ЛЕКЦИЯ №2 ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ Ч.1

Преподаватель: Оцоков Шамиль Алиевич

Москва, 2021 г.

Ряды Фурье

Если сигнал является чётной функцией , то для его разложения Иногда удобно пользоваться синусно-косинусной формой ряда Фурье

– в ней будет присутствовать только косинусные (чётные) слагаемые

$$a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos mx \, dx$$

$$(m = 1, 2, 3, \ldots).$$

$$b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin mx \, dx$$

$$(m = 1, 2, 3, \ldots).$$

Случайные сигналы и их характеристики

В отличие от детерминированных сигналов, значения которых в любой момент точно определены их параметрами, случайные сигналы представляют собой функции времени, мгновенные значения которых могут быть предсказаны лишь с некоторой долей вероятности.

В радиотехнике и теории связи имеют дело с двумя классами случайных сигналов. Во-первых, это *помехи или шумы* - любые случайные воздействия на сигнал, которые ухудшают верность воспроизведения передаваемых сообщений, то есть ухудшает качество связи. По своей физической природе помехи могут быть весьма разнообразны, общим же для них является то, что они носят случайный характер и их значения не могут быть предсказаны.

Во-вторых, случайными являются все сигналы, несущие информацию – речь, музыка, видео и телевизионные изображения, символы передаваемых текстов и т.д. Для их описания также должны использоваться статистические модели и характеристики.

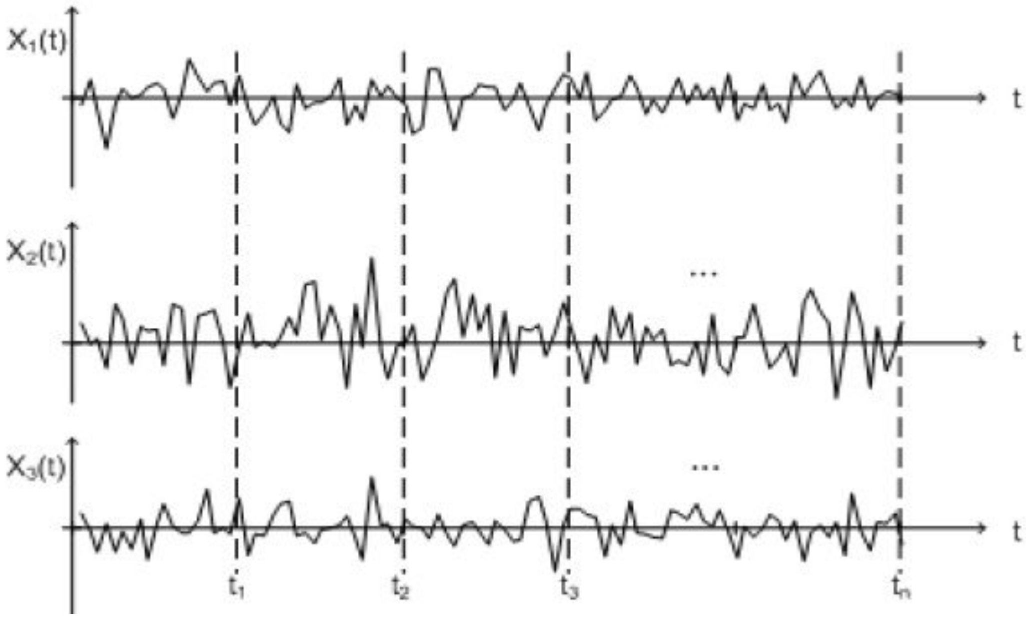
Терминология

Случайный процесс

Математическая модель изменяющегося во времени случайного сигнала называется *случайным процессом*. Случайный процесс *$X(t)$ – это функция,котрая случайным образом может принимать различный вид и значения которой в любой момент времени t, являются случайными величинами.* До регистрации случайный процесс $X(t)$ рассматривается как совокупность (*ансамбль*) функций $x_i(t)$, подчиняющийся общим для них статистическим закономерностям. Одна из этих случайных функций, реализовавшаяся в ходе приема, называется *реализацией случайного процесса*.

Случайный процесс

Пусть $X(t)$ – случайный процесс, заданный ансамблем своих возможных реализаций $\{x_1(t), x_2(t), \ldots x_k(t), \ldots\}$



Статистические характеристики случайного

Выбрав произвольный момент времени t_1 зафиксируем значения реализаций случайного процесса в этот момент времени $\{x_1(t_1),\ x_2(t_1),\ \dots x_k(t_1), \dots\}$. Совокупность этих значений дает нам *одномерное сечение случайного процесса* и представляет собой векторную случайную величину $X(t_1)$.

Эта случайная величина описывается следующими основными статистическими характеристиками:

- *Функция распределения вероятностей.* Обозначается как $F(x,t_1)$ и представляет собой *вероятность того, что в момент t_1 значения случайного процесса X не превышают величины x*

$$F(x) = P(X(t_1) \le x)$$

Функция распределения вероятностей – величина положительная, не убывающая, и принимающая значения $0 \le F(x) \le 1$. На границах интервла функция $F(x)$ принимает значения: $F(-\infty) = 0$, $F(\infty) = 1$. Вероятность попадания случайного процесса в некоторый интервал *(a,b) равна разности значений функции распределения на концах этого интервала*

$$P(a < X(t_1) \le b) = F(b) - F(a)$$

Статистические характеристики случайного процесса

- *Одномерная плотность вероятности.* Обозначается как $p(x,t_1)$ и представляет собой *вероятность попадания значений случайной величины $X(t_1)$ в интервал dx в окрестности x*

$$p(x,t_1)dx = P\{ x \le X(t_1) < x + dx \}$$

Плотность вероятности $p(x)$ – *неотрицательная функция* (как производная от неубывающей функции $F(x)$). *Вероятность попадания* случайного процесса в некоторый интервал (a,b) определяется, как

$$P(a < X(t_1) \le b) = \int_a^b p(x,t_1)dx$$

Поскольку случайная величина $X(t_1)$ должна принять какое либо значение, то есть вероятность ее попадания в интервал от $-\infty$ до ∞ равна единице, *функция плотности вероятности должна удовлетворять следующему условию (условие нормировки):*

$$\int_{-\infty}^{\infty} p(x,t_1)dx = P(-\infty < X(t_1) < \infty) = 1$$

Числовые характеристики случайных процессов

Наряду с функциональными характеристика случайных процессов, какими являются *функция распределения* вероятностей и *плотность вероятностей*, для их описания часто используют простейшие числовые характеристики. К ним относятся *математическое ожидание случайного процесса* $m_x(t)$ и его *дисперсия* $D_x(t)$.

Математическое ожидание – оценка среднего значения случайного процесса
Дисперсия – средняя мощность отклонения случайного процесса от его математического ожидания
Среднеквадратичное отклонение – корень квадратный из дисперсии

Источники сообщений

Источники непрерывных (аналоговых) сообщений – речь, музыка, данные измерений различных параметров физических объектов или явлений (температура, давление, скорость, напряжения, токи и т.д.), неподвижные изображения (фото, чертежи, схемы)), подвижные изображения (кино, видео) и т.д..

Первичные аналоговые сообщения - речь, музыка, изображения, параметры окружающей среды и т.д. – чаще всего представляют собой функции времени - $\lambda(t)$ или других аргументов - $\lambda(x, y)$ неэлектрической природы (акустическое давление, температура, распределение яркости т.п.). Для передачи по каналу связи эти, различные по своему характеру и физической природе сообщения *преобразуются в электрический сигнал*, изменения которого во времени – функция $\lambda(t)$ отображает передаваемую информацию. При этом, как аргумент этой функции, так и сама функция $\lambda(t)$ могут принимать *любые значения* из непрерывного интервала как по λ, так и по t :

$$\lambda \in (\lambda min, \lambda max), t \in (0, t),$$

Источники сообщений

Источники дискретных (цифровых) сообщений - тексты, цифровые данные, программы для ЭВМ, результаты цифровых измерений различных параметров, цифровые фотографии, текстовые, графические или иные файлы и т.п..

Можно выделить несколько разновидностей дискретных источников:

а) Если алфавит сообщения $A\ (\lambda_i)$ (то есть те элементы, из которых составлено сообщение, например – буквы, десятичные числа) представляют собой дискретное конечное множество

$$\lambda_i = \lambda_1,\ \lambda_2,\ \ldots,\ \lambda_k\ ,\ i = 1\ ,\ K$$

такие сообщения называются *дискретными по уровню или квантованными*.

Источники сообщений

б) Если сообщение принимает значения из непрерывного множества, но передается лишь в дискретные моменты времени

$$\lambda \in (\lambda min, \lambda max), \quad t = t_1, t_2, \ldots, t_m, i = 1, M,$$

такое сообщение называется *дискретным по времени.*

в) Наконец, если сообщение *состоит из дискретных элементов* $\lambda_i = \lambda_1, \lambda_2, \ldots, \lambda_k$, $i = 1, K$, которые *передаются только в дискретные моменты времени* $t = t_1, t_2, \ldots, t_m$, $i = 1, M$, такое сообщение называется *дискретным по времени и по уровню* или цифровым.

Таким образом, при всем разнообразии реальных сообщений (источников) большинство из них может быть отнесено к следующим видам:

- непрерывные по времени (*аналоговые*) сообщения;
- дискретные по времени (*дискретизованные*) сообщения;
- дискретные по уровню (*квантованные*) сообщения;
- дискретные по времени и уровню (*цифровые*) сообщения.

Цифровое представление

Таким универсальным представлением данных различных по характеру источников является их *цифровое* представление. Оказывается, любое сообщение можно (при соблюдении определенных условий) преобразовать в дискретное по времени и уровню (*дискретизованное и квантованное сообщение*), а затем абсолютно точно (или с любой необходимой точностью) восстановить в исходном виде. Иными словами, если Вам нужно передать с использованием системы связи, к примеру, фотографию, то ее можно преобразовать в последовательность чисел, передать эти числа, а затем по этим числам снова восстановить исходную фотографию. Точно таким же образом, можно передать речевое сообщение, видео, или текстовый файл.

Теорема Котельникова

Исключительно важным положением теории связи, на котором основана вся современная радиотехника, является так называемая теорема отсчетов, или теорема Котельникова. Эта теорема позволяет установить соотношение между *непрерывными сигналами - сообщениями*, какими являются большинство реальных информационных сигналов *и значениями этих сигналов лишь в отдельные моменты времени – так называемыми отсчетами.*

Теорема Котельникова формулируется следующим образом:

- *непрерывная функция $X(t)$ с ограниченным спектром, то есть не имеющая в своем спектре составляющих с частотами, лежащими за пределами полосы $f \in (-F_m , F_m)$, полностью определяется последовательностью своих отсчетов в дискретные моменты времени $X(t_i)$, следующих с шагом $\Delta t < 1/F_m$.*

Доказательство теоремы Котельникова

Человеческое ухо слышит в диапазоне от 20 Гц, до 20 КГц. Это подобно фильтру, вырезаем ограниченную часть спектра и воспринимаем только ограниченную часть спектра

Будем поэтому сразу считать, что передаваемый, или получаемый нами сигнал $f(t)$ (где t — время, $-\infty < t < \infty$) имеет финитный спектр, отличный от нуля лишь для частот ω, величина которых не превышает некоторого критического значения $a > 0$. Итак, $\hat{f}(\omega) \equiv 0$ при $|\omega| > a$, поэтому представление

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega t} d\omega$$

Доказательство теоремы

для функции с финитным спектром сводится к интегралу лишь по промежутку $[-a, a]$:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} \hat{f}(\omega)\, e^{i\omega t}\, d\omega. \tag{43}$$

На отрезке $[-a, a]$ функцию $\hat{f}(\omega)$ разложим в ряд Фурье

$$\hat{f}(\omega) = \sum_{-\infty}^{\infty} c_k(\hat{f})\, e^{i\frac{\pi\omega}{a}k} \tag{44}$$

$$c_k(\hat{f}) := \frac{1}{2a} \int_{-a}^{a} \hat{f}(\omega)\, e^{-i\frac{\pi\omega}{a}k}\, d\omega = \frac{\sqrt{2\pi}}{2a} f\left(-\frac{\pi}{a}k\right). \tag{45}$$

$$f\left(-\frac{\pi}{a}k\right) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} \hat{f}(\omega)\, e^{i\omega\cdot\left(-\frac{\pi}{a}k\right)}\, d\omega$$

Доказательство теоремы Котельникова

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} \sum_{-\infty}^{\infty} c_k(\hat{f}) e^{i\frac{\pi\omega}{a}k} e^{i\omega t} d\omega = \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} \sum_{-\infty}^{\infty} \frac{\sqrt{2\pi}}{2a} f\left(-\frac{\pi}{a}k\right) e^{i\frac{\pi\omega}{a}k} e^{i\omega t} d\omega =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} \sum_{-\infty}^{\infty} \frac{\sqrt{2\pi}}{2a} f\left(\frac{\pi}{a}k\right) e^{-i\frac{\pi\omega}{a}k} e^{i\omega t} d\omega =$$

$$= \frac{1}{2a} \sum_{k=-\infty}^{\infty} f\left(\frac{\pi}{a}k\right) \int_{-a}^{a} e^{i\omega\left(t-\frac{\pi}{a}k\right)} d\omega.$$

Доказательство теоремы Котельникова

$$\int_{-a}^{a} e^{i\omega\left(t-\frac{\pi}{a}k\right)}d\omega = \frac{1}{i\left(t-\frac{\pi}{a}k\right)}\int_{-a}^{a} e^{i\omega\left(t-\frac{\pi}{a}k\right)}d\left[i\omega\left(t-\frac{\pi}{a}k\right)\right] = \frac{1}{i\left(t-\frac{\pi}{a}k\right)}e^{i\omega\left(t-\frac{\pi}{a}k\right)}\Big|_{-a}^{a} =$$

$$= \frac{1}{i\left(t-\frac{\pi}{a}k\right)}\left[e^{ia\left(t-\frac{\pi}{a}k\right)} - e^{-ia\left(t-\frac{\pi}{a}k\right)}\right] = \frac{1}{i\left(t-\frac{\pi}{a}k\right)}2i\cdot\sin a\left(t-\frac{\pi}{a}k\right)$$

$$\sin(y) = \frac{e^{iy}-e^{-iy}}{2i}$$

Доказательство теоремы Котельникова

Вычислив эти элементарные интегралы, приходим к *формуле Котельникова*

$$f(t)=\sum_{k=-\infty}^{\infty} f\left(\frac{\pi}{a}k\right)\frac{\sin a\left(t-\frac{\pi}{a}k\right)}{a\left(t-\frac{\pi}{a}k\right)}. \qquad (46)$$

Формула (46) показывает, что для восстановления сообщения, описываемого функцией $f(t)$ с финитным спектром, сосредоточенным в полосе частот $|\omega|\leqslant a$, достаточно передать по каналу связи лишь значения $f(k\Delta)$ (называемые *отсчетными значениями*) данной функции через равные промежутки времени $\Delta=\pi/a$.

Это утверждение в совокупности с формулой (45) принадлежит В. А. Котельникову и называется *теоремой Котельникова* или *теоремой отсчетов*.

Теорема Котельникова

Таким образом, *по дискретной последовательности отсчетов функции $X(i\Delta t)$ можно абсолютно точно восстановить исходную непрерывную функцию $X(t)$,* если отсчеты брались с интервалом $\Delta t \le 1/2F_m$.

Это говорит о том, что не существует *принципиальных различий* между непрерывными и дискретными сигналами. Из любого непрерывного сигнала с ограниченным спектром можно взять его отсчеты в дискретные моменты времени, а затем по этим отсчетам *абсолютно точно* восстановить исходный непрерывный сигнал. При этом, *для абсолютно точного восстановления сигнала не нужно брать отсчеты бесконечно часто, достаточно, чтобы соблюдалось условие $\Delta t \le 1/2F_m$.*

$$\frac{\pi}{a} = \frac{\pi}{2\pi F_M}$$

Теорема Котельникова

Пример.

Передается речевой сигнал в полосе частот DF (от 0 до 3000 Гц). Время передачи Dt = 10 сек. Каждый дискретный отсчет кодируется 5 двоичными разрядами.

Определить минимальный объем памяти, требуемый для хранения информации Wзу .

T=1/2DF=1/6000=0,00016 с.,

следовательно, число отсчетов на интервале Dt:

N=Dt/T=10/0,00016=60000.

Объем ЗУ: Wзу =60000´ 5 = 300000 бит.

Теорема Котельникова

На практике теорему Котельникова для выбора шага дискретизации применяют следующим образом:

1. Определяют эффективную ширину спектра $f_э$.

2. Вычисляют шаг дискретизации $T=1/2f_э$.

3. На приемной стороне восстанавливается сигнал по следующей формуле

$$x(t) \approx \sum_{k=1}^{\beta} x(kT)\frac{\sin(2\pi f_э(t-kT))}{2\pi f_э(t-kT)},$$

где T - длительность сигнала; Dt - шаг дискретизации; b =Dt/T - база сигнала.

Теорема Котельникова

$$\Delta t \le \frac{1}{2f_{max}} = \frac{\pi}{\omega_{max}} \quad (4.2)$$

В зарубежной литературе теорему Котельникова чаще называют теоремой Найквиста или теоремой отсчётов.

Исходный сигнал в этом случае восстанавливается в следующем виде:

$$S(t) = \sum_{k=-\infty}^{\infty} S(k\Delta t) \frac{\sin\left(\pi \frac{t - k\Delta t}{\Delta t}\right)}{\pi \frac{t - k\Delta t}{\Delta t}} \quad (4.3).$$

Количественна мера информации

Всякая информация получается потребителем после приема сообщения. Сообщение, получаемое на приемной стороне, несет полезную информацию лишь в том случае, если у получателя имеется неопределенность относительно того – что будет передано. Если получатель *заранее знает то, что будет передано, то по результату этой передачи он не получает никакой информации.*

Предположим источник вырабатывает последовательность независимых дискретных сообщений $\{\lambda_i\}$, каждое из которых случайным образом выбирают из алфавита сообщения $A(\lambda_i) = \lambda_1, \lambda_2, \lambda_3, \dots \lambda_K$, где K - размер алфавита источника. Примером такого сообщения может быть, например, некоторый русский текст, где элементарными сообщениями являются буквы, или последовательность чисел из десятичного алфавита 0….9 и т.п.

В каждом элементарном сообщении λ_i для получателя содержится некоторая *информация*. Определим количество этой информации и выясним, от чего оно зависит.

До того, как связь состоялась, у получателя имеется большая или меньшая неопределенность относительно того, *какое сообщение* λ_i из числа возможных будет передано.

Количественная мера информации

Таким образом, очевидно, что количество информации, содержащейся в элементарном сообщении λ_i , является некоторой функцией от вероятности передачи этого сообщения $P(\lambda_i)$:

$$J(\lambda_i) = \varphi\{P(\lambda_i)\}.$$

Определим вид функции φ. Для этого потребуем, чтобы мера количества информации $J(\lambda_i)$ удовлетворяла двум интуитивным свойствам:

1. Если выбор сообщения λ_i заранее предопределен и мы знаем, какое сообщение будет передано ($P(\lambda_i) = 1$), то *количество информации* в этом сообщении равно нулю: $J(\lambda_i) = \varphi\{1\} = 0$.

2. Если источник последовательно выбирает сообщения λ_i и λ_j и вероятность такого выбора $P(\lambda_i, \lambda_j)$ есть совместная вероятность событий λ_i и λ_j, то *количество информации в этих двух* элементарных сообщениях будет равно *сумме количеств информации в каждом из них.*

Количественная мера информации

2. Если источник последовательно выбирает сообщения λ_i и λ_j и вероятность такого выбора $P(\lambda_i, \lambda_j)$ есть совместная вероятность событий λ_i и λ_j, то *количество информации в этих двух* элементарных сообщениях будет равно *сумме количеств информации в каждом из них.*

Вероятность совместного выпадения событий λ_i и λ_j $P(\lambda_i, \lambda_j)$, как известно, определяется по формуле полной вероятности

$$P(\lambda_i, \lambda_j) = P(\lambda_i) \cdot P(\lambda_j / \lambda_i) = P \cdot Q.$$

Тогда, в соответствии с требованием (2), должно выполняться условие

$$\varphi \{ P \cdot Q \} = \varphi(P) + \varphi(Q).$$

Нетрудно догадаться, что функцией, удовлетворяющей этим двум предъявляемым к ней условиям, является функция вида

$$J(\lambda_i) = a \log P(\lambda_i),$$

Количественная мера информации

Для удобства (чтобы количественная мера информации была положительной) принимают $a = -1$. Основание логарифма обычно выбирают равным двум, и тогда

$$J(\lambda_i) = -\log_2 P(\lambda_i).$$

Определенная таким образом единица измерения информации называется *битом информации.*

Например, если передаваемое элементарное сообщение λ_i выбрается из алфавита размером $K = 32$, в котором все буквы равновероятны, т .е. $P(\lambda_i) = 1/32$, то говорят, что в этом сообщении (в этой букве) содержится $\log_2 (1/32) = \mathbf{5}$ *бит* информации.

Количество информации

Количество информации, содержащееся в одном элементарном сообщении λ_i, еще не характеризует источник в целом. Одни элементарные сообщения могут нести много информации, но передаваться очень редко, другие - передаваться чаще, но нести меньше информации. Поэтому источник может быть охарактеризован *средним количеством информации, приходящимся на одно элементарное сообщение*, носящим название "*энтропия источника*" и определяемым следующим образом:

$$H(\lambda) = -\sum_{i=1}^{K} P(\lambda_i) \cdot \log P(\lambda_i), \qquad i = 1, K.$$

Энтропия, как количественная мера информативности источника, обладает следующими свойствами:

Количество информации

1. Энтропия есть величина вещественная, ограниченная и неотрицательная. Эти ее свойства вытекают из вида выражения для $H(\lambda)$, а также с учетом того, что $0 < P(\lambda_i) < 1$.

2. Энтропия детерминированных сообщений равна нулю, то есть $H(\lambda) = 0$, если хотя бы одно из сообщений имеет вероятность, равную единице.

3. Энтропия максимальна, если сообщения λ_i равновероятны, то есть

$P(\lambda_1) = P(\lambda_2) = ... P(\lambda_k) = 1/K$, и тогда

$$H(\lambda) = -\,{}^{1}\!/_{K} \sum_{i=1}^{K} \log {}^{1}\!/_{K} = \log K.$$

Энтропия сложных сообщений

Рассмотренные выше характеристики источника - количество информации и энтропия - относились к источнику, вырабатывающему поток независимых или *простых сообщений*.

Однако в реальных условиях независимость элементарных сообщений, вырабатываемых источником, - явление довольно редкое. Чаще бывает как раз обратное - достаточно сильная связь между элементами сообщения.

Например, при передаче текста вероятности появления отдельных букв зависят от того, какие буквы им предшествовали. Для русского текста, например, если передана буква "*П*", вероятность того, что следующей будет "*А*", гораздо выше, чем "*Н*", после буквы "*Ъ*" никогда не встречается "*Н*" и т.д. Подобная же картина наблюдается при передаче изображений - соседние элементы изображения имеют обычно почти одинаковые яркость и цвет.

Энтропия сложных сообщений

Сообщения, вырабатываемые такими источниками, называются *сложными сообщениями*, а сами источники - *источниками с памятью*.

Очевидно, что при определении энтропии и количества информации в сообщениях, элементы которых статистически связаны, нельзя ограничиваться только безусловными вероятностями - необходимо обязательно учитывать также условные вероятности появления сообщений.

Определим энтропию сложного сообщения, в котором имеется статистическая зависимость между двумя соседними элементарными сообщениями (*источник с памятью на одну букву*).

Пусть первое элементарное сообщение из пары принимает значения x_1, x_2, x_3,....x_k с вероятностями, соответственно, $P(x_1)$, $P(x_2)$,..... $P(x_k)$, второе - y_1, y_2,....y_m с вероятностями $P(y_1)$, $P(y_2)$,..... $P(y_m)$.

Энтропия сложных сообщений

Энтропию источника, вырабатывающего связанные символы X и Y можно определить следующим образом:

$$H(X,Y) = -\sum_{i=1}^{K}\sum_{j=1}^{m} P(x_i, y_j) * \log P(x_i, y_j) ,$$

Энтропия сложных сообщений

Совместная вероятность $P(x_i,y_j)$ по формуле Байеса определяется как

$$P(x_i, y_j) = P(x_i) * P(y_j / x_i) = P(y_j) * P(x_i / y_j),$$

тогда выражение для совместной энтропии можно записать в следующем виде:

$$H(X,Y) = -\sum_{i=1}^{k}\sum_{j=1}^{m} P(x_i) * P(y_j / x_i) * \log\{P(x_i) * P(y_j / x_i)\} =$$

$$= -\sum_{i=1}^{k} P(x_i) * \log P(x_i) * \sum_{j=1}^{m} P(y_j / x_i) - \sum_{i=1}^{k} P(x_i) * \sum_{j=1}^{m} P(y_j / x_i) * \log P(y_j / x_i).$$

Так как после первого сообщения x_i из пары обязательно будет передано второе (любое) из ансамбля Y, то

$$\sum_{j=1}^{m} P(y_j / x_i) = 1$$

Энтропия сложных сообщений

и совместная энтропия $H(X,Y)$ определится как

$$H(X,Y) = -\sum_{i=1}^{k} P(x_i) * \log P(x_i) - \sum_{i=1}^{k} P(x_i) * \sum_{j=1}^{m} P(y_j / x_i) * \log P(y_j / x_i) =$$

$$= H(X) + \sum_{i=1}^{k} P(x_i) * H(Y / x_i),$$

где $H(Y/x_i)$ - так называемая частная условная энтропия, отражающая энтропию сообщения Y при условии, что имело место сообщение x_i. Второе слагаемое в последнем выражении представляет собой усреднение $H(Y/x_i)$ по всем сообщениям x_i и называется *средней условной энтропией источника Y при условии передачи сообщения X.* И окончательно:

$$H(X,Y) = H(X) + H(Y/X).$$

Таким образом, *совместная энтропия пары связанных сообщений равна сумме безусловной энтропии первого из них и условной энтропии второго.*

Энтропия сложных сообщений

Следовательно, *при наличии связи между элементарными сообщениями энтропия источника снижается*, причем в тем большей степени, чем сильнее связь между элементами сообщения.

Таким образом, можно сделать следующие выводы относительно степени информативности источников сообщений:

1. *Энтропия* источника и количество информации *тем больше, чем больше размер алфавита* источника.

2. Энтропия источника зависит от статистических свойств сообщений. *Энтропия максимальна, если сообщения источника равновероятны и статистически независимы.*

3. *Энтропия источника, вырабатывающего неравновероятные сообщения, всегда меньше максимально достижимой.*

4. *При наличии статистических связей между элементарными сообщениями (памяти источника) его энтропия уменьшается.*

Сжатие с потерями и без потерь

Ниже приведена условная структура *системы сжатия данных*:

Данные источника → Кодер → Сжатые данные → Декодер → Восстановленные данные

Существуют два типа систем сжатия данных:

- системы сжатия без потерь информации (обратимое сжатие);
- системы сжатия с потерями информации (необратимое сжатие).

В системах сжатия без потерь декодер восстанавливает данные источника абсолютно точно, таким образом, процесс сжатия выглядит следующим образом:

Вектор данных **X** → Кодер → **B (X)** → Декодер → **X**

$$r = \frac{\text{размер данных источника в битах}}{\text{размер сжатых данных в битах}}$$

Сжатие с потерями и без потерь

Ниже приведена условная структура *системы сжатия данных*:

Данные источника → Кодер → Сжатые данные → Декодер → Восстановленные данные

Существуют два типа систем сжатия данных:

- системы сжатия без потерь информации (обратимое сжатие);
- системы сжатия с потерями информации (необратимое сжатие).

В системах сжатия без потерь декодер восстанавливает данные источника абсолютно точно, таким образом, процесс сжатия выглядит следующим образом:

Вектор данных **X** → Кодер → **B (X)** → Декодер → **X**

$$r = \frac{\textit{размер данных источника в битах}}{\textit{размер сжатых данных в битах}}$$

Арифметическое кодирование

Пусть нам нужно закодировать следующую текстовую строку: *РАДИОВИЗИР*.

Перед началом работы кодера исходный интервал составляет *[0; 1)*.

Алфавит кодируемого сообщения содержит следующие символы (буквы): { *Р, А, Д, И, О, В, З* }.

Определим количество каждой из букв алфавита в сообщении и назначим каждой из них интервал, пропорциональный вероятности этой буквы

Символ	Вероятность	Интервал
А	0.1	0 – 0.1
Д	0.1	0.1 – 0.2
В	0.1	0.2 – 0.3
И	0.3	0.3 – 0.6
З	0.1	0.6 – 0.7
О	0.1	0.7 – 0.8
Р	0.2	0.8 – 1

Арифметическое кодирование

Пусть нам нужно закодировать следующую текстовую строку: *РАДИОВИЗИР*.

Перед началом работы кодера исходный интервал составляет *[0; 1)*.

Алфавит кодируемого сообщения содержит следующие символы (буквы): { *Р, А, Д, И, О, В, З* }.

Определим количество каждой из букв алфавита в сообщении и назначим каждой из них интервал, пропорциональный вероятности этой буквы

Символ	Вероятность	Интервал
А	0.1	0 – 0.1
Д	0.1	0.1 – 0.2
В	0.1	0.2 – 0.3
И	0.3	0.3 – 0.6
З	0.1	0.6 – 0.7
О	0.1	0.7 – 0.8
Р	0.2	0.8 – 1

Арифметическое кодирование

Взяв первый символ сообщения - букву *Р*, кодер сужает исходный интервал до нового - [0.8; 1), который выделен этой букве.

Следующим символом сообщения, поступающим в кодер, будет буква *А*. Если бы эта буква была первой, ей был бы отведен интервал [0 - 0.1), но она следует за *Р* и поэтому кодируется новым *подынтервалом внутри уже выделенного для первой буквы интервала*. Начало и конец нового интервала определяются путем прибавления к началу предыдущего интервала числа, получаемого как *произведение ширины предыдущего интервала на значения нижней и верхней границ интервала, отведенного текущему символу.* В результате получим новый рабочий интервал [0.80 - 0.82).

Следующему символу *Д* выделен интервал [0.1 - 0.2), что применительно к уже имеющемуся рабочему интервалу [0.80 - 0.82) сужает его до величины [0.802 - 0.804).

Арифметическое кодирование

Очередным символом, поступающим на вход кодера, будет буква *И* с выделенным для нее интервалом [0,3 – 0,6). Применительно к уже имеющемуся рабочему интервалу получим [0,8026 - 0,8032).

Продолжая в том же духе, имеем:

вначале		[0.0 - 1.0)
после просмотра	*Р*	[0.8 - 1.0)
	А	[0.80 - 0.82)
	Д	[0.802 - 0.804)
	И	[0.8026 - 0.8032)
	О	[0.80302 - 0.80308)
	В	[0.803032 - 0.803038)
	И	[0.8030338 - 0.8030356)
	З	[0.80303488 - 0.80303506)
	И	[0.803034934 - 0.803034988)
	Р	[0.8030349772 - 0.8030349880)

Арифметическое кодирование

Теперь рассмотрим процедуру декодирования. Предположим, что все что декодер знает – это конечный интервал [0,8030349772 - 0,8030349880] и таблица распределения выделенных алфавиту интервалов.

Декодер сразу же может определить, что первый символ (буква) есть ***P***, так как результат кодирования лежит в интервале [0.8 - 1), выделенном согласно таблице символу ***P***.

Теперь повторим действия кодера:

вначале [0.0 - 1.0);

после просмотра [0.8 - 1.0).

Исключим из результата кодирования влияние теперь уже известного первого символа *P*, для этого вычтем из результата кодирования нижнюю границу диапазона, отведенного для *P*, – 0,8030349772 – 0.8 = 0,0030349772 – и разделим полученный результат на ширину интервала, отведенного для *P*, – 0.2. В результате получим 0,0030349772 / 0,2 = =0,015174886. Это число попадает в интервал, отведенный для буквы *A*, – [0 – 0,1) , следовательно, вторым символом декодированной последовательности будет *A*.

Арифметическое кодирование

Теперь рассмотрим процедуру декодирования. Предположим, что все что декодер знает – это конечный интервал [0,8030349772 - 0,8030349880] и таблица распределения выделенных алфавиту интервалов.

Декодер сразу же может определить, что первый символ (буква) есть ***P***, так как результат кодирования лежит в интервале [0.8 - 1), выделенном согласно таблице символу ***P***.

Теперь повторим действия кодера:

вначале [0.0 - 1.0);

после просмотра [0.8 - 1.0).

Исключим из результата кодирования влияние теперь уже известного первого символа *P*, для этого вычтем из результата кодирования нижнюю границу диапазона, отведенного для *P*, – 0,8030349772 – 0.8 = 0,0030349772 – и разделим полученный результат на ширину интервала, отведенного для *P*, – 0.2. В результате получим 0,0030349772 / 0,2 = =0,015174886. Это число попадает в интервал, отведенный для буквы *A*, – [0 – 0,1) , следовательно, вторым символом декодированной последовательности будет *A*.

Арифметическое кодирование

Поскольку теперь мы знаем уже две декодированные буквы - *РА*, исключим из итогового интервала влияние буквы *А*. Для этого вычтем из остатка 0,015174886 нижнюю границу для буквы *А* 0,015174886 – 0.0 = =0,015174886 и разделим результат на ширину интервала, отведенного для буквы *А*, то есть на 0,1. В результате получим 0,015174886/0,1=0,15174886. Результат лежит в диапазоне, отведенном для буквы *Д*, следовательно, очередная буква будет *Д*, и так далее, пока не декодируем все символы:

Арифметическое кодирование

Декодируемое число	Символ на выходе	Границы нижняя	Границы верхняя	Ширина интервала
0,8030349772	*Р*	0.8	1.0	0.2
0,015174886	*А*	0.0	0.1	0.1
0,15174886	*Д*	0.1	0.2	0.1
0,5174886	*И*	0.3	0.6	0.3
0,724962	*О*	0,7	0,8	0,1
0,24962	*В*	0,2	0,2	0,1
0,4962	*И*	0,3	0,6	0,3
0,654	*З*	0,6	0,7	0,1
0,54	*И*	0,3	0,6	0,3
0,8	*Р*	0,6	0,8	0,2
0.0	Конец декодирования			