Random Walk on Graphs

Applications to Ranking

Shaobo Han

Duke University

Learning Objectives

1. Understand the idea of Markov chains and transition probability matrix
2. Have a conceptual understanding of stationary distribution
3. Apply the ranking method to real-world applications

Problem: UEFA Euro 2016 Team Competence

Who will win Euro 2016?

Table 1 : 9 Selected European Teams

Index	Team	Abbreviation
1	Belgium	BEL
2	Spain	ESP
3	Germany	DEU
4	Portugal	PRT
5	England	ENG
6	Italy	ITA
7	Netherlands	NLD
8	Russia	RUS
9	France	FRA

Data: Recent Exhibition Games

Table 2 : Recent Matches within Last 5 Months

Date	Home Team	Score	Away Team	City
Nov.13 2015	Spain	2:0	England	Alicante, Spain
Nov.13 2015	France	2:0	Germany	Saint-Denis, France
Nov.13 2015	Belgium	3:1	Italy	Brussels, Belgium
Nov.13 2015	Russia	1:0	Portugal	Krasnodar, Russia
Nov.17 2015	England	2:0	France	London, England
Mar.24 2016	Italy	1:1	Spain	Udine, Italy
Mar.25 2016	Netherlands	2:3	France	Amsterdam, Netherlands
Mar.26 2016	Germany	2:3	England	Berlin, Germany
Mar.29 2016	England	1:2	Netherlands	London, England
Mar.29 2016	Germany	4:1	Italy	Munich, Germany
Mar.29 2016	Portugal	2:1	Belgium	Leiria, Portugal
Mar.29 2016	France	4:2	Russia	Saint-Denis, France

Data: Recent Exhibition Games

Table 3 : Recent Matches within Last 5 Months

Date	Home Team	Score	Away Team	City
Nov.13 2015	Spain	2:0	England	Alicante, Spain
Nov.13 2015	France	2:0	Germany	Saint-Denis, France
Nov.13 2015	Belgium	3:1	Italy	Brussels, Belgium
Nov.13 2015	Russia	1:0	Portugal	Krasnodar, Russia
Nov.17 2015	England	2:0	France	London, England
Mar.24 2016	Italy	1:1	Spain	Udine, Italy
Mar.25 2016	Netherlands	2:3	France	Amsterdam, Netherlands
Mar.26 2016	Germany	2:3	England	Berlin, Germany
Mar.29 2016	England	1:2	Netherlands	London, England
Mar.29 2016	Germany	4:1	Italy	Munich, Germany
Mar.29 2016	Portugal	2:1	Belgium	Leiria, Portugal
Mar.29 2016	France	4:2	Russia	Saint-Denis, France

Part I: Markov Chains, Transition Probability, and Stationary Distribution

X_n is a discrete time **Markov chain** with **transition matrix** $p(i, j)$ if for any j, i, $i_{n-1}, \ldots, i_0$

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = p(i, j) \tag{1}$$

1. $p(i, j) \geq 0$, since they are probabilities
2. $\sum_j p(i, j) = 1$, since when $X_n = i$, X_{n+1} will be in some state j

X_n is a discrete time **Markov chain** with **transition matrix** $p(i,j)$ if for any j, i, $i_{n-1}, \ldots, i_0$

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = p(i,j) \tag{1}$$

1. $p(i,j) \geq 0$, since they are probabilities
2. $\sum_j p(i,j) = 1$, since when $X_n = i$, X_{n+1} will be in some state j

Multistep Transition Probabilities: the probability of going from i to j in $m > 1$ steps:

$$P(X_{n+m} = j | X_n = i) = p^m(i,j) \tag{2}$$

The m step transition probability $P(X_{n+m} = j | X_n = j)$ is the mth power of the transition matrix p.

Example: Weather Chain

Let X_n be the weather on day n in Durham, NC, which we assume is either: $1 = \text{rainy}$, or $2 = \text{sunny}$, we propose a Markov chain model for the weather,

Table 4 : Transition Probability

	1	2
1	0.6	0.4
2	0.2	0.8

Suppose that the **initial distribution** is $q(1) = 0.3$, and $q(2) = 0.7$. After one step, the probability of being at space $(\text{rainy}, \text{sunny})$ is

$$qp = (0.3, 0.7) \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} = (0.32, 0.68) \tag{3}$$

Example: Weather Chain

Let X_n be the weather on day n in Durham, NC, which we assume is either: $1 = \mathrm{rainy}$, or $2 = \mathrm{sunny}$, we propose a Markov chain model for the weather,

Table 4 : Transition Probability

	1	2
1	0.6	0.4
2	0.2	0.8

Suppose that the **initial distribution** is $q(1) = 0.3$, and $q(2) = 0.7$. After one step, the probability of being at space $(\mathrm{rainy}, \mathrm{sunny})$ is

$$qp = (0.3, 0.7) \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} = (0.32, 0.68) \tag{3}$$

Question: Two steps?

Stationary Distributions

The probability at state j after n steps,

$$\begin{aligned} P(X_n = j) &= \sum_i P(X_0 = i, X_n = j) = \sum_i P(X_0 = i) P(X_n = j | X_0 = i) \\ &= \sum_i q(i) p^n(i, j) = q p^n \end{aligned} \tag{4}$$

Stationary Distributions

The probability at state j after n steps,

$$\begin{aligned} P(X_n = j) &= \sum_i P(X_0 = i, X_n = j) = \sum_i P(X_0 = i) P(X_n = j | X_0 = i) \\ &= \sum_i q(i) p^n(i,j) = qp^n \end{aligned} \tag{4}$$

Question: What is the **long-run fraction** of days that are sunny?

$$\lim_{n \to \infty} qp^n = ?$$

Stationary Distributions

Definition: If $\pi p = \pi$, then distribution π is called a **stationary distribution** for the Markov chain p.

Definition: If $\pi p = \pi$, then distribution π is called a **stationary distribution** for the Markov chain p.

Weather Chain Example:

$$\begin{cases} 0.6\pi_1 + 0.2\pi_2 &= \pi_1 \\ 0.4\pi_1 + 0.8\pi_2 &= \pi_2 \\ \pi_1 + \pi_2 &= 1 \end{cases} \tag{5}$$

Definition: If $\pi p = \pi$, then distribution π is called a **stationary distribution** for the Markov chain p.

Weather Chain Example:

$$\begin{cases} 0.6\pi_1 + 0.2\pi_2 &= \pi_1 \\ 0.4\pi_1 + 0.8\pi_2 &= \pi_2 \\ \pi_1 + \pi_2 &= 1 \end{cases} \tag{5}$$

Answer: $\pi_1 = 1/3$, $\pi_2 = 2/3$.

$$qp^{20} = (0.3, 0.7)\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}^{20} = (0.3333, 0.6667) \tag{6}$$

Part II: Exercise: Ranking Euro National Teams

Results of Recent Exhibition Games

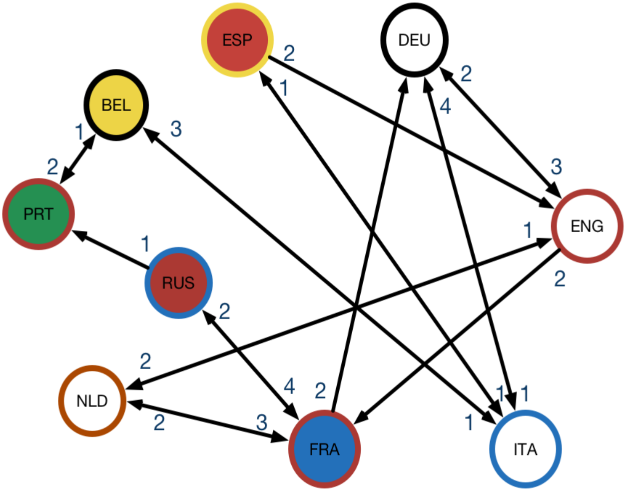

Figure 1 : Endorsement Graph Among Teams

Transition Probability Matrix

Table 5 : Weighted Adjacency Matrix

	BEL	ESP	DEU	PRT	ENG	ITA	NLD	RUS	FRA
BEL				2		1			
ESP						1			
DEU					3	1			2
PRT	1							1	
ENG		2	2				2		
ITA	3	1	4						
NLD					1				3
RUS									4
FRA					2		2	2	

Fill blank cells with zeros.

Transition Probability Matrix

Table 6 : Transition Probability Matrix

	BEL	ESP	DEU	PRT	ENG	ITA	NLD	RUS	FRA
BEL				2/3		1/3			
ESP						1			
DEU					3/6	1/6			2/6
PRT	1/2							1/2	
ENG		2/6	2/6				2/6		
ITA	3/8	1/8	4/8						
NLD					1/4				3/4
RUS									4/4
FRA					2/6		2/6	2/6	

Fill blank cells with zeros.

Group Exercise

1. Split into two groups of size 4 or 5
2. Download the file from the link `http://bit.ly/2016eurocup`
3. Choose a initial distribution q: either uniform or random
4. Arrive at the stationary distribution π
   - Multiplying the transition matrix p by itself over and over again
   - Hint: you may find the R package: **expm** useful
5. (Optional) Compare against:
   - Normalizing the first eigenvector of p to unit vector
   - Solving a series of linear equations $\pi p = \pi$ and $\sum \pi = 1$
6. Rank the teams according to the share of probability fluid

Discussion:

1. Does this Markov chain has a stationary distribution?
2. Does the initial distribution matter?

Results

Table 7 : Results: Stationary Distributions

	π	qp^{10}	qp^{50}
BEL	0.0591	0.0593	0.0591
ESP	0.0675	0.0663	0.0675
DEU	0.1069	0.1052	0.1069
PRT	0.0394	0.0390	0.0394
ENG	0.1632	0.1656	0.1632
ITA	0.1051	0.1056	0.1051
NLD	0.1313	0.1330	0.1313
RUS	0.0966	0.1001	0.0966
FRA	0.2308	0.2259	0.2308

Team competence ranking (according to the recent performance):

Team	FRA	ENG	NLD	DEU	ITA	RUS	ESP	BEL	PRT
Competence	0.2308	0.1632	0.1313	0.1069	0.1051	0.0966	0.0675	0.0591	0.0394

Brainstorming: Other Applications

Ranking via random walk on graphs:

- The popularity of food & dining places on Duke campus
- The most influential friend (scholar) on your social (academia) network
- Webpages according to the relevance (Google PageRank algorithm)

Next Class: A Markov chain p is **ergodic** if there exists a unique stationary distribution π and for every (initial) distribution q the limit

$$\lim_{n\to\infty} qp^n = \pi$$

More Readings

References:

1. Chapter 1. Essentials of Stochastic Processes (Springer Texts in Statistics), 2nd Edition, by Richard Durrett, 2012

Thank you!