

BERT模型中

BERT模型中，每个编码器由两层组成，即自注意力层和全连接层。