

分层自组织耦合振荡聚类方法

闫德勤¹ 迟忠先²

(辽宁师范大学计算机系 大连116029)¹ (大连理工大学计算机系 大连116026)²

Layer Self-Organization of Pulse-Coupled Oscillators Clustering Method

YAN De-Qin¹ CHI Zhong-Xian²

(Department of Computer Science, Liaoning Normal University, Dalian 116029)¹

(Department of Computer Science, Dalian University of Technology, Dalian 116026)²

Abstract A self-organization of pulse-coupled oscillators for clustering method is studied and the defectiveness of the method is analyzed. With modification to the method a new clustering method is presented. Experiments indicate the modified method is effective.

Keywords Pulse-coupled oscillators, Cluster, Layer cluster

1 引言

在自然界中有很多自组织共振现象,如多个蟋蟀的鸣叫、萤火虫群的闪光等^[1]。近年来人们通过对这一现象的研究产生了一些研究与应用模型。文[1]给出了自组织耦合振荡模型应用于聚类分析的方法。该方法聚类快且不需要预先给定聚类个数,但由于该方法基于生物的感受共振因此对于类间的区分力不强,尤其是类间有断续分布的样本时易于出现聚类错误。作者在文[2]中提出了一种分层聚类算法(CLC),该算法具有较好的分层能力。本文通过修改的 CLC 算法改进了文[1]的聚类方法,改进的算法继承了两种算法的优势,同时又克服了文[1]中自组织耦合振荡聚类算法的不足。

2 自组织耦合振荡聚类方法

在自组织振荡模型^[1,4]中一个系统包含一定数量的元素,每一个元素(样本)都是一个振荡器(振子),振子通过周期振荡对别的振子产生(或吸引或排斥的)作用。通过一定(振动周期)相互作用的时间系统达到一种平衡,相互同步振动的振子形成一个同步组,不同组的相位恒定,构成锁相。对聚类而言,每一同步组为一类。设具有 N 个振子的系统为 $O = \{O_1, O_2, \dots, O_N\}$, 每个 O_i 表示一个元素的振子,其振动状态由相角 θ_i 和状态变量 x_i 描述。在 x_i 和 θ_i 在由状态和相位构成的直角坐标系中满足函数关系式:

$$x_i = f_i(\theta_i), i = 1, \dots, N \quad (1)$$

其中 f_i 为光滑单调递增由 $[0, 1]$ 映射到 $[0, 1]$ 的实值函数,满足 $f_i(0) = 0, f_i(1) = 1$ 。振子 O_i 在状态 x_i 时的相位是 θ_i , 当状态 x_i 沿着函数 f_i 的曲线由小到大向 1 运动时,相位 θ_i 也相应地在相位坐标轴上由小到大向 1 运动。当一个振子 O_i 到达 (1, 1) 点时就向其它的振子发出一个触发,然后振子 O_i 复位到原点 (0, 0), 在重新向 (1, 1) 运动。设振子 i 到达 (1, 1) 点时向振子 $O_j (j \neq i)$ 发出的触发量为 $\epsilon_i(\theta_j)$, 则振子 O_j 运动状态变为:

$$x_j(t^+) = B(x_j(t) + \epsilon_i(\theta_j)) \quad (2)$$

在式(2)中函数 B 保证 $x_j(t)$ 限制在 $[0, 1]$ 中, 即

$$B(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x < 0 \\ 1 & \text{if } x > 1 \end{cases}$$

在自组织振荡聚类中 $\epsilon_i(\theta_j)$ 和函数 f_i 一般取为:

$$\epsilon_i(\theta_j) = \begin{cases} C_E \left[1 - \left(\frac{d_{i,j}}{\delta_0} \right)^2 \right] & \text{if } d_{i,j} \leq \delta_0 \\ C_I \left[\left(\frac{d_{i,j} - 1}{\delta_0 - 1} \right)^2 - 1 \right] & \text{if } d_{i,j} > \delta_0 \end{cases} \quad (j \neq i) \quad (3)$$

其中, $d_{i,j}$ 为归一化的样本 i 与 j 之间的距离, δ_0, C_E 和 C_I 为选取的常数。 δ_0 表示耦合范围, C_E 表示最大偶合量, C_I 表示排斥耦合最大量。

$$f_i(\theta) = f(\theta) = \frac{1}{b} \ln[1 + (e^b - 1)\theta] \quad (4)$$

其中 b 为常数, 决定着函数 f_i 的弯曲程度。

文[1]给出了自组织振荡的聚类算法 SOON1 和 SOON2, 其算法实质是一样的, 结构为(SOON):

Step1. 计算各样本间的距离并归一化, 随机选取每一振子 O_i 的相位 θ_i 。

Step2. Repeat

找出下一个产生触发的振子

$= \{O_i : \theta_i = \max_{j=1, \dots, N} \theta_j\}$; 使相位 θ_i 运动到端点 ($\theta_i = 1$), 调整其它相位: $\theta_k = \theta_k + (1 - \theta_i), k = 1, \dots, N$ 。

Step3. For 所有振子 $O_j (j \neq i)$ Do

由式(4)计算 $x_j = f(\theta_j)$

由式(3)计算 $\epsilon_i(\theta_j)$

根据式(2)调整 x_j

计算新的相位 $\theta_j = f^{-1}(x_j)$

End For

Step4. 找出各同步振子并记为同一同步组并复位它们的相位。把本组振子合为一个振子, 调整本组振子相应元素与非本组振子对应元素的距离, 使这些距离为同一值。Until 所有同步组稳定。

Step5. 每一同步组记为一类, 结束。

计算表明 SOON 算法速度快, 但该算法对有些在类间有些断续分布的样本易于出现聚类错误(见图2, 图2为 SOON 算法对图1数据的聚类结果)。

3 分层自组织耦合振荡聚类算法

自组织耦合振荡模型基于元素间近距离吸引趋于同步,

闫德勤 博士, 副教授, 主要研究方向为模式识别、图像与信息处理等。迟忠先 教授, 博士生导师, 主要研究领域为知识发现、数据仓库、数据挖掘等。

远距离排斥而趋于异步。系统通过元素(振子)间的不断作用最终达到一个稳定状态。事实上,对自组织耦合振荡系统而言从初始到稳定的演化过程应是并行相互作用的过程。即相应的算法应考虑到系统的整体性。在 SOON 算法中,如果一些振子一旦同步则不再分开,这样就限制了一定的竞争机制。由 SOON 聚类算法可以看到基于生物感染的共振因此对于类间的区分力不强,同时在一些同步组出现后其振动作用以及与其它元素的距离问题也需要一定的研究。

基于系统的整体性考虑作者在文[2]中提出了一种分层聚类算法(CLC)。分层聚类算法基于以下考虑:(a) 各聚类中心对周围点具有一定的吸引力。(b) 在同一类中各点的分布具有一定的连续度。(c) 类似万有引力,大的聚类块吸引周围的小块。

自组织耦合振荡聚类算法具有很好的关联性而分层聚类算法具有很好的整体性。基于改进的自组织耦合振荡聚类算法与改进的分层聚类算法,本文给出新的分层自组织耦合振荡聚类算法。新算法由自组织耦合振荡产生同步组,再根据系统的整体性进行分层处理。反复进行以上过程直到系统稳定。

分层自组织耦合振荡聚类算法如下:

Step 1. 计算样本元素 x_i 与 x_j 的距离 d_{ij} , 并存 d_{ij} 于 b_{ij} , $B = (b_{ij})$ 是 $N \times N$ 矩阵。对距离归一化处理并保存 B 到 U 。

Step 2. 若没有同步组则随机选取每一振子 O_i 的相位 θ_i 。否则,找出距触发振子最近的同步组,并以该同步组为触发振子,相应调整其它振子及同步组的相位与距离。

Step 3. 调用 SOON 算法结构的 Step 3. 得到 c 个同步组。

Step 4. 选择初始标尺 $EP = mb/(c+1)$, 并取 $c=0$ 。

Step 5. 构造连接矩阵。

```
for 1 ≤ i ≤ N and 1 ≤ j ≤ N do
  if bij < EP then bij = 1
  else
    bij = 0
```

enddo

Step 6. 处理连接元。

```
for 1 ≤ i ≤ n-1 do
  if bij = 0 then i = i+1
  for i+1 ≤ j ≤ n do
    if bij < > 0 then
      第 j 行加到第 i 行
    enddo
    for i+1 ≤ j ≤ n do
      if bij < > 0 then
        第 j 列加到第 i 列
      enddo
      for i+1 ≤ j ≤ n do
        if bij < > 0 then
          把 xj 归入 xi 所在的一类。
        enddo
      enddo
    enddo
  enddo
  置第 i 行与第 j 列所有元素为 0。
  C = C+1
enddo
计算并处理相应的连接度。
```

Step 7. 调整聚类规模。

```
若 C > 2c 则 EP = alpha × EP。
若 C < c 则 EP = beta × EP。
U 送到 B 转 step 4。
否则, 计算聚类中心。
```

Step 8. 若聚类稳定则结束, 否则转 step 2。

4 实验结果

我们用 SOON 算法与本文给出的分层自组织耦合振荡聚类算法进行了对比实验。图1为二维分布的样本数据, 图2由 SOON 算法得出的聚类结果, 图3为分层自组织耦合振荡聚类算法得出的聚类结果。从聚类结果看, 新的算法克服了 SOON

算法的不足, 得到了正确的聚类结果。在实验过程中参数 δ_0 , C_E 和 C_I 选取, 特别是 δ_0 的选取对 SOON 算法有较大影响, 一般 C_E 和 C_I 取 0.1 和 0.05, δ_0 取 0.05 到 0.3 之间, 三个参数有一定的关联性。其中, δ_0 的取值对聚类效果影响最大, 不同的 δ_0 值会有差别大的聚类结果。对新算法而言参数的取值范围较宽, 参数取值在一定范围的变化对聚类结果影响不大, 即新算法具有较好的稳定性。

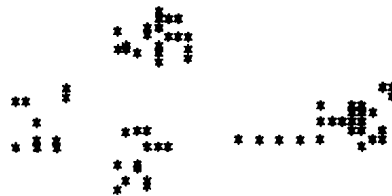


图1 原始数据

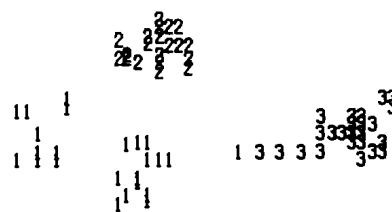


图2 SOON 算法聚类结果

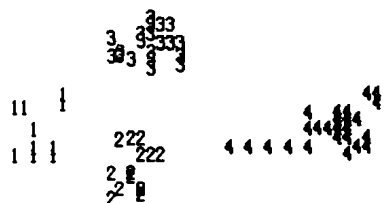


图3 新算法聚类结果

结束语 一些著名的模型和算法都来自生物与自然的启示, 如神经网络、遗传算法等。自组织耦合振荡模型是一种描述自然现象的方法, 并在生物、物理、化学与计算机科学等领域有着不断的应用与研究结果^[1]。自组织耦合振荡聚类方法即是一种新的应用。本文针对该算法所存在的问题进行了研究, 并提出了一种分层自组织耦合振荡聚类的改进算法。新算法增强了自组织耦合振荡的聚类能力与效果, 对进一步研究与应用具有意义。

参考文献

- 1 Rhouma M B H, Frigui H. Self-organization of pulse-coupled oscillators with application to clustering. IEEE Tran. PAMI, 2001, 23(2): 180~195
- 2 Deqin Y. A new connection layer clustering method. In: Proc. of the Sixth Intl. Conf. for Young Computer Scientist (ICYC-S'2001). Hangzhou, China, 2001. 331~333
- 3 郭桂蓉, 庄钊文. 信息处理中的模糊技术. 长沙: 国防科技大学出版社, 1996
- 4 Mirollo R E, Strogatz S H. Synchronization of pulse coupled biological oscillators. SIAM J. Applied Math., 1990, 50(6): 1645~1662