一种基于粗糙集的网络安全评估模型*)

陈志杰　王永杰　鲜　明
(国防科学技术大学电子科学与工程学院　长沙 410073)

摘　要　准确掌握计算机网络系统的安全水平对于保障网络系统的正常运行具有重要意义。当前大多数网络安全评估系统缺乏对数据的深入分析，难以形成对网络安全状况的整体认识。本文提出了一种利用粗糙集理论挖掘网络安全评估规则，进而利用评估规则构建网络安全评估决策系统的算法模型。研究了网络安全评估问题的粗糙集描述，给出了模糊属性决策表的约简方法。利用一个简化的网络安全评估数据集，验证了本文提出的决策规则提取方法，结果表明该方法可以得到与实际情况相符的决策规则。

关键词　网络安全，评估模型，决策规则，模糊粗糙集

An Evaluation Model of Computer Network Security Based on Rough Set

CHEN Zhi-Jie　WANG Yong-Jie　XIAN Ming
(School of Electronic Science and Engineering, NUDT, Changsha 410073)

Abstract　It is very important to know the security status of computer networks accurately. At present, most computer network security evaluation system don't analyze the datum thoroughly. Therefore it is difficult to acquire the security status of computer networks at the whole. An algorithm model with rough set theory to mine the rules of computer network security evaluation is proposed. A rough set knowledge system description of computer network security evaluation is studied. A fuzzy rough set attribute reduction method is given. The decision rules mining method presented in this paper is validated with a simplified network security evaluation data set. The experiment results show that decision rules acquired by the method are in accord with the fact.

Keywords　Network security, Evaluation model, Decision rules, Fuzzy rough set

1　引言

随着计算机和网络通信技术的发展，人类社会对计算机网络的依赖程度越来越高。与此同时，计算机网络的安全问题也变得日益突出，许多网络系统都由于安全问题而蒙受了大量的经济损失。对计算机网络系统的安全性进行先期评估是防范各种网络安全问题的一种有效手段。近些年来，市场上出现了一些计算机网络安全评估产品，对于加强网络系统的安全性起到了一定的作用，但是这些网络安全评估产品大都局限于对网络系统存在安全漏洞的探测与分析方面，缺乏对网络安全评估数据的深层分析，难以形成对计算机网络系统安全性的整体认识。

粗糙集理论是解决具有不确定、不完整信息的系统决策问题的有效工具，因此在本文中我们将可以应用该理论来构建计算机网络安全决策推理系统。基本思想是将网络安全评估指标集作为决策系统 $S=(U,C\cup D,V,f)$ 的条件属性集 C，将评估结果作为决策属性集 D，通过对其决策表的约简产生决策规则集，然后以决策规则集为基础建立网络安全评估的决策支持系统。

2　粗糙集理论的基本概念

粗集理论作为一种刻画不完整性和不确定性的数学工具，能有效地分析不精确、不一致、不完整等各种不完备的信息，还可以对数据进行分析和推理，从中发现隐含的知识、揭示潜在的规律。该理论的主要特点是：它无需提供问题所需处理的数据集合之外的任何先验信息，这是其与证据理论和模糊集合理论的最主要区别，也是最重要的优点。目前粗糙集理论已经在模式识别、决策支持与分析及智能控制等许多领域得到了广泛的应用[1]。

2.1　知识的含义与表示方法

知识是智能决策中一个非常重要的概念，所有的决策都依赖于知识。知识在不同的范畴中有不同的含义。在粗糙集理论中，知识被看作是关于论域的划分，是一种对对象进行分类的能力，因而提出知识是有粒度的，知识的不精确性是由于组成论域知识的颗粒而引起的。

定义 1　设 $U\neq\emptyset$ 是给定研究对象的有限集合，称为论域。$\forall X\subseteq U$，称为 U 中的一个概念(Concept)或范畴(Category)。U 中的一个概念族 $F=\{X_1,X_2,\cdots,X_n\}$ 称为关于 U 的一个知识，其中 $X_i\subseteq U, X_i\neq\emptyset, X_i\cap X_j=\emptyset, i\neq j, i,j=1,2,\cdots,n$，且有 $\bigcup X_i=U$。为规范起见，$\emptyset$ 也认为是一个概念。

定义 2　设 R 是 U 上的一个等价关系(Equivalence Relation)，$U/R=\{X_1,X_2,\cdots,X_n\}$ 表示 R 产生的分类，称为关于 U 的一个知识。$[x]_R=\{y\in U\mid xRy\}$ 表示关系 R 下包含元素 x 的等价类。(U,R) 称为近似空间(Approximation Space)。

等价关系也称为不可分辨关系(Indiscernibility Relation)。设 W 是 U 上的等价关系族，则知识库可表示为 $K=$

*)基金项目：国家自然科学基金项目(60372039)。陈志杰　副教授，主要研究方向为网络与信息系统；王永杰　博士研究生，主要研究方向为信息安全；鲜　明　副教授，博士。

(U,W)。

2.2 粗糙集

定义 3 令 $X\subseteq U$，R 是 U 上的一个等价关系。当 X 为 R 的某些等价类的并时，称 X 是 R 可定义的(R-definable)，否则称 X 为 R 不可定义的(R-undefinable)。R 可定义集称作为 R 精确集，R 不可定义集称为 R 粗糙集。粗糙集可以用两个精确集，即粗糙集的下近似和上近似来描述。

定义 4 包含在 X 中的最大可定义集称为 X 的 R 下近似(Lower Approximation)：

$$\underline{R}(X)=\{x\in U|[x]_R\subseteq X\} \tag{1}$$

包含 X 的最小可定义集称为 X 的 R 上近似(Upper Approximation)：

$$\overline{R}(X)=\{x\in U|[x]_R\cap X\neq\emptyset\} \tag{2}$$

$\underline{R}(X)$表示在知识 R 下 U 中所有一定能归入 X 的元素的集合，$\overline{R}(X)$表示在知识 R 下 U 中可能归入 X 的元素的集合。

定义 5 $\alpha_R(X)=|\underline{R}(X)|/|\overline{R}(X)|$称为关于 R 的近似精度，其中$|\cdot|$表示集合中元素的数目，称为集合的基数或势(Cardinality)。

显然 $0\leqslant\alpha_R(X)\leqslant 1$，当 $\alpha_R(X)=1$ 时，表示不存在边界域，则称集合 X 相对于 R 是清晰的；当 $\alpha_R(X)<1$ 时，表示存在边界域，则称集合 X 相对于 R 是粗糙的。$\alpha_R(X)$认为是在等效关系 R 下逼近集合 X 的精度。

定义 6 $\rho_R(X)=1-\alpha_R(X))$称为 X 的 R 粗糙度。如果 $\rho_R(X)=0$，则集合 X 关于 R 是普遍集合；如果 $\rho_R(X)>0$，则集合 X 关于 R 是粗糙的。

定义 7 信息系统论域中元素 x 对集合 X 的粗糙隶属函数定义为：

$$\mu_X^R(x)=\frac{|X\cap[x]_R|}{|[x]_R|},\mu_X^R(x)\in[0,1] \tag{3}$$

粗糙隶属函数体现了元素与集合间隶属关系的不确定性。利用粗糙隶属函数，可以定义集合 X 的上、下近似和边界：

$$\begin{aligned}&\underline{R}(X)=\{x\in U|\mu_X^R(x)=1\}\\&\overline{R}(X)=\{x\in U|\mu_X^R(x)>0\}\\&B_n(X)=\{x\in U|0<\mu_X^R(x)<1\}\end{aligned} \tag{4}$$

2.3 属性的约简与核

定义 8 设有决策系统 $S=(U,C\cup D,V,f)$，其中 C,D 分别表示条件属性和决策属性，则决策属性在条件属性下的正域可定义为：

$$POS_C(D)=\bigcup_{X\in U/D}\underline{C}(X) \tag{5}$$

$POS_C(D)$表明根据 C 的知识所进行的划分 U/C，能够确切地划入 U/D 类的对象集合。

定义 9 决策属性 D 对条件属性 C 的依赖度定义为：

$$k=\gamma_C(D)=\frac{|POS_C(D)|}{|U|} \tag{6}$$

3 基于粗糙集理论的网络安全评估模型

3.1 网络安全评估指标数据预处理

由于粗糙集是基于符号运算的离散知识推理系统，要求决策表中的值必须是离散数据(如整型、字符串型、枚举型)等。而网络安全评估指标数据通常为连续量或定性值，为此需要对网络安全评估的样本数据进行预处理。常用的离散化预处理方法有等距离划分法、等频率划分法、Naive Scaler 算法、Semi Naive Scaler 算法、Boolean 逻辑与 Rough 集理论相结合的离散化算法及基于属性重要性的离散化算法[3]等。结合网络安全评估指标数据的特点，在本文中我们采取如下方法进行样本数据离散化预处理：

(1)对于计算机网络安全的评估结果直接映射到一个评语集合 $V=\{v_1,v_2,\cdots,v_9\}$，并且以{1,2,3,4,5,6,7,8,9}来表示，其中 v_1：极差，v_2：很差，v_3：差，v_4：较差，v_5：一般，v_6：较好，v_7：好，v_8：很好，v_9：极好。

(2) 对于网络安全评估指标数据，主要考虑归一化后的情况，大部分评估指标可以看作是在[0,1]区间上连续变化。为了反映出不同评估指标值对安全的影响，本文中利用粗糙-模糊集集成的方法[5]，首先通过将连续属性值转化成模糊值，然后再进行粗糙集的决策分析。

3.2 网络安全评估的粗糙集描述

根据上面的分析，计算机网络安全评估可以看作一个属性值连续的决策系统 $S=\langle U,C\cup\{d\},V,f\rangle$，其中 $U=\{x_1,x_2,\cdots,x_n\}$是非空有限集合，表示网络安全评估样本集；$C=\{a_1,a_2,\cdots,a_m\}$是非空有限条件属性集，即网络安全评估指标集；d 表示结果属性，即网络安全的评估结果；$V=V_C\cup V_d$ 是属性值集合，$V_C=\{V_a|a\in C\}$是条件属性值集，V_d 是决策属性值集，第 i 个对象在第 j 个条件属性上的取值 v_{ij} $(i=1,2,\cdots,n;j=1,2,\cdots,m)$是连续变化的；$f:U\times C\cup\{d\}\to V$ 是一个信息函数，表示对 $\forall a\in C,x\in U$ 有 $f(x,a)\in V_a$。

对于每一个条件属性 a_j $(j=1,2,\cdots,m)$，确定其隶属函数和模糊区间数，设 E_j^k 为条件属性 a_j 的第 k 个模糊区域，h_j 为第 j 个条件属性的模糊区间数。r_{ij}^k 表示对象 x_i $(i=1,2,\cdots,n)$在模糊区域 E_j^k 中的隶属函数值，那么任意一个条件属性值 v_{ij} $(i=1,2,\cdots,n;j=1,2,\cdots,m)$均可以表示成：

$$\frac{r_{ij}^1}{E_j^1}+\frac{r_{ij}^2}{E_j^2}+\cdots+\frac{r_{ij}^{h_j}}{E_j^{h_j}} \tag{7}$$

据此，我们就可以将连续属性决策表转化为模糊属性决策表[4]。

3.3 模糊属性决策表约简

为了对模糊属性决策表进行约简分析，首先定义一个模糊关系 R[6]。

定义 10 $\forall x_s,x_t\in U,\forall a_j\in C(j=1,2,\cdots,m)$，定义模糊关系 R 为：

$$x_sRx_t=\{(x_s,x_t)\in U\times U|r_{sj}^k,r_{tj}^k>0\} \tag{8}$$

定义 11 $\forall x_s,x_t\in U,\forall a_j\in C(j=1,2,\cdots,m)$在模糊关系 R 下 x_s,x_t 之间的相关度为：

$$\mu_R(x_s,x_t)=1-\frac{1}{m}\sum_{j=1}^{m}|r_{sj}^k-r_{tj}^k| \tag{9}$$

可以证明模糊关系 R 满足自反性和对称性，但不满足传递性，因此 R 是模糊相容关系。

定义 12 所有与 $x_i\in U$ 模糊相容的对象集称为 x_i 的模糊相容类，用 $FSIM(x_i)$表示。即：

$$FSIM(x_i)=\{x_s\in U|r_{ij}^k,r_{sj}^k>0;j=1,2,\cdots,m\} \tag{10}$$

定义 13 对于 $X\subseteq U,P\subseteq C$，近似空间$(X,P)$在模糊相容关系下的下、上近似为：

$$\underline{P}(X)=\{x_i|x_i\in X,FSIM(x_i)\subseteq X,1\leqslant i\leqslant n\} \tag{11}$$

$$\overline{P}(X)=\{x_i|x_i\in X,FSIM(x_i)\cap X\neq\phi,1\leqslant i\leqslant n\} \tag{12}$$

属性约简为不含多余属性并保证分类正确的最小条件属性集，分类的正确性用一个决策属性对条件属性的依赖度来表示[2]，条件属性依赖度的定义如式(6)所示。

假设 $C'\subseteq C$ 是一个约简，可以通过依赖度来构建属性约

简的算法。其算法是以所有的条件属性作为初始约简集合，在保证依赖度不变的条件下，逐步缩减条件属性从而获取属性的约简，其具体步骤为：

(1)初始化：令 $C'=C$，计算 $\gamma(C',d)$；

(2)$j=1$；

(3)对于 $a_j\in C'$，计算 $\gamma(C'-\{a_j\},d)$；

(4)若 $\gamma(C'-\{a_j\},d)=\gamma(C',d)$，则 $C'=C'-\{a_j\}$；

(5)若 $j<m$，则 $j=j+1$，转(3)；

(6)输出属性约简 C'。

4 基于粗糙集理论的网络安全评估实例

为了说明方便，我们用一个简化的数据集来验证本节所提出的基于粗糙集理论的网络安全评估模型。

4.1 评估数据集样本

假设数据集样本包含 3 个条件属性(评估指标)、一个决策属性(评估结果)，即决策表的条件属性集为：a_1 安全漏洞，a_2 补丁情况，a_3 操作系统类型，相应的取值为归一化后的结果，对于越小越好的负向性指标要进行正向性归一化。决策属性 d 为相应的安全评估结果。假设数据集样本如表 1 所示。

表 1 网络安全评估信息表

	a_1	a_2	a_3	d
x_1	0.9318	0.5028	0.8600	0.7608
x_2	0.4660	0.7095	0.8537	0.6067
x_3	0.4186	0.4289	0.5936	0.4445
x_4	0.8462	0.3046	0.4966	0.5978
x_5	0.5252	0.1897	0.8998	0.4456
x_6	0.2026	0.1934	0.8216	0.2770
x_7	0.6721	0.6822	0.6449	0.6725
x_8	0.8381	0.3028	0.8180	0.6335
x_9	0.0196	0.5417	0.6602	0.1463
x_{10}	0.6813	0.1509	0.3420	0.4384
x_{11}	0.3795	0.6979	0.2897	0.4884
x_{12}	0.8318	0.3784	0.3412	0.5989

4.2 决策表约简与评估规则提取

将决策属性 d 等间隔地划分为 A,B,C,D 四个依次减弱的等级，所有条件属性按照图 1 给出的模糊隶属度函数进行离散化，可以得到如表 2 所示的模糊决策表。

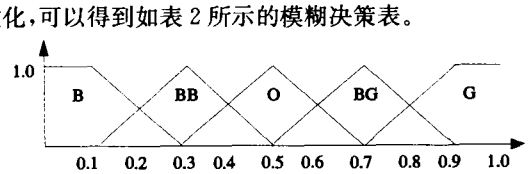

图 1 条件属性的模糊隶属函数

其中 B,BB,O,BG,G 分别表示差，较差，一般，较好，好等五个等级。

利用式(10)和(11)定义的模糊相容类，可以得到如下的决策规则：

(1)当$(a_1,a_2,a_3)=$(G,O,G)或(G,O,BG)时，$d=$A；

(2)当$(a_1,a_2,a_3)=$(O,BG,G)或(O,BG,BG)或(BB,BG,G)或(BB,BG,BG)(G,BB,O)或(BG,BB,O)或(BG,BG,BG)或(BG,BG,O)或(BG,O,BG)或(BG,O,O)或(O,BG,O)或(G,BB,G)或(G,BB,BG)或(BG,BB,G)或(BG,BB,BG)或(G,O,O)或(G,O,BB)或(G,BB,BB)或(BG,O,O)或(BG,O,BB)时，$d=$B；

(3)当$(a_1,a_2,a_3)=$(O,O,BG)或(O,O,O)或(BG,BB,BB)时，$d=$B 或 C；

表 2 网络安全模糊决策表

	a_1	a_2	a_3	d
x_1	1/G	1/O	0.8/G+0.2/BG	A
x_2	0.85/O+0.15/BB	1/BG	0.75/G+0.25/BG	B
x_3	0.6/O+0.4/BB	0.65/O+0.35/BB	0.45/BG+0.55/O	C
x_4	0.75/G+0.25/BG	1/BB	1/O	B
x_5	0.15/BG+0.85/O	0.45/BB+0.55/B	1/G	C
x_6	0.5/BB+0.5/B	0.45/BB+0.55/B	0.6/G+0.4/BG	C
x_7	0.85/BG+0.15/O	0.9/BG+0.1/O	0.7/BG+0.3/O	B
x_8	0.7/G+0.3/BG	1/BB	0.6/G+0.4/BG	B
x_9	1/B	0.7/BB+0.3/B	0.3/O+0.7/BB	D
x_{10}	0.9/BG+0.1/O	0.25/BB+0.75/B	0.2/O+0.8/BB	C
x_{11}	0.4/O+0.6/BB	1/BG	0.95/BB+0.05/B	C
x_{12}	0.65/G+0.35/BG	0.4/O+0.6/BB	0.2/O+0.8/BB	B

(4)当$(a_1,a_2,a_3)=$(O,BB,O)或(O,BB,BG)或(BB,O,BG)或(BB,O,O)或(BB,BB,O)或(BB,BB,BG)或(BG,BB,G)或(BG,B,G)或(O,BB,G)或(O,B,G)或(O,BG,BB)或(O,BG,B)或(BG,BG,BB)或(BG,BG,B)或(BB,BB,G)或(BB,B,G)或(BB,B,BG)或(B,BB,G)或(B,BB,BG)或(B,B,G)或(B,B,BG)或(BG,BB,O)或(BG,B,O)或(BG,B,BB)或(O,BB,BB)或(O,B,O)或(O,B,BB)时，$d=$C；

(5)当$(a_1,a_2,a_3)=$(B,BB,O)或(B,BB,BB)或(B,B,O)或(B,B,BB)时，$d=$D。

5 结果分析

由于使用的网络安全样本数量较少，得到的决策规则还不能全面反映计算机网络安全的决策过程，但基本体现了评估指标取值越大，安全性越好的规律。如果有足够多的网络攻击评估样本，就可以得到详细的决策规则，然后以此为基础构建计算机网络安全评估的决策支持系统。

参 考 文 献

1 熊丽君，许龙飞. Rough set 理论及其应用研究进展综述[J]. 暨南大学学报(自然科学版)，2003，24 (3)：70～75

2 Jensen R，Qiang Shen. Fuzzy-rough sets for descriptive dimensionality reduction[A]. Fuzz-IEEE'02[C]. In：Proceedings of the 2002 IEEE International Conference on Fuzzy Systems，2002

3 王国胤. Rough 集理论与知识获取[M]. 西安：西安交通大学出版社，2001

4 何亚群. 基于粗糙集的智能决策理论与应用研究[D]：[南京航空航天大学博士学位论文]. 2004

5 Anna M R，Etienne E K. A comparative study of fuzzy rough sets [J]. Fuzzy Sets and System，2002 (126)：137～155

6 Kankana C，Ranjit B，Sudarsan N. Fuzziness in rough sets[J]. Fuzzy Sets and System，2000 (110)：247～251