**小样本机器学习在NLP中的应用调研报告**

**论文：*Hybrid Attention-Based Prototypical Networks for Noisy Few-Shot Relation Classification***

**会议：**AAAI2019

**代码：**<https://github.com/thunlp/HATT-Proto>

**博客：**<https://blog.csdn.net/m0_38031488/article/details/90517760>

**摘要：**

关系分类（RC）是信息提取中的一项导入任务，旨在根据两个给定实体之间的相关上下文对它们之间的关系进行分类。 能够提取文本信息并使许多NLP应用领域受益（例如，信息检索，对话生成和问题回答）。

关系分类（RC）的现有方法主要依赖于远程监督（DS），因为尚无法获得大规模的监督训练数据集。通过将RC定义为小样本学习（FSL）问题，本文对RC提出了不同的看法，提出了基于混合Attention力的原型网络，用于解决噪声少的RC问题。 本文的Attention力模型加快了RC模型的收敛速度。

**相关工作及背景知识：**

远程监督（DS）机制，通过将现有知识图（KG）与文本对齐来自动标记训练实例。 DS是一种启发式规则：对于KG中的一个实体对，那些提及两个实体的句子将在KG中标有它们的关系。

原型网络易于实施，易于训练并且在多项小样本机器学习（FSL）任务上均达到了最新水平。 它为每个类计算原型，并通过计算其欧几里得距离对查询实例进行分类。 本文提出的方法是基于原型网络的。

**本文贡献：**

为嘈杂的小样本RC任务提出了基于混合Attention的原型网络。

类似于香草原型网络（vanilla prototypical networks），采用神经网络将所有实例嵌入支持集，并通过这些实例嵌入为每个关系计算特征向量（原型）。然后，通过测量查询实例嵌入和关系原型之间的距离，对查询实例中提到的实体对之间的关​​系进行分类。

**模型框架及算法：**

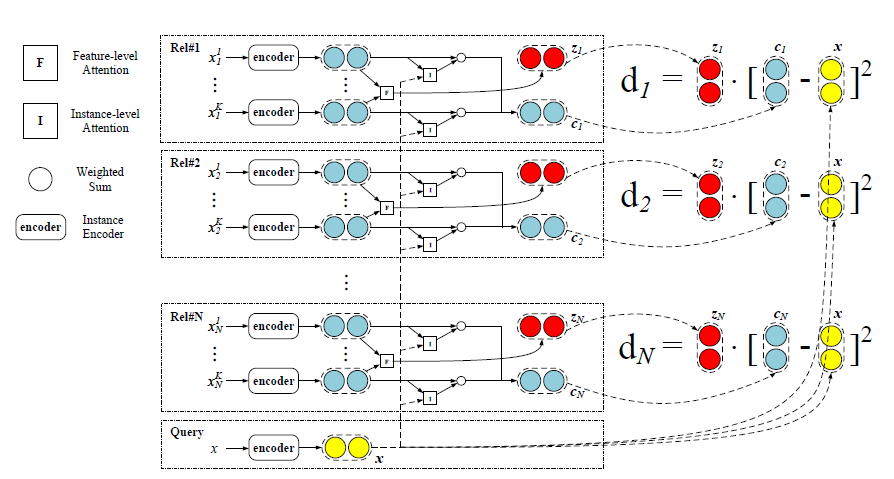


图1基于混合Attention力的原型网络的框架

模型采用了混合Attention，包括实例级Attention和特征级Attention。实例级别的关注模块能够在支持集中选择更多的信息实例，并在训练过程中对那些嘈杂的实例进行去噪。特征级关注模块可以突出显示特征空间中的重要维度，并针对不同关系制定特定的距离函数，这使本文的模型可以减轻特征稀疏性的问题。

实例编码器。给定一个实例及其提到的实体对，本文使用神经网络将实例语义编码为嵌入。 考虑到模型性能和时间效率，本文采用卷积神经网络（CNN）实现实例编码器。

原型网络。在计算实例嵌入之后，本文采用原型网络通过支持集中的实例嵌入为每个关系计算原型。 通过比较查询实例嵌入与每个关系原型之间的距离，本文最终可以对查询实例中提到的实体对之间的关​​系进行分类。

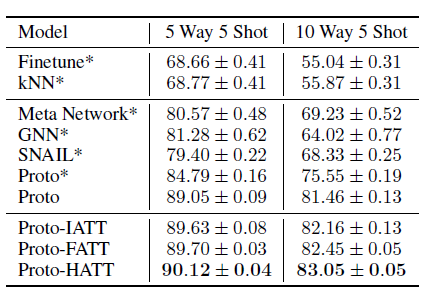
混合Attention。提出了一种混合Attention力机制，以提高分类性能和收敛速度。 本文的混合Attention机制包括两部分：实例级别的关注模块，用于帮助确定更多与查询相关的实例，以为每个关系计算出更好的原型；特征级别的关注模块，用于缓解特征稀疏性问题并测量对象中的空间距离比较合适的方法。 训练期间，两个部分会相互配合。混合Attention由两个模块组成，instance-level Attention模块用于在支持集中选择更多的信息实例，而feature-level Attention模块用于突出显示距离函数中的重要维度。

**实验结果及使用数据集：**

数据集为FewRel，<https://github.com/thunlp/FewRel>，它具有用于培训的64个关系，用于验证的16个关系和用于测试的20个关系。 训练和测试集之间没有重叠的关系。 每个关系在FewRel中都有700个实例。

实验使用了四个层次的随机噪声设置：无噪声数据、10%噪声数据、30%噪声数据和50%噪声数据。本文使用GloVe预先训练过的嵌入词（Pennington、Socher和Manning 2014）作为初始嵌入。在实践中，本文选择了嵌入集（Wikipedia 2014+Gigaword 5），它包含6B令牌、400K词汇和50维的单词嵌入。

实验结果。基于混合Attention的原型模型收敛速度比原始模型快，并且这种现象在数据噪声的影响下更加明显，通过使用混合Attention力并对实例和特征给予不同的评分，本文的模型可以知道训练时要关注的实例和特征的哪些部分，同时捕获正确的反向传播路径。 这有助于模型抵抗数据噪声的不利影响。在分类的准确率上达到了最好的结果。



**总结：**

在本文中，提出了一种基于混合Attention力的原型网络，用于嘈杂的小样本关系分类任务。 本文的混合Attention包括两个模块，一个实例级别的Attention（突出显示那些与查询相关的实例），以及一个特征级别的Attention力（减轻特征稀疏性的问题）。 在本文的实验中，本文使用几种随机噪声设置和少量触发设置评估模型，这表明本文的混合Attention力显着提高了FSL模型的鲁棒性和效率。