

# 管道流量泄漏时间序列分类的算法研究

周 鹏

ZHOU Peng

塔里木大学 信息工程学院, 新疆 阿拉尔 843300

College of information Engineering, Tarim University, Alar, Xinjiang 843300, China

E-mail: zpzqxy@163.com

ZHOU Peng. Arithmetic study on time series classification based on pipeline flux leak. *Computer Engineering and Applications*, 2009, 45(13): 166-169.

**Abstract:** Compared to traditional classification problems, time series classification poses additional difficulties. A major difficulty is due to the fact that the time sequences are variable in length, making many traditional classification methods unable to apply directly. In this paper, the data conversion based on the cluster model is presented, and then model-based cluster analysis is given, time series modeling is made with domain method, model parameters with composition of each vector is used to express sequence, the final time series matching algorithm analysis is given, classification algorithm is used for training and classification. Based on pipe flow of leakage point a time-series match, new methods is proposed, according continuity of similar samples time sequence is sorted, and the samples is inserted between adjacent samples. The new method is superior than the traditional method based on dynamic time bending.

**Key words:** time series; model based clustering; Markov model; dynamic time warping

**摘 要:** 时间序列分类比一般分类问题困难, 主要在于要分类的时间序列数据不等长, 因此不能直接应用一般的分类算法。首先提出基于聚类模型的数据转换, 然后进行基于模型的聚类分析, 用领域相关法对时间序列建模, 用模型参数组成等长向量来表示每条序列, 最后进行时间序列匹配算法分析, 用分类算法进行训练和分类。结合管道流量泄漏点提出一种时间序列匹配的新方法, 利用同类样本间的连续性规律, 将时间序列排序, 并在相邻的时间序列之间添加样本点, 新方法优于基于动态时间弯折的传统方法; 针对管道流量泄漏时间序列分类的算法研究观测到不同算法在不同因素影响下的性能表现, 为今后发展新的算法提供有力依据。

**关键词:** 时间序列; 基于模型聚类; 马尔可夫模型; 动态时间弯折

**DOI:** 10.3778/j.issn.1002-8331.2009.13.048 **文章编号:** 1002-8331(2009)13-0166-04 **文献标识码:** A **中图分类号:** TP311

## 1 引言

时间序列是指按照时间顺序取得的一系列观测值, 它蕴藏着很多有用的信息, 如何从中挖掘有价值的知识以及如何从数据库中查找与给定时间序列最相似的序列, 都是当前时间序列研究中很热门的课题。时间序列的匹配有两类方法: 一类是维数约简, 该类方法通过某些变换降低时间序列的维数, 或者是通过某些算法提取特征, 得到一个低维的特征空间, 然后再计算两个序列的距离; 一类是直接匹配方法, 最典型的动态时间弯折(Dynamic Time Warping, DTW)<sup>[1]</sup>, 该方法利用动态规划的原理, 将两个时间序列逐点匹配, 从而计算其距离。根据仿生模式识别中的连续性假设, 认为时间序列在同一类的样本空间呈某种流形分布, 根据一些标准对时间序列进行排序, 运用特定的插值算法, 在相邻的时间序列间, 插入新的参考样本作为已知样本, 最后用普通的最近邻方法对待分类的时间序列进行识别, 取得比较好的效果。

给定一个数据样本集合, 每个数据样本包括: 一个输入时间序列  $X_i = X_i^{(1)}, X_i^{(2)}, \dots$  及其离散的分类标签  $C_i$ , 其中  $X_i^{(t)} \in R^n$  是一个  $n$  维向量,  $C_i \in \{1, 2, \dots, N_c\}$ , 目标是预测新的时间序列的类标签。给定一个序列集合  $S = \{S_1, S_2, \dots, S_M\}$ , 其中每一个序列  $S_i$  由元素  $X_i(t) \in R^n (t=1, 2, \dots)$  组成, 而每一个元素  $X_i^{(t)}$  由一个属性值向量  $[X_i^{t,1}, X_i^{t,2}, \dots, X_i^{t,n}]$  组成, 时间序列分类的目标是得到测试集合  $S_{test}$  的分类结果, 使得该结果达到某个量度标准。

## 2 基于分类的时间序列数据变换

时间序列分类问题的输入是一些不等长的序列, 这使得大多数标准的分类方法<sup>[2]</sup>, 如  $k$ -近邻搜索、决策树、最大似然方法不能直接使用, 本文的方法首先将数据转化成等长向量, 利用一个基于模型的聚类方法来进行数据的转化, 建模利用序列前

**基金项目:** 新疆生产建设兵团工业科技攻关计划(Xinjiang Production and Construction Corps Industrial Technology Research Plans No.2007GG15); 塔里木大学校长基金青年资助项目(Tarim University Principal Youth Fund Grant No.TDZKQN05002)。

**作者简介:** 周鹏(1970-), 男, 副教授, 主要研究领域: 计算机应用技术, CCF 会员。

**收稿日期:** 2008-03-11 **修回日期:** 2008-05-26

后的依赖关系的特征,其次在经过转化的等长数据集上使用一般的分类算法作分类。

## 2.1 基于聚类模型的数据转换

算法首先根据已有的类标将训练集分为正例集和负例集,分别记为  $DB_+$  和  $DB_-$ , 然后再运用基于模型的聚类算法对正、负例集分别聚类,将正、负例集分别聚成  $p^+$  类和  $p^-$  类,然后将算法遍历所有的训练集,一次循环就是一条数据根据最大似然法得到它在每一个聚类中的度量,这个度量产生一系列的概率值,一个概率值与一个聚类对应,结果就是得到一条新的向量  $vector_i$ , 维数为  $p^++p^-$ , 然后将这些向量组合成一个向量集合  $Vectors$ , 这样就得到经过转换的新的数据集。

时间序列转换算法:

输入: 训练集  $DB$ , 正、负例分别所聚的类数  $p^+, p^-$

输出: 转换后的训练集向量

步骤:

- (1) 将训练集分为正例集  $DB_+$  和负例集  $DB_-$ ;
- (2) 正例的类=聚类( $DB_+, p^+$ ); 负例的类=聚类( $DB_-, p^-$ );
- (3)  $Model = (\text{正例的类}, \text{负例的类}); Vectors = \{ \}$ ;
- (4) For  $seq_i \in \text{TrainDB}$
- (5)  $Vector_i = \text{maxlikelihood}(seq_i, Model)$ ;
- (6)  $Vectors = Vectors \cup vector_i$ ;
- (7) end For
- (8) Return  $Vectors$ .

## 2.2 基于模型的聚类

任务首先是运用基于模型的聚类算法将数据统一转换为离散的状态,先将所有的数值特征离散化,可以通过一些标准的监督或非监督的离散化算法来实现,对数据集中的所有时间序列做离散化处理( $k$ -means 聚类算法来获得状态),得到  $Q$  个状态:  $S_i (i=1, 2, \dots, Q)$ , 其中  $Q$  是一个参数,可以通过调整得到恰当的值。从这些状态信息中可以看出,每个时间序列转换为一条状态转移链,运用基于 EM 的聚类算法分别得到正、负例集 ( $DB_+, DB_-$ ) 的聚类个数,即  $p^+, p^-$  的值。假设马尔可夫模型包含  $Q$  个状态,而在每个聚类中又有  $k$  条马尔可夫链,又假设先验状态分布  $\Pr(s_i) = a_i$ , 转移矩阵为:

$$v_p = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, T_p = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{bmatrix}$$

其中  $t_{ij}$  是表示从类  $s_i$  到类  $s_j$  的转移概率,那么就有

$$a_i = \frac{\sum_{t=1}^{t=k} I(C_t = i)}{K}, t_{ij} = \frac{\sum_{t=1}^{t=k} \sum_{s=1}^{s=i-1} I(C_t = i, C_{t(s+1)} = j)}{\sum_{t=1}^{t=k} \sum_{s=1}^{s=i-1} I(C_t = i)}$$

其中如果  $x$  为真,  $I(x) = 1$ , 否则  $I(x) = 0$ , 通过输入数据得到如下

形式的包含  $K$  个类的混合马尔可夫模型:  $p_k(\vdash|\theta) = \sum_{k=1}^k P(C_k|\theta)$

$P_k(\vdash|C_k, \theta)$ ,  $P(C_k|\theta)$  是指第  $k$  类的边沿概率并且  $\sum_{k=1}^k P(C_k|\theta) = 1$ ,

$P_k(\vdash|C_k, \theta)$  是指统计模型描述一条时间序列属于类  $k$  的概率,  $\theta$  表示模型的参数,  $v = v_1, v_2, \dots, v_L$  是一个随机长度的特征向量序列, 假设每个模型均是一阶马尔可夫模型, 可表示成:  $p_k(\vdash|C_k, \theta) =$

$P(\vdash|\theta'_k) \prod_{i=2}^L P(\vdash_i|v_{i-1}, \theta_k^T)$ , 其中  $\theta'_k$  和  $\theta_k^T$  均是概率分布的参数,

前一个概率分布是通过计算类  $k$  中所有时间序列的初始特征向量得到的, 后一个则是通过类  $k$  中一个特征向量到另一个特征向量的转换得到, 这个模型刻画时间序列中不同状态之间转换的特征, 包括他们的原始状态的特征和两个连续状态之间的转换关系。根据上述算法得到  $K = p^+ + p^-$  个聚类, 在文中取  $p^+ = 1$ , 对任意给定的时间序列, 都可以计算出这条序列在  $K$  个聚类中的概率分布, 然后用一个向量来表示  $V = (p_1, p_2, \dots, p_K)$ , 可以将所有的训练集和测试集转换为长度为  $K$  的向量, 这个概率度量将会用来对所有的测试集进行排序, 从而得到一条排好序的列表, 在实验中将会利用原有的类标和这个排好序的列表计算出 AUC 的值。在实际中可以尝试任意一个标准的分类算法, 例如  $k$ -近邻搜索、最大似然法等等, 对时间序列数据进行分类。首先将测试数据序列转换为状态向量形式:  $S_1, S_2, \dots, S_k$ , 进一步可以生成长度为  $K$  的向量  $V_i$ , 其中  $K$  表示聚类的个数, 这个向量  $V_i$  的每个元素就是这条序列属于不同聚类模型的概率, 这样把训练和测试数据转换为相同的向量, 根据这个转换后的向量, 可以用一般的分类算法进行分类。

## 3 时间序列匹配算法分析

### 3.1 动态时间弯折

动态规划方法是 20 世纪 50 年代美国的 Bellman 提出的, 它的优点在于只用较少的计算量就能给出最优结果, 将其用在时间序列的匹配上面即为动态时间弯折。设两个时间序列是  $A = a[1], a[2], \dots, a[m]$  和  $B = b[1], b[2], \dots, b[n]$  两个符号  $a[i]$  和  $b[j]$  的距离为  $d(i, j) = |a[i] - b[j]|$ , 则  $A$  和  $B$  之间总的匹配距离  $DIST(A, B)$  可由式(1)所示的动态规划的方法求出<sup>[3]</sup>:

$$\begin{cases} D(1, 1) = d(1, 1) \\ D(i, j) = d(i, j) + \min\{D(i-1, j), \\ D(i-1, j-1), D(i, j-1)\} \\ DIST(A, B) = d(1, 1) \end{cases} \quad (1)$$

$D(i, j)$  是  $A(i) = a[1], a[2], \dots, a[i]$  和  $B(j) = b[1], b[2], \dots, b[j]$  之间的匹配距离,  $DIST(A, B)$  越大, 序列之间的相似度就越小。

### 3.2 时间序列匹配算法和归一法

仿生模式识别认为, 同类样本在模式空间中是连续的, 因此需要首先对时间序列样本进行排序, 以获取样本间的连续性分布。提出的时间序列匹配的新算法主要包括如下的步骤: 首先将参考时间序列样本的长度归一化, 然后按照“正波能量中心”的数值进行排序<sup>[4]</sup>, 接着利用线性插值在相邻的时间序列间生成新的样本, 最后直接将两个序列相减, 得到两个序列的距离。

实际采集的同类时间序列样本, 其长度往往不一样, 因此需要先对其进行长度归一化, 设某一类时间序列的参考样本个数为  $n$ , 其长度分别为  $l[i], i=1, 2, \dots, n$ , 则归一化后的长度为:

$$\tilde{l} = \frac{1}{n} \sum_{k=1}^n l[k] \quad (2)$$

若某个序列  $C = c[1], c[2], \dots, c[m]$  归一化后的序列为  $\tilde{C} = \tilde{C}[1], \tilde{C}[2], \dots, \tilde{C}[n]$ , 利用线性插值, 则归一化后序列元素的计算公式为式(3):

$$\tilde{C}[k]=c[\text{int}(\max(\frac{mk}{n}))](\text{int}(\min(\frac{mk}{n}))- \frac{mk}{n})+ \\ c[\text{int}(\min(\frac{mk}{n}))](\frac{mk}{n}-\text{int}(\max(\frac{mk}{n}))) \quad (3)$$

$\text{int}(\max(\frac{mk}{n}))$ 表示不超过  $a$  的最大整数,  $\text{int}(\min(\frac{mk}{n}))$ 表示不小于  $a$  的最小整数。

### 3.3 时间序列排序和增加样本点

输入: 归一化后的时间序列  $c[i], i=1, 2, 3, \dots, n$

输出: 正波能量中心 Positive Gravity, ( $1 \leq \text{PositiveGravity} \leq n$ )

```
PositiveGravity=0;
Sum=0;
For k=1 to n
  If c[i]>0
    Sum=Sum+c[i]*c[i]
    PositiveGravity=PositiveGravity+k*c[i]*c[i]
  End
End
PositiveGravity=PositiveGravity/Sum
```

将同一类中的时间序列首先进行长度归一化, 按照正波能量中心的值由小到大的顺序进行排序, 将算法中的 If  $c[i]>0$  改成 If  $c[i]<0$ , 即可得到计算负波能量中心的方法<sup>[5]</sup>。

设相邻的两个时间序列为  $A=a[1], a[2], \dots, a[n]$  和  $B=b[1], b[2], \dots, b[n]$ , 要在其中插入  $k$  个参考样本, 第  $i$  个新样本的生成算法流程。

步骤 1 计算  $A, B$  的 PG 和 NG, 记  $a_{pg}, a_{ng}, b_{pg}, b_{ng}$ ;

步骤 2 对于第  $i$  个新样本  $C=c[1], c[2], \dots, c[n]$ , 其 PG 和 NG 计算公式见式(4):

$$\text{new\_pg} = ((k+1-i) \times a_{pg} + i \times b_{pg}) / (k+1) \\ \text{new\_ng} = ((k+1-i) \times a_{ng} + i \times b_{ng}) / (k+1) \quad (4)$$

步骤 3 基于线性插值, 计算样本中第  $m$  个数据的值。

两个时间序列  $A=a[1], a[2], \dots, a[n]$  和  $B=b[1], b[2], \dots, b[n]$  的距离可以用下式直接计算:

$$\text{DIST}(A, B) = \sum_{k=1}^n |a[k] - b[k]| / n$$

假设有  $N$  个参考时间序列样本  $T_i, i=1, 2, \dots, N$ , 测试样本为  $T$ , 如果  $k = \text{argminDIST}(T, T_i)$ , 即  $T$  与  $T_k (k \in \{1, 2, \dots, N\})$  的距离最小, 则  $T$  与  $T_k$  的距离最小。

## 4 实验与分析

合成的是一个两类的时间序列分类的数据集, 每一类由一个随机产生的隐马尔可夫模型产生 1 000 条序列, 隐马尔可夫模型隐层状态数量为 5, 隐层节点到观测节点由高斯分布产生。把每一类的 500 条序列用作训练样本, 其余的作为测试数据, 即训练和测试数据各包含 1 000 条序列。在第 1 组实验中变换训练数据的数量, 从 50~1 000, 测试 3 个算法在测试数据上的错误率, 从图 1 中发现两个算法 MC- $k$ -NN 与 MC-MLE 在这个数据集上比 DTW 算法的错误率小; 另外 MC- $k$ -NN 与 MC-MLE 算法受训练样本影响不大, 在训练样本很少的情况下能够得到理想的效果, 相比之下 DTW 算法随着训练样本减少, 误差上升较快。

在第 2 组实验中变换每条时间序列的长度范围是 50~

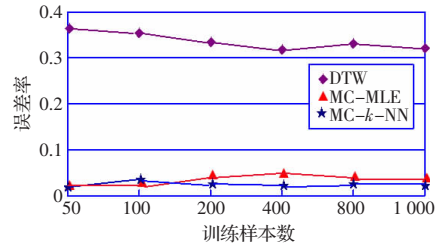


图1 训练样本大小对算法的影响

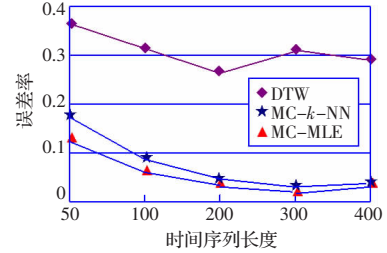


图2 序列长度对算法的影响

400, 训练样本与测试样本大小均为 1 000。从图 2 可以看到 MC- $k$ -NN 与 MC-MLE 算法的错误率在不同的序列长度上都比 DTW 算法小; 随着序列长度的增加, 3 个算法的错误率总体上都往下降, 并且 MC- $k$ -NN 与 MC-MLE 算法与 DTW 算法的错误率的差距增大。

在第 3 组实验中改变隐马尔可夫模型中高斯的标准方差范围为 0.125~2, 这相当于在数据中加入不同的噪声。在这组实验中训练样本与测试样本大小还是 1 000, 实验结果如图 3, 可以看出虽然 MC- $k$ -NN 与 MC-MLE 算法的错误率比 DTW 算法小, 但随着噪声增加, MC- $k$ -NN 与 MC-MLE 算法的误差越来越接近 DTW 的误差。

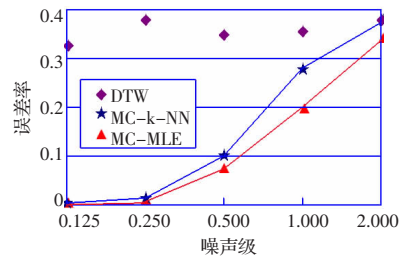


图3 噪声的影响

在第 4 组实验中采集管道中 30 个不同点, 每个点不同时间泄漏数据 50 个, 这样共得到 1 500 个时间序列样本, 用交叉验证的方法来设计实验: 将每个点的数据随机分成五组, 编号 1~5, 每组 10 个时间序列。第一次实验选择每个点的第一组数据作为参考序列, 其他 4 组数据作为测试序列; 第二次实验选择每个点的第二组数据作为参考序列, 其他 4 组测试, 以此类推总共进行 5 组实验。每次实验时, 首先将每个点的参考序列进行长度归一化, 然后利用算法进行排序和增加新的参考样本, 每两个相邻序列间新增的参考样本的个数的取值分别为 0, 2, 4, 6, 8, 10, 12, 15, 20, 30, 最后根据最近邻规则, 利用定义的距离来进行分类, 实验结果见表 1。

## 5 结束语

时间序列分类是时间序列数据分析中的重要任务之一, 它比一般分类问题困难的原因主要在于要分类的时间序列数据



表 1 仿生模式识别-识别错误率

(%)

组别	1		2		3		4		5		平均	
	排序	不排	排序	不排	排序	不排	排序	不排	排序	不排	排序	不排
K=0	20.8	25.2	20.8	25.3	20.5	22.3	20.6	23.5	20.7	24.7	20.7	24.2
K=2	16.4	19.9	16.5	20.0	16.2	18.1	16.2	18.7	16.4	19.7	16.3	19.3
K=4	12.7	16.0	12.7	15.9	12.4	13.8	12.5	14.5	12.6	15.0	12.6	15.0
K=6	9.8	12.9	9.8	12.8	9.6	11.2	9.6	11.4	9.7	11.7	9.7	12
K=8	8.0	11.0	7.9	11.2	7.7	9.9	7.8	10.3	7.9	10.5	7.9	10.6
K=10	6.7	9.6	6.7	9.6	6.5	8.2	6.6	8.5	6.6	9.3	6.6	9.0
K=12	6.3	9.2	6.3	9.5	6.2	8.0	6.2	8.3	6.3	9.2	6.3	8.8
K=15	6.1	9.1	6.1	9.3	6.0	7.7	6.1	8.0	6.1	8.9	6.1	8.6
K=20	6.0	9.0	6.0	9.3	6.0	7.7	6.0	8.0	6.0	8.8	6.0	8.6
K=30	6.0	9.0	6.0	9.2	5.9	7.7	6.0	7.9	6.0	8.8	6.0	8.2

不等长,这使得一般的分类算法不能直接应用。即使是等长的时间序列,由于不同序列在相同位置的数值一般不可直接比较,依然还不适合直接应用一般的分类算法。本文在不同的合成数据集和实际数据集上比较领域无关和领域相关的两类方法,发现在训练数据较少时,使用领域相关的算法比较合适,领域无关的算法受噪声的影响相对较少。本文提出一种新的时间序列匹配算法进行研究,利用“正波能量中心”对时间序列进行排序,并针对在管道泄漏点的数据集上与传统动态时间弯折算法的对比表明,新的匹配算法具有良好的分类性能,对于复杂的时间序列如何排序和怎样插入新的样本点进行的研究,利用领域知识的基于模型聚类的方法有较好的 AUC 性能。随着我国西部大开发及西气东输工程的实施,将有大量的油气管道投入建设和运行,为将泄漏事故造成的各种危害减少到最小,需要研究泄漏在线监测技术,管道流量泄漏时间序列分类算法的研究对提高泄漏监测灵敏度和定位精度具有直接影响和决定作用,本设计在线监测系统已在开封仪表有限公司已经投入生产,每年带来经济效益达 1 500 万元,完全符合工业现场要求,具

有很大的实际意义和应用价值。

### 参考文献:

- [1] Aach J, Church G. Aligning gene expression time series with time warping algorithms[J]. *Bioinformatics*, 2001, 17: 495-508.
- [2] Abou-Moustafa K T, Cheriet M, Suen C Y. A generative discriminative hybrid for sequential data classification[C]// *Proceedings of the IEEE International Conference on Acoustics and Signal Processing*. Montreal, 2004: 805-808.
- [3] Ling C X, Huang J, Zhang H. AUC: A statistically consistent and more discriminating measure than accuracy[C]// *Proceedings of the 18th International Joint Conference on Artificial Intelligence*. Acapulco, Mexico, 2003: 329-341.
- [4] 安冬, 王库, 王守觉. 高维空间点覆盖方法在物种计算机自动分类中的应用[J]. *电子学报*, 2006, 34(2): 277-281.
- [5] Wang Shou-jue, Lai Jiang-liang. Geometrical learning, descriptive geometry, and bio-mimetic pattern recognition[J]. *Neuro Computing*, 2005, 67: 9-28.
- [6] architecture for mobile Ad hoc networks[J]. *Journal of Electronics*, 2006, 23(3): 417-422.
- [7] Subhadrabandhu D, Sarkar S, Anjum F. A framework for misuse detection in Ad hoc networks-Part II[J]. *IEEE Journal on Selected Areas in Communications Special Issue on Security in Wireless Ad Hoc Networks*, 2006.
- [8] Anjum F, Talpade R. Packet-drop detection algorithm for Ad hoc networks[C]// *Proceedings of 60th IEEE Vehicular Technology Conference*, Sept. 2004.
- [9] architecture [J]. *Applied Mathematics and Computation* 2003, 136: 315-321.
- [10] Hwang M S, Lin I C, Eric J L Lu. A secure nonrepudiable threshold proxy signature scheme with known signers[J]. *International Journal of Informatics*, 2000, 11: 1-8.
- [11] Sun H M. An efficient nonrepudiable threshold proxy signature scheme with known signers [J]. *Computer Communications*, 1999, 22(8): 717-722.
- [12] Sun H M, Lee N Y, Hwang T. Threshold proxy signatures[J]. *IEEE Proceedings-Computers and Digital Techniques*, 1999, 146(5): 259-263.
- [13] Tzeng S F, Yang C Y, Hwang M S. A nonrepudiable threshold multi-proxy multi-signature scheme with shared verification[J]. *Future Generation Computer Systems*, 2004, 20(5): 887-893.

(上接 118 页)

- [17] Rajavaram, Sowjanya, Hiren Shah. Neighborhood watch: An intrusion detection and response protocol for mobile Ad-hoc networks[R/OL]. <http://userpages.umbc.edu/~hiren1/mobile/conferencecp.ps>.
- [18] Subhadrabandhu D, Sarkar S, Anjum F. A framework for misuse detection in Ad hoc networks-Part I[J]. *IEEE Journal on Selected Areas in Communications Special Issue on Security in Wireless Ad Hoc Networks*, 2006.
- [19] Yi Ping, Zhong Yiping, Zhang Shiyong. An immunity based security

(上接 129 页)

案的特殊形式。所以有代理的方案更具有广泛的实用性。在门限代理签名中加入有代理的思想,使得该方案不仅具有代理和门限的特点,更实现了代理签名者与原始签名者的联合签名,使得该方案可以在更广的范围内得到应用。

### 参考文献:

- [1] Mambo M, Usuda K, Okamoto E. Proxy signatures: delegation of the power to sign message[J]. *EICE Transactions on Fundamentals E79-A*, 1996(9): 1338-1354.
- [2] Huang H F, Chang C C. A novel efficient( $t, n$ ) threshold proxy signature scheme[J]. *Information Sciences*, 2006, 176: 1338-1349.
- [3] Hsu C L, Wu T S, Wu T C. Improvement of threshold proxy signa-