Arimoto-Blahut algoritmus csatornakapacitás számítására

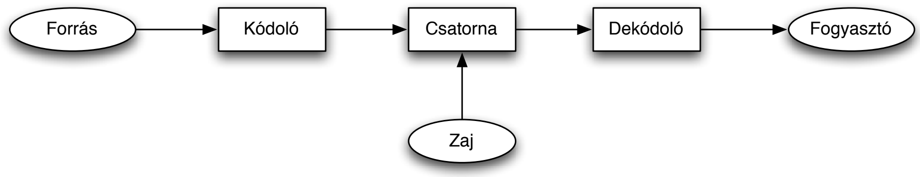
Kasza Péter

2009. május 3.

Tartalomjegyzék

1. Egyirányú hírközlési modell

Az alábbi egyirányú hírközlési modellt Shannon és Weaver alkották meg az 1949-as évek környékén. A modell egyirányúsága annyit jelent, hogy a forrás csak adóként, a fogyasztó pedig csak vevőként működhet. Az ilyen egyirányú csatorna a gyakorlatban ritka, de egyszerűsége miatt elméleti jelentősége számottevő.



1. ábra. Shannon-féle egyirányú hírközlés modellje

Legyen $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ a forrás ábécé, $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\}$ a kódszavak halmaza. Ekkor a $g : \mathcal{X} \to \mathcal{Y}$ leképzést kódolásnak nevezzük. Zajmentes esetben a kódszavak változás nélkül haladnak át a csatornán.

1.1. Definíció (Sztochasztikus és duplán sztochasztikus mátrix)**.** Az $A \in \mathbb{R}^{n \times m}$ mátrixot sztochasztikus mátrixnak nevezzük, ha $\forall(i,j)$-re $A_{ij} \geq 0$ és az a. vagy b. pontok közül legalább egy teljesül.

a.
$$\sum_{i=1}^{n} A_{ij} = 1, \quad (1 \leq j \leq m)$$

b.
$$\sum_{j=1}^{m} A_{ij} = 1, \quad (1 \leq i \leq n)$$

Ha az A illetve A^T mátrixok sztochasztikus mátrixok, akkor az A mátrixot duplán sztochasztikusnak nevezzük.

1.2. Definíció (Csatornamátrix)**.** A csatornamátrix egy olyan duplán sztochasztikus mátrix, amely az egyes szimbólumok átviteli valószínűségeit tartalmazza.

$$T_{ij} = P(\xi = x_i, \eta = y_j)$$

Megjegyzés. Szimmetrikus csatorna esetén a csatornamátrix ortogonális, azaz:

$$T = T^T$$

1.1. Jelölések és definíciók

Jelölés. A P és Q eloszlások együttes eloszlását független esetben a következő képpen jelöljük:

$$(p_1q_1, p_2q_2, \dots, p_nq_n) = (P \times Q)$$

1.3. Definíció (Információmennyiség)**.** A P eloszlás információmennyiségén, más néven entrópiáján az alábbi összefüggést értjük:

$$H(P) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p_i \log_2 p_i$$

1.4. Definíció (Kölcsönös információmennyiség)**.** Legyenek ξ, η valószínűségi változók.

$$I(\xi, \eta) = H(\xi) - H(\xi|\eta)$$

1.5. Definíció (I-divergencia)**.** A Q és P eloszlások I-Divergenciáján az alábbi várhatóértéket értjük:

$$D(Q\|P) = \sum_{i=1}^{n} q_i \log_2 \frac{q_i}{p_i}$$

1.6. Definíció (Csatornakapacitás)**.** A csatornakapacitás az a maximális információmennyiség, amelyet még megbízhatóan tudunk továbbítani a csatornán.

2. Csatornakapacitás zajos esetben

Legyen $\xi : Q = \{q_1, \dots, q_n\}$ a bementi, $\eta : Q' = \{q'_1, \dots, q'_n\}$ a kimineti eloszlás. A T csatornamátrix felhasználásával a kimeneti eloszlás az alábbi formában írható fel:

$$Q' = TQ$$

Jelölje $P = (q_1T_1, \dots, q_nT_n)$ az együttes eloszlást. Ekkor a kölcsönös információmennyiség az alábbi formában írható fel:

$$I(\xi, \eta) = D(P\|Q \times TQ) = v(Q)$$

A csatornakapcitást zajos esetben a bemeneti és a kimeneti eloszlások kölcsönös információmennyiségének maximumaként definiálhatjuk:

$$\max_{Q} v(Q) = C$$

3. Arimoto-Blahut algoritmus

Az Arimoto-Blahut algoritmus egy iteratív algoritmus a csatornakapcitás meghatározására. Az algoritmus a kölcsönös információmennyiség maximalizálásával határozza meg az optimális eloszlást és a csatornakapacitást. Az algoritmust elsősorban zajos csatorna esetén alkalmazzák, többek közt mivel zajmentes esetben léteznek sokkalta hatékonyabb megoldások, valamint a zajmentes esetet az algoritmus csak abban az esetben adja vissza, ha a bemeneti eloszlás egyenletes. [1]

3.1. Lemma (Log-szumma egyenlőtlenség)**.**

$$\sum_{i=1}^{n} a_i \log_2 \frac{a_i}{b_i} \geq a \log_2 \frac{a}{b} \tag{1}$$

$$a = \sum_{i=1}^{n} a_i, \quad b = \sum_{i=1}^{n} b_i > 0 \tag{2}$$

Állítás (Az I-divergencia nemnegativtása)*.* Az I-divergencia értéke bármely Q, P eloszlásra nemnegatív.

$$D(Q\|P) \geq 0$$

Bizonyítás. Log-szumma egyenlőtlenség felhasználásával:

$$a = \sum_{i=1}^{n} q_i = 1, \quad b = \sum_{i=1}^{n} p_i = 1 \tag{3}$$

$$\underbrace{\sum_{i=1}^{n} q_i \log_2 \frac{q_i}{p_i}}_{D(Q\|P)} \geq 0 \tag{4}$$

□

3.2. Lemma (Adatfeldolgozási lemma)**.**

$$D(TQ\|TQ') \leq D(Q\|Q') \tag{5}$$

Megjegyzés. Az adatfeldolgozási lemma állítása szerint, a bemeneti adatokra bármilyen adatfeldolgozási műveletet alkalmazva, az eredeti adat információmennyiségét csak csökkenteni tudjuk.

3.3. Lemma. *Legyenek* $\alpha, f_1, \ldots, f_m \in \mathbb{R}$*, ekkor:*

$$H(Q) \leq \alpha \sum_{i=1}^{m} q_i f_i + \log_2 \left(\sum_{i=1}^{m} 2^{-\alpha f_i} \right) \tag{6}$$

Jelölés. Vezessük be az alábbi jelöléseket:

$$D_j(R) := D(T_j \| R) \tag{7}$$

$$D(T\|T|Q) := \sum_{j=1}^{m} q_j D_j(R) \tag{8}$$

$$F(Q, Q') := D(T\|TQ'|Q) - D(Q\|Q') \tag{9}$$

Tekintsük az előbb bevezetett jelölések néhány tulajdonságát:

Állítás.

$$D(T\|R|Q) = v(Q) + D(TQ\|R) \tag{10}$$

Bizonyítás.

$$\sum_{j=1}^{m} q_j D_j(R) = \sum_{j=1}^{m} q_j D(T_j\|R) = \tag{11}$$

$$\sum_{j=1}^{m} q_j \sum_{i=1}^{k} t_{ij} \log_2 \frac{t_{ij}}{r_i} = \tag{12}$$

$$\sum_{j=1}^{m} \sum_{i=1}^{k} \underbrace{q_{ij} t_{ij}}_{p_{ij}} \log_2 \frac{t_{ij}}{r_i} = \tag{13}$$

$$\tag{14}$$

Az $r'_i = \sum_{j=1}^{m} q_j t_{ij}$ változók bevezetésével:

$$\underbrace{\sum_{j=1}^{m} \sum_{i=1}^{k} q_j t_{ij} \log_2 \frac{t_{ij} q_j}{r_i q_j}}_{D(P\|Q \times R)} \frac{r'_i}{r'_i} = \tag{15}$$

$$\underbrace{\sum_{j=1}^{m} \sum_{i=1}^{k} q_j t_{ij} \log_2 \frac{t_{ij} q_j}{q_j r'_i}}_{D(P\|Q \times TQ)} + \underbrace{\sum_{j=1}^{m} \sum_{i=1}^{k} q_j t_{ij} \log_2 \frac{r'_i}{r_i}}_{\sum_{i=1}^{k} r'_i \log_2 \frac{r'_i}{r_i}} = \tag{16}$$

$$v(Q) \quad + \quad D(TQ\|R) \tag{17}$$

$\square$

Állítás.

$$F(Q, Q') = v(Q) + D(TQ \| TQ') - D(Q \| Q') \leq v(Q) \tag{18}$$

Bizonyítás. A 10. egyenletet felhasználva:

$$F(Q, Q') = \underbrace{D(T \| TQ' | Q)}_{v(Q) + D(TQ \| TQ')} - D(Q \| Q')$$

Az adatfeldolgozási lemmát, valamint az I-divergencia nemnegativitását felhasználva:

$$v(Q) = D(P \| Q \times TQ) \geq 0, \quad D(Q \| Q') \geq 0, \quad D(TQ \| TQ') \geq 0 \tag{19}$$

$$D(TQ \| TQ') - D(Q \| Q') \leq 0 \tag{20}$$

$$v(Q) + D(TQ \| TQ') - D(Q \| Q') \leq v(Q) \tag{21}$$

□

Ekkor a 18. egyenletből:

$$D(Q \| Q') \leq v(Q) \tag{22}$$

$$v(Q) = F(Q, Q) \geq F(Q, Q') \tag{23}$$

A $v(Q)$ értékét tehát $F(Q, Q')$ maximalizálásával számíthatjuk ki.

$$F(Q, Q') \to \max$$

Megjegyzés. Az I-divergencia tulajdonságaiból adódóan egyenlőség a 22. egyenletben csak akkor következik be, ha:

$$D(TQ \| TQ') = 0, \quad TQ \equiv TQ'$$

Állítás. A csatornakapacitás értéke:

$$\max_Q F(Q, Q') = \log_2 S'$$

ha az iteráció során a következő eloszlásnak az alábbiakat választjuk:

$$q_j = \frac{q'_j 2^{D_j(TQ')}}{S'}, \quad S = \sum_{j=1}^{m} q'_j 2^{D_j(TQ')}$$

Bizonyítás. A 6. egyenletbe az alábbiakat helyettesítve:

$$\alpha = -1, \quad f_j = \log_2 q'_j + D_j(TQ')$$

a következők teljesülnek:

$$H(Q) \leq -\sum_{j=1}^{m} q_j f_j + \log_2 \sum_{j=1}^{m} 2^{f_i} = \tag{24}$$

$$-\sum_{j=1}^{m} q_j(\log_2 q'_j + D_j(TQ')) + \log_2 \left(\sum_{j=1}^{m} 2^{\log_2 q'_j + D_j(TQ')} \right) = \tag{25}$$

$$-\sum_{j=1}^{m} q_j \log_2 q'_j - \sum_{j=1}^{m} q_j D_j(TQ') + \log_2 \left(\sum_{j=1}^{m} q'_j 2^{D_j(TQ')} \right) \tag{26}$$

$$\sum_{j=1}^{m} q_j \log_2 \frac{q'_j}{q_j} \leq -D(T \| TQ' | Q) + \log_2 S' \tag{27}$$

$$\underbrace{D(T \| TQ' | Q) - D(Q \| Q')}_{F(Q,Q')} \leq \log_2 S' \tag{28}$$

$$\max_{Q} F(Q, Q') = \log_2 S' \tag{29}$$

□

Arimoto-Blahut($Q^{(0)}$, T)
[$Q^{(0)}$ - bemeneti eloszlás, T - csatornamátrix]

  $S^{(0)} \leftarrow 0$

  $N \leftarrow 1$

  WHILE *True* **DO**

    $S^{(N)} \leftarrow \sum_{j=1}^{m} q_j^{(N)} 2^{D_j\left(TQ^{(N)}\right)}$

    $q_j^{(N+1)} \leftarrow \frac{q_j^{(N)} 2^{D_j\left(TQ^{(N)}\right)}}{S^{(N)}}$

    IF $|\log_2 S^{(N)} - \log_2 S^{(N-1)}| < \varepsilon$ **THEN**

      BREAK

    ENDIF

    $N \leftarrow N+1$

  ENDWHILE

2. ábra. Arimoto-Blahut algoritmus

Hivatkozások

[1] Fegyverneki Sándor. Információelmélet jegyzet. *Miskolci Egyetem*, 2002.