

Communications Numériques TP 1

Polytech RICM3 - Bernard Tourancheau - Nicolas Basset

Vous utiliserez votre propre PC et les outils de programmation de votre choix. Les groupes de TP sont fixés pour l'ensemble du cours et des TP. Le compte-rendu doit être rédigé en Latex, cf. le modèle proposé sur MOODLE. Une archive zip ou tar contenant le compte rendu sous format .tex et .pdf ainsi que les programmes que vous aurez écrits devra être déposée sur MOODLE. NB: overleaf.com peut être une bonne option qui intègre gestion de version, cloud, partage et Latex. Les recherches sur le web sont encouragées mais attention à citer vos sources. NB, Latex permet de gérer des références avec un système très complet autour de bibtex. ATTENTION ! Les dates de rendu du TP sont fixées sur Moodle. Si vous utilisez le langage C: inclure la librairie "math.h" dans le programme pour les fonctions mathématiques et option "-lm" pour le compilateur gcc.

NB: Des QCMs individuels viendront compléter l'évaluation du contrôle continu durant les prochaines séances de TP.

1 Codage et Entropie

Soit une source X comportant 6 événements : e_1, \dots, e_6 chacun ayant une probabilité d'apparition $P_{e_1} = 0.39, P_{e_2} = 0.20, P_{e_3} = 0.14, P_{e_4} = 0.11, P_{e_5} = 0.09, P_{e_6} = 0.07$.

1. Calculer l'entropie $H(X)$.
2. Quelle serait la longueur des mots nécessaire avec un codage à taille fixe ?
3. Que pouvez-vous conclure pour ce codage par rapport à l'optimum ?
4. Ecrire un programme qui calcule $H(X)$ pour X décrit dans un fichier contenant les probabilités de ses événements.
5. Vérifier le résultat obtenu en première question.
6. Tester votre programme sur $X' = \{P_{e_1} = 0.24, P_{e_2} = 0.09, P_{e_3} = 0.30, P_{e_4} = 0.19, P_{e_5} = 0.08, P_{e_6} = 0.06, P_{e_7} = 0.03, P_{e_8} = 0.01\}$

2 5 a day

Monsieur et Madame B. ont 5 enfants. Chaque enfant doit manger 5 rations (ou plus) de fruits ou légumes par jour et leurs parents doivent donc savoir à tout moment de la journée combien de rations ont été administrées à chacun. Un des parents tient les comptes jusqu'à 19h avant de passer le relais à l'autre parent.

1. Combien de bits d'information est-il nécessaire pour transmettre le décompte des fruits et légumes des cinq enfants ? (indice: la réponse doit se baser sur la notion d'entropie mais laquelle?)
2. En utilisant la conversion nombre en base 6 vers le binaire, proposer un codage à longueur fixe qui utilise le moins de bits possible. Expliquer aussi le décodage. Donner quelques exemples.
3. Mêmes questions (quantité d'information, codage/décodage, exemples) sachant que les parents s'arrangent pour qu'il n'y ait jamais plus d'une ration d'écart entre chaque enfant.

4. On suppose maintenant que les enfants sont nourrit de manière probabiliste leurs alimentations sont indépendantes et identiquement distribuées comme suit: il est sûr qu'à 19h, chaque enfant a déjà manger au moins 3 rations. Que chacun a 20% de chances de n'en avoir manger que 3 et 25 % de chances d'en avoir manger 5 ou plus. Quantifier à nouveau l'information à transmettre. De quel entropie s'agit-il et quelle est sa valeur? On fera un codage de Huffman après avoir traité l'exercice correspondant par la suite. En attendant, donner un codage à longueur fixe avec le nombre minimum de bits. Donner quelques exemples.

3 Code de Huffman

1. Donner l'arbre d'un code de Huffman pour le premier exemple X de la section 1.
2. Quel est le codage résultant?
3. Quelle est sa longueur moyenne?
4. Ce codage atteint-il la borne inférieure ?
5. Quel est le taux de compression par rapport à un code avec mots de taille fixe?
6. Ecrire un programme avec le langage de programmation de votre choix qui construit le code de Huffman pour des événements simples et produit aussi la longueur moyenne des messages obtenus. Vérifier les résultats obtenus en première question.
7. Utiliser votre code pour le 2e exemple de la section 1.
8. On veut se rapprocher de la borne inférieure. L'idée est d'avoir la même démarche mais à partir de la liste des événements possibles pris deux par deux. Quelle est le nombre d'événements de cette liste pour l'exemple X ?
9. Donner la liste de ces événements doubles.
10. Etendre votre programme pour traiter ces événements doubles et donner les résultats de longueur moyenne des messages obtenus pour X et X'.
11. Donner un code de Huffman pour la question 4 de la section 2. Donner la longueur moyenne du code.

4 Codage arithmétique

Le principe est de coder chaque symbole source s par un nombre, ou plus exactement un intervalle de $[0;1[$, correspondant à la table des probabilité d'apparition du symbole s : $P(s)$. Par exemple dans WIKI (cf. wikipédia) :

Lettre	Probabilité	Intervalle
W	1/4	$[0;0,25[$
I	2/4	$[0,25;0,75[$
K	1/4	$[0,75;1[$

Soit l'algorithme de compression en Algorithme 1 et l'algorithme de décompression en Algorithme 2.

1. Expliquer le principe des algorithmes 1 et 2.
2. Programmer l'algorithme de compression.
3. Quel est $V_{message}$ pour "WIKI" ?

Algorithm 1 Compression arithmétique

Require: Les symboles du message en entrée et n le nombre de symboles du message.

Ensure: $V_{message}$ est la valeur représentant le message.

```
 $B_{inf} \leftarrow 0$   
 $B_{sup} \leftarrow 1$   
 $V_{message} \leftarrow 1$   
 $V_{ecart} \leftarrow B_{sup} - B_{inf}$   
Lire(symbole)  
for  $i = 1 \dots n$  do  
   $B_{sup} \leftarrow B_{inf} + V_{ecart} * B_{sup}^{symbole}$   
   $B_{inf} \leftarrow B_{inf} + V_{ecart} * B_{inf}^{symbole}$   
   $V_{message} \leftarrow \text{valeur} \in [B_{inf}; B_{sup}]$  par exemple  $\frac{B_{sup} + B_{inf}}{2}$   
   $V_{ecart} \leftarrow B_{sup} - B_{inf}$   
  Lire(symbole)  
end for
```

Algorithm 2 Décompression arithmétique

Require: $V_{message}$ la valeur représentant le message et n le nombre de symboles du message. **Lettre** une fonction qui pour un réel V entre 0 et 1 associe la lettre correspondante à l'intervalle qui contient x .

Ensure: Les symboles du message en sortie.

```
symbole  $\leftarrow$  Lettre( $V_{message}$ )  
for  $i = 1 \dots n$  do  
   $V_{message} \leftarrow (V_{message} - B_{inf}^{symbole}) / \text{Probabilité}(\text{symbole})$   
  symbole  $\leftarrow$  Lettre( $V_{message}$ )  
end for
```

4. Quel est $V_{message}$ pour "KIWI" ?
5. Quel est $V_{message}$ pour "KIKIWIWI" ?
6. Programmer l'algorithme de décompression.
7. Quel est le message pour $V_{message} = 0.008$?
8. Quel est le message pour $V_{message} = 0.517$?
9. Quel est le message pour $V_{message} = 0.164$?
10. Quel est le message pour $V_{message} = 0.312$?
11. Que se passe-t-il si le n utilisé lors de la décompression est plus petit que la taille du message ?
12. Que se passe-t-il si le n utilisé lors de la décompression est plus grand que la taille du message ?

5 Compression d'un texte

1. On appelle fréquence d'un caractère c dans un texte (où probabilité empirique d'occurrence de c) le nombres d'occurrences de c dans le texte. Quels sont les fréquences des différents caractères (lettres ou espace) dans le texte suivant: "some people feel the rain. others just get wet"
2. Ecrire un programme qui étant donné un fichier texte sous format input.txt, calcule la fréquence des caractères du texte et les imprime dans un fichier output.txt.

3. Utiliser ce programme ainsi que le programme de l'exercice précédent sur le codage de Huffman pour compresser n'importe quel fichier texte, votre programme prend en entrée un fichier et écrit dans un fichier le texte compressé et dans un troisième fichier le code de Huffman utilisé pour la compression.
4. Ecrire un programme pour décompresser un fichier donné dans un second fichier décrivant un code préfixe (par exemple un code de Huffman des questions précédentes). Votre programme devra écrire le texte décompressé dans un troisième fichier.
5. Testez vos programmes sur des fichiers texte de différentes tailles.