فهرست مطالب

زنجیره‌های مارکوف، به پاس‌داشتِ آندری مارکوف[1] نام‌گذاری شده است. مطالعات او زمینه‌ساز تحقیقات بر فرآیندهای تصادفی[2] شد. یک فرآیند تصادفی، ابزاری ریاضی برای مدل‌سازی پدیده‌های تصادفیِ وابسته به زمان است. به عنوان مثال می‌توان در فصل اول به توضیح نسخه کلاسیک و ساده زنجیره‌های مارکوف همراه با مدل‌ها و کاربرد (های ساده) آن می‌پردازیم.

زنجیره‌های مارکوف

به طور ساده می‌توان یک زنجیره مارکوف را یک مدل‌سازی[3] بر دنباله‌ای از متغیرهای تصادفی[4] دانست به طوری که هر متغیر تصادفی متناظر با یک حالت خاص از سیستم است و در آن وضعیت یک سیستم در حالت بعدی تنها بر وضعیت آن سیستم در لحظه کنونی بستگی دارد. به طور سادهتر می‌توان گفت در زنجیره‌های مارکوف آنچه بلافاصله رخ خواهد داد تنها به حالت فعلی سیستم بستگی دارد. برای افزایش دقت این تعریف بهتر است «مدل‌سازی» و «سیستم» را تعریف کنیم:

یک سیستم از بخش‌های در حال تعامل (تاثیرگذار و تاثیرپذیر) که به کمک هم اقداماتی را برای پاسخ‌گویی به یک مسئله انجام می‌دهند گفته می‌شود. یک مدل، سیستمی را انتزاعی می‌کند بدین معنا که مهم‌ترین و اساسی‌ترین واحدهای عملیاتی و روابط بین آن‌ها را به ساخت‌های ریاضی (مثل یک ماتریس) تبدیل می‌کند. یک مدل خوب مدلیست که نه تنها اساس عملیاتی یک سیستم را در برگیرد بلکه علاوه بر رفتار آن سیستم، کاری کند که آن مدل برای تصمیم‌گیری و استدلال آوردن مناسب باشد.

طبق تعریف‌های بالا، می‌توان گفت که یک زنجیره مارکوف از حالات سیستمی، حالات گذار[5] و مقادیری که به آن حالات گذار مربوط هستند، تشکیل شده است. در ادامه همراه با مثال‌های کاربردی، تعریف بالا را گسترش می‌دهیم اما قبل از آن بهتر است به انواع زنجیره‌های مارکوف از لحاظ زمانی بپردازیم. به طور کلی ما دو نوع زنجیره مارکوف داریم: زنجیره‌های مارکوف با زمان پیوسته[6] و زنجیره‌های مارکوف با زمان گسسته[7].

در بخش پیش رو به زنجیره‌های مارکوف با زمان گسسته خواهیم پرداخت.

مثال‌هایی از زنجیره‌های مارکوف

فرآیند تصادفیِ زیر را که بر مجموعه متناهی یا شمارای M تعریف شده است را در نظر بگیرید:

$$\{X^n, n=0,1,2,\ldots\}$$

مثال 1.1 . اگر X^n آب‌وهوای روز نهم باشد، یکی از مقادیر زیر را خواهد داشت:

$$M=\{sunny, windy, rainy, cloudy\}$$

در این صورت مقادیر ممکن X^n چنیین خواهد بود:

---

[1] Andrei A. Markov
[2] Stochastic Process
[3] Modeling
[4] Random Variables
[5] Transitions
[6] Continuous Time Markov Chains (CTMC)
[7] Discrete Time Markov Chains (DTMC)

$$X^0 = sunny, X^1 = windy, X^2 = rainy, X^3 = sunny, X^4 = cloudy, \ldots$$

نکته 1.2: برای سادگی ما فرض می‌کنیم که M که همان فضای حالات است، مجموعه اعداد صحیح باشد. یک عنصر از M را یک حالت از فرآیند می‌نامیم.

تعریف 1.3: فرض کنید که متغیر تصادفی P_{ij} که معین و مستقل از زمان است داریم. در این صورت

$$P\left(X^{(n+1)} = i \middle| X^{(n)} = j, X^{(n-1)} = i_{n-1}, \ldots, X^0 = i_0\right) = P_{ij}\ n \geq 0$$

که در آن $i, j\, i_0, i_1, \ldots, i_{n-1} \in M$ می‌باشد را یک فرآیند زنجیره مارکوف می‌نامیم.

نکته 1.4. می‌توان فرمول بالا را این گونه تفسیر نمود: اگر حالات قبلی سیستم به شکل زیر باشد

$$X^{(0)}, X^{(1)}, X^{(2)}, \ldots X^{(n-1)}$$

توزیع احتمال شرطی هر حالت بعدیِ $X^{(n+1)}$ تنها به حالت کنونیِ X^n بستگی دارد و از حالات قبلی سیستم مستقل است.

نکته 1.5. احتمال P_{ij} بیان‌گر احتمالی است که فرآیند از حالت کنونیِ j به حالت i گذر پیدا می‌کند. بدیهی است که

$$P_{ij} \geq 0, \sum_{i=0}^{\infty} P_{ij} = 1, j = 0,1, \ldots$$

تعریف 1.6. به ماتریسی که P_{ij} (احتمال‌های گذار) را در بر دارد، ماتریس احتمال گذار تک‌مرحله‌ای[8] فرآیند گفته می‌شود.

$$P = \begin{bmatrix} P_{00} & P_{01} & \ldots \\ P_{10} & P_{11} & \ldots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

مثال 1.7. این مسئله بازاریابی را در نظر بگیرید: تحقیقات نشان داده است که در یک شهر تنها دو سوپرمارکت به نام‌های الف و ب وجود دارد. یک مشتری از ب ممکن است خرید بعدی خود را از الف با احتمال $\alpha\,(>0)$ انجام دهد در حالی که یک مشتری از الف ممکن است خرید بعدی خود را از ب با احتمال $\beta\,(>0)$ انجام دهد. اگر $X^{(n)}$ یک فرآیند دو مرحله‌ای باشد (به عبارتی دیگر، مقادیری از مجموعه $\{0,1\}$ اختیار کند) که رفتار مشتری را توصیف می‌کند به طوری که اگر مشتری در روز nام از ب خرید کند داریم $X^{(n)} = 0$ و اگر مشتری در روز nام از الف خرید کند داریم $X^{(n)} = 1$. از آنجایی که حالات آینده (این که مشتری از چه مغازه‌ای خرید کند) تنها به حالت کنونی بستگی دارد، فرآیند مدل شده یک فرآیند زنجیره مارکوفی می‌باشد. مشخص است که:

$$P_{00} = 1-\alpha, P_{10} = \alpha, P_{11} = 1-\beta, P_{01} = \beta$$

در این صورت ماتریس گذار تک‌مرحله‌ای برای این فرآیند به شکل زیر است:

$$P = \begin{bmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{bmatrix}$$

[8] One-step transition probability matrix

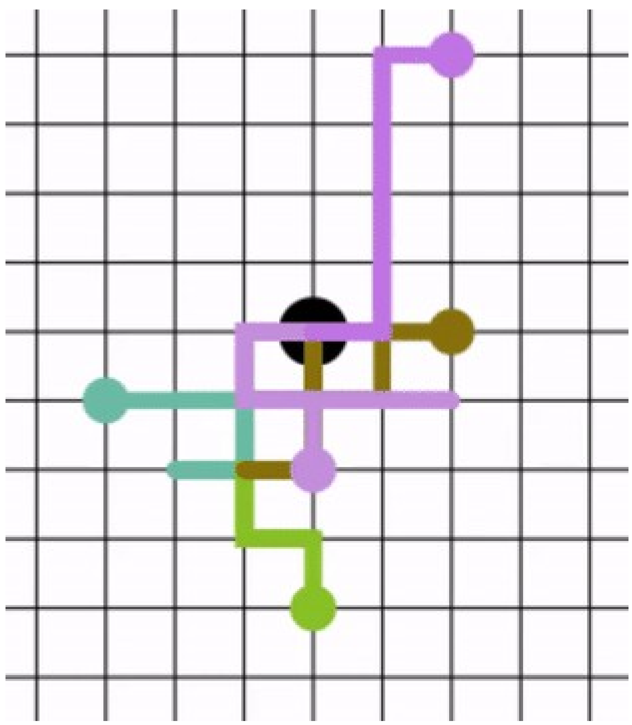
مثال 1.8. ولگشت یا گام تصادفی[9] یک مسئله معروف در ریاضیات است که در بسیاری از شاخه‌ها مانند علوم کامپیوتر و فیزیک نیز مورد مطالعه قرار می‌گیرد. ولگشت یک فرآیند تصادفی است که یک مسیر متشکل از دنباله‌ای از گام‌های تصادفی را بر یک فضای ریاضی را توصیف می‌کند. اگر فردی را که بر یک محور اعداد حسابی گام برمی‌دارد را در نظر بگیریم یعنی

$$\text{فضای حالت} = \{\ldots,-2,-1,0,1,2,\ldots\}$$

هر مرحله آن فرد در حالت i می‌تواند یک گام به جلو $(+1)$ یا یک گام به عقب (-1) بردارد که به ترتیب با احتمال‌های

$p\,(0<p<1)$ و $(1-p)$ همراه است. بدین ترتیب احتمال‌های گذار به شکل زیر است:

$$P_{ij}=\begin{cases} p, \wedge if\ i=j+1 \\ 1-p, \wedge if\ i=j-1\ for\ j=0,\pm 1,\pm 2,\ldots \\ 0\ otherwise \end{cases}$$



1- تصویری از یک مدل مسئله ولگشت

مثال 1.9. مسئله پاکباختگی قمارباز[10]. فرض کنید که قمار بازی در هر بازی با احتمال $p\,(0<p<1)$ یک دلار می‌برد و با احتمال $(1-p)$ یک دلار می‌بازد. بازی زمانی تمام می‌شود که یا او تمام پول خود را از دست دهد یا مجموعه N دلار به دست آورد. مشخص است که ثروت این قمار باز یک فرآیند زنجیره مارکوفی است. بدین ترتیب ما احتمال‌های گذار زیر را خواهیم داشت:

---
[9] Random Walk
[10] Gambler’s ruin

$$P_{ij}=\begin{cases} p, & if\ i=j+1 \\ 1-p, & if\ i=j-1 \\ 0\ otherwise \end{cases}$$

در اینجا حالت 0 و حالت N حالات جذاب[11] نامیده می‌شوند. اگر فرآیند به یکی از این حالات جذاب برسد، فرآیند تا ابد در حالت0 یا N باقی خواهند ماند.

ماتریس گذار n مرحله‌ای

در مثال‌های قبل از ماتریکس‌های گذار تک‌مرحله‌ای استفاده شده بود. اکنون ماتریکس‌های گذار را به n مرحله تعمیم می‌دهیم.

تعریف 1.9. احتمال این که یک فرآیند از حالت j طی چندین گذار به حالت i برسد را با $P_{ij}^{(n)}$ نشان می‌دهیم. بدیهی است که $P_{ij}^{(1)}=P_{ij}$.

قضیه 1.10. اگر $P^{(n)}$ ماتریس گذار n مرحله‌ای و P ماتریس گذار تک‌مرحله‌ای باشد، داریم $P^{(n)}=p^n$.

اثبات:

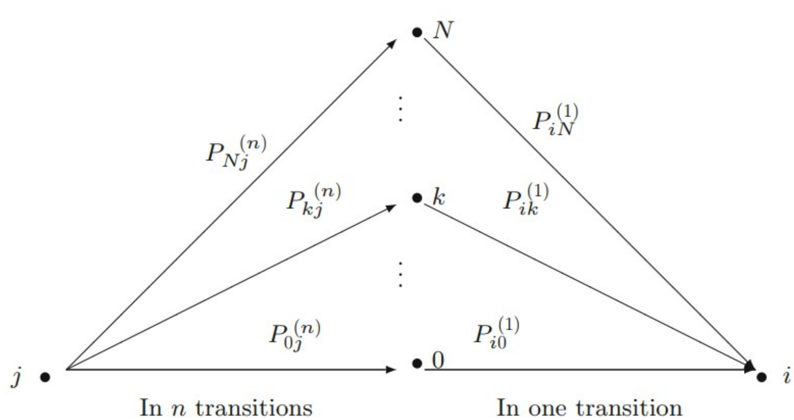
با استقرا اثبات می‌کنیم. مشخص است که حکم در حالتی که $n=1$ می‌باشد برقرار است. سپس فرض می‌کنیم که حکم برای n نیز برقرار است. با توجه به این که:

$$P^n=\underbrace{P\cdot P\cdot\ldots\cdot P}_{n\times ¿¿}$$

پس داریم:

$$P_{ij}^{(n+1)}=\sum_{k\in M}P_{ki}^{(n)}P_{jk}^{(1)}=\sum_{k\in M}P_{ki}^{n}P_{jk}=\left[P^{n+1}\right]_{ji}$$

که در تصویر ب این مورد مشخص است. طبق اصل استقرا چنین خاصیتی برای تمام مقادیر نامنفی برقرار است.



2- احتمال گذار مرحله $n+1$

[11] Absorbing states

نکته 1.11. می‌توان نشان داد که

$$P^{(m)} \cdot P^{(n)} = P^m \cdot P^n = P^{m+n} = P^{(m+n)}$$

مثال 1.12. مسئله بازاریابی را بار دیگر در نظر بگیرید. در مدلی که مطرح شد داشتیم که:

$$P = \begin{bmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{bmatrix}$$

اگر $\beta = 0.4$ و $\alpha = 0.3$ باشد، سپس داریم:

$$P^{(4)} = P^4 = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}^4 = \begin{bmatrix} 0.5749 & 0.5668 \\ 0.4351 & 0.4332 \end{bmatrix}$$

پس $P_{00}^{(4)} = 0.5749$ احتمال این است یک مشتریِ ب در چهارمین خرید خود باز از ب خرید کند و $P_{01}^{(4)} = 0.4332$ احتمال این که یک مشتری ب در چهارمین خرید خود از الف خرید کند را مشخص می‌کند.

نکته 1.13. یک فرآیند زنجیره مارکوف را با حالات $\{0,1,2,\ldots\}$ در نظر بگیرید. فرض کنید که در مرحله $n=0$ احتمال این که فرآیند در مرحله i باشد برابر با α_i باشد $(i=0,1,2,\ldots)$. یک پرسش مهم این است: احتمال این که فرآیند پس از n مرحله وارد حالت j شود چه می‌باشد؟ چنین احتمالی برابر است با $P_{ji}^{(n)} = [P^n]_{ji}$ که در آن P_{ji} احتمال گذار تک‌مرحله‌ای فرآیند از حالت i به حالت j است. در نتیجه احتمال مطلوب چنین محاسبه می‌شود:

$$\sum_{i=0}^{\infty} P(X^{(0)}=i) \cdot P_{ji}^{(n)} = \sum_{i=0}^{\infty} \alpha_i \cdot [P^n]_{ji}$$

اگر توزیع احتمال حالات در فرآیند زنجیره مارکوفی را پس از n مرحله گذار را به شکل زیر تعریف کنیم

$$X^{(n)} = (\widetilde{X}_0^{(\mathrm{n})}, \widetilde{X}_1^{(n)}, \ldots,)$$

سپس $\sum_{i=0}^{\infty} \widetilde{X}_i^{(n)} = 1$. به راحتی می‌توان نشان داد که $X^{(n+1)} = P\,X^{(n)}$ و

$$X^{(n+1)} = P^{(n+1)}\,X^{(0)} = P^n\,X^{(0)}$$

مثال 1.14. مثال قبل را در نظر بگیرید. این که در مرحله $n=0$ یک مشتری به الف تعلق داشته باشد را می‌توان به شکل زیر نشان داد:

$$X^{(0)} = X^{(n)} = (\widetilde{X}_0^{(\mathrm{n})}, \widetilde{X}_1^{(n)})^T = (0,1)^T$$

در پاسخ به این پرسش که مشتری مورد نظر با چه احتمالی هنگام خرید چهارم در مغازه الف یا ب است بدین صورت مشخص می‌گردد

$$X^{(4)} = P^{(4)}\,X^{(0)} = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}^4 (0,1)^T = (0.5668\,, 0.4332)^T$$

پس مشتری مورد نظر در چهارمین خرید خود با احتمال 0.4332 در مغازه الف و با احتمال 0.5668 در مغازه ب حضور خواهد داشت.

زنجیره مارکوفی تقلیل ناپذیر و کلاس حالات

در این بخش ما دو تعریف از حالات زنجیره مارکوف مطرح می‌کنیم.

تعریف 1.15. در یک زنجیره مارکوف، حالت i را قابل دسترسی از حالت j می‌نامیم اگر به ازای مقادیر $n>0$ داشته باشیم $P_{ij}^{(n)}>0$. این بدین معناست که اگر از حالت j فرآیند به حرکت بیافتد، محتمل است (مقدار احتمال مثبت) که پس از تعداد متناهی گذار به حالت i وارد شویم.

تعریف 1.16. حالت i و حالت j را مرتبط[12] با هم می‌نامیم اگر حالت i و حالت j به همدیگر قابل دسترسی باشند.

نکته 1.17. تعریف ارتباط یک رابطه هم‌ارزی را مشخص می‌کند.

i. حالت i با حالت i در هیچ مرحله گذار ارتباط دارد زیرا که

$$P_{ji}^{(0)}=P\left(X^0=i\middle| X^{(0)}=i\right)=1>0$$

ii. اگر حالت i با حالت j در ارتباط باشد پس حالت j نیز با حالت i در ارتباط است

iii. اگر حالت i با حالت j در ارتباط باشد و حالت j نیز با حالت k در ارتباط باشد پس حالت i با حالت k در ارتباط است. از آن‌جایی که به ازای مقادیر مختلف n و m داریم $P_{ji}^{(m)},P_{kj}^{(n)}>0$، به معادله زیر خواهیم رسید

$$P_{ki}^{(m+n)}=\sum_{h\in M} P_{hi}^{(m)}P_{kh}^{(n)}\geq P_{ji}^{(m)}P_{kj}^{(n)}>0$$

در نتیجه حالت k از طرف حالت i قابل دسترسی است. با تعویض i و k به نتیجه مشابه‌ای می‌رسیم. بنابراین i با k در ارتباط است و اثبات کامل می‌شود.

تعریف 1.17. دو حالت که با هم ارتباط دارند در یک کلاس[13] هستند.اگر تمام حالات یک فرآیند به یک کلاس تعلق داشته باشند (با هم دیگر ارتباط برقرار کنند)، آن زنجیره مارکوف را تقلیل‌ناپذیر[14] می‌نامیم.

مثال 1.18. ماتریس احتمال گذار زیر متعلق به یک زنجیره مارکوف تقلیل‌ناپذیر است.

$$\begin{bmatrix} 0.0 & 0.5 & 0.5 \\ 0.5 & 0.0 & 0.5 \\ 0.5 & 0.5 & 0.0 \end{bmatrix}$$

در حالی که ماتریس احتمال گذار زیر نمی‌تواند در خصوص یک زنجیره مارکوف تقلیل‌ناپذیر باشد.

$$\begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 1 & 0.0 & 0.5 \\ 0.5 & 0.5 & 0.0 \end{bmatrix}$$

---

[12] Communicate
[13] Class
[14] Irreducible

چرا که در این ماتریس نمی‌توان از حالت سطر 0 به حالت های 1 و 2 رسید. به عبارت دیگر

$$P_{01}^{(n)} = P_{02}^{(n)} = 0$$

متقابلا می‌توان این ماتریس را تقلیل‌پذیر [15] نامید.

تعریف 1.19. به ازای هر حالت i در زنجیره مارکوف، احتمال این که اگر از حالت i شروع به حرکت کنیم، فرآیند در نهایت پس از تعداد متناهی تغییر باز به حالت i برسد را با f_i نشان می‌دهیم. حالت i را بازگشت‌پذیر[16] می‌نامیم اگر $f_i = 1$ و گذرا[17] می‌نامیم اگر $f_i = 0$

قضیه پیش‌رو برای یک حالت بازگشت‌پذیر مطرح می‌شود.

قضیه 1.20. در یک زنجیره مارکوف متناهی، حالت i را بازگشت‌پذیر نامیم اگر و تنها اگر معادله زیر برقرار باشد.

$$\sum_{n=1}^{\infty} P_{ij}^{(n)} = \infty$$

این قضیه مطرح می‌کند که یک حالت گذرا به تعداد متناهی بازدید می‌شود. بنابراین می‌توان نشان در یک زنجیره مارکوف با تعداد متناهی حالات، همه حالات نمی‌توانند گذرا باشند. به کمک این قضیه، می‌توان قضیه زیر را اثبات کرد.

قضیه 1.21. در یک زنجیره مارکوف متناهی، اگر حالت i بازگشت‌پذیر (یا گذرا) باشد و حالت i با حالت j در ارتباط باشد پس حالت j نیز بازگشت‌پذیر (یا گذرا) است.

توزیع یکسان[18] یک زنجیره مارکوف متناهی

تعریف 1.22. می‌گوییم حالت i تناوب d را دارد اگر $P_{ii}^{(n)} = 0$ و n بر d بخش‌پذیر نباشد و d بزرگترین عدد حسابی با این خاصیت باشد. یک حالت با تناوب ۱ را بی‌تناوب[19] می‌نامیم.

مثال 1.23. ماتریس گذار حالات زیر را در نظر بگیرید

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

متوجه می‌شویم که

$$P^n = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^n = \frac{1}{2} \begin{bmatrix} 1+(-1)^n & 1+(-1)^{(n+1)} \\ 1+(-1)^{(n+1)} & 1+(-1)^n \end{bmatrix}$$

در نتیجه

[15] Reducible
[16] Recurrent
[17] Transient
[18] Stationary distribution
[19] Aperiodic

$$P_{00}^{(2n+1)} = P_{11}^{(2n+1)} = 0$$

پس هر دو حالت 0 و 1 تناوب 2 را دارند.

تعریف 1.24. حالت i را یک حالت بازگشت‌پذیر مثبت[20] می‌نامیم اگر بازگشت‌پذیر باشد و با شروع از حالت i، زمان مورد انتظار برای بازگشت به حالت i متناهی باشد.

تعریف 1.25. یک حالت را ارگادیک[21] می‌گوییم اگر بازگشت‌پذیر مثبت و بی‌تناوب باشد. به عنوان مثال در مسئله بازاریابی داشتیم:

$$X^{(0)} = (1,0)^T$$

مشاهده می‌شود که:

$$\mathbf{X}^{(1)} = P\mathbf{X}^{(0)} = \begin{pmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{pmatrix} (1,0)^T = (0.7, 0.3)^T,$$

$$\mathbf{X}^{(2)} = P^2\mathbf{X}^{(0)} = \begin{pmatrix} 0.61 & 0.52 \\ 0.39 & 0.48 \end{pmatrix} (1,0)^T = (0.61, 0.39)^T,$$

$$\mathbf{X}^{(4)} = P^4\mathbf{X}^{(0)} = \begin{pmatrix} 0.5749 & 0.5668 \\ 0.4251 & 0.4332 \end{pmatrix} (1,0)^T = (0.5749, 0.4251)^T,$$

$$\mathbf{X}^{(8)} = P^8\mathbf{X}^{(0)} = \begin{pmatrix} 0.5715 & 0.5714 \\ 0.4285 & 0.4286 \end{pmatrix} (1,0)^T = (0.5715, 0.4285)^T,$$

$$\mathbf{X}^{(16)} = P^{16}\mathbf{X}^{(0)} = \begin{pmatrix} 0.5714 & 0.5174 \\ 0.4286 & 0.4286 \end{pmatrix} (1,0)^T = (0.5714, 0.4286)^T.$$

[20] Positive recurrent
[21] Erogodic

در واقع می‌توان نشان داد که:

$$\lim_{n \to \infty} X^{(n)} = (0.5714, 0.4286)^T$$

این بدین معناست که در نهایت این که یک شخص مشتری مغازه الف باشد برابر با 0.5714 خواهد بود.

از آن جا که $X^{(n)} = PX^{(n-1)}$ پس اگر قرار دهیم

$$\lim_{n \to \infty} X^{(n)} = \pi$$

پس می‌توان نوشت

$$\pi = \lim_{n \to \infty} X^{(n)} = \lim_{n \to \infty} PX^{(n-1)} = P\pi$$

که این ما را به تعریف زیر می‌رساند.

تعریف 1.26. بردار $\pi = (\pi_0, \pi_1, \ldots, \pi_{(k-1)})^T$ را توزیع یکسان یک زنجیره مارکوف می‌نامیم اگر دو شرط زیر را داشته باشد:

- $\sum_{i=0}^{k-1} \pi_i = 1$ و $\pi_i \geq 0$ $\quad P(X_{n+1} = x)$
- $P\pi = \pi$ یا به عبارتی دیگر $\sum_{j=0}^{k-1} P_{ij} \pi_j = \pi_i$

زنجیره مارکوف مرتبه بالاتر

طبق ویژگی مارکوفی، در یک زنجیره مارکوف ما مدل‌مان را در هر حالت تنها وابسته به حالت کنونی می‌دانیم و از وضعیت حالات قبل‌تر صرف‌نظر می‌کنیم. ولیکن می‌توان مدل‌های مارکوفی را تنظیم کرد که علاوه بر حالت کنونی، به چند حالت قبل نیز وابسته باشد. به چنین زنجیره‌های مارکوفی، زنجیره‌های مارکوف با مرتبه بالاتر[22] گفته می‌شود. برای مثال برای یک زنجیره مارکوف با مرتبه ۲[23] داریم:

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

معمولا از زنجیره‌های مارکوف با مرتبه بالاتر از ۲ به ندرت استفاده می‌شود زیرا که در چنین زنجیره‌هایی، ویژگی مارکوفی کم‌رنگ‌تر می‌گردد. با این وجود استفاده از چنین زنجیره‌های مارکوف با مرتبه بالاتر، در موقعیت‌هایی که می‌خواهیم سیستم‌های پیچیده‌ای را که تنها به حالت کنونی وابسته نباشند خلق کنیم، می‌توانند مفید باشند.

---

[22]Higher order Markov chain
[23]Second order Markov chain

مثال برای زنجیره مارکوف ساده: تولید متن

یکی از ساده‌ترین کاربردهای زنجیره مارکوف در تولید متن است.این نوع تولید متن از ساده‌ترین الگوریتم‌های تولید متن است که بر اساس احتمال پیش‌آمد یک حالت به ازای حالات قبلی (ویژگی مارکوف) واژگان را پشت سر هم تولید می‌کند. برای این کار ما به کمک یک ساختمان داده (معمولا یک دیکشنری) به ازای هر حالت i تعداد تکرار حالت‌های j ای را که بعد از i آمده است را می‌شماریم. در یک متن ما با واژگان زیادی مواجه هستیم که هر واژه به مثابه یک حالت است. کافیست که به ازای هر بار تکرار آن واژه در جای جای متن، واژه بعدی را در نظر بگیریم، آن را بشماریم و به کمک ساختمان داده خود یک وزن به هر واژه ممکن بدهیم. به عنوان مثال واژه «تو» در سه جمله زیر در سه حالت تکرار شده است:

«تو باید قبل از شب خانه باشی.»

«تو بیدار هستی.»

«تو باید مرا بیدار کنی.»

در این صورت در ساختمان داده خود داریم:

$dict = \{[\text{تو}, \text{باید} = 2], [\text{تو}, \text{بیدار} = 1], ...\}$

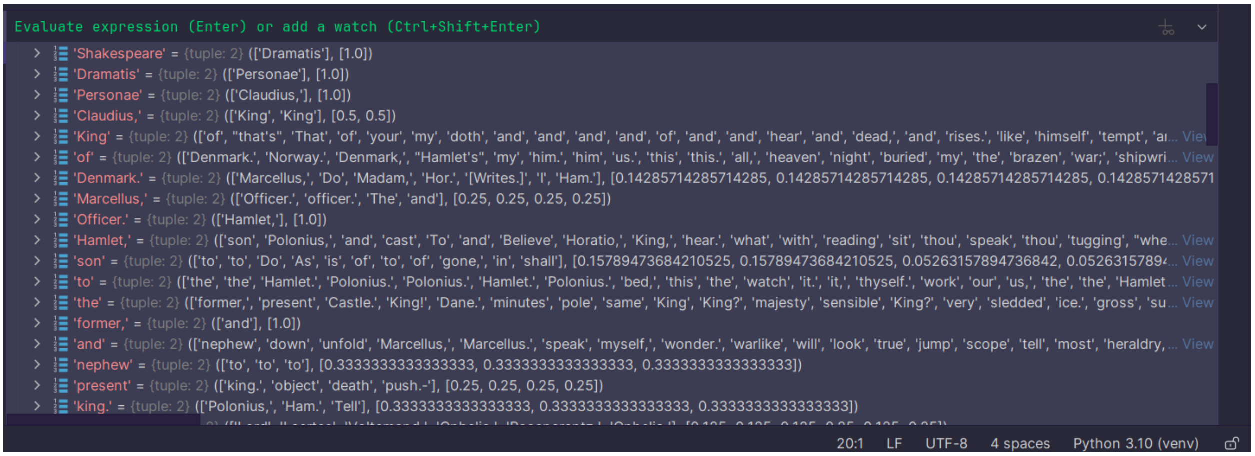
دیکشنری ما شامل تمام زوج‌واژه‌ها و تعداد تکرارهایشان در دیتاست ما خواهد بود. پس از شمردن تعداد تکرار به ازای هر زوج‌واژه یا به ازای هر زوج حالت، می‌بایست آن‌ها را با توجه به دفعات تکرارشان با مقادیر احتمالی به روز رسانی کرد. در دیتاست کوچک ما در دو حالت بعد از واژه «تو» واژه «باید» آمده است و در یک حالت بعد از واژه «تو» واژه «بیدار». در نتیجه طبق فرمول پیش رو: $P(j \mid i) = \frac{Number\ of\ ocurrences\ of\ j\ after\ i}{Number\ of\ ocurrences\ of\ j}$ دیکشنری را به شکل زیر به روز رسانی می‌کنیم:

$dict = \{[\text{تو}, \text{باید} = 0.66666], [\text{تو}, \text{بیدار} = 0.33333], ...\}$

ما همچنین می‌توانیم از یک زنجیره مارکوف مرتبه دوم نیز برای سه‌تایی $(i, j, k), i<j<k$ استفاده کنیم. در این صورت الگوریتم ما به جای بررسی کردن حالت بعدی یک واژه، دو حالت بعدی واژه را در نظر می‌گیرد. عوامل متعددی از جمله اندازه دیتاست ما در انتخاب مرتبه زنجیره مارکوف ما دخیل است؛ برای متون کوچک مرتبه‌های کوچک‌تر مناسب‌تر است چرا که تنوع جملات تولید شده را افزایش می‌دهد و از جملات تکراری جلوگیری می‌کند ولیکن برای متون بزرگ‌تر زنجیره‌های مارکوف با مرتبه‌های بالاتر پسندیده‌تر است چرا که احتمال تولید جملات قابل‌فهم‌تر را بیش‌تر می‌کند.

فایل first_order.py در پوشه text برنامه‌ای برای تولید ۲۰ جمله تصادفی بر اساس فایل تکست hamlet.txt یا saadi_norm.txt می‌باشد. توضیح کد بسیار ساده است: تابع build_markov_chain ابتدا متن را به آرایه‌ای از واژگان تبدیل می‌کند (به این قسمت Tokenize کردن متن می‌گویند) و سپس به

کمک یک دیشکنری حالات j که بلافاصله پس از حالت i می‌آید را ذخیره می‌کنیم و در نهایت طبق فرمولی که ارائه کرده‌ایم احتمال حالات را بدست می‌آوریم. خروجی این تابع دیکشنری زیر است:



در تصویرحالات j که پس از i می‌آیند به همراه احتمال هر حالت (که تابع احتمال توزیع یکنواخت است) مشخص است. دقت کنید که برخی حالات خاص هستند و تنها یک حالت پس از آن می‌آید و از همین رو احتمال ۱ به آن‌ها تخصیص داده می‌شود. این از نقص‌های تولید متن به کمک زنجیره مارکوف است که برای حل کردن آن می‌توان اندازه دیتاست را بزرگ کرد یا حالات خاص را نادیده گرفت.

تابع generate_sentence به کمک دیکشنری تولید شده و احتمالات محاسبه شده جملات را تولید می‌کند.

خروجی برنامه جملاتی را تولید می‌کنند که به ندرت قابل فهم و منطقی هستند ولی در اکثر مواقع می‌توان عبارت‌های معمول را در آن یافت. در تصویر زیر برخی از جملات تولیدی بر اساس اشعار سعدی را مشاهده می‌کنید:

```
Enter 1 for hamlet Enter 2 for sa'adi
2
نفرت کند ایام دشمن دلیران فرست کس آتش سوزان نشاندی
فاوثقنی المودت کاتبی قوما اسقیانی و لاله و گوی دولت
پیمبر زادگی قدرش چه غم دل خلقی به سر دل
جمالت کیوان برو که در آن که در آن که
پند گیرد بزن نوایی را گو مباش غره به دست
رویت ببینم که می بخشد با تو از سرش خواست
نخیزی به چوگان چنانکه شرط است و گفت و دین
دردناک باشد که در روی تو در سر زلف تو
باهم سگالند راز دل به هر که در این بنیان
العدی اتشمت اعدایی و گر تو را که در این
عراقم به میخانه در سر که در این سخن نمی
مردان نداری که ز هر که در بند پیش از
```
> seo > text > first_order.py

بدیهی است که پس از ابداع روش‌های پیچیده و منطقی‌تر برای تولید متون، دیگر کم‌تر از این روش متکی بر زنجیره مارکوف استفاده می‌شود. منتها این برنامه ساده همچنان در برخی از ویرایشگرها خصوصا برای پیش‌نهاد دادن کلمات پرکاربرد به کاربر هنگام نوشتن به کار می‌رود.

مثالی دیگر برای زنجیره مارکوف ساده: ساخت ملودی موسیقی

این جا نقص اساسی زنجیره‌های مارکوف ساده بهتر مشخص می‌شود. با آن که در تولید متن در تولید عبارات کوتاه با معنی موفق بوده‌ایم (و از همین رو برای یک سیستم autocorrect ساز و کاری ترتیب داده‌ایم) برای ساخت ملودی موسیقی کم‌تر با این موفقیت رو به رو خواهیم شد. ساخت ملودی از ساخت متن به مراتب پیچیده‌تر است و الگوریتم ما منطق درونی و مکانیسم‌های پنهان آهنگ‌سازی را در نظر نمی‌گیرد؛ تنها در تلاش است تا حالت‌های محتمل را پشت سر هم سوار کند.

محدودیت‌های زیادی پیش روست؛ برخلاف متن ما ممکن است در آن واحد با چندین جریان داده (چندین ملودی) سروکار داشته باشیم (مانند آکوردها). همچنین سیستم ما در مقابل سکوت اینرسی دارد و ممکن است هنگامی که با سکوت روبه‌رو شود مدام بین نت و سکوت در نوسان باشد. از همین جهت هم در برنامه mid_first.py از آکورد و سکوت صرف نظر شده است و تنها به یک خط ملودی توجه شده است. برنامه فایل‌ها با پسوند mid که درون پوشه midi می‌باشد را پردازش می‌کند، نت‌های خط اول (اصطلاحا در ترمینولوژی midi به آن ترک می‌گویند) را استخراج می‌کند و همه نت‌ها را در یک لیست به تابع generate_transition_matrix ارسال می‌کند تا همانند مثال تولید متن، شروع به تولید ماتریس گذار حالت شود. درون فولدر mid آثار آهنگ‌ساز اتریشی فرانز شوبرت قرار دارد که به نظر من به خاطر ذات رمانتیسیستی ملودی‌های آثار پیانواش، برای ساخت ملودی‌هایی شبیه به سبک و سیاق وی مناسب است.

زنجیره‌های مارکوف بالاتر برای ملودی مناسب‌تر هستند اما از آنجا که دیتاست یک آهنگ‌ساز نسبتا کوچک است به زنجیره مارکوف مرتبه اول بسنده شده است.

الگوریتم PageRank برای رتبه‌بندی صفحات وب

الگوریتمPageRank یک الگوریتم بر پایه زنجیره مارکوف است که توسط گوگل برای رتبه‌بندی صفحات وب استفاده می‌شود. چنین رتبه‌بندی‌ای برای مرتب‌سازی نتایج حاصل از یک جستجو مرحله‌ای کلیدی حساب می‌شود.

معمولا این الگوریتم به کمک مدل‌سازی وب به وسیله یک گراف جهت‌داربه صورتی که هر گره نمایانگر یک صفحه از وب و هر یال نمایانگر یک لینک از آن صفحه به صفحه مقصد است، صورت می‌گیرد. بدیهی است که وضعیت هر حالت (مشاهده نمودن یک صفحه وب از طریق یک لینک) تنها به حالت قبلی وابسته است، از این جهت چنین الگوریتمی خاصیت مارکوفی دارد و معمولا آن را به وسیله زنجیره مارکوف مرتبه اول مدل‌سازی می‌کنند. همچنین از آنجا که در ساختمان داده می‌توان یک گراف را به کمک یک ماتریس مجاورت ذخیره کرد، می‌توان گراف موردنظر را هم‌ارز با ماتریس گذار حالات در زنجیره مارکوف دانست. در ماتریس گذار متناظر با PageRank، هر چقدر صفحات بیش‌تری به یک صفحه مورد نظر لینک داده باشند، احتمال آن صفحه موردنظر بیش‌تر خواهد بود. برای روشن‌سازی بیش‌تر به مثال زیر دقت کنید:

فرض کنید A و B و C و D چهار صفحه در یک سایت باشند و چنین رابطه‌ای بین آن‌ها برقرار باشد:

$$\begin{gathered} A \rightarrow B \\ B \rightarrow C \\ C \rightarrow A, B, D \\ D \rightarrow A \end{gathered}$$

این بدین معناست که از صفحه A به B یک لینک وجود دارد، از صفحه B به C یک لینک، از صفحه C به سه صفحه دیگر یک لینک وجود دارد و از صفحه D به صفحه A یک لینک. می‌توان ماتریس مجاورت مرتبط به این گراف جهات دار را این‌گونه نمایش داد:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

سپس به نرمال‌سازی داده‌ها (تقسیم مقدار هر ستون بر جمع مقادیر آن ستون) می‌پردازیم. با این کار مقادیر را به مقادیر معتبر برای یک مدل‌سازی احتمالی تبدیل می‌کنیم:

$$\begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/2 & 1/2 & 0 & 1 \\ 1/2 & 0 & 0 & 0 \end{bmatrix}$$

معمولا برای مدل‌سازی بهتر چنین ماتریس گذاری عاملی را با نام ضریب استهلاک[24] (که معمولا آن را با d نشان می‌دهند) معرفی می‌کنیم. این ضریب حالتی را که ممکن است کاربر بدون پیروی از زنجیره لینک‌ها، به صورت تصادفی وارد یک صفحه دیگر بشود مدل‌سازی می‌کند. در الگوریتم PageRank ماتریس گذار در فرمول زیر ضرب می‌شود:

$$v' = d . P_v + \frac{1-d}{N}$$

که در آن v' مقدار به روزرسانی شده در ماتریس گذار، d ضریب استهلاک، P_V احتمال متناظر با ماتریس گذار آن مقدار، و N تعداد کل صفحات وب است.

اگر $d=0.7$، به کمک فرمول بالا $\frac{1-d}{N}=\frac{0.3}{4}$ و به مقدار 0.075 می‌رسیم. در نتیجه ماتریس موردنظر برابر خواهد بود با:

$$\begin{bmatrix} 0.075 & 0.35 & 0.075 & 0.075 \\ 0.075 & 0.075 & 0.775 & 0.075 \\ 0.425 & 0.425 & 0.075 & 0.775 \\ 0.425 & 0.075 & 0.075 & 0.075 \end{bmatrix}$$

به چنین ماتریسی، ماتریس دورنوردی[25] گفته می‌شود. هدف از ماتریس دورنوردی، اجتناب از حالت‌هاییست که ممکن است منجر به دور در وب‌گردی و یا صفحاتی که به دیگر صفحات لینک ندارد می‌باشد. همان‌طور که مشاهده می‌کنید به کمک چنین عملیاتی، مقادیر احتمالی ۰ (که می‌تواند به بن‌بست منجر شود) و ۱ (که می‌تواند به دور منجر شود) حذف می‌گردد.

برای رسیدن به این ماتریس گذا ر می‌توان از طریق‌های مختلفی (مانند web scrawling یا web scraping) این وابستگی‌ها را از یک سایت استخراج کرد. بدیهی است که رسیدن به یک مدل بزرگ داده نیازمند منابعی است که تنها در دسترس موتورهای جستجوی بزرگ است. در کد مود نظر ما، به کمک یک نمونه ساده ماتریس گذار، در تلاشیم که صفحات گوناگون را ردمبندی کنیم.

فایل pagerank.py را در نظر بگیرید. به علت عملیات‌های مربوط به جبر خطی، از کتابخانه scipy بهره جسته‌ایم.

متد load_chain در این برنامه، یک فایل را که در آن ماتریس گذار متناظر با گراف جهت‌داری که ارتباط صفحات را توصیف می‌کند ذخیره شده است، فراخوانی می‌کند. این متد سپس ماتریس دورنوری را به وسیله فرایندی که شرح داده شد، می‌سازد و برمی‌گرداند.

متد prob_trajectory، احتمال یک سری از لینک‌گردی‌ها را (مثلا ازصفحه ۱ به صفحه ۲ به صفحه ۶ به صفحه ۴) را محاسبه می‌کند.

متد stationary_dist به محاسبه توزیع یکسان زنجیره مارکوف می‌پردازد. در واقع از آن جا که ما به توزیع احتمالی مشاهده شدن صفحات مختلف وب در دراز مدت نیاز داریم، محاسبه توزیع یکسان قدمی کلیدی است. با توجه به تعریف 1.26 ما باید توزیع یکسان π را به گونه‌ای بیابیم که $P\pi=\pi$. از جبر خطی عددی

---

[24]Damping factor
[25]Teleportation matrix