

《心理学报》审稿意见与作者回应

题目：一种高效的 CD-CAT 在线标定新方法：基于熵的信息增益与 EM 视角

作者：谭青蓉，汪大勋，罗芬，蔡艳，涂冬波

第一轮

审稿人 1 意见：

本文针对 Q 矩阵和项目参数的在线标定问题，提出了 Q 矩阵在线标定的非参数标定方法，并与 EM 算法结合使用联合估计 Q 矩阵和项目参数。本研究虽然基于熵的信息增益视角提出的 q 向量确定方法，实质上是基于观察反应最小化熵的信息，即选择使

$$\left(- \sum_{h=1}^H \frac{n_{jh}}{n_j} \left(\sum_{x=0}^1 \frac{n_{jhx}}{n_{jh}} \log \frac{n_{jhx}}{n_{jh}} \right) \right)$$
最小的 q 向量作为项目的 q 向量，使一种非参数的方法，与

似然比 D 方统计量的思路类似，只是后者结合了观察和期望反应分布的熵。新方法能缩短运行时间，提高 q 向量判准率，具有一定得创新性。具体建议如下：

意见 1：文稿标题是否得当？基于熵的信息增益视角仅仅是其中 q 矩阵得标定方法。

回应：非常感谢专家的建议！我们已将题目修改为《一种高效的 CD-CAT 在线标定新方法：基于熵的信息增益与 EM 视角》

意见 2：论文写作格式需要调整，特别是参考文献的引用方法。

回应：非常感谢您的建议！根据您的建议，我们对论文格式进行修改，并仔细核对了参考文献的引用。

意见 3：关于 q 向量的估计方法，文中介绍得过于繁琐。建议提出熵的信息增益公式之后直

接指出其本质，就是最小化 $\left(- \sum_{h=1}^H \frac{n_{jh}}{n_j} \left(\sum_{x=0}^1 \frac{n_{jhx}}{n_{jh}} \log \frac{n_{jhx}}{n_{jh}} \right) \right)$ ，然后举例说明不同 q 向量的商

信息值不同的不同也过于繁琐，事实上稍加思考就知道是这样的。但是为了说明，我建议通

过列表的方式来呈现这一部分的内容。

回应：感谢您的建议！我们已简化了 q 向量估计方法，在提出熵的信息增益公式后，直接指出通过最小化条件熵 $E(R_j|q_j)$ 来估计新题 j 的 q 向量 ($\hat{q}_j = \arg \min_{q_j \in Q_j} (E(R|q_j^*))$)。另外，我们删除了文中的例子，使用列表来说明不同 q 向量下条件熵 $E(R_j|q_j)$ 和熵的信息增益 $g(R_j, q_j)$ 的变化，表格如下：

表1 不同 q 向量下 $E(R_j|q_j)$ 和 $g(R_j, q_j)$ 的计算

q 向量		掌握组	非掌握组	$E(R_j q_j)$	$g(R_j, q_j)$
$q_j^{correct} = [100]$	属性掌握模式	[100] [110] [101] [111]	[000] [101] [001] [011]		
	被试数目	$n_j/2$	$n_j/2$	0	0.690
	正确作答比	1	0		
	错误作答比	0	1		
$q_j^{incorrect} = [011]$	属性掌握模式	[011] [111]	[000] [100] [010] [001] [110] [101]		
	被试数目	$n_j/4$	$3n_j/4$	0.690	0.003
	正确作答比	0.500	0.500		
	错误作答比	0.500	0.500		

意见 4：本文无需再详细介绍 DINA 模型，太多研究太多应用都对其进行了介绍。对于 Q 矩阵和项目参数联合在线标定的方法和步骤建议放在一个部分，本文即详细介绍了基于熵的信息增益公式，还介绍了增益值反应不同 q 向量的合理性，最后还介绍了联合标定的步骤。因此，建议简化和让这三部分的逻辑、思路更清楚。

回应：感谢您的建议，我们从文中删除了对 DINA 模型的详细介绍。此外，我们将 IGEOCM 标定新题 Q 矩阵和项目参数的方法与新题标定步骤合并，重新梳理了该部分的内容，使逻辑思路更为清楚。具体细节请见 3.2 节。

意见 5：建议增加 SIE 方法的介绍。

回应：非常感谢您的建议，我们已在第 2 部分增加了 SIE 方法的介绍。具体如下：

目前，CD-CAT 中同时标定新题 Q 矩阵和项目参数的在线标定方法主要有 JEA 方法(陈平, 辛涛, 2011)和 SIE 方法(Chen et al., 2015)。SIE 方法基于 JEA 方法在决定型输入噪声与门模型(the Deterministic Input, Noisy and Gate Model, DINA; Junker & Sijtsma, 2001)模型下提出，其标定新题时考虑了被试属性掌握模式的估计误差，在标定新题 Q 矩阵和项目参数时充分利用被试的属性掌握模式后验分布。

SIE 方法标定新题时包含了 Q 矩阵标定和项目参数标定的两个部分。对于新题 Q 矩阵的标定，首先基于被试在旧题上的作答计算作答了新题 j 的被试的属性掌握模式后验分布。随后，根据被试属性掌握模式后验分布及每种属性掌握模式在 q 向量为 q_j 的新题 j 上的正确作答概率计算具有某一特定作答 R_{ij} 的被试 i 的后验预测分布：

$$P_i(q_j, \mathbf{g}_j, \mathbf{s}_j) = P(R_{ij} = 1 | q_j, \mathbf{g}_j, \mathbf{s}_j) = \sum_{c=1}^{2^K} \pi_i(\alpha_c) P_j(q_j, \mathbf{g}_j, \mathbf{s}_j, \alpha_c), \quad (1)$$

其中 K 为测验测量的属性个数， $\pi_i(\alpha_c)$ 表示被试 i 的属性掌握模式为 α_c 的概率，其基于被试 i 在旧题上的作答计算， $P_j(q_j, \mathbf{g}_j, \mathbf{s}_j, \alpha_c)$ 表示 DINA 模型下属性掌握模式为 α_c 的被试在项目 j 上的正确作答概率。最后，结合被试后验预测分布及其在新题 j 上的作答 R_{ij} 构建似然并最大化似然函数来估计新题的 q 向量，其表达式如下：

$$\hat{q}_j = \operatorname{argmax}_{q_j \in Q_j} L_j(q_j^*, \mathbf{g}_j, \mathbf{s}_j) = \operatorname{argmax} \left\{ \prod_{i=1}^{n_j} P_i(q_j^*, \mathbf{g}_j, \mathbf{s}_j)^{R_{ij}} [1 - P_i(q_j^*, \mathbf{g}_j, \mathbf{s}_j)]^{1-R_{ij}} \right\}, \quad (2)$$

其中 $Q_j = 2^K - 1$ 表示新题 j 所有可能 q 向量的集合。此外，SIE 方法使用 EM 算法来估计新题的项目参数。

意见 6：谭青荣使用了 BIC 和 RMSEA 方法在线标定 q 向量，SIE 使用 MLE 方法估计 q 向量，为何本文不与谭青荣的方法比较在线标定 q 向量的效率，相信二者的耗时都不会太长。建议增加非参数的 q 向量在线标定方法的实验。

回应：非常感谢专家的建议。根据专家的建议，我们补充比较了 SIE、RMSEA-N、SIE-R-BIC 和 IGEOCM 方法的标定精度和标定效率。模拟研究结果如下表 2 所示，多数条件下，IGEOCM 的项目标定精度都高于 SIE、RMSEA-N 和 SIE-R-BIC 方法。此外，尽管 RMSEA-N 方法和 SIE-R-BIC 方法的标定效率略优于 SIE 方法（RMSEA-N 方法和 SIE-R-BIC 方法的平均运行

时间 ART 略短于 SIE 方法), 但仍远不如 IGEOCM。RMSEA-N 方法在所有条件下的平均 ART 值约为 IGEOCM 的 39 倍, SIE-R-BIC 方法在所有条件下的平均 ART 值约为 IGEOCM 的 46 倍。总体而言, 相比于 RMSEA-N 和 SIE-R-BIC 方法, IGEOCM 在项目标定精度和标定效率上仍具有优势。

但限于文章的篇幅及表达的简洁, 我们在文章的讨论部分对该结果进行了补充说明。

表 2 不同条件下 SIE、RMSEA-N、SIE-R-BIC 和 IGEOCM 方法的标定精度及标定效率

分布		方法	40	80	120	160	200
AVCER	高阶分布	SIE	0.613	0.831	0.913	0.952	0.972
		RMSEA-N	0.400	0.674	0.799	0.882	0.931
		SIE-R-BIC	0.618	0.834	0.915	0.950	0.972
		IGEOCM	0.677	0.862	0.929	0.962	0.977
	均匀分布	SIE	0.798	0.948	0.987	0.993	0.998
		RMSEA-N	0.787	0.956	0.986	0.993	0.997
		SIE-R-BIC	0.805	0.952	0.990	0.993	0.998
		IGEOCM	0.821	0.952	0.985	0.993	0.998
	正态分布	SIE	0.382	0.505	0.577	0.630	0.677
		RMSEA-N	0.190	0.268	0.320	0.355	0.379
		SIE-R-BIC	0.395	0.511	0.581	0.635	0.678
		IGEOCM	0.460	0.586	0.666	0.728	0.774
RMSE	高阶分布	SIE	0.092	0.061	0.048	0.041	0.037
		RMSEA-N	0.132	0.098	0.084	0.077	0.074
		SIE-R-BIC	0.088	0.066	0.057	0.053	0.050
		IGEOCM	0.092	0.061	0.048	0.041	0.037
	均匀分布	SIE	0.122	0.069	0.052	0.045	0.039
		RMSEA-N	0.115	0.091	0.082	0.079	0.077
		SIE-R-BIC	0.098	0.069	0.056	0.050	0.046
		IGEOCM	0.119	0.069	0.053	0.045	0.039
	正态分布	SIE	0.122	0.076	0.059	0.050	0.045
		RMSEA-N	0.171	0.150	0.139	0.130	0.127

		SIE-R-BIC	0.097	0.073	0.064	0.058	0.055
		IGEOCM	0.121	0.076	0.060	0.050	0.045
ART	高阶分布	SIE	100.796	177.369	254.757	334.438	413.676
		RMSEA-N	63.319	122.240	180.282	237.766	294.350
		SIE-R-BIC	91.944	151.579	210.229	268.374	325.499
		IGEOCM	1.756	3.092	4.404	5.687	6.997
	均匀分布	SIE	110.866	188.895	268.312	346.661	423.562
		RMSEA-N	56.724	109.335	161.455	213.776	265.627
		SIE-R-BIC	85.361	141.107	196.119	250.612	304.330
		IGEOCM	1.724	3.003	4.315	5.592	6.862
	正态分布	SIE	102.043	177.113	255.225	336.941	414.950
		RMSEA-N	59.750	114.812	169.287	223.990	277.890
		SIE-R-BIC	88.154	144.714	201.128	256.935	308.703
		IGEOCM	1.804	3.167	4.474	5.798	7.110

意见 7：研究设计中标定样本间隔 20，似乎过于频繁？高阶能力中为何设置 $\lambda_0 = (-1, -0.6, -0.2, 0.2, 0.6, 1)$ ，且对所有属性 k 均有 $\lambda_{1k} = 1.5$ 。依据是什么？通常情况下，更常用的是分别从正态分布和对数正态分布中产生的。

回应：非常感谢专家的建议。在修改的文章中，我们将标定样本间隔修改为 40，标定样本修改为 40、80、120、160、200。本研究对于 λ_{0k} 和 λ_{1k} 的取值的设定主要是参考了 de la Torre 和 Chiu (2016) 研究，在他们研究中，高阶能力分布下， λ_{0k} 在区间 (-1,1) 以相等的间隔取值，并设置 $\lambda_{1k} = 1.5$ 。同时，根据您的建议，我们检验了 λ_{0k} 和 λ_{1k} 分别从正态分布和对数正态分布中产生时 SIE 方法和 IGEOCM 的标定精度，结果如表 3 所示。表 3 结果表明，不论在哪种实验条件下，IGEOCM 方法整体上均优于 SIE 方法。但限于文章的篇幅及表达的简洁，我们在文章的讨论部分对该结果进行了补充说明。

表 3 不同 λ_{0k} 和 λ_{1k} 产生方式下 SIE 和 IGEOCM 方法的项目标定精度

		方法	40	80	120	160	200
AVCER	条件 1	SIE	0.589	0.786	0.860	0.897	0.920
		IGEOCM	0.641	0.823	0.885	0.913	0.938
	条件 2	SIE	0.606	0.812	0.896	0.942	0.965
		IGEOCM	0.668	0.857	0.916	0.950	0.966
RMSE	条件 1	SIE	0.134	0.088	0.068	0.058	0.051
		IGEOCM	0.132	0.085	0.069	0.060	0.052
	条件 2	SIE	0.095	0.062	0.049	0.041	0.037
		IGEOCM	0.095	0.062	0.049	0.041	0.037

注：条件 1 表示 λ_{0k} 和 λ_{1k} 分别从正态分布和对数正态分布中产生；条件 2 表示设置 $\lambda_0 = (-1, -0.6, -0.2, 0.2, 0.6, 1)$ ，且对于所有属性 k 均有 $\lambda_{1k} = 1.5$ 。

意见 8：记得结果中可以采用标定效率作为评价指标，建议增加这一部分。另外，结果表格中似乎不用呈现所有标定被试的结果？

回应：非常感谢您的建议！CD-CAT 中同时标定新题 Q 矩阵和项目参数时使用 Q 矩阵和项目参数间的迭代估计次数作为标定效率评价指标(陈平, 辛涛, 2011)。此外，CD-CAT 中仅标定新题项目参数时使用 EM 算法循环次数和标定新题的平均运行时间作为标定效率评价指标(陈平, 2016)。考虑到 IGEOCM 方法直接基于被试的估计属性掌握模式来标定新题 Q 矩阵，继而直接标定项目参数，两者间无需循环估计，因此本研究使用平均运行时间作为标定效率指标。各方法的平均运行时间结果呈现在文章图 2（研究 1 结果）和表 3（研究 2 结果）中。此外，我们将标定样本间隔修改为 40，呈现标定样本为 40、80、120、160、200 时 SIE 方法和 IGEOCM 的标定结果。

陈平. (2016). 两种新的计算机化自适应测验在线标定方法. *心理学报*, 48(9), 1184-1198.

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.

意见 9：属性为 6 和 8 的实验为什么要分开写？建议结果中某些内容可以做成图的形式，运

用多种方法呈现结果。

回应：非常感谢您的建议。由于在测验属性为 6 的实验中考察了标定样本量，属性掌握模式分布和被试新题作答个数对两在线标定方法的影响，而在测验属性为 8 的实验中将被试作答新题的个数固定为 6（因被试新题作答个数对标定精度的影响较小，因此固定新题作答个数），仅考察了标定样本量和属性掌握模式分布对两在线标定方法的影响。因此，将两实验结果分开呈现。但根据审稿专家的意见及文章表达的简洁性，我们在修改的文章中，考虑到研究间的递进关系，删除了 8 属性下的实验，在研究 1 的基础上进一步检验了 CD-CAT 选题策略对两在线标定方法的影响。具体细节见研究 2。此外，我们将原表 1 和表 3 修改为图 1 和图 2，使用多种方法呈现结果。再次感谢审稿专家宝贵的意见及建议！

.....

审稿人 2 意见：

作者提出一种基于熵的信息增益的在线标定方法（记为 IGEOCM），并采用 Monte Carlo 模拟实验验证开发的新方法的效果，并同时与已有的在线标定方法 SIE 进行比较。研究结果表明，与 SIE 相比，IGEOCM 是 CD-CAT 题库中项目的增补的一种更为高效、准确的方法。论文具有一定的创新性，也有一定的前瞻性。但论文存在以下问题，供作者参考。

意见 1：文章太长了，建议缩短篇幅，建议删除例子。

回应：非常感谢专家的建议。根据您的建议，我们删除了文中的例子，但为了说明不同 q 向量下条件熵 $E(R_j|q_j)$ 和熵的信息增益 $g(R_j, q_j)$ 的变化，我们在 3.2.1 节中增加了表 1。

意见 2：建议不要使用第一人称，如本文、本研究等，建议使用第三人称。

回应：非常感谢您的建议，我们已将文章中第一人称修改为第三人称。

意见 3：对于作者 P. Chen，有时用 Chen，有时用 P. Chen，如在 MCAT 中，P. Chen（2017）对 Method A，OEM 和 MEM 方法进行拓展，用了全名，有些地方却用了姓 Chen，建议作者统一。

回应：非常感谢您指出这一问题。在修改后的文章中，我们已统一将“P. Chen”修改为“Chen”。

意见 4: 4.1 研究 1 实验设计, 改为实验设计; 4.2 研究 1 实验结果改为实验结果。

回应: 非常感谢您的建议! 根据您的建议, 我们已将“4.1 研究 1 实验设计”修改为“4.1 实验设计”; “4.2 研究 1 实验结果”修改为“4.2 实验结果”。

意见 5: 为什么只选择属性为 6 和 8? 研究一和研究是平行关系? 平行关系意义大吗? 为什么不在研究二用递进关系呢? 建议研究二在研究一的基础再探讨某个影响因素, 深度挖掘作用机制, 用递进似乎更合理。

回应: 非常感谢您的建议。在修改的文章中, 我们删除了 8 属性下的研究结果, 并在研究 1 的基础上探讨了选题策略对各在线标定方法的影响。具体细节请参见研究 2。

意见 6: ART 只可以作为参考, 目前计算机性能都比较高, 关键还是要看两种方法的性能差异, 但从结果来看不是非常明显。建议作者行文语气稍缓和。例如, 新方法使用非参数化方法标定新题 Q 矩阵, 极大地提高了标定的效率, 这句话的语气还是要斟酌的, 是否是极大地呢? 建议用“改善”。

回应: 非常感谢专家的建议, 我们已对文章中的措辞进行修改, 缓和行文语气。

意见 7: 差值达到多大才算比另外一种方法好呢? 有没有标准? 这个标准有没有文献支撑? 似乎作者在行文中没有交待。

回应: 非常感谢您的建议! 根据 Chen 等人(2015)的研究, 两种方法之间差值大于等于 0.01 表明一种方法优于另一种方法。我们已在“4.2 实验结果”节中给出相关说明。

Chen, Y., Liu, J., & Ying, Z. (2015). Online item calibration for Q-matrix in CD-CAT. *Applied psychological measurement, 39*(1), 5-15.

意见 8: 有些结果用表表达是可以的, 但更建议作者对有些结果用图来表达, 因为用图表达更为直观。

回应: 非常感谢您的建议! 在修改的文章中, 我们将原表 1 和表 3 的结果使用图 1 和图 2 来呈现, 使表达更为直观。

第二轮

审稿人 1 意见:

意见 1: 既然在相同实验条件下进行的模拟实验, 建议在研究 1 中直接将新方法与 SIE、RMSE-N、SIE-R-BIC 同类方法进行比较, 而无需在讨论部分呈现。

回应: 非常感谢专家的建议! 根据您的建议, 我们在研究 1 中新增了新方法与 RMSE-N 和 SIE-R-BIC 方法的比较且在文章第 2 部分增加了对 RMSE-N 和 SIE-R-BIC 方法的介绍。

意见 2: 表 2 中的数值有的是保留三位小数, 有的是两位, 是否可以保持全文统一。

回应: 非常感谢您的建议! 我们已将表 2 中的数值都修改为保留三位小数。

意见 3: 附表 1 既然是在研究 1 的实验条件下得到的结果, 表中并未体现被试作答新题个数 D 这一变量, 是否可以清楚交代?

回应: 感谢您的建议! 根据您的建议, 我们在附表 1 中增加了被试作答新题个数 D 这一变量 (即在原有 6 个新题的基础上, 我们增加了被试作答 4 和 8 个新题的结果), 且将该部分结果添加至研究 1 并与已有研究 1 的结果进行合并。

意见 4: 讨论部分提及与 RMSE-N 和 SIE-R-BIC 方法(谭青蓉, 2019) 的比较, 在引言部分是否可以补充该方法的综述?

回应: 非常感谢专家的建议! 根据您的建议, 我们在引言部分补充了 RMSE-N 和 SIE-R-BIC 方法的综述。

.....

审稿人 2 意见:

根据审稿专家意见, 作了较多修改。但论文仍然存在以下问题:

意见 1: 建议将主要的推导公式附文后, 而不是在第一次审稿基础直接删除, 因为新方法是要一定据理的。

回应: 非常感谢专家的宝贵建议! 我们在 3.2.1 中补充了 IGEOCM 方法标定新题 Q 矩阵合理性的公式推导; 同时, 我们还在附录中通过举例【例 1】进一步说明了 IGEOCM 最小化

条件熵估计新题 j 的 q 向量的合理性。即我们分别通过公式推导以及实例示例两方面来说明新方法的合理性及科学性。

意见 2: 建议模拟次数提高至 1000 次，或者 500 次也可以，如果对结果有影响，那也是要考虑的。

回应: 非常感谢您的建议！我们已在修改的文章中将模拟次数提高至 500 次。

意见 3: 论文实际是两种方法的比较，但论文题目似乎“文不对题”。

回应: 非常感谢专家的建议！本文主要是在认知诊断计算机自适应测验框架下，基于一种新的视角（即熵的信息增益和 EM 算法）开发了一种全新的在线标定方法即 IGEOCM，为了验证新开发方法的特点及优势，我们实验中将新方法与传统的方法（即 SIE 等方法）进行了综合比较。因此，本文的研究重点（或本研究的主要贡献）仍为新方法 IGEOCM 的开发。结合两位审稿专家的意见及本文研究重点，我们将本文的名称命名为“一种高效的 CD-CAT 在线标定新方法：基于熵的信息增益与 EM 视角”。

意见 4: 感觉两种方法结果相差不大，难以说明是“优于”，即便假定“优于”，也是要解释其原因的。

回应: 非常感谢您的建议！根据 Chen 等人(2015)的研究，两方法间标定精度的差值大于等于 1% 表明一种方法优于另一种方法。本研究中，SIE 方法和 IGEOCM 在项目参数标定精度上具有较为相似的性能（两方法标定精度之差小于 1%）。但在 Q 矩阵标定上，IGEOCM 在高阶分布和正态分布下具有较为明显的优势：在均匀分布下 SIE 方法和 IGEOCM 间的最大 AVCER 差值为 2.3%，而在属性掌握模式为高阶分布和正态分布时，两方法间的最大 AVCER 差值分别高达 6.5% 和 10%。因此，结合 Chen 等人(2015)的标准，整体而言，本文新开发的 IGEOCM 优于 SIE 方法。同时，根据审稿专家的意建，我们在“总结与讨论”部分从理论层面分析及解释了 IGEOCM 优于 SIE 方法的原因。谢谢审稿专家的建议。