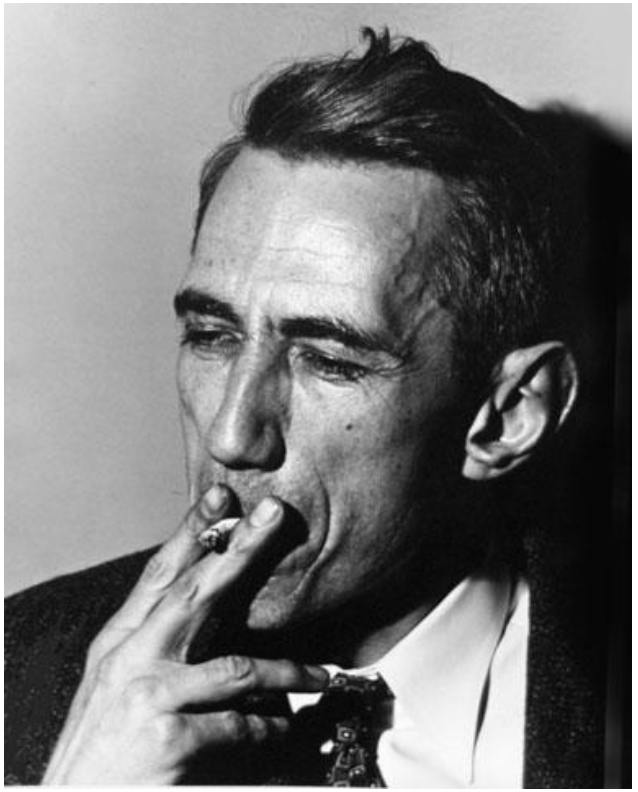
2. Понятие информации в теории Шеннона

2.1 Понятие энтропии

2.2 Энтропия и информация

2.3 Информация и алфавит



Клод Э́лвуд Ше́ннон
30 апреля1916
24 февраля2001

Случайное событие – означает отсутствие полной уверенности в его наступлении.

Пусть опыт имеет *n* равновероятных исходов.

Определение. Функция f(n) - мера неопределенности опыта.

Мера неопределенности является функцией числа исходов f(n).

Свойства функции f(n)

- *f*(1) = 0, если *n* = 1 исход опыта не является случайным и неопределенность отсутствует;
- *f(n)* возрастает с ростом *n*,

чем больше n, тем более затруднительным становится предсказание результата опыта;

Пусть α и β независимые опыты.

nα, $n\beta$ - число равновероятных исходов.

Рассмотрим сложный опыт, который состоит в одновременном выполнении опытов

α и β.

Число возможных его исходов равно:

$$n\alpha \cdot n\beta,$$

причем, все они равновероятны.

$f(n_\alpha \cdot n_\beta)$ - мера неопределенности сложного опыта.

α и β – независимы, т.е. в сложном опыте они не влияют друг на друга.

Следовательно,

$$f(n_\alpha \cdot n_\beta) = f(n_\alpha) + f(n_\beta) \qquad (3)$$

т.е. *мера неопределенности аддитивна.*

- $f(1) = 0$
- $f(n)$ возрастает с ростом n
- $f(n\alpha \cdot n\beta) = f(n\alpha) + f(n\beta)$

Этим свойствам удовлетворяет функция:

$$log(n)$$

можно доказать, что она единственная из всех существующих классов функций.

$$f(n) = log(n) \qquad (4)$$

мера неопределенности опыта, имеющего n равновероятных исходов.

Выбор основания логарифма значения не имеет.

$$\log_b n = \log_b a \cdot \log_a n ,$$

переход к другому основанию состоит во введении одинакового для обеих частей выражения постоянного множителя

$\log_b a$

Удобно, основание *2.*

За единицу измерения принимается неопределенность, содержащаяся в опыте, имеющем лишь два равновероятных исхода,

ИСТИНА *(True)* и ЛОЖЬ *(False)*.

$$f(2)=\log_2 2=1 \text{ бит.}$$

Определение. Единица измерения неопределенности при двух возможных равновероятных исходах опыта называется **бит.**

Название **бит** происходит от английского **binary digit,** «двоичный разряд» или «двоичная единица».

Определение. Мера неопределенности опыта, имеющего *n* равновероятных исходов равна

$$f(n)=\log_2(n). \qquad (4.1)$$

Эта величина – ***энтропия***, обозначается *H.*

Рассмотрим опыт с *n* равновероятными исходами.

Неопределенность, вносимую одним исходом?

$$\frac{1}{n}\log_2 n = -\frac{1}{n}\log_2 \frac{1}{n} = -p \cdot \log_2 p\,,$$

где $p = 1/n$ - вероятность любого из отдельных исходов.

Пусть исходы *неравновероятны,*
p(A1) и *p(A2)* – *вероятности исходов.*

$$H_1 = -p(A_1)\cdot\log_2 p(A_1) \text{ и } H_2 = -p(A_2)\cdot\log_2 p(A_2),$$

$$H = H_1 + H_2 = -p(A_1)\cdot\log_2 p(A_1) - p(A_2)\cdot\log_2 p(A_2).$$

Если опыт α имеет *n* неравновероятных исходов *A*1, *A*2... *An,* тогда:

$$H(\alpha) = -\sum_{i=1}^{n} p(A_i)\cdot\log_2 p(A_i). \qquad (5)$$

Используя формулу для среднего значения дискретных случайных величин:

$$< x >= \sum_{j=1}^{k} p_j x_j \ .$$

$$H(\alpha) = < -\log_2 p(A^{(\alpha)}) >,$$

A(α) - обозначает исходы, возможные в опыте α.

Определение. Энтропия *является мерой неопределенности опыта, в котором проявляются случайные события, и равна средней неопределенности всех возможных его исходов.*

ПРИМЕР

Имеются два ящика, в каждом из которых по 12 шаров. В первом - 3 белых, 3 черных и 6 красных; во втором - каждого цвета по 4. Опыты состоят в вытаскивании по одному шару из каждого ящика. Что можно сказать относительно неопределенностей исходов этих опытов?

Решаем :-).

$$H_\alpha = -\frac{3}{12}\log_2\frac{3}{12} - \frac{3}{12}\log_2\frac{3}{12} - \frac{6}{12}\log_2\frac{6}{12} = 1{,}50 \text{ бит},$$

$$H_\beta = -\frac{4}{12}\log_2\frac{4}{12} - \frac{4}{12}\log_2\frac{4}{12} - \frac{4}{12}\log_2\frac{4}{12} = 1{,}58 \text{ бит},$$

$H\beta > H\alpha$, т.е. неопределенность результата в опыте β выше и, следовательно, предсказать его можно с меньшей долей уверенности, чем результат опыта α.

Свойства энтропии

- 1) $H > 0$.

$H = 0$ в двух случаях:

(a) если $p(Aj) = 1$; т.е. один из исходов является *достоверным* (и общий итог опыта перестает быть случайным);

(b) все $p(Ai) = 0$, т.е. никакие из рассматриваемых исходов опыта невозможны.

2) Для двух *независимых* опытов α и β

$$H(\alpha \wedge \beta) = H(\alpha) + H(\beta)$$

Энтропия сложного опыта, состоящего из нескольких независимых, равна сумме энтропии отдельных опытов.

3) При прочих равных условиях наибольшую энтропию имеет опыт с равновероятными исходами.

$$-\sum_{i=1}^{n} p(A_i) \cdot \log_2 p(A_i) \leq \log_2 n . \qquad (6)$$

Условная энтропия

Найдем энтропию сложного опыта α ^ β (опыты не являются независимыми, на исход β оказывает влияние результат опыта α).

Пример, если в ящике два разноцветных шара и α – извлечение первого, а β - второго, тогда α полностью снимает неопределенность сложного опыта α ^ β,

т.е. $H(\alpha \wedge \beta) = H(\alpha)$,

а не сумме энтропий.

$$p(A_i \wedge B_j) = p(A_i) \cdot p_{A_i}(B_j),$$

$$\log_2 p(A_i \wedge B_j) = \log_2 p(A_i) + \log_2 p_{A_i}(B_j).$$

Подставим в (7)

$$H(\alpha \wedge \beta) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(A_i) p_{A_i}(B_j) \cdot \{\log_2 p(A_i) + \log_2 p_{A_i}(B_j)\} =$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(A_i) p_{A_i}(B_j) \cdot \log_2 p(A_i) - \sum_{i=1}^{n}\sum_{j=1}^{m} p(A_i) p_{A_i}(B_j) \cdot \log_2 p_{A_i}(B_j).$$

- В первом слагаемом индекс *j* имеется только у *B*; изменив порядок суммирования, получим члены вида:

$$\sum_{j=1}^{m} p_{A_i}(B_j).$$

$$\sum_{j=1}^{m} p_{A_i}(B_j) = p_{A_i}\left(\sum_{j=1}^{m} B_j\right) = 1$$

$$\sum_{j=1}^{m} B_j$$

$$-\sum_{i=1}^{n} p(A_i) \cdot \log_2 p(A_i) = H(\alpha)$$

Свойства условной энтропии

1. Условная энтропия является величиной *неотрицательной.*

   $H_\alpha(\beta) = 0$, если *любой* исход α полностью определяет исход β

2. Если α и β независимы, то $H_\alpha(\beta) = H(\beta)$, причем это оказывается *наибольшим* значением условной энтропии, т.е. *условная энтропия не превосходит безусловную.*

Пример 2.2.

В ящике имеются 2 белых шара и 4 черных. Из ящика извлекают последовательно два шара без возврата. Найти энтропию, связанную с первым и вторым извлечениями, а также энтропию обоих извлечений.

Задача 2.3.

- Имеется три тела с одинаковыми внешними размерами, но с разными массами *x*1, *x*2 и *x*3. Необходимо определить энтропию, связанную с нахождением наиболее тяжелого из них, если сравнивать веса тел можно только попарно.

2.2. Энтропия и информация

- ***Определение.*** *I - информацией относительно опыта β, содержащейся в опыте α*

$$I(\alpha,\beta)=H(\beta)-H_\beta(\alpha)$$

- Следствие 1. Единицы измерения количество информации – бит.
- Следствие 2. Пусть опыт α = β, тогда

$$H_\beta(\beta) = 0 \qquad \text{(свойство усл. энтропии)}$$

$$I(\beta,\beta) = H(\beta).$$

Определение. Энтропия опыта равна той информации, которую получаем в результате его осуществления.

Свойств информации:

1. $I(\alpha, \beta) \geq 0$, причем $I(\alpha, \beta) = 0$ тогда и только тогда, когда опыты α и β независимы.
2. $I(\alpha, \beta) = I(\beta, \alpha)$, т.е. информация симметрична относительно последовательности опытов.
3. *Информация опыта равна:*

$$I = -\sum_{i=1}^{n} p(A_i) \cdot \log_2 p(A_i)$$

Пример 2.4. Какое количество информации требуется, чтобы узнать исход броска монеты?

Пример 2.5. Виктор Сергеевич задумал «оценку» (целое число в интервале от 2 до 5). Опыт состоит в угадывании этого числа. На вопросы В.С. отвечает лишь *«Да»* или *«Нет».* Какое количество информации должны получить, чтобы узнать задуманную оценку? Как правильно построить процесс угадывания?

Количество информации численно равно числу вопросов с равновероятными бинарными вариантами ответов, которые необходимо задать, чтобы полностью снять неопределенность задачи.

Если все *n* исходов равновероятны

$$p(A_i) = \frac{1}{n} \qquad I = \sum_{i=1}^{n} \frac{1}{n} \cdot \log_2 n = \log_2 n$$

Формула Хартли (1928).

30.11.1888 - 1.05.1970 (81).
США

$$I = \log_2 n .$$

Связывает *количество равновероятных состояний (n) и (/), что любое из этих состояний реализовалось.*

Частным случаем применения формулы Хартли является ситуация, когда $n = 2^k$.

$$I = k \text{ бит.}$$

k равно количеству вопросов с бинарными равновероятными ответами, которые определяют количество информации.

- **Пример 2.6** В.С. **с**лучайным образом вынимает карта из колоды в 32 карты. Какое количество информации требуется, чтобы угадать, что это за карта? Как построить угадывание?

- **Пример 2.7** В некоторой местности имеются две близкорасположенные деревни: А и В. Жители А всегда говорят правду, а жители В - всегда лгут. Жители обеих деревень любят ходить друг к другу в гости, поэтому в каждой из деревень можно встретить жителя соседней деревни. Путешественник, оказался в одной из двух деревень и, заговорив с первым встречным, захотел выяснить, в какой деревне он находится и откуда его собеседник. Какое минимальное количество вопросов с бинарными ответами требуется задать путешественнику?

Выводы

1. Выражение $$I = -\sum_{i=1}^{n} p(A_i) \cdot \log_2 p(A_i)$$

является *статистическим* определением понятия *«информация»*, поскольку в него входят вероятности возможных исходов опыта.

Операционное определение новой величины, т.е. устанавливается процедура (способ) ее *измерения*.

2. Если начальная энтропия опыта H_1, а в результате сообщения информации I энтропия становится равной H_2 ($H_1 \geq H_2$), то

$$I = H_1 - H_2,$$

т.е. *информация равна убыли энтропии.*

Если изначально равновероятных исходов было n1, а в результате передачи информации ***I*** неопределенность уменьшилась, и число исходов стало *n*2 то

$$I = \log_2 n_1 - \log_2 n_2 = \log_2 \frac{n_1}{n_2}.$$

Определение. Информация *- это содержание сообщения, понижающего неопределенность некоторого опыта с неоднозначным исходом; убыль связанной с ним энтропии является количественной мерой информации.*

В случае равновероятных исходов информация равна логарифму отношения числа возможных исходов до и после (получения сообщения).

3. Аддитивность информации.

Пусть $I_A = \log_2 n_A$ *- первого опыта,*

$I_B = \log_2 n_B$ *- второго опыта,*

второй выбор никак не связан с первым.

При объединении число возможных состояний (элементов) будет

$n = n_A \cdot n_B$ *и* потребуется количество информации:

$$I = \log_2 n = \log_2 n_A \cdot n_B = \log_2 n_A + \log_2 n_B = I_A + I_B.$$

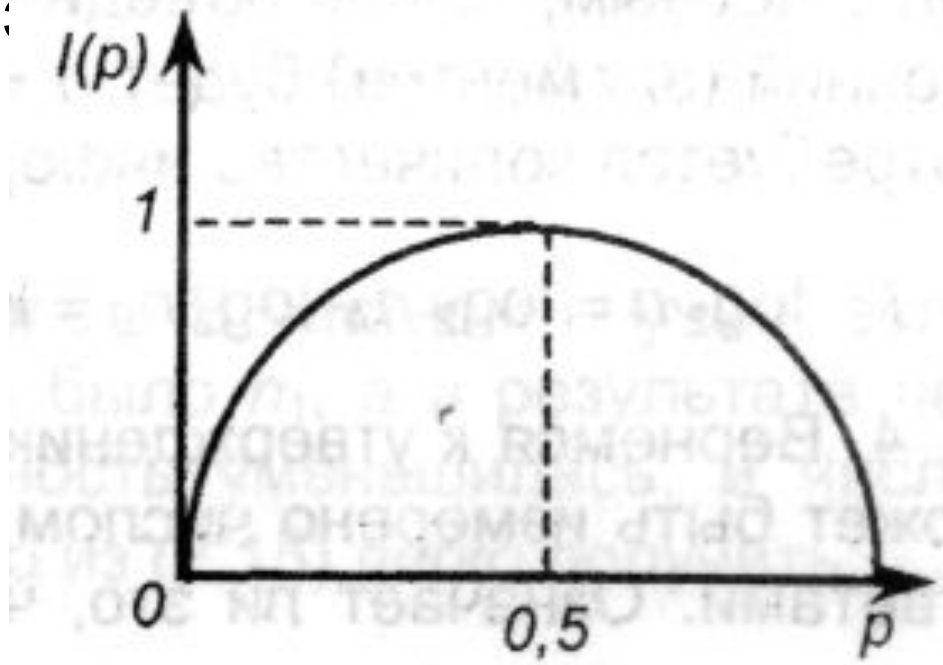
4. Рассмотрим опыт, реализующийся посредством двух случайных событий; если эти события равновероятны,

$p1 = p2 = 1/2$, и $I = 1$ бит, как следует из формулы Хартли.

Если их вероятности различны: $p1 = p$, то, $p2 = 1 - p$, и следовательно:

$$I(p) = -p \cdot \log_2 p - (1-p) \cdot \log_2 (1-p)$$

График



Ответ на бинарный вопрос может содержать не более 1 бит информации; информация равна 1 бит только для равновероятных ответов; в остальных случаях она меньше 1 бит.

Пример 2.8. При угадывании результата броска игральной кости задается вопрос *«Выпало 6?».* Какое количество информации содержит ответ?

На бытовом уровне, *«информация»* отождествляется с *«информированностью»,* т.е. человеческим *знанием.*

В «теории информации» информация является мерой нашего незнания чего-либо (но что в принципе может произойти); как только это происходит и узнаем результат, информация, связанная с данным событием, исчезает. Состоявшееся событие не несет информации, поскольку пропадает его неопределенность (энтропия становится равной нулю), и $I = 0$.

Глава 3. Кодирование символьной информации

- 3.1. Постановка задачи кодирования. Первая теорема Шеннона
- 3.2. Способы построения двоичных кодов

3.1. Постановка задачи кодирования. Первая теорема Шеннона

Код

(1) правило, описывающее соответствие знаков или их сочетаний первичного алфавита знакам или их сочетаниям вторичного алфавита.

(2) набор знаков вторичного алфавита, используемый для представления знаков или их сочетаний первичного алфавита.

Кодирование *- перевод информации, представленной сообщением в первичном алфавите, в последовательность кодов.*

- ***Декодирование*** *- операция, обратная кодированию, т.е. восстановление информации в первичном алфавите по полученной последовательности кодов.*
- ***Кодер*** *- устройство, обеспечивающее выполнение операции кодирования.*
- ***Декодер*** *- устройство, производящее декодирование.*
- *Операции кодирования и декодирования называются* ***обратимыми,*** *если их последовательное применение обеспечивает возврат к исходной информации без каких-либо ее потерь.*

Источник представляет информацию в форме дискретного сообщения, используя для этого алфавит - *первичным.*

Далее сообщение попадает в устройство, преобразующее и представляющее его в другом алфавите - *вторичным.*

Математическая постановка задачи кодировании. Пусть первичный алфавит

A состоит из ***N*** знаков со средней информацией на знак **I*(A),***

вторичный алфавит ***B*** - из ***M*** знаков со средней информацией на знак ***I(B).***

Пусть исходное сообщение, содержит **n** знаков, а закодированное сообщение - ***m*** знаков.

Если исходное сообщение содержит ***Ist(A)*** информации, а закодированное - ***Ifin(B),*** то условие обратимости кодирования:

$$I_{st}(A) \leq I_{fin}(B),$$

Операция обратимого кодирования может увеличить количество информации в сообщении, но не может его уменьшить.

$$n \cdot I^{(A)} \leq m \cdot I^{(B)}$$

$$I^{(A)} \leq \frac{m}{n} \cdot I^{(B)}.$$

m/n *- характеризует среднее число знаков вторичного алфавита, которое приходится использовать для кодирования одного знака первичного алфавита.*

Это - длина кода - ***K(A,B).***

$$K(A,B) \geq \frac{l^{(A)}}{l^{(B)}}.$$

Обычно $N > M$ и $I(A) > I(B)$, откуда $K(A,B) > 1$, т.е. один знак первичного алфавита представляется несколькими знаками вторичного.

Способов построения кодов при фиксированных алфавитах A и B множество, возникает проблема выбора (или построения) наилучшего варианта - *оптимального кода.*

Минимально возможным значением средней длины кода

$$K^{\min}(A,B) = \frac{l^{(A)}}{l^{(B)}}.$$

устанавливающее нижний предел длины кода.

Первая теорема Шеннона (*основная теорема о кодировании при отсутствии помех*).

1. При отсутствии помех всегда возможен такой вариант кодирования сообщения, при котором среднее число знаков кода, приходящихся на один знак первичного алфавита, будет сколь угодно близко к отношению средних информации на знак первичного и вторичного алфавитов.

Смысл теоремы: теорема открывает принципиальную возможность оптимального кодирования, т.е. построения кода со средней длиной $K\min(A,B)$.

Два пути сокращения *K*min*(A,B):*

1. уменьшение числителя - если при кодировании учесть различие частот появления разных знаков в сообщении, корреляции (двухбуквенные, трехбуквенные и т.д.)
2. увеличение знаменателя - найти такой способ кодирования, при котором появление знаков вторичного алфавита было бы равновероятным, т.е.

$$I(B) = \log_2 M.$$

Для первого приближения

$$K^{\min}(A,B) = \frac{l_1^{(A)}}{\log_2 M}.$$

Относительная избыточность кода:

$$Q(A,B) = \frac{K(A,B) - K^{\min}(A,B)}{K^{\min}(A,B)} = \frac{K(A,B)}{K^{\min}(A,B)} - 1 = \frac{K(A,B)\cdot l^{(B)}}{l^{(A)}} - 1.$$

Данная величина показывает, насколько операция кодирования увеличила длину исходного сообщения.

$Q(A,B) \rightarrow 0$ при $K(A,B) \rightarrow K\min(A,B)$.

Теорема Шеннона:

2. При отсутствии помех всегда возможен такой вариант кодирования сообщения, при котором избыточность кода будет сколь угодно близкой к нулю.

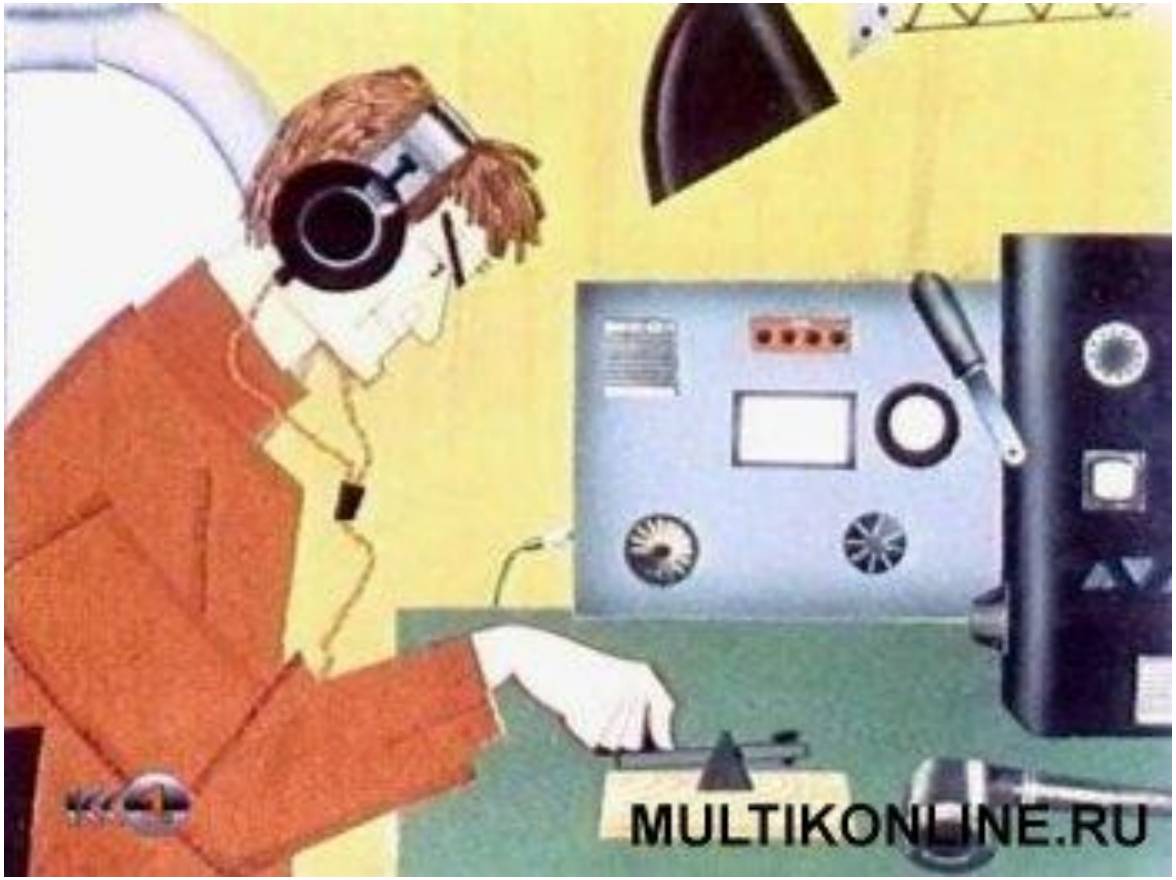
Наиболее важной для практики оказывается ситуация, когда $M = 2$, т.е. для представления кодов в линии связи используется лишь два типа сигналов.

Тогда $\log2(M) = 1$, и $K^{min}(A,2) = I_1^{(A)}$,

3. При отсутствии помех средняя длина двоичного кода может быть сколь угодно близкой к средней информации, приходящейся на знак первичного алфавита.

$$Q(A,2) = \frac{K(A,2)}{I_1^{(A)}} - 1.$$

3.2. Способы построения двоичных кодов



Возможны следующие особенности вторичного алфавита:

- элементарные сигналы (0 и 1) могут иметь одинаковые длительности (to=t1) или разные (to≠t1);
- длина кода может быть одинаковой для всех знаков первичного алфавита – ***равномерный код***
- коды разных знаков первичного алфавита могут иметь различную длину - ***неравномерный*** **код.**
- коды могут строиться для отдельного знака первичного алфавита *(**алфавитное кодирование**)* или для их комбинаций *(**кодирование блоков, слов**).*

3.2.1. Алфавитное неравномерное двоичное кодирование сигналами равной длительности. Префиксные коды

- знаки первичного алфавита (например, русского) кодируются комбинациями символов двоичного алфавита (т.е. 0 и 1).
- длина кодов и, соответственно, длительность передачи отдельного кода, могут различаться.
- Длительности элементарных сигналов при этом одинаковы (to = t1 = t).

Для передачи информации, в среднем приходящейся на знак первичного алфавита, необходимо время

$$T = K(A,2) \cdot t$$

Построить такую схему кодирования, в которой суммарная длительность кодов при передаче данного сообщения была бы наименьшей.

Решение: коды знаков первичного алфавита, вероятность появления которых в сообщении выше, следует строить из возможно меньшего числа элементарных сигналов, а длинные коды использовать для знаков с малыми вероятностями.

Пусть получен код

0010001000011101010111000011 0

Каким образом он может быть декодирован?

А) Неравномерный код с разделителем

Разделителем отдельных кодов букв будет последовательность *00.*

Разделителем слов-слов – *000.*

Правила построения кодов:

- код признака конца знака может быть включен в код буквы (т.е. коды всех букв будут заканчиваться *00*);
- коды букв не должны содержать двух и более нулей подряд в середине. (иначе они будут восприниматься как конец знака);
- код буквы (кроме пробела) всегда должен начинаться с *1;*
- разделителю слов (*000*) всегда предшествует признак конца знака; при этом реализуется последовательность *00000.*

Буква	Код	$p_i \cdot 10^3$	k_i	Буква	Код	$p_i \cdot 10^3$	k_i
пробел	*000*	174	3	я	*1011000*	18	7
о	*100*	90	3	ы	*1011100*	16	7
е	*1000*	72	4	з	*1101000*	16	7
а	*1100*	62	4	ь,ъ	*1101100*	14	7
и	*10000*	62	5	б	*1110000*	14	7
т	*10100*	53	5	г	*1110100*	13	7
н	*11000*	53	5	ч	*1111000*	12	7
с	*11100*	45	5	й	*1111100*	10	7
р	*101000*	40	6	х	*10101000*	9	8
в	*101100*	38	6	ж	*10101100*	7	8
л	*110000*	35	6	ю	*10110000*	6	8
к	*110100*	28	6	ш	*10110100*	6	8
м	*111000*	26	6	ц	*10111000*	4	8
д	*111100*	25	6	щ	*10111100*	3	8
п	*1010000*	23	7	э	*11010000*	3	8
у	*1010100*	21	7	ф	*11010100*	2	8

Среднюю длину кода $K(r,2)$ для данного способа кодирования:

$$K(r,2) = \sum_{j=1}^{32} p_j \cdot k_j = 4{,}964.$$

(по определению средней дискретной величины).

$I1(r)$ = 4,356 бит.

Избыточность данного кода:

$$Q(r,2) = 4{,}964 / 4{,}356 - 1 \approx 0{,}14 ,$$

При данном способе кодирования будет передаваться приблизительно на 14% больше информации, чем содержит исходное сообщение.

Рассмотрев один из вариантов двоичного неравномерного кодирования, возникают вопросы:

1) Возможно ли такое кодирование без использования разделителя знаков?

2) Существует ли наиболее эффективный (оптимальный) способ неравномерного двоичного кодирования?

Неравномерный код может быть однозначно декодирован, если никакой из кодов не совпадает с началом (префиксом) какого-либо иного более длинного кода – условие ФАНО.

Например, если имеется код *110,* то уже не могут использоваться коды *1, 11, 1101, 110101.*

Пример 3.1.

Пусть имеется следующая таблица префиксных кодов:

а	л	м	Р	у	ы
10	*010*	*00*	*11*	*0110*	*0111*

Требуется декодировать сообщение:

00100010000111010101110000110

Декодирование производится циклически

1. отрезать от текущего сообщения крайний левый символ, присоединить справа к рабочему кодовому слову;
2. сравнить рабочее кодовое слово с кодовой таблицей; если совпадения нет, перейти к 1.
3. декодировать рабочее кодовое слово, очистить его;
4. проверить, имеются ли еще знаки в сообщении; если «да», перейти к 1.

Шаг	Рабочее слово	Текущее сообщение	Распознанный знак	Декодированное сообщение
0	Пусто	**0**0100010000111010101110000110	–	–
1	*0*	**0**1000100001110101011100001 10	нет	–
2	***00***	**1**0001000011101010111000011 0	***м***	*м*
3	*1*	**0**0010000111010101110000110	нет	*м*
4	***10***	**0**0100001110101011100001 10	***а***	*ма*
5	*0*	**0**1000011101010111000011 0	нет	*ма*
6	***00***	**1**0000111010101110000110	***м***	*мам*
...				

а	л	м	Р	у	ы
10	*010*	*00*	*11*	*0110*	*0111*

Таким образом, использование префиксного кодирования позволяет делать сообщение более коротким.

Условие Фано не устанавливает способа формирования префиксного кода и, в частности, наилучшего из возможных.

В) Префиксный код Шеннона-Фано.

Данный вариант кодирования был предложен в 1948-1949 гг. независимо Р. Фано и К. Шенноном.

Пусть имеется первичный алфавит ***A***, состоящий из шести знаков ***a*1 ...*a*6** с вероятностями появления в сообщении, соответственно: **0,3; 0,2; 0,2; 0,15; 0,1; 0,05.**

Расположим эти знаки в таблице в порядке убывания вероятностей.

Разделим знаки на две группы таким образом, чтобы суммы вероятностей в каждой из них были бы приблизительно равными. Затем будем делить следующие группы.

Знак	p_i	Разряды кода				Код
		1	2	3	4	
a_1	0,30	0	0			00
a_2	0,20	0	1			01
a_3	0,20	1	0			10
a_4	0,15	1	1	0		110
a_5	0,10	1	1	1	0	1110
a_6	0,05	1	1	1	1	1111

Средняя длина кода равна:

$$K(A,2) = 0{,}3 \cdot 2 + 0{,}2 \cdot 2 + 0{,}2 \cdot 2 + 0{,}15 \cdot 3 + 0{,}1 \cdot 4 + 0{,}05 \cdot 4 = 2{,}45$$

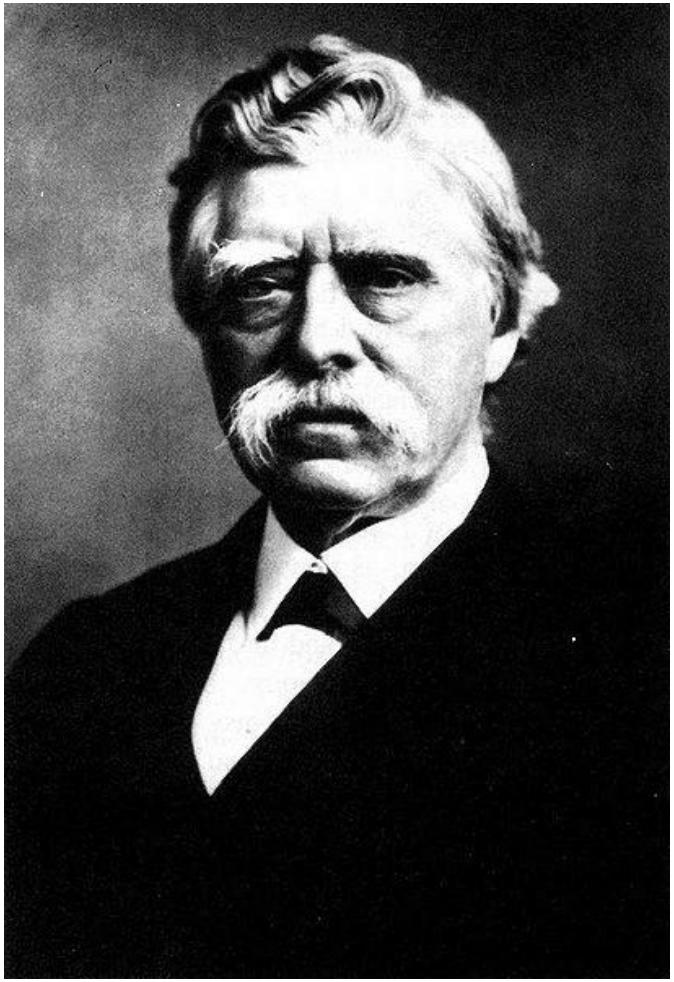
$I1(A)$ = 2,390 бит.

избыточность кода $Q(A,2)$ = 0,0249, т.е. около 2,5%.

Данный код нельзя считать оптимальным, поскольку вероятности появления 0 и 1 неодинаковы (6/17=0,35 и 11/17=0,65, соответственно).

Для русского алфавита избыточность кода Q=0,0147.

C) Префиксный код Хаффмана



Хаффман, Дэвид

David Albert Huffman

Дата рождения:	9 августа9 августа 1925
Место рождения:	АльянсАльянс, Огайо
Дата смерти:	7 октября7 октября 1999 (74 года)
Страна:	США
Научная сфера:	Теория информацииТеория информации, Алгоритмы

Способ *оптимального* префиксного двоичного кодирования был предложен Д. Хаффманом. Построение кодов Хаффмана рассмотрим на том же примере.

Создадим новый вспомогательный алфавит *A*1, объединив два знака с наименьшими вероятностями (*a*5 и *a*6) и заменив их одним знаком (например, *a*(1)); его вероятность будет равна сумме вероятностей т.е. 0,15; остальные знаки исходного алфавита включим в новый без изменений; общее число знаков в новом алфавите, очевидно, будет на 1 меньше, чем в исходном.

Аналогично продолжим создавать новые алфавиты, пока в последнем не останется два знака.

Количество таких шагов будет равно ***N* - 2**, где ***N*** - число знаков исходного алфавита (*N*=6, необходимо построить 4 вспомогательных алфавита).

В промежуточных алфавитах каждый раз будем переупорядочивать знаки по убыванию вероятностей.

№ знака	Вероятности				
	Исходный алфавит	Промежуточные алфавиты			
		$A^{(1)}$	$A^{(2)}$	$A^{(3)}$	$A^{(4)}$
1	0,3	0,3	0,3	0,4	0,6
2	0,2	0,2	0,3	0,3	0,4
3	0,2	0,2	0,2	0,3	
4	0,15	0,15	0,2		
5	0,1	0,15			
6	0,05				

Теперь в обратном направлении проведем процедуру кодирования.

№ знака	Вероятности									
	Исходный алфавит		Промежуточные алфавиты							
			$A^{(1)}$		$A^{(2)}$		$A^{(3)}$		$A^{(4)}$	
1	*0,3*	**00**	*0,3*	**00**	*0,3*	**00**	*0,4*	**1**	*0,6*	**0**
2	*0,2*	**10**	*0,2*	**10**	*0,3*	**01**	*0,3*	**00**	*0,4*	**1**
3	*0,2*	**11**	*0,2*	**11**	*0,2*	**10**	*0,3*	**01**		
4	*0,15*	**010**	*0,15*	**010**	*0,2*	**11**				
5	*0,1*	**0110**	*0,15*	**011**						
6	*0,05*	**0111**								

$K(A,2)$ = 0,3 · 2 + 0,2 · 2 + 0,2 · 2 +0,15 · 3 + 0,1 · 4 + 0,05 · 4 = 2,45.

$Q(A, 2)$ = 0,0249, однако, вероятности 0 и 1 сблизились (0,47 и 0,53, соответственно).

$K(r,2)$ = 4,395; избыточность кода $Q(r,2)$ = 0,0090, т.е. не превышает 1 %, что заметно меньше избыточности кода Шеннона-Фано.

Код Хаффмана важен в теоретическом отношении, поскольку можно доказать, что он является ***самым экономичным из всех возможных,*** т.е. *ни для какого метода алфавитного кодирования* ***длина кода не может оказаться меньше, чем код Хаффмана*****.**

Метод Хаффмана и его модификация - метод адаптивного кодирования *(динамическое кодирование Хаффмана)* - нашли широчайшее применение в **программах-архиваторах, программах резервного копирования файлов и дисков, в системах сжатия информации в модемах и факсах.**

3.2.2. Равномерное алфавитное двоичное кодирование. Байтовый код