【知识福利社】你的私人书院，微信：312005650 全网热门课程众筹共享！

AI
ChatGPT
进入课程 >
时长：12分31秒 大小：17.18MB
00:00 / 12:31
15
15
宝木 转述

第3讲 语言模型是如何进化的？

前沿课·吴军讲GPT
今天

🍅，你好！

欢迎来到《前沿课·吴军讲GPT》。

上一讲我们说了，ChatGPT的本质是语言模型，就是对人类的语言建立数学模型，也解释了语言模型的工作原理。今天各种语言模型的底层技术，其实出现在50年以前。

不过，不用我说你也知道，像GPT-3和GPT-4这样的语言模型，已经比早期的语言模型强大太多了。如同汽车、移动电话和计算机在不断进步一样，语言模型诞生之后也在持续进步。

这一讲，我们就讲讲语言模型进化的历史，说说自然语言处理的要点。

我知道，大部分同学不会做这方面的工作，但是了解一下自然语言处理的发展趋势和当前方法的局限性，可以帮我们理性看待ChatGPT，消除对于人工智能的迷信。

内容禁止商业，仅供社群内部免费交流使用
【知识福利社】微信:312005650

阶段一

语言模型发展的第一个阶段是上个世纪90年代之前，当时，贾里尼克等人用它解决了语音识别的问题。

在此之前，计算机只能识别百十来个英语单词，识别的错误率还非常高。在采用通信中的算法，并且引入语言模型之后，IBM的语音识别系统可以识别22000个英文单词，而且错误率从过去的30%左右降低到了10%以内。

这意味着什么？当识别的错误率控制在10%以内，也就是每识别十个词出错不到一个词的时候，就有应用价值了。因为这时候，人可以根据自己的知识纠正那一个错字。

这件事，对当时的一个年轻人触动很大，他就是还在卡内基梅隆大学读博士的李开复。

在征得导师同意后，李开复决定抛开卡内基梅隆大学所走的传统的人工智能道路，利用IBM的方法实现一个实用的语音识别系统。后来他做到了，并且成为当时美国的科技新星，还到CBS电视台向观众展示了他打造的语音识别系统。随后，很多家公司高薪挖他，这让他彻底离开学术界，进入工业界。

但是公平地讲，当时的语音识别系统离实用还差得远，和今天的更是完全无法相比。不过，这件事起码证明了，语言模型确实有用。

上世纪90年代还发生了一件事，就是贾里尼克的一个下属——彼得·布朗——利用语言模型实现了英语和法语之间的自动翻译。

相比语音识别，机器翻译的难点在于要准备两种语言的语言模型，而且最好利用两种语言一一对应的句子来做训练。如果你用狄更斯的英语小说和雨果的法语小说来训练，它们对应不上，效果自然就不会好。由于当时计算机上一一对应的多语种文本只有《圣经》和少量的联合国文件，因此布朗的翻译系统效果并不好。

不久之后，布朗干脆离开学术界，成为著名对冲基金文艺复兴技术公司的技术负责人，他建立数学模型解决问题的本事让基金和他本人都挣了很多钱。

在当时，布朗的成果并没有受到学术界的重视。直到十多年后，德国科学家弗朗茨·奥科在Google用类似的方法比较好地解决了机器翻译问题，布朗最初的那篇论文才成为那个领域被引用数量最多的论文之一。当然，奥科也用了比当时其它研究机构多上万倍、比布朗多上百万倍的数据。不过，这都是后话了。

阶段二

说回语言模型。

进入到90年代后，人们发现，单纯增加训练数据的做法并不是十分有效。贾里尼克自己做了一个统计，发现不论数据如何增加，都会有大量的语言学现象覆盖不了。于是，人们开始考虑用语言中更深的知识和信息建立语言模型。语言模型的发展由此进入到了第二阶段。

我就是在那个背景下进行语言模型研究的。当时，我们用了语言中的语法和语义信息，可以在不增加数据量的情况下大大提高语言模型的能力。

比如，读音相似的词，在不同的主题中，出现的可能性相差巨大。举个例子，和"事实"（不含音调）同音的词就有四十多个，在过去，到底是哪一个就很不好确定。但是加入语法和语义信息后，我们就知道——"时事""事实""失实"这些词会在时政新闻中经常出现，而"失事""史实""诗史"等词会在历史题材的文章中经常出现。很多时候，确定了主题，就知道该使用什么词语了。

当语言模型加入了语法和语义信息后，语言模型的训练，以及使用它计算概率时，就不再是简单的统计了。

在此之前的语言模型，稍微有一些概率论和统计学背景，学过一门自然语言处理课程的人基本都能构建。但是此后的语言模型，公式非常复杂，绝大部分人是搞不懂的。

当然，**把更多信息，比如语法信息、语义信息，加入语言模型后，人们面临的更大的挑战是，训练模型的计算量变得特别大。**

关于训练模型的计算量，我们后面还会讲到，这里举一个小例子：

在我之前，IBM有一位科学家，他是世界上第一个成功利用语义信息训练语言模型的人。他主要的工作，不是改进算法或者数学模型，而是写了一个程序，能够把IBM沃森实验室所有空闲的服务器都用上，然后计算了一年，完成了模型的训练。

要知道，沃森实验室是一个有上千名科学家的大型研究机构。这么大的研究机构，这么多的服务器，在空闲时间计算一年，你想想这是多大的计算量。

由于摩尔定律的作用，也就是计算机处理器的性能每18个月翻一番，我们有能力进行越来越复杂的计算，能处理越来越多的数据，也能利用和存储更多的信息。但是，这个计算量的增长是超过计算能力的提升的。

比如，2012年，我们在Google做机器问答时，就需要对几百亿个句子做深入的语法分析，提取各种有用的信息，而不是简简单单地进行统计。这个计算量，即便是Google也感到很吃力。

今天，ChatGPT的语言模型利用的语言信息更加丰富，计算量也更大，即使利用大量的计算能力更强的GPU，计算的成本也很高。

除了挖掘语言中更多的信息，从上世纪90年代起，科学家还提出了自适应的语言模型。**也就是说，一个人使用语言模型的过程中，他个人输入的信息和间接提供的反馈信息，能够被用来改进语言模型。**

比如，我们用语言模型构建一个拼音输入法，使用者在不断输入中文时，他输入的中文文本就被用来做训练数据了。这样一来，这个输入法软件就会越用越聪明。

当然，这种做法不仅需要个人数据，还需要大量的存储空间。因为这时候，不同的人使用的语言模型就已经不同了。

今天的ChatGPT其实就有这个功能，它会根据用户的输入不断迭代优化。你如果用它写几篇文章，然后不断对那些文章进行修改，它就会学到你的行文习惯，然后根据你的行文习惯输出内容。但是，如果人们刻意误导它，输入垃圾信息，它输出的文本的质量也会跟着下降。

刚才说了，挖掘语言中更深层的信息，能有效提高语言模型的能力，但与此同时，也让计算量变得巨大。因此，为了让计算机更有效率地工作，在2000年之后，很多语言模型的研究工作就转到提高算法效率上了。

比如，我在博士工作的后期，花了一年多的时间改进训练方法，将计算量减少了两到三个数量级。也就是说，只需要之前千分之一到百分之一的计算资源，就能做同样的事情。

阶段三

到了2010年前后，Google开发了深度学习的工具Google大脑：**一方面，能够更有效地利用计算资源，这使得语言模型能够越做越大；另一方面，也让模型计算出的概率越来越准确。**语言模型的发展由此进入第三阶段。

在随后的几年里，Google做了一件事，就是把之前的语言模型重新训练一遍。在训练数据没有增加的情况下，模型的准确性提高了很多，相应产品的质量也提高了。

由于数据量的增加和语言模型准确率的提高，到了本世纪的第二个10年，计算机就逐渐做到了从写一句话到写一个完整的段落。

在此之前，无论是语音识别，还是机器翻译，你都需要给它一个完整的句子，让它生产一个具有同等信息量的句子。比如，你给它一句中文，它产生相应的英文；或者你给它一个句子的语音，它产生相应的文字。输入输出是1:1。

而当语言模型足够强大之后，你给它一点输入，它就可以产生大量输出了。

比如，你问它一个问题，可能只有一句话，它生成的答案可能有一大段文字。你给它几个关键词，让它生成一份简历，它可能会给你产生一页纸的内容。

了解信息论原理的同学肯定会知道，输入的信息量少、输出的信息量多，这中间的差异必须有办法给补回来，否则输出的内容里一定会充满不确定性。

补充的信息从哪里来呢？实际上，都是存在于语言模型中的，需要先把信息输入语言模型。这时的语言模型就是所谓产生式的了。GPT中的G，就代表Generative，产生。

比如，GPT-3的训练数据就包括几十万本图书、几乎全部维基百科的内容。这么多的数据还只是它所用数据的五分之一不到。实际上，单是维基百科的内容，几乎就能回答今天绝大部分人提出的所有有意义的问题了。因此，ChatGPT的表现比很多人好并不奇怪，因为人一辈子学习的内容可能连它的千分之一都没有。

当然，光有信息还不够，为了保证一段文字中各个句子之间的连贯性，还需要语言模型能够计算出几个句子出现在一起的概率。在段落这个层级进行计算，要比在句子这个层级计算，计算量大出成千上万倍，而不是几倍几十倍。

同时，为了让语言模型产生的段落不太发散，通常还要用一些模版对生成的文字进行限制。这也是为什么ChatGPT写汇报和简历能写得很好，写不受限制的文章就显得非常枯燥的原因。

类似的，今天的写作软件通常只能应付有套路的写作要求，比如写财经报道、写大学申请材料等。所谓的"套路"，其实就是计算机使用的模版。今天，美国的作家群体普遍认为，ChatGPT只能按照一定的模版写作。

当然，要让ChatGPT能够回答好问题，或者和人进行对话，还需要让它理解人的问题。这件事相对写作本身要容易一些。

今天，**计算机对自然语言做语法分析，可以做得比人还准确。通过句子的语法结构，可以理解常见的实词，也就是名词、动词和形容词的含义。**通常来说，句子的含义是这些词的含义的叠加，于是，它就能理解这句话的含义。

当然，在丰富的人类语言中，也经常有例外，就是一句话中所有的词你都懂，但是这句话听不懂。对于这种情况，计算机也没有办法，这时候，它就会胡乱回答，甚至给出完全相反的答案。

下节预告

好，这一讲就到这里。

语言模型的算法其实不仅能处理和自然语言相关的问题，从理论上讲，用其它数据建立类似的模型，就能处理大量的、原本需要利用人类智能才能解决的问题。这些内容，我们下一讲再讲。我是吴军，我们下一讲见。

划重点　　添加到笔记

语言模型的发展大致经历了三个阶段：
1、上个世纪90年代之前，当时贾里尼克等人用它解决了语音识别的问题。
2、90年代后，语言模型加入了语法和语义信息，还出现了自适应的语言模型。
3、2010年前后，Google开发了深度学习的工具Google大脑，能够更有效地利用计算资源，也让模型计算出的概率越来越准确。

《前沿课·吴军讲GPT》
成为搞懂GPT的少数人
版权归得到App所有，未经许可不得转载
作家，计算机科学家 吴军

首次发布：今天

上一篇 第2讲 ChatGPT的本质是什么？

可免费赠送给20位朋友
微信　朋友圈

用户留言

最新　最热　只看作者回复

逸鹏
6小时前　+ 关注
老师，ai会一本正经胡说八道的这个特性，会不会导致网上搜索信息的模糊化（充斥着ai的回答）对于我们非专业人士怎么鉴别？
2　2　2　分享
作者回复：
简单地讲，很难判断！现在有一本很好的书《现实+》谈到这个问题讲得很好，包括谈到如何判断。但是它也给出结论，答案是非常难。不过它给出了一个提示，利用AI造假，背后常常有商业目的。

淡然
7小时前　+ 关注
Ghatgpt是工具不是神，使用者必须对于使用场景有一定的认知和理解，
换个角度，可以把ChatGpt当作一个学习的工具，如果想对一个新的领域进行深入了解，可以让Gtp先做最基础部分的调研。带着自己的问题和认知去论证，也可以少踩坑。
转发　评论　赞　分享

Zongchen
7小时前　+ 关注
吴老师，数学是我们宇宙的基本原理。那么人类的思想活动是不是也是基于某种算法？那么人类每天消耗2000卡就能够做到的事情，计算机消耗几十万度电是不是可以比人类做得更好？
转发　1　赞　分享
作者回复：
不能拿橘子和苹果对比

曼尼蓝
7小时前　+ 关注
吴老师好，我理解现在chatgpt和人对话，其实是应答式的。我假想一下如果它能主动和我们对话，至少也得我们先开腔点醒，它才能感知有一个对话个体。但是另一方面，机器人索菲亚似乎已经能主动思考，有表情了。我想知道这两者他们运行的底层代码是否差异不大。如果它们两者结合会不会有电影里那样强大的机器人，以至于远远超过人类。
收起
转发　6　赞　分享
作者回复：
不是它聪明，是你可能太。。，我不说了。

Daniel
7小时前　+ 关注
输入的质量决定了输出的质量
GPT学习小组
转发　评论　1　分享

佰明
7小时前　+ 关注
现在LLM有一个问题，就是回答问题时的知识不一定是正确的（"一本正经胡说八道"的幻觉问题）。
请问老师有没有什么思路能够来保证这个回答的正确性？或者退而求其次能提高回答的正确概率。
转发　1　赞　分享
作者回复：
没有

午时已到
7小时前　+ 关注
吴军老师您好，就我对世界观察，很多发明的事物，在使用过程中都产生了很多超出预期的结果（由于自身认知水平较低，一时间举不出一个很好的例子），因为人类一开始就以自己受限的认知，去框定了这个发明能做的什么，什么不能。但是当有出乎预料的结果产生的时候，人们就会惊呼：哇原来还能这样！
所以我对所谓的"涌现"一点也不惊讶，这种事情发生了无数次。
收起
转发　2　3　分享

夏虫冰语
7小时前　+ 关注
一切的恐惧、焦虑和盲从都来自于无知，不断更新认知方能淡定从容。
转发　评论　2　分享

赵棹
8小时前　+ 关注
2023年4月27日 吴军老师讲GPT
语言模型的进化:语音识别，听懂人的话；建立模型，给出答案；自适应，模型在应用中自我学习优化。
语言模型，是基于大量的公式变量，通过运算输出结果。公式变量是按照人的思维方式设置，千人千面，更有效需要人机互相滋养。
　　模型举例:和"事实"（不含音调）同音的词有四十多个，加入语法和语义变量因子，"时事""事实""失实"这些词会在时政新闻中经常出现，而"失事""史实""诗史"等词会在历史题材的文章中经常出现。

很有思考个性的人，通过GPT找到答案的难度会比较大。
同时。经过不断的学习、迭代，语言模型距离全覆盖目标，会更加接近。
收起
转发　评论　1　分享

Murphy
8小时前　+ 关注
哈哈，最后那句：所有的词都懂，组成的那句话不懂。我之前看文言文的时候，经常有这感觉。
转发　评论　赞　分享

机械芭蕉零零零
9小时前　+ 关注
说起来在跟4学数学的时候 问它数学入门它第一本推荐的就是数学之美...
转发　评论　赞　分享

小灰
10小时前　+ 关注
不管什么时候，问出好问题，筛选有效的信息的能力好像都是基层能力
转发　评论　1　分享

Jack
11小时前　+ 关注
吴军老师，您如何看待涌现这个现象？是否说明大型语言模型有了创造力？如果说大型语言模型其实本质上只是在计算概率，那么是否永远也不可能有创造力？（按照现在很多人的理解，涌现的意思是当一个复杂系统复杂到一定的程度，就会发生超越系统元素简单叠加的、自组织的现象。比如单个蚂蚁很笨，可是蚁群非常聪明；每个消费者都是自由的，可是整个市场好像是有序的；每个神经元都是简单的，可是大脑产生了意识……有些人认为大型语言模型，也会涌现出各种意想不到的能力。
2022年8月，谷歌大脑研究者发布一篇论文，专门讲了大型语言模型的一些涌现能力，包括少样本学习、突然学会做加减法、突然之间能做大规模、多任务的语言理解、学会分类等等……而这些能力只有当模型参数超过1000亿才会出现——
研究者并没有刻意给模型植入这些能力，这些能力是模型自己摸索出来的。
就如同孩子长大往往会出乎家长的预料。）
收起
转发　1　7　分享
作者回复：
分类，加减法早就会做了，只是没有发表，没有什么突然。

怪医浮士德
11小时前　+ 关注
对于"两位刚毕业的从事机器学习的博士和我讲了如何更好地利用ChatGPT写作"这个话题很感兴趣，非常期待了解更多的相关内容。听过吴军老师过去的一些课程，所以对计算机算法的基本原理略有了解，所以并不会觉得ChatGPT有多可怕。相对于AI威胁论，我个人更倾向于了解更多关于如何运用最新工具提升自己的思维意识、生活工作质量方面的信息，而比较警惕的是人能使用这些新工具对其他人做出来的事情。所以也想要问一下，在运用ChatGPT这个工具的过程中，是否有哪些潜在风险，如果有，我们该如何合理地保护自己。非常感谢~
收起
转发　评论　3　分享

贾旭
12小时前　+ 关注
请问老师，一些暂时还没有数据的领域，比如建筑施工，在从无到有积累数据的过程中，有什么底层原则么？

我能想到的，只有：
1. 开发可通过累积数量以满足需求的"标准接口"
2. 标准接口在物理上只有0和1两种状态的
3. 对接口编址并分类（类似快速排序的思路）
4. 对应用模块化（就是类似苍蝇飞行的逻辑，随意找一个方向，向前飞，如果碰到障碍，再随意找一个新方向向前飞，降低对算力和电的依赖）
5. 连接过程即数据收集，记录，更新过程，同时也确认安装合格与否
6. 连接点设计需要足够的冗余，既出于安全，也出于对安装过程和日后维护的考虑
7. 安装过程运作同一逻辑，可以通过不断重复完成
收起
转发　2　6　分享
作者回复：
做好你的本职工作，不要想多了。

甜小姐
14小时前　+ 关注
是的，我觉得ai对于人类世界的复杂性，特别是对于人性的理解是赶不上人类的，这就限制了它只能做纯数据分析战略，而不能像诸葛亮那样运用心理战略，比如，空城计。

打个比方，我有个日本同学，很喜欢问每个人要他们写的论文看。也许，他只是想学习一下，但对于没有发表的论文，我觉得随便给人看是有风险的。因为你不知道他会拿你的文章做什么，申请什么，参加什么，或者修修改改变成自己的东西。而且万一有问题，你可能有连带责任，以为你帮他cheat.比如，我一个同学的文章被另一个抄了，结果，两个人都零分。

我就问chatgpt,如果有同学要看你的文章怎么办？它一本正经地说，你检查好，附件粘好，发给他。确认收件人，附件及地址的正确。所以，它还是那种人家说的直男思维，你让洗碗，它就只洗碗，不洗锅。😂然后，我说，你有想过blabla的风险伐？然后，它说，你可以在文档上加个水印。这能有什么用？😂所以，还是too simple的思维，不够了解人类世界的复杂。从这点来说，用它做决策是代替不了人的心思的。

收起
转发　10　41　分享
作者回复：
谢谢你的留言

张伟明
15小时前　+ 关注
联想到了尤瓦尔赫拉利的《人类简史》里的数据主义
转发　评论　赞　分享

周明智
16小时前　+ 关注
学习的语言模型的背景和来源，这样不会被神话和恐惧
转发　评论　3　分享

Peter Ale
16小时前　+ 关注
不得不说现在ChatGPT写一些通用的文书类的时候，比我们很多人都写得好。如果有特殊文殊风格的训练素材估计写也能够学得很好。同样有一个人要写某一种类的文书，他也需要相应的素材来学习。图灵奖得主Yung Lecun一直说没有通用AI一说，因为连人都不通用，一个人从事某一项工作，都要受到相应的训练，才能真正的做好该类工作。
收起
吴军来信
1　2　6　分享
作者回复：
可能是那些人写得太烂，大家要加油啊！否则真的失去了存在的意义。

申彦
16小时前　+ 关注
1、一个人使用语言模型的过程中，他个人输入的信息和间接提供的反馈信息，能够被用来改进语言模型。
2、计算机通过句子的语法结构，可以理解常见的实词，也就是名词、动词和形容词的含义，从而理解这句话的含义。
3、在第三阶段，深度学习工具能够更有效地利用计算资源，使得语言模型能够越做越大，也让模型计算出的概率越来越准确。
收起
GPT学习小组
转发　评论　3　分享

以上留言由作者筛选显示

【知识福利社】你的私人书院，微信：312005650 全网热门课程众筹共享！