·开发研究与工程应用· 文章编号：1000-3428(2016)09-0310-05 文献标志码：A 中图分类号：TP393.06

# 网络故障管理中基于邻域粗糙集的规则自动生成

洪国栋[1]，闵卫东[2]

（1. 天津工业大学 计算机科学与软件学院，天津 300387；2. 南昌大学 信息工程学院，南昌 330031）

**摘 要：** 针对网络故障管理中的规则手工定义方法未考虑冗余和不准确数据对规则有效性和性能的影响问题，为提高规则匹配效率，提出一种规则自动生成方法。通过邻域粗糙集约简网络故障诊断属性并对约简结果限定阈值进而实现规则自动生成。针对监控数据的规则多匹配问题，设计基于值权重的规则匹配算法，可在发生多匹配时从规则中找出一条与当前监控数据匹配度最高的规则。实验结果表明，与规则手动定义方法相比，该方法能在不降低故障诊断率的情况下，使规则匹配效率平均提升 2.5 倍。

**关键词：** 网络故障管理；规则自动生成；邻域粗糙集；属性约简；规则匹配

**中文引用格式：** 洪国栋，闵卫东. 网络故障管理中基于邻域粗糙集的规则自动生成[J]. 计算机工程，2016，42(9)：310-314.

**英文引用格式：** Hong Guodong，Min Weidong. Automatic Rule Generation Based on Neighborhood Rough Set in Network Fault Management[J]. Computer Engineering，2016，42(9)：310-314.

# Automatic Rule Generation Based on Neighborhood Rough Set in Network Fault Management

HONG Guodong[1]，MIN Weidong[2]

（1. School of Computer Science and Software Engineering，Tianjin Polytechnic University，Tianjin 300387，China；
2. School of Information Engineering，Nanchang University，Nanchang 330031，China）

**【Abstract】** The current methods to define rules manually in the network fault management do not consider the influence of redundancy and inaccurate date on the effectiveness and performance of the rules. Aiming at the problem，an automatic rule generation method is proposed to improve the efficiency of rule matching. It uses neighborhood rough set to reduce the fault attributes of network faults，limits the threshold of reduction results and then generates rules automatically. To solve the problem of multiple rule matching about monitoring data，a rule matching algorithm based on value weight is proposed. It can find a rule that has the highest matching degree with the current monitoring data in the case of multiple matching. Experimental results demonstrate that compared with the method to define rules manually，the proposed method can increase the efficiency of rule matching by 2.5 times without reducing the rate of fault diagnosis.

**【Key words】** network fault management；automatic rule generation；neighborhood rough set；attribute reduction；rule matching

**DOI：** 10.3969/j.issn.1000-3428.2016.09.054

## 1 概述

网络资源管理一直是热门研究领域[1-4]，而网络故障管理是网络管理的重要研究方向。随着网络规模的不断扩大，网络信息资源发生故障不可避免。为了实现网络故障的自动化管理，研究人员提出基于规则的网络故障管理模型。该模型的基本原理是通过一定的故障诊断条件来定义相应的规则，利用监控数据匹配规则来诊断是否有故障发生。然而，在现有的管理模型中，规则通常是通过手工定义获得[5-6]。手工定义的规则无法准确地描述网络故障的特征信息，规则的诊断属性通常包含冗余、不准确的信息。由于手工定义的规则需要匹配冗余的诊断信息，必然会增加规则匹配的时间开销，导致故障诊断效率低下。在已有的规则自动生成研究中，文献[7]提出一种基于遗传编程的安全事件关联规则生成方法。文献[8]提出一种置信规则库的表示、生成和推理方法。文献[9]提出一种自动生成 Snort 内

**基金项目：** 天津市自然科学基金资助项目(13JCYBJC15500)。

**作者简介：** 洪国栋(1990－)，男，硕士研究生，主研方向为网络故障管理；闵卫东(通讯作者)，教授、博士生导师。

**收稿日期：** 2015-09-22 **修回日期：** 2015-11-12 **E-mail：** minweidong@ncu.edu.cn

容规则的方法来处理网络流量分析。但是上述规则生成方法没有考虑冗余和不准确的数据对规则有效性和性能的影响。针对该问题,本文提出一种基于邻域粗糙集的规则自动生成方法。邻域决策系统(Neighborhood Decision System,NDS)由网络故障数据样本构建。其中,决策系统的条件属性的属性值能够用于限定诊断规则的属性阈值,但生成规则的数据样本数量有限,不能完全覆盖诊断属性的阈值范围,不同的规则之间会有一定程度的阈值交叉,造成监控数据会匹配多个规则的问题。为解决该问题,本文提出基于值权重的规则匹配算法。

## 2　邻域粗糙集理论

粗糙集理论是一种描述不完整性和不确定性的数学工具[10],能有效地分析各种不完备的信息,在网络管理中已得到广泛应用[11-12]。传统的粗糙集只能处理离散数据,对于连续数据需要进行离散化才能进行数据约简,而数据离散化会带来一定的信息丢失。为处理连续数据,提出邻域粗糙集[13]。

**定义 1**(邻域决策系统)　给定 $NDS=(U,A,D)$,如果 $A$ 在论域上生成一组邻域关系,则称 $NDS$ 为邻域决策系统。其中,$U$ 为非空有限集,称为论域;$A$ 为条件属性;$D$ 为决策属性。

**定义 2**(邻域)　实数空间 $\Omega$ 上的非空有限集 $U=\{x_1,x_2,\cdots,x_n\}$,$\forall x_i$ 的邻域为 $\delta(x_i)=\{x\mid x\in U,\Delta(x,x_i)\leqslant\delta\}$。

**定义 3**(上下近似)　给定 $NDS=(U,A,D)$,决策属性 $D$ 将论域 $U$ 划分为 $N$ 个等价类($X_1,X_2,\cdots,X_n$),$\forall B\subseteq A$ 生成 $U$ 上的邻域关系 $N_B$,则决策属性 $D$ 关于子集 $B$ 的上下近似分别为 $\overline{N}_BD=\bigcup_{i=1}^{N}\overline{N}_BX_i$,$\underline{N}_BD=\bigcup_{i=1}^{N}\underline{N}_BX_i$,其中,$\overline{N}_BX=\{x_i\mid\delta_B(x_i)\cap X\neq\varnothing,x_i\in U\}$,$\underline{N}_BX=\{x_i\mid\delta_B(x_i)\subseteq X,x_i\in U\}$。

**定义 4**(依赖度)　决策属性 $D$ 对于条件属性 $B$ 的依赖度表示为 $k_D=\gamma_B(D)=\dfrac{|Pos(D)|}{|U|}$。

**定义 5**(重要度)　属性 $a_i$ 在属性集合 $A$ 中相对于决策属性 $D$ 的重要度为 $Sig(a_i,A,D)=\gamma_A(D)-\gamma_{A-\{a_i\}}(D)$。

## 3　基于邻域粗糙集的规则自动生成

由于造成网络故障的因素很难确定,通过管理员和专家知识也无法准确描述网络故障的特征,因此不可避免地带来了冗余和错误诊断信息。过多的诊断信息会增加规则匹配的复杂度,为了提高规则匹配的效率,本文提出基于邻域粗糙集的规则自动生成方法来约简与网络故障相关的诊断属性,然后自动生成规则。

本文利用邻域粗糙集对网络故障进行分析,给出一种规则生成算法。假设有 $n$ 种不同的故障诊断属性,在决策信息系统中这些属性构成条件属性集合,表示为 $A=\{C_1,C_2,\cdots,C_n\}$,故障作为决策信息系统属性的分类,表示为 $D=\{V_1,V_2,\cdots,V_n\}$。一个网络故障的数据样本表示网络故障在不同时刻的条件属性取值,网络故障的数据样本集合用 $U$ 表示。利用邻域粗糙集,直接处理连续数据就能得到属性的约简结果。

规则生成的主要步骤为:(1)计算邻域半径 $\delta$,基于邻域半径获得所有条件属性的邻域集合;(2)计算全体条件属性相对于决策属性的正域,根据正域计算全体条件属性的依赖度;(3)遍历计算每个条件属性的依赖度和重要度,若条件属性的重要度大于设定的重要度下限 $Sig\_ctrl$,将条件属性加入到约简结果 $red$ 中;(4)重复步骤(2)、步骤(3),直到条件属性的重要度小于设定的重要度下限;(5)对约简结果中的每一个属性限定阈值;(6)由约简结果及其阈值限定生成规则。

对于约简获得的规则诊断属性,根据属性值的类型,基于约简结果对应的数据样本来限定诊断属性的阈值。对于连续型数据,通过取最大值及最小值来限定诊断属性的阈值。对于离散型数据,根据不同的离散值出现的概率,获取最大概率的离散值。

**算法 1**　基于邻域粗糙集的规则自动生成算法

**输入**　邻域决策系统 $NDS=(U,A,D)$,邻域半径 $\delta$,重要度下限 $Sig\_ctrl$

**输出**　属性约简集合 $red$,规则集合 $Rules$

(1)$\forall a\in A$,计算邻域关系 $N_a$。

(2)初始化约简集合 $red=\varnothing$、数据样本 $smp=U$。

(3)对每个条件属性 $a_i\in A-red$,计算重要度 $Sig(a_i,red,D)=\gamma_{red\cup a}(D)-\gamma_{red}(D)$,$Sig(a_k,red,D)=\max(Sig(a_i,red,D))$。

(4)若 $Sig(a_k,red,D)>Sig\_ctrl$,$red=red\cup a_k$,则执行步骤(3);否则返回 $red$。

(5)限定 $red$ 中每个条件属性的阈值。

(6)若 $red\neq\varnothing$,则由 $red$ 生成规则集合 $Rules$。

## 4　基于值权重的规则匹配算法

在进行规则匹配时会出现一个诊断数据匹配多个规则的情况,其原因主要是由于生成规则的论域基数有限,不能准确地限定诊断属性的阈值范围,不同的规则之间会有一定程度的阈值交叉。为在不增加论域基数的前提下,提高规则匹配的准确率,本文提出一种基于值权重的规则匹配算法。该算法通过分析属性约简后得到的数据样本,计算规则诊断属

性的不同属性值相对于该属性的所有属性值的权重,通过比较规则相对于监控数据的值权重大小,从匹配的多个规则中选择一条匹配度最高的规则。

**定义 6**(值权重) 在生成的规则集 $rules=\{R_1,R_2,\cdots,R_n\}$ 中,每个规则都包含诊断属性集 $attribute=\{A_1,A_2,\cdots,A_n\}$。若在约简后的数据集中有与规则 $R_i$ 的属性 $A_j$ 相关的属性值 $value=\{V_1,V_2,\cdots,V_n\}$,则属性 $A_j$ 的属性值等于 $V_k$ 的值权重,表示为 $weight_{V_k}=\dfrac{count}{N_{R_i}}$,$count$ 为属性 $A_j$ 中属性值等于 $V_k$ 的数量,$N_{R_i}$为与规则 $R_i$ 有关的数据样本数量。

如表 1 所示的规则 $R_1$ 约简数据集中,诊断属性 $A_1$ 中属性值为 1 的数据样本分别为样本编号为 1 ~ 4 的数据,$count=4$。规则 $R_1$ 的数据样本数为 8,$N_{R_1}=8$。综上所述,属性 $A_1$ 的属性值等于 1 的值权重为 $weight_1=\dfrac{4}{8}=0.5$。

**表 1 规则 $R_1$ 的约简数据集**

| 编号 | $A_1$ | $A_2$ | $A_3$ |
|---|---|---|---|
| 1 | 1 | 1 | 3 |
| 2 | 1 | 1 | 2 |
| 3 | 1 | 1 | 1 |
| 4 | 1 | 1 | 3 |
| 5 | 2 | 0 | 0 |
| 6 | 2 | 0 | 0 |
| 7 | 3 | 0 | 0 |
| 8 | 3 | 0 | 0 |

**定义 7**(规则匹配度) 若监控数据 $data=\{D_1,D_2,\cdots,D_n\}$ 对于规则 $R$ 的属性集有 $\min|D_i-V|$,则监控数据的规则匹配度 $degree$ 表示为:$degree=\sum_{i=1}^{n}weight_i$,$weight_i$ 表示与 $D_i$ 差的绝对值最小的属性值 $V$ 的值权重。

假设输入的监控数据为($A_1$,$A_2$,$A_3$)=(1,1,0.9),得到 $\min|D_i-V|$诊断属性中的属性值分别为(1,1,1)。$A_1$ 的属性值中等于 1 的值权重为 0.5,$A_2$ 的属性值中等于 1 的值权重为 0.5,$A_3$ 的属性值中等于 1 的值权重为 0.125。监控数据对于 $R_1$ 规则的匹配度为 $degree=1.125$。

基于定义 6,对规则的每个诊断属性生成一个值权重集合 $weightSet$。规则引擎对监控数据进行规则匹配,当监控数据发生多匹配的情况时,将规则添加到监控数据对应的多匹配规则列表 $multiMatchRules$ 中。基于定义 7,计算监控数据对于多匹配规则列表中规则的匹配度,通过比较匹配度来确定唯一的匹配规则。

**算法 2** 基于值权重的规则匹配算法

**输入** 值权重集合 $weightSet$,多匹配数据 $datas$

**输出** 匹配规则集 $match\_rules$

(1)对每个多匹配数据 $data\in datas$:设置初始的规则匹配度 $degree=0$,获得对应的多匹配规则 $multiMatchRules$。

(2)对于多匹配规则 $rule\in multiMatchRules$: for $D\in data$ from 1 to $N$,计算值权重 $weight_{D_i}=weightSet(rule,min|D_i,V|)$,$data$ 对规则的匹配度为 $degree=\sum_{i=1}^{N}weight_{D_i}$。

(3)比较每个规则的 $degree$,获得最大匹配度对应的规则,返回 $match\_rule=rule$。

(4)得到 $match\_rules=match\_rules\cup match\_rule$。

## 5 实验与结果分析

为验证本文方法能够在保证故障诊断率的情况下提升规则匹配效率,通过对比实验进行验证。本节分别从规则的故障诊断率和规则的匹配时间来比较基于粗糙集及基于邻域粗糙集自动生成的规则与手工生成的规则在性能上的差异,并从匹配时间增长率方面分析论域基数对规则匹配时间的影响。

KDD 99 数据集被广泛应用于网络故障管理和入侵检测等研究领域,并取得了良好的效果[14-15]。本文使用 KDD 99 数据集作为测试数据,首先从数据集中随机生成决策类型为 $ipsweep$,$portsweep$,$satan$,$smurf$ 的数据来构造邻域决策系统作为算法 1 的输入,通过算法 1 的属性约简得到约简的数据集并生成规则。同时,针对规则匹配时产生的多匹配数据,利用算法 2 选择一个匹配度最高的规则。比较数据集由未约简的测试数据组成,其作为规则生成的输入数据解析获取规则。从数据集中随机生成数据当作比照数据,分析规则的故障诊断率和规则匹配时间。规则匹配时间为规则引擎启动到测试数据匹配完成所花费的时间。故障诊断率的计算公式为:

$$故障诊断率=\frac{正确诊断出的故障数据数量}{总的故障数据数量}$$

匹配时间增长率为当前论域基数的匹配时间与前一个论域基数的匹配时间的增长率,计算公式如下:

$$匹配时间增长率=\frac{当前匹配时间-前一个匹配时间}{前一个匹配时间}$$

实验在 Intel(R) Core(TM) i5-2400 CPU @

3.10 GHz处理器上进行,从 KDD 99 数据集中随机选取了 1 000 条正常数据和 4 000 条故障数据作为比照数据集,并分别随机生成论域基数为 100,200,300,400,500 的测试数据作为算法 1 的输入。从图 1可以看出,基于邻域自动生成规则的故障诊断率随着论域基数的增加而增加,并逐渐逼近手工定义的规则,而基于粗糙集的自动生成规则的故障诊断率都低于基于邻域粗糙集的自动生成方法。由此说明,当论域基数足够大,基于本文提出的方法自动生成的规则等价于手工定义规则,且基于邻域粗糙集的自动生成规则相对于传统的粗糙集方法的诊断率平均提升了 15% 。

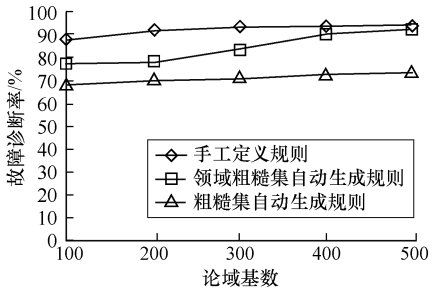


**图 1　论域基数和规则故障诊断率的关系**

从图 2 可以看出,在不同的论域基数下,自动生成规则由于经过了属性约简,规则匹配时间均在 350 ms以下,基于邻域粗糙集的生成规则匹配时间也优于传统粗糙集。而手工定义规则由于包含大量的冗余属性,匹配时间均超过了 700 ms,且随着论域基数的增加而增加,通过计算得出规则匹配效率平均提升了 2.5 倍,同时计算当前论域基数相对于前一个论域基数的匹配时间增长率。

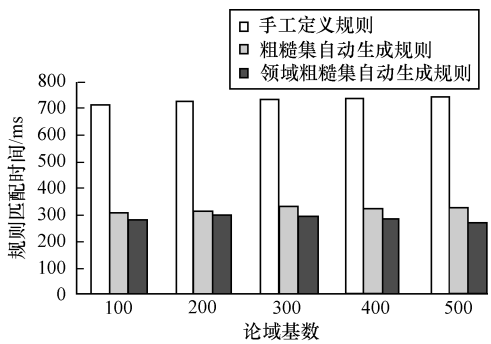


**图 2　论域基数和规则匹配时间的关系**

从图 3 可以看出,基于邻域粗糙集自动生成规则的匹配时间增长率随着论域基数的增加而减少,并且基数大于 200 时得到的增长率均小于 0,即匹配时间随着基数的增大而减少,说明论域基数越大,规则诊断信息越准确,匹配效率越高,并且稳定性也优于基于传统粗糙集的规则自动生成方法。

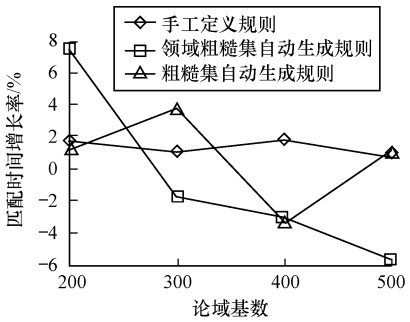


**图 3　论域基数和匹配时间增长率的关系**

## 6　结束语

本文提出一种基于邻域粗糙集的规则自动生成方法,解决了目前基于规则的网络故障管理中规则需要手动定义的问题。该方法通过约简故障诊断信息来消除冗余数据对故障诊断的影响,降低了规则匹配的复杂度。针对规则的多匹配问题,提出基于值权重的规则匹配算法,保证了自动生成规则具有较高的故障诊断率。实验结果表明,本文提出的方法具备较好的规则自动生成能力,生成的规则在保证故障诊断率的情况下,有效降低了规则匹配的处理时间。但由于实验环境和实际环境存在差异,实验结果不能完全代表实际环境中对故障的诊断能力,因此下一步研究主要是在完善自动生成规则诊断能力的同时,将提出的具体方法和理论应用到网络故障管理系统中验证其对实际网络故障的处理能力。

### 参考文献

[1] Min Weidong, Liu Yonghui, Ke Yongzhen, et al. Using Particle Swarm Optimization Algorithm to Improve Multi-agents Network Management [J]. Journal of Computational Information Systems, 2014, 10(2): 739-746.

[2] Liu Yonghui, Min Weidong. Network Management Based on Domain Partition for Mobile Agents [M]// Yin Hujun, Tang Ke, Yang Gao, et al. Intelligent Data Engineering and Automated Learning. Berlin, Germany: Springer, 2013: 153-160.

[3] Min Weidong, Chen Ke, Ke Yongzhen. A Matrix Grammar Approach for Automatic Distributed Network Resource Management [J]. Frontiers of Computer Science, 2013, 7(4): 583-594.

[4] Min Weidong. Distributed Network Resources Monitoring Based on Multi-agent and Matrix Grammar [C]//Proceedings of the 4th International Symposium on Parallel Architectures, Algorithms and Programming. Washington D. C., USA: IEEE Computer Society, 2011: 136-140.

[5] Chen Xu, Mao Yun. DECOR: Declarative Network Management and Operation [J]. ACM SIGCOMM Computer Communication Review, 2009, 40(1): 61-66.
[6] de Paola A, Fiduccia S, Gaglio S, et al. Rule Based Reasoning for Network Management[C]//Proceedings of Computer Architecture for Machine Perception. Washington D. C., USA: IEEE Computer Society, 2005: 25-30.
[7] Suarez-Tangil G, Palomar E, Fuentes J M D, et al. Automatic Rule Generation Based on Genetic Programming for Event Correlation[M]//Alvaro H, Paolo G, Rodolfo Z. Computational Intelligence in Security for Information Systems. Berlin, Germany: Springer, 2009: 127-134.
[8] Liu Jun, Martinez L, Calzada A, et al. A Novel Belief Rule Base Representation, Generation and Its Inference Methodology [J]. Knowledge-Based Systems, 2013, 53(9): 129-141.
[9] Shim K S, Yoon S H, Lee S K, et al. Automatic Generation of Snort Content Rule for Network Traffic Analysis[J]. Journal of Korean Institute of Communications & Information Sciences, 2015, 40(4): 666-677.
[10] Jiang Feng, Sui Yuefei. A Novel Approach for Discretization of Continuous Attributes in Rough Set Theory[J]. Knowledge-Based Systems, 2014, 73(1): 324-334.
[11] Peng Yuqing, Liu Gengqian, Geng Hengshan. Application of Rough Set Theory in Network Fault Diagnosis[C]//Proceedings of the 3rd International Conference on Information Technology and Applications. Sydney, Australia: IEEE Computer Society, 2005: 556-559.
[12] Qu Zhiming, Wang Xiaoli. Study of Rough Set and Clustering Algorithm in Network Security Management[C]//Proceedings of International Conference on Networks Security, Wireless Communications and Trusted Computing. Washington D. C., USA: IEEE Computer Society, 2009: 326-329.
[13] 胡清华,于达仁,谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. 软件学报, 2008, 19(3): 640-649.
[14] Meng You, Yu Lang, Luan Zhongzhi, et al. A Black-box Approach for Detecting the Failure Traces[J]. Communications in Computer & Information Science, 2014, 426: 252-259.
[15] 牟 琦,龚尚福,毕孝儒,等. 基于快速属性约简的网络入侵特征选择[J]. 计算机工程, 2011, 37(17): 113-115.

编辑 陆燕菲

(上接第 309 页)

[4] 李 雪,聂兰顺,齐文艳,等. 基于近似动态规划的动态车辆调度算法[J]. 中国机械工程, 2015, 26(5): 682-693.
[5] 陈 森,姜 江,陈英武,等. 一类非确定性车辆路径问题模型及其算法设计[J]. 计算机工程, 2011, 37(14): 186-188.
[6] Powell W B, Jaillet P, Odoni A. Chapter 3 Stochastic and Dynamic Networks and Routing[Z]. Handbooks in Operations Research & Management Science, 1995.
[7] 崔 丽,王笑丛. 需求驱动下的城市配送车辆动态调度研究[J]. 计算机工程与应用, 2015, 51(2): 241-244.
[8] Bent R, van Hentenryck P. A Two-stage Hybrid Local Search for the Vehicle Routing Problem with Time Windows[J]. Transportation Science, 2004, 38 (4): 515-530.
[9] 王晓博,李一军. 电子商务下基于改进两阶段算法的有时间窗车辆调度优化[J]. 中国管理科学, 2007, 15(6): 52-59.
[10] 柴宏建,高尚策. 基于聚类混合遗传算法的 LRP 问题研究[J]. 电子设计工程, 2015, 23(9): 1-4.
[11] 马小璐,李和成. 带容量约束车辆路径问题的一个新遗传算法[J]. 应用数学进展, 2014, 3(4): 222-230.
[12] 孙 琦,王 东. 具有粒子群特征的优化并行蚁群算法[J]. 计算机工程, 2008, 34(24): 208-210.
[13] 丰 伟,李雪芹. 基于粒子群算法的多目标车辆调度模型求解[J]. 系统工程, 2007, 25(4): 15-19.
[14] 田明才,于 东,吴 琼. 多目标遗传算法在 GPS 动态车辆调度中的应用研究[J]. 小型微型计算机系统, 2010, 31(3): 545-548.
[15] 金 叶,丁以中. 考虑总量和体积双重约束的时间窗车辆路径问题研究[J]. 物流科技, 2009, 32(4): 53-56.

编辑 索书志