

第 9 章 序列标注 – 条件随机场&神经网络

2/2

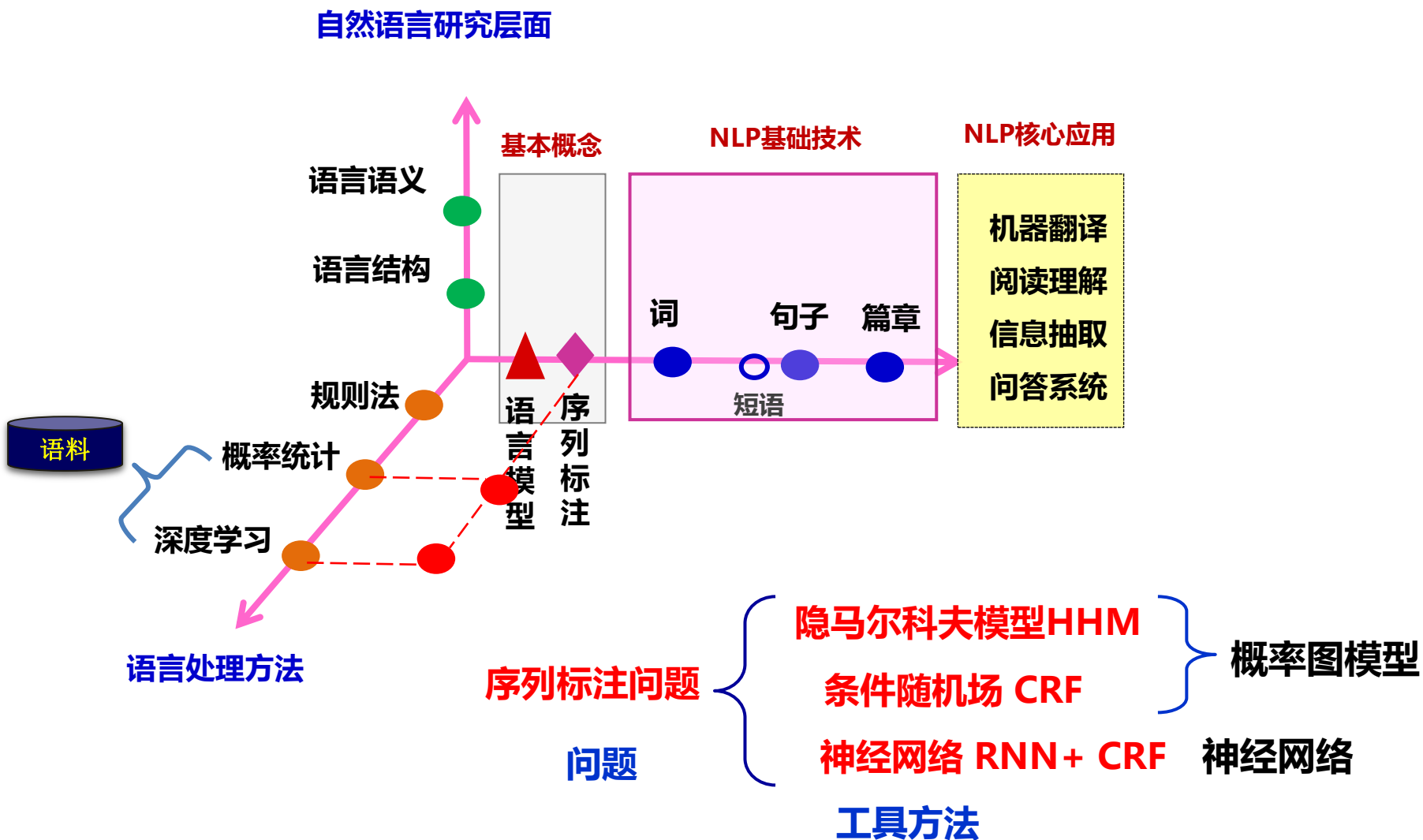
中科院信息工程研究所第二研究室

胡玥

huyue@iie.ac.cn

自然语言处理课程内容及安排

◇ 课程内容：



内 容 提 要

9.1 序列标注问题

9.2 隐马尔科夫模型HMM

9.3 条件随机场 CRF

9.3.1 条件随机场基本概念

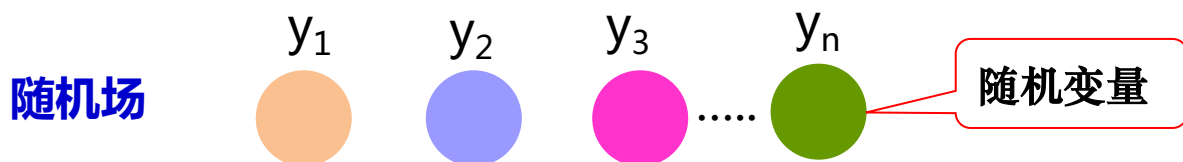
9.3.2 条件随机场应用

9.4 神经网络 RNN+ CRF

9.3.1 条件随机场基本概念

■ 马尔可夫随机场 (Markov Random Field)

- **随机场**：可看成是一组随机变量的集合（这组随机变量对应同一个样本空间）。这些随机变量之间可能有依赖关系，当给每一个位置中随机变量按照某种分布随机赋予相空间的一个值之后，其全体就叫做随机场。

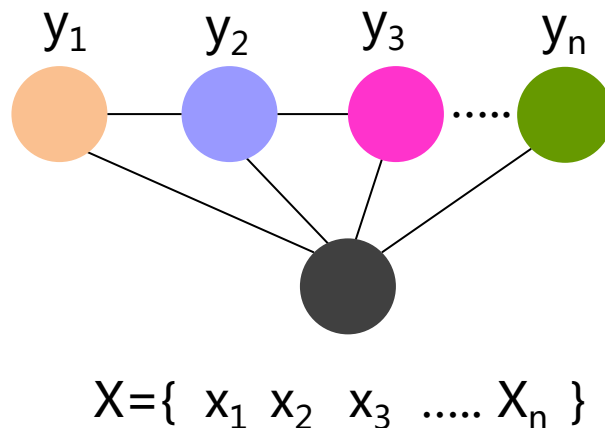


- **马尔可夫性质**：它指的是一个随机变量序列按时间先后关系依次排开的时候，第 $N+1$ 时刻的分布特性，与 N 时刻以前的随机变量的取值无关。
- **马尔可夫随机场 (MRF)**：具有马尔科夫性质的随机场。

9.3.1 条件随机场基本概念

■ 条件随机场 (CRF) :

如果给定的MRF中每个随机变量下面还有观察值，要确定的是给定观察集合下，这个MRF的分布，也就是条件分布，那么这个MRF就称为CRF(Conditional Random Field)。



链式条件随机场结构图

9.3.1 条件随机场基本概念

■ 线性条件随机场的基本形式：

$$p(y|x, \lambda) = \frac{1}{Z(x)} \exp\left(\sum_{ji} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{ki} \mu_k s_k(y_i, x, i)\right)$$

$$\text{其中：} \quad Z(x) = \sum_y \exp\left(\sum_{ji} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{ki} \mu_k s_k(y_i, x, i)\right)$$

输入： $x = \{x_1, x_2, \dots, x_n\}$ （观测序列）

输出： $y = \{y_1, y_2, \dots, y_n\}$ （标记序列）

参数： $\lambda_j \quad \mu_k$

式中 t_i 与 s_k 为特征函数， t_i 考虑当前位置 i 对应的标签值 Y_i 与前一位置对应的标签值 Y_{i-1} 对后验概率的影响，称为**边特征函数**； s_k 只考虑当前位置 i 对应的标签值对后验概率的影响，称为**节点特征函数**。 λ_j 与 μ_k 为特征函数对应的权值。 $Z(X)$ 是归一化函数，用于将计算得到的值归一化为概率形式。

9.3.1 条件随机场基本概念

模型中特征函数的形式定义：

$t_j(y_{i-1}, y_i, x, i)$ 是定义在边上的特征函数, 称转移特征

$$t_j(y_{i-1}, y_i, x, i) = \begin{cases} 1 & \text{满足条件} \\ 0 & \text{否则} \end{cases}$$

$s_k(y_i, x, i)$ 是定义在结点上的特征函数, 称状态特征, 依赖当前位置

$$s_k(y_i, x, i) = \begin{cases} 1 & \text{满足条件} \\ 0 & \text{否则} \end{cases}$$

条件随机场由特征函数 t_j, s_k 和对应的权值 λ_j, μ_k 确定.

内 容 提 要

9.1 序列标注问题

9.2 隐马尔科夫模型HMM

9.3 条件随机场 CRF

9.3.1 条件随机场基本概念

9.3.2 条件随机场应用

9.4 神经网络 RNN+ CRF

9.3.2 条件随机场应用

- 命名实体识别（人名识别）

目标：将给定的输入序列中的人名识别出来

如：输入序列： 新任总裁**罗建国**宣布了对部门经理**邓奇**的任免通知

↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓

输出序列： 0 0 0 0 **B I E** 0 0 0 0 0 0 0 0 **B E** 0 0 0 0 0

{ B I E O } 或 { B I O }

B - 词首
I - 词中
E - 词尾
O - 单个词

可利用开源工具包实现

目前基于CRF 的主要系统有： CRF、FlexCRF、**CRF++**

优点：使用方便

缺点：训练代价大、复杂度高

9.3.2 条件随机场应用

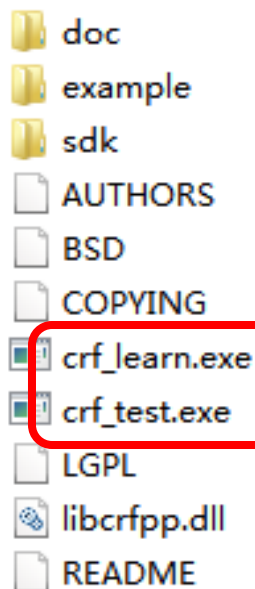
CRF++工具包介绍

CRF++ 是一个可用于分词/连续数据标注的简单、可定制并且开源的条件随机场(CRFs)工具。用于自然语言处理的实体识别, 信息抽取和语块分析等方面。

下载地址 : <http://pan.baidu.com/s/1kT2Jmhl>

9.3.2 条件随机场应用

- 工具包文件



crf_learn.exe: CRF++ 的训练程序
crf_test.exe: CRF++ 的预测程序

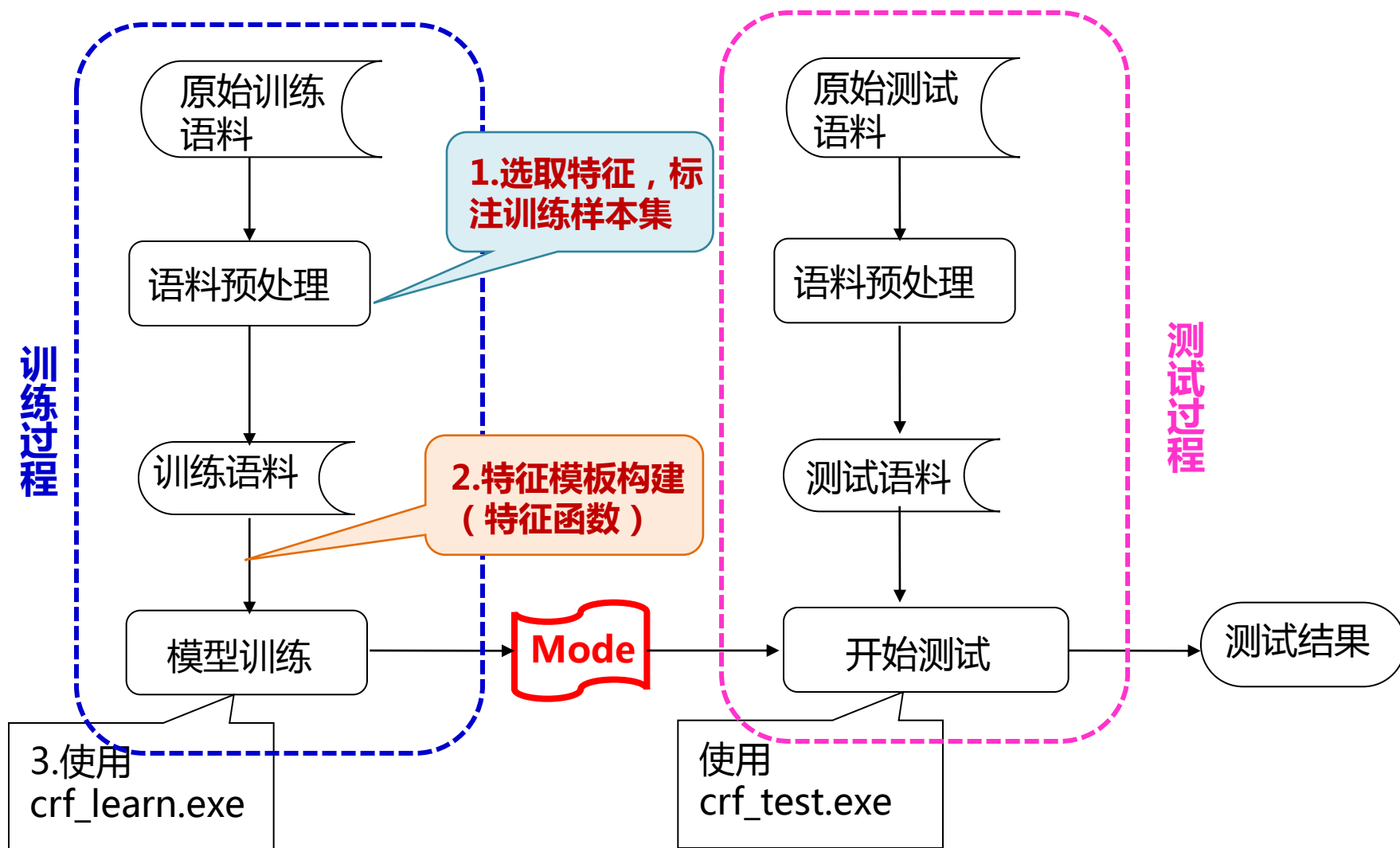
example 文件夹：有四个任务的训练数据、测试数据和模板文件。

crf_learn.exe：CRF++ 的训练程序。

crf_test.exe：CRF++ 的测试程序

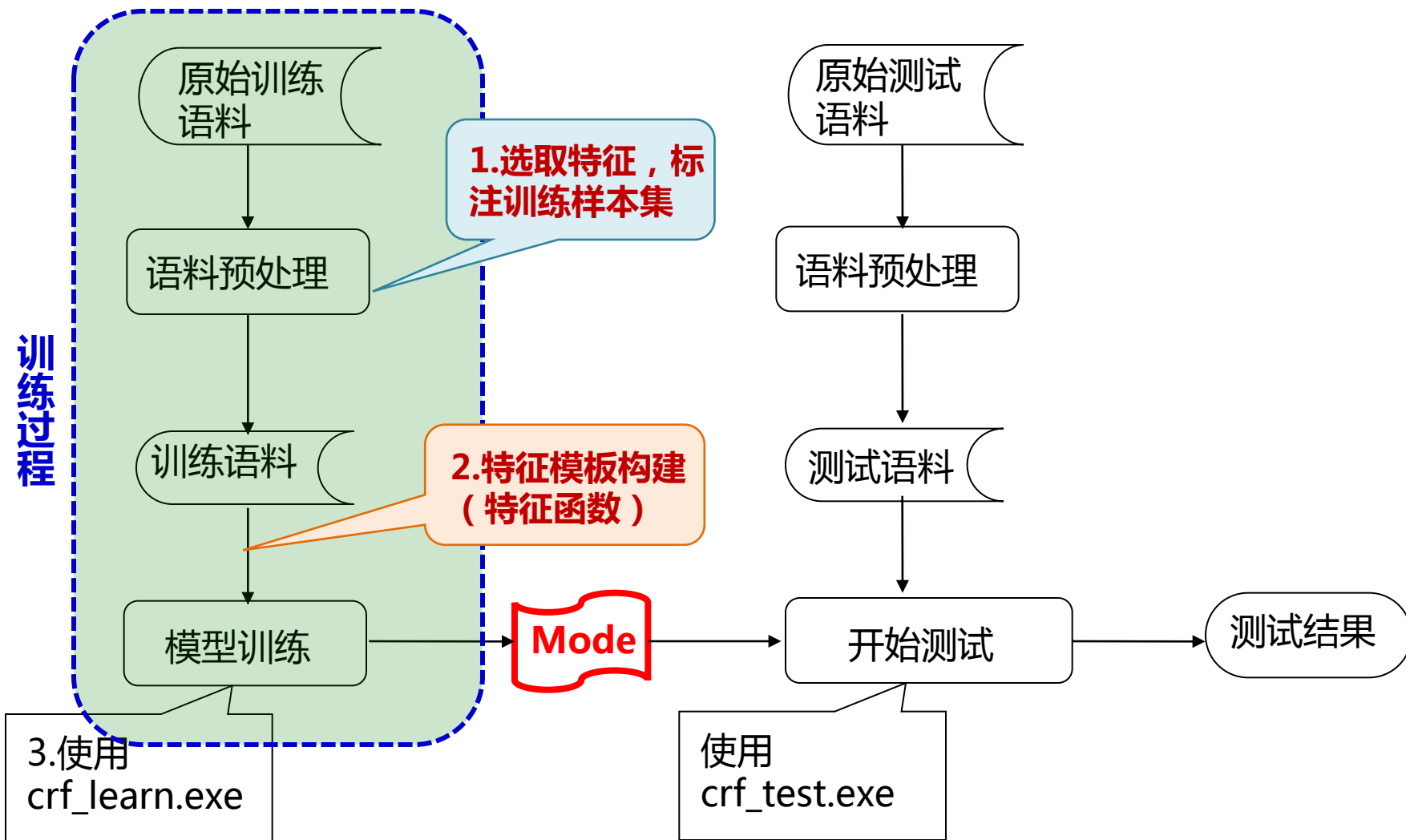
9.3.2 条件随机场应用

工具包使用结构图



9.3.2 条件随机场应用

工具包使用结构图



9.3.2 条件随机场应用

1. 语料处理

目标：构造训练样本集

任务：（1）对原始文本进行分词、词性标注等处理

（2）选择选择能表达随机过程统计特征的特征集合对语料进行标注

（3）标注分类标签

关键任务： 特征选择

9.3.2 条件随机场应用

◆ 特征选取问题

- 所有知识可信吗？所有知识都有用吗？
- 太多知识怎么办？

在所有的特征中，选择最有代表性的特征，构造约束集合

◆ 特征选取算法

1. 增量式特征选择算法

每一步用“贪心”原则：选“现在看来”最有用那个特征将其加入后熵减少最多

2. 基于频数阈值的特征选择算法

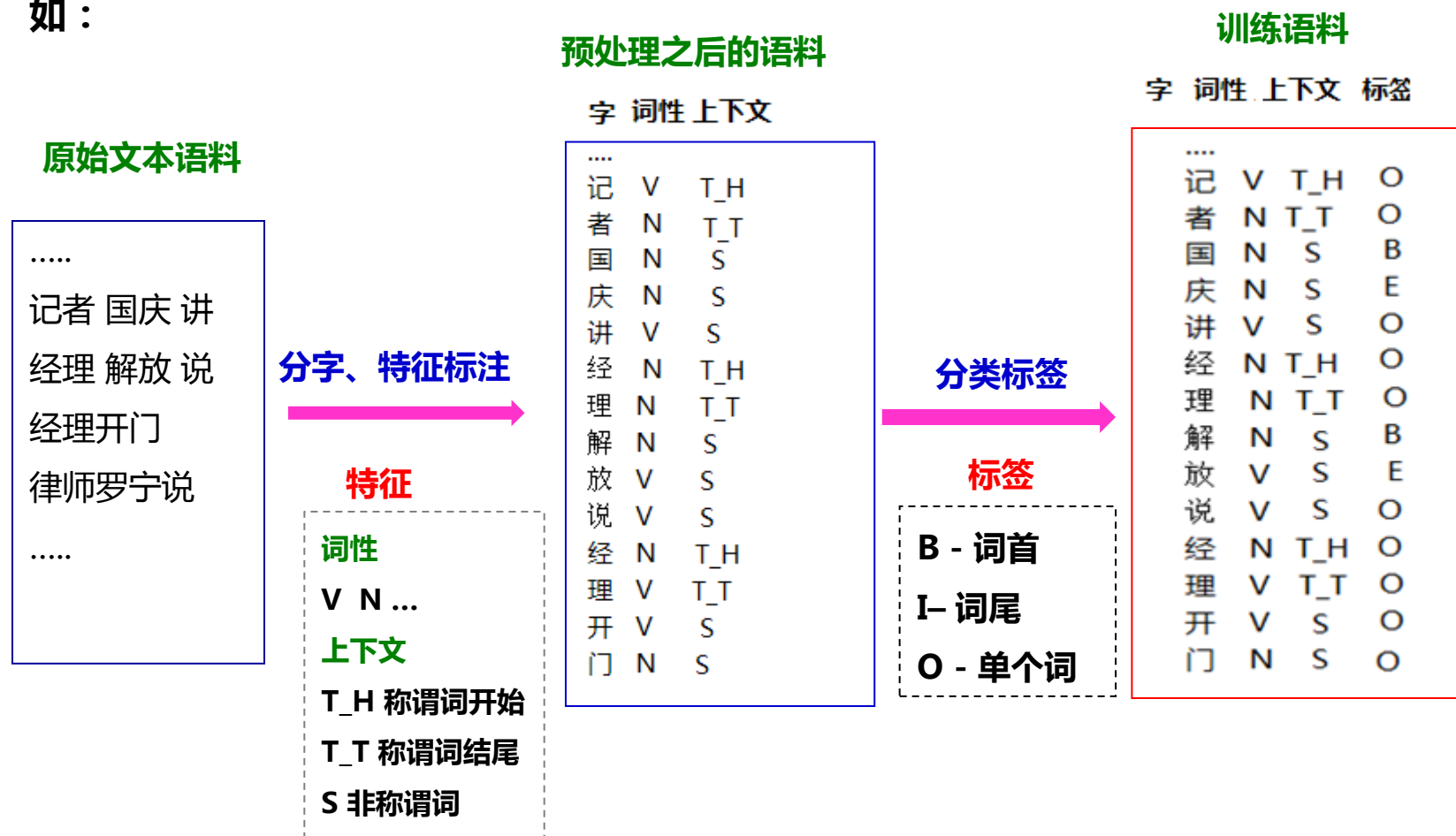
某个特征当其出现的次数大于一定的值时才将其加入

.....

在NLP里面：有很多种上下文信息可以作为特征

9.3.2 条件随机场应用

如：



9.3.2 条件随机场应用

2. 特征模板构建（关键）

特征模板作用：自动生成特征函数

模板文件中每一行代表一个 template，每一个 template 中，利用专门的宏 `%x[row,col]` 来确定输入数据中的一个 token，CRF++ 通过宏产生相应的特征函数集合。

CRF++ 中将特征分为两种类型，一种是 Unigram，以 “U” 开头，另一种是 Bigram， “B” 开头一般用以 “U” 开头

如：

U00:%x[-2,0]

U01:%x[-1,0]

U02:%x[0,0]

U03:%x[1,0]

U04:%x[2,0]

U05:%x[-1,0]/%x[0,0]

U06:%x[0,0]/%x[1,0]

U07:%x[0,1]

9.3.2 条件随机场应用

%x[row,col]含义:

row表示当前TOKEN的前后行 ;

col表示其在样本中的列 (特征)

如:

U00:%x[-2,0]

U01:%x[-1,0]

U02:%x[0,0]

U03:%x[1,0]

U04:%x[2,0]

U05:%x[-1,0]/%x[0,0]

U06:%x[0,0]/%x[1,0]

U07:%x[0,1]

9.3.2 条件随机场应用

如:

| | 字 | 词性 | 上下文 | 标签 |
|--------------------|------|----|-----|----|
| | | | | |
| | 记 | V | T_H | O |
| | 者 | N | T_T | O |
| | 国 | N | S | B |
| | 庆 | N | S | E |
| | 讲 | V | S | O |
| CURRENT TOKEN → | 经 | N | T_H | O |
| | 理 | N | T_T | O |
| | 解 | N | S | B |
| | 放 | V | S | E |

| | |
|----------------------|-----|
| U01:%x[-1,0] | 讲 |
| U02:%x[0,0] | 经 |
| U03:%x[1,0] | 理 |
| U05:%x[-1,0]/%x[0,0] | 讲经 |
| U06:%x[0,0]/%x[1,0] | 经理 |
| U07:%x[0,1] | N |
| U08:%x[0,2] | T_H |

9.3.2 条件随机场应用

特征模板构建：

| | | |
|-------|------|---|
| TEM-3 | 原子模板 | $word_{-1}, word_0, word_1$ |
| | 多元模板 | $word_{-1}word_0, word_0word_1, word_{-1}word_1$ |
| TEM-5 | 原子模板 | $word_i, i = -2, -1, 0, 1, 2$ |
| | 多元模板 | $word_iword_{i+1}, i = -2, -1, 0, 1$ $word_{i-1}word_iword_{i+1}, i = -1, 0, 1$ |
| TEM-7 | 原子模板 | $word_i, i = -3, -2, -1, 0, 1, 2, 3$ |
| | 多元模板 | $word_iword_{i+1}, i = -3, -2, -1, 0, 1,$ $word_{i-2}word_{i-1}word_i, i = 1, 2$ $word_{i-3}word_{i-2}word_{i-1}word_i, i = 0, 1$ |

窗口值：上下文信息的范围

原子特征：指仅适用单特征建立的模板（关注的对象是当前词）

多元特征：对原子模板中单特征的有效组合

9.3.2 条件随机场应用

特征模板作用： 自动生成特征函数

$$s_k(y_i, x, i)$$


如：

U01:%x[-1,0] { fun₁₁ =if(output=B and feature="U01:讲") return 1 else return 0
fun₁₂ =if(output=I and feature="U01:讲") return 1 else return 0
fun₁₃ =if(output=O and feature="U01:讲") return 1 else return 0
U02:%x[0,0]

$$t_j(y_{i-1}, y_i, x, i)$$

U03:%x[1,0]
U05:%x[-1,0]/%x[0,0] { fun_{m1} = if(y_{i-1}=O , y_i = O and x_i="讲") return 1 else return 0
.....
U06:%x[0,0]/%x[1,0]
U07:%x[0,1]

CURRENT
TOKEN



| | | | |
|------|---|-----|---|
| | V | T_H | O |
| 记 | N | T_T | O |
| 者 | N | S | B |
| 国 | N | S | E |
| 庆 | N | S | O |
| 讲 | N | T_H | O |
| 经 | N | T_T | O |
| 理 | N | S | B |
| 解 | V | S | E |
| 放 | | | |

9.3.2 条件随机场应用

特征模板作用： 自动生成特征函数

如: U01:%x[-1,0]

U02:%x[0,0]

U03:%x[1,0]

U05:%x[-1,0]/%x[0,0]

U06:%x[0,0]/%x[1,0]

U07:%x[0,1]

$s_k(y_i, x, i)$

fun₅₁=if(output=B and feature="U05:讲/经") return 1

fun₅₂=if(output=I and feature="U05:讲/经") return 1

fun₅₃=if(output=O and feature="U05:讲/经") return 1

$t_j(y_{i-1}, y_i, x, i)$

字 词性 上下文 标签

.....

CURRENT
TOKEN

| | | | |
|------|---|-----|---|
| | | | |
| 记 | V | T_H | O |
| 者 | N | T_T | O |
| 国 | N | S | B |
| 庆 | N | S | E |
| 讲 | V | S | O |
| 经 | N | T_H | O |
| 理 | N | T_T | O |
| 解 | N | S | B |
| 放 | V | S | E |

9.3.2 条件随机场应用

本例中特征模板构建：

训练语料

字 词性 上下文 标签

| | | | |
|------|---|-----|---|
| | | | |
| 记 | V | T_H | O |
| 者 | N | T_T | O |
| 国 | N | S | B |
| 庆 | N | S | E |
| 讲 | V | S | O |
| 经 | N | T_H | O |
| 理 | N | T_T | O |
| 解 | N | S | B |
| 放 | V | S | E |
| 说 | V | S | O |
| 经 | N | T_H | O |
| 理 | V | T_T | O |
| 开 | V | S | O |
| 门 | N | S | O |

特征模板：

f1: U01:%x[-1,2] 前一词的上下文

f2 : U02:%x[1,1] 后一词的词性

生成的特征函数？

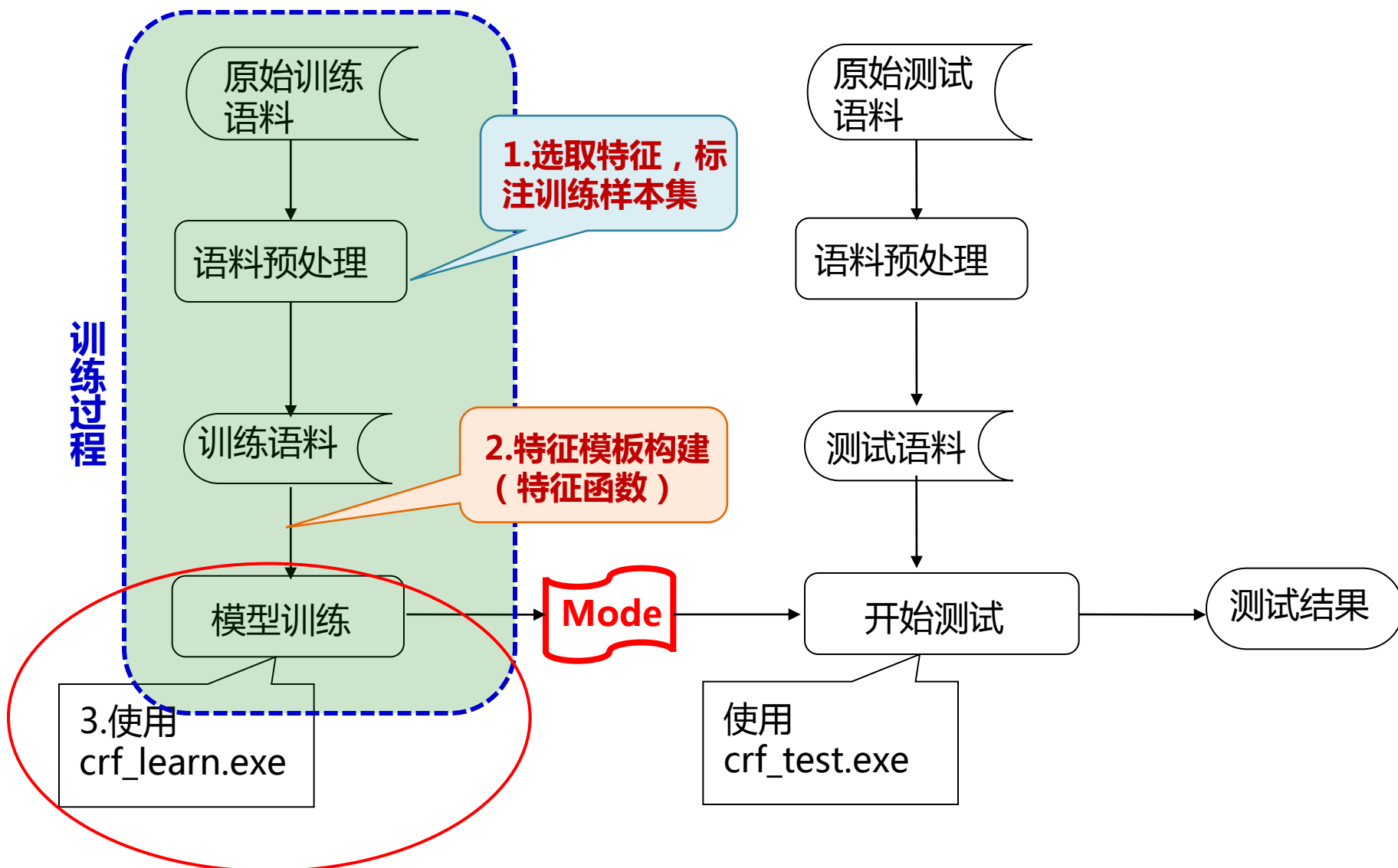
思考：

是否可写成以下模板？

f3 : U03:%x[-1,2] %x[1,1]

9.3.2 条件随机场应用

工具包使用结构图



9.3.2 条件随机场应用

3.使用crf_learn.exe

生成训练好的CRF模型：

$$p(y|x, \lambda) = \frac{1}{Z(x)} \exp\left(\sum_{ji} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{ki} \mu_k s_k(y_i, x, i) \right)$$

其中：

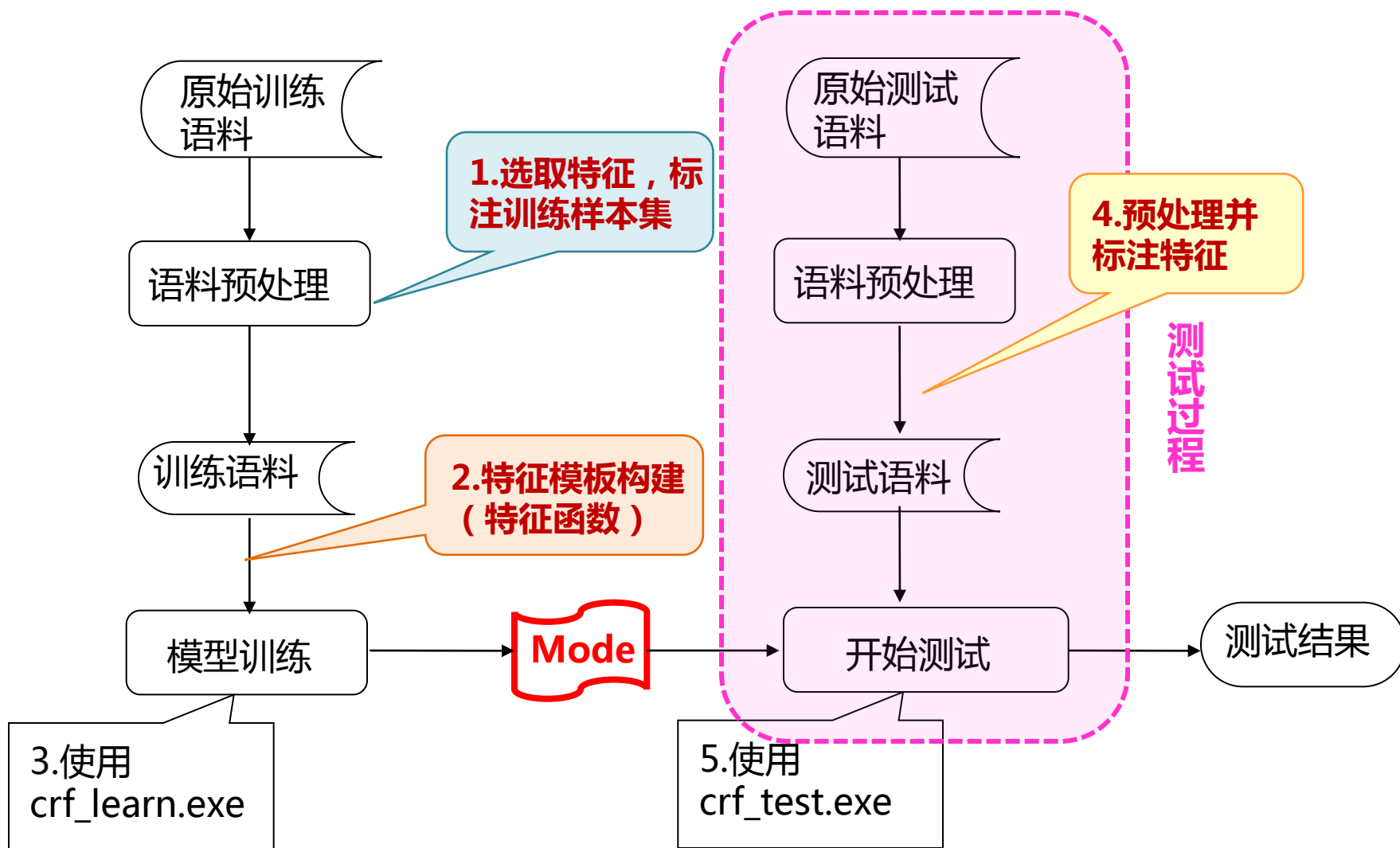
$$Z(x) = \sum_y \exp\left(\sum_{ji} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{ki} \mu_k s_k(y_i, x, i) \right)$$

确定参数

λ_j, μ_k

9.3.2 条件随机场应用

工具包使用结构图



9.3.2 条件随机场应用

4. 预处理并标注特征

原始预测文本

律师罗宁说
.....

分字、特征标注

特征

词性

V N ...

上下文

T_H 称谓词开始

T_T 称谓词结尾

S 非称谓词

预处理后预测文本

字 词性 上下文

| | | |
|-----|---|-----|
| 律 | N | T_H |
| 师 | N | T_T |
| 罗 | N | S |
| 宁 | N | S |
| 读 | V | S |
| ... | | |

预测结果

字 词性 上下文 结果

| | | | |
|-----|---|-----|---|
| 律 | N | T_H | O |
| 师 | N | T_T | O |
| 罗 | N | S | B |
| 宁 | N | S | E |
| 读 | V | S | O |
| ... | | | |

5. 使用 crf_test.exe

9.3.2 条件随机场应用

条件随机场理论 (CRFs) 可以用于 序列标记, 数据分割 ,组块分析等自然语言处理任务中。在中文分词、中文人名识别、歧义消解等汉语自然语言处理任务中都有良好的表现。

内 容 提 要

- 9.1 序列标注问题
 - 9.2 隐马尔科夫模型HMM
 - 9.3 条件随机场 CRF
 - 9.4 神经网络 RNN+ CRF

9.4 神经网络 RNN+ CRF

神经网络序列标注模型架构

标签解码



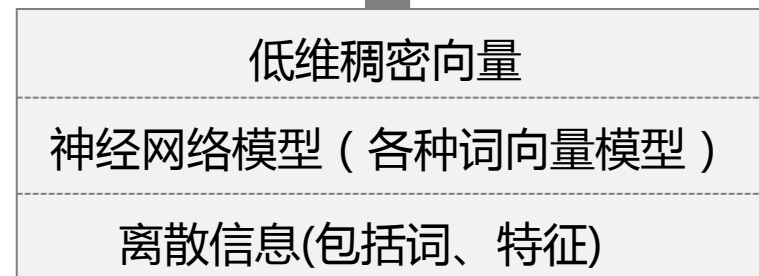
由特征表示生成
预测标签

表示层



由输入向量生成
各特征表示

输入层



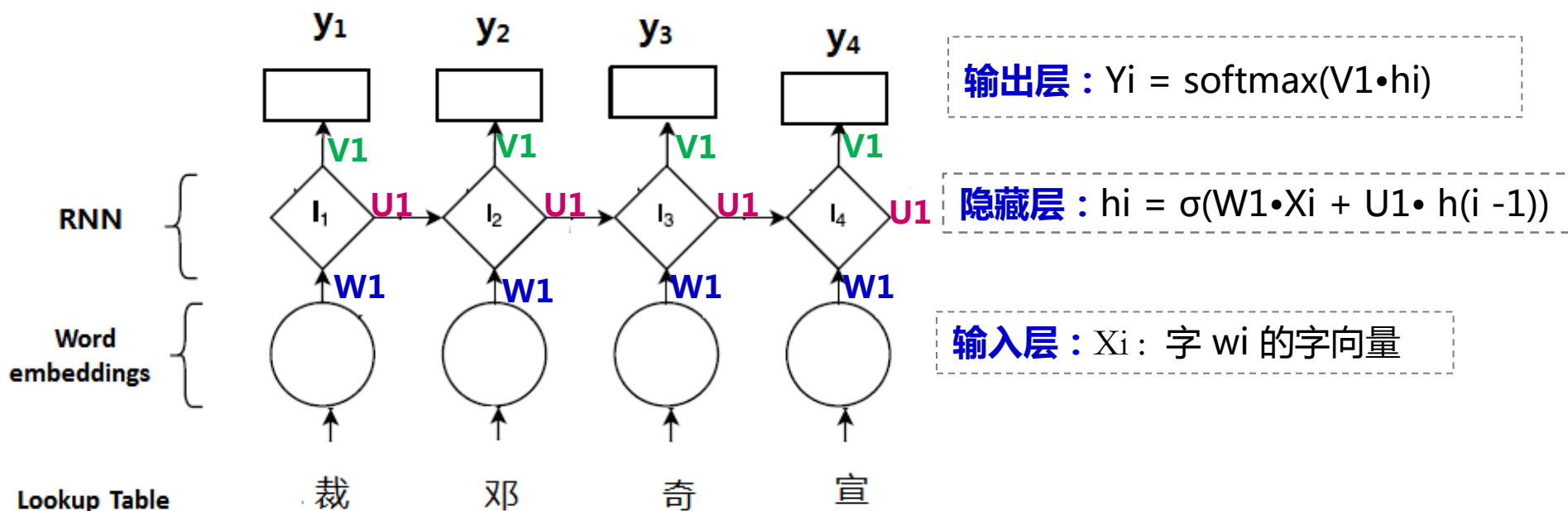
将离散特征(包括词、特征)
转变为低维稠密向量, 并完
成同一位置不同特征的组合

9.4 神经网络 RNN+ CRF

1. 用单向RNN结构：（以人名识别为例）

输入序列： 新任总裁 邓奇宣布了对部门经理罗建国的任免通知

■ 单向RNN模型结构：



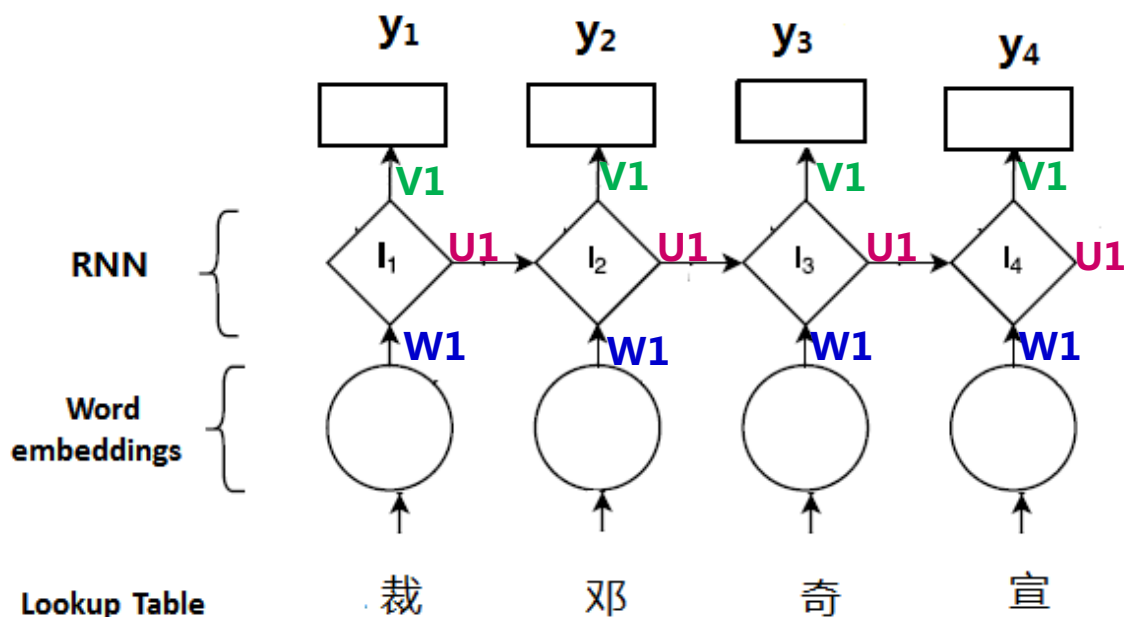
标签 $Y \in \{B, M, E, S\}$

9.4 神经网络 RNN+ CRF

1. 用单向RNN结构：（以人名识别为例）

输入序列： 新任总裁 邓奇宣布了对部门经理罗建国的任免通知

■ 单向RNN模型结构：



输入：中文字序列对应的字向量
输出：每一个中文字对应的标签

参数：

$W1$ ：输入层到隐藏层参数

$U1$ ：隐藏层之间的参数

$V1$ ：隐藏层到输出层参数

标签 $Y \in \{B, M, E, S\}$

9.4 神经网络 RNN+ CRF

■ 模型学习

● 定义损失函数

交叉熵损失：
$$J(\theta; x, y) = - \sum_{j=1}^4 y_j \log((y_{pred})_j)$$

$$\theta = \{ W1, U1, V1 \}$$

整体损失：
$$J(\theta) = \frac{1}{m} \sum_{i=1}^m J(\theta; x^{(i)}, y^{(i)})$$

● 用BPTT算法训练参数 $W1, U1, V1$

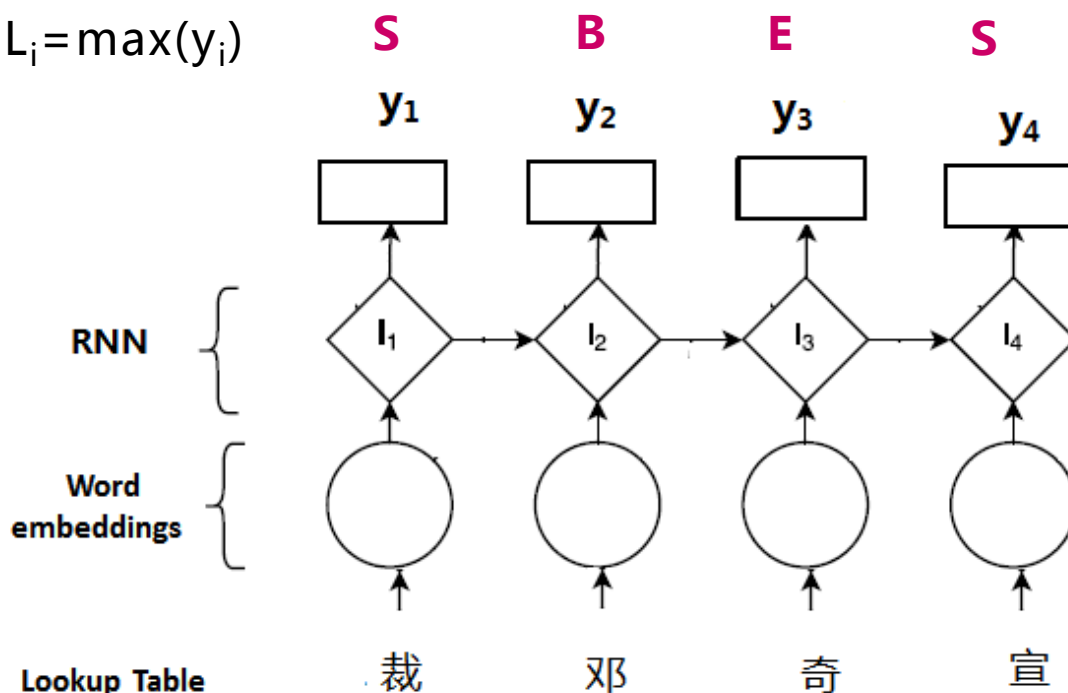
注：字向量对系统性能有影响，可做为模型参数学习

9.4 神经网络 RNN+ CRF

■ 模型预测：

输入序列： 新任总裁 邓奇宣布了对部门经理罗建国的任免通知

输出标签 $L_i = \max(y_i)$



标签 {B,M,E,S}

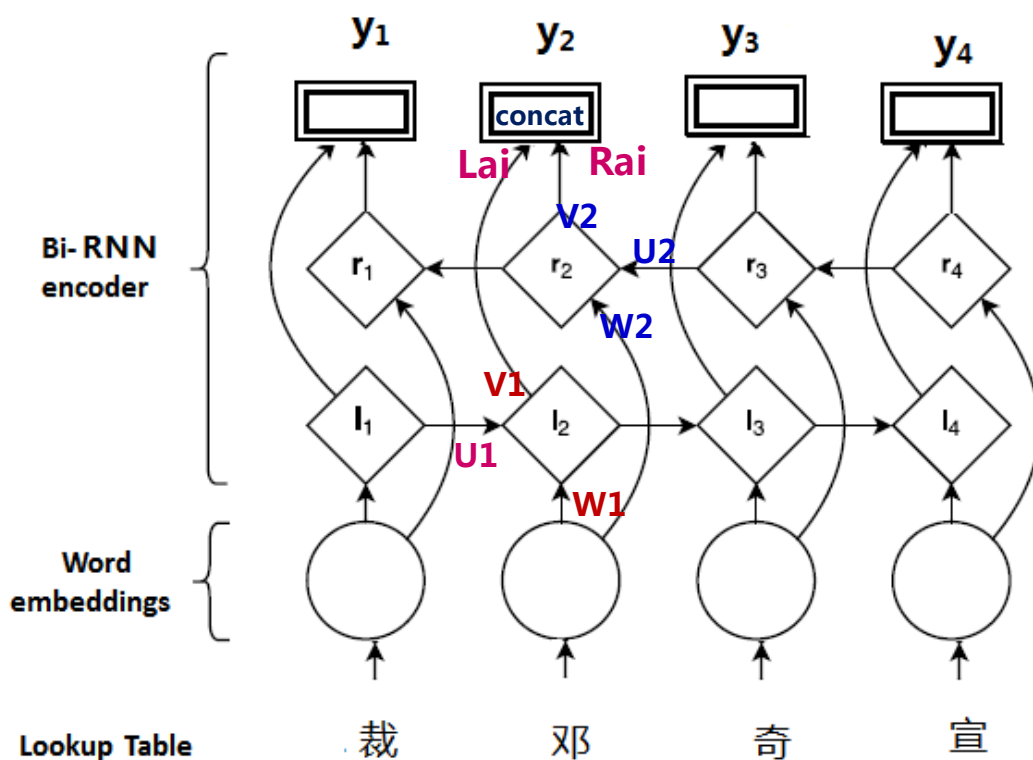
问题：只有上文信息，输出之间独立，有可能出现不合理的标注。如 B B

9.4 神经网络 RNN+ CRF

2. 用双向RNN结构：

输入序列： 新任总裁 邓奇宣布了对部门经理罗建国的任免通知

■ 双向RNN模型结构：



输出层： $Y_i = \text{softmax}(\text{concat}_i)$

$$\text{Concat}_i = \tanh(W_C [L_{ai}, R_{ai}])$$

$$L_{ai} = L_{hi} \times V1 \quad R_{ai} = R_{hi} \times V2$$

反向隐藏层：

$$R_{hi} = \sigma(W2 \cdot X_i + U2 \cdot R_{h(i-1)})$$

正向隐藏层：

$$L_{hi} = \sigma(W1 \cdot X_i + U1 \cdot L_{h(i-1)})$$

输入层： X_i ： 字 w_i 的字向量

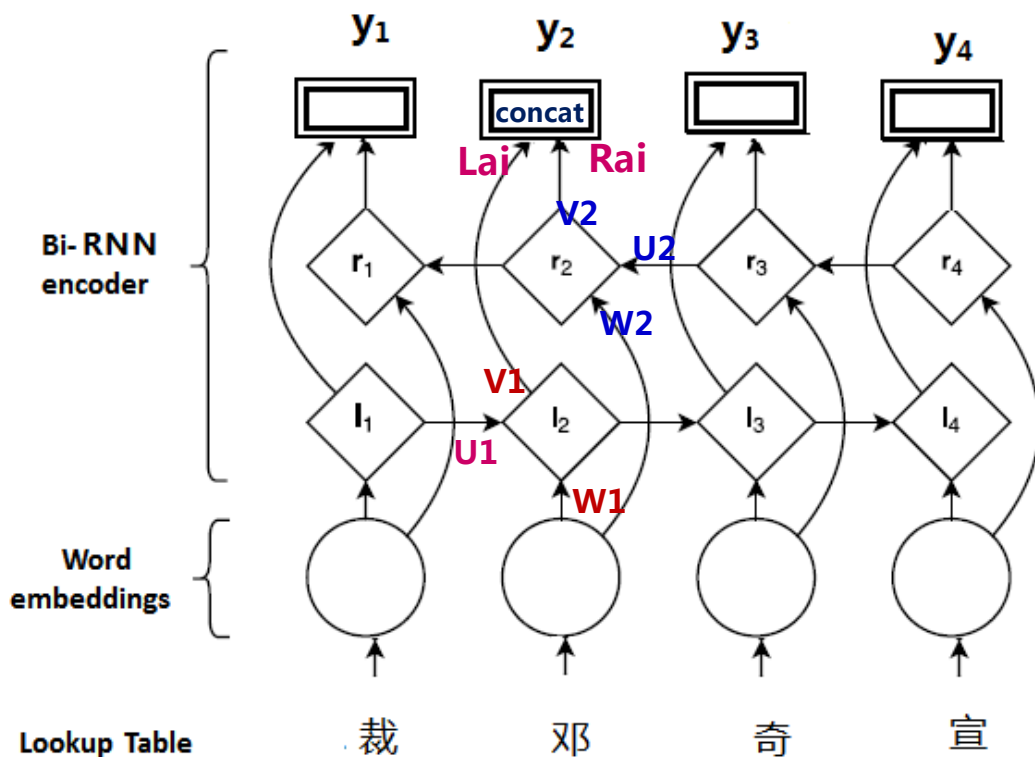
标签 $Y \in \{B, M, E, S\}$

9.4 神经网络 RNN+ CRF

2. 用双向RNN结构：

输入序列： 新任总裁 邓奇宣布了对部门经理罗建国的任免通知

■ 双向RNN模型结构：



输入：中文字序列对应的字向量

输出：每一个中文字对应的标签

参数：

$W1$ ：L 输入层到隐藏层参数

$U1$ ：L 隐藏层之间的参数

$V1$ ：L 隐藏层到输出层参数

$W2$ ：R 输入层到隐藏层参数

$U2$ ：R 隐藏层之间的参数

$V2$ ：R 隐藏层到输出层参数

Wc ：concat 层参数

标签 $Y \in \{B, M, E, S\}$

9.4 神经网络 RNN+ CRF

■ 模型学习

● 定义损失函数

$$\theta = \{W1, U1, V1, W2, U2, V2, Wc, \}$$

交叉熵损失：

$$J(\theta; x, y) = - \sum_{j=1}^4 y_j \log((y_{pred})_j)$$

整体损失：

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m J(\theta; x^{(i)}, y^{(i)})$$

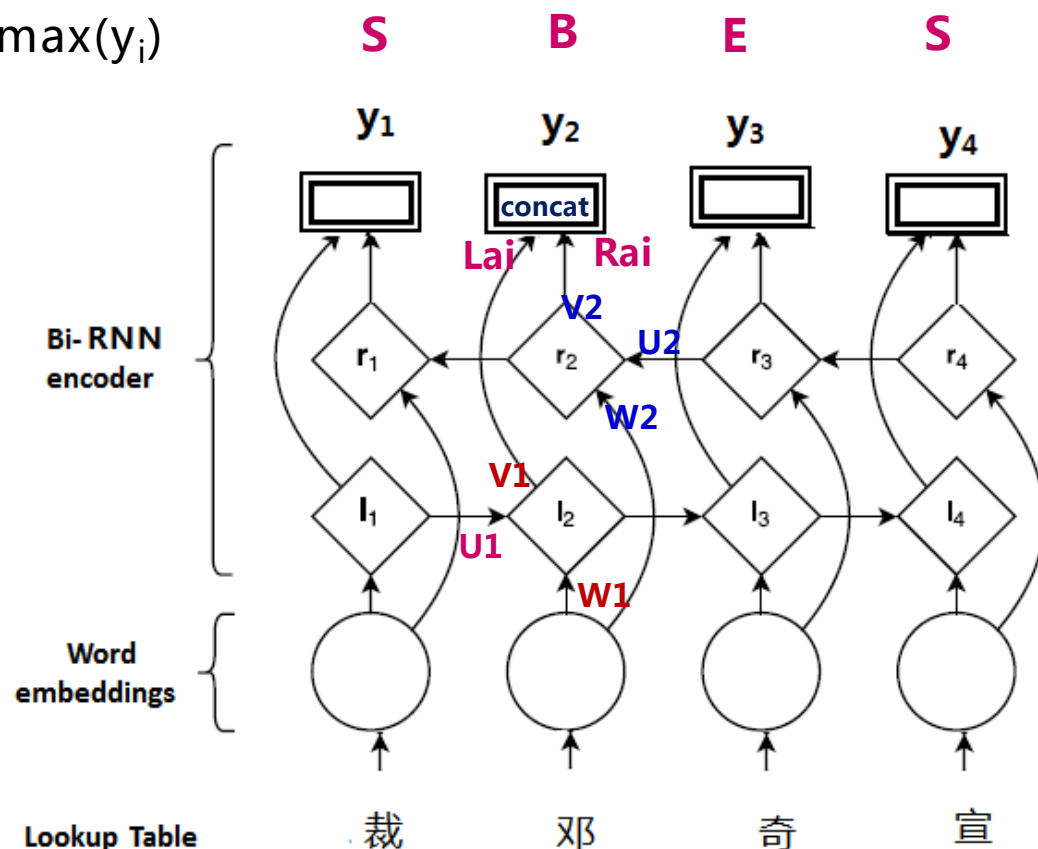
● 用BPTT算法训练参数 $W1, U1, V1, W2, U2, V2, Wc$

9.4 神经网络 RNN+ CRF

■ 模型预测：

输入序列： 新任总裁 邓奇宣布了对部门经理罗建国的任免通知

输出标签 $L_i = \max(y_i)$



标签 $Y \in \{B, M, E, S\}$

问题：输出之间独立，有可能出现不合理的标注。如 B B

9.4 神经网络 RNN+ CRF

3.用双向RNN+CRF：

■ 模型结构

A: $k \times k$ 方阵

B M E S

B
M
E
S

A_{i,j}: 表示从标签 i 转为
标签 j 的得分

输出: $y = (y_1, y_2, \dots, y_n)$

$$s(X, y) = \sum_{i=0}^L A_{y_i, y_{i+1}} + \sum_{i=1}^L P_{i, y_i}$$
$$p(y|X) = \frac{e^{s(X, y)}}{\sum_{\tilde{y} \in Y_X} e^{s(X, \tilde{y})}}$$

输入: $X = (x_1, x_2, \dots, x_n)$,

CRF模型：

$$p(y|x, \lambda) = \frac{1}{Z(x)} \exp \left(\sum_j \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_k \mu_k s_k(y_i, x, i) \right)$$

其中：

$$Z(x) = \sum_y \exp \left(\sum_j \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_k \mu_k s_k(y_i, x, i) \right)$$

P: $n \times k$ 矩阵

B M E S
B M E S
⋮
B M E S

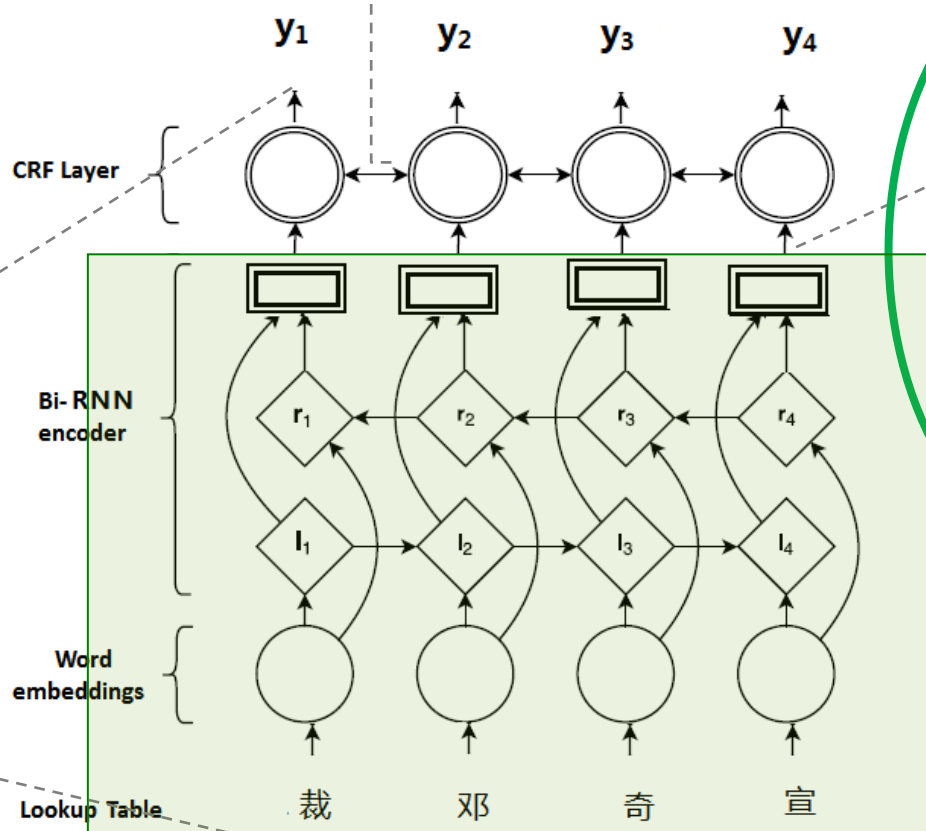
n: 输入长度

K: 标签数

$P_{i,j}$: 表示第 i 个字的
第 j 个标签的得分

模型参数:

A 及产生 P 的
(W1, U1, V1, W2,
U2, V2, Wc,)



9.4 神经网络 RNN+ CRF

■ 模型学习

- 优化目标：
$$p(y|X) = \frac{e^{s(X,y)}}{\sum_{\tilde{y} \in Y_X} e^{s(X,\tilde{y})}}.$$

最大化

$$\begin{aligned}\log(p(y|X)) &= s(X, y) - \log \left(\sum_{\tilde{y} \in Y_X} e^{s(X,\tilde{y})} \right) \\ &= s(X, y) - \text{logadd}_{\tilde{y} \in Y_X} s(X, \tilde{y}),\end{aligned}$$

其中, Y_X 是所有的输出序列

- 用BPTT算法训练参数

$$\theta = [A, W1, U1, V1, W2, U2, V2, Wc]$$

- 模型预测：
$$y^* = \operatorname{argmax}_{\tilde{y} \in Y_X} s(X, \tilde{y}).$$

9.4 神经网络 RNN+ CRF

模型应用：

A: $k \times k$ 方阵

| | B | M | E | S |
|---|---|---|---|---|
| B | × | × | × | × |
| M | × | × | × | × |
| E | × | × | × | × |
| S | × | × | × | × |

CRF模型：

$$p(y|x, \lambda) = \frac{1}{Z(x)} \exp \left(\sum_j \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_k \mu_k s_k(y_i, x, i) \right)$$

其中：

$$Z(x) = \sum_y \exp \left(\sum_j \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_k \mu_k s_k(y_i, x, i) \right)$$

P: $n \times k$ 矩阵

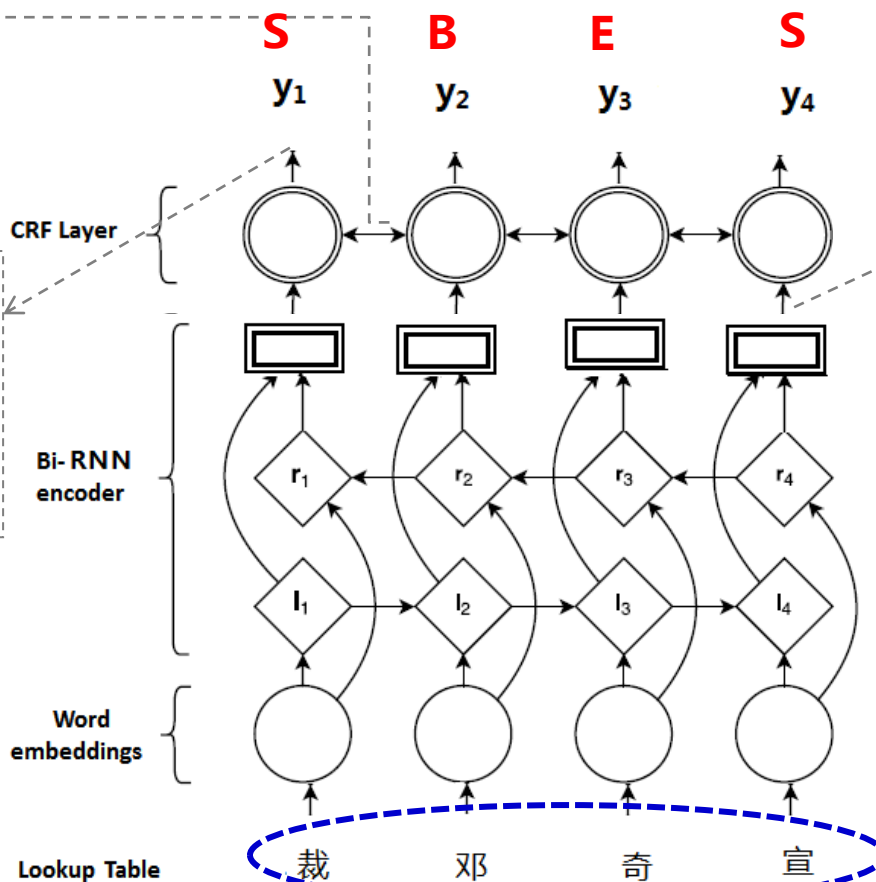
| B | M | E | S |
|-----|-----|-----|-----|
| B | M | E | S |
| ... | ... | ... | ... |
| B | M | E | S |

$$s(X, y) = \sum_{i=0}^L A_{y_i, y_{i+1}} + \sum_{i=1}^L P_{i, y_i}$$

$$p(y|X) = \frac{e^{s(X, y)}}{\sum_{\tilde{y} \in Y_X} e^{s(X, \tilde{y})}}$$

$$y^* = \operatorname{argmax}_{\tilde{y} \in Y_X} s(X, \tilde{y})$$

S B E S



模型参数:

A 及产生 **P** 的
($W_1, U_1, V_1, W_2, U_2, V_2, W_c$)

9.4 神经网络 RNN+ CRF

序列标注模型对比：

Comparison of tagging performance on POS, chunking and NER tasks for various models.

| | | POS | CoNLL2000 | CoNLL2003 |
|--------|-----------------------------------|--------------|--------------|----------------------|
| Random | Conv-CRF (Collobert et al., 2011) | 96.37 | 90.33 | 81.47 |
| | LSTM | 97.10 | 92.88 | 79.82 |
| | BI-LSTM | 97.30 | 93.64 | 81.11 |
| | CRF | 97.30 | 93.69 | 83.02 |
| | LSTM-CRF | 97.45 | 93.80 | 84.10 |
| | BI-LSTM-CRF | 97.43 | 94.13 | 84.26 |
| Senna | Conv-CRF (Collobert et al., 2011) | 97.29 | 94.32 | 88.67 (89.59) |
| | LSTM | 97.29 | 92.99 | 83.74 |
| | BI-LSTM | 97.40 | 93.92 | 85.17 |
| | CRF | 97.45 | 93.83 | 86.13 |
| | LSTM-CRF | 97.54 | 94.27 | 88.36 |
| | BI-LSTM-CRF | 97.55 | 94.46 | 88.83 (90.10) |

CRF效果好于只用LSTM或BI-LSTM

BI-LSTM与CRF结合的方法在多项任务上表现最好

参考文献：

<http://wenku.baidu.com/view/3cf29130f111f18583d05a57.html>

<http://wenku.baidu.com/view/9121f528bd64783e09122b80.html>

李航，统计学习方法 清华大学出版社

Guillaume Lample , Neural Architectures for Named Entity Recognition , NAACL- HLT 2016

在此表示感谢！

谢谢各位！



附录：

最大熵模型  条件随机场 CRF

问题引入

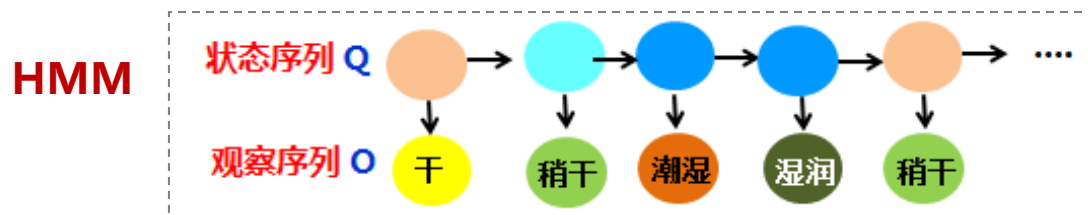
HMM 等生产式模型存在的问题

1. 由于生成模型定义的是联合概率，必须列举所有观察序列的可能值，这对多数领域来说是比较困难的。

在自然语言处理中，常知道各种各样但又不完全确定的信息，需要一个统一的模型将这些信息综合起来。

2. 输出独立性假设要求序列数据严格相互独立才能保证推导的正确性，导致其不能考虑上下文特征

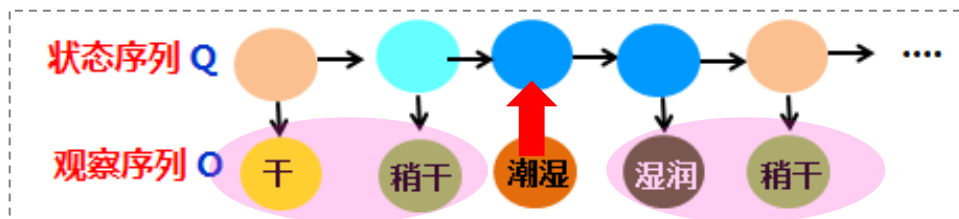
在自然语言处理中，常常需要考虑上下文关系。



问题：HMM对于自然语言中的上下文信息不能利用

问题引入

如何利用各种各样但又不完全确定的信息（上下文信息）？



最大熵模型：

$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中：

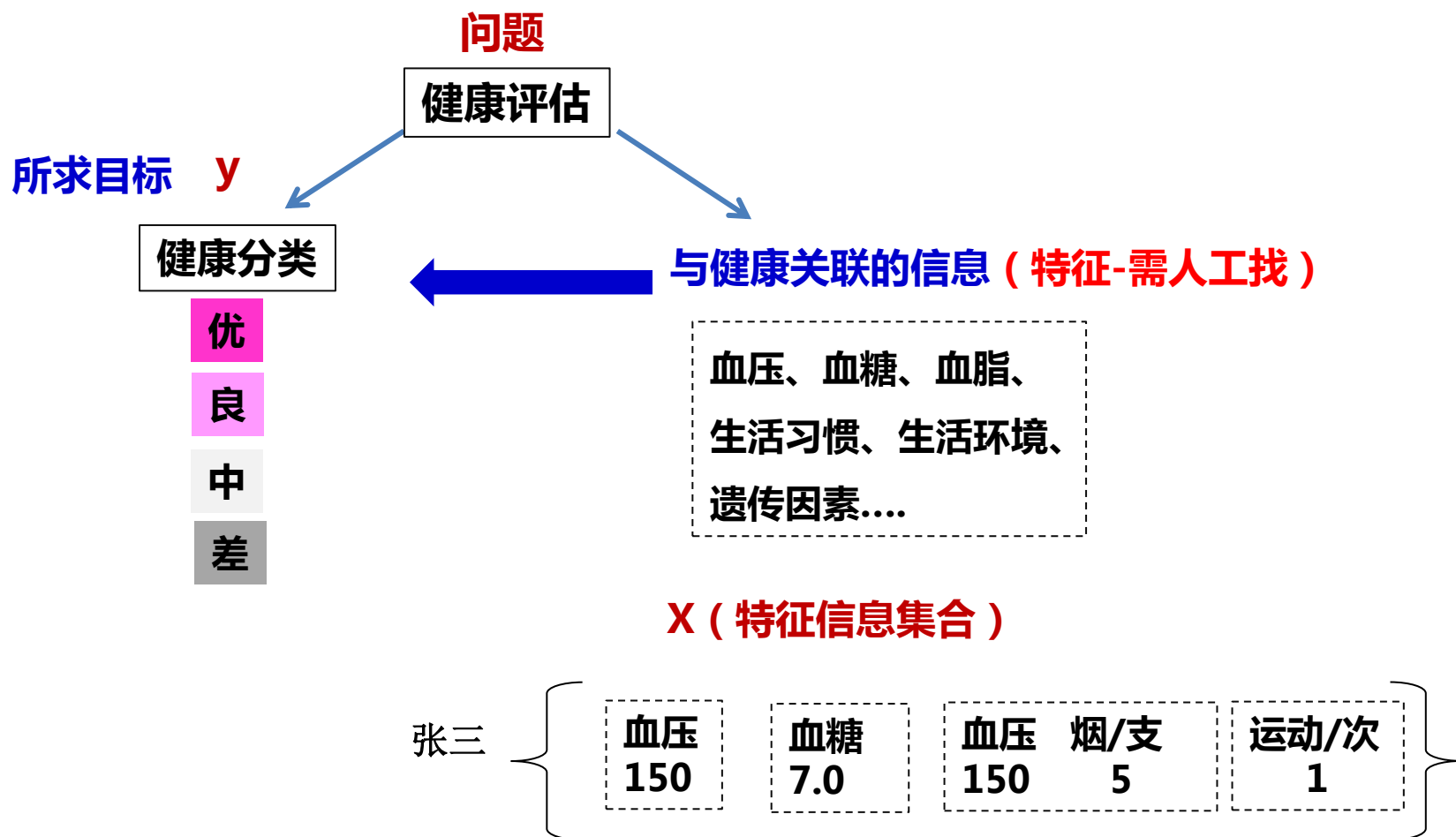
$$Z_{\lambda}(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

- 特点：**
- 可以综合多个不完全信息
 - 这些信息可以灵活表示

最大熵模型

最大熵模型说明（以健康问题为例）



最大熵模型

最大熵模型

X (特征信息集合) 张三

| | | | | |
|-----------|-----------|-----------|----------|-----------|
| 血压 150 | 血糖 7.0 | 血压 150 | 烟/支 5 | 运动/次 1 |
|-----------|-----------|-----------|----------|-----------|

$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中：

$$Z_{\lambda}(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

X中的一个元素

输出：**y** 的分布

| | |
|-------|------|
| y = { | P(优) |
| | P(良) |
| | P(中) |
| | P(差) |

Y(优)

$$\leftarrow \exp\left(\sum_i \lambda_i f_i(x, y_{\text{优}})\right)$$

Y(良)

$$\leftarrow \exp\left(\sum_i \lambda_i f_i(x, y_{\text{良}})\right)$$

特征函数

系统参数：
特征函数的权重 λ_i

最大熵模型

特征函数 $f_i(x, y)$

表示某个特征信息对某个输出是否有影响的事实

$$f_i(x, y) = \begin{cases} 1 & \text{满足条件} \\ 0 & \text{否则} \end{cases}$$

如： $x = \begin{cases} \text{血压} \\ 150 \end{cases}$

$$f_i(x, y_{\text{优}}) = \begin{cases} 1 & X = (\text{血压} : 150) \text{ 并且 } y = \text{"优"} \\ 0 & \text{否则} \end{cases}$$

$$f_i(x, y_{\text{良}}) = \begin{cases} 1 & X = (\text{血压} : 150) \text{ 并且 } y = \text{"良"} \\ 0 & \text{否则} \end{cases}$$

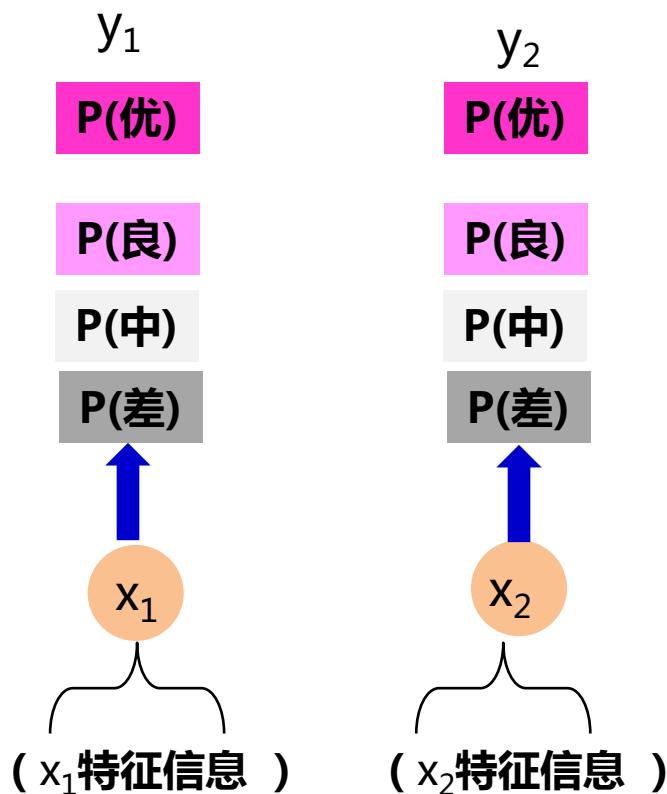
$$f_i(x, y_{\text{中}})$$

$$f_i(x, y_{\text{差}})$$

张三 $\left\{ \begin{array}{|c|c|c|c|} \hline \text{血压} & \text{血糖} & \text{血压} & \text{烟/支} \\ \hline 150 & 7.0 & 150 & 5 \\ \hline \end{array} \right\}$ $\left\{ \begin{array}{|c|} \hline \text{运动/次} \\ \hline 1 \\ \hline \end{array} \right\}$

最大熵模型

用最大熵模型解决序列问题的不足



$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中：

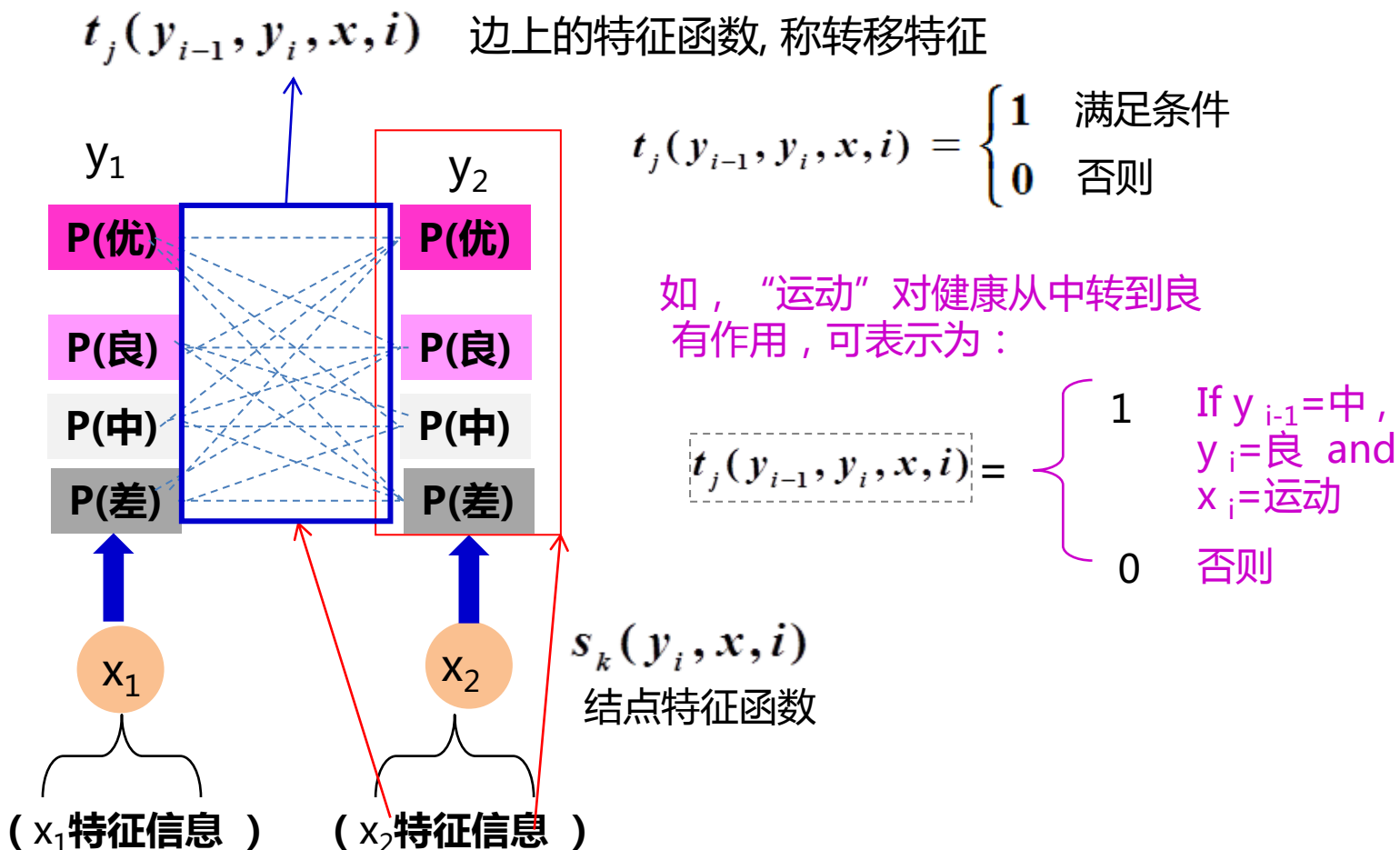
$$Z_{\lambda}(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

输出独立，可能输出 B B情况

条件随机场 CRF

改进： 最大熵模型 \longrightarrow 条件随机场 CRF



CRF :

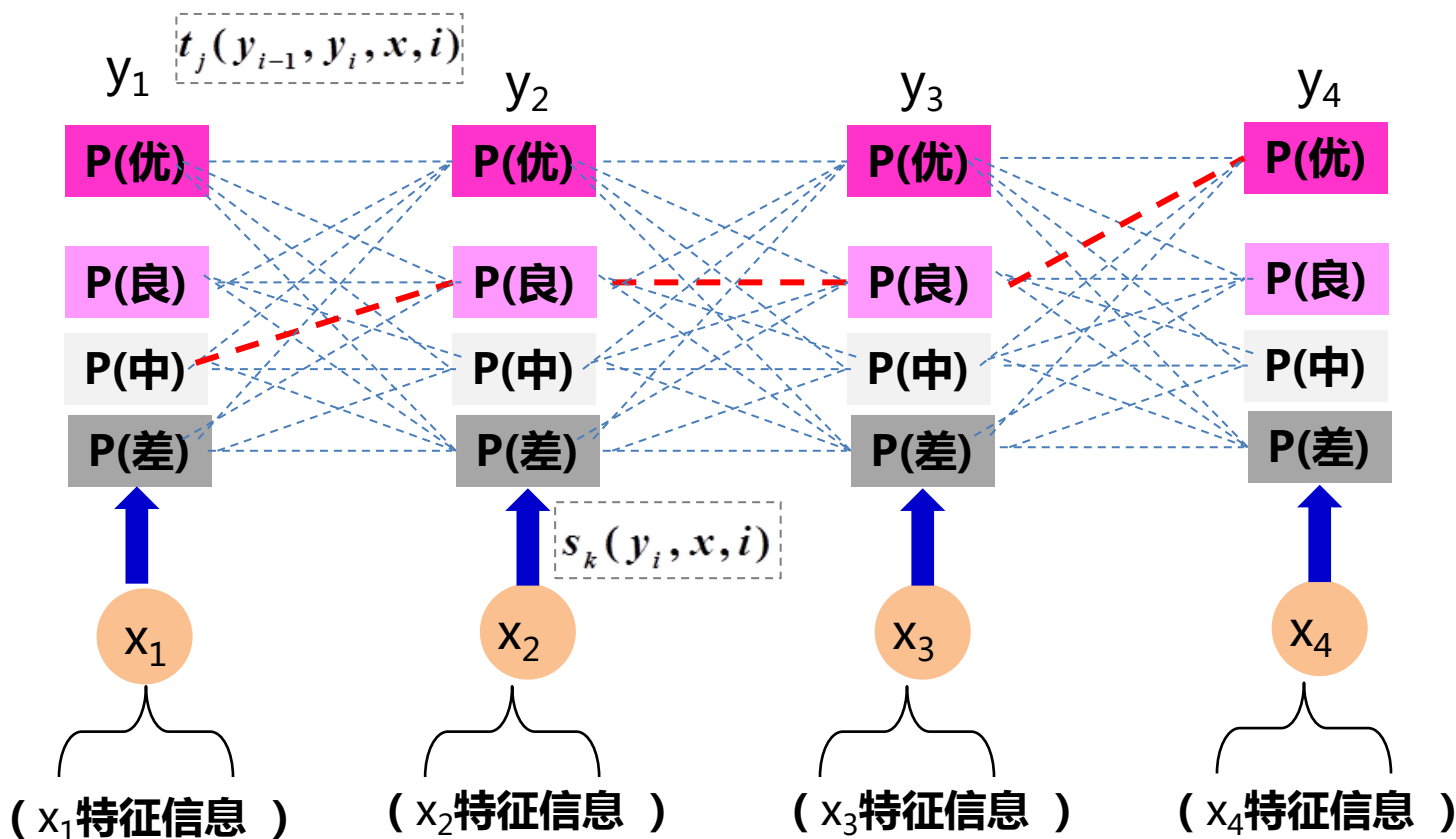
$$p(y|x, \lambda) = \frac{1}{Z(x)} \exp\left(\sum_{ji} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{ki} \mu_k s_k(y_i, x, i)\right)$$

其中：

$$Z(x) = \sum_y \exp\left(\sum_{ji} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{ki} \mu_k s_k(y_i, x, i)\right)$$

输出：

$$p(y_i|x, \lambda) = \exp\left(\sum_{ji} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{ki} \mu_k s_k(y_i, x, i)\right)$$



输入： $x_1 x_2 x_3 x_4$

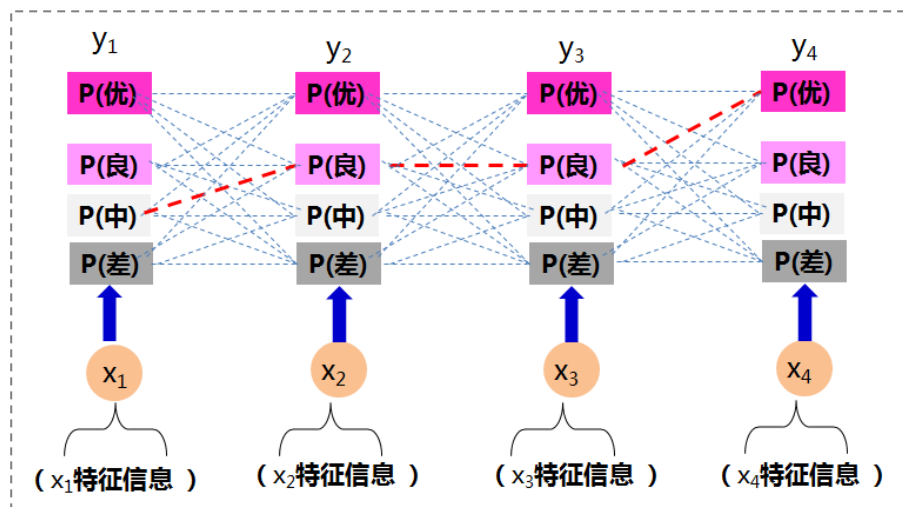
条件随机场 CRF

线性条件随机场的基本形式：

$$p(y|x, \lambda) = \frac{1}{Z(x)} \exp\left(\sum_{ji} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{ki} \mu_k s_k(y_i, x, i)\right)$$

其中：

$$Z(x) = \sum_y \exp\left(\sum_{ji} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{ki} \mu_k s_k(y_i, x, i)\right)$$



输入：

$x = \{x_1, x_2, \dots, x_n\}$ (观测序列)

输出：

$y = \{y_1, y_2, \dots, y_n\}$ (标记序列)

参数：

$\lambda_j \quad \mu_k$

附录完