

Une approche de classification non supervisée basée sur la détection de singularités et la corrélation de séries temporelles pour la recherche d'états: application à un bioprocédé fed-batch

Sébastien Régis

G.R.I.M.A.A.G.

Université Antilles-Guyane

Campus de Fouillole 97159 Pointe-à-Pitre Guadeloupe, France

sregis@univ-ag.fr

<http://grimaag.univ-ag.fr/>

Résumé. Nous proposons dans cet article une méthode de clustering qui combine l'analyse dynamique et l'analyse statistique pour caractériser des états. Il s'agit d'une méthode de fouille de données qui travaille sur des ensembles de séries temporelles pour détecter des états; ces états représentent les informations les plus significatives du système. L'objectif de cette méthode non supervisée est d'extraire de la connaissance à partir de l'analyse des séries temporelles multiples. Elle s'appuie sur la détection de singularités dans les séries temporelles et sur l'analyse des corrélations des séries entre les intervalles définis par ces singularités. Pour l'application présentée, les séries temporelles sont des signaux biochimiques mesurés durant un bioprocédé. Cette approche est donc utilisée pour confirmer et enrichir la connaissance des experts du domaine des bioprocédés sans utiliser la connaissance *a priori* de ces experts. Elle est appliquée à la recherche d'états physiologiques dans un bioprocédé de type fed-batch.

1 Introduction

L'extraction d'informations à partir de séries temporelles est un domaine de plus en plus important dans la fouille de données. En effet de nombreuses applications utilisant des signaux réels nécessitent l'utilisation des outils de ce champ d'étude. Ainsi, la fouille de données dans des ensembles de séries temporelles utilise divers outils pour extraire des informations intrinsèques ou d'interdépendance entre ces séries temporelles : recherche de similarités entre séries Marteau et al. (2006), corrélation entre séries Pelletier (2005), ou classification Nunez et al. (2002) par exemple. Les domaines d'application pour ces outils sont très variés : finance Takada et Bass (1998), données météorologiques Harms et Deogun (2004), données médicales Summa et al. (2006), biotechnologies, etc.

Par ailleurs les biotechnologies et plus particulièrement les bioprocédés offrent aujourd'hui des défis importants concernant l'extraction et la gestion des connaissances. L'analyse des états physiologiques survenant durant ces bioprocédés est un point essentiel pour leur contrôle et