文章编号：2095-2163(2020)02-0100-03 中图分类号：TP18 文献标志码：A

多任务孪生支持向量聚类算法

朱文文[1]，黄成泉[2]，阮 丽[1]
（1 贵州民族大学 数据科学与信息工程学院，贵阳 550025；2 贵州民族大学 工程技术人才实践训练中心，贵阳 550025）

摘 要： 多任务学习是一种利用其它任务来提高任务泛化性能的学习范式。孪生支持向量聚类是一种功能强大的聚类方法，为了提高孪生支持向量聚类模型的聚类性能，受多任务学习的启发，将模型扩展到多任务学习场景，提出多任务孪生支持向量聚类算法，并通过求解一系列二次规划问题，确定聚类中心平面。同时学习多个相关任务的经验和理论表明，相对于独立学习每个任务，多任务孪生支持向量聚类算法具有良好的聚类性能。

关键词： 多任务学习；孪生支持向量机；平面聚类

Multi-task twin support vector clustering algorithm

ZHU Wenwen[1], HUANG Chengquan[2], RUAN Li [1]
(1 School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China;
2 Engineering Training Center, Guizhou Minzu University, Guiyang 550025, China)

【Abstract】 Multi-task learning is a learning paradigm that uses other tasks to improve task generalization performance. The twin support vector machine clustering is one of the powerful clustering methods, in order to improve the performance of the twin support vector clustering, inspired by multi-tasking learning, the model is extended to multi-task learning scenarios, multi-task twin clustering support vector machine algorithm is proposed, cluster center planes is determined by solving a series of quadratic programming problems. The experience and theory of learning multiple related tasks at the same time show that the multi-task twin support vector clustering algorithm has good clustering performance compared with learning each task independently.

【Key words】 multi-tasking learning; twin support vector machine; plane-based clustering

0 引 言

传统机器学习方法，如分类和聚类，是假设要处理的数据必须来自于同一分布，当要处理的数据是来自于不同分布时，传统机器学习方法需要分别对每个分布下的数据、即每个任务进行学习，这样就导致较多的时间花费，且忽略了任务间的相关性，特别是当某一任务的数据有限时，采用传统机器学习技术并不能够获得很好的效果，多任务学习正是为了应对这种情况而被提出的。

聚类在计算机视觉、文本挖掘、生物信息学和信号处理等多个领域都有应用。聚类是机器学习中最基本的方法之一，其目的是将数据点划分为簇，使得同一个簇中的数据具有较大的相似性，不同簇之间的数据具有较大的差异性。考虑到传统的基于点的聚类方法、如 K-均值是根据数据集的分布将数据划分到所属集群中，当数据没有分布在多个集群点时，传统的基于点的聚类方法聚类性能很差。因此，本文在孪生支持向量聚类[1]模型基础上基于平面进行聚类。为了保持任务间的差异性、又充分利用任务间的相关性，从而整体上提高每个任务的聚类性能，本次研究把单任务孪生支持向量聚类扩展到多任务学习框架下，提出了多任务孪生支持向量聚类算法，多任务孪生支持向量聚类假设任务间共享一个公共的表示，同时学习多个相关任务，从而整体上提高所有任务的聚类性能。

1 孪生支持向量聚类

在孪生支持向量机的研究基础上，Wang 等人[1]提出了孪生支持向量聚类（twin support vector clustering，TWSVC），在 TWSVC 中，为了寻找 k 个聚类中心平面 $\omega_i^{\mathrm{T}}x_i + b_i = 0, i = 1, \cdots, k$，通过求解以下聚类模型：

$$\min_{\omega_i, b_i, \xi_i, X_i} \frac{1}{2} \| X_i\omega_i + b_i e \|^2 + ce^{\mathrm{T}}\xi_i$$

$$\text{s.t.} \left| \hat{X}_i\omega_i + b_i e \right| \geqslant e - \xi_i,\ \xi_i \geqslant 0. \quad (1)$$

其中，$c > 0$ 为惩罚参数；$\xi_i > 0$ 为松弛向量；ω_i 为超平面的法向量；b_i 为超平面的偏移量。

作者简介：朱文文（1993-），女，硕士研究生，主要研究方向：统计信号处理、计算机视觉、机器学习；黄成泉（1976-），男，博士，教授，主要研究方向：计算机软件与理论、计算机应用技术、嵌入式系统等；阮 丽（1994-），女，硕士研究生，主要研究方向：统计信号处理、计算机视觉、机器学习。

收稿日期：2019-11-25

分析可知，式(1)为一个二次规划问题。其模型的几何意义为：第 X_i 个样本点在 TWSVC 中需要尽可能靠近第 i 个聚类中心平面，而远离其他类的中心平面。

通过核技巧将 TWSVC 扩展到非线性情况下，非线性 TWSVC 在一个合适的内核生成空间中寻找 k 个聚类中平面，即：

$$K(x,X)\,u_i+\gamma_i=0,\ i=1,2,\cdots,k, \tag{2}$$

其中，$K(\cdot,\cdot)$ 是一个适当的核函数。

非线性孪生支持向量聚类模型为：

$$\min_{u_i,\gamma_i,\eta_i,X_i}\frac{1}{2}\|K(X_i,X)\,\boldsymbol{u}_i+\gamma_i\boldsymbol{e}\|^2+c\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\eta}_i$$
$$\text{s.t.}\ |K(\hat{X}_i,X)\,\boldsymbol{u}_i+\gamma_i|\geqslant\boldsymbol{e}-\boldsymbol{\eta}_i,\boldsymbol{\eta}_i\geqslant 0. \tag{3}$$

其中，$\boldsymbol{\eta}_i(i=1,2,\cdots,k)$ 为松弛向量。

2　多任务孪生支持向量聚类

基于前述工作，将孪生支持向量聚类扩展到多任务学习框架下，研究认为所有的任务都有一个公共的表示 $[\omega_i;b_i]$，$[\omega_{it};b_{it}]$ 表示任务 t 与共享的公共表示之间的偏差。多任务孪生支持向量聚类模型为：

$$\min_{\omega_i,\omega_{it},b_i,b_{it},\xi_{it}}\frac{1}{2}\|\boldsymbol{X}_i\boldsymbol{\omega}_i+b_i\boldsymbol{e}\|^2+\frac{1}{2}\sum_{t=1}^{T}\rho_t\|\boldsymbol{X}_{it}\boldsymbol{\omega}_{it}+b_{it}\boldsymbol{e}\|^2+c\sum_{t=1}^{T}\boldsymbol{e}_t^{\mathrm{T}}\boldsymbol{\xi}_{it}$$
$$\text{s.t.}\ |\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i+b_i\boldsymbol{e}|+|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}+b_{it}\boldsymbol{e}|\geqslant\boldsymbol{e}-\boldsymbol{\xi}_{it},\ \boldsymbol{\xi}_{it}\geqslant 0\quad(i=1,2,\cdots k). \tag{4}$$

其中，$\boldsymbol{X}_{it}$ 表示第 t 个任务的第 i 类样本，$\hat{\boldsymbol{X}}_{it}$ 表示其它远离 X_{it} 的样本；$\boldsymbol{X}_i$ 表示所有任务的第 i 类样本，$\hat{\boldsymbol{X}}_i$ 表示远离 $\boldsymbol{X}_i$ 的样本；$\boldsymbol{e}$ 为数值全为 1 的列向量。

类似于 TWSVC 求解方法，上述优化问题可以通过凹凸过程(CCCP)[2]求解，该过程将式(4)中的第 i 个问题分解为一系列具有初始 ω_i^0 和 b_i^0 的凸二次子问题，此时有：

$$\min_{\omega_i,\omega_{it},b_i,b_{it},\xi_{it}}\frac{1}{2}\|\boldsymbol{X}_i\boldsymbol{\omega}_i^{j+1}+b_i^{j+1}\boldsymbol{e}\|^2+\frac{1}{2}\sum_{t=1}^{T}\rho_t\|\boldsymbol{X}_{it}\boldsymbol{\omega}_{it}^{j+1}+b_{it}^{j+1}\boldsymbol{e}\|^2+c\sum_{t=1}^{T}\boldsymbol{e}_t^{\mathrm{T}}\xi_{it}^{j+1}$$
$$\text{s.t.}\ T|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^{j+1}+b_i^{j+1}\boldsymbol{e}|+T|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^{j+1}+b_{it}^{j+1}\boldsymbol{e}|\geqslant\boldsymbol{e}-\boldsymbol{\xi}_{it}^{j+1},$$
$$\boldsymbol{\xi}_{it}^{j+1}\geqslant 0. \tag{5}$$

其中，子问题的指数 $j=0,1,2,\cdots,T(\cdot)$ 定义为一阶泰勒展开式。

通过引入 $|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}|$ 关于 $\boldsymbol{\omega}_i^j,b_i^j$ 的次梯度[3]与 $|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^j+b_{it}^j\boldsymbol{e}|$ 关于 $\boldsymbol{\omega}_{it}^j,b_{it}^j$ 的次梯度，研究得到：

$$\nabla(|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}|)=diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}))\,[\hat{\boldsymbol{X}}_{it},\boldsymbol{e}]$$
$$\nabla(|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^j+b_{it}^j\boldsymbol{e}|)=diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^j+b_{it}^j\boldsymbol{e}))\,[\hat{\boldsymbol{X}}_{it},\boldsymbol{e}]$$

注意到：

$$|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}|=diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}))\,(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e})$$
$$|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^j+b_{it}^j\boldsymbol{e}|=diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^j+b_{it}^j\boldsymbol{e}))\,(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^j+b_{it}^j\boldsymbol{e})$$

由此可以得到：

$$T|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^{j+1}+b_i^{j+1}\boldsymbol{e}|=|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}|+\nabla(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+\boldsymbol{e})([\boldsymbol{\omega}_i^{j+1};b_i^{j+1}]-[\boldsymbol{\omega}_i^j;b_i^j])=diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}))\,[\hat{\boldsymbol{X}}_{it},\boldsymbol{e}]([\boldsymbol{\omega}_i^{j+1};b_i^{j+1}])+(|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}|-diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}))\,[\hat{\boldsymbol{X}}_{it},\boldsymbol{e}]\,[\boldsymbol{\omega}_i^j;b_i^j])=diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}))\,(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^{j+1}+b_i^{j+1}\boldsymbol{e}), \tag{6}$$

同理可得：

$$T|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^{j+1}+b_{it}^{j+1}\boldsymbol{e}|=diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^j+b_{it}^j\boldsymbol{e}))\cdot(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^{j+1}+b_{it}^{j+1}\boldsymbol{e})$$

因此模型(4)的约束为：

$$T|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^{j+1}+b_i^{j+1}\boldsymbol{e}|+T|\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^{j+1}+b_{it}^{j+1}\boldsymbol{e}|=diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}))\,(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^{j+1}+b_i^{j+1}\boldsymbol{e})+diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^j+b_{it}^j\boldsymbol{e}))\,(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^{j+1}+b_{it}^{j+1}\boldsymbol{e}), \tag{7}$$

从而，模型(4)等价为：

$$\min_{\omega_i,\omega_{it},b_i,b_{it},\xi_{it}}\frac{1}{2}\|\boldsymbol{X}_i\boldsymbol{\omega}_i^{j+1}+b_i^{j+1}\boldsymbol{e}\|^2+\frac{1}{2}\sum_{t=1}^{T}\rho_t\|\boldsymbol{X}_{it}\boldsymbol{\omega}_{it}^{j+1}+b_{it}^{j+1}\boldsymbol{e}\|^2+c\sum_{t=1}^{T}\boldsymbol{e}_t^{\mathrm{T}}\xi_{it}^{j+1}$$
$$\text{s.t.}\ diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^j+b_i^j\boldsymbol{e}))\,(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_i^{j+1}+b_i^{j+1}\boldsymbol{e})+diag(sign(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^j+b_{it}^j\boldsymbol{e}))\,(\hat{\boldsymbol{X}}_{it}\boldsymbol{\omega}_{it}^{j+1}+b_{it}^{j+1}\boldsymbol{e})\geqslant\boldsymbol{e}-\boldsymbol{\xi}_{it}^{j+1},$$
$$\boldsymbol{\xi}_{it}^{j+1}\geqslant 0, \tag{8}$$

受支持向量机[4-5]、孪生支持向量机[6-7]的启发，求解 $[\omega_i^{j+1};b_i^{j+1}]$ 与 $[\omega_{it}^{j+1};b_{it}^{j+1}]$，通过求解(8)的对偶问题：

$$\min\ -\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\left[\boldsymbol{G}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{G}^{\mathrm{T}}+\frac{1}{\rho_t}\boldsymbol{G}_t(\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{H}_t)^{-1}\boldsymbol{G}_t^{\mathrm{T}}\right]\boldsymbol{\alpha}+

$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{e}$

$$\text{s.t. } 0 \leqslant \boldsymbol{\alpha} \leqslant c\boldsymbol{e}, \tag{9}$$

其中,

$$G = diag(sign(\hat{\boldsymbol{X}}_{it}\omega_i^j + b_i^j\boldsymbol{e}))\ [\hat{\boldsymbol{X}}_{it} \quad \boldsymbol{e}],$$

$$G_t = diag(sign(\hat{\boldsymbol{X}}_{it}\omega_{it}^j + b_{it}^j\boldsymbol{e}))\ [\hat{\boldsymbol{X}}_{it} \quad \boldsymbol{e}],$$

$$\boldsymbol{H} = [\boldsymbol{X}_i \quad \boldsymbol{e}], H_t = [\boldsymbol{X}_{it} \quad \boldsymbol{e}],$$

并且 $\boldsymbol{\alpha} \in R$ 是拉格朗日乘子向量。

问题(9)是一个凸 QPP 问题,通过逐次超松弛[8]方法可以有效地解决,该方法是求解线性方程组的迭代方法,并成功地推广到求解上述问题[9],通过以下式子可得式(9)的解,从而得到式(8)的解:

$$[\omega_i^{j+1} + \omega_{it}^{j+1}; b_i^{j+1} + b_{it}^{j+1}] = (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\alpha} + \frac{1}{\rho_t}(\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{H}_t)^{-1}\boldsymbol{G}_t^{\mathrm{T}}\boldsymbol{\alpha}, \tag{10}$$

综上,对于 $i = 1,2,\cdots,k$, 式(4)可以通过以下步骤来求解:

(1)初始化 $[\omega_i^0 + \omega_{it}^0;\ b_i^0 + b_{it}^0]$。

(2)对于 $j = 0,1,2,\cdots$, 通过式(10)求 $[\omega_i^{j+1} + \omega_{it}^{j+1}; b_i^{j+1} + b_{it}^{j+1}]$。

(3)如果 $\| [\omega_i^{j+1} + \omega_{it}^{j+1}; b_i^{j+1} + b_{it}^{j+1}] - [\omega_i^j + \omega_{it}^j;\ b_i^j + b_{it}^j] \| \leqslant \varepsilon$, 停止迭代,并设置 $\omega_i = \omega_i^{j+1} + \omega_{it}^{j+1}$, $b_i = b_i^{j+1} + b_{it}^{j+1}$。

最后,数据样本的聚类标签可由 $y = \underset{i}{\operatorname{argmin}}\{|(\omega_i + \omega_{it})^{\mathrm{T}}x + (b_i + b_{it})| \ i = 1,\cdots,k\}$ 更新得到。

通过内核技巧将上面的线性多任务孪生支持向量机扩展到多任务非线性孪生支持向量机,即:

$$\min_{u_i,u_{it},\gamma_i,\gamma_{it},\eta_{it}} \frac{1}{2}\| K(\boldsymbol{X}_i,\boldsymbol{X})\boldsymbol{u}_i + \gamma_i\boldsymbol{e}\|^2 + \frac{1}{2}\sum_{t=1}^{T}\rho_t\| K(\boldsymbol{X}_{it},\boldsymbol{X})\boldsymbol{u}_{it} + \gamma_{it}\boldsymbol{e}\|^2 + c\sum_{t=1}^{T}\boldsymbol{e}_t^{\mathrm{T}}\boldsymbol{\eta}_{it}$$

$$\text{s.t. } |K(\hat{\boldsymbol{X}}_{it},\boldsymbol{X})\boldsymbol{u}_i + \gamma_i| + |K(\hat{\boldsymbol{X}}_{it},\boldsymbol{X})\boldsymbol{u}_{it} + \gamma_{it}| \geqslant \boldsymbol{e} - \boldsymbol{\eta}_{it},\ \boldsymbol{\eta}_{it} \geqslant 0(i = 1,2,\cdots k). \tag{11}$$

其中,$\boldsymbol{\eta}_{it}$ 为松弛向量,模型(11)的优化过程类似于上述线性情况的优化过程,此处不再赘述。

3 结束语

本文在孪生支持向量聚类模型上进行改进,将孪生支持向量聚类模型扩展到多任务学习框架下,提出了多任务孪生支持向量聚类算法,通过求解一系列二次规划问题确定聚类中心平面。同时学习多个相关任务的经验和理论表明,相对于独立学习每个任务,该算法利用任务间的相关性来提升所有任务的聚类性能。

参考文献

[1] WANG Z, SHAO Y H, BAI L, et al. Twin support vector machine for clustering[J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(10):2583.

[2] YUILLE A L, RANGARAJAN A. The concave-convex procedure (CCCP)[J]. Advances in Neural Information Processing Systems, 2002, 2:1033.

[3] CHEUNG P M, KWOK J T. A regularization framework for multiple-instance learning[C]// Proceedings of the Twenty-Third International Conference on Machine Learning (ICML 2006). Pittsburgh, Pennsylvania, USA:dblp, 2006:193.

[4] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3):273.

[5] DENG N, TIAN Y, ZHANG C. Support vector machines: Optimization based theory, algorithms, and extensions[M]. Boca Raton: Chapman and Hall/CRC, 2012.

[6] JAYADEVA, KHEMCHANDANI R, CHANDRA S. Twin support vector machines for pattern classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligent, 2007, 29(5):905.

[7] SHAO Yuanhai, ZHANG Chunhua, WANG Xiaobo, et al. Improvements on twin support vector machines[J]. IEEE Transactions on Neural Networks, 2011, 22(6):962.

[8] MANGASARIAN O L, MUSICANT D R. Successive overrelaxation for support vector machines[J]. IEEE Transactions on Neural Networks, 1999, 10(5):1032.

(上接第 99 页)

4 结束语

本文实现了一种货车位姿调整系统,尤其适用于大型重载货车的位姿调整。首先,根据系统的功能需求,指出了该系统的整体设计方向,继而给出设计方案,明确设计的核心主要为位姿测量系统和机械执行系统;其次,根据设计要求和功能需求,完成了机械执行系统的设计;最后,对位姿测量系统进行了研究,其中包括测量方案的设计和信息处理的原理设计。

参考文献

[1] 张建明. 现代物流管理[M]. 武汉:武汉大学出版社, 2013.

[2] 徐东,黄松和,熊楚良. 码垛机器人 Z 轴升降机构动力学分析[J]. 机械设计与制造,2015(9):25.

[3] 陈显龙,陈晓龙,罗新伟,等. 车辆尺寸的测量方法和装置: 中国,CN102679889[P]. 2012-09-19.

[4] 卢卓均. 基于 SCP 范式的中国啤酒产业分析[J]. 现代企业, 2017(10):27.

[5] 李敏. 衡水九州啤酒有限公司物流管理研究[D]. 石家庄:河北科技大学,2012.

[6] 薛长松. 基于 DM642 的双目视觉控制系统研究[D]. 开封:河南大学,2007.