

基于特征变换的 DGA 诊断范例推理方法

高明磊¹ 张钟江² 姬 波¹

(郑州大学信息工程学院 郑州 450001)¹ (中国电子科技集团公司第二十八研究所 南京 210007)²

摘 要 Pearson 相关系数是一种衡量变量间线性关系的方法,广泛用于变压器中油中气体故障诊断(DGA)的范例推理匹配算法。但是,现有方法存在偏袒数据区间较大的特征以及认为所有特征对相关系数判定的贡献相同这两个问题。因此,在深入分析 DGA 色谱数据的基础上,提出采用对数特征变换方法缩小特征值域来解决偏袒大数据区间特征的问题,采用均方差特征赋权区分特征贡献度的方法进一步提高 DGA 故障检测效果,并构造了基于特征变换和特征权重的 Pearson 相关系数 DGA 诊断(FTW_Pearson)算法。实验结果表明,FTW_Pearson 算法的 DGA 诊断正确率优于业界普遍使用的大卫三角形法、未考虑特征变换和权重的 Pearson 相关系数法以及贝叶斯算法和神经网络算法。

关键词 皮尔森相关系数,油中溶解气体,特征变换,特征权重,范例推理

中图法分类号 TP391.5 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.10.051

DGA Fault Diagnosis Based on CBR Method with Feature Transformation

GAO Ming-lei¹ ZHANG Zhong-jiang² JI Bo¹

(School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China)¹

(The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China)²

Abstract Pearson correlation coefficient is a way to measure the linear relationship between two variables, which is widely used as CBR matching algorithm for DGA fault diagnosis. However, the traditional application has two problems; discriminating in favor of the features which have larger data range and regarding equally the contributions of all features. To address these issues, the paper proposed the log-function feature transforming method to narrow the data range to solve the discrimination problem and proposed the mean square deviation feature weighting method to distinguish the contribution levels to improve the accuracy of DGA fault diagnosis. Experimental results show that the proposed FTW_Pearson algorithm is superior to David triangle method which is popularly used in real applications, the traditional Pearson algorithm without feature transforming/feature weighting, the Bayes algorithm and the BPNN algorithm.

Keywords Pearson correlation coefficient, DGA, Feature transforming, Feature weighting, CBR

电力变压器作为电力系统的关键设备之一,其工作状态影响整个电网的可靠性和稳定性^[1]。油中溶解气体分析(DGA)是分析电力变压器故障的有效手段。基于 DGA 的电力变压器故障智能化诊断方法包括粗糙集^[2]、3 层贝叶斯网^[3]、支持向量机^[4]、神经网络^[5]、范例推理^[6]等。其中,范例推理(CBR)^[7]模仿了人类的认知心理活动过程。在面临新的问题求解时,从过去处理相似问题的经验和知识出发,基于新的情况与已知范例的差异度,获取新问题的解。CBR 思想可以描述为:利用相似度衡量算法检索出与新问题差异度最小的老问题,然后利用已知的老问题的求解策略来解答新问题。与其他智能化诊断方法相比,CBR 拥有可以避免复杂的模型建立过程和快速求解两个明显优势,因而在多属性决策^[8]、空间地理学、交通路况分析、微博用户特征分析等领域得到了广泛的应用。Pearson 相关系数^[9,10]作为一种衡量变量间线性

关系的方法,具有简单高效的优点,在 DGA 分析中被广泛用于范例推理的匹配算法。但是,现有的 Pearson 相关系数 DGA 分析应用中存在两个问题:1)相关系数计算时偏袒数据区间较大的特征;2)认为所有特征对相关系数判定的贡献相同。因此,本文在深入分析 DGA 色谱数据的基础上,提出采用对数特征变换缩小特征值域来解决偏袒大数据区间特征的问题,采用均方差特征赋权^[11,12]对不同特征赋予不同重要度来区分特征贡献度的方法,以进一步提高 DGA 故障检测效果。在此基础上,构造了基于对数变换和均方差赋权的 Pearson 相关系数 DGA 故障诊断(FTW_Pearson)算法。实验表明,FTW_Pearson 算法的 DGA 诊断正确率优于业界普遍使用的大卫三角形法、未考虑特征变换和权重的 Pearson 相关系数法,以及贝叶斯算法和神经网络算法。

到稿日期:2014-05-22 返修日期:2014-07-13 本文受国家自然科学基金项目(61170223),国家自然科学基金河南人才培养联合基金(U1204610),河南省科技攻关计划项目(132102210404)资助。

高明磊(1963—),女,实验师,主要研究方向为数据分析、工控系统;张钟江(1972—),男,高级工程师,主要研究方向为系统工程、信息处理;姬 波(1972—),男,博士,副教授,CCF 会员,主要研究方向为模式识别、工控系统,E-mail:iebj@zzu.edu.cn(通信作者)。

1 DGA 范例推理方法

1.1 油中溶解气体分析方法

电力变压器在长期负荷运行中,受到热、电和机械力等的持续作用,会导致变压器中的绝缘材料(油或纸板等)逐渐老化并伴随产生 H_2 , CO , CO_2 和其他一些低分子烃类气体。当变压器内部潜伏性故障存在时,上述气体会加速产生。因此,对于油中气体的含量和趋势的分析(DGA)可以对变压器故障进行有效预警,从而提前制定检修或维护策略。DGA 诊断技术可以分为两类,即传统诊断方法和智能诊断方法。传统诊断方法主要包括特征气体法:依照故障产生气体的总烃、序和含量等进行诊断;三比值法:从 5 种主要的特征气体 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 和 C_2H_2 中选取溶解度与扩散系数均相近的两种特征气体组成 3 对比值, C_2H_2/C_2H_4 、 CH_4/H_2 及 C_2H_4/C_2H_6 , 然后查表得到编码并确定故障类型;大卫三角形法:计算 CH_4 、 C_2H_4 、 C_2H_2 3 种特征气体含量占总含量的百分比,然后利用大卫三角形图确定故障类型。智能诊断方法主要包括粗糙集:对数据进行属性约简,从而解决冗余信息下的故障诊断问题;贝叶斯网:建立基于概率推理的图形化网络并运用该网推理得出变压器故障模式;支持向量机:构造输入输出样本对,归一化应用 SVM 算法进行回归参数和偏差的计算,并建立故障诊断模型;神经网络:对输入故障数据与故障类型建立联系,先用故障样本对训练出神经网络,然后用该网络进行故障识别;范例推理:针对目标样本,从范例库中检索相似范例,并据此进行故障诊断。

1.2 范例推理和 Pearson 相关系数法

范例推理起源于人类对事物的认知过程。人们在求解新问题时,往往基于处理类似问题的经验和知识并针对新旧问题的差异性做出相应调整,从而得到新问题的解;同时,基于新问题及其解对范例库进行更新。范例推理(CBR)有以下 4 个组成成分:范例库(case base),经验知识的存储;检索机(indexing mechanism),管理范例的排列;匹配算法(matching algorithm),新旧范例相似度比对;调整机制(adaptation or inference mechanism),推演出最终的解。范例推理的主要步骤包括:1)检索。根据待定问题的信息从已有范例库中检索。2)重用。从检索到的范例中获取问题解。3)修正。从相似范例中修正问题解,使之适合于当前问题。4)保存。按照相关策略保存新范例到范例库。

Pearson 相关系数(见式(1))作为一种衡量变量间线性关系的方法,具有简单高效的优点,在 DGA 分析中被广泛用于范例推理的匹配算法。Pearson 相关系数 $corr$ 用于衡量两个数据变量是否共线,即给出变量间的线性关系,如衡量身高和体重、收入和消费变量间的线性相关关系。 $corr$ 绝对值越接近于 1,相关性越强; $corr$ 绝对值越接近于 0,相关度越弱。变量的相关强度的约定见表 1。 $corr$ 的一个显著数学特性是,两个变量位置和比例尺度的变化不会引起相关系数变化。例如将 X 变换为 $a+bX$,或者将 Y 变换为 $c+dY$, a 、 b 、 c 和 d 均是常数,并不改变相关系数值。

$$corr = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

表 1 变量的相关强度

相关性	相关系数绝对值
极弱相关或无相关	0.0~0.2
弱相关	0.2~0.4
中等程度相关	0.4~0.6
强相关	0.6~0.8
极强相关	0.8~1.0

2 FTW_Pearson 算法

2.1 特征变换

特征变换是指特征变量的全体或部分值的变换,主要分为两种变换类型:函数变换和规范化(标准化)。函数变换 $x' = f(x)$ 指将一个设定好的数学函数应用于特征的值,常用的变换函数有 xk , $\log x$, ex , $1/x$ 等。规范化(标准化)的目标是使得特征的所有值的集合具有特定的性质,如变量标准化 $x' = (x - \bar{x})/S_x$ ($\bar{x}$ 为特征值的均值, S_x 是标准差)方法将使得新变量 x' 具有均值 0 和标准差 1。特征变换的主要目的包括将不具备高斯分布的数据变换成具有高斯分布的数据,避免具有较大值域的特征左右分析结果等。根据 DGA 数据特征,本文选择了对数特征变换方法来缩小特征的数值范围,以避免偏袒大数据区间的特征。对数变换方法见式(2)。

$$x' = \log(x+c) \quad (2)$$

其中, c 是常数 1.01。常数 c 的引入是为了将函数值域范围从 $(-\infty, +\infty)$ 变化为 $(0, +\infty)$ 。

为了反映对数特征变换的效果,作为对比,实验中也给出了 Max-min 法和反正切函数法两种变换方法的结果。Max-min 法见式(3),变换后的值域为区间 $[0, 1]$;反正切函数法见式(4),变换后的值域为 $[-\pi/2, \pi/2]$ 。

$$x' = \frac{x - \text{Minvalue}}{\text{Maxvalue} - \text{Minvalue}} \quad (3)$$

$$x' = \frac{\arctan x \times 2}{\pi} \quad (4)$$

2.2 特征加权

在采用对数特征变换后,得到了更能反映数据本质的变换后特征空间。在此基础上,引入特征加权方法来对 DGA 数据的各个特征赋予不同的重要程度,从而达到提高重要特征权重、降低次要/冗余特征权重的目标。令属性权重为权值向量 w ,则引入特征加权的 Pearson 相关系数可以改写为:

$$corr(x, y; w) = \frac{cov(x, y; w)}{\sqrt{cov(x, x; w) cov(y, y; w)}} \quad (5)$$

其中, $cov(x, y; w)$ 为加权协方差。

本文在比较多种权重度量方案的基础上,提出采用均方差权重度量方案作为 DGA 数据的权重度量方法。均方差用来衡量样本波动的幅度,对于 DGA 数据,波动幅度大的特征比波动幅度小的特征对故障诊断更具指导意义,因而被赋予较高权重。根据实验对比需要,在此给出了另外两种常用的权重度量方案。熵权重:熵是信息的不确定性的度量,不确定性越大,熵值越高;互信息权重:互信息可以度量随机变量对 X 和 Y 的相互包含的信息量,可以用来描述变量的关联程度。令 DGA 属性集为 $\{A_1, A_2, \dots, A_n\}$, P_{A_i} 为属性 A_i 的概率分布, $p(a_i, a_j)$ 为属性 A_i 和 A_j 之间的联合概率分布, m 为属性数目,则属性 A_i 的均方差权值 $RMSD(A_i)$ 、熵权值 $H(A_i)$ 和互信息权重 $MI(A_i)$ 的定义见式(6)。

$$\begin{aligned}
\text{RMSD}(A_i) &= \sqrt{\text{MSE}(\bar{A}_i)} = \sqrt{E(A_i - \bar{A}_i)^2} \\
H(A_i) &= -\sum_{i=1}^n p_{A_i} \log p_{A_i} \\
\text{MI}(A_i) &= \frac{1}{m-1} \sum_{j=1, j \neq i}^m I(A_i; A_j) \\
I(A_i; A_j) &= \sum_{a_i \in A_i} \sum_{a_j \in A_j} p(a_i, a_j) \log \frac{p(a_i, a_j)}{p(a_i)p(a_j)}
\end{aligned} \quad (6)$$

2.3 FTW_Pearson 算法

2.3.1 算法伪码

基于特征变换和特征权重的 Pearson 相关系数法 DGA 故障诊断 (FTW_Pearson) 算法伪码如表 2 所列。

表 2 FTW_Pearson 算法伪码

输入: 已有范例集 Mtrain 待分类样本集 Mtest
输出: 待分类样本类标签 L
步骤 1. 对数特征变换 $x' = \log(x+c)$, 获得特征空间 Mtrain'
步骤 2. For loop = 1; 3
If loop == 1
计算均方差权重 $\text{RMSD}(A_i)$
Elseif loop == 2
计算熵权重 $H(A_i)$
Elseif loop == 3
计算互信息权重 $\text{MI}(A_i)$
End if
特征赋权, 获得特征空间 Mtrain''
计算 Pearson 相关系数, 得到 corr 绝对值最大的范例的类标签
作为待分类样本的类标签
End for

2.3.2 时间复杂度分析

FTW_Pearson 算法步骤 1 中的对数变换时间复杂度为 $O(|X||Y|)$, 步骤 2 中均方差权重时间复杂度为 $O(|X||Y|)$, 熵权重时间复杂度为 $O(|X||Y|)$, 互信息权重时间复杂度为 $O(|Y|^2)$, 特征赋权时间复杂度为 $O(|X||Y|)$, Pearson 相关系数时间复杂度为 $O(|X||Y|)$ 。其中 $|X|$ 是数据对象的个数, $|Y|$ 是属性值的个数。因此, FTW_Pearson 算法时间复杂度为 $O(|X||Y|)$ 。

3 实验

3.1 实验数据集及实验评估方法

实验数据集来源于变压器运行现场采集的 DGA 特征数据, 共 609 条记录。每条记录包括 8 个属性: (H_2 , CH_4 , C_2H_6 , C_2H_4 , C_2H_2 , ZTING, CO , CO_2)。首先对数据进行清理, 以得到适用于数据分析的数据。数据清理方法包括: 矛盾数据删除, 如特征值完全相同的两条记录的故障类型不同是一种明显的矛盾; 重复数据删除; 遗漏值 (某些特征值为空白) 删除等。最终选定了 150 条记录作为实验数据, 其中包括了 5 类典型故障。故障类型及记录数目见表 3。

表 3 实验数据集

序号	故障类型	记录数
1	低能放电	27
2	高能放电	34
3	中温过热	2
4	高温过热	85
5	放电及过热	2
合计		150

在此基础上生成了 10 组十折交叉验证 (10-fold cross-validation) 数据集当作实验数据集来测试 FTW_Pearson 算法

的准确性。十折交叉验证方法是将数据集随机分成 10 份, 轮流将其中 1 份作为测试集, 另外 9 份合并为训练集。最终的实验数据分为 10 组, 每组包括 10 个测试集 (含 15 个记录) 和 10 个训练集 (含 135 个数据)。

实验评估中采用 10 次十折实验的结果的平均正确率作为对 FTW_Pearson 算法的效果估计。平均正确率的计算方式如式 (7) 所示。

$$\text{Accuracy} = \frac{1}{m} \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n \frac{\text{rightnum}}{\text{num}} \quad (7)$$

其中, m 为实验组数; n 为每组实验中的实验次数; rightnum 为测试集中正确分类的记录数目; num 为测试集中的总记录数目。

3.2 实验结果

3.2.1 三类特征变换的实验结果

表 4 中给出了 10 次十折实验的正确率, 参与比较的算法包括大卫三角形方法、未进行特征变换的 Pearson 相关系数法、基于 Max-min 变换的 Pearson 相关系数法、基于反正切函数变换的 Pearson 相关系数法和本文提出的基于对数变换的 Pearson 相关系数法。从中可以看出:

1) 未进行特征变换的 Pearson 相关系数法的平均正确率和总正确率均优于大卫三角形。这说明 Pearson 相关系数作为范例推理方法的匹配算法, 对于 DGA 故障诊断是可行的和有效的。

2) 基于 Max-min 变换的 Pearson 相关系数法的正确率低于大卫三角形; 基于反正切函数变换和基于对数变换的 Pearson 相关系数法的正确率优于大卫三角形。这说明特征变换方法的不同会直接影响 DGA 故障诊断结果, 合理的特征变换方法能提高诊断正确率, 不合理的特征变换方法反而会降低诊断正确率。

3) 基于对数变换的 Pearson 相关系数法的平均正确率和总平均正确率均为最优。其总平均正确率比大卫三角形法高约 36%, 比无特征变换的方法高约 7%, 比 Max-min 变换高约 46%, 比反正切变换提高约 5%。这说明本文提出的基于对数特征变换的方法能对 DGA 数据特征进行最准确的描述, 从而得到最优 DGA 诊断效果。

表 4 10 次十折实验的正确率

	大卫 三角形	Pearson 相关系数			
		无变换	Max-min 变换	反正切变换	对数变换
1	0.4370	0.6257	0.3714	0.6486	0.7115
2	0.4371	0.6170	0.3401	0.6572	0.7056
3	0.4372	0.6400	0.3544	0.6569	0.6942
4	0.4373	0.6314	0.3802	0.6743	0.7000
5	0.4370	0.6315	0.3200	0.6457	0.7028
6	0.4372	0.6343	0.3114	0.6515	0.6942
7	0.4372	0.6256	0.3258	0.6429	0.6944
8	0.4371	0.6257	0.3655	0.6486	0.7000
9	0.4372	0.6629	0.3142	0.6599	0.7027
10	0.4370	0.6286	0.3601	0.6427	0.7058
总平均	0.4371	0.6323	0.3443	0.6528	0.7011

作为更细节的实验分析, 在图 1(a) 和图 1(b) 中给出了十折实验数据集第 1 组和第 2 组上的 Max-min 变换、反正切变换和对数变换的正确率曲线。从中可以看出:

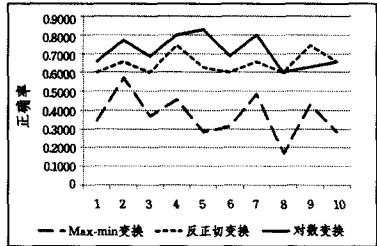
1) 在两组数据集上, Max-min 变换正确率均最低, 这进一步说明 Max-min 变换不适合于 DGA 数据的处理;

2) 在第 1 组的 10 个数据集上, 对数变换在其中 7 个数据

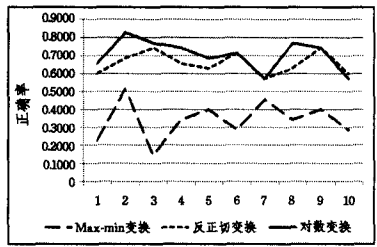
集上优于反正切变换,在 2 个数据集上与反正切变换相同,在 1 个数据集上劣于反正切变换;

3)在第 2 组数据集上,对数变换在其中 6 个数据集上优于反正切变换,在 3 个数据集上与反正切变换相同,在 1 个数据集上劣于反正切变换;

4)对数变换的效果在总体上优于 Max-min 变换和反正切变换,是更适合用于 DGA 数据的特征变换方法。



(a)第 1 组



(b)第 2 组

图 1 Max-min、反正切和对数变换的正确率对比

3.2.2 特征加权后的实验结果

上述实验说明了本文提出的对数特征变换是最佳特征变换方法。接下来,在对数变换基础上采用均方差权重方法对数据进行赋权,突出重要属性并抑制冗余/次要属性。表 5 给出了对数特征变换及特征加权后的实验结果。

表 5 对数特征变换并加权后的实验结果

	未加权	熵权重	互信息权重	均方差法
1	0.7115	0.6828	0.7171	0.7115
2	0.7056	0.6887	0.6971	0.7086
3	0.6942	0.6542	0.7058	0.6972
4	0.7000	0.6971	0.7228	0.7229
5	0.7028	0.6801	0.7055	0.7087
6	0.6942	0.6800	0.6972	0.7142
7	0.6944	0.6888	0.7029	0.6999
8	0.7000	0.6800	0.7000	0.6972
9	0.7027	0.6942	0.7200	0.7142
10	0.7058	0.6886	0.7173	0.7285
10 次十折总平均值	0.7011	0.6835	0.7086	0.7103

从表 5 可以看出,均方差法的 10 次十折总平均值优于未加权的方法、熵权重法和互信息权重法。其均方差法正确率比未加权方法高 0.92%,比熵权重法高 2.68%,比互信息权重法高 0.17%。这说明均方差法特征赋权可以进一步达到突出重要属性并抑制冗余/次要属性的目标,从而提高 DGA 故障诊断正确率。

3.2.3 与其他智能化诊断算法的实验对比

在此,将 FTW_Pearson 算法与贝叶斯算法和神经网络算法进行了比较。表 6 给出了实验对比结果。其中,FTW_Pearson 算法采用的是 log 变换和均方差加权的最终结果;贝叶斯算法的离散化处理采用的是等宽离散,表列 Bayes_2 和 Bayes_5 表示离散化为 2 和 5 箱后的 Bayes 分类结果;神经网络为 3 层,输入、隐含和输出层节点分别为 8、20 和 7 个,激励

函数为 tansig,表列 BPNN_0.01 和 BPNN_0.1 表示学习速率 η 为 0.01 和 0.1 时的分类结果。从中可以看出,采用了 log 变换和均方差加权的 FTW_Pearson 算法显著优于 Bayes 算法和 BPNN 算法。

表 6 FTW_Pearson 算法、Bayes 算法和 BPNN 算法实验对比

	Bayes_2	Bayes_5	BPNN_0.01	BPNN_0.1	FTW_Pearson
1	0.3343	0.2858	0.4172	0.4713	0.7115
2	0.3372	0.3286	0.4258	0.4686	0.7086
3	0.3371	0.3343	0.4285	0.4571	0.6972
4	0.3229	0.2942	0.3943	0.4314	0.7229
5	0.3372	0.3085	0.4457	0.4913	0.7087
6	0.3315	0.2915	0.4572	0.4599	0.7142
7	0.3370	0.3285	0.4200	0.4428	0.6999
8	0.3286	0.3085	0.4029	0.5000	0.6972
9	0.3372	0.3172	0.4257	0.5028	0.7142
10	0.3343	0.3142	0.4257	0.4886	0.7285
平均	0.3337	0.3111	0.4243	0.4714	0.7103

结束语 本文针对电力系统 DGA 故障诊断问题,分析了 Pearson 相关系数 DGA 故障诊断方法。指出现有应用存在两个不足:1)偏袒数据变化范围大的特征;2)假定所有特征贡献相同。针对这两个问题,本文提出了一种基于特征变换和特征权重的 Pearson 相关系数 DGA 故障诊断方法并构造了 FTW_Pearson 算法。该方法首先采用对数特征变换来缩小变化范围大的特征的值域,从而解决偏袒问题;然后采用均方差特征权重方案来为各个特征赋权,从而达到突出重要特征的目的。实际数据集上的实验表明,FTW_Pearson 算法的 DGA 诊断正确率比大卫三角形方法提高了 37%,比传统的 Pearson 相关系数法提高了 8%。同时,实验也表明对数变换优于 Max-min 变换和反正切变换;均方差权重方案优于熵值法权重和互信息权重方案。

参考文献

[1] 蔡红梅,陈剑勇,苏浩益.基于灰云模型的电力变压器故障诊断[J].电力系统保护与控制,2012,40(12):151-155
Cai Hong-mei, Chen Jian-yong, Su Hao-yi. Fault diagnosis of power transformer based on grey cloud model[J]. Power System Protection and Control, 2012, 40(12): 151-155

[2] 朱继,喻瑛,王辰炜,等.基于粗糙集和自适应遗传算法的电力变压器故障诊断[J].电测与仪表,2012,49(6):47-51
Zhu Ji, Yu Ying, Wang Chen-wei, et al. Application of Rough Set and Adaptive Genetic Algorithm to Transformer Fault Diagnosis[J]. Electrical Measurement & Instrumentation, 2012, 49(6): 47-51

[3] 白翠粉,高文胜,金雷,等.基于 3 层贝叶斯网络的变压器综合故障诊断[J].高电压技术,2013,39(2):330-335
Bai Cui-fen, Gao Wen-shen, Jin Lei, et al. Integrated Diagnosis of Transformer Faults Based on Three-layer Bayesian Network[J]. High Voltage Engineering, 2013, 39(2): 330-335

[4] 尹金良,朱永利.支持向量机参数优化及其在变压器故障诊断中的应用[J].电测与仪表,2012,49(5):11-16
Yin Jin-liang, Zhu Yong-li. Parameter Optimization for Support Vector Machine and Its Application to Fault Diagnosis of Power Transformers[J]. Electrical Measurement & Instrumentation, 2012, 49(5): 11-16

[5] 张思扬,匡芳君,徐蔚鸿.混沌动态模糊神经网络及变压器故障诊断的应用[J].湖南科技大学学报(自然科学版),2012,27(1):35-39
Zhang Si-yang, Kuang Fang-jun, Xu Wei-hong. Chaotic dynamic

fuzzy neural network and its application in power transformer fault Diagnosis[J]. Journal of Hunan University of Science & Technology(Natural Science Edition), 2012, 27(1): 35-39

- [6] 赵小霞,苑津莎,刘磊,等. 基于 CBR 与 RBR 变压器检修策略专家系统设计[J]. 电子科技, 2012, 25(7): 31-34
Zhao Xiao-xia, Yuan Jin-sha, Liu Lei, et al. Transformer Maintenance Strategy Expert System Based on CBR And RBR[J]. Electronic Science and Technology, 2012, 25(7): 31-34
- [7] 高俊杰,邓贵仕. 基于本体的范例推理系统研究综述[J]. 计算机应用研究, 2009, 16(2): 406-410, 418
Gao Jun-jie, Deng Gui-shi. Survey on ontology-based CBR system[J]. Application Research of Computers, 2009, 16(2): 406-410, 418
- [8] Negar A, Jean R. An application of multi-criteria decision aids models for Case-Based Reasoning[J]. Information Sciences, 2012, 210(22): 55-66
- [9] Wang J, Zheng N. A Novel Fractal Image Compression Scheme

With Block Classification and Sorting Based on Pearson's Correlation Coefficient[J]. IEEE Transactions on Image Processing, 2013, 22(9): 3690-3702

- [10] Wang Gang-jin, Xie Chi, Chen Shou, et al. Random matrix theory analysis of cross-correlations in the US stock market: Evidence from Pearson's correlation coefficient and detrended cross-correlation coefficient[J]. Physica A: Statistical Mechanics and its Applications, 2013, 392(17): 3715-3730
- [11] 王骏,王士同,邓赵红. 特征加权距离与软子空间学习相结合的文本聚类新方法[J]. 计算机学报, 2012, 35(8): 1655-1665
Wang Jun, Wang Shi-tong, Deng Zhao-hong. A Novel Text Clustering Algorithm Based on Feature Weighting Distance and Soft Subspace Learning[J]. Chinese Journal of Computers, 2012, 35(8): 1655-1665
- [12] He Z Y, Xu X F, Deng S C. Attribute Value Weighting in k-Modes Clustering [J]. Expert Systems with Applications, 2011, 38(12): 15365-15369

(上接第 221 页)

自动获取事件、自动分类事件、自动补全事件关系等,以使所提分类体系在机器翻译、信息自动提取、自然语言处理、文本挖掘等领域获得应用。

参考文献

- [1] Lenat D. CYC: A Large-scale Investment in knowledge Infrastructure[J]. Communications of the ACM, 1995, 38(11): 33-38
- [2] Singh P. The public acquisition of commonsense knowledge [C]// Proc. of the AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access. 2002
- [3] 陆汝铃,张松懋,石纯一,等. 面向 Agent 的常识知识库[J]. 中国科学 E 辑, 2000, 30(5): 453-463
Lu Ru-qian, Zhang Song-mao, Shi Chun-yi, et al. The Agent-oriented commonsense knowledge base [J]. Science China (Series E), 2000, 30(5): 453-463
- [4] 张松懋. 基于知识和推理的分布式交通控制系统[J]. 计算机研究与发展, 1992, 29(9): 35-41
Zhang Song-mao. A distributed traffic control system based on knowledge and reasoning[J]. Journal of Computer Research and Development, 1992, 29(9): 35-41
- [5] 李闪闪,曹存根. 事件前提和后果常识知识分析方法研究[J]. 计算机科学, 2013, 40(4): 185-192
Li Shan-shan, Cao Cun-gen. Commonsense Knowledge Analysis Approach Based on Event Preconditions and Effect[J]. Computer Science, 2013, 40(4): 185-192
- [6] 彭会良,曹存根. 相关事件挖掘与角色联系发现的研究[J]. 计算机科学, 2010, 37(12): 149-155
Peng Hui-liang, Cao Cun-gen. Research on Mining of Associated Events and Discovery of Roles Relationship[J]. Computer Science, 2010, 37(12): 149-155
- [7] 朱耀. 从大规模 Web 语料中获取常识语料[D]. 北京: 中国科学院, 2008
Zhu Yao. Acquiring Commonsense Corpora from Large Scale Web Corpora [D]. Beijing: Chinese Academy of Sciences, 2008
- [8] 田雯. 人类心理常识的形式化研究[D]. 北京: 中国科学院, 2004

Tian Wen. Research on the formalization of human psychological knowledge[D]. Beijing: Chinese Academy of Sciences, 2004

- [9] Zang Liang-jun, Cao Cong, Cao Ya-nan, et al. A Survey of Commonsense Knowledge Acquisition[J]. Journal of Computer Science and Technology, 2013, 28(4): 689-719
- [10] 鲁川. 知识工程语言学[M]. 北京: 清华大学出版社, 2010: 245-248
Lu Chuan. Linguistics for Knowledge Engineering [M]. Beijing: Tsinghua University Press, 2010: 245-248
- [11] 张文华,付晓琛,邓芳. 生物系统分类体系的建立和林奈的贡献[J]. 生物学通报, 2008, 43(5): 54-56
Zhang Wen-hua, Fu Xiao-chen, Deng Fang. Establishing of the classification system about biological systems and the contribution of Linnean [J]. Bulletin of Biology, 2008, 43(5): 54-56
- [12] 胡楠. 交易域商品买卖类框架语义分析研究[D]. 上海: 上海师范大学, 2009
Hu Nan. An Analysis of Frame Semantic in the Domain of Transaction[D]. Shanghai: Shanghai Normal University, 2009
- [13] IEEE Working Group. Suggested Upper Merged Ontology (SUMO)[DB/OL]. <http://www.ontologyportal>
- [14] Allen J F. Towards a General Theory of Action and Time[J]. Artificial Intelligence, 1984, 23(2): 123-154
- [15] 鲁川, 蔡瑞隆, 董丽萍. 现代汉语基本句模[J]. 世界汉语教学, 2000(4): 11-24
Lu Chuan, Gou Rui-long, Dong Li-ping. Modern Chinese Basic Sentence model [J]. Chinese Teaching in the World, 2000(4): 11-24
- [16] Liu H, Singh P. ConceptNet-A Practical Commonsense Reasoning Tool-Kit[J]. BT Technology Journal, 2004, 22(4): 211-226
- [17] 江利萍, 张再跃. 社会群体角色本体建模及其公理获取[J]. 计算机工程, 2012, 38(10): 41-44
Jiang Li-ping, Zhang Zai-yue. Social Group Role Ontology Modeling and Its Axioms Acquisition [J]. Computer Engineering, 2012, 38(10): 41-44
- [18] Collin B. Welcome to Release 1.3 of the FrameNet data[EB/OL]. <http://framenet.icsi.berkeley.edu/fnUsers/currentUsers>