

A Joint Named-Entity Recognizer for Heterogeneous Tag-sets Using a Tag Hierarchy 论文总结

本文章主要针对论文 A Joint Named-Entity Recognizer for Heterogeneous Tag-sets Using a Tag Hierarchy 阅读后自我的总结，主要从论文介绍、论文内容、论文总结三个方面进行展开，具体论文的原文可以查看底部的链接

1. 论文介绍和动机

本论文是发表在 ACL2019 的一篇 paper，主要针对的领域为命名实体识别，由于某些命名实体的数据集（文中交代特别是医药领域）的 label 可以是多样的，文中举了一个例子如 Figure 1，从图中可以看出分别有 Tag-set1、Tag-set2、Tag-set3 三种不同的 label。针对这样的问题，传统的做法可能是分别先建立起层次结构（tag hierarchy），训练是分别可以通过单独训练（每种 label 单独训练各自模型的参数）或者联合训练（共享提取特征层，一般为 bilstm 等序列模型），最后在测试集上预测时通过 tag hierarchy 来推理得到最后的命名实体。

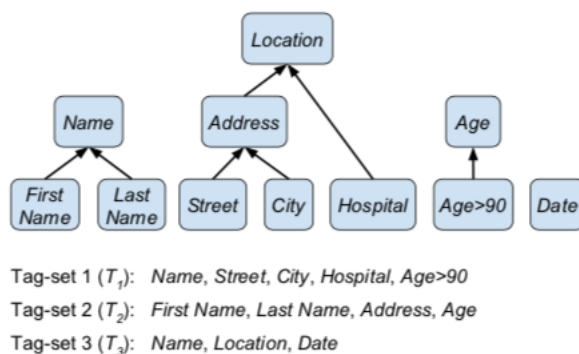


Figure 1: A tag hierarchy for three tag-sets.

为了解决上述的问题，本文提出 Tag Hierarchy Model，通过构建统一的 Tag Hierarchy。即将所有 label 统一为一种，这样训练时就不需要考虑单独或者联合训练，只需要训练一种命名实体模型就可以，构建内容如 Figure 4 所示，在下一节的论文内容中会详细讲解

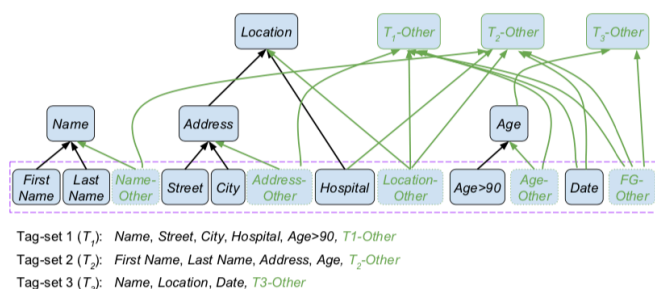


Figure 4: The tag hierarchy in Fig. 1 for three tag-sets after closure extension. Green nodes and edges were automatically added in this process. Fine-grained tags are surrounded by a dotted box.

2. 论文内容

本节主要讲解作者在证明其方法的时候所做各种实验的部分细节，其实为了证明其方法的优越性，本文分别介绍了单独训练（Indep, Independent Model）和联合训练（MTL, Multitasking for heterogeneous）的训练方式，并在最后对比了各种模型的指标，下面首先介绍一下 Indep 和 MTL 模型

2.1 Independent Model

其实 Independent Model 就是将 Figure 1 中不同的训练集分别构建单独的 NER 模型，模型框架如 Figure 2 所示。最后在组合预测阶段，首先将不同模型预测的 label 映射到测试集的标签上，然后将不同模型的预测结果统一到一起，文中分别用了三种方式进行统一（这里就不过多介绍了）

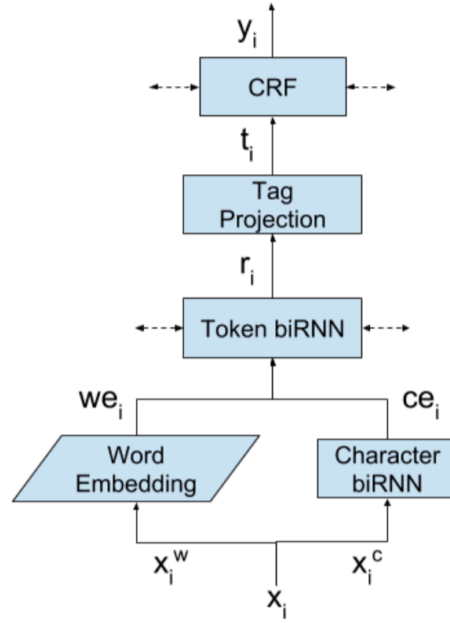


Figure 2: Neural architecture for NER.

2.2 Multitasking for heterogeneous

MTL 的模型如 Figure 3 所示，可以看到不同的 label 共享了输入层和特征提取层（biRNN），但是没共享 tagging layer(这也是本文创新点需要解决的事)，预测阶段和 Independent Model 在最后统一时做的一样。

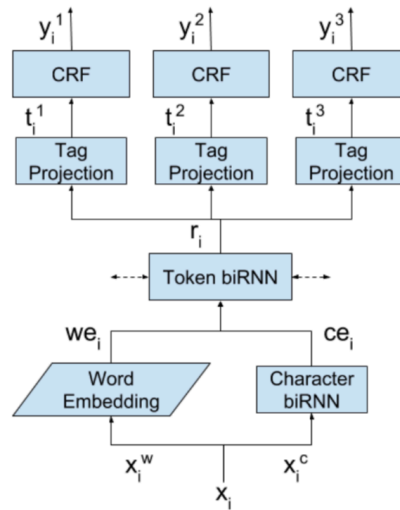


Figure 3: NER multitasking architecture for 3 tag-sets.

2.3 Tag Hierarchy Model

最后介绍下本文的创新点，本文创新点主要在与构建了一种（Tag Hierarchy）（1）将不同数据集的的 label 在预处理阶段统一了格式，而不需要像之前 model 一样需要在后处理阶段进行推理；（2）训练时只要训练一个如 Figure 2 一个单独的命名实体模型即可，即共享所有输入层、特征提取层和 tag layer。其中 Tag Hierarchy 如下图所示：

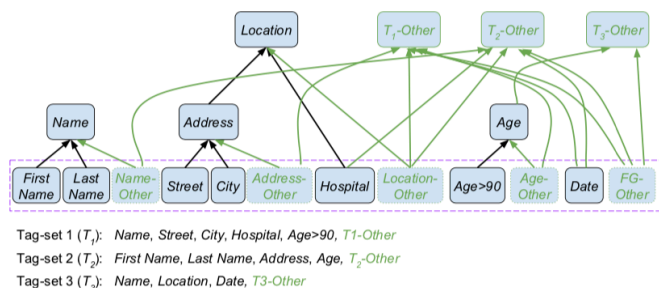


Figure 4: The tag hierarchy in Fig. 1 for three tag-sets after closure extension. Green nodes and edges were automatically added in this process. Fine-grained tags are surrounded by a dotted box.

其中在构建训练集 label 是分别从底往上寻找最外面的联合节点（这里联合是指几个数据集相同的根结点）作为 label。

3. 论文总结

其实本文在模型上并没有什么创新的地方，主要在预处理阶段组合 label 时应用了（Tag Hierarchy），共享所有参数且不需要进行后处理，论文创新主要在构建 label 的巧妙（语义相关性）。

论文原文：<https://www.aclweb.org/anthology/P19-1014.pdf>

本文讲解的 pdf 可以在我的 github 上查看：<https://github.com/GrinAndBear/NLP-Paper-Share/tree/master/KBQA>

看完麻烦大家点个赞或者 github 上 star 一下多谢啦