

UFSC: 基于统一特征空间构建的零样本关系抽取

刘雨辰^{1,2*}, 段建勇^{1,2*}, 孙康³, 张晴^{1,2}, 何丽^{1,2}, 王昊^{1,2}, 刘杰^{1,2*}

¹北方工业大学信息学院, 北京100144

²CNONIX 国家标准应用与推广实验室, 北京100144

³江苏自动化研究所, 江苏, 连云港222006

1870806242@mail.ncut.edu.cn; duanjy@ncut.edu.cn; liujxxy@126.com

摘要

零样本关系抽取 (ZSRE) 旨在从可见关系中学习提取不可见关系的能力。一些研究表明: 将样本语句与关系描述匹配进而预测不可见关系的方法, 可以有效完成零样本关系抽取任务。然而, 现有的匹配框架方法很少统一样本语句与关系描述的特征空间, 缺乏对二者特征进行对齐。因此, 本文提出一种为匹配框架零样本关系抽取而设计的统一特征空间构建方法。统一样本语句与关系描述的编码模块, 并在此基础上引入特征相似损失。同时, 为了减轻特征在空间上的聚合现象, 引入特征均匀化模块, 旨在构建特征更加均匀化的特征空间。本文所提出的方法实现了性能上的提升。与之前最佳的结果相比, 在FewRel和Wiki-ZSL数据集上 F_1 值平均提高1.6%和3.4%, 体现了统一特征空间构建以及特征均匀化模块的有效性。

关键词: 关系抽取; 零样本; 特征空间

UFSC: A Unified Feature Space Construction for Zero-Shot Relation Extraction

Yuchen Liu^{1,2*}, Jianyong Duan^{1,2*}, Kang Sun³, Qing Zhang^{1,2},
Li He^{1,2}, Hao Wang^{1,2}, Jie Liu^{1,2*}

¹School of Information Science and Technology, North China University of Technology, Beijing 100144, China

²CNONIX National Standard Application and Promotion Lab, Beijing 100144, China

³Jiangsu Automation Research Institute of CSIC, Lianyungang 222006
1870806242@mail.ncut.edu.cn; duanjy@ncut.edu.cn; liujxxy@126.com

Abstract

Zero-shot relation extraction (ZSRE) seeks to develop the capability to extract hidden relations from observable ones. Previous studies have highlighted that methods matching sample statements with relation descriptions to predict hidden relations can effectively tackle the zero-shot relation extraction task. Nevertheless, current matching framework approaches often fail to integrate the feature space of sample statements and relation descriptions, thus lacking feature alignment between the two. Therefore, this paper introduces a unified feature space construction method tailored for zero-shot relation extraction within a matching framework. We integrate the encoding modules of sample statements and relation descriptions, supplemented with a feature similarity loss. Additionally, to address feature spatial aggregation issues, we introduce a feature homogenization module to foster a more consistent feature space. The method introduced in this paper achieves a performance improvement. The F_1 values on the FewRel and Wiki-ZSL datasets are improved by an average of 1.6% and 3.4% compared to the previous best results, demonstrating the effectiveness of the unified feature space construction as well as the feature homogenization module.

Keywords: Relation Extraction , Zero-Shot , Feature Space

1 引言

关系抽取 (RE) 是自然语言处理 (NLP) 的一项基本任务, 目的在于识别文本当中一对实体之间的语义关系。在高质量标记的数据上学习时, 关系抽取模型取得了较好的性能。然而, 新兴关系类型不断出现, 无休止地标注数据成本较高也不切实际。

为了处理缺乏标注数据的新兴关系类型, 零样本关系抽取 (ZeroRE) 最近受到较多关注。两项研究 (Chen and Li, 2021; Zhao et al., 2023) 使用固定的Sentence-BERT编码器 (Reimers and Gurevych, 2019) 对关系描述进行向量化表示, 在零样本条件下利用匹配式模型完成关系抽取任务。两项研究均取得较大的性能提升。

但是, 固定的关系描述编码模块也存在着一些局限性。首先, 模型只能对样本语句编码器进行训练, 使其生成的特征尽可能拟合描述编码器的特征空间。在此过程中, 由于描述特征不可学习的特点, 模型无法从描述中学习较多的特征信息, 对其在特征空间上的表示并不充分。而零样本任务下的已知信息较少, 信息中含有的特征较为珍贵。特征不充分的表达, 是低资源场景下的一种资源浪费。

同时, 固定的关系描述编码模块对特征空间的构建也不完善, 不利于模型区分特征相近的关系。如Table 1所示, 关系标签“Country”与“Country of citizens”, 二者的关系标签、描述都较为接近。在描述中, 虽然“Country”标识该关系不能作用于人类, 但是二者含有“state”“country”等较为相似的词汇, 不利于模型对二者的关系描述特征进行区分。并且, 二者之间的语义相似度也较高, 模型对于关系标签“Country”与“Country of citizens”之中的语义特征也较易混淆。

关系标签	关系描述
Country	country is sovereign state of this item; don't use on humans
Country of citizens	country of citizenship is the object is a country that recognizes the subject as its citizen

Table 1: 相似关系标签示例。

在特征空间上, 相似度更高的文本信息, 特征更为临近。而在预测阶段, 对于样本语句与关系描述的匹配, 也是通过计算样本与每个描述特征之间的相邻度完成。若两组描述的特征临近, 模型难以分辨样本应该与哪一组描述匹配。因此, 描述相邻度越高, 越易发生样本特征的匹配错误。

基于以上问题, 本文取消此前固定的关系描述编码模块, 将其设置为可学习状态, 使得模型在训练过程中可以充分学习到描述中词汇、语义等特征, 尽可能利用零样本场景下的已知信息。并且, 在此基础上统一样本语句与关系描述的特征空间, 使二者的特征匹配处于相同的特征空间进行, 提升模型匹配框架性能。同时, 提出一种特征空间构建方法, 使得模型对于特征空间上的相似性判断更加准确, 空间上的特征分布也更加均匀。具体而言, 引入特征相似损失替代先前的交叉熵损失, 将训练阶段中对于关系标签的多分类任务转化为对于样本特征与描述特征相似度判断的二分类任务, 更加拟合预测阶段模型对关系预测的过程。同时, 针对特征聚合现象, 设置特征均匀化模块。该模块将特征空间上较为临近的特征进行分散, 使得一个样本语句周围较少出现两个甚至多个相邻的关系描述特征, 进而提升模型在匹配过程中的准确度。

如Figure 1所示, 模型UFSC在相似关系标签“Country”与“Country of citizens”上的抽取性能, 较RE-Matching (Zhao et al., 2023)有较大提升。错误样本如Table 2所示, UFSC在“Country”标签中的FN错误样本相较于RE-Matching有大幅减少, 即标签

*通讯作者

该工作受到国家科技部科技创新2030-“新一代人工智能”重大项目(2020AAA0109700), 国家自然科学基金(61972003,62076167) 资助。

©2024 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

为“Country”样本被识别为其他标签的情况有所减少。同时，UFSC在“Country of citizens”标签中的FP错误样本有所减少，即其他标签匹配为“Country of citizens”的情况有所减少。这得益于UFSC中的描述特征均匀化方法可以对关系描述的特征有区别性更强特征表示。并且，训练阶段负样本检索方法挑选出的负样本迷惑性更强，UFSC在预测阶段对于负样本的识别率更高。

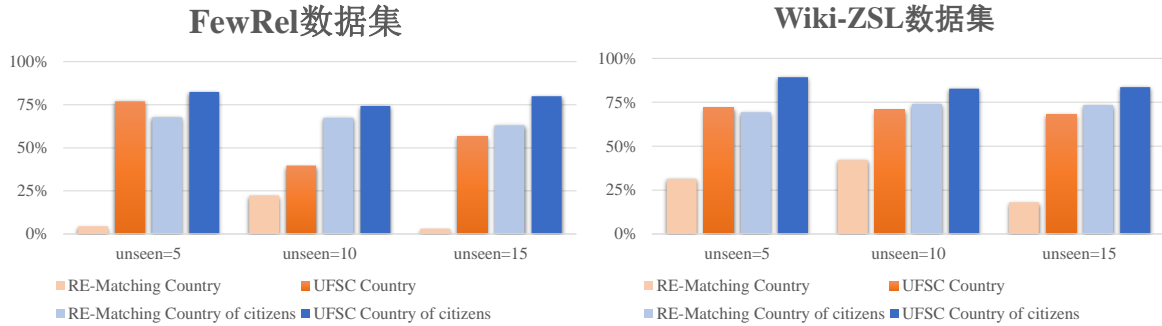


Figure 1: 相似关系标签的F1值对比。

数据集	不可见关系数量	关系标签	RE-Matching		UFSC	
			FP	FN	FP	FN
FewReL	unseen=5	Country	5	684	12	254
		Country of citizens	611	5	233	45
	unseen=10	Country	104	599	10	524
		Country of citizens	592	44	454	18
	unseen=15	Country	5	689	19	415
		Country of citizens	734	41	320	20
Wiki-ZSL	unseen=5	Country	170	781	171	342
		Country of citizens	854	16	232	7
	unseen=10	Country	84	711	191	348
		Country of citizens	658	24	397	15
	unseen=15	Country	73	893	228	362
		Country of citizens	662	37	317	54

Table 2: 相似关系标签的错误样本数量分析。

2 相关工作

2.1 关系抽取

关系抽取旨在根据相关上下文提取两个给定实体之间的关系，其在信息抽取和知识图谱的构建中发挥着关键作用。早期的方法涉及基于CNN、RNN (Zhang et al., 2017)和基于图 (Guo et al., 2021)的方法。随着预训练语言模型的发展 (Devlin et al., 2018)，使用预训练模型进行关系抽取任务已经成为一种主流范式。

2.2 零样本关系抽取

零样本关系抽取旨在识别出没有经过训练的新关系。Levy等人 (2017)将零样本关系抽取任务转化为问答形式。预先定义多个问题模板，将关系设置为问题答案，通过找到最符合给定问题的答案完成关系抽取任务。Obamuyide和Vlachos (2018)等人训练一个文本蕴含模型，来预测输入的关系描述是否可以从文本中推导出来。再通过训练阶段可见的语义关系推广到预测阶段不可见的语义关系。Chen和Li (2021)等人设计一个多任务学习模型，通过学习最小化样本与关系描述之间的距离，对关系进行分类。具体而言，将样本和关系描述映射到嵌入空间，利用最

邻近搜索来预测不可见的关系。Zhao和Zhan (2023)等人提出一种细粒度匹配方法，将句子间匹配分解为头尾实体匹配和上下文匹配。同时随机采样关系描述作为负样本对，基于匹配分数对不可见关系进行预测。

与上述方法相比，本文的方法有三个显著优势：

1. 统一样本语句与关系描述的特征空间，将先前样本语句特征单方向拟合关系描述特征空间的任务，变为构建一个样本语句与关系描述共享的特征空间。
2. 引入特征相似损失，使得模型在训练阶段学习判断相似特征的能力，进而提升预测阶段的泛化效果。
3. 模型对相似性较高的关系区分能力较强。通过设置特征均匀化模块，对特征空间上聚合度较高的特征进行分散。

3 方法

本节提出一种关系描述重构策略，丰富其中的可学习特征。同时，提出一种为匹配框架零样本关系抽取而设计的统一特征空间构建方法。通过设置可学习的关系描述编码模块将其与样本语句共享，并引入特征相似损失，构建统一的特征空间。提出一种特征均匀化方法。通过分散相似度较高的关系描述特征，并在负样本随机采样的基础上融合负样本检索，帮助模型推离错误匹配的描述特征，以得到更为均匀化的特征空间。

3.1 问题定义

零样本关系抽取 (ZeroRE) 任务的目标，是从可见关系集合 $R_s = \{r_1^s, r_2^s, \dots, r_n^s\}$ 学习到对应知识，并推理至不可见关系集合 $R_u = \{r_1^u, r_2^u, \dots, r_n^u\}$ 中。关系集合 R_s 和 R_u 是不相交的，即 $R_s \cap R_u = \emptyset$ 。并且在训练阶段只有可见关系集合 R_s 中的样本可以使用。本文将零样本关系抽取转化为特征匹配任务。在训练阶段，给定含有 n 个样本的训练集 $D_s = \{(t_i, t_i^{e1}, t_i^{e2}, r_i^s, d_i) | i = 1, \dots, n\}$ ，由样本语句 t_i 、头实体 t_i^{e1} 、尾实体 t_i^{e2} 、可见关系标签 r_i^s 及其对应的关系描述 d_i 组成。模型 M 在训练集上对可见关系 r_s 所对应的样本 t_i 与描述 d_i 进行特征匹配，学习统一的编码模块 E ，以得到更加准确的样本和描述的特征表示。在预测阶段，将含有 m 个样本的测试集 $D_u = \{(t_i, e_{1i}, e_{2i}) | i = 1, \dots, m\}$ 和 k 个描述 $d = \{d_1, d_2, \dots, d_k\}$ 输入至模型 M 中，找到与样本 t_i 特征相似度最高的描述 d_i ，将其对应关系 r_i^u 作为该样本的关系预测结果，组成关系三元组 $(t_i^{e1}, t_i^{e2}, r_i^u)$ 。模型整体结构如Figure 2所示。

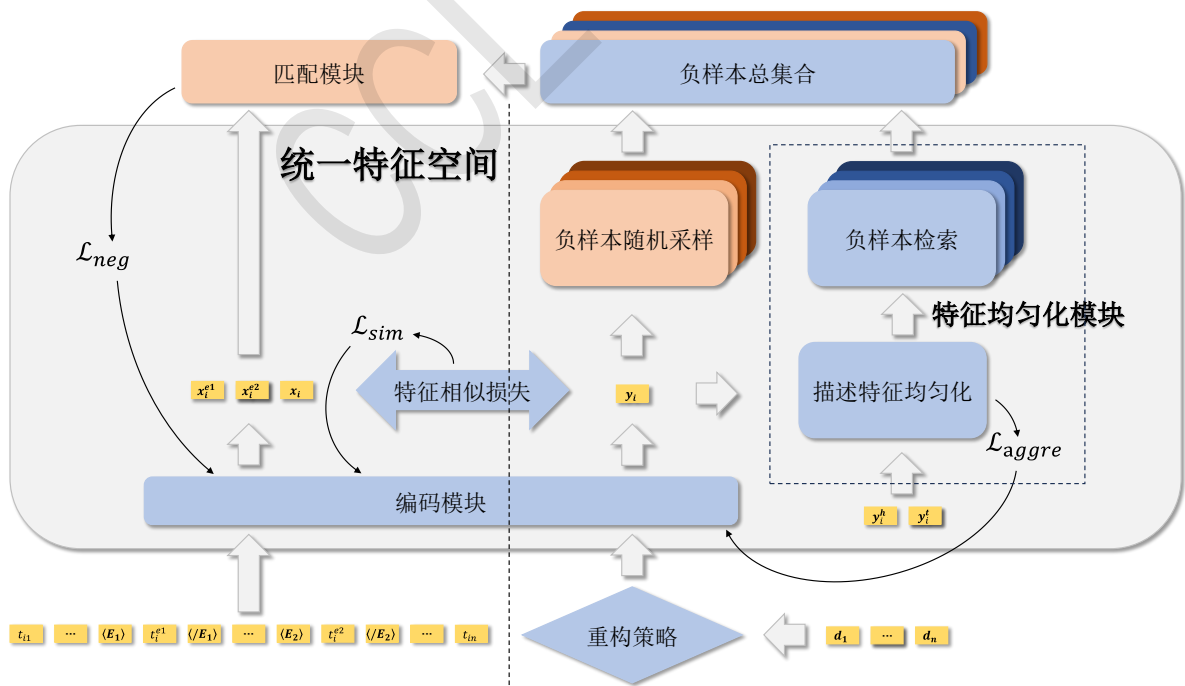


Figure 2: UFSC模型结构图。

3.2 关系描述重构策略

为了得到信息更为丰富的特征，模型对关系描述进行重新构建。将描述转化为提示与关系定义相结合的方式，在现有的基础上融合关系作用的信息。首先，设置提示模板，采用“the relation between head entity and tail entity is R_i ”的形式提示模型，头实体与尾实体之间的关系标签为 R_i 。之后，将标签与定义组合，采用“ R_i is D_i ”的形式，对标签的语义信息作进一步解释。

例如，对于标签“Mother”，给出新的描述“the relation between head entity and tail entity is mother, mother is female parent of the subject”。提示模板对描述的特征信息进行补充，在原有描述的基础上，加入关系作用，丰富描述的特征信息。定义对标签进行更深层次的解释，在特征空间上得到更加准确的描述特征，提高关系描述与样本语句的同义程度，使得二者的特征在特征空间上的相似度更高。

但是，丰富的特征在可学习样本较少的场景下，更易产生噪声。因此，模型描述重构策略为：只在不可见关系较少的场景下加入提示模板；对于不可见关系较多的场景，仅使用定义作为描述。

3.3 统一特征空间构建

模型将关系描述编码模块与样本语句编码模块统一，并引入特征相似损失，以帮助编码模块对关系标签相同的样本与描述，给出相似性更高的特征表示。

3.3.1 编码模块

为了尽可能利用关系描述中的特征信息，模型设置可学习的关系描述编码模块。假设输入关系描述 $d_i = [d_{i1}, d_{i2}, \dots, d_{in}]$ 包含 n 个词元。沿用Zhao和Zhan等人 (2023)的方法，对描述中的头尾实体进行特征表示，得到与 y_i^h 与 y_i^t 。然后使用Bert (Devlin et al., 2018)的`pooler_output`层来获得描述的特征表示 y_i 。

$$y_i = Bert_{encoder}(d_i) \quad (1)$$

统一的特征空间，更易于对样本与描述进行匹配。将描述编码模块与样本共享。假设输入文本 $t_i = [t_{i1}, t_{i2}, \dots, t_{in}]$ 包含 n 个词元， t_i^{e1} 和 t_i^{e2} 分别表示头实体和尾实体。首先，为了更好地从样本中提取实体特征信息，用实体特殊标记 $\langle E_1 \rangle$ 、 $\langle /E_1 \rangle$ 、 $\langle E_2 \rangle$ 、 $\langle /E_2 \rangle$ 分别标记 t_i^{e1} 和 t_i^{e2} 的起始和结束的位置 (Soares et al., 2019)。然后使用同一编码模块Bert来获得样本的隐藏层表示 h_i^t ，同时使用Maxpool和激活层tanh来获得头尾实体特征 x_i^{e1} 、 x_i^{e2} 。

$$h_{i1}^t, \dots, h_{in}^t = Bert_{encoder}(t_{i1}, \dots, \langle E_1 \rangle, t_i^{e1}, \langle /E_1 \rangle, \dots, \langle E_2 \rangle, t_i^{e2}, \langle /E_2 \rangle, \dots, t_{in}) \quad (2)$$

$$x_i^e = \tanh(\text{Maxpool}(h_{\langle E \rangle}, h_i^e, h_{\langle /E \rangle})) \quad (3)$$

与描述编码模块相同，将Bert的`pooler_output`层输出经过激活层后，作为样本的特征表示 x_i 。

$$x_i = \tanh(Bert_{encoder}(t_i)) \quad (4)$$

3.3.2 特征相似损失

为了更加拟合关系预测过程，模型引入特征相似损失 \mathcal{L}_{sim} 。计算关系描述 y_i 与样本语句 x_i 在特征空间中的相邻度 $\cos(x_i, y_i)$ ，将其作为特征相似性 $Sim(x_i, y_i)$ 。如果二者的特征相邻度越高，即特征在特征空间中夹角越小，就认为特征之间的相似性越高。

$$Sim(x_i, y_i) = \cos(x_i, y_i) \quad (5)$$

参照模型在预测阶段的匹配模式，设置特征相似损失 \mathcal{L}_{sim} 。具体而言，希望模型可以学习到关系标签相同的样本与描述之间更多的特征相似性。最理想的情况，标签相同的样本特征与描述特征完全重合，即相似性 Sim 取得相邻度最大值1。因此，将二者特征相似性向最大值1拟合，即计算特征相似性与最大相似性1之间的均方误差 MSE ，作为特征相似损失 \mathcal{L}_{sim} 。如此模

型 M 在训练过程中，通过不断减少均方误差值，优化其共享的编码模块。在预测阶段，编码模块对相同标签的样本与描述在特征空间上的特征表示，也更为相邻。由此提升模型的匹配准确率。

$$\mathcal{L}_{sim} = MSE(Sim, 1) \quad (6)$$

3.4 特征均匀化模块

如Figure 3所示，若描述特征 y_i 与 y_j 在特征空间中聚合度较高，与 y_i 标签相同的样本特征 x_{i1} 、 x_{i2} 在匹配预测过程中，可能会被模型错误认为与 y_j 相匹配。如果仅使用负样本对比机制，作为负样本的描述特征 y_j 可能只会在 x_{i1} 与 x_{i2} 之间反复波动，不跳出该范围而远离正样本 y_i 。针对以上情况，模型设置特征均匀化模块，由描述特征均匀化与负样本检索两部分组成，帮助模型区分聚合度较高的特征。

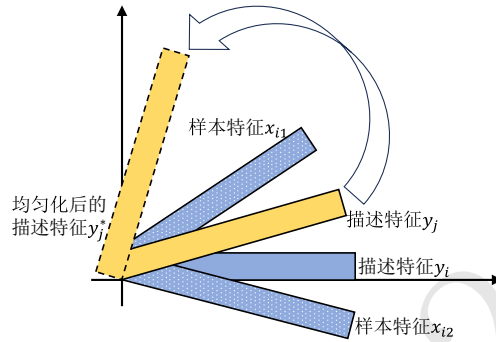


Figure 3: 描述特征聚合现象。

3.4.1 描述特征均匀化

模型通过计算特征夹角的方式，得到描述特征空间附近的特征聚合度 $Aggre$ 。

$$Aggre_{ij}(y_i, y_j) = (1 - 2 \cdot \alpha) \cdot \cos(y_i, y_j) + \alpha \cdot \cos(y_i^h, y_j^h) + \alpha \cdot \cos(y_i^t, y_j^t) \quad (7)$$

例如，对于关系描述 y_i ，将其余所有关系描述 $y_{aggre} = \{y_1, y_2, \dots, y_k | y_i \notin y\}$ 看作是易聚合描述集合。计算 y_i 与集合中每个负样本之间的特征聚合度 $Aggre_i = \{Aggre_{i1}(y_i, y_1), Aggre_{i2}(y_i, y_2), \dots, Aggre_{ik}(y_i, y_k)\}$ 。之后，从聚合度集合 $Aggre_i$ 中找到与 y_i 聚合度最高的特征 y_{max} 。将其聚合度与最高聚合限度 $margin$ 进行比较。若其值小于 $margin$ ，则表示 y_{max} 与 y_i 在特征空间聚合度较高，在之后匹配过程中易出现混淆现象。同时设置描述特征均匀损失 \mathcal{L}_{aggre} 。

$$\mathcal{L}_{aggre} = \max(Aggre_{imax}(y_i, y_{max}) - margin, 0) \quad (8)$$

3.4.2 负样本检索

先前任务的负样本采样机制为随机采样 (Zhao et al., 2023)。由于随机性的存在，模型并不能准确挑选出对于样本特征 x_i 最具迷惑性的描述特征 y_j ，如Figure 4所示。因此，若仅沿用负样本随机采样机制，匹配过程中仍易发生样本特征与错误描述特征相互匹配的现象。对此，模型在负样本随机采样的基础上融合负样本检索。

在原先含有 ra 个负样本随机采样集合 y_{random} 中加入负样本检索集合 $y_{retriever}$ ，共同组成负样本总集合 y_{neg} 。其中，集合 $y_{retriever}$ 由与描述特征 y_i 聚合度最高的 re 个描述特征组成。

$$y_{neg} = y_{random} \cup y_{retriever} \quad (9)$$

沿用Zhao和Zhan等人 (2023)的方法，使用细粒度匹配机制计算样本特征 x_i 与负样本集合 y_{neg} 中每个负样本的相似分数 $Score_{ij}(x_i, y_j)$ ，并计算负样本损失 \mathcal{L}_{neg} 。其中 $interval$ 为超参数，表示样本特征与描述特征之间的最小相邻间隔。

$$Score_{ij}(x_i, y_j) = (1 - 2 \cdot \beta) \cdot \cos(x_i, y_j) + \beta \cdot \cos(x_i^{e1}, y_j^h) + \beta \cdot \cos(x_i^{e2}, y_j^t) \quad (10)$$

$$\mathcal{L}_{neg} = \max(Score_{ii}(x_i, y_i) - Score_{imax}(x_i, y_{max}) + interval, 0) \quad (11)$$

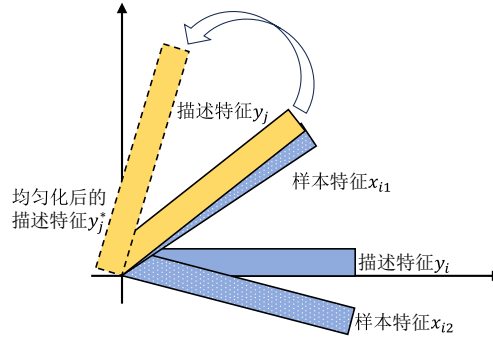


Figure 4: 样本描述特征聚合现象。

3.5 模型训练与预测

在训练阶段，模型按先后顺序分别对 \mathcal{L}_{sim} 、 \mathcal{L}_{aggre} 、 \mathcal{L}_{neg} 进行梯度优化。在预测阶段，模型计算样本特征 x_i 与关系描述特征集合 $y = \{y_1, \dots, y_i, \dots, y_k\}$ 中每个关系描述之间的特征相似分数，组成集合 $Score_i = \{Score_{i1}(x_i, y_1), \dots, Score_{ii}(x_i, y_i), \dots, Score_{ik}(x_i, y_k)\}$ 。之后，从集合 $Score_i$ 中找到最高分数 $Score_{imax}$ ，将 y_{max} 对应的关系标签 r_{max} 作为样本语句 t_i 的预测结果，并与其头尾实体 t_i^{e1} 、 t_i^{e2} 组成关系三元组 $(t_i^{e1}, t_i^{e2}, r_{max})$ 。

4 实验

4.1 实验设置

本文分别使用两个数据集对UFSC模型进行评估：FewRel (Han et al., 2018) 和 Wiki-ZSL (Chen and Li, 2021)。FewRel 是一个基于Wikipedia的开放数据集，由许多数据标注者手动标注而成，其包含80种关系标签，每种关系有700个样本。目前FewRel广泛用于小样本学习任务。本文使其在训练阶段与预测阶段的关系标签集合不相交，将其应用于零样本学习任务。Wiki-ZSL 是一个基于Wikidata知识库远程监督生成的数据集，其包含113种关系标签和93383个样本。与FewRel相比，Wiki-ZSL有着更多样的关系标签，但是其样本内部存在更多噪声。

与以往的研究者一样，本文选择5个关系标签作为验证集， $unseen \in \{5, 10, 15\}$ 为测试集中不可见关系标签的数量，并将剩余的关系标签作为可见关系标签对模型进行训练。同时，为了保证实验结果的可靠性，每组实验均重复进行5次，每次对关系标签重新选择，将平均结果作为最终实验结果。

4.2 基线模型

为了更好地评估方法的有效性，本文分别与较为主流的生成式方法和最先进的基于匹配的零样本关系抽取方法进行对比。同时，还比较了最近较为流行的基于大模型的零样本关系抽取方法。

- ZS-BERT (Chen and Li, 2021): 一种基于匹配的零样本关系抽取方法，分别采用BERT和Sentence-Bert作为编码器对样本和关系描述进行编码。除了分类损失外，还使用对比学习的方法来优化特征空间表示。
- PromptMatch (Sainz et al., 2021): 一种SOTA的生成式零样本关系抽取方法，对样本进行细粒度建模。

- REPrompt (Chia et al., 2022): 一种基于生成关系伪数据的零样本关系抽取方法。
- SUMASK (Li et al., 2023): 一种基于大模型的零样本关系抽取方法, 将关系抽取任务转换为多轮问答任务。通过设置不同的问题模板, 利用大模型进行样本信息总结、二分类关系判断和语句匹配, 最终对比语句匹配度完成对关系标签的抽取。
- RE-Matching (Zhao et al., 2023): 一种SOTA的基于匹配的零样本关系抽取方法, 提出一种细粒度匹配机制, 充分利用实体中的特征信息。

4.3 实验细节

基线模型实验中大部分使用Bert作为样本编码器, 为了公平比较并统一样本与描述的编码粒度, 本文使用Bert-base-uncased作为样本语句和关系描述的编码器。选取AdamW (Loshchilov and Hutter, 2017)作为优化器, batchsize为24。当 $unseen \in \{5, 10, 15\}$ 时, FewRel数据集上epoch分别为2、3、3, Wiki-ZSL数据集上epoch分别为2、2、3。训练过程中, 学习率为 $2e-5$ 。同时, 采用预热学习率策略, 预热系数为0.2。超参数 α 与 β 均为0.2, 负样本随机采样数量 ra 为7, 负样本检索数量 re 为5, 最高聚合限度 $margin$ 为0.95, 最小相邻间隔 $interval$ 为0.06。在 $unseen \in \{5, 10\}$ 时, 使用提示加关系定义组成关系描述; 在 $unseen = 15$ 时, 仅使用关系定义作为关系描述。所有实验均使用NVIDIA GeForce RTX 3090进行。

4.4 主要实验结果

$unseen$	方法	Wiki-ZSL			FewRel		
		<i>Prec.</i>	<i>Rec.</i>	F_1	<i>Prec.</i>	<i>Rec.</i>	F_1
5	ZS-Bert (Chen and Li, 2021)	71.54	72.39	71.96	76.96	78.86	77.90
	PromptMatch (Sainz et al., 2021)	77.39	75.90	76.63	91.14	90.86	91.00
	REPrompt (Chia et al., 2022)	70.66	83.75	76.63	90.15	88.50	89.30
	SUMASK (Li et al., 2023)	75.64	70.96	73.23	78.27	72.55	75.30
	RE-Matching (Zhao et al., 2023)	78.19	78.41	78.30	92.82	92.34	92.58
	UFSC(ours)	85.84	84.43	85.11	92.74	92.31	92.53
10	ZS-Bert (Chen and Li, 2021)	60.51	60.98	60.74	56.92	57.59	57.25
	PromptMatch (Sainz et al., 2021)	71.86	71.14	71.50	83.05	82.55	82.80
	REPrompt (Chia et al., 2022)	68.51	74.76	71.50	80.33	79.62	79.96
	SUMASK (Li et al., 2023)	62.31	61.08	61.69	64.77	60.94	62.80
	RE-Matching (Zhao et al., 2023)	74.39	73.54	73.96	83.21	82.64	82.93
	UFSC(ours)	75.37	76.58	75.95	85.82	85.97	85.89
15	ZS-Bert (Chen and Li, 2021)	34.12	34.38	34.25	35.54	38.19	36.82
	PromptMatch (Sainz et al., 2021)	62.13	61.76	61.95	72.83	72.10	72.46
	REPrompt (Chia et al., 2022)	63.69	67.93	65.74	74.33	72.51	73.40
	SUMASK (Li et al., 2023)	43.55	40.27	41.85	44.76	41.13	42.87
	RE-Matching (Zhao et al., 2023)	67.31	67.33	67.32	73.80	73.52	73.66
	UFSC(ours)	68.61	68.86	68.72	76.02	75.14	75.57

Table 3: 模型在两个零样本关系抽取数据集上的主要实验结果。

UFSC在Wiki-ZSL和FewRel数据集的实验结果如Table 3所示。从实验结果中可以发现, 在针对不同数量的不可关系 $unseen$ 时, UFSC在FewRel与Wiki-ZSL上的表现均优于现有基线。与之前最佳的结果相比, UFSC在FewRel上 F_1 值平均提高1.6%, 在Wiki-ZSL上 F_1 值平均提高3.4%。ZS-Bert是针对零样本关系抽取任务设计的, 但是关系描述编码模块在训练过程中固定不变, 对描述的特征表示较为单一, 导致其性能不佳。PromptMatch使用自注意力机制对样本特征进行建模, UFSC并未对样本特征进行去噪或细粒度建模操作, 但性能仍有较大提升。REPrompt在训练过程中使用新的关系伪数据对模型进行训练。与之相比, UFSC在训练过程中对数据的依赖性更小, 且能够达到更优的性能。与SOTA方法RE-Matching相

比，UFSC统一样本语句与关系描述的特征空间，使其在预测阶段取得更佳的匹配性能。同时，UFSC的特征均匀化模块也可以带来更加均匀的特征空间分布，对相似关系标签的分辨能力更佳。SUMASK在抽取过程中引入大模型交互问答，并与UFSC一样使用语句匹配的方式完成关系抽取任务。然而UFSC并未使用大模型的外部知识库对特征信息进行总结，也并未使用大模型识别语句间的匹配程度，但是其性能较SUMASK有很大提升。这得益于统一特征空间的构建与特征均匀化操作，使得样本特征与描述特征相统一并都有着更为准确的特征表示。

4.5 消融实验

4.5.1 关系描述重构策略消融实验

本节设置三组实验来讨论不同结构的关系描述对于模型性能的影响。Prompt组使用提示与关系定义组成关系描述，例如对于关系标签“Mother”，使用描述“the relation between head entity and tail entity is mother, mother is female parent of the subject”；Label组使用标签与关系定义组成关系描述，例如“mother is female parent of the subject”；NoPrompt组仅使用关系定义作为关系描述，如“mother: mother is female parent of the subject”。三组实验中，模型仅采用不同关系描述组成方法，其余部分均保持一致。实验结果如表Table 4所示。

<i>unseen</i>	方法	Wiki-ZSL			FewRel		
		<i>Prec.</i>	<i>Rec.</i>	F_1	<i>Prec.</i>	<i>Rec.</i>	F_1
5	UFSC-Prompt	85.84	84.43	85.11	92.74	92.31	92.53
	UFSC-NoPrompt	83.54	82.39	82.94	91.85	91.35	91.60
	UFSC-Label	83.60	81.90	82.71	91.30	90.67	90.99
10	UFSC-Prompt	75.37	76.58	75.95	85.82	85.97	85.89
	UFSC-NoPrompt	73.59	74.39	73.97	85.64	84.32	84.96
	UFSC-Label	73.58	74.48	74.01	85.72	84.80	85.25
15	UFSC-Prompt	67.44	68.94	68.16	74.92	74.70	74.79
	UFSC-NoPrompt	68.61	68.86	68.72	76.02	75.14	75.57
	UFSC-Label	66.62	67.52	67.03	74.92	74.33	74.62

Table 4: 关系描述不同构造方法的比较。

以数据集FewRel中的结果为例，在 $unseen \in \{5, 10\}$ 时，Prompt组相比于NoPrompt组，精准率、召回率和 F_1 值均有所提升。这得益于描述中所含特征信息更为多样。而在不可见关系较多时，NoPrompt组由于描述中的噪声较少，泛化能力优于Prompt组。因此，关系描述策略采用在预测空间较小时，重构关系描述以提升描述信息的丰富性；同时在预测关系较多时，仅采用关系定义作为关系描述，以避免不必要噪声的引入。

对比Prompt组与Label组可以发现，无论 $unseen$ 取何值，Prompt组的实验结果均优于Label组。相比于仅强调标签信息，引入关系作用信息可以帮助模型在低资源的场景下获得更多粒度的可学习信息。证明所引入提示的优越性。

关系标签	方法	Wiki-ZSL			FewRel		
		<i>Prec.</i>	<i>Rec.</i>	F_1	<i>Prec.</i>	<i>Rec.</i>	F_1
Country	UFSC-Prompt	77.32	65.17	70.72	94.92	34.71	50.84
	UFSC-NoPrompt	73.30	25.83	38.19	97.59	23.14	37.41
Country of citizens	UFSC-Prompt	71.23	98.50	82.67	63.78	95.86	76.60
	UFSC-NoPrompt	57.37	95.89	71.79	60.14	96.57	74.12

Table 5: $unseen = 10$ 时，关系描述重构对于相似关系标签的比较。

并且，引入关系作用信息可以帮助模型更准确地识别相似关系标签。如表Table 5所示，在 $unseen = 10$ 时，Prompt组对于关系标签“Country”的召回率相较NoPrompt组有着大幅提升，即Prompt组对于“Country”标签的正样本识别率更高。这得益于引入关系作用信息后模型对于“Country”标签的关系描述特征有着更准确地表示，进而提高“Country”标签的关系描述与

其正样本在特征空间上的同义程度，使得模型对于“Country”与“Country of citizens”这对相似标签的有着更强的区分能力。

4.5.2 特征均匀化模块消融实验

为探究描述特征均匀化方法与负样本检索方法对模型性能的影响，本节设置三组消融实验。实验结果如Table 6所示。NoHomogenization组完全去除特征均匀化模块；OnlyDistinguish组仅去除负样本检索方法；OnlyRetriever组仅去除描述均匀化方法。

<i>unseen</i>	方法	Wiki-ZSL			FewRel		
		<i>Prec.</i>	<i>Rec.</i>	F_1	<i>Prec.</i>	<i>Rec.</i>	F_1
5	UFSC	85.84	84.43	85.11	92.74	92.31	92.53
	NoHomogenization	83.74	81.47	82.55	89.76	88.98	89.37
	OnlyDistinguish	84.45	83.33	83.87	90.92	89.99	90.45
	OnlyRetriever	84.25	82.15	83.15	91.54	91.06	91.30
10	UFSC	75.37	76.58	75.95	85.82	85.97	85.89
	NoHomogenization	73.16	74.53	73.82	85.46	84.85	85.15
	OnlyDistinguish	74.70	75.68	75.18	84.78	84.34	84.56
	OnlyRetriever	73.72	74.26	73.98	86.04	85.44	85.73
15	UFSC	68.61	68.86	68.72	76.02	75.14	75.57
	NoHomogenization	64.91	64.72	64.79	76.11	74.55	75.31
	OnlyDistinguish	68.09	68.28	68.16	73.97	73.07	73.51
	OnlyRetriever	67.33	67.10	67.19	75.17	74.55	74.85

Table 6: 特征均匀化模块不同构造方法的比较。

实验结果表明，在大部分情况，使用描述特征均匀化方法与负样本检索均可以提升模型预测性能。并且，在一些场景下，两种方法相互补充，同时使用可以更加显著的提升模型性能。而在Wiki-ZSL数据集中，描述均匀化方法对模型性能有着更显著的提高。这是由于Wiki-ZSL中的样本存在更多噪声。因此，对描述特征的分辨可以将关系标签相同的样本特征进行聚类，不同标签的样本特征在空间上的分布间距更大，减小样本噪声对匹配过程的影响。

4.5.3 统一特征空间构建消融实验

<i>unseen</i>	方法	Wiki-ZSL			FewRel		
		<i>Prec.</i>	<i>Rec.</i>	F_1	<i>Prec.</i>	<i>Rec.</i>	F_1
5	UFSC	85.84	84.43	85.11	92.74	92.31	92.53
	UFSC-LabelLoss	73.85	70.48	72.10	78.66	76.37	77.48
	1-Encoder	64.72	60.34	62.34	73.88	60.60	66.10
	2-Encoder	57.44	55.32	56.33	83.45	81.16	82.26
10	UFSC	75.37	76.58	75.95	85.82	85.97	85.89
	UFSC-LabelLoss	51.01	51.97	51.47	63.45	61.80	62.56
	1-Encoder	50.72	42.78	46.13	64.01	46.12	53.56
	2-Encoder	63.40	61.31	62.24	71.25	65.94	68.45
15	UFSC	68.61	68.86	68.72	76.02	75.14	75.57
	UFSC-LabelLoss	51.98	50.92	51.43	56.09	54.19	55.08
	1-Encoder	51.41	41.05	45.45	53.10	41.13	46.24
	2-Encoder	55.04	52.46	54.65	61.44	56.58	58.82

Table 7: 是否统一特征空间方法的比较。

本节研究特征空间是否统一对模型性能的影响。消融实验的结果如Table 7所示。1-Encoder组使用可学习的Bert对样本进行编码，使用固定的Sentence-Bert (Reimers and Gurevych, 2019)对描述进行编码；2-Encoder组使用两个独立的Bert分别对样本与描述进行

编码。带有LabelLoss的实验组，使用关系标签多分类损失替代特征相似损失，以验证特征相似损失为模型性能带来的提升。

在所有场景下，UFSC的性能均优于1-Encoder组。这是由于统一的特征空间帮助相同标签的样本与描述获得较为统一的特征表示，使得二者在后续的特征相似性计算中取得较好的成绩。对比1-Encoder组和2-Encoder组结果可以发现，可学习描述编码模块的性能在大部分场景下优于固定的描述编码模块，可以更为充分地学习描述中的特征信息。通过对比UFSC和LabelLoss组结果可以发现，相较于关系标签多分类损失，特征相似损失提升标签相同的样本特征与描述特征在特征空间上的相似性，在统一特征空间的场景下对模型性能有着较大提升。

5 结论

本文提出了一种基于统一特征空间构建的零样本关系抽取方法。通过统一样本语句与关系描述的编码模块并引入特征相似度损失，模型构建了样本与描述相统一的特征空间。同时，设置特征均匀化模块并引入负样本检索，帮助模型实现特征分布更加均匀的特征空间构建。

与当今的生成式大模型相比，UFSC体量较小可以在本地部署，便于在现有基础上进行微调，可以较好应用于某一特定领域。但是，大模型作为一种强大的外部知识库，可以在模型零样本学习的过程中，为模型补充更多的外部知识。UFSC中并未使用任何外部知识，因此其在零样本关系抽取性能仍具备提升空间。

而更丰富的特征信息可能带来更多的噪声，如何进行特征去噪仍然是一个值得研究的方向。大模型所提供的外部知识或许可以帮助识别特征中的噪声，进一步提高UFSC在抽取过程中的特征简洁度。同时，零样本关系抽取要求模型具备较好的泛化能力。因此，特征去噪是否同样具备可迁移性也是一个值得探讨的问题，将此留作未来的工作。

参考文献

- Chih-Yao Chen and Cheng-Te Li. 2021. Zs-bert: Towards zero-shot relation extraction with attribute representation learning. *arXiv preprint arXiv:2104.04697*.
- Yew Ken Chia, Lidong Bing, Soujanya Poria, and Luo Si. 2022. Relationprompt: Leveraging prompts to generate synthetic data for zero-shot relation triplet extraction. *arXiv preprint arXiv:2203.09101*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Zhijiang Guo, Guoshun Nan, Wei Lu, and Shay B Cohen. 2021. Learning latent forests for medical relation extraction. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 3651–3657.
- Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. 2018. Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. *arXiv preprint arXiv:1810.10147*.
- Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. Zero-shot relation extraction via reading comprehension. *arXiv preprint arXiv:1706.04115*.
- Guozheng Li, Peng Wang, and Wenjun Ke. 2023. Revisiting large language models as zero-shot relation extractors. *arXiv preprint arXiv:2310.05028*.
- Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Abiola Obamuyide and Andreas Vlachos. 2018. Zero-shot relation classification as textual entailment. In *Proceedings of the first workshop on fact extraction and VERification (FEVER)*, pages 72–78.
- Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.
- Oscar Sainz, Oier Lopez de Lacalle, Gorka Labaka, Ander Barrena, and Eneko Agirre. 2021. Label verbalization and entailment for effective zero-and few-shot relation extraction. *arXiv preprint arXiv:2109.03659*.

- Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. *arXiv preprint arXiv:1906.03158*.
- Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D Manning. 2017. Position-aware attention and supervised data improve slot filling. In *Conference on empirical methods in natural language processing*.
- Jun Zhao, Wenyu Zhan, Xin Zhao, Qi Zhang, Tao Gui, Zhongyu Wei, Junzhe Wang, Minlong Peng, and Mingming Sun. 2023. Re-matching: A fine-grained semantic matching method for zero-shot relation extraction. *arXiv preprint arXiv:2306.04954*.