



《对自然语言处理成果GTP-3的简介》

|  |  |
| --- | --- |
| **专 业** | **计算机科学与技术** |
| **姓 名** | **蔡泽栋** |
| **日 期** | **2022/8/31** |

**摘 要**

近几年，自然语言处理遵循一个规律或者说方法论，大部分在做预训练，用更复杂的模型、更大的数据量、更多的计算资源进行模型预训练。然后针对领域或者特定任务进行微调，还会遇到解决样本量不足、小样本等问题。在落地应用时，如果使用前期预训练非常大的模型，对计算资源依赖度又非常高，如何通过知识蒸馏得到小模型，最后将小模型应用到实际领域，也是学术界和工业界重点研究的方向。

从 GPT-1 到 GPT-2，再到 GPT-3 让大家逐步认识到增大数据量、模型，以计算流的方式将计算发挥到极致，可以解决很多问题。当前，学术界、工业界的趋势之一是如何用更大的数据训练更复杂的算法系统去解决相关问题。

在 2020 年，GPT-3 的论文一经发表就引发了业内轰动，因为这一版本模型有着巨大的 1750 亿参数量，比GPT-2放大了 100 倍（96 层和 1,750 亿个参数），并且接受了更多数据的训练，支持的任务广泛且旨在测试快速适应不太可能直接包含在训练集中的任务的几个新任务。

本文将从GPT-3的应用、内容、模型和研究过程等方面来向你简要介绍这个自然语言处理最前沿的研究成果。

**目 录**

[**第1章 GPT-3的应用简介** - 1 -](#_Toc113265071)

[1.1 GPT-3的多种应用可能 - 1 -](#_Toc113265072)

[1.1.1 GPT-3将成为工作的助手 - 1 -](#_Toc113265073)

[1.1.2 GPT-3使我们进入人工文本生成的新时代 - 1 -](#_Toc113265074)

[**第2章 GPT-3的内容简介** - 2 -](#_Toc113265075)

[2.1 GPT-3是什么？ - 2 -](#_Toc113265076)

[2.2 GPT-3有什么限制？ - 2 -](#_Toc113265077)

[2.3 GPT-3的含义 - 3 -](#_Toc113265078)

[**第3章 GPT-3的研究过程及原理** - 4 -](#_Toc113265079)

[3.1 GPT-3研究过程介绍 - 4 -](#_Toc113265080)

[3.2 GPT-3模型简介 - 4 -](#_Toc113265081)

[3.2.1 输入与输出 - 4 -](#_Toc113265082)

[3.2.2 编码 - 5 -](#_Toc113265083)

[3.2.3 向量化 - 5 -](#_Toc113265084)

[3.2.4 位置信息编码 - 6 -](#_Toc113265085)

[3.2.5 注意力机制 - 6 -](#_Toc113265086)

[3.2.6 解码 - 6 -](#_Toc113265087)

[**第4章 GPT-3的结果展示** - 7 -](#_Toc113265088)

[4.1 生成应用和布局 - 7 -](#_Toc113265089)

[4.2 搜索和数据分析 - 7 -](#_Toc113265090)

[4.3 程序生成与分析 - 7 -](#_Toc113265091)

[**结 论** - 8 -](#_Toc113265092)

[**参考文献** - 9 -](#_Toc113265093)

**第1章 GPT-3的应用简介**

## 1.1 GPT-3的多种应用可能

自从OpenAI开放GPT-3的API以来，不少人相继在社交网络上推送自己的试用成果，引发了人们对GPT-3的热烈讨论，包括用GPT-3生成网页、图表、代码、文本以及推理。其中令人吃惊的是，GPT-3能生成Keras编写的卷积神经网络的代码，还能应对程序员面试、回复邮件、写积分表达式、回答物理问题。

### 1.1.1 GPT-3将成为工作的助手

随着GPT-3成为主流，它将为所有白领专业人士提供帮助，无论他们从事什么工作。这意味着生产力和创造力的新的文艺复兴时代。只需给出GPT-3相关提示，即可将想法变为现实，创建代码，撰写论文，生成方案。

### 1.1.2 GPT-3使我们进入人工文本生成的新时代

GPT-3将能够加快工作流程，帮助您产生想法，编写电子邮件，回复查询，将文本翻译成其他语言并提供灵感。它甚至可能带来新的写作指导。

我们现在生活在一个以机器和计算机为主的世界中。从Google搜索到Netflix推荐的算法已经占据了我们的生活。我们的设备已成为我们身体的延伸。智能扬声器和语音助手（例如，亚马逊的Alexa，苹果的Siri，谷歌的助手，三星的Bixby和微软的Cortana）正逐渐成为我们日常生活的一部分，因为他们每天都在学习和变得更聪明。随着语音革命的发展，这些设备有望获得更多的功能，甚至可以充当教师和心理学家。Siri现在可能很愚蠢，但是在部署GPT-3之后，我们将见证更加智能的语音助手。

**第2章 GPT-3的内容简介**

## 2.1 GPT-3是什么？

GPT-3代表“Generative Pre-trained Transformer 3”，是一种基于深度学习原理的语言预测模型。GPT-3是基于上下文的生成AI系统，当向GPT-3提供提示或上下文时，它可以填写其余内容。如果开始撰写文章，它将继续撰写文章。当写邮件时，知道会收到一些自动完成建议。GPT-3可以看作是一个巨大的自动完成程序。

GPT-3算法在1750亿个参数的基础上运行。每个参数都是一个公式，可帮助GPT-3算法做出精确的预测。使用云计算分析了45 TB的数据。完整的英语维基百科仅占GPT-3培训数据的0.6％。这意味着GPT-3已访问了迄今为止生成的许多Internet，新闻，故事，数字书籍，论坛帖子，小说，社交媒体，博客，手册，代码和人类文学。GPT-3掌握了所有这些知识，并获得了很多时间和资源来咀嚼这些资源。

由于GPT-3巧妙地在巨大的数据集中发现了模式，概率和一致性，因此它可以执行我们认为迄今为止尚无法实现的各种复杂任务。GPT-3是无监督的学习者。它已经学会并自己捡起了一切。可以创建原创文章，散文，诗歌，对话，模因，计算机代码等等。

## 2.2 GPT-3有什么限制？

尽管GPT-3令人印象深刻，但它甚至还不接近人类的智能和能力，并且它仍然有很多局限性。尽管具有175B参数，但GPT-3不能牢记长期目标。它无法在许多段落中保持逻辑一致的上下文。GPT-3没有自己的思想或语义，因为它不了解输入和输出的真正含义。它没有心智模型，记忆力或意义感。它计算相关性，但不能推理。

另一个巨大的限制是GPT-3始终会产生合理但不正确的输出。因此，GPT-3无法信任。不幸的是，它可以产生无限数量的似是而非的假新闻或网络钓鱼诈骗。这意味着GPT-3可能会使我们的生活比现在复杂得多。

OpenAI的Sam Altman也通过他的以下推文强调了这些局限性“GPT-3的宣传太多了。令人印象深刻，但仍然存在严重的弱点，有时还会犯非常愚蠢的错误。人工智能将改变世界，但是GPT-3只是很早的一瞥。我们还有很多事情要解决。”

## 2.3 GPT-3的含义

GPT-3现在已经成为世界的时代精神。当GPT-3成为每个可以访问Internet的人日常使用的工具时，将会发生什么？ GPT-3显然具有巨大的商业潜力。因此，整个行业和领先的跨国公司都试图加入潮流。

我们正在见证另一项突破性创新。任何人都可以使用此功能强大的语言工具来发挥创造性和概念性工作的作用。

随着摩尔定律的放慢，我们看到了一系列新技术的出现，包括GPT-3，大数据，云，量子计算，机器学习和人工智能。这些新技术将在未来几十年推动指数技术的新时代。

**第3章 GPT-3的研究过程及原理**

## 3.1 GPT-3研究过程介绍

最开始的GPT一代相对训练数据集比较小，它仅在几千本书和一台8 GPU机器上进行了培训。GPT-2则极大地扩展了训练的范围，它抓取了国外著名的论坛Reddit上3星以上的帖子作为训练数据，这样GPT-2的训练数据较前一个版本多了至少10倍。虽然这个该数据集相对还是有限的，但是已经有惊人的效果。比如2019年2月份OpenAI展示了一个用GPT-2生成文本的模型，用户只需要输入一段话，这个模型就能将这段话引申成一个新闻，这个新闻足以达到以假乱真的程度。这个引起了广泛的关注，当然也有很多人反对。但是OpenAI并没有停止他们的脚步，2020年7月又向一部分AI的研究人员开放了他们最新的GPT-3的beta测试版。

在规模方面，GPT模型再次大幅扩展，达到了1,750亿个内部节点，而GPT-2仅为15亿，是后者规模的116倍。GPT-3用了500G的书，文章和wiki进行训练，单词数量超过一万亿个单词。这个模型看过的书比我们任何人看过的书都多得多。而这么大规模训练的结果是，参与测试的研究人员在人使用GPT-3后，几乎都是惊叹于这个网络模型的强大。GPT-3惊人的能力来自于其海量的学习，将有监督的学习与无监督的结合，GPT-3采用最近比较流行的transformer模型。将采集到的文本里面的单词或短语随机删除，模型使用上下文试图填充删掉的单词或短语，通过这种训练方式就不需要人工最海量的数据进行训练，这样才使得GPT-3的大规模训练变得可能。

## 3.2 GPT-3模型简介

### 3.2.1 输入与输出

输入是N个单词（Token）的序列。输出是对最有可能在输入序列末尾放置的单词的预测。所有基于GPT模型的应用，如对话，故事和示例生成，都是通过这种简单的输入输出方案实现的：给它一个输入序列–得到下一个单词。比如：

***我在这里等着 -> 你***

得到下一个单词后，将其添加到序列中，得到下一个单词。比如：

***我在这里等着你 -> 来***

***我在这里等着你来 -> 看看***

***我在这里等着你来看看 -> 我的***

***我在这里等着你来看看我的 -> 微博***

重复所需的次数，最终得到需要的长文本。实际上还有两点要求：

1.输入序列实际上固定在2048个字（对于GPT-3）以内。仍然可以将短序列作为输入：只需用“空”值填充所有额外位置。

2. GPT输出不仅是一个预测，而是一系列预测每个可能单词的概率。序列中每个“next”位置都是一个预测。但是在生成文本时，通常只查看序列中最后一个单词的预测。

### 3.2.2 编码

GPT实际上无法理解单词。作为一种机器学习算法，它可以对数字向量进行运算。

第一步是将所有单词收集在一起构成词汇表，这使我们能够为每个单词赋予一个值（id）。比如Aardvark是0，aaron是1，依此类推，GPT的词汇表包含50257个单词。最终，我们可以将每个单词转换为大小为50257的单点编码矢量，其中仅索引i处的维（单词的值）为1，所有其他维度均为0。对序列的每个单词如此操作，得到一个2048 x 50257的1和0矩阵。

### 3.2.3 向量化

50257对于矢量来说相当大，并且大部分都用零填充。那是很多浪费的空间。为了解决这个问题，模型学习了一个embedding函数：一个神经网络，该神经网络采用50257长度的1和0的向量，并输出n长度的数字的向量。在这里，模型试图将单词含义的信息存储（或投影）到较小的空间中。例如，如果embedding维数为2，就好比将每个单词存储在2D空间中的特定坐标处。

从现在开始，将绘制二维矩阵，将其作为块，并在其旁边标注尺寸。如果适用，将矩阵行分开以明确表示每一行对应于序列中的一个单词。

### 3.2.4 位置信息编码

为了对当前Token在序列中的位置进行编码，作者采用了Token的位置（标量i，在[0-2047]中），并将其传递给12288个正弦函数，每个函数的频率都不同。对于每个Token，结果是12288个数字向量。与embedding一样，将这些向量组合成具有2048行的单个矩阵，其中每一行是序列中Token的12288列位置编码。最后，可以简单地将具有与序列embedding矩阵相同形状的该序列位置编码矩阵添加到该矩阵中。

### 3.2.5 注意力机制

Attention的目的是：对于序列中的每个输出，预测输入标记对输出影响程度。在这里，想象一个由3个Token组成的序列，每个Token都由512个值的embedding表示。该模型学习3个线性投影，所有这些投影都应用于序列embedding。换句话说，学习了3个权重矩阵，这些矩阵将序列embedding转换为3个单独的3x64矩阵，每个矩阵分别用于不同的任务。

现在，在作者提出的GPT模型中，他们使用了多头注意力。所有这些意味着，上述过程被重复了很多次（GPT-3中为96x），每个过程都有不同的可学习的query，key，value投影权重。

每个attention head的结果（单个2048 x 128矩阵）被串联在一起，产生2048 x 12288矩阵，然后将其乘以线性投影（不会改变矩阵形状），以达到良好的效果。

### 3.2.6 解码

通过所有96层GPT-3的注意力/神经网络机制后，输入数据已处理为2048 x 12288矩阵。对于该序列中的2048个输出位置中的每一个，该矩阵都应包含一个12288个向量的信息，其中应显示哪个单词。

此时可以反转embedding部分将单词转换为向量的映射，以将输出的12288向量矢量embedding转换回50257字编码。

**第4章 GPT-3的结果展示**

## 4.1 生成应用和布局

1.根据描述生成HTML布局和代码

2.根据描述创建UI设计

3.根据描述生成React代码创建待办事项清单应用

4.仅基于React变量名称生成component

5.根据颜色名称或表情符号生成色阶

6.根据描述创建网站

## 4.2 搜索和数据分析

1.问题解答和搜索引擎

2.扩充表中的信息

3.根据描述创建图表

4.根据描述生成代码并转换为电子表格

5.根据描述生成图表和代码

## 4.3 程序生成与分析

1.根据描述生成shell命令

2.阅读代码并回答相关问题

3.根据描述生成LaTeX表达式

4.根据问题描述生成SQL代码

**结 论**

GPT-3毫无疑问是NLP领域乃至AI领域取得重大突破的一项工作，甚至可以应用到数以百万计的上班族日常使用的生产力工具中。1750亿参数的超大规模，使得语言模型具备了生成难辨真假的新闻文章的能力。此后，人工智能将不会显得那么的智障了，而将真正的展现出一定的智能。

**参考文献**

[1] Eliza Strickland. IEEE Spectrum: OpenAI's GPT-3 Speaks! (Kindly Disregard Toxic Language).

[2] Chris Verdence. Be Unique: A Concise Description of GPT-3.

[3] Frank Chen and Sonal Chokshi. 16 Minutes on News #37: GPT-3, Beyond the Hype (Jul. 2020).