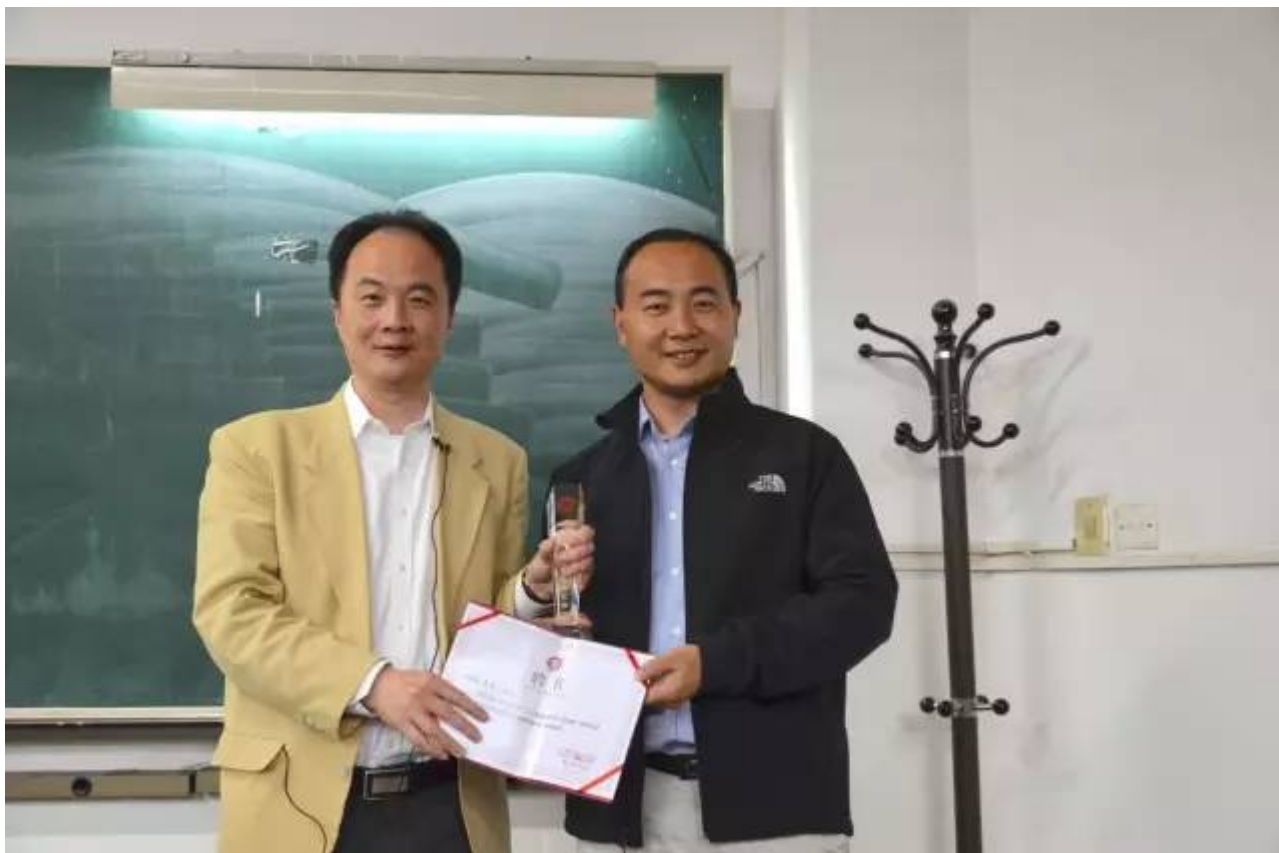


北大 AI 公开课第八讲：华为李航：自然语言处理 --- 理想与现实、机遇与挑战



华为诺亚方舟实验室主任李航老师亲临北大 AI 公开课第 8 讲，与北大人工智能创新中心主任、曾经的“百度七剑客”之一雷鸣老师一起，以“自然语言处理——理想与现实、机遇与挑战”为题，全面综述了 NLP 的任务、特点、最新技术，尤其是深度学习在 NLP 的应用。同时，两位老师就产品应用、处理难点、语义理解、知识库、通用人工智能、神经网络的记忆等多个大家关注的问题展开了深入的讨论和交流。

李航老师综述性地为大家介绍了 NLP 的任务、特点、最新技术以及发展趋势。李老师精辟地总结道：“给今天的讲座大概做一个总结，自然语言理解很难，自然语言处理现在用数据驱动的办法去做，有五个最基本的问题，即分类、匹配、翻译、结构预测和马尔可夫决策过程。

在具体的问题上，有了数据就可以跑 AI 的闭环，就可以不断提高系统的性能、算法的能力。深度学习在我刚说的五个大任务里的四个都能做得很好，特别是运用 seq to seq 的翻译和语音识别。单轮对话也能做的越来越好，但是多轮对话需要去研究和解决。”在对话部分，李航老师和北大人工智能创新中心主任、曾经的“百度七剑客”之一雷鸣老师一起，就产品应用、处理难点、语义理解、知识库、通用人工智能、神经网络的记忆等多个大家关注的问题展开了深入的讨论和交流。

李航老师的个人宣讲

Peking University
April 18, 2017

Natural Language Processing: An Overview

Hang Li
Noah's Ark Lab
Huawei Technologies

李航：今晚很高兴能够与大家交流，谈一下自己对于自然语言处理的一些看法，包括技术发展趋势和发展前景。讲座分为三部分，第一部分是自然语言处理主要解决什么问题，有哪些特点；第二部分是介绍自然语言处理的最新技术；第三部分是展望未来，自然语言处理方向的发展可能性。

Talk Outline

- *Introduction to Natural Language Processing*
- State-of-the-Art Technologies of Natural Language Processing
- Future Trends of Natural Language Processing

在座都对 AI 很感兴趣。人工智能就是希望计算机和人一样能够说话、理解文章内容、与人交互。自然语言理解是人工智能中很有挑战的领域，因为使用语言的能力是人所独有的，是最高智能的体现。

Natural Language Understanding

- Two definitions:
 - Representation-based: if system creates proper internal representation, then we say it “understands” language
 - Behavior-based: if system properly follows instruction in natural language, then we say it “understands” language, e.g., “bring me a cup of tea”
- We take the latter definition



对于自然语言理解，有两种定义。第一种是计算机能够将所说的语言映射到计算机内部表示；另一种是基于行为的，你说了一句话，计算机做出了相应行为，就认为计算机理解了自然语言。后者的定义，更为采用。

Five Characteristics of Human Language

- Incompletely Regular (Both Regular and Idiosyncratic)
- Compositional (or Recursive)
- Metaphorical
- Associated with Knowledge
- Interactive

为什么自然语言理解很难？其本质原因是语言是一种复杂的现象。自然语言有 5 个重要特点，使得计算机实现自然语言处理很困难：（1）语言是不完全有规律的，规律是错综复杂的。有一定的规律，也有很多例外。因为语言是经过上万年的时间发明的，这一过程类似于建立维基百科。因此，一定会出现功能冗余、逻辑不一致等现象。但是语言依旧有一定的规律，若不遵循一定的规范，交流会比较困难；（2）语言是可以组合的。语言的重要特点是能够将词语组合起来形成句子，能够组成复杂的语言表达；（3）语言是一个开放的集合。我们可以任意地发明创造一些新的表达。比如，微信中“潜水”的表达就是一种比喻。一旦形成之后，大家都会使用，形成固定说法。语言本质的发明创造就是通过比喻扩展出来的；（4）语言需要联系到实践知识；（5）语言的使用要基于环境。在人与人之间的互动中被使用。如果在外语的语言环境里去学习外语，人们就会学习得非常快，理解得非常深。

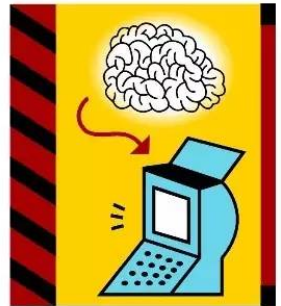
Natural Language Understanding by Computer Is Extremely Difficult

- It is still not clear whether it is possible to realize human language ability on computer
- On modern computer
 - The incomplete regularity and compositionality characteristics imply complex combinatorial computation
 - The metaphor, knowledge, and interaction characteristics imply exhaustive computation
- Big question: can we invent new computer closer to human brain?

这些现象都说明，在计算机里去实现与人一样的语言使用能力是一件非常具有挑战性的事情。首先，语言的不完全规律性和组合性，就意味着如果在目前的计算机上去实现，会产生组合爆炸；还有，如果需要语言做比喻，去联系到实践环境，就意味着要做全局的、穷举的计算。如果通过现代计算机来做，非常复杂，几乎不太可能。所以，**如果想让计算机像人一样使用语言，原理上需要完全不同的、与人脑更接近的计算机体系架构。**

Reason of Challenge

- A computer system must be constructed based on math
- Open question: whether it is possible to process natural language *as humans*, using math models
- Natural language processing is believed to be AI complete

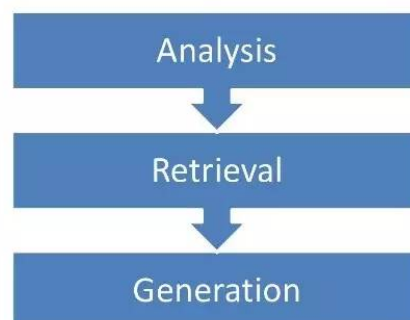
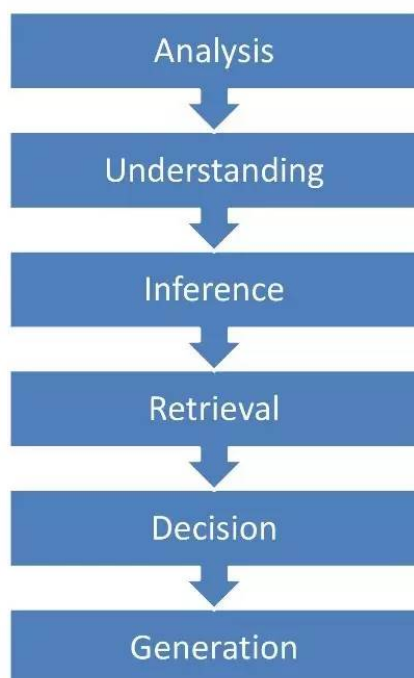


其本质原因是，目前在计算机上去实现东西一定需要数学模型。换句话说，计算机能够做的事情要通过数学形式化。但是，到目前为止，语言的使用还不清楚是否能够用数学模型去刻画。人工智能的终极挑战就是自然语言理解。现实当中，不能因为自然语言理解非常困难就放弃。我们还是希望能够使计算机越来越智能化，能够部分使用语言。因此，就形成了所谓自然语言处理这一领域。我们叫自然语言处理，而不是自然语言理解，因为真正的理解是太难了。

Simplified Problem Formulation



- Eg., Question Answering



Question answering, including search, can be practically performed, because it is simplified

自然语言处理做的第一件事情就是把问题简化。比如，知识问答中，问姚明身高是多少？朋友告诉你是 2 米 26。这是人与人之间的知识问答。那么，这其中有哪些步骤呢？首先是听，然后去理解问题，然后去做一定的推理，然后再去做信息检索，最后判断怎么去做回答，整个过程相当复杂。我们现在做自然语言处理时，也做这种知识问答，包括有名的 IBM 的 Watson，其整个步骤也是简化了自然语言处理的过程。一般而言，就是这几个步骤，先分析一下问句，接着去检索相关的知识或者信息，然后产生答案。

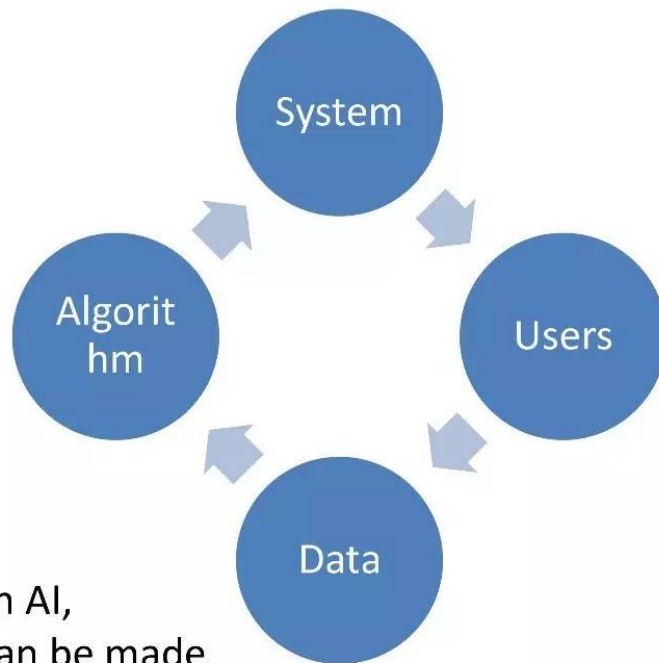
Data-driven Approach Works

- Hybrid is most realistic and effective for natural language processing, and AI
 - machine learning based
 - human-knowledge incorporated
 - human brain inspired
- Big data and deep learning provides new opportunity



自然语言处理领域发展的 60 多年中，总结出一个经验，就是目前最好的方法是机器学习，包括深度学习。也就是，基于机器学习，并在一定程度上把人的知识加进来，并参考人脑的机理，从而构建更好的机器学习办法。在短期内，自然语言处理很难突破这个框架。我们未来能够有更大的突破，人工智能能够完全超出目前基于机器学习的方法去做自然语言处理甚至自然语言理解。但是，目前的其他的路径都非常困难。现在，我们拥有大数据，使得我们能够去更好的做自然语言处理。

AI Loop



Advancement in AI,
including NLP can be made
through the closed loop

现在，无论是自然语言处理，还是人工智能的其他领域，都形成了一个闭环机制。比如，开始有一个系统，然后有用户产生大量的数据，之后基于数据，开发好的算法，提高系统的性能。如果能够闭环跑起来，就可以去收集更多的数据，可以开发出更好的机器学习算法，使得人工智能系统的性能能够不断提升。这个人工智能闭环是现代人工智能技术范式里最本质的一个现象，对于自然语言处理也不是例外。我们可以通过闭环，不断去开发新的算法，提高自然语言处理系统的性能。

Fundamental Problems of Statistical Natural Language Processing

- Classification: assigning a label to a string

$$s \rightarrow c$$

- Matching: matching two strings

$$s, t \rightarrow \mathbf{R}^+$$

- Translation: transforming one string to another

$$s \rightarrow t$$

- Structured prediction: mapping string to structure

$$\begin{array}{ccc} s & \xrightarrow{\quad} & s' \\ \uparrow & & \\ \mathbf{D} & & \end{array}$$

- Markov decision process: deciding next state given previous state and action

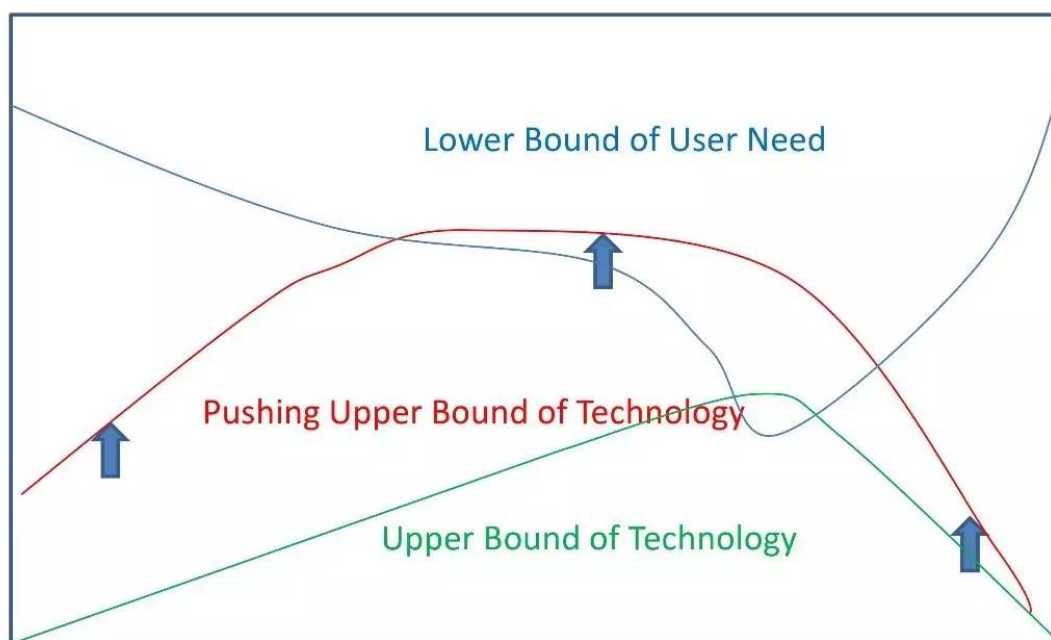
目前，所有的自然语言处理的问题都可以分类成为五大统计自然语言处理的方法或者模型，即分类、匹配、翻译、结构预测，马尔可夫决策过程。各种各样的自然语言处理的应用，都可以模型化为这五大基本问题，基本能够涵盖自然语言处理相当一部分或者大部分的技术。主要采用统计机器学习的方法来解决。第一是分类，就是你给我一个字符串，我给你一个标签，这个字符串可以是一个文本，一句话或者其他的自然语言单元；其次是匹配，两个字符串，两句话或者两段文章去做一个匹配，判断这两个字符串的相关度是多少；第三就是翻译，即更广义的翻译或者转换，把一个字符串转换成另外一个字符串；第四是结构预测，即找到字符串里面的一定结构；第五是马尔可夫决策过程，在处理一些事情的时候有很多状态，基于现在的状态，来决定采取什么样的行动，然后去判断下一个状态。我们也可以采用这样的模型，去刻画自然语言处理的一些任务。

Fundamental Problems of Statistical Natural Language Processing

- Classification
 - Text classification
 - Sentiment analysis
- Matching
 - Search
 - Question answering
 - Dialogue (single turn)
- Translation
 - Machine translation
 - Speech recognition
 - Hand writing recognition
 - Dialogue (single turn)
- Structured Prediction
 - Named entity extraction
 - Part of speech tagging
 - Sentence parsing
 - Semantic parsing
- Markov Decision Process
 - Dialogue (multi turn, task dependent)

分类主要有文本分类和情感分类，匹配主要有搜索、问题回答、对话（主要是单轮对话）；翻译主要有机器翻译，语音识别，手写识别，单轮对话；结构预测主要有专门识别，词性标注，句法分析，文本的语义分析；马可夫决策过程可以用于多轮对话。我们可以看到，自然语言处理里面有很多任务，在现实中我们已经开始使用最基本这五种最基本的模型它都去可以去刻画的。

Lower Bound of User Need vs Upper Bound of Technology



自然语言处理，在一定程度上需要考虑技术上界和性能下界的关系。现在的自然语言处理，最本质是用数据驱动的方法去模拟人，通过人工智能闭环去逼近人的语言使用能力。但是，这种技术并没有真正实现人的语言理解机制。可能会有这样的情况，这个技术的准确率（绿线）画了一个上界。比如，语音识别的上届是 95%，我们希望不断把这个技术做好，比如通过人工智能闭环，更好的深度学习方法，从而使得上界不断提高。但是，不可能一下子达到百分之百对，或者达到完全与人一样的水平。每个应用，对于下界的要求是不一样的。比如，在葡萄牙问路，对方也不会英语，我也不会葡萄牙语，交流非常困难，在这种环境下我其实就是听懂几个单词，让机器翻译给我翻译几个单词就行了，对性能的要求其实是比较低的，不需要去翻译一大段话。我们可以看到，不同的应用，用户对使用性能的要求不同，如果下界达到这个水平，用户就用了。再比如互联网搜索中排序第一的准确率不高，60% 多 - 70% 多，大家往往觉得，互联网搜索引擎已经达到要求了。当然因为搜索的时候，通过排序展示给用户多个结果，用户可以去逐个去看，一定程度上解决一些问题，这时候对性能要求下界相对就比较低。如果，现在的技术上届达到了用户要求的下界，就能够使用。所以，哪些自然语言处理的技术未来能够起飞，能够真正实用化，就可以通过这种关系来看。还是要看具体的应用的场景。在一些特定场景下，准确率达到 99% 都不行。我们相信，自然语言处理的技术会不断提高，但是是不是都能够达到我们每一个应用要求的性能的下界，就不好说了，要看未来的发展了。这是自然语言处理技术整个发展情况。

下面，给大家一起看一下我们自然语言处理领域里面都有哪些技术，有代表性的技术都大概达到什么样的水平，都是什么样的一些基础。假设大家对深度学习有一定的了解，如果这方面的知识还不够也没关系，过后你可以去再去看看书看论文，去了解一些相关的技术情况。

刚才，我介绍到站在一个很抽象的角度来看，自然语言处理就是五个问题。如果用各种方法包括深度学习把这五个问题做好了，就能够把自然语言做得很好。现实当中，我们就是通过深度学习，达到自然语言处理技术比较好的水平。

Applications

- Question Answering
- Image Retrieval
- Single Turn Dialogue
- Machine Translation

我们就看一下自然语言处理的几个应用。有一些也是华为诺亚方舟实验室开发的一些技术，也是业界代表性的技术。也给大家介绍一下谷歌的机器翻译技术。

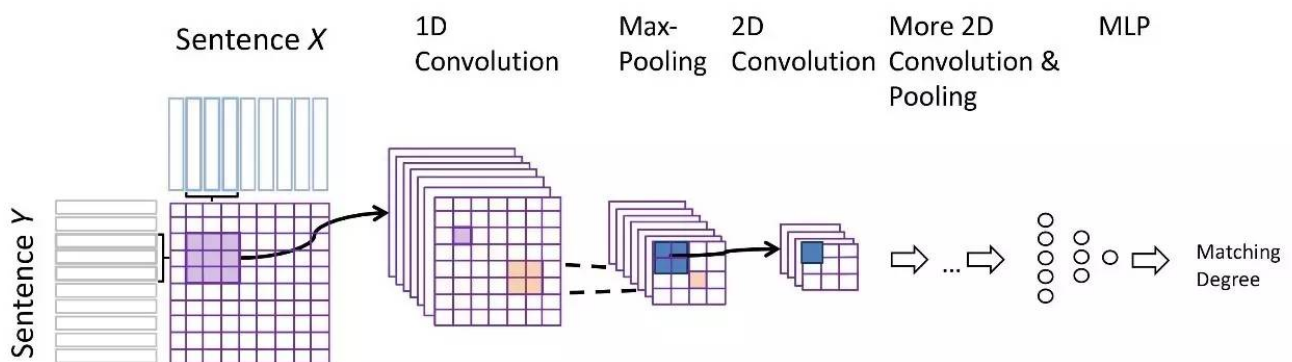
Question Answering

- DeepMatch CNN

首先，问答系统有很多，包括 IBM 的 Watson 也是一个问答系统，有大量的知识或者信息放在知识库。典型的办法就是把问答用 FAQ 索引起来，与搜索引擎相似，如果来了一个新问题，有一大堆已经索引好的 FAQ，然后去做一个检索（字符上的匹配），之后逐个去做匹配，判断问句与回答的匹配如何。往往匹配模型有多个，再去将候补做一个排序，把最有可能的答案排在前面，往往就取第一个作为答案返回给用户。

Deep Match Model CNN

- Represent and match two sentences simultaneously
- Two dimensional model
- State of art model for matching in question answering



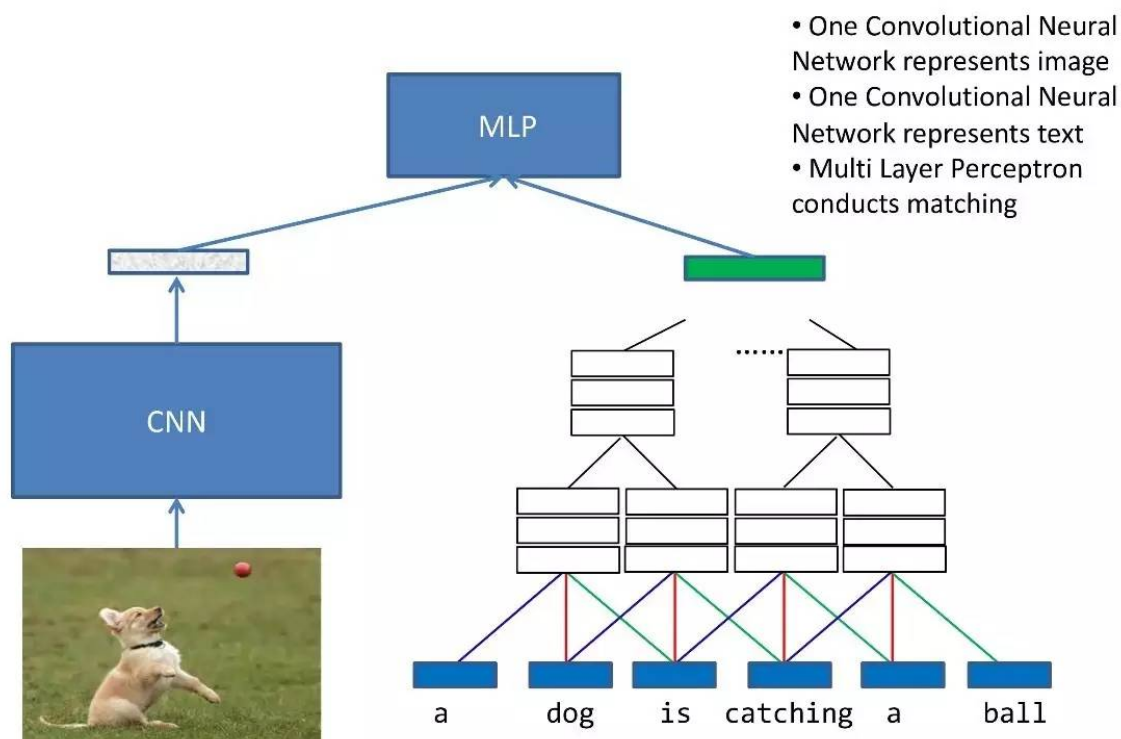
18

这里面牵扯到几个技术，我们在在线的时候要做匹配和排序，现在最先进的技术都是用机器学习，用深度学习技术。就是把问句和回答的可能的候选，用向量来表示，问句的每一个单词都可以用向量来表示。每一个词的语义都可以用一个实数值向量来表示，问句和候补都是实数值向量的序列。然后，用一个二维的卷积神经网络来判断两句话在语义上是不是相关，候选是否是很好的答案。通过二维卷积神经网络，可以判断两句话里面哪一些词语、词组是可以相互对应，最后可以做一个判断这两句话是不是相关的。整个模型的学习通过大量的数据、句对，去训练。如果卷积神经网络的参数学好，就可以判断任何给定的两句话是不是能够构成一轮问答。

Image Retrieval - Multimodal CNN

这样的模型不仅仅可以用到文本问答（知识问答）上，也可以用到图像检索上面。给大家演示一个 demo。

Multimodal CNN

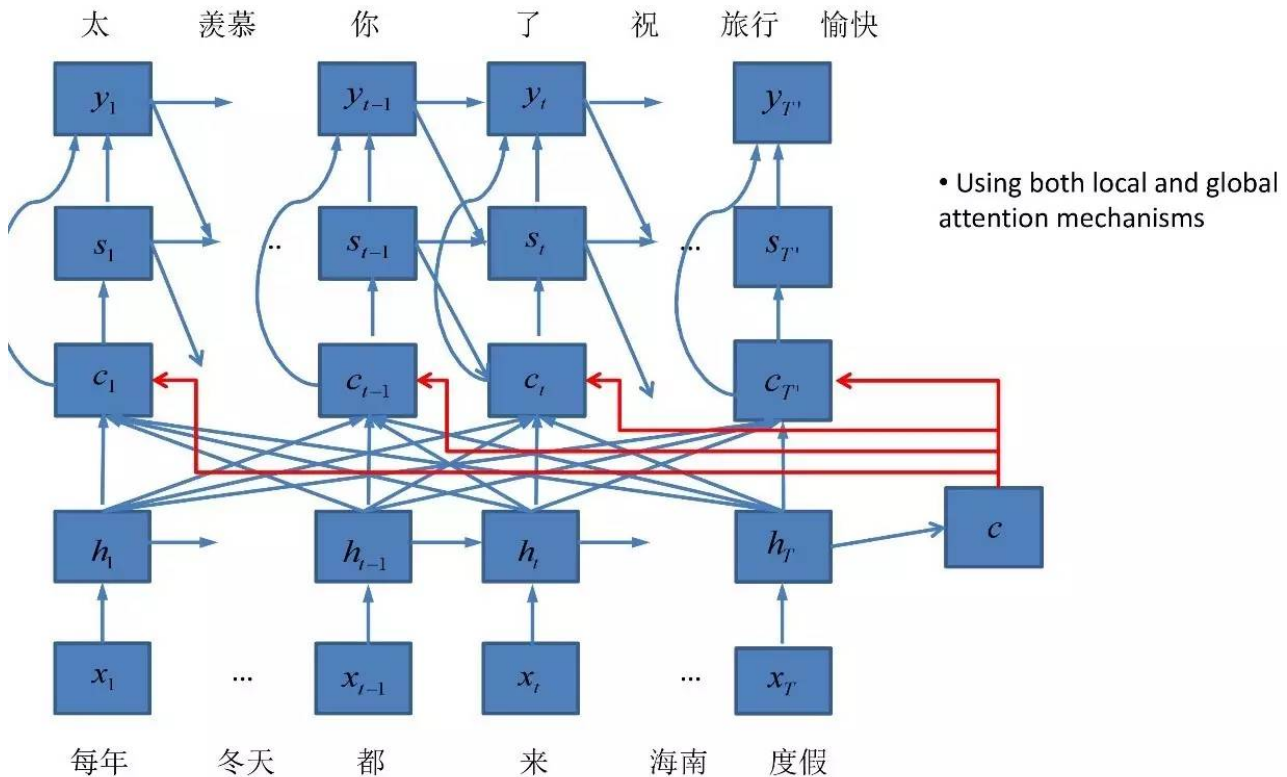


这种模型，可以跨模态的把文本和图片联系起来。在深度学习技术出现之前的话，这件事情是不可能的。因为他们是不同的模态。一个是符号表示的信息，一个是像素表示的信息，那么我们可以用深度学习的模型去做这种跨模态的匹配。比如，左边有一个卷积神经网络，他能够抽出左边图片的语义表示，表示成一个向量；右边也是一个卷积神经网络，能够把一段文字的内容抽取出来，表示成为一个向量，还有一个网络判断这两个向量在语义上是否能够匹配。这个模型可以通过大量的数据去训练。假设每一个照片有 3 到 5 个人给出描述。我们用大量这样的数据就可以学这样的神经网络，神经网络可以帮助我们，就是说任何给定的一句话，要去查找一个图片的内容，它就可以在这个图片库里帮你去匹配到最相关的图片，给你返回来。这个技术也是在深度学习出现之前应用的，因为我们不知道怎样把图片和文字匹配到一起。有了深度学习技术，我们可以做这样的事情。

Natural Language Dialogue - Neural Responding Machine

自然语言对话是用另外一种技术，用生成式的模型去做自然语言对话。大量的聊天系统是这么做的，输入一句话，里面准备了大量的 FAQ，搜索到一个最相关的回答，反馈给你。这叫做基于检索的自然语言问答系统。

Neural Responding Machine



我们这里看到是一种产生式，经过大量数据训练之后，输入一句话系统自动的产生一个回复，理论上产生出无穷多的不同的回复。下面先看一个实际系统的录像。

这个系统，我们在微博上爬了四百万的微博数据，微博数据可以看作是一种简单的单轮对话。我们用 400 万数据训练了这样一个系统能够去自动产生对话。（系统演示）

理论上它可以回答任何你输入的对话。用 400 万的微博数据就可以训练这样一个模型。系统产生一句话的比例是 96%，真正形成一个有意义的单轮对话的比例是 76% 左右。这个系统的一大特点是，可以回答没有见过的一句话。

第二个特点是能够记住训练数据。发现深度网络有一个共同的特点就是能够记住训练数据，同时也有去泛化的能力，能针对未知的新见到的东西去自动组织出一句话，并返回给你。这种能力很令人惊叹，是否实用并不清楚。在一个很固定的场景里，比如话务中心，如果话务员跟客户之间的交互都是简单的重复，大量的类似数据可以构建一个产生自动的回复系统，而且跟人的回复非常接近。

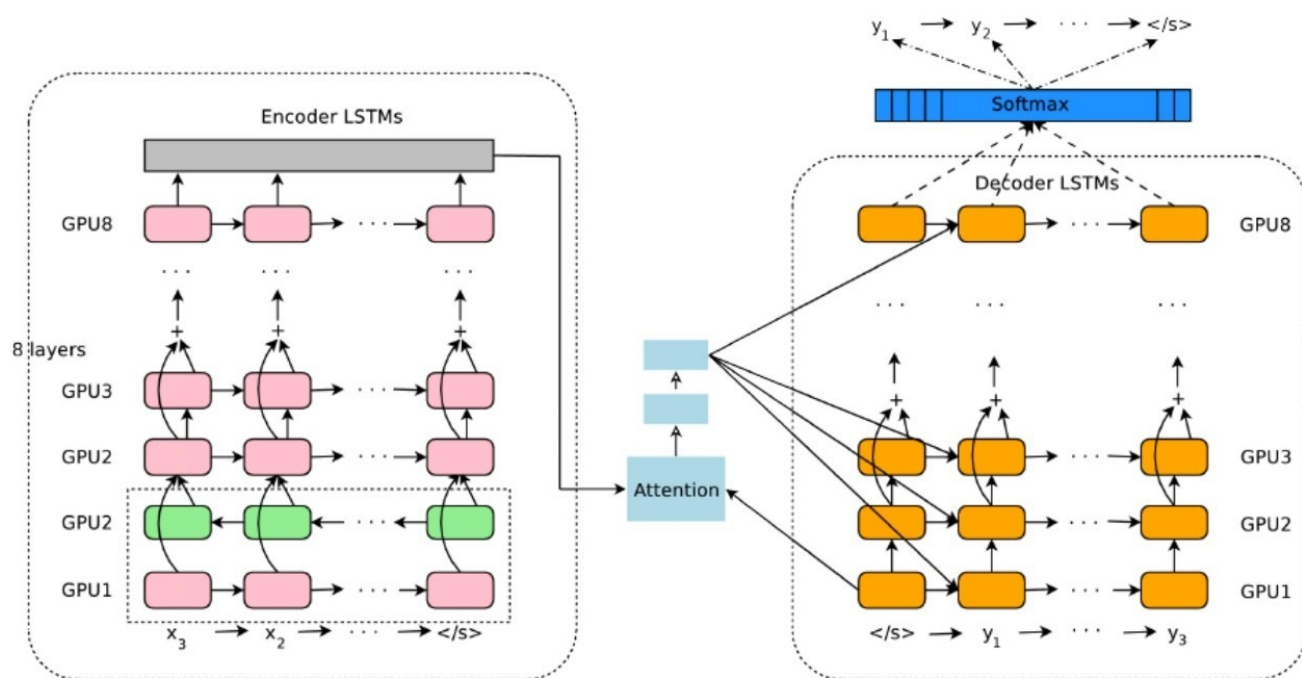
大家如果熟悉深度学习的话，刚才说自然语言处理有很多问题都是翻译的问题，即把一个文字的字符翻译成另外一个文字字符，那么单轮对话的产生也可以看成是机器翻译。序列对序列学习，sequence to sequence learning，可以用到这种单轮对话中。每个单词其实是用一个实数值向量表示，就是编码，之后用实数值向量分解成一个回复的一句，叫做解码。通过这种编码、解码这两个过程的话，我们把原始的数据转化成中间表示，再把中间表示，转换成为应该回复的话，可以产生对话系统。

Neural Machine Translation

- Google Neural Machine Translation

一个很重要的任务就是机器翻译，机器翻译的历史被认为与自然语言处理的历史是一样的。最近，大家也知道，深度学习，更具体的就是序列对序列学习，被成功地运用到机器翻译里，使得机器翻译的准确率要大幅度提升。

Architecture of Google Neural Machine Translation



谷歌的神经机器翻译系统是一个非常强大的系统，需要很多训练数据和强大计算资源。这个 seq2seq 模型有八层的编码器和八层的解码器，整个网络非常深。它还用了各种这个新的技术，比如注意力技术，并行处理技术，还有模型分割和数据分割等。目前，翻译的准确率已经超过了传统的统计机器翻译。

Performances

Task	Setting	Problem Formulation	Accuracy
Automatic Speech Recognition	Ideal Environment	Translation	95%
Dialogue	Single Turn	Classification or Structured Prediction	80%-90%
Dialogue	Multi Turn	Markov Decision Process	50%-70%
Question Answering	Single Turn	Matching	70%-80%
Machine Translation	Written Language Translation	Translation	70%-80% (derived from BLEU score)

下面，再看一下未来自然语言处理技术发展的前景和趋势。刚才我们看到技术上界和用户对于性能要求的下界，碰到一起就看到技术的使用化。那么，就预测一下未来自然语言处理技术的发展。目前，几个最基本的应用，包括语音识别，就是一个序列对序列学习的问题，就是翻译的问题，目前准确率是 95% 左右，那么已经比较实用了。单轮对话往往可以变成一个分类问题，或者结构预测问题，就是通过手写一些规则或者建一些分类器，可以做的比较准。很多手机上应用或者是语音助手像 siri，就是用这样的技术；多轮对话还很不成熟，准确率还远远达不到一般期待的要求，只有在特定场景下能做的比较好。单轮问答已经开始实用化，准确率一般来百分之七十八十，自动问答系统没有超过 80% 的这个准确率的情况。去年，我有一个报告就讲鲁棒的自动问答或者知识问答，并不要求准确率是百分之百。单轮自动问答会马上越来越实用化，因为我们看到很多成功的例子包括 Alexa 往往都是用单轮对话技术来做的。文本的机器翻译水平在不断提高，深度学习在不断进步，越来越接近人的专业水平，但只是在一些特定场景下。完全去替代人，还是不太可能。人的语言理解是一个非常复杂的过程，序列对序列实际上是一种近似，现在这种技术能够去无穷尽的逼近人，但是本质上还是跟人的做法不一样的。即使是这样，准确率可以达到百分之七十八十。在某些场景下，用户对性能要求并不是特别高。

总而言之，语音识别、机器翻译已经起飞，大家现在开始慢慢在用，但是真正对话的翻译还很困难，还有很长的路要走，但是说不定能够做得很好。并不是说序列对序列就没有问题需要解决了，还有细致的问题。一个典型的问题就是长尾现象。不常用的单词、语音识别、翻译还是做得不是很好。比如用中文语音输入，人名、地名这种专有名词识别率一下就下降，特殊的专业术语识也不好，讲中文中间夹杂一些英文单词也是一种长尾现象。因为现在机器学习的方法是基于统计的，原则上就是看到数据里面的规律，掌握数据的规律。需要看到甚至多次重复看到一些东西，才能够掌握这些规律。这块相信有很多技术能帮助解决一些问题，使得机器翻译或语音识别技术不断提高，但是完全彻底的解决还是比较困难，因为这是这种方法带来的一个局限性。

单轮的问答，特别是场景驱动的单轮的问答，可能慢慢会开始使用。但是多轮对话技术还是比较难。马尔可夫决策过程实际上是还是个统计学习模型，本质特点就是需要有大量的数据去学习。其实我们人在做多轮对话的时候，并不需要重复才能掌握这种天生能力。这些是否能够用马尔科夫决策过程去模拟或者近似还不是很清楚。还有一个重要的问题就是多轮对话的数据不够，不能够很好地去学习这样的模型，去研究这些问题。即使是特定任务，多轮对话还比较困难，如果是任务不特定，比如聊天机器人就更难了，我们都不知道该怎么去做，马尔科夫决策过程都用不上。现实当中的一些聊天机器人就是用单轮技术堆起来，但是形成不了一个很自然合理的多轮对话，变成大家用起来觉得很奇怪的东西。总结起来就是多轮对话，在任务驱动的简单场景，有了更多的数据，我们是有可能做的越来越好。

给今天的讲座大概做一个总结，自然语言理解很难，自然语言处理现在用数据驱动的办法去做，有五个最基本的问题，即分类、匹配、翻译、结构预测和马尔可夫决策过程。在具体的问题上，有了数据就可以跑 AI 的闭环，就可以不断提高系统的性能、算法的能力。深度学习在我刚说的五个大任务里的四个都能做得很好，特别是运用 seq toseq 的翻译和语音识别。单轮对话也能做的越来越好，但是多轮对话需要去研究和解决。

自然语言概括的那部分其实我也写过一些文章，大家感兴趣的话也可以去看一看，网上也能搜得到，然后还有就是我们相关的工作论文，包括谷歌的工作论文，我在这里列出来了。

最后，欢迎大家加入我们的实验室。方向有语音、语言处理、推荐搜索、大数据分析、智能通讯网络、计算机视觉、物联网、智能城市、智能终端。谢谢大家。

李航 VS 雷鸣 对话部分

雷鸣：特别感谢李老师精彩的讲座。今天这个讲座基本上对自然语言的整个发展能解决什么问题做了一个综述，而且对于技术、挑战和未来展望讲得特别全面，以致于我想到一个问题，后面就已经在回答了，今天讲得非常全面，非常仔细。我们想跟李航老师再探讨一下应用方面，我们知道华为的诺亚方舟做了很多计算语言方面的工作，能大概讲一下在落地方面做了哪些产品，大概现在处在一个什么水平上吗？



李航：好的。我们在语音、语言这方面做了两个应用，一个是机器翻译，一个是自然语言对话。机器翻译在我们公司内已经广泛使用，没有推到外面做产品。因为华为是一个非常国际化的公司，大概有不三、四万的非中国籍员工，所以中翻英，英翻中在公司内使用，我们的技术都应用在里面。还有云对话的应用场景就是手机。

雷鸣：助手。

李航：对，助手。手机的东西一直在做，现在不太方便说，欢迎大家到我们实验室访问，我们可以做进一步介绍。

雷鸣：李航老师也组织一次北大、清华、中科院的同学一起去参观，关起门来比较好说。提到对话这块很有意思，比较早的商用系统是 Siri, 包括在《生活大爆炸》里也看到调笑 Siri 的场景，最后发现也就是大家稍微玩一玩，后面就没有后面了。我们发现比较实用化的对话系统反而是亚马逊做的 Echo，客观地讲亚马逊在自然语言的技术积累以前没有见太多，比起苹果、谷歌、甚至百度都不那么强，那为什么它能先做出一个特别落地化的东西？它走了一个什么样的路径？对我们技术落地有什么启发？能大概解释一下吗？

李航：好。据我所知道的情况，亚马逊收购了几个公司，问答那部分是英国的剑桥做的，他们已经做了多年的这种问答，它们做得好的地方在于细节处理得非常好。也就是说问答的技术是亚马逊买来的。**还有麦克阵列那些好像也不是自己开发的，细节我不是很清楚。但是自动问答我很清楚，就是收购的技术。**

问答，或者说是广义的对话，刚才笼统地讲，有三类不同的技术，他们之间其实并不是包含和被包含的关系，是相互独立的。一种是分类或者结构预测，直观来说就是人手写规则，Siri 或者以前典型的场景都是基于这种技术做的。写好规则，这句话匹配上了，或者叫分类，分类对了，就去做了。还有一种技术就是问答，这种基于搜索、检索技术的比较多，有索引、排序这套东西，这你也是专家。再有就是多轮对话，比如说强化学习。这三套技术其实相互都比较独立。

说到匹配的话，模板、规则，或者说分类比较适合命令型的东西，未来就是家居各种场景，包括手机的命令，这种场景里面准确率相对也比较高，因为相对场景比较局限，能够达到百分之八、九十的识别准确率，甚至更高一些，应该是能够比较多地实用化。我们看到未来这是一个很好的场景。还有就是你刚才讲到 Alexa 的问答已经做得很好，这也是我觉得未来能够去实用化的一块，因为 Alexa 已经迈出非常好的一步了。它可以不断地跑人工智能闭环，收集更多的数据去把这个东西做得越来越好，就是内容不够填内容。如果是说“理解”用户的问法，还是用我们这种匹配、排序的技术，能看到哪个地方有问题就可以去改进，能够不断地把对话、问答这种东西做得越来越好。这两块我觉得都是未来能够起飞，能够用起来的技术。

多轮对话，刚才也说到，就是还需要很多研究了。数据也不够，大家都没有数据，是这样的一个状况。

雷鸣：可不可以理解为，他们虽然没什么积累，但是买的公司还挺厉害的。

李航：对，还挺厉害的。

雷鸣：有很牛的技术。第二点来讲，它进入到家庭场景里面，这个场景本身的限制导致说这个问题被降维了。

李航：对，就是实现做得比较好。

雷鸣：Siri 相当于一个开放式的，所以难度比较大一些。做到大家都满意就比较难一些，因为技术并不是一步到位的，选择技术能解决的问题去解决，解决实际问题然后再落地是更现实的事，可能会比较有挑战。谷歌成立很久，其实为全球培养了大量的科学家，出来了无数的创业公司，但是到现在好像还没有一个商业化成功的案例，所以他们走得有点儿太远了。

李航：是的。

雷鸣：刚才李航老师讲了好几次关于人类的语言，就是自然语言这一块它的复杂度是蛮高的，可能超过了用数学公式表达，或者是用概率就能搞定的。现在由于大量的数据积累，包括深度学习，使得在简单的语言问题上我们看到一种可解性。面对将来复杂的语言问题，我们现在能不能看到一个路径将来能够走到那一天？比如说这个路径大概是什么样子？现在在学术界，包括您这里有什么看法？比如说刚才讲这个多轮对话是个非常有挑战的问题，今天我们看到确实挺难的，多轮开放就更难了。但是有没有一种方向性的东西使得我们能够往那个方向走？

李航：这个问题问得很好，大家现在往往就是对人工智能过于乐观，包括自然语言处理也是一样，整个人工智能也是这样，还需要很多努力，还有漫长的路要走。面向未来的话，我们诺亚方舟实验室做研究一个大的方向就是怎么把知识和深度学习这样的技术结合起来。**换一个角度就是说把符号处理 symbolic processing 和神经处理 neural processing 结合起来，这能够帮助我们做很多事。**不一定能解决你说的多轮对话的问题，但是能帮助我们人类做很多事情。

你们可以这样想，计算机第一个是计算比咱们人类厉害，第二就是存储比人厉害。现在有互联网，各种信息库、知识库，但是我们觉得用起来还不是很方便，很多问题不是简单地通过搜索引擎就能去做。一个关于 symbolic neural processing 就是神经符号处理的想法就是，给计算机大量的文本，都是用符号表示的知识和信息，让它不加休息地去读，然后结合深度学习的技术，让它学到更好的知识表示、语义表示。对于每个人来说就是一个智能助手，帮助你去记忆各种东西，包括具体的信息，包括知识，这对我们人的能力是一种更大的延伸。不一定是多轮，但是是简单的多轮。实际上是把我们整个知识、信息的获取、检索的这样一个大的任务都解决，对我们每个人的能力是一种延伸。不光是知识，信息也可以融合起来。比如说，上次我和雷老师见面谈了什么都可以很快地导出来。这就扩大了我整个的能力，记忆的能力，存储的能力。这方面我们看到一些可能性，不好说是不是能突破，这也是自然语言知识问答的延伸，目前还做不到，做得不好，我们正在往这个方向一步步走，这是我们希望有突破的。

今年1月份我们去蒙特利尔访问了深度学习三大牛人之一 Bengio，请教了他关于 neuralsymbolic processing 的看法，他也是比较认可。他认可这种意义的结合，但是在一般意义上，如在深度网络里加一些 symbol，他认为是不对的。刚才说的这种意义上的结合他觉得还是有道理的，当然还有很多未知的问题，很多挑战，但是是值得进一步去探索的。我们现实中也开始在做一些研究。

雷鸣：这个回答大家去品味一下，技术发展无止境，有些时候一些特定的技术方式能解决一些特定的问题，但有时候也不是完全通用的。今天深度学习虽然很火，但是现在也有些反思，它是不是能解决所有问题。不同的学者有不同的想法，大家要用自己的智慧去理解。

刚才李航老师讲到 symbolic neural processing，从产业里我们有时候会从工程看问题。比如说自然语言这方面，第一是从语音到文字，如语音识别，现在专门有人做这个，做得还不错。第二就是从文字到语义，就是我知道你在说什么。第三层就是 response，就是我知道你在说什么，我再给你一个有效的反馈。从这三层来看，第一层做得还不错，第二层就是机器看到一段文字，到底理解没理解这个是否有个定义，咱们待会儿探讨一下。第三层就是来一个问题给一个正确的反馈，这和对问题的理解层面纠缠有多深？因为我看到现在都是把问题和答案对着训，把内涵加进去了。就比如说对牛弹琴有两种，一个是它根本听不懂你谈的东西，第二是听懂了但它不知道是什么意思。聊天也一样，对方说的每一个字我都懂，但是合起来不懂，或者说合起来说的我也都懂，但是不知道怎么回答。从学术上看，刚才讲到语音分开得比较清楚，就是语义理解和回答这两块的研究我看一体化比较严重，从长远来看应该是分开还是合并的？每一块有什么挑战？

李航：这个问题问得非常好。这就牵扯到自然语言处理的本质问题。**我个人观点，说到人工智能、自然语言处理还有很长的路要走的意思就在这儿。**一个就是要任务驱动才能去做，撇开任务单纯讲语义这是很难的，包括回答。深度学习的好处就是我们现在能做端到端的学习，输入、输出，里面都是黑箱，学习就好了。不好的地方就是中间发生什么都不知道。你刚才说对应人的语义，这个都不知道。深度学习肯定就有局限了。理想就是能够把人的支持加进来帮助语义的理解。刚才讲到 Bengio 也觉得这个事情不太好做，当然他也不一定绝对就对，但是这块是挺有挑战的，有太多不知道的事情了。

雷鸣：有一次我看到你提到谷歌的翻译，他们做的其实还是很牛的。以前我们都是端对端，英汉、汉英训练一个模型，诸如此类很多模型。但是谷歌是训练了个挺通用的模型，比如说英中对译训好了，中法对译训好了，然后英法之间就直接可以开始对译了。

李航：那是另外的一些工作，但是都是相通的。我今天介绍的不是太一样。业内有这样技术，有这样的研究。

雷鸣：这个我觉得挺有意思的，核心是我们没有训过英法之间的对译。这也就意味着，某种意义上来说，我们感觉机器在深度网络学习的时候对人类的语言做了一个内隐的表达。它在英中、中英、英法学习中间有个层面学会了英语的表达，然后就直接对上了。是不是说它找到了人类语言的某种内在表达方法？但是因为刚才说到的深度学习的特性导致我们读不出来，可不可以这样理解？

李航：我同意你的观点。就是这是做了一种表达。比如人做翻译其实也是一个很复杂的过程。你们有没有观察过专业的同声翻译，我只是从旁观察或跟他们交流，发现他们其实不思考的，他们已经形成了一种模式训练。不是所有外语好的人都能做同声翻译的，需要做一些特殊的训练。我的解释是他们其实是学了各种模式，但是他们很快，并不需要理解。他们就是有一种中间表示，很快就能转换成目标语言，我们现在的深度学习多少有点儿像那样一个机制，就是大量数据去训练。但往往我们一般人做翻译事实上是有语言的理解的，同声翻译至少局部就是一个模式。**所以我同意你的观点，就是让中法、中英这些都一起训练，可能针对某些语言映射到内部有一些表示，对于语义的理解是比较 universal 的，就可以通用。**

雷鸣：您刚才提到语义这一块您想说一下。

李航：对。有一篇文章是《迎接自然语言处理新时代》里面有写这个事情。你刚才谈到语音，语音只是人大脑里的一个模块，语言处理是整个大脑都会参与的，所以说语言本身就是很复杂。语音就是一个模块，所以语音处理不是一个人工智能完全的问题。语言处理是需要很多语言的知识模块参与在一起去做，而且大家可能也知道，脑区里面负责语言的部分都不止一个，多个脑区同时参与做这个事情。

雷鸣：占的面积也挺大的。

李航：对。现在大脑很多事情我们不知道。我们最早知道的一个有意思的事情是什么呢，就是给猴子的 premotor cortex 插上电极，发现有一个脑细胞在猴子自己吃香蕉和看到别人吃香蕉时都会有反应，说明在猴子的前运动域有一个脑细胞对应吃香蕉这个概念。然后人去做核磁共振这个实验，发现也是跟运动相关，比如说张开嘴或想象张开嘴，通过核磁共振去看脑区的反应的地方是一样的。让人去做某个动作和想象做某个动作是在大脑前运动皮质，而不是小脑，小脑是指挥你怎么样去运动，但是对应的大脑皮质有些运动的概念。有个假说就是有一个或多个脑细胞就是对应那个动作的概念。

现在有个我比较喜欢的假说，就是其实我们大脑皮质的一些细胞对应一些概念。比如喝水，看到人喝水，这个脑细胞可能就被激活，读小说读到有人在喝水，这个脑细胞也会被激活。所以说为什么读小说大家有身临其境的感觉，就是说视觉刺激和文字刺激都能刺激那个脑细胞，那一个或多个脑细胞就是对应这个概念。每个人理解语言的时候肯定是不完全一样的，因为每个人经验不一样。脑细胞被激活的过程是，成长过程中你的经历形成那些脑细胞激活的机制，相关的一些概念容易被联想出来，每个人容易联想的事情肯定是不一样的。但大家肯定有一些共性，如果差太多就没法交流了。脑细胞表示的概念还有很多共性的东西，使得我们能够做交流，能共同去做事情。既有共性，又有个性。

整个理解语言的过程就是激活相关的所有脑细胞对应的概念，把它们联系起来，然后还有自身体验的这种联系，这就是每个人对语言的理解。这个过程其实是非常复杂的。有人说意识占我们大脑处理的 2%，有个人极端的说法是下意识占 98%，就是说对语言的理解和处理实际上是在下意识中进行的，整个过程非常复杂，而且是并行处理，牵扯到里面的很多个模块，达到了所谓对语言的理解。说话、写和看到的東西都是 symbol，都是非常表层的东西。它背后牵扯到，产生或理解这个文字符号时背后的大部分东西都是在我们大脑里无意识的情况下进行的。这是非常复杂的，怎么去把它发掘出来，到底是怎么一回事儿，非常复杂。本身大脑规模又非常大，大家知道大脑有 10 的 11 次方的个神经元，15 次方的连接。

雷鸣：对，1000 亿的这么一个复杂的系统。

李航：对，就是这么一个复杂的系统，这么一个复杂的现象，我们要去再现这种理解的过程，从现在的技术来说是非常困难了。

雷鸣：其实对大脑的研究现在在相当初级的阶段，我也关注比较久，基本上就跟你说的一样，大概到脑细胞激活这个研究，只能对特别简单的一些低等生物做一做，对人还是做一些脑区分化和相关性的一些研究。

其实深度学习的发展跟对大脑的理解有很大关系。现在不知道还算不算，一段时间以前我记得还有两个派别，一个态度是说尽量要了解清楚大脑的结构，然后进入模拟大脑的过程，模拟得足够快，就会产生通用智能。还有一个就是说，造飞机不需要造出一个会扇翅膀的。我们大概了解清楚了，用机器的告诉照样可以超过。我们不知道哪个是最终答案，但我觉得研究大脑的结构肯定能够促进人工智能的发展，人工智能的发展反过来也会促进我们对大脑的研究，这是相辅相成的。

如果在座的大家对人工智能感兴趣的，还是比较建议大家去读一读神经生物学。神经的结构、大脑这些看一看会对大家很有帮助。大脑是怎么做决策的，意识、直觉、痛苦，行动这些都会讲到，挺有意思的。比如说我说一句话，不要想一个红色的苹果，你们脑子里有没有一个红色的苹果？所以劝人时不要说，你不要哭了，你别难受了，这完全起不到任何作用，因为他听到的就是难受，就是哭。小小地讲一下大脑很有意思的这一点。

雷鸣：自然语言研究里还有很大一部分，叫做知识库，就是尝试建立一个知识表达。现在知识库的研究在整个 NLP 中处于一个什么位置呢？在以深度学习方法为主流的 NLP 中，还会有它的位置吗？

李航：这是很热的一个领域，研究很多。但是这方面我看的不是特别多。我更关心的是结合应用。这跟你刚才提的另外一个问题相关，就是怎么样去定义知识。如果不是应用驱动的，而是纯粹去定义知识的话，到目前为止我们看到的都不理想。你建完了很大的知识库，也不知道该怎么用。就是说，知识表示是不是合理，如果没有一个明确的应用，就很难判断。目前对人类语言机制的理解并不清楚。应用驱动、数据驱动是我们的主要想法，知识库建设也应该是这样。大家现在在朝这个方向走，有一个大趋势。我不知道你们注意到没有，大概 4、5 年前知识图谱这个概念就很火。很多公司都在做。那时有些很有野心的项目，比如要做巨大的知识图谱等等。这些项目后来基本都停顿了。大家发现，真正要做这种通用的知识库还是很难的。就我了解的范围内的共识，是说其实可以做一些领域知识库，比如说医疗知识库。又比如说我们在华为做了通讯领域知识库。这是更现实的。在应用里面去使用，能够解决实际的问题。

这是知识库方面的一个趋势，或者说是动向吧。应该结合到实际的应用里面。你建了知识库无外乎就是希望大家去使用这些知识库。我们希望，**如果 Neural Symbolic Processing 这种新的技术有突破的话，如果知识的获取和检索能够解决的话，大家就能够更好地去使用知识库。**这是最基本最重要的应用。从这一角度来看，我们没有特别关注知识库本身，而还是以应用驱动、数据驱动，看神经和符号的结合。

雷鸣：李航老师从科研界出来，在产业界摸爬滚了一段时间，观点是典型的以应用为驱动。用得着的项目，我们就放进去用；用不着的，就先放在那里，先慢慢研究。我也是这样的人。（笑）

另外，刚才李航老师也说到，有一些我们认为很高大上的技术，现在作为通用的解决方案可能还非常困难。可能由于运算能力不够、数据不够，甚至是模型的复杂度不够，或是理论模型还不能支撑。这时我们可以在一个小领域里先应用它，降难度，限场景，先把它用起来，有时发现，在 **vertical** 的小领域里面，它还是挺好用的。

好，下面我们看一下同学、观众在我们的微信群、公众号和网上直播中提出的问题。一个同学问，主流的聊天机器人，比如小冰、小娜等等，现在主要是用什么技术实现现在的结果的？

李航：不说具体的系统，聊天的技术一般来说主要还是基于检索的。产生式的对话系统真正到了实用阶段的我们还没有看到，或者说看到的不多。因为有很多风险。有一个风险是，它说的话可能都是对的，但是它说的事实是错的。比如你问它，姚明身高多少？它说，1 米 2。这种时候你还能判断出它是错的。但有时候无法判断，就很码放了。深度学习不知道怎么去控制这个系统。不知道在什么时候能让它说出准确的答案。

我们现在其实也做了一些研究，还都是比较偏基础的一些东西，到使用阶段还有一些距离。即使聊天机器人，里面也还不是一些实用的技术，基本上都是基于检索的多轮对话的技术。简单说就是上下文对齐、指代消歧，这样的事情也能局部地做一做，但缺少一个整体的多轮对话的模型。我刚才也说到，现在任务驱动的时候，有马尔科夫决策过程，但如果是闲聊，是 **open** 的，那么都没有一个很好的数学模型去刻画这一过程。所以这还是非常难的一个事情。

雷鸣：下一个问题。为什么国际会议上都是用英文的数据集，大家比比去，为什么中文的 NLP 研究相对就少一些？

李航：没有啊，现在越来越多了。这是研究者的话语权的问题。现在做中文研究的学者越来越多，而且中文现在越来越重要。20 年前，中文的数据就更少了，现在中文数据已经越来越多了。同时英语也是作为国际性的语言，大家更容易去用。这不是什么大的问题。

雷鸣：好，有同学问，NLP 和创业相结合的话，有什么比较值得做的东西？

李航：这是个很好的问题。人工智能还是要跟具体的业务结合起来。自然语言处理也一样。这是第一个要定。第二个要点我刚才其实也讲到了，你需要去判断，你做的 NLP 系统其性能能达到的上界，和你面对的需求所要求的性能的下届是否能对上？这是非常重要的一个判断。如果你预测未来技术发展到某个阶段，能使上界提高到满足或超过需求下界的水平，就可以考虑结合实际的应用了。这个场景是非常多的。有很多场景里，我们都可以使用 NLP，来把它做得更好。**其实我整晚的课都在强调这两点：一个是应用驱动，或者说需求驱动；另一个就是这个上界下界的事情。**

雷鸣：李航老师高屋建瓴。我对创业这块比较熟悉一点。我们现在能看到 NLP 相关的创业，大体说有两类。一类是指令式的，比如智能家居，你下达指令，让它开灯关灯；车载环境下也是指令类。车载环境真的不方便用手了，智能用嘴去下达指令，你可能会说，给我老妈拨个电话，或是把刚刚收到的微信信息读一下。另一类是 QA 类的，基本应用在客服上。这一块最近用得非常多。大企业会自己做客服系统，借助自身大量的客户积累，去实现问题和回答的匹配。当然有一些中小企业和传统企业，没有这个能力，所以有一些创业公司就切进去了。这种情况也不少。另外还有一类，就是利用 NLP 技术对以前积累的数据和知识进行分析和处理，比如文本构成的知识库，像卷宗、病历等等，用 Watson 这样的系统去分析它，尝试发掘一些规则的知识。

大体上来讲，大公司，比如华为，有人力、财力、物力，再加上有数据，同时还有应用场景，你去跟它抢这个市场难度很大。也不是说完全没有可能，但难度很大。而有一些行业，比如医疗，大公司也没有数据，都要去抢。这相应来讲还有一定机会。

再问一个问题。上节课上徐伟老师讲过，有一个调研，调研了一百位科学家，这些人中有一半人都认为，2050 年之前，强人工智能有超过 50% 的可能性会实现。当然你可能不同意这种说法。显然，通用人工智能应该能理解人类语言。这是不是意味着自然语言处理在接下来的三十年间也会有很大的发展？或者说，到最后，NLP 的问题会等价于通用人工智能的问题？

李航：有一个说法我比较认可，就是未来的 5 年、10 年，我们可以预测。20 年的话，基本上就不能预测了。2050 年的事情，真的谁都很难预测了。返回头说，十年前我们能预料到语音识别会达到现在的水平吗？十年前可能没有人能够预测得到。很多东西不好预测，预测未来是一个挺难的事情。尤其是现在技术突飞猛进，发展这么快，各种路数都出来，你知道哪边突破了？这都很难说。

还有，通用人工智能也没有一个准确的定义。**在我看来，通用人工智能会在未来 10 年、20 年在一定程度上有突破，我也持这种观点。**如果我们把通用人工智能定义为把语言、视觉、听觉等所有这些能力综合起来的一种能力，那么是完全有可能的。因为传统上认为人工智能太难了，所以大家把它分而治之，研究视觉的，研究听觉的，等等。现在如果说要通过类似深度学习的方法把这些串起来，这是可能的。比如说你把图像识别和语言处理放在一起做。其实人在成长过程中，学习最基本概念的时候，图像、语言等等也是同时来学习的。从这种意义上说，未来甚至不用那么长的时间，就能看到一些成功的案例了。现实中我们已经看到一些多模态的智能结合了。但是这还不能说是达到语言理解了。我觉得从这个意义上说，语言理解就更难了。

雷鸣：对。上一次我跟徐伟探讨时，也说到了通用人工智能的定义问题。我们当时有一个简单的小共识，认为它不是用解决什么问题来定义的，而是可能要看它的学习能力。比如说跟我们人类一样的通用人工智能，你让它学开车，人用一个月学会，它也可以用一个月学会；然后还是同样的一段程序，去让它学围棋，它和人类一样，花了三年时间，变成了业余几段。从学习能力来考察，可以认为它是实现了通用人工智能。

李航：对对，这个我同意。可以从学习能力来判断。

雷鸣：对，就是同一套程序，干啥都行。这是蛮有意思的探讨。

李航：对，我同意。

雷鸣：未来的事情确实真的都不好说，你刚才说返回头看，其实就算往回推一年，去年的 4 月份，谁也不知道 AlphaGo 能赢人啊。那时候大部分人认为还是赢不了的。

再问一个问题的，刚才咱们也提了关于知识的问题。我们知道，神经网络没有一个明确的存储，虽然它存起来了，但不知道存在哪里了。现在有一种网络，叫 **memory network**，里面加了内存。未来发展它的潜力会不会比较大？加了 **memory** 之后，包括推理、自然语言处理等方面的能力会不会有一个比较大的提升？还是说现在看也就是一个架构而已？

李航：刚才我说的神经符号处理，就是其中一部分，我觉得这个东西是非常非常重要的。时间关系我今天不能讲太多。我今年还会做一两个报告来讲这个事情。人的一个很重要的特点是能够把过去的事情都记忆起来，可以把记忆里面这些知识和事实的信息都能够串联起来。机器现在就没有这样的能力。**AlphaGo** 就没有一个记忆的能力。如果能够把知识、信息不断往里存储，根据需要检索，如果有了这个能力，对机器智能来说是非常大的提升。

雷鸣：我们现在看不到 **neural network** 里显式的记忆点在哪里，但是我们发现它是有隐式的记忆的。就像你刚才说的，问答在它的里面跑了几回，它就记住了，好像是记在什么地方了。那么是不是一定要显式的记忆才会有效？还是说只要网络足够的复杂，其实它也通过网络结构实现了记忆呢？

李航：是这样的，它在记语言的使用模式的时候，比如句法啊、回答问题的反式啊，等等，隐式记忆也许是可以的。但像知识啊、信息啊，这种事实性的东西，我们还是希望它是显式的，让人能够看到，能够检验。

雷鸣：有可解释性。

李航：对，它需要是可解释的。如果长期记忆这种机制成功的话，神经网络能够不断去记住新的东西，那它就会越来越强大。

我再多说一句，好莱坞电影里面会有人和机器人谈恋爱。那太遥远了。第一步，机器首先需要有自己的意识。意识有很多定义，一个最基本的定义，是说如果一个系统能对外界的变化产生反应，那就是一种意识。最简单的意识，比如说温度计，或者向日葵，都可以看做是有最基本的意识的。

雷鸣：这么说人工智能已经有意识了？

李航：从这个意义上说，是的。但人工智能没有自我意识。它不知道自己是谁。第一，有了意识；第二，如果你能有记忆，你就又进一步了；第三，如果你有自我意识，你就有可能以此为基础生发出情感。这样才可能和人类恋爱。所以说，目前的人工智能可以说已经有了最基本的意识了；如果它现在又有了记忆，它能够把它整个的历史串起来了，我想这是整个智能机器在朝着自我意识的方向上又进化了一步。这件事是很激动人心的。

雷鸣：也有一点吓人（笑）。好，感谢李航老师的分享！

免责声明: 本站所有内容均来自网络, 我们对文中观点保持中立, 对所包含内容的准确性, 可靠性或者完整性不提供任何明示或暗示的保证, 请仅作参考。若有侵权, 请联系删除。

文章来源：新智元