# FaBR-CL : méthode de classification croisée de protéines

Walid Erray*, Faouzi Mhamdi**

*Laboratoire ERIC, Université Lumière Lyon 2
69500 Bron France
walid.erray@eric.univ-lyon2.fr,
http://eric.univ-lyon2.fr
**URPAH, faculté des Sciences de Tunis,Université d'El Manar
1060 Tunis Tunisie
faouzi.mhamdi@ensi.rnu.tn,
http://www.mes.tn/fst/index.html

**Résumé.** Dans cet article, nous proposons une méthode de classification croisée permettant de classer des protéines, d'une part, et de classer des descripteurs (3-grammes) selon leurs pertinences par rapport aux groupes de protéines obtenus, d'autres part.

## 1 Classification croisée de données biologiques

Afin d'étudier les séquences d'acides aminés représentant les protéines, nous avons utilisé des techniques de text mining afin d'extraire des descripteurs. Ces descripteurs nous permettrons de construire un tableau de données Protéines $\times$ Descripteurs. L'une des techniques les plus utilisées est l'extraction des x-grammes (Miller et al. (1999), Mhamdi et al. (2004)), x étant la taille d'un descripteur.

Plusieurs méthodes de classification croisée ont été proposées (Govaert (1977), Ritschard et Nicoloyannis (2000)). Récemment, des méthodes de classification croisée ont été appliquées aux données biologiques (Cheng et Church (2000)). Cependant, plusieurs de ces méthodes restent très coûteuses en temps de calcul.

## 2 FaBR-CL : méthode de classification croisée

Afin d'effectuer une classification croisée, nous nous sommes basé sur une méthode de classification peu coûteuse en temps de calcul (Erray (2005)). La méthode proposée, FaBR-CL, utilise FaUR dans une approche "Combinaison itérative de regroupement des lignes et des colonnes" afin d'obtenir un regroupement complet des protéines et des 3-grammes. Ainsi, nous effectuons le regroupement des protéines, dans un premier temps, et le regroupement des 3-grammes dans un deuxième temps. La complexité de cette méthodes est en O(l log l + p log p), l étant le nombre de protéines et p le nombre de descripteurs.