(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 108710948 B
(45)授权公告日 2021.08.31

(21)申请号 201810378299.2
(22)申请日 2018.04.25
(65)同一申请的已公布的文献号
申请公布号 CN 108710948 A
(43)申请公布日 2018.10.26
(73)专利权人 佛山科学技术学院
地址 528000 广东省佛山市南海区狮山镇仙溪水库西路佛山科学技术学院
(72)发明人 易长安 朱珍 黄营 胡明 邓波
(74)专利代理机构 广州嘉权专利商标事务所有限公司 44205
代理人 王国标
(51)Int.Cl.
G06N 20/00(2019.01)
G06K 9/62(2006.01)
(56)对比文件
CN 105023024 A,2015.11.04
US 2006047655 A1,2006.03.02
US 2017270387 A1,2017.09.21
US 2015116493 A1,2015.04.30
CN 107944716 A,2018.04.20
CN 107169511 A,2017.09.15
CN 107563410 A,2018.01.09
Yihui Luo .ect.Clustering Ensemble for Unsupervised Feature Selection.《2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery》.2009,
刘江涛.距离度量学习中的类别不平衡问题研究.《中国优秀硕士学位论文全文数据库》.2017,
曹丽君.基于迁移学习和特征融合的人脸识别算法的研究.《中国优秀硕士学位论文全文数据库》.2017,
庄广安.基于字典学习的无监督迁移聚类及其在SAR图像分割中的应用.《中国优秀硕士学位论文全文数据库》.2013,
审查员 李娜

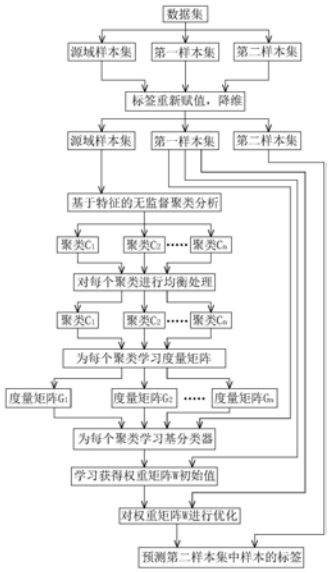
权利要求书2页 说明书6页 附图1页

(54)发明名称

一种基于聚类均衡和权重矩阵优化的迁移学习方法

(57)摘要

本发明公开了一种基于聚类均衡和权重矩阵优化的迁移学习方法,包括定义源域样本集和目标域样本集;对源域样本集以及目标域样本集样本的标签重新赋值;对源域样本集和目标域样本集中样本进行降维;对源域样本集中样本进行基于特征的无监督聚类分析;对每个聚类进行均衡处理;为每个聚类学习度量矩阵;根据聚类及度量矩阵,生成权重矩阵;对权重矩阵优化;利用权重矩阵预测目标域样本集中样本的标签。本发明通过无监督的聚类分析方法将源域样本集分为多个不同的聚类,使每个聚类具有相似的属性;同时基于各个聚类生成权重矩阵,并对其进行优化,更符合目标域样本集的实际情况,利用该权重矩阵对目标域样本集的标签进行预测,准确度更高。



CN 108710948 B

1.一种基于聚类均衡和权重矩阵优化的迁移学习方法,其特征在于,包括以下步骤:

步骤A.定义源域样本集D_S以及目标域样本集D_T,所述目标域样本集D_T分为两部分,分别为样本贴有标签的第一样本集D_{TL}以及样本没有贴标签的第二样本集D_{TU},所述第二样本集D_{TU}的样本数量远大于第一样本集D_{TL}的数量;

步骤B.对所述源域样本集D_S以及目标域样本集D_T中样本的标签进行重新赋值;

步骤C.对所述源域样本集D_S以及目标域样本集D_T中的样本进行降维操作;

步骤D.对所述源域样本集D_S中的样本进行基于特征的无监督聚类分析,生成若干个聚类;

步骤E.对每个聚类进行均衡处理;

步骤F.为每个聚类学习一个度量矩阵G;

步骤G.根据聚类及其度量矩阵G,以及第一样本集D_{TL},学习权重矩阵W的初始值;

步骤H.根据所述第一样本集D_{TL},对所述权重矩阵W进行优化;

步骤I.利用所述权重矩阵W预测第二样本集D_{TU}样本的标签;

所述步骤G包括以下步骤:

步骤G1.为每个聚类学习出一个基分类器$Model_i$,其中i表示第i个聚类;

步骤G2.设置所述基分类器$Model_i$的训练函数,所述训练函数如公式2所示;

$$Model_i = BaseLearner(C_i, Q_i, G_i) \quad \text{公式2}$$

其中C_i为第i个聚类,G_i为第i个聚类的度量矩阵,Q_i表示的是第一样本集D_{TL}中与第i个聚类最近的样本集合;

步骤G3.基于基分类器$Model_i$的训练函数,利用度量矩阵G_i,对聚类C_i和集合Q_i进行特征变换,并对特征进行归一化处理,完成对基分类器$Model_i$的训练;

步骤G4.利用函数$BaseLearnerPred(C_i, G_j, Model_j)$,求解所有$(C_i, G_j)$对第一样本集$D_{TL}$中样本的预测标签,其中$1 \leqslant i, j \leqslant n$,n为聚类的数量,各个$(C_i, G_j)$的预测准确率形成权重矩阵W的初始值$W_0$;

所述步骤H包括以下步骤:

步骤H1.根据所述权重矩阵W的初始值W_0,计算第一样本集合D_{TL}中样本的预测标签;

步骤H2.设置损失函数和正则项,所述损失函数如公式3所示;

$$norm(L_{pred} * w_t - L_{real}) \quad \text{公式3}$$

其中w_t是度量矩阵W第t列的值,是步骤H中需要优化的值,L_{pred}是经过权重矩阵W_0计算得到的预测标签,L_{real}是真实标签,所述正则项如公式4所示;

$$norm(w_t - b) \quad \text{公式4}$$

其中b是权重矩阵W_0第t列的值;

步骤H3.利用第一样本集合D_{TL}中样本,通过公式5求得w_t的最优值;

$$minimize(lamda * norm(w_t - b) + norm(L_{pred} * w_t - L_{real})) \quad \text{公式5}$$

其中lamda表示平衡因子。

2.根据权利要求1所述的一种基于聚类均衡和权重矩阵优化的迁移学习方法,其特征在于:所述步骤C中利用主成分分析法对所述源域样本集D_S以及目标域样本集D_T中的样本进行降维操作。

3.根据权利要求2所述的一种基于聚类均衡和权重矩阵优化的迁移学习方法,其特征

在于,所述步骤F包括以下步骤:

步骤F1.针对每个聚类,将聚类中的样本顺序随机化;

步骤F2.设置收敛条件,将度量矩阵G初始化为单位矩阵;

步骤F3.设置求解度量矩阵G的目标函数;

步骤F4.求解所述度量矩阵G的目标函数,直到符合收敛条件。

一种基于聚类均衡和权重矩阵优化的迁移学习方法

技术领域

[0001] 本发明涉及智能识别技术领域，更具体地说涉及一种迁移学习方法。

背景技术

[0002] 所述的迁移学习，对于人类来说，就是类似于举一反三的意思，利用已有的知识来学习新的知识，解决新的问题；而在机器学习来说，迁移学习在通俗意义上来讲就是能让现有的模型算法稍加调整即可应用于一个相似的领域和功能的一项技术。

[0003] 现有的迁移学习主要包括三种类型，基于特征的迁移学习、基于实例的迁移学习以及基于度量的迁移学习。其中，基于特征的迁移学习和基于实例的迁移学习是使用欧氏距离来衡量样本之间的距离，而欧氏距离不能够反应出样本的不同维度之间的关联。基于度量的迁移学习方法虽然考虑了样本的不同维度之间的关联，但这种学习方法和前两种类型一样，样本的类型完全取决于标签的种类，从而忽视了样本特征的本质属性，也就是忽略了不同标签的样本特征之间也可能存在的某些关联。

发明内容

[0004] 针对上述问题，本发明考虑了样本特征所隐含的本质属性，也就是考虑了不同标签的样本特征之间也可能存在的某种关联，提供一种基于聚类均衡和权重矩阵优化的迁移学习方法，能够将源域知识更好地迁移到目标域。

[0005] 本发明解决其技术问题的解决方案是：

[0006] 一种基于聚类均衡和权重矩阵优化的迁移学习方法，包括以下步骤：

[0007] 步骤A.定义源域样本集D_S以及目标域样本集D_T，所述目标域样本集D_T分为两部分，分别为样本贴有标签的第一样本集D_{TL}以及样本没有贴标签的第二样本集D_{TU}，所述第二样本集D_{TU}的样本数量远大于第一样本集D_{TL}的数量；

[0008] 步骤B.对所述源域样本集D_S以及目标域样本集D_T中样本的标签进行重新赋值；

[0009] 步骤C.对所述源域样本集D_S以及目标域样本集D_T中的样本进行降维操作；

[0010] 步骤D.对所述源域样本集D_S中的样本进行基于特征的无监督聚类分析，生成若干个聚类；

[0011] 步骤E.对每个所述聚类进行均衡处理；

[0012] 步骤F.为每个所述聚类学习一个度量矩阵G；

[0013] 步骤G.根据所述聚类及其度量矩阵G，以及第一样本集D_{TL}，学习权重矩阵W的初始值；

[0014] 步骤H.根据所述第一样本集D_{TL}，对所述权重矩阵W进行优化；

[0015] 步骤I.利用所述权重矩阵W预测第二样本集D_{TU}的样本的标签。

[0016] 作为上述技术方案的进一步改进，所述步骤C中利用主成分分析法对所述源域样本集D_S以及目标域样本集D_T中的样本进行降维操作。除此以外还可以采用特征选择法进行降维操作。

[0017] 作为上述技术方案的进一步改进，所述步骤F包括以下步骤：

[0018] 步骤F1.针对每个所述聚类，将聚类中的样本顺序随机化；

[0019] 步骤F2.设置收敛条件，将度量矩阵G初始化为单位矩阵；

[0020] 步骤F3.设置求解度量矩阵G的目标函数，记为公式1；

[0021] $$\left\{\begin{array}{l}\min\limits_{G}\int p(x;G_0)\log\dfrac{p(x;G_0)}{p(x;G)}dx \\ \text{subject to}\ \ d_{G(x_i,x_j)}\le\alpha\ \ (i,j)\in S \\ \qquad\qquad d_{G(x_i,x_j)}\ge\beta\ \ (i,j)\in D\end{array}\right\}$$ 公式1

[0022] 其中G_0表示单位矩阵，x_i和x_j是聚类中的样本，S表示x_i和x_j同类，D表示x_i和x_j是不同类，所述α和β分别代表第一阈值和第二阈值；

[0023] 步骤F4.求解所述度量矩阵G的目标函数，直到符合收敛条件。

[0024] 作为上述技术方案的进一步改进，所述步骤G具体包括以下步骤：

[0025] 步骤G1.为每个聚类学习出一个基分类器$Model_i$，其中i表示第i个聚类；

[0026] 步骤G2.设置所述基分类器$Model_i$的训练函数，所述训练函数如公式2所示；

[0027] $Model_i=BaseLearner(C_i,Q_i,G_i)$ 公式2

[0028] 其中C_i为第i个聚类，G_i为第i个聚类的度量矩阵，Q_i表示的是第一样本集D_{TL}中与第i个聚类最近的样本集合；

[0029] 步骤G3.基于基分类器$Model_i$的训练函数，利用度量矩阵G_i，对聚类C_i和集合Q_i进行特征变换，并对特征进行归一化处理，完成对基分类器$Model_i$的训练；

[0030] 步骤G4.利用函数$BaseLearnerPred(C_i,G_j,Model_j)$，求解所有$(C_i,G_j)$对第一样本集$D_{TL}$中样本的预测标签，其中$1\le i,j\le n$，n为聚类的数量，各个$(C_i,G_j)$的预测准确率形成权重矩阵W的初始值$W_0$。

[0031] 作为上述技术方案的进一步改进，所述步骤H具体包括以下步骤：

[0032] 步骤H1.根据所述权重矩阵W的初始值W_0，计算第一样本集合D_{TL}中样本的预测标签；

[0033] 步骤H2.设置损失函数和正则项，所述损失函数如公式3所示；

[0034] $norm(L_{pred}*w_t-L_{real})$ 公式3

[0035] 其中w_t是度量矩阵W第t列的值，是步骤H中需要优化的值，L_{pred}是经过权重矩阵W_0计算得到的预测标签，L_{real}是真实标签，所述正则项如公式4所示；

[0036] $norm(w_t-b)$ 公式4

[0037] 其中b是权重矩阵W_0第t列的值；

[0038] 步骤H3.利用第一样本集合D_{TL}中样本，通过公式5求得w_t的最优值；

[0039] $minimize(lamda*norm(w_t-b)+norm(L_{pred}*w_t-L_{real}))$ 公式5

[0040] 其中lamda表示平衡因子。

[0041] 本发明的有益效果是：本发明通过基于特征的无监督聚类分析方法将源域样本集D_S分为多个不同的聚类，使每个聚类具有相似的属性；同时基于各个聚类生成权重矩阵，并对其进行优化，更符合目标域样本集的实际情况，利用该权重矩阵对目标域样本集的第二

样本集D_{TU}的样本标签进行预测，预测效果更好。

附图说明

[0042] 为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单说明。显然，所描述的附图只是本发明的一部分实施例，而不是全部实施例，本领域的技术人员在不付出创造性劳动的前提下，还可以根据这些附图获得其他设计方案和附图。

[0043] 图1是本发明的方法流程示意图。

具体实施方式

[0044] 以下将结合实施例和附图对本发明的构思、具体结构及产生的技术效果进行清楚、完整的描述，以充分地理解本发明的目的、特征和效果。显然，所描述的实施例只是本发明的一部分实施例，而不是全部实施例，基于本发明的实施例，本领域的技术人员在不付出创造性劳动的前提下所获得的其他实施例，均属于本发明保护的范围。

[0045] 参照图1，本发明创造公开了一种基于聚类均衡和权重矩阵优化的迁移学习方法，该迁移学习方法可应用于智能机器人场景识别、药品识别以及智能监控等领域。

[0046] 一种基于聚类均衡和权重矩阵优化的迁移学习方法，包括以下步骤：

[0047] 步骤A.定义源域样本集D_S以及目标域样本集D_T，所述目标域样本集D_T分为两部分，分别为样本贴有标签的第一样本集D_{TL}以及样本没有贴标签的第二样本集D_{TU}，所述第二样本集D_{TU}的样本数量远大于第一样本集D_{TL}的数量；

[0048] 步骤B.对所述源域样本集D_S以及目标域样本集D_T中的样本标签进行重新赋值；

[0049] 步骤C.对所述源域样本集D_S以及目标域样本集D_T中的样本进行降维操作；

[0050] 步骤D.对所述源域样本集D_S中的样本进行基于特征的无监督聚类分析，生成若干个聚类；

[0051] 步骤E.对每个所述聚类进行均衡处理；

[0052] 步骤F.为每个所述聚类学习一个度量矩阵G；

[0053] 步骤G.根据所述聚类及其度量矩阵G，以及第一样本集D_{TL}，学习权重矩阵W的初始值；

[0054] 步骤H.根据所述第一样本集D_{TL}，对所述权重矩阵W进行优化；

[0055] 步骤I.利用所述权重矩阵W预测第二样本集D_{TU}的样本的标签。

[0056] 具体地，本发明通过基于特征的无监督聚类分析方法将源域样本集D_S分为多个不同的聚类，使每个聚类的样本特征之间具有某种关联；同时基于各个聚类生成权重矩阵，并对其进行优化，更符合目标域样本集的实际情况，利用该权重矩阵对目标域样本集中未知标签的样本的标签进行预测，预测效果更好。

[0057] 以下对所述迁移学习方法中的各个步骤进行详细说明。

[0058] 步骤A中，首先定义源域样本集D_S以及目标域样本集D_T，其中所述目标域样本集D_T分为两部分，分别为样本贴有标签的第一样本集D_{TL}以及样本没有贴标签的第二样本集D_{TU}，所述第二样本集D_{TU}的样本数量远大于第一样本集D_{TL}的数量，其中通常第一样本集D_{TL}数量是目标域样本集D_T数量的百分之五。实际应用中，本方法所述源域样本集D_S以及目标域样本

集D_T中的样本是服从不同的数据分布，但是第一样本集D_{TL}和第二样本集D_{TU}中样本是服从相同的数据分布，因此当得到第一样本集D_{TL}的预测模型，即可用该预测模型预测第二样本集D_{TU}中的样本。简单的说，本方法是利用源域样本集D_S和第一样本集D_{TL}生成初始的预测模型，并用第一样本集D_{TL}对预测模型进行优化，利用优化之后的预测模型对第二样本集D_{TU}中样本进行预测。

[0059] 步骤B中，对所述源域样本集D_S以及目标域样本集D_T中的样本标签进行重新赋值，本步骤中重新赋值前，若两个样本的标签是相同的，则重新赋值后这两个样本的标签依旧相同。重新赋值之后的标签是从1到n的整数值。本发明中对各个样本标签重新赋值，目的在于便于在后续步骤中学习、使用具有多分类功能的基分类器。

[0060] 步骤C中，需要对所述源域样本集D_S以及目标域样本集D_T中的样本进行降维操作，本发明具体实施例中，具体是利用主成分分析法或者特征选择法对所述源域样本集D_S以及目标域样本集D_T中的样本进行降维操作，通过以上两种降维函数，可将样本数据从几万维，甚至数百万维降低到几十维，同时保留样本的主要特性。

[0061] 步骤D中，对所述源域样本集D_S中的样本进行基于特征的无监督聚类分析，生成若干个聚类，其中，所生成聚类的数量可视实际情况而定。

[0062] 步骤E中，对每个所述聚类进行均衡处理。本发明具体实施例中，所述步骤E的具体操作如下：在某个聚类中，假设标签k对应的样本数量最多，记为S_k，对于任意的其他标签y，该聚类中标签为y的样本数量为S_y，该源域样本集D_S中标签为y的样本数量为d_y，从源域样本集D_S中随机抽取$\min\{(S_k-S_y),d_y\}$个标签为y的样本，添加到当前聚类中。进行均衡处理后，同一聚类中的某些样本可能会重复出现，进行均衡处理的目的在于防止某个标签的样本数量过少。

[0063] 步骤F中，为每个聚类学习一个度量矩阵G，其中本发明具体实施例中，步骤F具体包括以下步骤：

[0064] 步骤F1.针对每个所述聚类，将聚类中的样本顺序随机化，使得在后续步骤中被选中的样本更具有随机性；

[0065] 步骤F2.设置收敛条件，将度量矩阵G初始化为单位矩阵，本实施例中，所述收敛条件可以有两种，第一是以迭代次数超过某个阈值作为收敛条件，第二是以度量矩阵的变化幅度小于某个阈值作为收敛条件。本发明实施例优先采用的是以度量矩阵的变化幅度小于某个阈值作为收敛条件；

[0066] 步骤F3.设置求解度量矩阵G的目标函数，记为公式1；

[0067]
$$\left\{\begin{array}{l}\min\limits_{G}\int p(x;G_0)\log\dfrac{p(x;G_0)}{p(x;G)}dx \\ \text{subject to}\ \ d_{G(x_i,x_j)}\le\alpha\ \ (i,j)\in S \\ \qquad\qquad d_{G(x_i,x_j)}\ge\beta\ \ (i,j)\in D\end{array}\right\}\qquad \text{公式 1}$$

[0068] 其中G_0表示单位矩阵，x_i和x_j是聚类中的样本，S表示x_i和x_j同类，D表示x_i和x_j是不同类，所述α和β分别代表第一阈值和第二阈值；其中所述度量矩阵的作用是将样本从一个空间转换到另一个空间，在新的空间里，任意两个样本之间的距离都用马氏距离表示，如果

两个样本之间的距离小于第一阀值，那么它们就是相似的，如果它们之间的距离大于第二阀值，那么它们就是不相似的，本实施例利用如下公式表示马氏距离，$d_{G(x_i,x_j)}=\sqrt{(x_i,x_j)^T G(x_i,x_j)}$；本发明具体实施例中设置第一阈值和第二阈值的过程如下，从某个聚类中，随机选择若干个样本对，将它们之间的距离按从小到大的顺序排列，前5％对应的距离值就是第一阈值，前95％对应的距离值就是第二阈值(通常选取5％、95％为临界点)。例如，如果这100个(假设为100个)距离值都不重复，并且从1到100均匀分布，那么第一阈值为5，第二阈值为95；

[0069] 步骤F4.求解所述度量矩阵G的目标函数，直到符合收敛条件，获得度量矩阵G。通过公式1，若经过降维后样本维度是50，则求得的度量矩阵G是50*50的矩阵。

[0070] 步骤G中，根据所述聚类及其度量矩阵G，以及第一样本集D_{TL}，生成权重矩阵W并学习权重矩阵W的初始值，本发明具体实施例中，步骤G具体包括以下步骤：

[0071] 步骤G1.为每个聚类学习出一个基分类器$Model_i$，其中i表示第i个聚类；

[0072] 步骤G2.设置所述基分类器$Model_i$的训练函数如公式2所示；

[0073] $Model_i=BaseLearner(C_i,Q_i,G_i)$ 公式2

[0074] 其中C_i为第i个聚类，G_i为第i个聚类的度量矩阵，Q_i表示第一样本集D_{TL}中与聚类i的距离最近的样本集合；其中集合Q_i通过以下方法求得，首先计算每个聚类的聚类中心，针对第一样本集D_{TL}，计算第一样本集D_{TL}每个样本到各个聚类中心的欧氏距离，如果某个样本离聚类C_i最近，则将其存入集合Q_i中，集合Q_i的内容最初为空；

[0075] 步骤G3.基于基分类器$Model_i$的训练函数，利用度量矩阵G_i，对聚类C_i和集合Q_i进行特征变换，并对特征进行归一化处理，完成对基分类器$Model_i$的训练；

[0076] 步骤G4.利用函数$BaseLearnerPred(C_i,G_j,Model_j)$，求解所有$(C_i,G_j)$对第一样本集$D_{TL}$中样本的预测标签，其中$1\leq i,j\leq n$，n为聚类的数量，各个$(C_i,G_j)$的预测准确率形成权重矩阵W的初始值$W_0$。

[0077] 步骤G中，各个聚类C_i以及所有度量矩阵G_j之间形成表1所示关系。

[0078]

聚类 \ 度量矩阵	G_1	G_2	……	G_n
C_1	W_{11}	W_{12}	……	W_{1n}

[0079]

C_2	W_{21}	W_{22}	……	W_{2n}
……	……	……	……	……
C_n	W_{n1}	W_{n2}	……	W_{nn}

[0080] 表1

[0081] 其中表1中，W_{11}、W_{12}……W_{nn}组成权重矩阵W。

[0082] 步骤H中，需要对所述权重矩阵W进行优化，目的在于使第一样本集D_{TL}中样本的预测标签与真实标签差异最小，本发明具体实施例中，步骤H具体包括以下步骤：

[0083] 步骤H1.根据所述权重矩阵W的初始值W_0，计算第一样本集合D_{TL}中样本的预测标签；

[0084] 步骤H2.设置损失函数和正则项，所述损失函数如公式3所示；

[0085] $norm(L_{pred}*w_t-L_{real})$ 公式3

[0086] 其中w_t是度量矩阵W第t列的值，是步骤H中需要优化的值，L_{pred}是经过权重矩阵W_0计算得到的预测标签，L_{real}是真实标签，所述正则项如公式4所示；

[0087] $norm(w_t-b)$ 公式4

[0088] 其中b是权重矩阵W_0第t列的值；

[0089] 步骤H3.利用第一样本集合D_{TL}中样本，通过公式5求得w_t的最优值；

[0090] $minimize(lamda*norm(w_t-b)+norm(L_{pred}*w_t-L_{real}))$ 公式5

[0091] 其中lamda表示平衡因子。

[0092] 步骤I中，用所述权重矩阵W预测第二样本集D_{TU}中样本的标签。具体地，在步骤H中，利用第一样本集D_{TL}对权重矩阵进行优化，而由于第一样本集D_{TL}和第二样本集D_{TU}样本数据分布一致，因此优化后的权重矩阵同样适用于预测第二样本集D_{TU}样本的标签。所述第二样本集D_{TU}的每个样本必然和某一个聚类的距离最近。此处以聚类C_i为例子进行说明，假设第一样本集D_{TL}里与聚类C_i最近的样本集为R_i，首先采用函数BaseLearnerPred预测样本集R_i的标签，假设(C_i,G_j)对样本R_i的预测标签为$pred(R_i)$，那么对于同一个聚类C_i，样本集R_i中的每个样本都可以通过如下公式进行预测，$pred(R_i)=pred(R_i)+W(C_i,G_j)$，接着，样本集$R_i$里的每个样本都使用max函数(以MATLAB软件为例)求解使得$pred(R_i)$取得最大值的标签序号，最终得到相应的预测值$Final(R_i)$，即$Final(R_i)=max(pred(R_i))$。因为之前进行了重赋值操作，所以现在最后只需将预测值还原为真实值即可。

[0093] 以上对本发明的较佳实施方式进行了具体说明，但本发明创造并不限于所述实施例，熟悉本领域的技术人员在不违背本发明精神的前提下还可作出种种的等同变型或替换，这些等同的变型或替换均包含在本申请权利要求所限定的范围内。

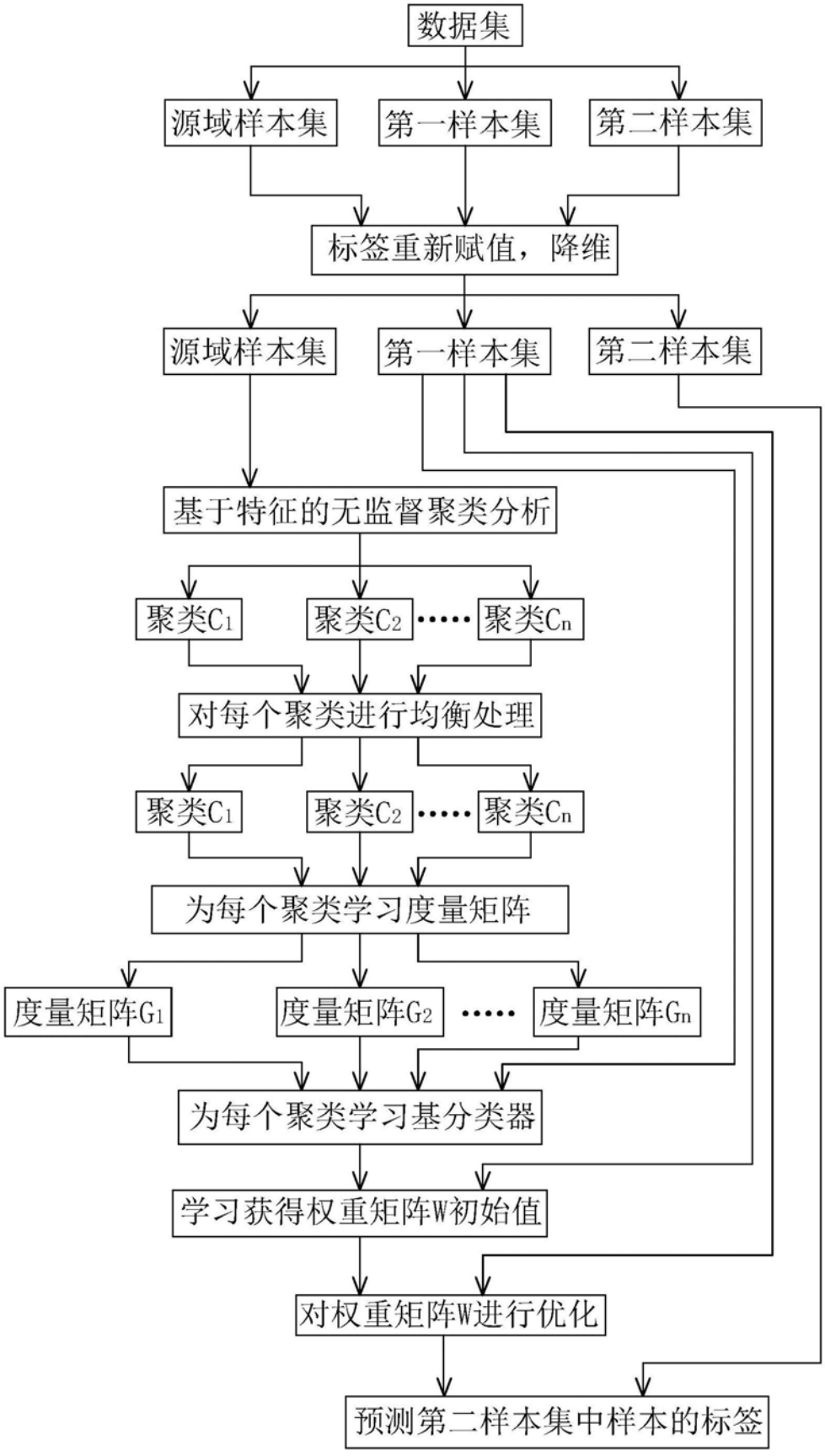

数据集
源域样本集
第一样本集
第二样本集
标签重新赋值，降维
源域样本集
第一样本集
第二样本集
基于特征的无监督聚类分析
聚类C1
聚类C2
聚类Cn
对每个聚类进行均衡处理
聚类C1
聚类C2
聚类Cn
为每个聚类学习度量矩阵
度量矩阵G1
度量矩阵G2
度量矩阵Gn
为每个聚类学习基分类器
学习获得权重矩阵W初始值
对权重矩阵W进行优化
预测第二样本集中样本的标签

图1