

大数据机器学习第九次理论作业

2026 年 1 月 22 日

一、 11.1 写出图 11.3 中无向图描述的概率图模型的因子分解式

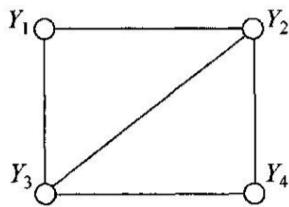


图 11.3 无向图的团和最大团

1. 问题分析与相关知识点

1). 条件随机场

条件随机场 (conditional random field, CRF) 是给定一组输入随机变量条件下另一组输出随机变量的条件概率分布模型。

2). 概念引入：概率图模型和无向图随机变量之间的性质

图 (graph) 是由结点 (node) 以及连接节点的边 (edge) 组成的集合。结点和边分别记作 v 和 e , 结点和边的集合分别记为 V 和 E , 图记为 $G = (V, E)$ 。无向图指的是边没有方向的图。

概率图模型 (probabilistic graphical model) 是由图表示的概率分布。我们设有联合概率分布 $P(Y)$, $Y \in \mathcal{Y}$ 是一组随机变量。由无向图 $G = (V, E)$ 表示概率分布 $P(Y)$, 即是在图 G 中, 结点 $v \in V$ 表示一个随机变量 Y_v , $Y = (Y_v)_{v \in V}$; 边 $e \in E$ 表示随机变量之间的概率依赖关系。

给定一个联合概率分布 $P(Y)$ 和它的无向图 G 。首先定义无向图表示的随机变量之间存在的成对马尔可夫性、局部马尔可夫性和全局马尔可夫性。

1. 成对马尔可夫性: 设 u 和 v 是无向图 G 中任意两个没有边连接的结点, 结点 u 和 v 分别对应随机变量 Y_u 和 Y_v 。其他所有结点记为 O , 对应的随机变量组 Y_O , 成对马尔可夫性指给定 Y_O 的条件下 Y_u 和 Y_v 是条件独立的。

$$P(Y_u, Y_v | Y_O) = P(Y_u | Y_O)P(Y_v | Y_O) \quad (1)$$

2. 局部马尔可夫性: 设 $v \in V$ 是无向图 G 任意一个结点, W 是与 v 有边连接的所有邻居结点, O 为 v 和 W 之外的所有结点。 v 表示的随机变量为 Y_v , W 表示的随机变量组为

Y_W, O 表示的随机变量组为 Y_O 。局部马尔可夫性是指在给定随机变量 Y_W 的条件下随机变量 Y_v 与随机变量组 Y_O 是独立的。

$$P(Y_v, Y_O | Y_W) = P(Y_v | Y_W)P(Y_O | Y_W) \quad (2)$$

在 $P(Y_O | Y_W) > 0$ 时, 等价地,

$$P(Y_v | Y_W) = P(Y_v | Y_W, Y_O) \quad (3)$$

3. 全局马尔可夫性: 设结点集合 A, B 是在无向图 G 中被结点集合 C 分开的任意结点集合, 如图 11.2 所示。结点集合 A, B 和 C 所对应的随机变量组分别是 Y_A, Y_B 和 Y_C 。全局马尔可夫性指在给定随机变量组 Y_C 的条件下, 随机变量组 Y_A 和 Y_B 是条件独立的。

$$P(Y_A, Y_B | Y_C) = P(Y_A | Y_C)P(Y_B | Y_C) \quad (4)$$

3). 概率无向图模型 (马尔可夫随机场)

概率无向图模型 (probabilistic undirected graphical model) 又称为马尔可夫随机场, 是一个可以由无向图表示的联合概率分布。其定义为:

A. 定义 (概率无向图模型) 设有联合概率分布 $P(Y)$, 由无向图 $G = (V E)$ 表示, 在图 G 中, 结点表示随机变量, 边表示随机变量之间的依赖关系, 如果联合概率分布 $P(Y)$ 满足成对、局部或全局马尔可夫性, 就称此联合概率分布为概率无向图模型, 或马尔可夫随机场。

4). 概率无向图模型的因子分解

A. 定义 (无向图中的团与最大团) 无向图 G 中任何两个结点均有边连接的结点子集称为团。如果 C 是无向图 G 的一个团, 并且不能再加进去任何一个 G 的结点让它成为一个更大的团, 就称 C 为最大团。

如图 11.3 表示的团中, 有 2 个最大团 $\{Y_1, Y_2, Y_3\}$ 和 $\{Y_2, Y_3, Y_4\}$ 。

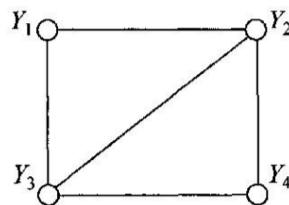


图 11.3 无向图的团和最大团

B. 定义 (无向图中的团与最大团) 将概率无向图模型的联合概率分布表示成其最大团上的随机变量的函数的乘积形式的操作, 称为概率无向图模型的因子分解。

给定概率无向图模型, 设其无向图为 G , C 为 G 上的最大团, Y_C 表示 C 对应的随机变量。那么:

$$P(Y) = \frac{1}{Z} \prod_C \Psi_C(Y_C) \quad (5)$$

其中, Z 为规范化因子, 有:

$$Z = \sum_Y \sum_C \Psi_C(Y_C) \quad (6)$$

函数 $\Psi_C(Y_C)$ 称为势函数 (potential function)。这里要求其严格为正数，通常定义为：

$$\Psi_C(Y_C) = \exp\{-E(Y_C)\} \quad (7)$$

概率图模型的因子分解由下述 Hammersley-Clifford 定理保证。

C: 定理 (Hammersley-Clifford 定理) 概率无向图模型的联合概率分布 $P(Y)$ 可以表示为如下形式：

$$P(Y) = \frac{1}{Z} \prod_C \Psi_C(Y_C)$$

$$Z = \sum_Y \sum_C \Psi_C(Y_C)$$

其中， C 为无向图最大团， Y_C 为结点 C 对应的随机变量， $\Psi_C(Y_C)$ 是 C 上定义的严格正函数，乘积是在无向图所有最大团上进行的。

2. 问题求解

解：本题中，图 11.3 有 2 个最大团 $C_1 = \{Y_1, Y_2, Y_3\}$ 和 $C_2 = \{Y_2, Y_3, Y_4\}$ 。我们根据公式 (5),(6) 将联合概率 $P(Y)$ 分解为：

$$P(Y) = \frac{1}{Z} \Psi_{\{Y_1, Y_2, Y_3\}}(Y_1, Y_2, Y_3) \Psi_{\{Y_2, Y_3, Y_4\}}(Y_2, Y_3, Y_4)$$

其中，

$$Z = \sum_Y \Psi_{\{Y_1, Y_2, Y_3\}}(Y_1, Y_2, Y_3) \Psi_{\{Y_2, Y_3, Y_4\}}(Y_2, Y_3, Y_4)$$

若基于 $\Psi_C(Y_C) = \exp\{-E(Y_C)\}$ 定义势函数，则：

$$\Psi_{\{Y_1, Y_2, Y_3\}}(Y_1, Y_2, Y_3) = \exp\{-E(Y_1, Y_2, Y_3)\}$$

$$\Psi_{\{Y_2, Y_3, Y_4\}}(Y_2, Y_3, Y_4) = \exp\{-E(Y_2, Y_3, Y_4)\}$$

二、11.4 参考图 11.6 的状态路径图，假设随机矩 $M_1(x), M_2(x), M_3(x), M_4(x)$ 分别如下所示。求以 $\text{start} = 2$ 为起点， $\text{stop} = 2$ 为终点的所有路径的状态序列 y 的概率及概率最大的状态序列。

$$M_1(x) = \begin{bmatrix} 0 & 0 \\ 0.5 & 0.5 \end{bmatrix}, \quad M_2(x) = \begin{bmatrix} 0.3 & 0.7 \\ 0.7 & 0.3 \end{bmatrix}$$

$$M_3(x) = \begin{bmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \end{bmatrix}, \quad M_4(x) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

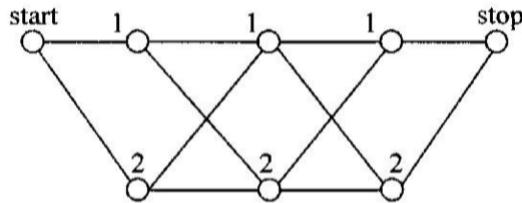


图 11.6 状态路径

1. 问题分析与相关知识点

1). 条件随机场

条件随机场是给定随机变量 X 的条件下，随机变量 Y 的马尔可夫随机场。在线性链上的特殊的条件随机场，称为线性链条件随机场。在条件概率模型 $P(Y|X)$ 中， Y 是输出变量，表示标记序列， X 是输入变量，表示需要标注的观测序列。在学习时，利用训练数据集通过极大似然估计或者正则化的极大似然估计的到条件概率模型 $\hat{P}(Y|X)$ ，预测时，对于给定的输入系列 x ，求出条件概率 $\hat{P}(y|x)$ 最大的输出序列 \hat{y} 。

A: 定义（条件随机场） 设 X 与 Y 是随机变量， $P(X|Y)$ 是在给定 X 的条件下 Y 的条件概率分布。若随机变量 Y 构成一个由无向图 $G = (V, E)$ 表示的马尔可夫随机场，即

$$P(Y_v|X, Y_w, w \neq v) = P(Y_v|X, Y_w, w \sim v) \quad (8)$$

对任意结点 v 成立，则称条件概率分布 $P(Y|X)$ 为条件随机场，式中 $w \sim v$ 表示在图 $G = (V, E)$ 中与结点 v 有边连接的所有结点 $w, w \neq v$ 表示结点 v 以外的所有结点， Y_v, Y_u 与 Y_w 为结点 v, u 与 w 对应的随机变量。即我们可以略去与 v 不相邻的结点。

我们一般假设 X 和 Y 有相同图结构，接下来我们考虑无向图为如图 11.4 和图 11.5 所示的线性链的情况。

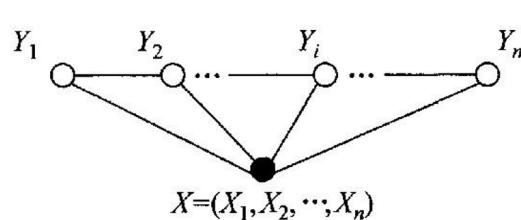


图 11.4 线性链条件随机场

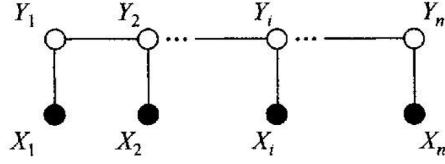


图 11.5 X 和 Y 有相同的图结构的线性链条件随机场

B: 定义（线性链条件随机场） 设 $X = (X_1, X_2, \dots, X_n), Y = (Y_1, Y_2, \dots, Y_n)$ 均为线性链表示的随机变量序列，若在给定随机变量序列 X 的条件下，随机变量序列 Y 的条件概率分布 $P(Y|X)$ 为线性链条件随机场，即满足马尔可夫性：

$$P(Y_i|X, Y_1, Y_2, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n) = P(Y_i|X, Y_{i-1}, Y_{i+1}) \quad (9)$$

$$i = 1, 2, \dots, n (\text{在 } i = 1 \text{ 和 } n \text{ 时只考虑单边}) \quad (10)$$

2). 条件随机场的参数化形式

则称 $P(Y|X)$ 为线性链条件随机场。在标注问题中， X 表示输入观测序列， Y 表示对应的输出标记序列或状态序列。

根据 Hammersley-Clifford 定理，我们可以给出线性链条件随机场的因子分解式，各个因子是定义在相邻两个结点（最大团）上的势函数。

C: 定理（线性链条件随机场的参数化形式） 设 $P(Y|X)$ 为线性链条件随机场，在随机变量 X 取值为 x 的条件下，随机变量 Y 取值为 y 的条件概率有如下形式：

$$P(y|x) = \frac{1}{Z(x)} \exp \left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i) \right) \quad (11)$$

其中，

$$Z(x) = \sum_y \exp \left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i) \right) \quad (12)$$

t_k, s_l 为特征函数， λ_k, μ_l 是对应的权值。条件随机场完全由特征函数和其对应的权值确定。 t_k 是转移特征， s_l 是状态特征，取值为 1 或 0。

3). 条件随机场的简化形式

我们可以用简化形式表现条件随机场，将其写成权值向量和特征向量的内积形式，即条件随机场的简化形式。我们设有 K_1 个转移特征， K_2 个状态特征， $K = K_1 + K_2$ ，记：

$$f_k(y_{i-1}, y_i, x, i) = \begin{cases} t_k(y_{i-1}, y_i, x, i), & k = 1, 2, \dots, K_1 \\ s_l(y_i, x, i), & k = K_1 + l; l = 1, 2, \dots, K_2 \end{cases} \quad (13)$$

然后对转移和状态特征在各个位置求和，记：

$$f_k(y, x) = \sum_{i=1}^n f_k(y_{i-1}, y_i, x, i), k = 1, 2, \dots, K \quad (14)$$

用 w_k 表示特征 $f_k(y, x)$ 的权值。即：

$$w_k = \begin{cases} \lambda_k, & k = 1, 2, \dots, K_1 \\ \mu_l, & k = K_1 + l; l = 1, 2, \dots, K_2 \end{cases} \quad (15)$$

条件随机场可表示为:

$$P(y|x) = \frac{1}{Z(x)} \exp \sum_{k=1}^K w_k f_k(y, x) \quad (16)$$

$$Z(x) = \sum_y \exp \sum_{k=1}^K w_k f_k(y, x) \quad (17)$$

令权值向量 $w = (w_1, w_2, \dots, w_K)^\top$, 全局特征向量 $F(y, x) = (f_1(y, x), f_2(y, x), \dots, f_K(y, x))^\top$ 。

条件随机场可以写成:

$$P_w(y|x) = \frac{\exp(w \cdot F(y, x))}{Z_w(x)} \quad (18)$$

其中,

$$Z_w(x) = \sum_y \exp(w \cdot F(y, x)) \quad (19)$$

4). 条件随机场的矩阵形式

假设 $P_w(y|x)$ 是线性链条件随机场, 表示给定观测序列 x 相应的标记序列 y 的条件概率。对每个标记序列我们引入特殊的起点和终点标记 $y_0 = \text{start}$, $y_{n+1} = \text{stop}$, 这时标注序列的概率 $P_w(y|x)$ 可以通过矩阵形式表示并且有效计算。

对 x 的每一个位置 $i = 1, 2, \dots, n+1$, 由于 y_{i-1} 和 y_i 在 m 个标记中取值, 可以定义一个 m 阶的矩阵随机变量, 表示两两的状态转移矩阵:

$$M_i(x) = [M_i(y_{i-1}, y_i | x)] \quad (20)$$

矩阵随机变量的元素为

$$M_i(y_{i-1}, y_i | x) = \exp(W_i(y_{i-1}, y_i | x)) \quad (21)$$

$$W_i(y_{i-1}, y_i | x) = \sum_{k=1}^K w_k f_k(y_{i-1}, y_i, x, i) \quad (22)$$

f_k 和 w_k 由式 (12) 和 (14) 计算。 y_{i-1} 和 y_i 表示标记随机变量 Y_{i-1} 和 Y_i 的取值。于是, 给定 x , 标记序列 y 的规范化条件概率, $P_w(y|x)$ 可以通过矩阵乘积表示:

$$P_w(y|x) = \frac{1}{Z_w(x)} \prod_{i=1}^{n+1} M_i(y_{i-1}, y_i | x) \quad (23)$$

其中

$$Z_w(x) = [M_1(x) M_2(x) \cdots M_{n+1}(x)]_{\text{start}, \text{stop}} \quad (24)$$

2. 问题求解

解: 参考图 11.6, 绘制出 $y_0 = \text{start} = 2, y_4 = \text{stop} = 2$ 时的状态路径图如下所示。

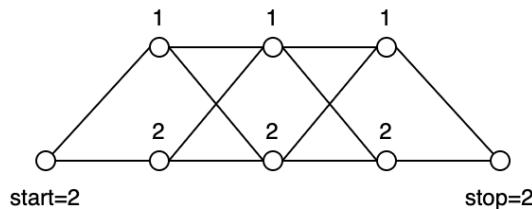


图 1: 状态路径图

我们首先计算图 1 中从 start 到 stop 对应于所有可能的 $2^3 = 8$ 种观测序列 $y = (1, 1, 1), y = (1, 1, 2), y = (1, 2, 1), y = (1, 2, 2), y = (2, 1, 1), y = (2, 1, 2), y = (2, 2, 1), y = (2, 2, 2)$ 各路径的非规范化概率分别为:

$$\begin{aligned}y &= (1, 1, 1) \text{ 时}, \quad 0.5 \times 0.3 \times 0.5 = 0.075 \\y &= (1, 1, 2) \text{ 时}, \quad 0.5 \times 0.3 \times 0.5 = 0.075 \\y &= (1, 2, 1) \text{ 时}, \quad 0.5 \times 0.7 \times 0.6 = 0.21 \\y &= (1, 2, 2) \text{ 时}, \quad 0.5 \times 0.7 \times 0.4 = 0.14 \\y &= (2, 1, 1) \text{ 时}, \quad 0.5 \times 0.7 \times 0.5 = 0.175 \\y &= (2, 1, 2) \text{ 时}, \quad 0.5 \times 0.7 \times 0.5 = 0.175 \\y &= (2, 2, 1) \text{ 时}, \quad 0.5 \times 0.3 \times 0.6 = 0.09 \\y &= (2, 2, 2) \text{ 时}, \quad 0.5 \times 0.3 \times 0.4 = 0.06\end{aligned}$$

计算 $Z_w(x)$:

$$\begin{aligned}Z_w(x) &= [M_1(x) M_2(x) \cdots M_{n+1}(x)]_{2,2} = \sum_y P(y) \\&= 0.075 + 0.075 + 0.21 + 0.14 + 0.175 + 0.175 + 0.09 + 0.06 \\&= 1\end{aligned}$$

故得 $P_w(y|x)$ 。以 start = 2 为起点, stop = 2 为终点的所有路径的状态序列 y 的概率为:

$$P_w(y|x) = \frac{1}{Z_w(x)} \prod_{i=1}^{n+1} M_i(y_{i-1}, y_i | x)$$

$$\begin{aligned}P_w(y = (1, 1, 1)|x) &= 0.5 \times 0.3 \times 0.5 = 0.075 \\P_w(y = (1, 1, 2)|x) &= 0.5 \times 0.3 \times 0.5 = 0.075 \\P_w(y = (1, 2, 1)|x) &= 0.5 \times 0.7 \times 0.6 = 0.21 \\P_w(y = (1, 2, 2)|x) &= 0.5 \times 0.7 \times 0.4 = 0.14 \\P_w(y = (2, 1, 1)|x) &= 0.5 \times 0.7 \times 0.5 = 0.175 \\P_w(y = (2, 1, 2)|x) &= 0.5 \times 0.7 \times 0.5 = 0.175 \\P_w(y = (2, 2, 1)|x) &= 0.5 \times 0.3 \times 0.6 = 0.09 \\P_w(y = (2, 2, 2)|x) &= 0.5 \times 0.3 \times 0.4 = 0.06\end{aligned}$$

3. 代码验证

```
import numpy as np

# y可能的取值为1, 2
# 用Y存储所有可能的观测序列y
Y= np.array([[1,1,1],[1,1,2],[1,2,1],[1,2,2],[2,1,1],[2,1,2],[2,2,1],[2,2,2]])
# 矩阵M1, M2, M3, M4
M1 = np.array([[0,0],[0.5,0.5]])
M2 = np.array([[0.3,0.7],[0.7,0.3]])
M3 = np.array([[0.5,0.5],[0.6,0.4]])
M4 = np.array([[0,1],[0,1]])
# start = 2,stop = 2
```

```

start = 2
stop = 2
def crf_matrix(Y,M1,M2,M3,M4,start,stop):
    # 计算规范化因子
    Z = (M1@M2@M3@M4)[start-1][stop-1]
    print(f'Z={Z}')
    # 用于存储概率最大的状态序列
    max_y = [1,1,1]
    max_P = 0
    # 对每个可能的序列，计算规范化概率
    for y in Y:
        # 对每个可能的观测序列
        print(f'y={y},',end = '')
        P_unnormalized= M1[2-1][y[0]-1]*M2[y[0]-1][y[1]-1]*M3[y[1]-1][y[2]-1]*M4[y[2]-1][2-1]
        print(f'非规范化概率: {P_unnormalized:.3f},',end=' ')
        P_normalized= P_unnormalized/Z
        print(f'规范化概率: {P_normalized:.3f}.')
        if P_normalized>max_P:
            max_P=P_normalized
            max_y=y
    print(f'概率最大的状态序列为: {max_y},最大概率为: {max_P}')
    return max_y,max_P
if __name__ == '__main__':
    max_y,max_P = crf_matrix(Y,M1,M2,M3,M4,start,stop)

```

输出结果

```

Z=1.0
y=[1 1 1],非规范化概率: 0.075,规范化概率: 0.075.
y=[1 1 2],非规范化概率: 0.075,规范化概率: 0.075.
y=[1 2 1],非规范化概率: 0.210,规范化概率: 0.210.
y=[1 2 2],非规范化概率: 0.140,规范化概率: 0.140.
y=[2 1 1],非规范化概率: 0.175,规范化概率: 0.175.
y=[2 1 2],非规范化概率: 0.175,规范化概率: 0.175.
y=[2 2 1],非规范化概率: 0.090,规范化概率: 0.090.
y=[2 2 2],非规范化概率: 0.060,规范化概率: 0.060.
概率最大的状态序列为: [1 2 1],最大概率为: 0.21

```

4. 结论

概率最大的状态序列为 $y = (1, 2, 1)$, 概率 $P_w(y = (1, 2, 1)|x) = 0.21$ 。