Школа лингвистики, 2023-24 уч. год
Дискретная математика для лингвистов
Кодирование - 2 (6, 13 и 16 февраля 2024 года)

В. В. Кочергин, Ю. Г. Кудряшов, А. В. Михайлович, И. В. Щуров, И. А. Хованская

Задача 1. (a) Построить оптимальный алфавитный код (код с минимальной избыточностью), буквы кодируемого алфавита имеют следующие частоты выпадения: 0.24, 0.23, 0.15, 0.09, 0.08, 0.07, 0.05, 0.03, 0.03, 0.02, 0.01, кодирующий алфавит состоит из двух букв.

(b) Построить оптимальный алфавитный код (код с минимальной избыточностью), буквы кодируемого алфавита имеют следующие частоты выпадения: 0.17, 0.16, 0.15, 0.12, 0.1, 0.1, 0.06, 0.05, 0.03, 0.02, 0.02, 0.01, 0.006, 0.004, кодирующий алфавит состоит из пяти букв.

(c) Построить оптимальный алфавитный код (код с минимальной избыточностью), буквы кодируемого алфавита имеют следующие частоты выпадения: $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{16}$, $\frac{1}{16}$, $\frac{1}{16}$, $\frac{1}{16}$, $\frac{1}{20}$, $\frac{1}{20}$, $\frac{1}{20}$, $\frac{1}{20}$, $\frac{1}{20}$ кодирующий алфавит состоит из четырёх букв.

Задача 2. Выяснить, существует ли q-ичный код с минимальной избыточностью (оптимальный код), обладающий заданной последовательностью L длин кодовых слов:

(a) $q = 2$, $L = (2, 3, 3, 3)$;
(b) $q = 2$, $L = (3, 3, 3, 3)$;
(c) $q = 2$, $L = (1, 3, 3, 3, 3)$;
(d) $q = 2$, $L = (1, 2, 3, 4)$;
(e) $q = 2$, $L = (1, 2, 3, 4, 4)$;
(f) $q = 2$, $L = (1, 2, 3, 4, 4, 4)$;
(g) $q = 2$, $L = (1, 2, 3, 4, 4, 4, 4)$;
(h) $q = 2$, $L = (1, 3, 3, 3, 3, 3, 3)$;
(i) $q = 2$, $L = (1, 4, 4, 4, 4, 4, 4, 4, 4)$;
(j) $q = 3$, $L = (1, 1, 2)$;
(k) $q = 3$, $L = (1, 1, 2, 2)$;
(l) $q = 3$, $L = (1, 1, 2, 2, 2, 2)$;
(m) $q = 3$, $L = (2, 2, 2, 2, 2, 2, 2)$;
(n) $q = 3$, L — 26 троек;
(o) $q = 3$, L — 25 троек;
(p) $q = 3$, $L = (1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 5)$;
(q) $q = 4$, L — 13 двоек, 4 тройки;
(r) $q = 4$, L — 11 двоек, 3 тройки.

Задача 3. Для префиксного кода C с заданным набором вероятностей P построить дерево, соответствующее коду. Выяснить, является ли код C оптимальным.

(a) $C = \{1, 00, 01, 02, 20, 21\}$, $P = (0.5; 0.1; 0.1; 0.1; 0.1; 0.1)$;
(b) $C = \{0, 1, 20, 21, 220, 221\}$, $P = \left(\frac{1}{3}; \frac{1}{3}; \frac{1}{9}; \frac{1}{9}; \frac{1}{18}; \frac{1}{18}\right)$;
(c) $C = \{0, 10, 11, 12, 20, 22\}$, $P = \left(\frac{1}{6}; \frac{1}{6}; \frac{1}{6}; \frac{1}{6}; \frac{1}{6}; \frac{1}{6}\right)$;
(d) $C = \{0, 10, 11, 120, 121, 122\}$, $P = \left(\frac{1}{6}; \frac{1}{6}; \frac{1}{6}; \frac{1}{6}; \frac{1}{6}; \frac{1}{6}\right)$;
(e) $C = \{0, 1, 20, 21, 220, 221\}$, $P = \left(\frac{1}{6}; \frac{1}{6}; \frac{1}{6}; \frac{1}{6}; \frac{1}{6}; \frac{1}{6}\right)$.

Задача 4. При оптимальном кодировании алфавита $A = \{a_1, a_2, a_3, a_4\}$ в алфавите $B = \{0, 1\}$ с набором вероятностей $p_1 \geq p_2 \geq p_3 \geq p_4 > 0$, $p_1 + p_2 + p_3 + p_4 = 1$, существует оптимальный код, в котором присутствует кодовое слово длины 3, и существует оптимальный код, в котором кодовые слова длины 3 отсутствуют. Найти все возможные значения, которые при этом может принимать величина

(a) p_1.
(b) p_2.
(c) p_3.
(d) p_4.

Для каждого такого значения привести пример набора вероятностей, на котором это значение достигается.

Задача 5. При оптимальном кодировании алфавита $A = \{a_1, a_2, a_3, a_4, a_5\}$ в алфавите $B = \{0, 1\}$ с набором вероятностей $p_1 \geq p_2 \geq p_3 \geq p_4 \geq p_5 > 0$, $p_1 + p_2 + p_3 + p_4 + p_5 = 1$, существует оптимальный код, в котором присутствует кодовое слово длины 1, и существует оптимальный код, в котором кодовые слова длины 1 отсутствуют. Найти все возможные значения, которые при этом может принимать величина

(a) p_2.
(b) p_3.
(c) p_4.

Для каждого такого значения привести пример набора вероятностей, на котором это значение достигается.

Задача 6. Построить q-ичный оптимальный код для распределения вероятностей $P = (0.4; 0.1; 0.1; 0.1; 0.1; 0.08; 0.06; 0.06)$ при $q = 3$ и $q = 4$.

Задача 7. Рассмотрим следующий алгоритм построения двоичного кода для заданного распределения вероятностей. Сначала разобьём все вероятности на две группы так, чтобы разница между суммами вероятносетй из этих групп была минимальной. Вероятностям из первой группы сопоставим коды, начинающиеся с 0, а вероятностям из второй группы — коды, начинающиеся с 1. Каждую группу аналогичным образом разобьём на две подгруппы, и в соответствии с этим разбиением для каждой вероятности определим второй символ сопоставленного этой вероятности кода. Данную процедуру разбиения вероятностей на подгруппы будем продолжать до тех пор, пока в каждой подгруппе не останется ровно одна вероятнотсть. Привести пример распределения вероятностей, для которого полученный в результате код не является оптимальным.

Задача 8. Построить по методу Хэмминга кодовое слово для сообщения

(a) 10101011;
(b) 111001111;
(c) 100010011;
(d) 01110111011.

Задача 9. По каналу передавалось кодовое слово, построенное по методу Хэмминга. Ошибок не более одной. Восстановить сообщение.

(a) 0101101;
(b) 11011100110;
(c) 1010101010100;
(d) 001011110111111.