Egzamin z Teorii informacji. Teoria

8 lutego 2019

Dla informacji podajemy uzasadnienia również tam, gdzie nie były wymagane.

1. Jaka jest pojemność kanału danego poniższą macierzą?

$$\begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \tag{1}$$

   Odpowiedź: $\log 3$. Nietrudno zauważyć, że gdy do A wybierzemy jedynie trzy symbole, powiedzmy $1, 2, 3$ (spośród $\{1, 2, 3, 4, 5\}$), to odczytujemy bezbłędnie, a zatem $I(A; B) = H(A)$. Jeśli wybieramy je z jednakowym prawdopodobieństwem, uzyskujemy informację $\log 3$.

   Że jest to optimum, można sprawdzić metodą rozkładu macierzy, ale można też elementarnie. Przyjmijmy

$$p(A = i) = p_i,\ p(B = i) = q_i,$$

   dla $i = 1, \ldots, 5$. Chcemy zmaksymalizować $H(B) - H(B|A)$.

   Mamy $q_1 = \frac{1}{2}(p_1 + p_5) = q_5$ i podobnie $q_2 = \frac{1}{2}(p_2 + p_4) = q_4$, oraz $q_3 = p_3$; dalej $H(B|A) = p_1 + p_2 + p_4 + p_5$. Przy ustalonym q_3 wielkość $H(B)$ będzie największa, gdy $q_1 = q_2$[1]. Zauważmy dalej, że wielkość $H(B) - H(B|A)$ zależy tylko od sumy $p_1 + p_5$ w tym sensie, że pozostanie niezmieniona, jeśli weźmiemy $p_1' + p_5' = p_1 + p_5$; podobnie dla $p_2 + p_4$. Tak więc bez zmniejszenia ogólności możemy założyć, że prawdopodobieństwo jest skoncentrowane tylko na jednej z dwóch liczb; niech $p_5 = p_4 = 0$. Przy uczynionym już wcześniej założeniu, że $q_1 = q_2$, mamy $p_1 = p_2$. Przyjmując $p = p_1$, $H(B) - H(B|A)$ staje się funkcją p

$$-4 \cdot \frac{1}{2} p \log\left(\frac{1}{2} p\right) - (1 - 2p)\log(1 - 2p) - 2p = \\ -2p \log p + (2p - 1)\log(1 - 2p)$$

   Licząc pochodną otrzymujemy, że maksimum jest osiągnięte dla $p = \frac{1}{3}$ i jest to dokładnie ten rozkład, który zgadliśmy na początku; a więc jest on optymalny.

2. Rozważmy rodzinę kodów Hamminga. Czy kod (15,11) poprawia więcej błędów niż kod (7,4)?

   Nie, każdy kod Hamminga poprawia dokładnie jeden błąd.

3. Rzucamy sprawiedliwą kostką do gry i jeśli wynik jest k, to rzucamy k razy sprawiedliwą monetą. Niech A będzie wynikiem rzutu kostką, a B

[1] Ta własność pojawia się w wielu miejscach; można ją np. wydedukować z faktu, że funkcja $x \log x - (1 - x) \log x$ osiąga ekstremum, gdy $x = \frac{1}{2}$.

liczbą wyrzuconych orłów i niech $B' = B \bmod 2$. Porównać ze sobą liczby $I(A;B), I(A;B'), H(A)$.

Odpowiedź: $I(A;B') < I(A;B) < H(A)$**.**

Gdy rzucamy $k \geq 1$ razy sprawiedliwą monetą, rozkład parzystości jest zawsze $\frac{1}{2}$, tak więc $I(A;B') = 0$. Natomiast $0 < I(A;B)$, bo A i B oczywiście są zależne.Dalej, $I(A;B) < H(A)$, bo A nie zależy funkcyjnie of B.

4. Niech $f : \{0,1\}^n \to \{0,1\}^n$ będzie funkcją wzajemnie jednoznaczną. Rozważamy kryptosystem podobny do szyfru Vernama (*One time pad*), a więc zmienne M (wiadomości, ang. *messages*) i K (klucze) przyjmują wartości w $\{0,1\}^n$ i są niezależne, przy czym K ma rozkład jednostajny. Natomiast zmienna C (szyfrogramy, *cipher-texts*) jest określona

$$C \quad = \quad f(M \oplus K)$$

gdzie $\oplus$ oznacza *xor* po współrzędnych. Czy ten kryptosystem jest doskonale bezpieczny (*perfectly secret*)? Odpowiedź proszę krótko uzasadnić (bez wykonywania rachunków).

Tak, bo $I(M; f(M \oplus K)) \leq I(M; M \oplus K) = 0$, dla dowolnej funkcji f. Założenie, że f jest wzajemnie jednoznaczna potrzebne jest po to, by deszyfrowanie było dobrze określone.

5. Wyciągamy losowo (z jednostajnym prawdopodobieństwem) kartę z talii. Niech A oznacza jej kolor (czerwony, czarny), B typ (pik, trefl, karo, kier), a C wartość (nominał lub rodzaj figury). Czy istnieje taka kombinacja różnych $X, Y, Z \in \{A, B, C\}$, dla której

$$I(X;Y|Z) \quad > \quad I(X;Y)\,?$$

Czy odpowiedź na powyższe pytanie może się zmienić, jeśli nie będziemy zakładać, że rozkład, z jakim wyciągamy kartę jest jednostajny?

Odpowiedź proszę krótko uzasadnić.

Otóż **nie**, postulowana kombinacja **nie istnieje dla żadnego rozkładu**. Oczywiście, kolor jest funkcją typu. Rozważmy ogólnie sytuację trzech zmiennych losowych: $X, f(X), Y$. Wzajemna informacja między X a $f(X)$ może po dodaniu warunku co najwyżej zmaleć:

$$I(X; f(X)|Y) = H(f(X)|Y) - \underbrace{H(f(X)|X,Y)}_{0} \leq H(f(X)) = I(X; f(X)).$$

Podobnie, gdy funkcja działa na warunku:

$$I(Y; f(X)|X) = \underbrace{H(f(X)|X)}_{0} - \underbrace{H(f(X)|X,Y)}_{0} \leq I(Y; f(X)).$$

Wreszcie

$$I(Y; X|f(X)) = H(Y|f(X)) - \underbrace{H(Y|X, f(X))}_{=H(Y|X)} \leq H(Y) - H(Y|X) = I(Y; X),$$

gdzie zaznaczona równość entropii wynika z równoważności zdarzeń $(Y = y \wedge X = x \wedge f(X) = f(x)) \Leftrightarrow (Y = y \wedge X = x)$.

Inne rozwiązanie.[2] Przypomnijmy, że różnica $I(X;Y) - I(X;Y|Z)$, która nas tu interesuje, nie zależy od kombinacji, tzn. jest taka sama dla każdego podstawienia (różnych) A, B, C za X, Y, Z. (Wielkość ta w diagramie Venna zajmuje „sam środek" figury i jest zwykle oznaczana $R(X;Y;Z)$.) Zatem wystarczy raz ją policzyć, by poznać jej znak. Wobec zależności funkcyjnej A od B, mamy

$$I(A;B|C) = H(A|C) - \underbrace{H(A|B,C)}_{0} \leq H(A) = I(A;B).$$

Stąd wnioskujemy, że po dodaniu warunku informacja wzajemna może się co najwyżej zmniejszyć.

Ł.Dębowski, T.Gogacz, D.Niwiński

[2] Wg pracy egzaminacyjnej W.Przybyszewskiego.