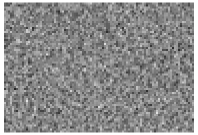
Информатика
для втузов

Сосни́н Владимир Валерьевич, к.т.н.
Бала́кшин Павел Валерьевич, к.т.н.

Краткое содержание курса

- Основы теории информации
- Сжатие компьютерных данных
- Помехоустойчивое кодирование
- Архитектура ЭВМ
- Организация компьютерных сетей
- Работа с офисными пакетами
- Программное обеспéчение профессионального программиста



Требования к слушателям: освоенный школьный курс информатики.

Определение термина «информатика»

Информатика – дисциплина, изучающая свойства и структуру информации, закономерности ее создания, преобразования, накопления, передачи и использования.

Англ: informatics = information technology + computer science + information theory

Важные даты

- 1956 – появление термина «информатика» (нем. Informatik, Штейнбух)
- 1968 – первое упоминание в СССР (информология, Харкевич)
- 197X – информатика стала отдельной наукой
- 4 декабря – день российской информатики

Терминология: информация и данные

Международный стандарт ISO/IEC 2382:2015
«Information technology – Vocabulary» (вольный пересказ):

> Информация – знания относительно фактов, событий, вещей, идей и понятий.
>
> Данные – форма представление информации в виде, пригодном для передачи или обработки.

- Что есть предмет информатики: информация или данные?
- Как измерить информацию? Как измерить данные?
  Пример: «Байкал — самое глубокое озеро Земли».

Некоторые признаки классификации информации

- **Точность.**
  IT-область: методы округления, сжатие медиа-данных.
- **Ценность.**
  IT-область: защита информации.
- **Стадия обработки** (исходная, промежуточная и результирующая).
  IT-область: программирование, базы данных.
- **Полнота** (недостаточная, достаточная, избыточная).
  IT-область: сжатие данных.
- **Близость к первоисточнику** (первичная, вторичная).
  IT-область: интеллектуальный анализ данных, Big Data, базы данных.

Некоторые признаки классификации информации (2)

- **Методы получения** (эмпирическая, теоретическая, эмпирико-теоретическая).
  IT-область: имитационное моделирование, компьютерная математика.
- **Актуальность.** IT-область: система обновлений ПО, системы управления содержимым.
- **Достоверность.** IT-область: MD5-суммы, открытые сертификаты (https).
- **Роль при передаче** (входная, выходная и внутренняя).
  IT-область: компьютерные сети, аппаратные интерфейсы компьютера.
- **Изменчивость** (постоянная, переменная, смешанная).
  IT-область: кэширование, системы хранения данных.
- **Доступность** (открытая, закрытая, конфиденциальная, секретная)
  IT-область: криптография.

ТЦ «СПБ—МАДРИД» – мнемотехническая аббревиатура для запоминания перечисленных признаков классификации, т.е. торговый центр «Санкт-Петербург—Мадрид».

Измерение количества информации

Количество информации ≡ информационная энтропия – это численная мера непредсказуемости информации. Количество информации в некотором объекте определяется непредсказуемостью состояния, в котором находится этот объект.

Пусть i (s) — функция для измерения количеств информации в объекте s, состоящем из *n* независимых частей s_k, где *k* изменяется от 1 до *n*. Тогда **свойства меры количества информации i(s)** таковы:

- Неотрицательность: $i(s) \geq 0$.
- Принцип предопределённости: если об объекте уже все известно, то $i(s) = 0$.
- Аддитивность: $i(s) = \sum i(s_k)$ по всем *k*.
- Монотонность: i(s) монотонна при монотонном изменении вероятностей.

Что такое вероятность события

- **Классическое определение** (существует только *n* равновозможных исходов эксперимента, из них *m* исходов приведут к событию А)

$$p(A)=m/n$$

- **Статистическое определение** (в результате проведённых *n* экспериментов события А возникло *m* раз)

$$p(A)=\lim_{n\to\infty}\frac{m}{n}$$

- **Свойства вероятности**

$$0\le p(A)\le 1,$$

сумма вероятностей всех возможных несовместных событий равна 1

Мера количества информации по Хартли

Система S может находиться в одном из N равновероятных состояний. Вероятность каждого из состояний $p = 1/N$. Передадим сообщение о выпавшем состоянии S, используя двоичное сообщение длины d:

$$2^d \geq N \quad \rightarrow \quad d \geq \log_2 N$$

Ральф Хартли
(1880--1970)

Значит, для однозначного описания системы требуется $\log_2 N$ бит. По определению Хартли, количество информации в системе S равно

$$i_{\text{H}}(s) = \log_x N = -\log_x p.$$

Единицы измерения количества информации:

i_{H} = (lb N бит = lb N Шн = lb N Sh) = $\log_3 N$ трит = (lg N харт = lg N Hart = lg N дит) = ln N нат

Какова этимология названий единиц измерения? Сколько дит содержится в 33 битах?

Ответ 1: (bit ⟶ binary digit), (dit ⟶ decimal digit), (Шн ⟶ Шеннон), (харт ⟶ Хартли) и т. д.

Ответ 2: т. к. 33 бит = $\log_2 N$, то $\log_{10} N = x$ дит, отсюда найдём x через N: $x = \log_{10} 2^{33} \approx 9{,}9$ дит.

Пример применения меры Хартли на практике

Пример 1. Ведущий загадывает число от 1 до 64. Какое количество вопросов типа «да-нет» понадобится, чтобы гарантировано угадать число?

- Первый вопрос: «Загаданное число меньше 32?». Ответ: «Да».
- Второй вопрос: «Загаданное число меньше 16?». Ответ: «Нет».
- …
- Шестой вопрос (в худшем случае) точно приведёт к верному ответу.
- Значит, в соответствии с мерой Хартли в загадке ведущего содержится ровно $\log_2 64 = 6$ бит непредсказуемости (т. е. информации).

Пример 2. Ведущий держит за спиной ферзя и собирается поставить его на произвольную клетку доски. Насколько непредсказуемо его решение?

- Всего на доске 8x8 клеток, а цвет ферзя может быть белым или чёрным, т. е. всего возможно 8x8x2 = 128 равновероятных состояний.
- Значит, количество информации по Хартли равно $\log_2 128 = 7$ бит.

Анализ свойств меры Хартли

Экспериментатор одновременно подбрасывает монету (М) и кидает игральную кость (К). Какое количество информации содержится в эксперименте (Э)?

Аддитивность:

i(Э) = i(М) + i(К) => i(12 исходов) = i (2 исхода) + i (6 исходов): $\log_x 12 = \log_x 2 + \log_x 6$

Неотрицательность:

Функция $\log_x N$ неотрицательно при любом $x > 1$ и $N \geq 1$.

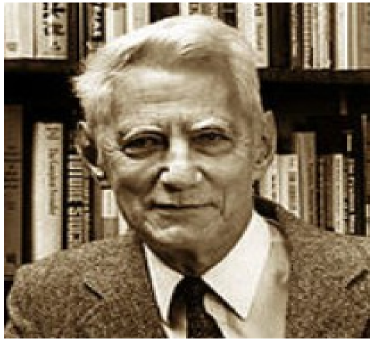
Монотонность:

С увеличением p(М) или p(К) функция i(Э) монотонно возрастает.

Принцип предопределённости:

При наличии всегда только одного исхода (монета и кость с магнитом) количество информации равно нулю: $\log_x 1 + \log_x 1 = 0$.

Мера количества информации по Шеннону



Клод Шеннон
(1916--2001)

Мера Хартли подходит лишь для систем с равновероятными состояниями. Если состояния системы S не равновероятны, используют меру Шеннона:

$$i(S) = -\sum_{i=1}^{N} p_i \cdot \log_2 p_i \,,$$

где N – число состояний системы,

p_i – вероятность того, что система S находится в состоянии i (сумма всех p_i равна 1).

Формула Хартли является частным случаем формулы Шеннона!

Пример 1. Количество информации в акте подбрасывания обычной монеты по формуле Хартли равно $\log_2 2 = 1$ бит. По формуле Шеннона получим то же: $i_{s1} = -0{,}5*\log_2 0{,}5 - 0{,}5*\log_2 0{,}5 = 1$ бит.

Пример 2. При подбрасывании монеты со смещённым центром тяжести количество непредсказуемости становится меньше: $i_{s2} = -0{,}75*\log_2 0{,}75 - 0{,}25*\log_2 0{,}25 \approx 0{,}8$ бит.

Пример использования меры Шеннона

Шулер наугад вытаскивает одну карту из стопки, содержащей 9 известных ему карт: 3 джокера, 3 туза, 1 король, 1 дама и 1 валет. Какое количество информации для шулера содержится в этом событии s?

$$\text{Вероятность вытащить} \left\{ \begin{array}{l} \text{джокера} \\ \text{туза} \\ \text{короля} \\ \text{даму} \\ \text{валета} \end{array} \right\} \text{равна} \left\{ \begin{array}{l} 3/9 = 1/3 \\ 3/9 = 1/3 \\ 1/9 \\ 1/9 \\ 1/9 \end{array} \right.$$

Количество информации, выраженное в тритах, равно:

$$i(s) = -\left(\frac{1}{3}\cdot\log_3\frac{1}{3} + \frac{1}{3}\cdot\log_3\frac{1}{3} + \frac{1}{9}\cdot\log_3\frac{1}{9} + \frac{1}{9}\cdot\log_3\frac{1}{9} + \frac{1}{9}\cdot\log_3\frac{1}{9}\right) =$$

$$= \frac{1}{3} + \frac{1}{3} + \frac{2}{9} + \frac{2}{9} + \frac{2}{9} = 1\frac{1}{3} \approx \log_3 5 \ \text{ vs } \ \log_3 14$$

Нестрогий вывод формулы Шеннона

Задача. Монета имеет смещённый центр тяжести. Вероятность выпадения «орла» – 0,25, вероятность выпадения «решки» – 0,75. Какое количество информации содержится в одном подбрасывании?

Решение

- Пусть монета была подброшена N раз ($N \to \infty$), из которых «решка» выпала K раз, «орёл» — M раз (очевидно, что $N = M + K$).
- Количество информации в N подбрасываниях: $i_N = M*i$(«решка») + $K*i$(«орёл»).
- Тогда среднее количество информации в одном подбрасывании: $i_1 = i_N/N = (M/N)*i$(«решка»)$+(K/N)*i$(«орёл») = p(«решка»)$*i$(«решка»)+p(«орёл»)$*i$(«орёл»).
- Подставив формулу Хартли для i, окончательно получим: i_1 = -p(«решка»)$*\log_x$p(«решка») - p(«орёл»)$*\log_x$p(«орёл») ≈ 0,8 бит.