**Deep contextualized word representations**

**Matthew E. Peters†, Mark Neumann†, Mohit Iyyer†, Matt Gardner†,**

{matthewp,markn,mohiti,mattg}@allenai.org

**Christopher Clark∗, Kenton Lee∗, Luke Zettlemoyer†∗**

{csquared,kentonl,lsz}@cs.washington.edu

†Allen Institute for Artificial Intelligence ∗Paul G. Allen School of Computer Science & Engineering, University of Washington

**Abstract**

We introduce a new type of deep contextualized word representation that models both (1) complex characteristics of word use (e.g., syntax and semantics), and (2) how these uses vary across linguistic contexts (i.e., to model polysemy). Our word vectors are learned functions of the internal states of a deep bidirectional language model (biLM), which is pre- trained on a large text corpus. We show that these representations can be easily added to existing models and significantly improve the state of the art across six challenging NLP problems, including question answering, textual entailment and sentiment analysis. We also present an analysis showing that exposing the deep internals of the pre-trained network is crucial, allowing downstream models to mix different types of semi-supervision signals.

**摘要**

本文介绍了一种新的深层情境化词表示模型包括(1) 复杂的词语用法特征 (如语法和语义), 以及 (2) 这些用法在不同的语言语境 (即模型多义性) 之间的差异。本文的词向量在一个大的文本语料库上进行预先训练，是一种学习深层双向语言模型 (biLM)的内部状态的函数。本文表明, 这些词表示可以很容易地添加到现有模型中, 并显著改善六个具有挑战性的 NLP 问题的目前为止最高状态, 包括问答、文本蕴涵和情感分析。本文还分析了暴露预先训练的深层网络内部的重要性, 以及下游模型混合不同类型的半监督方法的必要性。