

基于移相加权球面单簇聚类的周期时间序列异常检测

王 骏^{1,2} 钟富礼² 王士同¹ 邓赵红^{1,2}

摘要 针对传统的单分类器不适用于周期时间序列的异常检测, 提出了一种基于移相加权球面单簇聚类的单分类器 PS-WS1M-OCC. 通过在聚类过程中增加高效的循环移位操作, 解决了时间序列记录之间相似度计算的问题. 另一方面, 基于时间序列记录的权重分布, 提出了新的阈值自适应确定方法, 从而使单分类器对训练集包含的异常数据和参数设置不敏感. 实验表明, 本文提出的单分类器可以用于周期时间序列的异常检测; 与传统的单分类器相比, 可以成功地从包含异常数据的训练集中进行无监督学习, 对训练集包含的异常数据鲁棒, 并且对参数不敏感.

关键词 移相加权球面单簇聚类, 时间序列异常检测, 单分类器, 从包含噪声的数据中学习

DOI 10.3724/SP.J.1004.2011.00984

A Novel Discords Detector for Periodic Time Series Based on Weighted Spherical Single Means with Phase Shift

WANG Jun^{1,2} CHUNG Fu-Lai² WANG Shi-Tong³ DENG Zhao-Hong^{1,2}

Abstract The traditional one-class classifiers are not suitable for detecting discords in periodic time series. A novel one-class classifier PS-WS1M-OCC is proposed in this paper. In our method, the phase problem in time series is solved by introducing phase shift into the clustering procedure. Meanwhile, a novel criterion for adaptively choosing threshold is proposed. In this way, the proposed classifier is insensitive to noise in the training set. Experimental results show that our PS-WSKM-OCC is more robust than the existing one-class classifiers when it is applied to the problem of discord detection in the periodic time series.

Key words Weighted spherical single means with phase shift, discord detection, one-class classifier, learning from noise data

周期时间序列的异常检测是时间序列分析的重要研究内容, 其主要任务是检测周期时间序列中存在显著差异的、具有异常行为的区域 (如图 1 所示)^[1-5]. 目前经常使用的方法是, 使用滑动窗口技术对单条周期时间序列进行分割, 得到一系列等长或不等长的时间序列记录 (时间子序列), 生成时间序列数据集. 然后, 设计相应的算法检测其中的异常记录, 原时间序列中与之相对应的区域即为异常区域. 从数据挖掘的角度来看, 可以将时间序列数据集里的记录作为高维空间中的数据点来处理. 这样, 周期时间序列异常检测问题就转化为高维空间中的例外点检测问题来处理. 令人遗憾的是, 传统的单分类

器不能直接用来解决这类问题.

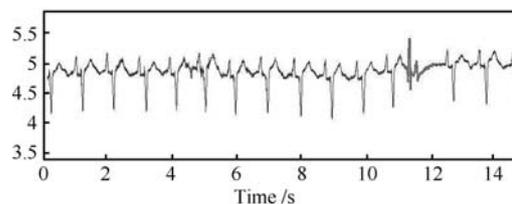


图 1 心电图时间序列及异常区域

Fig. 1 A human electrocardiogram with abnormal region

主要原因有:

1) 由于各时间序列记录之间存在着不同步现象, 传统的单分类器所使用的相似度计算方法不再适用. 例如, 我们使用滑动窗口技术分割某周期时间序列后, 得到如图 2 所示的两条记录. 由于分割时间序列时仅要求滑动窗口的宽度为周期 T , 各滑动窗口的起始位置均匀间隔, 所以图 2 中的时间序列记录并不是同步的, 这为计算它们之间的相似度带来困难. 解决此问题的一种方法是在预处理阶段对时间序列记录在水平方向进行移位, 使最大或最小值点的位置对齐. 但是, 当时间序列记录中存在若干个相等的最大或最小值点时, 这种方法将会失效.

收稿日期 2010-09-08 录用日期 2010-12-27
Manuscript received September 8, 2010; accepted December 27, 2010

国家自然科学基金 (90820002, 60903100), 江苏省自然科学基金 (BK2009067), 中央高校基本科研业务费专项基金 (JUSRP21128) 资助

Supported by National Natural Science Foundation of China (90820002, 60903100), Natural Science Foundation of Jiangsu Province (BK2009067), and the Fundamental Research Funds for the Central Universities (JUSRP21128)

1. 江南大学数字媒体学院 无锡 214122 2. 香港理工大学电子计算学系 香港

1. School of Digital Media, Jiangnan University, Wuxi 214122
2. Department of Computing, Hong Kong Polytechnic University, Hong Kong

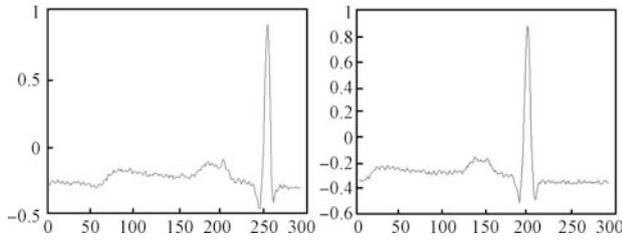


图2 周期时间序列的不同步现象 (由于误差累积, 根据周期划分生成的时间序列记录不严格同步)

Fig. 2 The unsynchronized phases of periodic time series (Because of error cumulation, the phases of the periodic time series are not synchronized)

2) 对周期时间序列进行分割后产生的数据集里包含异常记录, 由于选取的随意性, 训练集里不可避免地也会包含相当数量的异常记录. 可是, 传统的单分类器仅使用正常数据作为训练集, 它们依赖于用户对参数的合理设置来降低对训练集包含的异常数据的敏感性^[6]. 这就需要用户准确认识训练集里异常数据的组成情况. 不幸的是, 这往往很难实现. 最近, 学术界致力于研究如何从包含噪声的数据中进行学习, 通过异常数据来增强单分类器的判别能力^[7-8]. 文献 [7] 提出了训练集里带少量噪声的异常检测算法 (Novelty detection with few outlier training data, NDFO), 实现了从包含少量噪声的训练集里进行学习. 但是, 这是一个有监督的学习过程, 需要用户对训练集的样本进行类别标记, 这本身就会带来相当大的工作量.

基于聚类的单分类器设计方法为我们提供了新的研究思路. 其技术路线可概括为, 首先, 通过聚类算法得到聚类中心 (原型); 然后, 基于某一距离度量设定阈值, 形成关于目标类的数据描述. 典型的例子有 K 均值单分类器^[9]、 K 中心单分类器^[10]、基于单簇聚类的数据描述^[11] 等.

针对传统单分类器的不足, 本文基于单簇聚类的思想, 提出了一种适用于周期时间序列异常检测任务的新型单分类器 PS-WS1M-OCC (One-class classifier base on weighted spherical single means with phase shift). 与已有的单分类器相比, PS-WS1M-OCC 的特点在于:

1) 将时间序列记录之间的相似度计算与傅里叶变换建立关系, 通过在聚类过程中增加高效的循环移位操作, 解决了时间序列记录之间的同步问题以及由此引发的相似度计算问题;

2) 为参与聚类的时间序列记录引入权重信息, 使算法的训练过程对训练集包含的异常数据有很高的鲁棒性;

3) 基于单簇聚类得到的异常值, 提出新的阈值确定准则, 使算法能够自动地选择阈值. 因此, 单分

类器 PS-WS1M-OCC 在训练过程中具有自动识别目标类而拒绝异常数据的能力, 这是其他单分类器所欠缺的.

本文结构如下: 第 1 节给出了单位向量的若干性质, 这为本文算法提供了理论基础; 第 2 节提出了移相加权球面单簇聚类算法, 证明其收敛性, 并讨论了参数的选取; 第 3 节给出了基于 PS-WS1M 聚类的单分类器 PS-WS1M-OCC, 并提出一种确定相似度阈值的新方法; 第 4 节基于周期时间序列数据集, 将 PS-WS1M-OCC 和其他常用的单分类器进行比较, 给出实验结果与分析; 第 5 节给出结论.

1 单位向量的若干性质

定义 1. 设 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 为 s 维单位向量, 其权重为 w_1, \dots, w_n , 满足 $w_i > 0, i = 1, 2, \dots, n$. 则定义加权均值向量如下:

$$\mathbf{m} = \frac{\sum_{i=1}^n w_i \mathbf{x}_i}{\sum_{i=1}^n w_i} \quad (1)$$

将 \mathbf{m} 化为单位向量, 定义单位加权均值向量 \mathbf{c} 如下:

$$\mathbf{c} = \frac{\mathbf{m}}{\|\mathbf{m}\|} \quad (2)$$

定义 2. 假设 \mathbf{x} 为 s 维单位向量, \mathbf{x} 循环移动 τ 个位置后记为 $\mathbf{x}^{(\tau)}$, 满足:

$$\mathbf{x}^{(\tau)} = \mathbf{x}((i + \tau))_s, \quad i = 1, 2, \dots, s$$

其中, 符号 $((\cdot))_s$ 表示对 s 取余操作.

定义 3. 假设 \mathbf{x} 和 \mathbf{c} 为 s 维单位向量, \mathbf{x} 循环移动 τ 个位置后所得向量记为 $\mathbf{x}(\tau)$. 则定义 \mathbf{x} 相对于 \mathbf{c} 的最佳位移 $\tau_{\text{opt}}(\mathbf{x}, \mathbf{c})$ 如下:

$$\tau(\mathbf{x}, \mathbf{c}) = \arg \max_{\tau} \mathbf{c}^T \mathbf{x}^{(\tau)} \quad (3)$$

定理 1. 设 \mathbf{x}_i 为 \mathbf{R}^s 空间中的单位向量, w_i 为非负实数, $i = 1, 2, \dots, n$, \mathbf{c} 为单位向量, 且由式 (1) 和式 (2) 计算而得. 对于任何 \mathbf{R}^s 空间中的单位向量 \mathbf{z} , 如下不等式成立:

$$\sum_{i=1}^n w_i \mathbf{x}_i^T \mathbf{z} \leq \sum_{i=1}^n w_i \mathbf{x}_i^T \mathbf{c} \quad (4)$$

证明. 根据 Cauchy-Schwarz 不等式, 有:

$$\begin{aligned} \sum_{i=1}^n w_i \mathbf{x}_i^T \mathbf{z} &= \mathbf{z}^T \sum_{i=1}^n w_i \mathbf{x}_i \leq \\ &\|\mathbf{z}^T\| \cdot \left\| \sum_{i=1}^n w_i \mathbf{x}_i \right\| = \left\| \sum_{i=1}^n w_i \mathbf{x}_i \right\| \end{aligned}$$

另一方面:

$$\begin{aligned} \sum_{i=1}^n w_i \mathbf{x}_i^T \mathbf{c} &= \mathbf{c}^T \sum_{i=1}^n w_i \mathbf{x}_i = \\ \mathbf{c}^T \cdot \mathbf{m} \sum_{i=1}^n w_i &= \mathbf{c}^T \cdot \mathbf{c} \|\mathbf{m}\| \sum_{i=1}^n w_i = \\ \left\| \mathbf{m} \sum_{i=1}^n w_i \right\| &= \left\| \sum_{i=1}^n w_i \mathbf{x}_i \right\| \end{aligned} \quad (5)$$

因此, 不等式 (4) 成立. \square

定理 2. 假设 \mathbf{x} 和 \mathbf{c} 为 s 维单位向量, 则计算 \mathbf{x} 相对于 \mathbf{c} 的最佳位移 $\tau_{\text{opt}}(\mathbf{x}, \mathbf{c})$ 的时间复杂度为 $O(s \log s)$.

证明. 为了计算 \mathbf{x} 相对于 \mathbf{c} 的最佳位移 $\tau_{\text{opt}}(\mathbf{x}, \mathbf{c})$, 根据式 (3), 需要逐个计算所有 $\mathbf{x}(\tau)$ 和 \mathbf{c} 之间的相似度, $\tau = 1, 2, \dots, s$. 这等价于计算 \mathbf{x} 与 \mathbf{c} 的 s 点圆周卷积. 根据卷积定理, 序列傅里叶变换的乘积等于它们卷积后的傅里叶变换. 因此, 可以通过计算其相应的傅里叶变换的乘积的反变换来间接地求得 \mathbf{x} 与 \mathbf{c} 的圆周卷积. 令 F 表示傅里叶算子, F^{-1} 为反傅里叶算子, $F\{\mathbf{x}\}$ 和 $F\{\mathbf{c}\}$ 表示单位向量 \mathbf{x} 和 \mathbf{c} 的傅里叶变换. 则对 $\tau = 1, 2, \dots, s$, $\mathbf{x}(\tau)$ 和 \mathbf{c} 之间的相似度 $r_{\mathbf{x}, \mathbf{c}}(\tau)$ 计算如下:

$$r_{\mathbf{x}, \mathbf{c}}(\tau) = F^{-1}\{F\{\mathbf{x}\} \cdot F\{\mathbf{c}\}\}(\tau)$$

通过快速傅里叶变换 FFT, 傅里叶变换和反傅里叶变换可快速求得. 所以, 不同位移 $r_{\mathbf{x}, \mathbf{c}}(\tau)$ 可以同时计算, 其时间复杂度与快速傅里叶变换相同, 均为 $O(s \log s)$, 得到其最佳位 $\tau_{\text{opt}}(\mathbf{x}, \mathbf{c})$ 的时间复杂度亦为 $O(s \log s)$. \square

2 移相加权球面单簇聚类算法 PS-WS1M

2.1 算法推导及描述

由于时间序列是高维数据, 因此, 本文采用余弦相似度来计算时间序列记录之间的相似程度, 以避免欧氏距离计算公式对高维数据失效的问题. 但是, 时间序列记录之间存在的不同步现象, 导致了直接通过点积运算来计算相似度不能反映它们之间的差异. 为了使聚类过程中相似度的计算有实际意义, 本文在聚类过程中引入循环移位这一操作, 提出移相加权球面单簇聚类算法 PS-WS1M (Weighted spherical single means with phase shift).

令 $X = \{\mathbf{x}_i\}_{i=1}^n$ 为时间序列数据集, \mathbf{x}_i 为其中的时间序列记录, 且满足 $\|\mathbf{x}_i\| = 1, i = 1, \dots, n$, \mathbf{c} 为聚类中心, $\|\mathbf{c}\| = 1, \mathbf{B} = (\beta_1, \beta_2, \dots, \beta_n)$ 为权重

向量, $-1 < r < 0$ 为模糊指数. 构造目标函数如下:

$$J(X, \mathbf{c}, \mathbf{B}) = \sum_{i=1}^n \frac{1}{\beta_i^r} \mathbf{x}_i^T \mathbf{c} + \varepsilon \sum_{i=1}^n \frac{1}{\beta_i^r} \quad (6a)$$

$$\beta_i > 0, \quad \sum_{i=1}^n \beta_i = 1 \quad (6b)$$

PS-WS1M 算法搜索合适的聚类中心 \mathbf{c} 和权重向量 \mathbf{B} , 从而使式 (6a) 定义的目标函数 $J(X, \mathbf{c}, \mathbf{B})$ 在约束条件 (6b) 下最大化. 以上问题可分解为以下三个子问题:

问题 P_1 : 固定 $\mathbf{B} = \hat{\mathbf{B}}$ 和 $X = \hat{X}$, 求解问题 $J(\hat{X}, \mathbf{c}, \hat{\mathbf{B}})$.

问题 P_2 : 固定 $\mathbf{c} = \hat{\mathbf{c}}$ 和 $X = \hat{X}$, 求解问题 $J(\hat{X}, \hat{\mathbf{c}}, \mathbf{B})$.

问题 P_3 : 固定 $\mathbf{B} = \hat{\mathbf{B}}$ 和 $\mathbf{c} = \hat{\mathbf{c}}$, 求解问题 $J(X, \hat{\mathbf{c}}, \hat{\mathbf{B}})$.

由定理 1, 易得 $\mathbf{c} = \mathbf{m}/\|\mathbf{m}\|$, $\mathbf{m} = \sum_{i=1}^n \frac{\mathbf{x}_i}{\beta_i^r} / \sum_{i=1}^n \frac{1}{\beta_i^r}$ 为问题 P_1 的解.

使用拉格朗日法求解问题 P_2 , 构造约束条件 (6b) 下的拉格朗日函数如下:

$$L(\mathbf{B}, \lambda) = J(\hat{X}, \hat{\mathbf{c}}, \mathbf{B}) - \lambda \left(\sum_{i=1}^n \beta_i - 1 \right)$$

求导, 可得:

$$\frac{\partial L(\mathbf{B}, \lambda)}{\partial \mathbf{B}_i} = -r \frac{1}{\beta_i^{r+1}} \mathbf{x}_i^T \mathbf{c} - r \varepsilon \frac{1}{\beta_i^{r+1}} - \lambda = 0$$

且

$$\frac{\partial L(\mathbf{B}, \lambda)}{\partial \lambda} = \sum_{i=1}^n \beta_i - 1 = 0$$

易得:

$$\beta_i = \frac{(\mathbf{x}_i^T \mathbf{c} + \varepsilon)^{\frac{1}{\tau+1}}}{\sum_{i=1}^n (\mathbf{x}_i^T \mathbf{c} + \varepsilon)^{\frac{1}{\tau+1}}} \quad (7)$$

由于问题 P_3 在 \mathbf{R}^s 空间中是非连续的, 因此, 不能采用与问题 P_2 相同的方法求解其局部极值. 本文通过平移法进行求解. 即对于每一条时间序列记录 \mathbf{x}_i , 计算其相对于聚类中心 \mathbf{c} 的最佳位移 τ , 对其进行循环移位, 得到新的时间序列 \mathbf{x}_i^τ . 此时, 对任意位移, $(\mathbf{x}_i^\tau)^T \mathbf{c}$ 最大化. 相应的, 当 $\hat{\mathbf{B}}$ 和 $\hat{\mathbf{c}}$ 为定值时, 目标函数

$$J(X, \mathbf{c}, \hat{\mathbf{B}}) = \sum_{i=1}^n \frac{1}{\hat{\beta}_i^r} ((\mathbf{x}_i^\tau)^T \hat{\mathbf{c}} + \varepsilon) \quad (8)$$

亦最大化. 现在得到移相加权球面单簇聚类 PS-WS1M 算法描述如下:

输入: 时间序列数据集 $X = \{\mathbf{x}_i\}_{i=1}^n$, $\|\mathbf{x}_i\| = 1$, $i = 1, 2, \dots, n$

输出: 时间序列记录的权重向量 \mathbf{B} 和聚类中心

\mathbf{c} .

算法流程:

步骤 1. 随机生成时间序列记录的权重向量 $\mathbf{B}^{(0)}$ 且满足式 (6b), 设置迭代次数 $t = 0$;

步骤 2. 令 $\hat{\mathbf{B}} = \mathbf{B}^{(t)}$ 和 $\hat{X} = X^{(t)}$, 在数据集 $X^{(t)}$ 上求解问题 P_1 , 通过式 (1) 和式 (2) 得到 $\mathbf{c}^{(t+1)}$;

步骤 3. 令 $\hat{X} = X^{(t)}$ 和 $\hat{\mathbf{c}} = \mathbf{c}^{(t+1)}$, 在数据集 $X^{(t)}$ 上求解问题 P_2 , 使用式 (7) 计算 $\mathbf{B}^{(t+1)}$;

步骤 4. 令 $\hat{\mathbf{B}} = \mathbf{B}^{(t+1)}$ 和 $\hat{\mathbf{c}} = \mathbf{c}^{(t+1)}$, 对于数据集 $X^{(t)}$ 中的每个时间序列记录, 计算 $\mathbf{x}_i^{(t)}$ 相对于 $\hat{\mathbf{c}}$ 的最佳位移 $\tau_{\text{opt}}(\mathbf{x}_i^{(t)}, \hat{\mathbf{c}})$, 并对 $\mathbf{x}_i^{(t)}$ 根据 $\tau_{\text{opt}}(\mathbf{x}_i^{(t)}, \hat{\mathbf{c}})$ 移位得到 $\mathbf{x}_i^{(t+1)}$, 生成新数据集 $X^{(t+1)} = \{\mathbf{x}_i^{(t+1)}\}_{i=1}^n$;

步骤 5. 根据式 (6a) 计算 $J(X^{(t+1)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t+1)})$. 如果 $|J(X^{(t+1)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t+1)}) - J(X^{(t)}, \mathbf{c}^{(t)}, \mathbf{B}^{(t)})| < \varepsilon$, 输出 $(X^{(t+1)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t+1)})$, 结束. 否则, $t = t + 1$, 转步骤 2.

传统的聚类算法在执行过程中, 参与聚类的数据始终不发生变化. 但是对于 PS-WS1M, 由于步骤 4 的引入, 算法在执行过程中, 参与聚类的数据不断变化, 从而解决了时间序列记录的相位同步问题, 这是 PS-WS1M 与传统的聚类算法的不同之处.

另一方面, 对于两长度为 s 的时间序列记录 \mathbf{x} 和 \mathbf{c} , 如果采用穷举法计算 \mathbf{x} 相对于 \mathbf{c} 的最大相似度及对应位移, 其时间复杂度为 s^2 . 而定理 2 的引入, 提供了一种高效计算 \mathbf{x} 相对于 \mathbf{c} 的最大相似度及对应位移的新方法, 其时间复杂度也缩减为 $O(s \log s)$, 这对有效提高算法效率有着重要意义.

2.2 PS-WS1M 的收敛性

在 PS-WS1M 算法的执行过程中, 步骤 4 的加入使参与聚类的时间序列记录发生循环移位, 这是否会破坏算法的收敛性呢? 下面对其进行分析.

定理 3. 设集合 M_β 定义如下:

$$M_\beta = \left\{ \mathbf{B} \in \mathbf{R}^n \mid \beta_i \in [0, 1], \sum_{i=1}^n \beta_i = 1 \right\} \quad (9)$$

映射 $\Phi: M_\beta \rightarrow \mathbf{R}$ 定义如下:

$$\Phi(\mathbf{B}) = J(\hat{X}, \hat{\mathbf{c}}, \mathbf{B}) = \sum_{i=1}^n \frac{1}{\beta_i^r} (\hat{\mathbf{x}}_i^T \hat{\mathbf{c}} + \varepsilon) \quad (10)$$

其中, $\hat{X} \in \mathbf{R}^{ns}$, $\hat{\mathbf{c}} \in \mathbf{R}^s$, $\varepsilon > -\min \mathbf{x}_i^T \mathbf{c}$, $i = 1, 2, \dots, n$, 且 $-1 < r < 0$ 为定值. 则 \mathbf{B} 为 Φ 在 M_β 的全局最大化, 当且仅当 \mathbf{B} 由式 (7) 计算而得.

证明. 根据以上推导, 必要性易证. 下面证充分性. 设 $H(\mathbf{B})$ 为 $J(\hat{X}, \hat{\mathbf{c}}, \mathbf{B})$ 关于 \mathbf{B} 的海森矩阵, 则:

$$h_{i,j}(\mathbf{B}) = \frac{\partial}{\partial \beta_i} \left(\frac{\partial \phi(\mathbf{B})}{\partial \beta_i} \right) = \begin{cases} r(r+1)\beta_i^{-(r+2)}(\mathbf{x}_i^T \mathbf{c} + \varepsilon), & i = j \\ 0, & \text{其他} \end{cases}$$

显然 $H(\mathbf{B})$ 为对角阵. 由于 $-1 < r < 0$, β_i 由式 (7) 计算而得, 且 $\mathbf{x}_i^T \mathbf{c} + \varepsilon > 0$, 所以 $r(r+1)\beta_i - (r+2)(\mathbf{x}_i^T \mathbf{c} + \varepsilon) < 0$ 成立, 且海森矩阵 $H(\mathbf{B})$ 负定. 因此, 式 (7) 为 $J(\hat{X}, \hat{\mathbf{c}}, \mathbf{B})$ 最大化的充分条件. 因此, 证明了 \mathbf{B} 为局部极大值点当且仅当其由式 (7) 计算而得.

另一方面, 当 $-1 < r < 0$ 时 $\Phi(\mathbf{B})$ 为严格凹函数, 因此, 最多只有一个最大值点. \square

定理 4. 假设 $t \geq 0$ 为迭代次数, 则以下不等式成立:

$$J(X^{(t)}, \mathbf{c}^{(t)}, \mathbf{B}^{(t)}) \leq J(X^{(t+1)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t+1)})$$

证明. 从以上算法易得, 对于迭代次数 $t \geq 0$, $\mathbf{c}^{(t+1)}$ 为 $X^{(t)}$ 的加权平均值, 根据定理 1, 有:

$$\sum_{i=1}^n (\beta_i^{(t)})^{-r} (\mathbf{x}_i^{(t)})^T \mathbf{c}^{(t)} \leq \sum_{i=1}^n (\beta_i^{(t)})^{-r} (\mathbf{x}_i^{(t)})^T \mathbf{c}^{(t+1)}$$

因此, 如下不等式成立:

$$J(X^{(t)}, \mathbf{c}^{(t)}, \mathbf{B}^{(t)}) \leq J(X^{(t)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t)}) \quad (11a)$$

根据定理 3, 易得:

$$J(X^{(t)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t)}) \leq J(X^{(t)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t+1)}) \quad (11b)$$

另一方面, $X^{(t+1)}$ 由集合 $X^{(t)}$ 中的 $\mathbf{x}_i^{(t)}$ 移相而得, 从而保证 $(\mathbf{x}_i^{(t+1)})^T \mathbf{c}$ 局部最大化. 因此, 有:

$$J(X^{(t)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t+1)}) \leq J(X^{(t+1)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t+1)}) \quad (11c)$$

所以 $J(X^{(t)}, \mathbf{c}^{(t)}, \mathbf{B}^{(t)}) \leq J(X^{(t+1)}, \mathbf{c}^{(t+1)}, \mathbf{B}^{(t+1)})$ 成立. \square

定理 5. 以下极限存在:

$$\lim_{t \rightarrow \infty} J(X^{(t)}, \mathbf{c}^{(t)}, \mathbf{B}^{(t)})$$

证明. 使用式 (5) 和 Cauchy-Schwarz 不等式, 有:

$$\begin{aligned} \|\mathbf{m}\| &= \frac{\|\sum_{i=1}^n w_i \mathbf{x}_i\|}{\sum_{i=1}^n w_i} = \frac{\sum_{i=1}^n w_i \mathbf{x}_i^T \mathbf{c}}{\sum_{i=1}^n w_i} \leq \\ &= \frac{\sum_{i=1}^n \|w_i \mathbf{x}_i^T\| \|\mathbf{c}\|}{\sum_{i=1}^n w_i} = \frac{\sum_{i=1}^n w_i \|\mathbf{x}_i^T\| \|\mathbf{c}\|}{\sum_{i=1}^n w_i} = 1 \end{aligned}$$

根据定理 4, 易得由 PS-WS1M 产生的目标函数序列 $J(X^{(t)}, \mathbf{c}^{(t)}, \mathbf{B}^{(t)})$ 递增. 对于迭代次数 $t \geq 0$, 有

$$J(X^{(t)}, \mathbf{c}^{(t)}, \mathbf{B}^{(t)}) = \sum_{i=1}^n \frac{1}{\beta_i^r} \mathbf{x}_i^T \mathbf{c} + \varepsilon \sum_{i=1}^n \frac{1}{\beta_i^r} =$$

$$\|\mathbf{m}\| \sum_{i=1}^n \frac{1}{\beta_i^r} + \varepsilon \sum_{i=1}^n \frac{1}{\beta_i^r} \leq (1 + \varepsilon) \sum_{i=1}^n \frac{1}{\beta_i^r}$$

即算法 PS-WS1M 产生的递增序列存在上界, 因此序列收敛且极限存在. \square

2.3 参数选择

在算法 PS-WS1M 中, r 和 ε 是两个重要参数.

当 $r = 0$ 时, 每个时间序列记录的权重均为 1, 算法退化为单类模糊 C 均值聚类算法 (Fuzzy C means, FCM). 当 $r = -1$ 时, β_i 在迭代过程中保持不变. 此时, β_i 的值由先验信息决定. 然而, 这对用户来说往往很困难. 当 $r < -1$ 时, $\mathbf{x}_i^T \mathbf{c}$ 越大, $1/\beta_i^r$ 越小. 即与聚类中心的相似度越大的时间序列记录, 它对中心点的影响会越小, 这与加权原理相违背, 所以不能取 $r < -1$. 当 $r > 0$ 时, $\mathbf{x}_i^T \mathbf{c}$ 越大, β_i 也越大. 然而, $1/\beta_i^r$ 会越小, \mathbf{x}_i 对中心的影响也会越小. 这与加权原理相违背, 所以 $r > 0$ 也不能取. 当 $-1 < r < 0$ 时, $\mathbf{x}_i^T \mathbf{c}$ 越大, β_i 越小. 而 $1/\beta_i^r$ 因此会变大, 其相应的 \mathbf{x}_i 的效果会增强.

参数 ε 的作用是保证 $\mathbf{x}_i^T \mathbf{c} + \varepsilon > 0$. 显然, ε 应该满足 $\varepsilon > -\min \mathbf{x}_i^T \mathbf{c}$. 当 $\varepsilon \rightarrow \infty$ 时, β_i 趋于 $1/n$. 此时, PS-WK1M 退化为单类移相球面 k -means 算法^[12]. 此时, β_i 标识时间序列记录重要程度的作用失效. 反之, 当 ε 向相反方向变化时, β_i 对 $\mathbf{x}_i^T \mathbf{c}$ 值的变化会敏感, 这将有助于我们根据 β_i 的分布从数据集中区分相近的聚类.

3 基于移相加权球面单簇聚类的单分类器 PS-WS1M-OCC

单分类器 PS-WS1M-OCC 使用如下方法来获得数据描述: 首先, 在训练集上运行 PS-WS1M 进行单簇聚类, 得到聚类中心 \mathbf{c} , 然后, 确定相似度阈值 θ . 在分类测试时, 计算测试样本与聚类中心的最佳位移及其对应的相似度, 将相似度低于阈值 θ 的时间序列样本判为异常.

令人遗憾的是, 目前已有的时间序列异常检测算法大多由用户来指定阈值, 通常取异常值最大的 n 个时间序列记录作为异常记录, n 值由用户指定^[5, 12]; 从单分类器设计的角度出发, 文献 [11] 通过设置拒绝比来间接地设置隶属度阈值. 但是这些方法都依赖于用户对训练集的准确把握. 当用户缺乏训练集的相关知识时, 使用效果不理想.

本节中, 我们提出一种新方法来自动确定相似

度阈值 θ . 根据式 (7), PS-WS1M 收敛后得到的 β_i 值反映了训练样本 \mathbf{x}_i 与聚类中心 \mathbf{c} 的相似程度. 当数据集包含一个目标类和若干噪声时, PS-WS1M 收敛后所得 β_i 的分布直方图多呈典型的双峰特征, 两个波峰分别对应于数据集里的目标类和噪声. 新方法中把波峰之间的波谷作为划分数据集的依据. 下面我们将提出新的准则函数来确定 β_i 分布直方图上的波谷位置, 进而确定相应的相似度阈值 θ .

我们把 PS-WS1M 收敛后得到的 β_i 作为训练样本 \mathbf{x}_i 的异常值. 假设将异常值的分布范围等分为 L 个区间, n_l 表示落入第 l 个区间中异常值的数量. 则对于时间序列记录 \mathbf{x}_i 的异常值, 落入第 l 个区间的概率为

$$p_l = \frac{n_l}{n} \quad (13)$$

其中, $n = \sum_{i=1}^L n_i$ 为异常值的总数目.

假设存在阈值 t_{opt} , 将异常值分为两类, 用 C_1 代表正常类, C_2 代表异常类, 则两类的类概率计算如下:

$$\omega_1(t) = \sum_{i=1}^t p_i, \quad \omega_2(t) = \sum_{i=t+1}^L p_i \quad (14)$$

每一类的均值计算如下:

$$\mu_1(t) = \sum_{i=1}^t \frac{m_i p_i}{\omega_1(t)}, \quad \mu_2(t) = \sum_{i=t+1}^L \frac{m_i p_i}{\omega_2(t)} \quad (15)$$

其中, m_i 为第 i 个区间的中值.

根据最大类间方差准则, 并强调直方图的波谷对阈值的影响, 构造新的准则如下:

$$J(t) = (1 - p_t)^\alpha (\omega_1(t) \mu_1^2(t) + \omega_2(t) \mu_2^2(t)) \quad (16)$$

则相似度阈值 θ 计算如下:

$$t_{opt} = \arg \max_{1 \leq t \leq L} J(t)$$

$$\beta_{opt} = m_{t_{opt}}$$

$$\theta = \left(\beta_{opt} \sum_{i=1}^n (\mathbf{x}_i^T + \varepsilon)^{\frac{1}{r+1}} \right)^{r+1} - \varepsilon \quad (17)$$

其中, \mathbf{x}_i ($i = 1, 2, \dots, n$) 为训练集中的时间序列记录.

式 (16) 中, $\omega_1(t) \mu_1^2(t) + \omega_2(t) \mu_2^2(t)$ 部分体现了最大类间方差的原则. 而 $(1 - p_t)^\alpha$ 的加入, 更加强了直方图的波谷在确定阈值时的作用. 参数 α 控制了算法对式 (16) 中 p_t 变化的敏感程度. 当 α 过大时, 算法会对直方图中 p_t 的变化过于敏感. 根据反复实验, $\alpha = 0.1$ 时, 算法可以在大多数情况下得到很好的分割效果.

为了更清晰地表明 $(1 - p_t)^\alpha$ 的作用, 我们使用人工数据来模拟 PS-WS1M 得到的异常值. 方法如下: 生成集合 X_1 包含 500 个数值, 服从均值 0.8、标准差 0.08 的正态分布, 代表目标类的异常值; 生成集合 X_2 包含 30 个数值, 服从均值 0.5、标准差 0.02 的正态分布, 代表异常数据的异常值. 将 X_1 和 X_2 合并得到新的集合 X , 并归一化, 以满足约束条件 (6b). 以间隔 0.03 画出分布直方图, 并采用两种不同的分割准则计算 t_{opt} . 结果如图 3 所示, 所得阈值用虚线标出:

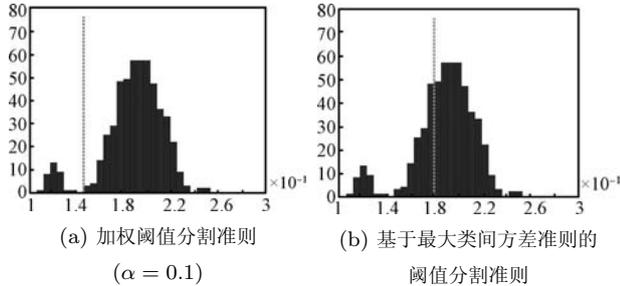


图 3 采用不同阈值分割准则函数的结果对比 (所得阈值用虚线标出)

Fig. 3 The comparison between the thresholding results obtained by different criterion functions (The broken line indicates the thresholds.)

从图 3 中不难看出, 对直方图进行分割时, 仅考虑最大化类间方差会将相当数量的 X_2 中的数值归为目标类中, 可见最大化类间方差不适用于两类大小差异明显的情况. 而采用本文的加权分割准则可以成功地将代表异常记录的异常值与代表正常记录的异常值分开. 可见 $(1 - p_t)^\alpha$ 的加入能够使算法有效地确定直方图中波谷的位置, 这对于包含少量异常数据的不平衡数据集有着重要的意义.

4 实验

4.1 数据集的获取

心电图时间序列是典型的周期时间序列, 对其进行异常检测具有重要的医学价值. 本文基于心电图时间序列对单分类器的性能进行评估. 实验中, 我们采用 Precision 和 Recall 这两个指标对单分类器的性能进行评估, 其计算方法如下:

$$Precision = \frac{TP}{TP + FP} \quad (18a)$$

$$Recall = \frac{TP}{TP + FN} \quad (18b)$$

其中, TP 、 FP 、 FN 如表 1 所示:

表 1 单分类可能的结果汇总

Table 1 The result for one-class classifiers

真实类别	分类为正常类	分类为异常类
正常类	True positive (TP)	False negative (FN)
异常类	False positive (FP)	True negative (TN)

本文实验所采用的数据集从 MIT-BIH ECG 数据库^[13] 中下载. MIT-BIH 数据库共包含 48 条超过 30 分钟的心电图数据. 选择编号为 100~109 的时间序列, 采用滑动窗口技术对它们进行分割. 对于每一条心电图时间序列, 首先, 确定周期 T , 然后, 使用宽度为 T 的滑动窗口, 以固定间隔得到等长的时间子序列. 这样我们就得到了 10 个时间序列数据集, 并对数据集里的每一条记录进行 z -score 归一化 (z -score normalization). 由于分割过程中的误差累积, 每个数据集里的时间序列记录的相位都存在着差异.

为了得到包含大量异常数据的数据集, 我们在领域专家的帮助下, 首先, 标识各数据集里的正常记录 and 异常记录, 并将正常记录作为目标类. 然后, 将这 10 个数据集里的所有异常记录作为异常类, 分别和每个数据集里的正常记录混合, 生成新的数据集. 例如, 选择编号为 100 的心电图时间序列, 对其进行分割后生成相应的时间序列数据集, 选取其中的正常记录作为目标类, 并将之与 10 个数据集包含的所有异常记录混合. 我们将新的数据集命名为 Blend100. 由于基于不同心电图得到的时间序列记录的长度各不相同, 所以, 我们采用线性插值的方法, 使它们具有相同的长度. 表 2 给出了我们构造的部分数据集的相关信息.

表 2 人工数据集相关信息

Table 2 The information on the artificial datasets

数据集名称	目标类大小	异常类大小	时间序列记录的长度	目标类来源 (记录号)
Blend100	220	360	150	100
Blend102	170	351	150	102
Blend104	100	331	150	104
Blend106	100	331	150	106
Blend108	120	301	150	108

4.2 训练集里时间序列记录与聚类中心的位移方差对单分类器性能的影响

我们首先研究训练集里时间序列记录与聚类中心的位移方差对单分类器性能的影响. 对于表 2 中的每个数据集, 我们分别构造两组包含相同训练样本的训练集, 并通过移位, 使这两组训练集相对于

聚类中心具有不同的位移方差. 表 3 给出了训练集里时间序列记录与训练算法收敛后所得聚类中心的位移方差和 PS-WS1M 收敛所需的迭代次数; 同时也给出了 PS-WS1M-OCC 使用不同的训练集进行训练后, 对其进行分类测试所得到的 *Precision* 和 *Recall* 指标值. 本实验中, 我们所使用的训练集大小均为 60, 其中, 包含 50 个目标类样本和 10 个异常类样本. 算法运行 10 次后取平均值.

表 3 PS-WS1M 收敛后训练集里时间序列记录位移的方差对 PS-WS1M-OCC 性能的影响

Table 3 The effect of the variance of the phase shift after PS-WS1M convergences on the performance of the PS-WS1M-OCC

数据集名称	目标类大小	异常类大小	时间序列记录的长度	目标类来源 (记录号)
Blend100	220	360	150	100
Blend102	170	351	150	102
Blend104	100	331	150	104
Blend106	100	331	150	106
Blend108	120	301	150	108

从表 3 中我们可以看出, 对于包含相同训练样本、但相对于聚类中心具有不同位移方差的训练集, 单分类器都有稳定的性能, 这说明单分类器 PS-WS1M-OCC 的性能不受训练集里每个样本相对于聚类中心位移方差的影响.

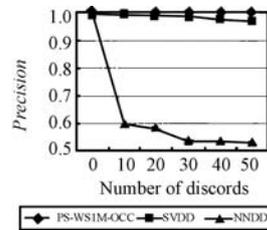
4.3 单分类器对训练集内异常数据的敏感性

本实验比较不同的单分类器对训练集包含的异常数据的敏感性. 我们选择支持向量数据描述 (Support vector domain description, SVDD)、最近邻单分类器 (Nearest neighbor data description, NNDD) 与 PS-WS1M-OCC 进行比较. 它们的实现代码均来自于 `dd_tools`^[6] 工具箱. 由于 `dd_tools` 中的单分类器并不是专为时间序列而设计, 所以不能解决时间序列记录的相位同步问题. 为了使它们适用于基于时间序列数据集的异常检测, 本文对人工数据集进行对齐处理, 使同一数据集里各个时间序列记录的最大值点在同一时间点上.

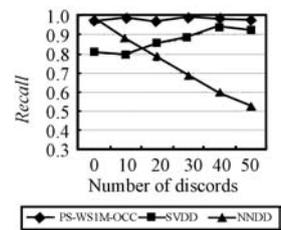
实验中, 我们随机从目标类中选择 50 个样本, 同时选择异常类中的若干样本, 从而构成训练集. 依次增加训练集里异常记录的数量, 考察其对单分类器性能的影响. 单分类器 PS-WS1M-OCC 进行训练时 PS-WS1M 的参数设置为: $\varepsilon = 3$, $r = -0.95$. 工具箱 `dd_tools` 采用参数 `fracrej` 来控制训练集里被拒绝的样本的数量. 为了优化 SVDD 和 NNDD 的性能, 如下设置 `fracrej` 参数: $fracrej = n_{discords} / (n_{normal} + n_{discords})$, 其中, $n_{discords}$ 和 n_{normal} 分别表示训练集里异常记录和正

常记录的数量. 此外, 我们设置 SVDD 的核带宽参数 $\sigma = 5$. 图 4 分别给出了训练集里异常记录数量变化的情况下, 测试不同单分类器所得的 *Precision* 和 *Recall* 指标的变化.

从图 4 所示实验结果中可以发现, 随着训练集里异常记录的增加, PS-WS1M-OCC 的 *Precision* 和 *Recall* 几乎不受影响, 始终接近 1, 而 SVDD 和 NNDD 的 *Precision* 和 *Recall* 指标均明显地受到



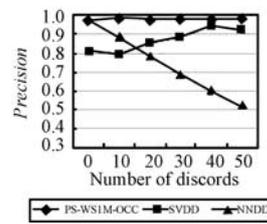
(a) 数据集 Blend100 上 *Precision* 指标比较



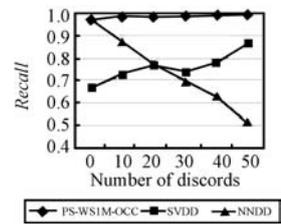
(b) 数据集 Blend100 上 *Recall* 指标比较

(a) Comparison of *Precision* indexes on dataset Blend100

(b) Comparison of *Recall* indexes on dataset Blend100



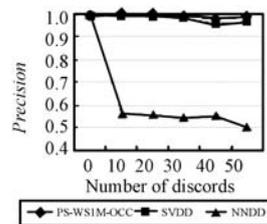
(c) 数据集 Blend102 上 *Precision* 指标比较



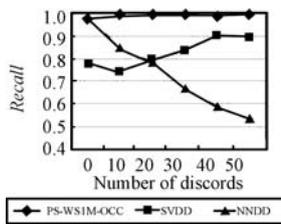
(d) 数据集 Blend102 上 *Recall* 指标比较

(c) Comparison of *Precision* indexes on dataset Blend102

(d) Comparison of *Recall* indexes on dataset Blend102



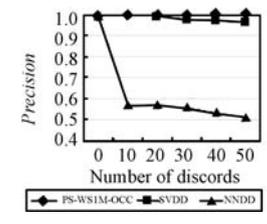
(e) 数据集 Blend104 上 *Precision* 指标比较



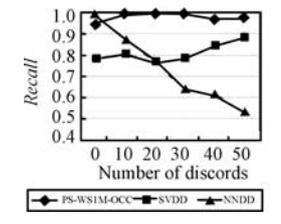
(f) 数据集 Blend104 上 *Recall* 指标比较

(e) Comparison of *Precision* indexes on dataset Blend104

(f) Comparison of *Recall* indexes on dataset Blend104



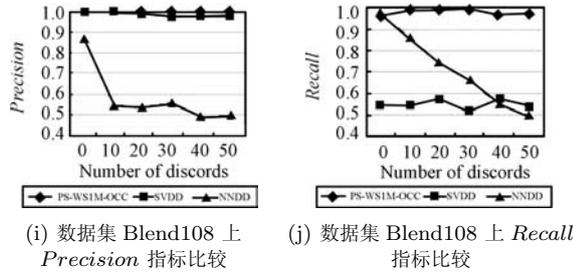
(g) 数据集 Blend106 上 *Precision* 指标比较



(h) 数据集 Blend106 上 *Recall* 指标比较

(g) Comparison of *Precision* indexes on dataset Blend106

(h) Comparison of *Recall* indexes on dataset Blend106



(i) 数据集 Blend108 上 Precision 指标比较
(j) 数据集 Blend108 上 Recall 指标比较
(i) Comparison of Precision indexes on dataset Blend108
(j) Comparison of Recall indexes on dataset Blend108
图 4 PS-WS1M-OCC、SVDD 和 NNDD 性能比较
Fig. 4 Performance comparison for PS-WS1M-OCC, SVDD and NNDD

训练集里异常记录数量变化的影响, 这说明 PS-WS1M-OCC 对训练集里的异常数据是鲁棒的. 其原因在于训练集里异常记录的增加不会对 PS-WS1M-OCC 的阈值产生明显的影响.

这里需要指出的是, 以上实验中, SVDD 和 NNDD 的实验结果是在假设用户已经知道数据集里异常记录数量的前提下获得的. 在实际应用中, 用户无法确切知道训练集里异常数据的数量, 因此, SVDD 和 NNDD 的 *fracrej* 的参数很难确定, 也很难使 SVDD 和 NNDD 得到最优的性能.

4.4 单分类器对参数的敏感性

参数的选择对单分类器的性能也有着重要影响. 本实验中, 我们在包含噪声的数据集上, 通过改变不同算法的相关参数, 来考查各个算法对参数的敏感程度. 实验中, 我们所用的训练集包含 50 个正常记录, 30 个异常记录. 以下是三个算法在各自参数的变化下, 其 Precision 和 Recall 指标的变化.

从图 5(a) 和 5(b) 中, 我们不难发现, PS-WS1M-OCC 对参数 ε 的变化不敏感. 虽然 ε 在较大的范围内变化, 但是单分类器所得结果的 Precision 和 Recall 指标基本保持不变, 这有利于用户在较大的范围内自由的选取参数. 当参数 r 变化时, 我们也可以得到类似结论.

相反, SVDD 和 NNDD 的分类结果对各自参数的不同取值较为敏感. 参数 *fracrej* 取决于训练集里异常数据所占的比例, 通常情况下用户无法得到关于训练集包含异常数据的确切数量, 因此, 一个理想的单分类器应该对参数 *fracrej* 不敏感. 令人遗憾的是, 图 5(d) 和 5(f) 表明, SVDD 和 NNDD 所得分类结果的 Recall 指标对 *fracrej* 的变化都比较敏感.

SVDD 的核带宽 σ 表明了数据描述边界的光滑程度. 当 σ 越大, 其边界越光滑; 反之, 当 σ 越小, 则越有可能发生过拟合. 图 5(g) 和 5(h) 给出了在核带宽 σ 不同取值的情况下 SVDD 对 Precision 和

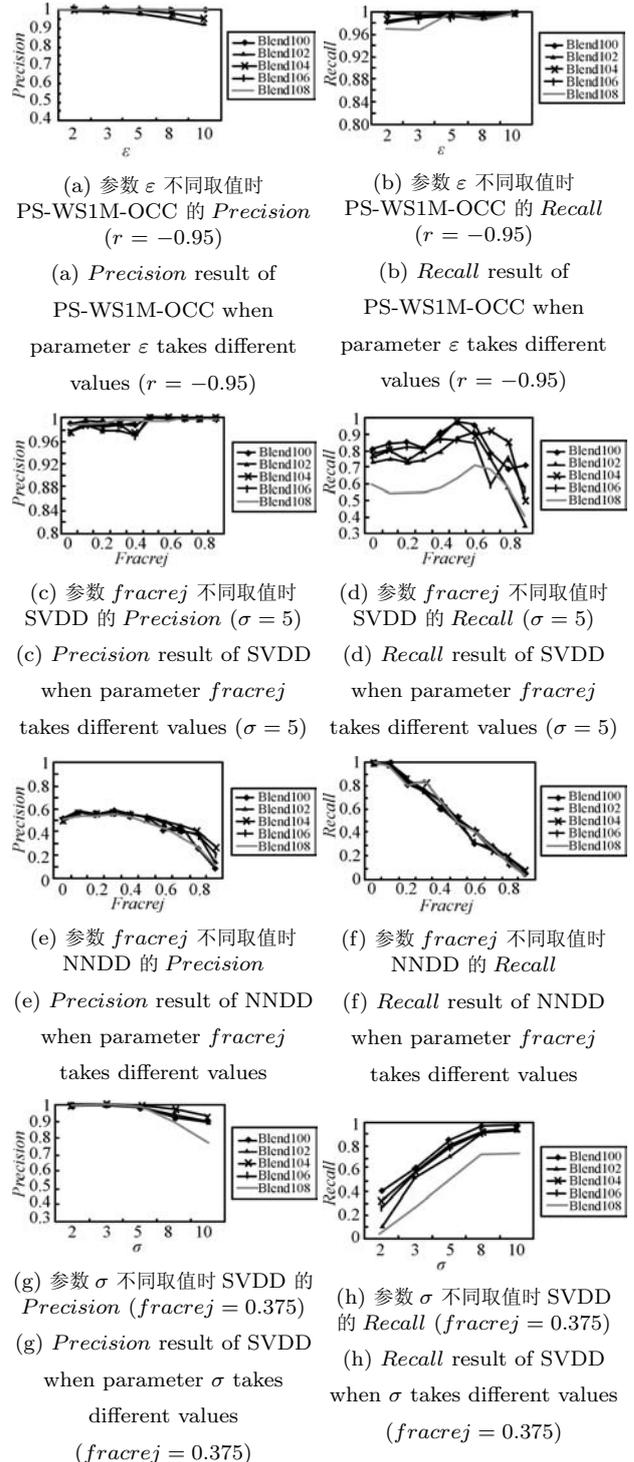


图 5 参数敏感性实验结果
Fig. 5 The experimental results for parametric sensitivity

Recall 的变化情况. 从中我们发现, 核带宽 σ 逐渐变大的过程中, Precision 有下降趋势. 原因在于, 核带宽 σ 越大, SVDD 生成的数据描述边界越光滑, 测试集里就有越多的异常记录被误判为正常, 即式 (18a) 中的 *FP* 的值变大. 反之, 核带宽 σ 逐渐变小的过程中, SVDD 发生过拟合的可能性就越来越大,

会有更多的正常记录被误判为异常, 即 FN 的值变大, 因此图 5(h) 中 $Recall$ 会随核带宽 σ 逐渐变小而减小. 总而言之, SVDD 的 $Precision$ 和 $Recall$ 指标不可能同时达到最优.

5 结论

针对单条周期时间序列的异常检测问题, 本文提出了一种基于移相加权球面单簇聚类的单分类器. 通过在聚类过程中增加高效的循环移位操作, 解决了时间序列记录之间相似度计算的问题; 基于时间序列记录的权重分布, 提出了新的阈值确定方法, 从而使单分类器在训练过程中具有自动识别目标类而拒绝异常数据的能力. 实验表明, 本文方法与传统的单分类器相比, 能够很好地实现周期时序序列的异常检测.

进一步地, 本文基于单簇聚类的思想来实现单分类器, 以及基于异常值的分布来实现阈值的自适应选取, 能够很好地实现从包含噪声的数据中进行无监督学习, 这进一步扩展了相关领域的研究思路.

但是, 本文实验中所使用的训练集里时间序列记录相对于聚类中心位移的方差尚处在一定的范围内. 在今后的工作中, 将进一步研究训练集里时间序列记录相对于聚类中心的位移方差对单分类器 PS-WS1M-OCC 性能的影响, 相关问题将另文探讨.

References

- 1 Ma J, Perkins S. Online novelty detection on temporal sequences. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2003. 613–618
- 2 Wei L, Kumar N, Lolla V N, Keogh E J, Lonardi S, Ratanamahatana C. Assumption-free anomaly detection in time series. In: Proceedings of the 17th International Scientific and Statistical Database Management. California, USA: Lawrence Berkeley Laboratory, 2005. 237–240
- 3 Keogh E J, Lonardi S, Chiu B Y. Finding surprising patterns in a time series database in linear time and space. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Alberta, Canada: ACM, 2002. 550–556
- 4 Yang J, Wang W, Yu P S. Mining surprising periodic patterns. *Data Mining and Knowledge Discovery*, 2004, **9**(2): 189–216
- 5 Keogh E J, Lin J, Fu A W, Herle H V. Finding unusual medical time-series subsequences: algorithms and applications. *IEEE Transactions on Information Technology in Biomedicine*, 2006, **10**(3): 429–439
- 6 Tax D M J. Data description tool-box (dd tools) [Online], available: <http://homepage.tudelft.nl/n9d04/dd.tools.html>, January 31, 2011
- 7 Wu M R, Ye J P. A small sphere and large margin approach for novelty detection using training data with outliers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(11): 2088–2092

- 8 Cao L J, Lee H P, Chong W K. Modified support vector novelty detector using training data with outliers. *Pattern Recognition Letters*, 2003, **24**(14): 2479–2487
- 9 Jiang M F, Tseng S S, Su C M. Two-phase clustering process for outliers detection. *Pattern Recognition Letters*, 2001, **22**(6–7): 691–700
- 10 Hochbaum D S, Shmoys D B. A best possible heuristic for the κ -center problem. *Mathematics of Operations Research*, 1985, **10**(2): 180–184
- 11 Chen Bin, Feng Ai-Min, Chen Song-Can, Li Bin. One-cluster clustering based data description. *Chinese Journal of Computers*, 2007, **30**(8): 1325–1332
(陈斌, 冯爱民, 陈松灿, 李斌. 基于单簇聚类的数据描述. 计算机学报, 2007, **30**(8): 1325–1332)
- 12 Rebbapragada U, Protopapas P, Brodley C E, Alcock C R. Finding anomalous periodic time series. *Machine Learning*, 2009, **74**(3): 281–313
- 13 Moody G B. MIT-BIH arrhythmia database directory [Online], available: <http://www.physionet.org/physiobank/database/html/mitdbdir/intro.htm>, January 31, 2011



王 骏 博士, 讲师. 主要研究方向为模式识别、数据挖掘、数字图像处理. 本文通信作者.

E-mail: wangjun_sytu@sina.com

(WANG Jun Ph.D., lecturer. His research interest covers pattern recognition, data mining, and digital image processing. Corresponding author of

this paper.)



钟富礼 博士, 香港理工大学电子计算学系副教授. 主要研究方向为模糊系统、模式识别、数据挖掘.

E-mail: cskchung@inet.polyn.edu.hk

(CHUNG Fu-Lai Ph.D., associate professor in the Department of Computing, Hong Kong Polytechnic University. His research interest covers fuzzy systems, pattern recognition, and data mining.)



王士同 教授. 主要研究方向为人工智能、模式识别、数字图像处理.

E-mail: wxwangst@yahoo.com.cn

(WANG Shi-Tong Professor. His research interest covers artificial intelligence, pattern recognition, and digital image processing.)



邓赵红 博士, 副教授. 主要研究方向为模式识别、模糊系统、智能计算.

E-mail: dzh666828@yahoo.com.cn

(DENG Zhao-Hong Ph.D., associate professor. His research interest covers pattern recognition, fuzzy systems, and intelligence computing.)