Fundamentalne twierdzenie Shannona

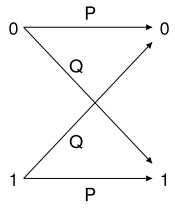
Na podstawie wykładu i notatki dr hab. Damiana Niwińskiego
przygotował Łukasz Chmielewski

22/04/2004

Przypuśćmy, że mamy do dyspozycji binarny kanał symetryczny Γ o macierzy:

$$\begin{pmatrix} P & Q \\ Q & P \end{pmatrix}$$

i grafie:



którym chcemy przesyłać wiadomości, stanowiące wartości pewnej zmiennej losowej $\mathcal{A}$. Przyjmijmy, że zbiorem wartości zmiennej $\mathcal{A}$ jest $A = \{a_1, \ldots, a_m\}$. Najogólniej biorąc, metoda polega na tym, by dobrać kod $\mathcal{C} : A \to \{0,1\}^*$, a następnie przesyłać słowa kodowe bit po bicie. Jednak, z wyjątkiem przypadku, gdy kanał jest wierny ($P = 1$), proste odczytanie otrzymanego słowa będzie na ogół prowadzić do przekłamań — prawdopodobieństwo bezbłędnego przesłania słowa w jest (zaledwie) $P^{|w|}$. Podobnie, zamiana przy odczytywaniu 0 i 1 nie będzie dobrą metodą, chyba że $Q = 1$. Dlatego też w tym zadaniu nie będzie nas interesował kod o minimalnej długości, przeciwnie — przyda się pewna redundancja.

Metoda będzie polegała na znalezieniu kodu

$$\mathcal{C} : A \to C \subseteq \{0,1\}^n$$

gdzie n jest sporo większe niż $\log_2 m$ i zastosowaniu przy odczytywaniu reguły dobrosąsiedzkiej. Zakładamy oczywiście, że kod $\mathcal{C}$ jest znany zarówno nadawcy jak odbiorcy. A zatem, jeśli odbiorca otrzyma z kanału słowo $w \in \{0,1\}^n$, to „domyśla się", że wysłane było takie słowo kodowe $\Delta(w) \in C$, które różni się od w na możliwie najmniejszej liczbie pozycji (a więc nadaną wiadomością było $\mathcal{C}^{-1}(\Delta(w))$).

Najważniejszą dla nas wielkością jest **średnie prawdopodobieństwo błędu** przy zastosowaniu kodu $\mathcal{C}$ i reguły dobrosąsiedzkiej Δ. Wyrażamy je wzorem:

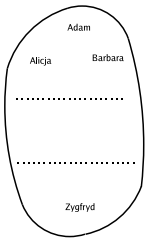
$$Pr_E(\Delta, \mathcal{CA}) = 1 - Pr_C(\Delta, \mathcal{CA})$$

gdzie

$$Pr_C(\Delta, \mathcal{CA}) = \sum_{u \in \{0,1\}^n} p(\mathcal{B} = u) \cdot p(\mathcal{CA} = \Delta(u) | \mathcal{B} = u)$$

W powyższym wzorze $\mathcal{CA}$ jest zmienną losową powstałą przez złożenie zmiennej $\mathcal{A}$ i funkcji $\mathcal{C}$ (a więc $p(\mathcal{CA} = w) = p(\mathcal{A} = \mathcal{C}^{-1}(w))$). Natomiast zmienna losowa $\mathcal{B}$ opisuje słowa (długości n) wychodzące z kanału, tzn.

$$p(\mathcal{B} = v) = \sum_{w \in C} p(\mathcal{CA} = w) \cdot P^{n - d(v,w)} \cdot Q^{d(v,w)}$$



Rysunek 1: Zbiór znajomych.
Sytuacja: rozmowa i przedstawianie się. Ponieważ kod nie wypełnia przestrzeni słów to duża jest szansa na domyślenie się poprawnego imienia w przypadku błędnego przedstawienia się.

gdzie $d(v, w)$ oznacza odległość Hamminga słów v i w (czyli liczbę pozycji gdzie się różnią).

Możemy ponownie sformułować:

Twierdzenie Shannona. *Niech Γ będzie binarnym symetrycznym kanałem o macierzy $\begin{pmatrix} P & Q \\ Q & P \end{pmatrix}$; zakładamy że $P > \frac{1}{2} > Q$.*
Niech $\delta, \epsilon > 0$.
Wtedy istnieje n_0 takie, że dla każdego $n \geqslant n_0$ istnieje zbiór słów $C \subseteq \{0,1\}^n$ o mocy $|C| = m$ spełniającej

$$\boxed{C_\Gamma - \epsilon \leqslant \frac{\log_2 m}{n} \leqslant C_\Gamma}$$

taki, że dla dowolnej zmiennej losowej $\mathcal{A}$, której zbiór wartości A ma tę samą moc m, istnieje kod $\mathcal{C} : A \to C$ spełniający

$$Pr_E(\Delta, \mathcal{CA}) < \delta.$$

Objaśnienia:

- $C_\Gamma = 1 - H(P)$ - pojemność kanału;
- Δ - reguła dobrosąsiedzka;

Uwaga 1. Oczywiście powyższy kod $\mathcal{C}$ jest bijekcją A na C.
Uwaga 2. W tej sytuacji stopa kodu $\mathcal{C}$, $R(\mathcal{C}) = \frac{\log_2 m}{n}$.

Wniosek 1. Niech Γ, δ, ϵ będą tak jak w Twierdzeniu Shannona. Wtedy istnieje m_0 takie, że dla każdej zmiennej losowej $\mathcal{A}$ o zbiorze wartości A mocy $|A| = m \geqslant m_0$, istnieje kod $\mathcal{C} : A \to \{0,1\}^n$ (dla pewnego n) taki, że $Pr_E(\Delta, \mathcal{CA}) < \delta$ oraz $C_\Gamma - \epsilon \leqslant \frac{\log_2 m}{n} \leqslant C_\Gamma$.

W szczególności, jeśli tylko $\epsilon < C_\Gamma$, to n spełnia nierówność $n \leqslant \frac{\log_2 m}{C_\Gamma - \epsilon}$ (a więc wydłużenie nie jest zbyt duże).

Dowód. Zastosujemy Twierdzenie Shannona dla Γ, δ i pewnego ϵ_1, od którego wymagamy $\epsilon_1 < \epsilon$ i $\epsilon_1 < C_\Gamma$. Wiemy, że dla n większych od pewnego n_0 możemy wskazać $m = m(n)$ spełniające

$$C_\Gamma - \epsilon_1 \leqslant \frac{\log_2 m(n)}{n} \leqslant C_\Gamma$$

i takie, że o ile zbiór wartości zmiennej losowej $\mathcal{A}$ ma moc $|A| = m(n)$, to istnieje kod $\mathcal{C} : A \to \{0,1\}^n$ spełniający $Pr_E(\Delta, \mathcal{CA}) < \delta$.

Przypuśćmy teraz, że mamy dowolną zmienną $\mathcal{A}$ i $|A| = m$. Jeżeli m jest postaci $m(n)$ dla pewnego n, to mamy szczęście: Twierdzenie Shannona (dla $\Gamma, \delta, \epsilon_1$) daje nam to, czego chcieliśmy dla Γ, δ, ϵ; w szczególności mamy $\frac{\log_2 m}{n} \geqslant C_\Gamma - \epsilon_1 \geqslant C_\Gamma - \epsilon$.

Jeśli nie, to wybierzmy najmniejsze $n \geqslant n_0$ takie, że $n(C_\Gamma - \epsilon_1) \geqslant \log_2 m$. Mamy więc

$$n(C_\Gamma - \epsilon_1) \geqslant \log_2 m \geqslant (n-1)(C_\Gamma - \epsilon_1)$$

przy czym ostatnia nierówność zachodzi o ile m będzie na tyle duże, by zagwarantować, że $n - 1 \geqslant n_0$. W tym celu wystarczy założyć, że $\log_2 m_0 \geqslant (n_0 + 1)C_\Gamma$.

Przekształcając ostatnią nierówność otrzymujemy

$$\frac{\log_2 m}{n} \cdot \left(\frac{n}{n-1}\right) > C_\Gamma - \epsilon_1$$

skąd

$$\frac{\log_2 m}{n} > (C_\Gamma - \epsilon_1)\left(\frac{n-1}{n}\right).$$

Jednak, o ile tylko $\epsilon_1 < \epsilon$, to dla dostatecznie dużych n mamy

$$(C_\Gamma - \epsilon_1)\left(\frac{n-1}{n}\right) \geqslant C_\Gamma - \epsilon.$$

Zatem, ewentualnie jeszcze zwiększając próg m_0, otrzymamy upragnioną nierówność

$$\frac{\log_2 m}{n} \geqslant (C_\Gamma - \epsilon_1)\left(\frac{n-1}{n}\right) \geqslant C_\Gamma - \epsilon.$$

Ta uwaga kończy dowód Wniosku. □

Teraz wyjaśnimy dlaczego można ograniczyć się do rozważania zmiennych $\mathcal{A}$ o **rozkładzie jednostajnym**, tzn. $p(\mathcal{A} = w) = \frac{1}{m}$, dla każdego $w \in C$.

W tym celu znajdziemy najpierw bardziej dogodne przedstawienie $Pr_E(\Delta, \mathcal{CA})$. Użyteczna transformacja, oparta na wzorze Bayesa, pozwala nam przejść od uśredniania po $\mathcal{B}$ do uśredniania po $\mathcal{A}$.

$$\begin{aligned}
Pr_E(\Delta, \mathcal{CA}) &= 1 - \sum_{u \in \{0,1\}^n} p(\mathcal{B} = u) \cdot p(\mathcal{CA} = \Delta(u) | \mathcal{B} = u) \\
&= 1 - \sum_{u \in \{0,1\}^n} p(\mathcal{B} = u \wedge \mathcal{CA} = \Delta(u)) \\
&= \sum_{u \in \{0,1\}^n} \sum_{w \neq \Delta(u)} p(\mathcal{B} = u \wedge \mathcal{CA} = w) \\
&= \sum_{w \in C} \sum_{w \neq \Delta(u)} p(\mathcal{B} = u \wedge \mathcal{CA} = w) \\
&= \sum_{w \in C} p(\mathcal{CA} = w \wedge \Delta \circ \mathcal{B} \neq w) \\
&= \sum_{w \in C} p(\mathcal{CA} = w) \cdot p(\Delta \circ \mathcal{B} \neq w | \mathcal{CA} = w)
\end{aligned}$$

Dwukrotnie pojawiające się w powyższym ciągu równości sumę wewnętrzną $\sum_{w \neq \Delta(u)}$ trzeba właściwie zrozumieć: w pierwszym przypadku, dla u ustalonego w sumie zewnętrznej, sumujemy po wszystkich $w \in C$ spełniających $w \neq \Delta(u)$; w drugim przypadku, dla w ustalonego w sumie zewnętrznej, sumujemy po wszystkich $u \in \{0,1\}^n$ spełniających tę nierówność.

Niech teraz $\mathcal{E}$ będzie zmienną losową o wartościach w $\{0,1\}^n$ powstałą jako konkatenacja (produkt) n niezależnych zmiennych losowych $\mathcal{X}_i$ o rozkładzie $p(\mathcal{X}_i = 1) = Q$ i $p(\mathcal{X}_i = 0) = P$. Mamy więc

$$p(\mathcal{E} = w) = \prod_{i=1}^{n} P^{1-w_i} Q^{w_i}$$

W naszej intencji $\mathcal{E}$ reprezentuje ciąg „zachowań" kanału przy przesyłaniu słowa n bitowego: 1 jeśli błąd, 0 jeśli transmisja poprawna. Wtedy $\mathcal{B} = \mathcal{CA} \oplus \mathcal{E}$ (gdzie $\oplus$ rozumiemy jak *XOR*, tzn. dodawanie bitów po współrzędnych *modulo* 2) i, jak łatwo sprawdzić, $p(\mathcal{B} = w|\mathcal{CA} = u) = p(\mathcal{E} = e)$, gdzie $u \oplus e = w$.

W konsekwencji otrzymujemy

$$p(\Delta \circ \mathcal{B} \neq w|\mathcal{CA} = w) = \sum_{v:\Delta(v)\neq w} p(\mathcal{B} = v|\mathcal{CA} = w) = \sum_{e:\Delta(w\oplus e)\neq w} p(\mathcal{E} = e) = p(\Delta(w \oplus \mathcal{E}) \neq w)$$

i wreszcie

$$Pr_E(\Delta, \mathcal{CA}) = \sum_{w\in C} p(\mathcal{CA} = w) \cdot p(\Delta(w \oplus \mathcal{E}) \neq w).$$

Jeżeli $\mathcal{A}$ ma rozkład jednostajny, to

$$Pr_E(\Delta, \mathcal{CA}) = \frac{1}{m} \sum_{w\in C} p(\Delta(w \oplus \mathcal{E}) \neq w).$$

Lemacik („lemat o skarbonce"). Niech będzie dany układ liczb rzeczywistych $\alpha_1, \ldots \alpha_m$ oraz liczby $p_1, \ldots, p_m \in [0, 1]$, takie, że $p_1 + \ldots + p_m = 1$. Wtedy istnieje permutacja $\sigma : \{1, \ldots, m\} \to \{1, \ldots, m\}$, taka że

$$\sum_{i=1}^{m} p_{\sigma(i)}\alpha_i \leqslant \frac{1}{m} \sum_{i=1}^{m} \alpha_i.$$

Dowód. Wykażemy przez indukcję po m, że o ile $\alpha_1 \leqslant \ldots \leqslant \alpha_m$ i $p_1 \geqslant \ldots \geqslant p_m$, to

$$\sum_{i=1}^{m} p_i\alpha_i \leqslant \frac{1}{m} \sum_{i=1}^{m} \alpha_i.$$

Jeśli nasz układ nie spełnia powyższego wstępnego założenia, możemy je oczywiście uzyskać przez odpowiednią permutację; stad teza **Lemaciku**.

Dla $m = 1$ nie ma czego dowodzić. W kroku indukcyjnym rozważmy układy $\alpha_1 \leqslant \ldots \leqslant \alpha_m$ i $p_1 \geqslant \ldots \geqslant p_m$. Z ustawienia ciągu p_i mamy, że $p_m \leqslant \frac{1}{m}$, powiedzmy $p_m = \frac{1}{m} - h$, dla pewnego $h \geqslant 0$. Natomiast z ustawienia ciągu α_i mamy, że $\alpha_m \geqslant \frac{1}{m-1} \sum_{i=1}^{m-1} \alpha_i$.

Z założenia indukcyjnego mamy

$$\frac{p_1}{p_1 + \ldots + p_{m-1}}\alpha_1 + \ldots + \frac{p_m}{p_1 + \ldots + p_{m-1}}\alpha_m \quad \leqslant \quad \frac{1}{m-1} \sum_{i=1}^{m-1} \alpha_i$$

skąd wnioskujemy dalej

$$\begin{aligned}
p_1\alpha_1 + \ldots + p_{m-1}\alpha_{m-1} + p_m\alpha_m &\leqslant \overbrace{(p_1 + \ldots + p_{m-1})}^{1-p_m} \cdot \frac{1}{m-1} \cdot \sum_{i=1}^{m-1} \alpha_i + p_m\alpha_m \\
&= \left(\frac{m-1}{m} + h\right) \cdot \frac{1}{m-1} \cdot \sum_{i=1}^{m-1} \alpha_i + \left(\frac{1}{m} - h\right) \cdot \alpha_m \\
&= \frac{1}{m} \sum_{i=1}^{m} \alpha_i + h \cdot \underbrace{\left(\frac{1}{m-1} \sum_{i=1}^{m-1} \alpha_i - \alpha_m\right)}_{\leqslant 0} \\
&\leqslant \frac{1}{m} \sum_{i=1}^{m} \alpha_i
\end{aligned}$$

jak chcieliśmy.

□

Przyjmując, że $C = \{w_1, \ldots, w_m\}$ i stosując Lemacik do układu liczb $\alpha_i = p(\Delta(w_i \oplus \mathcal{E}) \neq w_i)$ oraz $p_i = p(\mathcal{CA} = w_i)$, stwierdzamy, że dla pewnej permutacji σ zbioru C

$$\sum_{w \in C} p(\mathcal{CA} = \sigma(w)) \cdot p(\Delta(w \oplus \mathcal{E}) \neq w) \leqslant \frac{1}{m} \sum_{w \in C} p(\Delta(w \oplus \mathcal{E}) \neq w).$$

Jeśli teraz określimy nowy kod $\mathcal{C}' = \sigma^{-1} \circ \mathcal{C}$, to lewa strona powyższej nierówności przedstawia właśnie $Pr_E(\Delta, \mathcal{C}'\mathcal{A})$ (bo $\mathcal{CA} = \sigma(w) \Leftrightarrow \mathcal{C}'\mathcal{A} = w$). A zatem jeśli zbiór słów kodowych C jest dobry dla zmiennej losowej $\mathcal{A}$ z rozkładem jednostajnym, to jest dobry dla zmiennych o dowolnych rozkładach (z dokładnością do wyboru kodu $\mathcal{C}$).

Możemy rozumieć to tak: przypuśćmy, że chcemy przesyłać informację z błędem mniejszym niż δ, ale również nie za bardzo redundantnie, mianowicie ze stopą większą niż $C_\Gamma - \epsilon$. Twierdzenie mówi, że to jest możliwe dla wszystkich dostatecznie dużych n, i w konsekwencji dla wszystkich dostatecznie dużych m (wystarczy, że $\log_2 m \geqslant n_0(C_\Gamma - \epsilon)$; zauważmy, że kod „dobry" dla $m' \geqslant m$ jest tak samo dobry dla m).

$$\begin{aligned} p(\Delta \circ \mathcal{B} \neq w | \mathcal{CA} = w) &= \sum_{v : \Delta(v) \neq w} p(\mathcal{B} = v | \mathcal{CA} = w) \\ &= \sum_{e : \Delta(w \oplus e) \neq w} p(\mathcal{E} = e) \\ &= p(\Delta(w \oplus \mathcal{E}) \neq w) \end{aligned}$$

Uwaga 3. W powyższych rozważaniach mieliśmy zawsze na myśli Twierdzenie Shannona dla binarnych symetrycznych kanałów. W istocie zachodzi ono w ogólniejszej postaci dla dowolnych kanałów.