Presentation

Insincere Questions Classification

Cheng Zeng

CONTENTS

01

Introduction

THE PART ONE

01 Introduction

NLP

Let machines to understand the human language

Bert

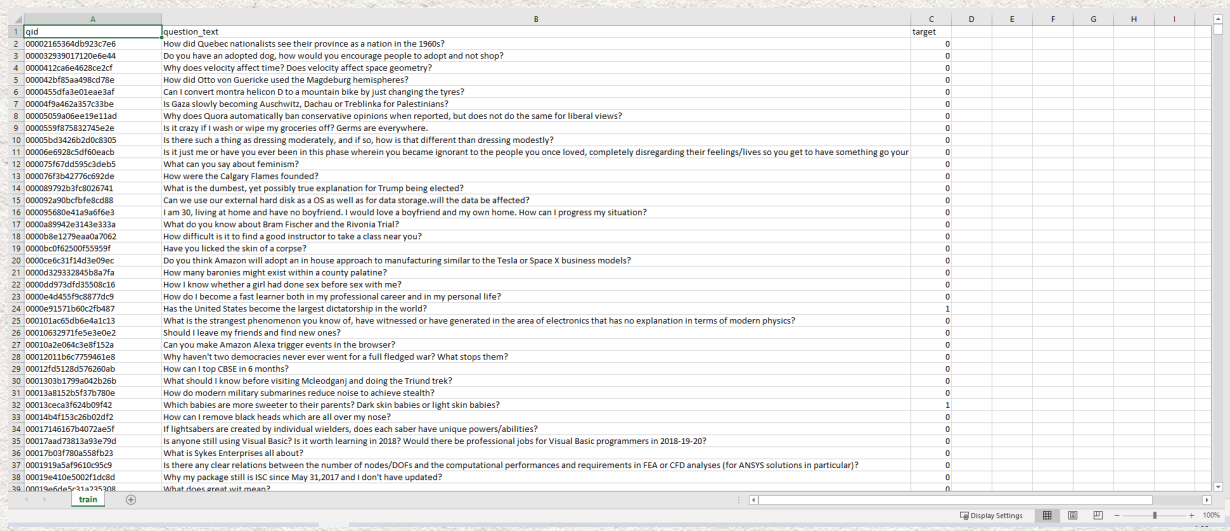
The popular model for the NLP nowadays

Detect the Insincere

Recognize the insincere context

01 Introduction

Dataset



	qid	question_text	target
2	000021653 64db923c7e6	How did Quebec nationalists see their province as a nation in the 1960s?	0
3	000032939017120e6e44	Do you have an adopted dog, how would you encourage people to adopt and not shop?	0
4	0000412ca6e4628ce2cf	Why does velocity affect time? Does velocity affect space geometry?	0
5	000042bf85aa498cd78e	How did Otto von Guericke used the Magdeburg hemispheres?	0
6	0000455dfa3e01eae3af	Can I convert montra helicon D to a mountain bike by just changing the tyres?	0
7	00004f9a462a357c33be	Is Gaza slowly becoming Auschwitz, Dachau or Treblinka for Palestinians?	0
8	00005059a06ee19e11ad	Why does Quora automatically ban conservative opinions when reported, but does not do the same for liberal views?	0
9	000055f875832745e2e	Is it crazy if I wash or wipe my groceries off? Germs are everywhere.	0
10	00005bd3426b2d0c8305	Is there such a thing as dressing moderately, and if so, how is that different than dressing modestly?	0
11	00006e6928c5df60eacb	Is it just me or have you ever been in this phase wherein you became ignorant to the people you once loved, completely disregarding their feelings/lives so you get to have something go your	0
12	000075f67dd595c3deb5	What can you say about feminism?	0
13	000076f3b42776c692de	How were the Calgary Flames founded?	0
14	000089792b3fc8026741	What is the dumbest, yet possibly true explanation for Trump being elected?	0
15	000092a90bcfbfe8cd88	Can we use our external hard disk as a OS as well as for data storage.will the data be affected?	0
16	00009568be41a9a6fee3	I am 30, living at home and have no boyfriend. I would love a boyfriend and my own home. How can I progress my situation?	0
17	0000a89942e3143e333a	What do you know about Bram Fischer and the Rivonia Trial?	0
18	0000b8e1279eaa0a7062	How difficult is it to find a good instructor to take a class near you?	0
19	0000bc0f62500f55959f	Have you licked the skin of a corpse?	0
20	0000ce6c31f14d3e09ec	Do you think Amazon will adopt an in house approach to manufacturing similar to the Tesla or Space X business models?	0
21	0000d329332845b8a7fa	How many baronies might exist within a county palatine?	0
22	0000dd973dfd3550dc16	How I know whether a girl had done sex before sex with me?	0
23	0000e4d455f9c8877dc9	How do I become a fast learner both in my professional career and in my personal life?	0
24	0000e91571b60c2fb487	Has the United States become the largest dictatorship in the world?	1
25	000101ac65db6e4a1c13	What is the strangest phenomenon you know of, have witnessed or have generated in the area of electronics that has no explanation in terms of modern physics?	0
26	00010632971fe5e3e0e2	Should I leave my friends and find new ones?	0
27	00010a2e064c3e8f152a	Can you make Amazon Alexa trigger events in the browser?	0
28	00012011b6c7759461e8	Why haven't two democracies never ever went for a full fledged war? What stops them?	0
29	00012fd5128d576260ab	How can I top CBSE in 6 months?	0
30	000130 3b1799a042b26b	What should I know before visiting Mcleodganj and doing the Triund trek?	0
31	00013a8152b5f37b780e	How do modern military submarines reduce noise to achieve stealth?	0
32	00013ceca3f624b09f42	Which babies are more sweeter to their parents? Dark skin babies or light skin babies?	1
33	00014b4f153c26b02df2	How can I remove black heads which are all over my nose?	0
34	000171461 67b4072ae5f	If lightsabers are created by individual wielders, does each saber have unique powers/abilities?	0
35	00017aad73813a93e79d	Is anyone still using Visual Basic? Is it worth learning in 2018? Would there be professional jobs for Visual Basic programmers in 2018-19-20?	0
36	00017b03f780a558fb23	What is Sykes Enterprises all about?	0
37	0001919a5af9610c95c9	Is there any clear relations between the number of nodes/DOFs and the computational performances and requirements in FEA or CFD analyses (for ANSYS solutions in particular)?	0
38	00019e410e5002f1dc8d	Why my package still is ISC since May 31,2017 and I don't have updated?	0
39	00019e6de5c31a225308	What does great wit mean?	0

train

02

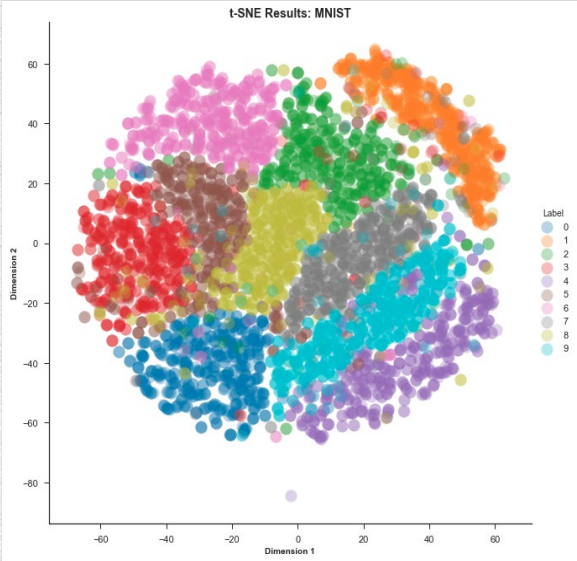
Background & Foundation

THE PART TWO

02 Background & Foundation

- Word embedding and word2vec
- Sequence-to-Sequence(seq2seq)
- Attention

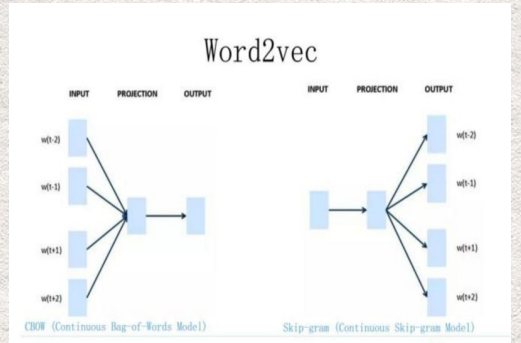
02 Background & Foundation



Embedding means using low dimensions vector to represent the articles

Thus word2vec is the basement for the Embedding

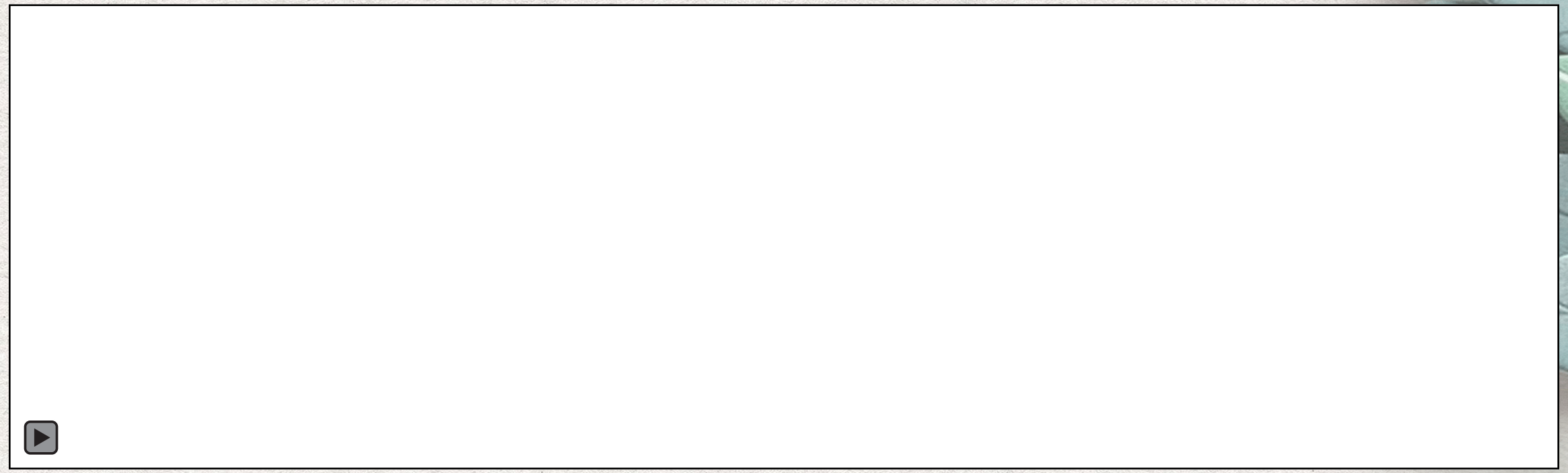
Two methods for Word2vec



Word embedding can transfer the discrete variable into vector variable as the input.

02 Background & Foundation

Sequence to Sequence model



02 Background & Foundation

Under the hood

02 Background & Foundation

Deeper under the hood

02 Background & Foundation

Transition to the Attention

02 Background & Foundation

Attention

02 Background & Foundation

Attention

03

Key Algorithms

THE PART THREE

03 Key Algorithms

Transformer:

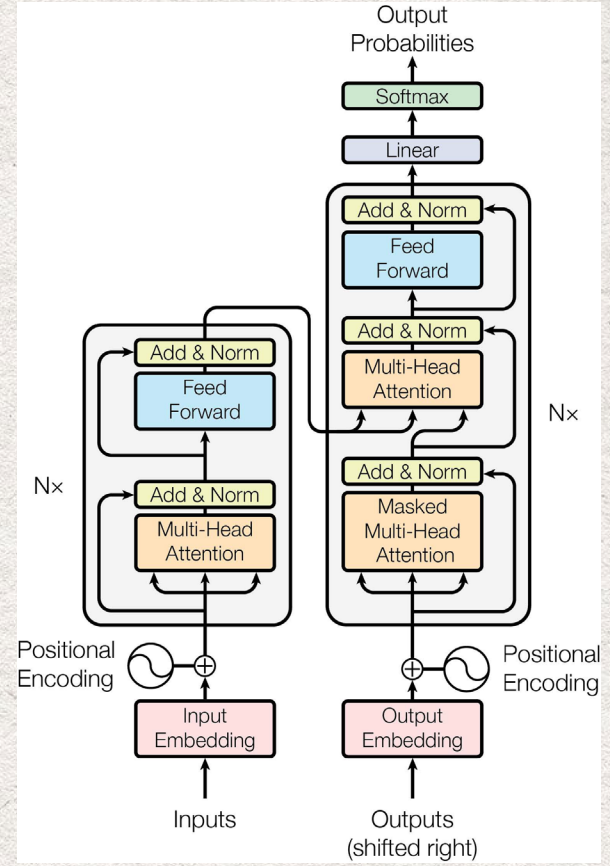
1. Self-Attention
2. Multi-Head Attention
3. Positional Encoding
4. Layer-normalization

Pre-training tasks:

1. Masked Language Model
2. Next Sentence prediciton

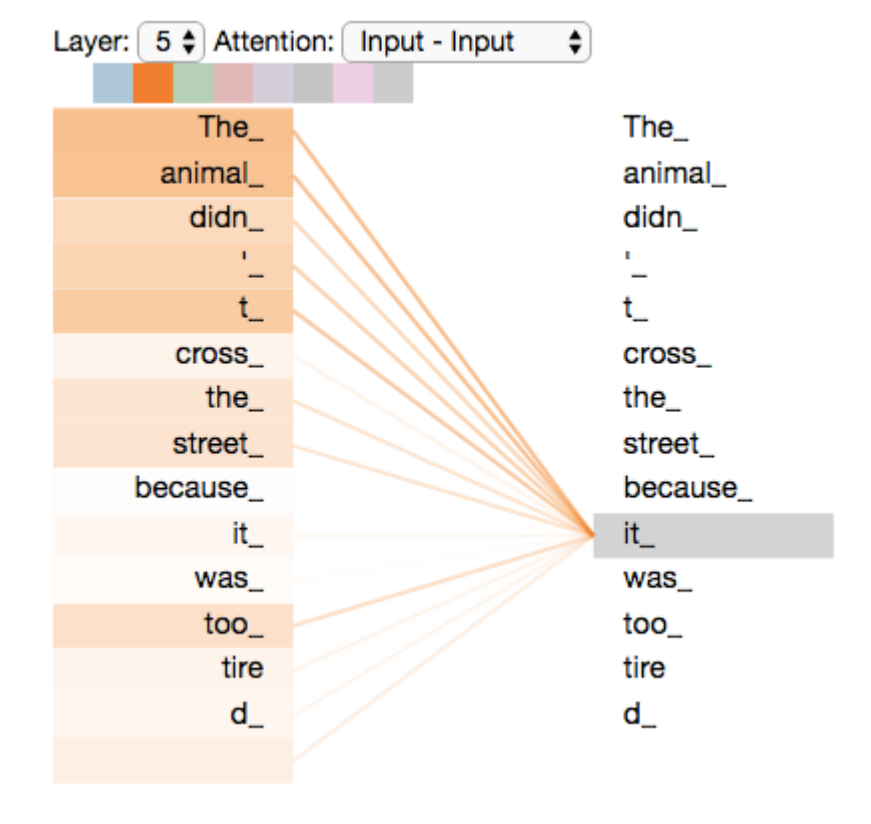
03 Key Algorithms

Architecture of Transformer



03 Key Algorithms

Self-Attention



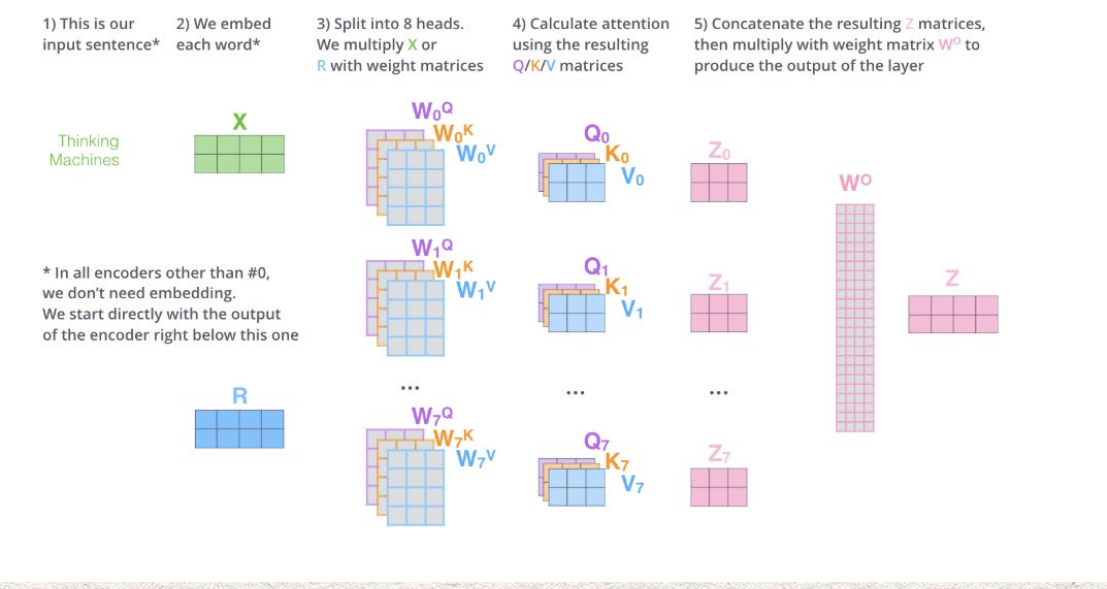
The self-attention calculation in matrix form

$$Z = \sum softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Three important vector: Query, Key and Value.

03 Key Algorithms

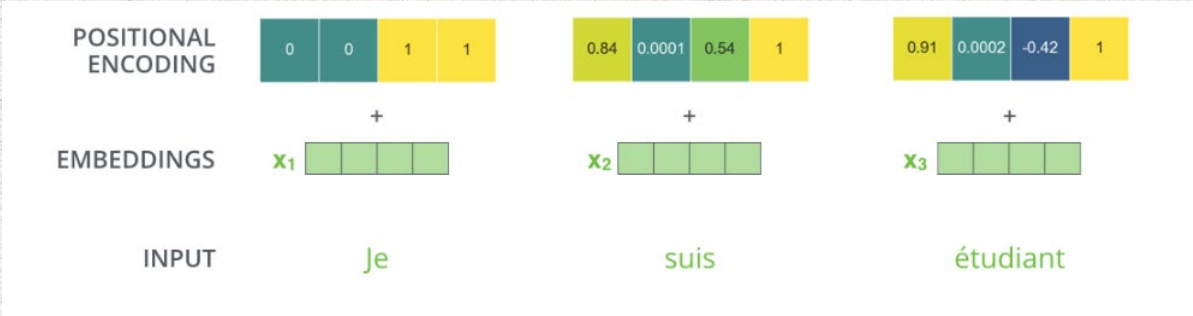
Multi-Head Attention



Multi-head attention just means multiple the self-attention several times and use a matrix transfer several Z into a Z in order to be the input of the feed-forward layer.

03 Key Algorithms

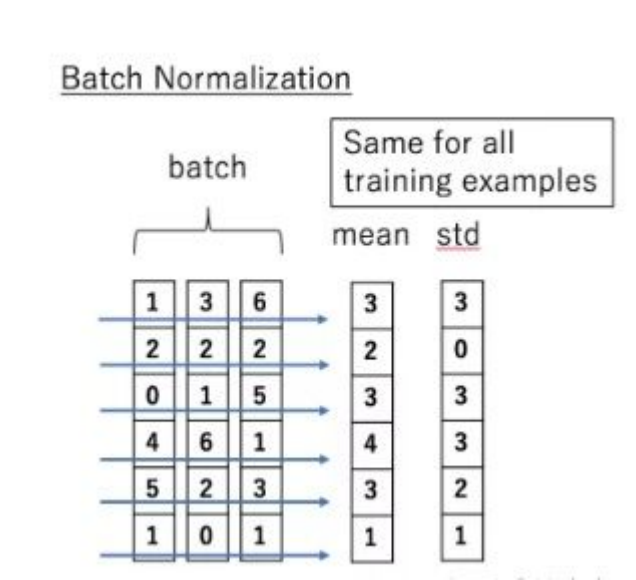
Positional Encoding

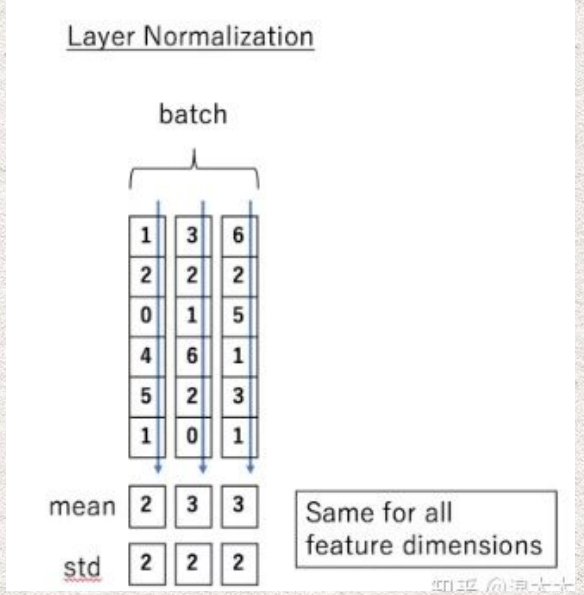


Embedding only give words itself vector but the position of words also needs vector. Thus, this is the positional encoding does.

03 Key Algorithms

Layer Normalization





Layer Normalization is similar BN, but norm at the vertical direction on the BN

04

Result & Learning

THE PART FOUR

04 Result

- Accuracy = 0.96
- F1-score = 0.65

04 Learning

This project show me a good view of NLP which I don't know before. Thus, now I have quite interesting about the further learning in this field.

Thanks

2020年4月15日