

模糊深隐马尔可夫模型研究

成科扬 文传军 詹永照

(江苏大学计算机学院 镇江 212013)

摘要 针对经典马尔可夫模型没有考虑模型应用中状态、观测量间的上下文相关性以及状态转移概率动态性、可变性,提出一种模糊深隐马尔可夫模型。该模型通过增加观测值间的相关性、解决概率转移问题中的不确定性和改进参数优化算法,使之能够较好地应用于强噪声、训练数据缺损等情形的模式识别中。理论证明,显式模糊深隐马尔可夫模型在同等模型复杂度下具有模型优化程度高、区分度好、误识率低、鲁棒性高的特性。

关键词 模糊深隐马尔可夫模型,最大互信息,模糊随机变量

Research on Fuzzy Buried Markov Model

CHENG Ke-yang WEN Chuan-jun ZHAN Yong-zhao

(School of Computer Science and Telecommunications Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China)

Abstract To avoid the disadvantage of hidden Markov model which doesn't consider the contextual relevant relationship of states and observations and the changeability of transfer probability, an improved model called fuzzy buried Markov model is put forward in this paper. Adding relationship among the different observations, resolving the problem of transfer probability uncertainty and ameliorating the parameter optimization arithmetic make the novel model be suitable to apply in pattern identification with much noise and losing of some training data. Compared with other model of the same complexity, fuzzy buried Markov model shows more good character such as optimum performance, better division degree, lower error rate and lustihood, which can be proved by graphical theory.

Keywords Fuzzy buried markov model, Maximum mutual information, Fuzzy random variable

1 引言

隐马尔可夫模型(Hidden Markov Model, HMM)作为一种用参数表示、用于描述随机过程统计特性的概率模型,以其十分健壮的数学结构和强大的模式分类能力,成为模式识别中的重要方法,被广大研究人员广泛地使用和研究^[1]。

然而,由于隐马尔可夫模型没有考虑应用中状态及观测量间的上下文相关性、转移概率动态性、可变性、非精确性和观测值可能存在缺值等不确定性因素,往往在具有较强噪声的模式识别中效果不佳。本文所提出的模糊深隐马尔可夫模型正是通过改良经典隐马尔可夫模型,力图进一步提高识别系统的识别率和鲁棒性。

2 模糊深隐马尔可夫模型

2.1 经典隐马尔可夫模型存在的问题

由于传统经典马尔可夫模型存在三个重要假设:

对于一个随机事件,有一个观测值序列 $X(t_1), \dots, X(t_T)$, 该事件隐含着—个状态序列 $Q(t_1), \dots, Q(t_T)$ 存在:

(1) 马尔可夫假设(状态构成—阶马尔可夫链) $p(Q_t | Q_{t-1} \dots Q_1) = p(Q_t | Q_{t-1})$;

(2) 不动性假设(状态与具体时间无关) $p(Q_t | Q_t) |_{t=m} = p(Q_t | Q_t) |_{t=n}$, 对任意 m, n 成立;

(3) 输出独立性假设(输出仅与当前状态有关) $p(X_1, \dots, X_T | Q_1, \dots, Q_T) = \prod p(X_t | Q_t)$ 。

三个假设—方面大大简化了模型的复杂性,节省了运算时间,推动了模型的应用;另一方面,由于以上模型假设的存在,导致了模型精度下降,限制了模型的应用范围。这是因为

经典隐马尔可夫模型没有考虑模型应用中状态及观测量间的上下文相关性以及状态转移概率的动态性、可变性。

本文提出的模糊深隐马尔可夫模型(Fuzzy Buried Markov Model, FBMM),将针对隐马尔可夫模型中观测值间相关性和状态、观测量及其关系可变性做进一步拓展,并改进参数优化算法,以期在高噪声背景或某些观测值缺省的情形下也能保证模型的有效性。

2.2 模糊深隐马尔可夫模型在观测值间相关性上的改进

经典隐马尔可夫模型是双重随机过程,状态间的转移是—重随机过程,特定状态下的以何种观测值出现是另一重随机过程。然而这种模型结构却割裂了不同时刻状态下观测值间的联系。模糊深隐马尔可夫模型正是基于此点考虑,增加了对观测值间相关性关系的研究:考虑到这种观测值间的相关性,应当是①建立在观测值间的,而不是观测值与状态间;②只建立在上下文有关的观测值间;③对应于特定的状态 $Q(t)$, 这种相关性关系应是特定的。因此,我们定义^[2]:对于特定的状态 $Q(t)$, 与当前观测值具有相关性的观测值集合为该观测值的观测值依赖集合,记为 $Z(Q(t))$ 。对于 $Z(Q(t))$ —般地有

$$Z(Q(t)) \subseteq \{X_1 \dots X_{t-1}, X_{t+1} \dots X_n\}$$

特别地,仅考虑上文相关性 $Z(Q(t)) \subseteq \{X_1 \dots X_{t-1}\}$;

更特别地,仅考虑最近、次近上文相关性 $Z(Q(t)) \subseteq \{X_{t-2}, X_{t-1}\}$ 。

如前文所述,对于特定时刻状态下的观测值由该时刻的状态和其前后时刻具有相关性的观测值依赖集合决定。但实质上,特定时刻状态下的观测值也应与前后时刻的相关状态集合相关。但我们考虑到观测值是易测的,而状态值隐藏较

* 国家自然科学基金资助项目(60673190)。成科扬 硕士,研究方向为模式识别;文传军 博士,主要研究方向为人工智能;詹永照 教授,博士生导师,主要研究方向为模式识别、计算机图像学。

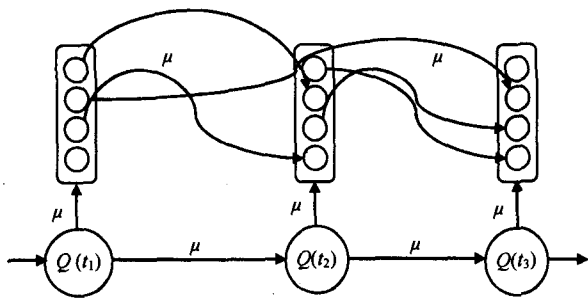


图1 模糊深隐马尔可夫模型结构示意图

深不易观测,并且这种前后时刻状态集合与当前时刻观测值的相关性可以表现为前后时刻观测值集合与当前时刻观测值间的相关性,因此我们所寻求的集合即为互信息(其定义将于下文给出):

$I(X(t); Z(Q(t)) | Q(t))$ 尽量大, $I(X(t); Z(Q(t)) | Q \in C_{Q(t)})$ 尽量小的 $Z(Q(t))$ 集合,即

$$Z(Q(t)) = \arg \max I(X(t); Z(Q(t)) | Q(t))$$

$$Z(Q(t)) = \arg \min I(X(t); Z(Q(t)) | Q \in C_{Q(t)})$$

其中 $C_{Q(t)}$ 是与 $Q(t)$ 类似,且能对 $X(t)$ 产生影响的非 t 时刻状态集合。

以下是 $Z(w(t))$ 的求解算法^[3]:

```

1 begin Initialize  $Z(Q(t)) = \{ \}$ 
2 While  $i < T$  (前溯的最大时间)
3   while  $j < N$  (观测值的最大个数)
4     if  $\{ I(X(t); X_j(i) | Q(t)) > \theta_1 \}$ 
5       &  $\{ I(X(t); X_j(i) | Q \in C_{Q(t)}) < \theta_2 \}$ 
6       &  $\{ I(X(t); X_j(i) | Q(t)) < \theta_3 I(X(t); Z(Q(t)) | Q(t)) \}$ 
7     }
8     then add  $X_j(i)$  to  $Z(Q(t))$ 
9 return
10 end

```

其中, $\theta_1, \theta_2, \theta_3$ 分别为对应阈值,其取值视具体应用而定。

2.3 模糊深隐马尔可夫模型在解决概率不确定性上的改进

从信息论的角度,存在两种不确定性:随机性和模糊性。随机性用随机变数来描述,模糊性用模糊集来描述。当所考虑的事件模糊性和随机性共存时,用模糊随机变量可以同时刻画这两种不确定性^[4](模糊随机变量最早由 Kwakernaak 在 1978 年引入)。

定义 对于给定概率测度空间中的任意 ω ,若总有 $\mu(\omega) \in [0, 1]$,则称 ω 为映射 $X: \omega \rightarrow \mu(\omega)$ 对应的模糊随机变量^[5]。

对 $\forall a \in [0, 1]$,存在 $\mu_a^l(\omega), \mu_a^u(\omega)$ 使得

$$\mu_a^l(\omega) = \inf \{ x \in \mathbb{U} | \mu(\omega) \geq x \}$$

$$\mu_a^u(\omega) = \sup \{ x \in \mathbb{U} | \mu(\omega) \geq x \}$$

则称 $\mu_a^l(\omega), \mu_a^u(\omega)$ 是 $\mu(\omega)$ 置信水平为 a 的上限和下限。

经典隐马尔可夫模型中两重随机过程中转移概率均为确定的概率值。但由于隐马尔可夫模型中的状态是人为划分的,而实际上各状态间往往是模糊无确定边界的,因此这种状态间的转移应该包含状态界限模糊性和状态转移随机性两重不确定性。另外,在如人脸表情识别、语音识别等应用中,这种概率的精确值往往无法求得,而且这些概率转移也会因时刻不同而有所变化^[6]。另外,由于待识样本的原因,有时一些训练数据还会缺失,从而无法直接用来求取相关参数。由此,我们引入模糊随机变量,即定义概率转移值为一概率区间, $p_{ij} \in [p_{ij}^l, p_{ij}^u]$,其中的 p_{ij}^l 和 p_{ij}^u 分别代表给定概率区间的上限和下限。

为了将问题转化为经典隐马尔可夫模型方法能够解决的问题,必须对模糊随机变量进行去模糊化。考虑到去模糊化后的概率值必须尽可能地反映去模糊化前的模糊随机变量,

而且能够反映实际问题的性质和便于计算,我们采用区间排序法来解决这一问题。

Givan 给出的两种不同的区间排序方法^[7]:

$$([l_1, u_1] \leq [l_2, u_2])? (u_1 \leq u_2 \text{ or } (u_1 = u_2 \wedge l_1 \leq l_2))$$

$$([l_1, u_1] \leq [l_2, u_2])? (l_1 \leq l_2 \text{ or } (l_1 = l_2 \wedge u_1 \leq u_2))$$

显然,上述两种排序都是部分序关系。Givan 证明了在这两排序方法下至少存在一个最优策略。另外,我们可以采用如下一种区间排序方法:

$$([l_1, u_1] \leq [l_2, u_2])? (\frac{l_1 + u_1}{2} \leq \frac{l_2 + u_2}{2})$$

这里,本文以 $([l_1, u_1] \leq [l_2, u_2])? (u_1 \leq u_2 \text{ or } (u_1 = u_2 \wedge l_1 \leq l_2))$ 为例说明区间排序去模糊的基本算法。

设集合 $R = \{ j_k | k=1, \dots, n \}$ 表示状态 i 对应的所有可转移到达状态。首先对集合 R 进行升序排列。因为 p_{ij} 在不确定模型中是以概率区间存在的,即 p_{ij} 属于 $[p_{ij}^l, p_{ij}^u]$,所以我们给序列中第一个状态 j_1 分配尽可能高的概率值,然后是 j_2 ,直至最后一个状态 j_n 。此过程等价于寻找一个最高的 $r \in [1, \dots, n]$, r 满足如下等式:

$$\sum_{k=1}^r p_{ij_k}^l + \sum_{k=r+1}^n p_{ij_k}^u \leq 1$$

我们给排在 j_r 前面的状态分配对应概率区间的上限,给排在 j_r 后面的状态分配对应概率区间的下限,最后将剩余的

$$p_{ij_r} = \begin{cases} p_{ij_r}^l & \text{if } k < r \\ p_{ij_r}^u & \text{if } k > r \end{cases}$$

$$p_{ij_r} = 1 - \sum_{k=1, k \neq r}^n p_{ij_k}$$

具体求解算法如下:

```

1 Initialize  $R = \{ j_k | k=1, \dots, n \}$ ;
2 对集合  $R$  进行升序排列;
3  $k := 1$ , Bound :=  $\sum_{k=1}^n p_{ij_k}^u$ 
4 While ( Bound -  $p_{ij_k}^l + p_{ij_k}^u < 1$ ) do
5   Bound := bound -  $p_{ij_k}^l + p_{ij_k}^u$ 
6    $p_{ij_k} := p_{ij_k}^l$ 
7    $K := k + 1$ 
8 end while
9  $r := k$ 
10  $p_{ij_r} := 1 - (\text{Bound} - p_{ij_r}^l)$ 
11 for all  $k \in \{ r+1, \dots, n \}$  do
12    $p_{ij_k} := p_{ij_k}^u$ 
13 end for
14 return (R,  $p_{ij_k}$ )

```

通过以上算法可计算出去模糊化后的状态转移概率矩阵 A ,同理可计算出去模糊化后的观测值概率分布矩阵 B 。

对于训练数据缺值问题,我们使用加权模糊填充法增补缺值。

设 $D = (D_1, D_2, \dots, D_n)$ 是训练数据集,对于其中任一样本 D_i, X_i 是 D_i 中所有缺值的集合。设 $\hat{\theta}$ 是关于参数 θ 的当前估计, D' 是基于 $\hat{\theta}$ 利用加权模糊填充法将 D 修补而得到的训练数据集。

定义 θ 的基于 D' 的对数似然函数为

$$Q(\theta | D') = Q(\theta | \hat{\theta}) = \sum_{i=1}^n \sum_{x_k \in \Omega_{X_i}} p(X_i = x_k | D_i, \hat{\theta}) \log p(D_i, X_i = x_k | \theta) \quad (x_k \text{ 是 } X_i \text{ 取值空间 } \Omega_{X_i} \text{ 中的一候选值})$$

这里, $p(X_i = x_k | D_i, \hat{\theta})$ 就是增补 $X_i = x_k$ 的权重 w_{x_k} 。当 $X_i = \emptyset$ 时,约定 $w_{x_k} = 1$, $\hat{\theta}$ 的值可由如下 EM 算法公式给出:

$$\text{若 } \sum_{i=1}^n \sum_{x_k \in \Omega_{X_i}} p(X_i = x_k | D_i, \hat{\theta}) > 0$$

$$\hat{\theta}^t = \frac{\sum_{i=1}^{r_i} \sum_{x_k \in \Omega X_i} p(X_i = x_k | D_i, \hat{\theta}^{t-1})}{\sum_{k=1}^{r_i} \sum_{x_k \in \Omega X_i} p(X_i = x_k | D_i, \hat{\theta}^{t-1})}$$

否则 $\hat{\theta}^t = \frac{1}{r_i}$

r_i 是 X_i 取值空间中候选者的个数。

(注:此时, $Q(\theta|\hat{\theta}^t)$ 达最大值)

因此,对于缺失的 X_i 可由 $X_i = \sum w_{x_k} x_k$ 增补填充。

2.4 模糊深隐马尔可夫模型在参数优化算法上的改进

由于 HMM 中 Baum-Welch 算法实际上是一种最大似然 (Maximum Likelihood, ML) 参数估计方法,即给定训练数据 O ,使 $P(O|\lambda)$ 最大时求出参数 λ 。而最大似然估计并不是唯一的,也不是所有情况下都适用的。特别是当初始模型选择不恰当时,其所得的局部最大将与全局最大相距甚远。这里,FBMM 采用最大互信息 (Maximum Mutual Information, MMI) 参数估计法进行参数优化,因为当事先假定模型不正确时,MMI 估值器要明显优于 ML 估值器^[8]。

对于训练序列 O 和模型 λ ,互信息定义为 $I(\lambda, O) = \lg \frac{p(O, \lambda)}{p(O)p(\lambda)} = \lg \frac{p(O|\lambda)}{p(O)} = \lg p(O|\lambda) - \lg p(O)$,令 λ_i 表示第 i 个候选模型,则

$$\begin{aligned} \hat{\lambda}_{MMI} &= \arg \max_{\lambda_i} \lg p(O|\lambda_i) - \lg p(O) \\ &= \arg \max_{\lambda_i} \lg p(O|\lambda_i) - \sum_j \lg p(O|\lambda_j) p(\lambda_j) \end{aligned}$$

当 $p(\lambda_j)$ 为均匀分布时,

$$\hat{\lambda}_{MMI} = \arg \max_{\lambda_i} \lg p(O|\lambda_i) - \sum_j \lg p(O|\lambda_j)$$

与最大似然参数估计方法 $\hat{\lambda}_{ML} = \arg \max_{\lambda_i} \lg p(O|\lambda_i)$ 比

较,很明显最大互信息方法不仅要求其所选模型的似然概率尽量大,而且要求其他备选模型概率尽量低,这就能保证优化结果是全局最优。

设 θ 表示模型 λ 中的参数, $\hat{\theta}$ 是 θ 的当前估计,则参数重估公式可定义为

$$\hat{\theta}^{t+1} = \hat{\theta}^t + \epsilon \frac{\partial \hat{\lambda}_{MMI}}{\partial \hat{\theta}^t}$$

其中优化因子 ϵ 可取 $1/t$,以保证算法的收敛性。

3 相关性能分析及理论证明

性质 1 具有等量时间步、状态数的 FBMM 和 HMM, FBMM 的计算复杂度与 HMM 的计算复杂度相比较无明显增长。

性质 1 证明的预备知识:

定义 1 对于一有向图,依次用无向边连接每两个具有共同子结点的结点,并去除所有边的方向而得到的无向图称为该有向图的端正图(moral graph),这一过程称为有向图的端正化^[9]。

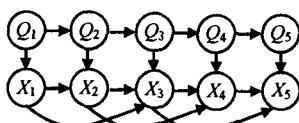


图2 FBMM的结构图

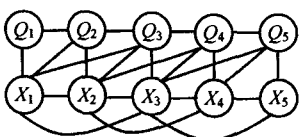


图3 FBMM的端正图

图 2,3 为一 FBMM 的结构图和端正图。

定义 2 对于有向图 G ,以下算法的输出结果定义为该有向图的消元复杂度^[10]。

输入:有向图 G ,消元顺序集 ρ

- 步骤:
1. $G^m \leftarrow G$ 的端正图
 2. 消元复杂度 $\theta \leftarrow 0$
 3. While($\rho \neq \Phi$)
 4. 从 ρ 队头移出一变量 z
 5. $\theta \leftarrow \theta + |z| \prod_{X \in nb(z)} |X|$

(注: X 为端正图 G^m 中与 z 相连的变量集合, $|t|$ 为变量 t 的取值维度)

6. End while

输出:有向图 G 的消元复杂度 θ

性质 1 证明:

设存在一时间步为 T 、状态数为 N 的 HMM 和一时间步为 T 、状态数为 N 、 K 步内上文相关的 FBMM。

先求时间步为 T 、状态数为 N 的 HMM 的计算复杂度。

设该 HMM 的结构图为 G_1 ,其消元顺序集为 $\{Q_1, X_1, Q_2, \dots, Q_T, X_T\}$

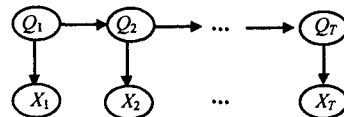


图4 HMM的结构图 G_1

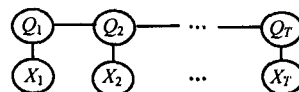


图5 G_1 的端正图 G_1^m

消元:

第一步消去 Q_1 ,此时消元复杂度

$$\theta = |Q_1| * |Q_2| * |X_1| = N^2$$

第二步消去 X_1 ,此时消元复杂度

$$\theta = N^2 + |X_1| = N^2 + 1$$

第三步消去 Q_2 ,此时消元复杂度

$$\theta = N^2 + 1 + |Q_2| * |Q_3| * |X_2| = 2N^2 + 1$$

第四步消去 X_2 ,此时消元复杂度

$$\theta = 2N^2 + 1 + |X_2| = 2N^2 + 2 = 2(N^2 + 1)$$

...

第 $2T-1$ 步消去 Q_T ,此时消元复杂度

$$\theta = (T-1)(N^2 + 1) + N^2$$

第 $2T$ 步消去 Q_T ,此时消元复杂度

$$\theta = T(N^2 + 1)$$

由此可见时间步为 T 、状态数为 N 的 HMM 的计算复杂度为 $O(TN^2)$

再求时间步为 T 、状态数为 N 、 K 步内上文相关的 BMM 的计算复杂度。

设该 FBMM 的结构图为 G_2 ,其消元顺序集为 $\{Q_1, X_1, Q_2, \dots, Q_T, X_T\}$

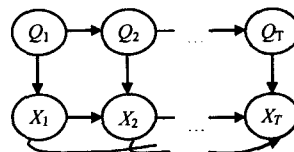


图6 FBMM的结构图 G_2

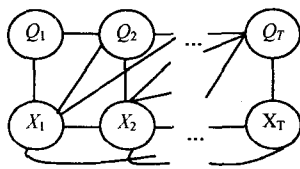


图7 G_2 的端正图 G_2^m

消元:

第一步消去 Q_1 , 此时消元复杂度为 $\theta = |Q_1| * |Q_2| * |X_1| = N^2$

第二步消去 X_1 , 此时消元复杂度 $\theta = N^2 + |Q_2| * |Q_3| * \dots * |Q_{1+k}| * |X_1| * |X_2| * \dots * |X_{1+k}| = N^2 + N^K$

第三步消去 Q_2 , 此时消元复杂度 $\theta = N^2 + N^K + |Q_2| * |Q_3| * |X_2| = N^2 + N^K + N^2 = 2N^2 + N^K$

第四步消去 X_2 , 此时消元复杂度 $\theta = 2N^2 + N^K + |Q_3| * |Q_4| * \dots * |Q_{2+k}| * |X_2| * |X_3| * \dots * |X_{2+k}| = 2(N^2 + N^K)$

...

第 $2T-1$ 步消去 Q_T , 此时消元复杂度

$$\theta = TN^2 + (T-1)N^K$$

第 $2T$ 步消去 Q_T , 此时消元复杂度

$$\theta = T(N^2 + N^K)$$

由于一般情况下 K 取值比较小, 如 $K=1, 2$, 因此一般情况下时间步为 T 、状态数为 N 、 K 步内上文相关的 FBMM 的计算复杂度也为 $O(TN^2)$

性质 1 得证。

性质 2 对相同状态数、观测值变量的 FBMM 和 HMM 利用同一训练数据训练, FBMM 的优化程度优于 HMM。

性质 2 证明预备知识:

定义 3 设 X 为一随机变量, 定义 $-\int p(X) \log p(X) dX$ 为 X 的信息熵, 记为 $H(X)$ 。

定义 4 设 X, Y 为两个随机变量, 定义 $-\int p(X, Y) \log p(X, Y) dXdY$ 为 X, Y 的联合熵, 记为 $H(X, Y)$ 。

定义 5 设 X, Y 为两个随机变量, 定义 $H(X) - H(X|Y)$ 为 X, Y 之间的互信息, 记为 $I(X, Y)$, $I(X, Y)$ 是对 Y 支持 X 信度的一个度量。

定义 6 设 X, Y, Z 为三个随机变量, 定义 $H(X|Z) - H(H|Z, Y)$ 为给定 Z 时 X 和 Y 之间的条件互信息, 记为 $I(X, Y|Z)$, $I(X, Y|Z)$ 是给定 Z 时 Y 支持 X 信度的度量。

定义 7 对定义于随机变量 X 的状态空间 R 上的两个概率分布 $P(X)$ 和 $Q(X)$, 定义 $\int p(X) \log \frac{p(X)}{Q(X)} dX$ 为 $P(X)$ 对 $Q(X)$ 的相对熵, 记为 $KL(P, Q)$, 也称为 $P(X)$ 和 $Q(X)$ 之间的 Kullback-Leibler 距离^[11]。

定理 1 设 $P(X)$ 和 $Q(X)$ 为定义在某个变量 X 的状态空间 R 上的两个概率分布, 则必有 $KL(P, Q) \geq 0$ 。

性质 2 证明:

对于 HMM, 由于观测值变量只与当前状态变量相关, 因此状态变量 Q 对观测值变量 X 的支持度由 X, Q 间的互信息表征:

$$I(X; Q) = H(X) - H(X|Q)$$

对于 FBMM, 由于观测值变量与相关上文观测值变量集合和当前状态变量相关, 因此对于给定状态变量 Q 时, 相关上文观测值变量集合 Z 对观测值变量 X 的支持度由给定 Q 时 Z 和 X 之间的条件互信息表征:

$$I(X; Z|Q) = H(X|Q) - H(X|Z, Q)$$

由于

$$\begin{aligned} I(X; Z|Q) - I(X; Q) &= (H(X|Q) - H(X|Z, Q)) - (H(X) - H(X|Q)) \\ &= 2H(X|Q) - H(X) - H(X|Z, Q) \\ &= -2 \int p(X, Q) \log p(X|Q) dXdQ + \int p(X) \log p(X) dX + \int p(X, Z, Q) \log p(X|Z, Q) dXdZdQ \\ &= - \int p(Z, Q) \log p(X|Q) dXdQ + \int p(X, Q) \log p(X) dXdQ - \int p(X, Z, Q) \log p(X|Q) dXdZdQ + \int p(X, Z, Q) \log p(X|Z, Q) dXdZdQ \\ &= \int p(X, Q) \log \frac{p(X)}{p(X|Q)} dXdQ + \int p(X, Z, Q) \log \frac{p(X|Z, Q)}{p(X|Q)} dXdZdQ \\ &= \int p(Q|X) p(X) \log \frac{p(X)}{p(X|Q)} dXdQ + \int p(Z, Q) p(X|Z, Q) \log \frac{p(X|Z, Q)}{p(X|Q)} dXdZdQ \\ &= KL(p(X), p(X|Q)) \int p(Q|X) dXdQ + KL(p(X|Z, Q), p(X|Q)) \int p(Z|Q) dZdQ \quad (*) \end{aligned}$$

因为 $KL(X, Y) \geq 0, \int p(Q|X) dXdQ \geq 0, \int p(Z, Q) dZdQ \geq 0$, 所以 $(*) \geq 0$ 。

这就证明了增加了观测值间依赖性的 FBMM 较之于 HMM 能提供更多的互信息, 从而对于同一训练数据, FBMM 模型能更有效地优化模型。

性质 3 具有等量状态数、观测值变量的 FBMM 和 HMM, FBMM 的模型区分度优于 HMM, 误识率低于 HMM。

性质 3 证明:

对于 FBMM,

设 $I_q(X_t; Z(q) | Q = q)$ 表示给定 t 时刻的状态 $Q = q$ 时, $Z(q)$ 与观测值变量 X 间的互信息 $I_r(X_t; Z(r) | Q = r)$ (其中 $r \neq q$) 表示对于非 t 时刻状态 $Q = r$ 时, $Z(r)$ 与 t 时刻观测值变量 X 间的互信息。

根据 FBMM 的定义知 $I_q(X_t; Z(q) | Q = q) > I_r(X_t; Z(r) | Q = r)$, 则要使 FBMM 的误识率低, 则须 $I_q(X_t; Z(q) | Q = q)$ 尽量大, 而 $I_r(X_t; Z(r) | Q = r)$ 尽量小。

对于 HMM:

设 $I_q(X_t; Q = q)$ 表示 t 时刻的状态 $Q = q$ 与观测值变量 X 间的互信息, 设 $I_r(X_t; Q = r)$ (其中 $r \neq q$) 表示非 t 时刻的状态 $Q = q$ 与 t 时刻观测值变量 X 间的互信息。

定义互信息差 $D = I_q - I_r$, 则模型区分度 $\propto D$, 误识率 $\propto \frac{1}{D}$ 。

由于

$$\begin{aligned} D_{\text{FBMM}} &= I_q(X_t; Z(q) | Q = q) - I_r(X_t; Z(r) | Q = r) \\ D_{\text{HMM}} &= I_q(X_t; Q = q) - I_r(X_t; Q = r) \\ D_{\text{FBMM}} - D_{\text{HMM}} &= [I_q(X_t; Z(q) | Q = q) - I_r(X_t; Z(r) | Q = r)] - [I_q(X_t; Q = q) - I_r(X_t; Q = r)] \\ &= [H(X|Q = q) - H(X|Q = q, Z = Z(q))] - [H(X|Q = r) - H(X|Q = r, Z = Z(r))] - [H(X) - H(X|Q = q)] + [H(X) - H(X|Q = r)] = \\ &= 2H(X|Q = q) - H(X|Q = q, Z = Z(q)) + H(X|Q \end{aligned}$$

$$\begin{aligned}
&= r, Z = Z(r) - 2H(X | Q = r) \\
&= 2 \int p(X, Q = q) \log \frac{1}{p(X | Q = q)} dX - \int p(X, Z = (q), \\
&\quad Q = q) \log \frac{1}{p(X | Q = q, Z = Z(q))} dX \\
&+ \int p(X, Q = Z(r), Q = r) \log \\
&\quad \frac{1}{p(X | Q = r, Z = Z(r))} dX - 2 \int p(X, Q = r) \log \\
&\quad \frac{1}{p(X | Q = r)} dX \\
&\geq \int p(X, Q = q) \log \frac{p(X | Q = q, Z = Z(q))}{p(X | Q = q) p(X | Q = q)} dX \\
&+ \int p(X, Z = Z(r), Q = r) \log \frac{1}{p(X | Q = r, Z = Z(r))} dX \\
&- 2 \int p(X, Q = q) \log \frac{1}{p(X | Q = r)} dX \\
&= \int p(X, Q = q) \log \\
&\quad \frac{p(X | Q = q, Z = Z(q)) p(X | Q = r) p(X | Q = r)}{p(X | Q = q) p(X | Q = q)} dX \\
&+ \int p(X, Z = Z(r), Q = r) \log \frac{1}{p(X | Q = r, Z = Z(r))} dX \\
&\geq \int p(X, Z = Z(r), Q = r) \log \\
&\quad \frac{p(X | Q = q, Z = Z(q)) p(X | Q = r) p(X | Q = r)}{p(X | Q = q) p(X | Q = q) p(X | Q = r, Z = Z(r))} dX \\
&\geq \int p(X, Z = Z(r), Q = r) \log \\
&\quad \frac{p(X | Q = r, Z = Z(r)) p(X | Q = r) p(X | Q = r)}{p(X | Q = q) p(X | Q = q) p(X | Q = r, Z = Z(r))} dX \\
&= \int p(X, Z = Z(r), Q = r) \log \\
&\quad \frac{p(X | Q = r) p(X | Q = r)}{p(X | Q = q) p(X | Q = q)} dX \\
&\geq 0
\end{aligned}$$

(注:以上推导中利用了 $p(X, Q = q) \geq p(X, Z = Z(q), Q = q)$, $p(X, Q = q) \geq p(X, Z = Z(q), Q = q) \geq p(X, Z = Z(r), Q = r)$, $p(X | Q = q, Z = Z(q)) \geq p(X | Q = r, Z = Z(r))$, $p(X | Q = q) \geq p(X | Q = r)$, 这是由于 X 是 t 时刻的观测值, $Q = q$ 是 t 时刻的当前状态, 其对应的各相关概率必显然大于非当前时刻 t 对应状态下的各相关概率值)

(上接第 134 页)

网的测试网来进行的。如何使模型在变换前后继续保持原有属性, 是一个需要进一步深入研究的问题。

参考文献

- [1] Girault C, Valk R. Petri Nets for Systems Engineering: A Guide to Modeling, Verification, and Applications. Berlin: Springer-Verlag, 2002
- [2] Murata T. Petri Nets: Properties, Analysis and Applications. Proceedings of the IEEE, 1989, 77(4): 541-580
- [3] 袁崇义. Petri 网原理与应用. 北京: 电子工业出版社, 2005
- [4] Padberg J. Petri net modules. Journal of Integrated Design and Process Science, 2002, 6(4): 105-120
- [5] 王培良, 赵义军, 叶志宝. Petri 网的并分解. 控制理论与应用, 2001, 18(1): 116-119
- [6] Huang H, Cheung T Y, Mak W M. Structure and behavior preservation by Petri-net-based refinements in system design. Theoretical Computer Science, 2004, 328: 245-269
- [7] 蒲飞, 陆维明. 同步合成 Petri 网系统可回复性的保持性. 系统

以上证明显示了 FBMM 模型的区分度优于 HMM, 误识率低于 HMM, 而这两者正是评价模型优劣的主要标准。

性质 4 FBMM 较之于 HMM 具有更高的鲁棒性。

由于 FBMM 采用模糊随机变量刻画状态间、观测值间及其之间的相关性, 其值用具有弹性的区间来表示, 更好地反映了模型应用的实际情况, 并且 FBMM 能够对训练数据中的部分缺失进行优化填充, 因此其鲁棒性优于 HMM 是显见的。

结束语 本文提出一种对经典隐马尔可夫模型改进的模型——模糊深隐马尔可夫模型。该模型充分考虑到观测值间的相关性和转移概率值的不确定性, 通过相关理论证明可知模糊深隐马尔可夫模型与经典隐马尔可夫模型相比较具有在同一训练数据下优化程度高、模式区分度好、鲁棒性强, 且模型复杂度没有明显增加的优点。相信具有这些特性的模糊深隐马尔可夫模型将能够在高性能要求的复杂模式识别中具有广泛的应用前景。

参考文献

- [1] 龚光鲁, 钱敏平. 应用随机过程教程[M]. 北京: 清华大学出版社, 2004: 249-255
- [2] Bilmes J A. Buried Markov models for speech recognition//Proceeding of the IEEE international Conference on Acoustics, Speech, and Signal Processing. Phoenix, AZ, March 1999: 713-716
- [3] Bilmes J A. Buried Markov models- a graphical-modeling approach to automatic speech recognition. Computer Speech and Language, 2003, 17: 213-231
- [4] 冯前进, 陈武凡. 模糊马尔可夫场模型与图像分割新算法[J]. 南方医科大学学报, 2006(5): 579-583
- [5] Avrachenkov K E, Sanchez E. Fuzzy Markov Chains: Specificities and Properties//Proc IPMU. Madrid, Spain, July 2000: 1851-1856
- [6] Tran D, Wagner M. Generalised Fuzzy Hidden Markov Models for Speech Recognition//Proceedings of the International Conference on Fuzzy Systems, Calcutta, India, 2002: 84-91
- [7] Givan R, Leach S, Dean T. Bounded parameter markov decision processes. Artificial Intelligence, 2000, 122(1/2): 71-109
- [8] 茅晓泉, 胡光锐. 基于最大互信息的离散隐马尔可夫模型训练方法[J]. 上海交通大学学报, 2001(11): 1713-1717
- [9] Pearl J. Probabilistic Reasoning in Intelligent Systems: Net2 works of Plausible Inference. San Mateo, California: Morgan Kaufmann, 1988: 71-78
- [10] 张连文, 郭海鹏. 贝叶斯网引论[M]. 北京: 科学出版社, 2006: 36-50
- [11] Information theory as an extension of the maximum likelihood principle//Petrov V, Csaki F, eds. Second International Symposium on Information Theory. Budapest: Akailseoniai Kiado

工程理论与实践, 2005, 25(10): 83-94

- [8] 蒲飞, 陆维明. 同步合成 Petri 网系统活性与无死锁性的保持性. 软件学报, 2003, 14(12): 1977-1988
- [9] Bernadsky M, Sharykin R, Alur R. Formal Techniques, Modeling and Analysis of Timed Systems. Lecture Notes in Computer Science, 2004, 3253: 309-325
- [10] Tsinarakis G J, Valavanis K P. Studying Multi-operational Production Systems with Modular Hybrid Petri Nets//Intl. Conference on Intelligent Robots and Systems. Las Vegas, Nevada, 2003, 3028-3046
- [11] 张继军, 吴哲辉. Petri 网的分层递归模型. 系统仿真学报, 2003, 15(1): 89-92
- [12] Kristensen L M, Christensen S, Jensen K. The practitioner's guide to coloured Petri nets. International Journal on Software Tools for Technology Transfer, Springer, 1998, 2(2): 98-132
- [13] Proth J M. A Class of Petri Nets for Manufacturing System Integration. IEEE Transactions on Robotics and Automation, 1997, 13(3): 317-326
- [14] 夏传良, 焦莉, 陆维明. Petri 网精细化操作及其在系统设计中的应用. 软件学报, 2006, 17(1): 11-20