

强迫选择实验中的心理测量函数研究

罗涛, 牟轩沁, 闫浩

(西安交通大学图像处理与模式识别研究所, 710049, 西安)

摘要: 为了评价观察者在视觉检测实验中的感知能力, 提出一种含约束条件的最大似然拟合方法. 该方法通过强迫选择实验获得不同视觉刺激强度下的主观识别率, 通过离差分析筛选出有效的实验数据, 然后采用最大似然准则拟合出识别率与刺激强度的函数关系. 实验建立了 2 选项和 4 选项强迫选择实验的心理测量函数, 其中刺激信号的面积覆盖 $100 \sim 400 \text{ mm}^2$. 实验结果表明: 拟合的均方误差在 0.05 附近, 决定系数近似为 0.9; 两位观察者的心理测量函数的最大误差是 0.046 6; 相比非线性最小二乘拟合, 最大似然拟合具有更高的拟合准确率.

关键词: 强迫选择实验; 心理测量函数; 最大似然拟合

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 0253-987X(2010)06-0104-05

Study on Psychometric Function in Forced-Choice Experiments

LUO Tao, MOU Xuanqin, YAN Hao

(Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A maximum likelihood estimation method is presented to evaluate the observers' performance in visual detection tasks. The percent of correct detection is obtained as a function of visual stimulus through forced-choice experiments. Effective experimental data is selected via deviance analysis. The maximum likelihood estimation with some constraints is employed to fit the psychometric function. The psychometric functions with two and four choices experiments are established respectively, and the stimulus area covers the range from 100 to 400 square millimeters. The mean square error of the fitting is about 0.05 and the R-square is around 0.9. The maximum error of the psychometric functions between two observers is 0.046 6. The results show that maximum likelihood estimation achieves a higher accuracy than the least squares nonlinear regression does.

Keywords: forced-choice experiment; psychometric function; maximum likelihood estimation

心理物理学研究的是物理刺激和人的主观感受之间的关系, 定量描述这种关系的函数被称为心理测量函数. 在视觉检测实验中, 心理测量函数揭示了人眼对特定视觉刺激的感知能力. 心理测量函数在视觉感知、视力异常诊断、信号检测和图像质量评价等方面具有很好的应用价值. 心理测量函数主要包括两方面的内容, 心理阈值和斜率. 很多心理物理学实验只考察差别阈值(difference limen, DL)或恰可分辨差异(just noticeable difference, JND), 而不关

心完整的心理测量函数及其斜率. 然而, 基于以下 3 点原因, 越来越多的实验重视考察完整的心理测量函数^[1-2]: ①不同的实验方法和标准获得的阈值存在差异且无可比性, 考察斜率信息就可以使不同方法获得的阈值具有可比性; ②斜率可以度量感知性能的可靠性; ③斜率是评价视觉异常的有价值的诊断信息^[3].

很多医学 X 射线图像中的病灶可以被认为是低对比度、小区域面积的圆盘信号, 为了考察这类信

号的视觉检测能力,本文通过强迫选择实验^[4]和带参数约束的最大似然估计准则^[5]研究了此类检测任务下的心理测量函数.同时,提出用离差指标来评价实验数据的不一致性,并推导出心理物理实验中样本容量与数据方差的关系.

1 理论基础

1.1 强迫选择实验

强迫选择实验具有可重复性、操作时间短的优点,被广泛应用于主观图像质量评价的研究中.在 M 选项的强迫选择实验中,每次同时呈现给观察者 M 幅图像,信号只出现在其中 1 幅图像内.受试观察者事先被告知信号的形状、大小、幅度、可能出现的位置等先验信息.对于同时呈现的 M 幅图像,受试者被要求判断信号所在的图像位置.相同强度的刺激信号出现若干次,定义观察者判断正确的次数与重复次数的比例为识别率 P_c ,检测能力为 D ,它们的关系为^[6]

$$P_c = \int_{-\infty}^{\infty} \Phi(t)^{M-1} \phi(t-D) dt \tag{1}$$

式中: $\Phi(t)$ 为标准正态累计分布函数; $\phi(t)$ 为标准正态概率密度.检测能力描述了图像输入信噪比经过人眼检测系统后滤波输出的结果.

1.2 心理测量函数

心理测量函数用来研究客观刺激量与主观受试者的感受之间的关系.本文将视觉刺激强度 x 定义为 Rose 模型^[7-8]计算出的信噪比 R_{SN} .假设图像中刺激信号的对比度为 C 、面积为 A ,背景亮度为 B ,背景中的加性高斯白噪声的标准差为 σ ,则 Rose 信噪比可以表示为^[9]

$$R_{SN} = CA^{1/2} \frac{B}{\sigma} \tag{2}$$

主观感受量 ψ 在本文中被定义为强迫选择实验中的正确识别率 P_c . ψ 的函数形式为^[10]

$$\psi(x;a,b,r,u) = r + (1-r-u)F(x;a,b) \tag{3}$$

式中: r 为猜测概率,是刺激量为 0 时的正确识别概率; u 为失误概率,描述刺激量超过一定值后,识别率不能完全达到 100% 的程度; F 被称作潜在在心理物理函数,表征 ψ 的阈值和斜率,与刺激量有关,本文中 F 采用 Weibull 函数表示

$$F(x;a,b) = 1 - \exp\left[-\left(\frac{x}{a}\right)^b\right] \tag{4}$$

1.3 方差与离差

强迫选择实验中,假设观察者的正确识别次数服从贝努利分布,则 P_c 的方差 V 与刺激信号的出现次数 n 满足如下关系

$$V = \frac{P_c(1-P_c)}{n} \tag{5}$$

设定方差上限,则可由上式计算出每次实验中刺激信号出现次数的最小值,即每组实验的样本容量.

为了降低受试观察者的实验误差,相同条件的强迫选择实验被重复 30 次.由于观察者的决策标准、反映倾向和情绪等因素在各次实验中存在波动,因此各重复实验的结果也存在较大误差.这些引起数据不一致性的主观因素被称作观察者的内部噪声.假设实验的重复组数是 J ,各组中刺激强度的种类是 K ,各种刺激强度出现的次数是 n_{jk} .定义第 j 组实验的离差为

$$T_j = 2 \sum_{k=1}^K \left[n_{jk} p_{jk} \log\left(\frac{p_{jk}}{\bar{p}_k}\right) + n_{jk} (1-p_{jk}) \log\left(\frac{1-p_{jk}}{1-\bar{p}_k}\right) \right] \tag{6}$$

式中: p_{jk} 为第 k 种刺激强度在第 j 组实验内的识别率; $\bar{p}_k$ 为 J 组间的平均识别率.

假设正确识别次数满足贝努利分布,则当 n_{jk} 趋于无穷大时, T_j 逼近于自由度为 K 的 χ^2 分布.给定显著因子 α ,若某组实验的 T_j 值大于上侧分位数 $\chi^2_{\alpha}(K)$,则表明该组实验误差较大.

1.4 最大似然拟合

在图像信噪比为 x_k 的 $n_k = \sum_j n_{jk}$ 次强迫选择实验中,我们计算出受试观察者在 J 组间的平均识别率是 $\bar{p}_k$.通过如下的最大似然拟合,由离散的 $(x_k, \bar{p}_k)$ 建立识别率与信噪比的函数关系式.假设在一组强迫选择实验中,正确识别次数满足贝努利分布,则 n_k 次实验中观察者识别率为 $\bar{p}_k$ 的概率 $f_{\bar{p}_k}$ 可以表示为

$$f_{\bar{p}_k} = C_{n_k}^{n_k \bar{p}_k} [\psi(x_k;a,b,r,u)]^{n_k \bar{p}_k} \cdot [1 - \psi(x_k;a,b,r,u)]^{(1-\bar{p}_k)n_k} \tag{7}$$

式中: $C_{n_k}^{n_k \bar{p}_k}$ 表示组合公式; a, b, r 和 u 是 4 个待拟合的参数.对于含有 K 种信噪比的实验,其似然函数 L 可以表示为

$$L(p;a,b,r,u) = \prod_k f_{p_k} = \prod_k C_{n_k}^{n_k \bar{p}_k} \psi^{n_k \bar{p}_k} (1-\psi)^{(1-\bar{p}_k)n_k} \tag{8}$$

由于对数运算不改变函数的单调性,所以似然函数变形为

$$l(p;a,b,r,u) = \log[L(p;a,b,r,u)] =$$

$$\sum_{k=1}^K \log C_{n_k}^{n_k \bar{p}_k} + \sum_{i=1}^K n_k \bar{p}_k \log \psi + \sum_{k=1}^K (1 - \bar{p}_k) n_k \log [1 - \psi]$$

(9)

为获得参数的估计量,可以采用求导或迭代方法求解似然函数的极大值. 本文采用 Nelder-Mead 方法求解极值,并且加入如下约束条件:令猜测概率 r 在 $1/M$ 附近变化,失误概率 u 在 0 附近变化且恒为正.

1.5 评价指标

为了评价最大似然拟合的优劣程度,定义决定系数为

$$R = \frac{E_{SSR}}{E_{SST}} = 1 - \frac{E_{SSE}}{E_{SST}}$$

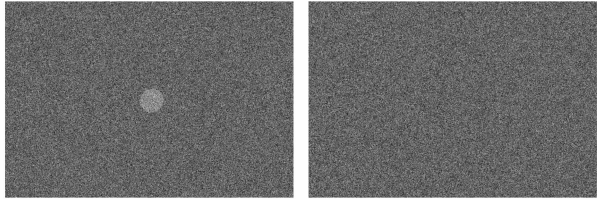
(10)

式中: $E_{SSE} = \sum_{j=1}^J \sum_{k=1}^K (p_{jk} - \hat{p}_{jk})^2$ 为剩余离差平方和
其中 $\hat{p}_{jk}$ 是由拟合的心理测量函数求得的识别率的估计值;
 $E_{SSR} = \sum_{j=1}^J \sum_{k=1}^K (\hat{p}_{jk} - \bar{p}_k)^2$ 为回归离差平方和;
 $E_{SST} = \sum_{j=1}^J \sum_{k=1}^K (p_{jk} - \bar{p}_k)^2$ 为总离差平方和. 决定系数 R 的值越接近于 1 ,则拟合所获得的心理测量函数与原始强迫选择实验数据就越相似,即拟合度越高.

2 实验方法

强迫选择实验中,刺激图像由计算机生成并随机出现. 观察者首先通过测试训练熟悉实验流程,了解各种强度的刺激和噪声在显示器上的显示效果. 待检测的刺激信号是圆盘形状,具有不同的对比度和面积. 图像背景中含有高斯白噪声,当光子数很大时,近似为泊松分布. 实验前,观察者被告知刺激信号的面积. 实验时,显示器上出现 M 块区域,不同信噪比水平的信号随机出现在某个位置. 观察者需要对信号是否出现以及位置做出判断,这一过程没有时间限制. 观察者距离显示器大约 50 cm ,实验环境选择暗室或亮度适中的房间. 实验采用的显示器亮度动态范围是 $0.5 \sim 80\text{ cd/m}^2$,分辨率是 $1\,024 \times 768$ 像素,像素间距是 0.3 mm . 图 1 描述的是当 M 选取 2 时呈现给观察者的 2 幅图像,信号随机出现在其中一幅图像的中心位置,观察者需要判断信号位于左图或右图. 当前图像中,信号位于图 1a 中,图 1b 是单纯背景的图像. 共有 2 位受试观察者参加了强迫选择实验,每位观察者在同一实验条件下的

重复次数为 30 次. 30 组重复实验分别在 3 个时间段完成,分别间隔 1 星期.



(a)含信号图像 (b)单纯背景图像

图 1 强迫选择实验中的刺激图像

3 实验结果

3.1 重复实验中的误差分布

本文的强迫选择实验包括 2 选项和 4 选项,其信号面积覆盖了 $100 \sim 400\text{ mm}^2$. 对相同面积的刺激信号,分别重复若干组实验,并由式(6)计算出各组实验对应的离差. 在拟合前,剔除较大离差值对应的实验数据. 图 2 描述了 2 选项强迫选择实验中信号面积为 100 mm^2 的 30 组实验的离差分布. 平行于横轴的直线表示 α 等于 0.975 对应的上侧分位数 $\chi^2_{\alpha}(K)$,此横线上方的点表示该组实验数据的离差较大. 因此,在建立心理物理测量函数时,应舍去这些误差较大的实验数据.

3.2 心理测量函数的拟合结果

将强迫选择实验的数据经过离差分析和最大似然拟合,建立了以 Rose 信噪比为自变量、识别率为因变量的心理测量函数. 图 3 描述的是 2 选项实验中信号面积为 100 mm^2 的数据的拟合结果. 图 3 中的 10 个离散数据点中,标准差的最大值是 $0.071\,6$,最小值是 $0.021\,9$,均值是 $0.050\,5$. 连续的曲线表示经过拟合获得的心理测量函数,拟合的均方误差是 $0.052\,1$,决定系数是 $0.875\,0$. 图 3 中的心理测量函数曲线在信噪比为 0 和 2.5 附近,分别渐进趋近于

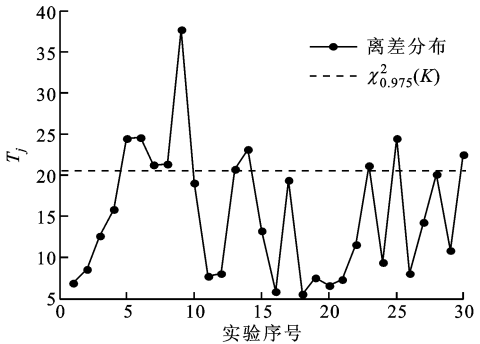


图 2 重复实验的离差分布

识别率等于 0.45 和 1. 这个结果符合强迫选择实验的约束,即当信噪比为 0 时识别率趋近于猜测概率(该图中,选项 M 为 2,因此猜测概率为 0.5),当信噪比超过一定值时识别率趋近于最大值 1.

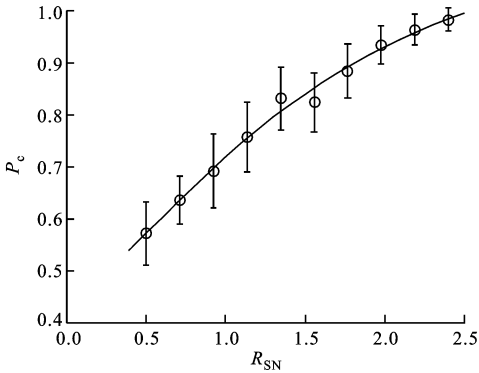


图 3 心理测量函数的拟合

3.3 方差分析

对特定信噪比,以拟合识别率为期望值,计算出 30 组强迫选择实验对应的识别率的方差.将该方差与由式(5)计算的方差理论值做比较,结果如图 4 所示.由图 4 可以看出,两种方差均随识别率的升高而减小,但是当识别率低于 0.7 时,实验方差小于理论值.这是因为在较低的信噪比下,受试观察者的决策标准在每组重复实验中会发生变化.例如, N 次实验中,观察者 A 在 Nq 次实验中的识别率是 p_1 ,在其余 $N(1-q)$ 次实验中的识别率是 p_2 ,则其识别率的期望值是 $p_A = qp_1 + (1-q)p_2$,方差为 $V_A = [qp_1(1-p_1) + (1-q)p_2(1-p_2)]/N$.另一观察者 B 在 N 次实验中采取相同决策标准,识别率是 p_B ,若两位观察者识别率的期望值相同,即 $p_B = qp_1 + (1-q)p_2$,则观察者 B 的识别率的方差为 $V_B = p_B(1-p_B)/N$.当 $p_1 \neq p_2$ 和 $q < 1$ 时, $V_B > V_A$.信噪比低时,观察者的决策标准在一组实验中会有波动,从而造成识别率的方差减小.

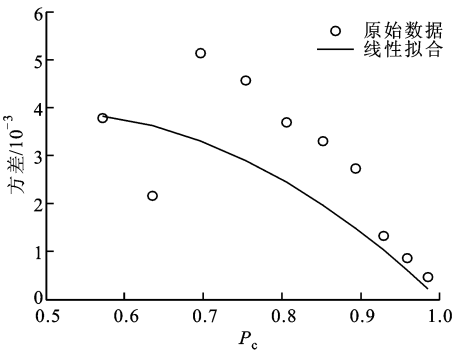


图 4 方差与识别率的关系

3.4 Rose 信噪比

将图 3 的识别率按照式(1)计算出相应的检测能力 D ,Rose 信噪比与检测能力的关系如图 5 所示.对比图 3 与图 5,图形由曲线变为直线.这表明在检测高斯噪声背景中的圆盘信号时,Rose 信噪比值可以很好地刻画检测能力.

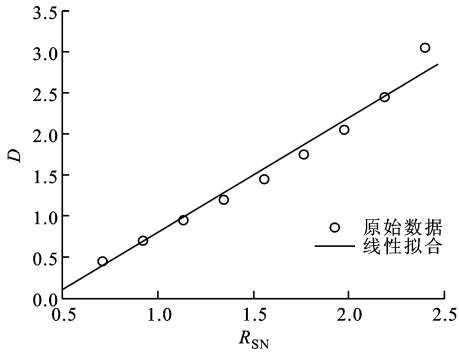


图 5 检测能力与信噪比的关系

3.5 观察者之间的比较

图 6 比较了两位观察者的检测结果,曲线分别是由各自实验数据拟合得到的识别率与信噪比的关系.两者的识别率的最大误差是 0.046 6,均方根误差是 0.028 1.结果表明,两位观察者识别率的差异随信噪比的增大而增大.

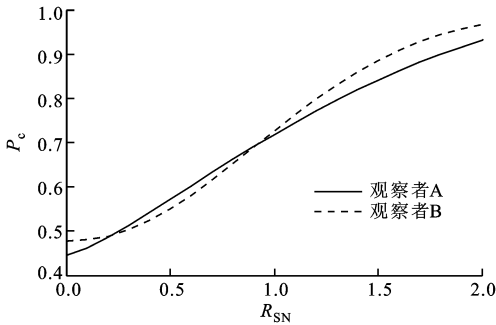


图 6 观察者检测结果的比较

4 结 论

本文通过强迫选择实验获得了视觉刺激的信噪比与识别率的关系.采用最大似然估计,建立了 2 选项和 4 选项强迫选择实验的心理测量函数,其刺激信号的面积覆盖了 $100 \sim 400 \text{ mm}^2$.拟合得到的心理测量函数与原始实验数据之间具有较小的偏差,能反映出实验数据的趋势.结果表明,最大似然准则适合于强迫选择实验中的心理测量函数的拟合.本文还推导出样本容量与数据方差的函数关系,以及识别率对方差的影响.对数据做离差分析,可以将误差

较大的实验组剔除. 在检测高斯噪声背景中的圆盘信号时, Rose 信噪比等价于检测能力. 本文的实验结果可以为医学图像中的病灶检测提供心理物理学依据.

参考文献:

- [1] KLEIN S K. Measuring, estimating, and understanding the psychometric function: a commentary [J]. *Perception & Psychophysics*, 2001, 63(8): 1421-1455.
- [2] STRASBURGER H. Converting between measures of slope of the psychometric function [J]. *Perception & Psychophysics*, 2001, 63(8): 1348-1355.
- [3] CHAUHAN B C, TOMPKINS J D, LEBLANC R P, et al. Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma [J]. *Investigative Ophthalmology & Visual Science*, 1993, 34(13): 3534-3540.
- [4] BROWN D G, BURGESS A E, METZ C E, et al. Medical imaging: the assessment of image quality, ICRU report 54 [R]. Bethesda, MD, USA: International Commission of Radiation Units and Measurements, 1996.
- [5] MIGUEL A. Sampling plans for fitting the psychometric function [J]. *The Spanish Journal of Psychology*, 2005, 8(2): 256-289.
- [6] BOCHUD F O. Search for lesions in mammograms: statistical characterization of observer responses [J]. *Medical Physics*, 2004, 31(1): 24-36.
- [7] BURGESS A E. The Rose model, revisited [J]. *J Opt Soc Am; A*, 1999, 16(3): 633-646.
- [8] WATTS R, WANG Yi, WINCHESTER P A, et al. Rose model in MRI: noise limitation on spatial resolution and implications for contrast enhanced MR angiography [C]//The 8th Scientific Meeting & Exhibition. Berkeley, CA, USA: ISMRM, 2000:277-288.
- [9] LUO Tao, MOU Xuanqin. An applicability research on JND model [C]//Medical Imaging: Image Perception, Observer Performance, and Technology Assessment. San Diego, CA, USA: SPIE, 2006:101-108.
- [10] MIGUEL A. Sampling plans for fitting the psychometric function [J]. *The Spanish Journal of Psychology*, 2005, 8(2):256-289.

(编辑 刘杨)