DOI:10.11992/tis.201509020
网络出版地址:http://www.cnki.net/kcms/detail/23.1538.TP.20160824.0928.006.html

SCADA 安全因素神经元的云推理机研究与仿真

熊柳[1],曹谢东[1],李杰[1],杨力[2],刘增良[3]
(1. 西南石油大学 电气信息学院,四川 成都 610500; 2. 西南石油大学 计算机科学学院,四川 成都 610500; 3. 中国人民解放军国防大学 信息作战研究所,北京 100091)

摘　要:为了解决 SCADA 系统信息安全的问题,本文提出了一种基于因素神经网络的主动防御方法。将 SCADA 系统信息安全的影响因素映射到因素空间坐标中,然后利用知识因素的因素神经元表示方法,通过云模型推理机实现了语言值表示的模糊概念到定量数据的转换,并通过云模型多规则多条件发生器进行规则推理,最后根据得到的期望值又可以转换为定性语言值,这样就实现了对未知恶意程序行为操作可能性的预测。本文着重于利用基于云模型的多条件多规则发生器实现推理,通过 MATLAB 进行算法设计和仿真,为油气 SCADA 系统信息安全防御的解决方法提供了一种思路。

关键词:SCADA 信息安全;因素空间;因素神经元;云模型;MATLAB 仿真
中图分类号:TP18　　**文献标志码**:A　　**文章编号**:1673-4785(2016)05-0688-08

中文引用格式:熊柳,曹谢东,李杰,等. SCADA 安全因素神经元的云推理机研究与仿真[J]. 智能系统学报, 2016, 11(5): 688-695.
英文引用格式:XIONG Liu, CAO Xiedong, LI Jie, et al. Study and simulation of the SCADA security factors neuron's cloud inference engine[J]. CAAI transactions on intelligent systems, 2016,11(5):688-695.

Study and simulation of the SCADA security factors neuron's cloud inference engine

XIONG Liu[1], CAO Xiedong[1], LI Jie[1], YANG Li[2], LIU Zengliang[3]
(1. School of Electrical Information Engineering, Southwest Petroleum University, Chengdu 610500, China; 2. School of Computer Science, Southwest Petroleum University, Chengdu 610500, China; 3. Institute of Information Operation, University of National Defense, Beijing 100091, China)

Abstract: Over recent years, with the amount of incidents involving industrial control information systems, the safety of these system has been given increased importance across the Globe, and several technical measures have been implented to improve this. This paper proposed an active defense method based on a factor neural network in reference to SCADA information security. The different aspects of SCADA information security were mapped to factor coordinates, and the factor neuron method for knowledge factors was then used to transform this from a fuzzy concept (represented by language) to quantitative data through a cloud inference engine. Inference was then conducted through generated multi conditions and multi rules based on a cloud model, so that it could then be transformed into a qualitative language (e.g. 'more likely' based on Ex) to be able to forecast the consequences of unknown malicious programs. This paper focus on using a generator with multiple conditions and rules based on a cloud model to achieve inference, thus providing an idea of the oil and gas SCADA information security's defense response using algorithmic design and MATLAB-based simulations.

Keywords: SCADA information security; factor space; factor neuron; cloud model; MATLAB simulation

SCADA(supervisory control and data acquistion,数据采集与监视控制系统)工业控制系统广泛运用于工业、能源、交通、水利以及市政等领域,主要用于生产数据采集和控制生产设备的运行。一旦工业控制系统信息安全出现漏洞,将对工业生产运行和国家经济安全造成重大隐患。一方面传统的病毒、木

收稿日期:2015-09-17.　网络出版日期:2016-08-24.
基金项目:国家自然科学基金项目(61175122);四川省科技支撑计划(2015GZ0345);四川省教育厅重点项目(15ZA0049).
通信作者:熊柳. E-mail:beartree1991@163.com.

马等会随之直接(如联网接入)或间接(如物理设备植入)地侵入工业控制系统,另一方面还有一些黑客组织或是敌对组织对国家的工控系统展开有组织的攻击如“震网”病毒事件,这种攻击具有未知性、隐秘性、持续时间长等特点,同时破坏效果更为强烈,而由于 SCADA 系统实时性、连续性和封闭性等特点,很多用在传统计算机安全方面的手段(如病毒库更新)又无法照搬过来用在 SCADA 系统上面,工控系统的信息安全问题日益突出。2002 年美国宣布将保护重要领域工业控制系统安全列入重要的工作内容,例如美国阿贡国家实验室(argonne national laboratory,ANL)的研究主要集中于美国天然气管道运输的 SCADA 系统安全领域,提出采用 SCADA 系统的蜜罐诱捕 SCADA 系统恶意攻击、进行脆弱性分析以及攻击行为分析。

本文首次提出基于因素神经网络[1]的油气集输 SCADA 安全防御模型,采用主动防御(active defense)的方法,对控制系统主机(工程师站、操作员站和 PLC 上位机等)的恶意程序行为进行分析来实现入侵检测的技术。主要研究 SCADA 安全因素神经元[2]的云推理机[3],实现对未知程序恶意行为的推测,并对其算法进行 MATLAB 仿真。

1　课题思路

对恶意程序行为的描述可以采用因素神经元的方法。因素神经元由刘增良教授提出,是建立在汪培庄教授的因素空间基础上提出的一种新的知识表示方法,因素空间理论[4]的主要目的是用于解释随机性的根源以及概率规律的数学实质,其作用就是“搭架子”,即建立一种广义坐标系,然后用以描述坐标系中的实际对象。因素神经元[1]是一个面向对象的综合知识表示与信息处理的单元模型,它能够获取各种类型的信息输入,神经元可以有选择性地进行感知,最后构成一个总体激发完成输出,能有效地作为模型的载体,这种表示方法具有语法和语义统一,定性与定量统一的知识表示特点。

完成了对恶意程序行为因素的表示和定性向定量转换后,需要通过一定的算法对这些行为因素进行分析,通过推理得到一个大致的结论,例如,这是一个什么样的恶意程序、其危害程度有多大。那么可以尝试采用云模型的方法来完成这个推理机。

云模型由李德毅院士在 1995 年提出[5],云是用自然语言值标识的某个定性概念与其定量表示之间的不确定性转换模型。云由许多云滴组成,每一个云滴就是这个定性概念在数域空间中的一次具体实现,这种实现带有不确定性。

2　基于因素神经元的 SCADA 系统恶意程序行为分析

2.1　SCADA 系统信息安全的因素空间构建

2.1.1　因素空间核心内容

“任何事物都是诸因素的交叉”,一个人可以由他在年龄、性别、身高、体重、职业、学历、性格、兴趣等因素方面的表现而加以确定[6],人就是上述因素的一种交叉。然而一个事物并非从任何因素都可以对之进行考察,如一块石头无从论性别,所谓事物 o 与因素 f 相关,是指从 f 讨论 o,有一个状态 $f(o)$ 与之对应。例如从 f(颜色因素)讨论 o(苹果),有一个状态 $f(o)$(红色)与之对应。

称 (O,V) 是一个配对,如果 O 与 V 分别是由一些对象和由一些因素所组成的集合,且对任意 $o\in O$,一切与 o 有关的因素都在 V 中,而对任意因素 $f\in V$,一切与 f 有关的事物也都在 O 中。

对于一个实际问题,假定有一个配对近似地存在着,对于给定的配对 (O,V),可以在 O 与 V 之间定义一个关系 R:

$R(o,f)=1$,当且仅当 o 与 f 有关。

称 R 为相关关系。为简便计把 R 定义为普通的(非 fuzzy,非模糊)关系。

$$D(f)=\{o\mid R(o,f)=1\}$$

$$V(o)=\{f\mid R(o,f)=1\}$$

因素 f 可以看作一个映射,作用在一定的对象 o 上可获得一定的状态,记为 $f(o)$:

$$f:D(f)\rightarrow X(f)$$

$$o\rightarrow f(o)$$

这里 $X(f)=\{f(o)\mid o\in O\}$ 叫做因素 f 的状态空间。

2.1.2　因素空间模型构建

给定论域 U 后,事物的因素分析一般步骤[6]为:

1)确定描述事物集合 O;

2)确定事物的因素集合 V;

3)对配对的 (O,V) 进行层次系统分析,确定其是否具有层次性、类别关系,是否需要构成因素分类树,从而实现优化分析;

4)确定各因素在论域 U 上的单因素状态空间;

5)构造离散因素空间;

6)构造原始事物描述离散模型。

大规模油气集输 SCADA 系统的信息安全防御是由多方面因素构成的,不是光从某一个因素入手就能解决的。对于 SCADA 系统信息安全防御,主要从恶意程序对系统的文件、注册表、进/线程操作这 3 方面的因素来考虑,每一种操作由多方面的因素构成,这样就可以把 SCADA 系统信息安全防御问题放在一个由有限个恶意程序行为因素构成的因素空间中,如图 1 所示。SCADA 系统信息安全就从

一个很笼统很抽象的问题对应到了一个多维的因素空间中,这是由模糊的定性概念转换为定量信息的第一步。

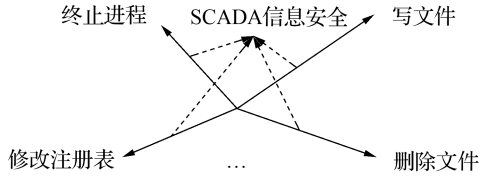

图 1 SCADA 信息安全因素空间

Fig.1 Factor space of SCADA information security

2.2 基于因素神经元的 SCADA 恶意程序行为因素表示方法

2.2.1 因素神经元核心内容

使用因素神经元的方法来表示 SCADA 恶意程序行为的因素。刘增良教授提出的"知识的因素表示模型"分为 2 种[4]:原子模型和关系模型。

1)原子模型

知识因素表示的原子模型是一个三元组即 $M(O)=\langle O,V,X\rangle$,也可记为 $M(O,V,X)$,其中 O 是事物集,V 是与 O 配对的因素集,X 是基于 V 的因素空间$\{X(f)\mid f\in F,F\subseteq V\}$。原子模型是描述事实、概念的模式。

2)关系模型

知识的关系模型是在原子模型基础上定义的,它可表示为:$R=\langle RM,M_i(O),XM\rangle,(i=1,2,\cdots)$也可记为 $R(RM,M_i(O),XM)$。其中 RM 表示原子模型 $M_i(O)$间的变换关系集,$M_i(O)$表示第 i 个原子模型,XM 表示原子模式 M 在知识模式集合 RM 上的状态及状态转换关系。关系模型是描述规则、推理、判断及知识动态过程的模式。

因素神经元分为解析型因素神经元和模拟型神经元[1],它们都是构成因素神经网络的基本单位。解析型因素神经元是对人类心理模式的模拟,模拟型因素神经元则是对人类生理模式的模拟。

1)解析因素神经元

一个具有推理功能的解析因素神经元可表述为:$M=\{\langle G,F,X\rangle,\langle p,q,r\rangle,\langle a,b\rangle\}$,其中$\langle G,F,X\rangle$共同表达了解析型因素神经元结构、因素及状态;$p,q,r$ 分别执行神经元的推理、判别与内部控制功能;a 是输入信息,b 是单元推理目标或响应。

一个解析型因素神经元就像一台微型自动机,通过一组用因素表达的状态,并有一套状态转换规则,当外部输入触发时,神经元按感知的信息执行相应的操作,进行状态的转换,并依据自身的输出响应函数,输出单元响应。解析型因素神经元基本结构包括输入/输出系统、判别系统、推理系统和控制系统 4 个基础系统构成,如图 2 所示。

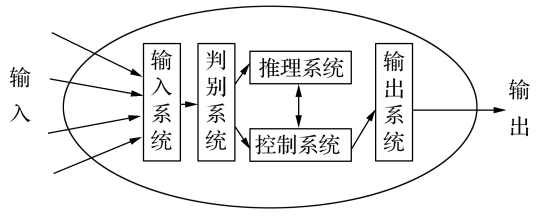

图 2 解析型因素神经元基本结构

Fig.2 Structure of analytic factor neuron

2)模拟型因素神经元

一个模拟型因素神经元可表述为:$Y=F(X,W,T)$,其中(X,Y)被称为模拟型因素神经元的输入输出模式对集合,X 被称为其激发或输入模式集合,Y 被称为其对应的响应或输出模式集合,W,T 为模拟型因素神经元内部网络模块的可控参数。

2.2.2 油气 SCADA 系统恶意程序行为因素表示方法

原子模型 $M(O)=\langle O,V,X\rangle$,$O$ 表示 SCADA 系统中恶意程序所有操作的集合,例如恶意程序对主机系统的文件、注册表、进/线程操作,可以表示为[2]:$O=\{$FileOperations, RegOperations, Process/ThreadOperations$\}$,对于其中的文件操作,又有查找、创建、打开、写入、复制、删除和属性设置等重要因素,即有因素集合 $F=\{$FindNextFile, CreateNewFile, OpenFile, WriteFile, CopyFile, DeleteFile, SetFileAttribute$\}$,$F\subseteq V$,V 为因素集,X 就是基于对应于因素集合的因素状态空间。

由此可以得到文件操作(FileOperations)知识因素表达的关系模型:$R=\langle R,M_i(O)\mid i=1,2,\cdots,n\rangle$,即 R(FileOperations)=$\langle RM$, M(FileOperations), $XM\rangle$,其中 RM 表示文件操作的一个知识模式集合,$RM=\langle\{$Rid$_1$: if XFile(Find Next File) then 遍历磁盘文件$\}$,$\{$Rid$_2$: if XFile(Create File) then 在指定路径下创建文件$\}$,$\{$Rid$_3$: if Xfile(Copy File) then 复制指定文件到指定路径$\}$,$\{$Rid$_4$: if XFile(Delete File) then 删除指定路径的指定文件$\}$,$\cdots\rangle$;M(FileOperations)为文件操作的知识因素表达原子模型;XM 表示原子模型 M(FileOperations)在知识模式集合 RM 上的状态及状态转换关系,$XM=\langle$Xid$_1$:if 在系统目录下创建文件 and 设置文件时间为系统文件时间 then 非法创建文件$\}$,$\{$Xid$_2$:if 查找所有文件 and 文件写操作 then 可疑遍历磁盘文件操作$\}$,$\{$Xid$_3$:if 创建打开文件 and 修改文件属性为系统文件 then 非法文件属性修改$\}$,$\{$Xid$_4$:if 删除 SCADA 主机系统目录文件 then 恶意文件操作$\}$,$\{$Xid$_5$:if 复制文件至 SCADA 系统目录 and 修改该目录下文件属性 then 可疑文件操作$\}$,$\cdots\rangle$。

3　基于云发生器的 SCADA 恶意程序行为因素推理机设计

在人工智能领域,对知识和推理的不确定性主要分为模糊性和随机性展开研究。作为处理模糊性问题的主要工具,模糊集理论用隶属度来刻画模糊食物的亦此亦彼性。然而,一旦用一个精确的隶属函数来描述模糊集,模糊概念被强行纳入到精确数学的王国,从此以后,在概念的定义、定理的叙述及证明等数学思维环节中,就不再有丝毫的模糊性了。由于传统模糊集理论的不彻底性[7],该推理机没有采用传统的隶属度算法,而是采用云模型理论。

3.1　云模型简介

云模型是一个以自然语言值为切入点,实现定性概念与定量值之间的不确定性转换的模型[7];它同时反映了客观世界中概念的两种不确定性,即随机性(发生的概率)和模糊性(亦此亦彼性),尤其随机性是其不同于传统隶属函数的特性。下面给出云模型的定义。

设 U 是一个用精确数值表示的定量论域,C 是 U 上的定性概念,若定量值 $x \in U$,且 x 是定性概念 C 的一次随机实现,x 对 C 的确定度 $u(x) \in [0,1]$ 是具有稳定倾向的随机数。若

$$u: U \to [0,1] \quad \forall x \in U \quad x \to u(x)$$

则 x 在论域 U 上的分布称为云,每一个 x 称为一个云滴。

云模型主要有以下 3 个数字特征[5]:

1)期望 Ex,云滴在论域空间分布的期望是概念在论域空间的中心值,是最能够代表定性概念的点;

2)熵 En,它是定性概念不确定性的度量,是由定性概念的随机性和模糊性共同决定的;

3)超熵 He,它是对熵的不确定性的度量,是熵的熵,反映了在论域空间代表该语言值的所有点的不确定度的凝聚性,它的大小间接地反映了它们之间的关联性。

由参数 Ex、En、He 得到云滴 drop(x,u) 的过程称为正向云发生器,如图 3 所示,这是从定性概念到定量表示的转换过程;反之,则为逆向云发生器。

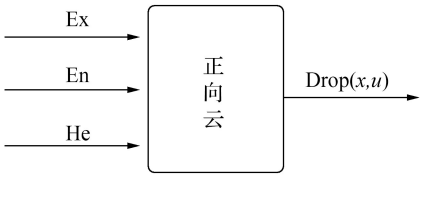

图 3　正向云发生器

Fig.3　F-cloud generator

由此可以衍生出 X 条件云发生器和 Y 条件云发生器[3]。即当给定云的 3 个数字特征和特定的 $x=x_0$ 时,产生满足条件的云滴 drop(x_0,u_i),X 条件云也称为前件云发生器;同理,当给定云的 3 个数字特征和特定的 $y = u_i$ 时,产生满足条件的云滴 drop(x_i,u_i),Y 条件云也称为后件云发生器。

3.2　云发生器的恶意程序行为因素推理机设计

3.2.1　多条件多规则发生器

一条定性规则[8]可以形式化地表示为:if A then B,其中 A、B 为语言值表示的云对象[7]。云发生器是运用云模型进行不确定性推理的基础。X 条件云和 Y 条件云可以组合构造成单条件单规则发生器,在前件云发生器中输入值 x_0 激活 CGx(x 条件云发生器)时,CGx 随机产生一个隶属度 u_i,这个值反映了 x_0 对此定性规则的激活强度,而 u_i 又作为 CGy(y 条件云发生器)的输入,随机地产生一个云滴 drop(x_i,u_i)。通过云模型构造的定性规则,使得这种推理系统对不确定性具有良好的继承性和传递性。

根据单条件单规则发生器的构造原理可以构造出多条件多规则发生器。以二维多规则生成器为例,如图 4 所示。

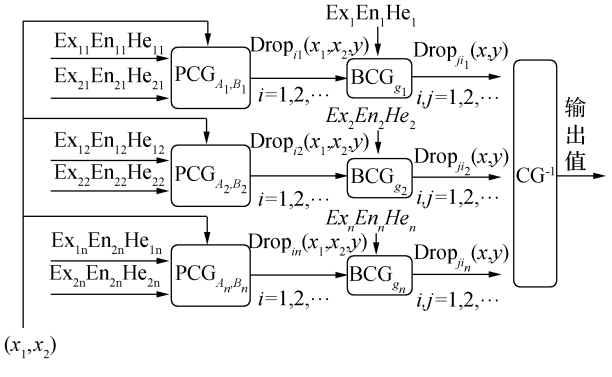

图 4　二维多规则生成器

Fig.4　Two dimensional multi rules generator

3.2.2　云推理机规则的因素神经元表示

前面在“SCADA 系统恶意程序行为因素表示方法[9]”中写出了关系模型的表达结构,其中 XM 表示原子模式 M 在知识模式集合 RM 上的状态及状态转换关系。以 XM 集合中的{Xid1:if 在系统目录下创建文件[10] and 设置文件时间为系统文件时间 then 非法创建文件}为例,“在系统目录下穿件文件(CreatFile)”和“设置文件时间为系统文件时间(SetFileTime)”为 SCADA 系统恶意程序行为[11]因素集合[12]中的两个因素,对应于云模型二维单规则“if A and B then C”中的条件 A 和 B,而“非法创建文件”则对应于规则结论 C。这样就实现了将模糊的定性概念转换为定量信息,便于推理机处理。

3.2.3　云推理机算法结构与设计

根据多条件多规则发生器的原理,以二维规则

发生器[13]为例,构建步骤如下:

1)给定各规则 2 个条件对象的参数,即 Ex_1、En_1、He_1、Ex_2、En_2、He_2;

2)由以上 2 组参数计算出单个规则对应的二维云发生器,MATLAB 程序如下:

```
%二维云的合成并画出其三维图像
clear:
cIc:
%输入单规则推理两个条件的参数
Exl=input('pleaseinputExl=');
Ex2=input('pleaseinputEx2=');
Enl=input('pleaseinputEnl=');
En2=input('pleaseinputEn2=');
Hel=input('pleaseinputHel=');
He2=input('pleaseinputHe2=');
N=input('pleaseinputN=');
temp=O:
%生成二维条件云并计算(X1 X2)时的隶属度
for i=1:N
[Ensl(1,i),Ens2(1,i)]=binormrand(En1,
En2,Hel,He2):
[X1(1,i),X2(1,i)]=binormrand(Exl,Ex2,
Ensl(1,i),Ens2(1,i)):
end
for i=1:N
q=(X1(1,i)一 Exl)^2/Ensl(1,i)^2:
d=(X2(1,i)-Ex2)^2/Ens2(1,i)^2:
Y(1,i)=exp(一(1/2)*(q+d)):
end
plot3(x1,X2,Y,'.').
```

生成二维隶属云如图 5 所示。

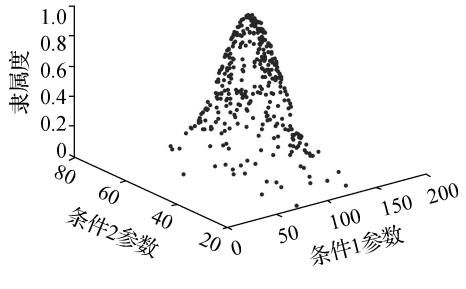

图 5 二维隶属云

Fig.5 Two dimensional cloud

3)输入待测条件参数 X_1 和 X_2,激活每一条规则,根据每一条规则条件参数生成的二维云发生器计算得到激活强度,即隶属度 u_i。MATLAB 程序如下:

```
%……单规则云发生器程序……%
function[Y]=onerule(Exl,Enl,Hel,Ex2,En2,He2,
X1,X2,N)
temp=O:
for i=1:N
    [Ensl(1,i),Ens2(1,i)]=binormrand(En1,
En2,Hel,He2):
end
for i=1:N
    q:(X1-Exl)^2/Ensl(1,i)^2:
    d=(X2-Ex2)^2/Ens2(1,i)^2:
    Y(1,i)=exp(-(1/2)*(q+d)):
    temp=temp+Y(1,i):
end
Y=temp/Y:
```

4)在计算出的 u_i 中选择最大 u_1 和次大的 u_2,它们对应的规则即为激活强度[14]最大的两条规则,MATLAB 程序图如下:

```
[u,IX]=sort(U):%将所有规则计算出来的隶属度
由小到大排列
%选择隶属度最大的 Ul 和次大的 U2,说明这两条
规则的激活强度最大
u1=u(M):
u2=u(M-1):
num1=IX(M):
num2=IX(M-1):
fprintf('触发强度最大规则为第%i 条\n',num1);
fprintf('触发强度最大规则为第%i 条\n',num2);
```

5)利用 Y 条件云由 U_1 和 U_2 分别计算出 2 组云滴,选择距离最小的 2 个云滴,通过逆向云发生器计算推理恶意度的期望和方差,MATLAB 程序如下:

```
%构造这两条规则对应的后件云发生器(Y 条件云)
[Y1(1),Y1(2)]=ycloud(d(numl,7),d(numl,8),
d(numl,9),ul,N);
[Y2(1),Y2(2)]=ycloud(d(num2,7),d(num2,
8),d(num2,9),u2,N);
%从 Y1 和 Y2 中分别选取一点使两点距离最小,
%由 ycloud 可知第一个大于第二个
    a=abs(Y1(1)-Y2(2)):
    b=abs(Y2(1)一 Y1(2)):
    if a>b
      Z1=Y2(1);
      zul=u2;
      z2=Y1(2);
      zu2=ul;
    else
      z1=Y1(1);
      zul=ul;
      z2=Y2(2);
      zu2=u2;
```

```
    end
    %利用逆向云发生器计算
    [Ex,En,He]=backc loud(z1,zu1,z2,zu2);
    com1=com1+Ex;
    com2=com2+En;
end
C=coml/5;
D=com2/5;
fprintf('云推理机预测程序行为恶意度期望为%f\n',C);
fprintf('云推理机预测程序行为恶意度期望为%f\n',D);
```

3.3　仿真实验与分析

3.3.1　建立对比实验

云模型理论建立在传统隶属函数[15]的基础上,其随机性是不同于传统隶属函数方法的特性。所以,利用传统隶属函数方法建立对比实验,可以加强对该云推理机优势和缺陷的分析,从而进行后续改进。传统隶属函数方法对比实验主要利用 Matlab 中的 FIS(模糊逻辑工具箱)编辑器,步骤如下:

1)建立隶属度函数。云推理机测试样本数据如表 1 所示。由数据可知,应该选择 Gauss 隶属函数(正态型),一共有 9 条规则,每一条规则有 2 个 input,1 个 output,均按照其期望(Ex)和熵(En)建立 Gauss 隶属函数。

表 1　云推理机测试样本数据

Table 1　Sample data

Rules	Factor A			Factor B			Factor C			Cla-les
	Ex_1	En_1	He_1	Ex_2	En_2	He_2	Ex_3	En_3	He_3	
1	99.6	18.3	0.18	49.7	9.8	0.1	75.5	7.3	0.15	*A*
2	47.2	9.7	0.1	98.9	14.5	0.15	74.4	7.6	0.15	*B*
3	74	11.8	0.12	72.6	13.1	0.13	65.2	4.1	0.08	*C*
4	69	12.2	0.13	25.3	12.3	0.12	45.5	3.15	0.06	*D*
5	47	11.9	0.12	50	12	0.12	50.6	3.1	0.06	*E*
6	23	12.9	0.12	74.8	12.7	0.12	58.9	3.0	0.06	*F*
7	27	12.1	0.12	27.6	13.2	0.13	31.0	6.2	0.12	*G*
8	3	18.7	018	52.4	11	0.1	20.8	7.5	0.15	*H*
9	52	10.8	0.1	1.6	1.7	0.17	19.0	7.5	0.15	*I*

2)编辑器建立推理规则。例如:If(条件 a is 1a)and (条件 b is 1b)then(推论 c is 1c)。条件 A、B、C 指广泛意义上的定义概念,其小写 a、b、c 指代具体的实际条件。

3)得到 Surface 三维图如图 6 所示,并在云推理机主程序中用 evalfis 函数调用,得到隶属函数法推理出的结论,并与云推理机得到结论进行对比分析。

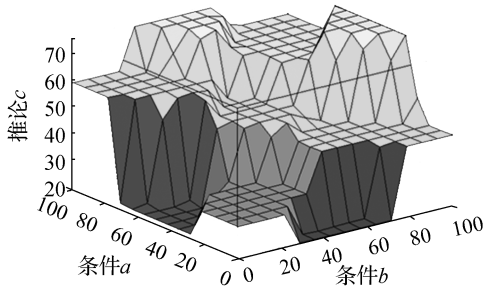

图 6　隶属函数法推理 Surface 图

Fig.6　FIS surface

3.3.2　仿真实验

已知数据:以“if 在系统目录下创建文件 and 设置文件时间为系统文件时间 then 非法创建文件”为例,因为各因素都是模糊语言值表示的概念,根据“知识因素表示的关系模型”原理,给各因素假设一个数字量的参考指数,指数满分为 100,指数越高表示确定度越大;反之越小。

待测数据:现有一未知恶意程序通过课题小组建立的感知神经元行为分解并分析后得到条件“在系统目录下创建文件因素指数 X_1”和“设置文件时间为系统文件时间指数 X_2”。

表 1 中,Factor A 代表“在系统目录下创建文件因素指数 A”,Factor B 代表“设置文件时间为系统文件时间因素指数 B”,Factor C 代表“非法创建文件因素指数 C”,Class 代表“每条规则对应的恶意程序种类”。将仿真结果统计如表 2 所示。

表 2　仿真结果数据统计表

Table 2　Reslult data

(x_1,x_2)	CG				FIS	
	Rules		Infer		Rules	Infer
	Max	Sec	Ex	Var		
(99,50)	1	3	75.78	0.85	1	75.50
(74,87.9)	3	2	61.39	12.17	Null	74.46
(35,83)	6	2	62.10	0.34	Null	60.02
(4.5,52.6)	8	6	36.75	21.67	8	20.80

表 2 中,(x_1,x_2) 表示待测样本数据的条件 a 和条件 b 的参数;CG 表示云模型推理机方法,FIS 表示利用模糊工具箱实现的隶属度函数法;Rules 表示触发规则,Max 表示激活强度最大的触发规则,Sec 表示次大,Null 表示无触发规则;Infer 表示推理机对待测样本的推测恶意度,Ex 表示期望,Var 表示方差。

云推理机得到的恶意度分布图和隶属函数法得到的规则图如图 7 ~ 10 所示(以(99,50)和(74,87.9)为例)。

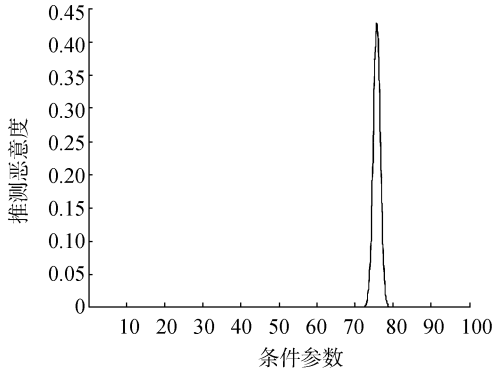

图 7 样本(99,50)云推理机推测恶意度

Fig.7 CG inference result of sample(99,50)

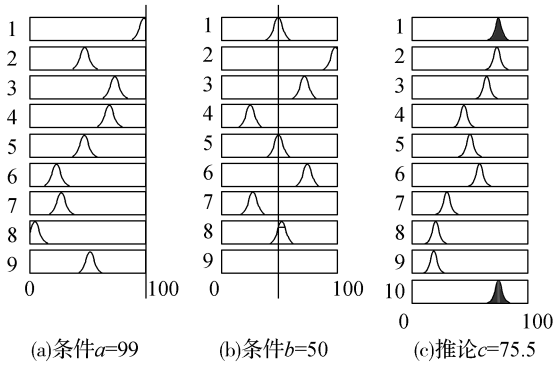

图 8 样本(99,50)隶属函数法推测恶意度

Fig.8 FIS inference result of sample(99,50)

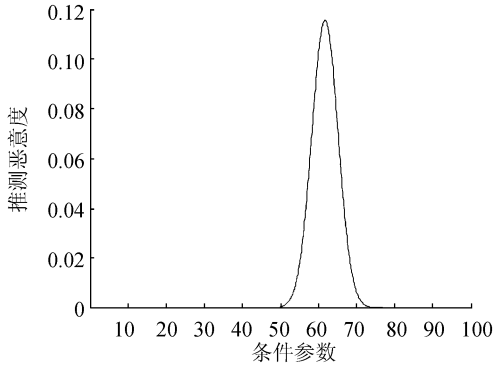

图 9 样本(74,87.9)云推理机推测恶意度

Fig.9 CG inference result of sample(74,87.9)

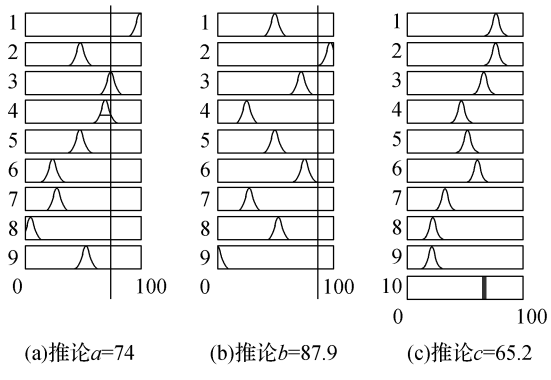

图 10 样本(74,87.9)隶属函数法推测恶意度

Fig.10 FIS inference result of sample(74,87.9)

3.3.3 实验分析

1)从云推理机得到的触发规则和已知样本数据来看,第一组样本(99,50)基本可以确定是 A 类(A,B,C,D,E,F,G,H,I 为 9 类已知的不同恶意程序,其数据参数如表 1 所示);第 2 组样本(74,87.9)非常可能是 C 类,如果不是 C 类,那么就应该不属于已知样本中的种类;第 3 组样本(35,83)可能是 F 和 B 类,F 类的可能性最大;第四组样本(4.5,52.6)基本可以确定是 H 类。从推测恶意度来看,前 3 组样本恶意程序非法创建文件的可能性比较大,第 4 组的可能性较小。

2)对比云推理机和隶属度函数法的触发规则可以发现,如果待测样本的 X_1 和 X_2 参数比较接近已知样本数据中的某一类,那么云推理机和隶属函数法都会激活对应的那条规则(如第 1 组和第 4 组),但是如果待测样本没有很明显的相似已知样本,那么传统的隶属函数法就不能得到触发规则,从而对结论推测没有很好地指引,而云推理机可以得到,从而进行推测结论。

3)对比云推理机和隶属函数法得到的推测恶意度可以看出,两种方法都可以对未知样本的恶意度进行推测。从数据上看,云推理机得到的恶意度并不是一个具体数据,而是服从正态分布的数据,体现了云模型理论区别于传统隶属函数的特性——随机性,但是隶属函数法得到结果基本都非常接近于已知样本的数据(Ex_3),所以很可能只是单纯按照已知样本数据构建的规则来计算。因此,云推理机得到的推测恶意度更为科学、更为可信。

4)云推理机最后只是得到了对推测有指引作用的数据,并没有直接得到具体结论,需要人工对这些指引数据进行进一步分析,这一点还需优化。总体来说,云推理机实现了对未知恶意程序非法创建文件恶意度的推测,达到了设计目的,但还需完善。

4 结束语

将云推理机的核心直观表现出来的其实就是“if A and B then C”规则,而 A、B、C 都是语言值表示的模糊概念,那么需要将语言值表示的模糊概念转换为定量信息,才能实现计算机数据推理,这就是知识因素的因素神经元表示方法在本文中的作用。对于 SCADA 系统信息安全的主动防御,通过因素神经元和云模型推理机结合的方法是一种尝试,通过文章中的仿真结果可以看出达到了预测效果,说明了这种方法实施的可能性与合理性,而且该算法还可结合一些优化算法来提升其效率和精度,这是作者今后着重研究的方向。

参考文献:

[1]刘增良. 因素神经网络理论及其应用[M]. 贵阳: 贵州科技出版社, 1994.
LIU Zengliang. Factor neural networks and its application [M]. Guiyang: Guizhou Science and Technology Publishing House, 1994.

[2]YANG Li, GAO Xiedong, LI Jie, et al. Research on FNN-based security defence architecture model of scada network [C]//Proceedings of the 2nd International Conference on Cloud Computing and Intelligence Systems. Hangzhou, China, 2012.

[3]QIN Yong, CAO Xiedong, LIANG Peng, et al. Research on the analytic factor neuron model based on cloud generator and its application in oil & gas scada security defense[C]// Proceedings of 2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems. Shenzhen, China, 2014: 5.

[4]刘增良. 知识的因素表示[C]//中国科学技术协会首届青年学术年会论文集(工科分册·上册). 北京, 1992: 5.
LIU Zengliang. The factor representation of knowledge [C]//Proceedings of the 1st National Association of Science and Technology of Youth (Engineering. First). Beijing, 1992: 5.

[5]杨朝晖, 李德毅. 二维云模型及其在预测中的应用[J]. 计算机学报, 1998, 21(11): 961-969.
YANG Zhaohui, LI Deyi. Planar model and its application in prediction[J]. Chinese journal of computers, 1998, 21(11): 961-969.

[6]汪培庄. 因素空间与概念描述[J]. 软件学报, 1992, 3(2): 30-40.
WANG Peizhuang. Factor space and description of concepts [J]. Journal of software, 1992, 3(2): 30-40.

[7]王树良. 基于数据场与云模型的空间数据挖掘和知识发现[D]. 武汉: 武汉大学, 2002.
WANG Shuliang. Data field and cloud model based spatial data mining and knowledge discovery[D]. Wuhan: Wuhan University, 2002.

[8]曹谢东. 模糊信息处理及应用[M]. 北京: 科学出版社, 2003: 103-104.
CAO Xiedong. Fuzzy information processing and application [M]. Beijing: Science Press, 2003: 103-104.

[9]YANG Li, CAO Xiedong, LI Jie, et al. A new formal description model of network attacking and defence knowledge of oil and gas field SCADA system[M]//WANG Hua, ZOU Lei, HUANG Guangyan, et al. Web Technologies and Applications. Berlin Heidelberg: Springer, 2012.

[10]YANG Li. A new factor state space model for SCADA network attack and defense[J]. International journal of security and its applications, 2014, 8(6): 303-314.

[11]王盼卿, 李晓辉. 一种基于知识因素表示理论的装备信息分类描述方法[J]. 国防科技大学学报, 2011, 35(5): 150-155.
WANG Panqin, LI Xiaohui. A description method on classification of equipment information based on knowledge factor expression theory[J]. Journal of national university of defense technology, 2011, 35(5): 150-155.

[12]刘增良, 刘有才. 因素神经网络理论及实现策略研究[M]. 北京: 北京师范大学出版社, 1992.
LIU Zengliang, LIU Youcai. Research on the theory of factor neural network and its realization[M]. Beijing: Research on the Theory of Factor Neural Network and Its Realization, 1992.

[13]付斌, 李道国, 王慕快. 云模型研究的回顾与展望[J]. 计算机应用研究, 2011, 28(2): 420-426.
FU Bin, LI Daoguo, WANG Mukuai. Review and prospect on research of cloud model[J]. Application research of computers, 2011, 28(2): 420-426.

[14]张飞舟, 范跃祖, 沈程智, 等. 利用云模型实现智能控制倒立摆[J]. 控制理论与应用, 2000, 17(4): 519-523.
ZHANG Feizhou, FAN Yuezu, SHEN Chengzhi, et al. Intelligent control inverted pendulum with cloud models[J]. Control theory and application, 2000, 17(4): 519-523.

[15]李德毅, 杜鹢. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005: 7.
LI Deyi, DU Yi. Artificial intelligence with uncertainty [M]. Beijing: National Defend Industry Press, 2005: 7.

作者简介:

熊柳,男,1991 年生,硕士研究生,主要研究方向为模式识别与智能控制,参加国家自然基金项目 1 项。

曹谢东,男,1954 年生,教授,主要研究方向为人工智能、工业控制信息系统安全、智能测控等。主持国家自然科学基金项目,承担国家 863、国家重大科技攻关项目和省部级项目多项,获四川省科技进步二等奖 1 项、三等奖 2 项,专著 3 部。