【NLP】2019年深度学习自然语言处理最新十大发展趋势

专知 产业智能官 Yesterday

【导读】自然语言处理在深度学习浪潮下取得了巨大的发展，FloydHub 博客上Cathal Horan介绍了自然语言处理的10大发展趋势，是了解NLP发展的非常好的文章。

https://blog.floydhub.com/ten-trends-in-deep-learning-nlp/

2018年是基于深度学习的自然语言处理(NLP)研究发展快速的一年。在此之前，最引人注目的是Word2Vec，它于2013年首次发布。

在此期间，深度学习模型在语言建模领域实现的方面出现了一种稳定的创新和突破的势头。

然而，2018年可能是所有这些势头最终结出硕果的一年，在NLP的深度学习方法方面出现了真正突破性的新发展。

去年的最后几个月，随着BERT模型的出现，出现了一场特别热闹的研究浪潮。2019年，一个新的挑战者已经通过OpenAI GTP-2模型出现，该模型“太危险”不适合发布。通过所有这些活动，很难从实

际的业务角度了解这意味着什么。

这对我意味着什么?

这项研究能应用于日常应用吗?或者，潜在的技术仍在如此迅速地发展，以至于不值得花时间去开发一种可能会被下一篇研究论文视为过时的方法?如果您想在自己的业务中应用最新的方法，了解NLP研究的趋势是很重要的。为了帮助解决这个问题，基于最新的研究成果，在这里预测10个关于NLP的趋势，我们可能会在明年看到。

NLP架构的趋势

我们可以看到的第一个趋势是基于深度学习神经网络架构，这是近年来NLP研究的核心。为了将它们应用到您的业务用例中，您不必详细地了解这些架构。但是，您需要知道，对于什么架构能够交付最佳结果，是否仍然存在重大疑问。

如果对最佳架构没有共识，那么就很难知道应该采用什么方法(如果有的话)。您将不得不投入时间和资源来寻找在您的业务中使用这些体系结构的方法。所以你需要知道2019年这一领域的趋势。

1. 以前的word嵌入方法仍然很重要

2. 递归神经网络(RNNs)不再是一个NLP标准架构

3. Transformer将成为主导的NLP深度学习架构

4. 预先训练的模型将发展更通用的语言技能

5. 迁移学习将发挥更大的作用

6. 微调模型将变得更容易

7. BERT将改变NLP的应用前景

8. 聊天机器人将从这一阶段的NLP创新中受益最多

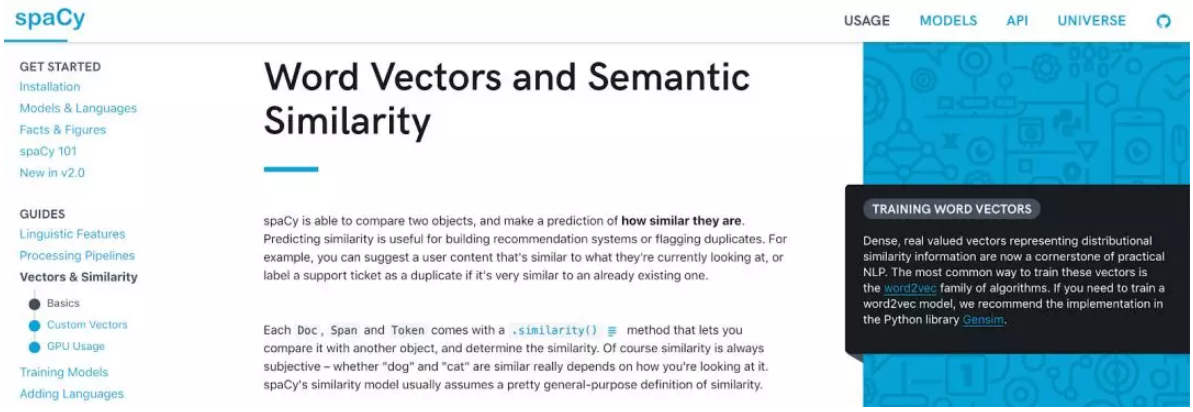
9. 零样本学习将变得更加有效

10. 关于人工智能的危险的讨论可能会开始影响NLP的研究和应用

1. 以前的word嵌入方法仍然很重要

Word2Vec和GLoVE是在2013年左右出现的。随着所有的新研究，你可能认为这些方法不再相关，但你错了。Francis Galton爵士在19世纪后期提出了线性回归的方法，但作为许多统计方法的核心部分，它今天仍然适用。

类似地，像Word2Vec这样的方法现在是Python NLP库(如spaCy)的标准部分，在spaCy中它们被描述为"实用NLP的基石"。如果你想快速分类常见的文本，那么word嵌入就可以了。

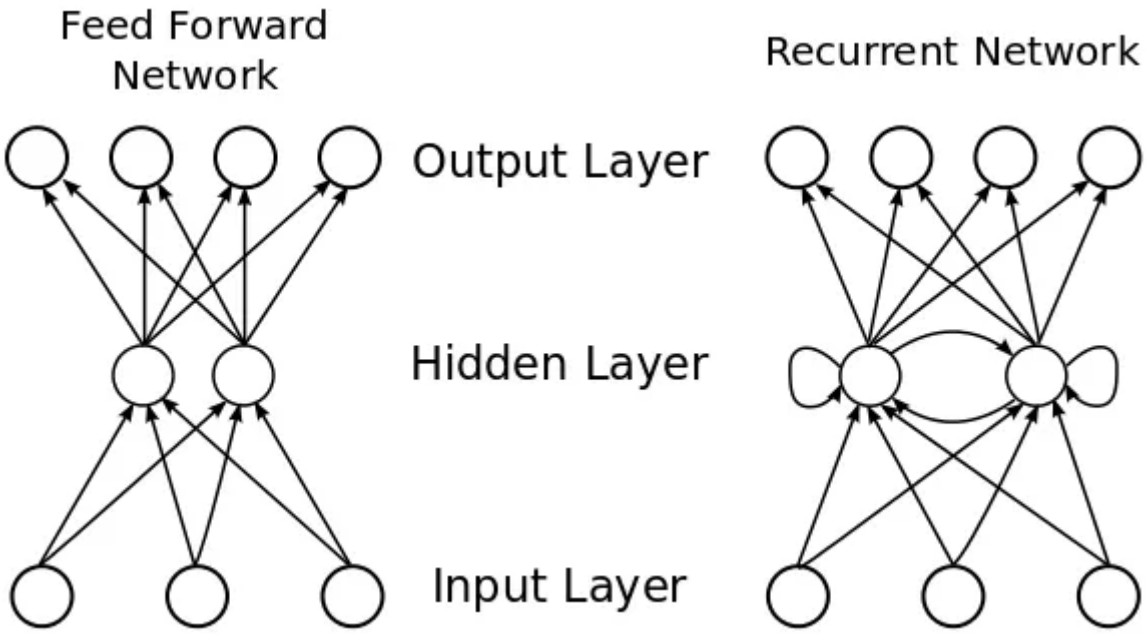


Word2Vec等方法的局限性对于帮助我们了解NLP研究的未来趋势也很重要。他们为所有未来的研究设定了一个基准。那么，他们在哪些方面做得不够呢?

- **每个词只能嵌入一个词**，即每个词只能存储一个向量。所以" bank "只有一个意思"我把钱存进了银行"和"河岸上有一条漂亮的长凳"
- 它们**很难在大型数据集上训练**
- 你无法**调整它们**。为了使他们适合你的领域，你需要从零开始训练他们
- 它们不是**真正的深度神经网络**。他们被训练在一个有一个隐藏层的神经网络上。

2. 递归神经网络(RNNs)不再是一个NLP标准架构

长期以来，RNNs一直是基于NLP的神经网络的基础架构。这些架构是真正的深度学习神经网络，是从早期的创新(如Word2Vec)设定的基准发展而来的。去年讨论最多的方法之一是ELMo(来自语言模型的嵌入)，它使用RNNs提供最先进的嵌入表示，解决了以前方法的大多数缺点。从下图中可以看出，与前馈网络不同，**RNNs允许隐藏层的循环返回到它们自己，并且以这种方式能够接受可变长度的序列输入**。这就是为什么它们非常适合处理文本输入。

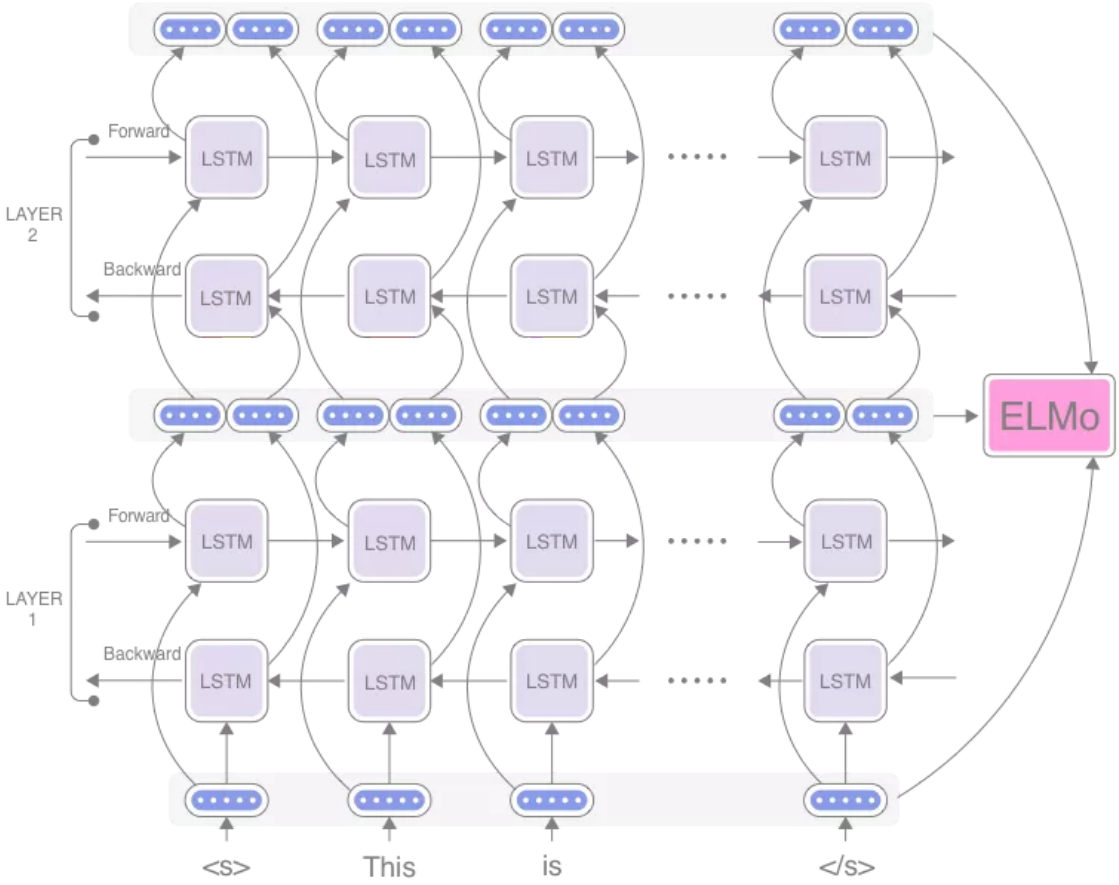


RNNs非常重要，因为它**提供了一种处理数据的方法，而时间和顺序非常重要。**例如，对于文本相关的数据，单词的顺序很重要。改变语序或单词可以改变一个句子的意思，或只是使它乱语。在前馈网络中，隐含层只能访问当前输入。它没有任何其他已经处理过的输入的“内存”。相比之下，RNN能够对其输入进行“循环”，看看之前发生了什么。

作为一个实际的例子，让我们回到我们的一个bank的例句，“I lodged money in the bank”。在前馈网络中，当我们到达“bank”这个词时，我们对之前的词已经没有“记忆”了。这使得我们**很难知道句子的上下文，也很难预测正确的下一个单词。**相比之下，在RNN中，我们可以参考句子中前面的单词，然后生成下一个单词是“bank”的概率。

RNNs和长短时记忆(LSTM)是RNN的一种改进类型，它们的详细信息不在本文讨论范围之内。但如果你真的想深入了解这个主题，没有比克里斯托弗•奥拉斯(Christopher Olahs)关于这个主题的精彩文章更好的起点了。

ELMo在多层RNN上接受训练，并从上下文学习单词嵌入。这使得它能够根据**所使用的上下文为每个单词存储多个向量。**它附带了一个预先训练好的模型，这个模型是在一个非常大的数据集上训练的，可以动态地创建基于上下文的词嵌入，而不是像以前的静态词嵌入方法那样简单地提供查找表。



这个图是一个两层ELMO架构的例子。**您拥有的层越多，就可以从输入中了解到越多的上下文。**低层识别基本语法和语法规则，而高层提取较高的上下文语义。ELMO使其更精确的另一个方面是它采用了双向语言建模。因此，不是简单地从开始到结束读取输入，而是从结束到开始读取输入。这使得它能够捕获句子中单词的完整上下文。如果没有这个，你必须假设一个特定单词的所有上下文都出现在单词之前或之后，这取决于你读它的方向。

它还允许进行微调，以便能够根据特定领域的数据进行调整。这导致一些人声称这是NLPs ImageNet时刻，**这意味着我们越来越接近拥有可用于下游NLP任务的一般训练模型的核心构件。**

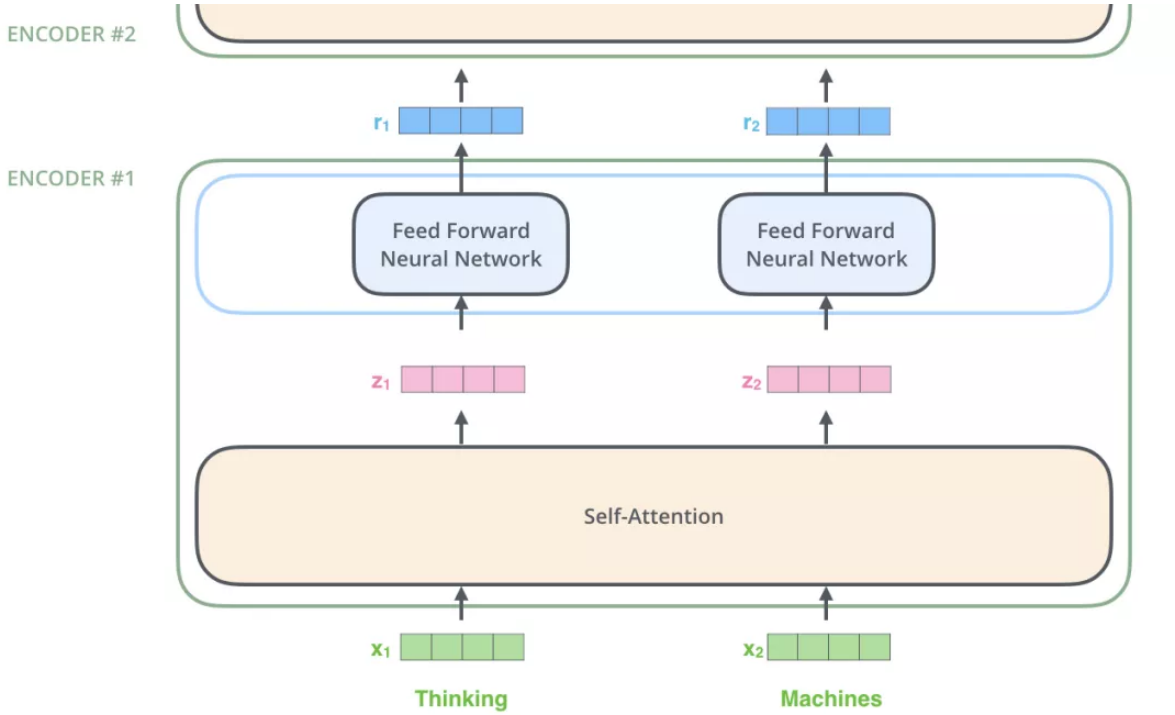
因此，RNN结构仍然是非常前沿的，值得进一步研究。直到2018年，它仍然是NLP的主要架构。一些评论家认为，**现在是我们完全放弃RNNs的时候了**，因此，无论如何，它们都不太可能成为2019年许多新研究的基础。相反，2019年深度学习NLP的主要架构趋势将是transformer。

3. Transformer将成为主导的NLP深度学习架构

虽然**ELMo能够克服以前的word嵌入式架构的许多缺点**，比如它只能记住一段文本的上下文，但它仍然必须按顺序处理它的输入，一个词一个词地处理，或者在ELMo的情况下，一个字符一个字符地处理。

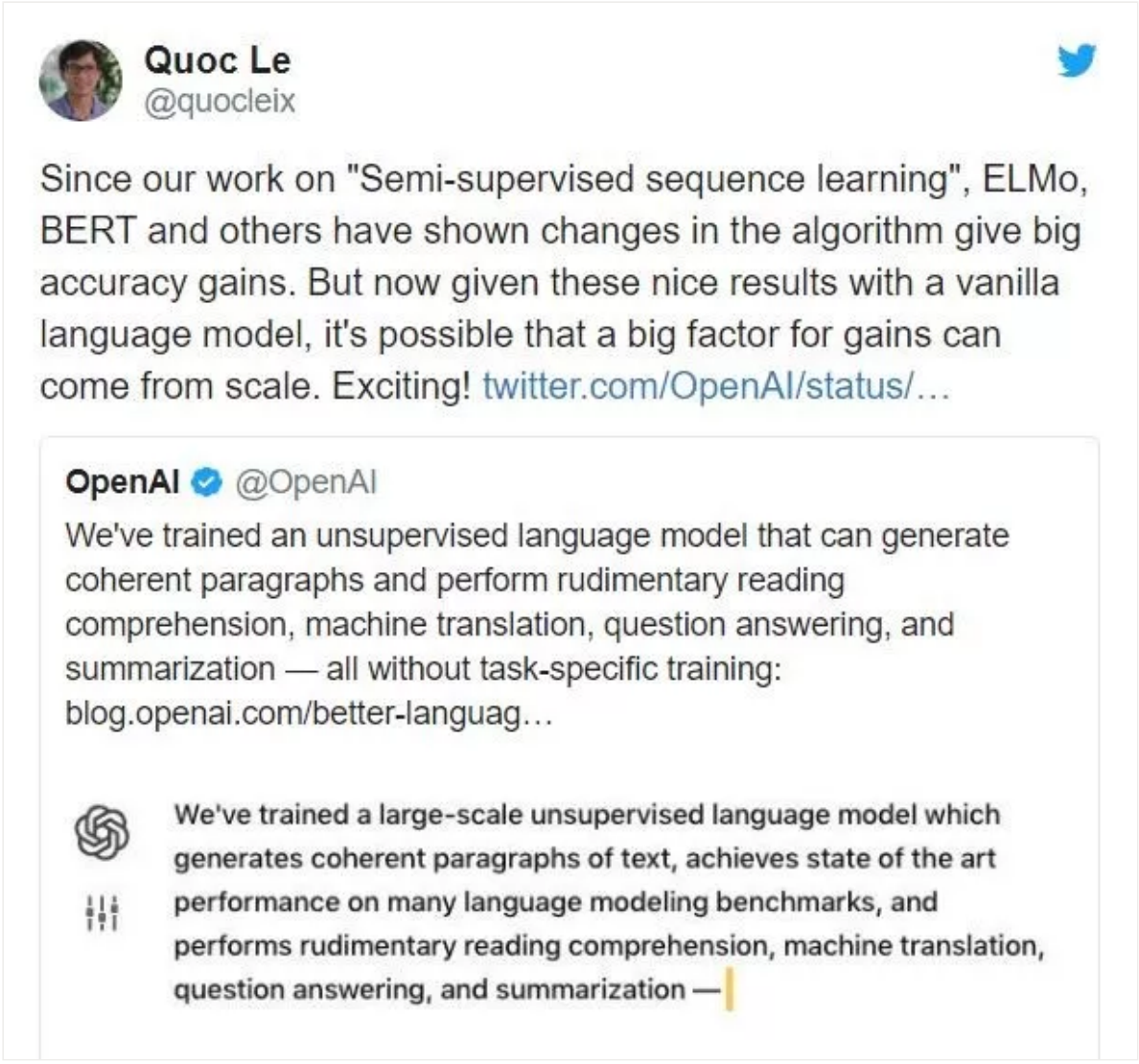
如前所述，这意味着**需要将文本流输入到输入层**。然后按顺序对每个隐层进行处理。因此，在处理文本以理解上下文时，体系结构必须存储文本的所有状态。**这使得学习较长的文本序列(如句子或段落)变得困难，也使得训练的速度变慢。**

最终，这限制了它可以训练的数据集的大小，而**这些数据集对任何训练它的模型的能力都有已知的影响**。在人工智能中，“生命始于十亿个例子”。语言建模也是如此。更大的训练集意味着您的模型输出将更准确。因此，在输入阶段的瓶颈可能被证明是非常昂贵的，就您能够生成的准确性而言。



Transformer架构在2017年底首次发布，它通过**创建一种允许并行输入的方法来解决这个问题**。每个单词可以有一个单独的嵌入和处理过程，这大大提高了训练时间，便于在更大的数据集上进行训练。

作为一个例子，我们只需要看看2019年的早期NLP感觉之一，**OpenAI的GTP-s模型。GTP-2模型的发布受到了很多关注**，因为创建者声称，考虑到大规模生成“虚假”内容的可能性，发布完整的预训练模型是危险的。不管它们的发布方法有什么优点，模型本身都是在Transformer架构上训练的。正如主要的AI专家Quoc Le所指出的，**GTP-2版本展示了普通Transformer架构在大规模训练时的威力……**



随着Transformer- xl的发布，**Transformer架构本身在2019年已经向前迈出了一步**。这建立在原始转换器的基础上，并允许一次处理更长的输入序列。**这意味着输入序列不需要被分割成任意固定的长度，而是可以遵循自然的语言边界**，如句子和段落。这有助于理解多个句子、段落和可能更长的文本(如冠词)的深层上下文。

通过这种方式，Transformer架构为新模型打开了一个全新的开发阶段。人们现在可以尝试训练更多的数据或不同类型的数据。或者，**他们可以在转换器上创建新的和创新的模型。这就是为什么我们将在2019年看到许多NLP的新方法**。

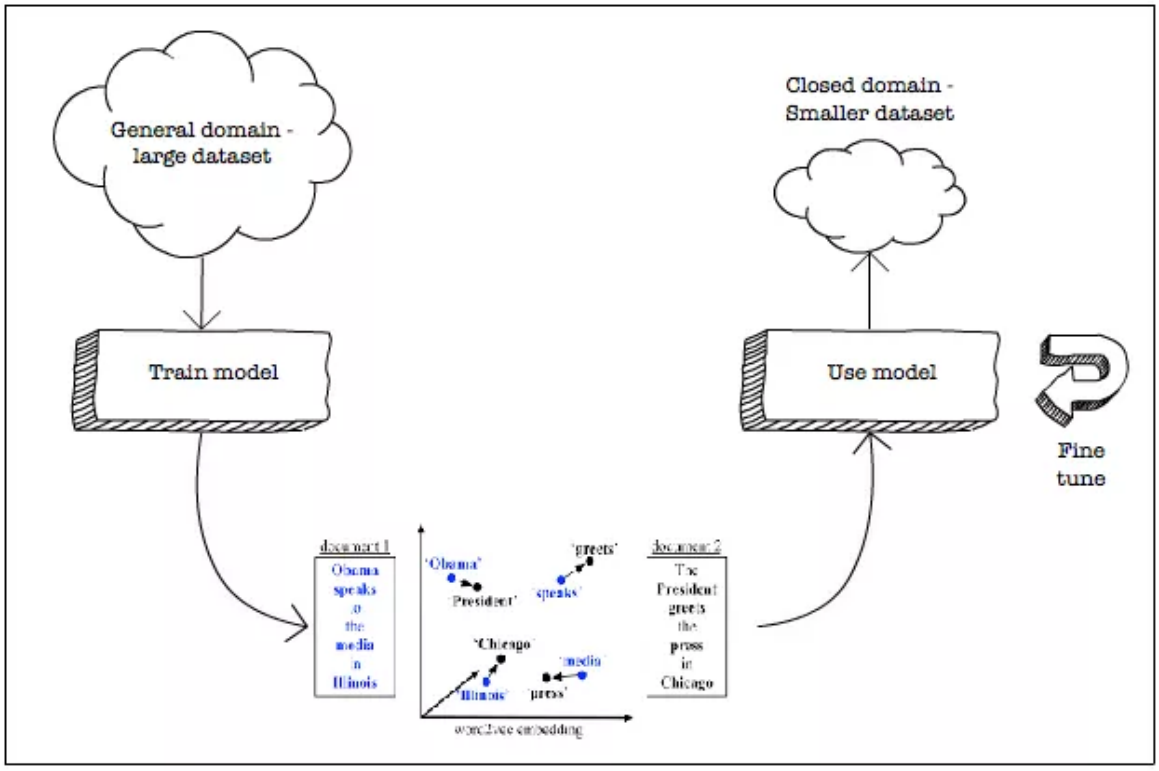
transformer架构的发布为NLP深度学习方法创建了一个新的基线。人们可以看到这种新体系结构所提供的潜力，并很快尝试寻找方法将其合并到新的更高级的NLP问题方法中。我们可以预计这些趋势将持续到2019年。

4. 预先训练的模型将发展更通用的语言技能

首先，像Transformer这样的新架构使得在数据集上训练模型变得更容易，而在此之前，人们认为数据集太大，而且学习数据集的计算开销太大。这些数据集对大多数人来说都是不可用的，即使新的体系结构使得重新训练他们自己的模型变得更容易，但对每个人来说仍然是不可行的。因此，这意味着人们需要使他们的预先训练的模型可用现货供应或建立和微调所需。

第二，TensorFlow Hub开启了，这是一个可重用机器学习模型的在线存储库。这使它很容易快速尝试一些先进的NLP模型，这也意味着你可以下载的模型，预先训练了非常大的数据集。这与ELMo和Universal Sentence Encoder (USE)的出版是一致的。使用的是一种新的模型，它使用转换器架构的编码器部分来创建句子的密集向量表示。

5. 迁移学习将发挥更大的作用



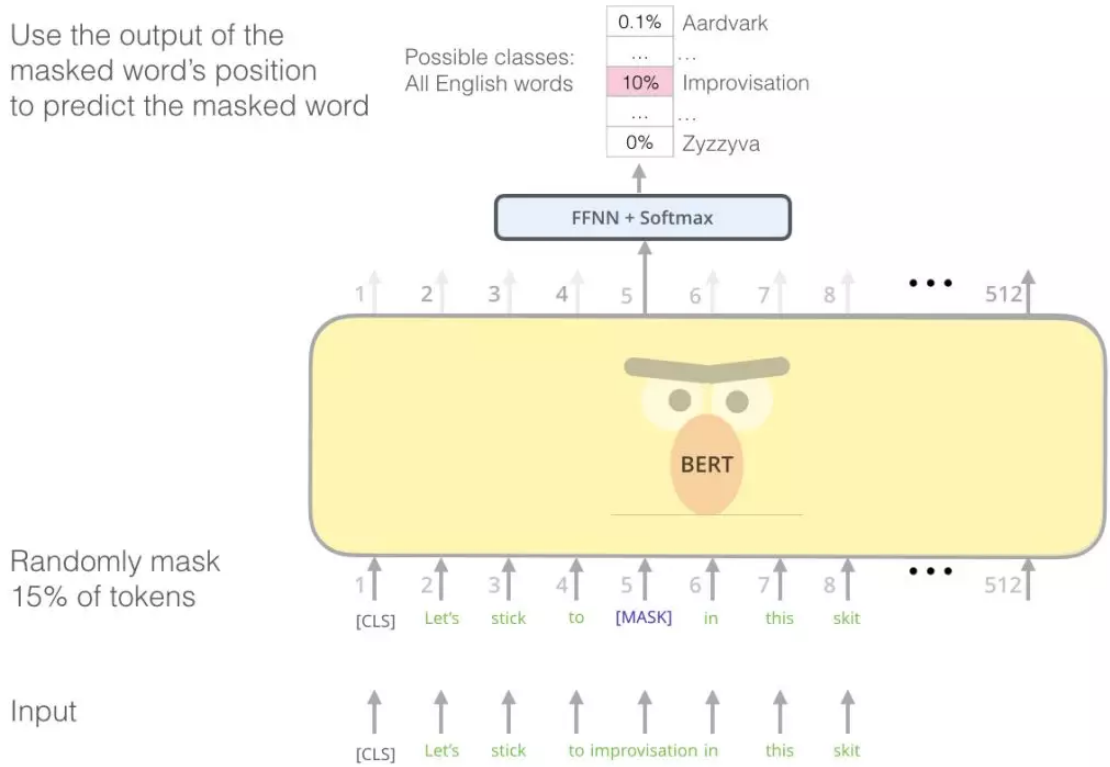
迁移学习允许您根据自己的数据对模型进行微调

随着更多的预先训练模型的可用性，实现您自己的NLP任务将变得更加容易，因为您可以使用下载的模型作为您的起点。这意味着您可以在这些模型的基础上构建自己的服务，并使用少量领域特定的数据对其进行快速培训。如何在您自己的生产环境中实现这些下游方法的一个很好的示例是将BERT作为服务提供的。

6. 微调模型将变得更容易

相反，原始模型的输出，BERTs和ELMos，是一个密集的向量表示，或嵌入。嵌入从它所训练的大的和一般的数据集中捕获一般的语言信息。您还可以对模型进行微调，以生成对您自己的封闭域更敏感的嵌入。这种形式的微调的输出将是另一种嵌入。因此，微调的目标不是输出情绪或分类的概率，而是包含领域特定信息的嵌入。

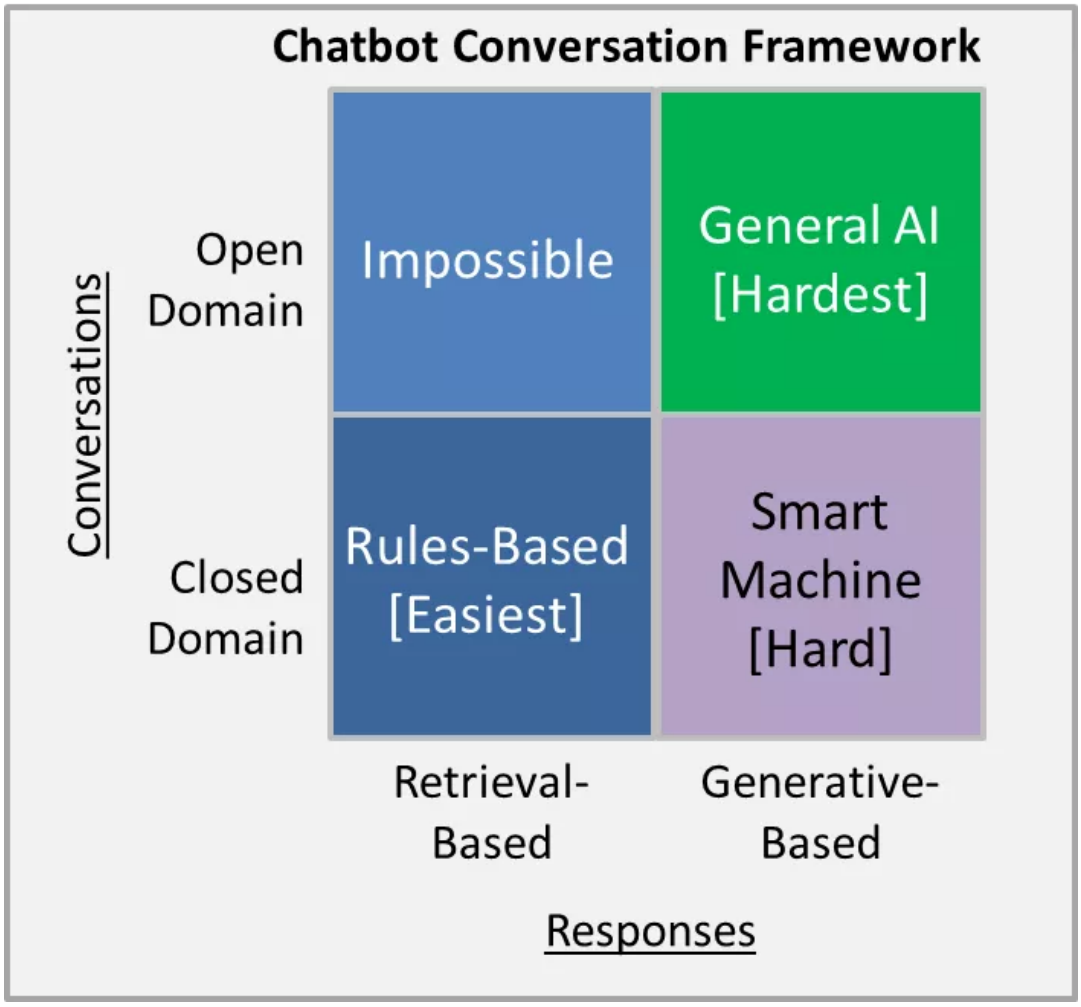
7. BERT将改变NLP的应用前景



BERT的预先训练的通用模型比它的任何前序都更强大。它已经能够通过使用双向方法将一种新技术纳入到NLP模型的训练中。这更类似于人类从句子中学习意义的方式，因为我们不只是在一个方向上

理解上下文。我们在阅读时也会提前投射以理解单词的上下文。

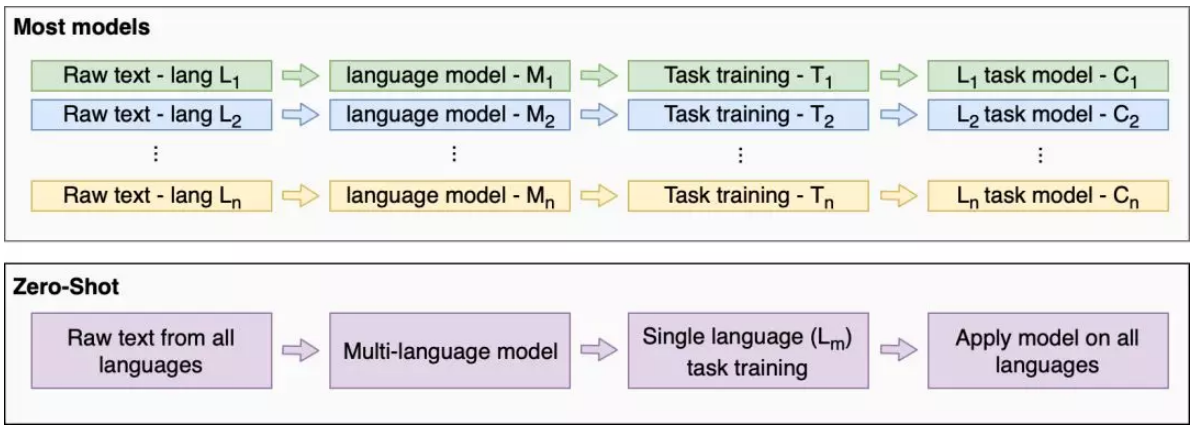
8. 聊天机器人将从这一阶段的NLP创新中受益最多



有了像GPT-2和BERT这样的方法，情况就不一样了。现在我们看到，一般训练的模型可以在接近人类的水平上产生反应。而特定的封闭域聊天机器人则比较困难，因为它们需要进行微调。**到2019年，将出现一种转变，即创建工具来更容易地对模型(如BERT)进行微调，以获得更小数量的领域特定数据。**未来一年的主要问题将是，是更容易生成响应，还是使用新的NLP模型将传入的客户问题与之前存储或管理的响应模板匹配起来。这种匹配将由发现问题和回答之间的相似性来驱动。调优越好，模型在识别新客户查询的潜在正确答案方面就越精确。

9. 零样本学习将变得更加有效

零样本学习是在一个非常大的数据集或一个非常不同的数据集上训练一个通用模型。然后您可以将此模型应用于任何任务。在翻译示例中，您将训练一个模型并将其用作其他语言的通用翻译程序。2018年底发表的一篇论文就做到了这一点，能够学习93种不同语言的句子表示。



10. 关于人工智能的危险的讨论可能会开始影响NLP的研究和应用

目前，深度学习NLP领域似乎是人工智能最令人兴奋的领域之一。有这么多事情要做，很难跟上最新的趋势和发展。这是伟大的，它看起来将继续和增长更快。唯一需要注意的是，经济增长的速度可能太过迅猛，以至于我们需要更多的时间来考虑潜在的影响。