TIVis：基于公司监控数据的威胁情报分析系统

徐劭斌，任珂，张慧杰

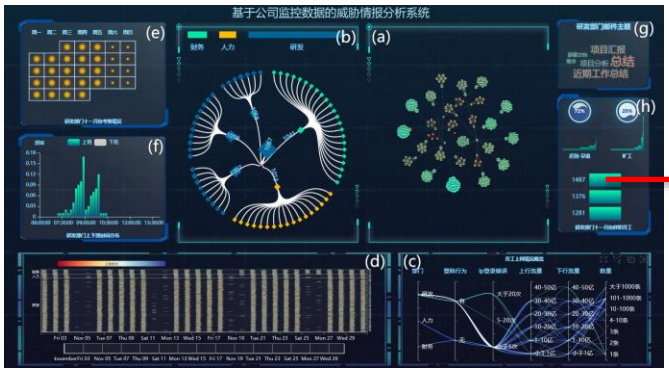

图 1. TIVis 可视分析系统基础模块

图 2. TIVis 可视分析系统个人信息模块

摘要：本文针对 ChinaVis 2018 的 HighTech 公司监控数据，利用随机森林算法对员工进行部分划分，同时设计并实现了一个支持分别从部门、员工角度探索的交互式可视分析系统 TIVis。系统支持工作时间、考勤情况、登陆日志、工作内容等方面探索各部门的正常工作模式，同时可以帮助用户从多角度提取单个员工工作模式的时变规律，及时发现具有威胁的异常行为。

关键词：组织关系、日常行为、威胁情报、可视分析

1 引言

本文将从四个方面介绍我们的可视化系统：第一部分介绍数据的清洗及预处理，第二部分详细描述可视化设计及分析策略，第三部分从多角度来评价我们的可视分析系统，第四部分将对本文主要内容进行总结。

2 数据清洗及预处理

2.1 数据清洗

首先我们将 30 天数据按表导入 mysql 数据库，基于邮件主题长度清理垃圾邮件，最后根据邮件数据将用户的 id 和 ip 对应，建立用户表。

2.2 员工部门划分

邮件主题反映了各员工的工作范围，因此它成为判别员工所属部门的重要特征。基于随机森林算法，我们选择出现频数最大的 90 个邮件主题作为特征，为每名员工构建特征向量。随机抽取 40 名员工，人工添加标签构成训练集，并进行随机森林训练，对剩下的员工进行部门划分。分类结果为财务部门有 24 名员工，人力资源部门有 18 员工，研发部门有 257 名员工。

3 可视化设计

基于挑战赛提出的分析任务，我们设计的 TIVis 可视分析系统由基础模块（图 1）和可弹出的个人信息模块（图 2）组成。基础模块用于探究部门（或小团体）内部结构和正常的工作模式。个人信息模块用于查看感兴趣员工的详细行为，帮助用户对异常事件进行细节探索。

3.1 节点连接图

根据员工间收发邮件的关系，我们构建各个部门的关系网络图（图1-a）。图中每个节点代表一名员工，节点的半径代表员工在部门内部接受和发送的邮件总数，边的粗细编码两名员工邮件往来数量。另外，我们使用信息熵衡量每个员工自我中心网络的混乱程度，即信息熵越大，该员工有更多的联系伙伴；反之，该员工只与个别人员关系紧密。我们使用信息熵编码节点的颜色，从绿到黄映射信息熵从小到大。由于领导管理整个部门，与所有员工均有大量收发邮件行为，所以信息熵比普通员工更大。通过探索该图我们可以确定财务部，人力资源部的领导和研发部的多级层次结构。

- 徐劭斌，东北师范大学，E-mail：2233935216@qq.com
- 任珂，东北师范大学，E-mail：renk205@nenu.edu.cn
- 张慧杰，东北师范大学，E-mail：zhanghj167@nenu.edu.cn

3.2 公司组织关系图

基于节点连接图发现的部门内部结构，我们使用树形图（图1-b）展示整个公司的组织关系，包括三个部门和299名员工。每个节点代表一名员工，节点的颜色标识员工的所属部门，节点的大小编码下属职员的数量。该图提供展开和收拢操作，当鼠标悬停某节点时系统显示其在公司所处的位置。

3.3 平行坐标图

为了分析不同部门登录日志和 TCPLOG 日志的信息间差异，我们设计平行坐标图进行多维信息展示。而传统的平行坐标在处理大规模数据时力不从心，存在严重的遮挡问题。为了解决这一缺陷，我们对 299 名员工 30 天共计 8970 条记录进行分段处理，并增加一个字段用于统计该记录重复的数量。如图 1-c 所示，前 5 个轴代表登录日志和 TCPLOG 日志的 5 个变量，最后一个轴代表这种记录的数量。系统支持对各轴进行刷取操作，并交互展示刷取记录的未分段原始数据。

3.4 时序热力图

我们用时序热力图（图 1-d）提供公司全体员工的 30 天的上下班的区间段的概览。每个员工一行，横轴为时间，若员工上班则为这个时间段填色，颜色映射该天的上班时长。系统提供刷取时间、员工的交叉筛选操作，方便用户放大视图探索细节信息。

3.5 基础模块其他视图

TIVis 可视分析系统基础模块中还包含多种其他视图。日历图（图 1-e）展示部门每天的上班人数；柱形图（图 1-f）用于展示上下班时间段的分布情况；文字云（图 1-g）展示部门常见邮件主题；水波柱形图（图 1-h）展示部门的考勤情况。

3.6 个人信息模块视图

用户可以点击节点连接图中节点、水波柱形图中柱形调出该员工的个人信息模块。该模块包括：阶梯瀑布图（图 2-a）展示该员工本月上下班的情况；文字云（图 2-b）展示该员工收发邮件的主题；流量图（图 2-c）展示该员工不同时刻的上下行流量数据；极坐标堆叠柱状图（图 2-d）展示多个协议登录成功和失败的数量；另外图 2-e 展示用户浏览网页的记录。我们可以利用个人信息模块视图总结该员工一个月内的总体工作模式，也可聚焦某一天探索用户的具体行为，从而发现对公司造成威胁的异常员工。

4 讨论

(1) 实用性：我们的可视分析系统由基础模块视图和个人信息模块视图两个层级构成，可以让用户对系统的架构有清晰的了解，很好解决了因为视图杂乱导致易用性降低的问题。

(2) 可交互性：TIVis 可视分析系统提供丰富的交互操作。在基础视图中，节点连接图的缩放功使用户能够更好地分析部门的组织结构；鼠标筛选感兴趣的团体后，上下班分布的柱状图、平行坐标图和时序热力图可进行多图联动，帮助用户分析部门内部不同团体工作模式的差异。在个人信息模块视图中，我们也提供时间选择、上下行流量缩放等交互功能，帮助用户分析威胁情报。

(3) 新颖性：我们利用信息熵衡量每个员工关系网络的混乱程度，并在节点连接图中使用节点颜色编码这一信息，使用户能够快速准确地发现部门领导，从而探究部门的组织结构。在平行坐标图中，我们增加重复记录的数量这一字段，突破了传统平行坐标对于展示大规模数据的局限性，同时用户可通过刷取数量极高或极低的记录发现常规模式和异常现象。

(4) 可扩展性：我们的可视化方案不仅能解决挑战一的所有问题，同时还具有良好的可扩展性，可用于其他具有多维、时序特征的数据集。同时，我们的系统是模块化编程，用户可以增加和修改 TIVis 可视分析系统模块，添加更多实用的交互，发现更深层的隐藏信息，以满足实际需求。

5 结论

本文针对挑战赛题目一的内容，设计了全面、直观的可视分析系统 TIVis。通过基础模块和个人信息模块两部分视图的联动分析，我们提取了公司的组织结构，总结了三个部门的正常工作模式，同时找到以 1487 员工为核心的三人团体窃取公司机密的异常事件。TIVis 能够对公司威胁情报分析提供强有力的帮助，具有一定的实用价值。