# 数据挖掘技术在网络病毒防范中的应用

2019329621004 梅雨欣 计算机科学与技术19(1)

**摘要：**数据挖掘技术通过分析每个数据，从大量数据中寻找其规律的技术，属于一种功能比较全面的挖掘技术和数据应用技术。为保障信息安全、构建更加完善的计算机网络病毒防御体系，需要分析计算机网络病毒防范中数据挖掘技术的应用情况。

**关键词**：计算机网络；数据挖掘技术；网络病毒；网络安全

Research艾瑞市场咨询根据公安部公共信息网络安全监察局统计数据显示：2006年中国（大陆）病毒造成的主要危害情况：“浏览器配置被修改”是用户提及率最高的选项，达20.9%。病毒造成的影响还表现为“数据受损或丢失”18%，“系统使用受限”16.1%，“密码被盗”13.1%。另外，“受到病毒非法远程控制”提及率为6.1%，“无影响”的只有4.2%。更为严重的是，传统的病毒、垃圾邮件还在出没，危害更大的间谍软件、“网络钓鱼”等又不断出现，网络信息安全形势愈加严峻。

为保障网络信息安全，必须采用相关病毒防范手段，其中数据挖掘技术可以对计算机网络中数据进行挖掘、处理和分析，进而为之后更加完善的网络病毒防御体系的构建做好理论数据基础。

## **一、数据挖掘技术的系统模型**

数据挖掘是指从大量的数据中通过算法搜索隐藏于其中信息的过程，如使用数据分类、估值、聚类等方式寻找病毒数据之间存在的具体规律。其主要有数据准备、规律寻找和规律表示三个步骤。规律寻找是用某种方法将数据集所含的规律找出；规律表示是尽可能以用户可理解的方式将找出的规律表示出来。数据挖掘的任务有关联分析、聚类分析、分类分析、异常分析、特异群组分析和演变分析等。网络病毒防御系统中的数据挖掘技术主要包含四个模块：数据源模块、规则库模块、数据挖掘模块、预处理模块。

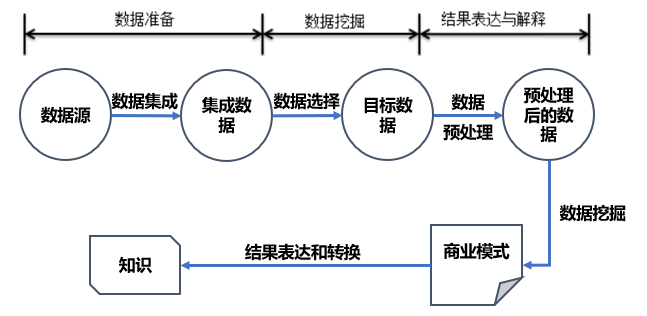


图1数据挖掘过程

数据挖掘过程模型步骤主要包括定义问题、建立数据挖掘库、分析数据、准备数据、建立模型、评价模型和实施：

1. 定义问题

在开始知识发现之前最先的也是最重要的要求就是了解数据和业务问题。必须要对目标有一个清晰明确的定义，即决定到底想干什么。

1. 建立数据挖掘库

建立数据挖掘库包括以下几个步骤：数据收集，数据描述，选择，数据质量评估和数据清理，合并与整合，构建元数据，加载数据挖掘库，维护数据挖掘库。

1. 分析数据

分析的目的是找到对预测输出影响最大的数据字段，和决定是否需要定义导出字段。

1. 准备数据

这是建立模型之前的最后一步数据准备工作。可以把此步骤分为四个部分：选择变量，选择记录，创建新变量，转换变量。

1. 建立模型

建立模型是一个反复的过程。先用一部分数据建立模型，然后再用剩下的数据来测试和验证这个得到的模型。有时还有第三个数据集，称为验证集，因为测试集可能受模型的特性的影响，这时需要一个独立的数据集来验证模型的准确性。训练和测试数据挖掘模型需要把数据至少分成两个部分，一个用于模型训练，另一个用于模型测试。

1. 评价模型

模型建立好之后，必须评价得到的结果、解释模型的价值。从测试集中得到的准确率只对用于建立模型的数据有意义。在实际应用中，需要进一步了解错误的类型和由此带来的相关费用的多少。

1. 实施

模型建立并经验证之后，可以有两种主要的使用方法。第一种是提供给分析人员做参考；另一种是把此模型应用到不同的数据集上。

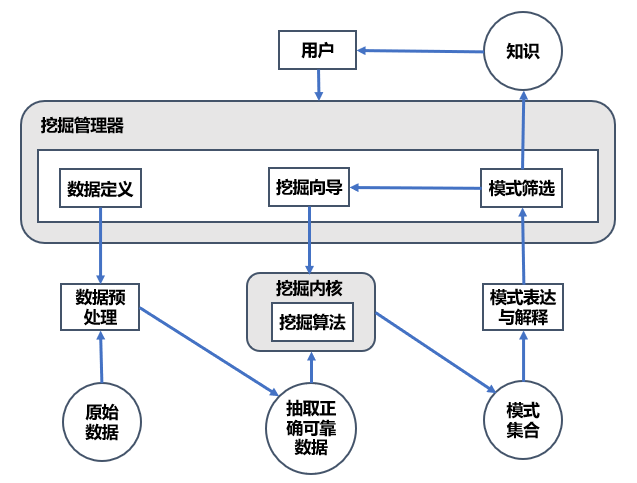


图2数据挖掘技术系统模型

## **二、病毒防范中数据挖掘技术的应用**

在使用数据挖掘技术的过程中，应着重考虑数据的收集、处理、分析和相应的模型匹配问题，通过具体的、有效的操作实现对计算机网络病毒的防范。

数据收集数据挖掘技术的应用本身和计算机技术的应用是相互匹配的，要基于计算机技术建立数据挖掘过程。为保证数据的安全性，加强对用户隐私的保护显得至关重要。通过对数据挖掘技术的应用，能够完成数据信息的快速收集、分析以及处理，同时也包含了数据信息的归类以及数据信息的筛选等对应的技术。

### 2.1用于网络病毒防御体系的可能性

进一步了解和分析网络病毒的传播过程可以为数据挖掘的应用提供依据。如果蠕虫病毒感染计算机，首先需要扫描网络上的主机。这个过程是检测蠕虫病毒的关键，也是建设计算机预防系统的突破口。以数据挖掘技术为基础，构建全新的网络病毒防御系统。全新的网络病毒防御系统是由数据源模块、预处理模块、数据挖掘模块、规则库模块、决策模块和防御模块等组成。其工作原理是:来自网络的、发向本地的数据包形成数据源之后，由预处理模块处理，并且记录网络信息传输的病毒，以便对后期同等性质的病毒形成免疫力。一旦这些病毒再次非法入侵，系统就会及时警报，并且启动相应的保护防御程序。数据挖掘结构示意图如下。

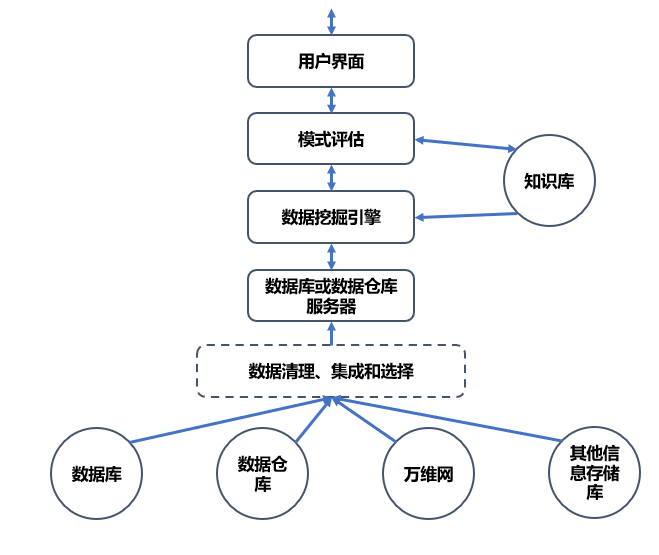


图3数据挖掘结构示意图

### 2.2统计学与模型匹配

数据处理和分析在数据收集完成之后，基于统计学的方法来初步的进行处理，分析阶段，则主要是强调对于关联挖掘法的实际应用。如在某个区域多发扫描侵袭的问题，并且其比例超出60%，这样就可以了解病毒传播主要是利用点-面的方式来进行的，同时在对病毒关联性进行深入挖掘，可以发现网络病毒的发源地主要集中在公共计算机领域，其占据的比例达到80%。因此，建议对公共计算机领域，例如，网吧、企业等领域的计算机做大规模的扫描，针对可能感染的移动设备，也需要将分析和数据处理结果作为基本依据，通过针对性的控制扫描传播，最终满足计算机网络病毒的防范效果，关于数据处理分析以及其防御过程，有利于针对未知攻击模块进行异常分析，从而明确未知攻击类型，并采取有效的防范措施。

模型匹配对于模型的匹配其本身属于一项具有较强针对性的网络安全处理方法，主要体现在：通过分析之后，发现某个区域存在特异性的木马病毒，将会实施模型匹配，并进行针对性处理。同时，可以对不同地区的计算机网络安全事件进行收集，然后将其中多件的网络病毒进行汇总与分析。通过分析之后发现，大约有70%左右的木马病毒是通过安包装方式进行传播，还有一些广告插件中也会带有隐藏的病毒，而大小通常在5MB之内。从后续的工作来看，侧重于带有类似特征的数据包，将其输入到默认的计算机管理软件当中，可以形成防火墙，在此基础上，可以对网络病毒进行有效拦截，并且提出相关的人员做好对应的处理，从而确保网络的安全性。

### 2.3网络病毒防控中的应用角度

基于数据挖掘技术形成的计算机网络病毒防御体系，是全新的计算机网络病毒防御方式，它能够更加全面、有效地处理计算机网络信息的安全问题。具体来讲，可以从以下几个角度探析其在计算机网络病毒防御过程中的应用：

(1)数据源模块。该模块是以抓包程序为基础，截获网络向主机发送的数据包信息。数据源模块中有最原始的数据包和数据结构，在获取这样的数据源之后，要将其交给预处理模块进行下一步操作。

(2)数据预处理模块。这是数据挖掘技术的基础性工作，会为后期数据分析打下良好的基础。它不仅关系到数据挖掘成效，还涉及到数据挖掘时间。

(3)规则库模块。其效能主要是存储检测到的病毒特征，并将其积累起来形成规则集，反映病毒的特点和连接数据的规则，为后期病毒数据特点分析和收集提供依据。

(4)数据挖掘模块。借助科学的算法分析事件库，并且在此基础上生成请求记录，完成此项工作后将其交给决策模块。

(5)决策模块。此模块的效能是分析数据挖掘结果和规则库规则的匹配度。如果发现两者之间的匹配度较高，则说明数据包中可能有病毒；如果发现两者之间的匹配度很低，则在此基础上发动预警机制，找到新的蠕虫病毒，并将其纳入规则库中，实现规则库扩展。

### 2.4基于数据挖掘技术的病毒防御系统的分析方式

基于数据挖掘技术的病毒防御系统会采取多样化的分析方式，主要如下：

(1)分类分析。将各个主体提前设定到几种类别中，以统计方法或者机械方法建立分类模型，并在数据库中实现特定类的数据映射，由此实现分类处理和分析。

(2)聚类分析。指分解得到的数据包，将其归纳总结到不一样的组别中，而处于同组别中的往往都有相似点，不同组别代表不同特点的事物。以聚类分析的方式处理数据，可找到数据分布的疏密情况，并在此基础上形成全局分布模式，进而展现数据之间的关系。

(3)异类分析。异类分析又被称为孤立点分析，指分析数据库中明显偏离其他数据、明显具有不同点的数据，特指偏离了一般模式的数据。异类分析包括孤立点的发现和孤立点的分析，孤立点的发现常会产生有违常理的结果。在分析孤立点的过程中，则可能会发现比一般数据更有价值的数据。

(4)序列分析。统计动态化的数据，研究随机数据序列归纳的情况，从而找到相应的病毒数据序列。一般情况下，序列分析的目的是最大限度地获取序列模式模型，以找到事件发生的时间序列。

## **三、总结**

互联网在人们生产生活中所应用的领域越来越广，涉及到金融、财产及人际关系的内容越来越多，同时，数据挖掘技术在计算机网络安全防御系统中的地位越来越重要。数据挖掘技术基于大数据模块更多的运用在计算机网络诈骗防护、危机评估等。数据挖掘技术充分反应了当前防病毒工具的现进性，能够准确预测和检测病毒的入侵形式和数量。在对计算机网络病毒防范中，数据挖掘技术扮演着“守护者”的重要角色。因此，相关技术人员对数据挖掘技术有一个清楚的认识，在具体的操作之中，能够通过各种有效的途径，如数据处理、模型匹配等，实现对计算机病毒的有效防范，确保计算机使用的安全性。

**参考文献：**

[1]陈春.基于数据挖掘技术的计算机网络病毒防御分析[J].信息通信,2015(05):121-122.

[2]石燕.数据挖掘技术在计算机网络病毒防御中的应用[J].中国高新科技,2020(17):55-56.

[3]李亮超.数据挖掘在计算机网络安全领域的应用[J].科技与创新,2015(21):109-110.

[4]唐永军.数据挖掘在计算机网络病毒防御中的应用研究[J].科技创新与生产力,2019(07):61-63.

[5]王海军.计算机网络安全病毒防御中的数据挖掘技术应用探析[J].信息与电脑(理论版),2019(12):229-230.

[6]刘春娟.数据挖掘技术在计算机网络病毒防御中的应用分析[J].电子测试,2014(05).

[7]王木.数据挖掘技术在网络安全态势分析中的应用[J].计算机产品与流通,2020(07):43.

[8] 百度文库.数据挖掘（计算机科学）[EB/OL]. <https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E6%8C%96%E6%8E%98/216477>