昆明学院学报　2016, 38(6):94~99　ISSN 1674-5639　CN 53-1211/G4
Journal of Kunming University

基于模糊集和粗糙集的支持向量聚类算法研究

郭晨晨,朱红康
(山西师范大学 数学与计算机科学学院,山西 临汾 041000)

摘要:目前很多已知的聚类算法对于异常点的处理存在不合理的问题,将模糊集和粗糙集的相关理论加入到支持向量聚类算法中,可增加异常点处理的合理性,并得到一种新的改进算法,将其称为模糊—粗糙支持向量聚类算法. 当支持向量集作为一个特殊的聚类,通过元素间的亲密程度,模糊边界的隶属度可以被计算出来. 而下近似集包含的样本点建立在算法训练阶段获得的超球体内. 在检测异常值和计算任意轮廓的聚类方面,该算法具有较大的优势和潜力.

关键词:粗糙集;模糊集;支持向量集;超球体;数据挖掘
中图分类号:TP18　**文献标识码**:A　**文章编号**:1674-5639(2016)06-0094-06
DOI:10.14091/j.cnki.kmxyxb.2016.06.021

Research on Support Vector Clustering Algorithm Based on Fuzzy Set and Rough Set

GUO Chenchen, ZHU Hongkang
(College of Mathematics and Computer Science, Shanxi Normal University, Linfen, Shanxi, China　041000)

Abstract: Due to many clustering algorithms known today are not reasonable in dealing with outliers, the theory of fuzzy set and rough set are added to the support vector clustering algorithm, which can increase the rationality of outlier processing to present a new improved algorithm called fuzzy-rough support vector clustering algorithm. When the support vector set is used as a special clustering, the membership degree of the fuzzy boundary can be calculated by the closeness degree between the elements. The lower approximation set contains the sample points set up in the training phase of the algorithm in the hyper sphere. The algorithm has a considerable advantage and potential in detecting outliers and calculating the clustering of arbitrary profiles.

Key words: rough set; fuzzy set; support vector set; hyper sphere; data mining

所谓聚类,就是将相似的事物聚集在一起,而将不相似的事物划分到不同的类的过程,它是数据挖掘中十分重要的一种手段. 在文献[1~6]已经提出的许多聚类算法,可以将其分为两大类:硬聚类和软聚类(模糊聚类). 然而现实中的很多分类问题并没有严格的分类界限,因此传统的硬聚类方法也许不能充分反映出问题的本质. 而软聚类可以给予更多的灵活性,得到足够的解决方案. 当歧义和不确定必须被处理[3],模糊集[7]和粗糙集[8]便显示出其独特的优势,模糊 C 均值[5]和粗糙 C 均值[9]也是常见的代表. 软聚类算法及其衍生算法在许多领域得到了广泛的应用. 然而,他们的使用仍然存在一定的限制,例如球形形状的簇,其对象的隶属度值的总和必须等于1(模糊 C 均值),并且需要事先知道簇的个数. 另外,异常点不能被分类. 支持向量机(SVM)[4]在解决非线性、小样本及高维模式识别中表现出许多特有的优势. 然而,对于不属于任何簇的样本点,该算法统一将其以异常值对待. 但是,异常值存在的情况下,如果不进行异常数据的区分,会造成很多重要信息的丢失. 为了克服这一缺陷,提出模糊—粗糙支持向量聚类算法(Fuzzy-Rough Support Vector Clustering). 这种聚类算法具有以下优点:与传统的支持向量聚类相似;在预先不知道簇状态的情况下,任何形状的簇均可以被识别. 此外,没有明确将异常点分配给任何聚类,而是根据这些异常点在高维

收稿日期:2016-05-18
作者简介:郭晨晨(1992—),男,山西长治人,在读硕士,主要从事计算机应用方向研究;朱红康(1975—),男,山西汾西人,副教授,博士,硕士生导师,主要从事数据挖掘方面的研究.

空间与所有簇之间的距离分配隶属值. 这些隶属值为异常点提供了很多重要的信息,而这些信息恰恰是很多应用中最为关键的部分.

1　超球体支持向量聚类算法概述

1.1　超球体支持向量聚类概述

令 $X = \{x^i \in R^d / i = 1,2,\cdots,N\}$ 是集合的 N 个样本点,R^d 是数据空间. 传统支持向量聚类算法将集合 X 中的元素分类,来解释支持向量域的解. 这是一个二阶段的算法,包括训练阶段和标记阶段. 在训练阶段,通过 Tax 和 Duin[10] 提出的超球体支持向量聚类算法,将原始数据空间中的点通过映射函数映射到高维空间,在高维空间中寻找一个超球体,该超球体以尽可能小的半径包含尽可能多的样本点. 通过计算最小超球体的球心和半径,权衡超球体半径和它所覆盖的样本数,即最大化聚类中样本点的密度. 在封闭的超球体上,包含 3 种向量:1) 支持向量(SV);2) 边界支持向量(BSV);3) 内部样本点(ID). 作为支持向量的样本点位于超球体的表面,边界支持向量的样本点位于超球体的外部,内部样本点位于超球体的内部. 样本点的图像再由高维空间映射回原始空间,形成聚类的轮廓,这样就完成了训练阶段. 最后,标记阶段确定在训练过程中发现的不同的聚类,建立一个 $\{0,1\}$ 隶属度矩阵. 整个映射过程如图 1 所示.

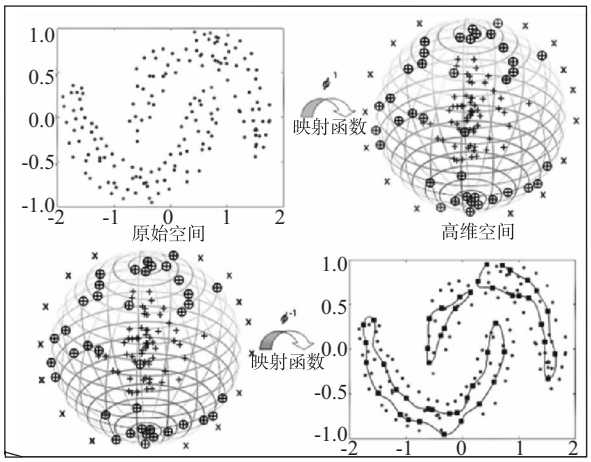

图1　超球体支持向量算法映射过程

1.2　数学描述

1.2.1　训练阶段

在训练阶段,是一个二次优化问题的求解过程. 该数学模型如下:

$$\min R^2 + C\sum_{i=1}^{N}\xi_i, \tag{1}$$

$$\text{s.t.}\ \|\varphi(x_i) - \alpha\|^2 \leqslant R^2 + \xi_i, \forall i = 1,\cdots,N, \tag{2}$$

$$\xi_i \geqslant 0, \forall i = 1,\cdots,N, \tag{3}$$

其中 $\|\cdot\|$ 是欧式范数,α 是超球体的球心,φ 是从原始空间到高维空间映射的非线性 Y 映射函数,ξ_i 是松弛变量,R 是超球体的半径,而 $C \in [0,1]$ 是一个惩罚参数.

应用模糊 — 粗糙支持向量算法获得一组关于 SV,BSV 和 ID 的数据是必要的,涉及的主要参数高斯核宽度参数 q,其影响支持向量的数量. 惩罚参数 $C = 1/N\upsilon$ 决定高维空间超球体中样本点所占全部样本点的百分比.

通过求解函数变量,可以得到原问题的解,这是一个函数的沃尔夫对偶形式:

$$\max\sum_{i=1}^{N}\beta_i K(x_i,x_j) - \sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i\beta_j K(x_i,x_j), \tag{4}$$

$$\text{s.t.}\ \sum_{i=1}^{N}\beta_i = 1, \tag{5}$$

$$0 \leqslant \beta_i \leqslant C, \forall i = 1,\cdots,N, \tag{6}$$

其中 β_i 是拉格朗日乘子,$K(x_i,x_j) = \varphi(x_i)\cdot\varphi(x_j)$ 是核函数.

1) $\beta_i \in (0,C)$,数据点位于超球体球面上,称为支持向量(SV);

2) $\beta_i = C$,数据点位于超球体外部,称为边界支持向量(BSV);

3) $\beta_i = 0$,数据点位于超球体的内部,称为内部数据点(ID).

一个被广泛使用的核函数,即高斯核函数,具体形式如下:

$$K(x_i,x_j) = e^{-q\|x_i - x_j\|^2}. \tag{7}$$

随着宽度参数 q 增大,支持向量的数量和簇的数量也相应增加. 对于任意数据点 x,其在高维空间中的图像中的位置到超球体中心的距离由下面公式得出:

$$R^2(x) = \|\varphi(x) - \alpha\|^2 = K(x,x) - 2\sum_{i=1}^{N}\beta_i K(x_i, x) + \sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i\beta_j K(x_i,x_j). \tag{8}$$

超球体半径计算如下:

$$R_s = \frac{1}{|SV|}\sum_{x_i \in SV} R(x_i). \tag{9}$$

最后将数据空间中簇的轮廓集合定义为:

$$\{x / R(x) = R_s\}. \tag{10}$$

轮廓的集合形成了如图 1 所示原始空间的聚类

边界. 边界支持向量均在聚类边界外部,而其余的样本点均在聚类内部.

在训练阶段,许多有效的算法使用支持向量数据描述模型来求解二次优化问题[11]. 例如,广义序列最小优化(GSMO) 算法,Keerthi 和吉尔伯特[12]算法.

1.2.2 标记阶段

虽然 SVC 算法的训练阶段提供了支持向量、边界支持向量和内部样本点,但是它不能区分同时属于两个簇的样本点. 要做到这一点,Ben-Hur 等人[4]给出了以下策略:给定一对样本点 x_i 和 x_j,它们分别属于不同的簇,则 $\exists \lambda \in [0,1]$,使得 $R(y_{i,j}) > R_S$,其中 $y_{i,j} = \lambda x_i + (1-\lambda) x_j$. 定义邻接矩阵 A,$a_{i,j}$ 是邻接矩阵中的元素,而点 x_i 和 x_j 是位于超球体上或者内部的点.

$$a_{i,j} = \begin{cases} 1, & \text{if} \quad \forall \lambda \in [0,1], R(y_{i,j}) \leqslant R_s, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

聚类现在被定义为一种连接超球体中样本点的封闭曲线,而边界支持向量是未经分类的,因为从高维空间图像中可以看出,边界向量位于超球体的外部. 我们可以将这些未经分类的边界支持向量保持未分类的状态或者应用一些关键的规则来将他们分到一些聚类中去.

Ben-Hur 引入的标记规则被称为支持向量曲线图. 由于其需要高昂的计算时间成本,很多相关的改进理论也被提出,诸如近似图建模[13]、稳定平衡点[14]、锥形聚类标签[15-16]、快速稳定标签[17] 等.

1.3 支持向量聚类算法的扩展

在本节,对支持向量聚类扩展进行简单介绍,并且阐述模糊 — 粗糙支持向量聚类算法所做的改进.

模糊支持向量聚类算法[6] 用多个超球体来表示高维空间中不同的簇,并且用遗传算法计算隶属度;García 等人[11] 提出的减法聚类算法用于发现每个簇的类中心,主要应用于支持向量机的密度估计领域;基于模糊 C 均值的元素隶属度算法用迭代思想来进行. 这些方法提供隶属度矩阵,即数据的模糊划分. 最后 Zheng 等人[18] 通过权衡分类器结构的相关性为每一个样本点分配隶属值,例如在输入电平领域应用模糊信息理论,使用这样的隶属函数可以实现数据的分区工作.

支持向量聚类算法的一个重要变化由 Asharaf 等人[2] 于2005 年提出,它的核心思想是在高维空间寻找包围样本点的两个超球体而不是仅仅一个. 由更小的超球体包围的元素被认为是聚类的下近似元素. 那些被较大球体包围但是没有被较小球体包围的样本点被假定为具有相同聚类边界的元素. 标记部分是由 Ben-Hur 的标记算法改进而成. 通过这种方法,得到数据集的软分区.

2 模糊 — 粗糙支持向量聚类算法

2.1 方法概述

在传统 SVC 算法的训练阶段,方程定义的模型如 1.2.1 节所述.

如果在这个模型中选择 $C = 1$,即没有样本点位于超球体之外,边界是空的. 那么该数据集的清晰的分区会出现(图1). 另一方面,如果 $C < 1$,一些样本点(BSV) 将位于超球体外部并且围绕在聚类周围(图 1). 球体内的样本点被使用标签规则建立成为聚类元素.

由传统 SVC 产生的清晰结构将由模糊 — 粗糙的结构替代,其特征由两个主要部分组成:即由超球体内部样本点所建立的元素下近似集和一个由边界支持向量组成的模糊边界(位于超球体外部的样本点). 模糊边界概念通过以下两个想法产生:

1) 某些边界支持向量非常接近支持向量构建的聚类边界;

2) 模糊集理论提供了用于获得聚类边界样本点隶属值的数学框架.

结合这些想法,提出了这一概念,它带给 BSV 更多的灵活性. 例如,如果 BSV“明显” 远离聚类边界(隶属度趋近于0),即被认为是异常值. 如果隶属度达到一定的阈值,则参与到聚类的过程中(图 2). 与其他的模糊聚类算法相比,本方法不需要保证样本点对于所有聚类的隶属度总和为 1. 另外预先假设聚类的形状和数目的约束条件,这样 SVC 的灵活性则不受影响.

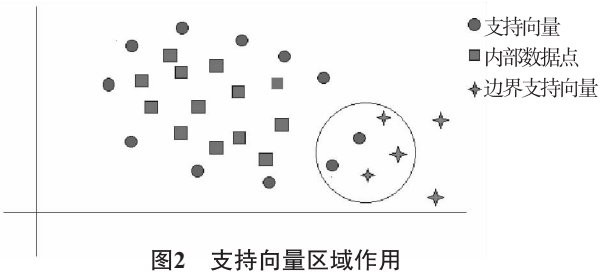

图2 支持向量区域作用

2.2　算法详细阐述

传统的SVC算法输出:支持向量(SV)、边界支持向量(BSV)、内部样本点(ID) 和聚类数目以及以{0,1} 为隶属度的隶属度矩阵. 值得注意的是,由于边界支持向量没有被分类,所以统一给定隶属度为0.

正如1.1 节所述,所有超球体中的内部样本点将被分到一个聚类中. 因为这些样本点的隶属度不存在不确定性,所以可以将每个簇的样本点组成的点集设定为各自聚类的下近似集.

SVC 算法对边界支持向量是不进行分类的,因为这些边界支持向量的隶属度存在不确定性. 而模糊边界是基于这样的思想:每个支持向量可以作为其圆形聚类的中心. 这样边界支持向量也会参与到支持向量的构建中来(图 2),该支持向量是这些边界向量的最佳代表.

基于这样的思想,通过计算边界支持向量与其聚类中心的距离获得相应的隶属度. 考虑到 SVC 算法形成的聚类结构是基于所有支持向量的,因此对于计算每一个边界支持向量隶属度时,需要对计算对象进行选择.

方法 1　取距离最近的支持向量作为计算目标;

方法 2　取所有支持向量作为计算目标.

对于边界支持向量 i 和支持向量 j,使用高斯核函数$\mu_{i,j}$作为他们的隶属度函数. 因为该核函数值可以映射到(0,1] 之间,公式如下:

$$\mu_{i,j} = \mu(x_i, SV_j) = K(x_i, SV_j) = e^{-q\|x_i - SV_j\|^2}.$$

当一个样本点十分接近于它所在聚类的支持向量时,隶属度趋近于 1,反之则趋近于 0.

使用这种内核函数作为隶属函数,也在于它保持高维空间与原始空间的相同判定规则,例如,支持向量在高位空间中很接近于超球体的表面,下面的定理可以表示为:

令 x_i, x_j, x_k 为数据空间 R^d 中的元素,$d(x,y)$ 为欧氏距离,φ 表示从数据空间 R^d 到高维空间的转化函数.

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) = e^{-q\|x_i - x_j\|^2}, (q > 0),$$

$$d(x_i, x_j) \leqslant d(x_i, x_k) \Leftrightarrow d(\varphi(x_i), \varphi(x_j)) \leqslant d(\varphi(x_i) \cdot \varphi(x_k)).$$

使用核函数作为隶属函数计算隶属度比使用任意隶属函数更自然,根据标准规则计算隶属度矩阵,从而得到最后的 BSV 隶属度如下:

1) 如果使用最近支持向量准则,边界支持向量 i 到聚类 j 的隶属度用 $K(BSV_i, SV_j)$ 的值表示当且仅当支持向量 i 是聚类 j 到边界支持向量 i 最近的支持向量.

2) 如果考虑一个聚类中所有支持向量到一个边界支持向量的平均距离,则需要下面的公式:

$$\mu_{i,j} = \frac{1}{|SV_j|} \sum_{x_k \in SV_j} K(BSV_i, x_k),$$

其中 SV_j 是聚类 j 的支持向量集.

一旦上面的思想被采用,即可得到隶属矩阵 U,$\mu_{i,j}$ 是边界支持向量 i 到聚类 j 的隶属度. 相比模糊 C 均值提供的隶属度矩阵,该方法形成的隶属度矩阵的每一行元素之和不一定是 1,因而更容易发现异常值的存在. 如果一个样本点是异常值,那么它对于所有聚类的隶属度都接近于 0,而非异常值显然不具有这样的性质.

为了实现粗糙 — 模糊聚类算法,将 1.2 节所有的思想进行结合以得到最终的隶属度矩阵和计算模糊粗糙聚类算法,算法 1 表示为:

输入　数据集 X,参数 $q > 0$,$\upsilon \in (1/N, 1)$,

输出　关于粗糙 — 模糊聚类的[0,1] 隶属度矩阵,聚类数 C.

步骤 1　计算惩罚参数 $C = 1/N\upsilon$;

步骤 2　SVC 算法的训练阶段,获得支持向量集(SV),边界支持向量(BSV),内部样本点(ID);

步骤 3　SVC 算法的标记阶段,获得清晰的数据集分区;

步骤 4　根据标记阶段获得的结果,将 SV 和 ID 分配到各自聚类的下近似集中;

步骤 5　将 BSV 分配给模糊边界集中;

步骤 6　对位于超球体以外的样本点生成隶属度矩阵(根据 BSV 到 SV 的距离);

步骤 7　按标记阶段划分距离矩阵;

步骤 8　对于每一个簇执行以下过程:任意 $x_i \in BSV$,获得最近的 $x_j \in SV$,则 x_i, x_j 属于同一聚类;

步骤 9　计算最终的隶属度.

其中,1 ~ 3 步是关于支持向量算法,4 ~ 9 步是本文提出的改进算法. 若 $C = 1$,则边界支持向量集将是空的,而模糊边界将不复存在. 因为所有元素组成聚类的下近似集,由此获得清晰的聚类划分.

3　实验

为了更加清晰的展示改进算法的实验效果,先通过对 Motivation 数据集来进行聚类测试,再通过与几种常见聚类算法的聚类比较来测试模糊 — 粗糙支持向量聚类算法的聚类性能.

3.1 Motivation 数据集聚类性能测试

Motivation 数据集是人工数据集,包含 316 个实例,其中有 16 个样本点位于两个主要聚类的外部. 对于这个样本数据集,通过若干次实验,调整得到支持向量算法参数 $q = 12$,$v = 0.074$,经过训练阶段,支持向量、边界支持向量、内部样本点显示如图 3 所示. 其中大圆点样本点为支持向量,三角形样本点为边界支持向量,小圆点为内部样本点.

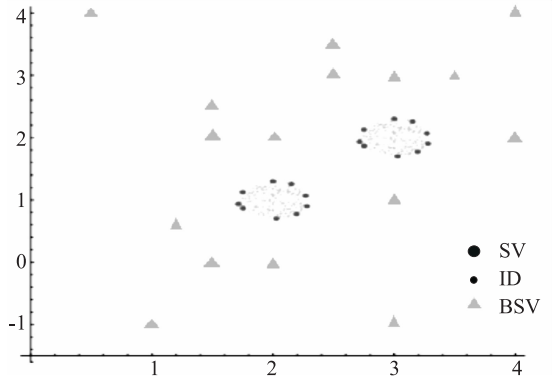

图3 Motivation数据集聚类图像

使用算法的 6 ~ 9 步骤计算的边界支持向量的隶属度矩阵如表 1 所示.

表 1 边界支持向量隶属度矩阵

边界支持向量 BSV	X	Y	$\mu_{i,1}$	$\mu_{i,2}$
1	4.0	4.0	9.617×10^{-58}	2.611×10^{-20}
2	2.5	3.0	4.182×10^{-17}	1.401×10^{-4}
3	3.5	3.0	4.916×10^{-26}	3.103×10^{-4}
4	3.0	3.0	2.611×10^{-20}	2.783×10^{-3}
5	2.5	3.5	2.902×10^{-27}	1.564×10^{-9}
6	4.0	2.0	7.185×10^{-21}	1.749×10^{-3}
7	1.2	0.6	1.111×10^{-2}	1.255×10^{-21}
8	1.5	2.0	1.401×10^{-4}	1.974×10^{-8}
9	2.0	2.0	2.782×10^{-3}	2.088×10^{-3}
10	1.5	2.5	1.564×10^{-9}	1.345×10^{-9}
11	3.0	1.0	1.749×10^{-3}	2.698×10^{-3}
12	1.5	0.0	9.638×10^{-5}	4.901×10^{-27}
13	2.0	0.0	2.698×10^{-3}	2.554×10^{-21}
14	3.0	-1.0	1.654×10^{-20}	9.120×10^{-39}
15	1.0	-1.0	2.554×10^{-21}	1.601×10^{-59}
16	0.5	4.0	1.946×10^{-50}	2.085×10^{-45}

从图 3 可知,本算法得到了相当清晰的聚类图像. 对于表 1,如果一个 BSV 的最大隶属度非常接近 0,那么就可以认定他为异常值. 对于 Motivation 数据集,所有的边界支持向量均可以被认为是异常点,这样符合该数据集的特性,但并不意味着所有数据集均为如此.

3.2 对比实验

在这一节通过与粗糙模糊 C 均值算法、可能性模糊 C 均值算法进行比较,以此来衡量本文算法的可行性和优越性. 测试数据集为 XO,XO Outlier,BankNote,Cancer,Glass 等经典数据集. 结果如表 2 所示.

表 2 算法性能对比实验

数据集名称	性能指标	RFSVC	RFCM	RPCM
XO	α	0.978 3	0.541 2	0.968 5
	α^*	0.978 5	0.532 9	0.983 5
XO Outlier	α	0.981 2	0.539 1	0.965 8
	α^*	0.981 3	0.531 1	0.983 3
Glass	α	0.944 6	0.859 6	0.162 9
	α^*	0.995 5	0.896 2	0.892 0
BankNote	α	0.982 5	0.782 3	0.821 0
	α^*	0.992 8	0.798 4	0.900 4
Cancer	α	0.811 7	0.808 6	0.448 3
	α^*	0.811 7	0.812 8	0.788 9

注:1)α 表示聚类的平均精度,α^* 表示聚类的下近似精度[19];2)RFSVC 表示模糊粗糙支持向量聚类算法,RFCM 表示粗糙模糊 C 均值算法,RPCM 表示可能性模糊 C 均值算法.

4 小结

本文对支持向量算法进行了一定的改进,图像和有效性指标表均表明,本算法显示出在人工数据集和现实数据集聚类方面的较大优势和潜力.

[参考文献]

[1] FILLIPPONE M, CAMASTRA F, MASULLI F, et al. A Survey of kernel and spectral methods for clustering[J]. Pattern Recognit, 2008, 41(1): 176 - 190.

[2] KOU G, PENG Y, WANG G. Evaluation of clustering algorithms for financial risk analysis using MCD M methods[J]. Information Sciences, 2014, 275: 1 - 12.

[3] PETERS G, CRSPO F, LINGRAS F, et al. Soft clustering: Fuzzy and rough approaches and their extensions and derivatives[J]. International Journal of Approximate Reasoning, 2012, 54(2): 307 - 322.

[4] VEGA-PONS S, RUIZ-SHULCOLOPER J. A survey of clustering ensemble algorithms[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2011, 25(3): 337 - 372.

[5] WANG Y Y, CHEN S C. Soft large margin clustering[J]. Information Sciences, 2013, 232: 116 - 129.

[6] XU Y, WUNSCH D. A survey of clustering algorithms[J]. IEEE

Transactions on Neural Networks,2005,16(3):645 – 678.
[7] ZADEH L. Fuzzy sets[J]. Information and Control,1965,8(3):338 – 353.
[8] PAWLAK Z. Rough sets [J]. International Journal of Approximate Reasoning,1982,11(5):341 – 356.
[9] LINGRAS P, WEST C. Interval set clustering of web users with rough *k*-means[J]. Journal of Intelligent Information Systems,2004,23(1):5 – 16.
[10] TAX D, DUIN R. Support vector domain description[J]. Pattern Recognition Letters,1999,20(11):1191 – 1199.
[11] MAHDAVI S, SHIRI M E, RAHNAMAYAN S. Metaheuristics in large-scale global continues optimization: Asurvey [J]. Information Sciences,2014,295:407 – 428.
[12] KEETHI S S, GILBERT E G. Convergence of a generalized SMO algorithm for SVM classifier design [J]. Machine Learning,2002,46(1):351 – 360.
[13] YANG J H, CASTRO V E, CHALUP S. Support vector clustering through proximity graph modelling[J]. Neural Information Processing,2002,2:898 – 903.
[14] LEE J, LEE D. An improved cluster labeling method for support vector clustering[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2005,27(3):461 – 464.
[15] LEE S H, DANIELS K M. Cone clustering labeling for support vector clustering [J]. Department of Computer Science,2011,26(5):484 – 488.
[16] LEE S H, DANIELS K M. Gaussian kernel width exploration and cone cluster labeling for support vector clustering [J]. Formal Pattern Analysis & Applications, 2012, 15(3):327 – 344.
[17] LI H. A fast and stable cluster labeling method for support vector clustering[J]. Journal of Computers,2013,8(12):3251 – 3256.
[18] ZHENG E H, YANG M, LI P, et al. Fuzzy support vector clustering[J]. Lecture Notes in Computer Science,2006,397:1050 – 1056.
[19] MAJI P, PAL S K. Rough set based generalized Fuzzy *C*-means algorithm and quantitative indices[J]. IEEE Transactions on Cybernetics,2007,37(6):1529 – 1540.

(上接第 93 页)

本文关于格桑花的渊源调查还存在局限性,仅以词义考证、历史及传说为线索进行分析推断,还不够完备,尚需进一步考证. 总之,研究还有待于从历史、民俗文化以及格桑花文化发源地生境入手,并结合植物的特性进行考证,一定会找到更科学严谨的解释.

[参考文献]

[1]苏发祥. 中国藏族[M]. 银川:宁夏人民出版社,2012.
[2]赤烈曲扎. 西藏风土志[M]. 拉萨:西藏人民出版社,2006:11 – 26.
[3]贺中. 西藏旅游探险手册[M]. 拉萨:西藏人民出版社,2001.
[4]祁连休,肖莉. 中国传说故事大辞典[M]. 北京:中国文联出版公司,1992.
[5]杨青松,特姆,高建国,等. 藏族明星植物“格桑梅朵”基原[C]//国际科学与工程中心. 细胞、分子生物学、生物物理学和生物工程国际会议. 2010:201 – 204.
[6]杨青松,赵艳. 格桑梅朵的民间传说[J]. 今日民族,2011(6):24 – 25.
[7]央金卓嘎. 藏族文化植物格桑梅朵[J]. 西藏科技,2012,237(12):73 – 76.
[8]张一诺. 你以为你以为的就是事实吗:植物学博士顾垒告诉你七个植物真相[N]. 中国绿色时报,2015 – 07 – 16(4).
[9]龙行天下. 青藏高原上的格桑花[J]. 花木盆景(花卉园艺),2012(12):6.
[10]韦素敏. 格桑花在园林绿化中的应用[J]. 现代园艺,2016(4):129.
[11]韩荣荣,杨桂军,王刚. 高原格桑花染料敏化剂稳定性研究[J]. 硅酸盐通报,2014,33(10):2679 – 2682.
[12]徐凤翔,郑维列. 西藏野生花卉[M]. 北京:中国旅游出版社,1999.
[13]中国科学院中国植物志编辑委员会. 中国植物志[M]. 北京:科学出版社,1998.
[14]吴征镒. 西藏植物志:第 1 卷[M]. 北京:科学出版社,1983.
[15]李永康. 贵州植物志:第 3 卷[M]. 贵阳:贵州人民出版社,1990.
[16]华夫. 中国古代名物大典:下册[M]. 济南:济南出版社,1993.
[17]张宗显. 西藏民俗[M]. 兰州:甘肃人民出版社,2004.
[18]中国农业百科全书总编辑委员会农作物卷编辑委员会. 中国农业百科全书:观赏园艺卷[M]. 北京:农业出版社,1996.