[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
H04L 12/56 (2006.01)
H04L 29/06 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200810021455.6

[43] 公开日 2009 年 1 月 14 日

[11] 公开号 CN 101345704A

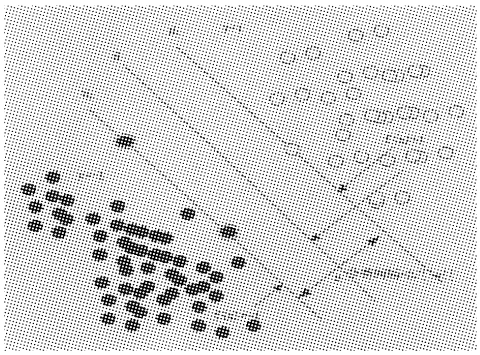
[22] 申请日 2008.8.15
[21] 申请号 200810021455.6
[71] 申请人 南京邮电大学
地址 210003 江苏省南京市新模范马路 66 号
[72] 发明人 王汝传 吴 敏 李玲娟 韩志杰
支萌萌 徐小龙 饶 元 李致远
[74] 专利代理机构 南京经纬专利商标代理有限公司
代理人 叶连生
权利要求书 2 页 说明书 9 页 附图 1 页

[54] 发明名称

基于支持向量机的对等网络流量检测方法

[57] 摘要

基于支持向量机的对等网络流量检测制方法提出了基于支持向量机的 P2P 流量检测方法，利用支持向量机技术来实现 P2P 流量检测问题，该方法将支持向量机技术应用到实际网络中的 P2P 流量检测应用中，解决 P2P 流量的检测问题，该方法所包含的步骤为两个阶段：第 1 阶段，支持向量机的训练阶段：第 2 阶段，支持向量机的实际 P2P 流量决策阶段：将机器学习领域的精度很高的模式分类器支持向量机技术应用于网络中的 P2P 流量检测中，解决 P2P 流量的检测问题。 较之其他的流量检测方法，该方法注重通过网络实际流量的挖掘发现规律，预测新未知数据的分类，该方法能够通过学习不断提高分类性能，适合完成流量较大时的识别工作，也能适合检测未知的和加密的 P2P 流量。



1.一种基于支持向量机的对等网络流量检测制方法，其特征在于该方法将支持向量机技术应用到实际网络中的 P2P 流量检测应用中，解决 P2P 流量的检测问题，该方法所包含的步骤为两个阶段：

第 1 阶段， 支持向量机的训练阶段：

a.从网络上截取网络数据包，统计 P2P 流量和正常网络流量的样本数目，得到两类样本集Ξ_1,Ξ_2，样本数目分别是N_1,N_2，其中Ξ_1表示 P2P 流量样本集，Ξ2 表示正常流量样本集，N_1表示 P2P 流量样本数目，N_2表示正常流量样本数目。

b.对这些已知的正常流量数据集、P2P 流量数据集进行特征处理，将之转化为数字向量形式，作为训练支持向量机的依据并存入数据库,

c.针对样本数据中 P2P 流量和非 P2P 流量数目不均衡情况下的支持向量机训练:

c1：对 P2P 流量，从Ξ_1中按照等概率的方法取得 P2P 流量数据样本集$\Xi 3$，样本数目为$N_1^{'}$，满足$N_1^{'}=N_2$；

c2：根据网格搜索的参数搜索方法，确定支持向量机的参数 C 和γ，其中 C 为对样本的惩罚系数，r 为核函数参数。对样本集$\{\Xi 3,\Xi 2\}$进行支持向量机设计，获取参数 W，ξ_i的值，其中 W 为最优超平面的法向量，ξ_i为松弛因子，i=1……n;

c3：根据公式$\frac{1}{2}(W\cdot W)=\sum_{i=1}^{n}C_i\xi_i$和$C_1/C_2=N_2/N_1$计算此时的样本惩罚因子 C_1、C_2，其中$C_i=\begin{cases}C_1 & X_i\in\Xi_1\\ C_2 & X_i\in\Xi_2\end{cases}$，$C_1$表示对 P2P 流量样本的惩罚因子，$C_2$表示对正常流量样本的惩罚因子。

c4： 根据新的 C_1、C_2，对样本集$\{\Xi 1,\Xi 2\}$进行支持向量机设计，获取此时参数 W，ξ_i的值，其中 W 为最优超平面的法向量，ξ_i为松弛因子，i=1……n;

c5：根据公式$\frac{1}{2}(W\cdot W)=\sum_{i=1}^{n}C_i\xi_i$和$C_1/C_2=N_2/N_1$重新计算新的 C_1、C_2，判断 C_1、

C_2 的变化情况，如果变化小于所设定的阈值，即满足收敛条件 $\frac{(C_{11}-C_{10})^2+(C_{21}-C_{20})^2}{C_{11}^2+C_{21}^2}<\delta$，其中 δ 为一百分比常数；$C_{i1},C_{i0}(i=1,2)$ 分别表示当前时刻与前一时刻的 C_i 值；则获得最终基于两类流量检测的 SVM 模型，得到最终分类决策函数，否则返回步骤 c4，

第 2 阶段， 支持向量机的实际 P2P 流量决策阶段：

d. 从网络上截取网络数据包，

e. 对获取流量数据进行特征处理，将之转化为数字向量形式，作为训练支持向量机的依据并存入数据库，

f. 根据 SVM 训练模型即最终分类决策函数得出分类结果并存入数据库，即如果最终分类决策函数 f(x)大于 0，表示属于该连接属于 P2P 流量，否则属于正常网络流量；

g. 根据数据库中保存的决策结果，结合网络实际运行情况，进行 P2P 流量流量分析并作出相应控制。

基于支持向量机的对等网络流量检测方法

技术领域

本发明提出了基于支持向量机的 P2P 流量检测方法，利用支持向量机技术来实现 P2P 流量检测问题，属于分布式计算安全领域。

背景技术

随着 P2 P 网络技术在 90 年代后期的兴起，P2P 流量逐渐成为了互联网流量的重要组成部分。精确地识别 P2P 流量对于有效地管理网络和合理地利用网络资源都具有重要意义。

目前 P2P 流量检测技术大致有以下三类:基于端口的检测技术，深层数据包检侧技术和基于流量特征的检测技术。

基于端口的分析方法是在网络流量中探测 P2P 用户最基本，最直接的方法。但由于现在大多数 P2P 应用允许用户手动选择随意的端口号来设置默认的端口号或使用随机的端口号，从而使得端口号不可预测，还有一些 P2P 应用使用默认端口号(例如 80 端口)来伪装自己的功能端口，因此基于端口号的分析方法的效率变得很差。

深层数据包检测技术，通过深入检测其数据包中的有效载荷来进行检测，即通过应用层数据包的正则表达式的匹配来完成探测工作，以确定特定的 P2P 应用。该方法识别准确度高，实现简单，维护方便。但该方法是高资源消耗的，由于必须读取处理所有网络流量，会严重地增加网络设备负担甚至会导致网络的崩溃，因而不适合大型网络。另外该方法对加密 P2P 流量捕获能力弱，对新的 P2P 应用必须升级后才能检测且该方法容易和隐私保护法律条款产生冲突。

基于流量特征的检测技术是利用 P2P 在传输层表现出来的流量特征来发现 P2P 应用。这类方法借用了统计学领域通用的一些概念，分析传输层的信息，不需要任何关于应用层协议的信息，几乎不需要任何额外的软件或者硬件并具有较强的加密和未知 P2P 流量的捕获能力，因而近年来关于流统计方式测量 P2P 流量得

到了国内外广泛的关注，被认为是最有前途的一种方法。目前主要包括以下几种识别方式：{IP, port}识别、TCP/UDP端口识别、BlockSize识别、基于会话(session)分类的识别、双向识别、流统计状态的识别等等，该方法虽然具有性能高、可扩展性好的有点，但由于准确性差，因此在实际应用部署在中也面临诸多困难。

支持向量机技术(Support Vector Machine, SVM)由 vapnik 及其合作者发明，在1992年计算机学习理论的会议上介绍进入机器学习领域，在20世纪90年代中后期得到了全面深入的发展，现已成为机器学习和数据挖掘领域的标准工具，在许多领域如手写识别、文本分类、入侵检测等已经取得相当出色的应用效果。支持向量机是统计学习理论中最年轻的内容，也是最实用的部分，已经被公认为是精度最高的模式分类器之一，它也是机器学习领域若干标准技术的集大成者，

支持向量机基本思想可用图 1 的二维情况说明。图中，实心点和空心点分别代表两类样本，H 为分类线，H_1、H_2 分别为过各类中离分类线最近的样本且平行于分类线的直线，它们之间的距离叫做分类间隔(margin)。所谓最优分类线就是要求分类线不但能将两类正确分开(训练错误率为 0)，而且使分类间隔最大。分类线方程为

$w \cdot X+b=0$， (1)

其中 w 为最优分类线的法向量，b 为偏置，样本集为 $(X_i, y_i), i=1,2,\cdots,n, X \in R^d, y_i \in \{-1,1\}$ 是类别标号，满足

$y_i [(w \cdot X_i)+b] - 1 \geq 0, i=1,2\ldots n$

此时，分类间隔为 $2/\|w\|$，使间隔最大等价于使 $\|w\|^2$ 最小。满足条件式(1)且使 $\|w\|^2$ 最小的分类面称为最优分类面，H_1, H_2 上的训练样本点称为支持向量。使分类间隔最大实际上就是对推广能力的控制，这是支持向量机的核心思想之一。根据统计学习理论求最优分类面的问题可转化为一个二次规划的优化问题，即在式(1)的约束下，求函数(2)的最小值。

$$\phi(w)=\frac{1}{2}\|w\|^2=\frac{1}{2}(w\cdot w) \qquad (2)$$

该问题可转化为在约束条件 $\sum_{i=1}^{n} a_i y_i = 0, a_i \geq 0, i=1,2,\ldots n$ 之下求式(3)的最大值，其中：a_i >0 为 Lagrange 系数。若 a_i^* 为最优解，求解上述问题后得到的最优

分类函数是:

$$sgn(w\cdot x)+b \qquad (4)$$

其中:sgn()为符号函数，b^*是分类的阈值,可以由任意一个支持向量用式(4)求得,或通过两类中任意一对支持向量取中值求得。对于给定的未知样本x,只需计算sgn(w·x)+b,即可判定x所属的分类。对于线性不可分情况，用内积$K(x_i\cdot x_j)$代替最优分类面中的点积,就相当于把原特征空间变换到了某一新的特征空间,此时优化函数变为:

$$w(a)=\sum_{i=1}^{n}a_i-\frac{1}{2}\sum_{i,j}^{n}a_ia_jy_iy_jK(x_i\cdot x_j) \qquad (5)$$

相应的判别函数也应变为:

$$f(x)=sgn\{\sum_{i=1}^{n}a_i^*y_iK(x_i\cdot x)+b^*\} \qquad (6)$$

实际上，从模式识别的角度而言，P2P流量的识别过程可以看作是一个二分类问题：即对流量数据进行分类，分为P2P流和非P2P流。因此本发明提出了一种新型的基于支持向量机的P2P流量识别办法，将支持向量机技术应用到流量检测的分类问题中去，事先离线获得大量训练正常流量和P2P流量样本数据，输入到支持向量机中构建最优分类面并以此作为网络P2P流量的检测方法。考虑到实际网络世界中所采集的样本流量数据中P2P流量的样本数目将会大大超过正常流量的样本数目，因为据统计，P2P应用已占ISP业务总量的60%～80%，已经跃然成为网络带宽最大的消费者，而一般的支持向量机技术在训练时对不同类别的错误惩罚是相同的，将会使得分类线偏向样本密度小的一方，从而降低分类的精度。因此本发明提出的基于支持向量机的P2P流量检测方法将根据实际训练的样本数目确定不同流量相应的惩罚因子，从而提高P2P流量识别的精度。因而本发明提出的基于观测数据(样本)的方法能够有效的进行P2P流量的模式识别问题。

发明内容

技术问题：本发明的目的是提供一种基于支持向量机的对等网络流量检测方法,将机器学习领域的精度很高的模式分类器支持向量机技术应用于网络中的P2P流量检测中，解决P2P流量的检测问题。较之其他的流量检测方法，该方法注重通过网络实际流量的挖掘发现规律，预测新未知数据的分类，该方法能够通过学

习不断提高分类性能，适合完成流量较大时的识别工作，也能适合检测未知的和加密的 P2P 流量。

技术方案：本发明的方法强调网络流量统计特性的挖掘和学习，从中发现规律，特别是考虑到实际样本训练时 P2P 流量和非 P2P 流量的样本悬殊带来的支持向量机识别精度下降问题，其目的是解决实际网络中的 P2P 流量识别问题。

本发明的基于支持向量机的对等网络流量检测制方法将支持向量机技术应用到实际网络中的 P2P 流量检测应用中，解决 P2P 流量的检测问题，该方法所包含的步骤为两个阶段：

第 1 阶段， 支持向量机的训练阶段：

a. 从网络上截取网络数据包，统计 P2P 流量和正常网络流量的样本数目，得到两类样本集Ξ_1,Ξ_2，样本数目分别是N_1,N_2，其中Ξ_1表示 P2P 流量样本集，Ξ2 表示正常流量样本集，N_1表示 P2P 流量样本数目，N_2表示正常流量样本数目。

b. 对这些已知的正常流量数据集、P2P 流量数据集进行特征处理，将之转化为数字向量形式，作为训练支持向量机的依据并存入数据库,

c. 针对样本数据中 P2P 流量和非 P2P 流量数目不均衡情况下的支持向量机训练：

c1：对 P2P 流量，从Ξ_1中按照等概率的方法取得 P2P 流量数据样本集$\Xi 3$，样本数目为$N_1^{'}$，满足$N_1^{'}=N_2$；

c2：根据网格搜索的参数搜索方法，确定支持向量机的参数 C 和γ，其中 C 为对样本的惩罚系数，r 为核函数参数。对样本集$\{\Xi 3,\Xi 2\}$进行支持向量机设计，获取参数 W，ξ_i的值，其中 W 为最优超平面的法向量，ξ_i为松弛因子，i=1……n；

c3：根据公式$\frac{1}{2}(W\cdot W)=\sum_{i=1}^{n}C_i\xi_i$和$C_1/C_2=N_2/N_1$计算此时的样本惩罚因子 C_1、C_2，其中$C_i=\begin{cases}C_1 & X_i\in\Xi_1\\ C_2 & X_i\in\Xi_2\end{cases}$，$C_1$表示对 P2P 流量样本的惩罚因子，$C_2$表示对正常流量样本的惩罚因子。

c4： 根据新的 C_1、C_2，对样本集$\{\Xi 1,\Xi 2\}$进行支持向量机设计，获取此时参数 W，ξ_i的值，其中 W 为最优超平面的法向量，ξ_i为松弛因子，i=1……n，

c5：根据公式 $\frac{1}{2}(W\cdot W)=\sum_{i=1}^{n}C_i\xi_i$ 和 $C_1/C_2=N_2/N_1$ 重新计算新的 C_1、C_2，判断 C_1、C_2 的变化情况，如果变化小于所设定的阈值，即满足收敛条件 $\frac{(C_{11}-C_{10})^2+(C_{21}-C_{20})^2}{C_{11}^2+C_{21}^2}<\delta$，其中 δ 为一百分比常数；$C_{i1},C_{i0}(i=1,2)$ 分别表示当前时刻与前一时刻的 C_i 值；则获得最终基于两类流量检测的 SVM 模型，得到最终分类决策函数，否则返回步骤 c4，

第 2 阶段， 支持向量机的实际 P2P 流量决策阶段：

d. 从网络上截取网络数据包，

e. 对获取流量数据进行特征处理，将之转化为数字向量形式，作为训练支持向量机的依据并存入数据库，

f. 根据 SVM 训练模型即最终分类决策函数得出分类结果并存入数据库，即如果最终分类决策函数 f(x)大于 0，表示属于该连接属于 P2P 流量，否则属于正常网络流量；

g. 根据数据库中保存的决策结果，结合网络实际运行情况，进行 P2P 流量流量分析并作出相应控制。

有益效果：

P2P 流量的实时识别能力：现有方法都以离线数据分析为主，缺乏 P2P 流量的实时识别能力。本发明提出的方法可以离线采集特定环境和时间范围内的实际网络流量，离线进行不平衡样本的支持向量机训练，产生适合特定时间范围内的 P2P 流量识别模型后能用于实时网络流量中 P2P 流量的检测

对加密流和新一代 P2P 数据流量的有效快速识别：基于深层数据包检测 DPI 技术面临着诸如如何提高检测算法的性能、如何支持对加密数据的分析、如何更新 P2P 应用特征库等问题。而本发明提出的方法是在 P2P 流统计特性的基础上的机器学习方法，和各种 P2P 应用的具体协议特征字无关，因而只要其流特性不发生变化，原有的支持向量机训练模型就一直有效。

P2P 流量新特征的学习能力：当 P2P 流量特性发生变化，则可重新采集网络流量数据，根据新的流量特征进行流量数据预处理后重新学习形成新的训练模型以适合于新条件下 P2P 流量识别。

识别精度高，结构简单，便于应用：支持向量机是统计学习理论中最年轻的内容，也是最实用的部分，已经被公认为是精度最高的模式分类器之一，而 P2P

流量检测本身就可以看成为两类的模式识别问题。传统的基于流量特征的 P2P 流量识别方法，虽然具有性能高、可扩展性好的有点，但由于准确性差，因此在实际应用部署在中也面临诸多困难。本发明创造性地把支持向量机技术应用到 P2P 流量检测的模式识别问题中，识别精度高，体系架构清晰，便于部署。

充分考虑了实际网络中 P2P 流量和正常流量可能的样本数目悬殊问题，学习速度快、范化能力强：该方法提出了针对不同类别样本数量不均衡时的不同惩罚因子算法，并且在利用网络样本流量进行支持向量机模型训练时，不需进行网络迭代训练，也不需要对实际网络流量建模，其求解速度明显高于神经网络，并具有较高的泛化能力。

附图说明

图 1 为支持向量机 SVM 的最优分类面示意图，图中，实心点和空心点分别代表两类样本，H_1、H_2 分别为过各类中离分类线最近的样本且平行于分类线的直线，H 代表最优分类面

图 2 是本方法的体系结构示意图，依次包括为数据采集模块、数据预处理模块、数据库模块、基于不平衡样本的 SVM 训练模块、SVM 决策模块和 P2P 流量控制模块。

具体实施方式

一、 体系结构

基于支持向量机的 P2P 流量检测方法实现的体系结构一般包括数据采集、数据预处理、数据库、基于不平衡样本的 SVM 训练、SVM 决策和 P2P 流量控制模块。该体系结构通过离线的样本流量学习、训练，实现实时的 P2P 流量的识别。图 2 给出了基于支持向量机的 P2P 流量检测体系结构，依次包括为数据采集模块、数据预处理模块、数据库模块、基于不平衡样本的 SVM 训练模块、SVM 决策模块和 P2P 流量控制模块。

下面我们给出结构中各个模块的具体说明：

数据采集模块：从网络上截取网络数据包。

数据预处理模块：主要包括特征的选择和适合于支持向量机训练的向量标准数据的处理以及导入数据库

数据库模块：其中保存了经过预处理后的训练数据、实时数据和检测识别结

果。

基于不平衡样本的SVM训练模块：根据数据库中的标准数据训练SVM分类器，确定SVM中核函数的参数，不同流量的惩罚因子以及SVM模型的确立等，最后通过训练得到一个用于决策的SVM分类器。该模块是本发明的核心。考虑到实际网络世界中所采集的样本流量中P2P流量的样本数目大大超过正常流量的样本数目，所以本发明提出对不同类别流量训练的样本数目应该确定不同的惩罚因子C_i。

SVM分类决策模块:根据SVM分类器确定的模型，对实时的未知流量进行决策识别，并把结果保存在数据库中并传递给P2P流量控制模块。

P2P流量分析控制模块：根据数据库中P2P流量的识别结果，结合实际网络中负载情况制定适合的P2P流量控制策略并进行控制。

二、 方法流程

本发明提出的基于支持向量机的P2P流量检测阶段的主要工作流程是：网络流量数据的采集和预处理阶段、支持向量机的训练阶段和支持向量机的实际流量检测阶段，P2P流量的分析和控制阶段。

1 网络流量数据的采集和预处理

从网络上截取网络数据包采集完数据后，必须对这些原始数据进行特征提取。提取的特征必须满足以下两点要求：第一要适合在SVM机制下进行分类；第二要能够体现出P2P和非P2P的正常流量的区别。如目前国内外研究提出来的P2P流量和非P2P流量在流统计方式上的一些区别，例如通过{IP,port}识别、TCP/UDP端口识别、BlockSize识别，基于会话（session）分类的识别、双向识别等流统计状态的识别方式等等。但由于样本数据集中提供是离散和不规范的，所以在模型处理数据之前，必须对数据进行量化处理和规格化处理，即将数据转换或者归并，以构成一个适合支持向量机的描述形式。数据转换的方法可以是平滑处理，去除数据中的噪声、数据泛化处理即用抽象的更高层次的概念来取代低层次或数据层的数据对象，例如根据流量样本数据构造出训练样本矩阵和类别样本矩阵，类别样本矩阵中的值为样本的分类类别，通常为1和一1。其中1表示P2P流量，-1表示非P2P流量。预处理过程结束后将相应的样本向量存储在数据库中。

2 基于不平衡样本的SVM训练

考虑到实际网络世界中所采集的样本流量中P2P流量的样本数目大大超过正常流量的样本数目，而传统的支持向量机技术比较适合于各类样本数目相等的分

类，这时 SVM 基本上能得到合理的分类线，且该分类线随着样本数目的增加，能逐渐逼近期望分类线，但是对于两类样本数目相差悬殊的情况，传统的支持向量机方法的分类线会明显偏向样本少的一方，造成了较大的分类误差。造成这种现象的根本原因在于对不同类别的错误划分采用了相同的惩罚系数 C，使得分类线偏向样本密度小的一方，这样可使错分样本数目减少。因此本发明提出针对这种 P2P 流量和非 P2P 的正常流量样本数目相差悬殊时的支持向量机改进方法，即对两类错分样本进行不同的惩罚。这时关键在于如何确定 $C_i(i=1,2)$ 值。如果 C_i 值太小，表明对样本的错分惩罚太小，可能会导致过多的错分样本，如果 C_i 值太大，表明对样本的错分惩罚大，同样起不到折中的作用。因此本发明采用迭代方法寻找到针对两类流量的最合适的惩罚因子。

3 支持向量机 svm 的实际流量预测

根据 SVM 训练模型即最终分类决策函数得出分类结果并存入数据库，即如果 f(x)大于 0，表示属于该连接属于 P2P 流量，否则属于正常网络流量；

4 P2P 流量的分析和控制

根据数据库中保存的决策结果，结合网络实际运行情况，进行 P2P 流量流量分析并作出相应控制。

具体实例：

1 训练阶段：

（1）从网络上截取网络数据包，统计 P2P 流量和正常网络流量的样本数目。得到两类样本集 Ξ_1，Ξ_2，样本数目分别是 N_1, N_2，其中 Ξ_1 表示 P2P 流量样本集，Ξ2 表示正常流量样本集，N_1 表示 P2P 流量样本数目，N_2 表示正常流量样本数目。

（2）对这些已知的正常流量数据、P2P 流量数据进行特征处理，将之转化为数字向量形式，作为训练支持向量机的依据并存入数据库。

（3）对 P2P 流量，按照等概率的方法取得样本集 $\Xi 3$，该样本集样本数目为 N_1' 满足 $N_1' = N_2$；

（4）根据网格搜索（grid-search）的参数搜索方法，确定支持向量机的参数 C 和 γ，对样本集{$\Xi 3, \Xi 2$}进行支持向量机设计，获取支持向量机参数 W，ξ_i 的

值；

（5）根据公式 $\frac{1}{2}(W \cdot W) = \sum_{i=1}^{n} C_i \xi_i$ （7）和 $C_1/C_2 = N_2/N_1$ （8），计算此时的不同样本集的惩罚因子 C_1, C_2，其中 $C_i = \begin{cases} C_1 & X_i \in \Xi_1 \\ C_2 & X_i \in \Xi_2 \end{cases}$

（6）根据 C_1, C_2，对样本集{Ξ1, Ξ2}进行支持向量机设计，获取 W，ξ_i 的值；

（7）根据公式（7）和（8）重新计算新的 C_1, C_2，判断 C_1, C_2 的变化情况，如果变化小于所设定的阈值，即满足收敛条件 $\frac{(C_{11} - C_{10})^2 + (C_{21} - C_{20})^2}{C_{11}^2 + C_{21}^2} < \delta$（其中 δ 为一百分比常数，如 5%。$C_{i1}, C_{i0} (i = 1,2)$ 分别表示当前时刻与前一时刻的 C_i 值。），则获得最终基于两类流量检测的 SVM 模型，得到最终分类决策函数。否则返回（6）。

2 决策阶段

（1）从网络上截取网络数据包，

（2）对获取流量数据进行特征处理，将之转化为数字向量形式，作为训练支持向量机的依据并存入数据库。

（3）根据 SVM 训练模型即最终分类决策函数得出分类结果并存入数据库，即如果最终分类决策函数 f(x)大于 0，表示属于该连接属于 P2P 流量，否则属于正常网络流量；

（4）根据数据库中保存的决策结果，结合网络实际运行情况，进行 P2P 流量流量分析并作出相应控制。

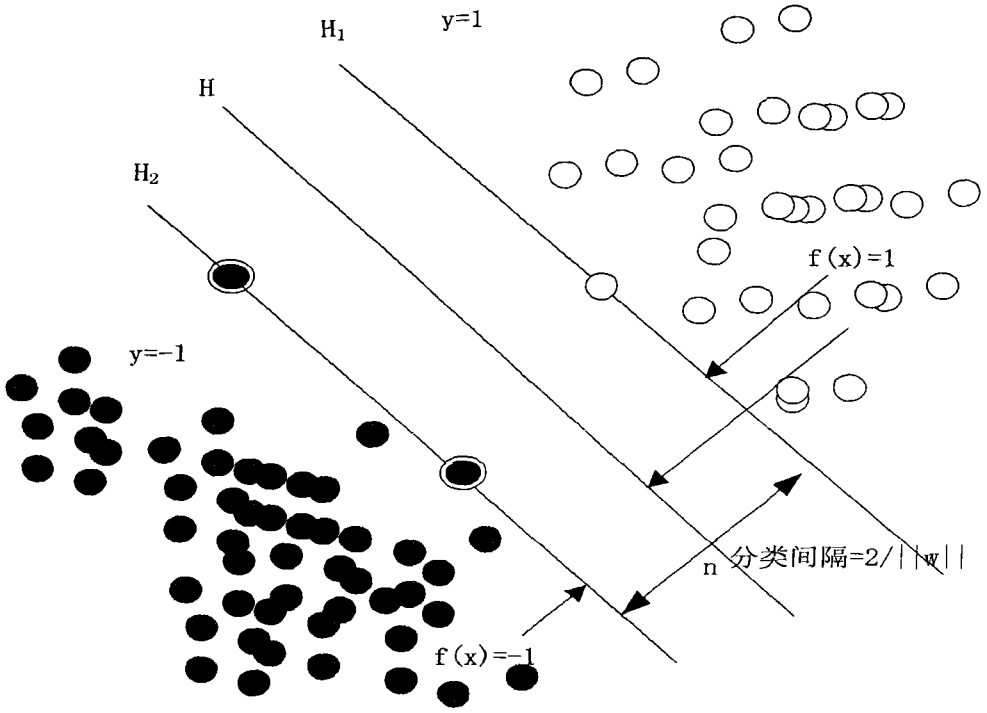

H1
y=1
H
H2
f(x)=1
y=-1
分类间隔=2/||w||
f(x)=-1

图1

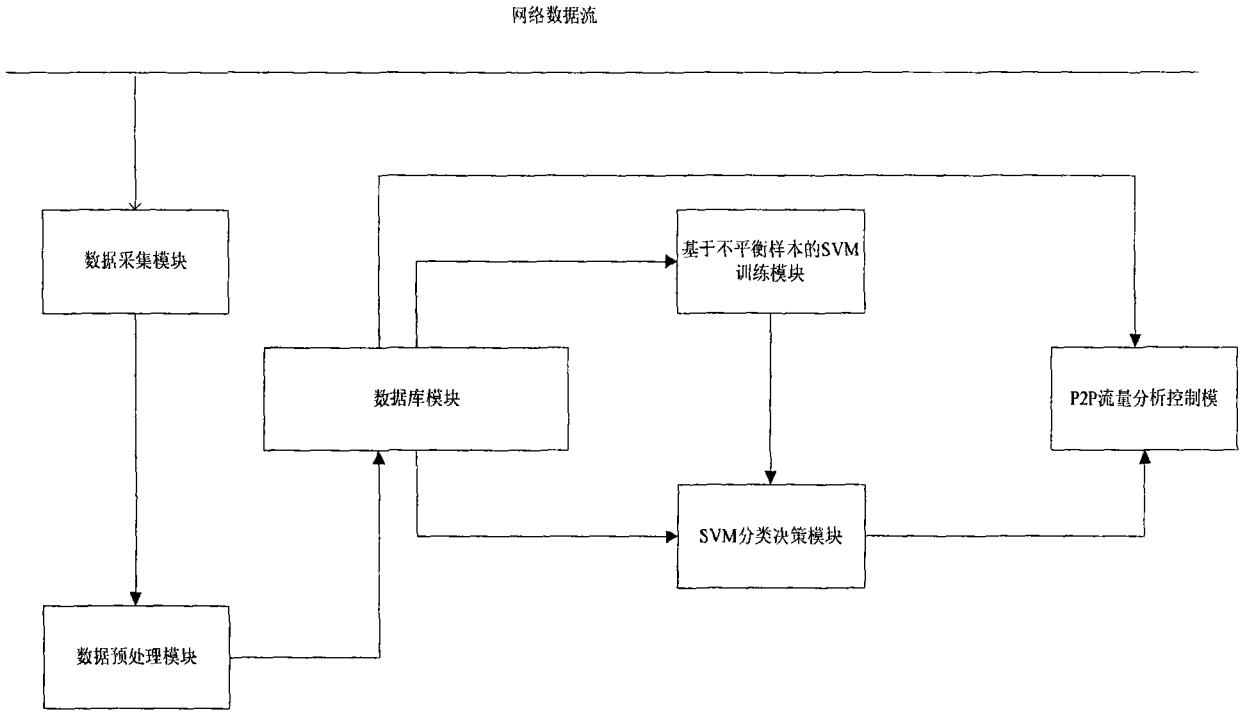

网络数据流
数据采集模块
基于不平衡样本的SVM
训练模块
数据库模块
P2P流量分析控制模
SVM分类决策模块
数据预处理模块

图 2