# Learned in Translation: Contextualized Word Vectors

**Bryan McCann**

bmccann@salesforce.com

**Caiming Xiong**

[cxiong@salesforce.com](mailto:cxiong@salesforce.com)

**James Bradbury**

james.bradbury@salesforce.com

**Richard Socher**

rsocher@salesforce.com

**Abstract**

Computer vision has benefited from initializing multiple deep layers with weights pretrained on large supervised training sets like ImageNet. Natural language processing (NLP) typically sees initialization of only the lowest layer of deep models with pretrained word vectors. In this paper, we use a deep LSTM encoder from an attentional sequence-to-sequence model trained for machine translation (MT) to contextualize word vectors. We show that adding these context vectors (CoVe) improves performance over using only unsupervised word and character vectors on a wide variety of common NLP tasks: sentiment analysis (SST, IMDb), question classification (TREC), entailment (SNLI), and question answering (SQuAD). For fine-grained sentiment analysis and entailment, CoVe improves performance of our baseline models to the state of the art.

**摘要**

计算机视觉已经从预先训练权重来初始化多个深层图中受益，比如在ImageNet这样的大型有监督训练集上。自然语言处理（NLP）通常仅使用预训练的单词向量来初始化深度模型中的最底层。在本文中，我们使用深层LSTM编码器，该编码器来自针对机器翻译（MT）训练的注意序列到序列模型，以对词向量进行语境化。我们表明，添加这些上下文向量（CoVe）可以对性能有所提高，而不仅仅只使用无监督的单词和特征向量来处理各种常见的NLP任务：情感分析（SST，IMDb），问题分类（TREC），文本蕴涵（SNLI）和问答（SQuAD）。对于细粒度情感分析和文本蕴涵，CoVe可将我们的基线模型的性能提升到最先进的技术水平。