**BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**

**Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova**

Google AI Language {jacobdevlin,mingweichang,kentonl,kristout}@google.com

**Abstract**

We introduce a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models (Peters et al., 2018; Radford et al., 2018), BERT is designed to pre-train deep bidirectional representations by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT representations can be fine-tuned with just one additional output layer to create state-of-the- art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications.

BERT is conceptually simple and empirically powerful. It obtains new state-of-the-art results on eleven natural language processing tasks, including pushing the GLUE bench- mark to 80.4% (7.6% absolute improvement), MultiNLI accuracy to 86.7% (5.6% absolute improvement) and the SQuAD v1.1 question answering Test F1 to 93.2 (1.5 absolute improvement), outperforming human performance by 2.0.

**摘要**

我们引入了一种名为BERT的新的语言表征模型，一个Transformer双向编码器表示。与最近的语言表示模型（Peters et al。，2018; Radford et al。，2018）不同，BERT旨在通过联合调节所有层中的左右上下文来预训练深度双向语言表示。因此，预训练的BERT表示可以通过一个额外的输出层进行微调，以创建适用于广泛任务的最先进模型，例如问答和语言推理任务，而无需具体的架构修改。

BERT概念简单且经验丰富。它在11项自然语言处理任务中获得了最新成果，包括将GLUE基准推至80.4％（绝对改进7.6％），MultiNLI准确率达到86.7％（绝对改进5.6％）和SQuAD v1 .1问题回答测试F1到93.2（1.5绝对改进），超过人类表现2.0。