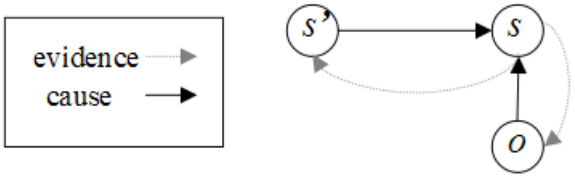
1

- HMM方法的局限性

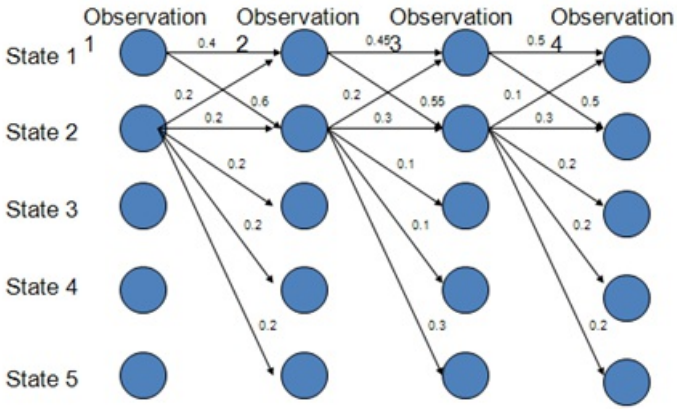
> 1 马尔科夫性（有限历史性）：实际上在NLP领域的文本数据，很多词语都是长依赖的关系。
> 2 □次性： 序列不同位置的状态转移矩阵可能会有所变化， 即位置信息会影响预测结果。
> 3 观测独□性： 观测值和观测值（字与字） 之间是有相关性的。
> 4 单向图： 只与前序状态有关， 和后续状态□关。 在NLP任务中， 上下□的信息都是必须的。
> 5 标记偏置LabelBias： 若状态A能够向N种状态转移， 状态B能够向M种状态转移。 若N<<M， 则预测序列更有可 能选择状态A， 因为A的转移概率较□。

借助MEMM最大熵马尔科夫模型理解标记偏执，但MEMM并没有解决标记偏执

最大熵马尔科夫模型是指将观测



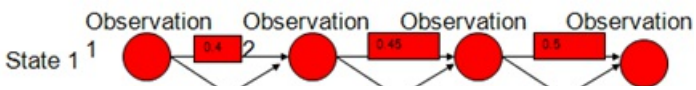
我们可以观察到MEMM与HMM最大的不同是指观测状态与隐藏状态的指向关系发生了变化。

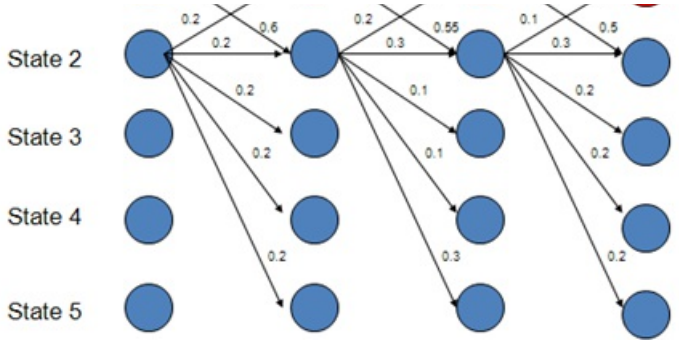


In [2]:

```
%%mark
路径: s1-s1-s1-s1的概率: 0.4*0.45*0.5=0.09

路径s2-s2-s2-s2的概率:0.2*0.3*0.3=0.018

路径s1-s2-s1-s2的概率:0.6*0.2*0.5=0.06

路径s1-s1-s2-s2的概率:0.4*0.55*0.3=0.066

由此可得最优路径为s1-s1-s1-s1
```

```
  File "<ipython-input-2-4cc06fc4c2d6>", line 2
    路径: s1-s1-s1-s1的概率: 0.4*0.45*0.5=0.09
       ^
SyntaxError: invalid character in identifier
```

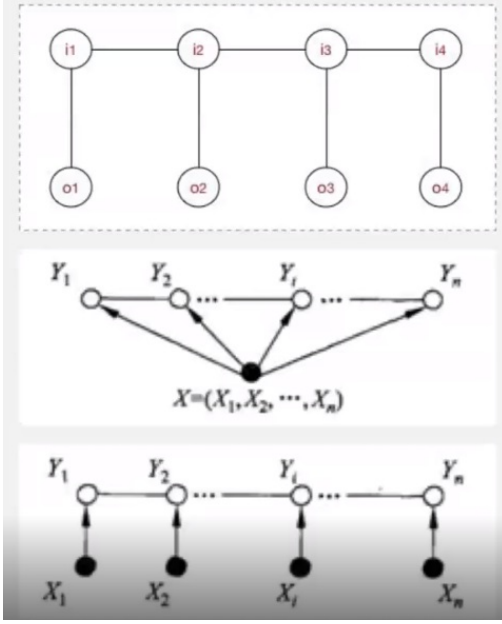




实际上，在上图中，状态1偏向于转移到状态2，而状态2总倾向于停留在状态2，这就是所谓的标注偏置问题，由于分支数不同，概率的分布不均衡，导致状态的转移存在不公平的情况。 由上面的两幅图可知，最大熵隐马尔科夫模型（MEMM）只能达到局部最优解，而不能达到全局最优解，因此MEMM虽然解决了HMM输出独立性假设的问题，但却存在标注偏置问题。

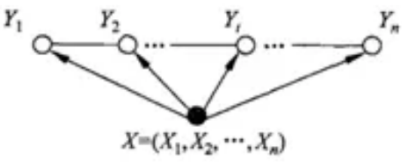
2 CRF

3.2、CRF条件随机场



- 随机场是一个图模型，是由若干个结点（随机变量）和边（依赖关系）组成的图模型，当给每一个结点按照某种分布随机赋予一个值之后，其全体就叫做随机场。
- 马尔可夫随机场是随机场的特例，它假设随机场中任意一个结点的赋值，仅仅和它的邻结点的取值有关，和不相邻的结点的取值无关。用学术语言表示是：满足成对、局部或全局马尔科夫性。
- 条件随机场CRF是马尔可夫随机场的特例，它假设模型中只有X（输入变量，观测值）和Y（输出变量，状态值）两种变量。输出变量Y构成马尔可夫随机场，输入变量X不具有马尔可夫性。
- 线性链条件随机场，是状态序列是线性链的条件随机场。

3.2、CRF条件随机场公式推“倒”



在linear-CRF中，特征函数分为两类。
第一类是定义在Y节点上的节点特征函数，这类特征函数只和当前节点有关，记为：

$$s_l(y_i,x,i),\ \ l=1,2,\dots L$$

第二类是定义在Y上下文的局部特征函数，这类特征函数只和当前节点和上一个节点有关，记为：

$$t_k(y_{i-1},y_i,x,i),\ \ k=1,2,\dots K$$

无论是节点特征函数还是局部特征函数，它们的取值只能是0或者1。即满足特征条件或者不满足特征条件。同时，我们可以为每个特征函数赋予一个权值，用以表达我们对这个特征函数的信任度。假设tk的权重系数是λk,sl的权重系数是μl,则linear-CRF由我们所有的tk,λk,sl,μl共同决定。此时我们得到了linear-CRF的参数化形式如下：

$$P(y|x)=\frac{1}{Z(x)}exp\Big(\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)+\sum_{i,l}\mu_l s_l(y_i,x,i)\Big)$$

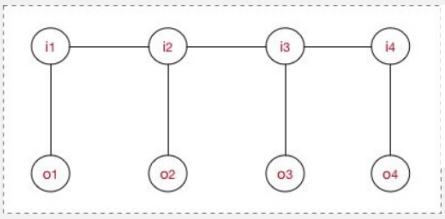
其中，Z(x)为规范化因子：

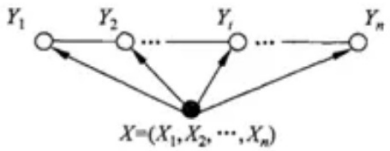
$$Z(x)=\sum_{y}exp\Big(\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)+\sum_{i,l}\mu_l s_l(y_i,x,i)\Big)$$

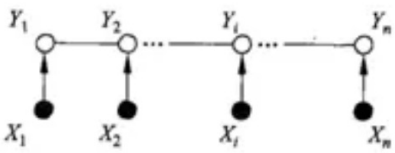
> 我们先来看下公式中的参数，不用说x, y。那么参数i肯定指的是状态变量的个数，那么k和l是指什么？

> - 我们可以这样理解，t_k是特征函数，数字k就代表X_i具有M个特征中的第几个特征。类比一个元素有M个神经元。

3.2、CRF条件随机场公式推“倒”







$$P(Y)=\frac{1}{Z(x)}\prod_c \psi_c(Y_c)$$

$$P(y|x)=\frac{1}{Z(x)}exp\Big(\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)+\sum_{i,l}\mu_l s_l(y_i,x,i)\Big)$$

$$P(I|O)=\frac{1}{Z(O)}e^{\sum_i^T\sum_k^M \lambda_k f_k(O,I_{i-1},I_i,i)}=\frac{1}{Z(O)}e^{[\sum_i^T\sum_j^J \lambda_j t_j(O,I_{i-1},I_i,i)+\sum_i^T\sum_l^L \mu_l s_l(O,I_i,i)]}$$

找一个例子来计算一下一个已知词性标注序列的非规范化概率

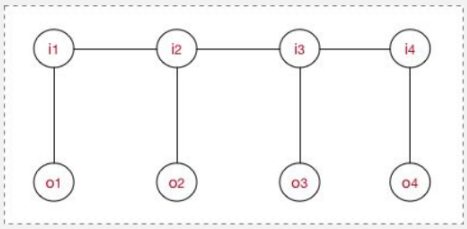
3.2、CRF条件随机场

一个linear-CRF用于词性标注的实例，为了方便，我们简化了词性的种类。假设输入的都是三个词的句子，即X=(X1,X2,X3),输出的词性标记为Y=(Y1,Y2,Y3),其中Y∈{1(名词)，2(动词)}

这里只标记出取值为1的特征函数如下：

t1 = t1(yi−1=1,yi=2,x,i), i=2,3, λ1=1
t2 = t2(y1=1,y2=1,x,2) λ2=0.5
t3 = t3(y2=2,y3=1,x,3) λ3=1
t4 = t4(y1=2,y2=1,x,2) λ4=1
t5 = t5(y2=2,y3=2,x,3) λ5=0.2
s1 = s1(y1=1,x,1) μ1=1
s2 = s2(yi=2,x,i), i=1,2, μ2=0.5
s3 = s3(yi=1,x,i), i=2,3, μ3=0.8
s4 = s4(y3=2,x,3) μ4=0.5

求标记(1,2,2)的非规范化概率。



$$P(y|x)\propto exp\Big[\sum_{k=1}^{5}\lambda_k\sum_{i=2}^{3}t_k(y_{i-1},y_i,x,i)+\sum_{l=1}^{4}\mu_l\sum_{i=1}^{3}s_l(y_i,x,i)\Big]$$

$$P(y_1=1,y_2=2,y_3=2|x)\propto exp(3.2)$$

标记一下重点：
已知的是观测结果Y=（1,2,2），代表的序列是名词之后连续接了2个动词，那么根据输入X序列得到上述观测结果的Y的非规范概率可以这样求

i = 2时，

$\lambda_1(\lambda_1=1)*t_1(y_2=2满足y_2-1=1; y_2=2; x, 2)= 1*1 =1$ $\lambda_2(\lambda_2=0.5)*t_2(y_1=1;但y_2=2不满足y_2=1; x, 2)= 0.5*0 =0$ $\lambda_3(\lambda_3=1)*t_3(y_2=2满足y_i-1=1; y_3=2不满足y_3=1; x, i=2不是3)= 1*0 =0$ $\lambda_4(\lambda_4=1)*t_4(y_1=1不满足y_1=2; y_2=2不满足y2=1; x, 2)= 1*0 =0$ $\lambda_5(\lambda_5=0.2)*t_5(y_2=2; y_3=2; x, i=2非3)= 0.2*0 =0$

i = 3时，
$\lambda_1(\lambda_1=1)*t_1(y_3-1=1; y_3=2; x, i=3非2)= 1*1 =1$ $\lambda_2(\lambda_2=0.5)*t_2(y_1=1;但y_2=2不满足y_2=1; x, i=3非2)= 0.5*0 =0$ $\lambda_3(\lambda_3=1)*t_3(y_2=2满足y_i-1=1; y_3=2不满足y_3=1; x, 3)= 1*0 =0$ $\lambda_4(\lambda_4=1)*t_4(y_1=1不满足y_1=2; y_2=2不满足y2=1; x, 2)= 1*0 =0$ $\lambda_5(\lambda_5=0.2)*t_5(y_2=2; y_3=2; x, 3)= 0.2*1 =0.2$

因此第二类特征函数的值是2.2

i =1时，
$\mu_1(\mu_1=1)*s_1(s_1=1)=1*1=1$,
$\mu_2(\mu_2=0.5)*s_2(s_2=0)=0.5*0=0$,
$\mu_3(\mu_2=0.8)*s_3(s_3=0)=0.8*0=0$,
$\mu_4(\mu_2=0.5)*s_4=0.5*0=0$,

i =2时，
$\mu_1(\mu_1=1)*s_1(s_1=1)=1*0=0$,
$\mu_2(\mu_2=0.5)*s_2(s_2=1)=0.5*1=0.5$,
$\mu_3(\mu_2=0.8)*s_3(s_3=0)=0.8*0=0$,
$\mu_4(\mu_2=0.5)*s_4=0.5*0=0$,

i =3时，
$\mu_1(\mu_1=1)*s_1(s_1=0)=1*0=0$,
$\mu_2(\mu_2=0.5)*s_2(s_2=0)=0.5*0=0$,
$\mu_3(\mu_2=0.8)*s_3(s_3=0)=0.8*0=0$,
$\mu_4(\mu_2=0.5)*s_4=0.5*1=0.5$,

因此第一类的特征函数值为2.0。因此非规范化概率为$\exp^{4.2}$

规范因子$Z(O)$就是指{1,2,2}进行全排列，求解上述非规范因子的和

In [ ]:

3.2、CRF条件随机场的优缺点

CRF相对于HMM的优点

(1) 规避了马尔可夫性（有限历史性），能够获取长文本的远距离依赖的信息。

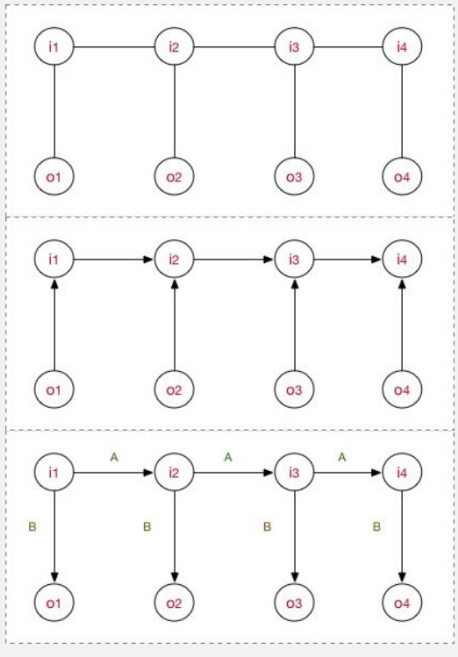
(2) 规避了齐次性，模型能够获取序列的位置信息，并且序列的位置信息会影响预测出的状态序列。

(3) 规避了观测独立性，观测值之间的相关性信息能够被提取。

(4) 不是单向图，而是无向图，能够充分提取上下文信息作为特征。

(5) 改善了标记偏置LabelBias问题，因为CRF相对于HMM能够更多地获取序列的全局概率信息。

(6) CRF的思路是利用多个特征，对状态序列进行预测。HMM的表现形式使他无法使用多个复杂特征。



3.2、概率图模型课后复习问题清单

- 有向图和无向图的P（Y）分别如何计算?
- HMM，HEMM，CRF的定义是什么？两两之间的区别是什么？优缺点是什么?
- 前向算法，后向算法如何计算?
- 标注偏置的原因是什么？如何解决?
- 维特比算法如何求解？维特比算法的DP公式如何写?
- EM算法是什么？如何求HMM参数?

问题清单解答

1

> 有向图： 后一个点只跟前一个点的状态有关。并且直接相乘即可。

> 无向图：后一个点的跟周围的点都有关系，利用最大团C可以写作它Y的联合概率可以表示为其最□团C上随机变量的函数的乘积的形t式。即：

公式3 $$P(Y)=\frac{1}{Z(x)}\prod_{c}\psi_c(Y_c)$$

公式4 $$Z(x)=\sum_{Y}\prod_{c}\psi_c(Y_c)$$

公式5 $$\psi_c(Y_c)=e^{-E(Y_c)}$$

2

HMM的基本定义： HMM是□于描述由隐藏的状态序列和显性的观测序列组合□成的双重随机过程。

条件随机场CRF是□尔可夫随机场的特例， 它假设模型中 只有X（输□变量， 观测值） 和Y（输出变量， 状态值） 两 种变量。 输出变量Y构成□尔可夫随机场， 输□变量X不具 有□尔可夫性。

3

前向算法是指已知HMM模型参数，求解观测序列的概率
后向算法是指已知HMM模型参数，求解观测序列的最有可能的隐藏状态

4

若状态A能够向N种状态转移， 状态B能够向M种状态转移。 若N<<M， 则预测序列更有可 能选择状态A， 因为A的转移概率较□。

5

如何求HMM的参数及EM最大似然估计

In [ ]:

```
```