(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105426762 A
(43)申请公布日 2016.03.23

(21)申请号 201510999378.1

(22)申请日 2015.12.28

(71)申请人 重庆邮电大学
地址 400065 重庆市南岸区黄桷垭崇文路2号

(72)发明人 尚凤军 邓小林

(74)专利代理机构 北京同恒源知识产权代理有限公司 11275
代理人 廖曦

(51)Int.Cl.
G06F 21/56(2013.01)

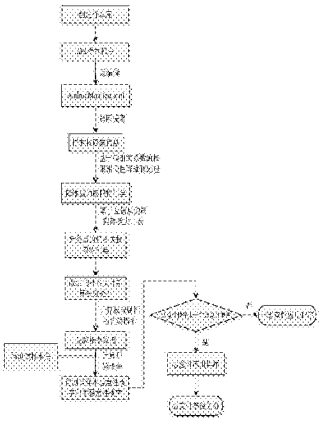
权利要求书2页 说明书7页 附图3页

(54)发明名称

一种android应用程序恶意性的静态检测方法

(57)摘要

本发明涉及一种android应用程序恶意性的静态检测方法，属于Android平台下应用程序安全性检测技术领域。该方法首先通过计算偏相关系数对Android应用程序权限特征属性进行相关性分析，达到对权限特征集进行降维预处理的目的；其次利用互信息和笛卡尔积方法，对降维后的权限特征集进行相关性聚类去冗余，并设定阈值，避免过拟合的现象，以此得到新的分类权限特征集的集合X_{new}，达到权限聚类后的权限特征集之间几乎是相互独立关系的目的；最后，在权限聚类后的基础上，构建朴素贝叶斯分类器，并对其进行改进，达到能使应用程序分类决策相关性高，进而提高Android应用程序恶意性检测的可靠性。



CN 105426762 A

1.一种android应用程序恶意性的静态检测方法，其特征在于：在该方法中，对选定的样本程序进行反编译得到AndroidManifest.xml文件，提取该文件的权限特征，并对其进行降维预处理，然后对降维后的权限特征集用互信息和笛卡尔积方法进行权限聚类去冗余，最后在此基础上构建朴素贝叶斯分类模型，以及对所检测到的恶意性应用程序进行恶意性等级的划分。

2.根据权利要求1所述的一种android应用程序恶意性的静态检测方法，其特征在于：该方法具体包括以下步骤：

步骤一：收取并创建恶意性应用程序和非恶意性应用程序的样本库，分别对其APK样本进行反编译处理得到AndroidManifest.xml文件，然后提取该文件的权限特征，获得权限特征集；

步骤二：利用Android权限特征属性变量之间的相关性关系，其中任意两个变量之间的相关性可能是由于第三个变量的存在所表现出来的，对此采用基于偏相关系数对权限特征属性进行相关性分析的方法，对权限特征集进行降维预处理；

步骤三：利用基于互信息理论和笛卡尔积方法，采用基于互信息和笛卡尔积的改进的朴素贝叶斯分类模型方法，对权限特征集降维预处理后获取的权限特征集进行聚类去冗余；

步骤四：基于分类属性集的集合Xnew构建朴素贝叶斯分类器，通过样本训练获得先验概率，然后用测试集样本通过计算后验概率判断所检测的Android应用程序是否具有恶意性，对具有恶意性的Android应用程序按概率方法进行等级划分。

3.根据权利要求2所述的一种android应用程序恶意性的静态检测方法，其特征在于：在步骤二中，所述基于偏相关系数对权限特征属性进行相关性分析的方法具体包括：

该方法首先通过计算两个权限特征属性变量之间的简单相关系数$r(x_i,x_j)=\frac{Cov(x_i,x_j)}{\sqrt{D(x_i)D(x_j)}}$，其中$Cov(x_i,x_j)$是$x_i$与$x_j$之间的协方差，$\sqrt{D(x_i)D(x_j)}$是$x_i$与$x_j$之间的标准差，将计算所得的简单相关系数做成相关系数矩阵R，计算|R|行列式中r_{ii},r_{ij},r_{jj}的数余子式A^{ii},A^{ij},A^{jj}然后带入特征权限属性变量之间的偏相关系数$\rho(x_i,x_j \mid x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_{j-1},x_{j+1},\cdots,x_n)=\frac{-A^{ij}}{\sqrt{A^{ii}A^{jj}}}$公式进行计算，根据得到的偏相关系数|ρ|的值判断权限特征属性之间的相关性大小，去除相关性低的权限特征属性，得到降维预处理后的权限特征集。

4.根据权利要求2所述的一种android应用程序恶意性的静态检测方法，其特征在于：在步骤三中，利用基于互信息理论和笛卡尔积方法，采用基于互信息和笛卡尔积的改进的朴素贝叶斯分类模型方法，对权限特征集降维预处理后获取的权限特征集进行聚类去冗余，聚类去冗余模型如下：

$$Cor(X_i,C)=\sum_{X_i,c}P(X_i,C)\log\frac{P(X_i,C)}{P(X_i)P(C)}$$

$$Cor(X_i,X_j)=\sum_{X_i,c}P(X_i,X_j)\log\frac{P(X_i,X_j)}{P(X_i)P(X_j)}$$

其中$Cor(X_i,C)$表示权限特征属性变量X_i和类别属性变量c之间的相关度，$Cor(X_i,X_j)$表

示权限特征属性变量X_i和X_j之间的相关度，计算方式如下：

1)计算预处理后权限特征属性变量X_i与类别变量C的相关度$Cor(X_i,C)$，按从大到小的顺序排列构成原始属性集X-ori；

2)计算X-ori中的第一个属性变量X-ori(1)与其它属性变量的相关度Cor(X-ori(1)，X_j)；

3)对X-ori中除X-ori(1)之外的其它变量X_j，若$Cor(X\text{-}ori(1),X_j)>Cor(X_j,C)$，则认为该变量与X-ori(1)高度相关，将其加入X-ori(1)的相关集中；

4)X-ori(1)及其相关集中的前m个变量的笛卡尔积X_{new}作为新属性集加入X_{new}，同时从X-ori中删除X-ori(1)及其相关集中的所有变量；

5)重复2)至4)，直到X-ori=ø为止。

5.根据权利要求2所述的一种android应用程序恶意性的静态检测方法，其特征在于：在步骤四中，基于分类属性集的集合X_{new}构建朴素贝叶斯分类器，通过样本训练获得先验概率，然后用测试集样本通过计算后验概率判断所检测的Android应用程序是否具有恶意性，基于权限分类属性集的集合X_{new}和类别C构建朴素贝叶斯的模型如下：

$$P(C_i|X_{new})=\frac{P(X_{new}|C_i)P(C_i)^{\alpha}}{P(X_{new})}$$

其中，$\alpha=1+\frac{count(X_k|C_i)}{count(X_k)count(X)}$，$count(X_k|C_i)$表示在类别$C_i$样本中权限特征属性$X_k$出现的次数，$count(X_k)$表示样本中权限特征属性$X_k$出现的次数，count(X)表示分类权限集集合$X_{new}$中权限特征集的个数，α表示不同权限特征属性对分类的影响程度，且量化了权限特征属性与其类别属性之间的关系，X_{new}为Android应用程序的权限特征属性集的集合，C_i是Android应用程序的类别，即非恶意性应用程序和恶意性应用程序两类，$P(X_{new})$对于所有类为常数，因此比较后验概率是只需要$P(X_{new}|C_i)P(C_i)^{\alpha}$最大即可判断应用程序是否具有恶意性；

对所得的具有恶意性的Android应用程序的权限特征集，对恶意性Android应用程序进行恶意性等级划分，计算恶意性等级如下：

$$T=\sum\frac{P_v}{P_m}$$

$$P_v=\prod_{i=1}^{n}P_v(X_i)$$

$$P_m=\prod_{i=1}^{n}P_n(X_i)$$

其中，P_v表示该待测试样本应用程序在恶意性程序中出现的概率；P_m表示该待测试样本应用程序在非恶意性程序中出现的概率；$P_v(X_i)$代表第i个权限特征集在恶意性程序中出现的概率；$P_n(X_i)$代表第i个权限特征集在非恶意性程序中出现的概率。

一种android应用程序恶意性的静态检测方法

技术领域

[0001] 本发明属于Android平台下应用程序安全性检测技术领域，涉及一种android应用程序恶意性的静态检测方法。

背景技术

[0002] 现代快速的生活工作节奏，使得人们对能够从网络上获取实时的信息和服务有了更高的要求，移动互联网应运而生。移动互联网的安全问题直接影响到用户使用和对移动互联网的信任，更关系到移动互联网产能的释放以及正面价值的正常发挥，更涉及到我们国家以及整个民族信息的安全产业。所以在这个信息化时代我们要时刻关注着移动互联网安全的新特性，了解最详细的移动互联网安全动态，时刻掌握和及时处理因移动互联网安全问题引发的一系列矛盾。要不断的改进和完善移动互联网安全的整体架构以及为防止安全问题出现所做的部署，分析和监控移动互联网时刻出现的流量恶意攻击，散播的不健康不科学的信息。经过不断的技术革新、安全设计改进、移动互联网安全部署等措施来确保移动互联网的安全，并委派专人实时监控，以及采用内容信息过滤等技术手段，来保证移动互联网的安全，确保给移动互联网带来一个干净健康的发展环境。借助于移动互联网的发展，现在用手机就可以做到以前必须使用电脑才可做到的事情，使得人们对智能手机需求有了极大的提升。移动互联网安全中Android的安全通信问题也越来越受关注，2007年11月，Google发布了基于Linux内核的开源智能移动操作系统Android。该系统拥有庞大的用户数量和应用市场:来自Gartner统计数据显示，2013年第3季度全球智能手机的销售量为2.5亿多台，其中Android系统占据了81.9％；而截止2014年1月8日仅Android官方应用市场Google Play上的应用数量就达到了103万。

[0003] 有数据显示，在2011年人们使用智能手机的比例还比较低，到2012年使用智能手机的比例就达到了46％。根据HIS统计的信息显示，预计在2013年智能手机在市场中所占的份额将达到55％，这些数据表明智能手机在正在改变着人们日常生活方式，成为很多人生活和工作的得力助手。

[0004] 智能手机功能不断的完善和发展，为人们日常生活带来很多的便利，但同时也成为各种手机病毒及恶意软件攻击的主要目标。智能手机的快速发展，针对智能手机的病毒也以大比例的数量增长。第一个智能手机病毒Cabir诞生于NOKIA大本营，经过短短几年的发展，针对智能终端的病毒便出现了上千种。当前主流的智能手机操作系统有:Symbian OS，苹果的iOS，微软的Windows phone，Google的Android。每个系统都有一套自身的安全防范措施，由于人们对手机隐私信息安全的重视，分析已有的智能手机操作系统安全规范，提高智能手机系统防范病毒行为成为了研究的重点。

发明内容

[0005] 有鉴于此，本发明的目的在于提供一种android应用程序恶意性的静态检测方法，该方法首先通过计算偏相关系数对Android应用程序权限特征属性进行相关性分析，达到

对权限特征集进行降维预处理的目的；其次利用互信息和笛卡尔积方法，对降维后的权限特征集进行相关性聚类去冗余，并设定阈值，避免过拟合的现象，以此得到新的分类权限特征集的集合X_{new}，达到权限聚类后的权限特征集之间几乎是相互独立关系的目的；最后，在权限聚类后的基础上，构建朴素贝叶斯分类器，并对其进行改进，达到能使应用程序分类决策相关性高，进而提高Android应用程序恶意性检测的可靠性。

[0006] 为达到上述目的，本发明提供如下技术方案：

[0007] 一种android应用程序恶意性的静态检测方法，在该方法中，对选定的样本程序进行反编译得到AndroidManifest.xml文件，提取该文件的权限特征，并对其进行降维预处理，然后对降维后的权限特征集用互信息和笛卡尔积方法进行权限聚类去冗余，最后在此基础上构建朴素贝叶斯分类模型，以及对所检测到的恶意性应用程序进行恶意性等级的划分。

[0008] 进一步，该方法具体包括以下步骤：

[0009] 步骤一：收取并创建恶意性应用程序和非恶意性应用程序的样本库，分别对其APK样本进行反编译处理得到AndroidManifest.xml文件，然后提取该文件的权限特征，获得权限特征集；

[0010] 步骤二：利用Android权限特征属性变量之间的相关性关系，其中任意两个变量之间的相关性可能是由于第三个变量的存在所表现出来的，对此采用基于偏相关系数对权限特征属性进行相关性分析的方法，对权限特征集进行降维预处理；

[0011] 步骤三：利用基于互信息理论和笛卡尔积方法，采用基于互信息和笛卡尔积的改进的朴素贝叶斯分类模型方法，对权限特征集降维预处理后获取的权限特征集进行聚类去冗余；

[0012] 步骤四：基于分类属性集的集合X_{new}构建朴素贝叶斯分类器，通过样本训练获得先验概率，然后用测试集样本通过计算后验概率判断所检测的Android应用程序是否具有恶意性，对具有恶意性的Android应用程序按概率方法进行等级划分。

[0013] 进一步，在步骤二中，所述基于偏相关系数对权限特征属性进行相关性分析的方法具体包括：

[0014] 该方法首先通过计算两个权限特征属性变量之间的简单相关系数 $r(x_i,x_j)=\frac{Cov(x_i,x_j)}{\sqrt{D(x_i)D(x_j)}}$，其中$Cov(x_i,x_j)$是$x_i$与$x_j$之间的协方差，$\sqrt{D(x_i)D(x_j)}$是$x_i$与$x_j$之间的标准差，将计算所得的简单相关系数做成相关系数矩阵R，计算|R|行列式中r_{ii}，r_{ij}，r_{jj}的代数余子式A^{ii}，A^{ij}，A^{jj}，然后带入特征权限属性变量之间的偏相关系数 $\rho(x_i,x_j|x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_{j-1},x_{j+1},\cdots,x_n)=\frac{-A^{ij}}{\sqrt{A^{ii}A^{jj}}}$公式进行计算，根据得到的偏相关系数|ρ|的值判断权限特征属性之间的相关性大小，去除相关性低的权限特征属性，得到降维预处理后的权限特征集。

[0015] 进一步，在步骤三中，利用基于互信息理论和笛卡尔积方法，采用基于互信息和笛卡尔积的改进的朴素贝叶斯分类模型方法，对权限特征集降维预处理后获取的权限特征集进行聚类去冗余，聚类去冗余模型如下：

[0016] $$Cor(X_i,C)=\sum_{x\in c}P(X_i,C)\log\frac{P(X_i,C)}{P(X_i)P(C)}$$

[0017] $$Cor(X_i,X_j)=\sum_{x\in c}P(X_i,X_j)\log\frac{P(X_i,X_j)}{P(X_i)P(X_j)}$$

[0018] 其中$Cor(X_i,C)$表示权限特征属性变量X_i和类别属性变量C之间的相关度，$Cor(X_i,X_j)$表示权限特征属性变量X_i和X_j之间的相关度，计算方式如下：

[0019] 1)计算预处理后权限特征属性变量X_i与类别变量C的相关度$Cor(X_i,C)$，按从大到小的顺序排列构成原始属性集X-ori；

[0020] 2)计算X-ori中的第一个属性变量X-ori(1)与其它属性变量的相关度Cor(X-ori(1),X_j)；

[0021] 3)对X-ori中除X-ori(1)之外的其它变量X_j，若$Cor(X\text{-}ori(1),X_j)>Cor(X_j,C)$，则认为该变量与X-ori(1)高度相关，将其加入X-ori(1)的相关集中；

[0022] 4)X-ori(1)及其相关集中的前m个变量的笛卡尔积X_{new1}作为新属性集加入X_{new}，同时从X-ori中删除X-ori(1)及其相关集中的所有变量；

[0023] 5)重复2)至4)，直到X-ori=ø为止。

[0024] 进一步，在步骤四中，基于分类属性集的集合X_{new}构建朴素贝叶斯分类器，通过样本训练获得先验概率，然后用测试集样本通过计算后验概率判断所检测的Android应用程序是否具有恶意性，基于权限分类属性集的集合X_{new}和类别C构建朴素贝叶斯的模型如下：

[0025] $$P(C_i|X_{new})=\frac{P(X_{new}|C_i)P(C_i)^{\alpha}}{P(X_{new})}$$

[0026] 其中，$\alpha=1+\frac{count(X_k|C_i)}{count(X_k)count(X)}$，$count(X_k|C_i)$表示在类别$C_i$样本中权限特征属性$X_k$出现的次数，$count(X_k)$表示样本中权限特征属性$X_k$出现的次数，count(X)表示分类权限集集合$X_{new}$中权限特征集的个数，α表示不同权限特征属性对分类的影响程度，且量化了权限特征属性与其类别属性之间的关系，X_{new}为Android应用程序的权限特征属性集的集合，C_i是Android应用程序的类别，即非恶意性应用程序和恶意性应用程序两类，$P(X_{new})$对于所有类为常数，因此比较后验概率是只需要$p(X_{new}|C_i)P(C_i)^{\alpha}$最大即可判断应用程序是否具有恶意性；

[0027] 对所得的具有恶意性的Android应用程序的权限特征集，对恶意性Android应用程序进行恶意性等级划分，计算恶意性等级如下：

[0028] $$T=\sum\frac{P_v}{P_m}$$

[0029] $$P_v=\prod_{i=1}^{n}P_v(X_i)$$

[0030] $$P_m=\prod_{i=1}^{n}P_m(X_i)$$

[0031] 其中，P_v表示该待测试样本应用程序在恶意性程序中出现的概率；P_m表示该待测试

样本应用程序在非恶意性程序中出现的概率;$P_v(X_i)$代表第i个权限特征集在恶意性程序中出现的概率;$P_n(X_i)$代表第i个权限特征集在非恶意性程序中出现的概率。

[0032] 本发明的有益效果在于:本发明通过对Android应用程序样本进行反编译得到其所使用的相关权限,为了后面朴素贝叶斯模型的建立,本发明采用了偏相关系数对Android应用程序权限特征属性进行相关度分析,对权限特征属性进行降维预处理,然后利用互信息和笛卡尔积方法对降维后的权限特征集进行相关性聚类去冗余,得到新的分类权限特征集,由于聚类后的权限集之间相关性很低,几乎是相互独立的关系,因此满足了朴素贝叶斯属性相互独立的条件,在此基础上构建朴素贝叶斯分类器,能使应用程序分类决策相关性高,另外对朴素贝叶斯所做改进进一步提高Android应用程序恶意性的检测率,在聚类过程中设定阈值也可以避免过拟合的现象,对恶意性进行等级划分,这在实际应用上提高了应用程序在安装时的安全性,本发明用于未来Android应用软件安装前的安全性检测,可以提醒用户应用程序是否具有恶意性以及恶意性强度和等级,这对应用程序使用的安全性研究具有深远的意义和广阔的研究。

附图说明

[0033] 为了使本发明的目的、技术方案和有益效果更加清楚,本发明提供如下附图进行说明:

[0034] 图1为本发明所述方法的流程示意图;

[0035] 图2是对权限特征进行降维预处理的示意图;

[0036] 图3为对降维预处理后的权限特征聚类去冗余的示意图。

具体实施方式

[0037] 下面将结合附图,对本发明的优选实施例进行详细的描述。

[0038] 图1为本发明所述方法的流程示意图,如图所示,本发明所述的android应用程序恶意性的静态检测方法主要包括以下四个步骤:步骤一:收取并创建恶意性应用程序和非恶意性应用程序的样本库,分别对其APK样本进行反编译处理得到AndroidManifest.xml文件,然后提取该文件的权限特征,获得权限特征集;步骤二:利用Android权限特征属性变量之间的相关性关系,其中任意两个变量之间的相关性可能是由于第三个变量的存在所表现出来的,对此提出一种基于偏相关系数对权限特征属性进行相关性分析的方法,对权限特征集进行降维预处理,该方法首先通过计算两个权限特征属性变量之间的简单相关系数$r(x_i,x_j)=\frac{Cov(x_i,x_j)}{\sqrt{D(x_i)D(x_j)}}$,其中$Cov(x_i,x_j)$是$x_i$与$x_j$之间的协方差,$\sqrt{D(x_i)D(x_j)}$是$x_i$与$x_j$之间的标准差,将计算所得的简单相关系数做成相关系数矩阵R,计算|R|行列式中r_{ii},r_{ij},r_{jj}的代数余子式A^{ii},A^{ij},A^{jj},然后带入特征权限属性变量之间的偏相关系数$\rho(x_i,x_j|x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_{j-1},x_{j+1},\cdots,x_n)=\frac{-A^{ij}}{\sqrt{A^{ii}A^{jj}}}$公式进行计算,根据得到的偏相关系数|ρ|的值判断权限特征属性之间的相关性大小,去除相关性低的权限特征属性,得到降维预处理后的权限特征集;步骤三:利用基于互信息理论和笛卡尔积方法,提出的一种基于互信息和笛卡尔积的改进的朴素贝叶斯分类模型方法,对权限特征集降维预处理后获取的权限特

征集进行聚类去冗余，(1)计算预处理后权限特征属性变量X_i与类别变量C的相关度$Cor(X_i, C)$，按从大到小的顺序排列构成原始属性集X-ori；(2)计算X-ori中的第一个属性变量X-ori(1)与其它属性变量的相关度$Cor(X\text{-}ori(1), X_j)$；(3)对X-ori中除X-ori(1)之外的其它变量X_j，若$Cor(X\text{-}ori(1), X_j) > Cor(X_j, C)$，则认为该变量与X-ori(1)高度相关，将其加入X-ori(1)的相关集中；(4)X-ori(1)及其相关集中的前m个变量的笛卡尔积X_{new1}作为新属性集加入X_{new}，同时从X-ori中删除X-ori(1)及其相关集中的所有变量；(5)重复(2)-(4)，直到X-ori=ø为止；步骤四：基于分类属性集的集合X_{new}构建朴素贝叶斯分类器，通过样本训练获得先验概率，然后用测试集样本通过计算后验概率判断所检测的Android应用程序是否具有恶意性，对具有恶意性的Android应用程序按概率方法进行等级划分。

[0039] 在步骤一中，对收取并创建的恶意性应用程序和非恶意性应用程序的样本库分别进行反编译处理得到AndroidManifest.xml文件，提取其权限特征，获得权限特征集；

[0040] 图2是对权限特征进行降维预处理的示意图，在步骤二中，该方法通过利用基于偏相关系数对权限特征属性变量之间的相关性关系进行分析，对权限特征属性进行降维预处理，分析权限特征属性之间相关性的方法模型如下：

[0041] $$r(x_i, x_j) = \frac{Cov(x_i, x_j)}{\sqrt{D(x_i)D(x_j)}}$$

[0042] $$\rho(x_i, x_j \mid x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) = \frac{-A^{ij}}{\sqrt{A^{ii}A^{jj}}}$$

[0043] $$A^{ij} = (-1)^{i+j} M^{ij}$$

[0044] $$R = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nn} \end{bmatrix}$$

[0045] 其中$r(x_i, x_j)$为简单相关系数；$Cov(x_i, x_j)$是x_i与x_j之间的协方差；$\sqrt{D(x_i)D(x_j)}$是x_i与x_j之间的标准差；A^{ii}，A^{ij}，A^{jj}为由简单相关系数做成矩阵R的|R|行列式中r_{ii}，r_{ij}，r_{jj}的代数余子式；M^{ij}是n阶行列式|R|的余子式，即去掉n阶行列式|R|中第i行第j列后剩下的n-1阶行列式即为M^{ij}。通过计算两个权限特征属性变量之间的简单相关系数$r(x_i, x_j) = \frac{Cov(x_i, x_j)}{\sqrt{D(x_i)D(x_j)}}$，将计算所得的简单相关系数做成相关系数矩阵R，计算|R|行列式中r_{ii}，r_{ij}，r_{jj}的代数余子式A^{ii}，A^{ij}，A^{jj}，然后带入特征权限属性变量之间的偏相关系数$\rho(x_i, x_j \mid x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) = \frac{-A^{ij}}{\sqrt{A^{ii}A^{jj}}}$公式进行计算，根据得到的偏相关系数|ρ|的值判断权限特征属性之间的相关性大小，去除相关性低的权限特征属性，得到降维预处理后的权限特征集。

[0046] 图3为对降维预处理后的权限特征聚类去冗余的示意图，在步骤三中，利用基于互信息理论和笛卡尔积方法，提出的一种基于互信息和笛卡尔积的改进的朴素贝叶斯分类模型方法，对权限特征集降维预处理后获取的权限特征集进行聚类去冗余，聚类去冗余模型如下：

[0047] $$Cor(X_i,C)=\sum_{x\in\{X_i,C\}}P(X_i,C)\log\frac{P(X_i,C)}{P(X_i)P(C)}$$

[0048] $$Cor(X_i,X_j)=\sum_{x\in\{X_i,X_j\}}P(X_i,X_j)\log\frac{P(X_i,X_j)}{P(X_i)P(X_j)}$$

[0049] 其中$Cor(X_i,C)$表示权限特征属性变量X_i和类别属性变量C之间的相关度，$Cor(X_i,X_j)$表示权限特征属性变量X_i和X_j之间的相关度，计算方式如下：

[0050] 1)计算各权限特征属性变量X_i与类别变量C的相关度$Cor(X_i,C)$，按从大到小的顺序排列构成原始属性集X-ori；

[0051] 2)计算X-ori中的第一个属性变量X-ori(1)与其它属性变量的相关度Cor(X-ori(1),X_j)；

[0052] 3)对X-ori中除X-ori(1)之外的其它变量X_j，若$Cor(X\text{-}ori(1),X_j)>Cor(X_j,C)$，则认为该变量与X-ori(1)高度相关，将其加入X-ori(1)的相关集中；

[0053] 4)X-ori(1)及其相关集中的前m个变量的笛卡尔积X_{new1}作为新属性集加入X_{new}，同时从X-ori中删除X-ori(1)及其相关集中的所有变量；

[0054] 5)重复(2)-(4)，直到X-ori=ø为止。

[0055] 在步骤四中，基于分类属性集的集合X_{new}构建朴素贝叶斯分类器，通过样本训练获得先验概率，然后用测试集样本通过计算后验概率判断所检测的Android应用程序是否具有恶意性，基于分类属性集的集合X_{new}和类别C构建朴素贝叶斯的模型如下：

[0056] $$P(C_i|X_{new})=\frac{P(X_{new}|C_i)P(C_i)^{\alpha}}{P(X_{new})}$$

[0057] 其中，$\alpha=1+\frac{count(X_k|C_i)}{count(X_k)count(X)}$，$count(X_k|C_i)$表示在类别$c_i$样本中权限特征属性$X_k$出现的次数，$count(X_k)$表示样本中权限特征属性$X_k$出现的次数，count(X)表示分类权限集集合$X_{new}$中权限特征集的个数，α表示不同权限特征属性对分类的影响程度，且量化了权限特征属性与其类别属性之间的关系，X_{new}为Android应用程序的权限特征属性集的集合，C_i是Android应用程序的类别，即非恶意性应用程序和恶意性应用程序两类，$P(X_{new})$对于所有类为常数，因此比较后验概率是只需要$P(X_{new}|C_i)P(C_i)^{\alpha}$最大即可判断应用程序是否具有恶意性。

[0058] 对所得的具有恶意性的Android应用程序的权限特征集，对恶意性Android应用程序进行恶意性等级划分，计算恶意性等级如下：

[0059] $$T=\sum\frac{P_v}{P_m}$$

[0060] $$P_v=\prod_{i=1}^{n}P_v(X_i)$$

[0061] $$P_m=\prod_{i=1}^{n}P_m(X_i)$$

[0062] 其中，P_v表示该待测试样本应用程序在恶意性程序中出现的概率；P_m表示该待测试

样本应用程序在非恶意性程序中出现的概率;$P_v(X_i)$代表第i个权限特征集在恶意性程序中出现的概率;$P_n(X_i)$代表第i个权限特征集在非恶意性程序中出现的概率。

[0063] 最后说明的是,以上优选实施例仅用以说明本发明的技术方案而非限制,尽管通过上述优选实施例已经对本发明进行了详细的描述,但本领域技术人员应当理解,可以在形式上和细节上对其作出各种各样的改变,而不偏离本发明权利要求书所限定的范围。

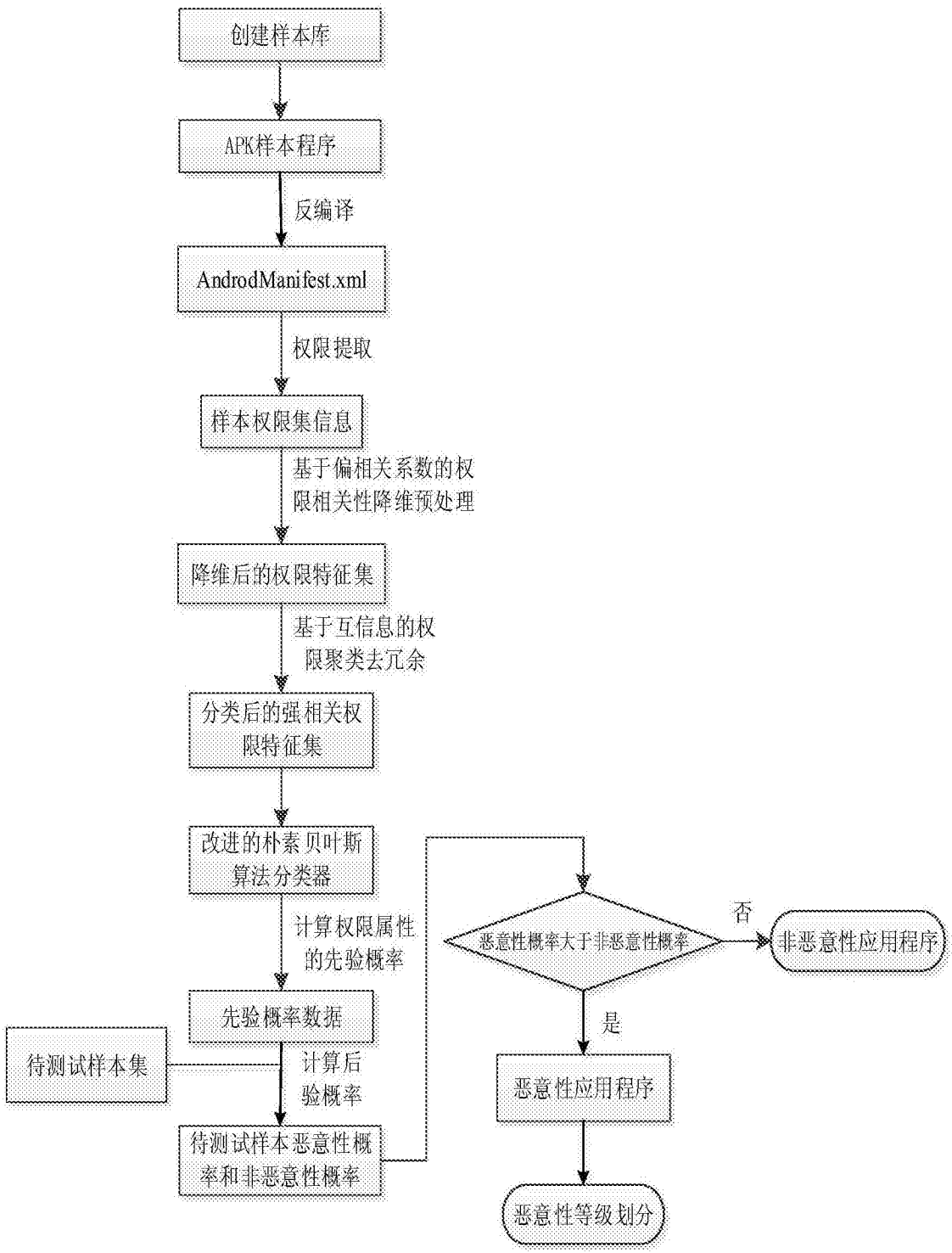

创建样本库
APK样本程序
反编译
AndrodManifest.xml
权限提取
样本权限集信息
基于偏相关系数的权限相关性降维预处理
降维后的权限特征集
基于互信息的权限聚类去冗余
分类后的强相关权限特征集
改进的朴素贝叶斯算法分类器
计算权限属性的先验概率
先验概率数据
待测试样本集
计算后验概率
待测试样本恶意性概率和非恶意性概率
恶意性概率大于非恶意性概率
否
非恶意性应用程序
是
恶意性应用程序
恶意性等级划分

图1

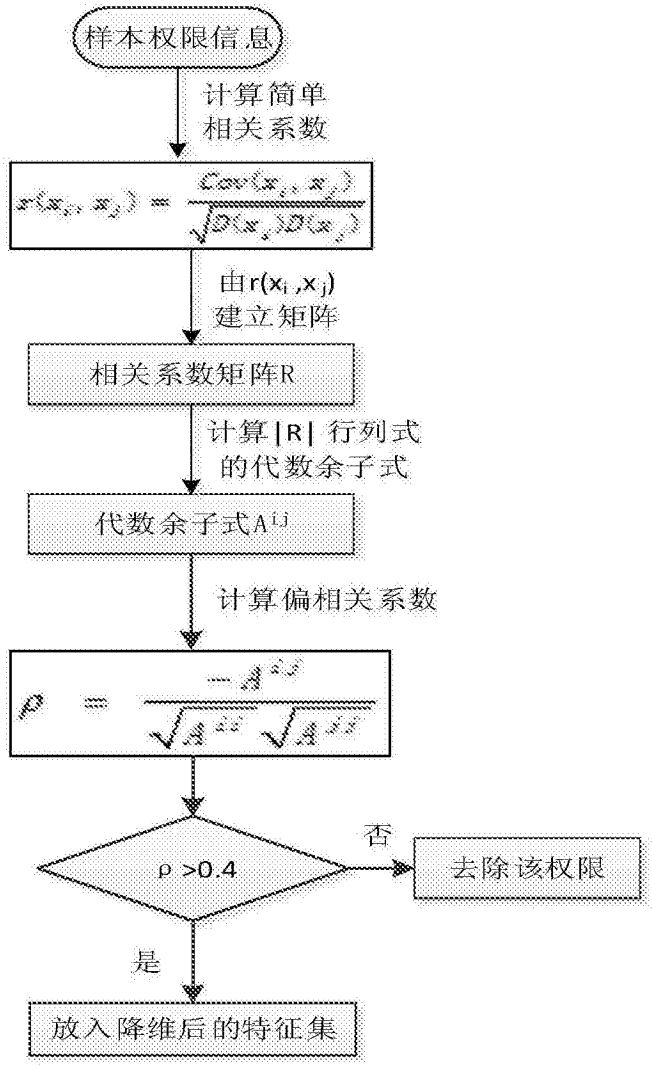

样本权限信息
计算简单
相关系数
由r(xi ,xj)
建立矩阵
相关系数矩阵R
计算|R| 行列式
的代数余子式
代数余子式A^ij
计算偏相关系数
ρ >0.4
否
去除该权限
是
放入降维后的特征集

图2

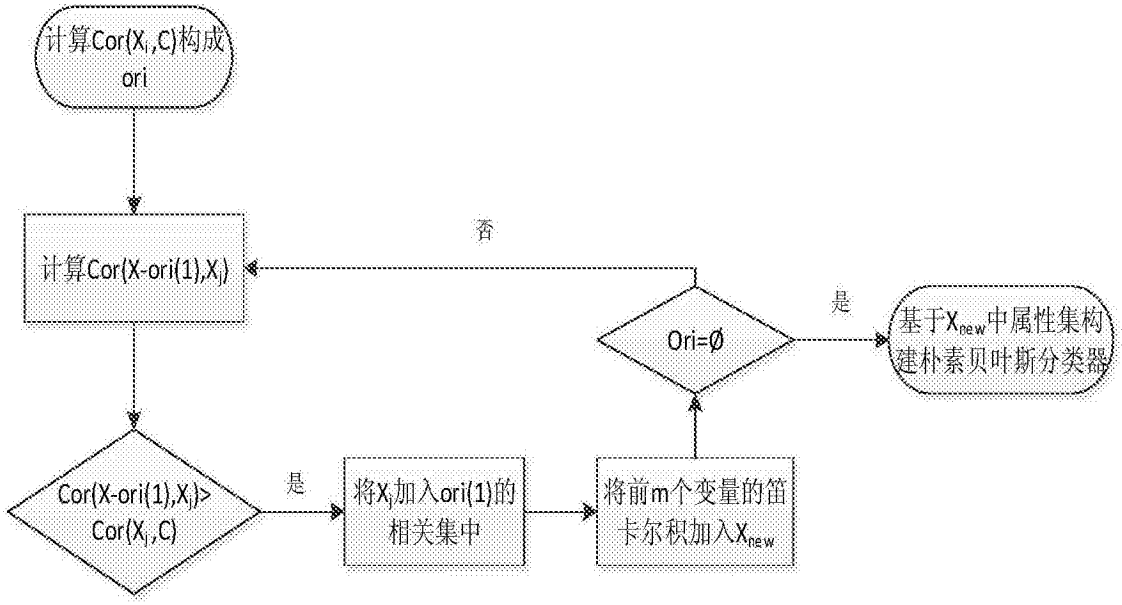

计算Cor(Xi,C)构成ori
计算Cor(X-ori(1),Xj)
否
Ori=Ø
是
基于Xnew中属性集构建朴素贝叶斯分类器
Cor(X-ori(1),Xj)>Cor(Xj,C)
是
将Xj加入ori(1)的相关集中
将前m个变量的笛卡尔积加入Xnew

图3