

基于对比学习和焦点损失的持续关系抽取^①

王苏越, 马丽丽, 陈金广

(西安工程大学 计算机科学学院, 西安 710600)
通信作者: 马丽丽, E-mail: xamll@163.com



摘要: 持续关系抽取旨在训练模型从不断变化的数据流中学习新关系, 同时保持对旧关系的准确分类。然而, 由于神经网络的灾难性遗忘问题, 模型在学习完新关系之后, 对旧关系的识别能力往往会大幅度降低。为了缓解灾难性遗忘对模型性能的影响, 本文提出了一种基于对比学习和焦点损失的持续关系抽取方法。首先, 在训练集与其增强样本集的并集上训练模型, 以学习新任务; 其次, 从训练集中, 为每个新关系选取并存储记忆样本; 然后, 将激活集中的示例与所有已知关系原型进行对比, 以学习新旧关系; 最后, 利用关系原型进行记忆再巩固, 并引入焦点损失提高模型对相似关系的区分。在 TACRED 数据集上进行实验, 结果表明本文方法能够进一步缓解灾难性遗忘问题, 提升模型的分类能力。

关键词: 持续关系抽取; 灾难性遗忘; 对比学习; 焦点损失; 关系原型

引用格式: 王苏越, 马丽丽, 陈金广. 基于对比学习和焦点损失的持续关系抽取. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9552.html>

Continual Relation Extraction Based on Contrastive Learning and Focal Loss

WANG Su-Yue, MA Li-Li, CHEN Jin-Guang

(School of Computer Science, Xi'an Polytechnic University, Xi'an 710600, China)

Abstract: Continual relation extraction aims to train models to learn new relations from evolving data streams while maintaining accurate classification of previously learned relations. However, due to the catastrophic forgetting problem of neural networks, the model's ability to recognize old relations tends to decrease drastically after being trained on new relations. To mitigate the impact of catastrophic forgetting on model performance, this study proposes a continual relation extraction method based on contrastive learning and focal loss. First, the model is trained on a concatenated set of the original training set and its augmented samples to learn a new task. Second, from the training set, memory samples are selected and stored for each new relation. Then, instances from the activation set are contrasted with all known relation prototypes to learn the old and the new relations. Finally, memory reconsolidation is performed using the relation prototypes and focal loss is introduced to improve the model's distinction between similar relations. Experiments are conducted on the TACRED dataset, and the results show that the method proposed can further alleviate catastrophic forgetting and improve the model's classification ability.

Key words: continual relation extraction; catastrophic forgetting; contrastive learning; focal loss; relation prototype

随着互联网技术的发展, 海量的半结构化和非结构化文本呈现在大众视界。如何迅速地从这些文本中抽取有价值的信息, 是学术界关注的重要问题。关系

抽取 (relation extraction, RE)^[1,2] 是信息抽取的重要子任务之一, 其目的是识别出文本中两个或多个实例之间的关系, 现阶段已被广泛应用于知识图谱^[3,4]、问答系

① 基金项目: 陕西省自然科学基金基础研究计划 (2023-JC-YB-568); 陕西省教育厅科研计划 (22JP028); 陕西省计算机学会&翔腾公司基金 (XT-QC-202309-119287)

收稿时间: 2024-01-03; 修改时间: 2024-02-04; 采用时间: 2024-02-26; csa 在线出版时间: 2024-05-31

统^[5,6]等诸多下游任务中。

传统的 RE 方法,通常假设待预测的关系属于一组固定的预定义关系集,并且要求模型在静态数据集上进行一次性训练.这种方式得到的模型,虽然能够完成简单的评估,但是不适用于关系和数据不断迭代更新的实际场景.此外,每当出现新关系时,存储所有数据并重新训练整个模型,需要高昂的计算成本.为了实现高效的学习,学者们提出了持续关系抽取 (continual relation extraction, CRE)^[7,8]范式.与传统的 RE 相比,CRE 侧重于在连续不断的数据流上训练模型,以学习新关系的同时保持对旧关系的稳定理解.而实现 CRE 需要通过持续学习 (continual learning, CL) 来进行建模.

持续学习^[9-12],也被称为终身学习 (lifelong learning, LL),其面临的主要挑战是灾难性遗忘 (catastrophic forgetting, CF)^[13,14],即在学习完新任务之后,模型在旧任务上的性能会严重下降.最近的研究尝试通过以下 3 种方式解决灾难性遗忘问题:(1) 基于记忆的方法^[15,16],存储先前任务的典型实例,并在学习新任务时重新训练;(2) 基于正则化的方法^[17,18],通过对旧任务的重要权重施加约束来缓解灾难性遗忘;(3) 基于架构的方法^[19],动态调整模型架构来学习新任务并防止遗忘旧任务.其中,基于记忆的方法已被证明在自然语言处理任务中很有前景.

近年来,基于记忆的持续关系抽取方法取得重大进展. Han 等人^[20]受人类长期记忆形成机制的启发,将情景记忆激活和再巩固 (episodic memory activation and reconsolidation, EMAR) 引入 CRE,并采用多轮联合训练机制来强化模型对旧关系的长期记忆. Cui 等人^[21]为了能够充分利用记忆样本的信息,将原型网络引入 CRE,并利用关系原型来改进样本嵌入,以更好地恢复嵌入空间的中断. Zhao 等人^[22]提出了一致性表示学习方法 (consistent representation learning, CRL),该方法利用监督对比学习和知识蒸馏来保持旧关系表示的一致性,从而实现持续学习. Hu 等人^[23]设计了原型对比网络 (continual relation extraction through prototypical contrastive learning, CRECL),该网络通过给定示例与所有已知关系的原型进行对比,充分利用负相关信息,来使新旧任务的数据分布更容易区分.

上述的 CRE 方法将灾难性遗忘问题归因于新关系的出现导致旧关系表征损坏,并隐晦地假设 CRE 模

型已经充分学习了旧关系.但是, Wang 等人^[24]通过实验证明这一假设可能不成立,并认为造成灾难性遗忘的一个重要原因是,学到的关系表示在后续学习过程中对于相似关系的出现没有很好的鲁棒性.为了应对这一挑战,他们提出了一种对抗类增强机制 (adversarial class augmentation, ACA),以加强模型对相似关系的区分能力.后来, Zhao 等人^[25]为了提高模型在相似关系上的性能,引入了综合训练和焦点知识蒸馏,并设计了记忆不敏感关系原型和记忆增强策略,以克服记忆样本过拟合问题.

受上述文献的启发,针对灾难性遗忘问题,本文以添加了对抗类增强机制 ACA 的 EMAR 模型 (EMAR-ACA) 为基础,提出了一种基于对比学习和焦点损失的持续关系抽取方法 CF-CRE (contrastive learning and focal loss for continual relation extraction). 主要贡献如下:(1) 设计了一种基于对比学习的记忆回放和激活,通过引入对比学习,使相同关系的样本嵌入在特征空间中彼此靠近,而不同关系的样本嵌入相互分离,以提高特征空间的对齐性和均匀性;(2) 设计了一种基于焦点损失的记忆再巩固,通过引入焦点损失函数来提高模型对相似关系的区分能力,以此来减轻相似关系混乱对模型灾难性遗忘的影响;(3) 在 TACRED 数据集上进行对比和消融实验,证明了本文方法能够改进样本的特征空间,并更好地区分相似关系,从而有效缓解模型的灾难性遗忘问题.

1 方法

本部分将详细阐述本文所提出的持续关系抽取方法 CF-CRE,该方法以添加了对抗类增强机制 ACA 的 EMAR 模型 (EMAR-ACA) 为基础,引入对比学习和焦点损失来提高特征空间的对齐性和相似关系的区分度,以缓解模型的灾难性遗忘.

1.1 任务形式化与总体框架

1.1.1 任务形式化

在持续关系抽取中,模型在一系列的任务 $\{T_1, T_2, \dots, T_K\}$ 上进行训练,其中第 k 个任务 T_k 包含自己的关系集 R_k 、训练集 D_k 和测试集 Q_k . 每个样本 $(x_i, y_i) \in D_k \cup Q_k$ 是由示例 x_i 及其对应的关系标签 $y_i \in R_k$ 组成.随着新关系不断地出现和积累,模型需要在第 k 个任务和前 $k-1$ 个任务上都表现良好.所以,训练完任务 T_k 之后,需要利用前 k 个任务测试集的并集 $\tilde{Q}_k = \bigcup_{i=1}^k Q_i$

对模型进行评估,而每个测试示例的关系应该被识别为关系 $r_j \in \tilde{R}_k$, 其中 $\tilde{R}_k = \bigcup_{i=1}^k R_i$ 是所有已知关系集. 因此, 随着任务数量增多, 评估的难度会逐渐提升.

为了处理持续关系抽取中的灾难性遗忘问题, 本文同 Cui 等人^[21], 设置了情景记忆模块为历史任务存储记忆样本 $\tilde{M}_k = \bigcup_{i=1}^k M_i = \bigcup_{r_j \in \tilde{R}_k} M^{r_j}$, 其中 M_i 是第 i 个任务 T_i 的记忆样本集, M^{r_j} 是关系 $r_j \in \tilde{R}_k$ 的记忆样本集, O 是每个关系限制存储的记忆样本数量.

1.1.2 总体框架

以任务 T_k 的训练过程为例, 简要说明本文方法

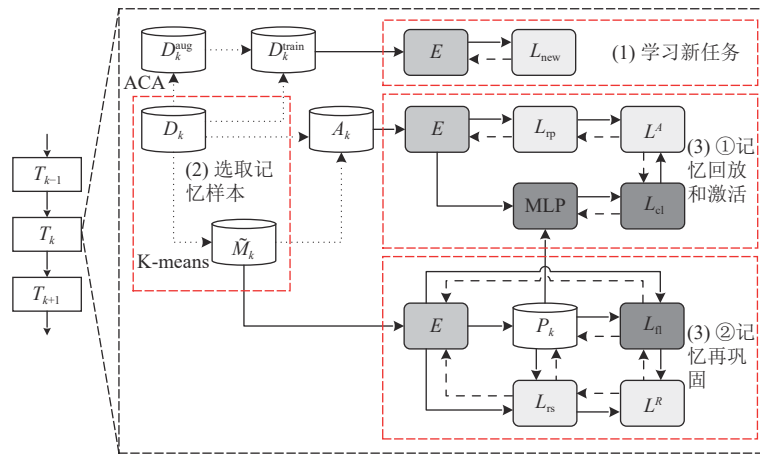


图1 本文方法 CF-CRE 的总体框架 (以任务 T_k 的训练过程为例)

1.2 示例编码器

获取示例的语义嵌入是实现持续关系抽取的关键. 本文采用预训练语言模型 BERT 来对实体对和上下文信息进行编码, 以获取示例的语义嵌入. 给定示例 x_i , 首先, 使用 $[E_{11}]$ 、 $[E_{12}]$ 、 $[E_{21}]$ 和 $[E_{22}]$ 这 4 个特殊的令牌扩充 x_i , 以增强示例中头尾实体的位置; 然后, 将新的令牌序列送入 BERT, 得到 $[E_{11}]$ 和 $[E_{21}]$ 的隐藏层表示; 最后, 将两者拼接, 输入全连接层和归一化层, 得到示例 x_i 的语义嵌入. 基于 BERT 的示例编码操作表达式如下:

$$h_i = f(x_i) \quad (1)$$

其中, h_i 是示例 x_i 的语义嵌入, 用于关系分类.

1.3 学习新任务

由于每个任务都是相互独立的, 所以当新任务 T_k 出现时, 编码器无法提取训练集 D_k 中示例的语义特征. 本文为了学习新任务 T_k , 首先, 利用对抗类增强机制 ACA 得到训练集 D_k 的增强样本集 D_k^{aug} , 然后, 在训练集 D_k 与增强样本集 D_k^{aug} 的并集 D_k^{train} 上, 微调示例编

CF-CRE 的总体框架. 如图 1 所示, 其中, 点划线表示数据流, 实线和短划线分别表示正向和反向传播, 主要分为 3 个步骤: (1) 学习新任务 T_k , 在训练集 D_k 与其增强样本集 D_k^{aug} 的并集 D_k^{train} 上学习新关系 R_k 并微调示例编码器 E ; (2) 选取记忆样本, 从训练集 D_k 中为关系集 R_k 选取部分记忆样本存入情景记忆模块; (3) 迭代地进行记忆回放、激活和再巩固, 在训练集 D_k 和记忆样本集 \tilde{M}_k 的并集 A_k 上, 利用损失函数 L^A 学习新旧关系, 再在记忆样本集 \tilde{M}_k 上, 利用关系原型集 P_k 与损失函数 L^R 巩固所有已知关系 \tilde{R}_k .

码器 E 和优化关系表示, 如图 1 的模块 (1) 所示. 所用的损失函数如下:

$$L_{new} = - \sum_{i=1}^{|D_k^{train}|} \sum_{j=1}^{|\tilde{R}_k|} \delta_{y_i=r_j} \times \log \frac{\exp(g(h_i, r_j))}{\sum_{l=1}^{|\tilde{R}_k|} \exp(g(h_i, r_l))} \quad (2)$$

其中, $|\cdot|$ 用于计算集合的基数, 即集合中所含元素的个数, $g(\cdot)$ 是计算嵌入之间的余弦相似度函数, h_i 是 D_k^{train} 中示例 x_i 的语义嵌入, r_j 是关系 $r_j \in \tilde{R}_k$ 的嵌入. 当 $y_i = r_j$ 时, $\delta_{y_i=r_j} = 1$; 当 $y_i \neq r_j$ 时, $\delta_{y_i \neq r_j} = 0$.

1.4 选取记忆样本

为了使模型能够长期有效地记忆旧关系, 需要在学习完新任务 T_k 之后, 为关系集 R_k 中的每个关系存储部分记忆样本, 如图 1 的模块 (2) 所示. 首先, 利用示例编码器 E 获取关系 $r_j \in R_k$ 所有训练样本的语义嵌入; 然后, 应用 K-means 算法对这些训练样本的语义嵌入进行聚类, 其中聚类数量 O 是每个关系限制存储的记忆样本数量; 最后, 在每个簇中, 选择最接近质心的语

义嵌入, 将其对应的训练样本作为记忆样本, 存储于情景记忆模块中。

1.5 记忆回放、激活和再巩固

在选取完记忆样本之后, 利用记忆样本集和关系原型集迭代地进行记忆回放、激活和再巩固, 以加强模型对新旧关系的区分。

1.5.1 计算关系原型

为了得到任务 T_k 的关系原型集 $P_k = \bigcup_{r_j \in \tilde{R}_k} p_{r_j}$, 首先, 结合情景记忆模块获取所有已知关系的记忆样本, 然后, 计算每个已知关系的原型嵌入。

$$p_{r_j} = \frac{1}{|M^{r_j}|} \sum_{i=1}^{|M^{r_j}|} \mathbf{h}_i \quad (3)$$

其中, M^{r_j} 是关系 $r_j \in \tilde{R}_k$ 的记忆样本集, \mathbf{h}_i 是 M^{r_j} 中示例 x_i 的语义嵌入。

1.5.2 基于对比学习的记忆回放和激活

记忆回放和激活的目的是加强模型对新旧关系的理解。在 EMAR-ACA 中, Han 等人^[20]利用交叉熵损失和 Softmax 分类器, 训练模型以识别新关系和区分旧关系。本文为了提高特征空间的对齐性和均匀性, 将有监督的对比学习引入该过程, 使具有相同关系示例的语义嵌入在特征空间中彼此映射得更近, 不同关系示例的语义嵌入映射得更远, 如图 1 的模块 (3) ①所示。具体来说, 首先利用双层 MLP 进行降维。

$$\mathbf{z}_x = \text{Norm}(\text{MLP}(\mathbf{h}_x)) \quad (4)$$

然后, 在由训练集 D_k 和记忆样本集 \tilde{M}_k 所整合的激活集 A_k 上, 利用所有已知关系的原型和 InfoNCE 损失函数来约束样本嵌入,

$$L_{\text{cl}} = -\frac{1}{|A_k|} \sum_{x_i \in A_k} \log \frac{\exp(\mathbf{z}_{x_i} \cdot \mathbf{z}_{y_i} / \tau)}{\sum_{r_j \in \tilde{R}_k} \exp(\mathbf{z}_{x_i} \cdot \mathbf{z}_{r_j} / \tau)} \quad (5)$$

其中, \mathbf{z}_{x_i} 、 \mathbf{z}_{y_i} 和 \mathbf{z}_{r_j} 分别表示激活集 A_k 中示例 x_i 的低维语义嵌入、关系 $y_i \in \tilde{R}_k$ 的低维嵌入和关系 $r_j \in \tilde{R}_k$ 的低维原型, 由式 (4) 得到。 $\tau \in \mathbb{R}^+$ 是控制分类的可调温度参数。

最后, 将记忆回放和激活阶段的目标函数定义为:

$$L^A = \lambda_1 L_{\text{cl}} + \lambda_2 L_{\text{tp}} \quad (6)$$

$$L_{\text{tp}} = -\sum_{i=1}^{|A_k|} \sum_{j=1}^{|\tilde{R}_k|} \delta_{y_i=r_j} \times \log \frac{\exp(g(\mathbf{h}_i, \mathbf{r}_j))}{\sum_{l=1}^{|\tilde{R}_k|} \exp(g(\mathbf{h}_i, \mathbf{r}_l))} \quad (7)$$

其中, $\lambda_1 \in [0, 1]$ 和 $\lambda_2 \in [0, 1]$ 是超参数, L_{tp} 是 EMAR-ACA 中记忆回放和激活阶段的损失函数, \mathbf{h}_i 是 A_k 中示

例 x_i 的语义嵌入。

1.5.3 基于焦点损失的记忆再巩固

由于仅进行记忆回放和激活, 模型最终会因新旧关系的样本数量不平衡, 而出现拟合现象。所以, 在 EMAR-ACA 中, Han 等人^[20]采用记忆再巩固来加强这一过程, 以保持对旧关系的稳定理解。

文献 [24] 指出, 当新任务出现与旧关系相似的新关系时, 模型对旧关系的预测能力会大幅度降低。但是, Han 等人^[20]提出的记忆再巩固是直接利用所有已知关系的原型嵌入来进行记忆重新整合训练, 没有考虑不同关系间的相似性对模型灾难性遗忘的影响。因此, 本文在该阶段引入了一种焦点损失函数如图 1 的模块 (3) ②所示, 使模型能够更加关注相似关系之间的区别, 减轻相似关系的混乱对灾难性遗忘的影响, 具体定义为

$$L_{\text{fl}} = -\sum_{(x_i, y_i) \in \tilde{M}_k} \log(p_s(r_i | x_i)) \quad (8)$$

其中, $p_s(r_i | x_i)$ 是记忆样本集 \tilde{M}_k 中示例 x_i 的分布, 计算方式如下:

$$p_s(r_i | x_i) = \frac{\exp(g(\mathbf{h}_i, \mathbf{p}_{r_i}))}{\sum_{l=1}^{|\mathcal{P}_{r_i}^{\text{sim}}|} \exp(g(\mathbf{h}_i, \mathbf{p}_{r_l}))} \quad (9)$$

其中, \mathbf{h}_i 是记忆样本集 \tilde{M}_k 中示例 x_i 的语义嵌入, $\mathbf{p}_{r_i} \in \mathcal{P}_{r_i}^{\text{sim}}$, \mathbf{p}_{r_i} 是 $r_i \in \tilde{R}_k$ 的关系原型, $\mathcal{P}_{r_i}^{\text{sim}} = [\mathbf{p}_{r_i}; \mathbf{p}_{r_i}^{\text{mn}}; \hat{\mathbf{p}}_{r_i}^{\text{sn}}]$ 是与示例 x_i 语义嵌入相似的关系原型集。具体来说, 首先, 得到示例 x_i 最相似的负原型 $\mathbf{p}_{r_i}^{\text{mn}} = \arg \max(g(\mathbf{h}_i, \mathbf{p}_s))$, 即最可能误算的关系原型; 然后, 设置一个阈值 α , 自动筛选一些容易混淆的负原型 $\hat{\mathbf{p}}_{r_i}^{\text{sn}} = \{\mathbf{p}_{r_i}^{\text{sn}} | g(\mathbf{h}_i, \mathbf{p}_{r_i}) - g(\mathbf{h}_i, \mathbf{p}_{r_i}^{\text{sn}}) < \alpha\}$, 其中包括与目标关系原型 p_{r_i} 最相似的负原型, 这里 $\mathbf{p}_s \neq \mathbf{p}_{r_i}$ 且 $\mathbf{p}_{r_i}^{\text{sn}} \neq \mathbf{p}_{r_i}$ 。

最后, 将记忆再巩固阶段的目标函数定义为:

$$L^R = \gamma_1 L_{\text{fl}} + \gamma_2 L_{\text{rs}} \quad (10)$$

$$L_{\text{rs}} = -\sum_{j=1}^{|\tilde{R}_k|} \sum_{i=1}^{|M^{r_j}|} \log \frac{\exp(g(\mathbf{h}_i, \mathbf{p}_{r_j}))}{\sum_{l=1}^{|\tilde{R}_k|} \exp(g(\mathbf{h}_i, \mathbf{p}_{r_l}))} \quad (11)$$

其中, $\gamma_1 \in [0, 1]$ 和 $\gamma_2 \in [0, 1]$ 是超参数, L_{rs} 是 EMAR-ACA 中记忆再巩固阶段的损失函数, \mathbf{h}_i 是关系 $r_j \in \tilde{R}_k$ 的记忆样本集 M^{r_j} 中示例 x_i 的语义嵌入, \mathbf{p}_{r_j} 是 $r_j \in \tilde{R}_k$ 的关系原型。

1.6 预测

在训练完任务 T_k 后, 从情景记忆模块中获取每个

关系 $r_j \in \tilde{R}_k$ 的记忆样本集 M^{r_j} , 计算最终的关系原型 \tilde{p}_{r_j} 进行预测.

$$\tilde{p}_{r_j} = \frac{r_j + \sum_{i=1}^{|M^{r_j}|} h_i}{1 + |M^{r_j}|} \quad (12)$$

其中, r_j 是关系 $r_j \in \tilde{R}_k$ 的嵌入, h_i 是 M^{r_j} 中示例 x_i 的语义嵌入. 最后, 预测示例 x_i 对应的关系标签 y_i , 由式 (13) 计算得到:

$$y_i = \arg \max(g(h_i, \tilde{p}_{r_j})) \quad (13)$$

其中, h_i 是测试集 $\tilde{Q}_k = \cup_{i=1}^k Q_i$ 中示例 x_i 的语义嵌入.

综上所述, 本文方法 CF-CRE 在任务 T_k 上的训练过程如算法 1 所示.

算法 1. CF-CRE 算法

输入: 任务 T_k 的训练集 D_k 、关系集 R_k 、学完前 $k-1$ 个任务得到的记忆样本集 $\tilde{M}_{k-1} = \cup_{i=1}^{k-1} M_i$ 和所有已知关系集 $\tilde{R}_{k-1} = \cup_{i=1}^{k-1} R_i$.
输出: 学完任务 T_k 之后存储的记忆样本集 $\tilde{M}_k = \cup_{i=1}^k M_i$ 和所有已知关系集 $\tilde{R}_k = \cup_{i=1}^k R_i$.

- 1) 利用 BERT 示例编码器 E 初始化 R_k 的关系嵌入;
- 2) $\tilde{R}_k \leftarrow \tilde{R}_{k-1} \cup R_k$; // 所有已知关系集 \tilde{R}_k
- 3) 利用对抗类数据增强策略 ACA 得到训练集 D_k 的增强样本集 D_k^{aug} ;
- 4) $D_k^{\text{train}} \leftarrow D_k \cup D_k^{\text{aug}}$; // 训练集 D_k 和增强样本集 D_k^{aug} 的并集 D_k^{train}
- 5) for ($i = 1$ to $epoch1$) // 学习新任务 T_k
- 6) 在 D_k^{train} 上, 利用 L_{new} 微调示例编码器 E 和优化关系嵌入;
- 7) end for
- 8) 在训练集 D_k 上, 应用 K-means 算法为关系集 R_k 中的每个关系选取记忆样本, 得到任务 T_k 的记忆样本集 M_k ;
- 9) $\tilde{M}_k \leftarrow \tilde{M}_{k-1} \cup M_k$; // 记忆样本集 \tilde{M}_k
- 10) $A_k \leftarrow \tilde{M}_k \cup D_k$; // 激活集 A_k
- 11) for ($i = 1$ to $epoch2$) // 记忆回放与激活、再巩固
- 12) 计算每个已知关系的原型嵌入, 得到任务 T_k 的关系原型集 P_k ;
- 13) for ($i = 1$ to $iter1$) // 记忆回放与激活阶段
- 14) 利用双层 MLP 获取激活集 A_k 中所有示例的低维嵌入、对应关系的低维嵌入和所有已知关系的低维原型;
- 15) 在激活集 A_k 上, 利用所有已知关系原型和 L^d 学习新旧关系;
- 16) end for
- 17) for ($i = 1$ to $iter2$) // 记忆再巩固阶段
- 18) 获取记忆样本集 \tilde{M}_k 中所有示例的相似关系原型集;
- 19) 在记忆样本集 \tilde{M}_k 上, 利用所有已知关系原型和 L^R 巩固旧关系;
- 20) end for
- 21) end for

2 实验分析

2.1 实验设置

2.1.1 数据集

为了验证本文方法 CF-CRE 的有效性, 在广泛使用的 TACRED^[26]数据集上进行实验, 将训练、验证和

测试集按 3:1:1 的比例划分. TACRED 数据集由斯坦福大学的自然语言处理研究组开发, 是一个大型的有监督关系抽取数据集. 数据来源于 TAC 知识库人口 (TAC KBP) 挑战赛的新闻电讯和网络文本, 总共包含 106264 个示例和 42 种不同的关系, 其中未定义的关系被标记为 no_relation. 实验为了与先前的研究保持一致, 删除了 no_relation, 并且限制每种关系的训练样本数量为 320 个, 测试样本数量为 40 个.

此外, 为了减轻因任务序列不同而对实验结果所造成的影响, 设置了 5 个与 Cui 等人^[21]完全相同的任务序列.

2.1.2 实验环境和参数设置

实验使用 RTX3090 (24 GB) 进行训练推理, CPU 为 Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz, 代码的运行环境为 Python 3.8、PyTorch 1.8.1 和 Cuda 11.1.

将 TACRED 数据集设置为 5 个不同的任务序列, 每个任务序列是由所有关系随机分成 10 组来模拟 10 个任务而构成, 每个任务包含 4 种关系. 示例编码采用预训练模型 BERT, 其词嵌入维度为 768, 学习率为 $1E-5$. 网络参数的优化采用 AdamW 优化器, 权重衰减为 $1E-2$. 每一步的批处理大小为 32, 词汇表的大小为 30522, 每个关系存储的记忆样本数量为 10. 学习新任务的训练轮数 $epoch1$ 为 2, 记忆回放与激活和再巩固的训练轮数 $epoch2$ 为 2, 其中记忆回放和激活、记忆再巩固的迭代次数 $iter1$ 、 $iter2$ 均为 1.

2.1.3 评价指标

继文献 [21] 之后, 将所有已学任务的平均准确率作为评价指标. 对于每个任务序列, 模型在当前任务 T_k 上的准确率是模型在测试集 $\tilde{Q}_k = \cup_{i=1}^k Q_i$ 上对所有已知关系的分类准确率.

2.2 实验结果与分析

2.2.1 总体性能比较

为了评估本文方法 CF-CRE 的性能, 在 TACRED 数据集上与 5 种新型的 CRE 模型进行比较. 其中, EMAR 将情景记忆激活和再巩固引入持续学习, 以保持对旧关系的稳定理解 (需要指出的是, 实验采用的是 BERT 示例编码器); RP-CRE 引入原型网络来细化样本嵌入, 以更好地恢复特征空间的中断; CRL 采用对比学习和知识蒸馏来保持关系表示的稳定; EMAR-ACA 在 EMAR 上添加对抗类增强机制 ACA, 以提高关系表示的鲁棒性; CEAR 设计记忆不敏感关系原型和记忆增强策略,

以克服模型的过拟合问题,并引入综合训练和焦点知识蒸馏来提高模型对相似关系的区分能力。

表 1 记录了上述 CRE 模型在所有已学任务上的平均准确率,并且用加粗字体表示当前任务的最高平均准确率,针对这些结果,得出以下结论:(1)从整体来看,随着任务数量增加,所有模型的性能都有所下降,说明灾难性地遗忘旧关系是不可避免的,解决灾难性遗忘问题是实现持续关系抽取的重要挑战。(2)与基线模型 EMAR-ACA 相比,本文方法 CF-CRE 几乎在所有任务上都取得更好的结果,并且在 T10 上提升了 1.26%,

说明引入对比学习和焦点损失是有效的,并进一步表明提高特征空间的对齐性和增强模型对相似关系的关注,有利于缓解模型的灾难性遗忘问题;(3)与其他各模型相比,本文方法 CF-CRE 在所有任务上都取得较好的结果,尤其是在后 4 个任务上效果显著,这表明本文方法在长期记忆方面具有优势.其原因是对抗类增强机制 ACA 帮助模型学习到更加精确和强大的关系表示,并且引入焦点损失提升了模型对相似关系的区分能力.其中,CEAR 在任务 T3-T6 上的平均准确率较高,但随着任务数量不断增加,CF-CRE 最终超过了 CEAR.

表 1 各模型在 TACRED 数据集上的对比实验结果 (%)

模型	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
EMAR ^[20]	97.62	92.24	88.92	85.16	84.26	81.06	79.13	78.57	78.09	77.35
RP-CRE ^[21]	97.80	91.61	88.71	84.64	83.03	80.69	78.46	77.57	77.26	76.63
CRL ^[22]	98.07	94.10	90.58	85.33	84.23	82.43	80.34	78.51	78.46	78.13
CEAR ^[25]	97.68	93.59	91.21	87.48	85.56	83.24	80.73	79.13	78.52	77.97
EMAR-ACA ^[24]	97.68	92.99	90.00	86.19	83.70	81.67	79.43	78.44	78.63	77.82
CF-CRE (本文)	97.88	93.79	90.66	86.00	85.37	83.06	81.11	80.28	79.93	79.08 (↑1.26%)

2.2.2 消融实验

为了验证对比学习和焦点损失的有效性和合理性,进行消融实验.表 2 记录了在 TACRED 数据集上,各模型在后 5 个任务上的平均准确率.其中,“+CL”是在 EMAR-ACA 的记忆回放和激活阶段引入对比学习,并且超参数 λ_1 和 λ_2 设置为 0.1 和 0.9;“+FL”是在 EMAR-ACA 的记忆再巩固阶段添加焦点损失函数,并且超参数 γ_1 和 γ_2 设置为 0.5 和 0.5.

表 2 各模型在 TACRED 数据集上的消融实验结果 (%)

模型	T6	T7	T8	T9	T10
Baseline (EMAR-ACA)	81.67	79.43	78.44	78.63	77.82
+CL	82.80	80.56	79.82	79.65	78.30
+FL	82.27	79.59	78.69	79.40	78.27
CF-CRE (+CL+FL)	83.06	81.11	80.28	79.93	79.08

针对表 2 的实验结果,得出以下结论:与基础模型 EMAR-ACA 相比,(1)引入对比学习(+CL)使得模型的性能有所提升(0.48%),这表明优化样本嵌入,提高特征空间的对齐性和均匀性,能有效缓解模型的灾难性遗忘问题,并提升其在长任务序列中的分类能力。(2)添加焦点损失(+FL)同样使得模型的性能有所提升(0.45%),这表明提高模型对相似关系的关注,能够减轻因相似度较高而导致的关系分类混乱,从而改善模型对旧关系的遗忘。

2.2.3 参数选取讨论

本节以 EMAR-ACA 为基础模型,通过调整超参数 λ_1 、 λ_2 、 γ_1 和 γ_2 来进一步分析对比学习与焦点损失对模型性能的影响,结果如图 2 所示。

观察图 2 的实验结果,可以得出如下结论:(1)当“ $\lambda_1=0.1$; $\lambda_2=0.9$ ”时,模型在最后一个任务上表现最佳.但是,随着对比损失的比重逐渐增加,模型性能开始下降.其原因是,虽然对比学习拉近了相同关系样本嵌入的距离,提高了特征空间的均匀性,但同时也导致模型在预测过程中对相似关系原型的敏感度增加。(2)当“ $\gamma_1=0.5$; $\gamma_2=0.5$ ”时,模型在各任务上的表现较好.但是,当继续增加焦点损失的比重时,模型的性能会出现下降.其原因是,虽然引入焦点损失可以减轻相似关系混乱对灾难性遗忘的影响,但由于已知关系中只有部分关系相似度较高,焦点损失比重过高会降低模型对非相似关系的关注度。

3 结论

针对持续关系抽取任务中面临的灾难性遗忘问题,本文以添加了对抗类增强机制 ACA 的 EMAR 模型(EMAR-ACA)为基础,提出了一种基于对比学习和焦点损失的持续关系抽取方法(CF-CRE).利用对比学习提高特征空间的对齐性和均匀性,使相同关系的样本

嵌入彼此映射得更近,不同关系的样本嵌入映射得更远;引入焦点损失函数,增强模型对相似关系的区分能力,减轻相似关系的混乱对灾难性遗忘的影响,提高关

系分类准确率.在 TACRED 数据集上,模型的性能有所提升,验证了本文方法 CF-CRE 的有效性.在未来,将深入研究少样本情况下的持续关系抽取.

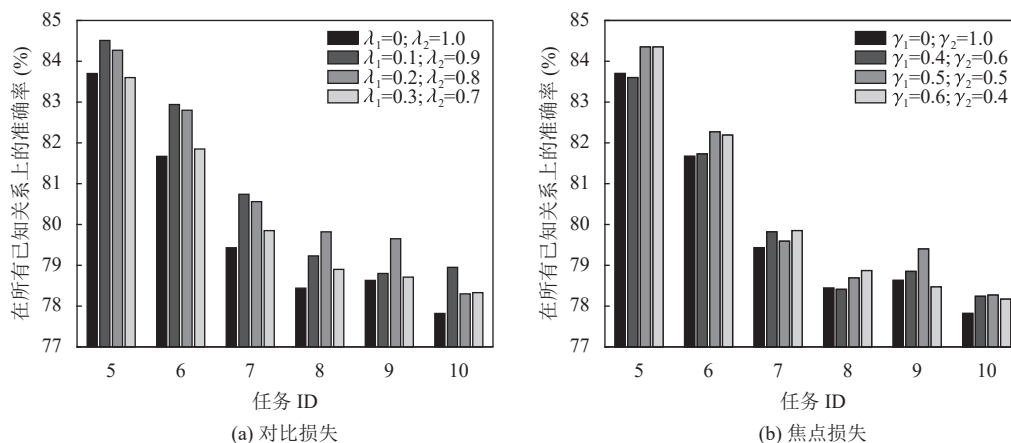


图2 不同超参数下两种损失对模型性能的影响

参考文献

- Liu K. A survey on neural relation extraction. *Science China Technological Sciences*, 2020, 63(10): 1971–1989. [doi: 10.1007/s11431-020-1673-6]
- 谢腾, 杨俊安, 刘辉. 融合多特征 BERT 模型的中文实体关系抽取. *计算机系统应用*, 2021, 30(5): 253–261. [doi: 10.15888/j.cnki.csa.007899]
- Wu F, Weld DS. Open information extraction using wikipedia. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Uppsala: Association for Computational Linguistics, 2010. 118–127.
- 赵宇博, 张丽萍, 闫盛, 等. 个性化学习中学科知识图谱构建与应用综述. *计算机工程与应用*, 2023, 59(10): 1–21.
- Mishra A, Jain SK. A survey on question answering systems with classification. *Journal of King Saud University-Computer and Information Sciences*, 2016, 28(3): 345–361. [doi: 10.1016/j.jksuci.2014.10.007]
- 姚元杰, 龚毅光, 刘佳, 等. 基于深度学习的智能问答系统综述. *计算机系统应用*, 2023, 32(4): 1–15. [doi: 10.15888/j.cnki.csa.009038]
- 吕鹏帆. 深度持续实体关系抽取算法的研究 [硕士学位论文]. 西安: 西安电子科技大学, 2022.
- 张志远. 基于小样本学习和终身学习的因果关系抽取 [硕士学位论文]. 长春: 吉林大学, 2022.
- Thrun S, Mitchell TM. Lifelong robot learning. *Robotics and Autonomous Systems*, 1995, 15(1-2): 25–46. [doi: 10.1016/0921-8890(95)00004-Y]
- De Lange M, Aljundi R, Masana M, *et al.* A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(7): 3366–3385.
- Parisi GI, Kemker R, Part JL, *et al.* Continual lifelong learning with neural networks: A review. *Neural Networks*, 2019, 113: 54–71. [doi: 10.1016/j.neunet.2019.01.012]
- 王松, 买日旦·吾守尔, 古兰拜尔·吐尔洪, 等. 基于知识架构的持续学习情感分类方法. *计算机工程*, 2023, 49(2): 112–118.
- Goodfellow IJ, Mirza M, Xiao D, *et al.* An empirical investigation of catastrophic forgetting in gradient-based neural networks. arXiv:1312.6211v3, 2015.
- Kirkpatrick J, Pascanu R, Rabinowitz N, *et al.* Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(13): 3521–3526.
- Lopez-Paz D, Ranzato MA. Gradient episodic memory for continual learning. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates, Inc., 2017. 6467–6476.
- Wang H, Xiong WH, Yu M, *et al.* Sentence embedding alignment for lifelong relation extraction. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technology*. Minneapolis: Association for Computational Linguistics, 2019. 796–806.
- Dhar P, Singh RV, Peng KC, *et al.* Learning without memorizing. *Proceedings of the 2019 IEEE/CVF Conference*

- on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 5138–5146.
- 18 Mallya A, Lazechnik S. PackNet: Adding multiple tasks to a single network by iterative pruning. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7765–7773.
- 19 Qin Q, Hu W P, Peng H, *et al.* BNS: Building network structures dynamically for continual learning. Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021. 20608–20620.
- 20 Han X, Dai Y, Gao TY, *et al.* Continual relation learning via episodic memory activation and reconsolidation. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Brussels: Association for Computational Linguistics, 2020. 6429–6440.
- 21 Cui L, Yang DQ, Yu JX, *et al.* Refining sample embeddings with relation prototypes to enhance continual relation 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. 232–243.
- 22 Zhao K, Xu H, Yang JG, *et al.* Consistent representation learning for continual relation extraction. Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022. Dublin: Association for Computational Linguistics, 2022. 3402–3411.
- 23 Hu CW, Yang DQ, Jin HL, *et al.* Improving continual relation extraction through prototypical contrastive learning. Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju: International Committee on Computational Linguistics, 2022. 1885–1895.
- 24 Wang PY, Song YF, Liu TY, *et al.* Learning robust representations for continual relation extraction via adversarial class augmentation. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Abu Dhabi: Association for Computational Linguistics, 2022. 6264–6278.
- 25 Zhao WZ, Cui YN, Hu W. Improving continual relation extraction by distinguishing analogous semantics. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto: Association for Computational Linguistics, 2023. 1162–1175.
- 26 Zhang YH, Zhong V, Chen DQ, *et al.* Position-aware attention and supervised data improve slot filling. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 35–45.

(校对责编: 孙君艳)