Конспект по Дискретной математике.

Чепелин В.А.

Содержание

1 Лекция 1.

1.1 Аксиоматическое вероятное пространство.

Пусть у нас есть Ω - **элементарные исходы** и связанная с ним функция $p : \Omega \to [0, 1]$ - **дискретная вероятностная мера (плотность вероятности)** - функция, которая по элементарному исходу возвращает **вероятность**.

А также $\sum\limits_{w\in\Omega} p(w) = 1$, а также $0 \le p_i \le 1$ А также мы считаем, что $|\Omega|$ не более чем счетно. Для множеств мощности континуума нам нужна более сложная теория.

Рассмотрим **примеры**:

1. Честная монета:

   $\Omega = \{0, 1\}$. $p(0) = p(1) = \frac{1}{2}$.
2. Нечестная монета или распределение Бернулли:

   $\Omega = \{0, 1\}$. $p(0) = 1 - p(1) = q$.
3. Честная игральная кость:

   $\Omega = \{1, 2, 3, 4, 5, 6\}$. $p(w) = \frac{1}{6}$. $p(w) = \dfrac{1}{52}$
4. Колода карт:

   $\Omega = \{\langle c, r\rangle\, 1 \le c \le 4, 1 \le r \le 15\}$
5. Геометрическое распределение:

   $\Omega = \mathbb{N}$, $p(i) = \dfrac{1}{2^i}$

Замечание. Не существует равномерного распределения на счетном множестве.

Событие — множество $A \subset \Omega$. $P(A) = \sum\limits_{w\in A} p(w)$. (Иногда используют Pr).

$P(A) = 1$ — достоверное событие.

$P(A) = 0$ — невозможное событие.

Рассмотрим примеры на честной игральной кости:

1. Только четные: $P(A) = \frac{3}{6} = \frac{1}{2}$.
2. Больше 4-ех: $P(A) = \frac{2}{6} = \frac{1}{3}$.

Замечание: нельзя с равной вероятностью выбрать случайное целое число.

Независимые события — A,B независимы, если $P(A \cap B) = P(A) \cdot P(B)$.

$\dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)}{P(\Omega)}$ — независимы (если выполнилось B, то вероятность не поменялась)

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ — вероятность A при условии B — **условная вероятность**.

Произведение вероятностных пространств.

Пусть у нас есть $\Omega_1.p_1$, а также Ω_2, p_2, тогда произведение вероятностных пространств:

$$\Omega = \Omega_1 \times \Omega_2$$
$$p(\langle w1, w2 \rangle) = p_1(w_1) \cdot p_2(w_2)$$

Утв. $\forall A \subset \Omega_1, B \subset \Omega_2$.

$A \times \Omega_2$ и $\Omega_1 \times B$ независимы.

Пусть у нас есть n событий: $A_1, A_2, \ldots, A_n$.

Тогда обычно **независимость n событий** подразумевает:

1. A_i, A_j - независимы $\forall i, j, \quad i \neq j$
2. $\forall I \subset \{1, 2, 3, \ldots, n\}$: $P(\bigcap\limits_{i \in I} A_i) = \prod\limits_{i \in I} P(A_i)$

Формула полной вероятности

$\Omega = A_1 \cup A_2 \cup \ldots \cup A_n, \forall i \neq j : A_i \cap A_j = \varnothing$ — **полная система событий**.

Возьму B - какое-то событие.

$P(B) = \sum\limits_{i=1}^{n} P(B \cap A_i) = \sum\limits_{i=1}^{n} P(B|A_i) \cdot P(A_i)$

Пример: урна с шариками. Сначала выбираете урну, потом достаете шарик.

Формула Байеса.

$P(A_i|B) = \dfrac{P(A_i \cap B)}{P(B)} = \dfrac{P(B|A_i) \cdot P(A_i)}{\sum\limits_{j=1}^{n} P(B|A_j) \cdot P(A_j)}$

2 Лекция 2.

2.1 Случайная величина.

Случайная величина или численная характеристика каждого элементарного исхода — это отображение $\xi : \Omega \to \mathbb{R}$, которое сопоставляет каждому элементарному исходу какое-то число.
Пример:

1. $D = \{1, 2, \dots, 6\}$. Возьмем $\Omega = D^2$. Например, человек бросает два игральных кубика. Тогда, очевидно, $p(\langle i, j\rangle) = \dfrac{1}{36}$. И тогда он задает функцию случайной величины, например, как $\xi(\langle i, j\rangle) = i + j$.
2. Возьмем случайный граф G на n вершинах. $\xi(G) =$ количеству компонент связности. Или $\xi(G) =$ количеству ребер в этом графе.
3. Давайте кидать игральный кубик и сопоставим каждой выпадающей грани число, равное количеству точек на этой грани.
   То есть $\Omega = \{1, 2, \dots, 6\}$, $\xi(i) = i$.
4. $\Omega = \{1, 2, \dots, 6\}; E = \{2, 4, 6\}$. $x_E(w) = \begin{cases} 1, w \in E \\ 0, w \notin E \end{cases}$

Возьмем какие-то Ω, p, ξ:

$[\xi = i] = \{w | \xi(w) = i\} \subset \Omega$ — **множество элементарных исходов**, случайная величина которых равна i.

def: $f_\xi : \mathbb{R} \to \mathbb{R}$ — **дискретная плотность вероятности ξ**.

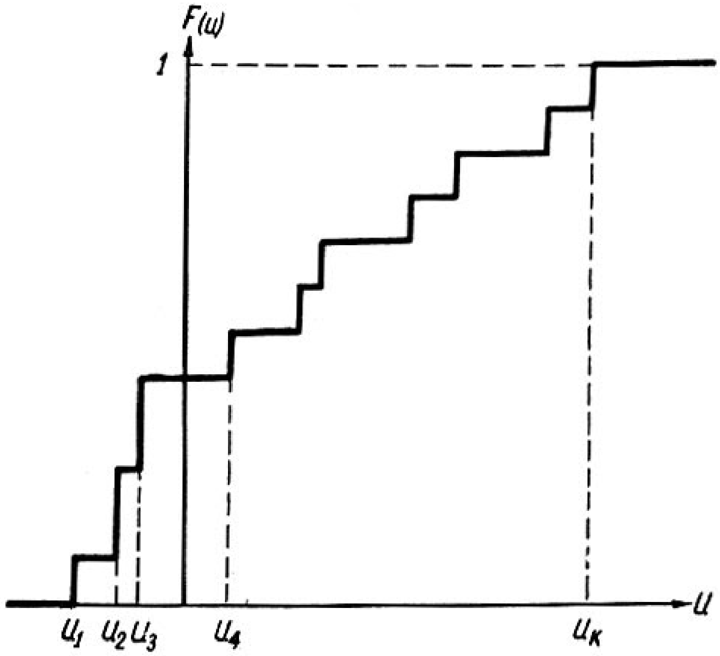
$$P([\xi = i]) = P(\xi = i) = f_\xi(i) = \sum_{w \in [\xi = i]} p(w)$$

Дискретная плотность вероятности — это функция, которая говорит нам, насколько вероятно каждое из этих отдельных значений, которые может принимать случайная величина. Другими словами, она присваивает вероятность каждому возможному исходу.

Немного поменяем и получим $[\xi \le i] = \{w | \xi(w) \le i\} \subset \Omega$.

$$P([\xi \le i]) = P(\xi \le i) = F_\xi(i)$$

def: $F_\xi : \mathbb{R} \to \mathbb{R}$ — **функция распределения**. У дискретной случайной величины функция распределения ступенчатая. Например:



2.2 Мат. ожидание.

Математическое ожидание — среднее значение случайной величины.

$E_\xi = \sum\limits_w p(w)\xi(w) = \sum\limits_i i \cdot P(\xi = i)$.

Дальше А.С. использует 3 вида обозначений:

1. E_ξ 2. $E(\xi)$ 3. $E\xi$ — не боимся, это одно и то же.

Теорема (линейность мат ожидания)

$$E\lambda\xi = \lambda E_\xi \quad E_{(\xi+\eta)} = E_\xi + E_\eta$$

Доказательство:

$$E\lambda\xi = \sum_w p(w) \cdot \lambda\xi(w) = \lambda \sum_w p(w)\xi(w) = \lambda E_\xi$$

$$E(\xi + \eta) = \sum_w p(w)(\xi(w) + \eta(w)) = \sum_w p(w)\xi(w) + \sum_w p(w)\eta(w) = E(\xi) + E(\eta)$$

Q.E.D.

МАТ. ОЖИДАНИЕ ВСЕГДА ЛИНЕЙНО!!!

2.3 Незав. случайные величины

ξ, η - **независимы**, если $[\xi = a], [\eta = b]$ — независимы $\forall a, b$.

Эквивалентное утверждение — $[\xi \leq a], [\eta \leq b]$ — независимы $\forall a, b$.

Иначе говоря, две случайные величины называются *независимыми*, если по значению одной нельзя сделать выводы о значении другой.

Теорема (о мультипликативности мат. ожидания)

ξ, η — независимы $\Rightarrow E(\xi \cdot \eta) = E_\xi \cdot E_\eta$.

Доказательство:

$$E_{(\xi\cdot\eta)} = \sum_a aP(\xi,\eta = a) = \sum_a a \sum_{\forall i,j:\, i\cdot j = a\, i\in R_\xi, j\in R_\eta} P(\xi = i, \eta = j) =$$

$$= \sum_a \sum_i \sum_j aP(\xi = i)P(\eta = j) = \sum_i iP(\xi = i) \cdot \sum_j jP(\eta = j) = E_\eta \cdot E_\xi$$

Q.E.D.

2.4 Дисперсия случайной величины.

$D_\xi = Var(\xi)$ — **дисперсия** случайной величины.

$$D_\xi = E((\xi - E_\xi)^2) = E_{\xi^2} - (E_\xi)^2$$

Дисперсия случайной величины — это мера того, насколько сильно разбросаны значения этой случайной величины вокруг её математического ожидания (среднего значения). Другими словами, она показывает, насколько "широко"распределение вероятностей случайной величины.

Теорема (свойства дисперсии). Если ξ, η - независимы:

$$D_{c\eta} = c^2 D_\eta \quad D_{\xi+\eta} = D_\xi + D_\eta$$

Доказательство тривиально из линейности мат. ожидания.

3 Лекция 3.

3.1 Ковариация

$$Cov(\xi, \eta) = E_{\xi\eta} - E_{\xi}E_{\eta}$$

Ковариация или **корреляционный момент** показывает на сколько зависимы случайные величины это мера зависимости двух случайных величин.

Если ξ, η - независимые случайные величины

$$Cov(\xi, \eta) = 0$$

:

$$Cov(\xi, \xi) = D_{\xi} = Var_{\xi} \text{ - вариация}$$

3.2 Корреляция

$$Corr(\xi, \eta) = \frac{E_{\xi\eta} - E_{\xi}E_{\eta}}{\sqrt{D_{\xi} \cdot D_{\eta}}} = \frac{Cov(\xi, \eta)}{\sqrt{D_{\xi} \cdot D_{\eta}}}$$

Корреляция - статистическая взаимосвязь двух случайных величин. Корреляция является **нормированной** версией ковариации, что позволяет сравнивать силу линейной зависимости между различными парами переменных, независимо от их масштаба.

Теорема (об ограниченности корреляции)

$$-1 \leq Cor(\xi, \eta) \leq 1$$

Доказательство:

Возьму $\alpha = \xi - \lambda\eta$:

$$D\alpha = D(\alpha) = E\xi^2 - 2\lambda E_{\xi\eta} + \lambda^2 E\eta^2 - (E\xi)^2 + 2\lambda E_{\xi}E_{\eta} - \lambda^2(E_{\eta})^2 \geq 0$$

$$D\xi - 2\lambda Cov(\xi, \eta) + \lambda^2 D\xi \geq 0$$

Откуда, если рассматривать это, как уравнение относительно λ, то $D \leq 0$, то есть:

$$4Cov(\xi, \eta) - 4D_{\eta}D_{\xi} \leq 0$$

А если присмотреться, то это и есть то, что нам надо.

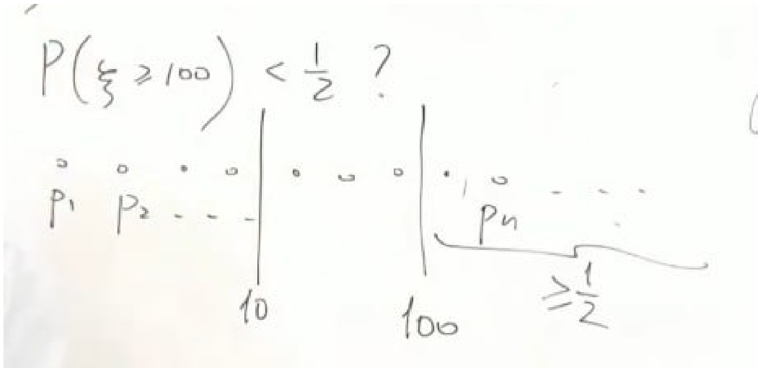
Q.E.D.

3.3 Хвостовые неравенства

Рассмотрим азартную игру. ~~не одобряем, не играем.~~

Проводится случайный эксперимент, смотрится значение ξ. Если оно получилось 100 или больше, то мы платим 100 рублей, а иначе наш друг платит нам 100 рублей. Мы знаем $E\xi = 10, \xi \geq 0$.

Хотим оценить $P(\xi \leq 100)$:



Давайте посмотрим, является ли наша вероятность меньше $\frac{1}{2}$. Тогда всё, что правее 100 имеет вероятность выпадения $\geq \frac{1}{2}$. Все левое оценивается нулем, откуда мат ожидание хотя бы 50. Такого быть не может. В общем случае:

Теорема (Неравенство Маркова)

$$\xi \not\equiv 0, \xi \geq 0 : \forall a \geq 1 : P(\xi \geq a \cdot E\xi) \leq \frac{1}{a}$$

Доказательство:

$$E_\xi = \sum_v v \cdot P(\xi = v) = \sum_{v < a \cdot E\xi} vP(\xi = v) + \sum_{v \geq a \cdot E\xi} vP(\xi = v) \geq \sum_{v \geq a \cdot E\xi} aE\xi P(\xi = v) = aE\xi \cdot P(\xi \geq a \cdot E\xi)$$

Q.E.D.

Теорема (Неравенство Чебышева)

Абсолютная версия и относительная версия ($\alpha = \lambda\sigma$):

$$P(|\xi - E\xi| \geq \alpha) \leq \frac{D\xi}{\alpha^2} \qquad P(|\xi - E\xi| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$

Доказательство:
Возьму вот такие величины:

$$D_\xi = E(\xi - E\xi)^2 \qquad \eta = (\xi - E\xi)^2$$

Заметим, что $E\eta = D\xi$. Используем неравенство Маркова для оценки дисперсии:

$$P(\eta \geq c \cdot E\eta) \leq \frac{1}{c}$$

Возьму $c = \dfrac{D_\xi}{\alpha^2}$ и получу искомое.

Q.E.D.

Нечестная монета. Вот вам дали домашку, вместе с вопросом $p > \frac{1}{2}$ или $p < \frac{1}{2}$. Что вы можете делать? Только кидать ее, но при этом бесконечное количество раз вы не кинете, у вас дедлайн домашки через час.

Пусть мы бросили n раз. Выпало c единиц и $n-c$ нулей. Пусть $c \le \frac{n}{2}$:

$$P(\xi = c) \le P(\xi \le c) \le P(|\xi - pn| \ge pn - c) \le P(|\xi - pn| \ge \frac{n}{2} - c) \le \frac{n}{4} \cdot \frac{1}{\left(\frac{n}{2} - c\right)^2}$$

Что это концептуально значит? На самом деле, это дает нам оценку на распределение. Зачем? Чтобы <u>**СДАТЬ**</u> домашку.

<u>**Теорема (Граница Чернова)**</u>

$$P(\xi \ge (1+\varepsilon)p) \le e^{-\frac{\varepsilon^2}{2+\varepsilon}np} \quad \Leftrightarrow \quad P(\xi \ge (1+\varepsilon)p) \le e^{-\frac{\varepsilon^2}{2}np}$$

$$e^{-\frac{\varepsilon^2}{3}np} \le \delta \quad \Leftrightarrow \quad -\frac{\varepsilon^2}{3}np \le \ln \delta$$

$$n \ge \frac{3}{p\varepsilon^2} \ln \frac{1}{\delta}$$

Не знаю, что это концептуально, напишите пж

4 Лекция 4.

4.1 Введение в теорию информации.

информация = - неопределенность - сказал дяденька Шеннон

Для осознания нам поможет рисунок АС:

Есть что-то - неизвестное - облачко. Затем, вы с помощью глаза заглядываете туда, и ваша неопределенность уменьшается. Соответственно вы получили информацию. То есть сначала была неопределенность H_1, потом H_2. $I = H_1 - H_2$, откуда и получается наша формула. У него есть глубокий смысл, но создается вопрос: «И че? И что это за неопределенность?»

Ну наличие глаза мешает, непонятно, фу фу фу. Поэтому хотим ввести что-то более формальное и менее абстрактное.

Пусть у нас есть какой-то случайный эксперимент Ω, с вероятностями $p_1, \ldots, p_n$. И вот мы получили информацию что выпало (например орел на монетке).

Случайный источник — черный ящик с красной кнопкой, который показывает номер эл. исхода, когда вы нажимаете на красную кнопку.

Возьмем монетку. Кинули, получили 0 или 1. Теперь возьмем кубик, получим число от 1 до 6. Когда мы кидаем кубик, мы получаем больше информации. И вот Шеннон решил систематизировать все это...

4.2 Энтропия

Пусть у нас есть случайный источник и вероятности $p_1, p_2, \ldots, p_n$. Мы хотим померить численно сколько информации содержится в одном эксперименте:

$$H(p_1, \ldots, p_n) : RS \to R^+$$

Энтропия Шеннона(H) - это мера неопределенности или случайности, связанная с случайной переменной. Она измеряет среднее количество информации, необходимое для описания результата случайной переменной. Иными словами, энтропия показывает, насколько непредсказуемым является источник информации.

Возьму пример $p_i = \dfrac{1}{n}$. Введем новое обозначение:

$$h(n) = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$$

Очевидно, что $h(n+1) > h(n)$.

Теперь рассмотрим вероятностное пространство и источник на нем:

$$\Omega = \{(1,1),(1,1),\dots,(1,m_1),(2,1),\dots,(k,1),\dots,(k,m_k)\}$$

И давайте теперь каждому причислим какую-то q_{ij}, так, что в сумме 1. $p_i = \sum\limits_{j=1}^{m_i} q_{ij}$. Пусть наш случайный источник сломан и показывает только одно число. Если я возьму сломанный случайный источник от Ω, то мы получим столько же информации сколько и у случайного источника сделанного из p.

Теперь давайте делить это на 2 части. Что вот мы сначала видим первую часть информации, а потом хоба и видим вторую часть информации. И того мы получаем, что когда мы открываем вторую часть мы получим $p_i H(\frac{q_{i1}}{p_i},\dots,\frac{q_{mi}}{p_i})$ информации. Откуда благодаря таким рассуждение получаем свойство, которое называется **аддитивностью энтропии**:

$$H(p_1,\dots,p_k) + \sum_{i=1}^{k} p_i H(\frac{q_{i1}}{p_i},\dots,\frac{q_{mi}}{p_i}) = H(q_{11},\dots,q_{mk})$$

Также для фиксированного n, H непр из $\mathbb{R}^n \to \mathbb{R}$.

Теорема.(Формула энтропии Шеннона)

$$H(p_1,\dots,p_n) = -\alpha \sum_{i=1}^{n} p_i \log_2 p_i$$

α отвечает за выбор единицы измерений.

Доказательство:

Лемма 1. $h(n\cdot m) = h(n) + h(m)$.

Доказательство:

Возьмем $k = n, m_i = m, p_= \frac{1}{n}, q_{ij} = \frac{1}{nm}$. Из утверждения сверху это верно!

Q.E.D.

Фиксируем $h(2) = \alpha$. Тогда:

Лемма 2. $h(2^k) = k\alpha$. тривиально из Леммы 1.

Лемма 2,5. $h(n^r) = rh(n)$. тривиально из Леммы 1.

Лемма 3. $h(n) = \alpha \log_2 n$

Доказательство:

Найду i такое, что $2^i \le n^r < 2^{i+1}$, где $r \in \mathbb{N}$.

Из монотонности h следует: $\alpha i \le h(n^r) < \alpha(i+1)$. Поэтому:

$$\alpha i \le rh(n) < \alpha \quad \Leftrightarrow \quad a\frac{i}{r} \le h(n) \le a\frac{i+1}{r}$$

Также мы знаем, что $i \le r\log_2 n < i+1$. Получим, что:

$$\alpha\frac{i}{r} \le \alpha log_2 n < \alpha\frac{i+1}{r}$$

То есть $\forall r : |h(n) - \alpha log_2 n| \le \frac{\alpha}{r}$. Откуда, получаем требуемое равенство.

Q.E.D.

Возвращаемся к доказательству теоремы. Пусть p_i рациональные. Приведем все p к общему знаменателю и пусть теперь $p_i = \frac{a_i}{b_i}$. Возьму $m_i = a_i$, $r_{ij} = \frac{1}{a_i}, q_{ij} = \frac{1}{n}$. Подставим во второе неравенство получим:

$$H\left(\frac{1}{b}, \frac{1}{b}, \dots, \frac{1}{b}\right) = H(p_1, p_2, \dots, p_k) + \sum_{i=1}^{k} p_i H\left(\frac{1}{a_i}, \dots, \frac{1}{a_i}\right)$$

Что тут происходит? Я разбиваю каждый исход изначальный, на a_i исходов по $\frac{1}{b_i}$. С одной стороны я получаю b исходов по $\frac{1}{b}$. С другой стороны я могу выбрать исход, а потом его разбить. Откуда по аддитивности и получается такая формула. А она в свою очередь уже удобная, так как в ней повторяются значения внутри H, так что можем заменить на h:

$$h(b) = H(p_1, \dots, p_k) + \sum_{i=1}^{k} p_i h(a_i)$$

Заметим, что $\sum_{i=1}^{n} p_i = 1$, так что левую часть на эту сумму:

$$\sum_{i=1}^{n} p_i h(b) = H(p_1, \dots, p_k) + \sum_{i=1}^{k} p_i h(a_i)$$

$$H(p_1, \dots, p_k) = \sum_{i=1}^{n} p_i (h(b) - h(a_i))$$

$$H(p_1, \dots, p_k) = \sum_{i=1}^{n} p_i (\alpha log_2 b - \alpha log_2 a_i) = -\alpha \sum_{i=1}^{n} p_i \log_2 p_i$$

Эта формула верна и не для рац. исходя непрерывности (любое не рац. можно зажать с двух сторон сходящимися последовательностями и мы победили)

Q.E.D.

α — бит, единица информации.

Обычно используется логарифм по основанию 2, тогда энтропия измеряется в битах (или "Шеннонах"). Если используется натуральный логарифм (основание e), то энтропия измеряется в натах. Использование логарифма по основанию 10 даёт единицы измерения в децитах (Hartleys). Выбор основания влияет только на масштаб энтропии, а не на её относительные значения.

Энтропия Шеннона имеет широкое применение в различных областях, включая:

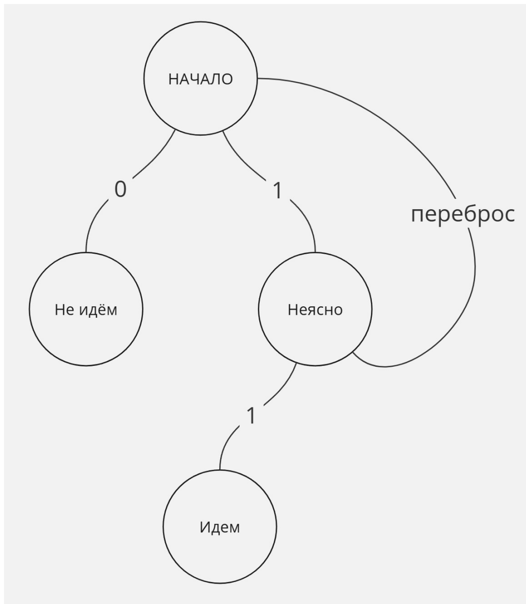
- Теория информации: Является фундаментальным понятием для измерения количества информации.
- Сжатие данных: Используется для оценки теоретического предела сжатия данных.
- Криптография: Оценка случайности ключей и стойкости шифров.
- Машинное обучение: В деревьях решений используется для выбора признаков, которые лучше всего разделяют данные.
- Обработка естественного языка (NLP): Оценка неопределенности в языковых моделях.
- Термодинамика: Аналогична термодинамической энтропии, отражает меру беспорядка в системе.

Также есть такие понятия, как взаимная энтропия и условная энтропия. Их определения появятся в конспекте после того, как пройдет неделя со сдачей домашки.

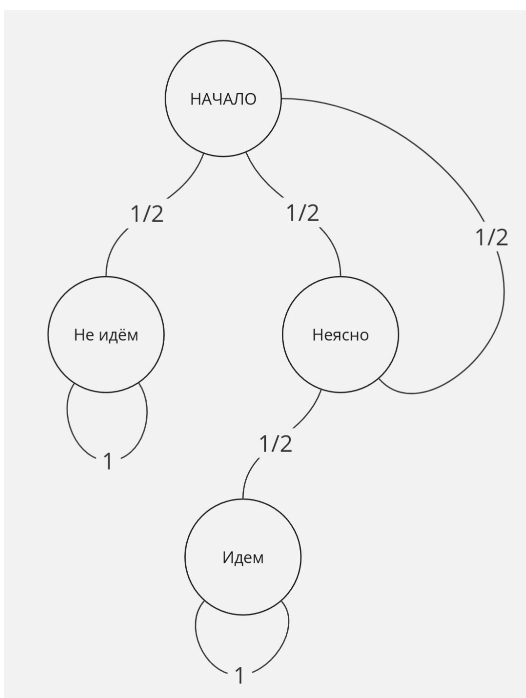
5 Лекция 5.

5.1 Марковские цепи

Вспоминаем задачку с Петей, монеткой и кинотеатром и визуализируем в виде ориентированного графа.



Заметим, что результаты нам не особо важны, важны только вероятности, перерисуем и дополнительно заведём ограничение на сумму значений вероятностей исходящих из вершины: сумма должна всегда должна быть равна единичке.



Давайте для такого графа введём определение.

Марковской цепью называется ориентированный граф с конечным числом вершин, у которых на исходяших рёбрах написаны вероятности, причём для каждой вершины верно, что сумма этих рёбер равна единице. Введём ещё парочку определений.

Вершина называется **полглощающей** если из неё исходит только петля.

Обозначим векторами $b_i^{(t)} = P(\xi_t = i)$, $b^t = (b_1^{(t)}, \dots, b_n^{(t)})$ - распределение ξ_t.
Давайте с такими данными построим матричку $P_{n\times n}$, где значение в ячейке – это вероятность перейти из i-той вершины в j-тую. Разберём пример:

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Хочется что-то с этой матрицей начать делать, начнём же:

Лемма 1. $b^{(t+1)} = b^{(t)}P$

Доказательство:

$$b_i^{(t+1)} = P(\xi_{t+1} = i) = \sum_{j=1}^{n} P(\xi_{t+1} = i \quad \& \quad \xi_t = j)P(\xi_t = j) = \sum_{j=1}^{n} b_j^{(t)} P_{ji} = (b^{(t)}P)_i$$

Q.E.D.

По коэффициенту очевидности $b^{(t)} = b^{(0)}P^t$, проверяйте сами

МЦ называется **поглощающей**, если $\forall u$ $\exists$ путь такой, что $u \to$ поглощаемая вершина

Введём отношение эквивалентности такое, что $u \sim v$ если $\exists$ пусть $u \to v$ и наоборот.

Можем заметить, что у нас образуется разбиение на классы эквивалентности, рассмотрим $[1; n] \; / \sim$, $\tilde{u}$ – класс эквивалентности

$\tilde{u}$ – **эргодический класс**. тем самым мы можем разделять задачи на две части: на наблюдение свойств в самом классе и между классами.

МЦ – **эргодическая**, если в ней ровно 1 эргодический класс

Эргодический класс называется **периодическим**, если $\exists d > 1$, что размер любого цикла делится на d – сама d называется **периодом**, очевидно, что период – нод длин всех циклов

Теорема о классификации Марковских Цепей.

В любой МЦ есть поглощающий эргодический класс, вероятность того, что МЦ окажется в одном таких классов стремится к единице

Вторая часть, я не могу разобрать свой почерк, на паре мы это не доказывали пока

Доказательство:

$$P = \left(\begin{array}{c|c} Q & R \\ \hline \mathbf{0} & E \end{array}\right), \text{ где } Q_{m\times m}, R_{m\times(n-m)}, \text{ где } m \text{ – это непоглощающие эргодические классы}$$

тогда каждый вектор $b^{(t)}$ делится на две части: первые m координат обзовём $a^{(t)}$, вторые назовём $c^{(t)}$. $a_i^{(t)} = P(\xi_t = i)$, где i непоглощающая

Лемма 2. $a^{(t)} = a^{(0)}Q^t$, лемма очевидна просто взгляните, на то, как умножается вектор на матрицу

Лемма 3. $Q^k \xrightarrow[k\to+\infty]{} \mathbf{0}$

Доказательство:

Для того, чтобы найти предел матрицы нам нужно ввести норму, норму введём такую $|A| = max_{ij}|a_{ij}|$, сами докажете, что это норма, не маленькие.

Тогда матрица стремится к нулю, если её норма стремится к нулю, наша задача показать, что $Q^n \xrightarrow[n\to+\infty]{} \mathbf{0}$

Пусть L - максимальная длина кратчайшего пути от i до поглощения, давайте посмотрим чему равно $X = Q^L$, рассмотрим $x_{ij} = \sum_{k_1,k_2,..k_{l-1}} q_{ik_1}q_{k_1,k_2}q_{k_2,k_3}q_{l-1,j}$. Чему равна эта сумма, неясно, каждое из этиъ произведений – это вероятность пройти оп пути длиной ровно L от вершины i до j, давайте рассмотрим такую сумму

$$\sum_{j\text{-не погл}} x_{ij} = \sum_{k_1,k_2,..k_{l-1},j} q_{ik_1}q_{k_1,k_2}q_{k_2,k_3}q_{l-1,j} = \delta < 1$$

Меньше единицы оно потому, что мы смотрим не на все пути, а лишь на часть. Посмотрим на матричку:

$$Q^n_{ij} = (Q^LQ^{n-L})_{ij} = \sum_k Q^L_{ik}Q^{n-L}_{kj} \le (\sum_k Q^L_{i,k})|Q^{n-L}| \le \delta|Q^{n-L}| \text{ т.к. } |Q^n| <= \delta^{\lfloor n/l \rfloor}$$

то наше выражение стремится к нулю

Лемму доказали, стало быть, и теоремка доказана.

Передём к вычислительным моментам, мы хотим понять, где мы поглотимся. Найдём мат. ожидание времени до поглощения Давайте введём T – случайная величина : число шагов до поглощения, тогда $T = \sum_{i=1}^m T_i$ T_i – число посещений i-го состояния

$T_i = \sum_{j=0}^\infty T_{ij}$

$$T_{ij} = \begin{cases} 1, & \text{если на шаге } j \text{ марковская цепь находится в состоянии } i, \\ 0, & \text{иначе.} \end{cases}$$

$$P(\text{м.ц на шаге } j \text{ в состоянии } i = (a^0Q^j)_i$$

$$ET = \sum_{i=1}^n ET_i = \sum_{i=1}^m \sum_{j=0}^\infty ET_{ij} = \sum_{i=1}^m \sum_{j=0} (a^0Q^j)_i$$

Воспользуемся линейной алгеброй, т.к. сумма матриц перестановочна относительно взятия i-го компонента

$$\sum_{i=1}^{m}\sum_{j=0}^{\infty}(a^0Q^j)_i = \sum_{i=1}^{m}(\sum_{j=0}^{\infty} a^0Q^j)_i$$

т.к. a^0 не зависит от j

$$\sum_{i=1}^{m}(\sum_{j=0}^{\infty} a^0Q^j)_i = \sum_{i=1}^{m} a^0(\sum_{j=0}^{\infty} Q^j)_i$$

Q.E.D.

Лемма $\sum_{j=0}^{\infty} Q^j = (I-Q)^{-1}$

Доказательство:

$$(E-Q)(E+Q+Q^2+Q^3+\ldots) = E-Q+Q-Q^2+Q^2+\ldots-Q^{n+1}$$

$$= E-Q^{n+1} = E \text{ т.к. } Q \text{ стремится к нулю}$$

Q.E.D.

Фундаментальная матрица полглощения м.ц. – это $N = (I-Q)^{-1}$

$$P(\text{погл в } j) = \sum_{i}^{m} P(\text{погл в } j \text{ из } i)P(j \text{ перед поглощением}) = \sum_{t}^{\infty}\sum_{i=1}^{m} P(\text{погл в } j \text{ из } i \text{ в } t)$$

$$= P(\text{быть в i в момент вермени t}) = a^0NR$$

Откуда получаем 2 крутые и очень полезные в дальнейшем формулы:

6 Лекция 6.

Поглощающий эргодическтий класс – класс, из которог в конденсации нет рёбер

МЦ **регулярная** если $\forall i, j\ P_{i,j} > 0$

Теорема: Эргодическая теорема для марковских цепей

$\exists b\ \forall b^{(0)} P^{(n)} \xrightarrow[n\to\infty]{} b$ и $bP = b$

Доказательство:

Давайте возьмём наш вектор $b^{(0)}$ и умножим на матрицу A такую, что $A \in M_{n\times n}$ и $\forall j\ a_{ji} = \tilde{a}$

$(b \cdot A)_i = \sum_{j=1}^{n} b_j a_{ji} = (\sum_{j=1}^{n} b^{(0)})\tilde{a_i} = \tilde{a_i}$. т.к. сумма в столбике $b^{(0)}$ всегда равна единичке

Докажем, что $P^n \xrightarrow[n\to\infty]{} A$, удовлетворяющей условиям.

Введем $M_i^{(t)} = max(P^t)_{ji},\ m_i^{(t)} = min(P^t)_{ji}$, посмотрим на разность максимума и минимума в столбце, правда ли он стремится к нулю?

Возьмём $\delta = \max P_{ij}$ и $\tilde{\delta} = \min P_{ij}$

$$(P^{(t+1)})_{ij} = (P\ \cdot\ P^t)_{ij} = \sum_{k=1}^{n} P_{jk}(P^t)_{ki} \le \sum_{n!=pos(min)} P_{jk}\ \cdot\ M_i^{(t)} + P_{j,pos(min)}m_i^{(t)} =$$
$$= \sum_{k=1}^{n} P_{ik}\ \cdot\ M_i^{(t)} + P_{j,pos(min)}(m_i^{(t)} - M_i^{(t)}) \le M_i^{(t)} - \delta A^{(t)}$$

с другой стороны

$$(P^{(t+1)})_{ij} = (P\ \cdot\ P^t)_{ij} = \sum_{k=1}^{n} P_{jk}(P^t)_{ki} \ge \sum_{k=1}^{n} P_{jk}m_i^{(t)} + P_{j,pos(max)}(M_i^{(t)} - m_i^{(t)}) = m_i^{(t)} + \tilde{\delta}A_i^{(t)}$$

$$M_i^{(t+1)} \le M_i^{(t)} - \delta A_i^{(t)}$$
$$m_i^{(t+1)} \ge m_i^{(t)} + \tilde{\delta} A_i^{(t)}$$

минусуем

$$A_i^{(t+1)} \le A_i^{(t)} - (\delta - \tilde{\delta}A_i^{(t)}) = (1 - \delta - \tilde{\delta})A_i^{(t)} \le \ldots \le (1 - \delta - \tilde{\delta})^{(t+1)} \to \mathbf{0}$$

Если разность стремится к 0, то в $b^{(0)}P^{(n)} \xrightarrow[n\to\infty]{} b^0 A = b$

Q.E.D.

Давайте научимся искать b, решим следующее уравнение: $(I - P)b = \mathbf{0}$. Казалось бы, это СЛОУ и она может иметь либо 0 решений, либо бесконечно много, тогда почему же у нас есть конкретное решение? На самом деле, $(I - P)$ – вырожденная, и она имеет ранг $n - 1$ и у нас одномерное пространство решений, но нам подойдёт только тот вектор, у которого сумма координат единичка. Тогда как в итоге найти b?

У нас есть два способа, мы либо возведём много раз матрицу P в квадрат, с какого-то момента это всё будет стремиться к A, мы найдём ответ и, тогда сможем вычислить b, все операции займут у нас $O(n^3c)$, понятно, что если использовать какие-то продвинутые алгоритмы, оценку можно улучшить.

Либо же просто решим систему уравнений, но это тоже имеет свои тонкости, как минимум нам понадобится работать с вещественными числами, а мы знаем, что вещественные числа в компьютере несовершенны и дорого с ними работать, хоть и асимтотика номинально $O(n^3)$, но она упирается в проблемы работы с даблами.

Наши теоремы мы доказали только для регулярных МЦ, но что происходит с нерегулярными?

Теорема

Если P – матрица нерегулярного эргодического класса, тогда $\exists t : P^t$ - регулярное Доказывать мы это не будет, просто поверим на слово

7 Лекция 7.

7.1 Формальные языки

Алфавит - Σ - конечное или пустое

Множество слов длины n: Σ^n

$\Sigma^* = \bigcup\limits_{k=0}^{\infty} \Sigma^k$

Язык - подмножество слов в алфавите $L \subset \Sigma^k$

Замечание. Множество языков 2^{Σ^*} несчетно.

7.2 Описание языка.

Вот мы хотим описать язык. И описать его словами например: все слова четной длины, они содержат четное число единиц. Но мы не можем дать эти слова компьютеру или человеку, говорящему на другом языке.

Поэтому существует несколько способов описания языков:

1. Перечисление: Явное перечисление всех слов в языке (возможно только для конечных языков). Например, $L = \{a, ab, aba\}$.
2. Порождающая грамматика (формальная грамматика): Набор правил, определяющих, какие слова принадлежат языку.
3. Распознающий автомат: Машина (например, конечный автомат), которая принимает или отклоняет слова в зависимости от того, принадлежат они языку или нет.
4. Регулярное выражение: Шаблон, описывающий структуру слов в языке.
5. Предикат: Математическое условие, которому должны удовлетворять слова, чтобы принадлежать языку.

Замечание. Всего описаний счетное множество

Фанфакт. Если вы захотите найти лексикографически минимальный язык без описания, то когда вы найдете его,у него появится описание. Мы не будем думать об этом и такого рода парадоксах

7.3 Регулярные (автоматные языки)

У нас есть операция конкатенации, определена она таким образом:

$\alpha, \beta \in \Sigma^* : \varphi = \alpha\beta$, причем $\alpha \in \Sigma^K, \beta \in \Sigma^l, \alpha\beta \in \Sigma^{k+l}$, а также $\varphi_i = \begin{cases} i \le k \Rightarrow \alpha_i \\ i > k \Rightarrow \beta_{i-k} \end{cases}$

Она ассоциативна и есть нейтральный элемент ε.

Но теперь хотим ввести понятие **конкатенации языков**:

$$AB = \{x | x = yz, y \in A, z \in B\}$$

Пример:

$A = \{0, 01\}, B = \{0, 10\}$, тогда $AB = \{00, 010, 0110\}$

А теперь обо всех операциях

Пусть L_1 и L_2 - языки над алфавитом Σ. Тогда можно определить следующие базовые операции:

1. Объединение: $L_1 \cup L_2 = \{w | w \in L_1 \mid\mid w \in L_2\}$
2. Конкатенация: $L_1 L_2 = \{w_1 w_2 | w_1 \in L_1, w_2 \in L_2\}$
3. Степень: $L_1^0 = \{\varepsilon\}, L_1^{n+1} = L_1 L_1^n$
4. Замыкание Клини: $A^* = \bigcup\limits_{k=0}^{\infty} A^k$

$Reg_0 = \{\varnothing, \{\varepsilon\}, \{c\} \text{for } c \text{ in } \Sigma\}$ - **базовые регулярные языки**

$Reg_{i+1} = Reg_i \cup \{A \cup B, AB, A^* | A, B \in Reg_i\}$

Семейство регулярных языков $Reg = \bigcup_{k=0}^{\infty} Reg_k$

Лемма:

$A, B \in Reg \Rightarrow A \cup B \in Reg, AB \in Reg, A^* \in Reg.$

Доказательство:

$A \in Reg_i, B \in Reg_j$, откуда по определению $A \cup B, AB, A^* \in Reg_{\max(i,j)+1}$

Q.E.D.

$X \in Good, X \subset 2^{\Sigma^*}, X - \text{set}\langle\text{language}\rangle, Good - \text{set}\langle\text{set}\langle\text{language}\rangle\rangle$

Good - множество множеств языков. Мы называем его **хорошим** если:

1. $Reg_0 \subset X$
2. X замкнут относительно базовых операций.

Теорема. $Reg = \bigcap\limits_{u \in Good} u$

Доказательство не будет.

$\varnothing$ - пустое множество, ε - пустая строка.

Пусть у языка A описание α. У языка B описание β. Тогда введем обозначение:

1. AB описание будет $\alpha\beta$ — средний приоритет
2. $A \cup B$ описание будет $\alpha | \beta$ — минимальный приоритет
3. A^* описание будет a^* — максимальный приоритет

В зависимости от приоритета будем заключать выражения в скобки. Такие записи называются **академическими регулярными выражениями**.

Также есть еще две обозначения:

1. $\alpha^+ = \alpha\alpha^*$.
2. $\alpha^k = \alpha \dots \alpha$ - k раз.

7.4 Детерминированный Конечный Автомат (ДКА)

В разделе марковских цепей у нас были графы с вероятностями на ребрах. А давайте теперь удалим вероятности.

$A = \langle \Sigma, Q, S, T, \sigma : Q \times \Sigma \to Q$, где

1. Σ - алфавит
2. Q - конечное множество состояний
3. $S \in Q$ - начальное состояние.
4. $T \subset Q$ - допускающее (конечное) состояние.
5. σ - функция перехода, которая по текущему состоянию переходит

$Snap = Q \times \Sigma^*$ — **множество мгновенных описаний** автомата или состояний.

$\vdash$: $\langle a, \alpha \rangle \vdash \langle r, \beta \rangle$

1. $\alpha = c\beta, c \in \Sigma$
2. $r = \delta(q, c)$

Это описание того, как мы бегаем по графу. То есть например пусть у нас есть автомат, который описывает четное или нечетное число единиц в строке и ему подали 0101. Тогда для него мгновенные описания:

$$\langle ch, 0101 \rangle \vdash \langle ch, 101 \rangle \vdash \langle nech, 01 \rangle \vdash \langle nech, 1 \rangle \vdash \langle ch, \varepsilon \rangle$$

Языком автомата $L(A) = \{w | \langle s, w \rangle \vdash^* \langle t, \varepsilon \rangle, t \in T\}$

Теорема (Клини)

$Reg = Aut$, где Aut - множество языков, задающихся ДНК.

Доказательства не будет

8 Информация о курсе

Поток — y2024.
Группы M3138-M3142.
Преподаватель — Станкевич Андрей Сергеевич.

В данном семестер фокусируются 2 темы: Дискретная теория вероятности и представление слов (токенов) в компьютере.