一种面向不确定数据流的模体发现算法

王 菊¹,刘付显¹,靳春杰²,李祯东³

(1. 空军工程大学防空反导学院 西安 710051; 2.93527部队 河北 张家口 075000; 3.93787部队 北京 丰台区 100076)

【摘要】借鉴生物信息学中序列模式发现思想,提出了基于MEME(multiple expectation-maximization for motif elicitation) 的不确定数据流模体发现算法。该算法根据不确定数据流的特点,设计了不确定滑动窗口的简化计算方法,改进了 SAX(symbolic aggregate approximation)的符号化策略,用防空反导情报传感器网络中的一组不确定数据流验证了其可行性,通 过植入不同数目模体的方法测试了其准确性,并在元组存在概率为1的条件下与已有算法进行比较,验证其有效性。

关 键 词 MEME算法; 模体发现; SAX; 不确定数据流; 不确定滑动窗口

中图分类号 TP391 文献标志码 A doi:10.3969/j.issn.1001-0548.2017.01.013

New Motif Discovery Algorithm for Uncertain Data Stream

WANG Ju¹, LIU Fu-xian¹, JIN Chun-jie², and LI Zhen-dong³ (1. Air and Missile Defense College, Air Force Engineering University Xi'an 710051; 2. 93527 Military Unit Zhangjiakou Heibei 075000; 3. 93787 Military Unit Fengtai Beijing 100076)

Abstract A new MEME-based motif discovery algorithm for uncertain data stream is proposed by using the idea of sequential pattern discovery in bioinformatics. According to features of uncertain data stream, the new algorithm designs a simplified calculation method for uncertain sliding window and modifies the SAX symbolic strategy. The feasibility of the proposed algorithm is verified by one uncertain test data stream from air and missile defense sensors. And its accuracy is measured through planting different number motifs. Furthermore, the proposed algorithm is validated by comparing with existing algorithms in the condition that the existence probability of tuples is set to 1.

Key words MEME algorithm; motif discovery; SAX; uncertain data stream; uncertain sliding window;

随着信息化时代的到来,通信、传感器和计算 机等技术发展迅猛,使得各类测量数据急剧膨胀, 催生出一个发展广阔且军事意义重大的应用研究领 域——数据流分析。作为一个连续到达的数据序列, 与传统时间序列相比,数据流具有无界性、高速性 等显著特点^[1]。携带有不完整、不精确信息的数据 流被称为不确定数据流,它在无线传感器网络、互 联网、战场态势等领域有极大的应用需求^[2-3]。目前, 有关不确定数据流的研究主要包括聚类、查询和分 类等方面,而关于不确定数据流内隐含的关系、规 则及模式等知识挖掘方面则很少提及^[4-6]。

本文借鉴生物信息学中序列模式发现思想,研 究不确定数据流中功能或行为相似的流段(模体),用 于分析或预测产生数据流的实体行为。为了发现不 确定数据流在不同时刻滑动窗口下的模体,实现对 不确定数据流的预测、规则挖掘、分类和异常检测, 设计了基于MEME的不确定数据流模体发现算法, 在传统时间序列模体发现的基础上,增加了处理不 确定性和动态性的功能。

1 MEME算法

在生物信息学中,模体发现的近似算法主要分为两类:一类是基于启发式或贪心技术的算法,主要有WEEDER、VINE、Pattern Branching算法等; 另一类是基于统计技术的算法,主要有EM、MEME、 吉布斯采样以及HMM算法等。

MEME算法^[7-9]是目前最流行的符号序列集合 模体发现算法,它可将模体从由背景成分和模体成 分组成的混合符号序列中辨识出来。对于由符号序 列 $S_i = s_{il}s_{i2}\cdots s_{il}$ (*i* = 1,2,…,*t*)组成的符号序列集合 $S = \{S_1, S_2, \cdots, S_t\}$ 而言,用*l*-mer表示一个长度为*l* 的碱基片段,即*l*-mer= $s_{i,j+1}s_{i,j+2}\cdots s_{i,j+l}$ 。MEME算

收稿日期: 2015-12-23; 修回日期: 2016-06-01

基金项目: 国家自然科学基金(61272011)

作者简介: 王菊(1991-), 女, 博士生, 主要从事数据挖掘方面的研究.

法就是从符号序列集合中识别出重复出现的 *l*-mer。

MEME算法的核心是定义了一个二元随机变量 $Z_{ij}(j=1,2,...,L_i)$,通过计算每一个l-mer的似然值 来寻找模体,具体见文献[5]和文献[7]。其中, Z_{ij} 表 示每一个l-mer 对应的背景成分或模体成分,即当 字符 s_{ij} 表示为一个结合位点时, $Z_{ij}=1$,否则 $Z_{ij}=0$ 。该算法将整个序列集合的似然值表示为:

$$\log(p(X, Z \mid \theta_0, \Theta)) = \sum_{i=1}^{t} \sum_{j=1}^{L_i - l + 1} Z_{ij} \times \log p(X \mid Z, \theta_i, \Theta)$$
(1)

式中, X_{ij} 表示第*i*行第*j*个*l*-mer; θ_0 表示背景分 布(此处采用零阶马尔科夫模型,即假设字符各自独 立的分布); $\Theta = (\theta_{lk}, \dots, \theta_{wk}, \dots, \theta_{lk})$ ($k \in \Omega$) 表示模体 分布; θ_{wk} 表示字符 *k* 在模体第 *w* 个位置出现的概 率。期望最大化算法正是通过更新潜在的隐变量 *Z* 使得似然值最大化,主要过程分为*E*步和*M*步。

E步可以表示为:

$$Z_{ij}^{(T)} = \frac{p(X | Z_{ij}, \Theta^{(T)})}{\sum_{L_i - l + 1} p(X | Z_{ij}, \Theta^{(T)})}$$
(2)

M步可以表示为:

$$\Theta^{(T+1)} = \arg \max \mathbb{E}[\log p(X, Z | \Theta^{(T)})]$$
(3)

E步和M步重复执行多次直至收敛。此外,通过描述 位点如何分布,MEME又将概率模型细分为OOPS、 ZOOPS和TCM这3类。OOPS模型表示每条序列中有 且只有一个模体出现,这是模体发现问题最基本的 假设;ZOOPS模型表示每条序列中含有一个或零个 模体;TCM模型允许一条序列中有零或多个模体 出现。

2 基于MEME的不确定数据流模体 发现算法

基于MEME的不确定数据流模体发现算法的核 心是如何将具有时序性、海量性、无界性、实时性、 高速性和连续性等特点的数据序列合理地转换为符 号序列集合。它的难点是如何对有效时间内的不确 定数据流进行合理划分以及对具有不确定性的数据 序列进行符号化。因此,在对不确定滑动窗口技术 和SAX符号化策略进行改进和扩展的基础上,本文 综合利用这两种思想的优点,提出了一种适用于不 确定数据流的符号化方法。利用MEME算法对符号 化的不确定数据流进行模体发现,进而解决不确定 数据流模体发现问题。

根据可能世界模型^[10],利用一组不确定数据流 U,假设各元组可以在离散域 B 中取多个规范化 值^[4],则该数据流可以被描述为:

 $(\langle t_1, i_1, p_1 \rangle, \langle t_2, i_2, p_2 \rangle, ..., \langle t_j, i_j, p_j \rangle, ...)$ (4) 式中, t_j 是元组到达的时刻; p_j 表示元组的存在概 率; $i_j \in B$ 。由式(4)可知,在对不确定数据流进行建 模与规范化的基础上,文中首先采用滑动窗口对有 效时间内的不确定数据流进行划分,然后对满足一 定条件的数据进行符号化,最后执行MEME算法, 进行模体发现。

2.1 不确定滑动窗口定义及简化计算

考虑到不确定数据流具有无限性,要处理从数 据出现到当前时刻的所有数据是不可能的。基于此, 文献[11]提出了不确定滑动窗口的概念,其核心思想 是根据一定的置信概率α使滑动窗口中存在的元组 数目为w,但该方法在应用中需要进行大量的复杂 概率计算。因此,本文提出了不确定滑动窗口的简 化计算方法,该方法可以大幅度减少更新滑动窗口 时的计算量,提高处理速度。

2.1.1 不确定滑动窗口的定义

为了便于观察,不确定滑动窗口的符号说明及 定义如表1所示。

表1 符号说明

符号名称	含义		
U	某不确定数据流		
W	滑动窗口的期望长度		
w'	滑动窗口的实际长度		
α	置信概率		
m	U 中最新元组的索引值		
W(U,w)	针对U的长度为w的滑动窗口		
$\hat{W}(U,w)$	W(U,w)的子窗口		
$ \hat{W}(U,w) $	ŵ(U,w) 中存在的元组数目		

定义 1 不确定滑动窗口:滑动窗口的实际长度 w'具有不确定性。因此,该窗口被定义为不确定 滑动窗口,且满足:

$$\begin{cases} \min & w' \\ \text{s.t.} & P(|\hat{W}(U, w')| \ge w) \ge \alpha \end{cases}$$
(5)

当新元组到来时,更新滑动窗口的计算流程 如下。

输入: *U*,*w*,*α*

输出: W(U,w',α)

 $W(U,w',\alpha)$ 被赋予空集

Loop

If (U中的新元组 u 到达) then $W(U,w',\alpha) \leftarrow W(U,w',\alpha) \cup \{u\}$ While $P(|\hat{W}(U,w'-1)| \ge w) \ge \alpha$ do $W(U,w',\alpha) \leftarrow W(U,w',\alpha) \setminus \{u'\}$ $(u' 是窗口W(U,w',\alpha) 中最早的元素)$ End while

End if

End loop

2.1.2 不确定滑动窗口的简化计算方法

从前面可以看出,不确定滑动窗口更新过程的 核心是 *P*(|*ŵ*(*U*,*w*')|≥*w*)≥α 的计算。基于此,根据 不确定数据流中元组的存在概率,本文提出了一种 不等式简化计算方法,其具体实现方法如下。

对于任意元组 u_i ,设定当随机变量 $X_i = 0$ 时, 表示 u_i 存在,存在概率为 $1 - p(u_i)$;当 $X_i = 1$ 时,表 示 u_i 不存在,不存在概率为 $p(u_i)$,即 X_i 服从以 $p(u_i)$ 为参数的[0,1]分布,因而有:

$$P(|\hat{W}(U,w')| \ge w) \ge \alpha \Leftrightarrow P\left(\sum_{i=1}^{w'} X_i \ge w\right) \ge \alpha \quad (6)$$

若
$$Y = \sum_{i=1}^{w} X_i$$
,则有:
 $P\left(\sum_{i=1}^{w'} X_i \ge w\right) \ge \alpha = P(Y \ge w) \ge \alpha$ (7)

曲于
$$P(Y = w) = \sum_{A \in F_w} \prod_{u_i \in A} p(u_i) \prod_{u_i \notin A} (1 - p(u_i))$$
, 因

此有:

$$P(Y \ge w) = 1 - \sum_{i=0}^{w-1} P(Y=i)$$
(8)

式中, F_w 表示在长度为 w'的滑动窗口中出现 w 个有效元组的所有可能组合; $A \neq F_w$ 中的一种组合方式。为了简化计算,文中采用均值为 $\mu = E[Y] = \sum_{i=1}^{w'} p(u_i)$,标准差 $\sigma = \sqrt{Var(Y)} = \sum_{i=1}^{w'} p(u_i) \times (1 - p(u_i))$ 的高斯近似^[12]来计算该概率,即满足:

$$P(Y \ge w) \approx 1 - \Phi\left(\frac{w + 0.5 - \mu}{\sigma}\right) \tag{9}$$

2.2 改进的SAX符号化策略

SAX^[13-14]是一种针对一元时间序列符号化的方法,可以用于聚类、分类、索引和异常检测等。为了将该方法拓展到不确定数据流,确保每个滑动窗口中存在元组的数目差别不大,本文将SAX的符号

化策略改进为以下3个步骤。

1) 根据2.1节的分析,在长度为w'的滑动窗口中,以置信概率α抽取该窗口中存在的w个元组,用向量 **D** = d₁,d₂,…,d_w进行表示;

2) PAA过程,即用一小段序列的平均值代表该 序列:

3) 符号化,即用不同的符号表示每小段的平均值。

上述步骤中,步骤1)的目的是为了确保每个滑动窗口中存在元组的数目近似为w。步骤2)是将原本的w维时间序列降为m维,即用一组向量 $\bar{D} = \bar{d}_1, \bar{d}_2, \dots, \bar{d}_m$ 表示原来的时间序列。其中,第*i*个元素的计算公式为:

$$\overline{d}_i = \frac{m}{w} \sum_{j=\frac{w}{m}(i-1)+1}^{\frac{w}{m}i} d_j$$
(10)

如图1所示,采用PAA过程后,一个128维的时间序列(浅线条)可以用的8维时间序列(粗线条)进行 表示。



图1 PAA过程示例

步骤3)的目的是为了选取合适的字符集进行符 号化,其关键是分界点的选取,文献[14]中给出了常 用的分界点值及所需字符集的数目,如表2所示。

表2 常用的分界点值及所需字符集的数目

β_i^{lpha}	3	4	5	6	7	8	9	10
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
eta_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
eta_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
$eta_{_6}$					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
$\beta_{_8}$							1.22	0.84
β_9								1.28

对于图1中所示的时间序列,当用3个字符进行 表示时,该序列的符号化情况如图2所示,符号化结 果为baabccbc。



2.3 算法设计

综合2.1节的方法和2.2节的改进策略,本节设计 了基于MEME算法的不确定数据流模体发现算法, 该算法可以对任意一组不确定数据流进行模体发 现,流程图如图3所示。





具体算法描述如下。

输入: U, w, α, \max _size

输出: $M = \{M_1, M_2, \dots\} / / 发现的模体集合$

- 1) 初始化参数 $w, \alpha, \max_{\text{size}}$, 设置 n = 0;
- 判断不确定数据流是否到来,是转步骤3), 否则算法结束;
- 3) 对*U*进行建模及规范化,得到: $U' = \{u'_1, u'_2, \dots\} = \{\langle t_1, i_1, p_1 \rangle, \langle t_2, i_2, p_2 \rangle, \dots\};$
- 4) 不确定滑动窗口的更新及数据缓存: LOOP
 - If (U'中的新元组u'到达) then $W(U',w',\alpha) \leftarrow W(U',w',\alpha) \cup \{u'\}$; If $P(|\hat{W}(U',w'-1)| \ge w) \ge \alpha$ then $S_n = \{u'_1,u'_2,\dots,u'_{w'}\}$; $W(U',w',\alpha)$ 被赋予空集;

$$n = n + 1$$
:

If $n = \max_{\text{size}}$ then

$$n=0;$$

另存
$$S_1, S_2, \dots, S_{\max_size}$$
为 S'_1, S'_2, \dots, S'_n ,

$$\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_{\text{max}_{\text{size}}}$$

转步骤5; Else

continue:

```
End if
```

Else

continue;

End if

End if

- 5) 对序列集 S'_1, S'_2, ..., S'_max size 进行符号化;
- 执行MEME算法,存储并输出所发现的模体 *M*₁,*M*₂,…;
- 7) 释放 S'₁, S'₂,…, S'_{max_size} 所占用的存储空间; End loop

综上所述,由于采用了并行计算和不确定滑动 窗口简化计算方法,文中算法计算时间较短,运算 效率高。

3 实验与分析

3.1 实例验证

防空反导情报传感器网络是军事中产生不确定 数据流的典型应用,本文对该网络中产生的一组距 离不确定数据流进行模体发现。

1) 设置窗口中有效元组数目w=100,置信概 率 $\alpha = 0.95$, n = 0, max_size = 3。

 用可能世界模型对距离不确定数据流进行 建模,再通过(x-μ)/σ对其规范化后,其部分数据 如表3所示。

表3 部分规范化不确定数据流示例

时间	规范化后的距离值	存在概率	
t_1	0.246 75	0.5	
t_2	0.122 22	0.9	
t_3	1.109 82	0	
t_4	-0.705 6	0.6	
:	:		

3) 继续采集诸如表3中的数据,根据存在概率 判断是否满足不等式 $P(|\hat{W}(U',w'-1)| \ge w) \ge \alpha$,直 到 $n = \max_$ size,得到 $S'_1, S'_2, \dots, S'_{\max_size}$ (具体如图4所 示,黑点表示元组存在,白点表示元组不存在),返 回初始状态。



b. S', 中规范化后的距离值

MOTIF LOCATIONS



图4 S'₁, S'₂, S'₃的图形表示

4) 对 *S*'₁,*S*'₂,…, *S*'_{max_size}进行符号化,字符集采用 {*A*,*G*,*C*,*T*},具体见2.2节的分析。根据图2中分界点 值的划设规范,确立3个分界点:-0.67、0和0.67, 设置 *m*=2,得到符号序列集合如表4所示。

表4 *S*₁',*S*₂',*S*₃'符号化结果

编号	上述三条综合指标序列所对应的字符串
seq1	CCCCAGTGCCGGCGGCCCCCCCGGAGGCCC GGGGGGCCGGCGACAAACGG
seq2	TTCCCCCTGCGAAGCCGGACCTAGGCGGCCG GGTCGCGCTACTCGCGTCG
seq3	CGCCCGGCCGGGAGGGCGGCGCGCGCGCC GCACCCGGCCCGGCCGCCCC

5) 根据MEME算法,采用DNA序列分析工具, 对表4中的符号序列集合进行模体发现,其结果如图 6所示。

由图5可知,在表3所示规范化后的距离值中, 共发现了两种模体,分别由两种粗细的线条进行表 示。进而根据所发现模体的排序及位置(如图5中方 框内所示),可以反推出原数据中模体的位置及形 状,如图6所示。

 Only Motif Sites O Motif Sites/Scan O All S
 Name ? p-value ? Motif Location 1. seq1 2.09e-2 1.14e-2 2. seq2 3. seq3 7.01e-1 CGGCGC Motif 7 value 5.6 Start 17 End 22 / 5 650-3 MEME versi 4.10.2 (Re date: Thu Sep 03 15:00:54 2015 -0700) 图5 模体发现结果 1.5 1.5 规范化后的距离值 规范化后的距离值 1.0 1.0 0.5 0.5 0 0 -0.5 -0.5 -1.0-1.0-15 -1.5^{L}_{0} 1.0 1.5 2.0 0.5 1.0 2.0 0 1.5 时间/s 时间/s a. S' 中规范化后的距离值及模体位置与形状 b. S', 中规范化后的距离值及模体位置与形状







图6 模体在原数据中的位置及形状

从图6可发现,所发现模体在形状上具有一定的 相似性,满足"模体"在时间序列中的定义,也验 证了文中算法的可行性。

6) MEME算法执行完毕后,释放出 S'₁, S'₂, S'₃ 所占 用的存储空间,并存储所发现的模体。

3.2 实验分析

通过3.1节的案例分析,文中算法的可行性得到 验证。为进一步分析文中算法的有效性,本节设计 了两个实验来进行验证。

实验1: 若对不确定数据流进行模体植入,当植入模体的数目依次为 x = 1,10,20,…,200 时,本文算法所发现模体的准确率如图7所示。



图7 本文算法所发现模体的准确率

从图7可知,在不同的模体植入情况下,文中算 法的模体发现准确率在0.8~0.9之间,稳定性强,验 证了该算法的有效性。

实验2:考虑到目前还没有关于不确定数据流模体发现的相关算法,文中设定不确定数据流中元组的存在概率都为1,设置滑动窗口长度为w=200。此时假设滑动窗口不更新,将不确定数据流转换为确定的时间序列。在相同的3个白噪声数据集上依次植入随机数目的模体,得到3组实验数据集,将文中算法与文献[15-16]提出的MK和MOEN算法进行比较,其模体发现准确率结果如表5所示。

表5 算法准确率比较

算法	数据集			
	数据集1	数据集2	数据集3	
MK	0.87	0.85	0.82	
MOEN	0.90	0.86	0.85	
本文算法	0.91	0.89	0.88	

从表5可知,文中算法的模体发现准确率高于 MK和MOEN算法,进一步验证了该算法的有效性。

4 结束语

本文在传统时间序列模体发现的基础上加入了 不确定性和动态性,建立起了序列数据挖掘和不确 定数据流挖掘之间的桥梁,并采用生物信息学算法 完成了对不确定数据流的模体发现。

1) 提出了基于MEME的不确定数据流模体发现算法,根据防空反导传感器网络对距离的实时测量数据进行模体发现,验证了其可行性;

2) 通过多次模体植入实验和算法性能对比实验,验证了本文算法的有效性。仿真分析表明,在同等仿真条件下,本文算法优于MK和MOEN算法;

3) 该文部分内容属于探索性的研究,相关理论和研究可以对不确定数据流的模体发现提供理论和应用支撑;

4)本文所建立的不确定数据流是基于离散属 性值,对具有连续属性的不确定数据流进行模体发 现是本文下一步的研究内容。

参考文献

- 梁春泉. 不确定数据流分类算法研究[D]. 西安: 西北农 林科技大学, 2014.
 LIANG Chun-quan. Classification algorithm based on uncertain data stream[D]. Xi'an: Northwest Agriculture and Forestry University, 2014.
- [2] THANH T L T, PENG L P, DIAO Y L, et al. CLARO: Modeling and processing uncertain data streams[J]. VLDB Journal, 2012, 21: 651-676.
- [3] JIN C Q, JEFFREY X Y, ZHOU A Y, et al. Efficient clustering of uncertain data streams[J]. Knowl Inf Syst, 2014, 40: 509-539.
- [4] 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(12): 1422-1430.
 ZHU Yue-long, PENG Li, LI Shi-jin, et al. Research on hydrological time series mining[J]. Hydraulic Engineering, 2012, 43(12): 1422-1430.
- [5] PUNEET A, GAUTAM S, SARMIMALA S, et al. Efficiently discovering frequent motifs in large-scale sensor data[EB/OL]. [2015-06-30]. https://www.researchgate.net/ publication/270454309_Efficiently_Discovering_Frequent_ Motifs_in_Large-scale_Sensor_Data.

- [6] 邹力鹍,张其善. 基于多最小支持度的加权关联规则挖掘算法[J]. 北京航空航天大学学报,2007,33(5):590-593. ZOU Li-pu, ZHANG Qi-shan. Algorithm of weighted association rules mining with multiple minimum supports [J]. Beijing University of Aeronautics and Astronautics Technology, 2007, 33(5): 590-593.
- [7] 张慈璞, 霍红卫, 于强, 等. 用于转录因子结合位点识别的定位投影求精算法[J]. 计算机学报, 2013, 36(12): 2545-2559.
 ZHANG Yi-pu, HUO Hong-wei, YU Qiang, et al. A novel fixed- position projection refinement algorithm for TFBS Identification[J]. Journal of Computers, 2013, 36 (12): 2545-2559.
- [8] TIMOTHY L B. Dreme: Motif discovery in transcription factor ChIP-seq data[J]. Original Paper, 2011, 17(12): 1653-1659.
- [9] DANIEL Q, XIE X H. Extreme: an online EM algorithm for motif discovery[J]. Original Paper, 2014, 30(12): 1667-1673.
- [10] 李明,张维明. 不确定数据流多维建模方法[J]. 国防科技大学学报, 2014, 36(5): 174-179.
 LI Ming, ZHANG Wei-ming. Multi-dimensional modeling method of uncertain data stream[J]. Journal of the National Defense University, 2014, 36 (5): 174-179.

- [11] MICHELE D. Modeling and querying data series and data streams with uncertainty[D]. The Autonomous Province of Trento: Universita' degli Studi di Trento, 2014,
- [12] HONG Y. On computing the distribution function for the sum of independent and non-identical random indicators[EB/OL]. [2015-10-10]. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2 20.8708.
- [13] 曲文龙,张克君,杨炳儒,等.基于奇异事件特征聚类的时间序列符号化方法[J].系统工程与电子技术,2006,28(8):1131-1134.
 QU Wen-long, ZHANG Ke-jun, YANG Bing-ru, et al. Time series symbolization based on singular event feature clustering[J]. Systems Engineering and Electronics, 2006, 28 (8): 1131-1134.
- [14] LIN J, KENOGH E J, WEI D L, et al. Experiencing SAX: a novel symbolic representation of time series[J]. Data Min Knowl Disc, 2007, 15: 107-144.
- [15] MUEEN A, KEOGH E J, ZHU Q, et al. Exact discovery of time series motif[C]//Society for Industrial and Applied Mathematics Conf. on Data Mining. [S.I.]: Springer, 2009.
- [16] ABDULLAH M, NIKAN C. Enumeration of time series motifs of all lengths[J]. Knowl Inf Syst, 2015, 45: 105-132.

编辑叶芳