

基于大语言模型的课程教学答疑系统设计

陆楚龙, 蔡国永, 肖微, 闭国欣

桂林电子科技大学, 计算机与信息安全学院, 广西 桂林

收稿日期: 2024年4月26日; 录用日期: 2024年5月24日; 发布日期: 2024年5月31日

摘要

在混合式教学中, 如何能对学生学习过程中的疑问进行及时回答对教师来说是一个比较头痛的问题。借助AI技术, 开发能自动回复学生问题的教学辅助系统是一个非常重要手段。本文介绍一个能协助教师, 帮助学生解决课程学习问题的会话系统ChatQA。该系统通过深度学习模型, 能够与学生进行自然、连贯的对话, 回答学生课程学习中的问题, 有效缓解了教师的答疑负担。论文详细讨论了该系统的设计原理、Prompt优化方法等, 并通过软件工程课程的问题进行了测试, 显示了该方法的可用性, 具有较好的自动答疑的效果。

关键词

大语言模型, 生成式对话, 模型优化, 课程答疑

Design of Course Teaching Q&A System Based on Large Language Model

Chulong Lu, Guoyong Cai, Wei Xiao, Guoxin Bi

School of Computer Science and Information Security, Guilin University of Electronic Science and Technology, Guilin Guangxi

Received: Apr. 26th, 2024; accepted: May 24th, 2024; published: May 31st, 2024

Abstract

In blended teaching, how to answer students' questions in the learning process in a timely manner is a headache for teachers. With the help of AI technology, it is very important to develop a teaching assistance system that can automatically respond to students' questions. This article introduces ChatQA, a conversational system that can help teachers and students solve course learning problems. Through the deep learning model, the system can have a natural and coherent dialogue with students, answer students' questions in course learning, and effectively alleviate the burden

of answering questions on teachers. The design principle and prompt optimization method of the system are discussed in detail, and the questions of the software engineering course are tested, which shows the usability of the method and has a good effect on automatic Q&A.

Keywords

Large Language Models, Generative Dialogue, Model Optimization, Course Q&A

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在现代教育领域，学生常常面临着各种各样的问题和困惑。尤其是在课程学习过程中，学生往往会遇到一些难以理解的概念、深奥的理论或者复杂的问题。为了解决学生的疑惑，教师通常提供答疑服务，以便学生能够更好地理解和掌握知识。然而，无论是线上教育或者线下教育，目前在教学过程中它们都存在一个共同的问题，即不能及时处理学生存在的课后遗留问题。现在的教学模式普遍都是一对多的形式，然而学生自身水平存在差异，在接收知识点时接受程度不一样，并且由于时间、地点、资源等因素，教师在面对众多学生提出的问题时，很难实施个性化教学，不能及时给出回应，所以在教育行业中，及时进行课后答疑一直都是难点。

随着技术的发展，人们总希望能用最短的时间获得最简洁的答案，通过搜索引擎搜索问题时得到的大多是与问题相关的网页，仍需要花费大量时间筛选正确答案；当下大部分能实现课程答疑任务的软件大多是基于传统的搜索方式，得到的结果大部分时候是不尽如人意的；相比之下，深度学习技术实现的语言模型可以生成与问题相关的自然且连贯的对话，模拟人类的回答过程。

同时，大语言模型逐渐成为自然语言生成领域的基本解决方法，模型参数数量越来越多，无论是国外 OpenAI 的 GPT3，或是国内百度的 Ernie3.5，清华的 GLM3.5 等等，如今流行的大语言模型的参数量级已经在亿以上，大语言模型的微调越来越困难。本文尝试将“轻量化”的高效参数微调思想[1]应用于对话系统的应用中，通过大语言模型，实现一套基于深度学习 NLP 算法的会话系统 ChatQA，进而降低因为答疑而造成的老师和同学的时间花销，减少老师工作量。

2. 相关工作

大量研究表明，无论如何对文本输入的形式进行优化，搜索式课程答疑软件在实际应用中是很难行得通的。李春明[2]讨论过基于全文检索的技术和关键词匹配的方法构建答疑系统，但以此类方法实现的答疑系统的答案准确度和智能性都与现实中答疑要求的标准存在较大的差距。而后，李印鹏[3]借助早期 NLP 技术，根据词形相似度、语句相似度、词序相似度三个指标来对学生输入的问题进行语句相似度的匹配，最后根据知识数据库搜索的方式完成答疑任务，但以此类方法构建的答疑系统仍然无法具备人类的思维能力，一旦知识库中缺少与问题匹配的相关内容，这种答疑系统无法对用户提出的疑问给出满意的答案。

近几年，学者和工程师们把目光转向了深度学习领域，开始讨论通过深度学习技术实现课程答疑任务的可行性。相较于传统的搜索式课程答疑软件，基于深度学习技术构建知识问答系统实现课程答疑任

务更具优势。随着大语言模型技术的兴起，深度学习模型对上下文具有更强大的理解和推理能力，能够理解复杂的问题，并且在回答中考虑相关的上下文信息，这使得知识问答系统能够更准确地回答复杂问题，而不仅仅局限于简单的关键词匹配。Luo 等提出大模型的生成和检索知识图谱相结合的问答框架，利用微调开源大模型进行自然语言问题到逻辑形式的转换，借助无监督实体关系检索生成图数据库检索语言，以生成再检索代替检索再生成的方式，解决了检索效率低、检索误导生成等痛点，并在 KBQA (知识库问答)任务上达到了最新的 SOTA 性能水平。同时，陈静远等[4]提出，预训练-提示(Pre-train-Prompt)这种全新的深度学习范式为人工智能在教育领域的应用提供了一种新的思路和方法。

3. ChatQA 系统方案设计

3.1. 系统结构与数据流向

ChatQA 是一个使用深度学习大语言模型来生成问题回复的系统，可以实现用户在课程答疑任务上的需求。用户在输入问题后，通过数据预处理技术将用户问题转换成问题向量，而后传入到系统使用的生成式大语言模型中，经过模型处理后得到模型生成的关联回答列表，将列表中的回答经过自然语言生成技术得到用户问题的答案。本文讨论将通过模型的优化和前后端开发实现 ChatQA 系统，ChatQA 系统大致由系统前端，系统后端业务以及深度学习模型三大部分组成，系统前端接收用户输入的问题，将文本字符串数据传入系统后端，由系统后端将文本字符串进行预处理形成问题向量后输入深度学习语言模型，语言模型根据问题向量生成关联问题列表后再次将其传入系统后端，选择合适的问题回复作为答案并进行二次处理，最终形成文本字符串并将其返回前端实现与用户进行交互，实现一次问题答疑流程，如图 1 所示。

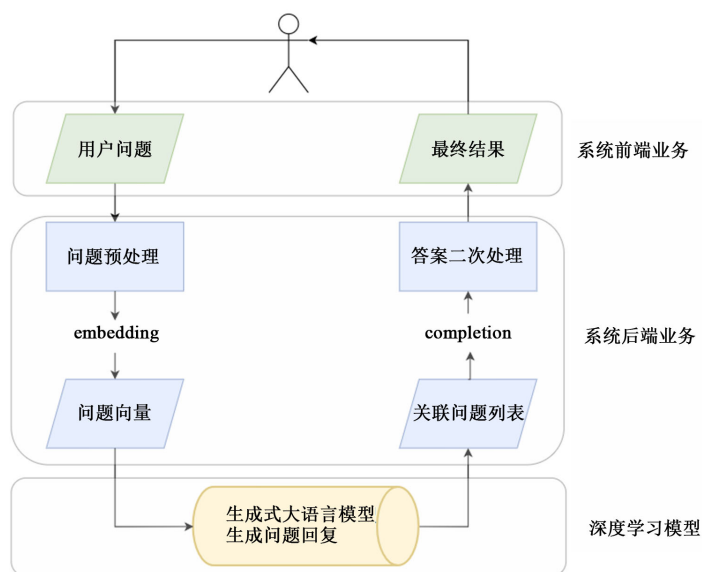


Figure 1. Overall structure and data flow illustration of ChatQA

图 1. ChatQA 系统整体结构与数据流向示意

3.2. 模型优化技术

深度学习生成式语言大模型中，“预训练 + 微调”范式[5]是经典的模型领域适配技术，可有效将预训练模型的丰富知识迁移到新的任务中。然而随着深度神经网络的规模不断增大，生成式大模型的参数规模往往动辄数十亿，而受硬件设备限制在新任务中使用常规的“全量微调”模型优化技术不再具有

可行性。因此，本文使用提示词微调法(Prompt Tuning)来适配课程答疑的任务，其主要的实现方式是在深度学习模型输入的词嵌入向量前端拼接一个针对某一特定任务的、连续的、可学习的张量 \mathbf{P} ，在训练中，只通过梯度下降法在下游任务上优化可学习张量 \mathbf{P} ，而冻结预训练模型的参数，从而更“节省”参数，示意如图 2。

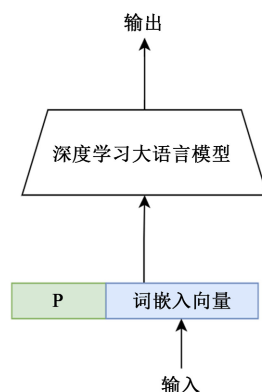


Figure 2. Prompt tuning illustration
图 2. 提示词微调法示意图

对于一条优质的 Prompt，Prompt = 任务 + 生成主体 + 细节(可选) + 形式(可选)。在对用户输入数据进行清洗和处理后，基于以上标准，根据课程答疑的场景，采取强化突出、迭代询问的策略分类设置训练集，帮助语言模型更好的适应在课程答疑的场景下生成对应优质 Prompt 的任务。利用 GlobalPointer 设计将用户输入的首尾视为一个整体去进行判别，利用全局归一化的思路来进行命名实体识别，以此实现关键实体抽取并优化，生成 Prompt 帮助大语言模型更好的实现课程答疑任务。

3.3. 文本生成控制

在大语言模型的文本生成中，Temperature 参数和 Top-p 参数是决定其生成文本创造性和多样性的重要参数。

Temperature 参数控制生成语言模型中生成文本的随机性和创造性，调整模型 Softmax 输出层中预测值的概率。人工智能论坛 Cohere 中有文章[6]解释，Temperature 值越大，预测词的概率方差减小，即很多词被选择的可能性增大，有利于文本多样化。当 Temperature 值为 0 时，其限制大语言模型必须使用可能性最高的词，这样会导致语言模型对同样的问题每次输出的结果都是相同的。随着 Temperature 值的提高，该限制会随之放宽，模型将被允许使用可能性更低的词语。

该文章通过实验，输入“The sky is ”测试大语言模型在不同 Temperature 的值时生成下文 S：

Temperature = 0，模型每次都会给出相同的答案——S = “the limit with this one”；

Temperature = 0.5，模型的输出更多样化，但仍基本符合规律——S = “the limit”/“blue”/“overcast”；

Temperature = 1，模型的输出开始不受限制——S = “not the limit” / “almost perfectly blue” / “grey and dreary today”；

当 Temperature 设置为最高值时，模型的输出完全不受限制——S = “clear, the water smooth, and it’s an unimaginably long way to go before the dolphins decide to give up their vertical quest.”

Top-p 参数定义了生成文本时模型考虑的概率分布范围，模型会考虑概率值在 Top-p 的值内的单词的结果。模型将单词按概率排序，并选择最低概率总和超过给定阈值的最小集合来控制生成的多样性，这

意味着在每个时间步选择的词汇仅限于整体概率分布中最可能的一部分。JarodYv [7]在松山湖开发者社区中发文,以“我喜欢漂亮的___”为例介绍 Top-p 参数对于大语言模型输出的采样情况,当 Top-p = 0.9 时,大语言模型对输出的采样结果如图 3 所示。

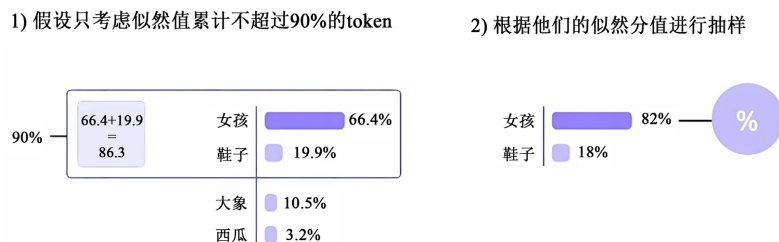


Figure 3. Top-p dynamic candidate output selection

图 3. Top-p 动态选择候补输出

本文通过大语言模型来实现课程答疑任务,需要优秀的准确性和可靠性,对大语言模型进行 PromptTuning 时,将 Temperature 值与 Top-p 值控制在 0.5 之内以获得更精确和可控的文本生成。

4. 系统实现与测试

4.1. 开发环境

本文采用搭载 Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz 2.59 GHz 处理器, NVIDIA GTX 1650 GPU 的计算机作为硬件实现设备,基线大语言模型选择 Eenie 3.5,模型 Prompt 参数 Temperature = 0.5, Top-p = 0.5。

4.2. 系统答疑测试

Q1. 基于数据流的建模如何实现?

Q2. 软件工程的三要素是什么?

4.2.1. 未经过优化的大语言模型答疑测试

以软件工程课程为例,本文设置 Q1、Q2 两个问题。通过对比 Ernie3.5 与 ChatQA 的问题答复,验证深度学习技术与课程答疑相结合的可行性。图 4、图 5 给出了 Ernie3.5 对 Q1 与 Q2 的回复。



Figure 4. The initial response of Ernie3.5 to Q1

图 4. Ernie3.5 对 Q1 的回复

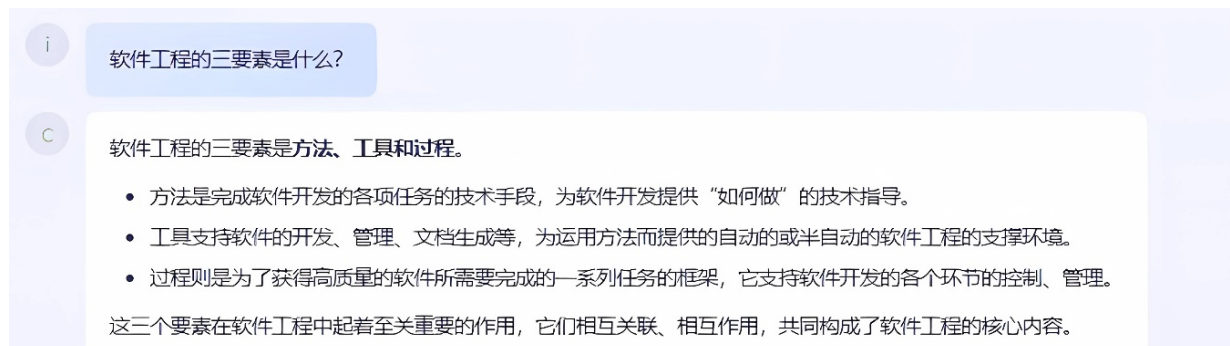


Figure 5. The initial response of Ernie3.5 to Q2

图 5. Ernie3.5 对 Q2 的回复

4.2.2. 经过模型优化的 ChatQA 答疑测试

参照本文 3.2 节提出的 Prompt 优化模板，以软件工程课程为例，本文对大语言模型进行问题回复前的 Prompt 设置为——“作为软件工程课程老师[x]，你需要详细解释软件工程中的关键概念、原理和实践，并提供实际案例以帮助你深入理解和掌握[y]。”明确了生成主体 x + 任务 y，ChatQA 使用 Prompt 优化后的 Ernie3.5 作为内置生成语言模型，图 6、图 7 给出 ChatQA 对 Q1、Q2 的答复。



Figure 6. The response of ChatQA to Q1 after prompt tuning

图 6. Prompt 优化后 ChatQA 对 Q1 的回复

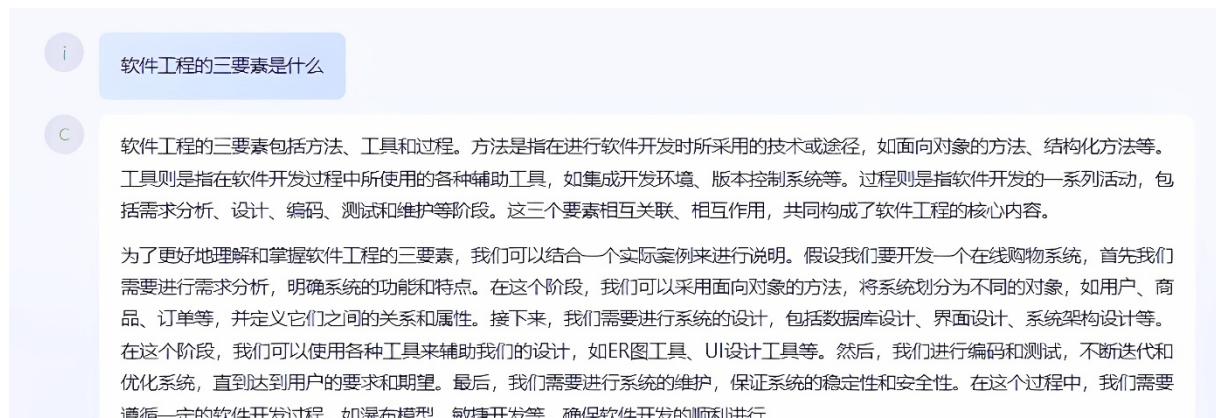


Figure 7. The response of ChatQA to Q1 after prompt tuning

图 7. Prompt 优化后 ChatQA 对 Q2 的回复

从图 6、图 7 与图 4、图 5 的对比观察中可以发现，与基线模型相比，通过本文提出的 Prompt 方法进行模型优化处理后，课程答疑任务的实现结果可以得到较为明显的提升，不仅能实现普遍的课程答疑任务，还进一步能通过大语言模型的学习能力举一反三地对相关问题进行解答。利用深度学习技术，能大幅降低因为答疑而造成的老师和同学的时间花销，减少老师工作量，成功验证深度学习技术与课程答疑相结合是可行的。

5. 结语

本文从系统结构，大语言模型的优化技术以及文本生成控制讨论了通过深度学习技术和大语言模型实现课程答疑任务的对话生成系统的开发方案，并以此方案开发了一套软件，取得了一定的成果，但仍然存在一些挑战和改进空间。例如，可以通过实体抽取和语义增强等技术对 Prompt 进行优化以提高模型在课程答疑任务上对话生成的准确性，系统的交互界面和用户体验也可以进一步优化。通过本文的研究可以看到 AI 技术于教育领域的巨大潜力，我们相信以大语言模型和小量参数的模型优化相结合的方法将在教育领域发挥越来越重要的作用。

基金项目

2022 年区级大学生创新训练计划立项项目——基于深度学习的生成式课程答疑会话系统 (S202210595165)。

参考文献

- [1] 丁鑫, 邹荣金, 潘志庚. 基于高效参数微调的生成式大模型领域适配技术[J]. 人工智能, 2023(4): 1-9.
- [2] 李春明. 基于全文检索的自动答疑系统的研究与设计[D]: [硕士学位论文]. 苏州: 苏州大学, 2005.
- [3] 李印鹏. 教学网站智能答疑系统设计及实现[D]: [硕士学位论文]. 石家庄: 河北科技大学, 2018.
- [4] 陈静远, 胡丽雅, 吴飞. ChatGPT/生成式人工智能促进以知识点为核心的教学模式变革研究[J]. 华东师范大学学报: 教育科学版, 2023, 41(7): 177-186.
- [5] 吴双. AI 大模型: 以“大规模预训练 + 微调”范式满足多元化需求[N]. 人民邮电, 2022-06-16(5).
- [6] Cohere Team (2022) LLM Parameters Demystified: Getting the Best Outputs from Language AI. <https://txt.cohere.com/llm-parameters-best-outputs-language-ai/>
- [7] JarodYv. ChatGPT 模型采用算法详解[EB/OL]. <https://blog.csdn.net/jarodyv/article/details/128994176>, 2024-03-08.