

人工智能专题研究

AIGC投资框架

西南证券研究发展中心
计算机研究团队
2023年5月

起点财经GPT搞钱交流群

ChatGPT不会淘汰你！
先驾驭ChatGPT的人会淘汰你！

- 1 各路大神 **畅聊AI** 使用指南和落地应用，分享商业化案例，碰撞思维火花
- 2 一次性领取 **649份** ChatGPT、AIGC **相关资料**，赠送 **80页** ChatGPT、AI绘画Midjourney保姆级教程，资料持续更新中
- 3 不定时分享AI智能、ChatGPT、AIGC、GPT-4等 **最新研报** 和相关资讯
- 4 不定期邀请行业 **大咖** 演讲互动交流学习
- 5 ChatGPT **初体验**，专属微信群与GPT互动提问！（目前开放的API为3.5版本）

识别二维码查看详情



核心观点

- **自然语言模型是大语言模型的核心。**自然语言模型是大模型的核心，支持许多类型的语言任务的执行。ChatGPT和各大主流AIGC模型主要基于大语言模型，通过深度学习算法来处理和生成自然语言。自然语言模型发展经历了基于规则、基于统计和基于神经网络三个阶段，Transformer的出现为以Bert和GPT-3为代表的大语言模型的诞生奠定了模型基础。
- **AI到AGI，ChatGPT开启人工智能新纪元。**ChatGPT结合自回归语言模型+zero shot prompting，带来AI范式突破，实现AI到AGI的转变。相比fine tuning技术，prompting技术能更好的完成生成类任务。ChatGPT在RLHF的帮助下，找到了GPT-3.5和人类自然语言的合理接口，解锁了模型的应用场景，改变了人机交互方式。
- **GPT跨越式发展，探索未来更多可能性。**各代GPT青出于蓝胜于蓝，更新迭代过程中涌现了诸多能力。GPT-4.0更是突破了GPT-3.5仅支持执行文本类任务的局限性，赋予ChatGPT更优异的上下文理解能力、推理能力、演绎能力，甚至图文理解能力。GPT也在不断的更新迭代中持续突破，解锁更多能力。
- **多行业结合，下游应用遍地开花。**ChatGPT以其高级的生成能力和复杂任务处理能力，以从重资产的模型层到轻资产的中间层和应用层为架构，为游戏行业、视频类行业、生物医药行业等其他行业锦上添花。在ChatGPT的帮助下，各行各业真正的做到了提高效率、减少成本并实现了功能上质的改变，让更多用户受益。
- **风险提示：**AI技术更新迭代缓慢、专业领域落地效果不及预期、版权纠纷、伦理问题、算力建设缓慢等风险。

目录



1 大模型与GPT

- 1.1 大模型篇
- 1.2 ChatGPT篇
- 1.3 GPT-4篇



2 国内篇

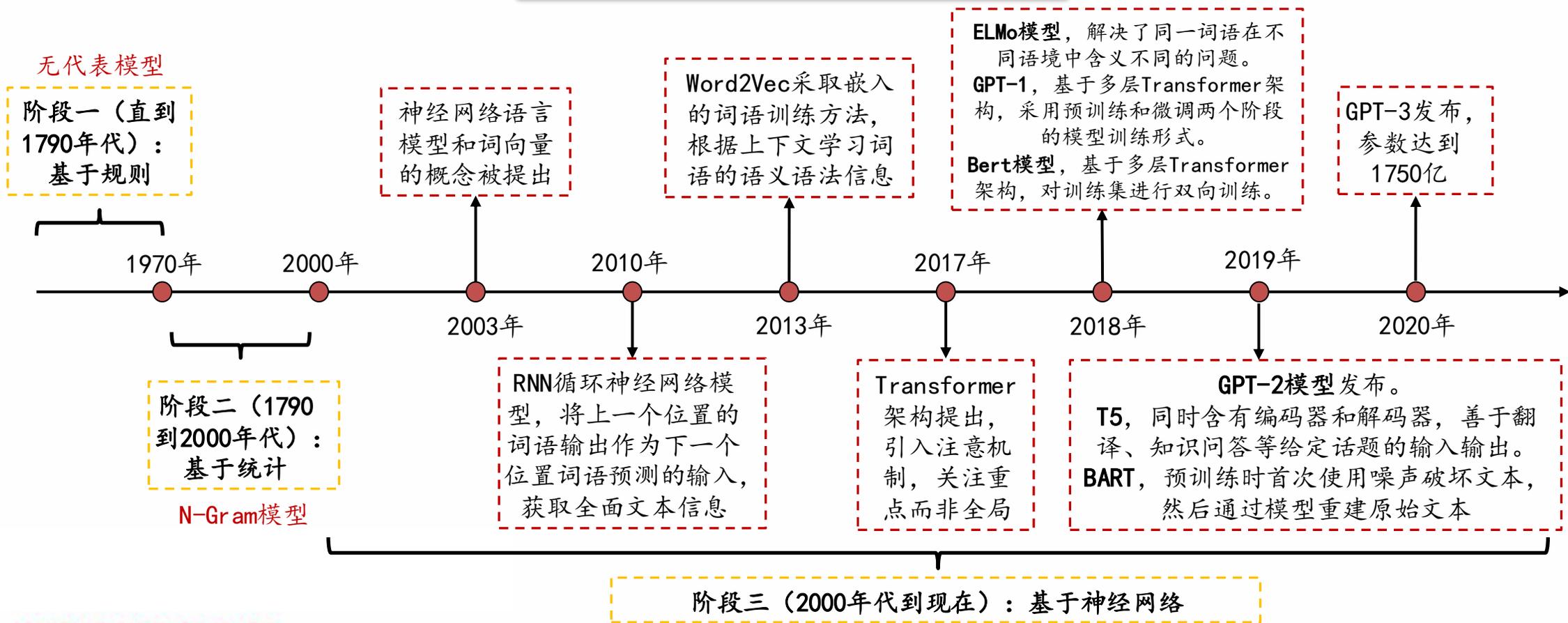


3 ChatGPT应用场景

1.1.1 大语言模型简介

大语言模型简称LLM (LargeLanguageModels)，是基础的机器学习模型，使用深度学习算法来处理和生成自然语言。这些模型在大量的文本数据上进行训练，再自动生成符合语言规律的文本。大语言模型可以执行许多类型的语言任务，如翻译语言、分析情感、聊天机器人对话等。

自然语言模型的发展历史



1.1.2 大语言模型的主要发展历程

➤ 第一阶段 (2013-2020)



在Bert和GPT模型问世之前：深度学习模型是自然语言处理（NLP）领域主流的技术，而深度学习在自然语言处理领域的应用主要依赖于以下关键技术：**使用大量改进的LSTM模型和少量改进的CNN模型作为典型的特征提取器；**使用序列到序列（Sequence to Sequence）（也称为编码器-解码器（Encoder-decoder））加上注意力作为各种具体任务的技术框架。**但这个技术拥有两大主要问题阻碍了深度学习领域的突破：第一，训练数据的数量有限。随着容量和深度的增加，模型需要更多的训练数据支持。第二，这项技术的表达能力不够强。即使拥有足量数据，也很难有效的吸收数据内容。**

在Bert和GPT模型问世之后，出现了两大主要范式转换：第一个范式转换体现在预训练模型出现并有两大主要影响，体现在：**其一，中间任务消失。**通过大量数据预训练，Bert和GPT把中间任务作为语言特征，吸收到Transformer的参数里，做到直接解决最终任务。**其二，统一了不同研究方向技术路线。**最终任务包括了自然语言理解类任务和语义生成类任务。**对于理解类任务，其技术体系统一到以Bert为代表的双向语言模型加Fine Tuning模式；对于生成类任务，其技术体系统一到以GPT2.0为代表的自回归语言模型加zero shot prompting（GPT-3）或few shot prompting（ChatGPT）模式。**

第二个范式转换为从预训练模型逐渐偏向于通用人工智能，简称AGI（Artificial General Intelligence），我们目前正处于这个转换的以GPT3.0（自回归语言模型+zero shot prompting）占据主导位置为代表的过渡过程中。自回归语言模型+zero shot prompting处于目前的主导地位的原因有两个：**其一，比如Google的T5模型，在自然语言生成任务中可以同时完成自然语言理解任务。**如仍使用类似Bert的理解类任务语言模型，则无法得到优质的生成内容。所以相比于理解类任务模型，目前来说大众更偏向于生成类任务模型。**其二，GPT模式效果更优。**若想以zero shot prompting或few shot prompting模式完成任务，必须采取GPT模式，也就是生成类任务模型。Prompting技术也更加符合人性，相比之下，fine tuning模式是被动适应模型。另外，zero shot prompting模式很难被大语言模型所理解，表现效果也差强人意，以目前的技术很难达到令人满意的效果。**所以目前科学家偏向于研究few shot prompting技术，起到过渡作用。**

➤ 第二阶段 (2020.06-至今)

1.1.3 大语言模型学习、存储、修正数据的方式

大语言模型学习数据的方式

- 大语言模型从海量的文本数据中学习了大量知识。对于其学习到的知识一共有两类：**语言类和世界知识类**。
- **浅层语言类知识**储存在Transformer的低层和**中层**，而**相对抽象的语言类知识**储存在Transformer的**中层或高层**。大语言模型在学习过程中不需要过多的语言类知识数据。
- **世界知识类知识**可以分为**事实型和常识型**。世界知识主要储存在Transformer的**中高层**，**同时大语言模型学习需要大量事实型世界知识**。

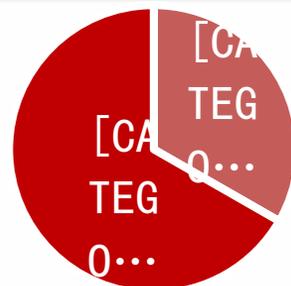
Transformer知识金字塔



大语言模型储存数据的方式

- 大语言模型会将学习过的知识储存在Transformer的模型参数里。
- Transformer的模型参数结构分为两个部分：其中，**多头注意力（MHA）的部分占参数总数的三分之一**；**FFN结构占参数总数的三分之二**。
- MHA主要用于计算单词与知识间的关联强度，不用来存储知识。**FFN存储着大量知识**，是大语言模型的key-value存储器。

Transformer参数结构分布



大语言模型修正数据的方式

- 大语言模型有很大可能会生成**错误答案**，所以修正数据非常必要。大模型有三种修正方法。
- **第一种是从训练数据源头去修正**。但在每次修正后需要重建模型，成本过高，无发展前景。
- **第二种是通过finetuning修正知识**。但这种方法容易误删正确数据，会影响效率。
- **第三种是直接修改知识对应的模型参数**。直接定位到储存旧知识的FFN节点并替换新知识。

精准定位FFN节点



1.1.4 大语言模型的分类

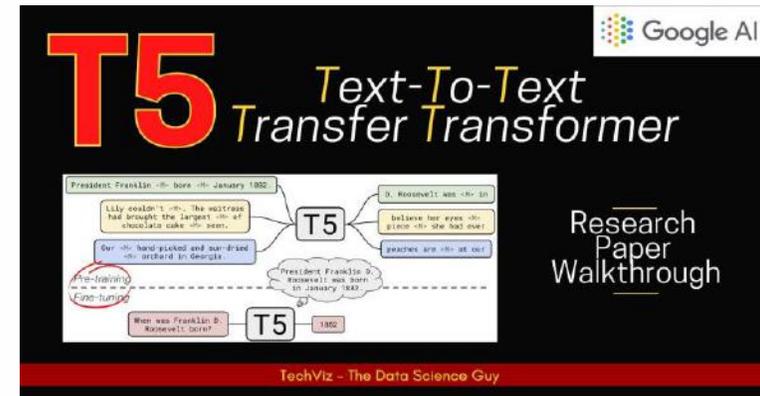
大语言模型按模型结构一共分为以下三类：

1.以BERT系列为代表的encoder-only模型：在每个阶段，注意力层可以访问初始句子中的所有单词并具有双向注意力。这类模型的预训练通常围绕着以某种方式破坏一个给定的句子，并让模型承担寻找或重建初始句子的任务。Encoder-only模型最适合于需要理解完整句子的任务。



2.以GPT为代表的decoder-only模型：在每个阶段，对于一个给定的词，注意力层只能访问句子中定位在它之前的词。这类模型通常被称为自回归模型。解码器模型的预训练通常围绕着预测句子中的下一个词进行。Decoder-only模型最适合于涉及文本生成的任务。

3.以T5为代表的既有encoder、又有decoder的模型：在每个阶段，encoder的注意力层可以访问初始句子中的所有单词，而decoder的注意力层只能访问输入中位于特定单词之前的单词。这类模型的预训练通常涉及到更复杂的任务。Encoder-decoder模型最适合于围绕根据给定的输入生成新的句子的任务，如总结、翻译或生成式问题回答。



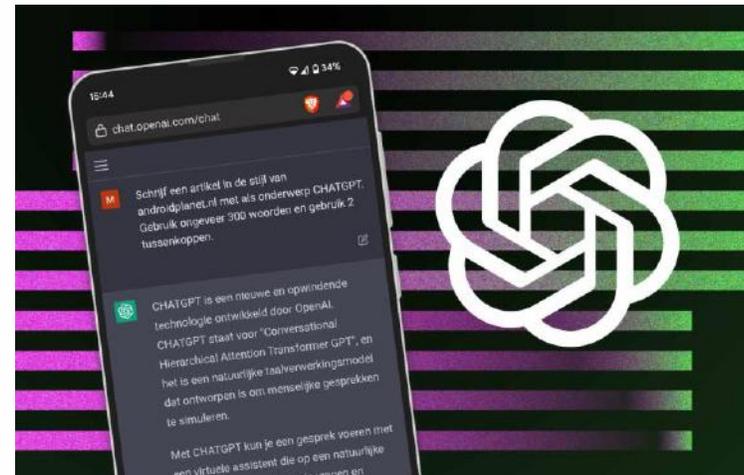
1.1.4 大语言模型的分类

随着深度学习技术的不断发展，大型模型在自然语言处理领域发挥着越来越重要的作用。大型模型通过使用更多的参数、更复杂的结构和更多的训练数据来提高性能，从而在自然语言处理任务中取得了显著的成果。**语义理解是自然语言处理模型的基础任务之一。**Bert和其变体模型被广泛用于各种语义理解任务，例如文本分类、命名实体识别和自然语言推理等。**文本生成也是自然语言处理模型中的另外一个重要任务。**在自然语言生成任务中，GPT已成为最典型的框架之一，因为它能够生成高质量的文本，并且能够处理更加复杂的任务。与语义理解相比，自然语言生成需要更多的计算资源和更大的模型来处理更多的数据和更复杂的任务。

需要注意的是，大型模型具有多元化和多样性，因此在选择适当的模型时需要考虑任务的性质和特点。对于简单的任务，不需要使用像ChatGPT这样复杂的大型模型，而对于更加复杂的任务，则需要使用更大、更复杂的模型来提高性能。

语义理解任务模型

文本生成任务模型



1.1.5 未来大模型发展的前提要素

拥有一个优质的网络架构是未来大模型发展的重要前提。缺少高质量的网络架构，大模型很难达到质的飞跃。而一个高质量的网络架构，不仅可以推动大模型的发展，还可以促进AGI达到一个质的飞跃。网络架构有三大主要种类，包括：CNN、LSTM、Transformer。

网络架构的三大主要种类

CNN (Convolutional NeuralNetwork)

CNN（卷积神经网络）是一种广泛应用于计算机视觉领域和自然语言处理领域的深度学习算法，是早期最常见的深度学习模型，可以有效的提取图像特征。

CNN主要有三层结构：卷积层、池化层和全连接层。相比较而言，Transformer的表现能力远超CNN。早期的CNN具有无法有效捕获长距离特征的问题，但随着网络深度的逐步增加，这个问题得到了很大的缓解但仍无法达到Transformer一样优秀的表现能力。

LSTM (LongShort- TermMemory)

LSTM（长短时记忆结构）是传统RNN模型的变体，能够缓解传统RNN模型对于过长序列存在的梯度消失以及遗忘或爆炸问题。

LSTM由三个门控构成，分别是输入门、遗忘门以及输出门。LSTM主要应用于自然语言处理领域和语音识别领域序列数据的处理，能够处理和预测时间序列中间隔较长的信息。通常情况下，与传统的RNN模型相比，它在处理需要长期依赖信息记忆的任务时表现更优。

Transformer

Transformer是一种新型的网络架构，用于自然语言处理（NLP）的深度学习模型，表现能力优异，远超CNN、RNN。

Transformer结构由两部分构成，分别是encoder（编码器）和decoder（解码器）。Transformer具有优秀的语义特征提取能力、远距离特征捕获能力、综合特征提取能力、并行计算能力及运行效率。

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

人工智能行业专题：AIGC投资框架（西南证券）.pdf

请登录 <https://shgis.com/post/1380.html> 下载完整文档。

手机端请扫码查看：

