# KDABert: Knowledge Distillation with Generative Adversarial Networks for Bert Model Compression

### **闫森** 2017111497

#### 摘要

这里是摘要。这里是摘要。这里是摘要。这 里是摘要。这里是摘要。这里是摘要。这里 是摘要。这里是摘要。这里是摘要。这里是 摘要。这里是摘要。这里是摘要。这里是摘 要。这里是摘要。这里是摘要。这里是摘

#### 1 介绍

这里是介绍。这里是介绍。这里是介绍。这 里是介绍。这里是介绍。这里是介绍。这里是 介绍。这里是介绍。这里是介绍。这里是介绍。 这里是介绍。这里是介绍。这里是介绍。 这里是介绍。这里是介绍。这里是介绍。 这里

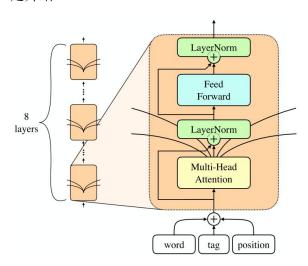


图 1: Transformer 结构。

#### 2 实验

这里是实验。这里是实验。这里是实验。这 里是实验。这里是实验。这里是实验。这里是 实验。这里是实验。这里是实验。这里是实验。 这里是实验。

模型	LR	LP	F1
单模型			
?	-	-	88.3
?	-	-	89.8
?	90.2	90.7	90.4
?	90.5	92.1	91.3
?	91.3	92.1	91.7
?	90.3	93.2	91.8
?	-	-	91.8
?	92.0	91.7	91.8
?	91.5	92.5	92.0
?	92.2	92.5	92.4
?	92.5	92.5	92.5
?	93.2	93.9	93.6
非单模型			
?	-	-	93.8
?	-	-	94.2
?	-	-	94.7
?	94.9	95.4	95.1
?	95.6	96.0	95.8

表 1: 不同模型在 PTB 测试集上的最终结果。

## 致谢