

IRESI : Algorithme Misra-Gries

Pour le 3 décembre 2013

Table des matières

1	Principe de l'algorithme	1
1.1	Objectif	1
1.2	Fonctionnement	1
1.3	Conditions de validité	2
2	Implémentation OCaml	2
3	Résultats	2

1 Principe de l'algorithme

1.1 Objectif

Cet algorithme a pour but de déterminer quels sont les éléments les plus fréquents d'un flux de données en temps réel, c'est à dire en temps linéaire par rapport à m , le nombre d'éléments du flux. Pour ce faire il n'utilise que k compteurs, il est donc constant en complexité spatiale par rapport à k et, puisque dans les faits $k \ll m$, l'algorithme est économe en mémoire.

1.2 Fonctionnement

Le fonctionnement de l'algorithme est le suivant :

1. On commence par initialiser k compteurs vides.
2. Pour chaque élément x du flux :
 - Si il y a déjà, parmi les k compteurs, un compteur associé à x alors on incrémente ce compteur.
 - Sinon, si au moins un des k compteurs est vide, on associe x à l'un des compteurs vides et on positionne ce dernier à 1 pour compter l'élément courant du flux.
 - Enfin, si les k compteurs sont déjà associés à des éléments du flux tous différents de x , on décrémente tous les compteurs. Un compteur devenant nul n'est plus associé à aucun élément du flux, il redevient un compteur vide.

3. À la fin de la mesure, les compteurs contiennent nécessairement les valeurs les plus fréquentes du flux, sous certaines conditions.

1.3 Conditions de validité

Tout élément j du flux dont la fréquence a été strictement plus grande que m/k , où m est le nombre d'éléments du flux et k le nombre de compteurs utilisés, sera nécessairement présent dans l'un des compteurs à la fin de la mesure. Pour les éléments du flux dont la fréquence a été inférieure à m/k , on ne peut rien dire avec certitude mais plus leur fréquence était grande plus ils ont de chance de se trouver dans un compteur à la fin si le flux n'était pas pathologique.

2 Implémentation OCaml

3 Résultats