

# 基于变分自编码器的伽马单中子出射反应截面 实验数据离群点研究\*

谢金辰<sup>1)</sup> 陶曦<sup>1)</sup> 续瑞瑞<sup>1)†</sup> 田源<sup>1)</sup> 邢康<sup>1)</sup> 葛智刚<sup>1)</sup> 牛一斐<sup>2)3)‡</sup>

1) (中国原子能科学研究院, 中国核数据中心, 北京 102413)

2) (兰州大学核科学与技术学院, 兰州 730000)

3) (兰州大学, 稀有同位素前沿科学中心, 兰州 730000)

(2024年12月25日收到; 2025年1月24日收到修改稿)

伽马单中子出射反应截面是核工程输运计算中的重要参数, 部分核素 ( $\gamma, n$ ) 的反应测量因来自不同实验室而分歧明显. 本文基于变分自编码器方法, 针对原子核质量数在 29—207 区域的伽马单中子出射反应截面实验测量数据进行分析, 有效识别多家测量之间的离群点. 首先, 研究变分自编码器方法, 建立伽马单中子光核测量数据离群点识别网络; 其次, 对  $^{29}\text{Si}$ ,  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{141}\text{Pr}$ ,  $^{181}\text{Ta}$ ,  $^{206}\text{Pb}$  和  $^{207}\text{Pb}$  的 29 家多能点测量数据进行离群点识别; 最后, 计算离群点识别前后的实验数据与国际原子能机构光核评价数据库 (IAEA-2019-PD) 评价价值之间的偏差, 检测变分自编码器的分析效果. 研究表明, 变分自编码器方法可以有效识别 ( $\gamma, n$ ) 反应实验测量离群点, 其中  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{181}\text{Ta}$ ,  $^{206}\text{Pb}$  和  $^{207}\text{Pb}$  的伽马单中子出射反应截面与 IAEA-2019-PD 评价结果一致性更高, 验证了该方法在核数据研究中的应用潜力.

**关键词:** 变分自编码器, 伽马单中子出射反应, 截面, 离群点

**PACS:** 25.20.-x

**DOI:** 10.7498/aps.74.20241775

**CSTR:** 32037.14.aps.74.20241775

## 1 引言

伽马单中子出射反应截面有助于理解核结构和核反应机制, 同时对屏蔽设计与放射性输运分析、核废料嬗变、医用同位素生产和天体物理应用等方面有着重要应用价值<sup>[1]</sup>.

一直以来, 由于高品质光子源实验条件稀缺和光子能量校准等技术难题的限制, 光中子反应实验数据质量参差不齐, 实验数据分散性大, 给核数据评价带来很多困难<sup>[2]</sup>. 20 世纪 60 年代, 美国的

Dietrich 和 Berman<sup>[3]</sup> 对准单能光源的光核反应实验数据开展系统分析与编撰, 为核数据评价做出了突出贡献. 为了进一步建立信息完整的光核数据库, 国际原子能机构 (IAEA) 分别在 1996 年和 2016 年组织国际协调合作, 给出了光中子核反应数据推荐值, 工作中部分分析了美国劳伦斯利弗莫尔实验室、法国萨克雷实验室以及俄罗斯国立大学等不同机构光中子测量的差异, 结合核反应理论方法, 先后建立了 IAEA 光核数据库 IAEA-1999-PD<sup>[1]</sup> 和 IAEA-2019-PD<sup>[4]</sup>. 然而, 大量早期测量数据因年代久或者文献报道不完整难以追溯; 国际韧致辐

\* 国家重点研发计划 (批准号: 2022YFA1602403, 2021YFA1601500)、国家自然科学基金 (批准号: 12005303, 12075104)、重点实验室重点项目 (批准号: JCKY2023201C153-5) 和持续支持基础科学研究项目 (批准号: BJ010261223282) 资助的课题.

† 通信作者. E-mail: xuruirui@ciae.ac.cn

‡ 通信作者. E-mail: niuyf@lzu.edu.cn

射装置、准单能装置等测量之间分歧难以澄清;同时众多核素尚缺少实测数据,无法标定核数据可靠性.上述问题在传统评价技术中尚无法得到处理,需要发展新的核数据评价方法以提升光中子数据质量.

实验数据评价是核数据研究的核心任务之一,其中评价测量数据中的异常数据点(即离群点)是提升核数据评价质量的重要环节.近年来,随着机器学习方法的快速发展,相关技术已经被广泛应用于核物理的多个领域,有效地促进了从理论模型到实验分析各方面的发展<sup>[5-10]</sup>.例如,在监督学习应用于核反应数据方面,兰州大学白景虎等<sup>[8]</sup>利用多任务神经网络提高了光核反应中巨偶极共振参数的精度,中国原子能科学研究院的邢康等<sup>[9]</sup>使用相移神经网络重现了<sup>235</sup>U(n, f)反应共振能区的低频和振荡截面,安徽大学的李伟峰等<sup>[10]</sup>采用贝叶斯神经网络方法改进了(n, 2n)反应截面的预测.相较于监督学习,非监督机器学习具有不用标签、适用于发现数据的隐藏结构和潜在模式等优势,在数据聚类分析、离群点检测等工作中具有一定优势<sup>[5,11]</sup>.目前,非监督学习方法在核反应数据研究中已经有一些应用工作,例如,亚利桑那州立大学的Mitra等<sup>[12]</sup>使用非监督学习的变分自编码器(variational autoencoder, VAE)等神经网络,学习和预测了2151个核素的截面关系,预测准确性高;美国阿贡国家实验室的Fox和Wendt<sup>[13]</sup>使用深度生成对抗网络开发了一个预测模型,学习基于核反应程序TALYS建立的评价核数据库(TENDL)中偶-偶核的中子非弹性散射趋势,为<sup>12</sup>C, <sup>16</sup>O, <sup>20</sup>Ne, <sup>172</sup>Yb和<sup>174</sup>Hf等多个核素的截面曲线提供了良好预测.然而,非监督学习尚未应用于开展核反应截面离群点分析.

变分自编码器(VAE)是一种生成模型,通过结合深度学习和贝叶斯推理的技术来学习输入数据的潜在表示<sup>[14]</sup>.与传统的自编码器不同,VAE在编码器部分引入了概率分布,使得生成的潜在空间具有良好的连续性和结构性,在数据清理方面具有更好的应用前景.为了有效查找光中子出射反应截面实验测量的离群点,本文提出采用变分自编码器<sup>[14]</sup>的非监督机器学习方法开展研究,从国际实验核反应数据库EXFOR中<sup>[15]</sup>获取实验光中子数据,挖掘其潜在变量进而筛选实验离群点,为核数据评价提供更为可靠的实验信息.

本文主要安排包括:第2节介绍伽马单中子实

验数据分析VAE网络模型原理与构建;第3节介绍<sup>29</sup>Si到<sup>207</sup>Pb光中子实验数据训练,以及离群点识别结果与分析;第4节对VAE方法在核数据离群点识别进行总结与展望.

## 2 基于变分自编码器(VAE)的离群点识别网络

### 2.1 VAE基本原理

VAE网络由编码器和解码器两部分组成<sup>[14]</sup>.假设 $x$ 为原始训练数据,其对应分布为 $p(x)$ , $z$ 为 $x$ 通过编码器确定的潜在变量, $\tilde{x}$ 为解码器重建的 $x$ 近似值.整体来讲,该网络首先建立编码器,通过学习原始数据 $x$ 提取其潜在变量 $z$ 的后验分布 $p(z|x)$ ,在 $p(z|x)$ 基础上抽样以获取潜在变量 $z$ .接着建立解码器,使用编码器所得潜在变量 $z$ 构建 $\tilde{x}$ 的条件概率分布 $p(\tilde{x}|z)$ ,再对 $p(\tilde{x}|z)$ 抽样获得重建数据 $\tilde{x}$ .具体原理如下.

根据贝叶斯定理,编码器 $p(z|x)$ 可以表示为

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}, \quad (1)$$

其中, $p(z)$ 为潜在变量 $z$ 的先验分布,一般选取 $p(z)$ 为标准正态分布 $\mathcal{N}(0, I)$ ;由于 $p(x)$ 和 $p(x|z)$ 难以直接确定,因此无法获得后验分布 $p(z|x)$ .在VAE网络中,采用变分推断法,使得编码器输出的变分后验分布 $q_\phi(z|x)$ 逼近真实的后验分布 $p(z|x)$ ,其中 $q_\phi(z|x)$ 多采用高斯形式, $\phi$ 为编码器参数.

变分推断法的目标是最小化KL散度(Kullback-Leibler divergence)<sup>[14]</sup>.其中,在VAE网络中定义KL散度 $D_{\text{KL}}(q_\phi(z|x)||p(z|x))$ 衡量 $q_\phi(z|x)$ 与 $p(z|x)$ 之间的距离,如(2)式:

$$D_{\text{KL}}(q_\phi(z|x)||p(z|x)) = \int_z q_\phi(z|x) \log \left[ \frac{q_\phi(z|x)}{p(z|x)} \right] dz, \quad (2)$$

通过最小化KL散度能够使编码器输出的分布逼近真实后验分布.对 $D_{\text{KL}}$ 进行简单变换,可转化为

$$\begin{aligned} D_{\text{KL}}(q_\phi(z|x)||p(z|x)) &= \int_z q_\phi(z|x) \log \left[ \frac{q_\phi(z|x)p(x)}{p(z,x)} \right] dz \\ &= \int_z q_\phi(z|x) \log \left[ \frac{q_\phi(z|x)}{p(z,x)} \right] dz \\ &\quad + \int_z q_\phi(z|x) \log p(x) dz, \end{aligned} \quad (3)$$

其中,  $\log p(x)$  与积分变量  $z$  无关, 考虑到  $\tilde{x}$  是由  $z$  生成的  $x$  的近似值, 且联合概率分布  $p(z, x)$  可以表示为  $p(z, \tilde{x}) = p_\theta(\tilde{x}|z)p(z)$ . 因此 (3) 式可以转换为

$$D_{\text{KL}}(q_\phi(z|x)||p(z|x)) + \mathbb{E}_{q_\phi(z|x)} \left[ \log \frac{p_\theta(\tilde{x}|z)p(z)}{q_\phi(z|x)} \right] = \log p(x), \quad (4)$$

其中, 左边第二项被称为证据下界 (evidence lower bound, ELBO);  $p_\theta(\tilde{x}|z)$  为解码器神经网络参数化后的潜在变量  $z$  的后验分布,  $\theta$  是解码器神经网络参数. 考虑到原始训练数据  $x$  确定的条件下,  $\log p(x)$  为固定值, 为了最小化  $D_{\text{KL}}$ , 需要最大化 ELBO.

ELBO 经过简单变换可以转化为

$$\text{ELBO} = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(\tilde{x}|z)] - D_{\text{KL}}(q_\phi(z|x)||p(z)). \quad (5)$$

一般将解码器的输出视为  $p_\theta(\tilde{x}|z)$  的均值, 即重建数据  $\tilde{x}$ , 无需再重复抽样. (5) 式右边第一项被称为重建项, 为了最大化重建项, 需最小化原始数据  $x$  与重建数据  $\tilde{x}$  的差异; 右边第二项被称为正则化项, 由于  $q_\phi(z|x)$  与  $p(z)$  均取为高斯分布, 因此  $D_{\text{KL}}(q_\phi(z|x)||p(z))$  有解析解, 可以转换

$$D_{\text{KL}}(q_\phi(z|x)||p(z)) = \frac{1}{2} \sum_k (\sigma_{\phi,k}^2 + \mu_{\phi,k}^2 - 1 - \log \sigma_{\phi,k}^2), \quad (6)$$

其中, 当潜在变量  $z$  数量大于 1 时, 通常采用参数化的均值  $\mu_k$  和方差  $\sigma_k^2$  来表示第  $z_k$  个潜在变量的高斯分布  $q_\phi(z_k|x)$ . 为了实现 VAE 网络中 ELBO 项最大化, 需要针对 (5) 式建立损失函数 (LOSS). 其中, (5) 式中的重建项通常通过计算均方误差 (MSE) 来实现, 同时通过在正则化项上乘以系数  $\beta$  共同建立 VAE 网络损失函数, 如 (7) 式所示<sup>[16]</sup>:

$$\text{LOSS} = \text{MSE}(x, \tilde{x}) + \beta \frac{1}{2} \sum_k (\sigma_{\phi,k}^2 + \mu_{\phi,k}^2 - 1 - \log \sigma_{\phi,k}^2). \quad (7)$$

基于上述公式体系, VAE 网络的训练过程遵循如下步骤: 首先, 将原始数据  $x$  输入编码器获得潜在变量  $z$  的分布  $q_\phi(z|x)$ ; 其次, 从  $q_\phi(z|x)$  中获得潜在变量  $z$  的值, 将  $z$  输入到解码器中产生重建数据  $\tilde{x}$ ; 最后, 通过计算损失函数 LOSS 获得损失, 并利用该损失进行反向传播以优化网络参数  $\phi$  和  $\theta$ . 训练过程涉及多次迭代, 直至网络达到设定的收敛条件, 训练结束. 其中需要注意, 在第二步计算中, 直接从  $q_\phi(z|x)$  中抽样得到  $z$  会导致 VAE 网络不可微分, 无法进行反向传播更新网络参数. 因此, 常常将  $z$  的值重建为一个关于随机噪声的可微函数, 如 (8) 式所示:

$$z_k = \mu_k + \sigma_k \cdot \epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0, I). \quad (8)$$

即潜在变量  $z_k$  被重构为均值  $\mu_k$  与方差  $\sigma_k$  的函数, 其中  $\epsilon_k$  为满足标准正态分布的随机值.

## 2.2 基于光中子反应截面的离群点识别网络构建

本研究基于光中子核反应截面的实验信息, 建立以 VAE 方法的离群点识别网络. 针对不同的原子核, 训练与质子数  $Z$  和质量数  $A$  相关的 VAE 网络参数  $\phi$  和  $\theta$ , 构建的 VAE 网络结构如图 1 所示, 其中每层均为全连接线性层, 图中括号内为神经元数量.

在图 1 的具体计算中, 首先将实验数据归一至 0—1 的范围内, 随后获取公共能量范围, 接着对公共能区使用核密度估计 (kernel density estimation, KDE)<sup>[17]</sup> 确定实验数据的能量分布概率密度  $\hat{f}(E)$ , 基于分布  $\hat{f}(E)$  随机获取 100 个能量点, 再将每家测量能量点对应的截面值组合为一批 (batch) 作为 VAE 网络的输入  $x$ . 将多家测量数据的能量  $E$  视为一维独立同分布的数据点, 则有如下 KDE 的表达式:

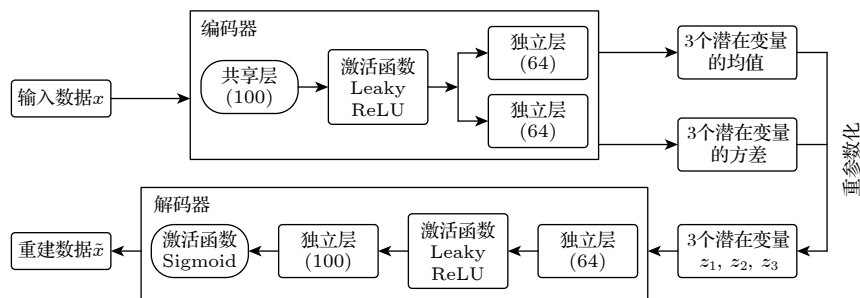


图 1 VAE 网络示意图, 括号内为该层神经网络神经元数量

Fig. 1. Schematic diagram of VAE network, with the number of neurons in this layer in parentheses.

$$\hat{f}(E) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{E - E_i}{h}\right), \quad (9)$$

其中,  $E$  为目标变量 (能量),  $E_i$  为第  $i$  个实验测量点的能量,  $\hat{f}(E)$  为  $E$  处的概率密度估计值,  $n$  为数据点数量,  $h$  为带宽参数 (平滑宽度),  $K(\cdot)$  为核函数. 本文选用高斯函数作为核函数  $K(u)$ , 其定义为

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right). \quad (10)$$

带宽  $h$  的选择对估计结果的平滑程度有显著影响. 为此, 本文采用了 Scott 规则 (11) 式自动计算带宽:

$$h = \sigma \cdot n^{-1/5}, \quad (11)$$

其中,  $\sigma$  为样本数据的标准差.

输入数据  $x$  通过一个共享层和两个独立层组成的编码器, 映射到三个潜在变量  $z_1, z_2$  和  $z_3$  的均值和方差上, 随后利用均值和方差重参数化生成潜在变量  $z$  的值, 再通过由两个独立层组成的解码器输出重建数据  $\tilde{x}$ . 为防止梯度消失, 编码器和解码器层间均采用 Leaky ReLU 激活函数 [18]. 同时, 解码器尾部额外增加 Sigmoid 激活函数使输出数据重建在 0—1 之间. 网络采用何凯明初始化方法对网络参数进行初置 [19], 偏置设置为 0.01, 以防止偏置节点失效. 优化器使用 Adam [20], 学习率为 0.01. 损失函数 LOSS 中正则化项的  $\beta$  参数设为 0.01.

当 VAE 训练在最后 100 次迭代的线性回归斜率绝对值小于 0.0005 时, 可以认为损失函数趋于稳定, 同时潜在变量  $z$  的分布近似标准正态分布, 重建数据  $\tilde{x}$  与原始数据  $x$  基本一致时, 可以认为网络训练完成, 所得 VAE 网络模型有效地捕捉到了数据的潜在结构特征.

### 2.3 实验数据离群点识别

在 VAE 网络训练完成后, 将每家测量数据潜在变量的均值 ( $\mu_1, \mu_2, \mu_3$ ) 进行聚类分析, 并选择具有最大成员数的簇作为判断标准. 处于簇之内的测

量为正常数据, 簇之外的测量为离群数据. 聚类采用  $k$ -means 算法 [21], 簇数量  $k$  通过轮廓系数法确定 [22], 令  $k$  的值至少为 2. 每个样本的轮廓系数  $s(i)$  通过 (12) 式计算, 取值范围为 -1—1,  $s(i)$  值越大表示聚类效果越好.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \quad (12)$$

其中,  $a(i)$  是聚类的内聚度, 表示样本  $i$  到其同一簇中所有其他样本的平均距离;  $b(i)$  是聚类的分离度, 表示样本  $i$  到其他簇中所有样本的平均距离.

### 3 $^{29}\text{Si}$ 到 $^{207}\text{Pb}$ 核区的离群点识别结果讨论

本工作针对国际实验核反应数据库 (EXFOR) 中重核区光中子核反应实验测量进行整理, 目标聚焦包含三家及以上多能点测量的核素  $^{29}\text{Si}$ ,  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{141}\text{Pr}$ ,  $^{181}\text{Ta}$ ,  $^{206}\text{Pb}$  和  $^{207}\text{Pb}$ , 系统分析离群点, 实验数据的年份、第一作者、能量区间等信息如表 1 所列.

以  $^{63}\text{Cu}$  为例, 图 2 给出训练过程中损失函数 LOSS 随迭代次数的变化, LOSS 在前 100 次迭代中由 112 逐渐下降至 5, 随后在至 1500 次迭代中缓慢下降并趋于收敛; 图 3 给出了 Katz [27] 和 Owen 等 [34] 的 VAE 重建数据  $\tilde{x}$  与输入的实验测量

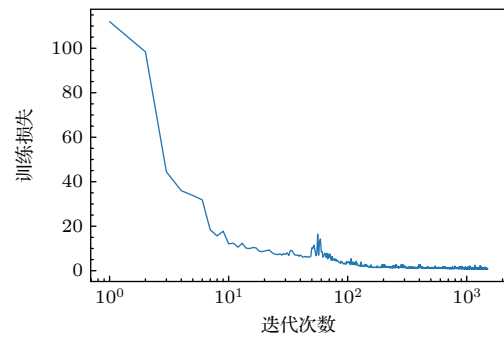


图 2 VAE 训练损失图

Fig. 2. VAE training loss chart.

表 1  $^{29}\text{Si}$  到  $^{207}\text{Pb}$  共 7 个核素的实验测量信息汇总表

Table 1. Summary table of experimental measurement information for 7 nuclides from  $^{29}\text{Si}$  to  $^{207}\text{Pb}$ .

核素	年份	第一作者	机构所属国家	机构	光源	能量区间	能点数	参考文献
$^{29}\text{Si}$	1973	K.Fukuda	日本	Radiation Centre of Osaka Prefecture	Bremsstrahlung	8.46—13.19	20	[23]
	1981	R.E.Pywell	澳大利亚	University of Melbourne	Bremsstrahlung	8.09—16.09	162	[24]
	1987	K.G.Mcneill	加拿大	University of Toronto	Annihilation radiation	9.52—18.71	29	[25]

表 1 (续)  $^{29}\text{Si}$  到  $^{207}\text{Pb}$  共 7 个核素的实验测量信息汇总表  
 Table 1 (continued). Summary table of experimental measurement information for 7 nuclides from  $^{29}\text{Si}$  to  $^{207}\text{Pb}$ .

核素	年份	第一作者	机构所属国家	机构	光源	能量区间	能点数	参考文献
$^{54}\text{Fe}$	1977	B.S.Ratner	俄罗斯	Institute for Nuclear Research of the Russian Academy of Sciences	Bremsstrahlung	16.00—26.50	43	[26]
	1951	L.Katz	加拿大	University of Saskatchewan	Bremsstrahlung	14.00—24.00	11	[27]
	1978	J.W.Norbury	澳大利亚	University of Melbourne	Bremsstrahlung	12.13—23.00	57	[28]
$^{63}\text{Cu}$	2012	C.Plaisir	法国	Centre d'Etudes Nucleaires de Bordeaux-Gradignan	Bremsstrahlung	11.00—19.00	81	[29]
	1968	R.E.Sund	美国	Gulf General Atomic	Annihilation radiation	10.12—25.06	66	[30]
	1979	L.Z.Dzhilavyan	俄罗斯	Institute for Nuclear Research of the Russian Academy of Sciences	Annihilation radiation	12.10—24.20	14	[31]
	1954	A.I.Berman	美国	Stanford University	Bremsstrahlung	13.00—29.00	9	[32]
	1951	L.Katz	加拿大	University of Saskatchewan	Bremsstrahlung	11.00—21.00	11	[27]
	1955	M.B.Scott	美国	University of Illinois	Bremsstrahlung	11.00—21.00	11	[33]
	1968	D.G.Owen	澳大利亚	University of Melbourne	Bremsstrahlung	10.80—23.60	129	[34]
	1951	P.R.Byerly Jr	美国	University of Pennsylvania	Bremsstrahlung	12.00—24.00	13	[35]
	1984	M.N.Martins	美国	National Institute of Standards and Technology	Virtual photons	11.00—30.00	5	[36]
$^{141}\text{Pr}$	1970	R.E.Sund	美国	Gulf General Atomic	—	8.89—23.69	100	[37]
	2006	H.Utsunomiya	日本	Konan University	Compton scattering	9.50—12.40	9	[38]
	1966	B.C.Cook	美国	Iowa State University	Bremsstrahlung	9.19—64.68	113	[39]
	1991	S.N.Belyaev	俄罗斯	Saratov Gosudarstvennyi University	Bremsstrahlung	7.98—23.52	185	[40]
$^{181}\text{Ta}$	1985	S.N.Belyaev	俄罗斯	Saratov Gosudarstvennyi University	Bremsstrahlung	7.68—11.63	80	[41]
	1951	L.Katz	加拿大	University of Saskatchewan	Bremsstrahlung	9.00—18.00	10	[27]
	2003	H.Utsunomiya	日本	Konan University	Compton scattering	7.80—12.00	13	[42]
	2006	S.Goko	日本	Konan University	Laser Compton Scattered Photons	9.20—12.30	6	[43]
$^{206}\text{Pb}$	2012	T.Kondo	日本	Konan University	Laser Compton Scattered Photons	8.14—12.24	11	[44]
	1985	S.N.Belyaev	俄罗斯	Saratov Gosudarstvennyi University	Bremsstrahlung	8.10—12.35	88	[41]
	1995	Y.Birenbaum	以色列	Ben Gurion University of the Negev	Monoenergetic photons	8.88—11.39	4	[45]
$^{207}\text{Pb}$	1995	Y.Birenbaum	以色列	Ben Gurion University of the Negev	Monoenergetic photons	7.37—9.72	5	[45]
	2012	T.Kondo	日本	Konan University	Laser Compton Scattered Photons	6.84—12.22	14	[44]
	1985	S.N.Belyaev	俄罗斯	Saratov Gosudarstvennyi University	Bremsstrahlung	6.77—11.18	95	[41]

$x$  比较图. 可以看出, 随着网络学习迭代次数的增加, 重建数据结果可以更好地再现实验测量, 两者截面之间的偏差

$$\Delta_{\text{偏差}} = \frac{|\sigma_{\text{实验数据}} - \sigma_{\text{重建数据}}| \times 100\%}{\sigma_{\text{实验数据}}} \quad (13)$$

也更接近零, 如图 3(b) 所示.

VAE 网络收敛后, 对多家测量对应的潜在变量的均值  $\mu_1$ ,  $\mu_2$  以及  $\mu_3$  进行聚类. 首先使用轮廓系数法获得  $k$ -means 聚类算法的最佳簇数量  $k$ , 随后选取最大成员数簇内的实验数据作为正常数据, 该簇外的实验数据为离群数据, 完成实验离群点识别. 其中  $^{63}\text{Cu}$  的聚类结果如图 4 所示, Plaisir 等<sup>[29]</sup>、

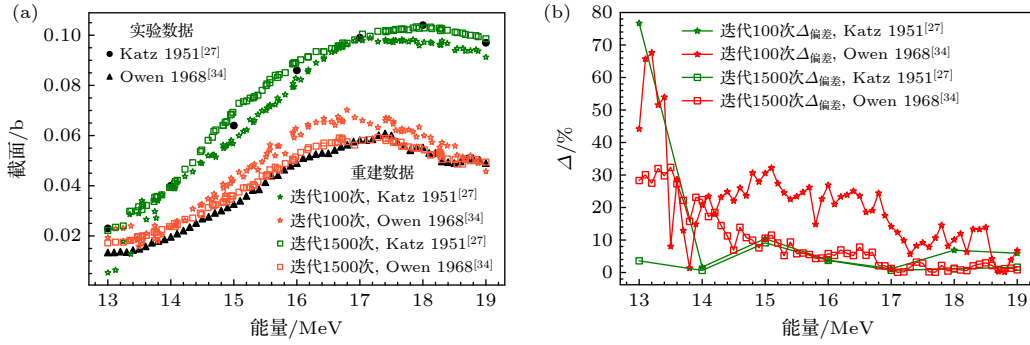


图 3  $^{63}\text{Cu}$  光中子核反应截面 VAE 网络 100 次迭代和 1500 次迭代训练结果与 Katz 和 Cameron<sup>[27]</sup>、Owen 等<sup>[34]</sup> 的实验测量比对 (a) VAE 数据重建结果; (b) 重建数据与实验数据的  $\Delta$  偏差

Fig. 3. Comparison of the  $^{63}\text{Cu}$  photoneutron reaction cross section training results of the VAE network after 100 and 1500 epochs with experimental measurements by Katz and Cameron<sup>[27]</sup>, and Owen et al.<sup>[34]</sup>: (a) The VAE data reconstruction results; (b) the  $\Delta$  偏差 between the reconstructed data and the experimental data.

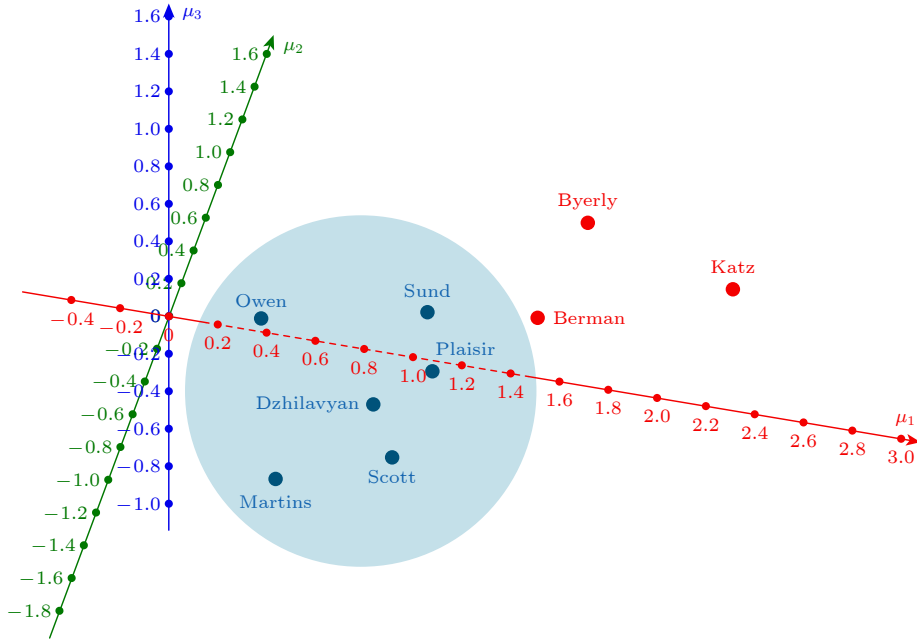


图 4 基于 Plaisir 等<sup>[29]</sup>、Sund 等<sup>[30]</sup>、Dzhilavyan 和 Kucher<sup>[31]</sup>、Scott 等<sup>[33]</sup>、Owen 等<sup>[34]</sup>、Martins 等<sup>[36]</sup>、Berman 和 Brown<sup>[32]</sup>、Katz 和 Cameron<sup>[27]</sup> 以及 Byerly Jr 和 Stephens 等<sup>[35]</sup> 的  $^{63}\text{Cu}$  实验测量的潜在变量的聚类图. 其中, 三个潜在变量的均值  $\mu_1$ ,  $\mu_2$  以及  $\mu_3$  作为坐标轴, 浅蓝色圆形阴影为最大成员簇的范围, 蓝色点为属于最大簇的实验测量, 红色为簇之外的实验测量

Fig. 4. Clustering plot of latent variables for experimental measurements  $^{63}\text{Cu}$ , based on the work of Plaisir et al.<sup>[29]</sup>, Sund et al.<sup>[30]</sup>, Dzhilavyan and Kucher<sup>[31]</sup>, Scott et al.<sup>[33]</sup>, Owen et al.<sup>[34]</sup>, Martins et al.<sup>[36]</sup>, Berman and Brown<sup>[32]</sup>, Katz and Cameron<sup>[27]</sup>, Byerly Jr and Stephens<sup>[35]</sup>. The plot uses three latent variables  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  as axes. The light blue circular shading represents the range of the largest cluster, with blue dots indicating experimental measurements belonging to the largest cluster, and red dots representing measurements outside the cluster.

Sund 等<sup>[30]</sup>、Dzhilavyan 和 Plaisir<sup>[31]</sup>、Scott 等<sup>[33]</sup>、Owen 等<sup>[34]</sup> 和 Martins 等<sup>[36]</sup> 的实验测量处于最大簇内, Berman 和 Brown<sup>[32]</sup>、Katz 和 Cameron<sup>[27]</sup> 以及 Byerly 和 Stephens<sup>[35]</sup> 的实验测量处于最大簇外, 为本工作 VAE 网络识别所得离群点. 图 5 给出了本工作  $^{29}\text{Si}$ ,  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{141}\text{Pr}$ ,  $^{181}\text{Ta}$  和  $^{207}\text{Pb}$  光中子核反应截面测量、VAE 离群点和

IAEA-2019 评价库中的结果比较. 由图可知, 针对光中子核反应测量, 本文的 VAE 网络有效识别了目标核素中的实验离群信息, 排除离群点之后的实验数据与 IAEA-2019 评价数据一致性提升.

为进一步评估变分自编码器在离群点识别中的效果, 在图 5 所示 IAEA-2019 数据与各家实验测量的公共区域, 定义排除离群点前后的实验室测

量数据与 IAEA-2019 评价数据之间的偏差, 如 (14) 式和 (15) 式所示.

$$D_{\text{积分偏差}} = \frac{1}{n} \sum_{i=1}^n \left| \int_{\text{IAEA-2019}_i} - \int_{\text{实验}_i} \right|, \quad (14)$$

其中,  $n$  代表该核素的测量家数总数,  $\int_{\text{IAEA-2019}_i}$  和  $\int_{\text{实验}_i}$  分别代表第  $i$  家测量 IAEA-2019 评价数据与实验数据公共能量范围内, 评价数据和实验数据的积分值. 积分过程首先对实验和评价数据进行线性插值, 随后采用梯形积分法进行积分.

$$P_{\text{实验点偏差}} = \frac{1}{n} \sum_{j=1}^n \left( \frac{1}{m_j} \sum_{i=1}^{m_j} \left| P_{\text{IAEA-2019}_{j_i}} - P_{\text{实验}_{j_i}} \right| \right), \quad (15)$$

其中,  $n$  代表该核素的测量家数总数,  $m_j$  代表第  $j$  家测量的测量总点数;  $P_{\text{IAEA-2019}_{j_i}}$  和  $P_{\text{实验}_{j_i}}$  分别表示第  $j$  家测量实验数据与 IAEA-2019 评价数据公共能量范围内, 第  $i$  个实验点能量对应的评价值和实验值.

根据 (16) 式和 (17) 式可计算排除离群点前后的变化量:

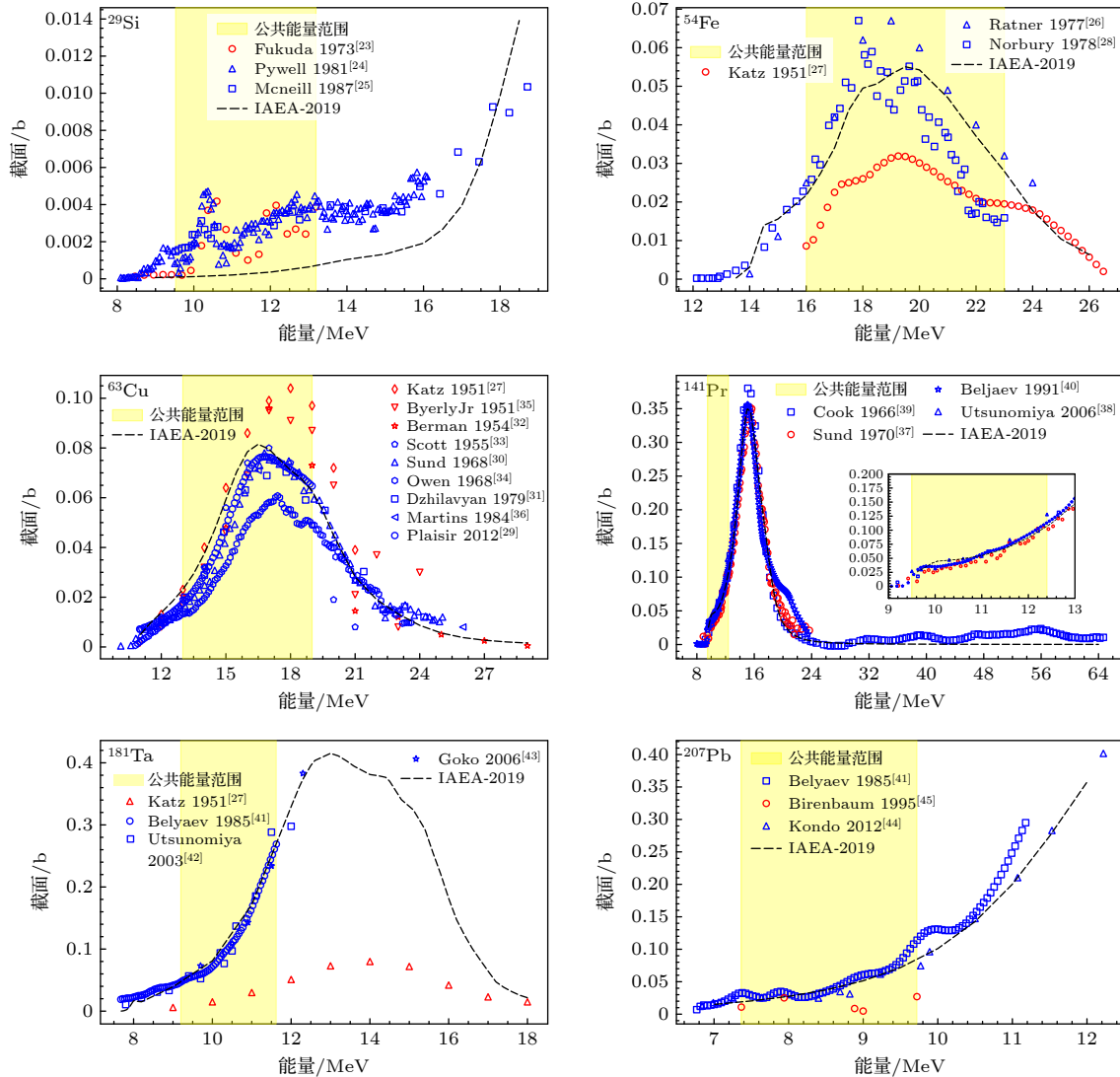


图 5  $^{29}\text{Si}$ ,  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{141}\text{Pr}$ ,  $^{181}\text{Ta}$  和  $^{207}\text{Pb}$  的光中子反应截面分布, 包括不同家实验数据 (散点) 和评价数据库 IAEA-2019 评估结果 (黑色虚线). 黄色区域为所有实验测量的公共能量范围, 红色数据点属于最大成员簇之外的离群数据, 蓝色数据为最大成员簇之内的正常数据. 其中  $^{141}\text{Pr}$  给出了 9—13 MeV 的放大图

Fig. 5. Photoneutron reaction cross section distributions for  $^{29}\text{Si}$ ,  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{141}\text{Pr}$ ,  $^{181}\text{Ta}$ , and  $^{207}\text{Pb}$ , including experimental data from different groups (scatter points) and the evaluated database IAEA-2019 results (black dashed line). The yellow region represents the common energy range of all experimental measurements. Red data points correspond to outliers outside the largest cluster, while blue data points belong to normal data within the largest cluster. A zoomed-in plot for  $^{141}\text{Pr}$  is provided for the 9 MeV to 13 MeV range.

$$\Delta_{\text{积分}} = \frac{D_{\text{积分偏差 (排除离群点前)}} - D_{\text{积分偏差 (排除离群点后)}}}{D_{\text{积分偏差 (排除离群点前)}}}, \quad (16)$$

$$\Delta_{\text{实验点}} = \frac{P_{\text{实验点偏差 (排除离群点前)}} - P_{\text{实验点偏差 (排除离群点后)}}}{P_{\text{实验点偏差 (排除离群点前)}}}. \quad (17)$$

如图 6 所示,  $^{54}\text{Fe}$ ,  $^{181}\text{Ta}$ ,  $^{206}\text{Pb}$  和  $^{207}\text{Pb}$  的  $\Delta_{\text{积分}}$  和  $\Delta_{\text{实验点}}$  均为正数, 表明在识别并排除离群点之前, 实验数据与 IAEA-2019 评价数据之间的偏差要大于排除离群点之后的偏差, 实验数据经过 VAE 排除离群点后与评价数据一致性更高;  $^{63}\text{Cu}$  的  $\Delta_{\text{积分}}$  为负但  $\Delta_{\text{实验点}}$  为正, 这表明实验数据点更接近评价数据, 但是整体较评价数据有偏移;  $^{29}\text{Si}$  的  $\Delta_{\text{积分}}$  和  $\Delta_{\text{实验点}}$  均为负数, 表明数据经过 VAE 排除离群点后与评价数据偏离, 结果仍待改进;  $^{141}\text{Pr}$  的  $\Delta_{\text{积分}}$  和  $\Delta_{\text{实验点}}$  同样为负, 这主要是因为实验数据的公共能量范围过小, 导致 VAE 无法有效地获取数据整体的潜在变量, 使得 VAE 方法评估不合理。

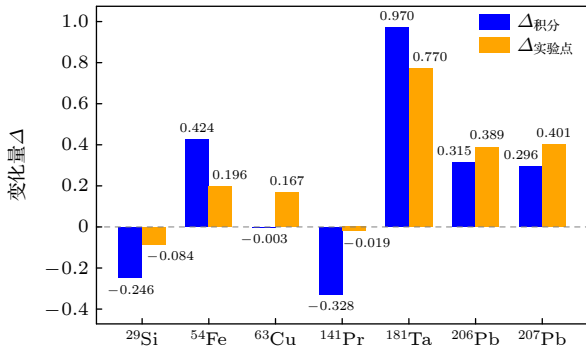


图 6 基于  $^{29}\text{Si}$ ,  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{141}\text{Pr}$ ,  $^{181}\text{Ta}$ ,  $^{206}\text{Pb}$  和  $^{207}\text{Pb}$  的伽马单中子反应道  $\Delta_{\text{积分}}$  和  $\Delta_{\text{实验点}}$  变化量图

Fig. 6. Graph of the amount of change in  $(\gamma, n)$  reaction channel  $\Delta_{\text{积分}}$  and  $\Delta_{\text{实验点}}$  based on  $^{29}\text{Si}$ ,  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{141}\text{Pr}$ ,  $^{181}\text{Ta}$ ,  $^{206}\text{Pb}$ , and  $^{207}\text{Pb}$ .

## 4 结论

本文研究了基于变分自编码器 (VAE) 的伽马单中子出射反应截面的离群点分析技术. 经过对 EXFOR 数据库中  $^{29}\text{Si}$ ,  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{141}\text{Pr}$ ,  $^{181}\text{Ta}$ ,  $^{206}\text{Pb}$  和  $^{207}\text{Pb}$  等 7 个核素的 29 家数据训练, 并结合基于  $k$ -means 方法的潜在变量聚类分析, 有效识别了实验数据离群点, 并且证明排除离群点前后数

据与其对应的 IAEA-2019 评价数据的偏差有效降低, 验证了 VAE 网络在核数据实验信息离群点识别领域的应用潜力. 尽管 VAE 方法在处理数据中显示出一定的优势, 但在数据量较少、公共能区过窄等情况下, 模型训练可能不充分, 导致潜在空间聚类结果不准确, 从而误筛选出有效数据. 后期有望开展裂变核中子共振截面、裂变截面等测量数据丰富反应截面的 VAE 网络应用尝试.

感谢中国原子能科学研究院张英逊研究员与北京大学赵鹏巍教授的热情讨论与帮助.

## 参考文献

- [1] International Atomic Energy Agency 2000 *Handbook on Photonuclear Data for Applications: Cross-sections and Spectra* (Vienna: IAEA) pp1-284
- [2] Obložinský P 2002 *J. Nucl. Sci. Technol.* **39** 31
- [3] Dietrich S S, Berman B L 1988 *At. Data Nucl. Data Tables* **38** 199
- [4] Kawano T, Cho Y, Dimitriou P, et al. 2020 *Nucl. Data Sheets* **163** 109
- [5] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, Zdeborová L 2019 *Rev. Mod. Phys.* **91** 045002
- [6] He W, Li Q, Ma Y, Niu Z, Pei J, Zhang Y 2023 *Sci. China Phys. Mech. Astron.* **66** 282001
- [7] He W B, Ma Y G, Pang L G, Song H C, Zhou K 2023 *Nucl. Sci. Tech.* **34** 88
- [8] Bai J, Niu Z, Sun B, Niu Y 2021 *Phys. Lett. B* **815** 136147
- [9] Xing K, Sun X J, Xu R R, Zou F L, Hu Z H, Wang J M, Tao X, Sun X D, Tian Y, Niu Z M 2024 *Phys. Lett. B* **855** 138825
- [10] Li W, Liu L, Niu Z, Niu Y, Huang X 2024 *Phys. Rev. C* **109** 044616
- [11] Bardhan J, Mandal T, Mitra S, Neeraj C, Patra M 2024 *Eur. Phys. J. Spec. Top.* **233** 2559
- [12] Mitra S, Choi H, Liu S, Glatt R, Wendt K, Schunck N 2024 arXiv: 2404.02332 [nucl-th]
- [13] Fox J M, Wendt K A 2024 arXiv: 2403.16389 [nucl-th]
- [14] Kingma D P 2013 arXiv: 1312.6114 [stat.ML]
- [15] Otuka N, Dupont E, Semkova V, et al. 2014 *Nucl. Data Sheets* **120** 272
- [16] Higgins I, Matthey L, Pal A, Burgess C P, Glorot X, Botvinick M M, Mohamed S, Lerchner A 2017 ICLR (Poster) 3
- [17] Parzen E 1962 *Ann. Math. Stat.* **33** 1065
- [18] Maas A L, Hannun A Y, Ng A Y, et al. 2013 *Proc. ICML* **30** 3
- [19] He K, Zhang X, Ren S, Sun J 2015 *Proceedings of the IEEE International Conference on Computer Vision* Santiago, Chile, December 13-16, 2015 p1026
- [20] Kingma D P 2014 arXiv: 1412.6980 [cs.LG]
- [21] MacQueen J 1967 *Proc. Fifth Berkeley Symp. Math. Stat. Prob.* **1** 281
- [22] Rousseeuw P J 1987 *J. Comput. Appl. Math.* **20** 53
- [23] Fukuda K, Okabe S 1973 *J. Phys. Soc. Jpn.* **34** 315
- [24] Pywell R, Berman B, Kean P, Thompson M 1981 *Nucl. Phys.*

- A 369 141
- [25] McNeill K, Pywell R, Berman B, Woodworth J, Thompson M, Jury J 1987 *Phys. Rev. C* **36** 1621
- [26] Ratner B, Sergiyevsky A, Verbitsky S 1977 *Nucl. Phys. A* **285** 71
- [27] Katz L, Cameron A 1951 *Can. J. Phys.* **29** 518
- [28] Norbury J, Thompson M, Shoda K, Tsubota H 1978 *Aust. J. Phys.* **31** 471
- [29] Plaisir C, Hannachi F, Gobet F, Tarisien M, Aléonard M, Méot V, Gosselin G, Morel P, Morillon B 2012 *Eur. Phys. J. A* **48** 68
- [30] Sund R, Baker M, Kull L, Walton R 1968 *Phys. Rev.* **176** 1366
- [31] Dzhilavyan L, Kucher N 1979 *Sov. J. Nucl. Phys.* **30** 151
- [32] Berman A, Brown K 1954 *Phys. Rev.* **96** 83
- [33] Scott M, Hanson A, Kerst D 1955 *Phys. Rev.* **100** 209
- [34] Owen D, Muirhead E, Spicer B 1968 *Nucl. Phys. A* **122** 177
- [35] Byerly Jr P R, Stephens W 1951 *Phys. Rev.* **83** 54
- [36] Martins M, Hayward E, Lamaze G, Maruyama X, Schima F, Wolyne E 1984 *Phys. Rev. C* **30** 1855
- [37] Sund R, Verbinski V 1970 *Phys. Rev. C* **2** 1129
- [38] Utsunomiya H, Makinaga A, Goko S, et al. 2006 *Phys. Rev. C* **74** 025806
- [39] Cook B, Hutchinson D, Waring R, Bradford J, Johnson R, Griffin J 1966 *Phys. Rev.* **143** 730
- [40] Belyaev S, Semenov V 1991 *Bull. Russ. Acad. Sci. Phys.* **55** 66
- [41] Belyaev S, Kozin A, Nechkin A, Semenov S S, Semenko S 1985 *Yad. Fiz.* **42** 1050
- [42] Utsunomiya H, Akimune H, Goko S, et al 2003 *Phys. Rev. C* **67** 015807
- [43] Goko S, Utsunomiya H, Goriely S, et al 2006 *Phys. Rev. Lett.* **96** 192501
- [44] Kondo T, Utsunomiya H, Goriely S, et al 2012 *Phys. Rev. C* **86** 014316
- [45] Birenbaum Y, Berant Z, Kahane S, Wolf A, Moreh R 1995 *Phys. Rev. C* **51** 3496

# Outliers identification of experimental $(\gamma, n)$ reaction cross section via variational autoencoder\*

XIE Jinchen<sup>1)</sup> TAO Xi<sup>1)</sup> XU Ruirui<sup>1)†</sup> TIAN Yuan<sup>1)</sup>  
 XING Kang<sup>1)</sup> GE Zhigang<sup>1)</sup> NIU Yifei<sup>2)3)‡</sup>

1) (*China Nuclear Data Center, China Institute of Atomic Energy, Beijing 102413, China*)

2) (*School of Nuclear Science and Technology, Lanzhou University, Lanzhou 730000, China*)

3) (*Frontiers Science Center for Rare Isotopes, Lanzhou University, Lanzhou 730000, China*)

( Received 25 December 2024; revised manuscript received 24 January 2025 )

## Abstract

The  $(\gamma, n)$  cross-section is important in nuclear engineering transport calculations. The measurements of the  $(\gamma, n)$  reaction for some isotopes show significant discrepancies among different laboratories. Since the analysis of experimental data is the primary task in the evaluation of nuclear data, identifying the measured outlier data is crucial for improving the quality of nuclear data. Therefore, a variational autoencoder (VAE) method is used in this work to analyze experimental measurements of  $(\gamma, n)$  cross sections for nuclear mass ranging from 29 to 207, in order to provide more reliable experimental information for evaluating nuclear data.

According to the proton number  $Z$  and nuclear mass  $A$ , we design a variational autoencoder network for outlier identification in the measurement of  $(\gamma, n)$ . The silhouette coefficient method and K-means algorithm are used to cluster the latent variables of VAE. Subsequently, the experimental data with and without the outliers are compared with those from the IAEA-2019-PD to assess the VAE in its application to the evaluation of photon-neutron measurements.

The results demonstrate that the VAE can effectively identify outliers in the measurements of  $(\gamma, n)$ . After excluding outliers, the  $(\gamma, n)$  cross-section for  $^{54}\text{Fe}$ ,  $^{63}\text{Cu}$ ,  $^{181}\text{Ta}$ ,  $^{206}\text{Pb}$ , and  $^{207}\text{Pb}$  showed higher consistency with the IAEA-2019-PD evaluation results. However,  $^{29}\text{Si}$  and  $^{141}\text{Pr}$  still deviate from the IAEA-2019-PD evaluation results, therefore requiring more analyses of the measurements themselves in future.

The VAE method can effectively identify outliers and extract the latent structures in experimental data of  $(\gamma, n)$  reaction. It provides more reliable experimental information for evaluating nuclear data and validating the potential application of this method in nuclear data research. However, the generalizability of VAE method still needs further developing, especially in addressing the issues of uneven energy distribution for various measurements.

**Keywords:** variational autoencoder,  $(\gamma, n)$  reaction, cross section, outlier

**PACS:** 25.20.-x

**DOI:** 10.7498/aps.74.20241775

**CSTR:** 32037.14.aps.74.20241775

\* Project supported by the National Key Research and Development Program of China (Grant Nos. 2022YFA1602403, 2021YFA1601500), the National Natural Science Foundation of China (Grant Nos. 12005303, 12075104), the Key Laboratory Key Project, China (Grant No. JCKY2023201C153-5), and the Continuous Support for Basic Scientific Research Project, China (Grant No. BJ010261223282).

† Corresponding author. E-mail: [xuruirui@ciae.ac.cn](mailto:xuruirui@ciae.ac.cn)

‡ Corresponding author. E-mail: [niuyf@lzu.edu.cn](mailto:niuyf@lzu.edu.cn)



基于变分自编码器的伽马单中子出射反应截面实验数据离群点研究

谢金辰 陶曦 续瑞瑞 田源 邢康 葛智刚 牛一斐

**Outliers identification of experimental ( $\gamma$ , n) reaction cross section via variational autoencoder**

XIE Jinchen TAO Xi XU Ruirui TIAN Yuan XING Kang GE Zhigang NIU Yifei

引用信息 Citation: *Acta Physica Sinica*, 74, 082501 (2025) DOI: 10.7498/aps.74.20241775

CSTR: 32037.14.aps.74.20241775

在线阅读 View online: <https://doi.org/10.7498/aps.74.20241775>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

## 您可能感兴趣的其他文章

### Articles you may be interested in

基于位移流U-Net和变分自动编码器的心脏电影磁共振图像左心肌运动追踪

Motion tracking of left myocardium in cardiac cine magnetic resonance image based on displacement flow U-Net and variational autoencoder

物理学报. 2021, 70(22): 228701 <https://doi.org/10.7498/aps.70.20210885>

14 MeV附近 $^{191}\text{Ir}(n,2n)^{190}\text{Ir}$ 反应截面实验研究

Measurement of  $^{191}\text{Ir}(n,2n)^{190}\text{Ir}$  cross section near 14 MeV

物理学报. 2022, 71(19): 192501 <https://doi.org/10.7498/aps.71.20220776>

采用薄靶方法测量低能电子致Al, Ti, Cu, Ag, Au元素K壳层电离截面与L壳层特征X射线产生截面

Measurements of K-shell ionization cross sections and L-shell X-ray production cross sections of Al, Ti, Cu, Ag, and Au thin films by low-energy electron impact

物理学报. 2022, 71(17): 173402 <https://doi.org/10.7498/aps.71.20220162>

中子诱发伽马产生截面测量中的谱分析技术

Spectral analysis techniques in measuring neutron-induced gamma production cross-section

物理学报. 2024, 73(7): 072901 <https://doi.org/10.7498/aps.73.20231980>