ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ И НАУКИ КЕМЕРОВСКОЙ ОБЛАСТИ

**ГПОУ «ЮРГИНСКИЙ ТЕХНОЛОГИЧЕСКИЙ КОЛЛЕДЖ»**

Цикловая методическая комиссия

отделения Автоматизации и информационных технологий

УТВЕРЖДАЮ

Заместитель директора по НМР

\_\_\_\_\_\_\_\_\_\_\_ И. Н. Тащиян

\_\_\_\_\_\_\_\_\_\_\_\_201\_ г.

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ СТУДЕНТОВ**

**по выполнению контрольной работы**

|  |  |
| --- | --- |
| Дисциплина | основы теории информации |
| Специальность | 09.02.02 Компьютерные сети |

201\_

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ СТУДЕНТОВ**

Задачи изучения дисциплины «Теория информации и кодирования» состоят

* в получении количественных оценок информации;
* в расчете информационных характеристик основных элементов систем передачи информации;
* в построении кодов.

В следующих методических указаниях приводятся основные определения и формулы теории информации, а также алгоритмы кодирования их оценки. Кроме того, рассмотрены примеры решения типовых задач, соответствующих задачам, предлагаемым студентам для самостоятельного решения в качестве контрольной работы.

1. Одно из направлений в измерении информации дает структурная теория, в которой количество информации определяется подсчетом информационных элементов или комбинаций из них.

Рассмотрим *аддитивную меру* (*меру Хартли*). Из комбинаторики известно, что число сочетаний с повторениями из *h* элементов по *l* равно

 .

Таким образом, число всех двоичных кодовых комбинаций длины *l* равно 2*l* (*h =*2). В качестве меры информации Хартли предложил взять

 (бит). (1)

Тогда 1 *бит* – это количество информации, содержащееся в двоичной кодовой комбинации единичной длины. Количество информации по Хартли эквивалентно количеству двоичных знаков «0» и «1» при кодировании сообщений по двоичной системе счисления.

 Пример 1. Рассмотрим систему, информационная емкость которой определяется десятичным числом *Q*= 121. Определим количество информации, содержащееся в системе, используя меру Хартли (1),

 ** (бит).

Заметим, что округление результата до целого необходимо проводить в сторону увеличения. Полученный результат означает, что при кодировании числа достаточно использовать 7 двоичных знаков (число возможных двоичных кодовых комбинаций равно 27 = 128)

 *.*

Замечание. Разложение по двоичной системе производится для числа на 1 меньше в силу того, что отсчет ведется от нуля, а число комбинаций равно 121.

 Для получения двоичного числа можно использовать метод последовательного деления числа на 2. При каждом делении определяется один двоичный знак кодовой комбинации: если деление без остатка – «0», в противном случае – «1». Ниже приведена таблица, иллюстрирующая метод, столбец справа показывает двоичные знаки кодовой комбинации.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 120 | 2 |  |  |  |  |  |  |  | «0» |
|  | 60 | 2 |  |  |  |  |  |  | «0» |
|  |  | 30 | 2 |  |  |  |  |  | «0 |
|  |  |  | 15 | 2 |  |  |  |  | «1» |
|  |  |  |  | 7 | 2 |  |  |  | «1» |
|  |  |  |  |  | 3 | 2 |  |  | «1» |
|  |  |  |  |  |  | 1 | 2 |  | «1» |
|  |  |  |  |  |  |  | 0 |  |  |

Двоичное число выписывается в обратном порядке, двигаясь снизу вверх.

2. В статистической теории вводится *энтропия* как мера неопределенности случайного события. Если случайное событие имеет *n* элементарныхисходов и известны их вероятности , то энтропия рассчитывается по формуле:

  (2)

Размерность энтропии определяется основанием логарифма: при основании 2

энтропия измеряется в битах.

Определим количество информации, соответствующее *i* исходу, как

 ,

тогда энтропию можно определить как среднее количество информации, приходящееся на один исход,

 *.*

Энтропия принимает максимальное значение в случае равновероятных исходов

 , (3)

где *n –* число исходов.

 Рассмотрим дискретный источник сообщений *X*. Пусть *xi* – независимые элементы алфавита сообщений, а *p*(*xi*) – их вероятности (*i =* 1, 2, …, *n*), тогда среднее количество информации на один элемент алфавита сообщений определяется через энтропию источника по формуле, аналогичной (2)

  (4)

Максимальное количество информации соответствует случаю равных вероятностей элементов алфавита и определяется по формуле (3).

Определим *абсолютную* и *относительную избыточность* передаваемого сообщенияпо следующим формулам

 **, (5)

 .(6)

Пример 2. Пусть в сообщении используются два независимых символа *x*1 и *x*2. Заданы вероятности появления символов *р*(*x*1)=0,3 и *р*(*x*2)=0,7.

Максимальное среднее количество информации на символ сообщения имеет место при равновероятном распределении и равно, согласно формуле (3),

 ** (бит).

 Рассчитаем среднее количество информации на символ сообщения при заданных вероятностях по формуле (4)

**,

(бит)

Оценим избыточность сообщения по формулам (5) и (6)

 ** (бит),

 .

3. Рассмотрим сложный опыт (*X*, *Y*), исходы которого обозначим через (*xi*, *j*). Энтропия вводится аналогичным образом

 . (7)

 Если составляющие *X*, *Y* сложного опыта независимы, то энтропия сложного опыта равна сумме энтропий составляющих

 . (8)

 В общем случае вводится понятие – условная энтропия *H*(*Y*|*X*)

 , (9)

и формула (8) приобретает следующий вид

 .

 Для дискретного источника сообщений с алфавитом, состоящим из зависимых попарно символов, энтропия источника определяется в соответствии с формулой (9)

  (10)

и характеризует среднее количество передаваемой информации на символ сообщения.

 Пример 3. Учтем зависимость между символами в примере 2, заданную матрицей условных вероятностей: .

 Рассчитаем энтропию источника по формуле (10)

 

Подставим числовые данные, используя пример 2,

 

Таким образом, среднее количество информации на символ сообщения равно 0,64 бит, что меньше 0,88 бит, полученных в примере 2. Это вызвано учетом известной зависимости между символами, что естественно уменьшает неопределенность опыта.

 4. Рассмотрим некоторые коды, позволяющие сделать передачу информации более эффективной, что достигается путем сжатия сообщения. Начнем с *кода Шеннона–Фано*. Алгоритм состоит в следующем: буквы исходного алфавита сообщения выписываются в столбец в порядке убывания их вероятностей; производится разбиение на две группы с равной по возможности суммарной вероятностью, всем буквам верхней группы в качестве первого символа кодовой комбинации приписывается «1», а нижней – «0»; затем производятся следующие разбиения до тех пор, пока в каждой подгруппе не останется одна буква (при каждом разбиении появляется новый символ кодовой комбинации по правилам, изложенным выше).

 *Эффективность кода* рассчитывается следующим образом

 ,

причем в случае двоичного алфавита формула имеет вид

 , (11)

где средняя длина кодовой комбинации равна

 , (12)

*n*(*zi*) – число символов в кодовой комбинации. Эффективность является безразмерной величиной и всегда меньше либо равна 1, т.е. æ ≤ 1. Чем ближе этот показатель к единице, тем эффективнее код.

 Пример 4. Проведем кодирование методом Шеннона–Фэно и рассчитаем характеристики кода. Пусть исходный алфавит состоит из восьми букв и заданы их вероятности. Проведем разбиения по алгоритму Шеннона–Фэно и составим кодовые комбинации.

|  |  |  |
| --- | --- | --- |
| *zi* | *p*(*zi*) | кодовые комбинации  |
| *z*1 | 0,25 |  | 1 |  | 1 |  |  |  |  |
| *z*2 | 0,20 |  | 1  |  | 0 |  |  |  |  |
| *z*3 | 0,15 |  | 0 |  | 1 |  | 1 |  |  |
| *z*4 | 0,10 |  | 0 |  | 1 |  | 0 |  |  |
| *z*5 | 0,10 |  | 0 |  | 0 |  | 1 |  | 1 |
| *z*6 | 0,10 |  | 0 |  | 0 |  | 1 |  | 0 |
| *z*7 | 0,06 |  | 0 |  | 0 |  | 0 |  | 1 |
| *z*8 | 0,04 |  | 0 |  | 0 |  | 0 |  | 0 |
|  |  |  |  |  |  |  |  |  |  |

Рассчитаем энтропию по формуле (4)

 

 Рассчитаем среднюю длину кодовой комбинации по формуле (12)

 

 Эффективность кода, согласно (11), равна

 

 Эффективность кодирования можно увеличить, если проводить кодирование блоков, состоящих из нескольких букв алфавита.

 Пример 5. Проведем блоковое кодирование по методу Шеннона–Фано. Пусть алфавит состоит из двух независимых букв с заданными вероятностями *p*(*z*1)=0,8 и *p*(*z*2)=0,2.

 Рассчитаем энтропию

 (бит).

Очевидно, что при кодировании по одной букве *lср=*1 и æ1=0,72.

 Проведем кодирование блоков, состоящих из двух букв. Ниже приведена таблица с разбиениями и соответствующими кодовыми комбинациями.

|  |  |  |
| --- | --- | --- |
| *zi zj* | *p*(*zi zj*) | код. комб. |
| *z*1 *z*1 | 0,64 | 1 |  |  |
| *z*1 *z*2 | 0,16 | 0 | 1 |  |
| *z*2 *z*1 | 0,16 | 0 | 0 | 1 |
| *z*2 *z*2 | 0,04 | 0 | 0 | 0 |

При расчете вероятностей блоков использовали теорему умножения вероятностей для независимых событий

 *p*(*zi zj*) = *p*(*zi*)∙*p(zj*).

Рассчитаем основные характеристики

 *lср=* 0,64∙1+0,16∙2+0,16∙3+0,04∙3 ≈ 1,56;

 .

В последней формуле учли, что энтропия увеличится в 2 раза, согласно (8),

 .

 При кодировании блоков из трех букв эффективность возрастает еще больше.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *z*1 *z*1 *z*1 | 0,512 | 1 |  |  |  |  |
| *z*1 *z*1 *z*2 | 0,128 | 0 | 1 | 1 |  |  |
| *z*2 *z*1 *z*1 | 0,128 | 0 | 1 | 0 |  |  |
| *z*1 *z*2 *z*1 | 0,128 | 0 | 0 | 1 |  |  |
| *z*1 *z*2 *z*2 | 0,032 | 0 | 0 | 0 | 1 | 1 |
| *z*2 *z*2 *z*1 | 0,032 | 0 | 0 | 0 | 1 | 0 |
| *z*2 *z*1 *z*2 | 0,032 | 0 | 0 | 0 | 0 | 1 |
| *z*2 *z*2 *z*2 | 0,008 | 0 | 0 | 0 | 0 | 0 |

 *lср=* 0,512∙1+0,128∙9+0,032∙15+0,008∙5 ≈ 2,184;

 .

Таким образом, имеем æ1 ‹ æ2 ‹ æ3 .

 Алгоритм *кода Хаффмана* состоит в следующем. Буквы исходного алфавита выписываются в столбец в порядке убывания их вероятностей. Последние две буквы столбца объединяются в одну – вспомогательную букву, которой приписывается суммарная вероятность. Затем формируется следующий столбец с учетом новой буквы по принципу убывания вероятностей. Процесс повторяется и продолжается до тех пор, пока останется одна буква с вероятностью, равной 1. Кодовые комбинации легко получить, построив кодовое дерево. Вершиной дерева является последняя буква, процесс ветвления проводится с учетом полученной таблицы, двигаясь в обратном направлении. Каждому из двух ребер, участвующих в объединении, приписывается кодовый символ: ребру с большей вероятностью – «1», с меньшей – «0». Двигаясь от вершины дерева до одной из букв алфавита по соответствующим ребрам, получаем ее кодовую комбинацию.

 Пример 6. Проведем кодирование по методу Хаффмена. Исходный алфавит состоит из шести букв с заданными вероятностями. Составим таблицу.

|  |  |  |
| --- | --- | --- |
| *zi* | *p*(*zi*) | Вспомогательные столбцы |
| *z*1 | 0,40 | 0,40 | 0,40 | 0,40 | 0,60 0,40 | 1,00 |
| *z*2 | 0,25 | 0,25 | 0,25 | 0,35 0,25 |  |
| *z*3 | 0,15 | 0,15 | 0,20 0,15 |  |  |
| *z*4 | 0,10 | 0,100,10 |  |  |  |
| *z*5 | 0,06 0,04 |  |  |  |  |
| *z*6 |  |  |  |  |  |

 Строим кодовое дерево и выписываем кодовые комбинации букв.

0,40 0,60

0

1

*z*1

0,35 0,25

*z*2

1

0

0,15 0,20

*z*3

0

1

0,10 0,10

0

1

*z*54

0,10 0,10

0,10 0,10

0,06 0,04

*z*44

1

0

*z*64

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *z*1 | *z*2 | *z*3 | *z*4 | *z*5 | *z*6 |
| 0 | 10 | 110 | 1111 | 11101 | 11100 |

Характеристики кода рассчитываются по тем же формулам, что и для кода Шеннона–Фано

 

 

 

 Код Хаффмана можно использовать и для кодирования блоков из букв так, как это было рассмотрено выше для кода Шеннона–Фано, что увеличит эффективность передачи информации.

 5. Важнейшей задачей кодирования является обеспечение достоверной передачи информации. Поэтому обнаружение и исправление ошибок приобретает актуальное значение. Рассмотрим некоторые коды, позволяющие решать эти задачи.

 *Код с проверкой четности* позволяет обнаружить однократную ошибку, т. е. ошибку в одном разряде двоичного слова. Алгоритм кодирования состоит в следующем: к передаваемому двоичному слову добавляем в конце один символ («0» или «1») с тем, чтобы сумма всех символов была равна 0. Суммирование проводится по модулю 2 (mod 2)

 0 ⊕ 0 = 0,

 0 ⊕ 1 = 1,

 1 ⊕ 0 = 1,

 1 ⊕ 1 = 0.

 При декодировании проверяем сумму символов принятого слова. Если она не равна 0, значит, имеем однократную ошибку.

 Пример 7. Проведем кодирование сообщения *a* = 010110 кодом с проверкой четности.

 Суммируем символы заданного сообщения, в результате получим

 0 ⊕ 1 ⊕ 0 ⊕ 1 ⊕ 1 ⊕ 0 = 1.

Следовательно, добавочный символ равен 1 и закодированное сообщение имеет вид

 *b =*0101101.

 Теперь рассмотрим код, позволяющий исправить однократную ошибку. Это – *код Хэмминга*. Разберем алгоритм кода. При кодировании сообщение разбивается на слова длиной *m*, к слову добавляется *r* контрольных символов. Таким образом, закодированное слово имеет длину *n = m + r*, причем выполнено

 2*r* ≥ *n* + 1.

Например, если *m =*4, то легко видеть, что

 23 ≥ (4 + 3) + 1 = 8

и таким образом, число контрольных символов *r =*3 и длина кодового слова *n =* 7. В этом случае имеем (4, 7) код Хэмминга. Контрольные символы размещаются в разрядах с номерами, равными степеням 2, т. е. 20 = 1, 21 = 2, 22 = 4, и т. д. Информационные символы располагаются по порядку в оставшихся разрядах. Если исходное слово *a* = *a*1 *a*2 *a*3 *a*4, а закодированное – *b = b*1*b*2*b*3*b*4*b*5*b*6*b*7, то для определения значений контрольных символов *b*1 , *b*2, *b*4 используем следующую систему

  (13)

 Пример 8. Закодируем двоичное слово *a* = 0101 кодом (4, 7) Хэмминга. Учитывая, что *b*3 = 0, *b*5 = 1, *b*6 = 0, *b*7 = 1. Для определения контрольных символов используем систему (13)

 

Следовательно, = 0, = 1, = 0. Таким образом, закодированное слово имеет вид *b =*0100101.

 Рассмотрим процесс декодирования. Если обозначить искаженное слово через *c = c*1*c*2*c*3*c*4*c*5*c*6*c*7. Тогда, используя следующую систему, определяем разряд, в котором произошла ошибка.

  (14)

Двоичное число (*e*1*e*2*e*3)2 определяет номер разряда, в котором произошла ошибка.

 Пример 9. Декодируем сообщение *c =*0101111, если при кодировании использовался (4, 7) код Хэмминга.

 Используя систему (12), определяем разряд, в котором произошла ошибка

 

Следовательно, ошибка в разряде с номером 2: 0102 = 210. Исправляя ошибку и исключая контрольные символы, получаем информационное сообщение *a* = 0111.

 6. Основной информационной характеристикой канала связи является его пропускная способность, которая определяется как наибольшее возможное количество информации, передаваемое по каналу в единицу времени. Пусть по каналу максимально передается *N* сигналов в единицу времени, тогда пропускная способность канала равна

 , (15)

где  есть среднее количество информации, передаваемое по каналу связи одним сигналом. В формуле (15) максимум берется по всем возможным вероятностным распределениям источника сообщений {*p*(*x*)}, *H*(*Y* | *X*) – условная энтропия, определяемая помехами в данном канале (9).

 Пример 10. Рассчитаем пропускную способность канала связи, потери которого определяются матрицей условных вероятностей

 .

Причем *N =* 10 сигн/сек.

 Рассчитаем условную энтропию по формуле (9)

 

так как *p*(*x*1) + *p*(*x*2) = 1.

 Используя формулу (15), имеем

 бит/сек.

Здесь использовано, что (бит).

**Домашняя контрольная работа**

Студенты должны выполнить контрольную работу, состоящую из 5 задач по разделам: измерение информации, кодирование информации.

**ЗАДАНИЕ НА КОНТРОЛЬНУЮ РАБОТУ**

**1 – 10.** Определить количество информации (по Хартли), содержащееся в системе, информационная емкость которой характеризуется десятичным числом *Q*. Закодировать это число по двоичной системе счисления.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| *Q* | 500 | 1000 | 750 | 1250 | 250 | 1500 | 650 | 900 | 1100 | 1600 |

**11 – 20.** Определить среднее количество информации, содержащееся в сообщении, используемом три независимых символа *S*1, *S*2, *S*3. Известны вероятности появления символов *p*(*S*1)=*p*1, *p*(*S*2)=*p*2, *p*(*S*3)=*p*3. Оценить избыточность сообщения.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| *p*1 | 0,1 | 0,2 | 0,3 | 0,1 | 0,15 | 0,1 | 0,2 | 0,2 | 0,05 | 0,15 |
| *p*2 | 0,15 | 0,1 | 0,15 | 0,3 | 0,2 | 0,4 | 0,25 | 0,3 | 0,15 | 0,25 |
| *p*3 | 0,75 | 0,7 | 0,55 | 0,6 | 0,65 | 0,5 | 0,55 | 0,5 | 0,8 | 0,6 |

**21 – 30.** Вусловии предыдущей задачи учесть зависимость между символами, которая задана матрицей условных вероятностей *P*(*Sj / Si*).

 **21.**  **22.**  **23.** 

 **24.**  **25.**  **26.** 

 **27.**  **28.**  **29.** 

 **30.** 

**31 – 40.** Провести кодирование по одной и блоками по две буквы, используя метод Шеннона–Фано. Сравнить эффективности кодов. Данные взять из задач №11 –20.

**41 – 50.** Алфавит передаваемых сообщений состоит из независимых букв *Si*. Вероятности появления каждой буквы в сообщении заданы. Определить и сравнить эффективность кодирования сообщений методом Хаффмана при побуквенном кодировании и при кодировании блоками по две буквы.

|  |  |  |  |
| --- | --- | --- | --- |
| № | *p*(*Si*) | № | *p*(*Si*) |
| 41 | (0,6;0,2;0,08;0,12) | 46 | (0,7;0,2;0,06;0,04) |
| 42 | (0,7;0,1;0,07;0,13) | 47 | (0,6;0,3;0,08;0,02) |
| 43 | (0,8;0,1;0,07;0,03) | 48 | (0,5;0,2;0,11;0,19) |
| 44 | (0,5;0,3;0,04;0,16) | 49 | (0,5;0,4;0,08;0,02) |
| 45 | (0,6;0,2;0,05;0,15) | 50 | (0,7;0,2;0,06;0,04) |

**51 – 60.** Декодировать полученное сообщение *c*, если известно, что использовался (7, 4) – код Хэмминга. Провести кодирование кодом с проверкой четности.

|  |  |  |  |
| --- | --- | --- | --- |
| № | *c* | № | *c* |
| 51 | 1100011 | 56 | 1011011 |
| 52 | 1010011 | 57 | 1010101 |
| 53 | 1101101 | 58 | 0110111 |
| 54 | 1101001 | 59 | 1110101 |
| 55 | 1100111 | 60 | 1000101 |

**61 – 70.** Определить пропускную способность канала связи, по которому передаются сигналы *Si*. Помехи в канале определяются матрицей условных вероятностей *P*(*Sj / Si*). За секунду может быть передано *N* = 10 сигналов.

**61.**  **62.**  **63.** 

**64.  65.**  **66.** 

 **67.**  **68.  69.** 

**70.** 