**Semi-supervised sequence tagging with bidirectional**

**language models**

**Matthew E. Peters, Waleed Ammar, Chandra Bhagavatula, Russell Power**

Allen Institute for Artificial Intelligence

{matthewp,waleeda,chandrab,russellp}@allenai.org

**Abstract**

Pre-trained word embeddings learned from unlabeled text have become a standard component of neural network archi- tectures for NLP tasks. However, in most cases, the recurrent network that operates on word-level representations to produce context sensitive representations is trained on relatively little labeled data. In this paper, we demonstrate a general semi-supervised approach for adding pre- trained context embeddings from bidi- rectional language models to NLP systems and apply it to sequence labeling tasks. We evaluate our model on two standard datasets for named entity recognition (NER) and chunking, and in both cases achieve state of the art results, surpassing previous systems that use other forms of transfer or joint learning with additional labeled data and task specific gazetteers.

**摘要**

从未标记文本中学习的预训练词向量已成为NLP任务的神经网络体系结构的标准组成部分。但是在大多数情况下，使用字级表示来生成对上下文敏感的表示的循环网络是在相对较少的标记数据上进行训练的。在本文中，我们演示了一种通用的半监督方法，用于将预训练的上下文嵌入从双向语言模型添加到NLP系统，并将其应用于序列标记任务中。我们在两个标准数据集上评估我们的模型，用于命名实体识别（NER）和分类，并且在两种情况下都实现了最先进的结果，超过了以前使用其他形式的迁移学习或联合学习的添加额外标记数据和特定任务的系统。