

基于模块转移和语义相似性推断的小样本关系三元组抽取^①



刘彤, 刘炳霄, 倪维健

(山东科技大学 计算机科学与工程学院, 青岛 266590)

通信作者: 倪维健, E-mail: niweijian@sdust.edu.cn

摘要: 针对现有的小样本关系三元组抽取方法难以解决单句话中存在多个三元组以及未考虑支持集和查询集之间语义相似性等问题, 提出了一种基于模块转移和语义相似性推断的小样本关系三元组抽取方法. 该方法采用一种在关系抽取、实体识别和三元组判别这3个模块不断转移的机制, 高效地提取出查询实例中存在的多个关系三元组. 在关系抽取部分, 将 BiLSTM 和注意力机制相融合, 以更好地捕捉应急预案文本的序列信息. 此外, 在实体识别部分设计了一种基于语义相似性推断的方法识别句子中存在的应急组织机构实体. 最终, 在应急预案领域数据集 ERPs⁺ 上进行了大量的实验. 实验结果显示, 相较于其他基线模型, 所提模型更适应于应急预案领域的关系三元组抽取任务.

关键词: 应急预案; 关系三元组抽取; 小样本学习; 模块转移; 语义相似性

引用格式: 刘彤, 刘炳霄, 倪维健. 基于模块转移和语义相似性推断的小样本关系三元组抽取. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9721.html>

Few-shot Relational Triple Extraction Based on Module Transfer and Semantic Similarity Inference

LIU Tong, LIU Bing-Xiao, NI Wei-Jian

(College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China)

Abstract: Existing few-shot relational triple extraction methods often struggle with handling multiple triples in a single sentence and fail to consider the semantic similarity between the support set and the query set. To address these issues, this study proposes a few-shot relational triple extraction method based on module transfer and semantic similarity inference. The method uses a mechanism that constantly transfers among three modules, namely relation extraction, entity recognition, and triple discrimination, to extract multiple relational triples efficiently from a query instance. In the relation extraction module, BiLSTM and a self-attention mechanism are integrated to better capture the sequence information of the emergency plan text. In addition, a method based on semantic similarity inference is designed to recognize emergency organizational entities in sentences. Finally, extensive experiments are conducted on ERPs⁺, a dataset for emergency response plans. Experimental results show that the proposed model is more suitable for relational triple extraction in the field of emergency plans compared with other baseline models.

Key words: emergency response planning; relational triple extraction; few-shot learning; module transfer; semantic similarity

^① 基金项目: 科技创新 2030—“新一代人工智能”重大项目 (2022ZD0119500); 山东省自然科学基金 (ZR2022MF319); 山东科技大学青年教师教学拔尖人才培养基金 (BJ20211110)

收稿时间: 2024-06-01; 修改时间: 2024-06-26; 采用时间: 2024-07-11; csa 在线出版时间: 2024-11-15

关系三元组抽取是自然语言处理和知识图谱信息抽取中的一项重要任务,其目的是从非结构化文本中抽取出一对实体和关系,即<头实体,关系,尾实体>。

在应急预案领域,应急组织机构作为应急预案的执行主体,在应急响应中扮演着至关重要的角色。因此,从应急预案文本中挖掘出应急组织机构及其之间的关系,即“关系三元组抽取”,对于应急组织机构快速、准确地应对突发事件具有极其重要的意义^[1]。

近年来,基于神经网络的方法如 TPLinker^[2]、CasRel^[3]和 PRGC^[4]等已广泛应用于关系三元组抽取任务,并取得了优异的性能。然而,这些方法通常依赖于大规模带标签的数据集才能展现出较好的性能。在应急预案领域,目前尚未有像 DuIE^[3]或 NYT^[4]这样的大规模带标签数据集,若依赖领域专家进行人工标注将耗费大量的人力和时间。为了解决这一问题,研究人员提出了远程监督方法^[5],用于自动标注大规模带标签的数据。然而,这种方法生成的数据中往往带有大量的噪声,严重影响了模型的性能。因此,如何利用少量标注样本有效地进行应急预案领域的关系三元组抽取任务仍是一个亟需解决的问题。

小样本学习方法^[6]旨在利用少量标记样本使模型学会解决问题的能力,以缓解特定领域中标签数据相对匮乏的问题。Yu 等人^[7]提出了一种基于小样本学习的多原型嵌入网络,该方法忽略了关系和实体之间的依赖关系,最终导致模型性能不佳。针对这一问题,Fei 等人^[8]提出了一种视角转移网络,采用“先关系后实体”的方式提取关系三元组,即独立地提取不同关系所对应的实体,最终取得了较为优异的性能。然而,该方法未能充分考虑支持集和查询集之间令牌的语义相似性问题。并且在面对语义较为复杂的应急预案文本时,关系抽取性能显著下降。此外,现有的小样本关系三元组抽取方法难以处理单句话中存在多个三元组的情况,而在应急预案文本中可能涉及多个三元组。

针对以上问题,本文提出了一种基于模块转移和语义相似性推断的小样本关系三元组抽取方法 (few-shot relational triple extraction method based on module transfer and semantic similarity inference, FS-MTS)。该方法采用一种在关系抽取、实体识别和三元组判别这3个模块之间不断转移的机制,解决了单句话中存在多个三元组难以识别的问题。在实体识别部分,FS-MTS模型为了充分考虑支持集和查询集之间令牌语义相似

性的问题,设计了一种基于语义相似性推断的方法。具体而言,该方法首先计算每个查询标记与相同关系类型的支持标记之间的相似性得分,然后将得分的平均值和最小值相加,以获得查询标记属于各类实体标签的概率。该方法的引入显著提高了实体识别的准确率。此外,FS-MTS模型还融合了 BiLSTM 和注意力机制,以更好地捕获应急预案文本的序列信息,从而提高模型关系抽取的准确性。

1 相关工作

1.1 关系三元组抽取

关系三元组抽取主要分成两个方向:流水线抽取和联合抽取。流水线抽取^[9,10]首先识别出句子中的实体,然后将实体对之间的关系进行分类。这种方法忽略了实体识别和关系抽取两个子任务之间的依赖关系,并且在实体识别任务中产生的错误会传递到关系分类任务中,导致误差累积的问题。联合抽取则是将实体识别和关系抽取任务整合到一个模型中进行联合建模。尽管这些方法已经取得了显著的性能提升,但它们大多数存在“数据饥渴”的问题,即通常需要大量的标签数据进行训练,才能取得较好的性能。然而,在某些特定领域,标签数据相对匮乏,这些方法的性能会受到一定的限制。

1.2 小样本关系三元组抽取

小样本关系三元组抽取是指利用小样本学习的方法来执行关系三元组抽取任务。相较于传统的关系三元组抽取方法,小样本关系三元组抽取旨在利用少量的标记样本从未标记的数据中提取关系三元组。此外,小样本关系三元组抽取可以利用较少的标记样本来处理未见过的实体类型或关系类别。近年来,研究人员开始广泛探索小样本学习在关系三元组抽取任务中的应用。Yu 等人^[7]首次将小样本学习方法应用于关系三元组抽取任务,并提出了一种多原型嵌入网络 MPE。该方法使用混合原型学习机制将文本和实体、关系的知识进行关联,以生成融合实体和关系信息的原型表示。然而,该方法在实体识别方面存在明显的局限性。因此,He 等人^[11]提出了一种基于最近邻匹配的小样本关系三元组抽取方法 NNM。最近,Cong 等人^[12]提出了一种关系引导的三元组抽取方法,该方法利用双重注意力机制来聚合关系的相关信息,然后利用对应的关系来引导模型识别头实体和尾实体。此外,Fei 等人^[8]提出了

一种视角转移网络 PTN. 该方法采用一种在关系、实体和三元组不同视图之间循环转移的方式, 从局部和全局上进行验证所提取的关系三元组是否合理. 尽管上述方法在 FewRel 数据集^[13]上表现出了不错的效果, 但这些方法主要专注于解决单个句子中仅包含一个关系三元组的情况. 在应急预案文本中, 单个句子中可能会涉及多个关系三元组. 因此, 为了更好地解决应急预案领域中存在的问题, 本文将深入研究更适应于该领域的小样本关系三元组抽取方法.

2 问题定义

在小样本关系三元组抽取任务中, 通常采用 episodic^[14]的训练方式, 将数据集划分成多个任务进行训练. 每个任务从数据集中随机选取 N 种关系类型, 并从每种关系类型中随机采样 K 个样本, 即 N -way K -shot 任务.

给定目标数据集 $T = \{D_{\text{train}}, D_{\text{test}}\}$, 训练集和测试集不相交, 即 $D_{\text{train}} \cap D_{\text{test}} = \emptyset$. 训练集 D_{train} 由支持集 $S_{\text{train}} = \{x_i^s, y_i^s\}_{i=1}^{s=N_s}$ 和查询集 $Q_{\text{train}} = \{x_j^q, y_j^q\}_{j=1}^{q=N_q}$ 组成. 每个支持集和查询集中都包含实例 x_i 和目标标签 y_i . 其中, $N_s = N \times K$ 和 $N_q = N \times K'$. 通常情况下, K 和 K' 对应的数值一般是相等的. 在训练阶段, 模型使用支持集中少量的标记样本进行学习类别之间的特征和关系. 而查询集则用于验证模型的性能并用于参数优化. 测试集 D_{test} 也被划分成两部分: 支持集 $S_{\text{test}} = \{x_i^s, y_i^s\}_{i=1}^{s=N_s}$ 和查询集 $Q_{\text{test}} = \{x_j^q\}_{j=1}^{q=N_q}$. 值得注意的是, 查询集中包含的测试样本不包含目标标签. 在测试阶段, 模型需要使用支持集中的少量支持实例标记查询集中的测试实例.

3 FS-MTS 模型

FS-MTS 模型的整体框架如图 1 所示, 主要包括 3 个部分: 关系抽取、实体识别和三元组判别. 该模型的输入数据为支持集和查询集, 目标是从查询集的句子中提取关系三元组. 在关系抽取部分, FS-MTS 模型首先标记支持实例中的实体对, 并将其与查询实例拼接后进行编码. 接着, 分别将实例的编码表示输入 BiLSTM 和注意力模块, 以更好地捕获应急预案文本的序列信息. 最终, 将全局编码表示拼接后输入二分类器, 以判断查询实例是否含有支持实例中的关系类型. 一旦检测到关系, 模型将进行实体识别任务, 利用基于语义相

似性推断的方法来识别实例中的头实体和尾实体. 然后, 模型转移到三元组判别部分, 利用识别出的头实体和尾实体来标记查询实例, 生成带有伪标签的查询实例. 接着, 将这些查询实例与标记的支持实例拼接后输入二分类器, 以判别预测的三元组是否合理. 最终, FS-MTS 模型输出可信的关系三元组, 并返回到关系抽取部分, 继续检测下一种关系类型. 接下来, 将详细介绍 FS-MTS 模型的每个部分.

3.1 关系抽取

关系抽取的目标是提取出查询实例中潜在的关系类型. 头实体和尾实体的信息在关系抽取任务中至关重要^[15]. 因此, 本文使用特殊标记 (如: $[h]$ 和 $[\backslash h]$) 等将头实体和尾实体的信息嵌入到支持实例中, 得到包含实体信息的支持实例 S'_{support} , 如式 (1) 所示:

$$S'_{\text{support}} = \{w_1, [h], w_{\text{head}}, [\backslash h], \dots, [t], w_{\text{tail}}, [\backslash t], w_n\} \quad (1)$$

其中, w_n 指的是句中的第 n 个字, $[h]$ 和 $[\backslash h]$ 用来标记头实体的开始和结束位置, $[t]$ 和 $[\backslash t]$ 用来标记尾实体的开始和结束位置. 接下来, 将支持实例 S'_{support} 和查询实例 S_{query} 进行结合, 如式 (2) 所示:

$$S_{\text{fusion}} = \{[\text{CLS}], S'_{\text{support}}, [\text{SEP}], S_{\text{query}}, [\text{SEP}]\} \quad (2)$$

其中, $[\text{CLS}]$ 和 $[\text{SEP}]$ 是特殊令牌. 然后, 将 S_{fusion} 输入到 BERT 编码器^[16]中, 如式 (3) 所示:

$$H = \text{BERT}(S_{\text{fusion}}) \quad (3)$$

其中, $H = [H_s; H_q]$ 是整个句子的词嵌入矩阵, H_s 和 H_q 分别是支持实例和查询实例的编码表示.

考虑到应急预案文本语义较为复杂的问题, 为了更好地捕捉应急预案文本的序列信息, 本文将 BiLSTM^[17] 和注意力机制^[18]进行融合, 以更全面地捕捉应急预案文本的上下文语义信息. 以支持实例 H_s 为例, 对于给定的输入序列 $H_s = (x_1^s, x_2^s, \dots, x_n^s)$, 每个时刻的输入 x_t 分别与前后两个时刻的隐藏状态 $\overrightarrow{h_{t-1}^s}$ 和 $\overleftarrow{h_{t+1}^s}$ 相结合, 得到两个方向的隐藏状态向量 $\overrightarrow{h_t^s}$ 和 $\overleftarrow{h_t^s}$. 最后, 将 $\overrightarrow{h_t^s}$ 和 $\overleftarrow{h_t^s}$ 进行拼接, 生成每个词的特征向量表示 h^s , 如式 (4) 所示:

$$\begin{cases} \overrightarrow{h_t^s} = \text{LSTM}(x_t, \overrightarrow{h_{t-1}^s}) \\ \overleftarrow{h_t^s} = \text{LSTM}(x_t, \overleftarrow{h_{t+1}^s}) \\ h^s = [\overrightarrow{h_t^s}; \overleftarrow{h_t^s}] \end{cases} \quad (4)$$

然后, 再利用注意力机制来计算句子中任意两个

词之间的相关性,以获取更加全面的上下文信息.以支持实例 $h^s = \{h_i^s \mid i = 1, 2, \dots, n\}$ 为例,其中 n 为序列长度,根据式(5)和式(6)计算出序列中每个向量 h_i^s 所对应的权重 α_i^s .

$$h^s = \tanh(w_s h^s + b_s) \quad (5)$$

$$\alpha_i^s = \frac{\exp(u_s h_i^s)}{\sum_i \exp(u_s h_i^s)} \quad (6)$$

其中, \tanh 为激活函数, u_s 为可训练参数.通过利用隐藏状态向量 h^s 和注意力分数得到句子的全局编码表示 s ,如式(7)所示:

$$s = \sum_i \alpha_i^s h_i^s \quad (7)$$

最终,将支持实例和查询实例的全局编码表示 s 和 q 进行拼接,得到 $H_{\text{global}} = [s : q]$.将 H_{global} 输入到二分类器中,以检测查询实例和支持实例中是否存在相同

的关系类型,计算公式如式(8)所示:

$$\begin{cases} p_r = \text{Softmax}(\text{MLP}(H_{\text{global}})) \\ \text{act} = \text{argmax}(p_r) \end{cases} \quad (8)$$

其中, p_r 是指从关系抽取模块转移到实体识别模块的概率, MLP 为多层感知器, $\text{act} \in \{0, 1\}$.若 act 为1,模型将从关系抽取模块转移到实体识别模块,并执行实体识别任务.若 act 为0,模型仍然处于关系抽取模块,并将查询实例与其他关系类型的支持实例对比.

在关系抽取部分,采用负对数似然损失函数,如式(9)所示:

$$L_{\text{rel}} = - \sum_{i=1}^N \log(p_r(\text{act}_i = y_i^r \mid S_{\text{fusion}})) \quad (9)$$

其中, N 为训练数据中句子的数量, $y_i^r \in \{0, 1\}$ 是第 i 个目标标签,代表查询实例是否包含关系 r .0表示“不包含”,1表示“包含”.

支持实例 S'_{support} [h]省减灾委[h]...向[t]省政府[v]报告...
 查询实例 S_{query} 市应急指挥部办公室...报告市应急指挥部...
 S'_{query} [h]市应急指挥部办公室[h]...报告[t]市应急指挥部[v]...
 S''_{query} [h]市应急指挥部办公室[h]...报告市应急[t]指挥部[v]...

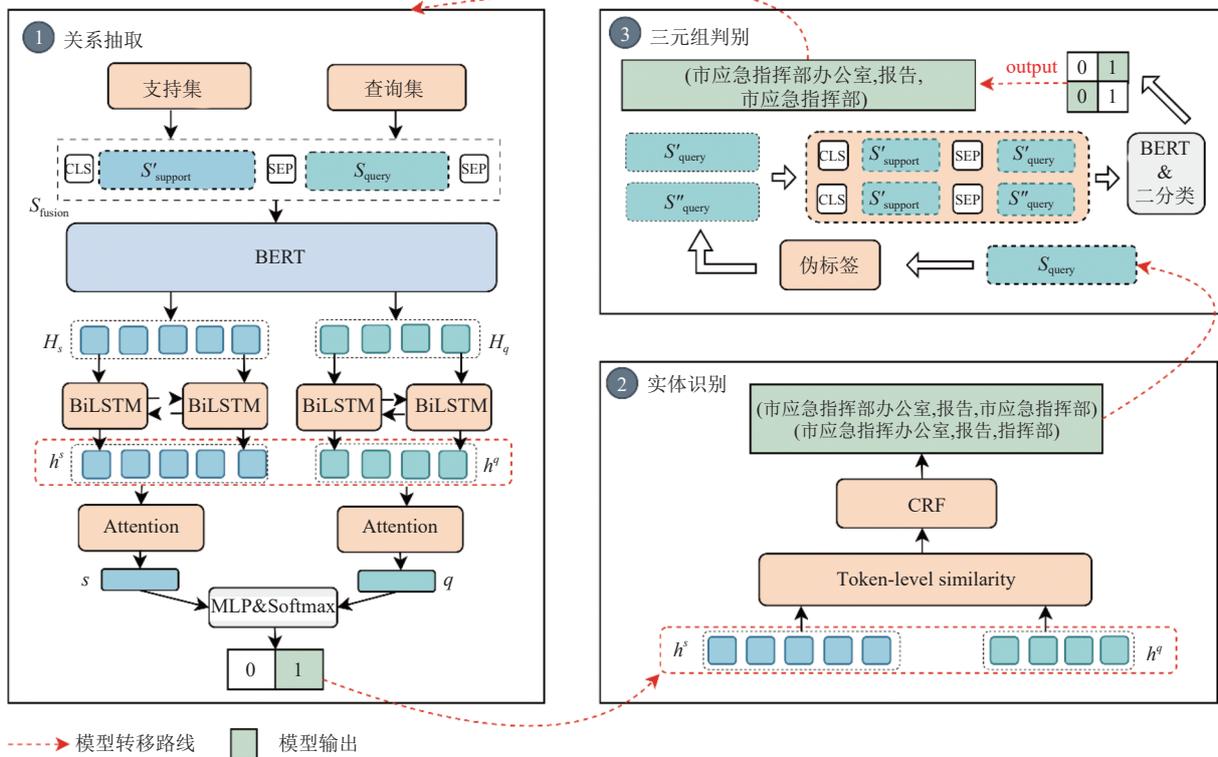


图1 FS-MTS模型框架

3.2 实体识别

在实体识别部分,本文采用基于条件随机场的方

法来识别查询实例中的头实体和尾实体.条件随机场由发射概率和转移概率组成,发射概率通常代表

一个标记属于各类实体标签的概率. 本文使用基于语义相似性方法的输出结果来代替条件随机场的发

射概率. 图2展示了查询实例的发射概率矩阵的计算结果.

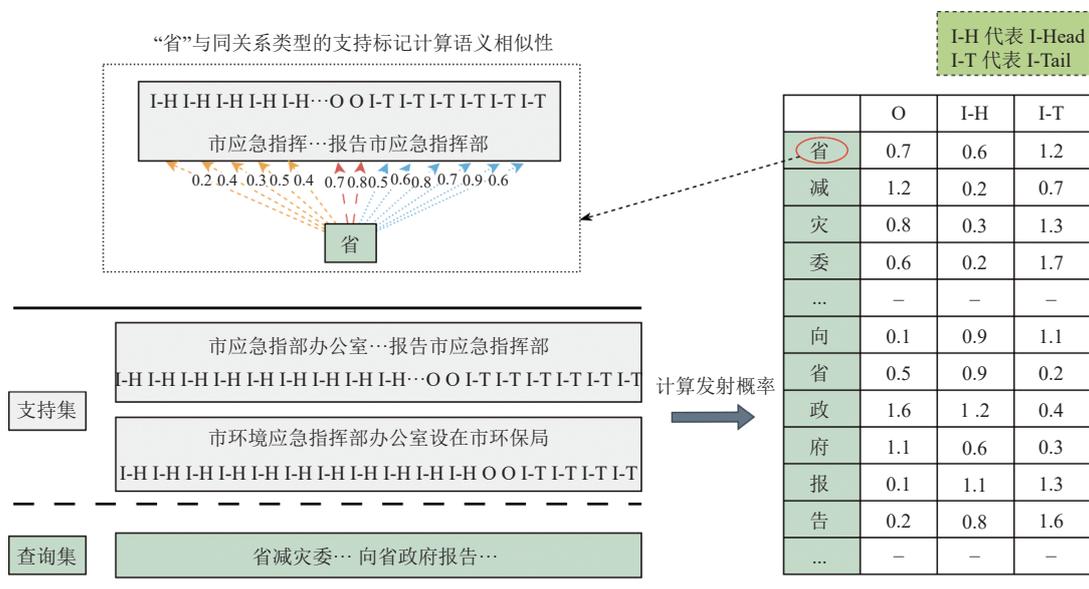


图2 发射概率矩阵

发射概率的计算过程如下, 首先从支持集中筛选出与查询实例具有相同关系类型的所有支持实例. 然后, 将查询实例中的每个标记 x_j^q 与所选支持实例的所有标记 x^s 计算相似性得分 $sim(x_j^q, x^s)$, 计算公式如式(10)所示:

$$sim(x_j^q, x^s) = d(x_j^q, x^s) = \|x_j^q - x^s\|_2^2 \quad (10)$$

其中, $d(\cdot, \cdot)$ 代表欧氏距离函数, 距离越小相似性越高. 在使用距离函数之前对句子的特征使用了L2正则化. 根据预定义的实体标签(I-Head、I-Tail和O), 将每个标记的相似性得分划分成3组, 相同实体标签的相似性得分被归为同一组^[19]. 然后, 计算出每一组相似性得分的平均值和最小值. 最终, 将每一组内的平均值和最小值相加, 该结果被视为每个查询标记属于各类实体标签的概率, 即“发射概率”. 计算公式如式(11)和式(12)所示:

$$m_{j,c} = \begin{cases} sim(x_j^q, x^s), & \text{if } Y_j = c \\ 0, & \text{其他} \end{cases} \quad (11)$$

$$P(y = c | x_j^q) = \min(m_{j,c}) + \text{avg}(m_{j,c}) \quad (12)$$

其中, Y_j 为所选支持实例中每个标记的实体标签, c 代表实体标签类别(I-Head、I-Tail、O), $P(y = c | x_j^q)$ 为查询标记 x_j^q 属于类别 c 的概率.

转移概率代表的是标签与标签之间的转移分数. 在有监督学习的实体识别任务中, 转移概率通常根据 BIO 标注法进行构建. 然而, 在小样本学习中, 通常采用 IO 标注法. Yang 等人^[20]根据 IO 标注法来构建维特比解码器, 以捕捉实体的转移概率分布. 然而, 现有方法中并未引入维特比解码器. 为了获得更优异的序列标注结果, 本文在实体识别任务的输出端引入维特比解码器. 通过统计 O、I 和 I-Other 标签之间的转移次数来估算源数据集中的抽象转移矩阵.

图3描述了如何从源数据集中的抽象转移矩阵扩张到目标领域的转移矩阵. 在源数据集中的抽象转移矩阵中, O代表非实体, I代表某个类型的实体, I-Other代表除该类型实体外的其他类型实体. 通过将源数据集中的抽象转移矩阵均匀地分割, 最终得到目标数据集的转移矩阵. 例如, 将O转移为I的概率均匀地分配到I-Head和I-Tail这两个类型. 通过将抽象转移矩阵扩展到目标测试集中, 使得最终标签序列的输出概率达到最优. 抽象转移矩阵的计算方法如式(13)所示:

$$P(Y | X) = \frac{N(X \rightarrow Y)}{N(\cdot \rightarrow Y)} \quad (13)$$

其中, $N(X \rightarrow Y)$ 和 $N(\cdot \rightarrow Y)$ 分别代表 X 转化为 Y 的频率和任何标签转化为 Y 的频率.

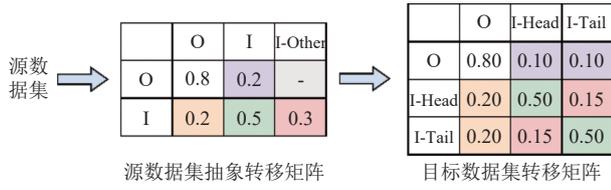


图3 抽象矩阵扩张

维特比解码器充分考虑了标签之间的转移概率和发射概率,以得出最优的标签预测结果.最终,模型对每个标签概率的计算方法如式(14)所示:

$$y^* = \underset{y}{\operatorname{argmax}} \prod_{t=1}^T P(y_t | x_t^q) \times P(y_t | y_{t-1}) \quad (14)$$

其中, $P(y_t | x_t^q)$ 为发射概率, $P(y_t | y_{t-1})$ 为转移概率.

实体识别部分的损失函数如式(15)所示:

$$L_{\text{entity}} = \frac{e^{(P(y_t|x_t^q)+P(y_t|y_{t-1}))}}{\sum_{y'} e^{(P(y_t|x_t^q)+P(y_t|y_{t-1}))}} \quad (15)$$

其中, $P(y_t | x_t^q)$ 为发射概率,表示在查询实例中第 i 个标记 x_t^q 的情况下,预测标签为 y_t 的概率. $P(y_t | y_{t-1})$ 为转移概率,表示在前一个标签 y_{t-1} 的情况下,当前标签为 y_t 的概率.

3.3 三元组判别

三元组判别部分主要用来验证预测的三元组是否合理.为此,本文对查询实例 S_{query} 预测出的实体对使用4个特殊令牌进行标记,以生成带有伪标记的查询实例表示 S'_{query} .标记方法如式(16)所示:

$$S'_{\text{query}} = \{w_1, [h], w_{\text{head}}, [\backslash h], \dots, [t], w_{\text{tail}}, [\backslash t], w_m\} \quad (16)$$

然后,将支持实例 S'_{support} 和带有伪标记的查询实例 S'_{query} 进行拼接,如式(17)所示:

$$S''_{\text{fusion}} = \{[\text{CLS}], S'_{\text{support}}, [\text{SEP}], S'_{\text{query}}, [\text{SEP}]\} \quad (17)$$

使用编码器 BERT 将 S''_{fusion} 进行编码, $H = \text{BERT}(S''_{\text{fusion}})$.然后,将编码结果 H 输入二分类器,用于验证预测的关系三元组是否合理,计算公式如式(18)所示:

$$\begin{cases} p_t = \text{Softmax}(\text{MLP}(H)) \\ pla = \operatorname{argmax}(p_t) \end{cases} \quad (18)$$

其中, p_t 指的是三元组是否合理的概率值,MLP 为多层感知器, $pla \in \{0, 1\}$.若 pla 为 1,则预测的关系三元组合理,并将其保留.若 pla 为 0,则关系三元组不合理,将其删除.

通过结合关系抽取和实体识别任务的损失函数,

得到三元组判别部分的损失函数 L_{triple} ,如式(19)所示.

$$L_{\text{triple}} = - \sum_{j=1}^M \sum_{i=1}^{N_j} \log(p_t(pla_i = y_i^t | S''_{\text{fusion}})) \quad (19)$$

其中, $y_i^t \in \{0, 1\}$ 是第 i 个句子的目标标签, M 为查询集中关系类别的数量, N 为训练数据中句子的数量.

3.4 损失函数

在关系三元组抽取任务中,关系抽取的目标是提取出查询实例中的关系类型.如果正确识别出查询实例中的关系类型,将大大地降低实体识别任务的难度.为了更好地平衡关系抽取和实体识别任务,设计了一种多任务动态学习的损失函数,使模型可以自动调整权重系数以优化整体损失,如式(20)所示:

$$L_{\text{total}} = e^{\lambda_1} L_{\text{rel}} + e^{\lambda_2} L_{\text{entity}} + e^{\lambda_3} L_{\text{triple}} \quad (20)$$

其中, λ_1 、 λ_2 和 λ_3 是可学习参数,初始值都设置为 0.

4 实验结果与分析

4.1 实验数据集

实验数据全部来源于中国各级政府官方网站,确保了数据来源的真实性与可靠性.本文成功爬取了 31 个省份,共计 902 份应急预案文件.在获取到原始应急预案文本后,首先采用多个正则表达式对原始文本进行数据清洗,将原始文本中包含的 HTML 标签和其他无关标签进行去除.随后,对处理好的文本进行句子级别的分割,以句号“。”为分隔符.根据专家的建议,定义了 20 种应急组织机构之间的关系类型,关系类型对应的数据分布如图 4 所示.

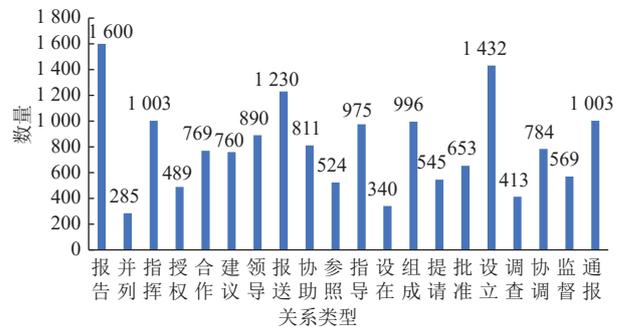


图4 ERPs+数据集中各种关系类型句子数量分布图

本文利用“阿里 NLP 自学习平台”对每个句子中的应急组织机构实体及其之间的关系进行标注.在公开数据集 FewRel 中,单个句子仅包含了一个三元组.为

了更贴近应急预案领域的实际情境, 本文在 FewRel 数据集标注规范的基础上进行了扩展. 在所标注的样本数据中, 单个句子可能涉及 1 个、2 个甚至更多个三元组, 这大大增加了模型进行关系三元组抽取的难度. 在实验中, 本文从 ERPs⁺数据集随机选择 11 种关系类型的数据作为训练集, 3 种关系类型的数据作为验证集, 剩下的 6 种关系类型作为测试集.

4.2 评价指标

本文采用 $F1$ 值作为评价指标, 如式 (21) 所示. 定义模型预测的三元组与目标三元组完全相同时才被视为正确结果.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (21)$$

4.3 实验参数设置

FS-MTS 模型使用 AdamW 优化器, 初始学习率被设置为 $1E-5$, 使用 Linear Warmup 在 200 个预热步骤内调整学习率. 批次大小设置为 2, 最大句子长度设置为 128.

本文遵循广泛使用的小样本训练范式, 在训练阶段模拟 N -way K -shot 场景. 在每个任务中, 从训练集中随机抽取 N 种关系类型, 每种关系类型随机抽取 K 个实例. 将 N 和 K 设置为 3 和 5, 从而生成两种类型的小样本关系三元组抽取任务: 3-way 3-shot 和 5-way 5-shot. 经实验验证, 本文确定在训练集上进行 18 000 次迭代用于训练, 在测试集上进行 600 次迭代用于评估. 所有实验使用不同的随机种子重复 5 次, 以提供可靠的实

验结果.

FS-MTS 模型在服务器 Ubuntu 18.04 操作系统上运行, 处理器为 Intel(R) Xeon(R) Platinum 8255C CPU@2.50 GHz, 显卡为 NVIDIA GeForce RTX 2080 Ti.

4.4 模型对比及结果分析

为了验证 FS-MTS 模型的有效性, 将其与监督学习方法和小样本学习方法的基线模型进行比较.

(1) CasRel^[3]: 该方法设计了一种新型级联二进制标记框架用于关系三元组抽取任务.

(2) PRGC^[4]: 该方法从全新的视角将关系三元组抽取任务分解为关系判断、实体抽取和主客体对齐.

(3) MPE^[7]: 该方法通过使用混合原型学习机制将文本和实体、关系的知识进行关联, 并且还通过使用原型感知正则化来学习更具有代表性的原型表示.

(4) NNM^[11]: 一种基于最近邻匹配的小样本关系三元组抽取模型.

(5) PTN^[8]: 该方法通过在关系视角、实体视角和三元组视角之间不断转移, 最终完成关系三元组抽取任务. 此外, 该方法还利用关系三元组信息来验证所提取的三元组是否合理.

(6) RelATE^[12]: 该方法是一种利用关系引导的小样本关系三元组抽取模型, 通过引入一种新的任务分解策略来提高模型性能.

(7) FS-MTS: 本文所提出的模型.

表 1 显示了 FS-MTS 模型与其他对比模型在 ERPs⁺测试集上的实验结果.

表 1 各模型在 ERPs⁺数据集上的 $F1$ 分数 (%)

类别	模型	3-way 3-shot			5-way 5-shot		
		Relation	Entity	Triple	Relation	Entity	Triple
监督学习	CasRel	54.9±1.5	6.4±0.9	1.3±0.8	58.2±1.0	7.6±0.6	1.9±0.7
	PRGC	56.2±0.7	6.8±0.6	1.5±0.6	58.3±0.6	8.1±0.5	2.1±1.1
小样本学习	MPE	64.3±0.3	21.8±1.1	14.8±1.2	66.0±0.6	24.2±0.7	16.2±0.9
	NNM	63.7±0.3	23.2±0.8	15.2±0.7	65.6±1.7	28.5±0.5	18.7±0.6
	RelATE	67.2±0.6	29.1±0.9	19.2±0.6	68.3±0.4	33.7±0.6	23.5±1.1
	PTN	67.8±0.5	29.4±0.4	19.6±0.7	69.1±0.8	34.3±0.7	24.6±0.4
	FS-MTS	69.1±0.7	30.2±0.9	22.4±0.8	71.6±0.8	36.5±0.8	26.9±0.9

接下来, 将对表 1 中的实验结果进行深入分析.

(1) 在 ERPs⁺数据集的两种设置下, 监督学习方法 CasRel 和 PRGC 的性能较差, $F1$ 值相对较低. 下面将从两个方面分析造成性能较差的原因. 一方面, 这两种方法通常为每种关系类型维护一个令牌表, 表中的数据表示具有特定关系的两个实体的起始和结束标记的位置. 这意味着模型需要利用相关的关系感知表来进

行三元组提取. 然而, 在小样本学习中, 训练集和测试集中的关系类型是不相交的. 因此, 在这种情况下使用这些关系感知表进行关系三元组抽取是无效的. 另一方面, 这两个模型存在不可转移性的问题. 尽管它们在训练集上进行了相应的训练, 但由于在测试集上仅使用了少量的标记实例进行微调, 模型仍然无法准确地提取句子中的关系三元组.

(2) 在 3-way 3-shot 和 5-way 5-shot 设置下, MPE 模型在关系抽取上的 $F1$ 值均高于 NNM 模型, 但在三元组抽取上的 $F1$ 值均低于 NNM. 这主要因为 MPE 模型通过混合原型学习将文本和实体、关系的相关知识进行关联, 生成了更具有代表性的原型表示, 从而在关系抽取任务中表现更优. NNM 模型通过利用最近邻匹配算法, 优化 MPE 模型在实体识别任务中存在的局限性, 从而在该任务中表现出更优异的性能.

(3) RelATE 模型在三元组抽取任务上的 $F1$ 值比 NNM 模型高约 4%, 证明了“先关系后实体”方式的有效性. MPE、NNM 和 RelATE 模型在 FewRel 数据集中均表现出较为优异的性能, 但 FewRel 数据集中每句话仅包含一个三元组. 在单句话涉及多个三元组的应急预案文本中, 以上 3 个模型的性能均有所下降.

(4) PTN 模型在所有的对比模型中性能最佳, 但在关系抽取、实体识别和三元组抽取任务上的 $F1$ 值仍低于 FS-MTS 模型. 这主要因为 PTN 模型在关系抽取部分仅将实体信息嵌入到支持实例和查询实例中, 并利用简单的 MLP 层来判断查询实例中存在的关系类型. 当处理语义较为复杂的应急预案文本时, 模型在关系抽取任务上的性能大大下降. 此外, PTN 模型并没有考虑支持集和查询集之间的令牌语义相似性问题, 这导致模型在实体识别任务中存在一定的局限性. FS-MTS 模型在关系抽取部分将 BiLSTM 与注意力机制融合, 可以更全面的捕获应急预案文本的序列信息. 此外, FS-MTS 模型在实体识别部分设计了一种基于语义相似性推断的方法进行识别句子中存在的头实体和尾实体. 与 PTN 模型相比, FS-MTS 模型在关系、实体和三元组抽取的 $F1$ 值上展现出巨大的优势.

4.5 消融实验及结果分析

本节旨在通过消融实验来评估 FS-MTS 模型中不同组件的有效性. 以下是 FS-MTS 模型的 3 个变体.

(1) -BiLSTM+Attention: 移除关系抽取部分的 BiLSTM 和 Attention 组合.

(2) -Semantic similarity inference: 用基于原型网络的方法代替基于语义相似性推断的实体识别方法.

(3) -Multi-task dynamically learning loss: 删除多任务动态学习的损失函数, 将各权重设置为相同的值.

从表 2 的实验结果可以看出, 上述 3 个非完整模型在关系、实体和三元组抽取任务上的性能均低于完整的 FS-MTS 模型. 下面将深入分析消融实验结果.

表 2 消融实验结果 (5-way 5-shot, $F1$ 值) (%)

模型	Relation	Entity	Triple
-BiLSTM+Attention	69.5	35.2	24.9
-Semantic similarity inference	67.6	30.2	19.7
-Multi-task dynamically learning loss	68.6	34.1	25.6
FS-MTS	71.6	36.5	26.9

(1) 当移除关系抽取部分中的 BiLSTM 和 Attention 的组合模块时, 关系抽取的 $F1$ 值下降了 2.1%, 三元组抽取的 $F1$ 值降低了 0.9%. 这一结果表明, 通过引入 BiLSTM 和 Attention 组合模块, FS-MTS 模型能够更有效地捕捉应急预案文本的序列信息, 并增强了词与词之间的关联性, 从而提高了模型在关系抽取任务上的性能.

(2) 将基于语义相似性推断的方法替换为基于原型网络的方法后, 实体识别任务上的 $F1$ 值下降了 6.3%. 这证明了语义相似性推断方法在应急预案领域实体识别任务中的有效性. 此外, 相较于普通的原型网络方法, 基于语义相似性推断的方法更好地考虑了支持集和查询集令牌之间的语义相似性, 提升了实体识别任务的性能.

(3) 当多任务动态学习损失函数被移除后, 实体、关系和三元组这 3 个任务的权重被设置成相等的值. 由表 2 中的结果显示, 三元组抽取的 $F1$ 值下降了 1.3%. 这证明了动态学习损失函数在 FS-MTS 模型中起到了一定的效果.

4.6 不同数量关系三元组的分析

为了进一步验证 FS-MTS 模型从单个句子中抽取多个关系三元组的能力, 本节选用监督学习模型 PRGC 和小样本学习模型 PTN 作为对比方法. 根据句子中包含三元组的数量, 将句子分成 4 类: $N=1$ 、 $N=2$ 、 $N=3$ 和 $N \geq 4$, 其中 N 代表单个句子中包含关系三元组的数量. 实验结果如表 3 所示.

表 3 句子中不同数量三元组的 $F1$ 值分析 (%)

模型	$N=1$	$N=2$	$N=3$	$N \geq 4$
PRGC	1.2	0.6	0.5	0.3
PTN	25.9	26.3	22.5	18.9
FS-MTS	30.1	30.6	27.8	24.4

从表 3 中的数据可以看出, 本文提出的 FS-MTS 模型在 4 种情况下均表现出了较为优异的性能. 虽然 PRGC、PTN 和 FS-MTS 模型的性能都随着关系三元组数量的增加而下降, 但仍然可以从不同模型的性能变化中得出一些结论. 与其他对比方法相比, FS-MTS

模型在句子复杂性不断增加时,其性能影响最小.在以上4种情况中, $F1$ 值的最大差值来自于句子最复杂的情况($N \geq 4$),这充分证明了本文所提模型在处理单个句子中存在多个三元组情况下的有效性.

5 结束语

本文提出了一种基于模块转移和语义相似度推断的小样本关系三元组抽取方法.针对现有的小样本关系三元组抽取方法难以处理单句话中存在多个三元组的问题,采用了一种在关系抽取、实体识别和三元组判别3个模块之间不断转移的机制,能够高效地识别出查询实例中存在的多个三元组.其次,针对应急预案文本语义复杂的问题,在关系抽取部分将BiLSTM和注意机制融合,以更好地捕捉应急预案文本的序列信息.然后,针对现有小样本关系三元组抽取方法未考虑支持集和查询集之间令牌语义相似性的问题,设计了一种基于语义相似性推断的方法来识别句子中的实体.此外,在ERPs⁺数据集上的实验结果验证了FS-MTS模型的有效性.

参考文献

- 1 Ni WJ, Shen QL, Liu T, *et al.* Generating textual emergency plans for unconventional emergencies—A natural language processing approach. *Safety Science*, 2023, 160: 106047. [doi: 10.1016/j.ssci.2022.106047]
- 2 Wang YC, Yu BW, Zhang YY, *et al.* TPLinker: Single-stage joint extraction of entities and relations through token pair linking. *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona: International Committee on Computational Linguistics, 2020. 1572–1582.
- 3 Wei ZP, Su JL, Wang Y, *et al.* A novel cascade binary tagging framework for relational triple extraction. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020. 1476–1488.
- 4 Zheng HY, Wen R, Chen X, *et al.* PRGC: Potential relation and global correspondence based joint relational triple extraction. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, 2021. 6225–6235.
- 5 Mintz M, Bills S, Snow R, *et al.* Distant supervision for relation extraction without labeled data. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Suntec: Association for Computational Linguistics, 2009. 1003–1011.
- 6 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述. *软件学报*, 2021, 32(2): 349–369. [doi: 10.13328/j.cnki.jos.006138]
- 7 Yu HY, Zhang NY, Deng SM, *et al.* Bridging text and knowledge with multi-prototype embedding for few-shot relational triple extraction. *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona: International Committee on Computational Linguistics, 2020. 6399–6410.
- 8 Fei JB, Zeng WX, Zhao X, *et al.* Few-shot relational triple extraction with perspective transfer network. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. Atlanta: ACM, 2022. 488–498.
- 9 鄂海红, 张文静, 肖思琪, 等. 深度学习实体关系抽取研究综述. *软件学报*, 2019, 30(6): 1793–1818. [doi: 10.13328/j.cnki.jos.005817]
- 10 Cai R, Zhang XD, Wang HF. Bidirectional recurrent convolutional neural network for relation classification. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin: Association for Computational Linguistics, 2016. 756–765.
- 11 He XL, Song H, Cheng D, *et al.* Few-shot relational triple extraction with nearest neighbor matching. *Proceedings of the 2022 International Conference on Computer Graphics, Artificial Intelligence, and Data Processing*. Harbin: SPIE, 2022. 262–266.
- 12 Cong X, Sheng JW, Cui SY, *et al.* Relation-guided few-shot relational triple extraction. *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Madrid: ACM, 2022. 2206–2213.
- 13 Han X, Zhu H, Yu PF, *et al.* FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels: Association for Computational Linguistics, 2018. 4803–4809.
- 14 Wang YQ, Yao QM, Kwok JT, *et al.* Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 2020, 53(3): 63.
- 15 Zheng SC, Xu JM, Zhou P, *et al.* A neural network framework for relation extraction: Learning entity semantic

- and relation pattern. Knowledge-based Systems, 2016, 114: 12–23. [doi: [10.1016/j.knosys.2016.09.019](https://doi.org/10.1016/j.knosys.2016.09.019)]
- 16 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional Transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019. 4171–4186.
- 17 Aytiran EF. Attention-based aspect sentiment classification using enhanced learning through CNN-BiLSTM networks. Knowledge-based Systems, 2022, 252: 109409. [doi: [10.1016/j.knosys.2022.109409](https://doi.org/10.1016/j.knosys.2022.109409)]
- 18 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc. , 2017. 6000–6010.
- 19 Jiang S, Zhu JZ, He LH. Few-shot relational triple extraction based on evaluation of token-level semantic similarity. Proceedings of the 32nd International Conference on Artificial Neural Networks. Heraklion: Springer, 2023: 232–242.
- 20 Yang Y, Katiyar A. Simple and effective few-shot named entity recognition with structured nearest neighbor learning. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2020. 6365–6375.

(校对责编: 张重毅)