《智能机器人设计》

智能机器人语音

NLP概述

自然语言处理（natural language processing，NLP）一直是人工智能研究的热点。自然语言是人类智慧的结晶，NLP是人工智能研究中最困难的课题之一，NLP的研究充满魅力和挑战。对智能机器人而言，NLP是实现机器人语音交互的基础，让机器人“听得懂”“说得清”，是衡量一个机器人是否真正具有智能的基本要求。

NLP发展历史

- 20世纪初，瑞士日内瓦大学的语言学教授Ferdinand de Saussure发明了一种将语言描述为“系统”的方法，即结构主义语言学，Saussure教授也被后人称为现代语言学之父。
- 1947年，美国科学家Weaver博士和英国工程师Booth提出利用计算机进行语言自动翻译的设想，机器翻译（machine translation）从此步入历史舞台，NLP也通过机器翻译的研究得以进一步发展。
- 1964年，首个自然语言对话程序ELIZA诞生，由于当时计算能力有限,ELIZA只是通过重新排列句子并遵循简单的语法规则，实现与人类的简单交流。这一时期的NLP(称为第一代NCP)。

NLP概述

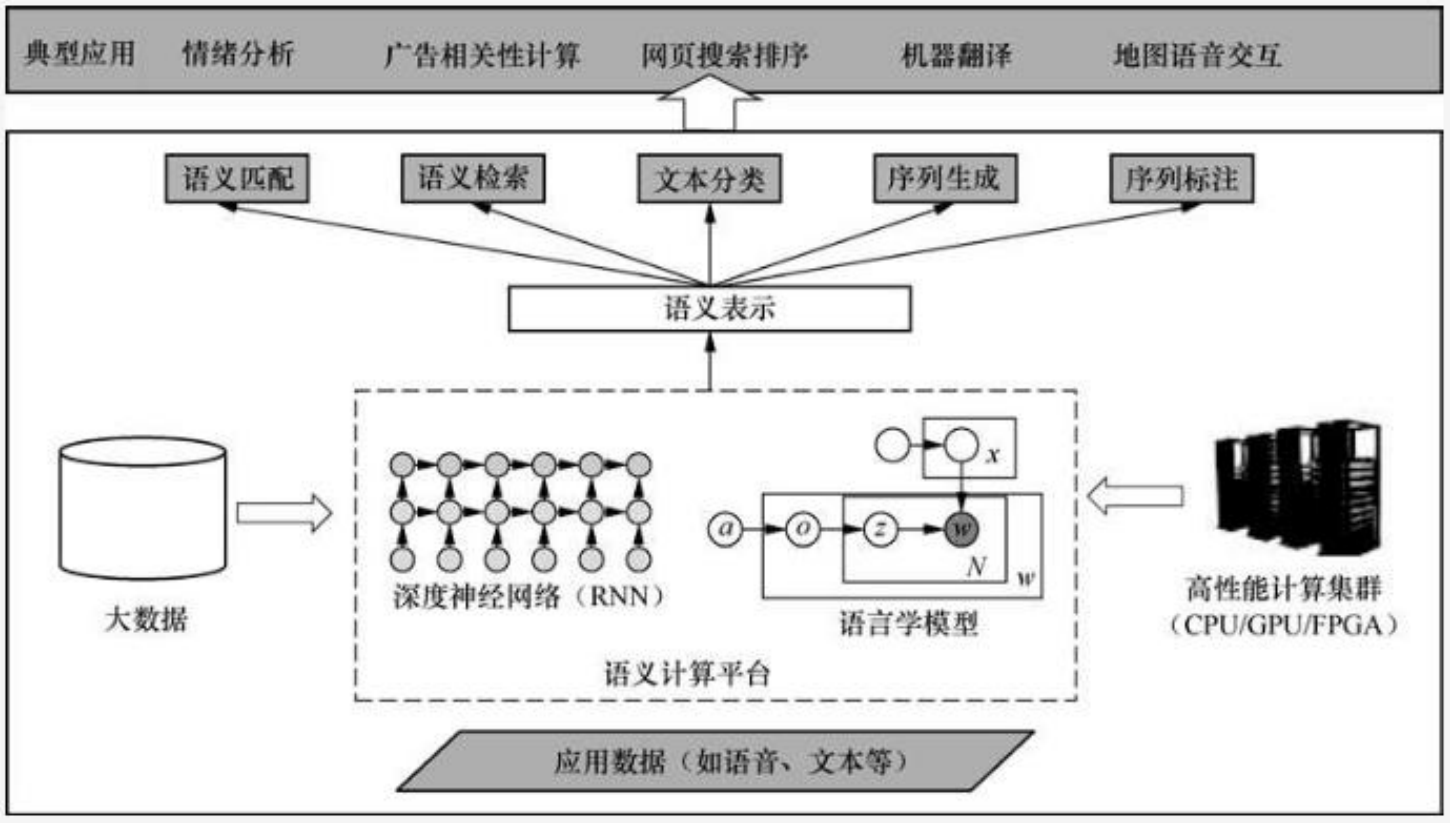
NLP发展历史

- 1966年，机器翻译的成本还是远高于人工翻译，并且没有任何计算机能够实现基本的对话。针对机器翻译的科研资助都停止了，NLP的发展陷入停滞状态。
- 20世纪80年代，得益于计算能力的稳定增长，以及机器学习的发展，早期的机器翻译概念被推翻，基于统计学和机器学习的NLP新流派（第二代NCP方法）诞生。
- 20世纪90年代，随着互联网的出现，用于NLP的统计模型迅速普及。纯粹的统计学NLP方法在网络自然语言处理中变得非常有价值。n元模型(n-gram)在跟踪大量的语言数据方面发挥了重要作用。为了缓解n元模型估算概率时遇到的数据稀疏问题，学者们提出神经网络语言模型。1997年，递归神经网络模型出现，并找到了语音和文本处理的利基市场。2001年，人工智能著名学者Yoshio Bengio教授发表了一篇论文，提出一种全新的语言神经网络模型，掀起了基于神经网络的NLP学术研究热潮。第三代的NLP研究进入新阶段，是传统机器学习方法在新一代人工智能背景下的延伸和拓展。

NLP概述

★ NLP 新技术框架

第三代NLP方法以深度学习作为机器学习的主要方法，实现了语言和文本的分布式特征表示，构建了基于大数据驱动的机器学习新模型，加速了NLP应用的落地，推动了NLP在实际生活场景中的广泛使用。第三代NLP整体技术框架如图所示。



该框架充分利用大数据时代的数据资源，依托云计算将深度学习与自然语言处理任务相结合，分析、挖掘、学习语音、文本等时间序列的深层语义特征和语音表达，提供语音识别、机器翻译、广告推荐、资讯引流、语音交互等典型应用的解决方案。该技术方案的典型特征是实现了大数据、云计算、深度学习的一体化——大数据为燃料，云计算为平台，深度学习为引擎，从而为各种边缘设备赋能。

NLP的深度学习模型和方法

★ 递归神经网络

基于深度学习的NLP新方法在当前的NLP领域取得了巨大的成功，这主要归功于递归神经网络（recurrent neural network，RNN）的发展。与CNN不同，RNN擅长处理“时间序列”数据，因此在NLP领域发挥了重要作用，主要应用领域包括语音识别、语义理解、机器翻译等。从特征工程的角度来说，RNN能够记住时间序列前面时序的特征，并根据记忆和遗忘机制推断后面时序的结果，如此不断循环，以拓展并优化模型的性能。RNN的原理与人类的NLP和记忆力机制不谋而合。

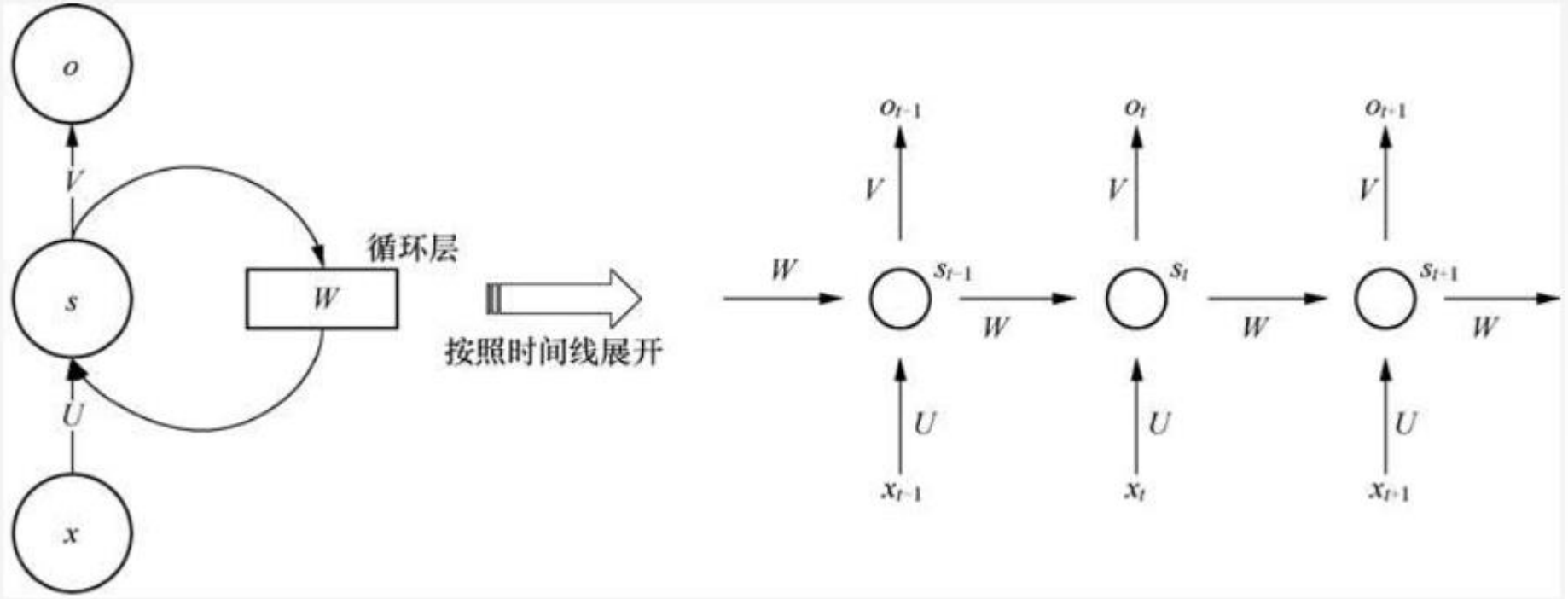
NLP的深度学习模型和方法

★ 递归神经网络

想要机器人听懂人类的语言，机器必须理解自然语言的上下文关系，才能准确理解某句话的含义，这就需要记忆能力。研究表明：反馈是实现记忆的核心手段。从原理上说，如果能够将当前网络的输出保存在一个记忆单元中，让这个记忆单元和下一时刻的输入一起进入下一个时序的神经网络，这是模仿人类记忆力机制的最朴素的思想。根据这个思想，科学家设计了最简单的记忆力模型，如图所示。将这个模型按照时间线展开，就可以得到一个递归表示的模型，如果其中的基本单元用神经网络构建，那就是递归神经网络。

$$O_t = g(V \cdot S_t)$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1})$$



NLP的深度学习模型和方法

★ 递归神经网络

RNN模型处理序列类型的数据具有天然优势，因为神经网络本身就是序列结构。但其也存在缺点，具体表现为：RNN可以很好地解决“短时依赖”问题，但对“长时记忆”却无能为力，这是简单版本RNN模型一直难以取得很好效果的根源。在深度学习诞生之前，研究者主要是针对具体问题人为挑选特定的参数来提升模型性能；但这样做不具有普适性，因为RNN无法决定挑选哪些参数。

NLP的深度学习模型和方法

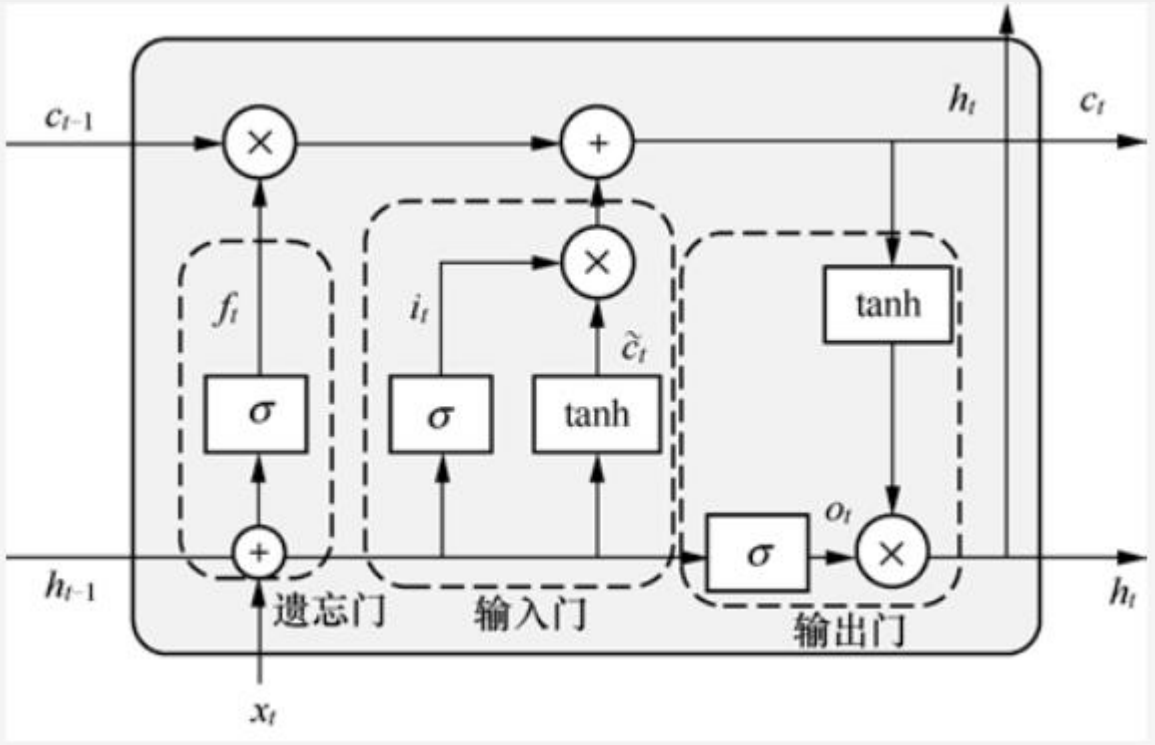
★ LSTM

为解决RNN的长时依赖问题，1997年人工智能领域著名学者Hochreiter Schmidhuber教授提出LSTM（long short term memory network）模型，字面意思是“长短时记忆网络”，本质上解决的仍然是短时记忆问题，只不过这种短时记忆比较长，能在一定程度上解决长时依赖问题。该模型一开始提出并没有引起学术界的重视，后来与深度学习思想相结合获得了新生，成为当下最主流的RNN模型之一，并且在NLP领域取得了突破性进展。

与简单RNN模型最大的不同在于，LSTM模型使用输入门、遗忘门和输出门控制网络，以实现长时记忆功能。研究表明，人类的记忆力机制同时具备记忆和遗忘两项能力，两者互为补充、缺一不可。人类的记忆力机制就是两者的平衡。如果一个模型只有记忆能力而没有遗忘能力，那么短时记忆将占据主导地位，长时记忆就无从谈起。LSTM模型成功地将记忆和遗忘两种能力融合于一个模型中，从而实现了长的短时记忆机制。

★ LSTM

LSTM的基本单元如图所示。其中，c_t表示时刻网络中的长时记忆（long term memory），h_t表示t时刻网络中的短时记忆（short term memory），网络具体要保留多少记忆是由前一时刻的输出和这一时刻的输入共同决定的。f_t是衰减系数。



$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$
$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$
$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t$$

由式可知，LSTM本质上是一个滤波器，是一个有关记忆的滤波器。经过滤波后，LSTM单元的输出可以表示为：

$$h_t = o_t \times \tanh(c_t)$$
$$y_t = \sigma(W_o \cdot h_t)$$

NLP的深度学习模型和方法

★ Word2Vec

在NLP问题中，如何生成紧凑、高效的词向量是一个核心问题，优秀的词向量表示形式对于自然语言的处理并行化、语义特征的层次化表达、语义理解的长时记忆具有重要意义。早期的词向量表示采用one-hot编码，其优点是简洁明了，其缺点是词向量过于稀疏，容易造成维度灾难。为了获取更紧凑、高效的词向量，学者们提出了词嵌入(word embedding)的概念,即将高维词向量嵌入一个低维空间。词嵌入有助于克服one-hot编码的缺陷，获得语义的分布式表达(distributed representation）。它的思路是：通过训练将每个词都映射到一个较短的词向量上。所有的这些词向量就构成了向量空间，进而可以用统计学的方法研究词与词之间的关系。词嵌入的方法有很多，但其基本思想是一致的：任何一个词的语义都跟它的上下文高度相关，任一词的含义可以用它的周边词表示。传统的词向量生成包括两大流派：基于语言模型的方法和基于统计的方法，分别对应之前所述的第一代和第二代NLP方法，在此不展开论述。

NLP的深度学习模型和方法

★ Word2Vec

第三代NLP方法中采用的词向量生成方法Word2Vec，顾名思义，就是“把词表示为向量”。该模型发布于2013年，在任何当代NLP任务中，它都是值得尝试的词向量生成首选项。虽然当前NLP的新模型层出不穷，但Word2Vec模型仍然是基础中的基础。Word2Vec的思想非常简单：词向量可以基于上下文词汇通过神经网络训练生成。例如，考虑如下语句：

```
Corpus = {I like Chinese green tea}
```

根据one-hot编码规则：

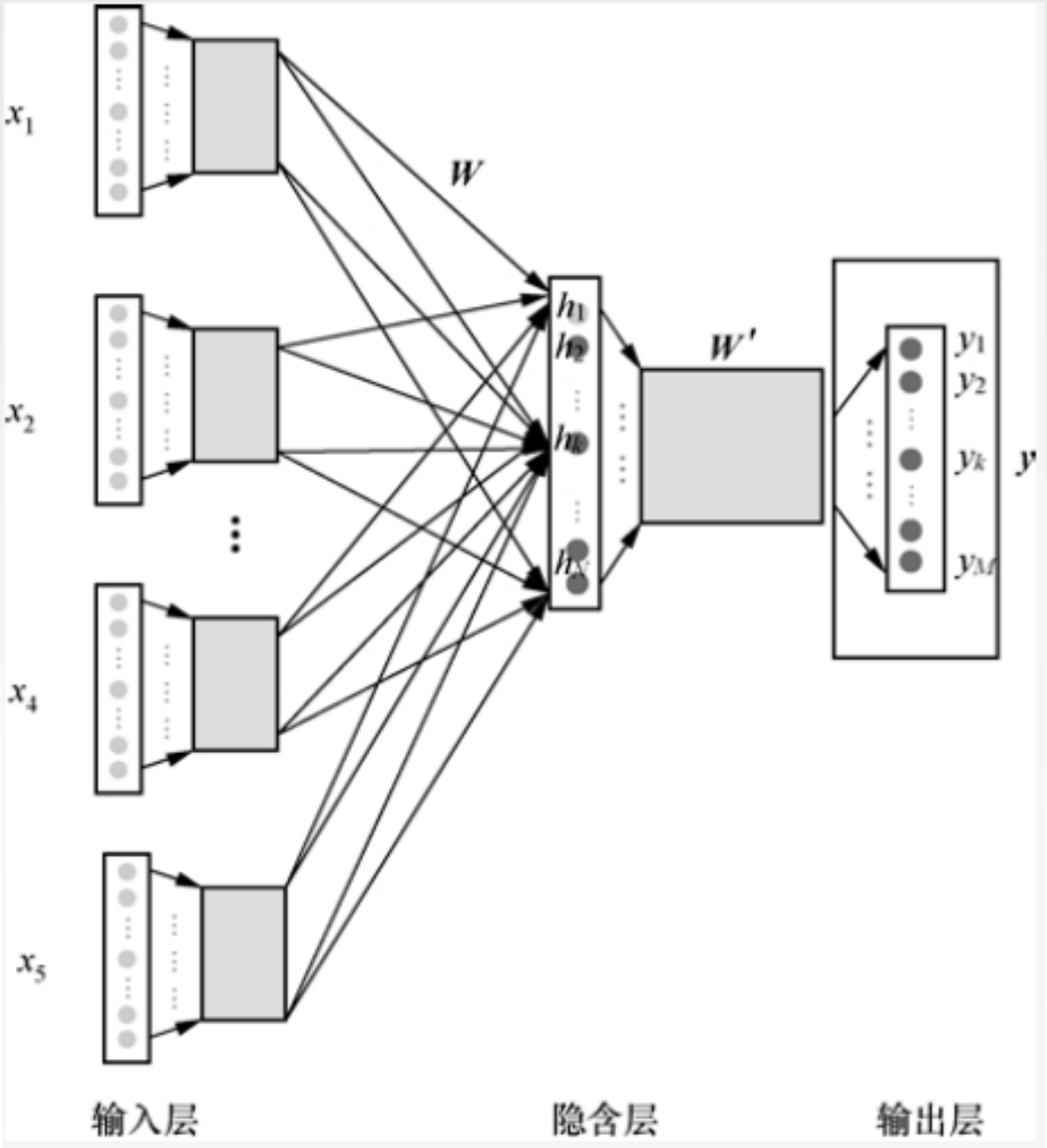
```
I:       X1 = [1 0 0 0 0]
Like:    X2 = [0 1 0 0 0]
Chinese: X3 = [0 0 1 0 0]
Green:   X4 = [0 0 0 1 0]
Tea:     X5 = [0 0 0 0 1]
```

NLP的深度学习模型和方法

★ Word2Vec

考虑如图所示的Word2Vec模型，现在欲得到Chinese这个词的嵌入式词向量，假设window size为2，就可以选取Chinese前面两个单词和后面两个单词的one-hot编码为神经网络的输入，以Chinese对应的one-hot编码为期望输出。

训练过程如下：采用梯度下降法训练神经网络并更新权重矩阵W，如果隐含层向量经过Softmax后的概率分布与目标单词的one-hot编码一致，那么对目标单词的训练就结束了。遍历整个词典，重复上述过程使得网络训练收敛，那么词典中任何一个单词的one-hot编码乘以矩阵W都将得到自己的word embedding。



NLP的深度学习模型和方法

★ Word2Vec

基于上述思想，学术界提出Word2Vec模型的两类具体实现，分别是CBOW（continuous bag-of-words）与skip-Gram模型。CBOW模型的训练输入是某一个特征词的上下文相关词对应的词向量，而输出就是这个特定的词的词向量。skip-gram模型和CBOW模型的思路正好相反。

CBOW是用周围词预测中心词，训练过程中其实是在从output的loss学习周围词的信息也就是embedding，但是在中间层是average的，一共预测V次。

skip-Gram是用中心词预测周围词，对每一个中心词都有K个词作为output，对一个词的预测有K次，所以能够更有效的从context中学习信息，共预测K*V次，因此，skip-gram的训练时间更长。

NLP的深度学习模型和方法

★ ELMO

在Word2Vec模型和2014年提出的GloVe模型中，每个词对应一个向量，但是它们对于多义词无能为力。2018年3月，ELMO（embedding from language models）给出一个较好的解决方案。不同于前述模型的一个词对应一个向量，ELMO不再给出固定的向量对应关系，而是给出一个预先训练好的模型。使用时，将一句话或一段话输入模型，模型会根据上下文推断每个词对应的词向量。因此，ELMO就可以结合前后语境对多义词进行理解。

ELMO具有两个优势：

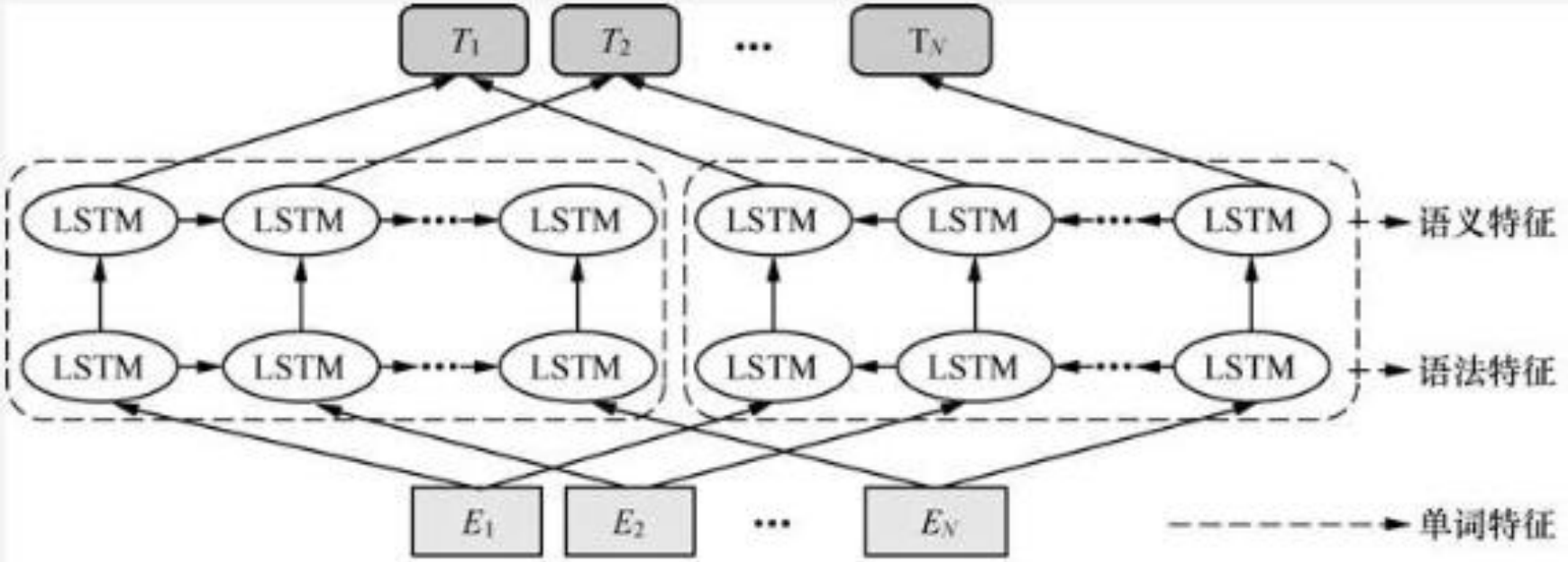
①能够学习到词汇用法的复杂性，即语法；

②能够学习到不同上下文语境中的词汇多义性，即语义。

NLP的深度学习模型和方法

★ ELMO

ELMO的核心思想是：每个词语的特征表达都是整个输入语句的函数。基于大量文本，词向量可以从深度双向语言模型(bidirectional language models, BiLM)中的内部状态学习而来。BiLM架构如图所示，采用的是经典的双层双向LSTM。图中，左端的前向双层LSTM是正向编码器，右端的逆向双层LSTM代表反向编码器，每个编码器都由两层LSTM叠加而成。类似的BiLM结构其实在NLP研究中很常见，只不过没有用双层LSTM编码器。



NLP的深度学习模型和方法

★ ELMO

语言模型训练的目标任务是根据单词T_i的上下文正确预测单词T_i。T_i之前的单词序列称为上文，T_i之后的单词序列称为下文。左端编码器的输入顺序是从左到右，不包含T_i的上文，右端编码器的输入是从右到左逆序的，不包含T_i的下文。

ELMO采用了NLP应用典型的两个阶段过程。第一个阶段是利用BiLM预训练；第二个阶段是在做下游任务时，利用预训练网络提取对应单词的各层word embedding作为新特征加入下游任务中。

ELMO充分利用了深度学习在分布式特征表达方面的优势。它的BiLM结构是“深”的，好处是能够提取到丰富的多层次特征，这是ELMO的核心思想。高层的LSTM可以捕捉词语意义中和语境相关的特征，而低层的LSTM可以找到语法方面的特征。它们结合在一起，在NLP任务中就会体现优势。

NLP的深度学习模型和方法

★ ELMO

使用BiLM结构利用大量语料就能训练好这个双向RNN网络，在预训练好的网络中输入一个新句子，句子中每个单词都能得到对应的3个embedding:

(1) 最底层是单词特征，即单词的word embedding，与Word2Vec模型类似;

(2)上一层提取的是语法特征，可以用于词性标注等任务;

(3)再上一层提取的是语义特征，可以用来做语义消歧的高阶任务。

ELMO模型也有自身的缺点，表现为:

(1)虽然采用深度神经网络作为特征抽取器，但是双层LSTM相比结构更优越的Transformer模型，其特征提取能力较弱。如果ELMO采用Transformer作为特征提取器，相信可以取得更好的性能。

(2)采取双向LSTM网络，这种BiLM结构拼接3个embedding,这种经典NLP结构虽然实现了多层特征的融合，但相比后续提出的一体化特征融合方式，网络结构和算法都比较复杂，限制其进一步发展。

NLP的深度学习模型和方法

★ Transformer

Transformer模型是由谷歌在2017年6月发表的论文Attention Is All Your Neea中提出的。这是一种称为seq2seq的模型，在机器翻译等应用中使用广泛。传统的seq2seq模型通常采用RNN,一般在网络结构中会用到Encoder和Decoder，要想提升效果，可以通过注意力(Attention）机制连接Encoder和Decoder。研究表明，如果使用RNN作为Encoder和Decoder，则存在两个问题：一是RNN的递归依赖难以并行化，早期版本的谷歌翻译系统（Google’s neural machine translation system,GNMT）需要96块GPU并行训练一周，而RNN无法提供这方面的支持；二是缺乏对全局语义信息的理解，尤其是在长时记忆、层级化语义表达两方面捉襟见肘。

Transformer模型摈弃了RNN，提出一种全新的并且更简单的网络结构，只需要Attention机制就能解决seq2seq的问题，并且能够一步到位获取全局语义信息。Transformer在机器翻译任务上的表现超过了RNN、CNN,其最大优点是可以高效地并行化。

NLP的深度学习模型和方法

★ Transformer

Transformer的核心是Attention机制。

(1)在编码当前词时，充分考虑上下文的信息。相比ELMO,Attention机制的独到之处是为不同的上下文分配不同的权重。例如，The bird didn’t fly because it was hurt by the cat,如果采用RNN或者LSTM作为编解码器，就是平等对待上下文的，因此不容易理解“it"是指代bird;而Attention机制会给bird分配较高的权重，这样就可以模拟人脑的Attention机制，从而准确地识别出“it”的含义。

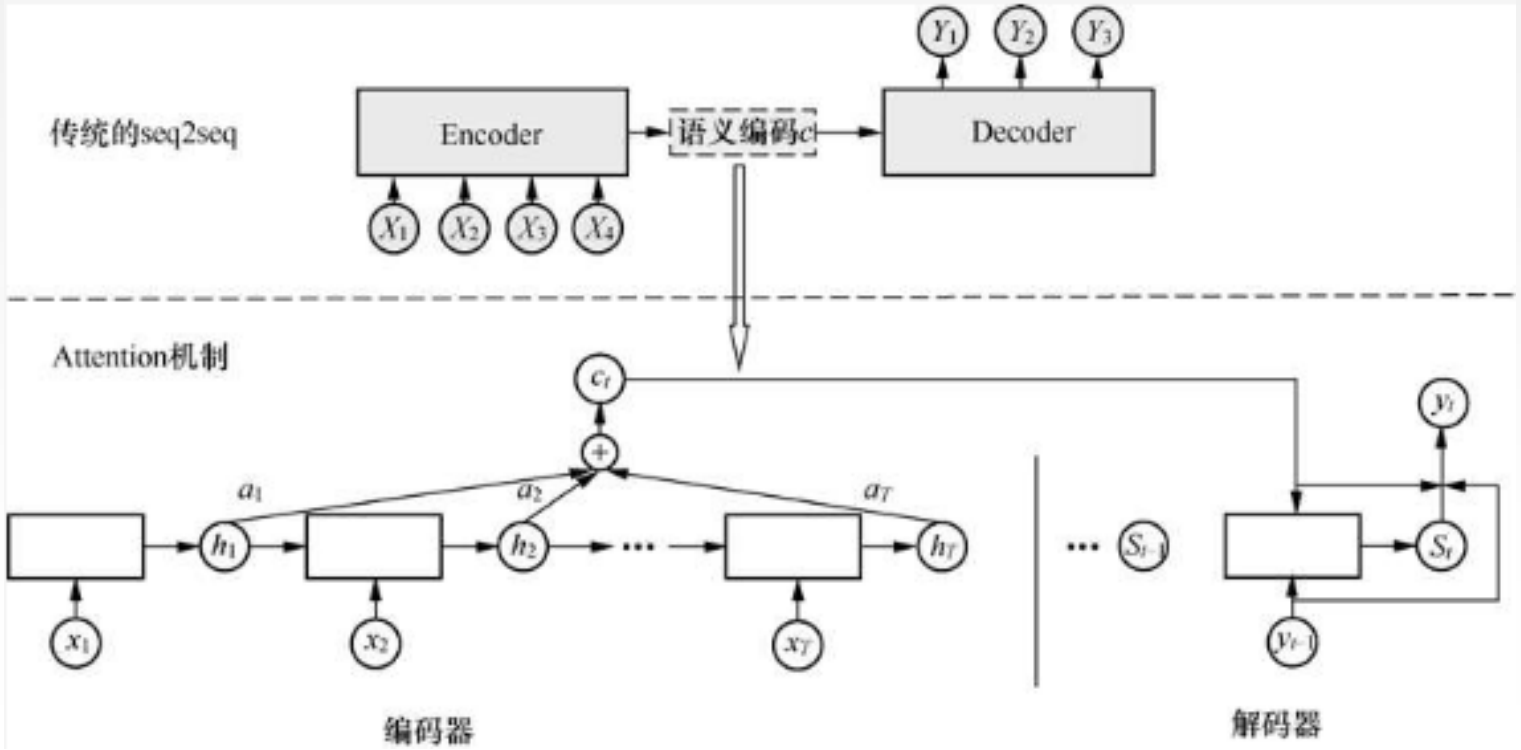
(2)在具体实现时，Transformer又加入了Self-Attention和Multi-Head Attention,通过多组权重参数优化上下文对当前词的影响，进一步提升了语义理解能力。

(3)除了在Encoder和Decoder加入Attention机制外，训练过程中，Decoder在每个时间步中还有单独一个Attention是从Encoder输入的，帮助当前词获取当前需要关注的重点内容。

NLP的深度学习模型和方法

★ Transformer

Attention机制的原理如图所示。传统的Seq2seq结构中，输入编码为一个定长语义编码，然后通过这个编码再生成对应的输出序列。针对这个问题，Bengio率先提出Attention机制，并因此获得2019年的图灵奖。区别在于，Encoder的输出不是一个语义向量，而是一个语义向量的序列，在解码阶段会有选择地从向量序列中选择一个子集，至于这个子集怎么选取，子集元素占比多少，这些都是Attention机制要解决的问题。



NLP的深度学习模型和方法

★ Transformer

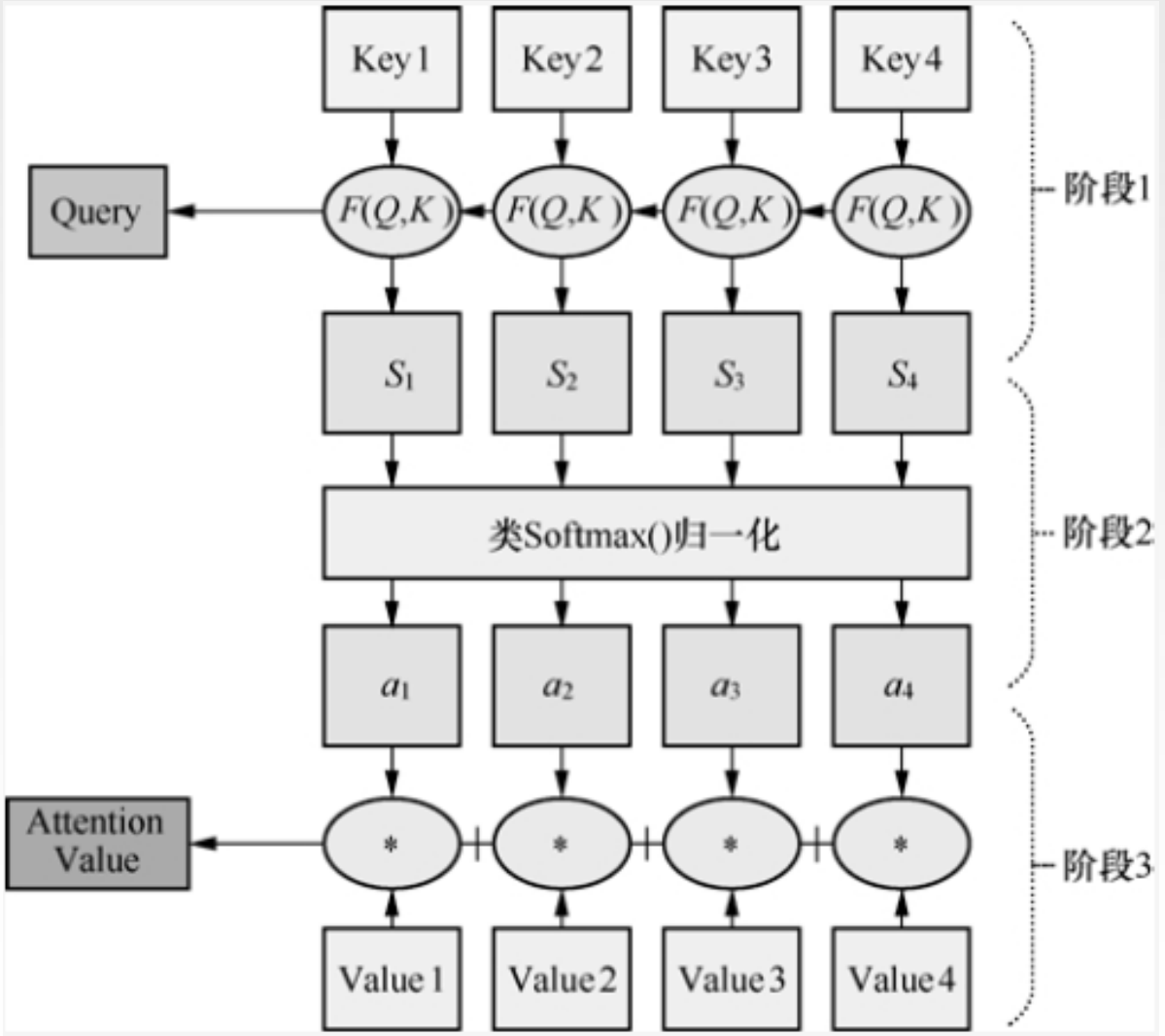
Attention机制本质上可以被描述为一个查询（Query）到一序列对（键Key/值Value）的映射过程。在计算Attention时，主要分为三步，如图所示。

第一步，将Query和每个Key进行相似度计算，得到权重，常用的相似度函数有点积、拼接、感知机等。

第二步，使用一个Softmax()函数对这些权重进行归一化。

第三步，将权重和相应的键值Value进行加权求和，得到最后的Attention。

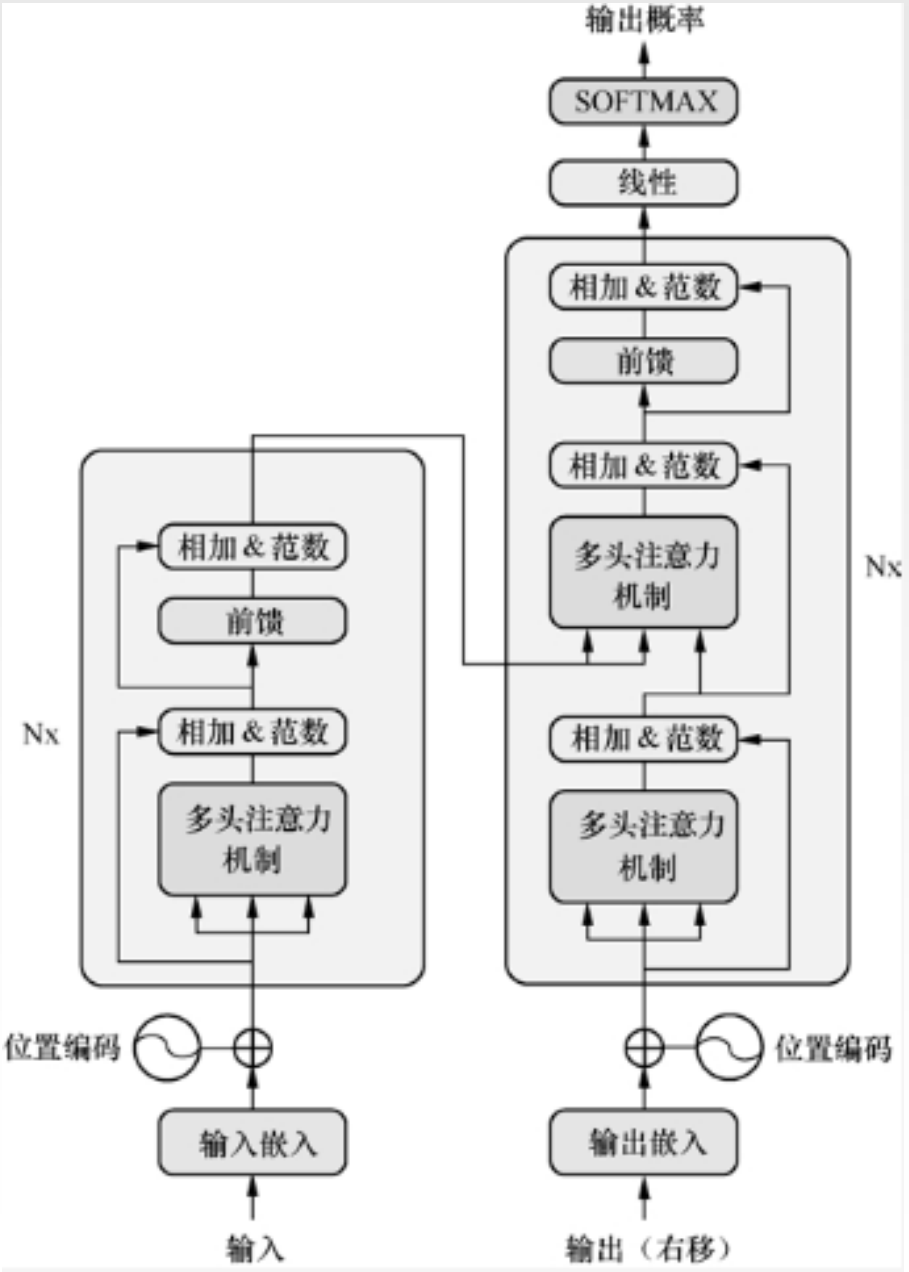
目前，在NLP研究中，Key和Value常常是同一个，即Key=Value。



NLP的深度学习模型和方法

★ Transformer

一个典型的Transformer模型的结构如图所示。左边的结构代表编码器，采用了N=6的重复结构，包含一个Multi-Head Attention和一个Position-wise feed-forward（一次线性变换后用ReLU激活，然后再线性变换）。右边的结构代表解码器，最下面是输出序列的tokens,在翻译任务中就是目标语言的词表，并且第一个Multi-Head Attention是带有Mask的，以消除右侧单词对当前单词Attention的影响，左边的Encoder编码后的输出将会插入右边Decoder的每一层，即Key和Value。



NLP的深度学习模型和方法

★ Transformer

Transformer相比RNN、LSTM等传统递归模型具有如下优点。

(1) 完全的并行计算。Transformer的Attention和feed-forward均可以并行计算，而LSTM则依赖上一时刻，必须串行。

(2)减少对长时记忆的依赖。利用self-attention将每个字之间的距离缩短为1，大大缓解了长距离依赖问题。

(3)提高网络深度。由于大大缓解了长距离依赖梯度衰减问题，Transformer网络可以很深，基于Transformer的网络可以做到20多层，而LSTM一般只有2~4层；根据深度学习的基本思想，网络越深，高阶特征提取能力越强，模型性能越好。

(4)真正的双向网络。Transformer可以同时融合前后位置的信息，而双向LSTM只是简单地将两个方向的结果相加，严格来说，双向LSTM仍然是单向的。

(5) 可解释性强。完全基于Attention的Transformer，可以表达字与字之间的相关关系,可解释性更强。

NLP的深度学习模型和方法

★ BERT

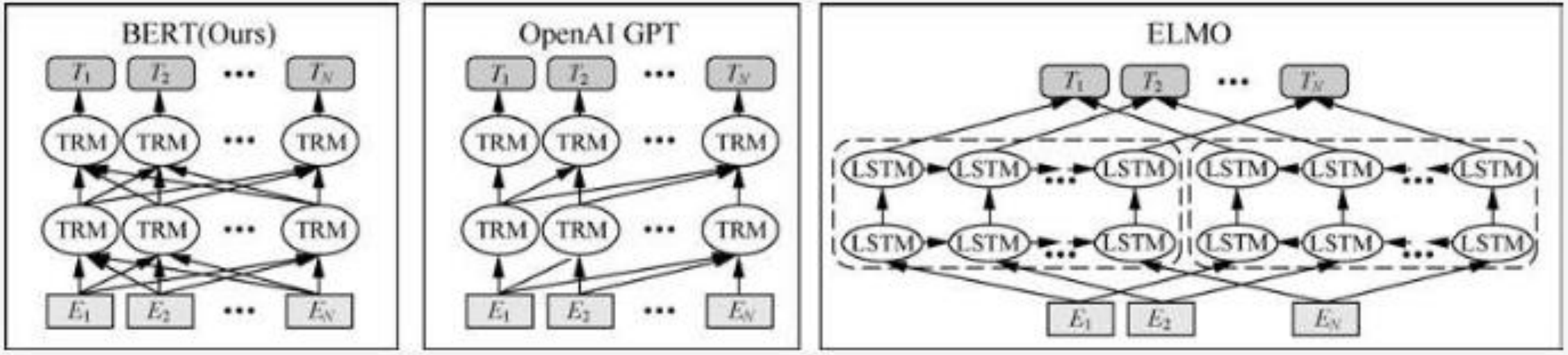
BERT是谷歌在2018年10月推出的深度语言表示模型，一经推出便席卷整个NLP领域，为NLP应用带来里程碑式的进步，是NLP领域SOTA(state-of-the-art)最重展。

BERT的全称是 bidirectional encoder representation fromtransformers，从名称上就可以看出BERT是近期NLP创新的集大成者：一是采用了Transformer作为基础模型；二是引入了ELMO的BiLM架构。客观地说，BERT在模型结构上的创新有限，主要创新点都在pre-training方法上。然而，有效果才是硬道理，BERT在NLP领域的成绩是有目共睹的，所有的赞誉都是应得的。在机器阅读理解顶级水平测试SquAD1.1中，BERT获得惊人成绩：在全部两个衡量指标上全面超越人类，并且在11种不同NLP测试中，BERT创造最佳成绩，包括将GLUE基准推至80.4%（绝对改进率7.6%），MultiNLI准确度达到86.7%（绝对改进率5.6%）等。

NLP的深度学习模型和方法

★ BERT

BERT模型的结构如图所示。对比OpenAI的GPT模型（当前NLP领域另一网红模型），BERT的语义表达是双向的，类似单向LSTM与双向LSTM的区别。BERT模型和ELMO模型都是“双向”的，但目标函数不同。ELMO分别以$P(w_i|w_1, w_2, \cdots w_{i-1})$和$P(w_i|w_{i+1}, w_{i+2} \cdots w_n)$作为目标函数，独立训练出两个representation然后拼接，而BERT是以$P(w_i|w_1, w_2. \cdots w_{i-1}, w_{i+1}, w_{i+2}, , w_n)$作为目标函数训练语言模型。BERT模型的结构及其与GPT、ELMO模型的区别如下图：



NLP的深度学习模型和方法

★ BERT

创新点1： masked language model

传统的语言模型学习语言特征都是以前文预测下一个词为训练目标，然而这个思路在双向模型中不可行，例如，在BiLM中反向encoding意味着“正向要预测的下一个词已知”，这显然是自相矛盾的。现有的语言模型（例如ELMO)号称是双向的，但是实际上是两个单向RNN语言模型拼接而成的。受A Neural Probabilistic Language Model论文的启发，BERT提出了masked language model，随机去掉句子中的部分token，然后利用模型预测被去掉的token是什么，其基本思想与Word2Vec类似。具体实行时，将语料库中15%的语料用[Mask] token代替，并通过Pre-training预测masked token，将masked token这一层对应输出的向量送入Softmax就能得到较理想的结果。

★ BERT

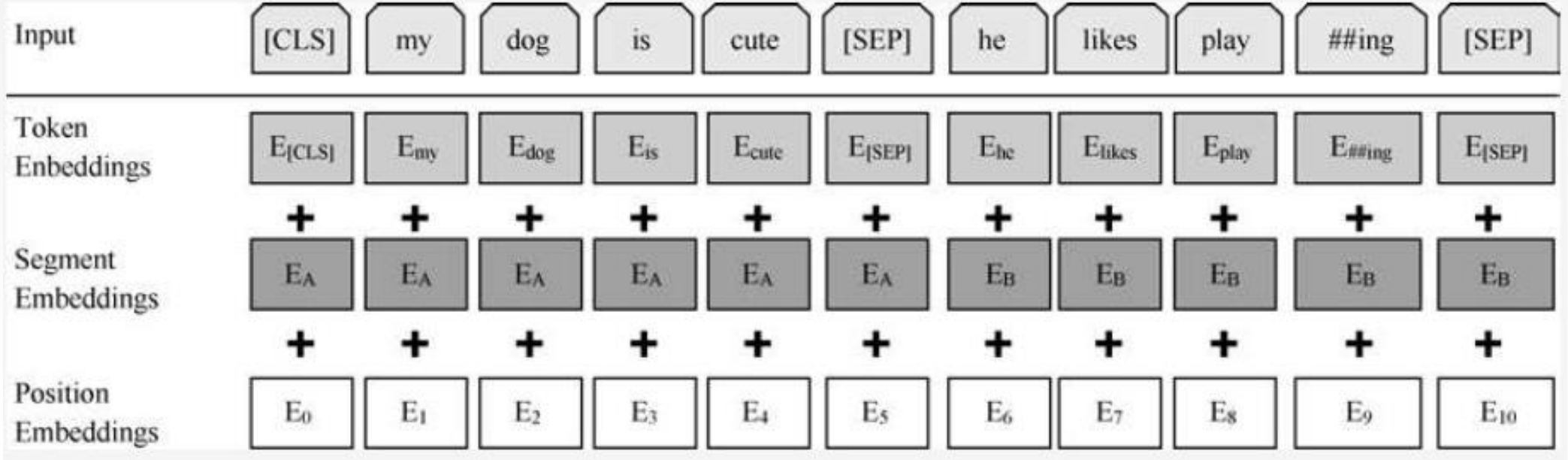
创新点2：next sentence prediction

很多语言任务都需要获取句子级别关系的representation，因此只有语言模型是不够的，还需要捕捉句子级的特征。所以BERT设计了一个“句子对”任务，该任务的训练语料是两句话，目标是预测第二句话是否是第一句话的下一句。例如，选择句子A和句子B为预训练样本，B有50%的可能性是A的下一句，也有50%的可能性是来自语料库中的随机句子。

★ BERT

创新点3：层次化embedding

与BiLM模型类似，BERT模型同样可以提供层次化的特征，如图所示，模型的输入是以下3个embedding向量的和。(1)token embedding:当embedding。(2) segment embedding：当前词所在句子的index embedding,是由BERT模型训练得到的。(3)position embedding：当前词所在位置的index embedding,是由BERT模型训练得到的。

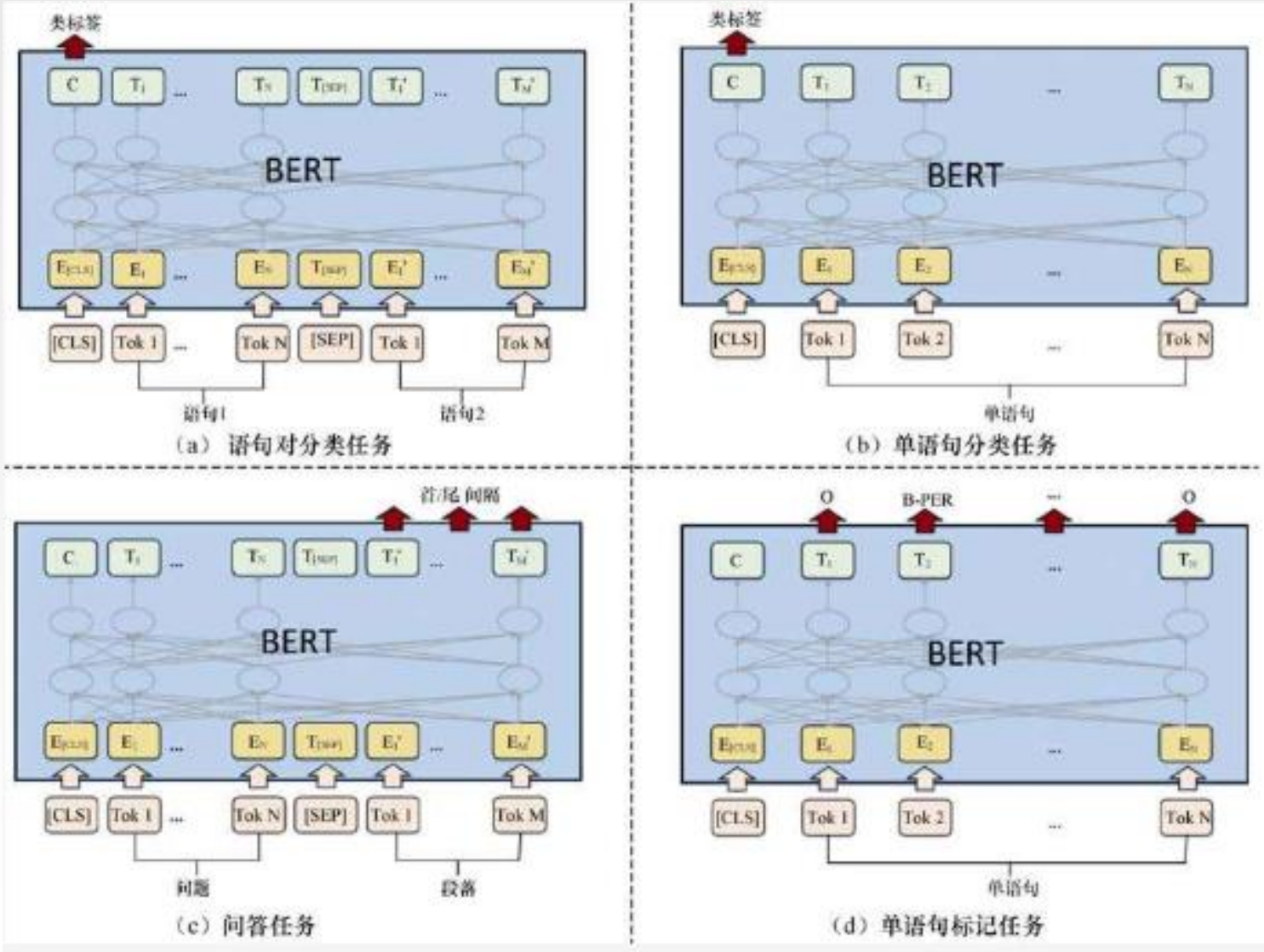


BERT的层次化特征表示

NLP的深度学习模型和方法

★ BERT

BERT采用了“Pre-training+Fine-tuning”的机制，一般只需要增加一层网络，如图所示。例如，在文本匹配任务和文本分类任务中，只需要在对应的representation处（即encoder在[CLS]词位的顶层输出)加上一层神经网络；对于阅读理解问题SQuAD,在原文span抽取上直接用两个线性分类器输出span的起点和终点；在序列标注任务中只需要增加softmax输出层。



BERT对下游任务的Fine-tuning

NLP的深度学习模型和方法

★ BERT

近十年来，学者一直致力于解决NLP领域的两个关键问题：一是构建一个真正的“深度”模型，特别是像CV领域著名的ResNet这样的深度网络；二是实现语言和语义的无监督学习。BERT在这两方面都得到了突破，其里程碑意义在于：证明了一个非常深的模型可以显著提升NLP任务的准确率，而且这个模型可以利用大量无标签数据集预训练得到。更进一步地，BERT模型给我们以下启发。

第一，BERT模型非常深(12层)但并不宽，中间层只有1024个单元，而Transformer模型的中间层有2048个单元。因此，从模型结构上来说，BERT与ResNet很像，可以类比为RNN中的ResNet，两者各自代表CV和NLP领域的SOTA水平，这似乎在印证一个朴素的观点：在深度学习领域，深而窄的模型比浅而宽的模型性能更好。

第二，我们已经知道，深度学习就是表征学习。无论是图像领域还是语音领域，如果一个模型能够通过“深”的网络分层次地表达特征，那么辅以“超大规模的数据+超强算力支持”就一定能获得强大的表征学习能力，取得SOTA的性能水平。

机器人语音技术AIUI开放平台

★ 机器人语音技术概述

科大讯飞在语音领域深耕多年，拥有声学处理、语音识别、语音合成、语音评测等核心技术。AIUI是科大讯飞提供的一套人机智能交互解决方案，从2015年发布至今，基于核心技术不断打磨效果，逐步成熟，是一套功能完善、易于接入的人机交互解决方案。

AIUI旨在实现人机交互无障碍，使人与机器可以通过语音、图像、手势等自然交互方式，进行持续、双向、自然的沟通。现阶段AIUI提供以语音交互为核心的交互解决方案，全链路聚合了语音唤醒、语音识别、语义理解、内容（信源）平台、语音合成等模块；可以应用于智能手机（终端）、机器人、音箱、车载、智能家居、智能客服等多个领域，让产品不仅能听会说，而且能理解会思考。

机器人语音技术AIUI开放平台

★ 机器人语音技术概述

AIUI开放平台主要包含语义技能(Skill)、问答库(Q&A)编辑及AIUI应用（硬件）云端配置的能力，并为不同形态产品提供了不同的接入方式，主要包括Android、iOS、Windows、Linux SDK、基于HTTP的WebAPI，以及软硬件一体的AIUI评估板（量产板）、讯飞魔飞智能麦克风。

AIUI将科大讯飞强大的单点交互能力（前端声学处理、语义理解、语音合成、丰富的内容信源)整合为全链路的交互方案提供给广大开发者，开发者可以根据实际的业务需求，利用热词、静态实体、动态实体、所见即可说等特性，进行个性化的优化和改进，提升交互准确率，让人机交互更加流畅，真正地满足和解决用户实际使用中遇到的问题。

机器人语音技术AIUI开放平台

★ 应用领域

AIUI解决方案可以应用于多个领域与产品，包括但不限于智能手机（终端）、服务型机器人、玩具机器人、音箱、玩具、手表、车载、智能家居、智能客服、医疗导诊。

在智能手机、手表或PC等终端中，AIUI可以与手机深度结合为全局的智能语音控制系统。在单个应用(App)中，可以帮助用户用语音完成复杂的交互，如导航/买票/订餐等。

在机器人、音箱、玩具、车载等产品中，AIUI可以化身个人智能助理或虚拟人物，执行用户的指令，如控制设备移动，多媒体的播放，天气、股票信息查询等能力。

当任意智能家居搭载AIUI后，开发者通过将AIUI的语义结果解析成对应的控制指令，不仅可以完成设备自身的状态控制，甚至可以化身为整个家庭的中控设备。

在智能客服领域，开发者可以利用AIUI的自定义问答和自定义技能能力，完成对用户表述的语义理解，极大地降低企业的人工成本。

AIUI还可应用于KTV场景下的点歌、播放控制，商超、政务、银行等场景下的大屏语音互动等领域，开发者可以在产品开发中释放AIUI的无限潜能。

机器人语音技术AIUI开放平台

★ 产品框架

AIUI语义信息透明开放，可云端接入，支持业务自由定制。AIUI的核心技术包括语音唤醒、语音识别、自然语言理解、语音合成、全双工交互及翻译，其产品架构如图所示。 AIUI开放平台希望给从事各个领域的开发者带来更多的可能性，

在支持自定义语义的基础上，平台未来将会支持开发者在技能商城上传自己的语义资源，如技能、问答库等，可供其他开发者使用。综上所述，AIUI应用的领域广泛，涵盖生活的各行各业，可以为广大开发者提供强有力的技术支持和引导，实现生态共享和互利共赢。