

深度学习在自然语言处理中的应用

邱锡鹏

复旦大学

2017年12月8日

<http://nlp.fudan.edu.cn/xpqi>

QCon

全球软件开发大会

成为软件技术专家
的必经之路

[北京站] 2018

2018年4月20-22日 北京·国际会议中心

7折 购票中, 每张立减2040元
团购享受更多优惠



识别二维码了解更多



极客时间

重拾极客精神·提升技术认知

下载极客时间App

获取有声IT新闻、技术产品专栏，每日更新



扫一扫下载极客时间App

AiCon

全球人工智能与机器学习技术大会

助力人工智能落地

2018.1.13 - 1.14 北京国际会议中心



扫描关注大会官网



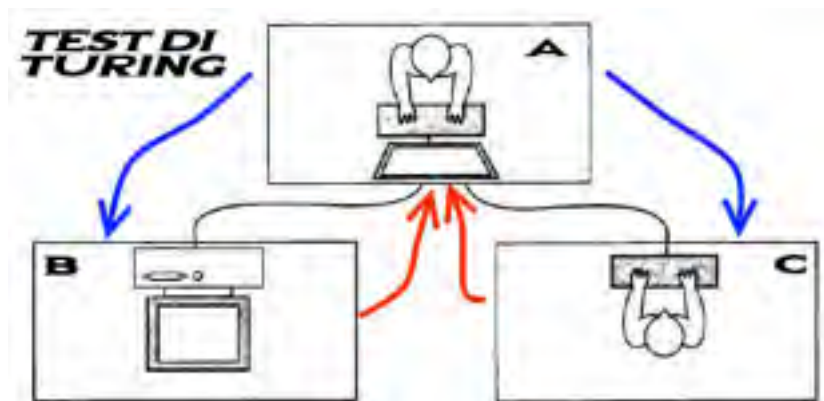
内容提纲

- ▶ 简介
 - ▶ 自然语言处理
 - ▶ 深度学习
- ▶ 语义表示学习
 - ▶ 词表示
 - ▶ 句子表示
- ▶ 自然语言处理的新范式
- ▶ 应用



自然语言处理

从人工智能开始



Alan Turing

自然语言处理：理解和生成



什么是自然语言？

- ▶ **语言**是指在一个有限的字符集上，产生的符合一定规则的字符串集合。
- ▶ **自然语言**通常是指一种自然地随文化演化的语言。
- ▶ 自然语言 VS 人工语言
 - ▶ 形式语言 (Chomsky, 1950)
 - ▶ 区别
 - ▶ 自然语言：歧义性
 - ▶ 人工语言：确定性





歧义：以中文分词为例

▶ 不同的语言环境中的同形异构现象，按照具体语言环境的语义进行切法。

▶ 交叉歧义

▶ 他／说／的／确实／在理

▶ 组合歧义

▶ 两个／人／一起／过去、个人／问题

▶ 从马／上／下来、马上／就／来

▶ 句子级歧义

▶ 白天鹅在水里游泳

▶ 该研究所获得的成果

} 伪歧义



自然语言处理

- ▶ 自然语言处理包括语音识别、自然语言理解、自然语言生成、人机交互以及所涉及的中间阶段。
- ▶ 是人工智能和计算机科学的子学科。

自然语言处理不等于研究语言学（计算语言学）、文学。

Every time I fire a linguist, the performance of our speech recognition system goes up.

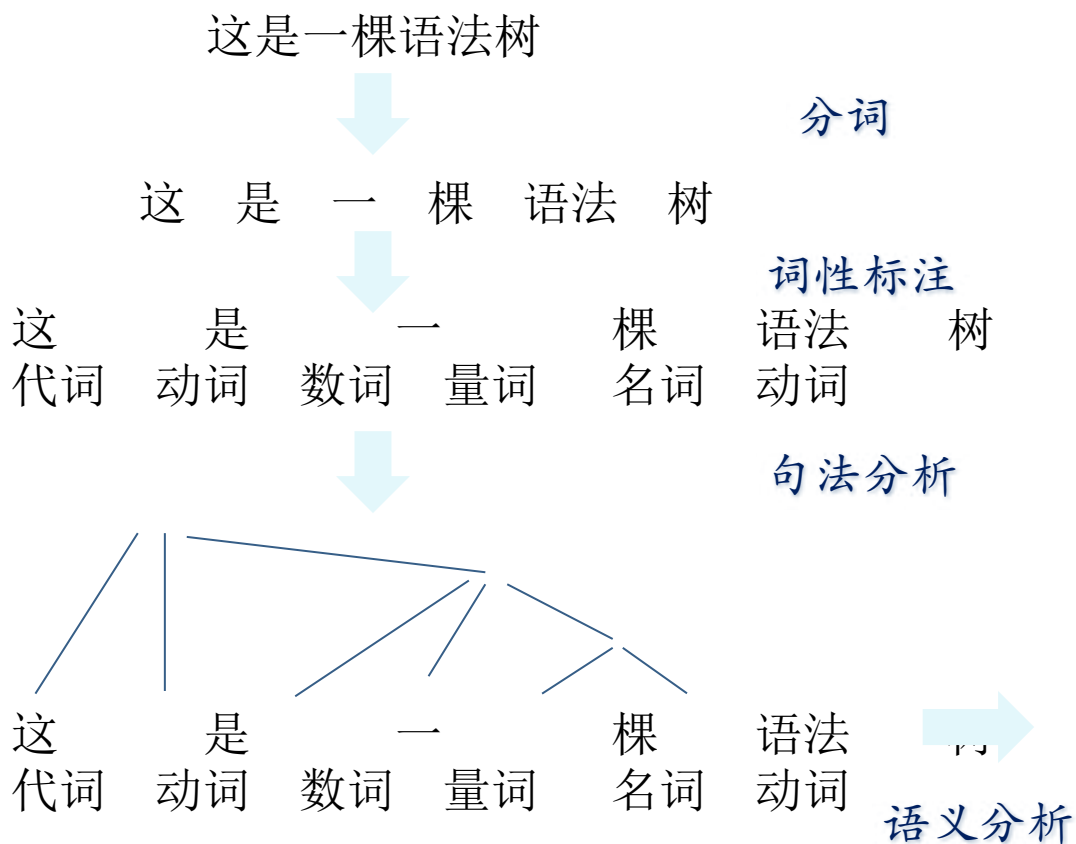
-- Frederick Jelinek, 1985

https://en.wikiquote.org/wiki/Fred_Jelinek





理想中的自然语言处理流程



应用

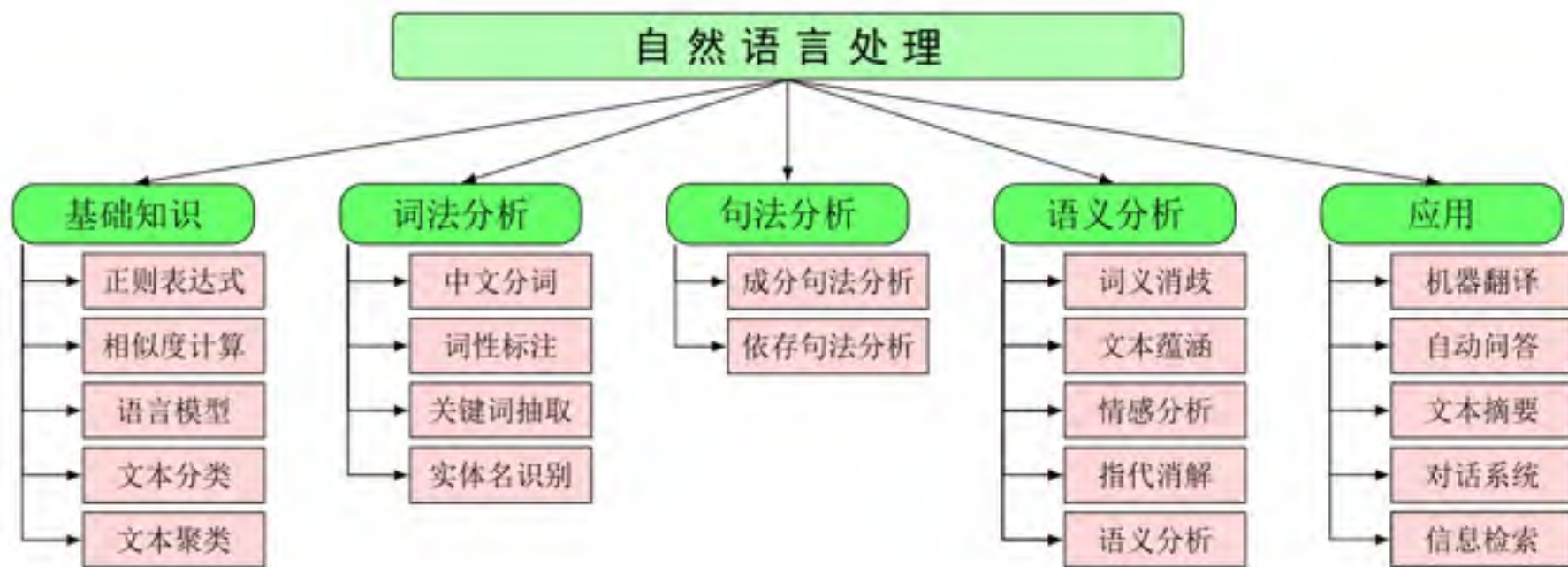
语义分析
机器翻译
自动问答
情感分析
... ..

{ 这, 是, 语法树 }

知识库

主要任务

- 自然语言处理任务可以分为四类：词法分析、句法分析、语义分析、应用。



日常生活的自然语言处理应用

► 搜索引擎



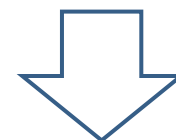
► 输入法



► 机器翻译



► 智能音箱



当出错时...

民主 ->  江泽民 主席
(注：早期google的搜索结果)



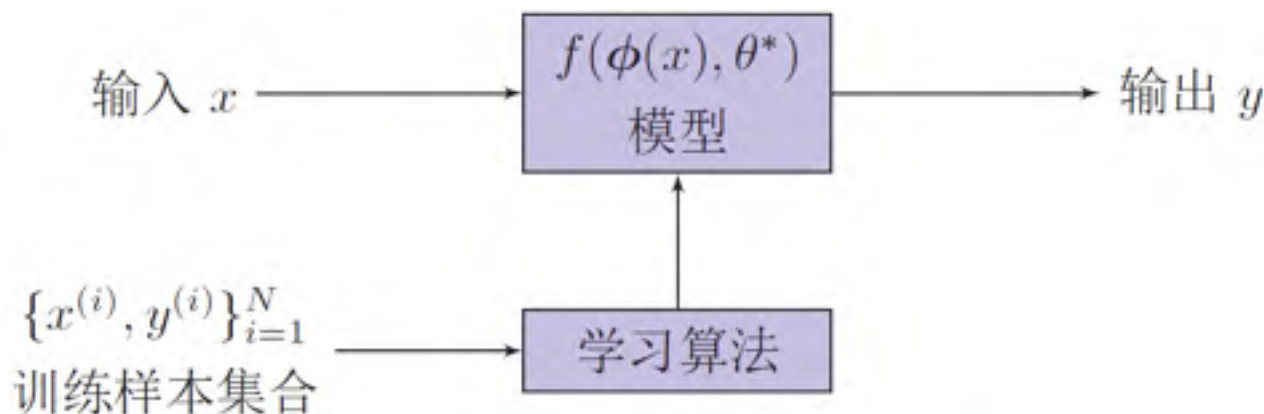


发展历程

- ▶ 1990年以前，基于**规则**（rule-based）的方法
 - ▶ 使用手写的规则
- ▶ 1990年以后，基于**语料库**（corpus-based）的方法
 - ▶ 也叫**实证**（empirical）方法或**数据驱动**（data-driven）方法
 - ▶ 大量使用统计或机器学习模型
 - ▶ 典型应用： *The mathematics of statistical machine translation: parameter estimation. 1993*
- ▶ 2011年以后，基于**神经网络**（neural-based）的方法
 - ▶ 端到端的神经网络模型
 - ▶ 典型应用： *Sequence to Sequence Learning with Neural Networks, 2014*

基于语料库的方法

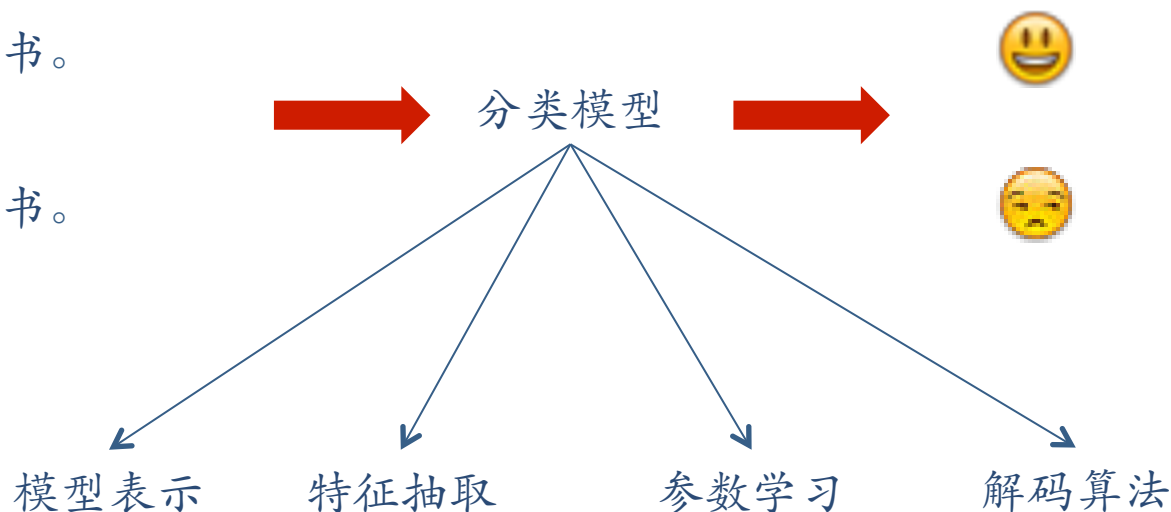
- ▶ 语料库: 文本数据的集合
- ▶ 技术手段:
 - ▶ 统计模型
 - ▶ 机器学习模型



实际的自然语言处理流程

我喜欢读书。

我讨厌读书。



情感分析

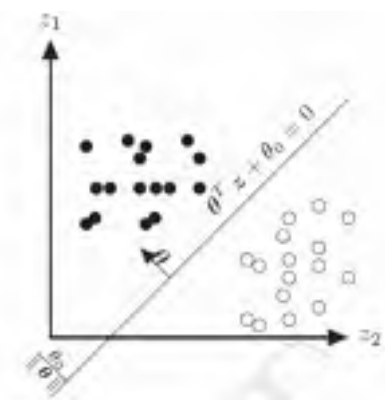
文本分类

根据文本内容来判断文本的相应类别

D_1 : “我喜欢读书”

D_2 : “我讨厌读书”

	我	喜欢	讨厌	读书
D_1	1	1	0	1
D_2	1	0	1	1



+

-



换个角度看中文分词

自	：	然	：	语	：	言	：	处	：	理
0		1		0		1		0		



窗口大小	样本 x	类别标签 y
2	“自：然”	0
	“然：语”	1
	“语：言”	0
4	“自然：语言”	1
	“然语：言处”	0
	“语言：处理”	1



单字符特征	$x_{-2}y_0, x_{-1}y_0, x_0y_0, x_1y_0, x_2y_0$ ^a
双字符特征	$x_{-1}x_0y_0, x_0x_1y_0, x_{-1}x_1y_0,$
三字符特征	$x_{-1}x_0x_1y_0$
马氏链特征	$y_{-1}y_0$

1/0

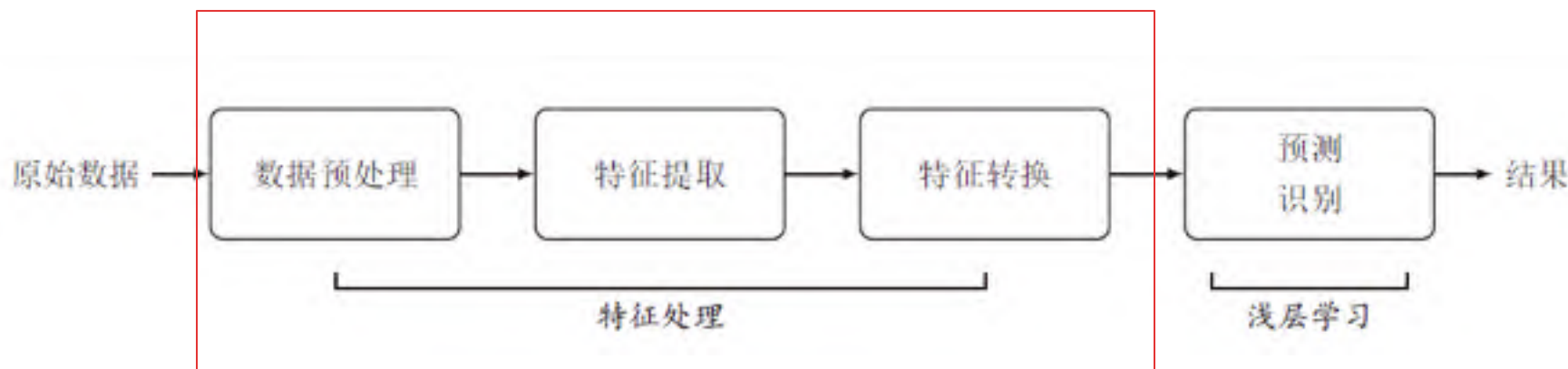
[000010001000100011001]



特征工程问题

▶ 在实际应用中，**特征**往往比分类器更重要

- ▶ 预处理：经过数据的预处理，如去除噪声等。比如在文本分类中，去除停用词等。
- ▶ 特征提取：从原始数据中提取一些有效的特征。比如在图像分类中，提取边缘、尺度不变特征变换特征等。
- ▶ 特征转换：对特征进行一定的加工，比如降维和升维。降维包括
 - ▶ 特征抽取 (Feature Extraction) : PCA、LDA
 - ▶ 特征选择 (Feature Selection) : 互信息、TF-IDF

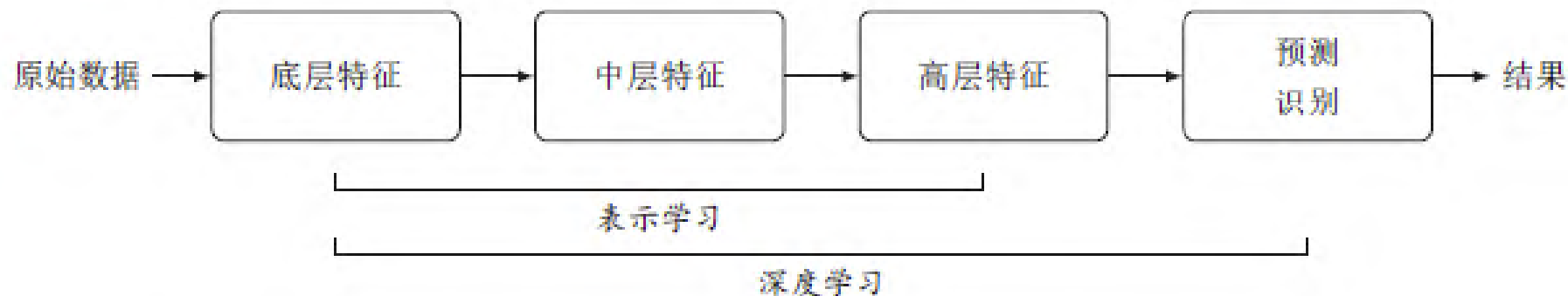




深度学习

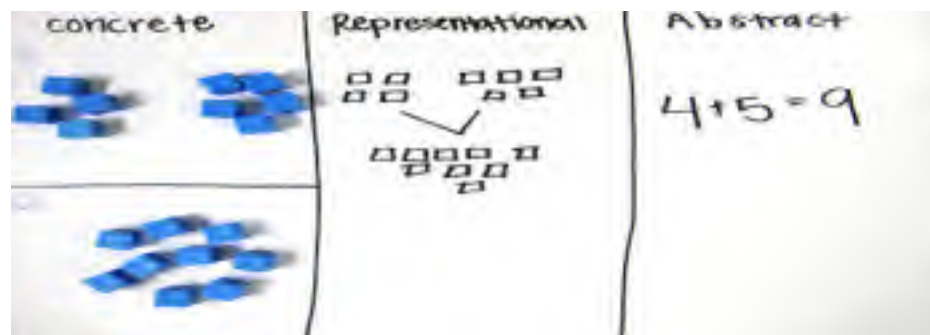
深度学习

- ▶ 深度学习 = 表示学习 + 浅层学习
- ▶ 难点：贡献度分配问题



表示学习与深度学习

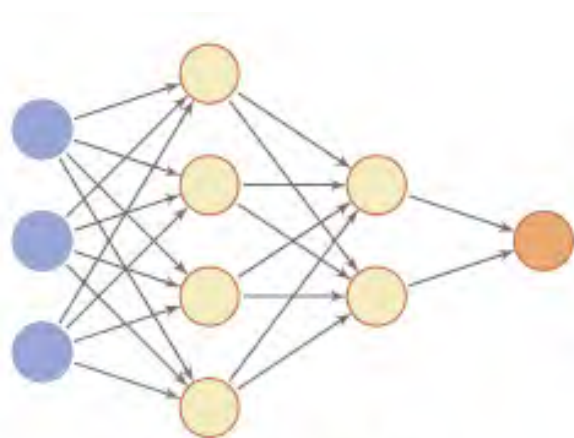
- ▶ 一个好的表示学习策略必须具备一定的深度
 - ▶ 特征重用
 - ▶ 指数级的表示能力
 - ▶ 抽象表示与不变性
 - ▶ 抽象表示需要多步的构造



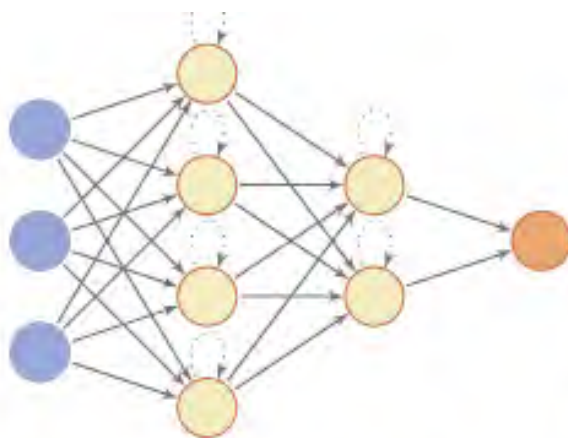
<https://mathteachingstrategies.wordpress.com/2008/11/24/concrete-and-abstract-representations-using-mathematical-tools/>

深度学习与神经网络

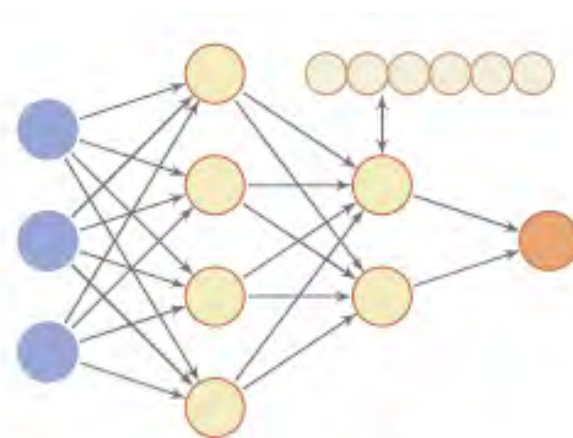
- 深度学习天然不是神经网络，但神经网络天然是深度学习！



(a) 前馈网络



(b) 反馈网络



(c) 记忆网络



语言表示学习

语义鸿沟

- ▶ 底层特征 VS 高层语义
 - ▶ 人们对文本、图像的理解无法从字符串或者图像的底层特征直接获得

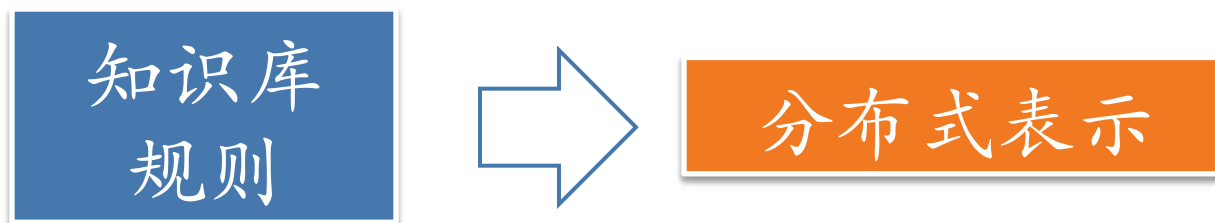
表示学习



床前明月光，
疑是地上霜。
举头望明月，
低头思故乡。

语言表示

► 如何在计算机中表示语言的语义?



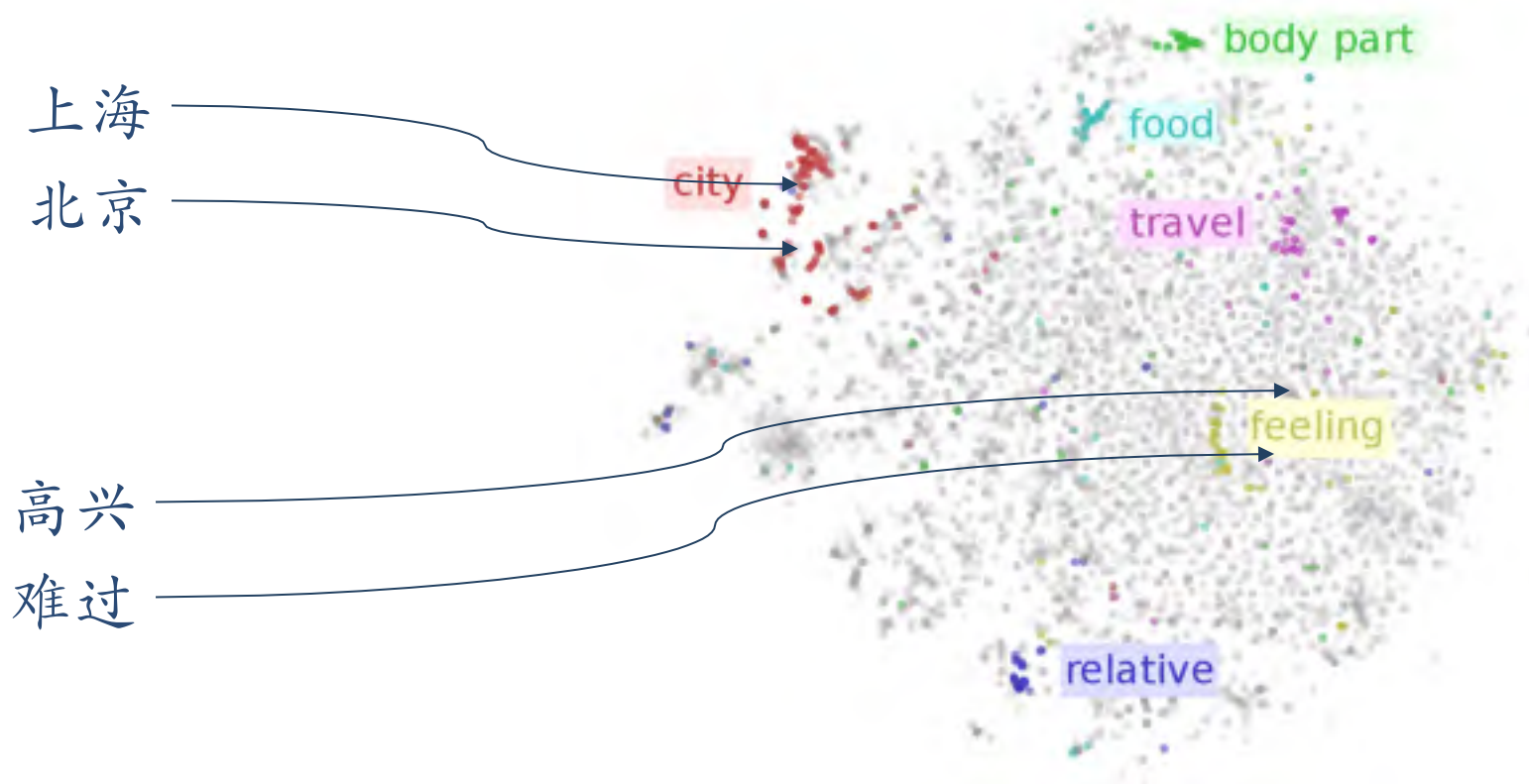
- 压缩、低维、稠密向量
- 用 $O(N)$ 个参数表示 $O(2^k)$ 区间
 - k 为非0参数, $k < N$

一个生活中的例子：颜色

命名	RGB值
红	[1,0,0]
绿	[0,1,0]
蓝	[0,0,1]
中国红	[0.67, 0.22, 0.12]
咖啡色	[0.64, 0.16, 0.16]



词嵌入 (Word Embeddings)



<https://indico.io/blog/visualizing-with-t-sne/>

分布式表示

--来自神经科学的证据



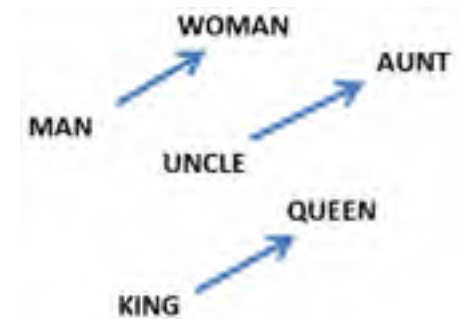
<http://www.nature.com/nature/journal/v532/n7600/full/nature17637.html>

词嵌入

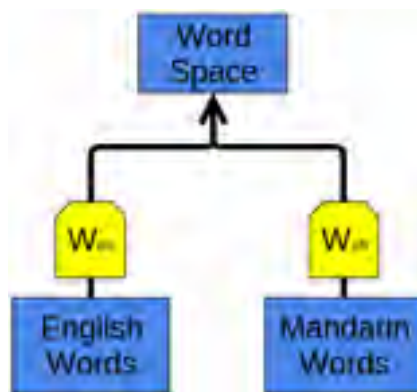
$W(\text{"woman"}) - W(\text{"man"}) \approx W(\text{"aunt"}) - W(\text{"uncle"})$

$W(\text{"woman"}) - W(\text{"man"}) \approx W(\text{"queen"}) - W(\text{"king"})$

$W(\text{"中国"}) - W(\text{"北京"}) \approx W(\text{"英国"}) - W(\text{"伦敦"})$



From Mikolov et al. (2013)



Socher et al. (2013)

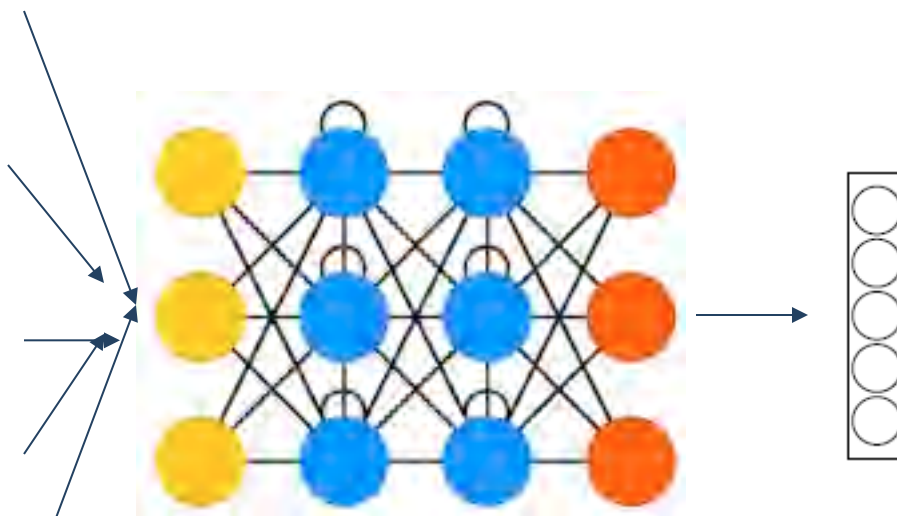


句子表示

语言表示学习

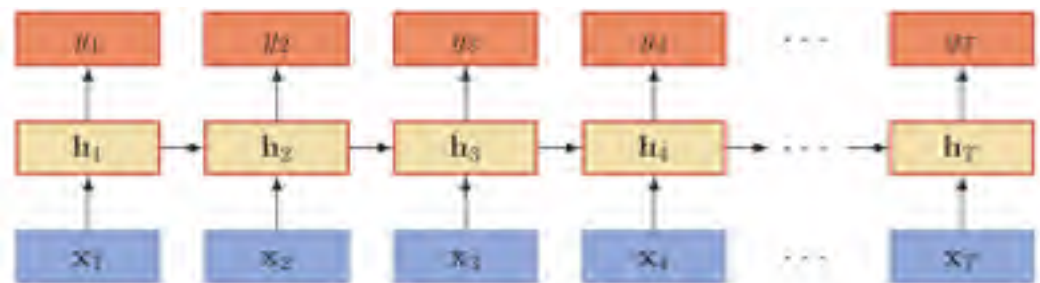
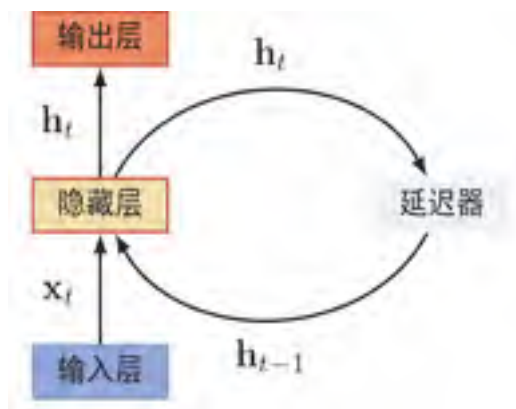
- ▶ 词
- ▶ 短语
 - ▶ 组合语义模型
- ▶ 句子
 - ▶ 连续词袋模型
 - ▶ 序列模型
 - ▶ 递归组合模型
 - ▶ 卷积模型
- ▶ 篇章
 - ▶ 层次模型

北京的天气真不错。



循环神经网络 (RNN)

缺点：长距离依赖问题



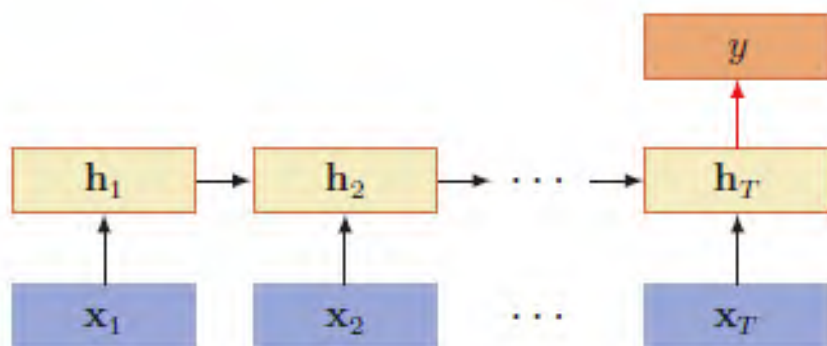
$$h_t = \begin{cases} 0 & t = 0 \\ f(h_{t-1}, x_t) & \text{otherwise} \end{cases}$$

RNN是图灵完全等价的 (Siegelmann and Sontag, 1995)

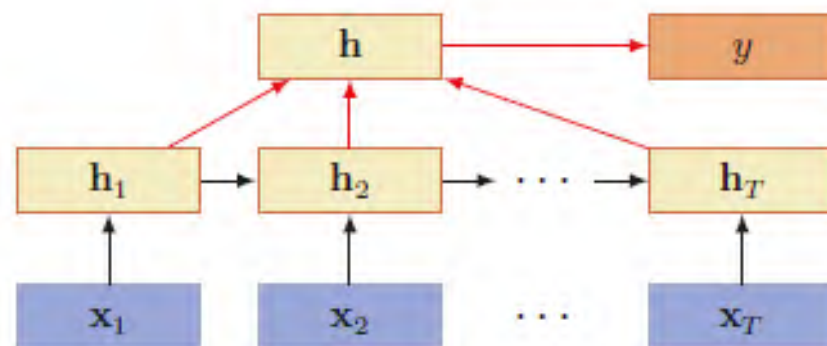
FNN: 模拟任何函数

RNN: 模拟任何程序 (计算过程)。

序列模型：RNN



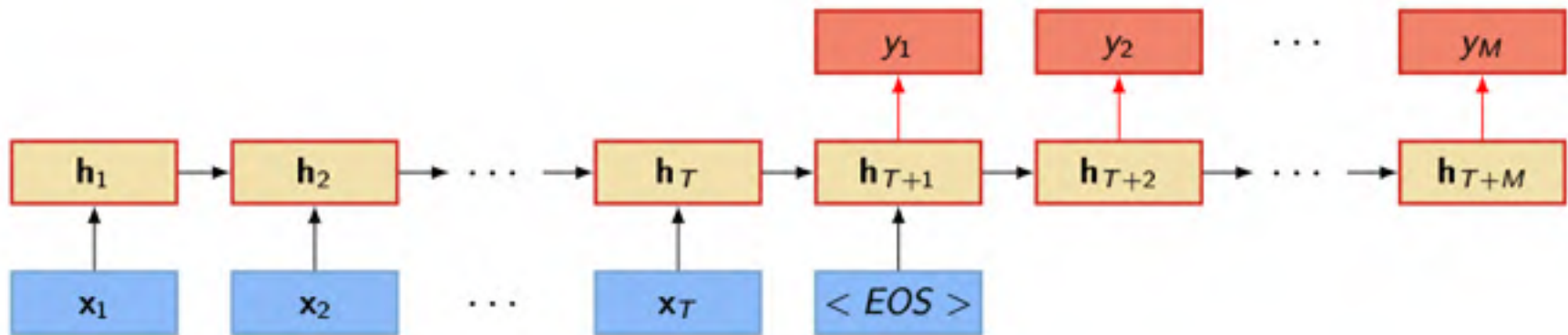
(a) 正常模式



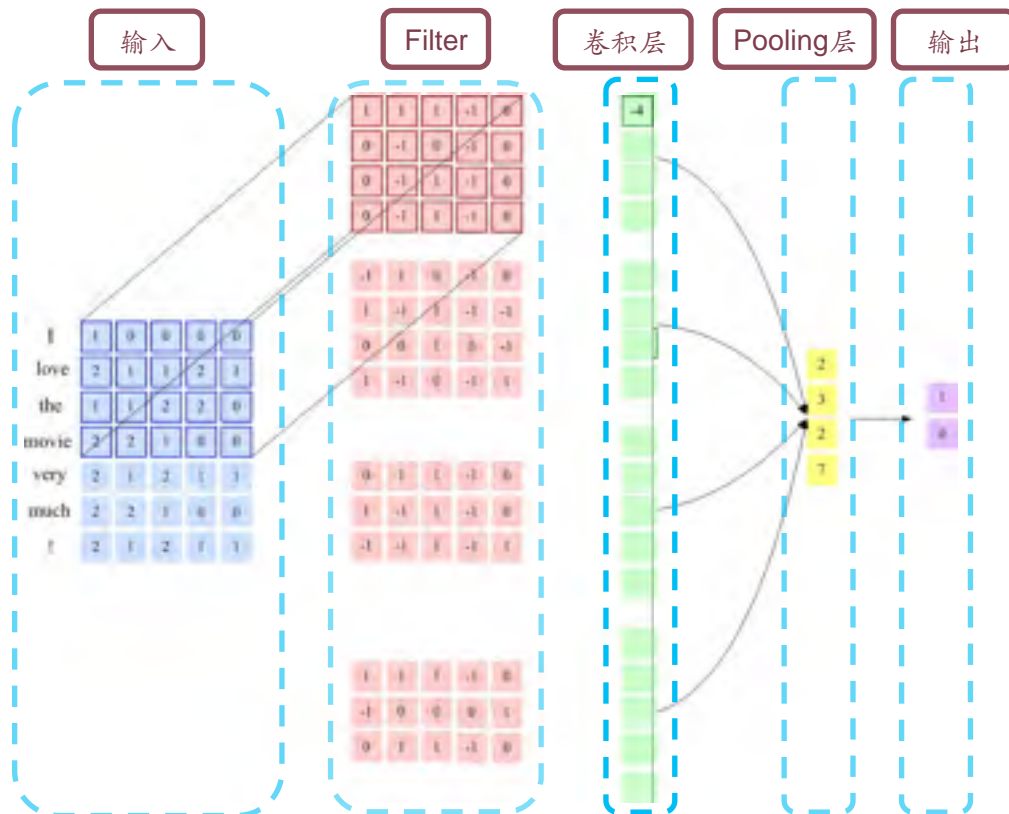
(b) 按时间进行平均采样模式



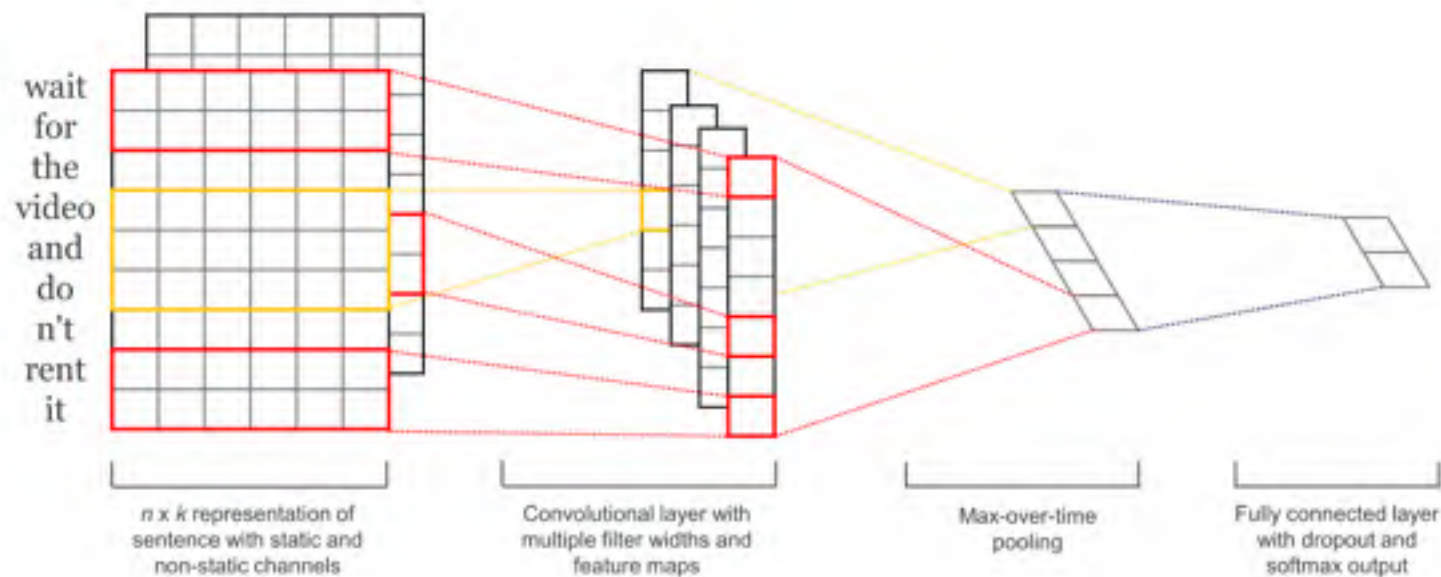
序列到序列模型



文本序列的卷积



基于卷积模型的句子表示



Y. Kim. "Convolutional neural networks for sentence classification" . In: *arXiv preprint arXiv:1408.5882* (2014).

递归神经网络

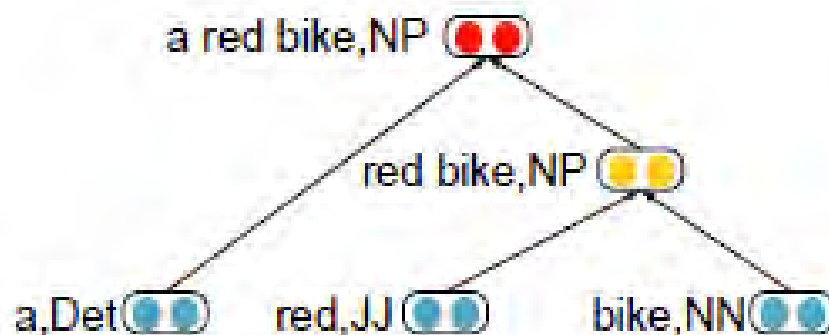
给定一个语法树,

$$p_2 \rightarrow ap_1,$$

$$p_1 \rightarrow bc.$$

$$p_1 = f\left(W \begin{bmatrix} b \\ c \end{bmatrix}\right),$$

$$p_2 = f\left(W \begin{bmatrix} a \\ p_1 \end{bmatrix}\right).$$





语言表示学习

		表示学习模型	
		词	句子、篇章
离散表示	符号表示	One-Hot表示	词袋模型 N元模型
	基于聚类的表示	Brown聚类	K-means聚类
连续表示	分布式表示	潜在语义分析 潜在狄利克雷分配	
	分散式表示	NNLM Skip-Gram模型 CBOW模型	连续词袋模型 序列模型 递归组合模型 卷积模型

为什么语言表示学习更难？

152 层



22层



Results:

- golden retriever: 0.97293
- Tibetan mastiff: 0.01576
- Irish setter: 0.00364
- redbone: 0.00152
- standard poodle: 0.00127

计算机视觉中的深层网络模型

对应NLP的最底层：词汇

语言表示的几个问题



长期依赖问题

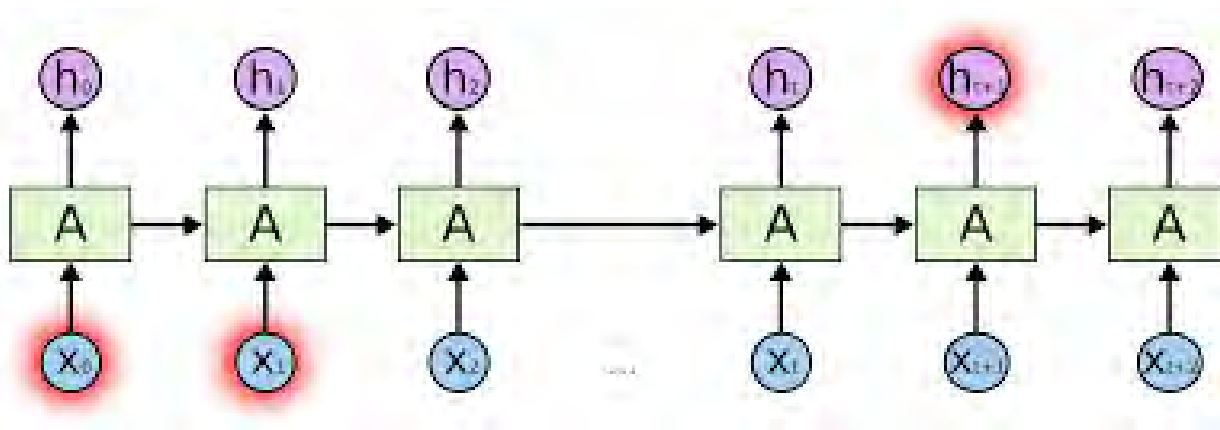
▶ 梯度消失/爆炸（主要因素）

▶ 改进：引入一个近似线性依赖的记忆单元来存储远距离的信息。

▶ 记忆容量（次要因素）

▶ 记忆单元的存储能力和其大小相关。

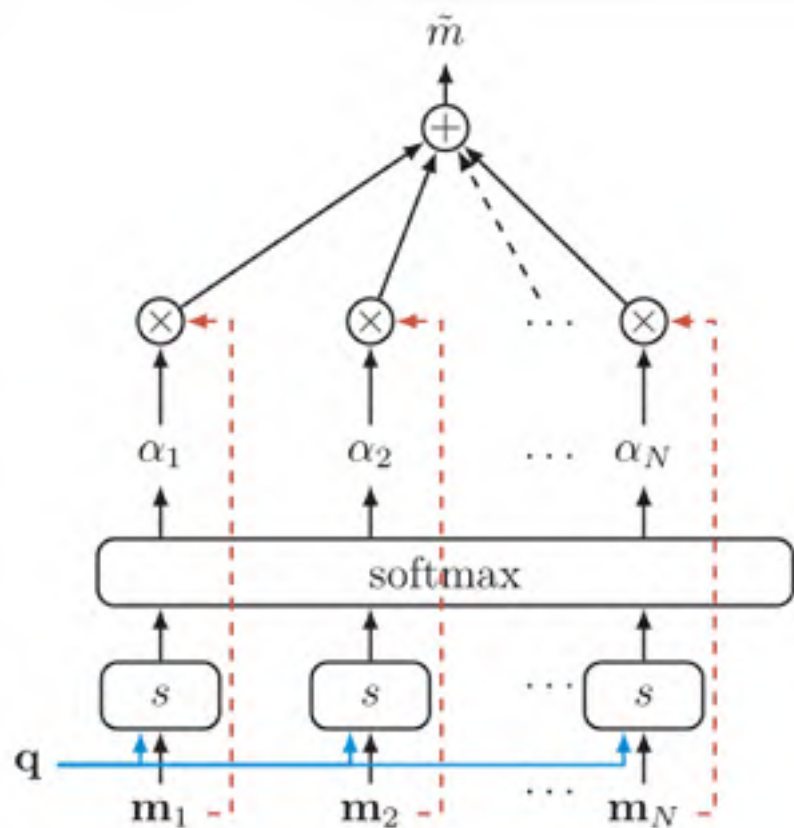
▶ 改进：注意力机制与外部记忆



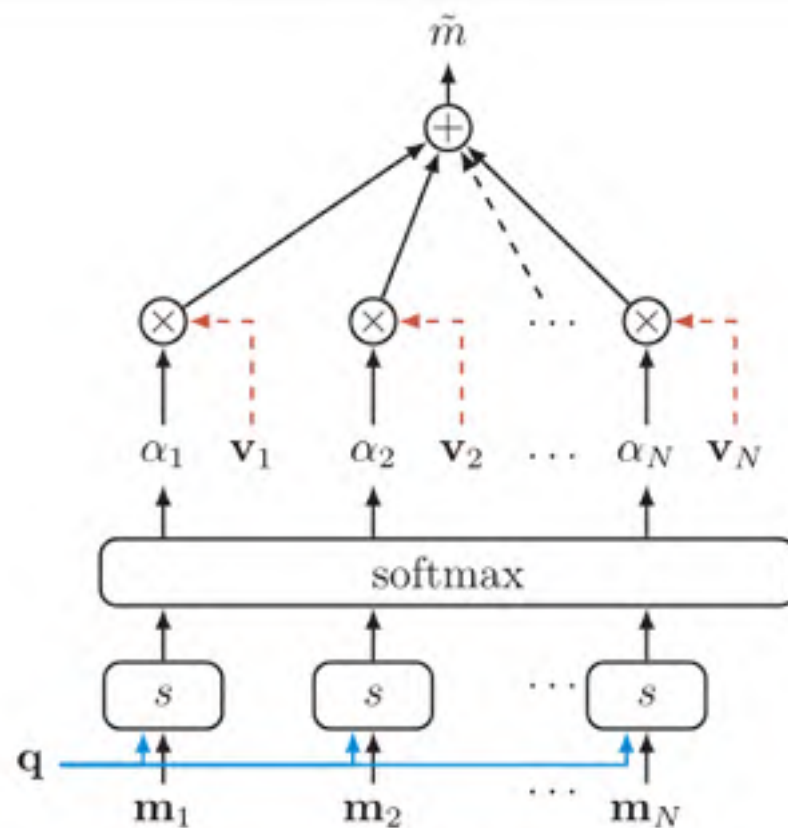
注意力



注意力模型



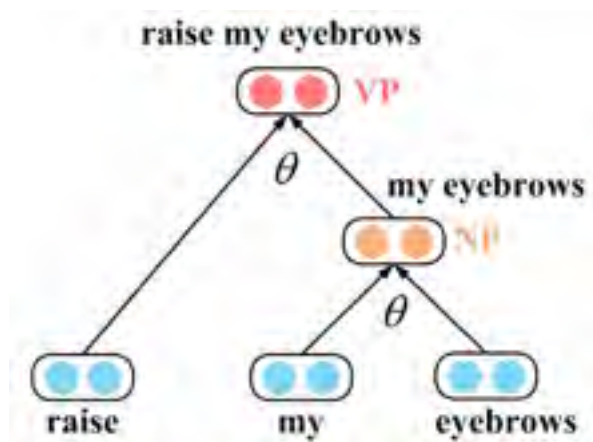
(a) 普通模式



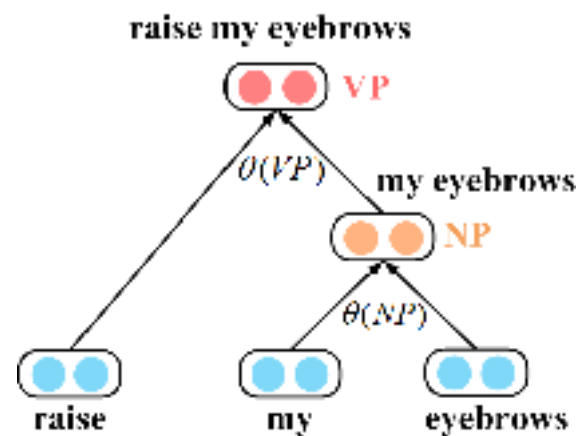
(b) 键值对模式

语言语义组合

- ▶ 如何组合自然语言的语义?
- ▶ 参数共享?



共享



不共享

动态语义组合网络

▶ 元网络 (Meta network)

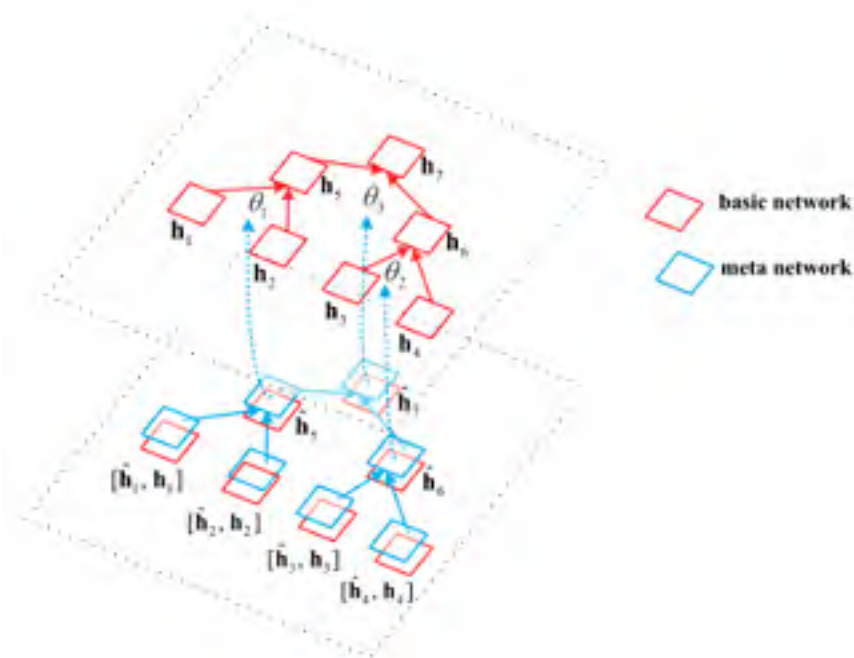
▶ 生成基网络参数

▶ 基网络 (Basic Network)

▶ 动态参数

Pengfei Liu, Xipeng Qiu, Xuanjing Huang, Dynamic Compositional Neural Networks over Tree Structure, In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI), pp. 4054-4060, 2017.

Pengfei Liu, Kaiyu Qian, Xipeng Qiu, Xuanjing Huang, Idiom-Aware Compositional Distributed Semantics, In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1215-1224, 2017.

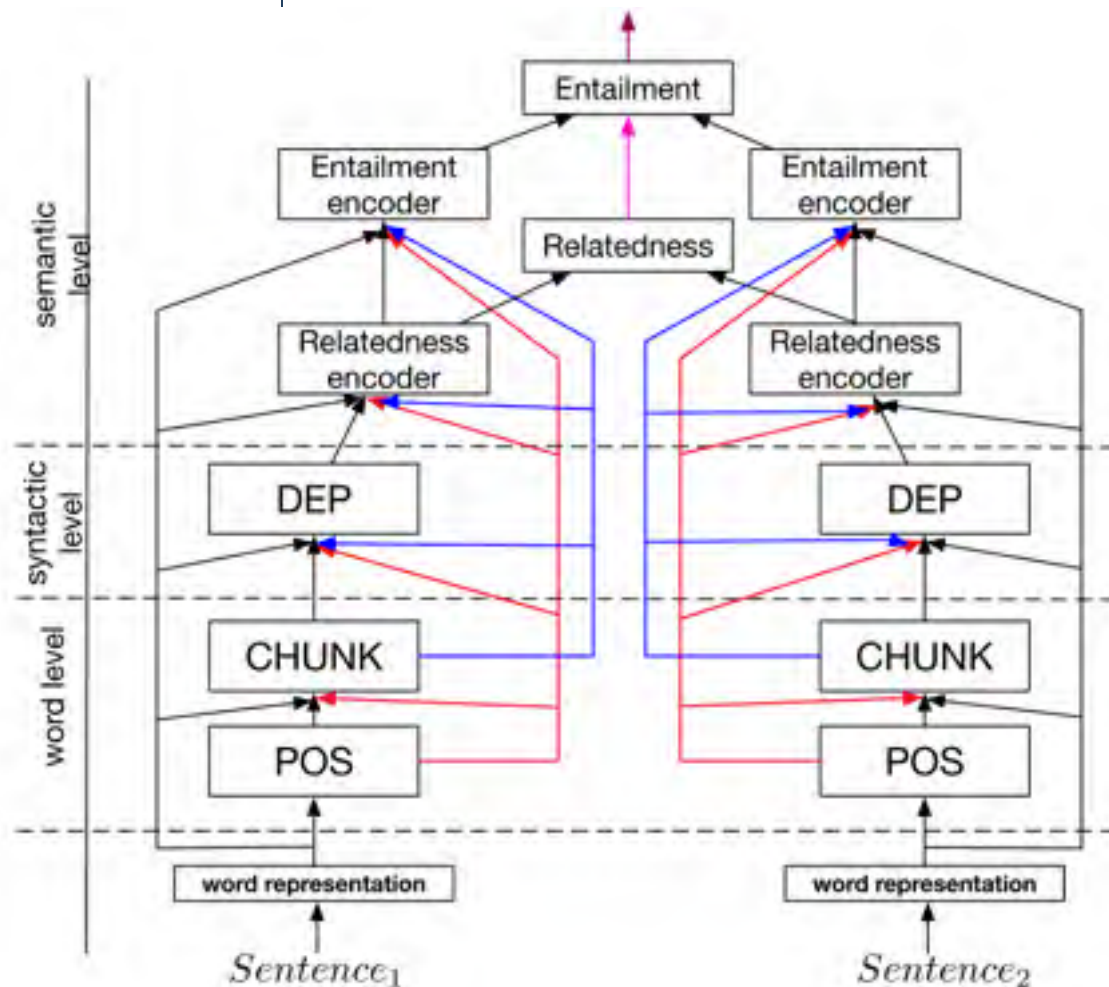




多任务学习

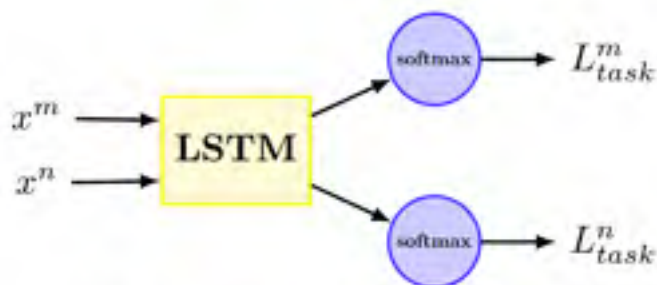
知识共享

- 词性标注
- 组块分析
- 依次句法分析
- 文本蕴涵

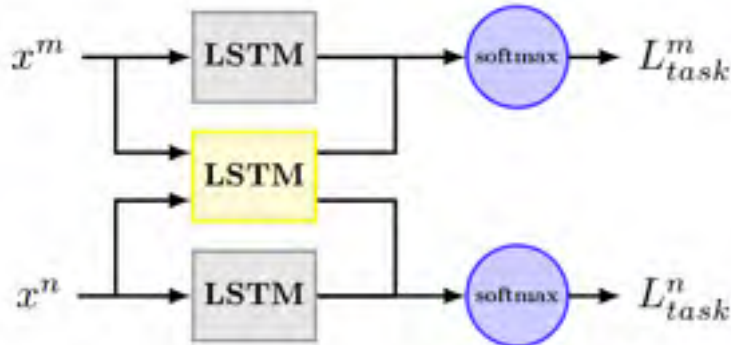


如何学习任务无关的共享表示

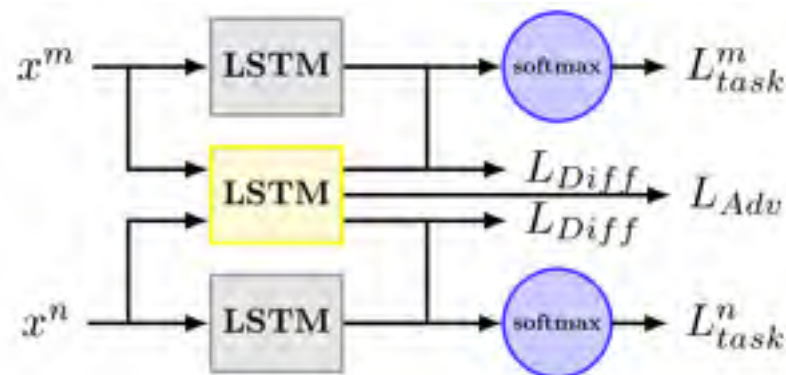
► 对抗学习



(a) Fully Shared Model (FS-MTL)



(b) Shared-Private Model (SP-MTL)



$$L = L_{Task} + \lambda L_{Adv} + \gamma L_{Diff}$$

Pengfei Liu, Xipeng Qiu, Xuanjing Huang, Adversarial Multi-task Learning for Text Classification, In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 1-10, 2017.

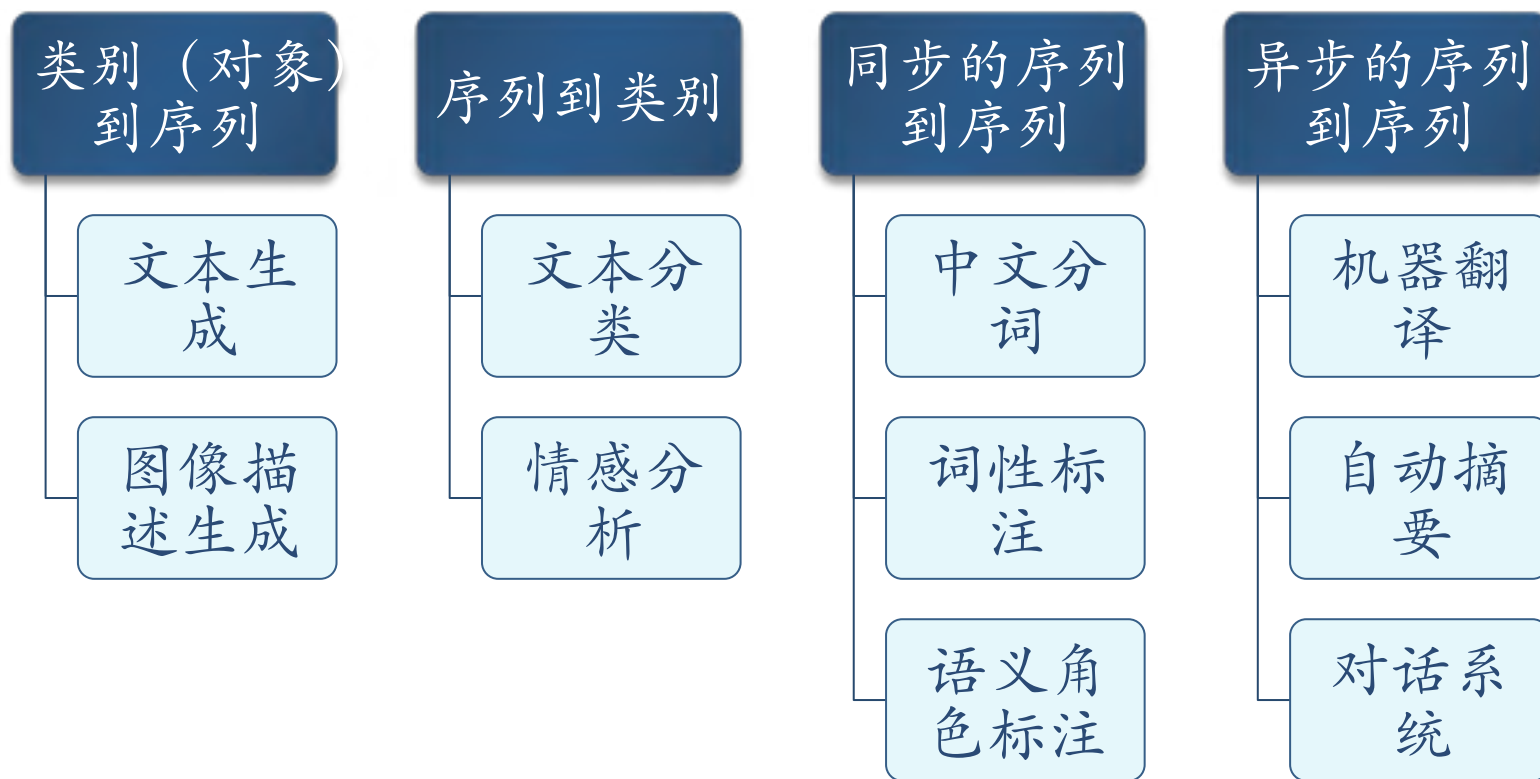


自然语言处理的新范式



自然语言处理任务

► 在得到字、句子表示之后，自然语言处理任务类型划分为



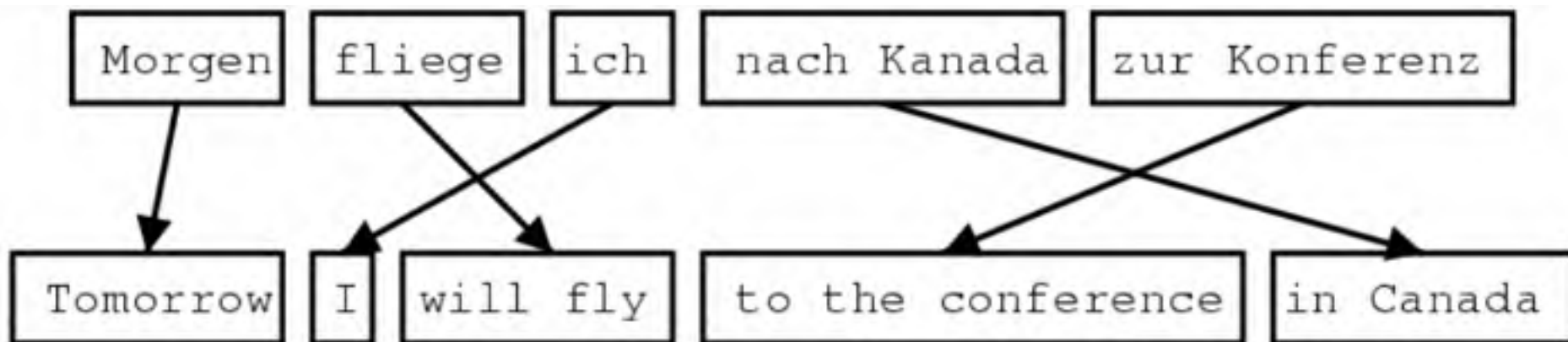
减轻了对特征工程的依赖!



应用例子

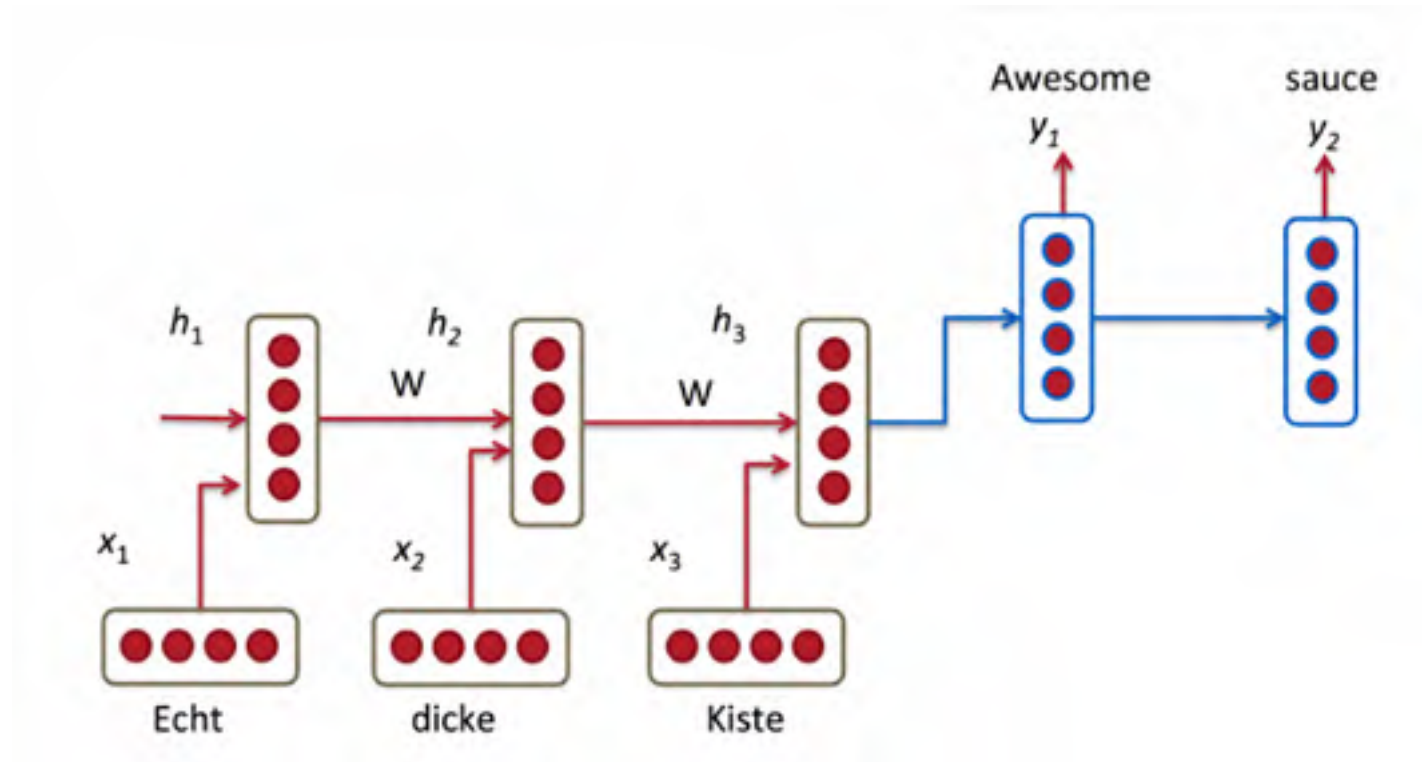
传统统计机器翻译

- ▶ 源语言: f
- ▶ 目标语言: e
 - ▶ 模型: $e = \operatorname{argmax}_e p(e|f) = \operatorname{argmax}_e p(f|e)p(e)$
 - ▶ $p(f|e)$: 翻译模型
 - ▶ $p(e)$: 语言模型

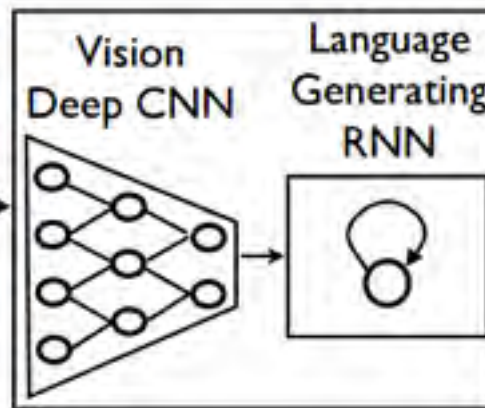


基于序列到序列的机器翻译

- ▶ 一个RNN用来编码
- ▶ 另一个RNN用来解码



看图说话



A group of people shopping at an outdoor market.

There are many vegetables at the fruit stand.

看图说话



Figure 5. A selection of evaluation results, grouped by human rating.



生成Linux内核代码

```
/*
 * If this error is set, we will need anything right after that BSD.
 */
static void action_new_function(struct s_stat_info *wb)
{
    unsigned long flags;
    int lel_idx_bit = e->edd, *sys & -((unsigned long) *FIRST_COMPAT);
    buf[0] = 0xFFFFFFFF & (bit << 4);
    min(inc, slist->bytes);
    printk(KERN_WARNING "Memory allocated %02x/%02x, "
        "original MLL instead\n"),
        min(min(multi_run - s->len, max) * num_data_in),
        frame_pos, sz + first_seg);
    div_u64_w(val, inb_p);
    spin_unlock(&disk->queue_lock);
    mutex_unlock(&s->sock->mutex);
    mutex_unlock(&func->mutex);
    return disassemble(info->pending_bh);
}

static void num_serial_settings(struct tty_struct *tty)
{
    if (tty == tty)
        disable_single_st_p(dev);
    pci_disable_spool(port);
}
```



作词机

▶ RNN在“学习”过汪峰全部作品后自动生成的歌词

- ▶ 我在这里中的夜里
- ▶ 就像一场是一种生命的意叶
- ▶ 就像我的生活变得在我一样
- ▶ 可我们这是一个知道
- ▶ 我只是一天你会怎吗
- ▶ 可我们这是我们是不要为你
- ▶ 我们想这有一种生活的时候

<https://github.com/phunterlau/wangfeng-rnn>

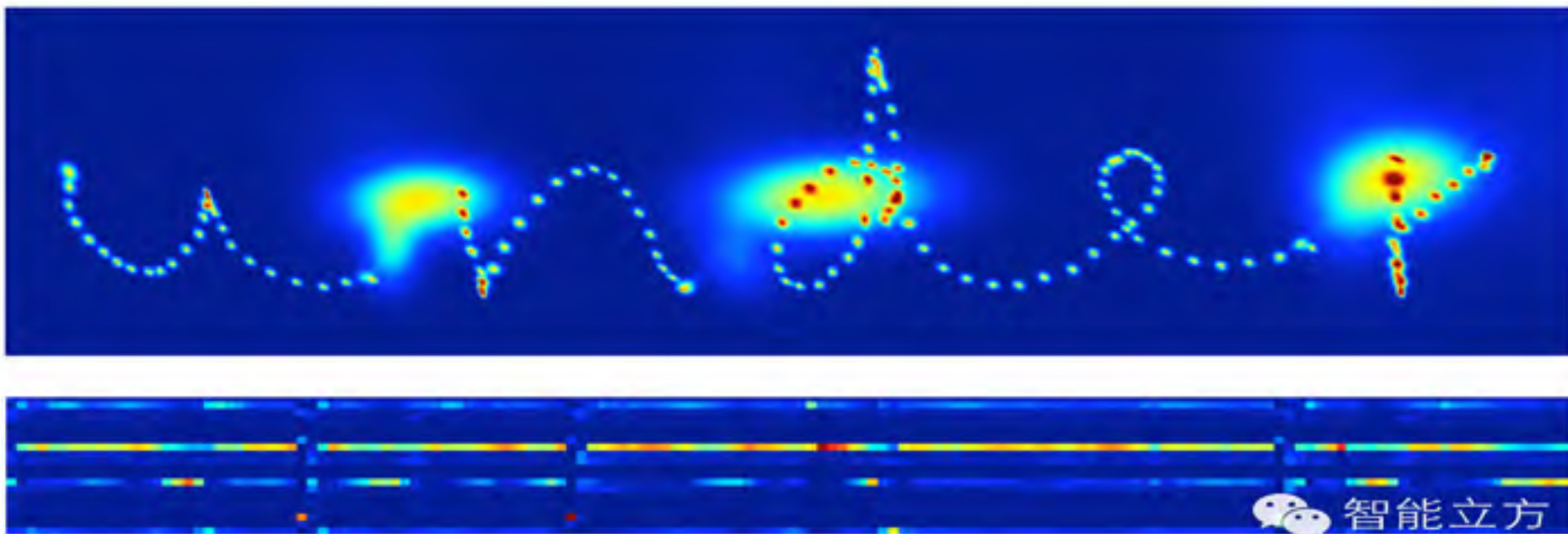


作诗

<p>白鹭窥鱼立， Egrets stood, peeping fishes. 青山照水开。 Water was still, reflecting mountains. 夜来风不动， The wind went down by nightfall, 明月见楼台。 as the moon came up by the tower.</p>	<p>满怀风月一枝春， Budding branches are full of romance. 未见梅花亦可人。 Plum blossoms are invisible but adorable. 不为东风无此客， With the east wind comes Spring. 世间何处是前身。 Where on earth do I come from?</p>
--	--

写字

- ▶ 把一个字母的书写轨迹看作是一连串的点。一个字母的“写法”其实是每一个点相对于前一个点的偏移量，记为(offset x, offset y)。再增加一维取值为0或1来记录是否应该“提笔”。



Making Neural Nets Great Again





阅读理解

- ▶ 三元组 (Q,D,A)
- ▶ 问题Q: $(q\downarrow 1, q\downarrow 2, \dots, q\downarrow m)$
- ▶ 文档D: $(x\downarrow 1, x\downarrow 2, \dots, x\downarrow n)$
- ▶ 答案A: $x\downarrow s, \dots, x\downarrow e$

Fred moved to the bedroom and Joe went to the kitchen then Joe took the milk there and Dan journeyed to the bedroom; Joe discarded the milk.

Where is the milk now ?

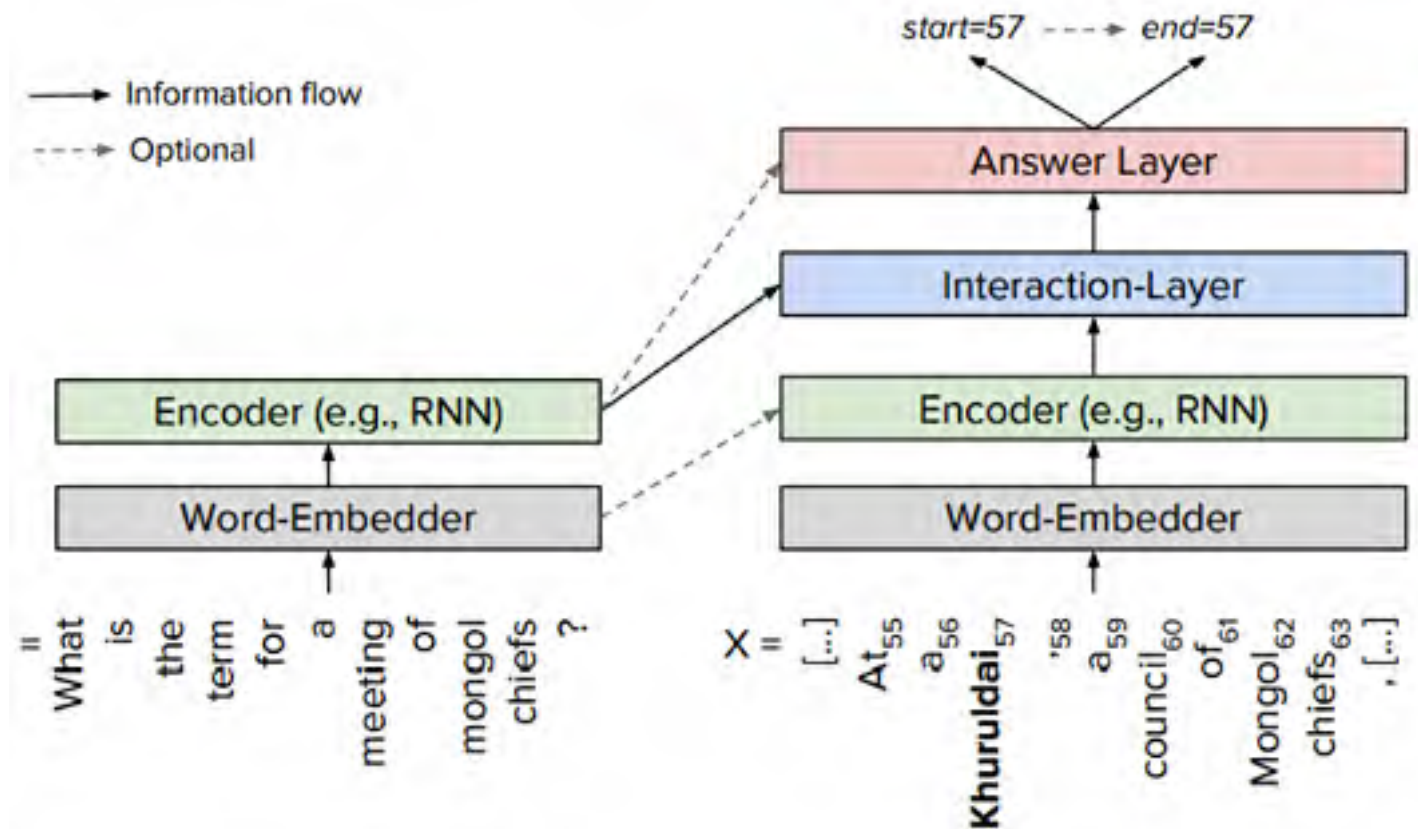
A: the milk is in the kitchen

Where is Dan now?

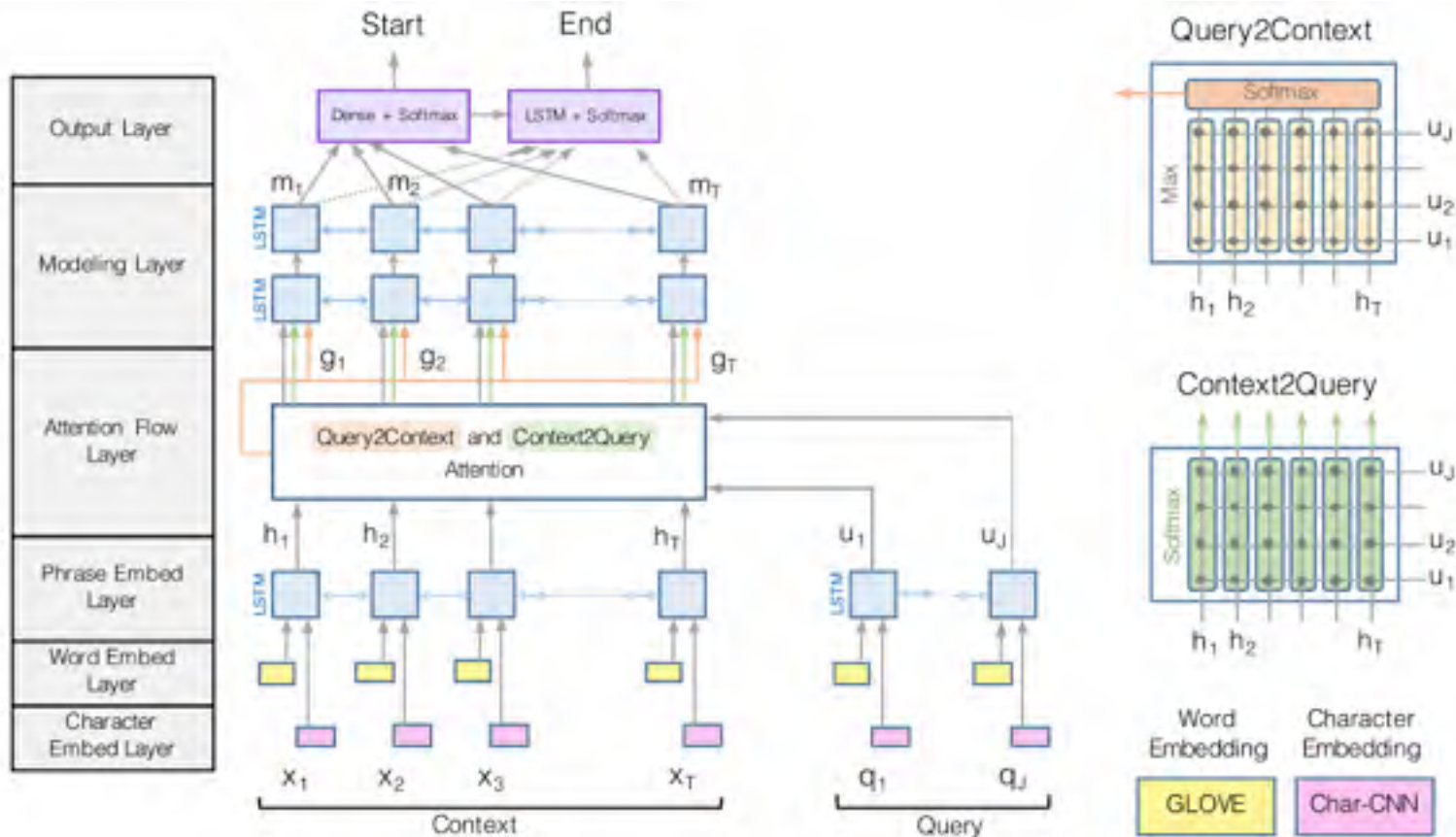
A: I think he is in the bedroom

SIMULATED WORLD QA

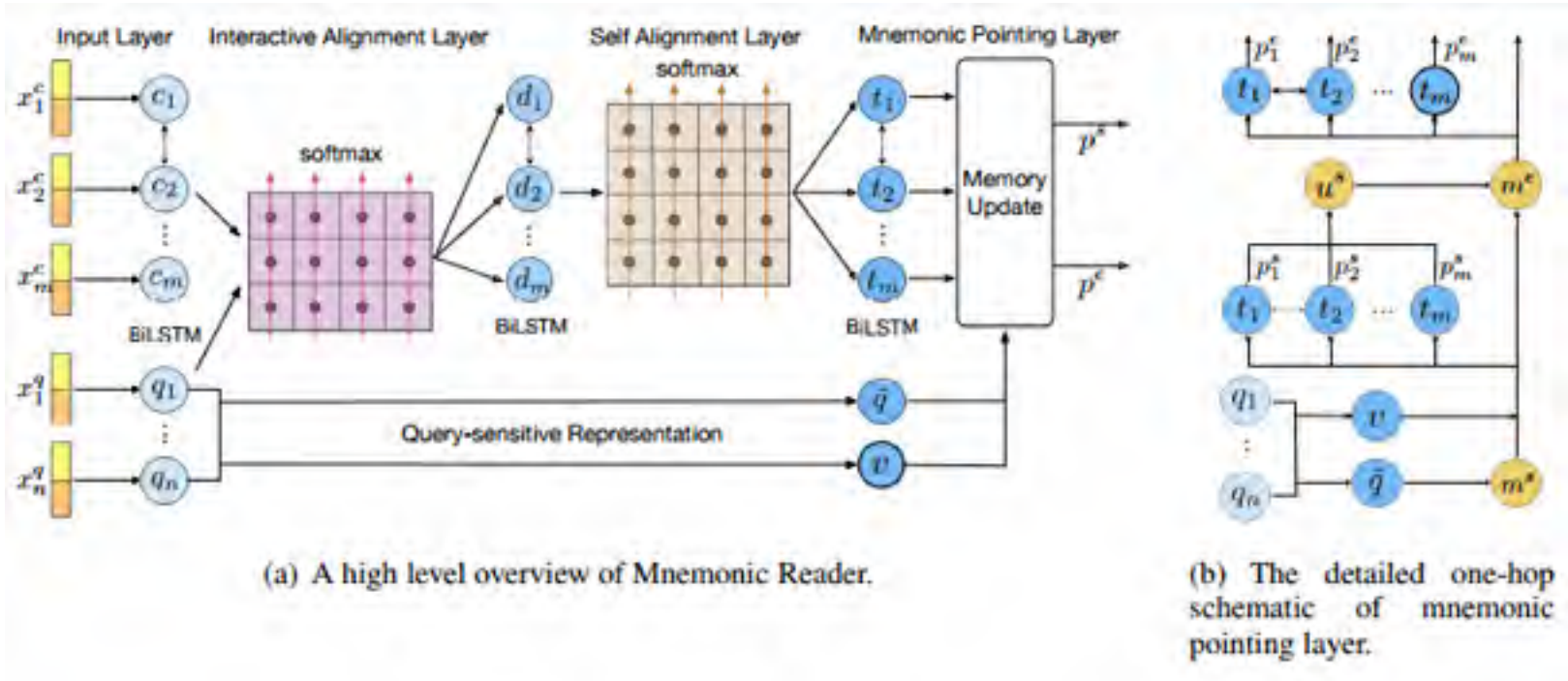
一般流程



Bidirectional Attention (Seo et al., 2016)

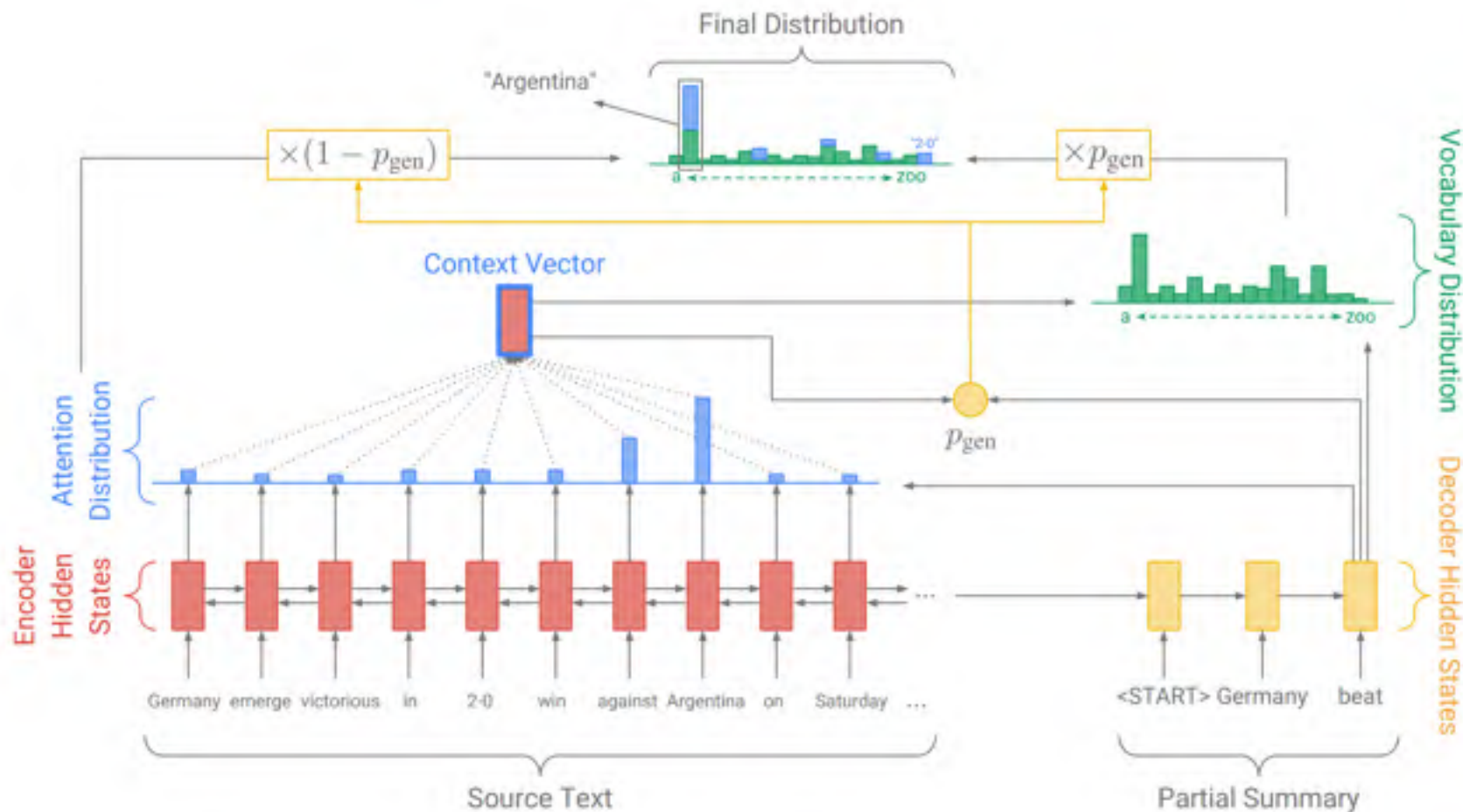


Mnemonic Reader



<https://arxiv.org/abs/1705.02798>

文本摘要





文本摘要

Reference summary

utility back francis sailli will join up with munster later this year .
the new zealand international has signed a two-year contract .
sailli made his debut for the all blacks against argentina in 2013 .

Sequence-to-sequence + attention summary

dutch international francis [UNK] has signed a two-year deal to join irish [UNK] super rugby side the blues .
[UNK] 's signature is something of a coup for munster and his head coach anthony foley believes he will be a great addition to their respective prospects .
[UNK] has been capped twice by new zealand .

Pointer-generator summary

new zealand international francis sailli will move to the province later this year .
utility back sailli made his all blacks debut against argentina in 2013 .
utility back sailli will move to the province later this year .

Pointer-generator model + coverage summary

francis sailli has signed a two-year deal to join munster later this year .
the 24-year-old was part of the new zealand under-20 side that won the junior world championship in italy in 2011 .
sailli 's signature is something of a coup for munster and head coach anthony foley .

对话

Li J, Monroe W, Ritter A, et al. Deep reinforcement learning for dialogue generation[J]. arXiv preprint arXiv:1606.01541, 2016.

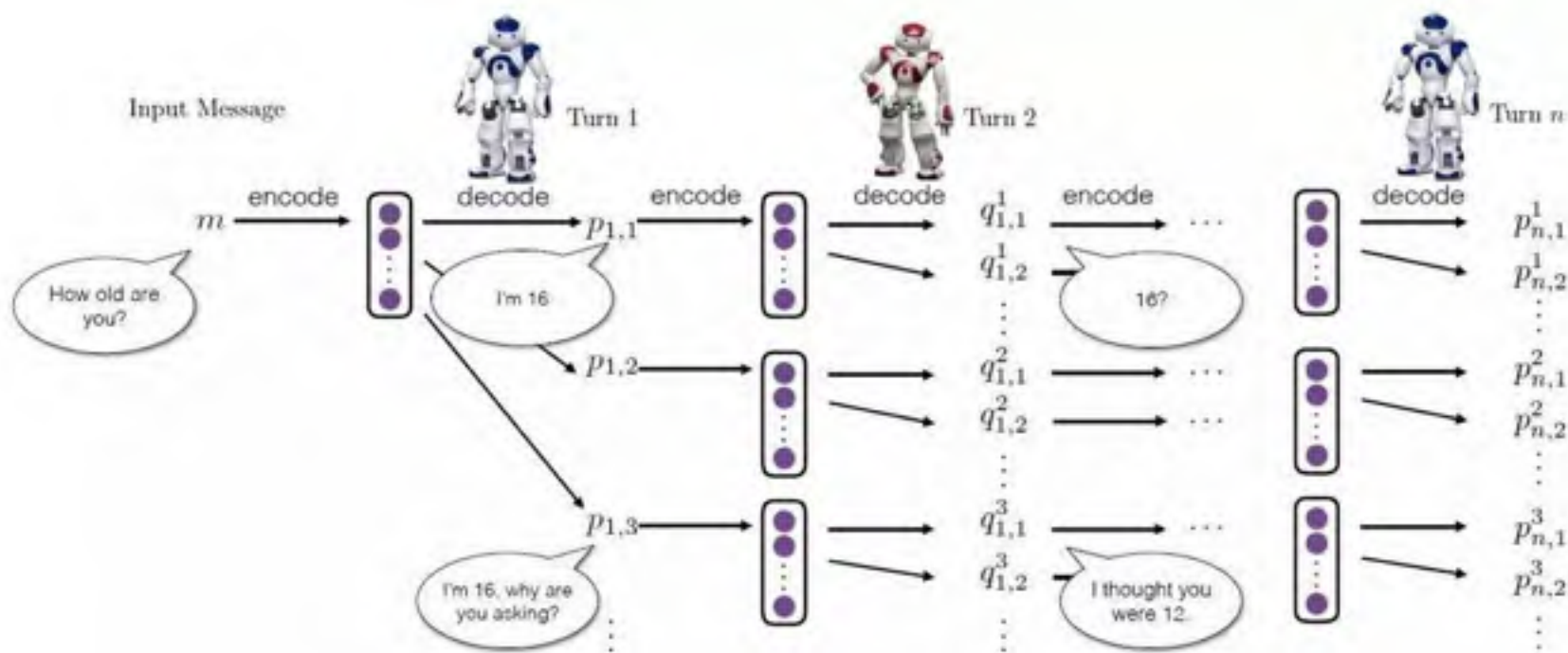


Figure 1: Dialogue simulation between the two agents.



总结

总结

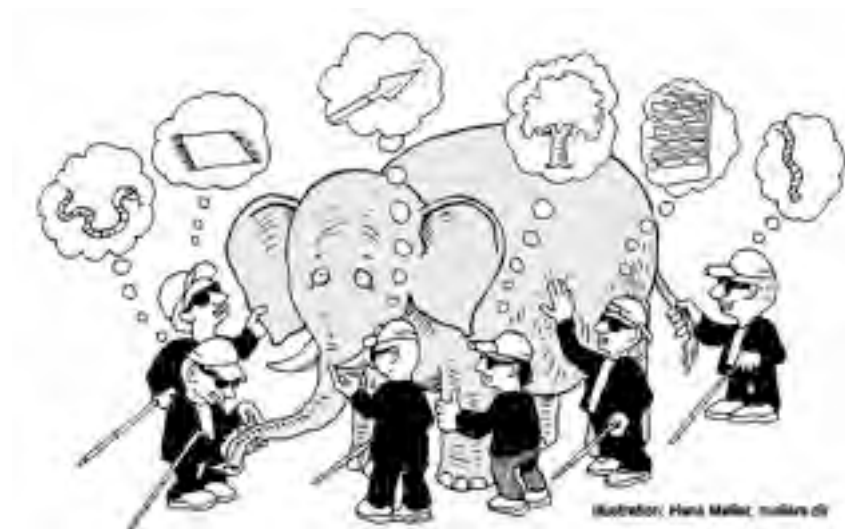
► 模型



► 任务



研究现状



Lapata's scream

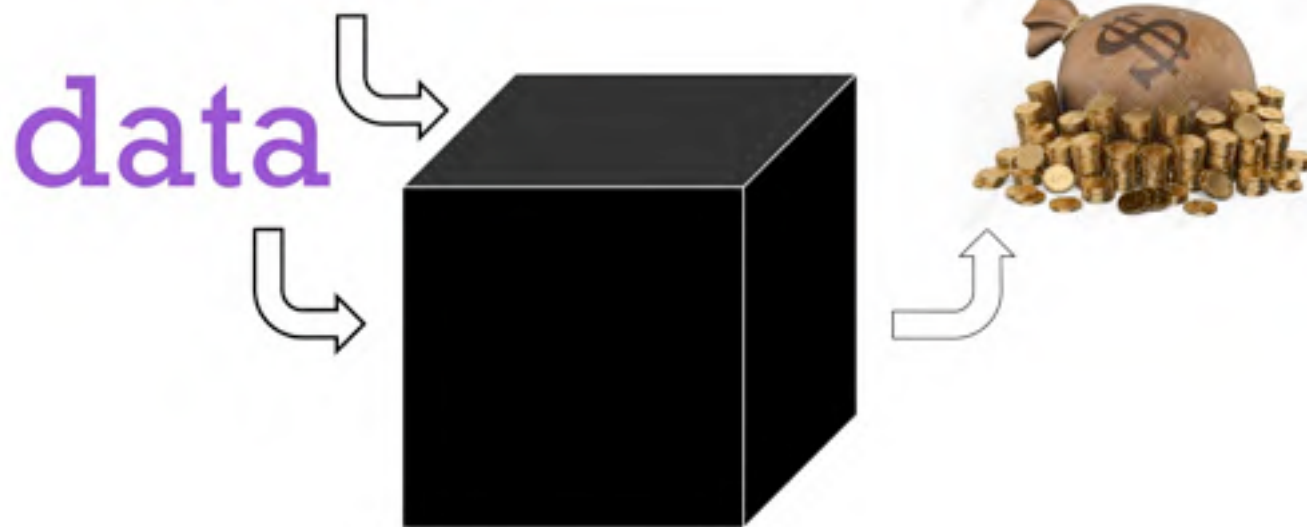
[Lapata ACL talk 2017]



Noah's Bias

[Smith ACL talk 2017]

(structural) **bias**



Putting things together

DL



RL



NLP



function
approximation

correct
training

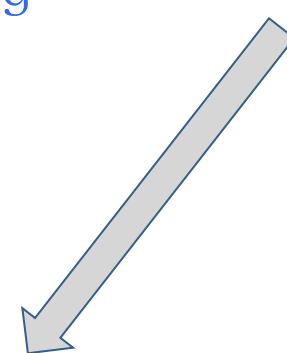
structural
bias



AlphaGo

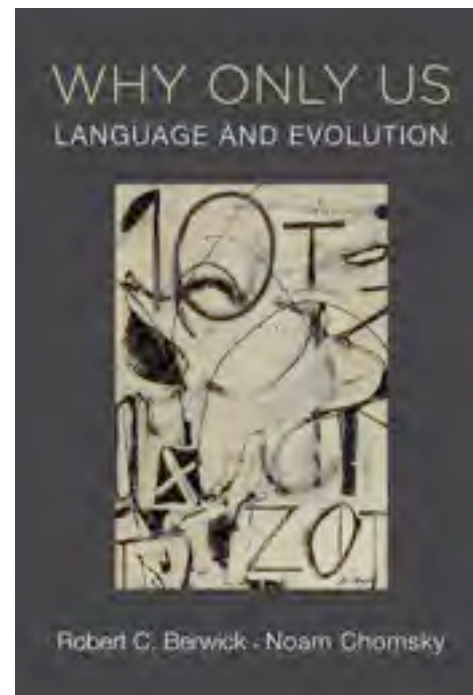


?



Language & reasoning

- ▶ 语言并不是为了交流而存在和发展，而是为了表征和操作人类的内部概念世界，
- ▶ 语言的统一性与多样性源于进化中的内化与外化过程。





谢 谢

<https://nndl.github.io/>