



(12)发明专利申请

(10)申请公布号 CN 108596737 A

(43)申请公布日 2018.09.28

(21)申请号 201810426131.4

(22)申请日 2018.05.07

(71)申请人 山东师范大学

地址 250014 山东省济南市历下区文化东路88号

(72)发明人 王红 狄瑞彤 房有丽 周莹
王露潼 刘海燕 王倩 宋永强

(74)专利代理机构 济南圣达知识产权代理有限公司 37221

代理人 杨哲

(51)Int.Cl.

G06Q 30/06(2012.01)

G06F 17/30(2006.01)

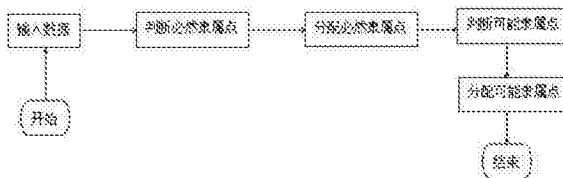
权利要求书2页 说明书9页 附图3页

(54)发明名称

基于电子商务评论数据的非聚类中心节点分配方法及装置

(57)摘要

本发明公开了一种基于电子商务评论数据的非聚类中心节点分配方法及装置,该方法包括:接收电子商务评论数据集,获取电子商务评论中心点集、近邻数和距离矩阵;建立与电子商务评论中心点集、近邻数和距离矩阵相对应的第一样本,判断所述第一样本中的电子商务评论为必然隶属点或可能隶属点;根据两步分配策略进行非聚类中心节点分配,所述两步分配策略为以必然隶属点和可能隶属点为度量,按照必然隶属点或可能隶属点准则进行样本点分配,得到真实和虚假两个不同类型的电子商务评论类簇。



1. 一种基于电子商务评论数据的非聚类中心节点分配方法,其特征在于,该方法包括:
接收电子商务评论数据集,获取电子商务评论中心点集、近邻数和距离矩阵;
建立与电子商务评论中心点集、近邻数和距离矩阵相对应的第一样本,判断所述第一样本中的电子商务评论为必然隶属点或可能隶属点;
根据两步分配策略进行非聚类中心节点分配,所述两步分配策略为以必然隶属点和可能隶属点为度量,按照必然隶属点或可能隶属点准则进行样本点分配,得到真实和虚假两个不同类型的电子商务评论类簇。
2. 如权利要求1所述的方法,其特征在于,在本方法中,判断所述第一样本中的电子商务评论为必然隶属点的方法为:
对所述基于中心点集,近邻数和距离矩阵的第一样本进行SNN的计算;
判断计算结构是否满足必然隶属点定义。
3. 如权利要求2所述的方法,其特征在于,所述必然隶属点定义为:
已被分配所述类簇的电子商务评论与未分配所述类簇的电子商务评论的共同邻居数量至少为设定近邻数量的一半,则未分配所述类簇的电子商务评论为已被分配所述类簇的电子商务评论所在簇的必然隶属点。
4. 如权利要求1所述的方法,其特征在于,判断所述第一样本中的电子商务评论为可能隶属点的方法为:
对所述基于中心点集,近邻数和距离矩阵的第一样本进行SNN的计算;
判断计算结构是否满足可能隶属点定义。
5. 如权利要求4所述的方法,其特征在于,所述可能隶属点定义为:
已被分配所述类簇的电子商务评论与未分配所述类簇的电子商务评论的共同邻居数量大于零小于设定近邻数量的一半,则未分配所述类簇的电子商务评论为已被分配所述类簇的电子商务评论所在簇的可能隶属点。
6. 如权利要求1所述的方法,其特征在于,所述必然隶属点准则进行样本点分配的具体方法为:
利用广度优先搜索,从聚类中心开始,依次考察其K近邻点;
若该点与考察的点之间存在较多共享近邻,则考察的点必然隶属于自身所在的簇,将其划分到所在簇,并加入队列,继续考察该点的近邻点。
7. 如权利要求1所述的方法,其特征在于,所述可能隶属点准则进行样本点分配的具体方法为:
遍历所有未分配点,统计其近邻中隶属于每个簇的点的数量,构成分配矩阵;
寻找分配矩阵中的最大值,再次遍历所有未分配点;
若其近邻中隶属于某一簇的点数量与前述最大值相同,则将该点分配至该簇,遍历过所有未分配点后更新识别矩阵,再次重复该过程,完成剩余点的分配。
8. 一种计算机可读存储介质,其中存储有多条指令,其特征在于,所述指令适于由终端设备的处理器加载并执行根据权利要求1-7中任一项所述的方法。
9. 一种终端设备,包括处理器和计算机可读存储介质,处理器用于实现各指令;计算机可读存储介质用于存储多条指令,其特征在于,所述指令用于执行根据权利要求1-7中任一项所述的方法。

10. 一种基于电子商务评论数据的聚类方法,其特征在于,一种基于电子商务评论数据的聚类方法,采用如权利要求1-7任一项所述的方法进行非聚类中心节点分配,得到真实和虚假两个不同类型的电子商务评论类簇进行聚类。

基于电子商务评论数据的非聚类中心节点分配方法及装置

技术领域

[0001] 本发明属于电子商务评论挖掘的技术领域,尤其是涉及一种基于电子商务评论数据的非聚类中心节点分配方法及装置。

背景技术

[0002] 在电子商务评论领域中,聚类分析作为研究其他问题的重要基础,在电子商务评论挖掘中扮演着重要角色,例如可以通过聚类分析得到两个不同类别的评论,进而识别出真实评论和虚假评论,以给潜在的客户提供更加可信的评论参考。

[0003] 在电子商务评论聚类中其非聚类中心的分配具有重要意义,是得到好的电子商务评论聚类效果的重要的环节。目前,分配非聚类中心的方法有基于密度的,定义局部密度和距离测度,将非聚类中心的样本点分配到密度比它高且距离它最近的样本点所在的簇中,进行一步分配。有的方法是基于距离的,在聚类中心确定的情况下,将非聚类中心的样本点划分到距离它最近的聚类中心所在的簇中。

[0004] 然而,电子商务评论数据具有不同类别数据在空间分布上非相互独立分离、数据空间密度分布不均匀,和数据的维度较高,即具有交叉缠绕、变密度和高维的特点。目前,基于现有分配非聚类中心的方法,为多尺度、交叉缠绕、变密度和高维的电子商务评论数据集进行正确分配非聚类中心的方法尚未出现。

[0005] 综上所述,在现有技术中针对电子商务评论中如何有效分配非聚类中心,从而快速、高效的对电子商务评论进行分类的问题,尚缺乏有效的解决方案。

发明内容

[0006] 针对现有技术中存在的不足,解决现有技术中针对电子商务评论中如何有效分配非聚类中心,从而快速、高效的对电子商务评论进行分类的问题,本发明提供了一种基于电子商务评论数据的非聚类中心节点分配方法及装置,基于两阶段非中心节点分配策略,使得能够快速、正确的进行多尺度、交叉缠绕、变密度和高维的电子商务评论数据集的非聚类中心的分配,提高聚类效果。

[0007] 本发明的第一目的是提供一种基于电子商务评论数据的非聚类中心节点分配方法。

[0008] 为了实现上述目的,本发明采用如下一种技术方案:

[0009] 一种基于电子商务评论数据的非聚类中心节点分配方法,该方法包括:

[0010] 接收电子商务评论数据集,获取电子商务评论中心点集、近邻数和距离矩阵;

[0011] 建立与电子商务评论中心点集、近邻数和距离矩阵相对应的第一样本,判断所述第一样本中的电子商务评论为必然隶属点或可能隶属点;

[0012] 根据两步分配策略进行非聚类中心节点分配,所述两步分配策略为以必然隶属点和可能隶属点为度量,按照必然隶属点或可能隶属点准则进行样本点分配,得到真实和虚假两个不同类型的电子商务评论类簇。

[0013] 作为进一步的优选方案,在本方法中,判断所述第一样本中的电子商务评论为必然隶属点的方法为:

[0014] 对所述基于中心点集,近邻数和距离矩阵的第一样本进行SNN的计算;

[0015] 判断计算结构是否满足必然隶属点定义。

[0016] 作为进一步的优选方案,所述必然隶属点定义为:

[0017] 已被分配所述类簇的电子商务评论与未分配所述类簇的电子商务评论的共同邻居数量至少为设定近邻数量的一半,则未分配所述类簇的电子商务评论为已被分配所述类簇的电子商务评论所在簇的必然隶属点。

[0018] 作为进一步的优选方案,判断所述第一样本中的电子商务评论为可能隶属点的方法为:

[0019] 对所述基于中心点集,近邻数和距离矩阵的第一样本进行SNN的计算;

[0020] 判断计算结构是否满足可能隶属点定义。

[0021] 作为进一步的优选方案,所述可能隶属点定义为:

[0022] 已被分配所述类簇的电子商务评论与未分配所述类簇的电子商务评论的共同邻居数量大于零小于设定近邻数量的一半,则未分配所述类簇的电子商务评论为已被分配所述类簇的电子商务评论所在簇的可能隶属点。

[0023] 作为进一步的优选方案,所述必然隶属点准则进行样本点分配的具体方法为:

[0024] 利用广度优先搜索,从聚类中心开始,依次考察其K近邻点;

[0025] 若该点与考察的点之间存在较多共享近邻,则考察的点必然隶属于自身所在的簇,将其划分到所在簇,并加入队列,继续考察该点的近邻点。

[0026] 作为进一步的优选方案,所述可能隶属点准则进行样本点分配的具体方法为:

[0027] 遍历所有未分配点,统计其近邻中隶属于每个簇的点的数量,构成分配矩阵;

[0028] 寻找分配矩阵中的最大值,再次遍历所有未分配点;

[0029] 若其近邻中隶属于某一簇的点数量与前述最大值相同,则将该点分配至该簇,遍历过所有未分配点后更新识别矩阵,再次重复该过程,完成剩余点的分配。

[0030] 本发明的第二目的是提供一种计算机可读存储介质。

[0031] 为了实现上述目的,本发明采用如下一种技术方案:

[0032] 一种计算机可读存储介质,其中存储有多条指令,所述指令适于由终端设备的处理器加载并执行以下处理:

[0033] 接收电子商务评论数据集,获取电子商务评论中心点集、近邻数和距离矩阵;

[0034] 建立与电子商务评论中心点集、近邻数和距离矩阵相对应的第一样本,判断所述第一样本中的电子商务评论为必然隶属点或可能隶属点;

[0035] 根据两步分配策略进行非聚类中心节点分配,所述两步分配策略为以必然隶属点和可能隶属点为度量,按照必然隶属点或可能隶属点准则进行样本点分配,得到真实和虚假两个不同类型的电子商务评论类簇。

[0036] 本发明的第三目的是提供一种终端设备。

[0037] 为了实现上述目的,本发明采用如下一种技术方案:

[0038] 一种终端设备,包括处理器和计算机可读存储介质,处理器用于实现各指令;计算机可读存储介质用于存储多条指令,所述指令适于由处理器加载并执行以下处理:

- [0039] 接收电子商务评论数据集,获取电子商务评论中心点集、近邻数和距离矩阵;
- [0040] 建立与电子商务评论中心点集、近邻数和距离矩阵相对应的第一样本,判断所述第一样本中的电子商务评论为必然隶属点或可能隶属点;
- [0041] 根据两步分配策略进行非聚类中心节点分配,所述两步分配策略为以必然隶属点和可能隶属点为度量,按照必然隶属点或可能隶属点准则进行样本点分配,得到真实和虚假两个不同类型的电子商务评论类簇。
- [0042] 本发明的第四目的是提供一种基于电子商务评论数据的聚类方法。
- [0043] 为了实现上述目的,本发明采用如下一种技术方案:
- [0044] 根据本发明的第四目的,本发明还提供了一种基于电子商务评论数据的聚类方法,采用所述一种基于电子商务评论数据的非聚类中心节点分配方法进行非聚类中心节点分配,得到真实和虚假两个不同类型的电子商务评论类簇进行聚类。
- [0045] 本发明的有益效果:
- [0046] 本发明所述的一种基于电子商务评论数据的非聚类中心节点分配方法及装置,以必然隶属点和可能隶属点为度量,基于两阶段非中心电子商务评论分配策略,使得能够快速、正确的进行多尺度、交叉缠绕、变密度和高维的复电子商务评论数据集的非聚类中心的分配,得到真实和虚假两个不同类型的电子商务评论类簇,进而进行电子商务评论数据的聚类,提高聚类效果。

附图说明

- [0047] 构成本申请的一部分的说明书附图用来提供对本申请的进一步理解,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。
- [0048] 图1是本发明的基于电子商务评论数据的非聚类中心节点分配方法流程图;
- [0049] 图2是本发明的实施例1的信息流转示意图;
- [0050] 图3是本发明的必然隶属点示例图;
- [0051] 图4是本发明的对电子商务评论数据集聚类结果图;
- [0052] 图5是本发明的分配策略示例图。

具体实施方式:

- [0053] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。
- [0054] 应该指出,以下详细说明都是例示性的,旨在对本申请提供进一步的说明。除非另有指明,本实施例使用的所有技术和科学术语具有与本申请所属技术领域的普通技术人员通常理解的含义。
- [0055] 需要注意的是,这里所使用的术语仅是为了描述具体实施方式,而非意图限制根据本申请的示例性实施方式。如在这里所使用的,除非上下文另外明确指出,否则单数形式也意图包括复数形式,此外,还应当理解的是,当在本说明书中使用术语“包含”和/或“包括”时,其指明存在特征、步骤、操作、器件、组件和/或它们的组合。

[0056] 需要注意的是,附图中的流程图和框图示出了根据本公开的各种实施例的方法和系统的可能实现的体系架构、功能和操作。应当注意,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,所述模块、程序段、或代码的一部分可以包括一个或多个用于实现各个实施例中所规定的逻辑功能的可执行指令。也应当注意,在有些作为备选的实现中,方框中所标注的功能也可以按照不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,或者它们有时也可以按照相反的顺序执行,这取决于所涉及的功能。同样应当注意的是,流程图和/或框图中的每个方框、以及流程图和/或框图中的方框的组合,可以使用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以使用专用硬件与计算机指令的组合来实现。

[0057] 在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合下面结合附图与实施例对本发明作进一步说明。

[0058] 实施例1:

[0059] 本实施例1的目的是提供一种基于电子商务评论数据的非聚类中心节点分配方法。

[0060] 为了实现上述目的,本发明采用如下一种技术方案:

[0061] 如图1-2所示,

[0062] 一种基于电子商务评论数据的非聚类中心节点分配方法,该方法包括:

[0063] 步骤(1):接收电子商务评论数据集,获取电子商务评论中心点集、近邻数和距离矩阵;

[0064] 步骤(2):建立与电子商务评论中心点集、近邻数和距离矩阵相对应的第一样本,判断所述第一样本中的电子商务评论为必然隶属点或可能隶属点;

[0065] 步骤(3):根据两步分配策略进行非聚类中心节点分配,所述两步分配策略为以必然隶属点和可能隶属点为度量,按照必然隶属点或可能隶属点准则进行样本点分配,得到真实和虚假两个不同类型的电子商务评论类簇。

[0066] 在本实施例的步骤(1)中,采用电子商务评论数据集总共有278条,每条数据有10个特征,共被分成了两类,分别为真实评论和虚假评论。表1所列的是针对电子商务评论数据所提出的10个虚假评论特征及其所表达的意义,通过本实施例中的分配策略进行聚类后,共得到两个簇,一个簇包含75条电子商务评论数据,另一个簇包含203条电子商务评论数据。

[0067] 表1

[0068]

特征	意义
评分	水军评分时一般给出 1 分或 5 分。
评论者信誉	评论者的信誉越低，虚假评论可能性越大
初评与追评 一致性	虚假评论者基本上会同时完成初评与追评
买家晒图	雇佣水军进行虚假评论时会要求他们晒商品图片
评论文本长 度	“水军”为了吸引浏览者的注意，一般会发比较丰富的内容，从而评论文本较长
点赞数	当点赞数少时，说明同意该评论的少，虚假的可能性越大。
转折词	水军评论时只会给出正面评论或者只会给出负面评论，不会有转折词的存在。
评论情感强 度	水军进行评论时，急于向用户推荐或者使用户对这个产品产生抵触情绪，往往会使用较多具有鲜明的感情色彩的词
评论与评分 一致性	特征刻画的是一条评论所表述的内容与此评论的评分的相符程度
评论与其他 评论连贯性	在电子商务评论数据集中同意一条评论的属性所占比例

[0069] 本实施例步骤 (2) 的具体步骤包括：

[0070] 步骤 (2-1)：建立与电子商务评论中心点集、近邻数和距离矩阵相对应的第一样本；

[0071] 步骤 (2-2)：判断所述第一样本中的电子商务评论是否为必然隶属点；

[0072] 步骤 (2-3)：判断所述第一样本中的电子商务评论是否为可能隶属点。

[0073] 在本实施例的步骤 (2-2) 中，判断所述第一样本中的电子商务评论为必然隶属点的方法为：

[0074] 步骤 (2-2-1)：对所述基于中心点集，近邻数和距离矩阵的第一样本进行 SNN 的计算；

[0075] 步骤 (2-2-2)：判断计算结构是否满足必然隶属点定义。

[0076] 所述必然隶属点定义为：

[0077] 已被分配所述类簇的电子商务评论与未分配所述类簇的电子商务评论的共同邻居数量至少为设定近邻数量的一半,则未分配所述类簇的电子商务评论为已被分配所述类簇的电子商务评论所在簇的必然隶属点。

[0078] 在本实施例中,对于 $\forall x, y \in X$ (X 是样本集),设电子商务评论 x 已经被分配所属的簇,电子商务评论 y 尚未分配所属类簇,则 y 是 x 所在簇的必然隶属点,当且仅当 x 与 y 的共同邻居数至少为 K 近邻中的一半,表示为公式(1)或(2)。

$$[0079] \quad |\{k | k \in \Gamma(A) \wedge k \in \Gamma(B)\}| \geq K/2 \quad (1)$$

[0080] 即

$$[0081] \quad |\text{SNN}(A, B)| \geq K/2 \quad (2)$$

[0082] 如图3所示。假设 A 被分配到相应的聚类, B 还没有分配,并且 l_A 和 l_B 是与 AB 线段垂直的两条线, $k=10$ 并且所有点都是分布的均匀。在图3a中, A 和 B 均有10个邻居点,其中两者共享5个。图表明 A 和 B 已经足够接近,可以被视为在同一聚类中。图3b的情况正好相反,虽然在 A 和 B 之间有许多点,但是, A 和 B 的邻域的交集是空集合,这表明 A 和 B 彼此远离,不被认为是在同一个聚类内。

[0083] 在本实施例的步骤(2-3)中,判断所述第一样本中的电子商务评论为可能隶属点的方法为:

[0084] 步骤(2-3-1):对所述基于中心点集,近邻数和距离矩阵的第一样本进行SNN的计算;

[0085] 步骤(2-3-2):判断计算结构是否满足可能隶属点定义。

[0086] 所述可能隶属点定义为:

[0087] 已被分配所述类簇的电子商务评论与未分配所述类簇的电子商务评论的共同邻居数量大于零小于设定近邻数量的一半,则未分配所述类簇的电子商务评论为已被分配所述类簇的电子商务评论所在簇的可能隶属点。

[0088] 在本实施例中,对于 $\forall x, y \in X$ (X 是样本集),设电子商务评论 x 已经被分配所属的簇,电子商务评论 y 尚未分配所属类簇,则 y 是 x 所在簇的可能隶属点,当且仅当满足公式(3)或(4):

$$[0089] \quad 0 < |\{k | k \in \Gamma(A) \wedge k \in \Gamma(B)\}| < K/2 \quad (3)$$

[0090] 即

$$[0091] \quad 0 < |\text{SNN}(A, B)| < K/2 \quad (4)$$

[0092] 本实施例步骤(3)的具体步骤包括:

[0093] 步骤(3-1):如果此电子商务评论是必然隶属点,则按照必然隶属点的准则进行样本点分配。

[0094] 步骤(3-2):如果此电子商务评论是可能隶属点,则按照可能隶属点的准则进行样本点分配,得到至少两个包含不同电子商务评论类型的类簇。

[0095] 在本实施例的步骤(3-1)中,所述必然隶属点准则进行样本点分配算法依赖于公式(1)或(2)定义的必然隶属点判定条件。必然隶属点分配算法利用广度优先搜索,从聚类中心开始,依次考察其 K 近邻点,若该点与考察的点之间存在较多共享近邻,则可认为考察的点必然隶属于自身所在的簇,将其划分到所在簇,并加入队列,以便后续继续考察该点的近邻点。

[0096] 必然隶属点的分配过程如下：

[0097] 步骤(3-1-1)：初始化队列Q，将所有聚类中心放入Q；

[0098] 步骤(3-1-2)：从Q中取出队首点this；

[0099] 步骤(3-1-3)：考察this的一个未被考察的K近邻点next，若next不属于任何簇，且 $|\text{SNN}(\text{this}, \text{next})| \geq K/2$ ，则将next纳入this所属簇，并将next放入Q的尾部；

[0100] 步骤(3-1-4)：若点this仍有尚未考察的近邻点，重复第3步；

[0101] 步骤(3-1-5)：若Q非空，重复第2步，否则结束算法2。

[0102] 步骤(3-1-6)：得到初步聚类结果 $\Omega = \{C_1, C_2, \dots, C_m\}$ 。

[0103] 在本实施例的步骤(3-2)中，所述可能隶属点准则进行样本点分配算法依赖于公式(3)或(4)定义的可能隶属点判定条件。可能隶属点分配算法包括：遍历所有未分配点，统计其近邻中隶属于每个簇的点的数量，构成一分配矩阵，之后寻找矩阵中的最大值，再次遍历所有未分配点，若其近邻中隶属于某一簇的点数量与前述最大值相同，则将该点分配至该簇，遍历过所有未分配点后更新识别矩阵，再次重复该过程，完成剩余点的分配。

[0104] 可能隶属点的分配过程如下：

[0105] 步骤(3-2-1)：查找所有未分配点，并按行为未分配点编号，列为簇的编号，构成分配矩阵recog；

[0106] 步骤(3-2-2)：统计每行对应点的K近邻点隶属的簇，将近邻中隶属于每个簇的点数量填入该行对应列中；

[0107] 步骤(3-2-3)：查找分配矩阵中的最大值，记为most

[0108] 步骤(3-2-4)：若 $\text{most} > 0$ ，则查找recog中所有的most值，并记录其所在行列分别形成向量 $\text{Row} = \{r_1, r_2, \dots, r_p\}$ 和 $\text{Col} = \{c_1, c_2, \dots, c_p\}$ ；将所有 r_i 对应的未分配点分配至 c_i 对应的簇，否则重复第1步；

[0109] 步骤(3-2-5)：若仍有未分配点，重复第1步，否则结束。得到最终聚类结果 $\Psi = \{C_1, C_2, \dots, C_m\}$

[0110] 通过图4的聚类结果表明在电子商务评论数据集运用本专利提出的面向电子商务评论数据的样本点分配策略聚类之后的效果都是比较好的，可以识别虚假评论和真实评论两个簇，其衡量指标AMI值，ARI，FMI值均为1，ARG值为5，可以看出聚类效果很好。

[0111] 以图5为例，我们更加详细的说明通过必然隶属点分配和可能隶属点分配的两步分配策略，可以使尽可能多的电子商务评论数据被分配到正确的类簇中，这种分配方法可以更加灵活地处理不同形状类簇。若两个样本点 i, j 必然隶属于同一类簇，则二者应该具有较近距离的。但是，若仅仅以距离作为判别标准，则极易受到样本点分布随机性的影响。

[0112] 图5中每个格子的边长为1，若样本点A与样本点B之间的距离略小于1，即 $\text{dist}(A, B) < 1$ ，按照传统的一步分配策略，在分配点B的归属时，由于距离B最近的点为A，则会错误地将B分配至点黑色类簇，若灰色点尚处于未分配状态，在后续的分配中，所有灰色点甚至都会被分配至黑色类簇，造成严重错分。

[0113] 而在必然隶属点和可能隶属点所示两步分配策略下，在 $K = 4$ 时，近邻情况为： $\Gamma(A) = \{A, A_1, A_2, A_3\}$ ， $\Gamma(B) = \{B, A, B_1, B_2\}$ ，

[0114] A, B两点的共享近邻为 $\text{SNN}(A, B) = \{A\}$ 。

[0115] 而由于 $|\text{SNN}(A, B)| < K/2$ ，故无法判定点B必然隶属于黑色类簇。

[0116] 在可能隶属点分配过程中,由于 $\Gamma(B) = \{B, A, B_1, B_2\}$,

[0117] 因此,在识别矩阵recog中,黑色类簇的列值为1,而灰色类簇的列值为2。此时即将B划入灰色类簇,没有出现错误。所以这可以说明通过必然隶属点分配和可能隶属点分配的两步分配策略,可以使尽可能多的电子商务评论数据被分配到正确的类簇中。

[0118] 实施例2:

[0119] 本实施例2的目的是提供一种计算机可读存储介质。

[0120] 为了实现上述目的,本发明采用如下一种技术方案:

[0121] 一种计算机可读存储介质,其中存储有多条指令,所述指令适于由终端设备设备的处理器加载并执行以下处理:

[0122] 步骤(1):接收电子商务评论数据集,获取电子商务评论中心点集、近邻数和距离矩阵;

[0123] 步骤(2):建立与电子商务评论中心点集、近邻数和距离矩阵相对应的第一样本,判断所述第一样本中的电子商务评论为必然隶属点或可能隶属点;

[0124] 步骤(3):根据两步分配策略进行非聚类中心节点分配,所述两步分配策略为以必然隶属点和可能隶属点为度量,按照必然隶属点或可能隶属点准则进行样本点分配,得到真实和虚假两个不同类型的电子商务评论类簇。

[0125] 实施例3:

[0126] 本实施例3的目的是提供一种基于电子商务评论数据的非聚类中心节点分配装置。

[0127] 为了实现上述目的,本发明采用如下一种技术方案:

[0128] 一种基于电子商务评论数据的非聚类中心节点分配装置,包括处理器和计算机可读存储介质,处理器用于实现各指令;计算机可读存储介质用于存储多条指令,所述指令适于由处理器加载并执行以下处理:

[0129] 步骤(1):接收电子商务评论数据集,获取电子商务评论中心点集、近邻数和距离矩阵;

[0130] 步骤(2):建立与电子商务评论中心点集、近邻数和距离矩阵相对应的第一样本,判断所述第一样本中的电子商务评论为必然隶属点或可能隶属点;

[0131] 步骤(3):根据两步分配策略进行非聚类中心节点分配,所述两步分配策略为以必然隶属点和可能隶属点为度量,按照必然隶属点或可能隶属点准则进行样本点分配,得到真实和虚假两个不同类型的电子商务评论类簇。

[0132] 这些计算机可执行指令在设备中运行时使得该设备执行根据本公开中的各个实施例所描述的方法或过程。

[0133] 在本实施例中,计算机程序产品可以包括计算机可读存储介质,其上载有用于执行本公开的各个方面的计算机可读程序指令。计算机可读存储介质可以是可以保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是一一但不限于一一电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、静态随机存取存储器(SRAM)、便携式压缩盘只读存储器(CD-ROM)、数字多功能盘

(DVD)、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身，诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波(例如，通过光纤电缆的光脉冲)、或者通过电线传输的电信号。

[0134] 本文所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备，或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令，并转发该计算机可读程序指令，以供存储在各个计算/处理设备中的计算机可读存储介质中。

[0135] 用于执行本公开内容操作的计算机程序指令可以是汇编指令、指令集架构 (ISA) 指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、或者以一种或多种编程语言的任意组合编写的源代码或目标代码，所述编程语言包括面向对象的编程语言—诸如 C++ 等，以及常规的过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络—包括局域网 (LAN) 或广域网 (WAN)—连接到用户计算机，或者，可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。在一些实施例中，通过利用计算机可读程序指令的状态信息来个性化定制电子电路，例如可编程逻辑电路、现场可编程门阵列 (FPGA) 或可编程逻辑阵列 (PLA)，该电子电路可以执行计算机可读程序指令，从而实现本公开内容的各个方面。

[0136] 应当注意，尽管在上文的详细描述中提及了设备的若干模块或子模块，但是这种划分仅仅是示例性而非强制性的。实际上，根据本公开的实施例，上文描述的两个或更多模块的特征和功能可以在一个模块中具体化。反之，上文描述的一个模块的特征和功能可以进一步划分为由多个模块来具体化。

[0137] 实施例4:

[0138] 本实施例4的目的是提供一种基于电子商务评论数据的聚类方法。

[0139] 为了实现上述目的，本发明采用如下一种技术方案:

[0140] 根据本发明的第四目的，本发明还提供了一种基于电子商务评论数据的聚类方法，采用所述一种基于电子商务评论数据的非聚类中心节点分配方法进行非聚类中心节点分配，得到真实和虚假两个不同类型的电子商务评论类簇进行聚类。

[0141] 以上所述仅为本申请的优选实施例而已，并不用于限制本申请，对于本领域的技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。因此，本发明将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

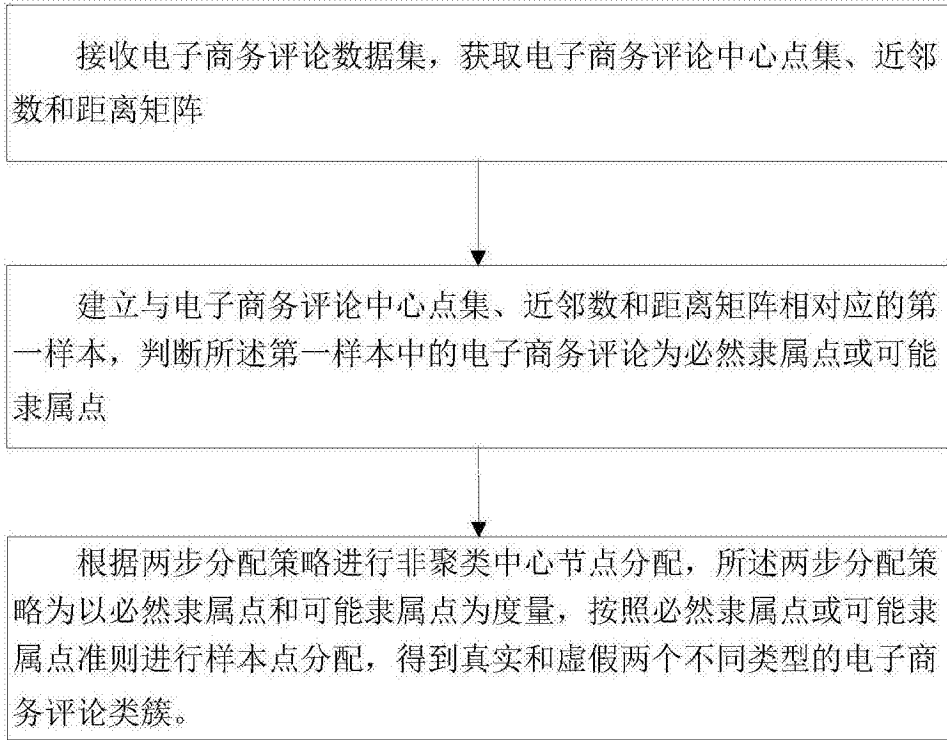


图1

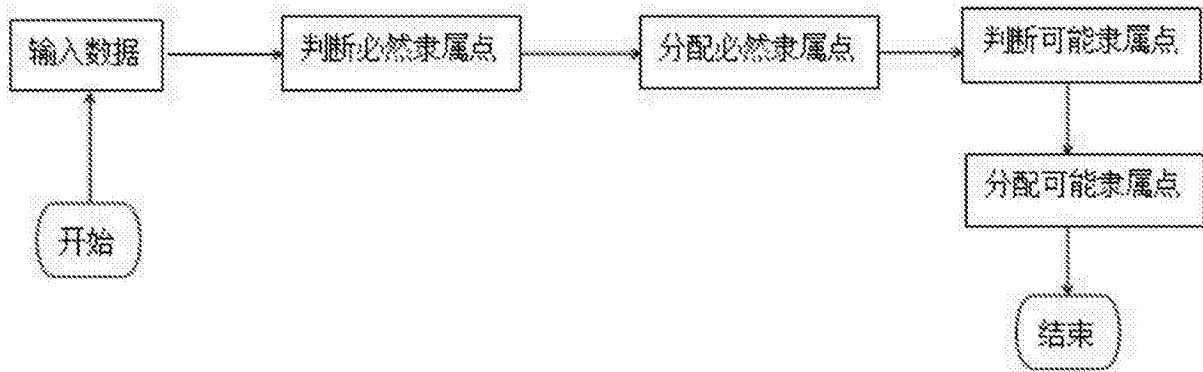


图2

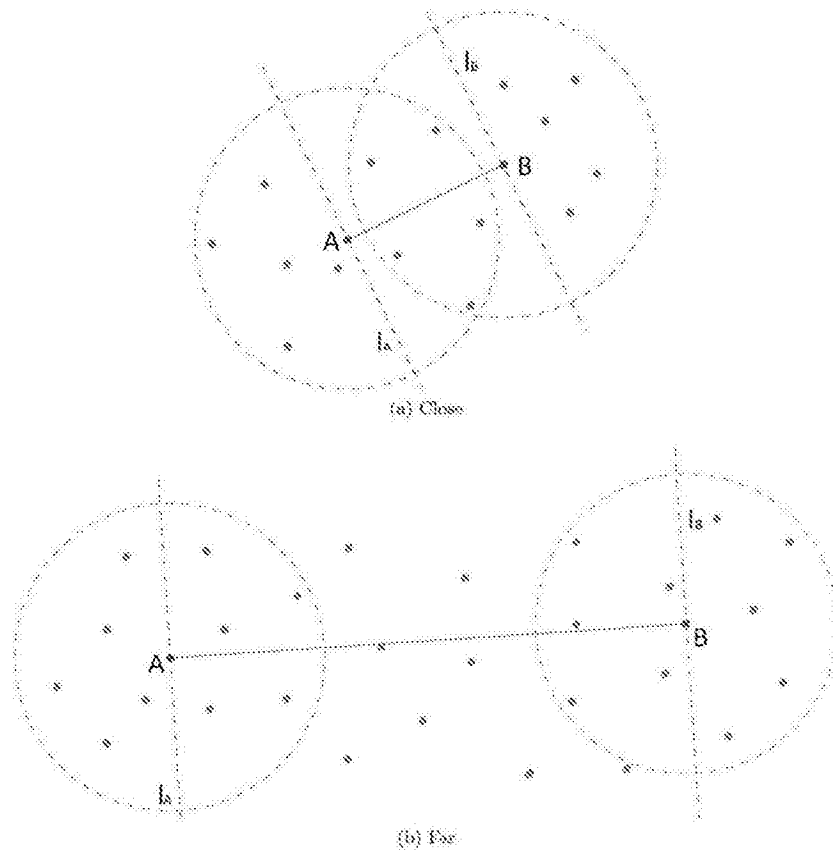


图3

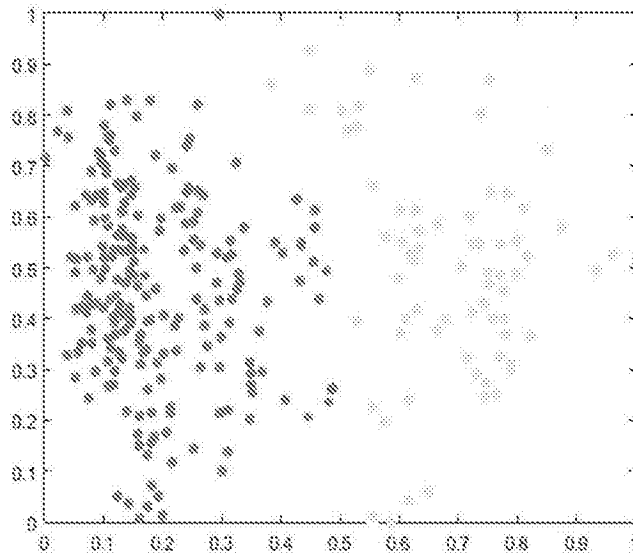


图4

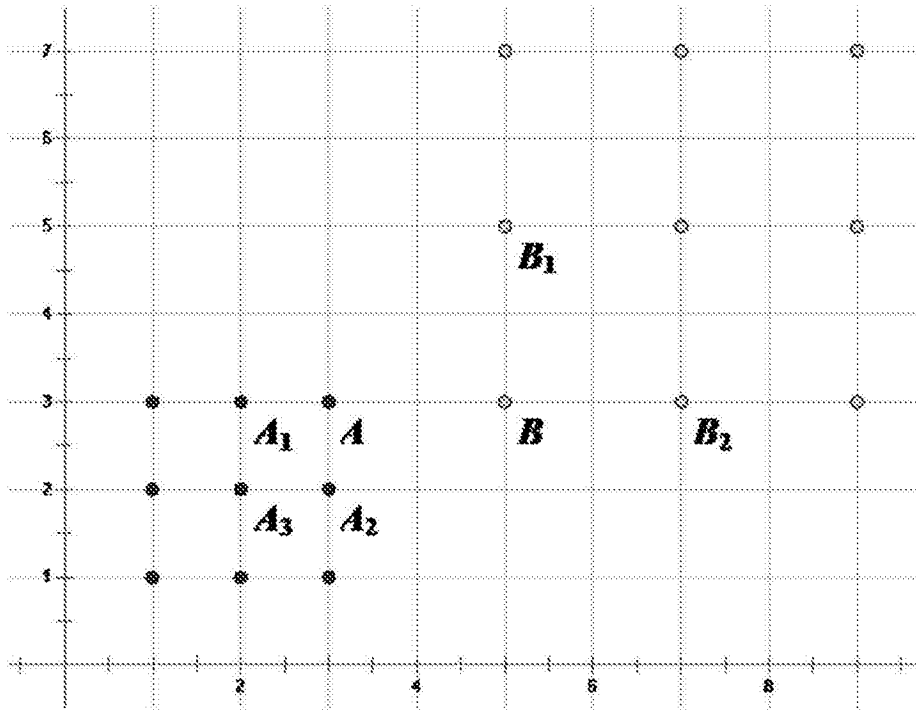


图5