**Few-shot Slot Tagging with Collapsed Dependency Transfer and Label-enhanced Task-adaptive Projection Network**

会议：ACL2020

代码：<https://github.com/AtmaHou/FewShotTagging>

<https://www.cnblogs.com/shona/p/11966557.html>

<https://blog.csdn.net/u013713117/article/details/54974363>

<https://blog.csdn.net/jesseyule/article/details/105105886>

<https://blog.csdn.net/fkyyly/article/details/83791833>

<http://blog.csdn.net/pipisorry/article/details/78397543>

摘要

在这篇文章中，我们探讨了只有几个标记的支持句（也就是说，小样本）的时隙标记。与其他Few-shot分类问题相比，小样本槽标记面临独特的挑战，因为它需要建模标签之间的依赖关系。但是，由于标签集的差异，很难将以前学习到的标签依赖项应用到未见的域。为了解决这个问题，我们在条件随机场（CRF）中引入了一个折叠的依赖转移机制，将抽象标签依赖模式作为转移分数进行转移。在小样本设置中，CRF的emission score可以计算为单词与每个标签表示的相似度。为了计算这种相似性，我们提出了一种基于最新的小样本分类模型-TapNet-标签增强任务自适应投影网络（L-TapNet）。实验结果表明，我们的模型在单样本设置下，显著优于最强的小样本学习基线14.64 F1分。

**1引言**

槽标记（Tur and De Mori，2011）是面向任务的对话系统中的一个关键模块（Young et al.，2013），通常被表述为序列标记问题（Sarikaya et al.，2016）。槽标记面对的是域的快速变化，对于样本较少的新域，标记数据通常比较稀少。小样本学习技术（Miller等人，2000；Fei-Fei等人，2006；Lake等人，2015；Vinyals等人，2016）在这种情况下很有吸引力，因为它学习的模型借鉴了旧域的经验，并快速适应新域，只有很少的例子（通常每个类有一个或两个例子）。

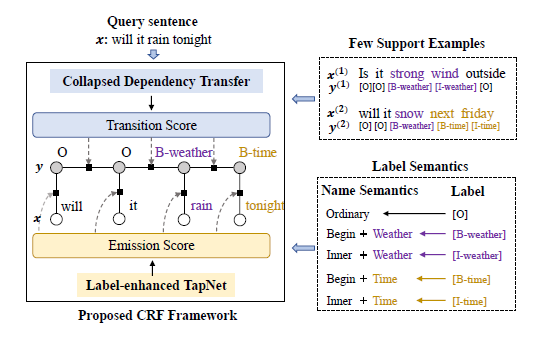


图1：我们的用于时隙标记的小样本CRF框架。

以往的小样本学习研究主要集中在分类问题上，这些问题已被基于相似性的方法广泛探讨（Vinyals等人，2016；Snell等人，2017；Sung等人，2018；Yan等人，2018；Yu等人，2018）。这些方法的基本思想是根据查询项与每个类的表示的相似性对其进行分类。相似函数通常是在先前的丰富资源域中学习的，每类表示是从少量的标记样本（支持集）中获得的。将小样本序列标记分解为一系列独立的小样本分类，并应用基于相似度的方法是非常简单的。然而，考虑到标签之间的依赖性，序列标签会带来好处（Huang等人，2015；Ma和Hovy，2016）。为了同时考虑项目相似性和标签依赖性，我们建议在小样本序列标签中利用条件随机字段（Lafferty等人，2001，CRF）（见图1）。本文将CRF的排放分数转化为基于相似度的方法的输出，并用一种特殊设计的传递机制来计算传递分数。

在学习CRF的发射和转换分数方面，小样本场景提出了独特的挑战。由于标签集的差异，源域中的标签依赖性和先前的标签依赖性无法直接传输。为了解决标签差异问题，我们引入了折叠依赖传递机制。它通过将特定于域的标签抽象为与域无关的抽象标签，并对这些抽象标签之间的标签依赖关系进行建模，将标签依赖信息从源域传输到目标域。

计算排放分数（在我们的例子中是单词标签相似性）也很有挑战性。流行的小样本模型，如Prototypic Network（Snell et al.，2017），将每个标签的支持示例的嵌入作为标签表示进行平均，这些支持示例通常紧密分布在嵌入空间中，从而导致错误分类。为了解决这个问题，Yoon等人。（2019）提出TapNet，学习将嵌入投影到不同标签的单词分隔良好的空间。我们将这一思想引入时隙标记，并进一步提出利用标签名称的语义来改进标签表示。我们认为标签名称通常在语义上与时隙词相关，可以帮助词标签相似性建模。例如，在图1中，单词rain和标签名weather高度相关。为了利用标签名称语义实现标签表示的良好分离，我们提出了标签增强TapNet（L-TapNet），它利用标签名称语义构造一个嵌入投影空间，其中标签表示与标签名称和槽字的嵌入分离并对齐。然后计算投影嵌入空间的相似度。此外，我们还引入了一种成对嵌入机制来表示具有特定领域上下文的单词。

在时隙标注和命名实体识别方面的一样本和五样本实验表明，我们的模型比强的小样本学习基线有显著的改进。消融试验显示LTapNet和塌陷依赖性转移都有改善。对标签依赖性的进一步分析表明，它捕获了非平凡的信息，并优于基于规则的转换。

本文的主要工作概括如下：

（1）提出了一个用于时隙标记的小样本CRF框架，该框架将发射分数计算为word label相似度，并通过转移先前学习到的标签依赖性来估计转换分数。

（2） 我们引入折叠依赖项传输机制来跨具有不同标签集的域传输标签依赖项。

（3） 我们提出L-TapNet来利用标签名称的语义来增强标签表示，这有助于建立单词标签相似性的模型。

**2 Problem Definition**

我们将句子定义为单词序列，并将句子的标签序列定义为。域是一组对。对于每个域，都有一个对应的特定于域的标签集。为了简化描述，我们假设所有域的标签数N相同。

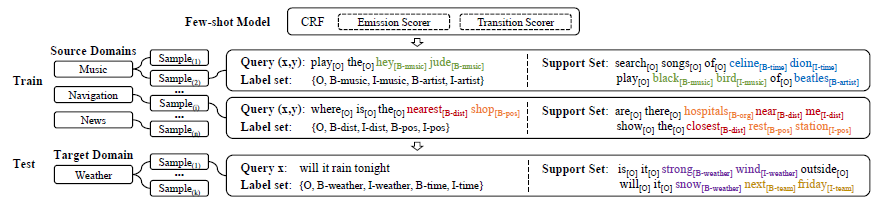


图2：培训和测试概述。此图说明了在一组源域上训练模型，并在只有支持集的未显示域上测试模型的过程。

如图2所示，小样本模型通常首先在一组源域上进行训练，然后直接在另一组未见的目标域上工作，而不进行微调。目标域只包含少量标记样本，称为支持集。S通常包括N个标签（N-way）中每个标签的k个示例（k-shot）。

K-shot序列标记任务定义如下：给定K-shot支持集S和输入查询序列，查找x的最佳标记序列：



**3 Model**

在这一节中，我们首先展示拟议的CRF框架（x3.1）的概述。然后讨论了如何用折叠依赖转移（x3.2）计算标签转移得分和用L-TapNet（x3.3）计算排放得分。<https://blog.csdn.net/cuipanguo/article/details/100747063>

**3.1 Framework Overview**

条件随机场（CRF）同时考虑转换分数和发射分数，为每个输入找到全局最优标签序列。遵循相同的思想，我们构建了包含两个组件的少数镜头槽标记框架：转换记分器和发射记分器。

通过对给定查询语句x和K-shot支持集S的标签y的标签概率建模，我们将线性CRF应用于Few-shot设置：





为转移得分器输出，为发射得分器输出。是平衡两个分数权重的缩放参数。

我们采用作为损失函数，并在源域数据上最小化它。在对模型进行训练之后，我们使用Viterbi算法（Forney，1973）来寻找每个输入的最佳标签序列。

**3.2 Transition Scorer**

转换记分器组件捕获标签之间的依赖关系。我们将标签依赖关系建模为两个标签之间的转换概率：



通常，这种概率是从训练数据中学习并存储在转移矩阵中的，其中N是标签的数目。例如，对应于 。但在few-shot设置中，模型在源域（train）和目标域（test）中面对不同的标签集。标签上的这种不匹配会阻止经过训练的转换记分器直接在目标域上工作。

**折叠的依赖项传输机制**我们通过直接建模抽象标签之间的转换概率来解决上述问题。直观地，我们将特定的标签折叠成三个抽象的标签：O、B和I。为了区分两个标签是在同一语义下还是在不同语义下，我们对从B和I到同一B（sB）、不同B（dB）、相同I（sI）和不同I（dI）的转换进行了建模。我们用表~T 3 5记录这种抽象的标签转换（见图3）。

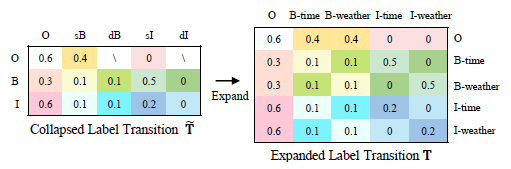


图3：折叠标签依赖项传输的示例。我们学习一个折叠的标签转换，并通过用相同颜色的值填充它的每个位置来获得特定的标签转换T。

例如，是两个相同B标签的转移概率。而是从B标签到不同类型I标签的转换概率，其中，和分别表示从O到任何B或I标签的转换概率。

为了计算一个新域的标签转移概率，我们用~T中的值来构造转移矩阵T。图3显示了填充过程，其中相同颜色的位置由相同的值填充。例如，我们用中的值填充。

**3.3 Emission Scorer**

如图4所示，排放记分器独立地为每个单词分配与每个标签相关的排放记分

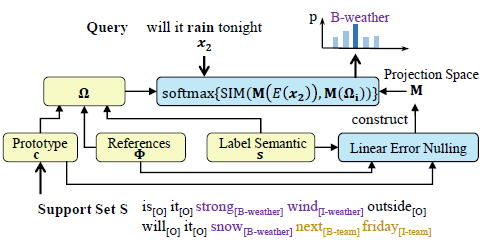


图4：带有L-TapNet的排放记分器。该算法首先对给定的区域进行线性误差置零，构造一个投影空间，然后在投影空间中预测一个词的发射分数及其与标签表示的距离。



在few-shot设置中，一个单词的发射分数是根据它与每个标签的表示的相似性来计算的。为了计算这种发射，我们提出了L-TapNet，通过使用标签语义和原型改进TapNet（Yoon et al.，2019）。

**3.3.1 Task-Adaptive Projection Network**

TapNet是目前最先进的few-shot分类模型。以往的few-shot模型，如原型网络，平均每个标签的嵌入支持示例作为标签表示，并直接计算单词嵌入空间中的单词标签相似度。不同于他们，

TapNet计算投影嵌入空间中的单词标签相似度，其中不同标签的单词被很好地分隔开。这使得TapNet可以减少错误分类。为了实现这一点，TapNet利用一组每个标签的参考向量作为标签表示。并在此基础上构造投影空间。然后，计算标签的单词x的发射分数，作为其与参考的相似性：



其中是投影函数，E是嵌入函数，SIM是相似函数。TapNet跨不同的域共享参考，并通过将引用随机关联到特定的标签来为每个特定的域构造M。

**任务自适应投影空间构造**在这里，我们对投影空间的构建做一个简要的介绍。设cj为支持集S中标签为的单词的嵌入特征的平均值。给定和支持集S，TapNet构造投影器，以便（1）当使用投影时，每个cj和相应的参考向量紧密对齐。（2） 在投影时，不同标签的单词被很好地分隔开。

为了达到这些目的，TapNet首先计算cj和在原始嵌入空间中的对齐偏差，然后找到一个投影M来消除这种对齐偏差，同时有效地分离不同的标签。具体来说，TapNet以线性误差置零过程linear error nulling的矩阵解作为嵌入方案。有关详细过程，请参阅原始文件。

**3.3.2 Label-enhanced TapNet**

如引言中所述，我们认为标签名称通常在语义上与槽字相关，并且可以帮助建立单词标签相似性模型。为了利用这些信息增强TapNet，我们在标签表示和投影空间的构造中都使用了标签语义。

**带有标签语义的投影空间**使原型 cj成为支持集中标签为的词的嵌入平均值。sj是标签的语义表示，第3.3.3节将详细介绍如何获取它。直观地说，时隙值（cj）和对应的标签名（sj）通常有相关的语义，它们应该在嵌入空间中接近。

所以，我们找到了一个将cj与和sj对齐的项目。TapNet的区别在于它只将cj与引用对齐，但我们也需要与标签表示对齐。标签增强参考计算如下：



其中是一个平衡因子。标签语义sj使每个域都是特定的。参考提供了跨域的泛化。

然后，我们按照TapNet的相同步骤，构造了标签增强参考和cj之间对准误差的线性零位。

**带有标签语义的发射分数**在排放评分计算中，与TapNet只使用领域不可知的引用作为标签表示相比，我们还考虑了标签语义，在标签表示中使用了标签增强的引用。

此外，我们还进一步结合了原型网络的思想，并使用原型参考cj表示标签，即。最后，计算x的发射分数作为其与标签表示的相似性：



其中SIM是点积相似函数，E是字嵌入函数，将在下一节介绍。

**3.3.3嵌入单词和标签名称**

对于单词嵌入函数E，我们提出了一种成对嵌入机制。如图5所示，当一个单词连接到不同的上下文时，其含义往往不同。为了解决相似度计算中的表示挑战，我们考虑了在few-shot学习中的特殊查询支持设置，并嵌入查询和支持词双管齐下。

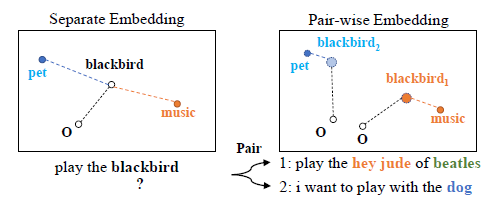


图5：成对嵌入的示例。当分别嵌入查询和支持句子时（左），很难根据黑鹂与标签的相似性来标记它。但是，如果我们通过将查询与不同的支持语句（右）配对来嵌入查询，则特定于域的上下文将分别为blackbird提供接近pet和song的特定含义。

这种成对嵌入可以在支持语句中利用领域相关上下文，为查询词提供领域自适应嵌入。这将进一步帮助建模查询词与特定于域的标签的相似性。为了实现这一点，我们在查询和支持词上用自我关注来表示每个词。我们首先将查询语句x复制次，并将它们与所有支持语句配对。然后将NS个对传递给BERT（Devlin et al.，2019），以获取每个查询词的NS个嵌入。我们将每个单词表示为NS个嵌入的平均值。现在，查询词的表示取决于特定于域的上下文。我们使用BERT，因为它可以自然地捕捉句子对之间的关系。

为了得到标签表示，我们首先连接抽象标签名（例如，begin and inner）和标签名（例如，天气）。然后，我们在第一个位置插入一个[CLS]标记，并将它们输入到一个BERT中。最后，使用[CLS]的表示作为标签语义嵌入。

**4 Experiment**

我们评估了所提出的时隙标记方法，并在一个类似的序列标记任务：名称实体识别（NER）上测试了其泛化能力。由于篇幅的限制，我们只给出了1-shot/5-shot时隙标记的详细结果，该标记将学习到的知识从源域（训练）转移到只包含1-shot/5-shot支持集的不可见目标域（测试）。NER的结果是一致的，我们在附录B中给出了它们。

**4.1设置**

数据集用于时隙标记，我们利用snips数据集（Coucke et al.，2018），因为它包含7个具有不同标签集的域，并且很容易模拟少数镜头的情况。这些领域包括天气（We）、音乐（Mu）、播放列表（Pl）、图书（Bo）、搜索屏幕（Se）、餐厅（Re）和创意作品（Cr）。原始数据集的信息见附录A。

为了模拟few-shot情况，我们从原始数据集构造了few-shot数据集，其中每个样本是一个查询数据和相应的K样本支持集的组合。表1显示了实验数据的概述。

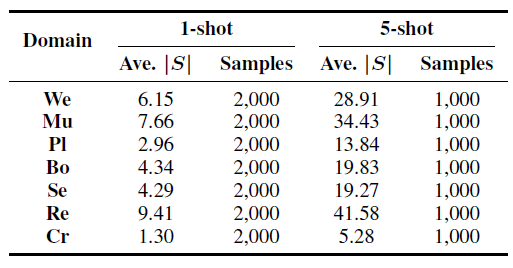
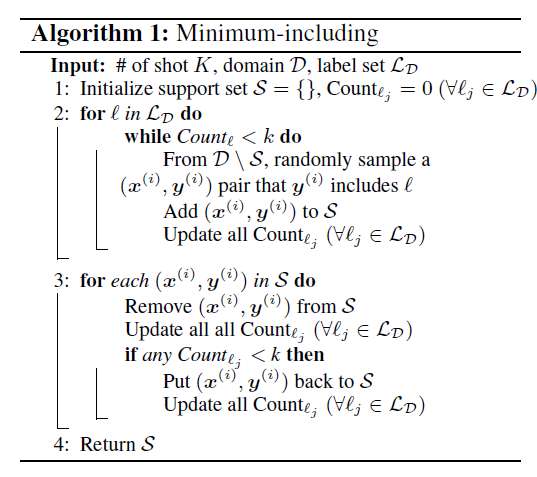


表1：few-shot槽标记数据概述。这里，“Ave “对应于每个域的平均支持集大小。而“Sample”代表我们从每个域构建的few-shot样本数。

**Few-shot Data Construction** 与简单的单字分类不同，时隙标注是一个覆盖整个句子的结构预测问题。因此，我们在每个标签下用句子而不是单个单词来构造支持集。

因此，常规的N-way-K-shotfew-shot定义不适用于few-shot时隙标注。我们不能保证每个标签在抽取支持句子时出现K次，因为不同的时隙标签随机地出现在一个句子中。例如，在图1中，在1-shot支持集中，label[B-weather]出现两次，以确保所有标签至少出现一次。因此，我们根据两个条件近似地构造K-shot支持集：（1）域中的所有标签在S中至少出现K次。（2）如果从中移除任何（x；y）对，则至少有一个标签在S中出现少于K次。算法1显示了详细的过程。



在这里，我们以1-shot时隙标记为例来说明数据构建过程。对于每个域，我们抽样100个不同的1-shot支持集。然后，对于每个支持集，我们将20个未包含的语句作为查询（查询集）进行采样。每个支持查询集对形成一个few-shot episode。最后，我们得到每个域100集和100X20个样本（1个查询语句和一个支持集）。

为了测试框架的健壮性，我们在不同的领域交叉验证了模型。每次，我们选择一个目标域进行测试，一个域进行开发，并使用其余域作为训练的源域。因此，对于时隙标记，所有模型都在10000个样本上进行训练，并分别在2000个样本上进行验证和测试。

在目标域上测试模型时，我们评估每个镜头集内的F1分数。然后我们从所有100个镜头集中平均100个F1分数作为最终结果，以对抗来自支持集的随机性。所有模型都在相同的支持查询集对上进行公平性评估。

为了控制神经网络训练的不确定性（Reimers和Gurevych，2017），我们报告了10个随机种子的平均得分。

超参数我们使用无基础的BERT基（Devlin et al.，2019）计算所有模型的上下文嵌入。我们使用ADAM（Kingma和Ba，2015）来培训批量大小为4、学习率为1e-5的模型。

对于CRF框架，我们在训练过程中学习尺度参数，这对于获得稳定的结果很重要。对于L-TapNet，我们将设为0.5，将设为0.7。我们用逐渐解冻的技巧微调BERT（Howard and Ruder，2018）。对于提议的模型和基线模型，，在固定步数下没有损耗衰减时， early stop训练和微调。

**4.2 Baselines**

Bi LSTM是一种双向LSTM（Schuster和Paliwal，1997年），嵌入GloVe（Pennington等人，2014年），用于插槽标记。它在支持集上进行训练，并在查询样本上进行测试。

SimBERT是一个根据non-finetuned-bert嵌入词的余弦相似性来预测标签的模型。对于每个单词xj，SimBERT在支持集中找到其最相似的单词x0 k，并且xj的标签被预测为x0k的标签。

transfer BERT是具有BERT的NER设置的域传输模型（Devlin等人，2019）。我们在源域上对它进行预处理，并在模型的同一开发集上选择最佳模型。我们只通过传输瓶颈特性来处理标签不匹配问题。在测试之前，我们在目标域支持集上对其进行微调。在培训和微调中，学习率设定为1e-5。

WarmProtoZero（WPZ）（Fritzler et al.，2019）是一种将序列标记视为每个单词分类的few-shot序列标记模型。它在源域上预先训练了一个原型网络（Snell等人，2017年），并利用它在不经过训练的情况下对目标域进行字级分类。Fritzler等人。（2019）使用随机初始化的单词嵌入。为了消除不同嵌入方法的影响，我们进一步使用预先训练的手套嵌入（Pennington等人，2014）和BERT来实现WPZ。

匹配网络（MN）类似于WPZ。唯一的区别是我们使用匹配网络（Vinyals等人，2016）和伯特嵌入进行分类。

**4.3 Main Results**

1-shot设置 表2的结果显示了1-shot槽标记结果。每列分别显示以某个域为目标域（test）和以其他域为源域（train&dev）的F1分数。如表所示，我们的L-TapNet+CDT实现了最佳性能。它的F1成绩平均为14.64分，优于最强的几杆学习基线WPZ+BERT。

我们的模型明显优于Bi LSTM和TransferBERT，这表明在很少的镜头设置下标记数据的数量对于传统的机器学习和传输来说都太少了。此外，SimBERT的性能证明了基于metric的方法比传统的机器学习模型在few-shot设置下的优越性。

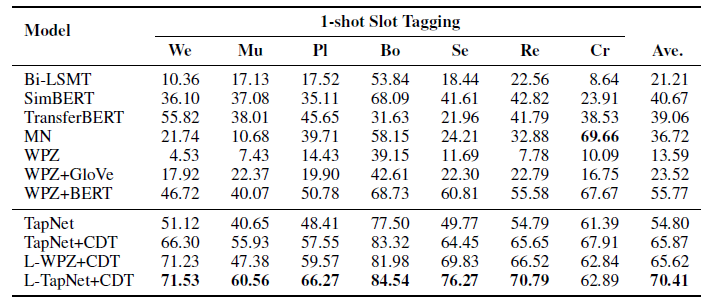


表2:F1得分。+CDT表示折叠的依赖项传输。中线以下的分数来自我们的方法，达到了最好的性能。Ave.显示平均得分。标准差结果见附录D。

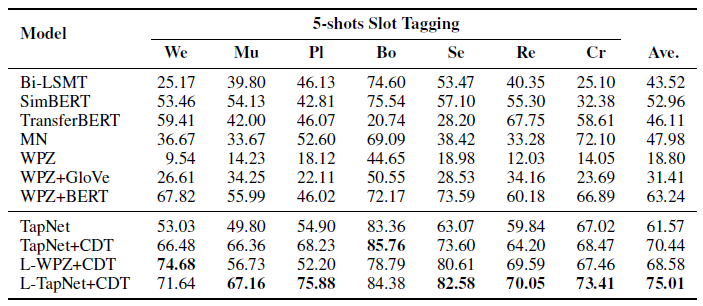


表3:5-shotF1得分结果。我们的方法取得了最好的效果。标准差结果见附录D。

原始的WarmProtoZero（WPZ）模型由于其单词嵌入的表示能力弱而受到影响。当我们使用GloVe和BERT词嵌入来增强它时，它的性能显著提高。这显示了few-shot设置中嵌入的重要性。匹配网络（MN）在这两种设置下都表现不佳。这在很大程度上是由于MN对所有支持词的关注程度相同，这使得它容易受到O-label数量不均衡的影响。

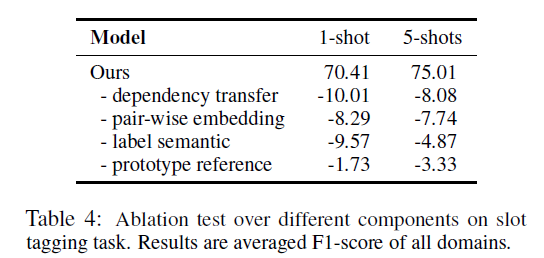
更具体地说，那些在支持集上进行微调的模型，比如Bi LSTM和TransferBERT，倾向于随机预测标签。这些系统只能处理从支持示例中易于概括的情况，例如专有名词标记的标记（例如城市名称和时间）。这表明，对极为有限的例子进行微调会导致泛化能力差和分类器训练不足。对于WPZ和MN等基于度量的方法，标签预测更为合理。然而，这些模型很容易被类似的标签混淆，比如当前位置和地理位置。它表明了良好分离的标签表示的必要性。此外，非法的标签转换也非常常见，这可以通过建议的折叠依赖项传输很好地解决。

为了消除由于标签名称中的附加信息引起的不公平比较，我们提出了L-WPZ+CDT，通过增强标签名称表示与L-TapNet相同的WarmProtoZero（WPZ）模型，并将其合并到所提出的CRF框架中。它结合了标签名称嵌入和原型作为每个标签表示。它对WPZ的改进主要来自于标签语义、折叠依赖传输和成对嵌入。L-TapNet+CDT比L-WPZ+CDT高4.79f1分，说明了嵌入投影的有效性。与TapNet+CDT相比，L-TapNet+CDT实现了平均提高了4.54f-score，这表明考虑标签语义和原型有助于改进排放评分计算。

5-shot设置 表3的结果显示了5-shot实验的结果，验证了该模型在更多镜头情况下的泛化能力。结果与一般趋势下的一次测斜结果一致。

**4.4 Analysis**

消融试验为了进一步了解我们的方法（L-TapNet+CDT）中的每个成分，我们对表4中的1次和5次消融设置进行了烧蚀分析。方法中的每个组件都被分别移除，包括：折叠依赖转移、成对嵌入、标签语义和原型引用。



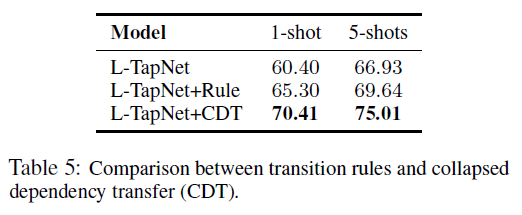
当折叠的依赖项转移被移除时，我们直接预测带有发射分数的标签，并且在所有设置中都看到巨大的F1分数下降。这种消融证明了考虑标签依赖性的必要性。对于没有成对嵌入的方法，我们独立地表示查询和支持语句。我们解决了支持语句可以提供领域相关上下文这一问题，而成对嵌入可以利用这种上下文并为查询语句中的单词提供领域自适应表示。这在计算单词与领域特定标签的相似性时有很大帮助。

当我们从LTapNet中去掉标签语义时，模型退化为TapNet+CDT，并在发射分数上增加了原型。结果表明，考虑标签名称可以提供更好的标签表示，有助于建立单词标签相似度模型。此外，我们还尝试删除标签表示中的内部和开头的单词，并观察到在一次截击中F1得分下降了0.97分。结果表明，在标签语义中区分B-I标签有助于标签的标注。

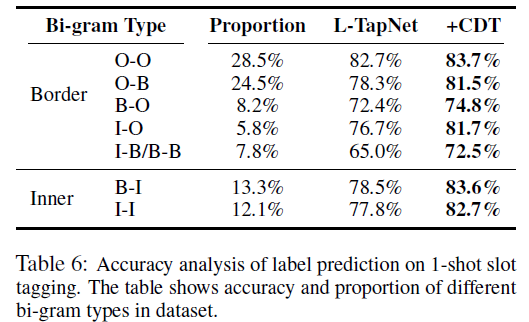
如果我们在没有原型参考的情况下计算排放分数，模型在5-shot设置中会失去更多的性能。这符合这样一种直觉，即原型允许模型从支持样本的增加中受益更多，因为原型直接从支持集派生。

分析折叠依赖转移（Collapsed Dependency Transfer）时，折叠依赖转移（Collapsed Dependency Transfer，CDT）带来了显著的改进，由此产生了两个自然的问题：CDT是否只是学习了简单的转移规则，以及它为什么工作。

为了回答第一个问题，我们用表5中的转换规则替换了CDT，这表明CDT比转换规则能带来更多的改进。



为了更深入地了解CDT的有效性，我们对其进行了精度分析。我们评估了不同类型标签的标签预测精度。结果如表6所示。我们进一步将bi grams归纳为两类：Border包括跨越槽跨边界的bigrams；Inner是槽跨内的bigrams。我们认为内部的改进成功地减少了CDT的非法标签转换。有趣的是，我们观察到，CDT还通过正确预测时隙跨度的第一个和最后一个标记带来了改进。边界分析的结果证实了CDT有助于更准确地确定槽跨的边界，而增加过渡规则是很难实现的。



**5 Related Works**

传统的few-shot学习方法高度依赖手工制作的特征（Fei Fei，2006；Fink，2005）。经典方法主要侧重于度量学习（Snell等人，2017；Vinyals等人，2016），该方法根据项目与每个类别表示的相似性对项目进行分类。最近的研究（Lu et al.，2018；Schwartz et al.，2019）提议利用类名的语义来增强类表示。然而，与我们不同的是，这些方法侧重于图像分类，其中名称语义的影响是隐式的，不需要标签依赖。

自然语言过程中的few-shot学习对分类任务进行了探索，包括文本分类（Sun et al.，2019；Geng et al.，2019；Yan et al.，2018；Yu et al.，2018）、实体关系分类（Lv et al.，2019；Gao et al.，2019；Ye and Ling，2019）和对话行为预测（Vlasov et al.，2018）。然而，针对时隙标记的few-shot学习研究较少。

罗等人。（2018）调查了使用附加正则表达式的few-shot槽标记，由于正则表达式的使用，这与我们的模型不可比。Fritzler等人。（2019）探索了fewshot命名实体识别的原型网络，其背景与我们相似。与之相比，该模型同时考虑了标签依赖传递和标签名称语义，取得了较好的性能。Zero-shot槽标记方法（Bapna et al.，2017；Lee and Jha，2019；Shah et al.，2019）在使用标签名称语义方面与我们有相似的想法，但有不同的设置，因为few-shot方法还受一些标签语句的支持。Chen等人。（2016）在意图检测中使用标签名称进行调查。除了直接从有限的示例中学习之外，解决NLP中数据稀缺问题的另一个研究方向是数据增强（Fader等人，2013；Zhang等人，2015；Liu等人，2017）。对于时隙标记的数据增强，探索基于句子生成的方法来创建附加标记样本（Hou等人，2018；Shin等人，2019；Yoo等人，2019）。

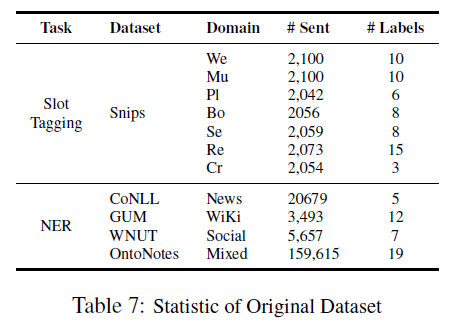
**6 Conclusion**

本文提出了一种面向任务的对话槽标注的few-shot CRF模型。为了计算few-shot设置下的转移得分，我们提出了折叠依赖转移机制，该机制将标签依赖的先验知识转移到具有不同标签集的域中。提出了L-TapNet来计算发射分数，改进了基于标签名称语义的标签表示方法。实验结果表明，折叠依赖转移和L-TapNet都能提高标签的正确率。

**Appendix**

A Detail of Dataset

表7显示了用于构建少量镜头实验数据的原始数据集的统计信息。



**B Few-shot experiments for Name entity recognition**

可将识别预定义名称实体（如人名、组织和位置）的名称实体识别（NER）建模为插槽标记任务。此外，新域的数据稀缺问题也存在于NER任务中。基于上述原因，我们很少进行shot-NER实验来测试模型的生成能力。

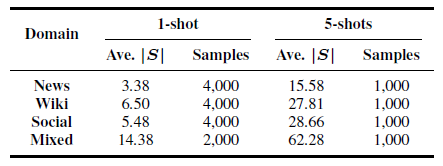


表8:NER实验的少量放炮数据概述。这里，“Ave.|S|”对应于每个域的平均支持集大小。而“Sample”代表我们从每个域构建的few-shot样本数。

对于命名实体识别，我们使用了4个不同的数据集：CoNLL-2003（Sang and Meulder，2003）、GUM（Zeldes，2017）、WNUT-2017（Derczynski et al.，2017）和Ontonotes（Pradhan et al.，2013），每个数据集仅包含一个域的数据。这四个领域是新闻，维基，社交和混合。原始数据集的详细情况见表7，构造的few-shot统计见表8。

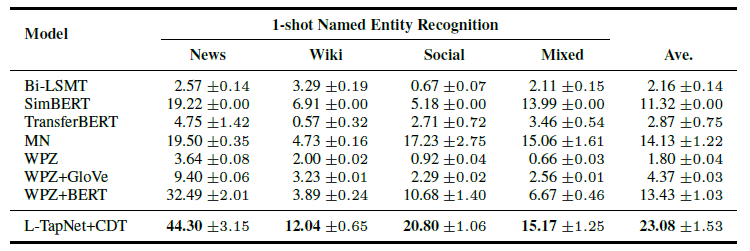


表9:F1在单镜头名称实体识别上的得分。CDT表示折叠的依赖项传输。中线以下的分数来自我们的车型，这些车型的性能最好。Ave.显示平均得分。

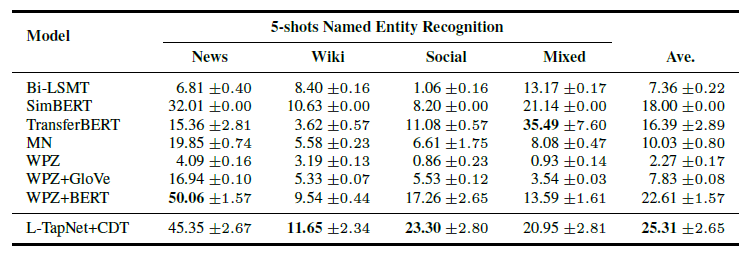


表10:。F1得分结果在5个镜头名称实体识别。我们的方法取得了最好的效果。

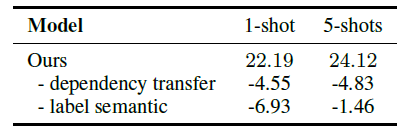


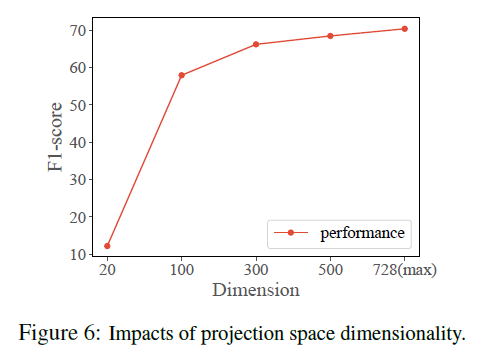
表11:NER任务中不同组分的消融试验。结果是所有领域的平均F1分数。

NER表9和表10的1-shot和5-shots结果分别显示了1-shot和5-shots名称实体识别结果。我们最好的模型在两种情况下都优于所有基线。

结果趋势与slot-tagging结果一致。但总分远低于插槽标记结果。这是因为NER域来自不同的数据集，并且域差距要大得多。

我们在5-shot上的改进在边距上缩小了。这是因为NER域有不同的类型和词汇。因此，与SNIPS相比，传递知识更困难，但依赖领域特定的支持示例更有利。这种趋势在更多的镜头中更加明显。在5样本设置中，最强的基线WPZ从增加的镜头中受益更多，因为它只使用支持集进行预测。但是对于我们的模型来说，更多的样本的好处是较弱的，因为它使用了更多的先验知识。

NER消融分析我们研究了内质网任务中折叠依赖转移和标记语义的有效性。我们对两个部件进行烧蚀并观察在单镜头和五镜头两种情况下，性能都有所下降，这证明了所提出的两种机制的泛化能力。



**C 投影空间维数分析**

图6显示了在LTapNet中使用不同投影空间维度时，1-shot的性能。

如图中趋势所示，随着映射空间维数的增加和逐渐稳定，模型的性能变得更好。这显示了在不损失太多性能的情况下降低维度的可能性（Yoon等人，2019）。

**D 带标准偏差的槽标记结果**

表12和13显示了插槽标记任务的标准差的完整结果。

