Введение в теорию информации

Тронов Антон
Высшая школа экономики

2021 г.

Цитата Н. Винера

«Информация есть информация,
а не материя или энергия»
Н. Винер

Общее понятие информации

Слово «информация» происходит от лат. informatio, что в переводе обозначает сведение, разъяснение, ознакомление.

На сегодняшний день существует большое число различных определений для понятия «информация», различные области научного знания предлагают свои варианты данного определения.

Несмотря на широкую распространённость, понятие информации остаётся одним из самых дискуссионных в науке.

Разные определения информации

- В кибернетике под информацией понимает ту часть знаний, которая используется для ориентирования, активного действия, управления, т.е. в целях сохранения, совершенствования, развития системы (**Н. Винер**);
- Информация — это сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности, неполноты знаний (**Н.В. Макарова**);
- Информация — это отрицание энтропии (**Леон Бриллюэн**);

Разные определения информации

- Информация — это мера сложности структур (**Моль**);
- Информация — это отраженное разнообразие (**Урсул**);
- Информация — это содержание процесса отражения (**Тузов**);
- Информация — это вероятность выбора (**Яглом**);
- Информация — это снятая неопределенность наших знаний о чем-то (**Клод Шеннон**).

Понятие информации в математике

Исторически, в математике понятие информации стало предметом изучения в контексте теории кодирования и теории связи.
Основоположники математической теории информации:

- Ральф Хартли
- Клод Шеннон
- Андрей Николаевич Колмогоров

Отцы основатели

Вопрос о количестве информации

С появлением первых носителей цифровой информации возник вопрос об оценке количества информации.
Например, сколько информации на перфокарте (один из первых способов ввода программного кода)?

Перфокарта

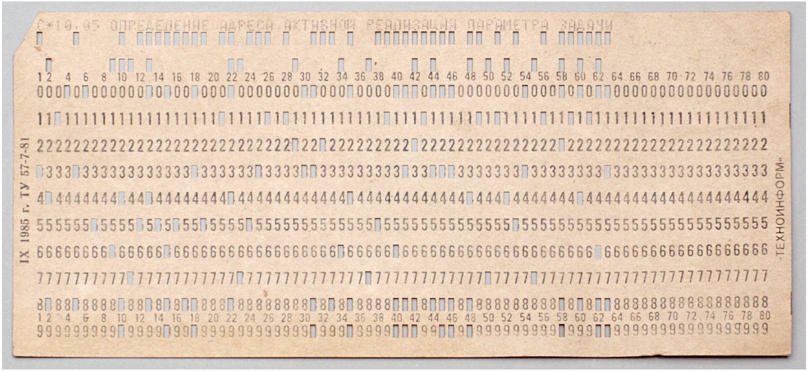


Рис. 1: Перфокарта с кодом

Количество информации

Первоначально были связать количество информации с числом возможных состояний носителя.
Пусть мы имеем перфокарты с N позициями, тогда на каждая такая перфокарта может иметь $C = 2^N$ возможных состояний.
Очевидно, что количество состояний двух таких же первокарт

$$C = 2^N \cdot 2^N = 2^{2 \cdot N}$$

Это противоречит здравому смыслу.

Количество информации двух носителей

Очевидно, две перфокарты должны содержать вдвое больше информации чем каждая по отдельности. Значит, количество информации I и число состояний C для двух носителей должны подчиняться условию:

$$I(C_1 \cdot C_2) = I(C_2) + I(C_2)$$

Сразу напрашивается логарифм на роль искомой функции

Количество информации по Хартли

Определение. Пусть дан алфавит $A = \{a_1, a_2, \ldots, a_L\}$ мощности L и сообщение M длины N символов из алфавита A, тогда хартлиевское количество информации(еще называют логарифмическая мера информации или мера Хартли) определяется формулой:

$$I(M) = \log_2 L^N = N \cdot \log_2(L)$$

Проще говоря, мера Хартли - это двоичный логарифм от числа состояний носителя информации.

Мера Хартли двоичного носителя

Применим меру Хартли для вычисления количества информации двоичного носителя (напр. перфокарты). Мера Хартли для сообщения M длины N символов из двоичного алфавита $A = \{0, 1\}$ равна:

$$I(M) = \log_2(2^N) = N \cdot \log_2(2) = N \text{ Бит}$$

Бит - количество информации в сообщении, состоящем из одного символа двоичного алфавита $\log_2(2) = 1$ Бит.

Мера Хартли. Пример

Из формулы Хартли следует, что алфавит, содержащий только 1 символ не может быть использован для передачи информации:

$$I(M) = N \cdot \log_2(1) = 0$$

т.к количество передаваемой информации нулевое.

Информация по Шеннону

Шеннон предложил новое концептуальное математическое определение информации, согласно Клоду Шеннону:
"информация — это снятая неопределенность то есть то, что уменьшает меру неопределенности.
В своей работе "Математическая теория связи"он вводит понятие информационной энтропии, как коичественную меру неопределенности и информации.

Информация по Шеннону

Итак, информация - это устранение неопределенности. Следовательно, чем более неопределенно было состояние системы, тем больше информации мы получаем узнав это состояние.
Например, при подбрасывании монеты реализуются два элементарных события {Орел, Решка}, вероятность каждого равна $\frac{1}{2}$
при бросании игральной реализуются шесть элементарных событий $\{1, 2, 3, 4, 5, 6\}$, вероятность каждого равна $\frac{1}{6}$.
Следовательно, при бросани кости мы получаем большую информацию чем при подбрасывании монеты.

Информация по Шеннону

При подбрасывании двух игральных костей число вариантов P есть произведение, числа вариантов для каждой отдельной кости P_1 и P_2:

$$P = P_1 \cdot P_2 = 6 \cdot 6 = 36$$

Количество информации, очевидно, суммируется:

$$I(P) = I(P_1) + I(P_2) = \log_2 6 + \log_2 6 = \log_2 6^2 = 2 \cdot \log_2 6$$

Энтропия. Общее определение. Виды

Энтропия(от греч. $\epsilon\nu\tau\rho o\pi$ - «обращение, превращение) в самом общем случае есть мера неупорядоченности системы.
Данное понятие в разных вариациях используется во многих областях науки:

- Топологическая и метрическая энтропии в эргодической теории и теории динамических систем;
- Информационная энтропия — в теории информации мера неопределённости источника сообщений, определяемая вероятностями появления тех или иных символов при их передаче.

Виды энтропии

- В статистической физике — характеризует вероятность осуществления некоторого макроскопического состояния системы.
- Термодинамическая энтропия — термодинамическая функция, характеризующая меру необратимой диссипации энергии в ней.
- Дифференциальная энтропия — формальное обобщение понятия энтропии для непрерывных распределений

Подход Клода Шеннона

Клод Шеннон применил к понятию информации статистический подход. Он переформулировал задачу следующим образом: пусть дан конечный автомат без памяти, который последовательно выдает символы из некторого конечного алфавита, причем вероятности выпадения каждого символа нам известны. Сколько общих вопросов(предполагающих ответ Да или Нет) нужно задать, чтобы определить один символ такой машины. Это и будет количеством информации на один символ в данной системе.

Информационная энтропия

Определение. Пусть X источник(автомат) без памяти, генерирующий символы из алфавита $\{x_1, x_2, \ldots, x_N, \}$ с соответствующими вероятностими $(p_1, p_2, \ldots, p_N)$, тогда его информационная энтропия равна:

$$H(x) = -\sum_{i=1}^{N} [p_i \cdot \log_2(p_i)]$$

Информация по Хартли vs по Шенону

Формула Хартли - частный случай формулы Шенона, когда вероятности одинаковы:

$$p_1 = p_2 = \ldots = p_M = \frac{1}{M}, \text{ где } M = L^N$$

$$-\sum_{i=1}^{M} [p_i \cdot \log_2(p_i)] = \sum_{i=1}^{M} \left[p \cdot \log_2(\frac{1}{p})\right] = L^N \cdot \frac{1}{L^N} \cdot \log_2(L^N)$$

Нахождение энтропии. Пример

Пусть дан конечный автомат F_1 без памяти, который выдает символы алфавита $\{a, b, c, d\}$ с вероятностями $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ соответственно, его энтропия равна:

$$H(F_1) = -4 \cdot \frac{1}{4} \cdot \log_2(\frac{1}{4}) = \log_2(4) = 2 \text{ Бит}$$

Нахождение энтропии. Пример

Также, дан второй источник F_2 (автомат без памяти), который выдает символы из того же алфавита $\{a, b, c, d\}$ с вероятностями $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{4})$ соответственно, его энтропия равна:

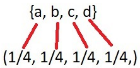
$$H(F_2) = -\left[\frac{1}{2}\log_2(\frac{1}{2}) + 2\frac{1}{8}\log_2(\frac{1}{8}) + \frac{1}{4}\log_2(\frac{1}{4})\right] =$$

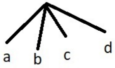
$$\frac{1}{2}\log_2(2)+2\frac{1}{8}\log_2(8)+\frac{1}{4}\log_2(4) = \frac{1}{2}+\frac{3}{4}+\frac{1}{2} = 1.75 \text{ Бит}$$

Сравнение источников

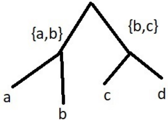
Как мы видим, первый источник обладает большей неопределенностью, и следовательно призводит больше информации, так для определения 100 выдаваемых им элементов нужно задать 200 вопросов. Для второго источника неопределенность меньше, для определения 100 выдаваемых им элементов нужно задать только 175 вопросов.

Пример реализаций схем выбора



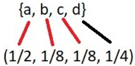


4 вопроса да/нет

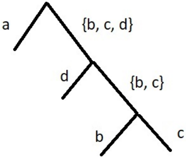


2 вопроса да/нет

Пример последовательности: (a, b, a, c, d, c, b, d)



Пример последовательности: (a, b, c, a, d, d, a, a)



Максимально 3 вопроса да/нет, но в среднем менее 2 вопросов

Игра Бар-Кохба

Исчисление информации в битах означает ответы на вопросы в форме «да» или «нет». В книге венгерского математика А. Реньи «Трилогия о математике» такая игра в вопросы и ответы называется игрой «Бар-Кохба». Происхождение названия следующее. В 135 г. н. э. произошло восстание иудеев против римского господства, руководимое Бар-Кохбой. В римский лагерь был послан лазутчик. Он был пойман, ему отрезали язык и отрубили руки. Он вернулся к своим, но не мог ни говорить, ни писать. Однако Бар-Кохба получил всю нужную информацию, задавая вопросы в форме «да или нет», на что несчастный лазутчик мог отвечать кивком головы.

Максимальная энтропия. Ф-я Шеннона

Как видно из предыдущих примеров, равновероятный источник обладает большей энтропией.
Покажем это на примере функции Шеннона, описывающий энтропию двоичного источника как функцию от вероятности p реализации одного из элементарных событий, вероятность реализации второго будет,очевидно, равна $1 - p$.
Энтропия достигает максимума при $p = \frac{1}{2}$.

$$H(p) = -\left[p\log_2(p) + (1-p)\log_2(1-p)\right]$$

График функции Шеннона

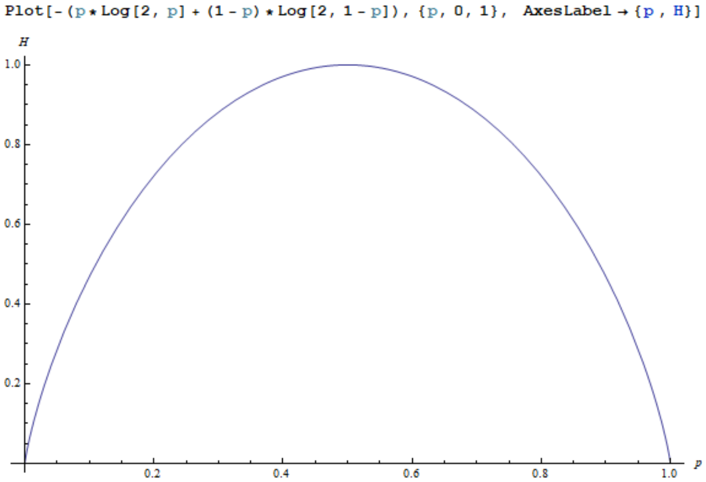


Рис. 2: График функции Шеннона

Свойства энтропии

Как и раньше, X источник(автомат) без памяти, генерирующий символы из алфавита $\{x_1, x_2, \dots, x_N, \}$ с соответствующими вероятностими $(p_1, p_2, \dots, p_N)$, тогда $H(X)$ обладает следущими свойствами:

- Неотрицательность: $H(X) \geq 0$
- Ограниченность: $H(X) = \sum_{i=1}^{N} \left[p_i log_2(\frac{1}{p_i}) \right] \leq \log_2(N)$
- Если X и Y независимы, то:

$$H(X \cdot Y) = H(X) + H(Y)$$

Свойства энтропии

Как и раньше, X источник(автомат) без памяти, генерирующий символы из алфавита $\{x_1, x_2, \dots, x_N, \}$ с соответствующими вероятностими $(p_1, p_2, \dots, p_N)$, тогда $H(X)$ обладает следущими свойствами:

- $H(X)$ выпуклая вверх функция
- Если X и Y имеют одинаковое распределения вероятностей, то они имеют одинаковые энтропии:

$$H(X) = H(Y)$$

Литература:

- 1. М. В. Волькенштейн Энтропия и информация. — М.: Наука, 2006.
- 2. П. Биллингслей. Эргодическая теория и информация. М.: Мир, 1969
- 3. М. Вернер Основы кодирования. — М.: Наука, 2004.

Доклад закончен

Спасибо за внимание!