
Sujet n°5 : *Étude d'une méthode de sélection d'attributs en vue de son utilisation sur des données de puces ADN*

Résumé

Le but de ce projet est d'étudier les propriétés d'un algorithme de sélection d'attributs.

L'analyse du transcriptome (données issues de puces à ADN) repose cruciallement sur des méthodes de sélection d'attributs. Pourtant celles-ci sont généralement utilisées sans grande justification ni précaution.

Dans une publication récente, nous (groupe de travail à Orsay) avons proposé l'utilisation d'une nouvelle méthode de sélection d'attributs que nous pensons bien adaptée à l'analyse des données du transcriptome. Il reste cependant à préciser les propriétés de cette nouvelle méthode.

Nous proposons donc un travail en deux temps

1. Une première étape du projet consistera à appliquer cette méthode (programmée en Matlab) sur des bases de données soumises sur un site américain pour un challenge sur l'analyse du transcriptome.
2. La deuxième étape consistera à analyser les propriétés de cet algorithme de sélection d'attributs sur des données synthétiques afin en particulier d'étudier
 - i. La capacité à détecter les attributs pertinents (les gènes actifs) même lorsque le concept cible est de type XOR généralisé
 - ii. L'effet de «Boosting» sur les attributs faiblement pertinents.

Si ces propriétés se confirmaient, cela constituerait une indication forte de la bonne adéquation de cette méthode aux besoins des biologistes.

On rendra un rapport final rendant compte de la démarche suivie, des expériences réalisées et des leçons à en tirer. Ce rapport fera de 10 à 20 pages.

Une démonstration sur machine aura lieu devant l'enseignant chargé de l'encadrement ainsi qu'une soutenance orale de 15' pour présenter le travail réalisé devant les autres étudiants.