·统计、测量与方法·

一种简单有效的 Q 矩阵估计方法开发：基于非参数化方法视角*

汪大勋 高旭亮 韩雨婷 涂冬波**
（江西师范大学心理学院，南昌，330022）

摘 要 相对于参数化的 Q 矩阵估计方法，本研究将海明距离（Hamming Distance，HD）用于 Q 矩阵估计，开发出一种简单有效的非参数化的 Q 矩阵估计方法。模拟和实证研究发现（1）基于海明距离的 Q 矩阵估计法具有较高的估计正确率。（2）该方法简单易懂，运算时间短，是一种简单而有效的 Q 矩阵估计方法。（3）新方法对于 Tatsuka（1990）分数减法测验的 Q 矩阵的估计准确率尚可，说明新方法在实测数据中也具有较好的效果。

关键词 认知诊断 Q 矩阵 海明距离 DINA 模型

1 引言

作为新一代测量理论的核心，认知诊断能报告被试在测验领域上知识的掌握情况，受到越来越多人的关注。在认知诊断中将被试完成测验所需要的潜在特质，包括知识、技能、策略等称作属性（Leighton，Gierl，& Hunka，2004）。属性和项目之间的关系用 Q 矩阵来表述。研究发现 Q 矩阵错误会增大参数估计误差和降低被试诊断正确率（涂冬波 , 蔡艳 , 戴海崎，2012；de la Torre, 2009; Rupp & Templi, 2008）。在国内，丁树良等人对 Q 矩阵理论进行了深入的研究（丁树良 , 罗芬 , 汪文义 , 2012；丁树良 , 汪文义 , 罗芬 , 2012；丁树良 , 杨淑群 , 汪文义 , 2010）。而在实际中，Q 矩阵往往是通过专家来界定，这容易受到专家主观因素的影响。并且不同专家界定的 Q 矩阵可能还不尽相同。例如分数减法（Tatsuoka,1990）测验的 Q 矩阵界定，到现在还存在着争议。

除了专家界定属性以外，研究者尝试在已知作答数据基础上进行 Q 矩阵界定或修正，思路是在已有作答数据和部分已知 Q 矩阵题目的基础上，通过各种指标来为新题确定测量模式，由此来进行 Q 矩阵估计和修正。在对 Q 矩阵进行估计上，汪文义、丁树良和游晓锋（2011）以及陈平和辛涛（2011）分别对新题的题目属性标定进行过研究。喻晓锋等人（2015）提出使用似然比 D^2 统计量来进行题目属性定义。以上方法均需要进行复杂的参数估计，计算复杂且耗时。在 Q 矩阵修正方面，研究者提出了 δ 法（de la Torre，2008），γ 法（涂冬波等，2012），*RSS* 法（Chiu，2013）等。δ 法和 γ 法同样需要进行参数估计，相对于参数化的方法，*RSS* 法通过计算理想作答与实际作答之间的距离来进行Q矩阵修正,计算相对简单,并且修正效果较好，但该方法目前还无法实现对 Q 矩阵的估计。

目前国内外学者开发的 Q 矩阵估计方法都是基于复杂的参数化方法，都需要对被试参数和项目参数进行估计，包含了大量的数学运算，并且运算时间很长；同时这类方法的参数估计还会受到数据和模型不匹配的影响。因此开发出一种计算简单，效果稳定的 Q 矩阵估计方法对认知诊断 Q 矩阵的界定有重要意义。本研究借鉴 *RSS* 法的思想，用海明距离进行 Q 矩阵估计，开发出一种非参数化的 Q 矩阵估计方法。并通过 Monte Carlo 模拟研究和实测数据研究，探查该方法进行 Q 矩阵估计时的效果。

2 基于海明距离的 Q 矩阵估计方法开发

* 本研究得到国家自然科学基金（31660278，31300876，31100756）、江西省高校人文社科项目（XL1507，XL1508）和武汉市卫计委支撑课题（WG16C08）的资助。

** 通讯作者：涂冬波。E-mail: tudongbo@aliyun.com

DOI:10.16719/j.cnki.1671-6981.20180127

2.1 海明距离（HD）用于估计被试掌握模式

罗照盛、李喻骏、喻晓锋、高椿雷和彭亚风（2015）对海明距离估计被试掌握模式进行了研究，并根据不同的判别方式分为 R 法和 B 法。海明距离定义如下：

$$HD(Y_i, I_t) \stackrel{\Delta}{=} \sum_{j=1}^{J} HD(Y_{ij}, I_j^{(t)}) \quad (1)$$

其中

$$HD(Y_{ij}, I_j^{(t)}) \stackrel{\Delta}{=} \left| Y_{ij} - I_j^{(t)} \right| \quad (2)$$

式中$HD(Y_{ij}, I_j^{(t)})$表示项目j上被试i的观察反应Y_{ij}与项目j上第t种理想反应$I_j^{(t)}$的海明距离。$HD(Y_i, I_t)$表示所有项目的海明距离之和。

海明距离法对被试属性掌握模式的估计步骤如下：

（1）构建所有理想掌握模式（IMP）在所有题目上的理想反应矩阵。

（2）计算被试i的观察反应模式（ORP）与每种掌握模式的理想反应模式（IRP）之间的海明距离，将最小海明距离所在的掌握模式作为被试i的掌握模式。

（3）对无法判别的被试使用 R 方法或 B 方法进行判别（罗照盛等，2015）。

2.2 基于海明距离的 Q 矩阵估计方法

在进行 Q 矩阵估计时，需要部分已经界定好 Q 矩阵的题目作为基础，这部分题目的 Q 矩阵为 Q_base，未界定 Q 矩阵的题目称为新题。定义H为所有可能的测量模式集合，若属性个数为k个且相互独立，则H中共有 2^k-1 种测量模式。

在进行 Q 矩阵估计时，首先根据已知 Q 矩阵的题目使用海明距离来估计被试掌握模式，这里采用 R 方法，即当某被试在两种掌握模式上的海明距离都最小时，采用随机的方式进行被试分类，由此得到所有被试的掌握模式。然后根据估计的被试掌握模式构造所有被试在所有测量模式上的理想作答矩阵。再分别计算新题j的观察得分向量与每种测量模式的理想作答向量之间的距离。最后将海明距离最小的测量模式作为新题j的测量模式。

新题j的观察反应模式与第t种测量模式的理想作答之间的海明距离为：

$$HD(Y_j, I_t) = \sum_{i=1}^{N} \left| Y_{ij} - \eta_{it} \right| \quad (3)$$

η_{it}为被试i在第t种测量模式上的理想作答，Y_{ij}为被试i在项目j上的实际得分即观察作答。上式表示所有被试在项目j的实际得分与所有被试在第t种测量模式的理想作答之间的距离。

新题j的测量模式为：

$$Q_j = \min_{H_t}\left(HD(Y_j, I_t) = \sum_{i=1}^{N} \left| Y_{ij} - \eta_{it} \right|, t = 1, 2, \cdots 2^k - 1 \right) \quad (4)$$

即在所有测量模式中选择海明距离最小的测量模式作为新题j的测量模式。

海明距离进行 Q 矩阵估计的步骤如下：

第一步：估计新题的测量模式

（1）根据已知的部分项目 Q 矩阵即 Q_base，使用海明距离估计所有被试的掌握模式。

（2）根据步骤（1）估计的被试掌握模式，然后构建所有被试在所有测量模式（2^k-1 种）上的理想得分矩阵。

（3）计算新题j的实际得分与每一种测量模式的理想得分之间的海明距离。

（4）将海明距离最小的测量模式作为新题j的测量模式。

（5）将新题j加入到 Q_base 中，作为界定下一个新题的基础题。

重复步骤（1）-（5），由此界定所有新题的测量模式。

第二步：对新定义的 Q 矩阵进行循环修正。

记第一步得到的 Q 矩阵为$\hat{Q}_0$，这个$\hat{Q}_0$可能包含一些错误。

（1）从$\hat{Q}_0$中依次挑选一个新定义 Q 矩阵的题目j，以其余题目的 Q 矩阵作为基础，估计被试的掌握模式。

（2）根据步骤（1）估计的被试掌握模式，并构建所有被试在所有测量模式（2^k-1 种）上的理想得分矩阵。

（3）计算题目j的实际得分与每一种测量模式的理想得分之间的海明距离。

（4）对于题目j，寻找海明距离最小的测量模式，看其是否与当前题目j的测量模式相同，如果不同，则将海明距离最小的测量模式赋予题目j。

重复步骤（1）-（4），将所有新定义 Q 矩阵的题目循环一次，得到$\hat{Q}_1$，则完成一次迭代。当迭代次数达到 20 次或两次迭代间 Q 矩阵相同，则停止循环。

算法结束。

3 基于海明距离的 Q 矩阵估计方法示例

为了便于理解，这里给出一个示例展示基于海明距离的Q矩阵估计方法的估计过程。假设有30个题目，测量属性为3个，已知前4个题目的测量属性（即Q_base）如下，

$$Q_{base} = \begin{bmatrix} 100 \\ 010 \\ 001 \\ 110 \end{bmatrix}$$

图1 示例中已知部分题目的Q矩阵

现需根据这4题已知的测验Q矩阵来估计剩余30-4 = 26题的测验Q矩阵。此处只示例第5题的估计过程,其余题类推。其中括号内为题目测量模式。数据格式如表1：

具体估计步骤如下：

第一步：逐题估计每题的Q矩阵（从第5题开始估计）

（1）根据被试在已知Q矩阵的4个题目上的作答数据，估计被试的掌握模式。首先构建所有理想掌握模式（IMP）在4个题目上的理想作答矩阵如下：

表1 *N*名被试在前5题的观察作答数据

	第1题（100）	第2题（010）	第3题（001）	第4题（110）	第5题 ?
被试1	1	1	0	1	1
被试2	1	0	0	1	1
被试3	1	0	0	1	0
被试4	0	1	0	0	1
被试5	1	1	1	1	1
…	…	…	…	…	…
被试N	1	1	1	1	1

表2 所有掌握模式在4个题目上的理想作答矩阵

	第1题（100）	第2题（010）	第3题（001）	第4题（110）
IMP_1（000）	0	0	0	0
IMP_2（100）	1	0	0	0
IMP_3（010）	0	1	0	0
IMP_4（001）	0	0	1	0
IMP_5（110）	1	1	0	1
IMP_6（101）	1	0	1	0
IMP_7（011）	0	1	1	0
IMP_8（111）	1	1	1	1

计算被试1的观察得分与每种掌握模式在这4个题目上的理想得分之间的海明距离，本例中被试1的观察得分与IMP_1-IMP_8的海明距离分别为（3、2、2、4、0、3、3、1），因此被试1的掌握模式为IMP_5（110）。其余被试同理，由此估计出前5位被试的掌握模式分别为：（110）（100）（100）（010）（111），限于篇幅这里未显示所有被试估计的属性掌握模式。

（2）根据步骤（1）估计的被试掌握模式，构建被试在所有测量模式的理想作答矩阵。如下：

根据公式（3）分别计算第5题的观察作答向量与表3每种测量模式的理想作答之间的海明距离。

表3 被试在所有测量模式上的理想作答矩阵

	H_1（100）	H_2（010）	H_3（001）	H_4（110）	H_5（101）	H_6（011）	H_7（111）
被试1（110）	1	1	0	1	0	0	0
被试2（100）	1	0	0	0	0	0	0
被试3（100）	1	0	0	0	0	0	0
被试4（010）	0	1	0	0	0	0	0
被试5（111）	1	1	1	1	1	1	1
…	…	…	…	…	…	…	…
被试*N*（111）	1	1	1	1	1	1	1

示例中第 5 题的观察作答向量分别与 H_1—H_7 之间的海明距离为 2、1、3、2、3、3、3。因此判定第 5 题的测量模式为 H_2（010）。

（3）重复步骤（1）-（3）定义其它待估计题目的 Q 矩阵。

第二步：对所有新估计项目的 Q 矩阵进行修正。

第二步的具体方法（见 2.2 部分）与第一步的方法相似，区别在于第二步要考察题目 j 的测量模式时，是将其余所有题目作为基础题。

算法结束。

4 研究 1： 基于海明距离的 Q 矩阵估计法效果验证

为了验证本研究开发的 Q 矩阵估计方法的科学性及有效性，采用 Mont Carlo 的实验方法，实验设置如下：

4.1 研究设计

4.1.1 Q 矩阵

实验采用 Liu, Xu 和 Ying（2012）的部分 Q 矩阵，采用其中 3 和 5 个属性的 Q 矩阵，增加 7 个属性的 Q 矩阵。其中 3 和 5 个属性的 Q 矩阵题目数量为 20。为了保证新增加的 Q 矩阵对每个属性测量的次数一致，7 个属性的 Q 矩阵题目数量为 21。Q 矩阵如图 2 所示。

4.1.2 题目参数和被试参数模拟

本研究采用 DINA 模型，题目参数和被试参数均采用均匀分布产生。题目参数 s 和 g 的取值区间为 [.05,.25]。被试掌握模式按照均匀分布产生，分别产生 400、500、800、1000 人。

4.1.3 被试作答模拟

根据模拟的被试参数和题目参数分别计算被试 i 在题目 j 上的答对概率 P_{ij}，以 P_{ij} 为概率在贝努力分布（bernoulli distribution）中产生被试 i 在题目 j 上的 0-1 作答反应得分 $response(i, j)$。即，$response(i, j)=Bernoulli(P_{ij})$。

4.1.4 基础题个数

本研究设置基础题个数为 6、8、10、12 个。基础题从真实 Q 矩阵中随机选取。

4.1.5 评价指标

采用成功估计次数（$N_{successful}$）作为评价指标，即随机生成 100 批数据，使用上述方法完全正确估计 Q 矩阵的次数作为评价指标。计算每次估计的 Q 矩阵中所有题目的测量模式与真实 Q 矩阵题目测量模式的一致性作为题目模式判准率（pattern match ratio, PMR），并计算 100 次实验的平均值。根据每次实验估计的 Q 矩阵来估计被试的掌握模式，计算所有被试的平均模式判准率（pattern match ratio, PMR）。

$$
Q_1=\left\{\begin{matrix}
1&0&0\\0&1&0\\0&0&1\\1&0&0\\0&1&0\\0&0&1\\1&0&0\\0&1&0\\0&0&1\\1&1&0\\1&0&1\\1&0&1\\0&1&1\\1&0&1\\0&1&1\\1&1&0\\0&1&1\\1&0&1\\0&1&1\\1&1&1
\end{matrix}\right\}
Q_2=\left\{\begin{matrix}
1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\\1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\\1&1&0&0&0\\1&0&1&0&0\\1&0&0&1&0\\1&0&0&0&1\\0&1&1&0&0\\0&1&0&1&0\\0&1&0&0&1\\0&0&1&1&0\\0&0&1&0&1\\0&0&0&1&1
\end{matrix}\right\}
Q_3=\left\{\begin{matrix}
1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\\1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\\1&1&0&0&0&0&0\\0&0&1&1&0&0&0\\0&0&0&0&1&1&0\\1&0&0&0&0&0&1\\0&1&1&0&0&0&0\\0&0&0&1&1&0&0\\0&0&0&0&0&1&1
\end{matrix}\right\}
$$

图 2 模拟的真实 Q 矩阵（部分引自 Liu 等（2012））

$$N_{\text{successful}}=\sum_{r=1}^{100}\left(n_{r\_correct}\right) \tag{5}$$

$$PMR_{item}=\frac{\sum_{j=1}^{J}n_{j\_correct}}{J} \tag{6}$$

$$PMR_{subject}=\frac{\sum_{i=1}^{N}n_{i\_correct}}{N} \tag{7}$$

式（5）中，$n_{r\_correct}$ 表示第 r 次实验估计的 Q 矩阵与真实 Q 矩阵是否完全一致，完全一致则为 1，否则为 0。式（6）中，J 为题目个数，$n_{j\_correct}$ 为估计的第 j 题 Q 矩阵是否与真实 Q 矩阵中第 j 题一致，完全一致则为 1，否则为 0。式（7）中，N 为被试人数，$n_{i\_correct}$ 表示估计的被试 i 的掌握模式是否与被试 i 的掌握模式一致，如果一致则为 1，否则为 0。

4.2 研究结果

表 4 呈现了基于海明距离的 Q 矩阵估计方法的成功估计次数。表 5 呈现了不同实验条件下的平均题目模式判准率。表 6 是根据估计的 Q 矩阵计算被试的模式判准率。

表 4 的结果显示，基于海明距离的 Q 矩阵估计方法在属性个数为 3 和 5 个时总体上具有很高的成功率，但在属性个数为 7 个时，估计效果很差。随着属性个数的增加，成功率有所下降。分析 3 个 Q 矩阵的估计效果，可以看出当属性个数在 5 个以下时，只要保证基础题个数，海明距离就能有较高的

估计效果。而当属性个数增加到 7 个，该方法完全估计正确 Q 矩阵的可能性不高。这是因为属性个数越多，需要更多的基础题，才能准确估计被试的掌握模式，从而构建正确的理想得分矩阵。从人数上来看，在不同 Q 矩阵和基础题个数上，海明距离的估计效果均相近。如在 Q2 和 Q3 条件下，当基础题为 6 个时，人数从 400 增加到 1000，估计成功率并没有随人数增加的趋势,说明海明距离的估计效果，相对于人数而言，更受基础题个数的影响。

表 5 是从题目层面对估计效果的反应。表 5 的结果和表 4 的结果相对应。从表 5 的结果可以看出，海明距离进行 Q 矩阵估计在 Q1 条件下均能达到 .99 以上的平均题目模式判准率。同样随着属性个数越多，海明距离方法的效果有所降低。随着基础题个数增加，平均题目模式判准率有所增加，当属性个数越多时，这种增加趋势越明显。

表 6 是基于作答分数和估计的 Q 矩阵进行分析得到的平均被试模式判准率，这里使用的方法是海

表 4 基于海明距离的 Q 矩阵估计法 100 次实验成功估计的次数

Q 矩阵	被试人数	基础题个数			
		6	8	10	12
Q1	400	99	99	100	100
	500	97	98	100	100
	800	97	98	99	100
	1000	98	99	100	99
	Mean	97.75	98.5	99.75	99.75
Q2	400	58	79	94	99
	500	44	82	92	96
	800	41	80	92	97
	1000	46	76	95	98
	Mean	47.25	79.25	93.25	97.5
Q3	400	2	18	35	53
	500	0	14	38	56
	800	1	5	35	71
	1000	2	13	41	64
	Mean	1.25	12.5	37.25	61

表 5 基于海明距离的 Q 矩阵估计法的题目模式判准率（PMR_{item}）

Q 矩阵	被试人数	基础题个数			
		6	8	10	12
Q1	400	.998	.998	1	1
	500	.993	.995	1	1
	800	.992	.994	.998	1
	1000	.993	.998	1	.998
	Mean	.994	.996	.999	.999
Q2	400	.858	.950	.985	1
	500	.808	.953	.984	.992
	800	.791	.945	.981	.995
	1000	.815	.936	.988	.995
	Mean	.818	.946	.984	.995
Q3	400	.584	.768	.876	.926
	500	.593	.770	.869	.923
	800	.607	.734	.863	.952
	1000	.597	.762	.877	.936
	Mean	.595	.758	.871	.935

表 6 基于海明距离的 Q 矩阵估计法的被试平均模式判准率（$PMR_{subject}$）

Q 矩阵	被试人数	基础题个数			
		6	8	10	12
Q1	400	.920	.918	.923	.920
	500	.915	.917	.922	.921
	800	.916	.918	.922	.920
	1000	.914	.920	.922	.919
Q2	400	.616	.710	.743	.759
	500	.581	.705	.733	.748
	800	.559	.712	.742	.745
	1000	.576	.690	.745	.746
Q3	400	.171	.301	.419	.472
	500	.170	.311	.395	.460
	800	.169	.260	.390	.492
	1000	.172	.302	.403	479

明距离判别法。对于估计出的 Q 矩阵，如果越接近于真实的 Q 矩阵，模式判准率应该越高。从表 6 中可以发现，Q 矩阵的属性个数越多，模式判准率越低。这是由于 Q 矩阵属性个数越多时，所估计出的 Q 矩阵与真实 Q 矩阵之间差异较大，从而影响了模式判准率。此外模式判准率还受到属性个数的影响，属性个数越

表 7　Tatsuoka 分数减法测验 Q 矩阵（摘自 de la Torre，2008）

题号	题目内容	考核属性				
		A1	A2	A3	A4	A5
1	$\frac{3}{4}-\frac{3}{8}$	1	0	0	0	0
2	$3\frac{1}{2}-2\frac{3}{2}$	1	1	1	1	0
3	$\frac{6}{7}-\frac{4}{7}$	1	0	0	0	0
4	$3-2\frac{1}{5}$	1	1	1	1	1
5	$3\frac{7}{8}-2$	1	0	1	0	0
6	$4\frac{4}{12}-2\frac{7}{12}$	1	1	1	1	0
7	$4\frac{1}{3}-2\frac{4}{3}$	1	1	1	1	0
8	$\frac{11}{8}-\frac{1}{8}$	1	1	0	0	0
9	$3\frac{4}{5}-3\frac{2}{5}$	1	0	1	0	0
10	$2-\frac{1}{3}$	1	0	1	1	1
11	$4\frac{5}{7}-1\frac{4}{7}$	1	0	1	0	0
12	$7\frac{3}{5}-\frac{4}{5}$	1	0	1	1	0
13	$4\frac{1}{10}-2\frac{8}{10}$	1	1	1	1	0
14	$4-1\frac{4}{3}$	1	1	1	1	1
15	$4\frac{1}{3}-1\frac{5}{3}$	1	1	1	1	0

注：A1，分数减法运算基础；A2，化简分数或代分数；A3，从分数中分离出整数；A4，向代分数的整数借位；A5，将整数转换为分数

多，判准率越低。

5 研究 2： 基于海明距离的 Q 矩阵估计法在实测数据中的研究

为进一步研究基于海明距离的 Q 矩阵估计法在实测数据中的效果，本研究采用 Tatsuoka 的分数减法数据，该数据是由 Tatsuoka（1984）收集的，包括 15 个分数减法的题目和 536 个被试作答。共测量了 5 个属性。分数减法的 Q 矩阵如表 7。

为保证选取的基础题是更有把握的，我们选取 s 和 g 平均值最小的题目作为基础题，因为 s 和 g 小说明题目既没有属性多余也没有属性缺失，则可以认为题目的测量模式是正确的。根据已有研究（de la Torre，2008）估计的项目参数，将 s 和 g 平均值从小到大进行排序。分别选取前 6、7、8、9、10 个题目为基础题，以剩余的题目作为待估计的题目。用本研究提出的方法进行估计，计算估计的 Q 矩阵与原有的 Q 矩阵之间的一致性。结果如下，其中括号内为相同元素的个数占总元素个数的比例，本例中总元素为 15 × 5=75 个。

表 8 基于海明距离的 Q 矩阵估计方法对分数减法数据估计的结果

相同元素的个数	基础题个数				
	6	7	8	9	10
	59（.786）	62（.826）	63（.84）	63（.84）	67（.893）

从表 8 可以看出，基于海明距离的 Q 矩阵估计方法在实测数据中有较高的成功率。随着基础题个数增加，估计的成功率更高。由于分数减法的数据只有 15 个题目，所以当选定了基础题之后需要估计的题目数量不多，因此不能充分展示基于海明距离的 Q 矩阵估计方法的效果。

为了比较专家界定的 Q 矩阵和估计的 Q 矩阵的差异，分别计算不同 Q 矩阵下的分类一致性（classification consistency） 信 度（Cui， Gierl, & Chang，2012）、 分 类 准 确 性（classification accuracy）信度（Cui et al.，2012）和模型拟合指标（如负 2 倍的对数似然、AIC 和 BIC 指标）。结果如下。

表 9 使用海明距离法估计 Q 矩阵与专家界定 Q 矩阵的比较

		原始 Q 矩阵	估计的 Q 矩阵（基础题个数）		
			8	9	10
信度	分类一致性	.611	.781	.777	.646
	分类准确性	.721	.839	.840	.692
模型拟合度	-2*log（L）	6911.60	6925.89	6880.96	6923.77
	AIC	7033.60	7047.89	7002.96	7045.77
	BIC	7294.93	7309.22	7264.29	7307.10

注：信度指标，数值越大代表信度越高；模型拟合度指标，数值越大代表拟合越差。

从表 9 可以看出，当基础题个数为 9 个时，用海明距离的方法估计出的 Q 矩阵在信度和模型拟合上均比专家界定的 Q 矩阵要好。因此从实测数据的分析结果来看，基于海明距离的 Q 矩阵方法在实测数据中仍具有一定的合理性。

6 讨论与小结

6.1 讨论

（1）基于海明距离的 Q 矩阵估计法具有较高的成功率，属性个数越少或基础题个数越多成功率更高。

（2）相对于现有的题目属性定义方法，该方法计算简单，运行时间快。如海明距离方法在估计 Q1 下，基础题个数为 8 个，人数为 1000 人时，平均用时不到 30 秒。相对于喻晓锋等人（2015）的方法和 Liu 等人（2012）的算法，节约了大量的时间。

（3）该方法对于 Tatsuoka（1990）分数减法测验的 Q 矩阵的估计准确率尚可，说明该方法在实测数据中也具有较好的效果。

6.2 研究结论

（1）基于海明距离的 Q 矩阵估计方法的优点和局限

相对于参数化的方法来说，该方法不需要进行大量的计算，方法计算简便。此外由于不需要进行参数估计，不会受到数据和模型之间不拟合造成参数估计异常的影响。但是该方法需要有部分已知 Q

矩阵的题目作为基础来估计其他题目的 Q 矩阵，并且需要保证基础题的 Q 矩阵（即 Q_base）是正确的。因此在实际中选择基础题（即 Q_base 中的题）时应该慎重，对于专家们属性标定不是很确定的试题不宜作为 Q_base 试题，我们应当选择测量属性经专家们确认且无争议的题目进入 Q_base，以保证 Q_base 的科学性；同时，也可以辅助其它方法来保证 Q_base 的正确性，如口语报告法（verbal report）、证据中心设计（evidence-centered design, ECD）、效标组法等。

（2） Q 矩阵界定需要参考专家的意见以及各种估计方法结合

在使用客观方法进行 Q 矩阵界定时，建议与专家定义 Q 矩阵结合起来，这样既有利于避免客观方法因为随机误差和方法限制导致的误差，也能避免专家界定 Q 矩阵的主观性。此外不同的 Q 矩阵估计或修正的方法，虽然效果和适用范围不相同，但是都是从不同的角度来考察 Q 矩阵的合理性。因此可以采用多种方法来进行 Q 矩阵估计，将各种方法估计结果不相同的题目提取出来再进行专家讨论。这样既提高了估计的 Q 矩阵的可信度，也减少了专家的工作量。

（3）该方法在 CD-CAT 新题 Q 矩阵标定与估计中的价值

在 CD-CAT 下题库随着时间的推移，需要不断补充新题。对于新补充的题目，需要为其标定Q矩阵。而题库中有大量题目的 Q 矩阵是已知的，因此当获得了被试在新题上的作答数据以后，可以根据本研究提出的方法来为新题标定 Q 矩阵。由此可以实现题库新题的自动标定。

由于篇幅和时间有限，本研究还有许多需要进一步探究的地方，如题目质量水平，属性间有关系时，基础题测量模式的特征等对方法效果的影响等等。总之本方法是 Q 矩阵估计简单而有效的方法，希望能为 Q 矩阵估计及认知诊断的发展发挥一点作用。

参考文献

陈平 , 辛涛 . (2011). 认知诊断计算机化自适应测验中在线标定方法的开发 . *心理学报* , *43(6)*, 710–724.

丁树良 , 罗芬 , 汪文义 . (2012). Q 矩阵理论的扩展 . *心理学探新* , *32(5)*, 417–422.

丁树良 , 汪文义 , 罗芬 . (2012). 认知诊断中 Q 矩阵和 Q 矩阵理论 . *江西师范大学学报 (自然科学版)*, *36(5)*, 441–445.

丁树良 , 杨淑群 , 汪文义 . (2010). 可达矩阵在认知诊断测验编制中的重要作用 . *江西师范大学学报 (自然科学版)*, *34(5)*, 490–494.

罗照盛 , 李喻骏 , 喻晓锋 , 高椿雷 , 彭亚风 . (2015). 一种基于 Q 矩阵理论朴素的认知诊断方法 . *心理学报* , *47(2)*, 264–272.

涂冬波 , 蔡艳 , 戴海崎 . (2012). 基于 DINA 模型的 Q 矩阵修正方法 . *心理学报* , *44(4)*, 558–568.

汪文义 , 丁树良 , 游晓锋 . (2011). 计算机化自适应诊断测验中原始题的属性标定 . *心理学报* , *43(8)*, 964–976.

喻晓锋 , 罗照盛 , 高椿雷 , 李喻骏 , 王睿 , 王钰彤 . (2015). 使用似然比 D2 统计量的题目属性定义方法 . *心理学报* , *47(3)*, 417–426.

Chiu, C. Y. (2013). Statistical refinement of the Q–matrix in cognitive diagnosis. *Applied Psychological Measurement, 37(8)*, 598–618.

Cui, Y., Gierl, M. J., & Chang, H. H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement, 49(1)*, 19–38.

de la Torre, J. (2008). An empirically based method of Q–matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement, 45(4)*, 343–362.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34(1)*, 115–130.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule–space approach. *Journal of Educational Measurement, 41(3)*, 205–237.

Liu, J. C., Xu, G. J., & Ying, Z. L. (2012). Data–driven learning of Q–matrix. *Applied Psychological Measurement, 36(7)*, 548–564.

Rupp, A. A., & Templin, J. L. (2008). The effects of Q–matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement, 68(1)*, 78–96.

Tatsuoka, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems: Computer–based Education Research Laboratory*, University of Illinois.

Tatsuoka, K. K. (1990). Toward an integration of item–response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453–488). New Jersey: Lawrence Erlbaum Associates Publishers.

A Simple and Effective Q-matrix Estimation Method: From Non-Parametric Perspective

Wang Daxun, Gao Xuliang, Han Yuting, Tu Dongbo
(School of Psychology,Jiangxi Normal University,Nanchang,330022)

Abstract As the core of a new generation test theory, cognitive diagnosis(CD) attracts more and more people's attention. Since it can reveal the result from a microscopic perspective, such as individuals' knowledge structures, processing skills and cognitive procedure, it would help us to provide individualized teaching and promote students ' development. Cognitive diagnosis assessments infer the attribute mastery pattern of the respondents by item responses based on the Q-matrix. The Q-matrix plays the role of a bridge between items and respondents. Many studies have shown that misspecification of the Q-matrix can affect the accuracy of model parameters and result in the misclassification of the respondents. In practice, the Q-matrix is established by experts. However, with the application of cognitive diagnosis, more and more researchers found that specification of a Q-matrix was very hard. To avoid the subjectivity from experts in Q-matrix specification and ensure a correct Q-matrix, researchers are trying to look for objective methods. Many researchers have found a number of methods to estimate and validate the Q-matrix. Nevertheless, existing methods need information from parameters and a large amount of computation.

To simplify the method of Q-matrix estimation, this article introduces a new Q-matrix estimation method based on the Hamming Distance(HD) which is simple and non-parametric. The method is as follow: Firstly , we infer the attribute mastery pattern of the respondents by the Hamming Distance. Secondly, we can establish an Expected Response Pattern(ERP) matrix by the relationship between the attribute mastery pattern of all the respondents and each measurement pattern. Finally, the method measures the distance between all respondents' Observed Response Pattern（ORP） and Expected Response Pattern(ERP)in each measurement pattern, and chooses a measurement pattern with the minimum Hamming Distance to items. In this way, we can infer the measurement pattern of items. When there is more than one measurement pattern with the same minimum Hamming Distance, we pick one randomly. In order to explore the effect of the method, we considered different numbers of participants, different numbers of base items and different Q-matrices whose attribute numbers are different.

The Monte Carlo simulation study and real data study showed that: generally, the Hamming Distance method can recover the real Q-matrix with a high rate of success, especially when item attribute is 3 and the number of base items is more than 10. When the attribute is 3, no matter how many base items and participants there are, the rate of success of the method can reach at least 97%. When the number of base items is more than 10, no matter how many participants there are, the rate of success can reach 90% in 3 Q-matrices. Relative to the sample size, the number of base items is more important. Furthermore, the method is easier to understand and needs less computation. The real data study also showed that the Hamming Distance method can estimate the Q-matrix with a high success rate. Compared to the existing methods, the Hamming Distance method is faster and superior. Besides, without the needs of parameters estimation, the method is not affected by the deviation caused by the misfit between model and data. In a word, the Hamming Distance method is simple and effective in Q-matrix Estimation, which is meaningful to the simplification of cognitive diagnosis.

Key words cognitive diagnosis, Q-matrix, Hamming Distance, DINA model