Arithmetic Coding

Arithmetic coding is a variable-length source encoding technique. In traditional entropy encoding techniques such as Huffman coding, each input symbol in a message is substituted by a specific code specified by an integer number of bits.

Arithmetic coding deviates from this paradigm. In arithmetic coding, a sequence of input symbols is represented by an interval of real numbers between *0.0* and *1.0* (i.e a sequence of source symbols (or message) is assigned a single arithmetic code word. The code word itself defines an interval of real numbers between *0* and *1*).

For example we consider a four symbol alphabet *A={a,b,c,d}* with the fixed symbol probabilities *p(a)=0.3 p(b)=0.2 p(c)=0.4* and *p(d)=0.1* respectively. The symbol probabilities can be expressed in terms of partition of the half-open range *[0.0 , 1.0]* as shown in the following table.

Source symbol	*Probability*	*Range (initial subinterval)*
a	*0.3*	*0.0 – 0.3*
b	*0.2*	*0.3 – 0.5*
c	*0.4*	*0.5 – 0.9*
d	*0.1*	*0.9 – 1.0*

Probability model

Example: *Use arithmetic coding algorithm to encode CLASS*

Symbol	*probability*		*range*
A	1/5	0.2	0.0 – 0.2
C	1/5	0.2	0.2 – 0.4
L	1/5	0.2	0.4 – 0.6
S	2/5	0.4	0.6 – 1.0

Coding

	New char	***arithmetic code***
0.2	C	0.2
0.2	L	0.4*0.2+0.2=0.28
0.2	A	0.0*0.2*0.2+0.28=0.28
0.4	S	0.6*0.2*0.2*0.2+0.28=0.2848
0.4	S	0.6*0.4*0.2*0.2*0.2+0.2848= **0.28672**

Decoding

C	0.28672	- 0.2	0.08672	÷ 0.2
L	0.4336	- 0.4	0.0336	÷ 0.2
A	0.168	- 0.0	0.168	÷ 0.2
S	0.84	- 0.6	0.24	÷ 0.4
S	0.6	- 0.6	0	

Example: ***Use arithmetic encoding to encode "KING HENRY"***

Symbol	***Probability***		***Range***
Space	1/10	0.1	0.0-0.1
E	1/10	0.1	0.1-0.2
G	1/10	0.1	0.2-0.3
H	1/10	0.1	0.3-0.4
I	1/10	0.1	0.4-0.5
K	1/10	0.1	0.5-0.6
N	2/10	0.2	0.6-0.8
R	1/10	0.1	0.8-0.9
Y	1/10	0.1	0.9-1.0

Coding

	Newchar	***Arithmetic code***
0.1	K	0.5
0.1	I	0.4*0.1+0.5=0.54
0.2	N	0.6*0.1*0.1+0.54=0.546
0.1	G	0.2*0.2*0.1*0.1+0.546=0.5464
0.1	Space	0.0*0.1*0.2*0.1*0.1+0.5464=0.5464
0.1	H	0.3*0.1*0.1*0.2*0.1*0.1+0.5464=0.546406
0.1	E	0.1*0.1*0.1*0.1*0.2*0.1*0.1+0.546406=0.5464062
0.2	N	0.6*0.1*0.1*0.1*0.1*0.2*0.1*0.1+0.5464062=0.54640632
0.1	R	0.8*0.2*0.1*0.1*0.1*0.1*0.2*0.1*0.1+0.54640632=0.546406352
0.1	Y	0.9*0.1*0.2*0.1*0.1*0.1*0.1*0.2*0.1*0.1+0.546406352=***0.5464063556***

Decoding

K	0.5464063556	- 0.5	÷ 0.1
I	0.464063556	- 0.4	÷ 0.1
N	0.64063556	- 0.6	÷ 0.2
G	0.2031778	- 0.2	÷ 0.1
Space	0.031778	- 0.0	÷ 0.1
H	0.31778	- 0.3	÷ 0.1
E	0.1778	- 0.1	÷ 0.1
N	0.778	- 0.6	÷ 0.2
R	0.89	- 0.8	÷ 0.1
Y	0.9	- 0.9	0