

Détection et caractérisation d'un bioprocédé fed-batch par l'analyse de l'exposant de Hölder

Sbastien RÉGIS¹, Andrei DONCESCU², Lætitia FAURE³, Jean-Louis URIBELARREA³

¹Groupe de Recherche en Informatique et Mathématiques Appliquées des Antilles et de la Guyane
Université Antilles-Guyane 97159 Pointe-à-Pitre Guadeloupe France

²Laboratoire d'Analyse et d'Architecture des Systèmes CNRS
7, Av. du Colonel Roche 31077 Cedex 04 Toulouse France

³Laboratoire de Biotechnologies Bioprocédés GBA-INSA
Toulouse 31077 Cedex 04 Toulouse France
sregis@univ-ag.fr, adoncesc@laas.fr

Résumé – Nous cherchons une méthode non supervisée qui puisse détecter des états physiologiques dans les bioprocédés pour confirmer et enrichir la connaissance d'expert. Pour cela nous proposons une méthode de clustering basée sur la détection et la sélection de singularités par l'exposant de Hölder à partir des coefficients de la Transformée en Ondelettes.

Abstract – We search an unsupervised method for the detection of physiological states of bioprocesses in order to confirm and increase knowledge. We propose a method of clustering based on the detection and selection of singularities by Hölder exponent by using wavelets transform.

1 Introduction

Aujourd'hui les systèmes biologiques sont au centre de la recherche appliquée et fondamentale. L'explication à ce grand intérêt pour la biologie systémique vient de la nature de l'information véhiculée par ce type de système. D'abord les mécanismes de transport et de production qui se réalisent à l'intérieur de la cellule et qui s'organisent dans un système complexe de réactions biologiques peuvent être décrites par un système d'équations dynamiques non-linéaires. La résolution d'un système non-linéaire est une tâche complexe que nous voulons éviter. Pour caractériser un système biologique nous avons choisi l'analyse des états physiologiques. L'analyse des états survenant durant ces bioprocédés est un point essentiel pour le contrôle et l'optimisation de ces bioprocédés. Des méthodes basées sur des modèles existent, mais ces techniques basées sur des simulations peuvent conduire à des conclusions erronées car ces modèles ne tiennent pas compte de tous les paramètres en jeu et ne peuvent prévoir certaines situations inattendues. C'est pourquoi les méthodes qui ne sont pas basées sur des modèles et qui s'appuient sur l'analyse de signaux biochimiques mesurés durant l'expérience sont de plus en plus utilisées. La détection et la caractérisation des états physiologiques survenant durant le bioprocédé sont basées sur le traitement des signaux biochimiques et sur leur analyse statistique en utilisant des méthodes de classification. Dans ce papier nous proposons une méthode utilisant le maximum du module de la transformée en ondelettes, l'exposant (ou coefficient) de Hölder et le produit de corrélation pour détecter et caractériser les états d'un bioprocédé de type fed-batch. Dans la deuxième partie, nous présentons quelques-uns des travaux existants sur la détection et la caractérisation des états d'un bioprocédé. Dans la troisième par-

tie, la méthode du maximum du module de la transformée en ondelettes et l'exposant de Hölder qui sont utilisés pour détecter et sélectionner les singularités les plus significatives, sont présentés succinctement. La corrélation et la méthode de clustering qui sont utilisées pour caractériser les états, sont présentées dans la quatrième partie. La cinquième partie est consacrée aux résultats expérimentaux et enfin la dernière partie présente la conclusion et les perspectives.

2 Travaux existants

Les plupart des travaux sur la détection des états physiologiques ont montré que les singularités des signaux biochimiques mesurés durant un bioprocédé (quelque soit son type) correspondent au début et à la fin d'un état physiologique du système. Ainsi plusieurs auteurs utilisant des méthodes très différentes les unes des autres sont arrivés à la conclusion que ces singularités représentaient les limites des états physiologiques: c'est le cas de Steyer et al. [1] (en utilisant la logique floue et un système expert), Bakshi and Stephanopoulos [2] (en utilisant les ondelettes et un système expert) et Doncescu et al. [3] (en utilisant la logique inductive). La détection des états physiologiques peut être suivie d'une phase de caractérisation automatique des ces états, souvent à partir des propriétés statistiques des signaux mesurés. Une des particularités des bioprocédés de type fed-batch est qu'un état peut apparaître plusieurs fois et à des moments différents de l'expérience. Il est donc nécessaire de caractériser ces états pour savoir s'ils réapparaissent au cours du temps. Plusieurs méthodes statistiques ont été proposées pour caractériser ces états. Par exemple, des méthodes de classification basées sur l'Analyse en Composantes

Principales (ACP) [4], l'ACP adaptative [5], ou la kernel ACP [6] permettent de distinguer et de caractériser les différents états d'un procédé.

La méthode du maximum du module de la transformée en ondelettes [7][8] est utilisée dans cet article pour détecter les singularités des signaux (afin de définir les limites d'un état) et l'étude des coefficients de corrélation entre ces signaux permet de caractériser ces différents états.

3 Détection et sélection des singularités par les ondelettes et l'exposant de Hölder

Les singularités des signaux biochimiques correspondent aux limites d'un état. Ces signaux biochimiques sont des signaux non stationnaires et non symétriques. Plusieurs méthodes utilisent les ondelettes pour détecter les singularités de ces signaux afin de détecter les états: par exemple, Bakshi and Stephanopoulos [2] et plus récemment Jiang et al. [9]. En effet ces singularités correspondent à des maxima des coefficients de la transformée en ondelettes des signaux biochimiques. Jiang et al. [9] proposent de sélectionner les maxima en utilisant un seuil. Cependant le choix de ce seuil reste empirique. Nous proposons d'utiliser l'exposant de Hölder pour sélectionner ces maxima. En effet la valeur de cet exposant de Hölder change en fonction de la nature de la singularité (voir [7][8]). Ainsi il est possible de sélectionner les singularités les plus significatives partir de leur exposant de Hölder. Récemment une méthode utilisant les algorithmes génétiques a été proposée et semble fournir des résultats plus précis que les méthodes classiques [10].

4 Caractérisation des états par corrélation et classification

Après avoir déterminé les limites des états par la détection des singularités, ces états physiologiques sont caractérisés par l'analyse des corrélations entre les signaux biochimiques. Sur chaque intervalle temporel défini à partir des singularités, le coefficient de corrélation est calculé entre les signaux deux à deux. Ce coefficient de corrélation (aussi appelé coefficient de Bravais-Pearson voir [11]) est donné par l'équation suivante:

$$\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \quad (1)$$

où x_i représente les valeurs du premier paramètre biochimique (sur un intervalle temporel donné), y_i les valeurs du deuxième paramètre (sur le même intervalle temporel), n le nombre d'éléments, \bar{x} la valeur moyenne des éléments x , \bar{y} la valeur moyenne des éléments y , et σ_x et σ_y les écarts type pour chacun des deux paramètres.

Ce coefficient de corrélation est en fait équivalent au cosinus du produit scalaire de deux variables projetées sur le cercle de corrélation pour une Analyse en Composantes Principales (ACP) effectuée entre ces deux variables.

Sur chaque intervalle on garde le signe des coefficients de corrélation entre deux signaux. Chaque intervalle est ainsi caractérisé par un ensemble de signes positifs et négatifs. Les intervalles

ayant la même série de signes sont regroupés dans la même classe. Ruiz et al. [4] proposent également une méthode basée sur l'ACP également pour une application concernant le traitement des eaux usées: la méthode consiste à classer les données projetées préalablement dans l'espace défini par les deux premières composantes principales. Cette méthode réduit la dimension de l'espace analysé mais l'ACP ne tient pas compte du temps: l'évolution des signaux n'est pas prise en compte. Pour pallier ce problème, Ruiz et al. proposent d'utiliser une fenêtre d'analyse de taille fixe contenant des données consécutives dans le temps. Cependant du fait que la taille de la fenêtre d'analyse est fixe, cette méthode ne tient pas réellement compte des changements survenant dans le procédé. Ainsi, la méthode basée sur la segmentation temporelle à partir de l'exposant de Hölder des singularités, semble mieux adaptée si l'on veut tenir compte de la dynamique du système.

5 Résultats expérimentaux

Les tests ont été effectués sur un bioprocédé fermentaire de type fed-batch. Ce bioprocédé utilisant des micro-organismes (levures) appelés *Saccharomyces Cerevisiae* a duré environ 34 heures. 11 signaux biochimiques ont été utilisés pour la classification. L'échelle maximum utilisée a été 2^{10} [7]. La méthode a permis de détecter et de caractériser une action externe réalisée durant l'expérience. La classification est composée au total de 24 classes mais c'est surtout la classe numéro 8 qui est intéressante, caractérisée par un coefficient d'Hölder de 0.4. En effet cette classe 8 correspond à l'ajout d'un acide dans le milieu. Toutes les apparitions de la classe 8 correspondent exactement à l'ajout de cet acide. Autant que nous le sachions, c'est la première qu'une méthode qui n'est pas basée sur un modèle permet de trouver automatiquement l'addition d'un acide dans un bioprocédé fed-batch. Les résultats sont prometteurs et une analyse biologique approfondie est donc nécessaire.

6 Conclusion et perspectives

Nous avons proposé une méthode de classification non supervisée basée sur les ondelettes, l'exposant de Hölder et le coefficient de corrélation pour la détection et la caractérisation des états dans un bioprocédé. La détection des états est basée sur la détection et la sélection de singularités des signaux biochimiques en utilisant le maximum du module de la transformée en ondelettes et l'évaluation de l'exposant de Hölder. La caractérisation des états s'appuie sur le coefficient de corrélation entre les signaux. Les prochains travaux concernent l'utilisation d'autres ondelettes. L'utilisation des valeurs du coefficient de corrélation (au lieu du signe de ce coefficient) pour caractériser les états est une voie à explorer. La prochaine étape est la mise en ligne de la méthode pour une utilisation en temps réel.

Références

- [1] J.P. Steyer, J.B. Pourciel, D. Simoes, and J.L. Uribelarrea, "Qualitative knowledge modeling used in a real time expert system for biotechnological process control," in

- [2] B.R. Bakshi and G. Stephanopoulos, "Representation of process trends-III. multiscale extraction of trends from process data," *Computer and Chemical Engineering*, vol. 18, no. 4, pp. 267–302, 1994.
- [3] A. Doncescu, J. Waissman, G. Richard, and G. Roux, "Characterization of bio-chemical signals by inductive logic programming," *Knowledge-Based Systems*, vol. 15, no. 1-2, pp. 129–137, Janvier 2002.
- [4] G. Ruiz, M. Castellano, W. González, E. Roca, and J.M. Lema, "Algorithm for steady states detection of multivariate process: application to wastewater anaerobic digestion process," in *AutMoNet 2004*, 2004, pp. 181–188.
- [5] J. Lennox and C. Rosen, "Adaptative multiscale principal components analysis for online monitoring of wastewater treatment," *Water Science and Technology*, vol. 45, no. 4-5, pp. 227–235, 2002.
- [6] J.-M. Lee, C. Yoo, I.-B. Lee, and P. Vanrolleghem, "Multivariate statistical monitoring of nonlinear biological processes using kernel PCA," in *IFAC CAB'9*, Nancy, France, March 2004.
- [7] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. on PAMI*, vol. 14, no. 7, pp. 710–732, 1992.
- [8] S. Mallat and W.-L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp. 617–643, March 1992.
- [9] T. Jiang, B. Chen, X. He, and P. Stuart, "Application of steady-state detection method based on wavelet transform," *Computer and Chemical Engineering*, vol. 27, no. 4, pp. 569–578, 2003.
- [10] L. Manyri, S. Regis, A. Doncescu, J. Desachy, and JL Urribelarea, "Holder coefficient estimation by differential evolutionary algorithms for *saccharomyces cervisiae* physiological states characterisation," in *ICPP-HPSECA*, Kaohsiung, Taiwan, Octobre 2003.
- [11] G. Saporta, *Probabilités, et Analyse des données et Statistique*, Technip, 1990.