Теория информации

Лекция 9.

Теорема кодирования Шеннона для канала с шумом

Данная теорема имеет три части, одно утверждение положительное и два отрицательных.

Теорема.

1) Для каждого дискретного канала без памяти с емкостью канала

$$C = \max_{P_x} I(X;Y) \tag{9.1}$$

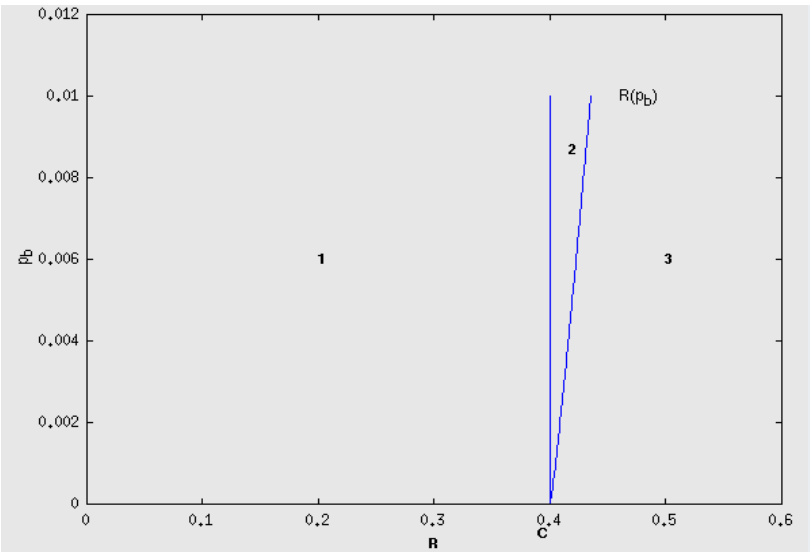
выполняется следующее: для каждого $\varepsilon > 0$ и $R < C$, для достаточно большого N существует блочный код (N,K) с долей кода на бит $\frac{K}{N} \geq R$ такие что $p_{BM} < \varepsilon$.

2) Если вероятность битовой ошибки p_b приемлема, то доля кода на бит достижима вплоть до величины

$$R(p_b) = \frac{C}{1 - H_2(p_b)} \quad . \tag{9.2}$$

3) Для любой p_b доля кода на бит больше $R(p_b)$ недостижима.

Ниже приведен рисунок иллюстрирующий достижимые и недостижимые области в плоскости (R, p_b). Области 1 и 2 достижимы, а область 3 нет.



Перед тем как доказывать эту теорему, рассмотрим понятие совместной типичности.

Совместная типичность

Рассмотрим кодовое слово $\boldsymbol{x}^{(s)}$ как элемент ансамбля X^N и ему соответствующее значение на выходе канала $\boldsymbol{y}$ и, таким образом определим совместный ансамбль $(XY)^N$. Рассмотрим теперь типический декодер, то есть такой, который интерпретирует сигнал $\boldsymbol{y}$ на выходе канала как сигнал s если кодовое слово $\boldsymbol{x}^{(s)}$ и сигнал на выходе $\boldsymbol{y}$ являются *совместно типичными*.

Совместная типичность. Пара последовательностей (строк) $\boldsymbol{x}$ и $\boldsymbol{y}$ длиной N являются совместно типичными с точностью β по отношению к распределению $P(\boldsymbol{x},\boldsymbol{y})$ если

$\boldsymbol{x}$ типичная для $P(\boldsymbol{x})$, т.е. $\left|\frac{1}{N}\log_2\frac{1}{P(\boldsymbol{x})} - H(X)\right| < \beta$,

$\boldsymbol{y}$ типичная для $P(\boldsymbol{y})$, т.е. $\left|\frac{1}{N}\log_2\frac{1}{P(\boldsymbol{y})} - H(Y)\right| < \beta$,

пара $(\boldsymbol{x},\boldsymbol{y})$ типичная для $P(\boldsymbol{x},\boldsymbol{y})$, т.е. $\left|\frac{1}{N}\log_2\frac{1}{P(\boldsymbol{x},\boldsymbol{y})} - H(X,Y)\right| < \beta$.

Совместно типическое множество $J_{N\beta}$ это множество всех совместно типичных пар $\boldsymbol{x},\boldsymbol{y}$ длины N.

Теорема о совместной типичности. Пусть пара $\boldsymbol{x},\boldsymbol{y}$ извлечена из ансамбля $(XY)^N$ определенного вероятностью

$$P(\boldsymbol{x},\boldsymbol{y}) = \prod_{n=1}^{N} P(x_n, y_n) .$$

Тогда

1. вероятность того, что $\boldsymbol{x}$ и $\boldsymbol{y}$ совместно типичные (с точностью β) стремится к 1 при $N \to \infty$;
2. число совместно типичных пар строк $|J_{N\beta}|$ близко к $2^{NH(X,Y)}$, а именно

$$|J_{N\beta}| \le 2^{N(H(X,Y)+\beta)} \quad ; \tag{9.3}$$

3. если $\boldsymbol{x}'$ и $\boldsymbol{y}'$ независимые последовательности (строки) из ансамблей X^N и Y^N соответственно с той же маргинальной вероятностью $P(\boldsymbol{x},\boldsymbol{y})$, то вероятность того, что пара $(\boldsymbol{x}',\boldsymbol{y}')$ принадлежит совместно типическому множеству близка к $2^{-NI(X;Y)}$, а именно

$$P((\boldsymbol{x}',\boldsymbol{y}')\in J_{N\beta}) \le 2^{-N(I(X;Y)-3\beta)} \quad . \tag{9.4}$$

Доказательство:
В курсе теории вероятности при изучении законов больших чисел доказывается неравенство Чебышева

$$P\left(|x-\langle x\rangle|\ge\alpha\right) \le \frac{\sigma_x^2}{\alpha^2} \quad , \tag{9.5}$$

где α - положительное число. Неравенство Чебышева можно переписать в виде более удобном для данного доказательства:

$$P\left((x-\langle x\rangle)^2 \ge \alpha\right) \le \frac{\sigma_x^2}{\alpha} \quad . \tag{9.6}$$

Если случайная величина x_N это среднее арифметическое N независимых случайных величин $h_1,h_2,h_3,\ldots,h_N$ имеющих общие среднее $\langle h\rangle$ и дисперсию σ_h^2, то неравенство Чебышева формулируется как закон больших чисел[1]

$$P\left((x_N-\langle h\rangle)^2 \ge \alpha\right) \le \frac{\sigma_h^2}{\alpha N} \quad . \tag{9.7}$$

Неравенство (9.7) легко доказывается с учетом того, что $\sigma_x^2=\sigma_h^2/N$.

Рассмотрим теперь случайную величину

$$\xi = \frac{1}{N}\log_2\frac{1}{P(\boldsymbol{x},\boldsymbol{y})} = \frac{1}{N}\sum_{n=1}^{N}\log_2\frac{1}{P(x_n,y_n)} \quad ,$$

которая есть среднее арифметическое случайных величин $\log_2\frac{1}{P(x_n,y_n)}$, для

1 Закон больших чисел формулируется как свойство устойчивости среднего арифметического. Теорема Чебышева формулируется следующим образом: *при достаточно большом числе независимых опытов среднее арифметическое случайной величины сходится по вероятности к ее математическому ожиданию*. В математической форме это утверждение обычно записывается в виде формулы

$$P(|x_N-\langle h\rangle|<\varepsilon)>1-\delta \quad ,$$

где ε,δ - произвольно малые положительные числа.

которых, в свою очередь, известно среднее

$$\left\langle \log_2 \frac{1}{P(x_n, y_n)} \right\rangle = H(X,Y) = H \quad .$$

Таким образом, очевидно, что $\langle \xi \rangle = H$. Закон больших чисел для этой переменной можно записать в виде

$$P\left\{ (\xi - \langle \xi \rangle)^2 \geq \beta^2 \right\} \leq \frac{\sigma^2}{\beta^2 N} \quad , \tag{9.8}$$

где вместо α использована β^2 , а σ^2 это дисперсия случайной величины $\log_2(1/P(x_n, y_n))$. Теперь подставляя значения ξ и $\langle \xi \rangle$ в (9.8) можно получить выражение

$$P\left\{ \left(\frac{1}{N} \log_2 \frac{1}{P(\boldsymbol{x}, \boldsymbol{y})} - H \right)^2 \geq \beta^2 \right\} \leq \frac{\sigma^2}{\beta^2 N} \quad , \tag{9.9}$$

где слева стоит вероятность того, что пара $(\boldsymbol{x}, \boldsymbol{y})$ не входит в совместно типическое множество $J_{N\beta}$. Исходя из этого

$$1 - P\left\{ (\boldsymbol{x}, \boldsymbol{y}) \in J_{N\beta} \right\} \leq \frac{\sigma^2}{\beta^2 N}$$

и, следовательно,

$$P\left\{ (\boldsymbol{x}, \boldsymbol{y}) \in J_{N\beta} \right\} \geq 1 - \frac{\sigma^2}{\beta^2 N} \quad ,$$

откуда видно, что при $N \to \infty$ $P\left\{ (\boldsymbol{x}, \boldsymbol{y}) \in J_{N\beta} \right\} \to 1$. Таким образом, доказано утверждение 1 теоремы.

По определению совместной типичности принадлежность пары $(\boldsymbol{x}, \boldsymbol{y})$ совместно типичному множеству $J_{N\beta}$ означает выполнение неравенства

$$\left| \frac{1}{N} \log_2 \frac{1}{P(\boldsymbol{x}, \boldsymbol{y})} - H \right| < \beta \quad ,$$

что, в свою очередь, означает, что вероятность $P(\boldsymbol{x}, \boldsymbol{y})$ для такой пары заключена в пределы, определяемые неравенством

$$2^{-N(H+\beta)} < P(\boldsymbol{x}, \boldsymbol{y}) < 2^{-N(H-\beta)} \quad . \tag{9.10}$$

Таким образом величина $2^{-N(H+\beta)}$ представляет собой оценку снизу вероятности $P(\boldsymbol{x},\boldsymbol{y})$ для типичной пары. Это позволяет записать следующее неравенство:

$$|J_{N\beta}|2^{-N(H+\beta)} < \sum_{(\boldsymbol{x},\boldsymbol{y})\in J_{N\beta}} P(\boldsymbol{x},\boldsymbol{y}) < 1 ,$$

откуда непосредственно следует, утверждение 2 теоремы, т.е.

$$|J_{N\beta}| < 2^{N(H+\beta)} .$$

Что касается утверждения 3, то здесь можно воспользоваться независимостью переменных $\boldsymbol{x}'$ и $\boldsymbol{y}'$, а также оценками вида (9.10):

$$P\left((\boldsymbol{x}',\boldsymbol{y}')\in J_{N\beta}\right) = \sum_{(\boldsymbol{x},\boldsymbol{y})\in J_{N\beta}} P(\boldsymbol{x})P(\boldsymbol{y}) \tag{9.11}$$

$$\leq |J_{N\beta}|2^{-N(H(X)-\beta)}2^{-N(H(Y)-\beta)} \tag{9.12}$$

$$\leq 2^{N(H(X,Y)+\beta)}2^{-N(H(X)+H(Y)-2\beta)} = \tag{9.13}$$

$$= 2^{-N(I(X;Y)-3\beta)} \tag{9.14}$$

#

Доказательство теоремы кодирования для канала с шумом

Идея Шеннона заключалсь в том, чтобы не конструировать систему кодирования и декодирования чтобы оценить вероятность ошибки, а в том чтобы определить среднюю вероятность сразу для *всех* кодов и доказать, что она мала.

Рассмотрим следующую кодирующую систему с долей кода на бит равной R':

1. Зафиксируем распределение $P(x)$ и создадим $S=2^{NR'}$ кодовых слов блочного кода $(N, NR')=(N,K)$ с распределением

$$P(\boldsymbol{x}) = \prod_{n=1}^{N} P(x_n) . \tag{9.15}$$

2. Пусть данный код (обозначим его как C) известен и передатчику и приемнику по обе стороны шумящего канала.
3. Сообщение s выбирается из множества $\{1,2,3,\ldots,2^{NR'}\}$,

соответствующее ему кодовое слово $\boldsymbol{x}^{(s)}$ посылается в канал. Принятый сигнал $\boldsymbol{y}$ характеризуется вероятностью

$$P(\boldsymbol{y} \mid \boldsymbol{x}^{(s)}) = \prod_{n=1}^{N} P(y_n \mid x_n^{(s)}) \quad . \tag{9.16}$$

4. Сигнал декодируется *типическим декодером*, что означает интерпретировать принятое слово $\boldsymbol{y}$ как сигнал $\hat{s}$ если пара $(\boldsymbol{x}^{(\hat{s})}, \boldsymbol{y})$ совместно типичная и нет другого s', такого что пара $(\boldsymbol{x}^{(s')}, \boldsymbol{y})$ тоже совместно типичная. В противном случае считать, что произошла ошибка, т.е. $\hat{s}=0$. И хотя такой декодер не является оптимальным, он удобен для анализа.
5. Ошибка декодирования случается когда $\hat{s} \neq s$.

Имеется три вероятности ошибки, которые стоит рассмотреть. Во-первых, вероятность блочной ошибки для какого-то определенного кода C:

$$p_B(C) = P(\hat{s} \neq s \mid C) \quad . \tag{9.17}$$

Эту величину труднее всего оценить. Во-вторых, усредненная по всем кодам вероятность блочной ошибки:

$$\langle p_B \rangle = \sum_C P(C) P(\hat{s} \neq s \mid C) \quad . \tag{9.18}$$

И, в-третьих, максимальная вероятность блочной ошибки для заданного кода C:

$$p_{BM} = \max_s P(\hat{s} \neq s \mid s, C) \quad . \tag{9.19}$$

Последняя наиболее интересна, поскольку для доказательства теоремы кодирования необходимо показать, что существует код C, для которого эта ошибка мала.

Имеется два источника ошибок типического декодера. Либо a) выходное слово $\boldsymbol{y}$ не является совместно типичным с переданным кодовым словом $\boldsymbol{x}^{(s)}$, либо b) имеется другое кодовое слово в C совместно типичное с принятым словом $\boldsymbol{y}$. В силу способа конструирования кода, средняя вероятность ошибки не должна зависеть от того, какой именно сигнал s рассматривается. Поэтому пусть для определенности и без потери общности рассмотрения $s=1$.

a) Вероятность того, что входное кодовое слово $\boldsymbol{x}^{(1)}$ и выходное кодовое слово $\boldsymbol{y}$ не являются совместно типичными исчезающе мала при $N\to\infty$ в силу утверждения 1 теоремы о совместной типичности. Обозначим как δ верхнюю границу этой вероятности (оценка сверху) удовлетворяющую условию $\delta\to 0$ при $N\to\infty$. Таким образом для каждого произвольно малого δ можно найти достаточно большую длину кодового слова $N(\delta)$, такую что $P\left((\boldsymbol{x}^{(1)},\boldsymbol{y})\in J_{N\beta}\right)\le\delta$.

b) Вероятность того, что $\boldsymbol{x}^{(s')}$ и $\boldsymbol{y}$ совместно типичные для заданного $s'\neq 1$ не превышает $2^{-N(I(X;Y)-3\beta)}$ в силу утверждения 3 теоремы о совместной типичности. И имеется всего $2^{NR'}-1$ возможно ошибочных s'.

Таким образом, средняя вероятность блочной ошибки удовлетворяет неравенству:

$$\langle p_B\rangle \le \delta+\sum_{s'=2}^{2^{NR'}} 2^{-N(I(X;Y)-3\beta)} \le \delta+2^{-N(I(X;Y)-R'-3\beta)}. \qquad (9.20)$$

Средняя вероятность блочной ошибки (9.20) может быть уменьшена до $<2\delta$ с помощью увеличения длины кодового слова N, если, конечно, выражение в скобках в показателе степени двойки положительно или

$$R'<I(X;Y)-3\beta. \qquad (9.21)$$

Выполним следующие модификации:

1. Выберем оптимальное входное распределение $P(x)$, тогда условие (9.21) превратиться в неравенство $R'<C-3\beta$.
2. Поскольку средняя всем кодам вероятность ошибки меньше 2δ, значит среди этих кодов есть такой, для которого $p_B(C)<2\delta$.
3. Чтобы показать, что не только средняя, но и максимальная вероятность ошибки p_{BM} может быть сделана малой, модифицируем код отбрасывая худшую половину кодовых слов, тех, которые с наибольшей вероятностью могут быть источником ошибки. Те кодовые слова, которые останутся должны все иметь условную вероятность ошибки менее чем 4δ. Оставшиеся кодовые слова формируют новый код, имеющий $2^{NR'-1}$ кодовых слов, что означает уменьшение доли кода на бит с величины R' до величины $R'-1/N$. Если N велико, то уменьшение скорости незначительно. Для этого кода достигнута $p_{BM}<4\delta$. Результирующий код не является оптимальным, но достаточно хорош для доказательства теоремы.

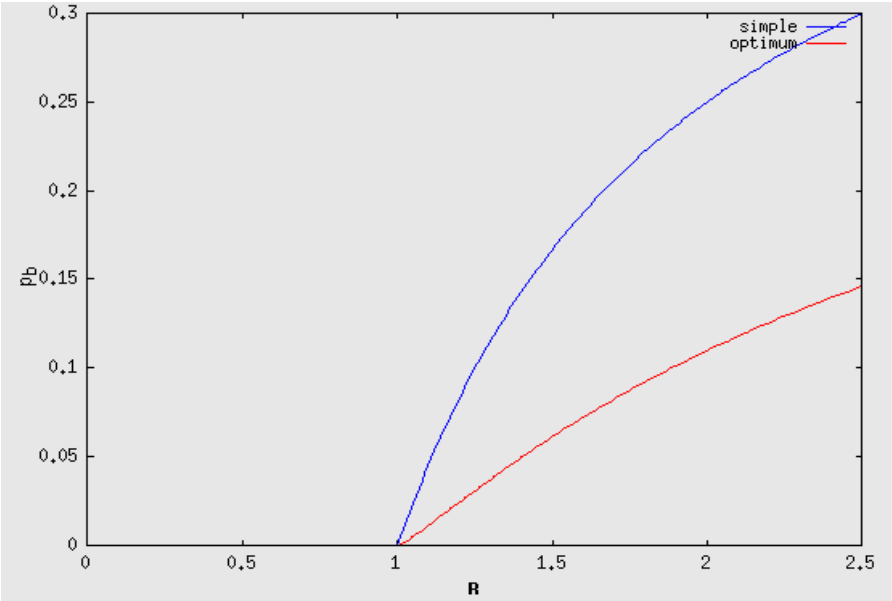
Таким образом, мы можем сконструировать код с долей кода на бит $R'-\frac{1}{N}$, где $R'<C-3\beta$ с $p_{BM}<4\delta$. Утверждение 1 теоремы кодирования можно

получить, если задать $R' = (R+C)/2$, $\delta = \varepsilon/4$, $\beta < (C-R')/3$ и N достаточно большим, чтобы все условия были выполнены.

Таким образом, доказана достижимость области 1 в плоскости R, p_b . Это означает, что канал с шумом может быть использован как двоичный канал без шума до скорости вплоть до доли кода C на бит. Теперь представим, что мы пытаемся достичь скорости передачи с $R=2$. Для этого посылающая сторона должна эффективно отбрасывать половину информации. Как сделать это наилучшим образом, чтобы минимизировать вероятность битовой ошибки? Самая простая стратегия это посылать часть равную $1/R$ бит источника игнорируя остальные. Приемник должен случайным образом угадывать оставшуюся часть бит равную $1-1/R$, средняя вероятность битовой ошибки будет в этом случае

$$p_b = \frac{1}{2}(1-1/R) \quad . \tag{9.22}$$

Можно даже достичь меньшей ошибки распределяя риск повреждения между всеми битами, фактически можно достичь $p_b = H_2^{-1}(1-1/R)$ (здесь $H_2^{-1}()$ это функция обратная к $H_2()$). Ниже на графиках показаны зависимость (9.22) и оптимальная Шенноновская граница достижимости.



Воспользуемся только что разработанным кодом (N,K) для канала с шумом используя декодер, чтобы задать сжатие с потерями (упаковщик с потерями). Возьмем превосходный (N,K) код с долей кода на бит $R'=K/N$ для двоичного симметричного канала с переходной вероятностью q . Асимптотически для такого кода $R' \simeq 1-H_2(q)$. Вспомним, что если такой код

достигает емкости канала, то (а) это означает равномерное распределение на входе и примерно равномерное распределение на его выходе, поскольку энтропия на выходе равна энтропии источника источника (NR') плюс энтропия шума ($NH_2(q)$), и (б) оптимальный декодер такого кода обычно отображает принятый вектор длиной N на переданный вектор в котором отличаются в среднем qN бит.

Итак сигнал, который необходимо передать через канал делится на блоки по N бит (вместо того, чтобы отправляться блоками длиной K бит) и отправляется на сжатие с потерями, после чего уже сжатый блоками по K бит отправляется в канал, который уже работает в области 1 со скоростью равной его емкости. На выходе декодер берет код блоками по K бит и реконструирует (распаковывает) сообщение, в котором в среднем qN битовых ошибок. Таким образом, вероятность битовой ошибки будет $p_b = q$. Отношение, с которым осуществляется сжатие с потерями $R = N/K = 1/R' = 1/(1-H_2(p_b))$. Применение такого сжатия с потерями к каналу емкостью C , работающему в области без ошибок и приводит к ограничению достижимости кода, задаваемому границей

$$R = \frac{C}{1-H_2(p_b)} \quad . \tag{9.23}$$

Таким образом достигается область 2, соответствующая утверждению 2 теоремы кодирования.

Источник, кодер, шумящий канал, декодер определяют цепь Маркова:

$$P(s, \boldsymbol{x}, \boldsymbol{y}, \hat{s}) = P(s)P(\boldsymbol{x} \mid s)P(\boldsymbol{y} \mid \boldsymbol{x})P(\hat{s} \mid \boldsymbol{y}) \quad . \tag{9.24}$$

Очевидно, что для цепи Маркова $I(s;\hat{s}) \le I(\boldsymbol{x};\boldsymbol{y})$. Более того, по определению емкости канала $I(\boldsymbol{x};\boldsymbol{y}) \le NC$, а следовательно и $I(s;\hat{s}) \le NC$. Предположим теперь, что наша система достигла точки R, p_b , тогда взаимная информация $I(s;\hat{s}) \ge NR(1-H_2(p_b))$. Но $I(s;\hat{s}) > NC$ недостижима, а следовательно и $R > \frac{C}{1-H_2(p_b)}$ недостижима. #