

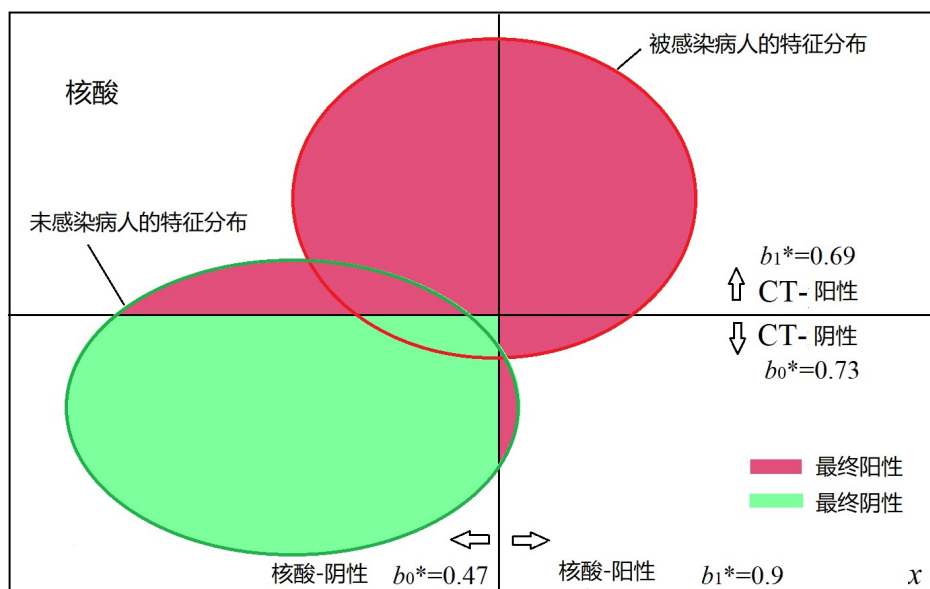
信道确证和预测确证——从医学检验到乌鸦悖论

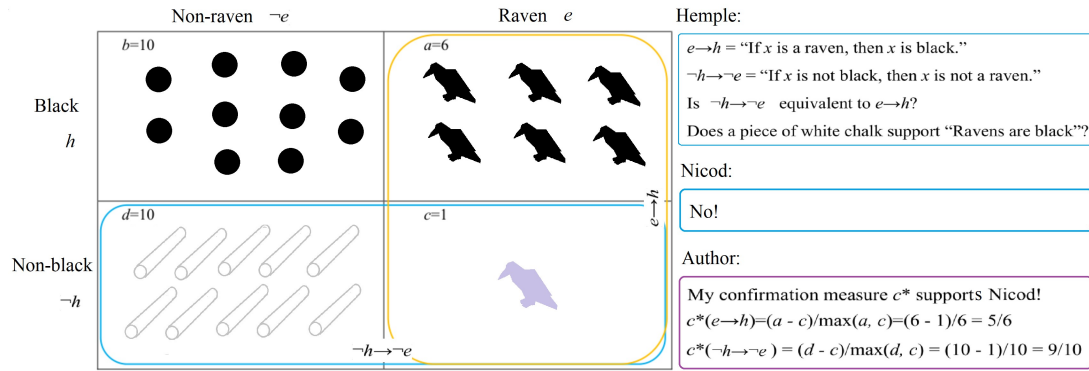
鲁晨光

lcguang@foxmail.com

摘要: 在证伪和证实之间的长期争论之后, 全称假设的证实被不确定大前提的确证所取代。不幸的是, Hempel 提出乌鸦悖论(又叫确证悖论)——在等价条件和尼科德准则之间存在矛盾。然后 Carnap 提出用逻辑概率的增量作为确证测度。为了确证大前提并消除乌鸦悖论, 研究者们提出多种确证测度。其中由 Kemeny 和 Oppenheim 提出的 F 测度具有 Elles 和 Fitelson 提出的对称性和不对称性、Crupi 等人提出归一性(确证度在-1 和 1 之间变化)、Greco 等人提出单调性。基于语义信息方法并以医学检验为例, 作者推导出一个和 F 类似的确证测度 b^* 。 b^* 和 F 同似然比类似, 能体现信道或检验手段有多好, 但是不能体现概率预测(根据阳性或阴性预测有病或没病)有多好。并且, 用 b^* 、 F 或其他测度还是不能清楚解释如何消除乌鸦悖论。为此, 作者推导出类似于正确率的确证测度 c^* 。 c^* 有形式简单: $(a-c)/\max(a,c)$, 它明确支持尼科德准则并反对等价条件, 因此用它可以消除乌鸦悖论。一个例子表明 F 和 b^* 有助于同时用核酸试剂和 CT 诊断新冠病毒, 而其它流行的确证测度不行。另一个例子揭示: 所有流行的确证测度都不能用来解释一个黑乌鸦比一支白粉笔更好支持“乌鸦是黑的”。 F 、 b^* 和 c^* 都表明, 较少反例比较多正例更重要, 它们因此兼容 Popper 的证伪思想。

图片式摘要:





关键词: 相对熵, 确证, 语义信息, 医学检验, 乌鸦悖论, 归纳推理, Popper思想

1 序言

一个全称假设或推理规则 $e \rightarrow h$, 比如“所有乌鸦是黑的”, 等价于一个假言判断, 比如“对于每个 x , 如果 x 是乌鸦, 则 x 是黑的.”两者都可以作为三段论的大前提。演绎逻辑需要大前提, 但是涉及经验推理的大前提不能从演绎得到, 只能从归纳得到支持。逻辑经验主义者起初肯定全称假设(或与之等价的大前提)可以通过经验证实. Popper 反驳说, 全称假设不能被证实, 只能被证伪。经过长时间争论, Popper 和逻辑经验主义代表得出一致结论: 不确定全称假设可以通过证据得到确证[1,2]。

1945 年, Hempel [3]提出确证悖论——即乌鸦悖论。按照经典逻辑的等价条件(EC), “如果 x 是乌鸦则 x 是黑的”(推理 1)和 “如果 x 不是黑的则 x 就不是乌鸦”(推理 2)等价(结论 1)。而根据 Nicod 准则[4]或 Nicod-Fisher 准则(NFC) (2.3 节详解): 黑乌鸦支持推理 1; 不黑的乌鸦否定推理 1; 非乌鸦物体比如黑猫和白粉笔和推理 1 不相关(结论 2)。结论 1 和结论 2 之间或 EC 和 NFC 之间存在悖论。

为了量化确证, 逻辑经验主义代表Carnap提出贝叶斯确证方法和确证测度[1]。Popper 也提出自己的确证测度[2], 但是Carnap的确证测度更加著名。现在, 研究者们已经提出很多确证测度[1, 5-13]。归纳问题现在似乎已转变为确证问题或确证度计算问题. 为了筛选合理的确证测度, Elles 和 Fitelson[14]提出四个对称不对称标准, Crupi 等人[8]和 Greco 等人[15]提出归一化标准(要求确证度在-1 和 1 之间变化), Greco 等人[16]提出单调标准。上面那些测度中, 只有 Kemenv 和 Oppenheim 提出的测度 F [12]和 Crupi 等人提出的测度 Z [8]符合上述标准. 测度 Z 和 Shortliffe 和 Buchanan 提出的确定性因子相关[7]。(中文版补充: 国内学者对确证和乌鸦悖论也有研究[37-40], 但是作者没见过有人提出新的确证测度)。

笔者[17]使用语义信息方法优化(模糊)真值函数时发现, 一个二选一预测的真值函数可以看作清晰句和永真句的不同比例组合, 用样本分布优化的永真句的比例就可以看做否定度 b^* ——使用最大语义信息准则或最大似然准则, 1 减否定度就可以作为确证度 b^* 。 b^* 和 F 类似, 也具有上述多个所希望的数学性质。

好的确证测度应该不仅具有好的数学性质, 还应该具有实用性. 我们可以通过医学检验来检查确证测度的实用性. 我们用可信度表示我们对一个大前提的相信程度, 用样本优化的信任度作为确证度(即可信度). 医学检验用阳性和阴性预测感染和未感染(有病或无病), 阳性和阴性都有信任度和确证度. 医学检验中有个重要问题比如: 在核酸检验(NAT 即 Nucleic Acid Test)和 CT (Computed Tomography)同时被采用的时候, 如果核酸检验呈阴性而 CT 检验

呈阳性, 我们应该相信哪一个? 根据一个检验的敏感性和特异性[18]和感染的先验概率, 我们可以用任一测度算出阴性和阳性的确证度. 使用流行的确证测度, 我们能算出合理的有助于诊断(从核酸阴性和 CT 阳性中选一个更好的)的确证度? 这些确证度能反映有病无病的概率?

本文将显示, 只有和似然比兼容或是似然比函数的确证测度, 比如 F 和 b^* , 能帮助我们诊断感染, 并且诊断结果可以被医学界接受. 然而, 用 b^* 、 F 或 Z 还是不容易清楚地消除乌鸦悖论。

笔者最新研究发现, 医学检验问题和乌鸦悖论性质不同, LR 、 F 和 b^* 反映的是检验手段有多好, 不能反映概率预测有多好。为了澄清乌鸦悖论, 我们需要能反映概率预测有多好的确证测度。为此, 本文推导出确证测度 c^* ——称之为预测确证测度, 并称 b^* 为信道确证测度。信道确证和预测确证的区分类似但却不同于贝叶斯确证和似然确证的区分[19]。测度 c^* 符合 Nicod-Fisher (NF) 准则并否定等价条件, 从而可以用来消除乌鸦悖论。

本文主要目的是:

- 区分信道确证测度(和似然比兼容)和预测确证测度(反映概率预测);
- 推导出一个预测确证测度 c^* , 用以消除乌鸦悖论);
- 解释确证和证伪可以兼容。

本文使用方法和流行方法主要不同在于:

- b^* 和 c^* 都是用语义信息方法和最大似然准则推导出来的而不是定义的[17, 20];
- 确证和统计学习相互支持, 求确证度不只是为了评价大前提或推理规则, 还为了方便做概率预测。

本文的主要贡献是:

- 澄清了: 我们不能用一个确证测度完成两个不同任务: 1) 评估通信信道, 比如作为检验手段的医学检验; 2) 评估概率预测, 比如评估“乌鸦是黑的”和“检验呈阳性则很可能有病”。
- 提出和正确率性质类似的并能反映 Nicod-Fisher 准则的确证测度 c^* , 从而澄清乌鸦悖论。

本文组织如下: 第 2 节是背景, 包括回顾已有确证公式及其筛选, 介绍了相关语义信息方法, 澄清一些问题. 第 3 节用语义信息方法推导出新的确证测度 b^* 和 c^* , 提供了有不同前后件的大前提的确证公式。第 4 节是结果。其中用一个例子说明新的确证度的特点——反例少比正例多更加重要, 用另一个关于新冠肺炎诊断的例子说明 b^* 和 F 比其他确证测度实用, 还用另一个例子说明 a 和 d 的增量对不同确证测度的影响(用于澄清乌鸦悖论)。第 5 节讨论了为何只有测度 c^* 才能消除乌鸦悖论, 也讨论了流行确证方法中的一些概念混淆和误解, 以及新的确证测度如何兼容 Popper 的证伪思想。最后是结论。

2. 背景

2.1 统计概率, 逻辑概率, Shannon 信道, 和语义信道

首先我们区分一个假设或标签的逻辑概率和统计概率. 逻辑概率是它被判定为真的概率, 而统计概率是它被选择的概率。

假定有一万个有不同年龄的人走过来，让门卫来判断来人是否是老年人。如果有 2000 人被判定为“老年人”，那么“老年人”的逻辑概率就是 0.2。再假设有一万个有不同年龄的人走过来；请门卫给每个人贴上四个标签——“小孩”，“年轻人”，“成年人”，“老年人”——中的一个，贴上“老年人”标签可能有 1000 人，则统计概率是 0.1。它小于逻辑概率 0.2。为什么不是 2000 人？因为有的老年人被贴上了“成年人”标签。一个永真句的逻辑概率是 1，而它的统计概率通常是 0。因为我们极少选择永真句。统计概率是归一化的，而逻辑概率通常不是归一化的[17]。由于这些原因，我们要用“ T ”和“ P ”区分逻辑概率和统计概率。

现在我们考虑 Shannon 信道[21]和语义信道——在人的年龄和标签（“小孩”，“大人”，“年轻人”，“中年人”，“老人”…）之间。

令 X 是一随机变量，表示年龄； Y 是一随机变量，表示一个标签。 X 取值 $x \in \{\text{年龄}\}$ ； Y 取值 $y \in \{\text{“小孩”，“大人”，“年轻人”，“中年人”，“老人”}\}$ 。Shannon 称先验概率分布 $P(X)$ (或 $P(x)$) 是信源，称 $P(Y)$ 是信宿。从 X 到 Y 的 Shannon 信道是 $P(Y|X)$ 。它是一个转移概率矩阵：

$$P(Y|X) \Leftrightarrow \begin{bmatrix} P(y_1|x_1) & P(y_1|x_2) & \dots & P(y_1|x_m) \\ P(y_2|x_1) & P(y_2|x_2) & \dots & P(y_2|x_m) \\ \dots & \dots & \dots & \dots \\ P(y_n|x_1) & P(y_n|x_2) & \dots & P(y_n|x_m) \end{bmatrix} \Leftrightarrow \begin{bmatrix} P(y_j|x) \\ P(y_j|x) \\ \dots \\ P(y_n|x) \end{bmatrix}, \quad (1)$$

其中 $\Leftrightarrow$ 表示等价。该矩阵由一组条件概率 $P(y_j|x_i)$ ($j=0,1,\dots,n$; $i=0,1,\dots,m$) 或一组转移概率函数 $P(y_j|x)$ ($j=0,1,\dots,n$) 组成。Shannon 称呼 $P(y_j|x)$ 是转移概率函数[21]，其中 y_j 是常量， x 是变量。

相应地也有语义信道，它由一组真值函数构成。令 $T(\theta_j|x)$ 是 y_j 的真值函数，其中 θ_j 是模型参数，用它可以构成 $T(\theta_j|x)$ 。比如， $y_j = \text{“}x \text{ 是年轻人”}$ 。其真值函数可以是

$$T(\theta_j|x) = \exp[-(x-20)^2/25], \quad (2)$$

其中 20 和 25 就是模型参数。设 $y_k = \text{“}x \text{ 是老年人(elderly)”}$ ，其真值函数可以是一个 Logistic 函数：

$$T(\theta_k|x) = 1/[1 + \exp[-0.2(x-65)]], \quad (3)$$

其中 0.2 和 65 就是模型参数。两个真值函数如图 1 所示。我们也可以把 θ_j 看成 x 的论域上的一个模糊集合，那么真值函数就是隶属函数[17]。

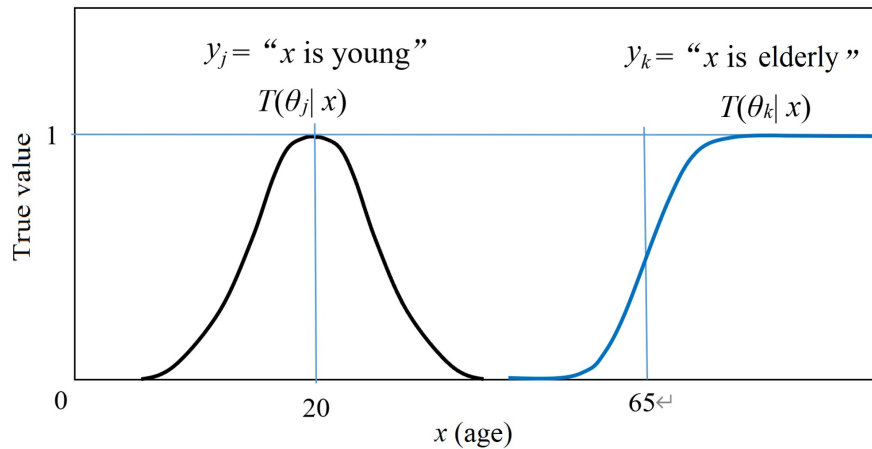


Figure 1. 两个关于年龄的谓词的真值函数

根据 Tarski 的真理论[22]和 Davidson 的真值条件语义学[23]，一个真值函数代表一个假设的语义。因此，我们称下面由真值函数构成的矩阵是语义信道：

$$T(\theta | X) \Leftrightarrow \begin{bmatrix} T(\theta_1 | x_1) & T(\theta_1 | x_2) & \dots & T(\theta_1 | x_m) \\ T(\theta_2 | x_1) & T(\theta_2 | x_2) & \dots & T(\theta_2 | x_m) \\ \dots & \dots & \dots & \dots \\ T(\theta_n | x_1) & T(\theta_n | x_2) & \dots & T(\theta_n | x_m) \end{bmatrix} \Leftrightarrow \begin{bmatrix} T(\theta_1 | x) \\ T(\theta_2 | x) \\ \dots \\ T(\theta_n | x) \end{bmatrix}. \quad (4)$$

使用转移概率函数 $P(y_j|x)$, 我们可以做概率预测:

$$P(x|y_j)=P(x)P(y_j|x)/P(y_j). \quad (5)$$

使用真值函数 $T(\theta_j|x)$, 我们也可以做概率预测或产生似然函数:

$$P(x | \theta_j) = P(x)T(\theta_j | x) / T(\theta_j), \quad (6)$$

其中 $T(\theta_j)$ 是逻辑概率. 我们有

$$T(\theta_j) = \sum_i P(x_i)T(\theta_j | x_i). \quad (7)$$

公式(6)可谓语义贝叶斯公式[17]. 似然函数是主观的,它可以看作是逻辑概率和统计概率的杂交.

当信源 $P(x)$ 改变时, 上面概率预测公式仍然可用. 容易证明, $P(x|\theta_j)=P(x|y_j)$ 当 $T(\theta_j|x) \propto P(y_j|x)$ 时. 因为 $T(\theta_j|x)$ 最大值是 1, 令 $P(x|\theta_j)=P(x|y_j)$, 我们可以得到优化的真值函数 [17]:

$$T^*(\theta_j|x) = [P(x|y_j)/P(x)]/\max[P(x|y_j)/P(x)] = P(y_j|x)/\max[P(y_j|x)], \quad (8)$$

其中 x 是以变量, $\max(.)$ 是括号中的函数的最大值.

2.2 回顾流行的确证测度及筛选

我们用 h_1 表示一个假设, h_0 是它的否定; h 是变量, 表示 h_1 和 h_0 中的一个; e_1 表示和 h_1 相关的假设, e_0 是相反假设; e 同理. 我们且用 c 表示任一确证测度, 用 $c(e, h)$ 表示证据对大前提的支持度——即确证度(注意: 本文采用文献[8]中写法, e 在左边, h 在右边).

已有的比较著名的确证测度如下:

$$D(e_1, h_1)=P(h_1|e_1)-P(h_1) \text{ (Carnap, 1962 [1])},$$

$$M(e_1, h_1) = P(e_1|h_1)-P(e_1) \text{ (Mortimer, 1988 [5])},$$

$$R(e_1, h_1)=\log[P(h_1|e_1)/P(h_1)] \text{ (Horwich, 1982[6])},$$

$$C(e_1, h_1)=P(h_1, e_1)-P(e_1)P(h_1) \text{ (Carnap, 1962 [1])},$$

$$Z(h_1, e_1) = \begin{cases} [P(h_1 | e_1) - P(h_1)] / (1 - P(h_1)), & \text{当 } P(h_1 | e_1) \geq P(h_1), \\ [P(h_1 | e_1) - P(h_1)] / P(h_1), & \text{否则} \end{cases} \quad (\text{Shortliffe and Buchanan, 1975[7]; Crupi et al., 2007[8]}),$$

$$S(e_1, h_1)=P(h_1|e_1)-P(h_1|e_0) \text{ (Christensen, 1999 [9])},$$

$$N(e_1, h_1)=P(e_1|h_1)-P(e_1|h_0) \text{ (Nozik, 1981 [10])},$$

$$L(e_1, h_1)=\log[P(e_1|h_1)/P(e_1|h_0)] \text{ (Good, 1984 [11])}, \text{ and}$$

$$F(e_1, h_1)=[P(e_1|h_1)-P(e_1|h_0)]/[P(e_1|h_1)+P(e_1|h_0)] \text{ (Kemeny and Oppenheim, 1952[12])}.$$

从文献[8,24]可见更多确证测度. 测度 F 被有些作者称为 I^* [13]、 L [8]或 k [24].

首先, 我们要澄清: 确证需要什么样的证据支持上面样的假设. 我们看下面三种假设:

- **假设 1:** $h_1(x)$ =" x 是年老的" (x 是年龄). 这是一个谓词; 如果 $x=70$ 岁(作为证据), 就有命题 $h_1(70)$ 的真值, 比如 1; 如果 $x=50$ 岁, $h_1(50)$ 的(模糊)真值较小, 比如说是 0.3.

- **假设 2:** h_1 = “如果 $x \geq 60$ 岁, 则 x 被称为老年人”。这是一个假言判断, 或大前提。注意: $x=70$ 和 $x \geq 60$ 只是结论的证据, 不是大前提的证据。大前提的证据是含有许多多样例的样本。

- **假设3:** $e_1 \rightarrow h_1$ = “如果 $x \geq 60$ 岁, 则 x 被称为老年人”。和假设 2 相同, 但是符号不同, 其中 e_1 = “ $x \geq 60$ 岁”, h_1 = “ x 是老年人”。其证据是含四种样例的样本, 用统计学习的方法表达是: $\{(e_1, h_1), (e_1, h_0), \dots\}$

假设 1 只有真值函数和逻辑概率——在 0 和 1 之间变化, 不需要确证; 而假设 2 或假设 3 需要确证, 其确证度在 -1 和 1 之间变化。

关于 $c(e, h)$, 存在两种理解:

- 理解 1: h 是要确证的大前提, 如**假设 1**, e 是支持大前提的证据(Elles和Fitelson [14]);
- 理解 2: 要确证的大前提就是 $e \rightarrow h$, 如**假设 2**, 即“若 e 则 h ” (Kemeny 和 Oppenheim[12])。 e 只是支持 h 的证据, 并不是支持大前提的证据。

但是, 尽管存在不同理解, 大多数人都赞成用四个样例(e_1, h_1), (e_0, h_1), (e_1, h_0), (e_0, h_0)的个数 a, b, c, d 作为大前提的证据; 用四种样例的个数 a, b, c, d 计算确证度(参看表 1)。下面讨论都基于这种共识。

表1. 支持大前提的 $e_1 \rightarrow h_1$ 的四种样例的个数¹

	e_0	e_1
h_1	b	a
h_0	d	c

表1中 a 是(e_1, h_1)的个数, 比如 e_1 = “乌鸦”(可看做是标签, 也可以看做是“ x 是乌鸦”的简写, 其他同理), h_1 = “黑的”; 则 a 是黑乌鸦的个数; b, c, d 同理。由 a, b, c, d 可以算出上面不同确证测度中不同概率。

为了明确任务, 本文按理解 2, 把 $e \rightarrow h$ = “若 e 则 h ” 看作是要确证的大前提, 把 $e \rightarrow h$ 的确证度写成 $c(e \rightarrow h)$. 研究确证方法就是构造或选择确证测度 $c(e \rightarrow h) = f(a, b, c, d)$. 确证就是计算 $f(a, b, c, d)$ 。

为筛选出更合理的确证测度, Elles和Fitelson[14]提出用几种对称性:

HS(假设对称性): $c(e_1 \rightarrow h_1) = -c(e_1 \rightarrow h_0)$ (大前提后件相反),

ES(证据对称性): $c(e_1 \rightarrow h_1) = -c(e_0 \rightarrow h_1)$ (前件相反),

MS(互换对称性): $c(e_1 \rightarrow h_1) = c(h_1 \rightarrow e_1)$ (前后件互换),

TS(整体对称性): $c(e_1 \rightarrow h_1) = c(e_0 \rightarrow h_0)$ (前后件都相反).

他们的分析表明, 只有 HS 是需要的(desirable), 其他三种对称性是不需要的。这一结论得到大多数研究者的支持。因为 TS 可以通过 HS 和 ES 产生, 只要检查前面三种对称性就行。按照他们的对称性标准, 只有测度 L, F, Z 测度符合对称性要求。测度 N 是否符合, 不能确定。关于确证测度对称性的更多讨论见文献[14,25,26].

Greco 等人[15](包括粗糙集理论创始人 Pawlak)在他们的文章中还提出单调性标准。如果 $f(a, b, c, d)$ 不随 a 或 d 增加而减小, 也不随 b 或 c 增加而增加, 则我们称 $f(a, b, c, d)$ 具有单

¹ 表1的设计考虑到配合图2.

调性。 L 、 F 和 Z 具有单调性而 D 、 M 、 N 不具有。所以，使用对称不对称标准和单调性标准，只有 L 、 F 和 Z 符合要求。如果再按照确证度的归一化（在-1和1之间）要求[8, 12]，最后只有测度 F 和 Z 符合所有要求。Crupi等人[8]还提出其他一些性质，比如 L 性质（兼容经典逻辑的性质 **Logicity**，即无反例时 $c(e \rightarrow h)=1$ ；无正例时 $c(e \rightarrow h)=-1$ ）。按照 L 性质要求，也可以排除大多数测度并保留 F 和 Z 。

考虑医学检验，比如新冠肺炎(COVID-19)检验。设 e_1 ="阳性"(即" x 是阳性", x 是一个检验样品)， e_0 ="阴性"， h_1 ="被感染"(即" x 被感染")， h_0 ="没被感染"。则阳性似然比是 $LR^+=P(e_1|h_1)/P(e_1|h_0)$ ，它反映阳性判断有多可靠。 L 和 F 同 LR 之间存在一一对应关系：

$$L(e_1 \rightarrow h_1) = \log LR^+; \quad (9)$$

$$F(e_1, h_1) = (LR^+ - 1)/(LR^+ + 1). \quad (10)$$

所以 L 和 F 可以用作医学检验中推理“阳性 $\rightarrow$ 有病”和“阴性 $\rightarrow$ 无病”的确证度。和 LR 或 L 相比， F 能更加直观地反映该检验(F)和最好检验(1)最坏检验(-1)的差距。但是 LR 用作有病无病概率预测比 F 方便[27]。 Z 和似然比 LR^+ 没有一一对应关系。

2.3 区分大前提的证据和结论的证据

三段论中结论的证据是小前提，而确证大前提用到的证据是样本或样本分布 $P(e, h)$ 。在有些人的研究中， e 有时被看做是三段论的小前提，有时又被看作是样例序列或样本； h 有时被看做大前提的结论，有时又被看做大前提。大家之所以不用 $c(e \rightarrow h)$ 而用 $c(e, h)$ 或 $c(h, e)$ ，原因就是为回避两种理解之间的矛盾。如果区分两种证据，使用 $c(e \rightarrow h)$ 毫无问题；我们只需强调大前提 $e \rightarrow h$ 的证据不是 e 而是样本或样本分布 $P(e, h)$ 。

用 h 表示大前提同时用 e 表示大前提的证据(Elles和Fitelson[14,28])。这让人很难理解 $\neg e$ 。因为可能的样例有四种，而不是两种。比如，设 $h=p \rightarrow q$ ， e 是 (p, q) 、 $(p, \neg q)$ 、 $(\neg p, q)$ 和 $(\neg p, \neg q)$ 中的一个。如果认为 $(p, \neg q)$ 作为 $\neg e$ 反对 h ，其他三个样例，包括 $(\neg p, q)$ 和 $(\neg p, \neg q)$ 都作为 e 支持 $p \rightarrow q$ ，这就有问题了。因为按同样的道理， $(\neg p, q)$ 和 $(\neg p, \neg q)$ 也支持 $p \rightarrow \neg q$ 。而按HS或SS(后件对称性)，这是不应该的。另外， e 的更一般情况是包含许多样例的样本，一个样本的否定和概率是什么？这是令人费解的。

可惜的是，虽然很多作者把 e 理解为大前提的证据，但是所用具体计算公式中还是把 e 和 h 看做大前提中的前后件。因为只有这样，才好谈 e 和 h 的概率。既然如此，为了明确任务，笔者建议，干脆用 $c(e \rightarrow h)$ 取代 $c(e, h)$ 。5.3节将说明，这样做还可以避免误解对称性。

2.4 增量确证还是绝对确证？

因为使用一个或数个样例也能得到确证度，所以很多作者把流行的确证度公式算出的确证度解释为确证度增量，称这样的确证是增量确证[14,15]。但是也有反对观点，称我们需要的是绝对确证[29]。本文支持绝对确证。

增量确证的问题是，一两个样例也能算出很大的确证度(增量)，而且和过去的知识(即已知的 a, b, c, d)无关。这是不合理的。

比如，设 h_1 ="x 是老年人"的逻辑概率是0.2，给定证据 $x=65$ 或 $x>60$ ，则 h_1 的条件逻辑概率是0.9。用测度 D 得到 $D=0.9-0.2=0.7$ ，这就是证据对结论的冲击。0.7太大了。

在上面计算 F 、 b^* 和 c^* 的函数 $f(a, b, c, d)$ 中， a 、 b 、 c 、 d 应分别是过去已知的和当前新增的四个样例之和，确证度反映绝对确证。确证度的增量应是：

$$\Delta f = f(a+\Delta a, b+\Delta b, c+\Delta c, d+\Delta d) - f(a, b, c, d). \quad (11)$$

由新样例带来的确证度的增量和旧样例个数火显眼知识紧密相关。5.2节将进一步讨论。

2.5 医学检验的Shannon信道和可信度

考虑医学检验， h 和 e 之间关系如图1所示。 $x \in U$ 表示化验数据。

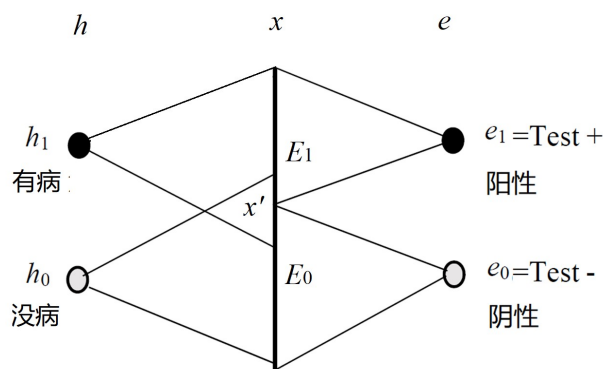


图2. 医学检验中阳性/阴性和有病/无病之间关系

x' 是判断阳性还是阴性的分界点。老年人分类、西瓜分类、垃圾邮件分类...类似。在老年人分类时, h 表示真标签, 即样本中的标签; x 是年龄; e 是分类标签——根据划分点 x' (比如60岁)得到的标签。优化分类可以使用最大正确率准则, 也可以使用最大似然比准则——笔者证明过最大互信息准则与之等价[32]; 我们可以使用信道匹配迭代算法实现最大互信息分类[17,32]。

h 和 x 的关系可用 x 的两个后验概率分布 $P(x|h_0)$ 和 $P(x|h_1)$ 表示, e 和 x 的关系可以用两个清晰集合 E_0 、 E_1 表示。参看图 2。

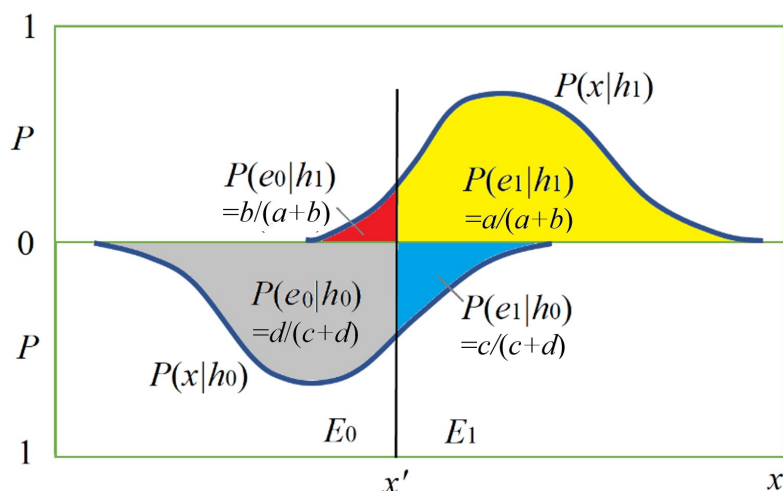


图3. 计算似然比的四个条件概率(等于四个区域面积)

医学检验中把 $P(e_1|h_1)$ 就做敏感性, 把 $P(h_0|e_0)$ 叫做特异性[18], 则它们确定了一个香农信道[25] $P(h|e)$ 如表 2 所示。

表2 医学检验的敏感性和特异性构成Shannon信道 $P(e|h)$

	阴性 e_0	阳性 e_1
有病 h_1	$P(e_0 h_1)=1-\text{敏感性}$	$P(e_1 h_1)=\text{敏感性}$
无病 h_0	$P(e_0 h_0)=\text{特异性}$	$P(e_1 h_0)=1-\text{特异性}$

下面我们定义医学检验的语义信道[17]。令 $e_1(h)$ 是 h 使 e_1 为真(比如感染和没感染使阳性为真)的谓词, 其真值函数是 $T(\theta_{e_1}|h)$, 其中 θ_{e_1} 是模型参数或 $\{h_0, h_1\}$ 上的模糊集合。我们把 $e_1(h)$ 看做是可信部分和不可信部分相加(参看图 4), 可信部分的真值函数是 $T(E_1|h) \in$

$\{0,1\}$, 不可信部分可看作是永真句, 其真值是常数 1。则谓词 $e_1(h)$ 和 $e_0(h)$ 的真值函数可以写成:

$$T(\theta_{e_1}|h) = b_1' + b_1 T(E_1|h); \quad T(\theta_{e_0}|h) = b_0' + b_0 T(E_0|h). \quad (12)$$

b_1' 是不可信部分比例, 也是以反例 h_0 为条件时 e_1 的真值 $T(\theta_{e_1}|h_0)$ 。

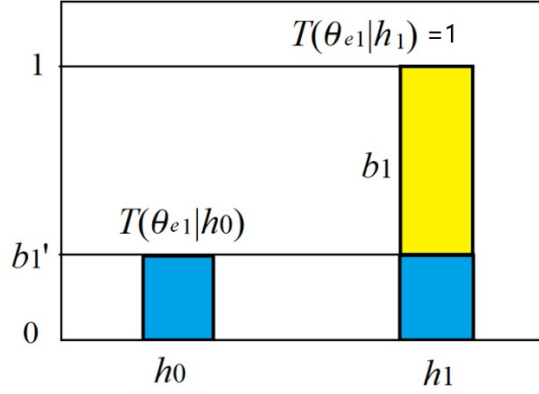


图4 真值函数 $T(\theta_{e_1}|h)$ 写成可信部分 (b_1) 和不可信部分 ($b_1' = 1 - |b_1|$) 相加四个真值构成一个语义信道, 如表 3 所示。

表3. 两个不信心度 b_1' 和 b_0' 确定医学检验的语义信道

	阴性 e_0	阳性 e_1
有病 h_1	$T(\theta_{e_0} h_1) = b_0'$	$T(\theta_{e_1} h_1) = 1$
无病 h_0	$T(\theta_{e_0} h_0) = 1$	$T(\theta_{e_1} h_0) = b_1'$

e_1 的逻辑概率是

$$T(\theta_{e_1}) = \sum_i P(h_i) T(\theta_{e_1} | h_i) = P(h_1) + b_1' P(h_0), \quad (13)$$

似然函数是

$$P(h | \theta_{e_1}) = P(h) T(\theta_{e_1} | h) / T(\theta_{e_1}). \quad (14)$$

$P(h|\theta_j)$ 也是根据 $T(\theta_{e_1}|h)$ 或阳性 e_1 的语义对有病无病的预测。

要度量主观或语义信息, 我们需要主观概率或逻辑概率[17]。要度量确证, 我们需要统计概率。

2.6 语义信息公式和 Nicod-Fisher 准则

根据语义信息 G 理论[17], y_j 传递的关于 x_i 的语义信息(量)被定义为标准对数似然度:

$$I(x_i; \theta_j) = \log \frac{P(x_i | \theta_j)}{P(x_i)} = \log \frac{T(\theta_j | x_i)}{T(\theta_j)}, \quad (15)$$

其中 $T(\theta_j | x_i)$ 是命题 $y_j(x_i)$ 的真值, $T(\theta_j)$ 是 y_j 的逻辑概率。如果 $T(\theta_j | x)$ 总是 1, 该公式就变成 Carnap 和 Bar-Hillel 的语义信息公式 [30]。

在语义通信中, 我们经常看到这样的预测: “温度大约 10 °C”, “时间大概是 7 点钟”, 或 “股市下个月将上升约 10%”。每一个都可以表达为 $y_j = “x \text{ 大约是 } x_j.”$ 我们可以用下面真值函数表示:

$$T(\theta_j | x) = \exp[-(x - x_j)^2 / (2\sigma^2)]. \quad (16)$$

把公式(16)带入(15), 我们有

$$I(x_i; \theta_j) = \log[1/T(\theta_j)] - (x_i - x_j)^2 / (2\sigma^2), \quad (17)$$

用它我们可以解释:语义信息等于 Carnap-Bar-Hillel 语义信息减相对偏差平方. 其图解见图 5.

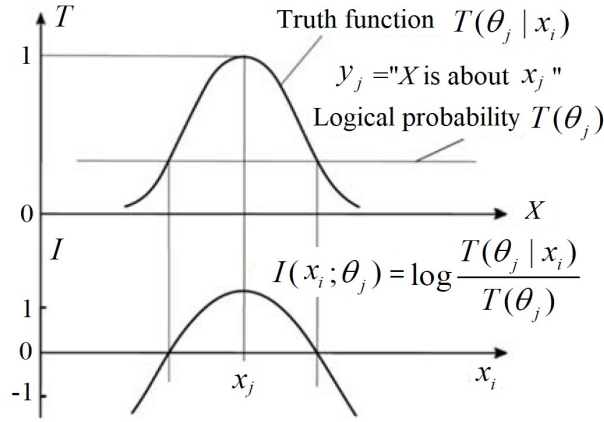


图 5. y_j 提供的关于 x_i 的语义信息

图 5 表明:逻辑概率越小,信息量越大;偏差越大,信息量越小;错误假设提供的信息是负的. 这些结论符合 Popper 思想(见 [2], p. 294).

平均 $I(x_i; \theta_j)$ 可以得到广义 Kullback-Leibler 信息或相对熵:

$$I(X; \theta_j) = \sum_i P(x_i | y_j) \log \frac{P(x_i | \theta_j)}{P(x_i)} = \sum_i P(x_i | y_j) \log \frac{T(\theta_j | x_i)}{T(\theta_j)}, \quad (1)$$

其中 $P(x | y_j)$ 是样本分布, $P(x | \theta_j)$ 是似然函数. If $P(x | \theta_j)$ 等于 $P(x | y_j)$, 则 $I(X; \theta_j)$ 达到其最大值——相对熵或 Kullback-Leibler 离散度.

考虑医学检验, e_1 提供关于 h 的语义信息量是

$$I(h; \theta_{e_1}) = \log \frac{P(h_i | \theta_{e_1})}{P(h_i)} = \log \frac{T(\theta_{e_1} | h)}{T(\theta_{e_1})}. \quad (19)$$

e_1 提供关于 h 的平均语义信息量是:

$$I(H; \theta_{e_1}) = \sum_{i=0}^1 P(h_i | e_1) \log \frac{T(\theta_{e_1} | h_i)}{T(\theta_{e_1})} = \sum_{i=0}^1 P(h_i | e_1) \log \frac{P(h_i | \theta_{e_1})}{P(h_i)}, \quad (20)$$

其中 H 表示取值为 h 的随机变量; $P(h_i | e_1)$ 是来自样本的条件概率.

下面看似然度和平均信息 $I(H; \theta_{e_1})$ 之间的关系. 设 $\mathbf{D}$ 是样本 $\{(h(t), e(t)) | t = 1 \text{ to } N; h(t) \in \{h_0, h_1\}; e(t) \in \{e_0, e_1\}\}$, 它包括两个子样本: 带有 e_0 的 $\mathbf{H}_0$ 和带有 e_1 的 $\mathbf{H}_1$. 当 N 个数据点来自独立同分布随机变量时, 我们有似然度:

$$\begin{aligned} L(\theta_{e_1}) &= \log P(\mathbf{H}_1 | \theta_{e_1}) = \log P(h(1), h(2), \dots, h(N) | \theta_{e_1}) = \log \prod_{i=0}^1 P(h_i | \theta_{e_1})^{N_{1i}} \\ &= N_1 \sum_{i=0}^1 P(h_i | e_1) \log P(h_i | \theta_{e_1}) = -N_1 H(H | \theta_{e_1}). \end{aligned} \quad (21)$$

其中 N_{1i} 是 $\mathbf{D}$ 中 (h_i, e_1) 的个数, N_1 是 $\mathbf{H}_1$ 中样例的个数. 比较上面两个公式可见

$$I(H; \theta_{e_1}) = L(\theta_{e_1}) / N_1 - \sum_{i=0}^1 P(h_i | e_1) \log P(h_i). \quad (22)$$

因为上式第二项和模型无关，所以最大似然准则等价于最大平均语义信息准则。容易看出，正例(e_1, h_1)增加平均对数似然度 $L(\theta_{e_1})/N_1$ ，反例(e_0, h_1)减少它，(e_0, h_0)和(e_0, h_1)对它没影响。

关于确证的 Nicod 准则(NC)是：正例(e_1, h_1)支持推理 $e_1 \rightarrow h_1$ ，反例(e_1, h_0)否定该推理。已有文献没有表明 Nicod 是否肯定 (e_0, h_1)和(e_0, h_1)对 $e_1 \rightarrow h_1$ 没有影响。如果 Nicod 没有肯定，我们可以加上它。我们可以称相应的准则是 Nicod-Fisher 准则 (因为 Fisher 提出最大似然估计)，简称 NFC。

2.7 选择假设和确证大前提——统计学习的两个任务

研究者们已经注意到确证测度和信息测度的相似性。一种解释是[31]信息是确证冲击的平均。然而，本文解释不同。

统计学习中有三个任务：标签学习、分类和计算可靠性(提供似然比或正确率)。归纳推理中有类似任务：

- 归纳——就是标签学习；对于不确定假设，标签学习就是求似然函数 $P(x|\theta_j)$ 或真值函数 $T(\theta_j|x)$ (其中 θ_j 是使 h_j 为真的 x 构成的模糊集合。常见的 Logistic 函数就可以看做二分类时的真值函数)。
- 选择假设——就是根据正确率、似然度，或信息准则建立分类器 $e=f(x)$ 为 x 分类。
- 确证——计算确证度就是计算可靠性，经典方法是提供似然比和正确率(包括错误率)。

分类和计算可靠性是两个不同任务；同样，选择假设和提供确证度是两个不同任务。分类需要准则，常见的有最大后验概率准则(它等价于最大正确率准则)，分类器是：

$$e(x) = \begin{cases} e_1, & \text{如果 } P(\theta_1 | x) \geq P(\theta_0 | x) \text{ 或 } P(x | \theta_1) \geq P(x | \theta_0); \\ e_0, & \text{否则.} \end{cases} \quad (23)$$

分类后可以用信息测度 $I(h_j; \theta_{e_j})$ 或 $I(h_j; e_j)$ 评价预测。如果预测模型 θ_{e_j} 是用样本优化的，则

$$\begin{aligned} I(h_j; \theta_{e_j}) &= I(h_j; e_j) = \log \frac{P(h_j | e_j)}{P(h_j)} = \log \frac{P(e_j | h_j)}{P(e_j)} \\ &= \log P(h_j | e_j) - \log P(h_j) = \log P(e_j | h_j) - \log P(e_j) \\ &= \log P(h_j, e_j) - \log [P(h_j)P(e_j)]. \end{aligned} \quad (24)$$

其中预测模型 θ_{e_j} 是用样本分布 $P(h|e_j)$ 优化的，使得 $P(h|\theta_{e_j})=P(h|e_j)$ 且 $I(h_j; \theta_{e_j})=I(h_j; e_j)$ 。

我们可以优化分类器[17]用下面公式：

$$e_j^* = f(x) = \arg \max_{e_j} [P(h_0 | x)I(h_0; \theta_{e_j}) + P(h_1 | x)I(h_1; \theta_{e_j})]. \quad (2)$$

新的分类器将提供新的 Shannon 信道 $P(e|h)$ 。重复公式(23)和(25)就可以实现最大互信息分类即最大似然比分类[32, 17]。

有了上述分类器，我们就可以根据输入 x 输出 e —— e_j 是对 h_j 的预测。为了告诉信息接受者预测有多准确，我们还要提供似然比 LR 表示信道(因为有对错，应该是语义信道)有多好，或者提供正确率表示预测有多好。确证是类似的，就是要提供和 LR 相似的确证度，比如 F 和 b^* ，或提供和正确率相似的确证度 c^* 。但是要保证确证度在-1 和 1 之间变化。

预测的信息评价和大前提的确证的主要区别是，信息评价用信息测度分别评价每一次预测(用 e 预测 h)，而确证是用一个确证测度和一个样本(包括许多样例(e, h)作为事实)评价一个大前提。

统计学习中, 信息接受者根据收到的 e_j 、 LR_j 和 $P(h)$ 做概率预测; 或根据 e_j 和正确率做概率预测。类似地, 接受者可以根据收到的 e_j 、 b_j^* (或 F_j) 和 $P(h)$ 做概率预测, 或根据 e_j 和 c_j^* 做概率预测。

从上面分析可见。流行的测度中, 测度 D 、 M 、 R 和 C 用概率的增量或增量的对数, 看起来都更像是信息测度或语义信息测度 (因为度量语义信息用逻辑概率或主观概率[17]), 或评价和选择预测的测度, 而不是确证大前提的测度。Z 是这些测度的归一化[7], 似乎介于两者之间。

而确证和评价预测不同, 要用到正反例比例, 和计算似然比或正确率类似。

3. 两个新的确证测度

3.1 推导信道确证度 b^*

我们使用最大语义信息准则——它和最大似然准则一致——推导信道确证测度。根据公式(13) 和 (18), e_1 传递关于 h 的平均语义信息是

$$I(h; \theta_{e_1}) = P(h_0 | e_1) \log \frac{b_1'}{P(h_1 + b_1' P(h_0))} + P(h_1 | e_1) \log \frac{1}{P(h_1 + b_1' P(h_0))} \quad (3)$$

令 $dI(h; \theta_{e_1})/db_1' = 0$, 可以得到优化的不信心度 b_1' :

$$b_1'^* = \frac{P(h_0 | e_1)}{P(h_0)} \bigg/ \frac{P(h_1 | e_1)}{P(h_1)}, \quad (4)$$

其中 $P(h_1 | e_1) / P(h_1) \geq P(h_0 | e_1) / P(h_0)$. $b_1'^*$ 可谓否定测度. 分子和分母同乘以 $P(e_1)$, 上面公式就变为:

$$b_1'^* = P(e_1 | h_0) / P(e_1 | h_1) = (1 - \text{特异性}) / \text{敏感性} = 1 / LR^+. \quad (5)$$

根据语义信息 G 理论[17], 当真值函数正比于转移概率函数时, 即 $T^*(\theta_{e_1} | h) \propto P(e_1 | h_1)$, 平均语义信息最大. 令真值函数 $T^*(\theta_{e_1} | h)$ 正比于转移概率函数 $P(e_1 | h_1)$, 我们也可以得到

$$\frac{b_1'^*}{P(e_1 | h_0)} = \frac{1}{P(e_1 | h_1)} \quad (6)$$

和公式 (28). 我们称

$$b_1^* = 1 - b_1'^* = [P(e_1 | h_1) - P(e_1 | h_0)] / P(e_1 | h_1) \quad (30)$$

是大前提 $e_1 \rightarrow h_1$ 的确证度。上面假设 $P(h_1 | e_1) \geq P(h_0 | e_1)$, 否则, $b_1'^* = P(e_1 | h_1) / P(e_1 | h_0)$, 确证度是负的:

$$b_1^* = b_1'^* - 1 = [P(e_1 | h_0) - P(e_1 | h_1)] / P(e_1 | h_0). \quad (31)$$

结合上面两个公式得到

$$b_1^* = b^*(e_1 \rightarrow h_1) = \frac{P(e_1 | h_1) - P(e_1 | h_0)}{\max[P(e_1 | h_1), P(e_1 | h_0)]} = \frac{LR^+ - 1}{\max(LR^+, 1)}. \quad (32)$$

因为

$$b_1^* = b^*(e_1 \rightarrow h_0) = \frac{P(e_1 | h_0) - P(e_1 | h_1)}{\max[P(e_1 | h_0), P(e_1 | h_1)]} = -b^*(e_1 \rightarrow h_1), \quad (7)$$

b_1^* 具有 HS 即后件对称性.

同理可得

$$b_0^* = b^*(e_0 \rightarrow h_0) = \frac{P(e_0|h_0) - P(e_0|h_1)}{\max[P(e_0|h_0), P(e_0|h_1)]} = \frac{LR^- - 1}{\max(LR^-, 1)}. \quad (34)$$

其中 $LR^- = P(h_0|e_0)/P(h_1|e_0)$ 是阴性似然比。应用后件对称性, 可以得到另外两种信道确证测度: $b^*(e_1 \rightarrow h_0) = -b^*(e_1 \rightarrow h_1)$ 和 $b^*(e_0 \rightarrow h_1) = -b^*(e_0 \rightarrow h_0)$ 。

用 b^* 或 F , 我们可以回答问题: 当核算检验是阴性, CT 检查是阳性, 我们应当信哪个? 4.2 节将提供和改进的武汉 COVID-19 诊断方法相一致的结论。

和 F 相比, b^* 用于概率预测更加方便。比如从 b_1^* 和 $P(h)$, 可以得到

$$P(h_1|\theta_{e1}) = P(h_1) / [P(h_1) + b_1^* P(h_0)] = P(h_1) / [1 - b_1^* P(h_0)]. \quad (35)$$

如果 $b_1^* = 0$, 则 $P(h_1|\theta_{e1}) = P(h_1)$ 。如果 $P(h_1|\theta_{e1}) < 0$, 则我们可以利用假设对称性 (HS) 得到 $b_{10}^* = b_1^*(e_1 \rightarrow h_0) = |b_1^*(e_1 \rightarrow h_1)|$ 。然后获得

$$P(h_0|\theta_{e1}) = P(h_0) / [P(h_0) + b_{10}^* P(h_1)] = P(h_0) / [1 - b_{10}^* P(h_1)]. \quad (36)$$

从 $F_1 = F(e_1 \rightarrow h_1)$ 可以算出 $b_1^* = 2F_1 / (1 + F_1)$,

所以 F 测度也可以用作概率预测, 但是计算要复杂些。

然而, 用 b^* 或 F 解释乌鸦悖论还是有问题。比如, d 增加 1 和 a 增加 1 引起 $b^*(e_1 \rightarrow h_1)$ 的增量可能很接近, 这意味着一支白粉笔和一只黑乌鸦一样支持“乌鸦是黑的”。所以 b^* 和 F 不符合 Nicod-Fisher 准则。 b^* 和 Z 也不符合。

为什么用最大似然准则得到 b^* 测度不能直接反映 NF 准则? 原因 $L(\theta_{e1})$ 和 $P(h)$ 相关, 而 b^* 和 $P(h)$ 无关。

3.2 推导预测确证度 c^*

统计理论中用似然比反映检验手段(作为信道)的好坏, 还用正确率反映概率预测好坏。 F 、 b^* 和似然比类似, 也不能直接反映概率预测好坏。

比如, 设用于 COVID-19 的核酸检验 [33] 的敏感性是 $P(e_1|h_1) = 0.5$, 特异性是 $P(e_1|h_0) = 0.95$ 。可以算出 $b_1^* = 0.1$ 和 $b_1^* = 0.9$ 。当基础概率(即先验概率) $P(h_1)$ 变化时, 根据阳性预测的概率 $P(h_1|\theta_{e1})$ (又叫阳性预测值) 也会变化。根据公式(35)得到的结果见表 4。用经典统计方法可以得到同样结果(见公式(5)), 只是计算少许复杂。

表4 确证度 $b_1^* = 0.9$ 时预测的有病概率 $P(h_1|\theta_{e1})$ 随先验概率 $P(h_1)$ 变化

	安全人群	普通人群	高危人群
$P(h_1)$	0.001	0.1	0.25
$P(h_1 \theta_{e1})$	0.002	0.19	0.77

可见 b^* 或 F 不能直接反映概率预测质量。为此, 我们需要利用 $P(h)$ 定义和正确率相似的确证度。

我们把 $P(h|\theta_{e1})$ 看作是有用部分和无用部分的组合, 称有用部分的比例 c_1 是规则 $e_1 \rightarrow h_1$ 的预测信任度, 参看图 6。

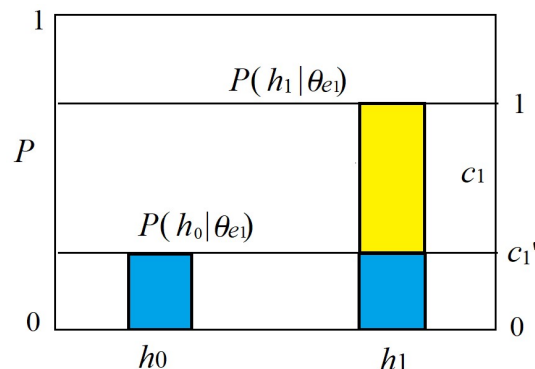


图6. 似然函数 $P(h|\theta_{e1})$ 可以写成可信部分 c_1 和不可信部分 c_1' 相加.

当预测和事实符合即 $P(h|\theta_{e1}) = P(h|e_1)$ 时, c_1 就变成预测确证度 c_1^* . 于是可以推导出预测否证度

$$c^*(e_1 \rightarrow h_1) = \min(P(h_0|e_1)/P(h_1|e_1), P(h_1|e_1)/P(h_0|e_1)). \quad (37)$$

进一步可以得到

$$\begin{aligned} c^*(e_1 \rightarrow h_1) &= \frac{P(h_1|e_1) - P(h_0|e_1)}{\max(P(h_1|e_1), P(h_0|e_1))} \\ &= \frac{2P(h_1|e_1) - 1}{\max(P(h_1|e_1), 1 - P(h_1|e_1))} = \frac{2CR_1 - 1}{\max(CR_1, 1 - CR_1)}. \end{aligned} \quad (38)$$

其中 $CR_1 = P(h_1|\theta_{e1}) = P(h_1|e_1)$, 它是大前提 $e_1 \rightarrow h_1$ 的正确率, 表示 x 在 E_1 中时, $h=h_1$ 的概率是 CR_1 . 上式分子分母同乘以 $P(e_1)$, 于是得到

$$c_1^* = c^*(e_1 \rightarrow h_1) = \frac{P(h_1, e_1) - P(h_0, e_1)}{\max(P(h_1, e_1), P(h_0, e_1))} = \frac{a - c}{\max(a, c)}. \quad (39)$$

图7中两条曲线覆盖的四个区域面积就分别和 a 、 b 、 c 、 d 成正比。

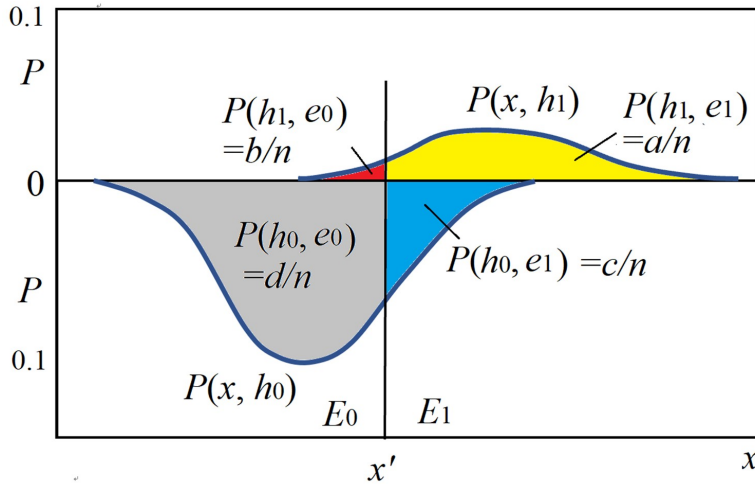


图7. 预测确证度 $c^*(e_0 \rightarrow h_0)$ (看左边) 和 $c^*(e_1 \rightarrow h_1)$ (看右边) 的正例和反例个数

同理, 我们可以得到

$$\begin{aligned} c_0^* = c^*(e_0 \rightarrow h_0) &= \frac{P(h_0|e_0) - P(h_1|e_0)}{\max(P(h_0|e_0), P(h_1|e_0))} \\ &= \frac{P(h_0, e_0) - P(h_1, e_0)}{\max(P(h_0, e_0), P(h_1, e_0))} = \frac{d - b}{\max(d, b)}. \end{aligned} \quad (40)$$

应用后件对称性, 我们可以得到另外两个预测确证度: $c^*(e_1 \rightarrow h_0) = -c^*(e_1 \rightarrow h_1)$ 和 $c^*(e_0 \rightarrow h_1) = -c^*(e_0 \rightarrow h_0)$.

从图7可见, 计算预测确证度 c^* 所用的两条函数曲线覆盖的两个区域面积不同, 分别是 $P(h_1)$ 和 $P(h_0)$. 如果 $P(h_1) = P(h_0) = 0.5$, 预测确证度 c^* 就等于信道确证度 b^* .

c^* 也可以用来评估概率预测的质量. 比如, 在 $P(h_1|\theta_{e1}) = 0.77$ (见表4), 我们有 $c_1^* = (0.77 - 0.23)/0.77 = 0.701$. 我们也能从 c^* 得到概率预测. 设 $e=e_1$ 且 c_1^* 大于0. 则由式(39)可得规则 $e_1 \rightarrow h_1$ 的正确率:

$$P(h_1 | \theta_{e1}) = CR_1 = 1 / (1 + c_1^*) = 1 / (2 - c_1^*). \quad (41)$$

比如, 如果 $c_1^* = 0.701$, 则 $CR_1 = 1 / (2 - 0.701) = 0.77$. 如果 $c^*(e_1 \rightarrow h_1) = 0$, then $CR_1 = 0.5$. 如果 $c^*(e_1 \rightarrow h_1) < 0$, 我们可以利用假设对称性(HS)或后件对称性得到 $c_{10}^* = c^*(e_1 \rightarrow h_0) = |c_1^*|$, 然后做概率预测:

$$\begin{aligned} P(h_0 | \theta_{e1}) &= 1 / (2 - c_{10}^*), \\ P(h_1 | \theta_{e1}) &= 1 - P(h_0 | \theta_{e1}) = (1 - c_{10}^*) / (2 - c_{10}^*). \end{aligned} \quad (42)$$

我们也可以定义和 F 类似的预测确证度——记为 c_F^* (用 $+$ 代替 $\vee$):

$$\begin{aligned} c_{F1} = c_F^*(e_1 \rightarrow h_1) &= \frac{P(h_1 | e_1) - P(h_0 | e_1)}{P(h_1 | e_1) + P(h_0 | e_1)} = \frac{P(h_1 | e_1) - P(h_0 | e_1)}{P(e_1)} \\ &= \frac{P(h_1, e_1) - P(h_0, e_1)}{P(e_1)} = \frac{a - c}{a + c}. \end{aligned} \quad (43)$$

和 b^* 相比, c_F^* 在 $P(h)$ 固定时做概率预测也很方便。我们有

$$\begin{aligned} P(h_1 | \theta_{e1}) &= CR_1 = (1 + c_{F1}^*) / 2; \\ P(h_0 | \theta_{e1}) &= 1 - CR_1 = (1 - c_{F1}^*) / 2. \end{aligned} \quad (44)$$

但是在 $P(h)$ 变化时, 还是要用信道确证度 b^* 和 $P(h)$ 做概率预测.

容易证明 $c^*(e_1 \rightarrow h_1)$ 和 $c_F^*(e_1 \rightarrow h_1)$ 具有上面所有对称不对称性、单调性、逻辑性和归一性。

3.3 反信道确证度 $b^*(h \rightarrow e)$ 和反预测确证度 $c^*(h \rightarrow e)$

Greco 等人[19]把确证测度分为:

贝叶斯确证测度——条件概率是 $P(h|e)$, 大前提是 $e \rightarrow h$;

和似然确证测度——条件概率是 $P(e|h)$, 大前提是 $e \rightarrow h$;

反贝叶斯确证测度——条件概率是 $P(h|e)$, 大前提是 $h \rightarrow e$;

反似然确证测度——条件概率是 $P(e|h)$, 大前提是 $h \rightarrow e$.

本文的确证测度划分是:

信道确证测度 $b^*(e \rightarrow h)$;

预测确证测度 $c^*(e \rightarrow h)$;

反信道确证测度 $b^*(h \rightarrow e)$;

反预测确证测度 $c^*(h \rightarrow e)$.

我们先考虑 $c^*(h_1 \rightarrow e_1)$. 如果我们已知有病的特征分布 $P(x|h_1)$ (图 3 中上部分函数曲线) 并且需要向看不到函数曲线 $P(x|h_1)$ 的人解释它, 我们可以用自然语言 “如果 x 有病, 则 x 很可能检验呈阳性” 传递信息。要解释 “老年人” 的含义, 我们可能说 “如果一个被称为老年人, 他很可能大于 60 岁”。这种语句可以抽象为 $h_1 \rightarrow e_1$. 相应的确证度应该是反预测确证测度 $c^*(h_1 \rightarrow e_1)$. 这时候我们需要在图 7 中横向找正反比例。于是得到:

$$c^*(h_1 \rightarrow e_1) = \frac{P(e_1 | h_1) - P(e_0 | h_1)}{\max(P(e_1 | h_1), P(e_0 | h_1))} = \frac{a - b}{\max(a, b)}. \quad (45)$$

反预测确证测度 $c^*(h_1 \rightarrow e_1)$ 反映的正确率就是敏感性 $P(h_1|e_1)$; $c^*(h_1 \rightarrow e_1)$ 反映的正确率就是特异性 $P(h_0|e_0)$.

现在考虑反信道确证测度 $b^*(h_1 \rightarrow e_1)$. 现在 $P(e)$ 被认为是信源, $P(h)$ 被认为是信宿。我们需要把 $b^*(e_1 \rightarrow h_1)$ 中 h 和 e 互换, 或 a 和 d 互换及 b 和 c 互换。于是得到:

$$b^*(h_1 \rightarrow e_1) = \frac{P(h_1 | e_1) - P(h_1 | e_0)}{P(h_1 | e_1) \vee P(h_1 | e_0)} = \frac{ad - bc}{a(b+d) \vee b(a+c)}. \quad (46)$$

其中用模糊逻辑取大运算 $\vee$ 代替了 $\max()$ 函数。

反信道确证测度和反预测确证测度也有4种(参看表7)。考虑后件对称性,一共有8种。

3.4 前后件不同的八种确证公式

表5显示了确证测度 b^* 需要的正例和反例的比例。

表5. 计算 $b^*(e \rightarrow h)$ 和 $c^*(e \rightarrow h)$ 用到的四个正反例比例

	e_0 (阴性)	e_1 (阳性)
h_1 (感染)	$P(e_0 h_1) = b/(a+b)$	$P(e_1 h_1) = a/(a+b)$
h_0 (没感染)	$P(e_0 h_0) = d/(c+d)$	$P(e_1 h_0) = c/(c+d)$
h_1 (感染)	$P(h_1 e_0) = b/(b+d)$	$P(h_1 e_1) = a/(a+c)$
h_0 (没感染)	$P(h_0 e_0) = d/(b+d)$	$P(h_0 e_1) = c/(a+c)$

表6提供了规则 $e \rightarrow h$ 的4种确证公式。

表6 信道确证测度 $b^*(e \rightarrow h)$ 和预测确证测度 $c^*(e \rightarrow h)$

	信道确证度 $b^*(e \rightarrow h)$ (参看图3)	预测确证度 $c^*(e \rightarrow h)$ (参看图7)
$e_1 \rightarrow h_1$	$\frac{P(e_1 h_1) - P(e_1 h_0)}{P(e_1 h_1) \vee P(e_1 h_0)} = \frac{ad - bc}{a(c+d) \vee c(a+b)}$	$\frac{P(h_1 e_1) - P(h_0 e_1)}{P(h_1 e_1) \vee P(h_0 e_1)} = \frac{a - c}{a \vee c}$
$e_0 \rightarrow h_0$	$\frac{P(e_0 h_0) - P(e_0 h_1)}{P(e_0 h_0) \vee P(e_0 h_1)} = \frac{ad - bc}{d(a+b) \vee b(c+d)}$	$\frac{P(h_0 e_0) - P(h_1 e_0)}{P(h_0 e_0) \vee P(h_1 e_0)} = \frac{d - b}{d \vee b}$

这些公式和预测规则 $e \rightarrow h$ 的误报率相关。比如, $b^*(e_1 \rightarrow h_1)$ 或 $c^*(e_1 \rightarrow h_1)$ 更小意味着更多未感染者检验为阳性。

表7提供了规则 $h \rightarrow e$ 的4种确证公式。这些公式和推理 $h \rightarrow e$ 的漏报率相关。

表7 反信道确证测度 $b^*(h \rightarrow e)$ 和反预测确证测度 $c^*(h \rightarrow e)$

	反信道确证度 $b^*(h \rightarrow e)$	反预测确证度 $c^*(h \rightarrow e)$ (参看图7)
$h_1 \rightarrow e_1$	$\frac{P(h_1 e_1) - P(h_1 e_0)}{P(h_1 e_1) \vee P(h_1 e_0)} = \frac{ad - bc}{a(b+d) \vee b(a+c)}$	$\frac{P(e_1 h_1) - P(e_0 h_1)}{P(e_1 h_1) \vee P(e_0 h_1)} = \frac{a - b}{a \vee b}$
$h_0 \rightarrow e_0$	$\frac{P(h_0 e_0) - P(h_0 e_1)}{P(h_0 e_0) \vee P(h_0 e_1)} = \frac{ad - bc}{d(a+c) \vee c(b+d)}$	$\frac{P(e_0 h_0) - P(e_1 h_0)}{P(e_0 h_0) \vee P(e_1 h_0)} = \frac{d - c}{d \vee c}$

这些确证侧度和规则 $h \rightarrow e$ 的漏报率相关。比如, 更小的 $b^*(h_1 \rightarrow e_1)$ 或 $c^*(h_1 \rightarrow e_1)$ 意味着更多感染者检验显示为阴性。漏报是更加严重的问题。

上面8种确证测度, 每个测度都有其对称形式, 所以一共有16个包含 a 、 b 、 c 、 d 的确证公式。

在预测和反预测确证测度中,其中条件概率的条件都是推理的前件,从而使得确证测度只和正反例个数相关。因此,这些测度符合 Nicod-Fisher 准则。

对于测度 b^* 和 c^* , 我们把取大运算 $\vee$ 换成加运算 $+$, 相应的 b^* 测度就变为 $b_F^*=F$; c^* 测度变为 c_F^* . 比如:

$$c_F^*(e_1 \rightarrow h_1) = (a-c)/(a+c). \quad (47)$$

3.5 确证测度 b^* 和 F 的关系

b^* 和 F 非常类似, 唯有分母不同, b^* 和 F 一样符合所有对称性要求以及上述其他要求。两者随似然比的变化如图8所示。

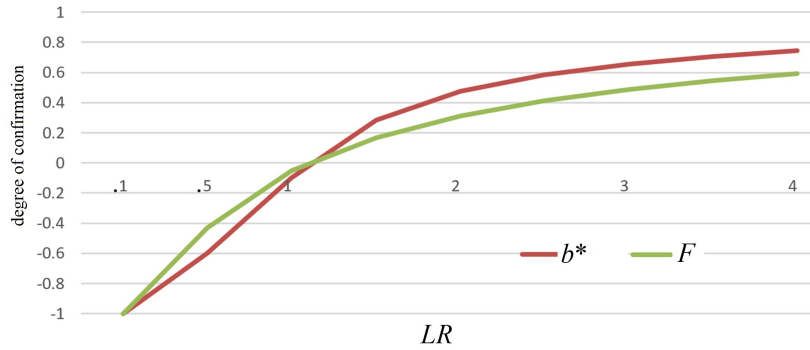


图8. b^* 和 F 随似然比 LR 的变化。

测度 F 有四种形式(参看 Greco 等人的文章[8]):

$$F(e_1 \rightarrow h_1) = \frac{P(e_1 | h_1) - P(e_1 | h_0)}{P(e_1 | h_1) + P(e_1 | h_0)} = \frac{ad - bc}{ad + bc + 2ac} = b_F^*(e_1 \rightarrow h_1), \quad (48)$$

$$F(h_1 \rightarrow e_1) = F(e_1 \rightarrow h_1) = \frac{P(h_1 | e_1) - P(h_1 | e_0)}{P(h_1 | e_1) + P(h_1 | e_0)} = \frac{ad - bc}{ad + bc + 2ab} = b_F^*(h_1 \rightarrow e_1), \quad (49)$$

$$F(e_0 \rightarrow h_0) = \frac{P(e_0 | h_0) - P(e_0 | h_1)}{P(e_0 | h_0) + P(e_0 | h_1)} = \frac{ad - bc}{ad + bc + 2bd} = b_F^*(e_0 \rightarrow h_0), \quad (50)$$

$$F(h_0 \rightarrow e_0) = \frac{P(h_0 | e_0) - P(h_0 | e_1)}{P(h_0 | e_0) + P(h_0 | e_1)} = \frac{ad - bc}{ad + bc + 2cd} = b_F^*(h_0 \rightarrow e_0). \quad (51)$$

F 和 b_F^* 等价, 和 b^* 类似。测度 b^* 和 F 一样, 具有上述各种所希望的性质。差别是: b^* 的绝对值更大; b^* 用于概率预测更加方便(见公式(35))。

3.6 预测确证度 c^* 和各种医学检验指标的关系

信道确证度和似然比相关, 而预测确证度和各种正确率错误率相关。为帮助我们理解各种预测确证度的意义, 表 8 包含预测确证度(包括反预测确证度)和医学检验各种指标(正确率和错误率)的关系。

表 8. 预测确证测度在医学检验中和不同正确率[27]和错误率相关

预测确证度	正相关的正确率	负相关的错误率
$c^*(e_1 \rightarrow h_1)$	$P(h_1 e_1)$: 阳性预测值	$P(h_0 e_1)$: 错误发现率
$c^*(e_0 \rightarrow h_0)$	$P(h_0 e_0)$: 阴性预测值	$P(h_1 e_0)$: 错误遗漏率
$c^*(h_1 \rightarrow e_1)$	$P(e_1 h_1)$: 敏感性或真阳性率	$P(e_0 h_1)$: 假阴性率

$c^*(h_0 \rightarrow e_0)$	$P(e_0 h_0)$: 特异性或真阴性率	$P(e_1 h_0)$: 假阳性率
和预测确证测度相关的错误率都是推理 $e \rightarrow h$ 的误报率，而反预测确证度相关的错误率都是推理 $h \rightarrow e$ 的漏报率。比如,错误发现率 $P(h_0 e_1)$ 就是规则 $e_1 \rightarrow h_1$ 的误报率，而假阴性率 $P(e_0 h_1)$ 就是其漏报率。		

4. 结果

4.1 用两个例子比较不同确证测度

在中国抵抗新冠病毒的战役中,人们经常问,既然核酸检验的真阳性率或敏感性这么低(低于 0.5),为什么我们还是相信它? 专家回答说:虽然核酸检验敏感性低,但是它的特异性高,因而它的阳性是非常可信的。

我们用两个极端的例子(参看图 9)解释为什么一个低敏感性的检验能比一个高敏感性的检验提供更加可信的阳性, 以及流行的确证侧度是否支持这个结论。

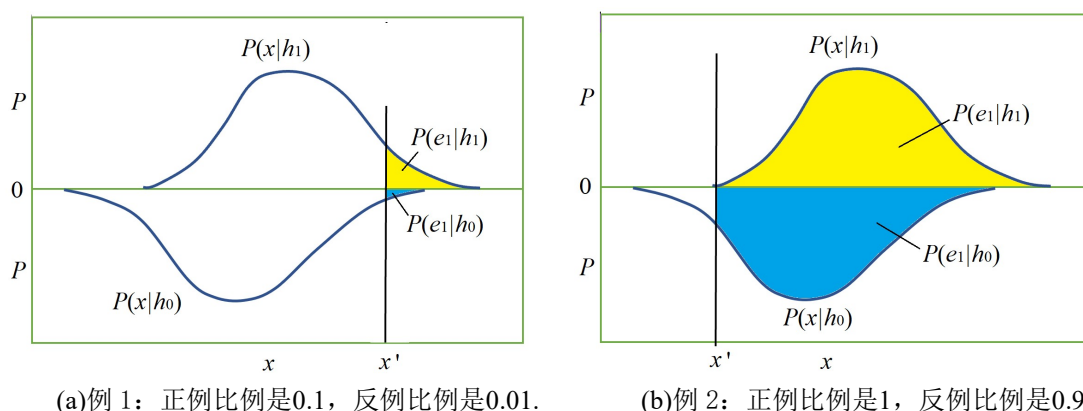


图9. 正反例比例对确证测度 b^* 的影响

对于例 1, $b^*(e_1 \rightarrow h_1) = (0.1 - 0.01) / 0.1 = 0.9$, 较大。对于例 2, $b^*(e_1 \rightarrow h_1) = (1 - 0.9) / 1 = 0.1$, 较小。可见要增大 b^* , 较少的反例比较多正例更加重要。确证测度 c^* 、 F 和 c_F^* 同 b^* 一样, 也对反例比例更加敏感。而其他大多数确证测度不具有这个性质。

进一步假定, 对于两个例子都是 $P(h_1) = 0.2$, $P(h_0) = 0.8$, 于是可用不同确证测度算出例 1 和例 2 的确证度(参看表 9)。其中 $\log$ 以 2 为底。表中也包含了例 3(其中 $P(h_1) = 0.01$), 它揭示了 Z 和 b^* 或 F 的差别。

表 9 从三个例子看不同确证测度的差别(补充材料提供了一个 excel 文件, 它包含表 9, 12, 13 的数据计算)。

Ex.	a, b, c, d	D	M	R	C	Z	S	N	L	F	b^*	c^*
1	20, 180, 8, 792	.514	<u>.072</u>	1.84	<u>.014</u>	.643	.529	<u>.09</u>	3.32	.818	.9	0.8
2	200, 0, 720, 80	.017	<u>.08</u>	0.12	<u>.016</u>	.022	.217	<u>.1</u>	.152	.053	.1	-.722
3	10, 0, 90, 900	.09	.9	3.32	.009	.091	.1	.091	3.46	.833	.91	-.9

例 1 和例 2 的数据显示, F 和 b^* 给与例 1 比例 2 更高评价; 而 M, C, N 给与例 2 比例 1 更高的评价(见红色数字)。

对于例 2 和例 3, $c > a$, 只有 $c^*(e_1 \rightarrow h_1)$ 的值是负的。在反例多于正例时, 负值应该是合理的。例 3 的数据显示,

第三行显示, 在 $P(h_0) = 0.99 \gg P(h_1) = 0.01$ 时, Z 和 F 及 b^* 非常不同(见蓝色数字), 因为 F 和 b^* 独立于 $P(h)$ 。

3.2 用测度 b^* 解释为什么及怎样 CT 和核算同时用于 COVID-19 检验

COVID-19 爆发于武汉, 感染了很多。早期只有用核酸检验诊断感染, 后来医生们发现核酸检验经常漏报感染。因为检验敏感性低(可能低于 0.5)而特异性高, 所以检验呈阳性时我们可以确认感染, 但是检验呈阴性时, 我们不好确认未感染。这意味着核酸检验阴性不可信。为减少漏报, CT 得到更多重视, 因为 CT 有更高敏感性。

在核酸和 CT 同时使用时, 武汉医生改进诊断如图 10 和表 11 所示。如果根据确证测度 b^* , 我们能和武汉医生一样诊断感染吗? 除了核酸和 CT, 病人症状比如发热和咳嗽也被用来诊断。为了简化问题, 我们假设所有病人症状相同, 诊断只根据核酸和 CT 检验结果。

文献[34]介绍了作者使用 CT 的敏感性和特异性。根据[33, 34]和其他网上报告, 本文作者假定的核酸和 CT 的敏感性和特异性如表 10 所示。

表 10. 核酸检验 (NAT) 和 CT 对新冠肺炎的敏感性和特异性

	敏感性	特异性
NAT	0.5	0.95
CT	0.8	0.75

图 10 是根据表 10 画的。图 10 也显示了敏感性和特异性。比如红色圆圈有一半在右侧, 表示核酸检验的敏感性是 0.5。

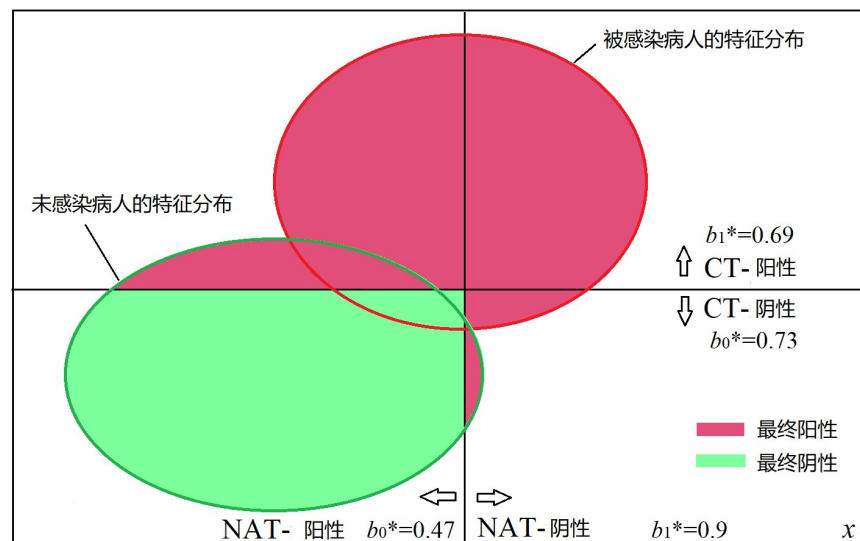


图 10. 借助于确证测度 b^* , 同时使用核酸(NAT)和 CT 诊断新冠病毒感染。

我们用 $c(\text{NAT}+)$ 表示使用任一确证测度 c 时的核酸阳性的确证度, 使用 $c(\text{NAT}-)$, $c(\text{CT}+)$, 和 $c(\text{CT}-)$ 同理。于是我们有

$$b^*(\text{NAT}+) = [P(e_1|h_1) - P(e_1|h_0)]/P(e_1|h_1) = [0.5 - (1 - 0.95)]/0.5 = 0.9;$$

$$b^*(\text{NAT}-) = [P(e_0|h_0) - P(e_0|h_1)]/P(e_0|h_0) = [0.95 - (1 - 0.5)]/0.95 = 0.47.$$

同理, 我们也可以获得 $b^*(\text{CT}+) = 0.69$ 和 $b^*(\text{CT}-) = 0.73$ (见表 Table 11)。

表 11. 根据核酸检验和 CT 所做的改进诊断 (得到最终阳性和阴性)。

	NAT-Negative, $b_0^*=0.47$	NAT-Positive, $b_1^*=0.9$
CT-阳性, $b_1^*=0.69$	最终阳性(变了)	最终阳性
CT-阴性, $b_0^*=0.73$	最终阴性	最终阳性

如果我们仅仅使用核酸检验决定最终阳性和最终阴性, 当核算显示为阴性时我们诊断为未感染. 同时使用核酸和 CT 两者时, 如果根据 b^* , 当核酸显示阴性而 CT 显示阳性时, 我们选择最终阳性 (见表 11 中蓝色), 因为 $b^*(CT+) = 0.69$ 大于 $b^*(NAT-) = 0.47$. 这样诊断和武汉改进的诊断就一样.

作者假定阳性的先验概率是 $P(e_1) = 0.25$, 用各种确证测度得到的确证度(用同样的敏感性和特异性)如表 12 所示.

表 12. 用各种确证测度评价核酸检验和 CT 的结果.

	D	M	Z	S	C	N	F	b^*	c^*
$c(NAT-)$	.10	.11	.40	.62	.08	.45	.31	.47	.83
$c(NAT+)$	.52	.34	.69	.62	.08	.45	.82	.90	.70
$c(CT-)$	.17	.14	.67	.43	.10	.55	.58	.73	.91
$c(CT+)$	.27	.41	.36	.43	.10	.55	.52	.69	.06
$c(CT+) > c(NAT-)$			No	No					No
$c(NAT+) > c(CT-)$					No	No			No

如果有“No”在一个测度下面, 则使用该测度会导致和武汉改进诊断不同的诊断. 红色意味着 $c(CT+) < c(NAT-)$ 或 $c(NAT+) < c(CT-)$, 是不合理的. 测度 D, M, F 和 b^* 一样, 和改进的诊断一致. 如果 $P(h_1)$ 从 0.1 到 0.6 变化, 我们将发现, 测度 M 也和武汉改进的诊断不一致. 如果检验阳性或阴性的确证度大于 0.2 才可信, 则 D 也不理想, 只有测度 F 和 b^* 满足我们要求.

表 10 中敏感性和特异性并不是特意选出来的. 如果核酸敏感性在 0.3 和 0.7 之间变化, CT 敏感性在 0.6 和 0.9 之间变化, 结论同样是: 只有 D, F 和 b^* 和武汉改进的诊断一致.

测度 c^* 也不合适用于诊断, 因为它反应正确率而不反应漏保比例, 而感染漏保比误报损失更大.

4.3 Δa 和 Δd 怎样影响各种确证测度

下面例子用来检查我们是否可以用流行的确证测度解释一只黑乌鸦能比一支白粉笔更好支持“乌鸦是黑的”. 表 13 显示的是用 9 个确证测度算出的确证度. 首先我们用 $a=d=20, b=c=10$ 算出九个确证度; 然后用 $a+1$ 代替 a 算出九个确证度; 再用 $d+1$ 代替 d 算出九个确证度.

表 13. 各种确证测度怎样受 $\Delta a=1$ 和 $\Delta d=1$ 影响.

	$f(a, b, c, d)$	$a=d=20$ $b=c=10$	$\Delta a=1$ $\Delta d=0$	$\Delta d=1$ $\Delta a=0$	$\Delta f/\Delta a - \Delta f/\Delta d$
$D(e_1 \rightarrow h_1)$	$a/(a+c) - (a+b)/n$	0.167	0.169	0.175	-0.006
$M(e_1 \rightarrow h_1)$	$a/(a+b) - (a+c)/n$	0.167	0.169	0.175	-0.006
$C(e_1 \rightarrow h_1)$	$a/n - (a+c)(a+b)/n^2$	0.083	0.086	0.086	0
$Z(e_1 \rightarrow h_1)$	$D(e_1 \rightarrow h_1)/[(c+d)/n]$	0.333	0.344	0.344	0
$S(e_1 \rightarrow h_1)$	$a/(a+c) - b/(b+d)$	0.333	0.334	0.344	0
$N(e_1 \rightarrow h_1)$	$a/(a+b) - c/(c+d)$	0.333	0.334	0.344	0
$F(e_1 \rightarrow h_1)$	$(ad-bc)/(ad+bc+2ac)$	0.333	0.340	0.348	-0.007
LR^+	$[a/(a+b)]/[c/(c+d)]$	2	2.03	2.07	-0.034
$c^*(e_1 \rightarrow h_1)$	$(a-c)/\max(a, c)$	0.5	0.524	0.5	0.024 > 0

结果一定超出大多数研究者预料. 表 13 表明, 所有测度除了 c^* (见蓝色数字) 不能保证 $\Delta a=1$ 比 $\Delta d=1$ 更能增大 $f(a, b, c, d)$. 如果我们让 b 和 c 在 1 和 19 之间变化, 除了 c^*, S, N , 其他测度都不能保证 $\Delta f/\Delta a \geq \Delta f/\Delta d$. 在 $b > c$ 时, 测度 S 和 N 也不能保证 $\Delta f/\Delta a \geq \Delta f/\Delta d$. 对于 D 和 M , 原因是 $\Delta d=1$ 减小 $P(h_1)$ 和 $P(e_1)$ 超过增加 $P(h_1|e_1)$ 和 $P(e_1|h_1)$. 对于除 c^* 外的其他测度, 原因类似.

5. 讨论

5.1 澄清乌鸦悖论

为了消除乌鸦悖论，很多研究者——包括 Hemple [3]——肯定 EC 而否定 Nicod-Fisher 准则（NFC）；也有很多研究者——比如 Scheffler 和 Goodman [35]——否定 EC 而肯定 NFC。还有其他一些人，不完全肯定或否定 EC，或不完全肯定 NFC。

首先我们看是否能使用 F 测度消除乌鸦悖论。 $F(e_1 \rightarrow h_1)$ 和 $F(h_0 \rightarrow e_0)$ 的不同之处在于，两者反例相同，但是正例不同。当 d 增大1时， $F(e_1 \rightarrow h_1)$ 和 $F(h_0 \rightarrow e_0)$ 都会增加，但是增加幅度不同。所以：

- F 测度虽然否定 EC，但是还是肯定 $F(e_1 \rightarrow h_1)$ 和 $F(h_0 \rightarrow e_0)$ 正相关；
- F 测度不符合 Nicod-Fisher 准则。

结论是，测度 F 不能消除我们关于乌鸦悖论的困惑。

Greco[15]等人从粗糙集理论角度得出结论：NFC是对的。但是他们分析很多确证测度后，没有发现一个合适的且符合NFC的确证测度。也有很多研究者相信：NFC不对，它符合直觉只是因为 $c(e_1 \rightarrow h_1)$ 随 a 增加较大，随 d 增加较小。Fitelson 和 Hawthorne [28]在比较不同测度后认为似然比 LR 可以用来说明 a 支持 $e_1 \rightarrow h_1$ 超过 d 。

不幸的是，表 13 显示，除 c^* 以外的所有确证测度由 $\Delta d = 1$ 引起的增量超过或等同于由 $\Delta a = 1$ 引起的增量，这意味着这些测度支持结论：一支白粉笔对“乌鸦是黑的”支持超过或等于一只黑乌鸦。因此这些测度不能用来澄清乌鸦悖论。

然而测度 c^* 不同，因为 $c^*(e_1 \rightarrow h_1) = (a - c)/(a \vee c)$ 且 $c^*(h_0 \rightarrow e_0) = (d - c)/(d \vee c)$ ，等价条件不成立，测度 c^* 完全符合 NFC。所以根据测度 c^* ，乌鸦悖论不再存在。

5.2 关于增量确证和绝对确证

在表 13 中，如果初始数字 $a=d=200$ 和 $c=d=100$ ， $\Delta a = 1$ 引起的不同测度的增量会很小。比如， $D(e_1 \rightarrow h_1)$ 由 0.1667 增加到 0.1669； $c^*(e_1 \rightarrow h_1)$ 由 0.5 增加到 0.5025。增幅是表 13 中增幅的约 1/10。因此，由新样例带来的确证度增量和原有样例个数或先验知识密切相关。

绝对确证要求：

- 样本足够大，即 n 足够大；
- 每个样例是独立产生的；
- 样例具有代表性。

如果 a 、 b 、 c 、 d 都很小，所得确证度是不可靠的。我们需要用确证度区间代替确证度，比如用 $[0.5, 1]$ 代替 1。具体做法有待进一步探讨。

5.3 假设对称(HS)还是后件对称是需要的？

Elles 和 Fitelson 提出的 HS 对称性其实是大前提的后件对称性，即 $c(\text{前件}, \text{后件}) = -c(\text{前件}, \neg \text{后件})$ ；而 ES 对称性是前件对称性。因为在他们眼里， e 和 h 不是大前提的前后件，所以他们不能说后件或前件对称。考虑 $c(e, h)$ 互换为 $c(h, e)$ ，如果按 HS 的字面理解，就会出现误解(参看表 14)。

表 14. 对假设对称性 HS 和证据对称性 ES 的误解

	HS 对称即后件对称性	ES 对称即前件对称性
误解的 HS 对称性	$c(e, h) = -c(e, \neg h)$	$c(h, e) = -c(\neg h, e)$
误解的 ES 对称性	$c(h, e) = -c(h, \neg e)$	$c(e, h) = -c(\neg e, h)$

文献[8,19]就出现了这种误解，其中称 $c(h, e) = -c(h, \neg e)$ 是证据对称(ES)。而实际上是 HS 或后件对称。文献[19]中， $F(H, E)$ (左边是假设，右边是证据)应符合 HS 对称性(这是对的)，

而 $F(E, H)$ 应符合 ES 对称性(这一说法不符合 ES 原意)。实际上两者都符合 HS 对称性, 即后件对称性。误解的更严重后果是, 文献[18]认为 ES 及 EHS(即 $c(H, E)=c(\neg H, \neg E)$)和 HS 一样是有价值的, 从而选出 S 、 N 和 C 等测度。

笔者赞成 Elles 和 Fitelson 的结论, 只有 HS 对称性(即后件对称性)是需要的。因此, 我们有必要明确 $c(e, h)$ 中前后件关系, 把 $c(e, h)$ 写成 $c(e \rightarrow h)$, 把假设对称性 HS 改为后件对称性, 把证据对称性 ES 改为前件对称性。

4.7 关于贝叶斯确证和似然确证

Carnap 提出的 D 测度常被称之为标准贝叶斯确证测度。但是上面分析表明, D 适合用作选择假设的测度, 不适合作为确证大前提的测度。Carnap 开辟了贝叶斯确证方向。但是他对 D 的解释容易让人混淆支持大前提的证据(样本)和结论的证据(小前提)。

Greco 等人[19]把条件概率 $P(h|e)$ 构成的测度叫做贝叶斯确证度, 把 $P(e|h)$ 构成的测度叫做似然确证度, 把 $c(h \rightarrow e)$ 这样的确证测度称为反贝叶斯或反似然确证测度。这样划分很有启发。然而本文对确证测度的分类不取决于符号, 而是取决于方法——把真值函数中有效部分比例优化后当做信道确证度 b^* ——它类似于似然比, 反映信道有多好; 把似然函数中有效部分比例优化后当做预测确证度 c^* ——它类似于正确率, 反映概率预测有多好。 b^* 可谓逻辑贝叶斯确证测度(因为使用的是逻辑贝叶斯推理[17])——虽然其中可能使用 $P(e|h)$; c^* 可谓似然确证测度——虽然其中可能使用 $P(h|e)$ 。

本文也提供了反信道确证测度 $b^*(h \rightarrow e)$ 和反预测确证测度 $c^*(h \rightarrow e)$ 。确证测度 $b^*(e \rightarrow h)$ 和 $c^*(e \rightarrow h)$ 和错报率相关, 而反确证测度 $b^*(h \rightarrow e)$ 和 $c^*(h \rightarrow e)$ 和漏报率相关。

4.8. 关于确定性因子用于概率专家系统

确定性因子(CF 即 Certainty Factor) 是 Shortliffe 和 Buchanan 提出用于专家系统 MYCIN [7]的。CF 用来表示推理 “If e then h ” 有多好, CF 和测度 Z 的关系是 $CF(h \rightarrow e) = Z(e \rightarrow h)$ [36]。

如 Heckerman 和 Shortliffe [32]指出, CF 模型被广泛用于基于规则的专家系统, 但是它也有理论和实践局限性; 信念网表示可以克服其中大多数局限性, 而 CF 模型更加简单; 结合两者有可能发展出更简单的概率专家系统。

测度 $b^*(e \rightarrow h)$ 和谓词 $e_1(h)$ 的真值函数的可信部分相关, 含义和 CF 类似。差别是 $b^*(e \rightarrow h)$ 独立于 $P(h)$, 而 CF 和 $P(h)$ 相关; b^* 兼容统计概率理论, 而 CF 并不兼容。是否可以把 b^* 或 c^* 当做确定性因子, 从而简化信念网或概率专家系统? 这个问题值得探讨。

4.9 确证测度 F , b^* , 和 c^* 如何兼容 Popper 的证伪思想

Popper 肯定一个反例就可以证伪一个全称假设或一个大前提。但是对于不确定大前提, 反例如何影响其确证度? 确证测度 F , b^* 和 c^* 就能体现反例的重要性。

表9含有两个例子中, 例1中正例比例少, 反例比例更少, 所以确证度大。这个例子说明, 要提高确证度, 增大后验概率 $P(e_1|h_1)$ (对于 b^*) 或 $P(h_1|e_1)$ (对于 c^*) 不是必要的。例2中虽然正例比例大, 反例比例也不小, 所以确证度 b^* 很小。这个例子说明, 要提高确证度, 增大后验概率也不是充分的。只有增加正反例比例的相对差才是必要且充分的。

Popper 肯定一个反例就可以证伪一个全称假设——可以解释为, 对于全称假设的证伪, 没有反例是重要的。现在对于不确定假设的确证, 这话就变成: 较少反例是重要的。可见这些确证测度和 Popper 证伪思想兼容。

Scheffler 和 Goodman [35] 根据 Popper 的证伪思想, 提出 “选择确证”, 认为黑乌鸦之所以支持 “乌鸦 $\rightarrow$ 黑的”, 是因为它否定(在一定程度上证伪) “乌鸦 $\rightarrow$ 不黑”, 从而支持相反假设 “乌鸦 $\rightarrow$ 黑的”。而不黑的乌鸦之所以支持 “乌鸦 $\rightarrow$ 不黑”, 是因为它否定相反假设 “乌鸦 $\rightarrow$ 黑的”。

他们的解释是很有意义的. 但是他们没有提出相应的确证测度。 $c^*(e_1 \rightarrow h_1)$ 正好体现他们的解释。

5. 结论

用语义信息方法和统计学习方法, 以医学检验为例, 我们可以推导出一个确证测度 $b^*(e \rightarrow h)$, 它和 Kemeny 和 Oppenheim 很早提出的 F 测度类似, b^* 和 F 能像似然比 LR 那样反映医学检验的信道特性, 表明检验手段有多好, 所以 b^* 被称之为信道确证测度。为了得到反映概率预测有多好, 我们还可以推导出预测确证测度 $c^*(e \rightarrow h)$ ——它类似于正确率, 但是也在 -1 和 1 之间变化。 b^* 和 c^* 都可以用于概率预测, 而且 b^* 适合有病的先验概率变化时根据阳性或阴性对有病或无病做概率预测。 b^* 和 c^* 可以写成四种样例个数 a, b, c, d 的函数 $f(a, b, c, d)$ 。互换前后件, 我们还可以得到反信道确证测度 $b^*(h \rightarrow e)$ 和反预测确证测度 $c^*(h \rightarrow e)$ 。考虑后件对称性, b^* 和 c^* 各有 8 种形式, 都符合 Elles 和 Fitelson 提出的对称性标准, Greco 等人提出的单调标准, 并且在 -1 和 1 之间变化。

本文显示, 大多数确证测度不能用来帮助我们诊断新冠肺炎, 但是测度 F 和 b^* 等测度(它们是似然比的函数)可以。文中证明流行的确证测度都不能支持结论: 一只黑乌鸦比一支白粉笔更加支持“乌鸦是黑的”。而 c^* 支持, 并且 c^* 完全否定等价条件并确切符合 Nicod 准则, 因此可以用来消除乌鸦悖论。新的确证测度表明: 反例少比正例多更重要; 所以新的确证测度兼容 Popper 的证伪思想。

新的确证测度支持绝对确证而不是增量确证。在样本较小时, 用任何确证测度算出的确证度都是不可靠的, 应当用确证度区间代替确证度——具体做法有待进一步研究。用新的确证测度作为确定性因子或许可以简化概率专家系统或信念网, 这一问题值得进一步研究。

补充材料: 用于计算表 9, 12, 13 中数据的 Excel 文件下载地址:

<http://survivor99.com/lcg/Table9-12-13NAT.zip>。我们可以通过改变 a, b, c, d 检查不同确证测度。

基金资助: 自我。

感谢: 作者感谢复旦大学张志林教授和长沙大学周健勇教授, 因为本研究得益于和他们的交流。作者感谢辽宁工程技术大学的汪培庄教授, 因为他的长期支持和鼓励。作者也感谢期刊 Entropy 的匿名审稿人, 因为他们的建议大大改进了本文。

利益冲突: 没有。

参考文献:

1. Carnap, R. *Logical Foundations of Probability*, 2nd ed.; University of Chicago Press: Chicago, IL, USA, 1962.
2. Popper, K. *Conjectures and Refutations*, 1st ed.; Routledge: London, UK; New York, NY, USA, 2002.
3. Hempel, C.G. Studies in the Logic of Confirmation. *Mind* **1945**, *54*, 1–26, 97–121.
4. Nicod, J. *Le Problème Logique De L'induction*; Alcan: Paris, France, 1924; p. 219. (Engl. Transl. The logical problem of induction. In *Foundations of Geometry and Induction*; Routledge: London, UK, 2000.)
5. Mortimer, H. *The Logic of Induction*; Prentice Hall: Paramus, NJ, USA, 1988.
6. Horwich, P. *Probability and Evidence*; Cambridge University Press: Cambridge, UK, 1982.
7. Shortliffe, E.H.; Buchanan, B.G. A model of inexact reasoning in medicine. *Math. Biosci.* **1975**, *23*, 351–379.

8. Crupi, V.; Tentori, K.; Gonzalez, M. On Bayesian measures of evidential support: Theoretical and empirical issues. *Philos. Sci.* **2007**, *74*, 229–252.
9. Christensen, D. Measuring confirmation. *J. Philos.* **1999**, *96*, 437–461.
10. Nozick, R. *Philosophical Explanations*; Clarendon: Oxford, UK, 1981.
11. Good, I.J. The best explicatum for weight of evidence. *J. Stat. Comput. Simul.* **1984**, *19*, 294–299.
12. Kemeny, J.; Oppenheim, P. Degrees of factual support. *Philos. Sci.* **1952**, *19*, 307–324.
13. Fitelson, B. *Studies in Bayesian Confirmation Theory*. Ph.D. Thesis, University of Wisconsin: Madison, WI, USA, 2001.
14. Eells, E.; Fitelson, B. Symmetries and asymmetries in evidential support. *Philos. Stud.* **2002**, *107*, 129–142.
15. Greco, S.; Slowiński, R.; Szczęch, I. Properties of rule interestingness measures and alternative approaches to normalization of measures. *Inf. Sci.* **2012**, *216*, 1–16.
16. Greco, S.; Pawlak, Z.; Slowiński, R. Can Bayesian confirmation measures be useful for rough set decision rules?. *Eng. Appl. Artif. Intell.* **2004**, *17*, 345–361.
17. Lu, C. Semantic information G theory and Logical Bayesian Inference for machine learning. *Information* **2019**, *10*, 261.
18. Sensitivity and specificity. Wikipedia the Free Encyclopedia. https://en.wikipedia.org/wiki/Sensitivity_and_specificity (accessed on 27 February 2020).
19. Greco, S.; Slowiński, R.; Szczęch, I. Measures of rule interestingness in various perspectives of confirmation. *Inf. Sci.* **2016**, *346–347*, 216–235.
20. Lu, C. A generalization of Shannon's information theory. *Int. J. Gen. Syst.* **1999**, *28*, 453–490.
21. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–429, 623–656.
22. Tarski, A. The semantic conception of truth and the foundations of semantics. *Philos. Phenomenol. Res.* **1994**, *4*, 341–376.
23. Davidson, D. Truth and meaning. *Synthese* **1967**, *17*, 304–323.
24. Tentori, K.; Crupi, V.; Bonini, N.; Osherson, D. Comparison of confirmation measures. *Cognition* **2007**, *103*, 107–119.
25. Glass, D.H. Entailment and symmetry in confirmation measures of interestingness. *Inf. Sci.* **2014**, *279*, 552–559.
26. Susmaga, R.; Szczęch, I. Selected group-theoretic aspects of confirmation measure symmetries. *Inf. Sci.* **2016**, *346–347*, 424–441.
27. Thornbury, I.R.; Fryback, D.G.; Edwards, W. Likelihood ratios as a measure of the diagnostic usefulness of excretory urogram information. *Radiology* **1975**, *114*, 561–565.
28. Fitelson, B.; Hawthorne, J. How Bayesian confirmation theory handles the paradox of the ravens. In *The Place of Probability in Science*; Eells, E., Fetzer, J., Eds.; Springer: Dordrecht, German, 2010; pp. 247–276.
29. Huber, F. What Is the Point of Confirmation? *Philos. Sci.* **2005**, *72*, 1146–1159.
30. Carnap, R.; Bar-Hillel, Y. *An Outline of a Theory of Semantic Information*; Technical Report No. 247; Research Lab. of Electronics, MIT: Cambridge, MA, USA, 1952.
31. Crupi, V.; Tentori, K. State of the field: Measuring information and confirmation. *Stud. Hist. Philos. Sci.* **2014**, *47*, 81–90.
32. Lu, C. Semantic channel and Shannon channel mutually match and iterate for tests and estimations with maximum mutual information and maximum likelihood. In *Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing*, Shanghai, China, 15 January 2018; IEEE Computer Society Press Room: Washington, DC, USA, 2018; pp. 15–18.
33. Available online: <http://news.cctv.com/2020/02/13/ARTIHIHFAHyTYO6NEovYRMNh200213.shtml> (accessed on 13 February 2020).
34. Wang, S.; Kang, B.; Ma, J.; Zeng, X.; Xiao, M.; Guo, J.; Cai, M.; Yang, J.; Li, Y.; Meng, X.; et al. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *medRxiv* **2020**, doi:10.1101/2020.02.14.20023028.
35. Scheffler, I.; Goodman, N.J. Selective confirmation and the ravens: A reply to Foster. *J. Philos.* **1972**, *69*, 78–83.

-
36. Heckerman, D.E.; Shortliffe, E.H. From certainty factors to belief networks. *Artif. Intell. Med.* **1992**, 4, 35–52.



© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

国内有关研究：

37. 顿新国. 归纳悖论研究[M], 北京：人民出版社，2012.
38. 任晓明,陈晓平等. 决策、博弈与认知——归纳落实的理论与应用[M]. 北京：北京师范大学出版集团，2014.
39. 熊立文.现代归纳逻辑的发展[M]，人民出版社，2004.
40. 周文华.一个新的解决乌鸦悖论的方案[J]，自然辩证法，2015, 37(5):26-32.

作者主页：<http://www.survivor99.com/>

关于语义信息论和机器学习研究：<http://www.survivor99.com/lcg/books/GIT/>