

灰度关联理论在 CBR 中的应用研究

王宏宇,倪志伟,严 军,韩 丹

(合肥工业大学 管理学院,安徽 合肥 230009;

合肥工业大学 过程优化与智能决策教育部重点实验室,安徽 合肥 230009)

摘 要:针对基于规则推理技术(RBR)知识获取困难、自学习能力差等缺陷,将基于案例推理技术(CBR)引入故障诊断系统中。介绍了基于案例推理的故障诊断方法的工作机理和过程模型,阐述了案例表示、案例检索、案例保存和案例库维护机制,然后简单介绍了灰色关联理论知识,并把灰色关联理论应用到故障案例相似度的计算中。根据实验结果可知,该方法有效地改进了案例检索算法,提高了故障案例匹配的准确度和检索效率,同时具有较好的分辨率。

关键词:案例推理;灰色关联理论;故障诊断

中图分类号:TP399

文献标识码:A

文章编号:1673-629X(2010)05-0096-04

Research on Application of Grey - Relational Theory in CBR

WANG Hong-yu, NI Zhi-wei, YAN Jun, HAN Dan

(School of Management, Hefei University of Technology, Hefei 230009, China;

Ministry of Education Key Lab. of Process Optimization and Intelligent Decision - making,

Hefei University of Technology, Hefei 230009, China)

Abstract: Because of the defects of rule - based reasoning (RBR) technology, such as knowledge acquisition difficulties and poor self - learning ability, introduces case - based reasoning into fault diagnosis system, and explains the mechanism and process model of the CBR - based fault diagnosis system. It explains case expression, search mechanism, case preservation and case - based maintenance in detail, and then uses grey - relational theory in calculation of similarity degree in CBR. According to the experimental results, know that the method effectively ameliorates the case retrieval algorithm, and greatly improves the accuracy of the matching degree of fault cases and the search efficiency, and also has a good resolution.

Key words: case - based reasoning; grey - relational theory; fault diagnosis

0 引 言

在故障诊断领域中,基于规则推理(Rule Based Reasoning, RBR)系统的应用时间较长,其利用领域专家知识建立相应规则库进行求解问题。但它通常面临知识获取困难、专业知识难以清楚表达以及缺乏自学习能力等问题^[1]。而基于案例推理(Case Based Reasoning, CBR)技术正好克服了知识获取和自学习的瓶颈问题,成为当今故障诊断领域的研究热点。CBR是用案例来表达知识并把问题求解和学习相融合的一种推理方法,一般适用于理论抽象较弱而经验性因素较

强的应用领域,但随着人工智能领域的发展已越来越受重视并被广泛应用于各个领域之中^[2]。文中将CBR应用到汽车故障诊断的推理中,并针对实际诊断中故障特征信息不完整的问题,在相似度计算中引入了灰色关联理论,有效地提高了检索效率和检索分辨率。

1 CBR 故障诊断过程

CBR不仅是一种理论或技术,更是一种从人类认知学角度模拟人类解决复杂问题的最原始的思维方式。一般,CBR系统把过去处理过的问题,描述成由问题特征集和解决方案组成的事例(case),存储在系统的事例库(case base)中。当新的问题到来时,系统从事例库中检索出与新问题最相似的旧事例。如果旧事例和新事例完全一致,则自然可以把旧事例的解决方案作为新问题的答案;否则需要对最为近似的旧事例进行修改,从而得到新问题的解决方案^[3]。这样,就完

收稿日期:2009-09-09;修回日期:2009-12-10

基金项目:国家高技术研究发展计划(863)(2007AA04Z116);国家自然科学基金(70871033)

作者简介:王宏宇(1984-),男,硕士生,研究方向为案例推理与案例决策;倪志伟,教授,博士生导师,研究方向为人工智能、机器学习应用。

成了对新问题的解答。在大多数情况下,新解决的问题将被加入到事例库中,实现 CBR 的学习过程。一个典型的 CBR 故障诊断过程的模型^[4]如图 1 所示。

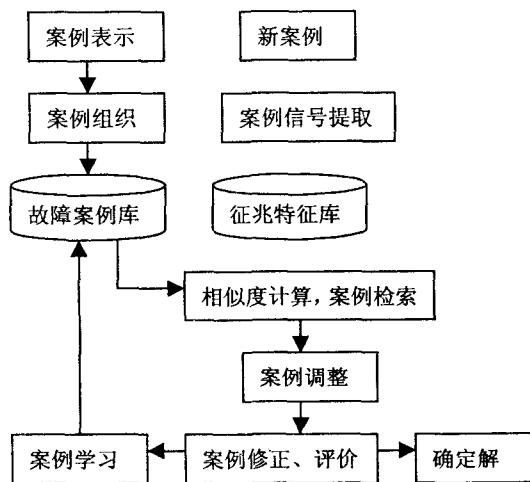


图1 CBR故障诊断过程模型

2 CBR故障诊断的实现

2.1 案例表示与组织

输入的案例首先必须被描述为系统可以识别的方式,然后才可以进行检索、修改等。案例表示主要是用一定的数据结构来描述案例的特征以及案例之间的关系,一般涉及这样几个问题:选择什么信息存放在一个案例中;如何选择合适的案例内容描述结构;案例库如何组织和索引等。案例的表示方式不仅决定着现实世界问题及其相关知识向案例的转换是否合理,而且影响着整个 CBR 系统的各个环节。

一般来说,一个典型的案例可以结构化为案例的原始描述、案例的特征向量和案例的结果集三部分。常用的案例表示方法有:剧本表示法、框架表示法、谓词逻辑表示法、语义网络表示法、自然语言表示法、面向对象的表示方法等^[5]。根据不同的问题,案例的表示选择不同的方法。在该故障诊断系统中将案例模型表示为:

$\text{faultCase} = \{\text{CaseID}, \langle f_1, s_1 \rangle, \dots, \langle f_i, s_i \rangle, \dots, \langle f_n, s_n \rangle, r, e\}$

式中 CaseID 表示案例号, $\langle f_i, s_i \rangle$ 分别表示案例的第 i 个主要特征以及相应的匹配权值, r 表示可能的故障原因和解决方法, e 表示故障诊断的其他一些解释内容。

案例的组织是在案例表示的基础上,根据案例的特征和检索的需要,对案例进行整理和归类,通常包括案例索引结构和案例的存储结构^[6]。案例的索引反映了不同案例之间的区别,对于检索效率有着重要作用。文中根据汽车故障的特点及对故障案例特征的分析,

选择故障部位作为案例索引,并采取层次结构将案例存储在案例库中,从而确保了案例检索的效率。

2.2 基于灰色关联理论案例检索

20 世纪 80 年代,邓聚龙创立了灰色系统理论,其中的灰色关联分析理论是灰色系统分析和处理随机量的一种方法,也是一种数据到数据的“映射”^[7]。它能够依据空间理论的数学基础,按照规范性、偶对称性、整体性和接近性的原则,根据系统动态过程发展态势来确定参考序列和若干比较序列之间的关联系数和关联度。而且灰色关联系数的计算公式为案例间相似度的计算提供了一种新方法。

在基于 CBR 的故障诊断系统中,案例检索是最为核心的步骤。CBR 的检索要达到以下两个目标:一是检索出来的事例应尽可能的少;二是检索出来的事例应尽可能与当前事例相关或相似^[8]。如何在 CBR 中高效地完成事例的检索是十分重要的,对问题求解的性能有直接影响。案例的检索策略通常有最近相邻策略、归纳推理策略和知识引导策略等。文中首先利用关键指标对案例库进行初步检索,然后采用灰色关联理论与欧几里德距离相结合的方法来计算待分析案例与案例库中的案例之间的相似度,最后取相似度最大的案例作为诊断结果。

获得待分析故障的征兆向量后,利用其中权重系数最大的关键征兆在案例库中进行初步检索,筛选出含有该故障征兆的案例集。设论域 X 为初步检索得到的 m 个案例的故障特征向量集, x_0 为当前故障的特征向量, $x_0 = (x_0(1), x_0(2), \dots, x_0(n))$, 则 X 可以表示为 $m \times n$ 维的案例征兆指标集矩阵:

$$X = [x_1 \ x_2 \ \dots \ x_n]^T = \begin{bmatrix} x_1(1) & x_1(2) & \dots & x_1(n) \\ x_2(1) & x_2(2) & \dots & x_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ x_m(1) & x_m(2) & \dots & x_m(n) \end{bmatrix}$$

为消除不同量纲对指标值的影响,案例匹配计算之前需要对 X 和 x_0 中的元素进行归一化处理。文中采用向量归一化法,即:

$$x'_i(k) = \frac{x_i(k)}{\sqrt{\sum_{i=0}^m x_i^2(k)}} \quad 0 \leq i \leq m, 1 \leq k \leq n \quad (1)$$

经过向量归一处理后,得到新的案例指标集矩阵 X' 和新的当前待分析故障征兆向量 x'_0 , 其中各指标值均满足 $0 \leq x'_i(k) \leq 1$ 。这样,案例检索问题就转化为 x'_0 与 X' 中征兆向量值的相似度计算问题。

由于各个征兆向量在故障中的重要程度不一样,在计算相似度时需要将权重纳入比较环境的计算中。

根据灰色关联理论^[7],当前故障征兆向量 x_0 与案例集内第 i 个案例向量 x_i 在第 k 个指标值上的灰关联相似度见式(2)。

式中, $i \in m, k \in n; \zeta \in [0, 1]$ 为分辨系数, 作用为调整比较环境的大小, 在最小信息原理下取 $\zeta = 0.5$; $\min_k (\omega_k |x'_0(k) - x'_i(k)|)$ 和 $\max_k (\omega_k |x'_0(k) - x'_i(k)|)$ 为 x'_0 和 x'_i 在第 k 个征兆特征指标的比较环境; ω_k 为第 k 个征兆特征指标的所占权重。

相似度和距离都是判断相似性的尺度。依据距离可很容易地剔除那些与待分析案例仅在某一个特征方面差别较大的案例, 但是它在特征值不明的情况下难以计算相似距离。根据文献[9]相似度和距离之间的转换公式, 当前故障征兆向量 x_0 与案例集内第 i 个案例向量 x_i 在第 k 个指标值上的灰关联距离为:

$$\text{DIST}(x_0(k), x_i(k)) = \frac{1}{\text{SIM}(x_0(k), x_i(k))} - 1 \quad (3)$$

显然, 相似度的取值范围为 $(0, 1]$, 距离的取值范围为 $[0, \infty)$ 。函数 $\text{DIST}(x_0(k), x_i(k))$ 反映了当前故障征兆向量 x_0 与案例集内第 i 个案例向量 x_i 在第 k 个指标上基于灰色关联理论的相对距离。在 n 维空间上根据欧几里德距离公式, 则当前向量 x_0 与案例向量 x_i 的灰色距离为:

$$\text{DIST}(x_0, x_i) = \sqrt{\sum_{k=1}^n \text{DIST}^2(x_0(k), x_i(k))} \quad (4)$$

然后根据式(3), 当前向量 x_0 与案例向量 x_i 在整个 n 维指标空间上的复合灰色相似度为:

$$\text{SIM}(x_0, x_i) = \frac{1}{1 + \text{DIST}(x_0, x_i)} \quad (5)$$

根据上述公式可以计算待分析故障与各案例之间的灰色相似度。在实际使用时, 可以根据具体情况设定一个阈值, 只有当相似度大于该阈值时, 才认为当前故障与这一案例相似。如果检索结果有多个案例的相似度大于阈值, 则应考虑推荐相似度最大或者使用频率最高的案例作为结果案例。当没有任何案例与当前待分析故障的相似度大于阈值时, 则认为此故障为新故障。

2.3 案例修正与保存

在 CBR 系统中, 从案例库中检索到与当前案例完全匹配的旧案例的情况并不常见, 一般只能找到若干与当前案例比较相似的旧案例, 然后再进行修正, 使其

能适应新情况。通常, 对案例修正分为两步进行: 首先重用旧案例的解决方案, 并对复用结果进行评估, 分析新旧案例之间的不同之处, 然后以新案例为基础进行修正, 即在旧的解决方案中增加、删除或替换某些内容^[5]。修正后的案例经过验证, 如果是可行的或正确的, 就可作为新的案例存储到案例库中。

如果在检索过程中, 没有任何案例与当前案例相似或相似度大于设定的阈值, 则认为此案例为新的案例。对于新案例, 可以重用得出以前案例解决方案的规则或公式, 根据新的条件推导出新的解决方案, 也可以向专家询问完成, 然后应用到实际环境依据反馈进行调整。当问题得到解决, 就可将其存入案例库中, 以便用于将来与之相似的案例的检索重用。

2.4 案例库维护

在案例推理系统中, 系统的增量式学习会使案例库无限增大, 一般说案例库越大, 案例越丰富, 能解决更多的问题, 体现较高的智能。但随之而来也会导致很多问题: 超出案例库的设计的大小限制, 检索效率下降(如沼泽问题), 知识出现冗余, 而且还可能出现矛盾等等, 因此需要对案例库进行维护。这也是 CBR 目前研究的一个热点。常见的案例维护方法有: 随机删除法、维护规则方法、基于 Agent 的维护方法以及基于案例分类的删除策略等^[6]。

3 应用实例

文中对文献[10]中提供的实例进行了计算, 并采用最近邻检索算法和余弦函数算法对计算结果进行了验证分析。故障特征值如表 1。

表 1 故障特征值

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9
x_1	1318	12	84	100	590	98	55	104	275
x_2	131	8	24	55	13	6	12	6	7
x_3	240	10	51	42	94	10	13	10	9
x_4	1466	11	128	157	863	84	51	59	113
x_5	655	7	102	364	98	18	20	19	27
x_0	456	10	77	93	176	15	23	32	30

利用式(1)对表 1 中的数据作量纲归一化处理, 可以得到 x'_0 和 X' (见式(6)、(7)):

$$\text{SIM}(x_0(k), x_i(k)) = \frac{[\min_k (\omega_k |x'_0(k) - x'_i(k)|) + \zeta \max_k (\omega_k |x'_0(k) - x'_i(k)|)]}{[\omega_k |x'_0(k) - x'_i(k)| + \zeta \max_k (\omega_k |x'_0(k) - x'_i(k)|)]} \quad (2)$$

$$x'_0 = [0.21266 \ 0.41595 \ 0.37154 \ 0.21885 \ 0.16466 \ 0.11389 \ 0.27753 \ 0.25443 \ 0.09992] \quad (6)$$

$$X' = \begin{bmatrix} 0.61465 & 0.49913 & 0.40532 & 0.23532 & 0.55199 & 0.74411 & 0.66366 & 0.82691 & 0.91589 \\ 0.06109 & 0.33276 & 0.11581 & 0.12943 & 0.01216 & 0.04556 & 0.14480 & 0.04771 & 0.02331 \\ 0.11192 & 0.41595 & 0.24609 & 0.09884 & 0.08794 & 0.07592 & 0.15687 & 0.07951 & 0.02997 \\ 0.68367 & 0.45754 & 0.61763 & 0.36945 & 0.80740 & 0.63781 & 0.61540 & 0.46911 & 0.37635 \\ 0.30546 & 0.29116 & 0.49217 & 0.85657 & 0.09169 & 0.13667 & 0.24133 & 0.15107 & 0.08992 \end{bmatrix} \quad (7)$$

然后利用式(2)进行相似度计算,其中设定各个故障特征的权重为 $\omega_k = (0.25, 0.025, 0.025, 0.1, 0.2, 0.05, 0.05, 0.1, 0.2)$,取分辨系数 $\zeta = 0.5$;再利用式(3)~式(5)进行欧氏距离和相似度之间的转换,可得当前案例与以往案例集中各个案例之间的相似度。如表2所示。

表2 当前故障与各案例之间的相似度

	$SIM(x_0, x_1)$	$SIM(x_0, x_2)$	$SIM(x_0, x_3)$	$SIM(x_0, x_4)$	$SIM(x_0, x_5)$
相似度	0.273	0.590	0.673	0.303	0.536

根据表2计算的结果,可知案例 x_3 与案例 x_0 的相似度最大,即当前案例 x_0 与案例 x_3 特征向量最为相似,因此在解决此故障时可以借鉴案例 x_3 的解决方案。

同时,利用最近邻检索算法和余弦函数算法^[11]计算出的相似度如表3所示。

表3 最近邻与余弦函数算法计算的相似度

	$SIM(x_0, x_1)$	$SIM(x_0, x_2)$	$SIM(x_0, x_3)$	$SIM(x_0, x_4)$	$SIM(x_0, x_5)$
最近邻算法	0.546	0.868	0.905	0.612	0.877
余弦函数算法	0.515	0.868	0.908	0.572	0.900

根据表3的结果,同样可知案例 x_3 与案例 x_0 的相似度最大,从而证明了文中算法是正确有效的;但就某一算法而言,表2数据比表3数据更加分散,即文中算法比另外两种算法具有更好的分辨率,在同等条件下可以更容易地得到正确的结果。

4 结束语

文中将灰色关联理论引入到 CBR 中,并结合欧氏距离计算故障案例之间的相似度问题,提高了检索效

率。同时, CBR 作为新兴的人工智能推理技术,与其它人工智能推理和学习方法的有效结合和应用成为其近些年来一个发展方向。

参考文献:

[1] 许六一,李建璜,胡柏青,等.一种改进的 CBR 故障诊断方法研究[J].微计算机信息,2007,23(3):216-218.

[2] LIU Jia li, YAN Xiang bin, QI Wei. A Case-based Reasoning System for Mechanical Design[C]//International Conference on Management Science & Engineering. California: IEEE,2008.

[3] 艾芳菊. CBR 中的检索模型研究[J].计算机工程与应用,2005(19):77-79.

[4] 李小青.基于案例推理的故障诊断方法[J].计算机测量与控制,2007,15(9):1130-1131.

[5] 杨善林,倪志伟.机器学习与智能决策支持系统[M].北京:科学出版社,2004:79-112.

[6] 倪志伟,李锋刚,毛雪岷.智能管理技术与方法[M].北京:科学出版社,2007:34-39.

[7] 邓聚龙.灰理论基础[M].武汉:华中科技大学出版社,2002:135-150.

[8] 郭茂祖,苏晓红,王亚东,等.基于 IBL 算法的 CBR 系统中索引与检索机制研究[J].计算机工程与应用,2001(5):67-69.

[9] Burkhard H D. Similarity and Distance in Case Based Reasoning[J]. Fundamenta Informaticae,2001,47:201-215.

[10] 王东,刘怀亮,徐国华.案例推理在故障诊断系统中的应用[J].计算机工程,2003(12):10-12.

[11] 陈富民,吴垌沅,林志航.用于故障诊断的案例匹配算法分析[J].计算机应用研究,2008,25(5):1352-1354.

(上接第95页)

[M]. [s.l.]: Software Quality Research Laboratory,2000:49-53.

[13] Kullback S. Information theory and statistics[M]. New York: Wiley,1958:10-83.

[14] Sayre K, Poore J H. Stopping criteria for statistical testing[J]. Information and Software Technology,2000,42(12):851-857.

[15] Dulz W, Zhen F, Ma Te Lo. Statistical usage testing by annotated sequence diagrams, Markov chains and TTCN-3[C]//Proceedings of the Third International Conference on Quality Software (QSIC'03). [s.l.]: [s.n.],2003:336-342.

[16] Guen H L, Thelin T. Practical experiences with statistical usage testing[C]//Proceedings of the 11th Annual International Workshop on Software Technology and Engineering Practice (STEP'04). [s.l.]: [s.n.],2004:87-93.