



(12) 发明专利

(10) 授权公告号 CN 108614536 B

(45) 授权公告日 2020.10.27

(21) 申请号 201810597191.2	CN 102013039 A, 2011.04.13
(22) 申请日 2018.06.11	CN 107872354 A, 2018.04.03
(65) 同一申请的已公布的文献号 申请公布号 CN 108614536 A	CN 104704499 A, 2015.06.10
(43) 申请公布日 2018.10.02	CN 107391961 A, 2017.11.24
(73) 专利权人 云南中烟工业有限责任公司 地址 650231 云南省昆明市红锦路367号	CN 107806690 A, 2018.03.16
(72) 发明人 周冰 曾仲大 唐军 文里梁 何邦华 陈文 崔亚娟	CN 106940758 A, 2017.07.11
(74) 专利代理机构 昆明正原专利商标代理有限公司 53100 代理人 金耀生 亢能	CN 104092609 A, 2014.10.08
(51) Int. Cl. G05B 19/418 (2006.01)	CN 101110095 A, 2008.01.23
(56) 对比文件	CN 104572985 A, 2015.04.29
CN 105006119 A, 2015.10.28	CN 107103160 A, 2017.08.29
CN 103313288 A, 2013.09.18	CN 106203633 A, 2016.12.07
CN 105184373 A, 2015.12.23	CN 107729440 A, 2018.02.23
CN 105740952 A, 2016.07.06	CN 101965566 A, 2011.02.02
CN 101256641 A, 2008.09.03	US 2005132051 A1, 2005.06.16
CN 108062360 A, 2018.05.22	US 2002147937 A1, 2002.10.10
CN 107395440 A, 2017.11.24	US 9904961 B2, 2018.02.27
	US 2005278437 A1, 2005.12.15

李硕豪.“贝叶斯网络结构学习综述”.《计算机应用研究》.2015,第641-646页.

钟文焱.“基于多因素分析的烘丝机人口含水率预测模型的建立与应用”.《烟草科技》.2015,

(续)

审查员 叶双清

权利要求书2页 说明书8页 附图3页

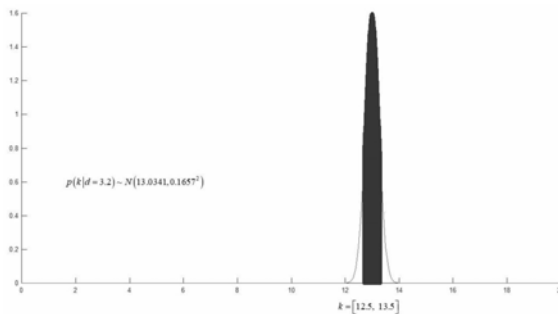
(54) 发明名称

一种卷烟制丝工艺关键因素的复杂网络构建方法

(57) 摘要

本发明公开了一种卷烟制丝工艺关键因素的复杂网络构建方法,属于烟草领域。该方法首先,基于不同属性数据组成复杂网络的节点集合;其次,采用协方差方法计算节点间相关性,确定网络路径,构建初步的复杂网络结构;然后,采用最大最小爬山算法,实现复杂网络结构优化;最后,采用最大似然估计法,确定复杂网络参数。根据上述方法所构建的复杂网络,可以系统性地解析制丝过程中工艺参数与工艺质量指标的关系;更可以全局性的分析制丝过程中关键工艺参

数及工艺质量指标对卷烟质量的影响,为制丝工艺标准的制修订提供数据支撑,进而保障卷烟质量稳定性。



CN 108614536 B

[接上页]

**(56) 对比文件**

李金艳.“信息不完备情况下多因素工序质量诊断方法”.《哈尔滨工业大学学报》.2016,第88-93页.

李秀芳.“烟片松散回潮关键工艺参数过程控制系统的优化设计”.《中国烟草学报》.2015,

裴平.“基于贝叶斯网络的P2P网贷借款人信用评价模型”.《中国经济问题》.2017,

YANG L.“Bayesian belief network-based approach for diagnostic and prognostic of semiconductor manufacturing systems”.《Robotic and Computer-Integrated Manufacturing》.2012,

1. 一种卷烟制丝工艺关键因素的复杂网络构建方法,其特征在于:包括以下步骤:

步骤(1)、构建复杂网络的节点集合

根据卷烟制丝线工艺流程,筛选制丝数据产生节点,采集数据,构建复杂网络的节点集合;

步骤(2)、数据预处理

进行数据预处理,去除制丝工艺不稳定时采集的数据;

步骤(3)、初步复杂网络构建

根据预处理后的数据,计算节点间的独立性,分析复杂网络节点集合中所有节点间的两两相关性,确定节点间是否有边存在,完成复杂网络边集合的确定,结合复杂网络节点集合,完成初步复杂网络构建;

步骤(4)、复杂网络结构优化

采用最大最小爬山法,对不同的复杂网络结构加入惩罚值计算网络得分,评分高的复杂网络结构优于评分低的复杂网络结构,去除掉不会增加整个网络信息的递进边,和确定网络结构中边的方向,加入惩罚值的贝叶斯网络得分计算公式如下:

$$Score_{BIC}(G:D) = Score_L(G:D) - \frac{\log M}{2} Dim[G];$$

其中,Score<sub>L</sub>(G:D)为基于信息论方法计算所得的网络得分, $\frac{\log M}{2} Dim[G]$ 为所加入的惩罚值,两者差值用于确定复杂网络结构中的边方向,和去除掉不增加网络整体信息量的递进边;

步骤(5)、计算复杂网络每条边的条件概率参数

针对复杂网络的每条边,采用最大似然估计法计算复杂网络概率参数,复杂网络的最大似然函数如下:

$$\begin{aligned} l_k(D:U,\theta) &= -\frac{1}{2} \sum_{m=1}^M \left[ \log(2\pi) + \log(\sigma^2) + \frac{1}{\sigma^2} (K[m] - g(u[m]))^2 \right] \\ &= -\frac{1}{2} \left[ M \log(2\pi) + M \log(\sigma^2) + \frac{1}{\sigma^2} \sum_{m=1}^M (K[m] - g(u[m]))^2 \right]; \end{aligned}$$

其中,U是其中一个节点的父节点集合, $\theta = \langle \beta_0, \beta_1, \beta_2, \beta_k, \delta \rangle$ , $g(u[m]) = (\beta_1 u_1, \dots, \beta_m u_m : \theta) = \sum_j \beta_j u_j + \theta_0$ 为线性映射,u[m]表示包含其中一个节点及其所有父节点数据的,第m个样本的实例,K[m]表示u[m]中某一个节点的值,K为a、b、...、p中的一个;

根据上式分别对 $\beta_0, \beta_1, \dots, \beta_k, \delta$ 求偏导,得到k+2个等式,求解方程组,得到该边的条件概率参数 $\beta_0, \beta_1, \dots, \beta_k, \delta$ ;

依据连续型贝叶斯网络的特性,若在连续型贝叶斯网络中,两节点的条件概率满足正态分布;

对复杂网络的每条边重复上述操作,计算得到复杂网络每条边的条件概率参数值。

2. 根据权利要求1所述的卷烟制丝工艺关键因素的复杂网络构建方法,其特征在于:步骤(1)中,相关数据产生节点包括制丝工艺在线监测节点、制丝过程烟丝质量离线检测节

点、制丝工艺调控记录节点、所有制丝生产环境记录节点和所有卷烟质量评价结果节点。

3. 根据权利要求1所述的卷烟制丝工艺关键因素的复杂网络构建方法,其特征在於:步骤(2)中,数据预处理的方法具体包括:

A、采用箱型图数据质量评价方法,采用数据的上四分位数、中位数、下四分位数,和上边缘、下边缘的统计数据,将箱型图外数据组作为误差数据去除;

B、采用移动窗口动态评价法,具体步骤如下:

步骤(1)、将步骤(1)的待处理的数据进行排序: $X = \{x_1, x_2, \dots, x_n\}$ ;

步骤(2)、设置数据序列X的中位值  $\bar{x} = x_{\lfloor \frac{n+1}{2} \rfloor}$ ;

步骤(3)、确定窗口大小: $l < \lfloor \frac{n+1}{2} \rfloor$ ,或根据实际情况调整窗口大小;

步骤(4)、在步骤(1)的数据序列 $X = \{x_1, x_2, \dots, x_n\}$ 中,截取长度为 $2l+1$ 的子数据序列

$\hat{X} = \left\{ x_{\lfloor \frac{n+1}{2} \rfloor - l}, \dots, x_{\lfloor \frac{n+1}{2} \rfloor}, \dots, x_{\lfloor \frac{n+1}{2} \rfloor + l} \right\}$ ,计算的 $\hat{X}$ 平均值mean,和标准偏差sd;

步骤(5)、对于数据序列 $X = \{x_1, x_2, \dots, x_n\}$ 的每个数值进行处理,若 $\text{mean} - n \times \text{sd} \leq x_i \leq \text{mean} + n \times \text{sd}$ ,其中,n为根据实际情况设置的常数,则 $x_i$ 正常值;否则, $x_i$ 为奇异值,需删除。

4. 根据权利要求1所述的卷烟制丝工艺关键因素的复杂网络构建方法,其特征在於:步骤(3)中,节点间独立性的计算方法包括如下步骤:

步骤(1)、针对需要进行独立性检测的两个节点,节点a和节点b,分别采集n批制丝过程数据,其对应的数据分别为 $a_1, \dots, a_i, \dots, a_j, \dots, a_n$ 和 $b_1, \dots, b_i, \dots, b_j, \dots, b_n$ ,其中 $a_i$ 和 $b_i$ 均为列向量数据;

步骤(2)、构建矩阵A、B,矩阵元素计算公式如下:

$$A_{j,k} = ||a_j - a_k||,$$

其中, $a_j$ 表示节点a的第j个数据实例, $a_k$ 表示节点a的第k个数据实例, $A_{j,k}$ 表示节点a的第j个数据实例与第k个数据实例之间的距离,即欧几里得范数;

采用同样的处理方法,对节点b相应的采集数据进行处理,构建矩阵B;

步骤(3)、计算矩阵A和矩阵B的协方差,并根据该值计算节点a和节点b间独立性,

按如下公式计算:

$$dCov_n^2(X, Y) = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n A_{j,k} B_{j,k},$$

$dCov_n^2(X, Y)$ 即表示节点a和节点b的相关性;

步骤(4)、按照上述的方法,分析复杂网络节点集合中所有节点间的两两相关性,并按照5%置信水平,确定节点间是否有边存在,完成复杂网络边集合的确定,结合复杂网络节点集合,完成初步复杂网络构建。

## 一种卷烟制丝工艺关键因素的复杂网络构建方法

### 技术领域

[0001] 本发明涉及一种复杂网络构建方法,尤其是一种卷烟制丝工艺关键因素的复杂网络构建方法,属于烟草领域。

### 背景技术

[0002] 制丝是卷烟加工过程的核心部分,是决定卷烟加工质量好坏的关键环节之一,对提高卷烟产品综合品质起着关键作用。然而,制丝过程的加工工序多、工艺流程长、加工方式复杂,其对卷烟质量的影响是各制丝工艺关键因素及制丝环境因素相互作用、相互影响及相互调控的系统性综合结果。目前,制丝过程中的工艺技术研究很多,如加工方式中的松散回潮、叶丝干燥工序、如质量指标中的烟丝水分、烟丝结构等,均对卷烟质量有显著影响作用,但相关工艺技术研究均聚焦于某一关键工序或某重要质量指标,缺乏系统性的能纳入所有影响因素的方法。

### 发明内容

[0003] 为了解决上述技术问题,本发明采用系统科学与网络数据分析的方法,对制丝过程关键工序参数数据、工艺质量数据及卷烟质量数据构建复杂网络,提供了一种卷烟制丝工艺关键因素的复杂网络构建方法,对卷烟加工过程所采集到的所有制丝工艺在线监测数据、所有制丝过程烟丝质量离线检测数据、所有制丝工艺调控记录数据、所有制丝生产环境记录数据、所有卷烟质量评价结果数据等进行分析,构建一个层次递进的复杂网络,该复杂网络可用于分析制丝过程关键因素对卷烟质量的影响关系。本发明的具体方案如下:

[0004] 一种卷烟制丝工艺关键因素的复杂网络构建方法,包括以下步骤:

[0005] 步骤(1)、构建复杂网络的节点集合

[0006] 根据卷烟制丝线工艺流程,筛选制丝相关数据产生节点,采集数据,构建复杂网络的节点集合;

[0007] 步骤(2)、数据预处理

[0008] 进行数据预处理,去除制丝工艺不稳定时采集的MES数据;数据预处理的目的在于去除“脏数据”,即去除制丝工艺不稳定时采集的MES数据以及去除数据采集时,因环境噪声引入的数据误差。

[0009] 步骤(3)、初步复杂网络构建

[0010] 根据预处理后的数据,计算节点间的独立性,分析复杂网络节点集合中所有节点间的两两相关性,确定节点间是否有边存在,完成复杂网络边集合的确定,结合复杂网络节点集合,完成初步复杂网络构建;

[0011] 步骤(3)构建完成的初步复杂网络无法被直接用于制丝工艺关键因素对卷烟质量影响分析,其原因如下:

[0012] 1、初步复杂网络间的边,由两两节点间的相关系数得到,因为相关系数的相互性,无法确定复杂网络中的边方向。

[0013] 2、同样由于相关性的特点,导致初步的复杂网络结构中大量存在递进关系,例如:对于a,b,c三个节点,通过相关性分析,和边方向的确定,得到 $a \rightarrow b$ , $b \rightarrow c$ 和 $a \rightarrow c$ 三条有向边,在卷烟制丝工艺关键因素影响卷烟质量的分析过程中,边 $a \rightarrow c$ 并不增加额外的有用信息,反而会增加网络复杂度,因此这种递进边需要从初步的复杂网络结构中删除。

[0014] 为解决上述的问题,我们加入步骤(4),通过最大最小爬山法优化初步复杂网络,实现初步复杂网络的边反转和边删除操作。其具体思路是:对不同的复杂网络结构进行评分,评分高的复杂网络结构要优于评分低的复杂网络结构,在此搜索过程中使用最大最小爬山法,将最终得分高的复杂网络结构作为优化后的复杂网络结构。

[0015] 步骤(4)、复杂网络结构优化

[0016] 采用最大最小爬山法,对不同的复杂网络结构加入惩罚值计算网络得分,评分高的复杂网络结构优于评分低的复杂网络结构,去除掉不会增加整个网络信息的递进边,和确定网络结构中边的方向,加入惩罚值的贝叶斯网络得分计算公式如下:

$$[0017] \quad Score_{BIC}(G:D) = Score_L(G:D) - \frac{\log M}{2} Dim[G]$$

[0018] 其中, $Score_L(G:D)$ 为基于信息论方法计算所得的网络得分, $\frac{\log M}{2} Dim[G]$ 为所加入的惩罚值,该惩罚值用于确定复杂网络结构中的边方向,和去除掉不增加网络整体信息量的递进边;

[0019] 复杂网络的边方向不同,其 $Score_L$ 得分不同,因此借助 $Score_L$ 得分,可确定复杂网络结构中的边方向。 $Score_{BIC}$ 在 $Score_L$ 的基础上,增加了网络结构复杂度惩罚值,即网络结构越复杂,其惩罚值越高, $Score_{BIC}$ 综合得分越低,因此借助惩罚值,可去除复杂网络结构中的递进边。

[0020] 至此,结合复杂网络的节点集合,和优化后的复杂网络边集合,可确定复杂网络结构。步骤(5)、计算复杂网络每条边的条件概率参数。

[0021] 优化后的复杂网络结构可确定节点间相关密切程度,但无法确定节点因相关节点变动的具体变动范围。即复杂网络结构中边 $a \rightarrow b$ ,可解释节点的a从取值 $a_1$ 变动到取值 $a_2$ 的过程中,其相关节点b会因节点的a的取值变动,其取值相应发生改变的概率,但无法解释节点b会因节点的a的最终取值为 $a = a_2$ ,时,节点b的最取值是否处于 $[b_1, b_2]$ 区间内,或 $b \in [b_1, b_2]$ 的概率。

[0022] 针对上述问题,在完成复杂网络结构优化后,我们采用最大似然估计法确定复杂网络的条件概率参数。

[0023] 进一步地,步骤(1)中,相关数据产生节点包括制丝工艺在线监测节点、制丝过程烟丝质量离线检测节点、制丝工艺调控记录节点、所有制丝生产环境记录节点和所有卷烟质量评价结果节点。

[0024] 进一步地,步骤(2)中,数据预处理的方法具体包括:

[0025] A、采用箱型图数据质量评价方法,采用数据的上四分位数、中位数、下四分位数,和上边缘、下边缘的统计数据,将箱型图外数据组作为误差数据去除;

[0026] B、采用移动窗口动态评价法,具体步骤如下:

[0027] 步骤(1)、对步骤A的数据进行排序: $X = \{x_1, x_2, \dots, x_n\}$ ;

[0028] 步骤(2)、设置数据序列X的中位值  $\bar{x} = x_{\lfloor \frac{n+1}{2} \rfloor}$ ;

[0029] 步骤(3)、确定窗口大小:  $l < \lfloor \frac{n+1}{2} \rfloor$ , 或根据实际情况调整窗口大小;

[0030] 步骤(4)、在步骤(1)的数据序列  $X = \{x_1, x_2, \dots, x_n\}$  中, 截取长度为  $2l+1$  的子数据序列  $\hat{X} = \left\{ x_{\lfloor \frac{n+1}{2} \rfloor - l}, \dots, x_{\lfloor \frac{n+1}{2} \rfloor}, \dots, x_{\lfloor \frac{n+1}{2} \rfloor + l} \right\}$ , 计算的  $\hat{X}$  平均值mean, 和标准偏差sd。

[0031] 步骤(5)、对于数据序列  $X = \{x_1, x_2, \dots, x_n\}$  的每个数值进行处理, 若  $\text{mean} - n \times \text{sd} \leq x_i \leq \text{mean} + n \times \text{sd}$ , 其中,  $n$  为根据实际情况设置的常数, 则  $x_i$  正常值; 否则,  $x_i$  为奇异值, 需删除。

[0032] 进一步地, 步骤(3)中, 节点间独立性的计算方法包括如下步骤:

[0033] 步骤(1)、针对需要进行独立性检测的两个节点, 节点a和节点b, 分别采集  $n$  批制丝过程数据, 其对应的数据分别为  $a_1, \dots, a_i, \dots, a_j, \dots, a_n$  和  $b_1, \dots, b_i, \dots, b_j, \dots, b_n$ , 其中  $a_i$  和  $b_i$  均为列向量数据;

[0034] 步骤(2)、构建矩阵A、B, 矩阵元素计算公式如下:

$$[0035] \quad A_{j,k} = ||a_j - a_k||,$$

[0036] 其中,  $a_j$  表示节点a的第  $j$  个数据实例,  $a_k$  表示节点a的第  $k$  个数据实例,  $A_{j,k}$  表示节点a的第  $j$  个数据实例与第  $k$  个数据实例之间的距离, 即欧几里得范数;

[0037] 采用同样的处理方法, 对节点b相应的采集数据进行处理, 构建矩阵B;

[0038] 步骤(3)、计算矩阵A和矩阵B的协方差, 并根据该值计算节点a和节点b间独立性, 按如下公式计算:

$$[0039] \quad dCov_n^2(X, Y) = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n A_{j,k} B_{j,k},$$

[0040]  $dCov_n^2(X, Y)$  即表示节点a和节点b的相关性;

[0041] 步骤(4)、按照上述的方法, 分析复杂网络节点集合中所有节点间的两两相关性, 并按照5%置信水平, 确定节点间是否有边存在, 完成复杂网络边集合的确定, 结合复杂网络节点集合, 完成初步复杂网络构建。

[0042] 进一步地, 步骤(5)中, 针对复杂网络的每条边, 采用最大似然估计法计算复杂网络概率参数, 复杂网络的最大似然函数如下:

$$[0043] \quad \begin{aligned} l_b(D:U, \theta) &= -\frac{1}{2} \sum_{m=1}^M \left[ \log(2\pi) + \log(\sigma^2) + \frac{1}{\sigma^2} (K[m] - g(u[m]))^2 \right] \\ &= -\frac{1}{2} \left[ M \log(2\pi) + M \log(\sigma^2) + \frac{1}{\sigma^2} \sum_{m=1}^M (K[m] - g(u[m]))^2 \right], \end{aligned}$$

[0044] 其中,  $U$  是其中一个节点的父节点集合,  $\theta = \langle \beta_0, \beta_1, \beta_2, \beta_k, \delta \rangle$ ,  $g(u[m]) = (\beta_1 u_1, \dots, \beta_m u_m : \theta) = \sum_j \beta_j u_j + \theta_0$  为线性映射,  $u[m]$  表示包含其中一个节点及其所有父节点数据的, 第  $m$  个样本的实例,  $K[m]$  表示  $u[m]$  中某一个节点的值;

[0045] 根据上式分别对 $\beta_0, \beta_1, \dots, \beta_k, \delta$ 求偏导,得到 $k+2$ 个等式,求解方程组,得到该边的条件概率参数 $\beta_0, \beta_1, \dots, \beta_k, \delta$ ;

[0046] 依据连续型贝叶斯网络的特性,若在连续型贝叶斯网络中,两节点的条件概率满足正态分布;例如:以 $a, b$ 节点为例:

[0047]  $p(b|a=a_0) \sim N(\beta_0 + \beta^T a_0, \sigma^2)$

[0048] 其中, $\beta^T = [\beta_1, \dots, \beta_k]$ 。

[0049] 即,在确定节点 $a$ 处于状态 $a=a_0$ 后,可根据正态分布的特征,通过积分方式计算得到节点 $b$ 处于状态 $[b_1, b_2]$ 的概率;

[0050] 对复杂网络的每条边重复上述操作,计算得到复杂网络每条边的条件概率参数值。

[0051] 与现有技术相比,本发明的有益效果如下:

[0052] (1)、本发明的方法不需要进行数据区间离散化操作,解决了离散型贝叶斯方法中,数据离散化区间难以优选的问题;更符合制丝线连续生产,相关数据联系采集的特征。

[0053] (2)、本发明构建的复杂网络可以系统性地解析制丝过程中工艺参数与工艺质量指标的关系;还可以全局性的分析制丝过程中关键工艺参数及工艺质量指标对卷烟质量的影响,为制丝工艺标准的制修订提供数据支撑,进而保障卷烟质量稳定性。

[0054] 构建贝叶斯网络的过程中,将工艺设置参数、环境因素(温湿度)、过程质量检测结果、烟支卷制质量都吸收进入网络形成节点,通过数学计算方法建立的复杂网络,可以描述出制丝过程中影响工艺质量指标的关键参数,也可以描述出这些关键参数之间的相互影响;是一种能够有效分析制丝过程关键因素对卷烟质量的影响关系的方法,可以分析影响卷制质量的关键因素之间的关系及影响权重,以达到对卷烟质量的精准调控。

## 附图说明

[0055] 图1为实施例中的网络节点集合示意图;

[0056] 图2为实施例中采用箱型图去除“脏数据”的示意图;

[0057] 以某一批制丝工序数据的“叶丝干燥-出料含水率”节点为例,黑色圆点为通过数据预处理,去除掉的“脏数据”,黑线为数据预处理后保留的“正常”数据。

[0058] 图3为实施例中通过独立性检验后所构建的初步复杂网络结构示意图;

[0059] 图4为实施例中通过最大最小爬山法确定初步复杂网络结构中的边方向后的复杂网络示意图;

[0060] 图5为实施例中通过最大最小爬山法删除递进边后,得到最终的优化后的复杂网络结构示意图;

[0061] 图6为实施例中,在节点 $d$ (松散回潮-加水比例)已知取值(或预期取值) $d=3.2$ 情形下,其相关节点 $k$ (叶丝干燥-出料含水率)取值在 $[12.5, 13.5]$ 范围内的概率计算示意图。

## 具体实施方式

[0062] 下面将结合本发明实施例中的附图,对本实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是对本发明一部分实例,而不是全部的实例。基于本发明中的实施例,本领域普通技术人员在没有付出创造性劳动前提下所获得的所有其他实施



例,都属于本发明保护的范围。

[0063] 本实例针对某烟厂牌号制丝线,并采集该烟厂该牌号2017年整年共2393批制丝数据,卷烟制丝工艺关键因素的复杂网络构建方法,其特征在于:包括以下步骤:

[0064] 步骤(1)、构建复杂网络的节点集合

[0065] 实施例中所采集的烟厂制丝线相关数据采集节点如下:

[0066] A、制丝工艺在线监测节点,包括松散回潮-工艺流量、松散回潮-加水比例、松散回潮-加水流量、松散回潮-出料含水率、松散回潮-出料温度、叶丝干燥-热风温度、叶丝干燥-出料含水率、叶丝干燥-出料温度,总计8制丝工艺在线检测节点;

[0067] B、制丝工艺调控记录节点,包括松散回潮-气水混合自动阀门开度、叶丝干燥-排潮阀门开度,总计2个制丝工艺调控记录节点;

[0068] C、所有制丝生产环境记录节点,包括会制丝过程生产环境的温度数据和湿度数据,总计2个环境节点;

[0069] D、所有卷烟质量评价结果节点,包括烟支含水率(m)、烟气水分(n),总计2个卷烟质量评价节点。

[0070] 上述14个数据采集节点,构成复杂网络结构中的节点集合,分别记为:c,d,⋯,o,p。实施例的节点集合如图1所示。

[0071] c:松散回潮-工艺流量

[0072] d:松散回潮-加水比例

[0073] e:松散回潮-加水流量

[0074] f:松散回潮-气水混合自动阀门开度

[0075] g:松散回潮-出料含水率

[0076] h:松散回潮-出料温度

[0077] i:叶丝干燥-排潮阀门开度

[0078] j:叶丝干燥-热风温度

[0079] k:叶丝干燥-出料含水率

[0080] l:叶丝干燥-出料温度

[0081] m:烟支含水率

[0082] n:烟气水分

[0083] o:环境湿度

[0084] p:环境温度

[0085] 步骤(2)、数据预处理

[0086] 在从MES系统中所采集到的制丝工艺在线监测节点和制丝工艺调控记录节点数据中,在生产开始和结束后,所产生的不稳定数据,会影响最终的复杂网络模型结果。为此,采用箱型图方法,去除掉所采集数据中因工艺不稳定所产生的“脏数据”。

[0087] 数据预处理所采用的方法具体如下:

[0088] A、箱型图数据质量评价方法,采用数据的上四分位数、中位数、下四分位数,和上边缘、下边缘的统计数据,将箱型图外数据组作为误差数据去除。

[0089] B、移动窗口动态评价法,移动窗口动态评价法的具体实施步骤如下:

[0090] (1)、对步骤A的数据进行排序: $X = \{x_1, x_2, \dots, x_n\}$ ;

[0091] (2)、设置数据序列X的中位值  $\bar{x} = x_{\lfloor \frac{n+1}{2} \rfloor}$ ;

[0092] (3)、确定窗口大小:  $l < \lfloor \frac{n+1}{2} \rfloor$ , 或根据实际情况调整窗口大小;

[0093] (4)、在的数据序列  $X = \{x_1, x_2, \dots, x_n\}$  中, 截取长度为  $2l+1$  的子数据序列

$\hat{X} = \left\{ x_{\lfloor \frac{n+1}{2} \rfloor - l}, \dots, x_{\lfloor \frac{n+1}{2} \rfloor}, \dots, x_{\lfloor \frac{n+1}{2} \rfloor + l} \right\}$ , 计算的  $\hat{X}$  平均值mean, 和标准偏差sd。

[0094] (5)、对于数据序列  $X = \{x_1, x_2, \dots, x_n\}$  的, 每个数值进行处理, 若  $\text{mean} - n \times \text{sd} \leq x_i \leq \text{mean} + n \times \text{sd}$  (其中,  $n$  为根据实际情况设置的常数), 则  $x_i$  正常值; 否则,  $x_i$  为奇异值, 需删除。

[0095] 其中, 以叶丝干燥-出料含水率数据预处理为例, 结果如图2所示。

[0096] 步骤(3)、初步复杂网络构建

[0097] 对于复杂网络中某节点a的  $n_a$  个实例两两间的距离计算公式如下:

[0098]  $a_{j,k} = ||a_j - a_k||$

[0099] 其中,  $a_j$  表示节点a的第j个数据实例,  $a_k$  表示节点a的第k个数据实例,  $A_{j,k}$  表示节点a的第j个数据实例与第k个数据实例之间的距离, 即欧几里得范数;

[0100] 根据上述得到的  $a_{j,k}$ , 按照如下的方式, 获得矩阵A:

[0101]  $A_{j,k} = a_{j,k} - \bar{a}_{j\cdot} - \bar{a}_{\cdot k} + \bar{a}_{\cdot\cdot}$

[0102] 其中,  $A_{j,k}$  为矩阵A中, 第j行第k列的元素,  $\bar{a}_{j\cdot}$  为行平均值,  $\bar{a}_{\cdot k}$  为列平均值,  $\bar{a}_{\cdot\cdot}$  表示总平均值。

[0103] 根据上述方法, 分别得到节点a和节点b所对应的矩阵A和B, 按如下方式, 矩阵间的协方差。

[0104]  $dCov_n^2(X, Y) = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n A_{j,k} B_{j,k}$

[0105]  $dCov_n^2(X, Y)$  即表示节点a和节点b的相关性, 按照上述的方法, 分析复杂网络节点集合中所有节点间的两两相关性, 并按照5%置信水平, 确定节点间是否有边存在, 完成复杂网络边集合的确定, 结合复杂网络节点集合, 完成初步复杂网络构建。

[0106] 步骤(4)、复杂网络结构优化

[0107] 如上所示, 对复杂网络进行评分, 是采用最大最小爬山法优化初步复杂网络的关键。本项目中, 复杂网络评分公式如下所示:

[0108]  $Score_{BIC}(G: D) = Score_L(G: D) - \frac{\log M}{2} Dim[G]$

[0109] 复杂网络的边方向不同, 其  $Score_L$  得分不同, 因此借助  $Score_L$  得分, 可确定复杂网络结构中的边方向。 $Score_{BIC}$  在  $Score_L$  的基础上, 增加了网络结构复杂度惩罚值, 即网络结构越复杂, 其惩罚值越高,  $Score_{BIC}$  综合得分越低, 因此借助惩罚值, 可去除复杂网络结构中的递进边。

[0110] 至此, 结合复杂网络的节点集合, 和优化后的复杂网络边集合, 可确定复杂网络结构。

[0111] 1) 优化过程中,不断调整图3所示的初步复杂网络中的边方向,并计算相应复杂网络的得分后,确定初步复杂网络中所有边的方向,确定了边方向的复杂网络结构如图4所示。

[0112] 2) 在确定了复杂网络边方向基础上,继续通过复杂网络结构的优化,删除复杂网络中的递进边,得到优化后的复杂网络结构如图5所示。

[0113] 步骤(5)、计算复杂网络每条边的条件概率参数。

[0114] 优化后的复杂网络结构可确定节点间相关密切程度,但无法确定节点因相关节点变动的具体变动范围。例如:复杂网络结构中边 $d \rightarrow k$ ,可解释节点的 $d$ 从取值 $d_1$ 变动到取值 $d_2$ 的过程中,其相关节点 $k$ 会因节点的 $d$ 的取值变动,其取值相应发生改变的概率,但无法解释节点 $k$ 会因节点的 $d$ 的最终取值为 $d = d_2$ ,时,节点 $k$ 的最取值是否处于 $[k_1, k_2]$ 区间内的概率。

[0115] 针对上述问题,在完成复杂网络结构优化后,本发明采用最大似然估计法确定复杂网络的条件概率参数。

[0116] 以 $d, k$ 两节点为例,复杂网络的最大似然函数如下:

$$[0117] \quad l_k(D:U, \theta) = -\frac{1}{2} \sum_{m=1}^M \left[ \log(2\pi) + \log(\sigma^2) + \frac{1}{\sigma^2} (d[m] - g(u[m]))^2 \right]$$

$$= -\frac{1}{2} \left[ M \log(2\pi) + M \log(\sigma^2) + \frac{1}{\sigma^2} \sum_{m=1}^M (d[m] - g(u[m]))^2 \right]$$

[0118] 其中, $U$ 是 $k$ 节点的所有父节点集合, $\theta = \langle \beta_0, \beta_1, \beta_2, \dots, \beta_k, \delta \rangle$ ,

$g(u[m]) = (\beta_1 u_1, \dots, \beta_m u_m : \theta) = \sum_j \beta_j u_j + \theta_0$ 为线性映射, $u[m]$ 表示包含节点 $d$ 及其所有父节点数据,第 $m$ 个样本的实例, $d[m]$ 表示 $u[m]$ 中节点 $e$ 的值。

[0119] 在上式中,对 $\beta_0$ 求偏导:

$$[0120] \quad \frac{\partial}{\partial \beta_0} l_k(D:U, \theta) = -\frac{1}{\sigma^2} \sum_{m=1}^M (\beta_0 + \beta_1 u_1[m] + \dots + \beta_k u_k[m] - x[m])$$

$$= -\frac{1}{\sigma^2} \left( M \beta_0 + \beta_1 \sum_m u_1[m] + \dots + \beta_k \sum_m u_{1k}[m] - \sum_m x[m] \right) = 0$$

[0121] 在上式中,两边同乘 $\frac{\sigma^2}{M}$ ,并移项后得到:

$$[0122] \quad \frac{1}{M} \sum_m a[m] = \beta_0 + \beta_1 \frac{1}{M} \sum_m u_1[m] + \dots + \beta_k \frac{1}{M} \sum_m u_k[m]$$

[0123] 其中, $E_D(U_j) = \frac{1}{M} \sum_m u_j[m]$ 为平均值,该值可根据所采集的数据计算得到。

[0124] 即上式可表示为如下等式:

$$[0125] \quad E_D(X) = \beta_0 + \beta_1 E_D(U_1) + \dots + \beta_k E_D(U_k)$$

[0126] 同样的,对 $\beta_1, \dots, \beta_k, \delta$ 进行上述操作,可得到 $k+2$ 个等式,通过求解方程组,可得 $\beta_0, \beta_1, \dots, \beta_k, \delta$ 的值。

[0127] 依据连续型贝叶斯网络的特性,若在连续型贝叶斯网络中,两节点的条件概率满足正态分布。

[0128]  $p(k|d=d_0) \sim N(\beta_0 + \beta^T d_0, \sigma^2)$

[0129] 其中,  $\beta^T = [\beta_1, \dots, \beta_k]$ 。

[0130] 在节点d(松散回潮-加水比例)处于状态 $d=3.2$ 时,根据如上方法计算得出d、k两节服从分布 $p(k|d=3.2) \sim N(13.0341, 0.1657^2)$ ,得出在此状态下,可根据正态分布的特征,通过积分方式计算其相关节点k(叶丝干燥-出料含水率)在 $[12.5, 13.5]$ 范围内的概率为95%,如图6所示。

[0131] 重复上述步骤,直到复杂网络中所有边的条件概率参数计算完毕。

[0132] 至此,卷烟制丝工艺关键因素影响卷烟质量的复杂网络构建完成。

[0133] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

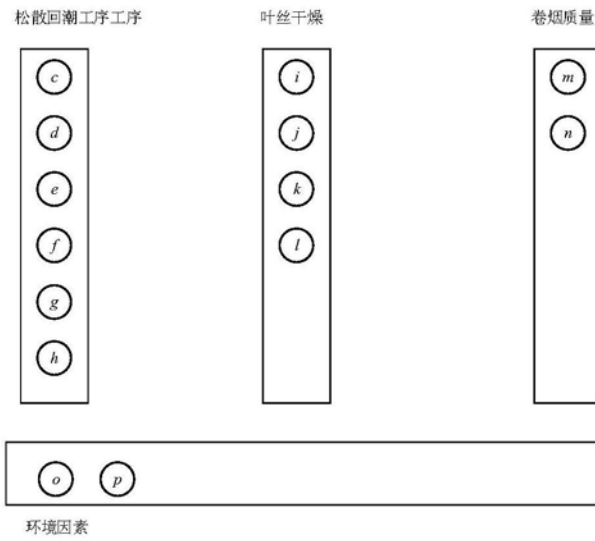


图1

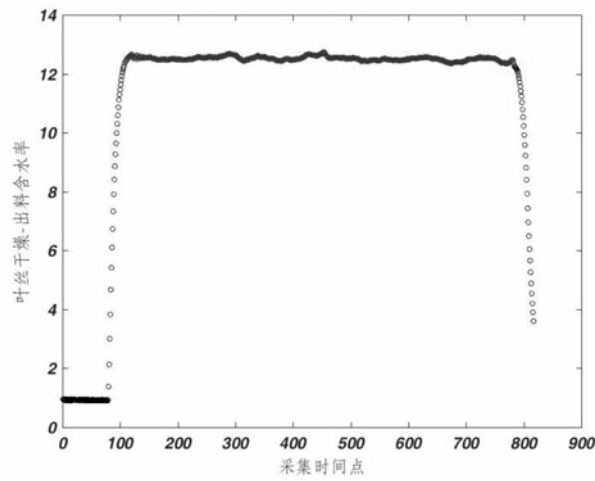


图2

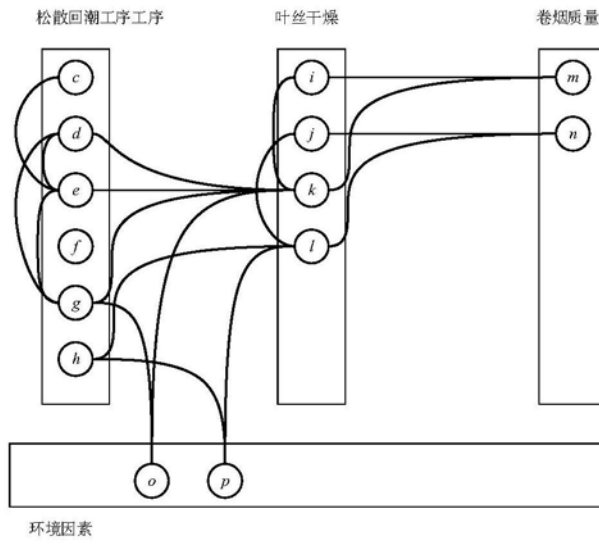


图3

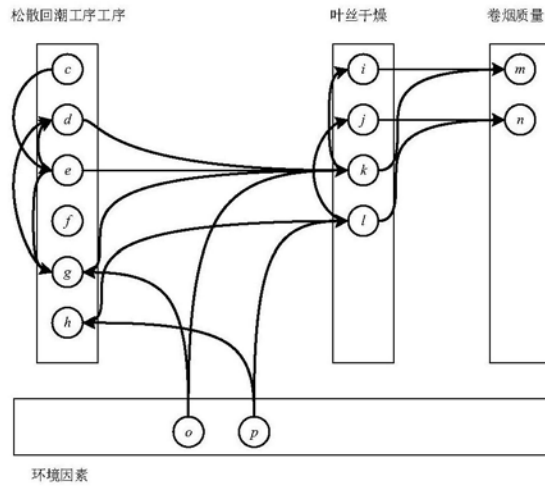


图4

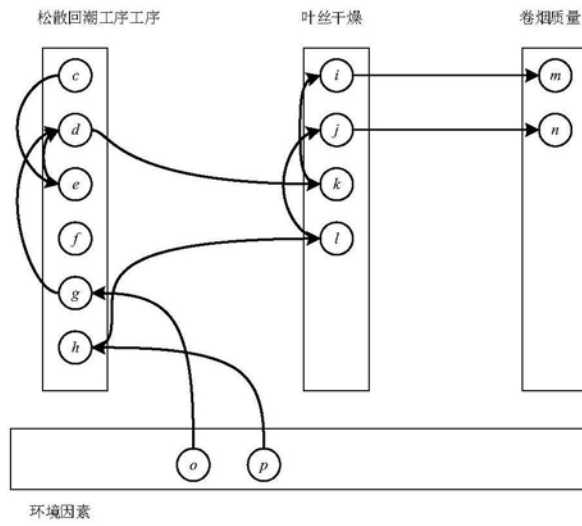


图5

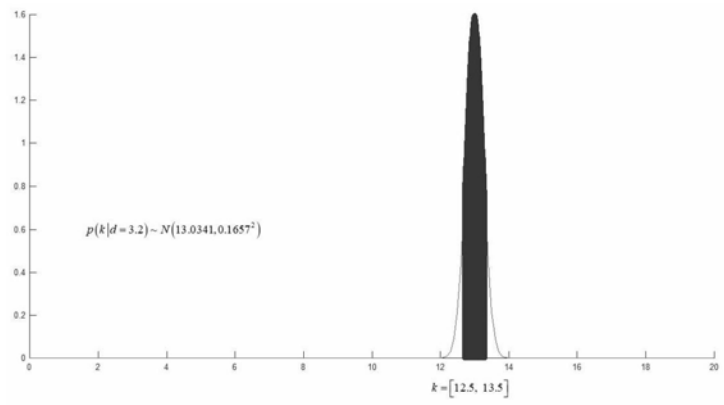


图6