|  |  |
| --- | --- |
| Череповецкий государственный университет  Кафедра «Математического и программного обеспечения ЭВМ» | |
| ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ  по дисциплине «Теория информации»  ИССЛЕДОВАНИЕ СВОЙСТВ ЭНТРОПИИ ДИСКРЕТНОГО ИСТОЧНИКА | |
|  | Принял:  преподаватель Е.Н. Руденко    подпись, дата  Выполнил: Антипичев Никита  студент гр. 1Пиб-02-1оп-2022    подпись, дата |

Череповец, 2022

Реферат

Предметом исследования являются формулы К. Шеннона для вычисления количества информации в сообщениях дискретного источника и его энтропии, а также простейшие модели дискретных источников.

Цель работы – исследование свойств энтропии как количественной меры неопределенности дискретного источника.

В ходе работы проводились теоретические исследования энтропии, а также численное моделирование простейших дискретных стационарных и нестационарных источников.

В результате аналитических исследований были найдены условия минимального и максимального значения энтропии. Численно были получены значения энтропии стационарных источников двух и нескольких видов сообщений. Произведено моделирование и исследована динамика изменения энтропии нестационарного источника.

Количественная оценка значения энтропии применяется при моделировании и кодировании источников.

Введение

Понятие информации предполагает наличие двух объектов: *источника информации* и *потребителя* [1, 2]. Информация представляется в виде специальных знаков, символов; характерным носителем информации является *сообщение,* под которым обычно понимают все то, что подлежит передаче. Статистический подход к оценке качества информации представлен в обширном разделе кибернетики – *теории информации*, которая занимается математи­ческим описанием и оценкой методов передачи, хранения, извлече­ния и классификации информации.

Основы теории информации были заложены в 1948 г. американским математиком К. Шенноном, который ввел понятие *энтропии* как меры неопределенности источника и *количества информации* через изменение этой неопределенности.

Пусть имеется дискретный источник, заданный *ансамблем* сообщений *X*={*x*1, *x*2, … *xN*} и вероятностями формирования этих сообщений *P*={*p*1, *p*2, … *pN*}. В силу свойств ансамбля, сообщения множества *X* являются несовместными событиями и

∑*pi* =1. (1)

Количество собственной информации *Ii*, содержащееся в конкретном сообщении *xi*, может быть найдено по следующей формуле:

*Ii*=–log*pi*, (2)

где *pi* – вероятность появления этого сообщения. Единицы измерения количества информации определяет основание логарифма. Использование логарифма по основанию два дает результат в битах.

Среднее значение (математическое ожидание) собственной ин­формации назы­вается *энтропией*. Для дискретного источника сообщений случайная величина собственной ин­формации принимает значения *I*1, *I*2, … *IN* c вероятностями *p*1, *p*2, … *pN* соответственно, и ее мат. ожидание может быть найдено следующим образом:

*H*=M[*Ii*]=∑*pj**Ij*=–∑*pj*log*pj*, (3)

где *j*=1..*N*. В случае *pj*=0 слагаемое *pj*log*pj*принимается равным нулю. Единицы измерения определяются основанием логарифма.

Малые значения энтропии источника говорят о его малой информативности; большие – о неопределенности того, какое именно сообщение будет сформировано источником в определенный момент. Значение энтропии в битах определяет минимальный размер двоичного кода (на одно сообщение в среднем), необходимого для взаимнооднозначного кодирования сообщений источника.

Целью данной лабораторной работы является исследование свойств энтропии, предложенной Шенноном, как количественной меры неопределенности дискретного источника.

Оформление отчета по лабораторной работе было выполнено согласно требованиям ГОСТ 7.32–2001 «Отчет о научно-исследовательской работе. Структура и правила оформления».

1. Основные свойства энтропии дискретного источника

Минимальное значение энтропии

Формула энтропии (3) зависит только от вероятностей *P*. Рассмотрим функцию (–*pi*)⋅log*pi* в случаях *pi* =0, *pi* ∈(0; 1) и *pi* =1.

В случае *pi* =0 произведение (–*pi*)⋅log*pi* (неопределенность вида 0⋅∞) полагается равным своему предельному значению – нулю. В случае *pi* =1 произведение (–*pi*)⋅log*pi* обращается в ноль в силу свойств логарифма. При *pi* ∈(0; 1) величина log*pi* всегда отрицательна, а произведение (‑*pi*)⋅log*pi* больше нуля.

Следовательно, энтропия как сумма слагаемых вида (–*pi*)⋅log*pi* является неотрицательной функцией. Нулевое значение энтропии возможно только при обращении в ноль всех ее слагаемых, что возможно в случае, когда вероятность одного из сообщений равна единице, а другие сообщения невозможны. Таким образом, минимальным значением энтропии является ноль.

Максимальное значение энтропии

Энтропия достигает максимального значения, когда вероятности появления возможных сообщений одинаковы (*p*1 =*p*2 =…= *pN* =1/*N*), что может быть доказано методом неопределенных множителей Лагранжа [4]. Следовательно, максимальное значение энтропии (3) составит

*H*max =–∑(1/*N*)log(1/*N*)=log*N*. (4)

График энтропии источника с двумя состояниями

Пусть источник формирует всего два вида сообщений с вероятностью *P*={*p*1, *p*2}. Из (1) следует, что *p*2 = 1 – *p*1, следовательно, энтропия (3) является функцией одной переменной. График энтропии такого источника представлен на рисунке 1.

Гладя на рисунок, можно понять что при p=0,5 определённость максимальна.

2 Вычисление энтропии простейших систем

2.1 Идеальная монета

Результат броска идеальной монеты может быть представлен в виде стационарного источника с двумя равновероятными состояниями. Имея *P*={0,5; 0,5}, вычислим значение энтропии в битах по формуле (3):

*H*=–∑*pj*log2 *pj* = – 0,5 log2 0,5 – 0,5 log2 0,5 = 0,5 бит + 0,5 бит = 1 бит. (5)

Согласно (4) полученное значение является максимально возможным значением энтропии системы с двумя состояниями.

2.2 Фальшивая монета

Фальшивая монета имеет смещенный центр тяжести, из-за чего вероятности выпадения «орла» или «решки» различны. Представим результат броска такой монеты в виде источника с двумя неравновероятными состояниями, положим *P*={0,57; 0,43} и вычислим значение энтропии в битах по формуле (3):

*H*=–∑*pj*log2 *pj* =

 (6)

Полученное в (6) значение энтропии (больше или меньше?) значения энтропии (5), что говорит о (большей или меньшей?) неопределенности результата броска фальшивой монеты. Действительно, результат броска фальшивой монеты более предсказуем: сторона, ближе к которому смещен центр тяжести, чаще будет оказываться снизу.

2.3 Игральная кость

Результатом броска игральной кости является случайная дискретная величина, принимающая шесть равновероятных значений.

P {1/6; 6}

H= –∑*pj*log2 *pj =* (1/6)\*LOG2(1/6)=2,584963 бит.



2.4 Фальшивая игральная кость

Предположим, что в игральной кости центр тяжести смещен так, чтобы шестерка выпадала более часто (единица на противоположной грани кости будет выпадать реже; значения 2÷5 выпадают по-прежнему с вероятностью 1/6). Пусть вероятности составят, например, *P*={0,12; 0,17; 0,17; 0,17; 0,17; 0,20}.

H = -Log2(0,12)\* 0,12 + 4\*(-log2(0,17)\*0,17)+ (-log2(0,20)\*0,20)=2,5698 бит.



**3. Вычисление энтропии нестационарного дискретного источника**

Для стационарных источников вероятности сообщений не изменяются во времени, следовательно, значение энтропии также остается постоянным. Рассмотрим нестационарный источник, о котором априорно известно:

- мощность ансамбля сообщений (число возможных сообщений) N = 12;

- множество сообщений (в данном случае, букв) X = {"а", "н", "т", "и", "п", "ч", "е", "в", "\_", "к", "о", "л"};

- множество абсолютных частот этих сообщений F0 = {3; 3; 2; 6; 1; 2; 2; 2; 2; 2; 1; 1;}, которые следует уменьшать на единицу для каждого принятого сообщения.

Пусть таким источником была сформирована строка "антипичев\_никита\_николаевич". Перед принятием первой буквы частоты составляли F0 . После принятия буквы "ю" частота, соответствующая этой букве, уменьшается на единицу; F1 ={2; 3; 2; 6; 1; 2; 2; 2; 2; 2; 1; 1;} Таблица А.1 приложения А содержит последовательные вычисления Fk, где k – количество принятых сообщений (букв), fk∑ – сумма всех частот Fk (кол-во сообщений, которые следует ожидать), sk – (k + 1)-е принятое сообщение.

Перед принятием (k + 1)-го сообщения на основе имеющихся частот Fk и свойства полноты вероятностей (1) может быть найдено распределение вероятностей Pk следующим образом:

pki = fki / fk∑ , (7)

где i = 1 .. N ; fk∑ – сумма всех fki в Fk . Имея вероятности Pk перед принятием (k + 1)-го сообщения по формуле (3) может быть найдено значение энтропии источника Hk. Построим таблицу А.2 значений Pk и Hk, подобную таблице А.1, где pk∑ – сумма вероятностей для проверки условия (1).

Полученные значения Hk, представленные в таблице А.1, отображены графически на рисунке 2.



