

# 基于多目标蚁群算法的稳定参考点选择

张磊<sup>1</sup>, 曹建军<sup>2</sup>, 刘艺<sup>1</sup>, 郑奇斌<sup>1</sup>, 李红梅<sup>1</sup>, 冯钦<sup>1</sup>

(1. 陆军工程大学 指挥控制工程学院, 江苏 南京 210007;

2. 国防科技大学 第六十三研究所, 江苏 南京 210007)

**摘要:**为提高异构数据实体分辨的参考点选择稳定性,提出了基于多目标蚁群算法的异构数据实体分辨参考点稳定选择方法。异构数据的实体分辨可选取若干匹配的数据对象作为参考点,将不同数据集中的对象转换为到各自参考点的距离向量,依据空间结构的相似性进行实体分辨;为选择出较为稳定的参考点子集并且确保实体分辨的准确性,以实体分辨准确性与参考点集的稳定性最优为目标建立模型。以文中所提方法选择的参考点子集与通过集成 Filter 方法选择的参考点子集的扩展昆彻瓦指标值作为参考点子集稳定性评价指标,以  $F_1$  值作为实体分辨准确性的评价指标;利用多目标蚁群算法求解该模型,同时将结合参考点之间距离的均值、方差和最大信息系数作为蚁群算法的启发式信息,提高了蚂蚁搜索较好解的能力,实现了参考点子集的优化选择。实验结果表明,该方法能够在实体分辨准确性和参考点子集稳定性方面达到较好的平衡。

**关键词:**异构数据;实体分辨;蚁群算法;参考点选择;稳定性

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2019)08-0001-06

doi:10.3969/j.issn.1673-629X.2019.08.001

## Stable Reference Data Selection Based on Multi-objective Ant Colony Algorithm

ZHANG Lei<sup>1</sup>, CAO Jian-jun<sup>2</sup>, LIU Yi<sup>1</sup>, ZHENG Qi-bin<sup>1</sup>,

LI Hong-mei<sup>1</sup>, FENG Qin<sup>1</sup>

(1. Institute of Command and Control Engineering, Army Engineering University, Nanjing 210007, China;

2. The 63rd Institute, National University of Defense Technology, Nanjing 210007, China)

**Abstract:** In order to improve the stability of reference data selection for heterogeneous data entity resolution, a method of stable reference data selection for heterogeneous data entity resolution based on multi-objective ant colony algorithm is proposed. Heterogeneous data entity resolution can select numbers of matching the data objects as reference data, other objects are converted to distance vector based on the reference data for the entity resolution based on the spatial structure of similarity. In order to select more stable reference data and ensure the accuracy of entity resolution, the optimal model of the entity resolution accuracy and stability reference data is established. We choose extended Kuncheva evaluation index of the reference data of the proposed method and that of integrated Filter method selection as an optimization objective, with  $F_1$  value as the other optimization objective. Using ant colony algorithm to solve this model, at the same time we combine with the reference data of the distance between the mean value, variance and coefficient of maximum information as heuristic information of ant colony algorithm, improving the ability of the ant searching better solutions and achieving the optimal selection of reference data. The experiment shows that the method can achieve a great balance between the accuracy of solid resolution and the stability of reference data.

**Key words:** heterogeneous data; entity resolution; ant colony algorithm; reference data selection; stability

## 0 引言

实体分辨是找到同一实体在同一数据源或不同数据源的不同描述,正确识别出不同实体的过程<sup>[1]</sup>。随

着互联网数据产生能力的增强,异构数据成为大多数互联网应用和平台的基础数据结构<sup>[2]</sup>。例如一个互联网分享平台可能包括了视频、图片、文本标签等数据,

收稿日期:2018-08-30

修回日期:2019-01-04

网络出版时间:2019-03-27

基金项目:国家自然科学基金(61371196)

作者简介:张磊(1989-),男,硕士研究生,研究方向为数据工程;曹建军,博士,副研究员,研究方向为数据质量控制与治理、数据智能分析。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190327.1624.022.html>

并形成了图片分享、图片标注等关联交互关系。这些异构数据处在不同的特征空间,因此传统的直接比较特征相似度的方法不能适用于异构数据的实体分辨。

现有异构数据实体分辨方法主要是通过学习一个公共的空间,并将异构的数据映射到该空间,从而使得异构的数据能够在一个公共的空间进行实体分辨<sup>[3]</sup>。但该方法需要大量的训练数据集,而异构数据的训练数据集获取成本高昂<sup>[4]</sup>。为了降低对训练数据的需求,同时保持较高的实体分辨准确性,Zheng 提出了一种基于空间结构相似性的实体分辨方法<sup>[5]</sup>。该方法主要从两个异构的数据源中分别选择表示为同一实体的数据对象作为参考点,将异构的数据对象转换成其到各自参考点的距离向量,依据空间结构的相似性来判断是否为同一实体。在该方法中,参考点的选择将影响数据的空间结构表征,进而影响实体分辨的准确性。同时,参考点选择的稳定性也成为影响数据空间结构表征的重要因素。

参考点选择稳定性是指选择方法对训练数据的微小扰动具有一定的鲁棒性,即能够生成较为稳定的参考点子集。在基于空间结构相似性的异构数据实体分辨中,由于参考点集的选择将影响数据的空间结构表征,稳定的参考点集具有较好的数据空间结构表征能力,从而提高了实体分辨的准确性。提高参考点集选择的稳定性可以选择出较为稳定的参考点集,从而提高数据的空间结构表征能力,提高实体分辨的准确性,增强结果的可信度。同时,如果仅采用实体分辨准确性评价参考点集的效果,在未知数据集上训练模型上易产生不稳定的泛化错误,通过稳定的参考点选择方法可以减少泛化错误,进一步提高模型的泛化能力。

文中主要对异构数据实体分辨参考点选择稳定性问题进行研究。为提高参考点选择的稳定性,进而选择出较优的参考点集,提高异构数据实体分辨的准确性,以实体分辨准确性和参考点子集稳定性最优为目标建立模型,使用多目标蚁群算法对模型进行求解,从而选择出较为稳定的参考点集。通过实验验证了该方法的有效性和优越性。

## 1 相关工作

为了解决异构数据实体分辨问题,近年来学者们提出了大量的解决异构数据实体分辨的方法,其中以利用学习公共空间的方法最为广泛,因此文中主要介绍学习公共空间的异构数据实体分辨方法。

利用学习公共空间的方法进行实体分辨主要是通过学习一个公共的空间,通过将异构的数据映射到该空间进行实体分辨的过程。Rasiwasia 等<sup>[6]</sup>运用典型相关分析(canonical correlation analysis, CCA)进行异

构数据的实体分辨,该方法主要通过提取异构数据的底层特征,找到一个统一的线性空间,使得异构的数据对象之间的相关性最大。Andrew 等<sup>[7]</sup>提出了利用深度学习的方法来找到异构数据的潜在联系,从而使得异构数据关联起来。Kang 等<sup>[8]</sup>提出了一种利用条件随机场的方法来学习异构数据的统一表示,从而使得异构的数据关联起来。

Pereira 等<sup>[9]</sup>提出了三种异构数据实体分辨方法:相关性匹配(correlation matching, CM)、语义匹配(semantic matching, SM)、语义相关匹配(semantic correlation matching, SCM)。CM 是一种利用底层特征来学习异构数据之间关系的方法,即针对每条异构数据对象学习一个映射到公共空间的变换,通过最大化公共空间的相互联系来保证数据表示之间的语义联系。SM 是通过对数几率回归将异构数据映射到同一个语义空间,然后直接在这个语义空间中进行数据的拟合与匹配。而 SCM 则是两者的结合,即建立一个联系最大的语义空间,并通过对数几何回归将异构数据映射到该语义空间,从而更好地反映数据的语义之间的关系。

为了降低对训练数据的需求同时能够保持较高的实体分辨准确性,Zheng 等<sup>[5]</sup>提出了一种基于空间结构相似性的异构数据实体分辨方法。通过从异构的数据集中选取表示为同一实体的对象作为参考点,将数据对象转换为其到参考点的距离向量,之后依据空间结构的相似性进行实体分辨。在该方法中,较好的参考点集具有更好的空间表征能力,从而提高实体分辨的准确性。为了使得参考点集具有更好的空间结构表征能力,需要选择较为稳定的参考点集,将选择稳定性方法用于上述参考点选择问题,从而提高参考点集的稳定性。

## 2 多目标蚁群算法稳定参考点选择

本节主要以实体分辨的准确性和参考点集的稳定性作为优化目标,使得在异构数据实体分辨中能够选择出较为稳定且能够使得实体分辨准确性较高的参考点。

### 2.1 优化目标及数学模型

基于空间结构相似性的实体分辨方法主要是从两个异构的数据源中分别选择表示为同一实体的数据对象作为参考点,将异构的数据对象转换成其到各自参考点的距离向量,依据空间结构的相似性来判断是否为同一实体。在该方法中,需要选择出能够使得实体分辨准确性较高且较为稳定的参考点集。

在上述问题描述中,优化目标有 2 个,一是实体分辨的准确性,二是参考点集的稳定性。文中把实体分

辨看作二分类问题,正类表示待分辨的对象为同一实体,使用  $F_1$  值作为实体分辨准确性指标。

设参与分类的样本总数为 Num,正类正确区分为正类的样本数为  $P\_num$ ,负类正确区分为负类的样本数为  $N\_num$ ,正类区分为负类的样本数为  $FN\_num$ ,负类区分为正类的样本数为  $FP\_num$ ,则查全率  $R$ 、查准率  $P$ 、 $F_1$  指标分别为:

$$R = \frac{P\_num}{P\_num + FN\_num} \tag{1}$$

$$P = \frac{P\_num}{P\_num + FP\_num} \tag{2}$$

$$EK(s, s') = \frac{|s \cap s'| - \frac{|s| \cdot |s'|}{c}}{\max[-\max(0, |s| + |s'| - c) + \frac{|s| \cdot |s'|}{c}; \min(|s|, |s'|) - \frac{|s| \cdot |s'|}{c}]} \tag{4}$$

其中,  $c$  为参考点的对数。  
文献[12]分别通过集成卡方检验、信息增益和 ReliefF 和 SVM-RFE 等 Filter 方法对特征子集进行选择,发现采用卡方检验、信息增益和 ReliefF 的集成方法(简称 Ensemble1)选择特征子集的稳定性较好,其主要步骤是分别采用上述三种 Filter 方法对样本进行排序,最后通过中值法集成多种排序的结果。

文中采用 Ensemble1 方法来选择参考点子集。假定存在  $c$  对参考点集合对  $R = \{U, V\}$ , 其中  $U = \{u_1, u_2, \dots, u_c\}$ ,  $V = \{v_1, v_2, \dots, v_c\}$  分别来自两个异构数据集。为了从  $R$  中选择出较为稳定的参考点子集,分别利用卡方检验、信息增益和 ReliefF 方法对参考点集  $R$  中的对象进行排序选择,之后通过中值法集成多种结果,得到 Ensemble1 方法选择出来的参考点子集  $R^n = \{U^n, V^n\}$ , 其中  $1 \leq n \leq c$ 。为了确保文中方法能够选择出较为稳定的参考点子集,令  $s' = R^n$ , 即用 Ensemble1 方法选择出来的参考点子集  $s'$  与文中方法选择的参考点子集  $s$  进行比较,以指导文中的参考点稳定性选择方法,使其能够选择出较为稳定的参考点子集。

扩展昆彻瓦相似度指标的取值在  $[0, 1]$  之间<sup>[13]</sup>, 指标取值越大,两个参考点集合相似度越高,说明选择

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{3}$$

稳定性的度量指标<sup>[10]</sup>主要包括谷元距离、邓恩稳定性指标、权重一致性指标和扩展昆彻瓦相似度度量指标等<sup>[11]</sup>。由于扩展昆彻瓦相似度度量指标能够度量不同规模的子集,具有良好的可扩展性,因此文中采用扩展昆彻瓦相似度指标对参考点集的稳定性进行度量。设由同一选择方法生成的两个参考点集合为  $s$  和  $s'$ , 它们的扩展昆彻瓦指标值的计算如下:

方法的稳定性越好。  
因此,建立如下两目标优化数学模型  
$$\max F_1 \tag{5}$$
$$\max EK(s, s') \tag{6}$$
$$s. t. \quad 1 \leq |s| \leq c, 1 \leq |s'| \leq c$$

上述模型中,  $s$  和  $s'$  为同一参考点选择方法生成的两个参考点集合。式 5、式 6 表示基于参考点集  $s$  的  $F_1$  指标最高同时子集的稳定性最好。上述优化模型可以看作一个两目标子集问题进行求解。

2.2 模型分析及多目标蚁群算法求解

由于两目标优化问题通常并不存在各目标都为最优的解,而存在一非劣解集,称为 Pareto 解集。多目标优化的目标就是找出一组解,尽可能全面地逼近 Pareto 解集,决策者可按需求选择出自己的满意解。

对于求解多目标优化模型,文献[14]提出了一种求解子集问题的基于图的蚂蚁系统。针对子集问题,定义了构造图和等效路径,提出了基于等效路径增强的信息素更新策略,克服了蚂蚁寻找解的有序性与问题解的无序性之间的矛盾,并以多维背包问题为例验证了该系统的有效性和优越性。文中将该蚁群算法用于求解式 5、式 6。

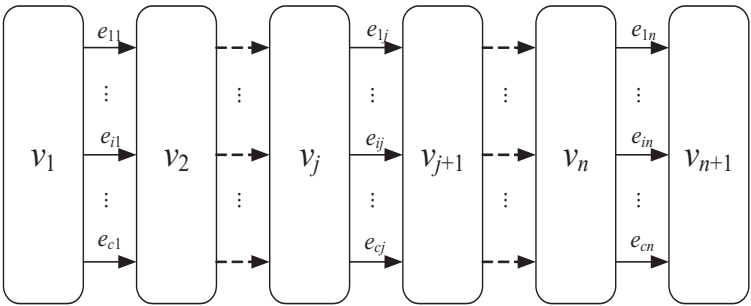


图 1 基于图的蚂蚁系统

如图 1 所示,在基于图的蚂蚁系统中,  $t$  时刻在节点  $v_1$  生成  $m$  只蚂蚁,每只蚂蚁根据边上的信息素和启



发式信息独立地选择某一边向下一节点移动。 $c$  为可选边的数量,即为蚂蚁可选择的参考点对的数量, $n$  为蚂蚁在一次搜索过程中寻找解的个数,即蚂蚁在一次搜索中找到  $n$  对参考点, $e_{ij}$  表示蚂蚁在第  $j$  步选择第  $i$  条边。

在基于图的蚂蚁系统中, $t$  时刻第  $m'$  只蚂蚁使用式 7 进行路径选择。

$$e_{ij}(t) = \begin{cases} \frac{\tau_{ij}^\alpha(t-1)\eta_i^\beta}{\sum_{e_{ij} \notin \text{tabu}_m} \tau_{ij}^\alpha(t-1)\eta_i^\beta} & e_{ij} \notin \text{tabu}_m \\ 0 & \text{其他} \end{cases} \quad (7)$$

其中,禁忌表  $\text{tabu}_m$  为第  $m'$  只蚂蚁走过的边,即蚂蚁已选择的参考点对; $\alpha$  和  $\beta$  为信息素量和启发式信息的重要程度;启发式信息  $\eta_i$  是外部信息,表示选择第  $i$  对参考点的希望程度; $\tau_{ij}(t)$  为在  $t(t=0,1,2,\dots)$  时刻边  $e_{ij}$  上的信息素量。

采用多个信息素矩阵能够获得较好的 Pareto 解,一般的设置方式为每个目标对应一个信息素矩阵,由于需要同时优化两个目标: $F_1$  指标与稳定性指标,因此需要设置 2 个信息素矩阵。2 个信息素矩阵的设置意味着在计算条件转移概率的过程中,需要对 2 个信息素值进行聚合,因此需要将两个信息素值聚合成单个值计算路径选择概率,聚合采用式 8:

$$\tau_{ij} = (\tau_{ij}^1)^{(1-\lambda)} \cdot (\tau_{ij}^2)^\lambda \quad (8)$$

其中, $\lambda (0 \leq \lambda \leq 1)$  是权重参数。

当蚂蚁每次迭代完毕后,使用帕累托档案解更新 2 个信息素矩阵时采用目标最好解方法,在  $t$  时刻拟对路径  $\text{tabu}^t$  的信息素增强,信息素更新如下:

$$\tau_{ij}(t) = \begin{cases} (1-\rho) \times \tau_{ij}(t-1) + \Delta'(\text{tabu}^t) & e_{ij} \in \Gamma(\text{tabu}^t) \\ (1-\rho) \times \tau_{ij}(t-1) & \text{其他} \end{cases} \quad (9)$$

其中, $\Delta'(\text{tabu}^t)$  为信息素增量; $\rho (0 \leq \rho \leq 1)$  为信息素挥发系数。

在蚁群算法中,启发式信息表示蚂蚁在路径选择中的先验偏好,能够有效提升蚂蚁获得较好的解的能力。基于异构数据实体分辨背景,提出了基于参考点之间的距离、方差和最大信息系数(maximal information coefficient, MIC)评估值相结合的启发式信息定义方法。

在基于空间结构的异构数据实体分辨中,参考点之间的距离较大或参考点之间的距离方差较大时,其实体分辨准确性较好<sup>[15]</sup>。同时, MIC 是一种度量变量之间任意关系的方法, MIC 的评估值越大,表明变量之间的相关性越强。文中利用 MIC 值度量参考点两两之间的相关性, MIC 的取值在  $(0,1)$  之间, MIC 的

评估值越大,表明两个参考点之间的相关性越强,同时有  $\text{MIC}(u_i, u_j) = \text{MIC}(u_j, u_i)$ 。因此,将参考点  $u_i$  的 MIC 的评估值设置为:

$$\text{MI}(u_i) = \frac{\text{MIC}(u_i, u_1) + \text{MIC}(u_i, u_2) + \dots + \text{MIC}(u_i, u_k)}{c} \quad (10)$$

其中, $c$  为数据集中参考点的数量。

综上,第  $i$  对参考点的信息素设置为:

$$\eta_i = \omega[\mu_u^i \times (\sigma_u^i)^\gamma] + [\mu_v^i \times (\sigma_v^i)^\gamma] + \varepsilon[\omega \text{MI}(u_i) + \text{MI}(v_i)] \quad (11)$$

其中, $\mu_u^i$ 、 $\mu_v^i$  为参考点  $u_i$ 、 $v_i$  到其所在样本中的其他对象的距离均值; $\sigma_u^i$ 、 $\sigma_v^i$  为参考点到样本中的其他对象的距离方差; $\text{MI}(u_i)$ 、 $\text{MI}(v_i)$  为参考点  $u_i$ 、 $v_i$  的 MIC 评估值; $\varepsilon$  为两优化目标的权重; $\omega$  为参考点集合中两个异构数据集的权重; $\gamma$  为均值与方差的调和因子。

文中提出的蚁群算法选择参考点的伪代码如下所述:

算法:蚁群算法选择稳定参考点。

(1) BEGIN

(2) 在参考点训练样本中采用卡方检验、信息增益和 ReliefF 方法进行排序选择参考点子集

(3) 集成步骤 2 中的三种参考点集选择结果(Ensemble1)

(4) 计算训练样本中每个对象与其他对象之间的距离均值和距离方差

(5) 计算训练样本中每个对象与其他对象的最大信息系数评估值

(6) 利用式 11 集成上述两种信息作为启发式信息

(7) 以  $F_1$  值和蚁群算法选择的子集与 Ensemble1 选择的子集的扩展昆彻瓦相似度指标作为蚁群算法的优化目标,采用多目标蚁群算法搜索较好的参考点子集

(8) END

上述算法中,步骤 2 和步骤 3 主要是利用 Ensemble1 方法对参考点集进行排序选择,以此作为蚁群算法选择参考点的指导信息;步骤 4~6 是计算蚁群算法的启发式信息,以便蚂蚁能够选择较好的解;步骤 7 为蚂蚁根据图上的信息素值和启发式信息值搜索较优的参考点子集并更新其相应路径。

### 2.3 算法复杂度分析

在异构数据参考点选择问题上,蚂蚁寻找最优解的时间复杂度(最坏情况)为  $O(\text{NC} \times m \times c^2)$ ,其中 NC 为算法的最大迭代次数, $c$  为备选解的个数。因此蚂蚁在搜索参考点子集时最坏情况下的时间复杂度为  $O(\text{NC} \times m \times c^2)$ 。

上述方法中,蚁群算法中信息素表是占用存储资源最大的部分,其空间复杂度为  $O(c^2)$ ,因此蚂蚁寻找最优解的空间复杂度为  $O(c^2 + m)$ 。

3 实验与分析

在本节使用真实数据集进行对比实验,验证了文中方法的有效性。

3.1 实验数据

实验数据来源于公开数据集 NUS-WIDE 中的子集 NUS-WIDE-LITE, NUS-WIDE-LITE 是一个从 Flickr 上获取的图片数据集,包括大概 27 000 张图片。其中图像的颜色直方图数据 Normalized\_CH\_Lite\_Test 中对象的特征个数为 64,图像的文本标签数据 Lite\_Tags1k\_Test 中对象的特征个数为 1 000。实验分别从以上两组数据中选取代表同一客观实体、并且文本标签数据不全为 0 的数据对象作为训练数据集。

3.2 实验设计

实验对以下三种方法进行比较。

方法 1(文中方法):利用两目标蚁群算法对参考点进行选择,并考虑参考点集的稳定性,简称 SSACO-SSM(space structure matching);

方法 2(文献[16]所提方法):在异构数据实体分辨中利用蚁群算法对参考点进行选择,但未考虑参考点集的稳定性,简称 ACO-SSM(ant colony optimized

space structure matching);

方法 3(文献[5]所提方法):在异构数据实体分辨中使用随机选择参考点的方法进行异构数据实体分辨,简称 SSM(space structure matching)。

共设计了 1 组对比实验,对比实验中共 2 组数据集,分别称为数据集 1 和数据集 2。这 2 组数据中,匹配的数据对象数量  $c$  分别为 600 对和 1 000 对,分别用三种方法在 2 组数据集中测试其实体分辨结果和参考点稳定性。该对比实验主要验证 SSACO-SSM 的  $F_1$  值和参考点集的稳定性指标  $EK(s,s')$  要好于 ACO-SSM 和 SSM。

蚁群算法的参数设置如下:初始时刻信息素  $\tau_{ij}(0)=100$ ,信息素和启发式信息的重要程度系数  $\alpha=2$ ,  $\beta=1$ ,信息素挥发系数  $\rho=0.2$ ,蚂蚁数量  $M=30$ 。

实验选用 SVM 作为分类器,SVM 中的核函数选用线性核函数。

3.3 实验结果

数据集 1 的  $F_1$  值指标和稳定性分别如图 2、图 3 所示。

数据集 2 的实验结果分别如图 4、图 5 所示。

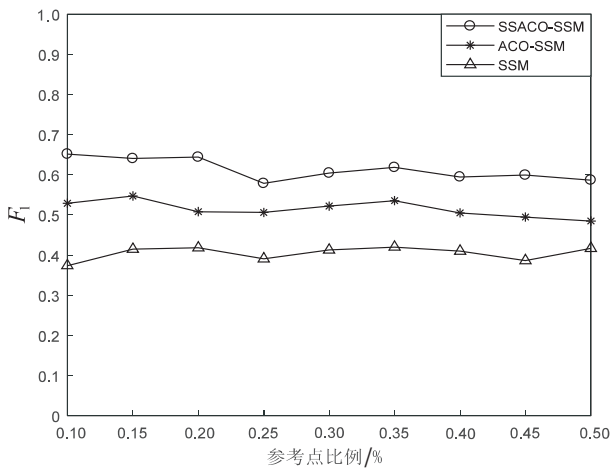


图 2 数据集 1 中的参考点比例与  $F_1$  值关系曲线

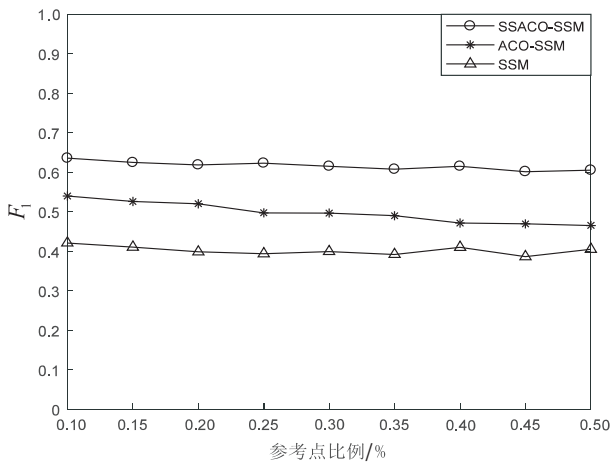


图 4 数据集 2 中参考点比例与  $F_1$  值关系曲线

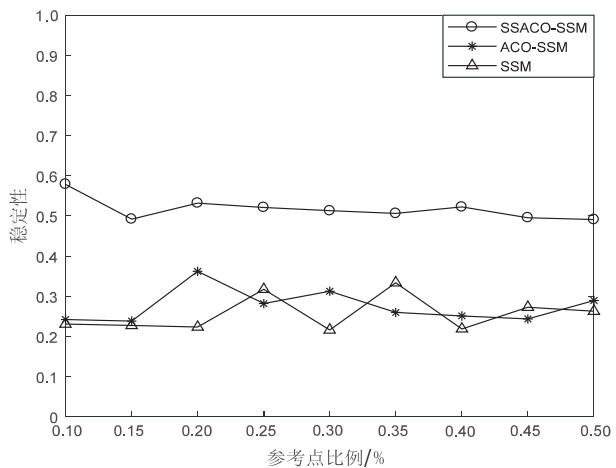


图 3 数据集 1 中参考点比例与稳定性关系曲线

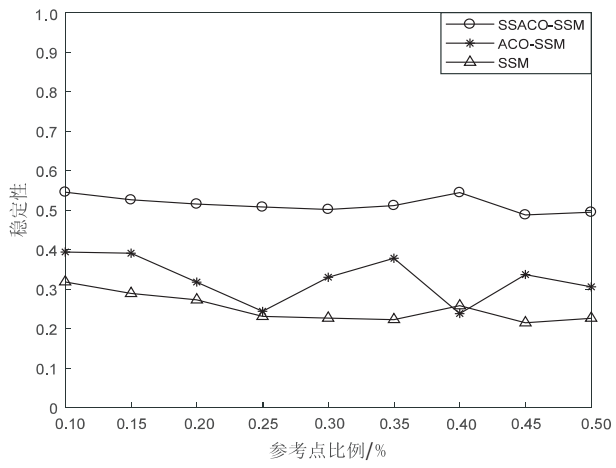


图 5 数据集 2 中参考点比例与稳定性关系曲线

从图 2、图 4 中可以看出,SSACO-SSM 方法 2 个数据集上的  $F_1$  值都要好于其他两种方法,说明 SSACO-SSM 方法能够有效提高异构数据实体的准确性。从图 3、图 5 中可以看出,SSACO-SSM 方法在稳定性指标方面都要好于其他两种方法,说明 SSACO-SSM 方法能够有效提高参考点集的稳定性。分别对比图 2 和图 3、图 4 和图 5,可以看出,当参考点集稳定性提高时能够有效提高异构数据实体分辨的  $F_1$  值,即实体分辨准确性。

综上,文中方法在异构数据实体分辨选择参考点子集时具有良好的稳定性,同时也具有较高的实体分辨准确性。

## 4 结束语

为了提高异构数据实体分辨中参考点选择的稳定性,提出了一种基于多目标蚁群算法的参考点选择方法,实现了参考点子集稳定性和实体分辨准确性的综合最优。得到如下结论:

(1)在基于空间结构相似性的异构数据实体分辨中提高参考点子集的稳定性,能够有效提高异构数据实体分辨的准确性;

(2)使用实体分辨准确性和最大信息度量系数作为蚁群算法的启发式信息,使得算法能够有效利用蕴含的先验信息,从而提高算法获得更好解的概率;

(3)异构数据实体分辨的参考点选择多目标优化模型,综合考虑了实体分辨准确性和参考点子集的稳定性,实现了效果的最优。

该方法与业务领域无关,具有较强的通用性。

## 参考文献:

- [1] 谭明超,刁兴春,曹建军. 实体分辨研究综述[J]. 计算机科学,2014,41(4):9-12.
- [2] 欧明栋. 面向大规模异构数据的哈希表征学习研究[D]. 北京:清华大学,2017.
- [3] 霍 静. 面向异构人脸识别的跨模态度量学习研究[D]. 南京:南京大学,2017.
- [4] GAO Nengneng, HUANG Shengjun, YAN Yifan, et al. Cross modal similarity learning with active queries [J]. Pattern Recognition, 2018, 75: 214-222.

- [5] ZHENG Qibin, DIAO Xingchun, CAO Jianjun, et al. Multi-modal space structure: a new kind of latent correlation for multi-modal entity resolution [D]. Ithaca, New York: Cornell University, 2018.
- [6] RASIWASIA N, PEREIRA J C, COVIELLO E, et al. A new approach to cross-modal multimedia retrieval [C]//ACM international conference on multimedia. Firenze, Italy: ACM, 2010: 251-260.
- [7] ANDREW G, ARORA R, BLIMES J, et al. Deep canonical correlation analysis [C]//Proceedings of the 30th international conference on international conference on machine learning. Atlanta, GA, USA: [s. n.], 2013: 1247-1255.
- [8] KANG Cuicui, LIAO Shengcai, HE Yonghao, et al. Cross-modal similarity learning: a low rank bilinear formulation [C]//Proceedings of the 24th ACM international on conference on information and knowledge management. [s. l.]: ACM, 2015: 1251-1260.
- [9] PEREIRA J C, COVIELLO E, DOYLE G, et al. On the role of correlation and abstraction in cross-modal multimedia retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(3): 521-535.
- [10] ZHOU Qifeng, DING Jianchao, NING Yongpeng, et al. Stable feature selection with ensembles of multi-relief [C]//Proceedings of the 2014 10th international conference on natural computation. Xiamen, China: IEEE, 2014: 742-747.
- [11] KAMKER I, GUPTA S K, PHUNG D, et al. Stabilizing 1-norm prediction models by supervised feature grouping [J]. Journal of Biomedical Informatics, 2015, 59: 149-168.
- [12] 刘 艺, 曹建军, 刁兴春, 等. 特征选择稳定性研究综述 [J]. 软件学报, 2018, 29(9): 2559-2579.
- [13] FAHAD A, TARI Z, KHALIL I, et al. An optimal and stable feature selection approach for traffic classification based on multi-criterion fusion [J]. Future Generation Computer Systems, 2014, 36(7): 156-169.
- [14] 曹建军, 张培林, 王艳霞, 等. 一种求解子集问题的基于图的蚂蚁系统 [J]. 系统仿真学报, 2008, 20(22): 6146-6150.
- [15] 高能能. 面向多标记/多模态数据的主动学习研究 [D]. 南京: 南京航空航天大学, 2017.
- [16] 张 磊, 曹建军, 郑奇斌, 等. 异构数据实体分辨中蚁群算法参考点选择 [J]. 兵工自动化, 2018, 38(7): 23-30.