Федеральное государственное автономное

образовательное учреждение

высшего профессионального образования

«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий

Кафедра вычислительной техники

**ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №5**

**«Подсчёт энтропии и избыточности алфавита»**

Преподаватель Пушкарёв К.В.

подпись, дата инициалы, фамилия

Студент КИ15-08Б Войченко В.В.

подпись, дата инициалы, фамилия

Красноярск 2016

**Цели работы:**

1.Изучение энтропии и избыточности алфавита.

**Порядок выполнения работы:**

1.Выполнить все задания.

2.Продемонстрировать выполнение заданий преподавателю.

3.Подготовить отчёт.

4.Защитить лабораторную работу перед преподавателем.

**Указания:** Работу выполнять индивидуально.

**Задания**

1. Написать функцию, подсчитывающую энтропию заданного алфавита: h = alph\_entropy(P), где P -- вектор вероятностей символов алфавита.

2. Написать функцию, подсчитывающую избыточность заданного алфавита:

r = alph\_redundancy(P), где P -- вектор вероятностей символов алфавита.

Указание: функция alph\_redundancy() должна использовать функцию alph\_entropy().

3. Подсчитать энтропию и избыточность алфавитов, приложенных к заданию.

4. Сгенерировать случайный вектор вероятностей для алфавита из трёх символов (сумма вероятностей должна быть равна единице!). Подсчитать энтропию и избыточность этого алфавита.

**Указание:** в качестве генератора случайных чисел использовать функцию rand() в MATLAB. Программный код, использованный для генерации случайного вектора вероятностей, привести в отчёте. Сгенерированный вектор вероятностей привести в отчёте не менее чем с тремя знаками после запятой.

Результаты работы:

Отчёт, включающий программный код, входные данные и результаты проверки.

1. Функция для вычисления энтропии алфавита (входной параметр: вектор вероятностей алфавита):

function [H] = alph\_entropy (P)

H=0;

for i=1:length (P)

if (P(i) ~= 0)

H=H-P(i)\*(log2(P(i)));

end

end

Используемая формула:

1. Функция для вычисления избыточности алфавита входные параметры: (вектор вероятностей алфавита)

function [x] = alph\_redundancy (P)

x=1-(alph\_entropy (P)/log2(length (P)));

end

Используемая формула:

, где , где М- число букв в алфавите

1. Инициализация векторов, взятых из файлов; инициализация случайного вектора, состоящего из 3- х элементов и сумма координат которого равна 1;

COIN = [0.5 0.5];

UNFAIR = [0.999 0.001];

CRIME = [0.166 0.066 0.014 … ];

VENTSEL = [0.145 0.064 0.015 … ];

Результат работы кода,

генерирующего вектор:



THEvector = rand (1, 3);

THEvector = THEvector/sum(THEvector);

1. Вызов функции для вычисления энтропии разных алфавитов (COIN, CRIME, UNFAIR, VENTSEL- векторы вероятности алфавита, данных в файлах \*.csv; THEvector - сгенерированный в шаге III вектор вероятности алфавита):

COIN\_entropy = alph\_entropy (COIN)

CRIME\_entropy = alph\_entropy (CRIME)

UNFAIR\_entropy = alph\_entropy (UNFAIR)

VENTSEL\_entropy = alph\_entropy (VENTSEL)

THEvector\_entropy = alph\_entropy (THEvector)

1. Вызов функции для вычисления избыточности разных алфавитов:

COIN\_redundancy = alph\_redundancy (COIN)

CRIME\_redundancy = alph\_redundancy (CRIME)

UNFAIR\_redundancy = alph\_redundancy (UNFAIR)

VENTSEL\_redundancy = alph\_redundancy (VENTSEL)

THEvector\_redundancy = alph\_redundancy (THEvector)

**Энтропия и избыточность алфавитов:**

**Алфавит "coin.csv"**

|  |  |
| --- | --- |
| Символ | Вероятность появления |
| h | 0,5 |
| t | 0,5 |

Энтропия данного алфавита : 1 бит

Избыточность данного алфавита : 0 бит/символ

**Алфавит "crime.csv"**

|  |  |
| --- | --- |
| Символ | Вероятность появления |
| 0 | 0,166 |
| а | 0,066 |
| б | 0,014 |
| в | 0,039 |
| г | 0,014 |
| д | 0,027 |
| е | 0,072 |
| ё | 0,001 |
| ж | 0,009 |
| з | 0,013 |
| и | 0,055 |
| й | 0,009 |
| к | 0,028 |
| л | 0,038 |
| м | 0,026 |
| н | 0,055 |
| о | 0,096 |
| п | 0,023 |
| р | 0,035 |
| с | 0,045 |
| т | 0,054 |
| у | 0,024 |
| ф | 0,001 |
| х | 0,007 |
| ц | 0,002 |
| ч | 0,015 |
| ш | 0,007 |
| щ | 0,003 |
| ъ | 0 |
| ы | 0,014 |
| ь | 0,018 |
| э | 0,003 |
| ю | 0,005 |
| я | 0,018 |

Энтропия данного алфавита : 4.3717 бит

Избыточность данного алфавита : 0.1407 бит/символ

**Алфавит "unfair.csv"**

|  |  |
| --- | --- |
| Символ | Вероятность появления |
| X | 0,999 |
| Y | 0,001 |

Энтропия данного алфавита : 0.0114 бит

Избыточность данного алфавита : 0.9886 бит/символ

**Алфавит "ventsel.csv"**

|  |  |
| --- | --- |
| Символ | Вероятность появления |
| 0 | 0,145 |
| а | 0,064 |
| б | 0,015 |
| в | 0,039 |
| г | 0,014 |
| д | 0,026 |
| е | 0,074 |
| ж | 0,008 |
| з | 0,015 |
| и | 0,064 |
| й | 0,01 |
| к | 0,029 |
| л | 0,036 |
| м | 0,026 |
| н | 0,056 |
| о | 0,095 |
| п | 0,024 |
| р | 0,041 |
| с | 0,047 |
| т | 0,056 |
| у | 0,021 |
| ф, | 0,002 |
| х | 0,009 |
| ц | 0,004 |
| ч | 0,013 |
| ш | 0,006 |
| щ | 0,003 |
| ъ, ь"," | 0,015 |
| ы | 0,016 |
| э | 0,003 |
| ю | 0,007 |
| я | 0,019 |

Энтропия данного алфавита : 4.4197 бит

Избыточность данного алфавита : 0.1161 бит/символ

**Алфавит "THEvector" (сгенерированный)**

|  |
| --- |
| Вероятность появления |
| 0,440986528005884 |
| 0,490279156108804 |
| 0,0687343158853116 |

Энтропия данного алфавита : 1,2906 бит

Избыточность данного алфавита : 0,1857 бит/символ