École polytechnique — Majeure informatique
Promotion 2008

Cours *Conception et analyse d'algorithmes*
TD 6 – Algorithmes d'approximation, exploration
20 octobre 2010

1. Algorithme approché pour un recouvrement minimal

Le problème Recouvrement d'ensemble prend comme donnée un ensemble fini E, une famille $\mathcal{S} = \{S_i\}_{1\leq i\leq k}$ de sous-ensembles de E et pour chaque S_i un poids $w(S_i)$ qui est un entier positif. Il s'agit de déterminer un ensemble d'indices $J \subseteq \{1, \ldots, n\}$ tel que l'union des sous-ensembles $\cup_{j\in J} S_j$ soit E et que la somme des poids $\sum_{j\in J} w(S_j)$ soit minimale.

I. Recouvrement d'ensemble est $\mathcal{NP}$-complet.

En utilisant un paramètre supplémentaire W, donner une version de décision de ce problème. Démontrer que le problème posé est $\mathcal{NP}$-complet par réduction du problème recouvrement par sommets.

On rappelle que Recouvrement par sommets prend comme données un graphe $G = (V, E)$ et un entier k et répond à la question *existe-t-il un sous-ensemble de sommets $X \subseteq V$ de cardinal au plus k tel que pour toute arête $(u, v) \in E$, on a $u \in X$ ou $v \in X$.*

II. Algorithme approché glouton.

On considère l'algorithme glouton suivant qui calcule un recouvrement J. Ici et dans la suite, le cardinal d'un ensemble X est noté $|X|$ et $X \setminus Y$ désigne l'ensemble des éléments de X qui ne sont pas dans Y.

- Initialiser $F \leftarrow \emptyset$, $J \leftarrow \emptyset$.
- Tant que $F \neq E$, faire :
  - Calculer les valeur $c(S_i) = w(S_i)/|S_i \setminus F|$ pour tout $i = 1, \ldots, n$.
  - Trouver i tel que $c(S_i)$ soit minimal.
  - Mettre à jour $F \leftarrow F \cup S_i$, et $J \leftarrow J \cup \{i\}$.
- Retourner l'ensemble J.

Soit $(T_i)_{1\leq i\leq k}$ les ensembles du recouvrement ordonnés dans l'ordre dans lequel ils sont ajoutés par l'algorithme glouton.

1. Soit W_0 la somme des poids des sous-ensembles S_i qui figurent dans la solution optimale.
Montrer qu'au moins un de ces sous ensembles S_i satisfait $\frac{w(S_i)}{|S_i|} \leq \frac{W_0}{n}$. En déduire que $\frac{w(T_1)}{|T_1|} \leq \frac{W_0}{n}$

Soit $x_1, \ldots, x_n$ les éléments de E dans l'ordre dans lequel ils sont ajoutés dans F.

2. Montrer que le sous-ensemble T_i contenant x_j à l'étape où il entre dans F satisfait lors de cette étape : $c(T_i) \leq W_0/(n - j + 1)$.

3. En déduire que l'algorithme glouton est approché à un facteur H_n, où $H_n = 1 + 1/2 + \ldots + 1/n$.

4. Soit les $n + 1$ sous-ensembles suivants de $E = \{1, 2, \ldots, n\}$, $S_i = \{i\}$ pour $1 \leq i \leq n$ et $S_{n+1} = E$ et soit $\varepsilon > 0$. Donner des poids à ces sous-ensembles pour que le résultat de l'algorithme glouton diffère de l'optimum d'un facteur $H_n/(1 + \varepsilon)$.

III. Recouvrement de mots.

Soit Σ un alphabet fini. Un mot est une suite finie de lettres de Σ. La longueur d'un mot f est notée $\ell(f)$. On dit qu'un mot g est *facteur* du mot f si il existe des mots f' et f'' tels que $f = f'gf''$.

On s'intéresse dans la suite au problème Recouvrement de mots qui consiste, étant donné un ensemble $S = \{f_1, \ldots, f_n\}$ de mots, à déterminer le plus court mot f contenant tous les f_i comme facteurs. On se propose d'appliquer les résultats qui précèdent à ce problème.

Deux mots f_i et f_j tels que les k dernières lettres de f_i forment un mot égal à celui constitué par les k premières de f_j sont dits *k-compatibles*. On note dans ce cas σ_{ijk} le mot de longueur $n_i + n_j - k$ obtenu par concaténation de f_i et des $n_j - k$ derniers caractères de f_j (n_i et n_j sont les longueurs de f_i et f_j).

1. Soit $S = \{f_1 = ab^k, f_2 = b^kc, f_3 = b^{k+1}\}$. Que vaut $\sigma_{1,2,k}$? Quel est le f optimal ?

2. On considère l'algorithme glouton qui à chaque étape, détermine deux mots f_i et f_j de S qui sont k-compatibles pour un k maximal, et qui remplace alors f_i et f_j dans S par $\sigma_{i,j,k}$.

L'algorithme termine quand S est réduit à un seul mot. Montrer sur un exemple que cet algorithme est au mieux (1/2)-approché.

On associe à un problème P de Recouvrement de mots, donné par $S = \{f_1, \ldots, f_n\}$, un problème Q de Recouvrement d'ensemble. Pour Q on pose $E = S$ et on construit des sous-ensembles S_u pour tous les mots u obtenus de la façon suivante : u est ou bien dans S ou bien égal à σ_{ijk} pour un couple (f_i, f_j) k-compatible ; S_u est alors l'ensemble des facteurs de u qui sont dans S, il a pour poids $w(S_u) = \ell(u)$ la longueur de u.

3. Soit $S_{\theta_1}, \ldots, S_{\theta_p}$ les ensembles donnant une solution optimale au problème Q, et W_0 le poids de cette solution. Donner une solution de longueur W_0 au problème P.

Soit f la solution optimale du problème P, dans lequel on suppose que deux mots de S ne sont pas facteurs l'un de l'autre. On renumérote les f_i de sorte que f_i apparaît pour la première fois avant f_j dans f si et seulement si $i < j$. *Par abus de notation, f_i désignera uniquement sa première occurrence dans f.*

On considère la suite d'indice $1 = b_1 < e_1 < b_2 < e_2 < \cdots < b_t < e_t = n$ telle que pour tout $i = 1, \ldots, t$, $b_i = e_{i-1} + 1$ et e_i est le plus grand indice d'un mot de S qui a une intersection avec f_{b_i}.

4. On note k_i la longueur de l'intersection entre f_{b_i} et f_{e_i}, et $\pi_i = \sigma_{b_i e_i k_i}$. Montrer que $S_{\pi_1}, \ldots, S_{\pi_t}$ est une solution au problème Q.

5. Montrer que π_i et π_{i+2} sont des facteurs de f qui ne s'intersectent pas dans f. En déduire que $\sum_i \ell(\pi_i) \leq 2\ell(f)$.

6. En déduire un algorithme approché à un facteur $2H_n$ qui trouve un mot f contenant tous les f_i comme facteurs.

2. Réseau avec relais, NP-complétude, approximation

Soit un graphe $G = (X, E)$ et un sous-ensemble S de sommets. On cherche à construire dans G un réseau le plus petit possible qui connecte les sommets de S entre eux, en utilisant éventuellement certains autres sommets de G comme relais. Plus formellement on considère le problème suivant :

Réseau avec relais

Donnée : Un graphe $G = (X, E)$, un sous-ensemble $S \subset X$ de sommets et un entier k.

Problème : Trouver un sous-ensemble T d'au plus k arêtes tel que pour tout x, y dans S il existe un chemin de x à y n'utilisant que des arêtes de T.

A. Un algorithme d'approximation pour Réseau avec relais.

Soit G_S le graphe complet pondéré sur S avec, pour tout x, y dans S, le poids de l'arête (x, y) égal à la longueur d'un plus court chemin entre x et y dans le graphe G.

1. Montrer que le poids d'un arbre couvrant de poids minimum de G_S est au plus deux fois la taille d'un réseau de taille minimale connectant S dans G.
2. Donner un algorithme approché à un facteur 2 pour le problème Réseau avec relais. Quelle est la complexité de votre algorithme ?

B. NP-complétude du problème Réseau avec relais.

Nous allons montrer que le problème réseau avec relais est NP-complet. Pour cela nous considérons une variante du problème de recouvrement d'ensembles vu dans l'exercice précédent :

Couverture par triplets

Donnée : Un ensemble Y de cardinal $3n$ et une famille $C = \{C_1, \ldots, C_m\}$ de triplets non ordonnés d'éléments de Y : pour tout i, $C_i \subset Y$ et $|C_i| = 3$.

Problème : Existe-t-il un sous-ensemble C' de C qui forme une partition de Y : les éléments de C' sont disjoints et leur union est Y ?

Nous admettrons que la variante Couverture par triplets est encore NP-complète.

1. Montrer que le problème Réseau avec relais est dans la classe NP.
2. Montrer que le problème Réseau avec relais est NP-complet.