

# Self-Weighted Clustering With Adaptive Neighbors

Feiping Nie , Danyang Wu , Rong Wang , and Xuelong Li

## 摘要

许多现代聚类模型可以分为两个独立的步骤，即在样本上构建相似性图（SG），并基于 SG 将每个样本划分为相应的聚类。因此，学习合理的 SG 已成为聚类领域的一个热点问题。已经提出了许多以前的工作，重点是构建更好的 SG。然而，它们中的大多数都遵循一个理想的假设，即不同特征的重要性是相等的，这在实际应用中并不适用。为了缓解这一问题，本文提出了一种自适应邻居自加权聚类（SWCAN）模型，该模型可以为不同的特征分配权重，学习 SG，并同时样本划分为簇。在实验中，我们观察到 SWCAN 可以为不同的特征合理分配权重，并且在合成数据集和实际数据集上优于相比较的聚类模型。

**关键词：** 自适应邻域；块对角相似矩阵；聚类；加权特征

## 1 引言

聚类是机器学习和数据挖掘领域的一项基本技术，例如图像处理、人类动作识别和多任务/多视图学习。在过去几十年中，已经提出了许多聚类模型，包括 k 均值聚类、分层聚类、光谱聚类（SC）、最大边际聚类方法、支持向量聚类、基于图的聚类模型、自适应邻居聚类（CAN）[29] 和许多子空间聚类模型。然后，我们简要介绍了先前模型的核心思想；最广泛使用的模型是 k-means 聚类，它根据样本和质心之间的距离将每个样本分配给其最近的聚类质心。分层聚类是一种聚类分析方法，旨在建立聚类的层次结构。最大边缘聚类的目的是在低密度区域中找到一个决策边界，将数据点划分为两个不同的组。

## 2 相关工作

对于矩阵  $Q$ ， $Q$  的转置用  $Q^T$  表示， $Q$  的迹用  $\text{Tr}(Q)$  表示。对于向量  $q$ ，我们将  $\|q\|_2$  表示为  $q$  的 2-范数。 $q$  的期望和偏差分别由  $E(q)$  和  $D(q)$  表示。现在，我们在本文中给出一些具体的符号。我们定义  $x_i \in R^d$  作为每个样本和  $X \in R^{n \times d}$  作为数据矩阵，其第  $i$  行是向量  $x_i^T$ 。 $A \in R^{n \times n}$  是 SG（或相似矩阵），其每个元素  $a_{ij}$  表示  $x_i$  和  $x_j$  之间的相似性。 $D \in R^{n \times n}$  是一个对角矩阵，其第  $i$  个对角元素是  $\sum_j (a_{ij} + a_{ji})/2$ ， $L \in R^{n \times n}$  称为拉普拉斯矩阵，可通过  $L = D - (A + A^T)/2$  计算和  $S_t \in R^{n \times n}$  由  $XX^T$  计算， $XX^T$  可被识别为  $X$  的协方差矩阵的比例。 $I_n \in R^{n \times n}$  是一个  $n$  维单位矩阵。此外，我们表示向量  $\theta \in R^d$  作为特征的加权向量， $\Theta \in R^{d \times d}$  是一个对角矩阵，其主要对角向量是向量  $\theta$ 。我们将  $\sigma_i(L)$  表示为矩阵  $L$  的第  $i$  个最小特征值。

### 2.1 原始相似图

原始相似图（OSG） $A$  定义如下：

$$a_{ij} = \begin{cases} K(d(x_i, x_j)), & x_i \in \text{KNN}(x_j) \\ 0, & \text{otherwise.} \end{cases}$$

其中  $d(x_i, x_j)$  是  $x_i$  和  $x_j$  之间的距离， $K(\bullet)$  是核算子，例如高斯核、 $\text{poly}(\bullet)$  核和  $\cos(\bullet)$ 。如果我们设置最近邻居的数量  $k = n$ ，KNN 结构将是完全连通的。

## 2.2 光谱聚类

SC 模型利用 OSG 的拉普拉斯矩阵的谱（特征值）在聚类之前进行降维。为了实现这一目标，提出了 R-cut SC 模型<sup>[1]</sup>。由于 R-cut 导致 NP 难题，研究人员将其放宽为可解形式，可写成如下：

$$\min_H \text{Tr}(H^T L H) \quad s.t. \quad H^T H = I_k$$

其中  $H \in R^{n \times k}$  是单位矩阵， $k$  是降维。为了获得更鲁棒的身份矩阵  $H$ ，提出了 N-cut<sup>[2]</sup>，目标可以放宽为以下问题：

$$\min_H \text{Tr}(H^T L H) \quad s.t. \quad H^T D H = I_k$$

基于这两个 SC 模型，我们得到了一个单位矩阵  $H \in R^{n \times k}$ 。 $H$  的每一行都是一个新样本，聚类结果可以通过利用  $H$  的  $k$  均值来获得。

## 2.3 基于图的聚类

基于图的聚类模型旨在学习一种新的 SG，以在一些约束条件下近似 OSG。在这部分中，我们主要介绍了一种最先进的基于图的聚类模型，即约束拉普拉斯秩（CLR）聚类模型。CLR 有两个具有不同损失函数的版本，即 2 范数版本和 1 范数版本，它们可以分别写为：

$$\begin{aligned} \min_A \quad & \|A - A_0\|_F^2 \quad \text{or} \quad \min_A \quad \|A - A_0\|_1 \\ s.t. \quad & \sum_j a_{ij} = 1, \quad a_{ij} \geq 0, \quad \text{rank}(L) = n - c. \end{aligned}$$

由于学习的 SG  $A$  具有块对角结构，因此可以直接将其划分为  $c$  个子图。为了将基于图的模型推广到大规模数据中，提出了一种基于二分图的聚类模型。

## 2.4 具有自适应邻居的聚类

SC 和基于图的聚类模型的性能取决于 OSG 的质量。如果 OSG 更差，集群性能将受到很大影响。为了缓解这一问题，提出了一种具有自适应邻居（CAN）的聚类模型<sup>[3]</sup>，该模型直接从数据空间学习图。CAN 的目标定义为：

$$\begin{aligned} \min_A \quad & \sum_{i=1}^n \sum_{j=1}^n (\|x_i - x_j\|_2^2 a_{ij} + \gamma a_{ij}^2) \\ s.t. \quad & \forall i, \quad a_i^T = 1, \quad 0 \leq a_{ij} \leq 1, \quad \text{rank}(L) = n - c. \end{aligned}$$

学习的 SG  $A$  具有块对角性质，对 OSG 不敏感。此外，为了将 CAN 应用于高维数据，提出了 PCAN<sup>[29]</sup>。PCAN 的目标定义为：

$$\begin{aligned} \min_A \quad & \sum_{i=1}^n \sum_{j=1}^n (\|W^T x_i - W^T x_j\|_2^2 a_{ij} + \gamma a_{ij}^2) \\ s.t. \quad & \forall i, \quad a_i^T = 1, \quad 0 \leq a_{ij} \leq 1, \quad \text{rank}(L) = n - c, \quad W^T S_t W = I_k. \end{aligned}$$

其中  $W \in R^{d \times k}$  是投影矩阵， $k$  是降维，矩阵  $S_t$  被识别为  $X$  的协方差矩阵的比例。

### 3 自适应邻域的自加权聚类

#### 3.1 方法概述

自适应邻域的自加权聚类与其他聚类方法对比，如图 1所示：

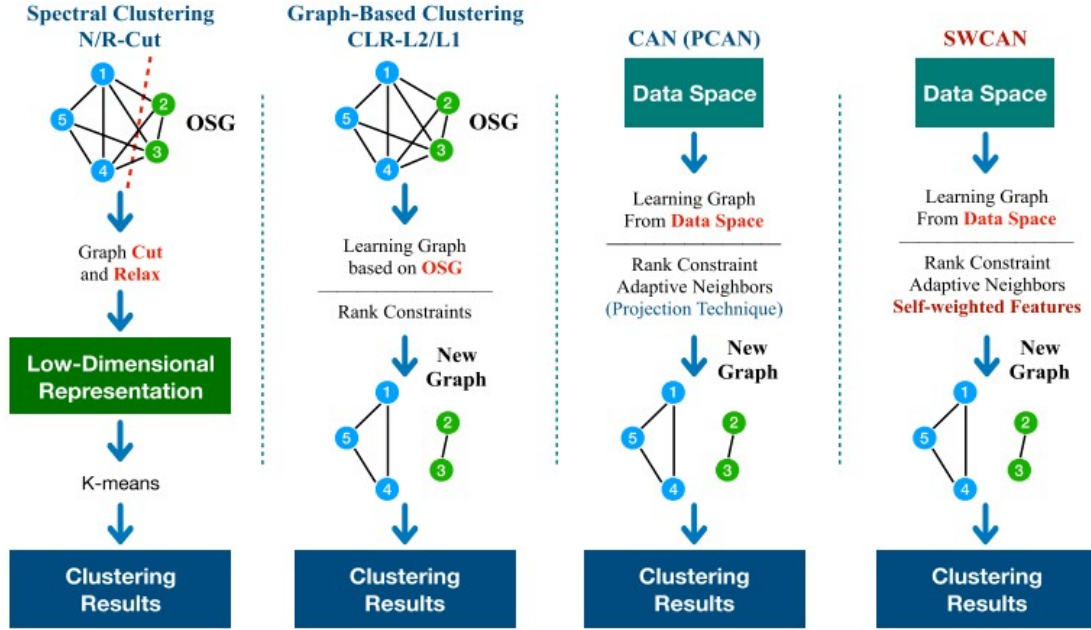


图 1: SC、基于图的集群、CAN、PCAN 和 SWCAN 的核心思想

在以前的工作中，每个特征同等重要的理想假设将使模型对于具有噪声特征的数据不具有鲁棒性。研究人员总是利用特征选择方法对降维模型在聚类之前解决有噪声的特征。

与特征选择和降维不同，在本文中，作者提出了一个简单而明亮的想法<sup>[4]</sup>。显然，如果我们对不同特征的重要性有先验知识，我们可以为每个特征分配先验权重。然而，在实际应用中，获取这些先验知识非常困难。为了缓解这个问题，我们提出了一个 SWCAN 模型。SWCAN 模型可以学习不同特征的权重，学习 SG，并同时执行聚类。为了自适应地学习权重，我们将约束  $\theta \geq 0$ ,  $\sum_i \theta_i = 1$  嵌入到 CAN 模型中。SWCAN 的目标可以写成以下问题：

$$\begin{aligned} \min_A \quad & \sum_{i=1}^n \sum_{j=1}^n (\|\Theta x_i - \Theta x_j\|_2^2 a_{ij} + \gamma a_{ij}^2) \\ \text{s.t.} \quad & \forall i, \quad a_i^T = 1, \quad 0 \leq a_{ij} \leq 1, \quad \theta^T 1 = 1 \\ & 0 \leq \theta_k \leq 1, \quad k = 1, 2, \dots, d \\ & \Theta = \text{diag}(\theta), \quad \text{rank}(L) = n - c. \end{aligned}$$

其中  $\theta \in R^d$  是自加权向量， $\Theta \in R^{d \times d}$  是一个对角矩阵，其主要对角向量是向量  $\theta$ 。其他参数的定义与 CAN 相同。从<sup>[3]</sup>中，我们知道拉普拉斯矩阵  $L$  中特征值 0 的数量等于 SG  $A$  中连通分量的数量，因此， $A$  必须具有  $c$  连通分量和块对角结构。除了处理噪声特征外，SWCAN 的自加权方法不会改变数据的结构信息。具体来说，对特征权重进行一定程度的微调有时会使模型性能更好。

#### 3.2 方法优点

SWCAN 在实践中的优势总结如下：

1) SWCAN 模型可以自适应地学习不同特征的重要性，同时学习 SG 和聚类。

2) 如果数据具有噪声或低辨别特征，SWCAN 可以自适应地学习噪声特征的低权重和辨别特征的高权重。

3) 如果大多数特征是有区别的，则 SWCAN 可以精细地调整权重以更好地执行聚类。

方法的核心技术如下图 2所示：

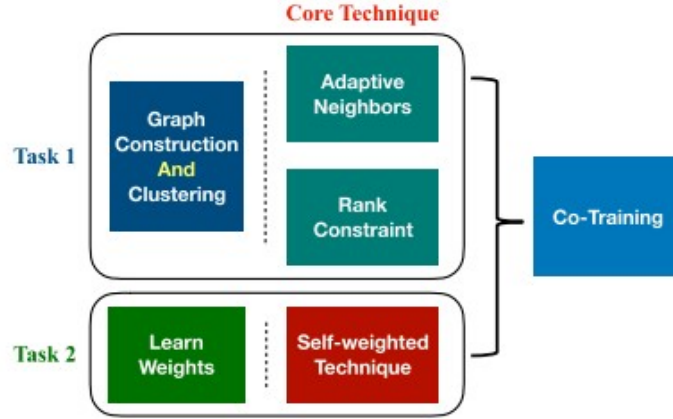


图 2: 技术核心

增加权重后方法对噪音特征处理的大致效果如下图 3所示：

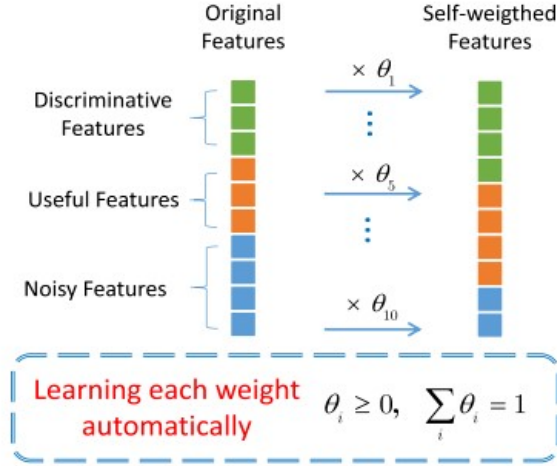


图 3: 降噪效果图

### 3.3 目标函数

对前面展示得到的 SWCAN 的目标函数，可转化为如下函数：

$$\begin{aligned} \min_{A, \Theta, F} \quad & \sum_{i=1}^n \sum_{j=1}^n (||\Theta x_i - \Theta x_j||_2^2 a_{ij} + \gamma a_{ij}^2) + 2\lambda \text{Tr}(F^T L F) \\ \text{s.t.} \quad & \forall i, \quad a_i^T = 1, \quad 0 \leq a_{ij} \leq 1, \quad \theta^T 1 = 1 \\ & 0 \leq \theta_k \leq 1, \quad k = 1, 2, \dots, d, \quad \Theta = \text{diag}(\theta) \\ & F \in R^{n \times c}, \quad F^T F = I_c. \end{aligned}$$

函数中的 F 由对应于 c 个最小特征值的 L 的 c 个特征向量形成。

## 4 复现细节

### 4.1 与已有开源代码对比

我所复现论文没有任何相关源代码，代码均为自己所写。在本论文的复现工作中，我完成了论文所提出的聚类模型代码的编写，所用编程语言为 python，算法的伪代码如下：

---

**Procedure 1** Algorithm for Solving the Problem of Goal function.

---

**Input:** Data  $X=[x_1, x_2, \dots, x_n]^T \in R^{n \times d}$ , Cluster number  $c$ , The tradeoff parameters  $\lambda$  and  $\gamma$

**Ensure :** The similarity graph  $A \in R^{n \times n}$ , Weighted vector  $\theta$ .

Initialize a similarity graph  $A \in R^{n \times n}$ .

**Output:** The similarity graph  $A$

**repeat :**

    Update F,      Update each element of  $\theta$ ,      Update A

**until** Convergence

---

### 4.2 算法迭代过程

在算法的伪代码中，针对更新 F、更新  $\theta$  的每个元素以及更新 A，具体的更新方法如下：

(1) 更新 F:

$$\min_{\mathbf{F} \in \mathbb{R}^{n \times c}, \mathbf{F}^T \mathbf{F} = \mathbf{I}} \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F}).$$

函数中的 F 由对应于 c 个最小特征值的 L 的 c 个特征向量形成。

(2) 更新  $\theta$  的每个元素:

$$\theta_i = \frac{1}{m_i^* \sum_{i=1}^d \frac{1}{m_i^*}}$$

$m_i^*$  是 M 矩阵的第 i 个对角元素，M 矩阵为:

$$\mathbf{M} \in \mathbb{R}^{d \times d} = \mathbf{X}^T \mathbf{L} \mathbf{X}$$

(3) 更新 A:

$$\begin{aligned} & \min_{\mathbf{a}_i} \sum_{j=1}^n (\gamma a_{ij}^2 + r_{ij} a_{ij}) \\ \Rightarrow & \min_{\mathbf{a}_i} \sum_{j=1}^n \left( a_{ij}^2 + \frac{1}{\gamma} r_{ij} a_{ij} + \frac{1}{4\gamma^2} r_{ij}^2 \right) \\ \Rightarrow & \min_{\mathbf{a}_i} \left\| \mathbf{a}_i + \frac{1}{2\gamma} \mathbf{r}_i \right\|_2^2 \\ \text{s.t. } & \mathbf{a}_i^T \mathbf{1} = 1, 0 \leq a_{ij} \leq 1. \end{aligned}$$

其中， $\mathbf{r}_i$  为一个向量，且其第 j 个元素为:

$$r_{ij} = \left\| \Theta^T \mathbf{x}_i - \Theta^T \mathbf{x}_j \right\|_2^2 + \lambda \left\| \mathbf{f}_i - \mathbf{f}_j \right\|_2^2$$

## 5 实验结果分析

该论文用了多种算法针对多个数据进行了实验，充分验证了其算法的有效性，样例如下图 4 所示:

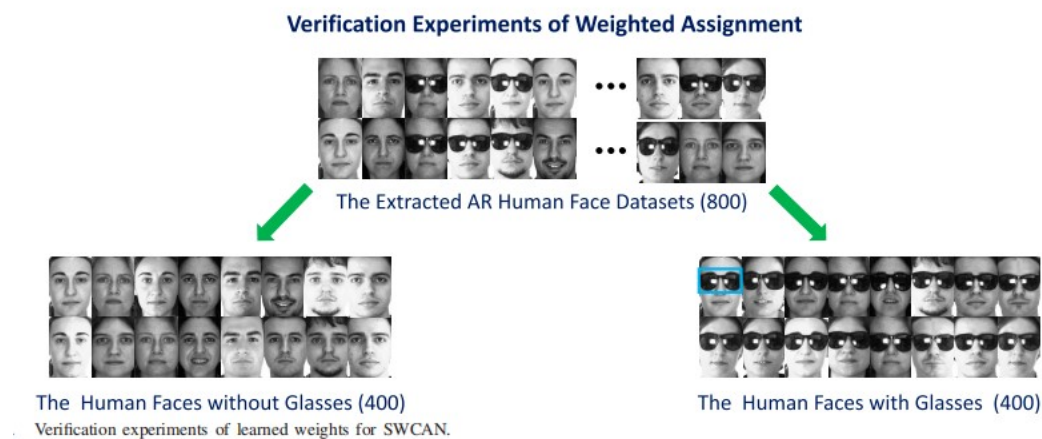


图 4: SWCAN 学习权重的验证实验

对此，为验证我所编写的代码的有效性，我也进行了实验，但得到的实验结果并不良好，代码整体没有问题，预计是局部偏差导致效果不佳，如在我的实验中得到的目标函数的变化曲线如下图 5 所示：

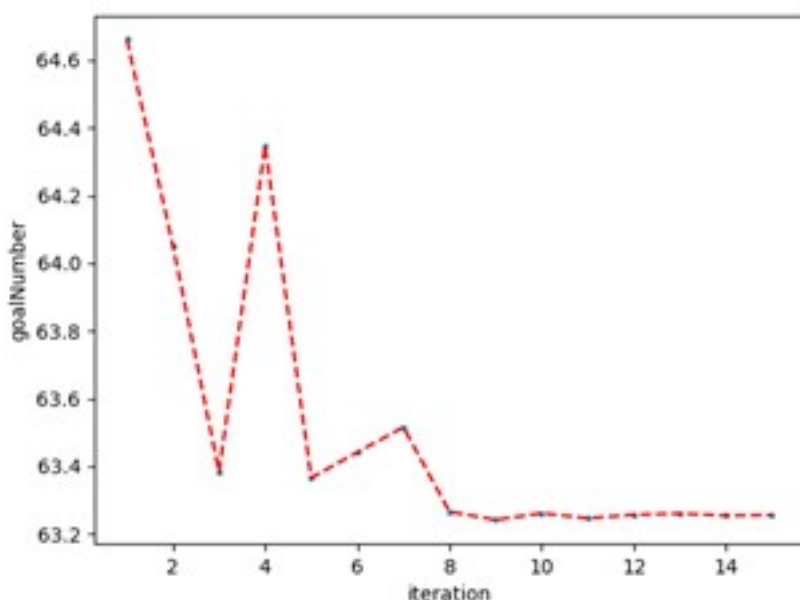


图 5: 目标函数变化曲线

## 6 总结与展望

许多先前的聚类工作遵循一个理想的假设，即不同特征的权重应该相等。由于真实世界的数据集总是具有一些噪声特征，因此该假设不适用于实际应用。论文作者提出的新的聚类模型 SWCAN，它可以学习不同特征的权重，同时学习 SG 和聚类。SWCAN 中的自加权方法可以有效地抑制噪声特征，从而获得更好的聚类性能。

在本次复现过程中，我理解与分析了算法原理以及公式推导的含义，此论文并没有公开代码，故用 python 自行复现了代码，复现代码的过程也让我得到了很大的锻炼，代码整体上复现完整良好，但在代码复现的细节之处并没有处理妥当，以致于实验效果并不理想，这是我此次复现过程中的不足之处。

## 参考文献

- [1] HAGEN L, KAHNG A B. New spectral methods for ratio cut partitioning and clustering[J]. Comput.-Aided Des. Integr. Circuits Sys, 1992, 11(09): 1074-1085.
- [2] SHI J, MALIK J. Normalized cuts and image segmentation[J]. Pattern Anal. Mach. Intell., 2000, 22(08): 888-905.
- [3] F. NIE X W, HUANG H. Clustering and projected clustering with adaptive neighbors[J]. Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2014, pp: 977-986.
- [4] F. NIE X W, HUANG H. Self-Weighted Clustering With Adaptive Neighbors[J]. IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, 2020, 31(09): 3428-3440.