Некоторые аспекты теоремы Шеннона

В.А.Зорич

Раздел I

1. *Информация и её количественное описание.*

Бит неопределённости и информации.

Неопределённость одного из M равновозможных событий, выраженная в битах. Указание одного такого события требует $\log_2 M$ бит. (Вектор из нулей и единиц).

Сколько различных объектов можно описать вектором длиной n из нулей и единиц: 2^n.

Если вероятность события p, то оно встречается один раз среди $M = \frac{1}{p}$ событий. Значит, его неопределённость, измеренная в битах, есть $\log_2 M = \log_2 \frac{1}{p} = -\log_2 p$.

Систему случайных событий (функцию со случайными значениями) в математике принято называть случайной величиной. Пусть X случайная величина. Средняя на одно значение (на одно событие) неопределённость в битах системы случайных событий — это математическое ожидание $EX = -\sum_{i=1}^{M} p_i \log_2 p_i$ случайной величины X. Это математическое ожидание называется *энтропией* системы или, точнее, *энтропией случайной величины* X и обозначают символом $H(X) = -\sum_{i=1}^{M} p_i \log_2 p_i$. (Греч. $\eta\nu\tau\rho\phi\pi\iota\alpha$, от др.-греч. $\eta\nu$ — в + $\tau\rho\phi\pi\eta$ — превращение; обращение. Термин *энтропия* исходно появился в термодинамике. В теорию информации введён Шенноном.)

Если имеется другая случайная величина (система возможных событий) Y, то появление её событий может дать информацию о событиях исходной случайной величины X и, тем самым, изменить её энтропию $H(X)$ на некоторую величину $H(X|Y)$. Величину $H(X|Y)$ называют *условной энтропией случайной величины X по отношению к случайной*

величине Y, а разность $I(X;Y) = H(X) - H(X|Y)$ считают *средней информацией о системе* X*, которую несёт появление одного из событий системы* Y .

Эти понятия подробно разрабатываются и описываются в общей теории информации.

Если событие системы X имело вероятность появления p, то его осуществление несёт $-\log_2 p$ бит информации. Но длительное наблюдение за появлениями такого события в единицу времени приносит всего $-p\log_2 p$ бит информации. Это же относится и к среднему числу бит информации $H(X) = -\sum_{i=1}^{M} p_i \log_2 p_i$, которое за единицу времени приносит появление одного из событий системы X (значений случайной величины X). В этом смысл информационной энтропии.

Этот статистический смысл энтропии можно описать следующим образом.

Пусть X — произвольная дискретная случайная величина, которая может принимать M различных значений x_i с вероятностями p_i соответственно.

Статистический характер энтропии в точной формальной записи состоит в следующем: для любых положительных чисел $\varepsilon,\ \delta$ найдется такое число $n_{\varepsilon\delta}$, что при $n \geq n_{\varepsilon\delta}$ выполняется неравенство

$$P\{|-\frac{1}{n}\sum_{i=1}^{n}\log_2 p_{x_i} - H(X)| < \delta\} > 1-\varepsilon, \tag{1}$$

где, как обычно, P — вероятность указанного в скобках события, но теперь $x_i,\ i = 1, ..., n$ — n независимых значений случайной величины X, а p_{x_i} — вероятности этих значений.

Как связана энтропия с кодированием?

Рассмотрим сообщения-слова-векторы $\bar{x} = (x_1, ..., x_n)$, образованные n последовательными независимыми значениями случайной величины X. Вероятность $p_{\bar{x}}$ появления слова $\bar{x}$ равна $p_{\bar{x}} = p_{x_1} \cdot ... \cdot p_{x_n}$. В силу соотношения (1) при $n \geq n_{\varepsilon\delta}$ с вероятностью, большей чем $1-\varepsilon$, будем иметь

$$2^{-n(H(X)+\delta)} \leq p_{\bar{x}} \leq 2^{-n(H(X)-\delta)}. \tag{2}$$

Слово $\bar{x}$ называют δ-*типичным*, если для него выполнены эти оценки. Ясно, что существует не более $2^{n(H(X)+\delta)}$ таких δ-типичных слов, а если $n \geq n_{\varepsilon\delta}$, то их ещё и не меньше чем $(1-\varepsilon)2^{n(H(X)-\delta)}$, и при этом всё множество не δ-типичных слов имеет вероятность, не большую чем ε.

В принципе теперь уже можно использовать двоичные последовательности длиной $n(H(X)+\delta)$, чтобы закодировать все δ-типичные слова. Даже, если все остальные слова закодировать одним символом, вероятность ошибки при передаче слов $\bar{x}$ длины n, вызванная таким кодом, будет меньше ε.

С другой стороны (и это эффект неустойчивости экономных кодов), любой код, использующий в той же ситуации двоичные последовательности относительно чуть меньшей длины $n(H(X)-\delta)$ (например, $2\delta n$ из $n(H(X)+\delta)$ посланных символов потерялось в шуме), будет иметь асимптотически неисчезающую вероятность ошибки, стремящуюся к единице при $n \to +\infty$.

Итак, связь энтропии и кодирования информации состоит, например, в том, что эффективное кодирование асимптотически при $n \to +\infty$ требует $N \sim 2^{nH(X)}$ слов и энтропия $H(X)$ может интерпретироваться как мера количества информации в битах на передаваемый символ, т. е. на одно значение случайной величины X.

Отсюда, в частности, следует, что энтропия источника информации не должна превышать пропускную способность канала связи, если мы хотим адекватно и без задержек передавать поступающую информацию по этому каналу связи.

2. *Скорость передачи информации.*

Если передаётся M сообщений и на передачу каждого затрачивается время T секунд, то это равносильно тому что скорость передачи в битах в секунду равна $\frac{1}{T}\log_2 M$.

Раздел II

3. *Дискретизация непрерывного сигнала.*

Разложения в ряды.
Ряд Фурье. Спектр сигнала.
Интеграл Фурье. Сигналы с конечным спектром. (Зрение, слух.)
Теорема отсчётов (формула Котельникова — Шеннона) и интервал Найквиста.
Оценка размерности n телевизионного сигнала ($n \approx 10^{10} \gg 1$).

Теорема отсчётов (формула Котельникова — Шеннона).

Это пример красивого комбинирования ряда и интеграла Фурье. Формула имеет прямое отношение к теории передачи информации по каналу связи. Чтобы пример не показался искусственным, напомним, что в силу ограниченных возможностей наших органов чувств мы способны воспринимать сигналы только в определенном диапазоне частот. Например, ухо слышит. в диапазоне от 20 герц до 20 килогерц. Таким образом, какие бы ни были сигналы, мы, подобно фильтру, вырезаем только ограниченную часть их спектра и воспринимаем их как сигналы с финитным спектром. 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

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

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

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} \hat{f}(\omega) e^{i\omega t} d\omega. \tag{3}$$

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

$$\hat{f}(\omega) = \sum_{-\infty}^{\infty} c_k(\hat{f}) e^{i\frac{\pi\omega}{a}k} \tag{4}$$

по системе $\{e^{i\frac{\pi\omega}{a}k}; k \in \mathbb{Z}\}$, ортогональной и полной на этом отрезке. Учитывая формулу (3), для коэффициентов $c_k(\hat{f})$ этого ряда получаем сле-

дующее простое выражение:

$$c_k(\hat{f}) = \frac{1}{2a}\int_a^a \hat{f}(\omega)e^{-i\frac{\pi\omega}{a}k}d\omega = \frac{\sqrt{2\pi}}{2a}f\left(-\frac{\pi}{a}\ k\right). \qquad (5)$$

Подставляя ряд (4) в интеграл (3), с учетом соотношений (5) находим

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

Или

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

Вычислив эти элементарные интегралы, приходим к формуле

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

Это формула Котельникова — Шеннона или формула (теорема) отсчётов. Она показывает, что для восстановления сообщения, описываемого функцией $f(t)$ с финитным спектром, сосредоточенным в полосе частот $0 \le \omega \le a$, достаточно передать по каналу связи лишь значения $f(k\Delta)$ (называемые отсчётными значениями) данной функции через равные промежутки времени $\Delta = \frac{\pi}{a} = \frac{2\pi}{2a}$.

Как интерполяционная формула (специальный случай формулы Лагранжа для целых функций конечной степени) она уже была известна математикам. Заслуга Котельникова и Шеннона состоит в интерпретации этой формулы с точки зрения кодирования сигнала и передачи информации по каналу связи.

Отсчётный интервал Δ, вслед за Шенноном, называют *интервалом Найквиста*, обратившего на него особое внимание. Напомним, что частота ν колебаний в секунду, измеряемая в герцах, и круговая частота ω, измеряемая в радианах в секунду, связаны соотношением $\omega = 2\pi\nu$. Таким образом, если спектр частот $0 \le \nu \le W$ ограничен величиной W, то есть период колебаний максимальной частоты есть $1/W$ секунды, то отсчётный интервал должен быть вдвое короче $\Delta = \frac{1}{2W}$.

Оценим, вслед за Шенноном, размерность вектора отсчётных значений TV-сигнала времён Шеннона. Пусть, например, как и Шеннон, мы

рассматриваем TV-сигнал, имеющий частоту $W = 5$ MHz (1 мегагерц $= 10^6$ Герц) и продолжительность $T = 1$ час. Подсчитаем, какой длины должен быть вектор отсчётных значений, отвечающий такому сигналу, т. е. найдем количество $N = T/\Delta = 2WT$ отсчетных значений: $N = 2 \cdot 5\text{МГц} \cdot 1\text{час} = 2 \cdot 5 \cdot 10^6 \cdot 60^2 = 3,6 \cdot 10^{10}$. Это вектор в пространстве $\mathbb{R}^N$ огромной размерности. Геометрия такого пространства имеет свою специфику, которую Шеннон понимал и использовал в своей теореме.

4. *Мощности сигнала и помехи.*

Работа передатчика по созданию электромагнитного сигнала $f(t)$ продолжительностью T выражается интегралом

$$\int_0^T f^2(t)dt.$$

Тогда мощность передатчика (и сигнала) есть

$$P = \frac{1}{T}\int_0^T f^2(t)dt.$$

Если

$$f(t) \simeq \sum_1^n x_k e_k,$$

где $< e_i, e_j >= \frac{1}{2W}\delta_{ij}$, то

$$\int_0^T f^2(t)dt = \frac{1}{2W}\sum_1^n x_k^2$$

и, если $x = (x_1, ..., x_n)$, то

$$P = \frac{1}{2WT}||x||^2.$$

Поскольку $2WT = n$,

$$||x||^2 = nP, \qquad x_i^2 \simeq P.$$

Таким образом, P можно трактовать как мощность, отнесённую к одному отсчётному значению сигнала $f(t)$.

Аналогично, если $\xi = (\xi_1, ..., \xi_n)$ — вектор помехи, то

$$||\xi||^2 = nN, \qquad \xi_i^2 \simeq N,$$

где nN — полная мощность помехи, а N — мощность, отнесённая к каждой координате вектора помехи.

Раздел III

5. *Идея случайного кода в пространстве большой размерности.*

Помеха раздувает каждое индивидуальное сообщение, выраженное вектором (точкой) x, до шара с центром x и радиусом $||\xi||$.

Чтобы гарантированно избежать ошибок декодирования надо заниматься плотной упаковкой таких шаров, избегая их пересечений. Но, если размерность очень большая, то, даже при наличии пересечений, вероятность ошибки декодирования становится малой, и тем меньшей, чем больше размерность.

В этой связи вместо поиска плотной упаковки вызванных помехами шаров, центры шаров распределяют случайным образом по области сигнала. Полученное распределение центров принимают за исходный код. Итак, исходный код случаен! Но, оказывается, он очень эффективен. Проделаем нужный расчёт.

Количество сообщений при случайном распределении по объёму шара (определяемого мощностью сигнала и объёмом шара, определяемого мощностью помехи) при $n \gg 1$.

Оценка вероятности ошибки при декодировании такого кода.

Итак, рассчитаем случайный код с учётом помехи.

Учитывая, что размерность n очень большая, можно считать, что почти все точки случайно брошенные в шар $||x||^2 \leq nP$ возможных сообщений x, окажутся в непосредственной близости граничной сферы, то есть можно считать, что $||x|| = \sqrt{nP}$.

Далее, случайный вектор помехи ξ независим от вектора x, поэтому при $n \gg 1$ можно считать, что $< x, \xi >= 0$.

Тогда

$$||x + \xi||^2 = ||x||^2 + ||\xi||^2 = n(P + N).$$

И, значит, если центр x шара помехи находится на сфере $||x|| = \sqrt{nP}$, то вектор (точка) $x + \xi$ находится на сфере радиуса $\sqrt{n(P + N)}$.

Если шары помехи, имеющие радиус $\sqrt{nN}$ распределять случайным образом по объёму шара радиуса $\sqrt{n(P + N)}$ до исчерпания его объёма, то получим случайный набор (код) из

$$M = \left(\frac{n(P + N)}{nN}\right)^{\frac{n}{2}} = \left(1 + \frac{P}{N}\right)^{\frac{n}{2}} = \left(1 + \frac{P}{N}\right)^{WT}$$

точек, кодирующих этим случайным кодом M сообщений.

6. *Скорость передачи информации в канале при наличии помех.*

Теперь можно найти скорость передачи информации по такому каналу связи в битах в секунду

$$C = \frac{1}{T}\log_2 M = W\log_2\left(1+\frac{P}{N}\right).$$

Заметим, что даже если вместо равенства

$$M = \left(1+\frac{P}{N}\right)^{WT}$$

написать приближённое равенство с некоторым ограниченным коэффициентом пропорциональности

$$M = k\left(1+\frac{P}{N}\right)^{WT},$$

то при $T \gg 1$ мы получили бы ту же скорость C передачи информации.

7. *Оценка вероятности ошибки.*

Допустим, что на приёмном конце получен вектор $x+\xi$. Надо восстановить переданное сообщение x.

Известно, что $||x|| = \sqrt{nP}$ и $||x+\xi||^2 = n(P+N)$. Ошибка декодирования может наступить, если в луночку пересечения шара $||x|| \le \sqrt{nP}$ и шара радиуса $\sqrt{N}$ с центром $x+\xi$, кроме x, попадёт ещё хотя бы одна кодовая точка. Поскольку кодовые точки распределены по площади случайным образом, вероятность такого события меньше отношения объёма этой луночки к объёму всего шара $||x|| \le \sqrt{nP}$. Если диаметр луночки равен $2h$, то объём луночки меньше, чем объём шара радиуса h.

Величину h находим из элеменратной геометрии, записывая двумя способами площадь прямоугольного треугольника с катетами x, ξ и гипотенузой $x+\xi$:

$$h||x+\xi|| = ||x||\cdot||\xi|| \qquad\qquad h = \frac{\sqrt{nP}\sqrt{nN}}{\sqrt{n(P+N)}}$$

Значит, вероятность ошибки декодирования не превосходит величины

$$\left(\frac{h}{||x||}\right)^n = \left(\frac{\sqrt{nN}}{\sqrt{n(P+N)}}\right)^n = \left(\frac{N}{P+N}\right)^{\frac{n}{2}} = \left(1+\frac{P}{N}\right)^{-WT} \asymp \frac{1}{M}.$$

Увеличивая T, мы увеличиваем $n = 2WT$ и $M = \left(1 + \frac{P}{N}\right)^{WT}$, а вероятность ошибки $\frac{1}{M}$ уменьшается при сохранении найденной скорости передачи $C = W \log_2 \left(1 + \frac{P}{N}\right)$. Таким образом, скорость передачи определяется прежде всего шириной W полосы допустимых частот. Это вполне аналогично тому, как конечая скорость ракеты прежде всего определяется эффективностью горючего, а не соотношением масс ракеты и топлива.

8. *Формулировка теоремы Шеннона.*

Итак, мы пояснили содержание, смысл и справедливость следующего утверждения.

Теорема.

Пусть имеется канал связи со следующими характеристиками: P — удельная (на одно отсчётное значение сигнала) мощность передатчика; N — удельная мощность случайной помехи (белого шума); W — ширина полосы частот передатчика.

В таком канале можно асимптотически достичь передачи информации со скоростью

$$C = W \log_2 \left(1 + \frac{P}{N}\right)$$

бит в секунду при сколь годно малой вероятности ошибки декодирования.

Эта теорема была доказана Клодом Шенноном в работе 1949 года, оригинал которой будет приложен к этому файлу. Во многих отношениях полезно просмотреть оригинал. Как и почти каждая фундаментальная работа, эта статья написана просто, ясно и без пустых слов.

Небольшой заключительный комментарий.

Следует обратить внимание на то, что любое превышение указанной скорости передачи ведёт к тому, что ошибки декодирования появляются с вероятностью, стремящейся к единице. Это родовой грех любого оптимального кода, как и грех неустойчивости любого максимума потенциальных возможностей (например, максимума потенциальной энергии).

Ещё следует обратить внимание на то, что теорема говорит не о скорости передачи самих сообщений, которые могут быть очень длинными, а о скорости передачи информации в битах в секунду. Передача одного сообщения может занимать время $T \gg 1$. Более того, для уменьшения

вероятности ошибки декодирования, сообщения приходится делать длинными. То есть теорема даёт предельную скорость передачи информации при малой вероятности ошибки декодирования как раз, когда $T \to \infty$. Это означает большую задержку в передаче самих сообщений, ели они очень длинные и передаются последовательно.