

用模式识别法优化微生物培养基的营养条件

陈 敏 刘万卉 王静馨

(烟台大学,烟台 264005)

摘要 应用模式识别方法,对微生物培养的营养条件进行优化。以培养基组成构筑模式空间,通过主成份分析(PCA)揭示模式空间的可视优化区,建立模式分类器并逆推回到高维空间得到最佳培养基组成。以锌酵母为例,介绍优化全过程。

关键词 培养基,优化,模式识别,锌酵母

研究微生物生长的营养和环境条件,已有许多成熟的方法。最常用的是用于机理研究的单因素法和多因素同时研究的实验设计法。模式识别在该领域中的应用尚不多见。

模式识别方法,是从空间区域划分和属性类别判断角度出发,处理多元数据的一种非函数方法。该方法用一组表示被研究对象特征的变量构成模式空间,按“物以类聚”的观点分析数据结构,划分出具有特定属性模式类别的空间聚集区域,并辨认每一个模式的类别。由计算机按模式识别原理处理数据信息,作出最优决策。

本文研究模式识别在微生物培养的营养条件优化中的应用。以锌酵母培养为例,说明优化全过程。本方法以发酵培养基为特征变量构筑模式空间,以糖转化率和含锌量为评价目标,将样本划分为优与非优两类。通过 PCA 线性映射,将高维模式空间降至二维可视区,求出线性分类器。再以优化拟合方法从二维逆推回原高维空间,得到优化培养基条件。

1 材料与方法

1.1 模式识别优化的原理及模型

模式识别应用于微生物生长培养基优化的基本出发点是以培养基组成为特征变量构筑模式空间,一种培养基组合即一个样本,对应于模式空间中的一个点,称为样本点。以优化目标作为划分样本类别属性的依据,将目标值分为

若干档次,每一档次对应一种类别。培养基组成往往是多变量的,因此构成的模式空间是高维的,这时分类的模式无法显示出来。运用降维的方法,将高维空间映射到可视的二维或三维空间中,确定各类聚集的区域,再以适应的数学方法,将优类逆推回原高维空间而得到优化的培养基组成。

主成份分析(PCA)^[1]是模式识别中广泛应用的一种线性映射方法,实现降维和面向显示分析,得到模式空间的可视优化区。特征值和特征向量的求解有多种方法,本文采用幂法^[2]。低维映射空间(二维或三维)中的一个点,可对应高维特征空间中的无限多样本点(培养基组成的一种组合)。如何从低维优化区回复到高维特征空间,从而得到最优培养基组成,是模式识别用于优化需要解决的重要问题。本文研究了多因素坐标轮换优化法^[3]用于从低维到高维原始空间的回复,取得了满意的结果。

1.2 菌种

啤酒酵母 (*Saccharomyces cerevisiae*) CY26,由上海酵母厂烟台分厂提供。

1.3 培养基

斜面、种子培养基为麦芽汁培养基(含糖 9.5%, pH5.2)。发酵培养基由糖化液添加硫酸锌、硫酸铵、硫酸钾、磷酸、硫酸镁组成,其含量

按表 1 组合, pH 调至 4.6, 共 27 组试验组合, 每组试验重复 3 次。

1.4 培养方法

1.4.1 摇瓶培养: 取一环菌接种到含 9ml 种子培养基的试管中, 30℃ 静止培养 24h, 取其 1ml 接至内装 10ml 种子培养基试管中, 30℃ 静止培养 24h 后, 转入 100ml 发酵培养基(组成及浓度按试验设计变化, pH 均为 4.6, 含 1.8ml 新鲜玉米浆)的 250ml 三角摇瓶中, 于恒温培养箱内 30℃ 振荡培养 12h, 每 2h 补充一次培养液。

1.4.2 6L 玻璃发酵罐流加培养: 采用上海制药机械四厂生产 6L 玻璃发酵罐, 发酵培养基为经摇瓶培养、模式识别法优化的组成, 初始装量 2.95L。流加培养基(糖浓度 10%, 其它同发酵培养基)1.05L, 流加时间为发酵第 2 至第 8h。搅拌速度 300r/min, 通气量 5L/min, 培养温度 30℃, 总发酵时间为 12h。

1.5 分析方法

菌体干重用发酵液以 4000r/min 离心, 105℃ 烘干恒重后称重。

还原糖用斐林试剂还原法测定。

细胞含锌量用原子吸收光谱法^[4]测定。

2 结果与讨论

2.1 变量选择及数据预处理

选择培养基(%)中的葡萄糖、硫酸铵、硫酸锌、硫酸钾、硫酸镁、磷酸为待优化控制变量, 其试验设计组合如表 1。为消除数据差异的影响, 对原始数据 $\{X'_{ij}\}$ 进行如下标准化处理:

$$X_{ij} = \frac{X'_{ij} - \bar{X}_i}{S_i} \quad \bar{X}_i = \sum_{j=1}^n X'_{ij} / n,$$

$$S_i = \frac{1}{n} \left[\sum_{j=1}^n (X'_{ij} - \bar{X}_i)^2 \right]^{1/2}$$

式中, i 为变量数, $i=1, \dots, m$; j 为试验次数, $j=1, \dots, n$ 。

2.2 评判标准

模式识别是以一定的评判标准作为分类的依据。本文以锌酵母培养为课题, 根据儿童及成人对锌的日摄入需求量, 确定锌酵母最佳锌

含量在 0.5%~0.8%, 锌酵母培养的最佳目标即为满足锌含量基础上提高糖转化率。本文以细胞中锌含量和糖转化率双目标作为评判标准, 将样本分为两类: 细胞含锌量在 0.5%~0.8%、糖转化率>30%为第一类样本, 即优类样本; 其余均为第二类样本, 差类样本, 其中包括细胞含锌量和糖转化率均不符合一类指标及其中一项不符合的样本。

2.3 模式分析

应用 PCA 方法后, 一般问题能在两维或三维主成份空间中形成明显分类趋势, 但本文对 6 个特征变量进行 PCA 映射后, 两维主成份空间没有形成趋势。经研究发现, 葡萄糖含量与糖转化率之间, 在试验设定区间内, 有抛物线关系的倾向, 由此, 增加葡萄糖含量的平方项为一个独立的特征变量, 这样特征变量增至 7 个, 再进行 PCA 分析, 在主成份 Y_1 - Y_2 平面上得到了分类趋势。

以第一、二主成份 y_1 、 y_2 为坐标轴构成一个模式空间, 二维映射如图 1 所示。主成份分析列于表 2。

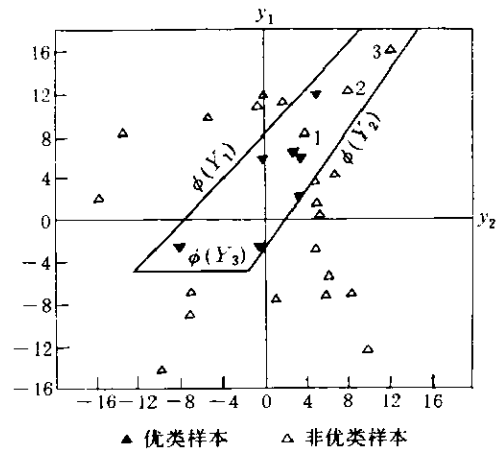


图 1 主成份分析 y_1 - y_2 映射图

从图 1 可以看出分类趋势, 优类样本集中在坐标中上方, 在坐标下方、左上及右下为非优类样本。为在 y_1 - y_2 平面上建立优类与非优类间的线性分类器, 将非优类样本划分为三个部分: 左上方六个样本点, 下部三个样本点及右下十个样本点, 分别与优类样本按重心距离分

表 1 试验设计及结果

No	C ₆ H ₁₂ O ₆ (%)	(NH ₄) ₂ SO ₄ (%)	ZnSO ₄ (%)	K ₂ SO ₄ (%)	H ₃ PO ₄ (%)	MgSO ₄ (%)	CRS (%)	Zn (%)
1	3	0.3	1.6	0.4	0.1	0.05	32.13	0.503
2	3	0.3	0.1	0.5	0.2	0.1	32.27	0.137
3	3	0.3	0.8	0.7	0.4	0.3	29.93	0.630
4	3	0.6	0.1	0.4	0.2	0.3	30.20	0.064
5	3	0.6	0.8	0.5	0.4	0.05	33.00	0.555
6	3	0.6	1.6	0.7	0.1	0.1	30.90	0.690
7	3	1.2	0.8	0.4	0.4	0.1	29.60	0.536
8	3	1.2	1.6	0.5	0.1	0.3	29.50	0.615
9	3	1.2	0.1	0.7	0.2	0.05	29.40	0.109
10	4	0.3	0.8	0.4	0.1	0.05	33.60	0.504
11	4	0.3	1.6	0.5	0.2	0.1	34.10	0.687
12	4	0.3	0.1	0.7	0.4	0.3	33.22	0.187
13	4	0.6	1.6	0.4	0.2	0.3	32.20	0.722
14	4	0.6	0.1	0.5	0.4	0.05	32.48	0.286
15	4	0.6	0.8	0.7	0.1	0.1	33.55	0.601
16	4	1.2	0.8	0.4	0.4	0.1	28.95	0.558
17	4	1.2	1.6	0.5	0.1	0.3	29.72	0.656
18	4	1.2	0.1	0.7	0.2	0.05	23.95	0.056
19	5	0.3	1.6	0.4	0.1	0.05	25.66	0.713
20	5	0.3	0.1	0.5	0.2	0.1	21.36	0.263
21	5	0.3	0.8	0.7	0.4	0.3	24.30	0.612
22	5	0.6	0.1	0.4	0.2	0.3	24.84	0.044
23	5	0.6	0.8	0.5	0.4	0.05	21.50	0.575
24	5	0.6	1.6	0.7	0.1	0.1	25.88	0.636
25	5	1.2	0.8	0.4	0.4	0.1	23.14	0.540
26	5	1.2	1.6	0.5	0.1	0.3	24.30	0.683
27	5	1.2	0.1	0.7	0.2	0.05	23.64	0.168

CRS: 糖转化率 Zn: 锌含量

表 2 主成份分析表

No.	特征值	特征向量						
		X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
1	λ ₁ = 63.6198	-0.4430	-3.0598	1.0394	3.5852	2.9725	-4.5381	3.2707
2	λ ₂ = 40.2386	-4.5066	-0.3607	-2.6156	0.9466	0.9292	-0.8804	-3.2285
3	λ ₃ = 31.0503	0.8249	-3.3229	0.6053	-2.2226	3.0975	1.5325	-1.4418
4	λ ₄ = 23.3821	-1.9057	-0.3693	4.1092	-0.3256	-1.2315	-0.3711	-0.9735
5	λ ₅ = 12.5857	0.9988	-1.4820	-0.0526	2.3136	-1.2507	1.0638	-1.1577
6	λ ₆ = 10.6027	0.2551	1.7612	0.9073	1.3584	1.9041	0.8940	-0.5853
7	λ ₇ = 0.5208	-0.3460	-0.1036	-0.0230	0.0795	0.0058	0.4621	0.4123

X₁ = C₆H₁₂O₆, X₂ = (NH₄)₂SO₄, X₃ = ZnSO₄, X₄ = K₂SO₄, X₅ = MgSO₄, X₆ = H₃PO₄, X₇ = (C₆H₁₂O₆)²

类法建立分段线性分类器^[5]。以左上非优类与优类样本为例,说明建立线性分类器方法。首先分别求出不同类别样本的重心坐标:

优类重心坐标:

$$C_{11} = -5.394, C_{12} = 9.445, C_1^T = \begin{bmatrix} C_{11} \\ C_{12} \end{bmatrix}$$

差类重心坐标:

$$C_{21} = 0.679, C_{22} = 4.245, C_2^T = \begin{bmatrix} C_{21} \\ C_{22} \end{bmatrix}$$

采用线性判别函数 Φ(Y₁), 则有:

$$\Phi(Y_1) = (C_1 - C_2)y - 0.5(|C_1|^2 - |C_2|^2)$$

令 Φ(Y₁) = 0, 就是一个判别面, 即是:

$$\Phi(Y_1) = -0.6073y_1 + 5.190y_2 - 49.869 = 0$$

用同样方法求出优类与右下非优类样本的线性分类器为:

$$\Phi(Y_2) = 5.857y_1 - 9.213y_2 - 24.433 = 0$$

下部三个非优类样本与优类的分类器目视为:

$$\Phi(Y_3) = y_2 + 5 = 0$$

从图 1 看出, 优类样本集中在三条线性分类器所夹的区域, 并有向右上方延伸的趋势。

对于一个未知类别的样本点 [X]^T, 经数据预处理后求得对应的 $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, 然后用判别函数 Φ(Y) 判别其属性。当 Φ(Y₁) > 0、Φ(Y₂)

< 0、Φ(Y₃) > 0 时, 该样本点属于优类; 否则属于非优类。

2.4 确定优化的培养基组成

由图 1 已清楚显示了优类样本点在 y₁-y₂ 主成份空间中的区域和趋势。为此在优类样本区域内向右上外延方向设定 Δ₁、Δ₂、Δ₃ 三个点。为求解这三个点所对应的培养基组成, 需将二维平面上的点回复到多维原始空间。本文研究应用了多因素轮换优化法^[3], 取得满意的结果。以最小二乘目标函数 Q = [(y₁' - y₁)² + (y₂' - y₂)²]^{1/2} 为最小, 优化求解得到培养基组成。式中 y₁、y₂ 为 Δ₁、Δ₂、Δ₃ 各点在 y₁-y₂ 空间的坐标值; y₁'、y₂' 为设定 X₁~X₇ 值后按主成份分析的计算值。优化结果列于表 3。表 3 中每点中的第一行数据为标准化数据, 第二行为对应的原始数据。

根据上述模拟计算的优化培养基组成, 在 6L 玻璃发酵罐上扩大培养, 细胞含锌量均在优类 0.5%~0.8% 范围内, 糖转化率有明显提高。具体结果已列于表 3 之中。

模式识别调优技术, 为微生物培养的营养条件选择提供了新的途径。特别在十分复杂的情况下, 如优化标准表现为多目标(例如本文要求的 Zn 含量在一定区间和糖转化率尽量高的双目标), 多目标间甚至互相矛盾, 这时应用单因素试验或试验设计方法不仅要进行大量试验, 而且由于目标函数的矛盾可能很难进行; 此外, 当目标函数与营养条件的变量间回归相关

表3 模拟计算的优化参数及试验结果

选择点	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Q	CRS	Zn
Δ_1	-1.42	0.62	1.38	-0.03	1.30	-0.92	-1.31	0.637	41.8	0.705
	2.82	0.94	1.60	0.53	0.40	0.05	7.94			
Δ_2	-1.70	-1.05	1.10	-0.05	1.31	-0.91	-1.50	0.004	42.2	0.690
	2.59	0.30	1.40	0.52	0.40	0.05	6.68			
Δ_3	-2.30	-1.05	1.38	1.20	1.31	-0.91	-1.84	0.063	40.4	0.744
	2.09	0.30	1.60	0.69	0.40	0.05	4.35			

* $X_1 \sim X_7$ 注同表2, "Q"优化计算的目标函数。

关系不显著时,单因素和试验设计方法往往得不到结论。而采用本文提出的模式识别方法,处理多目标问题简便,可在可视模式空间中确定优化区域,易于处理相关不显著问题,同时还可外延至试验范围外的区域。

参 考 文 献

[1] 方开泰. 实用多元统计分析, 上海: 华东师范大学出版

社, 1989, 291~293.

[2] 清华大学应用数学系. 现代应用数学手册计算方法分册, 北京: 北京出版社, 1990, 447~449.

[3] 中科院数学研究所优选法组. 优选法, 北京: 科学出版社, 1978, 45.

[4] 黄伟坤. 食品检验与分析, 北京: 轻工业出版社, 1989, 208~209.

[5] 王碧泉, 陈祖荫. 模式识别, 北京: 地震出版社, 1989, 89.

OPTIMIZATION OF NUTRITIONAL CONDITIONS OF A MICROBE CULTURE BY A PATTERN RECOGNITION METHOD

Chen Min Liu Wanhui Wang Jingxing

(Yantai University, Yantai 264005)

Abstract Nutritional conditions of a microbe culture were optimized by the pattern recognition method. The medium compositions were used as feature variables to construct the pattern space, on where the visual optimization region were found by using PCA method, it was trasformed into the original data space and the optimum medium composition were got. This method were described with zinc yeast culture.

Key words Medium, Optimization, Patternrecognition, Zinc yeast