



Demystifying Deep Learning (Especially ChatGPT)

Prof Richard Xu

Department of Mathematics, HKBU

2023-02-27

AI history up to 2017

HKBU :: MATH :: Prof Xu



Deep Learning Recent Development timeline

HKBU :: MATH :: Prof Xu

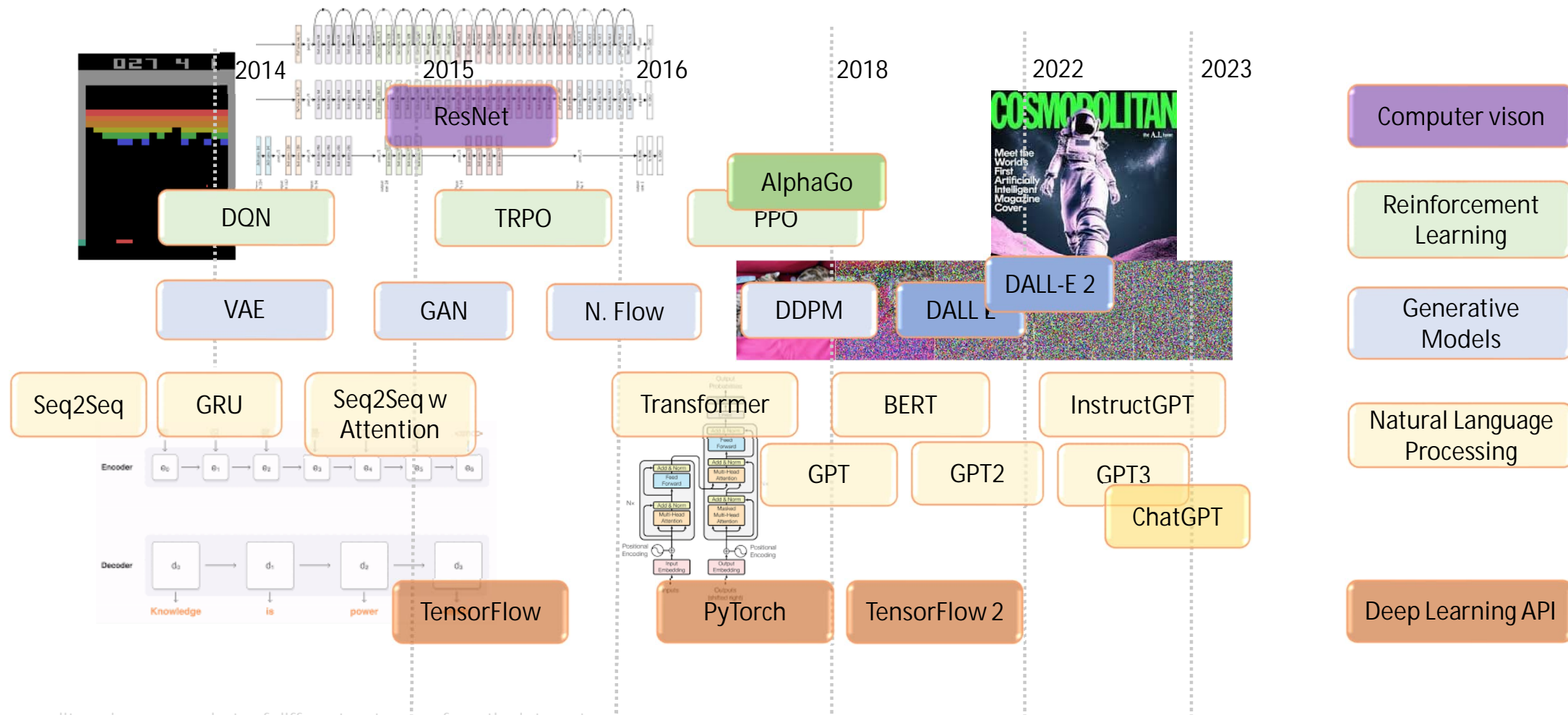


Image credit: various screenshots of different system are from the internet

AI and Big Data: other ingredients

HKBU :: MATH :: Prof Xu



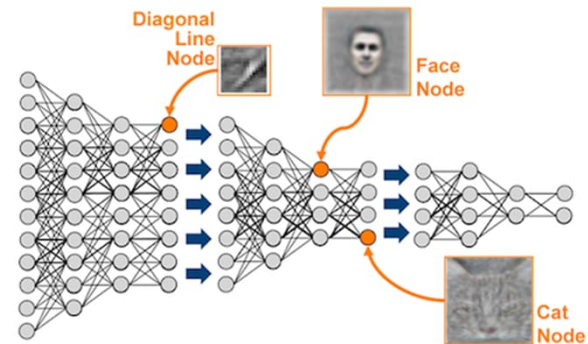
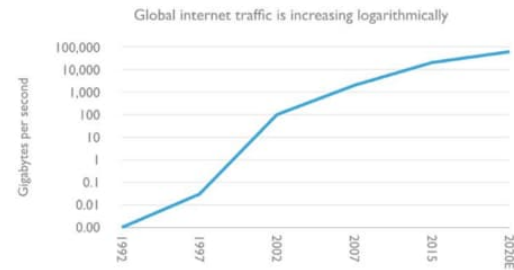
Cloud accessibly for Big Data



Better algorithms



Much powerful GPU



Buzzwords: AI, ML & DL

HKBU :: MATH :: Prof Xu

Artificial Intelligence

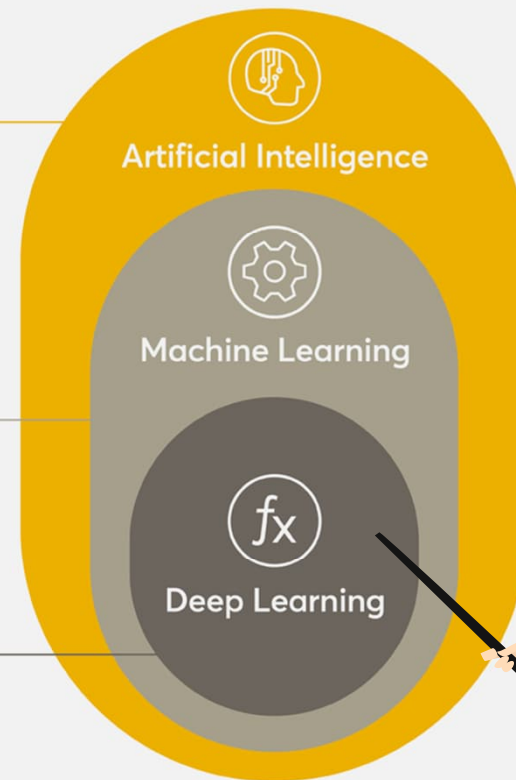
Any technique which enables computers to mimic human behaviour.

Machine Learning

Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

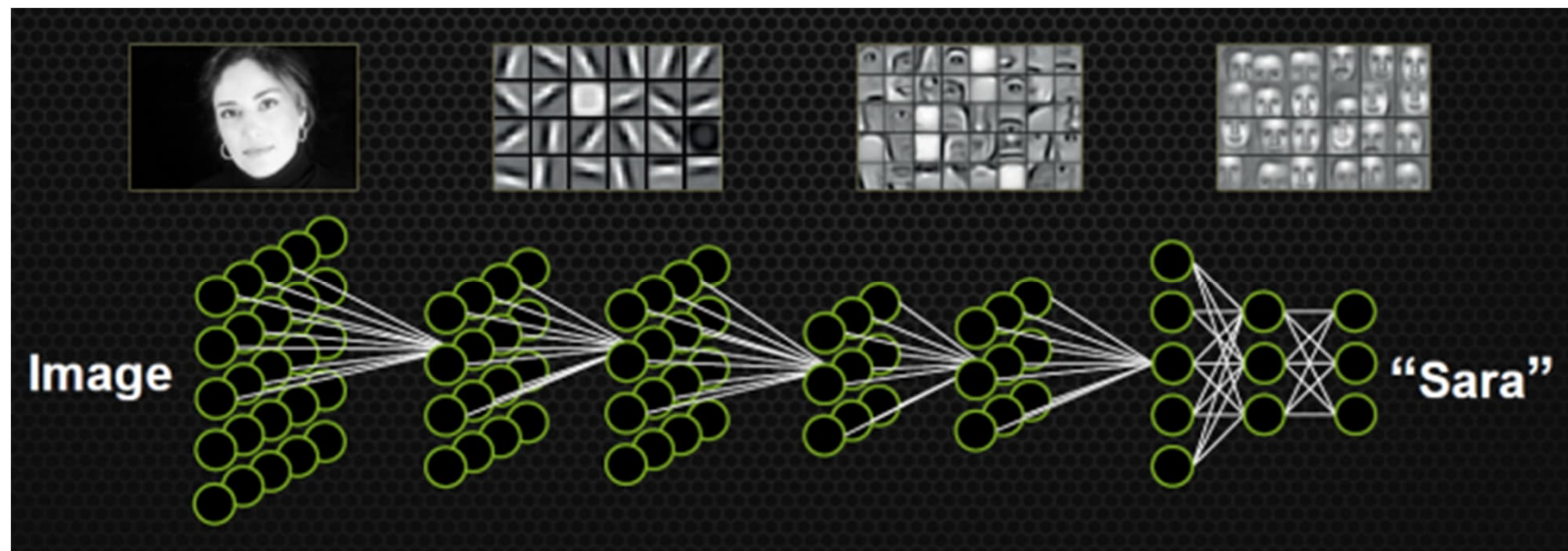
Deep Learning

Subset of ML which makes the computation of multi-layer neural networks feasible.



Early Deep Learning: Convolutional Neural Network:

HKBU :: MATH :: Prof Xu



More details are found on my Machine Learning notes:

https://github.com/roboticcam/machine-learning-notes/blob/master/cnn_beyond.pdf

Early Deep Learning: Recurrent Neural Network

HKBU :: MATH :: Prof Xu



Elaborate at chatGPT

02

Current Deep Learning technologies

Computer vision

Reinforcement
Learning

Generative
Models

Natural Language
Processing

Deep Learning
API

DL technologies: before we start...

HKBU :: MATH :: Prof Xu

Take home Message

- ML is not that far away from University research as you might think
- Still a lot of research needs to be done
- I select ML applications that I have worked: so you can relate to it easily
- Too many DL areas, I only concentrate on a few

Let's start

DL technologies: 3D computer vision

HKBU :: MATH :: Prof Xu

Computer vision

Previous work without DL



How do I coordinate cameras?

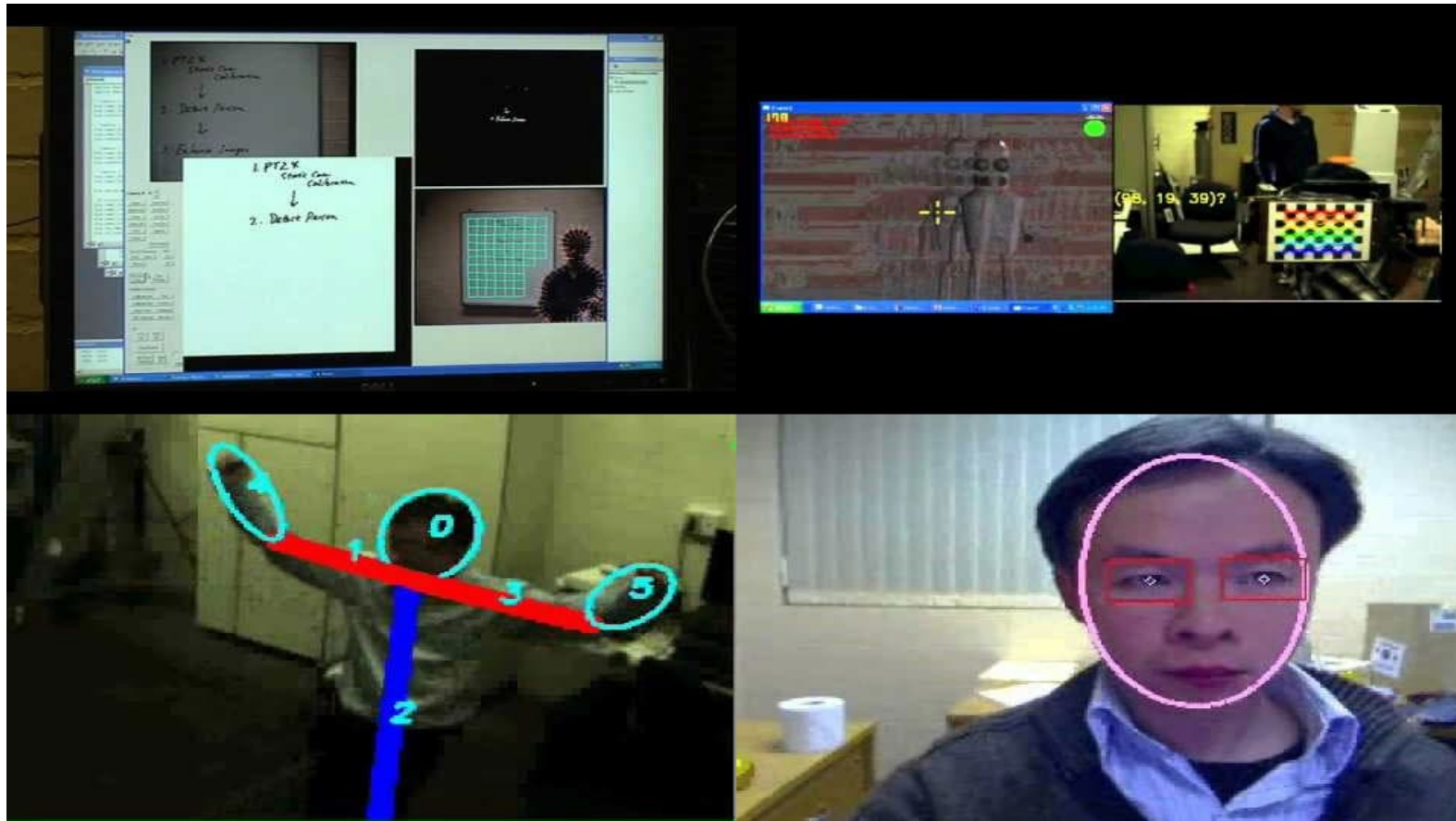


DL technologies: 3D computer vision

HKBU :: MATH :: Prof Xu

Computer vision

Previous pre-Deep Learning research

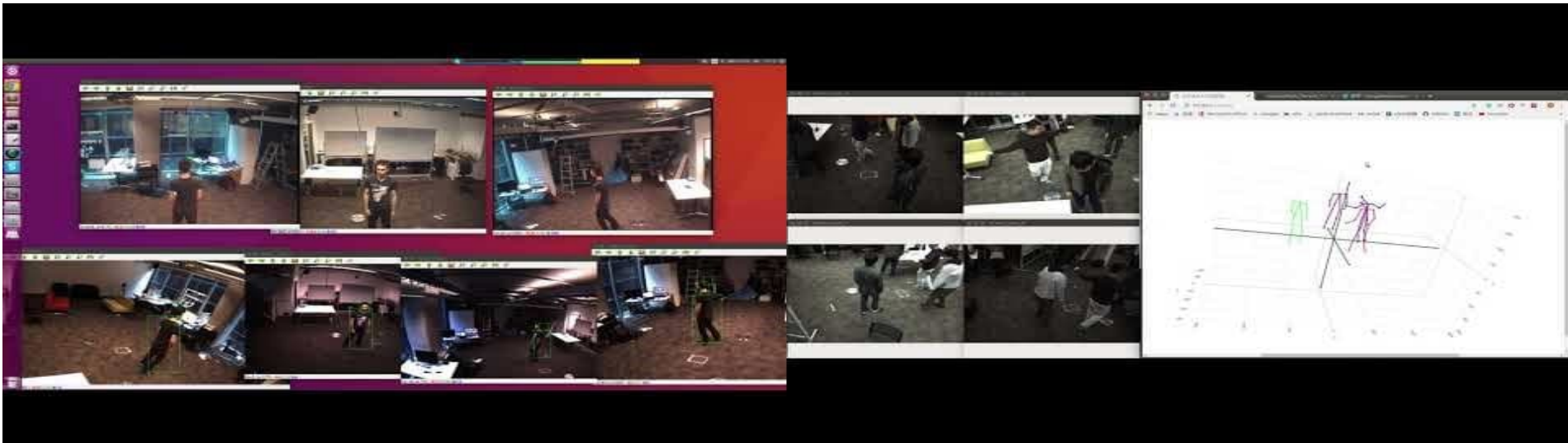


DL technologies: 3D computer vision

HKBU :: MATH :: Prof Xu

Computer vision

Newer Research with Deep Learning



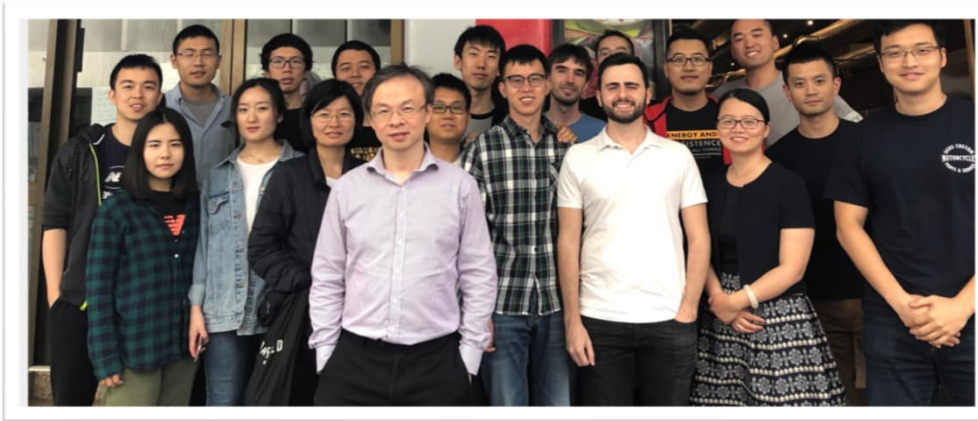
DL technologies: 3D computer vision

HKBU :: MATH :: Prof Xu

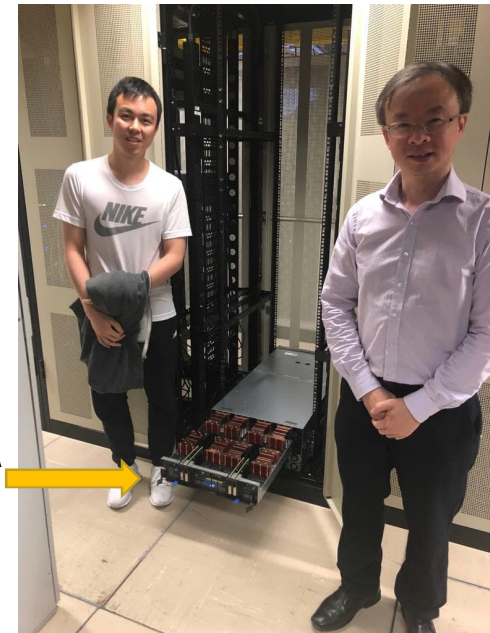
Computer vision

Just DL technology is not enough We also need:

- Many talented PhD students and engineers
- Lots of high-end GPUs



NVIDIA
DGX-1



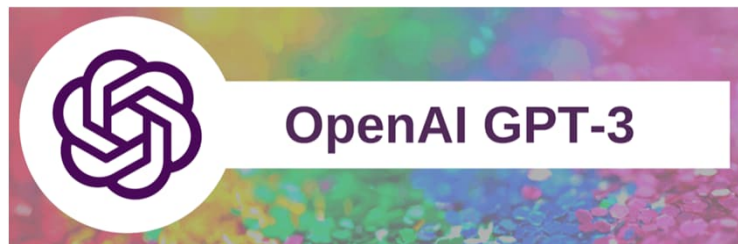
DL technologies: language model

HKBU :: MATH :: Prof Xu

Natural Language
Processing

Present in NLP:

- ChatGPT is dominating the news!
- In essence, it is a huge language model



Future in NLP:

- Produce the world's best selling (in cryptocurrency maybe) novel

What we have done

JD Auto Completion Tool

Classification

Number of
words

[Submit Form](#)

Initial Text

a global brand and an industry leader that is based sydney urgently require a ux designer to join their team of highly talented and creative individuals this is a very high profile project that will be used nation wide . we are looking for **an experienced digital developer to**

- Anyone can train their own (tailored) language model

Chatbot:

- Google Duplex



Image credit: BuzzFeed

Future AI-chatbot:

- Chat naturally with a super-knowledgeable human – chatGPT is just the start!

What we have done

- My former PhD student's research in Aspect-based sentiment analysis

"The décor (negative) is not special at all but their amazing food (positive) makes up for it"

DL technologies: cross domain translation

HKBU :: MATH :: Prof Xu

Generative
Models

Present:

- DALL-E 2



Future:

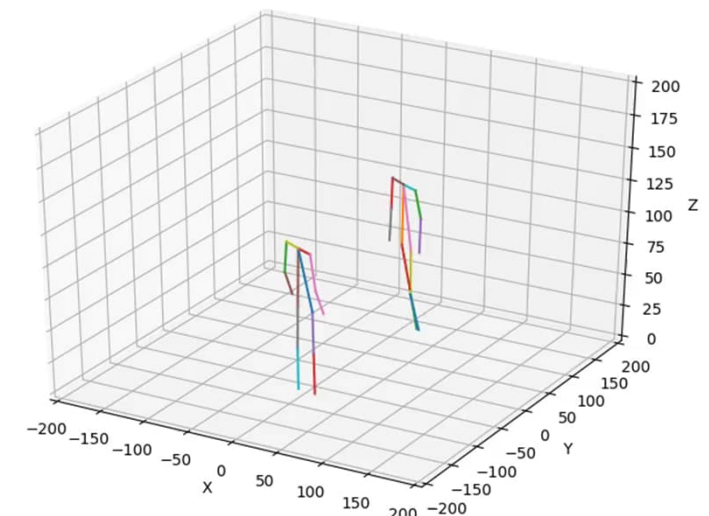
- AI create full-feature film

Image credit: Times <https://time.com/collection/best-inventions-2022/6225486/dall-e-2/>

What we have done

- My teams' work ins script to animation

I meet a friend, check my
watch and suggest we leave



DL technologies: music generation

HKBU :: MATH :: Prof Xu

Generative
Models

Present AI music:

- SONY CSL Research Lab created first-ever entire songs composed by robot: Daddy's Car



- Google's AI song



Future AI music:

- AI compose a set of viral songs, i.e., AlanWalkerBot

What we have done

- Two hand's piano play form one-hand play input

- One hand



- Left hand + Right hand



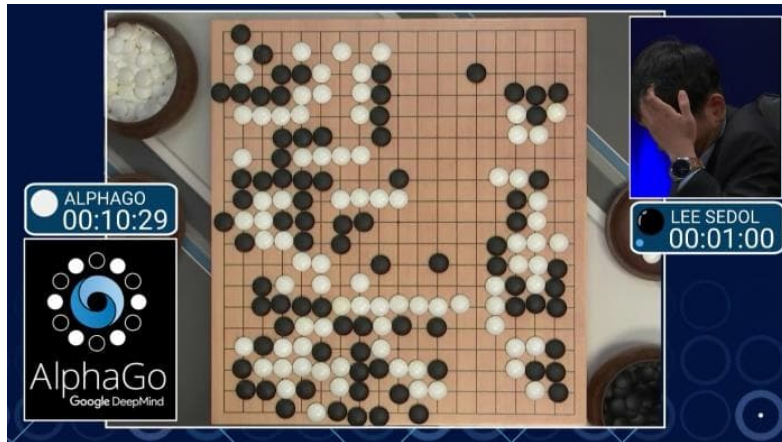
DL technologies: RTS strategy

HKBU :: MATH :: Prof Xu

Reinforcement
Learning

Present AI-powered gaming

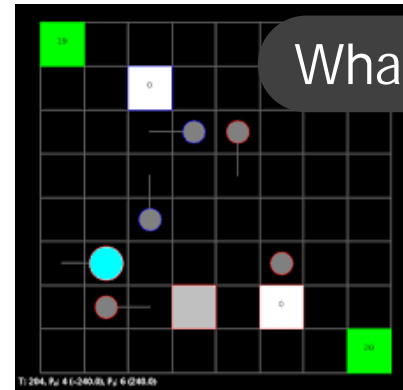
- Google DeepMind AlphaGo



Future AI-powered gaming:

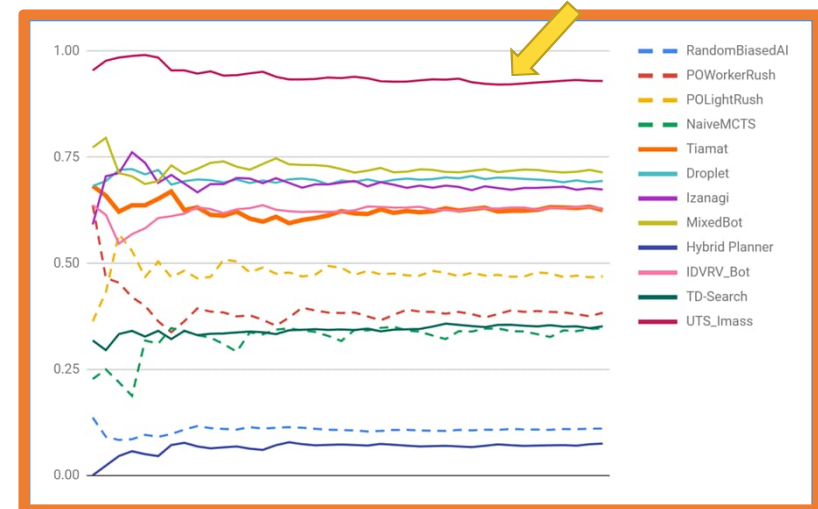
- DeepMind alphaStar beat best human player in StarCraft

What we have done



- My team's work in MicroRTS competition (2019 results)

My team @UTS



DL technologies: some background on RL

HKBU :: MATH :: Prof Xu

Reinforcement Learning

Supervised Learning

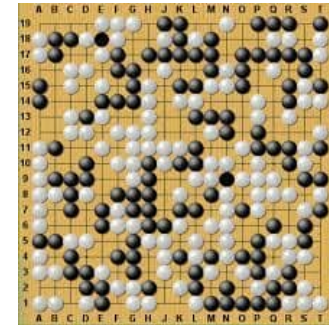


Human can easily label this image as "elephant"

- Has instructional label
- Data is usually all available for training

Elaborate at chatGPT

Reinforcement Learning



Human cannot easily label this image as "win"

- only reward signal
- feedback is delayed, i.e., not instantaneous
- data are not i.i.d., (consecutive frames are similar)
- agent's actions affect subsequent data it receives
- find the policy most likely to maximize expected future returns by taking current actions

DL & AI in real world:

HKBU :: MATH :: Prof Xu

My industry focused research (Australia)

6 years, 14 companies, 24 individual AI projects developing following technologies:

- Product range modelling
- Customer segmentation
- Predictive inventory/maintenance
- Recommendation system
- Personalized email/campaign
- Product correlations
- Transport passenger prediction
- HR performance indicator prediction
- Sentiment analysis
- ...
- Smart city applications
- CityRail Bus replacement optimization
- Injury prediction
- Peer-to-peer lending ranking
- Process optimization
- Education-to-job alignment
- Real-time strategy for defense
- Network traffic modelling
- ...

Talent hunting in Data Mining

HKBU :: MATH :: Prof Xu

JUMP
mingpao.com

Home Career News Market Trends Daily News 求職啟事 Work-life Balance Video
CAREER NEWS / INFORMATION TECHNOLOGY

AI 行業新興工種多 早着人工智能先機 實踐
「智慧」工作
(資訊科技) IT AI FOCUS
2020-01-17

AI will eliminate 1 of 8 jobs in Asia by 2024
by Alison DeNisco Rayome in CMO
on May 13, 2019, 6:16 AM PST

Artificial intelligence and automation will steal more jobs than it enhances, according to an MIT Technology Review report.

Kelly Services发布《2021香港薪资指南》，在疫情的摧残下房地产行业也败下阵来，今年最赚钱的职业TOP2都被Data岗位收归囊下。数据架构师最低薪资都有HK\$900K

Data Architect	HK\$900K - 1.0M
Data Scientist	HK\$660K - 1.0M
Machine Learning Specialist	HK\$540K - 720K
Blockchain Backend Developer	HK\$360K - 540K
IT Project Manager	HK\$600K - 1.5M
IT Business Analyst	HK\$600K - 1.5M
IT Solution Architect	HK\$600K - 1.3M
IT Programmers / Analyst Programmers / System Analyst	HK\$200K - 720K
BIM Modeler - Construction Design Consultancy	HK\$247K - 325K

JobsDB by SEEK

【新興工作】全球人工智能未成熟 港急直追造就AI人材
搶手

HKSTP HK10X

AI AND ROBOTICS TO REVITALISE HONG KONG INDUSTRY, WORKFORCE AND SOCIETY
PIONEERS TRANSFORMATION OPPORTUNITIES MEDIA STORIES
JOIN US 簡 / 繁 / 英

Data Scientist
Tsam Sha Tsui, Kowloon
電話查詢: 2,854,0000

Position Overview

We are seeking a driven and innovative Data Scientist to join our growing Data team and help generate, design, and implement solutions that solve the company's day-to-day business and operational challenges. In this role, you will utilise your analytical, statistical, and programming skills to collect, analyse, and interpret large data sets, support data experiments, rapid data insights and drive data-driven decision making. You will be able to work with hundreds of millions of customer data points and directly influence how internal clients operate.

Essential Duties & Responsibilities

- Act as an internal data consultant that identifies long-standing bottlenecks, advises on which technologies, platforms and methods to use and helps creating solutions from project inception to completion
- Work closely with cross-functional teams to develop data-driven systems that organize, collect and standardise data to solve complex business problems
- Mine and analyse data from various sources to drive automation through developing custom data models and processes, such as sales forecasting, customer clustering or assortment planning
- Drive automation through developing custom data models and processes, such as sales forecasting, customer clustering or assortment planning
- Using large data sets to find opportunities for process optimisation as well as using models to test the effectiveness of different courses of action

Experience, Skills & Knowledge

- Bachelor's degree or higher in Statistics, Computer Science, Mathematics, Software Engineering, or related quantitative field (preferred, but not required)
- Strong expertise with RBA programming in Excel for prototyping and maintaining current processes until transition to other systems
- Proven track record with C++, Python, VBA, R, SAS or other related languages and platforms
- Hands-on experience with data mining / data analysis method using a variety of data tools, building and implementing models, using / creating algorithms and treating / running simulations
- Strong ability using a variety of data tools, building and implementing models, using / creating algorithms and treating / running simulations

Senior Data Analyst (Data Mining/ETL)

Swing Consulting
Hong Kong

Apply

Glassdoor

28 days ago Full-time No degree mentioned

- Develop and implement end to end data pipeline from collection, analysis to visualization
- Manage data warehouse and apply data analysis for data modelling, solution design and
- Provide support for data issues, maintain best practices for data management and provide support
- Collaborate with business product, design, operations, marketing and growth team to strengthen data



03

Finally... chatgpt

intuitive explanation

Difference between



image

VS *“there is a unicorn standing at the river bank”*

piece of natural language

- Image is comprised by pixels, which are already in numeric form
- Natural language comprised of words; they are nominal, incomparable tokens
- Therefore, first step is convert word tokens to numerical representations.

Easiest encoding

$$\begin{bmatrix} \text{"a"} & 1 & 0 & \dots & 0 \\ \text{"abbreviate"} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{"zoology"} & 0 & 0 & \dots & 1 \end{bmatrix}$$

- structure is huge and sparse.
- as long as total number of words in vocabulary
- every pair of words are exactly $\sqrt{2}$ apart. cannot measure similarity or dissimilarity between them
- Beautiful visualization: <https://projector.tensorflow.org/>

word2vec encoding

Mikolov et. al., (2013)

- Find numerical representation of each word such that their probability conditional on their neighbourhood is maximized.
- In an effort that words have similar representations if their neighbourhood are similar.
 - a very "*nice*" picture
 - a very "*beautiful*" picture
- Then, possible to perform arithmetic on words:



Deep Natural Language Process word2vec example

HKBU :: MATH :: Prof Xu

Natural Language
Processing

" the cat sit on the mat"

window = 3

Step 1: find all windows

- (a) "the", "cat", "sit", target: cat
- (b) "cat", "sit", "on", target: sit
- (c) "sit", "on", "the", target: on
- (d) "on", "the", "mat", target: the

Step 2: find all context target pairs

"the" ← "cat"
"sit" ← "cat"
"cat" ← "sit"
"on" ← "sit"
"sit" ← "on"
"the" ← "on"
"on" ← "the"
"mat" ← "the"

Step 3: maximize probabilities:

$$\Pr("the"|"cat") \times \Pr("sit"|"cat") \times \Pr("cat"|"sit") \times \Pr("on"|"sit") \times \\ \Pr("sit"|"on") \times \Pr("the"|"on") \times \Pr("on"|"the") \times \Pr("mat"|"the")$$

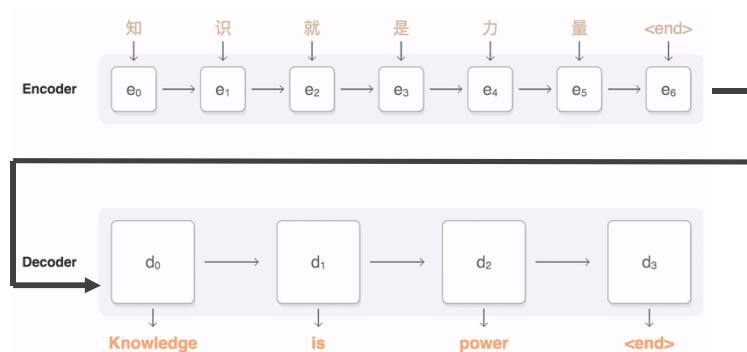
Deep Natural Language Process attention

HKBU :: MATH :: Prof Xu

Natural Language
Processing

Pure Seq2Seq

Sutskever et al., (2014)



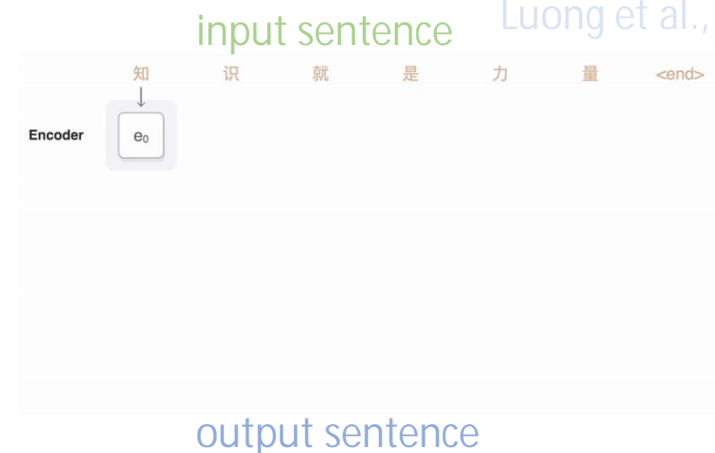
- Uses bi-directional Recurrent Neural Network

- Translation analogy:

A person first reads input from left to right (and right to left) to understand entire sentence. He then translate word for word into the output. The translation of each output word depends on the preceding translated words.

Seq2Seq with Attention

Luong et al., (2015)



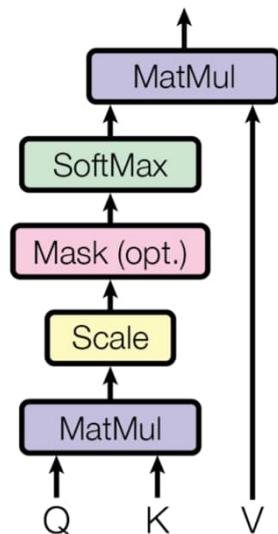
- added attention so elements of decoder can access individual elements of encoder directly.
- Translation analogy:
In addition to fully understanding input, when translating each word for output, the person also considers the amount of contribution (attention) from each individual input word.

different way to understand sentence

- Idea 1: One read a sentence, finding out how each individual word is related to the rest of words. Thus, he/she gets the deeper meaning of each word.
- Idea 2: With their deeper meanings, the person revisits the relationship between the words so he/she obtains an even deeper meaning for each word.
- Idea 3: The process is repeated many times until the person is satisfied with the depth of understanding
- Idea 4: Even better, we ask many people examine the same sentence using the same mechanism, since everyone may interpret it differently
- Idea 5: Finally, their combined understanding of each word becomes the final word representation

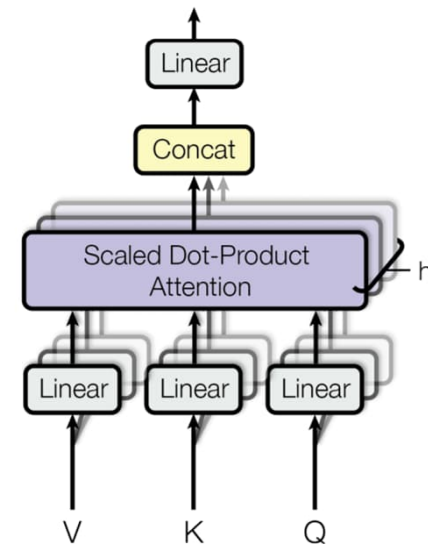
After this process, the understanding is so thorough!

Scaled Dot-Product Attention



encapsulate Idea 1, Idea 2, Idea 3

multi-head



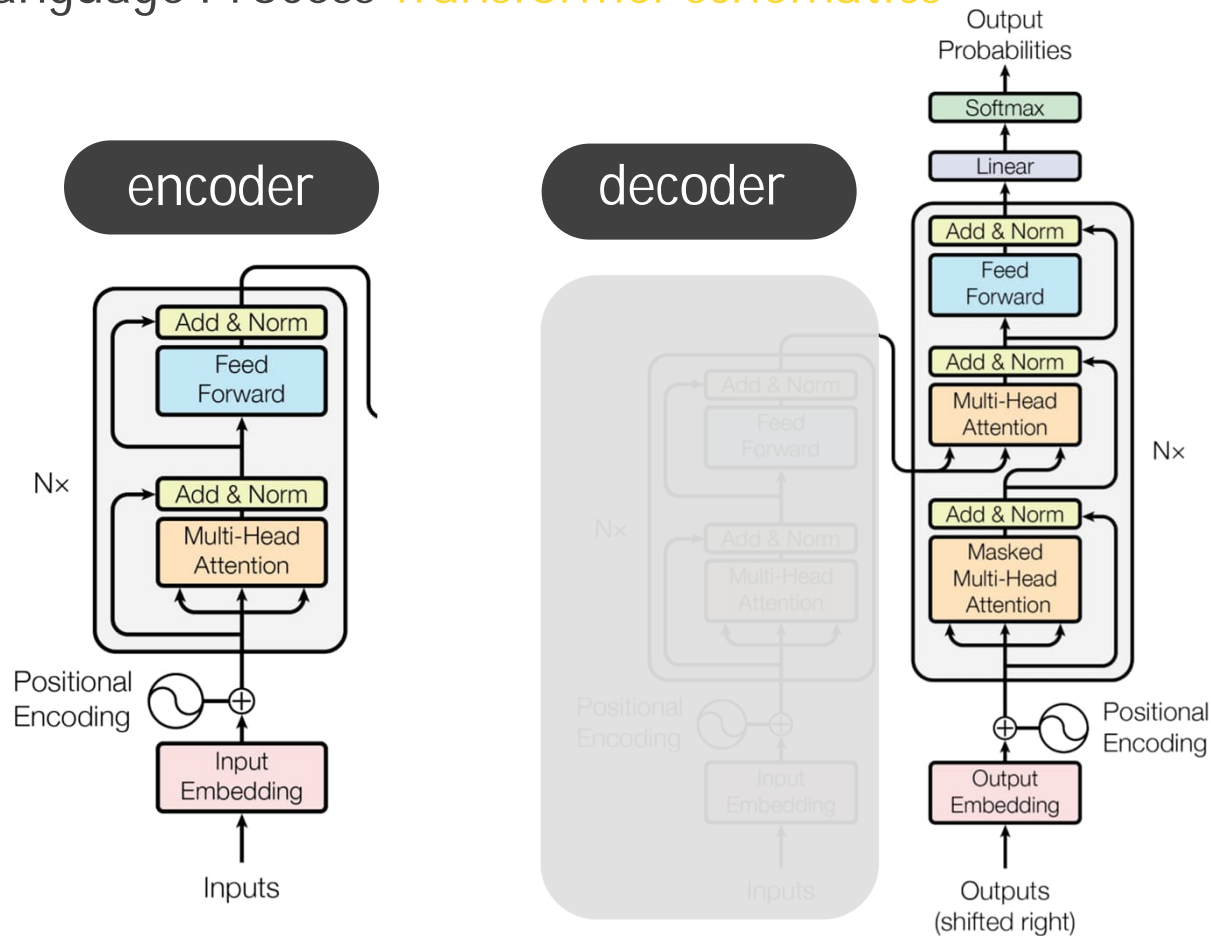
encapsulate Idea 4, Idea 5

■ For mathematical explanation, please refer to the accompanying notes in Moodle

Deep Natural Language Process Transformer schematics

HKBU :: MATH :: Prof Xu

Natural Language Processing



Put Ideas into seq2seq setting

Bidirectional Encoder Representations from Transformers

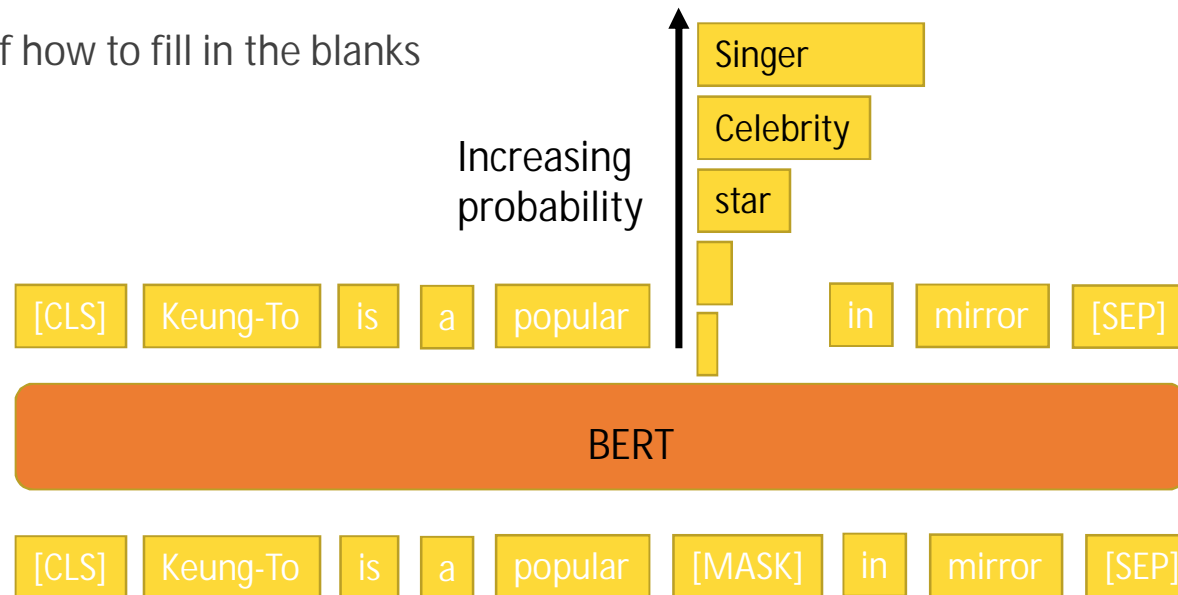
Devlin et al., (2018)

- Now that sentence embedding is also context dependant

*she walks along the river **bank***

*she works at an investment **bank***

- BERT trains itself of how to fill in the blanks



Bidirectional Encoder Representations from Transformers

- Now that sentence embedding is also context dependant

*she walks along the river **bank***

*she works at an investment **bank***

- BERT also trains itself of recognize if the adjacent sentence is indeed the next to each other

[CLS] Keung-To is a popular [MASK] in mirror [SEP] so is Yeung-lok-man [SEP]

Label = IsNext

[CLS] Keung-To is a popular [MASK] in mirror [SEP] weather is very nice [SEP]

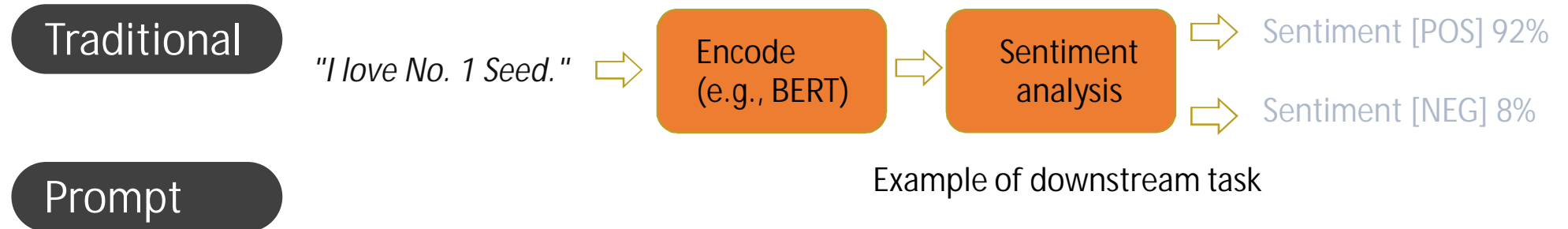
Label = NotNext

- In both cases, the training labels do not require human

Deep Natural Language Process concept of prompt

HKBU :: MATH :: Prof Xu

Natural Language
Processing



- So if design a template: "[X] it was a [Z] song."

$X = \text{"I love No. 1 Seed."}$



$X' = \text{" I love No. 1 Seed. it was a [Z] song."}$

$Z = \{\text{"excellent", "great", "wonderful" ...}\}$ [POS sentiment]
 $Z = \{\text{"bad", "awful", "silly" ...}\}$ [NEG sentiment]

- Language models are so good that you can just generate word(s), instead of solving an additional problem (sentiment analysis)
- Can perform zero-shot learning

Deep Natural Language Process concept of prompt

HKBU :: MATH :: Prof Xu

Natural Language
Processing

with powerful language model

- Language model *understands* prompt

[CLS] how many members in the band mirror [SEP] there are 12 members

- the answer is generated to correspond to input sentence (input)

Unsupervised pre-training

- Maximize the probabilities of the words in sentences

$\text{Prob}(2^{\text{nd}} | 1^{\text{st}}, \text{param})$
 $\times \text{Prob}(3^{\text{rd}} | 2^{\text{nd}}, 1^{\text{st}}, \text{param})$
 $\times \text{Prob}(4^{\text{th}} | 3^{\text{rd}}, 2^{\text{nd}}, 1^{\text{st}}, \text{param})$
 $\times \dots$

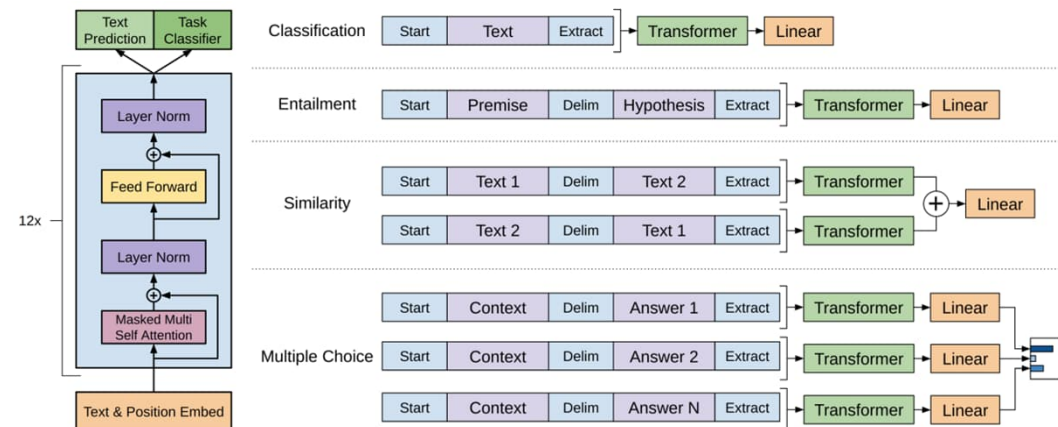
- For each head:

first layer: word tokens, word embedding, position embedding $\Rightarrow h_0$
 repeat: $h_l = \text{transformer}(h_{l-1})$ last layer h_n

- Can perform zero-shot learning:

$(h_n, \text{word embeddings}) \Rightarrow \text{words probabilities}$

Supervised fine-tuning



Deep Natural Language Process InstructGPT

HKBU :: MATH :: Prof Xu

Natural Language
Processing

GPT3 + InstructGPT = chatGPT

- When grab data from the internet, answer may not always be useful.

e.g., "who sings the song "No. 1 Seed"?"

Keung-To.

← useful answer

I don't know, please tell me!

← not a useful answer

- Need human (teacher) guidance

"who sings the song "No. 1 Seed"?"



A teacher model



Keung-To.



I don't know, please tell me!

high

Reward
signal in RL

low

Where to find more material?

HKBU :: MATH :: Prof Xu

- My never-ending Machine Learning knowledge dissemination notes:

<https://github.com/roboticcam/machine-learning-notes>

- All my digital footprint (links to things I talked about today)

https://github.com/roboticcam/demo_links

- Mathematics supplementary notes on Moodle:

Live Machine Learning Class:

中文机器学习研究线上课

2022年我坚持每周日晚上8:30直播机器学习研究课程系列 (微信二维码在这个链接)- From 2022, I hold regular 8:30pm Sunday Night live (SNL) broadcast on Machine Learning theory.

English version

From April 2022, I started a machine learning research seminar series every 2-3 weeks in English via Zoom. It's at 7pm Hong Kong Time. I will continue to explain machine learning using an intermediate level mathematics. The current topic is: "Gradient Descent Research". You need a solid understanding of linear algebra, calculus, probability and statistics. You can register via meetup <https://www.meetup.com/machine-learning-hong-kong/> (Back in Australia, I also conducted research training to all machine learning PhD students at Australian universities, with over 100 students participating via Zoom.)

Learning Theory Classes

- [Class 1: Introduction](#)

Mathematics in Modern Natural Language Modeling

Richard Xu

February 27, 2023

1 A few words

To start this topic, I will first discuss the mathematics in modern natural language processing. By the way, Natural Language Processing (NLP) is one of the most important/exciting applications of artificial intelligence, machine learning and data mining. Contrary to computer vision, NLP will influence the work of many people in the future.

Although neural networks play an important role in NLP. However, we will cover neural networks later in this topic. Therefore, we temporarily avoid talking about N-N.

Also note that while this topic is about techniques in NLP, these techniques can also be applied to other machine learning and data science settings.