Przedmiot:	Matematyczne narzędzia komputerowe w zastosowaniach telekomunikacyjnych
Prowadzący:	Piotr Chołda piotr.cholda@agh.edu.pl
Kierunek:	Elektronika i Telekomunikacja
Specjalność:	Sieci i usługi
Semestr:	I sem. (letni) studiów magisterskich

Konspekt

Piotr Chołda

7 czerwca 2018

1 Zmienne losowe, modelowanie probabilistyczne i kodowanie źródłowe (entropijne)

1.1 Modelowanie źródeł wiadomości z użyciem entropii

1. Miara informacji.
2. Entropia zmiennej losowej.
3. Właściciowości entropii.
4. Użycie entropii w systemach telekomunikacyjnych i teleinformatycznych.
5. Entropia binarnego źródła wiadomości
6. Rozszerzenie źródła wiadomości i entropia źródła rozszerzonego.
7. Entropia dla wielu zmiennych losowych: entropia łączna, entropia warunkowa.

1.2 Kodowanie entropijne

1. Kod źródłowy.
2. Kod jednoznacznie dekodowalny. Nierówność McMillana-Krafta.
3. Średnia długość słowa kodowego $\overline{L}$. Ograniczenie dolne na $\overline{L}$.
4. Kod zwięzły, kod optymalny. Kod Huffmana.
5. Twierdzenie Shannona dla kanałów bezszumowych (twierdzenie Shannona o kodowaniu źródłowym).

1.3 Zadania

1. Która z poniższych sekwencji niesie większą informację:
   (a) 10 liter alfabetu łacińskiego (zakładamy, że wystąpienie każdej z 32 liter jest równie prawdopodobne, nie interesuje nas rozróżnienie między wielkimi a małymi literami),
   (b) 32 cyfry (zakładamy, że wystąpienie każdej z dziesięciu cyfr jest równoprawdopodobne)?

2. Jaką maksymalną nieokreśloność może zawierać fotografia $6 \times 9\,\mathrm{cm}^2$, jeśli rozmiar ziarna jest równy $0,01 \times 0,01\,\mathrm{cm}^2$, a każde ziarno może mieć trzy odcienie: biały, czarny i szary?

3. Źródło generuje wiadomości według następującego schematu:

   (a) losowane są 0 oraz 1 (odpowiednio z prawdopodobieństwem $p_0 = p$, $p_1 = 1 - p$);

   (b) losowanie kończy się w chwili wylosowania pierwszej jedynki;

   (c) źródło wysyła informację, za którym razem to nastąpiło.

   Proszę podać entropię takiego źródła.

4. Ile najwięcej słów kodowych może liczyć kod binarny jednoznacznie dekodowalny, którego najdłuższe słowo ma siedem liter?

5. Czy można w alfabecie ternarnym skonstruować kod jednoznacznie dekodowalny, jeśli długości słów kodowych miałyby wynosić:

$$1, 2, 2, 2, 2, 2, 3, 3, 3, 3?$$

   A jeśli długości słów kodowych miałyby wynosić:

$$1, 2, 2, 2, 2, 2, 3, 3, 3?$$

   Jak wiele takich kodów (różnych) można skonstruować?

6. Źródło generuje równoprawdopodobne 4-symbolowe słowa kodowe kodu zwięzłego trójsymbolowego. Ile maksymalnie informacji o źródle można uzyskać po odebraniu ośmiu słów?

7. Proszę udowodnić, że dla każdego kodu zwięzłego (ewentualnie kodu optymalnego) dla źródła o rozkładzie prawdopodobieństwa (p_i) zachodzi następująca zależność między prawdopodobieństwami wystąpienia wiadomości a długościami reprezentujących je słów kodowych l_i:

$$p_j > p_k \Rightarrow l_j \leq l_k.$$

8. Źródło generuje trzy wiadomości z prawdopodobieństwami: $p_1 \geq p_2 \geq p_3$. Pokazać, że średnia długość binarnego kodu zwięzłego $\mathcal{C}$ dla tego źródła wynosi:

$$L(\mathcal{C}) = 2 - p_1.$$

   Ile będzie wynosiła średnia długość słowa kodu tego typu dla źródła generującego cztery wiadomości z prawdopodobieństwami: $p_1 \geq p_2 \geq p_3 \geq p_4$?

9. Obliczyć entropię źródła X, które generuje nieskończenie wiele wiadomości $m_1, m_2, \ldots, m_i, \ldots$ (zbiór wiadomości jest równoliczny ze zbiorem liczb naturalnych) o prawdopodobieństwie:

$$p_i = 2^{-i} \qquad (i = 1, 2, 3, \ldots).$$

   Następnie podać optymalny binarny kod jednoznacznie dekodowalny i oblicz jego średnią długość.

10. Proszę pokazać, że dla żadnego źródła o siedmiu wiadomościach nie da się z użyciem kodowania Huffmana uzyskać kodu o podanym niżej zbiorze słów kodowych:

$$01 \quad 100 \quad 101 \quad 1110 \quad 1111 \quad 0011 \quad 0001.$$

11. Źródło X generuje wiadomości $x_1, x_2, \dots, x_n$ z prawdopodobieństwami $p_1 \geq p_2 \geq \cdots \geq p_n$ oraz dodatkowo:

$$\forall_{i=1,\dots,n-3}\colon p_i > p_{i+2} + \cdots + p_n.$$

Pokazać, jakie będą długości słów kodowych dla dowolnego binarnego kodu Huffmana tego źródła. Ile istnieje różnych binarnych kodów Huffmana dla tego źródła?

12. Źródło generuje osiem wiadomości elementarnych według następującego rozkładu prawdopodobieństwa:

$$\Pr\{x_1\} = \frac{2}{5}; \Pr\{x_2\} = \frac{1}{5}; \Pr\{x_3\} = \Pr\{x_4\} = \frac{1}{10}; \Pr\{x_5\} = \Pr\{x_6\} = \Pr\{x_7\} = \Pr\{x_8\}.$$

Podać:

a) jednoznacznie dekodowalny kod zwięzły dla tego źródła, gdy alfabet kodu to: $\{*, \otimes, \bullet, \triangle\}$;

b) zbiór list długości ciągów kodowych dla wszystkich kodów zwięzłych jednoznacznie dekodowalnych tego źródła dla wszystkich możliwych długości używanych alfabetów.

13. Odczytać tekst 0101001111000110100000 zakodowany binarnym kodem Huffmana wiedząc, że względna częstość występowania symboli w tekście jest następująca: A(3), B(1), K(1), R(4), M(1) [tj. na trzy pojawienia się A mamy średnio cztery pojawienia się R itd.]. Gdyby w trakcie budowy drzewa Huffmana miała miejsce sytuacja, że do wyboru są więcej niż dwa wierzchołki przypisane konkretnemu symbolowi, wybrać wierzchołek związany z 0 jako przyporządkowany symbolowi występującemu wcześniej w alfabecie.

14. Na podstawie analizy kodowania Huffmana dla bezpamięciowego źródła generującego $N \geq 2$ wiadomości, pokazać że tworzone w alfabecie binarnym słowa kodowe o długościach l_i spełniają nierówność:

$$\sum_{i=1}^{N} l_i \leq \frac{1}{2}\left(N^2 + N - 2\right).$$

15. Mamy bezpamięciowe źródło wiadomości:

$$\begin{bmatrix} m_1 & m_2 & m_3 & m_4 & m_5 & m_6 & m_7 & m_8 \\ 0{,}03 & 0{,}16 & 0{,}11 & 0{,}31 & 0{,}07 & 0{,}07 & 0{,}19 & 0{,}06 \end{bmatrix}.$$

Zakodować wiadomości pochodzące z tego źródła kodem jednoznacznie dekodowalnym w taki sposób, żeby zminimalizować średni czas nadawania zakodowanej wiadomości, jeśli alfabet składa się z trzech liter: a, b, c. Czasy nadawania poszczególnych liter wynoszą: a — 3 sekundy, b — 2 sekundy oraz c — 3 sekundy. Podać średni czas nadawania zakodowanej wiadomości.

16. Mamy dane bezpamięciowe źródło wiadomości:

$$\begin{bmatrix} m_1 & m_2 & m_3 & m_4 & m_5 & m_6 & m_7 & m_8 & m_9 & m_{10} & m_{11} & m_{12} & m_{13} \\ 0{,}04 & 0{,}04 & 0{,}2 & 0{,}008 & 0{,}2 & 0{,}04 & 0{,}008 & 0{,}2 & 0{,}04 & 0{,}008 & 0{,}2 & 0{,}008 & 0{,}008 \end{bmatrix}.$$

Proszę znaleźć (jeśli jest to w ogóle możliwe) jednoznacznie dekodowalny optymalny kod źródłowy o średniej długości większej niż 1. Odpowiedni alfabet kodowy proszę zaproponować samemu.

1.4 Lektury

1.4.1 Materiał wykładu

Zagadnienia omówione w ramach tego wykładu są w dużym stopniu opisane w następujących książkach:

- Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., New York, NY, 1991: chapters 2, 4, 5, 8, 12.10, 16.
- Gareth A. Jones and J. Mary Jones. *Information and Coding Theory.* Springer-Verlag London Ltd., London, UK, 2000: chapters 2.6, 3.1-3.2, 3.5-3.7, 4.1-4.4, 4.6-4.8.
- Stefan M. Moser and Po-Ning Chen. *A Student's Guide to Coding and Information Theory.* Cambridge University Press, Cambridge, UK, 2012: chapters 1.1-1.4, 2.4, 4, 5.1-5.8, 6, 7.7.

1.4.2 Bibliografia uzupełniająca

- Dominic Welsh. *Codes and Cryptography.* Clarendon Press, Oxford, UK, 1988: klasyczna pozycja z teorii informacji, kodowania oraz kryptografii.
- Todd K. Moon. *Error Correction Coding.* John Wiley & Sons, Inc., Hoboken, NJ, 2005: książka poświęcona przede wszystkim kodowaniu nadmiarowego, ale z krótkim wstępem nt. modelowania źródeł informacji.