

Przepustowość kanału, odczytywanie wiadomości z kanału, poprawa wydajności kanału.

Wiktor Miszuris

2 czerwca 2004

1 Przepustowość kanału

Zacznijmy od wprowadzenia równości $I(\mathcal{A}, \mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}) = H(\mathcal{A}) - H(\mathcal{A}|\mathcal{B})$. Można ją intuicyjnie zilustrować poniższym rysunkiem 1.



Rysunek 1: Intuicyjne przedstawienie związków pomiędzy entropią i informacją

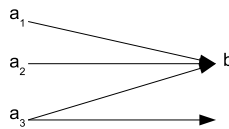
Intuicyjnie informacja $I(\mathcal{A}, \mathcal{B})$ jest to miara tego, co musimy jeszcze wiedzieć o $\mathcal{A}$, jeśli wiemy wszystko o $\mathcal{B}$. I na odwrót.

Założmy, że mamy dwa alfabetu A oraz B i kanał z A do B reprezentowany przez macierz

$$\begin{pmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \ddots & \vdots \\ p_{m,1} & \cdots & p_{m,n} \end{pmatrix}$$

gdzie $p_{i,j} = P(a_i \rightarrow b_j)$.

Prawdopodobieństwo wystąpienia symbolu b na wyjściu z kanału można opisać równaniem: $p(b) = \sum_{a \in \mathcal{A}} p(a)p(a \rightarrow b)$. Intuicyjnie można ten wzór zailustrować poniższym rysunkiem (prawdopodobieństwo wystąpienia na wyjściu symbolu b jest sumą iloczynów prawdopodobieństwa wystąpienia symbolu a na wejściu oraz prawdopodobieństwa przejścia w kanale z a na b).



Zdefiniujemy przepustowość kanału jako:

$$C = \max_{\mathcal{A}} I(\mathcal{A}, \mathcal{B})$$

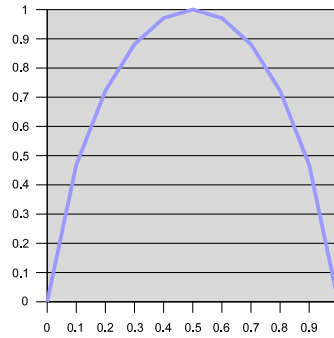
Rozkład $\mathcal{A}$, dla którego otrzymywane jest maksimum, nazywany jest optymalnym rozkładem wejściowym.

Przykład 1: Obliczyć przepustowość binarnego kanału symetrycznego.

Dla przypomnienia, binarnym kanałem symetrycznym nazywamy kanał charakteryzujący się następującymi cechami:

- Alfabetem wejściowym jest $A = \{0, 1\}$, alfabetem wyjściowym jest również $B = \{0, 1\}$.
- Macierzą reprezentującą kanał jest $\begin{pmatrix} P & \bar{P} \\ \bar{P} & P \end{pmatrix}$, gdzie $\bar{P} = 1 - P$

Wprowadźmy funkcję pomocniczą $\mathcal{H}_b$, nazywaną funkcją binarnej entropii $\mathcal{H}_b(x) = -x \log(x) - (1-x) \log(1-x)$. Jej wykres zaprezentowany jest poniżej.



Założmy, że rozkładem alfabetu $\mathcal{A}$ jest $P(\mathcal{A} = 0) = q$ oraz $P(\mathcal{A} = 1) = \bar{q}$. Naszym zadaniem jest zmaksymalizować wartość $I(\mathcal{A}, \mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A})$. Zauważmy, że:

$$\begin{aligned} H(\mathcal{B}) = \sum_{x \in \mathcal{B}} P(x) \log \frac{1}{P(x)} &= (Pq + \bar{P}\bar{q}) \log \frac{1}{Pq + \bar{P}\bar{q}} + (P\bar{q} + \bar{P}q) \log \frac{1}{P\bar{q} + \bar{P}q} = \\ &= \mathcal{H}_b(Pq + \bar{P}\bar{q}) \end{aligned}$$

oraz

$$\begin{aligned} H(\mathcal{B}|\mathcal{A}) &= P(\mathcal{A} = 0) [P(\mathcal{B} = 0|\mathcal{A} = 0) \log (P(\mathcal{B} = 0|\mathcal{A} = 0)^{-1}) + \\ &\quad + P(\mathcal{B} = 1|\mathcal{A} = 0) \log (P(\mathcal{B} = 1|\mathcal{A} = 0)^{-1})] + \\ &\quad + P(\mathcal{A} = 1) [P(\mathcal{B} = 0|\mathcal{A} = 1) \log (P(\mathcal{B} = 0|\mathcal{A} = 1)^{-1}) + \\ &\quad + P(\mathcal{B} = 1|\mathcal{A} = 1) \log (P(\mathcal{B} = 1|\mathcal{A} = 1)^{-1})] = \\ &= P(\mathcal{A} = 0) \left(P \log \frac{1}{P} + \bar{P} \log \frac{1}{\bar{P}} \right) + P(\mathcal{A} = 1) \left(\bar{P} \log \frac{1}{\bar{P}} + P \log \frac{1}{P} \right) = \\ &= P \log \frac{1}{P} + \bar{P} \log \frac{1}{\bar{P}} = \mathcal{H}_b(P) \end{aligned}$$

A zatem szukana przepustowość kanału binarnego, jest równa

$$\max_q I(\mathcal{A}, \mathcal{B}) = \mathcal{H}_b(Pq + \bar{P}\bar{q}) - \mathcal{H}_b(P) = \mathcal{H}_b(0.5) - \mathcal{H}_b(P)$$

Wartość 0.5 nie jest wybrana przypadkowo. Funkcja $\mathcal{H}_b(x)$ osiąga swoje maksimum w tym punkcie (patrz rysunek powyżej).

2 Odczytywanie wiadomości z kanału

Założmy, że dostaliśmy z kanału wiadomość b . W jaki sposób można się dowiedzieć, co miał na myśli nadawca? Lub inaczej mówiąc, jaka wiadomość a została oryginalnie włożona do kanału? W

tym celu zdefiniujemy pojęcie reguły decyzyjnej $\Delta : B \rightarrow A$. Funkcja ta ma za zadanie zwrócić jak najdokładniej symbol włożony przez nadawcę na wejściu kanału (przetłumaczyć wiadomość).

Rozpatrzmy tutaj dwa przypadki. Pierwszy – łatwy, polegający na założeniu, że wiemy z jakim prawdopodobieństwem jego symbole mogą się pojawiać na wejściu do kanału, znamy rozkład $\mathcal{A}$.

Możemy policzyć

$$p(\mathcal{A} = a | \mathcal{B} = b) = \frac{P(a \cap b)}{b} = \frac{P(a \rightarrow b)}{\sum_{a' \in \mathcal{A}} P(a') P(a' \rightarrow b)}$$

Jakością reguły decyzyjnej Δ dla danego symbolu b jest wartość $P(\Delta(b)|b)$. Jakością reguły Δ jest uśrednienie po wszystkich możliwych wartościach b .

Wprowadzimy miarę jakości reguły decyzyjnej w zależności od rozkładu $\mathcal{A}$.

$$\begin{aligned} P_{rC}(\Delta, \mathcal{A}) &= \sum_{b \in \mathcal{B}} P(b) P(\Delta(b)|b) = \sum_{b \in \mathcal{B}} \frac{P(b) P(\Delta(b) \cap b)}{P(b)} = \\ &= \sum_{b \in \mathcal{B}} P(\Delta(b) \cap b) \end{aligned}$$

Miarą dualną do miary jakości będzie $P_{rE}(\Delta, \mathcal{A}) = 1 - P_{rC}(\Delta, \mathcal{A})$. Jest to wielkość określająca średnie prawdopodobieństwo błędu.

Będziemy docelowo zmierzać w tym wykładzie do odpowiedzi, jaką regułę decyzyjną należy przyjąć, jeśli nie znamy rozkładu $\mathcal{A}$.

Zauważmy, że wygodniej jest patrzeć na macierz

$$\begin{pmatrix} P(a_1 \cap b_1) & \cdots & P(a_1 \cap b_n) \\ \vdots & \ddots & \vdots \\ P(a_m \cap b_1) & \cdots & P(a_m \cap b_n) \end{pmatrix} = \begin{pmatrix} P(a_1) & & \\ & \ddots & \\ & & P(a_m) \end{pmatrix} \begin{pmatrix} P(a_1 \rightarrow b_1) & \cdots & P(a_1 \rightarrow b_n) \\ \vdots & \ddots & \vdots \\ P(a_m \rightarrow b_1) & \cdots & P(a_m \rightarrow b_n) \end{pmatrix}$$

W aktualnie rozpatrywanym, prostym przypadku, dobrą regułą decyzyjną jest „reguła idealnego obserwatora”. Funkcja Δ w tym przypadku maksymalizuje wartość $P(a)P(a \rightarrow b)$.

W przypadku binarnego kanału symetrycznego (BSC)

$$\begin{pmatrix} Pq & \bar{P}q \\ \bar{P}\bar{q} & P\bar{q} \end{pmatrix} = \begin{pmatrix} q & 0 \\ 0 & \bar{q} \end{pmatrix} \begin{pmatrix} P & \bar{P} \\ \bar{P} & P \end{pmatrix}$$

reguła idealnego obserwatora będzie wyglądała następująco:

$$\Delta(0) = \begin{cases} 0, & \text{gdy } Pq \geq \bar{P}\bar{q} \\ 1, & \text{wpp} \end{cases}, \quad \Delta(1) = \begin{cases} 1, & \text{gdy } P\bar{q} \geq \bar{P}q \\ 0, & \text{wpp} \end{cases}$$

Przejdźmy do przypadku trudniejszego, nie znamy rozkładu $\mathcal{A}$. Pierwszym narzucającym się rozwiązaniem jest założenie, że $\mathcal{A}$ ma rozkład jednostajny i traktowanie tego zadania, jak poprzednio. Innym rozwiązaniem jest zdefiniowanie funkcji $\Delta(a) = b$ jako „reguły maksymalnego podobieństwa”, maksymalizującej $P(a \rightarrow b)$.

Reguła ta jest optymalna, argumentem za tym są następujące rozważania. Zdefiniujemy przestrzeń wszystkich możliwych rozkładów prawdopodobieństw $\mathcal{A} \in \mathcal{P}$.

$$\mathcal{P} = \{ \langle P_1, \dots, P_m \rangle \in [0..1]^m, P_1 + \dots + P_m = 1 \}$$

Obliczymy następującą całkę (reprezentującą średnie prawdopodobieństwo poprawnego odkodowania informacji) i zobaczymy, czy rzeczywiście będziemy osiągać maksimum w średnim przypadku dla reguły „maksymalnego podobieństwa”.

$$\begin{aligned}
\int_{\vec{p} \in \mathcal{P}} P_{rC}(\Delta, \vec{p}) d\vec{p} &= \int_{\vec{p} \in \mathcal{P}} \sum_{b \in \mathcal{B}} P(\Delta(b)) P(\Delta(b) \rightarrow b) d\vec{p} = \\
&= \sum_{b \in \mathcal{B}} \underbrace{P(\Delta(b) \rightarrow b)}_I \underbrace{\int_{\vec{p} \in \mathcal{P}} P(\Delta(b)) d\vec{p}}_{II}
\end{aligned}$$

Człon I nie zależy od $\vec{p}$, więc można było go wyprowadzić przed całkę. Człon II jest stały i nie zależy od permutacji liczb – przechodzi on całą dziedzinę $\mathcal{P}$, zależy on jedynie od m . A więc maksimum jest faktycznie osiągane w przypadku, gdy użyjemy reguły „maksymalnego podobieństwa” (człon I ma maksymalną wartość).

3 Poprawa wydajności kanału

Kolejnym zagadnieniem jest polepszenie kanału w ten sposób, żeby stał się on jak najbardziej niezawodny. Jedną z prostszych technik jest powtarzanie tego samego symbolu kilka razy. Rozpatrzmy przykład, w którym mamy do czynienia z symetrycznym kanałem binarnym $\begin{pmatrix} p & q \\ q & p \end{pmatrix}$, gdzie $q = \bar{p}$. Załóżmy, że każdy symbol powtarzamy trzy razy. I tak możemy napisać prawdopodobieństwa przejścia $P(0 \rightarrow 000) = ppp$, $P(0 \rightarrow 010) = pqp$ itd. Powstaje więc w ten sposób nowy kanał zdefiniowany następującą macierzą

$$\begin{array}{c}
\begin{pmatrix} p^3 & p^2q & p^2q & q^2p & p^2q & q^2p & q^2p & q^3 \\ q^3 & q^2p & q^2p & p^2q & q^2p & p^2q & p^2q & p^3 \end{pmatrix} \\
\text{Dla} \quad 000 \quad 001 \quad 010 \quad 011 \quad 100 \quad 101 \quad 110 \quad 111
\end{array}$$

Zdefiniujmy regułę decyzyjną dekodującą wiadomość w ten sposób, że dla komunikatu o większej ilości jedynek będzie to jedynka, zaś o większej ilości zer – zero. W ten sposób zdefiniowaliśmy trzeci kanał z $\{0, 1\}$ w $\{0, 1\}$ o macierzy

$$\begin{pmatrix} p^3 + 3p^2q & 3q^2p + q^3 \\ q^3 + 3q^2p & p^3 + 3p^2q \end{pmatrix}$$

Porównajmy $P_{rC,1}$ dla zwykłego binarnego kanału symetrycznego (z prawdopodobieństwem poprawnego przesłania symbolu p) oraz dla nowo otrzymanego kanału. Dla zwykłego kanału, i rozkładu normalnego $\mathcal{A}$,

$$P_{rC,1} = \sum_{i \in \{0,1\}} P(\mathcal{B} = i) P(\mathcal{A} = id(i) | \mathcal{B} = i) = \sum_{i \in \{0,1\}} P(\mathcal{A} = id(i)) P(\mathcal{B} = i | \mathcal{A} = id(i)) = p$$

A więc prawdopodobieństwo błędu wynosi dla zwykłego kanału $P_{rE,1} = 1 - PrC, 1 = q$.

Natomiast w przypadku potrójnego powtarzania symbolu, odpowiednio $P_{rE,3} = q^3 + 3q^2p$. Widać, że $P_{rE,1} < P_{rE,3}$ dla $q < 1/4$.

W ogólnym przypadku, jeśli ten sam symbol będzie powtarzany n razy, prawdopodobieństwo błędu będzie wynosiło $P_{rE,n} = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{i} p^i q^{n-i}$. Oszacujmy, jak wielki jest to błąd w przypadku granicznym. Obliczenia będziemy przeprowadzać dla sytuacji $q < p$, gdyż w sytuacji $p = q = 0.5$ prawdopodobieństwo błędu wynosi 0.5 (co jest łatwe do obliczenia). Nie rozpatrujemy tego przypadku, gdyż wskazuje on na to, że kanał kompletnie ignoruje wejście i wyjście nie jest w żaden sposób zależne od symbolu wejściowego. Równie dobrze moglibyśmy podrzucić monetę i podać wynik losowania.

Wróćmy do oszacowania. Załóżmy, że $q < p$ (a więc kanał jest bardziej prawdziwy niż zakłócający). Zauważyć można, że $P_{rE,n} \leq \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{i} p^{\lceil n/2 \rceil} q^{\lfloor n/2 \rfloor}$. Dla ułatwienia, załóżmy, że $n = 2k$. $P_{rE,n} \leq p^k q^k \sum_{i=0}^k \binom{2k}{i} = p^k (1-p)^k \sum_{i=0}^k \binom{2k}{i} = (p(1-p))^k \left(\frac{2^{2k}}{2} \right)$. W przypadku $n = 2k + 1$ rozumowanie jest bardzo podobne.

Funkcja $p(1-p)$ przyjmuje maksimum w punkcie $p = 0.5$. Jak założyliśmy $p \in [0, 0.5)$, a więc $p(1-p) \in [0, 1/4)$. Połóżmy $p(1-p) = \frac{t}{4}$, gdzie $t \in [0, 1)$. Przepiszmy $P_{rE,n} \leq \left(\frac{t}{4}\right)^k 2^{2k-1} = t^k \frac{1}{2}$.

Z twierdzenia o trzech ciągach wynika, że w przypadku granicznym, $\lim_{n \rightarrow \infty} P_{rE,n} = 0$.

Faktem więc jest, że wydłużając sztucznie komunikat, zwiększamy szanse poprawnego jego odczytania.