Марковские цепи и спектральная теория

А. Родоманов

Факультет ВМК
МГУ им. М. В. Ломоносова

3 октября 2014 г.

Спецсеминар "Байесовские методы машинного обучения"

Декодирование шифра

- Рассмотрим подстановочный шифр:
  АБВГДЕЁЖЗИКЛМНОПРСТУФХЦЧШЩЪЭЮЯ
  БАВГДЕЁЖЗИКЛРНСПОМТУФХЦЧШЩЪЭЮЯ
  МОСКВА $\mapsto$ РСМКВБ
- Как по зашифрованному тексту восстановить исходный?

```
!ЧЕ1У9П\ЧНЧР1ЖU9ЦЧMRЖЧ4Р9ЖР/ТЧЩ/1VЧ!Ч4ЕК
РО;'РЖЧЯ1ЖРЦЧ/К9'РЮJ0Ч'1ЕК14ЧMRЖЧЦ9ЧЦ1'0
Ч41Ч'К9П9Ц;Ч4'19Я1ЧЕ1Н'Ж9Ц;НЧЦРЧЯ99Ч1ЦЧУ
Р4/1Ч4ЕКРО;'РЖЧ49МНЧ1МЧЩ/1П0Ч1ЦЧП1ЯЧЕК1М
\Р;/Т4НЧ1/ЧИК9ЕИ1Я1Ч4ЦРЧ'ЧЕК;Н/Ц1ЮЧЕК1СЖ
РР9ЧЦ1У;Ч;Ч1МЦРК\U;/ТЧ'ЧЯ1Ж1'9ЧМ9JП1Ж'Ц1
Ч4/\УРL;ЮХЧЕ1Р1МЦ1ЧИК1С1/Ц1П\ЧМРКРМРЦ\ХЧ
'1ЕК14IЧ<Е1У9П\ЧНЧР1
```

Вероятностный подход

- Зададим вероятностное распределение на множестве всех перестановок:

$$\pi(\sigma) = \frac{1}{Z} \prod_{i=1}^{|T|-1} K(\sigma(T_i), \sigma(T_{i+1})),$$

  где
  - $K(a, b) =$ частота встречаемости пары символов (a, b)
  - Z —нормировочная константа

  Матрицу $K(a, b)$ можно построить с помощью любого большого текста (например, «Война и мир»).
- Тогда «правильная» перестановка

$$\sigma^* = \mathrm{argmax}_\sigma\, \pi(\sigma)$$

Связь с марковскими цепями

- Число всевозможных перестановок $= n!$, где n —размер алфавита (87 символов в примере). Обычный перебор не сработает.
- Заметим, что распределение $\pi(\sigma)$, скорее всего, сконцентрировано в одной точке (для не слишком маленького текста).
- Поэтому вместо поиска максимума $\pi(\sigma)$ будем генерировать образцы из этого распределения. Они должны быть очень близки к «правильной» перестановке.
- Две проблемы:
  - Число состояний огромное
  - Нормировочная константа Z неизвестна
- Нам поможет марковская цепь!

Марковские цепи

- Будем рассматривать марковские цепи, имеющие конечное число состояний (но, может быть, очень большое).
- Обозначения:
  - $\mathcal{X}$ —множество всевозможных состояний
  - $x_0 \in \mathcal{X}$ —начальное состояние
  - $P(x, y) \geq 0$ —вероятность перехода из x в y, $\sum_{y \in \mathcal{X}} P(x, y) = 1$
- Тогда марковская цепь —это последовательность случайных величин $X_0 = x_0, X_1, X_2, \ldots$, таких что
  - Мы начинаем из состояния $X_0 = x_0$
  - Затем на каждой итерации переходим из состояния $X_i = x$ в состояние $X_{i+1} = y$ с вероятностью $P(x, y)$
- Из определения

$$\mathbf{P}(X_2 = z \mid X_0 = x_0) = \sum_{y \in \mathcal{X}} P(x_0, y) P(y, z) = P^2(x_0, z)$$

Аналогично $\mathbf{P}(X_t = y \mid X_0 = x_0) = P^t(x_0, y)$

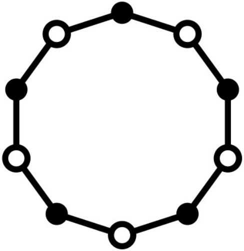
Неупрощаемость и непериодичность

- Каждой марковской цепи с матрицей переходов $P(x, y)$ можно поставить в соответствие граф $\mathcal{G}_P$, в котором
  - вершины соответствуют состояниям из $\mathcal{X}$
  - вершины x и y соединены ребром $\Leftrightarrow P(x, y) > 0$
- Марковская цепь называется неупрощаемой, если граф $\mathcal{G}_P$ сильно связный (т. е. из любой вершины можно добраться в любую другую).
- Рассмотрим множество моментов времени, через которое можно вернуться в начальное состояние x:

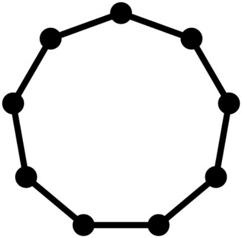
  $$\mathcal{T}(x) = \{t \geq 1 \mid P^t(x, x) > 0\}$$

  Периодом состояния x называется $\gcd \mathcal{T}(x)$.
- Если цепь неупрощаемая, то периоды всех состояний совпадают [2, лемма 1.6], т. е. можно говорить о периоде цепи.
- Неупрощаемая марковская цепь называется непериодичной, если ее период равен 1.

Пример: случайное блуждание по циклу



$\mathbf{Z}_{10}$



$\mathbf{Z}_9$

- Начальное состояние: $x_0 = 0$
- Матрица переходов:

$$P(x,y) = \begin{cases} 1/2, & y \equiv x+1 \pmod n \\ 1/2, & y \equiv x-1 \pmod n \\ 0, & \text{иначе} \end{cases}$$

- Обе цепи неупрощаемые
- Первая цепь периодичная (период 2), вторая непериодичная

Сходимость марковской цепи

- Напоминание: $\mathbf{P}(X_t = y \mid X_0 = x_0) = P^t(x_0, y)$, т. е. случайная величина, генерируемая цепью на t-ом шаге, из распределения $P^t(x_0, \cdot)$
- Хотим, что цепь генерировала величины из некоторого заранее заданного распределения π, т. е. хотим, чтобы она сходилась:
$$P^t(x_0, \cdot) \xrightarrow[t \to \infty]{} \pi$$
- Периодические цепи могут не сходиться! Например, в $\mathbf{Z}_{10}$:
  - $P^{2t+1}(x_0, \cdot)$ сосредоточено на нечетных состояниях
  - $P^{2t}(x_0, \cdot)$ сосредоточено на четных состояниях

Ленивая версия марковской цепи

- Каждую периодичную цепь можно сделать непериодичной, добавив к каждому состоянию петлю:

$$Q(x,y) = \begin{cases} 1/2P(x,y), & x \neq y \\ 1/2, & x = y \end{cases}$$

- Новая цепь действует следующим образом:
  - В состоянии x побрасывается монетка
  - Если выпал орел, то переходим в y согласно $P(x,y)$
  - Если выпала решка, то остаемся на месте
- В матричном виде:

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

- Новая цепь Q называется ленивой версией P

Стационарное распределение

- Распределение π называется стационарным, если
$$\sum_{x \in \mathcal{X}} \pi(x) P(x, y) = \pi(y)$$
В матричном виде: $\pi P = \pi$.
Таким образом, π —это левый собственный вектор P, отвечающий собственному значению 1.
- Фундаментальная теорема марковских цепей [2, теор. 4.9]
Любая неупрощаемая и непериодическая марковская цепь имеет единственное стационарное распределение π, причем
$$P^t(x_0, \cdot) \xrightarrow[t \to \infty]{} \pi$$
- Заметим, что ленивая версия цепи $Q = 1/2(P + I)$ не меняет стационарное распределение

Уравнения детального баланса и обратимость

- Как построить неупрощаемую марковскую цепь, которая будет иметь π в качестве стационарного распределения?
- Для стационарности π достаточно, чтобы матрица P удовлетворяла т. н. уравнениям детального баланса:

  $$\pi(x)P(x,y) = \pi(y)P(y,x)$$

  Действительно:

  $$\sum_x \pi(x)P(x,y) = \sum_x \pi(y)P(y,x) = \pi(y)\sum_x P(y,x) = \pi(y)$$

- Цепь, удовлетворяющая уравнениям детального баланса, называется обратимой

Алгоритм Метрополиса–Хастингса

- Пусть уже имеется некоторая марковская цепь с матрицей переходов J (но ее стационарное распределение, может быть, не π)
- Модифицируем цепь следующим образом:

$$P(x,y) = \begin{cases} J(x,y)a(x,y), & x \neq y, \\ 1 - \sum_{y:y\neq x} J(x,y)a(x,y), & x = y, \end{cases}$$

  где $a(x,y) = \min\{1, \pi(y)J(y,x)/\pi(x)J(x,y)\}$
- Новая цепь работает по следующему принципу:
  - В состоянии x выбирается y согласно $J(x,y)$
  - Далее побрасываем монетку с вероятностью орла $a(x,y)$
  - Если выпал орел, то переходим в состояние y
  - Если выпала решка, то остаемся в состоянии x
- Такая модификация матрицы переходов превращает цепь в обратимую

Строим марковскую цепь для задачи декодирования

- Хотим генерировать перестановки из распределения

$$\pi(\sigma) = \frac{1}{Z} \prod_{i=1}^{|T|-1} K(\sigma(T_i), \sigma(T_{i+1})),$$

- Построим марковскую цепь с помощью алгоритма Метрополиса–Хастингса
- Для этого нам нужна вспомогательная цепь J, которая будет случайно блуждать по всем перестановкам
- Будем использовать случайные транспозиции:

$$J(\sigma, \phi) = \begin{cases} \frac{1}{\binom{n}{2}}, & \sigma \text{ и } \phi \text{ отличаются на транспозицию,} \\ 0, & \text{иначе} \end{cases}$$

- Заметим, что $J(\sigma, \phi) = J(\phi, \sigma)$, поэтому $a(\sigma, \phi) = \frac{\pi(\phi)}{\pi(\sigma)}$. Неизвестная нормировочная константа Z сокращается!

Итоговый алгоритм

- Выбираем начальную перестановку σ (например, $\sigma = e$)
- Далее по циклу:
  - Меняем σ на ϕ с помощью случайной транспозиции
  - Вычисляем $\tilde{\pi}(\phi)$
  - Если $\tilde{\pi}(\phi) > \tilde{\pi}(\sigma)$, то переходим в ϕ
  - Иначе побрасываем $\tilde{\pi}(\phi)/\tilde{\pi}(\sigma)$ монетку
    - Если орел, то переходим в ϕ
    - Если решка, то остаемся в σ

Заметим, что полученная цепь является неупрощаемой и непериодичной. Значит, для нее справедлива фундаментальная теорема о сходимости.

Результаты работы

Основание, Айзек Азимов:

```
[00000] ЙМЙРМJРОЗАJ ИОВМЙОМ:R -МЗРХМЙАМ? АОМ 2J08 ОМЙРМЖ ВХЙОМ2ҐВОМЖОХРМ? ЪЙНRКЪХМЪ 8ЙОЙАХМRJРИОЫ8ОСМАМЖВХМЖ ООМВН5,АОМЪ
[01001] АИАСИЬСЕЗ?Ь ЖЕЮИАЕИ.Л ЩИЗСЧИА?ИШ ?ОИ РЬЕ" ОИАСИД Ю:АЕИРБЮЕИДЕ:СИШ ТАНЛКТЧИТ "АЕА?ЧИЛЬСЖЕВ"ЕРИ?ИДЮЧИД ОЕИЮНХУ?ОИТ
[10010] ИВ ВЕ ШЕОРАШИТОЧ ВО ЦЛИ? РЕК ВА ХИАМ ИЗШОДИМ ВЕ ПИЧ"ВО ЗЬЧО ПО"Е ХИСВУЛНСК СИДВОВАК ЛШЕТОЙДО. А ПЧК ПИМО ЧУБЖАМ
[20317] ОН НЕ ЛЕАҐИЛОТАС НА ЭРО. ҐЕЯ НИ ВОИМ ОЗЛАЧОМ НЕ ШОСКНА ЗЬСА ШАКЕ ВОДНУРЫДЯ ДОЧНАНИЯ РЛЕТАЙЧА, И ШСЯ ШОМА СУБЦИМ
[30975] ОН НЕ РЕАҐИРОВАЛ НА ЭТО. ҐЕЯ НИ ХОИМ ОБРАКОМ НЕ ДОЛЖНА БЬЛА ДАЖЕ ХОСНУТЫСЯ СОКНАНИЯ ТРЕВАЙКА, И ДЛЯ ДОМА ЛУЧЦИМ
[43090] ОН НЕ РЕАҐИРОВАЛ НА ЭТО. ҐЕЯ НИ КОИМ ОБРАЗОМ НЕ ДОЛЖНА БЫЛА ДАЖЕ КОСНУТЬСЯ СОЗНАНИЯ ТРЕВАЮЗА, И ДЛЯ ДОМА ЛУЧЦИМ
[50596] ОН НЕ РЕАҐИРОВАЛ НА ЭТО. ҐЕЯ НИ КОИМ ОБРАЗОМ НЕ ДОЛЖНА БЫЛА ДАЖЕ КОСНУТЬСЯ СОЗНАНИЯ ТРЕВА?ЗА, И ДЛЯ ДОМА ЛУЧШИМ
[53248] ОН НЕ РЕАҐИРОВАЛ НА ЭТО. ҐЕЯ НИ КОИМ ОБРАЗОМ НЕ ДОЛЖНА БЫЛА ДАЖЕ КОСНУТЬСЯ СОЗНАНИЯ ТРЕВАЙЗА, И ДЛЯ ДОМА ЛУЧШИМ
```

451 градус по Фаренгейту, Рэй Брэдбери:

```
[00000] ДВ^Д5ШВҐ^ДВUД0В'ДВ*ЪLҐФ*ВЫРДМ=ДL^MGВД'В'ФВМ5ФЪ^ВVД*Ы0Ф5Ъ'М^Ъ0ЬВ<ВМВДСД'ЬВВФU'ДВ)РД*ФЪ^*МВ'ФВUД5ШВДКРФ2МVФМВVЪҐ
[01000] АМЛАЫЯМЮЛАМРАВФ АМО"ҐЮЕОМПУА?БАҐЛ?:МА М ЕФ?ЫЕ"ЛМДАОПВЕЫ" ?Л"ВСМШМ?МАКА СМФЕР АМТУАОЕ"ЛОХМ ЕМРАЫЯМАНУЕИ?ДЕХМД"Ю"
[05005] А ДАЗ! БДА МАТХЛА ВЕЮБОВ ЦНАИПАЮДИ) АЛ ЛОХИЗОЕД РАВЦТОЗЕЛИДЕТ. Х И АҐАЛ. ХОМЛА "НАВОЕДВ? ЛО МАЗ! АКНОШИРО? РЕБЕНЛЕЕ
[10136] О ДОХ. "ДО МОЛЧНО СЕЙ"АС БРОИЖОЙДИ; ОН НАЧИХАЕД ТОСБЛАХЕНИДЕЛУ N И ОҐОНУ ЧАМНО ЗРОСАЕДСЯ НА МОХ. ОКРАШИТАЯ ТЕ"ЕРНЕЕ
[15039] О ТОВ. ШТО ДОЛЖНО СЕЙШАС БРОИХОЙТИ: ОН НАЖИВАЕТ МОСБЛАВЕНИТЕЛЬ З И ОҐОНЬ ЖАДНО ПРОСАЕТСЯ НА ДОВ. ОКРАФИМАЯ МЕШЕРНЕЕ
[20489] О ТОВ, ЧТО ДОЛЖНО СЕЙЧАС ПРОИХОЙТИ: ОН НАЖИВАЕТ МОСПЛАВЕНИТЕЛЬ - И ОҐОНЬ ЖАДНО БРОСАЕТСЯ НА ДОВ, ОКРАШИМАЯ МЕЧЕРНЕЕ
[24030] О ТОМ, ЧТО ДОЛЖНО СЕЙЧАС ПРОИЗОЙТИ: ОН НАЖИМАЕТ ВОСПЛАМЕНИТЕЛЬ - И ОҐОНЬ ЖАДНО БРОСАЕТСЯ НА ДОМ, ОКРАШИВАЯ ВЕЧЕРНЕЕ
```

Психология, Википедия

```
[000000] НЕЗ0ОЕ^-^ВФ\ФЁОЪФ^ЦЯЦ^-ФОВ^ШВОЮЕЗЕGО4^ВНВЗВ^ВВFЕВНЕ4НФЗ0ЪЕЕ^ЪВ(ОЕЕК^ЪЕ^ЕНЁФЗО-КОВ0^ЕН^ТОЗЕВЕТООJ^ЕЪВ^Ш\ЕЁЕЗДВ
[010139] ТИМЫШИ Д СОКОЖЕНО ОRО ДОША ЗСЕҐИМИЧЕЯ СТАМА САРИСТИЯТОМЫНИЙ НАЛШИЙХ НИ ИТЖОМЕД?ЕСЫ ИТ ЩЕМИСИЩЕЕ; ИНА ЗКИЖИМ"АОТ
[020060] ТОМЫЛО Д СИКИЧЕНИ ОМО ДИЛА БСЕҐОМОРЕЯ СТАМА САВОСТОЯТИМЫНОЙ НАУЛОЙ. НО ОТЧИМЕД"ЕСЫ ОТ ПЕМОСОПЕЕ, ОНА БКОЧОМФАИТ
[030096] ТОЛЫКО Д СЕРЕЧИНЕ ТЕТ ДЕКА ВСИХОЛОҐИЯ СТАЛА САМОСТОЯТЕЛЫНОЙ НАУКОЙ. НО ОТЧЕЛИДЦИСЫ ОТ ФИЛОСОФИИ, ОНА ВРОЧОЛЖАЕТ
[050160] ТОЛЬКО В СЕРЕЧИНЕ ОСО ВЕКА ДСИХОЛОҐИЯ СТАЛА САМОСТОЯТЕЛЬНОЙ НАУКОЙ. НО ОТЧЕЛИВПИСЬ ОТ ФИЛОСОФИИ, ОНА ДРОЧОЛЖАЕТ
[125506] ТОЛЬКО В СЕРЕЧИНЕ IVI ВЕКА ДСИХОЛОҐИЯ СТАЛА САМОСТОЯТЕЛЬНОЙ НАУКОЙ. НО ОТЧЕЛИВПИСЬ ОТ ФИЛОСОФИИ, ОНА ДРОЧОЛЖАЕТ
[141874] ТОЛЬКО В СЕРЕЧИНЕ VIV ВЕКА ДСИХОЛОҐИЯ СТАЛА САМОСТОЯТЕЛЬНОЙ НАУКОЙ. НО ОТЧЕЛИВПИСЬ ОТ ФИЛОСОФИИ, ОНА ДРОЧОЛЖАЕТ
[158439] ТОЛЬКО В СЕРЕЧИНЕ IVI ВЕКА ДСИХОЛОҐИЯ СТАЛА САМОСТОЯТЕЛЬНОЙ НАУКОЙ. НО ОТЧЕЛИВПИСЬ ОТ ФИЛОСОФИИ, ОНА ДРОЧОЛЖАЕТ
```

Скорость сходимости

- Основная задача теория марковских цепей состоит в оценке скорости сходимости $P^t(x_0, \cdot) \to \pi$
- Для этого на вероятностных распределениях над множеством $\mathcal{X}$ вводится некоторая мера близости $\rho(\cdot, \cdot)$ (конкретная функция рассматривается далее)
- Чтобы получить скорость сходимости для произвольного начального состояния, обычно оценивают величину

$$d(t) = \max_{x \in \mathcal{X}} \rho(P^t(x, \cdot), \pi)$$

- Временем сходимости цепи называется число

$$t_{\mathrm{mix}}(\varepsilon) = \min\{t : d(t) \le \varepsilon\}$$

Расстояние по вариации

- В качестве меры близости $\rho(\cdot, \cdot)$ обычно используют расстояние по вариации (англ. total variation distance):

$$\rho(\mu, \pi) = \|\mu - \pi\|_{\mathrm{TV}} = \max_{A \subset \mathcal{X}} |\mu(A) - \pi(A)|$$

- Можно показать, что

$$\|\mu - \pi\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \pi(x)|$$

Спектральный радиус матрицы переходов

- Все собственные значения матрицы переходов P по модулю не превосходят единицы: если ψ —собственный вектор для λ, то

$$|\lambda| \, \|\psi\|_\infty = \|\lambda\psi\|_\infty = \|P\psi\|_\infty \le \|P\|_\infty \, \|\psi\|_\infty = \|\psi\|_\infty \,,$$

  т.е.

$$|\lambda| \le 1$$

- Вектор-столбец из всех единиц $\mathbf{1}$ является собственным вектором, отвечающим собственному значению $\lambda = 1$:

$$P\mathbf{1} = \mathbf{1}$$

- Таким образом, спектральный радиус матрицы переходов равен одному:

$$\rho(P) := \max_i |\lambda_i| = 1$$

Собственные значения и векторы матрицы переходов

- Матрица P, вообще говоря, несимметричная
- Введем матрицу $\Pi := \mathrm{diag}(\pi(x))_{x\in\mathcal{X}}$
- Для обратимых цепей матрица $\Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}}$ симметричная, поэтому для нее справедливо спектральное разложение:

$$\Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}} = Q \Lambda Q^\top$$

  Значит

$$P = \Psi \Lambda \Psi^\top \Pi,$$

  где $\Psi := \Pi^{-\frac{1}{2}} Q$
- Заметим, что $\Psi^\top \Pi \Psi = \Psi \Psi^\top \Pi = I$.
- Столбцы Ψ являются правыми собственными векторами для P, отвечающими собственным значениям, Λ.
  Кроме того, из предыдущего пункта следует, что они образуют ортонормированную систему относительно скалярного произведения $\langle f, g \rangle_\pi := f^\top \Pi g$

Теорема Перрона–Фробениуса

Пусть неотрицательная матрица A является неупрощаемой и имеет спектральный радиус $\rho(A) = r$. Тогда верны следующие утверждения:

- Число r вещественное и положительное и является собственным значением матрицы P; оно называется собственным значением Перрона–Фробениуса
- Собственное значение Перрона–Фробениуса является простым: оба пространства левых и правых собственных векторов A являются одномерными
- Все другие собственные значения A строго меньше, чем r: $|\lambda| < r$
- Матрица A имеет левый и правый собственные векторы, соответствующие собственному значению r; все компоненты этих векторов положительные.

Разложение по базису

- Согласно собственному разложению:

$$P^t = \Psi \Lambda^t \Psi^\top \Pi$$

  В поэлементной записи:

$$P^t(x, y) = \pi(y) \sum_{i=1}^{|\mathcal{X}|} \lambda_i^t \psi_i(x) \psi_i(y)$$

- По теореме Перрона–Фробениуса
  - $\lambda_1 = 1 > \lambda_2 \geq \ldots \lambda_{|\mathcal{X}|} > -1$
  - пространство собственных векторов для $\lambda_1 = 1$ одномерное

  Отсюда следует, что в спектральном разложении можно взять $\psi_1 \equiv \mathbf{1}$
- В итоге

$$P^t(x, y) = \pi(y) \left(1 + \sum_{i=2}^{|\mathcal{X}|} \lambda_i^t \psi_i(x) \psi_i(y) \right)$$

Верхняя оценка

- Оценим $d(t) := \max_{x \in \mathcal{X}} \left\| P^t(x, \cdot) - \pi \right\|_{\mathrm{TV}}$:

$$4 \left\| P^t(x, \cdot) - \pi \right\|_{\mathrm{TV}}^2 = \left(\sum_{y \in \mathcal{X}} |P^t(x, y) - \pi(y)| \frac{\pi(y)^{\frac{1}{2}}}{\pi(y)^{\frac{1}{2}}} \right)^2$$

$$\leq \left(\sum_{y \in \mathcal{X}} \frac{(P^t(x, y) - \pi(y))^2}{\pi(y)} \right) \left(\sum_{y \in \mathcal{X}} \pi(y) \right)$$

$$= \sum_{y \in \mathcal{X}} \pi(y) \left(\sum_{i=2}^{|\mathcal{X}|} \lambda_i^t \psi_i(x) \psi_i(y) \right)^2$$

$$= \sum_{i=2}^{|\mathcal{X}|} \lambda_i^{2t} \psi_i^2(x)$$

Верхняя оценка 2

- Мы получили:

$$4\left\|P^t(x,\cdot)-\pi\right\|_{\mathrm{TV}}^2 \le \sum_{i=2}^{|\mathcal{X}|} \lambda_i^{2t}\psi_i^2(x)$$

- Введем $\lambda_* := \max_{i=2,\dots,|\mathcal{X}|} |\lambda_i|$
- Из ортонормированности столбцов Ψ ($\Psi\Psi^\top = \Pi^{-1}$) следует, что

$$\sum_{i=2}^{|\mathcal{X}|} \psi_i^2(x) \le \pi(x)^{-1}$$

- Введем $\pi_{\min} := \min_x \pi(x) > 0$. Тогда

$$4\left\|P^t(x,\cdot)-\pi\right\|_{\mathrm{TV}}^2 \le \frac{\lambda_*^{2t}}{\pi_{\min}}$$

Верхняя оценка: итог

- Мы получили:

$$4\left\|P^t(x,\cdot)-\pi\right\|_{\mathrm{TV}}^2 \le \frac{\lambda_*^{2t}}{\pi_{\min}}$$

- Если ввести т. н. абсолютный спектральный зазор $\gamma_* := 1-\lambda_* > 0$, то

$$d(t) \le \frac{\lambda_*^t}{2\pi_{\min}^{1/2}} = \frac{(1-\gamma_*)^t}{2\pi_{\min}^{1/2}} \le \frac{e^{-\gamma_* t}}{2\pi_{\min}^{1/2}}$$

(Здесь мы воспользовались оценкой $e^x \ge 1+x$.)

- В итоге мы получили верхнюю оценку на время сходимости:

$$t_{\mathrm{mix}}(\varepsilon) \le \log\left(\frac{1}{2\pi_{\min}^{1/2}\varepsilon}\right) t_{\mathrm{rel}},$$

где $t_{\mathrm{rel}} := 1/\gamma^*$ —т. н. время релаксации

Нижняя оценка

- Пусть ψ —собственный вектор матрицы P^t, отвечающий собственному значению $\lambda^t < 1$.
- Поскольку ψ ортогонален $\mathbf{1}$ относительно $\langle \cdot, \cdot \rangle_\pi$, то $\sum_{y \in \mathcal{X}} \pi(y)\psi(y) = 0$ и для любого $x \in \mathcal{X}$

$$
\begin{aligned}
|\lambda^t \psi(x)| = |P^t \psi(x)| &= \left| \sum_{y \in \mathcal{X}} P^t(x,y)\psi(y) \right| \\
&= \left| \sum_{y \in \mathcal{X}} [P^t(x,y)\psi(y) - \pi(y)\psi(y)] \right| \\
&\le \|\psi\|_\infty 2 \left\| P^t(x,\cdot) - \pi \right\|_{\mathrm{TV}} = \|\psi\|_\infty 2d(t)
\end{aligned}
$$

- Взяв x, для которого $|\psi(x)| = \|\psi\|_\infty$, получаем

$$|\lambda|^t \le 2d(t)$$

Нижняя оценка 2

- Итак, $|\lambda|^t \leq 2d(t)$. Взяв $t = t_{\text{mix}}(\varepsilon)$, получаем

$$|\lambda|^{t_{\text{mix}}(\varepsilon)} \leq 2\varepsilon$$

- После взятия логарифма:

$$t_{\text{mix}}(\varepsilon) \log\left(\frac{1}{|\lambda|}\right) \geq \log\left(\frac{1}{2\varepsilon}\right)$$

- Воспользуемся неравенством $x - 1 \geq \log x$ при $x > 0$:

$$t_{\text{mix}}(\varepsilon)\left(\frac{1}{|\lambda|} - 1\right) \geq \log\left(\frac{1}{2\varepsilon}\right)$$

- Подставляя $\lambda = \lambda_*$, в итоге получаем

$$t_{\text{mix}}(\varepsilon) \geq (t_{\text{rel}} - 1)\log\left(\frac{1}{2\varepsilon}\right)$$

Результат

Мы получили, что для неупрощаемой, непериодичной и обратимой марковской цепи справедливы следующие оценки на время сходимости:

$$(t_{\mathrm{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right) \le t_{\mathrm{mix}}(\varepsilon) \le \log\left(\frac{1}{2\pi_{\mathrm{min}}^{1/2}\varepsilon}\right) t_{\mathrm{rel}},$$

где

- $t_{\mathrm{rel}} = 1/\gamma_* = 1/(1 - \lambda_*)$
- $\lambda_* =$ второе наибольшее по модулю с.з. матрицы P
- $\pi_{\mathrm{min}} = \min_x \pi(x)$

Вывод: чем меньше λ_*, тем скорость сходимости выше.

Ленивая версия цепи

- Собственные значения матрицы P:

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\mathcal{X}|} > -1$$

- Второе наибольшее по модулю с.з.:

$$\lambda_* = \max\{|\lambda_2|, |\lambda_{|\mathcal{X}|}|\}$$

- Если $|\lambda_2| < |\lambda_{|\mathcal{X}|}|$, то $\lambda_* = |\lambda_{|\mathcal{X}|}|$.
- Рассмотрим ленивую версию цепи $Q = (P + I)/2$. Тогда у Q все собственные значения станут положительные и

$$\lambda_*(Q) = (\lambda_2 + 1)/2$$

- Если λ_2 сильно дальше от 1, чем $\lambda_{|\mathcal{X}|}$ от -1, то у ленивой версии цепи скорость сходимости будет больше!

Ссылки

Stephen Boyd, Persi Diaconis, and Lin Xiao.
Fastest Mixing Markov Chain on a Graph.
SIAM Review, 46(4):667–689, January 2004.

DA Levin, Y Peres, and EL Wilmer.
Markov chains and mixing times.
2009.