Applications of the Markov Chains

Coursera. Stochastic Processes

May 26, 2021

Markov chains are known to have a variety of applications in different fields, ranging from physics to linguistics and art. In what follows, we would like to consider a few examples of the use of Markov chains in more detail.

We would like to begin with the application discovered by A. A. Markov himself. In his 1913 paper [3], Markov analysed the sequence of the first 20000 letters of the A. S. Pushkin poem «Eugeny Onegin». As we would say now, he viewed this sequence as a Markov chain with the state space $\mathcal{S} = \{1, 2\}$, where 1 stands for the letter being a vowel and 2 for a consonant. Dividing the whole sample into 200 subsamples of size 100, Markov performed some statistical analysis and estimated, in particular, the probabilities $p = \mathbb{P}\{X_i = 1\}$, $p_1 = \mathbb{P}\{X_i = 1|X_{i-1} = 1\}$ and $p_0 = \mathbb{P}\{X_i = 1|X_{i-1} = 2\}$, $i = \{\overline{1, 100}\}$. The main outcome of these calculations was that the probability of a letter being a vowel clearly depends on the preceding letter, since it turned out that $p_0 - p_1 = 0.535$. What is probably more important, Markov has also demonstrated that the obtained empirical results are consistent with the theoretical developments from his other paper [2], leading to the conclusion that a Markov chain can successfully describe the considered experiment.

Seeming overly simple nowadays, Markov's study was a breakthrough a century ago, since before Markov time dependence has never been taken into account [6]. Also, despite being very specific at the first glance, the application discovered by A. A. Markov turned out to be extremely valuable for practice almost 30 years later, when the American mathematician C. E. Shannon has written his prominent work [5]. Proceeding in the similar manner, he proposed to use this model not only to describe the events, but also to generate them, and has shown that a Markov chain can create sequences of symbols which resemble an ordinary English text. This allowed Shannon to describe an information source as a Markov process and introduce the notion of entropy which measures the amount of produced information and has the form

$$H = -\sum_{i=1}^{n} p_i \log p_i,$$

with $p_i, i = \{\overline{1, n}\}$ being the probability of occurrence of some information unit. Beside laying the foundation of the whole information theory, Shannon's discovery opened up the opportunity of using Markov chains to generate objects, which found applications even in such areas as art. For instance, the musician Iannis Xenakis suggested using Markov chains to create music [7], and this idea is nowadays widely used in the so-called algorithmic music composition.

Another application of the Markov chain is much more recent and is given by the PageRank algorithm presented in the paper by Brin and Page [4] and used by Google to rank websites. The algorithm measures importance of a webpage by its link structure and describes the behaviour of a «random surfer» visiting websites by randomly clicking the links on a current website. The $M \in \mathbb{N}$ existing webpages constitute the state space $\mathcal{S} = \{1, \ldots, M\}$, and the transition matrix P consists of elements $p_{ij}, i, j = \{\overline{1, M}\}$ representing the probability of moving from page i to page j in one click. Particularly, the «Google matrix» P is defined as

$$P = \alpha P_1 + (1 - \alpha) P_2,$$

where P_1 and P_2 are $M \times M$ matrices and $\alpha \in (0, 1)$. The parameter α is usually taken equal to 0.85 and stands for the proportion of time a surfer uses hyperlinks, while $(1 - \alpha)$ depicts the fraction of time he manually enters a URL instead. The elements $p_{ij}^{(1)}, i, j = \{\overline{1, M}\}$ of the matrix P_1 have the form

$$p_{ij}^{(1)} = \begin{cases} \frac{1}{\sum\limits_{m=1}^{M} \mathbb{1}\{i \to m\}}, & \textit{if } i \to j, \\ 0, & \textit{otherwise} \end{cases}$$

and represent the link structure. The matrix P_2 is defined as

$$P_2 = ev^T,$$

where e is a $M \times 1$ unit vector and $v^T > 0$ is a $1 \times M$ «personalisation» vector with i-th element being the probability of typing i-th URL; for more details on the construction of P, see [1]. Now, since $p_{ij} > 0$ for all $i, j = \{\overline{1, M}\}$, the Markov chain is ergodic, and there exists a unique limit $\lim\limits_{k \to \infty} \vec{\pi}^{(k)} = \vec{\pi}^*$, where $\vec{\pi}^*$ is a stationary distribution, i.e., $\vec{\pi}^* P = \vec{\pi}^*$. This is exactly the PageRank vector, with i-th element being the proportion of time a user spends on the i-th webpage in the long run.

As can be seen from the above examples, the applications of Markov chains are very diverse and have a long history. Appeared in the early XXth century as a novel model with quite specific application, they developed to a powerful tool which can be used in many practical instances. It is worth mentioning, however, that the applications of Markov chains go further beyond the direct results presented here. More importantly, they give rise to a variety of other models and methods, such as Markov decision processes, Hidden Markov Models or Markov Chain Monte Carlo, which can be successfully employed to even more complicated problems.

References

[1] Langville, A.N. and Meyer, C.D. (2005). A survey of eigenvector methods for web information retrieval. *SIAM review*, 47(1):135–161.

[2] Markov, A.A. (1907). Issledovanie zamechatel'nogo sluchaya zavisimyh ispytanij. *Izvestiya Akademii Nauk, SPb, VI seriya*, 1(93):61–80.

[3] Markov, A.A. (1913). Primer statisticheskogo issledovanija nad tekstom «Evgenija Onegina», illjustrirujuschij svjaz' ispytanij v tsep (An example of statistical study on the text of «Eugene Onegin» illustrating the linking of events to a chain). *Izvestiya Akademii Nauk, SPb, VI seriya*, 7(93):153–162.

[4] Page, L. and Brin, S. and Motwani, R. and Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

[5] Shannon, C.E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

[6] Von Hilgers, P. and Langville, A.N. (2006). The five greatest applications of Markov chains. In *Proceedings of the Markov Anniversary Meeting*, pages 155–158. Citeseer.

[7] Xenakis, I. (1992). *Formalized music: thought and mathematics in composition*. Number 6. Pendragon Press.