

# 文本复制检测报告单(全文对照)

№:ADBD2017R\_2017042800163820170512185652401191430881

检测时间:2017-05-12 18:56:52

检测文献: 11491915\_刘力铭

作者: 刘力铭

检测范围: 中国学术期刊网络出版总库

中国博士学位论文全文数据库/中国优秀硕士学位论文全文数据库

中国重要会议论文全文数据库

中国重要报纸全文数据库

中国专利全文数据库

互联网资源(包含贴吧等论坛资源)

英文数据库(涵盖期刊、博硕、会议的英文数据以及德国Springer、英国Taylor&Francis 期刊数据库等)

港澳台学术文献库

优先出版文献库

互联网文档资源

图书资源

CNKI大成编客-原创作品库

学术论文联合比对库

个人比对库

时间范围: 1900-01-01至2017-05-12

## 检测结果

总文字复制比: **17.5%**

跨语言检测结果: **0%**

去除引用文献复制比: **6.7%**

去除本人已发表文献复制比: **17.5%**

单篇最大文字复制比: **10.7%**

重复字数: [8952]

总段落数: [5]

总字数: [51079]

疑似段落数: [5]

单篇最大重复字数: [5461]

前部重合字数: [1485]

疑似段落最大重合字数: [3400]

后部重合字数: [7467]

疑似段落最小重合字数: [222]



指标: ☐ 疑似剽窃观点 ☒ 疑似剽窃文字表述 ☐ 疑似自我剽窃 ☐ 疑似整体剽窃 ☐ 过度引用

表格: 1 脚注与尾注: 0

14.9% ( 1650 ) 11491915\_刘力铭\_第1部分 ( 总11104字 )

20.5% ( 1979 ) 11491915\_刘力铭\_第2部分 ( 总9639字 )

37% ( 3400 ) 11491915\_刘力铭\_第3部分 ( 总9178字 )

2.4% ( 222 ) 11491915\_刘力铭\_第4部分 ( 总9378字 )

14.4% ( 1701 ) 11491915\_刘力铭\_第5部分 ( 总11780字 )

( 注释: ■ 无问题部分 ■ 文字复制比部分 ■ 引用部分 )

## 1. 11491915\_刘力铭\_第1部分

总字数: 11104

相似文献列表 文字复制比: 14.9%(1650) 疑似剽窃观点: (0)

|   |  |                          |
|---|--|--------------------------|
| 1 | 恶意网页医疗数据应急通信移动通信<br>- 《学术论文联合比对库》- 2015-01-09                                | 9.8% ( 1088 )<br>是否引证: 否 |
| 2 | 王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15<br>王维光 - 《学术论文联合比对库》- 2015-01-15 | 9.5% ( 1056 )<br>是否引证: 否 |
| 3 | 基于分类算法的恶意网页检测技术研究<br>王维光(导师: 王枏) - 《北京邮电大学硕士论文》- 2015-03-16                  | 9.3% ( 1032 )<br>是否引证: 是 |

|   |   |                        |
|---|---|------------------------|
| 4 | 于俊_基于等级容灾模型<br>于俊 - 《学术论文联合比对库》 - 2014-12-31                  | 7.4% ( 818 )<br>是否引证：否 |
| 5 | 基于决策树的搜索引擎恶意网页检测研究与实现<br>周浩(导师：孙建华) - 《湖南大学硕士论文》 - 2013-05-27 | 3.6% ( 396 )<br>是否引证：否 |
| 6 | 浅谈防火墙的控制列表与透明模式<br>刘卫;杨扬;刘善春;- 《甘肃科技》 - 2010-11-30            | 0.4% ( 40 )<br>是否引证：否  |
| 7 | 校园网中蠕虫病毒分析浅谈<br>肖忠良;张玲峰;罗斌;- 《硅谷》 - 2011-07-23                | 0.3% ( 29 )<br>是否引证：否  |

| 原文内容 |   | 相似内容来源   |
|------|---|--|
| 1    | <p>此处有 30 字相似</p> <p>中的各种应用链接到移动互联网。传统互联网下的恶意链接问题也同样蔓延到了移动互联网当中，并产生了大量的变种。</p> <p>因此，如何</p> <p>有效地检测出恶意链接，并且及时的识别出新出现的恶意链接样本，</p> <p>是目前国内外专家学者以及网络安全工作者探讨和研究的重点。目前，国内外针对恶意链接的检测技术，主要有防火墙过滤、利用黑/白</p>                           | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.性、易用性和跨平台性等优点而被越来越多的人所熟悉，随之而来的恶意网页攻击也逐渐成为网络信息安全的主要威胁之一。有效进行恶意网页检测，尤其是对新出现的恶意网页样本能够及时识别和检测，是当前web 安全检测领域研究的重要内容。通过调研国内外专家学者在恶意网页检测领域的研究动态和</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究 2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p> <p>1.性、易用性和跨平台性等优点而被越来越多的人所熟悉，随之而来的恶意网页攻击也逐渐成为网络信息安全的主要威胁之一。有效进行恶意网页检测，尤其是对新出现的恶意网页样本能够及时识别和检测，是当前web 安全检测领域研究的重要内容。通过调研国内外专家学者在恶意网页检测领域的研究动</p> |
| 2    | <p>此处有 79 字相似</p> <p>algorithms in the field of Naive Bayesian algorithm, C4.5, classification and regression tree and machine learning support vector machine algorithm, and based on the experience of malicious links fr</p> | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.的最佳分类算法，我们分别使用了朴素贝叶斯算法Na?ve Bayes、决策树算法Decision Tree、分类回归树算法Classification And Regression Tree、支持向量机算法Support Vector Machine来进行实验，通过对实验结果进行比较，最终选择分类效果最好的分类算法作为网页恶意检测系统的分类学习算法。这四种分类算法</p>  |
| 3    | <p>此处有 56 字相似</p> <p>ructure</p> <p>1 绪论</p> <p>1.1 课题的研究背景</p> <p>在过去的十年里，互联网的飞速发展极大的改变了人们的生活。</p> <p>从日常办公到在线购物，从社交网络到在线课堂等待，我们无时无刻不在享受互联网给我们的生活带来的便捷。但是，互联网在极大的丰富和方便了人们的生活、为社会带来了无限的商机和可能的同时，其所蕴含的宝贵财富也招致了黑产行业的势力涌入，从而埋下</p> | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》 - 2013-05-27 ( 是否引证：否 )</p> <p>1. 绪 论1.1 研究意义及背景随着互联网的飞速发展，互联网逐渐成为人们获取信息和资源的首要渠道。从桌面办公到网上购物，网络社交以及在线课堂，无不体现互联网带来的便捷。互联网在改变人们生活方式的同时，其潜在的安全隐患也不容小觑。大量的木马、病毒，蠕虫，无时无刻不在威胁着浏览器的安全。用户甚</p>   |
| 4    | <p>此处有 36 字相似</p>   | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》 - 2013-05-27 ( 是否引证：否 )</p>   |

|   |   |   |
|---|---|---|
|   | <p>着互联网的木马、病毒等，给互联网用户带来了极大的安全隐患，严重干扰了正常的网络环境。其突出的一个传播途径，就是恶意链接。</p> <p>用户不需要下载和打开任何文件，一旦用户点击了恶意链接，用户的个人隐私信息</p> <p>就会轻易的被泄露而不被用户所察觉，或者恶意的软件就会自动的被安装，进而用户的整个电脑都将被攻击者所控制。</p> <p>国民经济的持</p>   | <p>1.联网在改变人们生活方式的同时，其潜在的安全隐患也不容小觑。大量的木马、病毒，蠕虫，无时无刻不在威胁着浏览器的安全。用户甚至不需要下载任何文件，一旦用户访问这些恶意网页，个人隐私信息将泄漏，恶意软件将强制安装，甚至机器将成为肉鸡完全由攻击者控制。互联网推动了国民经济的快速繁荣，仅2012 年淘宝</p>  |
| 5 | <p>此处有 131 字相似</p> <p>市，层出不穷的网络安全事故让人触目惊心。众多的安全事故凸显出Web安全的重要性，利用Web漏洞进行攻击是黑客的主要手段。</p> <p>2016 年绿盟科技数据显示：平均每个网站拥有网页489.4 个，每个网页中Web 漏洞出现的概率达46.9%，其中高危的 Web 漏洞为 1.9%，尤其以安全配置错误、跨站脚本漏洞为主。</p> <p>根据360公司 2016 年发布的2016年中国互联网安全报告显示</p> <p>：2016年360互联网安全中心共截获各类新增钓鱼网站196.9万个，共拦截钓鱼攻击279.5亿次。在新增的钓鱼网站中，</p> | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》- 2013-05-27 (是否引证：否)</p> <p>1.泄泄露、到美国金融机构接连遭受黑客攻击，每一次都使人胆战心惊。与此同时，Web 攻击成为众矢之的。2012 年绿盟科技数据[1]显示：平均每个网站拥有网页489.4 个，每个网页中Web 漏洞出现的概率达46.9%，其中高危的 Web 漏洞为 1.9%，尤其以安全配置错误、跨站脚本漏洞为主。来自赛门铁克公司 2011 年发布的安全威胁报告[2]显示：它们阻止了 55 亿个恶意攻击，相比 2010 年增长了 81%，基于 Web 网页的攻击增长了 36%，相当于</p> |
| 6 | <p>此处有 98 字相似</p> <p>法完整显示URL地址等缺陷。此外用户对待移动设备使用时的安全防范意识不强，这些都是导致移动互联网众多安全问题的潜在因素。</p> <p>Sophos 2016年安全威胁报告显示：黑客的攻击范围已经覆盖到更广的平台，从传统互联网领域到云服务平台再到移动终端领域，同时，黑客能够对新的安全防范措施做出更迅捷的响应，更频繁的采用0-Day漏洞进行攻击。其中60%以上的黑客攻击行为发生着移动互联网当中。</p> <p>面对如此数量庞大的恶意链接攻击行为，对用户的网络使用</p>   | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》- 2013-05-27 (是否引证：否)</p> <p>1.的途径如聊1基于决策树的搜索引擎恶意网页检测研究与实现天、社交平台，分类资源网络、讨论、微博等。Sophos 2013 年安全威胁报告[5]显示：攻击扩展覆盖到更广的平台，从社交网络到云服务平台到安卓移动设备，同时对新出现的安全研究发现做出的响应更迅速，更频繁的采用 Zero-day[6]攻击，80%以上的网站威胁采用重定向技术。面对如此庞大的恶意网页群体，如何在浏览网页过程中进行有效的防御变得十分重要</p>                              |
| 7 | <p>此处有 40 字相似</p> <p>进行攻击。其中60%以上的黑客攻击行为发生着移动互联网当中。</p> <p>面对如此数量庞大的恶意链接攻击行为，对用户的网络使用行为</p> <p>进行安全有效的保护就变得至关重要。目前成熟使用的安全产品大多是基于黑白名单机制，</p> <p>其原理是通过手工或蜜罐对拦截和爬取的URL进行鉴别和标记后，将数据按照黑白名单库分类号存放在数据库中，该方法的优点是对于</p>   | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》- 2013-05-27 (是否引证：否)</p> <p>1. Zero-day[6]攻击，80%以上的网站威胁采用重定向技术。面对如此庞大的恶意网页群体，如何在浏览网页过程中进行有效的防御变得十分重要。现有的安全防护产品大都基于黑白名单机制，对手工或蜜罐检测过的网页标记好存放在数据库中，数据库更新周期缓慢。当用户试图访问恶意网页时，它们能迅速阻止其</p>  |
| 8 | <p>此处有 286 字相似</p>  | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p>  |

|   |   |
|---|---|
| <p>，攻击方式种类繁多，这些都给给这种防御机制带来了巨大的困扰。因此如何有效而准确的对恶意网页进行实时检测具有深远意义。</p> <p>本文在分析上述研究成果的基础上，提出了在基于特征匹配的基础上使用分类学习算法进行恶意网页检测的方案。该方案首先使用特征匹配方式进行恶意网页匹配，以检测出已知特征类型的恶意网页；然后使用分类算法对第一步中未知类型的网页进行分类预测，并根据新增加的恶意网页样本集对分类器进行自适应学习和更新。本文重点研究使用机器学习中的分类算法来进行恶意网页的检测，可以有效检测网页中的暗链和恶意代码，准确区分恶意网页和正常网页，这对于维护网站的稳定与安全具有重要意义和价值。</p> <p>1.2 国内外研究现状</p> <p>目前国内外对网页恶意代码的检测与防治技术手段主要有：特征码匹配、防火墙拦截、入侵检测、蜜罐检测、虚拟机脱壳、机器学习检测等技术。</p> <p>1.特征码匹配技术[7]：该技术将待网页的源码转换成字符串后输入到检测系统</p> | <p>1.码。在恶意网页特征发生变化或者出现新形式的恶意代码时，这些技术都不能很好的进行恶意网页检测能力的自适应更新。本文在分析上述研究成果的基础上，提出了在基于特征匹配的基础上使用分类学习算法进行恶意网页检测的方案。该方案首先使用特征匹配方式进行恶意网页匹配，以检测出已知特征类型的恶意网页；然后使用分类算法对第一步中未知类型的网页进行分类预测，并根据新增加的恶意网页样本集对分类器进行自适应学习和更新。本文重点研究使用机器学习中的分类算法来进行恶意网页的检测，可以有效检测网页中的暗链和恶意代码，准确区分恶意网页和正常网页，这对于维护网站的稳定与安全具有重要意义和价值。1.4论文结构与主要工作本文在调研当前网络环境中存在的各种恶意网页攻击威胁的基础上，分析现有的基于特征匹配的恶意</p> <p>2.恶意网页代码特征的选取，基于恶意代码特征匹配的恶意网页检测及使用分类学习算法进行恶意检测的基础。1.2国内外研究动态目前国内外对网页恶意代码的检测与防治技术手段主要有：防火墙技术，入侵检测技术，特征码匹配技术，虚拟机脱壳引擎技术，蜜罐检测技术和机器学习技术等。Egele等人通过检测在JavaScript字符</p>   |
|   | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p> <p>1.意代码。在恶意网页特征发生变化或者出现新形式的恶意代码时，这些技术都不能很好的进行恶意网页检测能力的自适应更新。本文在深入分析上述研究成果的基础上，提出了在基于特征匹配的基础上使用分类学习算法进行恶意网页检测的方案。该方案首先使用特征匹配方式进行恶意网页匹配，以检测出已知特征类型的恶意网页；然后使用分类算法对第一步中未知类型的网页进行分类预测，并根据新增加的恶意网页样本集对分类器进行自适应学习和更新。本文重点研究使用机器学习中的分类算法来进行恶意网页的检测，可以有效检测网页中的暗链和恶意代码，准确区分恶意网页和正常网页，维护网站的稳定与安全，这对于网站运营者和网站浏览者具有重要的安全价值。1.4 论文结构与主要工作本文在调研当前网络环境中存在的各种恶意</p> <p>2.，是后面进行恶意网页代码特征的选取，基于恶意代码特征匹配的恶意网页检测及使用分类学习算法进行恶意检测的基础。1.2 国内外研究动态当前阶段，国内外对网页恶意代码的检测与防治技术手段主要有：防火墙技术，入侵检测技术，特征码匹配技术，虚拟机脱壳引擎技术，蜜罐检测技术和机器学习技术等。Egele 等人通过检测在 JavaScript</p> <p>于俊_基于等级容灾模型_于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.恶意代码。在恶意网页特征发生变化或者出现新形式的恶意代码时，这些技术都不能很好的进行恶意网页检</p> |



|   |            |   |
|---|------------|---|
|   |            | <p>测能力的自适应更新。本文在深入分析上述研究成果的基础上，提出了在基于特征匹配的基础上使用分类学习算法进行恶意网页检测的方案。该方案首先使用特征匹配方式进行恶意网页匹配，以检测出已知特征类型的恶意网页；然后使用分类算法对第一步中未知类型的网页进行分类预测，并根据新增加的恶意网页样本集对分类器进行自适应学习和更新。本文重点研究使用机器学习中的分类算法来进行恶意网页的检测，可以有效检测网页中的暗链和恶意代码，准确区分恶意网页和正常网页，维护网站的稳定与安全，这对于网站运营者和网站浏览者具有重要的安全价值。1.4 论文结构与主要工作本文在调研当前网络环境中存在的各种</p> <p>2.，机器学习算法设计及相关分类阈值设定的基础。图1-1 网页恶意代码攻击的一般过程1.2 国内外研究动态当前阶段，国内外对网页恶意代码的检测与防治技术手段主要有：防火墙技术，入侵检测技术，特征码匹配技术，启发式检测，虚拟机脱壳引擎技术，蜜罐检测技术，机器学习技术等。Egele等人通过检测在JavaScript</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比<br/>对库》 - 2015-01-09 (是否引证：否)</p> <p>1.意代码。在恶意网页特征发生变化或者出现新形式的恶意代码时，这些技术都不能很好的进行恶意网页检测能力的自适应更新。本文在深入分析上述研究成果的基础上，提出了在基于特征匹配的基础上使用分类学习算法进行恶意网页检测的方案。该方案首先使用特征匹配方式进行恶意网页匹配，以检测出已知特征类型的恶意网页；然后使用分类算法对第一步中未知类型的网页进行分类预测，并根据新增加的恶意网页样本集对分类器进行自适应学习和更新。本文重点研究使用机器学习中的分类算法来进行恶意网页的检测，可以有效检测网页中的暗链和恶意代码，准确区分恶意网页和正常网页，维护网站的稳定与安全，这对于网站运营者和网站浏览者具有重要的安全价值。1.4 论文结构与主要工作本文在调研当前网络环境中存在的各种恶意</p> <p>2.基本特征，是后面进行恶意代码特征的选取，机器学习算法设计及相关分类阈值设定的基础。1.2 国内外研究动态当前阶段，国内外对网页恶意代码的检测与防治技术手段主要有：防火墙技术，入侵检测技术，特征码匹配技术，虚拟机脱壳引擎技术，蜜罐检测技术，机器学习技术等。Egele 等人通过检测在 JavaScript</p> <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》 - 2013-05-27 (是否引证：否)</p> <p>1.防御变得迫在眉睫。因此如何有效而准确的对搜索引擎恶意网页进行实时检测具有深远意义。1.2 恶意网页检测技术及国内外研究现状现有的恶意网页检测技术主要分为两大类：动态检测和静态检测。其中动态检测又分为高交互的蜜罐和低交互的蜜罐[7]检测系统。它们各有优缺点，适用于</p> |
| 9 | 此处有 96 字相似 | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比</p>  |

|    |   |   |
|----|---|---|
|    | <p>恶意代码的检测与防治技术手段主要有：特征码匹配、防火墙拦截、入侵检测、蜜罐检测、虚拟机脱壳、机器学习检测等技术。</p> <p>1.</p> <p>特征码匹配技术[7]：该技术将待网页的源码转换成字符串后输入到检测系统，使用字符串匹配的算法将网页的源码片段与已知的恶意脚本特征码进行匹配，针对匹配成功的源码，系统会将该脚本在网页源码中的位置进行标记的同时给用户发出警告信息提示。特征码扫描法是病毒扫描的主要方法。通过使用每一种病毒体含有的特定病毒码对被检测的对</p> | <p>对库》- 2015-01-09 (是否引证：否)</p> <p>1.测和监控，并及时报告违反安全策略的异常行为，IDS 系统通常采用硬件和软件相结合的方式进行局域网安全检测。3. 特征码匹配技术：该技术将待检测网页源码以字符串的形式输入，使用字符串匹配的算法将网页源码中的脚本片段与已知的恶意脚本特征码进行匹配，如果匹配成功，则标记该恶意脚本在网页源码中的位置，并给出警告信息[11]，特征码匹配技术是当前传统反病毒软件使用的主要检测手段。由于使用简单的字符串匹配原</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.测和监控，并及时报告违反安全策略的异常行为，IDS 系统通常采用硬件和软件相结合的方式进行局域网安全检测。3. 特征码匹配技术：该技术将待检测网页源码以字符串的形式输入，使用第一章绪论4字符串匹配的算法将网页源码中的脚本片段与已知的恶意脚本特征码进行匹配，如果匹配成功，则标</p> <p>2.检测。3. 特征码匹配技术：该技术将待检测网页源码以字符串的形式输入，使用第一章绪论4字符串匹配的算法将网页源码中的脚本片段与已知的恶意脚本特征码进行匹配，如果匹配成功，则标记该恶意脚本在网页源码中的位置，并给出警告信息[11]，特征码匹配技术是当前传统反病毒软件使用的主要检测手段。由于使用简单的字符串匹配原</p> <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士学位论文》- 2015-03-16 (是否引证：是)</p> <p>1.测和监控，并及时报告违反安全策略的异常行为，IDS系统通常采用硬件和软件相结合的方式进行局域网安全检测。3.特征码匹配技术：该技术将待检测网页源码以字符串的形式输入，使用3北京邮电大学工程硕士学位论文 字符串匹配的算法将网页源码中的脚本片段与已知的恶意脚本特征码进行</p> <p>2.术：该技术将待检测网页源码以字符串的形式输入，使用3北京邮电大学工程硕士学位论文 字符串匹配的算法将网页源码中的脚本片段与已知的恶意脚本特征码进行匹配，如果匹配成功，则标记该恶意脚本在网页源码中的位置，并给出警告信息["]，特征码匹配技术是当前传统反病毒软件使用的主要检测手段。由于使用简单的字符串匹配原理，所</p> |
| 10 | <p>此处有 30 字相似</p> <p>扫描的主要方法。通过使用每一种病毒体含有的特定病毒码对被检测的对象进行扫描来发现病毒码所代表的病毒，是传统杀毒软件的主要</p> <p>利器，用来区分一个文件是否是病毒。与此相对的技术是主动防御。</p> <p>由于基于字符串匹配，检测方法相对简单，因此特征码匹配技术的检测速度和准确性比较好，但是特征码的提取需要提前准备，所以无法</p>   | <p>校园网中蠕虫病毒分析浅谈 肖忠良;张玲峰;罗斌;-《硅谷》- 2011-07-23 (是否引证：否)</p> <p>1.也是前所未有的。近年来蠕虫病毒造成的危害如表</p> <p>2。3常用病毒分析技术手段3.1特征码扫描技术特征码扫描是传统杀毒软件的主要利器,用来区分一个文件是否是病毒。与此相对的技术是主动防御,主动防御此处特指通过对计算机病毒的行为进行分析来实现检测的技术。特征码扫描主要是提取病毒文件的特征,注意:这并不一定是病</p>   |

|    |  |   |
|----|--|---|
|    |  |   |
| 11 | <p>此处有 37 字相似</p> <p>行扫描来发现病毒码所代表的病毒，是传统杀毒软件的主要利器，用来区分一个文件是否是病毒。与此相对的技术是主动防御。由于基于</p> <p>字符串匹配，检测方法相对简单，因此特征码匹配技术的检测速度和准确性比较好，</p> <p>但是特征码的提取需要提前准备，所以无法完成实时监测，对于加密和混淆变形后的恶意代码片段更是无能为力。</p> <p>2.防火墙技术[</p> | <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1.是将待检测的文件数据与特征码相匹配，如果匹配成功就可判定为该文件中包含恶意代码片段[11]。特征码匹配技术由于使用简单的字符串匹配原理，所以有较高的检测速度和准确性，但其弊端是提取特征码过程较慢，且不能检测出经过加密或混淆变形的恶意代码片段。</p> <p>4.启发式检测法：该方法采用探索与发</p> <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.码中的位置，并给出警告信息[11]，特征码匹配技术是当前传统反病毒软件使用的主要检测手段。由于使用简单的字符串匹配原理，所以特征码匹配技术有较高的检测速度和准确性，但其弊端是提取特征码过程较慢，且不能检测出经过加密或混淆变形的恶意代码片段。</p> <p>4.虚拟机脱壳引擎技术：该技术主要</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.页源码中的位置，并给出警告信息[11]，特征码匹配技术是当前传统反病毒软件使用的主要检测手段。由于使用简单的字符串匹配原理，所以特征码匹配技术有较高的检测速度和准确性，但其弊端是提取特征码过程较慢，且不能检测出经过加密或混淆变形的恶意代码片段。</p> <p>4.虚拟机脱壳引擎技术：该技术主要</p> <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.脚本在网页源码中的位置，并给出警告信息[11]，特征码匹配技术是当前传统反病毒软件使用的主要检测手段。由于使用简单的字符串匹配原理，所以特征码匹配技术有较高的检测速度和准确性，但其弊端是提取特征码的速度较慢，且不能检测出经过加密或混淆变形的恶意代码片段。</p> <p>4.虚拟机脱壳引擎技术：该技术</p> |
| 12 | <p>此处有 41 字相似</p> <p>，是一种描述网络安全系统安全性方法的形象说法。防火墙让不同的局域网之间建立起安全网关，从而避免内部网遭受外部非法用户的入侵。</p> <p>服务访问规则、验证工具、包过滤和应用网关是防火墙的主要组成部分。一旦建立起防火墙，</p> <p>该局域网内的计算机所有的出入流量都需要经过此防火墙的过滤和验证。因此，防火墙是私有网络的内部保护屏障，主要对外部网络发送</p> | <p>浅谈防火墙的控制列表与透明模式 刘卫;杨扬;刘善春;-《甘肃科技》- 2010-11-30 (是否引证：否)</p> <p>1.部网和外部网之间、专用网与公网之间的界面上构造的保护屏障。该计算机流入流出的所有网络通信均要经过此防火墙。防火墙主要由服务访问规则、验证工具、包过滤和应用网关4个部分组成,因此防火墙不但可以保护内部网络在Internet中的安全,同时还可以保护若干主机在一个内部网络中的安全。在每一个被防火墙分割的网络</p>  |
| 13 | <p>此处有 125 字相似</p> <p>是防火墙的主要组成部分。一旦建立起防火墙，该局域网内的计算机所有的出入流量都需要经过此防火墙的过滤和验证。因此，防火墙是</p> <p>私有网络的内部保护屏障，主要对外部网络发送的请求数据包进行特征匹配和安全检查，并对请求包的头部源</p>   | <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.该技术主要是通过中间过滤层将内部网络与外部网络相隔离，并对外部申请消息进行安全检测和访问控制[9]。作为内部私有网络的安全屏障，防火墙对外部网络发送的请求数据包进行特征匹配和安全检查，并对请求</p>   |

|    |  |  |
|----|--|--|
|    | <p>进行安全策略审计，一旦发现存在恶意特征或者未通过安全策略审计，则该请求数据包将会被直接丢弃。防火墙技术可以有效防御已知种类的非法攻击，但对新出现的变种攻击</p> <p>无法做出很好的过滤。</p> <p>3.蜜罐检测技术[33]：蜜罐的定义是一个安全资源，它的价值在于被探测、攻击和损害。因此，我们可</p>   | <p>包的头部源进行安全策略审计，一旦发现存在恶意特征或者未通过安全策略审计，则直接将该请求数据包丢弃。防火墙技术可以有效抵抗已知类型的恶意攻击，但对新类型的恶意攻击却表现不好。2. 入侵检测系统：Intrusion Detection System，简称 IDS，即识别与判断</p>  |
|    |  | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.该技术主要是通过中间过滤层将内部网络与外部网络相隔离，并对外部申请消息进行安全检测和访问控制[9]。作为内部私有网络的安全屏障，防火墙对外部网络发送的请求数据包进行特征匹配和安全检查，并对请求包的头部源进行安全策略审计，一旦发现存在恶意特征或者未通过安全策略审计，则直接将该请求数据包丢弃。防火墙技术可以有效抵抗已知类型的恶意攻击，但对新类型的恶意攻击却表现不好。2. 入侵检测系统：Intrusion Detection System，简称 IDS，即识别与判断恶意</p>   |
|    |  | <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.障，该技术主要是通过中间过滤层将内部网络与外部网络相隔离，并对外部申请消息进行安全检测和访问控制[9]。作为内部私有网络的安全屏障，防火墙对外部网络发送的请求数据包进行特征匹配和安全检查，并对请求包的头部源进行安全策略审计，一旦发现存在恶意特征或者未通过安全策略审计，则直接将该请求数据包丢弃。防火墙技术可以有效抵抗已知类型的恶意攻击，但对新类型的恶意攻击却表现不好。2.入侵检测系统 (Intrusion Detection System, IDS)：即识别与判断恶意使</p>  |
|    |  |  |
| 14 | <p>此处有 41 字相似</p> <p>术[33]：蜜罐的定义是一个安全资源，它的价值在于被探测、攻击和损害。因此，我们可以把蜜罐类比作一个情报收集系统，它故意</p> <p>引诱黑客攻击。一旦攻击者入侵并发生攻击动作，我们就可以观察、分析其攻击手段和目标，</p> <p>随时了解针对计算机发动的最新的攻击和漏洞，并为真正需要保护的计算机资源制定及时而有效的安全规则和策略。此外，我们还可以监</p> | <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.对于潜伏期较长的恶意木马检测效果并不理想。5. 蜜罐检测技术：蜜罐是一个提供服务仿真和漏洞仿真的虚拟服务环境，以引诱攻击者进行入侵攻击，并观察、分析和研究攻击者的攻击手段和真实企图的技术[14]。蜜罐技术是一项新兴的网络安全技术，通过观察和记录攻击者在虚拟环境中的攻击手段和操作，</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.对于潜伏期较长的恶意木马检测效果并不理想。5. 蜜罐检测技术：蜜罐是一个提供服务仿真和漏洞仿真的虚拟服务环境，以引诱攻击者进行入侵攻击，并观察、分析和研究攻击者的攻击手段和真实企图的技术[14]。蜜罐技术是一项新兴的网络安全技术，通过观察和记录攻击者在虚拟环境中的攻击手段和操</p> <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.于潜伏期较长的恶意木马检测效果并不理想。5.蜜罐</p> |



|    |   |   |
|----|---|---|
|    |   | <p>检测技术：蜜罐是一个提供服务仿真和漏洞仿真的虚拟服务环境，以<b>引诱攻击者进行入侵攻击，并观察、分析和研究攻击者的攻击手段和真实企图的技术</b>[14]。蜜罐技术是一项新兴的网络安全技术，通过观察和记录攻击者在虚拟环境中的攻击手段和操作，</p>  |
| 15 | <p><b>此处有 30 字相似</b></p> <p>黑客攻击。一旦攻击者入侵并发生攻击动作，我们就可以观察、分析其攻击手段和目标，随时了解针对计算机发动的最新的攻击和漏洞，<br/><b>并为真正需要保护的计算机资源制定及时而有效的安全规则和策略。</b></p> <p>此外，我们还可以监听攻击者之间的通信，收集收集攻击者的个人信息。蜜罐技术和虚拟机脱壳技术核心思路相同，都是通过仿真环境诱</p>                                  | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.项新兴的网络安全技术，通过观察和记录攻击者在虚拟环境中的攻击手段和操作，全面掌握攻击行为所用的技术及暴露的弱点，<b>并为真正需要保护的服务资源制定有效的防御策略。</b>该技术同虚拟机脱壳引擎技术相似，可以对恶意攻击行为观察透彻，但是检测周期长，占用计算资源较大。6. 基于机器</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.是一项新兴的网络安全技术，通过观察和记录攻击者在虚拟环境中的攻击手段和操作，全面掌握攻击行为所用的技术及暴露的弱点，<b>并为真正需要保护的服务资源制定有效的防御策略。</b>该技术同虚拟机脱壳引擎技术相似，可以对恶意攻击行为观察透彻，但是检测周期长，占用计算资源较大。6. 基于机器学习</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.一项新兴的网络安全技术，通过观察和记录攻击者在虚拟环境中的攻击手段和操作，全面掌握攻击行为所用的技术及暴露的弱点，<b>并为真正需要保护的服务资源制定有效的防御策略。</b>该技术同虚拟机脱壳引擎技术相似，可以对恶意攻击行为观察透彻，但是检测周期长，占用计算资源较大。6.基于机器学习</p> |
| 16 | <p><b>此处有 64 字相似</b></p> <p>需要保护的计算机资源制定及时而有效的安全规则和策略。此外，我们还可以监听攻击者之间的通信，收集收集攻击者的个人信息。蜜罐<br/><b>技术和虚拟机脱壳技术核心思路相同，都是通过仿真环境诱使攻击行为发生，不同的是，蜜罐技术的检测周期长，占用的计算机资源比较多。</b></p> <p>4.虚拟机脱壳引擎技术[11]：一旦病毒成功运行，那么用户计算机就已经被感染。为了做到防患于未然，虚拟机脱壳引擎 (VUE</p> | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1.一个拥有很多漏洞的环境，以引诱攻击者发起攻击，从而对攻击者的所有操作和行为进行监控和记录，最终掌握攻击者的企图和手段。<b>该技术同虚拟机脱壳引擎技术，可以对恶意攻击行为观察透彻，但是检测周期长，占用计算资源较大。</b>7. 基于机器学习的恶意代码检测技术：通过提取恶意网页的代码特征，利用训练样本训练出分类器模型，然后用该分类模型对待检测</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.虚拟环境中的攻击手段和操作，全面掌握攻击行为所用的技术及暴露的弱点，并为真正需要保护的服务资源制定有效的防御策略。<b>该技术同虚拟机脱壳引擎技术相似，可以对恶意攻击行为观察透彻，但是检测周期长，占用计算资源较大。</b>6. 基于机器学习的恶意网页检测技术：通过提取恶意网页的代码特征，利用训练样本训练出分类器模型，然后用该分类模型对待</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p>   |

|    |  |   |
|----|--|---|
|    |  | <p>1.拟环境中的攻击手段和操作，全面掌握攻击行为所用的技术及暴露的弱点，并为真正需要保护的服务资源制定有效的防御策略。该技术同虚拟机脱壳引擎技术相似，可以对恶意攻击行为观察透彻，但是检测周期长，占用计算资源较大。6. 基于机器学习的恶意网页检测技术：通过提取恶意网页的代码特征，利用训练样本训练出分类器模型，然后用该分类模型对待</p> <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.拟环境中的攻击手段和操作，全面掌握攻击行为所用的技术及暴露的弱点，并为真正需要保护的服务资源制定有效的防御策略。该技术同虚拟机脱壳引擎技术相似，可以对恶意攻击行为观察透彻，但是检测周期长，占用计算资源较大。6.基于机器学习的恶意网页检测技术：通过提取恶意网页的代码特征，利用训练样本训练出分类器模型，然后用该分类模型对待</p>  |
| 17 | <p>此处有 73 字相似</p> <p>计算机资源比较多。</p> <p>4.虚拟机脱壳引擎技术[11]：一旦病毒成功运行，那么用户计算机就已经被感染。为了做到防患于未然，虚拟机脱壳引擎（VUE）技术被提出，该技术的核心思想是给病毒搭建一个虚拟的仿真环境，主动诱骗木马、病毒等在虚拟环境中自觉脱掉“外壳”，并观察和记录其表现出的感染目标和攻击手段。由于“仿真环境”是与用户的实际操作环境相隔离，因此，病毒在的任何攻击操作都不会对计算机产生</p> | <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1. 启发式检测法的优点是误判率低，其规则可以及时调整；缺点是启发式学习效率低，相关参数需靠主观经验来进行设置。5. 虚拟机脱壳引擎技术：该技术主要是搭建一个虚拟的服务器环境，主动诱骗木马、病毒等在虚拟环境中自觉脱掉“外壳”，观察并记录木马、病毒的攻击目标和手段[14]。跟踪病毒在虚拟机中的行为，且保证该病毒虚拟环境与计算机相互隔离，因此，病毒在虚拟机上的运作</p> <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.有较高的检测速度和准确性，但其弊端是提取特征码过程较慢，且不能检测出经过加密或混淆变形的恶意代码片段。4. 虚拟机脱壳引擎技术：该技术主要是搭建一个虚拟的服务器环境，主动诱骗木马、病毒等在虚拟环境中自觉脱掉“外壳”，观察并记录木马、病毒的攻击目标和手段[12-13]。该技术虽然可以检测到恶意病毒、木马，但是虚拟环境的搭建成本高，且对于</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究 2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.技术有较高的检测速度和准确性，但其弊端是提取特征码过程较慢，且不能检测出经过加密或混淆变形的恶意代码片段。4. 虚拟机脱壳引擎技术：该技术主要是搭建一个虚拟的服务器环境，主动诱骗木马、病毒等在虚拟环境中自觉脱掉“外壳”，观察并记录木马、病毒的攻击目标和手段[12-13]。该技术虽然可以检测到恶意病毒、木马，但是虚拟环境的搭建成本高，且对于</p> <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.有较高的检测速度和准确性，但其弊端是提取特征码的速度较慢，且不能检测出经过加密或混淆变形的恶意代码片段。4.虚拟机脱壳引擎技术：该技术主要是搭建一个虚拟的服务器环境，主动诱骗木马、病毒等在虚拟</p> |

|    |   |  |
|----|---|--|
|    |   | 环境中自觉脱掉“外壳”，观察并记录木马、病毒的攻击目标和手段该技术虽然可以检测到恶意病毒、木马，但是虚拟环境的搭建成本高，且对于潜伏期较长的恶意木马   |
| 18 | <p>此处有 52 字相似</p> <p>”，并观察和记录其表现出的感染目标和攻击手段。由于“仿真环境”是与用户的实际操作环境相隔离，因此，病毒在任何攻击操作都不会对计算机产生影响。该技术虽然可以检测到恶意病毒、木马，但是搭建仿真环境耗时间且成本较高，同时，对于某一类长期潜伏在用户计算机的、只在特定的时间出发的病毒难以达到较好的检测效果。</p> <p>5.机器学习检测技术[13]：该技术的</p> | <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1.病毒的攻击目标和手段[14]。跟踪病毒在虚拟机中的行为，且保证该病毒虚拟环境与计算机相互隔离，因此，病毒在虚拟机上的运作不会响响用户计算机。该技术虽然可以检测到恶意病毒、木马，但是虚拟环境的搭建成本高，且对于潜伏期较长的恶意木马检测效果并不理想。6.蜜罐检测技术：蜜罐是一个被严密监控的计算资源，提供虚拟的服务器环</p> <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.，主动诱骗木马、病毒等在虚拟环境中自觉脱掉“外壳”，观察并记录木马、病毒的攻击目标和手段[12-13]。该技术虽然可以检测到恶意病毒、木马，但是虚拟环境的搭建成本高，且对于潜伏期较长的恶意木马检测效果并不理想。5.蜜罐检测技术：蜜罐是一个提供服务仿真和漏洞仿真的虚拟服务环境，以引诱</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.，主动诱骗木马、病毒等在虚拟环境中自觉脱掉“外壳”，观察并记录木马、病毒的攻击目标和手段[12-13]。该技术虽然可以检测到恶意病毒、木马，但是虚拟环境的搭建成本高，且对于潜伏期较长的恶意木马检测效果并不理想。5.蜜罐检测技术：蜜罐是一个提供服务仿真和漏洞仿真的虚拟服务环境，以引诱</p> <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士学位论文》- 2015-03-16 (是否引证：是)</p> <p>1.个虚拟的服务器环境，主动诱骗木马、病毒等在虚拟环境中自觉脱掉“外壳”，观察并记录木马、病毒的攻击目标和手段该技术虽然可以检测到恶意病毒、木马，但是虚拟环境的搭建成本高，且对于潜伏期较长的恶意木马检测效果并不理想。5.蜜罐检测技术：蜜罐是一个提供服务仿真和漏洞仿真的虚拟服务环境，以引</p> |
| 19 | <p>此处有 31 字相似</p> <p>13]：该技术的第一步和特征码检测技术有些相似，即分析恶意脚本源码后提取恶意代码的特征，利用机器学习算法和大量的样本进行分类训练后得到一个分类模型，之后利用该模型对需要检测的网页进行判断和分类。</p> <p>自机器学习被应用到恶意链接检测以来，许多学者专家提出了多种特征提取思路。针对HTML标签在页面中的使用频</p>                    | <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1.检测周期长，占用计算资源较大。7. 基于机器学习的恶意代码检测技术：通过提取恶意网页的代码特征，利用训练样本训练出分类器模型，然后用该分类模型对待检测的网页进行学习 and 分类。Hou等人根据动态HTML中不同标签的使用频率高低，将网页特征分层并在每层选择一定量的特征集，使用不同分类算</p> <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.期长，占用计算资源较大。6. 基于机器学习的恶意网页检测技术：通过提取恶意网页的代码特征，利用训练样本训练出分类器模型，然后用该分类模型对待检测的</p>   |

|    |   |  |
|----|---|--|
|    |   | <p>网页进行学习 and 分类。Hou 等人根据动态 HTML 中不同标签的使用频率高低，将网页特征分层并在每层选择一定量的特征集，</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.测周期长，占用计算资源较大。6. 基于机器学习的恶意网页检测技术：通过提取恶意网页的代码特征，利用训练样本训练出分类器模型，然后用该分类模型对待检测的网页进行学习 and 分类。Hou 等人根据动态 HTML 中不同标签的使用频率高低，将网页特征分层并在每层选择一定量的特征集，</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.周期长，占用计算资源较大。6.基于机器学习的恶意网页检测技术：通过提取恶意网页的代码特征，利用训练样本训练出分类器模型，然后用该分类模型对待检测的网页进行学习 and 分类。Hou等人根据动态HTML中不同标签的使用频率高低，将网页特征分层并在每层选择一定量的特征集，使</p>   |
| 20 | <p>此处有 109 字相似</p> <p>应用到恶意链接检测以来，许多学者专家提出了多种特征提取思路。针对HTML标签在页面中的使用频率不同，Hou[44]等人将</p> <p>网页特征分层并在每层选择一定量的特征集，使用不同分类算法对恶意网页进行学习 and 分类预测，并比较了不同算法在恶意检测中的优缺点。ChitraS[3]等人提出将支持向量机算法和遗传进化模糊规则相结合来进行恶意网页检测的方案。</p> <p>基于机器学习的恶意链接检测相比传统的检测方法在检测效果和应对变种攻击方便提升明显，但机器学习存在训练和学习过程太过复杂，</p> | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1.本训练出分类器模型，然后用该分类模型对待检测的网页进行学习 and 分类。Hou等人根据动态HTML中不同标签的使用频率高低，将网页特征分层并在每层选择一定量的特征集，使用不同分类算法对恶意网页进行学习 and 分类预测，并比较了不同算法在恶意检测中的优缺点[16]。Fette 等人利用统计方法进行恶意电子邮件的分类检测[17]。Chitra S等人提出将支持向量机算法和遗传进化模糊规则</p> <p>2.分类预测，并比较了不同算法在恶意检测中的优缺点[16]。Fette 等人利用统计方法进行恶意电子邮件的分类检测[17]。Chitra S等人提出将支持向量机算法和遗传进化模糊规则相结合来进行恶意网页检测的方案[18]，但方案算法过于复杂，生成最终决策规则的效率过低，不适于处理日益增长的恶意网页检测情况。1.3 课题研究的目的与意义</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.类器模型，然后用该分类模型对待检测的网页进行学习 and 分类。Hou 等人根据动态 HTML 中不同标签的使用频率高低，将网页特征分层并在每层选择一定量的特征集，使用不同分类算法对恶意网页进行学习 and 分类预测，并比较了不同算法在恶意检测中的优缺点[15]。Fette 等人利用统计方法进行恶意电子邮件的分类检测[16]。Chitra S 等人提出将支持向量机算</p> <p>2.同算法在恶意检测中的优缺点[15]。Fette 等人利用统计方法进行恶意电子邮件的分类检测[16]。Chitra S 等人提出将支持向量机算法和遗传进化模糊规则相结合来</p> |



|    |   |   |
|----|---|---|
|    |   | <p>进行恶意网页检测的方案[17]，但方案算法过于复杂，生成最终决策规则的效率过低，不适于处理日益增长的恶意网页检测情况。1.3 课题研究的目的</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 ( 是否引证：否 )</p> <p>1.类器模型，然后用该分类模型对待检测的网页进行学习和分类。Hou 等人根据动态 HTML 中不同标签的使用频率高低，将网页特征分层并在每层选择一定量的特征集，使用不同分类算法对恶意网页进行学习和分类预测，并比较了不同算法在恶意检测中的优缺点[15]。Fette 等人利用统计方法进行恶意电子邮件的分类检测[16]。Chitra S 等人提出将支持向量机算法和</p> <p>2.同算法在恶意检测中的优缺点[15]。Fette 等人利用统计方法进行恶意电子邮件的分类检测[16]。Chitra S 等人提出将支持向量机算法和遗传进化模糊规则相结合来进行恶意网页检测的方案[17]，但方案算法过于复杂，生成最终决策规则的效率过低，不适于处理日益增长的恶意网页检测情况。1.3 课题研究的目的与意</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》- 2015-03-16 ( 是否引证：是 )</p> <p>1.出分类器模型，然后用该分类模型对待检测的网页进行学习和分类。Hou等人根据动态HTML中不同标签的使用频率高低，将网页特征分层并在每层选择一定量的特征集，使用不同分类算法对恶意网页进行学习和分类预测，并比较了不同算法在恶意检测中的优缺点Fette等人利用统计方法进行恶意电子邮件的分类检测[16]。Chitra S等人提出将支持向量机算法和遗传进化模糊规则相</p> <p>2.较了不同算法在恶意检测中的优缺点Fette等人利用统计方法进行恶意电子邮件的分类检测[16]。Chitra S等人提出将支持向量机算法和遗传进化模糊规则相结合来进行恶意网页检测的方案[17]，但方案算法过于复杂，生成最终决策规则的效率过低，不适于处理日益增长的恶意网页检测情况。1.3课题研究的目的与</p> |
| 21 | <p>此处有 165 字相似</p> <p>机器学习存在训练和学习过程太过复杂，生成的决策规则效率较低等问题。</p> <p>1.3 课题的研究意义</p> <p>经过上文的分析可以看出，</p> <p>这些技术基本都是基于模式匹配或者代码特征签名的方法进行恶意网页检测，而且检测的恶意网页形式单一、特征明显，当网页恶意代码进行变形、加密、混淆或者更新后，这些技术就很难有效检测恶意代码。在恶意网页特征发生变化或者出现新形式的恶意代码时，这些技术都不能很好的进行恶意网页检测能力的自适应更新。因此，本文在前人研究成果的基础上，提出基于支持向量机算法[1] ( Support Vector Machine , SVM ) 和文档对象模型[5] ( Document Ob</p> | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》- 2014-12-31 ( 是否引证：否 )</p> <p>1.加到黑名单中，以提高恶意网页URL的覆盖范围，但这种方案检测形式单一，很难检测出加密或新出现的恶意网页URL。上述这些技术基本都是基于模式匹配或者代码特征签名的方法进行恶意网页检测，而且检测的恶意网页形式单一、特征明显，当网页恶意代码进行变形、加密、混淆或者更新后，这些技术就很难有效检测恶意代码。在恶意网页特征发生变化或者出现新形式的恶意代码时，这些技术都不能很好的进行恶意网页检测能力的自适应更新。本文在深入分析上述研究成果的基础上，提出了在基于特征匹配的基础上使用分类学习算法进行恶意网页检测的方案。该方案首先使用特征匹配方式进行恶意网页匹配，以检测出已知特征类型</p>   |

|  |   |
|--|---|
|  | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 (是否引证：否)</p> <p>1.加到黑名单中，以提高恶意网页 URL 的覆盖范围，但这种方案检测形式单一，很难检测出加密或新出现的恶意网页 URL。这些技术基本都是基于模式匹配或者代码特征签名的方法进行恶意网页检测，而且检测的恶意网页形式单一、特征明显，当网页恶意代码进行变形、加密、混淆或者更新后，这些技术就很难有效检测恶意代码。在恶意网页特征发生变化或者出现新形式的恶意代码时，这些技术都不能很好的进行恶意网页检测能力的自适应更新。本文在深入分析上述研究成果的基础上，提出了在基于特征匹配的基础上使用分类学习算法进行恶意网页检测的方案。该方案首先使用特征匹配方式进行恶意网页匹配，以检测出已知</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证：否)</p> <p>1.加到黑名单中，以提高恶意网页 URL 的覆盖范围，但这种方案检测形式单一，很难检测出加密或新出现的恶意网页 URL。这些技术基本都是基于模式匹配或者代码特征签名的方法进行恶意网页检测，而且检测的恶意网页形式单一、特征明显，当网页恶意代码进行变形、加密、混淆或者更新后，这些技术就很难有效检测恶意代码。在恶意网页特征发生变化或者出现新形式的恶意代码时，这些技术都不能很好的进行恶意网页检测能力的自适应更新。本文在深入分析上述研究成果的基础上，提出了在基于特征匹配的基础上使用分类学习算法进行恶意网页检测的方案。该方案首先使用特征匹配方式进行恶意网页匹配，以检测出已知</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士学位论文》 - 2015-03-16 (是否引证：是)</p> <p>1.态增加到黑名单中，以提高恶意网页URL的覆盖范围，但这种方案检测形式单一，很难检测出加密或新出现的恶意网页URL。这些技术基本都是基于模式匹配或者代码特征签名的方法进行恶意网页检测，而且检测的恶意网页形式单一、特征明显，当网页恶意代码进行变形、加密、混淆或者更新后，这些技术就很难有效检测恶意代码。在恶意网页特征发生变化或者出现新形式的恶意代码时，这些技术都不能很好的进行恶意网页检测能力的自适应更新。本文在分析上述研究成果的基础上，提出了在基于特征匹配的基础上使用分类学习算法进行恶意网页检测的方案。该方案首先使用特征匹配方式进行恶意网页匹配，以检测出</p> |
|--|---|

| 指 标  |
|--|
| 疑似剽窃文字表述   |
| <div> <div>1. 从日常办公到在线购物，从社交网络到在线课堂等待，我们无时无刻不在享受互联网给我们的生活带来的便捷。但是，互联网在</div> <div>2. 2016 年绿盟科技数据显示：平均每个网站拥有网页 489.4 个，每个网页中Web 漏洞出现的概率达 46.9%，其中高危的 Web 漏洞为 1.9%，尤其以安全配置错误、跨站脚本漏洞为主。<br/>根据360公司 2016 年发布的2016年中国互联网安全报告显示</div> <div>3. Sophos 2016年安全威胁报告显示：黑客的攻击范围已经覆盖到更广的平台，从传统互联网领域到云服务平台再到移动</div> </div> |

- 终端领域，同时，黑客能够对新安全防范措施做出更迅捷的响应，更频繁的采用0-Day
- 进行安全有效的保护就变得至关重要。目前成熟使用的安全产品大多是基于黑白名单机制，
  - 服务访问规则、验证工具、包过滤和应用网关是防火墙的主要组成部分。一旦建立起防火墙，

## 2. 11491915\_刘力铭\_第2部分

总字数：9639

相似文献列表 文字复制比：20.5%(1979) 疑似剽窃观点：(0)

|    |  |                        |
|----|--|------------------------|
| 1  | JavaScript脚本恶意代码检测系统<br>赵梓旭(导师：张炜;董志明) - 《哈尔滨工业大学硕士论文》 - 2016-06-01  | 8.2% ( 786 )<br>是否引证：否 |
| 2  | 王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15<br>王维光 - 《学术论文联合比对库》 - 2015-01-15  | 7.6% ( 737 )<br>是否引证：否 |
| 3  | 恶意网页医疗数据应急通信移动通信<br>- 《学术论文联合比对库》 - 2015-01-09   | 7.6% ( 730 )<br>是否引证：否 |
| 4  | 基于分类算法的恶意网页检测技术研究<br>王维光(导师：王枏) - 《北京邮电大学硕士论文》 - 2015-03-16  | 7.0% ( 674 )<br>是否引证：是 |
| 5  | 基于互联网营销的B&R品牌推广策略研究<br>李东敏(导师：郭毅) - 《华东理工大学硕士论文》 - 2015-04-08  | 2.7% ( 262 )<br>是否引证：否 |
| 6  | 二维码怎样制作_欢乐214122<br>- 《网络 ( <a href="http://blog.sina.com">http://blog.sina.com</a> ) 》 - 2015   | 2.7% ( 262 )<br>是否引证：否 |
| 7  | 河北省建二维码注册解析服务中心 揭其背后“玄机”_播报天下<br>- 《网络 ( <a href="http://www.gywb.cn/c">http://www.gywb.cn/c</a> ) 》 - 2016                                | 2.7% ( 262 )<br>是否引证：否 |
| 8  | 计控学院2220140267董阳<br>董阳 - 《学术论文联合比对库》 - 2015-10-10  | 2.7% ( 262 )<br>是否引证：否 |
| 9  | 基于JavaEE的设备管理系统<br>1247172270-张玉 - 《学术论文联合比对库》 - 2015-09-14  | 2.7% ( 262 )<br>是否引证：否 |
| 10 | 1988087_变电站智能巡检系统设计与应用_XXX论文排版2014.05.02<br>变电站智能巡检系统设计与应用 - 《学术论文联合比对库》 - 2015-05-02  | 2.7% ( 260 )<br>是否引证：否 |
| 11 | 20111103178送审--变电站智能巡检系统设计与应用0505<br>- 《学术论文联合比对库》 - 2015-05-05  | 2.7% ( 260 )<br>是否引证：否 |
| 12 | 激光直接标记二维码(Data Matrix)及识别用户操作系统设计_图文<br>- 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》 - 2016                    | 2.6% ( 255 )<br>是否引证：否 |
| 13 | 密集型区域中物流终端的安全分析及解决策略<br>杨兴波;张林; - 《科技视界》 - 2016-05-15  | 2.6% ( 251 )<br>是否引证：否 |
| 14 | 二维码技术在高校教学信息化中的应用实践探索<br>宋艳艳;颜月南; - 《现代经济信息》 - 2016-11-25  | 2.6% ( 251 )<br>是否引证：否 |
| 15 | Android手机僵尸网络技术及防御方法研究<br>王钰淇(导师：董永强) - 《东南大学硕士论文》 - 2015-10-21  | 2.6% ( 251 )<br>是否引证：否 |
| 16 | 跨行二维码支付方案研究<br>赵东林; - 《2016年第二届今日财富论坛论文集》 - 2016-05-03   | 2.6% ( 251 )<br>是否引证：否 |
| 17 | &#34;二维码&#34;、&#34;扫码关注&#34;用英语怎么说-搜狐教育<br>- 《网络 ( <a href="http://learning.sohu">http://learning.sohu</a> ) 》 - 2016                      | 2.6% ( 251 )<br>是否引证：否 |
| 18 | 助威详情   [原创]揭秘河南大桥爆炸调查报告中的弥天大谎   华声论坛   全球中文论坛前十强 - 华声论坛<br>- 《互联网文档资源 ( <a href="http://bbs.voc.com.c">http://bbs.voc.com.c</a> ) 》 - 2016 | 2.6% ( 251 )<br>是否引证：否 |
| 19 | 二维码扫一扫手机就会中毒是真的吗？<br>- 《互联网文档资源 ( <a href="http://www.360doc.co">http://www.360doc.co</a> ) 》 - 2016                                       | 2.6% ( 251 )<br>是否引证：否 |
| 20 | 物联网+智能停车平台项目计划书(高校版)_图文<br>- 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》 - 2016                                 | 2.6% ( 251 )<br>是否引证：否 |
| 21 | 94_张悦_基于RFID组网分割的大棚温度监测方法研究<br>张悦 - 《学术论文联合比对库》 - 2015-05-02   | 2.6% ( 251 )<br>是否引证：否 |
| 22 | 优惠券分发系统的设计与实现<br>刘正刚 - 《学术论文联合比对库》 - 2015-11-16  | 2.6% ( 251 )<br>是否引证：否 |
| 23 | 二维码在建筑工程物资管理过程中的应用<br>丁立欢;窦市鹏; - 《《工业建筑》2016年增刊II》 - 2016-06-20  | 2.2% ( 215 )<br>是否引证：否 |
| 24 | QR码在Android系统中的研究与应用<br>阚钊玉; - 《网络安全技术与应用》 - 2016-03-15  | 2.2% ( 213 )<br>是否引证：否 |

|    |   |                        |
|----|---|------------------------|
| 25 | 我们身边的条形码<br>- 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》 - 2016 | 2.2% ( 213 )<br>是否引证：否 |
| 26 | 测绘学院-工程硕士-赵永<br>- 《学术论文联合比对库》 - 2014-11-05  | 0.8% ( 74 )<br>是否引证：否  |
| 27 | HTML5之荣耀与苦涩<br>刘宗凡;邱元阳;谭剑仪; - 《中国信息技术教育》 - 2012-12-15                                       | 0.6% ( 54 )<br>是否引证：否  |
| 28 | 跨平台移动应用开发框架的研究与设计<br>王才双(导师：周华) - 《云南大学硕士论文》 - 2016-05-01                                   | 0.4% ( 36 )<br>是否引证：否  |

| 原文内容 |   | 相似内容来源   |
|------|---|--|
| 1    | <p>此处有 32 字相似</p> <p>证。<br/>第6章对课题所做的工作进行了总结，同时分析了系统方案中存在的不足之处，对未来的进一步的研究工作进行展望。</p> <p>2</p> <p><b>恶意链接相关技术介绍</b><br/>本章节的主要内容是对恶意网页的研究和恶意链接攻击的底层原理。同时，本章还详细分析了作为恶意链接主要攻击手段之一的Javascript脚本是如何注入到攻击页面从而</p>                     | <p>JavaScript脚本恶意代码检测系统 赵梓旭 - 《哈尔滨工业大学硕士论文》 - 2016-06-01 ( 是否引证：否 )</p> <p>1.测系统的不足之处，以及完善的方法。 哈尔滨工业大学工程硕士学位论文 - 9 -第2章 <b>恶意代码检测相关理论知识本章介绍作者对恶意网页的研究，以及恶意 Java Script 攻击实现的原理，其能产生的危害，并详细分析如何通过 Java Script 对网页进行恶</b></p>  |
| 2    | <p>此处有 71 字相似</p> <p>et应用程序、包含恶意ActionScript 的Flash动画以及IE浏览器下常用的ActiveX控件等的页面，并生成<b>恶意链接后在用户中传播。通常，恶意链接会伪装成正常的网页链接，或者通过正常页面引入，用户访问这些 Web 页面时，会触发被攻击的 B 及漏洞，</b>蠕虫、木马等恶意软件就会在用户毫无察觉的情况下控制用户的电脑，盗取用户隐私，破坏用户数据，更严重的甚至会将用户的电脑控制</p> | <p>JavaScript脚本恶意代码检测系统 赵梓旭 - 《哈尔滨工业大学硕士论文》 - 2016-06-01 ( 是否引证：否 )</p> <p>1.恶意软件可以在用户不知情的情况下获取用户计算机控制权，在后台盗取用户隐私，破坏计算机系统，甚至组件僵尸网络来实现<b>恶意软件的传播复制。通常，恶意网页会伪装成正常网页，或者通过正常页面引入，用户访问这些 Web 页面时，会触发被攻击的 bug 及漏洞</b>，使主机被恶意攻击人员攻击。2.1.2 恶意网页的危害<br/>大量浏览器（尤其是 IE 浏览器）相关漏洞</p>   |
| 3    | <p>此处有 41 字相似</p> <p>者是“肉鸡”从而更快速的传播自己。</p> <p>2、恶意脚本<br/>网页恶意代码[11]是注入到网页中通过利用系统或者后台漏洞达到攻击<b>目的的代码片段，常见的恶意代码通常使用 JavaScript、TypeScript、CSS等编写，并且经过一定的编码和代码混淆技术伪装来逃避系统的判别。因此，利用这些语言编写的恶意攻击代码片断叫做恶意脚本</b></p>                     | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.意网页攻击的关键在于能够发现并利用客户端操作系统或浏览器的漏洞，并利用这些漏洞在用户机器上隐蔽的下载并执行恶意程序的<b>目的。其中，恶意攻击脚本一般使用 JavaScript、VBScript 等语言编写，利用操作系统和浏览器中存在的漏洞，使用精心构造的恶意代码触发操作系统的异常调用并获取一定的系统权限，</b></p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p> <p>1.意网页攻击的关键在于能够发现并利用客户端操作系统或浏览器的漏洞，并利用这些漏洞在用户机器上隐蔽的下载并执行恶意程序的<b>目的。其中，恶意攻击脚本一般使用 JavaScript、VBScript 等语言编写，利用操作系统和浏览器中存在的漏洞，使用精心构造的恶意代码触发操作系统的异常调用并获取一定的系统权限，</b></p> |
| 4    | <p>此处有 41 字相似</p> <p>来说，恶意脚本的攻击氛围直接攻击和间接攻击两种。</p>   | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.言编写的恶意攻击代码片段称为攻击脚本；包含恶意</p>   |



|   |  |   |
|---|--|---|
|   | <p>( 1 ) 直接攻击：是指恶意代码直接包含在恶意网页里面，用户通过恶意链接访问恶意网页后，嵌入其中的恶意脚本将被触发并发生攻击行为，如篡改用户首页信息、弹出钓鱼网站的伪装页面、窃取用户隐私数据、伪造用户网银身份进行资金诈骗、向网站后台注入非法SQL语句夺得站点控制权等。</p> <p>(</p>  | <p>攻击脚本的网页称为攻击页面，其所在的网站为恶意攻击宿主站点。当用户访问含有恶意代码的网页时，其中被嵌入的恶意代码就会被触发而执行攻击行为，包括篡改网站首页、非法弹出对话框、强制修改用户操作系统文件、非法占用用户资源、窃取用户隐私数据等，甚至可能将用户主机变为僵尸机器以实现对其他服务器</p>   |
|   |  | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p>  |
|   |  | <p>1.言编写的恶意攻击代码片段称为攻击脚本；包含恶意攻击脚本的网页称为攻击页面，其所在的网站为恶意攻击宿主站点。当用户访问含有恶意代码的网页时，其中被嵌入的恶意代码就会被触发而执行攻击行为，包括篡改网站首页、非法弹出对话框、强制修改用户操作系统文件、非法占用用户资源、窃取用户隐私数据等，甚至可能将用户主机变为僵尸机器以实现对其他服务器</p>                  |
|   |  | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p>   |
| 5 | <p>此处有 72 字相似</p> <p>脚本注入到漏洞网页中，当用户浏览该网页时将会触发恶意脚本的执行，通过重定向、页面替换等方式诱导用户通过恶意链接进入真正的恶意页面。具体的攻击过程如下图2-1所示：</p> <p>图 21 恶意链接攻击的一般流程</p> <p>由上图可以看出，具体的攻击过程分为以下几个步骤：</p> <p>Step</p> <p>1：黑客向漏洞页面注入恶意脚本；</p> <p>Step 2：用户通过浏览器或者内嵌WebView引擎的客户端发送浏览被感染页面的请</p> | <p>1.写的恶意攻击代码片段称为攻击脚本；包含恶意攻击脚本的网页称为攻击页面，其所在的网站为恶意攻击宿主站点。当用户访问含有恶意代码的网页时，其中被嵌入的恶意代码就会被触发而执行攻击行为，包括篡改网站首页、非法弹出对话框、强制修改用户操作系统文件、非法占用用户资源、窃取用户隐私数据等，甚至可能将用户主机变为僵尸机器以实现对其他服</p>                      |
|   |  | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p>   |
|   |  | <p>1.正常网页中后，通过各种手段诱使用户点击超链接等形式将恶意脚本植入到用户主机的被动攻击[24-25]。网页恶意脚本攻击的一般流程如下图 2-1 所示：图 2-1 网页恶意脚本攻击一般流程由上图可知，一般的恶意脚本攻击过程有如下几个阶段:Step 1: 客户端访问包含特定漏洞类型恶意代码的页面；Step 2: 服务器响应访问请求，并将嵌入恶意脚本的页面发送给客户端；</p> |
|   |  | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p>  |
|   |  | <p>1.攻击[24-25]。网页恶意脚本攻击的一般流程如下图 2-1 所示：第二章网页恶意代码概述9图 2-1 网页恶意脚本攻击一般流程由上图可知，一般的恶意脚本攻击过程有如下几个阶段:Step 1: 客户端访问包含特定漏洞类型恶意代码的页面；Step 2: 服务器响应访问请求，并将嵌入恶意脚本的页面发送给客户端；</p>                             |
|   |  | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p>   |
|   |  | <p>1..恶意'号，. l 4攻失歌 &amp;lt; ? pT攻击成恶意《字下截结点 \S执行恶#^序j图2-1网页恶意脚本攻击一般流程由上图可知，一般的恶意脚本攻击过程有如下几个阶段：Step 1:客户端访问包含特定漏洞类型恶意代码的页面；Step 2;服务器响应访问请求，并将嵌入恶意脚本的页面发送给客户端</p>                               |

|   |   |   |
|---|---|---|
| 6 | <p>此处有 75 字相似</p> <p>行相关的介绍。</p> <p>2.3.1 JavaScript简介</p> <p>JavaScript[11]是一种非编译式脚本语言，在1995由Netscape公司的Brendan Eich在网景导航者浏览器上首次设计实现而成。JavaScript语言是一种动态类型、弱类型、基于原型的语言，通过WebView引擎中内置的JavaScript引擎解析执行，它的主要设计目的是应用在HTML语言编写的网页中来增加网</p>  | <p>测绘学院-工程硕士-赵永 -《学术论文联合比对库》- 2014-11-05 (是否引证：否)</p> <p>1.的类均可COM组件化，以组件使用。[32]3.3.2.3 JavaScript开发语言JavaScript，是由Netscape公司的Brendan Eich与1995年在网景导航者浏览器上首次设计实现而成的，是一种集直译式、动态类型、弱类型、基于原型于一体的开发语言，其解释器 JavaScript引擎作为浏览器的一部分，被广泛应用于客户端的脚本语言，用以增加网页的动态功能。JavaSc</p>  |
| 7 | <p>此处有 37 字相似</p> <p>有对变量类型做严格的类型区分和限定，因此，相比于C++和Java等强数据类型语言，JavaScript语言编写程序时较为简单。</p> <p>(4) 动态性：JavaScript是一种采用事件驱动的脚本语言，其执行不必经过服务器的解析和帮助，只要解析引擎按照ECMA标准设计实现，那么JavaScript程序可以响应鼠标移动、屏</p>   | <p>跨平台移动应用开发框架的研究与设计 王才双 -《云南大学硕士论文》- 2016-05-01 (是否引证：否)</p> <p>1.量类型的语言，它没有像C、Java等语言严格的对变量类型分出整型、浮点型等变量类型，它的设计与使用都比较简单。JavaScript是一种采用事件驱动的脚本语言，事件驱动也就是当用户点击或移动W沈页面等时，JavaScript能根据这些事件进行相应的相应，JavaScript</p>  |
| 8 | <p>此处有 55 字相似</p> <p>执行JavaScript程序的环境包含其解析引擎即可。因此一个JavaScript脚本在编写后可以在多种端设备上执行不同于PHP与ASP等服务器端的脚本语言，JavaScript主要被作为客户端脚本语言在用户的浏览器上运行。</p> <p>因此最初程序员们喜欢将许多逻辑用JavaScript在端实现从而达到对服务器减去负载的目的，但与此同时也引入了安全性这一</p>   | <p>HTML5之荣耀与苦涩 刘宗凡;邱元阳;谭剑仪;-《中国信息技术教育》- 2012-12-15 (是否引证：否)</p> <p>1.他场合,如服务器端编程。完整的JavaScript实现包含三个部分:ECMAScript、文档对象模型、浏览器对象模型。不同于服务器端脚本语言,如PHP与ASP,JavaScript主要被作为客户端脚本语言在用户的浏览器上运行,不需要服务器的支持。所以早期的程序员比较青睐于JavaScript以减少对服务器的负担,而与此同时也带来另一个问题:安全</p>   |
| 9 | <p>此处有 180 字相似</p> <p>me引擎等。</p> <p>2.3.3 页面加载JavaScript代码的方式</p> <p>目前，在网页中引入 JavaScript脚本的方式主要有三种：</p> <p>(1) 通过&lt;script&gt;标签块内嵌 JavaScript 代码通过编写&lt;script&gt;标签，content 为 JavaScript 代码，是 HTML 动态化最常见最通用的方法。这种方法是最常见的。示例代码如下：</p> <pre>&lt;script&gt; document.write('Attack'); &lt;/script&gt;</pre> <p>(2) 通过src链接外部 JavaScript脚本</p> <p>在实际的网页开发中，为了实现代码的解耦和组件的复用，开发者们可以将常用的JavaScript</p> | <p>JavaScript脚本恶意代码检测系统 赵梓旭 -《哈尔滨工业大学硕士论文》- 2016-06-01 (是否引证：否)</p> <p>1.HTML 文档中载入 Java Script，并用其完成动态功能。在 HTML 中载入 Java Script 主要有三种方式：</p> <p>(1) 通过&lt;script&gt;标签块内嵌 Java Script 代码通过编写&lt;script&gt;标签，content 为 Java Script 代码，是 HTML 动态化最常见最通用的方法。这种方法是最常见的。示例代码如下：</p> <pre>&lt;script&gt;document.write('Hello World');&lt;/script&gt;</pre> <p>(2) 加载外部的 js 文件在 HTML 页面开发编写之中，为了更好的明确程序逻辑关系，以及为了更好地实现代码复用，我们可</p> |

|    |   |   |
|----|---|---|
| 10 | <p>此处有 60 字相似</p> <p>请求伪造、页面重定向以及跨站脚本攻击等等，本文接下来简要的介绍一下JavaScript攻击常用的漏洞。</p> <p>(1) XSS</p> <p>注入式攻击</p> <p>跨站脚本[34] ( Cross-site scripting , 通常简称为 XSS ) 是利用漏洞的恶意行为之一。</p> <p>攻击者将首先检测站点存在的可利用漏洞，然后尝试在页面的输入框内提交特殊的用户文本，多为变形后的JavaScript代码，</p>       | <p>JavaScript脚本恶意代码检测系统 赵梓旭 - 《哈尔滨工业大学硕士论文》 - 2016-06-01 ( 是否引证：否 )</p> <p>1.些漏洞包括脚本挟持、Cookie 篡改、URL 重定向以及 XSS 注入式攻击。2.2.5.1 XSS 注入式攻击跨站脚本 ( Cross-site scripting , 通常简称为 XSS ) [36]是利用漏洞的恶意行为之一。恶意进攻人员将 Java Script 代码植入用户加载的程序中，在传输过程中，很大一部分情况下不会被检测和编码</p> |
| 11 | <p>此处有 66 字相似</p> <p>用户浏览的网页上执行，达到攻击目的。XSS主要有反射型 XSS、存储型XSS、基于DOM的XSS三种。</p> <p>(2) URL</p> <p>重定向</p> <p>URL 重定向[35] ( URL redirection , 或称网址重定向或网域名称转址 ) 的概念是指当使用者点击的是某个网址，</p> <p>但是进入该网址对应的页面后又会被转入其他的页面。攻击者经常利用该漏洞来进行网络钓鱼、发送匿名邮件等操作来让用户在毫无察觉</p> | <p>JavaScript脚本恶意代码检测系统 赵梓旭 - 《哈尔滨工业大学硕士论文》 - 2016-06-01 ( 是否引证：否 )</p> <p>1.攻中威胁最强，威胁范围最大。哈尔滨工业大学工程硕士学位论文 - 16 -2.2.5.2 URL 重定向URL 重定向 ( URL redirection , 或称网址重定向或网域名称转址 ) [39]的概念是指当使用者浏览某个网址时，将他导向到另一个网址的技术。这个行为经常被攻击者利用，产生很多恶意行为。比较常见的如钓鱼、木马、群发邮件</p>    |
| 12 | <p>此处有 33 字相似</p> <p>的页面后又会被转入其他的页面。攻击者经常利用该漏洞来进行网络钓鱼、发送匿名邮件等操作来让用户在毫无察觉和准备的情况下通过</p> <p>恶意链接进入恶意网页。通过将原域名、路径和恶意域名重定向到路径后，</p> <p>用户就会访问恶意网页，导致个人信息的泄露。图 2-2为一种 URL 重定向的方法。</p> <p>图 22 重定向JavaScript</p>   | <p>JavaScript脚本恶意代码检测系统 赵梓旭 - 《哈尔滨工业大学硕士论文》 - 2016-06-01 ( 是否引证：否 )</p> <p>1.为经常被攻击者利用，产生很多恶意行为。比较常见的如钓鱼、木马、群发邮件等。用户往往在不知情的条件下被跳转到恶意链接。它可以将原域名、路径和恶意域名、路径进行跳转，进而诱骗用户跳转恶意 URL，并盗取用户信息，进而进行更严重的恶意行为。图 2-5 为一种 URL 重定向的方法</p>                                    |
| 13 | <p>此处有 44 字相似</p> <p>的情况下通过恶意链接进入恶意网页。通过将原域名、路径和恶意域名重定向到路径后，用户就会访问恶意网页，导致个人信息的泄露。</p> <p>图 2-2为一种 URL 重定向的方法。</p> <p>图 22 重定向JavaScript脚本</p> <p>页面跳转向指定的新链接，这些链接存储在 window.location、document.location、documen</p>                                    | <p>JavaScript脚本恶意代码检测系统 赵梓旭 - 《哈尔滨工业大学硕士论文》 - 2016-06-01 ( 是否引证：否 )</p> <p>1.、路径和恶意域名、路径进行跳转，进而诱骗用户跳转恶意 URL，并盗取用户信息，进而进行更严重的恶意行为。图 2-5 为一种 URL 重定向的方法。图 2-5 重定向 Java Script 脚本页面跳转向指定的新链接，这些链接存储在 document.location、</p>                                      |
| 14 | <p>此处有 122 字相似</p> <p>。图 22 重定向JavaScript脚本</p> <p>页面跳转向指定的新链接，这些链接存储在 window.location、</p>   | <p>JavaScript脚本恶意代码检测系统 赵梓旭 - 《哈尔滨工业大学硕士论文》 - 2016-06-01 ( 是否引证：否 )</p> <p>1.跳转向指定的新链接，这些链接存储在 document.location、document.URL 或 document.referrer 等 DOM 元素的属性值中。例如访问链接</p>  |

|    |   |  |
|----|---|--|
|    | document.location、document.referrer等DOM元素的属性值中。例如访问链接<br>: <a href="https://baidu.com/redirect.jsp?url=http://www.attacks.com">https://baidu.com/redirect.jsp?url=http://www.attacks.com</a> 后, 请求将会被自动转发到url后面对应的恶意网址, 即<br><a href="http://www.attacks.com">http://www.attacks.com</a> 。更严重的是, 攻击者将url进行base64等   | : <a href="http://localhost:8080/redirect.jsp?url=http://www.attack.site">http://localhost:8080/redirect.jsp?url=http://www.attack.site</a> 后, 请求将会被自动转发到对应的恶意网站。<br>URL 重定位的攻击行为较 Server 端重定向攻击隐蔽性更高, 恶意 URL 不会被 Cli  |
| 15 | <p>此处有 171 字相似</p> <p>页面的JSON数据为己用, 来帮助自己完成恶意攻击。具体的攻击过程如图 2-3 所示。</p> <p>图 23 CSRF攻击过程</p> <p>以</p> <p>用户跳转 <a href="http://www.normal.com">http://www.normal.com</a> 为例, 说明恶意脚本 JavaScript 劫持攻击步骤:</p> <p>(1) 用户输入目标 URL 链接 <a href="http://www.normal.com">http://www.normal.com</a>, 登录网站。</p> <p>(2) 用户输入正确的验证信息进行身份验证, 身份验证成功后, 目标网站返回对应的临时访问信息 Cookie, 并被浏览器存储。</p> <p>(3) 用户</p> <p>保持<a href="http://www.normal.com">http://www.normal.com</a>页面登陆的情况下访问了含有恶意脚本的网站<a href="http://www.attac">http://www.attac</a></p> | <p>JavaScript脚本恶意代码检测系统 赵梓旭 - 《哈尔滨工业大学硕士论文》- 2016-06-01 (是否引证: 否)</p> <p>1.PI, 取得返回的 JSON 结果。如图 2-6 所示。图 2-6 Java Script 劫持以用户跳转 <a href="http://www.tmall.com">http://www.tmall.com</a> 为例, 说明恶意脚本 Java Script 劫持攻击步骤: (1) 用户输入目标 URL 链接 <a href="http://www.tmall.com">http://www.tmall.com</a>, 登录网站。(2) 用户输入正确的验证信息进行身份验证, 身份验证成功后, 目标网站返回对应的临时访问信息 Cookie, 并被浏览器存储。(3) 用户跳转 URL : <a href="http://www.malicious.com">http://www.malicious.com</a> (4) URL 请求跳转 <a href="http://www.tmal">http://www.tmal</a></p>  |
| 16 | <p>此处有 187 字相似</p> <p>。进而实现了 JavaScript 脚本劫持技术。</p> <p>图 24 JavaScript劫持攻击的恶意脚本代码</p> <p>(4) 基于</p> <p>二维码进行隐藏攻击</p> <p>二维条码/二维码[36] (2-dimensional bar code) 是用某种特定的几何图形按一定规律在平面 (二维方向上) 分布的黑白相间的图形记录数据符号信息, 在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念, 使用若干个与二进制相对应的几何形体来表示文字数值信息, 通过图象输入设备或光电扫描设备自动识读以实现信息自动处理。</p> <p>QR 码[17] (Quick Response Code) 是二维条码的一种, 其内容可以被移动设备等快速被解码, QR码具有</p>   | <p>二维码怎样制作_欢乐214122 - 《网络 ( <a href="http://blog.sina.com">http://blog.sina.com</a> ) 》- (是否引证: 否)</p> <p>1.想要说的话,请给小编评论!如果你认为本经验可能对你的朋友有帮助,请点击页面右下方的箭头分享给你的朋友! 四、相关信息: 二维码介绍: 二维条码/二维码 (2-dimensional bar code) 是用某种特定的几何图形按一定规律在平面 (二维方向上) 分布的黑白相间的图形记录数据符号信息的; 在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念, 使用若干个与二进制相对应的几何形体来表示文字数值信息, 通过图象输入设备或光电扫描设备自动识读以实现信息自动处理: 它具有条码技术的一些共性: 每种码制有其特定的字符集; 每个字符占有一定的宽度; 具有一定的校验功能等。同时还具有对不同行的</p> <p>河北省建二维码注册解析服务中心 揭其背后“玄机”_播报天下 - 《网络 ( <a href="http://www.gywb.cn/c">http://www.gywb.cn/c</a> ) 》- (是否引证: 否)</p> <p>1.注册解析认证应用示范工作, 联合多家单位组织开展的“中国二维码产业促进工程标准规范应用全国行”的系列活动之一。一起来了解二维码 定义 二维条码/二维码 (2-dimensional bar code)是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的; 在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念, 使用若干个与二进制相对应的几何形体来表示文字数值信息, 通过图象输入设备或光电扫描设备自动识读以实现信息自动处理: 它具有条码技术的一些共性: 每种码制有其特定的字符集; 每个字符占有一定的宽度; 具有一定的校验功</p> |



|  |  |   |
|--|--|---|
|  |  | 能等。同时还具有对不同行的   |
|  |  | 基于互联网营销的B&R品牌推广策略研究 李东敏 - 《华东理工大学硕士论文》 - 2015-04-08 ( 是否引证：否 )  |
|  |  | 1.10万人;而且2014年还获得了5000万美元的投资,成为移动医疗投资的典型性案例,并在投资市场掀起了巨大的风波。(8) 二维码营销二维条码/二维码(2-diinensional bar code)是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的[421;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理,它具有条码技术的一些共性,每种码制有其特定的字符集,每个字符占有一定的宽度,具有一定的校验功能等。同时还具有对不同行的 |
|  |  | 1988087_变电站智能巡检系统设计与应用 XXX论文排版 2014.05.02 变电站智能巡检系统设计与应用 - 《学术论文联合比对库》 - 2015-05-02 ( 是否引证：否 )  |
|  |  | 1.检终端二维码关联设备台账信息,移动终端数据库开发,以及相关技术在智能巡检系统中的应用。3.1 二维码3.1.1 二维码原理二维条码/二维码(2-dimensional bar code)是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的           |
|  |  | 20111103178送审--变电站智能巡检系统设计与应用0505 - 《学术论文联合比对库》 - 2015-05-05 ( 是否引证：否 )   |
|  |  | 1.检终端二维码关联设备台账信息,移动终端数据库开发,以及相关技术在智能巡检系统中的应用。3.1 二维码3.1.1 二维码原理二维条码/二维码(2-dimensional bar code)是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的           |
|  |  | 基于JavaEE的设备管理系统 1247172270-张玉 - 《学术论文联合比对库》 - 2015-09-14 ( 是否引证：否 )   |
|  |  | 1.ervlet展现页面。后续操作不需要继续解析和编译可执行代码,也大大提高了JVM运行的速度和效率[27]。<br>2.11 二维码概述二维条码/二维码(2-dimensional bar code)是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的  |

|  |  |   |
|--|--|---|
|  |  | <p>；在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集；每个字符占有一定的宽度；具有一定的校验功能等。同时还具有对不同行的</p>  |
|  |  | <p>计控学院2220140267董阳 董阳 - 《学术论文联合比对库》 - 2015-10-10 ( 是否引证：否 )</p>  |
|  |  | <p>1.主宰者[]。可以说，HTML5是WEB的未来，不仅在电脑端，而且在移动端也一定会得到广泛的应用[]。2.4.3 二维码技术二维条码/二维码 ( 2-dimensional bar code ) 是用某种特定的几何图形按一定规律在平面 ( 二维方向上 ) 分布的黑白相间的图形记录数据符号信息的 []；在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集；每个字符占有一定的宽度；具有一定的校验功能等。同时还具有对不同行的</p> |
|  |  | <p>&amp;#34;二维码&amp;#34;、&amp;#34;扫码关注&amp;#34;用英语怎么说-搜狐教育 - 《网络 ( <a href="http://learning.sohu">http://learning.sohu</a> ) 》 - ( 是否引证：否 )</p>   |
|  |  | <p>1.是越来越高。微信二维码、支付二维码基本天天会用到，但你知道用英语该怎么说吗？首先，让我们看看什么是二维码：二 维 码 二维条码/二维码(2-dimensional bar code)是用某种特定的几何图形按一定规律在平面 (二维方向上)分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的</p>             |
|  |  | <p>94_张悦_基于RFID组网分割的大棚温度监测方法研究 张悦 - 《学术论文联合比对库》 - 2015-05-02 ( 是否引证：否 )</p>   |
|  |  | <p>1.文章进行整体总结，重点是本文的前景展望，在未来结合高校不同领域的需求，增加二维码管理系统的功能更好的为高校师生服务。二维条码/二维码 ( 2-dimensional bar code ) 是用某种特定的几何图形按一定规律在平面 ( 二维方向上 ) 分布的黑白相间的图形记录数据符号信息的；在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念[4]，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集；每个字符占有一定的宽度；具有一定的校验功能等。同时还具有对不同行的信</p>   |

|  |   |
|--|---|
|  | <p>Android手机僵尸网络技术及防御方法研究 王钰淇 - 《东南大学硕士论文》 - 2015-10-21 (是否引证: 否)</p> <p>1.球手机安全报告》显示,2012年上半年查杀到手机恶意软件H676款,而其中二维码技术成为手机病毒、钓鱼网站传播的新渠道。<b>二维条码/二维码(2-dimensional bar code)</b>是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信</p> <p>优惠券分发系统的设计与实现 刘正刚 - 《学术论文联合比对库》 - 2015-11-16 (是否引证: 否)</p> <p>1.SDK可以开发适用于移动设备的地图应用,通过接口,可以轻松访问百度服务和数据,构建功能丰富、交互性强的地图应用程序。<b>二维条码/二维码 ( 2-dimensional bar code )</b>是用某种特定的几何图形按一定规律在平面 ( 二维方向上 ) 分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的</p> <p>跨行二维码支付方案研究 赵东林; - 《2016年第二届今日财富论坛论文集》 - 2016-05-03 (是否引证: 否)</p> <p>1.额是未来银行的立身之本。本文研究二维码支付模式,探讨实现跨行二维码支付的方法和路径。二、认识二、维码(一)二维码及其特点<b>二维条码/二维码(2-dimensional bar code)</b>是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信</p> <p>密集型区域中物流终端的安全分析及解决策略 杨兴波;张林; - 《科技视界》 - 2016-05-15 (是否引证: 否)</p> <p>1.端的应用而产生并应用到电子商务相关领域的新技术,顿时,“扫一扫”为代表的应扑而来,充斥在生活的方方面面。<b>二维码/二维条码(2-dimensional bar code)</b>是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理;它具有条码技术的一些</p> |
|--|---|

|  |  |
|--|--|
|  | <p>共性,每种码制有其特定的字符集,每个字符占有一定的宽度,具有一定的校验功能等。同时还具有对不同行的信</p>  |
|  | <p>二维码扫一扫手机就会中毒是真的吗? - 《互联网文档资源 ( <a href="http://www.360doc.co">http://www.360doc.co</a> ) 》 - 2016-11-21 10:34:22 ( 是否引证: 否 )</p>  |
|  | <p>1.类似的活动么?( 图片来源网络 ) 二维码是啥?它安全么? 要弄明白问题,我们先得来看看“二维码”这个家伙到底是什么来的? <b>二维条码/二维码 ( 2-dimensional bar code )</b> 是用某种特定的几何图形按一定规律在平面 ( 二维方向上 ) 分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的</p> |
|  | <p>二维码技术在高校教学信息化中的应用实践探索 宋艳艳;颜月南;-《现代经济信息》- 2016-11-25 ( 是否引证: 否 )</p>   |
|  | <p>1.引言 <b>二维条码/二维码(2-dimensional bar code)</b>是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信</p>   |
|  | <p>物联网+智能停车平台项目计划书(高校版)_图文 - 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》 - 2016/12/27 14:14:58 ( 是否引证: 否 )</p>  |
|  | <p>1.口或DVI接口的画面及视频显示信息,置入帧存储器,按分区驱动方式生成LED显示屏所能识别的串行显示数据和扫描控制时序。 <b>二维条码V二维码 ( 2-dimensional bar code )</b> : 是用某种特定的几何图形按一定规律在平面 ( 二维方向上 ) 分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的</p> |
|  | <p>激光直接标记二维码(Data Matrix)及识别用户操作系统设计_图文 - 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》 - 2016/8/10 11:40:30 ( 是否引证: 否 )</p>  |
|  | <p>1.条码并通过二维条码实现数据的自动交换。 5. 银行票据管理。 6. 行包、货物的运输和邮递。 1.3 课题研究的意义 <b>二维条码V二维码 ( 2-dimensional bar code )</b> 是用某种特定的几何图形按一定规律在平面 ( 二维方向上 ) 分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、</p>   |



|  |   |
|--|---|
|  | <p>“1”比特流的概念，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集；每个字符占有一定的宽度；具有一定的校验功能等。同时还具有对不同行的</p>  |
|  | <p>湖南现代物流职业技术学院 湖南现代 湖湖湖湖南湖南现代物流职业业技术学院 湖湖湖湖代湖 湖湖湖... - 《互联网文档资源 ( <a href="http://www.worlduc.c">http://www.worlduc.c</a> ) 》 - 2016/8/9 4:27:19 ( 是否引证：否 )</p>  |
|  | <p>1.地提供点、线、面要素的精确三维坐标以及其他相关信息，GPS技术可以极大提高相对定位精度，在小范围内可以达到厘米级精度。 二维条码/二维码 ( 2-dimensional bar code ) 是用某种特定的几何图形按一定规律在平面 ( 二维方向上 ) 分布的黑白相间的图形记录数据符号信息的；在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集；每个字符占有一定的宽度；具有一定的校验功能等。同时还具有对不同行的</p> |
|  | <p>二维码在建筑工程物资管理过程中的应用 丁立欢;甯市鹏; - 《《工业建筑》2016年增刊II》 - 2016-06-20 ( 是否引证：否 )</p>  |
|  | <p>1.码,扫一扫即可进行信息的传递。鉴于二维码如此之多的应用,相信在建筑行业中的发展也是大有可为。1二维码定义二维条码/二维码是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的;在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信</p>  |
|  | <p>我们身边的条形码 - 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》 - 2016-11-1 13:14:25 ( 是否引证：否 )</p>  |
|  | <p>1.，和为125；将125除以10，取余数，即5；用10减去5，得到的数值，就是校验码，即本例中的第4段，5。 二维条码 它是用某种特定的几何图形按一定规律在平面 ( 二维方向上 ) 分布的黑白相间的图形记录数据符号信息的；在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集；每个字符占有一定的宽度；具有一定的校验功能等。同时还具有对不同行的</p>                                     |
|  | <p>QR码在Android系统中的研究与应用 阚钊玉; - 《网络安全技术与应用》 - 2016-03-15 ( 是否引证：否 )</p>  |

|    |   |  |
|----|---|--|
|    |   | <p>1.一个近几年来移动设备上超流行的一种编码方式,它比传统的Bar Code条形码能存更多的信息,也能表示更多的数据类型。它用某种特定的几何图形按一定的规律在平面上分布黑白相间的图形记录数据符号信息,在代码编制上巧妙地利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等,同时还具有对不同行的信</p>  |
| 17 | <p>此处有 82 字相似</p> <p>QR 码[17] ( Quick Response Code ) 是二维条码的一种,其内容可以被移动设备等快速被解码,QR码具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。和传统的</p> <p>条形码相比,QR码具有存储信息更多,识别更加快捷简单邓特点,因此在移动互联网时代得到了迅速的普及和应用。</p> <p>正是在这样的</p> | <p>&amp;#34;二维码&amp;#34;、&amp;#34;扫码关注&amp;#34;用英语怎么说-搜狐教育 - 《网络 ( <a href="http://learning.sohu">http://learning.sohu</a> ) 》 - ( 是否引证: 否 )</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。 二维码概念太抽象了,太难懂了,有木有!小编看着都觉得累,话虽这样说,但标准的英文表达还木学会,真的有点不甘心啊。“二</p> <p>二维码怎样制作_欢乐214122 - 《网络 ( <a href="http://blog.sina.com">http://blog.sina.com</a> ) 》 - ( 是否引证: 否 )</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。</p> <p><a href="http://www.wm23.cn/214122">http://www.wm23.cn/214122</a> 我的网络营销能力秀主页</p> <p><a href="http://blog.sina.com">http://blog.sina.com</a>.</p> <p>河北省建二维码注册解析服务中心 揭其背后“玄机”_播报天下 - 《网络 ( <a href="http://www.gywb.cn/c">http://www.gywb.cn/c</a> ) 》 - ( 是否引证: 否 )</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。 功能 ●信息获取(名片、地图、WIFI密码、资料) ●网站跳转(跳转到微博、手机网站、网站) ●广告推送(用户扫码,直</p> <p>基于互联网营销的B&amp;R品牌推广策略研究 李东敏 - 《华东理工大学硕士论文》 - 2015-04-08 ( 是否引证: 否 )</p> <p>1.的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理,它具有条码技术的一些共性,每种码制有其特定的字符集,每个字符占有一定的宽度,具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。随着互联网及智能手机的飞速发展,3G网络及4G网络也日益普及。因此,具</p> |

|  |  |   |
|--|--|---|
|  |  | 有拍照功能的智能手机通过对二维码的扫描,可以让消  |
|  | 94 张悦 基于RFID组网分割的大棚温度监测方法研究 张悦 - 《学术论文联合比对库》 - 2015-05-02 ( 是否引证 : 否 ) | 1.], 使用若干个与二进制相对应的几何形体来表示文字数值信息, 通过图象输入设备或光电扫描设备自动识读以实现信息自动处理: 它具有条码技术的一些共性: 每种码制有其特定的字符集; 每个字符占有一定的宽度; 具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。[5]二维码作为一种信息容量大、可标识文字网址等多种信息、成本低廉的自动识别技术, 目前已经在我国众多行业取得规模化应      |
|  | 基于JavaEE的设备管理系统 1247172270-张玉 - 《学术论文联合比对库》 - 2015-09-14 ( 是否引证 : 否 )  | 1.概念, 使用若干个与二进制相对应的几何形体来表示文字数值信息, 通过图象输入设备或光电扫描设备自动识读以实现信息自动处理: 它具有条码技术的一些共性: 每种码制有其特定的字符集; 每个字符占有一定的宽度; 具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。ZXing是一个出自Google的, 开放源码的, 用Java实现的多种格式的1D/2D条码图像处理库, 它包含了联系到其他语 |
|  | 计控学院2220140267董阳 董阳 - 《学术论文联合比对库》 - 2015-10-10 ( 是否引证 : 否 )            | 1.概念, 使用若干个与二进制相对应的几何形体来表示文字数值信息, 通过图象输入设备或光电扫描设备自动识读以实现信息自动处理: 它具有条码技术的一些共性: 每种码制有其特定的字符集; 每个字符占有一定的宽度; 具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。二维码技术具备以下特性: 1. 高密度编码, 信息容量大: 可容纳多达1850个大写字母或2710个数字或1108个字节,   |
|  | Android手机僵尸网络技术及防御方法研究 王钰淇 - 《东南大学硕士论文》 - 2015-10-21 ( 是否引证 : 否 )      | 1.的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。二维码具有W下特点(1)高密度编码,信息容量大:可容纳多达1850个大写字母或2710个数字或1108个字节,或500多          |
|  | 优惠券分发系统的设计与实现 刘正刚 - 《学术论文联合比对库》 - 2015-11-16 ( 是否引证 : 否 )              | 1.概念, 使用若干个与二进制相对应的几何形体来表示文字数值信息, 通过图象输入设备或光电扫描设备自动识读以实现信息自动处理: 它具有条码技术的一些共性: 每种码制有其特定的字符集; 每个字符占有一定的宽度; 具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点。我国对二维   |

|  |  |  |
|--|--|--|
|  |  | <p>码技术的研究开始于1993年。中国物品编码中心对几种常用的二维码PDF417、QRCCode、Data Mat</p> <p>QR码在Android系统中的研究与应用 阚钊玉; - 《网络安全技术与应用》 - 2016-03-15 ( 是否引证 : 否 )</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等;同时还具有对不同行的信息自动识别功能、及处理图形旋转变化点。随着智能手机的发展,手机与二维码进行结合,进一步拓展了二维码应用价值,促进了行业的融合,手机二维码将为通信、媒体以及其他</p> <p>跨行二维码支付方案研究 赵东林; - 《2016年第二届今日财富论坛论文集》 - 2016-05-03 ( 是否引证 : 否 )</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变化点。二维码具有如下优点:1.高密度编码,信息容量大:可容纳多达1850个大写字母或2710个数字或1108个字节,或500多</p> <p>密集型区域中物流终端的安全分析及解决策略 杨兴波;张林; - 《科技视界》 - 2016-05-15 ( 是否引证 : 否 )</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理;它具有条码技术的一些共性,每种码制有其特定的字符集,每个字符占有一定的宽度,具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变化点。[3]2.2二维码的特点及应用在物联网时代,人们选择更便捷、更安全的方式来为自己的生活服务。相比较条码和RFID,二维码</p> <p>二维码在建筑工程物资管理过程中的应用 丁立欢;窦市鹏; - 《《工业建筑》2016年增刊II》 - 2016-06-20 ( 是否引证 : 否 )</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变化点。现在二维码主要应用于2.二维码的特点2.1高密度编码,信息容量大可容纳多达1 850个大写字母或2 710个数字或1 10</p> <p>二维码扫一扫手机就会中毒是真的吗? - 《互联网文档资源 ( <a href="http://www.360doc.co">http://www.360doc.co</a> ) 》 - 2016-11-21 10:34:22 ( 是否引证 : 否 )</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信</p> |
|--|--|--|



|  |  |
|--|--|
|  | <p>息自动识别功能、及处理图形旋转变点。二维码条码可以容纳1848个字母字符或2729个数字字符，约500个汉字信息。这种二维条码比普通条码信息容量高几十倍。</p>   |
|  | <p>二维码技术在高校教学信息化中的应用实践探索 宋艳艳;颜月南; - 《现代经济信息》 - 2016-11-25 (是否引证: 否)</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变点。国外对二维码技术的研究起于20世纪80年代末,其信息密度比传统的一维码有很大的提高。二维码作为一种全新的信息存储、传递和</p>   |
|  | <p>物联网+智能停车平台项目计划书(高校版) 图文 - 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》 - 2016/12/27 14:14:58 (是否引证: 否)</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变点。2.3服务器端 图2-16服务器端系统结构图 2.3.1系统功能需求 停车位信息管理子后台: 1、车位停车信息显示: 所</p>                       |
|  | <p>激光直接标记二维码(Data Matrix)及识别用户操作系统设计 图文 - 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》 - 2016/8/10 11:40:30 (是否引证: 否)</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变点。2 二维码的出现是为了解决一维条码无法解决的问题而产生。因而它具有高密度、高可靠性等特点,所以可以用它表示数据文件(包括汉字文件)、</p>     |
|  | <p>湖南现代物流职业技术学院 湖南现代 湖湖湖湖南湖南现代物流职业业技术学院 湖湖湖湖代湖 湖湖湖... - 《互联网文档资源 ( <a href="http://www.worlduc.c">http://www.worlduc.c</a> ) 》 - 2016/8/9 4:27:19 (是否引证: 否)</p> <p>1.概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理:它具有条码技术的一些共性:每种码制有其特定的字符集;每个字符占有一定的宽度;具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变点。目前,安卓开发技术、GPS二维码等技术已经很成熟了,所以要开发这个软件在技术上是可行的。5 第2章 总体设计 2.1</p> |
|  | <p>1988087_变电站智能巡检系统设计与应用_XXX论文排版 2014.05.02 变电站智能巡检系统设计与应用 - 《学术论文</p>  |

|    |   |   |
|----|---|---|
|    |   | <p>联合比对库》- 2015-05-02 (是否引证：否)</p> <p>1.概念，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集；每个字符占有一定的宽度；具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点[19]。3.1.2 二维码发展历程从上个世纪的80年代末开始，国际上已经开始了二维码技术的研究工作，现已开发研制出了多种码制</p> <p>20111103178送审--变电站智能巡检系统设计与应用0505 -《学术论文联合比对库》- 2015-05-05 (是否引证：否)</p> <p>1.概念，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集；每个字符占有一定的宽度；具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换点[19]。3.1.2 二维码发展历程从上个世纪的80年代末开始，国际上已经开始了二维码技术的研究工作，现已开发研制出了多种码制</p> <p>我们身边的条形码 -《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》- 2016-11-1 13:14:25 (是否引证：否)</p> <p>1.概念，使用若干个与二进制相对应的几何形体来表示文字数值信息，通过图象输入设备或光电扫描设备自动识读以实现信息自动处理：它具有条码技术的一些共性：每种码制有其特定的字符集；每个字符占有一定的宽度；具有一定的校验功能等。同时还具有对不同行的信息自动识别功能、及处理图形旋转变换等。 二维条码的特点 1. 信息容量大 2. 编码范围广 3. 保密防伪性能好 4. 译码可靠性高 5. 修改错误能力强 6. 成本低</p> |
| 18 | <p>此处有 494 字相似</p> <p>著，自然地，网络安全工作者开始尝试将这些技术引入到Web安全领域。接下来，本文将介绍自动分类和机器学习中常用的一些技术和</p> <p>算法。</p> <p>3 分类算法介绍</p> <p>分类学习算法是根据已知样本集特征分布规律来对未知样本进行分类预测的机器学习算法。分类学习算法的一般过程是首先对已有的数据集进行特征抽取及学习，并在此基础上训练出分类模型，然后将该分类模型应用到新的样本类型预测中。分类学习算法分为分类模型的训练和已生成模型的测试两个阶段：分类模型的训练和对已生成分类模型的测试。在分类模型训练阶段，首先从待分类样本中抽取能够明显区分不同类别的属性集作为分类算法训练的特征向量集，并根据样本中属性特征分布规律生成样本分类模型；在分类模型测试阶段，使用已生成的分类模型来预测未知样本的类型，并计算其分类结果的准确率及召回率，根据分类结果对分类模型进行参数调整和优化。为了能够准确地对恶意网页进行类别预测与判定，我们分别使用了朴素贝叶斯算法、决策</p> | <p>王维光_2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1. 浏览器和用户安全意识四个方面给出了防御网页恶意代码攻击的有效措施。第三章分类学习算法概述15第三章分类学习 算法概述分类学习算法是根据已知样本集特征分布规律来对未知样本进行分类预测的机器学习算法。分类学习算法的一般过程是首先对已有的数据集进行特征抽取及学习，并在此基础上训练出分类模型，然后将该分类模型应用到新的样本类型预测中。分类学习算法分为分类模型的训练和已生成模型的测试两个阶段：分类模型的训练和对已生成分类模型的测试。在分类模型训练阶段，首先从待分类样本中抽取能够明显区分不同类别的属性集作为分类算法训练的特征向量集，并根据样本中属性特征分布规律生成样本分类模型；在分类模型测试阶段，使用已生成的分类模型来预测未知样本的类型，并计算其分类结果的准确率及召回率，根据分类结果对分类模型进行参数调整和优化。为了能够准确地对恶意网页进行类别预测与判定，我们分别使用了朴素贝叶斯算法、决策树 C4.5 算法、分类回归树算法</p>  |

|   |  |
|---|--|
| <p>树C4.5算法、分类回归树算法 ( CART )、支持向量机算法 ( SVM ) 进行社交网络短文本和恶意网页类型的分类判定。本章介绍了四种分类学习算法的基本概念、核心理论、一般步骤及各分类算法的优缺点和适用场景，并给出了评价分类算法结果的具体参数及意义。</p> <p>3.1 朴素贝叶斯算法</p> <p>朴素贝叶斯算法</p> <p>[37]是基于贝叶斯定理与特征条件独立假设的分类方法，其核心是根据贝叶斯定理来对一个未知样本属于某个类别的概率进行预测分</p> | <p>( CART )、支持向量机算法 ( SVM ) 来进行恶意网页类型的分类判定。本章系统的介绍了四种分类学习算法的基本概念、核心理论、一般步骤及各分类算法的优缺点和适用场景，并给出了评价分类算法结果的具体参数及意义。3.1 朴素贝叶斯算法贝叶斯分类算法[29]是一种基于贝叶斯定理的统计学分类算法，其主要思想是根据贝叶斯定理来预测一个未知类别的样本属于各个类别的概率</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p> <p>1.全意识四个方面给出了防御网页恶意代码攻击的有效措施。14第三章分类学习算法概述 第三章分类学习算法概述分类学习算法是根据已知样本集特征分布规律来对未知样本进行分类预测的机器学习算法。分类学习算法的一般过程是首先对已有的数据集进行特征抽取及学习，并在此基础上训练出分类模型，然后将该分类模型应用到新的样本类型预测中。分类学习算法分为分类模型的训练和已生成模型的测试两个阶段：分类模型的训练和对已生成分类模型的测试。在分类模型训练阶段，首先从待分类样本中抽取能够明显区分不同类别的属性集作为分类算法训练的特征向量集，并根据样本中属性特征分布规律生成样本分类模型；在分类模型测试阶段，使用已生成的分类模型来预测未知样本的类型，并计算其分类结果的准确率及召回率，根据分类结果对分类模型进行参数调整和优化。为了能够准确地对恶意网页进行类别预测与判定，我们分别使用了朴素贝叶斯算法、决策树C4.5算法、分类回归树算法 ( CART)、支持向量机算法 ( SVM)进行恶意网页类型的分类判定。本章介绍了四种分类学习算法的基本概念、核心理论、一般步骤及各分类算法的优缺点和适用场景，并给出了评价分类算法结果的具体参数及意义。3.1朴素贝叶斯算法贝叶斯分类算法[29]是一种基于贝叶斯定理的统计学分类算法，其主要思想是根据贝叶斯定理来预测一个未知类别的样本属于各个类别的概率</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.时候可能需要改变训练分类器时的样本集属性特征向量来适应新的恶意网页检测，这将是我们的后期工作的重点。第七章分类学习算法概述分类学习算法是根据已知样本集特征分布规律来对未知样本进行分类预测的机器学习算法。分类学习算法的一般过程是根据对已有的数据集进行特征抽取及学习，在此基础上训练出分类模型，并将该分类模型应用到新的样本类型预测中。分类学习算法分为分类模型的训练和已生成模型的测试两个阶段：分类模型的训练和对已生成分类模型的测试。在分类模型训练阶段，首先从待分类样本中抽取能够明显区分不同类别的属性集作为分类算法训练的特征向量集，并根据样本中属性特征分布规律生成样本分类模型；在分类模型测试阶段，使用已生成的分类模型来预测未知样本的类型，并计算其分类结果的准确率及召回率，根据分类结果对分类模型进行参数调整和优化。为了能够准确地对恶意网页进行类别预测与判定，我们分别</p> |
|---|--|

|    |  |   |
|----|--|---|
|    |  | <p>使用了朴素贝叶斯算法、决策树 C4.5 算法、分类回归树算法 ( CART )、支持向量机算法 ( SVM ) 来进行恶意网页类型的判定。本章系统的介绍了四种分类学习算法的基本概念，核心理论和一般运行步骤，并给出了评价分类算法结果的具体参数和计算公式。3.1 朴素贝叶斯算法贝叶斯分类算法[29]是一种基于</p> <p>2.Na ve Bayes )、决策树 C4.5 算法、分类回归树算法 ( CART )、及支持向量机算法 ( SVM ) 的概念、基本原理和优缺点，并给出了评价分类学习结果的主要参数标准。( 4 ) 设计和实现了基于分类算法的恶意网页检测系统 ( CA-MWDS )，介绍了该系统的总体结构和 workflows。CA-MWDS</p>   |
| 19 | <p>此处有 88 字相似</p> <p>的分类方法，其核心是根据贝叶斯定理来对一个未知样本属于某个类别的概率进行预测分析，找到概率最大的类别后将未知样本归类到该</p> <p>类别中去。贝叶斯分类算法主要包括两类：一类是基于条件独立性假设的朴素贝叶斯算法，即各属性之间相互独立互不影响；一类是考虑了实际决策中各属性间相互关系的贝叶斯网络。朴素贝叶斯算法</p> <p>具有分类过程简单、原理通俗易懂、稳定性比较高、模型训练快分类迅速等优点，朴素贝叶斯算法的具体原理如下：</p> <p>对于一个未知</p> | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.类算法，其主要思想是根据贝叶斯定理来预测一个未知类别的样本属于各个类别的概率大小，并将该样本分配给具有最大概率的类别。贝叶斯分类算法主要包括两类：一类是基于条件独立性假设的朴素贝叶斯算法 [30]，即各属性之间相互独立，互不影响；一类是考虑了实际决策中各属性间相互关系的贝叶斯网络[31]。朴素贝叶斯算法在许多实际分类决策问题中能够达到复杂分类算法的分类效果，并且该算法具有分类原理简单易于理解，分类模型生成速度快且</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究 2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p> <p>1.类算法，其主要思想是根据贝叶斯定理来预测一个未知类别的样本属于各个类别的概率大小，并将该样本分配给具有最大概率的类别。贝叶斯分类算法主要包括两类：一类是基于条件独立性假设的朴素贝叶斯算法 [30]，即各属性之间相互独立，互不影响；一类是考虑了实际决策中各属性间相互关系的贝叶斯网络[31]。朴素贝叶斯算法在许多实际分类决策问题中能够达到复杂分类算法的分类效果，并且该算法具有分类原理简单易于理解，分类模型生成速度快且</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p> <p>1.法，其主要思想是根据贝叶斯定理来预测一个未知类别的样本属于各个类别的概率大小，并将该样本分配给具有最大概率的类别。贝叶斯分类算法主要包括两类一：类是基于条件独立性假设的朴素贝叶斯算法，即各属性之间相互独立，互不影响一；类是考虑了实际决策中各属性间相互关系的贝叶斯网络[31]。朴素贝叶斯算法在许多实际分类决策问题中能够达到复杂分类算法的分类效果，并且该算法具有分类原理简单易于理解，分类模型生成速度</p> |

|   |
|---|
| 指 标   |
| 疑似剽窃文字表述  |
| 1. 恶意链接后在用户中传播。通常，恶意链接会伪装成正常的网页链接，或者通过正常页面引入，用户访问这些 Web 页 |



- 面时，会触发被攻击的 B 及漏洞，
2. 不同于PHP与ASP等服务器端的脚本语言，JavaScript主要被作为客户端脚本语言在用户的浏览器上运行。
  3. 图 2-2为一种 URL 重定向的方法。
- 图 22 重定向JavaScript脚本
4. document.location、document.referrer等DOM元素的属性值中。例如访问链接  
: <https://baidu.com/redirect.jsp?url=http://www.attacks.com>后，请求将会被自动转发

### 3. 11491915\_刘力铭\_第3部分 总字数：9178

| 相似文献列表 文字复制比：37%(3400) 疑似剽窃观点：(0) |  |                          |
|-----------------------------------|--|--------------------------|
| 1                                 | 基于分类算法的恶意网页检测技术研究<br>王维光(导师：王枏) - 《北京邮电大学硕士论文》 - 2015-03-16  | 25.4% ( 2331 )<br>是否引证：是 |
| 2                                 | 王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15<br>王维光 - 《学术论文联合比对库》 - 2015-01-15  | 22.9% ( 2102 )<br>是否引证：否 |
| 3                                 | 恶意网页医疗数据应急通信移动通信<br>- 《学术论文联合比对库》 - 2015-01-09   | 21.9% ( 2008 )<br>是否引证：否 |
| 4                                 | 于俊_基于等级容灾模型<br>于俊 - 《学术论文联合比对库》 - 2014-12-31   | 20.2% ( 1850 )<br>是否引证：否 |
| 5                                 | 基于决策树的搜索引擎恶意网页检测研究与实现<br>周浩(导师：孙建华) - 《湖南大学硕士论文》 - 2013-05-27  | 6.4% ( 585 )<br>是否引证：否   |
| 6                                 | 谁能诠释下"单实例存储" TechTarget技术社区 - 技术论坛 - TechTarget中国<br>- 《网络 ( <a href="http://club.techtarget">http://club.techtarget</a> ) 》 - | 1.6% ( 147 )<br>是否引证：否   |
| 7                                 | 基于虚拟磁带库的容灾系统设计与实现<br>- 《学术论文联合比对库》 - 2012-11-06  | 1.6% ( 147 )<br>是否引证：否   |
| 8                                 | 基于双布鲁姆过滤器的数据排重算法及其应用<br>席晔文(导师：杨金民;刘绪崇) - 《湖南大学硕士论文》 - 2013-04-23  | 1.6% ( 146 )<br>是否引证：否   |
| 9                                 | 重复数据删除技术在船舶信息化建设中的应用<br>杜阳; - 《天津航海》 - 2014-06-30  | 1.6% ( 146 )<br>是否引证：否   |
| 10                                | 重复数据删除技术的研究与探讨<br>成杰;李恒志;祁悦; - 《中国无线电》 - 2012-10-26  | 1.3% ( 119 )<br>是否引证：否   |
| 11                                | 基于双布鲁姆过滤器的数据排重技术<br>席晔文，杨金民 - 《计算机工程与应用 ( 优先出版 ) 》 - 2013-04-08 16:46  | 1.2% ( 112 )<br>是否引证：是   |
| 12                                | 基于双布鲁姆过滤器的数据排重技术<br>席晔文;杨金民; - 《计算机工程与应用》 - 2013-04-08 1   | 1.1% ( 105 )<br>是否引证：是   |

| 原文内容 |  | 相似内容来源   |
|------|--|--|
| 1    | 此处有 29 字相似<br>然后，我们用<br>朴素贝叶斯分类算法分别计算样本X属于每一个类别的概率大小，<br>我们称做后验概率，将得到的结果按照高低大小排序后，将X分配到具有最高结果的类别C中，即当且仅当 $PC_iX > PC_jX, 1 \leq j$ | 于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )<br>1.中。该算法原理如下：假设每个样本的个属性值以的维向量形式表示，若数据样本集共有个不同的类别。对于给定的某个样本集，朴素贝叶斯分类算法分别计算样本属于每一个类别的后验概率，通过比较后验概率大小后将该样本分配到具有最高后验概率的类别中，即当且仅当时，朴素贝叶斯算法才会将样本集判定为类别。 |
|      |  | 恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )<br>1.n 维向量形式表示，若数据样本集共有m个不同的类别 $mC, C_1, C_2$ 。对于给定的某个样本集 X，朴素贝叶斯分类算法分别计算样本属于每一个类别的后验概率，通过比较后验概率大小后将该样本分配到具有最高后验概率的类别 C 中，即当且仅当时，朴素贝叶斯算法才会将样本集判定  |
|      |  | 王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )<br>1.n 维向量形式表示，若数据样本集共有m个不同的类  |
|      |  |  |

|   |  |  |
|---|--|--|
|   |  | <p>别mC ,C,,C12。对于给定的某个样本集 X , 朴素贝叶斯分类算法分别计算样本属于每一个类别的后验概率 , 通过比较后验概率大小后将该样本分配到具有最高后验概率的类别 C 中 , 即当且仅当PCXPCXjmjijj(</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证 : 是 )</p> <p>1.{1, , 12, ?,1?}的《维向量形式表示, 若数据样本集共有m个不同的类别, ???, 。。?对于给定的某个样本集Z, 朴素贝叶斯分类算法分别计算样本X属于每一个类别的后验概率, 通过比较后验概率大小后将该样本Z分配到具有最高后验概率的类别 ϕ :中, 即当且仅当且时, 朴素贝叶斯算法才会将样本</p>   |
| 2 | <p>此处有 80 字相似</p> <p>, 2 , .. , n。</p> <p>需要注意的是, 在实际应用中, 朴素贝叶斯算法计算未知样本的后验概率时经常会得到多个最高结果, 这表明未知</p> <p>样本的类别特征模糊, 此时, 可以调整或者添加分类特征向量中样本特有属性来减少这种情况的发生。</p> <p>3.2 决策树算法</p> <p>决策树方法[4]是一种逼近离散函数值的方法, 最早可以追溯到19世纪60年代。先是J Ross Quinlan提出了ID3算法, 该算法减少了决策树的深度, 代价是减少对</p> | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证 : 否 )</p> <p>1.别中的后验概率, 并将具有最高后验概率的类别作为该样本的预测类别。当计算的后验概率中同时有两个及两个以上的最高值时, 说明该样本的类别特征并不明显 , 此时, 可以向分类特征向量中添加具有区分度的样本属性来避免这种情况的再次发生。3.2 决策树算法决策树算法[32]是一种常用的树形决策算法, 它利用生成的树形结构, 按一定的分割原则把训练样本集分为特征更为集中的子集, 这些子集即为决策树的内部节点。一</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证 : 否 )</p> <p>1.验概率, 并将具有最高后验概率的类别作为该样本的预测类别。当计算的后验概率中同时有两个及两个以上的最高值时, 说明该样本的类别特征并不明显, 此时 , 可以向分类特征向量中添加具有区分度的样本属性来避免这种情况的再次发生。3.2 决策树算法决策树算法 [32]是一种常用的树形决策算法, 它按一定的划分规则将训练样本集分成属性特征更为集中的子节点, 并最终生成一棵分类决策树, 然后可以使用</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证 : 否 )</p> <p>1.验概率, 并将具有最高后验概率的类别作为该样本的预测类别。当计算的后验概率中同时有两个及两个以上的最高值时, 说明该样本的类别特征并不明显, 此时 , 可以向分类特征向量中添加具有区分度的样本属性来避免这种情况的再次发生。3.2 决策树算法决策树算法 [32]是一种常用的树形决策算法, 它按一定的划分规则将训练样本集分成属性特征更为集中的子节点, 并最终生成一棵分类决策树, 然后可以使用</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证 : 是 )</p> <p>1.率, 并将具有最高后验概率的类别作为该样本的预测类别。当计算的后验概率中同时有两个及两个以上的最高值时, 说明该样本的类别特征并不明显, 此时, 可以向分类特征向量中添加具有区分度的样本属性以避免这种情况的再次发生。3.2决策树算法决策树算法[32]是一种常用的树形决策算法, 它按一定的划分规则将训练样</p> |

|   |  |  |
|---|--|--|
|   |  | 本集分成属性特征更为集中的子节点，并最终生成一棵分类决策树，然后可以   |
| 3 | <p>此处有 57 字相似</p> <p>后的C4.5算法针对ID3的问题预测变量的缺值处理、剪枝技术、派生规则等方面做了较大既适合于分类问题，又适合于回归问题。</p> <p>1966年 Hunt. E. B等人提出的概念学习系统 [39](Concept Learning System, CLS)得到了极大的普及和应用，现在流行的决策树算法基本是对CLS的提升和改进。</p> <p>CLS 的核心原理为：决策树其实就利</p>        | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》 - 2013-05-27 (是否引证：否)</p> <p>1.浏览器的负荷，不利于检测的效率。2.4.4 决策树算法 决策树是目前分类应用中广泛采用的模型之一，源于 1966 年 Hunt. E. B[38]等人提出的概念学习系统 CLS(concept learning system)，现在众多的决策树算法都可以看作是 CLS 的优化和改进。CLS 算法的大体思想是：初始为空的一个决策树，不断</p> |
| 4 | <p>此处有 65 字相似</p> <p>利用该节点的分类规则进行分类验证，直到达到最终的叶子节点，此时决定出的分类决策就是该样本所属的类别。构造决策树时首先预置一个空的决策树，决策时不断向其中添加新的叶子节点作为分类节点，重复该过程，直到构造出的决策树可以将训练集样本正确的分好类为止。</p> <p>ID3和C4.5是经典而常用的决策树算法，同时，针对决策树算法存在的不足也有很多改进算法，分类回归树算法是典型代表。接下</p>         | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》 - 2013-05-27 (是否引证：否)</p> <p>1.ng system)，现在众多的决策树算法都可以看作是 CLS 的优化和改进。CLS 算法的大体思想是：初始为空的一个决策树，不断往其中加入新的判决节点，该判决节点是当前最有判决力的节点，直到该决策树可以将训练样本正确分类为止。这里的决策树，实际上就是一个树状结构的预测模型。按照树的层次可以将其分为树根，分支节点和叶子节点。决策过程从</p> |
| 5 | <p>此处有 38 字相似</p> <p><math>yS=-i=1npxilogbpxi</math> ( 3-3 )</p> <p>根据香农的信息论，系统的熵与系统的有序性呈反比关系，即系统越稳定，熵越小，也就是系统包含的信息越少，因此信息熵也是对系统状态有序程度的一个度量，信息增益和信息熵类似。假定分类训练样本集合为 S，其中的某个属性为D，其信息增益Gain(S,D)定义如式3-4：</p> <p>Ga</p>              | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》 - 2013-05-27 (是否引证：否)</p> <p>1. <math>n(i)logb(i)(2.17)i</math> 1pxpx当一个系统越有序，其信息熵越低，包含的信息量也越少，因此信息熵也是对系统状态有序程度的一个度量。信息增益和信息熵具有一致性，对于分类训练样本集合 S 中的一个属性 A，其信息增益 Gain(S,A)定义如下：</p>                           |
| 6 | <p>此处有 41 字相似</p> <p><math>nS,D=EntropyS-v\in Values(D)Sv/SEntropySv</math> ( 3-4 )</p> <p>其中Values(D)是S中属性D的所有可能取得的值，Sv是集S中样本属性D值为v的集合。</p> <p>ID3 分类算法</p> <p>的主要思想就是计算样本集中所有可能属性对应的信息熵和信息增益，将信息最大增益的属性用作构建决策树时的根节点，此时的根节点</p> | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》 - 2013-05-27 (是否引证：否)</p> <p>1.ntropy ( Sv)v Values ( A)(2.18)其中 Values(A)是一个集合代表属性 A 所有可能的值，Sv是集 S 中样本属性A 值为 v 的集合。ID3 分类算法过程就是计算训练样本集合中每个属性的信息熵和信息增益，选择信息增益最大的属性作为决策树的树根，由于该属性针对分类提</p>                   |
| 7 | <p>此处有 47 字相似</p> <p>Values(D)是S中属性D的所有可能取得的值，Sv是集 S 中样本属性D值为v的集合。</p>   | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 (是否引证：否)</p> <p>1.实现随机森林的构建，首先选择属性数量k，然后采用无放回的随机抽样方式抽取随机森林中不同决策树的输</p>  |



|    |  |  |
|----|--|--|
|    | <p>ID3 分类算法的主要思想就是计算样本集中所有可能属性对应的信息熵和信息增益，将信息最大增益的属性用作构建决策树时的根节点，此时的根节点包含了样本信息的最大信息量，故而可以将样本数据按照最大化的程度进行分开来构建决策树，对于根节点后的子节点同样</p>  | <p>入属性，随后根据基尼系数计算属性的信息增益，依次迭代，找到信息增益最大的属性作为分裂节点，使用同样的方式为该节点构建子树。令attIndex代表当前选择的属性，bestIndex代表当前具有最大信息增益的属性，</p> <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》- 2013-05-27 (是否引证：否)</p> <p>1.所有可能的值，Sv是集合 S 中样本属性A 值为 v 的集合。ID3 分类算法过程就是计算训练样本集合中每个属性的信息熵和信息增益，选择信息增益最大的属性作为决策树的树根，由于该属性针对分类提供了最大的信息量，因此它能最大限度的将样本分开，接下来根据树根属性的分类规则确定左右子树</p> |
| 8  | <p>此处有 51 字相似</p> <p>]主要用来对ID3无法处理的连续值属性进行分类。因此，C4.5构建决策树时将最大的信息增益率作为根节点和叶子节点。具体的处理过程如下：</p> <p>Step1: 首先利用离散算法将训练集内的属性的连续指进行离散化处理，假定样本集S，其某个属性为D，将其离散化处理后得到一个离散值序列，根据值的大小做升序排列，得到离散属性序列 D'={D1， D2， D3</p>  | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》- 2013-05-27 (是否引证：否)</p> <p>1.问题，节点选择标准改为最大信息增益率。当属性为离散值时处理方式和 ID3 一致，对属性为连续取值时，C4.5 算法处理流程如下：Step1: 将训练集中为连续值的属性进行离散化处理，对样本总数为 n 的属性A 按大小进行升序排列，形成数值序列{A0， A1， ..., An-1}；接着将生成的数值序列中插入 n</p>  |
| 9  | <p>此处有 95 字相似</p> <p>小做升序排列，得到离散属性序列D'={D1， D2， D3， ...， Dn-1}；</p> <p>Step2：向D'中利用分割法平均插入n个点序列进行分割，对于分割点i，其取值为Di和 Di+1的平均值，这样一来，分割点将D'中的每个值分割成了两个子区间。</p> <p>Step3：分别计算每个分割点的信息增益以及增益率，信息增益</p> <p>的计算法和ID3是已知的，这里不再赘述，这里主要一下针对信息增益的改进——信息增益率，其值可以通过式得到式3-5：</p> <p>G</p> | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》- 2013-05-27 (是否引证：否)</p> <p>1.总数为 n 的属性A 按大小进行升序排列，形成数值序列{A0， A1， ..., An-1}；接着将生成的数值序列中插入 n 个分割点，对于分割点 i，其取值为 Ai和 Ai+1的平均值，数值序列被每个分割点划分成两个子区间。Step2: 对每个分割点分别计算其信息增益和信息增益率，信息增益计算方法同 ID3 算法，属性 A 的信息增益率是对信息增益的修正，计算公式为：GainRatio ( S</p>   |
| 10 | <p>此处有 34 字相似</p> <p>InfoS,D=GainS,D-i=1cSiS•log2SiS ( 3-5 )</p> <p>SplitInfoS,D描述了属性D对S进行分类时的广度和均匀性。得到全部分割点的信息增益率后，值最大的分割点作就是D的最优分割点。计算所有的属性，选择增益率最大的属性作为决策树的树根构建决策树。</p> <p>Step4：构建根节点后，将</p>   | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》- 2013-05-27 (是否引证：否)</p> <p>1.i 1  S   log     S SplitInfo(S,A)描述了用属性 A 对样本集合 S 进行分类时的广度和均匀性。选择信息增益率值最大的分割点作为该属性的最优分割点。将拥有最大信息增益率的属性作为决策树的树根。Step3:对树根节点属性划分的取值区</p>  |
| 11 | <p>此处有 57 字相似</p> <p>区间用作新的属性离散序列，重复Step2和Step3，来构建子树，当所有的子区间中的样本分类相同时，决策树</p>   | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》- 2013-05-27 (是否引证：否)</p> <p>1.Step3:对树根节点属性划分的取值区间递归的调用</p>   |



|    |   |  |
|----|---|--|
|    | <p>构建完成。</p> <p>Step5：利用训练好的决策树对未知样本进行分类。</p> <p>C4.5算法可以说是目前决策树模型中的比较重要和核心的算法，</p> <p>具有决策规则决策准确度高、简单、构建过程进行剪枝处理从而平衡性好等特点。但是，反复对训练集进行离散、分割和扫描必然会增加</p>   | <p>Step2，直到满足所有的子区间中都为同类的样本为止。</p> <p>Step4:利用训练好的决策树对新的测试样本进行分类。</p> <p>C4.5 算法作为决策树中的核心算法，生成的决策规则简单，精度较高，采用修剪技术避免了树的不平衡性。</p> <p>其缺点是：决策树生成过程中，频繁的对训练的数据</p>   |
| 12 | <p>此处有 90 字相似</p> <p>。总结起来，CART算法的运行过程就是先利用样本集依据判定标准构建一颗初始二叉树，之后通过交叉和测试集验证来对二叉树</p> <p>进行剪枝，得到最终优化后的分类决策树。CART构建初始决策树时根据基尼 ( Gini ) 系数来衡量不同类别下每个节点相互的差异程度。在数据集S上定义的Gini系数公式如式3-6：</p> <p><math>Gini(s) = 1 - \sum_{k=1}^K p_k^2</math> ( 3-6 )</p> <p>其中：K是我们选取的样本集类的个数，<math>K_i</math>是具体的每个类，<math>S_i</math>是训练集中</p> | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p> <p>1.首先根据一定规则在现有的所有样本集上生成一棵结构复杂的二叉树,然后使用交叉验证和测试集验证的方法对生成的决策回归树进行剪枝，最终得到简化的二叉分类树。CART生成原始决策树时使用基尼 ( Gini)系数测量每个节点在不同类别之间的差异程度。在数据集S上定义的Gini系数公式如下：<math>Gini(s) = 1 - \sum_{j=1}^C p_j^2</math>其中：C是预定义的类个数，<math>\phi_j</math>是类，是属于类C的样本数，<math>P = \sum_{i=1}^n I(S_i = C_i)</math></p> <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.，首先根据一定规则在现有的所有样本集上生成一棵结构复杂的二叉树，然后对树中每个分支使用交叉验证和测试集验证的方法对决策树进行剪枝，最终得到简化的二叉分类树。CART生成原始决策树时使用基尼 ( Gini ) 系数来测量每个节点在不同类别之间的差异程度。在数据集S上定义的Gini系数公式如下：式 ( 3-1 ) 其中：是预定义的类个数，是类，是属于类的样本数，是类在数据集中的相对频度，。Gini系数表</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.根据一定规则在现有的所有样本集上生成一棵结构复杂的二叉树，然后使用交叉验证和测试集验证的方法来对生成的决策回归树进行剪枝，最终得到简化的二叉分类树。CART生成原始决策树时使用基尼 ( Gini ) 系数来测量每个节点在不同类别之间的差异程度。在数据集S上定义的 Gini 系数公式如下： ( 3-1 ) 其中：c是预定义的</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p> <p>1.根据一定规则在现有的所有样本集上生成一棵结构复杂的二叉树，然后使用交叉验证和测试集验证的方法来对生成的决策回归树进行剪枝，最终得到简化的二叉分类树。CART生成原始决策树时使用基尼 ( Gini ) 系数来测量每个节点在不同类别之间的差异程度。在数据集S上定义的 Gini 系数公式如下：<math>Gini(s) = 1 - \sum_{i=1}^c p_i^2</math></p> |
| 13 | <p>此处有 104 字相似</p> <p><math>i</math>是训练集中属于<math>K_i</math>的样本的个数，<math>P_i = S_i / S</math>是<math>K_i</math>在训练集中的相对频度，<math>i = 1, \dots, K-1</math>。Gini系数体现的是S的分区纯度。例如，我们对两个类分别进行分支预测的话，<math>Gini(s) \in [0, 0.5]</math>，如果S中的所有数据都属</p>   | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.<math>p_{S_i}</math>是类 <math>i</math>在数据集中的相对频度，<math>i = 1, \dots, c</math>。Gini系数表示数据集S的分区纯度。例如对于两个类的分支预测，<math>Gini(s) \in [0, 0.5]</math>，如果S中的所有数据都属于同一</p>  |

|    |  |   |
|----|--|---|
|    | <p>于同一个类，那么Ginis=0，我们称此时的S是纯的数据集。如果Ginis=0.5，则代表S中的数据值在这两个类上是均匀分布的。把S根据某个属性分区为m个子集Si的Gini系数就是子集的Gini系数的加权求和，其计算</p>  | <p>个类，则 <math>Gini(s) = 0</math>，表示S是纯的。如果 <math>Gini(s) = 0.5</math>，则表示中的所有观测值在两个类上是平均分布。根据一个属性把数据集S分区为k个子集iS的Gini指数即为所得子集</p>  |
|    |  | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证：否)</p>  |
|    |  | <p>1. <math>p_{Si}/n</math> 是类 iC 在数据集中的相对频度，<math>i = 1, \dots, c</math>。Gini 系数表示数据集S的分区纯度。例如对于两个类的分支预测，<math>Gini(s) \in [0, 0.5]</math>，如果S中的所有数据都属于同一个类，则 <math>Gini(s) = 0</math>，表示S是纯的。如果 <math>Gini(s) = 0.5</math>，则表示中的所有观测值在两个类上是平均分布。根据一个属性把数据集S分区为k个子集iS的Gini指数即为所得子集</p>           |
|    |  | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士学位论文》 - 2015-03-16 (是否引证：是)</p>   |
|    |  | <p>1. <math>n_i</math> 是属于类C<sub>i</sub>的样本数，<math>P_i = n_i/n</math> 是类C<sub>i</sub>在数据集中的相对频度，<math>i = 1, \dots, c</math>。Gini系数表示数据集S的分区纯度。例如对于两个类的分支预测，<math>e \in [0, 0.5]</math>，如果S中的所有数据都属于同一个类，则 <math>G(s) = 0</math>，表示S是纯的。如果 <math>G(s) = 0.5</math>，则表示&gt;5中的所有观测值在两个类上是平均分布。</p> |
| 14 | <p>此处有 52 字相似</p> <p>那么Ginis=0，我们称此时的S是纯的数据集。如果Ginis=0.5，则代表S中的数据值在这两个类上是均匀分布的。把S根据某个属性分区为m个子集Si的Gini系数就是子集的Gini系数的加权求和，其计算公式为：</p> <p>Gini</p> <p><math>Gini = \sum_{i=1}^m \frac{n_i}{n} Gini(s_i)</math> (3-7)</p> <p>其中m表示分区后子集的样本数，n表示总样本量。算法</p> | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 (是否引证：否)</p>   |
|    |  | <p>1. 对于两个类的分支预测，如果中的所有数据都属于同一个类，则，表示是纯的。如果，则表示中的所有观测值在两个类上是平均分布。根据一个属性把数据集分区为个子集的Gini指数即为所得子集的Gini指数的加权和，其计算公式为：式(3-2) 其中表示分区后子集的样本数，表示总样本量。在对样本集进行分割时，算法从根节点开始分割，对样本中的</p>  |
|    |  | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 (是否引证：否)</p>   |
|    |  | <p>1. <math>Gini(s) = 0</math>，表示S是纯的。如果 <math>Gini(s) = 0.5</math>，则表示中的所有观测值在两个类上是平均分布。根据一个属性把数据集S分区为k个子集iS的Gini指数即为所得子集的Gini指数的加权和，其计算公式为：(3-2) 其中 <math>n_i</math> 表示分区后子集的样本数，n表示总样本量。在对样本集进行分割时，算法从根节点开始分</p>   |
|    |  | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证：否)</p>  |
|    |  | <p>1. <math>Gini(s) = 0</math>，表示S是纯的。如果 <math>Gini(s) = 0.5</math>，则表示中的所有观测值在两个类上是平均分布。根据一个属性把数据集S分区为k个子集iS的Gini指数即为所得子集的Gini指数的加权和，其计算公式为：()</p>   |
|    |  | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士学位论文》 - 2015-03-16 (是否引证：是)</p>   |
|    |  | <p>1. 则 <math>G(s) = 0</math>，表示S是纯的。如果 <math>G(s) = 0.5</math>，则表示&gt;5中的所有观测值在两个类上是平均分布</p>   |

|    |  |  |
|----|--|--|
|    |  | <p>布。根据一个属性把数据集S分区为个子集的Gini指数即为所得子集的Gini指数的加权和，其计算公式为：</p> $G^* = \sum_{i=1}^n \text{Gini}(S_i) \cdot \frac{ S_i }{ S }$ <p>其中表示分区后子集的样本数，<math> S </math>表示总样本量。在对样本集进行分割时，算法从</p>  |
| 15 | <p>此处有 168 字相似</p> <p>后子集的样本数，<math>n</math>表示总样本量。</p> <p>算法首先从根节点开始对训练集进行分割操作，训练集中的所有属性求解Ginisplit后</p> <p>选择结果最小的属性用作根节点和下一个分割点，然后利用递归重复以上操作构建其子树。</p> <p>3.3 支持向量机算法</p> <p>支持向量机 ( Support Vector Machines , SVM)是一种建立在统计学习理论基础上的新型机器学习分类算法，具有很好的学习反馈能力和推广适用性，因此，SVM已经成为了现阶段分类学习算法中新的研究热点和重点[35]。</p> <p>在以往的分类算法实践中，传统的分类算法暴露出了训练集样本不足、维度过高、或者不满足线性均匀分布时造成的分类偏差、模型训练</p> | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.进行分割时，算法从根节点开始分割，对样本中的所有属性计算splitGini，并选择splitGini 最小的属性作为分割点，然后递归地对每个子节点重复进行以上操作。</p> <p>3.4 支持向量机分类算法支持向量机 ( SVM ) 是一种基于统计学习理论的新型机器学习分类算法，其良好的学习能力和推广适应性使得支持向量机算法成为分类学习算法</p> <p>2.最小的属性作为分割点，然后递归地对每个子节点重复进行以上操作。3.4 支持向量机分类算法支持向量机 ( SVM ) 是一种基于统计学习理论的新型机器学习分类算法，其良好的学习能力和推广适应性使得支持向量机算法成为分类学习算法中研究的重点[35]。支持向量机算法专门研究和解决在对实际样本集分类过程中出现的样本量有限、样本属性呈非线性特征和特征向量维度高困难，很</p> <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.其中表示分区后子集的样本数，表示总样本量。在对样本集进行分割时，算法从根节点开始分割，对样本中的所有属性计算，并选择最小的属性作为分割点，然后递归地对每个子节点重复进行以上操作。3.4 支持向量机分类算法支持向量机SVM是一种基于统计学习理论 ( SLT ) 的新型机器学习分类算法，其良好的学习能力和推广适应性使得支持向量机算法成为分类学</p> <p>2.算，并选择最小的属性作为分割点，然后递归地对每个子节点重复进行以上操作。3.4 支持向量机分类算法支持向量机SVM是一种基于统计学习理论 ( SLT ) 的新型机器学习分类算法，其良好的学习能力和推广适应性使得支持向量机算法成为分类学习算法中研究的重点 [35]。支持向量机算法对于小样本、非线性、高维度和局部极小点等实际问题有较好的解决能力，由于它可以人为的进行核函数选择和相关参数</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p> <p>1.G?并选择最小的属性作为分割点，然后递归地对每个子节点重复17北京邮电大学工程硕士学位论文 进行以上操作。3.4支持向量机分类算法支持向量机 ( Suppotr Vector Machines , SVM)是一种基于统计学习理论的新型机器学习分类算法，其良好的学习能力和推广适应性使得支持向量机算法成为分类学习算法中研究的重点 [35]。支持向量机算法专门研究和解决在对实际样本集分类过程中出现的样本量有限、样本属性呈非线性特征和特征向量维度高</p> |

|    |  |   |
|----|--|---|
|    |  | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证:否)</p> <p>1.每个子节点重复第三章分类学习算法概述18进行以上操作。3.4 支持向量机分类算法支持向量机 (SVM) 是一种基于统计学习理论的新型机器学习分类算法,其良好的学习能力和推广适应性使得支持向量机算法成为分类学习算法中研究的重点[35]。支持向量机算法专门研究和解决在对实际样本集分类过程中出现的样本量有限、样本属性呈非线性特征和特征向量维度高等困难,很</p>   |
| 16 | <p>此处有 37 字相似</p> <p>算法暴露出了训练集样本不足、维度过高、或者不满足线性均匀分布时造成的分类偏差、模型训练资源耗费过大、过拟合等问题。SVM</p> <p>针对不同的训练集特征分布情况会采用多种不同的核函数进行分类模型训练,很好的</p> <p>解决了上述存在的问题,目前SVM已被大量使用于文本分类、图像分类等领域,取得了可喜额成果。</p> <p>3.3.1 统计学习理论</p>  | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 (是否引证:否)</p> <p>1.集分类过程中出现的样本量有限、样本属性呈非线性特征和特征向量维度高等困难,很好的解决了“维数灾难”问题。并且算法针对不同的样本特征分布规律提供了多种核函数进行分类模型训练的优化,有效解决了模型选择和过拟合等机器学习中难题,因此被广泛应用在文本分类、基因分类及其他分类预测领域。3.4.1 统</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证:否)</p> <p>1.集分类过程中出现的样本量有限、样本属性呈非线性特征和特征向量维度高等困难,很好的解决了“维数灾难”问题。并且算法针对不同的样本特征分布规律提供了多种核函数进行分类模型训练的优化,有效解决了模型选择和过拟合等机器学习中难题,因此被广泛应用在文本分类、基因分类及其他分类预测领域。3.4.1 统</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 (是否引证:是)</p> <p>1.集分类过程中出现的样本量有限、样本属性呈非线性特征和特征向量维度高等困难,很好的解决了“维数灾难”问题。并且算法针对不同的样本特征分布规律提供了多种核函数进行分类模型训练的优化,有效解决了模型选择和过拟合等机器学习中难题,因此被广泛应用在文本分类、基因分类及其他分类预测领域。3.4.1</p> |
| 17 | <p>此处有 157 字相似</p> <p>不同的训练集特征分布情况会采用多种不同的核函数进行分类模型训练,很好的解决了上述存在的问题,目前SVM已被大量使用于文本</p> <p>分类、图像分类等领域,取得了可喜额成果。</p> <p>3.3.1 统计学习理论</p> <p>统计学习理论[30] (Statistical Learning Theory, SLT) 顾名思义,就是借助统计学的研究方法来帮助完成机器学习的研究,以便发现无法被目前的技术所解释但是又真实的存在于实际的应用场景中的规律,并将这些规律总结推广用在</p> <p>实际的学习预测当中。该理论力求通过有限的样本数据归纳出合适的可以被泛化和推广的统计模型以使模型在根据相同的数据进行预测时</p> | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证:否)</p> <p>1.函数进行分类模型训练的优化,有效解决了模型选择和过拟合等机器学习中难题,因此被广泛应用在文本分类、基因分类及其他分类预测领域。3.4.1 统计学习理论 SLT统计学习理论 (Statistical Learning Theory, SLT) 主要是利用统计学的方法来研究机器学习规律,目的是发掘出一些当前理论无法解释但真实存在于样本集中的规律,并将这些规律用在分析客观事物中。该理论在有限样本条件下建立的统计模型在取得最优解的同时保证模型具有较好的泛化性和推广性。统计学习理</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 (是否引证:是)</p> <p>1.行分类模型训练的优化,有效解决了模型选择和过拟</p>  |



|    |  |  |
|----|--|--|
|    |  | <p>合等机器学习中难题，因此被广泛应用在文本分类、基因分类及其他分类预测领域。3.4.1统计学习理论统计学习理论 ( Staitstical Learning Theory, SLT)主要是利用统计学的方法来研究机器学习规律，目的是发掘出一些当前理论无法解释但真实存在于样本集中的规</p> <p>2.3.4.1统计学习理论统计学习理论 ( Staitstical Learning Theory, SLT)主要是利用统计学的方法来研究机器学习规律，目的是发掘出一些当前理论无法解释但真实存在于样本集中的规律，并将这些规律用在分析客观事物中。该理论在有限样本条件下建立的统计模型在取得最优解的同时保证模型具有较好的泛化性和推广性。统计学习理</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.函数进行分类模型训练的优化，有效解决了模型选择和过拟合等机器学习中难题，因此被广泛应用在文本分类、基因分类及其他分类预测领域。3.4.1 统计学习理论 SLT统计学习理论 ( Statistical Learning Theory , SLT ) 主要是利用统计学的方法来研究机器学习规律，目的是发掘出一些当前理论无法解释但真实存在于样本集中的规律，并将这些规</p> <p>2.ical Learning Theory , SLT ) 主要是利用统计学的方法来研究机器学习规律，目的是发掘出一些当前理论无法解释但真实存在于样本集中的规律，并将这些规律用在分析客观事物中。该理论在有限样本条件下建立的统计模型在取得最优解的同时保证模型具有较好的泛化性和推广性。统计学习理</p> <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.效避免了“维数灾难”和过拟合等机器学习中难题，因此被广泛应用在文本分类、基因分类及其他分类预测领域。7. 3.4.1 统计学习理论SLT统计学习理论statistical learning theory是一种利用统计学方法专门研究机器学习在有限样本情况下的学习规律的理论，目的是发掘出一些当前理论无法解释但真实存在于</p> <p>2.arning theory是一种利用统计学方法专门研究机器学习在有限样本情况下的学习规律的理论，目的是发掘出一些当前理论无法解释但真实存在于样本集中的规律，并将这些规律用在分析客观事物中。该理论在有限样本统计问题上建立的统计规则综合考虑了对渐近性能的要求和在现有样本条件下的获得最优解的要求。</p> |
| 18 | <p>此处有 38 字相似</p> <p>测时可以得到最优解。SLT的核心理论是求解一个函数集，其所能打散的最大样本个数被称为该函数集的VC维[30]，反应的是该函数集的适用性。因此函数的VC维越高，则该函数集的推广适用性也就越好。</p> | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.维，将一个函数集能够打散的最大样本数作为该函数集的VC 维，其反映了该函数集的推广适应能力。由统计学习理论可知，一个函数集的 VC 维越高，则该函数集的推广适应能力也就越好[35]。统计学习理论研究各类函数集的经验风险与实际风险的关系，针对二分类问题，该理论指出分类函数集中所有函数的经验风险 <math>R(\mathbf{f})</math></p>   |

|    |   |  |
|----|---|--|
|    | SLT在确定函数集后，需要对函数集的经验风险与实际风险进行验证分析，在处理二分的问题上，分类函数集中所有函数的经验风险R  | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证：否)</p> <p>1.维，将一个函数集能够打散的最大样本数作为该函数集的VC维，其反映了该函数集的推广适应能力。由统计学习理论可知，一个函数集的VC维越高，则该函数集的推广适应能力也就越好[35]。统计学习理论研究各类函数集的经验风险与实际风险的关系，针对二分类问题，该理论指出分类函数集中所有函数的经验风险R(</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 (是否引证：是)</p> <p>1.维，将一个函数集能够打散的最大样本数作为该函数集的VC维，其反映了该函数集的推广适应能力。由统计学习理论可知，一个函数集的VC维越高，则该函数集的推广适应能力也就越好[35]。统计学习理论研究各类函数集的经验风险与实际风险的关系，针对二分类问题，该理论指出分类函数集中所有函数的经验风险及和</p>   |
| 19 | <p>此处有 40 字相似</p> <p>数集的推广适用性也也就越好。</p> <p>SLT在确定函数集后，需要对函数集的经验风险与实际风险进行验证分析，在处理二分的问题上，分类函数集中所有函数的经验风险<math>R_{emp}(w)</math>和实际风险<math>R(w)</math>需要满足下面的公式</p> <p>3-8：</p> $R_w \leq R_{emp} + h \ln 2nh + 1 - \ln(\delta/4) n \quad (3-8)$ <p>其中h是函数集的VC维，n是样本数。</p> | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 (是否引证：否)</p> <p>1.应能力也就越好[35]。统计学习理论研究各类函数集的经验风险与实际风险的关系，针对二分类问题，该理论指出分类函数集中所有函数的经验风险<math>R(w)_{emp}</math>和实际风险<math>R(w)</math>需要满足下面的公式[36]：(3-3)其中h是函数集的VC维，n是样本数。由上可知，在二分类问题中的实际风险可由经验风</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证：否)</p> <p>1.应能力也就越好[35]。统计学习理论研究各类函数集的经验风险与实际风险的关系，针对二分类问题，该理论指出分类函数集中所有函数的经验风险<math>R(w)_{emp}</math>和实际风险<math>R(w)</math>需要满足下面的公式</p> <p>[36]：<math>nhnR_w R_{wemp}(\ln(2)/1)\ln(4/())(3-3)</math>其中</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 (是否引证：是)</p> <p>1.广适应能力也就越好[35]。统计学习理论研究各类函数集的经验风险与实际风险的关系，针对二分类问题，该理论指出分类函数集中所有函数的经验风险及和实际风险<math>R(w)</math>需要满足下面的公式[36]：1?...-明(34)其中2是函数集的VC维，《是样本数。由上可知，在二分类问题中的实际风险可由经验风</p> <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 (是否引证：否)</p> <p>1.好[35]。统计学习理论主要研究各类函数集的经验风险与实际风险之间的内在关系，针对两类分类问题，该理论指出在指示函数集中所有函数的经验风险和实际风险需要满足下面的公式[36]：式(3-3)其中是函数集的VC维，是样本数。根据上述结论可以推出，在机器学习中的实际风险可由</p> |
| 20 | 此处有 32 字相似  | 于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 -   |

|    |   |  |
|----|---|--|
|    | <p>n ( 3-8 )</p> <p>其中h是函数集的VC维，n是样本数。</p> <p>从上式可以发现，二分问题中，经验风险和置信区间是构成实际风险的</p> <p>两个部分，并且实际风险、训练样本的数量这二者的大小与VC维的高低</p> <p>程度存在关联。其具体的关系如式3-9：</p> $R_w \leq R_{emp} + \phi(h/n) \quad (3-9)$ <p>公式 ( 3-5 ) 表明样本数量有限时</p>                                     | <p>2014-12-31 ( 是否引证：否 )</p> <p>1. ( 3-3 ) 其中是函数集的VC维，是样本数。根据上述结论可以推出，在机器学习中的实际风险可由经验风险和置信范围两个部分组成，并且实际风险的大小和训练样本的数量和VC维的高低有关。其关系可以表示为如下公式：. 式 ( 3-4 ) 该公式表明在有限的样本条件下，机器学习算法的置信范围大小和算法VC</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 ( 是否引证：否 )</p> <p>1.： ( 3-3 ) 其中h是函数集的 VC 维，n是样本数。由上可知，在二分类问题中的实际风险可由经验风险和置信范围两个部分组成，并且实际风险的大小和训练样本的数量和 VC 维的高低有关。其关系可以表示为如下公式：. ( 3-4 ) 该公式表明在有限的样本条件下，机器学习算法的置信范围大小和算法 VC</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 ( 是否引证：否 )</p> <p>1.) ( 3-3 ) 其中h是函数集的 VC 维，n是样本数。由上可知，在二分类问题中的实际风险可由经验风险和置信范围两个部分组成，并且实际风险的大小和训练样本的数量和 VC 维的高低有关。其关系可以表示为如下公式：R(w)R(w)(h/n)emp. ( 3-4 ) 该公式表明在有限的样本条件</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》- 2015-03-16 ( 是否引证：是 )</p> <p>1.明] (34)其中/2是函数集的VC维，《是样本数。由上可知，在二分类问题中的实际风险可由经验风险和置信范围两个部分组成，并且实际风险的大小和训练样本的数量和VC维的高低有关。其关系可以表示为如下公式：R(w)^R?ln^(w)+&amp;lt;i&gt;{h}- (3-5)公式 ( 3-5)表明在有</p> |
| 21 | <p>此处有 72 字相似</p> <p>体的关系如式3-9：</p> $R_w \leq R_{emp} + \phi(h/n) \quad (3-9)$ <p>公式 ( 3-5 ) 表明样本数量有限时，SLT的置信区间大小和函数集的VC维高低正相关，即函数集的VC维越高，SLT的复杂程度和置信区间也就越大，从而导致实际风险增加，SLT就会出现模型过学习的现象。</p> <p>SLT指出在分类模型构建时，要同时尽可能的降低VC维和减小置信区间并最小化训练集的经验风险，从而让分类模型具有较低的实际</p> | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》- 2014-12-31 ( 是否引证：否 )</p> <p>1.维的高低有关。其关系可以表示为如下公式：. 式 ( 3-4 ) 该公式表明在有限的样本条件下，机器学习算法的置信范围大小和算法VC维高低成正相关，即算法的VC维越高，则算法的复杂度和置信范围也就越大，进而使算法的实际风险也相应增加而出现分类模型过学习的现象。统计学习理论说明在分类学习过程中，要使得分类模型具有较小的实际风险并具有良好的分类效果和推广性，需要同时降低VC维以缩小</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 ( 是否引证：否 )</p> <p>1. 维的高低有关。其关系可以表示为如下公式：. ( 3-4 ) 该公式表明在有限的样本条件下，机器学习算法的置信范围大小和算法 VC维高低成正相关，即算法的 VC维越高，则算法的复杂度和置信范围也就越大，进而使算法的实际风险也相应增加而出现分类模型过学习的现象。统计学习理论说明在分类学习过程中，要使得分类模型具有较小的实际风险并具有良好的分类效果和推广性，需要同时降低 V</p>   |

|    |  |   |
|----|--|---|
|    |  | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证:否)</p> <p>1.式: <math>R(w)R(w)(h/n)\text{emp.}</math> (3-4) 该公式表明在有限的样本条件下,机器学习算法的置信范围大小和算法VC维高低成正相关,即算法的VC维越高,则算法的复杂度和置信范围也就越大,进而使算法的实际风险也相应增加而出现分类模型过学习的现象。统计学习理论说明在分类学习过程中,要使得分类模型具有较小的实际风险并具有较好的分类效果和推广性,需要同时降低V</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 (是否引证:是)</p> <p>1.<math>R(w)^R\ln(w)+\&amp;lt;i\&amp;gt;\{h\}</math>- (3-5)公式 (3-5)表明在有限的样本条件下,机器学习算法的置信范围大小和算法VC维高低成正相关,即算法的VC维越高,则算法的复杂度和置信范围也就越大,进而使算法的实际风险也相应增加而出现分类模型过学习的现象。统计学习理论说明在分类学习过程中,要使得分类模型具有较小的实际风险并具有较好的分类效果和推广性,需要同时降低</p>  |
| 22 | <p>此处有 120 字相似</p> <p>维和减小置信区间并最小化训练集的经验风险,从而让分类模型具有较低的实际风险和较好的分类与推广性</p> <p>，。</p> <p>3.3.2 SVM</p> <p>算法</p> <p>1.广义最优分类面</p> <p>支持向量机算法的核心思想取自线性可分样本集中求解最优分类面的算法,并将该思想推广到在非线性高维空间的样本集中求解最优分类面的场景下,下面以在二维平面下的样本寻找最有分类为例子来简要说明该思想。</p> <p>图3-1所示,平面</p> <p>S中分布着大量的黑点和白点,这些点表示了样本集中两种不同的数据样本类别,直线L将这两类样本分开</p> <p>,L1、L2分别表示经过各</p> | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 (是否引证:是)</p> <p>1.维以缩小置信范围并使训练样本的经验风险最小化。</p> <p>18第三章分类学习算法概述 3.4.2支持向量机算法(1)广义最优分类面支持向量机算法的思想来自于在线性可分样本集中求解最优分类面的算法,并将该思想成功推广到求解非线性高维空间的样本最优分类面的问题中,下面通过在二维平面对样本进行分类的例子来简单说明。图3-1所示,在二维平面S中,实心点和空心点分别代表两种类型的样本分布,直线H将这两类样本分开</p> <p>,<math>l_1</math>、<math>l_2</math>分别为过各类中离H最近的样本点且平</p> <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 (是否引证:否)</p> <p>1. (1) 广义最优分类面支持向量机算法Support Vector Machines (SVM) 的思想来自于在线性可分样本集中求解最优分类面的算法,并将该思想成功推广到求解非线性高维空间的样本最优分类面的问题中,下面通过在二维平面对样本进行分类的例子来简单说明。图3-1所示,在二维平面s中,实心点和空心点分别代表两种类型的样本分布,其中直线H为这两类样本的分类线</p> <p>,<math>l_1</math>、<math>l_2</math>分别为过各类中离H最近的样本点且平行于H的直</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 (是否引证:否)</p> <p>1.M (1) 广义最优分类面支持向量机算法 Support Vector Machines (SVM) 的思想来自于在线性可分样本集中求解最优分类面的算法,并将该思想成功推广到求解非线性高维空间的样本最优分类面的问题中,下面通过在二维平面对样本进行分类的例子来简单说明。图 3-1 所示,在二维平面 s 中,实心点和空心点分别代表两种类型的样本分布,直线 H 将这两类样本分开</p> <p>,<math>l_1H</math>、<math>l_2H</math> 分别为过</p> |



|    |  |   |
|----|--|---|
|    |  | <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p> <p>1.M ( 1 ) 广义最优分类面支持向量机算法 Support Vector Machines ( SVM ) 的思想来自于在线性可分样本集中求解最优分类面的算法，并将该思想成功推广到求解非线性高维空间的样本最优分类面的问题中，下面通过在二维平面中对样本进行分类的例子来简单说明。图 3-1 所示，在二维平面 s 中，实心点和空心点分别代表两种类型的样本分布，直线 H 将这两类样本分开，1H、2H 分别为过</p>   |
| 23 | <p>此处有 93 字相似</p> <p>黑点和白点，这些点表示了样本集中两种不同的数据样本类别，直线L将这两类样本分开，L1、L2分别表示经过各自类别样本的直线</p> <p>中离L最近的样本点和L平行不相交的两条直线，L1、L2间的垂直距离即为这两种数据样本的分类间隔。直线L1、L2经过的点就是将两类数据样本分离开来的支持向量 ( Support Vector , SV )。在求得L1、L2的时候要保证在完全分离两类样本数据的条件下还要具备最大的垂直距离也就是分类间隔，保证其最大也就</p> | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p> <p>1.在二维平面S中，实心点和空心点分别代表两种类型的样本分布，直线H将这两类样本分开，//、//,分别为过各类中离H最近的样本点且平行于H的直线，它们之间的距离为这两类样本的分类间隔。在直线与//,上的样本点称作将这两类样本分开的支持向量(Support Vector, SV)。最优样本分类线要求在正确划分两类样本点分布的同时具有最大的分类间隔，而使分类间隔最大就是使分类面对未知样</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.中，实心点和空心点分别代表两种类型的样本分布，直线 H 将这两类样本分开，1H、2H 分别为过各类中离 H 最近的样本点且平行于 H 的直线，它们之间的距离为这两类样本的分类间隔。在直线1H 与2H上的样本点称作将这两类样本分开的支持向量 Support Vector ( SV )。最优样本分类线要求</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p> <p>1.中，实心点和空心点分别代表两种类型的样本分布，直线 H 将这两类样本分开，1H、2H 分别为过各类中离 H 最近的样本点且平行于 H 的直线，它们之间的距离为这两类样本的分类间隔。在直线1H 与2H上的样本点称作将这两类样本分开的支持向量 Support Vector ( SV )。最优样本分类线要求</p> <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.。图3-1所示，在二维平面s中，实心点和空心点分别代表两种类型的样本分布，其中直线H为这两类样本的分类线，、分别为过各类中离H最近的样本点且平行于H的直线，它们之间的距离为这两类样本的分类间隔 ( margin )，在直线与上的样本点称作支持向量 Support Vector ( SV )。最优样本分类线要求在正确划分两类样</p> |
| 24 | <p>此处有 116 字相似</p> <p>， yi ) ， i=1 ， ... ， n ， x∈Rd ， y∈{+1 ， -1}满足式3-10：<br/> <math>y_i w \cdot x_i + b - 1 \geq 0</math>， i=1 ， ... ， n<br/> ( 3-10 )</p>  | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.本集分类预测的推广能力达到最大。假设分类线方程为，对其进行归一化，使得对线性可分的样本集满足如下关系：；式 ( 3-5 ) 此时分类面的分类间隔等于，而</p>   |

|    |  |  |
|----|--|--|
|    | <p>此时平面的分类间隔等于<math>2w</math>，当<math>W_2</math>取得最小值的时候分类间隔<math>2w</math>将取得最大的结果，所求得分类面就是最优分类面。</p> <p>图 3-1 两种类型样本集的最优分类线</p> <p>2. 支持向量机</p> <p>对于<math>N</math>维空间中的线性函数，其<math>VC</math>维等于<math>N+1</math>，当<math>N</math>取值较大时，会出现样本集维度过高的问题此时需要对<math>VC</math>维做降维处理，以保证求得的最优分类面具有高适用性。同时还可以通过</p> | <p>使值达到最小的时候分类间隔可取得最大值，此时所求得分类面即为最优分类面。图3-1 两种类型样本集的最优分类线 ( 2 ) 支持向量机对于<math>N</math>维空间中的线性函数，其<math>VC</math>维等于<math>N+1</math>，我们可以将高维空间的<math>VC</math>维通过一定方法进行降维，以保证求得的最优分类面具有推广适应性。同时在求解函数时可以将原问题转化为</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1. 对线性可分的样本集<math>(x_1, 1), \dots, (x_n, 1)</math>满足如下关系：； ( 3-5 ) 此时分类面的分类间隔等于<math>w_2</math>，而使<math>2w</math> 值达到最小的时候分类间隔可取得最大值，此时所求得分类面即为最优分类面。图 3-1 两种类型样本集的最优分类线 ( 2 ) 支持向量机对于 <math>N</math> 维空间中的线性函数，其 <math>VC</math> 维等于 <math>N+1</math>，我们可以将高维空间的 <math>VC</math> 维通过一定方法进行降维，以保证求得的最优分类面具有推广适应性。同时在求解函数时可以将原</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究 2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p> <p>1. <math>(x_1, 1), \dots, (x_n, 1)</math>满足如下关系：<math>yw \leq 1, 0, 1, \dots</math>； ( 3-5 ) 此时分类面的分类间隔等于<math>w_2</math>，而使<math>2w</math> 值达到最小的时候分类间隔可取得最大值，此时所求得分类面即为最优分类面。图 3-1 两种类型样本集的最优分类线 ( 2 ) 支持向量机对于 <math>N</math> 维空间中的线性函数，其 <math>VC</math> 维等于 <math>N+1</math>，我们可以将高维空间的<math>VC</math> 维通过一定方法进行降维，以保证求得的最优分类面具有推广适应性。同时在求解函数时可以将原</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p> <p>1. <math>\{x_i\}</math>，满足如下关系：<math>yw \geq 0, y = 1, \dots, tr</math>； (3-6) 此时分类面的分类间隔等于！，而使<math>2W</math>值达到最小的时候分类间隔可取得最<math>W</math>大值，此时所求得分类面即为最优分类面。<math>\gamma = \frac{1}{2} \min_i \frac{y_i}{\ z_i\ }</math>图3-1两种类型样本集的最优分类线(2)支持向量</p> <p>2. 大值，此时所求得分类面即为最优分类面。<math>\gamma = \frac{1}{2} \min_i \frac{y_i}{\ z_i\ }</math>图3-1两种类型样本集的最优分类线 (2)支持向量机对于<math>N</math>维空间中的线性函数，其<math>VC</math>维等于<math>N+1</math>，可以将高维空间的<math>VC</math>维通过一定方法进行降维，以保证求得的最优分类面具有推广适应性。同时在求解函数时可以将原问</p> |
| 25 | <p>此处有 162 字相似</p> <p>支持向量机</p> <p>对于<math>N</math>维空间中的线性函数，其<math>VC</math>维等于<math>N+1</math>，当<math>N</math>取值较大时，会出现样本集维度过高的问题此时需要对<math>VC</math>维做</p> <p>降维处理，以保证求得的最优分类面具有高适用性。同时还可以通过在求解函数时将原问题转化为对偶问题来让问题的复杂度由样本集中的数据支持向量的个数来决定，从而简化原问题的复杂程度。通过这些方式可以把求解高维的空间中最优分类面的情况转化为在较低维度</p>                | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1. ) 支持向量机对于 <math>N</math> 维空间中的线性函数，其 <math>VC</math> 维等于 <math>N+1</math>，我们可以将高维空间的<math>VC</math> 维通过一定方法进行降维，以保证求得的最优分类面具有推广适应性。同时在求解函数时可以将原问题转化为对偶问题，使得求解问题的复杂度直接由样本中支持向量的个数来决定，以达到简化原问题的复杂度的目的。这些特点将使高维空间求解最优分类面的问题转化为在低维空间求解最优分类面的问题进行求解成为可能。对于非线性划分问</p>  |

|   |  |
|---|--|
| <p>空间进行求解。</p> <p>对于非线性划分问题，可以首先将其升维到高维空间转变成线性可分的问题</p> <p>后在通过降维后利用核函数简化高维空间的向量运算来求解最优分类面内积。已知在低维空间中凡是满足Mercer定理的函数对应着</p> | <p>题，可以通过将其转换为高维空间的线性可分问题来求解分类模型的最优分类面。一般使用低维空间中一些核函数求解转换后的高维空间复杂的向量内积，这使得高维空间的向量</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证：否)</p> <p>1.) 支持向量机对于 N 维空间中的线性函数，其 VC 维等于 <math>N+1</math>，我们可以将高维空间的 VC 维通过一定方法进行降维，以保证求得的最优分类面具有推广适应性。同时在求解函数时可以将原问题转化为对偶问题，使得求解问题的复杂度直接由样本中支持向量的个数来决定，以达到简化原问题的复杂度的目的。这些特点将使高维空间求解最优分类面的问题转化为在低维空间求解最优分类面的问题进行求解成为可能。对于非线性划分问题，可以通过将其转换为高维空间的线性可分问题来求解分类模型的最优分类面。一般使用低维空间中一些核函数求解转换后的高维空间复杂的向量内积，这使得高维空间的向量</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 (是否引证：是)</p> <p>1.线(2)支持向量机对于N维空间中的线性函数，其VC维等于<math>N+1</math>，可以将高维空间的VC维通过一定方法进行降维，以保证求得的最优分类面具有推广适应性。同时在求解函数时可以将原问题转化为对偶问题，使得求解问题的复杂度直接由样本中支持向量的个数来决定，以达到简化原问题的复杂度的目的。这些特点将使高维空间求解最优分类面的问题转化为在低维空间求解最优分类面的问题进行求解成为可能。对于非线性划分问题，可以通过将其转换为高维空间的线性可分问题来求解分类模型的最优分类面。一般使用低维空间中一些核函数求解转换后的高维空间复杂的向量内积，这使得高维空间的</p> <p>于俊_基于等级容灾模型_于俊 - 《学术论文联合比对库》 - 2014-12-31 (是否引证：否)</p> <p>1.类线 ( 2 ) 支持向量机对于N维空间中的线性函数，其 VC 维等于<math>N+1</math>，我们可以将高维空间的 VC 维通过一定方法进行降维，以保证求得的最优分类面具有推广适应性。同时在求解函数时可以将原问题转化为对偶问题，使得求解问题的复杂度与样本空间的维度大小无关，而直接由样本中支持向量的个数来决定，这将在很大程度上简化了原问题的求解复杂</p> <p>2.分类面具有推广适应性。同时在求解函数时可以将原问题转化为对偶问题，使得求解问题的复杂度与样本空间的维度大小无关，而直接由样本中支持向量的个数来决定，这将在很大程度上简化了原问题的求解复杂度。这些特点将使高维空间求解最优分类面的问题转化为在低维空间求解最优分类面的问题进行求解成为可能。对于非线性划分问题，通过使用相关方法将非线性转化为高维空间中的线性来</p> |
|---|--|

|    |   |  |
|----|---|--|
| 26 | <p>此处有 175 字相似</p> <p>较低维度空间进行求解。</p> <p>对于非线性划分问题，可以首先将其升维到高维空间转变成线性可分的问题后在通过降维后利用核函数简化高维空间的向量运算来求解最优分类面内积。已知在低维空间中凡是满足Mercer定理的函数对应着其向某一高维空间变换的内积，而该变换函数就称为核函数，转换后的高维空间称为再生核希尔伯特空间（Reproducing Kernel Hilbert Space，RKHS）。下面为 Mercer 定理的定义：</p> <p>定理3.1：Mercer定理核函数<math>K : K(u, v) = \varphi(u) \cdot \varphi(v)</math>，当且仅当对于任意满足公式为有限值的函数<math>g(x)</math> <math>\int_2 dx</math>为有限值的函数<math>g(x)</math>，则：<math>K(x, y)</math></p> | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士学位论文》 - 2015-03-16 (是否引证：是)</p> <p>1.性可分问题来求解分类模型的最优分类面。一般使用低维空间中一些核函数求解转换后的高维空间复杂的向量内积，这使得高维空间的向量运算得以简化。已知在低维空间中凡是19北京邮电大学工程硕士学位论文 满足Mercer定理的函数对应着其向某一高维空间变换的内积，而该变换函数</p> <p>2.的向量内积，这使得高维空间的向量运算得以简化。已知在低维空间中凡是19北京邮电大学工程硕士学位论文 满足Mercer定理的函数对应着其向某一高维空间变换的内积，而该变换函数就称为核函数[31]，转换后的高维空间称为再生核希尔伯特空间（Reproducing Kernel Hilbetr Space, RKHS)。下面为 Mercer 定理的定义：定理 3.1 Mercer定理核函数尤：<math>A:(m,v)=0(m)-\&amp;lt;D(v)</math>,当且仅当对于任意满足公式为有限值的函数<math>g(c)</math>，则/<math>K(x,y)g(x)</math></p> |
|    |   | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 (是否引证：否)</p> <p>1.的最优分类面。一般使用低维空间中一些核函数求解转换后的高维空间复杂的向量内积，这使得高维空间的向量运算得以简化。已知在低维空间中凡是满足 Mercer 定理的函数对应着其向某一高维空间变换的内积，而该变换函数就称为核函数[31]。下面为 Mercer 定理的定义：定理 3.1 Mercer 定理[37]核函数<math>K</math>：，当且仅当对于任</p>  |
|    |   | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证：否)</p> <p>1.高维空间复杂的向量内积，这使得高维空间的向量运算得以简化。已知在低维空间中凡是第三章分类学习算法概述20满足 Mercer 定理的函数对应着其向某一高维空间变换的内积，而该变换函数就称为核函数[31]。下面为 Mercer 定理的定义：定理 3.1 Mercer 定理[37]核函数<math>K : K(u, v) (</math></p>  |
|    |   |  |
| 27 | <p>此处有 113 字相似</p> <p><math>(v)</math>，当且仅当对于任意满足公式为有限值的函数<math>g(x)</math> <math>\int_2 dx</math>为有限值的函数<math>g(x)</math>，则：<math>K(x, y) g(x) g(y) dx dy \geq 0</math> (3-11)</p> <p>因此，在求解非线性分类的最优分类面问题中可以使用内积核函数<math>K(x_i, x_j)</math> 将低维空间中的非线性分类转化为高维空间中的线性分类，从而降低低维空间非线性分类最优平面求解问题的复杂度，此时的目标函数变为：</p> <p><math>Qa=i=1n\alpha-12i, j=1naiajyijK(x_i, x_j)</math> (3-12)</p> <p>相应的分类函数变为：<math>f_x=</math></p>                     | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 (是否引证：否)</p> <p>1.定义：定理3.1 Mercer定理[37] 核函数可以表示为：，当且仅当对于任意满足为有限值的函数，则。式 (3-6) 因此，在求解非线性分类的最优分类面问题中可以使用内积核函数将低维空间中的非线性分类转化为高维空间中的线性分类，从而降低低维空间非线性分类最优平面求解问题的复杂度，此时的目标函数变为：；式 (3-7) 相应的分类函数变为：。式 (3-8) 综上所述，支持向量机算法通过内积核函数将低维空间中的</p>   |
|    |   | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 (是否引证：否)</p> <p>1.37]核函数<math>K</math>：，当且仅当对于任意满足公式<math>g(x)dx2(</math>)为有限值的函数<math>g(x)</math>，则。(3-6) 因此，在求解非线性分类的最优分类面问题中可以使用内积核函数<math>(,ij)Kxx</math>将低维空间中的非线性分类转化为高维空间中的</p>  |
|    |   |  |



|    |  |   |
|----|--|---|
|    |  | <p>线性分类，从而降低低维空间非线性分类最优平面求解问题的复杂度，此时的目标函数变为：；（3-7）相应的分类函数变为：（3-8）综上所述，支持向量机算法通过内积核函数将低维空间中的非线性问题转为</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究 2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15（是否引证：否）</p> <p>1.意满足公式<math>\int_0^1 g(x)g(y)dx dy</math>（3-6）因此，在求解非线性分类的最优分类面问题中可以使用内积核函数<math>(,j)K_{xx}</math>将低维空间中的非线性分类转化为高维空间中的线性分类，从而降低低维空间非线性分类最优平面求解问题的复杂度，此时的目标函数变为：<math>(,j)21(1,1)ijijijiniQ yyK_{xx}</math>；（3-7）相</p> <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16（是否引证：是）</p> <p>1.7)因此，在求解非线性分类的最优分类面问题中可以使用内积核函数将低维空间中的非线性分类转化为高维空间中的线性分类，从而降低低维空间非线性分类最优平面求解问题的复杂度，此时的目标函数变为：<math>1/G(1n,?,X_j)</math>，(3-8)/=1相应的分类函数变为：<math>fix</math></p>  |
| 28 | <p>此处有 132 字相似</p> <p>，<math>x_j</math>）（3-12）</p> <p>相应的分类函数变为：</p> <p><math>fx=sgn(i=1nai*yiK_{xi}, x_j+b^*)</math>（3-13）</p> <p>综上所述，SVM算法通过内积核函数将低维空间中的非线性问题转为高维空间中的线性划分问题来求解，并在高维空间中通过降维等方式简化运算复杂度，进而求解最优分类面的过程。</p> <p>3.支持向量机的核函数</p> <p>常用的SVM核函数有以下三种类型：</p> <p>—是多项式核函数，其表达式为：</p> <p><math>Kx, x'=ax \cdot x'+\beta q, \beta \geq 0</math>；（3-14）</p> <p>二是径向基核函数（RBF）：<math>Kx, x'=\exp-\gamma x \cdot x'</math></p> | <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16（是否引证：是）</p> <p>1.=1相应的分类函数变为：<math>fix = Sgp(^a yMx^X_j)+b'</math>（3-9）1=1综上所述，SVM算法通过内积核函数将低维空间中的非线性问题转为高维空间中的线性划分问题来求解，并在高维空间中通过降维等方式简化运算复杂度，进而求解最优分类面的过程。（3）支持向量机的核函数常用的SVM核函数有以下三种类型：—是多项式核函数，其表达式为：<math>\gamma(x=, x, ) (ca \geq x, ; 3 \geq 0</math>；（3-10）二是径向基核函数（RBF）：<math>K(x=,x-) \exp(-r) x</math></p> <p>于俊 基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31（是否引证：否）</p> <p>1.度，此时的目标函数变为：；式（3-7）相应的分类函数变为：。式（3-8）综上所述，支持向量机算法通过内积核函数将低维空间中的非线性问题转为高维空间中的线性划分问题来求解，并在高维空间中通过降维等方式简化运算复杂度，进而求解最优分类面的过程。（3）支持向量机的核函数常用的SVM核函数有如下三种类型：—是多项式核函数，其表达式为：；式（3-9）二是径向基核函数（RBF）；；式（3-10）三是Sigmoid核函数：。式（3-11）</p> <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09（是否引证：否）</p> <p>1.面求解问题的复杂度，此时的目标函数变为：；（3-7）相应的分类函数变为：（3-8）综上所述，支持向量机算法通过内积核函数将低维空间中的非线性问题转为高维空间中的线性划分问题来求解，并在高维空间中通过降维等方式简化运算复杂度，进而求解最优分类面的过程。（3）支持向量机的核函数常用的SVM核函数有以下三种类型：—是多项式核函数，其表达式为：；</p> |

|    |  |  |
|----|--|--|
|    |  | <p>( 3-9 ) 二是径向基核函数 ( RBF ) ; ; ( 3-10 ) 三是 Sigmoid 核函数 : 。 ( 3-11 )</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证 : 否 )</p> <p>1.)sgn((,))*1*fxyKxxbjniii ( 3-8 ) 综上所述,支持向量机算法通过内积核函数将低维空间中的非线性问题转为高维空间中的线性划分问题来求解,并在高维空间中通过降维等方式简化运算复杂度,进而求解最优分类面的过程。( 3 ) 支持向量机的核函数常用的 SVM 核函数有以下三种类型:一是多项式核函数,其表达式为:<br/>(,)(,0,,qK xxx; ( 3-9 ) 二是径向基核函数 ( RBF );<br/>(,)exp(    ),0</p>   |
| 29 | <p>此处有 62 字相似</p> <p><math>x_p - \gamma x - x'^2</math>, <math>\gamma &gt; 0</math> ;<br/>( 3-15 )</p> <p>三是 Sigmoid 核函数: <math>Kx</math>, <math>x' = \tanh(\gamma x \cdot x + c)</math>。<br/>( 3-16 )</p> <p>其中最常使用的是径向基核函数,该核函数公式简洁,推广性好,在小样本分类中具有较高的分类准确率和召回率。</p> <p>3.3.3 SVM 算法的主要优点</p> <p>相比于传统的分类算法,SVM 具有以下几个显著的优点:</p> <p>( 1 ) 通过使用 SLT 的最小化</p> | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证 : 否 )</p> <p>1. 数,其表达式为: ; ( 3-9 ) 二是径向基核函数 ( RBF ) ; ; ( 3-10 ) 三是 Sigmoid 核函数: 。 ( 3-11 ) 其中最常使用的是径向基核函数,该核函数公式简洁,推广性好,在小样本分类中具有较高的分类准确率和召回率。3.4.3 支持向量机的主要优点支持向量机算法的优点主要有以下三点[38]: (4) 支持向量机算法基于统</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证 : 否 )</p> <p>1. <math>xx \times x</math> ; ( 3-10 ) 三是 Sigmoid 核函数:<br/>(,)tanh(),,K <math>xx \times x \times c</math>。 ( 3-11 ) 其中最常使用的是径向基核函数,该核函数公式简洁,推广性好,在小样本分类中具有较高的分类准确率和召回率。3.4.3 支持向量机的主要优点支持向量机算法的优点主要有以下三点 [38]: (1) 支持向量机算法基于统</p> <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证 : 否 )</p> <p>1., 其表达式为: ; 式 ( 3-9 ) 二是径向基核函数 ( RBF ) ; ; 式 ( 3-10 ) 三是 Sigmoid 核函数: 。 式 ( 3-11 ) 其中最常使用的是径向基核函数,该核函数公式简洁,推广性好,在小样本分类中具有较高的分类准确率和召回率。9. 3.4.3 支持向量机的主要优点支持向量机算法的优点主要有以下三点[38]: (1) 支持向量机算法基于</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证 : 是 )</p> <p>1.- f),<math>r &gt; 0</math>; (3-11)三是 Sigmoid 核函数: <math>Ar(x = , x')</math><br/><math>\tanh Oa.jc' + c</math> )。 ( 3-12 ) 其中最常使用的是径向基核函数,该核函数公式简洁,推广性好,在小样本分类中具有较高的分类准确率和召回率。3.4.3 支持向量机的主要优点支持向量机算法的优点主要有以下三点[38]: (1) 支持向量机算法基于统计学习理</p> |
| 30 | <p>此处有 64 字相似</p> <p>。</p>   | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证 : 否 )</p> <p>1. 向量机的主要优点支持向量机算法的优点主要有以下</p>  |

|  |   |  |
|--|---|--|
|  | <p>3.3.3 SVM算法的主要优点</p> <p>相比于传统的分类算法，SVM具有以下几个显著的优点：</p> <p>(1) 通过使用SLT的<br/>最小化结构风险原理使其不仅具备传统分类算法的经验风险最小化的优点，还能够通过求解最大分类间隔而避免分类模型出现过拟合的问题。</p> <p>(2) 求得的最优解既是局部也是全局最优解，避免了基于规则的分类和基于贪心算法的神经网络由于求解过程基于局部最优原理而出现</p>               | <p>三点[38]：(1) 支持向量机算法基于统计学习理论的结构风险最小化原理，因此其不仅具有传统分类算法的经验风险最小化的优点，同时也通过求解最大分类间隔使得分类模型避免过拟合的问题。(2) 支持向量机算法是一个凸优化问题，所求解的局部最优解同时也是全局最优解，避免了基于规则的分类器和基于贪心算法的神经网络</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1. 机的主要优点支持向量机算法的优点主要有以下三点[38]：(4) 支持向量机算法基于统计学习理论的结构风险最小化原理，因此其不仅具有传统分类算法的经验风险最小化的优点，同时也通过求解最大分类间隔使得分类模型避免过拟合的问题。(5) 支持向量机算法求解的局部最优解同时也是全局最优解，避免了基于规则的分类器和基于贪心算法的神经网络中出现的局部</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1. 机的主要优点支持向量机算法的优点主要有以下三点[38]：(1) 支持向量机算法基于统计学习理论的结构风险最小化原理，因此其不仅具有传统分类算法的经验风险最小化的优点，同时也通过求解最大分类间隔使得分类模型避免过拟合的问题。(2) 支持向量机算法求解的局部最优解同时也是全局最优解，避免了基于规则的分类器和基于贪心算法的神经网络中出现的局部</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1. 持向量机的主要优点支持向量机算法的优点主要有以下三点[38]:(1)支持向量机算法基于统计学习理论的结构风险最小化原理，因此其不仅具有传统分类算法的经验风险最小化的优点，同时也通过求解最大分类间隔使得分类模型避免过拟合的问题。(2)支持向量机算法求解的局部最优解同时也是全局最优解，避免了基于规则的分类器和基于贪心算法的神经网络中出现的局部</p> |
|  | <p>此处有 68 字相似</p> <p>其不仅具备传统分类算法的经验风险最小化的优点，还能够通过求解最大分类间隔而避免分类模型出现过拟合的问题。</p> <p>(2) 求得的最优解既是局部也是全局最优解，避免了基于规则的分类和基于贪心算法的神经网络由于求解过程基于局部最优原理而出现结果不是全局最优的问题。</p> <p>(3) 利用低维度内积核函数将低维空间的非线性分类问题转化为高维空间的线性分类问题却不增加算法的复杂程度，提高了算法的适用</p> | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1. 的经验风险最小化的优点，同时也通过求解最大分类间隔使得分类模型避免过拟合的问题。(2) 支持向量机算法求解的局部最优解同时也是全局最优解，避免了基于规则的分类器和基于贪心算法的神经网络中出现的局部最优而非全局最优的问题。(3) 支持向量机算法通过使用内积核函数技术在不增加算法复杂度的情况下将低维空间的非线性分类问题转换为高维空间的线性</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1. 经验风险最小化的优点，同时也通过求解最大分类间隔使得分类模型避免过拟合的问题。(2)支持向量机算法求解的局部最优解同时也是全局最优解，避免了基于规则的分类器和基于贪心算法的神经网络中出现的局部最优而非全局最优的问题。(3)支持向量机算法通过使用内</p>  |
|  |   |  |
|  |   |  |
|  |   |  |

|    |   |   |
|----|---|---|
|    |   | 积核函数技术在不增加算法复杂度的情况下将低维空间的非线性分类问题转换为高维空间的线性  |
| 32 | <p>此处有 35 字相似</p> <p>最优解，避免了基于规则的分类和基于贪心算法的神经网络由于求解过程基于局部最优原理而出现结果不是全局最优的问题。</p> <p>(3)</p> <p>利用低维度内积核函数将低维空间的非线性分类问题转化为高维空间的线性分类</p> <p>问题却不增加算法的复杂程度，提高了算法的适用性。</p> <p>3.4 本章小结</p> <p>本章简单介绍了本文用于恶意网页分类检测的四种分类</p>  | <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1.表示为：，当且仅当对于任意满足为有限值的函数，则。式(3-6)因此，在求解非线性分类的最优分类面问题中可以使用内积核函数将低维空间中的非线性分类转化为高维空间中的线性分类，从而降低低维空间非线性分类最优平面求解问题的复杂度，此时的目标函数变为：；式(3-7)相应的分类函数变为：</p> <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.xdx2()为有限值的函数g(x)，则。(3-6)因此，在求解非线性分类的最优分类面问题中可以使用内积核函数(,ijKxx将低维空间中的非线性分类转化为高维空间中的线性分类，从而降低低维空间非线性分类最优平面求解问题的复杂度，此时的目标函数变为：；(3-7)相应的分类函数变为：</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.(x)，则K(x,y)g(x)g(y)dxdy0(3-6)因此，在求解非线性分类的最优分类面问题中可以使用内积核函数(,ijKxx将低维空间中的非线性分类转化为高维空间中的线性分类，从而降低低维空间非线性分类最优平面求解问题的复杂度，此时的目标函数变为：(,21)1,1</p> <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.c)，则/Kix,y)gix)g(y)dxdy &gt;0 (3-7)因此，在求解非线性分类的最优分类面问题中可以使用内积核函数将低维空间中的非线性分类转化为高维空间中的线性分类，从而降低低维空间非线性分类最优平面求解问题的复杂度，此时的目标函数变为：/1 G(^ 1 n ,?),X</p> |
| 33 | <p>此处有 157 字相似</p> <p>用低维度内积核函数将低维空间的非线性分类问题转化为高维空间的线性分类问题却不增加算法的复杂程度，提高了算法的适用性。</p> <p>3.4 本章小结</p> <p>本章简单介绍了本文用于恶意网页分类检测的四种分类学习算法，主要包括朴素贝叶斯算法、决策树C4.5算法、分类回归树算法(CART)及支持向量机算法(SVM)。分别从分类学习算法的基本原理、主要思想、一般步骤、关键技术和算法的优缺点及适用场景等方面进行描述，并使用公式描述了各个算法的核心原理。</p> <p>4 恶意链接检测系统的设计和实现</p> <p>本章内容在基于前三章介绍的基础上，来具体介绍和实现恶意链接检测系统通过制定系统的设计</p> | <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.点，对现有的机器学习算法进行改进使其具有增量学习、在线学习的能力是研究增量机器学习算法的一种有效途径。3.6本章小结本章简单介绍了本文用于恶意网页分类检测的四种分类学习算法，主要包括朴素贝叶斯算法、决策树C4.5算法、分类回归树算法(CART)及支持向量机算法(SVM)。分别从分类学习算法的基本原理、主要思想、一般步骤、关键技术和算法的优缺点及适用场景等方面进行描述，并使用公式描述了各个算法的核心原理。介绍了评价分类预测实验结果的正率率(TPR)、假正率(FPR)、准确率A、召回率R、精确率P等相关概念、计算公式及</p> <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1.意网页的样本量中实际为真正恶意网页所占的比例，反应分类模型对恶意网页的查准率。3.6 本章小结本章简单介绍了本文用于恶意网页检测分类学习的四种分</p>   |



|    |   |   |
|----|---|---|
| 34 |   | <p>类学习算法，主要包括朴素贝叶斯算法Navie Bayes、决策树算法C4.5、分类回归树算法CART及支持向量机算法SVM。分别从分类学习算法的基本原理、主要思想、一般步骤、关键技术和算法的优缺点及适用场景等方面进行描述，并使用公式描述了各个算法的核心原理。介绍了评价分类预测实验结果的真正率TPR、假正率FPR、准确率A、召回率R、精确率P等相关概念、计算公式及对分类学习算法</p>  |
|    |   | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p>  |
|    |   | <p>1.进行改进使其具有增量学习、在线学习的能力是研究增量机器学习算法的一种有效途径。3.6 本章小结本章简单介绍了本文用于恶意网页检测分类学习的四种分类学习算法，主要包括朴素贝叶斯算法 Navie Bayes、决策树算法 C4.5、分类回归树算法 CART 及支持向量机算法 SVM。分别从分类学习算法的基本原理、主要思想、一般步骤、关键技术和算法的优缺点及适用场景等方面进行描述，并使用公式描述了各个算法的核心原理。介绍了评价分类预测实验结果的真正率 TPR、假正率 FPR、准确率 A、召回率 R、精确率 P 等相关概念、计算公式</p> |
|    |   | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p>   |
| 34 | <p>此处有 84 字相似</p> <p>文设计和实现了基于SVM和DOM检测算法的恶意链接检测系统(SD-MLDS)，SD-MLDS系统利用基于黑/白名单库测出</p> <p>已知类型的恶意链接，然后对未知类型的链接使用分类算法进行特征学习和类别预测。SD-MLD系统共分为三个子模块，分别为:基于黑/白名单库的检测模块，基于SVM的检测模块和</p> <p>基于TDD算法进行恶意链接截取的检测模块。SD-MLD系统结构如图4-1所示-</p> <p>图 41 基于基于SVM和DOM检测算</p> | <p>1.率及召回率，根据分类结果对分类模型进行参数调整和优化。为了能够准确地对恶意网页进行类别预测与判定，我们分别使用了朴素贝叶斯算法、决策树 C4.5 算法、分类回归树算法 ( CART )、支持向量机算法 ( SVM ) 来进行恶意网页类型的分类判定。本章系统的介绍了四种分类学习算法的基本概念、核心理论、一般步骤及各分类算法的优</p>  |
|    |   | <p>2.贝叶斯算法 ( Navie Bayes )、决策树 C4.5 算法、分类回归树算法 ( CART ) 及支持向量机算法 ( SVM )。分别从分类学习算法的基本原理、主要思想、一般步骤、关键技术和算法的优缺点及适用场景等方面进行描述，并使用公式描述了各个算法的核心原理。介绍了评价分类预测实验结果的真正率 TPR、假正率 FPR、准确率 A、召回率 R、精确率 P 等相关概</p>  |
|    |   | <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p>  |
| 34 |   | <p>1.实现了基于分类算法的恶意网页检测系统(CA-MWDS), CA-MWDS系统首先使用基于知识库的特征匹配技术检测出已知类型的恶意网页，然后对未知类型的网页使用分类算法进行样本特征学习和类别预测。CA-MWDS系统共分为三个子模块，分别为：基于知识库的特征匹配检测子模块，基于分类算法的恶意检测子模块和根据新采集的恶意网页样本集对系统当前分类器的自适应学习子模块。CA-MWDS系统结构如图4-1所示- CA-M</p>  |
|    |   | <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究</p>   |

|    |  |   |
|----|--|---|
|    |  | <p>究 2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证: 否)</p> <p>1.能力,我们设计和实现了基于分类算法的恶意网页检测系统(CA-MWDS),该系统首先使用基于知识库的特征匹配技术检测出<b>已知类型的恶意网页</b>,然后对<b>未知类型的网页使用分类算法进行样本特征学习和类别预测</b>。基于分类算法的恶意网页检测系统(CA-MWDS)共分为三个子模块,分别为:基于知识库的特征匹配检测子模块,基于分</p>  |
| 35 | <p>此处有 99 字相似</p> <p>M和DOM检测算法的恶意链接检测系统</p> <p>以下是几个模块具体功能的介绍:</p> <p>(1) 基于黑白名单库的特征匹配检测模块:该模块利用包含恶意链接地址的黑/白名单库对已知类型的恶意链接进行检测,通过利用黑/白名单库包含的知识特征对待检测的URL样本集进行字符串和特征匹配,从而检测出已知类型的恶意链接。</p> <p>(2) 基于SVM的检测</p> <p>模块:将黑白名单库中的恶意链接作为训练样本集并用SVM算法对样本集内数据的属性特征进行学习构建分类预测模型,并以此为基础</p> | <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 (是否引证: 否)</p> <p>1. 使用链接地址正则表达式从待检测网页源码中抽取所有的链接地址 URL;3) 计算每条链接地址 md5 值;4) <b>使用链接地址白名单,黑名单对链接的 md5 值进行过滤</b>。首先进行链接地址的白名单过滤,如果该链接 md5 值在白名单中,则直接忽略该链接地址 URL;如果不在,则进行链</p> <p>2.,其中知识库包括链接地址URL 黑/白名单、恶意代码特征库。该模块主要进行暗链、恶意链接与恶意代码(恶意脚本)的<b>检测,使用知识库的知识特征对待检测网页源代码进行恶意特征匹配,直接检测出已知特征的恶意网页代码。</b>(2) <b>基于分类算法的恶意检测子模块</b>:本模块使用分类学习算法对恶意网页样本集的属性特征分布进行机器学习,并在此基础上训练和生成恶意网页分类器,</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究 2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证: 否)</p> <p>1. 使用链接地址正则表达式从待检测网页源码中抽取所有的链接地址 URL;3) 计算每条链接地址 md5 值;4) <b>使用链接地址白名单,黑名单对链接的 md5 值进行过滤</b>。首先进行链接地址的白名单过滤,如果该链接 md5 值在白名单中,则直接忽略该链接地址 URL;如果不在,则进行链</p> <p>2.,其中知识库包括链接地址URL 黑/白名单、恶意代码特征库。该模块主要进行暗链、恶意链接与恶意代码(恶意脚本)的<b>检测,使用知识库的知识特征对待检测网页源代码进行恶意特征匹配,直接检测出已知特征的恶意网页代码。</b>(2) <b>基于分类算法的恶意检测子模块</b>:本模块使用分类学习算法对恶意网页样本集的属性特征分布进行机器学习,并在此基础上训练和生成恶意网页分类器,</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 (是否引证: 是)</p> <p>1.特征库。该模块主要进行暗链、恶意链接与恶意代码(恶意脚23北京邮电大学工程硕士学位论文 本)的<b>检测,使用知识库的知识特征对待检测网页源代码进行恶意特征匹配,直接检测出已知特征的恶意网页代码。</b>(2)基于分类算法的恶意检测子模块:使用分类学习算法对恶意网页样本集的属性特征分布进行机</p> |

|    |  |   |
|----|--|---|
|    |  | <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证:否)</p> <p>1.白名单、恶意代码特征库。该模块主要进行暗链、恶意链接与恶意代码(恶意脚本)的检测,使用知识库的知识特征对待检测网页源代码进行恶意特征匹配,直接检测出已知特征的恶意网页代码。(2)基于分类算法的恶意检测子模块:本模块使用分类学习算法对恶意网页样本集的属性特征分布进行机器学习,并在此基础上训练和生成恶意网页分类器,然后使用</p>   |
| 36 | <p>此处有 39 字相似</p> <p>意链接。</p> <p>(2)基于SVM的检测模块:将黑白名单库中的恶意链接作为训练样本集并用SVM算法对样本集内数据的属性特征进行学习构建分类预测模型,并以此为基础生成恶意链接分类器,使用该分类器对未知类型的链接进行分类判断,并将结果更新黑白名单库,并用该库作为新的样本集更新分类预测模型。</p> <p>(3)基于TDD算法的检测模块:通</p>   | <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证:否)</p> <p>1.征的恶意网页代码。(2)基于分类算法的恶意检测子模块:本模块使用分类学习算法对恶意网页样本集的属性特征分布进行机器学习,并在此基础上训练和生成恶意网页分类器,然后使用分类器对未知类别的待检测网页进行分类预测。(3)分类器自适应学习子模块:本模块使用自适应分类学习算法在基于分类算法的恶意检测子模块中</p> <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证:否)</p> <p>1.的恶意网页代码。(2)基于分类算法的恶意检测子模块:本模块使用分类学习算法对恶意网页样本集的属性特征分布进行机器学习,并在此基础上训练和生成恶意网页分类器,然后使用分类器对未知类别的待检测网页进行分类预测。(3)分类器自适应学习子模块:本模块使用自适应分类学习算法在基于分类算法的恶意检测子模块</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究 2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证:否)</p> <p>1.的恶意网页代码。(2)基于分类算法的恶意检测子模块:本模块使用分类学习算法对恶意网页样本集的属性特征分布进行机器学习,并在此基础上训练和生成恶意网页分类器,然后使用分类器对未知类别的待检测网页进行分类预测。(3)分类器自适应学习子模块:本模块使用自适应分类学习算法在基于分类算法的恶意检测子模块</p> |
| 37 | <p>此处有 119 字相似</p> <p>面DOM结构的改变,将新增的DOM节点识别出来,提取其中JavaScript脚本中存在的暗链接,并将该链接送入前两个模块进行链接检测。</p> <p>(3)分类器自适应学习子模块:使用自适应分类学习算法在基于分类算法的恶意检测子模块中使用的分类器基础上,根据新采集的恶意网页样本集进行自适应学习,训练并生成能够识别新恶意网页样本特征的新分类器。</p> <p>2、系统的工作流程</p> <p>本系统的工作流程图如图 4-2 所示:</p> <p>Step1:系统读取格式化后的恶意链接样本集,经过数据排重过滤掉重复的数据。</p> | <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证:是)</p> <p>1.样本集的属性特征分布进行机器学习,并在此基础上训练和生成恶意网页分类器,然后使用分类器对未知类别的待检测网页进行分类预测。(3)分类器自适应学习子模块:使用自适应分类学习算法在基于分类算法的恶意检测子模块中使用的分类器基础上,根据新采集的恶意网页样本集进行自适应学习,训练并生成能够识别新恶意网页样本特征的新分类器。4.1.2工作流程编出基于知K库的#^“S于知识库的特征匹配 恶麵翻结果‘亚在古中A=i 检测子模块 未知类型的I</p> <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证:否)</p> <p>1.上训练和生成恶意网页分类器,然后使用分类器对未</p>   |

|    |   |   |
|----|---|---|
|    |   | <p>知类别的待检测网页进行分类预测。(3) 分类器自适应学习子模块：本模块使用自适应分类学习算法在基于分类算法的恶意检测子模块中使用的分类器基础上，根据新采集的恶意网页样本集进行自适应学习，训练并生成能够识别新恶意网页样本特征的新分类器。11. 4.1.2 基于分类算法的恶意网页检测系统CA-MWDS工作流程图4-2 基于分类算法的恶意网页检测系统C</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.训练和生成恶意网页分类器，然后使用分类器对未知类别的待检测网页进行分类预测。(3) 分类器自适应学习子模块：本模块使用自适应分类学习算法在基于分类算法的恶意检测子模块中使用的分类器基础上，根据新采集的恶意网页样本集进行自适应学习，训练并生成能够识别新恶意网页样本特征的新分类器。4.1.2 CA-MWDS 恶意网页检测系统工作流程基于知识库的特征匹配检测子模块基于分类算法的恶意检测</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.的待检测网页进行分类预测。(3) 分类器自适应学习子模块：本模块使用自适应分类学习算法在基于分类算法的恶意检测子模块中使用的分类器基础上，根据新采集的恶意网页样本集进行自适应学习，训练并生成能够识别新恶意网页样本特征的新分类器。4.1.2 CA-MWDS 恶意网页检测系统工作流程基于知识库的特征匹配检测子模块基于分类算法的恶意检测</p> |
| 38 | <p>此处有 115 字相似</p> <p>，采用的工具为：</p> <p>开发环境：VS2012</p> <p>数据库：Postgres和Redis</p> <p>发布程序：Tomcat</p> <p>本</p> <p>实验所在运行环境为台式计算机一台，酷睿双核 CPU，主频 3.2GHZ，</p> <p>内存 8GB，网络带宽 100Mbps，操作系统 Windows 7，分类平台 Weka 以及 Google Chrome 20.0 浏览器。</p> <p>4.4</p> <p>数据排重模块设计</p> <p>从本系统设计的角度来看，数据排重[35]的目标是挖掘相似文本，用于数据去重，降低后续流程计算的数据量</p> | <p>基于决策树的搜索引擎恶意网页检测研究与实现 周浩 - 《湖南大学硕士论文》- 2013-05-27 (是否引证：否)</p> <p>1.图 4.1 Chrome 扩展获取的域名信息4.2 实验环境及分类平台 Weka4.2.1 实验环境本实验所在运行环境为台式计算机一台，Pentium 双核 CPU，主频 3.2GHZ，内存 2GB，网络带宽 100Mbps，操作系统 Windows XP，编程环境 Eclipse，分类平台 Weka 以及 Google Chrome 20.0 浏览器。4.2.2 分类平台 WekaWeka 全称为怀卡托智能分析环境，作为一款完全免费并且基于 Java 环境下的开源数据挖</p>   |
| 39 | <p>此处有 263 字相似</p> <p>个事件的文本内容不完全相等，需要柔性比较。从目前技术的普及程度来看，很容易想到使用基于Hash的 MD5排重技术。</p> <p>但是</p> <p>查找时间方面，其查找过程实际是将传入数据的MD5码与全集 S 的 MD5 码元素按集合索引一条条比对，时间复杂度为 <math>O(n)</math>，若传入文件全为新文件，则每次查找</p>   | <p>谁能诠释下"单实例存储" TechTarget技术社区 - 技术论坛 - TechTarget中国 - 《网络(<a href="http://club.techtarget">http://club.techtarget</a>)》- (是否引证：否)</p> <p>1.sh) 与备份设备上散列索引中的一个散列匹配，表明该数据已经被备份，设备只更新它的表，以说明在这个新位置上也存在该数据。基于散列(hash)的方法存在内置的可扩展性问题。为了快速识别一个数据块是否已经被备份，这种基于散列(hash)的方法会在内存中</p>  |



|   |   |
|---|---|
| <p>都要将集合从头查到尾，所以该算法的查找时间并不尽人意。</p> <p>在存储代价方面，基于 hash 值的排重方法存在内置的可扩展性问题，为了快速识别一个数据块是否已经重复，这种基于 hash 值的方法会在内存中保存指纹值索引。当被备份的数据块数量增加时，该索引也随之增长，一旦索引增长到超过了设备在内存中保存它所支持的容量，将增加大量的磁盘 IO，因为磁盘搜索会比内存搜索更慢，系统性能会急速下降。</p> | <p>拥有散列 ( hash ) 索引。当被备份的数据块数量增加时，该索引也随之增长。一旦索引增长超过了设备在内存中保存它所支持的容量，性能会急速下降，同时磁盘搜索会比内存搜索更慢。因此，目前大部分基于散列 ( hash ) 的系统都是独立的，可以保持存储数据所需的内存量与磁盘空间量的平衡，这样，散列 ( hash ) 表</p> <p>基于虚拟磁带库的容灾系统设计与实现 - 《学术论文联合比对库》 - 2012-11-06 ( 是否引证：否 )</p> <p>1.与备份设备上散列索引中的一个散列匹配，表明该数据已经被备份，设备只更新它的表，以说明在这个新位置上也存在该数据。基于散列(hash)的方法存在内置的可扩展性问题。为了快速识别一个数据块是否已经被备份，这种基于散列(hash)的方法会在内存中拥有散列(hash)索引。当被备份的数据块数量增加时，该索引也随之增长。一旦索引增长超过了设备在内存中保存它所支持的容量，性能会急速下降，同时磁盘搜索会比内存搜索更慢。因此，目前大部分基于散列(hash)的系统都是独立的，可以保持存储数据所需的内存量与磁盘空间量的平衡，这样，散列(hash)表</p> <p>基于双布鲁姆过滤器的数据排重算法及其应用 席晔文 - 《湖南大学硕士论文》 - 2013-04-23 ( 是否引证：否 )</p> <p>1.的指纹值在备份设备上指纹集合中的找到匹配，表明该数据已经被备份，设备只需更新它的元数据增加该元数据的引用数。基于 hash 值的排重方法存在内置的可扩展性问题，为了快速识别一个数据块是否已经重复，这种基于 hash 值的方法会在内存中保存指纹值索引。当被备份的数据块数量增加时，该索引也随之增长，一旦索引增长到超过了设备在内存中保存它所支持的容量，将增加大量的磁盘 IO，因为磁盘搜索会比内存搜索更慢，系6工程硕士学位论文统性能会急速下降。因此，目前大部分基于 hash 值的排重系统都是独立的，以此</p> <p>重复数据删除技术在船舶信息化建设中的应用 杜阳; - 《天津航海》 - 2014-06-30 ( 是否引证：否 )</p> <p>1.ash)与备份设备上散列索引中的一个散列匹配,表明该数据已经被备份,设备只更新它的表,以说明在这个新位置上也存在该数据。基于散列(hash)的方法存在内置的可扩展性问题。为了快速识别一个数据块是否已经被备份,这种基于散列(hash)的方法会在内存中拥有散列(hash)索引。当被备份的数据块数量增加时,该索引也随之增长。一旦索引增长超过了设备在内存中保存它所支持的容量,性能会急速下降,同时磁盘搜索会比内存搜索更慢。因此,目前大部分基于散列(hash)的系统都是独立的,可以保持存储数据所需的内存量与磁盘空间量的平衡,这样,散列(has</p> <p>基于双布鲁姆过滤器的数据排重技术 席晔文, 杨金民 - 《计算机工程与应用 ( 优先出版 ) 》 - 2013-04-08 16:46 ( 是否引证：是 )</p> <p>1.合 S 中查找，若发现集合中存在相同的 MD5 码，说明该传入文件是重复文件，删除该文件，就达到了数据排重的目的。在效率方面，其查找过程就是将传入文件</p> |
|---|---|

|  |  |
|--|--|
|  | <p>MD5 码与全集 S 的 MD5 码元素按集合索引一条条比对，时间复杂度为 <math>O(n)</math>，若传入文件全为新文件，则每次查找都要将集合从头查到尾，所以该算法的查找时间并不尽人意，没有对 MD5 码进行任何处理直接进行查找，对其利用率还处于原始阶段。而通过使用布鲁姆过滤器技术来处理 MD5 码可以帮</p> <p>基于双布鲁姆过滤器的数据排重技术 席晔文;杨金民; - 《计算机工程与应用》- 2013-04-08 1 (是否引证：是)</p> <p>1.合S中查找,若发现集合中存在相同的MD5码,说明该传入文件是重复文件,删除该文件,就达到了数据排重的目的。在效率方面,其查找过程就是将传入文件MD5码与全集S的MD5码元素按集合索引一条条比对,时间复杂度为<math>O(n)</math>若传入文件全为新文件,则每次查找都要将集合从头查到尾,所以该算法的查找时间并不尽人意,没有对MD5码进行任何处理直接进行查找,对其利用率还处于原始阶段。而通过使用布鲁姆过滤器技术来处理MD5码可以帮助减少查</p> <p>重复数据删除技术的研究与探讨 成杰;李恒志;祁悦; - 《中国无线电》- 2012-10-26 (是否引证：否)</p> <p>1.ash)与备份设备上散列索引中的一个散列匹配,表明该数据已经被备份,设备只更新它的表,以说明在这个新位置上也存在该数据。基于散列(hash)的方法存在内置的可扩展性问题。为了快速识别一个数据块是否已经被备份,这种基于散列(hash)的方法会在内存中拥有散列(hash)索引。当被备份的数据块数量增加时,该索引也随之增长。一旦索引增长超过了设备在内存中保存它所支持的</p> <p>2.快速识别一个数据块是否已经被备份,这种基于散列(hash)的方法会在内存中拥有散列(hash)索引。当被备份的数据块数量增加时,该索引也随之增长。一旦索引增长超过了设备在内存中保存它所支持的容量,性能会急速下降。因此,目前大部分基于散列(hash)的系统都是独立的,可以保持存储数据所需的内存量与磁盘空间量的平衡,这</p> |
|--|--|

| 指 标      |   |
|----------|---|
| 疑似剽窃文字表述 |   |
| 1.       | 一个空的决策树，决策时不断向其中添加新的叶子节点作为分类节点，重复该过程，直到构造出的决策树可以将训练集样本正确的分好类为止。   |
| 2.       | 计算样本集中所有可能属性对应的信息熵和信息增益，将信息最大增益的属性用作构建决策树时的根节点，   |
| 3.       | 处理过程如下：<br>Step1: 首先利用离散算法将训练集内的属性的连续指进行离散化处理，假定样本集S，   |
| 4.       | 插入n个点对序列进行分割，对于分割点i，其取值为Di和 Di+1的平均值，这样一来，分割点将D'中的每个值分割成了两个子区间。   |
| 5.       | Step5：利用训练好的决策树对未知样本进行分类。<br>C4.5算法可以说是目前决策树模型中的比较重要和核心的算法，   |
| 6.       | 实验所在运行环境为台式计算机一台，酷睿双核 CPU，主频 3.2GHZ，<br>内存 8GB，网络带宽 100Mbps，操作系统 Windows 7，分类平台 Weka 以及 Google Chrome 20.0 浏览器。                                     |
| 7.       | 基于 hash 值的排重方法存在内置的可扩展性问题，为了快速识别一个数据块是否已经重复，这种基于 hash 值的方法会在内存中保存指纹值索引。当被备份的数据块数量增加时，该索引也随之增长，一旦索引增长到超过了设备在内存中保存它所支持的容量，将增加大量的磁盘 IO，因为磁盘搜索会比内存搜索更慢， |

| 4. 11491915_刘力铭_第4部分              |  | 总字数：9378               |
|-----------------------------------|--|------------------------|
| 相似文献列表 文字复制比：2.4%(222) 疑似剽窃观点：(0) |  |                        |
| 1                                 | 于俊_基于等级容灾模型<br>于俊 - 《学术论文联合比对库》 - 2014-12-31   | 1.7% ( 156 )<br>是否引证：否 |
| 2                                 | 基于分类算法的恶意网页检测技术研究<br>王维光(导师：王枳) - 《北京邮电大学硕士论文》 - 2015-03-16  | 1.2% ( 114 )<br>是否引证：是 |
| 3                                 | 王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15<br>王维光 - 《学术论文联合比对库》 - 2015-01-15                            | 1.2% ( 114 )<br>是否引证：否 |
| 4                                 | 恶意网页医疗数据应急通信移动通信<br>- 《学术论文联合比对库》 - 2015-01-09   | 1.2% ( 114 )<br>是否引证：否 |
| 5                                 | 布隆过滤器(Bloom Filter)详解<br>- 《互联网文档资源 ( <a href="http://www.360doc.co">http://www.360doc.co</a> ) 》 - 2015 | 0.7% ( 65 )<br>是否引证：否  |
| 6                                 | 一种面向用户的无线局域网冗余消除机制<br>周新生(导师：薛广涛) - 《上海交通大学硕士论文》 - 2015-06-30  | 0.5% ( 51 )<br>是否引证：否  |

| 原文内容 |   | 相似内容来源  |
|------|---|---|
| 1    | <p>此处有 65 字相似</p> <p><math>-k \approx 0.6185m/n</math></p> <p>而当p的大小确定是，位向量的最优位数m可以通过<math>m = -n \ln p / \ln 2</math>求得。</p> <p>以上分析表明，</p> <p>位数组的大小最好与插入元素的个数成线性关系，对于给定的m，n，k，假正例概率最大为：<math>1 - e^{-kn} + 0.5/m - 1k</math>。</p> <p>布隆过滤器</p> <p>牺牲了很小的准确率，换来的是存储成本的极大降低和查找速度的较大提高。因此，在对误判率要求不高，而对存储和查找速度要求很高</p> | <p>布隆过滤器(Bloom Filter)详解 - 《互联网文档资源 ( <a href="http://www.360doc.co">http://www.360doc.co</a> ) 》 - 2015-6-18 19:26:42 ( 是否引证：否 )</p> <p>1.tives的概率为：而对于给定的False Positives概率p，如何选择最优的位数组大小m呢，上式表明，位数组的大小最好与插入元素的个数成线性关系，对于给定的m，n，k，假正例概率最大为：下图是布隆过滤器假正例概率p与位数组大小m和集合中插入元素个数n的关系图，假定Hash函数个数</p>   |
| 2    | <p>此处有 43 字相似</p> <p>意链接，同时，将该链接记录到系统的关系型数据库当中去。</p> <p>该模块使用的黑/白名单库中的黑名单存储的是已经被确定检测为恶意</p> <p>链接的木马、钓鱼、病毒、诈骗等链接。白名单中存储的是经过安全认证的正常站点的链接地址。</p> <p>下面将具体描述这两个子模块的设计与实现。</p> <p>具体的检测过程是，将待检测的未知链接样本集输入到该模块的入口后，利用上一小节</p>  | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.上很难被发现的链接，其危害已经在前面章节中介绍过。使用的知识库为链接地址黑/白名单，链接地址黑名单中存储已知的暗链、木马链接、钓鱼链接、病毒链接、其它链接等，链接地址白名单中存储已知的安全链接地址。暗链与恶意链接检测分为网页暗链检测子模块和网页恶意链接检测子模块，下面将具体描述这两个子模块的设计与实现。(1)</p> <p>恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )</p> <p>1.发现的链接，其危害已经在前面章节中介绍过。使用的知识库为链接地址黑/白名单，链接地址黑名单中存储已知的暗链、木马链接、钓鱼链接、病毒链接、其它链接等，链接地址白名单中存储已知的安全链接地址。暗链与恶意链接检测分为网页暗链检测子模块和网页恶意链接检测子模块，下面将具体描述这两个子模块的设计与实现。(1)</p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p> <p>1.的链接，其危害已经在前面章节中介绍过。使用的知识库为链接地址黑/白名单，链接地址黑名单中存储已知的暗链、木马链接、钓鱼链接、病毒链接、其它链接等</p> |

|   |   |  |
|---|---|--|
|   |   | <p>，链接地址白名单中存储已知的安全链接地址。暗链与恶意链接检测分为网页暗链检测子模块和网页恶意链接检测子模块，下面将具体描述这两个子模块的设计与实现。1</p>   |
| 3 | <p>此处有 76 字相似</p> <p>全检测。由于链接的判定具有二分性，即要么是恶意要么是非恶意，因此，本系统采用SVM算法来构建系统的分类引擎。</p> <p>SVM的</p> <p>本质是在样本训练集上构造一个高维度超平面来作为正负样本的分界面，并使它们之间的边界距离最大化。并且SVM通过使用核函数可以处理正负样本线性不可分的情况。</p> <p>具体的原理本文在第三章有详细的介绍，这里不再赘述。实际应用场景下，我们的样本集数据量是有限的，SVM通过有限的样本集训练</p> | <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1.图4.4 分类器自适应学习子模块18. 4.4.1 标准支持向量机算法的不足支持向量机SVM算法思想的本质是在给定的训练样本集上构造一个高维度超平面来作为正负样本的分界面，并使他们之间的边界距离最大化。并且SVM通过使用核函数可以处理正负样本线性不可分的情况。所谓的支持向量就是那些非常接近最近分界超平面的数据点，它们在整个分类模型中起着决定性的作用。由于该算法在小样本集的分类学</p> <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1.值比较直方图4.4 分类器自适应学习子模块4.4.1 标准支持向量机算法的不足支持向量机 SVM 算法思想的本质是在给定的训练样本集上构造一个高维度超平面来作为正负样本的分界面，并使他们之间的边界距离最大化。并且 SVM通过使用核函数可以处理正负样本线性不可分的情况。所谓的支持向量就是那些非常接近最近分界超平面的数据点，它们在整个分类模型中起着决定性的作用。由于该算法在小样本集</p> <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)</p> <p>1.果均值比较直方图4.4分类器自适应学习子模块</p> <p>4.4.1标准支持向量机算法的不足标准支持向量机算法思想的本质是在给定的训练样本集上构造一个高维度超平面来作为正负样本的分界面，并使它们之间的边界距离最大化。并且SVM通过使用核函数可以处理正负样本线性不可分的情况。所谓的支持向量就是那些非常接近最优分界超平面的数据点，它们在整个分类模型中起着决定性的作用。由于该算法在有限</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1.4.4.1 标准支持向量机算法的不足标准支持向量机算法思想的本质是在给定的训练样本集上构造一个高维度超平面来作为正负样本的分界面，并使他们之间的边界距离最大化。并且 SVM通过使用核函数可以处理正负样本线性不可分的情况。所谓的支持向量就是那些非常接近最优分界超平面的数据点，它们在整个分类模型中起着决定性的作用。由于该算法在有限样本</p> |
| 4 | <p>此处有 38 字相似</p> <p>加权合并生成新的分类器。</p> <p>以上的思路都有一个共同的不足之处，就是需要重新用样本集训练，这在数据量很大时会造成系统不小的开销。Yang等人[30]借鉴规范损失最小化理论，提出了自适应支持向量机算法</p> <p>(Adaptive Support Vector Machine , ASVM)。ASVM无</p>   | <p>于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》- 2014-12-31 (是否引证：否)</p> <p>1.算法相比具有较好的适应性，较高的分类器训练效率和灵活性。20. 4.4.3 ASVM自适应支持向量机算法原理及时空开销分析Yang等人[46]在规范损失最小化原理的基础上提出了自适应学习的支持向量机算法ASVM，其可以在当前已经存在的分类器上直接进行新样本的再学</p>   |



|  |                  |  |
|--|------------------|--|
|  | 需更新原有的经验样本集，可以用新 | 习及分类决策函数的调整，进而生成适应新分类样本特征的分类器。ASVM算  |
|  |                  | 恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 )   |
|  |                  | 1.学习算法相比具有较好的适应性，较高的分类器训练效率和灵活性。4.4.3 ASVM 自适应支持向量机算法原理及时空 <b>开销分析Yang 等人[46]在规范损失最小化原理的基础上提出了自适应学习的支持向量机算法</b> ASVM，其可以在当前已经存在的分类器上直接进行新样本的再学习及分类决策函数的调整，进而生成适应新分类样本特征的分类       |
|  |                  | 王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )  |
|  |                  | 1.有较高的分类器训练效率和对新恶意网页样本具有较好的自适应性。4.4.3 ASVM 自适应支持向量机算法原理及时空 <b>开销分析Yang 等人[46]在规范损失最小化理论的基础上提出了自适应学习的支持向量机算法</b> ASVM，其可以在当前已经存在的分类器上直接进行新样本的再学习及分类决策函数的调整，进而生成适应新分类样本特征的分类       |
|  |                  | 基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )   |
|  |                  | 1.检测中具有较高的分类器训练效率和对新恶意网页样本具有较好的自适应性。4.4.3自适应支持向量机算法原理及时空 <b>开销分析Yang等人[46]在规范损失最小化理论的基础上提出了自适应学习的支持向量机算法 ( ASVM )</b> ，可以在当前已经存在的分类器上直接进行新样本的再学习及分类决策函数的调整，进而生成适应新分类样本特征的分类器。ASV |

| 指 标  |
|--|
| 疑似剽窃文字表述   |
| <div>1. 位数组的大小最好与插入元素的个数成线性关系，对于给定的m，n，k，假正例概率最大为：1-e-kn+0.5/m-1k。<br/>布隆过滤器</div> <div>2. 链接的木马、钓鱼、病毒、诈骗等链接。白名单中存储的是经过安全认证的正常站点的链接地址。</div> |

| 5. 11491915_刘力铭_第5部分  | 总字数：11780                |
|---|--------------------------|
| 相似文献列表 文字复制比：14.4%(1701) 疑似剽窃观点：(0)   |                          |
| 1 王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15<br>王维光 - 《学术论文联合比对库》 - 2015-01-15 | 12.3% ( 1452 )<br>是否引证：否 |
| 2 恶意网页医疗数据应急通信移动通信<br>- 《学术论文联合比对库》 - 2015-01-09                                | 11.5% ( 1359 )<br>是否引证：否 |
| 3 于俊_基于等级容灾模型<br>于俊 - 《学术论文联合比对库》 - 2014-12-31                                  | 11.2% ( 1320 )<br>是否引证：否 |
| 4 基于分类算法的恶意网页检测技术研究<br>王维光(导师：王枏) - 《北京邮电大学硕士论文》 - 2015-03-16                   | 10.4% ( 1220 )<br>是否引证：是 |
| 5 基于决策树的搜索引擎恶意网页检测研究与实现<br>周浩(导师：孙建华) - 《湖南大学硕士论文》 - 2013-05-27                 | 0.7% ( 88 )<br>是否引证：否    |

| 原文内容          | 相似内容来源   |
|---------------|--|
| 1 此处有 202 字相似 | 恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》 - 2015-01-09 ( 是否引证：否 ) |

|   |   |
|---|---|
| <p>ame不被浏览器渲染引擎渲染，但其通过src引入的外部脚本依旧会被解析引擎加载和解析执行。针对这种形式的暗链接，我们构建</p> <p>如下正则表达式：</p> <p><code>/[^\&amp;gt;!]*((width\s*?=\s*?0) (height\s*?=\s*?0) (visibility:\s*?hidden) (display:\s*?none) (style\s*?=\s*?"hidden") (top\s*?:\s*?-?[0-9]{3,}\s*?px;) (left\s*?:\s*?-?[0-9]{3,}\s*?px;))[^\&amp;gt;]*ig。</code></p> <p>( 2 )</p> <p>超长字符串：</p> <p>经过前文分析可以知道，恶意链接和正常的网页链接相比在长度上有很大的不同，常见的恶意链接往往具有很长的地址</p> | <p>1. 与 left 属性值为负数等情况，被隐藏的网页地址链接就是暗链。iframe/frame 框架隐藏的正则表达式形式如下</p> <p><code>: &amp;lt;[^\&amp;gt;!]*((width\s*?=\s*?0) (height\s*?=\s*?0) (visibility:\s*?hidden) (display:\s*?none) (style\s*?=\s*?"hidden") (top\s*?:\s*?-?[0-9]{3,}\s*?px;) (left\s*?:\s*?-?[0-9]{3,}\s*?px;))[^\&amp;gt;]*&amp;gt;; ( 2 ) Windows 操作系统 CLSID 唯一标识码：CLSID 是指 windows 操作系统对于不同的文件类型、应用程序、OLE 对</code></p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 ( 是否引证：否 )</p> <p>1. 与 left 属性值为负数等情况，被隐藏的网页地址链接就是暗链。iframe/frame 框架隐藏的正则表达式形式如下</p> <p><code>: &amp;lt;[^\&amp;gt;!]*((width\s*?=\s*?0) (height\s*?=\s*?0) (visibility:\s*?hidden) (display:\s*?none) (style\s*?=\s*?"hidden") (top\s*?:\s*?-?[0-9]{3,}\s*?px;) (left\s*?:\s*?-?[0-9]{3,}\s*?px;))[^\&amp;gt;]*&amp;gt;; ( 2 ) Windows 操作系统 CLSID 唯一标识码：CLSID 是指 windows 操作系统对于不同的文件类型、应用程序、OLE 对</code></p> <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.n”，top与left属性值为负数等情况，被隐藏的网页地址链接就是暗链。iframe/frame框架隐藏的正则表达式形式如下</p> <p><code>: &amp;lt;[^\&amp;gt;!]*((width\s*?=\s*?0) (height\s*?=\s*?0) (visibility:\s*?hidden) (display:\s*?none) (style\s*?=\s*?"hidden") (top\s*?:\s*?-?[0-9]{3,}\s*?px;) (left\s*?:\s*?-?[0-9]{3,}\s*?px;))[^\&amp;gt;]*&amp;gt;;2. Windows操作系统CLSID唯一标识码：CLSID是指windows操作系统对于不同的文件类型、应用程序、特殊</code></p> <p>基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》 - 2015-03-16 ( 是否引证：是 )</p> <p>1.top与left属性值为负数等情况，被隐藏的网页地址链接就是暗链。irfame/frame框架隐藏的正则表达式形式如下</p> <p><code>: &amp;lt;[^\&amp;gt;!]*((width\s*?=\s*?0) (height\s*?=\s*?0) (visibility:\s*?hidden) (display:\s*?none) (style\s*?=\s*?"hidden") (top\s*?:\s*?-?[0-9]{3,}\s*?px;) (left\s*?:\s*?-?[0-9]{3,}\s*?p(2) Windows操作系统CLSID唯一标识码：CLSID是指windows操作系统对于不同的文件类型、应用程序</code></p> |
| <p>2</p> <p>此处有 40 字相似</p> <p>&amp;gt;</p> <p>其中http://www.attacks.index/attack_index返回的是一个恶意页面，该页面中会内嵌</p> <p>恶意脚本对用户进行攻击，框架里面就是恶意网页的内容。由于document.wri</p> <p>te这一API频发使用会对浏览器的开销较大，因此该函数很少在正常网页的开发中使用，但是经常会被黑客利</p>   | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 ( 是否引证：否 )</p> <p>1.意代码，如果将以上代码保存为html格式，直接运行就会在浏览器中显示出一个框架，框架里面就是恶意网页的内容。在各式各样的恶意JS脚本攻击，大多数的恶意网页会使用document.write函数来执行恶意攻击脚本，并且该函数的出现次数多于正常网页，因此我们使用Document.write在网页源码中的出现</p>   |

|   |  |  |
|---|--|--|
|   | 用来进行攻击，因此可   |  |
| 3 | <p>此处有 840 字相似</p> <p>性比较</p> <p>MSHASH 柔性比较两两比较耗时多</p> <p>MSHASH-BF 柔性比较，相对耗时较少结果依赖分桶粒度</p> <p>5.3</p> <p>分类结果评价标准</p> <p>由于恶意网页类型识别是一个二分类问题，即通过分类器将未知类型的网页分为恶意网页和正常网页，因此我们使用二值分类评价标准对分类结果进行评价。本文主要目的是检测未知网页样本集并发现其中包含的恶意网页，所以我们将恶意网页作为正例，正常网页作为负例，则恶意网页分类模型分类结果的列联表如表5-4所示：</p> <p>表 54 恶意网页检测分类结果列联表</p> <p>实际是恶意网页实际是正常网页</p> <p>被判定为恶意网页 TP FP</p> <p>被判定为正常网页 FN TN</p> <p>设样本测试集中共有N个样本，即<math>N = TP + FN + FP + TN</math>，其中各变量的含义如下所示：</p> <p>TP：True Positive，表示被模型预测为真的正样本，可以称作判断为真的正确率；</p> <p>TN：True Negative，表示被模型预测为真的负样本，可以称作判断为真的正确率；</p> <p>FP：False Positive，表示被模型预测为真的负样本，可以称作误报率；</p> <p>FN：False Negative，表示被模型预测为真的正样本，可以称作漏报率。</p> <p>则有如下几个评价标准：</p> <p>（1）真正率（True Positive Rate）TPR：<math>TPR = TP / (TP + FN)</math>，该值主要是说明分类模型检测出的恶意网页占恶意网页实际总量的比例，反应了分类器识别恶意网页的能力。</p> <p>（2）假正率（False Positive Rate）FPR：<math>FPR = FP / (FP + TN)</math>，主要说明正常网页被判定为恶意网页的比率，反应了分类器的误判能力。</p> <p>（3）召回率R：<math>R = TP / (TP + FN) \times 100\%</math>，含义同真正率。</p> <p>（4）准确率A：<math>A = TP / (TP + FP) \times 100\%</math>，主要说明分类模型正确分辨的恶意网页和正常网页数占总样本量的比例，反应分类器对所有样本类别的整体分类能力。</p> <p>（5）精确率P：<math>P = TP / (TP + FP) \times 100\%</math>，主要说明分类模型在判定为恶意网TP+FP页的样本量中实际为真正恶意网页所占的比例，反应分类模型对恶意网页的查准率。</p> <p>5.4</p> <p>测试结果分析</p> <p>5.4.1 基于黑/白名单的匹配结果</p> <p>前文可知，本次试验中恶意链接和正常样本分别为6012和5337，</p> | <p>于俊_基于等级容灾模型 于俊 - 《学术论文联合比对库》 - 2014-12-31 (是否引证：否)</p> <p>1.增加算法复杂度的情况下将低维空间的非线性分类问题转换为高维空间的线性分类问题，使得算法具有较大的应用范围。3.5 分类结果评价标准由于恶意网页类型识别是一个二分类问题，即通过分类器将未知类型的网页分为恶意网页和正常网页，因此我们使用二值分类评价标准对分类结果进行评价。本文主要目的是检测未知网页样本集并发现其中包含的恶意网页，所以我们将恶意网页作为正例，正常网页作为负例，则恶意网页分类模型分类结果的列联表如表3-1所示：表</p> <p>2.页，因此我们使用二值分类评价标准对分类结果进行评价。本文主要目的是检测未知网页样本集并发现其中包含的恶意网页，所以我们将恶意网页作为正例，正常网页作为负例，则恶意网页分类模型分类结果的列联表如表3-1所示：表3-1 恶意网页分类器列联表 实际是恶意网页实际是正常网页被判定为恶意网页 TP FP被判定为正常网页 FN TN设样本测试集中共有个样本，即，其中各变量的含义如下所示：TP: True Positive, 表示被模型预测为真的正样本；可以称作判断为真的正确率；TN: True Negative，表示被模型预测为真的负样本；可以称作判断为真的正确率；FP: False Positive，表示被模型预测为真的负样本；可以称作误报率；FN: False Negative，表示被模型预测为真的正样本。可以称作漏报率则有如下几个评价标准：（1）真正率(True Positive Rate):，该值主要是说明分类模型检测出的恶意网页占恶意网页实际总量的比例，反应了分类器识别恶意网页的能力。（2）假正率(False Positive Rate):，主要说明正常网页被判定为恶意网页的比率，反应了分类器的误判能力。（3）召回率:，含义同真正率。（4）准确率:，主要说明分类模型正确分辨的恶意网页和正常网页数占总样本量的比例，反应分类器对所有样本类别的整体分类能力。</p> <p>3.网页被判定为恶意网页的比率，反应了分类器的误判能力。（3）召回率:，含义同真正率。（4）准确率:，主要说明分类模型正确分辨的恶意网页和正常网页数占总样本量的比例，反应分类器对所有样本类别的整体分类能力。（5）精确率:，主要说明分类模型在判定为恶意网页的样本量中实际为真正恶意网页所占的比例，反应分类模型对恶意网页的查准率。3.6 本章小结本章简单介绍了本文用于恶意网页检测分类学习的四种分类学习算法，主要包括朴素贝叶斯算法Navie Bayes</p> <p>王维光_2012140293 基于分类算法的恶意网页检测技术研究 2015-01-15 王维光 - 《学术论文联合比对库》 - 2015-01-15 (是否引证：否)</p> <p>1.线性分类问题转换为高维空间的线性分类问题，使得算法具有较大的应用范围。第三章分类学习算法概述 213.5 分类结果评价标准由于恶意网页类型识别是一个</p> |

二分类问题，即通过分类器将未知类型的网页分为恶意网页和正常网页，因此我们使用二值分类评价标准对分类结果进行评价。本文主要目的是检测未知网页样本集并发现其中包含的恶意网页，所以我们将恶意网页作为正例，正常网页作为负例，则恶意网页分类模型分类结果的列联表如表 3-1 所示：表 3-1 恶意网页检测分类结果列联表实际是恶意网页实际是正常网页被判定为恶意网页 TP FP被判定为正常网页 FN TN设样本测试集中共有 N 个样本，即 N TP FN FP TN，其中各变量的含义如下所示：TP: True Positive, 表示被模型预测为正的样本，可以称作判断为真的正确率；TN: True Negative，表示被模型预测为负的负样本，可以称作判断为假的正确率；FP: False Positive，表示被模型预测为正的负样本，可以称作误报率；FN: False Negative，表示被模型预测为负的正样本，可以称作漏报率。则有如下几个评价标准：（1）真正率(True Positive Rate)TPR:TPFNTPTPR，该值主要是说明分类模型检测出的恶意网页占恶意网页实际总量的比例，反应了分类器识别恶意网页的能力。（2）假正率(False Positive Rate)FPR:FPTNFFTPR，主要说明正常网页被判定为恶意网页的比率，反应了分类器的误判能力。（3）召回率R：100%TPFNTPTPR，含义同真正率。（4）准确率 A: 100%NTPTNA，主要说明分类模型正确分辨的恶意网页和正常网页数占总样本量的比例，反应分类器对所有样本类别的整体分类能力。（5）精确率P：100%TPFFTPPP，主要说明分类模型在判定为恶意网页的样本量中实际为真正恶意网页所占

2.类器对所有样本类别的整体分类能力。（5）精确率P：100%TPFFTPPP，主要说明分类模型在判定为恶意网页的样本量中实际为真正恶意网页所占的比例，反应分类模型对恶意网页的查准率。上述四种分类算法为当前使用最广泛的分类学习算法，并且每个算法具有各自不同的适应场景，本文针对恶意网页分类问题，通过在

恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 (是否引证：否)

1.法复杂度的情况下将低维空间的非线性分类问题转换为高维空间的线性分类问题，使得算法具有较大的应用范围。3.5 分类结果评价标准由于恶意网页类型识别是一个二分类问题，即通过分类器将未知类型的网页分为恶意网页和正常网页，因此我们使用二值分类评价标准对分类结果进行评价。本文主要目的是检测未知网页样本集并发现其中包含的恶意网页，所以我们将恶意网页作为正例，正常网页作为负例，则恶意网页分类模型分类结果的列联表如表 3-1 所示：表 3-1 恶意网页检测分类结果列联表实际是恶意网页实际是正常网页被判定为恶意网页 TP FP被判定为正常网页 FN TN设样本测试集中共有 N 个样本，即 N TP FN FP TN，其中各变量的含义如下所示：TP: True Positive, 表示被模型预测为正的样本；可以称作判断为真的正确率；TN: True Negative，表示被模型预测为负的负样本；可以称作判断为假的正确率；FP: False Positive，表示被模型预测



为正的负样本；可以称作误报率；FN: False Negative，表示被模型预测为负的正样本。可以称作漏报率则有如下几个评价标准：（1）真正率(True Positive Rate)TPR:TPFNTPTPR，该值主要是说明分类模型检测出的恶意网页占恶意网页实际总量的比例，反应了分类器识别恶意网页的能力。（2）假正率(False Positive Rate)FPR:FPTNFPTPR，主要说明正常网页被判定为恶意网页的比率，反应了分类器的误判能力。（3）召回率R

2.实际总量的比例，反应了分类器识别恶意网页的能力。（7）假正率(False Positive Rate)FPR：，主要说明正常网页被判定为恶意网页的比率，反应了分类器的误判能力。（8）召回率R：，含义同真正率。（9）准确率 A：，主要说明分类模型正确分辨的恶意网页和正常网页数占总样本量的比例，反

3.页被判定为恶意网页的比率，反应了分类器的误判能力。（3）召回率R：100%TPFNTPR，含义同真正率。（4）准确率 A：100%NTPTNA，主要说明分类模型正确分辨的恶意网页和正常网页数占总样本量的比例，反应分类器对所有样本类别的整体分类能力。（5）精确率P：100%TPFPTPP，主要说明分类模型在判定为恶意网页的样本量中实际为真正恶意网页所占的比例，反应分类模型对恶意网页的查准率。上述四种分类算法为当前使用最广泛的分类学习算法，并且每个算法具有各自不同的适应场景，本文针对恶意网页分类问题，通过在

基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》- 2015-03-16 (是否引证：是)

1.换为高维空间的线性分类问题，使得算法具有较20第三章分类学习算法概述 大的应用范围。3.5分类结果评价标准由于恶意网页类型识别是一个二分类问题，即通过分类器将未知类型的网页分为恶意网页和正常网页，因此我们使用二值分类评价标准对分类结果进行评价。本文主要目的是检测未知网页样本集并发现其中包含的恶意网页，所以我们将恶意网页作为正例，正常网页作为负例，则恶意网页

2.页和正常网页，因此我们使用二值分类评价标准对分类结果进行评价。本文主要目的是检测未知网页样本集并发现其中包含的恶意网页，所以我们将恶意网页作为正例，正常网页作为负例，则恶意网页分类模型分类结果的列联表如表3-1所示：表3-1恶意网页检测分类结果列联表实际是恶意网页|实际是正常网页被判定为恶意网页 TTT ^被判定为正常网页| FN | TN设样本测试集中共有AT个样本，即 $N = TP + m + FP + IN$ ，其中各变量的含义如下所示：TP: True Positive,表示被模型预测为正确的正样本，可以称作判断为真的正确率；TN: True Negative,表示被模型预测为负的负样本，可以称作判断为假的正确率；FP: False Positive,表示被模型预测为正的负样本，可以称作误报率；FN: False Negative,表示

|   |  |  |
|---|--|--|
|   |  | <p>被模型预测为负的正样本，可以称作漏报率。则有如下几个评价标准：真TP(1)正率(True Positive Rate)7P_R: TPR = ~~, 该值主要是</p> <p>3.Negative,表示被模型预测为负的正样本，可以称作漏报率。则有如下几个评价标准：真TP(1)正率(True Positive Rate)7P_R: TPR = ~~, 该值主要是说明分TP+FN类模型检测出的恶意网页占恶意网页实际总量的比例，反应了分类器识别恶意网页的能力。PP(2)假正率(False Posiitive Rate)TPR = ,主要说明正常网FP + TN页被判定为恶意网页的比率，反应了分类器的误判能力。率TP(3)</p> <p>4.常网FP + TN页被判定为恶意网页的比率，反应了分类器的误判能力。率TP(3)召回；? : R=含义同真正率。(r4)准确率j = ?气100%,主要说明分类模型正确分辨的恶意N网页和正常网页数占总样本量的比例，反应分类器对所有样本类别的整体分类能力。(5)精确率尸：P = - ^X100%,主要说明分类模型在判定为恶意网TP+FP页的样本量中实际为真正恶意网页所占的比例，反应分类模型对恶意网页的查准率。21北京邮电大学工程硕士学位论文 上述四种分类算法为当前使用最广泛的分类学习算法，并且每个算法具有各</p>  |
| 4 | <p>此处有 29 字相似</p> <p>次随机取其中的9份集合黑名单库进行分类模型训练，取剩下的一份用来进行测试，将恶意链接标识为1，正常连接标识为0，分别验证</p> <p>每次实验的准确性和召回率，并取十次结果的平均期望作为最终的</p> <p>预测精度。</p> <p>根据表5-6的分类标准以及F-Measure，本实验从五个角度来测试和验证朴素贝叶斯、C4.5、分类回归树</p>   | <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士学位论文》- 2015-03-16 (是否引证：是)</p> <p>1.叉验证的方法进行实验，即将样本集平均分为10份，每次轮流以其中的9份作为训练样本集，剩下的一份作为测试样本集，计算每次实验的准确率和召回率。把10次实验结果的平均值作为算法精度的最终估计值。表4-7分别从真正率 ( TPR)、假正率 ( FPR)、精确率 ( Precision)、召回率(Reca</p>   |
| 5 | <p>此处有 250 字相似</p> <p>所示：</p> <p>表 57 四种分类算法结果比较</p> <p>分类算法判定 TPR FPR P R F-Measure</p> <p>朴素贝叶斯 1</p> <p>0.612 0.14 0.831 0.62 0.693</p> <p>0 0.86 0.388 0.7120.84 0.79</p> <p>C4.5 1 0.81 0.13 0.85 0.822 0.833</p> <p>0 0.87 0.19 0.848 0.87 0.851</p> <p>CART 1 0.881 0.09 0.932 0.876 0.909</p> <p>0 0.91 0.119 0.881 0.93 0.92</p> <p>SVM 1 0.898 0.04 0.95 0.90 0.932</p> <p>0 0.96 0.102 0.913 0.</p> <p>955 0.939</p> <p>由上表可以看出：</p> <p>( 1 ) SVM的TPR是0.898，使所有的算法中最高的，朴素贝叶斯算法的TP</p> | <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》- 2015-01-09 (是否引证：否)</p> <p>1. TPR FPR Precision Recall F-MeasureNaive Bayesyes 0.603 0.11 0.832 0.603 0.699no 0.89 0.397 0.712 0.89 0.791CARTyes 0.82 0.128 0.853 0.82 0.836no 0.872 0.18 0.842 0.872 0.857C4.5yes 0.879 0.06 0.935 0.879 0.906no 0.94 0.121 0.887 0.94 0.913SVMyes 0.897 0.047 0.949 0.897 0.922no 0.953 0.103 0</p> <p>王维光 2012140293 基于分类算法的恶意网页检测技术研究 2015-01-15 王维光 -《学术论文联合比对库》- 2015-01-15 (是否引证：否)</p> <p>1. TPR FPR Precision Recall F-MeasureNaive Bayesyes 0.603 0.11 0.832 0.603 0.699no 0.89 0.397 0.712 0.89 0.791CARTyes 0.82 0.128 0.853 0.82 0.836no 0.872 0.18 0.842 0.872 0.857C4.5yes 0.879 0.06 0.935 0.879 0.906no 0.94 0.121 0.887 0.94 0.913SVMyes 0.897 0.047 0.949 0.897 0.922no 0.953</p> |

|   |   |   |
|---|---|---|
|   |   | 0.103 0   |
|   |   | 于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》 - 2014-12-31 (是否引证:否)   |
|   |   | 1. 0.82 0.836no0.872 0.18 0.8420.872 0.857C4.5<br>yes0.879 0.06 0.9350.879 0.906no 0.940.121 0.887<br>0.94 0.913SVM yes0.8970.047 0.9490.897<br>0.922no0.9530.103 0.9030.953 0.927表4-3-3-1分别从<br>真正率TPR、假正率FPR、精确率Precision、召回率<br>Recall和F-Meas  |
| 6 | <p>此处有 56 字相似</p> <p>模型。</p> <p>Step 3: 利用这9个分类器分别对测试样本集T进行测试, 并求所有的测试结果的均值作为最终的分类结果。</p> <p>以上</p> <p>实验, 9个分类器的分类结果如表5-9所示:</p> <p>表 59利用ASVM进行自适应学习的9个分类器分类结果</p> <p>分类器</p> <p>f(a)1 f(a)2 f(a)3 f(a)3 f(a)4 f(a)5 f(a)6 f(a)7 f(a)</p> | <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》 - 2015-01-09 (是否引证:否)</p> <p>1.1、C2、C3、C4; Step 3: 使用更新的分类器 C1、C2、C3、C4 对测试样本集 S 进行分类预测。实验结果如表 5-1 所示: 表 5-1 系统当前分类器与自适应学习分类器分类准确率分类器 C0 C1 C2 C3 C4准确率<br/>84.53% 89.63% 90.31% 90.20% 89.87%从表 5</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》 - 2015-01-15 (是否引证:否)</p> <p>1.ep 3: 使用更新的分类器 C1、C2、C3、C4 对测试样本集 S 进行分类预测。第五章实验结果分析46实验结果如表 5-1 所示: 表 5-1 系统当前分类器与自适应学习分类器分类准确率分类器 C0 C1 C2 C3 C4准确率<br/>84.53% 89.63% 90.31% 90.20% 89.87%从表 5</p>   |
| 7 | <p>此处有 47 字相似</p> <p>证。算法依旧存在可以优化的细节。</p> <p>6 结语</p> <p>6.1 本文的研究工作</p> <p>本文通过对当前互联网环境下存在的恶意链接泛滥的问题进行介绍, 并分析了目前国内外专家和网络安全工程师在恶意链接检测领域各自研究成果和应用方案的优势和缺陷。针对这些缺陷, 本文在研究了恶意链接的主要攻击形式和核心技术后, 介绍了JavaScript脚本的发展、运行原理</p>                       | <p>基于分类算法的恶意网页检测技术研究 王维光 -《北京邮电大学硕士论文》 - 2015-03-16 (是否引证:是)</p> <p>1.检测技术的基础上, 发现当前基于特征匹配的恶意网页检测技术虽然可以快速准确地检测出已知类型的恶意网页, 但对新出现的恶意网页却很难有效识别的问题。针对这一问题, 本文分析和比较了现阶段国内外专家学者在恶意网页检测领域研究成果的优点和不足。本文在调研有关恶意网页攻击研究成果的基础上, 总</p> <p>恶意网页医疗数据应急通信移动通信 -《学术论文联合比对库》 - 2015-01-09 (是否引证:否)</p> <p>1.效进行恶意网页检测的方法措施变得尤为重要, 并通过列举实际数据和攻击实例说明本课题的研究背景和意义。通过调研现阶段国内外专家学者在恶意网页检测领域的研究成果和技术进展, 总结了当前恶意网页检测技术的优点和不足, 并提出了本文的研究目的和方向。第二章阐述了网页恶意代码的概念及</p> <p>王维光_2012140293_基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 -《学术论文联合比对库》 - 2015-01-15 (是否引证:否)</p> <p>1.页检测, 尤其是对新出现的恶意网页样本能够及时识别和检测, 是当前 web 安全检测领域研究的重要内容。通过调研国内外专家学者在恶意网页检测领域的研究动态和主要技术手段, 分析现有的基于特征匹配的恶意网页检测技术虽然可以快速准确地检测出已知类型的恶意网页, 但存在对混淆变</p> |
| 8 | <p>此处有 35 字相似</p>   | 于俊_基于等级容灾模型 于俊 -《学术论文联合比对库》 -   |

|   |   |   |
|---|---|---|
|   | <p>要有以下几点:</p> <p>(1) 系统调研和学习了恶意链接的由来、常见的攻击形式, 分析了其在当前互联网背景下的造成问题的严重性。并</p> <p>在学习和借鉴了国内外专家和网络安全工作者对恶意链接检测做出的研究成果,</p> <p>在此基础上针对其存在的不足, 提出了利用黑/白名单库过滤、自适应支持向量机分类学习算法和DOM结构改变监测算法进行分类预测</p>               | 2014-12-31 (是否引证: 否)  |
|   |   | 1.为的资料背景, 通过具体恶意脚本攻击案例和网络安全权威组织机构的调查统计数据分析目前的web安全面临的主要威胁和安全形势。 <b>在学习和研究了现阶段国内外专家学者在网络安全检测方面的研究动态和学术成果</b> 的基础上, 分析现有恶意网页检测技术的不足, 提出了本文的研究目的和意义。(2) 阐述了网页恶意代码攻击的 |
|   |   | 恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 (是否引证: 否)  |
|   |   | 1.点总结如下: (1) 系统调研了当前典型的恶意网页攻击的资料背景, 分析了当前 web 安全面临的主要威胁和安全形势。 <b>在学习和研究了现阶段国内外专家学者在网络安全检测方面的研究动态和学术成果</b> 的基础上, 分析现有恶意网页检测技术的不足, 提出了在基于知识库的恶意网页特征匹配基础上, 对检测后未能判断  |
| 9 | <p>此处有 33 字相似</p> <p>介绍了各个模块的实现原理。</p> <p>6.2 本文存在的不足和未来的改进</p> <p>本文提出的恶意链接检测系统(SD-MLDS)经过实验验证, 可以通过自适应向量机对新出现的恶意链接形式进行增量特征学习,</p> <p>并利用DSCD算法提取出页面隐藏的暗链接, 提高了对恶意链接的识别率和准确度。但是, 本论文依旧存在一些问题以及可以改进的地</p> | 王维光 2012140293 基于分类算法的恶意网页检测技术研究_2015-01-15 王维光 - 《学术论文联合比对库》- 2015-01-15 (是否引证: 否)   |
|   |   | 1.点总结如下: (1) 系统调研了当前典型的恶意网页攻击的资料背景, 分析了当前 web 安全面临的主要威胁和安全形势。 <b>在学习和研究了现阶段国内外专家学者在网络安全检测方面的研究动态和学术成果</b> 的基础上, 分析现有恶意网页检测技术的不足, 提出了在基于知识库的恶意网页特征匹配基础上, 对检测后未能判断  |
|   |   | 恶意网页医疗数据应急通信移动通信 - 《学术论文联合比对库》- 2015-01-09 (是否引证: 否)  |
|   |   | 1.攻击。图 5-4 CA-MWDS 系统基于知识库特征匹配检测结果5.3.2 自适应支持向量机算法分类效果为了验证 <b>自适应支持向量机模块对新恶意网页样本特征</b> 分布的自适应学习效果, 我们从专业安全检测网站上新采集了 500 个恶意网页样本, 并将其分为 5组, 每组各 100                |
|   |   | 基于分类算法的恶意网页检测技术研究 王维光 - 《北京邮电大学硕士论文》- 2015-03-16 (是否引证: 是)  |
|   |   | 1.*rite docinent.write(...) 12图5-4CA-MWDS系统基于知识库特征匹配检测结果 <b>5.3.2自适应支持向量机算法分类效果</b> 为了验证自适应支持向量机模块对新恶意网页样本特征分布的自适应学习效果, 从专业安全检测网站上新采集了 500个恶意网                    |

|   |  |         |
|---|--|---------|
| 指 标   |  |         |
| 疑似剽窃文字表述  |  |         |
| 1. 进行特征提取。实际中许多恶意网页都无法访问或拥有相同的域名, 最后提炼的恶意网页特征数为 2100 个。 |  |         |
| 2. 实验, 9个分类器的分类结果如表5-9所示:                               |  |         |
| 表 59利用ASVM进行自适应学习的9个分类器分类结果                             |  |         |
| 分类器   |  |         |
| 表格检测详细结果  |  |         |
| 原文表格1: 表 5-4 恶意网页检测分类结果列列表 共有1个相似表格                     |  |         |
| 实际是恶意网页   |  | 实际是正常网页 |



|          |    |    |
|----------|----|----|
| 被判定为恶意网页 | TP | FP |
| 被判定为正常网页 | FN | TN |

相似表格1：表3-1 恶意网页分类器列联表

相似度：100.00%

来源：于俊\_基于等级容灾模型-于俊-《》-2014-12-31

|          |         |         |
|----------|---------|---------|
|          | 实际是恶意网页 | 实际是正常网页 |
| 被判定为恶意网页 | TP      | FP      |
| 被判定为正常网页 | FN      | TN      |

说明：1.总文字复制比：被检测论文总重合字数在总字数中所占的比例

2.去除引用文献复制比：去除系统识别为引用的文献后，计算出来的重合字数在总字数中所占的比例

3.去除本人已发表文献复制比：去除作者本人已发表文献后，计算出来的重合字数在总字数中所占的比例

4.单篇最大文字复制比：被检测文献与所有相似文献比对后，重合字数占总字数的比例最大的那一篇文献的文字复制比

5.指标是由系统根据《学术论文不端行为的界定标准》自动生成的

6.红色文字表示文字复制部分;绿色文字表示引用部分

7.本报告单仅对您所选择比对资源范围内检测结果负责



 [amlc@cnki.net](mailto:amlc@cnki.net)

 <http://check.cnki.net/>

 <http://e.weibo.com/u/3194559873>