arXiv:2305.03503v1 [cs.CL] 2 May 2023

理性思考所见内容：用于关系抽取的连续理由提取

Xuming Hu Zhaochen Hong Chenwei Zhang 清华大学 清华大学 亚马逊
hxm19@mails.tsinghua.edu.cn hongzc20@mails.tsinghua.edu.cn cwzhang910@gmail.com

Irwin King 香港中文大学
king@cse.cuhk.edu.hk

Philip S. Yu 伊利诺伊大学芝加哥分校
psyu@cs.uic.edu

摘要

关系抽取（RE）旨在根据两个实体的上下文提取潜在关系，因此，从句子中推导出合理的上下文至关重要。以往的工作要么侧重于如何利用实体信息（例如实体类型、实体verbalization）来推断关系，却忽视了以上下文为中心的内容；要么采用反事实思维来消除模型对实体中潜在关系的偏差，但无关内容仍会妨碍关系推理过程。因此，如何保留相关内容并从句子中去除噪声片段是一项关键任务。此外，保留的内容还需要足够流畅，以维持语义连贯性和可解释性。在这项工作中，我们提出了一种新颖的理由提取框架 RE2，利用连续性和稀疏性两个因子，从句子中提取相关且连贯的理由。为解决金标准理由未标注的问题，RE^2 对句子中的每个词元应用可优化的二值掩码，并根据关系标签调整需要选取的理由。在四个数据集上的实验表明，RE^2 优于基线方法。

CCS 概念

• **计算方法 → 信息抽取**。

关键词

连续理由提取，关系抽取 ion

ACM 参考格式：
Xuming Hu、Zhaochen Hong、Chenwei Zhang、Irwin King 和 Philip S. Yu。2023年。理性看待所见：面向关系抽取的连续理由提取。载于第*46*届国际*ACM*信息检索研究与发展会议论文集（*SIGIR '23*），*2023*年*7*月*23*日至*27*日，中国台湾台北。ACM，美国纽约州纽约市，共5页。https://doi.org/10.1145/3539618.3592072

1 引言

关系抽取（RE）是许多信息检索（IR）系统的重要组成部分，能够从句子中抽取实体之间的关系。这些结构化三元组，例如 *(Ryan, Yaz,per/per/alumni)*

（图 1）来自异构来源的数据可能会受益于



图 1：不同模型通过不同的思考方式，在句子中"看到"不同的内容。理性思考通过看到相关且正确的内容，预测出实体 Ryan 和 Yaz 之间正确的关系标签 per/per/alumni。

多个下游应用，如问答 [19, 20] 和自然语言理解 [14, 21, 22]。为了获得结构化三元组，我们需要从实体上下文中提取相关且无噪声的句子，从而在两个给定实体之间正确地抽取关系。实体思维 方法，例如 Hu 等人 [15, 16, 17]，在实体前后分别插入保留的特殊标记 <e> 和 </e>，并通过这些特殊标记关注实体的上下文化特征。然而，实体的语义信息无法通过特殊标记来明确指定。因此，Zhou 和 Chen [37] 以及 Lu 等人 [24] 分别引入了实体类型和实体 verbalization，以更好地揭示实体的上下文语义表示，并推断实体之间的关系。尽管实体思维方法能够更好地捕获实体的上下文语义，但无法自动移除噪声内容和无关内容。这类噪声内容往往会破坏正确的关系推理。以图 1 为例，实体思维方法无法判断诸如"他们在学校时就认识彼此"和"我和我的朋友"之类的内容与预测实体之间关系的相关性。因此，模型可能会受到"朋友"一词的误导，并错误地将关系预测为 Per/per/peer。为了消除这些内容对实体间关系抽取的潜在影响，反事实思维 方法 [25, 33] 消除了模型对不同词语和实体的偏差。然而，这些方法并未专注于显式移除带噪声的上下文内容，因此，模型的预测仍可能受到误导，错误地预测为 Per/per/siblings。

为了去除句子中的无关内容和噪声内容，我们首先提出了理性思维方法，以提取相关的

本作品在知识共享署名国际 4.0 许可协议授权。*SIGIR '232023*年7月 台北2023年 版权归所有者/作者所有。ACM ISBN 978-1-4503-9408-6/23/07。https://doi.org/10.1145/3539618.3592072

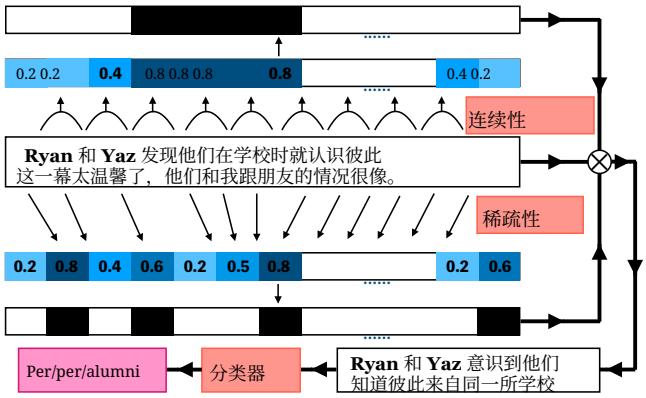

图2：架构：理由提取器使用由连续性和稀疏性因子组成的二值掩码，从输入中获取理由。

且无噪声的关系抽取任务理由。尽管理性思维方法已在问答等信息检索下游任务中得到验证 [34]，，但将理性思维方法用于关系抽取任务仍面临两个关键挑战：(1) 没有与句子中的关系标签相关的金标准理由，因此我们无法通过监督学习的方式训练理由提取器；(2) 鼓励提取出的理由保持连续，这不仅能够提升理由的可解释性，还能表达连贯的语义，以预测实体之间的关系标签。

为应对这两项主要挑战，我们在本研究中提出了两种新的连续性和稀疏性因子，用于管理所选理由词元的连贯性和数量。施加稀疏性有助于在消除无关材料与保留相关内容之间取得平衡。促进连续性有利于获得连续的理由，从而形成更连贯的语义表示。此外，我们采用可调节的二值掩码来选择理由，并利用关系标签修改关系抽取任务所需的理由词元。因此，可以通过端到端训练解决缺少金标准理由的问题。我们的主要贡献包括：(1) 提出一种新颖的端到端训练系统 RE2，将理由抽取视为关系抽取任务的可调节二值掩码，并通过连续性和稀疏性因子保留相关且无噪声的理由。(2) 在四个常用数据集上的实验表明，RE^2 在完整数据和低资源设置下均显著优于目前报告的最佳基线。

2 提出模型

2.1 连续理由提取器

在模型的连续理由提取器模块中，我们屏蔽与关系抽取任务无关的词元，并保留连续的词元以提升抽取性能。

2.1.1 句子表示与重要性矩阵。 我们可以通过其上下文化的句子表示获得每个词元的语义嵌入。在实践中，我们采用 BERT [8] 对词元表示进行编码： $\boldsymbol{x_{sent}} \in \mathbb{R}^{D\times L}$，其中 D 是嵌入的维度， L 是词元的数量

在句子中。为了在关系抽取任务中选择与句中实体最相关的词元，我们首先利用从 $\boldsymbol{x_{sent}}$ 中提取的两个实体 $\boldsymbol{x_{e_1}}$ 和 $\boldsymbol{x_{e_2}}$ 的词元嵌入，计算重要性矩阵： $\boldsymbol{s} = \boldsymbol{x_{sent}}^\top (\boldsymbol{x_{e_1}} + \boldsymbol{x_{e_2}})$。我们将 $\boldsymbol{s} = (s_1, ..., s_L)^\top$ 记为，并得到重要性得分 s_i，该得分表示 i^{th} 词元对于关系抽取任务的重要性。

2.1.2 因子图。 我们使用一个二元向量 $\boldsymbol{m} = (m_1, m_2, ..., m_L)^\top$表示词元选择，其中 $m_i \in \{0, 1\}$。取值 0 或 1 用于指示是否选择了该词元。通过这种方式，我们可以将理由序列生成的结构化预测问题转化为对多个变量赋值的问题。为了保持词元选择的语义连贯性，必须考虑词元选择的连续性。此外，为了强调词元与实体的重要性和相关性，我们需要通过稀疏性限制词元数量。因此，我们引入因子图 F，并将这些要求分解为多个局部因子，以实现最优词元选择。具体而言，我们采用成对因子 CONTINUITY 和 L 元因子 SPARSITY。在接下来的章节中，我们将形式化这两个因子，并给出它们的评分函数。

连续性（CON）：为了提高为关系抽取任务选取的词元的连续性，我们采用连续性（CON）因子，该因子可以检查每对相邻词元是否都被选中。我们采用因子 CON $(m_i, m_{i+1}; r_{i, i+1})$ 来表示对 i^{th} 和 $(i+1)^{th}$ 词元连续选择的约束。如图2所示，如果两个词元都被选中，我们会在评分函数中添加边评分 $r_{i, i+1} \geq 0$，以鼓励这种连续选择。形式上，CON 因子的评分函数可表示为：

$$\text{score}_{\text{CON}}(m_i, m_{i+1}; r_{i,i+1}) = m_i m_{i+1} r_{i,i+1}. \tag{1}$$

如第 2.1.1 节所示，我们采用重要性矩阵 $\boldsymbol{s}$ 来衡量与实体相关的词元，这是词元选择的一项关键指标。我们将所选 i^{th} 和 $(i+1)^{th}$ 词元的分数加入评分函数，并最终将评分函数定义为：

$$\text{score}_{\text{CON}}(m_i, m_{i+1}; r_{i,i+1}) = m_i m_{i+1} r_{i,i+1} + m_i s_i + m_{i+1} s_{i+1}. \tag{2}$$

我们可以利用成对因子的组合，对原句的词元选择施加连续性约束。形式上，因子图可以表示为：

$$\mathcal{F} = \{\text{CON}(m_i, m_{i+1}; r_{i,i+1}) : 1 \leq i < L\}. \tag{3}$$

稀疏性（SPA）：为了控制关系抽取任务中词元选择的稀疏性，我们采用 L 元因子稀疏性（SPA），通过对所选词元的最大数量施加限制 K 作为约束。在实践中， K 也可以是句子中所有词元的比例。根据定义，SPA 因子是一种硬约束。我们可以用以下评分函数来表示 SPA 因子：

$$\text{score}_{\text{SPA}}(m_1, m_2, \cdots, m_L, K) = \begin{cases} 0, & \sum m_i \leq K, \\ -\infty, & \sum m_i > K. \end{cases} \tag{4}$$

总之，为了同时考虑连续性和稀疏性，我们通过实例化 L 二元变量并结合 CON 和 SPA 因子，得到因子图 F：

$$\mathcal{F} = \{\text{SPA}(m_1, ..., m_L; K)\} \cup \{\text{CON}(m_i, m_{i+1}; r_{i,i+1}). \tag{5}$$

为了利用连续性和稀疏性约束，并找到词元选择的最优解，因子图 F 的评分函数需要将局部子问题 {CON(m_i, m_{i+1}; $r_{i,\,i+1}$) : $1 \le i < L$} 和 {SPA(m_1, ..., m_L; K)} 求和，如下所示：

$$\text{score}(m;s) = \begin{cases} \sum_{i=1}^{L} m_i s_i + \sum_{i=1}^{L-1} m_i m_{i+1} r_{i,i+1}, & \sum m_i \le K, \\ -\infty, & \sum m_i > K. \end{cases} \tag{6}$$

F 的硬约束继承自 SPA 因子，该因子规定所选词元的总数不得超过 K。软约束继承自 CON 因子，旨在鼓励选择与关系抽取任务相关的连续词元。为了找到能够较好地满足这些约束的解，我们将求解这些变量的问题视为一个最大后验（MAP）推断问题，以最大化评分函数：score ($\boldsymbol{m}; \boldsymbol{s}$)。我们可以将该问题表示为在满足 $|\boldsymbol{m}|_1 \le K$ 的约束下最大化评分函数：

$$\hat{\boldsymbol{m}} = \arg \max_{|\boldsymbol{m}|_1 \le K, \boldsymbol{m} \in \{0,1\}^L} \underbrace{\left(\boldsymbol{s}^\top \boldsymbol{m} + \sum_{i=1}^{L-1} m_i m_{i+1} r_{i,i+1}\right)}_{\text{score}(\boldsymbol{m};\boldsymbol{s})}. \tag{7}$$

事实上，最大化评分函数本质上是一个复杂的结构化问题，涉及具有相互关联的全局一致性约束的子问题，因此难以找到精确的最大化算法 [26]。为解决这一问题，我们考虑采用拉格朗日乘子边缘推断。

2.1.3 拉格朗日乘子边缘推断。 我们可以利用吉布斯分布求解最大化问题，并得到一个近似解。我们构造一个吉布斯分布，使得 $p(\boldsymbol{m}; \boldsymbol{s}) \propto exp(\text{score}(\boldsymbol{m}; \boldsymbol{s}))$。这样，我们就可以从 $\hat{\boldsymbol{m}} \sim p(\boldsymbol{m}; \boldsymbol{s})$ 中采样，并获得近似最优解。然而，获取无偏样本颇具挑战。为解决这一问题，我们采用扰动与最大后验 [27],这一近似采样策略。

另一个问题是，式 6 中的得分函数是分段函数，导致吉布斯分布 $p(\boldsymbol{m}; \boldsymbol{s}) \propto exp(\text{score}(\boldsymbol{m}; \boldsymbol{s}))$ 不连续。由于不连续马尔可夫随机场中的边缘推断难以求解，我们使用拉格朗日乘子重新表述硬约束：式 6 中的 SPA，即用连续函数表示硬约束。具体而言，我们使用一个拉格朗日乘子 $\lambda > 0$，并在式 7 的目标函数中加入 $\lambda(K - |\boldsymbol{m}|_1)$。最终，式 7 可写为：

$$\hat{\boldsymbol{m}} = \arg \max_{|\boldsymbol{m}|_1 \le K, \boldsymbol{m} \in [0,1]^L} \left(\boldsymbol{s}^\top \boldsymbol{m} + \sum_{i=1}^{L-1} m_i m_{i+1} r_{i,i+1} + \lambda(K - |\boldsymbol{m}|_1)\right), \tag{8}$$

其中，吉布斯分布应重新表述为 $p\ \boldsymbol{m}\ \boldsymbol{s}\ \propto exp\ \boldsymbol{m}\ \boldsymbol{s}\ +\lambda\ K - |\boldsymbol{m}|_1$ (;) (score(;) ( ))。因此，重新表述后的吉布斯分布变为连续分布，使我们能够计算出使评分函数最大化的最优 $\boldsymbol{m}$，并获得与关系抽取任务相关的理由。

2.2 关系分类器

最后，分类器根据选定的理由和实体 $\boldsymbol{e}$ 进行关系预测 $\hat{\boldsymbol{y}} = \text{pred}(\boldsymbol{m} \odot \boldsymbol{x} \,\|\, \boldsymbol{e})$，以获得关系标签分布。 $\odot$ 和 $\|$ 表示逐元素的分别表示乘积和拼接。关系分类损失可计算为：$\mathrm{L} = -\sum_{i=1}^{N} \boldsymbol{y}_i \log \hat{\boldsymbol{y}}_i$，其中 N是一个轮次中的训练句子数量， $\boldsymbol{y}_i$ 是句子 x_i 的真实标签向量。关系分类损失可通过端到端方式联合训练连续理由提取器模块和关系分类器模块。

3 实验与分析

3.1 实验设置与基线模型

设置： 我们在四个广泛使用的关系抽取数据集上评估模型：SemEval [13], 包含 6,507/1,493/2,717 个训练集/开发集/测试集样本和 19 种关系类型。TACRED [36]和 TACRED-Revisit[1], 包含 68,124/22,631/15,509 个样本和 42 种关系类型。Re-TACRED [32], 包含 58,465/19,584/13,418 个样本和 42 种关系类型。遵循先前的工作 [33], , 我们采用微平均F1作为评估指标。在低资源设置下，我们随机抽取训练集的 10%、25% 和 50 % 作为小规模训练集进行评估，并在测试集上评估模型。我们使用 BERT-Base 默认分词器，并将最大长度设为 128 来预处理数据。我们将 K 设为句中所有词元的 60%。对于分类器，我们将各层维度设为 768-384-标签数。我们使用 BertAdam [18] ，学习率为 3e-5，预热比例为 0.06，以优化损失，并将批次大小设为 16。**基线：** 我们首先介绍关系抽取任务中的 SOTA 模型作为基础模型，然后采用各种基线。我们采用 SURE [24]作为基础模型。我们将 RE^2 与以下基线进行比较：实体思维 基线：(1) MTB [31], (2) 实体掩码 [36], (3) 类型化实体标记 [37]。反事实思维 基线采用因果推断来消除关系抽取任务中的偏差：(4) CFIE [25], (5) CORSAIR [29](6) CORE [33]。理由思维 基线可以针对关系抽取任务，预测输入词元上的稀疏二值掩码：(7) HardKuma [2], (8) IB目标 [28], (9) UNIREX [3]。请注意，SURE 中也使用了实体思维方法，所有实体思维基线都用于替换 SURE 中的方法。

3.2 结果与分析

总体性能。表 1 展示了在四个数据集上进行 5 次训练和测试所得结果的均值与标准差。我们发现，与其他实体思维方法相比，使用实体信息口语化方法（SURE [24]）在所有数据集上的 F1 平均提升了 0.6%。因此，我们采用 SURE 作为基础模型。对于反事实思维和理由思维，我们发现二者都使所有数据集上的 F1 性能提升了 0.3%。我们提出的理性思维方法解决了两项主要挑战：（1）对理由抽取和关系分类器进行端到端训练；（2）抽取连续的理由。因此，RE^2 在所有关系抽取数据集上（包括低资源关系抽取设置）均实现了显著的 F1 提升，增幅为 0.9%。与先前的 SOTA 方法 UNIREX 相比，RE^2 的 F1 性能又提升了 0.4%。一个有趣的发现是，在低资源设置下（例如，仅使用 10% 的训练集），RE^2 的性能提升高于使用完整数据的设置：1.1% 对比 0.9%。这表明 RE^2 在训练数据有限的情况下仍具有足够的鲁棒性。使用 RE^2 获得的低噪声且相关性强的理由，能够更显著地提升基础模型的 F1 性能。

表1：四个关系抽取数据集上的微平均F1平均结果。"re."表示我们将用相应的实体思维基线替换 SURE 中的实体信息口语化。我们标注了结果的（标准差）。

方法/数据集	SemEval				TACRED				TACRED-Revisit				Re-TACRED			
	10%	25%	50%	100%	10%	25%	50%	100%	10%	25%	50%	100%	10%	25%	50%	100%
SURE	77.2	81.5	83.9	86.3	67.9	70.4	71.9	73.3	72.3	75.1	77.4	79.2	78.5	82.6	84.7	88.2
re. MTB [31]	76.3	80.8	82.9	85.4	67.2	69.8	71.1	72.3	71.4	74.3	76.6	78.4	77.6	81.8	83.8	87.2
re. 实体掩码 [36]	76.7	80.9	83.3	86.0	67.3	70.0	71.4	72.5	71.8	74.7	76.9	78.5	77.8	82.2	84.3	87.6
re. 类型标记 [37]	77.0	81.3	83.7	86.3	67.8	70.4	71.8	73.2	72.1	75.0	77.2	79.0	78.2	82.4	84.5	88.1
+CFIE [25]	77.3	81.7	84.1	86.6	68.2	70.5	72.2	73.5	72.5	75.2	77.6	79.4	78.6	82.8	85	88.4
+CORSAIR [29]	77.4	81.8	84.1	86.7	68.3	70.7	72.4	73.7	72.7	75.4	77.7	79.6	78.9	83.1	85.1	88.6
+CORE [33]	77.5	82.0	84.2	86.7	68.3	70.8	72.3	73.7	72.8	75.6	77.8	79.6	78.9	83.2	85.3	88.7
+HardKuma [2]	77.4	81.4	84.0	86.5	68.0	70.6	72.2	73.5	72.5	75.3	77.6	79.4	78.7	82.7	85.0	88.3
+IB目标 [28]	77.5	81.7	84.3	86.6	68.3	70.7	72.3	73.6	72.6	75.3	77.8	79.5	79.0	83.0	85.1	88.6
+UNIREX [3]	77.8	81.9	84.4	86.7	68.5	70.9	72.5	73.8	72.8	75.7	78.0	79.7	79.2	83.1	85.4	88.8
+**RE^2**	**78.2(0.2)**	**82.3(0.1)**	**84.8(0.1)**	**87.2(0.2)**	**68.9(0.3)**	**71.3(0.2)**	**73.0(0.3)**	**74.2(0.1)**	**73.3(0.2)**	**76.0(0.2)**	**78.4(0.1)**	**80.1(0.1)**	**79.6(0.3)**	**83.6(0.2)**	**85.7(0.1)**	**89.1(0.2)**
无连续性	77.8(0.3)	81.9(0.4)	84.2(0.2)	86.8(0.4)	68.5(0.3)	70.7(0.2)	72.5(0.3)	73.8(0.2)	72.8(0.4)	75.5(0.3)	77.9(0.3)	79.7(0.2)	79.1(0.4)	83.2(0.3)	85.3(0.2)	88.8(0.3)
无稀疏性	77.6(0.2)	81.8(0.2)	84.3(0.1)	86.6(0.3)	68.4(0.1)	70.5(0.2)	72.4(0.3)	73.6(0.1)	72.7(0.2)	75.5(0.3)	77.8(0.3)	79.8(0.2)	79.0(0.2)	83.1(0.3)	85.1(0.1)	88.6(0.1)
不添加实体	78.0(0.2)	82.2(0.1)	84.6(0.2)	87.1(0.2)	68.8(0.3)	71.1(0.2)	72.8(0.2)	74.0(0.1)	73.1(0.2)	75.9(0.2)	78.2(0.1)	80.0(0.1)	79.3(0.3)	83.4(0.2)	85.6(0.2)	89.0(0.1)

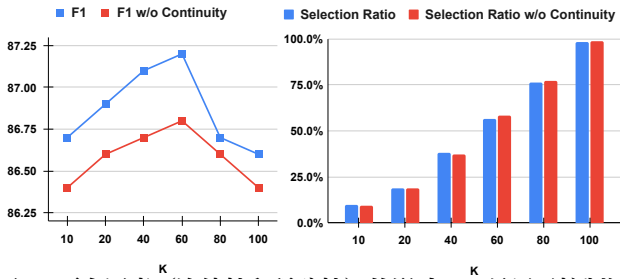

图3：两个因素（连续性和稀疏性）的影响。K是用于控制词元选择稀疏性的超参数。施加连续性以提高连贯性。

消融研究。我们在测试集上进行消融研究，以展示模型各个模块的有效性。RE^{2}不含连续性和 RE^2 不含稀疏性分别移除了理由抽取中因子图里的连续性和稀疏性要素。RE^2 不添加实体则移除了关系分类器模块中添加的实体，仅使用理由进行关系分类。表 1 总体表明，所有模块都有助于提升性能。具体而言，缺少连续性会导致理由不连续，影响语义表示的连贯性，使 F1 性能下降 0.4%。移除稀疏性会选出噪声更多的理由，使 F1 性能降低 0.5%。有趣的是，移除添加的实体对 F1 性能影响很小（0.1%）。我们发现，89% 的理由包含两个实体，97 % 的理由至少包含一个实体，这表明添加实体几乎不会提供额外信息。**两个因子的影响：**如图 3 所示，我们展示了 SemEval 上 F1 分数和词元选择率随 K 取值变化的情况。随着 K 增大，句子中更多词元会被选作理由。不过，由于引入了不相关的理由，RE^2 的 F1 分数并未随 K 增大而持续上升。在 K = 60处性能最佳，这意味着平均有 60% 的词元被选作理由。完全移除稀疏性因子会使模型的 F1 分数从 87.2 降至 86.6。此外，连续性约束也有利于模型，因为采用连续性约束的 RE^2 始终能取得更好的结果。**理由连贯性分析**。RE^2 利用连续性因子来控制理由的生成，使其在语义上更加连贯，从而表达得更加流畅。我们通过困惑度分析理由的连贯性，基于

表2：提取的理由的困惑度。Original表示原始句子。困惑度越低越好。

方法 / 数据集	SemEval	TACRED	TACRED-Revisit	Re-TACRED
HardKuma [2]	11.37	13.35	12.21	11.93
IB目标 [28]	10.37	12.23	11.56	10.74
UNIREX [3]	13.42	12.64	14.22	13.87
RE^2	**5.13**	**5.68**	**6.02**	**5.75**
原始句子	3.75	4.02	3.83	3.68

表3：原始句子和提取的理由的人工评估（微平均F1 / 信息充分性）。

数据集	SemEval	TACRED	TACRED-Revisit	Re-TACRED
提取的理由	95.5 / 4.	89.3 / 4.0	93.5 / 4.2	96.8 / 4.5
原始句子	94.3 /4.5	87.6/4.3	92.3 / 4.2	96.0 / 4.6

在 GPT-3 [30] 上。根据表2，RE^2 可获得最低的平均困惑度，接近原始句子的水平。

人工评估。 我们组建了一支由15名成员组成的标注团队（其中5名成员参与数据验证），对理由进行人工评估。标注员使用原始句子和提取的理由预测关系标签，并分别对两者的信息充分性进行评分（1至5分）。分数越高，表示信息越充分。为确保一致性，我们评估了标注者间一致性，并进行了人工验证。表3显示，即使理由的信息充分性低于原始句子，标注员仍能给出更准确的关系标签；这表明，从句子中去除无关细节可以降低噪声并提高关系预测准确率。

4 结论与未来工作 K

本文提出了一种新颖的理由提取框架 RE2，该框架采用连续性和稀疏性这两个因素，来控制理由与关系抽取任务的相关性并提高连贯性。我们引入了带有拉格朗日乘子的边缘推断，以解决同时包含这两个因素的评分函数最大化问题。因此，在没有理由金标准标注的情况下，我们可以以端到端方式联合训练理由提取和关系分类任务。在四个数据集上的实验表明，RE^{2}效果显著。未来，我们可以将关系抽取研究拓展到知识图谱的构建 [6, 7, 9, 35], 、知识图谱的匹配[10–12, 23], 以及信息检索的加速 [4, 5]。

5 致谢

Chenwei Zhang 为通讯作者。Xuming Hu 和 Zhaochen Hong 对本项工作贡献相同。源代码可在此处获取：https://github.com/THU-BPM/RE2。本文所述工作部分受到中国国家重点研发计划（编号：2018AAA0100204）、中国香港特别行政区研究资助局（RGC GRF 2151185；CUHK 14222922）以及美国国家科学基金会（NSF）III-1909323 项目资助。

参考文献

[1] Christoph Alt、Aleksandra Gabryszak 和 Leonhard Hennig。2020年。《TACRED再探：对TACRED关系抽取任务的全面评估》。载于*ACL*论文集。1558–1569。[2] Jasmijn Bastings、Wilker Aziz 和 Ivan Titov。2019年。《利用可微二元变量实现可解释的神经预测》。载于*ACL*论文集。2963–2977。[3] Aaron Chan、Maziar Sanjabi、Lambert Mathias、Liang Tan、Shaoliang Nie、Xiaochang Peng、Xiang Ren 和 Hamed Firooz。2022年。《UNIREX：语言模型理由提取的统一学习框架》。载于*ICML*论文集。PMLR，2867–2889。[4] Yankai Chen、Yixiang Fang、Yifei Zhang 和 Irwin King。2023年。《用于汉明空间高效Top-N搜索的二部图卷积哈希》。载于*WWW*论文集。ACM。[5] Yankai Chen、Huifeng Guo、Yingxue Zhang、Chen Ma、Ruiming Tang、Jingjie Li 和 Irwin King。2022年。《通过多面量化强化学习二值化图表示，实现top-k推荐》。载于*SIGKDD*论文集。168–178。[6] Yankai Chen、Menglin Yang、Yingxue Zhang、Mengchen Zhao、Ziqiao Meng、Jianye Hao 和 Irwin King。2022年。《利用双曲几何对无标度图进行建模，以实现知识感知推荐》。载于*WSDM*论文集。94–102。[7] Yankai Chen、Yaming Yang、Yujing Wang、Jing Bai、Xiangchen Song 和 Irwin King。2022年。《协同引导下用于个性化推荐的注意力知识感知图卷积网络》。载于*ICDE*论文集。IEEE，299–311。[8] Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。2019年。《BERT：用于语言理解的深度双向Transformer预训练》。载于*NAACL-HLT*论文集。4171–4186。[9] Yixiang Fang、Reynold Cheng、Yankai Chen、Siqiang Luo 和 Jiafeng Hu。2017年。《有效且高效的属性社区搜索》。*VLDB*期刊 26（2017年），803–828。[10] Yunjun Gao、Xiaoze Liu、Junyang Wu、Tianyi Li、Pengfei Wang 和 Lu Chen。2022年。《ClusterEA：采用随机训练和归一化小批量相似度的可扩展实体对齐》。载于*KDD*。421–431。[11] Congcong Ge、Xiaoze Liu、Lu Chen、Baihua Zheng 和 Yunjun Gao。2021年。《化繁为简：一种有效的端到端实体对齐框架》。载于*SIGIR*。777–786。[12] Congcong Ge、Xiaoze Liu、Lu Chen、Baihua Zheng 和 Yunjun Gao。2022年。《LargeEA：面向大规模知识图谱的实体对齐》。*PVLDB* 15，2（2022年），237–245。[13] Iris Hendrickx、Su Nam Kim、Zornitsa Kozareva、Preslav Nakov、Diarmuid O Séaghdha、Sebastian Padó、Marco Pennacchiotti、Lorenza Romano 和 Stan Szpakowicz。2010年。《Semeval-2010任务8：名词对之间语义关系的多路分类》。载于*SemEval*论文集。33–38。[14] Xuming Hu、Zhijiang Guo、Yu Fu、Lijie Wen 和 Philip S. Yu。2022年。《将场景图修改作为增量结构扩展》。载于*COLING*论文集。5707–5720。[15] Xuming Hu、Fukun Ma、Chenyao Liu、Chenwei Zhang、Lijie Wen 和 Philip S Yu。2021年。《通过增量元自训练实现半监督关系抽取》。载于*EMNLP*：*Findings*论文集。[16] Xuming Hu、Lijie Wen、Yusong Xu、Chenwei Zhang 和 Philip Yu。2020年。《SelfORE：用于开放关系抽取的自监督关系特征学习》。载于论文集 EMNLP论文集。线上会议，3673–3682。[17] Xuming Hu、Chenwei Zhang、Yawen Yang、Xiaohe Li、Li Lin、Lijie Wen 和 Philip S. Yu。2021年。《用于低资源关系抽取的梯度模仿强化学习》。载于EMNLP论文集。2737–2746。[18] Diederik P Kingma 和 Jimmy Ba。2015年。《Adam：一种随机优化方法》。载于ICLR论文集。[19] Aiwei Liu、Xuming Hu、Li Lin 和 Lijie Wen。2022年。《通过迭代学习模式链接图，实现语义增强的文本到SQL解析》。载于KDD论文集。1021–1030。[20] Aiwei Liu、Xuming Hu、Lijie Wen 和 Philip S Yu。2023年。《对ChatGPT零样本文本到SQL能力的全面评估》。arXiv预印本 arXiv:2303.13547（2023年）。[21] Aiwei Liu、Honghai Yu、Xuming Hu、Shu'ang Li、Li Lin、Fukun Ma、Yawen Yang 和 Lijie Wen。2022年。《通过可附加子词替换对Transformer实施字符级白盒对抗攻击》。载于EMNLP论文集。[22] Shuliang Liu、Xuming Hu、Chenwei Zhang、Shu'ang Li、Lijie Wen 和 Philip S. Yu。2022年。《HiURE：用于无监督关系抽取的分层示例对比学习》。载于NAACL-HLT论文集。5970–5980。[23] Xiaoze Liu、Junyang Wu、Tianyi Li、Lu Chen 和 Yunjun Gao。2023年。《时序知识图谱的无监督实体对齐》。载于WWW。[24] Keming Lu、I Hsu、Wenxuan Zhou、Mingyu Derek Ma、Muhao Chen 等。2022年。《将摘要生成作为关系抽取的间接监督》。载于EMNLP论文集。[25] Guoshun Nan、Jiaqi Zeng、Rui Qiao、Zhijiang Guo 和 Wei Lu。2021年。《揭示长尾信息抽取中的主要因果关系》。载于EMNLP论文集。9683–9695。[26] Vlad Niculae 和 André F. T. Martins。2020年。《LP-SparseMAP：用于稀疏结构化预测的可微松弛优化》。载于第37届国际机器学习会议论文集（ICML 2020），2020年7月13日至18日，线上会议（《机器学习研究论文集》第119卷）。PMLR，7348–7359。[27] George Papandreou 和 Alan L Yuille。2011年。《扰动映射随机场：利用离散优化学习能量模型并从中采样》。载于ICCV论文集。IEEE，193–200。[28] Bhargavi Paranjape、Mandar Joshi、John Thickstun、Hannaneh Hajishirzi 和 Luke Zettlemoyer。2020年。《通过信息瓶颈方法控制理由抽取的简洁性》。载于EMNLP论文集。1938–1952。[29] Chen Qian、Fuli Feng、Lijie Wen、Chunping Ma 和 Pengjun Xie。2021年。《用于文本分类去偏的反事实推断》。载于ACL-IJCNLP论文集。5434–5445。[30] Alec Radford、Jeffrey Wu、Rewon Child、David Luan、Dario Amodei 和 Ilya Sutskever。2019年。《语言模型是无监督的多任务学习者》。OpenAI博客 1，8（2019年），9。[31] Livio Baldini Soares、Nicholas FitzGerald、Jeffrey Ling 和 Tom Kwiatkowski。2019年。《匹配空白：用于关系学习的分布相似性》。载于ACL论文集。2895–2905。[32] George Stoica、Emmanouil Antonios Platanios 和 Barnabás Póczos。2021年。《重新审视TACRED：解决TACRED数据集的不足》。载于AAAI论文集，第35卷。13843–13850。[33] Yiwei Wang、Muhao Chen、Wenxuan Zhou、Yujun Cai、Yuxuan Liang、Dayiheng Liu、Baosong Yang、Juncheng Liu 和 Bryan Hooi。2022年。《我们应该依赖实体提及进行关系抽取吗？通过反事实分析消除关系抽取中的偏差》。载于NAACL-HLT论文集。3071–3081。[34] Lei Zhang、Shuai Wang 和 Bing Liu。2018年。《情感分析中的深度学习：综述》。《Wiley跨学科评论：数据挖掘与知识发现》8，4（2018年），e1253。[35] Xinni Zhang、Yankai Chen、Cuiyun Gao、Qing Liao、Shenglin Zhao 和 Irwin King。2022年。《面向冷启动推荐的知识感知神经网络与个性化特征引用》。arXiv预印本 arXiv:2209.13973（2022年）。[36] Yuhao Zhang、Victor Zhong、Danqi Chen、Gabor Angeli 和 Christopher D Manning。2017年。《位置感知注意力和监督数据有助于槽位填充》。载于EMNLP论文集。35–45。[37] Wenxuan Zhou 和 Muhao Chen。2022年。《句子级关系抽取的改进基线》。AACL-IJCNLP论文集（2022年），161。