DATA ANALYTICS

Unit 5: Advanced Techniques

Bharathi R
Department of Computer Science and Engineering

DATA ANALYTICS

Unit 5: Markov chains in Predictive Analytics

Bharathi R

Department of Computer Science and Engineering

Markov Chains in Predictive Analytics

One of the primary applications of Markov chain is predicting the values of Xn in the future.
For example, assume that the initial distribution of customers in 4 states is

$$P_I = (450, 225, 175, 150).$$

Assume that the one-step transition matrix P is as

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \end{array} \begin{pmatrix} 0.8189 & 0.0882 & 0.0472 & 0.0457 \\ 0.1128 & 0.7180 & 0.0902 & 0.0789 \\ 0.2077 & 0.0984 & 0.6011 & 0.0929 \\ 0.0663 & 0.0964 & 0.0964 & 0.7410 \end{pmatrix}$$

Using Chapman–Kolmogorov relationship [Eq. (16.12)], we can show that the distribution of customers after n periods is given by $P_I \times P^n$, where P_I is the initial distribution of customers across various states and P is the one-step transition matrix. For example, the distribution of customers after 4 weeks is $P_I \times P^4$. That is

$$\begin{pmatrix} 450 & 225 & 175 & 150 \end{pmatrix} \times \begin{pmatrix} 0.8189 & 0.0882 & 0.0472 & 0.0457 \\ 0.1128 & 0.7180 & 0.0902 & 0.0789 \\ 0.2077 & 0.0849 & 0.6011 & 0.0929 \\ 0.0663 & 0.0964 & 0.0964 & 0.7410 \end{pmatrix}^4 = \begin{pmatrix} 417.84 & 243.06 & 150.69 & 188.41 \end{pmatrix}$$

So after 4 periods, the distribution of customers will be

State 1: 417.84; State 2: 243.06; State 3: 150.69; and State 4: 188.41

Stationary Distribution in a Markov Chain

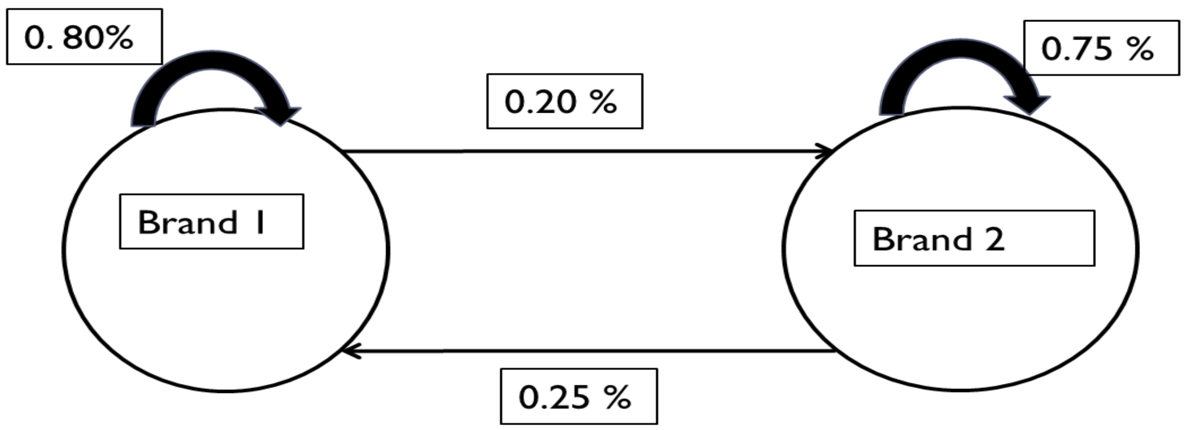
Consider brand switching between two brands (B1 and B2) and let the initial market share be as shown in the following vector:

$P_I = (0.2\ 0.8)$

Transition probability

	Brand 1	Brand 2
Brand 1	0.80	0.20
Brand 2	0.25	0.75

State transition diagram between brands.



Stationary Distribution in a Markov Chain

In Table given, both rows of the matrix **P^n** converge to 0.555556 and 0.444444 as the value of n increases.

The market share of brands 1 and 2 converges to 0.555556 and 0.444444, respectively.

The values (0.555556, 0.444444) are the stationary probability distribution of the Markov chain or equilibrium probabilities.

The values can be interpreted as long-run market shares of the brands.

TABLE 16.7 Shows the values of P^n and the market share of brands after n periods ($P_I P^n$)

		Brand 1	Brand 2	Market Share *n* Periods		
P^I		0.2	0.8			
		Brand 1	Brand 2		Brand 1	Brand 2
P	Brand 1	0.8	0.2	1 ($P_I P^1$)	0.36	0.64
	Brand 2	0.25	0.75			
P^2	Brand 1	0.69	0.31	2 ($P_I P^2$)	0.448	0.552
	Brand 2	0.3875	0.6125			
P^4	Brand 1	0.596225	0.403775	4 ($P_I P^4$)	0.52302	0.47698
	Brand 2	0.504719	0.495281			
P^8	Brand 1	0.559277	0.440723	8 ($P_I P^8$)	0.552578	0.447422
	Brand 2	0.550904	0.449096			
P^{16}	Brand 1	0.555587	0.444413	16 ($P_I P^{18}$)	0.555531	0.444469
	Brand 2	0.555517	0.444483			
P^{32}	Brand 1	0.555556	0.444444	32 ($P_I P^{32}$)	0.555556	0.444444
	Brand 2	0.555556	0.444444			
P^{64}	Brand 1	0.555556	0.444444	64 ($P_I P^{64}$)	0.555556	0.444444
	Brand 2	0.555556	0.444444			

Stationary Distribution in a Markov Chain

Let $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$ be the stationary distribution. Then it satisfies the following system of equations:

$$\pi_j = \sum_{k=1}^{m} \pi_k P_{kj} \quad (16.20)$$

$$\sum_{k=1}^{m} \pi_k = 1 \quad (16.21)$$

The system of equations in Eq. (16.20) can be written as

$$\pi = \pi \mathbf{P} \quad (16.22)$$

Stationary Distribution in a Markov Chain

The stationary distribution equation for the matrix in Transition Table is given by

$$(\pi_1 \quad \pi_2) = (\pi_1 \quad \pi_2)\begin{pmatrix} 0.80 & 0.20 \\ 0.25 & 0.75 \end{pmatrix}$$

That is

$$\pi_1 = 0.80\pi_1 + 0.25\pi_2$$
$$\pi_2 = 0.20\pi_1 + 0.75\pi_2$$

Since π_1 and π_2 are probabilities, we have

$$\pi_1 + \pi_2 = 1$$

$$0.20\pi_1 - 0.25\pi_2 = 0$$
$$\pi_1 + \pi_2 = 1$$

Solving the above system of equations, we get $\pi_1 = 0.555556$ and $\pi_2 = 0.444444$. That is, in the long run, the markets shares of brand 1 and brand 2 will converge to 0.555556 and 0.444444, respectively. The stationary distribution will be independent of the initial probability distribution $\mathbf{P}_\mathbf{I}$.

Regular Matrix

A matrix P is called a regular matrix, when for some *n*, all entries of P^n will be greater than zero, that is for some *n* $P_{ij}^{n} > 0$

Consider the matrix:

$$\mathbf{P} = \begin{pmatrix} 0.2 & 0 & 0.8 \\ 0.5 & 0 & 0.5 \\ 0.3 & 0.7 & 0 \end{pmatrix}$$

Then

$$\mathbf{P}^2 = \begin{pmatrix} 0.28 & 0.56 & 0.16 \\ 0.25 & 0.35 & 0.4 \\ 0.41 & 0 & 0.59 \end{pmatrix} \text{ and } \mathbf{P}^3 = \begin{pmatrix} 0.384 & 0.112 & 0.504 \\ 0.345 & 0.280 & 0.375 \\ 0.259 & 0.413 & 0.328 \end{pmatrix}$$

Note that although the matrix P has zero entries ($P_{12} = P_{22} = P_{33} = 0$), in $\mathbf{P}^3$ all entries are greater than zero, thus matrix **P** is a **regular matrix**. A regular matrix will have stationary distribution and satisfy the system of equations as shown below,

$$\pi_j = \sum_{k=1}^{m} \pi_k P_{kj} \qquad \sum_{k=1}^{m} \pi_k = 1$$

Example

The number of flights cancelled by an airline daily is modelled using a Markov chain. The states of the chain and the description of states are given in Table 16.8. The revenue loss (in millions of rupees) due to cancellation of flights in various states is given in Table 16.9. The transition probability matrix between states is shown in Table 16.10.

(a) If there are no cancellations initially, what is the probability that there will be at least one cancellation after 2 days?

(b) Calculate the steady-state expected loss due to cancellation of flights.

TABLE 16.8 States representing cancellation of flights

State	Description
0	No cancellations
1	One cancellation
2	Two cancellations
3	More than 2 cancellations

TABLE 16.9 Revenue loss due to cancellations

State	0	1	2	3
Loss	0	4.5	10.0	16.0

TABLE 16.10 State transition matrix between flight cancellations

	0	1	2	3
0	0.45	0.30	0.20	0.05
1	0.15	0.60	0.15	0.10
2	0.10	0.30	0.40	0.20
3	0	0.10	0.70	0.20

Example

Solution:

(a) If there are no cancellations initially, then the initial state vector is $\mathbf{P}_1 = [1\ 0\ 0\ 0]$. The probability distribution after two days is

$$\mathbf{P}_1\mathbf{P}^2 = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.45 & 0.30 & 0.20 & 0.05 \\ 0.15 & 0.60 & 0.15 & 0.10 \\ 0.10 & 0.30 & 0.40 & 0.20 \\ 0 & 0.10 & 0.70 & 0.20 \end{pmatrix}^2 = \begin{pmatrix} 0.2675 & 0.38 & 0.25 & 0.1025 \end{pmatrix}$$

Probability that there will be at least one cancellation after 2 days $= 0.38 + 0.25 + 0.1025 = 0.7325$.

Example

(b) To calculate the steady-state expected loss, we have to calculate the steady-state distribution. The steady-state distribution will satisfy the following system of equations:

$$\pi_0 = 0.45\pi_0 + 0.15\pi_1 + 0.10\pi_2$$
$$\pi_1 = 0.30\pi_0 + 0.60\pi_1 + 0.30\pi_2 + 0.10\pi_3$$
$$\pi_2 = 0.20\pi_0 + 0.15\pi_1 + 0.40\pi_2 + 0.70\pi_3$$
$$\pi_3 = 0.05\pi_0 + 0.10\pi_1 + 0.20\pi_2 + 0.20\pi_3$$
$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

TABLE 16.10 State transition matrix between flight cancellations

	0	1	2	3
0	0.45	0.30	0.20	0.05
1	0.15	0.60	0.15	0.10
2	0.10	0.30	0.40	0.20
3	0	0.10	0.70	0.20

Solving the above system of equations we get

$$\begin{pmatrix}\pi_0 & \pi_1 & \pi_2 & \pi_3\end{pmatrix} = \begin{pmatrix}0.163 & 0.390 & 0.311 & 0.137\end{pmatrix}$$

The steady-state expected loss is $\sum_{i=0}^{3} \pi_i \times L_i$, where L_i is the expected revenue loss in state i (Table 16.9). Hence

$$\sum_{i=0}^{3} \pi_i \times L_i = 0.163 \times 0 + 0.390 \times 4.5 + 0.311 \times 10 + 0.137 \times 16 = 7.05$$

Classification of States in a Markov Chain

Not all Markov chains will have stationary probability distribution. To derive the necessary and sufficient conditions for existence of stationary distribution of a Markov chain, we have to understand different classes of states that exist in a Markov chain.

The classes of states are

- Accessible state
- Communicating state
- Recurrent and Transient state
- Positive recurrent and Null-Recurent state
- Periodic and Aperiodic state

Accessible State

A state *j* is accessible (or reachable) from state *i* if there exists a *n* such that

$$P_{ij}^{n} > 0.$$

That is, there exists a path from state *i* to state *j*.

Communicating States

Two states i and j are communicating states when there exists n and m such that

$$P_{ij}^{n} > 0 \text{ and } P_{ji}^{m} > 0.$$

That is, state j can be reached (accessible) from state i and similarly state i can be reached from state j.

A Markov chain is called irreducible if all states of the chain communicate with each other.

Recurrent and Transient States

A state i of a Markov chain is called a recurrent state when

$$\sum_{n=1}^{\infty} P_{ii}^{n} = \infty$$

That is, if the state i is recurrent then the Markov chain will visit state i infinite number of times in the long run.

If state i is recurrent and states i and j are communicating states, then state j is also a recurrent state.

A state k of a Markov chain is called a transient state when

$$\sum_{n=1}^{\infty} P_{kk}^{n} < \infty$$

That is, state k is called a transient state when $\sum_{n=1}^{\infty} P_{kk}^{n} < \infty$ is finite.

This means it is possible that the Markov chain may not return to state k in the long run.

First Passage Time and Mean Recurrence Time

First passage time is the probability that the Markov chain will enter state *i* exactly after *n* steps for the first time after leaving state *i*, that is

$$f_{ii}^{n} = P\left[X_n = i, X_k \neq i, k = 1, 2, \ldots, n-1 \mid X_0 = i\right]$$

Mean recurrence time is the average time taken to return to state *i* after leaving state *i*. Mean recurrence time μ_{ii} is given by

$$\mu_{ii} = \sum_{n=1}^{\infty} n \times f_{ii}^{n}$$

If the mean recurrence time is finite (μ_{ii} is finite), then the recurrent state is called a **positive recurrent state** and if it is infinite then it is called **null-recurrent state**.

Periodic and Aperiodic State

Periodic state is a special case of recurrent state in which *d*(*i*) is the greatest common divisor of *n* such that

$$P_{ii}^{n} > 0$$

If *d*(*i*) = 1, it is called aperiodic state and if *d*(*i*) $\geq$ 2, then it is called a periodic state.

In the matrix shown in Table, for state 1

$$P_{11}^{3} = 1, P_{11}^{6} = 1, P_{11}^{9} = 1, \text{ and } P_{11}^{n} = 0$$

when n is not a multiple of 3. That is, the greatest common divisor is 3 which means that the periodicity is 3.

TABLE 16.11 Transition matrix

		1	2	3
$P=$	1	0	1	0
	2	0	0	1
	3	1	0	0

Ergodic Markov Chain

A state *i* of a Markov chain is ergodic when it is positive recurrent and aperiodic. Markov chain in which all states are positive recurrent and aperiodic is called an ergodic Markov chain.
For an ergodic Markov chain, a stationary distribution exists that satisfies the system of equations

$$\pi_j = \sum_{k=1}^{m} \pi_k P_{kj}$$

$$\sum_{k=1}^{m} \pi_k = 1$$

Limiting Probability

In a Markov chain, the limiting probability is given by:

$$\underset{n \to \infty}{Lt} \; p_{ij}^{n}$$

The main difference between limiting probability and stationary distribution is that, stationary distribution when exists is unique. Whereas limiting probability may not be unique.

DATA ANALYTICS

Unit 5: Markov chains contd.

Bharathi R

Department of Computer Science and Engineering

Markov Chains with Absorbing States

A state *i* of a Markov chain is called an absorbing state when $P_{ii} = 1$, that is if the system enters this state, it will remain in the same state. Many real-life problems such as bad debt (non-performing assets), bankruptcy, customer churn, employee attrition and so on can be modelled using absorbing state Markov chain.

Absorbing state Markov chain is a Markov chain in which there is at least one state k such that $P_{kk} = 1$. Non-absorbing states in an absorbing state Markov chain are transient states.

While using absorbing state Markov chains in analytics problem solving, we would like to learn the following from an absorbing state Markov chain:

1. The probability of eventual absorption to a specific absorbing state (when there are more than one absorbing states) from various transient states of the Markov chain.

2. The expected time to absorption from a transient state to absorbing states.

The above questions are answered using canonical form of the transition matrix.

Canonical Form of the Transition Matrix of an Absorbing State Markov Chain

The rows of the transition probability matrix of an absorbing state Markov chain can be rearranged such that the top rows are assigned for absorbing states followed by transient states (the idea here is to group the absorbing state and non-absorbing states).

Let A and T be the set of absorbing and transient states, respectively, in the Markov chain.

Then the transition probability matrix can be arranged such that

P =		A	T
	A	I	0
	T	R	Q

The Matrix P

The matrix P is divided into 4 matrices I, 0, R, and Q, where

- Matrix **I** is the identity matrix. It corresponds to transition within absorbing states.

- Matrix **0** is a matrix in which all elements are zero. Here the elements correspond to transition between an absorbing state and transient states.

- Matrix **R** represents the probability of absorption from a transient state to an absorbing state.

- Matrix **Q** represents the transition between transient states.

Expected time to absorption

To calculate the eventual probability of absorption, we would like to calculate the long-run (limiting probability) value of R in the above matrix. When we multiply the canonical form of the matrix, we get

$$\mathbf{P}^n = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \left(\sum_{k=0}^{n-1} \mathbf{Q}^k\right)\mathbf{R} & \mathbf{Q}^n \end{pmatrix}$$

For large n, the matrix $\left(\sum_{k=0}^{n-1} \mathbf{Q}^k\right)\mathbf{R}$ will give the probability of eventual absorption to an absorbing state.

As $n \to \infty$, we can show that $\sum_{k=0}^{n-1} \mathbf{Q}^k = \mathbf{F} = (\mathbf{I} - \mathbf{Q})^{-1}$. The matrix F is called the fundamental matrix and the matrix FR is the probability of eventual absorption into an absorbing state from a transient state. The expected time to absorption is given by

Expected time to absorption = Fc

where c is a unit vector. That is, the row sum of the fundamental matrix gives the expected duration for absorption (that is, expected time it takes to reach an absorbing state from a transient state).

Example 16.5

Airwaves India (AI) is a mobile phone service provider based in Allahabad, India that provides several value-added services such as mobile data, video conferencing, etc. The market is highly competitive and AI faces high churn rate among its customers. The customers of AI are categorized into different states as listed below:

STATE 1 : Customer churn that generated no revenue/profit
STATE 2 : Customer churn that generated INR 200 profit per month on average (customer uses the service only for incoming calls and data)
STATE 3 :Customer state that generated INR 300 profit per month on average
STATE 4 :Customer state that generated INR 400 profit per month on average
STATE 5 :Customer state that generated INR 600 profit per month on average
STATE 6 :Customer state that generated INR 800 profit per month on average

Example 16.5

The transition probability values between different states are shown in Table 16.12.

TABLE 16.12 Transition probability matrix (based on monthly data)

	1	2	3	4	5	6
1	1	0	0	0	0	0
2	0	1	0	0	0	0
3	0.05	0.05	0.90	0	0	0
4	0.10	0.05	0	0.80	0.05	0
5	0.20	0.10	0	0.05	0.60	0.05
6	0.10	0.20	0	0	0	0.70

(a) If a customer is in state 6, calculate the probability of eventual absorption in state 2?
(b) Calculate the expected value of time taken to absorption if the current state is 4.

Solution: 16.5

(a) To calculate the probability of absorption of a customer in state 6 to state 2, we have to calculate **FR**.

TABLE 16.12	Transition probability matrix (based on monthly data)					
	1	2	3	4	5	6
1	1	0	0	0	0	0
2	0	1	0	0	0	0
3	0.05	0.05	0.90	0	0	0
4	0.10	0.05	0	0.80	0.05	0
5	0.20	0.10	0	0.05	0.60	0.05
6	0.10	0.20	0	0	0	0.70

The matrix Q is given by

$$Q = \begin{bmatrix} 0.9 & 0 & 0 & 0 \\ 0 & 0.8 & 0.05 & 0 \\ 0 & 0.05 & 0.6 & 0.05 \\ 0 & 0 & 0 & 0.7 \end{bmatrix}$$

$$I - Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.9 & 0 & 0 & 0 \\ 0 & 0.8 & 0.05 & 0 \\ 0 & 0.05 & 0.6 & 0.05 \\ 0 & 0 & 0 & 0.7 \end{bmatrix} = \begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.2 & -0.05 & 0 \\ 0 & -0.05 & 0.4 & -0.05 \\ 0 & 0 & 0 & 0.3 \end{bmatrix}$$

$$F = (1 - Q)^{-1} = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 5.1613 & 0.6452 & 0.1075 \\ 0 & 0.6452 & 2.5806 & 0.4301 \\ 0 & 0 & 0 & 3.3333 \end{bmatrix}$$

Solution: 16.5

Probability of absorption FR is given by

$$FR = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 5.1613 & 0.6452 & 0.1075 \\ 0 & 0.6452 & 2.5806 & 0.4301 \\ 0 & 0 & 0 & 3.3333 \end{bmatrix} \times \begin{bmatrix} 0.05 & 0.05 \\ 0.1 & 0.05 \\ 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.6559 & 0.3441 \\ 0.6237 & 0.3763 \\ 0.3333 & 0.6667 \end{bmatrix}$$

That is, if the current customer state is 6, the probability of absorption into churn state 2 is 0.6667.

(b) Expected value of time to absorption is given by Fc:

$$Fc = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 5.1613 & 0.6452 & 0.1075 \\ 0 & 0.6452 & 2.5806 & 0.4301 \\ 0 & 0 & 0 & 3.3333 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 5.91 \\ 3.65 \\ 3.33 \end{bmatrix}$$

Expected value of time to absorption when the current state is 4 is 5.91 months.

References

Text Book:

“Business Analytics, The Science of Data-Driven Decision Making”, U. Dinesh Kumar, Wiley 2017

Markov chains contd (absorbing states, expected duration to reach a state) [ch 16.4.5 - ch 16.8]

THANK YOU

Prof Bharathi R
Professor, Department of Computer Science
rbharathi@pes.edu