Теория информации

Лекция 8

Вероятность входного сигнала при условии принятого на выходе

Предполагая то, что случайная величина x на входе принадлежит ансамблю X (с его A_x и $P(x)$) и имея все переходные вероятности $P(y|x)$ (компоненты матрицы $\boldsymbol{Q}$), можно для совместного ансамбля XY определить совместное распределение

$$P(x,y) = P(y|x)P(x) \quad . \tag{8.1}$$

Теперь можно определить вероятность посылки символа x при условии, что на выходе канала получен символ y. Это можно сделать если воспользоваться формулой Байеса:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} = \frac{P(y|x)P(x)}{\sum_{x'} P(y|x')P(x')} \quad . \tag{8.2}$$

Пример

Пусть $A_x=\{0,1\}$, $A_y=\{0,1\}$, $P_x=\{0.9,0.1\}$ и передача осуществляется через двоичный симметричный канал с $f=0.15$. Тогда, подставляя в (8.2),

$$P(x=1|y=1)=\frac{P(y=1|x=1)P(x=1)}{\sum_{x'} P(y=1|x')P(x')}=\frac{0.85\times0.1}{0.15\times0.9+0.85\times0.1}\approx 0.39 \quad ,$$

$$P(x=0|y=1)=\frac{P(y=1|x=0)P(x=0)}{\sum_{x'} P(y=1|x')P(x')}=\frac{0.15\times0.9}{0.15\times0.9+0.85\times0.1}\approx 0.61 \quad .$$

В этом примере $x=0$ при принятом $y=1$ все еще более вероятное значение, чем $x=1$, однако оно уже более вероятное чем было в ансамбле X. Если же этот канал сделать в 10 раз менее шумящим, т.е. $f=0.015$, то

$$P(x=1|y=1)\approx 0.88 \quad ,$$

$$P(x=0|y=1)\approx 0.12 \quad .$$

Таким образом, можно заметить, что условные вероятности зависят от соотношения между вероятностями P_x и f .

Информация переносимая каналом

Рассмотрим как много информации может быть передано через тот или иной канал связи. Взаимная информация

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

характеризует уменьшение неопределенности относящейся к случайной переменной x после того как принята случайная переменная y , в этом смысле понятно первое равенство, однако практически вычислять легче используя последнее равенство. В качестве примера рассмотрим тот же двоичный симметричный канал.

Пример

$$A_x = \{0,1\} \quad , \quad P_x = \{0.9, 0.1\} \quad , \quad f = 0.15 \quad .$$

Для вычисления безусловной распределения вероятности P_y для ансамбля $A_y = \{0,1\}$ можно воспользоваться (7.12) или результатом предыдущего примера (знаменатель в (8.2)):

$$\boldsymbol{p}_y = \boldsymbol{Q}\,\boldsymbol{p}_x = \begin{pmatrix} 1-f & f \\ f & 1-f \end{pmatrix} \boldsymbol{p}_x = \begin{pmatrix} 0.85 & 0.15 \\ 0.15 & 0.85 \end{pmatrix} \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix} = \begin{pmatrix} 0.78 \\ 0.22 \end{pmatrix} \quad .$$

Теперь можно вычислить взаимную информацию:

$$I(X;Y) = H(Y) - H(Y|X) = \sum_{y \in A_y} P(y) \log_2 \frac{1}{P(y)} - \sum_{x \in A_x} P(x) \sum_{y \in A_y} P(y|x) \log_2 \frac{1}{P(y|x)} \quad .$$

Первое слагаемое в этом выражении легко вычислить, поскольку

$$\sum_{y \in A_y} P(y) \log_2 \frac{1}{P(y)} = 0.78 \log_2 \frac{1}{0.78} + 0.22 \log_2 \frac{1}{0.22} = H_2(0.22) = H_2(0.78) \quad .$$

Условную энтропию (второе слагаемое в выражении для взаимной информации) можно представить как

$$H(Y|X) = \sum_{x \in A_x} P(x) H(Y|x) \quad ,$$

где

$$H(Y \mid x=0) = (1-f) \log_2 \frac{1}{1-f} + f \log_2 \frac{1}{f} = H_2(1-f) = \quad ,$$

$$= H(Y \mid x=1) = f \log_2 \frac{1}{f} + (1-f) \log_2 \frac{1}{1-f} = H_2(f) = \quad ,$$

$$= H_2(0.15) = H_2(0.85) \quad .$$

Поскольку в сумме можно вынести за знак суммирования $H_2(f)$, то и $H(Y|X)=H_2(f)$. Таким образом,

$$I(X;Y) = H_2(0.22) - H(0.15) = 0.76 - 0.61 = 0.15 \quad \text{бит.}$$

При этом $H(X) = H_2(0.1) \approx 0.47$. Если посчитать взаимную информацию $I(X;Y)$ для Z-канала, то можно получить величину 0.36 бит, что больше чем величина полученная для двоичного симметричного канала. Это означает, что Z-канал более надежен при той же вероятности ошибки f .

Максимизация взаимной информации

Ясно, что взаимная информация между переменными на входе и на выходе канала зависит от выбора входного ансамбля. Таким образом, можно предположить, что можно выбрать такой входной ансамбль, что $I(X;Y)$ окажется максимальной. Определим *емкость канала* $C(Q)$ как максимально возможная для него взаимная информация, т.е.

$$C(Q) = \max_{P_x} I(X;Y) \quad , \qquad (8.3)$$

а распределение P_x на котором достигается этот максимум назовем *оптимальным распределением*, обозначив как P_x^* . Следует отметить, что это распределение не обязательно единственное.

В рассмотренном выше примере двоичного симметричного канала с $f=0.15$ можно осуществить максимизацию взаимной информации явно. Зададим ансамбль вероятностями $P_x=\{p_0, p_1\}$ и посмотрим при каком наборе вероятностей взаимная информация окажется максимальной. С учетом уже

сделанных выкладок, можно записать

$$I(X;Y) = H_2\left((1-f \quad f)\cdot\begin{pmatrix} p_0 \\ p_1 \end{pmatrix}\right) - H_2(f) =$$

(здесь в качестве аргумента первой функции H_2 матричное произведение первой строки матрицы $\boldsymbol{Q}$ на вектор $\boldsymbol{p}_x$ согласно выражению (7.12))

$$= H_2((1-f)(1-p_1)+f\,p_1) - H_2(f) \quad .$$

Это выражение достигает своего максимума в максимуме функции $H_2(0.5)=1$, вероятность p_1 можно найти из уравнения

$$(1-f)(1-p_1)+f\,p_1 = 0.5 \quad ,$$

решением которого является $p_1=0.5$, и следовательно, $p_0=1-p_1=0.5$. Таким образом, для двоичного симметричного канала (дск)

$$C(Q_{дск}) = I(X;Y)$$

если $P_x^* = \left\{\frac{1}{2}, \frac{1}{2}\right\}$. Для $f=0.15 \qquad C(Q_{дск})\approx 0.39$.

Если провести вычисления для Z-канала и той же вероятности ошибки, что несколько сложнее, то можно получить оптимальное распределение $P_x^*=\{0.555,\ 0.445\}$ и емкость канала $C(Q_Z)=0.685$. Можно сказать что максимальная скорость передачи может быть достигнута, если использовать 0 чуть чаще чем 1.

Теорема кодирования для канала с шумом (предварительные замечания)

Хотя взаимная информация кажется подходящей мерой информации переносимой каналом, совершенно не очевидно, что эта величина действительно характеризует связь с помощью блочных кодов через канал с *пренебрежимо малой вероятностью ошибки*. Для начала некоторые определения.

Блочный код (N,K) для канала *Q* это список $S=2^K$ кодовых слов длиной *N*

$$\{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(2^K)}\}, \qquad \boldsymbol{x}^{(s)} \in A_x^N \quad .$$

При этом число кодовых слов целое, однако $K \equiv \log_2 S$ не обязательно целое.

Используя этот код можно закодировать сигнал $s \in \{1,2,3,\ldots,2^K\}$ кодовым словом $\boldsymbol{x}^{(s)}$. Скорость работы этого канала можно характеризовать отношением $R = K/N$ или долей кода в каждом бите передаваемом через канал.

Декодер блочного кода *(N,K)* это отображение множество строк длиной *N*, A_y^N на номер кодового слова $\hat{s} \in \{0,1,2,3,\ldots,2^K\}$. Здесь по сравнению с s появился дополнительный 0, означающий ошибку декодирования.

Вероятность блочной ошибки кода и декодера для заданных канала и распределения вероятности кодируемого сигнала $P(s_{вх})$ это

$$p_B = \sum_{s_{вх}} P(s_{вх}) P(s_{вых} \neq s_{вх} \mid s_{вх}) \quad . \tag{8.4}$$

Максимальная вероятность блочной ошибки это

$$p_{BM} = \max_{s_{вх}} P(s_{вых} \neq s_{вх} \mid s_{вх}) \tag{8.5}$$

Оптимальный декодер минимизирует вероятность блочной ошибки. Он декодирует слово на выходе канала $\boldsymbol{y}$ как входной сигнал s так, что достигается максимум апостериорной вероятности

$$P(s \mid \boldsymbol{y}) = \frac{P(\boldsymbol{y} \mid s) P(s)}{\sum_{s'} P(\boldsymbol{y} \mid s') P(s')} \quad , \tag{8.6}$$

то есть

$$\hat{s}_{опт} = \arg\max P(s \mid \boldsymbol{y}) \quad . \tag{8.7}$$

В случае равномерного распределения $P(s)$ такой декодер называют *декодером максимального правдоподобия*, поскольку он соответствует максимуму величины $P(\boldsymbol{y} \mid s)$.

Вероятность битовой ошибки p_b определяется в предположении, что кодовое слово номер s представлено двоичным вектором $\boldsymbol{s}$ (строкой, числом) длиной K бит. Это средняя вероятность того, что бит в $\boldsymbol{s}_{вых}$ не совпадет с соответствующим битом в $\boldsymbol{s}_{вх}$ (усреднение по всем K битам).

В заключение, сформулируем первую часть *теоремы кодирования Шеннона для канала с шумом*. Для каждого дискретного канала без памяти существует

неотрицательное число C, называемое емкостью канала, такое что для каждого малого $\varepsilon > 0$ и каждого $R < C$, для достаточно большого N существует блочный код длиной N с долей кода $\frac{K}{N} \geq R$ и алгоритм декодирования такие что $P_{BM} < \varepsilon$.