Теория информации

Лекция 1.

Предварительные замечания из теории вероятностей.

Биномиальное распределение.

Если слегка согнутая монета при бросании имеет вероятность выпадения герба f и вероятность выпадения решки $(1-f)$, то вероятность выпадения r гербов из N бросков

$$P(r|f,N) = C_N^r f^r (1-f)^{N-r} \quad , \tag{1.1}$$

где $C_N^r = \binom{N}{r} = \frac{N!}{(N-r)!\,r!}$ - биномиальный коэффициент. Поскольку есть распределение (1.1), то можно вычислить среднее значение

$$\langle r \rangle = E[r] = \sum_{r=0}^{N} r\,P(r|f,N) \tag{1.2}$$

и дисперсию

$$D_r = E\left[(r-E[r])^2\right] = E[r^2]-(E[r])^2 = \sum_{r=0}^{N} r^2 P(r|f,N) - (E[r])^2 \quad . \tag{1.3}$$

Чтобы вычислить $\langle r \rangle$ и D_r не прибегая к сложностям суммирования рядов, заметим, что речь идет о независимых событиях и r это сумма N независимых случайных величин (число гербов в каждом единичном броске с первого по N -ый, причем это число может быть либо 0, либо 1). Следовательно,

$$\langle r \rangle = \langle r_1 \rangle + \langle r_2 \rangle + ... + \langle r_N \rangle \quad \text{и} \quad D_r = D_{r_1} + D_{r_2} + ... + D_{r_N} \quad ,$$

причем $\langle r_i \rangle = 1\cdot f + 0\cdot(1-f) = f$ и $D_{r_i} = \langle r_i^2 \rangle - \langle r_i \rangle^2 = [1^2\cdot f + 0^2\cdot(1-f)] - f^2 = f(1-f)$. Отсюда имеем

$$\langle r \rangle = Nf \quad \text{и} \quad D_r = Nf(1-f) \quad . \tag{1.4}$$

Аппроксимация $x!$ и C_N^r .

Формула Стирлинга (-Муавра):

$$x! = \sqrt{2\pi x}\cdot x^x e^{-x} e^{\frac{\theta}{12x}}, \quad 0<\theta<1 \ .$$

В различных порядках приближения используют

$$x! \approx x^x e^{-x} \ , \tag{1.5a}$$

$$x! \approx \sqrt{2\pi x}\ x^x e^{-x} \ , \tag{1.5b}$$

$$\ln x! \approx x\ln x \ - \ x \ , \tag{1.5c}$$

$$\ln x! \approx x\ln x \ - \ x \ + \ \frac{1}{2}\ln(2\pi x) \ . \tag{1.5d}$$

Воспользуемся приближением (1.5c) для вычисления логарифма биномиального коэффициента:

$$\ln C_N^r = \ln\frac{N!}{(N-r)!r!} = \ln N! - \ln(N-r)! - \ln r! \approx (N-r)\ln\frac{N}{N-r} + r\ln\frac{N}{r} \ . \tag{1.6}$$

Выражение (1.6) верно для логарифмов по любому основанию, а стало быть верно и для логарифма по основанию 2 (можно просто воспользоваться тождеством $\log_2 x = \frac{\ln x}{\ln 2}$, чтобы убедиться в этом). Если ввести *функцию двоичной энтропии*

$$H_2(x) = x\log_2\frac{1}{x} + (1-x)\log_2\frac{1}{1-x} \ , \tag{1.7}$$

то аппроксимация для логарифма по основанию 2 биномиального коэффициента может быть записана как

$$\log_2 C_N^r \approx N H_2\left(\frac{r}{N}\right) \ , \tag{1.8}$$

или для самого биномиального коэффициента как

$$C_N^r \approx 2^{NH_2(r/N)} \ . \tag{1.9}$$

Если бы мы воспользовались аппроксимацией (1.5d), то вместо (1.8) получили бы более точное приближение:

$$\log_2 C_N^r \approx NH_2\left(\frac{r}{N}\right) - \frac{1}{2}\log_2\left[2\pi N \frac{N-r}{N}\,\frac{r}{N}\right] \quad . \qquad (1.10)$$

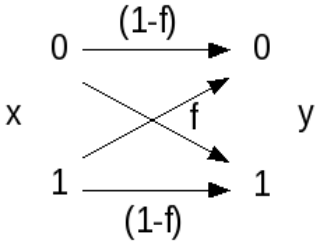
Введение в теорию информации

Основная проблема в области передачи и приема информации это как достичь идеально точную передачу информации через неидеальный шумящий канал связи. Вот несколько примеров когда встает подобная проблема:

- аналоговая телефонная линия, через которую два модема передают и принимают двоичную информацию;
- канал радиосвязи между межпланетным зондом (станцией) и центром управления на Земле;
- воспроизведение живой клетки по информации содержащейся в ДНК, которая в свою очередь скопирована из родительской ДНК;
- запись и чтение информации с жесткого диска компьютера (гибкого диска, CD, DVD и т.п.).

Телефонная линия подвержена перекрестным помехам от других линий в кабеле, радиоканал с межпланетным зондом подвержен влиянию фонового излучения космоса, шумам от звезд и планет, ДНК может мутировать, а проблемы с компьютерными носителями информации уже знакомы каждому. Во всех наших примерах принятое сообщение с какой-то вероятностью не будет идентично посланному. Нам же однако хотелось, чтобы эта вероятность была практически равна нулю.

Рассмотрим дисковод, который читает каждый бит правильно с вероятностью $(1-f)$ правильно и неправильно с вероятностью f . Эта модель известна как двоичный симметричный канал (binary symmetric channel). Обозначим как x отправленное сообщение, а как y - сообщение принятое на другом конце канала связи. Ниже приведено схематическое изображение этого канала и соответствующие вероятности передачи двоичного сообщения.

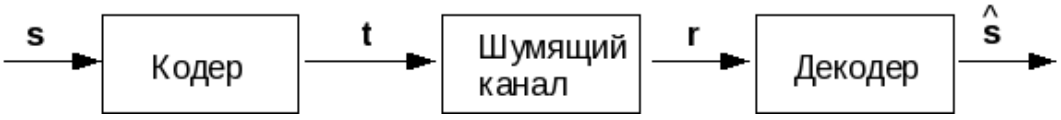


$$P(y=0|x=0) = 1-f; \quad P(y=0|x=1) = f;$$

$$P(y=1|x=0) = f; \quad P(y=1|x=1) = 1-f.$$

Предположим для определенности, что вероятность ошибки в нашем канале $f=0.1$ (10%). Если наш канал обеспечивает передачу данных между компьютером и жестким диском, то естественно ожидать, что вероятность

ошибки будет ничтожно мала, скажем при передаче гигабайта данных в день в течении 10 лет. Для этого неплохо бы иметь вероятность ошибки порядка 10^{-15} . Имеется 2 способа решить проблему. Первый способ это физическое решение, т.е. улучшить аппаратную часть дисковода, использовать более качественные электронные компоненты, откачать воздух, использовать больший размер домена на магнитной поверхности, использовать большую мощность сигнала и т.п. Все эти меры серьезно удорожат дисковод. Теория информации предлагает второй способ решения задачи, при котором шумящий канал связи остается таким как он есть, но к нему добавляется система (устройство), которая может обнаруживать и исправлять ошибки. Ниже схематически показано системное решение.



Здесь s - исходное сообщение, t - избыточное передаваемое сообщение, r - принятое на выходе канала сообщение и $\hat{s}$ - декодированное сообщение. Теория информации имеет дело с изучением теоретических ограничений подобного способа передачи информации. Теория же кодирования имеет дело с практическим воплощением подобных систем.

Коды с исправлением ошибок для двоичного симметричного канала

Рассмотрим наш пример. Что значит добавить избыточность? Мы не рассматриваем случай многократной передачи одного и того же сообщения. Ограничимся только однократной передачей и однократным приемом одного сообщения.

Коды повторения

Самое простое, что приходит в голову, это просто повторить каждый бит сообщения некоторое число раз (заранее заданное). Пусть это число 3, обозначим такой код как R_3. Пусть наше сообщение $s = 0\,0\,1\,0\,1\,1\,0$ передается через канал с $f = 0.1$, тогда вектор закодированного сообщения

$$t = 000\ 000\ 111\ 000\ 111\ 111\ 000 ,$$

к нему добавится случайный сигнал реализующий заданный уровень вероятности ошибки

$$\boldsymbol{n} = 000\ 00\mathbf{1}\ 000\ 000\ \mathbf{1}0\mathbf{1}\ 000\ 000$$

и на выходе канала связи будет получено сообщение

$$\boldsymbol{r} = 000\ 001\ 111\ 000\ 010\ 111\ 000 \quad .$$

Остается только его декодировать, например выбрать то число, которое встречается больше в каждой тройке чисел («голосование большинством»). Тогда получим следующий декодированный вектор:

$$\hat{\boldsymbol{s}} = 0\,\mathbf{0}\,1\,0\,\mathbf{0}\,1\,0 \quad .$$

Сравнивая декодированное сообщение с исходным и с вектором шума видим, что во втором числе сообщения ошибка была обнаружена и исправлена, однако в пятом числе она пропущена.

Пусть исходное сообщение $\boldsymbol{s}$ состоит только из одного числа (0 или 1), которое кодируется вектором $\boldsymbol{t}(\boldsymbol{s})$, на выходе из канала имеем вектор $\boldsymbol{r}=r_1r_2r_3$. Согласно формуле Байеса (или теореме гипотез), *апостериорная вероятность* **s**

$$P(\boldsymbol{s}|r_1r_2r_3) = \frac{P(r_1r_2r_3|\boldsymbol{s})P(\boldsymbol{s})}{P(r_1r_2r_3)} \quad . \tag{1.11}$$

Можно записать это выражение для каждой альтернативы отдельно:

$$P(s=1|r_1r_2r_3) = \frac{P(r_1r_2r_3|s=1)P(s=1)}{P(r_2r_2r_3)} \quad , \tag{1.12}$$

и

$$P(s=0|r_1r_2r_3) = \frac{P(r_1r_2r_3|s=0)P(s=0)}{P(r_2r_2r_3)} \quad . \tag{1.13}$$

Данная апостериорная вероятность определяется двумя множителями: *априорной вероятностью* $P(s)$ и множителем, зависящим от принятых данных $P(r_1r_2r_3|s)$, который называется *правдоподобием* s. Нормирующий множитель $P(r_1r_2r_3)$ в данном рассмотрении вычислять не понадобится, поскольку оптимальное решение принимается при сравнении (1.12) и (1.13), а этот множитель общий. Итак, решение принимается $\hat{s}=0$, если $P(s=0|\boldsymbol{r}) > P(s=1|\boldsymbol{r})$ и $\hat{s}=1$ в противном случае. Для того, чтобы с точностью до множителя определять эти сравниваемые апостериорные вероятности, необходимо иметь некоторое предположение об априорном распределении $P(s)$. Предположим, что априорные вероятности равны, т.е.

$P(s=0) = P(s=1) = 0.5$. Тогда максимизация апостериорной вероятности $P(s|\boldsymbol{r})$ эквивалентна максимизации правдоподобия $P(\boldsymbol{r}|s)$. Кроме того, предполагается что канал - двоичный симметричный канал с $f<0.5$, так что правдоподобие легко вычисляется:

$$P(\boldsymbol{r}|s) = P(\boldsymbol{r}|\boldsymbol{t}(s)) = \prod_{n=1}^{N} P(r_n|t_n(s)) \quad , \tag{1.14}$$

где в нашем случае $N=3$, каждый множитель в произведении (1.14)

$$P(r_n|t_n) = \begin{cases} (1-f), & если \;\; r_n=t_n \\ f, & если \;\; r_n\neq t_n \end{cases} \quad . \tag{1.15}$$

Таким образом, отношение правдоподобия двух гипотез

$$\frac{P(\boldsymbol{r}|s=1)}{P(\boldsymbol{r}|s=0)} \;=\; \prod_{n=1}^{N} \frac{P(r_n|t_n(1))}{P(r_n|t_n(0))} \quad . \tag{1.16}$$

В (1.16) каждый множитель в произведении равен $\frac{(1-f)}{f}$, если $r_n=1$ и $\frac{f}{(1-f)}$ если $r_n=0$. Отношение $\gamma = \frac{(1-f)}{f} > 0$, так что каждый «голос» в произведении (1.16) учитывается множителем γ .

Ошибка совершается кодом R_3, если 2 и более бит в блоке перевернутся. Вероятность ошибки суммируется из вероятность переворота всех 3-х бит и вероятности 2-х перевернутых бит. Используя биномиальное распределение (1.1) можно записать вероятность блочной ошибки:

$$P_b \;=\; P(r=3|f,N=3) + P(r=2|f,N=3) \;=\; f^3 + 3f^2(1-f) \;=\; 3f^2-2f^3 \quad . \tag{1.17}$$

Для малых f доминирует слагаемое $3f^2$, поэтому для $f=0.1$ вероятность ошибки кода R_3 $P_b \approx 0.03$. Однако в случае кода повторения, уменьшение ошибки втрое одновременно втрое уменьшило скорость передачи.

Оценим теперь сколько раз нужно повторять бит в каждом блоке кода повторения, чтобы достичь $P_b \sim 10^{-15}$. Следует отметить, что для кода повторения R_N, N должно быть нечетным, чтобы отношение правдоподобия никогда не было бы равным $\gamma^0 = 1$ и не возникала неопределенность что же выбирать «голосованием». Вероятность ошибки кода R_N определяется суммой аналогичной (1.17), в которой доминирующая вероятность:

$$P_b \;=\; C_N^{(N+1)/2} \, f^{(N+1)/2} \, (1-f)^{(N-1)/2} \quad . \tag{1.18}$$

Используя оценку (1.9) для биномиального коэффициента можно получить

$$C_N^{(N+1)/2} \approx 2^{NH_2\left(\frac{(N+1)/2}{N}\right)} = 2^{NH_2(0.5)} = 2^N ,$$

что в свою очередь дает следующую достаточно грубую оценку для вероятности ошибки

$$P_b \approx 2^N (f(1-f))^{N/2} = (4f(1-f))^{N/2} . \quad (1.19)$$

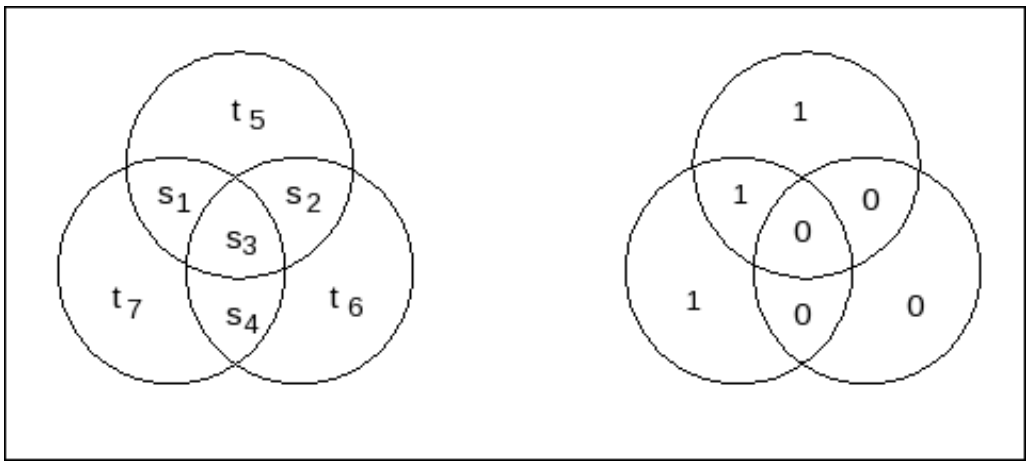
Решая уравнение

$$(4f(1-f))^{N/2} = 10^{-15} \quad (1.20)$$

для заданного уровня вероятности ошибки в канале $f = 0.1$, получим $N \approx 67.6$. Более точные численные расчеты с привлечением точной формулы Стирлинга-Муавра дают значение $N \approx 61$. Данное число в случае жесткого диска можно интерпретировать как увеличение фактической емкости диска. То есть вместо 1 Гб данных, нужно фактически хранить 61 Гб. Дорого и неэффективно.

Блочные коды — код Хэмминга (7,4)

Блочный код представляет собой правило преобразования исходного слова $\boldsymbol{s}$ длиной в K бит в передаваемую последовательность $\boldsymbol{t}$ длиной N бит. Для придания избыточности N должно быть больше K. В линейном коде дополнительные $N-K$ бит вычисляются как линейная функция исходных K бит. Простейшим примером блочного кода является *код Хэмминга*, который передает 7 бит на каждые 4 бита исходного сообщения. Удобно изобразить этот код графически, как это сделано ниже.



Первые четыре передаваемых бита $t_1 t_2 t_3 t_4$ совпадают с исходными битами $s_1 s_2 s_3 s_4$, а биты четности $t_5 t_6 t_7$ выбираются так, чтобы сумма внутри каждого круга была четной. Преобразование исходного слова $\boldsymbol{s}$ в передаваемое $\boldsymbol{t}$ можно выполнить с помощью линейной операции

$$\boldsymbol{t} = \boldsymbol{G}^T \boldsymbol{s} \ , \tag{1.21}$$

где векторы $\boldsymbol{t}$ и $\boldsymbol{s}$ рассматриваются как векторы столбца, $\boldsymbol{G}$ - *матрица-генератор* кода,

$$\boldsymbol{G}^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{I_4} \\ \boldsymbol{P} \end{pmatrix} \ , \tag{1.22}$$

где $\boldsymbol{I_4}$ - единичная матрица размера 4×4, а $\boldsymbol{P}$ - оставшаяся часть матрицы $\boldsymbol{G}^T$ размера 3×4. Все арифметические операции предполагаются по модулю 2 (т.е. $1+1=0$, $0+1=1$, ...). Если в процессе передачи вектора $\boldsymbol{t}$ произошла одна ошибка и в векторе $\boldsymbol{r}$ перевернут один бит, либо информационный либо бит четности, то несоответствие в кругах в графическом представлении выше по четности будет однозначно характеризовать какой именно бит перевернулся. Разница между принятыми битами $r_5 r_6 r_7$ и вычисленными на основе принятых $r_1 r_2 r_3 r_4$ дает вектор синдрома $\boldsymbol{z}$, который в свою очередь однозначно характеризует какой именно бит перевернулся в процессе передачи. Если синдром 000, то это означает, что ошибок нет и вектор $r_1 r_2 r_3 r_4$ с наибольшей вероятностью и есть вектор исходного сообщения $s_1 s_2 s_3 s_4$. Ниже приведена таблица соответствия синдромов и перевернутых бит.

Синдром $\boldsymbol{z}$	000	001	010	011	100	101	110	111
Перевернутый бит	нет	r_7	r_6	r_4	r_5	r_1	r_2	r_3

Вычисление синдрома это линейная операция

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{r} \ , \tag{1.23}$$

где проверочная матрица определяется как

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{P} & \boldsymbol{I}_3 \end{pmatrix} \quad . \tag{1.24}$$

Все переданные слова $\boldsymbol{t} = \boldsymbol{G}^T \boldsymbol{s}$ удовлетворяют условию

$$\boldsymbol{H}\,\boldsymbol{t} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad . \tag{1.25}$$

Поскольку принятый вектор может быть представлен как $\boldsymbol{r} = \boldsymbol{G}^T \boldsymbol{s} + \boldsymbol{n}$, декодирование с помощью синдромов эквивалентно нахождению такого шума $\boldsymbol{n}$, который удовлетворяет уравнению

$$\boldsymbol{H}\,\boldsymbol{n} = \boldsymbol{z} \quad . \tag{1.26}$$

Декодирующий алгоритм, решающий данную задачу можно назвать *декодером максимального правдоподобия*. Вероятность ошибки в блоке (из 4 бит) это вероятность того, что один или несколько бит в декодированном слове не совпадут с битами исходного слова, т.е.

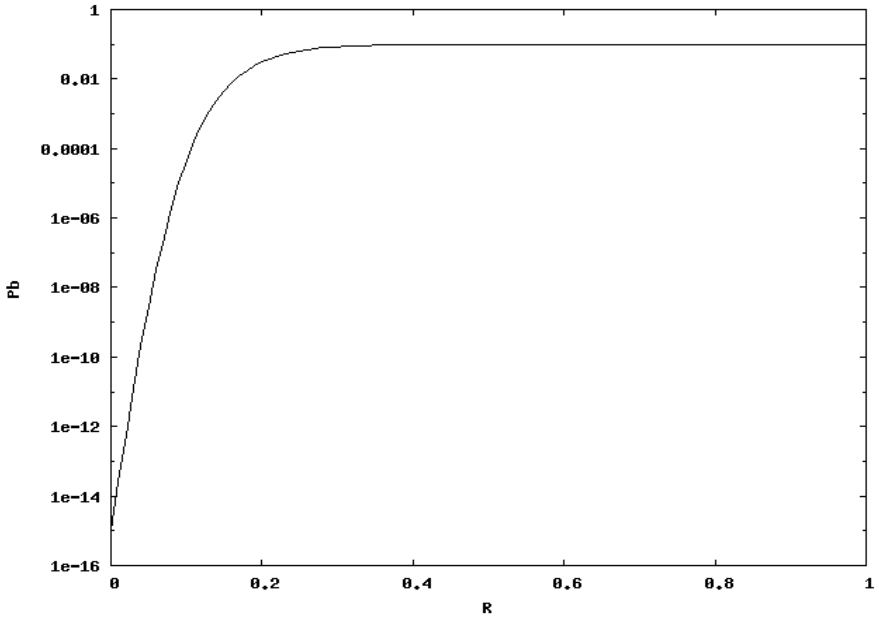
$$P_b = P(\hat{\boldsymbol{s}} \neq \boldsymbol{s}) \quad . \tag{1.27}$$

Вероятность битовой ошибки представляет собой среднюю вероятность битовой ошибки в каждом из K битов, т.е.

$$p_b = \frac{1}{K} \sum_{k=1}^{K} P(\hat{s}_k \neq s_k) \quad . \tag{1.28}$$

В случае кода Хэмминга (7,4) ошибка происходит тогда, когда в блоке из 7 бит происходит переворачивание 2-х и более бит. Вероятность ошибки имеет тот же порядок $O(f^2)$ что и для кода R_3, однако передача осуществляется заметно быстрее и замедление составляет всего 4/7.

Обобщение кода Хэмминга дает семейство кодов, которые лучше кодов повторения, однако достижение желаемой вероятности блочной ошибки в 10^{-15} все равно приводит к существенному замедлению передачи. Ниже схематически изображена зависимость вероятности блочной ошибки P_b от замедления R (отношение числа переданных битов к числу битов в исходном слове) для кодов повторения.



В плоскости этого графика точки соответствующие кодам Хэмминга находятся под кривой, поскольку они лучше кодов повторения. Можно предположить, что плоскость $R-P_b$ может быть разделена некоторой кривой на достижимую (слева сверху) и недостижимую (справа снизу) области. До Клода Шеннона (в 1948 г.) считалось, что такая кривая проходить через начало координат. Однако Шеннон доказал, что граница между достижимыми и недостижимыми кодами пересекает ось R в точке $C>0$. Максимальное замедление $R=C$, достижимое для данного канала, при которой достигается пренебрежимо малая вероятность ошибки, называется *емкостью канала*.

Для двоичного симметричного канала

$$C(f) = 1-H_2(f) = 1-\left[f\log_2\frac{1}{f} + (1-f)\log_2\frac{1}{1-f}\right] \quad . \tag{1.29}$$

Подставляя в (1.29) значение $f=0.1$, получим $C(0.1)\approx 0.53$. Это значит, что для достижения ошибки в 10^{-15} чтения/записи для диска в 1Гб, понадобится не 61 шумящий диск, а всего 2. Однако нет никакой гарантии, что длина кода не окажется длиннее емкости диска.

Теорему кодирования через шумящий канал Шеннона можно сформулировать следующим образом:

Информация может быть передана через шумящий канал с ненулевой скоростью с пренебрежимо малой вероятностью ошибки.