Прямая и обратная теорема побуквенного кодирования

~~5 января 2023 г. 1:21~~

Средняя длина кодового слова - матожидание от длины кодового слова:

$\overline{l} = E[l_i] = \sum_{i=1}^{M} p_i l_i$

H(X) - нижняя граница для средней длины кодового слова

Прямая теорема неравномерного побуквенного кодирования отражает связь энтропии и длины кодового слова, формулируется следующим образом:

Theorem

Для ансамбля $X = \{x, p(x)\}$ с энтропией $H(X) = H$ существует побуквенный неравномерный префиксный код со средней длиной кодовых слов $\bar{l} < H + 1$.

Пусть имеется код с длинами кодовых символов $l_i = \lceil -log\,g(p_i) \rceil$. *(код Шеннона?)*

Докажем, что для него выполняется неравенство Крафта (т.е. что существует префиксный код объемом M с заданными длинами кодовых слов):

$\sum_{i=1}^{M} 2^{-l_i} = \sum_{i=1}^{M} 2^{-\lceil -log\,g(p_i) \rceil} \le \sum_{i=1}^{M} 2^{log\,g(p_i)} = \sum_{i=1}^{M} p_i = 1$

Итак, такой префиксный код существует

Определим среднюю длину кодовых слов:

$\overline{l} = \sum_{m=1}^{M} p_m l_m < \sum_{m=1}^{M} p_m(-log\,g(p_m) + 1) = H + 1$

$\overline{l} < H + 1$, доказано

Обратная теорема неравномерного побуквенного кодирования

<, т.к. заменяем ⌈⌉ на +1

Theorem

Для любого однозначно декодируемого кода для дискретного источника $\{X, p(x)\}$ с энтропией H, $\bar{l} \ge H$.

Доказательство.

$$
\begin{aligned}
H - \bar{l} &= -\sum_{x \in X} p(x) \log p(x) - \sum_{x \in X} p(x) l(x) \\
&= \sum_{x \in X} p(x) \log \frac{2^{-l(x)}}{p(x)} \le \log e \sum_{x \in X} p(x) \left(\frac{2^{-l(x)}}{p(x)} - 1 \right) \\
&\le \log e \left(1 - \sum_{x \in X} p(x) \right) = 0
\end{aligned}
$$

□

При сжатии средние затраты в битах на символ (средняя длина кодового слова) должны находиться в диапазоне от H до H+1

Для префиксного:

$H \le \bar{l} < H + 1$

Для однозначно декодируемого?

$H \le \bar{l} \le H + 1$

Будем рады получить ваш отзыв!

Не могли бы вы ответить всего на два вопроса?