

# 《The road to AGI》

YKY 甄景贤

December 14, 2019

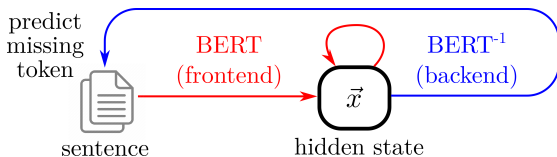
# Table of contents

- 1 BERT 的革命性意义
- 2 从 BERT 过渡到 AGI
- 3 「集」结构 带来麻烦
- 4 “Attention is all you need” 是什么意思？

多谢 支持 😊

# BERT 的革命性意义

- BERT 利用平常的文本 induce 出知识，而这 representation 具有 通用性 (universality) :



(1)

换句话说：隐状态的 representation 压缩了句子的意思，而它可以应用在别的场景下

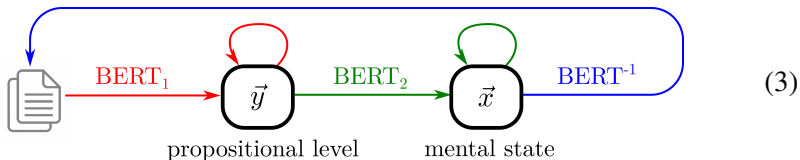
- This implies that human-level AI can be *induced* from existing corpora, 而不需要重复像人类婴儿成长的学习阶段
- Such corpora can include items such as images, movies with dialogues / subtitles
- 这种训练方法是较早的另一篇论文提出，它并不属于 BERT 的内部结构

# 从 BERT 过渡到 AGI

- 词语 组成 句子，类比於 逻辑中，概念 组成 逻辑命题
- 抽象地说，逻辑语言 可以看成是一种有 2 个运算的 代数结构，可以看成是 加法  $\wedge$  和 乘法  $\cdot$ ，其中 乘法 是不可交换的，但加法 可交换
- 例如 两个命题：

我  $\cdot$  爱  $\cdot$  妳  $\wedge$  妳  $\cdot$  爱  $\cdot$  我 (2)

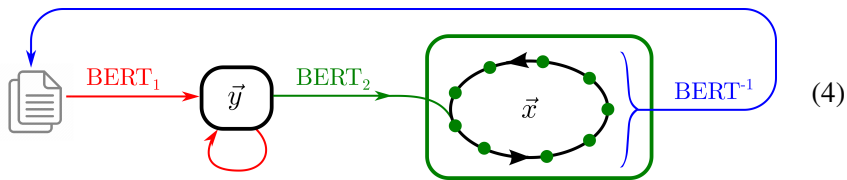
- 这种逻辑结构 可以用 两层 的 BERT 模型 处理：



- 我发现 最难处理的问题，是在第 2 层 的状态  $\vec{x}$ . 它是一个 逻辑命题 的 集合，集合中元素有可交换性，亦即 **permutation invariance**. 这看似简单的特性，其实带来很大的麻烦

# 「集」结构带来麻烦

- Word2Vec 也是革命性的；由 Word2Vec 演变成 Sentence2Vec 则比较容易，基本上只是向量的延长 (**concatenation**)；逻辑命题 类似於 sentence
- 假设 全体逻辑命题的空间是  $\mathbb{P}$ ，则 **命题集合** 的空间是  $2^{\mathbb{P}}$ ，后者非常庞大
- 如果限制 状态  $\vec{x}$  = working memory 只有 10 个命题，这空间是  $\mathbb{P}^{10}/\sim$  其中  $\sim$  是交换群的等价关系。这空间虽然是  $\mathbb{P}^{10}$  的子空间，但  $\sim$  结构很难用神经网络实现
- 现时 比较可行的办法，是将 状态  $\vec{x}$  实现成一个时间上的「轮盘」，每个 ● 表示一个命题：



- 有趣的是，如果用「轮盘」方法，BERT 的 **注意力机制** 有特殊意义....

# “Attention is all you need” 是什么意思？

- 注意力机制 最初起源於 Seq2seq architecture, 后来 BERT 引入 self-attention
- 在 Seq2seq 中, 编码器 (encoder) 由下式给出, 它将输入的词语  $x_i$  转化成一连串的 隐状态  $h_i$ :

$$h_t = f(x_t, h_{t-1}) \quad (5)$$

多谢收看 😊