Designing a Course Unit on Large Language Models (LLMs)

CSE598 Master Project (Advisor: Marion Neumann)

Ziwei Wang
Washington University in St. Louis
ziwei.wang@wustl.edu

Abstract

This project focuses on the development of a comprehensive course unit dedicated to Large Language Models (LLMs) for advanced undergraduate students. Recognizing the pivotal role of LLMs in AI, the course aims to impart a fundamental understanding of LLMs and provide hands-on experience with their applications. The course design is tailored to fit within the constraints of lecture volume and prerequisite knowledge. It includes lectures, written assignments, and a project involving Pytorch, facilitating an intuitive understanding of LLMs. Future enhancements include identifying an optimal prerequisite for student preparation and addressing the practicality of the project format to the availability of computational resources. This is the course note link: course note link. This is the project code link.

1 Introduction

Large language models (LLMs) refer to Transformer language models that contain hundreds of billions (or more) of parameters4, which are trained on massive text data, such as GPT-3, PaLM, Galactica, and LLaMA. (Zhao et al., 2023) Housing hundreds of billions of parameters, these models are trained on vast text data, empowering them to comprehend natural language and tackle intricate tasks via text generation. Their proficiency in understanding and generating human-like text has made them indispensable in numerous contemporary AI applications. Acknowledging their escalating importance, universities globally are progressively integrating LLMs into their academic programs.

This project endeavors to construct a course unit centered around LLMs for advanced undergraduate students. The course will encompass lecture notes, homework tasks, project code, and other supplementary materials.

2 Related Work

Universities adopt varied strategies when incorporating LLMs into their curriculum, contingent on the course objectives. The first category of courses offers a comprehensive introduction to language models, with LLMs forming a part of the curriculum. An example is Stanford University's CS324 *Natural Language Processing with Deep Learning* course, which devotes substantial time to elucidating the theoretical underpinnings and development process of language models. LLMs are introduced as a distinct unit later in the course. The second category of courses zeroes in exclusively on LLMs. Princeton University's COS 597G *Understanding Large Language Models* course exemplifies this approach. This course familiarizes students with various typical LLMs and delves into cutting-edge topics in the LLM field, such as in-context learning.

3 Course Design

The primary objective of this course is to provide students with a brief understanding of the main fundamental theory behind LLMs and initial exposure to practical applications to gain an intuitive understanding of LLMs according to the volume of 2 to 3 lectures of the unit and the limited prerequisite. The course content is selected based on a review of similar courses offered by peer universities and surveys conducted to identify the most relevant and useful topics.

3.1 Teaching and learning activities

The first lecture will cover the Language Model section and half of the Transformer section. The second lecture will cover the rest of the Transformer section, and the Large Language Model section.

Time	00:00-00:10
Objects	Introduction of the unit
Goals	Students will have a general understanding of the overall content of the unit
Activities	Interactive poll
Time	00:10-00:20
Objects	Language model
Goals	Students will understand the task of the language model, and know what the n-gram model is.
Activities	Animation, diagrams
Time	00:20-00:45
Objects	The neural language model
Goals	Students will understand how the neural language model works and what the advantage of this model
Activities	Animation, diagrams
Time	00:45-00:50
Objects	Evaluation of language model
Goals	Students will master how to use perplexity to evaluate language model
Time	00:50-01:00
Objects	Word representation
Goals	Students will know the principles of representing words, and what is word2vec
Time	01:00-01:20
Objects	RNN and LSTM
Goals	Students will understand how RNN and LSTM work, the difference between them, and their improvement from the neural language model
Activities	Diagrams

Table 1: Lecture 1

Time	00:00-00:10
Objects	Review
Goals	Students will recall the fundamental knowledge discussed in the last lecture and grasp the main development routes of language models
Time	00:10-00:40
Objects	Transformer
Goals	Students will understand what is the attention mechanism, key-query-values, why we need it, how the Transformer works, what is multi-head self-attention, and some different structures of the Transformer.
Activities	Animation, diagrams
Time	00:40-00:55
Objects	The pretraining-finetuning paradigm
Goals	Students will understand what this paradigm is by learning a specific application example. In the example of machine translation, what is the base model, what is the pre-training task and dataset, what is the fine-tuning task and dataset. They will also understand why it is the foundation of the large language model, and what are the advantages of it.
Time	00:55-01:20
Objects	Large language model
Goals	Students will understand the model size, data size and amount of computing these factors affect the models' performance. They will also understand the emergent abilities along with the increase of these factors.
Activities	Diagrams

Table 2: Lecture 2

3.2 Homeworks and Assessment

Report the quantitative results you have found. To compare results and compare against baselines, use a table or plot.

- **Written assignments**

  Several written questions are designed to help the students understand the properties of neural networks. The content includes cross entropy and perplexity, vanishing gradient and gradient explosion problems, and computing a simple neural network example

- **Project:**

  For the project, students will be provided with the project code. They will be required to read and run the code to do some experiments, and then observe the results. This will give them a preliminary hands-on experience working on LLMs by Pytorch. They will achieve a more intuitive understanding of the LLMs properties taught in the lecture.

  Task 1: Translate the text from French to English with different size models. Compare the results of using the small-size model T5-small with 60 Million parameters, with the results of using the medium-size model T5-3b with 3 Billion parameters.

  Task 2: Fine-tuning T5-small with the dataset from French to English. Compare the results of using the model T5-small without fine-tuning, with the results of the fine-tuned model.

4 Future Work

The current structure of the course unit has been established, but some areas need enhancement. A key concern is ensuring students possess the foundational knowledge required to grasp the course content. Ideally, students should have prior exposure to deep learning before undertaking this unit. However, the current course structure in the CS department does not include a course specifically dedicated to deep learning. Given this, it's crucial to identify an optimal prerequisite to prepare students adequately without imposing excessive requirements. Secondly, we must consider the practicality of the project format and the availability of computational resources. Training a large language model requires substantial computational power, which comes at a cost. To enable students to delve deeply into the process of training large language models at minimal expense, we need to devise an appropriate method for assigning project homework or demonstrating the project in class. Alternatively, we may need to source computational resources from the school or other institutions.

References

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A survey of large language models.