Introduction to Natural Language Processing

HW3

TA : HISU-HUNG LEE, WEI-LING HSU
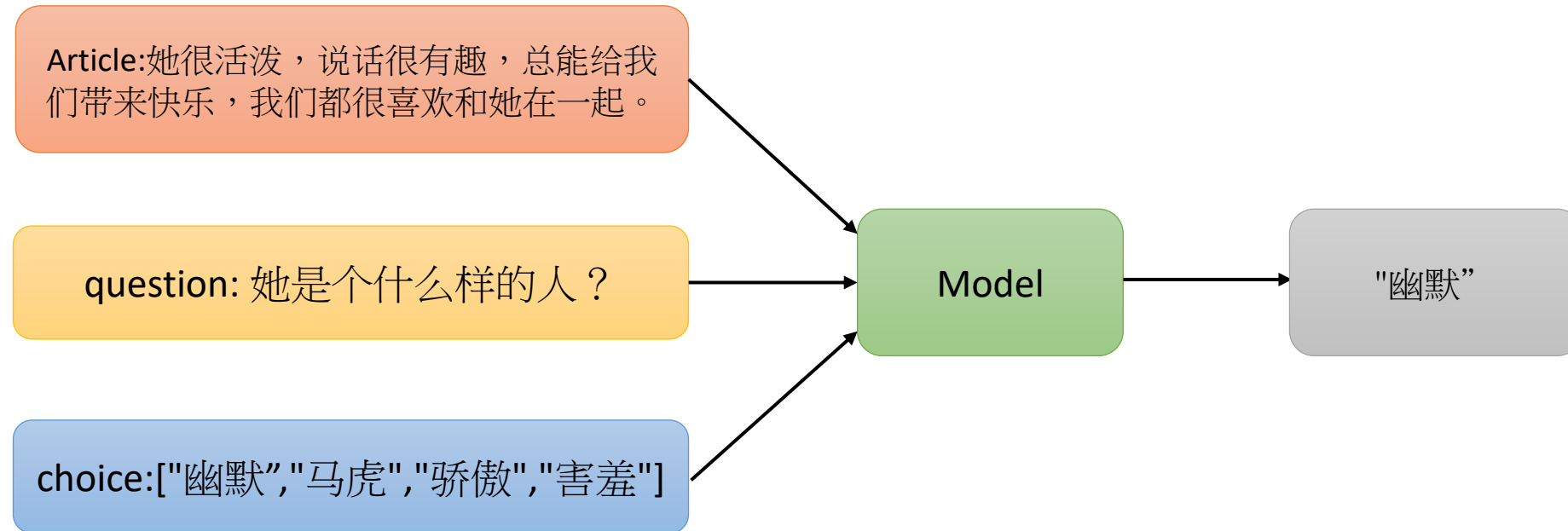
buddy0960391119@gmail.com

weiling.hsu.cs11@nycu.edu.tw

Task introduction

- **Machine Reading Comprehension** with neural network

 Given an article, a question and a set of choices, determine which choice is the answer.

Article:她很活泼，说话很有趣，总能给我们带来快乐，我们都很喜欢和她在一起。

question: 她是个什么样的人？

choice:["幽默","马虎","骄傲","害羞"]

Model

"幽默"

Task introduction

- You can use any external package you want, e.g. Pytorch, Keras…
- Any method is allowed in hw3.
- Requirement:

    1.Submit a report and your source code to E3

    2.Upload your submission to Kaggle

Dataset

https://drive.google.com/drive/folders/1XhtvX2_2Q3qPmWXwc3v_F3DfYUsraHEE?usp=sharing

You can get the dataset from the link or E3,

it contains : < train_HW3dataset.json >

< dev_ HW3dataset.json >

< test_ HW3dataset.json >

You can use < train_ HW3dataset > and < dev_ HW3dataset > to tune your model,

then predict the answers in < test_ HW3dataset >

Output

- Your model needs to predict which choice is correct in
  < test _HW3dataset.csv >
- For example:

1."question": "动物的器官感觉与人的相比有什么不同?",
  "choice": ["没有人的灵敏", "和人的差不多", "和人的一样好", "比人的灵敏"],
  "answer": "比人的灵敏"
  output: 4

2. "question": "他喜欢骑自行车。",
  "choice": ["正确","错误"],
  "answer": "正确"
  output:1

- Write your prediction in a CSV file and upload it to Kaggle.

index	answer
0	1
1	1
2	1
3	1
4	1
5	1
6	1

Kaggle submission(50%)

- Kaggle link: https://www.kaggle.com/t/b7ab308a6e6a411ebbbbeba0a2e40189

- Display team name:<student ID> (important！)

- Submission format:
  - A 2003 X 2 CSV file, the first row is for the column name, and the last 2002 rows are for your result.
  - Column names should be "index" and "answer".
  - The result should be the index and predicted answer(1~4), please make sure the order of your result is right!

- There will be one baseline.

Kaggle submission(50%)

- You can submit at most 5 times each day.

- You can choose 2 of the submissions to be considered for the private leaderboard, or will otherwise default to the best public scoring submissions.

Report Submission(50%)

- Submit a report containing 3 questions:

    1. Describe all the methods you have implemented. (60%)

    2. Did you preprocess your data from the dataset? Why? And how?
    (Did you encounter the problem that the input length is longer than the maximum sequence length of the model you use? How did you solve this problem?) (30%)

    3. What difficulties did you encounter in this assignment? How did you solve it? (10%)

    Please answer the questions in detail!

E3 Submission

- Deadline:
    - Submit Zip to E3 before 1/ 5 11:59 PM
    - No Late Submission!
- Format:
    - Source code : Hw3_<student ID>.py
    - Report file : Hw3_<student ID>.pdf
    - Zip file : Hw3_<Student ID>.zip

Grading policy

- Kaggle(50%):

  - Basic score (according to the public leaderboard):

    under baseline: 40

    over baseline: 60

  - Ranking score (according to the private leaderboard) :

    score= 40-(40/N)*(ranking-1), N=number of people

- Source code and Report(50%):

  - the more detail you make, the higher score you get

Timeline

Hw3 release

Deadline

12/8

1/5
23:59

If you have any question about HW3, please feel free to contact with TA:HSIU_HUNG LEE or WEI-LING HSU, through email buddy0960391119@gmail.com or weiling.hsu.cs11@nycu.edu.tw