Теорія кодів

Модуль 4 Лекція 4

План

- Вступ до теорії кодів
- Породжуючі матриці
- Код Хемінга

Умовні позначення

- визначення

- приклад

- примітка

- важливо!

- теорема

Вступ до теорії кодів

Визначемо ***код*** як представлення множини символів рядків, що складається з нулів та одиниць. Ця множина символів зазвичай включає букви алфавіту, цифри і, як правило, певні контрольні символи. Коди представляються бінарними рядками з метою використання їх комп’ ютерами для зберігання та передачі даних. Кодування всіх символів двійковими рядками однієї довжини називається ***блокуванням***.

Для кодування кожного символа використовується 8 біт, то відомо, що кожні 8 біт представляють один символ передаваємого повідомлення.

Блоковий код є особливо корисним для обмеження довжини кода для кожного відправленого символа або букви.

При використанні ***кома-коду*** кожний символ кодується рядком з одиниць, в кінці якого стоїть нуль. Множина рядків коду має вигляд {0, 10, 110, 11110, 11111110, ...}. Цей код має явний недолік: елементи коду можуть бути дуже довгими і займати великий об'єм пам'яті.

Для мінімізації об'єма пам'яті найбільш ефективним є ***код Хафмана.*** Найбільша перевага цього виду кодування є в тому, що це миттєвий код.

Прикладом коду, що мінімізує час передачі, є ***код Морзе***. Букви і символи, які зустрічаються набільш часто, мають більш короткий код. В коді Морзе букви розділені «пробілами», а слова – трьома «пробілами». В даному випадку пробіл – це одиниця часу.

Недоліки: в процесі передачі можуть виникати помилки. Причиною помилок – називається невизначеним терміном «шум».

Код Морзе

A	· –	J	· – – –	S	· · ·
B	– · · ·	K	– · –	T	–
C	– · – ·	L	· – · ·	U	· · – –
D	– · ·	M	– –	V	· · · –
E	·	N	– ·	W	· – –
F	· · – ·	O	– – –	X	– · · – –
G	– – ·	P	· – – ·	Y	– · – –
H	· · · ·	Q	– – · –	Z	– – · ·
I	· ·	R	· – ·		

План

Коди, що мають властивість визначення наявності помилок, називаються ***кодами, виявляючими помилки***.

Коди, що дозволяють виправляти помилки, називаються кодами, ***виправляючими помилки***.

Проблема використання кодів з виправленням помилок і кодів з виявленням помилок полягає в тому, що вони повинні включати в себе додаткову інформацію, тому вони являються менш ефективними у відношенні мінімізації об'єма пам'яті.

Десяткова позиційна система числення – це спосіб кодування натуральних чисел. Римські цифри – інший спосіб кодування натуральних чисел, при чому є більш наглядним: палець – I, п'ятерня – V, дві п'ятерні – X. Проте при цьому способі кодування складніше виконувати арифметичні дії над великими числами, тому він був витіснутий позиційною десятковою системою.

Декартові координати – спосіб кодування геометричних об’єктів числами.

Кодування є центральним питанням у розвитку різних (практично всіх) задач програмування:

- Представлення даних довільної природи (чисел, текста, графіки) в пам’яті комп’ютера
- Захист інформації від несанкціонованого доступу
- Забезпечення перешкодостійкості під час передачі даних по каналам зв’язку.
- Стискання інформації в базах даних.

Породжуючі матриці

Будемо вважати, що всі рядки мають фіксовану довжину n, і будемо розглядати рядки як вектори або (1 × n)-матриці. Визначимо додавання по модулю 2, так що 1 + 1 = 0. Таким чином, 11110001 + 10100111 = 01010110

Скалярний добуток векторів $u = (u_1, u_2, \ldots, u_n)$ та $v = (v_1, v_2, \ldots, v_n)$ позначається $u \bullet v$ і дорівнює $u_1v_1 + u_2v_2 + \ldots + u_n v_n$.

Вагою рядку кода c, що позначається wt(c), називається кількість одиниць в рядку.

Якщо c = 1011010, то wt(c) = 4

Припустимо, що є така матриця (k x n)-матриця G, що її перші k стовпців і рядків утворюють одиничну матрицю I_k розміру (k x k), всі стовпці якої різні. Таким чином, матриця G має вигляд $[I_k \mid A_{n-k}]$. Наприклад, наступна матриця

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

є породжуючею матрицею.

S – множина рядків. S ={100101, 010110, 001011}.

Нехай C – код, утворений всіма векторами, які являються кінцевими сумами рядків з S (група C є породженою групою S).

ТЕОРЕМА 19.1. [n, k] –код C містить 2^k рядків.

Для передачі рядків повідомлення довжини k необхідно закодувати їх, помноживши справа на матрицю G. Закодуємо (1, 1, 0)

$$(1,1,0)\begin{bmatrix}1&0&0&1&0&1\\0&1&0&1&1&0\\0&0&1&0&1&1\end{bmatrix}=(1,1,0,0,1,1)$$

Вектор називається ***ортогональним*** іншому вектору, якщо їх скалярний добуток дорівнює нулю.

Двоїстим кодом до коду C, що позначається $C^{\perp}$, називається множина всіх рядків з B_n, ортогональних кожному рядку з коду C.

ТЕОРЕМА 19.2. Нехай C – груповий код, $C^{\perp}$- двоїстий йому код. Рядок t належить коду $C^{\perp}$ тоді і тільки тоді, коли він ортогональний кожному рядку з S, множини, що породжує елементи коду C.

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} = [I_3 | A_3]$$

$$G^{\perp} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} = [A_3^t | I_3]$$

A_3^t – транспонована матриця матриці A_3. Матриця $G^{\perp}$ - матриця контролю парності.

Код Хемінга

- Нехай матриця $G_H^{\perp}$ - матриця, в якій r таких рядків, що стовпці складаються з різних рядків довжини r, за виключенням рядку, що складається з нулів.

Матриця Хемінга – це $((n - r) \times n)$ матриця вигляду $[I_{n-r}| A]$, де A - $((n - r) \times n)$. Код, утворений рядками матриці Хемінга, називається ***кодом Хемінга***.

Нехай $r = 3$ і $G_H^{\perp}$ - матриця $\begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$.

Тоді G_H – матриця $\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$

ТЕОРЕМА 19.3. Для рядків c і c' вага $\mathrm{wt}(c + c') \leq \mathrm{wt}(c) + \mathrm{wt}(c')$.

Відстань Хемінга, або просто відстань між двома рядками кода c і c', що мають однакову довжину, - це число відповідних блоків в рядку, де один рядок має цифру 1, а інший 0. Позначимо функцію відстані $\delta(c, c')$.

Якщо $c = 101011$ і $c' = 110010$, то $\delta(c, c') = 3$, оскільки два рядки відрізняються у другій, третій та шостій позиціях.

ТЕОРЕМА 19.4. Функція відстані Хемінга має наступні властивості:

а) для рядків c і c' відстань $\delta(c, c') = 0$ тоді і тільки тоді, коли $c = c'$

б) для рядків c і c' відстань $\delta(c, c') = \delta(c', c)$

в) для рядків c, c' і c'' виконується співвідношення $\delta(c, c'') \leq \delta(c, c') + \delta(c', c'')$

Якщо C – код, то мінімальна відстань кода C, що позначається D(C), дорівнює найменшій відстані між двома рядками з C.

ТЕОРЕМА 19.5. Для кода C:

а) якщо $D(C) = k + 1$, то використання кода дозволяє виявляти до k помилок.

б) якщо $D(C) = 2k + 1$, то використання кода дозволяє виправляти до k помилок.

ТЕОРЕМА 19.6. Мінімальна відстань D(C) кода C дорівнює $W(C) = \min\{wt(c) : c \in C \text{ і } c \neq 0\}$.

План

Література

- Андерсон Д.А. Дискретная математика и комбинаторика: Пер. с англ.. – М.: Изд. дом «Вильямс», 2003. – 960 с.
- Новиков Ф.А. Дискретная математика для программистов: Учебник для вузов. 3-е изд. – Спб.: Питер, 2008. – 384 с.

Дякую за увагу