一种基于DTW聚类的水文时间序列相似性挖掘方法

杨艳林 叶 枫 吕 鑫 余 霖 刘 璇
(河海大学计算机与信息学院 南京 210098)

摘 要 水文时间序列相似性挖掘是水文时间序列挖掘的重要方面,对洪水预报、防洪调度等具有重要意义。针对水文数据的特点,提出了一种基于DTW聚类的水文时间序列相似性挖掘方法。该方法先对数据进行小波去噪、特征点分段以及语义划分,再基于DTW距离对划分后的子序列做层次聚类并符号化;然后根据符号序列间的编辑距离筛选候选集;最后通过序列间的DTW距离进行精确匹配,获取相似水文时间序列。以滁河六合站的日水位数据进行实验,结果表明,所提方法能够有效地缩小候选集,提高查找语义相似的水文时间序列的效率。

关键词 水文时间序列,语义相似,DTW距离,层次聚类,编辑距离
中图法分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.2.051

DTW Clustering-based Similarity Mining Method for Hydrological Time Series

YANG Yan-lin YE Feng LV Xin YU Lin LIU Xuan
(College of Computer and Information, Hohai University, Nanjing 210098, China)

Abstract Similarity mining of hydrological time series is an importance aspect of hydrological time series mining. It will be of great importance in flood forecasting and flood control scheduling. According to the characteristics of hydrological data, this paper proposed a DTW clustering-based similarity mining method over hydrological time series. Firstly, on the premise of wavelet denoising, feature point segmentation and semantic classification, hierarchical cluster analysis is used to the classified sub-sequences based on DTW distance and the sub-sequences are symbolized. Then, candidate sets of time series are filtered according to the edit distance between symbol sequences. Finally, the similar hydrological time series are got precisely from the candidate sets by DTW exact matching. Experiments on the water level of Chuhe Liuhe station show that the proposed method can narrow the candidate sets effectively and improve the efficiency of searching for semantic similarity of hydrological time series.

Keywords Hydrological time series, Semantic similar, DTW distance, Hierarchical clustering, Edit distance

1 引言

水文时间序列的相似性分析可以回答防汛指挥中经常会问到的"当前水文过程相当于历史上哪一时期的同类过程"等问题,同时也是研究时序关联规则挖掘、聚类、模体挖掘以及异常发现等问题的基础,因而在洪水预报、防洪调度等方面有着重要的意义[1]。

在水文时间序列相似性查询中,时间序列降维和相似性度量方式是相似性分析研究的热点问题。基于特征点的分段表示法能在保留子序列语义的基础上实现有效的时间序列降维,在水文时间序列相似性挖掘中被广泛应用[3-5,8]。Lin等[6]提出时间序列符号化的经典算法SAX,SAX采用PAA算法对时间序列进行降维,用不同的符号代表相应平均值的子序列,具有简单易用的特点。然而,PAA算法等长划分时间序列往往会破坏序列语义,导致符号化效果不理想,因此后期的符号化算法多是基于SAX算法并结合数据特点进行改进。闫秋燕[7]等提出了一种基于关键点的SAX改进算法,该算法选取序列中满足一定筛选条件的极值点作为关键点来对时间序列进行分段,很好地保留了序列的形态特征。朱跃龙等[8]通过引入语义相似的概念,基于极值点对时间序列进行划分并赋予语义符号,实现水文时间序列的语义符号化。李迎[9]提出了一种基于DTW的符号化时间序列聚类算法,该算法将DTW距离用于关键点提取后的符号序列间的相似性度量,再利用Normal矩阵和FCM方法进行聚类分析,说明了将DTW距离用于时间序列聚类分析具有较高的准确率。但是该算法对初始聚类中心敏感,容易产生局部最优解,且无法避免孤点造成的影响。现有的水文时间序列的相似性度量方式主要有欧氏距离[2]、动态模式匹配[3]、FastDTW[4]距离

到稿日期:2015-05-14 返修日期:2015-10-20 本文受国家自然科学基金面上项目(61272543),国家科技支撑计划(2013BAB06B04),国家自然科学基金委-广东联合项目(U1301252),江苏省博士后科研资助计划(1401001C),江苏水利科技项目:"智慧河流"研究及其在六合滁河管理中的应用(2013025),国家科技支撑计划项目数字流域关键技术(2013BAB05B00),基于物联网的流域信息获取技术研究(2013BAB05B01)资助。
杨艳林(1990—),男,硕士,主要研究方向为数据挖掘,E-mail:383939536@qq.com;叶 枫(1980—),男,博士,讲师,主要研究方向为云计算、水信息学;吕 鑫(1983—),男,博士后,主要研究方向为密码学、网络信息安全;余 霖(1991—),女,硕士,主要研究方向为数据挖掘;刘 璇(1989—),女,博士,主要研究方向为数据挖掘。

以及 DTW 距离[10,11]等，这些度量方式各有所长，针对不同的相似性分析需求可选择相应的度量方式。目前水文时间序列相似性度量应用最多的是动态扭曲时间(Dynamic Time Warping，DTW)距离。

快速挖掘出相似水文时间序列具有重要意义，本文在文献[8]所提出的 DTW_SS 方法的基础上，针对 DTW_SS 方法在挖掘过程中产生的候选集过大从而导致在采用 DTW 距离进行精确匹配时耗时过长的问题，结合凝聚层次聚类方法，提出一种基于 DTW 聚类的水文时间序列相似性挖掘方法(DTW Clustering-based Similarity Mining，DTW_CSM)。DTW_CSM方法采用凝聚层次聚类方法，对语义划分后的子序列做聚类分析并符号化，根据符号化结果在历史数据中获取与查询符号序列编辑距离最小的候选集，最后在候选集中进行 DTW 距离精确匹配以获取相似时间子序列，从而提高查询效率。

2 时间序列语义相似与聚类

2.1 时间序列语义相似

语义是数据在某个领域上的解释和逻辑表示。时间序列语义化是对水文时间序列进行离散化，将所得子序列看作一系列的符号，并对每个符号进行语义解释。

对水文时间序列 $X=\{x_1,x_2,\cdots,x_n\}$进行特征点提取，得到特征点序列 $X'=\{x_1',x_2',\cdots,x'_{m+1}\}$，对基于 X'分段后得到的子序列进行语义符号化，可将时间序列降维表示为 $S=\{s_1,s_2,\cdots,s_m\}$，$m\leqslant n$。其中，

$$s_i=\begin{cases}D, & x'_{i+1}<x_i'\\ B, & x'_{i+1}=x_i',i=1,2,\cdots,m\\ U, & x'_{i+1}>x_i'\end{cases}$$

D、B、U 分别代表下降、平稳和上升，S 就是时间序列$X=\{x_1,x_2,\cdots,x_n\}$的语义模式表示。所谓时间序列语义相似，就是指两个时间序列的语义模式表示的相似程度。如果两个时间序列的语义模式表示均为 U、D、B(上升、下降、保持)，那么称这两个时间序列是语义相似的。

2.2 层次聚类

层次聚类方法将数据对象聚集成聚类树，通常分为凝聚和分裂两种层次聚类方法[12]。凝聚层次聚类是一种自底向上的策略，首先将每个对象看作一个类，然后根据不同的类间距离度量准则合并初始类，直至满足终止条件。分裂层次聚类是一种自顶向下的策略，与凝聚层次聚类的策略正好相反，先是将所有对象看作一个类，然后再不断分解，直到满足终止条件。水文时间序列中往往存在异常点，导致语义划分后的子序列容易出现孤点，采用文献[9]的算法进行聚类分析的效果不好，而凝聚层次聚类法无需事先确定聚类中心，且在结果生成后可剔除数据中的孤点，能达到更好的聚类效果。因此，本文采用 DTW 距离作为时间子序列间的相似性度量，通过凝聚层次聚类法对单一语义的时间子序列进行聚类，具体算法步骤如下：

输入：单一语义时序子序列集 $X=\{X_1,X_2,\cdots,X_m\}$，目标类的类数 $N(0<N\leqslant n)$；

输出：语义相同的 N 类子序列集。

(1)初始化。把每个子序列 $X_i(i=1,2,\cdots,m)$归为一类，计算每两个序列间的 DTW 距离作为类间距离。

(2)合并。寻找各个类之间距离最近的两个类，将这两个类归为一类。

(3)重新计算类间距离。采用 average-linkage 作为两个类之间的距离准则，即两个类对象之间的平均距离。假设类 A 和类 B 是已有类，则类 A 和类 B 的类间距离为：

$$avgDis(A,B)=\frac{1}{|A|\cdot|B|}\sum_{X_i\in A}\sum_{X_j\in B}DTW(X_i,X_j) \quad (1)$$

其中，$|A|$为类 A 的大小，$|B|$为类 B 的大小，$1\leqslant i\leqslant m$，$1\leqslant j\leqslant m$，$i\neq j$。

(4)重复步骤(2)和(3)，直到满足终止条件。

聚类完成后，需要定义每个类的中心子序列，用于对任意查询序列进行符号化。对于类 U 中的子序列 u_i，若 u_i 使得式(2)达到最小，则将 u_i 作为类 U 的中心子序列。

$$\sum_{u_j\in U,i\neq j}DTW(u_i,u_j) \quad (2)$$

3 基于 DTW 聚类的相似性挖掘

3.1 挖掘流程

基于 DTW 聚类的相似性挖掘流程可以分为两个主要的阶段：相似性挖掘准备阶段和相似性挖掘查询阶段。

3.1.1 相似性挖掘准备阶段

水文时间序列相似性挖掘准备阶段主要包括以下 4 个步骤。

步骤 1：数据预处理。对原始时间序列进行数据缺失和数据平滑处理。

步骤 2：特征点提取。根据特征点定义从原始时间序列中获取特征点。

步骤 3：语义划分。根据特征点对应数值的大小，将每两个特征点之间的子序列划分为相应的语义符号。

步骤 4：聚类并符号化。结合 DTW 距离，通过凝聚层次聚类法分别对单一语义的符号集进行聚类，根据聚类结果实现时间子序列的符号化。

3.1.2 相似性挖掘查询阶段

相似性挖掘准备阶段完成后，首先结合中心子序列对查询序列进行符号化，然后把历史符号序列中与查询符号序列等长且编辑距离最短的子序列归到候选集，最后在候选集中通过 DTW 距离精确匹配获取相似子序列。鉴于计算 DTW 距离的时间复杂度为 $O(mn)$，在用 DTW 距离进行精确匹配时，候选集越大则查询时间越长，因此缩小候选集能够有效提高获取相似水文时间序列的效率。

根据水文时间相似性查询流程，给出相似性挖掘查询阶段的算法框架，如算法 1 所示。

算法 1 水文时间序列相似性查询算法

输入：历史序列 X，查询序列 Q，历史符号序列 SX

输出：相似水文时间序列 Sim_X

Step1：查询序列符号化

1. 查询序列进行小波去噪并提取特征点；
2. 基于特征点对查询序列进行划分并语义化；
3. 获取查询符号序列 SQ，分别将语义划分后的每个单一语义时序子序列用与它之间 DTW 距离最小的中心子序列所在类的符号表示。

Step2：根据编辑距离筛选候选集

//初始化符号序列最小编辑距离

4. editDisMin=SQ. size()；

```
5. Cand_X=∅;//初始化候选集
6. for i=1:(SX. size()-SQ. size()+1)
7.   if editDis(SQ,SX(i:i+SQ. size-1)) < editDisMin
8.     Cand_X= ∅;//清空候选集
       //将符号序列对应的时间序列归入候选集
9.     Cand_X ←Xi;
10.  else if editDis(SQ,S(i:i+SQ. size-1)) == editDisMin
       //将符号序列对应的时间序列归入候选集
11.    Cand_X ←Xi
12.  end if
13. end for
Step3:基于 DTW 距离精确匹配
14. dtwDisMin=10000;//为最小 DTW 距离赋初值
15. for k=1:Cand_X. size()
16.   if dtwDis(Q,Cand_X(k)) < dtwDisMin
17.     dtwDisMin=dtwDis(Q,Cand_X(k));
18.     Sin_X=Cand_X(k);
19.   end if
20. end for
21. return Sim_X;//输出相似时间序列
```

3.2 数据预处理

水文时间序列数据通常是不完整且存在缺失值和噪声的，在对历史水文数据进行挖掘前，往往先要对数据进行预处理。通过数据预处理，能使数据更适合进行数据挖掘，从而更好地从中挖掘出有价值的信息。

3.2.1 数据缺失处理

水文数据在采集入库的过程中，可能会出现数据丢失的情况，导致数据库存储值为 0 或 null，这些缺失的数值对挖掘结果会产生一定的影响，因此要对数据进行缺失处理。数据缺失处理的方式主要有丢弃数据和填补数据两种，使用较多的主要是填补数据的方式，即通过使用一些统计值（如平均值）来填补缺失数据。对于水文时间序列 $X=\{x_1,x_2,\cdots,x_n\}$，X 首尾不缺失，若 x_i $(1<i<n)$ 缺失，则用 $x_i=\frac{x_{i-1}+x_{i+1}}{2}$ 填补。

3.2.2 数据平滑

时间序列的整体波动趋势是水文时间序列数据中最受关注的部分。通过数据去噪来消除序列中的短期波动能够很好地显示数据整体变化趋势，所以结合数据特点进行数据平滑处理是数据预处理过程中必不可少的。已有研究表明，小波变换能够在很好地保持原序列整体变化趋势的情况下最大限度地去除噪声，适用于水文时间序列平滑处理[8,13]。图 1 是用 bior3.7 小波变换对本文查询序列进行数据平滑的效果图，由图可知，bior3.7 小波平滑很好地保留了数据整体变化的趋势。

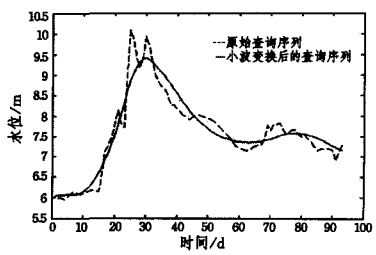


图 1 查询序列 bior3.7 小波变换效果图

3.3 特征点选取

特征点是指水文时间序列中对其形态以及整体趋势变化影响较大的数据点。基于特征点的分段表示法能够在不破坏子序列语义的基础上实现有效的时间序列降维，提高查询速度。时间序列 $X=\{x_1,x_2,\cdots,x_n\}$ 的特征点获取原则如下：

1. 时间序列的起始点和终止点，即 $i=1$ 或 $i=n$；

2. 数据点 x_i 是极值点且满足以下条件中的一个：

- x_i 与上一个特征点 x_j 的距离大于指定阈值 C，即：

$$|j-i|\geqslant C \tag{3}$$

- x_i 与相邻两点的连线形成的夹角小于指定筛选角度 α_0，即：

$$\cos\alpha_0<\cos\langle\vec{a},\vec{b}\rangle=\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|*|\vec{b}|}\leqslant 1 \tag{4}$$

其中，$\vec{a}=(x_{i-1}-x_i,-1)$，$\vec{b}=(x_{i+1}-x_i,1)$。

3.4 时间序列符号化

时间序列符号化是将时间序列转化为符号序列的过程，符号序列的每一个符号都表示一段子序列。结合水文数据特点，本文首先基于特征点对水文时间序列进行划分，然后根据语义概念对划分后的时间子序列进行语义符号化，最后通过凝聚层次聚类对单一语义的子序列做进一步聚类分析并符号化，在不破坏时间子序列语义的前提下实现时间序列符号化。

3.5 符号序列度量

符号序列可直接通过比较来计算两个符号序列的距离，较常用的比较方式是计算两个符号序列的编辑距离，即两个符号序列通过插入、删除和替换符号达到完全一致的最小编辑次数，如符号序列 BBDU 与 BBDB 的编辑距离为 1，将 U 替换为 B。本文对查询序列和历史序列符号化后，在历史符号序列中以从头到尾遍历进行筛选，每次都取与查询符号序列等长的历史符号子序列进行比对，最终获取编辑距离最小的子序列集作为候选集。通过符号序列间的编辑距离筛选候选集可以对文献[8]中的筛选过程做出优化，避免因找不到符号序列完全一致的子序列而导致查询失败。两个等长的符号序列 $X=\{x_1,x_2,\cdots,x_n\}$ 和 $Y=\{y_1,y_2,\cdots,y_n\}$ 的编辑距离为：

$$editDis(X,Y)=\sum_{i=1}^{n}dis(x_i,y_i) \tag{5}$$

其中：

$$dis(x_i,y_i)=\begin{cases}0, & x_i=y_i\\ 1, & x_i\neq y_i\end{cases} \tag{6}$$

4 实验分析

在 DTW_SS 方法的基础上，结合 DTW 距离，采用凝聚层次聚类对语义划分后的子序列做进一步聚类并符号化，以达到缩小候选集的目的，进而提高查询效率。为了验证DTW_CSM 方法的有效性及准确性，以滁河六合站的日平均水位为例，在相同的实验环境下，将 DTW_CSM 方法与DTW_SS 方法进行实验对比分析。

4.1 查询准备

本文实验软件环境为 Matlab R2010b，Windows 7 旗舰版操作系统，硬件环境为 2.30GHz CPU、4GB 内存的笔记本电脑。以滁河六合站 1972－2003 年共 32 年的日平均水位数据

作为数据集，选取 2003 年 6 月 12 日—2003 年 9 月 12 日长度为 93 的日平均水位时间序列作为查询序列。

首先用以平均值填补缺失值的方法实现数据缺失处理，然后用 bior3.7 小波变换对原时间序列进行数据平滑，接着设相邻特征点阈值 C 为 4，筛选角度 α_0 为 135°，下降语义聚类类数 D_n 为 2，上升语义聚类类数 U_n 为 2，对小波平滑后的水文时间序列进行特征点提取、语义划分、聚类并符号化。在聚类过程中，由于时间序列经过小波平滑后，导致平稳语义的时间子序列被弱化，因此只对上升和下降语义的时间子序列进行凝聚层次聚类。最后根据水文时间序列相似性查询算法进行相似性挖掘。查询序列在不同方法下的符号化结果如表 1 所列。

表 1　查询序列符号化结果

方法	符号序列
DTW_SS	UDUD
DTW_CSM($C=4,\alpha_0=135°,D_n=2,U_n=2$)	$U_2D_1U_2D_1$

4.2　时间分析

本文将相似性挖掘查询阶段产生的候选集大小和查询时间作为时间分析的指标。表 2 是查询序列在整个数据集中的查找结果。从表 2 可以看出，在适当的参数下，DTW_CSM 方法的候选集大小和查询时间都比 DTW_SS 方法小。

表 2　DTW_CSM 与 DTW_SS 方法运行结果的对比

方法	特征点个数	候选序列个数	查询时间/s
DTW_SS	439	159	5.52
DTW_CSM	403	25	1.12

在查询时间序列不变的情况下，随着历史序列集的不断增加，分别对 DTW_SS 方法和 DTW_CSM 方法的查询时间和候选集大小进行对比。图 2 和图 3 分别是在不同长度的历史序列集下两种方法查询时间和候选集大小的对比图。从中可以发现，随着历史序列长度的不断增长，查询时间和候选集大小也不断增长，而与 DTW_SS 相比，DTW_CSM 方法的查询时间和候选集大小都比 DTW_SS 方法的小，增长幅度更加缓慢，说明了 DTW_CSM 方法的有效性。同时，也可以看出查询时间与候选集大小成正比关系，即候选集越大，精确匹配获取相似水文时间序列消耗的时间越多。

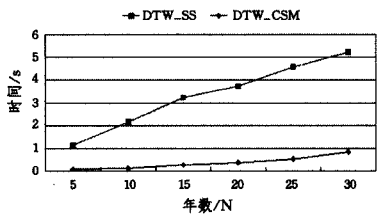


图 2　两种方法查询时间的对比

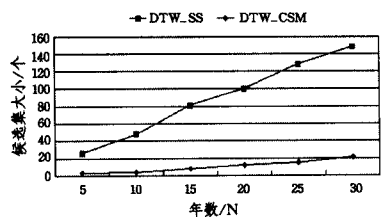


图 3　两种方法候选集大小的对比

4.3　准确性分析

本文将 DTW_SS 方法和 DTW_CSM 方法查询得到的相似水文时间序列进行对比，以分析 DTW_CSM 方法的准确性。表 3 和表 4 分别列出了 DTW_SS 方法和 DTW_CSM 方法在精确匹配时按 DTW 距离排序得到的前 3 个相似序列，可知两种方法的前 3 个相似序列有 2 个是相同的，说明了 DTW_CSM 方法的有效性。

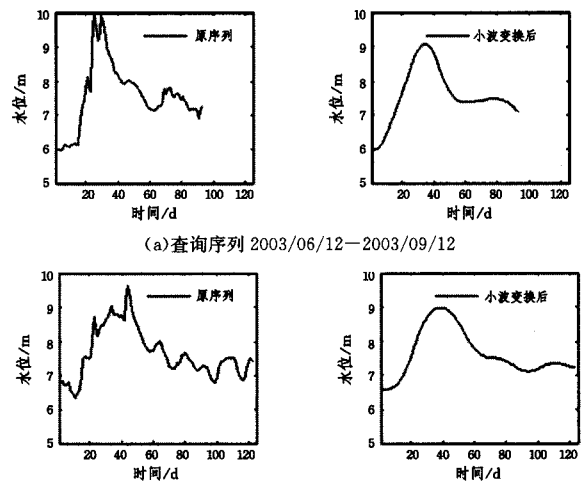
表 3　DTW_SS 方法匹配结果

起始日期	结束日期	DTW 距离
1983 年 6 月 10 日	1983 年 10 月 19 日	0.7592
1999 年 4 月 26 日	1999 年 9 月 30 日	1.9604
1996 年 6 月 20 日	1996 年 9 月 19 日	2.8124

表 4　DTW_CSM 方法匹配结果

起始日期	结束日期	DTW 距离
1983 年 6 月 10 日	1983 年 10 月 19 日	0.7592
1999 年 4 月 26 日	1999 年 9 月 30 日	1.9604
1991 年 5 月 13 日	1996 年 8 月 6 日	2.8650

此外，DTW_CSM 方法查询得到的最相似水文时间序列与 DTW_SS 方法的一致，说明了 DTW_CSM 方法的准确性。图 4 是查询序列在两种方法下的查询结果。



(a)查询序列 2003/06/12—2003/09/12

(b)查询序列 1983/06/10—1983/10/19

图 4　相似水文时间序列查询结果

综上所述，DTW_CSM 方法与 DTW_SS 方法的查询结果相当，但是 DTW_CSM 方法通过基于 DTW 距离的凝聚层次聚类方法对单一语义符号做进一步划分，在查询过程中有效地缩小了候选集，从而提高了查询效率。

结束语　水文时间序列相似性挖掘在水文研究领域具有重要意义。本文在已有方法的基础上，提出了一种基于 DTW 聚类的水文时间序列相似性挖掘方法，并以六合水库数据对该方法进行了实验验证，证明其有效地缩小了候选集，提高了相似水文时间序列查找效率，具有较好的准确性。此外，该方法通过对单一语义的符号进行聚类，在保留子序列语义的前提下实现离散化，对于挖掘有意义的水文时间序列时序关联规则具有重要意义。如何进一步提高符号序列聚类准确率以及实现水文时间序列的增量聚类，是下一步的研究重点。

参 考 文 献

[1]　Wang Ji-min, Zhu Yue-long, Li Wei, et al. Multi-measure Similarity Analysis of Hydrological Time Series[J]. Journal of China

Hydrology,2014,34(4):15-20(in Chinese)
王继民,朱跃龙,李薇,等. 多度量水文时间序列相似性分析[J]. 水文,2014,34(4):15-20

[2] Serrà J, Li J, Arcos. Anempirical evaluation of similarity measures for time series classification[J]. Knowledge-Based Systems,2014,61(9):305-314

[3] Li Wei,Sun Hong-lin. Analysis and study on hydrological time series similarity search[J]. Journal of China Hydrology,2009,29(6):76-80(in Chinese)
李薇,孙洪林. 水文时间序列相似性查询的分析与研究[J]. 水文,2009,29(6):76-80

[4] Gu Xin-chen,Wan Ding-sheng,Fan Long. Research and Application of Hydrological Time Series Similarity Based on Hadoop [J],Computer & Digital Engineering,2014,42(11):1-5(in Chinese)
顾昕辰,万定生,樊龙. 基于 Hadoop 的水文时间序列相似性研究与应用[J]. 计算机与数字工程,2014,42(11):1-5

[5] Cheng Xi-feng, Wan Ding-sheng, Wang Ya-ming. Similarity search optimization algorithm in hydrological time series[J]. Computer Engineering and Design,2013,34(11):4046-4050(in Chinese)
程习锋,万定生,王亚明. 水文时间序列相似性查询优化算法[J]. 计算机工程与设计,2013,34(11):4046-4050

[6] Lin J,Keogh E,Lonardi S,et al. A symbolic representation of time series, with implications for streaming algorithms[C]// Proc of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. New York:ACM Press, 2003:2-11

[7] Yan Qiu-yan,Meng Fan-rong. A key point based SAX improving algorithm[J]. Journal of Computer Research and Development, 2009,46(Suppl.):483-490(in Chinese)
闫秋艳,孟凡荣. 一种基于关键点的 SAX 改进算法[J]. 计算机研究与发展,2009,46(Suppl.):483-490

[8] Zhu Yue-long,Wang Yong-mei,Wan Ding-sheng,et al. Similarty mining of hydrological time series based on semantic similarity measures[J]. Journal of China Hydrology,2011,31(1):35-40(in Chinese)
朱跃龙,王咏梅,万定生,等. 基于语义相似的水文时间序列相似性挖掘[J]. 水文,2011,31(1):35-40

[9] Li Ying. Symbolization time series clustering based on DTW[J]. Microcomputer & Its Applications,2011,30(18):3-5(in Chinese)
李迎. 一种基于 DTW 的符号化时间序列聚类算法[J]. 微型机与应用,2011,30(18):3-5

[10] Ouyang Ru-lin,Ren Li-liang,Zhou Cheng-hu. Similarity search in hydrological time series[J]. Journal of Hohai University (Natural Sciences),2010,38(3):241-245(in Chinese)
欧阳如琳,任立良,周成虎. 水文时间序列的相似性搜索研究[J]. 河海大学学报(自然科学报),2010,38(3):241-245

[11] Ouyang Ru-lin,Ren Li-liang,Cheng Wei-ming,et al. Similarity search and pattern discovery in hydrological time series data mining[J]. Hydrol. Process,2010,24(9):1198-1210

[12] Zhang Xiao-hang,Liu Jia-qi,Du Yu,et al. A novel clustering method on time series data[J]. Expert Systems With Applications,2011,38(9):11891-11900

[13] Zhu Yue-long,Peng Li,Li Shi-jin,et al. Research on hydrological time series motifs mining[J]. Journal of Hydraulic Engineering,2012,43(11):1422-1430(in Chinese)
朱跃龙,彭力,李士进,等. 水文时间序列模体挖掘[J]. 水利学报,2012,43(11):1422-1430

(上接第 234 页)

[5] Zhou Jin-zhu, Huang Jin. Multiple kernel liner programming support vector regression incorporating prior knowledge[J]. Acta Automatica Sinica,2011,37(3):360-370(in Chinese)
周金柱,黄进. 集成先验知识的多核线性规划支持向量回归[J]. 自动化学报,2011,37(3):360-370

[6] Khemchandani R,Chandra S. Twin support vector machines for pattern classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2007,29(5):905-910

[7] Shao Y H,Deng N Y,Yang Z M,et al. Probabilistic outputs for twin support vector machines[J]. Knowledge-Based Systems, 2012,33:145-151

[8] Shao Y H,Deng N Y,Yang Z M. Least squares recursive projection twin support vector machine for classification[J]. Pattern Recognition,2012,45(6):2299-2307

[9] Peng X. TSVR:an efficient twin support vector machine for regression[J]. Neural Networks,2010,23(3):365-372

[10] Shao Y H,Zhang C H,Yang Z M,et al. An ε-twin support vector machine for regression[J]. Neural Computing and Applications,2013,23(1):175-185

[11] Peng X. Primal twin support vector regression and its sparse approximation[J]. Neurocomputing,2010,73(16):2846-2858

[12] Lu G F,Zou J,Wang Y. Incremental learning of complete linear discriminant analysis for face recognition[J]. Knowledge-Based Systems,2012,31:19-27

[13] Deguchi T, Ishii N. On Memory Capacity in Incremental Learning with Appropriate Refractoriness and Weight Increment[C]// 2011 First ACIS/JNU International Conference on Computers, Networks,Systems and Industrial Engineering (CNSI). IEEE, 2011:427-430

[14] Ross D A,Lim J,Lin R S,et al. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008,77(1-3):125-141

[15] Khreich W,Granger E,Miri A,et al. A survey of techniques for incremental learning of HMM parameters[J]. Information Sciences,2012,197:105-130

[16] Yuan Ya-xiang,Sun Wen-yu. Optimization theory and methods [M]. Beijing:Science Press,1997(in Chinese)
袁亚湘,孙文瑜. 最优化理论与方法[M]. 北京:科学出版社,1997

[17] Zhang Xiao-yun,Liu Yun-cai. Performace analysis of support vector machines with Gauss kernel[J]. Computer Engineering, 2003,29(8):22-25(in Chinese)
张小云,刘允才. 高斯核支撑向量机的性能分析[J]. 计算机工程,2003,29(8):22-25