Jak zadać dobre pytanie, czyli czym jest informacja i jak ją zmierzyć

Adam Doliwa
doliwa@matman.uwm.edu.pl

WYKŁAD Z CYKLU NIEZWYKŁA MATEMATYKA
WYDZIAŁ MATEMATYKI I INFORMATYKI UWM

Olsztyn, 28 września 2016 r.

BIT – binary digit – cyfra dwójkowa

Problem

Ile bitów potrzeba do zapisania wyniku pięciu rzutów monetą?

ORROR $\leftrightarrow$ 01101

Tyle samo, ile trzeba zadać pytań binarnych, tzn. mających za możliwą odpowiedź

TAK lub NIE

Przed doświadczeniem przyjmujemy każdy z możliwych $2^5 = 32$ wyników za jednakowo prawdopodobny. Liczba 5 jest tu miarą naszej niepewności co do wyniku rzutów

BIT – binary digit – cyfra dwójkowa

Problem

Ile bitów potrzeba do zapisania wyniku pięciu rzutów monetą?

$$\text{ORROR} \leftrightarrow \ 01101$$

Tyle samo, ile trzeba zadać pytań binarnych, tzn. mających za możliwą odpowiedź

TAK lub NIE

Przed doświadczeniem przyjmujemy każdy z możliwych $2^5 = 32$ wyników za jednakowo prawdopodobny. Liczba 5 jest tu miarą naszej niepewności co do wyniku rzutów

Informacja jako redukcja niepewności

Problem

Ile bitów informacji niesie wiadomość, że w poprzednim doświadczeniu za każdym razem uzyskaliśmy ten sam wynik?

ALBO

Teraz do opisania wyniku doświadczenia wystarczy znać odpowiedź na TYLKO JEDNO PYTANIE

Wiadomość ta dostarcza nam cztery bity informacji

Informacja jako redukcja niepewności

Problem

Ile bitów informacji niesie wiadomość, że w poprzednim doświadczeniu za każdym razem uzyskaliśmy ten sam wynik?

ALBO

Teraz do opisania wyniku doświadczenia wystarczy znać odpowiedź na TYLKO JEDNO PYTANIE

Wiadomość ta dostarcza nam cztery bity informacji

Logarytmy informatyczne $\lg = \log_2$

$\ldots$	$\ldots$
$1 = 2^0$	$\lg 1 = 0$
$2 = 2^1$	$\lg 2 = 1$
$4 = 2^2$	$\lg 4 = 2$
$8 = 2^3$	$\lg 8 = 3$
$16 = 2^4$	$\lg 16 = 4$
$32 = 2^5$	$\lg 32 = 5$
$\ldots$	$\ldots$

Uwaga: $\lg 3 = 1,584962501...$, $\lg 5 = 2,321928095...$

Zadanie 1

Ile bitów informacji ma wiadomość, że rzucając osiem razy monetą uzyskaliśmy raz orła i siedem razy reszkę?

Kod Morse'a

A	·−	J	·−−−	S	···	2	··−−−
B	−···	K	−·−	T	−	3	···−−
C	−·−·	L	·−··	U	··−	4	····−
D	−··	M	−−	V	···−	5	·····
E	·	N	−·	W	·−−	6	−····
F	··−·	O	−−−	X	−··−	7	−−···
G	−−·	P	·−−·	Y	−·−−	8	−−−··
H	····	Q	−−·−	Z	−−··	9	−−−−·
I	··	R	·−·	1	·−−−−	0	−−−−−

Stworzony w 1832 przez Samuela Morse'a i Alfreda Vaila sposób reprezentacji alfabetu, cyfr i znaków specjalnych za pomocą dźwięków, błysków światła, impulsów elektrycznych lub znaków popularnie zwanych kreską i kropką. Alfabetem źródłowym jest alfabet łaciński z cyframi i znakami specjalnymi, a alfabetem kodowym $M = \{\cdot, -, \ \}$ (trzecim elementem zbioru M jest spacja). Spacja oddziela słowa kodowe, grupy znaków oddzielamy trzema spacjami.

ASCII – *American Standard Code for Information Interchange*

Dec	Bin	Char	Dec	Bin	Char	Dec	Bin	Char	Dec	Bin	Char
0	0000000	NUL	32	0100000	space	64	1000000	@	96	1100000	‘
1	0000001	SOH	33	0100001	!	65	1000001	A	97	1100001	a
2	0000010	STX	34	0100010	”	66	1000010	B	98	1100010	b
3	0000011	ETX	35	0100011	#	67	1000011	C	99	1100011	c
4	0000100	EOT	36	0100100	$	68	1000100	D	100	1100100	d
5	0000101	ENQ	37	0100101	%	69	1000101	E	101	1100101	e
6	0000110	ACK	38	0100110	&	70	1000110	F	102	1100110	f
7	0000111	BEL	39	0100111	’	71	1000111	G	103	1100111	g
8	0001000	BS	40	0101000	(	72	1001000	H	104	1101000	h
9	0001001	TAB	41	0101001	)	73	1001001	I	105	1101001	i
10	0001010	LF	42	0100010	*	74	1001010	J	106	1101010	j
11	0001011	VT	43	0100011	+	75	1001011	K	107	1101011	k
12	0001100	FF	44	0100100	,	76	1001100	L	108	1101100	l
13	0001101	CR	45	0100101	-	77	1001101	M	109	1101101	m
14	0001110	SO	46	0100110	.	78	1001110	N	110	1101110	n
15	0001111	SI	47	0100111	/	79	1001111	O	111	1101111	o
16	0010000	DLE	48	0101000	0	80	1010000	P	112	1110000	p
17	0010001	DC1	49	0101001	1	81	1010001	Q	113	1110001	q
18	0010010	DC2	50	0101010	2	82	1010010	R	114	1110010	r
19	0010011	DC3	51	0101011	3	83	1010011	S	115	1110011	s
20	0010100	DC4	52	0101100	4	84	1010100	T	116	1110100	t
21	0010101	NAK	53	0101101	5	85	1010101	U	117	1110101	u
22	0010110	SYN	54	0101110	6	86	1010110	V	118	1110110	v
23	0010111	ETB	55	0101111	7	87	1010111	W	119	1110111	w
24	0011000	CAN	56	0110000	8	88	1011000	X	120	1111000	x
25	0011001	EM	57	0110001	9	89	1011001	Y	121	1111001	y
26	0011010	SUB	58	0110010	:	90	1011010	Z	122	1111010	z
27	0011011	ESC	59	0110011	;	91	1011011	[	123	1111011	{
28	0011100	FS	60	0110100	¡	92	1011100	\	124	1111100	—
29	0011101	GS	61	0110101	=	93	1011101	]	125	1111101	}
30	0011110	RS	62	0110110	¿	94	1011110	ˆ	126	1111110	˜
31	0011111	US	63	0110111	?	95	1011111	_	127	1111111	DEL

Opowieść o dziennikarzu (część pierwsza)

Z Centusiowa do Słoneczkowa został wysłany młody dziennikarz, którego zadaniem było przesyłanie do centrali raz na tydzień wiadomości zbiorczej o pogodzie jaka była każdego dnia

słonecznie 	zachmurzenie małe 	zachmurzenie duże 	deszcz 
00	01	10	11

Przykładowa wiadomość: 010011010010 01

Opowieść o dziennikarzu (część druga)

Młody dziennikarz zauważył, że w Słoneczkowie (średnio rzecz biorąc) raz na dwa dni jest zachmurzenie małe, raz na cztery dni słonecznie, raz na osiem dni jest zachmurzenie duże oraz raz na osiem dni pada deszcz. Postanowił to wykorzystać do skonstruowania lepszego kodowania wiadomości

$p_1 = \frac{1}{4}$	$p_2 = \frac{1}{2}$	$p_3 = \frac{1}{8}$	$p_4 = \frac{1}{8}$
			
0	1	10	11

$$p_1 + p_2 + p_3 + p_4 = \frac{2}{8} + \frac{4}{8} + \frac{1}{8} + \frac{1}{8} = 1$$

Odpowiadająca  wiadomość: 101110101

może być odczytana (rozkodowana) jako: 

lub nawet jako: 

W poszukiwaniu optymalnego kodowania

$p_1 = \frac{1}{4}$	$p_2 = \frac{1}{2}$	$p_3 = \frac{1}{8}$	$p_4 = \frac{1}{8}$
			
01	1	001	000

Jest to kod jednoznaczny i natychmiastowy (słowo kodowe nie jest początkiem innego słowa kodowego)

Odpowiadająca pogodzie  wiadomość ma postać: 1010001010011

Czy nowy kod jest lepszy od pierwszego?

Ile zaoszczędzi dziennikarz używając przez 40 tygodni nowy kod zamiast pierwszego, jeśli przesłanie bitu kosztuje 1 PLN ?

Użytkowanie starego kodu kosztuje: $40 \times 7 \times 2 = 560$ [PLN]

Rozwiązanie zadania

Koszty związane z używaniem nowego kodu w przeciągu 40 tygodni:

- średnio $20 \times 7 = 140$ dni z małym zachmurzeniem: $1 \times 140 = 140$
- średnio $10 \times 7 = 70$ dni słonecznych: $2 \times 70 = 140$
- średnio $5 \times 7 = 35$ dni z dużym zachmurzeniem: $3 \times 35 = 105$
- średnio $5 \times 7 = 35$ dni deszczowych: $3 \times 35 = 105$

Razem przesłano (średnio rzecz biorąc) $140 + 140 + 105 + 105 = 490$ bitów za sumę 490 PLN.

Dziennikarz oszczędził 560 [PLN] - 490 [PLN] = 70 [PLN]

Uwaga: Średnia długość słowa starego kodu wynosi $E(K_S) = 2$ bity, a nowego kodu wynosi $E(K_N) = 490/280 = 7/4$ bitu

Pytanie

Czy jest to najlepsze (optymalne) kodowanie?

Entropia źródła

Jeśli źródło informacji wysyła wiadomości z częstościami (prawdopodobieństwem) $p_1, p_2, \ldots, p_n$ to jego entropia jest zdefiniowana wzorem

$$\mathcal{H}(p_1, p_2, \ldots, p_n) = p_1 \lg \frac{1}{p_1} + p_2 \lg \frac{1}{p_2} + \ldots p_n \lg \frac{1}{p_n}$$

Przykład: Entropia pogody w Słoneczkowie

$$\mathcal{H}\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}\right) = \frac{1}{4} \lg 4 + \frac{1}{2} \lg 2 + \frac{1}{8} \lg 8 + \frac{1}{8} \lg 8 = \frac{7}{4}$$

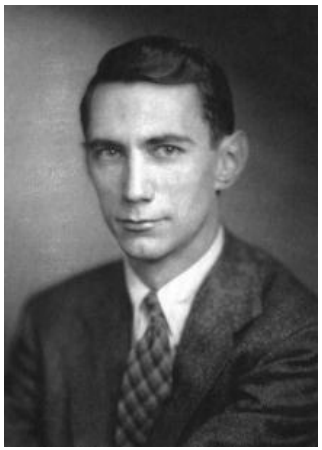
$p_1 = \frac{1}{4}$	$p_2 = \frac{1}{2}$	$p_3 = \frac{1}{8}$	$p_4 = \frac{1}{8}$
			
01	1	001	000

$$E(K) = \frac{1}{4} \cdot 2 + \frac{1}{2} \cdot 1 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = \frac{7}{4}$$

Twierdzenie Shannona

Dla dowolnego źródła i dowolnego odpowiadającego mu binarnego kodu natychmiastowego średnia długość $E(K)$ słowa kodowego jest nie mniejsza niż entropia tego źródła

$$E(K) \geq \mathcal{H}$$



Claude Elwood Shannon (1918-2001) – twórca matematycznej teorii informacji

Wniosek: Ponieważ w kodowaniu wymyślonym przez młodego dziennikarza średnia długość słowa kodowego jest równa entropii źródła informacji, więc jest to kodowanie optymalne

Entropia jako „średnia niepewność”

W przypadku wyboru jednego elementu ze zbioru n-elementowego z równomiernym prawdopodobieństwem $p_i = \frac{1}{n}, \quad i = 1, 2, \ldots, n$ entropia jest równa $\mathcal{H}(\frac{1}{n}, \ldots, \frac{1}{n}) = \frac{1}{n} \lg n + \cdots + \frac{1}{n} \lg n = \lg n$. Jest to maksymalna jej wartość spośród wszystkich możliwych rozkładów prawdopodobieństw $0 \leq p_i \leq 1, \quad p_1 + p_2 + \cdots + p_n = 1$

Na gruncie nauki o cieple entropię jako *tajemniczą wielkość* opisującą stan układu wprowadził w 1854 roku Rudolf Clausius (1822-1888). Jej sens z punktu widzenia atomowej teorii materii wyjaśnił w 1877 roku Ludwig Boltzmann (1844-1906), jako wielkości mierzącej stopień nieuporządkowania układu

Zadanie 2

Średnie dane pogody w Deszczykowie przedstawione są w tabeli poniżej. Wyznacz odpowiedni optymalny binarny i natychmiastowy kod do przesyłania informacji o pogodzie.

$p_1 = \frac{1}{16}$	$p_2 = \frac{1}{16}$	$p_3 = \frac{1}{8}$	$p_4 = \frac{1}{2}$	$p_5 = \frac{1}{4}$
				

Pierwsza osoba, która nadeśle prawidłowe odpowiedzi na oba zadania na adres doliwa@matman.uwm.edu.pl otrzyma w nagrodę monetę

wybitą z okazji Europejskiego Kongresu Matematyki (Kraków 2012) przedstawiającą najwybitniejszego matematyka polskiego XX wieku Stefana Banacha (1892-1945)

Siedziba Agencji Bezpieczeństwa Narodowego (National Security Agency) w Fort Meade w stanie Maryland — największe skupienie matematyków na świecie.

Które pytanie najlepiej zadać?

Alicja wybiera jedną liczbę naturalną od 1 do 12. Odpowiedź na które z pytań zmniejszy (średnio) w największym stopniu naszą niepewność na temat jej wyboru?

1. Czy jest to liczba podzielna przez trzy?
2. Czy jest to liczba parzysta?
3. Czy jest to liczba 10?

Zanim zadamy pytanie, nasza niepewność co do wybranej liczby wynosi $\lg 12 \simeq 3,585$ bitów

Analiza odpowiedzi na pytanie 1

p(TAK) = 4/12 = 1/3, pozostaje niepewność $\lg 4 = 2$ bitów
p(NIE) = 8/12 = 2/3, pozostaje niepewność $\lg 8 = 3$ bitów

Średnia niepewność po otrzymaniu odpowiedzi na pytanie 1 wynosi

$$\frac{1}{3} \cdot \lg 4 + \frac{2}{3} \cdot \lg 8 = \frac{8}{3} \simeq 2,667$$

Informacja = redukcja niepewności $\simeq 0,918$ bitu

Które pytanie najlepiej zadać?

Alicja wybiera jedną liczbę naturalną od 1 do 12. Odpowiedź na które z pytań zmniejszy (średnio) w największym stopniu naszą niepewność na temat jej wyboru?

1. Czy jest to liczba podzielna przez trzy?
2. Czy jest to liczba parzysta?
3. Czy jest to liczba 10?

Zanim zadamy pytanie, nasza niepewność co do wybranej liczby wynosi $\lg 12 \simeq 3,585$ bitów

Analiza odpowiedzi na pytanie 1

p(TAK) = 4/12 = 1/3, pozostaje niepewność $\lg 4 = 2$ bitów
p(NIE) = 8/12 = 2/3, pozostaje niepewność $\lg 8 = 3$ bitów

Średnia niepewność po otrzymaniu odpowiedzi na pytanie 1 wynosi

$$\frac{1}{3} \cdot \lg 4 + \frac{2}{3} \cdot \lg 8 = \frac{8}{3} \simeq 2,667$$

Informacja = redukcja niepewności $\simeq 0,918$ bitu

Analiza odpowiedzi na pytanie 2

p(TAK) = 6/12 = 1/2, pozostaje niepewność $\lg 6 \simeq 2,585$ bitów
p(NIE) = 6/12 = 1/2, pozostaje niepewność $\lg 6 \simeq 2,585$ bitów

Średnia niepewność po otrzymaniu odpowiedzi na pytanie 2 wynosi

$$\frac{1}{2} \cdot \lg 6 + \frac{1}{2} \cdot \lg 6 = \lg 6 \simeq 2,585$$

Informacja = redukcja niepewności = $\lg 12 - \lg 6 = \lg 2 = 1$ bit

Analiza odpowiedzi na pytanie 3

p(TAK) = 1/12, pozostaje niepewność $\lg 1 = 0$ bitów
p(NIE) = 11/12, pozostaje niepewność $\lg 11 \simeq 3,459$ bitów

Średnia niepewność po otrzymaniu odpowiedzi na pytanie 3 wynosi

$$\frac{1}{12} \cdot \lg 1 + \frac{11}{12} \cdot \lg 11 \simeq 3,171$$

Informacja = redukcja niepewności =$\lg 12 - \frac{11}{12} \lg 11 \simeq 0,414$ bitu

Analiza odpowiedzi na pytanie 2

p(TAK) = 6/12 = 1/2, pozostaje niepewność $\lg 6 \simeq 2,585$ bitów
p(NIE) = 6/12 = 1/2, pozostaje niepewność $\lg 6 \simeq 2,585$ bitów

Średnia niepewność po otrzymaniu odpowiedzi na pytanie 2 wynosi

$$\frac{1}{2} \cdot \lg 6 + \frac{1}{2} \cdot \lg 6 = \lg 6 \simeq 2,585$$

Informacja = redukcja niepewności = $\lg 12 - \lg 6 = \lg 2 = 1$ bit

Analiza odpowiedzi na pytanie 3

p(TAK) = 1/12, pozostaje niepewność $\lg 1 = 0$ bitów
p(NIE) = 11/12, pozostaje niepewność $\lg 11 \simeq 3,459$ bitów

Średnia niepewność po otrzymaniu odpowiedzi na pytanie 3 wynosi

$$\frac{1}{12} \cdot \lg 1 + \frac{11}{12} \cdot \lg 11 \simeq 3,171$$

Informacja = redukcja niepewności =$\lg 12 - \frac{11}{12} \lg 11 \simeq 0,414$ bitu

Odpowiedź

Najwięcej informacji uzyskamy jeśli zadamy pytanie numer 2

Podsumowanie

- W najprostszym przypadku logarytm z liczby możliwych odpowiedzi mierzy naszą niepewność co do wyniku obserwacji
- Jeśli prawdopodobieństwa wyników nie są równe to miarą niepewności jest entropia
- Informacja przekazana jest równa redukcji niepewności/entropii podczas transmisji
- Entropia pozwala wyznaczyć granice naszych możliwości w bezstratnej kompresji danych (opowieść o młodym dziennikarzu)
- W termodynamice entropia jest miarą nieuporządkowania układu i rośnie podczas procesów spontanicznych
- Organizmy żywe zmniejszają swoją entropię wykorzystując energię z zewnątrz i zwiększając entropię otoczenia
- Czy jest możliwa zamiana informacji w energię?