



universidade de aveiro
theoria poiesis praxis

Universidade de Aveiro

Departamento de Electrónica, Telecomunicações e Informática

Algorithmic Information Theory (2021/22)

Lab work nº1

Grupo nº 10

Fernando Lopes nº 106358

Maria João Sousa n.º 109488

João Carvalho nº 89059

Índice

1.	Capa	Página 1
2.	Índice.....	Página 2
3.	Finite-context model	Páginas 3-7
4.	Gerador.....	Página 7
5.	Links utilizados.....	Página 7

Finite-context model

O processo de desenvolvimento do programa que usa o finite-context model para calcular a entropia de um texto foi dividido em quatro etapas:

1. Leitura do texto e extração do seu alfabeto;

- Abertura do ficheiro e iteração sobre todos os símbolos do texto. Todos os símbolos diferentes são guardados numa lista.

2. Criação da tabela contendo o número de vezes que cada símbolo aparece em determinado contexto;

- Iterar o conteúdo do ficheiro obtendo todos os contextos e a quantidade dos seus símbolos e armazenar essa informação num dicionário (dictionary = {context: {symbol: num}}).

3. Cálculo da entropia de cada contexto;

- Aplicação da fórmula matemática para o cálculo da entropia de cada contexto.

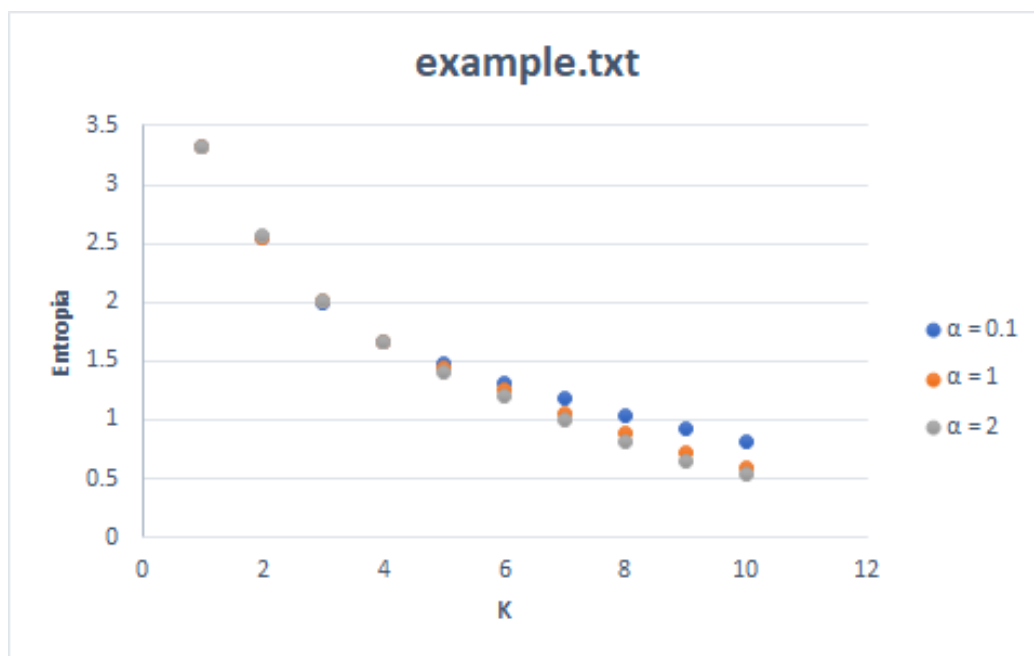
4. Cálculo da entropia geral;

- Aplicação da fórmula matemática para o cálculo da entropia geral.

Utilizando o Finite Context Model desenvolvido, foram calculadas as entropias dos ficheiros: example.txt, dorian.txt e gatsby.txt. Para cada ficheiro fez-se variar os valores de K , entre 1 e 10, para os valores de $\alpha=0.1$, $\alpha=1$ e $\alpha=2$. Foi feita uma representação gráfica com os resultados obtidos para cada ficheiro.

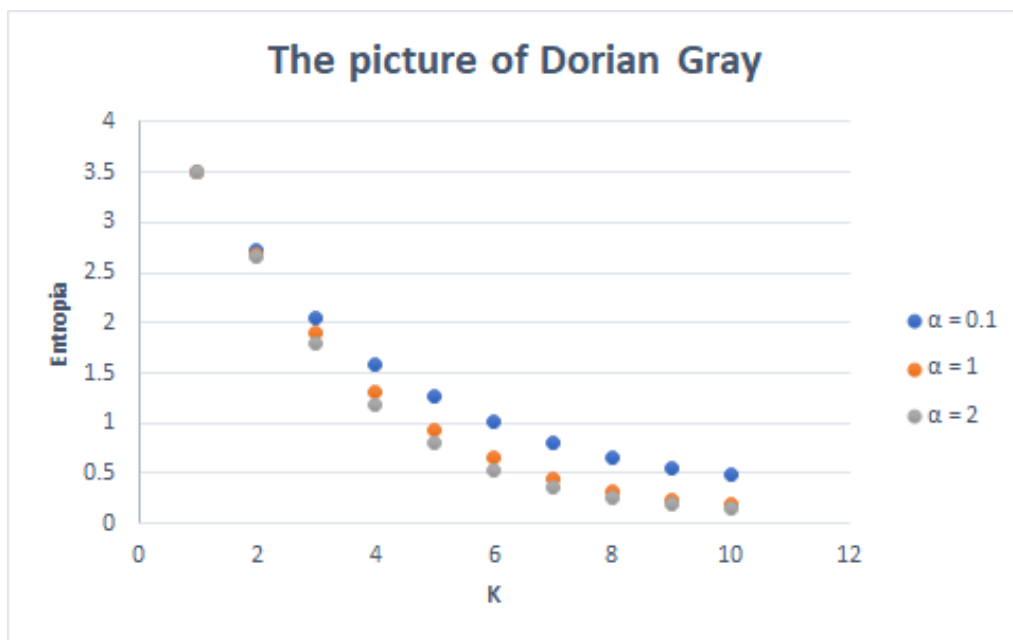
Example.txt

K	Entropy		
	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 2$
1	3.32	3.32	3.32
2	2.55	2.55	2.56
3	2	2.01	2.01
4	1.67	1.67	1.66
5	1.47	1.44	1.41
6	1.32	1.25	1.2
7	1.18	1.06	1
8	1.04	0.88	0.81
9	0.92	0.72	0.65
10	0.82	0.6	0.53



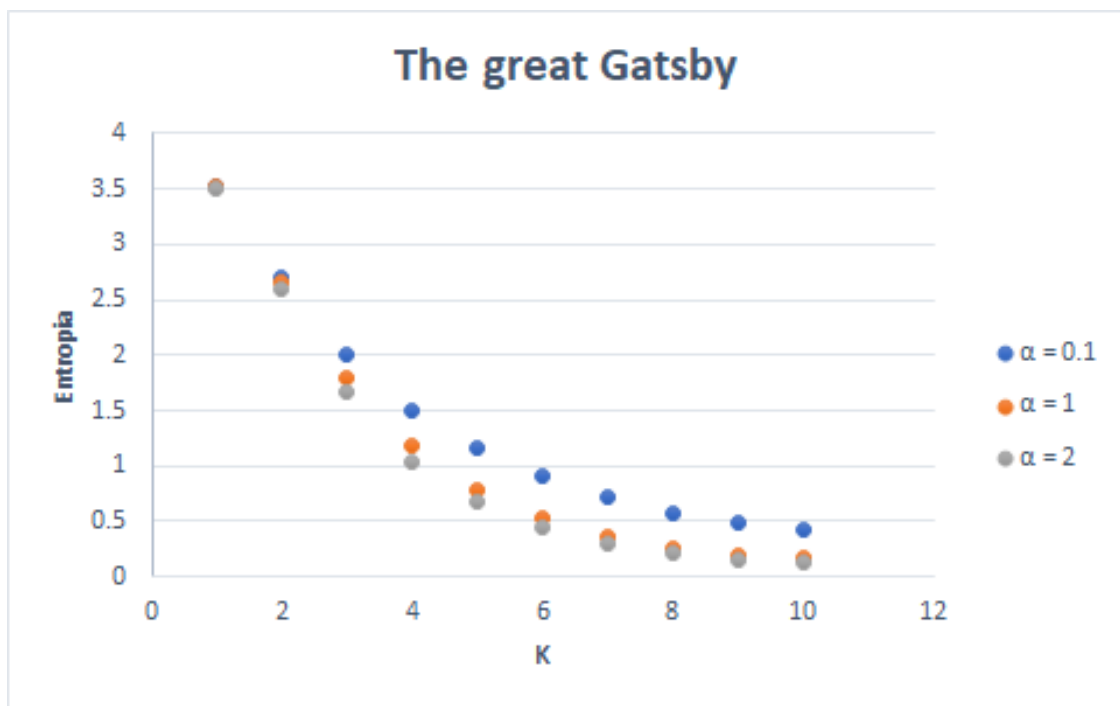
Dorian.txt

K	Entropy		
	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 2$
1	3.51	3.51	3.51
2	2.73	2.69	2.66
3	2.05	1.9	1.8
4	1.59	1.31	1.18
5	1.26	0.92	0.8
6	1.01	0.65	0.54
7	0.8	0.45	0.36
8	0.65	0.32	0.25
9	0.55	0.24	0.19
10	0.48	0.19	0.15



Gatsby.txt

K	Entropy		
	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 2$
1	3.52	3.52	3.51
2	2.71	2.65	2.6
3	2.01	1.8	1.67
4	1.51	1.18	1.04
5	1.16	0.79	0.67
6	0.91	0.54	0.44
7	0.71	0.36	0.29
8	0.58	0.26	0.21
9	0.49	0.2	0.16
10	0.43	0.17	0.13



Observando os gráficos pode-se concluir que à medida que o K aumenta a Entropia diminui. Isso deve-se ao facto de o aumento do K fazer aumentar a profundidade da tabela do modelo, sendo assim possível estabelecer mais relações entre os valores precedentes. Devido a isso, a probabilidade de prever os valores futuros aumenta.

Analisando o gráfico conclui-se que o aumento do α faz diminuir a entropia. O α resolve o problema de haver casos em que apenas existe uma letra que sucede uma determinada sequência. Nesses casos a probabilidade é 1 mas com um α superior a 0 reduz essas probabilidades para valores inferiores a 1.

Gerador

O gerador recebe 3 parâmetros, o dicionário de probabilidades, dictionary, o prior que são as primeiras letras do texto a gerar e o comprimento do texto a gerar lenText. O gerador verifica se o prior é válido, ou seja, se têm o mesmo comprimento que K e está contido no dicionário e caso esteja, cria uma lista. Após este processo, cria o novo elemento baseado na probabilidade que ele tem de acontecer após os últimos K caracteres e adiciona à lista. A geração do elemento seguinte utiliza um número aleatório entre 0 e 1 que escolhe o primeiro elemento da lista cuja soma de probabilidades dos elementos anteriores a si, mais a sua própria probabilidade é maior do que o número aleatório.

Percentagens de participação

Fernando Lopes nº 106358 - 33%

Maria João Sousa n.º 109488 - 33%

João Carvalho nº 89059 - 33%

Links utilizados

The Picture of Dorian Gray(dorian.txt) - <https://www.gutenberg.org/ebooks/174>

The Great Gatsby(gatsby.txt) - <https://www.gutenberg.org/ebooks/64317>