The road to AGI

YKY

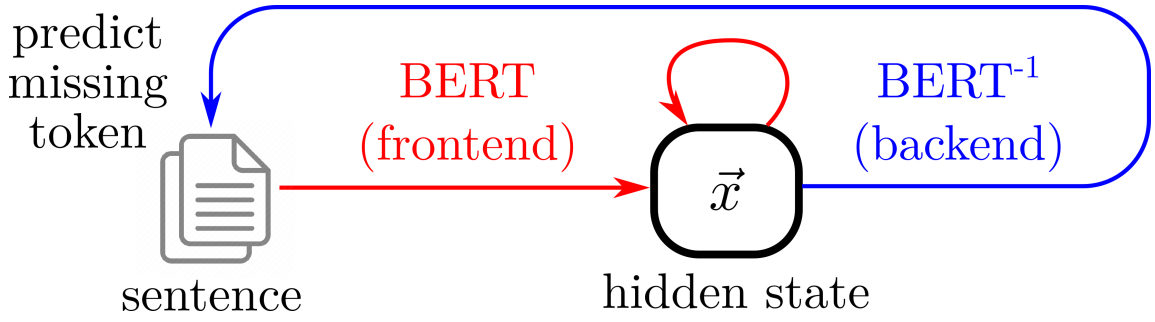
December 18, 2019

Table of contents

☺️

BERT

- BERT induce representation (universality) :



(1)

representation

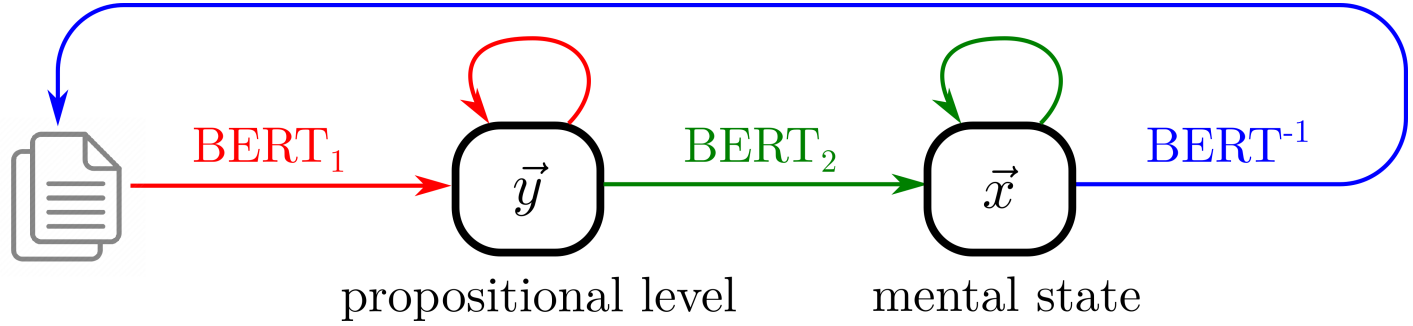
- This implies that human-level AI can be *induced* from existing corpora,
- Such corpora can include items such as images, movies with dialogues / subtitles
- BERT

BERT AGI

-
- 2 ∧ ·
-

· · ∧ · · (2)

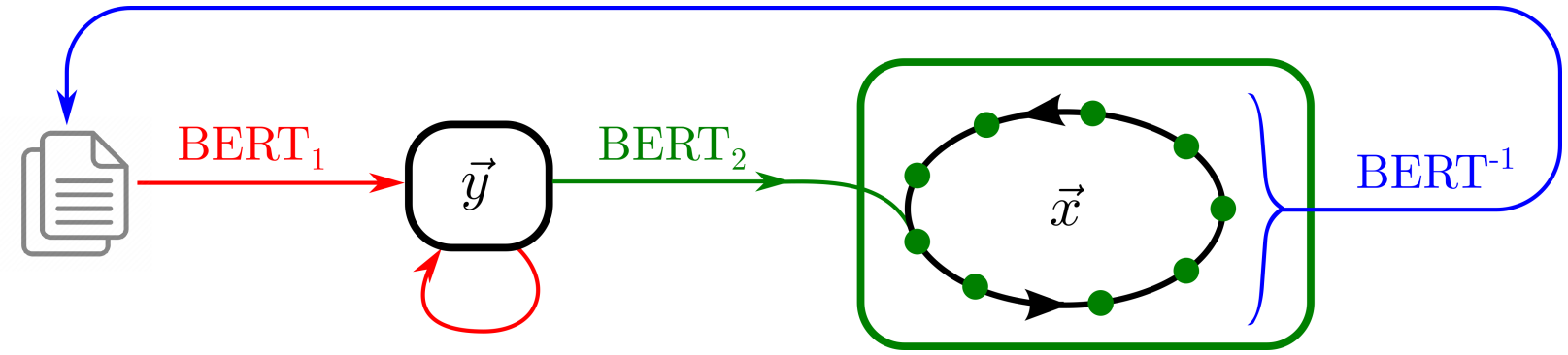
- BERT



(3)

- 2 $\vec{x}$. permutation invariance.

- Word2Vec Word2Vec Sentence2Vec (concatenation) sentence
- $\mathbb{P}$ $2^{\mathbb{P}}$
- $\vec{x}$ = working memory 10 $\vec{x}$ $\mathbb{P}^{10}/\sim$ $\sim$ $\mathfrak{S}_{10}$ $2^{\mathbb{P}} \cong \prod_{n=0}^{\infty} \mathbb{P}^n/\mathfrak{S}_n$
- $\mathbb{P}^n/\mathfrak{S}_n$ $\mathbb{P}^n$ $\mathfrak{S}_n$-
- $\vec{x}$ ●



(4)

- BERT ....

Attention

- Seq2seq architecture BERT self-attention
- Seq2seq (encoder) x_i h_i

$$h_t = \text{RNN}_{encode}(x_t, h_{t-1}) \tag{5}$$

- h_i $c = q(h_1, \ldots, h_n)$. c
- s_t y_t

$$s_t = \text{RNN}_{decode}(y_t, s_{t-1}, c_t) \tag{6}$$

- c_t h_j

$$c_t = \sum_j \alpha_{ij} h_j \tag{7}$$

- α_{ij}

$$\alpha_{ij} = \text{softmax}\{\langle s_i, h_j \rangle\} \tag{8}$$

Attention

- attention
- s_t
- Attention $\geq 1 \sum h_j$ information retrieval
- M N N^M mappings attention mappings
- “attention is all you need” inductive bias mapping
- mapping f : $\rightarrow$
- logic rules (applicable) rules
- s_t “search state” “” s_t
-

Thanks for watching ☺️