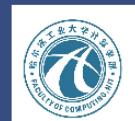




基于分段注意力匹配网络的跨领域少样本关系分类



戴尚峰; 孙承杰; 单莉莉; 林磊; 刘秉权

哈尔滨工业大学智能技术与自然语言处理实验室

<http://insun.hit.edu.cn>

研究背景

在互联网中存在着大量的文本数据，信息抽取可以从文本中提取出有效信息为我们所用，而关系抽取是信息抽取中重要的步骤之一，现有的关系抽取方法依赖于大量的标注语料，这需要大量的人力物力，而远程监督的方法又存在齐噪声和长尾分布问题。由此，少样本关系抽取问题被提出。

方法

■ 分段注意力匹配网络(PAMN)模型

- 模型结构如图1所示
- 编码层
 - ◆ 使用BERT对句子进行单词级别的编码
- 句子匹配层
 - ◆ 使用分段注意力机制计算两个句子间的相似度

■ 动态段长

- 三个数据集的句长和段长分布均有较大差异
 - ◆ 对于FewRel2.0验证集和FewRel1.0训练集、验证集这三个数据集的句长均值、标准差以及句子根据实体分为三段后的段长均值和标准差进行统计，统计结果如图2所示
- 用动态段长实现领域的自适应
 - ◆ 实验中为FewRel2.0设置较长的句长和段长，这样对不同批次的数据段长长度均为较为合理的值

表1. 不同模型在FewRel2.0测试集上准确率

模型	5-1	5-5	10-1	10-5
Proto(BERT)	41.90%	54.74%	27.36%	37.40%
Proto-ADV(CNN)	42.21%	58.71%	28.91%	44.35%
BERT-PAIR	67.41%	78.57%	54.89%	66.85%
PAMN	77.54%	90.40%	65.98%	82.03%

表2. 不同模型在FewRel1.0验证集上准确率

模型	5-1	5-5	10-1	10-5
MLMAN	82.98%	92.66%	73.59%	87.29%
BERT-PAIR	88.32%	93.22%	80.63%	87.02%
MTB	93.86%	97.06%	89.20%	94.27%
PAMN	90.10%	95.22%	84.13%	90.58%

表3. 不同分段数类型在FewRel2.0验证集上准确率

模型	5-1	5-5	10-1	10-5
PAMN (one piece)	72.73%	82.38%	62.15%	73.13%
PAMN (three piece)	77.78%	86.88%	68.07%	79.26%
PAMN (four piece)	76.43%	86.66%	65.77%	78.34%

表4. 十二学生与十生经向量模型在FewRel2.0验证集上准确率

模型	5-1	5-5	10-1	10-5
PAMN (un-siamese)	77.78%	86.88%	68.07%	79.26%
PAMN (siamese)	74.98%	83.61%	64.75%	75.49%

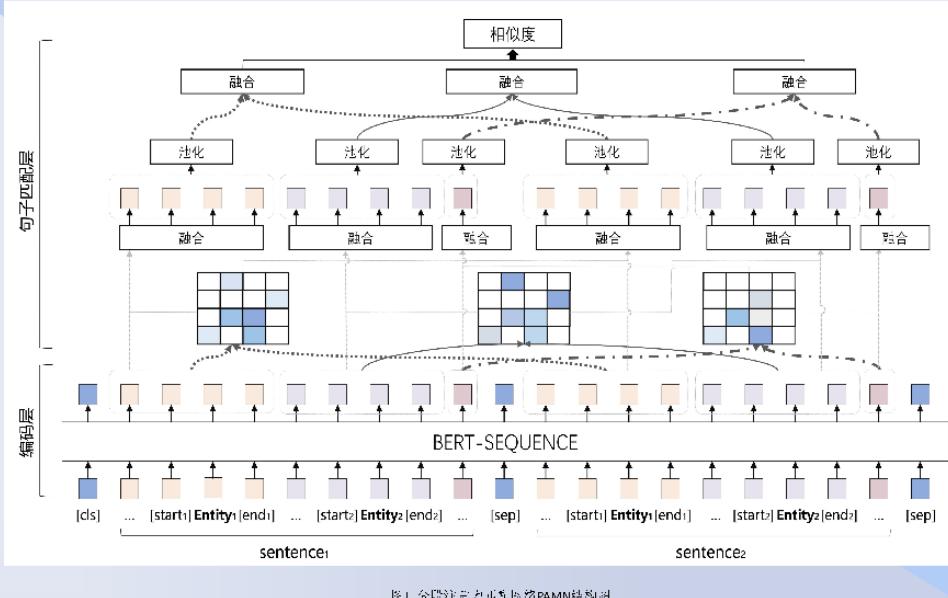


图1. 分段注意力匹配网络PAMN结构图

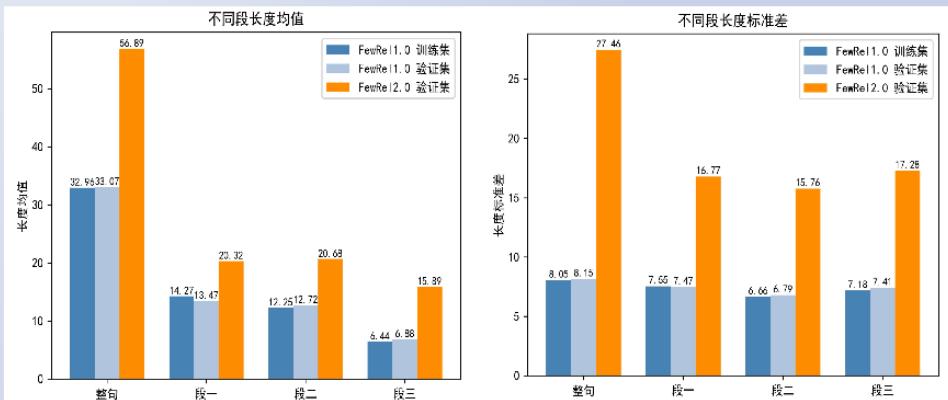


图2. FewRel1.0训练集、验证集和FewRel2.0验证集不同段的长度均值、标准差

实验结果

• FewRel2.0测评

- 与FewRel2.0论文中的模型Proto(BERT), Proto-ADV(CNN), BERT-PAIR进行对比，如表1所示，PAMN在FewRel2.0领域适应任务测评中取得了当前榜单上的最好效果
- FewRel1.0测评
 - 与FewRel1.0测评中效果最好的模型BERT-PAIR, MLMAN进行对比，实验结果如表2所示，PAMN仍然表现出较好的结果

实验数据

■ 数据来源

- 训练集： FewRel1.0训练集
- 验证集： FewRel1.0验证集和FewRel2.0领域适应任务验证集
- 测试集： FewRel2.0领域适应任务测评数据集

■ 分类数据的构造

- 将数据构建为少样本N way K shot分类数据

对比实验

• 分段注意力机制对模型的影响

- 结果如表3所示，可见将句子分为四段比分为三段模型准确率要低，推测是因为不同段向特征中噪音较多，导致真正能对文本匹配起到效果的特征混在噪音中
- 学生结构对模型的影响
 - 结果如表4所示，可见十生经向量比十生网络的效果要好，推测是使用相同的参数对向量进行映射会导致向量相似度增加，向量的部分特征会被减弱，导致在句子匹配层误差增加，模型性能降低

结论

本文提出了基于分段注意力机制的跨领域少样本关系分类方法PAMN，通过句子相似度计算的方法计算查询实例和支撑集合实例间的相似度，具有良好的领域适应性，同时针对关系分类实例使用分段注意力机制进行分段匹配，使得模型可以更准确的计算关系分类实例间的句子相似度，最后针对不同领域间段长分布差异的问题，使用动态段长进行段长领域自适应。PAMN取得了目前FewRel2.0领域适应测评榜单的最好效果。

致谢

本项目由国家重点研发计划（2018YFC0830700）、国家自然科学基金（61672192）、深圳市技术创新计划-基础研究项目（JCYJ20170808110410773）支持。