

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 3 月 19 日 (19.03.2020)



(10) 国际公布号
WO 2020/051795 A1

(51) 国际专利分类号:
G06Q 10/04 (2012.01)

(21) 国际申请号: PCT/CN2018/105197

(22) 国际申请日: 2018 年 9 月 12 日 (12.09.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 大连理工大学 (DALIAN UNIVERSITY OF TECHNOLOGY) [CN/CN]; 中国辽宁省大连市凌工路 2 号大连理工大学专利中心, Liaoning 116024 (CN)。

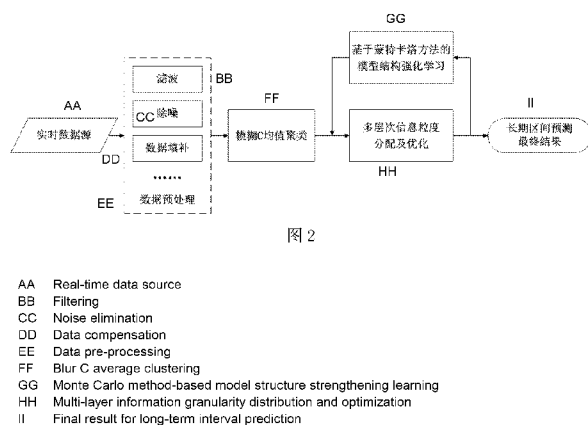
(72) 发明人: 韩中洋 (HAN, Zhongyang); 中国辽宁省大连市凌工路 2 号大连理工大学专利中心, Liaoning 116024 (CN)。 赵琨 (ZHAO, Jun); 中国

辽宁省大连市凌工路 2 号大连理工大学专利中心, Liaoning 116024 (CN)。 王伟 (WANG, Wei); 中国辽宁省大连市凌工路 2 号大连理工大学专利中心, Liaoning 116024 (CN)。 王霖青 (WANG, Linqing); 中国辽宁省大连市凌工路 2 号大连理工大学专利中心, Liaoning 116024 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: LONG-TERM INTERVAL PREDICTION FOR STEEL COAL GAS SYSTEM AND STRUCTURE LEARNING METHOD THEREFOR

(54) 发明名称: 一种钢铁煤气系统长期区间预测及其结构学习方法



(57) Abstract: Long-term interval prediction for steel coal gas system and structure learning method therefor, belonging to the field of information technology. The method employs real industry data: first, constructing a multi-layer information granularity unequal length distribution structure, and establishing a corresponding optimization model; next, considering the importance of model structure to prediction accuracy, carrying out strengthening learning on structural parameters of the multi-layer model by means of a Monte Carlo method; finally, computing a structure on the basis of optimal multi-layer granularity, and using a parallel computing policy to obtain a long term interval prediction result for the yield and consumption of coal gas. The accuracy of a result obtained by the present method is relatively high, and the computing efficiency meets actual application requirements; in addition, the method may be popularized and applied in the field of steel for other energy medium systems.

(57) 摘要: 一种钢铁煤气系统长期区间预测及其结构学习方法, 属于信息技术领域。该方法采用工业真实数据, 首先构造多层次的信息粒度非等长分配结构, 建立相应优化模型; 进而, 考虑到模型结构对预测精度的重要性, 借助蒙特卡洛方法, 对多层次模型的结构参数进行强化学习; 最终基于最优的多层粒度计算结构, 运用并行计算策略, 求得煤气产消量的长期区间预测结果。该方法所得到结果精度较高, 且计算效率符合实际应用要求, 在钢铁工业其它能源介质系统中亦可推广应用。

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

- (84) 指定国(除另有指明，要求每一种可提供的地区
保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

- 包括国际检索报告(条约第21条(3))。

一种钢铁煤气系统长期区间预测及其结构学习方法

技术领域

本发明属于信息技术领域，涉及到模糊建模、强化学习、并行计算等技术，是一种粒度计算与强化学习相结合的钢铁工业煤气系统长期区间预测及其结构学习方法。本发明采用工业真实数据，首先构造多层次的信息粒度非等长分配结构，建立相应优化模型；进而，考虑到模型结构对预测精度的重要性，本发明借助蒙特卡洛方法，对多层次模型的结构参数进行强化学习；最终基于最优的多层粒度计算结构，运用并行计算策略，求得煤气产消量的长期区间预测结果。此方法所得到结果精度较高，且计算效率符合实际应用要求，在钢铁工业其它能源介质系统中亦可推广应用。

背景技术

节能减排始终是钢铁工业日常生产的重要目标之一。作为随生产过程产生的副产煤气这一二次能源，能否实现对其合理利用将直接关系到企业经济利益和降耗效果。钢铁工业的副产煤气主要包括高炉煤气、焦炉煤气和转炉煤气三种，由于生产需求变化、设备切换操作等原因，现场常出现产消失衡情况，此时需制定调度方案以使管网达到新的产消平衡，从而保障生产顺畅、避免资源浪费。在上述过程中，产销量变化趋势是能源调度人员做出决策的重要依据，因此对副产煤气的预测工作具有重要的实际应用意义。（熊超，钢铁企业煤气系统节能探讨[C]. (2011). 中国金属学会冶金技术经济学会学术年会.）

随着研究应用工作的广泛开展，以粒度计算为代表的预测模型已可实现超过 4 小时的长期趋势估计（J. Zhao, Z.Y. Han, W. Pedrycz, W. Wang. (2016). Granular model of long-term prediction for energy system in steel industry[J]. IEEE transactions on cybernetics, 46(2), 388-400）（Z.Y. Han, J. Zhao, Q. Liu, W. Wang. (2016). Granular-computing based hybrid collaborative fuzzy clustering for long-term prediction of multiple gas holders levels[J]. Information Sciences, 330, 175-185）。然而，这些方法的预测结果均以点为呈现形式，无法满足现场对结果可靠性衡量的需求。鉴于此，除了支持向量机（Support Vector Machine, SVM）、基于统计学的一些模型均可被用来实现区间预测以外（C.Y. Sheng, J. Zhao, W. Wang, H. Leung. (2013). Prediction intervals for a noisy nonlinear time series based on a bootstrapping reservoir computing network ensemble[J]. IEEE Transactions on neural networks and learning systems, 24(7), 1036-1048）（A. Khosravi, S. Nahavandi, D. (2010). Creighton. Construction of optimal prediction intervals for load forecasting problems[J]. IEEE Transactions on Power Systems, 25(3), 1496-1503.），将粒度计算中的聚类中心纵向拓展为区间值、进而求解相应优化模型，也是一个可行的解决方案（Z.Y. Han, J. Zhao, H. Leung, W. Wang. (2018). Construction of prediction intervals for gas flow systems in steel industry based on granular computing[J]. Control Engineering Practice, 78, 79-88）。

然而，这些方法存在明显不足：首先，由于以迭代机制构造样本，SVM、基于统计学等模型在长期预测方面精度难以令人满意，一般仅能给出 60 个点以内的好预测结果；其次，虽然单层的信息粒度分配方式可计算出较长时间的区间估计结果，但大量的待优化参数导致其运算时间和精度均存在不确定性，若能将这些参数分层次处理，则有望在计算效率和平均精度上获得明显提升；此外，若应用多层次粒度计算模型，其结构将直接左右长期区间预测结果的精度，因此需要设计方法以高效、合理地获取最优模型结构参数。

发明内容

本发明主要解决钢铁工业副产煤气系统的产销量长期区间预测及模型结构学习问题。方法使用采集自现场的真实工业数据，首先建立多层次的信息粒度非等长分配结构及相应优化模型；为自适应地确定模型结构，本发明借助蒙特卡洛方法对结构参数进行强化学习；最终借助并行计算进行分层次的高效求解，得到长期区间预测结果。

本发明的技术方案：

一种钢铁煤气系统长期区间预测及其结构学习方法，步骤如下：

(1) 由现场实时数据库采集副产煤气系统的产销量数据，经过除噪、滤波及填补等预处理后作为基础数据样本；

(2) 应用模糊 C 均值 (FuzzyC-Means, FCM) 算法进行聚类，得到聚类中心及隶属度；

(3) 逐层次的对聚类中心分配信息粒度，将其由点值拓展为区间值，并建立各层次的信息粒度优化模型；

(4) 利用并行计算策略，逐层求解最优结构参数下的信息粒度分配优化模型，并结合概率、模糊建模等手段，得到初步的长期区间预测结果；

(5) 将长期区间预测模型的当前结构定义为“状态”、参数变化定义为“动作”，借助蒙特卡洛方法，以强化学习的方式求取近似的策略值函数，从而确定模型的最优结构参数，并再次运用步骤 4，得到最终的长期区间预测结果；

本发明的有益效果：本发明的长期区间预测模型将信息粒度分层优化，克服了传统单层方法需求解参数过多、平均精度较低的问题。所建立的粒度分配优化模型将信息的覆盖性描述为约束条件，而目标函数仅为专一性一个目标，从而避免了求解多目标问题的繁琐。此外，蒙特卡洛方法的运用，为长期区间预测模型的结构学习提供了强化学习机制，从而可自适应地确定多层次粒度计算结构。加之并行计算在上述优化模型求解、强化学习过程中的运用，保障了方法的计算效率可符合实际应用之要求。

附图说明

图 1 为钢铁工业副产煤气系统示意图。

图 2 为本发明应用流程图。

图 3 为多层次信息粒度分配及优化结构示意图。

图 4(a)为 MVE 方法对#2 高炉煤气发生量的长期区间预测结果图。

图 4(b)为单层粒度计算方法对#2 高炉煤气发生量的长期区间预测结果图。

图 4(c)为本发明方法对#2 高炉煤气发生量的长期区间预测结果图。

图 5(a)为 MVE 方法对#1 焦炉煤气使用量的长期区间预测结果图。

图 5(b)为单层粒度计算方法对#1 焦炉煤气使用量的长期区间预测结果图。

图 5(c)为本发明方法对#1 焦炉煤气使用量的长期区间预测结果图。

图中的 MVE 指均值-方差估计 (Mean-Variance Estimation)

具体实施方式

为了更好地理解本发明的技术路线和实施方案,下面以国内钢铁工业自动化水平较高的上海宝山钢铁厂副产煤气系统做进一步说明。由附图 1 所示的宝钢煤气系统示意图可看出,四座高炉、六座焦炉和六座转炉构成了三种主要副产煤气的发生单元,而消耗单元则包括冷/热轧、烧结等,其中低压锅炉和电厂常作为可调节单元。管网内还包含多座煤气柜,起到暂存和缓冲的作用。此外,煤气混合站和加压站作为输配系统,负责将煤气压送至各消耗单元。在日常生产中,维持产消平衡既可以保障生产的顺利进行,也有助于达成节能减排的目标,因此是能源调度人员的工作重点。由于副产煤气网络复杂,遍布炼铁、炼钢、轧钢等多个生产区域,具有非线性、大时滞等显著特点,能源调度人员难以准确判断产消量的未来趋势。本发明即针对此问题,开展副产煤气产消量的预测方法研究与应用工作。

本发明的具体实施步骤如下:

步骤 1: 数据预处理

从工业现场实时关系数据库,读取副产煤气系统发生和消耗单元数据,并做基本的除噪、滤波和填补等预处理工作。

步骤 2: FCM

将数据分割为数段等长度的片段,即 $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N\}$, 其中 $\mathbf{z}_i \in \mathbb{R}^n$, n 代表每个数据段所含数据点个数, N 则为数据段个数。应用 FCM 聚类算法,获得聚类中心 $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c\}$ 和对应隶属度 $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N\}$, 其中 $\mathbf{v}_i \in \mathbb{R}^n$, $\mathbf{u}_i \in \mathbb{R}^c$, c 是聚类中心维度。

步骤 3: 建立多层次粒度计算模型

如附图 1 所示,由下至上逐层次地对聚类中心 $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c\}$ 分配信息粒度 $\alpha_{i,j}$ 和 β_i , 其中 $i = 1, 2, \dots, m, j = 1, 2, \dots, n_i$, 且 $n_1 \neq n_2 \neq \dots \neq n_m$ 。如此即可将聚类中心由点值延展为

区间值。为优化求解上述信息粒度参数，首先定义覆盖性 cov 和专一性 $spec$ 两个衡量指标如下：

$$cov \doteq \frac{1}{T} \sum_{i=1}^T \lambda_i \quad (1)$$

$$spec \doteq range - \frac{1}{T} \sum_{i=1}^T |\bar{z}_i - \underline{z}_i| \quad (2)$$

其中， T 表示样本所含数据点个数； λ_i 是标识变量，即当区间覆盖样本的数据点时等于 1，否则等于 0； $range$ 指样本数据的最大值与最小值之差， $\bar{z}_i$ 和 $\underline{z}_i$ 分别表示预测区间上下限。

信息粒度模型的优化目标是使得（1）和（2）均取极大值，其中 cov 至少应等于目标置信度 $(1 - \rho) \times 100\%$, $\rho \in [0,1]$ 是显著水平。为避免求解多目标问题的困难与繁琐，本发明将（1）作为约束条件考虑，即 cov 必须大于或等于目标置信区间。信息粒度的优化顺序与分配顺序相反，各层优化模型建立如下：

① 第二层

$$\begin{aligned} \max \quad & range^{(2)} - \frac{1}{m} \sum_{i=1}^m |\bar{z}_i^{(2)} - \underline{z}_i^{(2)}| \\ \text{s.t.} \quad & \sigma_{min}^{(2)} \varepsilon \leq \frac{1}{m} \sum_{i=1}^m \beta_i \leq \sigma_{max}^{(2)} \varepsilon \\ & \frac{1}{m} \sum_{i=1}^m \lambda_i^{(2)} \geq (1 - \rho) \times 100\% \end{aligned} \quad (3)$$

其中， $range^{(2)}$ 指第二层对应数据样本的最大值与最小值之差， $\bar{z}_i^{(2)}$, $\underline{z}_i^{(2)}$ 表示第二层所得区间结果的上下限； ε 是控制总体信息粒度的超参数； $\sigma_{min}^{(2)}$ 和 $\sigma_{max}^{(2)}$ 用来控制 β_i ，以使其不至于过分偏离 ε ； $\lambda_i^{(2)}$ 是类似 λ_i 的标识变量，即当第二层所得区间覆盖样本的数据点时为 1，否则为 0。

② 第一层

与第二层仅处理一个优化问题不同，第一层要计算合计为 m 个的一系列优化问题，其中任意一个可表述如下：

$$\begin{aligned} \max \quad & range_i^{(1)} - \frac{1}{n_i} \sum_{j=1}^{n_i} |\bar{z}_{ij}^{(1)} - \underline{z}_{ij}^{(1)}| \\ \text{s.t.} \quad & \sigma_{min}^{(1)} \beta_i \leq \frac{1}{n_i} \sum_{j=1}^{n_i} \alpha_{ij} \leq \sigma_{max}^{(1)} \beta_i \\ & \frac{1}{n_i} \sum_{j=1}^{n_i} \lambda_{ij}^{(1)} \geq (1 - \rho) \times 100\% \end{aligned} \quad (4)$$

其中， $range_i^{(1)}$ 指第一层各优化问题对应数据样本的最大值与最小值之差， $i = 1, 2, \dots, m$ ； $\bar{z}_{ij}^{(1)}$ 和 $\underline{z}_{ij}^{(1)}$ 分别为第一层各优化问题求解所得区间上下限； $\sigma_{min}^{(1)}$ 和 $\sigma_{max}^{(1)}$ 则用来控制 α_{ij} 不过分偏离 β_i 。

考虑到收敛性和求解速度，本发明运用差分进化（Differential Evolution, DE）算法求解上述优化问题。需要特别指出的是，第一层的各个优化问题之间相互独立，因此本发明采用并

行策略处理，如此便可较大程度地缩短计算时间，以符合现场对于实时性的要求。

步骤 4：长期区间预测

基于粒度计算的长期区间预测本质上是预测模糊隶属度，即 $\hat{\mathbf{u}}_k = f(\mathbf{u}_{k-n_l}, \dots, \mathbf{u}_{k-1})$ ，其中 $\hat{\mathbf{u}}_k$ 为隶属度预测值， $\mathbf{u}_{k-n_l}, \dots, \mathbf{u}_{k-1}$ 是隶属度矩阵 $\mathbf{U}$ 的一部分， n_l 表示隶属度关系的输入个数。本发明将通过概率估计实现隶属度的预测。为便于理解，以下定义及叙述以点形式给出。首先定义粒度计算框架下的聚类中心概率 $p(\mathbf{v}_{i1}, \dots, \mathbf{v}_{in_l})$ 、数据段概率 $p(\mathbf{z}_{k-n_l}, \dots, \mathbf{z}_{k-1})$ 和共生矩阵 $\mathbb{P}$ 如下：

$$p(\mathbf{v}_{i1}, \dots, \mathbf{v}_{in_l}) \doteq \frac{\sum_{j=1}^N \mathbb{O}_j}{N} \quad (5)$$

$$p(\mathbf{z}_{k-n_l}, \dots, \mathbf{z}_{k-1}) \doteq \{p(\mathbf{v}_{i1}, \dots, \mathbf{v}_{in_l}) | \{i1, \dots, in_l\} \in [1, c]\} \quad (6)$$

$$\mathbb{P} \doteq \{p(\mathbf{v}_i | \mathbf{v}_{j1}, \dots, \mathbf{v}_{jn_l}) | i = 1, \dots, c; \{j1, \dots, jn_l\} \in [1, c]\} \quad (7)$$

其中， $\{\mathbf{v}_{i1}, \dots, \mathbf{v}_{in_l}\} (\{i1, i2, \dots, in_l\} \in [1, c])$ 是聚类中心矩阵 $\mathbf{V}$ 中的一部分， $\mathbf{z}_{k-n_l}, \dots, \mathbf{z}_{k-1}$ 是数据片段集合 $\mathbf{Z}$ 中的一部分； $\mathbb{O}_j$ 是标识型变量，即考虑 $\{u_{h1j} = \max(\mathbf{u}_{h1}), \dots, u_{hn_lj} = \max(\mathbf{u}_{hn_l})\}$ ($\{u_{h1j}, u_{h2j}, \dots, u_{hn_lj}\}$ 分别是隶属度 $\mathbf{u}_{h1}, \mathbf{u}_{h2}, \dots, \mathbf{u}_{hn_l}$ 中的元素，且 $\{h1, h2, \dots, hn_l\} \in [1, c]$)，当满足 $h1 = i1, \dots, hn_l = in_l$ 时为 1，否则为 0； $p(\mathbf{v}_i | \mathbf{v}_j)$ 是条件概率，即当 $\{\mathbf{z}_{k-n_l}, \dots, \mathbf{z}_{k-1}\}$ 的最大隶属度出现在聚类中心集合 $\{\mathbf{v}_{j1}, \dots, \mathbf{v}_{jn_l}\}$ 时， $\mathbf{z}_k$ 的最大隶属度出现在 $\mathbf{v}_i$ 的概率。

基于上述定义，数据段 $\mathbf{z}_k$ 的概率可估计为 $\hat{\mathbf{p}}(\mathbf{z}_k) = p(\mathbf{z}_{k-n_l}, \dots, \mathbf{z}_{k-2}, \mathbf{z}_{k-1})^T \times \mathbb{P}$ ，对应预测值 $\hat{\mathbf{z}}_k$ 可由中心法得到，即：

$$\hat{\mathbf{z}}_k = \frac{\sum_{i=1}^c \hat{\mathbf{p}}_i \mathbf{V}}{\sum_{i=1}^c \hat{\mathbf{p}}_i} \quad (8)$$

其中， $\hat{\mathbf{p}}_i$ 是 $\hat{\mathbf{p}}(\mathbf{z}_k)$ 中元素。

步骤 5：模型结构参数的强化学习

本发明将多层次粒度计算的模型结构确定问题看做单步马尔科夫决策过程，进而采用蒙特卡洛方法，强化学习包括 m 和 $n_i (i = 1, 2, \dots, m)$ 在内的结构参数。首先定义状态 $\mathbf{S}$ 、动作 $\mathbf{A}$ 和奖赏 $\mathbf{R}$ 如下：

$\mathbf{S}$ – 确定的多层次粒度计算模型结构

$\mathbf{A}$ – 变更参数 m 和 $n_i (i = 1, 2, \dots, m)$

$\mathbf{R}$ – 所得预测区间的 *spec*

由于待确定结构参数数量较庞大，本发明采用梯度下降近似函数策略，学习策略值函数 $\pi_\omega(s, a)$ 。设 $\pi_\omega(s, a)$ 为多层感知机神经网络：

$$\pi_{\omega}(s, a) = f(\omega^T \cdot \phi(s, a) + b) \quad (9)$$

其中, $\phi(s, a)$ 是表征状态-动作对的特征向量, 本发明定义为 $\phi(s, a) = (m, n_1, n_2, \dots, n_m)^T$; b 表示多层感知机的偏置量, f 表示激活函数, 本发明采用的是 sigmoid 函数。

定义可导的性能函数如下:

$$J(\omega) = q_{\pi_{\omega}}(s_0) \quad (10)$$

其中, $q_{\pi_{\omega}}$ 是 $\pi_{\omega}(s, a)$ 的真值函数, s_0 是起始状态。对 $J(\omega)$ 求关于 ω 的梯度并应用策略梯度定理, 可最终得出 ω 的更新公式为:

$$\omega_{t+1} = \omega_t + \tau \gamma \nabla_{\omega_t} \ln \pi_{\omega_t}(s, a) r_t \quad (11)$$

其中, τ 表示变化步长, γ 是贴现因子, r_t 是 t 时刻获得的奖赏, 即:

$$r_t = \frac{1}{T} \sum_{i=1}^T |\bar{z}_i - \underline{z}_i| \quad (12)$$

设结构参数的搜索空间维度为 L , 模型参数确定过程可总结如下:

- ① 初始化 $\tau > 0$, $\gamma > 0$, $\omega \in \mathbb{R}^{max(m)}$, 在 L 中取 l 个结构参数样本作为训练子集;
- ② 取 $t = 1$ 到 l , 利用步骤四表述过程循环先后计算 (12) 和 (11), 获得学习到的策略值函数参数 ω_{opt} , 考虑到问题的相互独立性, 此步中 (12) 的计算可借助并行策略加速;
- ③ 取 $t = 1$ 到 L , 计算 (9), 选取其中策略值函数最大值对应的特征向量 $\phi_{opt}(s, a)$ 为最优结构参数, 即:

$$\phi_{opt}(s, a) = \arg \max_{t=1,2,\dots,L} (\pi_{\omega_{opt}}(s_t, a_t)) \quad (13)$$

- ④ 利用获得最优结构参数, 再次计算 (3) - (8), 即可获得最终的长期区间预测结果。

由以上过程可以看出, 本发明一方面将信息粒度分层次分配和并行优化, 在提高运算效率的同时保证预测精度; 另一方面还通过强化学习方式, 自适应地确定多层粒度计算的结构参数。

图 4、图 5 分别是针对 #2 高炉煤气发生量、#1 焦炉煤气发生量的长期区间预测结果, 预测时长为 480 个点, 即 8 个小时, 其中(a)为统计学中的均值-方差估计(Mean-Variance Estimation, MVE)方法, (b)为一般的单层粒度计算长期区间预测模型, (c)则是本发明方法。虚线为真实值, 灰色带状区域为构造的预测区间。表 1 给出预测区间精度及运算效率比较, 衡量指标包括预测区间覆盖率(Prediction Intervals Coverage Probability, PICP)、正则化平均区间宽度(Prediction Intervals Normalized Average Width, PINAW)、区间得分(Interval Score, IS)以及计算耗时(Computing Time, CT), 其中 PICP、PINAW 和 IS 的定义如下:

$$PICP = \frac{1}{T_{test}} \sum_{i=1}^{T_{test}} \lambda_i \times 100\% \quad (14)$$

$$PINAW = \frac{1}{T_{test} (d_{max} - d_{min})} \sum_{i=1}^{T_{test}} |\bar{z}_i - \underline{z}_i| \quad (15)$$

$$IS(\bar{z}_i, \underline{z}_i, t_i) = \frac{1}{T_{test}} \sum_{i=1}^{T_{test}} e_i \times 100\% \quad (16)$$

其中, T_{test} 是测试集所含数据点总数, λ_i 为标识变量, 当测试集中的数据点在预测区间内时 $\lambda_i = 1$, 否则 $\lambda_i = 0$; $\bar{z}_i$ 和 $\underline{z}_i$ 分别是预测区间上下限; 测试集中的极大值、极小值分别是 d_{max} 和 d_{min} ; e_i 是个分段定义变量:

$$e_i = \begin{cases} -2\rho(\bar{z}_i - \underline{z}_i) - 4(\underline{z}_i - d_i), & \text{当 } d_i < \underline{z}_i \\ -2\rho(\bar{z}_i - \underline{z}_i), & \text{当 } \underline{z}_i \leq d_i \leq \bar{z}_i \\ -2\rho(\bar{z}_i - \underline{z}_i) - 4(d_i - \bar{z}_i), & \text{当 } d_i > \bar{z}_i \end{cases} \quad (17)$$

其中, d_i 是测试集中的数据点, 实验中取显著水平 $\rho = 0.1$ 。综合图表结果可以明显看出, 本发明在精度表现和运算效率上均优于其它工业常用的区间预测方法。

表 1 三种方法在单次长期区间预测的精度及耗时结果比较

预测对象	预测方法	$PICP$ (%)	$PINAW$	IS	CT (s)
#2 高炉煤气 发生量	MVE	67.76	0.33	-63.08	11.34
	单层粒度计算模型	65.71	0.30	-64.35	53.17
	本发明	96.46	0.26	-18.33	9.81
#1 焦炉煤气 发生量	MVE	77.50	0.87	-1.63	10.54
	单层粒度计算模型	88.54	0.88	-1.27	59.93
	本发明	95.13	0.73	-0.92	7.98

权利要求书

1.一种钢铁煤气系统长期区间预测及其结构学习方法，其特征在于，步骤如下：

步骤 1：数据预处理

从工业现场实时关系数据库，读取副产煤气系统发生和消耗单元数据，并做基本的除噪、滤波和填补；

步骤 2：FCM

将步骤 1 读取的数据分割为数段等长度的片段，即 $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N\}$ ，其中 $\mathbf{z}_i \in \mathbb{R}^n$ ， n 代表每个数段所含数据点个数， N 则为数段个数；应用 FCM 聚类算法，获得聚类中心 $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c\}$ 和对应隶属度 $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N\}$ ，其中 $\mathbf{v}_i \in \mathbb{R}^n, \mathbf{u}_i \in \mathbb{R}^c$ ， c 是聚类中心维度；

步骤 3：建立多层次粒度计算模型

由下至上逐层次地对聚类中心矩阵 $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c\}$ 分配信息粒度 α_{ij} 和 β_i ，其中 $i = 1, 2, \dots, m$ ； $j = 1, 2, \dots, n_i$ ；且 $n_1 \neq n_2 \neq \dots \neq n_m$ ；如此即将聚类中心由点值延展为区间值；为优化求解上述信息粒度参数，首先定义覆盖性 cov 和专一性 $spec$ 两个衡量指标如下：

$$cov \doteq \frac{1}{T} \sum_{i=1}^T \lambda_i \quad (1)$$

$$spec \doteq range - \frac{1}{T} \sum_{i=1}^T |\bar{z}_i - \underline{z}_i| \quad (2)$$

其中， T 表示数据样本所含数据点个数； λ_i 是标识变量，即当区间覆盖样本的数据点时等于 1，否则等于 0； $range$ 指样本数据的最大值与最小值之差， $\bar{z}_i$ 和 $\underline{z}_i$ 分别表示预测区间上限和下限；

信息粒度模型的优化目标是使得覆盖性 cov 和专一性 $spec$ 均取极大值，其中 cov 至少等于目标置信度 $(1 - \rho) \times 100\%$ ， $\rho \in [0, 1]$ 是显著水平；将式 (1) 作为约束条件考虑，即 cov 必须大于或等于目标置信区间；信息粒度的优化顺序与分配顺序相反，各层优化模型建立如下：

(1) 第二层

$$\begin{aligned} \max \quad & range^{(2)} - \frac{1}{m} \sum_{i=1}^m |\bar{z}_i^{(2)} - \underline{z}_i^{(2)}| \\ \text{s.t.} \quad & \sigma_{min}^{(2)} \varepsilon \leq \frac{1}{m} \sum_{i=1}^m \beta_i \leq \sigma_{max}^{(2)} \varepsilon \\ & \frac{1}{m} \sum_{i=1}^m \lambda_i^{(2)} \geq (1 - \rho) \times 100\% \end{aligned} \quad (3)$$

其中, $range^{(2)}$ 指第二层对应数据样本的最大值与最小值之差, $\bar{z}_i^{(2)}$, $\underline{z}_i^{(2)}$ 表示第二层所得区间结果的上限和下限; ε 是控制总体信息粒度的超参数; $\sigma_{min}^{(2)}$ 和 $\sigma_{max}^{(2)}$ 用来控制 β_i , 以使其不至于过分偏离 ε ; $\lambda_i^{(2)}$ 是类似 λ_i 的标识变量, 即当第二层所得区间覆盖样本的数据点时为 1, 否则为 0;

(2) 第一层

与第二层仅处理一个优化问题不同, 第一层要计算合计为 m 个的一系列优化问题, 其中任意一个表述如下:

$$\begin{aligned} \max \quad & range_i^{(1)} - \frac{1}{n_i} \sum_{j=1}^{n_i} |\bar{z}_{i,j}^{(1)} - \underline{z}_{i,j}^{(1)}| \\ \text{s.t.} \quad & \sigma_{min}^{(1)} \beta_i \leq \frac{1}{n_i} \sum_{j=1}^{n_i} \alpha_{i,j} \leq \sigma_{max}^{(1)} \beta_i \\ & \frac{1}{n_i} \sum_{j=1}^{n_i} \lambda_{i,j}^{(1)} \geq (1 - \rho) \times 100\% \end{aligned} \quad (4)$$

其中, $range_i^{(1)}$ 指第一层各优化问题对应数据样本的最大值与最小值之差, $i = 1, 2, \dots, m$; $\bar{z}_{i,j}^{(1)}$ 和 $\underline{z}_{i,j}^{(1)}$ 分别为第一层各优化问题求解所得区间的上限和下限; $\sigma_{min}^{(1)}$ 和 $\sigma_{max}^{(1)}$ 则用来控制 $\alpha_{i,j}$ 不过分偏离 β_i ;

步骤 4: 长期区间预测

基于粒度计算的长期区间预测本质上是预测模糊隶属度, 即 $\hat{\mathbf{u}}_k = f(\mathbf{u}_{k-n_l}, \dots, \mathbf{u}_{k-1})$, 其中 $\hat{\mathbf{u}}_k$ 为隶属度预测值, $\mathbf{u}_{k-n_l}, \dots, \mathbf{u}_{k-1}$ 是隶属度矩阵 $\mathbf{U}$ 的一部分, n_l 表示隶属度关系的输入个数; 将通过概率估计实现隶属度的预测; 便于理解, 以下定义及叙述以点形式给出, 首先定义粒度计算框架下的聚类中心概率 $p(\mathbf{v}_{i1}, \dots, \mathbf{v}_{in_l})$ 、数据段概率 $p(\mathbf{z}_{k-n_l}, \dots, \mathbf{z}_{k-1})$ 和共生矩阵 $\mathbf{P}$ 如下:

$$p(v_{i1}, \dots, v_{in_l}) \doteq \frac{\sum_{j=1}^N \mathbb{O}_j}{N} \quad (5)$$

$$p(z_{k-n_l}, \dots, z_{k-1}) \doteq \{p(v_{i1}, \dots, v_{in_l}) | \{i1, \dots, in_l\} \in [1, c]\} \quad (6)$$

$$\mathbf{P} \doteq \{p(v_i | v_{j1}, \dots, v_{jn_l}) | i = 1, \dots, c; \{j1, \dots, jn_l\} \in [1, c]\} \quad (7)$$

其中, $\{v_{i1}, \dots, v_{in_l}\}$, $\{i1, i2, \dots, in_l\} \in [1, c]$, 是聚类中心矩阵 $\mathbf{V}$ 中的一部分;
 $z_{k-n_l}, \dots, z_{k-1}$ 是数据片段集合 $\mathbf{Z}$ 中的一部分; $\mathbb{O}_j$ 是标识型变量, 即考虑
 $\{u_{h1j} = \max(u_{h1}), \dots, u_{hn_lj} = \max(u_{hn_l})\}$, $\{u_{h1j}, u_{h2j}, \dots, u_{hn_lj}\}$ 分别是隶属
 度 $u_{h1}, u_{h2}, \dots, u_{hn_l}$ 中的元素, 且 $\{h1, h2, \dots, hn_l\} \in [1, c]$; 当满足 $h1 =$
 $i1, \dots, hn_l = in_l$ 时为 1, 否则为 0; $p(v_i | v_j)$ 是条件概率, 即当 $\{z_{k-n_l}, \dots, z_{k-1}\}$ 的
 最大隶属度出现在聚类中心集合 $\{v_{j1}, \dots, v_{jn_l}\}$ 时, z_k 的最大隶属度出现在 v_i 的概
 率;

基于上述定义, 数据段 z_k 的概率估计为 $\hat{p}(z_k) = p(z_{k-n_l}, \dots, z_{k-2}, z_{k-1})^T \times$
 $\mathbf{P}$, 对应预测值 $\hat{z}_k$ 由中心法得到, 即:

$$\hat{z}_k = \frac{\sum_{i=1}^c \hat{p}_i V}{\sum_{i=1}^c \hat{p}_i} \quad (8)$$

其中, $\hat{p}_i$ 是 $\hat{p}(z_k)$ 中元素;

步骤 5: 模型结构参数的强化学习

将多层次粒度计算的模型结构确定问题看做单步马尔科夫决策过程, 进而
 采用蒙特卡洛方法, 强化学习包括 m 和 n_i , $i = 1, 2, \dots, m$ 在内的结构参数; 首先
 定义状态 $\mathbf{S}$ 、动作 $\mathbf{A}$ 和奖赏 $\mathbf{R}$ 如下:

$\mathbf{S}$ 为确定的多层次粒度计算模型结构;

$\mathbf{A}$ 为变更参数 m 和 n_i , $i = 1, 2, \dots, m$;

$\mathbf{R}$ 为所得预测区间的 *spec*;

由于待确定结构参数数量较庞大, 采用梯度下降近似函数策略, 学习策略
 值函数 $\pi_\omega(s, a)$; 设 $\pi_\omega(s, a)$ 为多层感知机神经网络:

$$\pi_{\omega}(s, a) = f(\omega^T \cdot \phi(s, a) + b) \quad (9)$$

其中， $\phi(s, a)$ 是表征状态 - 动作对的特征向量，定义为 $\phi(s, a) = (m, n_1, n_2, \dots, n_m)^T$ ； b 表示多层感知机的偏置量， f 表示激活函数，本发明采用的是 sigmoid 函数；

定义可导的性能函数如下：

$$J(\omega) \doteq q_{\pi_{\omega}}(s_0) \quad (10)$$

其中， $q_{\pi_{\omega}}$ 是 $\pi_{\omega}(s, a)$ 的真值函数， s_0 是起始状态；对 $J(\omega)$ 求关于 ω 的梯度并应用策略梯度定理，最终得出 ω 的更新公式为：

$$\omega_{t+1} = \omega_t + \tau \gamma \nabla_{\omega_t} \ln \pi_{\omega_t}(s, a) r_t \quad (11)$$

其中， τ 表示变化步长， γ 是贴现因子， r_t 是 t 时刻获得的奖赏，即：

$$r_t = \frac{1}{T} \sum_{i=1}^T |\bar{z}_i - \underline{z}_i| \quad (12)$$

设结构参数的搜索空间维度为 L ，模型参数确定过程总结如下：

(1) 初始化 $\tau > 0$ ， $\gamma > 0$ ， $\omega \in \mathbb{R}^{max(m)}$ ，在 L 中取 l 个结构参数样本作为训练子集；

(2) 取 $t = 1 \sim l$ ，利用步骤四表述过程循环先后计算 (12) 和 (11)，获得学习到的策略值函数参数 ω_{opt} ，考虑到问题的相互独立性，此步中 (12) 的计算借助并行策略加速；

(3) 取 $t = 1 \sim l$ ，计算 (9)，选取其中策略值函数最大值对应的特征向量 $\phi_{opt}(s, a)$ 为最优结构参数，即：

$$\phi_{opt}(s, a) = \arg \max_{t=1,2,\dots,L} (\pi_{\omega_{opt}}(s_t, a_t)) \quad (13)$$

(4) 利用获得最优结构参数，再次计算 (3) - (8)，即获得最终的长期区间预测结果。

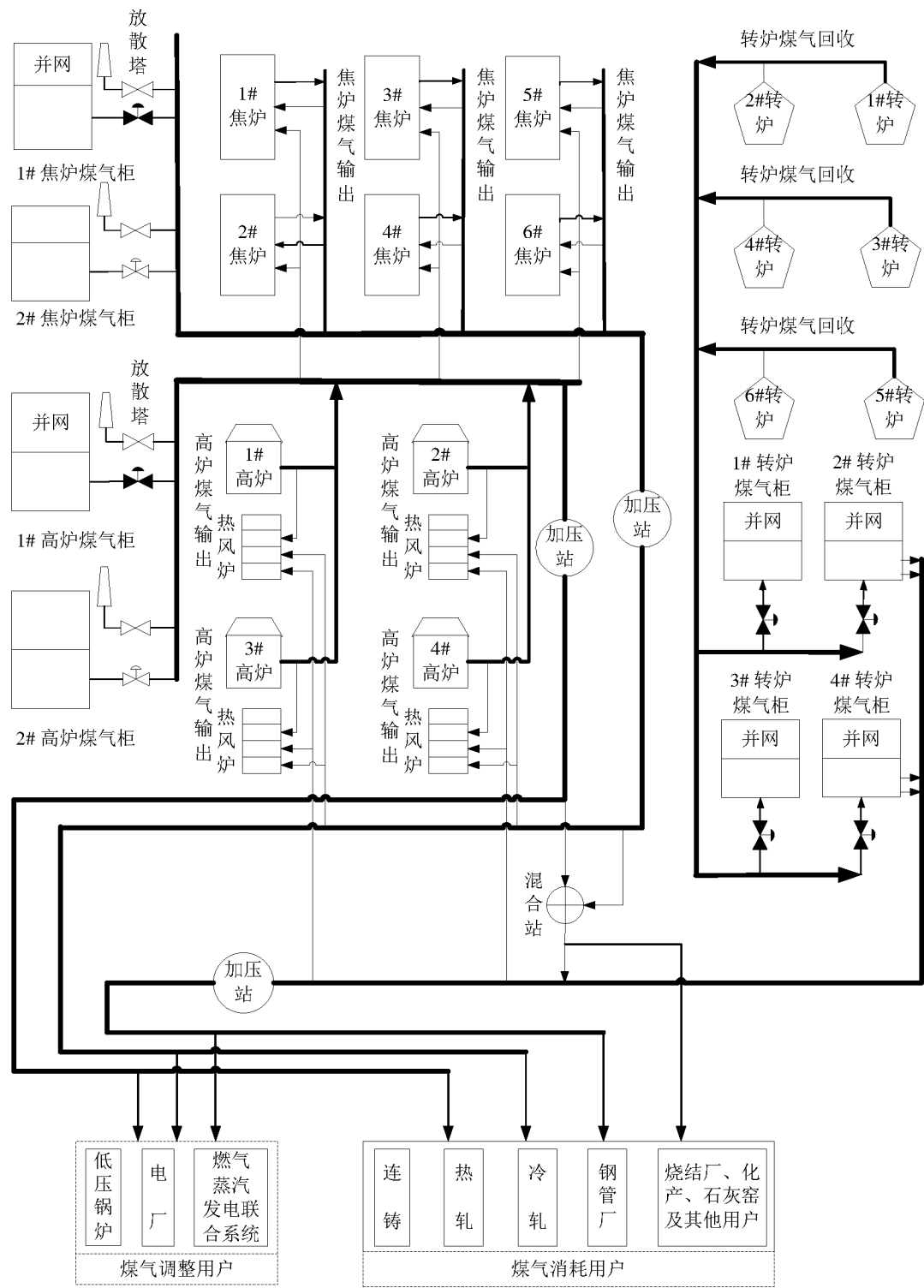


图 1

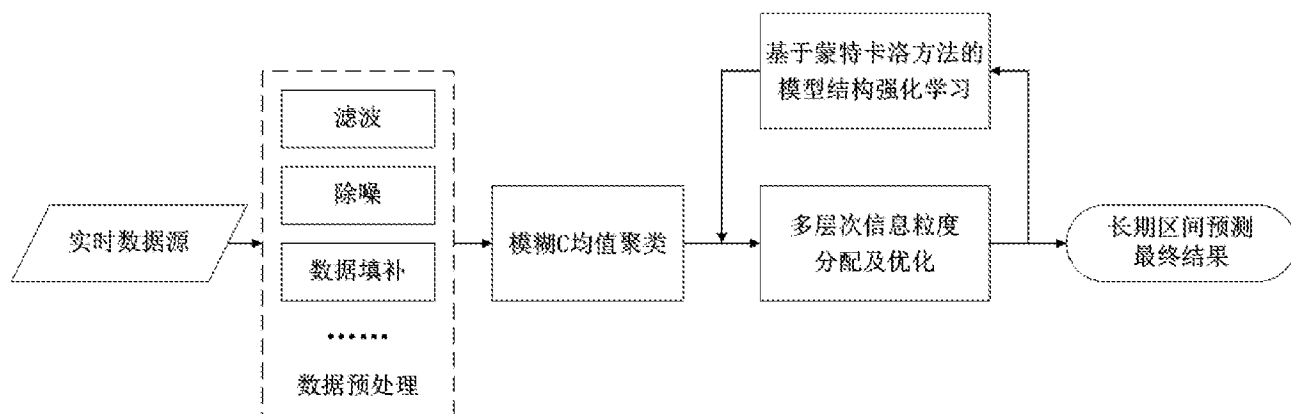


图 2

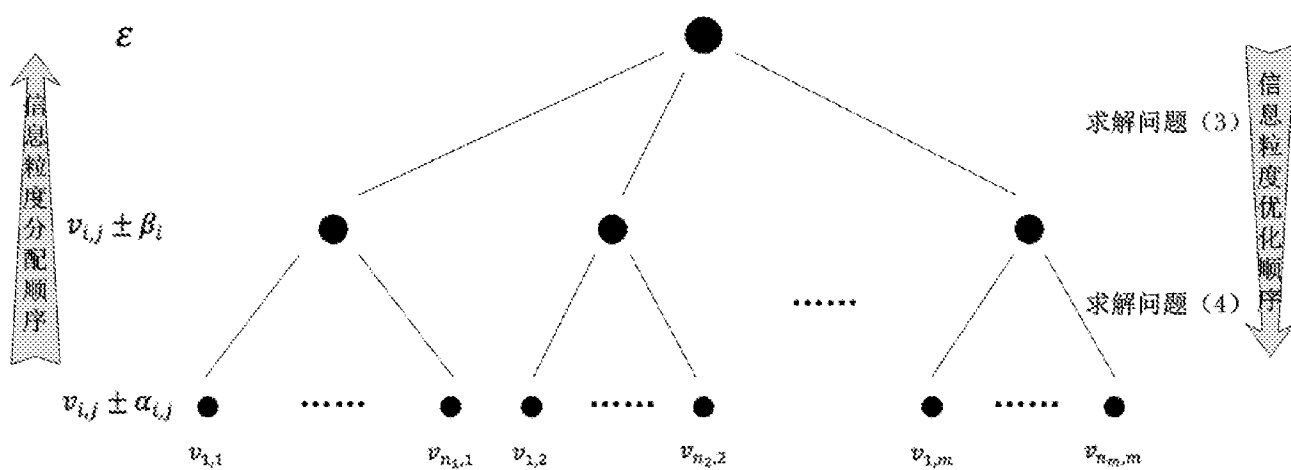


图 3

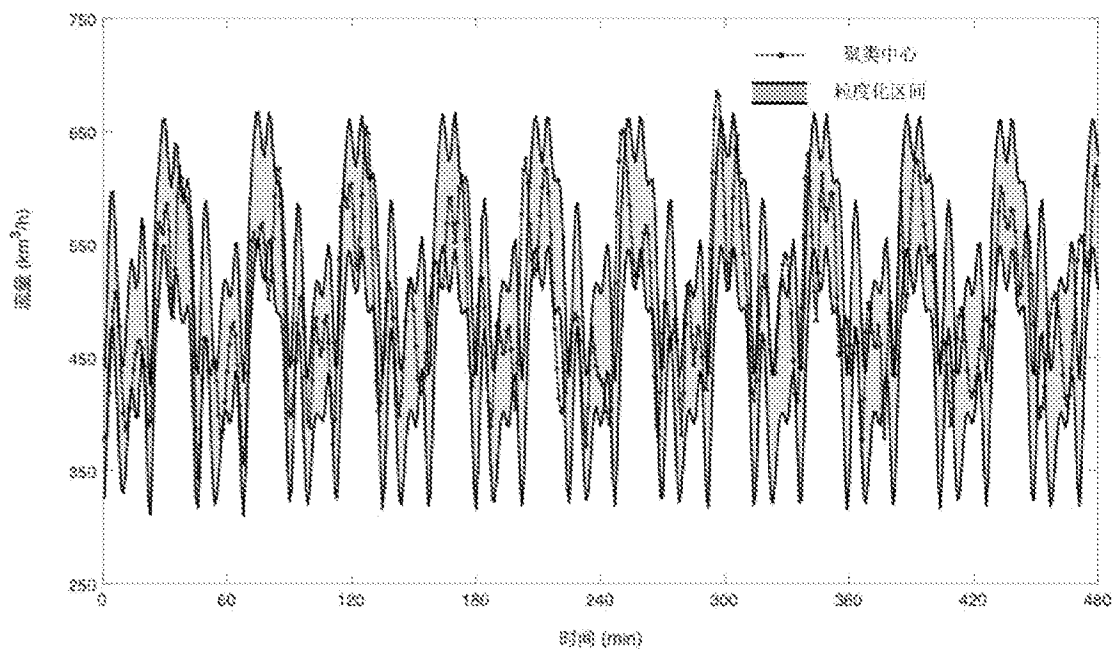


图 4 (a)

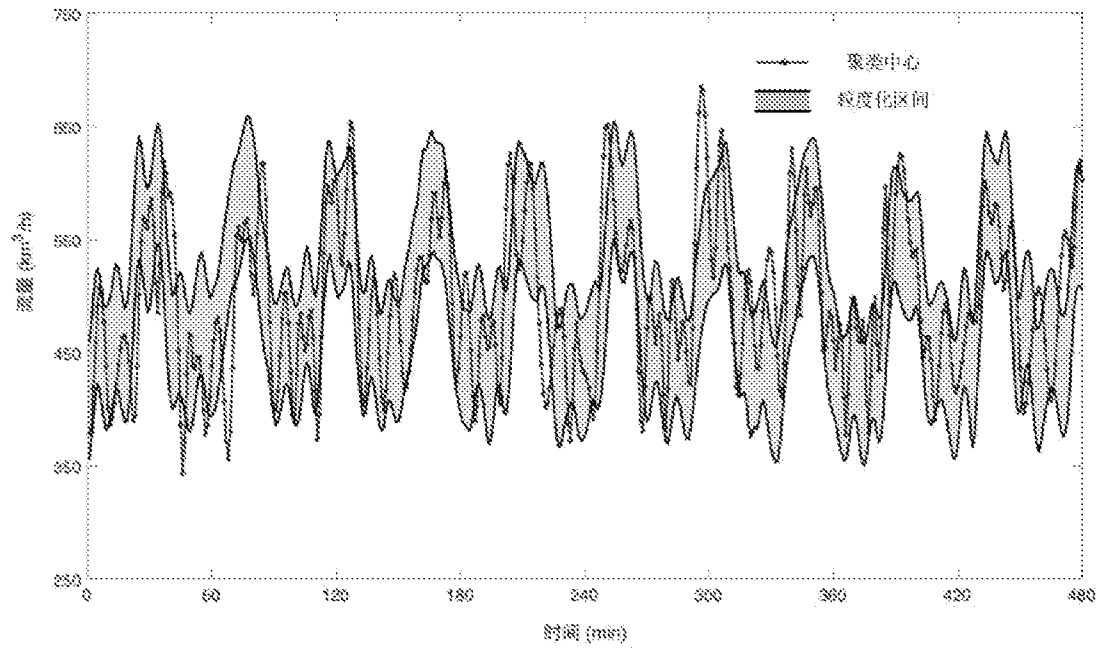


图 4 (b)

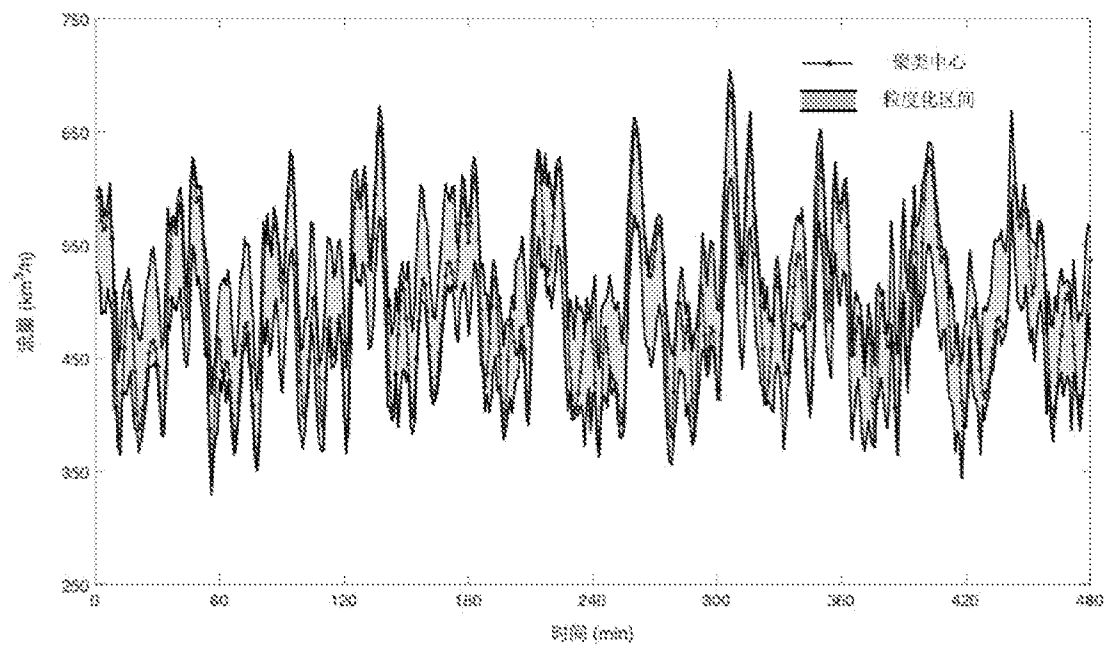


图 4 (c)

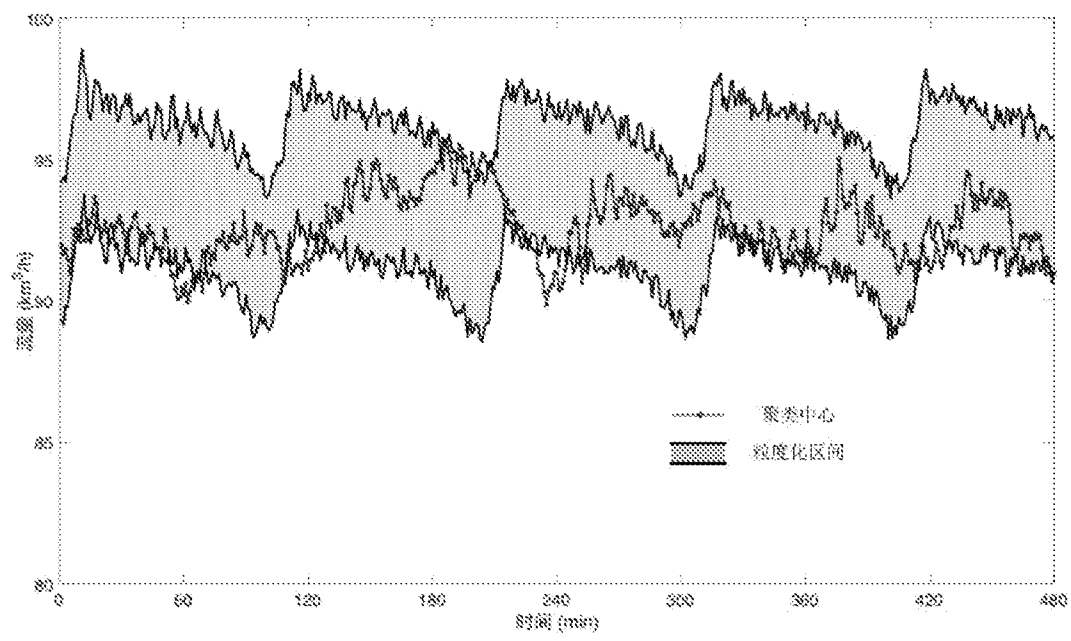


图 5 (a)

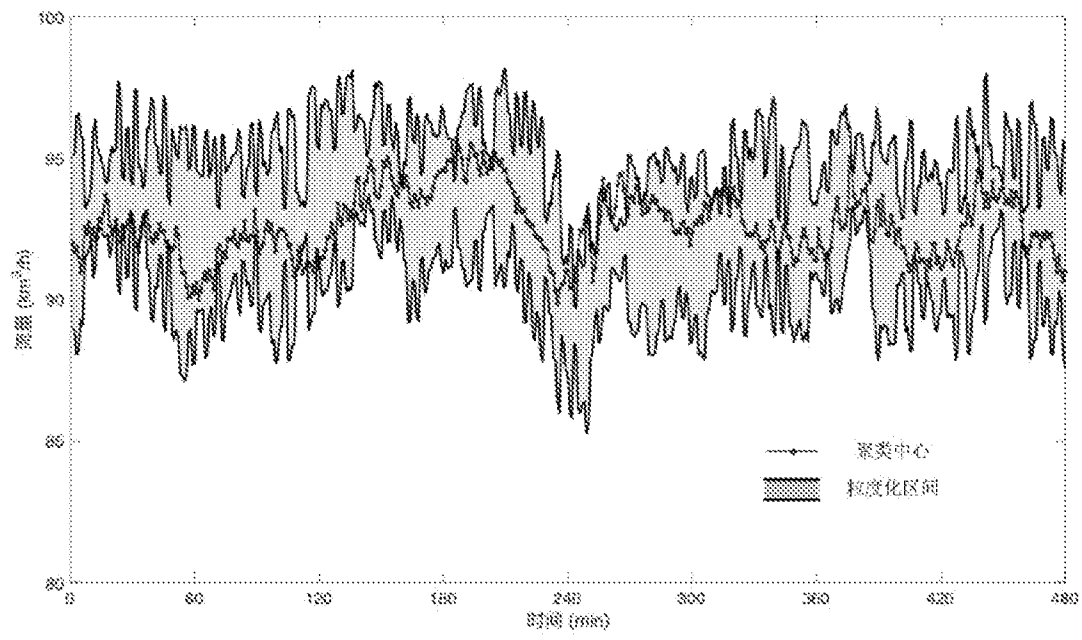


图 5 (b)

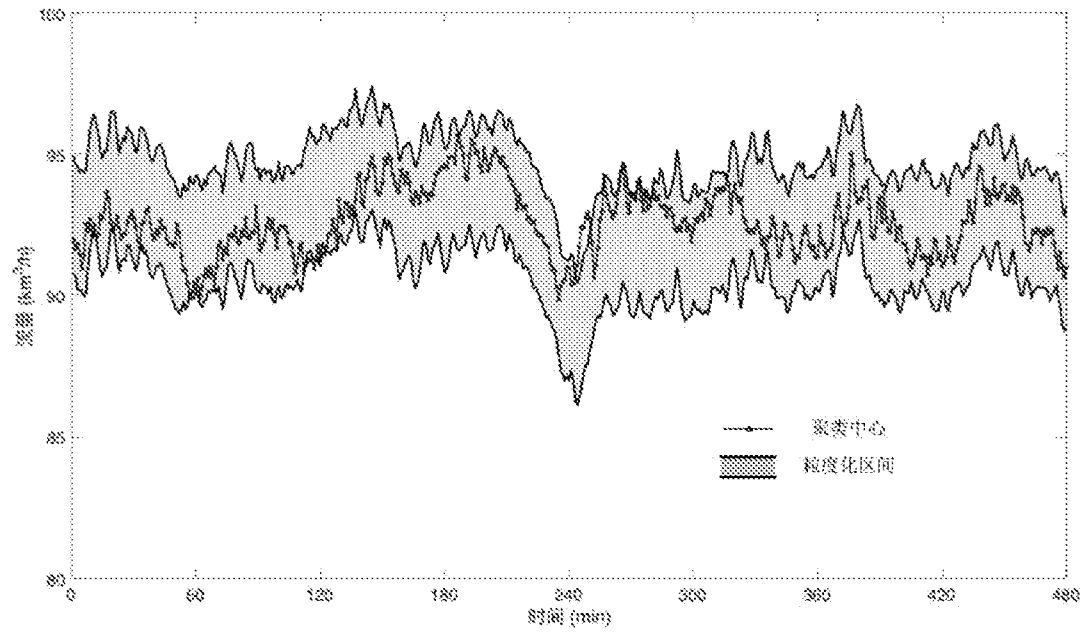


图 5 (c)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/105197

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/04(2012.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT; EPODOC; WPI; CNKI; IEEE: 韩中洋, 王伟, 赵瑀, 王霖青, 预测, 区间, 长期, 钢铁, 煤气, 聚类, 层次, 粒度, 隶属度, 马尔科夫, prediction, interval, long-term, steel, gas, cluster, layer, granularity, markov

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106779384 A (DALIAN UNIVERSITY OF TECHNOLOGY ET AL.) 31 May 2017 (2017-05-31) description, paragraphs [0005]-[0058]	1
A	CN 103473469 A (NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS) 25 December 2013 (2013-12-25) entire document	1
A	CN 103426035 A (ZHEJIANG UNIVERSITY) 04 December 2013 (2013-12-04) entire document	1
A	CN 103514486 A (SHANGHAI BAOSIGHT SOFTWARE CO., LTD.) 15 January 2014 (2014-01-15) entire document	1
A	US 7013249 B1 (KINDER MORGAN, INC.) 14 March 2006 (2006-03-14) entire document	1



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

29 May 2019

Date of mailing of the international search report

13 June 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/ CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/105197

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106779384	A	31 May 2017	None			
CN	103473469	A	25 December 2013	CN	103473469	B	22 June 2016
CN	103426035	A	04 December 2013	None			
CN	103514486	A	15 January 2014	None			
US	7013249	B1	14 March 2006	US	2006100834	A1	11 May 2006
				US	7231331	B2	12 June 2007

国际检索报告

国际申请号

PCT/CN2018/105197

A. 主题的分类 G06Q 10/04 (2012.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																				
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06Q 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, EPDOC, WPI, CNKI, IEEE: 韩中洋, 王伟, 赵琚, 王霖青, 预测, 区间, 长期, 钢铁, 煤气, 聚类, 层次, 粒度, 隶属度, 马尔科夫, prediction, interval, long-term, steel, gas, cluster, layer, granularity, markov																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 106779384 A (大连理工大学 等) 2017年 5月 31日 (2017 - 05 - 31) 说明书第[0005]-[0058]段</td> <td>1</td> </tr> <tr> <td>A</td> <td>CN 103473469 A (南京航空航天大学) 2013年 12月 25日 (2013 - 12 - 25) 全文</td> <td>1</td> </tr> <tr> <td>A</td> <td>CN 103426035 A (浙江大学) 2013年 12月 4日 (2013 - 12 - 04) 全文</td> <td>1</td> </tr> <tr> <td>A</td> <td>CN 103514486 A (上海宝信软件股份有限公司) 2014年 1月 15日 (2014 - 01 - 15) 全文</td> <td>1</td> </tr> <tr> <td>A</td> <td>US 7013249 B1 (KINDER MORGAN, INC.) 2006年 3月 14日 (2006 - 03 - 14) 全文</td> <td>1</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 106779384 A (大连理工大学 等) 2017年 5月 31日 (2017 - 05 - 31) 说明书第[0005]-[0058]段	1	A	CN 103473469 A (南京航空航天大学) 2013年 12月 25日 (2013 - 12 - 25) 全文	1	A	CN 103426035 A (浙江大学) 2013年 12月 4日 (2013 - 12 - 04) 全文	1	A	CN 103514486 A (上海宝信软件股份有限公司) 2014年 1月 15日 (2014 - 01 - 15) 全文	1	A	US 7013249 B1 (KINDER MORGAN, INC.) 2006年 3月 14日 (2006 - 03 - 14) 全文	1
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
A	CN 106779384 A (大连理工大学 等) 2017年 5月 31日 (2017 - 05 - 31) 说明书第[0005]-[0058]段	1																		
A	CN 103473469 A (南京航空航天大学) 2013年 12月 25日 (2013 - 12 - 25) 全文	1																		
A	CN 103426035 A (浙江大学) 2013年 12月 4日 (2013 - 12 - 04) 全文	1																		
A	CN 103514486 A (上海宝信软件股份有限公司) 2014年 1月 15日 (2014 - 01 - 15) 全文	1																		
A	US 7013249 B1 (KINDER MORGAN, INC.) 2006年 3月 14日 (2006 - 03 - 14) 全文	1																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																				
国际检索实际完成的日期 2019年 5月 29日		国际检索报告邮寄日期 2019年 6月 13日																		
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 马丽莉 电话号码 86-(10)-53961409																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/105197

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	106779384	A	2017年 5月 31日	无	
CN	103473469	A	2013年 12月 25日	CN 103473469	B 2016年 6月 22日
CN	103426035	A	2013年 12月 4日	无	
CN	103514486	A	2014年 1月 15日	无	
US	7013249	B1	2006年 3月 14日	US 2006100834	A1 2006年 5月 11日
				US 7231331	B2 2007年 6月 12日