Randomisierte Algorithmen

8. Markov-Ketten

Thomas Worsch

Fakultät für Informatik
Karlsruher Institut für Technologie

Wintersemester 2019/2020

Überblick

Grundlegendes zu Markov-Ketten

Irreduzible und ergodische Markov-Ketten

Überblick

Grundlegendes zu Markov-Ketten

Irreduzible und ergodische Markov-Ketten

Markov-Kette

- stochastischer Prozess in diskreter Zeit
- schrittweiser Übergang von einem Zustand zum nächsten
- festgelegt durch $M = (S, \mathbf{P})$:
  - S: (bei uns immer) endliche Menge von *Zuständen*
  - $\mathbf{P} = (P_{ij})$: zeilenstochastische $S \times S$-Matrix von *Übergangswahrscheinlichkeiten*:
    für $i, j \in S$ ist $0 \leq P_{ij} \leq 1$ und $\sum_j P_{ij} = 1$
  - P_{ij} ist Wahrscheinlichkeit (W.keit), dass M von Zustand i in Zustand j übergeht.
- Beachte:
  - P_{ij} hängt nur von i und j ab
  - nicht etwa von noch früheren Zuständen oder Anzahl Schritte oder …

Markov-Ketten ↔ Graphen

- jeder Markov-Kette M entspricht ein gerichteter Graph G_M:
  - Kante $i \longrightarrow j$ genau dann, wenn $P_{ij} > 0$
    u. U. gewichtet mit P_{ij}
- jedem Graph G entspricht Markov-Kette M_G
  (*einfacher Random Walk*)
  - $P_{ij} = 0$, falls keine Kante zwischen i und j
  - $P_{ij} = 1/|\{ j \mid (i, j) \in E \}|$ sonst

Vereinbarung

- X_t: Zufallsvariable für Zustand zum Zeitpunkt t,
- bei Markovketten also

$$\begin{aligned} &\mathbf{P}(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) \\ &= \mathbf{P}(X_{t+1} = j \mid X_t = i) \\ &= P_{ij} \end{aligned}$$

- X_0 … *Anfangszustand* …
  - im allgemeinen randomisiert
  - manchmal egal …

Rechnung

- wenn zum Zeitpunkt t
  - $\mathbf{q}$ Zeilenvektor
  - $\mathbf{q}_i$ W.keit für Zustand i
- dann zum Zeitpunkt $t+1$
  - $\mathbf{qP}$ entsprechender Zeilenvektor:

$$\begin{aligned}\mathbf{P}(X_{t+1} = j) &= \sum_i \mathbf{P}(X_t = i)\,\mathbf{P}(X_{t+1} = j \mid X_t = i) \\ &= \sum_i \mathbf{q}_i P_{ij} = (\mathbf{qP})_j\end{aligned}$$

- $\mathbf{qP}^k$ die Verteilung nach k Schritten
  - W.keit $P_{ij}^{(k)}$ in k Schritten von i nach j überzugehen

$$P_{ij}^{(k)} = (\mathbf{P}^k)_{ij}$$

Abgeschlossene und irreduzible Teilmengen

- nichtleere Teilmenge $C \subseteq S$ von Zuständen *abgeschlossen*, falls $\forall\, i \in C \colon \forall\, j \in S \setminus C \colon P_{ij} = 0$.
- S ist immer abgeschlossen
- C heißt *irreduzibel*, falls C abgeschlossen, aber keine echte Teilmenge von C abgeschlossen
- Markov-Kette *irreduzibel*, falls ganz S irreduzibel

- verschiedene irreduzible Teilmengen sind disjunkt

Transiente und rekurrente Zustände

Es seien $C_1, \ldots, C_r$ alle irreduziblen Teilmengen einer Markov-Kette S und $T = S \setminus (C_1 \cup \cdots \cup C_r)$.

- Die Zustände in T heißen *transient*,
- die Zustände in den C_k *rekurrent* oder *persistent*.

Notation

- Wahrscheinlichkeit, von i nach t Schritten *erstmals* nach j überzugehen:

$$f_{ij}^{(t)} = \mathbf{P}(X_t = j \wedge \forall\, 1 \le s \le t-1 : X_s \ne j \mid X_0 = i)$$

- Wahrscheinlichkeit von Zustand i aus irgendwann Zustand j zu erreichen:

$$f_{ij}^{*} = \sum_{t>0} f_{ij}^{(t)}$$

- Erwartungswert für die benötigte Anzahl Schritte, um von Zustand i irgendwann erstmals Zustand j zu erreichen:

$$m_{ij} = \begin{cases} \sum_{t\ge1} t \cdot f_{ij}^{(t)} & \text{falls } f_{ij}^{*} = 1 \\ \infty & \text{sonst} \end{cases}$$

Charakterisierung transienter Zustände

Für endliche Markov-Ketten gilt:
Ein Zustand i ist genau dann *transient*, wenn eine der folgenden (äquivalenten) Bedingungen erfüllt ist:

- $f_{ii}^* < 1$.
- $\sum_{t \geq 0} p_{ii}^{(t)} < \infty$.
- Ein Random Walk, der in i startet, kehrt mit Wahrscheinlichkeit 0 unendlich oft nach i zurück.

Charakterisierung rekurrenter Zustände

Für endliche Markov-Ketten gilt:
Ein Zustand i ist genau dann *rekurrent*, wenn eine der folgenden (äquivalenten) Bedingungen erfüllt ist:

- $f_{ii}^* = 1$.
- $\sum_{t \geq 0} p_{ii}^{(t)} = \infty$.
- Ein Random Walk, der in i startet, kehrt mit Wahrscheinlichkeit 1 unendlich oft nach i zurück.

Überblick

Irreduzible Markov-Ketten

für uns vor allem irreduzible Markov-Ketten interessant

- ganz S die einzige irreduzible Teilmenge
- es gibt keine transienten Zustände
- zugehöriger Graph streng zusammenhängend

Perioden und Aperiodizität

- *Periode* d_i eines Zustandes i:
  größter gemeinsamer Teiler aller Zahlen in
  $N_i = \{k \in \mathbb{N}_+ \mid P_{ii}^{(k)} > 0\}$.
- Zustand mit Periode 1 heißt auch *aperiodisch*.
- Ein aperiodischer rekurrenter Zustand heißt auch *ergodisch*.

Aperiodische und ergodische Markov-Ketten

- Eine Markov-Kette ist *aperiodisch*, wenn alle ihre Zustände aperiodisch sind.
- Eine irreduzible und aperiodische Markov-Kette heißt auch *ergodisch*.

Für aperiodische Zustände gilt *nicht* automatisch,
dass $P_{ii}^{(k)} > 0$ ist für *alle* k.

Aber immerhin …

Lemma

- Es sei $M \subseteq \mathbb{N}$ eine Menge natürlicher Zahlen mit der Eigenschaft, dass
  - $M + M = \{k + \ell \mid k, \ell \in M\} \subseteq M$ und $\gcd M = 1$.
- Dann gibt es ein $k_0 \in \mathbb{N}$ mit
  - $\{k_0\} + \mathbb{N}_0 = \{k_0, k_0 + 1, k_0 + 2, \ldots\} \subseteq M$, d. h.
  - M enthält ab irgendeinem k_0 *alle* natürlichen Zahlen.

(Übungsaufgabe)

Konstruktion aperiodischer Markov-Ketten

- Aus *nicht* aperiodischer Markov-Kette M mit Matrix $\mathbf{P}$ kann man aperiodische Markov-Kette M' konstruieren:

Konstruktion aperiodischer Markov-Ketten

- Aus *nicht* aperiodischer Markov-Kette M mit Matrix $\mathbf{P}$ kann man aperiodische Markov-Kette M' konstruieren:

$$\mathbf{P}' = \frac{1}{2}(\mathbf{I} + \mathbf{P})$$

  $\mathbf{I}$ bezeichne die Einheitsmatrix.
- diese Vorgehensweise erhält folgende Eigenschaften
  - Ist $\mathbf{wP} = \mathbf{w}$, dann ist auch $\mathbf{wP}' = \mathbf{w}$ und umgekehrt.
  - Die beiden Matrizen haben die gleichen Eigenvektoren.
  - Für die Eigenwerte gilt: Ist λ Eigenwert von $\mathbf{P}$, dann ist $1/2 + \lambda/2$ Eigenwert von $\mathbf{P}'$.

wird später hier benutzt

8.16 Satz

Potenzen ergodischer Markov-Ketten

Satz

Es sei $\mathbf{P}$ die Matrix einer ergodischen Markov-Kette. Dann gilt:

- $\mathbf{W} = \lim_{t \to \infty} \mathbf{P}^t$ existiert.
- $\mathbf{W}$ besteht aus identischen Zeilen $\mathbf{w}$.
- Alle Einträge von $\mathbf{w} = (w_1, \ldots, w_n)$ sind echt größer 0 und $\sum_{i=1}^{n} w_i = 1$.

Beweis (1)

- Da Markov-Kette ergodisch, gibt es eine Potenz $\mathbf{P}^k$, deren Einträge *alle* echt positiv sind.
  - (Übungsaufgabe)
- O. B. d. A. habe schon $\mathbf{P}$ diese Eigenschaft (sonst: arbeite mit $\mathbf{P}^k$).
- Sei $d > 0$ der kleinste in $\mathbf{P}$ vorkommende Eintrag.
- Sei zunächst $\mathbf{y}$ ein beliebiger Spaltenvektor.

1. Zeige: Wenn
   - m_0 und M_0 der kleinste resp. der größte Wert eines Vektors $\mathbf{y}$ und
   - m_1 und M_1 der kleinste resp. der größte Wert von $\mathbf{Py}$,

   dann
   - $m_0 \leq m_1 \leq M_1 \leq M_0$ und
   - $M_1 - m_1 \leq (1 - 2d)(M_0 - m_0)$

Beweis (2)

1. $M_1 - m_1 \leq (1-2d)(M_0 - m_0)$:
    - Die Einträge jeder Zeile von P addieren sich zu 1.
    - Für jedes i ist $(\mathbf{Py})_i = \sum_j P_{ij} y_j$. Offensichtlich ist
        - $m_1 = \min_i \sum_j P_{ij} y_j \geq dM_0 + (1-d)m_0 \geq m_0$
        - $M_1 = \max_i \sum_j P_{ij} y_j \leq dm_0 + (1-d)M_0 \leq M_0$
    - Also
        - $$\begin{aligned} M_1 - m_1 &\leq (dm_0 + (1-d)M_0) - (dM_0 + (1-d)m_0) \\ &= (1-2d)(M_0 - m_0) \end{aligned}$$
        - $m_0 \leq m_1 \leq M_1 \leq M_0$.

Beweis (3)

2. Induktion für kleinste und größte Einträge m_k und M_k von $\mathbf{P}^k\mathbf{y}$:
   - $M_k - m_k \leq (1-2d)^k(M_0 - m_0)$ und
   - $m_0 \leq m_1 \leq \cdots m_k \leq M_k \leq \cdots \leq M_1 \leq M_0$.
   - Die Folgen m_k und M_k sind beschränkt und monoton,
   - sie besitzen Grenzwerte $m = \lim_{k\to\infty} m_k$ bzw. $M = \lim_{k\to\infty} M_k$.

Beweis (4)

3. O. B. d. A. habe $\mathbf{P}$ mindestens 2 Zeilen und Spalten.
   - Dann ist $0 < d \leq 1/2$ und damit $0 \leq 1 - 2d < 1$.
   - $M_k - m_k \leq (1 - 2d)^k (M_0 - m_0)$,
   - also $\lim_{k\to\infty} M_k - m_k = 0$ und daher $M = m$.

4. Es sei $u = M = m$.
   - Alle Einträge in $\mathbf{P}^k\mathbf{y}$ liegen zwischen m_k und M_k,
   - Also ist $\lim_{k\to\infty} \mathbf{P}^k\mathbf{y} = \mathbf{u}$, wobei $\mathbf{u}$ der konstante Vektor ist, dessen Einträge alle gleich u sind.

Beweis (5)

5. Betrachte $\mathbf{y} = \mathbf{e}_j$ (j-ter Einheitsvektor):
   - $\mathbf{P}^k \mathbf{e}_j$ ist die j-te Spalte von $\mathbf{P}^k$.
   - Folge der $\mathbf{P}^k \mathbf{e}_j$ konvergiert gegen einen konstanten Vektor
   - also existiert $\lim_{k\to\infty} \mathbf{P}^k = \mathbf{W}$ und
   - besteht aus lauter konstanten Spalten, d. h.
   - aus lauter gleichen Zeilen $\mathbf{w}$

Beweis (6)

6. Alle Einträge in $\mathbf{w}$ sind echt größer 0:
   - $\mathbf{P}$ hat keine Nulleinträge.
   - Also gilt für jedes j: $\mathbf{Pe}_j$ enthält nur echt positive Werte,
   - d. h. $m_1 > 0$ und daher auch $m > 0$.
   - Dieses m ist die j-te Komponente von $\mathbf{w}$.
7. $\sum_{i=1}^{n} w_i = 1$:
   - alle Potenzen $\mathbf{P}^k$ sind stochastische Matrizen,
   - d. h. haben Zeilensumme 1

Stationäre Verteilung

Eine Verteilung $\mathbf{w}$ heißt *stationär*, falls $\mathbf{w} = \mathbf{w}\mathbf{P}$ ist.

8.18 Satz

Stationäre Verteilung ergodischer Markov-Ketten

Satz

Für jede ergodische Markov-Kette mit Matrix $\mathbf{P}$ und $\mathbf{w}$ wie eben gilt:

1. $\mathbf{wP} = \mathbf{w}$ stationäre Verteilung
2. Falls $\mathbf{vP} = \mathbf{v}$ ist, ist $\mathbf{v} = (\sum_j v_j)\mathbf{w}$.
3. Es gibt genau eine Wahrscheinlichkeitsverteilung $\mathbf{v}$ mit $\mathbf{vP} = \mathbf{v}$, nämlich $\mathbf{v} = \mathbf{w}$.

Beweis

1. $\mathbf{WP} = (\lim_{k\to\infty} \mathbf{P}^k) \cdot \mathbf{P} = \lim_{k\to\infty} \mathbf{P}^{k+1} = \mathbf{W}$
   Insbesondere gilt also für jede Zeile $\mathbf{w}$ von $\mathbf{W}$: $\mathbf{wP} = \mathbf{w}$.
2. Wenn $\mathbf{vP} = \mathbf{v}$ ist,
   dann $\mathbf{vP}^k = \mathbf{v}$ für jedes k und
   $\mathbf{vW} = \mathbf{v}$.
3. $r = \sum_j v_j$ die Summe der Komponenten von $\mathbf{v}$,
   dann $\mathbf{vW} = r\mathbf{w}$, also $\mathbf{v} = r\mathbf{w}$.
4. Unter allen Vektoren $r\mathbf{w}$ gibt es offensichtlich genau einen,
   für den die Summe aller Einträge gleich 1 ist.

Beobachtung

Graph $G = (V, E)$ mit $|V| = n \geq 2$ und $|E| = m$ sei endlich, zusammenhängend, ungerichtet und nicht bipartit.

- M_G ist irreduzibel:
  - G zusammenhängend
- M_G ist aperiodisch:
  - jeder Knoten in Zyklus der Länge 2
    - zu einem Nachbarn und zurück
  - jeder Knoten von G in einem Zyklus ungerader Länge:
    - G zusammenhängend und
    - ein Knoten in einem Zyklus ungerader Länge, da G nicht bipartit
- Also ist M_G ergodisch.

8.22 Lemma

In der stationären Verteilung $\mathbf{w}$ von M_G gilt für alle $v \in V$:

$$\mathbf{w}_v = d(v)/2m \ .$$

Insbesondere ist die stationäre Verteilung regulärer Graphen die Gleichverteilung.

8.23 Beweis

- stationäre Verteilung gegebenenfalls eindeutig
- rechne nach, dass $\mathbf{q}$ mit $q_v = d(v)/2m$ stationäre Verteilung ist:

$$
\begin{aligned}
(\mathbf{qP})_v &= \sum_{u \in V} q_u P_{uv} = \sum_{(u,v) \in E} q_u P_{uv} \\
&= \sum_{(u,v) \in E} \frac{d(u)}{2m} \cdot \frac{1}{d(u)} \\
&= \sum_{(v,u) \in E} \frac{1}{2m} = \frac{d(v)}{2m} = q_v \ .
\end{aligned}
$$

Stationäre Verteilung irreduzibler Markov-Ketten

- Wegen früher angemerkter Erhaltungseigenschaften gilt der dritte Teil der von Satz 8.18 für irreduzible Markov-Ketten, auch bei Nichtaperiodizität:
- *Jede irreduzible Markov-Kette* $\mathbf{P}$ *besitzt genau eine stationäre Verteilung* $\mathbf{w}$*.*
- Aber: $\lim_{k\to\infty} \mathbf{P}^k$ existiert für irreduzible Markov-Ketten im allgemeinen nicht.

Stationäre Verteilung irreduzibler Markov-Ketten

- Wegen früher angemerkter Erhaltungseigenschaften gilt der dritte Teil der von Satz 8.18 für irreduzible Markov-Ketten, auch bei Nichtaperiodizität:
- *Jede irreduzible Markov-Kette* $\mathbf{P}$ *besitzt genau eine stationäre Verteilung* $\mathbf{w}$*.*
- Aber: $\lim_{k\to\infty} \mathbf{P}^k$ existiert für irreduzible Markov-Ketten im allgemeinen nicht.
- Beispiel: $\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ und alle k: $\mathbf{P}^{2k} = \mathbf{I}$ und $\mathbf{P}^{2k+1} = \mathbf{P}$.

Bemerkung

- Für ergodische Markov-Ketten existiert $\lim_{t\to\infty} \mathbf{P}^t = \mathbf{W}$.
- Also existiert auch der Cesàro-Grenzwert $\lim_{t\to\infty} \mathbf{A}_t$, mit $\mathbf{A}_t = \frac{1}{t+1} \sum_{k=0}^{t} \mathbf{P}^k$
- und es ist $\lim_{t\to\infty} \mathbf{A}_t = \mathbf{W}$.
  - (Übungsaufgabe)
- $P_{ij}^{(k)}$ ist die Wahrscheinlichkeit, in k Schritten von i nach j zu gelangen.
- Also ist $(\mathbf{A}_t)_{ij}$ der erwartete Anteil von Zeitpunkten zwischen 0 und t, zu denen man in Zustand j ist, wenn man in Zustand i startet.
- Das ist nicht nur für ergodische Markov-Ketten so …

8.26 Satz

Es sei $\mathbf{P}$ die Übergangsmatrix einer irreduziblen Markov-Kette M. Dann gilt:

- $\lim_{t\to\infty} \mathbf{A}_t = \mathbf{W}$ existiert.
- Alle Zeilen von $\mathbf{W}$ sind gleich.
- Die Zeile $\mathbf{w}$ ist die eindeutig bestimmte stationäre Verteilung von M.

(ohne Beweis)

8.27 Satz

Für jede ergodische Markov-Kette $\mathbf{P}$ und jede Verteilung $\mathbf{v}$ gilt:

$$\lim_{k\to\infty} \mathbf{v}\mathbf{P}^k = \mathbf{w} \ .$$

8.28 Beweis

- Es ist $\lim_{k\to\infty} \mathbf{v}\mathbf{P}^k = \mathbf{v}\mathbf{W}$.
- Da sich die Einträge in $\mathbf{v}$ zu 1 summieren und alle Zeilen von $\mathbf{W}$ gleich $\mathbf{w}$ sind, ist $\mathbf{v}\mathbf{W} = \mathbf{w}$.

8.29 Satz

Für jede irreduzible Markov-Kette mit stationärer Verteilung $\mathbf{w} = (w_1, \ldots, w_n)$ gilt für alle i:

$$w_i = 1/m_{ii}$$

8.30 Beweis

1. $i \neq j$: $m_{ij} = P_{ij} \cdot 1 + \sum_{k \neq j} P_{ik}(m_{kj} + 1) = 1 + \sum_{k \neq j} P_{ik} m_{kj}$
2. $i = j$: $m_{ii} = P_{ii} \cdot 1 + \sum_{k \neq i} P_{ik}(m_{ki} + 1) = 1 + \sum_{k \neq i} P_{ik} m_{ki}$
3. Bezeichne $\mathbf{E}$ die Matrix, deren Einträge alle 1 seien,
   $\mathbf{M}$ die Matrix mit

$$\mathbf{M}_{ij} = \begin{cases} m_{ij} & \text{falls } i \neq j \\ 0 & \text{falls } i = j \end{cases}$$

   und $\mathbf{D}$ die Matrix mit

$$\mathbf{D}_{ij} = \begin{cases} 0 & \text{falls } i \neq j \\ m_{ii} & \text{falls } i = j \end{cases}$$

8.30 Beweis (2)

- Dann lassen sich die eben genannten Gleichungen ausdrücken als Matrixgleichung
$$\mathbf{M} + \mathbf{D} = \mathbf{E} + \mathbf{P}\mathbf{M} \; .$$
- Multiplizieren mit $\mathbf{w}$ von links ergibt
$$\mathbf{w}\mathbf{M} + \mathbf{w}\mathbf{D} = \mathbf{w}\mathbf{E} + \mathbf{w}\mathbf{P}\mathbf{M} \; .$$
- Es ist $\mathbf{w}\mathbf{P} = \mathbf{w}$, also
$$\mathbf{w}\mathbf{M} + \mathbf{w}\mathbf{D} = \mathbf{w}\mathbf{E} + \mathbf{w}\mathbf{M}$$
- und folglich $\mathbf{w}\mathbf{D} = \mathbf{w}\mathbf{E}$.
- Das bedeutet aber ausgeschrieben nichts anderes als
$$(w_1 m_{11}, w_2 m_{22}, \ldots, w_n m_{nn}) = (1, 1, \ldots, 1)$$