Engineering Applications of Artificial Intelligence 28 (2014) 165–173

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Bioprocess diagnosis based on the empirical use of distance measures in the theory of belief functions

Sébastien Régis [a,*], Andrei Doncescu [b], Nathalie Goma [c]

[a] LAMIA, Department of Computer Science and Mathematics, French West Indies and Guyana University, 97159 Pointe-à-Pitre, Guadeloupe, France
[b] LAAS-CNRS 7, Robotics and Interactions Group, Avenue du Colonel Roche, 31077 Toulouse Cedex 4, France
[c] IPBS-CNRS, Route du Narbone Toulouse Cedex 4, France

ARTICLE INFO

Article history:
Received 18 December 2012
Received in revised form
13 October 2013
Accepted 18 October 2013
Available online 6 December 2013

Keywords:
Theory of belief functions
Conflict
Distance measures
Relevance of source
Diagnosis
Bioprocess

ABSTRACT

Microorganisms plays a central role in the production of a wide range of industrial chemicals, enzymes and antibiotics. The rate of product formation in a given industrial process is directly related to the rate of biomass formation which is influenced directly or indirectly by a whole host of different environmental factors. In this paper we propose to use distance measures between basic belief assignment in the context of the belief functions theory, in order to diagnosis the relevance of bioprocess sensors and actors which measure the environmental factors.

1. Introduction

The establishment of reliable processes with increased efficiency and cost reduction is of primary importance in the fermentation industry. Nowadays, the increase of the number of sensors in biotechnology field provides an important amount of heterogenous data. This data with various natures and various levels of precision require the use of automatic method of data fusion. The goal of data fusion is to diagnosis here the correctness of the bioprocess. The role of fermentation diagnosis is to detect the anomalies able to modify the optimal conditions maximizing out of product, in our case biomass. Anomalies can occur in biochemical parameters such as temperature, aeration, pH and dissolved oxygen. Monitoring and diagnosis of bioprocesses has been tackled in three ways:

1. Adaptive state estimators: These approaches are designed to adapt to the time varying characteristics of the process, for example to the changes in growth and metabolite expression rates when the nutrient levels are depleted. The measurements are used in conjunction with the process model and need to be carefully tuned to achieve accurate reconstruction of fermentation parameters and states. There are two mains reasons. First, because the dynamics of biological processes are both non-linear and non-stationary; secondly, because classical methods have proved inadequate in describing the overall behavior of biological process.
2. Artificial intelligence based algorithms: The second approach to the development of monitoring and fault diagnosis strategies is based on artificial intelligence based algorithms. Such approaches rely on the construction of a qualitative and quantitative database of regular and faulty modes of plant operation. The event-tracking is classified as "normal" or "faulty" using a supervised classifier. Heuristics based expert systems rely on capturing knowledge and know-how) on the growth conditions of microorganism, and take into account the different fault occurrences and process variable interactions that the plant personnel can envision, by capturing them a rule base (Steyer, 1991; Doncescu et al., 2002).
3. Model-based statistical signal processing: In these approaches, statistical models are developed using event-tracking data collected during the routine normal operation of fermentation. The data from the batch is compared with the template of normal conditions established in the statistical model and diagnosed for process upsets and sensor failures.

The contradictions between observation and predictions allow more than the task detection but provide information on the localization of these faults. Basically, the predictions obtained from models (analytical or logical) lead to contradictions with the observations

* Corresponding author.
E-mail addresses: sregis@univ-ag.fr (S. Régis), adoncesc@laas.fr (A. Doncescu).
URLS: http://lamia.univ-ag.fr (S. Régis), http://www.laas.fr (A. Doncescu).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.10.006

produced by the different sources of information if these are not correct. This reasoning by absurd lets us regroup these conflicts in order to localize the faults. The detection of conflicts is the first step in the diagnostic. The second consists in generating the diagnosis from the set of conflicts.

In this frame work the theory of belief functions used in this paper is an Artificial Intelligence Method of Diagnosis. The originality of the approach presented in this paper is its capacity to manage imprecise and uncertain information. The final purpose of our approach is, first, to have an automatic method to evaluate the relevance of an information source, instead of using the subjective knowledge of an human expert, and secondly, to improve the result of classification (by detecting non-relevant sources of information). More precisely, we use the notion and values of distance measures in order to assess the relevance of a source. This approach enables to estimate the relevance without *a priori* informations on the source: the estimation of the relevance is based only on the intrinsic informations. We use two measures of conflict: a measure based on a norm between the mass function of the sources (Régis et al., 2004a, 2007) and a measure called the Jousselme distance (Jousselme et al., 2001). We test the evaluation of the relevance with the two measures on a practical application of fermentation bioprocess and the results show that this approach improves the results of classification.

The paper is organized as follows: Section 2 presents the context and problematic of bioprocess diagnosis, Section 3 presents the notion of relevance; Section 4 provides the basic notion of the theory of belief functions. In Section 5 we make the connection between the relevance and the theory of belief functions and Section 6 presents the first experimental results for a bioprocess.

2. Bioprocess diagnosis

There are several kind of bioprocesses. One of the most well known is the batch, but there are also fed-batch and single CSTR (chemostat) bioreactors which also fed with sterile nutrient medium. The difference between them is in the manner of operate without or with external sources of biomass after inoculation. When we speak of by external source of biomass, we mean the biomass introduced into the bioreactor after the inoculation. After this process of inoculation, the culture is maintained at conditions that are compatible with growth (e.g. at suitable temperatures) and often kept in an agitated state.

Three biological reactions need to be diagnosed in view of control:

Growth:

$$Glucose + Salts + Vitamins + O_2 \rightarrow Y_{S,X} Biomass + CO_2 + H_2O \quad (1)$$

Ethanol production:

$$Glucose \rightarrow Y_{S,P} Ethanol + CO_2 \quad (2)$$

Maintenance:

$$Glucose + O_2 \rightarrow CO_2 + H_2O \quad (3)$$

where Y represents the yield.

Depending on the organism being cultivated, the fermentation is typically carried out at volume ranging from a few liters up to a few hundred thousand liters, and lasting for a period of several hours or up to a few weeks. Furthermore, there may be additional requirements for amino acids, vitamins, purines, or pyrimidine. It may be necessary to add precursors according to the metabolites that we want to obtain. Oxygen may or may not be required as the terminal electron acceptor, and the fermentation will need to be carried out at an appropriate pH. The medium components used to satisfy an organism's nutritional requirements are partially influenced by the nature of the fermentation process used developed.

The biological system studied in this article is the yeast *Saccharomyces cerevisiae*. Yeast is one of the smallest eukaryotic systems sequenced and is unparalleled for the level of molecular investigations that have been carried out and the range of possible manipulations. It is an ideal target for a comprehensive study at the system-level. Two directions have been explored:

1. *on-line* analysis: It does not permit diagnosis in an instantaneous manner or with certainty regarding the physiological state of the yeast.
2. *off-line* analysis: It makes possible to soundly characterize the current state, but unfortunately after too late to take into account this information or to adjust the process on the fly by actions of regulators allowing to adjust some critical parameters such that pH, temperature (addition of basis, heats, cooling).

To remedy these drawbacks, computer scientists in collaboration with micro-biologists have developed tools for supervised control of the bioprocess. They use the totality of information provided by the sensors during a set of sample processes to infer some general rules to which the biological process obeys. These rules can be used to control the next processes. This is exactly the problem we tackle in this paper. To sum up, our application focuses on the evolutive behavior of a *bio-reactor* (yeast fermentation), that is to say an evolutive biological system whose interaction with the physical world, as described by pH, pressure, temperature, etc…, generates an observable reaction. This reaction is studied by way of a set of sensors providing a large amount of (generally) numerical data.

Physiological state diagnosis uses the relevant information provided by sensors in a process of classification. This “relevant classification” associates classes to physiological states. Faults or abnormal behavior means apparitions of physiological states different from the fermentation goals. Therefore our goal is to avoid the reinforcement of the faults or abnormal behavior in the classification process.

The data used in this paper was obtained from the biomass production. *Saccharomyces cerevisiae* was studied under oxidative regime (i.e. no ethanol production). Two different protocols have been applied: a batch procedure that is followed by a continuous procedure. The batch procedure is composed by a sequence of biological stages. This phase can be thought of a start-up procedure. Biotechnologists state that the behavior in the batch procedure induced influences later phenomena in the continuous phase. So complete knowledge of the batch phase is of great importance for the biotechnologist. The traditional way of getting acquainted with such knowledge is at present carried out through offline measurements and analysis which most of the time produce results when the batch procedure has ended, thus lacking real time performance. Contrast the proposed methodology allows for real time implementation. This example deals with the batch procedure. The expert chooses among the set of available on-line signals which, according to the expert knowledge contain the most relevant information to diagnose the physiological state:

1. DOT: Partial oxygen pressure in the medium.
2. O_2: Oxygen percent in the output gas
3. CO_2: Carbon dioxide percent in the output gas
4. pH.
5. OH-ion consumption: Derived from control action of the pH regulator and the index of reflectivity.

The consumption of negative OH ions is evaluated from the control signal of the pH regulator. The actuator is a pump, switch on by an hysteresis relay, that inoculates a basic solution (NaOH).