

文章编号: 1005-8451 (2005) 11-0006-04

铁路信号设备故障诊断知识表示方法的研究

殷亚平, 张 喜

(北京交通大学 交通运输学院, 北京 100044)

摘 要: 针对铁路车站信号设备复杂故障分类决策问题, 提出并建立基于ID5R算法的故障诊断知识表示与知识库开发的实际应用方法。最后, 通过实例分析验证应用ID5R算法建立故障诊断专家系统知识库的有效性。

关键词: 熵; ID5R 算法; 决策树学习; 故障诊断

中图分类号: U284.7 **文献标识码:** A

Research on knowledge representation way of railway signals malfunctions diagnosis

YIN Ya-ping, ZHANG Xi

(School of Traffic and Transportation of Beijing Jiaotong University, Beijing 100044, China)

Abstract: According to the classification decision questions of the complex railway signals malfunctions diagnosis, it was given the practical application way of knowledge representation and knowledge base development based on ID5R algorithm. Through an instance, it was proved the algorithm which was useful to form the knowledge base of Malfunctions Diagnosis Expert System.

Key words: entropy; ID5R algorithm; decision tree learning; malfunctions diagnosis

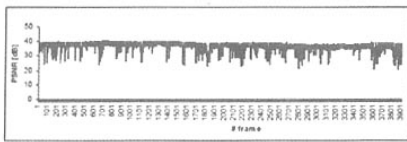
铁路信号系统是一个由多种机电设备组成的复杂的控制系统。实际信号系统的故障现象具有多样性, 很多复杂故障产生的原因往往具有模糊性、随机性和组合性等特点。如何基于大量的信号设备故

障实例信息获取领域知识, 并采用适当的知识形式化表示方法建立知识库, 是建造实际铁路信号设备故障诊断专家系统的一个重要内容, 也是需待研究的一项课题。

决策树学习算法是以实例信息为基础的机器学习方法, 也称为归纳学习算法^[1]。它着眼于从环境提

收稿日期: 2005-05-17

作者简介: 殷亚平, 在读硕士研究生; 张 喜, 教授。



(b) 采用区分服务的视频 PSNR 值

图2 区分服务与 best-effort 服务的视频 PSNR 值对比

4 结束语

本文提出了一种在 DS 域中传输 MPEG-4 视频的 QoS 映射机制, 并对提出的方案进行 NS 仿真。在网络带宽不足的情况下, 该机制能尽量保证重要的数据的转发, 达到较好的视频质量。仿真结果表明, 与传统的尽力而为服务相比, 区分服务在视频的 PSNR 方面有很大提高。

参考文献:

- [1] S. Floyd, and V. Jacobson, Random Early Detection Gateways for Congestion Avoidance [J]. IEEE/ACM Transactions on Networking, 1993, 1(4): 397 — 413.
- [2] K. Fall, and K. Varadhan, The NS Manual-Notes and Documentation [EB/OL]. <http://www.isi.edu/nsnam/ns/tutorial/>, 2001.
- [3] F. H. P. Fitzek, and M. Reisslein, MPEG-4 and H.263 Video Traces for Network Performance Evaluation [EB/OL]. <http://www.eas.asu.edu/trace/TRACE/trace>, 2002.
- [4] K. Bo-Kyoung, Simulation Study of Weighted Round Robin Queuing Policy [J]. Masters Thesis, University of Massachusetts Lowell, 1993, 22: 63 — 72.
- [5] Jirka Klaue, Berthold Rathke, and Adam Wolisz, EvalVid - A Framework for Video Transmission and Quality Evaluation [EB/OL]. <http://www.tkn.tu-berlin.de/research/evalvid/>, 2003.

供的一组无次序、无规则的实例（特殊知识）信息中归纳推理出领域知识的一般规则，即决策树知识表示形式的分类规则（产生式规则），从而实现知识库的建立。本文结合铁道部计划项目“铁路车站信号设备故障诊断专家系统”的研制开发工作，应用 ID5R 决策树学习算法，对铁路信号系统复杂故障诊断的知识表示方法和知识库的建立进行了分析研究。

1 基于决策树的故障分类知识表示方法

决策树概念中的树（tree）是由结点（node）和分支（branch）组成的层次数据结构，如图 1 所示。结点用于存储信息或知识，分支连接各结点，有时分支也称为连接或边，而结点称为顶点（vertex）。在一棵有向树（oriented tree）中，根结点处于最顶层（hierarchy），而叶结点（leafs）处于最底层。其中，除根结点外，每个结点只有一个双亲结点（parent）以及有 0 个或多个子结点（child node）。

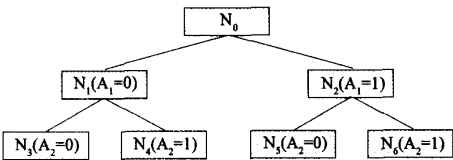


图 1 故障分类决策树

当用树来做决策判定时，称为决策树（decision tree）。例如，表 1 所示的故障分类决策问题，它需要根据对象设备的属性 A1 和 A2 的取值，确定故障的类别 C0、C1、C2 和 C3。这一决策问题可由一个如图 1 所示的决策树实现。故障分类决策从结点 N0 开始，测试属性 A1，根据它的取值（0 或 1）决定到达子结点 N1 还是 N2，在结点 N1 或 N2，再测试属性 A2，根据 A2 的取值分别决定到达 N3（分类 C0）或 N4（分类 C1）或 N5（分类 C2）或 N6（分类 C3），从而对故障类别做出决策。

表 1 故障分类决策表

设备属性值		故障分类
A1	A2	
0	0	C0
0	1	C1
1	0	C2
1	1	C3

由故障分类决策树的例子可见，决策树的决策过程是从根结点开始，采用自顶向下的递归方式，在决策树的内部结点（非叶结点）进行属性值的比

较并根据不同的属性值判断从该结点向下的分支，在决策树的叶结点得到结论。根结点表示决策的开始条件，内部非叶结点表示基于判断标准的决策信息处理机构，叶结点表示决策结果。所以，从根结点到叶结点的一条路径就对应着求解决策问题的一条规则，整棵决策树就对应着一组规则，称为决策树分类规则。一条决策树分类规则就是确定故障分类决策的一个知识，也就是说基于这个知识即可判定故障的类型。这正是利用决策树建立故障诊断专家系统知识库的基本依据。基于决策树分类规则的知识表示方法，一般可采用产生式规则法，即通过结构（决策树）进行广度优先搜索，并在每个结点生成“IF...THEN”规则来实现。

2 复杂故障分类决策树学习算法的基本构成

基于决策树分类规则的知识表示方法，是建立故障诊断专家系统知识库的有效方法。但对于实际铁路信号系统的复杂故障，由于设备属性值与故障分类间的关系存在模糊性和随机性（不具有确定的一一对应关系），一般并不具备明显的决策树结构，也就不可能直接采用决策树分类规则的知识表示方法进行故障分类决策。所以，针对这种复杂故障分类决策问题，需要寻求新的解决方法。

ID5R 决策树学习算法是求解复杂分类决策问题的一种方法。其基本思想是：基于大量的故障统计实例信息（训练例集合：故障分类决策统计表），通过学习归纳操作（以信息熵函数作为学习归纳操作的评价标准），从实例信息中获得以近似决策树形式表示的知识（决策树分类规则）。

ID5R 算法以信息熵函数作为学习归纳操作的评价标准，设复杂分类决策问题的实例（训练例）信息集为 $S=\{A, \delta\}$ ，其中 A 表示设备属性信息集合，单个属性用 a_i 表示，则有 $A=\{a_i, i \in (1,2,\cdots, |A|)\}$ ， $|A|$ 是 A 所包含的属性数目，对每个属性 a_i ，其可能的取值集合用 V_i 表示，单个值用 V_{ij} 表示，则有 $V_i=\{V_{ij}, j \in (1,2,\cdots, |V_i|)\}$ ， $|V_i|$ 是属性 a_i 可能取的值的数目； δ 表示分类信息集合，共有 m 种分类，记为 $\delta_i, i=1,2,\cdots,m$ 。令 $S_j^i (i=1,2,\cdots,m)$ 为 S 中对应属性 a_i 的值 V_{ij} 且分类为 δ_i 的训练例的数目，则可构造如下式所示的评价函数。

$$E(a_i)=-\sum_{j=1}^{|V_i|}\sum_{i=1}^m S_j^i \log_2 (S_j^i/(S_j^1+S_j^2+\cdots+S_j^m))$$

该评价函数以分类概率为依据,是一个基于熵(平均信息量)的函数,可用于评价选择各属性值进行分类所能获得的信息量,选择一个E值最小的属性实际上就是选择一个获得信息量最大的属性。

基于上述概念的ID5R决策树学习算法的基本构成如下:

步骤1:如果根结点为空,则把实例放入根结点的实例集,并把该实例的分类作为根结点的分类。此时根结点的实例集中只有一个实例,并且此时的根结点可以被看成是一个叶结点;

步骤2:否则,将实例沿树进行匹配,直到到达一个叶结点;

步骤3:如果实例和该叶结点中的其他实例是同一分类,则把这个实例加入该叶结点的实例集;

步骤4:否则,

(1)将实例加入该叶结点的实例集中,随意选择一个属性对该结点进行划分;

(2)对所有结点的所有属性,根据新增加的实例更新计数(以有序对的形式表示);

(3)如果某结点已用于划分的属性对该结点来说熵函数E值不是最小,则:

a.调用“上拉”过程使该结点用对它来说E值最小的属性进行划分;

b.对它的各子树递归地进行更新,使每个子结点用于划分的属性对该子结点来说都是最优的,其中不包括新加入实例的结点;

(4)对新加入实例的结点,递归地更新该结点的子树,必要时可以生成新的分枝。

上述ID5R算法中的“上拉”过程如下:

(1)如果被“上拉”的属性 a_{new} 已在根结点,则停止;

(2)否则:

a.递归地把属性 a_{new} 沿其与根结点之间的路径逐级“上拉”,必要时可以用 a_{new} 对叶结点进行划分,使其扩展一层;

b.在每一级都对子树进行变换,使子树的根结点用 a_{new} 进行划分,而原来所用的属性 a_{old} 成为它的子代所用的划分属性。

3 铁路信号设备复杂故障诊断专家系统知识库建立的实例应用分析

3.1 铁路信号设备复杂故障诊断决策树的生成

设通过故障统计数据已获得铁路信号设备复杂故障的实例集共有8个训练例,每个训练例有3个属性(3种设备工作状态),分别用“Dev1, Dev2, Dev3”表示,每种属性有3种可能的取值“N”、“Y”和“F”,“N”表示“接通”,“Y”表示“断开”,“F”表示“虚接”;有3种故障分类“ $\delta 1$ ”、“ $\delta 2$ ”、“ $\delta 3$ ”,分别表示“无故障”、“短路故障”、“断线故障”。实例集的训练例信息(故障统计表)如表2所示。

表2 铁路信号设备复杂故障实例集信息表

训练例 序号	信号系统对象设备属性			故障分类 δ
	Dev1	Dev2	Dev3	
1	N	N	N	$\delta 1$
2	Y	N	N	$\delta 3$
3	N	Y	N	$\delta 2$
4	N	F	Y	$\delta 2$
5	N	Y	Y	$\delta 2$
6	Y	N	Y	$\delta 2$
7	Y	Y	N	$\delta 2$
8	Y	Y	Y	$\delta 1$

另设算法中故障分类 $\delta 1$ 、 $\delta 2$ 、 $\delta 3$ 的计数用一个有序对来表示,其形式是 $[\delta 1, \delta 2, \delta 3]$ 。

应用ID5R算法求解该复杂故障诊断问题决策树的实现过程如下:

(1)把第1个训练例加入决策树根结点集,Dev1=N、Dev2=N和Dev3=N属于 $\delta 1$ 类,此时的根结点可以作一个叶结点。

(2)第2个训练例,Dev1=Y、Dev2=N和Dev3=N属于 $\delta 3$ 类,于是树要扩展一层,根据ID5R算法的步骤4(1),随意选择一个属性Dev1进行划分,生成新的决策树。

(3)第3个训练例,Dev1=N、Dev2=Y和Dev3=N属于 $\delta 2$ 类,沿Dev1=N进行匹配,到达叶结点,发现它与其他训练例不是同一分类,于是又根据算法的步骤4(1),随意选择一个属性Dev2进行划分,生成新的决策树。

(4)第4个训练例,Dev1=N、Dev2=F和Dev3=Y属于 $\delta 2$ 类,加入后,沿Dev1=N进行匹配,到达结点Dev2,对于属性Dev2来说,F是一个新出现的取值,因此要生成一个新的分枝,得到新的决策树。经过计算发现 $E(Dev2)=2 < E(Dev1)=2.76 < E(Dev3)=4.77$,于是将属性Dev2“上拉”到根节点,得到新的决策树。

(5)第5个训练例,Dev1=N、Dev2=Y和Dev3=Y加入后,经过计算得 $E(Dev2)=2 < E(Dev1)=3.25 < E(Dev3)=4.77$,所以决策树的结构不用发生变化。

(6) 第 6 个训练例, $Dev1 = Y$ 、 $Dev2 = N$ 、 $Dev3 = Y$ 加入后, 经过计算得 $E(Dev2) = E(Dev3) = 4.77 < E(Dev1) = 5.25$, 于是将属性 $Dev3$ “上拉”, 生成新的决策树。

(7) 第 7 个训练例, $Dev1 = Y$ 、 $Dev2 = Y$ 和 $Dev3 = N$ 加入后, 经过计算知 $E(Dev2) = 4.77 < E(Dev3) = E(Dev1) = 6$, 决策树的结构不用发生变化。

(8) 第 8 个训练例, $Dev1 = Y$ 、 $Dev2 = Y$ 、 $Dev3 = Y$ 加入后, 决策树结构不会发生变化, 得到如图 2 所示的最终决策树。

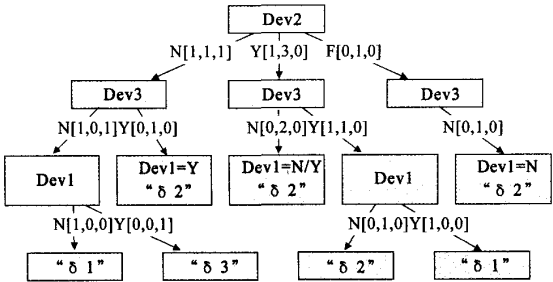


图 2 最终生成的决策树

此实例分析表明, 应用 ID5R 算法不仅能找到描述训练例集合尽量简单的决策树, 而且对于新增训练例无需改变决策树的原有结构, 可以较好地实现增量式学习, 减少算法信息处理的系统开销, 提高决策树生成速度, 具有较强的实用性和有效性。

3.2 铁路信号设备复杂故障诊断专家系统知识库的实现

基于 ID5R 算法生成的决策树, 通过深度优先搜索并在每个结点生成 “IF…THEN” 规则, 即可得到故障诊断知识的产生式规则表示形式, 从而实现故障诊断专家系统知识库的建造。由表 2 所示的铁路信号系统复杂故障诊断分类决策问题的近似决策树图 2, 可得到如下的复杂故障诊断专家系统知识库的程序原码。

```
rule:                                     // 规则 1
if Dev2 is N                             // 前提
    and Dev3 is N
    and Dev1 is N
then Conclude 电路故障 is 无故障        // 结论
rule:                                     // 规则 2
if Dev2 is N                             // 前提
    and Dev3 is N
    and Dev1 is Y
then Conclude 电路故障 is 断线故障      // 结论
rule:                                     // 规则 3
if Dev2 is N                             // 前提
    and Dev3 is Y
```

```
    and Dev1 is Y
then Conclude 电路故障 is 短路故障      // 结论
rule:                                     // 规则 4
if Dev2 is Y                             // 前提
    and Dev3 is N
    and Dev1 is N
then Conclude 电路故障 is 短路故障      // 结论
rule:                                     // 规则 5
if Dev2 is Y                             // 前提
    and Dev3 is N
    and Dev1 is Y
then Conclude 电路故障 is 短路故障      // 结论
rule:                                     // 规则 6
if Dev2 is Y                             // 前提
    and Dev3 is Y
    and Dev1 is N
then Conclude 电路故障 is 短路故障      // 结论
rule:                                     // 规则 7
if Dev2 is Y                             // 前提
    and Dev3 is Y
    and Dev1 is Y
then Conclude 电路故障 is 无故障        // 结论
rule:                                     // 规则 8
if Dev2 is F                             // 前提
    and Dev3 is Y
    and Dev1 is N
then Conclude 电路故障 is 短路故障      // 结论
```

4 结束语

铁路车站信号设备故障种类繁多且分类复杂, 不仅要求能够自动获取知识, 也要求所表示的故障诊断知识具有很高的推理效率, 因此基于实例信息知识库的生成就显得尤为重要。基于 ID5R 算法的知识获取与表示方法, 不仅具有自学习, 自组织能力, 更具有增量式学习功能, 因此可满足复杂故障诊断(分类决策)的需要。本文针对铁路车站信号设备复杂故障分类决策问题, 提出并建立了基于 ID5R 算法的故障诊断知识表示与知识库开发的实际应用方法。该方法已经被应用于铁道部计划项目“铁路车站信号设备故障诊断专家系统”的实际研制开发中, 并取得了较好的效果。

参考文献:

[1] 田盛丰, 黄厚宽. 人工智能与知识工程[M]. 北京: 中国铁道出版社, 1999.

[2] 陈世福, 陈兆乾. 人工智能与知识工程[M]. 南京: 南京大学出版社, 1997.

[3] 史忠植. 知识发现[M]. 北京: 清华大学出版社, 2001.

[4] 王书强, 张喜. 铁路车站信号设备故障诊断专家系统的研究[J]. 铁路计算机应用, 2005 (1): 9—12.