文章编号:1671-4598(2018)07-0041-03 DOI:10.16526/j.cnki.11—4762/tp.2018.07.009 中图分类号:TP273 文献标识码:A

基于 KPCA—DFNN 海洋微生物发酵过程软测量建模

孙丽娜[1]，黄永红[2]，蒋星红[1]，冯培燕[1]
(1. 苏州工业园区职业技术学院，江苏 苏州 215123；2. 江苏大学 电气信息工程学院，江苏 镇江 212013)

摘要：针对海洋微生物发酵过程中关键生物参量（基质浓度、菌体浓度、产物浓度等）在线测量困难，离线化验滞后大，难以实现实时控制的问题，提出了一种基于核主元分析（KPCA）与动态模糊神经网络（DFNN）相结合的软测量方法；以典型的海洋微生物—海洋蛋白酶发酵过程为例，通过 KPCA 提取输入数据空间中的非线性主元，将提取的主元作为 DFNN 的输入，基质浓度、菌体浓度、相对酶活作为 DFNN 的输出，建立了基于 KPCA—DFNN 的海洋蛋白酶发酵过程生物参量软测量模型；仿真结果表明，KPCA—DFNN 模型比 DFNN 和 PCA—DFNN 建模的测量精度高，跟踪性能强，能很好地满足发酵过程中生物参量的测量要求。

关键词：海洋微生物；生物参量；核主元分析；动态模糊神经网络

Soft Sensor Modeling for Marine Microbe Fermentation Process Based on KPCA and DFNN

Sun Lina[1], Huang Yonghong[2], Jiang Xinghong[1], Feng Peiyan[1]
(1. Suzhou Industrial Park Institute of Vocational Technology, Suzhou 215123, China;
2. College of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: To overcome the difficulty that crucial biological variables (such as substrate concentration, biomass concentration, product concentration, etc.) cannot be effectively controlled during the marine microbe fermentation process due to a lack of real—time on—line instrumentation, a soft sensor method is proposed by combining the Kernel Principal Component Analysis (KPCA) with the Dynamic Fuzzy Neural Network (DFNN). The typical marine microbe fermentation process (the marine protease fermentation process) was taken as an example. Firstly, KPCA was applied to choose the nonlinear principal component of the model input data space. And then its result was taken as input of the DFNN, substrate concentration , biomass concentration and relative enzyme activity were taken as output of the DFNN. Finally, the soft sensor model of biological parameters based on KPCA—DFNN is established in the marine protease fermentation process. Simulation results indicate that the KPCA—DFNN model has a higher accuracy, better tracking performance when compared with DFNN model and the PCA—DFNN model. Therefore, the proposed method can satisfy the requirements of on—line measurement of biological variables in the marine microbe fermentation process .

Keywords: marine microbe; biological parameters; kernel principal component analysis; dynamic fuzzy neural network

0 引言

海洋微生物作为微生物的一类，因其生存的海洋具有特殊的环境，故其所产生的酶与其他微生物所产生的酶相比具有更加独特的性质，如耐低温，耐碱性，PH 作用范围宽等，这使得海洋微生物在食品加工、酶工业、添加剂和医药等发酵行业具有极大的开发潜力和应用前景[1-5]。在海洋微生物发酵过程中，为保证发酵产物的品质和质量，需要实时检测一系列生物参数，尤其是基质浓度、菌体浓度及产物浓度（酶活）。当前，在线测量仪器仅能检测发酵过程中某些物理和化学参数，还没有成熟实用的仪器来测量这些关键生物参数[6-8]。在此背景下，许多关于微生物发酵过程中生物参数的软测量方法应运而生，其中，基于神经网络[9]的预测方法成为软测量领域的研究热点。然而，对于海洋微生物发酵过程这一类十分复杂的非线性系统来说，如果没有先验知识，就盲目应用神经网络方法，关键生物参数的测量问题就不能很好的解决。为此，文中将核主元分析法（Kernel Principal Component Analysis，KPCA）[10-12]和动态模糊神经网络（Dynamic Fuzzy Neural Network，D—FNN）[13-15]相结合，提出了一种基于 KPCA—DFNN 的软测量方法。

以典型的海洋微生物—海洋蛋白酶发酵过程为研究对象，首先，确定基质浓度、菌体浓度、相对酶活（能够更好地描述酶活的变化趋势）这三个参量为软测量模型的输出变量，环境变量为模型的初始输入变量。然后，利用

收稿日期:2017-11-13； **修回日期**:2017-12-15。
基金项目:江苏高校优势学科建设工程资助项目(PAPD)；“十二五”国家 863 计划重点科技项目(2011AA09070301)；江苏省自然科学基金面上项目(BK20151345)；江苏高校品牌专业建设工程资助项目(PPZY2015A088)。
作者简介:孙丽娜(1986-)，女，山东聊城人，硕士，讲师，主要从事复杂过程的智能检测与控制方向的研究。

KPCA 对输入变量进行数据压缩和信息抽取，将所提取的主元作为 DFNN 的输入，以上三个变量作为 DFNN 的输出，建立了基于 KPCA－DFNN 的海洋蛋白酶发酵过程生物参量软测量模型。仿真结果表明，该模型较基于 DFNN 和 PCA－ DFNN 建模具有学习速度快、预测精度高等优势，有益于海洋蛋白酶的高效、高质量生产。

1 核主元分析与动态模糊神经网络的原理

1.1 核主元分析

核主元分析是主元分析法（Principal Component Analysis，PCA）的非线性推广，其具体算法如下：

给定 n 个样本，样本集 $X=\{x_1,x_2,\cdots,x_n\}$，$x_k\in \boldsymbol{R}^m$，由非线性函数 $\varphi(\cdot)$ 将输入数据从原空间映射到高维特征空间 F，F 中的样本记为 $\varphi(x_k)$，且满足：

$$\sum_{k=1}^{n}\varphi(x_k)=0 \tag{1}$$

在 F 空间中样本的协方差矩阵 C 为

$$C=\frac{1}{n}\sum_{i=1}^{n}\varphi(x_i)\varphi(x_i)^T \tag{2}$$

对 C 进行特征值分解

$$\lambda V=CV \tag{3}$$

式中，V 是与 λ 对应的特征向量，特征值 $\lambda\geqslant 0$。将（3）式的两边左乘以核样本 $\varphi(x_k)$：

$$\lambda\varphi(x_k)\cdot V=\varphi(x_k)\cdot CV,k=1,2,\cdots,n \tag{4}$$

解方程可得与非零特征值对应的特征向量 V。其解一定处于 $\varphi(x_1),\varphi(x_2),\cdots,\varphi(x_n)$ 张成的空间中，所以 V 可以表示为：

$$V=\sum_{i=1}^{n}\alpha_i\varphi(x_i)=\varphi(x)\alpha \tag{5}$$

其中，

$\varphi(x)=[\varphi(x_1),\varphi(x_2),\cdots,\varphi(x_n)]$，$\alpha$ 表示 $a_1,\cdots,a_n$ 中的一个列向量。

引入核函数：

$$K_{ij}=K(x_i,x_j)=[\varphi(x_i),\varphi(x_j)],i,j=1,2,\cdots,n \tag{6}$$

本文选用径向基核函数 $K(x,x_i)=\exp\left[-\frac{\|x-x_i\|^2}{2\sigma^2}\right]$，其中，$\sigma$ 为核宽度。可以转变为求核矩阵 $\boldsymbol{K}$ 的特征值和特征向量：

$$n\lambda\alpha=\boldsymbol{K}\alpha \tag{7}$$

归一化特征向量 $\boldsymbol{V}$，即 $(V,V)=1$，即可得样本 x 在特征空间中的第 k（$k=1,2,\cdots,n$）个主元分量 $t_k(x)$：

$$t_k(x)=\boldsymbol{V}\cdot\varphi(x)=\sum_{i=1}^{n}\alpha_i^k K(x,x_i) \tag{8}$$

特征值 λ_k 小的主元 t_k 可以认为是噪声引起的。比值 $\lambda_k/\sum_{i=1}^{n}\lambda_i$ 反映了分量 t_k 对整体方差的贡献率，较重要的分量对应较大的值，主元数量的选取一般依据：

$$\left(\sum_{k=1}^{p}\lambda_k/\sum_{k=1}^{n}\lambda_k\right)>E \tag{9}$$

式（9）表示前 p 个 λ_k 的和与总和比值大于 E，通常选取 $E>85\%$。

以上推导特征空间变量均值都是以 $\sum_{k=1}^{n}\varphi(x_k)=0$ 为假设条件的，然而实际中的样本数据并不一定满足等式 $\sum_{k=1}^{n}\varphi(x_k)=0$，此时，可通过 $\tilde{\boldsymbol{K}}$ 取代 $\boldsymbol{K}$ 来实现，

$$\tilde{\boldsymbol{K}}=\boldsymbol{K}-\boldsymbol{LK}-\boldsymbol{KL}+\boldsymbol{LKL} \tag{10}$$

其中：$\boldsymbol{L}$ 为 $\frac{1}{n}$ 的 $n\times n$ 阶单位矩阵。

1.2 动态模糊神经网络

图 1 为动态模糊神经网络的结构。图中 $x_1,x_2,\ldots x_r$ 为输入变量，MF_{ij} 是第 i 个输入变量的第 j 个隶属函数，R_j 是第 j 条模糊规则，N_j 是第 j 个归一化节点，w_j 是第 $x'=\frac{x_{\max}-x}{x_{\max}-x_{\min}}$ 个规则的连接权，x' 是系统的总规则数，y 是系统的输出。下面介绍 DFNN 各层的含义。

第 1 层：输入层，每个节点代表一个输入变量。

第 2 层：隶属函数层，每个节点代表一个隶属函数，该隶属函数可用式（11）表示：

$$u_{ij}(x_i)=\exp\left[-\frac{(x_i-c_{ij})^2}{\sigma_j^2}\right] \tag{11}$$

其中，$i=1,2,\cdots,r$，$j=1,2,\cdots,u$，r 是输入变量数，u 是隶属函数的数量，即系统的总规则数，c_{ij} 和 σ_j 分别是 x_i 的第 j 个高斯隶属函数的中心和宽度，$x_{\min}$ 是 x_i 的第 j 个高斯隶属函数。

第 3 层：T－范数层，每个节点代表一个可能的模糊规则中的 IF－部分，即该层的节点数反映了模糊规则数。第 j 个规则 R_j 的输出为：

$$\varphi_j=\exp\left[-\frac{\sum_{i=1}^{r}(x_i-c_{ij})^2}{\sigma_j^2}\right]=\exp\left[-\frac{\|X-C_j\|^2}{\sigma_j^2}\right] \tag{12}$$

其中：$j=1,2,\cdots,u$，$X=(x_1,x_2,\ldots,x_r)\in\boldsymbol{R}^r$，$C_j=(c_{1j},\ldots,c_{rj})\in\boldsymbol{R}^r$ 是第 j 个 RBF 单元的中心。

第 4 层：归一化层，这些节点被称为 N 节点。其数目等于模糊规则的节点数。第 j 个节点 N_j 的输出为：

$$\varphi_j=\frac{\varphi_j}{\sum_{i=1}^{u}\varphi_k},j=1,2,\cdots,u \tag{13}$$

第 5 层：输出层，每个节点代表一个输出量，此输出是所有输入信号的叠加：

$$y(X)=\sum_{k=1}^{u}w_k\cdot\varphi_k \tag{14}$$

式中，w_k 是 THEN－部分或者称为第 k 个规则的连接权，$w_k=a_{k0}+a_{k1}x_1+\cdots a_{kr}x_r$，$k=1,2,\cdots u$，$y$ 是变量的输出。

2 KPCA－DFNN 模型的构建与验证

2.1 模型的构建

基质浓度 S、菌体浓度 X、相对酶活 P 对海洋蛋白酶发酵过程的优化控制非常重要，因此，选择这三个变量作为 KPCA－DFNN 软测量模型的输出变量。通过分析海洋蛋白

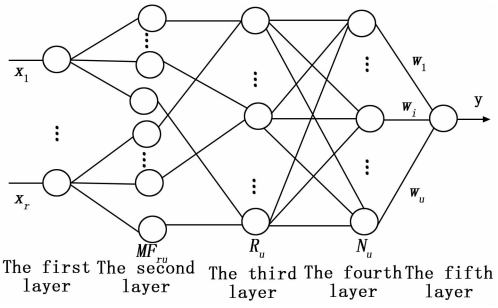

图 1　DFNN 的结构图

酶发酵机理并结合发酵过程的实验数据，选取了 10 个影响生物参数 S、X、P 的主要因素作为初始样本变量，分别是：时间 t、温度 T、搅拌速度 r、溶解氧浓度 DO、空气流量 l、pH 值、CO_2 浓度、基质进给速率 u、发酵罐压力 p、反应器体积 v。

采集了 10 批发酵数据，前 9 批作为模型的训练样本集，第 10 批作为模型的测试样本集，由于采集到的样本数据变化范围较大，如果直接使用原始测量数据进行计算，不仅会夸大量纲数据的作用，而且还可能导致信息丢失或引起数值计算的不稳定。需要对样本数据进行归一化处理。公式如下：

$$x' = \frac{x_{\max} - x}{x_{\max} - x_{\min}} \tag{15}$$

式中：$x_{\max}$ 为样本数据的最大值，$x_{\min}$ 为样本数据的最小值，x 为原始样本数据，x' 为归一化后的数据。归一化后样本数据在［0，1］之间。

根据核主元分析法和动态模糊神经网络的基本原理，构建了基于 KPCA－DFNN 的海洋蛋白酶发酵过程生物参量的软测量模型，建模过程如图 2 所示。建模步骤如下。

步骤 1：根据建模对象选取适当的输入输出样本数据。

步骤 2：利用式（15）对输入输出数据进行预处理。

步骤 3：根据 KPCA 算法对输入变量进行数据压缩和信息抽取，消除输入变量之间的相关性，进行特征选取。文中按累积方差百分比大于 95％，KPCA 选定了 2 个特征主元，PCA 选定了 6 个特征主元。

步骤 4：将提取非线性主元作为 DFNN 的输入，X、S、P 作为模型的输出变量，利用训练样本集对 DFNN 模型进行训练，选取最佳模型构建参数。

步骤 5：利用测试样本集对建好的 KPCA－DFNN 软测量模型进行验证。

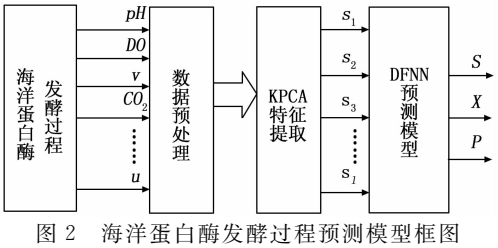

图 2　海洋蛋白酶发酵过程预测模型框图

2.2　模型验证

用测试样本集对建好的 KPCA－DFNN 模型进行仿真验证。仿真结果如图 3、图 4 和表 2 所示。

图 3 显示出了海洋蛋白酶发酵过程 X、S、P 的离线化验值（真实值）和软测量值（预测值）对比结果。

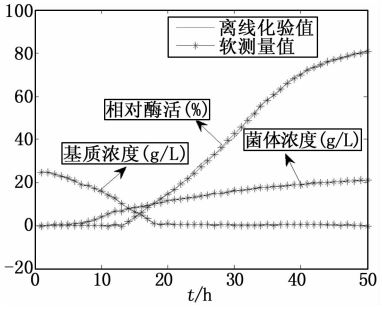

图 3　KPCA－DFNN 生物参量预测曲线

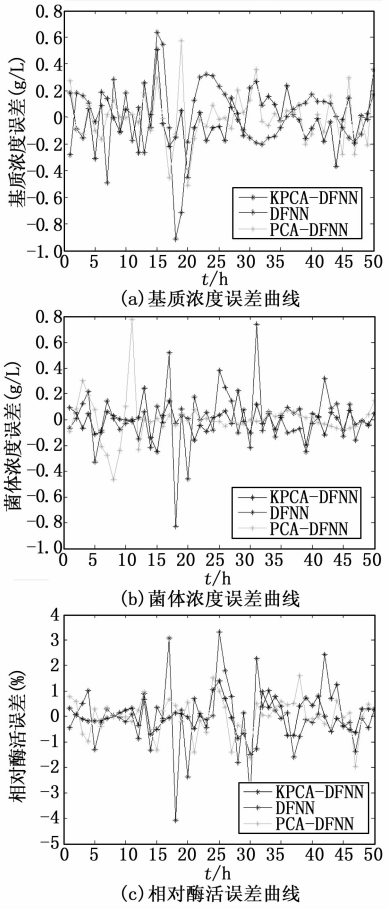

图 4　DFNN、PCA－DFNN、KPCA－DFNN 生物参量预测值误差曲线

虽然试验过程中采集到的样本值分散性很大和重复性很小，但从图 3 中可以看出，对于 X、S、P，基于 KPCA－DFNN 的软测量模型，输出的软测量值都能够很好的追踪离线化验值，这说明，KPCA－DFNN 具有较好的逼近能力。

采用均方根误差（RMSE）和平均绝对误差（MAD）更直观的反映 DFNN、PCA－DFNN、KPCA－DFNN 建模方式的预测效果，如表 1 所示。

（下转第 98 页）