Résumé AICC II

Arthur Herbette

Lundi 02 juin 2025

1 Théorème est définition

Un code de source, ou encodage, Γ est une application bijective $\Gamma : \mathcal{A} \to \mathcal{C}$. Soit Γ un code de source. Γ est à décodage unique si pur toute suite de symbols de code, qui résulte de l'encodage d'une suite de symbole de la source, il existe un décodage unique. Nous disons qu'un code est instantané

- s'il est à décodage unique
- et si à mesure que les séquences de symboles de l'alphabet du code sont reçus, les mots du code peuvent être déterminés sans s'inquiéter des symboles de code suivants.

On dit qu'un mot de code $c = x_1 x_2 \dots x_k$ est préfixe d'un autre mot de code c' si on peut écrire $c' = x_1 x_2 x_k x_{k+1} \dots x_l$ pour un certain $l \geq k+1$. Sur l'arbre du code, cela veut dire que c' est un descendant de c. On dit que le code Γ est sans préfixe si aucun mot de code n'est prefixe d'un autre mot de code. Sur l'arbre du code, cela veut dire que aucun mot de code n'est descendant d'un autre mot de code. Un code est sans préfixe si et seulement s'il est instantané

1.1 Kraft-MCMillan

Soit Γ un code D'aire dont les longueurs des M mots de code sont $l_1, \dots, l_M$. Si Γ est à décodage unique alors il satisfait l'inégalité de Kraft:

$$D^{-l_1} + \cdots + D^{-l_M} \leq 1$$

Réciproquement, si des nombres $l_1, \dots, l_M$ satisfont l'inégalité de Kraft, il existe un code D'aire instantané (donc à décodage unique) dont le dictionnaire possède M mots de code et dont les longueurs des mots de code sont $l_1, \dots, l_M$.

Pour tout code à décodage unique, il existe un code instantané sur les mêmes alphabets de source et de code qui a les mêmes longueurs de mot

1.2 Efficacité d'un code de source

Soit une source S d'alphabet $\mathcal{A}$ et de densité de probabilité p, et soit Γ un code D'aire de la source S. La longueur moyenne du code Γ est:

$$L(\Gamma) = \sum_{s \in \mathcal{A}} p(s)\, l(\Gamma(s))$$

L'unité est le symbole de code par symbole de source (si $D = 2$ on dit bits par symbole de source) Soit une source S d'entropie $H(S)$ et soit Γ un code D'aire de la source s. Si Γ est à décodage unique, sa longueur moyenne satisfait:

$$L(\Gamma) \geq \frac{H(S)}{\log_2(D)}$$

1.2.1 Question 1

Bon je sais pas pourquoi mais j'ai de la peine avec ça.
Donc on reprends, dans le meilleur cas on aurait que $q = 1,0$. dans ce cas la on a juste l'arbre de base avec deux possibilités, donc quelque soit les probabilité (qu'on connaît en l'occurrence) alors la on a que la longueur est de 1.
Dans le pire des cas on aurait que $P(S_2)$ est uniforme est que donc $q = \frac{1}{2}$. Dans ce cas la on liste les quatre possibilités:

$$\begin{aligned} p(ac) &= \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8} \\ p(ad) &= \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8} \\ p(bc) &= \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8} \\ p(bd) &= \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8} \end{aligned}$$

Donc maintenant on construire l'arbre de ça, on calcule la longueur moyenne:

$$\begin{aligned} L(S, \Gamma_H) &= 2 \cdot \frac{3}{8} + \frac{6}{8} + \frac{3}{8} \\ &= \frac{15}{8} \leq 2 \end{aligned}$$

Donc oui la première proposition est vraie.
La question maintenant est sur, est ce que $\mathit{length}(\Gamma_H(bc)) = 3 \forall q < \frac{1}{4}$?
Si on prends $q = \frac{1}{4}$ on a

$$\begin{aligned} p(ac) &= \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16} \\ p(ad) &= \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16} \\ p(bc) &= \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16} \\ p(bd) &= \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16} \end{aligned}$$

Donc ici le seul cas où bc pourrait avoir 3 branche serait celui ou il est plus petit que ad, or ce n'est pas le cas ici, il peut être avec 3 branches comme avec deux donc ce n'est pas forcémment vrai.

1.2.2 Question 2

Donc ici on doit faire un arbre de probablité (en tout cas ça fonctionne pour trouver la réponse). On aura donc un arbre qui a en premier lieu trois branche uniforme entre X_1, X_2, X_3 et ensuite quatre branches qui représenteront chaque cas de figures. On peut ensuite toute les additionner (leur probabilités) pour donner le résultat pour chaque cas de figure.

$$\begin{aligned} P(111) &= \frac{3}{12} \\ P(001) &= \frac{1}{12} \\ P(010) &= \frac{1}{12} \\ P(100) &= \frac{1}{12} \\ P(110) &= \frac{2}{12} \\ P(101) &= \frac{2}{12} \\ P(011) &= \frac{2}{12} \end{aligned}$$

Si on additionne tout on va bien que c'est égal à 1. Il reste plus qu'à calculer l'entropie:

$$\begin{aligned}
H\left(X_1, X_2, X_3\right) &= \frac{3}{12}\log\left(\frac{12}{3}\right) + \frac{3}{12}\log\left(\frac{12}{1}\right) + \frac{3\cdot 2}{12}\log\left(6\right) \\
&= \frac{3\cdot 2}{12} + \frac{3}{12}\left(2+\log\left(3\right)\right) + \frac{1}{2}\left(\log\left(2\right)+\log\left(3\right)\right) \\
&= 1 + \frac{\log\left(3\right)}{4} + \frac{1}{2} + \frac{\log\left(3\right)}{2} \\
&= \frac{3}{2} + \frac{3}{4}\log\left(3\right)
\end{aligned}$$

1.2.3 Question 3

Il me semble que c'est un théorème direct qui dit que n'importe quelle D-ary code qui est uniquely decodable a l'égalité

$$L\left(S, \Gamma_D\right) \geq H_D\left(S\right)$$

Donc ici il suffit de faire avec $D = 4$ et on voit que l'égalité n'est pas respecté et on peut le déviner pour la suite car il faudrait que les probablité soit toute égal au D comme ceci $p\left(\cdot\right) = \frac{1}{D}$. Maintenant si on fait avec $D = 3$, on a donc:

$$\begin{aligned}
H_D\left(S\right) &= 3\left(\frac{1}{9}\log_3 9\right) + 2\left(\frac{1}{3}\log_3\left(3\right)\right) \\
&= \frac{2}{3} + \frac{2}{3} = \frac{4}{3}
\end{aligned}$$

Maintenant si on calcule la longueur moyenne on a:

$$\begin{aligned}
L = 3\cdot 2\cdot\left(\frac{1}{3}\right) + 2\cdot 1\cdot\left(\frac{1}{3}\right) \\
= \frac{2}{3} + \frac{2}{3} = \frac{4}{3}
\end{aligned}$$