

基于光滑表示的半监督分类算法

王省¹ 康昭²

1 电子科技大学信息与软件工程学院 成都 610054

2 电子科技大学计算机科学与工程学院 成都 611731

(wangxing1027@gmail.com)



摘要 近年来,基于图的半监督分类是机器学习与数据挖掘领域的研究热点之一。该类方法一般通过构造图来挖掘数据中隐含的信息,并利用图的结构信息来对无标签样本进行分类。因此,半监督分类的效果严重依赖于图的质量。文中提出了一种基于光滑表示的半监督分类算法。具体来说,此方法通过应用一个低通滤波器来实现数据的平滑,然后将光滑数据用于半监督分类。此外,所提方法将常见的图构造和标签传播集成到一个统一的优化框架中,使它们互相促进,从而避免低质量图导致的次优解。对人脸和物品数据集进行大量实验,结果表明,所提 SRSSC 算法在大部分情况下都优于其他算法,从而证明了光滑表示的重要性。

关键词 半监督分类;信号过滤;图方法;相似度计算;表征学习

中图法分类号 TP181

Smooth Representation-based Semi-supervised Classification

WANG Xing¹ and KANG Zhao²

1 School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

2 School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract Graph-based semi-supervised classification is a hot topic in machine learning and data mining. In general, this method discovers the hidden information by constructing a graph and predicts the labels for unlabeled samples based on the structural information of graph. Thus, the performance of semi-supervised classification heavily depends on the quality of graph. In this work, we propose to perform semi-supervised classification in a smooth representation. In particular, a low-pass filter is applied on the data to achieve a smooth representation, which in turn is used for semi-supervised classification. Furthermore, a unified framework which integrates graph construction and label propagation is proposed, so that they can be mutually improved and avoid the sub-optimal solution caused by low-quality graph. Extensive experiments on face and subject data sets show that the proposed SRSSC outperforms other state-of-the-art methods in most cases, which validates the significance of smooth representation.

Keywords Semi-supervised classification, Signal filtering, Graph-based method, Similarity measure, Representation learning

1 引言

由于传统的监督分类需要对已标记数据进行训练,而标记样本通常需要耗费大量时间或人力。与此同时,未标记数据相对容易收集,因此研究能处理它们的算法尤为重要。半监督学习旨在通过学习大量未标记的数据以及小部分有标签数据来构建更好的分类器,从而解决上述问题。相对而言,半监督学习只需要较少的人力就能达到更高的准确性,近年来受到了广泛的关注^[1]。

在众多半监督分类方法^[2-4]中,基于图方法的研究是近年来机器学习与数据挖掘领域的研究热点之一。这类方法通常由两步构成:首先,将数据集映射为一个图,数据集中每个样本对应于图中的一个节点,若两个样本之间的相似度很高(或

相关性很强),则对应的节点之间存在一条边,边的强度正比于两样本之间的相似度(或相关性)。其次,通过图上的局部平滑程度来推断缺失的标签,即两个样本越相似,它们具有相同标签的可能性就越大。因此,大量的算法关注于构建图或标签传播。例如,Jebara等^[5]提出了一种基于 b-matching 的图构造方法;Cheng等^[6]通过将每个数据点分解为 L_1 稀疏线性组合来衡量数据点之间的相似性,从而构造图的相似度矩阵;Li等^[7]提出了一种基于低秩子空间的图构造方法;Wang等^[8]提出了一种基于线性邻域的标签传播方法;Bo等^[9]提出了一种用于多视图的半监督学习算法,在多视图数据的潜在空间上构建图。

尽管现有的算法已经在各种实际应用中取得了不错的效果,但其在某些方面依然受到限制。1)大多数图都是直接基

收稿日期:2020-07-11 返修日期:2020-10-01 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61806045)

This work was supported by the National Natural Science Foundation of China(61806045).

通信作者:康昭(zkang@uestc.edu.cn)

于原始数据构建的。但由于原始数据受到污染,建立的图可能无法准确反映样本之间的潜在关系^[10]。同时,大量的研究表明,图的质量对于后续的任务执行至关重要^[11-12]。因此,如何构造具有鲁棒性的图至关重要。2)现有方法通常将图的构造与标签传播视为两个单独的步骤,但这样可能难以得到全局最优解,即在第一步中构建的图对于后续任务处理可能并不是最优的。

面对上述限制,本文提出了一种基于光滑表示的半监督分类算法(SRSSC)。它由3个部分组成:首先,对原始数据进行图滤波,使原始数据更加光滑;其次,在光滑表示的数据上学习一个图,并在图上进行标签传播,对这一过程进行联合优化;最后,根据分类结果进行迭代更新,从而同时优化图结构和标签传播。本文的主要贡献如下:1)提出了一种用于半监督分类的图过滤框架。这种框架可以对原始数据进行过滤,从而获得一种更加光滑的数据表示,这也为表征学习提供了一种新的策略。2)提出了一种基于光滑表示的半监督分类算法。其将常见的图构造和标签传播集成到一个统一的框架中。3)对数据集进行了广泛的实验。与具有代表性的半监督分类方法相比,本文方法在某些方面展现了其优越性。

为了避免混淆,此处将给出本文主要使用的部分符号。矩阵用大写字母表示,数据点(样本点)使用列向量并用小写字母表示。将半监督分类问题的训练数据矩阵表示为 $\mathbf{X} = [x_1, \dots, x_l, \dots, x_{l+u}]^T$, 其中 $l+u=N$, l 和 u 分别是有标签和无标签数据的数目, $x_i \in \mathcal{R}^m$ 为数据样本(数据点), m 为特征数。 c 为数据的类别总数, \mathbf{y} 为标签矩阵, 当第 i 个样本属于第 j 个类别时, $y_{ij} = 1$, 否则 $y_{ij} = 0$ 。向量 $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}] \in \mathcal{R}^m$ 的 ℓ_2 -范数定义为:

$$\|\mathbf{x}_i\|_2 = \sqrt{\sum_{j=1}^m (x_{ij})^2}$$

2 相关工作

近年来,基于图的半监督分类得到了广泛的关注。例如, Zhu 等^[13]设计了一种基于高斯场和谐波函数(GFHF)的半监督分类,其利用了图上的高斯随机场的谐波特性。尽管该算法已经得到了普及,但其分类性能在很大程度上仍然取决于输入图。

一些半监督分类算法考虑到了构造图的鲁棒性对分类性能的影响。例如, Nie 等^[14]提出了一种基于最小化谱嵌入的 ℓ_1 -范数的半监督分类算法; Gu 等^[15]提出了一种基于放射子空间稀疏表示的、图构造方法,这种方法能够快速对新来样本点进行分类,并且继承了稀疏表示的、能够自适应进行邻域选择以及具有较高判别性的优点; Kang 等在 Robust Graph Learning From Noisy Data(RGC)^[16]中提出了一种鲁棒性的图学习方法。通过将原始数据分解为干净和有噪声的两部分,在干净数据上构建一个图,并根据构建的图进行标签传播。在构建图的部分, RGC 旨在解决以下问题:

$$\begin{aligned} \min_{D, E, S} & \|D\|_* + \alpha \|E\|_1 + \gamma \|S\|_F^2 \\ \text{s. t. } & X = D + E, S1 = 1, 0 \leq S \leq 1 \end{aligned} \quad (1)$$

其中, \mathbf{S} 为图的相似度矩阵, D 为数据干净的部分, E 为数据噪声(错误)的部分, α 和 γ 用于平衡各项之间的影响。构建好鲁棒性的图后, RGC 利用一种局部和全局一致的方法(LGC)来执行分类任务^[17], 即通过以下优化方法获得数据标签 F :

$$\min_F \text{Tr} \{F^T L F + \gamma (F - Y)^T (F - Y)\} \quad (2)$$

尽管上述算法在很多方面取得了成效,可以避免直接从嘈杂数据中构造图,但由于图构建和标签传播是分开进行的,其分类性能仍然可能受到影响。

随着信息技术的发展,数据量剧增且数据变得更加复杂,如何有效地提取数据特征并将其表示出来非常重要。近年来,如何将原始数据转换为能够被机器学习有效开发的一种形式也引起了广泛的关注。Kang 等^[18]提出了一种新的表征学习方法,即充分挖掘样本之间的关系信息。具体来说,他们使用以下方式来构建样本间的相似度矩阵 \mathbf{Z} 。

$$\begin{aligned} \min_Z & \frac{1}{2} \text{Tr}(\mathbf{K} - 2\mathbf{K}\mathbf{Z} + \mathbf{Z}^T \mathbf{K}\mathbf{Z}) + \\ & \alpha \|\mathbf{K} - \mathbf{Z}^T \mathbf{K}\mathbf{Z}\|_F^2 + \beta \rho(\mathbf{Z}) \\ \text{s. t. } & \text{diag}(\mathbf{Z}) = 0 \end{aligned} \quad (3)$$

其中, \mathbf{K} 是关于数据集 X 的核矩阵, α 和 β 用于平衡各项之间的影响。通过解决上述问题,可以得到一个低秩或稀疏的矩阵 \mathbf{Z} 。该矩阵可以用于解决其他常见问题,尤其是对样本相似度敏感的问题,如聚类、分类等。

3 图滤波

本节提出了一种图过滤器,这种过滤器能够将原始数据过滤,从而获得一种更加光滑的数据表示。

在原始数据矩阵 \mathbf{X} 上建立一个图 $G = (V, S)$, 其中 $\mathbf{S} = \{s_{ij}\} \in \mathcal{R}^{N \times N}$ 为图的相似度矩阵, $\mathbf{D} = \text{diag}(d_1, \dots, d_N)$ 为度矩阵, $d_i = \sum_{j=1}^N s_{ij}$ 为顶点 v_i 的度。归一化图拉普拉斯矩阵 $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{S} \mathbf{D}^{-\frac{1}{2}}$ 。由于 \mathbf{L} 是一个半正定矩阵,因此 \mathbf{L} 可以分解为 $\mathbf{L} = \mathbf{U}\mathbf{A}\mathbf{U}^{-1}$, 其中对角矩阵 $\mathbf{A} = \text{diag}(\lambda_1, \dots, \lambda_N)$ 按照特征值从小到大依次排序,正交矩阵 $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_N]$ 是其对应的特征向量(或者标准正交基)。根据文献^[19]所述, \mathbf{L} 的特征向量集可视为图的傅里叶基,特征值 λ_i 为对应的频率。

定义信号 $\mathbf{h} = [h(v_1), h(v_2), \dots, h(v_N)]^T$ 可以表示为特征向量的线性组合,即:

$$\mathbf{h} = \sum_{i=1}^N t_i \mathbf{u}_i = \mathbf{U}\mathbf{t} \quad (4)$$

其中, $\mathbf{t} = [t_1, t_2, \dots, t_N]^T$ 是相关系数, t_i 的绝对值表示基信号 u_i 的强度。 \mathbf{h} 的平滑程度可以由式(5)来衡量:

$$\begin{aligned} E_h & = \frac{1}{2} \sum_{i,j=1}^N s_{ij} \left\| \frac{h_i}{\sqrt{d_i}} - \frac{h_j}{\sqrt{d_j}} \right\|_2^2 = \mathbf{h}^T \mathbf{L} \mathbf{h} \\ & = (\mathbf{U}\mathbf{t})^T \mathbf{L} \mathbf{U}\mathbf{t} = \sum_{i=1}^N t_i^2 \lambda_i \end{aligned} \quad (5)$$

式(5)表明,较小的 λ_i 将对应更小的 E_h , 基信号也就更平滑,故光滑信号 \mathbf{h} 应包含更多的低频信号而非高频信号^[20]。

因此,可以设计一种低通的图滤波器 g 来使得图信号更加平滑。假设 $g(\lambda_i)$ 为一个低频通过的响应函数,则通过滤波器后的信号 $\bar{\mathbf{h}}$ 为:

$$\begin{aligned} \bar{\mathbf{h}} & = \sum_{i=1}^N g(\lambda_i) t_i \mathbf{u}_i = \mathbf{U} \mathbf{G}(\Lambda) \mathbf{t} \\ & = \mathbf{U} \mathbf{G}(\Lambda) \mathbf{U}^{-1} \mathbf{h} \end{aligned} \quad (6)$$

其中, $\mathbf{G}(\Lambda) = \text{diag}(g(\lambda_1), g(\lambda_2), \dots, g(\lambda_N))$ 。为了保留低频信号而去除高频信号,较小的 λ_i 应对应较大的 $g(\lambda_i)$, 而较大的 λ_i 应对应较小的 $g(\lambda_i)$ 。由于拉普拉斯算子 L 的特征值 λ_i

的范围为 $[0, 2]$, 滤波器 g 的其中一种选择是 $g(\lambda_i) = \left(1 - \frac{\lambda_i}{2}\right)^k$, 其中正整数 k 用于捕获 k 跳邻域关系^[21-22]。通过滤波器后的信号可以表示为:

$$\bar{h} = U \left(I - \frac{\Lambda}{2} \right)^k U^{-1} h = \left(I - \frac{L}{2} \right)^k h \quad (7)$$

由于数据矩阵 \mathbf{X} 的每一列都可以看作一个信号 h , 因此基于光滑表示的训练数据 $\bar{\mathbf{X}}$ 可以通过式(8)来实现。

$$\bar{\mathbf{X}} = \left(I - \frac{L}{2} \right)^k \mathbf{X} \quad (8)$$

4 基于光滑表示的半监督分类算法

同一个类中的样本一般连接更加密集, 它们自然会有更加相似的特征表示。因此, 可以使用图过滤来获得更加光滑的训练数据表示形式。本节提出的 SRSSC 旨在对原始数据进行图滤波, 以获得更光滑的数据 $\bar{\mathbf{X}}$, 然后构造一个图构造和标签传播的统一化框架, 即将训练数据矩阵 \mathbf{X} 中构造图的部分和在图上进行标签传播的部分统一, 并进行联合优化。

4.1 获得光滑表示的数据 $\bar{\mathbf{X}}$

由于计算 $\bar{\mathbf{X}}$ 需要事先知道相似度矩阵 \mathbf{S} , 为了解决这个问题, 本文采用了一种基于连接性构建相似度矩阵的方法 (Clustering and Projected Clustering with Adaptive Neighbors, CAN)^[23], 该方法为每个数据点自适应地分配最佳邻居, 即通过式(9)计算相似度矩阵 \mathbf{S} 。

$$\begin{aligned} \min_{\mathbf{S}} \sum_{i,j=1}^N (\|x_i - x_j\|_2^2 s_{ij} + \gamma s_{ij}^2) \\ \text{s. t. } \forall i, s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1, \\ \text{rank}(L_S) = N - c \end{aligned} \quad (9)$$

其中, L_i 是 \mathbf{S} 对应的拉普拉斯矩阵, γ 可以由邻居数确定, 然后使用式(8)计算获得数据的光滑表示 $\bar{\mathbf{X}}$ 。

4.2 图构建和标签传播的统一化框架

得到平滑的数据后, 仍然还存在着两个重要的问题。第一, 如何基于训练数据构建一个高质量的图? 第二, 如何将图的构造和标签传播过程联合起来? 为了解决上述两个问题, 本节提出了一个统一化框架。

4.2.1 图构建

本节将在平滑表示的数据矩阵 $\bar{\mathbf{X}}$ 上建立一个图。 $\bar{\mathbf{X}}$ 中每个样本(即每行)对应于图中的一个节点, 若两节点之间的相似度很高(或者相关性很强), 则对应的节点之间存在一条边, 边的权重正比于样本之间的相似度(或相关性)。

为了方便起见, 此处定义 $\bar{\mathbf{X}} = \bar{\mathbf{X}}^T$ 。定义图的邻接矩阵为 $\mathbf{Z} = \{Z_{ij}\}$, 其中 Z_{ij} 表示样本 i 和样本 j 之间的相似度。估计两样本间相似度的一个典型方法为 $z_{ij} = \exp\left(-\frac{\|x_i - x_j\|_2^2}{\sigma}\right)$, $\sigma > 0$ 。

但是, 由于数据中通常会有很多噪声, 这种方式通常不能精确地反映样本间的关系。本文借鉴子空间聚类的思想来建立相似度矩阵 \mathbf{Z} ^[21], 其核心思想是数据来自多个子空间, 每个样本都可以用同一个子空间的样本的线性组合来表示。数学上, 通过以下形式求解 \mathbf{Z} :

$$\min_{\mathbf{Z}} \|\bar{\mathbf{X}} - \bar{\mathbf{X}}\mathbf{Z}\|_F^2 + \alpha \|\mathbf{Z}\|_F^2 \quad (10)$$

其中, 参数 $\alpha > 0$, 用于平衡两部分之间的影响。

4.2.2 标签传播

方程(10)自动地从数据中学到了一个图, 但我们不能保

证其对于接下来的分类来说是最优的。理想情况下, 如果数据中有 c 的类, 图 Z 应该恰好拥有 c 个连通成分。使用 σ_i 表示拉普拉斯矩阵 L 中第 i 个最小的特征值。由于 L 是一个半正定矩阵, 故 $\sigma_i > 0$ 。为了解决这个问题, 可以采用定理 1。

定理 1 图 2 连通成分的个数 c 等于其拉普拉斯矩阵 L 的零特征值出现的次数^[24]。

根据定理 1, 当 $\sum_{i=1}^c \sigma_i(L) = 0$ 时, 可以确定 L 的秩 $N - c$ 。根据 Ky Fan's Theorem^[20], 有:

$$\sum_{i=1}^c \sigma_i = \min_{\mathbf{F}, \mathbf{F}^T \mathbf{F} = I} \text{Tr}(\mathbf{F}^T L \mathbf{F}) \quad (11)$$

在半监督学习中, 矩阵 \mathbf{F} 可以被分解成 $\mathbf{F} = [\mathbf{F}_l; \mathbf{F}_u] = [\mathbf{Y}_l; \mathbf{F}_u]$ 。根据文献[11], 上述等式的右边部分其实就是半监督分类的标签传播目标函数。因此, 在标签传播过程中, 需要优化的目标函数如下:

$$\min_{\mathbf{F}, \mathbf{F}^T \mathbf{F} = I} \text{Tr}(\mathbf{F}^T L \mathbf{F}) \quad (12)$$

4.2.3 统一优化的目标函数

为了将图构建过程和标签传播过程相结合, 基于式(10)和式(12)可以构成一个统一的优化目标函数:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{F}} \|\bar{\mathbf{X}} - \bar{\mathbf{X}}\mathbf{Z}\|_F^2 + \alpha \|\mathbf{Z}\|_F^2 + \beta \text{Tr}(\mathbf{F}^T L \mathbf{F}) \\ \text{s. t. } \mathbf{F}_l = \mathbf{Y}_l \end{aligned} \quad (13)$$

可以观察到, 式(13)将同时优化图和标签矩阵, 从而有助于实现整体的全局最优解。

4.2.4 求解最优化问题

本节基于一种交替迭代的策略来求解问题(13), 即固定某一个变量的同时确定另一个变量。

(1)更新相似度矩阵

当固定标签指示矩阵 \mathbf{F} 时, 问题(13)可以逐列求解, 即转化为:

$$\min_{Z_{:,i}} \|\bar{\mathbf{X}}_{:,i} - \bar{\mathbf{X}}\mathbf{Z}_{:,i}\|_2^2 + \alpha Z_{:,i}^T Z_{:,i} + \frac{\beta^T d_i^T Z_{:,i}}{2} \quad (14)$$

其中, $d_{ij} = \|F_{i,:} - F_{j,:}\|_2^2$ 。因此, 式(14)可以变换为:

$$Z_{:,i} = (2\alpha I + 2\bar{\mathbf{X}}^T \bar{\mathbf{X}})^{-1} (2\bar{\mathbf{X}}^T \bar{\mathbf{X}}_{:,i} - \frac{\beta d_i}{2}) \quad (15)$$

(2)更新标签指示矩阵 \mathbf{F}

当固定 \mathbf{Z} 后, 问题(13)转换为:

$$\min_{\mathbf{F}} \text{Tr}(\mathbf{F}^T L \mathbf{F}) \quad \text{s. t. } \mathbf{F}_l = \mathbf{Y}_l \quad (16)$$

为了解以上问题, 可以将拉普拉斯矩阵 L 分块为 $L = \begin{bmatrix} L_{ll} & L_{lu} \\ L_{ul} & L_{uu} \end{bmatrix}$, 接着对矩阵 \mathbf{F} 进行求导, 并将其置为 0, 可以得到:

$$\begin{bmatrix} L_{ll} & L_{lu} \\ L_{ul} & L_{uu} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_l \\ \mathbf{F}_u \end{bmatrix} = 0 \quad (17)$$

式(17)给出^[13]:

$$\mathbf{F}_u = -L_{uu}^{-1} L_{ul} \mathbf{Y}_l \quad (18)$$

将上述过程迭代多次, 直至 \mathbf{F} 的变化程度小于阈值 ϵ 。

最后, 未标记数据点的标签可以通过以下决策函数得到:

$$y_i = \arg \max_j F_{ij}, \forall i = l+1, l+2, \dots, N, \forall j = 1, 2, \dots, c \quad (19)$$

完整的基于光滑表示的半监督分类算法如算法 1 所示。

算法 1 基于光滑表示的半监督分类算法

输入: 数据矩阵 \mathbf{X} , 标签矩阵 \mathbf{Y}_l

输出: Y

1. 通过式(9)计算 S
 2. 通过式(10)计算 \bar{X}
- REPEAT
- 步骤 1 和步骤 2
- Until $\|F_t - F_{t-1}\|_F^2 < \epsilon$

5 实验

5.1 数据集

为了评价本文算法的分类性能,本节在以下 5 个标准数据集上进行了分类实验,图 1 给出了一些示例图片。

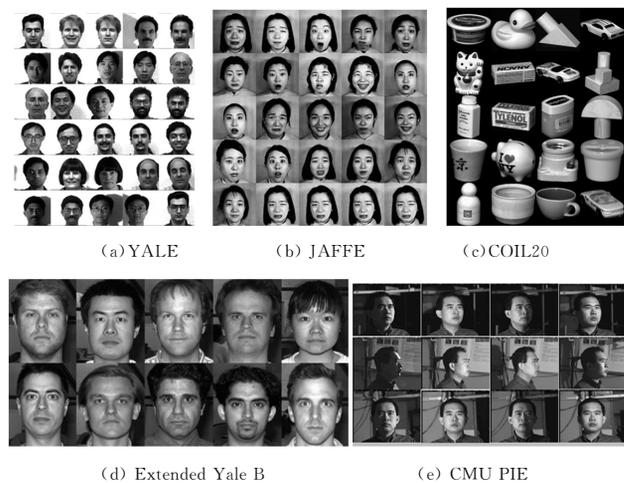


图 1 5 个数据集样本的示例

Fig. 1 Sample images of five datasets

(1)YALE 人脸数据库:由耶鲁大学计算视觉与控制中心创建,包含 15 位志愿者的 165 张图片,每个对象采集的样本包含 11 张有明显光照变化的近额图像。

(2)JEFFE 人脸数据集:包含 10 位日本女性志愿者的 213 张图片,每个对象的样本包含 7 种不同的面部表情。

(3)COIL20 图像数据集:由哥伦比亚大学图像库发布,包含 20 个物体在 360° 旋转中不同角度成像的图片,每个对象包含 72 种姿势。

(4)Extended Yale B (EYaleB) 人脸数据集:包含 38 名被试者的 2 414 张面部图像,每个对象约有 64 张面部图像。我们将所有样本的图像大小调整为 32×32 。

(5)CMU PIE^[25]:该数据集包含了 68 名受试者的 4 万多张面部图像。我们使用的数据集包括了 12 名对象,每名对象有 170 张来自不同照明和不同面部表情的图像,其中每张图像被调整为 32×32 的大小。

5.2 对比算法

本节将 SRSSC 与 4 种现有算法进行了比较。

(1)Learning with local and global consistency(LGC)^[17]:一种广泛使用的半监督分类算法。

(2)Gaussian Field and Harmonic Function(GFHF)^[13]:除了 LGC 外的另一种流行的标签传播算法。

(3)Self-Expression Graph-Based Semi Supervised Learning^[26]:提出了一种基于自表达的方法来构建半监督学习图。相似度矩阵和类别指示矩阵交替迭代更新,从而达到了互相学习和提高的目的。低秩和稀疏表示分别对应 S2LRR 和 S3R 模型。

(4)Semi-Supervised Classification With Adaptive Neighbors(SCAN)^[11]:一种基于图的方法,使用自适应邻近点的方法构造相似度矩阵,并将图构造和标签传播集合到一个框架中联合优化。

在上述 4 种方法中,LGC 与 GFHF 均以拉普拉斯矩阵为输入。为了获得更好的性能,我们基于 7 种核矩阵计算拉普拉斯矩阵,其中有 4 个形式分别为 $t \in \{0.1, 1, 10, 100\}$ 的高斯核,1 个形式为 $K(x, y) = x^y$ 的线性核,以及 2 个形式为 $K(x, y) = (a + x^T y)^2, a \in \{0, 1\}$ 的多项式核。另外,两种方法则直接从原始数据中构建图。

5.3 实验结果

为了测试数据是否光滑对结果的影响,我们将汇报 $k=0$ 和 $k>0$ 两种情况的结果, $k=0$ 即没有采用滤波。所有方法均选择 10%,30% 和 50% 有标签的数据,重复实验 20 次,将分类准确度的平均值和标准差记录于表 1 中。

表 1 各种方法在标准数据集上的分类精度(mean±standard deviation)

Table 1 Classification ACC of different methods on benchmark data sets (mean±standard deviation)

(单位:%)

Data	Labeled Percentage/%	GFHF	LGC	S ³ R	S ² LRR	SCAN	SRSSC	
							$k>0$	$k=0$
YALE	10	38.00±11.91	47.33±13.96	38.83±8.60	28.77±9.59	45.07±1.30	25.87±8.54	9.8±1.96
	30	54.13±9.47	63.08±2.20	58.25±4.25	42.58±5.93	60.92±4.03	45.13±4.80	20.50±4.79
	50	60.28±5.16	69.56±5.42	69.00±6.57	51.22±6.78	68.94±4.57	53.33±5.59	23.61±4.49
JAFFE	10	92.85±7.76	96.68±2.76	97.33±1.51	94.38±6.23	96.92±1.68	98.39±0.69	94.84±3.74
	30	98.50±1.01	98.86±1.14	99.25±0.81	98.82±1.05	98.20±1.22	99.51±0.51	99.87±0.27
	50	98.94±1.11	99.29±0.94	99.82±0.60	99.47±0.59	99.25±5.79	99.60±0.45	99.96±0.20
COIL20	10	87.74±2.26	85.43±1.40	93.57±1.59	81.10±1.69	90.09±1.15	93.71±1.26	89.38±6.14
	30	95.48±1.40	87.82±1.03	96.52±0.68	87.69±1.39	95.27±0.93	97.26±1.06	92.49±4.10
	50	98.62±0.71	88.47±0.45	97.87±0.10	90.92±1.19	97.53±0.82	98.74±0.18	94.75±1.95
EYaleB	10	64.74±7.74	66.72±3.45	72.86±9.24	66.12±5.92	72.86±9.24	73.31±5.72	56.00±1.46
	30	69.31±3.72	71.13±1.67	81.92±6.13	71.08±2.98	81.92±6.13	83.20±4.13	70.08±1.28
	50	76.08±1.28	79.32±1.80	89.10±3.54	78.48±1.96	88.71±4.34	88.87.±1.54	79.08±1.20
CMU PIE	10	75.89±6.16	59.49±6.34	56.64±4.42	73.27±7.13	69.23±5.10	75.98±4.92	73.12±3.20
	30	82.43±2.23	74.08±3.08	78.81±6.46	86.03±0.49	82.72±3.21	91.17±1.68	86.09±1.07
	50	87.34±1.08	80.08±1.20	80.15±3.98	92.15±0.96	87.53±1.84	93.83±0.12	91.70±0.49

注:最优的结果以加粗的方式表示

可以看出,在大部分情况下,使用更光滑的训练数据($k > 0$)比原始数据($k = 0$)的结果更好,这证明了数据光滑的重要性。与其他现有算法相比,本文提出的 SRSSC 算法在大部分情况下都优于其他算法,这证明了该算法的有效性。

与此同时,我们发现 JAFFE 数据在 30% 和 50% 标签率这两种情况下, $k = 0$ 的效果略优于光滑数据的结果。由于算法涉及求逆,因此其数值不稳定性可能带来小的偏差;另外一个可能的原因是 $k > 0$ 导致数据过光滑,失去了原有数据的判别能力,从而加大了分类难度。

图 2 给出了相似度矩阵 Z 在 YALE, JAFFE, COIL20 这 3 个数据集上的分布,可以发现相似度矩阵 Z 几乎可看作是块对角矩阵,这说明我们学到的图能够很好地反映样本间的关系。另外,我们可以发现 YALE 图上有很多噪声,而噪声会影响图的构建,即式(9)会产生一个不精确的图 S ,从而导致光滑表示不准确,进而降低本文算法的分类性能,使其比其他方法差。我们将在 5.6 节中提出一种优化方法。

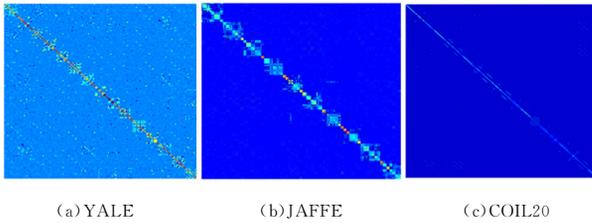


图 2 相似度矩阵 Z 在 3 个数据集上的可视化

Fig. 2 Visualization of learned similarity Graph Z on YALE, JAFFE, COIL20

5.4 消融实验

为了验证相比两者单独优化,将图构造和标签传播过程联合优化的结果更好,我们将 JAFFE 数据集中图构造与标签传播过程分开进行优化,并与联合优化的结果进行了对比,最后的结果如表 2 所列。由表 2 可以发现,联合优化的效果比单独优化的效果好很多,这验证了本文模型的有效性和正确性。

表 2 联合优化与单独优化在 JAFFE 数据集上的比较结果

Table 2 Performance Comparison of Separate and Joint Optimization on JAFFE (mean \pm standard deviation)

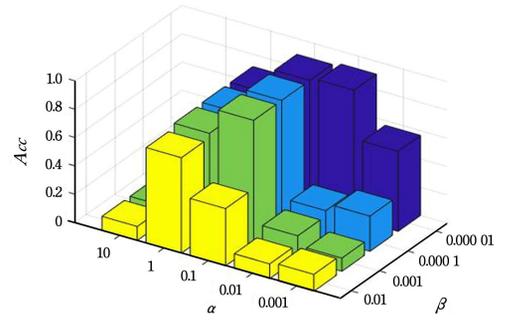
(单位: %)

Data	Labeled Percentage	Joint	Separate
JAFFE	10%	98.39 \pm 0.69	42.02 \pm 12.59
	30%	99.51 \pm 0.51	63.39 \pm 17.56
	50%	99.60 \pm 0.45	87.43 \pm 13.31

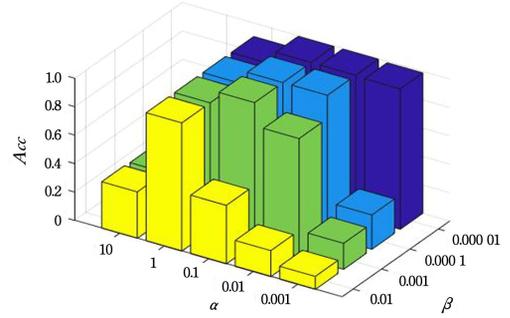
注:最优的结果以加粗的方式表示

5.5 参数敏感性

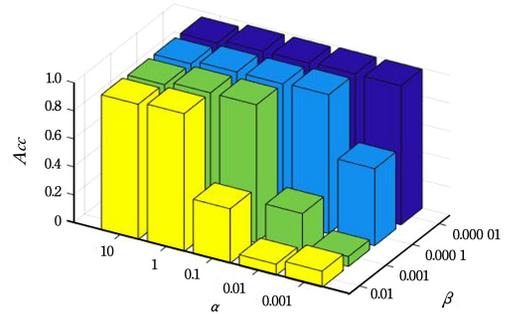
为了测试 SRSSC 算法对参数 α 和 β 的敏感性,我们以 JAFFE 数据集为例,图 3 给出了不同参数值下的精度。我们可以发现,在标记样本较少时, α 受到的影响较大。图 4 给出了在 YALE 数据集上的分类性能对参数 k 的敏感性。根据图 4, $k > 0$ 时,分类效果都有所提升,并且 k 在 5~10 范围内通常都能取得不错的结果,这再次说明了滤波的重要性。



(a) Labeled percentage (10%)



(b) Labeled percentage (30%)



(c) Labeled percentage (50%)

图 3 参数 α 和 β 在 JAFFE 数据集上的影响

Fig. 3 Influence of α and β on JAFFE data set

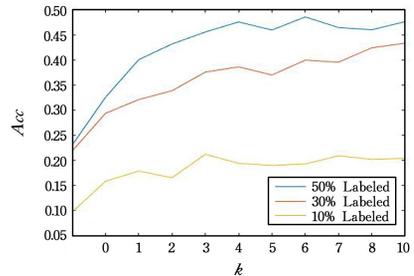


图 4 参数 k 在 YALE 数据集上的影响

Fig. 4 Influence neighbor number k on YALE data set

5.6 滤波优化

由表 1 可以看出,SRSSC 在 YALE 数据集上的分类情况相对较差,为了提高其性能,我们可以借鉴神经网络中的重整化技巧^[11]:将原本计算滤波的拉普拉斯矩阵 $L = I - D^{-\frac{1}{2}}SD^{-\frac{1}{2}}$ 变为 $L = I - D^{-\frac{1}{2}}(S + I)D^{-\frac{1}{2}}$,从而提高模型的稳定性。得到的优化结果如表 3 所列。我们可以看到,通过这一修正,SRSSC 在 YALE 上的性能得到了大大的提高。性能得到提升的原因在于该修正消除了数值的不稳定性,并缩小了谱图域^[27]。

表3 改进后的SRSSC方法在YALE上的实验结果
(mean±standard deviation)

Table 3 Optimized classification ACC on YALE (mean±standard deviation)

(单位:%)

Data	Labeled Percentage	Original	Optimized
YALE	10%	25.87±8.54	43.53±17.77
	30%	45.13±4.80	57.67±4.02
	50%	53.33±5.59	66.83±5.22

注:最优的结果以加粗的方式表示

结束语 本文提出了一种基于光滑表示的半监督分类算法,该算法通过将原始数据进行滤波,并将一个集成了图构造和标签传递的半监督分类框架应用在光滑表示的数据上。实验结果表明,经过滤波后更光滑的数据较原始数据所能达到的分类效果更好,这证明了数据的光滑性对于分类算法的重要性。本文提出的SRSSC算法在大部分情况下都优于其他分类算法,这也说明了所提方法的有效性。

参 考 文 献

- [1] ZHU X J. Semi-supervised learning literature survey[Z]. University of Wisconsin-Madison Department of Computer Sciences, 2005.
- [2] WANG F, LIU J C, WEI I. Semi-supervised Feature Selection Algorithm Based on Information Entropy[J]. Computer Science, 2019, 45(11A): 427-430.
- [3] ZHU C B, CHEN G, GAO Q. Research on Image Classification Algorithm Based on Semi-supervised Deep Belief Network[J]. Computer Science, 2018, 43(26): 46-50.
- [4] LIN J C, AI H J. Noise Tolerant Label Combination Semi-supervised Learning Algorithm[J]. Computer Engineering, 2019, 45(4): 157.
- [5] JEBARA T, WANG J, CHANG S. Graph construction and b-matching for semi-supervised learning[C]// Proceedings of the 26th Annual International Conference on Machine Learning, 2009: 441-448.
- [6] CHENG H, LIU Z, YANG J. Sparsity induced similarity measure for label propagation[C]// 2009 IEEE 12th International Conference on Computer Vision, 2009: 317-324.
- [7] LI S, FU Y. Learning balanced and unbalanced graphs via low-rank coding[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 27(5): 1274-1287.
- [8] WANG F, ZHANG C. Label propagation through linear neighborhoods[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 20(1): 55-67.
- [9] BO X, KANG Z, ZHAO Z, et al. Latent Multi-view Semi-Supervised Classification[C]// Proceedings of The Eleventh Asian Conference on Machine Learning, 2019: 348-362.
- [10] KANG Z, WEN L, CHEN W, et al. Low-rank kernel learning for graph-based clustering[J]. Knowledge-Based Systems, 2019, 163: 510-517.
- [11] NIE F, CAI G, LI X. Multi-view clustering and semi-supervised classification with adaptive neighbours[C]// Thirty-First AAAI Conference on Artificial Intelligence, 2017: 2408-2414.
- [12] KANG Z, GUO Z, HUANG S, et al. Multiple Partitions Aligned Clustering[C]// the 28th International Joint Conference on Artificial Intelligence(IJCAI-19), 2019: 2701-2707.
- [13] ZHU X, GHAHRAMANI Z, LAFFERTY J D. Semi-supervised

learning using gaussian fields and harmonic functions[C]// Proceedings of the 20th International Conference on Machine Learning (ICML-03), 2003: 912-919.

- [14] NIE F, WANG H, HUANG H, et al. Unsupervised and semi-supervised learning via ℓ_1 -norm graph[C]// 2011 International Conference on Computer Vision, 2011: 2268-2273.
- [15] GUN N, FANM Y, WANF D I, et al. Semi-supervised classification based on affine subspace sparse representation[J]. Entia Sinica, 2015, 45(8): 985-1000.
- [16] KANG Z, PAN H, HOI S C H, et al. Robust Graph Learning From Noisy Data[J]. IEEE Transactions on Cybernetics, 2020, 50(5): 1833-1843.
- [17] ZHOU D, BOUSQUET O, LAL T N, et al. Learning with local and global consistency[C]// Advances in Neural Information Processing Systems, 2004: 321-328.
- [18] KANG Z, LU X, LU Y, et al. Structure learning with similarity preserving[J]. Neural Networks, 2020, 129: 138-148.
- [19] SHUMAN D I, NARANG S K, FROSSARD P, et al. The emerging field of signal processing on graphs; Extending high-dimensional data analysis to networks and other irregular domains[J]. IEEE Signal Processing Magazine, 2013, 30(3): 83-98.
- [20] CHUNG F R, GRAHAM F C. Spectral graph theory [M]. American Mathematical Society, 1997.
- [21] LU C, MIN H, ZHAO Z, et al. Robust and efficient subspace segmentation via least squares regression[C]// European Conference on Computer Vision, 2012: 347-360.
- [22] ZHANG X, LIU H, LI Q, et al. Attributed Graph Clustering via Adaptive Graph Convolution[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 2019: 4327-4333.
- [23] NIE F, WANG X, HUANG H. Clustering and projected clustering with adaptive neighbors[C]// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014: 977-986.
- [24] MOHAR B, ALAVI Y, CHARTRAND G, et al. The Laplacian spectrum of graphs[J]. Graph theory, Combinatorics and Applications, 1991, 2(12): 871-898.
- [25] TERENCE SIM S, BAKER M B. The CMU Pose, Illumination, and Expression (PIE) database[C]// Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition, 2002: 53-58.
- [26] LI C, LIN Z, ZHANG H, et al. Learning semi-supervised representation towards a unified optimization framework for semi-supervised learning[C]// Proceedings of the IEEE International Conference on Computer Vision, 2015: 2767-2775.
- [27] KIPF T N, WELLLING M. Semi-Supervised Classification with Graph Convolutional Networks[C]// Proceedings of the 5th International Conference on Learning Representations, 2017.



WANG Xing, born in 2000, undergraduate. Her main research interests include machine learning, data mining and deep learning.



KANG Zhao, born in 1983, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include machine learning and pattern recognition.