wiadomość – komunikat - informacja

Caius Julius Cesar

Człowiek zasztyletowany przez senatorów na forum Romanum w Idy Marcowe roku DCCIX ab urbe condita

Wojna Bambadocji przeciwko Alandii i Cezji

Alandia:
Czuję się dobrze i niczego mi nie brak

Cezja:
Czuję się dobrze i niczego mi nie brak
Jestem zupełnie zdrów
Nie martwcie się o mnie, niczego nie potrzebuje

wiadomość – komunikat - informacja

komunikat
zakodowana **wiadomość** zawierająca pewną ilość **informacji**

Teoria informacji (ilościowa)
Claude Shannon, 1948-49.

Teoria informacji charakteryzuje w sposób matematyczny zapis, przesyłanie i odtwarzanie informacji. Dąży przy tym do pogodzenia dwóch przeciwstawnych celów:

- zapisywania wiadomości jak najzwięźlej,
- chronienia wiadomości przed przekłamaniami podczas transmisji.

Podstawowe założenie ilościowej teorii informacji:

Komunikat zawiera tym więcej informacji, im mniejsze jest prawdopodobieństwo jego wystąpienia.

Przykład: Wojna Bambadocji przeciwko Alandii i Cezji

Alandia:
Czuję się dobrze i niczego mi nie brak

Cezja:
Czuję się dobrze i niczego mi nie brak
Jestem zupełnie zdrów
Nie martwcie się o mnie, niczego nie potrzebuję

1.4. Reprezentacja informacji

Teoria informacji - dział matematyki na pograniczu statystyki i informatyki, mający również olbrzymie znaczenie w współczesnej telekomunikacji, dotyczący przetwarzania informacji oraz jej transmisji, kompresji, kryptografii itd.

Informacją nazywamy wielkość abstrakcyjna, która może być przechowywana w pewnych obiektach, przesyłana miedzy nimi, przetwarzana i stosowana do sterowania nimi (obiektami). Przez obiekty należy rozumieć organizmy żywe, urządzenia techniczne oraz systemy takich obiektów.

1.4. Reprezentacja informacji

Badaniem problemów ilości informacji, sposobów kodowani i przesyłania informacji zajmuje się nauka – teoria informacji.

Pojęcia teorii informacji:

bit: najmniejsza jednostka informacji potrzebna do zakodowania zdarzeń, które z dwóch możliwych zdarzeń zaszło.

1.4. Reprezentacja informacji

•**entropia**: najmniejsza średnia ilość informacji potrzebna do zakodowania faktu zajścia zdarzenia ze zbioru zdarzeń o danych prawdopodobieństwach.

•**Entropia** w ramach teorii informacji jest definiowana jako średnia ilość informacji, przypadająca na znak symbolizujący zajście zdarzenia z pewnego zbioru. Zdarzenia w tym zbiorze mają przypisane prawdopodobieństwa wystąpienia.

Entropia informacyjna źródła informacji

Def.

Jeśli źródło może nadawać n różnych komunikatów z prawdopodobieństwami p_i , (i=1, .., n), to średnia (ważona) ilość informacji w komunikatach z tego źródła wynosi

$$H = \sum_{i=1}^{n} p_i \, log_2\left(\frac{1}{p_i}\right)$$

1.4. Reprezentacja informacji

Def.

Jeśli p= Pr(komunikat *a*), to *a* zawiera (niesie)

$k = \log_2(1/p) = - \log_2 p$ jednostek informacji

Np. Możliwy tylko jeden komunikat *a* ze źródła:

Pr(*a*) = 1 $\rightarrow$ komunikat *a* niesie $\log_2 (1/1) = 0$

jednostek informacji

Np. komunikat *a* ze źródła:

Pr(*a*) = 0,5 $\rightarrow$ komunikat *a* niesie $\log_2 (1/2) = 1$

jednostek informacji

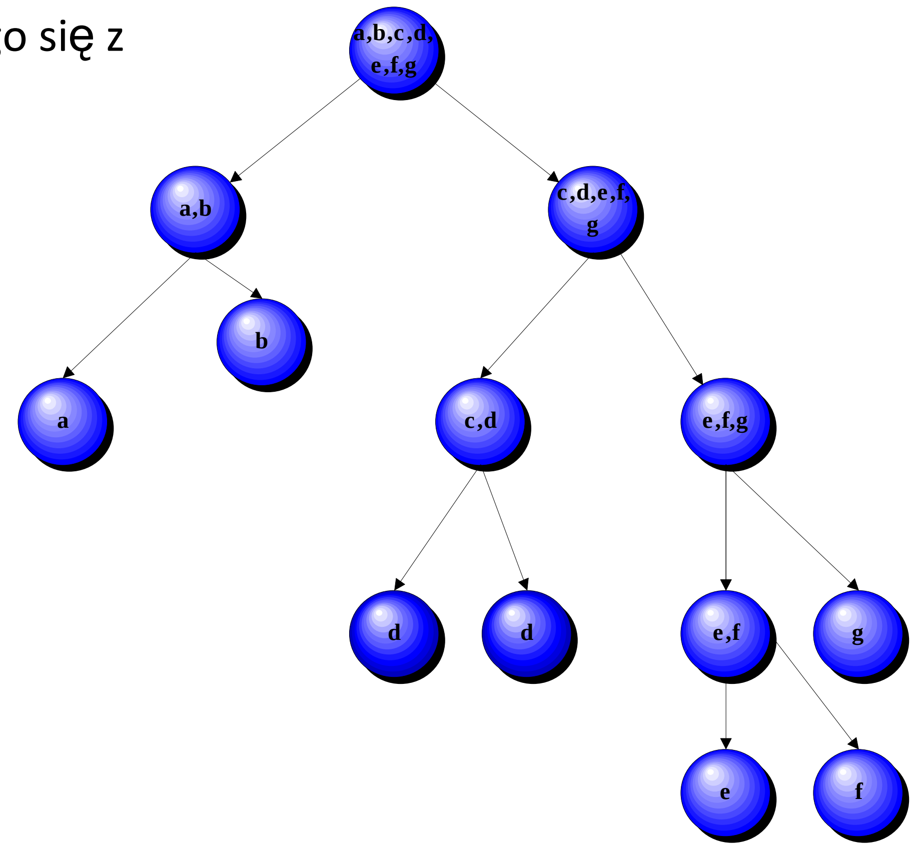
Jednostka informacji:
ilość informacji uzyskana po stwierdzeniu, że zaszło jedno z dwóch jednakowo prawdopodobnych wydarzeń.

1.4. Reprezentacja informacji

Kodowanie komunikatu

Zależnie od przyjętego sposobu kodowania, ciągi kodowe tego samego komunikatu mogą być różnej długości

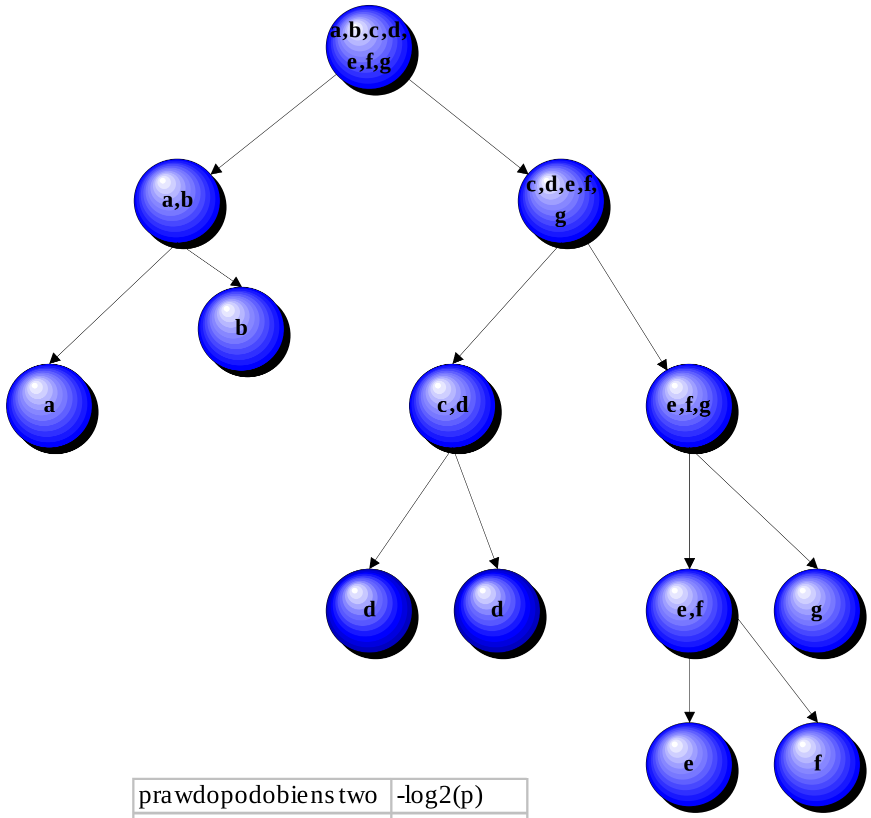
Przyjęto, że elementy alfabetu składającego się z dwóch znaków oznacza się przez 0 i 1.



Kodowanie komunikatu

Dla kodowania z użyciem alfabetu dwuznakowego minimalna długość słowa kodowego dla komunikatu o prawdopodobieństwie wystąpienia p_i wynosi

$$N = -\log_2 p_i$$



Kod danego komunikatu nazywa się ciągiem albo słowem kodowym tego komunikatu.

Liczba elementów (znaków) w nich występujących nazywa się długością słowa kodowego.

prawdopodobienstwo	-log2(p)
0.1	3.32
0.2	2.32
0.5	1.00
1	-

Kodowanie komunikatu

Przykład: komunikaty cyfrowe (cyfry dziesiętne) z jednakowym prawdopodobieństwem wystąpienia p_i =0,1.

liczba	p	-log2(p)
0	0.1	3.32
1	0.1	3.32
2	0.1	3.32
3	0.1	3.32
4	0.1	3.32
5	0.1	3.32
6	0.1	3.32
7	0.1	3.32
8	0.1	3.32
9	0.1	3.32
10	0.1	3.32

$$N^* = -log_2(0{,}1)$$

$$N \approx 3{,}322$$

Uwaga 1: prawdopodobieństwa wystąpienia komunikatów nie zawsze są całkowitymi potęgami liczby ½ (tak jak w analizowanym przykładzie)

Uwaga 2: słowa kodowe muszą się składać z całkowitej liczby znaków

Kodowanie komunikatu

Jeśli komunikatom źródła przypisano słowa kodowe o długościach N_i, $i=1,..,n$,

to wielkość

$$L = \sum_{i=1}^{n} p_i N_i$$

liczba	p	N=-log2(p)	
0	0.1	3.32	0.33
1	0.1	3.32	0.33
2	0.1	3.32	0.33
3	0.1	3.32	0.33
4	0.1	3.32	0.33
5	0.1	3.32	0.33
6	0.1	3.32	0.33
7	0.1	3.32	0.33
8	0.1	3.32	0.33
9	0.1	3.32	0.33
10	0.1	3.32	0.33
		L	3.65
			4

określana jest jako średnia (ważona) długość słowa kodowego.

Kodowanie komunikatu

Dla danego źródła i danego sposobu kodowania wielkość

$$R = L - H$$

określana jest jako redundancja sposobu kodowania

Redundancja (łac. *redundantia – powódź, nadmiar*, *zbytek*), inaczej *nadmiarowość* w stosunku do tego, co konieczne lub zwykłe. Może odnosić się zarówno do nadmiaru zbędnego lub szkodliwego, niecelowo zużywającego zasoby, jak i do pożądanego zabezpieczenia na wypadek uszkodzenia części systemu.

1.4. Reprezentacja informacji

Kodowanie komunikatu

Przykład (cd.): komunikaty cyfrowe (cyfry dziesiętne) z jednakowym prawdopodobieństwem wystąpienia p_i =0,1.

Kod Aikena

(ciągi jednakowej długości)

0: 0000 5: 1011

1: 0001 6: 1100

2: 0010 7: 1101

3: 0011 8: 1110

4: 0100 9: 1111

$$H = \sum_{i=1}^{10} 0{,}1 \log_2\left(\frac{1}{0{,}1}\right) \quad \rightarrow \quad \underline{H \approx 3{,}322}$$

$$\underline{L = 4}$$

$$R = 4 - 3{,}322 \quad \rightarrow \quad \underline{R = 0{,}678}$$

Kodowanie komunikatu

Przykład (cd.): komunikaty cyfrowe (cyfry dziesiętne) z jednakowym prawdopodobieństwem wystąpienia p_i =0,1.

Kod Aikena (ciągi jednakowej długości)

Wada: przypadkowe przekłamanie (zmiana znaku) może przekształcić dane słowo kodowe w inne dopuszczalne słowo kodowe; odbiorca nie ma żadnego sygnału o zaistniałym przekłamaniu

0: 0000	5: 1011
1: 0001	6: 1100
2: 0011	7: 1101
3: 0011	8: 1110
4: 0100	9: 1111

Kodowanie komunikatu

Przykład (cd.): komunikaty cyfrowe (cyfry dziesiętne) z jednakowym prawdopodobieństwem wystąpienia p_i =0,1.

Kod Aikena (ciągi jednakowej długości)

Rozwiązanie: zwiększenie redundancji

Kodowanie komunikatu

Kod Aikena (ciągi jednakowej długości)

Rozwiązanie: zwiększenie redundancji poprzez

zwiększenie długości słów kodowych o 1

0: 0000 1 5: 1011 0

1: 0001 0 6: 1100 1

2: 0010 1 7: 1101 0

3: 0011 0 8: 1110 0

4: 0100 0 9: 1111 1

Kodowanie komunikatu

Kod Aikena (ciągi jednakowej długości)

Rozwiązanie: zwiększenie redundancji poprzez

zwiększenie długości słów kodowych o 1

Uwaga 1: opisany sposób to kontrola parzystości.

0: 0000 1 - OK.

0: 0001 1 - NIE OK

Uwaga 2: przy dalszym zwiększeniu długości słowa kodowego można także rekonstruować z przekłamanego słowa słowo prawidłowe (kody samokorygujace się).

Kodowanie komunikatu

Dla danego źródła i danego sposobu kodowania wielkość

$$R = L - H$$

Redundancja to ilość informacji przekraczająca wymagane do rozwiązania problemu minimum. Bardziej formalnie – ilość bitów w wiadomości minus ilość bitów faktycznej informacji.

Celowa redundancja danych jest stosowana w celu ułatwienia odtworzenia danych po ich częściowej utracie czy uszkodzeniu lub też do wykrycia takiego uszkodzenia.

Redundancja ma zastosowanie głównie w przypadku bardzo ważnych, strategicznych dla danego systemu informacji. Szczególnie często mamy do czynienia z redundancją danych w systemach telekomunikacyjnych, gdzie niezawodność przesyłania odgrywa kluczową rolę podczas transmisji. Obecnie nadmiarowość jest cechą każdego systemu informacyjnego przesyłającego jakieś dane cyfrowe.

1.4. Reprezentacja informacji

Kodowanie komunikatu

Dla danego źródła i danego sposobu kodowania wielkość

$$R = L - H$$

Usuwanie redundancji to **kompresja** danych. Paradoksalnie, wiele programów kompresujących może dodawać niewielkie informacje nadmiarowe, pozwalające wykryć uszkodzenie skompresowanych danych (sumy kontrolne).

1.4. Reprezentacja informacji