Engineering Applications of Artificial Intelligence 72 (2018) 340–349

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

# Operation space design of microbial fuel cells combined anaerobic–anoxic–oxic process based on support vector regression inverse model☆

Jing Wang[a,*], Qilun Wang[a], Jinglin Zhou[a], Xiaohui Wang[b], Long Cheng[c]

[a] College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China
[b] Beijing Engineering Research Center of Environmental Material for Water Purification, College of Chemical Engineering, Beijing University of Chemical Technology, Beijing, 100029, China
[c] State key laboratory of management and control for complex systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

A R T I C L E I N F O

*Keywords:*
MFC-A$^2$/O
Operation space design
Support vector regression
Inverse model
Prediction uncertainty estimation

A B S T R A C T

Microbial Fuel Cells (MFCs) can produce power at the same time of wastewater treatment, which is a new technique for environmental protection and new energy. An appropriate space design of operation variables is very important to improve the performance of MFC process. This paper presents a space design method based on data-driven model but not the traditional mechanism model, which is easy to accomplish in a fast and cost-effective mode. The support vector regression (SVR) forward and inverse model are deduced with the quadratic kernel function, in which the quadratic kernel function is suitable for the mathematical formula in the inversion stage. And the space design of operation variables are proposed to calculate directly from the inverse model with the effect of confidence interval when the model prediction uncertainty are considered. The proposed design method is verified in the real MFC-A$^2$/O equipment. It is shown that the designated operation space is a narrow and effective region of the knowledge space which brackets the entire fraction of the MFC experiment space. And in general terms, the possible product quality from the designated operation space is more densely concentrated on the desired value compared to the tradition forward model design method.

## 1. Introduction

Microbial fuel cells (MFC) have become a promising but challenging technology in recent years. As the main type of the bio-electrochemical system, MFCs can convert biomass into electrical energy through the metabolic activities of microorganisms (Pant et al., 2010). Initially, MFC combining with the treatment of wastewater is proposed by Liu et al. (2004). MFC is considered to be a technology to meet urgent energy needs, especially using wastewaters as substrates, which can get new energy and solving environmental problems (Logan, 2009). At present, most researchers are committed to work on how to improve the performance of microbial fuel cells and application in order to put into practical production (Yousefi et al., 2016).

An appropriate design in MFCs is very important and researchers are always concentrating on coming up with designs of MFCs with improved performance. A majority of researchers are about the work on designs of structure and material in MFCs. Recent years, some adaptations have been made in MFCs design and structure, which are discussed by Du et al. in detail (Du et al., 2007). Meanwhile,optimization of operational parameters and design of the electrode materials and reactor configurations all can enhance the performance of MFCs (Haishu et al., 2016). For the operation conditions design, there are some finished work which are the effect of C/N ratio on contaminant removal (Huang et al., 2013), the effect of PH on biofilm activity (Kelly and He, 2014) and the effect of hydrogen peroxide and chemical oxygen demands concentration (Tamakloe, 2015). Most of the research results are based on the experimental design or the improvement of the device material. There still are some limitations on the factors of time and cost required to finish the experiments. The operation condition design based on MFC system model is still an open task for quality attributes (COD, ammonia nitrogen, output voltage, etc.) located within a specific range of values.

To overcome the limitation of MFC experiment, several studies focused on developing mathematical models for investigating the

☆ This work is supported by the National Natural Science Foundation of China (61573050, 61473025 and 61633016), Beijing Natural Science Foundation (4162066) and the open-project grant funded by the State Key Laboratory of Synthetical Automation for Process Industry at the Northeastern University (PAL-N201702).
* Corresponding author.
*E-mail address:* jwang@mail.buct.edu.cn (J. Wang).

https://doi.org/10.1016/j.engappai.2018.04.005
Received 26 October 2017; Received in revised form 8 March 2018; Accepted 5 April 2018
Available online 8 May 2018
0952-1976/

influence of various factors on MFC performance in addition to being cost effective and less time consuming. The mathematical models for several kinds of MFC system have been developed which are based on the Nernst–Monod type equations to quantify the kinetic characteristics of substrate consumption and bacterial growth (Oliveira, 2013; Perlman, 2013). Actually, it is an extra load to develop the mechanism model due to its highly complexity and long-time calculation. The data-driven computing or learning methods can be used as an alternative method for modeling complex physical non-linear systems in the fields of energy (Garg et al., 2014), steel, metallurgy (Lei et al., 2016), chemical industry (Wu et al., 2012), bio-pharmacy (Liu et al., 2016), insurance and finance (Lu and Kao, 2016; Wei et al., 2017), and so on. For example, SVR is applied to analyze operating cost of laser cutting through the process parameters (Jovi et al., 2016b), to estimate contact forces of the robotic finger (Jovi et al., 2016a)and to predict the precipitation concentration index according to the monthly precipitation data during the period of 1946–2012 (Gocic et al., 2016). Otherwise, daily dewpoint temperature predictive models of 10 years data sets (Mohammadi et al., 2016)and short-term multistep-ahead predictive models of heat load for consumers are developed using support vector machine (Al-Shammari et al., 2016). The traditional data-driven modeling methods (Garg et al., 2014; Geng et al., 2017; Chen et al., 2017; Garg et al., 2017b, a) include support vector machine (SVM), fuzzy logic, extreme learning machine (ELM), particle swarm optimization (PSO), automated neural search(ANS), genetic programming(GP) and artificial neural network (ANN). They offer the advantage of a fast and cost-effective explicit formulation based on multiple variables without any pre-definition of non-linear structure of the model. So the data-driven model can overcome the shortcomings of mechanism model when it is applied to operation condition design. Under the condition that many methods can solve nonlinear problems, SVM is more suitable for small sample data, but also suitable for high-dimensional samples and the mathematical theory foundation of SVM is quite perfect. Furthermore, the data collected from the real equipment are not enough to support big data learning. So SVM is supposed to be a good choice and will show better performance than other methods.

In this paper, we try to determine the operation conditions based on the data-driven model for the new MFC developers. We presents a methodology for the design of the operation space in the view of data-driven inversion model and uncertainty analysis. First, we develops a SVR model according to the history data acquired from the existing MFC device. An SVR inversion model is developed to locate the normal operation point for the desired MFC performance. Considering the effect of model prediction uncertainty, the reasonable domain for the operation variables can be bracketed around the normal point, who can be used to guide the new MFC development. The proposed method is different from the previous experiment method that the inverse data-driven model is applied to design the operation conditions. Traditional operation condition design is determined using experiments or first-principles model carried out within a domain of input combinations (e.g., raw materials properties and process operating conditions) that result from similar products already developed. It is difficult to realize due to the strict demanding for the related experimentations especially if the number of inputs is large. Additionally, it also is an alternative method for the traditional optimization design without solving any complex optimization problem.

Compared to the traditional design method based on experiments or first-principles model, the proposed methodology has the following advantages,

(1) The proposed methodology relies on the exploitation of historical databases on products that have already been obtained. It is based on the data-driven model, such as soft computing model (for instance, neural network, multi-gene genetic programming and other artificial intelligent methods) or latent-variable model (for instance, principal component analysis, projection to latent structures and other multivariate statistical methods), which is not limited to SVR method.

(2) This model gives an explicit formulation for the output voltage as a function of materials properties and process conditions. It is usually much simpler to develop than first-principles ones, and the appropriate number of samples does not necessarily need to be very large. The data-driven model can be conveniently generalized to predict a new product for the determination of the operation condition or design space without extra test experiments.

(3) The operation condition or design space are directly calculated from the inverse model with the effect of model prediction uncertainty. It will give a narrower region within which to design and carry out the experiments without solving any complex optimization problems. So taken in this sense, the model inversion problem will be called as product development problem.

The remainder of this paper is organized as follows. Mathematical backgrounds and methodology is given in Section 2. Here we discuss the SVR forward/inversion model and their prediction uncertainty, then present the operation space design based on the uncertainty analysis of inversion model. In Section 3, the proposed methodology is applied on a real MFC case, a combination system of MFC and Anaerobic–Anoxic–Oxic units (MFC-A$^2$/O). The experimental details are discussed to collect data required for modeling. SVM and ANN techniques are used to model the MFC performance with five operation variables as inputs and voltage as output, respectively. The prediction performances of ANN and SVR method are compared, and SVR shows better performance due to the small samples problem. The operation space design based on SVR inversion is shown to verify the effectiveness of the proposed method. Finally, Section 4 draws the conclusions and recommendations for future work.

## 2. Operation condition design based on SVR inverse model

This section provides the theoretical basis for the modeling and model inversion. According to the experiment data we can establish a relatively accurate data-driven model and then calculate the model inverse to order to obtain the window of operating conditions. In the following sections manipulated and output variables arranged as the columns in an $X$ and $Y$-matrix respectively. From now on there are more manipulated variables than output variables, and hence, that $X$ has more columns than $Y$.

### 2.1. Support vector Regression (SVR) modeling

MFC-A$^2$/O is high complex and strong nonlinear, which is difficult to establish its mathematical model based on the first-principles. So data-driven model, such as artificial intelligent methods and multivariate statistical methods, is acceptable to obtain the relationship between operation variables and system output. Data-driven models are usually much simpler to develop than first-principles ones, but their development requires a fairly large amount of data. Here Support vector machine(SVM) for regression analysis, i.e., support vector regression (SVR) is selected due to its advantage in small sample learning.

Suppose the training data is $\{x_i, y_i\}_1^N$ with process variables $x_i \in R^n$ and output $y_i \in R^1$. The optimal relationship between input and output in high dimensional feature space is:

$$f(x) = \omega^T \phi(x) + b \tag{1}$$

where $\omega$ is support vector weight, $b$ is the bias. And $\phi(x)$ is the kernel function, in which the nonlinear transformation $\phi(\cdot)$ is to map the original data to the high dimension feature space. The linear estimation (1) is performed in the high dimension space. An insensitive loss function is introduced,

$$|y - f(x)|_\varepsilon = \begin{cases} 0, & if\ |y - f(x)| \le \varepsilon \\ y - f(x) - \varepsilon, & otherwise \end{cases} \tag{2}$$

Then the vector machine regression problem is

$$\min_{\omega,b,\xi_i,\xi_i^*} \quad \frac{1}{2}\omega^T \omega + C\sum_{i=1}^{N} \xi_i + C\sum_{i=1}^{N} \xi_i^* \tag{3}$$