

纤维素酶水解动力学的人工神经网络模型研究

张 宇^{1,2}, 许敬亮¹, 袁振宏¹, 庄新姝¹, 吕鹏梅¹

¹中国科学院广州能源研究所, 中国科学院可再生能源与天然气水合物重点实验室, 广东广州 510640;

²中国科学院研究生院, 北京 100049

摘要: 酶作用机制的模糊以及影响异相体系因素的大量存在, 使得纤维素水解的酶催化过程高度复杂, 很难为之建立理论模型. 采用非理论模型人工神经网络模拟和预测了纤维素酶水解反应, 并与常用的响应面模型进行了比较. 选取加酶量 X_1 , 底物浓度 X_2 和反应时间 X_3 作为自变量, 还原糖浓度 Y_1 和原料转化率 Y_2 作为响应值. 结果表明, 人工神经网络模型比响应面模型更适合作为研究纤维素酶水解的动力学工具. 在模拟过程中, 除中心试验点外, 只有 1 个试验点上人工神经网络模拟值 Y_2 产生的误差大于响应面模型. 在预测过程中, 人工神经网络模型的预测值都比响应面模型更接近实验值.

关键词: 酶催化动力学; 纤维素酶水解; 人工神经网络; 响应面模型; 异相催化

中图分类号: O643 文献标识码: A

Kinetic Model Study on Enzymatic Hydrolysis of Cellulose Using Artificial Neural Networks

ZHANG Yu^{1,2}, XU Jingliang¹, YUAN Zhenhong^{1,*}, ZHUANG Xinshu¹, LÜ Pengmei¹

¹Key Laboratory of Renewable Energy and Gas Hydrate, Guangzhou Institute of Energy Conversion, Chinese Academy of Sciences, Guangzhou 510640, Guangdong, China;

²Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Enzymatic hydrolysis of cellulose was highly complex because of the unclear enzymatic mechanism and many factors that affect the heterogeneous system. Therefore, it is difficult to build a theoretical model to study cellulose hydrolysis by cellulase. Artificial neural network (ANN) was used to simulate and predict this enzymatic reaction and compared with the response surface model (RSM). The independent variables were cellulase amount X_1 , substrate concentration X_2 , and reaction time X_3 , and the response variables were reducing sugar concentration Y_1 and transformation rate of the raw material Y_2 . The experimental results showed that ANN was much more suitable for studying the kinetics of the enzymatic hydrolysis than RSM. During the simulation process, relative errors produced by the ANN model were apparently smaller than that by RSM except one and the central experimental points. During the prediction process, values produced by the ANN model were much closer to the experimental values than that produced by RSM. These showed that ANN is a persuasive tool that can be used for studying the kinetics of cellulose hydrolysis catalyzed by cellulase.

Key words: enzymatic kinetics; enzymatic hydrolysis of cellulose; artificial neural network; response surface model; heterogeneous catalysis

纤维素可在由葡聚糖内、外切酶和 β -葡萄糖苷酶 3 种组分组成的纤维素酶的协同作用下水解成葡萄糖等还原性糖^[1,2]. 3 种酶复杂的作用机制以及大量影响纤维素-纤维素酶这一异相反应体系因素(如颗粒尺寸分布、搅拌方式等)的存在, 严重阻碍了动力学理论模型的建立^[3,4]. 研究表明, 米氏(Henri-

Michaelis-Menten) 方程不能适应于异相催化反应, 特别是存在扩散限制的纤维素酶水解反应^[5,6]. 因此, 目前研究者倾向于用经验模型或其他非理论模型来研究纤维素酶水解反应动力学^[7].

人工神经网络(artificial neural network, ANN)由基本神经元相互连接, 能模拟人脑的神经处理信息

收稿日期: 2008-11-24.

联系人: 袁振宏. Tel: (020)87057735; Fax: (020)87057737; E-mail: yuanzh@ms.giec.ac.cn.

基金来源: 国家高技术研究发展计划(863 计划, 2007AA100702-4 和 2007AA05Z406); 中国科学院知识创新工程重大项目(KSCX1-YW-11-A3)和重要方向项目(KSCX2-YW-G-063-1).

方式进行信息并行处理和非线性转换,是目前生物学领域中最受欢迎的学习工具,其应用包括基因、蛋白质序列的功能分析以及基因表达图谱的模式认知等^[8-13].它作为一种“黑箱子”模型,不需要清晰明了的理论作为支持,十分有助于建立纤维素酶水解的预测模型^[14].目前国内外还没有关于 ANN 模拟和预测纤维素酶水解反应的相关报道.

本文采用 ANN 模型研究了纤维素酶水解动力学,并与常用的响应面模型(response surface model, RSM)作了相应比较.

1 实验部分

准确称取 10.0 g 稻秆(取自湖南省望城县境内,粉碎过 40 目筛,干燥)于盛有 300 ml NaOH 溶液(0.5 mol/L)的圆底烧瓶内,于 100 °C 下常压回流 60 min,然后将残渣用沸水反复洗涤至中性,于 80 °C 烘干至恒重备用.

将纤维素酶(购自 Sigma 公司,活力约 60 FPU/g, FPU = filter paper unit)、碱处理过的稻秆残渣以及 25 ml 的 HAc-NaAc 缓冲溶液(0.2 mol/L, pH 5.0)混匀,于 50 °C, 100 r/min 的条件下反应一定时

间后,用 DNS (3,5-二硝基水杨酸)法测定还原糖浓度 Y_1 (g/L), 计算原料转化率 Y_2 , $Y_2 = 100\% \times 0.9Y_1/X_2$, 其中 X_2 (g/L) 为底物浓度.

ANN 模型的建立通过软件 Matlab 7.0 实现, RSM 由软件 Design-Expert 7.1.3 完成.

2 结果与讨论

2.1 响应面模型

本试验中,我们选取加酶量 X_1 , 底物浓度 X_2 和反应时间 X_3 作为自变量,以还原糖浓度 Y_1 和原料转化率 Y_2 作为响应值.采用 Box-Behnken 原则选取试验样本,其中共有 12 个析因点,5 个中心点,如表 1 所示,最后对试验结果进行二次项回归分析,得到的回归方程为:

$$Y_1 = -4.41902 + 0.15090X_1 + 0.88044X_2 + 0.67885X_3 - 0.0000169474X_1X_2 + 0.00617167X_1X_3 + 0.00973526X_2X_3 - 0.00283350X_1^2 - 0.027370X_2^2 - 0.044999X_3^2$$

$$Y_2 = 6.9671 + 1.28258X_1 + 0.98759X_2 + 10.24693X_3 - 0.10367X_1X_2 + 0.0555X_1X_3 + 0.083246X_2X_3 + 0.04365X_1^2 - 0.054708X_2^2 - 0.59708X_3^2$$

表 1 建立(训练)模型的试验样本以及实测值与模型模拟值比较

Table 1 Training samples for building models and comparison of experimental and model values

X_1 (FPU)	X_2 (g/L)	X_3 (h)	Y_1 (g/L)						Y_2 (%)					
			EV	RSM		ANN		EV	RSM		ANN			
				MV	RE(%)	MV	RE(%)		MV	RE(%)	MV	RE(%)		
10	5	9	3.4144	3.9821	6.63	3.5828	4.93	61.46	67.65	10.07	60.44	1.66		
5	10	3	4.0600	4.3474	7.08	4.3062	6.06	36.54	37.40	2.35	35.28	3.45		
15	20	6	7.8310	8.0395	2.66	7.9099	1.01	35.24	37.79	7.24	35.53	0.82		
15	10	9	7.7333	7.4459	3.72	7.7518	0.24	69.60	68.74	1.24	69.66	0.09		
15	5	6	4.4088	4.2243	4.18	4.5550	3.32	79.36	74.31	6.36	78.62	0.93		
10	10	6	6.2799	6.2799		6.2167		56.52	56.52		56.93			
10	10	6	6.2799	6.2799		6.2167		56.52	56.52		56.93			
10	10	6	6.2799	6.2799	0	6.2167	1.01	56.52	56.52	0	56.93	0.73		
10	10	6	6.2799	6.2799		6.2167		56.52	56.52		56.93			
10	20	3	6.2488	5.8649	6.14	6.2196	0.47	28.12	23.33	17.03	27.99	0.46		
15	10	3	5.2097	5.4732	5.06	5.2067	0.06	46.89	50.25	7.71	47.00	0.23		
5	10	9	6.2133	5.9498	4.26	6.2848	1.15	55.92	52.56	6.39	55.53	0.70		
5	5	6	3.1289	2.9125	6.92	2.9908	4.41	56.32	54.61	3.04	56.72	0.07		
10	10	6	6.2799	6.2799	0	6.2167	1.01	56.52	56.52	0	56.93	0.73		
10	5	3	2.6533	2.4865	6.29	2.4881	6.23	47.76	48.33	1.20	48.72	0.08		
5	20	6	6.5377	6.7303	2.95	6.5591	0.33	29.42	33.64	14.34	29.50	0.27		
10	20	9	8.2533	8.2366	0.20	8.1645	1.08	37.14	35.16	5.33	36.67	4.29		

X_1 : cellulase amount; X_2 : substrate concentration; X_3 : reaction time; Y_1 : reducing sugar concentration; Y_2 : transformation rate, $Y_2 = 100\% \times 0.9Y_1/X_2$; RSM: response surface model; ANN: artificial neural network; EV: experimental value; MV: model value; RE: relative error.

2.2 人工神经网络

最典型的ANN是前馈(back-propagation, BP)神经网络, 由1个输入层、1个输出层和1个或更多的隐含层构成. 研究表明, 3层BP神经网络能够模拟任意精度的连续函数^[6,12], 因此本试验中仅考虑含有1个隐含层的网络. 同样选取上述3个参数和2个响应值分别作为输入和输出变量, 建立BP网络来模拟与预测纤维素酶解反应(图1).

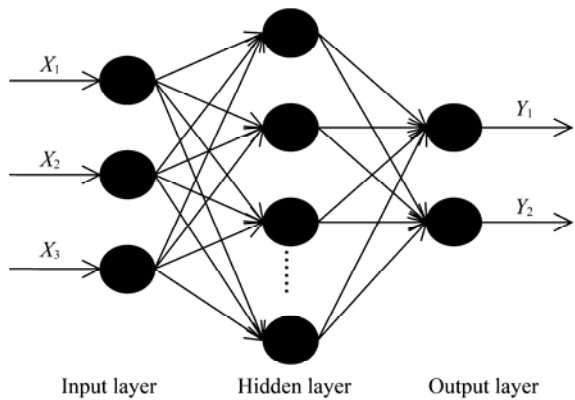


图1 纤维素酶水解的人工神经网络模型

Fig. 1. Schematic representation of the artificial neural network for simulating enzymatic hydrolysis of cellulose.

BP神经网络的基本思想是, 学习过程由信号的正向传播和误差的反向传播两个过程组成. 若输入样本经正向传播得到的输出值与期望值(试验值)不符, 则转入误差的反向传播阶段, 即将误差分摊给各层, 以此作为误差信号来修正各单元权(阈)值. 周而复始的正向传播与误差反向传播就是网络的学习训练(建立)过程. 此过程一直进行到误差减少到

可接受的程度或预先设定的学习次数为止^[15].

本试验采用Matlab编写程序, 隐含层和输出层的传递函数分别为S型的正切函数 *tansig* 和纯线性函数 *purelin*, 学习函数为梯度下降权值/阈值的 *learngd*, 用BFGS (Broyden-Fletcher-Goldfarb-Shanno) 准牛顿算法函数 *trainbfg* 来训练(调整)网络, 以均方误差函数 $mse = 1e-3$ 为训练目标, 最大训练次数 *net.trainParam.epochs* 为1000, 学习速率 *lp.lr*、训练时间 *net.trainParam.time* 和最小性能梯度 *net.trainParam.min_grad* 等为软件默认值. 通过有限次尝试, 得到1个恰当的3-4-2的BP网络能很好模拟本试验中的酶解纤维素反应, 模型结构即各层的权值与阈值如下所示:

$$iw\{1\} = \begin{pmatrix} -5.2102 & -11.4810 & 0.04090 \\ 0.0443 & 6.09380 & -0.0685 \\ -8.5103 & -32.8037 & 0.0527 \\ -0.1574 & -0.8645 & -0.2277 \end{pmatrix}$$

$$lw\{2\} = \begin{pmatrix} 0.0680 & -28.1383 & -27.5048 & -1.7965 \\ -0.3358 & -53.7954 & -51.8258 & -0.8758 \end{pmatrix}$$

$$b\{1\} = (2.0715 \quad 3.8990 \quad -21.9867 \quad 0.4672)^T$$

$$b\{2\} = (0.6734 \quad 0.4816)^T$$

2.3 模拟与预测比较

由表1可见, 除中心点($X_1 = 10, X_2 = 10, X_3 = 6$)外, 只有在试验($X_1 = 5, X_2 = 10, X_3 = 3$)中, ANN模拟 Y_2 的误差(3.45%)大于RSM的模拟误差(2.35%), 其余试验中ANN的模拟误差均小于RSM的模拟误差. 从误差分布看, ANN的模拟 Y_1 在6.23%以内, 模拟 Y_2 的误差大多在0.1%之内; RSM模拟 Y_1 的误差大多在2.5%以上, 最大值为16.63%, 模拟 Y_2 的误差

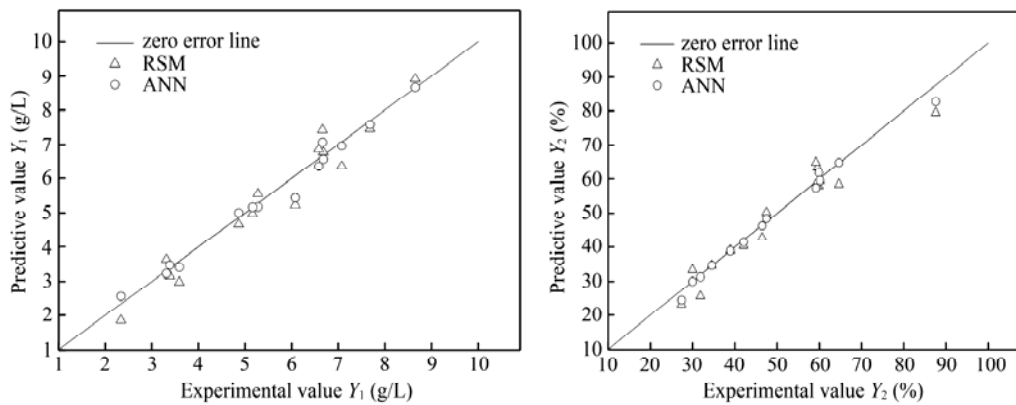


图2 纤维素酶水解反应实验值与模型预测值比较

Fig. 2. Comparison of experimental and model predictive values.

都在 1.0% 以上, 最大值为 17.03%。因此, ANN 能比 RSM 更好地模拟纤维素酶解反应。

建立模型的样本是 3 因子 3 水平中的 13 组试验 (表 1), 其余的 14 组试验用来检验两种模型的预测准确性。由图 2 可以看出, 整体上 ANN 的模拟点明显比 RSM 的模拟点更加靠近零误差线。ANN 的许多模拟点都相当接近零误差线, 脱离零误差线较远的模拟点也都由 RSM 产生。因此, ANN 预测纤维素酶水解反应的准确性也大大优于 RSM。

3 结论

建立合适的模型研究纤维素酶水解反应长期以来都是相当困难的, 许多研究者建立的动力学模型适应范围小, 而且涉及的参数众多, 工作量大^[3,6,16]。在人工智能快速发展的今天, 用 ANN 研究纤维素酶水解动力学是非常便捷的。本文分别用 ANN 和 RSM 研究了纤维素酶水解反应动力学。结果表明, ANN 能更适合模拟与预测纤维素酶水解反应, 是研究纤维素酶水解动力学的有力工具。

参 考 文 献

- 1 张宇, 许敬亮, 王琼, 徐明忠, 庄新姝, 李东, 袁振宏. 农业工程学报(Zhang Y, Xu J L, Wang Q, Xu M Zh, Zhuang X Sh, Li D, Yuan Zh H. *Trans CSAE*, 2008, **24**: 186
- 2 张宇, 许敬亮, 李东, 袁振宏. 中国生物工程杂志(Zhang Y, Xu J L, Li D, Yuan Zh H. *Chin Biotechnol*), 2008, **28**(6s): 311
- 3 Gan Q, Allen S J, Taylor G. *Process Biochem*, 2003, **38**: 1003
- 4 Movagarnejad K, Sohrabi M, Kaghazchi T, Vahabzadeh F. *Biochem Eng J*, 2000, **4**: 197
- 5 刘森林, 宗敏华. 催化学报(Liu S L, Zong M H. *Chin J Catal*), 2006, **27**: 25
- 6 Carrillo F, Lis M J, Colom X, Lopez-Mesas M, Valdeperas J. *Process Biochem*, 2005, **40**: 3360
- 7 袁振宏. [博士学位论文]. 上海: 华东理工大学(Yuan Zh H. [PhD Dissertation]. Shanghai: East China University of Science Technology), 2006
- 8 Almeida J S. *Curr Opinion Biotechnol*, 2002, **13**: 72
- 9 Dutta J R, Dutta P K, Banerjee R. *Process Biochem*, 2004, **39**: 2193
- 10 Manohar B, Divakar S. *Process Biochem*, 2005, **40**: 3372
- 11 Singh A, Majumder A, Goyal A. *Bioresource Technol*, 2008, **99**: 8201
- 12 刘迎新, 未作君, 陈吉祥, 张继炎. 催化学报(Liu Y X, Wei Z J, Chen J X, Zhang J Y. *Chin J Catal*), 2005, **26**: 20
- 13 Corma A, Serra J M, Serna P, Valero S, Argente E, Botti V. *J Catal*, 2005, **229**: 513
- 14 Corma A, Serra J M, Serna P, Moliner M. *J Catal*, 2005, **232**: 335
- 15 韩立群. 人工神经网络理论、设计及应用. 北京: 化学工业出版社(Han L Q. *Theory, Design and Application of Artificial Neural Networks*. Beijing: Chem Ind Press), 2007. 38
- 16 计红果, 庞浩, 张容丽, 廖兵. 化学通报(Ji H G, Pan H, Zhang R L, Liao B. *Chemistry*), 2008: 329