TD : Plus longue séquence croissante

Dominique Michelucci, Université de Dijon

7 décembre 2012

Un tableau non trié d'entiers $E[0], \ldots E[n-1]$ est donné. Le problème est de calculer la longueur de la séquence croissante la plus longue. Note : dans cette séquence, tout élément (sauf le dernier) est inférieur ou égal à son élément suivant. Par exemple, si $E = [0; 300; 100; 200; 1000; 400; 500; 1100; 900; 800; 600; 700; -100]$, alors les séquences croissantes les plus longues ont 7 éléments. L'une d'elles est $[0; 100; 200; 400; 500; 600; 700]$.

Proposez une méthode en temps polynomial ($O(n^2)$). Par exemple, définir récursivement $LT[i]$, comme étant la longueur de la séquence croissante la plus longue qui se termine (et utilise) E_i. $LT[0] = 1$. Définissez $LT[i]$ en fonction de $LT[0], \ldots LT[i-1]$. Exemple :

i	0	1	2	3	4	5	6	7	8	9	10	11	12
E_i	0	300	100	200	1000	400	500	1100	900	800	600	700	-100
LT_i	1	2	2	3	4	4	5	6	6	6	6	7	1

Cette méthode est en temps $O(n^2)$. Donnez une méthode en $O(n \log n)$. Piste : stockez dans un tableau $V[l]$ la dernière valeur de la séquence de longueur l. Quand vous cherchez quelle est la plus longue séquence croissante que peut prolonger E_i, vous pouvez procéder par dichotomie dans le tableau V. Il faut aussi gérer L, la plus grande longueur courante des séquences croissantes. N'oubliez pas de mettre à jour le tableau V. Exemple :

i	0	1	2	3	4	5	6	7	8	9	10	11	12
E_i	0	300	100	200	1000	400	500	1100	900	800	600	700	-100
LT_i	1	2	2	3	4	4	5	6	6	6	6	7	1
V_i	$-$	-100	100	200	400	500	600	700	$-$	$-$	$-$	$-$	$-$

La suite suppose pour simplifier que tous les éléments E_i sont distincts.

Q : Soit S_k la sous séquence formée par les E_i dont la LT (longueur terminale) vaut k. Que constatez-vous ? Prouvez le.

Q. Soit E_i le premier (i minimum) élément tel que $LT[i] = l$, avec $l > 1$. Prouver que E_i prolonge E_k, où k est l'élément le plus à gauche (k maximum, $k < i$) tel que $LT[k] = l - 1$; autrement dit, il est inutile de vérifier que $E[k] < E[i]$: il l'est.

Q : Quelle est la longueur de la séquence commune la plus longue entre une séquence croissante et une séquence décroissante ? (Tous les E_i sont distincts)

Q : Soit L la longueur de la séquence croissante la plus longue.

Q : quelle application de cette partition minimale en séquences décroissantes ? En déduire que toute partition en séquences décroissantes a au minimum L séquences décroissantes.

Q : En déduire une méthode pour partitionner une séquence en un nombre minimum de séquences décroissantes.

Q : Soit S_k la sous séquence formée par les E_i dont la LT (longueur terminale) vaut k. Que constatez-vous ? Prouvez le.

R : Elles sont décroissantes. Preuve triviale.

Q. Soit E_i le premier (i minimum) élément tel que $LT[i] = l$, avec $l > 1$. Prouver que E_i prolonge E_k, où k est l'élément le plus à gauche (k maximum, $k < i$) tel que $LT[k] = l - 1$; autrement dit, il est inutile de vérifier que $E[k] < E[i]$: il l'est.

R : trivial, cf Q précédente.

Q : Quelle est la longueur de la séquence commune la plus longue entre une séquence croissante et une séquence décroissante ? (Tous les E_i sont distincts)

R : la séquence commune la plus longue a soit 0 soit 1 seul élément en commun. Si elle en a deux (a, b dans cet ordre, ou plus), alors $a < b$ et $a > b$, contradiction.

Q : soit L la longueur de la séquence croissante la plus longue. En déduire que toute partition en séquences décroissantes a au minimum K séquences décroissantes.

R : soit K la séquence croissante la plus longue, de longueur L. Soient $D_1, D_2, \dots D_d$ une partition en séquences décroissantes (il y a donc d séquences décroissantes dans cette partition). Tout élément E_i de K se trouve dans exactement une seule des D_1 ou D_2 ou $\dots D_d$. Mais chaque D_j ne peut contenir qu'un seul des éléments de K. Donc $d \geq L$.

R : Soit $D_1, D_2, \dots D_d$ une partition en séquences décroissantes. On a vu que $d \geq L$. Il manque une preuve d'existence pour que le plus petit des d soit vraiment égal à L. Mais c'est donné par une question précédente. Donc : la longueur de la séquence croissante la plus longue est le nombre minimum de séquences décroissantes des partitions en séquence décroissantes.

Q : En déduire une méthode pour partitionner une séquences en un ensemble de séquences décroissantes.

R : calculez la LT_i maximum : L par la méthode en $O(n \log n)$. Soient $S_1, \dots S_L$ (S_k la sous séquence formée par les E_i dont la LT vaut k) : elles partitionnent la séquence en L séquences décroissantes.

Q : quelle application de cette partition minimale en séquences décroissantes ?

R : le tri par monotonie, pour des ensembles gigantesques qui ne rentrent pas en mémoire centrale ; ils sont d'abord partitionnés en séquences décroissantes ; chacune est sauvegardé sur disque ; ensuite ces fichiers triés sont fusionnés : c'est comme le tri par fusion mais il peut y avoir bien plus que 2 listes (ou fichiers) à fusionner ; il faut parcourir séquentiellement ces fichiers ; un tas (heap) est utilisé pour stocker les "têtes" de ces fichiers ; quand la tête du fichier F_k est traitée, il faut avancer la tête de lecture dans le fichier F_k et insérer (ou remplacer) dans le tas. On espère que le tas rentre en mémoire centrale. Pour un ensemble aléatoire de n éléments, que vaut L ? Question mathématiquement non triviale, mais il est facile de mesurer. Sujet de TP possible...