Лекция 13. Алфавитные коды. Неравенство Макмиллана. Теорема о существовании префиксного кода с заданными длинами кодовых слов. Дерево префиксного кода.

Лектор — Селезнева Светлана Николаевна
selezn@cs.msu.ru

факультет ВМК МГУ имени М.В. Ломоносова

Лекции на сайте https://mk.cs.msu.ru

Неравенство Макмиллана

Теорема 13.1 (неравенство Макмиллана). *Пусть $C_\varphi = \{B_1, \ldots, B_r\}$ — алфавитный код в кодирующем алфавите B, $|B| = q$, и $|B_i| = l_i$, $i = 1, \ldots, r$. Если код C_φ — разделим, то верно неравенство:*

$$\sum_{i=1}^{r} \frac{1}{q^{l_i}} \leqslant 1.$$

Неравенство Макмиллана

Доказательство. Пусть $n \geqslant 1$. Рассмотрим выражение:

$$\left(\sum_{i=1}^{r} \frac{1}{q^{l_i}}\right)^n .$$

Получаем:

$$\left(\sum_{i=1}^{r} \frac{1}{q^{l_i}}\right)^n = \left(\sum_{i_1=1}^{r} \frac{1}{q^{l_{i_1}}}\right) \cdot \left(\sum_{i_2=1}^{r} \frac{1}{q^{l_{i_2}}}\right) \cdot \ldots \cdot \left(\sum_{i_n=1}^{r} \frac{1}{q^{l_{i_n}}}\right) =$$
$$= \sum_{i_1=1}^{r} \sum_{i_2=1}^{r} \ldots \sum_{i_n=1}^{r} \frac{1}{q^{l_{i_1}+l_{i_2}+\ldots l_{i_n}}} = \sum_{k=1}^{n \cdot l_{\max}} \frac{c_k}{q^k},$$

где $l_{\max} = \max\limits_{1 \leqslant i \leqslant r} l_i$ и c_k равно числу таких наборов $(i_1, \ldots, i_n)$, что $l_{i_1} + \ldots + l_{i_n} = k$ (для каждого $k = 1, \ldots, n \cdot l_{\max}$).

Вспомогательная лемма

Лемма 13.1. *Если* C_φ *— разделимый алфавитный код, то* $c_k \leqslant q^k$.

Доказательство. Итак, c_k равно числу таких наборов $(i_1, \ldots, i_n)$, что $l_{i_1} + \ldots + l_{i_n} = k$.

Каждому такому набору $(i_1, \ldots, i_n)$ соответствует слово $\alpha = a_{i_1} \ldots a_{i_n} \in A^*$ (где A — исходный алфавит).

Далее:

$$\varphi(\alpha) = \varphi(a_{i_1} \ldots a_{i_n}) = B_{i_1} \ldots B_{i_n} = \beta,$$

причем $|\beta| = l_{i_1} + \ldots + l_{i_n} = k$.

Вспомогательная лемма

Но код C_φ — разделим, поэтому если $\beta \in B^*$, то найдется не более одного такого слова $\alpha \in A^*$, что $\varphi(\alpha) = \beta$.

Поэтому любому слову $\beta \in B^*$, $|\beta| = k$, соответствует не более одного такого слова $\alpha = a_{i_1} \ldots a_{i_n} \in A^*$, что $\beta = \varphi(\alpha)$.

А значит, число таких наборов $(i_1, \ldots, i_n)$, что $l_{i_1} + \ldots + l_{i_n} = k$, не превосходит числа слов длины k в алфавите B, т. е. $c_k \leqslant q^k$.

$\square$

Неравенство Макмиллана

Доказательство теоремы 13.1 (продолжение). Итак,

$$\left(\sum_{i=1}^{r} \frac{1}{q^{l_i}}\right)^n = \sum_{k=1}^{n \cdot l_{\max}} \frac{c_k}{q^k}.$$

По лемме 13.1 верно $c_k \leqslant q^k$, поэтому

$$\left(\sum_{i=1}^{r} \frac{1}{q^{l_i}}\right)^n \leqslant \sum_{k=1}^{n \cdot l_{\max}} 1 \leqslant n \cdot l_{\max},$$

или

$$\sum_{i=1}^{r} \frac{1}{q^{l_i}} \leqslant \sqrt[n]{n \cdot l_{\max}}.$$

Неравенство Макмиллана

Неравенство

$$\sum_{i=1}^{r} \frac{1}{q^{l_i}} \leqslant \sqrt[n]{n \cdot l_{\max}}$$

выполняется для любого $n \geqslant 1$. Переходя в нем к пределу при $n \to \infty$, получаем:

$$\sum_{i=1}^{r} \frac{1}{q^{l_i}} \leqslant 1.$$

$\square$

Неравенство Макмиллана

Пример. Является ли разделимым алфавитный код

$$C_{\varphi_1} = \{00, 01, 10, 001, 011, 100\}?$$

Находим кодирующий алфавит: $B = \{0, 1\}$.

Получаем:

$$\frac{3}{2^2} + \frac{3}{2^3} = \frac{3}{4} + \frac{3}{8} = \frac{9}{8} > 1.$$

Если бы код C_{φ_1} был разделим, то сумма в левой части не превосходила бы единицу, что не так. Значит, код C_{φ_1} не является разделимым.

Неравенство Макмиллана

Пример. Является ли разделимым алфавитный код

$$C_{\varphi_2} = \{00, 01, 10, 001, 011\}?$$

Находим кодирующий алфавит: $B = \{0, 1\}$.

Получаем:

$$\frac{3}{2^2} + \frac{2}{2^3} = \frac{3}{4} + \frac{2}{8} = 1.$$

Пока невозможно сделать вывод о разделимости или неразделимости кода C_{φ_2}.

Неравенство Макмиллана

Пример (продолжение). Построим граф $G_{\varphi_2} = (V_{\varphi_2}, E_{\varphi_2})$ для кода $C_{\varphi_2} = \{00, 01, 10, 001, 011\}$. Получаем: $V_{\varphi_2} = \{\Lambda, 0, 1\}$.

Далее:

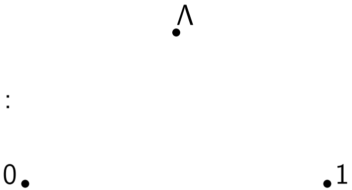
G_{φ_2} :

Неравенство Макмиллана

Пример (продолжение). Построим граф $G_{\varphi_2} = (V_{\varphi_2}, E_{\varphi_2})$ для кода $C_{\varphi_2} = \{00, 01, 10, 001, 011\}$. Получаем: $V_{\varphi_2} = \{\Lambda, 0, 1\}$.

Далее:

G_{φ_2} :

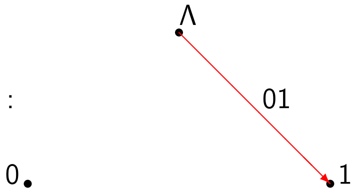


Неравенство Макмиллана

Пример (продолжение). Построим граф $G_{\varphi_2} = (V_{\varphi_2}, E_{\varphi_2})$ для кода $C_{\varphi_2} = \{00, 01, 10, 001, 011\}$. Получаем: $V_{\varphi_2} = \{\Lambda, 0, 1\}$.

Далее:

G_{φ_2} :

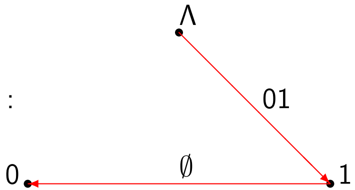


Неравенство Макмиллана

Пример (продолжение). Построим граф $G_{\varphi_2} = (V_{\varphi_2}, E_{\varphi_2})$ для кода $C_{\varphi_2} = \{00, 01, 10, 001, 011\}$. Получаем: $V_{\varphi_2} = \{\Lambda, 0, 1\}$.

Далее:

G_{φ_2} :

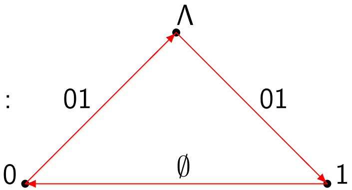


Неравенство Макмиллана

Пример (продолжение). Построим граф $G_{\varphi_2} = (V_{\varphi_2}, E_{\varphi_2})$ для кода $C_{\varphi_2} = \{00, 01, 10, 001, 011\}$. Получаем: $V_{\varphi_2} = \{\Lambda, 0, 1\}$.

Далее:

G_{φ_2} :

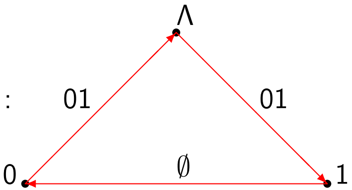


Неравенство Макмиллана

Пример (продолжение). Построим граф $G_{\varphi_2} = (V_{\varphi_2}, E_{\varphi_2})$ для кода $C_{\varphi_2} = \{00, 01, 10, 001, 011\}$. Получаем: $V_{\varphi_2} = \{\Lambda, 0, 1\}$.

Далее:

G_{φ_2} :



В графе G_{φ_2} найдется направленный цикл, проходящий через вершину Λ, значит, код C_{φ_2} — не является разделимым.

Неравенство Макмиллана

Пример. Существует ли разделимый алфавитный код в кодирующем алфавите из $q = 3$ букв с длинами кодовых слов:

$$1,\ 1,\ 2,\ 2,\ 2,\ 2?$$

Получаем:

$$\frac{2}{3^1} + \frac{4}{3^2} = \frac{2}{3} + \frac{4}{9} = \frac{10}{9} > 1.$$

Если бы такой код существовал, то сумма в левой части не превосходила бы единицу, что не так. Значит, такого разделимого кода не существует.

Неравенство Макмиллана

Пример. Существует ли разделимый алфавитный код в кодирующем алфавите из $q = 3$ букв с длинами кодовых слов:

$$1,\ 2,\ 2,\ 3,\ 3,\ 3?$$

Получаем:

$$\frac{1}{3^1} + \frac{2}{3^2} + \frac{3}{3^3} = \frac{1}{3} + \frac{2}{9} + \frac{3}{27} = \frac{2}{3} \leqslant 1.$$

Противоречия нет. Но найдется ли такой разделимый код?

Префиксный код с заданными длинами кодовых слов

Теорема 13.2 (о существовании префиксного кода с заданными длинами кодовых слов). *Пусть* $q, l_1, \ldots, l_r$ *— такие натуральные числа, что выполняется неравенство:*

$$\sum_{i=1}^{r} \frac{1}{q^{l_i}} \leqslant 1.$$

Тогда существует такой **префиксный** *код* $C = \{B_1, \ldots, B_r\}$ *в любом кодирующем алфавите из* q *букв, что* $|B_i| = l_i$ *для всех* $i = 1, \ldots, r$*.*

Префиксный код с заданными длинами кодовых слов

Доказательство. Итак, пусть $q, l_1, \ldots, l_r \geqslant 1$,

$$\sum_{i=1}^{r} \frac{1}{q^{l_i}} \leqslant 1$$

и B — произвольный кодирующий алфавит из q букв.

Пусть $m_1, \ldots, m_k$ — все различные числа среди чисел $l_1, \ldots, l_r$, $1 \leqslant k \leqslant r$, причем чисел m_i среди $l_1, \ldots, l_r$ ровно r_i, $i = 1, \ldots, k$. Отметим, что

$$r_1 + \ldots + r_k = r.$$

Значит,

$$\sum_{j=1}^{k} \frac{r_j}{q^{m_j}} \leqslant 1.$$

Пусть, для определенности, $m_1 < m_2 < \ldots < m_k$.

Префиксный код с заданными длинами кодовых слов

Следовательно, выполняется система неравенств:

$$\begin{cases} \frac{r_1}{q^{m_1}} \leqslant 1, \\ \frac{r_1}{q^{m_1}} + \frac{r_2}{q^{m_2}} \leqslant 1, \\ \frac{r_1}{q^{m_1}} + \frac{r_2}{q^{m_2}} + \frac{r_3}{q^{m_3}} \leqslant 1, \\ \ldots, \\ \frac{r_1}{q^{m_1}} + \frac{r_2}{q^{m_2}} + \ldots + \frac{r_k}{q^{m_k}} \leqslant 1, \end{cases}$$

откуда

$$\begin{cases} r_1 \leqslant q^{m_1}, \\ r_2 \leqslant q^{m_2} - r_1 q^{m_2 - m_1}, \\ r_3 \leqslant q^{m_3} - r_2 q^{m_3 - m_2} - r_1 q^{m_3 - m_1}, \\ \ldots, \\ r_k \leqslant q^{m_k} - r_{k-1} q^{m_k - m_{k-1}} - \ldots - r_1 q^{m_k - m_1}. \end{cases}$$

Префиксный код с заданными длинами кодовых слов

Итак, $r_1 \leqslant q^{m_1}$.

Выберем r_1 различных слов $B_1, \dots, B_{r_1}$ длины m_1 в алфавите B.

Всего различных слов длины m_1 в алфавите B найдется q^{m_1}.

Из $r_1 \leqslant q^{m_1}$ следует, что r_1 различных слов длины m_1 в алфавите B можно найти.

Из дальнейшего рассмотрения исключим все слова в алфавите B с префиксами $B_1, \dots, B_{r_1}$.

Префиксный код с заданными длинами кодовых слов

Теперь, $r_2 \leqslant q^{m_2} - r_1 q^{m_2 - m_1}$.

Выберем r_2 различных слов $B_{r_1+1}, \ldots, B_{r_1+r_2}$ длины m_2 в алфавите B, не начинающихся с $B_1, \ldots, B_{r_1}$.

Всего различных слов длины m_2 в алфавите B найдется q^{m_2}. Из них содержат одно из слов $B_1, \ldots, B_{r_1}$ как префикс в точности $r_1 q^{m_2 - m_1}$ слов.

Но $r_2 \leqslant q^{m_2} - r_1 q^{m_2 - m_1}$, поэтому r_2 различных слов длины m_2 с такими условиями можно найти.

Из дальнейшего рассмотрения исключим все слова в алфавите B с префиксами $B_1, \ldots, B_{r_1}$, $B_{r_1+1}, \ldots, B_{r_1+r_2}$.

Префиксный код с заданными длинами кодовых слов

Далее, $r_3 \leqslant q^{m_3} - r_2 q^{m_3 - m_2} - r_1 q^{m_3 - m_1}$.

Выберем r_3 различных слов $B_{r_1+r_2+1}, \dots, B_{r_1+r_2+r_3}$ длины m_3 в алфавите B, не начинающихся с $B_1, \dots, B_{r_1}$, $B_{r_1+1}, \dots, B_{r_1+r_2}$.

Всего различных слов длины m_3 в алфавите B найдется q^{m_3}. Из них содержат одно из слов $B_1, \dots, B_{r_1}$, $B_{r_1+1}, \dots, B_{r_1+r_2}$ как префикс в точности $r_1 q^{m_3 - m_1} + r_2 q^{m_3 - m_2}$ слов.

Но $r_3 \leqslant q^{m_3} - r_2 q^{m_3 - m_2} - r_1 q^{m_3 - m_1}$, поэтому r_3 различных слов длины m_3 с такими условиями можно найти.

Из дальнейшего рассмотрения исключим все слова в алфавите B с префиксами $B_1, \dots, B_{r_1}$, $B_{r_1+1}, \dots, B_{r_1+r_2}$, $B_{r_1+r_2+1}, \dots, B_{r_1+r_2+r_3}$.

И т. д.

Префиксный код с заданными длинами кодовых слов

Повторив эти рассуждения k раз, получим слова:

$$B_1, \dots, B_{r_1}, B_{r_1+1}, \dots, B_{r_1+r_2}, \dots, B_{r_1+\ldots+r_{k-1}+1}, \dots, B_{r_1+\ldots+r_{k-1}+r_k}.$$

По построению ни одно из этих слов не является префиксом никакого другого из этих слов.

Поэтому эти слова образуют искомый префиксный (а значит, и разделимый) алфавитный код.

$\square$

Префиксный код с заданными длинами кодовых слов

Пример. Существует ли разделимый алфавитный код в кодирующем алфавите из $q = 3$ букв с длинами кодовых слов:

$$1,\ 2,\ 2,\ 3,\ 3,\ 3?$$

Получаем:

$$\frac{1}{3^1} + \frac{2}{3^2} + \frac{3}{3^3} = \frac{1}{3} + \frac{2}{9} + \frac{3}{27} = \frac{2}{3} \leqslant 1.$$

По доказательству теоремы 13.2 построим префиксный код с такими длинами кодовых слов в кодирующем алфавите $B = \{0, 1, 2\}$:

$$B_1 = 0,\ B_2 = 10,\ B_3 = 11,$$
$$B_4 = 120,\ B_5 = 121, B_6 = 122.$$

Префиксные коды

Теорема 13.3 (о существовании префиксного кода с теми же длинами кодовых слов). *Если* $C = \{B_1, \dots, B_r\}$ *— разделимый алфавитный код в кодирующем алфавите* B*, то найдется такой* **префиксный** *код* $C' = \{B'_1, \dots, B'_r\}$ *в том же алфавите* B*, что* $|B'_i| = |B_i|$ *для всех* $i = 1, \dots, r$.

Доказательство. Пусть $|B| = q$ и $|B_i| = l_i$, $i = 1, \dots, r$.

Код C — разделимый, поэтому по теореме 13.1 верно:

$$\sum_{i=1}^{r} \frac{1}{q^{l_i}} \leqslant 1.$$

Значит, по теореме 13.2 найдется такой префиксный код $C' = \{B'_1, \dots, B'_r\}$ в кодирующем алфавите B, что $|B'_i| = l_i$, $i = 1, \dots, r$. Он и есть искомый.

$\square$

Дерево префиксного кода

Префиксный код C_φ можно задавать в виде корневого дерева D_φ.

Пусть задан префиксный код $C_\varphi = \{B_1, \ldots, B_r\}$ в кодирующем алфавите B. Пусть $b_1, \ldots, b_t \in B$ — все буквы, являющиеся префиксами хотя бы одного кодового слова из C_φ, и

$$C_{\varphi_i} = \{\beta \in B^* \mid \exists B_j \in C_\varphi : B_j = b_i\beta\},$$

где $i = 1, \ldots, t$. Отметим, что C_{φ_i} также является префиксным кодом для всех таких i, $1 \leqslant i \leqslant t$, что $C_{\varphi_i} \neq \{\Lambda\}$.

Тогда корневое дерево D_φ кода C_φ содержит корень v_0, ребра (v_0, v_i), помеченные буквой b_i, $i = 1, \ldots, t$, и поддеревья D_{φ_i} с корнем v_i для всех таких i, $1 \leqslant i \leqslant t$, что $C_{\varphi_i} \neq \{\Lambda\}$.

Дерево префиксного кода

Если D_φ — дерево префиксного кода $C_\varphi = \{B_1, \dots, B_r\}$ с корнем v_0, то у дерева D_φ ровно r листьев.

Более того, все листья дерева D_φ можно так занумеровать $u_1, \dots, u_r$, что если записать последовательно пометки ребер вдоль единственной простой (v_0, u_i)-цепи P_i в этом дереве, то получим кодовое слово B_i, $i = 1, \dots, r$.

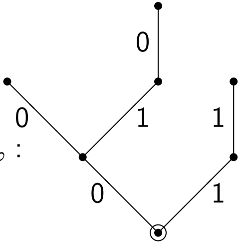
Дерево префиксного кода

Пример. Построим дерево D_φ префиксного кода $C_\varphi = \{00, 11, 010\}$ в кодирующем алфавите $B = \{0, 1\}$:

Дерево префиксного кода

Пример. Построим дерево D_φ префиксного кода $C_\varphi = \{00, 11, 010\}$ в кодирующем алфавите $B = \{0, 1\}$:

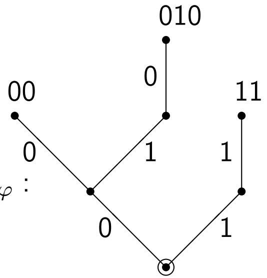
D_φ :



Дерево префиксного кода

Пример. Построим дерево D_φ префиксного кода $C_\varphi = \{00, 11, 010\}$ в кодирующем алфавите $B = \{0, 1\}$:

D_φ :



Дерево префиксного кода

Обратно, пусть D — корневое дерево с корнем v_0 с ребрами, помеченными буквами из некоторого алфавита B.

Кроме того, пусть для любых двух ребер e_1 и e_2, исходящих из одной и той же вершины v и помеченных одной и той же буквой $b \in B$, верно, что одно из этих ребер принадлежит единственной простой (v_0, v)-цепи.

Тогда D можно рассматривать как дерево префиксного кода C, построенного следующим образом.

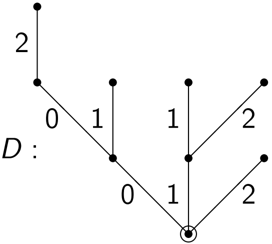
Если $u_1, \ldots, u_r$ — все листья в дереве D, то кодовое слово B_i получаем, записывая последовательно пометки ребер вдоль единственной простой (v_0, u_i)-цепи P_i в дереве D, $i = 1, \ldots, r$.

Далее: $C = \{B_1, \ldots, B_r\}$.

Дерево префиксного кода

Пример. Построим префиксный код C по дереву D:

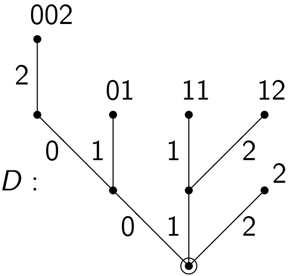
D :



Дерево префиксного кода

Пример. Построим префиксный код C по дереву D:

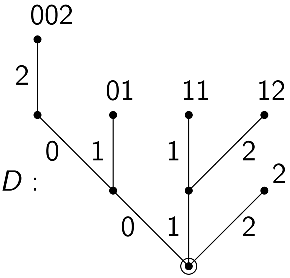
D :



Дерево префиксного кода

Пример. Построим префиксный код C по дереву D:

D :



Получаем: $C = \{002, 01, 11, 12, 2\}$.

Задачи для самостоятельного решения

1. Постройте дерево префиксного кода из примера после теоремы 13.2.

Литература к лекции

1. Алексеев В. Б. Лекции по дискретной математике. М.: Инфра-М, 2012. С. 48–50.

2. Яблонский С. В. Введение в дискретную математику. М.: Высшая школа, 2001. С. 272–276.

3. Гаврилов Г. П., Сапоженко А. А. Задачи и упражнения по дискретной математике. М.: Физматлит, 2004. Гл. VIII 1.6, 1.7, 1.8.