



## 不确定成对约束的双对抗流形传播方法

孙乐, 王士同

引用本文:

孙乐,王士同. 不确定成对约束的双对抗流形传播方法[J]. 智能系统学报, 2023, 18(2): 270–281.

SUN Le,WANG Shitong. Doubly adversarial manifold propagation on uncertain pairwise constraints[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(2): 270–281.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202202025>

## 您可能感兴趣的其他文章

### 采用编码输入的生成对抗网络故障检测方法及应用

Fault detection method and its application using GAN with an encoded input

智能系统学报. 2022, 17(3): 496–505 <https://dx.doi.org/10.11992/tis.202102003>

### 基于二进制生成对抗网络的视觉回环检测研究

Visual loop closure detection based on binary generative adversarial network

智能系统学报. 2021, 16(4): 673–682 <https://dx.doi.org/10.11992/tis.202007007>

### 对抗样本三元组约束的度量学习算法

Metric learning algorithm with adversarial sample triples constraints

智能系统学报. 2021, 16(1): 30–37 <https://dx.doi.org/10.11992/tis.202009050>

### 生成对抗网络辅助学习的舰船目标精细识别

Fine-grained inshore ship recognition assisted by deep-learning generative adversarial networks

智能系统学报. 2020, 15(2): 296–301 <https://dx.doi.org/10.11992/tis.201901004>

### 基于知识距离的粗糙粒结构的评价模型

Evaluation model of rough granular structure based on knowledge distance

智能系统学报. 2020, 15(1): 166–174 <https://dx.doi.org/10.11992/tis.201904037>

### 基于混合距离学习的鲁棒的模糊C均值聚类算法

Robust FCM clustering algorithm based on hybrid-distance learning

智能系统学报. 2017, 12(4): 450–458 <https://dx.doi.org/10.11992/tis.201607019>

DOI: 10.11992/tis.202202025

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20221008.1134.008.html>

# 不确定成对约束的双对抗流形传播方法

孙乐, 王士同

(江南大学人工智能与计算机学院, 江苏 无锡 214122)

**摘要:** 成对约束传播 (pairwise constraint propagation, PCP) 通常研究的是在初始给定精确的成对约束基础上通过传播学习来增加成对约束的数量, 从而给机器学习任务提供较多的监督信息。可是, 在现实场景中, 有时还有一些不精确的成对约束, 因此, 如何利用这些不精确的成对约束来提高成对约束传播学习的效果是一个有待解决的问题。针对这一问题, 本文提出了一种不确定成对约束的传播方法。主要思想是用两个矩阵分别表示必须链接和不能链接的可能性, 两种可能性之间形成对抗, 同时两种成对约束之间也存在对抗关系, 两类对抗相结合形成一种双对抗结构, 作用于必须链接和不能链接的传播过程, 使二者的对抗强度在竞争中趋于最小化。我们将该方法称为不确定成对约束传播 (uncertain pairwise constraint propagation, UPCP)。在多个数据集上的实验结果表明, 不确定成对约束的传播效果不超过但近似于理想化传播效果, 在增强现实应用性的同时尽可能地保证了传播精度。

**关键词:** 不确定成对约束; 双对抗关系; 流形正则化; 成对约束传播; 约束可能性;  $k$  近邻; 可区分性; 弱监督  
**中图分类号:** TP181    **文献标志码:** A    **文章编号:** 1673-4785(2023)02-0270-12

中文引用格式: 孙乐, 王士同. 不确定成对约束的双对抗流形传播方法 [J]. 智能系统学报, 2023, 18(2): 270-281.

英文引用格式: SUN Le, WANG Shitong. Doubly adversarial manifold propagation on uncertain pairwise constraints[J]. CAAI transactions on intelligent systems, 2023, 18(2): 270-281.

## Doubly adversarial manifold propagation on uncertain pairwise constraints

SUN Le, WANG Shitong

(School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China)

**Abstract:** Pairwise constraint propagation aims to increase the number of accurate pairwise constraints through propagation learning on the basis of the initial given exact pairwise constraints to provide additional supervision information for the machine learning task. However, inaccurate pairwise constraints occasionally exist in real scenes. Therefore, one problem to be solved is the use of these inaccurate pairwise constraints to enhance pairwise constraint propagation. A propagation method of uncertain pairwise constraints is proposed in this paper to solve this problem. The main idea is to use two matrices to represent the possibility of must-links and cannot-links. A confrontation between the two possibilities and a confrontation relationship between two kinds of pairwise constraints exist. The combination of two types of confrontation forms a double adversarial structure, which acts on the propagation process of must-links and cannot-links; therefore, the adversarial intensity between them is minimized in the competition. This method is called uncertain pairwise constraint propagation. Experimental results on multiple datasets show that the propagation effect of uncertain pairwise constraints does not exceed but is similar to the ideal propagation effect, which enhances the practical applicability and ensures the propagation accuracy completely.

**Keywords:** uncertain pairwise constraint; doubly adversarial relationship; manifold regularization; pairwise constraint propagation; possibility of constraints;  $k$ -nearest neighbor; distinguishability; weak supervision

收稿日期: 2022-02-26. 网络出版日期: 2022-10-09.

基金项目: 国家自然科学基金项目 (61972181).

通信作者: 王士同. E-mail: [wxwangst@aliyun.com](mailto:wxwangst@aliyun.com).

成对约束被认为是一种弱监督信息<sup>[1]</sup>, 因为相比于标签信息, 成对约束不能提供显式的类信

息,而是一种综合的监督信息。具体来说,成对约束不是标记每一个数据样本属于哪一类,而是标记一对数据样本之间的关系,即由必须链接和不能链接两类约束信息来说明两个数据样本是否属于同一类。因此成对约束信息不具备推演传导的特性<sup>[2]</sup>,例如,给定3个数据点a、b、c,已知a和b之间存在必须链接,b和c之间存在必须链接,但a和c之间的约束关系并无法推导出。在很多机器学习任务中,成对约束比类标签使用起来更简单。

成对约束可以作为一种先验知识改进模型来提高聚类<sup>[3-7]</sup>、降维<sup>[8-10]</sup>、特征选择<sup>[11-12]</sup>、度量学习<sup>[13-17]</sup>和半监督分类<sup>[18]</sup>等任务的性能。但是可用的成对约束通常是有限的,因此成对约束的传播就至关重要。成对约束传播就是将有限可用的成对约束关系从已知约束的数据样本对传播到未知约束的数据样本对,从而获得大量可靠的成对约束信息,显著提高机器学习任务的性能。

为了将初始成对约束信息传播到整个数据集中,前人做了大量的工作,提出很多方法。例如,Lu等<sup>[19]</sup>通过高斯核函数将已知的约束传播到其他相似的未知约束的数据点,但是该方法做了特定假设,仅针对两个类的聚类问题。Li等<sup>[20]</sup>将成对约束传播表示为半定规划(semi-definite programming, SDP)过程,不局限于两个类的问题,但是SDP求解过程会产生极大的开销,计算复杂度达到 $O(n^4)$ 。Yu等<sup>[21]</sup>把成对约束传播表示为一个约束优化问题,但只局限于必须链接的优化。Lu等<sup>[22]</sup>为了克服上述问题,提出一种不限于两类和必须链接,并且能够降低计算复杂度的方法,将成对约束传播分解为一组相互独立的半监督分类子问题。为了增强传播中两种约束的可辨别性,Jia等<sup>[23]</sup>提出一种基于对抗流形正则化的成对约束传播模型,将必须链接和不能链接分开传播,并且加入其对抗关系。上述方法都是应用于单模态数据成对约束传播问题的,现实生活中多模态数据已经非常普遍,因此一些扩展方法也研究了多模态数据的成对约束传播问题<sup>[24-25]</sup>。

前面提到的很多成对约束传播模型都是利用精确的初始成对约束来传播,而在现实世界中,想要获取精确的成对约束存在一定的困难,很多约束关系来自现实环境,其具有弱关系,所以关系的不确定性是可变的<sup>[26]</sup>。很显然只追求精确的初始成对约束时并没有考虑到现实中获取成对约束的精准性,导致在现实情境下的传播会忽略一部分不精确信息。例如当初始成对约束信息是从用

户反馈中获得时,就存在很大的不确定性。假设现有两个病例的症状信息,不同个体出现的症状差异和表达能力的偏差等都会成为判断两个病例是否患同一种病的不确定因素,我们不能得出两个病例一定患或不患同一种病的结论,但可以判断二者患或不患同一种病的可能性。

为了解决这个问题,本文提出了不确定成对约束这个新概念,并在其基础上构造出一种双对抗的约束传播模型。这种传播通过已知信息获得两个数据样本属于或不属于同一类的可能性,两种可能性之间形成一个对抗,另外必须链接和不能链接本身存在一个对抗,综合两个对抗并使其作用于判断其他数据样本间的约束类别。不仅较大幅度地利用了有限数量初始成对约束的潜力,也增强了成对约束传播的现实应用性。

本文的主要贡献如下:1)关注到从现实场景中推断出成对约束关系时的准确性问题,首次提出不确定成对约束这个新的概念。2)在不确定成对约束的传播过程中,与文献[23]不同,在考虑了约束对抗的基础上,引入了可能性的对抗,形成双对抗的新结构,调整对抗强度。3)本文提出模型的交替迭代算法在理论上保证了收敛性。4)在8个数据集上的实验结果验证了不确定成对约束的双对抗传播效果的有效性,在现实场景的应用中有着明显的优越性。

## 1 相关知识

### 1.1 流形正则化

假设给定一个由 $n$ 个 $d$ 维数据样本组成的数据集 $\mathbf{X}$ ,将其第 $i$ 列记为 $\mathbf{x}_i$ ,那么输入数据表示为 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 。假定数据分布的几何形状由基础流形支持,可以将其映射为一个无向加权图 $G = \langle \nu, \varepsilon \rangle$ ,每一个数据样本代表一个顶点,所以 $\nu$ 是由 $\mathbf{X}$ 中所有数据样本组成的顶点集, $\varepsilon$ 是描述成对的数据样本之间存在约束情况的边集,边的权值大小通过计算相关的两个数据样本之间的相似度获得。 $k$ 近邻( $k$ -nearest neighbor,  $k$ -NN)图通常用来构造相似矩阵,其形成的稀疏图可以带来更好的性能。构造过程如下:首先基于欧氏距离找到与样本 $\mathbf{x}_i$ 最靠近的 $k$ 个样本,并记录为 $N_k(\mathbf{x}_i)$ <sup>[27]</sup>,两个顶点之间的边权通常使用径向基核函数(radial basis function, RBF)计算,生成一个位于区间 $[0, 1]$ 的真实值 $\mathbf{W}_{i,j}$ ( $1 \leq i, j \leq n$ ),表示为

$$\mathbf{W}_{i,j} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), & \mathbf{x}_j \in N_k(\mathbf{x}_i) \\ 0, & \text{其他} \end{cases} \quad (1)$$



其中 $\sigma > 0$ 是高斯核的宽度。因为 k-NN 图是对称的, 需要用公式 $\mathbf{W} = (\mathbf{W} + \mathbf{W}^T)/2$ 对初步获得的权值矩阵进行对称处理。

图的正则化将迫使嵌入的数据样本与临近点具有相似的性质, 从而保持原来的流形结构。已知上述 $\mathbf{W}_{i,j}$ 表示 $\mathbf{x}_i$ 和 $\mathbf{x}_j$ 两个数据样本之间在流形结构上的相似度, 给定 k-NN 图, 流形正则化表示为

$$\begin{aligned} & \sum_{i,j} \mathbf{W}_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \\ & \sum_{i,j} \mathbf{W}_{i,j} \mathbf{x}_i^T \mathbf{x}_i + \sum_{i,j} \mathbf{W}_{i,j} \mathbf{x}_j^T \mathbf{x}_j - 2 \sum_{i,j} \mathbf{W}_{i,j} \mathbf{x}_i^T \mathbf{x}_j = \\ & \sum_i \mathbf{D}_{i,i} \mathbf{x}_i^T \mathbf{x}_i + \sum_j \mathbf{D}_{j,j} \mathbf{x}_j^T \mathbf{x}_j - 2 \sum_{i,j} \mathbf{W}_{i,j} \mathbf{x}_i^T \mathbf{x}_j = \\ & 2 \sum_i \mathbf{D}_{i,i} \mathbf{x}_i^T \mathbf{x}_i - 2 \sum_{i,j} \mathbf{W}_{i,j} \mathbf{x}_i^T \mathbf{x}_j = \end{aligned} \quad (2)$$

$$2\text{tr}(\mathbf{X}\mathbf{D}\mathbf{X}^T) - 2\text{tr}(\mathbf{X}\mathbf{W}\mathbf{X}^T) = 2\text{tr}(\mathbf{X}\mathbf{L}\mathbf{X}^T)$$

式中:  $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$  是一个对角矩阵, 它的第  $i$  个对角元素计算为  $\mathbf{D}_{i,i} = \sum_{j=1}^n \mathbf{W}_{i,j}$ , 也就是相似度矩阵  $\mathbf{W}$  对应行的元素的和;  $\mathbf{L} = \mathbf{D} - \mathbf{W}$  是拉普拉斯矩阵;  $\text{tr}(\cdot)$  表示矩阵的迹。

## 1.2 对抗的 PCP

成对约束传播已经被证明是利用有限数量的成对约束的一种有效方法。对于给定的数据集, 已知其中一定数量成对的数据样本之间的关系, 即初始成对约束信息, 将必须链接保存在集合  $M$  中, 表示为

$$M = \{(\mathbf{x}_i, \mathbf{x}_j) | (\mathbf{x}_i, \mathbf{x}_j) \text{ 属于同类} \} \quad (3)$$

同理不能链接的集合  $C$  表示为

$$C = \{(\mathbf{x}_i, \mathbf{x}_j) | (\mathbf{x}_i, \mathbf{x}_j) \text{ 属于不同类} \} \quad (4)$$

成对约束提供的监督信息被广泛应用于许多机器学习任务中, 而在现实环境中, 我们可以获得的初始成对约束信息相当有限, 因此提出成对约束传播模型。具体来说, 就是在基础数据分布的支持下, 将初始成对约束关系传播到整个数据集中, 使任意一对数据样本之间都产生一个成对约束关系的预测值, 以增加成对约束的数量。

在已知的成对约束传播 (pairwise constraint propagation, PCP) 模型中, Jia 等<sup>[23]</sup> 提出的一种对抗流形正则化的成对约束传播方法, 是能够兼顾必须链接和不能链接传播的有效方法之一。它使用相似矩阵和不相似矩阵两个非负矩阵, 分别记录传播中的必须链接和不能链接, 采用流形正则化, 将必须链接和不能链接的传播分别表示为

$$\begin{aligned} & \min \text{tr}(\mathbf{S}\mathbf{L}\mathbf{S}^T) \\ & \text{s.t. } \mathbf{S} \geq 0, \mathbf{S}_{i,j} = \mathbf{M}_{i,j}(\mathbf{x}_i, \mathbf{x}_j) \in M \cup C \end{aligned} \quad (5)$$

$$\min \text{tr}(\mathbf{D}\mathbf{L}\mathbf{D}^T)$$

$$\text{s.t. } \mathbf{D} \geq 0, \mathbf{D}_{i,j} = \mathbf{C}_{i,j}, (\mathbf{x}_i, \mathbf{x}_j) \in M \cup C \quad (6)$$

此外, 加入必须链接和不能链接的对抗关系  $\|\mathbf{S} \odot \mathbf{D}\|_1$ , 将提出的模型总结为

$$\begin{aligned} & \min \|\mathbf{S} \odot \mathbf{D}\|_1 + \alpha \text{tr}(\mathbf{D}\mathbf{L}\mathbf{D}^T) + \beta \text{tr}(\mathbf{S}\mathbf{L}\mathbf{S}^T) \\ & \text{s.t. } \mathbf{D} \geq 0, \mathbf{S} \geq 0 \\ & \mathbf{D}_{i,j} = \mathbf{C}_{i,j}, \mathbf{S}_{i,j} = \mathbf{M}_{i,j}, (\mathbf{x}_i, \mathbf{x}_j) \in M \cup C \end{aligned} \quad (7)$$

式中:  $\mathbf{S}$  是必须链接的传播矩阵;  $\mathbf{D}$  是不能链接的传播矩阵;  $\mathbf{L}$  是拉普拉斯矩阵;  $\mathbf{M}$  和  $\mathbf{C}$  矩阵分别保存了初始必须链接和不能链接的信息, 用来保证  $\mathbf{S}$  和  $\mathbf{D}$  均满足初始的成对约束。

该模型利用了必须链接和不能链接之间的对抗关系, 从而获得了更加准确的传播结果。然而, 模型中所输入的初始成对约束信息都是以必须链接或不能链接理想化存在为前提, 即提供的初始约束信息仅包含 0 和 1 两个值, 但是现实场景中想要确定两个样本之间一定存在必须链接或不能链接并不容易, 这导致该 PCP 模型在实际中的应用能力并不理想。

## 2 不确定成对约束的双对抗流形传播模型

如前所述, 在现有的成对约束的对抗流形传播方法中, 关于精确初始约束信息的要求忽略了其在实际中的应用性, 只是在给定精确的初始成对约束的理想假设下得到较好的结果, 并未考虑到实际情况下可以获得的初始成对约束信息的精准性。为此, 我们提出了一种不确定成对约束的双对抗流形传播方法不确定成对约束传播 (uncertain pairwise constraint propagation, UPCP), 能够利用两个样本数据之间存在必须链接和不能链接其中一种的可能性来将成对约束信息泛化到整个数据集, 不仅可以充分挖掘其在实际应用中的潜力, 还能大大减少人工标注监督信息的成本。

### 2.1 UPCP 模型的建立

不同信息对于提高成对约束传播的准确性都至关重要, 必须链接和不能链接的传播都不可忽略, 因此, 给定一个由  $n$  个  $d$  维数据样本组成的输入数据矩阵, 首先根据已获得的初始成对约束信息, 将其理想化处理后分为两个矩阵  $\mathbf{A}$  和  $\mathbf{B}$ , 分别表示理想状态下为必须链接和不能链接的成对约束:

$$\mathbf{A}_{i,j} = \begin{cases} 1, & (\mathbf{x}_i, \mathbf{x}_j) \text{ 理想状态下为必须链接} \\ 0, & \text{其他} \end{cases} \quad (8)$$

$$\mathbf{B}_{i,j} = \begin{cases} 1, & (\mathbf{x}_i, \mathbf{x}_j) \text{ 理想状态下为不能链接} \\ 0, & \text{其他} \end{cases} \quad (9)$$

下面, 我们定义一个新的概念:

**定义 1** 当两个数据样本  $\mathbf{x}_i$  和  $\mathbf{x}_j$  在理想化状

态下存在必须链接(或不能链接)时, 根据客观经验, 判断这个必须链接(或不能链接)实际存在的可能性为  $p_{i,j}$ , 以此表示的成对约束称为不确定成对约束。

根据定义 1, 在用数学公式来表达不确定成对约束时, 定义相似可能性矩阵  $U$  和不相似可能性矩阵  $V$  来分别表示初始必须链接的可能性和初始不能链接的可能性:

$$U_{i,j} = \begin{cases} u_{i,j}, & A_{i,j} = 1 \\ 0, & \text{其他} \end{cases} \quad (10)$$

$$V_{i,j} = \begin{cases} v_{i,j}, & B_{i,j} = 1 \\ 0, & \text{其他} \end{cases} \quad (11)$$

其中  $u_{i,j}$  和  $v_{i,j}$  都是根据与现实场景相关的知识判断得到, 例如用户反馈等途径, 其数值范围在  $(0.5, 1]$ 。具体来说,  $u_{i,j}$  或  $v_{i,j}$  所在位置关联的两个数据样本  $x_i$  和  $x_j$  之间存在初始成对约束,  $u_{i,j}$  和  $v_{i,j}$  的大小表示成对约束存在的可能性。因为只有在可能性大于 0.5 时, 两个数据样本之间的约束类型才相对可靠, 故其取值设定在  $(0.5, 1]$ 。

为了记录传播过程中的必须链接和不能链接, 我们使用两个非负矩阵  $K$  和  $F$ , 在基于数据样本构建的  $k$ -NN 图指导下, 根据流形正则化, 将必须链接和不能链接的传播分别表示为

$$\begin{aligned} & \min \text{tr}(KLK^T) \\ & \text{s.t. } K \geq 0, K_{i,j} = A_{i,j}(x_i, x_j) \in M \cup C \end{aligned} \quad (12)$$

$$\begin{aligned} & \min \text{tr}(FLF^T) \\ & \text{s.t. } F \geq 0, F_{i,j} = B_{i,j}(x_i, x_j) \in M \cup C \end{aligned} \quad (13)$$

对于同一对样本, 其间存在必须链接或不能链接是相对的, 从而形成一种对抗关系, 数学表示为  $\|K \circ F\| = \sum_{i,j=1}^n K_{i,j} F_{i,j}$ , 其值的大小表示对抗的强弱。当同一对样本之间的  $K_{i,j}$  和  $F_{i,j}$  差异较大时,  $K_{i,j} F_{i,j}$  的值越小, 对抗越弱, 反之则对抗越强。

对应地, 定义两个非负矩阵  $\tilde{U}$  和  $\tilde{V}$  分别记录传播过程中的相似可能性和不相似可能性。当一对样本之间的相似可能性越大时, 其不相似可能性相应就越小, 所以相似可能性和不相似可能性之间也存在对抗关系, 同样使用 Hadamard 乘积可以将二者的对抗表示为  $\|\tilde{U} \circ \tilde{V}\|$ 。

图 1 是一对样本间双对抗结构的表达。  $x_i$  和  $x_j$  表示两个数据样本,  $K_{i,j}$  和  $F_{i,j}$  分别记录传播时两个样本间的必须链接和不能链接,  $\tilde{U}_{i,j}$  和  $\tilde{V}_{i,j}$  分别记录传播时两个样本间的相似可能性和不相似可能性。如标注所示, 上半部分的对抗流是两种约束的对抗, 下半部分则是两种约束可能性的对抗。在数学表达中这种双对抗关系可以有多种形式, 而  $\|\tilde{U} \circ \tilde{V}\| \|K \circ F\|$  就是其中的一种, 这种形式不

仅能够产生更真实的对抗强度, 而且便于交替优化求解, 可以有效调节成对约束的传播效果。

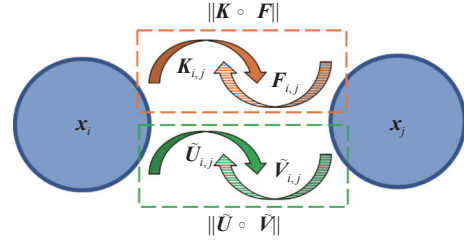


图 1 双对抗结构

Fig. 1 Doubly adversarial structure

将双对抗过程加入到成对约束的传播过程中, 并且使其最小化, 可使相近的成对约束具有相似的标记:

$$\begin{aligned} & \min \|\tilde{U} \circ \tilde{V}\| \|K \circ F\| + \alpha \text{tr}(KLK^T) + \beta \text{tr}(FLF^T) \\ & \text{s.t. } K \geq 0, F \geq 0, \tilde{U} \geq 0, \tilde{V} \geq 0 \\ & K_{i,j} = A_{i,j}, F_{i,j} = B_{i,j}, (x_i, x_j) \in M \cup C \end{aligned} \quad (14)$$

其中  $\alpha$  和  $\beta$  是两个用来控制必须链接和不能链接的传播强度的参数, 避免因为数据集潜在成对约束中的必须链接和不能链接的比例差距使传播失衡。

很明显, 在成对约束传播的过程中, 相似可能性和不相似可能性也在传播, 利用初始的两个可能性矩阵  $U$  和  $V$  中的信息, 可以通过添加  $\|\tilde{U} - U\|^2$  和  $\|\tilde{V} - V\|^2$  两项将可能性信息泛化到未知可能性的成对约束中, 这样可以将模型表述为

$$\begin{aligned} & \min \|\tilde{U} \circ \tilde{V}\| \|K \circ F\| + \alpha \text{tr}(KLK^T) + \beta \text{tr}(FLF^T) + \\ & \lambda_1 \|\tilde{U} - U\|^2 + \lambda_2 \|\tilde{V} - V\|^2 \\ & \text{s.t. } K \geq 0, F \geq 0, \tilde{U} \geq 0, \tilde{V} \geq 0 \\ & K_{i,j} = A_{i,j}, F_{i,j} = B_{i,j}, (x_i, x_j) \in M \cup C \end{aligned} \quad (15)$$

其中  $\lambda_1$  和  $\lambda_2$  是调节初始可能性信息贡献权重的参数。

## 2.2 算法的优化

为了有效地处理模型中的等式约束条件, 我们定义了一个矩阵  $P$  来协助模型的优化, 数学表达如下:

$$P_{i,j} = \begin{cases} 1, & (x_i, x_j) \in M \cup C \\ 0, & \text{其他} \end{cases} \quad (16)$$

矩阵  $P$  记录了初始成对约束的位置, 当  $P_{i,j}$  的值为 1 时, 表明样本  $x_i$  和  $x_j$  之间的成对约束已知, 否则就是未知。因为已知成对约束很少而未知成对约束很多, 为缓解过拟合, 利用  $P$  将两个等式约束表示成  $\|P \circ (K - A)\|^2$  和  $\|P \circ (F - B)\|^2$  的形式, 引入已知初始成对约束信息, 迫使预测情况与真实情况尽可能接近, 将模型优化为

$$\begin{aligned} & \min_{K, F, \tilde{U}, \tilde{V}} \|\tilde{U} \circ \tilde{V}\| \|K \circ F\| + \alpha \text{tr}(KLK^T) + \beta \text{tr}(FLF^T) + \\ & \gamma (\|P \circ (K - A)\|^2 + \|P \circ (F - B)\|^2) + \\ & \lambda_1 \|\tilde{U} - U\|^2 + \lambda_2 \|\tilde{V} - V\|^2 \\ & \text{s.t. } K \geq 0, F \geq 0, \tilde{U} \geq 0, \tilde{V} \geq 0 \end{aligned} \quad (17)$$

为求解传播结果矩阵,使用拉格朗日乘子法处理不等式约束条件,将式(17)的拉格朗日函数表示为

$$\begin{aligned} H(\mathbf{K}, \mathbf{F}, \tilde{\mathbf{U}}, \tilde{\mathbf{V}}) = & \|\tilde{\mathbf{U}} \circ \tilde{\mathbf{V}}\| \|\mathbf{K} \circ \mathbf{F}\| + \alpha \text{tr}(\mathbf{K} \mathbf{L} \mathbf{K}^T) + \\ & \beta \text{tr}(\mathbf{F} \mathbf{L} \mathbf{F}^T) + \gamma (\|\mathbf{P} \circ (\mathbf{K} - \mathbf{A})\|^2 + \\ & \|\mathbf{P} \circ (\mathbf{F} - \mathbf{B})\|^2) + \lambda_1 \|\tilde{\mathbf{U}} - \mathbf{U}\|^2 + \\ & \lambda_2 \|\tilde{\mathbf{V}} - \mathbf{V}\|^2 + \text{tr}(\Phi^T \mathbf{K}) + \text{tr}(\varphi^T \mathbf{F}) \end{aligned} \quad (18)$$

其中 $\Phi$ 和 $\varphi$ 是两个拉格朗日乘子矩阵。求 $H$ 关于 $\mathbf{K}$ 和 $\mathbf{F}$ 的一阶偏导,可得

$$\begin{aligned} \frac{\partial H}{\partial \mathbf{K}_{i,j}} = & (\|\tilde{\mathbf{U}} \circ \tilde{\mathbf{V}}\| \mathbf{F} + \alpha \mathbf{K}(\mathbf{L} + \mathbf{L}^T) + 2\gamma \mathbf{P} \circ (\mathbf{K} - \mathbf{A}) + \Phi)_{i,j} = \\ & (\|\tilde{\mathbf{U}} \circ \tilde{\mathbf{V}}\| \mathbf{F} + 2\alpha \mathbf{K} \mathbf{L} + 2\gamma \mathbf{P} \circ \mathbf{K} - 2\gamma \mathbf{P} \circ \mathbf{A} + \Phi)_{i,j}, \forall \{i, j\} \end{aligned} \quad (19)$$

$$\begin{aligned} \frac{\partial H}{\partial \mathbf{F}_{i,j}} = & (\|\tilde{\mathbf{U}} \circ \tilde{\mathbf{V}}\| \mathbf{K} + \beta \mathbf{F}(\mathbf{L} + \mathbf{L}^T) + 2\gamma \mathbf{P} \circ (\mathbf{F} - \mathbf{B}) + \varphi)_{i,j} = \\ & (\|\tilde{\mathbf{U}} \circ \tilde{\mathbf{V}}\| \mathbf{K} + 2\beta \mathbf{F} \mathbf{L} + 2\gamma \mathbf{P} \circ \mathbf{F} - 2\gamma \mathbf{P} \circ \mathbf{B} + \varphi)_{i,j}, \forall \{i, j\} \end{aligned} \quad (20)$$

强制 $\partial H / \partial \mathbf{K}_{i,j}$ 和 $\partial H / \partial \mathbf{F}_{i,j}$ 等于0,并且等式两边分别乘以 $\mathbf{K}_{i,j}$ 和 $\mathbf{F}_{i,j}$ 矩阵,可得等式:

$$\begin{aligned} (\|\tilde{\mathbf{U}} \circ \tilde{\mathbf{V}}\| \mathbf{F})_{i,j} \mathbf{K}_{i,j} + 2\alpha (\mathbf{K} \mathbf{L})_{i,j} \mathbf{K}_{i,j} + 2\gamma (\mathbf{P} \circ \mathbf{K})_{i,j} \mathbf{K}_{i,j} - \\ 2\gamma (\mathbf{P} \circ \mathbf{A})_{i,j} \mathbf{K}_{i,j} + \Phi_{i,j} \mathbf{K}_{i,j} = 0, \forall \{i, j\} \end{aligned} \quad (21)$$

$$\begin{aligned} (\|\tilde{\mathbf{U}} \circ \tilde{\mathbf{V}}\| \mathbf{K})_{i,j} \mathbf{F}_{i,j} + 2\beta (\mathbf{F} \mathbf{L})_{i,j} \mathbf{F}_{i,j} + 2\gamma (\mathbf{P} \circ \mathbf{F})_{i,j} \mathbf{F}_{i,j} - \\ 2\gamma (\mathbf{P} \circ \mathbf{B})_{i,j} \mathbf{F}_{i,j} + \varphi_{i,j} \mathbf{F}_{i,j} = 0, \forall \{i, j\} \end{aligned} \quad (22)$$

根据KKT(Karush-Kuhn-Tucker)条件可知 $\Phi_{i,j} \mathbf{K}_{i,j} = 0$ 和 $\varphi_{i,j} \mathbf{F}_{i,j} = 0$ ,并且依据关系 $\mathbf{L} = \mathbf{D} - \mathbf{W}$ 将 $\mathbf{L}$ 矩阵替换,可分别得到矩阵 $\mathbf{K}$ 和 $\mathbf{F}$ 的更新规则,如下:

$$\mathbf{K}_{i,j} = \mathbf{K}_{i,j} \frac{(2\gamma \mathbf{P} \circ \mathbf{A} + 2\alpha \mathbf{K} \mathbf{W})_{i,j}}{(\|\tilde{\mathbf{U}} \circ \tilde{\mathbf{V}}\| \mathbf{F} + 2\alpha \mathbf{K} \mathbf{D} + 2\gamma \mathbf{P} \circ \mathbf{K})_{i,j}} \quad (23)$$

$$\mathbf{F}_{i,j} = \mathbf{F}_{i,j} \frac{(2\gamma \mathbf{P} \circ \mathbf{B} + 2\beta \mathbf{F} \mathbf{W})_{i,j}}{(\|\tilde{\mathbf{U}} \circ \tilde{\mathbf{V}}\| \mathbf{K} + 2\beta \mathbf{F} \mathbf{D} + 2\gamma \mathbf{P} \circ \mathbf{F})_{i,j}} \quad (24)$$

传播后的 $\mathbf{K}$ 和 $\mathbf{F}$ 应该是对称矩阵,需要在每次传播后用公式 $(\mathbf{K} + \mathbf{K}^T)/2$ 和 $(\mathbf{F} + \mathbf{F}^T)/2$ 将其处理以保持对称性。

同样地,求 $H$ 关于 $\tilde{\mathbf{U}}$ 和 $\tilde{\mathbf{V}}$ 的一阶偏导,可得:

$$\frac{\partial H}{\partial \tilde{\mathbf{U}}_{i,j}} = (\tilde{\mathbf{V}} \|\mathbf{K} \circ \mathbf{F}\| + 2\lambda_1 \tilde{\mathbf{U}} - 2\lambda_1 \mathbf{U})_{i,j}, \forall \{i, j\} \quad (25)$$

$$\frac{\partial H}{\partial \tilde{\mathbf{V}}_{i,j}} = (\tilde{\mathbf{U}} \|\mathbf{K} \circ \mathbf{F}\| + 2\lambda_2 \tilde{\mathbf{V}} - 2\lambda_2 \mathbf{V})_{i,j}, \forall \{i, j\} \quad (26)$$

强制 $\partial H / \partial \tilde{\mathbf{U}}_{i,j}$ 和 $\partial H / \partial \tilde{\mathbf{V}}_{i,j}$ 等于0,可以求出相似可能性矩阵 $\tilde{\mathbf{U}}$ 和不相似可能性矩阵 $\tilde{\mathbf{V}}$ 的更新规则为

$$\tilde{\mathbf{U}}_{i,j} = \mathbf{U}_{i,j} - \frac{(\tilde{\mathbf{V}} \|\mathbf{K} \circ \mathbf{F}\|)_{i,j}}{2\lambda_1} \quad (27)$$

$$\tilde{\mathbf{V}}_{i,j} = \mathbf{V}_{i,j} - \frac{(\tilde{\mathbf{U}} \|\mathbf{K} \circ \mathbf{F}\|)_{i,j}}{2\lambda_2} \quad (28)$$

由于可能性矩阵是非负的对称矩阵,更新后的 $\tilde{\mathbf{U}}$ 和 $\tilde{\mathbf{V}}$ 需要归一化为0~1,并处理成对称矩阵。

优化算法的详细求解过程总结为如下算法:

**输入** 参数 $\alpha, \beta, \gamma, \lambda_1, \lambda_2$ , 数据集的k-NN图结构的边权矩阵 $\mathbf{W}$ , 初始成对约束的位置矩阵 $\mathbf{P}$ , 理想状态下的初始必须链接矩阵 $\mathbf{A}$ 和不能链接矩阵 $\mathbf{B}$ , 初始相似可能性矩阵 $\mathbf{U}$ 和不相似可能性矩阵 $\mathbf{V}$

初始化:  $\mathbf{K} = \text{rand}^+(n, n)$ ,  $\mathbf{F} = \text{rand}^+(n, n)$

$\tilde{\mathbf{U}} = \text{rand}^+(n, n)$ ,  $\tilde{\mathbf{V}} = \text{rand}^+(n, n)$

( $\text{rand}^+(n, n)$ 返回一个正的 $n \times n$ 的随机矩阵)

1) 计算对角矩阵 $\mathbf{D}$

计算拉普拉斯矩阵 $\mathbf{L} = \mathbf{D} - \mathbf{W}$ 。

2) 迭代更新 $\mathbf{K}$ 、 $\mathbf{F}$ 、 $\tilde{\mathbf{U}}$ 、 $\tilde{\mathbf{V}}$ 矩阵

While 不收敛 do

固定 $\mathbf{F}$ 、 $\tilde{\mathbf{U}}$ 、 $\tilde{\mathbf{V}}$ , 根据式(23)更新 $\mathbf{K}$ ;

固定 $\mathbf{K}$ 、 $\tilde{\mathbf{U}}$ 、 $\tilde{\mathbf{V}}$ , 根据式(24)更新 $\mathbf{F}$ ;

固定 $\mathbf{K}$ 、 $\mathbf{F}$ 、 $\tilde{\mathbf{V}}$ , 根据式(27)更新 $\tilde{\mathbf{U}}$ ;

将 $\tilde{\mathbf{U}}$ 归一化到[0, 1];

固定 $\mathbf{K}$ 、 $\mathbf{F}$ 、 $\tilde{\mathbf{U}}$ , 根据式(28)更新 $\tilde{\mathbf{V}}$ ;

将 $\tilde{\mathbf{V}}$ 归一化到[0, 1];

$(\mathbf{K} + \mathbf{K}^T)/2$ ;

$(\mathbf{F} + \mathbf{F}^T)/2$ ;

$(\tilde{\mathbf{U}} + \tilde{\mathbf{U}}^T)/2$ ;

$(\tilde{\mathbf{V}} + \tilde{\mathbf{V}}^T)/2$ ;

end while

**输出**  $\mathbf{K}$ 、 $\mathbf{F}$ 、 $\tilde{\mathbf{U}}$ 、 $\tilde{\mathbf{V}}$ 矩阵

### 2.3 收敛性分析

**定理1** 目标函数式(17)在算法下收敛。

定理1的证明可以等价地分解为5个命题:

**命题1** 式(17)中的目标函数在式(23)中的更新规则下递减,直到到达驻点;

**命题2** 式(17)中的目标函数在式(24)中的更新规则下递减,直到到达驻点;

**命题3** 式(17)中的目标函数在式(27)中的更新规则下递减,直到到达驻点;

**命题4** 式(17)中的目标函数在式(28)中的更新规则下递减,直到到达驻点;

**命题5** 式(17)是下界。

只要以上5个命题都成立,定理1即可成立。

首先,引入上界辅助函数的定义,辅助命题1的证明。

**定义2**<sup>[28]</sup> 当同时满足条件 $g(x, x') \geq f(x)$ ,  $g(x, x) = f(x)$ 时,  $g(x, x')$ 是 $f(x)$ 的一个上界辅助函数。

**引理1** 如果 $g$ 是 $f$ 的一个上界辅助函数,那么 $f$ 在按照 $x^{t+1} = \arg \min_i g(x, x')$ 更新时不递增。



引理 1 的证明:

$$f(x^{t+1}) \leq g(x^{t+1}, x') \leq g(x^t, x') = f(x^t) \quad (29)$$

当  $F$ 、 $\tilde{U}$ 、 $\tilde{V}$  固定时, 如果能够证明式 (23) 中的更新规则可以降低式 (17) 的上界函数, 则命题 1 得证。去掉与  $K$  无关的项, 将式 (17) 中的目标函数重新表示为

$$H_K = \|\tilde{U} \circ \tilde{V}\| \|K \circ F\| + \alpha \text{tr}(KLK^T) + \gamma \|P \circ (K - A)\|^2 \quad (30)$$

引理 2 式 (30) 关于  $K_{i,j}$  的上界辅助函数为

$$g(K_{i,j}, K_{i,j}^t) = H_K(K_{i,j}^t) + \left( \frac{\partial H_K}{\partial (K^t)} \right)_{i,j} (K - K^t)_{i,j} + \frac{(\|\tilde{U} \circ \tilde{V}\| F)_{i,j} + 2\alpha(K^t D)_{i,j} + 2\gamma(P \circ K^t)_{i,j}}{2K_{i,j}^t} (K - K^t)_{i,j}^2 \quad (31)$$

引理 2 的证明: 式 (30) 在点  $K_{i,j}^t$  处的泰勒级数表示为

$$H_{K_{i,j}} = H_K(K_{i,j}^t) + \left( \frac{\partial H_K}{\partial (K^t)} \right)_{i,j} (K - K^t)_{i,j} + \frac{1}{2} H_K''(K^t)_{i,j} (K - K^t)_{i,j}^2 \quad (32)$$

其中  $(\partial H_K / \partial (K^t))_{i,j}$  和  $H_K''(K^t)_{i,j}$  分别是  $H_K$  关于  $K_{i,j}$  的一阶导数和二阶导数, 具体表示为

$$\left( \frac{\partial H_K}{\partial (K^t)} \right)_{i,j} = (\|\tilde{U} \circ \tilde{V}\| F + 2\alpha K^t L + 2\gamma P \circ K^t - 2\gamma P \circ A)_{i,j} \quad (33)$$

$$H_K''(K^t)_{i,j} = 2\alpha L_{j,j} + 2\gamma P_{i,j} \quad (34)$$

若  $g(K_{i,j}, K_{i,j}^t) = H_{K_{i,j}}$ , 根据定义 1, 要证明  $g(K_{i,j}, K_{i,j}^t)$  为  $H_{K_{i,j}}$  的上界辅助函数, 就等价于证明  $g(K_{i,j}, K_{i,j}^t) \geq H_{K_{i,j}}$ , 代入式 (31) 和式 (32) 即为

$$\frac{(\|\tilde{U} \circ \tilde{V}\| F)_{i,j} + 2\alpha(K^t D)_{i,j} + 2\gamma(P \circ K^t)_{i,j}}{2K_{i,j}^t} (K - K^t)_{i,j}^2 \geq \frac{1}{2} (2\alpha L_{j,j} + 2\gamma P_{i,j}) (K - K^t)_{i,j}^2 \quad (35)$$

根据推导可知  $(K^t D)_{i,j} = \sum_{k=1}^n K_{i,k}^t D_{k,j} \geq K_{i,j}^t D_{j,j}$ , 又因为  $L = D - W$ , 可推出  $K_{i,j}^t D_{j,j} \geq K_{i,j}^t L_{j,j}$ , 综合可得  $(K^t D)_{i,j} \geq K_{i,j}^t L_{j,j}$ 。同时  $(\|\tilde{U} \circ \tilde{V}\| F)_{i,j} \geq 0$ , 由此可得式 (35) 成立。即引理 2 也成立。

由于式 (31) 是凸二次函数, 其全局最优值在点  $K_{i,j}^{t+1}$ , 其求解过程为

$$K_{i,j}^{t+1} = K_{i,j}^t - \frac{\left( \frac{\partial H_K}{\partial (K^t)} \right)_{i,j}}{2 \times \frac{(\|\tilde{U} \circ \tilde{V}\| F)_{i,j} + 2\alpha(K^t D)_{i,j} + 2\gamma(P \circ K^t)_{i,j}}{2K_{i,j}^t}} = \frac{K_{i,j}^t - K_{i,j}^t \frac{(\|\tilde{U} \circ \tilde{V}\| F + 2\alpha K^t D + 2\gamma P \circ K^t)_{i,j}}{(\|\tilde{U} \circ \tilde{V}\| F + 2\alpha K^t D + 2\gamma P \circ K^t)_{i,j}}}{1} = K_{i,j}^t$$

$$K_{i,j}^t \frac{(2\gamma P \circ A + 2\alpha K^t W)_{i,j}}{(\|\tilde{U} \circ \tilde{V}\| F + 2\alpha K^t D + 2\gamma P \circ K^t)_{i,j}} = K_{i,j}^t \frac{(2\gamma P \circ A + 2\alpha K^t W)_{i,j}}{(\|\tilde{U} \circ \tilde{V}\| F + 2\alpha K^t D + 2\gamma P \circ K^t)_{i,j}} \quad (36)$$

式 (36) 与必须链接矩阵  $K$  的更新规则式 (23) 相同, 所以可证明命题 1。

按照同样的流程, 式 (24) 中的更新规则可以在  $K$ 、 $\tilde{U}$ 、 $\tilde{V}$  固定时降低式 (17) 的上界函数, 由此命题 2 也可以证明。

对于命题 3, 当  $K$ 、 $F$ 、 $\tilde{V}$  固定时, 去掉与  $\tilde{U}$  无关的项, 式 (17) 中的目标函数可表示为

$$H_{\tilde{U}} = \|\tilde{U} \circ \tilde{V}\| \|K \circ F\| + \lambda_1 \|\tilde{U} - U\|^2 \quad (37)$$

观察可知, 式 (37) 是关于  $\tilde{U}$  的凸二次函数, 其全局最优值在  $\tilde{U}_{i,j} = U_{i,j} - \frac{(\tilde{V} \|K \circ F\|)_{i,j}}{2\lambda_1}$  处, 与更新规则式 (27) 相同, 所以可证明命题 3。同理, 命题 4 也可以证明。

最后, 因为已知  $K \geq 0$ 、 $F \geq 0$ 、 $\tilde{U} \geq 0$ 、 $\tilde{V} \geq 0$ , 很显然可以判断

$$\|\tilde{U} \circ \tilde{V}\| \|K \circ F\| + \alpha \text{tr}(KLK^T) + \beta \text{tr}(FLF^T) + \gamma (\|P \circ (K - A)\|^2 + \|P \circ (F - B)\|^2) + \lambda_1 \|\tilde{U} - U\|^2 + \lambda_2 \|\tilde{V} - V\|^2 \geq 0 \quad (38)$$

所以命题 5 得到了证明。

综上所述, 定理 1 成立, 即可证明算法的收敛性。

### 3 实验与结果

#### 3.1 数据集

为了展现 UPCP 方法传播效果的可靠性, 我们将其与没有加入可能性时精确的单对抗 PCP 方法传播效果进行比较。为了更好地评估传播性能, 实验在 8 个常用的 UCI 数据集上进行, 其都来源于实际的应用问题。如表 1 所示, 表中介绍了实验所用数据集的样本数、特征维度和类别数。其中, Chart 数据集是用于测试时间序列数据库中的索引方案。Dermatology 数据集是确定红细胞鳞状病变的类型。Glass 数据集是根据玻璃的化学成分来判断玻璃类型、确定玻璃用途, 多用于犯罪证据的取证。Ionosphere 数据集包含了雷达站的信号数据。Libras 数据集是对 4 个天秤座的人在视频中的手部运动路径的映射。Mnist 数据集是从 70 000 个初始样本中按照每类样本数相同的原则随机选取 500 个进行实验。Wine 数据集是关于红酒的化学成分分析的数据。Zoo 数据集包含了用于动物分类或聚类分析的一些动物

特征,例如是否有毛发、是否有腿等布尔值。

表 1 用于实验的数据集的特性

Table 1 Characteristics of the datasets used for evaluation

| 数据集         | 样本数 | 特征维度 | 类别数 |
|-------------|-----|------|-----|
| Chart       | 600 | 60   | 6   |
| Dermatology | 366 | 34   | 6   |
| Glass       | 214 | 9    | 6   |
| Ionosphere  | 351 | 34   | 2   |
| Libras      | 360 | 90   | 15  |
| Mnist       | 500 | 784  | 10  |
| Wine        | 178 | 13   | 3   |
| Zoo         | 101 | 16   | 7   |

### 3.2 参数设置

对于实验中用到的参数,我们进行了如下设置:首先在正则化部分涉及的 k-NN 图结构,每个数据集选取  $k$  值的大小由公式  $k = \lfloor \log_2 n \rfloor + 1$  来确定<sup>[23]</sup>,其中  $n$  是数据集的样本数量, $\lfloor \cdot \rfloor$  表示取对数结果的下整数。在用 RBF 核函数计算已知 k-NN 图的边权值时,依据斯科特法则设置高斯核的宽度  $\sigma$ <sup>[29]</sup>,公式为  $\sigma = \sqrt{5sn^{-1/(d+4)}}$ ,其中  $s$  为数据集的标准差, $n$  为数据集的样本数量, $d$  为数据集的特征维度。另外,为了避免求解过程中出现自增强效应,权值矩阵主对角线上的元素都设定为  $W_{ii} = 0$ 。两个约束正则化项的参数  $\alpha$  和  $\beta$  在  $\{0.1, 0.3, 1, 3, 10\}$  中选择,可能性矩阵泛化项的参数  $\lambda_1$  和  $\lambda_2$  从  $\{2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0\}$  中选择,通过穷举搜索,约束矩阵的泛化项参数  $\gamma$  设置为 10。

### 3.3 评价指标

成对约束传播追求的是模型的预测性能。实验中采用  $F_1$  值、精确率 (precision) 和召回率 (recall) 作为评价指标。以必须链接为例,precision 表示的是在预测为必须链接的成对约束中,有多少是真正的必须链接,追求 precision 高,目的在“精”,即希望预测的必须链接中为真的比例越高越好,即使还有很多必须链接没有预测到。recall 则指的是真正的必须链接有多少被预测准确了,追求 recall 高,目的在“全”,即希望所有必须链接都被预测到,即使同时有很多约束被预测错误。仅追求 precision 或 recall 某一个高,可能会使预测结果走向极端,而  $F_1$  值是二者的调和均值,计算公式为  $F_1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ ,它不仅能够兼顾“精”和“全”,又可以克服准确率 (accuracy) 不能解决的两种链接数量不平衡的问题,因此将  $F_1$  值

作为最重要的衡量指标。

为了方便计算  $F_1$  值、precision 和 recall,对于传播后的约束矩阵  $K$  和  $F$ ,定义两个结果矩阵  $Q = K - F$  和  $R = F - K$  分别来辅助必须链接和不能链接的结果统计,对于每个矩阵,采用不同的阈值离散,以  $Q$  为例,离散方法为

$$\hat{Q}_{i,j} = \begin{cases} 1, & Q_{i,j} > \text{TH} \\ -1, & Q_{i,j} < -\text{TH} \\ 0, & \text{其他} \end{cases} \quad (39)$$

其中,TH 表示阈值大小,变化范围在 0.01~0.09,步长为 0.01。将离散后的矩阵与 ground-truth 值比较,计算出 99 种情况下的  $F_1$  值、precision 和 recall,并且取  $F_1$  值最大时的一组数据作为必须链接(或不能链接)最终的评价值。

由于每个数据集中潜在的必须链接和不能链接的比例不同,所以在综合二者的  $F_1$  值时需要调节两个  $F_1$  值的比重,数学表示为

$$F_1 = \mu F_{1_m} + (1 - \mu) F_{1_c} \quad (40)$$

式中: $F_{1_m}$  和  $F_{1_c}$  分别表示必须链接和不能链接的  $F_1$  值, $\mu$  为所有潜在不能链接占有潜在成对约束的比值。precision 和 recall 也采用同样的方法计算。

### 3.4 实验结果

对每一个数据集,首先从 ground-truth 值中随机选取 1000 个初始成对约束,作为精确信息,并且在其他条件改变时这 1000 个成对约束固定不变。为了减少随机抽样的影响,随机选择初始成对约束 20 次,并计算这 20 次的平均性能,作为实验结果。在模拟现实场景中获得的不确定成对约束时,我们将初始可能性矩阵  $U$  和  $V$  进行人工设定,实验设计为两组。

第 1 组是从已选的 1000 个精确初始成对约束中随机选取 50% 为其设定可能性值,这部分成对约束的可能性大小在  $\{1, 0.8, 0.6, 0\}$  中变化,观察在不确定成对约束的数量相同时可能性大小对成对约束传播效果的影响。当可能性取 1 时,实验就简化为这 1000 个精确的初始成对约束的单对抗传播,如文献 [23] 中的方法,同样取 0 时则表示剩下的 500 个精确的初始成对约束的单对抗传播。实验结果如图 2 所示。

图 2 直观地展现了在各个数据集上可能性大小对成对约束传播效果的影响。在取 50% 初始成对约束加入可能性时,不确定成对约束双对抗传播所得到的  $F_1$  值虽然略低于 1000 个精确的初始成对约束单对抗传播的水平,但是通常明显高于去掉这 50% 初始成对约束的 500 个精确的初始成对约束单对抗传播的  $F_1$  值,在 Ionosphere、Lib-



ras、Mnist、Wine 和 Zoo 数据集中尤为明显。当赋给这 50% 的可能性不同时,随着可能性的增大,传播结果的  $F_1$  值会升高,并且在超过一半的数据集中呈现较明显的逐步升高趋势。但由于赋值可能性的部分比重较大且为随机,可能会出现一些

成对约束的赋值与实际有差距的情况,影响传播效果,导致可能性为 0.6~0.8 时的  $F_1$  值升高不明显。综合来看,UPCP 中的可能性信息在成对约束传播时发挥了一定的作用,并且可能性的提升有利于提高传播性能。

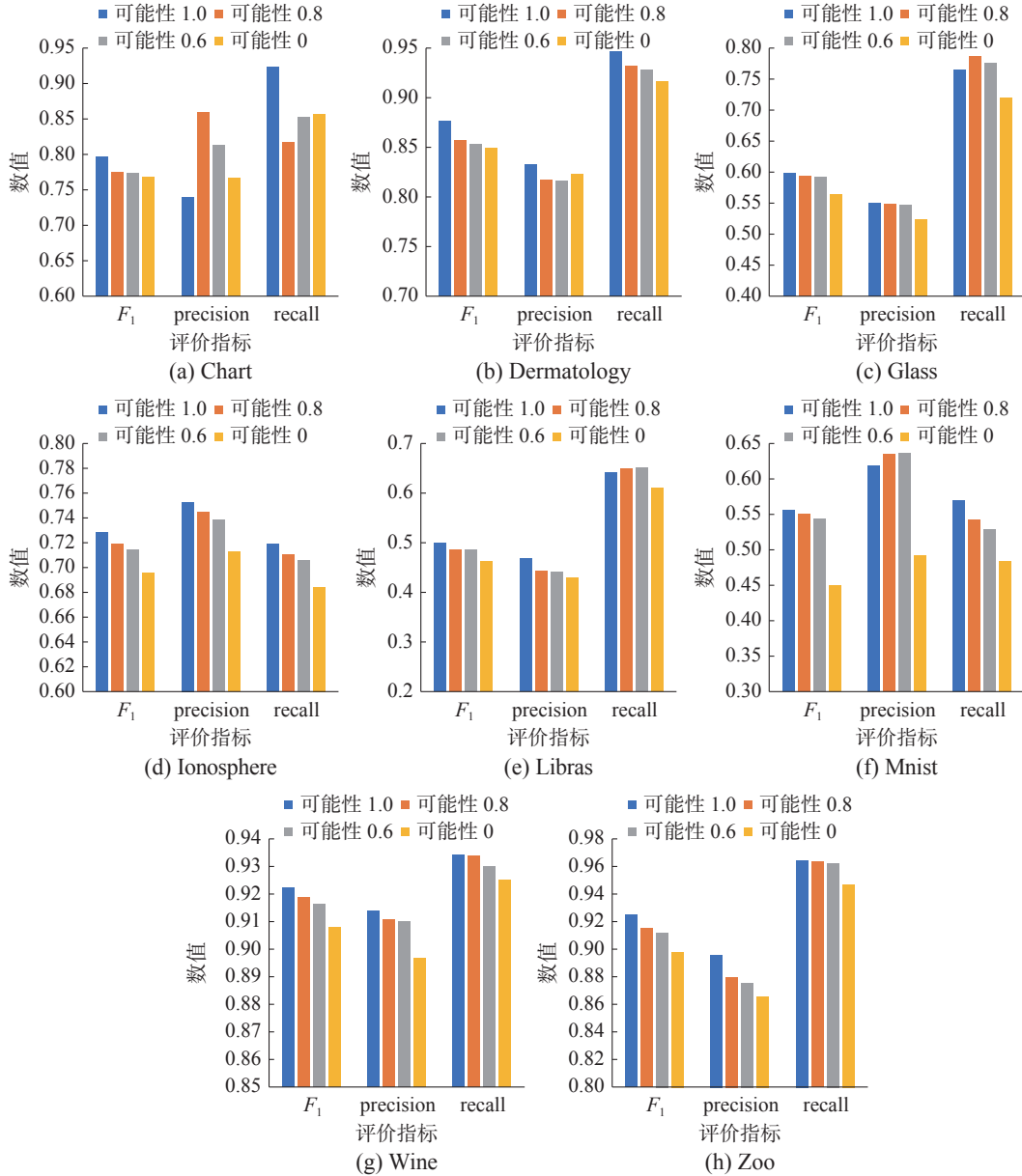


图2 可能性大小不同时  $F_1$ 、precision 和 recall 的值

Fig. 2  $F_1$ , precision and recall against probability

另外,实验还选取了两个先进的无对抗的 PCP 模型进行比较,分别为穷举且有效的约束传播模型 (exhaustive and efficient constraint propagation, E2CP)<sup>[22]</sup> 和对称图正则化约束传播模型 (symmetric graph regularized constraint propagation, SRCP)<sup>[30]</sup>,这两个模型分别选取 500 和 1000 个初始成对约束进行测试,并与上述可能性大小为 0.6 和 0.8 时的实验结果作比较。E<sup>2</sup>CP 和 SRCP 都是基于图的

方法,其涉及的图构造过程均与 UPCP 的实验设置相同,两个模型中的  $\alpha$  参数均在 {0.2, 0.4, 0.6, 0.8} 中选择。结果矩阵统计方法也与上述实验统一,所得的  $F_1$  值如表 2 所示,最优  $F_1$  值加粗标注。同时,讨论 UPCP 模型自身的鲁棒性,研究在大量给初始成对约束赋可能性值时,随机选取不同的成对约束对传播结果的影响。表 2 中在 UPCP 模型的  $F_1$  值后添加了随机 20 次实验的标准差。

表 2 中,UPCP 在可能性为 0.6 和 0.8 时的传播性能都明显优于  $E^2CP$  和 SRCP 在初始成对约束为 500 时的传播性能,而当  $E^2CP$  和 SRCP 初始成对约束为 1000 时,除 Chart 和 Dermatology 数据集外,UPCP 在可能性为 0.6 和 0.8 时的  $F_1$  值也高

于  $E^2CP$  和 SRCP 模型。由此可见,相对于传统的 PCP 方法,UPCP 在精确的初始成对约束较少的情况下,也能表现出较好的传播性能。另外,UPCP 在初始信息大量不确定时的标准差也都比较理想,表现出较好的鲁棒性。

表 2 不同模型上成对约束传播性能的比较 ( $F_1$  值)

Table 2 Comparison of pairwise constraint propagation performance on different models ( $F_1$  value)

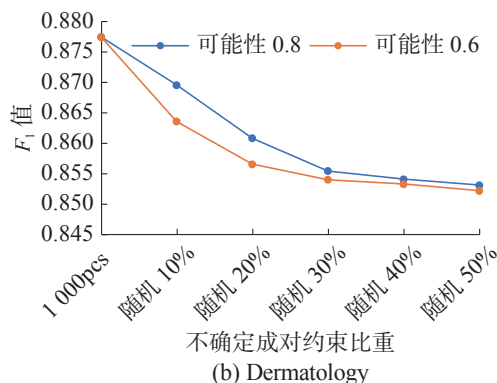
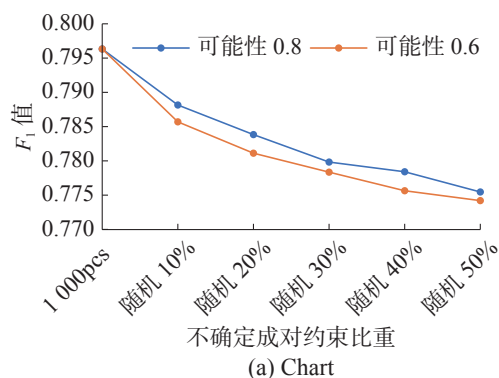
| 数据集         | SRCP   | SRCP    | $E^2CP$ | $E^2CP$       | UPCP               | UPCP                  |
|-------------|--------|---------|---------|---------------|--------------------|-----------------------|
|             | 500pcs | 1000pcs | 500pcs  | 1000pcs       | 1000pcs (50%, 0.6) | 1000pcs (50%, 0.8)    |
| Chart       | 0.3783 | 0.5854  | 0.7181  | <b>0.8047</b> | 0.7742±0.0044      | 0.7755±0.0036         |
| Dermatology | 0.6916 | 0.7597  | 0.8332  | <b>0.8755</b> | 0.8539±0.0021      | 0.8570±0.0012         |
| Glass       | 0.4733 | 0.4815  | 0.4966  | 0.5178        | 0.5919±0.0013      | <b>0.5937</b> ±0.0006 |
| Ionosphere  | 0.6362 | 0.6621  | 0.6814  | 0.7118        | 0.7137±0.0058      | <b>0.7183</b> ±0.0029 |
| Libras      | 0.3058 | 0.4417  | 0.4068  | 0.4895        | 0.5145±0.0026      | <b>0.5154</b> ±0.0015 |
| Mnist       | 0.2646 | 0.4372  | 0.3938  | 0.4487        | 0.5435±0.0040      | <b>0.5507</b> ±0.0034 |
| Wine        | 0.6125 | 0.7495  | 0.6932  | 0.8128        | 0.9167±0.0011      | <b>0.9192</b> ±0.0007 |
| Zoo         | 0.6718 | 0.7584  | 0.8225  | 0.8904        | 0.9121±0.0017      | <b>0.9149</b> ±0.0018 |

第 2 组实验是从 1000 个初始成对约束中,分别随机选取 10%、20%、30%、40% 和 50% 加入可能性,设置选到的必须链接对应的  $U_{i,j}$  和不能链接对应的  $V_{i,j}$  大小在 {0.6,0.8} 中变化,并与初始成对约束数量为 1000 时的精确传播作比较,观察在可能性大小相同时不确定成对约束数量对成对约束传播效果的影响。

从图 3 可以看到,不确定成对约束的数量也是影响 UPCP 传播性能的重要因素。以  $F_1$  值为观测目标,在每个数据集中,随着不确定成对约束比重的增加,在两种可能性下传播结果的  $F_1$  值都逐渐下降,且下降的趋势大致相同。可能性为 0.8 时的  $F_1$  值不会大跨度地下降,总体上,在大部分数据集中,下降的速率存在减缓的趋势。而可能性为 0.6 时的  $F_1$  值会比可能性为 0.8 时下降得

大,同样下降速率也有所减缓。但因为数据集中必须链接和不能链接的不平衡,导致对不确定信息所占比重的敏感度存在差异,下降速率也存在一些波动。由此可以得出结论,当不确定成对约束数量越少时,UPCP 的传播效果越好,即便在不确定成对约束数量较多的情况下,UPCP 的传播性能也并不会急剧下降,而是尽可能保证传播结果的精准。另外,在 8 个数据集中,可能性为 0.8 时的  $F_1$  值总是高于可能性为 0.6 时的  $F_1$  值,符合第 1 组实验的结论。

综合以上两组实验,相比于精确的成对约束传播方法的适用范围,UPCP 还可用于精确的初始成对约束较少,但存在一定量的不确定初始成对约束的现实场景中,充分挖掘不确定成对约束的潜力,表现出较好传播效果。



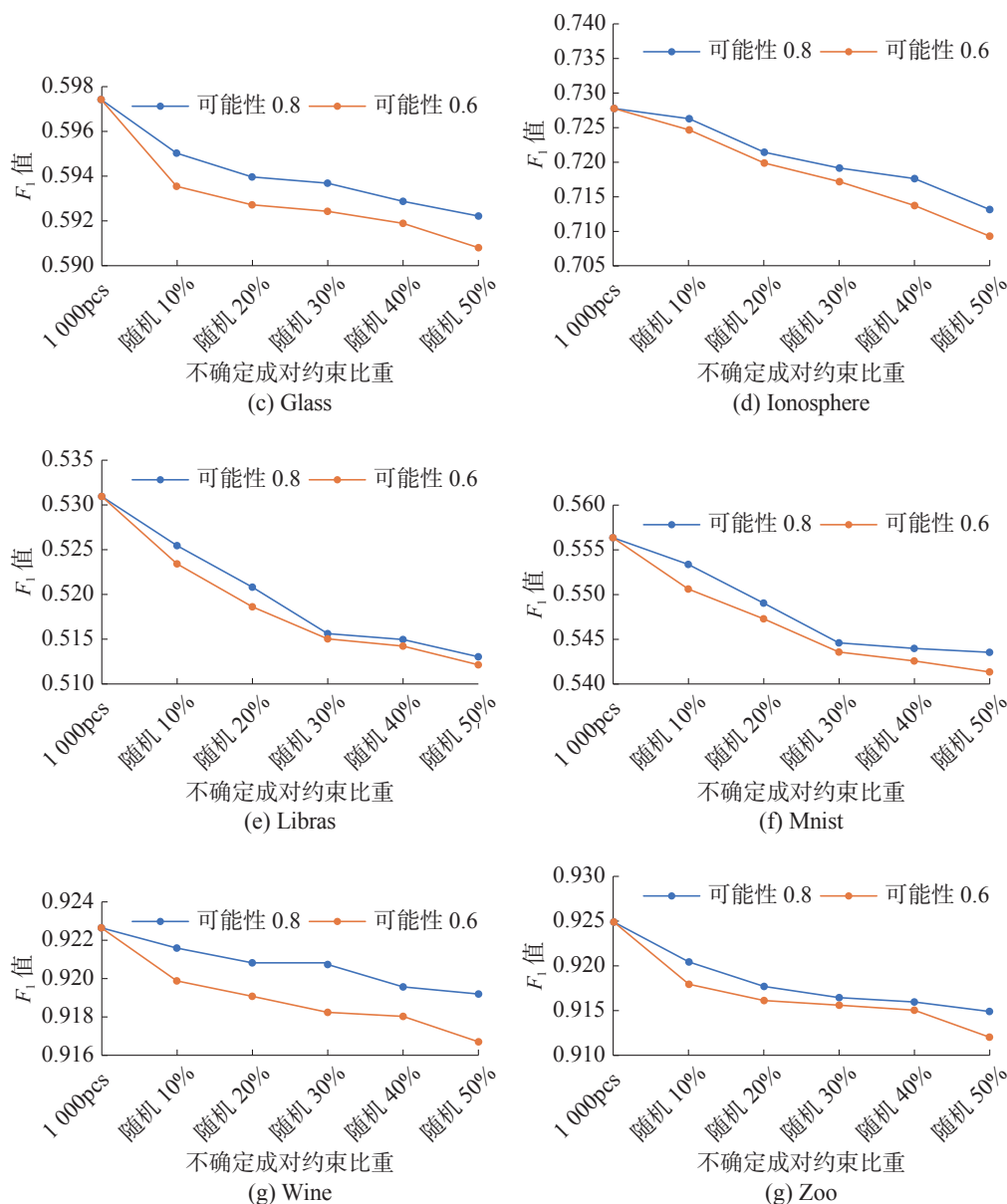


图 3 加入可能性比重不同时 F1 值的变化

Fig. 3 Change of F1 value when the proportion of possibility is different

## 4 结束语

本文提出了一种不确定成对约束的双对抗流形传播模型 UPCP, 来解决现有 PCP 模型中仅使用精确的初始必须链接和不能链接进行传播的问题。在已被提出的 PCP 方法中, 通常需要给出一定量的精确的初始成对约束, 在这种理想条件下进行传播。考虑到初始成对约束获取过程中受各种客观因素的影响, 可获得的精确约束信息很少, 我们定义了不确定成对约束这个新概念来表示不精确的初始成对约束。根据现实场景中相关知识的判断, 将初始必须链接和不能链接存在的可能性分别保存在两个矩阵中, 并让两个可能性矩阵之间形成对抗, 同时结合必须链接和不能链接之

间的对抗, 通过双对抗结构调节对抗强度进行传播, 从而有效利用不确定成对约束。实验结果表明, 本文提出的方法能够接近于等量初始约束信息情况下的精确传播的效果。接下来, 将关注提高不确定成对约束传播的精确度并将其应用于机器学习任务中。

## 参考文献:

- [1] FU Z Y, LU Z W. Pairwise constraint propagation: a survey[J]. Computer science, 2015, 1(3): 161–164.
  - [2] 付振勇. 成对约束传递方法及其在约束聚类问题中的应用研究 [D]. 上海: 上海交通大学, 2012.
- FU Zhenyong. Studies on pairwise constraint propagation and its applications in the constrained clustering



- tasks[D]. Shanghai: Shanghai Jiao Tong University, 2012.
- [3] WAGASTAFF K, CARDIE C, ROGERS S, et al. Constrained k-means clustering with background knowledge[C]// Eighteenth International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc, 2001: 577–584.
  - [4] KLEIN D. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering[J]. *Proceedings of the Nineteenth International Conference on Machine Learning*, 2002, 12(7): 307–314.
  - [5] BASU S, BILENKO M, MOONEY R J. A probabilistic framework for semi-supervised clustering[C]// *Proceedings of the 10th Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) International Conference on Knowledge Discovery and Data Mining*. Seattle: Dept. of Computer Sciences University of Texas at Austin, 2004: 59–68.
  - [6] KULIS B, BASU S, DHILLON I, et al. Semi-supervised graph clustering: a kernel approach[J]. *Machine learning*, 2009, 74(1): 1–22.
  - [7] 卞则康, 王士同. 基于混合距离学习的鲁棒的模糊 C 均值聚类算法 [J]. *智能系统学报*, 2017, 12(4): 450–458.  
BIAN Zekang, WANG Shitong. Robust FCM clustering algorithm based on hybrid-distance learning[J]. *CAAI transactions on intelligent systems*, 2017, 12(4): 450–458.
  - [8] WANG D, GAO X B, WANG X M. Semi-supervised nonnegative matrix factorization via constraint propagation[J]. *IEEE transactions on cybernetics*, 2016, 46(1): 233–244.
  - [9] CHEN S G, ZHANG D Q. Semi-supervised dimensionality reduction with pairwise constraints for hyperspectral image classification[J]. *IEEE geoscience and remote sensing letters*, 2011, 8(2): 369–373.
  - [10] 钱进, 罗鼎. 基于成对约束半监督降维的高光谱遥感影像特征提取 [J]. *遥感技术与应用*, 2014, 29(4): 681–688.  
QIAN Jin, LUO Ding. Feature extraction from hyperspectral remote sensing imagery based on semisupervised dimensionality reduction with pairwise constraints[J]. *Remote sensing technology and application*, 2014, 29(4): 681–688.
  - [11] LIU Mingxia, ZHANG Daoqiang. Pairwise constraint-guided sparse learning for feature selection[J]. *IEEE transactions on cybernetics*, 2017, 46(2): 298–310.
  - [12] 赵帅, 张雁, 徐海峰. 基于成对约束分的特征选择及稳定性评价 [J]. *计算机与数字工程*, 2019, 47(6): 1441–1445.  
ZHAO Shuai, ZHANG Yan, XU Haifeng. Feature selection and stability evaluation based on pairwise constraint score[J]. *Computer & digital engineering*, 2019, 47(6): 1441–1445.
  - [13] XING E P, NG A Y, JORDAN M I, et al. Distance metric learning with application to clustering with side-information[C]// *Advances in Neural Information Processing Systems*. Cambridge, Boston: Massachusetts Institute of Technology Press, 2002: 318–326.
  - [14] HOI S C H, LIU Wei, LYU M R, et al. Learning distance metrics with contextual constraints for image retrieval [C]// *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York, USA: 2072–2078.
  - [15] LIU Wei, MA Shiqian, TAO Dacheng, et al. Semi-supervised sparse metric learning using alternating linearization optimization[C]// *Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*. New York: ACM, 2010: 1139–1148.
  - [16] FU Zhenyong, LU Zhiwu, IP H H S, et al. Local similarity learning for pairwise constraint propagation[J]. *Multi-media tools and applications*, 2015, 74(11): 3739–3758.
  - [17] 刘笑, 陈家炜, 胡峻林. 用于亲属关系鉴别的成对约束组合度量学习 [J]. *山东大学学报(工学版)*, 2022, 52(2): 50–56.  
LIU Xiao, CHEN Jiawei, HU Junlin. Pairwise constrained compositional metric learning for kinship verification[J]. *Journal of Shandong University (engineering science edition)*, 2022, 52(2): 50–56.
  - [18] YAN Rong, ZHANG Jian, YANG Jie, et al. A discriminative learning framework with pairwise constraints for video object classification[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2006, 28(4): 578–593.
  - [19] LU Zhengdong, CARREIRA-PERPINAN M A. Constrained spectral clustering through affinity propagation[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage: IEEE, 2008: 1–8.
  - [20] LI Z, LIU J, TANG X. Pairwise constraint propagation by semidefinite programming for semi-supervised classification[C]// *Machine Learning, Proceedings of the Twenty-Fifth International Conference*. Helsinki, Finland, 2008: 264–272.
  - [21] YU S X, SHI J B. Segmentation given partial grouping constraints[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2004, 26(2): 173–183.
  - [22] LU Z, PENG Y. Exhaustive and efficient constraint propagation: a graph-based learning approach and its applications[J]. *International journal of computer vision*,

- 2013, 103(3): 306–325.
- [23] JIA Y, LIU H, HOU J. Pairwise constraint propagation with dual adversarial manifold regularization[J]. *IEEE transactions on neural networks and learning systems*, 2020, 31(12): 5575–5587.
- [24] LU Z, PENG Y. Heterogeneous constraint propagation with constrained sparse representation[C]// *IEEE International Conference on Data Mining*. Dallas: IEEE, 2013.
- [25] FU Z, IP H S, LU H, et al. Multi-modal constraint propagation for heterogeneous image clustering[C]// *Proceedings of the 19th ACM international conference on Multimedia*. Scottsdale: ACM, 2011: 143–152.
- [26] BERAHMAND K, ROSTOMI M, FOROUZANDEH S. A novel method of constrained feature selection by the measurement of pairwise constraints uncertainty[J]. *Journal of big data*, 2020, 83(7): 483–504.
- [27] 郑静, 熊伟丽. 基于互信息的多块 k 近邻故障监测及诊断 [J]. *智能系统学报*, 2021, 16(4): 717–728.
- ZHENG Jing, XIONG Weili. Multiblock k-nearest neighbor fault monitoring and diagnosis based on mutual information[J]. *CAAI transactions on intelligent systems*, 2021, 16(4): 717–728.
- [28] LEE D D, SEUNG H S. Algorithms for non-negative matrix factorization[C]// *International Conference on Neural Information Processing Systems*. Cambridge: MIT Press, 2000: 535–541.
- [29] 卞则康, 张进, 王士同. 基于类标感知的 KNN 分类算法 [J]. *模式识别与人工智能*, 2021, 34(10): 873–884.
- BIAN Zekang, ZHANG Jin, WANG Shitong. Class-aware based KNN classification method[J]. *Pattern recognition and artificial intelligence*, 2021, 34(10): 873–884.
- [30] FU Z Y, LU Z W, IP H, et al. Symmetric graph regularized constraint propagation[J]. *Proceedings of the AAAI conference on artificial intelligence*, 2011, 25(1): 350–355.

#### 作者简介:



孙乐, 硕士研究生, 主要研究方向为人工智能、模式识别。



王士同, 教授, 博士生导师, 主要研究方向为人工智能与模式识别。发表学术论文近百篇。