arXiv:2303.01844v1 [cs.LO] 3 Mar 2023

学习用于描述逻辑概念的置换不变嵌入

Caglar Demir 和 Axel-Cyrille Ngonga Ngomo

数据科学研究组，帕德博恩大学

摘要。概念学习旨在从背景知识和输入示例中学习描述逻辑概念。其目标是学习一个涵盖所有正例、同时不涵盖任何负例的概念。这一非平凡任务通常被表述为在无限拟序概念空间中的搜索问题。尽管最先进模型已成功应用于解决这一问题，但由于过度探索导致运行时间不切实际，其大规模应用受到严重阻碍。为解决这一局限性，我们在此提出一种补救方法。我们将学习问题重新表述为多标签分类问题，并提出一种神经嵌入模型（NERO），该模型为示例集合学习置换不变嵌入，专门用于预测预先选定的描述逻辑概念的 F_1 分数。通过按照预测分数降序排列这些概念，可以在少量检索操作内检测到一个可能的目标概念，即无需进行过度探索。重要的是，排名靠前的概念可用于在概念空间中的多个有利区域启动最先进符号模型的搜索过程，而不是从最一般概念 $\top$开始。我们在包含770个学习问题的5个基准数据集上进行的实验有力表明，就 F_1 分数、已探索概念数量和总运行时间而言，NERO显著优于最先进模型（p值 $<$ 1%）。我们提供了该方法的开源实现。[1]

关键词：描述逻辑、概念学习、置换不变性

1 引言

基于深度学习的模型已被有效应用于处理各种图相关问题，包括问答、链接预测 [19,30]。然而，它们的预测结果无法由人类解释，并且局限于固定词汇表中的术语 [9,11]。相比之下，描述逻辑（DLs）能够在无限设定下进行可人工解释的推理 [1,16,27]。长期以来，人们一直知道如何为描述逻辑概念推导解释 [5]。例如，可以利用包含层级，将解释推导为一系列二元

[1] https://github.com/dice-group/Nero

分类器，其方式类似于沿着决策树中的路径进行分类 [35,4]。利用描述逻辑被认为是构建可解释人工智能（AI）的潜在基础 [33]。尽管描述逻辑已成为形式化知识库（KB）的标准技术 [19,20,29], ，知识库的不完整性以及符号模型不切实际的运行时间一直是实现其潜力的挑战。当前最先进的概念学习（CL）模型已成功应用于从知识库和输入示例中学习描述逻辑概念 [24,27]。然而，其不切实际的运行时间严重阻碍了实际应用。这一局限源于对短视启发式函数的依赖，该函数往往会导致对概念进行过度探索 [16,22,35]。描述逻辑概念的探索包括检索其个体，并根据输入知识库和示例计算其质量（见第 2 节）。随着输入知识库规模的增长，过度探索已成为实际应用中的计算瓶颈。在此，我们针对这一局限提出一种解决方案。我们将学习问题重新表述为多标签分类问题，并提出 NERO——一种神经置换不变嵌入模型。给定一组正例 E^+ 和一组负例 E^-，NERO 会预测预先选定的描述逻辑概念的 F_1 得分，如图 1 所示。

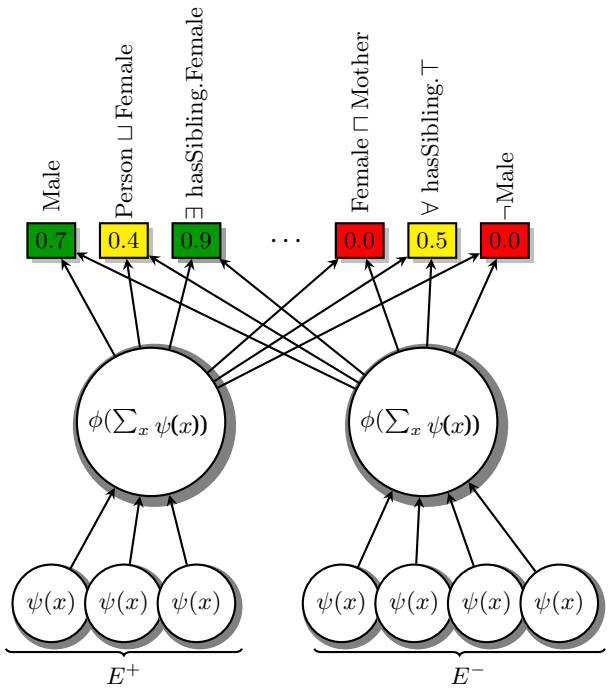


图 1。 NERO的可视化。方框和值分别表示预先选定的唯一描述逻辑概念及其预测的 F_1 分数。

通过按预测分数降序排列预先选定的描述逻辑概念，只需探索排名靠前的少数概念，就能找到目标概念。重要的是，如果未找到目标概念，排名靠前的概念可用于初始化最先进模型的标准搜索过程。这样，最先进的概念学习模型便能够从更有利的状态开始搜索，而不是从最一般概念 $\top$ 开始。我们在 5 个基准数据集上的 770 个学习问题上进行的实验表明，NERO 在 F_1 分数、已探索概念数量和总运行时间等标准指标上显著优于最先进模型（p值为 $< 1\%$）。重要的是，在 NERO 中配备最先进模型（CELOE），还能进一步提升其在基准数据集上的 F_1 分数，且运行时间成本较低。Wilcoxon符号秩检验的结果证实，NERO 的优越性能具有显著性。我们提供了 NERO 的开源实现，其中包括预训练模型、评估脚本以及网络服务。[2]

2 背景

知识库：知识库（KB）是一个二元组 $\mathcal{K} = (\mathcal{T}box, \mathcal{A}box)$，其中$\mathcal{T}box$ 是一个术语公理集合，用于描述命名概念 N_C [33]之间的关系。术语公理的形式为 $A \sqsubseteq B$ 或 $A \equiv B$，其中 A、$B \in N_C$。$\mathcal{A}box$ 是一个断言集合，用于描述个体a、$b \in N_I$ 之间通过角色 $r \in N_R$ 建立的关系，以及N_I 与 N_C之间的概念成员关系。$\mathcal{A}box$ 中的每条断言都必须采用 $A(x)$ 和 $r(x, y)$ 的形式，其中$A \in N_C$、$r \in N_R$，且 x、$y \in N_I$。示例见图 2。

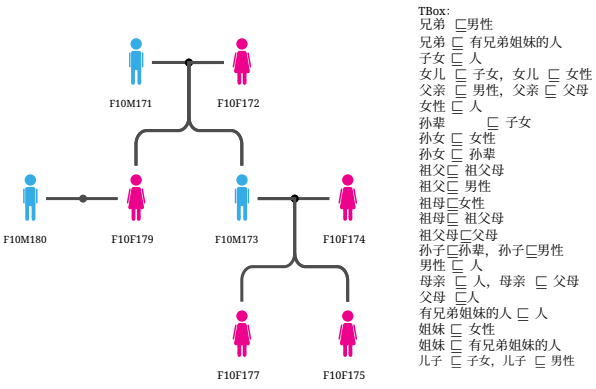


图 2. 家庭知识库的可视化，其中包含 $\mathcal{T}box$ 以及 $\mathcal{A}box$的一个子集。颜色表示概念断言，而（·）和从（·）分出的分支分别表示角色断言。

[2] https://github.com/dice-group/Nero

描述逻辑：描述逻辑（DLs）是一阶谓词逻辑的片段，仅使用一元谓词和二元谓词。一元谓词、二元谓词和常量分别称为概念、角色和个体 [1]。描述逻辑已成为许多应用领域中形式化背景知识的标准技术，包括语义网 [20,29]。利用基于描述逻辑定义的知识库具有成为可解释人工智能基石的潜力 [33]。与许多其他研究（见第 3 节）一样，本文考虑描述逻辑 $\mathcal{ALC}$ （带补集的属性语言） [1] 中的知识库。 $\mathcal{ALC}$ 的模型论语义见表 1。

表 1。 $\mathcal{ALC}$ 语法与语义。 $\mathcal{I}$ 表示一种解释， $\Delta^{\mathcal{I}}$ 表示其论域。

构造	语法	语义
原子概念	A	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
Role	r	$r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
顶概念	$\top$	$\Delta^{\mathcal{I}}$
底概念	$\bot$	$\emptyset$
合取	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
析取	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
否定	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
存在限制	$\exists\, r.C$	$\{x \mid \exists\, y.(x,y) \in r^{\mathcal{I}}$ 和 $y \in C^{\mathcal{I}}\}$
全称限制	$\forall\, r.C$	$\{x \mid \forall\, y.(x,y) \in r^{\mathcal{I}}$ 蕴含 $y \in C^{\mathcal{I}}\}$

概念学习：给定 $\mathcal{K}$ 上的 $\mathcal{ALC}$、正例集合 $E^+ \subset N_I$ 以及负例集合 $E^- \subset N_I$ 。描述逻辑概念学习问题定义如下

$$\forall p \in E^+, \forall n \in E^- \big(\mathcal{K} \models \mathtt{H}(p)) \wedge (\mathcal{K} \not\models \mathtt{H}(n)\big), \tag{1}$$

其中 $\mathtt{H} \in \mathcal{C}$ 表示一个 $\mathcal{ALC}$ 概念，而 $\mathcal{C}$ 表示依据构造规则所构成的所有有效 $\mathcal{ALC}$ 概念： $\mathtt{C} ::= \mathtt{A} \mid \neg \mathtt{C} \mid \mathtt{C} \sqcap \mathtt{C} \mid \mathtt{C} \sqcup \mathtt{C} \mid \exists r.\mathtt{C} \mid \forall r.\mathtt{C} \mid$，其中 $\mathtt{A} \in N_C$和 $r \in N_R$。 $\mathcal{K} \models \mathtt{H}(p)$ 意味着，关于类别成员关系 $\mathtt{H}(p)$ 的推理是 $\mathcal{K}$的逻辑后果。判断 H 是否满足方程（1），需通过检索函数 $\mathcal{R}: \ \mathcal{C} \rightarrow 2^{N_I}$ ，该函数基于开放世界假设（OWA）或封闭世界假设（CWA）定义。这一非平凡的学习问题通常会转化为准序概念空间 $(\mathcal{S}, \preceq)$ [7,13,27,34]中的搜索问题。遍历 $\mathcal{S}$ 通常通过自顶向下细化算子来进行，其定义为 $\rho: \ \mathcal{S} \rightarrow 2^{\mathcal{S}}$ ，其中

$$\forall \mathtt{A} \in \mathcal{S} : \rho(\mathtt{A}) \subseteq \{\mathtt{B} \in \mathcal{S} \mid \mathtt{B} \preceq \mathtt{A}\}. \tag{2}$$

最先进的概念学习模型在以最一般描述逻辑概念（$\top$）作为根节点初始化搜索树后，开始朝着 H 进行搜索。这棵搜索树

通过迭代构建：选择包含启发式值最高的准序描述逻辑概念的节点，并将其符合条件的细化作为子节点添加到搜索树 [27] 中。

启发式方法：启发式函数是在 $\mathcal{S}$ 中高效搜索 H [26] 的关键。已探索概念的数量和运行时间被用作效率的代理指标。人们研究了各种启发式函数 [26,35]。最先进模型中的大多数启发式函数都可以视为短视函数，倾向于偏好语法上较短且准确的概念。因此，它们容易陷入局部最优 [35]。例如，CELOE 的启发式函数定义为

$$\phi_{\mathrm{CELOE}}(\mathtt{A},\mathtt{B}) = \mathrm{Q}(\mathtt{B}) + \lambda \cdot \big[\mathrm{Q}(\mathtt{B}) - \mathrm{Q}(\mathtt{A})\big] - \beta \cdot |\mathtt{B}|, \tag{3}$$

其中 A $\in \mathcal{S}$，B $\in \rho$(A)。 $\beta > \lambda \geq 0$ ，Q(·) 表示质量函数（例如 F_1分数或准确率）。通过 Q(·) 和 $|\cdot|$，搜索仅依据 A 和 B，引导至更准确且语法上更短的概念。 F_1(·) 定义为

$$F_1(\mathtt{A}) = \frac{|\ E^+ \cap \mathcal{R}(\mathtt{A})\ |}{|\ E^+ \cap \mathcal{R}(\mathtt{A})\ | + 0.5(|\ E^- \cap \mathcal{R}(\mathtt{A})\ | + |\ E^+ \setminus \mathcal{R}(\mathtt{A})\ |)}. \tag{4}$$

随着知识库规模的增长，执行检索操作 $\mathcal{R}$(·) 的运行时间会增加 [3,4,25]。因此，在 $\mathcal{S}$ 中遍历会成为计算瓶颈。由此，减少已探索概念的数量对于解决知识库上的概念学习至关重要。尽管最先进模型（例如 CELOE）采用冗余消除和表达式简化规则来减少已探索概念的数量，但最先进模型过长的运行时间仍然阻碍了大规模应用 [17]。此外，解决大型知识库上的概念学习问题时， $\mathcal{R}$() 所依据的假设也会发挥作用。由于知识库具有不完整性，OWA 似乎是更合适的假设 [31]。然而，采用 OWA 往往会使成员查询在计算上更具挑战性 [12,26]。因此，许多近期研究通常采用 CWA [16,22,34]。

3 相关工作

大量研究探讨了如何根据知识库和输入示例学习描述逻辑概念。有关介绍，请参阅 [1,18,23] 。大多数符号系统在启发式函数的使用和细化算子的设计方面有所不同 [2,7,12,14,26,27,21,34]。DL-Learner [24] 被视为用于概念学习的最成熟且最新的系统 [32]。DL-Learner包含多个最先进模型，包括ELTL、OCEL和CELOE。ELTL基于描述逻辑 $\mathcal{EL}$ 的细化算子，并使用偏好语法较短概念的启发式函数。CELOE以OCEL和ELTL为基础，并采用更复杂的启发式函数。目前，CELOE是DL-Learner中表现最佳的概念学习模型，其性能往往优于包括OCEL和ELTL在内的许多最先进模型

在学习表达式的质量、已探索概念的数量和运行时间方面 [35,28]。上述方法采用冗余消除和表达式简化规则，以减少已探索概念的数量。尽管应用冗余消除和表达式简化规则通常能减少已探索概念的数量，但这些操作会增加计算开销，而较长的运行时间仍然阻碍了大规模应用 [17]。近期的大多数研究都致力于解决概念学习中不切实际的运行时间问题。CLIP [22]是一种神经方法，可作为基于细化方法的补充，并通过预测可能目标状态的长度来支持剪枝搜索空间。EvoLearner [16] 将概念表示为抽象语法树，对应于进化算法中的一个个体。初始个体群体通过从 E^+出发的偏置随机游走获得。West-phal等人 [35] 设计了一种基于模拟退火的元启发式方法，以平衡搜索过程中的探索与利用。在本研究中，我们主要将 NERO 与DL-Learner提供的CELOE进行比较，原因有二：（1）DL-Learner被视为用于概念学习的最成熟且最新的系统 [32]；（2）近期开发的模型通常根据概念质量和运行时间进行评估。然而，不报告已探索概念的数量，就无法量化 NERO带来的潜在改进究竟源于我们的新颖思路，还是高效实现。因此，在实验中，我们主要将 NERO 与CELOE进行比较，比较指标包括已探索概念的数量、学习所得概念的质量以及运行时间。

4 方法论

动机：概念学习问题的目标是找到一个最大化公式（4）的描述逻辑概念 $\mathrm{H} \in \mathcal{C}$。在此，我们希望通过学习专为预测预先选定概念的 F_1 分数而设计的置换不变嵌入来实现这一目标。首先探索排名靠前的概念，旨在通过少量检索操作找到目标概念。如果在排名靠前的概念中未找到目标状态，则可使用排名靠前的概念及其对应的启发式值初始化先进的概念学习模型的搜索树。这样，标准搜索过程便可从比最一般概念 $\top$ 更有利的状态开始。

方法：公式（4）表明，$F_1(\cdot)$ 对 E^+、E^- 和 $\mathcal{R}(\cdot)$ 中个体的顺序具有不变性。此前，Zaheer 等人 [36] 已证明，所有对输入顺序具有不变性的函数都可以分解为

$$f(\mathbf{x}) = \phi\Big(\sum_{x \in \mathbf{x}} \psi(x)\Big), \tag{5}$$

其中，$\mathbf{x} = \{x_1, \ldots, x_m\} \in 2^{\mathcal{X}}$and $\phi(\cdot)$ 和 $\psi(\cdot)$ 分别表示一个输入集合和两个参数化连续函数。通过公式 (5) 定义的置换不变神经网络仍然满足通用逼近

定理 [36]。我们推测，这种神经网络能够学习个体集合（例如 E^+ and E^-）的置换不变嵌入，以便预测预先选定概念的 F_1 分数。通过准确预测预先选定描述逻辑概念的 F_1分数，可以从预先选定的概念中检测出可能的目标概念，而无需使用 $F_1(\cdot)$ 和 $\mathcal{R}(\cdot)$。基于这些考虑，我们将 NERO 定义如下

$$\text{NERO}(E^+, E^-) = \sigma\Big(\phi\Big(\sum_{x\in E^+}\psi(x)\Big) - \phi\Big(\sum_{x\in E^-}\psi(x)\Big)\Big), \tag{6}$$

其中，$\psi(\cdot): N_I \to \mathbb{R}^m$ 和 $\phi: \mathbb{R}^m \to [0,1]^{|\mathcal{T}|}$ 分别表示嵌入查找操作和仿射变换。$\mathcal{T}$ 表示预先选定的描述逻辑概念。平移操作的结果记为 $\mathbf{z} \in \mathbb{R}^m$，并通过逻辑 Sigmoid 函数 $\sigma(\mathbf{z}) = \frac{1}{1+\exp(-\mathbf{z})}$进行归一化。因此，NERO：$2^{N_I} \times 2^{N_I} \mapsto [0,1]^{|\mathcal{T}|}$ 可视为将两个个体集合映射到 $|\mathcal{T}|$ 单位区间的函数。NERO 可视为一种多任务学习方法，它利用多个任务之间的相似性；在我们的案例中，一个任务对应于准确预测预先选定的描述逻辑概念 F_1 的分数 [8]。

针对相关任务量身定制表示的重要性已得到充分研究 [15,8]。受此启发，我们在算法 1 中阐明了选择描述逻辑概念的过程。我们选择其规范解释并非完全重叠的概念（见第 4 行）。如其中所示，只要给定 $\mathcal{R}(\cdot)$ 和$\rho(\cdot)$，便可在任何描述逻辑上定义的知识库上训练 NERO。

算法 1 构建目标描述逻辑概念

输入：$\mathcal{R}(\cdot)$ $\rho(\cdot)$ d，最大长度 输出：$\mathcal{T}$

1: $\mathcal{T} := \{\mathsf{C} \mid \mathsf{C} \in \rho(\top) \wedge |\mathsf{C}| \leq$ 最大长度 $\wedge\, 0 < |\mathcal{R}(\mathsf{C})|\}$
2: 对每个 $\mathsf{A} \in \mathcal{T}$执行
3: 对每个 $\mathsf{B} \in \mathcal{T}$执行
4: 如果 $\mathcal{R}(\mathsf{A}) \neq \mathcal{R}(\mathsf{B})$，则
5: 对每个 $\mathsf{X} \in \{\mathsf{A} \sqcap \mathsf{B}, \mathsf{A} \sqcup \mathsf{B}\}$ 执行
6: 如果 $|\mathcal{R}(\mathsf{X})| > 0 \wedge \mathcal{R}(\mathsf{X}) \notin \{\mathcal{R}(\mathsf{E}) \mid \mathsf{E} \in \mathcal{T}\}$ 那么
7: 将 X 添加到 $\mathcal{T}$。
8: 结束如果
9: 如果 $|\mathcal{T}| = d$ 那么
10: 返回 $\mathcal{T}$
11: 结束如果
12: 结束循环
13: 结束条件判断
14: 结束循环
15: 结束循环
16: 如果 $|\mathcal{T}| < d$ 那么
17: 转到步骤 (2)。
18: 结束 if

训练过程：令 $\mathcal{D} = \{(E_i^+, E_i^-, \mathrm{y}_i)\}_{i=1}^N$ 表示一个训练数据集，其中一个数据点 (E^+, E^-, y) 通过以下连续四个步骤获得：(i) 从 $\mathcal{T}$ 中均匀随机采样 C；(ii) 从 $E^+ \subset \mathcal{R}(\mathrm{C})$ 中均匀随机采样 k 个个体；(iii) 从 $E^- \subset N_I \setminus E^+$ 中均匀随机采样 k 个个体；(iv) 根据公式（4），针对 E^+, E^- 计算 F_1 分数 y，其中 $\mathcal{T}$。对于给定的 (E^+, E^-, y) 和预测值 $\hat{\mathrm{y}} := \textsc{NERO}(E^+, E^-)$，可计算相应的二元交叉熵损失。需要注意的是，训练过程结束后，无需在 NERO 中执行平移操作，即可直接获得任意 $\mathcal{ALC}$ 描述逻辑概念的置换不变嵌入。例如，描述逻辑概念（如 Male⊓∃hasSibling.Female）的嵌入可通过 $\phi\big(\sum_{x \in \mathcal{R}(\mathrm{Male} \sqcap \exists \mathrm{hasSibling.Female})} \psi(x)\big)$ 获得。在我们的项目页面上，我们提供了家庭知识库的学习到的嵌入的二维可视化。

5 实验

我们的实验设置基于 [6,26,7]，并使用了其中提供的学习问题。表 2 概述了这些数据集。为了对模型进行全面比较，我们通过随机抽样 E^+ 和 E^- 生成了额外的学习问题。我们确保评估中使用的所有学习问题都未在无监督训练阶段中使用。在实验中，我们在相同的硬件上评估了 $\mathcal{ALC}$ 中用于类表达式学习（CEL）的所有模型。

表 2. 类表达式学习基准数据集概览。

数据集	$\lvert N_I \rvert$	$\lvert N_C \rvert$	$\lvert N_R \rvert$
Family	202	18	4
致癌作用	22372	142	21
诱变	14145	86	11
Biopax	323	28	49
淋巴造影	148	49	1

我们通过 F_1 分数、运行时间和已探索概念数量来评估模型。F_1 分数用于衡量所发现概念相对于正例和负例的质量，而运行时间和已探索概念数量则用于衡量效率。我们测量了完整的计算时间，其中包括输入数据预处理以及解决学习问题所花费的时间。此外，我们对最先进的模型采用了两种标准的停止准则。(i) 我们将最长运行时间设为 10 秒，尽管模型通常会在 1.5 秒内找到良好的解 [27]。(ii) 模型被配置为一旦找到目标概念便立即终止。在我们的实验中，我们在相同硬件上评估了 $\mathcal{ALC}$ 中用于概念学习（CL）的所有模型。训练期间，我们设置了 $|\mathcal{T}| = 1000$、$N = 50$，并在 NERO中使用 Adam 优化器。我们仅考虑排名前 100 的概念来评估 NERO。

6 结果

基准学习问题的结果：表3报告了基准学习问题上的概念学习结果。表3表明，为 NeRo 配备标准搜索过程，能够进一步提升其在 F_1 分数方面的最先进性能，而运行时间仅略有增加。CELOE 和 ELTL 在 Family 上找到准确概念所需的时间至少比 NeRo 多 $\mathbf{14.7}\times$。这是因为 NeRo 平均只探索 21 个概念，而 CELOE 探索了 1429 个。在 Mutagenesis 和 Carcinogenesis 上，NeRo 找到的概念更准确，同时探索的概念更少，因此运行时间表现更好。运行时间的提升源于 NeRo 至少少探索了 $\mathbf{2.3}\times$ 个概念。

表3. 基准学习问题的结果。F_1、T 和实验分别表示 F_1 分数、以秒为单位的总运行时间以及已探索概念的数量。NeRo† 表示为 NeRo 配备 CELOE。ELTL 不报告实验。

数据集	NeRo†			**NeRo**			CELOE			**ELTL**	
	F_1	T	实验	F_1	T	实验	F_1	T	实验	F_1	T
Family	.987	.83	26	.984	**.68**	**21**	.980	4.65	1429	.964	4.12
诱变	.714	17.30	200	.704	13.18	100	.704	23.05	516	.704	21.04
致癌作用	.725	32.23	200	.720	26.26	100	.714	37.18	230	.719	36.29

需要特别指出的是，我们没有在 NeRo 中使用并行处理，并且针对每个单独的学习问题，都会重新加载 NeRo 的参数。为了进行更全面的评估，我们在五个基准数据集上总共生成了 750 个随机学习问题。由于淋巴造影和 Biopax 数据集中不包含任何学习问题，因此它们未列入表3。

随机学习问题的结果：表4报告了随机学习问题上的概念学习结果。表4表明，CELOE 探索的概念至少比 NeRo多 $\mathbf{3.19}\times$ 。重要的是，NeRo 找到的概念准确度不逊于甚至高于其他方法，同时探索的概念更少。这里，我们对每个数据集仅加载一次 NeRo 的参数，并依次用于解决学习问题。这使 NeRo在 Family、诱变和致癌作用基准数据集上的总计算时间减少了 $\mathbf{3}-\mathbf{6}\times$ 。尽管 NeRo 可以并行解决学习问题（例如通过多进程），但我们没有使用任何并行处理，因为 CELOE 和 ELTL 不支持并行处理 [7]。以标准小批量方式加载学习问题并使用多GPU，或许能进一步缩短 NeRo的运行时间。这些结果表明，对于要求低延迟的应用，NeRo 可能比 CELOE 和 ELTL 更合适。

表4. 每个基准数据集不同规模的随机学习问题。每行报告在50个学习问题中取得的均值和标准差。 $|E|$ 表示 $|E^+| + |E^-|$。

数据集	$\|E\|$	**NeRo**			CELOE			**ELTL**	
		F_1	T	Exp.	F_1	T	Exp.	F_1	T
Family	10	**.913 ± .06**	**.16 ± .51**	**74 ± 43**	.903 ± .06	11.61 ± 3.58	5581 ± 2375	.718 ± .01	4.45 ± 2.84
	20	**.807 ± .04**	**.16 ± .49**	**100 ± 00**	.795 ± .05	13.28 ± 1.47	7586 ± 645	.678 ± .02	3.59 ± 1.27
	30	**.775 ± .03**	**.15 ± .41**	**100 ± 00**	.760 ± .03	13.24 ± 1.42	7671 ± 575	.672 ± .01	3.46 ± 1.59
淋巴造影	10	**.968 ± .07**	**.12 ± .43**	**75 ± 41**	.968 ± .07	6.63 ± 4.29	5546 ± 5169	.733 ± .09	3.07 ± .30
	20	**.828 ± .04**	**.13 ± .40**	**100 ± 00**	.826 ± .05	13.01 ± 1.23	11910 ± 1813	.678 ± .02	3.08 ± .50
	30	**.780 ± .04**	**.13 ± .01**	**100 ± 00**	.780 ± .04	13.02 ± 1.69	13138 ± 2601	.672 ± .01	3.09 ± .72
Biopax	10	**.859 ± .08**	**.19 ± .71**	**86 ± 34**	.806 ± .07	13.26 ± 1.94	4752 ± 2153	.685 ± .06	3.71 ± .10
	20	**.793 ± .05**	**.19 ± .52**	**100 ± 00**	.746 ± .04	13.63 ± .10	4151 ± 748	.668 ± .06	3.72 ± .10
	30	**.749 ± .03**	**.18 ± .52**	**100 ± 00**	.718 ± .02	13.91 ± .44	3843 ± 963	.668 ± .06	3.90 ± .22
诱变	10	**777 ± .05**	**3.47 ± 1.61**	**100 ± 00**	.753 ± .06	20.27 ± 1.39	546 ± 613	.670 ± .02	10.29 ± .40
	20	**.746 ± .05**	**3.09 ± 1.75**	**100 ± 00**	.712 ± .02	20.38 ± 1.30	430 ± 28	.667 ± .00	10.73 ± 1.10
	30	**.721 ± .03**	**2.89 ± 1.60**	**100 ± 00**	.700 ± .02	20.39 ± 1.06	429 ± 38	.667 ± .00	11.74 ± .97
致癌作用	10	**.768 ± .06**	**5.39 ± 2.98**	**98 ± 14**	.764 ± .06	29.90 ± 1.02	401 ± 125	.673 ± .05	19.99 ± .67
	20	**.722 ± .03**	**5.40 ± 1.87**	**100 ± 00**	.713 ± .02	30.30 ± .19	318 ± 152	.667 ± .00	20.00 ± 1.11
	30	**.704 ± .05**	**4.70 ± 2.78**	**100 ± 00**	.697 ± .02	29.99 ± .58	319 ± 43	.667 ± .00	20.38 ± .85

表5。不同数量已探索概念下的性能比较。每行报告在50个学习问题中取得的均值和标准差。

数据集	$\|E\|$	NeRo-1		NeRo-10		NeRo-1000	
		F_1	T	F_1	T	F_1	T
Family	10	.906 ± .07	**.08 ± .06**	.910 ± .05	.09 ± .06	**.916 ± .06**	.81 ± .50
	20	.793 ± .05	**.08 ± .05**	.806 ± .04	.09 ± .05	**.807 ± .04**	1.17 ± .50
	30	.742 ± .05	**.08 ± .05**	.773 ± .03	.09 ± .05	**.775 ± .03**	1.15 ± .50
淋巴造影	10	.882 ± .07	**.08 ± .06**	.905 ± .05	.08 ± .06	**.916 ± .06**	.77 ± .50
	20	.793 ± .05	**.07 ± .05**	.827 ± .04	.08 ± .05	**.828 ± .04**	1.03 ± .50
	30	.738 ± .05	**.07 ± .06**	.777 ± .04	.08 ± .05	**.780 ± .03**	1.00 ± .60
Biopax	10	.853 ± .08	**.09 ± .06**	.856 ± .05	.97 ± .59	**.868 ± .08**	1.31 ± .80
	20	.779 ± .05	**.09 ± .06**	.791 ± .04	.10 ± .62	**.793 ± .04**	1.35 ± .60
	30	.708 ± .07	**.09 ± .06**	.742 ± .04	.10 ± .63	**.749 ± .03**	1.39 ± .60
诱变	10	.733 ± .07	**.32 ± 2.03**	.785 ± .06	.57 ± 1.81	**.803 ± .06**	36.98 ± 5.81
	20	.689 ± .08	**.31 ± 1.99**	.734 ± .05	.52 ± 1.82	**.751 ± .04**	34.29 ± 5.91
	30	.673 ± .08	**.31 ± 2.03**	.712 ± .04	.49 ± 1.77	**.728 ± .03**	32.88 ± 6.13
致癌作用	10	.717 ± .09	**.41 ± 2.53**	.740 ± .09	.89 ± 2.89	**.783 ± .02**	56.941 ± 9.55
	20	.680 ± .06	**.40 ± 2.49**	.707 ± .05	.82 ± 2.45	**.731 ± .02**	57.205 ± 5.08
	30	.610 ± .11	**.41 ± 2.83**	.671 ± .06	.77 ± 2.66	**.716 ± .02**	52.872 ± 7.58

有限探索下的结果：表5报告了在五个基准数据集上有限探索条件下的概念学习结果。表5显示，即使仅探索排名前10的概念，NERO-10在所有指标上通常也优于CELOE和ELTL（见表4）。

显著性检验：为验证结果的显著性，我们针对 F_1 得分、运行时间和已探索概念数量进行了Wilcoxon符号秩检验（单侧和双侧）。我们的零假设是，NERO 和CELOE的性能来自相同的分布。在所有数据集上，我们均能以 $< 1\%$的p值拒绝零假设，因此， NERO 的优越性能在统计上具有显著性。

6.1 讨论

我们的结果支持了这一假设：通过学习个体集合的置换不变嵌入，可以准确预测描述逻辑概念的 F_1 得分。首先考虑排名靠前的描述逻辑概念，可以减少为找到目标概念而执行大量检索操作的需求。在我们的实验中，就 F_1 得分、已探索概念数量和总计算时间而言， NERO 始终优于最先进的模型。更重要的是，在这些排名靠前的概念上启动标准搜索过程，能够进一步改善结果。因此，NERO 可应用于最先进的模型，以缩短其运行时间。不过，需要指出的是，Lehmann等人 [26] 此前已证明，CELOE在概念学习问题中具有完备性；也就是说，对于给定的学习问题，只要时间和内存需求没有上限，CELOE就能在目标表达式存在时找到它。尽管这些需求在实践中并不可行，但要使概念学习具备完备性，就必须为 NERO 配备CELOE搜索过程。

7 结论

我们提出了一种置换不变的神经嵌入模型（NERO），以高效解决描述逻辑概念学习问题。对于给定的学习问题， NERO 以多标签分类的方式准确预测预先选定的描述逻辑概念的 F_1 分数。通过按预测的 F_1 分数降序排列概念，只需少量检索操作即可学习到目标概念。实验表明，在5个基准数据集上的770个概念学习问题中，就预测质量、已探索概念数量和总计算时间而言， NERO 优于最先进的模型。为 NERO 配备标准搜索过程后，所有学习问题和基准数据集上的 F_1 分数均进一步提升。

我们认为，在概念学习问题中引入神经模型值得进一步探索。未来，我们将致力于在 NERO 更具表达力的描述逻辑上开展研究，并将非短视启发式方法 [10]中的概念嵌入整合起来。

致谢

本研究得到欧盟 Horizon Europe 研究与创新计划（GA 编号 101070305）、北莱茵-威斯特法伦州文化与科学部 תרגסמב SAIL 项目（GA 编号 NW21-059D）以及德国科学基金会（GA 编号 TRR 318/1 2021–438445824）的支持。

参考文献

1. Baader, F., Calvanese, D., McGuinness, D., Patel-Schneider, P., Nardi, D., 等：描述逻辑手册：理论、实现与应用。剑桥大学出版社（2003） 2. Badea, L., Nienhuys-Cheng, S.H.：描述逻辑的细化算子。载于：ILP（2000） 3. Bin, S., B¨uhmann, L., Lehmann, J., Ngomo, A.C.N.：面向基于 SPARQL 的大规模资源描述框架数据集归纳。载于：ECAI（2016） 4. Bin, S., Westphal, P., Lehmann, J., Ngonga, A.：在 SAKE 项目中实现面向大数据的可扩展结构化机器学习。载于：2017 IEEE 国际大数据会议（Big Data），第 1400–1407 页。IEEE（2017） 5. Borgida, A., Franconi, E., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F.：解释 ALC 包含关系。载于：ECAI，第 209–213 页（2000） 6. B¨uhmann, L., Lehmann, J., Westphal, P.：DL-Learner——语义网上归纳学习的框架。《网络语义学杂志》39，15–24（2016） 7. B¨uhmann, L., Lehmann, J., Westphal, P., Bin, S.：DL-Learner：语义网数据上的结构化机器学习。载于：2018年万维网大会伴随论文集，第 467–471 页。WWW ’18，国际万维网会议指导委员会，瑞士日内瓦共和国和州（2018） 8. Caruana, R.：多任务学习。Springer（1998） 9. Demir, C., Moussallem, D., Heindorf, S., Ngomo, A.C.N.：用于链接预测的卷积超复数嵌入。载于：亚洲机器学习会议，第 656–671 页。PMLR（2021） 10. Demir, C., Ngomo, A.：Drill——用于 ALC 细化算子的深度强化学习。CoRR abs/2106.15373 223, 224（2021） 11. Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S.：卷积二维知识图谱嵌入。载于：AAAI（2018） 12. Fanizzi, N., d’Amato, C., Esposito, F.：描述逻辑中的 DL-FOIL 概念学习。载于：国际归纳逻辑编程会议，第 107–121 页。Springer（2008） 13. Fanizzi, N., Rizzo, G., d’Amato, C.：提升描述逻辑概念学习器。载于：欧洲语义网会议，第 68–83 页。Springer（2019） 14. Fanizzi, N., Rizzo, G., d’Amato, C., Esposito, F.：DLfoil：重新审视类表达式学习。载于：欧洲知识获取研讨会，第 98–113 页。Springer（2018） 15. Goller, C., Kuchler, A.：通过结构反向传播学习任务相关的分布式表示。载于：国际神经网络会议论文集（ICNN’96），第 1 卷，第 347–352 页。IEEE（1996）

16. Heindorf, S., Bl¨ubaum, L., D¨usterhus, N., Werner, T., Golani Nandkumar, V., Demir, C., Ngonga Ngomo, A.C.：EvoLearner：使用进化算法学习描述逻辑。载于：万维网。ACM（2022） 17. Hitzler, P., Bianchi, F., Ebrahimi, M., Sarker, M.K.：神经符号集成与语义网。《语义网》11(1)，3–11（2020） 18. Hitzler, P., Krotzsch, M., Rudolph, S.：语义网技术基础（2009） 19. Hogan, A., Blomqvist, E., Cochez, M., d' Amato, C., Melo, G.d., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., 等：知识图谱。《ACM计算调查》（CSUR）54(4)，1–37（2021） 20. Horrocks, I., Patel-Schneider, P.F., Van Harmelen, F.：从 SHIQ 和资源描述框架到网络本体语言：网络本体语言的形成。《网络语义学杂志》1(1)，7–26（2003） 21. Iannone, L., Palmisano, I., Fanizzi, N.：一种基于反事实的语义网概念学习算法。《应用智能》26(2)，139–159（2007） 22. Kouagou, N., Heindorf, S., Demir, C., Ngomo, A.C.N.：学习概念长度可加速 ALC 中的概念学习。载于：第十九届扩展语义网会议——研究方向。Springer（2022） 23. Kr¨otzsch, M., Simancik, F., Horrocks, I.：描述逻辑入门。arXiv 预印本 arXiv:1201.4089（2012） 24. Lehmann, J.：DL-Learner：学习描述逻辑中的概念。《机器学习研究杂志》10，2639–2642（2009） 25. Lehmann, J.：学习 OWL 类表达式，第 22 卷。IOS 出版社（ 2010 ） 26. Lehmann, J., Auer, S., B¨uhmann, L., Tramp, S.：用于本体工程的类表达式学习。《网络语义学杂志》9(1)，71–81（2011） 27. Lehmann, J., Hitzler, P.：使用细化算子进行描述逻辑中的概念学习。《机器学习》78(1-2)，203（2010） 28. Lehmann, J., 等：DL-Learner手册（2016） 29. Michel, F., Turhan, A.Y., Zarrieß, B.：利用值限制进行高效 TBox 推理——介绍 Flower推理器。载于：第三届规则与推理联合国际会议论文集（RuleML+ RR 2019）。《计算机科学讲义》，意大利博尔扎诺。Springer（2019） 30. Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E.：知识图谱关系机器学习综述。《IEEE会刊》104(1)，11–33（2015） 31. Rudolph, S.：描述逻辑基础。载于：推理网国际暑期学校，第 76–136 页。Springer（2011） 32. Sarker, M.K., Hitzler, P.：描述逻辑的高效概念归纳。载于：AAAI人工智能大会论文集，第 33 卷，第 3036–3043 页（2019） 33. Schockaert, S., Ibanez-Garcia, Y., Gutierrez-Basulto, V.：一种用于类比推理的描述逻辑。载于：IJCAI-21论文集，第 2040–2046 页（2021年8月） 34. Tran, A.C., Dietrich, J., Guesgen, H.W., Marsland, S.：并行对称类表达式学习。《机器学习研究杂志》18，64:1–64:34（2017） 35. Westphal, P., Vahdati, S., Lehmann, J.：一种用于描述逻辑概念学习的模拟退火元启发式。载于：国际归纳逻辑编程会议，第 266–281 页。Springer（2021） 36. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., Smola, A.J.：深度集合。载于：《神经信息处理系统进展》，第 30 卷。Curran出版社（2017）