下，样本容量是否需要相应的变化？按照什么样的要求进行变化？样本容量究竟如何计算？它是无法解决的；

2) 所得结果的置信度 以上做法不能够给出所得结果的置信度，同时，如果对置信度的要求有变化，这种做法也不能够给出新的估计值和置信区间；

3) 识别系统的稳定性 按照上面的做法，如果两个识别系统在某给定条件下对同一目标的识别率相等，则可以认为二者对该目标识别能力相同，这存在很多潜在问题！首先，二者原始测试结果所

反映出的系统稳定性可能不同，例如，对A, B两个识别系统在某给定条件下对同一目标进行识别，70次测试结果如表1所示。这里，A, B两系统识别体制相同，仅仅从识别率来看，二者相差甚微，从表1来看，人们可能从直观上认为二者识别的稳定性略有差异，但是如何严格地界定？传统做法不能够给出回答。

表1 A, B两个识别系统在相同条件下对同一目标的识别结果

Table 1 Recognition result of the same target by ATR system A and B

	识别结果
A	1011111110111011111011101101010111111111011110011011011111110111101
B	101011110011111011111011111111111111011110111111110111001111111011110

另外，A, B两系统原始测试结果的方差可能不一致，这一指标也反映识别系统的性能，如何区分识别率相等而识别结果的方差不相同的系统？显然，采用传统做法，无论什么识别系统，在总的测试次数相同、正确识别的次数相同时，根据方差的定义会计算出同样的方差结果，这种结论肯定太粗糙，也无法精确衡量识别系统的性能。

更进一步，假设进行了大样本测试，比如说1500次，传统方法仅仅有一个识别率，如何从这里出发获取有关系统识别结果的分布特性及稳定性的认识？它是无法解决的。

3.2.2 识别率测试结果概念的提出 解决以上问题的思路是：第一步，生成一种测量工具，用它可度量（或表征）识别系统某个方面的特性；第二步，利用生成的工具，获取被评价系统的样本；第三步，考察样本各种特性，从而获得关于系统特性的深入的认识。

这种工具就是识别率测试（MRR, measurement of recognition rate）结果。具体操作：比如说1500次测试，将总的测试分组， $1500 = 50$ （个/组） $\times 30$ （组），每组取平均得出一个MRR。MRR是一个变量，有分布、均值、方差，也可以考察与外界条件的独立性，从而能够准确评价一个系统，还可根据概率论和数理统计的有关知识严格地设计实验。

如果是对非训练目标的鉴别则成为鉴别率测试（MDR, measurement of discrimination rate）结果。

另外，利用传统识别率的概念来考察系统识别

性能, 不管测试多少次, 只要样本容量是有限个, 所得到的“识别率”实际上就是一个“识别率测试结果”。

3.2.3 MRR 呈正态分布的证明 在对识别系统的测试中满足样本容量的要求时, MRR 是呈正态分布的。将  $n$  次测试所得结果看成是  $n$  个随机变量,  $x_i, i = 1, 2, \dots, n$ 。显然,  $x_i$  独立同分布, 对于  $n$  次测试的总体来说, 是一个典型的二项分布。令  $\zeta_n = \sum_{i=1}^n x_i$ , 根据独立同分布的中心极限定理

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\zeta_n - np}{(np(1-p))^{1/2}} \leq x \right\} =$$

$$\lim_{n \rightarrow \infty} P \left\{ \left[ \sum_{i=1}^n x_i - np \right] / (np(1-p))^{1/2} \leq x \right\} =$$

$$\lim_{n \rightarrow \infty} P \left\{ \left[ \frac{1}{n} \sum_{i=1}^n x_i - p \right] / (p(1-p)/n)^{1/2} \leq x \right\} =$$

$$\int_{-\infty}^x (2\pi)^{-1/2} e^{-t^2/2} dt \quad (1)$$

式(1)说明  $\zeta_n$  以及  $\zeta_n/n$  的极限分布是正态分布, 即当测试次数较多时, MRR 呈正态分布。式(1)最后一步的理论依据见文献[11]。文献[11]中使用独立同分布的中心极限定理时, 认为  $n$  大于 50 时即可满足大样本的要求, 这里取  $n = 55$ 。

### 3.3 生成 MRR 的样本

为考察识别系统的性能, 首先生成 MRR (或 MDR) 的样本。于是涉及到一个问题, 到底需要生成多少个 MRR 的样本? 根据对 MRR 所作检验的不同, 以及检验要求的不同, 就需要不同的 MRR 的样本容量。假设根据要求需要的 MRR 样本容量为  $m$ , 按照下面的方法生成 MRR 的样本

$$y_1 = \frac{1}{55} \sum_{k=1}^{55} x_k, y_2 = \frac{1}{55} \sum_{k=56}^{110} x_k, \dots,$$

$$y_m = \frac{1}{55} \sum_{k=55(m-1)+1}^{55m} x_k \quad (2)$$

式(2)中,  $x_i$  含义同前,  $x_i$  为 0 或 1,  $i = 1, 2, \dots, 55m$ 。

### 3.4 建立评估指标

1) 识别训练目标的若干评估指标 对训练目标而言, 正确识别率测试结果的均值、错误识别率测试结果的均值、不能识别率测试结果的均值即拒判率测试结果的均值是首先需要考虑的, 由此得出 3 个评估指标: 识别均值、错误均值和识别均值。正确识别率测试结果的方差也是很重要的, 由此得

出评估指标——识别方差。

2) 识别非训练目标的若干评估指标 对非训练目标而言, 正确鉴别率测试结果的均值、错误鉴别率测试结果的均值是首先需要关注的, 由此得出 2 个评估指标: 鉴别均值、错鉴均值。正确鉴别率测试结果的方差也是很重要的, 由此得出评估指标——鉴别方差。

3) 关于代价的指标 识别系统进行处理花费的代价必须考虑, 实际上, 代价包含的内容很多, 设备费用、设备体积、系统复杂度、处理时间、训练目标的模板存储代价都很重要, 这里仅从处理时间、存储代价来建模。所有的识别和鉴别处理时间的均值称为处理时间, 训练目标的模板存储代价称为模板代价, 非训练目标不存在模板存储代价。

## 4 采用模糊综合评判知识构筑目标识别效果评估模型

在前面工作基础上, 构筑目标识别效果评估模糊综合评判模型。

### 4.1 确定因素集

这里的因素集选为 {识别均值、识别方差、鉴别均值、鉴别方差、处理时间、模板代价}, 记之为  $U = \{u_i, i = 1, 2, \dots, 6\}$ 。

### 4.2 确定因素集元素的重要性

给定  $U$  上的向量  $A = (a_1, a_2, a_3, a_4, a_5, a_6)$  为  $\{u_1, u_2, \dots, u_6\}$  分配权重, 权重集可以不遵守归一化要求, 在得出评估结果后再归一化。

### 4.3 选择评语集

评语集选为 {优, 良, 中, 及, 差}, 记之为  $B = \{b_1, b_2, b_3, b_4, b_5\}$ 。

### 4.4 为每个因素确定评语变化范围

为了从测试结果中得到单因素评判矩阵, 研究中使用的做法是给每个因素确定评语变化范围, 而后采用“投票法”的工作思路。按照表 2 为每个因素选择对应的评语变化范围。

简要解释使用表 2 的工作机制。假设一个识别系统对一种训练目标在某条件范围内进行 3 组测试, 正确识别率测试结果的均值 1 次落在区间  $[P_2^y, P_1^y]$  上, 2 次落在区间  $\geq P_1^y$  上, 则该系统在这一条件范围内对这一目标在“识别均值”上对“良”的隶属度是 1/3, 对“优”的隶属度是 2/3, 同时还要记录处理时间和模板代价并进行评判。

### 4.5 模糊综合评判

表 2 每个因素的评语变化范围

Table 2 Variation interval of comments for each factor

评语	优	良	中	及	差
识别均值	$\geq P_1^r$	$[P_2^r, P_1^r)$	$[P_3^r, P_2^r)$	$[P_4^r, P_3^r)$	$\leq P_4^r$
识别方差	$\leq \sigma_4^r$	$[\sigma_4^r, \sigma_3^r)$	$[\sigma_3^r, \sigma_2^r)$	$[\sigma_2^r, \sigma_1^r)$	$\geq \sigma_1^r$
鉴别均值	$\geq P_1^d$	$[P_2^d, P_1^d)$	$[P_3^d, P_2^d)$	$[P_4^d, P_3^d)$	$\leq P_4^d$
鉴别方差	$\leq \sigma_4^d$	$[\sigma_4^d, \sigma_3^d)$	$[\sigma_3^d, \sigma_2^d)$	$[\sigma_2^d, \sigma_1^d)$	$\geq \sigma_1^d$
处理时间	$\geq T_4^d$	$[T_4^d, T_3^d)$	$[T_3^d, T_2^d)$	$[T_2^d, T_1^d)$	$\leq T_1^d$
模板代价	$\leq F_4^r$	$[F_4^r, F_3^r)$	$[F_3^r, F_2^r)$	$[F_2^r, F_1^r)$	$\geq F_1^r$

依据上面的工作机制，得到模糊综合评判的单因素评判矩阵为

$$R = [r_{ik}]_{6 \times 5} \quad (3)$$

模糊综合评判的结果为  $B = A \cdot R = (b_k)$ ,  $k = 1, 2, \dots, 5$ , 其中,  $b_k = \bigvee_{i=1}^6 (a_i \wedge r_{ik})$ ,  $k = 1, 2, \dots, 5$ 。

## 5 仿真结果

### 5.1 选择被评估的识别方法

被评估的识别方法有 4 种，基于一维平均距离像矩特征的目标识别方法和基于一维散射中心匹配的目标识别方法<sup>[12]</sup>，Prony 方法<sup>[13]</sup>和矩阵束方

法<sup>[14]</sup>。这 4 种识别方法分别简称方法 A1, A2, A3, A4。

### 5.2 选择评估中使用的目标及测试条件

评估实验中有 5 种目标实测数据，它们是不同的真实目标的缩比模型，分别称为目标 F, H, J, Y 和目标 W。缩比模型尺寸 1.5~2 m，测量系统是毫米波步进雷达，工作频率 34.7~35.7 GHz，带宽 1 GHz，标称的距离分辨率为 0.15 m，步进频率间隔为 2 MHz，极化状态为垂直发射-水平接收。目标俯仰角 0°，方位角范围 0~30°。方位角每变化 1°进行一次测量。

目标 F, H, J 用作训练目标，目标 Y, W 用作非训练目标。

### 5.3 评估模型各参数设置

1) 每个评语对应的因素变化区间 根据实际情况，认为在该测试条件下，每个评语对应每个因素变化区间的如表 3 所示。

表 3 中，识别时间或鉴别时间仅仅包括识别算法对目标数据处理及得出识别结果的时间，因而比实际的识别系统识别时间短很多。另外，识别代价是所有训练目标在这一条件下的平均代价，从识别代价看出，这 4 种识别方法目标模板的存储量都很小。

表 3 每个因素的评语变化范围

Table 3 Variation interval of comments for each factor

评语	优	良	中	及	差
识别均值	$\geq 0.875$	$[0.845, 0.875]$	$[0.805, 0.845]$	$[0.755, 0.805]$	$\leq 0.685$
识别方差 $\times 10^{-3}$	$\leq 0.65$	$[0.65, 0.75]$	$[0.75, 0.85]$	$[0.85, 0.95]$	$\geq 0.95$
鉴别均值	$\geq 0.675$	$[0.645, 0.675]$	$[0.615, 0.645]$	$[0.595, 0.615]$	$\leq 0.545$
鉴别方差 $\times 10^{-3}$	$\leq 1.25$	$[1.25, 1.45]$	$[1.45, 1.65]$	$[1.65, 1.85]$	$\geq 1.85$
处理时间/ms	$\leq 23.5$	$[23.5, 25.5]$	$[25.5, 28.5]$	$[28.5, 30.5]$	$\geq 33.5$
模板代价/kB	$\leq 1.05$	$[1.05, 1.55]$	$[1.55, 2.05]$	$[2.05, 2.55]$	$\geq 3.05$

2) 权重向量 设置权重向量出发点是对该测试情况下不同因素关注的程度，模糊综合评判的权重向量组为 (0.78, 0.88, 0.68, 0.71, 0.69, 0.48)。省略中间计算过程，在选用的数据下，对 4 种被评估的识别方法归一化之后的评估结果如表 4 所述。

评估结论：在该测试的情况下，方法 A1, A2 的性能强于方法 A4，方法 A3 性能较其余三种方法逊色。

表 4 对四种被评估的识别方法的评估结果

Table 4 Evaluation result for 4 ATR algorithm being evaluated

评语	优	良	中	及	差
方法 A1	0.53	0.32	0.13	0.02	0
方法 A2	0.50	0.31	0.17	0.02	0
方法 A3	0.31	0.36	0.31	0.02	0
方法 A4	0.37	0.34	0.22	0.06	0.01

## 6 结语

笔者仅给出了一个模糊综合评判的目标识别效果评估模型,但是,该评估模型是开放性的模型,可以随评估者关注内容的变化而配置模型,因素集、评语集包含更多或者更少元素时可参照该模型建模。

目标识别效果评估是一项难度较大的系统工程,该方法虽然可以通用于对所有识别系统的评估并结合识别条件给出量化的结果,但与整个研究相比,还有很多工作有待完成。这一领域发展方向是建立更完备的、科学的评估指标,使用有效的评估模型给出合理的评估结论。

致谢:在研究过程中受到华中理工大学于寅教授;国防科技大学李永乐副教授的热情指导,使得论文得以顺利完成,在此谨致谢意。

### 参考文献

- [1] Nasr H N. Automatic object recognition [A]. SPIE Proceedings 1990 SPIE Automatic Object Recognition Conference, Vol IS7 [C]. Cocoa Beach, Florida, 1990. 2~3
- [2] 郁文贤. 智能化识别方法及其在舰船雷达目标识别系统中的应用 [D]. 长沙: 国防科技大学 ATR 国家重点实验室, 1992
- [3] Arnold C W. Improve ATR performance evaluation via mode seeking [A]. SPIE Proceedings Signal Processing, Sensor Fusion, and Target Recognition IV Conference, Vol 2484 [C]. Orlando, FL, Bellingham, WA, 1995. 574~582
- [4] Mohd M A. Performance characterization and sensitivity analysis of ATR algorithms to scene distortions [A]. SPIE Proceedings Architecture, Hardware, and Forward-looking Infrared Issues in Automatic Target Recognition, Vol 1957 [C]. Orlando, FL, Bellingham, WA, 1993. 203~214
- [5] Nasr H N. Automated instrumentation evaluation and diagnostics of automatic target recognition systems [A]. SPIE Proceedings 1990 SPIE Automatic Object Recognition Conference, Vol IS7 [C]. Cocoa Beach, Florida, 1990. 202~213
- [6] Bassham C B. Automatic target recognition classification system evaluation methodology [D]. US: Air Force Institute of Technology, 2002
- [7] Mossing J C, Ross T D, Bradley J. Evaluation of SAR ATR Algorithm Performance Sensitivity to MSTAR Extended Operating Conditions [R]. Available From NTIS No ADA357055/XAB, Arp 1998
- [8] Dudgeon D E. ATR Performance Modeling and Estimation [R]. Available From NTIS No ADA357723/XAB, 1998
- [9] 贺仲雄. 模糊数学及其应用 [M]. 天津: 天津科学技术出版社, 1982. 188~218
- [10] 李洪兴, 汪群, 段钦治, 等. 工程模糊数学方法及其应用 [M]. 天津: 天津科学技术出版社, 1993. 132~136
- [11] 盛骤, 谢式千, 潘承毅. 概率论与数理统计 [M]. 北京: 高等教育出版社, 1989. 72~179
- [12] 付耀文. 基于信息融合的雷达目标识别研究 [D]. 长沙: 国防科技大学 ATR 国家重点实验室, 2003
- [13] 王保义, 时振栋. 电磁场在目标识别中的应用 [M]. 北京: 电子工业出版社, 1995. 72~77
- [14] 姜卫东. 光学区雷达目标结构成像的理论及其在雷达目标识别中的应用 [D]. 长沙: 国防科技大学 ATR 国家重点实验室, 2000

## Performance Evaluation in Automatic Target Recognition Based on Fuzzy Comprehensive Evaluation and Statistics

Li Yanpeng, Li Xiang, Wang Hongqiang, Zhuang Zhaowen

(Institute of Space Electronics in Electronic Science and Engineering School of National University of Defense Technology, Changsha 410073, China)

**[Abstract]** Aiming at performance evaluation of Automatic Target Recognition (ATR), the evaluation indexes and their measurement method are chosen with the application of statistics. Then, with the application of fuzzy comprehensive evaluation, an opening performance evaluation model is formed. This model can offer quantitative object conclusion. In the end, simulation is carried out by test data.

**[Key words]** statistics; fuzzy comprehensive evaluation; automatic target recognition; performance evaluation