
Project Proposal

Machine Learning, 2023 Fall

Due: 2023-11-1 11.59 PM. You should submit a **PDF** format proposal to **Tsinghua Web Learning** written by **TEX** in **NeurIPS conference paper format** (which is in line with the final report's requirement. Visit NeurIPS official website^a or Overleaf^b for templates). The proposal should be no longer than **2 pages (excluding references)** and every team (2-3 students) only needs to submit one proposal.

^a<https://nips.cc/Conferences/2020/PaperInformation/StyleFiles>

^b<https://www.overleaf.com/latex/templates/neurips-2022/kxymzbjpsqx>

1 Requirements

In this assignment, you should propose a research proposal based on Machine Learning. You need to develop new machine learning methods for established problems or newly defined topics. Specifically, the following suggestions may be helpful to your proposal writing:

- **Background:** What's the background of the problem? Is it theory-driven, or deeply rooted in some useful application situations? Is it important or necessary? What impact will it bring if you finally solve it?
- **Definition:** Is there any formal or mathematical definition for your problem? Explain the symbols you may want to use in the proposal writing.
- **Related Work:** Is your problem a well-established one? If so, review existing approaches and discuss their advantages and disadvantages; if not, survey and describe related problems in the field and list some potentially applicable baseline methods. Remember to cite related work properly.
- **Proposed Method:** What are the motivations for you to choose it? Which datasets do you propose to experiment on? What baseline approaches do you plan to compare with? How do you implement your proposed method based on the dataset? It is ok to change and improve it later but now try to describe it as detailed as possible.

Additionally, the proposal PDF should include your team member's name(s) and student ID(s). **Plagiarism is strictly prohibited**, or you will fail the course.

2 Examples and Resources

There are some potentially good topics and related sources for your references, and you can also

- **Kaggle or Other Public Competition:** Kaggle provides many datasets and established baselines, while their quality could be quite different from each other. The competition track in NeurIPS 2022¹ could be another good but challenging option.
- **Journal or Conference Publications:** Find some recently published ML papers (e.g., from ICML/NeurIPS/ICLR/KDD/ACL/CVPR) to follow. They usually provide detailed datasets and baselines results, but remember that mere reproducing is not quite enough for our final project. Try to make some improvements to the paper you want to follow.
- **Large Language Model (LLM) Applications:** Recently, the community has witnessed the rise of LLMs such as ChatGPT. Different from conventional end-to-end supervised learning, most APPs^{2 3} created by LLMs are API-based zero-shot or few-shot learning, without the need of local GPUs. We provide a certain amount of free API budget for an LLM—ChatGLM⁴—a bilingual chat LLM from zhipu.ai. If you would like to have a try, please register your personal account in <https://open.bigmodel.cn/>. We will provide one billion free tokens and you have to provide your username in the questionnaire to get the free budget https://docs.google.com/forms/d/e/1FAIpQLSfG0sZlGyBopDCv4GMurCK9w19vy5zn5Zq8BzR0zIMgi0pu1w/viewform?usp=pp_url. You can also try to develop other LLM applications or improve LLMs, e.g., acceleration, and finetuning, if you have enough resources.

¹<https://neurips.cc/Conferences/2023/CompetitionTrack>

²<https://beta.openai.com/examples/>

³<https://www.langchain.com/>

⁴<https://chatglm.cn/>