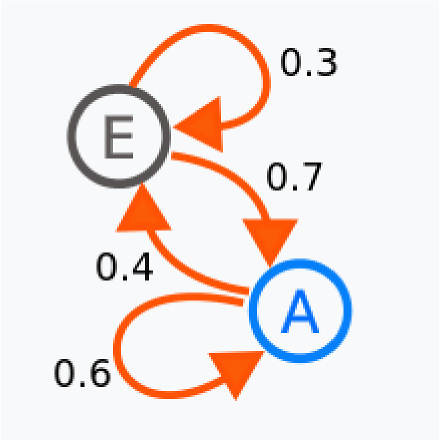
WIKIPEDIA

马尔可夫链

维基百科，自由的百科全书

马尔可夫链（英语：Markov chain），又稱**離散時間馬可夫鏈**（discrete-time Markov chain，縮寫為**DTMC**[1]），因俄國數學家安德烈·马尔可夫（俄语：Андрей Андреевич Марков）得名，为狀態空間中经过从一个状态到另一个状态的转换的随机过程。该过程要求具备“无记忆”的性质：下一状态的概率分布只能由当前状态决定，在时间序列中它前面的事件均与之无关。这种特定类型的“无记忆性”称作馬可夫性質。马尔科夫链作为实际过程的统计模型具有许多应用。



一个具有两个转换状态的马尔可夫链

在马尔可夫链的每一步，系统根据概率分布，可以从一个状态变到另一个状态，也可以保持当前状态。状态的改变叫做转移，与不同的状态改变相关的概率叫做转移概率。随机漫步就是马尔可夫链的例子。随机漫步中每一步的状态是在图形中的点，每一步可以移动到任何一个相邻的点，在这里移动到每一个点的概率都是相同的（无论之前漫步路径是如何的）。

目录

介绍

形式化定义

- 变种

瞬态演变

性质

- 可还原性
- 周期性
- 重现性
- 各态历遍性
- 律动性
- 平稳状态分析和极限分布
- 平稳状态分析和时齐马尔可夫链

有限状态空间

- 稳定分布与特征向量和单纯形的关系
- 有限状态空间内的时齐马尔可夫链
- 趋向稳定分布的收敛速度

可反转马尔可夫链

伯努利方案

一般的状态空间

- Harris链
- 局部交互的马尔可夫链

应用

- 物理
- 化学
- 测试
- 语音识别
- 信息科学
- 排队理论
- 互联网应用
- 统计
- 经济和金融
- 社会科学
- 生物数学
- 遗传学
- 游戏
- 音乐
- 棒球
- 马尔可夫文本生成器

Fitting

历史

隱馬可夫模型

参看

注释

参考文献

外部連結

介绍

马尔可夫链是满足马尔可夫性质的随机过程。

形式化定义

马尔可夫链是满足马尔可夫性质的随机变量序列$X_1, X_2, X_3, \ldots$，即给出当前状态，将来状态和过去状态是相互独立的。从形式上看，

> 如果两边的条件分布有定义（即如果 $\Pr(X_1 = x_1, \ldots, X_n = x_n) > 0$），则
> $\Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_n = x_n)$ 。

X_i的可能值构成的可数集S叫做该链的“状态空间”。

通常用一系列有向图来描述马尔可夫链，其中图n的边用从时刻n的状态到时刻$n+1$的状态的概率$\Pr(X_{n+1} = x \mid X_n = x_n)$来标记。也可以用从时刻$n$到时刻$n+1$的转移矩阵表示同样的信息。但是，马氏链常常被假定为时齐的（见下文的变种），在这种情况下，图和矩阵与n无关，因此也不表现为序列。

这些描述强调了马尔可夫链与初始分布$\Pr(X_1 = x_1)$无关这一结构。当时齐的时候，可以认为马氏链是分配从一个顶点或状态跳变到相邻一个的概率的状态机。可以把机器状态的概率$\Pr(X_n = x | X_1 = x_1)$作为仅有元素x_1的状态空间为输入的机器的统计行为分析，或作为初始分布为$\Pr(X_1 = y) = [x_1 = y]$状态为输入的机器行为，其中$[P]$是艾佛森括号。

一些状态序列可能会有零概率的事件，对应多连通的图，而我们禁止转移概率为0的边。例如，若a到b的概率非零，但a到x位于图的不同连通分量，那么$\Pr(X_{n+1} = b | X_n = a)$有意义，而$\Pr(X_{n+1} = b | X_1 = x, \ldots, X_n = a)$无意义。

变种

- 连续时间马尔可夫过程具有连续索引。
- **时齐马尔可夫链**（或**静态马尔可夫链**）是对于所有n

  $$\Pr(X_{n+1} = x \mid X_n = y) = \Pr(X_n = x \mid X_{n-1} = y)$$

  的过程。转移概率与n无关。

- ***m*阶马尔可夫链**（或记忆为m的马尔可夫链），其中m有限，为满足

  $$\begin{aligned} &\Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1) \\ =&\Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_{n-m} = x_{n-m}) \text{ for } n > m \end{aligned}$$

  的过程。换句话说，未来状态取决于其前m个状态。It is possible to construct a chain（Y_n）from (X_n) which has the 'classical' Markov property by taking as state space the ordered m-tuples of X values, ie. $Y_n = (X_n, X_{n-1}, \ldots, X_{n-m+1})$。

瞬态演变

用n步从状态i到状态j的概率为

$$p_{ij}^{(n)} = \Pr(X_n = j \mid X_0 = i)$$

而单步转移是

$$p_{ij} = \Pr(X_1 = j \mid X_0 = i).$$

对于一个时齐马尔可夫链来说：

$$p_{ij}^{(n)} = \Pr(X_{k+n} = j \mid X_k = i)$$

而且

$$p_{ij} = \Pr(X_{k+1} = j \mid X_k = i).$$

n步转移概率满足Chapman-Kolmogorov等式，对任意k使得$0 < k < n$，

$$p_{ij}^{(n)} = \sum_{r \in S} p_{ir}^{(k)} p_{rj}^{(n-k)}$$

其中S为此马尔可夫链的状态空间。

边缘分布$\Pr(X_n = x)$为第n次状态的分布。初始分布为$\Pr(X_0 = x)$。用一步转移把过程演变描述为

$$\Pr(X_n = j) = \sum_{r \in S} p_{rj} \Pr(X_{n-1} = r) = \sum_{r \in S} p_{rj}^{(n)} \Pr(X_0 = r)。$$

注意：上标（n）是索引而非指数。

性质

可还原性

马尔可夫链是由一个条件分布来表示的

$$P(X_{n+1}|X_n)$$

这被称为是随机过程中的“转移概率”。这有时也被称作是“一步转移概率”。二、三，以及更多步的转移概率可以导自一步转移概率和马尔可夫性质：

$$P(X_{n+2}|X_n) = \int P(X_{n+2}, X_{n+1}|X_n)\, dX_{n+1} = \int P(X_{n+2}|X_{n+1})\, P(X_{n+1}|X_n)\, dX_{n+1}$$

同样，

$$P(X_{n+3}|X_n) = \int P(X_{n+3}|X_{n+2}) \int P(X_{n+2}|X_{n+1})\, P(X_{n+1}|X_n)\, dX_{n+1}\, dX_{n+2}$$

这些式子可以通过乘以转移概率并求$k-1$次积分来一般化到任意的将来时间$n+k$。

周期性

边缘分布$P(X_n)$是在时间为n时的状态的分布。初始分布为$P(X_0)$。该过程的变化可以用以下的一个时间步幅来描述：

$$P(X_{n+1}) = \int P(X_{n+1}|X_n)\,P(X_n)\,dX_n$$

这是Frobenius-Perron equation的一个版本。这时可能存在一个或多个状态分布π满足

$$\pi(X) = \int P(X|Y)\,\pi(Y)\,dY$$

其中Y只是为了便于对变量积分的一个名义。这样的分布π被称作是“平稳分布”（Stationary Distribution）或者“稳态分布”（Steady-state Distribution）。一个平稳分布是一个对应于特征值为1的条件分布函数的特征方程。

平稳分布是否存在，以及如果存在是否唯一，这是由过程的特定性质决定的。“不可约”是指每一个状态都可来自任意的其它状态。当存在至少一个状态经过一个固定的时间段后连续返回，则这个过程被称为是“周期的”。

重现性

各态历遍性

具有重现性而不具有周期性叫做遍历的。重现性包括局部重现性。

律动性

平稳状态分析和极限分布

平稳状态分析和时齐马尔可夫链

有限状态空间

若状态空间是有限的，则转移概率分布可以矩阵表示，该矩阵称为转移矩阵，记为$\mathbf{P}$。其中处于(i,j)的元素等于

$$p_{ij} = \Pr(X_{n+1} = j \mid X_n = i).$$

由于$\mathbf{P}$的每一行各元素之和为1，且$\mathbf{P}$中所有的元素都是非负的，因此$\mathbf{P}$是一个右随机矩阵。

稳定分布与特征向量和单纯形的关系

稳定分布$\boldsymbol{\pi}$是一个（行）向量，它的元素都非负且和为1，不随施加$\mathbf{P}$操作而改变，定义为

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}.$$

通过将这个定义与特征向量进行比较，我们看到，这两个概念是相关的，并且

$$\boldsymbol{\pi} = \frac{e}{\sum_i e_i}$$

是由（$\sum_i \pi_i = 1$）归一化的转移矩阵$\mathbf{P}$的左特征向量$\mathbf{e}$的倍数，其特征值为1。如果有多个单位特征向量那么相应稳态的加权和也是稳态。但对于马尔可夫链来说，我们更关心的是作为一些对初始分布的序列分布的极限的稳态。

稳定分布π_i的值与状态空间$\mathbf{P}$有关，它的特征向量保留各自的相对比例。因为$\boldsymbol{\pi}$的成分为正且满足约束条件$\sum_i 1 \cdot \pi_i = 1$，我们发现$\boldsymbol{\pi}$与一个成分全为1的向量的点积是统一的，、$\boldsymbol{\pi}$位于一个单纯形上。

有限状态空间内的时齐马尔可夫链

对于一个离散状态空间，k步转移概率的积分即为求和，可以对转移矩阵求k次幂来求得。就是说，如果$\mathbf{P}$是一步转移矩阵，$\mathbf{P}^k$就是k步转移后的转移矩阵。

如果转移矩阵$\mathbf{P}$不可约，并且是非周期的，则$\mathbf{P}^k$收敛到一个每一列都是不同的平稳分布π^*，并且，

$$\lim_{k\to\infty} \mathbf{P}^k = \pi^*,$$

独立于初始分布π。这是由Perron-Frobenius theorem所指出的。

正的转移矩阵（即矩阵的每一个元素都是正的）是不可约和非周期的。矩阵被称为是一个随机矩阵，当且仅当这是某个马尔可夫链中转移概率的矩阵。

注意：在上面的定式化中，元素(i,j)是由j转移到i的概率。有时候一个由元素(i,j)给出的等价的定式化等于由i转移到j的概率。在此情况下，转移矩阵仅是这里所给出的转移矩阵的转置。另外，一个系统的平稳分布是由该转移矩阵（每列的和为1）的右特征向量给出的，而不是左特征向量。

转移概率独立于过去的特殊况为熟知的Bernoulli scheme。仅有两个可能状态的Bernoulli scheme被熟知为伯努利过程

趋向稳定分布的收敛速度

可反转马尔可夫链

可反转马尔可夫链类似于应用贝叶斯定理来反转一个条件概率：

$$\begin{aligned}\Pr(X_n = i \mid X_{n+1} = j) &= \frac{\Pr(X_n = i, X_{n+1} = j)}{\Pr(X_{n+1} = j)} \\ &= \frac{\Pr(X_n = i)\Pr(X_{n+1} = j \mid X_n = i)}{\Pr(X_{n+1} = j)}.\end{aligned}$$

以上就是反转的马尔可夫链。因而，如果存在一个π，使得：

$$\pi_i p_{ij} = \pi_j p_{ji}.$$

那么这个马尔可夫链就是可反转的。

这个条件也被称为细致平衡（detailed balance）条件。

对于所有的i求和：

$$\sum_i \pi_i p_{ij} = \pi_j$$

所以，对于可反转马尔可夫链，π总是一个平稳分布。

伯努利方案

伯努利方案是马尔可夫链的一种特殊情形，其转移概率矩阵有相同的行，即下一状态均匀独立于当前状态（除了独立于过往状态以外）。仅有两个可能状态的伯努利方案是伯努利过程。

一般的状态空间

对于一般状态空间上的马尔可夫链的概述，详见文章状态空间可测的马尔可夫链

Harris链

局部交互的马尔可夫链

应用

物理

马尔可夫系统广泛出现在热力学和统计力学中，

化学

测试

语音识别

信息科学

排队理论

互联网应用

谷歌所使用的网页排序算法（PageRank）就是由马尔可夫链定义的。马尔可夫模型也被应用于分析用户浏览网页的行为。一阶或者二阶的马尔可夫模型可以用于对一个用户从某一网络链接转移到另一链接的行为进行建模，然后这些模型可以用于对用户之后的浏览行为进行预测。

统计

经济和金融

社会科学

生物数学

马尔可夫链也有众多的生物学应用，特别是增殖过程，可以帮助模拟生物增殖过程的建模。隐蔽马尔可夫模型还被用于生物信息学，用以编码区域或基因预测（見哈代-溫伯格定律。）

遗传学

游戏

音乐

棒球

马尔可夫文本生成器

马尔可夫过程，能为给定样品文本，生成粗略，但看似真实的文本：他们被用于众多供消遣的“模仿生成器”软件。马尔可夫链还被用于谱曲。

Fitting

历史

马尔可夫在1906年首先做出了这类过程。而将此一般化到可数无限状态空间是由俄國數學家柯尔莫果洛夫（俄语：Андре́й Никола́евич Колмого́ров）在1936年给出的。马尔可夫链与布朗运动以及遍历假说这两个二十世纪初期物理学重要课题是相联系的，但马尔可夫寻求的似乎不仅于数学动机，名义上是对于纵属事件大数法则的扩张。

俄国数学家安德雷·安德耶维齐·马尔可夫.

隱馬可夫模型

- 語音辨識

参看

- 马尔可夫性质
- 马尔可夫
- 隐马尔可夫模型
- 马尔科夫蒙特卡洛

注释

1. Norris, James R. Markov chains. Cambridge University Press. 1998.

参考文献

- A.A. Markov. "Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga". *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, 2-ya seriya, tom 15, pp 135-156, 1906.
- A.A. Markov. "Extension of the limit theorems of probability theory to a sum of variables connected in a chain". reprinted in Appendix B of: R. Howard. *Dynamic Probabilistic Systems, volume 1: Markov Chains*. John Wiley and Sons, 1971.
- Leo Breiman. *Probability*. Original edition published by Addison-Wesley, 1968; reprinted by Society for Industrial and Applied Mathematics, 1992. ISBN 978-0-89871-296-4. *(See Chapter 7.)*
- J.L. Doob. *Stochastic Processes*. New York: John Wiley and Sons, 1953. ISBN 978-0-471-52369-7.

外部連結

- 免费的《概率论入门》书中有关马尔可夫链的章节
- 世界上最大的矩阵计算 *(Google's PageRank as the stationary distribution of a random walk through the Web.)*

- Disassociated Press in Emacs approximates a Markov process
- [1] Markov chains used to produce semi-coherent English.
- 有关马尔可夫链的资源列表