计算机工程与应用 2009 45 (36): 99-100 ISSN: 1002-8331 CN: 11-2127/TP

网络、通信、安全

基于集对分析的半监督ISODATA聚类

魏小涛

北京交通大学 软件学院，北京 100044

收稿日期 2009-3-31 修回日期 2009-10-23 网络版发布日期 2009-12-30 接受日期

摘要 提出一个基于集对分析的半监督ISODATA聚类算法，用于网络异常检测。在三方面进行了改进：首先，算法能够直接处理字符数字混合属性的数据，并使用集对分析来计算数据记录之间的距离；其次，算法同时处理有标号和无标号的数据，并利用少量的有标号数据来指导算法的分裂过程；最后，将算法的输入参数减少到只有两个。在KDD99入侵检测数据集上的实验结果显示，该算法获得了95.62%的检测率和1.29%的误报率。

关键词 集对分析 网络异常检测 半监督聚类 迭代自组织数据分析方法（ISODATA）

分类号 TP18

Semi-supervised ISODATA clustering based on set pair analysis

WEI Xiao-tao

Software School，Beijing Jiaotong University，Beijing 100044，China

Abstract

A semi-supervised ISODATA clustering algorithm based on the Set Pair Analysis（SPA） is proposed for network anomaly detection.This paper improves the original ISODATA algorithm mainly in three aspects.Firstly，the modified algorithm can directly process the mixed attributes of symbolic and numeric values，and employ the SPA to calculate the distance between data records.Secondly，the algorithm can process both labeled and unlabeled samples.The small portion of labeled samples is used to supervise the clustering process in the splitting stage.Thirdly，the initial parameters needed to be input into the algorithm are reduced to only two.Experimental result on the KDD 99 intrusion detection datasets shows that the algorithm has high detection rate（95.62%） while maintaining a low false positive rate（1.29%）.

Key words set pair analysis network anomaly detection semi-supervised clustering Iterative Self-Organizing Data Analysis Technique（ISODATA）

DOI: 10.3778/j.issn.1002-8331.2009.36.029

通讯作者 魏小涛 weixt@bjtu.edu.cn