

国防科技大学研究生 800 级讲座 从统计理论角度看多视图聚类算法

报告人：计算机学院博士后 梁伟轩

2025 年 5 月 10 日



国防科技大学
National University of Defense Technology



Pattern Recognition & Machine Intelligence Lab, NUDT

目录

1 研究背景及研究现状

2 聚类算法统计理论：泛化性和一致性

3 多核聚类算法的大规模扩展

4 问答

目录

1 研究背景及研究现状

- 研究背景
- 研究现状

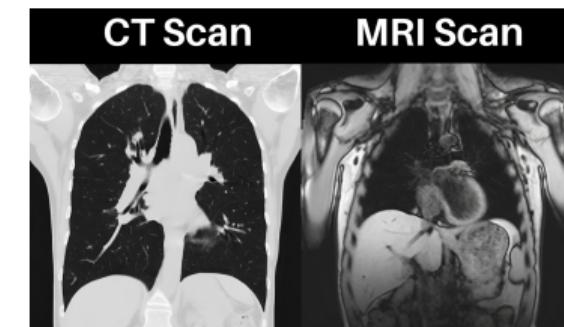
2 聚类算法统计理论：泛化性和一致性

3 多核聚类算法的大规模扩展

4 问答

研究背景

随着传感技术和信息收集技术的进步，对于同一个物体的描述，其数据类型也在不断丰富。



图：多视图数据示例

从同一个物体的**不同角度或不同模态**获取的数据集，一般被称为**多视图或者多模态**数据集。

研究背景

具备处理多视图数据能力的算法，一般被称为**多视图学习算法**，这类算法已经在现实的问题中有广泛的应用。



图：多视图学习算法用于图像分割任务示例^[1]

^[1] Abdelaziz Djelouah et al. "Multi-View Object Segmentation in Space and Time". In: ICCV. 2013, pp. 2640–2647.

研究背景

具备处理多视图数据能力的算法，一般被称为**多视图学习算法**，这类算法已经在现实的问题中有广泛的应用。

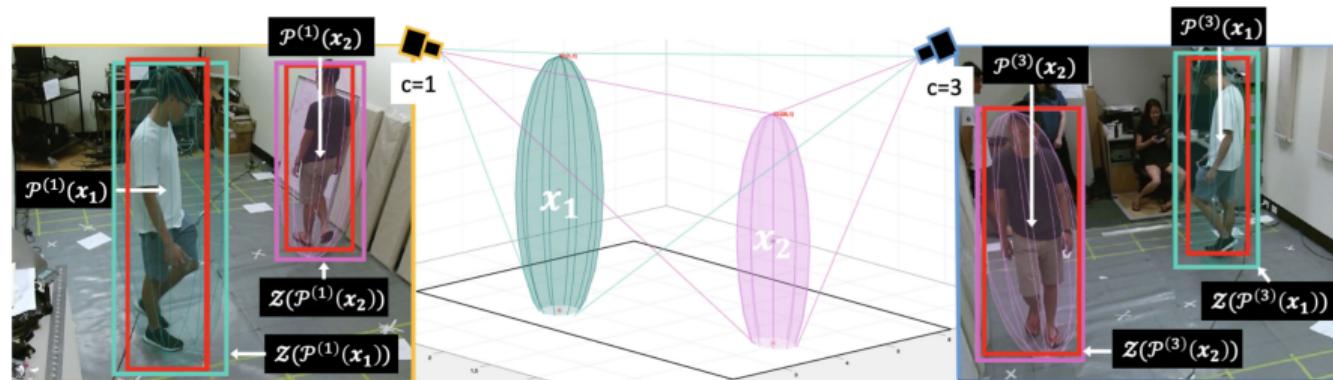


图: 多视图学习算法用于目标追踪任务示例 [1]

¹ Jonah Ong et al. "A Bayesian Filter for Multi-View 3D Multi-Object Tracking with Occlusion Handling". In: TPAMI (2020).

研究背景

具备处理多视图数据能力的算法，一般被称为**多视图学习算法**，这类算法已经在现实的问题中有广泛的应用。

图：多视图学习算法用于自动驾驶任务示例^[1]

^[1]Yuqi Wang et al. "Driving into the Future: Multiview Visual Forecasting and Planning with World Model for Autonomous Driving". In: CVPR (2024).

研究背景

与此同时，现实中绝大部分数据集是**无标签的**。处理无标签数据集的关键手段之一是**聚类**。



“人类无法完全依赖有监督学习的方法完成所有神经元训练，而需要更多来自于**无监督学习**的帮助。”



“无监督学习是人工智能的下一站。”

图灵奖得主Geoffrey E. Hinton

图灵奖得主Yann LeCun

聚类算法的目的是将数据划分为若干个子集，使得同一子集中的数据与其他子集中的数据相比，具有更高的相似性。对多视图数据进行聚类的算法，被称为**多视图聚类算法**，这是此次报告的核心内容。

研究背景

现有的绝大部分多视图聚类算法存在**计算复杂度过高**的问题，很多方法的复杂度高达 $\mathcal{O}(n^3)$ 。

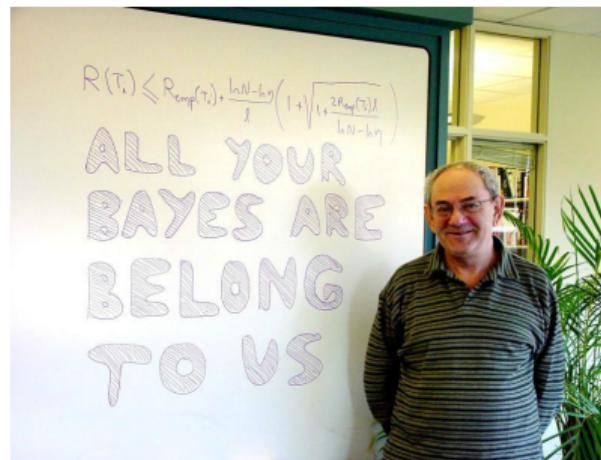


图：多视图数据集 Covertype，是基于海拔、坡向、坡度、遮阳、土壤类型等属性进行划分的森林覆盖种类的数据集，包含 581012 个样本

在 Covertype 上运行复杂度 $\mathcal{O}(n^3)$ 的算法，每次迭代需要**19.61 亿亿次的运算**，而“天河二号”超级计算机的持续计算速度为**每秒 3.39 亿亿次**。

研究背景

针对复杂度高的问题，不少学者尝试利用**锚点法**对现有的多视图聚类算法进行了加速。然而，这些方法缺少基于**统计学理论的可解释性研究**。



*Nothing is more practical
than a good theory.*

— Vladimir N. Vapnik

对于逐渐兴起的大规模多视图聚类算法，利用统计理论**对学习算法的统计性质进行研究**，也是一项至关重要的研究工作。

目录

1 研究背景及研究现状

- 研究背景
- 研究现状

2 聚类算法统计理论：泛化性和一致性

3 多核聚类算法的大规模扩展

4 问答

研究现状

现有的多视图聚类算法基于**不同的学习机制**，可以大致分为：

a 非深度多视图聚类算法

- ① 基于协同训练的多视图聚类算法
- ② 多核聚类算法
- ③ 多视图子空间聚类算法
- ④ 基于图的多视图聚类算法
- ⑤ 基于矩阵分解的多视图聚类算法

b 深度多视图聚类算法

研究现状

现有的多视图聚类算法基于**不同的学习机制**，可以大致分为：

a 非深度多视图聚类算法

- ① 基于协同训练的多视图聚类算法
- ② **多核聚类算法**
- ③ 多视图子空间聚类算法
- ④ 基于图的多视图聚类算法
- ⑤ 基于矩阵分解的多视图聚类算法

b 深度多视图聚类算法

研究现状——多核聚类算法

k 均值聚类算法

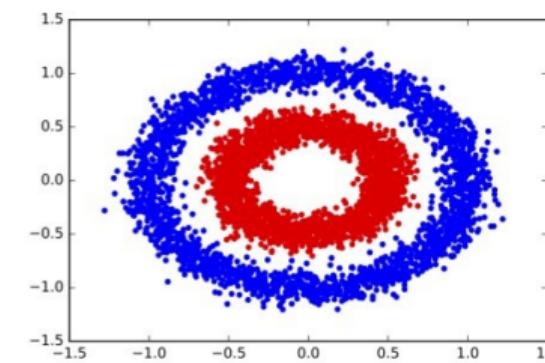
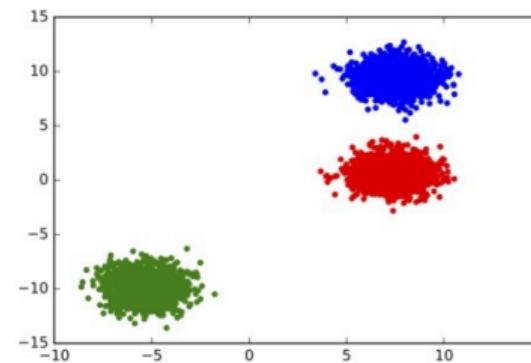
$$\mathcal{W}_n(\mathbf{C}, \rho_n) = \int_{\mathcal{X}} \min_{j \in [k]} \|x - \mathbf{c}_j\|^2 d\rho_n(x) = \frac{1}{n} \sum_{i=1}^n \min_{j \in [k]} \|x_i - \mathbf{c}_j\|^2. \quad (1)$$

研究现状——多核聚类算法

k 均值聚类算法

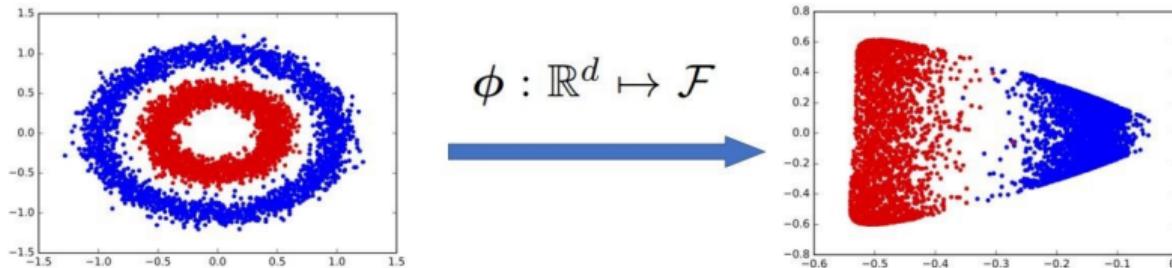
$$\mathcal{W}_n(\mathbf{C}, \rho_n) = \int_{\mathcal{X}} \min_{j \in [k]} \|x - \mathbf{c}_j\|^2 d\rho_n(x) = \frac{1}{n} \sum_{i=1}^n \min_{j \in [k]} \|x_i - \mathbf{c}_j\|^2. \quad (1)$$

标准的 k 均值聚类算法只能处理线性可分的数据。



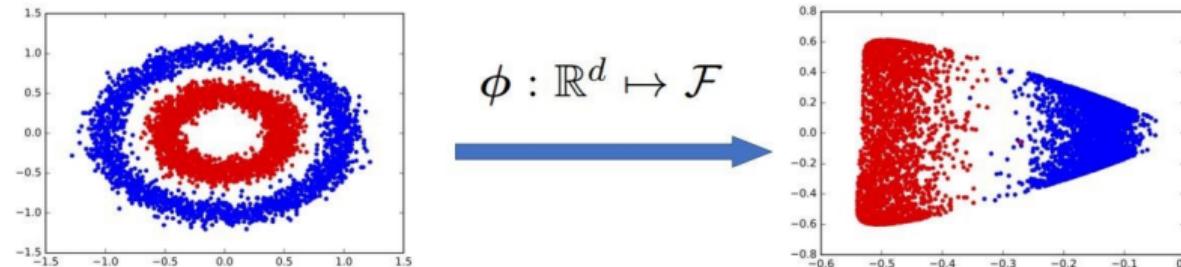
研究现状——多核聚类算法

为了使 k 均值聚类算法能够处理非线性数据，利用核函数将数据投射到特征空间。



研究现状——多核聚类算法

为了使 k 均值聚类算法能够处理非线性数据，利用核函数将数据投射到特征空间。



根据 Mercer 定理，可知对于任意核函数 $K(\cdot, \cdot)$ ，都存在特征映射 $\phi(\cdot)$ 及某个函数空间 \mathcal{F} ，满足 $K(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{F}}$ ，其中 x, y 为样本空间中任意两个样本。

核 k 均值聚类

$$\mathbf{x}_i \rightarrow \phi(\mathbf{x}_i) \implies \mathcal{W}_n(\mathbf{C}, \rho_n) = \frac{1}{n} \sum_{i=1}^n \min_{j \in [k]} \|\phi(\mathbf{x}_i) - \mathbf{c}_j\|^2. \quad (2)$$

研究现状——多核聚类算法

根据 Mercer 定理，式(2)可以写成

$$\min_{\mathbf{Z}} \frac{1}{n} \text{tr}(\mathbf{K}_n) - \frac{1}{n} \text{tr}(\mathbf{K}_n \mathbf{Z} \mathbf{L}^{-1} \mathbf{Z}^\top), \quad (3)$$

其中 $\mathbf{Z} \in \{0, 1\}^{n \times k}$ 是一个指示矩阵，且 $\mathbf{L} = \text{diag}(\mathbf{Z}^\top \mathbf{1}_n)$ 。

研究现状——多核聚类算法

根据 Mercer 定理，式(2)可以写成

$$\min_{\mathbf{Z}} \frac{1}{n} \text{tr}(\mathbf{K}_n) - \frac{1}{n} \text{tr}(\mathbf{K}_n \mathbf{Z} \mathbf{L}^{-1} \mathbf{Z}^\top), \quad (3)$$

其中 $\mathbf{Z} \in \{0, 1\}^{n \times k}$ 是一个指示矩阵，且 $\mathbf{L} = \text{diag}(\mathbf{Z}^\top \mathbf{1}_n)$ 。

松弛的核 k 均值聚类算法

令 $\mathbf{H} = \mathbf{Z} \mathbf{L}^{-1/2}$ ，并松弛其离散约束，可得

$$\min_{\mathbf{H}} \frac{1}{n} \text{tr}(\mathbf{K}_n) - \frac{1}{n} \text{tr}(\mathbf{K}_n \mathbf{H} \mathbf{H}^\top), \text{ s.t. } \mathbf{H}^\top \mathbf{H} = \mathbf{I}_k. \quad (4)$$

研究现状——多核聚类算法

在实际中，难以选择合适的核函数，这样就需要多核聚类算法。假设有 m 个基核函数 $\{K_p(\cdot, \cdot)\}_{p=1}^m$ ，其中核函数 $K_p(\cdot, \cdot)$ 对应的特征映射为 $\phi_p(\cdot)$ 。在多核条件下，样本 x 的特征映射为

$$[\phi_1^\top(x), \dots, \phi_m^\top(x)]^\top.$$

为了突出每个视图不同的重要性，利用系数对其进行线性组合，具体如下

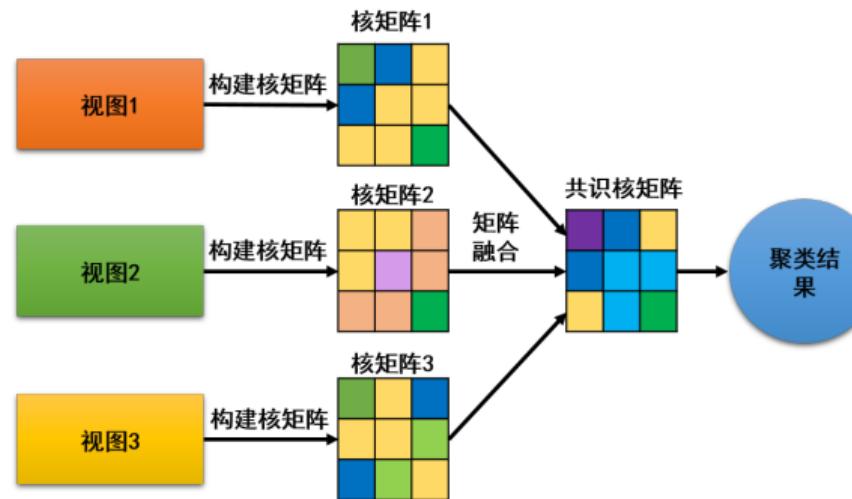
$$[\gamma_1\phi_1^\top(x), \dots, \gamma_m\phi_m^\top(x)]^\top.$$

假设由样本集计算的第 p 个基核矩阵为 $\frac{1}{n}\mathbf{K}_p$ ，整个核矩阵可表示为

$$\frac{1}{n}\mathbf{K}_\gamma = \frac{1}{n} \sum_{p=1}^m \gamma_p^2 \mathbf{K}_p.$$

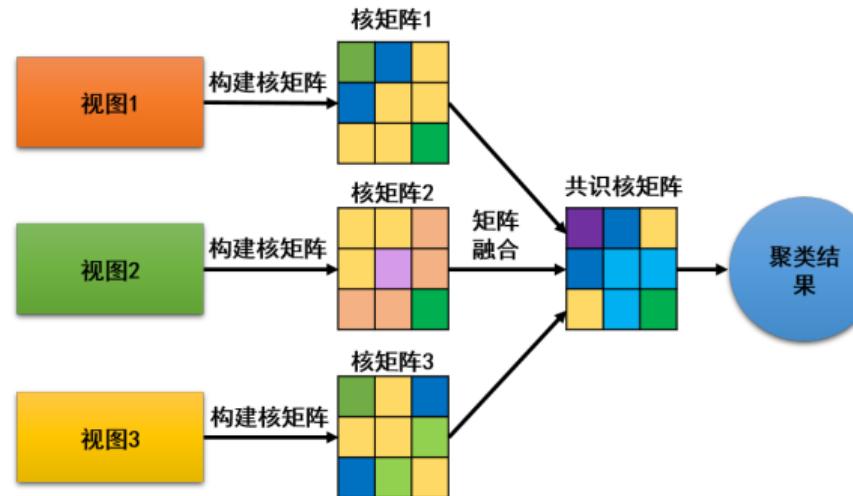
研究现状——多核聚类算法

在多视图聚类领域的多核聚类则是针对不同视图，利用相同或者不同的核函数构造相应的基核矩阵，再利用核系数对基核矩阵进行组合，其示意图如下所示。



研究现状——多核聚类算法

在多视图聚类领域的多核聚类则是针对不同视图，利用相同或者不同的核函数构造相应的基核矩阵，再利用核系数对基核矩阵进行组合，其示意图如下所示。



在目前的多核聚类算法中，基核矩阵构造的**空间复杂度为 $\mathcal{O}(mn^2)$** ，而后的特征分解导致其**时间复杂度为 $\mathcal{O}(n^3)$** 。同时，对于多核聚类算法的统计性质的**理论研究比较匮乏**。

目录

1 研究背景及研究现状

2 聚类算法统计理论：泛化性和一致性

3 多核聚类算法的大规模扩展

4 问答

聚类算法的泛化性分析

为什么要进行泛化性分析？

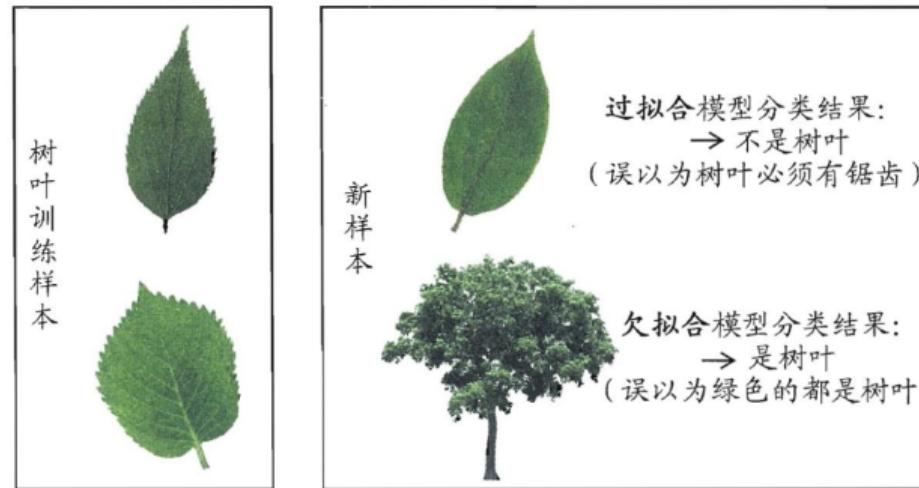


图: 泛化性能研究

聚类算法的泛化性分析

在统计学习理论中，**算法的泛化性**是核心问题之一。首先引入如下两个概念。

聚类算法的经验风险

$$\mathcal{W}_n(\mathbf{C}, \rho_n) = \frac{1}{n} \sum_{i=1}^n \min_{j \in [k]} \|\phi(x_i) - \mathbf{c}_j\|^2, \quad (5)$$

其中 $\mathbf{C} = \{\mathbf{c}_j\}_{j=1}^k$ 为 k 个聚类中心， $\phi(\cdot)$ 为某个核函数 $K(\cdot, \cdot)$ 对应的特征映射。

给定一组聚类中心，通常用期望风险衡量其在全空间 \mathcal{X} 上的表现。

聚类算法的期望风险

$$\mathcal{W}(\mathbf{C}, \rho) = \int_{\mathcal{X}} \min_{j \in [k]} \|\phi(x) - \mathbf{c}_j\|^2 d\rho(x). \quad (6)$$

聚类算法的泛化性分析

经验风险与期望风险最小项

设使经验风险与期望风险最小的聚类中心分别为：

$$\mathbf{C}_n = \arg \min_{\mathbf{C} \in \mathcal{H}^k} \mathcal{W}_n(\mathbf{C}, \rho_n), \quad \mathbf{C}^* = \arg \min_{\mathbf{C} \in \mathcal{H}^k} \mathcal{W}(\mathbf{C}, \rho). \quad (7)$$

通常用下式衡量聚类算法的泛化性。

聚类算法的额外风险

$$\mathcal{E} := \mathbb{E}_{S_n}[\mathcal{W}(\mathbf{C}_n, \rho)] - \mathcal{W}^*(\rho), \quad (8)$$

其中 $\mathcal{W}^*(\rho) = \mathcal{W}(\mathbf{C}^*, \rho)$ 。

分析多核聚类算法的泛化性，将面临在样本集和样本空间上运行多核算法得到的系数不一致的困难。因此，**要对多核聚类算法进行泛化性分析，就要先对其核系数的一致性进行分析。**

聚类算法的一致性分析

一致性是学习算法的重要统计性质，一致性分析旨在研究**经验参数是否能收敛于真实参数。**

聚类算法的一致性分析

一致性是学习算法的重要统计性质，一致性分析旨在研究**经验参数是否能收敛于真实参数**。



图：“吾生也有涯，而知也无涯”



图：机器学习的过程

简单多核 k 均值聚类算法

多核 k 均值聚类算法 (Multiple Kernel k -means, MKKM)^[2] 是较为流行的多核聚类算法。

MKKM

$$\min_{\gamma, \mathbf{H}} \frac{1}{n} \text{tr}(\mathbf{K}_\gamma (\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)), \text{ s.t. } \gamma \in \Delta, \mathbf{H}^\top \mathbf{H} = \mathbf{I}_k. \quad (9)$$

简单多核 k 均值聚类算法 (Simple Multiple Kernel k -means, SimpleMKKM)^[3] 是对 MKKM 的一种改进，其聚类效果好，并且不含超参数，故将其作为主要分析对象。

SimpleMKKM

$$\min_{\gamma} f_n(\gamma), \text{ s.t. } \gamma \in \Delta, \quad (10)$$

其中 $f_n(\gamma) = \max_{\mathbf{H}} \frac{1}{n} \text{tr}(\mathbf{K}_\gamma \mathbf{H}\mathbf{H}^\top), \text{ s.t. } \mathbf{H}^\top \mathbf{H} = \mathbf{I}_k.$

²Hsin-Chien Huang, Yung-Yu Chuang, and Chu-Song Chen. “Multiple Kernel Fuzzy Clustering”. In: *TFS* (2012).

³Xinwang Liu. “SimpleMKKM: Simple Multiple Kernel K-Means”. In: *TPAMI* (2022).

简单多核 k 均值聚类算法

在原文献中，式(10)中的 $f_n(\gamma)$ ，被证明了是可微的，并可利用归约梯度法进行优化。固定某个分量 u 使 $\gamma_u > 0$ ， $f_n(\gamma)$ 的归约梯度有如下定义：

$f_n(\gamma)$ 的归约梯度 $\bigtriangledown f_n(\gamma)$

$$[\bigtriangledown f_n(\gamma)]_p = \frac{\partial f_n(\gamma)}{\partial \gamma_p} - \frac{\partial f_n(\gamma)}{\partial \gamma_u}, \text{ 若 } p \neq u; \quad [\bigtriangledown f_n(\gamma)]_u = \sum_{t \neq u} \left(\frac{\partial f_n(\gamma)}{\partial \gamma_u} - \frac{\partial f_n(\gamma)}{\partial \gamma_t} \right), \quad (11)$$

其中 $\frac{\partial f_n(\gamma)}{\partial \gamma_p} = \frac{2\gamma_p}{n} \text{tr} (\mathbf{K}_p \mathbf{H}_\gamma \mathbf{H}_\gamma^\top)$ ， $\mathbf{H}_\gamma = \arg \max_{\mathbf{H}^\top \mathbf{H} = \mathbf{I}_k} \text{tr} (\mathbf{K}_\gamma \mathbf{H} \mathbf{H}^\top)$ 。

简单多核 k 均值聚类算法

在原文献中，式(10)中的 $f_n(\gamma)$ ，被证明了是可微的，并可利用归约梯度法进行优化。固定某个分量 u 使 $\gamma_u > 0$ ， $f_n(\gamma)$ 的归约梯度有如下定义：

$f_n(\gamma)$ 的归约梯度 $\bigtriangledown f_n(\gamma)$

$$[\bigtriangledown f_n(\gamma)]_p = \frac{\partial f_n(\gamma)}{\partial \gamma_p} - \frac{\partial f_n(\gamma)}{\partial \gamma_u}, \text{ 若 } p \neq u; \quad [\bigtriangledown f_n(\gamma)]_u = \sum_{t \neq u} \left(\frac{\partial f_n(\gamma)}{\partial \gamma_u} - \frac{\partial f_n(\gamma)}{\partial \gamma_t} \right), \quad (11)$$

其中 $\frac{\partial f_n(\gamma)}{\partial \gamma_p} = \frac{2\gamma_p}{n} \text{tr} (\mathbf{K}_p \mathbf{H}_\gamma \mathbf{H}_\gamma^\top)$ ， $\mathbf{H}_\gamma = \arg \max_{\mathbf{H}^\top \mathbf{H} = \mathbf{I}_k} \text{tr} (\mathbf{K}_\gamma \mathbf{H} \mathbf{H}^\top)$ 。

为了保持单纯形约束，目标函数的下降方向 $\hat{\mathbf{d}}$ 的第 p 个分量应为

$$\hat{d}_p = \begin{cases} 0, & \text{若 } \gamma_p = 0, [\bigtriangledown f_n(\gamma)]_p \geq 0 \text{ 且 } p \neq u, \\ -\frac{1}{m-1} \sum_{t \neq u, \gamma_t > 0} \left(\frac{\partial f_n(\gamma)}{\partial \gamma_u} - \frac{\partial f_n(\gamma)}{\partial \gamma_t} \right), & \text{若 } p = u, \\ -\frac{1}{m-1} \left(\frac{\partial f_n(\gamma)}{\partial \gamma_p} - \frac{\partial f_n(\gamma)}{\partial \gamma_u} \right), & \text{其他情况.} \end{cases} \quad (12)$$

SimpleMKKM 的期望版本

对于任意核函数 $K(\cdot, \cdot)$ 及样本空间 \mathcal{X} , 定义其相应的积分算子为

$$L_K : C(\mathcal{X}) \rightarrow C(\mathcal{X}), \quad L_K f(x) = \int_{\mathcal{X}} K(x, y) f(y) d\rho(y), \quad (13)$$

其中 $C(\mathcal{X})$ 代表连续函数空间。上述积分算子是核矩阵 $\frac{1}{n} \mathbf{K} \in \mathbb{R}^{n \times n}$ 在样本空间上的推广。

SimpleMKKM 的期望版本

$$\min_{\gamma} f(\gamma), \text{ s.t. } \gamma \in \Delta, \quad (14)$$

其中 $f(\gamma) = \max_{\{h_j\}_{j=1}^k \in \Gamma} \sum_{j=1}^k \iint_{\mathcal{X}} K_{\gamma}(x, y) h_j(x) h_j(y) d\rho(x) d\rho(y)$, Γ 是 $C(\mathcal{X})$ 上的单位正交约束^a。

^a若函数组 $\{h_j\}_{j=1}^k \in \Gamma$, 则对于 $i \neq j (i, j \in [k])$, $\int_{\mathcal{X}} h_i(x) h_j(x) d\rho(x) = 0$, 且有 $\int_{\mathcal{X}} h_i^2(x) d\rho(x) = 1$ 。

SimpleMKKM ——一致性问题

设 $\hat{\gamma}_n = \arg \min_{\gamma} f_n(\gamma)$, $\gamma^* = \arg \min_{\gamma} f(\gamma)$, SimpleMKKM 系数一致性问题如下所示。

SimpleMKKM 系数一致性问题

能否得到 $\|\hat{\gamma}_n - \gamma^*\|_\infty$ 的非渐近上界?

SimpleMKKM ——一致性问题

设 $\hat{\gamma}_n = \arg \min_{\gamma} f_n(\gamma)$, $\gamma^* = \arg \min_{\gamma} f(\gamma)$, SimpleMKKM 系数一致性问题如下所示。

SimpleMKKM 系数一致性问题

能否得到 $\|\hat{\gamma}_n - \gamma^*\|_\infty$ 的非渐近上界?

定理 1 (SimpleMKKM 一致性结论)

若期望版本和经验版本的 SimpleMKKM 初始化相同，学习步长 η 小于某个正常数 c ，经 T 迭代收敛后，则有

$$\|\gamma^* - \hat{\gamma}\|_\infty \lesssim k \sqrt{\frac{\log(4kT/\delta)}{n}}, \quad (15)$$

以至少 $1 - \delta$ 的概率成立。

SimpleMKKM 额外风险界

对于某组核系数 $\gamma \in \mathbb{R}^m$ 和任意样本 $x \in \mathcal{X}$, 关于 x 的多核映射可以写为

$$\phi_{\gamma}(x) = [\gamma_1 \phi_1^\top(x), \dots, \gamma_m \phi_m^\top(x)]^\top. \quad (16)$$

假设其对应的再生核希尔伯特空间为 \mathcal{H}_γ 。对于某组质心 $\mathbf{C} = \{\mathbf{c}_j\}_{j=1}^k \in \mathcal{H}_\gamma^k$, SimpleMKKM 的标准聚类损失函数为

$$l(x, \mathbf{C}, \gamma) = \min_{j \in [k]} \|\phi_{\gamma}(x) - \mathbf{c}_j\|^2. \quad (17)$$

SimpleMKKM 经验风险和期望风险

SimpleMKKM 经验风险: $\mathcal{W}_n(\mathbf{C}, \gamma, \rho_n) = \frac{1}{n} \sum_{i=1}^n \min_{j \in [k]} \|\phi_{\gamma}(x_i) - \mathbf{c}_j\|^2.$

SimpleMKKM 期望风险: $\mathcal{W}(\mathbf{C}, \gamma, \rho) = \int_{\mathcal{X}} \min_{j \in [k]} \|\phi_{\gamma}(x) - \mathbf{c}_j\|^2 d\rho(x).$

SimpleMKKM 额外风险界

设算法经验版本的聚类中心为 $\hat{\mathbf{C}} = \{\hat{\mathbf{c}}_j\}_{j=1}^k \in \mathcal{H}_{\hat{\gamma}}^k$, 则 $\hat{\mathbf{C}} = \operatorname{argmin}_{\mathbf{C} \in \mathcal{H}_{\hat{\gamma}}^k} \mathcal{W}_n(\mathbf{C}, \hat{\gamma}, \rho_n)$ ⁴。同时设算法期望版本的聚类中心为 $\mathbf{C}^* = \operatorname{argmin}_{\mathbf{C} \in \mathcal{H}_{\gamma^*}^k} \mathcal{W}(\mathbf{C}, \gamma^*, \rho)$ 。

定理 2

SimpleMKKM 的额外风险的上界为

$$\mathbb{E}_{S_n}[\mathcal{W}(\hat{\mathbf{C}}, \hat{\gamma}, \rho)] - \mathcal{W}(\mathbf{C}^*, \gamma^*, \rho) \leq \tilde{\mathcal{O}}(k/\sqrt{n}). \quad (18)$$

⁴这里不讨论 k 均值聚类算法的最优解与实际得到的解之间的差异。

证明思路

进行如下分解：

$$\begin{aligned} & \mathbb{E}_{S_n}[\mathcal{W}(\hat{\mathbf{C}}, \hat{\gamma}, \rho)] - \mathcal{W}(\mathbf{C}^*, \gamma^*, \rho) \\ &= \underbrace{\mathbb{E}_{S_n}[\mathcal{W}(\hat{\mathbf{C}}, \hat{\gamma}, \rho) - \mathcal{W}_n(\hat{\mathbf{C}}, \hat{\gamma}, \rho_n)]}_{\mathcal{A}} + \underbrace{\mathbb{E}_{S_n}[\mathcal{W}_n(\hat{\mathbf{C}}, \hat{\gamma}, \rho_n) - \mathcal{W}_n(\mathbf{C}^*, \gamma^*, \rho_n)]}_{\mathcal{B}} \\ &\quad + \underbrace{\mathbb{E}_{S_n}[\mathcal{W}_n(\mathbf{C}^*, \gamma^*, \rho_n)] - \mathcal{W}(\mathbf{C}^*, \gamma^*, \rho)}_{\mathcal{C}}. \end{aligned} \tag{19}$$

证明思路

设聚类质心 $\hat{\mathbf{C}}$ 对应的聚类指示矩阵为 $\hat{\mathbf{H}}_{\hat{\gamma}} \in \mathbb{R}^{n \times k}$, 设其元素为 $[h_{ij}]_{i \times j \in [n] \times [k]}$ 。当 x_i 属于第 j 个簇时, $h_{ij} = 1/\sqrt{|\mathcal{C}_j|}$, 否则 $h_{ij} = 0$ 。类似地, 将 \mathbf{C}^* 对应的聚类指示矩阵设为 $\mathbf{H}_{\gamma^*} \in \mathbb{R}^{n \times k}$ 。因为 $\hat{\mathbf{C}} = \operatorname{argmin}_{\mathbf{C} \in \mathcal{H}_{\hat{\gamma}}^k} \mathcal{W}_n(\mathbf{C}, \hat{\gamma}, \rho_n)$, 可知

$$\begin{aligned}\mathcal{B} &= \mathbb{E}_{S_n} [\mathcal{W}_n(\hat{\mathbf{C}}, \hat{\gamma}, \rho_n) - \mathcal{W}_n(\mathbf{C}^*, \gamma^*, \rho_n)] \\ &= \mathbb{E}_{S_n} [\mathcal{W}_n(\hat{\mathbf{C}}, \hat{\gamma}, \rho_n) - \mathcal{W}_n(\mathbf{C}^*, \hat{\gamma}, \rho_n)] + \mathbb{E}_{S_n} [\mathcal{W}_n(\mathbf{C}^*, \hat{\gamma}, \rho_n) - \mathcal{W}_n(\mathbf{C}^*, \gamma^*, \rho_n)] \quad (20) \\ &\leq \mathbb{E}_{S_n} [\mathcal{W}_n(\mathbf{C}^*, \hat{\gamma}, \rho_n) - \mathcal{W}_n(\mathbf{C}^*, \gamma^*, \rho_n)]\end{aligned}$$

证明思路

根据经验聚类风险的定义，可知

$$\begin{aligned}
 \mathcal{B} &= \mathbb{E}_{S_n} \left[\frac{1}{n} \text{tr} \left((\mathbf{K}_{\hat{\gamma}} - \mathbf{K}_{\gamma^*})(\mathbf{I}_n - \mathbf{H}_{\gamma^*} \mathbf{H}_{\gamma^*}^\top) \right) \right] \\
 &= \mathbb{E}_{S_n} \left[\sum_{p=1}^m (\hat{\gamma}_p^2 - (\gamma_p^*)^2) \cdot \frac{1}{n} \text{tr} \left(\mathbf{K}_p (\mathbf{I}_n - \mathbf{H}_{\gamma^*} \mathbf{H}_{\gamma^*}^\top) \right) \right] \\
 &\leq \sum_{p=1}^m (\hat{\gamma}_p + \gamma_p^*) |\hat{\gamma}_p - \gamma_p^*| \\
 &\lesssim \|\hat{\gamma} - \gamma^*\|_\infty \\
 &\lesssim k \sqrt{\frac{\log(4kT/\delta)}{n}}.
 \end{aligned} \tag{21}$$

证明思路

\mathcal{A} 和 \mathcal{C} 可以由核 k 均值聚类算法的泛化风险界^[5] 得到，其上界均为 $\mathcal{O}(k/\sqrt{n})$ 。

定理 3 (核 k 均值聚类的泛化风险界)

对于任意的系数 γ 和聚类中心 \mathbf{C} ，存在常数 $c > 0$ ，使得

$$|\mathbb{E}_{S_n}[\mathcal{W}(\mathbf{C}, \gamma, \rho) - \mathcal{W}_n(\mathbf{C}, \gamma, \rho_n)]| \leq c \cdot \frac{k}{\sqrt{n}}. \quad (22)$$

⁵Gérard Biau, Luc Devroye, and Gábor Lugosi. “On the Performance of Clustering in Hilbert Spaces”. In: *IEEE Transactions on Information Theory (TIT)* (2008), pp. 781–790.

目录

1 研究背景及研究现状

2 聚类算法统计理论：泛化性和一致性

3 多核聚类算法的大规模扩展

4 问答

基于 Nyström 的核 k 均值聚类算法

为了加速核 k 均值聚类算法，如下方法被提出^[6]。

Algorithm 1 基于 Nyström 近似的核 k 均值聚类算法

- 1: **输入：**样本集 $S_n = \{x_i\}_{i=1}^n$; 锚点集 $\{a_t\}_{t=1}^s$ (从 S_n 中不带放回地均匀随机采样而得); 核函数 $K(\cdot, \cdot)$; 聚类簇数 k 。
 - 2: **输出：**聚类结果。
 - 3: 计算核相似度矩阵 \mathbf{G} , 其元素为 $\mathbf{G}(i, t) = K(x_i, a_t)(i \in [n], t \in [s])$ 。相应地, 计算锚点集的核矩阵 \mathbf{W} , 其元素为 $\mathbf{W}(i, t) = K(a_i, a_t)(i \in [s], t \in [s])$ 。
 - 4: 对 \mathbf{W} 进行特征分解, 记特征向量为 $\mathbf{U} \in \mathbb{R}^{s \times s}$, 特征值组成的对角矩阵为 $\Lambda \in \mathbb{R}^{s \times s}$ 。
 - 5: 对样本集所有样本, 计算其近似的谱嵌入 $\tilde{\phi}_i = \mathbf{G}(i, :) \mathbf{U} \Lambda^{-1/2} \in \mathbb{R}^s$, 其中 $\mathbf{G}(i, :)$ 为 \mathbf{G} 的第 i 行。
 - 6: 对所有的 $\tilde{\phi}_i$ 进行 k 均值聚类, 获取最终的聚类结果。
-

⁶Daniele Calandriello and Lorenzo Rosasco. “Statistical and Computational Trade-Offs in Kernel K-Means”. In: NeurIPS. 2018.

基于 Nyström 的核 k 均值聚类算法

上述方法可有效降低将核 k 均值聚类算法的复杂度。

- ① 空间复杂度: $\mathcal{O}(n^2) \rightarrow \mathcal{O}(ns)$ 。
- ② 矩阵分解时间复杂度: $\mathcal{O}(n^3) \rightarrow \mathcal{O}(s^3 + ns + ns^2)$ 。
- ③ 获取聚类结果时间复杂度: $\mathcal{O}(n^2k) \rightarrow \mathcal{O}(nsk)$ 。

定理 1 (Nyström 法的理论保证)

给定一个正数 μ , 若 $s = \Omega(n/\mu \log(n/\delta))$, 则下式以 $1 - \delta$ 的概率成立

$$\mathcal{W}(\mathbf{C}_{n,s}, \rho) - \mathcal{W}(\mathbf{C}_n, \rho) \leq \mathcal{O}\left(k\left(\frac{1}{\sqrt{n}} + \frac{\mu}{n}\right)\right). \quad (23)$$

问题: 当上述方法应用于**多核聚类算法**时, 存在**近似前后系数不一致**的问题, 理论分析将难以进行。

基于均匀采样的大规模扩展法

接下来还是以 SimpleMKKM 为主要研究对象，对其进行大规模扩展。首先介绍基于均匀采样的大规模扩展法。

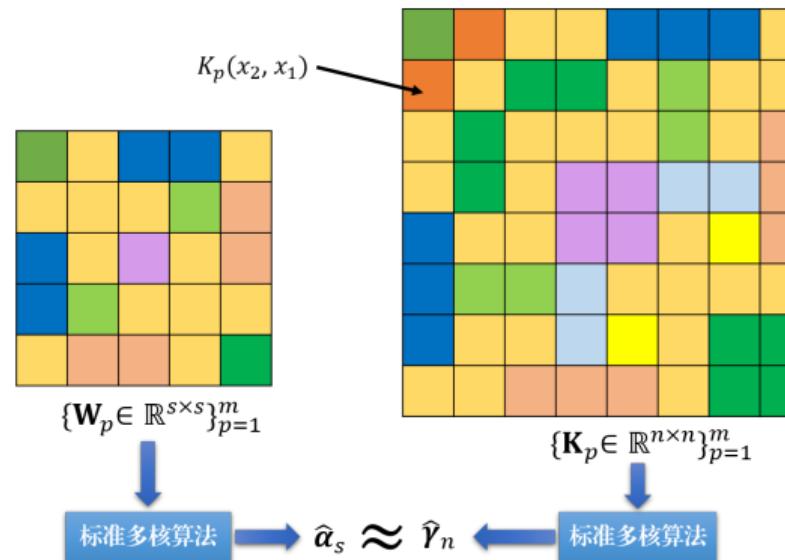


图: 基于均匀采样的大规模扩展方法中的核系数近似

基于均匀采样的大规模扩展法

基于均匀采样的大规模 SimpleMKKM 算法流程如下所示：

Algorithm 2 基于均匀采样的大规模 SimpleMKKM

- 1: **输入:** 样本集 $S_n = \{x_i\}_{i=1}^n$; 锚点集 $\{a_t\}_{t=1}^s$ (从 S_n 中不带放回地均匀随机采样而得); m 个基核函数 $\{K_p(\cdot, \cdot)\}_{p=1}^m$; 聚类簇数 k 。
 - 2: **输出:** 聚类结果。
 - 3: 计算 m 个锚点核矩阵 $\{\mathbf{W}_p\}_{p=1}^m$, 其元素为 $\mathbf{W}_p(i, t) = K_p(a_i, a_t)(i \in [s], t \in [s])$ 。在锚点核矩阵上运行 SimpleMKKM, 设输出的核系数为 $\hat{\alpha}_s$ 。
 - 4: 在第 p 个核中, 记由 n 样本和 s 个锚点构造的核相似度矩阵为 $\mathbf{G}_p \in \mathbb{R}^{n \times s}$ 。然后, 利用核系数 $\hat{\alpha}_s$ 按照原始 SimpleMKKM 算法的方式对上述矩阵进行加权组合, 分别表示为 $\mathbf{G}_{\hat{\alpha}_s}$ 和 $\mathbf{W}_{\hat{\alpha}_s}$ 。
 - 5: 对 $\mathbf{W}_{\hat{\alpha}_s}$ 进行特征分解, 得到 $\mathbf{W}_{\hat{\alpha}_s} = \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^\top$ 。
 - 6: 对 $\mathbf{G}_{\hat{\alpha}_s} \mathbf{U}_s \mathbf{\Lambda}_s^{-1/2}$ 进行标准的 k 均值聚类算法以获得最终的聚类结果。
-

基于均匀采样的大规模扩展法——核系数近似效果

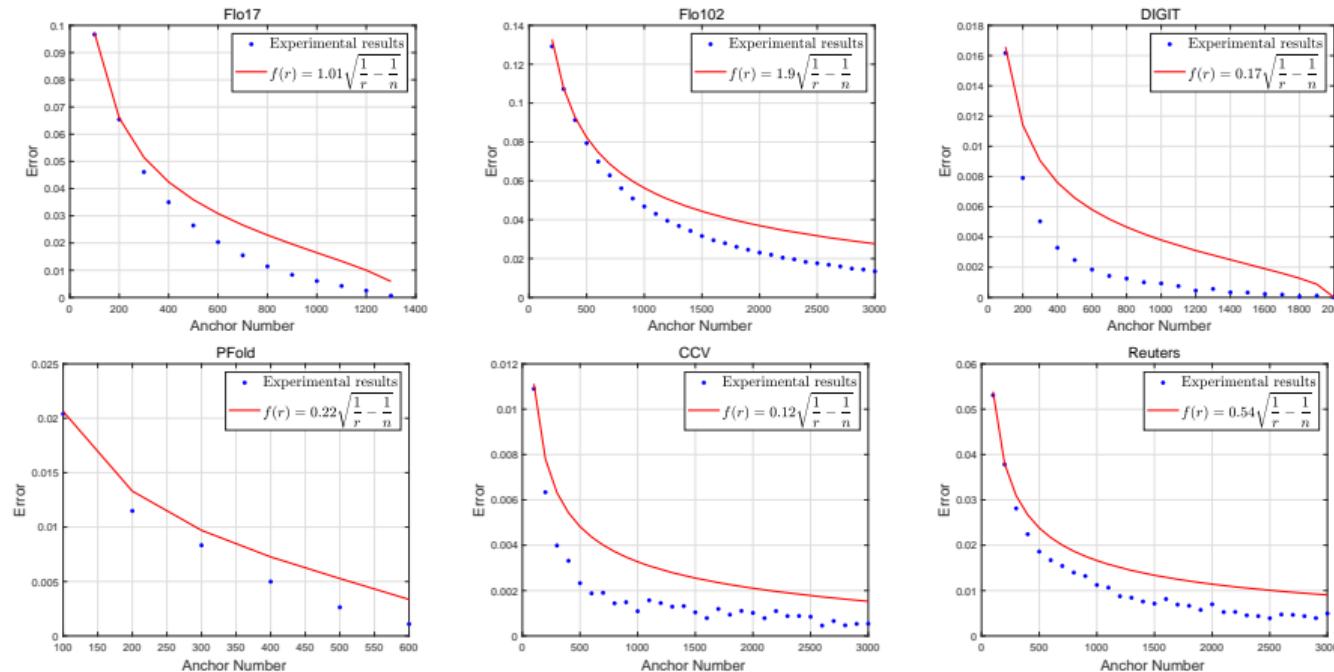


图: 基于均匀采样的近似算法与原始算法学习到的核系数的差异。蓝点记录了数值实验的真实值，而红色曲线是 $f(s) = a\sqrt{1/s - 1/n}$ 的图像，其中 a 是对应不同数据集的常数

基于均匀采样的大规模扩展法——理论分析

将从以下三个方面对上述方法进行理论分析

- ① 复杂度分析。
- ② 核系数近似一致性分析。
- ③ 额外聚类风险上界分析。

基于均匀采样的大规模扩展法——理论分析

1. 复杂度分析

- ① 步骤 3 中，对 m 个规模为 $s \times s$ 基核矩阵执行 SimpleMKKM 所消耗的时间为 $\mathcal{O}(Ts^3 + Tms^2)$ ，所占用的空间大小为 $\mathcal{O}(ms^2)$ 。
- ② 步骤 4 中，计算 $\mathbf{G}_{\hat{\alpha}_s}$ 和 $\mathbf{W}_{\hat{\alpha}_s}$ 将消耗 $\mathcal{O}((m-1)(ns + s^2))$ 的时间，相应的空间复杂度为 $\mathcal{O}(mns + ms^2)$ 。
- ③ 步骤 5 中，对 $\mathbf{W}_{\hat{\alpha}_s}$ 的特征分解，其耗时为 $\mathcal{O}(s^3)$ 。
- ④ 步骤 6 中，计算 $\mathbf{G}_{\hat{\alpha}_s} \mathbf{U}_s \mathbf{\Lambda}_s^{-1/2}$ 消耗 $\mathcal{O}(ns^2)$ 的时间，后续的 k 均值聚类消耗 $\mathcal{O}(nskT_1)$ 的时间， T_1 是 k 均值聚类的迭代次数。

基于均匀采样的大规模扩展法——理论分析

2. 核系数近似一致性分析

定理 4 (基于均匀采样的 SimpleMKKM 的近似一致性分析)

如果基于均匀采样的近似算法与原算法初始化相同，且学习步长 $\eta \leq c$ ，其中 c 是某个大于 0 的常数。设 *SimpleMKKM* 在 S_n 及其 s 个样本子集 (无放回采样) 上运行，其经历最多 T 次迭代后，算法收敛得到的解分别为 $\hat{\gamma}_n, \hat{\alpha}_s$ ，则有

$$\|\hat{\alpha}_s - \hat{\gamma}_n\|_\infty \lesssim k \sqrt{\left(\frac{1}{s} - \frac{1}{n}\right) \log\left(\frac{4kT}{\delta}\right)}, \quad (24)$$

至少以 $1 - \delta$ 的概率成立。

基于均匀采样的大规模扩展法——理论分析

3. 额外聚类风险上界分析

假设通过基于均匀采样的近似方法学习到的聚类中心为 $\hat{\mathbf{C}}_{n,s}$, 其属于的空间是 $\mathcal{H}_{\hat{\alpha}_s}^k$ 。可利用定理4中关于核系数近似一致性的结果推导出定理5。

定理 5

当随机选择的样本点数量为 s 时, 则

$$\mathbb{E}_{S_n}[\mathcal{W}(\hat{\mathbf{C}}_{n,s}, \hat{\boldsymbol{\alpha}}_s, \rho)] - \mathcal{W}(\mathbf{C}^*, \boldsymbol{\gamma}^*, \rho) \quad (25)$$

有上界为

$$\tilde{\mathcal{O}}\left(\frac{k}{s} + \frac{k}{\sqrt{s}} + \frac{k}{\sqrt{n}}\right). \quad (26)$$

基于奇异值分解的大规模扩展法

下面是基于奇异值分解的大规模扩展法中实现核系数近似的示意图。

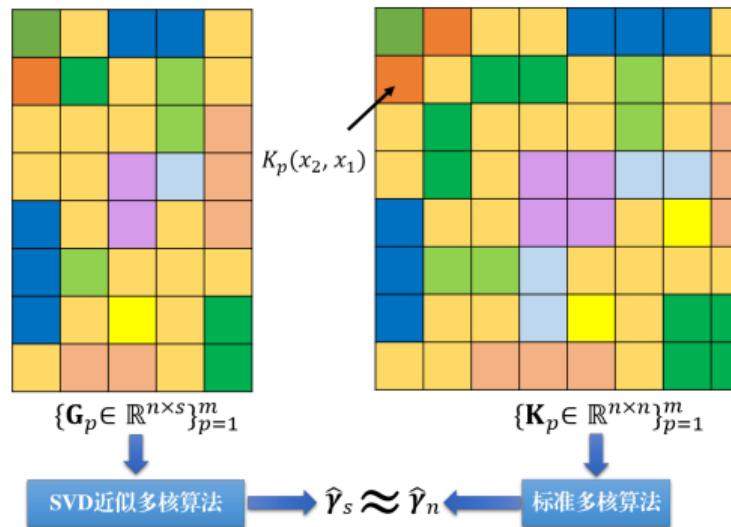


图: 基于奇异值分解的大规模扩展法中的核系数近似

基于奇异值分解的大规模扩展法

为了实现对 $\{\mathbf{G}_p\}_{p=1}^m$ 的融合，提出如下目标式：

目标式

$$\min_{\gamma} g(\gamma), \text{ s.t. } \gamma \in \Delta, \quad (27)$$

其中

$$g(\gamma) = \max_{\mathbf{U}, \mathbf{V}} \frac{1}{\sqrt{ns}} \text{tr}(\mathbf{U}^\top \mathbf{G}_\gamma \mathbf{V}),$$

且 $\mathbf{G}_\gamma = \sum_{p=1}^m \gamma_p^2 \mathbf{G}_p$, 且 $\mathbf{U} \in \mathbb{R}^{n \times k}, \mathbf{V} \in \mathbb{R}^{s \times k}$ 满足 $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_k, \mathbf{V}^\top \mathbf{V} = \mathbf{I}_k$ 。

相应地，第 p 个分量的梯度为

$$\frac{\partial g(\gamma)}{\partial \gamma_p} = \frac{2\gamma_p}{\sqrt{ns}} \text{tr}(\mathbf{U}_*^\top \mathbf{G}_p \mathbf{V}_*). \quad (28)$$

其中 $\mathbf{U}_*, \mathbf{V}_* = \text{argmax}_{\mathbf{U}, \mathbf{V}} \text{tr}(\mathbf{U}^\top \mathbf{G}_\gamma \mathbf{V})$ 。可用与 SimpleMKKM 相同，可用归约梯度法求解。

基于奇异值分解的大规模扩展法

基于奇异值分解的大规模 SimpleMKKM 算法流程如下所示：

Algorithm 3 基于奇异值分解的大规模 SimpleMKKM

- 1: **输入:** 样本集 $S_n = \{x_i\}_{i=1}^n$; 锚点集 $\{a_t\}_{t=1}^s$; m 个基核函数 $\{K_p(\cdot, \cdot)\}_{p=1}^m$; 聚类簇数 k .
 - 2: **输出:** 聚类结果.
 - 3: 计算 m 个基核相似度矩阵 $\{\mathbf{G}_p\}_{p=1}^m$, 其元素为 $\mathbf{G}_p(i, t) = K_p(x_i, a_t)$ ($i \in [n], t \in [s]$).
 - 4: 初始化 $\gamma^{(0)} = \frac{1}{m}[1, \dots, 1]$.
 - 5: **while** 未收敛 **do**
 - 6: 对 $\mathbf{G}_{\gamma^{(l)}} = \sum_{p=1}^m (\gamma_p^{(l)})^2 \mathbf{G}_p$ 运行秩 k 的截断奇异值分解, 并令 \mathbf{U}_* , \mathbf{V}_* 为其左右奇异向量.
 - 7: 利用归约梯度法计算其下降方向 $\hat{\mathbf{d}}^{(l)}$.
 - 8: $\gamma^{(l+1)} \leftarrow \gamma^{(l)} + \eta \hat{\mathbf{d}}^{(l)}$.
 - 9: **end while**
 - 10: 在 \mathbf{U}_* 上运行标准的 k 均值聚类算法获取最终的聚类结果.
-

基于奇异值分解的大规模扩展法——理论分析

将从以下三个方面对上述方法进行理论分析

- ① 复杂度分析。
- ② 核系数近似一致性分析。
- ③ 聚类指示矩阵的近似程度分析。

基于奇异值分解的大规模扩展法——理论分析

1. 复杂度分析

- ① 需要存储 m 个大小为 $n \times s$ 的矩阵。因此，其存储复杂度为 $\mathcal{O}(msn)$ 。
- ② 对规模为 $n \times s$ 的矩阵进行奇异值分解，其时间复杂度为 $\mathcal{O}(s^2 n)$ 。
- ③ 计算 m 个规模为 $n \times s$ 的矩阵之和，时间复杂度为 $\mathcal{O}(msn)$ 。

基于奇异值分解的大规模扩展法——理论分析

2. 核系数近似程度

定理 6

在相同的初始化，且学习步长不大于某个常数的情况下，经 T 迭代收敛后，设 *SimpleMKKM* 和所提出的基于奇异值分解的近似算法得到的解分别为 $\hat{\gamma}_n, \hat{\gamma}_s$ 。那么，

$$\|\hat{\gamma}_s - \hat{\gamma}_n\|_\infty \lesssim k \sqrt{\left(\frac{1}{s} - \frac{1}{n}\right) \log\left(\frac{4kT}{\delta}\right)}, \quad (29)$$

以至少 $1 - \delta$ 的概率成立。

基于奇异值分解的大规模扩展法——理论分析

3. 聚类指示矩阵的近似程度

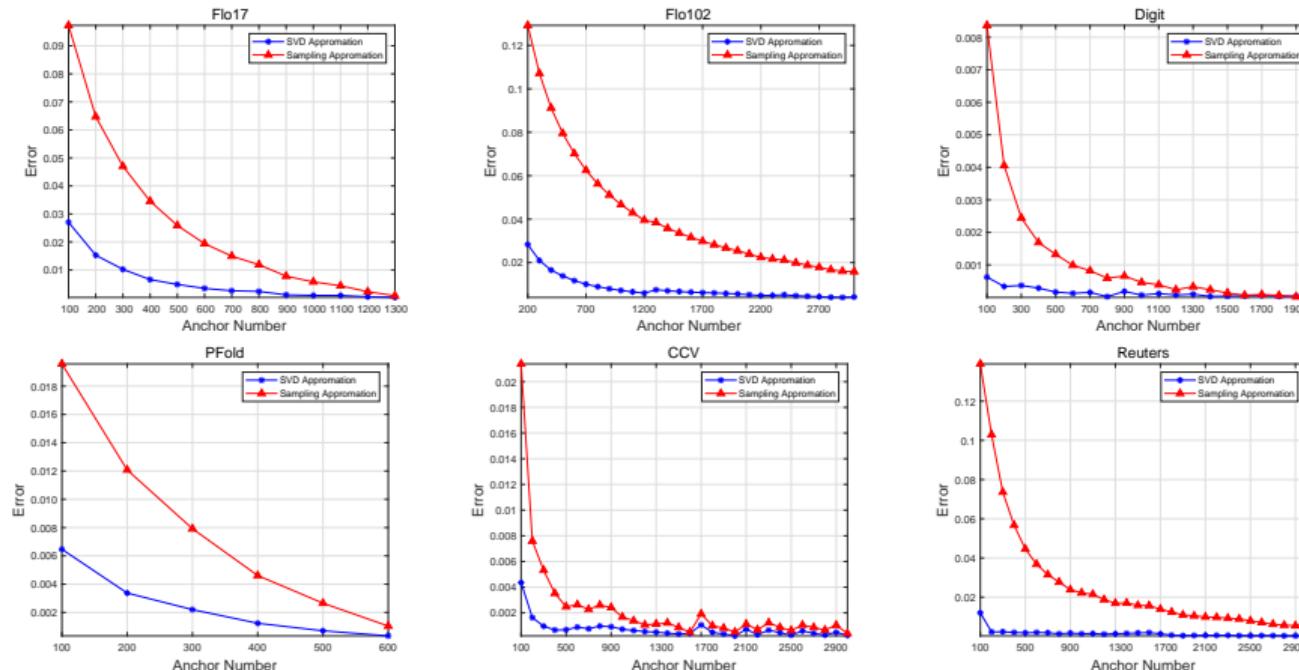
定理 7

基于与定理6相同的假设，当 f_n, g 都收敛时，假设原 SimpleMKKM 和基于奇异值分解的近似方法得到的聚类指示矩阵分别为 \mathbf{H}_* 和 \mathbf{U}_* 。那么，不等式

$$\|\mathbf{H}_* \mathbf{H}_*^\top - \mathbf{U}_* \mathbf{U}_*^\top\|_F \lesssim k \sqrt{\left(\frac{1}{s} - \frac{1}{n}\right) \log\left(\frac{4kT}{\delta}\right)} \quad (30)$$

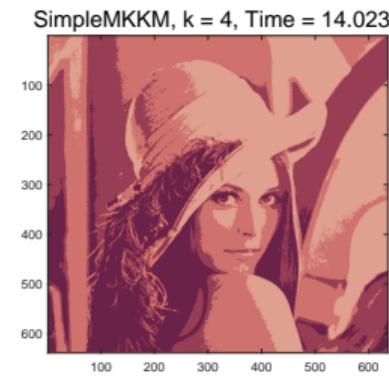
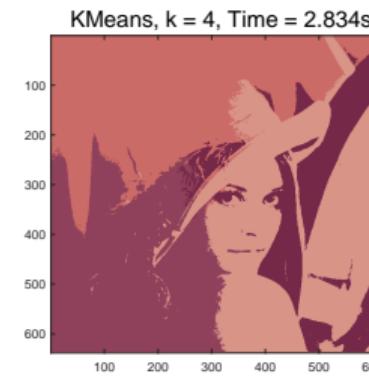
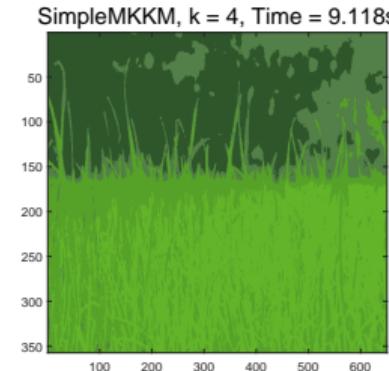
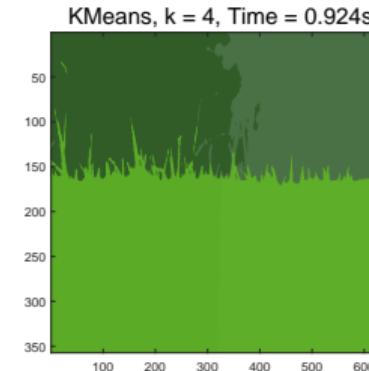
至少以 $1 - \delta$ 的概率成立。

基于奇异值分解的大规模扩展法——核系数近似效果



图：基于奇异值分解的近似算法与原始算法学习到的核系数的差异。蓝色曲线和红色曲线分别是基于奇异值分解和基于均匀采样近似算法关于原始算法核系数差异的值

在图像分割任务中的应用



目录

1 研究背景及研究现状

2 聚类算法统计理论：泛化性和一致性

3 多核聚类算法的大规模扩展

4 问答

问答

请各位批评指正！