

关系抽取技术预研

作者：纽约的自行车

日期：2021 年

摘 要

本文介绍了关系抽取领域的三个主流研究方向：有监督学习、少样本学习和远程监督学习。

1 背景介绍

关系抽取是信息抽取领域的重要分支，是构建知识图谱的重要技术环节。本次调研采用文献调研法。收集顶会上发表的论文，然后按照研究方向归纳论文。

2 研究现状

关系抽取（Relation extraction，RE）就是从非结构化文本中提取实体之间的关系。依据实体是否在文本中被标记，关系抽取方法可分为联合抽取和流水线式抽取。联合抽取是指从文本中完成实体识别和关系分类任务，流水线式抽取是指先使用实体识别模型识别文本中的实体对，然后判断实体对的关系。一个完整的流水线式关系抽取系统包括命名实体识别、实体链接和关系分类，本次调研限定关系分类方法。为了简述方便，下文中仍使用关系抽取表述指代关系分类。关系抽取模型大致可分为三类 [1]：基于 pattern 的方法 [2-3]、统计机器学习 [4] 和神经网络。其中神经网络方法效果更好，本文介绍的关系抽取方法以神经网络为主。依据抽取语料形式，关系抽取模型又可分为句子级关系抽取和段落级关系抽取。句子级关系抽取指两个实体在一个句子中，段落级关系抽取指两个实体不在同一个句子中。句子级关系抽取在实际业务中更常见，本次调研仅限于句子级关系抽取。按照关系类型是否预定义，关系抽取模型又可分为基于 schema 的关系抽取和开放关系抽取。前者是指实体对的关系只能从预先定义好的类别中选取，后者是指实体对关系没有限定。在构建领域知识图谱中，schema 是固定的，因此实体对的关系类型是预定义的，本次调研针对基于 schema 的关系抽取。

神经网络可以从大规模数据中自动学习特征，此类方法的研究多集中在设计模型结构捕获文本语义。当前最好的关系抽取模型都是有监督模型，需要大量标注数据并且只能抽取预定义的关系。这种方法不能面对复杂的现实场景，例如少样本。目前已有不少工作探索在现实场景下的关系抽取任务，本文介绍三个研究方向：有监督神经网络方法、远程监督和少样本学习。

2.1 有监督神经网络方法

有监督神经网络方法是指采用深度学习方法在大规模有监督数据集上训练模型，此类方法是目前效果最好且研究最深入的。本文归纳总结了部分有代表性的模型结构，如图 1 所示，模型

通常由输入层、编码层、注意力层和分类层组成。

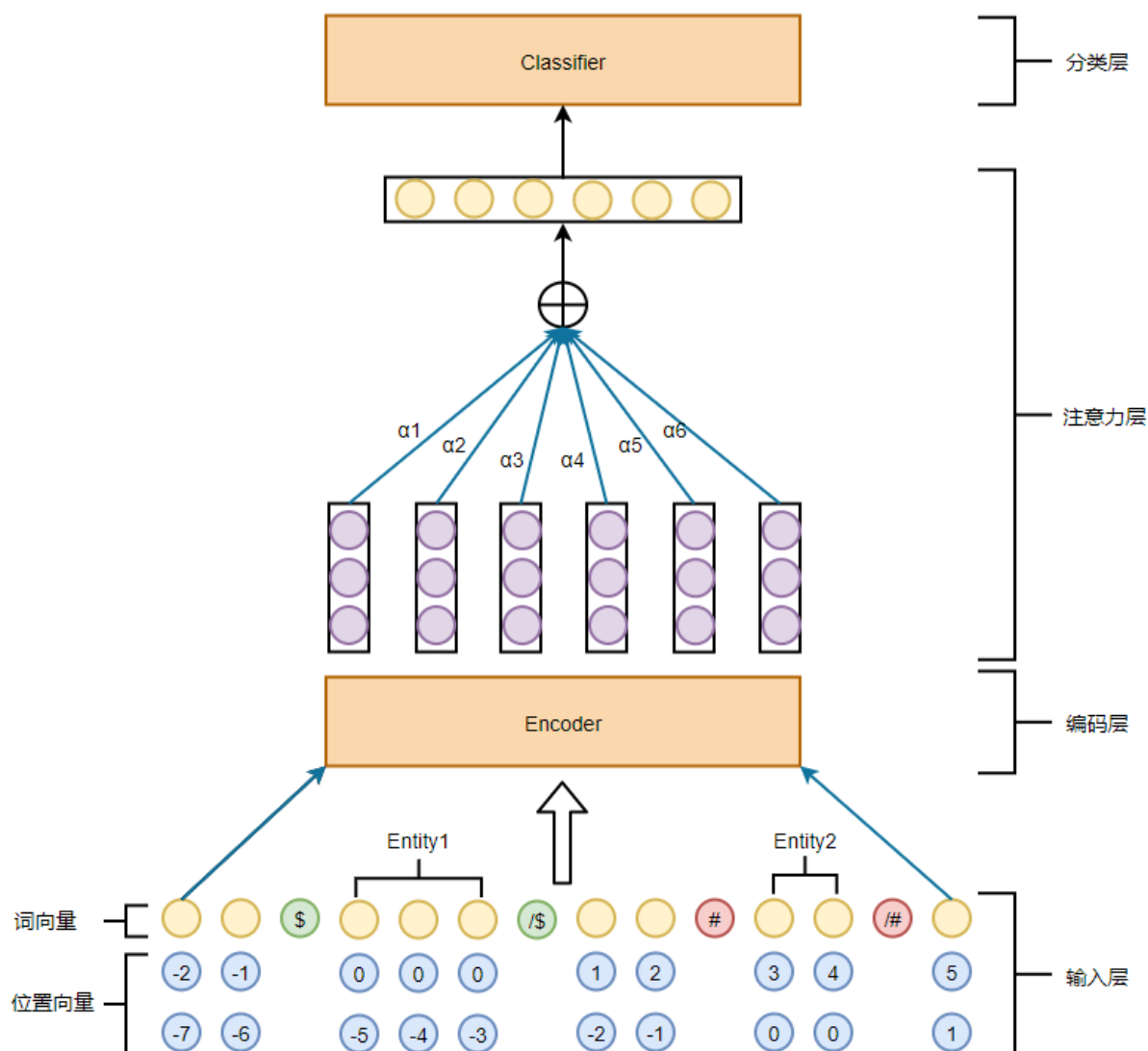


图 1: 有监督神经网络模型结构

输入层将输入字符转化为向量。输入句子包含两个标记的实体，除去实体外的字符称为上下文。Peng Hao [5] 分析了实体信息和上下文信息对关系抽取模型的重要度。实体信息包括实体类型、实体属性信息等，上下文信息是指上下文字符。作者通过实验发现，上下文信息和实体信息对关系抽取任务都提供了重要信息，只选用其中一种信息会导致模型性能下降严重，作者还观察到实体信息中最重要的是实体类型信息。实体位置信息对关系抽取任务同样很重要，在 CNN 模型中通常使用相对位置向量 [6]，在 RNN 模型中通常使用实体标记表示实体位置 [7]。除了实体信息、上下文信息、实体位置信息外，其他比较重要的特征还有句子最短依赖树 [8]，依存句法特征 [9]。编码层从输入向量中提取特征。早期使用 CNN、RNN 模型，Peng Hao [5] 的对比结果表明这些模型的编码能力不足，目前主流使用 BERT 等预训练模型作为编码器。Wu [10] 使用 BERT+MLP 结构取得了当时最好结果。句子中不同位置的字符对实体对关系分类的重要度不同，使用注意力方法对不同字符赋予不同权重。Zhang Yuhao [11] 在注意力计算中加入位置信息，Wang Linlin [12] 使用 CNN 模型抽取句子特征，并在此基础上增加了实体注意力和关系类别

注意力。注意力层可有可无，在以 CNN、RNN 为编码器的模型中使用居多。分类层对编码层/注意力层输出的特征向量分类，主流使用 MLP 结构分类。损失函数一般有两种，交叉熵损失和排序损失。交叉熵损失计算过程简单，排序损失需要采样负样本，Santos [13] 使用排序损失取得了相比交叉熵损失很好的结果。有监督神经网络方法在大规模语料下取得了很好的效果，本文选取 Wu [10] 作为此类方法的 baseline 方法。Wu [10] 使用位置标记表示实体位置，使用 BERT 作为编码器，按照分段池化的思想，根据实体位置将 BERT 输出分为三部分，然后池化拼接，使用 MLP+softmax 分类，取得了当时最好结果。模型细节将在 3.1 节介绍。

2.2 少样本学习

少样本是实际业务中经常遇到的问题，少样本是指只有很少的标注数据提供给模型训练。神经网络方法在大量高质量标注数据上具有很好的效果，但在少样本场景下效果很差。少样本学习是专门针对少样本场景提出的方法，少样本关系抽取有很多解决方案，例如 meta learning [14-15]、transfer learning [16-17]、prototypical network [18] 等，其中 prototypical network 效果最好且模型简单，本文主要介绍 prototypical network 在少样本关系抽取中的应用。

Prototypical network 对每一个类别的所有样例特征向量求均值，将这个均值作为该类别的 prototype。对测试样例进行预测的时候，计算测试样例特征向量和这些 prototype 的距离，距离最近的 prototype 对应的类别作为该样例的类别。如图 2 所示，C1, C2, C3 表示 prototype，x 表示测试样本，其他颜色小圆圈表示训练样本，不同颜色表示不同关系类别。

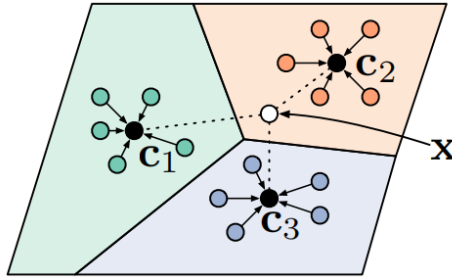


图 2: few shot.

少样本方法训练过程和常规的模型训练稍有不同，少样本训练过程是将训练集分成 train-support 和 train-query 两部分，模型将 train-support 中每个样本编码成一个向量，然后对每类关系的所有样本向量求均值，将这个均值向量作为该类别的 prototype，类似于聚类中心。然后从 train-query 中选取 query 样本，模型将 query 样本编码成向量，计算该向量与每个 prototype 的距离，选取距离最近的 prototype 对应的关系类别作为该 query 样本的关系类别，损失函数一般选用欧式距离。

原生 prototypical network 在计算 prototype 时使用全部数据，无法解决 train-support 中样本标注错误问题。Tianyu Gao [19] 提出使用 Instance-level attention 从训练数据中筛选与 query 相关的样本计算 prototype，从而减少错误样本的影响。少样本带来的另一个问题是有用特征稀疏。每个样本能提供的有用特征本来就少，再加上每个类别的样本数量不多，导致 prototype 中对判别关系有用的特征少，称之为特征稀疏。Tianyu Gao [19] 使用 feature-level attention 从 prototype 中选择有效特征，从而减少特征稀疏的影响。ZhiXiu Ye [20] 使用和 Tianyu Gao [19] 类似的方法，

在 prototype 形成过程中加入 query 的注意力，取得了更好的效果。

Soares [21] 想训练一个通用关系抽取模型，类似于 BERT 这样的通用语言模型。作者将句子中的实体对使用 “[blank]” 代替，将实体 mask 的目的是让模型更加关注于上下文，通过上下文来判断句子对是否包含相同关系类型，而不是利用实体推断。将 mask 后的句子称为 relation statement。使用 BERT 在 relation statement 数据集上预训练，然后在有实体的数据集上训练，效果好于直接在有实体的数据集上训练，尤其是在少样本数据集上效果明显。基于这个方法，作者训练一个 relation statement 编码器用于学习 relation statement 向量表征。如果两个 relation statement 表达相同的关系，那么等于 1，否则等于 0。训练数据来源于远程标注，使用 Wikipedia 作为文本库，wikidata 为知识库。两个句子含有相同实体对则标记为正例，否则为负例。作者共远程标注了 6 亿条数据。在实际操作中，并不是每个实体都被 mask，而是根据一定概率，这类似于 BERT 对 token 的 mask。

Soares [21] 中的模型使用预训练方法在少样本数据集上取得很好的效果，但是预训练是一件非常困难的事情，需要大量数据、显卡和训练技巧。ZhiXiu Ye [20] 在 FewRel1.0 数据集上取得了仅次于 Soares [21] 的分数，且模型简单无需额外数据，本文选取该方法作为 baseline 模型。

2.3 远程监督

训练强大的关系抽取模型需要更多的高质量数据，但是构建这样的数据集需要大量人工标注，耗时费力。Mike Mintz [22] 首次使用远程监督方法生成标注数据，远程监督的假设是：如果两个实体有关系，那么任何包含这两个实体的句子都可以表达这种关系。例如， $Rae1, e2$ 表示实体 $e1$ 和 $e2$ 具有关系 Ra ，若存在一个句子同时包含 $e1$ 和 $e2$ ，则认为该句子表达了关系 Ra ，将该句子标注为关系 Ra 的一个正样本。使用这种方法只需一个知识库和一个文本库便可自动得到标注数据。远程监督似乎是一种解决监督数据不足的完美解决方案，但实际上远程监督数据存在如下问题：1) 并非所有同时包含 $e1$ 和 $e2$ 的句子都可以表达关系 Ra ，因此数据集中存在大量标注错误问题；2) 无法解决一对实体含有多关系的情况；3) False negative 问题，标记为负样本的实例实际存在关系，但这种知识在知识图谱中不存在，导致标记错误。第二个问题是远程监督无法解决的问题，因为在知识图谱中两个节点间只能存在一条边，所以无法建模一对实体存在多个关系的情况。第三个问题可以通过更好的生成负样本方式缓解，例如选择包含两个明显不存在关系的实体对的句子作为负样本。第一个问题是最严重的，目前大量的研究多集中于此，本文接下来介绍目前的主流解决方案。

缓解远程监督数据噪声有三种方法：1) 多实例学习，从多个实例中选择最有效的实例作为训练样本；2) 使用外部信息选择有效实例；3) 使用复杂的模型和训练方法，例如 soft label、增强学习、对抗学习。

多实例学习 Riedel [23] 放松了远程监督假设，将任务建模为多实例学习问题。多实例学习的假设是：如果实体对存在某种关系，那么包中至少有一个句子反映了该关系。这里的包是指由存在该实体对的所有句子组成，即一个实体对对应一个包。如何从包中选择最能反映关系类别的句子作为训练样本是多实例学习的核心问题。在 PCNN [24] 模型中，作者使用 CNN+softmax 计算包中每个样本的分类得分，选择得分最高的样本作为该包中最能反映关系类别的样本。PCNN 模型是关系抽取中常用的编码器，用于抽取句子特征，模型结构如下图所示。依据两个实体的

位置，将句子分割为三个片段，左边部分包含了左边实体信息，中间部分包含了两个实体信息，右边部分包含了右边实体信息。对全局进行池化操作无法捕获这种结构信息，因此作者分别对三部分进行池化操作。

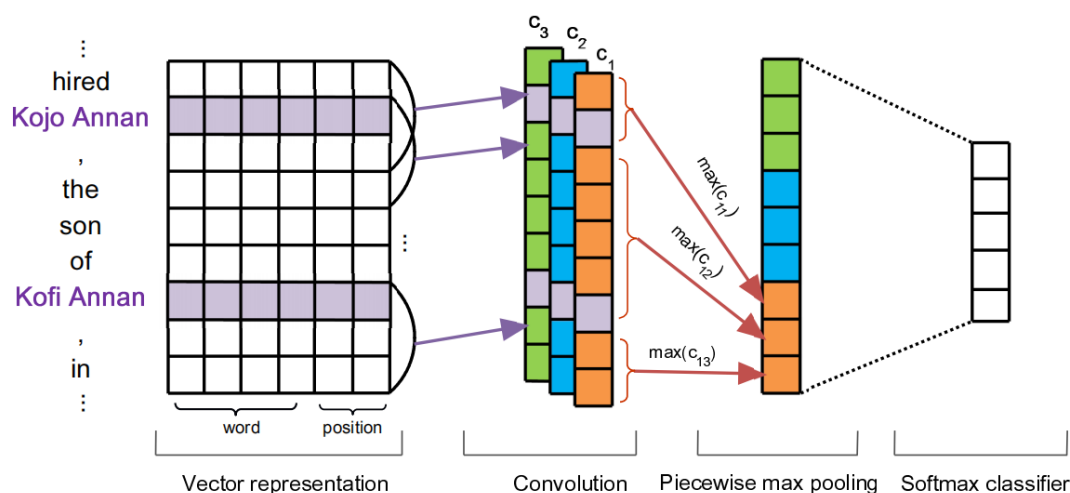


图 3: PCNN.

Yankai Lin [25] 认为 PCNN 模型只选用包中一个句子作为训练样本会导致训练不够充分，因为包中可能存在多个句子都能有效表达关系，只用一个样本会遗漏其他样本所表达的特征。因此，作者使用选择性注意力对包中所有句子学习权重，对能够表达该关系的句子赋予高权重，无法表达该关系的句子赋予低权重。包中所有句子特征向量加权得到包的向量表征，使用包向量表征完成最终的关系分类。使用外部信息如果一个实例表达了某种关系，那么该实例的句向量应该和这种关系的特征向量相似度很高。基于这种想法，许多研究人员开始使用关系特征向量参与选择有效实例。这种方法首要解决的问题是如何获取关系的向量表征。Han xu [26] 借鉴 TransE 的思想，直接使用知识图谱中训练得到的关系向量表征。随后在此基础上改进，使用多任务学习方法，联合学习关系抽取和知识图谱知识表示 [27]。Guoliang Ji [28] 利用 Wikipedia 中实体的描述文本训练得到实体向量，然后两个实体向量相减得到关系特征向量。模型结构如图 4 所示， b_i 表示包中第 i 个实例向量，由 PCNN 模型编码得到。使用关系特征向量 v 计算实例向量 b_i 权重，实例向量加权得到包的向量表征。

复杂网络有部分学者引入增强学习 [29-30] 和对抗训练 [31-32] 完成关系抽取，由于本人对这两个领域完全陌生，在此不展开调研。减少远程监督数据中的噪声是正确的研究方向，但是目前的远程监督方法效果并不好，精确率和召回率都很低，在主流 benchmark 上 F1 值在 0.6 左右。在第十三届全国知识图谱与语义计算大会 CCKS2019¹任务 3 中，最好成绩的 F1 值仅 0.54。远程监督方法的低效目前还难以应用在实际业务中。因此本文没有选取该领域的论文复现。

¹https://www.cast.org.cn/art/2019/1/22/art_44_86133.html

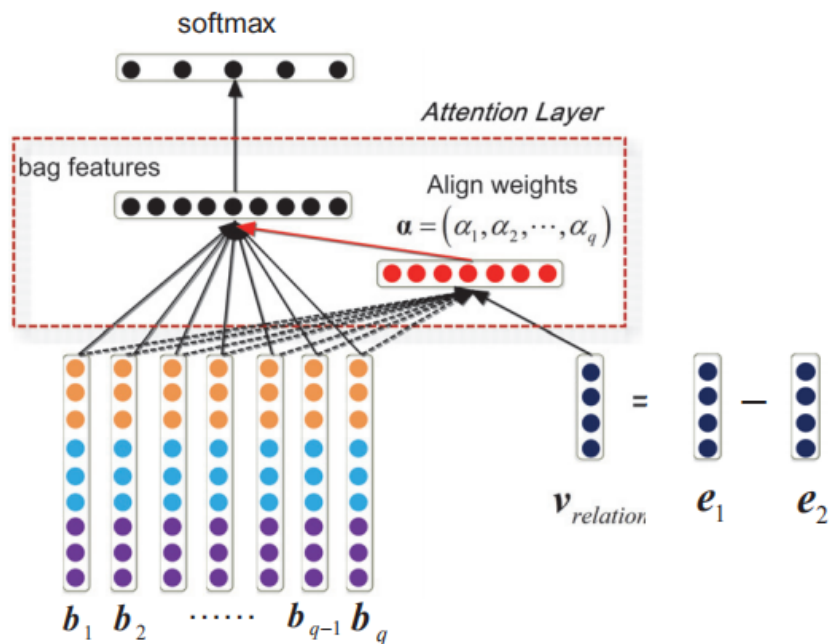


图 4: Sentence-level Attention.

3 baseline 论文介绍及复现

3.1 Enriching Pre-trained Language Model with Entity Information for Relation Classification

3.1.1 论文介绍

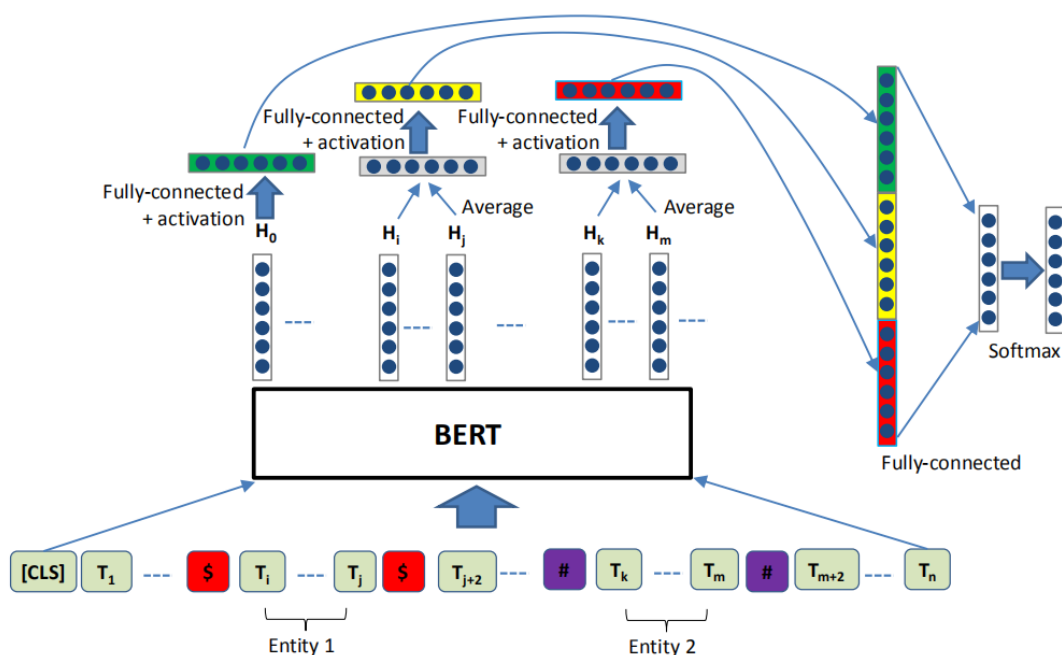


图 5: The model architecture.

本文使用 BERT 抽取句子和实体特征，然后使用 MLP 分类实体关系。结构非常简单，是基

本的使用语言模型做关系分类的方法。

模型结构如图 5 所示。输入是单个句子，在开头加上 “[CLS]” 标志，在第一个实体的开始和结束位置添加 “\$” 标记，在第二个实体的开始和结束位置添加 “#” 标记。从 BERT 的输出中选取三个部分，第一个是 [CLS] 位置的输出，第二个是实体 1 对应位置的输出，第三个是实体 2 对应位置的输出。在图中词 T_i 对应的输出是 H_i ，同理词 T_j 对应的输出是 H_j ， T_i 到 T_j 组成实体 1。实体 1 和实体 2 的输出分别取平均得到一维向量，这样加上 [CLS] 位置的输出就有三个一维向量。这三个向量分别做激活和全连接操作，如下所示：

$$H'_0 = W_0 (\tanh(H_0)) + b_0 \quad (1)$$

$$H'_1 = W_1 \left[\tanh\left(\frac{1}{j-i+1}\right) \sum_{t=i}^j H_t \right] + b_1 \quad (2)$$

$$H'_2 = W_1 \left[\tanh\left(\frac{1}{m-k+1}\right) \sum_{t=k}^m H_t \right] + b_1 \quad (3)$$

H_0 表示 [CLS] 位置输出的一维向量，第二个和第三个等式分别表示对实体 1 和实体 2 做激活和全连接操作，注意后面两个全连接操作的 W 相同，说明使用同一个全连接。最后拼接这三个输出，完成一个全连接后输入 softmax 中分类。

$$h' = W_3 [\text{concat}(H'_0, H'_1, H'_2)] + b_3 \quad (4)$$

$$p = \text{softmax}(h') \quad (5)$$

模型在 SemEval-2010 task 8 数据集上测试，取得 SOTA。为了分析哪些操作对模型有用，消融实验结果如表 1。

表 1: 消融实验

method	F1
R-BERT-NO-SEP-NO-ENT	81.09
R-BERT-NO-SEP	87.98
R-BERT-NO-ENT	87.99
R-BERT	89.25

R-BERT 是本文的模型，NO-ENT 表示只使用 [CLS] 位置的向量，NO-SEP 表示输入中去除实体标记 “\$” 和 “#”。实验结果说明让模型知道实体及其位置非常重要。[CLS] 位置向量建模整个句子特征，使用这个向量做分类是常规操作。在句子中插入实体标记让模型可以定位实体的位置，拼接实体位置的向量做分类可以让模型更精确的捕获实体语义。

3.1.2 论文复现

采用开源数据集 DuIE2.0²，DuIE2.0 是规模最大的开源的中文关系抽取数据集，包含 43 万三元组数据、21 万中文句子及 43 个预定义的简单值关系类型。训练集共 290280 个三元组。

在测试集上测试结果表 3 所示，共 34958 条三元组。训练代码和数据已上传服务器³。

²<https://aistudio.baidu.com/aistudio/competition/detail/46>

³[/home/team3user/knowledge_graph/relation_extract/RelationExtract-BERT](https://github.com/Team3user/knowledge_graph/relation_extract/RelationExtract-BERT)

表 2: DuE12.0 训练集关系统计情况

关系类型	样本数量	关系类型	样本数量	关系类型	样本数量	关系类型	样本数量
主演	42186	作者	29796	歌手	22034	导演	20182
父亲	15030	成立日期	14774	妻子	12469	丈夫	12456
国籍	10891	母亲	9486	作词	9194	作曲	8775
毕业院校	8314	所属专辑	7746	董事长	7415	朝代	6683
嘉宾	6136	出品公司	6080	编剧	6055	简称	4641
主持人	4280	主题曲	3394	校长	3007	总部地点	2806
主角	2341	创始人	2233	制片人	1735	号	1616
祖籍	1550	注册资本	977	代言人	923	气候	908
所在城市	748	行政区域面积	745	占地面积	666	人口数量	536
改编自	441	首都	435	海拔	382	官方语言	145
邮政编码	26	专业代码	24	修业年限	19		

表 3: 测试结果

关系类型	P	R	F1	样本数量	关系类型	P	R	F1	样本数量
主演	0.96	0.93	0.94	5323	校长	0.99	1	1	341
作者	0.97	0.97	0.97	3679	总部地点	1	0.99	1	351
歌手	0.88	0.92	0.9	2550	主角	0.93	0.92	0.92	223
导演	0.94	0.95	0.95	2475	创始人	0.93	0.91	0.92	232
父亲	0.96	0.98	0.97	1809	制片人	0.89	0.88	0.88	203
成立日期	1	1	1	1748	号	1	1	1	204
妻子	0.96	0.96	0.96	1484	祖籍	1	0.98	0.99	190
丈夫	0.96	0.96	0.96	1484	注册资本	1	1	1	126
国籍	1	0.99	0.99	1423	代言人	0.95	0.97	0.96	105
母亲	0.97	0.93	0.95	1084	气候	1	1	1	103
作词	0.79	0.8	0.79	1104	所在城市	0.98	0.98	0.98	81
作曲	0.74	0.64	0.69	1039	行政区域面积	0.92	0.95	0.93	83
毕业院校	1	1	1	988	占地面积	0.95	0.92	0.94	87
所属专辑	0.99	0.99	0.99	915	人口数量	1	1	1	53
董事长	0.97	0.98	0.97	872	改编自	0.96	1	0.98	68
朝代	0.98	1	0.99	749	首都	1	1	1	51
嘉宾	0.9	0.92	0.91	656	海拔	1	1	1	43
出品公司	1	1	1	797	官方语言	1	1	1	15
编剧	0.8	0.78	0.79	792	邮政编码	1	1	1	4
简称	1	1	1	520	专业代码	1	0.5	0.67	2
主持人	0.91	0.91	0.91	503	修业年限	1	0.5	0.67	2
主题曲	0.98	0.98	0.98	397					

在大规模数据集上，模型效果总体很好。在部分关系类型上，即使训练集中样本数量很多，但测试效果很差，例如关系作词、作曲和编剧。这三种关系常常同时出现，容易导致模型混淆。邮政编码、专业代码和修业年限由于测试数量太少，测试结果不可信度。首都和海拔的上下文中存在明显的触发词暗示了实体类型，例如句子“<e1> 五阆山 </e1> 在湖头镇西部，海拔 <e2>1256.7 米 </e2>”，实体 e2 前有触发词“海拔”。这种类型的关系提取相对简单，在训练样本数量少的情况下效果也很好。

3.2 Multi-Level Matching and Aggregation Network for Few-Shot Relation Classification

3.2.1 论文介绍

本文采用 Prototypical network 方法，在生成 prototype 过程中加入 query 的注意力，以此减少训练集中数据标注错误带来的影响。模型结构如图 6 所示。模型分为四个阶段：Encoder、local matching and aggregation、instance matching and aggregation 和 class matching。

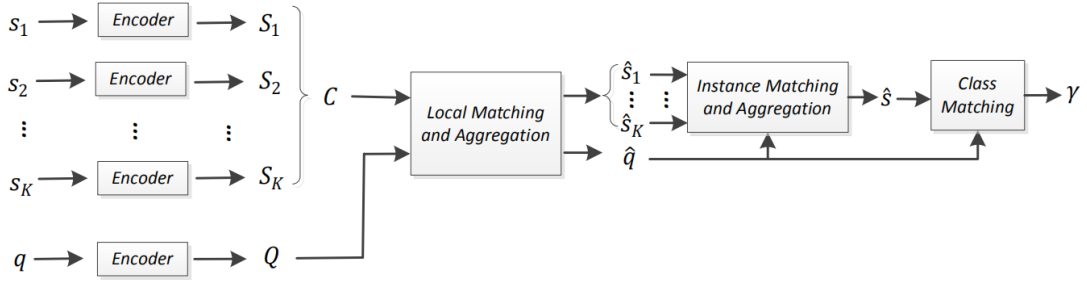


图 6: MLMAN architecture.

将数据分为 support 和 query 两部分。Encoder 是句子编码器，使用 CNN 模型。作者使用 d_c 个卷积核，最后得到一个矩阵而不是向量。query 句子经过 Encoder 后得到矩阵 $Q \in R^{T_q \times d_c}$ ， T_q 表示句子长度。support 中每个类别的句子经过 Encoder 后得到矩阵 $S_i \in R^{T_i \times d_c}$ ， T_i 表示句子长度， i 表示类别。local matching and aggregation 阶段收集 query 和 support 实例的局部匹配信息。将 support 中每个类别中的句子拼接成一个句子，得到 $C = \text{concat}(S_{i_k} = 1^K)$ ，其中 $C \in R^{T_s \times d_c}$ ， $T_s = \sum_{k=1}^K T_k$ 。然后计算局部注意力，得到新的矩阵 \tilde{Q} 和 \tilde{C} ，计算过程如下式：

$$\alpha_{mn} = q_m^T c_n \quad (6)$$

$$\tilde{q}_m = \sum_{n=1}^{T_s} \frac{\exp(\alpha_{mn})}{\sum_{n'=1}^{T_s} \exp(\alpha_{mn'})} c_n \quad (7)$$

$$\tilde{c}_n = \sum_{m=1}^{T_q} \frac{\exp(\alpha_{mn})}{\sum_{m'=1}^{T_q} \exp(\alpha_{m'n})} q_{m'} \quad (8)$$

q_m 表示矩阵 Q 的第 m 行， \tilde{q}_m 表示矩阵 \tilde{Q} 的第 m 行。 c_n 表示矩阵 C 的第 n 行， \tilde{c}_n 表示矩阵 \tilde{C} 的第 n 行。拼接 Q 和 \tilde{Q} ，然后全连接激活，得到 \bar{C} ，如下式所示：

$$\bar{Q} = \text{ReLU}([Q; \tilde{Q}; |Q - \tilde{Q}|; Q \cdot \tilde{Q}]W_1) \quad (9)$$

$$\bar{C} = \text{ReLU}([C; \tilde{C}; |C - \tilde{C}|; C \cdot \tilde{C}]W_1) \quad (10)$$

然后将 \bar{C} 切分成 K 份，也就是还原成 K 个实例矩阵 $\bar{S}_k \in R^{T_k \times d_n}$ 。将 \bar{S}_k 和 \tilde{Q} 分别输入单层

双向 LSTM，得到 $\bar{S}_k \in R^{T_k \times 2d_h}$, $\bar{Q} \in R^{T_q \times 2d_h}$ 。使用最大池化和平均池化将矩阵降维成向量。计算过程如下所示：

$$\hat{s}_k = [\max((\hat{S})_k); \text{ave}(\hat{S}_k)], \forall k \in 1, \dots, K \quad (11)$$

$$\hat{q} = [\max(\hat{Q}); \text{ave}(\hat{Q})] \quad (12)$$

instance matching and aggregation 就是计算全局注意力。用 query 和每个实例计算，得到关联度作为实例的权重，然后对实例加权得到类别的原型。如下所示：

$$\beta_k = v^T (\text{ReLU}(W_2[\hat{s}_k; \hat{q}])) \quad (13)$$

$$\hat{s} = \sum_{k=1}^K \frac{\exp(\beta_k)}{\sum_{k=1}^K \exp(\beta_k)} \hat{s}_k \quad (14)$$

最后的 class matching 就是分类，query 和每个原型计算匹配度，这里没有用欧氏距离损失，用的是交叉熵损失，如下所示：

$$J_{\text{match}} = -\frac{1}{R} \sum_{(q,l) \in Q} P(l|S, q) \quad (15)$$

$$P(S|l, q) = \frac{\exp(f(\{s_k^l\}_{k=1}^K, q))}{\sum_{i=1}^N \exp(f(\{s_k^i\}_{k=1}^K, q))} \quad (16)$$

$$f(\{s_k^l\}_{k=1}^K, q) = v^T (\text{ReLU}(W_2[\hat{s}; \hat{q}])) \quad (17)$$

最后一个匹配度函数，第二个表示 query 和类别实例集合 S 匹配概率，l 是关系类别，也就是标签。损失函数中除了交叉熵，还加上了聚合度损失 J。J 表示每个原型和对应类别实例的聚合程度。

$$J_{\text{incon}} = \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K \|\hat{s}_k^i - \hat{s}^i\|_2 \quad (18)$$

最终损失函数是两者相加 $J = J_{\text{match}} + \lambda J_{\text{incon}}$ 。

3.2.2 论文复现

采用开源数据集 DuIE2.0，按照少样本方法划分数据。随机选择 26 个关系作为训练集，6 个关系作为验证集，8 个关系作为测试集，每类关系随机选择 100 个样本。选择 5-way-1-shot、5-way-5-shot、5-way-10-shot、10-way-1-shot 四种设置测试模型性能，由于 GPU 内存不足无法运行 10-way-10-shot 场景。关于 N-way-K-shot 的解释可参考博客⁴。测试结果如表 4 所示。在 5-way-10-shot 场景下可接近有监督学习方法，但是当样本数量增加到 10 个时，性能急剧下降，10-way-1-shot 相比 5-way-1-shot 少了 20 个百分点。因此少样本学习方法在样本类别少的情况下表现较好。

4 总结

关系抽取技术主要有监督学习方法、少样本学习和远程监督。在大量有标签数据场景下，深度学习取得了很好的效果。然而我们面对的通常只有少量有标签数据，在这种情况下人们

⁴https://blog.csdn.net/qq_27668313/article/details/118192492?spm=1001.2014.3001.5501

表 4: 测试结果

	P	R	F1
5-way-1-shot	87.76	87.76	87.76
5-way-5-shot	94.34	94.34	94.34
5-way-10-shot	96.53	96.53	96.53
10-way-1-shot	67.75	67.75	67.75

提出了少样本学习和远程监督方法。远程监督方法可以自动生成大量有标签数据，这种方法理论上能取得和有监督学习相媲美的效果，但是由于数据中存在大量噪声，导致实际效果与监督学习方法相差甚远，在实际业务中难以应用。少样本学习是解决样本稀缺的另一种方法，尤其是原型网络方法，在图像分类领域取得不错的效果。但这种方法训练效率极慢，尤其是在多类别情况下，计算原型非常耗时。从长远来看，远程监督方法仍是最具前景的方法，如何从含有大量噪声数据中寻找有效样本是这种方法的最大难点。

参考文献

- [1] XU H, TIANYU G, YANKAI L, et al. More Data, More Relations, More Context and More Openness: A Review and Outlook for Relation Extraction[J]. arXiv, 2020.
- [2] MENG J, JINGBO S, TAYLOR C, et al. Metapad: Meta pattern discovery from massive text corpora[C]. [S.l.: s.n.], 2017: 877-886.
- [3] ZHENG S, YU P, CHEN L, et al. Diag-nre: A deep pattern diagnosis framework for distant supervision neural relation extraction[C]. [S.l.: s.n.], 2018: 1419-1429.
- [4] PAWAR S, PALSHIKAR G K, BHATTACHARYYA P. Relation extraction: A survey[J]. arXiv, 2017.
- [5] HAO P, TIANYU G, XU H, et al. Learning from Context or Names? An Empirical Study on Neural Relation Extraction [C]. [S.l.]: Association for Computational Linguistics, 2020.
- [6] D Z, K L, S L, et al. Relation Classification via Convolutional Deep Neural Network[C]. [S.l.: s.n.], 2014.
- [7] DONGXU Z, DONG W. Relation Classification via Recurrent Neural Network[C]. [S.l.: s.n.], 2015.
- [8] YANG L, FURU W, SUJIAN L, et al. A Dependency-Based Neural Network for Relation Classification[C]. [S.l.]: Association for Computational Linguistics, 2015.
- [9] SHU Z, DEQUAN Z, XINCHEN H, et al. Bidirectional Long Short-Term Memory Networks for Relation Classification [C]. [S.l.: s.n.], 2015.
- [10] SHANCHAN W, YIFAN H. Enriching Pre-trained Language Model with Entity Information for Relation Classification [C]. [S.l.: s.n.], 2019.
- [11] YUHAO Z, VICTOR Z, DANQI C, et al. Position-aware Attention and Supervised Data Improve Slot Filling[C]. [S.l.]: Association for Computational Linguistics, 2017.
- [12] LINLIN W, ZHU C, DE MELO G, et al. Relation Classification via Multi-Level Attention CNNs[C]. [S.l.]: Association for Computational Linguistics, 2016.
- [13] DOS SANTOS C, BING X, BOWEN Z. Classifying Relations by Ranking with Convolutional Neural Networks[C]. [S.l.]: Association for Computational Linguistics, 2015.
- [14] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning[C]. [S.l.: s.n.], 2016.
- [15] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]. [S.l.: s.n.], 2016.
- [16] BENGIO Y. Deep learning of representations for unsupervised and transfer learning[C]. [S.l.: s.n.], 2012.

- [17] RICH C. Learning many related tasks at the same time with backpropagation[C]. [S.l.: s.n.], 1995.
- [18] JAKE S, KEVIN S, RICHARD S Z. Prototypical Networks for Few-shot Learning[C]. [S.l.: s.n.], 2017.
- [19] TIANYU G, XU H, ZHIYUAN L, et al. Hybrid Attention-Based Prototypical Networks for Noisy Few-Shot Relation Classification[C]. [S.l.]: AAAI Conference on Artificial Intelligence, 2019.
- [20] ZHIXIU Y, ZHENHUA L. Multi-level matching and aggregation network for few-shot relation classification[C]. Florence, Italy: Association for Computational Linguistics, 2019: 2872-2881.
- [21] SOARES L B, NICHOLAS F, JEFFREY L, et al. Matching the Blanks: Distributional Similarity for Relation Learning [C]. [S.l.]: Association for Computational Linguistics, 2019.
- [22] MIKE M, STEVEN B, RION S, et al. Distant supervision for relation extraction without labeled data[C]. [S.l.]: Association for Computational Linguistics, 2009.
- [23] SEBASTIAN R, LIMIN Y, ANDREW M. Modeling relations and their mentions without labeled text[C]. [S.l.: s.n.], 2010.
- [24] DAOJIAN Z, KANG L, YUBO C, et al. Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks[C]. [S.l.]: Association for Computational Linguistics, 2015.
- [25] YANKAI L, SHIQI S, ZHIYUAN L, et al. Neural Relation Extraction with Selective Attention over Instances[C]. [S.l.]: Association for Computational Linguistics, 2016.
- [26] XU H, PENGFEI Y, ZHIYUAN L, et al. Hierarchical Relation Extraction with Coarse-to-Fine Grained Attention[C]. [S.l.]: Association for Computational Linguistics, 2018.
- [27] XU H, ZHIYUAN L, MAOSONG S. Neural Knowledge Acquisition via Mutual Attention between Knowledge Graph and Text[C]. [S.l.]: AAAI Conference on Artificial Intelligence, 2018.
- [28] GUOLIANG J, KANG L, SHIZHU H, et al. Distant Supervision for Relation Extraction with Sentence-Level Attention and Entity Descriptions[C]. [S.l.]: AAAI Conference on Artificial Intelligence, 2017.
- [29] J F, M H, L Z, et al. Reinforcement learning for relation classification from noisy data[C]. [S.l.]: AAAI Conference on Artificial Intelligence, 2018.
- [30] ZENG X, HE S, LIU K, et al. Large scaled relation extraction with reinforcement learning[C]. [S.l.]: AAAI Conference on Artificial Intelligence, 2018.
- [31] YI W, DAVID B, STUART R. Adversarial training for relation extraction[C]. [S.l.]: Association for Computational Linguistics, 2017.
- [32] XU H, ZHIYUAN L, MAOSONG S. Denoising distant supervision for relation extraction via instance-level adversarial training[J]. arXiv, 2018.
- [33] XU H, ZHIYUAN L, MAOSONG S. Denoising distant supervision for relation extraction via instance-level adversarial training[C]. [S.l.: s.n.], 2018.
- [34] 韩先培, 王志春, 张江涛, 等. CCKS2019 知识图谱测评技术报告: 实体、关系、事件及问答[J]. 中文信息学报, 2019.