2023 Introduction to Artificial Intelligence

Machine Learning Project

Schedule

12/6
2023
Competition start

12/20
2023
Competition ends

12/22
2023
Report due (eeclass)

12/22
2023
Peer evaluation

Outline

- Introduction
- Kaggle
- Report

Introduction

Data

Data contents

- # examples: 500,000
- # attributes: 35
- # class (label 0/1): 2

Data

Data provided

- Training data (train.csv)
  - Data with labels, for model training
- Testing data (test.csv)
  - Data without labels, you need to predict the label of each data point
  - **Each data point has an id number**
  - The testing data is divided into two parts:
    - One for the public leaderboard
    - And the other for the private leaderboard
- Sample submission
  - The prediction you submit on Kaggle should follow the same format as the sample_submission.csv

Performance indexes

F-score

- $Precision\ score = \frac{tp}{tp+fp}$
- $Recall\ score = \frac{tp}{tp+fn}$
- $F-score = 2 \cdot \frac{precision \cdot recall}{precision+recall}$

Macro F-score

- The metric used for Kaggle leaderboard
  - Can refer to the code

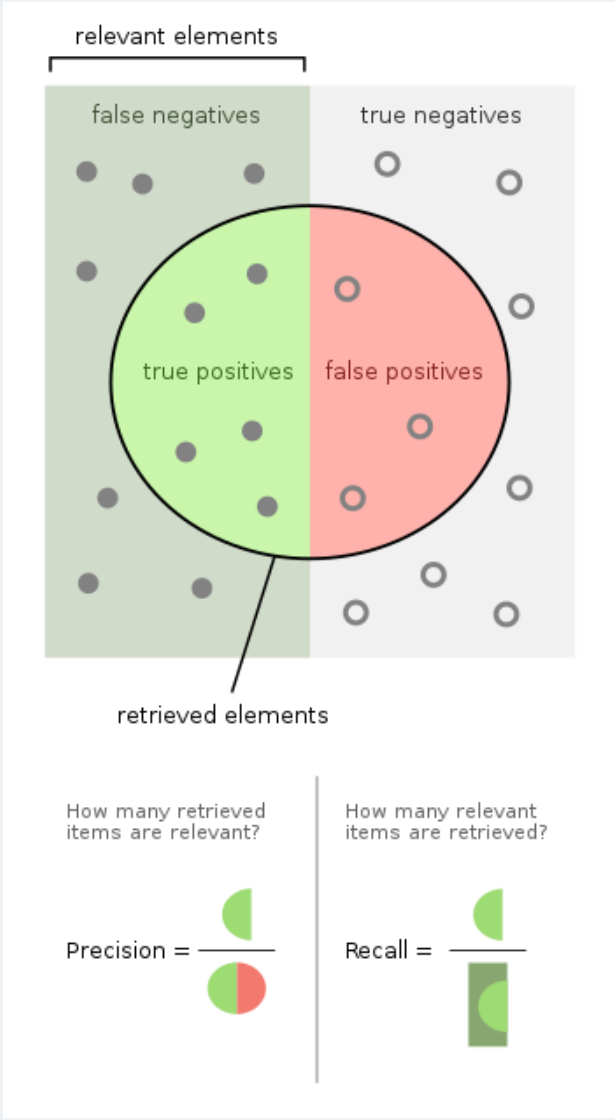


Image from https://en.wikipedia.org/wiki/Precision_and_recall

Kaggle

Join the Kaggle contest

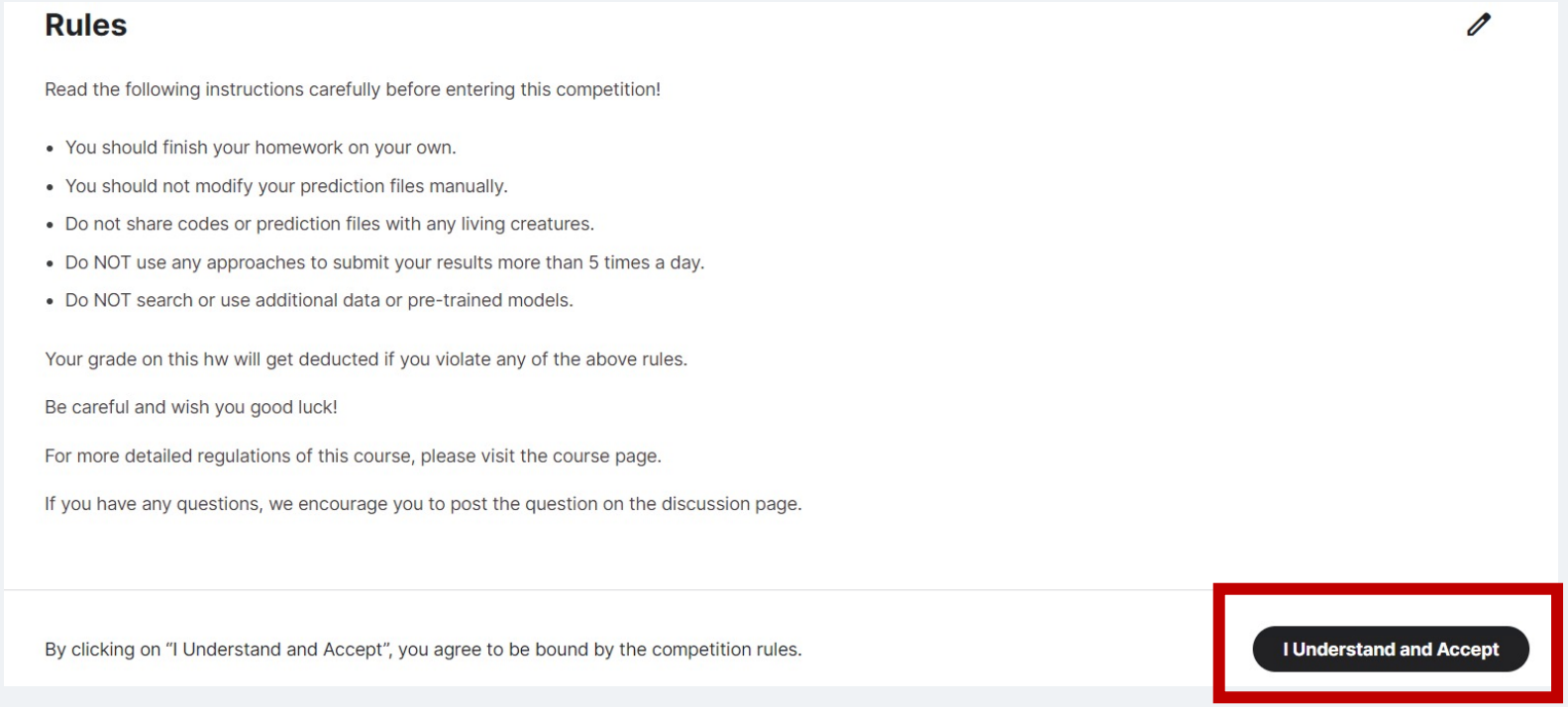
Join through the Kaggle link

- You can get access to the contest if you've already filled out the team up google form
- The Kaggle link
  - Please contact TA 黃恩平 if you do not have access to the contest

Join the Kaggle contest

Accept the rules

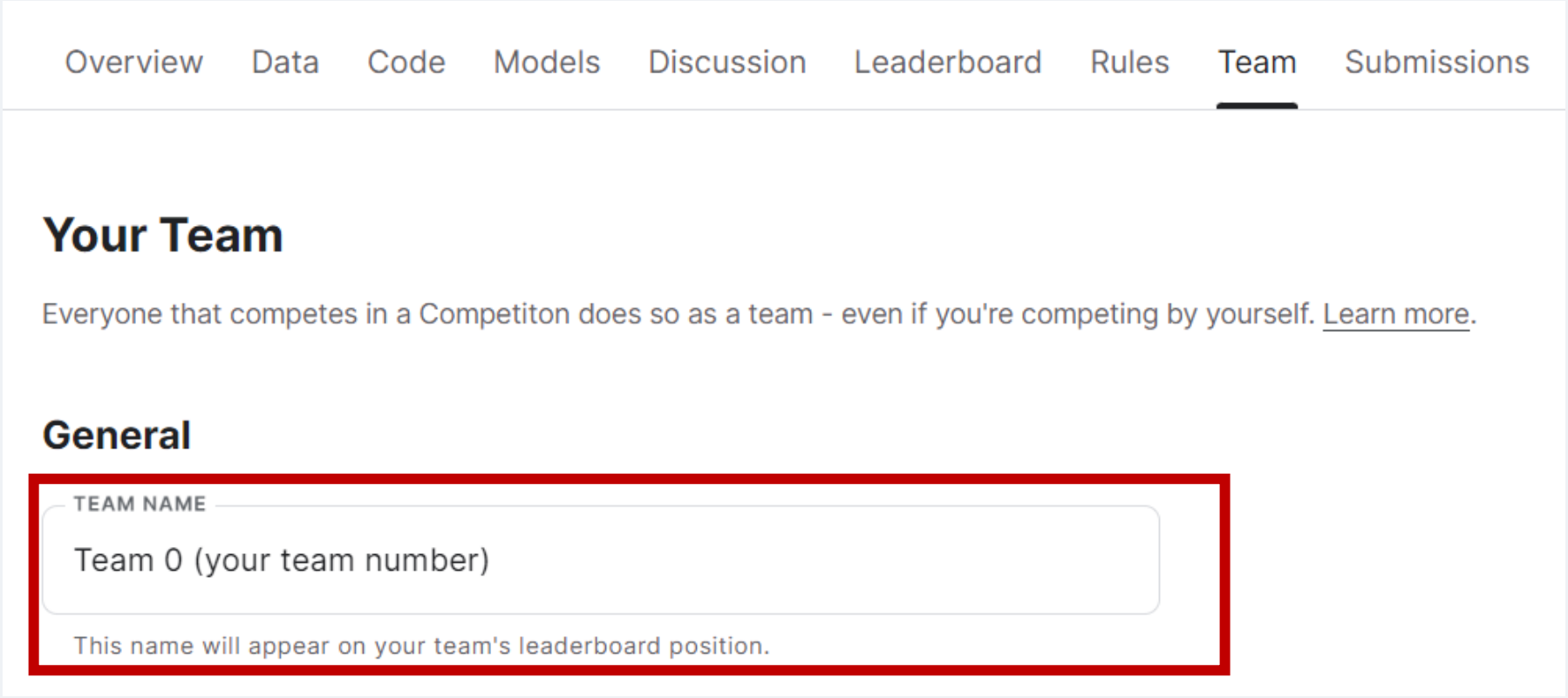
- Click “I understand and Accept” to start competing on the “Rules” page



Team up (Important!!)

Change your team's name

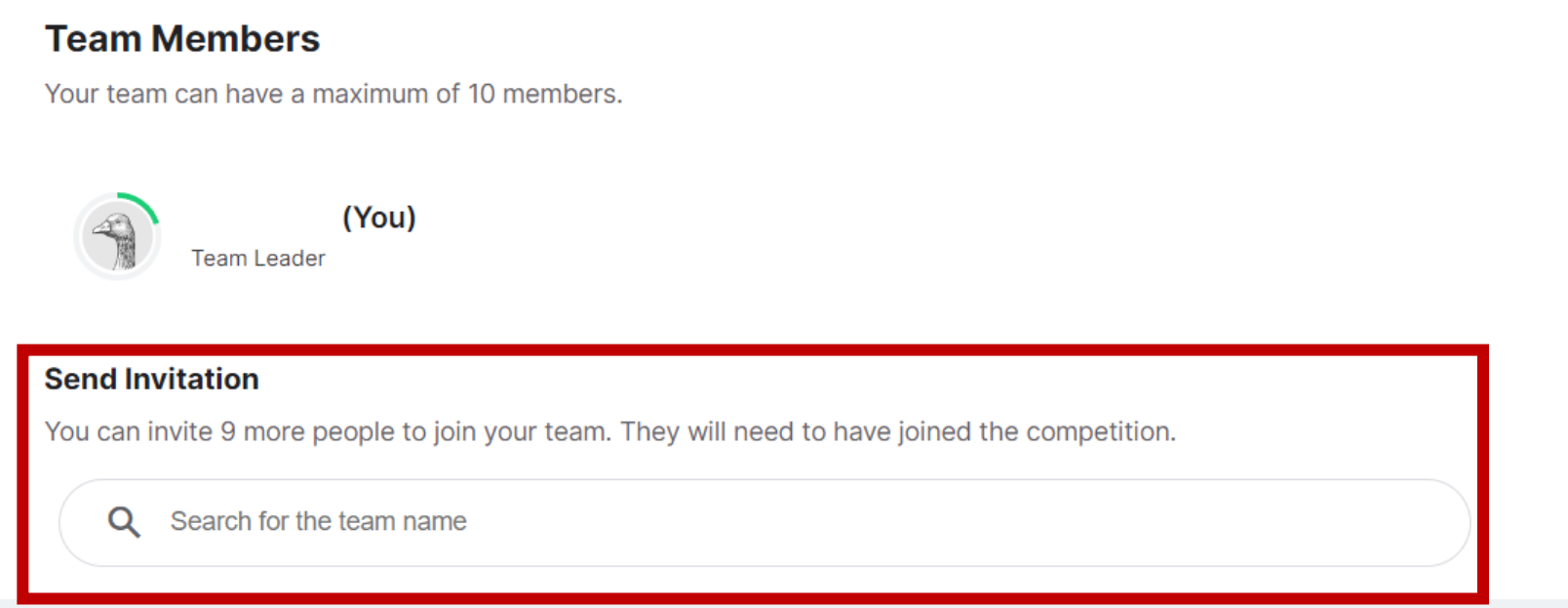
- See your team number on eeclass 小組專區



Team up (Important!!)

Invite your teammate(s)

- If you participate in this project by yourself, you don't need to do this step
- Search your teammates by their Kaggle account email
  - They must be added to the contest and "accept the rules" first.



Team up (Important!!)

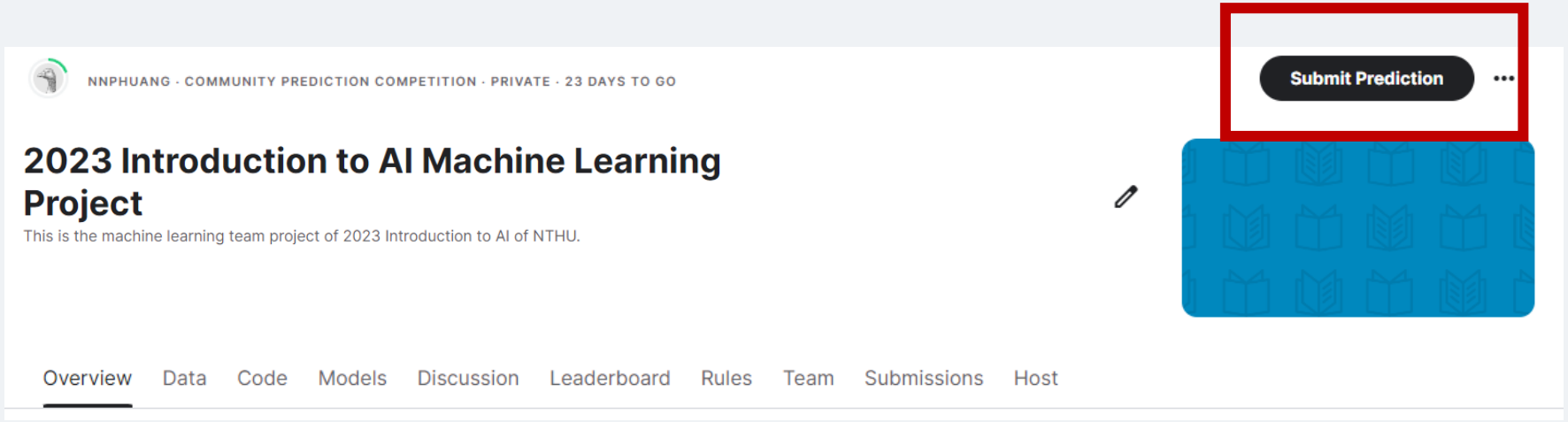
Anyone who doesn't follow these regulations will be revoked from the contest!

You need to do these before you make any submissions.

Kaggle submission

Public leaderboard submission

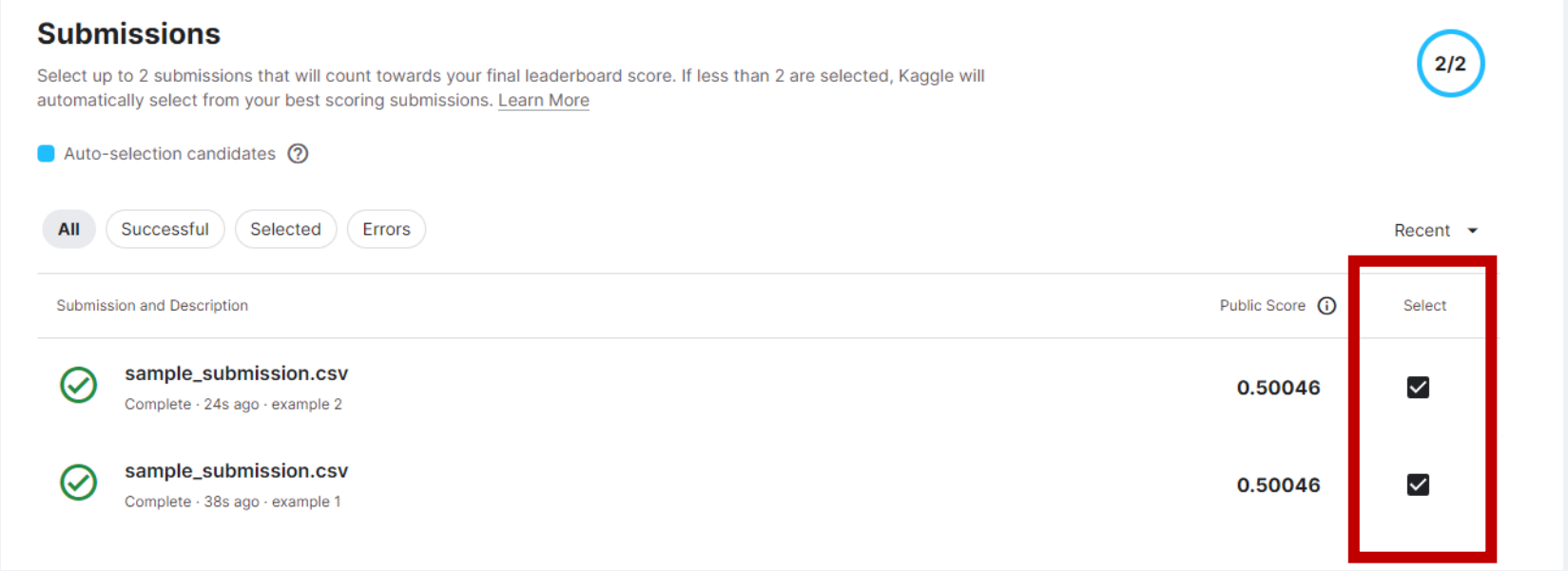
- You can submit up to 5 times per day as a team
- Your submission file must follow the format of sample_submission.csv



Kaggle submission

Private leaderboard submission

- Select 2 submissions as the private leaderboard submission
- The private leaderboard ranking will be released after the competition ends



Report

Report

Things you should write in your report

- Explain your methods
- Describe how you improve the results

Format

- Save as a pdf file

Grading

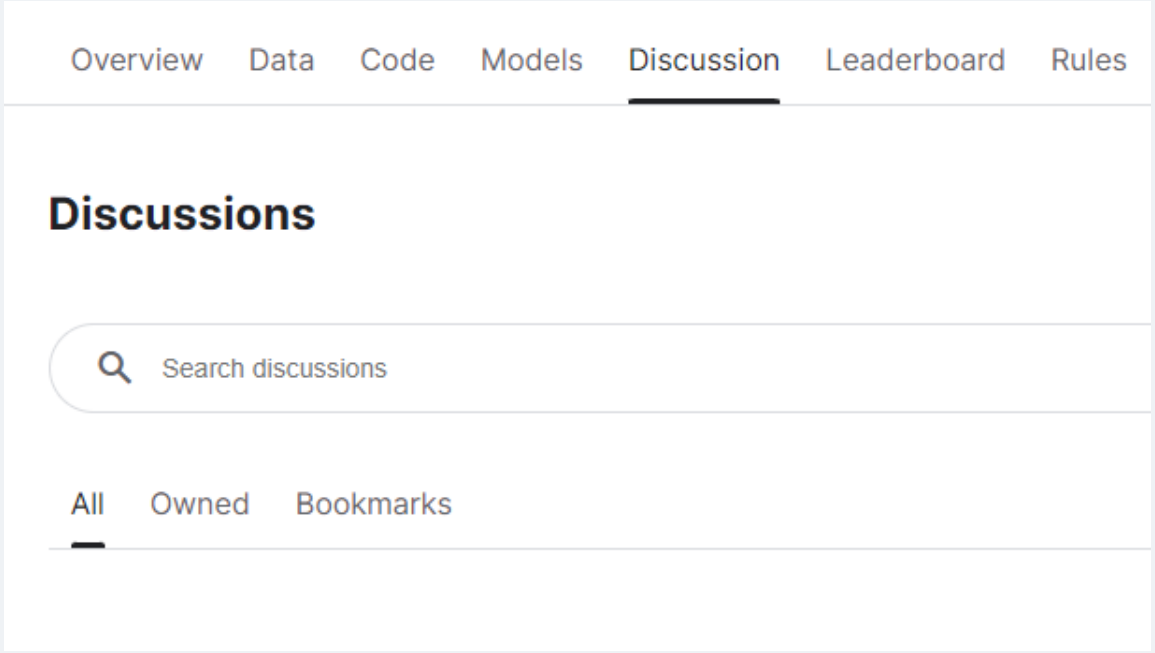
Evaluation

- Report 60%
- Performance 60%
  - Private leader board F1 score 30%
  - Private leader board rank 30%
- Cooperation 30%
  - Teammates' evaluation by Google form. (The form will be announced later)
- No need to be stressed out about your grade ☺
  - If you aren't satisfied with your performance on the private leaderboard, you can still make up by writing an informative report.

Contact us if you have problems…

Kaggle discussion page

- We strongly encourage you to post your question here
- You can also search the competition forums in case a similar question has been answered

Email

- Questions regarding the dataset
  - 郭宥成 nelson891224@gmail.com
- Questions regarding Kaggle
  - 黃恩平 huangenping0512@gmail.com