

控制与决策

Control and Decision

基于云模型的煤矿安全大数据多粒度表示方法及应用

代劲, 张磊, 王国胤

引用本文:

代劲, 张磊, 王国胤. 基于云模型的煤矿安全大数据多粒度表示方法及应用[J]. *控制与决策*, 2021, 36(10): 2359–2368.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.0325>

您可能感兴趣的其他文章

Articles you may be interested in

[基于云模型和多层权重求解的多粒度语言大群体决策方法](#)

Multi-granularity linguistic large group decision-making based on cloud model and multi-layer weight determination

控制与决策. 2021, 36(9): 2257–2266 <https://doi.org/10.13195/j.kzyjc.2020.0102>

[基于正态云模型的状态转移算法求解多目标柔性作业车间调度问题](#)

State transition algorithm based on normal cloud model for solving multi-objective flexible job shop scheduling problem

控制与决策. 2021, 36(5): 1181–1190 <https://doi.org/10.13195/j.kzyjc.2019.1233>

[基于知识粒度特征的多目标粗糙集属性约简算法](#)

Multi objective rough set attribute reduction algorithm based on characteristics of knowledge granularity

控制与决策. 2021, 36(1): 196–205 <https://doi.org/10.13195/j.kzyjc.2019.0490>

[结合注意力机制的循环神经网络复述识别模型](#)

Recurrent neural networks based paraphrase identification model combined with attention mechanism

控制与决策. 2021, 36(1): 152–158 <https://doi.org/10.13195/j.kzyjc.2019.0638>

[基于联合知识表示学习的多模态实体对齐](#)

Multi-modal entity alignment based on joint knowledge representation learning

控制与决策. 2020, 35(12): 2855–2864 <https://doi.org/10.13195/j.kzyjc.2019.0331>

基于云模型的煤矿安全大数据多粒度表示方法及应用

代 劲^{1,2}, 张 磊¹, 王国胤^{1†}

(1. 重庆邮电大学 计算智能重庆市重点实验室, 重庆 400065; 2. 重庆邮电大学 软件工程学院, 重庆 400065)

摘 要: 以大规模物联网为支撑的新一代信息技术的深入应用, 为基于海量大数据挖掘的煤矿安全知识发现提供了实现的可能. 现有的针对煤矿安全大数据的研究大多基于定量模型, 其解决问题的角度单一且一定程度上忽略了煤矿监管中多时空、多粒度的管控需求, 使得数据中蕴含的煤矿风险知识未得到客观、全面的发现. 从煤矿监管中的多粒度需求出发, 借助云模型定量数据与定性概念间良好的转换能力, 从煤矿监管中的时间、空间监管架构角度, 提出基于自适应混合云变换的面向煤矿安全大数据的多粒度表示方法. 该方法能够有效满足煤矿监管中其基于宏观、微观, 不同时间、空间维度的变粒度需求, 实现煤矿安全大数据在不同粒度认知结构中特性的深入挖掘. 通过在煤矿数据概念提取中的应用并与高斯云变换算法对比, 其提取的概念覆盖度更全且更客观, 验证了所提方法的合理性; 在煤矿监测数据预测应用中, 其预测精度相较于 ARIMA 算法更高, 验证了所提方法的可行性.

关键词: 煤矿安全; 大数据; 知识发现; 云模型; 多粒度; 概念提取

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.0325

开放科学(资源服务)标识码(OSID):



引用格式: 代劲, 张磊, 王国胤. 基于云模型的煤矿安全大数据多粒度表示方法及应用 [J]. 控制与决策, 2021, 36(10): 2359-2368.

Multi-granularity representation method of big data in coal mine safety based on cloud model and its application

DAI Jin^{1,2}, ZHANG Lei¹, WANG Guo-yin^{1†}

(1. Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: The in-depth application of information technology supported by the large-scale internet of things provides the possibility for the discovery of coal mine safety knowledge based on massive data mining. The existing researches on big data of coal mine safety are mostly based on quantitative models, which solve problems with a single perspective and neglect the multi-granularity management needs in coal mine supervision to some extent, so that the risk knowledge of coal mine contained in the data has not been found objectively and comprehensively. From the view of multi-granularity demand in coal mine supervision, with the advantage that cloud models can have a good conversion between quantitative data and qualitative concepts, a multi-granularity representation method for coal mine safety big data based on adaptive hybrid cloud transformation is proposed. This method can effectively meet the variable granularity requirements of coal mine supervision based on macro, micro, different time and space dimensions, and realize the deep mining of the characteristics of coal mine safety big data in different granularity cognitive structure. Through the application in the concept extraction of coal mine data and the comparison with the Gaussian cloud transformation algorithm, the concept coverage of the extraction is more complete and more objective, which verifies the rationality of the proposed method. In the application of coal mine monitoring data prediction, the prediction accuracy is higher than that of the ARIMA algorithm, which verifies the feasibility of the proposed method.

Keywords: coal mine safety; big data; knowledge discovery; cloud model; multi-granularity; concept extraction

0 引 言

我国煤炭分布范围广泛, 埋藏地质条件复杂. 长期以来, 煤炭工业是我国安全事故发生最严重的行

业, 随着我国对煤矿监管监察力度的加强, 数字矿山的智能化也在稳步发展. 在矿业的信息化建设过程中, 虽然积累了海量的煤矿安全生产数据, 但基于安

收稿日期: 2020-03-20; 修回日期: 2020-06-30.

基金项目: 国家重点研发计划项目(2018YFC0808305).

责任编委: 阳春华.

†通讯作者. E-mail: wanggy@cqupt.edu.cn.

全生产方面相应的方法或模型较少,无法为煤矿安全管理提供可靠的决策信息支撑^[1].如何对这些海量的煤矿风险数据资源进行分析和处理,发现其中潜在、未知的知识和规律,是目前急需解决的问题.近年来,越来越多的学者开始投入到针对煤矿安全生产数据的研究中来,并取得了较好的应用效果.现有的针对煤矿安全生产数据的研究按照其应用场景可主要分为以下3类:

1) 煤矿数据建模与降维: Hua等^[2]对煤矿安全监管历史数据进行信息挖掘,提出了煤矿安全监管历史数据的相似度分析模型,有效地实现了煤矿监管监测数据降维,并保证了较高的精度,为矿山监测数据挖掘提供了一种新的方法. Yilmaz等^[3]给出了基于煤矿数据集建模的经典方法,其通过研究发现采用二分量混合指数分布模型(2MED)对煤矿数据集有着更好的解释性,并利用该模型在191个煤矿监测数据集上进行了验证,表明其模型良好的适用性.

2) 煤矿数据不确定性挖掘: Zhou等^[4]通过粗糙集理论对煤矿监测数据领域中的不完整、不精确和不确定性知识进行定量分析,解决了其煤矿监测数据信息不完整、不确定情况下的推理与决策问题,并设计了相关系统将其应用到地下瓦斯涌出的数据挖掘分析中,取得了良好效果. 邵良杉^[5]针对煤矿瓦斯灾害的特点,提出了用粗糙集理论对瓦斯灾害进行预测,通过实际应用证实了粗糙集理论在瓦斯灾害预测中的有效性和实用性. 王军号等^[6]针对煤矿安全监控的复杂性和不确定性,运用置信距离测度与采集数据的时间戳相结合的动态限幅滤波算法对数据进行预处理以消除误差,有效地消除了干扰数据等特征,在安全监测领域有着良好的实用性.

3) 煤矿安全评价与分析: Lu等^[7]分析了煤矿数据具有数据量大、信息量大、动态、随机、模糊等特点,强调数据挖掘技术在煤矿安全中的应用是提高煤矿安全工作认识和理解能力的重要途径;在此基础上,提出了煤矿安全数据挖掘系统原型结构,根据煤矿安全管理的特点,设计了相应的数据挖掘方法并基于此开发了煤矿安全预警系统软件. 邢玉忠等^[8]通过对瓦斯监测数据的分析,建立了矿井通风系统合理性评价模型,通过对实际矿井通风系统的合理性评价,取得了良好的应用效果. Shao等^[9]基于煤矿安全信息动态变化及与空间密切相关等特征,基于空间数据挖掘和GIS技术构建了煤矿安全监测系统,其一定

程度上减少了煤矿灾害的发生,提高了煤矿安全管理水平. Sun等^[10]发现煤矿监测数据实质为数据流且随着环境变化其隐含着概念漂移这一特点,基于随机决策树模型采用Hoeffding界限不等式和信息熵代替随机选择来确定分界点,进一步检测其概念漂移. 实验结果表明,对煤矿数据流具有较好的分类精度,为煤矿安全评价提供了一种新的实用方法.

上述对煤矿数据的挖掘普遍在固定粒度层面展开,且其对定量分析与定性评价之间的相互转换关系缺乏研究,导致定性的煤矿风险评估难以有效指导定量的生产实践工作. 本文借助粒计算中的多粒度思想,对原始煤矿数据进行定量到定性、细粒度到粗粒度间的转换,抽象出较高的概念层,可在煤矿安全管理和决策分析时,提供更有指导意义的知识. 云模型^[11]是粒计算中定量到定性转换的经典模型,本研究主要基于云模型相关理论,开展面向煤矿安全大数据的多粒度表示方法的研究.

1 相关理论

多粒度计算的思想来源于Hobbs所提出的“人类问题求解的基本特征之一就是多粒度计算的思想,具有从不同的粒度上观察世界,且很容易地从一个抽象层次转换到其他层次的能力,即分层次地处理它们”^[12]. 其进行多粒度计算的目的是为了降低处理复杂问题的复杂性,能够在复杂问题求解、大数据挖掘和不确定性信息处理等问题上提供有力帮助.

1.1 多粒度表示方法

现有的多粒度表示方法主要基于粒计算相关模型,其大体分为两大类:一类以处理不确定性为主要目标,如以模糊处理为基础的计算模型,以粗糙集为基础的模型;另一类则以多粒度计算为主要目标,如商空间理论. 这两类的侧重点有所不同,前者在粒化过程中更侧重于计算对象的不确定性处理. 以云模型为代表的粒计算模型介于两类模型之间,其既能够很好地处理不确定性数据间的定性与定量间的相互转换问题,又能够实现数据论域间粒度的抽象与细化,能够很好地适应煤矿安全大数据中的定性与定量共存、多元表征等相关问题.

1.2 基于云模型的多粒度表示方法

云模型是定性定量间的转换模型,并可通过云发生器实现定性概念(概念内涵)和定量数据(概念外延)之间的相互转换. 云模型理论发展20余年来,在

理论研究上不断被完善,云模型发生器、逆向云算法、云规则发生器、云变换和云模型的粒计算等理论相继被提出^[13].同时,云模型在数据挖掘、系统评测、智能控制等领域的应用都得到了不错的效果^[14-18].本节主要介绍云模型的相关理论及算法.

1.2.1 云模型

云模型用期望 Ex 、熵 En 、超熵 He 作为数字特征表示定性概念:期望值 Ex 为最能代表当前定性概念的数值,反映了相应的定性知识的信息中心值;熵 En 用于度量定性概念的亦此亦彼性,体现其模糊性,反映了论域空间中可被概念接受的云滴的取值范围;超熵 He 为熵的熵,反映数值隶属于定性概念的随机程度,超熵的大小间接地反映云的厚度,超熵越大,云的厚度越大^[13].云滴的数字特征如图1所示,其纵坐标轴为云滴对定性概念的确定度 μ ,表示当前云滴对其概念的隶属程度.

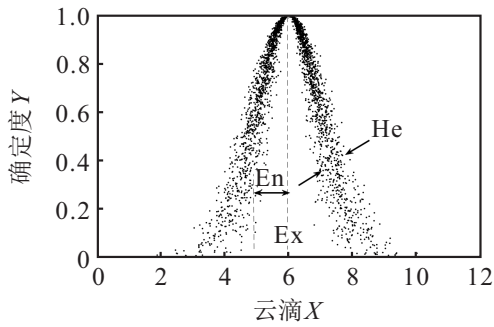


图1 云的数字特征

1.2.2 云规则发生器

云规则发生器由前件云和后件云发生器组合而成,通过给定一个前件云作为规则的前件,给定一个后件云作为规则的后件来完成不确定性推理.

算法1 一维前件云发生器(FCLG)^[19].

输入: 一维定性概念的数字特征 (Ex, En, He) 及定量值 x ;

输出: 定量值 x 属于的确定度 y .

step 1: 生成以 En 为期望值、 He 为均方差的正态随机数 En' ;

step 2: 计算确定度 $y = e^{-\frac{(x-Ex)^2}{2En'^2}}$.

算法2 一维后件云发生器(BCLG).

输入: 一维定性概念的数字特征 (Ex, En, He) 及确实度 $y, y \in [0, 1]$;

输出: 满足确定度 y 的定量值 x .

step 1: 生成以 En 为期望值、 He 为均方差的正态随机数 En' ;

step 2: 计算定量值 $x = Ex \pm En' \times \sqrt{-2 \ln y}$.

1.2.3 高斯云变换

高斯云变换^[20]是一种基于云模型的连续数据离散化方法,能够将问题域中的数据分布自动转化为多粒度的不同概念.其算法如下.

算法3 适应高斯云变换算法(A-GCT).

输入: 数据集样本集 $\{x_i | i = 1, 2, \dots, N\}$,概念含混度上限 β ;

输出: 由高斯云模型表示的定性概念.

step 1: 概念数量初始化. 统计数据样本的频度分布 $p(x_i)$ 的波峰数量作为概念数量初始值.

step 2: 高斯混合聚类. 调用高斯混合模型(GMM)将其转换为 M 个高斯分布 $G(\mu_k, \delta_k) | k = 1, 2, \dots, m$.

step 3: 计算高斯云参数. 对于第 k 个高斯分布,计算其标准差缩放比 α_k ,则第 k 个表征概念的云参数为 $Ex_k = \mu_k, En_k = (1 + \alpha_k)\sigma_k/2, He_k = (1 - \alpha_k)\sigma_k/6, CD_k = (1 - \alpha_k)(1 + \alpha_k)$,得到 M 个高斯云 $C(Ex_k, En_k, He_k), k = 1, 2, \dots, m$.

step 4: 概念数计算. 按顺序对每个高斯云的划分混乱指数 CD_k 进行判断,如果 $CD_k > \beta, k = 1, 2, \dots, M$,则概念数 $M = M - 1$.

step 5: 循环迭代. 循环step 3 ~ step 5,形成 m 个含混度小于等于 β 的高斯 $C(Ex_k, En_k, He_k), k = 1, 2, \dots, m$.

step 6: 输出. 按含混度对 M 个高斯云进行排序输出.

2 面向煤矿安全大数据的多粒度表示

煤矿安全大数据为煤矿作业生产时传感器监测节点所记录的动态监测数据,安全管理时记录的隐患数据和矿区生产相关的统计数据等,其中以动态监测数据体量最大,具有高时空维度(工作面、采区、矿井关系,区域内多个矿井)、定性与定量共存、灾害分区特征明显、多元表征等特点.本文主要针对煤矿传感器的动态监测数据,基于其高时空维度,定性与定量共存等特点,开展面向煤矿安全大数据多粒度表示方法的相关研究.

2.1 煤矿安全大数据多粒度表示框架

煤矿监管体系中,不同的粒度对风险态势的管控需求不同.比如,从宏观粒度的认知角度,对煤矿的态势认知只需关注不同程度风险的分布情况;从微观

粒度的认知角度,需要发现不同时间维度和空间维度中的数据特性. 基于此,为满足煤矿监管中的多粒度、变粒度需求,借助云发生器、云变换等相关理论,提出

宏观、微观、时间、空间的概念表示方法,实现煤矿安全大数据的定性与定量间转换,同时满足安全监管的变粒度需求. 整体框架如图2所示.

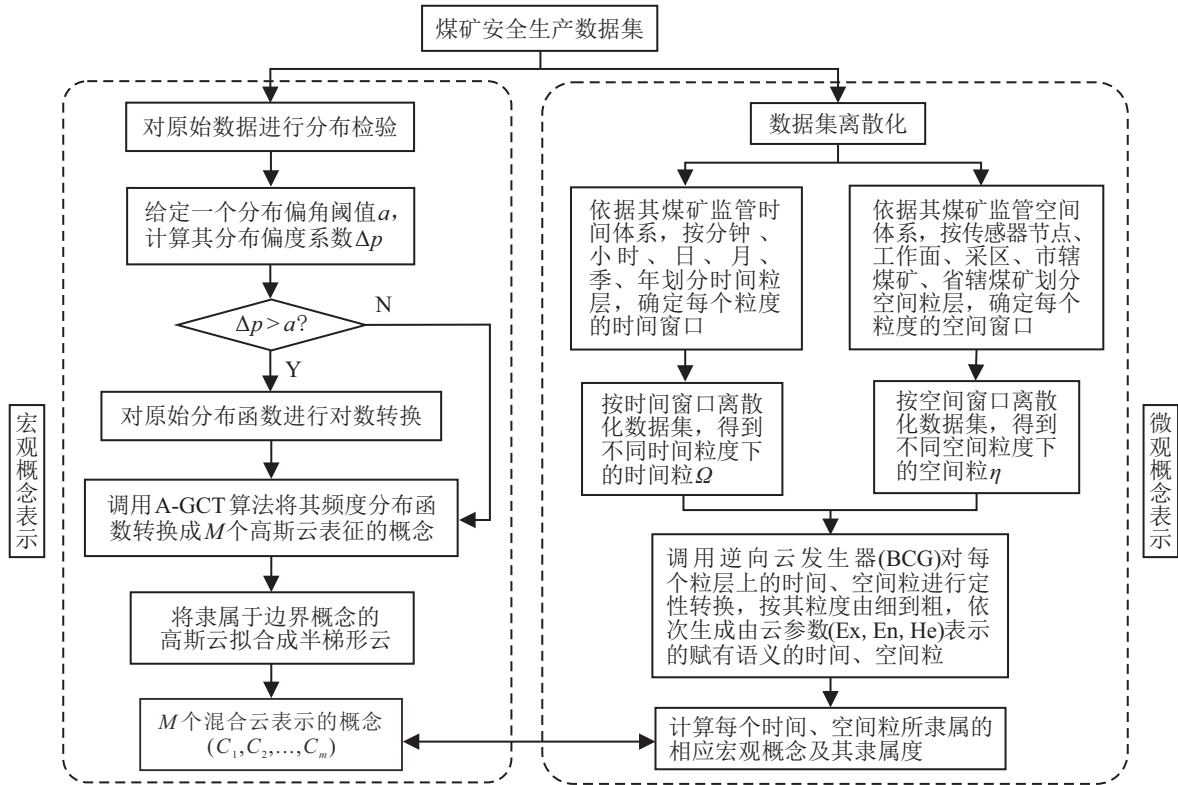


图2 煤矿大数据多粒度表示方法整体框架

2.2 煤矿安全数据宏观概念表示

考虑到高斯云变换在对整体论域的概念提取上有着显著优势,在煤矿数据的宏观概念表示上,主要基于高斯云变换算法,针对煤矿数据的实际分布情况对其进行改进.

高斯云变换算法强调其频度分布越高对总体概念贡献越大,因此在针对煤矿这类偏态分布严重的数据集时,很容易出现概念提取不全的问题. 图3为基于高斯云变换对煤矿瓦斯浓度概念提取的结果.

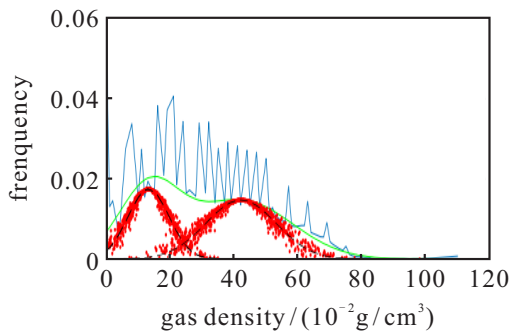


图3 基于高斯云变换的煤矿瓦斯概念提取

提取到期望浓度在0.2的低浓度概念和期望浓度在0.5的高浓度概念,在原始数据中,浓度大于1的

数据共116条,占总样本的万分之五左右,这一小部分数据对煤矿风险定性有着决定性作用. 在最终生成的概念中,由于都是高斯云表征的概念,边界概念的隶属度不符合正常的认知,浓度在0.2对于低浓度的隶属度要高于浓度在0.1对低浓度的隶属度,这意思就是浓度为0.2比浓度为0.1更属于低浓度这个概念,这显然是不符合认知的. 因此,对于论域中的边界概念,高斯云的表征适用性在一定程度上减弱了.

由于煤矿数据集普遍属于偏态分布,即随着监测值升高,样本量急速递减. 基于这个特点,先对原始的监测数据进行频度分布统计,如果其属于偏态分布,则可以在不影响整体频度分布的情况下,对原始分布函数进行对数转换,再调用高斯混合模型将问题域中的整个频度分布函数转换成多个高斯分布的叠加,在生成的高斯分布上加入区间约束,得到首尾区间的半云形态,并将其拟合成分梯形云,解决其边界隶属度的认知问题,最终形成多类型云混合的自适应概念提取方法. 该算法如下.

算法4 自适应混合云变换算法(A-MCT).

输入: 数据集样本集 $\{x_i | i = 1, 2, \dots, N\}$, 概念

含混度上限 β ,数据分布偏度阈值 γ ;

输出: 由多种云模型表示的定性概念.

step 1: 概念数量初始化. 统计数据样本的频度分布 $p(x_i)$ 的波峰数量作为概念数量初始值.

step 2: 偏度计算. 计算数据分布偏度 Δp , 如果 $\Delta p > \gamma$, 则取其频度 $p'(x_i) = \log_2(p(x_i) + 1)$.

step 3: 高斯混合聚类. 调用高斯混合模型(GMM)将其转换为 m 个高斯分布 $G(\mu_k, \delta_k)|k = 1, 2, \dots, m$.

step 4: 计算高斯云参数. 对于第 k 个高斯分布, 计算其标准差缩放比 α_k , 则第 k 个表征概念的云参数为 $Ex_k = \mu_k, En_k = (1 + \alpha_k)\sigma_k/2, He_k = (1 - \alpha_k)\sigma_k/6, CD_k = (1 - \alpha_k)(1 + \alpha_k)$, 得到 M 个高斯云 $C(Ex_k, En_k, He_k), k = 1, 2, \dots, m$.

step 5: 概念数计算. 按顺序对每个高斯云的划分混乱指数 CD_k 进行判断, 如果 $CD_k > \beta, k = 1, 2, \dots, M$, 则概念数 $M = M - 1$.

step 6: 循环迭代. 循环step 3~step 5, 形成 m 个含混度小于等于 β 的高斯 $C(Ex_k, En_k, He_k), k = 1, 2, \dots, m$.

step 7: 半云拟合. 判断当前高斯云是否位于其论域边界, 如果 $\text{loc}(C_k) = \text{loc}(C_1)$ 或 $\text{loc}(C_k) = \text{loc}(C_M)$, 则用梯形云进行拟合, $C_1 = \text{RHT_C}(C_1)$ (右半梯形云), $C_M = \text{LHT_C}(C_M)$ (左半梯形云).

step 8: 排序输出. 按含混度对 M 个混合云进行

排序输出.

2.3 煤矿安全数据微观概念表示

煤矿安全数据微观概念表示分为时间粒表示方法、空间粒表示方法.

2.3.1 时间粒表示方法

从粒计算的观点看, 时间也能够被粒化, 时间序列的离散化操作在本质上是对时间变量进行粒化^[21]. 基于煤矿安全大数据时间序列的粒化是压缩煤矿数据规模, 便于后续分析解释、建模的基础. 对于煤矿安全生产领域, 其不同时间粒度下的数据应结合矿区实际生产从不同的认知角度去表示和分析. 基于时间维度, 分别从分钟、小时、天、月、年多个粒度去进行风险态势的表示, 不同的时间粒度上, 根据生产领域的侧重点不同分别对不同粒层赋予语义表示. 针对煤矿监测数据而言, 分钟粒度上主要反映监测数值的实时变化情况, 小时粒度上主要反映当前小时段内的态势波动和趋势, 天粒度上主要反映当前一天时段内的态势分布情况, 月粒度上主要反映当月每天的态势波动, 年粒度上主要反映当年的总体浓度态势. 以此为目标, 对各个粒层的时间粒分别构建语义表示. 时间粒度的形式化表示如表1所示, 其中 α 为时间粒度水平, 其粒度的形式化表述为 Ω_i^α , 表示在 α 粒度水平下的第 i 个时间粒.

表 1 煤矿安全大数据时间粒形式化表示

α	时间窗口 w	粒层	粒度的形式化表述
0	1	分钟	$\Omega_i^0 = \{x_i, \Delta x_i\}$
1	60	小时	$\Omega_i^1 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$
2	1440	天	$\Omega_i^2 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$
3	43200	月	$\Omega_i^3 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$
4	129600	季	$\Omega_i^4 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$
5	518400	年	$\Omega_i^5 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$

基于上述时间粒度表示, 提出基于时间粒概念提取算法.

算法5 时间粒概念提取算法(T-CE).

输入: 数据集样本集 $\{x_i|i = 1, 2, \dots, N\}$, 时间窗口宽度 w ;

输出: 按时间粒度划分的概念粒及其隶属度 μ .

step 1: 数据集离散化. 基于粒化标准将数据集按 w 划分成 p 个时间片 T_1, T_2, \dots, T_p .

step 2: 云转换. 调用逆向云发生器算法(BCG)将小时粒层以上的时间片 T_i 转换成赋有语义表示的小时粒层时间粒 $\Omega_i^1 = \{C_1^i(Ex_1, En_1, He_1), \dots,$

$C_n^i(Ex_n, En_n, He_n)\}$.

step 3: 时间粒生成. 按分钟粒层(原始数据样本集 $\{x_i|i = 1, 2, \dots, N\}$), 小时粒层, 天粒层, 月粒层, 季粒层, 年粒层输出时间粒.

step 4: 风险粒判定. 判定各粒层粒的表示云 $\{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$ 的期望值 Ex 隶属于基于A_MCT提取的概念区间, 如果其隶属于相应概念的论域区间, 则计算其期望在隶属概念中的隶属度 $\mu = \exp(-(Ex' - Ex)^2/2 \times En'^2)$. 其中: Ex 为时间粒期望, Ex' 为所属的宏观概念的期望, En' 为所属宏观概念的熵.

2.3.2 空间粒表示方法

空间粒的表示主要基于矿区风险监管监察体系,依据矿区生产大数据的来源,按照空间粒度,将煤矿空间大数据按照传感器、工作面、采区、矿区、市辖所有矿区、省辖所有矿区进行空间粒度的逻辑粒化。

煤矿安全生产大数据主要依托于传感器的实时记录,因此单一传感器为整个空间粒度中的最细粒度,一个工作面包含多个传感器节点,一个采区又有着多个工作面,一个矿区有着多个采区,再依据煤矿监管监察体系,市区内多个煤矿同属市煤监局监察监管,省煤监局又辖管多个市区的煤矿。因此,依据传感器的布局分布和煤矿监管监察管理体系,完成空间粒度的多粒度表示。

$A_i^0 = \{x_1, x_2, \dots, x_n\}$ 表示第 i 个传感器中

的 n 个数据样本,设 λ 为空间粒度水平,取 $\lambda = 0$,将数据集按传感器划分为由 m 个传感器构成的空间粒 $\eta_1, \eta_2, \dots, \eta_m$,表示为处于传感器粒层。用云变换将每个传感器节点中的所有数据转换为具有定性含义的云模型,有 $\eta_i = C_1(Ex_1, En_1, He_1), \dots, C_n(Ex_n, En_n, He_n)$ 。

取 $\lambda = 1$,将数据集按照工作面划分,将同属于一个工作面的所有传感器中的数据用一个空间粒 B_i 表示, $B_i = \eta_1, \eta_2, \dots, \eta_m$ 。传感器 1 到传感器 m 隶属于同一个工作面,采用云模型对每个工作面内的数据进行云变换表示,有 $B_i = C_1(Ex_1, En_1, He_1), \dots, C_n(Ex_n, En_n, He_n)$ 。

参照传感器与工作面的粒化方法,对采区、矿区、市辖区、省辖区采用相同的粒化方法对空间粒度进行表示,其空间多粒度表示见表 2。

表 2 煤矿安全大数据空间粒的形式化表示

λ	粒层	空间粒度的形式化表述
0	传感器	$\eta_i^0 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$
1	工作面	$\eta_i^1 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$
2	采区	$\eta_i^2 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$
3	矿区	$\eta_i^3 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$
4	市辖矿区	$\eta_i^4 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$
5	省辖矿区	$\eta_i^5 = \{C_1^i(Ex_1, En_1, He_1), \dots, C_n^i(Ex_n, En_n, He_n)\}$

3 煤矿安全大数据多粒度表示仿真实验与分析

3.1 实验准备

本文提出的方法主要应用于煤矿监测数据的概念提取以及短期监测数据的预测,数据主要来源于重庆煤炭科学研究院提供的煤矿瓦斯浓度监测数据(重庆石壕煤矿,2019年1月至6月),共 209 730 条。预期在本文提出的煤矿安全大数据多粒度表示方法的基础上,能够对煤矿生产过程中的瓦斯浓度进行合理的概念识别以及对其短期内的瓦斯浓度预测达到较好的精度。其监测数据概念提取的对比算法采用自适应高斯云变换算法,监测数据预测的对比算法采用 ARIMA 算法。

本实验选用的实验平台为 Jupyter Notebook + Python 3.6 + Matlab 2016a; 操作系统是 Win10 专业版 64 位; 处理器为 Intel (R) Core (TM) i5-6500 CPU @ 3.20 GHz; 安装内存是 20.0 GB。

3.2 主要实验流程

实验主要基于本文提出的煤矿安全大数据多粒度表示方法,对瓦斯浓度进行概念提取,并对瓦斯浓

度预测进行验证。主要步骤如下:

- step1: 宏观概念表示;
- step2: 微观概念表示。

3.2.1 宏观概念表示

煤矿安全数据的宏观概念提取主要基于数据的原始分布情况,通过调用相应的云变换算法,得出相关的概念表示结果。具体流程如下:

1) 分布检验。首先对数据进行分布检验,取偏度 (Skewness) 为描述变量取值分布对称性的统计量。偏度计算公式如下:

$$\text{Skewness}(\Delta p) = \sum \left(\frac{x - E(x)}{\sqrt{D(x)}} \right)^3.$$

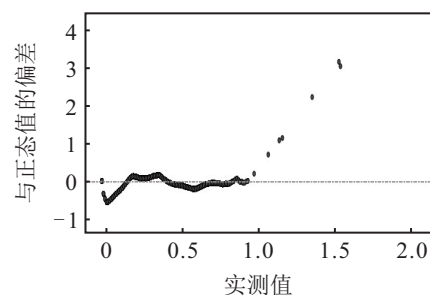


图 4 去趋势正态偏差

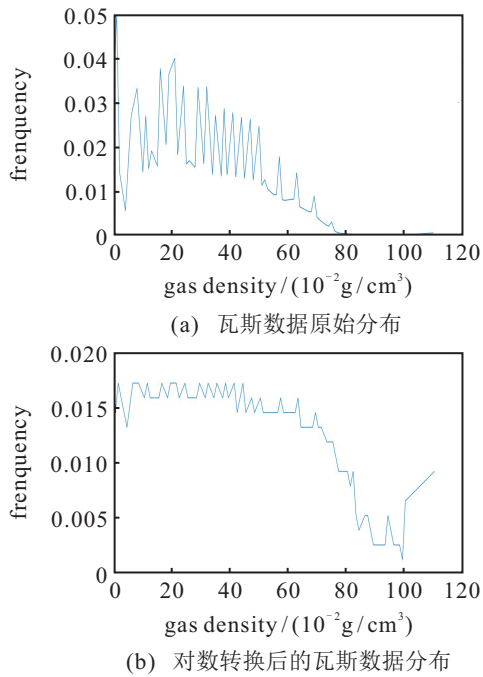


图 5 瓦斯数据分布

计算得到数据分布偏度为0.05,且通过其浓度的去趋势正态偏差图4可发现,当浓度大于1时,数据分布右偏严重.

标准正态分布偏度为0,以其标准正态分布98%的置信区间来判断当前数据分布是否属于偏态分布,将偏度阈值设为 $|0.02|$,计算其数据分布偏度,当前偏度 $\Delta p = 0.05 > 0.02$,发现其满足偏态分布.

2)对数转换.在不影响原始频度分布函数的情况下,对其进行对数转换,如图5所示.

3)概念提取.调用A-GCT和A-MCT算法对煤矿瓦斯数据进行自适应概念聚类.

输入概念含混阈值为0.5004,调用A-GCT算法,最终生成如下两个概念,如图6所示.输入概念含混阈值为0.5004,调用A-MCT算法,最终生成如下4个概念,如图7所示.表示结果如表3所示.

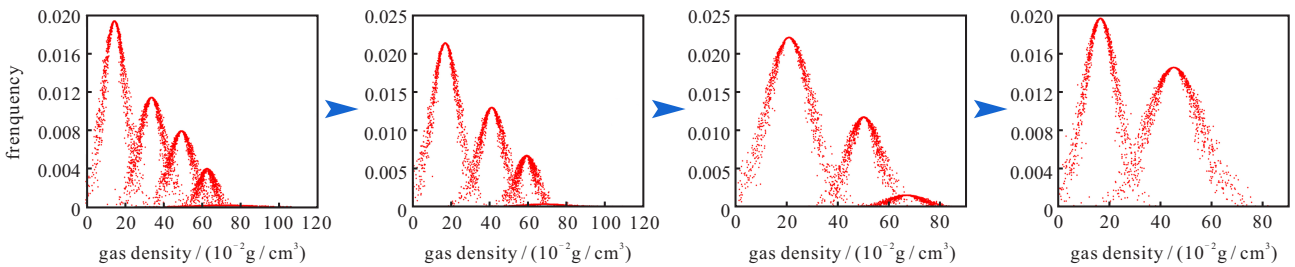


图 6 高斯云变换瓦斯浓度概念提取

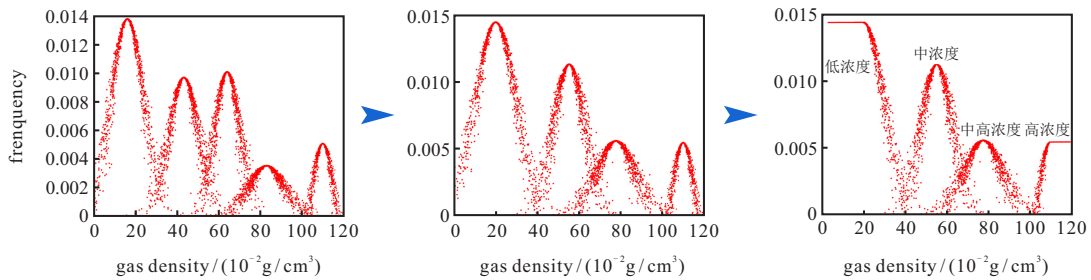


图 7 混合云变换瓦斯浓度概念提取

表 3 不同算法的概念表示结果

算法	概念	期望/(g/cm ³)	熵/(g/cm ³)	超熵/(g/cm ³)	含混度
A-MCT	低浓度	0~0.03	4.2	0.47	0.5004
	中浓度	0.52	3.3	0.55	0.5004
	中高浓度	0.78	3.9	0.75	0.5004
	高浓度	1~∞	2.7	0.45	0.5004
A-GCT	低浓度	0.35	5.6	0.69	0.5004
	中浓度	0.60	4.8	0.82	0.5004

4)对比评价指标选取.以提取到的概念在当前客观论域的覆盖度CR和针对煤矿安全领域其概念

隶属度的客观性SR作为评价指标,覆盖度CR主要考量当前概念表示是否符合其整体的论域分布状态,其

计算公式如下:

$$CR = \frac{\max(X') - \min(X')}{\max(X) - \min(X)}, X \in D, X' \in D'.$$

其中: D 表示原数据集客观论域, D' 表示概念提取后的数据论域.

概念隶属度的客观性(SR)主要考察概念隶属范围是否合理,因为人们在对边界概念的认知上其实是有一定共性的.原算法中,未对边界概念的隶属度给出一个合理的说明(比如其最终生成的两个概念,年轻院士和年老院士,50岁要比40岁对年轻概念的隶属度更高,80岁要比90岁对年老的隶属更高,从常规的认知角度看是不够合理的).SR的计算公式如下:

$$SR = \frac{N(C')}{N(C)}.$$

其中: $N(C')$ 为论域中隶属度合理的概念个数, $N(C)$ 为最终提取概念总个数

5) 实验评估. 对比实验评估如表4所示.

表4 实验结果评估

算法	CR	SR
A-GCT	0.67	0
A-MCT	1	1

由CR和SR两个指标可以看出,本文方法针对煤矿数据概念表示中覆盖论域更全且概念表示结果更为合理.

3.2.2 微观概念表示

煤矿安全数据的微观概念表示主要基于煤矿安全时间粒化结构,通过调用相应的云发生器将离散的时间数据转化为赋有语义表示的时间粒,并依据其时间粒的表示结果构建云规则发生器,应用到短期瓦斯浓度监测数据的预测中,其具体流程如下.

1) 时间粒概念提取. 以天粒层为代表,其云参数变化如图8所示.基于T-CE算法将时间片内的数据转换为具有语言表示的时间粒云概念.

基于时间粒化后,各个时间片内的数据均被云模

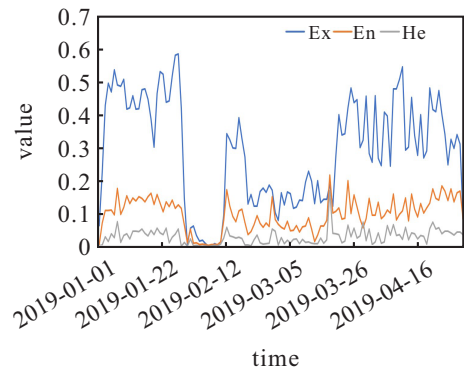


图8 天粒层时间粒内的云参数变化

型的3个参数所表示,期望Ex反映当前时间粒内的浓度量化代表值,熵En反映当前粒度内浓度的波动情况,超熵则为其粒度内浓度波动不确定性的度量值.

2) 风险粒判别. 与宏观概念匹配后,隶属于低浓度的定义其风险粒度为1,表示低风险;隶属于中浓度的定义其风险粒度为2,表示中等风险;隶属于中高浓度的定义其风险粒度为3,表示中高风险;隶属于高浓度的定义其风险粒度为4,表示高风险.隶属度反映当前概念属于该风险的不确定程度.选取当前煤矿数据集一月到五月时间段,基于时间粒概念提取结果,其天粒层的云参数及概念表示见表5.

表5 天粒层概念表示结果

天粒层	Ex	En	He	风险粒度	隶属度
2019年1月1日	0.0647	0.0066	0.0026	1	0.333596
2019年1月2日	0.2257	0.0717	0.0053	2	0.487882
2019年1月3日	0.4284	0.1102	0.0309	2	0.710186
⋮	⋮	⋮	⋮	⋮	⋮
2019年5月1日	0.086	0.0857	0.0368	1	0.350332

如图9所示,各个时间段内的监测数据被转换为相应的风险语义信息.风险粒体积的大小和颜色深度对应相应的风险程度,其颜色深度越深、体积越大,风险度越高.煤矿监测数据时间序列的风险态势得到较为直观的展现.

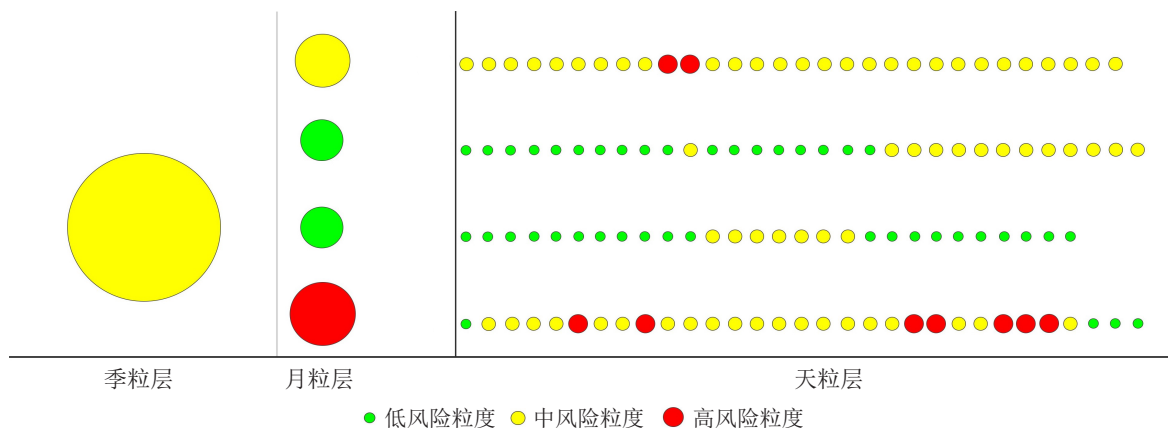


图9 时间粒层对应的风险粒

3) 构建基于时间粒的云规则发生器. 随机选取小时粒层中一天内 6 小时的概念表示云参数, 并基于小时粒度, 将当前分钟粒层的浓度值 x 作为前件云的输入, 调用前件云发生器 (FCRG), 返回确定度 y , 并将确定度 y 作为后件云发生器 (BCRG) 的输入, 最终返回一个预测浓度 x' . 表 6 为基于时间粒所构造的前件云和后件云.

表 6 前后件云参数

前/后件云	time-hour	期望/(g/cm ³)	熵/(g/cm ³)	超熵/(g/cm ³)
前件	3:30	0.5427	0.1148	0.023
	4:30	0.5047	0.1595	0.0669
	5:30	0.5959	0.1158	0.0255
	6:30	0.6291	0.1259	0.0324
	7:30	0.7153	0.1077	0.0275
	8:30	0.5963	0.1043	0.0088
后件	3:31	0.5737	0.0762	0.0099
	4:31	0.5258	0.1145	0.0186
	5:31	0.6478	0.1045	0.0273
	6:31	0.6854	0.1058	0.0352
	7:31	0.5234	0.1006	0.0135
	8:31	0.5234	0.1006	0.0135

4) 实验对比. 微观概念提取的评估主要基于所提取的时间粒构建云规则推理, 进而预测其瓦斯浓度, 并与当下经典的时序预测模型 ARIMA^[22] 对比, 验证其表示方法的有效性.

图 10 为基于云规则预测和 ARIMA 预测的对比图.

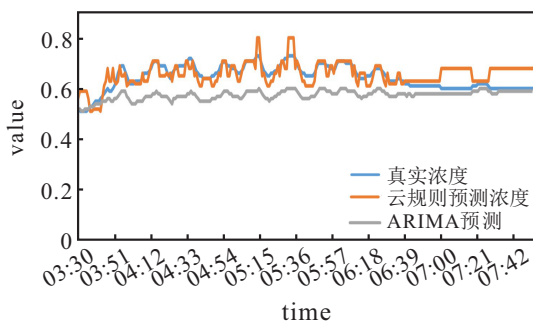


图 10 不同算法浓度预测对比

5) 实验评估与分析. 分别选取均方误差 (MSE)、平均绝对误差 (MAE) 和平均绝对百分比误差 (MAPE) 作为预测评估指标, 对比结果如表 7 所示.

表 7 预测精度对比

算法	MSE	MAE	MAPE
ARIMA	0.0813	0.0699	10.4537
时间粒云规则	0.0018	0.0327	5.0257

由 MSE、MAE、MAPE 三个指标可以看出, 本文方法相较于经典的 ARIMA 预测算法在煤矿数据集上精度更高, 一定程度上验证了多粒度表示方法的优越性.

4 结 论

本文在对煤矿大数据多粒度表示方法的研究中, 首先分析了已有面向煤矿安全大数据的研究现状; 其次, 给出了相关基础算法理论的定义; 再次, 结合煤矿实际监管结构, 提出了基于云变换理论的面向煤矿安全大数据的宏观粒度表示方法和基于逆向云发生器的多时空的微观粒度的表示方法, 并将两者结合, 构建面向煤矿安全大数据的多粒度表示方法; 最后, 将此方法应用到概念提取和煤矿监测数据预测中, 验证了该方法的可行性和有效性. 目前的多粒度表示模型还未考虑煤矿监测数据中不同风险类别这一要素, 如何结合不同煤矿风险类别进行多粒度表示及管控都是后续值得深入研究的问题.

参考文献 (References)

- [1] 崔亚仲, 白明亮, 李波. 智能矿山大数据关键技术与发展研究[J]. 煤炭科学技术, 2019, 47(3): 66-74. (Cui Y Z, Bai M L, Li B. Key technology and development research on big data of intelligent mine[J]. Coal Science and Technology, 2019, 47(3): 66-74.)
- [2] Hua G, Fu L Y, Ren K, et al. The research on historical data similarity of coal mine safety supervision[C]. 2017 International Conference on Information Acquisition. Seogwipo: IEEE, 2007: 143-147.
- [3] Yilmaz M, Potas N, Buyum B. A classical approach to modeling of coal mine data[J]. Chaos, Complexity and Leadership, DOI: 10.1007/978-3-319-18693-1_7.
- [4] Zhou T P. Application of data mining in coal mine safety decision system based on rough set[C]. Intelligent Computing Technology. Berlin Heidelberg: Springer, 2012: 34-41.
- [5] 邵良杉. 基于粗糙集理论的煤矿瓦斯预测技术[J]. 煤炭学报, 2009, 34(3): 371-375. (Shao L S. Disaster Prediction of coalmine gas based on rough theory[J]. Journal of China Coal Society, 2009, 34(3): 371-375.)
- [6] 王军号, 孟祥瑞. 基于物联网感知的煤矿安全监测数据级融合研究[J]. 煤炭学报, 2012, 37(8): 1401-1407. (Wang J H, Meng X R. Reasearch on the data levels fusion of mine safe monitoring based on the perception of internet of things[J]. Journal of China Coal Society, 2012, 37(8): 1401-1407.)
- [7] Lu G, Sun Y B, Hai K Q, et al. Resolution to DM technology for coal mine safety data[C]. 2009

- International Conference on Information Management, Innovation Management and Industrial Engineering. Xi'an: IEEE, 2009: 30-33.
- [8] 邢玉忠, 范红伟, 王彦凯, 等. 基于瓦斯监测数据的矿井通风系统合理性评价[J]. 采矿与安全工程学报, 2010, 27(4): 522-526.
(Xing Y Z, Fan H W, Wang Y K, et al. Ationality evaluation of coal mine ventilation system based on gas monitoring data[J]. Journal of Mining and Safety Engineering, 2010, 27(4): 522-526.)
- [9] Shao C A, Wu Q, Guan X. The research on safety monitoring system of coal mine based on spatial data mining[C]. 2009 International Workshop on Knowledge Discovery and Data Mining (WKDD). Moscow: IEEE, 2009: 126-129.
- [10] Sun G, Wang Z X, Zhao J, et al. A coal mine safety evaluation method based on concept drifting data stream classification[C]. 2016 International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Changsha: IEEE, 2016: 1125-1129.
- [11] 李德毅, 孟海军, 史雪梅. 隶属云和隶属云发生器[J]. 计算机研究与发展, 1995, 32(6): 15-20.
(Li D Y, Meng H J, Shi X M. Membership clouds and membership cloud generators[J]. Journal of Computer Research and Development, 1995, 32(6): 15-20.)
- [12] 王国胤. 云模型与粒计算[M]. 北京: 科学出版社, 2012: 1-4.
(Wang G Y. Cloud model and granular computing[M]. Beijing: Science Press, 2012: 1-4.)
- [13] 杨洁, 王国胤, 刘群, 等. 正态云模型研究回顾与展望[J]. 计算机学报, 2018, 41(3): 724-744.
(Yang J, Wang G Y, Liu Q, et al. Retrospect and prospect of research of normal cloud model[J]. Chinese Journal of Computers, 2018, 41(3): 724-744.)
- [14] Wang J Q, Peng L, Zhang H Y, et al. Method of multi-criteria group decision-making based on cloud aggregation operators with linguistic information[J]. Information Sciences, 2014, 274: 177-191.
- [15] 陈晖, 李德毅, 沈程智, 等. 云模型在倒立摆控制中的应用[J]. 计算机研究与发展, 1999, 36(10): 1180-1187.
(Chen H, Li D Y, Shen C Z, et al. A clouds model applied to controlling inverted pendulum[J]. Journal of Computer Research and Development, 1999, 36(10): 1180-1187.)
- [16] Wang G Y, Xu C L, Li D Y. Generic normal cloud model[J]. Information Sciences, 2014, 280: 1-15.
- [17] Shang X B, Ma P, Chao T. Performance evaluation of electromagnetic railgun exterior ballistics based on cloud model[J]. IEEE Transactions on Plasma Science, 2017, 45(7): 1614-1621.
- [18] Jin D. Construction of “one belt and one road” intelligent analysis system based on cloud model data mining algorithm[C]. 2018 International Conference on Intelligent Computation Technology and Automation (ICICTA). Changsha: IEEE, 2018: 282-285.
- [19] 李德毅, 杜鹞. 不确定性人工智能[M]. 第2版. 北京: 国防工业出版社, 2014: 2-20.
(Li D Y, Du Y. Artificial intelligence[M]. The 2nd edition. Beijing: National Defense Industry Press, 2014: 2-20.)
- [20] 刘玉超. 一种自适应的多粒度概念提取方法——高斯云变换[J]. 计算机工程与应用, 2015, 51(9): 1-8.
(Liu Y C. Adaptive concept abstraction method on multi-granularity—Gaussian cloud transformation[J]. Computer Engineering and Applications, 2015, 51(9): 1-8.)
- [21] Lu W, Pedrycz W, Liu X D, et al. The modeling of time series based on fuzzy information granules[J]. Expert Systems with Applications, 2014, 41(8): 3799-3808.
- [22] 汤筠筠, 郭忠印. 基于自回归求和移动平均的冬季路温短临预测[J]. 同济大学学报: 自然科学版, 2017, 45(12): 1824-1829.
(Tang J J, Guo Z Y. Pavement temperature short-impending prediction based on ARIMA in winter[J]. Journal of Tongji University: Natural Science, 2017, 45(12): 1824-1829.)

作者简介

代劲(1978—), 男, 教授, 博士生导师, 从事大数据知识工程、智能信息处理等研究, E-mail: daijin@cqupt.edu.cn;

张磊(1995—), 男, 硕士生, 从事智能信息处理、数据挖掘的研究, E-mail: s180231955@stu.cqupt.edu.cn;

王国胤(1970—), 男, 教授, 博士生导师, 从事粒计算、认知计算、智能信息处理等研究, E-mail: wanggy@cqupt.edu.cn.

(责任编辑: 齐 粟)