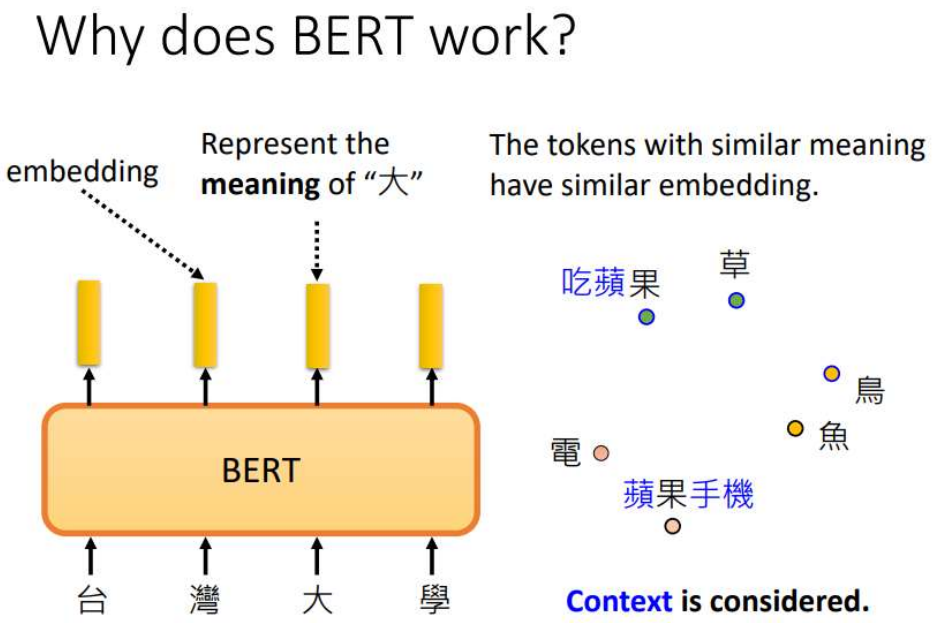
自督導式學習 (Self-supervised Learning) (下)

Create at 2022/06/26

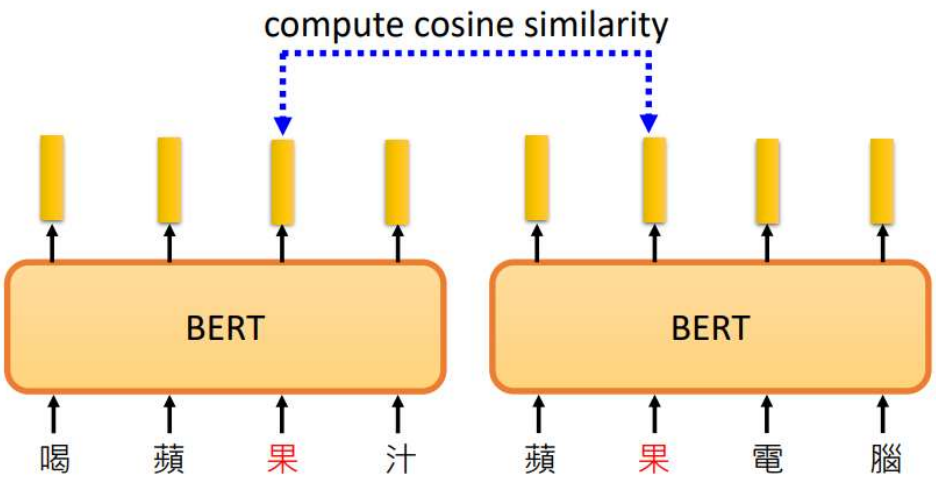
- 上課資源：
  1. 自督導式學習 (Self-supervised Learning) (三) – BERT的奇聞軼事 (https://www.youtube.com/watch?v=ExXA05i8DEQ)
  2. 自督導式學習 (Self-supervised Learning) (四) – GPT的野望 (https://www.youtube.com/watch?v=WY_E0Sd4K80)

BERT的奇聞軼事

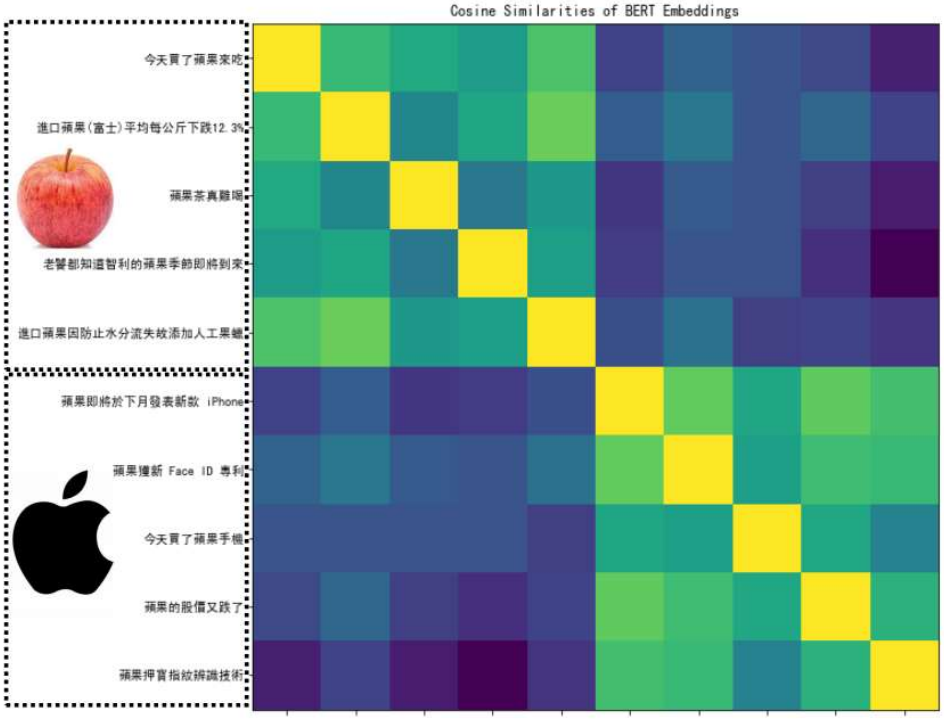


- 為甚麼 BERT 有用呢？
  - 輸入一段文字之後產生一段向量稱它為 embedding
  - 這些向量代表輸入的字的意思
  - 意思越相近，向量就越像

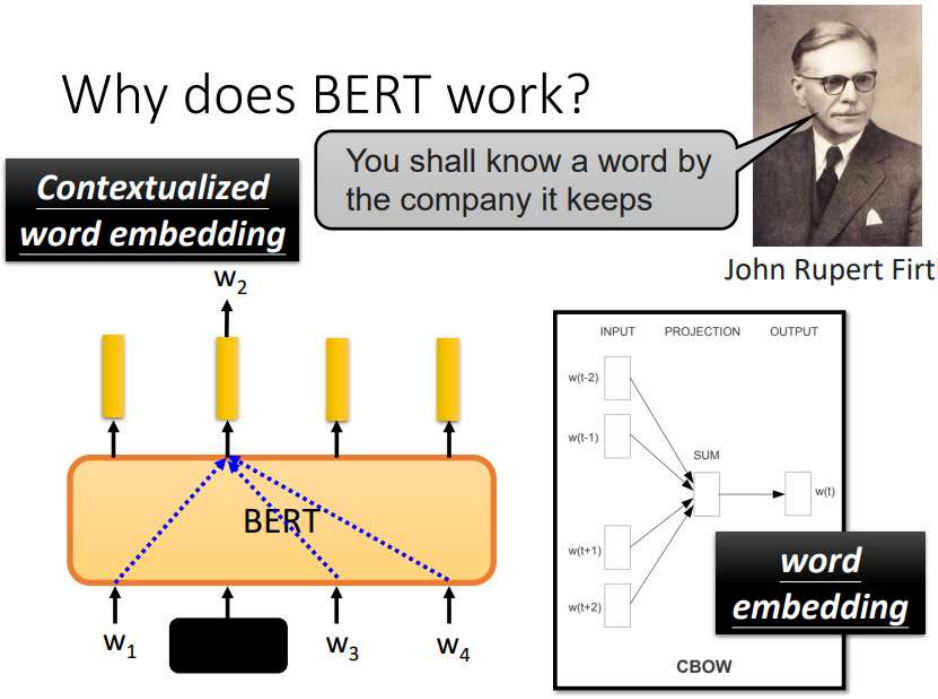
Why does BERT work?



- 把兩個句子都丟到 BERT 裡面，接著分別計算兩個 “果” 之間的 cosine similarity



- 這邊有 10 個 “果”，兩兩之間計算相似度
- 前五個 “果” 相似度比較高，後五個 “果” 相似度比較高

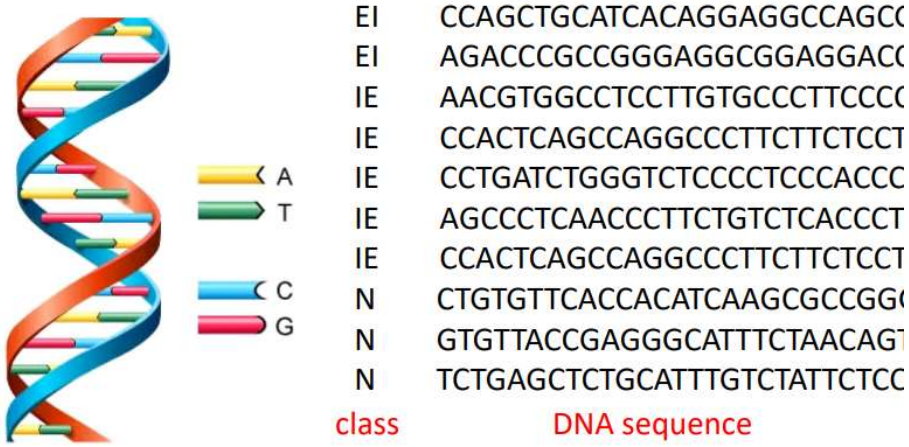


- BERT 輸出的每一個向量，就代表了那個字的意思
- BERT 在學填空題的過程中，學會了每個中文字的意思，所以可以在接下來的任務做的更好
- 要知道一個詞彙的意思取決於它的上下文

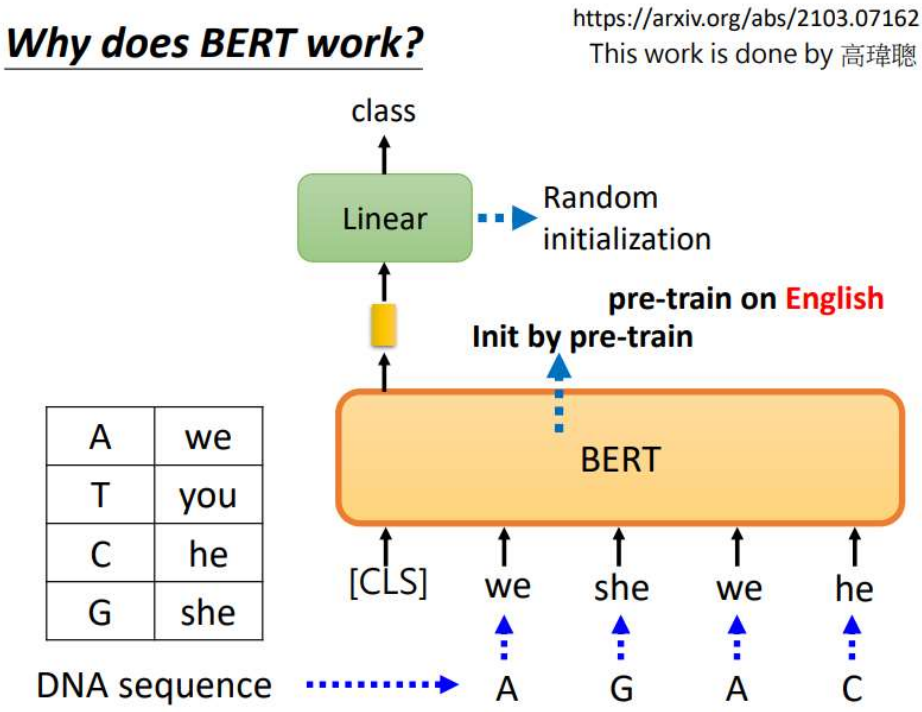
https://arxiv.org/abs/2103.07162
This work is done by 高瑋聰

Why does BERT work?

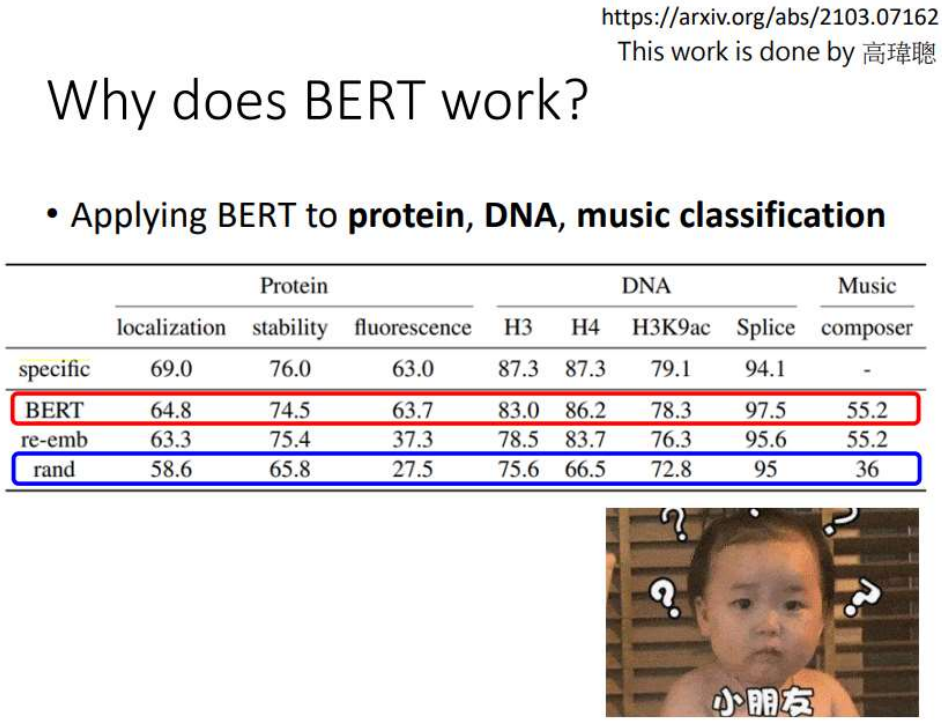
- Applying BERT to **protein, DNA, music classification**



- 把訓練在文字上的 BERT，拿來做蛋白質的分類、DNA 的分類、音樂的分類
- 以 DNA 的分類為例

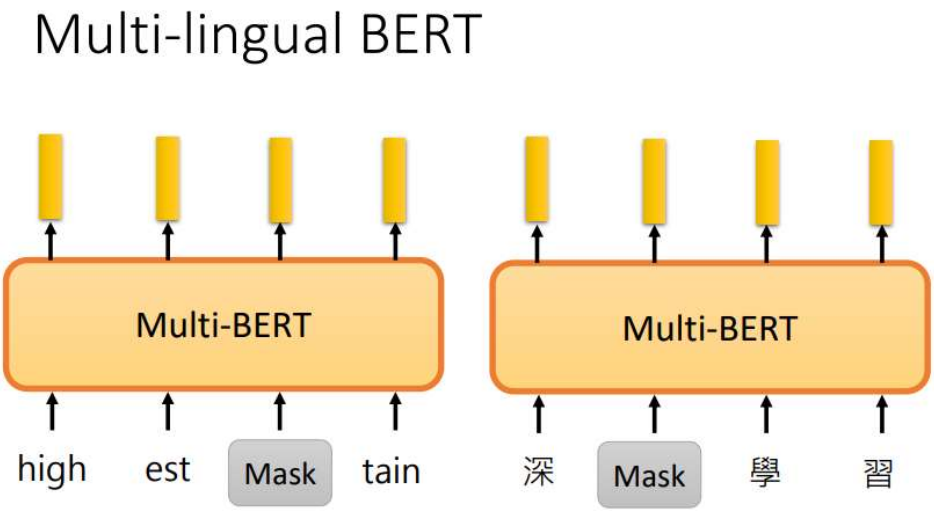


- DNA 用 A、T、C、G 來表示
- 現在要把 BERT 用在 DNA 的分類上
- 把 A、T、C、G 分別對應到隨便一個英文的詞彙

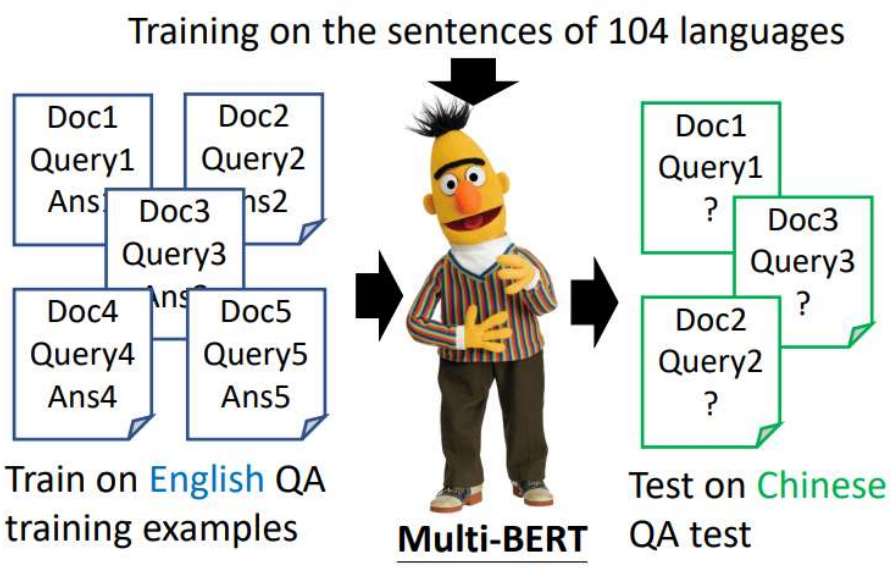


- BERT 的能力不完全來自於它看得懂文章

Multi-lingual BERT



- 用不同的語言給他做填空



- 拿英文的 QA 資料做訓練，它就會自動學會做中文的 QA

Zero-shot Reading Comprehension

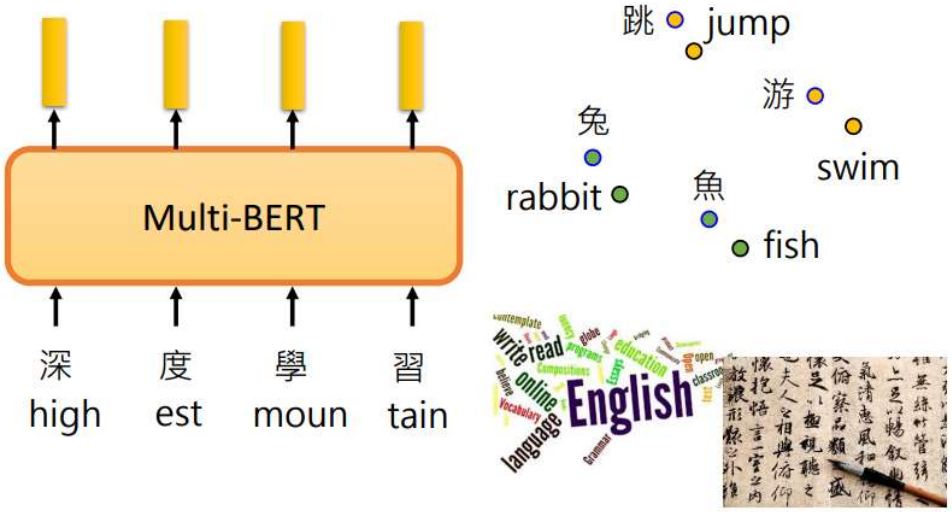
- English: SQuAD, Chinese: DRCD

Model	Pre-train	Fine-tune	Test	EM	F1
QANet	none	Chinese	Chinese	66.1	78.1
BERT	Chinese	Chinese		82.0	89.1
	104 languages	Chinese		81.2	88.7
		English		63.3	78.8
		Chinese + English		82.6	90.1

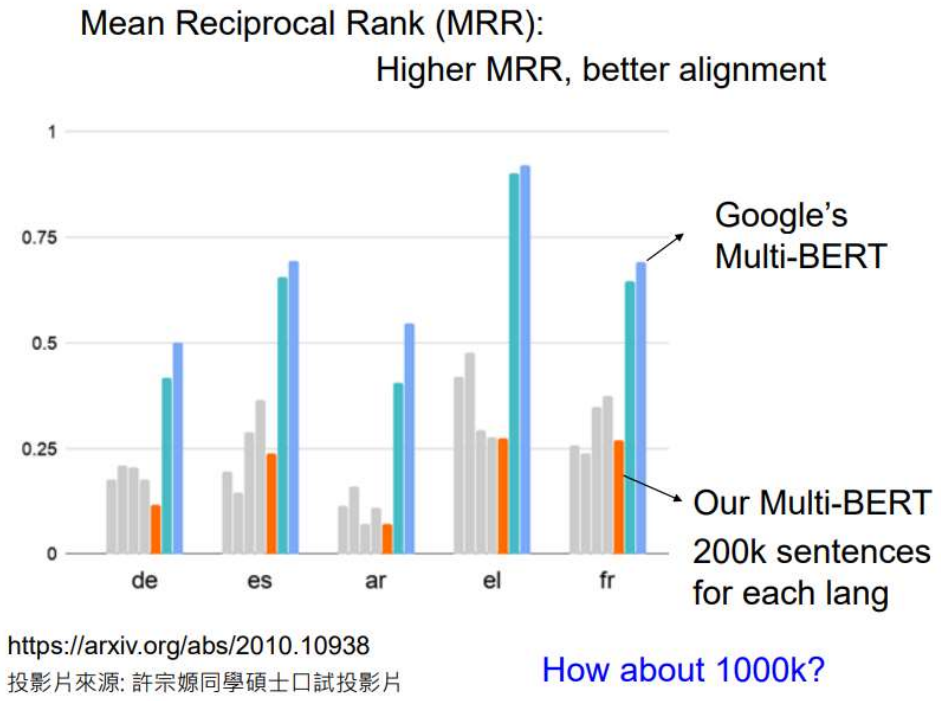
F1 score of Human performance is 93.30%

This work is done by 劉記良、許宗嫄
https://arxiv.org/abs/1909.09587

Cross-lingual Alignment?

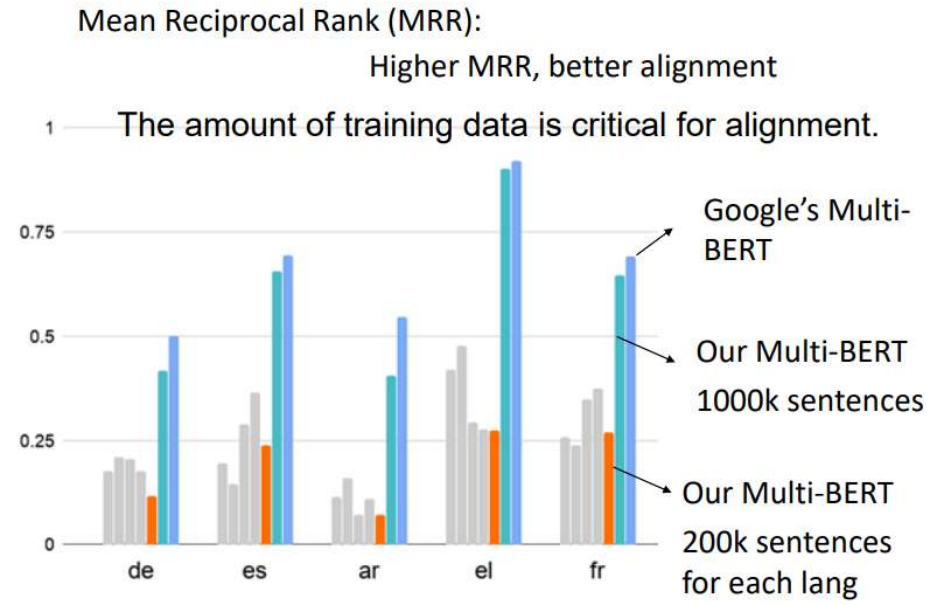


- 也許對 Multi-BERT 而言，不同語言之間沒有什麼差異



- MRR 的值越高，代表兩個不同語言它們的 align 越好 (向量越接近)
- 藍色的線是 google 試出來的 104 種語言的 Multi-BERT 得到的 MRR
- 橘色的是我們自己訓練的 Multi-BERT 使用 200k sentences

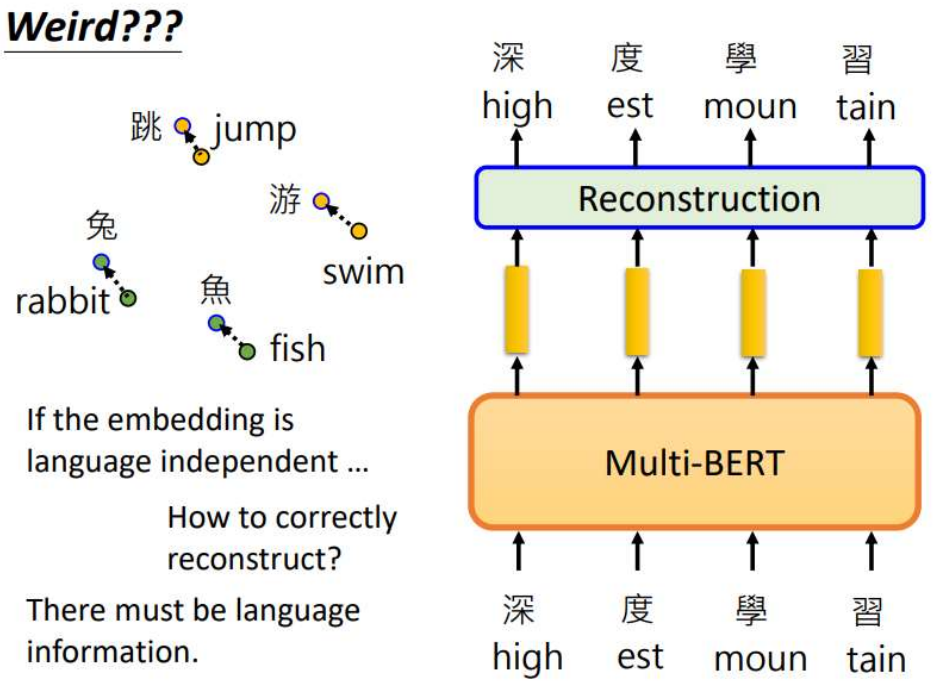
- 資料變成 100k sentences



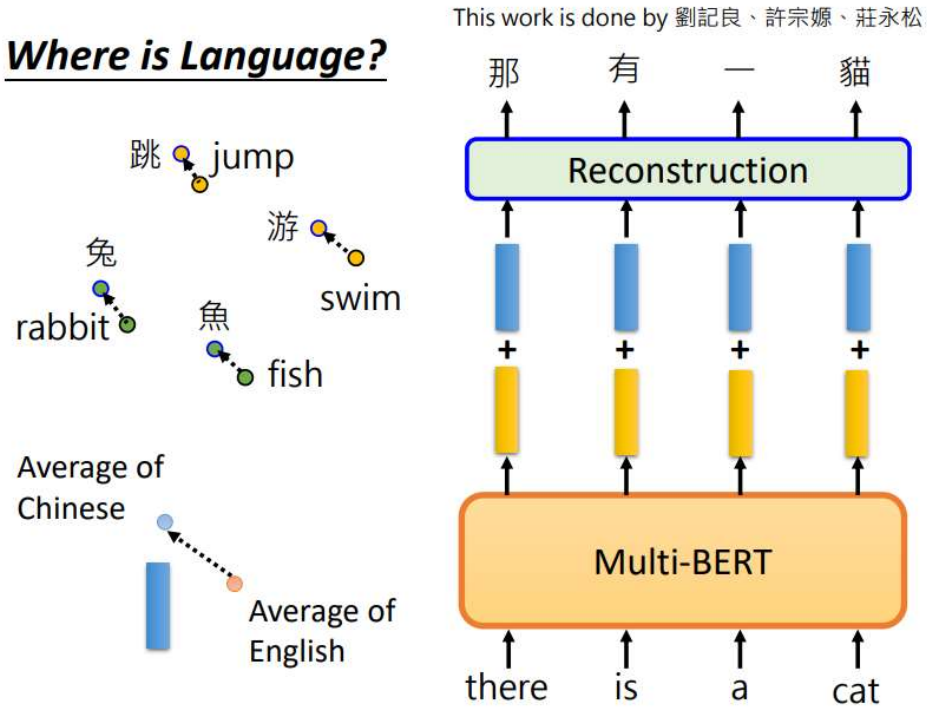
https://arxiv.org/abs/2010.10938
投影片來源: 許宗嫄同學碩士口試投影片

- 綠色的線是資料量改成 1000k 的 MRR

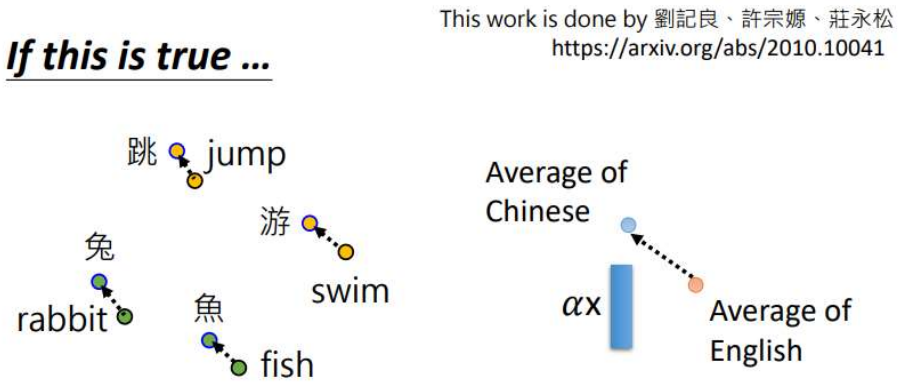
https://arxiv.org/abs/2010.10041



- 它知道語言的資訊



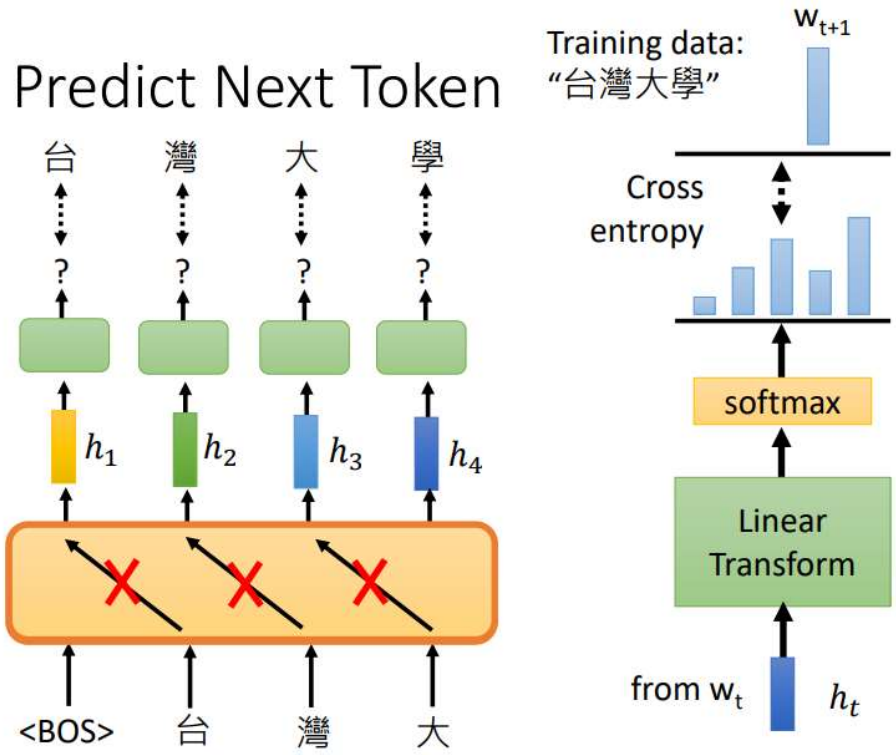
- 把所有英文的詞彙都丟到 Multi-BERT，把英文的 embedded 計算平均，再計算所有中文的 embedded 平均
- 兩者相減就是中文跟英文之間的差距 (藍色向量)
- 給 Multi-BERT 一句英文得到 embedded，把這個 embedded 加上藍色的向量，最後的向量對 Multi-BERT 來說就變成了中文的句子



Input (en)	The girl that can help me is all the way across town. There is no one who can help me.
Ground Truth (zh)	能帮助我的女孩在小镇的另一边。没有人能帮助我。。
en→zh, $\alpha = 1$	. 孩，can 来我是all the way across 市。。There 是无人人can help 我。
en→zh, $\alpha = 2$	. 孩的的家我是这个人的市。。他是他人人的到我。
en→zh, $\alpha = 3$	。，的的的他是的个的的，。：他是他人，的。他。

Unsupervised token-level translation ☺

GPT 的野望

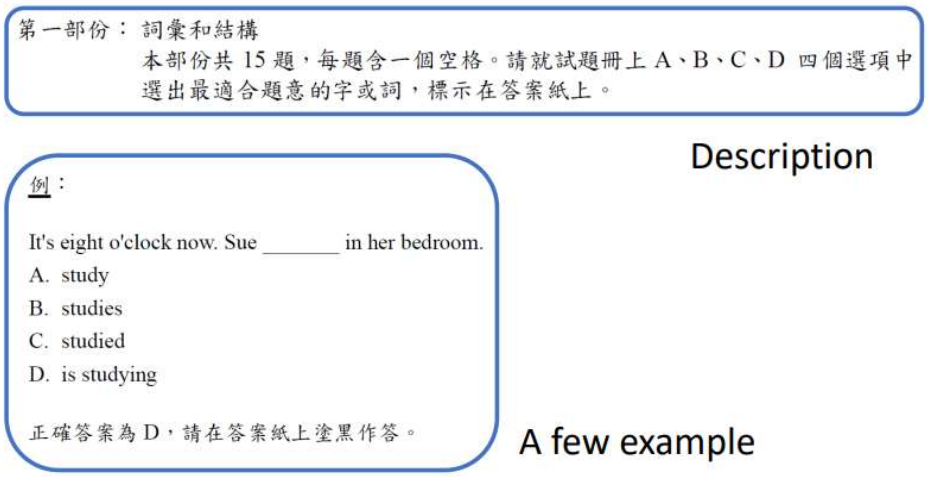


- GPT 就是改一下我們在做 self-supervised learning 的時候要模型做的任務
- GPT 要做的任務是預測接下來會出現的 token 是什麼
- GPT 模型像是 transformer 的 decoder

https://talktotransformer.com/

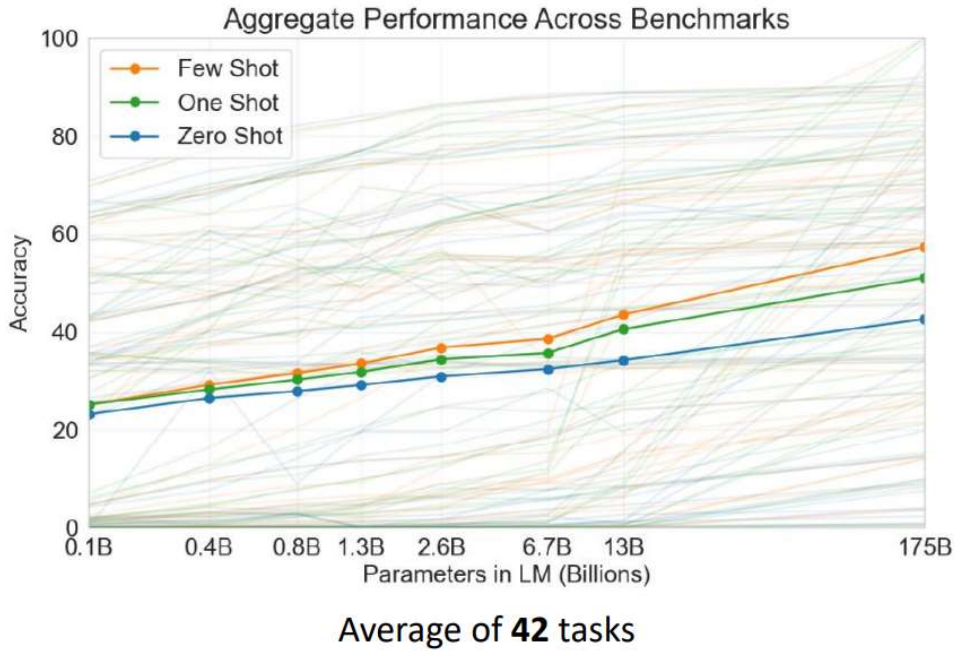
- https://app.inferkit.com/demo (https://app.inferkit.com/demo)

How to use GPT?

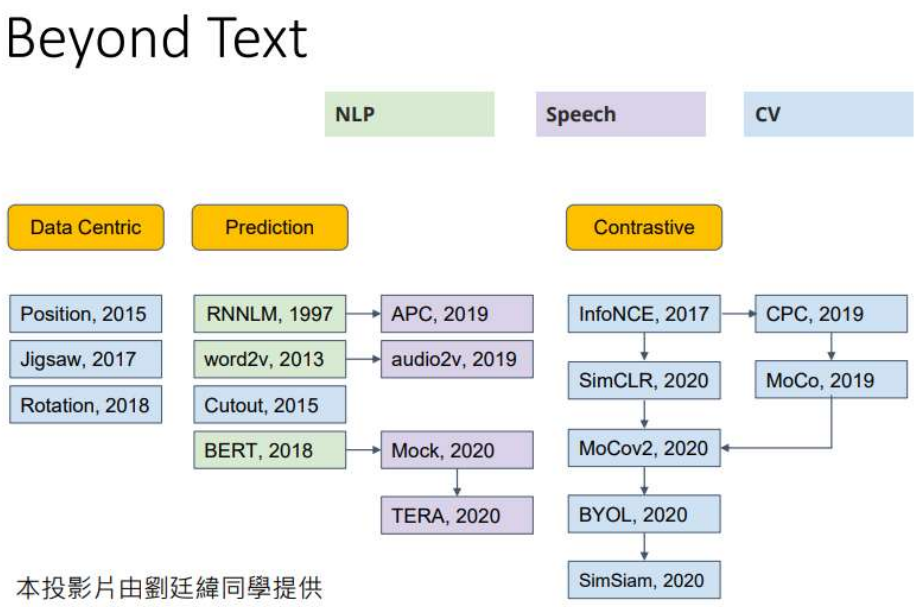


- GPT 是如何運作？
  - 先給問題一個描述
  - 再給一個範例

- Few-show Learning : 給一點例子
- In-context Learning (不是一般的 learning，它連 gradient descent 都沒有做)
- One-shot Learning : 給一個例子
- Zero-shot Learning : 完全不給例子



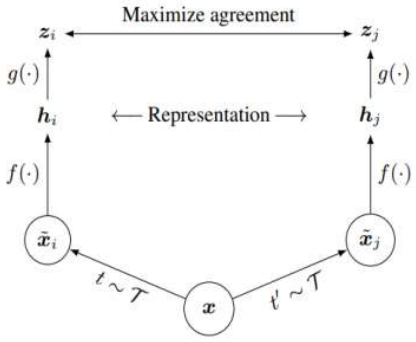
- 三條線是 42 個任務的平均正確率



- 在語音、影像都可以用 self-supervised learning 的技術

Image - SimCLR

https://arxiv.org/abs/2002.05709
https://github.com/google-research/simclr

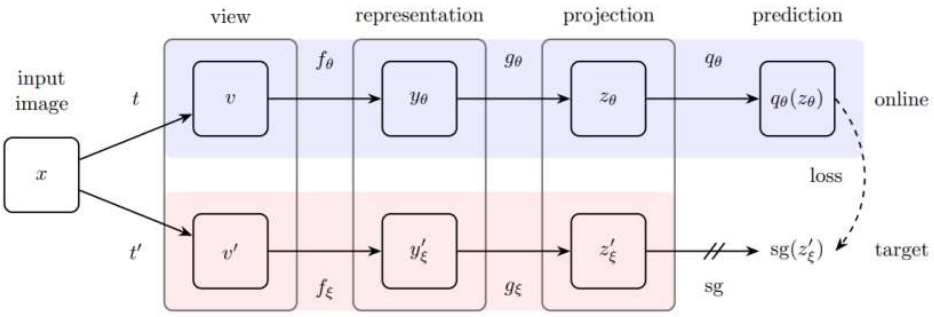


- Image - SimCLR

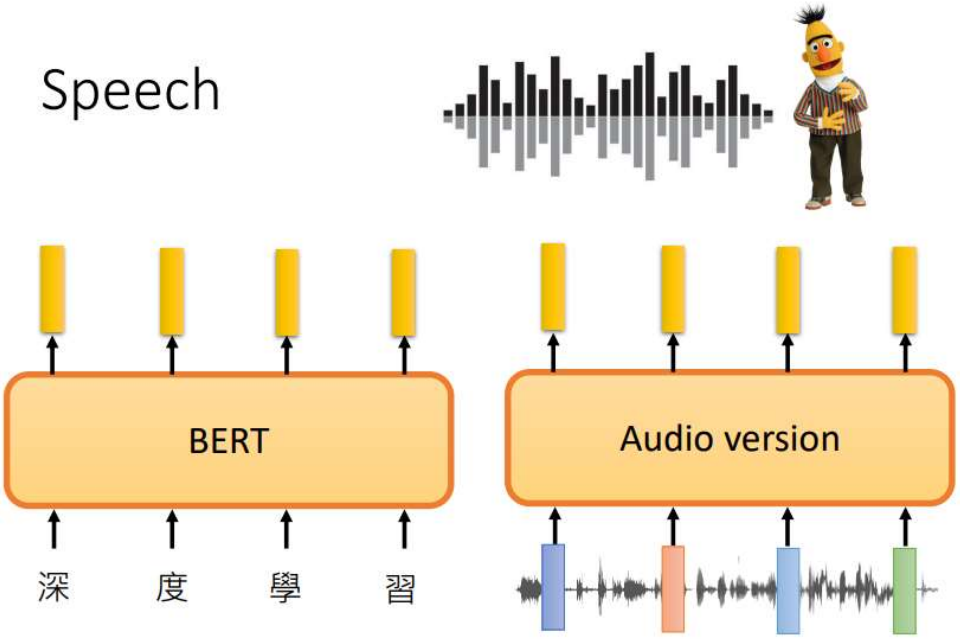
Image - BYOL

Bootstrap **y**our **o**wn **l**atent:
A new approach to self-supervised Learning

https://arxiv.org/abs/2006.07733



- Image - BYOL

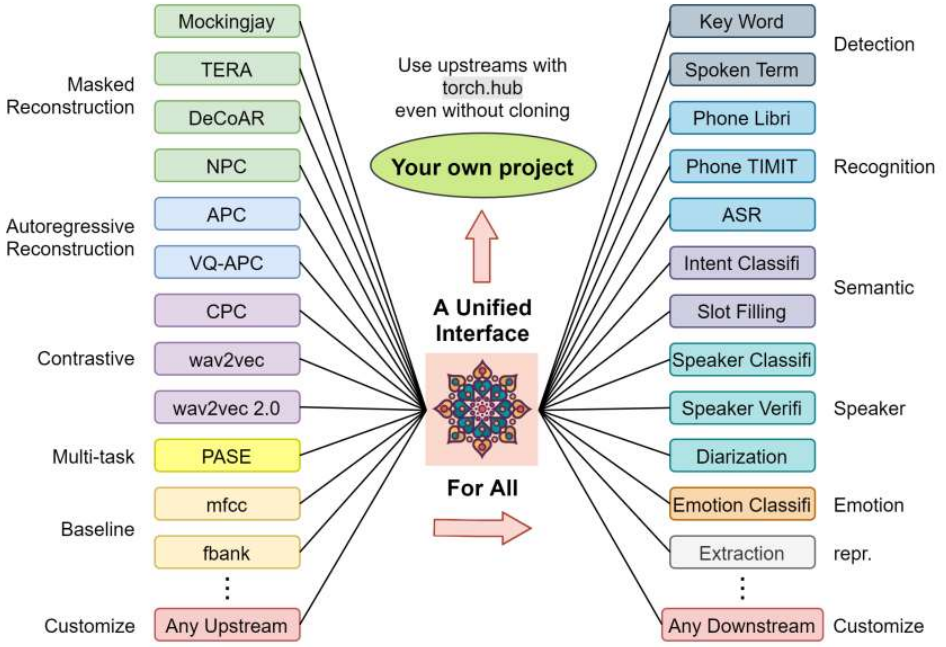


- 也可以做語音版的 GPT、BERT

Speech GLUE - SUPERB

- **S**peech processing **U**niversal **PER**formance **B**enchmark
  - Will be available soon
- **Downstream:** Benchmark with 10+ tasks
  - The models need to know how to process content, speaker, emotion, and even semantics.
- **Toolkit:** A flexible and modularized framework for self-supervised speech models.
  - https://github.com/s3prl/s3prl

- 在語音上還沒有類似 GLUE 基準的資料庫
- **SUPERB**



- Toolkit 裡面包含各式各樣 self-supervised learning 模型
- 以及這些 self-supervised learning 模型可以做的各式各樣的下游的任務

- self-supervised learning 技術，不只能用在文字上，在影像上、語音上都仍然有非常大的空間可以使用 self-supervised learning 的技術

tags: 2022 李宏毅_機器學習