Foundations of Natural Language Processing

Peking University, 2024

Lab 4
For Graduating Students: Due Sunday, June 30, 2024 by 23:59
For Other Students: Due Sunday, July 14, 2024 by 23:59

1. Directions

PLEASE read the instructions to ensure you receive full credit on your homework.

LATE SUBMISSION POLICY (Not applicable to graduating students)

• Late homework will have 5% deducted from the final grade for each day late.

• <u>No submission will be accepted after July 21</u>, one week after the due date. Your homework submission time will be based on the time of your last submission to Course.

Therefore, do not re-submit after midnight on the due date unless you are confident that the new submission is significantly better to overcompensate for the credits lost. You can resubmit as much as you like, but each time you resubmit be sure to upload all files you want to be graded! Submission time is non-negotiable and will be based on the time you submitted your last file to Course. The number of points deducted will be rounded to the nearest integer.

2. General Descriptions

In this lab, you are going to complete **one of the following** NLP projects according to your interest. See the detailed instructions for each project in their respective PDFs.

<u>Project 1:</u> Low-Resource Translation

<u>Project 2:</u> Dish Name Recognition

<u>Project 3:</u> Open-Domain Fact Verification

In general, you should follow the following steps:
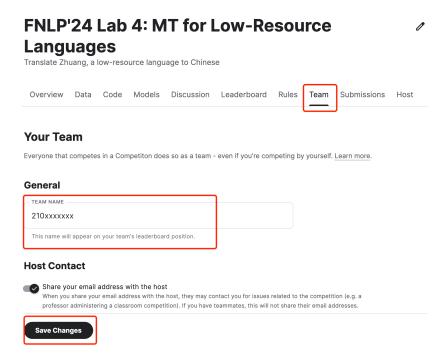
Step 1: Choose your project

First, choose one of the three projects list above and join the Kaggle leaderboard for this task by the following links.

Project 1: https://www.kaggle.com/t/9e936b029ff6438fa68b23f6d4b54c94

Project 2: https://www.kaggle.com/t/601026e0eaff4aab899638a490223a3d

Project 3: https://www.kaggle.com/t/5b984bbb62534fc483f528bfec7e2e5a

Please change your Team Name to your student ID. Otherwise, we cannot identify your scores on the leaderboard and you will get zero grade.

You can change your choice of project before the deadline as you wish. However, make sure that you clearly indicate your final choice in your report. If you write about more than one project in your report, we will only grade the first one and ignore the rest.

Step 2: Carry out your project

You should carefully read the instructions of your chosen project. Here is some general information that you should be informed of.

Submission Trials:

Please actively use our leaderboard system to evaluate your methods. **Note that you can only submit your prediction results up to five times every 24 hours.** Do not delay your project to the last few days!

LLM APIs:

According to each project description, you might conduct some experiments with LLMs. **To ensure a fair comparison between students, please use the API from Deepseek** (https://platform.deepseek.com/api-docs/zh-cn/) **for your experiments with LLM APIs.**

We will provide each student with a fund of 10 RMB. It can cover a usage limit of 10 million tokens, which is sufficient for this project. The funds will be handed out by the professor in the class. You should create your own API account and recharge it with the fund by yourself.

Step 3: Submit your project

Submission Deadline:

By July 14, 2024, you should submit all the material required in the instructions to course.pku.edu.cn.

For graduating students, your deadline is earlier (June 30). Don't miss the deadline, or it would affect your graduation. **The late submission policy does not apply to the graduating students.**

Grading Criterion:

We will grade your project based on the following criterion:

1. Report (80%): In the report, you should

a) describe your method clearly;

b) report the implementation details of your method;

c) evaluate your method on the leaderboard;

d) compare your method with a suitable baseline (such as a simple zero-shot baseline for LLMs) to evidence the effectiveness of your method;

e) analyze the experiment results in depth.

2. Implementation (5%): We will check whether you have submitted all the required material. The submitted code should meet the requirement of reproducibility.

3. Leaderboard (15%): Part of the grade will be determined by your ranking on the leaderboard.

3. Contact

If you have any general questions about Lab 4, please contact zhangch@pku.edu.cn.

If you have questions for a specific project, please contact the following TAs:

Project 1: zhangch@pku.edu.cn

Project 2: lxlisa@pku.edu.cn

Project 3: ziruiwu@pku.edu.cn

Good luck with your final project!