

[机组]

基于粗糙集和聚类PSO-RBF的水电机组故障诊断

文/湖北工业大学电气与电子工程学院 范秀香
深圳康必达控制技术有限公司 林峰平

将聚类PSO优化粗糙集-神经网络算法引入水电机组故障诊断中，利用聚类算法和PSO算法的优点改进粗糙集属性约简方法并对水电机组故障的检测信息进行约简，提取对故障分类起主要作用的信息，并用BP神经网络对粗糙集处理后的故障信息进行诊断。该方法降低了神经网络的输入信息空间维数，简化神经网络结构，有效提高故障诊断的准确性。通过对水电机组振动故障数据进行诊断分析，证明了该诊断方法的有效性和优越性。

水电机组的可靠性直接影响电站发电的安全与稳定。面对庞杂的监测数据，需要控制系统实时应对并且能够快速准确地判定故障。然而，机组的各种异常与故障受到机、电、水力等因素影响，表现出非线性、复杂性、耦合性等特点，使得水电机组故障诊断的准确性与可靠性大受影响，成为近年来的研究热点。针对该难题，学者们提出了基于支持向量机、人工神经网络及Petri网络等的智能诊断算法，但是在处理具有不确定、不一致与不完备性故障信息时，上述方法仍存在局限性。

求解最小属性约简和属性值约简问题是NP问题，粗糙集理论由于其无需先验知识或者其他条件，在属性约简和规则提取方面有较强优势，即可高效处理各种不完备数据，而且在复杂数据信息的分析决策、模式分类、机器学习与知识发现等方面取得成功，目前已经有效应用于各类电气故障诊断。

粒子群优化算法（PSO）是一种基于种群的随机优化技术，由Eberhart和Kennedy于1995年提出，模仿昆虫、鸟和鱼等群集以合作模式搜寻食物，种群个体通过学习其自身经验和其他成员的经验来不断优化搜索模式。

本文综合传统属性约简和属性值约简的特点，提出一种粗糙集-聚类粒子群算法。通过粗糙集建立规则提取数据决策表，采用聚类算法将数据组依据决策属性进行属性分类，利用粒子群算法的群聚特点有效融合区间，优化初始断点集，约简属性值并实现属性约简，最后应用粗糙集理论对离散约简后的决策表提取规则。实例验证了该算法的有效性。

粗糙集理论离散化

决策表 $T = (U, R, V, f)$ ，其中 U 是论域， $R = C \cup D$ 是属性集合， C 、 D 分别是条件属性和决策属性子集。 $V = \bigcup V_a$ ， $a \in A$ 是属性值， A 是属性集， V_a 是属性值域； $f: U \times R \rightarrow V$ 是一个信息函数，它指定 U 中每一个对象 x 的属性。

离散化是通过选取合适的断点划分条件属性空间，把 n 维（ n 为条件属性个数）空间划分成 m 档，每档对象的决策值相同，采用以下标准衡量断点集 P 合理性：

- （1）一致性，离散后数据集一致性与离散前数据一致性相同。
- （2）不可约简性，即不存在 $P' \subset P$ ，满足一致性。

(3) 最小离散性, 对于任意满足一致性的断点集 P' , 均有 $\text{card}(P) < \text{card}(P')$, card 是集合“基数”, 则 P 是最优断点集。

聚类PSO算法

1. PSO算法

粒子群算法属于智能优化算法的范畴, 主要设计思想与进化算法和人工智能两个方面的研究密切相关, 与大多数演化计算技术具有相同的特点:

- (1) 随机确定初始种群。
- (2) 对种群内每一样本求取适应值 (fitness value), 适应值评价与最优解的距离。
- (3) 基于适应值情况更新种群。
- (4) 若终止条件达成则停止, 反之转步骤2。

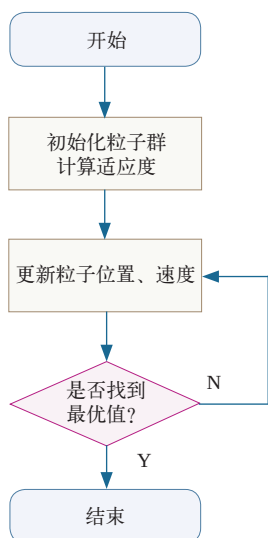


图1 标准PSO计算流程

可见, PSO的核心之处在于随机选取起始种群并通过适应值计算评价可行解, 但不保证一定找到最优解。可以将PSO算法分解为如下步骤:

(1) 设立初始种群 $S(t)$, t 为群体规模, 每个粒子是 q 维的。

(2) 由群体 S 对 q 维空间搜索。每个粒子 $x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{iq})$, 每个粒子速度为 $v_i = (v_{i1}, v_{i2}, v_{i3}, \dots, v_{iq})$, 每个粒子搜索时遵循两个规律:

a. 自己搜索到的历史最优值 p_i , $p_i = (p_{i1}, p_{i2}, \dots, p_{iq})$, $i = 1, 2, 3, \dots, t$;

b. 全部粒子搜索到的最优值 p_j , $p_j = (p_{j1}, p_{j2}, \dots, p_{jq})$, p_j 只有一个。

(3) 粒子位置更新速度

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 \zeta (p_{id}^k - x_{id}^k) + c_2 \eta (p_{jd}^k - x_{id}^k) \quad (1)$$

$$x_{id}^{k+1} = x_{id}^k + r v_{id}^{k+1} \quad (2)$$

式(1)中 ω 是惯性权重, 反映原有速度; c_1 是粒子跟踪自己历史最优值的权重系数, 反映粒子对自身的认知, 一般置为2; c_2 反映粒子跟踪群体最优值的权重, 也叫“社会知识”, 即粒子对群体知识的认识, 一般设为2; ζ, η 是 $[0, 1]$ 区间内均匀分布的随机数; r 是位置更新时对速度的约束因子, 一般为1。图1是基本粒子群算法的过程表示。

判断终止条件可是设置适应值到达一定的数值或者循环一定的次数, 粒子同时跟踪自己的历史最优值与全局(群体)最优值来改变自己的位置预速度。

2. 聚类算法

聚类方法按照某种相似度原则将大量 q 维样本(d 个个体, 每个个体有 e 个属性)分成 K 类($K < d$), 保证同类样本间相似度最大, 非同类样本间相似度最小。相似度评价方法较多, 本文以每个样本到其聚类中心的距离和最小作为目标函数

$$A(s, c) = \min \sum_{j=1}^k \sum_{i=1}^{n_j} \sum_{p=1}^m \|x_{ip} - c_{jp}\|^2 \quad (3)$$

$$c_{id} = \frac{\sum_{j=1}^k s_{ij} x_{ip}}{\sum_{i=1}^{n_j} s_{ij}} \quad (j=1, \dots, k; p=1, \dots, m) \quad (4)$$

$$s_{id} = \begin{cases} 1, & \text{样本} i \text{属于} j \text{类} \\ 0, & \text{样本} i \text{不属于} j \text{类} \end{cases} \quad (j=1, \dots, k; i=1, \dots, n_j) \quad (5)$$

式中, x_{ip} 为第 i 个样品的第 p 个属性, c_{jp} 为第 j 个类中心的第 p 个属性。

3. 聚类PSO算法

粒子群算法本质是利用粒子搜索的极值和全局极值两个信息不断更新查找, 算法过程虽然可以找到较优值, 但由于前期缺乏先验知识及对数据信息的引导, 算法后期不可避免地陷入收敛速度变慢和精度变低的问题, 从而造成决策表及决策规则复杂化。为克服这些问题, 本文综合聚类和PSO算法, 提出一种混合优化算法。前期利用聚类算法对决策表数据信息预处理, 根据决策属性分类划分断点区间, 减少断点数; 后期应用PSO算法对预处理后的决策表离散化, 弥补PSO不足, 得到最简决策规则。

本文采用二进制编码方式对粒子编码, 每个粒子代表决策表连续属性的一个初始断点。用长度为 P 的二进制串来表示样本种群, 该串由 N 个子串组成, 每个子串对应决策表中一个条件属性的

初始断点集合，每一位对应一个断点。其值“1”和“0”分别代表该断点不可融合和可融合。

最简决策规则提取取决于两点：所包含的属性和属性值个数尽量少；条件属性对决策属性的依赖度尽量大。因此定义由断点数和依赖度决定的目标函数

$$F(x)=\left(1-\frac{I_x}{n}\right)+\alpha\times r_{ci}(D) \tag{6}$$

式（6）中， N 表示属性 x 的断点总数， I_x 表示个体 x 约简后包含断点数； α 为权重调节因子， $r_{ci}(D)$ 为粒子所包含的断点划分得到的离散化决策表中决策属性对条件属性集的依赖度

$$r_c(D)=\frac{\text{card}(\text{pos}_c(D))}{\text{card}(U)} \tag{7}$$

式（7）中， $r_c(D)$ 为条件属性 C 对决策属性 D 的依赖度； $\text{card}(U)$ 表示集合的势； $\text{pos}_c(D)$ 为等价类 U/D 关于的 C 正域。基于上述聚类PSO模型描述，聚类PSO算法决策规则提取步骤：

Step1初始化PSO算法中的离散、属性约简粒子的各参数值。

Step2输入决策表 $T=(U, R, V, f)$ ，按照决策属性进行聚类排序。

Step3计算条件属性 C_i 对决策属性 D 的支持度 $r_{ci}(D)$ ，重复迭代多次使 $r_{ci}(D)$ 的值连续多次保持最大值不变。

Step4若 $i < N$ ，令 $i=i+1$ ，转Step3，否则转Step5。

Step5根据式（3）计算目标函数值，若目标函数值连续多次保持最大值不变则转Step6，反之，令，转Step3继续寻优。

Step6记录各粒子所表示的断点集，根据当前状态确定离散区间，对决策表进行离散化。

Step7记录当前各粒子所表示的约简属性集，对离散化后的决策表进行属性约简。

故障诊断决策表的建立

水电机组故障成因复杂，振动是最常见的故障诱发因素，如：质量不平衡、轴系不对中、定子组合缝松动、碰摩、尾水管内偏心涡带引起振动、叶片出口卡门涡列等。总结了水电机组的常见振动频率特征，本文在此基础上利用文献中的数据作为算法测试的决策表，对数据进行分析，得到各个频段的特征向量作为粗糙集的条件属性，对应的故障类型作为决策属性，形成故障决策表，并运用聚类PSO优化粗糙集的方法进行诊断。该决策表含有8组数据，选取振动信号频谱中的6个频率特征量： $0.18-0.2f$ 、 $1/6-1/2f$ 、 $1f$ 、 $2f$ 、 $\geq 3f$ 、50 Hz或100 Hz（ f 为机组转频）以及振动与转速关系、振动与负荷关系、振动与流量关系作为条件属性，分别用 $c_1\sim c_9$ 表示。选取5类故障类型作为决策属性。构成相应于9种条件属性和5类故障的决策表，建立水电机组振动故障决策表如表1所示。

表1 故障决策表

U	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	D
1	0.01	0.03	0.61	0.97	0.85	0.05	0.99	0.91	0.38	转子不对中
2	0.02	0.07	0.71	0.98	0.79	0.02	0.91	0.97	0.54	转子不对中
3	0.04	0.01	0.82	0.95	0.77	0.01	0.99	0.95	0.07	动静碰摩
4	0.07	0.02	0.44	0.97	0.81	0.04	0.98	0.92	0.11	动静碰摩
5	0.06	0.01	0.96	0.46	0.51	0.01	0.91	0.11	0.12	动静碰摩
6	0.01	0.1	0.97	0.03	0.22	0.02	0.92	0.25	0.03	转子质量不平衡
7	0.05	0.96	0.05	0.01	0.01	0.9	0.11	0.91	0.02	尾水管偏心涡带
8	0.04	0.01	0.97	0.01	0.01	0.02	0.95	0.05	0.97	磁极不均匀

采用本文提出的离散化方法对故障决策表进行离散化处理，得到离散化后的决策系统，如表2所示。

表2 离散化后的决策表

U	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	D
1	1	1	2	3	3	1	3	3	2	转子不对中
2	1	1	2	3	3	1	3	3	2	转子不对中
3	1	1	3	3	3	1	3	3	1	动静碰摩
4	1	1	2	3	3	1	3	3	1	动静碰摩
5	1	1	3	2	2	1	3	1	1	动静碰摩
6	1	1	3	1	2	1	3	2	1	转子质量不平衡
7	1	3	1	1	1	3	1	3	1	尾水管偏心涡带
8	1	1	3	1	1	1	3	1	3	磁极不均匀

采用基于属性重要性的方法进行决策表的属性约简，得到以下信息：属性{c₉}为核属性，总的分类质量为1，约简属性为属性c₅和属性c₉，提取的决策规则总数为5条，如表3所示。

表3 PSO约简得到的规则表

U	c ₅	c ₉	D
1	3	2	转子不对中
2	3	1	动静碰摩
3	2	1	转子质量不平衡
4	1	1	尾水管偏心涡带
5	1	3	磁极不均匀

用聚类PSO算法离散化所得到的决策规则较少，更加简洁。

1. 水电机组故障诊断分析

分别构造BP神经网络、RS-BP神经网络和聚类PSO优化RS-BP神经网络进行训练。给定一组待诊断故障数据样本作为测试样本进行测试，数据如表4所示。

表4 故障诊断测试样本

c ₁	c ₂	c ₃	c ₄	c ₅	c ₆	c ₇	c ₈	c ₉	实际故障
0.01	0.06	0.88	0.11	0.15	0.06	0.93	0.02	0.01	转子质量不平衡

将表4作为待诊断测试样本，分别输入已经训练好的3个神经网络进行诊断，诊断比较结果如表5所示。

表5 各算法诊断结果

故障诊断 分类结果	诊断方法		
	BP神经网络	RS-BP神经网络	本文方法
1	0.0523	0.0612	0.0153
2	0.1647	0.1736	0.0217
3	0.7853	0.9163	0.9887
4	0.1312	0.0357	0.0112
5	0.1137	0.0541	0.0237

神经网络阈值取0.5，小于0.5该故障不存在，大于0.5该故障存在。3种诊断方法得到的诊断结果都是转子质量不平衡，诊断结果与水电机组实际故障相符，前两者的可信度分别为0.7853和0.9163，基于本文的方法得到的诊断结果的可信度为0.9887，由此可以看出，采用本文的方法进行诊断时，诊断结果的准确性更高。

结束语

为解决水电机组故障中难以精确诊断的问题，本文将聚类PSO优化的粗糙集-神经网络算法应用于水电机组的故障诊断中，通过实例分析，表明了该方法的有效性。EA

(收稿日期：2015.02.27)

(编辑：白文亭)

东润环能携手阿里云共建“能源云”探索能源互联网

2016年4月20日，北京东润环能科技股份有限公司与阿里云计算有限公司签署合作协议，推动能源服务与云计算大数据技术的融合，共同探索基于阿里云技术的“能源云”服务。

根据协议，双方将在新能源、能源互联网领域中云计算、大数据、互联网技术和生态能力方面达成合作，通过共享各自领域的优质资源，在新能源、能源互联网业务领域形成合力，开展应用创新、数据价值创新、能源互联网服务创新。

据悉，东润环能将发挥行业优势和业务经验，基于阿里云进行应用研发，对新能源、能源互联网产业各方用户提供全面的专业应用和数据服务；阿里云将为东润环能提供云计算、大数据、互联网技术支持，赋予东润环能开发、创新应用环境。双方通过不断的业务合作与技术交互，本着发挥双方技术优势、促进能源互联网发展、推动新能源大数据云应用，共同打造基于阿里云技术的“能源云”服务。

作为一家创新能源互联网服务公司，东润环能在新能源产业规划、电站并网、发电预测、新能源气象、分布式微网、智能用电等技术和数据服务方面，已经发展成为行业领跑者。此次，阿里云和东润环能的联手，无疑将为产业客户提供实用、便捷、独特、互动的新能源互联网技术服务，为能源互联网+带来创新动力。

据透露，该平台将结合新能源门、能量魔方应用服务，以地理数据、资源数据、气象服务、设备参数、能源规划、消纳限电评价等数据服务为基础，构建新能源产业互联网云应用服务平台，为新能源投资企业、地方新能源开发、电网、装备技术贸易、产业金融服务提供深度数据挖掘和专业应用服务。