



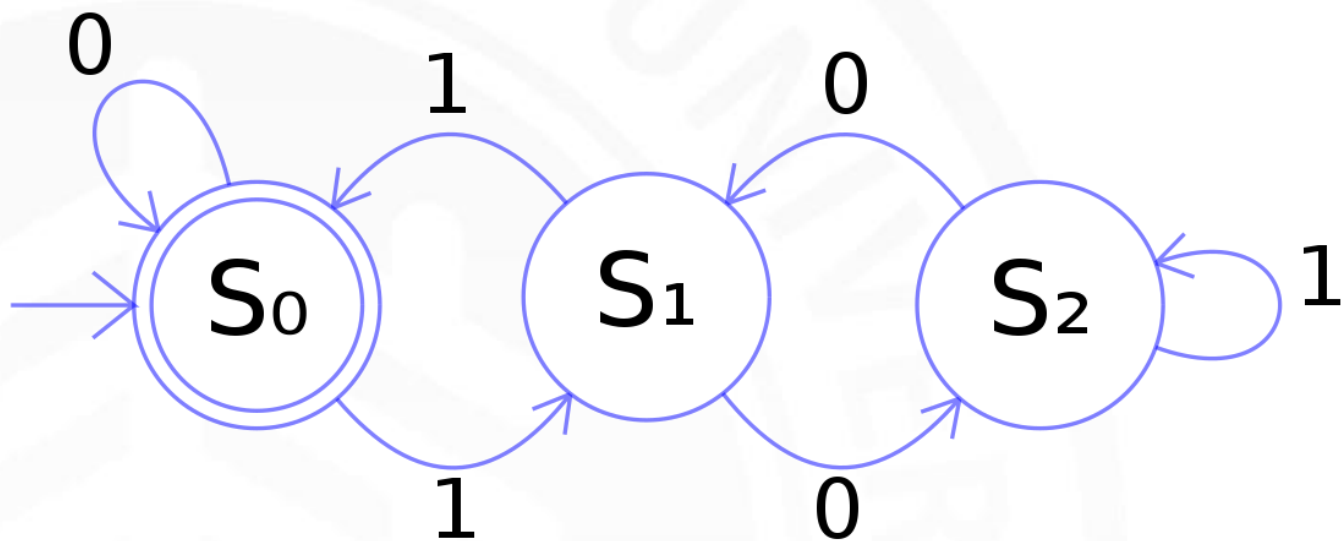
華東師範大學
EAST CHINA NORMAL UNIVERSITY

第 10章 循环神经网络

前馈网络

- 连接存在层与层之间，每层的节点之间是无连接的。（无循环）
- 输入和输出的维数都是固定的，不能任意改变。无法处理变长的序列数据。
- 假设每次输入都是独立的，也就是说每次网络的输出只依赖于当前的输入。

有限状态自动机 (Finite Automata)

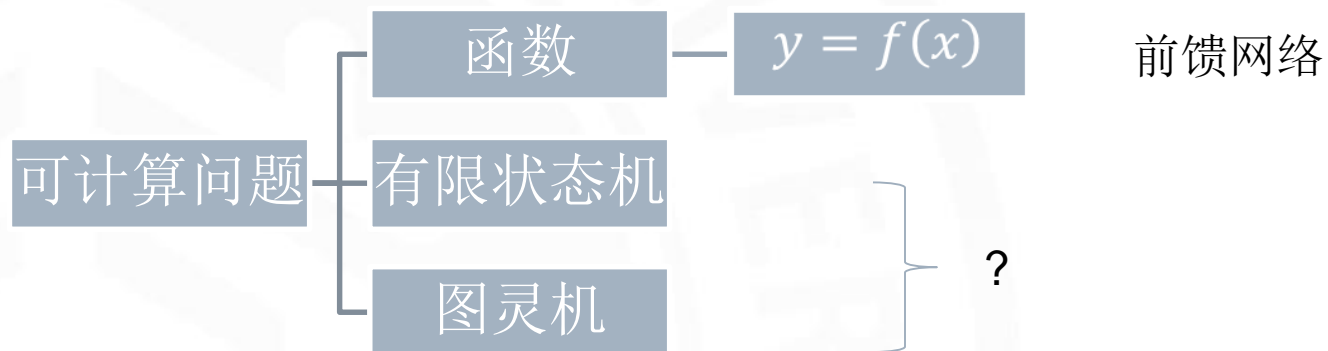


如何用**FNN**去模拟一个有限状态自动机？

可计算问题

■ 前馈网络

■ 通用近似定理

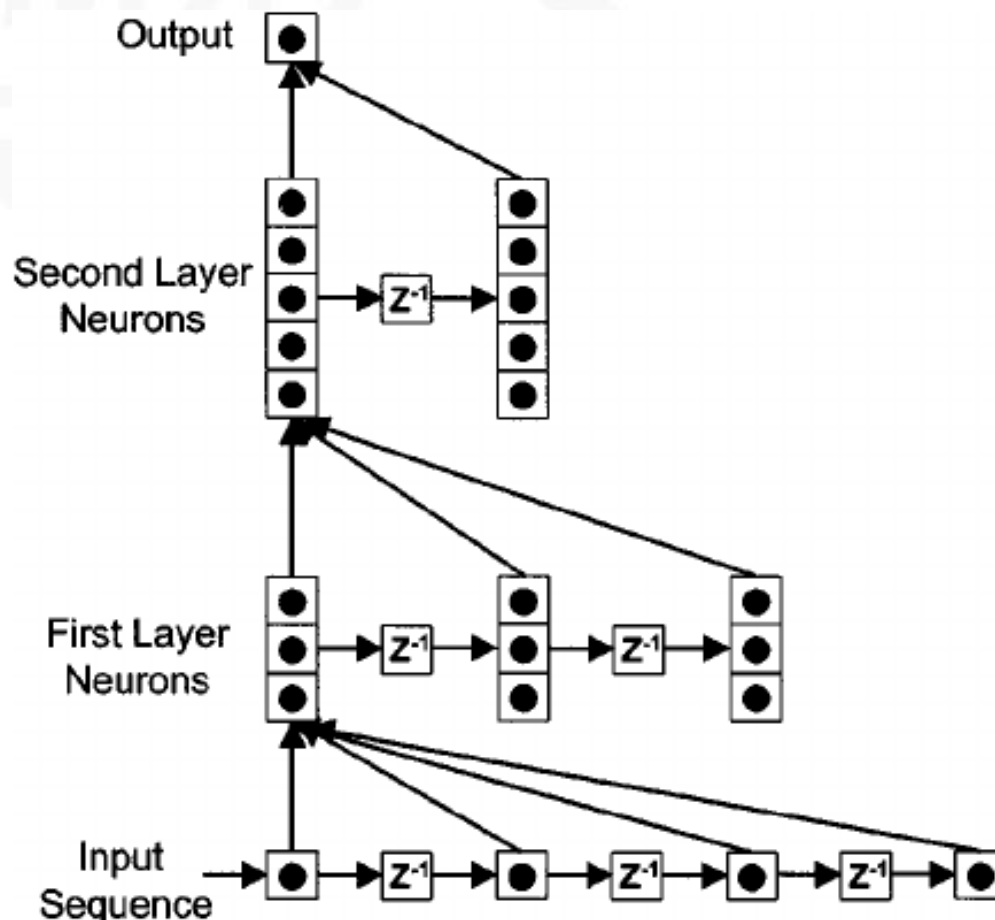


如何给网络增加记忆能力？

- 延时神经网络 (Time Delay Neural Network, TDNN)
- 建立一个额外的延时单元，用来存储网络的历史信息（可以包括输入、输出、隐状态等）

$$\mathbf{h}_t^{(l+1)} = f(\mathbf{h}_t^{(l)}, \mathbf{h}_{t-1}^{(l)}, \dots, \mathbf{h}_{t-p}^{(l)})$$

延时神经网络



<https://www.researchgate.net/publication/12314435> Neural system identification model of human sound localization

如何给网络增加记忆能力？

- 自回归模型 (Autoregressive Model, AR)

- 一类时间序列模型，用一个变量 y_t 的历史信息来预测自己

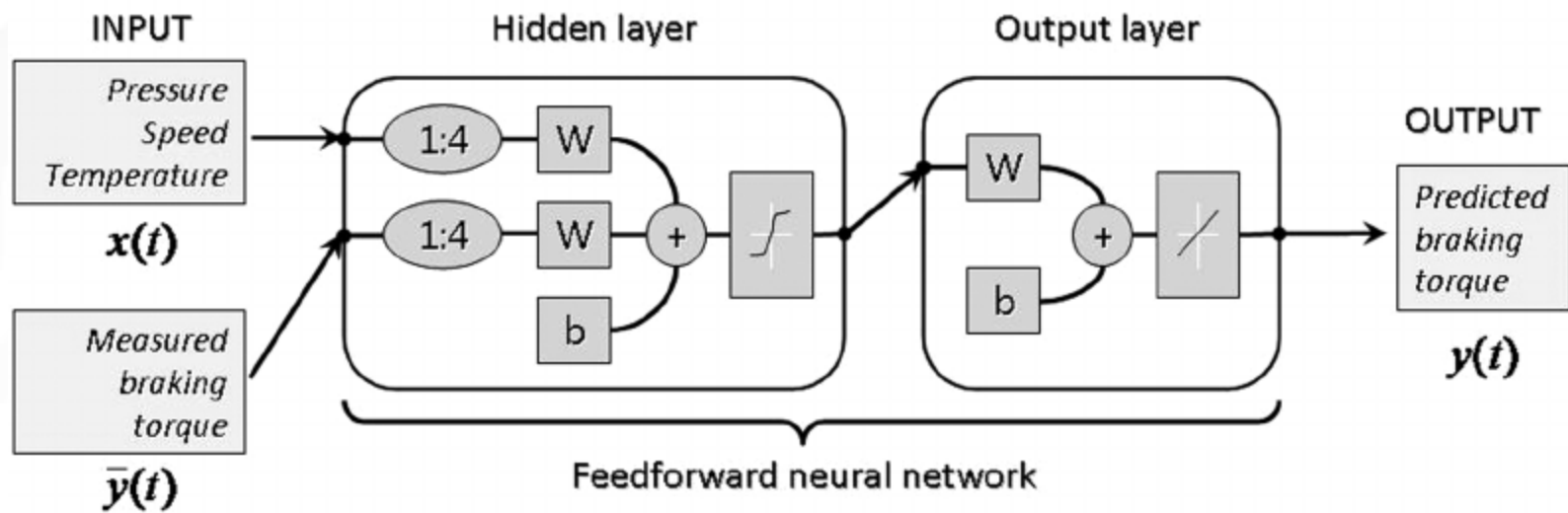
$$y_t = w_0 + \sum_{i=1}^p w_p y_{t-i} + \epsilon_t$$

- $\epsilon_t \sim N(0, \sigma^2)$ 为第 t 个时刻的噪声

- 有外部输入的非线性自回归模型 (Nonlinear Autoregressive with Exogenous Inputs Model,

$$N y_t = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{t-q})$$

非线性自回归模型

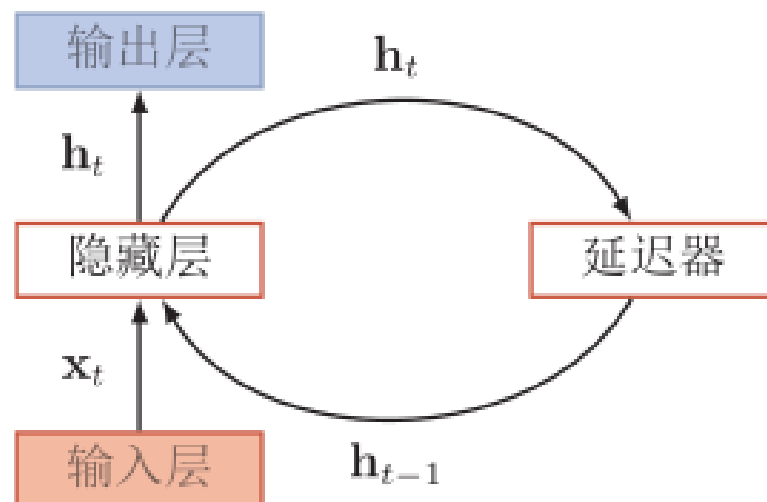


<https://www.researchgate.net/publication/234052442> Braking torque control using reccurent neural networks

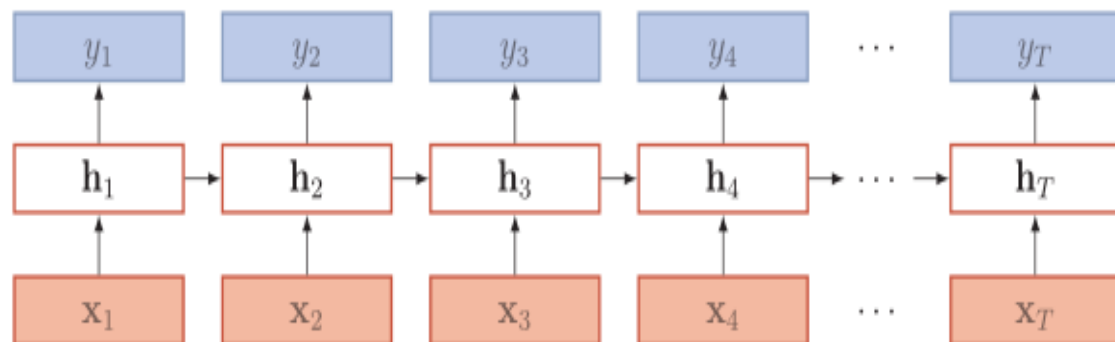
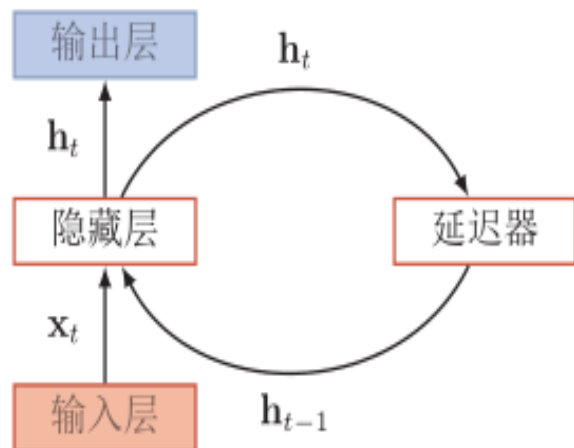
循环神经网络

给定一个输入序列 $\mathbf{x}_{1:T} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t, \dots, \mathbf{x}_T)$ ，循环神经网络通过下面公式更新带反馈边的隐藏层的活性值 \mathbf{h}_t ：

$$\mathbf{h}_t = f(\mathbf{h}_{t-1}, \mathbf{x}_t)$$



按时间展开



循环神经网络

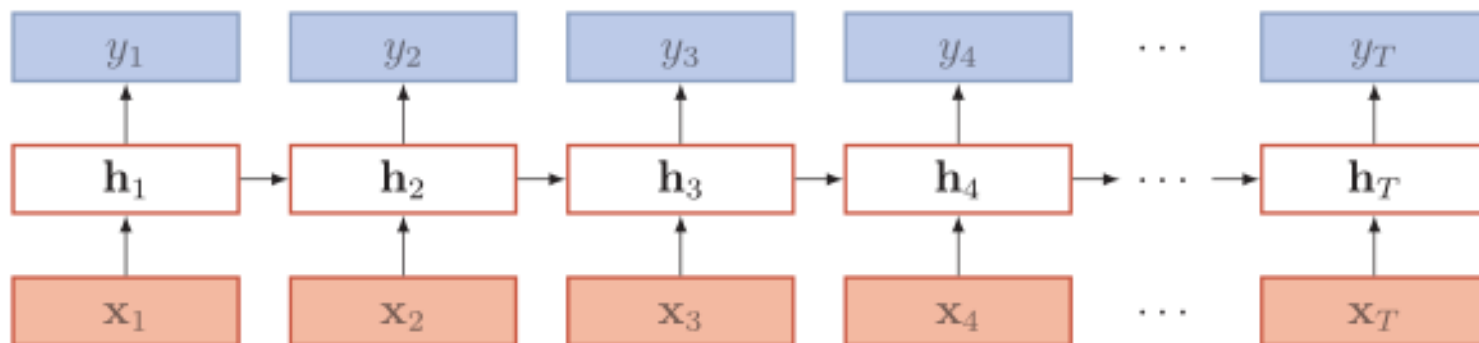
■ 循环神经网络

- 循环神经网络通过使用带自反馈的神经元，能够处理任意长度的序列。
- 循环神经网络比前馈神经网络更加符合生物神经网络的结构。
- 循环神经网络已经被广泛应用于语音识别、语言模型以及自然语言生成等任务上。

简单循环网络

■ 状态更新:

$$\mathbf{h}_t = f(U\mathbf{h}_{t-1} + W\mathbf{x}_t + \mathbf{b}),$$



通用近似定理

- 一个完全连接的循环网络是任何非线性动力系统的近似器。

定理 6.1 – 通用近似定理 [Haykin, 2009]: 如果一个完全连接的循环神经网络有足够数量的 sigmoid 型隐藏神经元, 它可以以任意的准确率去近似任何一个非线性动力系统

$$\mathbf{s}_t = g(\mathbf{s}_{t-1}, \mathbf{x}_t), \quad (6.10)$$

$$\mathbf{y}_t = o(\mathbf{s}_t), \quad (6.11)$$

其中 \mathbf{s}_t 为每个时刻的隐状态, \mathbf{x}_t 是外部输入, $g(\cdot)$ 是可测的状态转换函数, $o(\cdot)$ 是连续输出函数, 并且对状态空间的紧致性没有限制。

图灵完备

- 图灵完备 (Turing Completeness) 是指一种数据操作规则，比如一种计算机编程语言，可以实现图灵机的所有功能，解决所有的可计算问题。

定理 6.2 – 图灵完备 [Siegelmann and Sontag, 1991]: 所有的图灵机都可以被一个由使用 sigmoid 型激活函数的神经元构成的全连接循环网络来进行模拟。

- 一个完全连接的循环神经网络可以近似解决所有的可计算问题。

循环神经网络

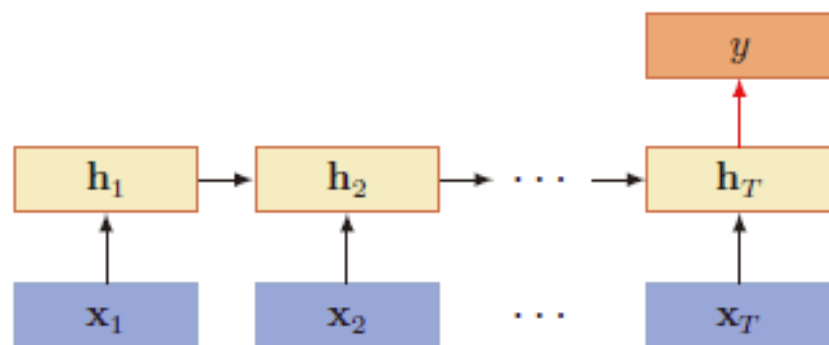
■ 作用

- 输入-输出映射（机器学习）

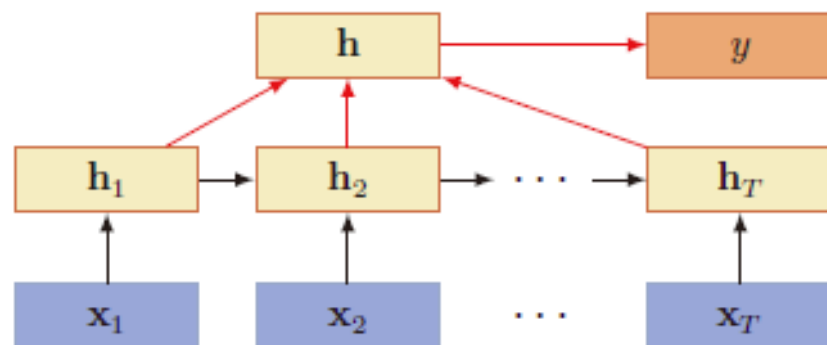
- 存储

应用到机器学习

■ 序列到类别



(a) 正常模式



(b) 按时间进行平均采样模式

序列到类别

情感分类

带着愉悦的心情
看了这部电影

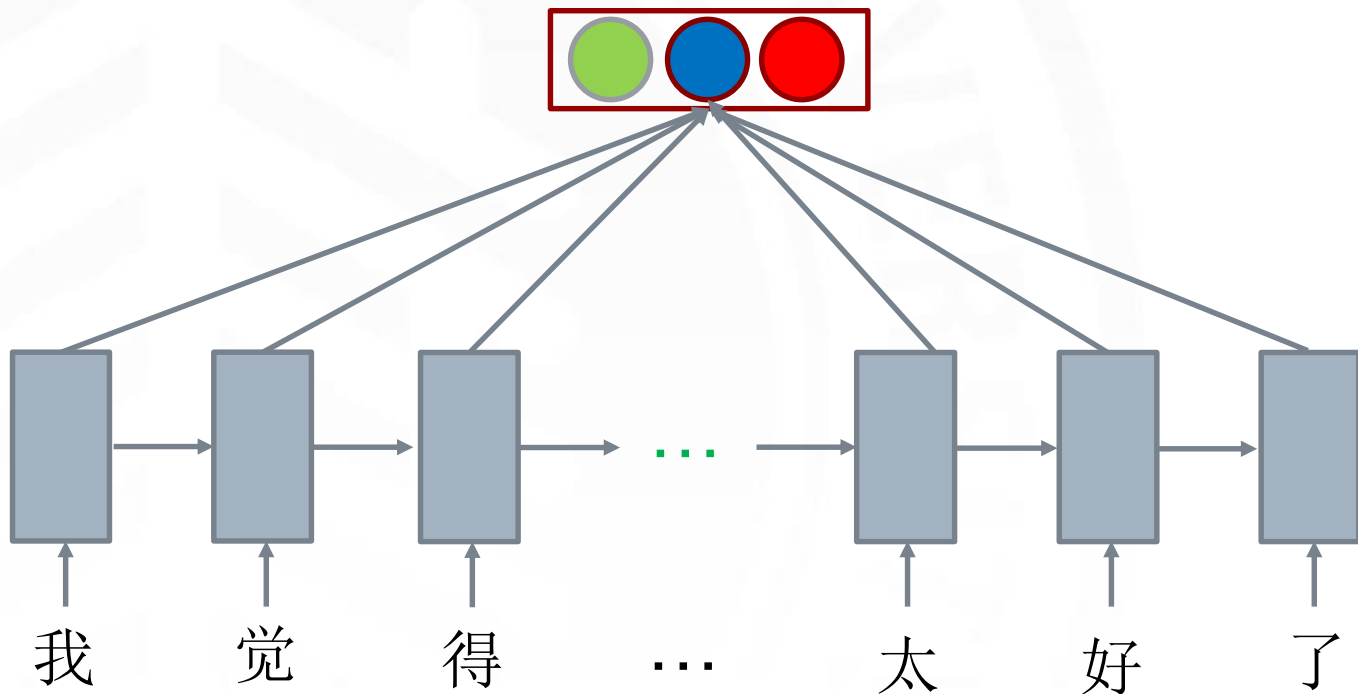
Positive (正面)

这部电影太糟了

Negative (负面)

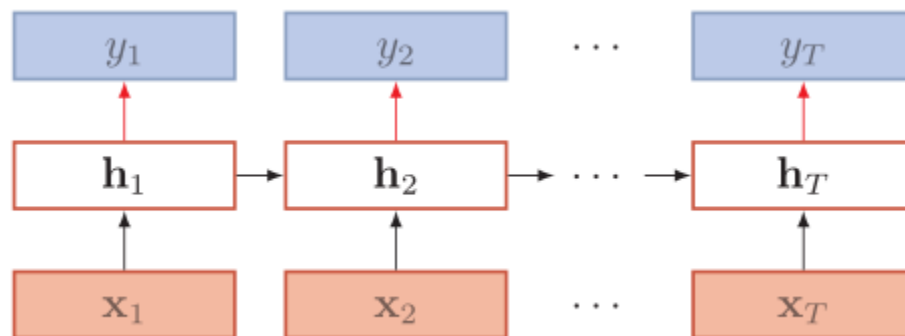
这部电影很棒

Positive (正面)



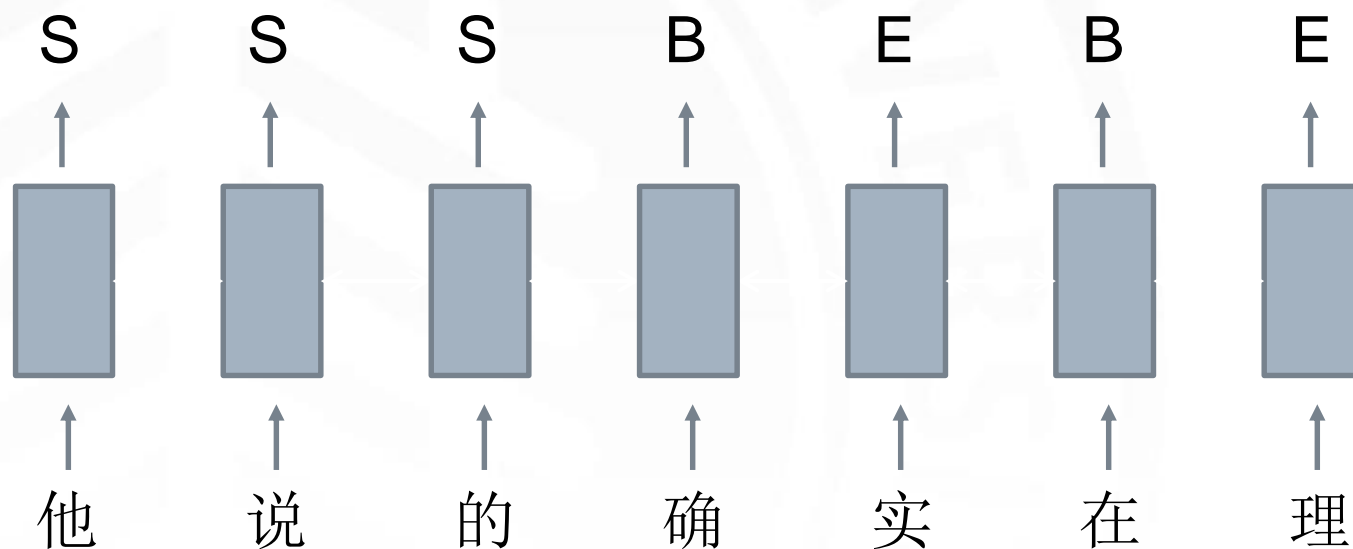
应用到机器学习

■ 同步的序列到序列模式



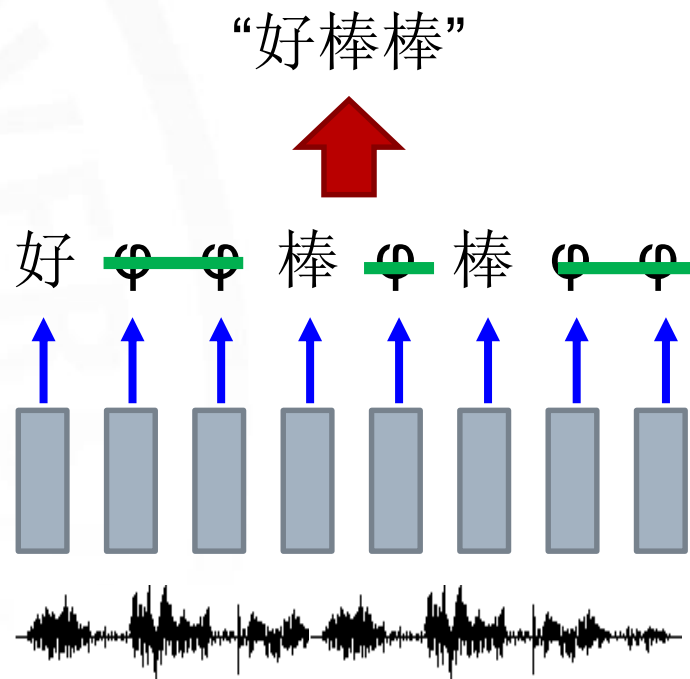
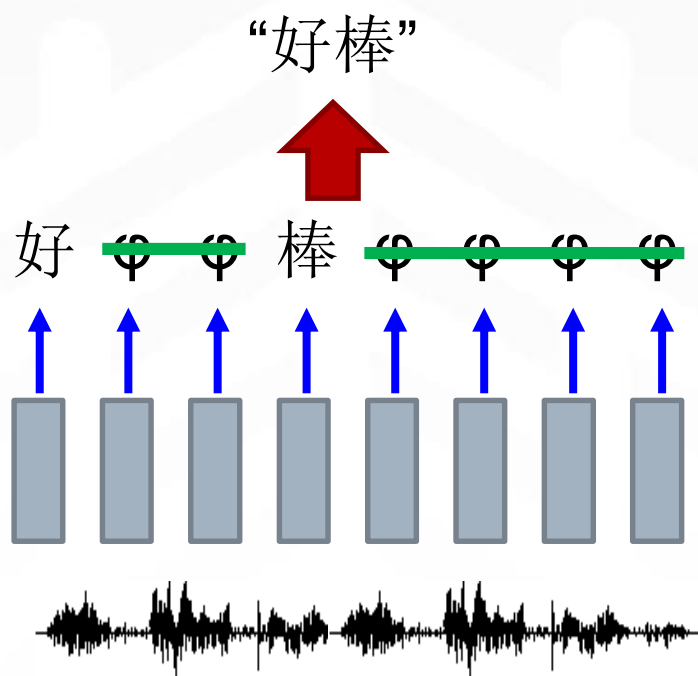
同步的序列到序列模式

■ 中文分词



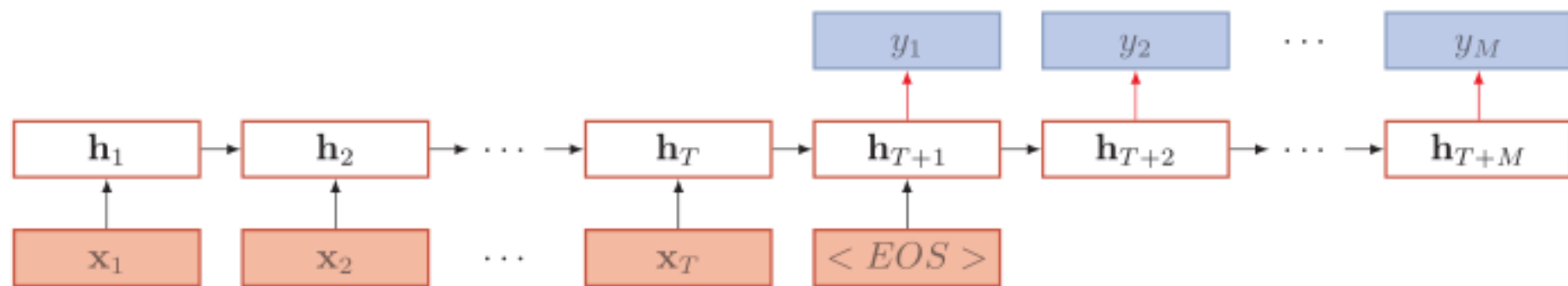
同步的序列到序列模式

- Connectionist Temporal Classification (CTC) [Alex Graves, ICML' 06][Alex Graves, ICML' 14][Haşim Sak, Interspeech' 15][Jie Li, Interspeech' 15][Andrew Senior, ASRU' 15]



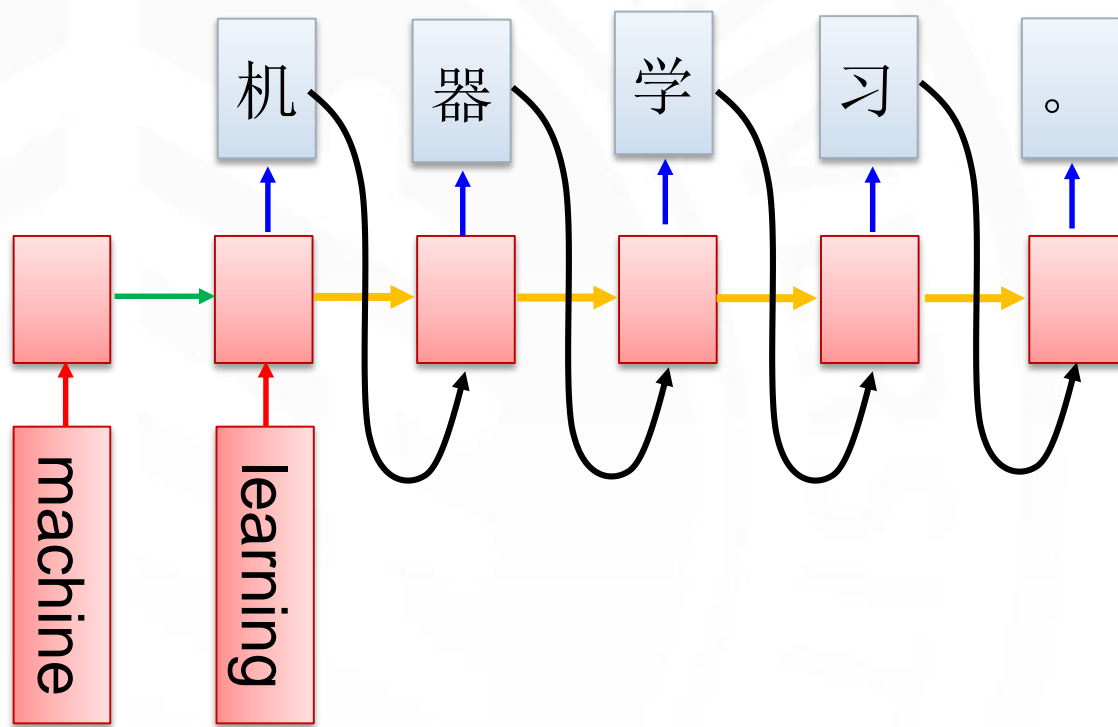
应用到机器学习

■ 异步的序列到序列模式



异步的序列到序列模式

■ 机器翻译



参数学习

■ 机器学习

■ 给定一个训练样本 (\mathbf{x}, \mathbf{y}) ,

□ 其中 $\mathbf{x} = (x_1, \dots, x_T)$ 为长度是 T 的输入序列,

□ $\mathbf{y} = (y_1, \dots, y_T)$ 是长度为 T 的标签序列。

■ 时刻 t 的瞬时损失函数为

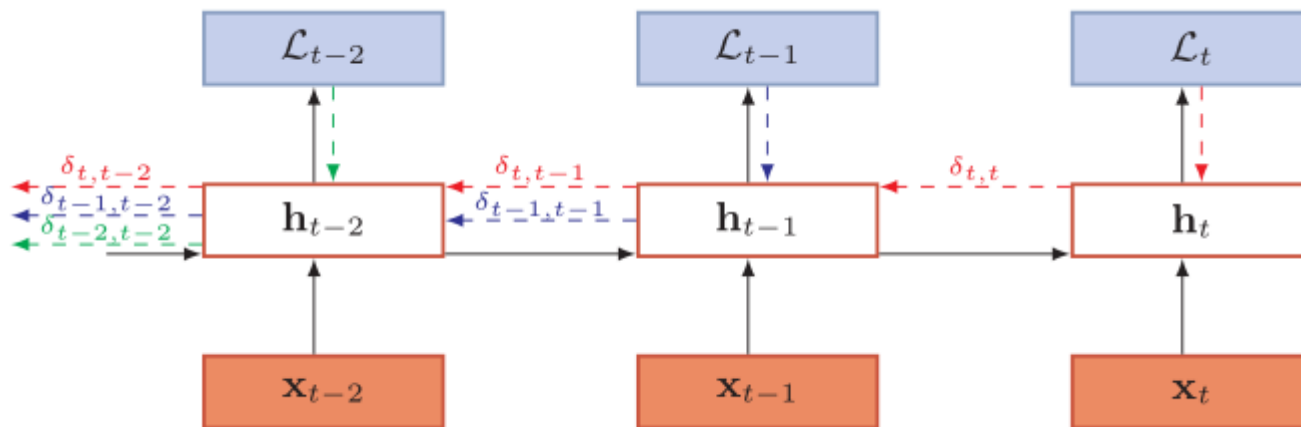
$$\mathcal{L}_t = \mathcal{L}(\mathbf{y}_t, g(\mathbf{h}_t)),$$

■ 总损失函数

$$\mathcal{L} = \sum_{t=1}^T \mathcal{L}_t.$$

梯度

■ 随时间反向传播算法



梯度

链式法则

$$\begin{aligned}\frac{\partial \mathcal{L}_t}{\partial U_{ij}} &= \sum_{k=1}^t \text{tr} \left(\left(\frac{\partial \mathcal{L}_t}{\partial \mathbf{z}_k} \right)^\top \frac{\partial^+ \mathbf{z}_k}{\partial U_{ij}} \right) \\ &= \sum_{k=1}^t \left(\frac{\partial^+ \mathbf{z}_k}{\partial U_{ij}} \right)^\top \frac{\partial \mathcal{L}_t}{\partial \mathbf{z}_k},\end{aligned}$$

$$\frac{\partial^+ \mathbf{z}_k}{\partial U_{ij}} = \begin{bmatrix} 0 \\ \vdots \\ [\mathbf{h}_k]_j \\ \vdots \\ 0 \end{bmatrix} \triangleq \mathbb{I}_i([\mathbf{h}_k]_j),$$

$\delta_{t,k}$ 为第 t 时刻的损失对第 k 步隐藏神经元的净输入 \mathbf{z}_k 的导数

$$\begin{aligned}\delta_{t,k} &= \frac{\partial \mathcal{L}_t}{\partial \mathbf{z}_k} \\ &= \frac{\partial \mathbf{h}_k}{\partial \mathbf{z}_k} \frac{\partial \mathbf{z}_{k+1}}{\partial \mathbf{h}_k} \frac{\partial \mathcal{L}_t}{\partial \mathbf{z}_{k+1}} \\ &= \text{diag}(f'(\mathbf{z}_k)) U^\top \delta_{t,k+1}.\end{aligned}$$

梯度消失/爆炸

■ 梯度

$$\frac{\partial \mathcal{L}}{\partial U} = \sum_{t=1}^T \sum_{k=1}^t \delta_{t,k} \mathbf{h}_k^T.$$

■ 其中

$$\delta_{t,k} = \prod_{i=k}^{t-1} \left(\frac{\mathbf{diag}(f'(\mathbf{z}_i)) U^T}{\lambda} \right) \delta_{t,t}.$$

由于梯度爆炸或消失问题，实际上只能学习到短周期的依赖关系。这就是所谓的长期依赖问题。

长期依赖问题

- 循环神经网络在时间维度上非常深!

- 梯度消失或梯度爆炸

- 如何改进?

- 梯度爆炸问题

- 权重衰减

- 梯度截断

- 梯度消失问题

- 改进模型

长期依赖问题

■ 改进方法

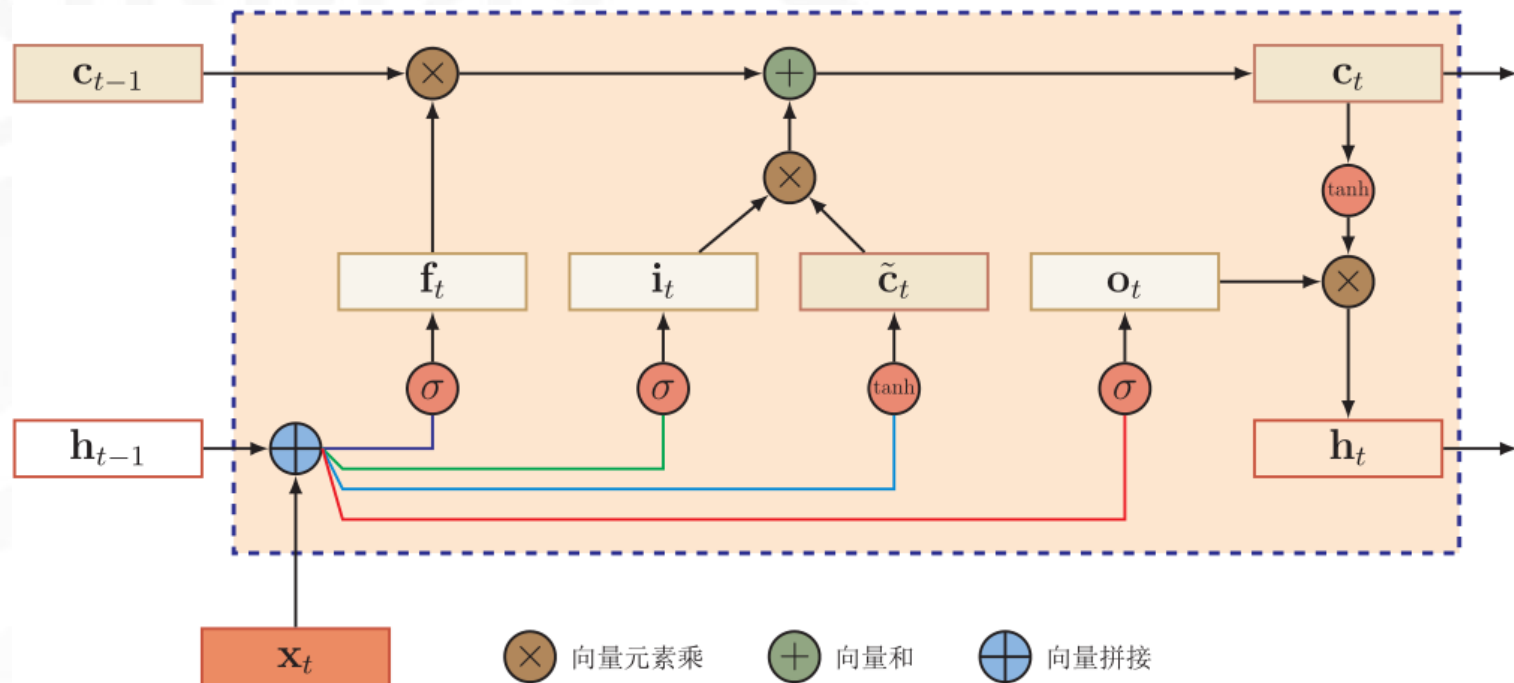
■ 循环边改为线性依赖关系

$$\mathbf{h}_t = \mathbf{h}_{t-1} + g(\mathbf{x}_t; \theta),$$

■ 增加非线性

$$\mathbf{h}_t = \mathbf{h}_{t-1} + g(\mathbf{x}_t, \mathbf{h}_{t-1}; \theta),$$

长短时记忆神经网络：LSTM



$$\mathbf{i}_t = \sigma(W_i \mathbf{x}_t + U_i \mathbf{h}_{t-1} + \mathbf{b}_i),$$

$$\mathbf{f}_t = \sigma(W_f \mathbf{x}_t + U_f \mathbf{h}_{t-1} + \mathbf{b}_f),$$

$$\mathbf{o}_t = \sigma(W_o \mathbf{x}_t + U_o \mathbf{h}_{t-1} + \mathbf{b}_o),$$

$$\tilde{\mathbf{c}}_t = \tanh(W_c \mathbf{x}_t + U_c \mathbf{h}_{t-1} + \mathbf{b}_c)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t,$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t),$$

LSTM的各种变体

■ 没有遗忘门

$$\mathbf{c}_t = \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t.$$

■ 耦合输入门和遗忘门

$$\mathbf{f}_t + \mathbf{i}_t = \mathbf{1}.$$

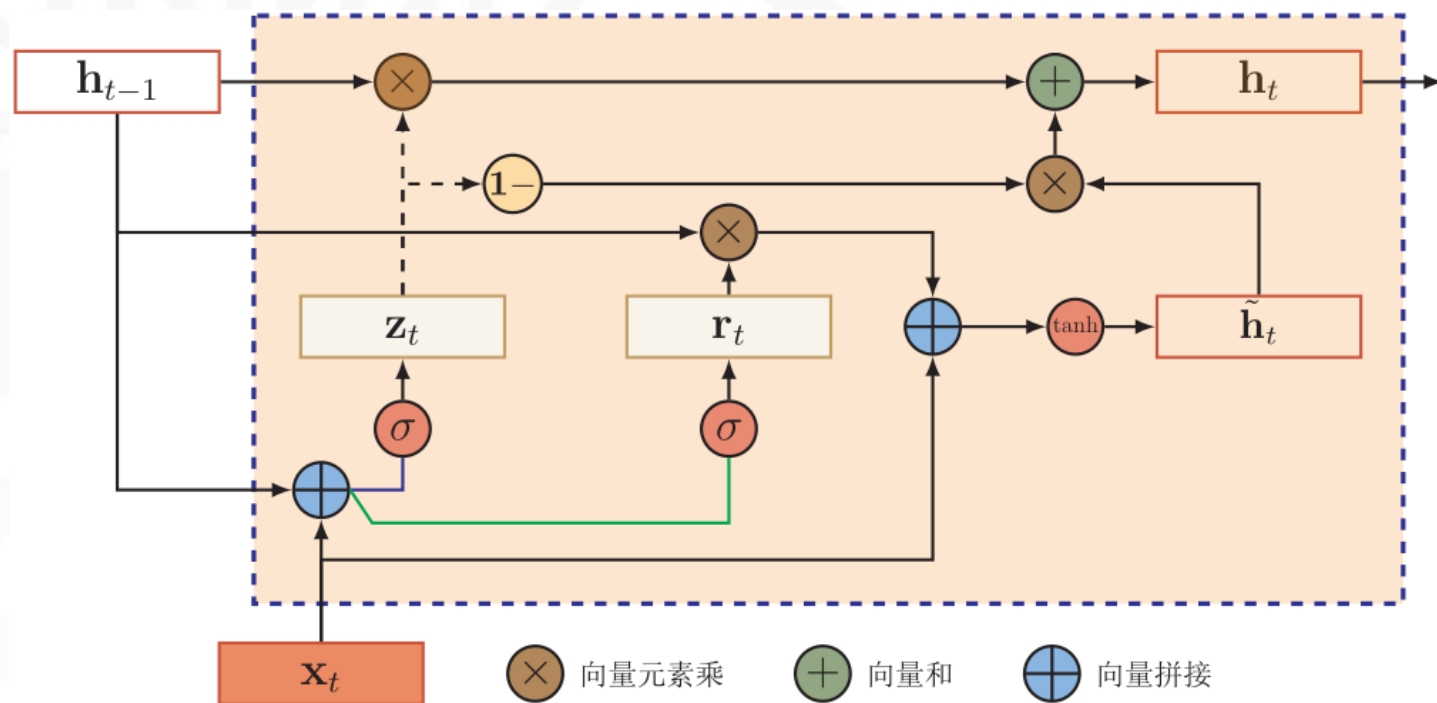
■ peephole连接

$$\mathbf{i}_t = \sigma(W_i \mathbf{x}_t + U_i \mathbf{h}_{t-1} + V_i \mathbf{c}_{t-1} + \mathbf{b}_i),$$

$$\mathbf{f}_t = \sigma(W_f \mathbf{x}_t + U_f \mathbf{h}_{t-1} + V_f \mathbf{c}_{t-1} + \mathbf{b}_f),$$

$$\mathbf{o}_t = \sigma(W_o \mathbf{x}_t + U_o \mathbf{h}_{t-1} + V_o \mathbf{c}_t + \mathbf{b}_o),$$

GRU



重置门

$$\mathbf{r}_t = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r),$$

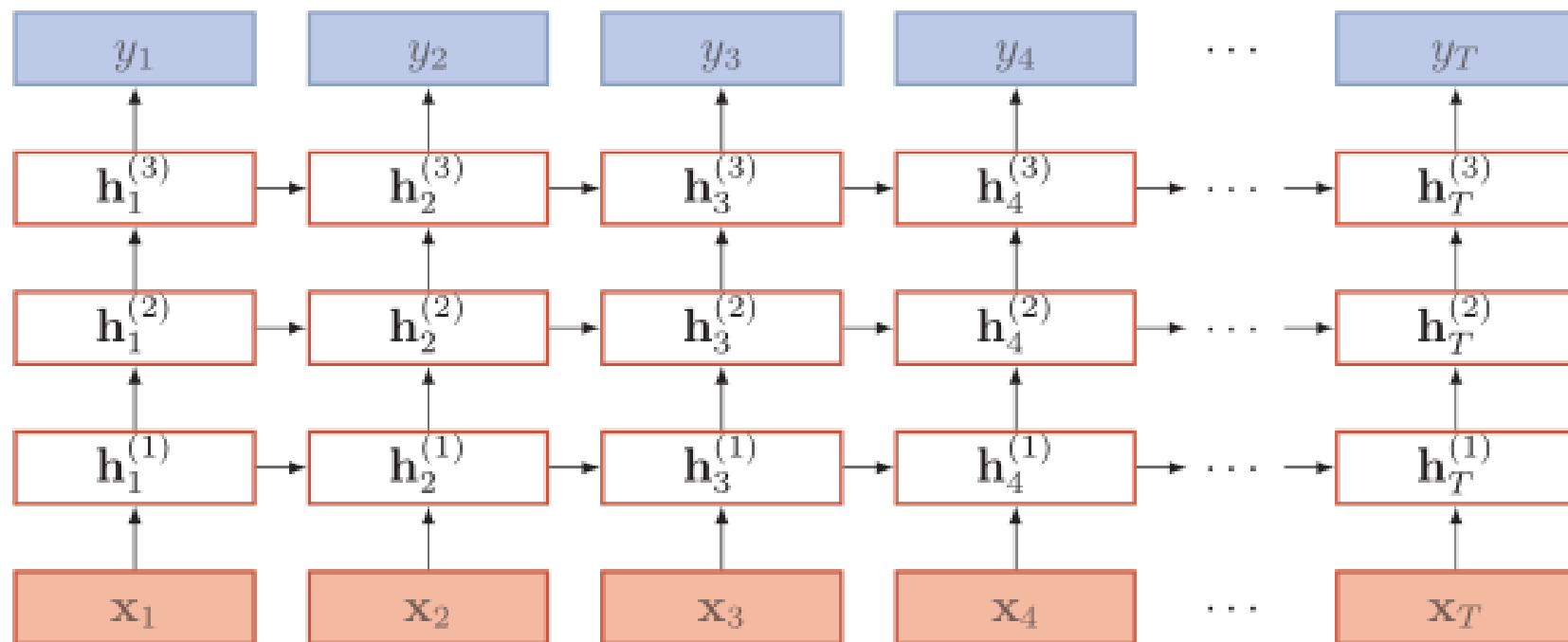
$$\mathbf{z}_t = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z),$$

更新门

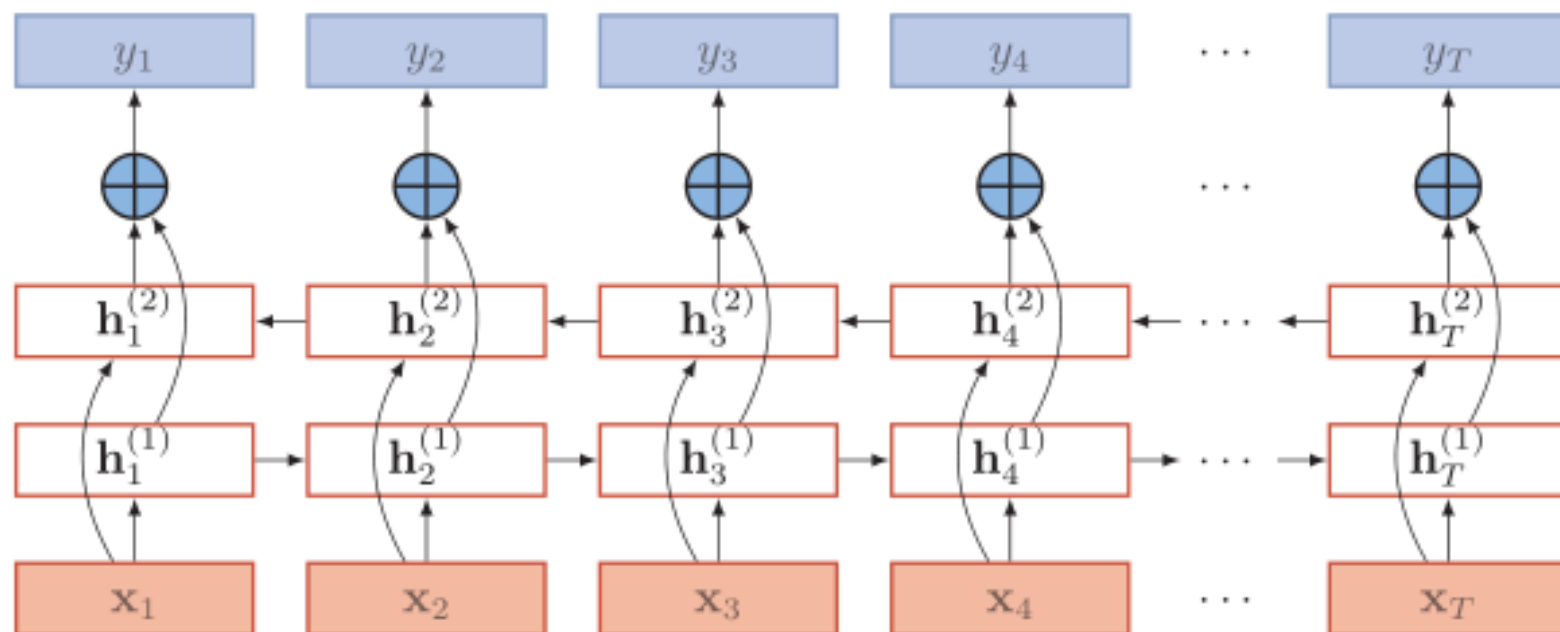
$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}(\mathbf{r}_t \odot \mathbf{h}_{t-1}))$$

$$\mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t,$$

堆叠循环神经网络



双向循环神经网络

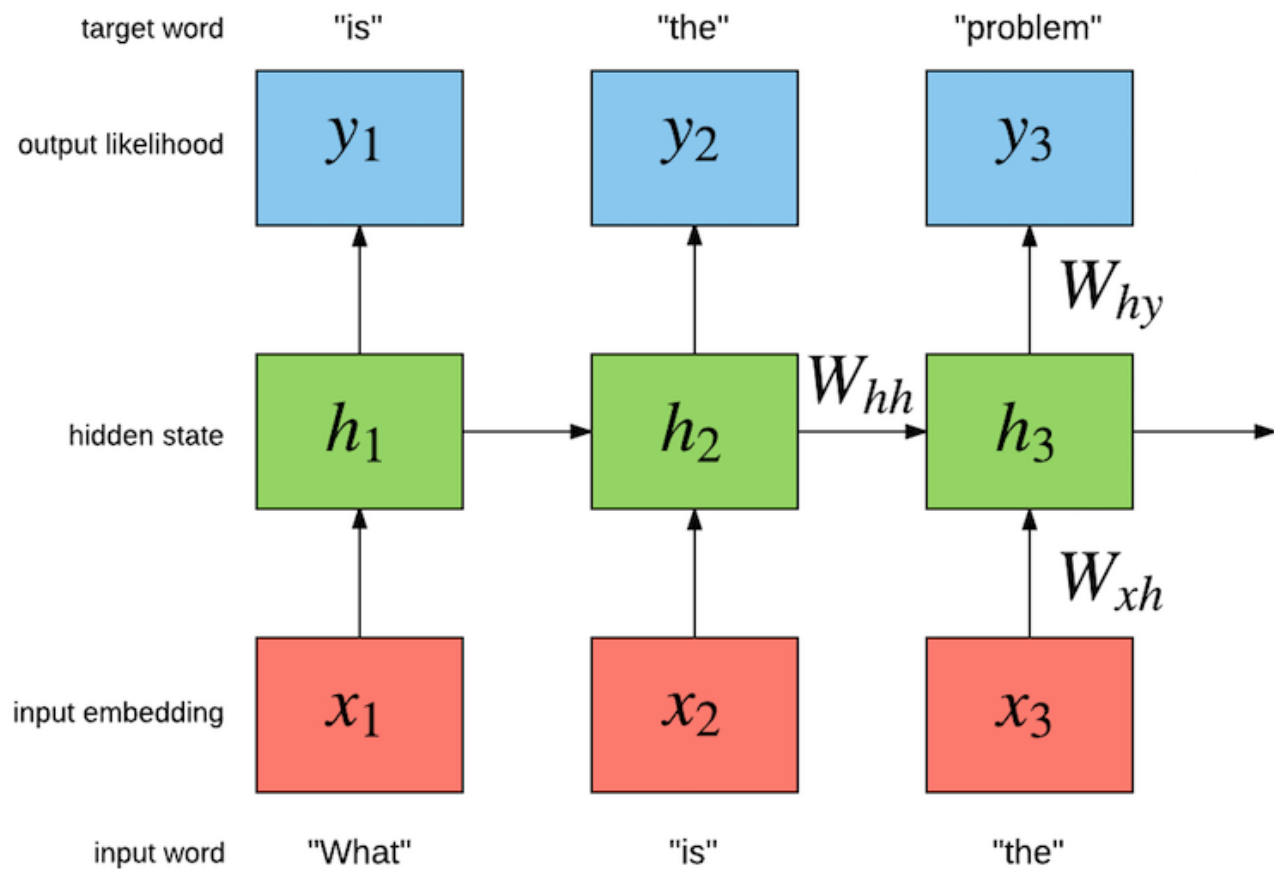




華東師範大學
EAST CHINA NORMAL UNIVERSITY

循环网络应用

语言模型



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

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.

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，

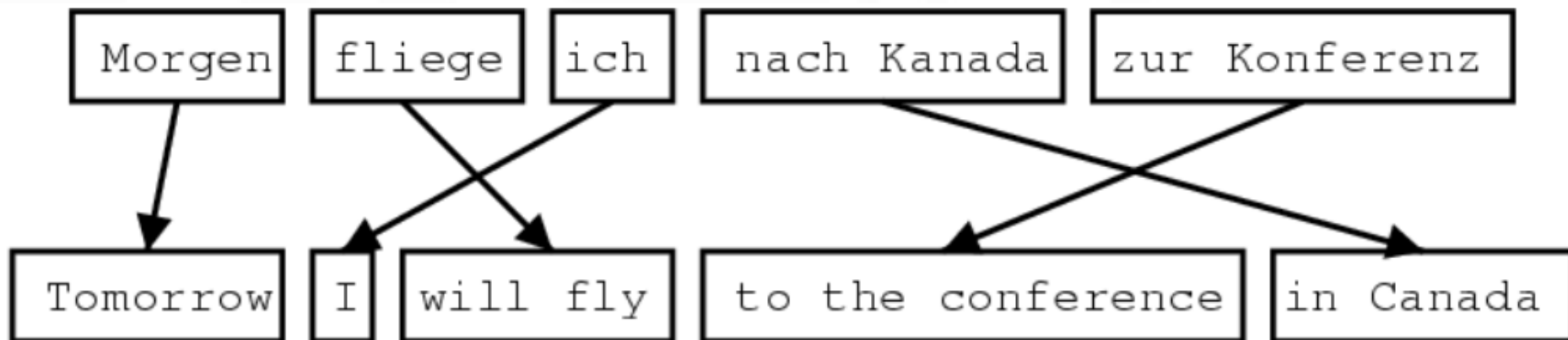
With the east wind comes Spring.

世间何处是前身。

Where on earth do I come from?

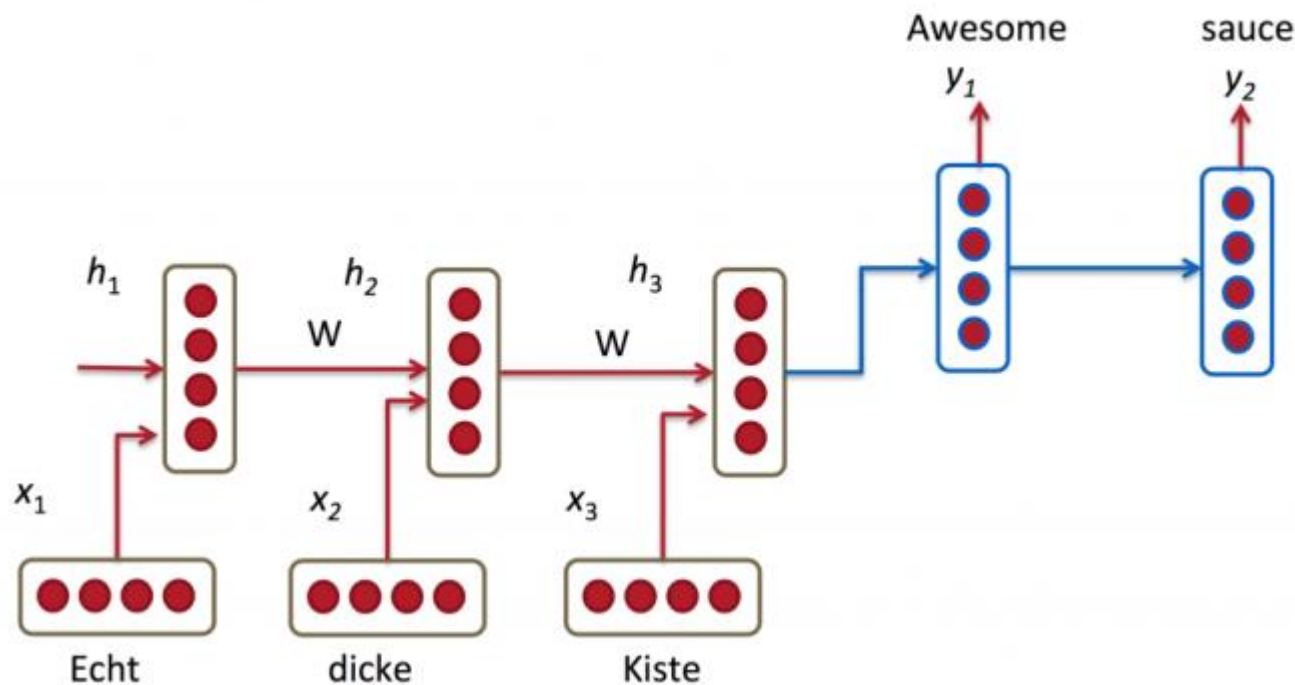
传统统计机器翻译

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

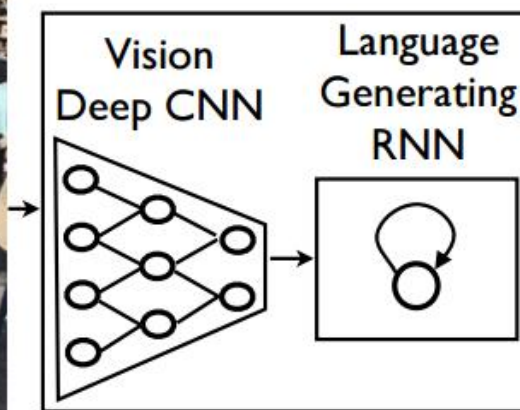


基于序列到序列的机器翻译

- 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

A person riding a motorcycle on a dirt road.



Two dogs play in the grass.



A skateboarder does a trick on a ramp.



A dog is jumping to catch a frisbee.



A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.



A little girl in a pink hat is blowing bubbles.



A refrigerator filled with lots of food and drinks.



A herd of elephants walking across a dry grass field.



A close up of a cat laying on a couch.



A red motorcycle parked on the side of the road.



A yellow school bus parked in a parking lot.



Describes without errors

Describes with minor errors

Somewhat related to the image

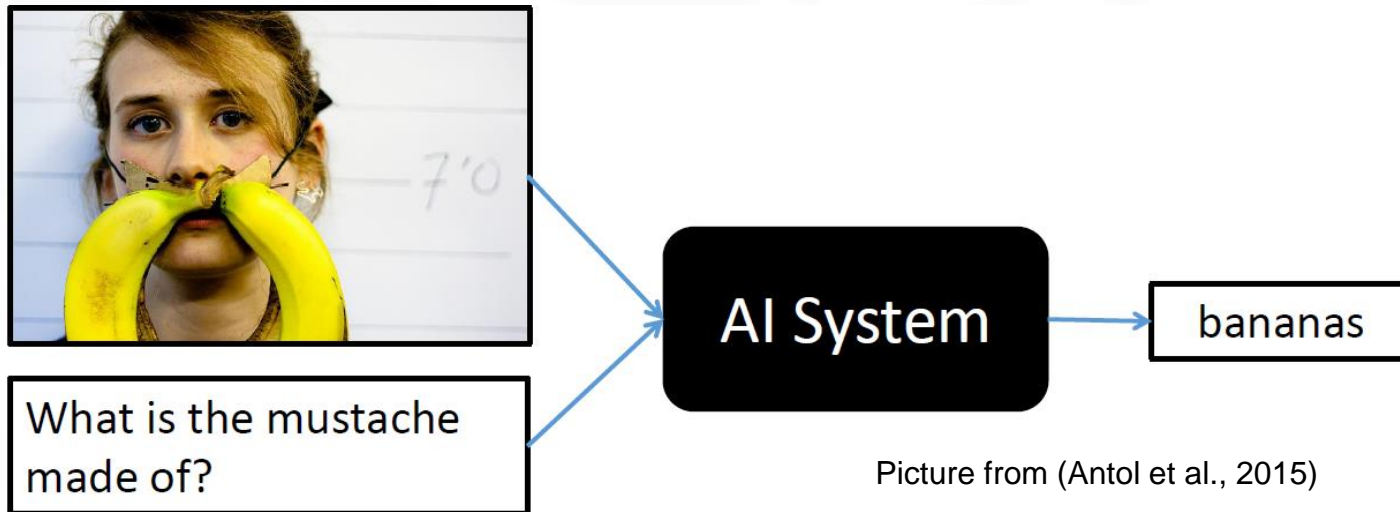
Unrelated to the image

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

Visual Question Answering (VQA)

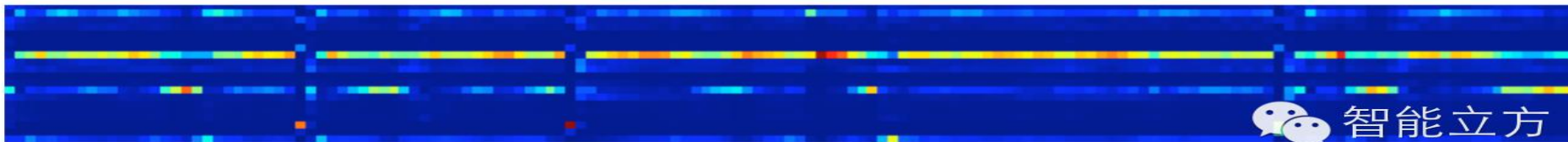
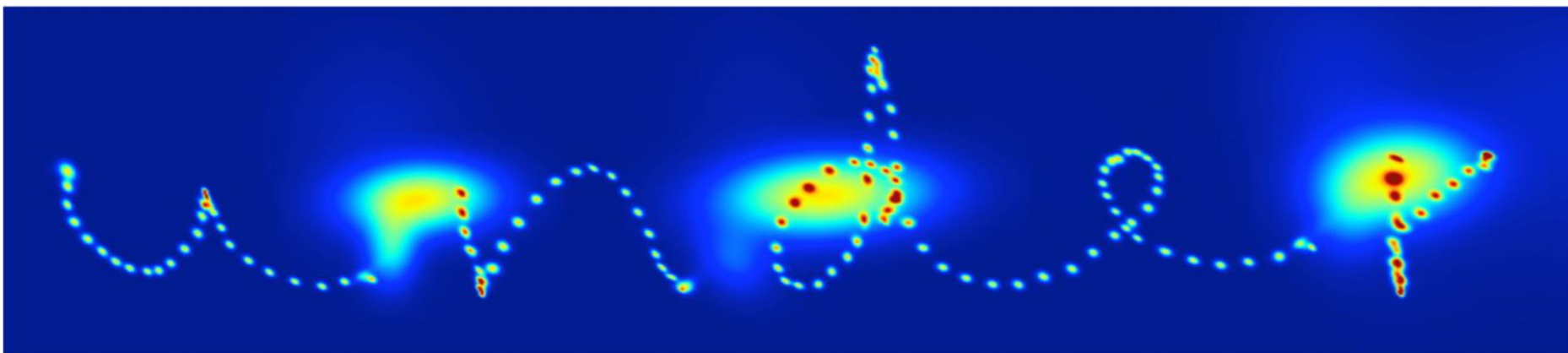
Demo Website

VQA: Given an image and a natural language question about the image, the task is to provide an accurate natural language answer



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



Donald J. Trump @realDonaldTrump · May 30

The people are such a wonderful mistake in decades of my speech at the @nytimes yesterday. Big crowd! #Trump2016 pic.twitter.com/XW060pZbmm

7.2K 26K 97K



Donald J. Trump @realDonaldTrump · May 28

Last night was one of the worst things that Obama was a trillion dollar budget deficit with a single defeat. I won't report the truth and thanks.

50K 28K 106K



Donald J. Trump @realDonaldTrump · May 28

Congratulations to @HuffingtonPost poll with @MittRomney today. Be sure to watch the Trump Tower atrium. It should not be the most beautiful money on me. They will be a big loser and special interest money.

21K 14K 66K



Donald J. Trump @realDonaldTrump · May 28

The Trump Tower atrium is so great honor to be indeceing on chockey supporters at the Miss Universe Pageant. I think it should be dead, finally, billions of incredible!

18K 19K 74K



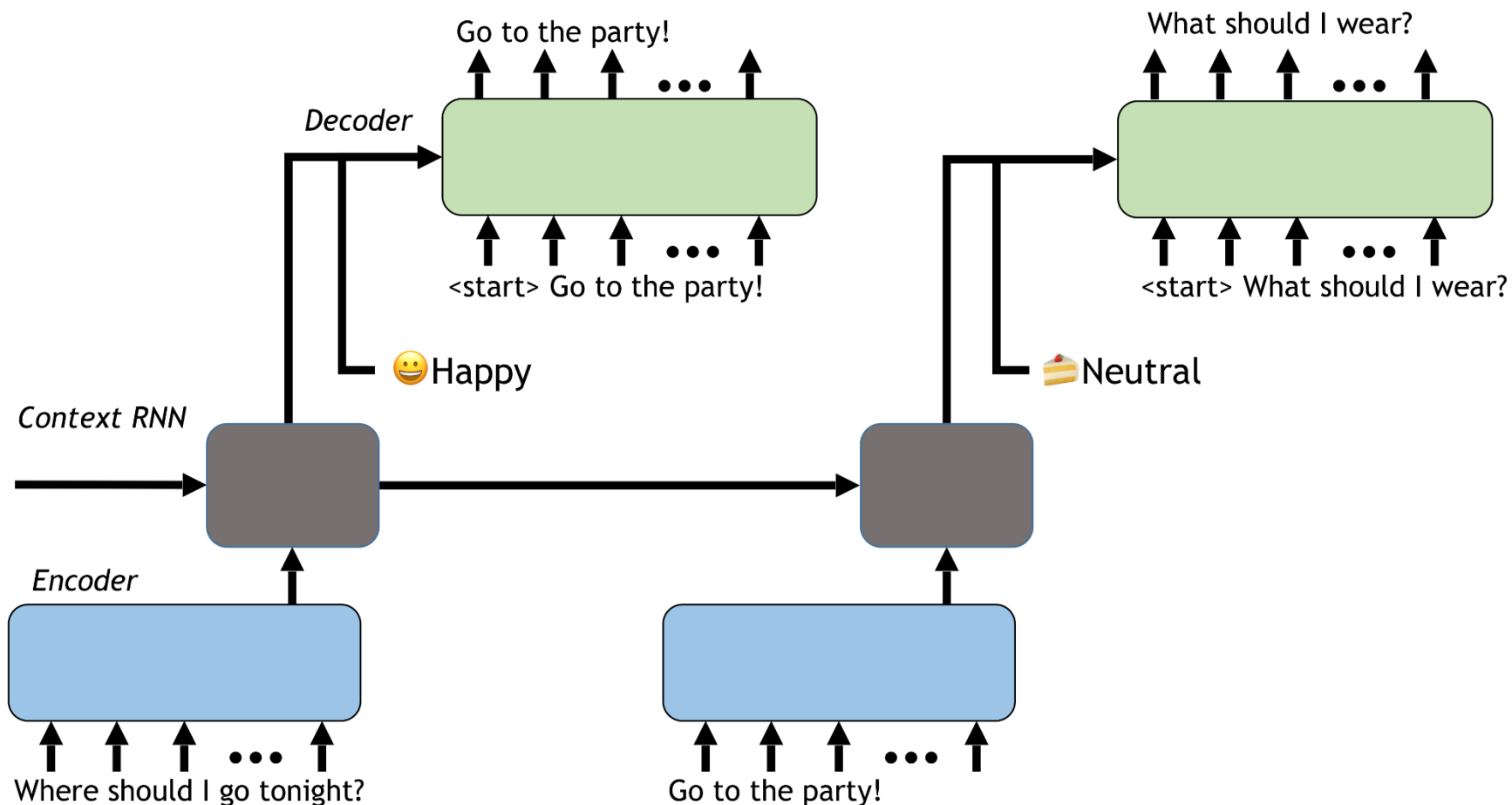
Donald J. Trump @realDonaldTrump · May 28

Why would the fact that I left the Democrats that he has a show that is a complete and money to a bad deal. Stop congratulating the U.S. Starting the bankruptcy proud. What a festing career!

21K 14K 79K

对话系统

<https://github.com/lukalabs/cakechat>



循环神经网络总结

■ 优点：

- 引入记忆

- 图灵等价

■ 缺点：

- 长期依赖问题

- 记忆容量问题

- 并行能力



THE END