(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109143848 A
(43)申请公布日 2019.01.04

(21)申请号 201710498512.9

(22)申请日 2017.06.27

(71)申请人 中国科学院沈阳自动化研究所
地址 110016 辽宁省沈阳市东陵区南塔街114号

(72)发明人 尚文利 赵剑明 万明 崔君荣 刘贤达 曾鹏 于海斌

(74)专利代理机构 沈阳科苑专利商标代理有限公司 21002
代理人 王倩

(51)Int.Cl.
G05B 13/04(2006.01)

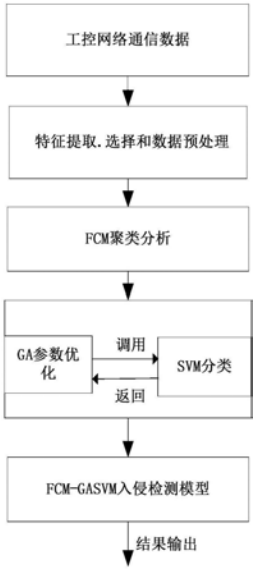
权利要求书2页 说明书8页 附图2页

(54)发明名称

基于FCM-GASVM的工业控制系统入侵检测方法

(57)摘要

本发明涉及基于FCM-GASVM的工业控制系统入侵检测方法，具体为基于FCM-GASVM算法提出了一种工业控制系统应用层网络入侵检测方法，该方法将无监督的模糊C-均值聚类和有监督的支持向量机相结合，提取工业控制系统Modbus/TCP协议的通信流量数据，设计了一种先将通信数据利用FCM聚类，后将满足阈值条件的部分数据进一步由遗传算法优化的支持向量机分类的方法。该方法将无监督学习和有监督学习完美结合，并且在不需要提前知道类别标签的前提下即可有效的降低训练时间，提高分类精度。



CN 109143848 A

1.一种基于FCM-GASVM的工业控制系统入侵检测方法,其特征在于,包括以下步骤:

特征提取:获取通信流量数据包,并提取最能反映数据特征的属性;

特征构造:根据入侵模式构造工业基本特征;

数据预处理:按时间顺序对所提取和构造的数据包进行排序,随机分割成不同的序列,去除各序列中存在的冗余数据包后,对每个序列里的数据进行归一化得到数据向量;

FCM聚类:将数据向量分簇并计算簇的聚类中心,根据每一个数据向量与聚类中心的距离得到数据集作为训练集A,形成训练模型;

GASVM:将不满足阈值条件的数据集通过遗传算法分别得到惩罚因子C和核函数参数g,通过支持向量机检测,得到检测模型;

FCM-GASVM:根据训练模型和检测模型建立入侵检测模型进行异常检测。

2.根据权利要求1所述的基于FCM-GASVM的工业控制系统入侵检测方法,其特征在于,所述特征提取包括以下步骤:

抓取Modbus/TCP通信流量数据包,对需要提取的Modbus/TCP属性标号,查询所标号的属性的数据帧,计算数据帧头中的地址,提取标号属性的数据值。

3.根据权利要求1所述的基于FCM-GASVM的工业控制系统入侵检测方法,其特征在于,所述特征构造包括:若干秒内功能码请求次数,若干秒内访问地址次数,若干秒内连接同一设备次数。

4.根据权利要求1所述的基于FCM-GASVM的工业控制系统入侵检测方法,其特征在于,所述归一化包括以下步骤:

采用最小最大标准化的方法将不同单位和量纲的数据归一成统一的形式:

$$v' = \frac{v - \min}{\max - \min}(\max' - \min') + \min'$$

其中,Max和min分别表示某序列中数据的最大值、最小值;max'和min'分别表示映射新空间的区间(min',max');v为该序列内的每个数据,表示输入向量,;v'为数据向量,表示输出向量。

5.根据权利要求1所述的基于FCM-GASVM的工业控制系统入侵检测方法,其特征在于,所述FCM聚类包括以下步骤:

对数据向量进行FCM聚类,得到每个簇的聚类中心O,所有的正常聚类中心标记为O+,所有的异常聚类中心标记为O-,正常集合标记为A+,异常集合标记为A-,设置类别标签;λ表示阈值;

对于每个数据向量xi,计算与聚类中心的距离,判定数据向量的隶属度和目标函数,若满足xi与O+的距离小于λ,则标记该数据向量xi∈A+,否则标记xi∈A-;

构成训练集A=A+∪A-。

6.根据权利要求5所述的基于FCM-GASVM的工业控制系统入侵检测方法,其特征在于,所述聚类中心为:

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i=1}^{n} u_{ij}^{m}}$$

隶属度为：

$$u_{ij}=\begin{cases}\left[\sum_{k=1}^{c}\frac{\left\|x_i-v_j\right\|^{\frac{2}{m-1}}}{\left\|x_i-v_k\right\|^{\frac{2}{m-1}}}\right]^{-1}\\ 1\\ 0\end{cases}$$

目标函数为：

$$J=\sum_{i=1}^{n}\sum_{j=1}^{c}\left(u_{ij}\right)^{m}\left\|x_i-v_j\right\|$$

其中，u_{ij}为个体x_i属于第j类的模糊隶属度；m为模糊权重指数；v_j为第j类的聚类中心；n为数据向量总个数；c为类别数。

7.根据权利要求1所述的基于FCM-GASVM的工业控制系统入侵检测方法，其特征在于，所述GASVM算法包括以下步骤：

对不满足阈值条件的数据集进行参数初始化，计算个体适应度，对其进行选择、交叉、变异操作，得到最优的惩罚因子C和核函数参数g；

将最优的惩罚因子C和核函数参数g带入支持向量机；

将FCM聚类得到的类别标签赋予给支持向量；

根据构造对偶问题和决策函数得到分类。

8.根据权利要求7所述的基于FCM-GASVM的工业控制系统入侵检测方法，其特征在于，所述对偶问题为：

$$\min Q(a)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K\left(x_i,\ x_j\right)-\sum_{i=1}^{n}\alpha_i$$

$$s.t.\sum_{i=1}^{l}\alpha_i y_i=0,\qquad 0\le\alpha_i\le C,i=1,\ldots,n$$

其中，Q(α)表示对偶运算，α＝($\alpha_1,\alpha_2,\ldots,\alpha_n$)表示拉格朗日算子，K($x_i,x_j$)表示高斯径向基核函数，得解$\alpha^*$＝($\alpha_1^*,\alpha_2^*,\ldots,\alpha_n^*$)；n表示数据向量总个数；l＝n；

决策函数为：

$$b^*=y_j-\sum_{i=1}^{n}y_i\alpha_i^*K(x_i,x_j)$$

$$f(x)=\operatorname{sgn}(\omega\cdot\Phi(x)+b)=\operatorname{sgn}(\sum_{i=1}^{n}\alpha_i y_i K(x_i\cdot x)+b)$$

其中，b^*是支持向量机的最终决策函数的阈值，sgn()表示符号函数；b＝b*，y_j表示分类的标签。

9.根据权利要求1所述的基于FCM-GASVM的工业控制系统入侵检测方法，其特征在于，所述FCM-GASVM算法包括以下步骤：

根据训练模型和检测模型得到入侵检测模型的分类准确率。

基于FCM-GASVM的工业控制系统入侵检测方法

技术领域

[0001] 本发明涉及一种基于FCM-GASVM的工业控制系统入侵检测方法，利用模糊C均值和遗传算法优化的支持向量机对异常行为进行检测，属于工业控制网络安全领域。

背景技术

[0002] 传统的工业控制系统一般以厂区为单位，相互之间是独立的，与外界之间没有物理连接。但是随着工业信息化和网络技术的迅速发展，工业控制系统越来越多的采用通用硬件和通用软件，工控系统的开放性与日俱增，系统安全漏洞和缺陷容易被病毒所利用，然而工业控制系统又应用于国家的电力、交通、石油、取暖、制药等多种大型制造行业，一旦遭受攻击会带来巨大的损失，因此需要有效的方法确保工控系统的网络安全。

[0003] 保护工业控制系统的网络安全有多种方式，最常用的是采用防火墙、日志处理等联动方式，然而防火墙是基于第三方的路由访问控制，不能检测来自系统内部的攻击，只能起到过滤的作用，无法有效降低系统的安全风险。入侵检测(Intrusion Detection,ID)作为一种主动防御技术，在检测外部攻击的同时又能很好的检测系统内部攻击，将防护、检测、响应有效的融为一体，为工控网络的安全提供更加可靠的保障。

[0004] 入侵检测技术可以有效的应用在工业控制系统中，国内外研究学者和专家也对其进行了大量的研究，本文通过分析Modbus的通信行为，研究基于半监督分簇策略的工业控制网络入侵检测算法，建立工业控制系统的入侵检测模型，对异常入侵做出及时的检测，从而实现对工控系统的保护。

发明内容

[0005] 有鉴于此，本发明的目的是提供一种基于FCM-SVM的工业控制入侵检测方法，实现了对工控攻击行为的检测。

[0006] 本发明解决其技术问题所采用的技术方案是：一种基于FCM-GASVM的工业控制系统入侵检测方法，包括以下步骤：

[0007] 特征提取：获取通信流量数据包，并提取最能反映数据特征的属性；

[0008] 特征构造：根据入侵模式构造工业基本特征；

[0009] 数据预处理：按时间顺序对所提取和构造的数据包进行排序，随机分割成不同的序列，去除各序列中存在的冗余数据包后，对每个序列里的数据进行归一化得到数据向量；

[0010] FCM聚类：将数据向量分簇并计算簇的聚类中心，根据每一个数据向量与聚类中心的距离得到数据集作为训练集A，形成训练模型；

[0011] GASVM：将不满足阈值条件的数据集通过遗传算法分别得到惩罚因子C和核函数参数g，通过支持向量机检测，得到检测模型；

[0012] FCM-GASVM：根据训练模型和检测模型建立入侵检测模型进行异常检测。

[0013] 所述特征提取包括以下步骤：

[0014] 抓取Modbus/TCP通信流量数据包，对需要提取的Modbus/TCP属性标号，查询所标

号的属性的数据帧，计算数据帧头中的地址，提取标号属性的数据值。

[0015] 所述特征构造包括：若干秒内功能码请求次数，若干秒内访问地址次数，若干秒内连接同一设备次数。

[0016] 所述归一化包括以下步骤：

[0017] 采用最小最大标准化的方法将不同单位和量纲的数据归一成统一的形式：

[0018] $$v'=\frac{v-\min}{\max-\min}(\max'-\min')+\min'$$

[0019] 其中，Max和min分别表示某序列中数据的最大值、最小值；max’和min’分别表示映射新空间的区间(min’,max’)；v为该序列内的每个数据，表示输入向量，；v’为数据向量，表示输出向量。

[0020] 所述FCM聚类包括以下步骤：

[0021] 对数据向量进行FCM聚类，得到每个簇的聚类中心O，所有的正常聚类中心标记为O+，所有的异常聚类中心标记为O-，正常集合标记为A+，异常集合标记为A-，设置类别标签；λ表示阈值；

[0022] 对于每个数据向量xi，计算与聚类中心的距离，判定数据向量的隶属度和目标函数，若满足xi与O+的距离小于λ，则标记该数据向量xi∈A+，否则标记xi∈A-；

[0023] 构成训练集A＝A+∪A-。

[0024] 所述聚类中心为：

[0025] $$v_j=\frac{\sum_{i=1}^{n}u_{ij}^{m}x_i}{\sum_{i=1}^{n}u_{ij}^{m}}$$

[0026] 隶属度为：

[0027] $$u_{ij}=\begin{cases}\left[\sum_{k=1}^{c}\frac{\|x_i-v_j\|^{\frac{2}{m-1}}}{\|x_i-v_k\|^{\frac{2}{m-1}}}\right]^{-1}\\ 1\\ 0\end{cases}$$

[0028] 目标函数为：

[0029] $$J=\sum_{i=1}^{n}\sum_{j=1}^{c}(u_{ij})^{m}\|x_i-v_j\|$$

[0030] 其中，u_{ij}为个体x_i属于第j类的模糊隶属度；m为模糊权重指数；v_j为第j类的聚类中心；n为数据向量总个数；c为类别数。

[0031] 所述GASVM算法包括以下步骤：

[0032] 对不满足阈值条件的数据集进行参数初始化，计算个体适应度，对其进行选择、交叉、变异操作，得到最优的惩罚因子C和核函数参数g；

[0033] 将最优的惩罚因子C和核函数参数g带入支持向量机；

[0034] 将FCM聚类得到的类别标签赋予给支持向量；

[0035] 根据构造对偶问题和决策函数得到分类。

[0036] 所述对偶问题为：

[0037] $$\min Q(a)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,\ x_j)-\sum_{i=1}^{n}\alpha_i$$

[0038] $$s.t.\sum_{i=1}^{l}\alpha_i y_i=0,\qquad 0\le\alpha_i\le C,i=1,...,n$$

[0039] 其中，Q(α)表示对偶运算，α＝(α1,α2,...,αn)表示拉格朗日算子，K(xi,xj)表示高斯径向基核函数，得解α*＝(α1*,α2*,...,αn*)；n表示数据向量总个数；l＝n；

[0040] 决策函数为：

[0041] $$b^*=y_j-\sum_{i=1}^{n}y_i\alpha_i^*K(x_i,x_j)$$

[0042] $$f(x)=\mathrm{sgn}(\omega\cdot\Phi(x)+b)=\mathrm{sgn}(\sum_{i=1}^{n}\alpha_i y_i K(x_i\cdot x)+b)$$

[0043] 其中，b*是支持向量机的最终决策函数的阈值，sgn()表示符号函数；b＝b*，yi∈R＝{-1,1}，R表示实数，yj表示分类的标签，正常记为1，异常记为-1。

[0044] 所述FCM-GASVM算法包括以下步骤：

[0045] 根据训练模型和检测模型得到入侵检测模型的分类准确率。

[0046] 本发明具有以下有益效果及优点：

[0047] 1.本发明选取工控通信协议Modbus/TCP为主要研究对象，对工控数据进行了提取和构造，提出了一种基于无监督聚类和遗传算法优化的有监督支持向量机的工控入侵检测方法，建立了半监督的工控入侵检测模型，该模型适合于处理小样本数据的分类问题。

[0048] 2.本发明提出的入侵检测模型在不需要提前知道标签的前提下就可对工控数据进行有效的检测，打破了传统的必须知道类别标签的局限性。

[0049] 3.本文提出的FCM-GASVM模型将无监督学习和有监督学习完美结合，有效的降低了训练时间，提高了分类精度。

附图说明

[0050] 图1是基于FCM-GASVM的工控入侵检测算法整体框架图；

[0051] 图2是FCM-GASVM入侵检测模型。

具体实施方式

[0052] 下面结合实施例对本发明做进一步的详细说明。

[0053] 基于FCM-SVM的工业控制入侵检测方法，包括以下步骤：

[0054] 步骤一：首先用wireshark抓取Modbus/TCP通信流量数据包，对于每一个Modbus TCP/IP协议均有多种属性，从中提取出最能反映数据特征的属性。

[0055] 步骤二：根据入侵模式，构造工业基本特征，10秒内功能码请求次数，20秒内访问地址次数，10秒内连接同一设备次数。

[0056] 步骤三：按时间顺序对提取和构造的数据包进行排序，随机分割成不同的序列，去除冗余数据，对数据进行归一化、采用最小最大标准化的方法将不同单位和量纲的数据归

一成统一的形式。

[0057] 步骤四:将工控网络数据分簇,计算簇的聚类中心,靠近聚类中心的数据向量认为是正确分类的,因此计算每一个数据向量与聚类中心的距离,得到训练集A,形成训练模型。

[0058] 步骤五:给定阈值ε,将满足阈值条件的数据集传送GA进行惩罚因子C和核函数参数g的优化,将满足参数优化停止准则的参数设定为SVM最优参数,SVM继续检测,得到检测模型。

[0059] 步骤六:建立工业控制入侵检测模型进行异常检测。

[0060] 特征提取是用wireshark抓取Modbus/TCP通信流量数据包,对需要提取的Modbus/TCP属性标号,查询所标号的属性的数据帧,计算数据帧头中的地址,提取标号属性的数据值。

[0061] 特征构造是主机发送正常请求时会读取功能码数据信息,攻击者可能会利用这一行为访问功能码数据,则利用功能码特征可能就无法判断是不是入侵行为。功能码03的功能是读取保持寄存器当前二进制值,若10秒内连续6次产生功能码03的请求信息,则不符合工业控制系统周期性的操作模式,则认为此请求为入侵行为。根据入侵模式,构造工业基本特征,10秒内功能码请求次数,20秒内访问地址次数,10秒内连接同一设备次数。

[0062] 所述数据预处理包括以下步骤:

[0063] 按时间顺序对提取和构造的数据包进行排序,随机分割成不同的序列。保证样本的代表性。

[0064] 去除冗余数据,对数据进行归一化、采用最小最大标准化的方法将不同单位和量纲的数据归一成统一的形式。

[0065] 对通信数据先利用FCM聚类,其步骤如下:

[0066] 对提取和构造出的工控通信流量数据进行FCM聚类,得到每个簇的聚类中心O,判定数据的隶属度和目标函数,所有的正常聚类中心标记为O+,所有的表示入侵的异常聚类中心标记为O-,正常集合标记为A+,异常集合标记为A-,设置类别标签。

[0067] 对于每个数据向量xi,计算与聚类中心的距离,若满足distance(xi,O+)<λ,则标记该数据向量xi∈A+,否则标记xi∈A-。

[0068] 重复上述步骤,直至数据集X中的每个数据向量标记入集合中。

[0069] 训练集A=A+∪A-。

[0070] 所述的GASVM算法检测,其步骤如下:

[0071] 设置GA算法在无法满足参数优化停止准则情况下的最大迭代次数和遗传次数。

[0072] 对不满足阈值条件(distance(xi,O+)大于等于λ)的数据集进行参数初始化,计算个体适应度,对其进行选择、交叉、变异操作,得到最优参数。

[0073] 设定惩罚因子C和核函数参数g的最优值。

[0074] 将FCM得到的类别标签赋予给支持向量。

[0075] 构造对偶问题和决策函数。

[0076] 基于FCM-GASVM算法建立入侵检测模型,其步骤如下:

[0077] 根据第五步介绍的FCM聚类步骤得到训练模型,根据第六步介绍的GASVM算法步骤得到检测模型。将两种算法进行结合,得到FCM-GASVM入侵检测模型的分类准确率。

[0078] 一种基于FCM-GASVM的工业控制系统入侵检测方法,包括以下步骤:

[0079] 特征提取:首先用wireshark抓取Modbus/TCP通信流量数据包,对于每一个Modbus TCP/IP协议均有多种属性,从中提取出最能反映数据特征的属性。

[0080] 特征构造:根据入侵模式,构造工业基本特征,10秒内功能码请求次数,20秒内访问地址次数,10秒内连接同一设备次数。

[0081] 数据预处理:按时间顺序对提取和构造的数据包进行排序,随机分割成不同的序列,去除冗余数据,对数据进行归一化、采用最小最大标准化的方法将不同单位和量纲的数据归一成统一的形式。

[0082] FCM聚类:将工控网络数据分簇,计算簇的聚类中心,靠近聚类中心的数据向量认为是正确分类的,因此计算每一个数据向量与聚类中心的距离,得到训练集A,形成训练模型。

[0083] GASVM:给定阈值ε,将满足阈值条件的数据集传送GA进行惩罚因子C和核函数参数g的优化,将满足参数优化停止准则的参数设定为SVM最优参数,SVM继续检测,得到检测模型。

[0084] FCM-GASVM:建立工业控制入侵检测模型进行异常检测。如图2所示。

[0085] 特征提取包括以下步骤:

[0086] 用wireshark抓取Modbus/TCP通信流量数据包,对需要提取的Modbus/TCP属性标号,查询所标号的属性的数据帧,计算数据帧头中的地址,提取标号属性的数据值。

[0087] 特征构造包括以下步骤:

[0088] 主机发送正常请求时会读取功能码数据信息,攻击者可能会利用这一行为访问功能码数据,则利用功能码特征可能就无法判断是不是入侵行为。

[0089] 功能码03的功能是读取保持寄存器当前二进制值,若10秒内连续6次产生功能码03的请求信息,则不符合工业控制系统周期性的操作模式,则认为此请求为入侵行为。

[0090] 根据入侵模式,构造工业基本特征,10秒内功能码请求次数,20秒内访问地址次数,10秒内连接同一设备次数。

[0091] 数据预处理包括以下步骤:

[0092] 按时间顺序对提取和构造的数据包进行排序,随机分割成不同的序列。保证样本的代表性。

[0093] 去除冗余数据,对数据进行归一化、采用最小最大标准化的方法将不同单位和量纲的数据归一成统一的形式。

[0094] $$v' = \frac{v - \min}{\max - \min}(\max' - \min') + \min'$$

[0095] Max和min代表的是数据集中的最大值、最小值;max'和min'代表的是映射新空间的区间(min',max');v为输入向量;v'为输出向量。

[0096] 所述FCM聚类包括以下步骤:

[0097] 对提取和构造出的工控通信流量数据进行FCM聚类,得到每个簇的聚类中心O,所有的正常聚类中心标记为O+,所有的异常聚类中心标记为O-,正常集合标记为A+,异常集合标记为A-,设置类别标签。

[0098] 对于每个数据向量xi,计算与聚类中心的距离,判定数据的隶属度和目标函数,若满足distance(xi,O+)<λ,则标记该数据向量xi∈A+,否则标记xi∈A-。

[0099] 重复上述步骤，直至数据集X中的每个数据向量标记入集合中。

[0100] 训练集A＝A+∪A-。

[0101] GASVM算法包括以下步骤：

[0102] 设置GA算法在无法满足参数优化停止准则情况下的最大迭代次数和遗传次数。

[0103] 对不满足阈值条件的数据集进行参数初始化，计算个体适应度，对其进行选择、交叉、变异操作，得到最优参数。

[0104] 设定惩罚因子C和核函数参数g的最优值。

[0105] 将FCM得到的类别标签赋予给支持向量。

[0106] 构造对偶问题和决策函数。

[0107] FCM-GASVM算法包括以下步骤：

[0108] 根据第五步介绍的FCM聚类步骤得到训练模型，根据第六步介绍的GASVM算法步骤得到检测模型。将两种算法进行结合，得到FCM-GASVM入侵检测模型的分类准确率。

[0109] 聚类中心为：

[0110] $$v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i=1}^{n} u_{ij}^{m}}$$

[0111] 隶属度为：

[0112] $$u_{ij} = \begin{cases} \left[\sum_{k=1}^{c} \frac{\left\| x_i - v_j \right\|^{\frac{2}{m-1}}}{\left\| x_i - v_k \right\|^{\frac{2}{m-1}}} \right]^{-1} \\ 1 \\ 0 \end{cases}$$

[0113] 目标函数为：

[0114] $$J = \sum_{i=1}^{n} \sum_{j=1}^{c} \left(u_{ij} \right)^{m} \left\| x_i - v_j \right\|$$

[0115] 其中，u_{ij}为个体x_i属于第j类的模糊隶属度；m为模糊权重指数；v_j为第j类的聚类中心；n为数据向量总个数；c为类别数；

[0116] 对偶问题为：

[0117] $$\min Q(a) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K\left(x_i, x_j \right) - \sum_{i=1}^{n} \alpha_i$$

[0118] $$s.t. \sum_{i=1}^{l} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le C, i = 1,\ldots,n$$

[0119] 其中，Q(α)表示对偶运算，α＝(α_1,α_2,...,α_n)表示拉格朗日算子，K(x_i,x_j)表示高斯径向基核函数，得解α*＝(α_1*,α_2*,...,α_n*)。l表示i＝1,…,n中目前取到的值。

[0120] 决策函数为：

[0121] $$b^* = y_j - \sum_{i=1}^{n} y_i \alpha_i^* K(x_i, x_j)$$

[0122] $$f(x)=\mathrm{sgn}(\omega\cdot\Phi(x)+b)=\mathrm{sgn}(\sum_{i=1}^{n}\alpha_i y_i K(x_i\cdot x)+b)$$

[0123] 其中，b^*是支持向量机的最终决策函数的阈值，sgn()表示符号函数。ω和b*一样表示支持向量机的最终决策函数的阈值，Φ(x)表示x的约束函数。b＝b*，yi∈R＝{-1,1}，y_j表示分类的标签，正常记为1，异常记为-1。

[0124] 如图1所示，基于FCM-GASVM的工控入侵检测方法，包括：

[0125] a.特征提取、构造和预处理部分

[0126] 1、Modbus/TCP应用数据单元主要包括Modbus应用协议报文头(MBAP)和协议数据单元(PDU)。MBAP包括事务处理标识码符、协议标识符、长度、单元标识符。PDU包括功能码和数据。

[0127] 2、首先用wireshark抓取Modbus/TCP通信流量数据包，对于每一个Modbus TCP/IP协议均有多种属性，从中提取出最能反映数据特征的属性。

[0128] 3、根据入侵模式，构造工业基本特征，10秒内功能码请求次数，20秒内访问地址次数，10秒内连接同一设备次数。

[0129] 4、按时间顺序对提取和构造的数据包进行排序，随机分割成不同的序列。保证样本的代表性。

[0130] 去除冗余数据，对数据进行归一化、采用最小最大标准化的方法将不同单位和量纲的数据归一成统一的形式。

[0131] $$v'=\frac{V-\min}{\max-\min}(\max'-\min')+\min'$$

[0132] Max和min代表的是数据集中的最大值、最小值；max'和min'代表的是映射新空间的区间(min',max')；v为输入向量；v'为输出向量。

[0133] b.训练模型

[0134] 1、对提取和构造出的工控通信流量数据进行FCM聚类，得到每个簇的聚类中心O，所有的正常聚类中心标记为O+，所有的异常聚类中心标记为O-，正常集合标记为A+，异常集合标记为A-，设置类别标签。聚类中心为：

[0135] $$v_j=\frac{\sum_{i=1}^{n}u_{ij}^{m}x_i}{\sum_{i=1}^{n}u_{ij}^{m}}$$

[0136] 其中，u_{ij}为个体x_i属于第j类的模糊隶属度。

[0137] 2、对于每个数据向量xi，计算与聚类中心的距离，判定数据的隶属度和目标函数，若满足distance(xi,O+)<λ，则标记该数据向量xi∈A+，否则标记xi∈A-。隶属度和目标函数公式如下：

[0138] $$u_{ij}=\begin{cases}\left[\sum_{k=1}^{c}\frac{\|x_i-v_j\|^{\frac{2}{m-1}}}{\|x_i-v_k\|^{\frac{2}{m-1}}}\right]^{-1}\\ 1\\ 0\end{cases}$$

[0139] $$J=\sum_{i=1}^{n}\sum_{j=1}^{c}\left(u_{ij}\right)^{m}\left\|x_i-v_j\right\|$$

[0140] 其中，u_{ij}为个体x_i属于第j类的模糊隶属度；m为模糊权重指数；v_j为第j类的聚类中心；n为数据向量总个数；c为类别数；

[0141] 3、重复上述步骤，直至数据集X中的每个数据向量标记入集合中。训练集A＝A+∪A-。

[0142] c.检测模型：

[0143] 1对不满足阈值条件的数据集进行参数初始化，计算个体适应度，对其进行选择、交叉、变异操作，得到最优参数。

[0144] 2设定惩罚因子C和核函数参数g的最优值。

[0145] 3将FCM得到的类别标签赋予给支持向量。

[0146] 4构造对偶问题和决策函数

[0147] $$\min Q(a)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K\left(x_i,\ x_j\right)-\sum_{i=1}^{n}\alpha_i$$

[0148] $$s.t.\sum_{i=1}^{l}\alpha_i y_i=0,\qquad 0\le\alpha_i\le C,i=1,\ldots,n$$

[0149] 其中，Q(α)表示对偶运算，α＝($\alpha_1,\alpha_2,\ldots,\alpha_n$)表示拉格朗日算子，K($x_i,x_j$)表示高斯径向基核函数，得解$\alpha^*$＝($\alpha_1^*,\alpha_2^*,\ldots,\alpha_n^*$)。l表示i＝1，…，n中目前取到的值。

[0150] 决策函数为：

[0151] $$b^*=y_j-\sum_{i=1}^{n}y_i\alpha_i^*K(x_i,x_j)$$

[0152] $$f(x)=\mathrm{sgn}(\omega\cdot\Phi(x)+b)=\mathrm{sgn}(\sum_{i=1}^{n}\alpha_i y_i K(x_i\cdot x)+b)$$

[0153] 其中，b^*是支持向量机的最终决策函数的阈值，sgn()表示符号函数。

[0154] 5根据FCM得到训练模型，GASVM得到检测模型。计算FCM-GASVM入侵检测模型的分类准确率。

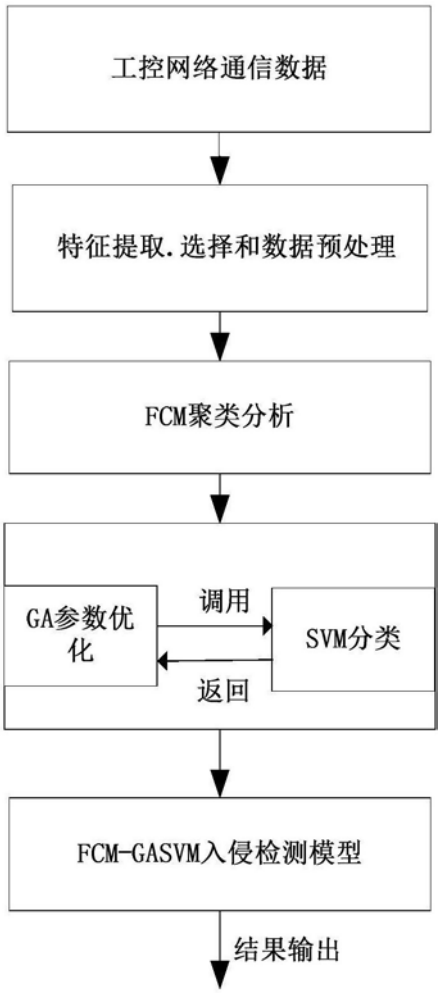

工控网络通信数据
特征提取.选择和数据预处理
FCM聚类分析
GA参数优化
调用
SVM分类
返回
FCM-GASVM入侵检测模型
结果输出

图1

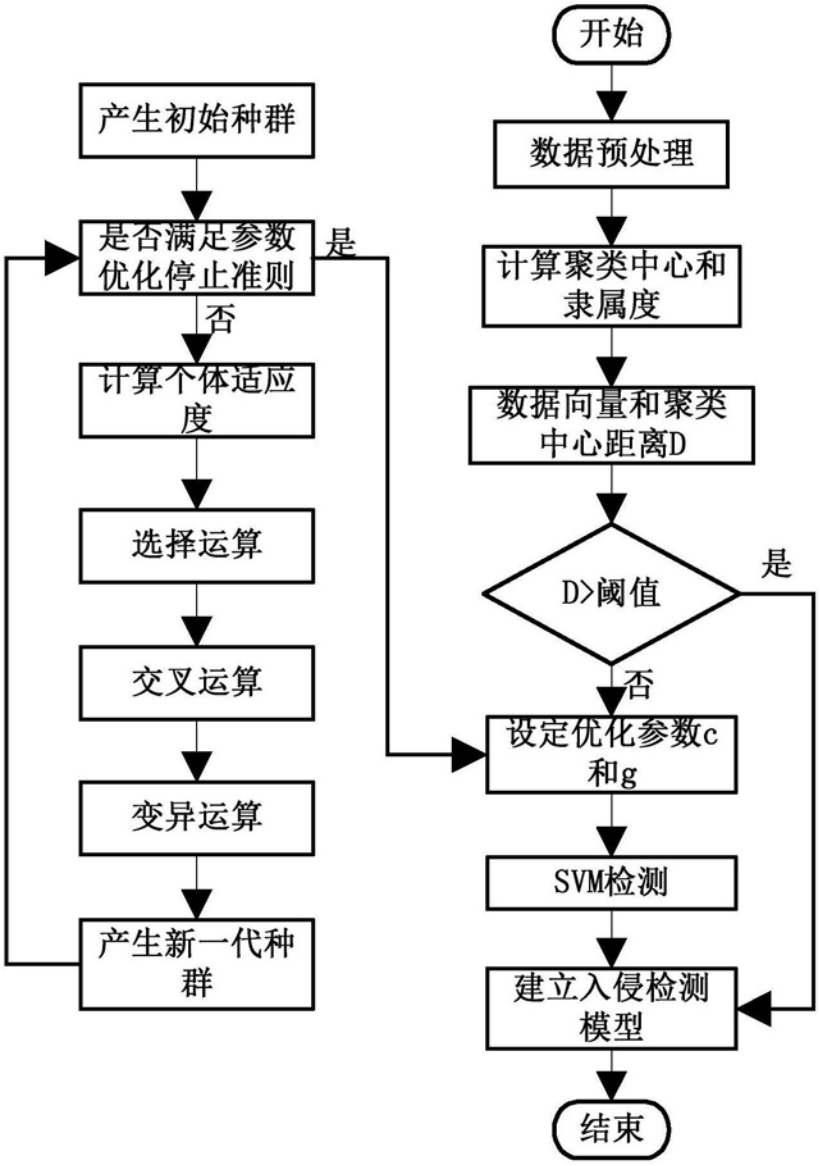

开始
数据预处理
计算聚类中心和隶属度
数据向量和聚类中心距离D
D>阈值
是
否
设定优化参数c和g
SVM检测
建立入侵检测模型
结束
产生初始种群
是否满足参数优化停止准则
是
否
计算个体适应度
选择运算
交叉运算
变异运算
产生新一代种群

图2