

# 贝叶斯分类算法应用于回转窑烧结温度预测模型<sup>①</sup>

丁钢坚, 张小刚

(湖南大学 电气与信息工程学院, 长沙 410082)

**摘要:** 针对氧化铝回转窑烧结过程中存在的多变量、强耦合、非线性、环境恶劣、煤粉和料浆成分不稳定等因素造成的烧结带温度不稳定, 熟料质量低等难题, 提出一种基于朴素贝叶斯分类(NBC)算法的回转窑烧结温度预测模型。采用 FastICA 算法, 找到回转窑热工数据的独立成分, 从而满足 NBC 的属性之间相互独立的条件, 并通过 AdaBoost 算法对模型进行提升。实验结果表明, 这种模型具有较好的控制效果。

**关键词:** 朴素贝叶斯; 预测模型; 回转窑; ICA; AdaBoost

## Prediction Model of Bayesian Classification Algorithm Applied to Control Sintering Temperature in Rotary Kiln

DING Gang-Jian, ZHANG Xiao-Gang

(College of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

**Abstract:** The sintering temperature in rotary kiln is usually hard to be stable because of complex industrial environment. In this paper, we present a prediction model based on Bayesian classification algorithm to predict the trend of the amount of feed coal for controlling the sintering temperature in rotary kiln. To avoid the influence of attribute independence assumption of Bayesian classification algorithm, the FastICA algorithm used to find the independent components of the working condition data set in rotary kiln. Then we use AdaBoost algorithm to find a best classifier. The final simulation results show that the model has better control performance.

**Key words:** naive bayes; prediction model; rotary kiln; ICA; AdaBoost

在建材、冶金等许多生产行业中, 回转窑被广泛地用于对物料进行物理或化学处理。全国各类回转窑约有两千台左右。回转窑的最大问题是能耗高、热效率低, 是我国工业中重要耗能设备。虽然造成回转窑能耗高、效率低的原因是多方面的, 但是其中的一个重要因素是回转窑工作过程中的检测与控制手段还很落后。而烧结带温度的检测与控制, 是保证氧化铝熟料质量, 提高产能, 降低功耗的关键任务。

由于回转窑控制情况复杂; 工况条件非常恶劣, 工况数据带有许多不确定的噪声; 烧结过程特性复杂, 一些关键工况参数如: 喂煤量、窑体转数、料浆流量、鼓风机流量、窑内各带温度等耦合严重。所以, 常规的线性数学模型很难解决回转窑的过程控制问题。文献[1]采用计算机图像处理的方法将工业视频摄像机摄取的回转窑窑内煅烧火焰图像转换成火焰温度, 然后

将火焰温度作为主要参数来构建RBF网络预测控制模型, 改变了原始的人工看火操作方法, 实现了正常工况下对烧结温度的稳定控制。然而由于粉尘、烟雾以及灯光等因素, 摄像机采集的图像质量不能时刻得到保证。文献[2]提出基于多传感器数据融合的模糊专家控制模型, 并成功用于实践。为了进一步提高回转窑烧结过程中熟料的质量, 稳定烧结带的温度, 综合回转窑正常工况数据的特点, 本文提出了一种统计模式识别<sup>[3]</sup>中的贝叶斯分类算法的回转窑烧结温度预测模型。

## 1 基于ICA的朴素贝叶斯分类器优化模型

### 1.1 问题描述

贝叶斯分类模型是一种典型的基于统计方法的分类模型。贝叶斯学习方法的理论基础是贝叶斯定理,

<sup>①</sup> 基金项目:国家自然科学基金(60874096,50704016)

收稿时间:2011-01-22;收到修改稿时间:2011-03-04

它将事件的先验概率与后验概率巧妙地联系起来,充分利用先验信息和样本数据信息确定事件的后验概率。贝叶斯分类器主要分为两种:朴素贝叶斯分类器<sup>[4]</sup>和贝叶斯网络分类器。作为一种最简单、有效的而且在实际应用中很成功的分类器,朴素贝叶斯分类器具有坚实的理论基础和较优秀的分类性能,其性能与神经网络、决策树相当,实际应用中降低了贝叶斯网络构建的复杂性。结合回转窑的工艺特点,作者认为朴素贝叶斯分类器可以较好的用于氧化铝的烧结温度控制当中。由于回转窑工艺复杂的特性,难以用实验的方法得到样本数据,本文使用在正常的工况下采集的回转窑的热工数据作为特征属性和分类属性,通过对参数的变化趋势进行离散化,然后在此数据集上学习贝叶斯分类器模型。由当前的数据,结合先验数据信息,来预测下一时刻回转窑给煤量的变化趋势,以此调节进给燃料,稳定烧结带的温度。

## 1.2 构建贝叶斯分类器模型

离散后的回转窑数据集有 8 个特征属性  $A_1, A_2, \dots, A_8$ , 每个属性的取值为:上升、不变和下降。用特征向量  $X = (x_1, x_2, \dots, x_8)$  表示回转窑数据集  $D$  中一个样本的 8 个属性的具体取值。朴素贝叶斯分类模型将训练实例分解成特征向量  $X$  和决策类标变量  $C$ 。

贝叶斯分类器的预测依据就是取后验概率:

$$P(C_i | X) = \frac{P(C_i)P(X | C_i)}{P(X)} \quad (1)$$

$$1 < i < 3$$

最大的类别。记  $C(X)$  为  $X$  所属的分类标签。则朴素贝叶斯分类器模型为:

$$C(X) = \arg \max_{C_i} P(C_i)P(X | C_i) \quad (2)$$

假设属性  $x_k$  之间相互独立,则:

$$P(X | C_i) = \prod_{k=1}^8 P(x_k | C_i) \quad (3)$$

式(2)可表示为:

$$C(X) = \arg \max_{C_i} P(C_i) \prod_{k=1}^8 P(x_k | C_i) \quad (4)$$

其中,先验概率  $P(C_i) = d_i / d$ , 条件概率  $P(x_k | C_i) = d_{ik} / d_i$ , 其中是属于类  $C_i$  的训练样本数,  $d$  是训练样本总数。 $d_{ik}$  是训练样本中属于类  $C_i$  且属性  $A_k$  取值为  $x_k$  的样本个数。

朴素贝叶斯分类模型假定特征向量的分量间相对于决策变量时相对独立的,也就是各分量独立的作用

于决策变量。但是在实际应用中,例如:窑头温度和窑尾温度可能存在一定程度上的线性相关,这种属性的独立性假设可能并不能完全成立,从而导致分类精确度不高。因此有必要对属性进行优化。

## 1.3 属性优化

为了满足朴素贝叶斯分类器所要求的独立性假设,通过独立成分分析 ICA<sup>[5]</sup>计算得到的分离矩阵  $W$ , 可以将数据集映射到另一个多维特征空间,使得各分量相互独立。ICA 的基本思想是将观察到的数据进行某种线性分解,使其分解成统计独立的成分。设  $x$  为观测值,对应于一个  $n$  维离散时间信  $X = [x_1, x_2, \dots, x_n]^T$ , 是源信  $S = [s_1, s_2, \dots, s_m]^T$  号的线性组合

$$X = AS \quad (5)$$

式(5)称为独立成分分析模型。ICA 的任务就是找到这样的矩阵  $W$ , 使得  $Y = WX$ , 其中  $Y = [y_1, y_2, \dots, y_m]^T$ ,  $W$  是  $A$  的逆矩阵。如果能保证  $y_i$  之间相互独立,可以认为  $y_i$  近似等于  $s_i$ 。

信号的独立性是通过它的非高斯性程度来进行度量的,文献[6]介绍的多种准则。本文基于负熵度量准则,采用 Aapo Hyvärinen 等人提出的 FastICA<sup>[7]</sup>算法来计算分离矩阵  $W$ 。

FastICA 算法步骤:

将训练样本的属性作为行向量,构成观测数据矩阵  $X$ :

- 1) 中心化观测数据  $X$ , 使其均值为 0
- 2) 再对数据白化处理,  $X \rightarrow \tilde{X}$ ; 选择要估计的分量个数  $m$
- 3) 选择选择初始随机权矢量  $W_p$ ; 设迭代次数  $v \leftarrow 1$
- 4) 令  $W_p = E\{\tilde{X} g(W_p^T \tilde{X})\} - E\{g(W_p^T \tilde{X})\}W_p$ , 非线性函数  $g$  通常取  $g(u) = u \exp(-u^2 / 2)$
- 5)  $W_p = W_p - \sum_{j=1}^{p-1} (W_p^T W_j) W_j$
- 6) 令  $W_p = W_p / \|W_p\|$
- 7) 如果  $W_p$  不收敛, 则返回第 4) 步
- 8) 令  $p = p + 1$ , 如果,  $p < m$ , 返回第 5) 步

## 1.4 对单个模型的提升

提升方法<sup>[8,9]</sup>(boosting)是一种多分类器集成学习算法,算法的目标是提高任何给定算法的分类准确率。它的思想是学习一系列分类器,在这个序列中每一个

分类器对它前一个分类器导致的错误分类例子给予更大的权重。在学习完分类器  $H_k$  之后，增加了由  $H_k$  导致分类错误的训练例子的权重，并且通过重新对训练例子计算权值，再学习下一个分类器  $H_{k+1}$ 。重复这个过程  $T$  次。最终的分类器从这一系列的分类器中综合得出。本文采用由 Freund 和 Schapire 提出的 AdaBoost<sup>[10]</sup>(Adaptive Boosting)方法对多个朴素贝叶斯分类器进行提升。算法步骤如下：

$N$  个训练实例： $\langle (X_1, C_1), \dots, (X_N, C_N) \rangle$

$N$  个训练实例上的分布  $D:W$ ， $W$  为训练实例的权向量

$T$  为训练重复的次数

- 1) 初始化
- 2) 初始化训练实例的权向量。 $W_i = 1/N, i = 1, \dots, N$
- 3) for  $t = 1$  to  $T$
- 4) 给定权值  $w_i^t$  得到一个假设  $H^{(t)}: X \rightarrow C$
- 5) 估计假设  $H^{(t)}$  的总体误差

$$e^{(t)} = \sum_{i=1}^N w_i^{(t)} I(C_i \neq h_i^{(t)}(x_i)),$$

- 6) 计算  $\beta^{(t)} = e^{(t)} / (1 - e^{(t)})$
- 7) 计算下一轮样本的权值

$$w_i^{(t+1)} = w_i^{(t)} (\beta^{(t)})^{1 - I(C_i = h_i^{(t)}(x_i))}$$

- 8) 正规化  $w_i^{(t+1)}$ ，使其总和为 1
- 9) end for
- 10) 输出：

$$h(x) = \arg \min_{c \in C} \sum_{t=1}^T (\log \frac{1}{\beta^{(t)}}) I(h^t(x) = c)$$

其中，如果  $\phi = T, I(\phi) = 1$ ；否则  $I(\phi) = 0$ 。

## 2 模型学习步骤及框图

分类器模型的学习步骤如下：

**Step 1:** 首先，对原始数据集  $D$  进行均值滤波和中值滤波，得到训练数据集  $D_t$  和验证数据集  $D_v$ ；

**Step 2:** 对训练集和验证集进行一次曲线拟合离散，得到各数据的变化趋势离散值；

**Step 3:** 在离散训练集上使用 FastICA 算法，求得分离矩阵  $W$ ；

**Step 4:** 用分离矩阵  $W$  将离散数据集变换到另一个特征空间，得到训练集  $D_{ti}$  和验证集  $D_{vi}$ ；

**Step 5:** 在训练集  $D_{ti}$  上学习贝叶斯分类器参数。

基于贝叶斯分类算法的回转窑烧结温度预测模型总体框图如图 1 所示。其中的分类器是提升算法得出的最优分类器，其输出为下一时刻给燃料的变化趋势，即增煤、减煤和煤量不变，从而达到稳定回转窑烧结温度的目的。

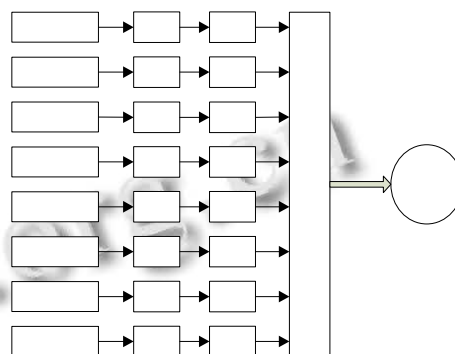


图 1 模型总体框图

## 3 实验仿真结果

实验是使用 matlab7.0, 在一台内存容量为 1G, CPU 为英特尔双核 2.0GHz 的计算机进行的, 实验数据库是贵州铝厂某台窑一个月内正常运行的热工数据。

作者使用文章所述方法为自动看火系统设计自动给煤预测控制器，使用了以下几个热工参数：窑炉转速、窑头温度、火焰温度、窑尾温度、料浆流量、鼓风机流量、主机负荷、冷却机负荷和给煤量。数据采集周期是 3 分钟，参数经过中值滤波和均值滤波。当前时刻的数据和 7 个历史时刻的数据，经过曲线拟合离散化后构成一个样本。滑窗移动 3 个采集周期对数据进行曲线拟合离散化。实验中发现曲线斜率阈值为 0.001，独立分量个数为 6，提升算法迭代次数只要 9 次，分类器的分类效果即可达到最佳，为 79.12%。

表 1 给煤量变化趋势预测精度表

样本数	离散斜率 阈值	NBC	ICANBC	ICANBCAda
4582	0.0012	68.33	70.06	75.17
4582	0.0011	70.05	71.78	76.89
4582	0.001	73.24	74.23	79.12
4582	0.0009	71.49	72.16	77.26
4582	0.0008	69.87	70.32	75.43

表 1 是最佳独立量个数和迭代次数下，不同的斜率阈值下离散后的数据集对应的预测精度值。NBC 代表朴素贝叶斯分类器的预测精度值，ICANBC 代表用

分离矩阵将数据集转换后的分类器预测精度值, ICANBCAda 代表在多个 ICANBC 上提升后的最佳分类器的预测精度值。在最佳离散斜率阈值时, 图 2 为提升算法迭代次数为 9 时, 不同独立分量个数对应的预测精度曲线图, 图 3 是当独立分量个数为 6 时, 不同迭代次数对应的预测精度曲线图。

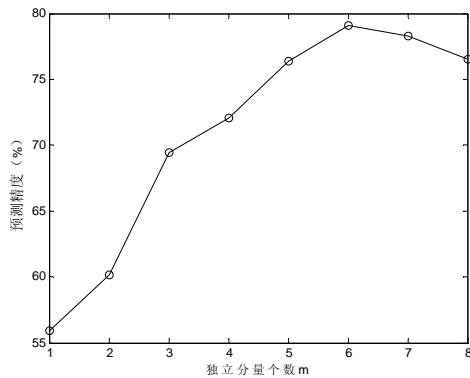


图 2 独立分量个数  $m$  与预测精度关系图

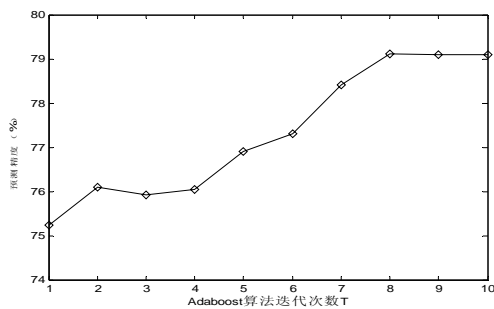


图 3 adaboost 算法迭代次数  $T$  与预测精度关系图

#### 4 结语

从贝叶斯分类算法中可以看出, 即使样本中出现孤立的噪声点, 由于分类计算能够将这些噪声点平均分配, 朴素贝叶斯分类器仍能保持较好的精确度。本文提出的预测模型对窑炉烧结温度的稳定效果接近于神经网络预测控制器的效果, 优于只含有单个朴素贝

叶斯分类器的控制效果。试验中发现 ICA 算法对分类精度的提高并不明显, 而且 AdaBoost 算法对样本中的噪声点比较敏感, 所以分类器预测的精度比较依赖于样本质量以及数据的离散化细节。作者下一步的研究是对回转窑数据构建贝叶斯网络, 继续探索更好的模型, 实现回转窑氧化铝烧结过程的最优控制。

#### 参考文献

- 1 张小刚, 陈华, 章兢, 刘小燕. 基于图像反馈的回转窑烧结温度智能预测控制. 控制理论与应用, 2007, 24(6): 995-998.
- 2 张小刚, 陈华, 章兢. 基于多传感器数据融合的回转窑烧结温度检测和控制方法. 控制与决策, 2002, 17(6): 867-870, 875.
- 3 Jain AK, Duin RPW, Mao JC. Statistical pattern recognition: A review. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000, 22(1): 4-37.
- 4 林士敏, 田凤占, 陆玉昌. 用于数据挖掘的贝叶斯分类器研究. 计算机科学, 2000, 27(10): 73-76.
- 5 Hyv RA, Karhunen J, Oja E. Independent component analysis. A Wiley-Interscience Publication. John Wiley & Sons, 2001.
- 6 Hyv RA. Independent component analysis: algorithms and applications. Neural Networks, 2000, 13(4-5): 411-430.
- 7 Hyv RA. Fast and robust fixed-point algorithm for independent component analysis. IEEE Trans. on Neural Networks, 1999, 10(3): 626-634.
- 8 Schapire RE. The Strength of Weak Learnability. Machine Learning, 1990, 5(2): 197-227.
- 9 Freund Y. Boosting a Weak Learning Algorithm by Majority. Information and Computation, 1995, 121(02).
- 10 Freund Y, Schapire RE. A Decision Theoretic Generalization of On-Line Learning and an Application to Boosting. Computational Learning Theory, 1997, 904/1995: 23-27.