# An efficient framework for learning sentence representations

**Lajanugen Logeswaran∗ & Honglak Lee†∗**

∗University of Michigan, Ann Arbor, MI, USA  †Google Brain, Mountain View, CA, USA {llajan,honglak}@umich.edu,honglak@google.com

**Abstract**

In this work we propose a simple and efficient framework for learning sentence representations from unlabelled data. Drawing inspiration from the distributional hypothesis and recent work on learning sentence representations, we reformulate the problem of predicting the context in which a sentence appears as a classifica- tion problem. Given a sentence and its context, a classifier distinguishes context sentences from other contrastive sentences based on their vector representations. This allows us to efficiently learn different types of encoding functions, and we show that the model learns high-quality sentence representations. We demonstrate that our sentence representations outperform state-of-the-art unsupervised and supervised representation learning methods on several downstream NLP tasks that involve understanding sentence semantics while achieving an order of magnitude speedup in training time.

**摘要**

在本文的工作中，我们提出了一个简单而有效的框架，其用于从未标记的数据中学习句子表示。从分布式假设和最近的学习句子表示的工作中汲取到灵感，我们将预测句子出现的背景问题重新设计为分类问题。给定句子及其上下文，分类器基于其向量表示将上下文句子与其他对比句子区分开。这使我们能够有效地学习不同类型的编码函数，并且我们证明该模型能够学习到高质量的句子表示。我们证明我们的句子表示在几个下游NLP任务上优于最先进的无监督和监督表示学习方法，这些任务涉及理解句子语义，同时在训练时间内速度可达到一个数量级的提升。