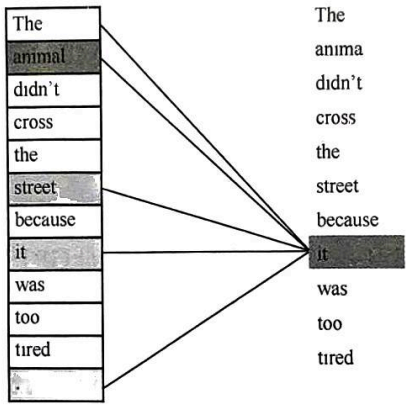
频等多个领域。这一飞跃式的进步不仅极大地提升了内容生产的效率，而且降低了创作的门槛，使得更多人能够参与内容创造。

1.1.1　生成原理

大模型基于 Transformer 架构进行构建，由多层神经网络架构叠加而成，能够根据输入内容预测输出内容。

大模型的核心生成原理是将输入的语句以词向量的表征形式传递给神经网络，通过编码器/解码器（Encoder/Decoder，详见第 3 章）、位置编码和自注意力机制建立单词（或字）之间的联系。从宏观的视角来看，输入的每个单词（或字）首先会与已经编码在模型中的单词（或字）进行相关性计算，然后把这种相关性以编码的形式叠加在每个单词（或字）中。如图 1-2 所示，经过计算后，“it”与输入句子中的其他单词的相关性权重将会增加，颜色越深代表相关性越高。



The animal didn't cross the street because it was too tired.

图 1-2　相关性权重可视化示例

在获得各个单词间的相关性之后，模型以概率分数标记序列中下一个输出的单词的可能性（也称概率），并选择最佳选项。如图 1-3 所示，由于“movie”的概率最大，因此模型的最终输出结果为“movie”。

Interstellar is a very excellent

song	3 2%
movie	5.2%
cartoon	2 8%
animation	3.1%

图 1-3　不同单词的输出概率

虽然模型会选择下一个最合适的单词，但是由多个最佳单词组成的句子可以并不通顺。为了解决这个问题，Transformer 使用了 Beam Search（束搜索）[1]等方法以提高生成质量。这

1　束搜索是处理文本生成任务时常用的解码策略。

些方法不是只关注序列中的下一个单词，而是将更大的一组单词作为一个整体来考虑，同时考虑多个序列上的联合概率。如图 1-4 所示，我们同时考量 4 个序列上的联合概率（为了方便理解，此处以一组单词的颜色深浅来表示输出概率，单词的颜色越深，代表其被选择并输出的概率越大），将一组单词作为整体进行评估，可以有效提高模型的生成质量。

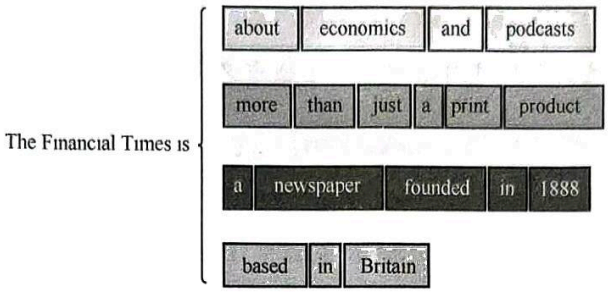


图 1-4 通过束搜索方法生成最佳输出

综上所述，可以将大模型看作概率模型。不同于通过数据库对数据进行检索，大模型通过大量学习世界知识，依据概率生成足够准确的回答。

1.1.2 关键技术

大模型（如 GPT-4、LLaMA2 等）的优异性能离不开多种技术的支持。本节将介绍大模型的常用技术，这些技术在大模型的研究过程中提供了重要的理论支撑。

1. 迁移学习

迁移学习（Transfer Learning）最早于 2005 年由加拿大工程院和皇家科学院院士杨强教授提出[14]。作为机器学习的重要分支，迁移学习是指利用在源领域中训练好的模型，将其知识和能力迁移到新的目标领域，以解决该领域的问题。通常，我们会首先在数据量大的领域训练模型，然后将其应用于数据量较小的领域。

换言之，迁移学习通过将模型已学习的知识迁移到新的任务上，以提高模型在新任务上的性能。在大模型的开发过程中，开发者常常将在大型文本数据集上训练好的模型作为基座，通过微调等手段让模型适应新的下游任务。这一应用的核心是运用已有的基础知识来学习更专业的知识。

2. 零样本学习

2009 年，Lampert 等人发布 Animals with Attributes 数据集（该数据集已在由 Lampert 领导的奥地利科技学院机器学习与计算机视觉小组网站开源），并提出一种基于属性的类间迁移学习机制。该机制对于零样本学习（Zero-shot Learning）的发展起到重要的奠基作用[15]。零样本学习的核心在于使模型能够识别那些从未在训练集中出现过的数据类别，从而扩展模

型的应用范围并增强其泛化能力。

在大模型研究中，模型的零样本学习能力已成为评估模型性能的重要指标之一。与此同时，提示词（Prompt）作为大模型的一种关键输入形式，经常与零样本学习协同使用，以优化模型的性能。提示词可以被视作用户向大模型发送的指令。通过精心设计提示词，用户可以引导大模型生成符合期望的内容。例如下面的示例。

模型输入：你现在需要从这句话中抽取出城市名称和目的地信息。我想去上海的外滩，那里有壮观的夜景。

模型输出：上海，外滩

零样本学习存在使用限制，只有当用户的目标任务与模型已具备的能力相匹配时才能获得最优的零样本学习效果。

用户在实际使用大模型时，通常会根据任务的复杂性选择不同的提示词策略。提示词策略主要包含3种：零样本提示（Zero-shot Prompting）适用于简单任务，无须向大模型提供任何示例，直接下达指令即可；在面对较为复杂的任务时，小样本提示（Few-shot Prompting）更为有效，它通过提供少量示例来帮助大模型更好地理解并执行任务；当面临复杂的推理任务（如算术推理、常识推理和符号推理等）时，思维链提示（Chain-of-Thought Prompting）则成为首选，它通过向大模型提供带有中间推理步骤的示例来引导大模型逐步完成推理。

3. 小样本学习

小样本学习（Few-shot Learning）是指大模型通过使用很少的样本来进行学习。与零样本学习相似，小样本学习也经常与提示词配合使用。如果大模型无法理解用户提出的问题，或者用户需要大模型按照特定的格式进行输出时，用户可以使用小样本学习方法，在提示词中加入引导信息并给出一定的样例，指导大模型生成满足业务要求的输出结果。

接下来对比小样本学习和零样本学习在构造提示词时的异同。在下面的示例中，我们使用零样本学习方法构造提示词，要求模型对句子进行分词。从下面的示例可以看到，模型的输出结果并不正确。

模型输入：请对这句话进行分词。我一把把车把把住了

模型输出：我/一把把/车把/把/住/了

如果在提示词中给出部分示例，模型将可以生成更好的输出结果。在下面的示例中，我们使用小样本学习来构造提示词。

模型输入：请参照下面的示例，对给定语句进行分词，示例数据如下：
我想过过儿过过的生活

对上述示例进行解析，“我”是主语，“想过”是指想要尝试，“过儿”指的是《神雕侠侣》中的杨过，“过过的生活”指的是曾经生活过的生活。因此，最终的分词结果应该是：我/想过/过儿/过过的/生活。

向模型输入由小样本学习构造的提示词后，对于相同的分词任务，模型可以生成更好的结果。

模型输入：我一把把车把把住了

模型输出：我/一把/把/车把/把住了

4. 持续学习

在一般情况下，当开发者的需求或所在的应用场景发生变化时，往往需要重新训练模型。作为一种增量学习，持续学习（Continual Learning）的核心思想是在一系列任务上训练模型，模型保留在旧任务上学习的知识并继续在新任务上进行训练。该方法在深度神经网络时代被赋予了更大的实用价值[16]。以百度公司在 2019 年 7 月推出的 ERNIE 2.0 模型[17]为例，研究者使用持续学习方法来训练模型，引入了大量的预训练任务。ERNIE 2.0 模型在学习新任务的同时保留对旧任务的记忆，渐进式地学习词语、句法和语义表征知识。在多项自然语言处理任务上，它都取得了超过 BERT 模型与 XLNet 模型的表现。

5. 多任务学习

传统的机器学习主要基于单任务的模式进行学习。对于复杂的任务，首先将其分解为多个独立的单任务并进行处理，然后对学习的结果进行组合。多任务学习（Multi-Task Learning）是一种联合学习方法[18]。在这种方法中，模型通过对多个任务进行并行学习，共享表征信息，可以取得比训练单任务更好的表现。此时模型具有更好的泛化能力。

多任务学习的关键在于寻找任务之间的关系。如果多个任务之间的关系搭配恰当，那么不同任务能够提供额外的有用信息，进而可以训练出表现更好、更鲁棒的模型。例如，GPT-2 模型采用多任务学习，通过在 WebText 数据集（40GB 大规模数据集）上进行自监督预训练，在多项任务上取得 SOTA（State-Of-The-Art，在指定领域最高水平的技术）结果。

6. RLHF

强化学习（Reinforcement Learning，RL）是指通过不断与环境交互、试错，最终完成特定目的或者使得整体行动收益最大化的技术。强化学习不需要标注数据集，但是需要在每一步行动后得到环境给予的反馈，基于反馈不断调整训练对象的行为。

2017 年，OpenAI 公司和 DeepMind 公司的研究人员在论文“Deep Reinforcement Learning from Human Preference”中提出基于人类偏好的强化学习概念。研究人员通过实验证明，将非专家标注的少量数据作为反馈，可以提高模型在雅达利游戏中的性能[19]。

2022 年，OpenAI 公司在 InstructGPT 模型的训练过程中引入 RLHF（Reinforcement Learning from Human Feedback，基于人类反馈的强化学习）。该技术在大模型训练中发挥了巨大作用，有效减少了模型输出中的有害内容，力图实现模型与人类的价值观对齐[20]。RLHF 是涉及多个模型和不同训练阶段的复杂技术，这里将其分成 3 个阶段进行讲解。

第一阶段，OpenAI 公司将 GPT-3 模型作为 InstructGPT 模型的预训练模型，借助数十名人工标注师为训练数据集中的问题编写预期输出结果（人工编写每个问题的答案），利用标注数据对 GPT-3 模型进行监督训练。模型首先通过前向推理生成多个输出结果，然后通过人工对模型的输出结果进行打分和排序，并将打分和排序数据用于训练奖励模型（Reward Model）。

第二阶段，目标是训练奖励模型。奖励模型应能评判 InstructGPT 模型的输出结果是否符合人类偏好。如图 1-5 所示，奖励模型接收一系列输入并返回标量奖励，标量奖励与人类的反馈数据共同参与损失函数的计算。在模型的选择上，奖励模型可以是经过微调或根据偏好数据重新训练的语言模型。

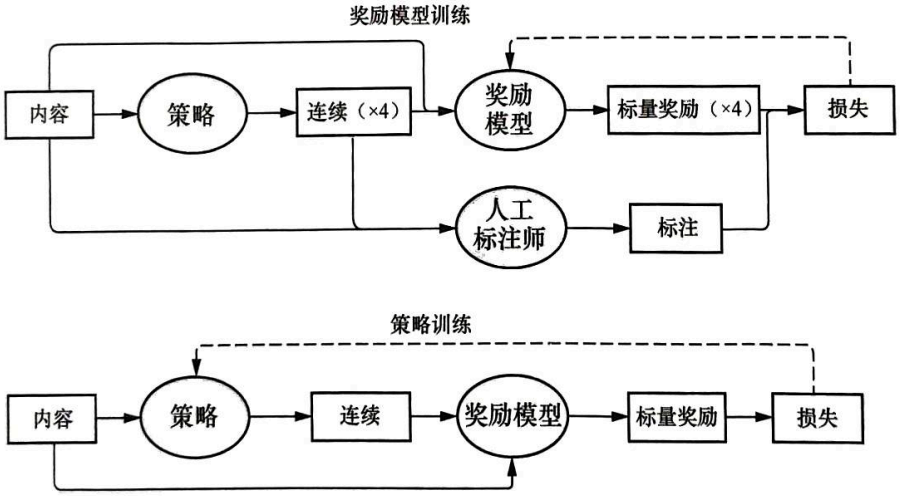


图 1-5　RLHF 训练过程[20]

第三阶段，采样新的输入句子，首先利用 PPO（Proximal Policy Optimization，近端策略优化）网络生成输出结果，然后奖励模型计算反馈，并将结果作用于 PPO 网络，以此反复，最终训练出与人类价值观对齐的模型。

PPO 算法由 OpenAI 公司于 2017 年提出，它是一种基于策略的强化学习算法[21]。它将智能体（Agent）当前的状态输入神经网络，可以得到相应的下一步行动（Action）和奖励（Reward），并更新智能体的状态。OpenAI 公司的 John Schulman 等人在一系列基准任务上对 PPO 算法进行测试，发现该算法比其他算法在样本复杂性、简单性和运行时间上具有更好的平衡性。

2023 年，Google 公司提出 RLAIF（Reinforcement Learning from AI Feedback，基于 AI 反馈的强化学习）。该技术使用人工智能模型来取代 RLHF 中的人工标注师。与 RLHF 相比，模型经过 RLAIF 训练后，可以在无害内容生成、文本总结等任务上达到与 RLHF 相近的水平[22]。

7. 上下文学习

2020 年 6 月，OpenAI 公司在发布 GPT-3 模型的同时提出上下文学习（In Context Learning）[1] 概念。基于上下文学习，模型不根据下游任务来调整参数，而是连接下游任务的输入输出，以此作为提示词引导模型根据测试集的输入生成预测结果。该方法的实际效果大幅超越无监督学习。

8. 思维链

思维链（Chain of Thought）最早由 Google 公司的高级研究员 Json Wei 等人于 2022 年提出。思维链是一种离散式的提示学习方法，可以提高模型在复杂任务上的性能[23]。如图 1-6 所示，为了指导大模型进行数学运算，研究人员给出少量人工编写的推理示例，并将步骤解释清晰，引导大模型对相似问题进行推理。此处将包含人工编写的详细推理过程的提示词称为思维链提示。思维链可以激发大模型的多步推理能力。这个过程类似于人类通过学习他人的思维方式来进行深度思考以解决复杂任务。

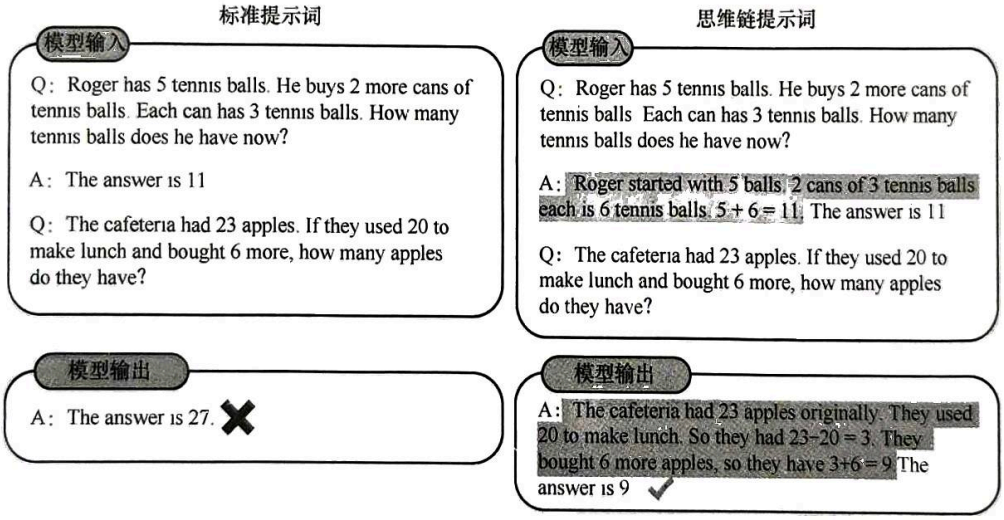


图 1-6 以思维链方法指导模型生成正确答案

9. 提示工程

在人工智能领域，尤其是大模型中，提示词对于模型的生成结果和质量具有重要影响。一个好的提示词可以帮助模型更好地理解用户的意图和需求，并生成更加准确、有针对性的回复。所以，也可以将提示工程看作一种优化和改进人工智能模型性能的方法。前面提到的零样本提示和小样本提示都属于提示工程的范畴。这类方法并不涉及对模型参数的修改或重

1 上下文学习又称情境学习。

新训练，而是通过特殊设计的提示词引导模型生成更好的结果。

在实际使用中，如果我们需要使模型快速实现特定的任务目标，或者需要以一定的格式生成内容，则可以使用提示工程方法，通过少量样例或具有一定格式的样例引导模型生成内容。与微调相比，提示工程不需要用户拥有大量的人工智能领域专业知识，只需要对特定任务有一定的了解，设计合适的提示文本即可。

1.1.3　关键术语

本节将详细介绍与大模型相关的 3 个关键术语——Token、Prompt 和 Embedding。这些术语在自然语言处理和机器学习领域中扮演着重要的角色，对于理解和应用大模型至关重要。

1. Token

在大模型中，Token（词元）是文本中的最小单位，可以代表一个单词、一个标点符号、一个数字等。Tokenization（分词）是将一个句子或文本分成多个 Token 的过程。常用的分词方法包括 BPE（Byte-Pair Encoding，字节对编码）算法、WordPiece 算法和 ULM（Uni-gram Language Model，一元语言模型）算法。在第 12 章中进行开源模型的微调实践时，在训练模型前，不仅需要加载模型的权重参数，还需要导入对应的 Tokenizer（分词器）。

在大模型的训练和应用中，模型将接收的大量 Token 作为输入，并对下一个最有可能出现的 Token 进行预测。如今，很多模型会将 Token 处理为词向量（Embedding，也称为词嵌入）的形式，这种形式的数据便于在神经网络中处理。

2. Prompt

这里讲解的 Prompt 偏向于模型的训练过程，另外，Prompt 的形式不拘泥于自然语言，也可以是向量，不同于用户与模型沟通交互时传递的提示词。此处的 Prompt 将会给模型提供输入的上下文信息。在有监督或无监督训练过程中，Prompt 可以帮助模型更好地理解输入内容并响应。

具体而言，针对情感分类任务，可以在输入“I love this song”句子后增加提示词“This song is ___”，将情感分类任务转变为完形填空任务。模型将输出“Wonderful”“Moving”等具有情感偏好的形容词。这种方式可以引导模型输出更加正确的分类标签。将上述问题转化为更一般的形式，对于输入文本 x，通常有专用的函数 $f_{\text{forward}}(x)$，可以将 x 转换为所需的形式。应用模板如下。

```
[X] This is a [Z] song.
```

其中，[X]指代输入文本 x 的位置，[Z]指代生成文本的位置。

在研究过程中，一般会留下空位置以便模型填充答案。如果空位置在句子中间，一般称此类提示词为 Cloze Prompt；如果空位置在句末，则称为 Prefix Prompt。总之，通过设计合

理的 Prompt，经过无监督预训练的模型也可以处理多种下游任务。

Prompt 主要有两种设计方法——手工设计模板和自动学习模板。

在 Prompt 发展初期，一般将其称为一种输入形式或模板。手工设计模板一般基于自然语言而设计，例如经典的 LAMA 数据集中包含的 Cloze Templates[24]。手工设计模板的优势是直观、简单，但缺点是需要设计人员拥有大量的相关知识和经验，并且构造速度较慢。

为了解决手工设计模板的缺点，研究人员提出了自动学习模板。自动学习模板可细分为离散提示词（Discrete Prompt）和连续提示词（Continuous Prompt）两类。离散提示词是指自动生成由自然语言组成的提示词，因其搜索空间是离散的而得名，常用方法包括 Prompt Mining、Prompt Paraphrasing、Gradient-based Search 等。而连续提示词则不把提示词的形式拘泥于自然语言，向量也可以作为提示词（通常是可训练的），因为自然语言并不是模型或机器能直接理解的语言。第 6 章将要介绍的 Prefix tuning 方法就属于连续提示词类的经典方法。

3. Embedding

在机器学习和自然语言处理领域中，Embedding（词嵌入，也称词向量）是一种将高维度离散数据（如单词、短语或整个句子等文本数据）映射到低维度连续向量空间的技术。这种映射的目的是捕捉和表征数据的语义与句法特征，使得原本在高维度空间中表示的数据在低维度空间中能够更加有效地进行处理和分析。

随着机器学习和自然语言处理技术的快速发展，Embedding 得到进一步改进。例如，BERT、ELMo 和 GPT 等模型可以生成上下文相关的向量表示。这些向量不仅捕获单词的语义信息，而且融入了上下文信息，可以提高模型对语言的理解能力。

此外，Embedding 与向量数据库的结合使用为大模型的知识检索和知识补充提供了强大支持。Embedding 可以将用户提问转化为向量表示，并在向量数据库（已存入大量知识）中进行相似度计算，取出相似度最高的知识并将其作为提示词输入模型中，以实现对模型输入的有益补充。这种技术在信息检索、问答系统等方面具有广泛的应用前景。在 11.3.3 节中，我们将详细介绍这一技术的原理及应用案例。

在实际开发过程中，比较常用的生成词向量的方法包括 Word2Vec、GloVe[25]以及 OpenAI 公司提供的 Embedding 工具等。

1.2 大模型分类

本节将按照模型结构、模态、微调方式对大模型进行分类，读者通过对本节的学习，可