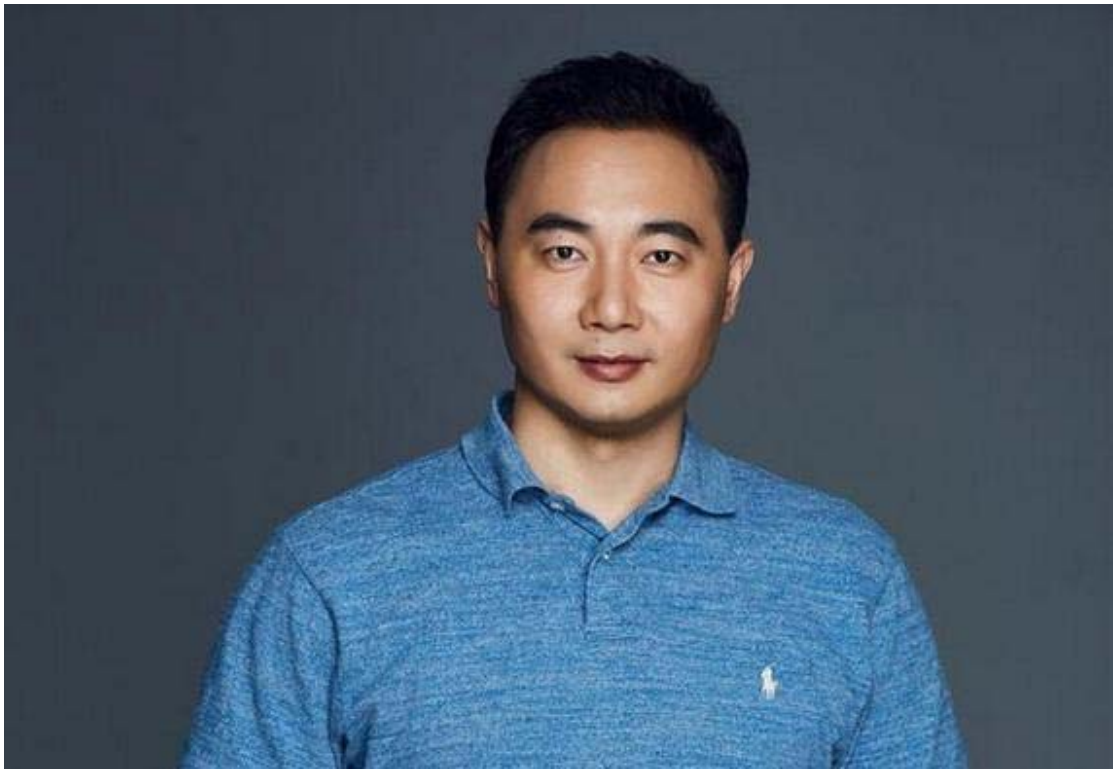
周明：NLP 进步将如何改变搜索体验

2019-04-09 | 作者：微软亚洲研究院

编者按：4 月 3 日，微软亚洲研究院副院长周明受邀参加北大 AI 公开课，与大家分享了近期自然语言技术取得的进展和创新成果，并探讨了自然语言技术和搜索引擎如何进一步结合并创造新的可能。在课后问答环节，周明解读了当前自然语言技术比较重要的研究方向，并为想要进入这一领域的同学提供了一些实用建议。本文由 AI 前线（ID：ai-front）独家整理首发，未经授权请勿转载。



课程导师：雷鸣， 天使投资人，百度创始七剑客之一，酷我音乐创始人，北大信科人工智能创新中心主任，2000 年获得北京大学计算机硕士学位，2005 年获得斯坦福商学院 MBA 学位。

特邀讲者：周明博士， 1999 年加入微软研究院，现任微软亚洲研究院副院长，也是现任国际计算语言学会（ACL）会长，中国计算机学会理事、中文信息技术专委会（即 NLP 专委会）主任、中国中文信息学会常务理事。他长期领导 NLP 的研究，包括输入法、在线词典

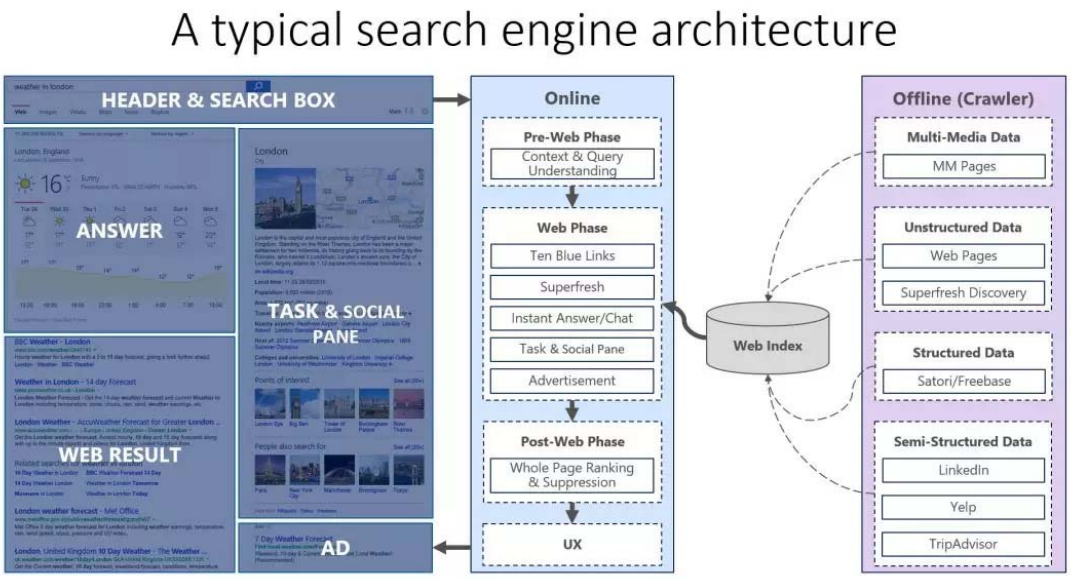
（必应词典）、下一代搜索、机器翻译、问答、聊天机器人、计算机对联（微软对联）、知识图谱、语义分析、文本挖掘、文本生成、用户画像和推荐系统等。主编《机器翻译》、《智能问答》等 NLP 技术专著。他的团队对微软产品（譬如 Office、Windows、必应搜索、Azure、小冰等）的 NLP 技术做出了不可替代的贡献。

以下为 AI 前线（ID：ai-front）独家整理的周明老师课程内容（略有删减）

对于搜索引擎来说，最重要的是两件事，第一是智能程度，指的是理解用户意图和文档，然后快速找出答案，这是智能部分；第二是自然程度（Naturalness），指的是根据用户输入的搜索请求，把搜索结果很自然地展现给用户，整体表现就是搜索非常流畅。自然语言从搜索引擎出现开始一直到今天为止，都对搜索引擎的智能和自然这两个方面起到了极为重要的作用。

搜索引擎背后的 NLP 技术

下图是一个典型的搜索引擎，我们以微软 Bing 搜索为例回顾一下搜索引擎的工作过程，再看看其中涉及到哪些自然语言技术。



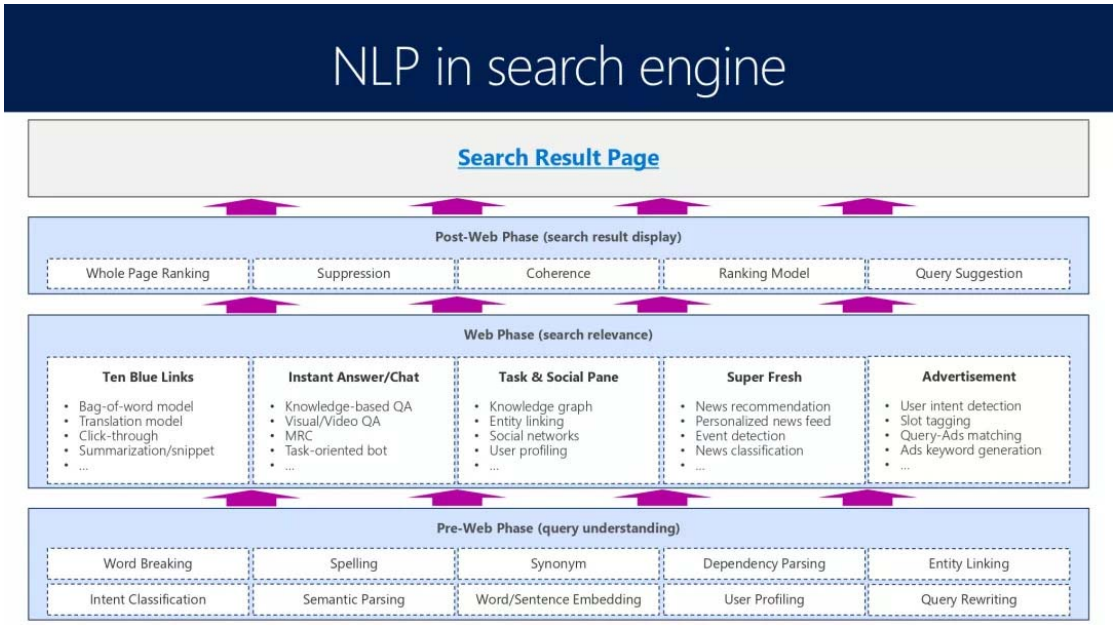
首先会有一个输入框，我们叫做 Header&Search Box，用于输入 Query。用户输入 Query 之后，它就要干一件事，叫 Context&Query Understanding。搜索引擎首先要理解 Context，就是什么人用了什么样的设备、在什么地点输入这样的 Query，其次要理解用户的意图，就是用户输入这个 Query 是想搜什么。

基于理解的结果，再到网上去搜索。主要会搜出几件事，第一个就是 Ten Blue Links 即十个最重要的匹配文档。还有 Super Fresh 内容，就是新鲜的一些事物或者文档，比如新闻的有关结果搜出来。还有一些 Instant Answer，就是涉及到天气、股票、交通等垂直领域的信息，我们一般都叫 Instant Answer。还有一些 Task&Social Pane，Task 指的是基于搜索结果的页面上还可能要做新的任务，比如订票；Social Pane 是列出相关的微信、微博或 LinkedIn 的各种信息。最后在以上的各种搜索结果基础上配上合适的广告。

这就是我们的搜索引擎基于一个用户 Query 到最后输出搜索结果的一个过程。然后我们要把这些结果体现在一个网页上，再对页面进行整体优化，适应于不同的设备、浏览器和屏幕（比如桌面和手机），页面布局要美观大方。

为了支持这个搜索过程，我们需要一些 Offline 的操作，最主要的就是 Crawler 和 Index。Crawler 指的是到网上把浩如烟海的各种文档爬下来，爬的越多越好；然后是 Index，把重要的文档选出来，同时把每篇文档中其中的重要信息摘出来，一般是用关键词来做索引，然后入库；这个过程中需要把一些有极端倾向或者黄色的文档过滤掉。这些都是 Offline 要做的工作，最后要把结果体现到 web index 里面，供搜索的前端系统使用。

我们可以看到，整个搜索过程背后用到了很多自然语言技术，具体如下图。



在搜索引擎初期，自然语言技术用的比较浅。随着自然语言技术快速发展并变得越来越成熟，我们把越来越多的自然语言技术（NLP）试探性地放到搜索引擎中，看它能起到什么样的效果，并不断加以改进直到稳定。NLP 在搜索中的作用越来越重要。

今天的讲座主要回答两个问题，第一是自然语言处理到底有哪些新的进展；第二是这些新的进展给我们的搜索引擎带来了什么新的变化，或者未来可能会带来什么新的变化。

自然语言技术的重要组成

自然语言技术覆盖的领域可以总结为三方面，包括 NLP 基础、NLP 核心技术和 NLP 应用。

NLP 基础包括词的表示，比如现在热门的 word Embedding。词的表示包括上下文无关的表示和上下文有关的表示，前者就是静态的 Word Embedding，后者现在一般使用各种预训练模型，根据当前的句子来体现一个词在特定上下文的的语境里面该词的语义表示。同样一个词，在不同的语境下，其语义表示也不一样。基于词的表示，我们就可以做很多应用，比如语言模型、分词、语言模型、句法语义分析、篇章分析、等等，这些都是 NLP 的基础。

基于 NLP 基础，我们又有很多 NLP 的核心技术，包括机器翻译、问答、信息检索、信息抽取、对话、知识工程，还有自然语言生成、推荐系统，等等。

基于 NLP 核心技术，我们就可以把 NLP 用在一些具体的应用中，比如搜索引擎、客服、商业智能和语音助手。

为了完成这些任务还需要很多底层支撑技术，包括用户画像建模、用于实现个性化的推荐技术、大数据能力、计算能力、机器学习和深度学习的能力、知识库、常识及推理的能力。

深度学习对自然语言技术的影响

深度学习先后对图像、语音、自然语言这些领域都产生了重要的影响。其中，深度学习对自然语言的影响主要体现在以下 6 个方面：

1. 端到端训练（End-end training）

过去做统计自然语言处理的时候，都是由专家去定义各种 Feature，需要很多领域知识。有的时候不容易找到很好的 Feature。而有了端对端的训练，只要有输入和输出的对照（输入-输出），把输入对应的输出标注好，形成训练数据集合。然后用神经网络通过自动训练就可以得到学习系统，不需要人为设定和优选 Feature。这改变了很多自然语言技术的发展，大大降低了自然语言处理的技术门槛。这意味着，你只要有算力和标注数据，基本上可以“傻瓜式”地实现一个自然语言模型的学习，从而推动了自然语言处理技术的普及。

2. 语义表示（Embedding）和预训练模型（Pretrained Model）

一是上下文无关的 Embedding（表示），就是不管上下文是什么，一个词的表示是固定的（用多维向量来表示）。第二个，根据上下文有关，在不同的句子里，同一个词的意思可能不一样，那么它的 embedding 也是不一样的。现在利用 Bert 和 GPT-2 这样的模型，可以根据一个词的上下文训练这个词的动态 Embedding。在做其他任务时候，预训练模型可以用来强化输入信息。有了 Embedding 这个东西就可以计算词与词之间的距离；基于词的 Embedding，又可以得到句子的 Embedding，也就可以计算句子与句子之间的距离。这就使得搜索引擎中 Query 对 Document 匹配程度的计算得以改进。

3. Attention（注意力模型）

Attention 指的是不同的输入信号源之间可以做相应的修正，来动态地体现当前层对网络的下一层或者对网络输出层的最佳输入信号。有了 Attention，就可以对受多输入路信号，然后动态计算信号之间产生的互相影响。

4. 句子的编码方法（RNN/LSTM/GRU/Transformer）

对于一个不定长的句子，可以通过 RNN、LSTM/GRU 或者 Transformer 技术表示其编码，表现为若干个隐含状态的序列。一个隐含状态对应句子的一个词汇。虽然以上对句子的几种编码方式都可行，但是发展到目前，更多是用 Transformer 来对句子编码。对句子编码之后，就可以做翻译、问答、检索等各种应用。

5. 编码 - 解码模型（Encoder-Decoder）

NLP 中，很多任务都可以定义成一个输入和一个输出的对应。所以编码-解码模型有普遍的适用意义。比如，机器翻译任务，源语言句子是输入，目标语言句子是输出。这样就存在输入和输出的对应。如果是单轮任务，就是输入和输出直接对应，不需要中间推理，可以用编码和解码的技术来进行建模。除了机器翻译，词性标注、分词、句法分析、语义分析、问答、摘要、阅读理解等许多任务都可以通过编码-解码模型进行建模。

6. 强化学习

系统根据用户的反馈或者环境的反馈信号，会迭代地修正参数，整个系统得以不断改进。比如对话系统很多用到了强化学习。不过在很多其他 NLP 任务中，如何体现强化学习是一个还在不停探索的问题。

自然语言技术的进展和趋势

接下来逐一介绍自然语言技术在不同方向上的进展，并讨论每一项进展对于搜索的影响。

问答技术（QA）

当用户提出问题或者 Query 的时候，搜索引擎或问答系统需要到它所能够掌握的资源里去找到相应的答案。一般有如下几项资源以及相对应的 QA 技术：

1. Community-QA，就是常见的 FAQ 表。对于一个问题，可寻找历史上类似的问题，然后把其对应的答案输出。

2. KBQA，就是到知识图谱里把相应的答案找出来或者推理出来。

3. TableQA，针对问题在网络上查找对应的表格，然后把表格的相关信息抽取出来作为答案。

4. PassageQA，针对问题，在无结构的文档中寻找答案。

5. VQA，从视频或者图像中把答案抽取出来。

目前利用多源数据流或者知识库进行 QA 的技术已经越来越普及，而且相应的语义分析技术和排序技术也比以前大大提高了。

过去的 QA 都是用的传统的，像手工编辑的基于规则的语义分析，比如说 CCG，但由于它存在各种问题，最近三年以来人们更多使用 Encoder-Decoder 技术来做语义分析，在分层语义分析、上下文感知的语义分析上都取得了新的突破。

有了很好的 QA 之后，搜索引擎的智能水平和自然程度都提高了。但是在具体做搜索引擎的时候，比如在某些垂直领域，或者使用某些设备时，怎么用 QA 的结果，可能是仁者见仁智者见智。可信度极高的时候可以使用 QA 的结果，可信度不高的时候还是要回归到原来的 Ten Blue Links 上面，这需要拿捏一定的尺度并跟 UI 很好地结合。

多语言处理能力（Multi-lingual capability）

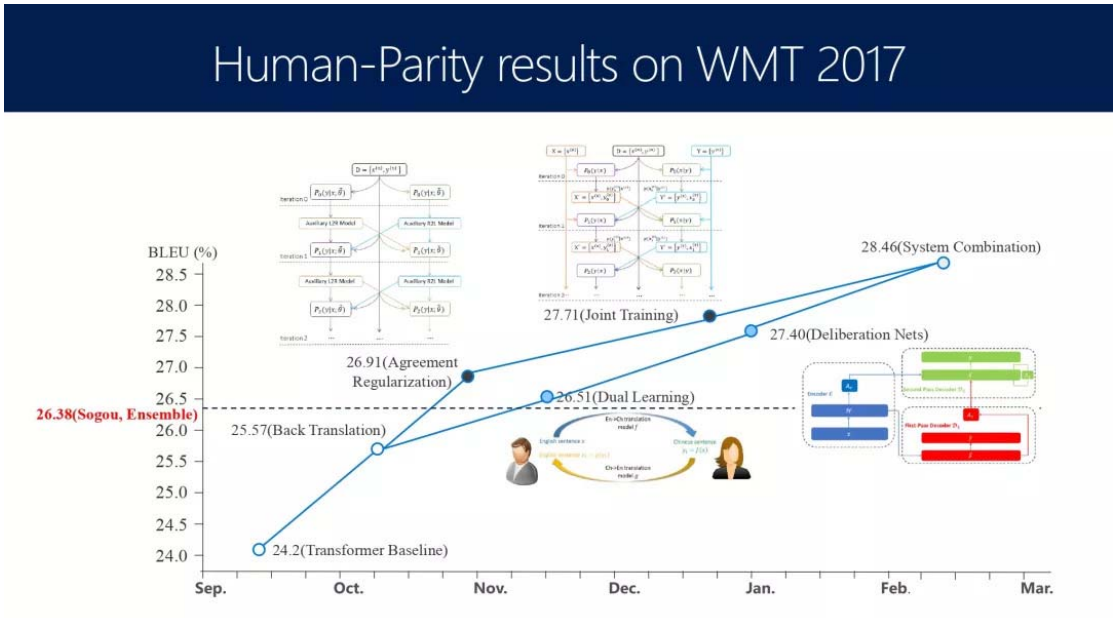
多语言的处理能力对于搜索引擎非常重要。假设我们有很好的机器翻译技术，就可以做多语言的搜索，将用户想要搜索的信息、哪怕是来自其他语言的也可以找出来，在搜索结果中呈现，并通过机器翻译技术把页面翻译成用户的母语。

机器翻译其实一直都进展缓慢，近几年由于深度学习技术的发展，神经网络机器翻译相比统计机器翻译已经有了大幅度的进展。机器翻译技术可以分成两类，一种是 Rich Resource NMT，也就是双语语料丰富的语言对（比如中文 - 英文）；另一种叫 Low Resource NMT，即缺少足够的双语语料（比如中文 - 希伯来语）。

目前的机器翻译在 Rich Resource 上已经做得非常好了，甚至在某些训练集下已经可以达到或超过人工翻译水平。但是 Low Resource 现在才刚刚开始，有很多有趣的研究，整体水平还处于比较低级的阶段。

机器翻译在搜索上已经有很多的应用，现在每一个搜索引擎都会有机器翻译应用，用户可以动态地把网页翻译成所需要的某种语言。

神经网络机器翻译最开始是用 RNN 来做，后来引入了注意力模型，过去两年又出现了 Transformer 技术，大大提升了并行能力。现在大部分神经网络机器翻译都是用 Transformer 来做的，最近业界也有了很多新的进展，包括微软亚洲研究院的最新技术等，使得机神经器翻译有了长足的进步。



上图所示是微软 2018 年在神经机器翻译的进展。传统 Transformer 的 Baseline 只有 24.2，加上单方向的 Back Translation 之后可以达到 25.57，再加上联合学习、对偶学习、多次解码和双向一致性解码等技术让系统的表现不断提高。当前将所有成果结合起来已经在 WMT 新闻语料上得到了一个最佳翻译结果，而且这个结果达到了人工翻译在这个数据集合的水平。

下面是一些句子翻译的示例展示，第一行是输入的句子，第二行就是机器翻译的句子，第三行是人工翻译的句子。虽然有些词用法不一样，但是所有句子相互之间都是等价的。

News translation results

- Sampled from WMT2017 Chinese-English task

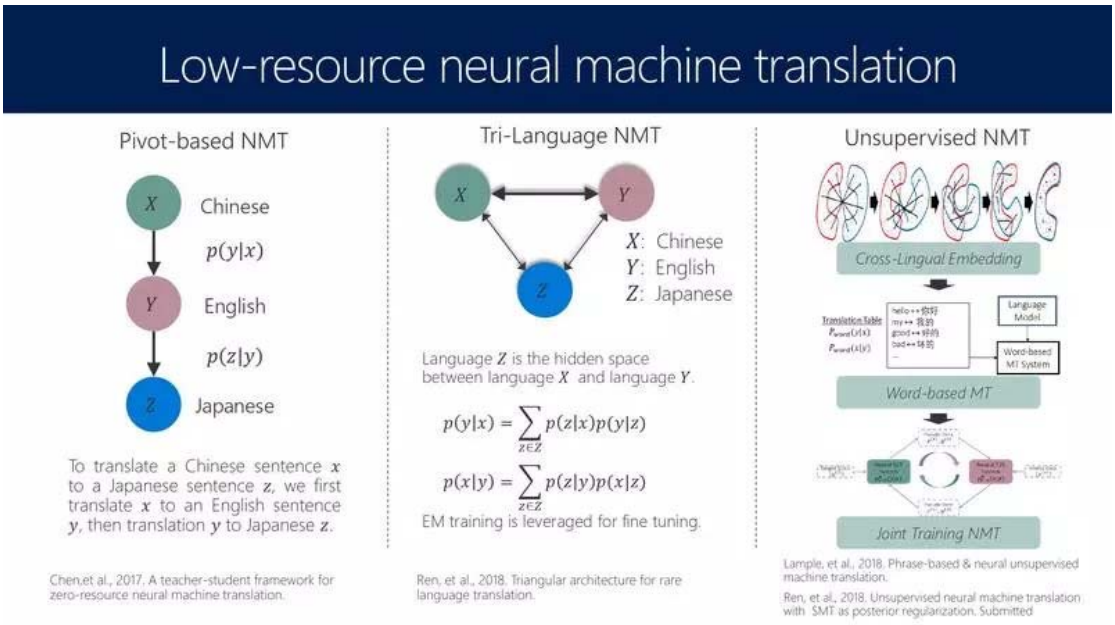
Source input	**有 线索 人士** 请 拨打 旧金山 警察局 举报 电话 4 15- 575 - 44 44 。
NMT output	For clues, call the San Francisco Police Department at 415-575 - 4444.
Human reference	Anyone with information is asked to call the SFPD Tip Line at 415-575-4444 .

Source input	他 的 职业 生涯 如 过 山 车 一般 。
NMT output	It has been a rollercoaster ride .
Human reference	His career is like a roller coaster.

Source input	霍夫 施泰特尔 表示 : " **这 将 由 检察官 来 确定** " 。
NMT output	" That 's what the prosecutor must determine , " said Hofstetter .
Human reference	Mr Hoff Steitel said: "It will be up to the prosecutors to determine.

当没有那么足够多语料的时候，就要考虑 Low Resource 的机器翻译技术。Low Resource 的机器翻译现在主要有三个思路。

Low-resource neural machine translation



枢轴式翻译（Pivot-based NMT）：比如说要翻中日，可以通过先中翻英、再英翻日这样一个两步走的过程来实现，因为中文 - 英文、英文 - 日文的双语语料比较多。

Tri-Language NMT：这是一个三角形的机器翻译架构。假设有一个 Rich Resource 的语料对，比如中英，但是要翻译一个 Low Resource 的语言，比如希伯来语。中文和希伯来语、英文和希伯来语的预料对相对都比较少，那么可以利用中英已经很强大的机器翻译和对应的

语料，来把希伯来语和中文，与希伯来语和英文的翻译来强化，通过一个 EM 迭代的过程来体现这样的带动作用。

Unsupervised NMT：有时候可能什么双语料也没有，只有一些简单的小辞典，体现源语言词与目标语言词的对应关系。那么可以利用这个小辞典做一些工作。首先做一个所谓的跨语言 Word Embedding，把不同语言的词，如果它们表达相近或者相同的意思，试图通过一种方式把它们聚在一起。抽取高可信度的词汇对应形成一个双语对照辞典。基于这个翻译辞典，再加上目标语言的语言模型，就可以做一个词汇级的统计机器翻译。基于这个统计机器翻译，就可以把源语言翻译得到目标语言，或反之，虽然翻译质量不高。再利用这样的双语料，就可以分别去训练神经网络机器翻译，然后再通过类似我们在做 WMT 的一些技术，比如实现源到目的、目的到源的翻译系统的互相迭代，进一步强化翻译结果。

多模态搜索

多模态搜索指的就是将语言、语音、文字、图像等各种模态集成来进行搜索。

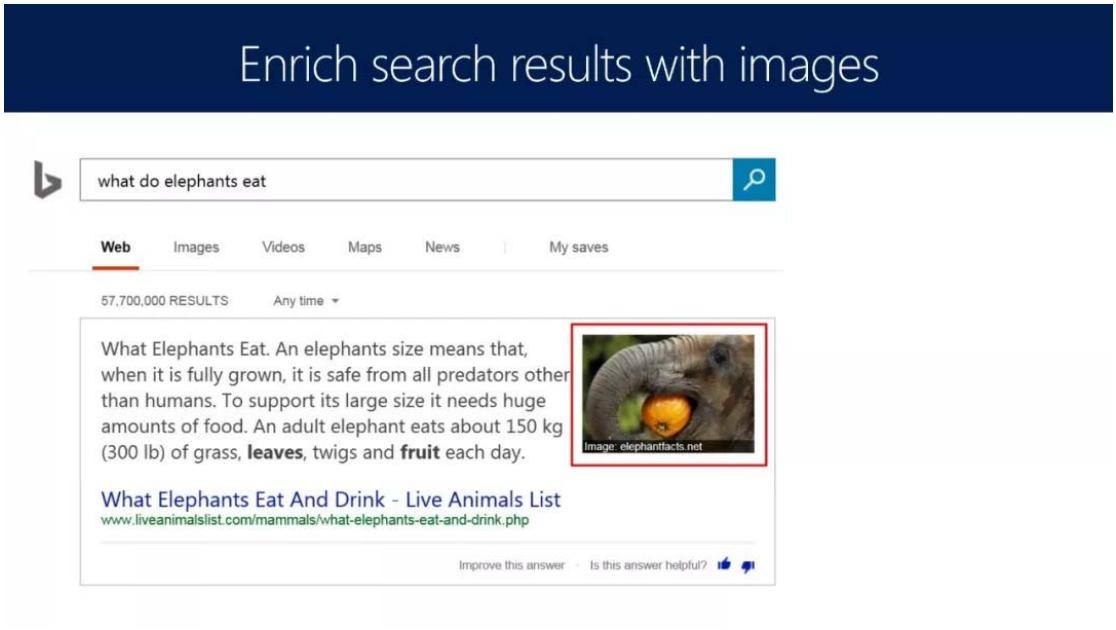
近几年 ImageNet 数据集将图像识别的水平大幅度提高，而 image captioning 和 video captioning 技术可以用自然语言来描述图像和视频的内容。这些技术的进展激发了研究人员对多模态搜索的更多尝试来提高搜索的用户体验。

VQA 指的是基于图片对用户提出的问题进行解答。有了 VQA 数据集之后，研究人员就可以实验性地去做一些基于图像的问答系统，背后涉及到如何对图像跟自然语言进行编码，如何通过注意力模型把常识引入进去，以提高整体水平。目前仍处于初期阶段，其中还有许多有意思的挑战，比如怎么使用推理和常识。最近斯坦福做了一个叫做 GQA 的数据集，用来体现 VQA 的推理过程，比如对一个问题经过哪几个步骤进行了推理得到答案。研究人员可以用它来训练 VQA 系统的推理能力。

基于多模态技术，可以做出很多新的搜索体验。比如用户输入一个 Query，可以直接输出图像结果，甚至图像中每一个人在知识图谱中对应的 ID 可以找回来，提示给用户，可以链接知识图谱的描述。其中也用到了人脸识别技术。

另外，也可以直接输入图像进行 Query，比如手机照相，经过图形识别，得到相关图像和文档。

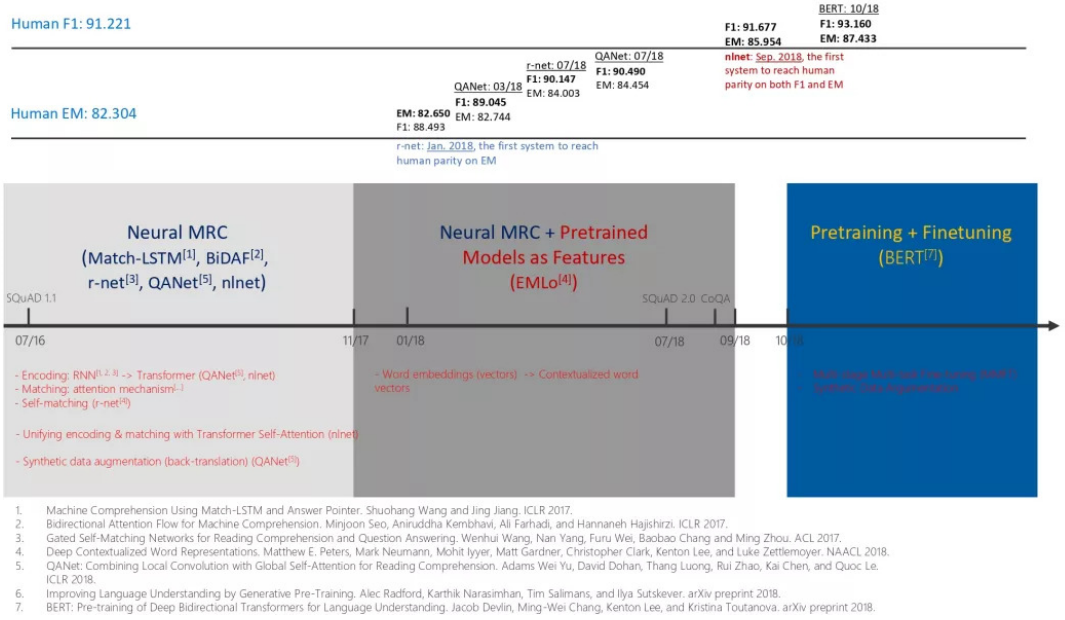
图像搜索的结果也可以强化普通的文本搜索结果。比如在输入引擎中输入一个 Query：大象吃什么？可得到文本搜索的结果以及图像的搜索结果。这两个结果可以互相增强，来提高用户的搜索体验。



机器阅读理解

机器阅读理解（Machine Reading Comprehension），简称 MRC。简单地说，就是针对一篇文章，如果问机器一个问题，看机器能不能把这个问题回答出来，有时候是直接从文章中找出一个答案，有时候可能要根据上下文进行推理。

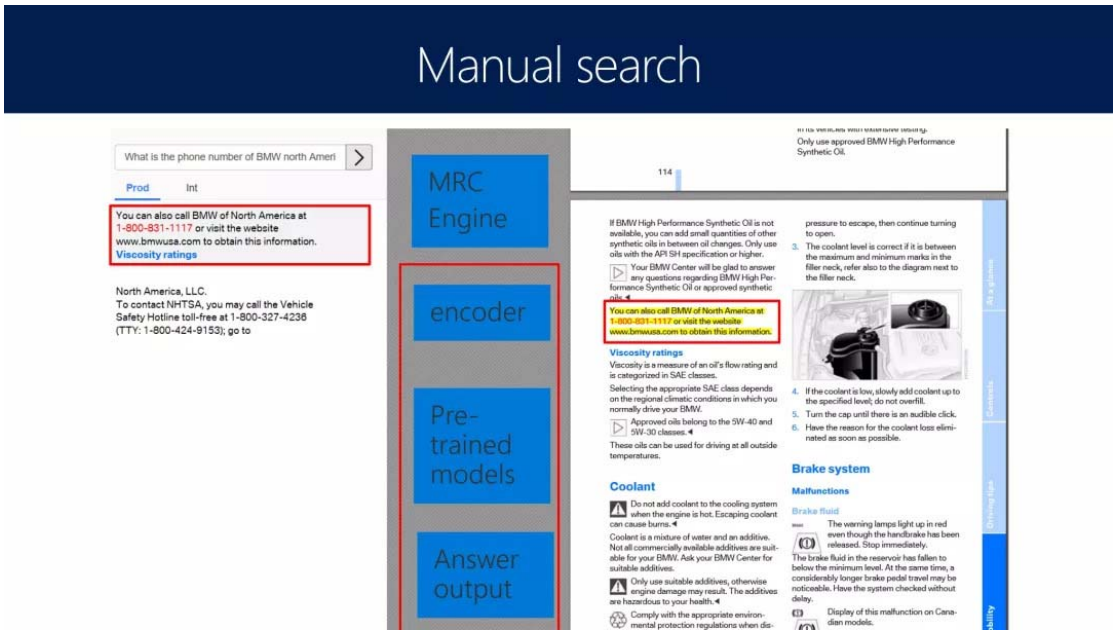
过去这几年，SQuAD 1.1、SQuAD 2.0、CoQA 这些经典的机器阅读理解数据集驱动着 MRC 快速发展。而大量涌现出的很多优秀的 Pre-trained Model，像 ELMo、Bert 等也大幅提高了 MRC 的水平，主要体现在能够将一些开放领域的人类知识（隐含）进行编码，强化输入信号（问题和文章），并参与到一个整体的端对端训练过程中。学术界和产业界有很多团队在不断刷新着 MRC 的记录，甚至超越了人工水准。如下图所示。



有了更高的机器阅读理解水平，我们可将其应用在搜索上做一些新的尝试，MRC 对搜索的帮助主要体现在以下几方面：

1. 首先可以对搜索结果的摘要进行改进，抽取出更好的摘要。对于摘要中最匹配答案的部分，可以更好地 highlight 出来。

2. 优化手册阅读理解。一般使用手册都很长，没有人愿意看，只需要将手册的 PDF 文件数字化，就可以做机器阅读理解。而用户只要发出一个问题，可直接找出它的答案。



3. 加速网站全站搜索。在某一个网站中，比如客服网站或某一个产品的介绍网站，如果用户有问题，只要把问题输入进去，机器可以对整个网站进行解析，并把答案直接抽取出来。同时用户还可以通过 Conversational QA 连续对网站进行连续提问。通过 MRC 找到精准的答案，以实现一个交互式的搜索过程。

未来 MRC 如果要进一步提升，一方面在 Pre-trained Model 上还有很多可以改进的地方，另外还要加强上下文推理的能力，以及更好地融合常识和知识库，增强推理过程。

个性化推荐

对于搜索引擎来说，推荐系统变得越来越重要。所谓推荐系统指的就是用户不用（显式）输入 Query，系统会根据用户过去的行为，直接把他可能喜欢的内容推荐过去。现在这种方式在手机端越来越流行。

这背后涉及很多技术：

1. 第一个技术是用户画像（User Modeling）。即根据用户的各种行为，比如搜索行为、使用地图的行为、电子商务网站的各种行为，这些行为构成的异构数据，对某个用户形成了一个全面的了解。通过对多种异构数据的融合建模，来体现用户画像。

2. 第二是怎么将用户画像的结果表达出来。一种是显示的表达，比如男女、性格、年龄等用关键词或数字表示出来。如果涉及隐私问题，可考虑隐式的表达，通过 User Embedding，用多维向量（一串数字）来代表一个用户的整体特征。 虽然不能显式体现用户的特征（从而保护用户隐私），但是却能够体现出很好的推荐效果。

3. 根据知识图谱和朋友圈对推荐内容进行扩展。

推荐内容拓展之后，再加上用户画像，最后就会变成一个简单的匹配或者 Ranking 的过程。也就是将用户画像作为一个 Embedding，待推荐的事情（比如说新闻、博客、Video、电影等）也做 Embedding，通过神经网络计算他们的相似度，相似度高就推荐给用户，这样就可以实现一个推荐的过程。

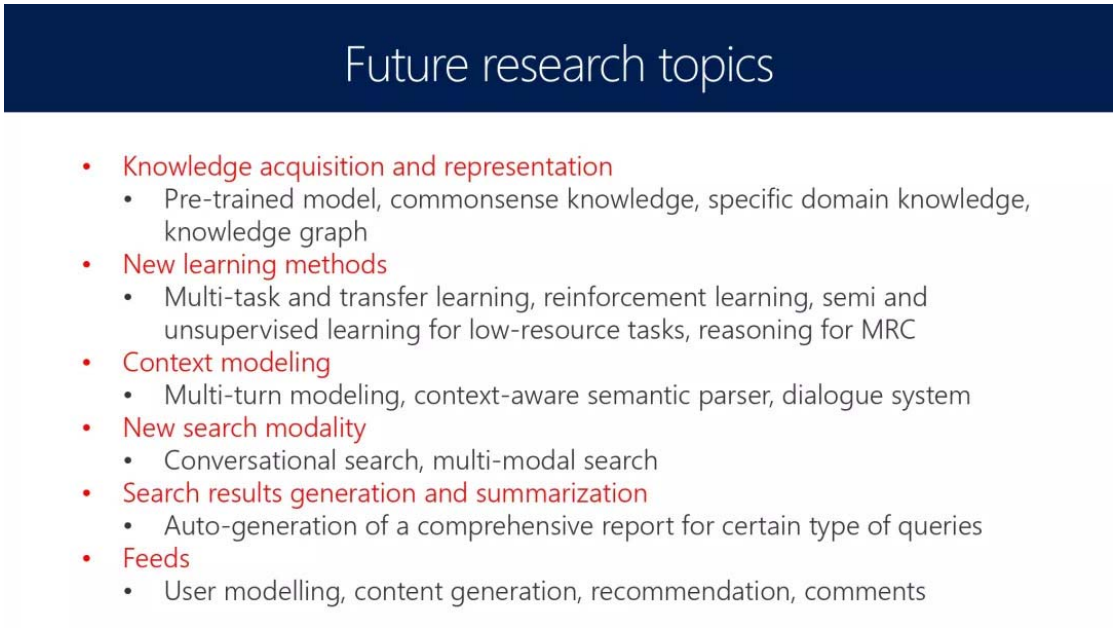
上面展示的是微软亚洲研究院在个性化推荐系统上做过的一些工作。

未来，个性化推荐系统有几个方向值得关注：

1. 做聪明的推荐，既能找到用户以前喜欢的内容，又能预测用户未来可能喜欢的内容，及时推荐给用户。

2. 推荐系统的可解释性，做推荐不能盲目推荐，还需要给用户一个解释，为什么要把这样的内容推荐给他呢？可能因为他的朋友某某喜欢，或者因为通过用户早前的搜索行为预测用户可能喜欢这个被推荐内容。这种解释要以自然语言形式附着在被推荐内容上面，来帮助用户理解。

未来研究方向



自然语言处理未来比较重要的研究课题包括：

知识获取和知识表示，尤其是前面提到的 Pre-trained Model，一是怎么用，二是怎么改进。还有像常识知识如何获取，如何加入到数据训练过程中，以及如何融入领域知识和 Open Domain 的知识等。

新的学习方法，比如多任务学习、强化学习、半监督和无监督的学习，还有 Low Resource 资源的学习。此外，推理是未来关注焦点。如何把推理很好地建立起来。用在机器阅读理解、多轮对话、法律、医疗诊断等方面。

上下文相关的建模（Context Modeling）。多轮对话的时候，如何把历史信息存储起来，又如何用在当前句子的解析里面。

新的搜索模态：除了文字，用户用语音、图像、手势、触摸等进行搜索。而且多个模态可以自然融合。

搜索结果的生成和摘要。这方面做的相对比较少，比如将不同方面的内容收集出来，体现出鲜明的观点，甚至做一些对比，生成深度好文等。

信息流。信息流现在无论是工业界还是学术界都越来越热。如何进行用户画像建模，如何获得丰富的内容（抓取、授权、翻译、生成等），如何做各种推荐，如何提供推荐的解释等都是未来很重要的研究方向。

雷鸣对话周明

雷鸣：多轮对话一直是研究上的一个热点，也是一个难点，到现在应该说解决的也不是特别好，这块的话，你感觉它的最大的挑战在哪？未来的几年会有什么样的进展？可能在什么技术上能支撑它做得更好？

周明： 多轮对话问题确实很难，现在来讲最难的就是，上下文信息记录下来之后，什么信息可以用在当前的这个句子里，什么信息应该遗忘，这在目前是不够清楚的，没有那么强的信号。所以有时候语义分析结果会出现一些错误，通用的多轮对话还是很难。因此具体应用的时候，多轮对话一定要考虑场景，如果把场景定义清楚了，你就可以很容易地定义状态，而在每个状态下可以提问的形式也是有限的，就可以做相应的推理。这样一来，多轮对话可能相应会容易一些。现在对话系统都是面向具体任务或者具体场景来设计的。

如果一个对话系统对应多个场景，就需要判断场景之间是否出现切换。只要判断进了某个场景，就调用那个领域的对话引擎（知识图谱、对话状态图谱），根据当前的状态来判断那个下一个回复应该怎么进行，等等。

当然有一些聊天机器人，比如微软小冰，不是完成某一个任务的对话驱动，它的技术跟面向任务的多轮对话有很多不同。这里就不多介绍了。

雷鸣：最近强化学习之父发了一篇文章，是关于算力推动整个计算机领域包括算法和技术发展的论点，文中主要观点是说我们的科研要跟着算力走。这点正好也映射了之前我们在课上提到过的，如 Bert、GPT 2.0 等依靠巨大的算力建立起预训练模型，在很多方面帮助自然语言提升各项能力或解决了一些问题。从一定意义上，你怎么看算力和自然语言下一步发展之间的关系？现在很多自然语言问题没有解决，会不会是算力没达到，还是说算法不够精巧？未来算力和算法之间会出现什么样的交替关系，或者有没有可能因为算力的高度提升，最终 NLP 能够计算机视觉一样得到终极解决？

周明： 首先，算力永远是重要的。现在好多 NLP 评测任务（机器翻译、阅读理解等），如果没有算力根本不可能上得来。算力的背后，体现的一种对知识（特征）的不断抽取过程，比如神经网络四层和八层，区别在哪？层数越高对特征抽取的能力就越强。当然也需要更强的算力。从这一点上来看，算力强，对输入信号表示能力和特征抽取的能力就强，当然其对应的解题的能力就更强。

但是有些问题只凭算力也无从下手。比如我们刚刚提到的 Low Resource 问题，在缺少训练语料的情况下，搞一千层、一万层也没用。这时候光靠算力的蛮力解决不了太多问题，可能要引入一些建模上的能力，或者引入人类知识。比如说将人类专家的几条翻译规则融入到神经网络机器翻译之中，进行冷启动或者增强基于数据的学习系统。这时候需要考虑的是，人类的知识模型或知识库如何跟数据驱动的模型巧妙地融合起来，发挥各自的特长。这时候

算力当然还是越强越好，但已经不是唯一重要的。如何建立人类知识体系（开放和领域相关的）并将其融入到原来的基于数据驱动的方法之中，这里还有很多值得研究的问题。

第二个问题就是多轮问题。应用神经网络方法，现在只要算力足够强、数据足够大，对单轮任务（比如单句级机器翻译任务就是典型的单论 NLP 任务，一个输入句子，对应一个输出句子）的能力已经非常强大，但多轮依然很难。因为多轮任务会出现动态的变化。比如多轮对话，用户会根据前一个回答再提出新的问题，我们无法提前把训练的输入、中间轮可能的回答、最后轮的输出等这些数据，大规模地标注出来，做成足够大的训练集合。多轮建模的时候还会涉及到 Memory 的问题（存储前后轮上下文得到的信息）、建立人类常识等。这些还没有到简单地凭算力就可以很好地解决的地步。

未来可能有两条路，一是基于数据驱动的。两件事，一把数据掌握好，二是把算力掌握好，就能把模型很好地训练出来。还有一条路就是基于知识，以及推理的这条线，它背后也要靠一些算力，但我们现在还没有到以算力取胜的阶段。当前可能如何对知识进行建模、如何获取知识、如何推理，整个理论体系并没有完全地形成起来。也许未来某一天，理论体系已经起来了，那时候大家又要比算力。这件事如果不恰当地比喻一下，就相当于革命在不同的阶段：数据驱动已经快接近共产主义社会了，而基于知识和推理还处于社会主义初级阶段。两者焦点不一样。等后者也搞清楚了，这时候加上算力推动，也许真有希望进入共产主义了，也就是说对多轮 NLP 任务可以很好地解决了。这样认知智能会进入实用阶段。

雷鸣：关于知识图谱和深度学习相结合，现在有没有什么新的研究在尝试，或者说未来会有什么发展？还是说这两个有点水火不容，很难真的融合起来？

周明：首先在 Offline 阶段，建知识图谱的时候会用到一些深度学习的方法，比如说信息抽取、分类问题、Relation 等，背后的技术可以用深度学习来做，但是建完之后就成为一个知识图谱了，它又变成符号化了。第二个就是在 Run time 的时候，怎么把知识图谱也融入到刚才所说的数据驱动里面。现在有一种办法，就把知识图谱也做 Embedding，即 Entity Embedding，可以根据知识图谱的前后左右周围的节点和边，对知识图谱中的每个节点和节点关系，用一个多维语义向量来表达。这跟词的 Word Embedding 是一样的，那么如果这两个 Embedding 是一样的，再往上走的时候也可以做 Attention，也可以计算 Encoder、Decoder。现在有很多任务都在沿着这个方向走，也有一定的效果，但我不认为目前这个领域取得了多么大的突破，可能还有一些新的探索的余地。

雷鸣：能否大胆地假设，多轮对话和计算机视觉只是信息接口不同，其实背后能够落到一个相似的领域中去？计算机视觉再往后发展，是否会跟自然语言的融合性会更强？

周明：未来人类跟机器最自然的交互是多模态的，有的是图像，有的是文字，有的是语音。现在做研究大都是针对每个单模态，先做好研究，比如图像识别、语音识别、自然语言处理。将来是多模态一起融合来编码和解码，比如对一个图片连续问答，实际上就是多模态处理。

遗憾的是这个方面尚缺乏相应的评测集和数据集来推动。我们刚刚提到的斯坦福大学建立的 GQA 数据集，实际上就是想把自然语言的问题提问跟图像理解融合到一起，考察背后的推理能力和答案抽取能力。如果这个数据集能有效地推动相关研究，我们以后就可以做更大胆的尝试，要么更大的数据集，要么把某些技术用到一个很狭窄的垂直领域里去看看结果。比如说地图领域，对着地图指指点点、说说话，看看是不是能做新一代的智能地图。可以做一些这样的尝试慢慢推动这个领域的发展。

雷鸣：未来三五年，你觉得自然语言在哪些领域会有比较好的进展？这个进展指的是能够真正落地，能做出来一些我们作为终端用户感受得到的产品或者服务的，或者有没有什么地方可能适合同学们未来几年创业之类的？

周明： 首先我们要考虑两件事，一个是研究，研究可以有自己的 Vision，可能短期实现不了，但是长期必然要走到某一个地方，那从研究角度就应该大胆地去研究。也许一两年没做出来很了不起的成果，但长期它总是驱动人类认知提升的一个动力，我们未来一定要走到那里。比如多模态问答，我觉得它就是人工智能一个终极目标，我们一定要做到。至于怎么做，可以先从单模态做起，再加双模态，再多模态融合；从简单的单轮问答，再到多轮问答；从一开始不需要推理，再逐渐需要推理，一点点来推进整个过程。

其次就是要去思考某些技术是否可以找到一个垂直领域把它用起来。可能是很窄的一个领域，但是用的特别巧妙。

比如说文本生成，现在已经可以做到给几个关键词，就把一篇文章，或一首诗，或歌词生成出来。现在的 Demo 都已经做得挺漂亮了，但是仔细去看其实前后的句子或者段落，不合逻辑，或者不合事实。目前需要研究的是如何把文本生成跟事实融合起来，使它生成的句子既逻辑合理，也体现事实。这件事如果往前推动，我认为是能做出来的。做出来之后可以快速生成大规模文本，可以做深度好文。做完之后，可以再由人工专家，就是编辑或者作家，来润色和修改确认。我认为这会对整个人类的文档生产过程产生巨大的影响。

还有信息抽取。对一个垂直领域，比如说金融、法律、医疗，做信息抽取，抽取之后形成知识图谱，基于知识图谱进行问答、搜索或者推理，甚至建立某一个垂直领域的专家系统，我认为一旦对某一个垂直领域做成知识图谱和推理，将会产生巨大的落地效果。诸如此类的场景，大家都可以去考虑，这是仁者见仁智者见智。另外不完善技术的应用，需要运用之妙存乎一心。技术不必也不可能非得达到百分之一百的好，也许某些场景下，巧妙的设计，对技术的要求百分之六七十就足够。如果用的特别巧妙，也可能在某一个领域产生相应的经济效应。

雷鸣：随着 BERT 和 GPT 2.0 的出现，NLP 是不是进入了比算力的阶段？另外 NLP 最近在挑战图灵测试吗？还有多远的距离？

周明：图灵测试从某种意义上是不是已经算解决了，要看怎么定义。以聊天机器人为例，在很多场合，比如微软小冰现在能聊 23 轮以上，我们没有去做图灵测试，如果真要去做，也有可能是突破了图灵测试。但是我认为真正的人类智能光凭传统的图灵测试是不能完全体现的，比如刚才提过的多轮事实类问答，事实不能错。除此之外，多模态对话、需要复杂推理的阅读理解、自然语言交互的专家系统（比如医疗诊断、法律咨询）等等，这些任务的智能水平，离突破图灵测试还需要很长的时间。

雷鸣：还有就是 NLP 是不是只比算力了，没点大机器就研究不了了？

周明：第一大家要尊重算力，过去很多搞人工智能的人都不服算力。但实际上，我认为要尊重这件事，算力体现了刚才所说的建模能力、信息抽取能力、解码能力，它不是简单的速度快了、容量大了的问题，而是有一个由量变到质变的过程。第二，我们要尊重算力但也不唯算力，要体现人类建模的能力、知识抽取、常识推理各方面，而恰恰常识知识推理这块没有一个人能说清楚，也没有一套成熟的理论和工具包，这块恰好是我们未来可以深入研究的。对高校的同学来讲，可能会觉得学校没有公司的算力强，要搞研究就要吃亏，但我觉得应该多去做一些刚才我说的后者，就是建模、知识推理、知识获取这方面的研究，这样跟以算力取胜的很多公司可以很好地配合。

雷鸣：未来同声传译有可能会被取代吗？如果有可能，需要具备什么前提呢？

周明： 似乎有人认为同声传译在一些场合有可能是可以被取代的，但是好好思考一下其实还有很长的路要走。目前语音翻译还有几个问题：第一，针对不同人语音特点的语音识别已经不错了，但是还有很大提升空间；第二，背景噪声对语音识别影响还是很大；第三，专业术语、新词影响对语音识别和翻译影响非常大；第四，凡是用同声传译的场合都是重要场合，它对错误的容忍度是非常低的。这不像网页上只需要把大概的意思翻译出来，即使有点错误，用户是可以容忍的，而同声传译的场合，只要一个重要人物的名字翻译错了，整个翻译就算失败了，而且可能有重大的影响。从这个意义上来看，要达到这么高标准的要求，自动同声传译还有很长的路要走。

这里其实有很多技术上的考量、实用上的考量、政治上的考量、投资回报上的考量，我认为不能简单用 yes 或 no 来回答，它是很漫长的一个过程。但是我们做技术的人，应该继续关注技术。把语音识别做得更好，把翻译做得更好，把 TTS 做得更好，更个性化。不过至于未来它能不能采用，有很多非技术的因素在起作用，现在不能一概而论。

雷鸣：时间关系，我们最后再代表同学问一个问题。很多同学现在正在学 NLP 或者对 NLP 感兴趣，包括专业的和非专业的两种学生，如果他们未来想做从事这方面的工作，你觉得他们在大学阶段或者研究生阶段应该怎么做，将来才能在这一块有所建树？

周明： 大家可能会觉得自然语言好像听起来很复杂。其实我想跟大家讲的是，进入了深度学习时代之后，自然语言的门槛一下就降低了。只要你会 Python 编程，网上找到训练语料，基本上就能把自然语言的第一个模型走出来了。请大家不要被一大堆公式所迷惑，还是要找一个具体的任务试一试。我建议大家拿机器翻译为例，把端到端的训练过程玩起来，沉浸其

中，很快就能理解整个自然语言的精髓了。第一件事做好了，比什么都重要。比如你把第一个机器翻译模型训练好，确实 Work 了，水平也还可以，至少和别人发表的水平差不多，你这时就会信心大涨。

只要有一个任务通了，其他自然语言的任务也可以通的。现在因为深度学习的原因，你会做机器翻译，就会做问答，就可能会做搜索，它背后的原理全部都一样。以前就不敢这么说，以前可能某位著名专家可能就是问答做得好，他做了一辈子。有的人 summarization 做得好，做了一辈子，它们之间不容易直接借鉴。所以那时候的门槛就非常高。现在只要懂了深度学习、比如编码-解码技术，把 NLP 主要领域都熟悉一遍是没有太大问题的。这样就有了“全栈”自然语言处理能力。这时候再考虑延伸到图像处理、语音识别，发现他们背后也是同样的编码-解码这些东西。所以又可以从自然语言走向其他领域，或者多模态融合。当然要做到世界顶级那还是要花点工夫的，但是做到普及，对它不怵，把它用在自己的工作场合或者应用之中，大家是应该有信心的。

观看本期公开课视频：https://v.qq.com/x/page/s0857c3z4d4.html