文章编号：2095-2163(2019)03-0319-03 中图分类号：TP393.08 文献标志码：A

基于 RS-SVM 的无线传感器网络入侵检测模型研究

张志霞
(广东工商职业学院(本科)，广东 肇庆 526020)

摘 要：本文对于无线传感器网络安全进行了分析，阐述了基于 RS-SVM 的无线传感器网络入侵检测模型所涉及到的理论及技术，即粗糙集理论、支持向量机和狼群算法，最后对模型进行了整体阐述，总结了仿真实验的结果，以期为无线传感器网络的入侵检测提供实用模型。

关键词：粗糙集；支持向量机；网络入侵检测

Research on intrusion detection model of wireless sensor networks based on RS-SVM

ZHANG Zhixia
(Guangdong College of Business and Technology(undergraduate), Zhaoqing Guangdong 526020, China)

【Abstract】 In this paper, the security of Wireless Sensor Networks is analyzed, and the theory and technology involved in the intrusion detection model of Wireless Sensor Networks based on RS-SVM, namely Rough Set theory, Support Vector Machine and wolf swarm algorithm, are elaborated. Based on the above, the model is described as a whole, and the results of simulation experiments are summarized in order to provide a practical model for intrusion detection of wireless sensor networks.

【Key words】 Rough Set; Support Vector Machine; network intrusion detection

1 无线传感器网络安全性分析

无线传感器网络(Wireless Sensor Networks, WSN)，即由众多无线传感器组成的网络。无线传感器网络在安全性上存在诸多问题，是因为网络中的节点采用无线通信的方式来传输信息，因此节点间的信息传输不能和有线通道一样防护严密。无线传感器网络主要面临着 5 方面的安全问题。对此可做分述如下。

(1)安全机制缺失问题。现在常用的安全处理机制都是针对某一方面，且往往只能解决一方面的问题，而无线传感器网络遭遇的攻击却会多种多样，所以无线传感器网络尚未配备系统性的安全机制。

(2)节点能量限制问题。由于传感器经常被部署在较难实施监控的地方，能够给节点提供动力的电池能量有限。

(3)节点随机组织问题。节点的随机组织有可能导致安全防护滞后。

(4)节点物理安全问题。

(5)通信不稳定问题。

粗糙集(Rough Set)是一种数据挖掘方法，可以发现数据之间隐含的特征关系，将粗糙集引入到无线传感器网络入侵检测中，可以将网络特征进行约简，减少网络分类器输入的向量数。同时，利用改进的支持向量机建立网络分类器，实践表明，本文所采用的方法可以有效地增强无线传感器网络的抗攻击能力。

2 RS-SVM 理论简介

2.1 粗糙集理论

粗糙集理论认为知识是一种分类能力，而人们的各种行为都是基于分辨对象的能力而发生的。将知识理解为划分数据，划分得到的每一集合则称为概念。那些根据事务的特征差别将其分门别类的能力都可以看作是某种“知识”。论域中，相互间不可分辨的对象组成的集合是组成知识的颗粒(granule)。知识是有粒度的，粒度越小，能精确表达的概念越多。粒度的形式表示：不可分辨关系/等价类，粒度是知识的最小单位。粗糙集理论的主要思想是利用已有的知识库来辨别不确定的知识，或者利用已有的知识来替代其他的某些相似的知识。

设 R 是一个等价的关系族，若：

基金项目：肇庆市科技创新指导类项目(201604030902)。

作者简介：张志霞(1979-)，女，硕士，计算机讲师，主要研究方向：计算机科学与技术、计算机应用。

收稿日期：2019-02-27

$$IND(R) = IND(R - \{R\}), \quad (1)$$

那么称 R 为关系族中可省的,否则即为不可省的。

设 R 中任意一个等价关系都是不可省的,此时 R 为独立,相反则称为依赖。

此处令 $Q \subseteq P$,若 Q 是独立的,且 $IND(Q) = IND(P)$,则称 Q 是等价关系族 P 的一个约简。

P 中所有不可省关系的集合记作 $CORE(P)$,P 可以有多个约简,以 $RED(P)$ 表示 P 的所有约简集合,则有:

$$CORE(P) = \cap\ RED(P). \quad (2)$$

设 $S \subseteq P$,称 S 是 P 的 Q 约简,当且仅当 S 是 P 的 Q 独立的,且有 $POS_p(Q) = POS_s(Q)$,当满足此约简表时说明决策表的约简是成功的。

2.2 支持向量机理论

支持向量机的核心是寻找一个超平面,超平面能够将尽可能多的点分割开来。解决方法就是构造一个在约束条件下的优化问题。具体地说,就是一个约束二次规划问题,求解该问题,得到分类器。

假设有线性可分样本集 $\{(x_i, d_i)\}_{i=1}^{M}$,其中 M 表示输入样本的数目,则对于线性可分模式所构造的最优分类超平面公式为:

$$\boldsymbol{w}^{\mathrm{T}}x + b = 0, \quad (3)$$

其中,$\boldsymbol{w}$ 为可调的权值向量,b 为偏值。

支持向量的样本点需要满足的条件如下:

$$\min(\boldsymbol{w}) = \min \frac{1}{2} \| \boldsymbol{w} \|^2 = \min(\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}), \quad (4)$$

为了处理不可分离数据点和数据噪音,引入松弛变量 $\{\varepsilon_i\}_{i=1}^{M}$,则有:

$$d_i(\boldsymbol{w}^{\mathrm{T}}x_i + b) \geqslant 1 - \varepsilon_i, \quad (5)$$

为了提高泛化能力和达到结构风险最小化,研究推得计算公式为:

$$\phi(\boldsymbol{w}, \varepsilon) = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + c\sum_{i=1}^{M}\varepsilon_i, \quad (6)$$

为了解决约束最优问题,引进拉格朗日因子,最终可求得决策函数为:

$$f(x) = \operatorname{sgn}\{(\boldsymbol{w} \cdot x) + b\} = \operatorname{sgn}\{\sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i d_i k(x_i, x_j) + b\}. \quad (7)$$

支持向量机适用于二分类问题,而无线传感器网络的入侵检测是一个多分类问题,利用二叉树与森林的转换原理,可以将多分类问题变换为二分类问题进行处理,其基础原理如图 1 所示。

2.3 狼群算法

狼群算法是 2013 年提出的一种新型仿生算法,因为具有控制参数少,搜索路径优,全局优化能力强等优点,在很多优化问题中都取得了良好的应用效果。

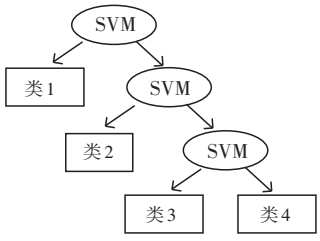

图 1 多分类问题转变为二分类问题原理图

Fig. 1 The schematic of multi-classification problem transforming into a two-classification problem

假设狼群的空间是一个 $N \times D$ 的欧几里得空间,其中 N 表示人工狼的数量,D 为待寻优的变量数。某一人工狼 i 的状态可表示为:

$$X_i = (X_{i_1}, X_{i_2}, \ldots, X_{i_D}). \quad (8)$$

其中,X_{i_d} 为第 i 匹人工狼在欲寻优的第 d 维变量空间中所处位置;人工狼所感知到的猎物气味浓度可表示为 $Y = f(x)$,而 Y 就是目标函数值;人工狼 p 和 q 之间的距离定义为其状态向量间的 Manhatan 距离。

3 RS-SVM 的无线传感器网络入侵检测模型

基于 RS-SVM 的无线传感器网络入侵检测模型的核心思想为:先利用粗糙集理论对于收集到的网络状态信息进行约简,以便去除掉冗余信息,再将约简后得到的网络状态信息向量作为输入信息输入到无线传感器网络入侵分类器中,输出网络状态。这里,给出了该模型的设计原理如图 2 所示。

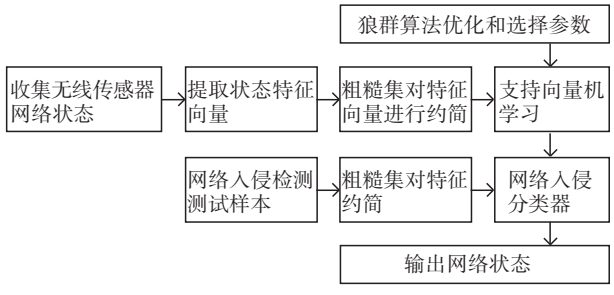

图 2 入侵检测模型工作原理

Fig. 2 The schematic of intrusion detection model

通过利用狼群算法进行入侵检测分类器的参数优化,一方面可以降低误警率,使得检测结果更加可靠,另一方面可以提高网络入侵检测率,加快网络入侵检测的速度,降低了算法的复杂程度,使得算法的运行效率更高,从而消耗更少的能量。

(下转第 323 页)