

Part  
03

# 巨人来袭

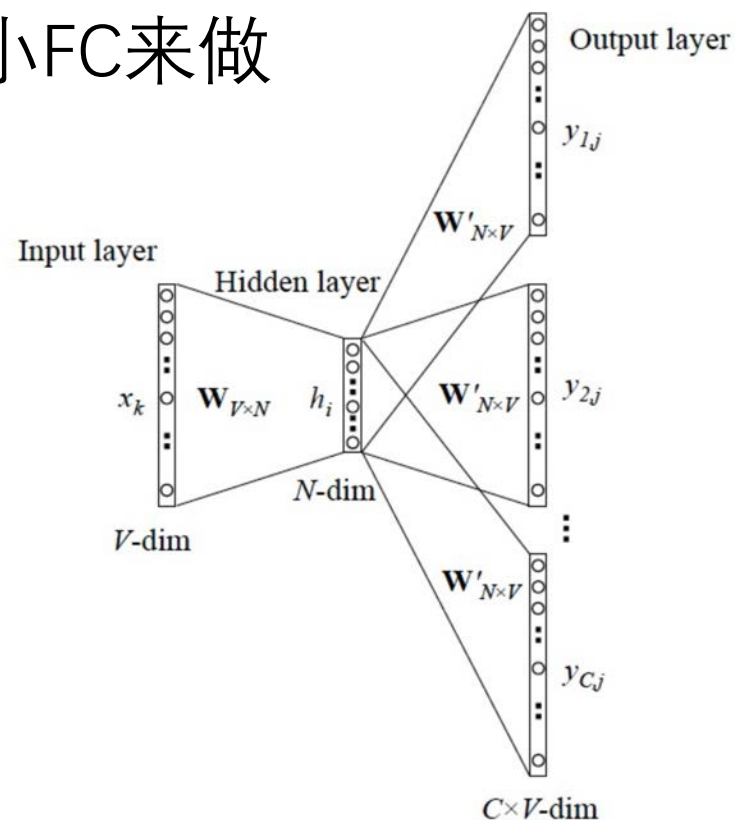
- BERT
- 百变魔君



## 3.1

## Attention到底给了我们什么？

- CNN、RNN：一个兼容并包、横扫天下的合体
- **Full Connected**：一个可以被训练出来的我
- 如果用很深、很大的**Transformer Encoder**替代小FC来做 Embedding？
- 更多信息（位置、变长度）可以被挖掘
- 那就来吧，先来340M参数



## 3.1

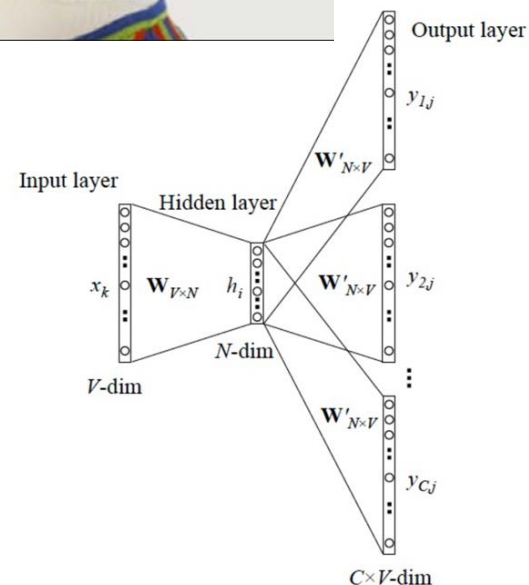
# Bidirectional Encoder Representations from Transformers

- Self-supervised Learning 自监督学习
- 网络结构是Transformer的encoder
- 训练的标签是原始数据中的一部分
- 我们一切的历史、一切的过去都可以是学习资料

My name is Li Hua . I'm thirteen. I'm a middle school student . I am 1 Class Five, Grade One. My English teacher 2 Mr Lin. He is 3 old teacher. I 4 a pen, a ruler and two 5 in my pencil-box. I have a bike 6.

Liu Ping is in my class. She is a girl. We 7 good friends . She is not 8 today. I think she is at home. Look ! Here is a 9 . But it is not my book, I think it is 10 book.

- |                    |              |               |               |
|--------------------|--------------|---------------|---------------|
| ( ) 1.A. in        | B. at        | C. do         | D. not        |
| ( ) 2.A. are       | B. am        | C. is         | D. ×          |
| ( ) 3.A. a         | B. an        | C. this       | D. very       |
| ( ) 4.A. am        | B. think     | C. know       | D. have       |
| ( ) 5.A. boxes     | B. pencils   | C. buses      | D. desk       |
| ( ) 6.A. too       | B. or        | C. much       | D. very       |
| ( ) 7.A. have      | B. am        | C. are        | D. all        |
| ( ) 8.A. right     | B. where     | C. at home    | D. at school  |
| ( ) 9.A. licence   | B. book      | C. picture    | D. map        |
| ( ) 10.A. Liu Ping | B. Liu Pings | C. Liu Ping's | D. Liu Pings' |



## 3.1

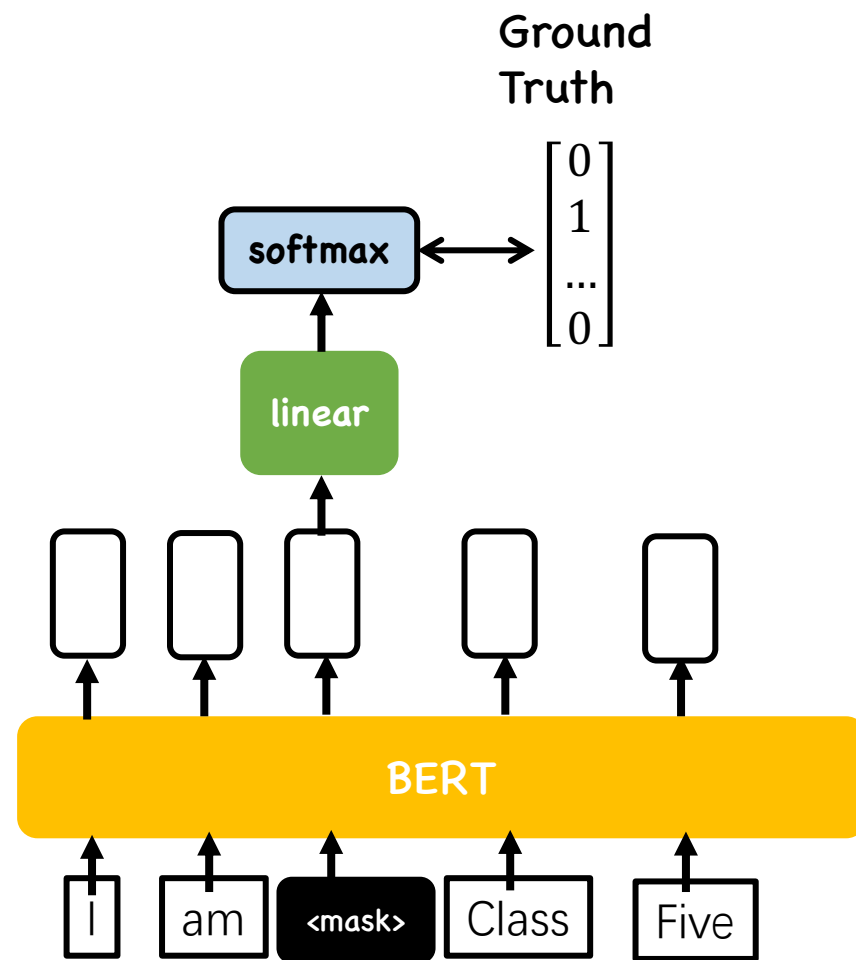
## Self-supervised Learning 自监督学习

My name is Li Hua. I'm thirteen. I'm a middle school student. I am 1 Class Five, Grade One. My English teacher 2 Mr Lin. He is 3 old teacher. I 4 a pen, a ruler and two 5 in my pencil-box. I have a bike 6.

Liu Ping is in my class. She is a girl. We 7 good friends. She is not 8 today. I think she is at home. Look! Here is a 9. But it is not my book, I think it is 10 book.

- |                     |              |               |               |
|---------------------|--------------|---------------|---------------|
| ( ) 1. A. in        | B. at        | C. do         | D. not        |
| ( ) 2. A. are       | B. am        | C. is         | D. ×          |
| ( ) 3. A. a         | B. an        | C. this       | D. very       |
| ( ) 4. A. am        | B. think     | C. know       | D. have       |
| ( ) 5. A. boxes     | B. pencils   | C. buses      | D. desk       |
| ( ) 6. A. too       | B. or        | C. much       | D. very       |
| ( ) 7. A. have      | B. am        | C. are        | D. all        |
| ( ) 8. A. right     | B. where     | C. at home    | D. at school  |
| ( ) 9. A. licence   | B. book      | C. picture    | D. map        |
| ( ) 10. A. Liu Ping | B. Liu Pings | C. Liu Ping's | D. Liu Pings' |

- 特殊输入 <mask>
- 原始编码是one-hot



- BERT训练完毕后是一个训练好的网络
- 在此基础上进行微调训练以适应多样化任务
- Word2vec后的新起手式
- 需要训练**BERT**和**Linear**的参数
- **BERT**是一个**预训练**很大模型（开源的）
- **Linear**是随机初始化的

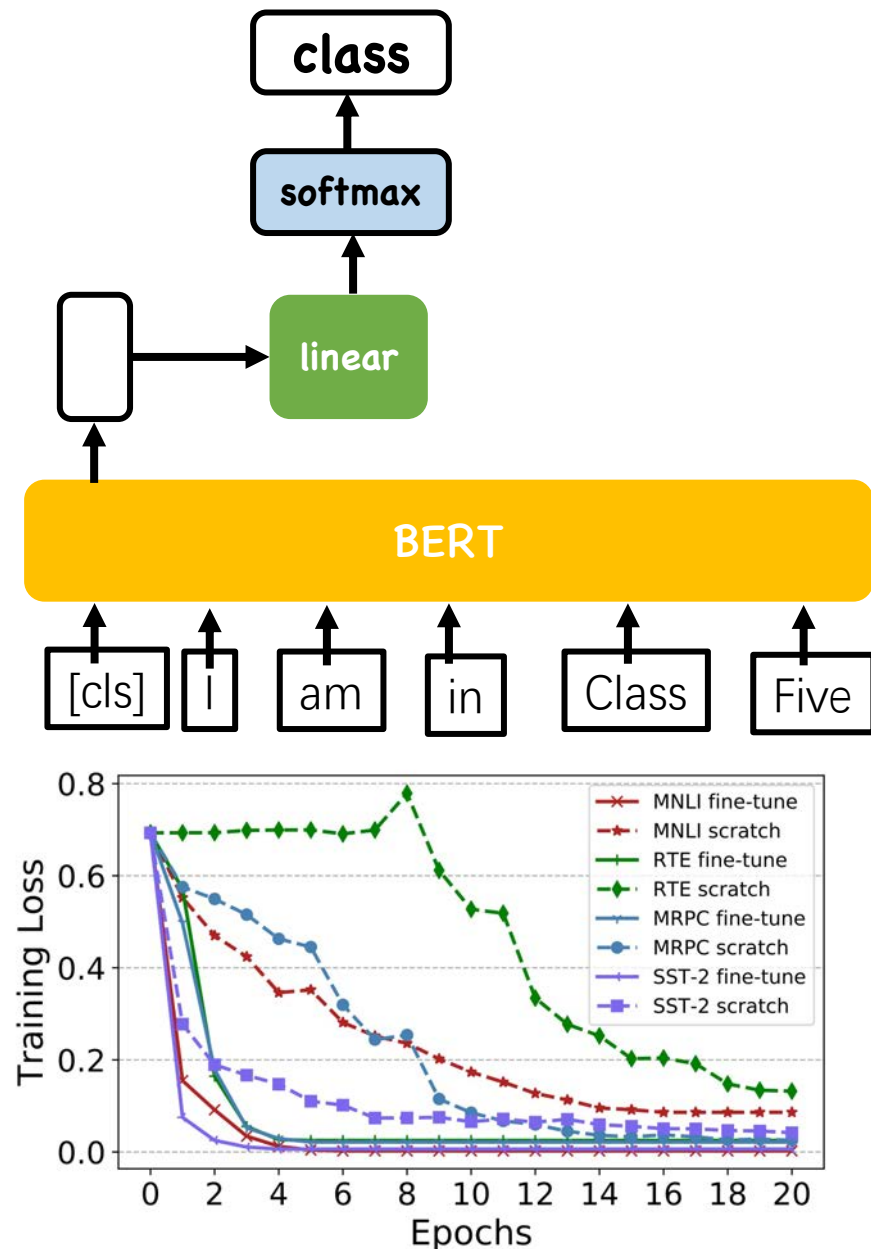


Figure 2: Training loss of fine-tuning BERT and training from scratch on four datasets.

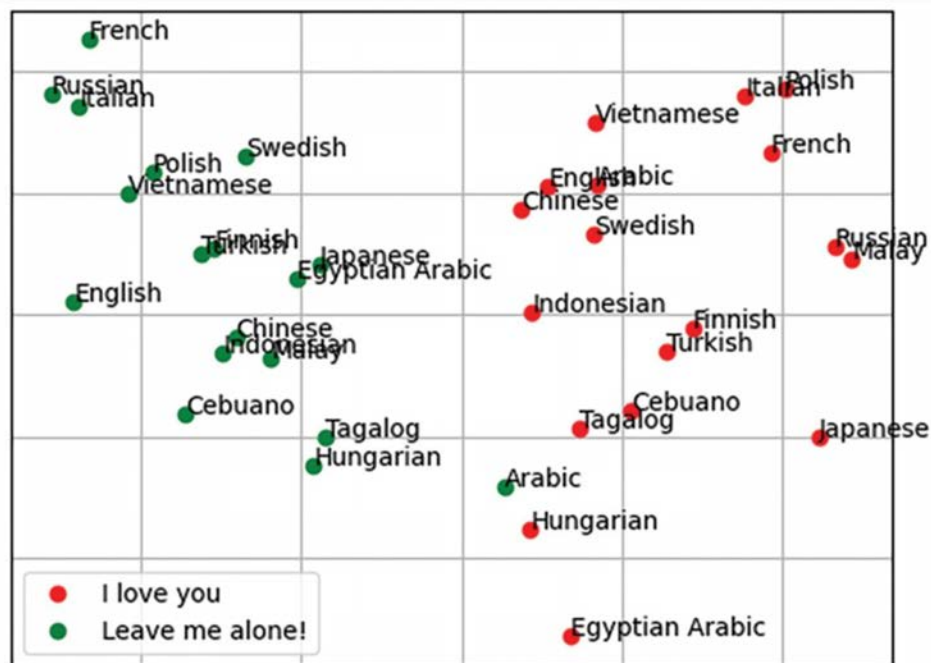
<https://arxiv.org/abs/1908.05620>



## 3.3

## BERT 的能力强在哪里

- 相较于Word2Vec
  - 序列长度长的多: 512 vs 10
  - 位置信息&上下文关系
  - 网络要深, 参数多得多
  - W2v 后面需要更深网络
- 包含**上下文**的信息
- Embedding的含义



How translations of the phrases "I love you" and "Leave me alone!" in several languages are represented by m-BERT when projected on a scatter plot.

海水朝朝朝朝朝朝朝落  
浮云长长长长长长长消

