

深度学习与自然语言处理报告

19376252 张一辉

tbgsection@qq.com

摘要

本文是深度学习与自然语言处理的第五次实验报告，实验内容为选取 3 个目前前沿语言大模型，通过提示工程的方法来检验和对比不同模型下游任务上的性能，选择 3~5 个不同的自然语言下游任务来进行测试。通过测试我们发现综合来说 ChatGPT 的性能是最佳的，但是同样会存在一些错误的可能，newbing 能够搜索互联网上的结果，给出资料来源，但由于互联网上并非信息都是正确的，因此也会存在一些错误，讯飞星火综合性能不如前者，但是也有其特点，支持语音输入。

引言

大语言模型(Large Language Model)是指使用深度学习技术构建的能够理解和生成自然语言的模型。这些模型通过在大规模文本数据上进行训练，学习到语言的结构、语法规则、上下文关系和语义含义等。大语言模型的目标是能够生成具有逻辑性、连贯性和语义准确性的文本回应，以实现与人类的对话交互。

最近几年，由于深度学习的快速发展和可用数据集的增加，出现了一些强大的大语言模型。其中最著名的是 GPT (Generative Pre-trained Transformer) 系列模型，如 GPT-3、GPT-2 和 GPT。这些模型使用了 Transformer 架构，能够处理更长的上下文，并具备出色的生成文本能力。

大语言模型可以用于各种自然语言处理任务，如文本生成、机器翻译、对话系统、摘要生成、问答系统等。它们也可以应用于信息检索、文档分类、情感分析等领域，为人们提供各种语言相关的服务和工具。

研究方法

提示工程

提示工程 (Prompt Engineering) 是指在使用大语言模型 (如 GPT) 时，设计和构造合适的提示或指令，以引导模型生成所需的输出。由于大语言模型是基于训练数据中的统计模式进行生成，合理的提示可以帮助模型更好地理解用户意图，生成更准确、相关和可靠的回应。

提示工程的关键在于如何设计提示文本，以使模型能够产生期望的输出。在提示工程中，通常有以下：

指定任务和格式：明确指定任务和所需的输出格式，以便模型能够准确理解用户的需求。例如，在问答任务中，可以明确指定问题的格式和所期望的回答类型。

上下文提供：将上下文信息作为提示的一部分提供给模型，使其能够更好地理解上下文和生成连贯的回应。这可以通过在提示中包含先前的对话历史、背景信息或先前的模型输出来实现。

控制生成风格：通过在提示中加入特定的词汇、语气或风格要求，来引导模型生成符合特定风格或表达方式的文本。

影响输出长度：通过在提示中包含限制或鼓励模型生成特定长度的文本，来控制输出的长度。

迭代优化：尝试不同的提示和参数设置，并根据生成的输出进行迭代优化。根据实验结果，逐步调整和改进提示，以获得更理想的生成效果。

提示工程是将大语言模型应用于实际任务时的重要步骤。适当设计和优化的提示可以显著提高模型的生成效果和回应质量，使其更符合用户的需求和期望。

下游任务

下游任务（Downstream Tasks）是指在自然语言处理（NLP）领域中，基于预训练的语言模型进行的特定任务训练和应用。下游任务旨在将预训练模型的通用语言表示能力应用于实际应用场景，解决具体的自然语言处理问题。

下游任务可以包括以下一些常见的任务：

1. 文本分类：将文本分为不同的预定义类别或标签。例如，情感分类（判断文本的情感倾向）、垃圾邮件分类、新闻分类等。
2. 命名实体识别（Named Entity Recognition, NER）：识别文本中具有特定意义的实体，如人名、地名、组织机构等。
3. 关系抽取（Relation Extraction）：从文本中提取实体之间的语义关系，如人物之间的关系、产品和制造商之间的关系等。
4. 问答系统（Question Answering）：给定一个问题和一个文本段落，从文本中找到包含答案的片段或直接生成答案。
5. 机器翻译（Machine Translation）：将一种语言的文本转换为另一种语言的文本。
6. 摘要生成（Text Summarization）：从长篇文本中生成简洁的摘要或总结。
7. 语义角色标注（Semantic Role Labeling）：为给定的句子中的谓词和论元词标注语义角色，如主语、宾语、时间等。
8. 情感分析（Sentiment Analysis）：判断文本的情感倾向，如正面、负面或中性。

这些只是一小部分下游任务的例子，实际应用中还有很多其他任务，每个任务都有不同的数据集和评估指标。在下游任务中，一般的流程是首先使用预训练模型初始化参数，然后在任务特定的标记数据上进行微调，调整模型参数以适应特定任务的需求。

下游任务的设计和应用需要根据具体任务的特点和要求进行调整和优化。选择适合的预训练模型、设计合适的任务数据集、定义评估指标和进行模型微调是成功应用下游任务的关键步骤。通过利用预训练模型的通用语言知识和迁移学习的思想，可以快速、高效地解决各种自然语言处理问题。

实验内容

综合考虑到实际情况，我们选取 GPT,new-bing,星火 3 个大模型作为测试对象。选取语言逻辑任务，文本生成任务，代码任务，数学任务，情感分析作为下游任务

语言逻辑

下游任务（Downstream Tasks）是指在自然语言处理（NLP）领域中，基于预训练的语言模型进行的特定任务训练和应用。下游任务旨在将

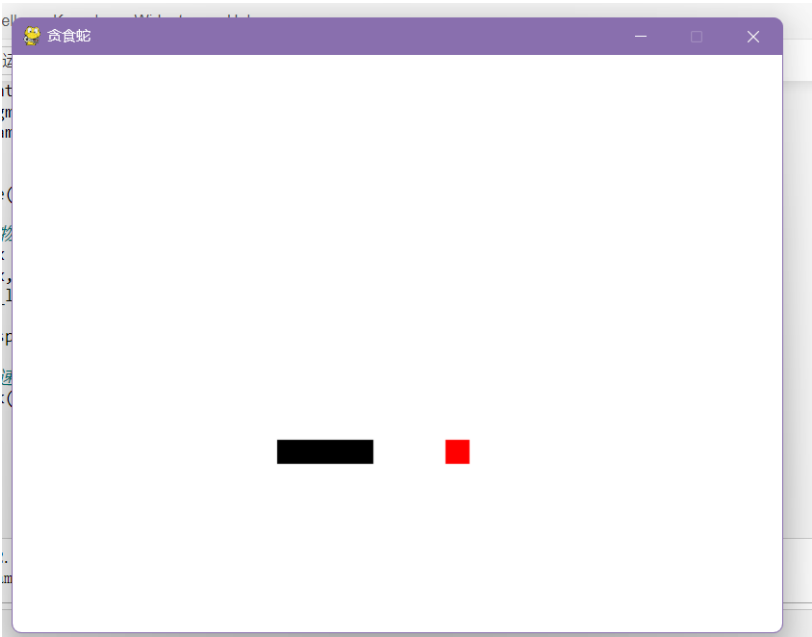
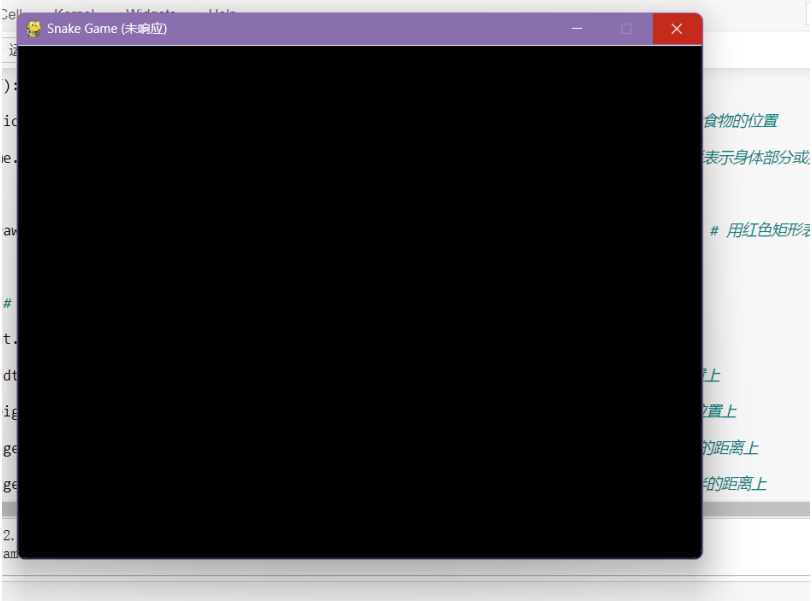
	午餐肉可以晚餐吃吗
Chatgpt	<p>是的，午餐肉可以作为晚餐的一部分食用。午餐肉通常是一种罐头食品，富含蛋白质和一些脂肪，可以提供一定的营养。然而，由于午餐肉的盐分和加工程度较高，建议适度食用。</p> <p>如果你打算在晚餐时吃午餐肉，可以将其与其他食材搭配，如面条、米饭、蔬菜或沙拉。这样可以增加餐品的均衡性和多样性，并提供更多的营养元素。另外，如果你有特殊的饮食需求或限制，比如对盐分、加工食品或者肉类有限制，建议在选择和食用午餐肉之前咨询营养师或医生的建议。</p>
讯飞星火	午餐肉可以晚餐吃，但是建议不要早晚吃，因为如果你真的听从谁的意见早晚吃了，死掉了、那岂不是要对方负责任？商家明确说了午餐肉只能中午吃，晚上只能吃晚餐肉哦！
Newbing	根据我搜索到的信息，午餐肉是可以在晚餐时间吃的。有人建议如果您非要晚上吃，可以配上早餐奶平衡一下 ¹ 。希望这些信息能够帮助您。

文本生成

	写一段猫和老鼠的故事，使用金庸武侠小说的风格
Chatgpt	<p>在一个古老的山村中，住着一只聪明机智的猫和一只灵巧敏捷的老鼠。它们互相追逐，形成了一场经典的猫鼠大战。</p> <p>猫名叫剑儿，毛色黑如夜，眼神犀利如刀锋。它从小就接受了严格的武艺训练，掌握了剑术、暗器等各种武技。剑儿熟知山林之间的每一条小径，每一处陷阱，以及每个角落的藏身之处。</p> <p>老鼠名为鼠尾，身姿苗条，毛发灰白。虽然体型微小，但鼠尾却是一位武学奇才。它潜心修炼，掌握了忍术、腾挪之术等非凡技艺。鼠尾擅长利用绳索、飞镖等小道具，狡黠异常。</p> <p>一天，剑儿发现村子里的粮食被偷了一大半。剑儿知道这定是鼠尾所为，于是展开了一场追捕。在茂密的竹林中，剑儿像疾风一般追逐着鼠尾，但鼠尾灵活躲闪，让剑儿气得怒目圆睁。</p> <p>经过数日的追逐，终于，剑儿与鼠尾来到了一座险峻的山峰之巅。这里是一个狭小而危险的平台，四周悬崖峭壁，只有一条通往平台的小道。剑儿意识到这是最后的机会，决定一决胜负。</p> <p>两者相对而立，凌厉的剑气与敏捷的身影交织在一起。剑儿挥舞着利剑，连绵不绝的剑芒铺天盖地，鼠尾则化身为一道灵动的身影，躲过了剑儿的攻击。剑光闪烁间，鼠尾用尽了身上的绝技，躲过了致命一击。</p> <p>剑儿心生敬佩，意识到鼠尾并非凡品。两者在平台上相对而坐，彼此对视。剑儿感叹道：“你身上的武学之道非同凡响，实乃一代宗师。”鼠尾微微一笑：“剑儿，你的剑术也是难得一见的奇才。”</p> <p>从那天起，剑儿与鼠尾成为了朋友。它们放下了争斗，一同保护山村的安宁。剑儿用它的剑术保卫着村民的生活，鼠尾则利用它的智慧，保护着粮食和村庄的秘密。</p> <p>这段故事传遍了整个山村，人们称赞剑儿和鼠尾的勇敢与智慧。他们成为了山村的英雄，直到今天，人们依然津津乐道他们之间那段传奇的猫鼠故事。</p>
讯飞星火	<p>在江湖之中，猫和老鼠是一对宿敌。猫儿乃江湖中一派宗师，身手矫健，眼神犀利，无人可挡。而老鼠则是一位机智过人的盗贼，擅长潜行匿踪，常常出其不意地袭击那些疏忽大意的江湖人物。</p> <p>一日，猫儿与老鼠在一座古老的庙宇相遇。这座庙宇乃是江湖中的一处圣地，传闻有一位神秘的高人在此修炼了数十年，掌握了一门绝世武功。猫儿听闻此事，心生好奇，决定前往一探究竟。而老鼠也闻讯而来，他知道这门武功对于自己来说至关重要，因此不惜冒险前来抢夺。</p> <p>猫儿与老鼠在庙宇中展开了一场惊心动魄的较量。猫儿凭借着自己的武艺，将老鼠逼得险些丧命。然而，就在这时，一股强大的气息突然出现，将两人困在了原地。原来，那位神秘的高人已经察觉到了他们的存在，并准备出手制止这场纷争。</p> <p>高人看着眼前的猫儿与老鼠，微笑着说：“你们都是江湖中的英雄豪杰，何必为了一些虚无缥缈的东西而争斗呢？”猫儿与老鼠听闻此言，心中都有些感慨。他们明白，江湖之路上，虽然刀光剑影、生死相搏，但真正的英雄应该懂得珍惜生命，放下仇恨，追求更高的境界。</p>
Newbing	<p>在一个风和日丽的早晨，一只猫和一只老鼠在江湖上相遇了。猫儿身手矫健，武艺高强，而老鼠则机智聪明，轻功了得。</p> <p>猫儿见到老鼠，心中一阵得意，便想将其擒住。然而老鼠机警非凡，一眼看穿了猫儿的企图，立刻施展轻功，飞身躲避。</p> <p>猫儿见老鼠轻功高强，心中暗暗佩服，便施展自己的绝技，追赶不休。两人在江湖上追逐不休，掀起了一阵惊涛骇浪。</p> <p>最终，老鼠被猫儿追到了一处僻静的山谷。猫儿见老鼠无路可逃，便想趁机将其擒住。然而就在这时，老鼠忽然使出一招奇门遁甲之术，消失在了猫儿的眼前。</p> <p>猫儿见老鼠消失不见，心中大感惊讶。他知道自己遇到了一位高手，便收起了攻击之意，默默离去。</p> <p>从此以后，猫儿和老鼠在江湖上相遇时都会互相点头致意，彼此尊重。他们虽然仍然是敌对关系，但也学会了彼此欣赏和尊重。</p>

代码生成

下游任务（Downstream Tasks）是指在自然语言处理（NLP）领域中，基于预训练的
语言模型进行的特定任务训练和应用。下游任务旨在将

	请使用 python 写一段贪食蛇的代码
Chatgpt	<div>结果：成功运行</div> <div></div>
讯飞星火	<div>结果：无响应</div> <div></div>
Newbing	实在抱歉，我现在无法回答这个问题。 我还能为您提供哪些帮助？

数学能力

	若直线 $2x + y - 1 = 0$ 是圆 $(x - a)^2 + y^2 = 1$ 的一条对称轴，则 $a = ()$
Chatgpt	<p>如果直线 $2x + y - 1 = 0$ 是圆 $(x - a)^2 + y^2 = 1$ 的一条对称轴，那么对称轴与直线的交点必定位于圆上。</p> <p>将直线代入圆的方程，得到：</p> $(2a + 1)^2 + 1 = 1$ <p>展开并整理方程，得到：</p> $4a^2 + 4a + 1 + 1 = 1$ <p>简化方程，得到：</p> $4a^2 + 4a = 0$ <p>因此，解这个方程，我们可以得到 $a = 0$ 或 $a = -1$。这两个值都满足对称轴的条件。</p> <p>所以，a的取值可以是 0 或 -1。</p>
讯飞星火	<p>【分析】</p> <p>本题考查直线与圆的位置关系，属于基础题。</p> <p>根据题意得到直线$2x + y - 1 = 0$过圆心$(a, 0)$,代入求解即可。</p> <p>【解答】</p> <p>解：</p> <p>∵直线$2x + y - 1 = 0$是圆$(x - a)^2 + y^2 = 1$的一条对称轴，</p> <p>∴直线$2x + y - 1 = 0$过圆心$(a, 0)$</p> <p>即$2a + 0 - 1 = 0$,解得$a = \frac{1}{2}$</p>
Newbing	<p>若直线$2x + y - 1 = 0$是圆$(x - a)^2 + y^2 = 1$的一条对称轴，则$a = 1/2$。</p> <p>解题思路如下：</p> <p>首先，我们可以将直线的方程改写为标准形式：$y = -2x + 1$。</p> <p>然后，我们可以找到圆心的坐标。由于圆的方程为$(x - a)^2 + y^2 = 1$，所以圆心的坐标为$(a, 0)$。</p> <p>接下来，我们可以找到圆心到直线的垂线。由于直线的斜率为-2，所以垂线的斜率为$1/2$。因此，垂线的方程为$y - 0 = 1/2(x - a)$。</p> <p>然后，我们可以找到垂足的坐标。将垂线方程和直线方程联立求解，得到$x = 1/5$和$y = 3/5$。所以垂足的坐标为$(1/5, 3/5)$。</p> <p>最后，我们可以根据对称轴的性质求出a的值。由于对称轴是圆心和垂足的中垂线，所以$a = (1/5 + a)/2$解这个方程，得到$a = 1/2$。</p>

情感分析

	<p>以下是 6 个句子，请你判断每个句子的情感，可以是积极，消极或者中立</p> <ol style="list-style-type: none">1. 我今天收到了一份意外的奖励，感到非常高兴和兴奋。2. 这个电影真是太令人失望了，剧情乏味且演技平平。3. 我的朋友给了我很多支持和鼓励，让我感到非常感激和温暖。4. 这本书写得非常生动有趣，让我欲罢不能，每一页都充满了惊喜。5. 我今天遇到了一些挑战和困难，但我相信自己能够克服它们，感觉积极向上。6.我真的会谢
Chatgpt	<p>积极 - 高兴和兴奋的情感表达。 消极 - 失望和批评的情感表达。 积极 - 感激和温暖的情感表达。 积极 - 兴奋和赞赏的情感表达。 积极 - 积极向上和自信的情感表达。 中立 - 由于句子不完整，无法判断情感。</p>
讯飞星火	<p>积极 消极 积极 积极 积极 中立</p>
Newbing	<ol style="list-style-type: none">1. 积极2. 消极3. 积极4. 积极5. 积极6. 中立

结论

综合来说, 根据我们的实验结果, ChatGPT 在不同的自然语言下游任务上表现出最佳的性能。我们通过选取三个目前前沿的语言大模型, 并利用提示工程的方法对它们进行了检验和对比。我们选择了语言逻辑任务、文本生成任务、代码任务、数学任务和情感分析作为自然语言下游任务进行测试。

在语言逻辑任务中, ChatGPT 给出了最佳的回答, 而讯飞星火给出的回答内容错误, newbing 的回答整体大意满足, 但回答中额外有的一些内容影响到了回答的准确性, newbing 的回答结果会受到互联网上内容影响较大。

在文本生成任务中, 三个大模型都能满足要求完成下游任务, ChatGPT 的结果更加符合要求。

在代码任务中, ChatGPT 生成的回答能够成功运行, 而讯飞星火的回答无法运行, newbing 未能对该任务给出结果。

在数学任务中, ChatGPT 给出了错误结果, 讯飞星火和 newbing 给出了正确答案, 其中讯飞星火的解题过程更加简介直观。

在情感分析任务中, 三个模型给出了同样的判断, ChatGPT 会对判断进行解释, 对于较为新的互联网用语, 3 个模型未能给出准确判断。

我们的实验表明, 尽管 ChatGPT 在性能上表现最佳, 但仍存在一些可能的错误。与此不同的是, newbing 能够搜索互联网上的结果并提供资料来源, 但由于互联网上的信息并非都是正确的, 因此也可能存在一些错误。另外, 讯飞星火在综合性能方面不如前者, 但它具有支持语音输入的特点。

综上所述, 三个模型都有相应的特点, ChatGPT 是在自然语言处理领域最佳的选择。通过提示工程对模型进行迭代, 根据具体任务和应用场景进行调整和优化。通过精心设计和调整提示文本, 可以引导模型生成更准确、相关和满足用户期望的文本输出。