



(12) 发明专利申请

(10) 申请公布号 CN 104020845 A

(43) 申请公布日 2014. 09. 03

(21) 申请号 201410119897. X

(22) 申请日 2014. 03. 27

(71) 申请人 浙江大学

地址 310027 浙江省杭州市西湖区浙大路
38 号

(72) 发明人 陈岭 侯仓健

(74) 专利代理机构 杭州天勤知识产权代理有限
公司 33224

代理人 胡红娟

(51) Int. Cl.

G06F 3/01 (2006. 01)

G06F 3/0346 (2013. 01)

G06K 9/62 (2006. 01)

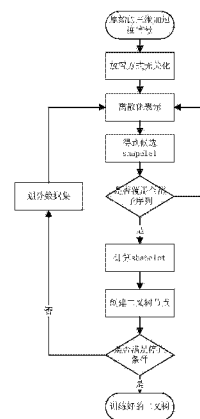
权利要求书2页 说明书9页 附图4页

(54) 发明名称

基于 shapelet 特征的加速度传感器放置无关化运动识别方法

(57) 摘要

本发明公开了一种基于 shapelet 特征的加速度传感器放置无关化运动识别方法,通过处理加速度传感器的运动信号识别人体运动。本方法分为数据训练和运动识别两个部分,在数据训练部分中首先将原始的三维加速度信号处理成和放置方式无关的一维信号,然后抽取一维信号中与放置位置无关的 shapelet 特征,建立识别模型;在运动识别部分中对实时采集的三维加速度信号进行处理得到和放置方式无关的一维信号,然后利用数据训练部分训练出的识别模型识别出运动。该方法可运行在装有加速度传感器的设备如智能手机、智能平板等设备上,在任意的放置方式和位置下均可提供准确的运动识别结果。



1. 一种基于 shapelet 特征的加速度传感器放置无关化运动识别方法,其特征在于,包括:

步骤 1,获取若干原始的三维加速度信号,将每个原始三维加速度信号处理成与放置方式无关的时间序列,得到数据集,其中原始的三维加速度信号带有类别标注;

步骤 2,从数据集中获取设定长度的时间子序列并进行离散化表示,得到离散化序列的集合;

步骤 3,利用随机映射从离散化序列的集合中得到候选 shapelet,并更改设定长度,返回步骤 2,直至数据集中所有长度的时间子序列均已被获取;

步骤 4,从候选 shapelet 中计算得到 shapelet;

步骤 5,判断步骤 2 中所处理的数据集是否符合预设的停止条件:

符合,对符合停止条件的数据集停止操作;

不符合,根据步骤 4 所得 shapelet 将不符合停止条件的数据集划分为两个数据集,针对每个划分后得到的数据集重复步骤 2 至 4;

步骤 6,利用所得的全部 shapelet 构建分类器;

步骤 7,获取实时的三维加速度信号,将实时的三维加速度信号处理为与放置方式无关的时间序列,将时间序列与分类器中的 shapelet 进行距离比较,识别出实时的三维加速度信号所属的类别。

2. 如权利要求 1 所述基于 shapelet 特征的加速度传感器放置无关化运动识别方法,其特征在于,将三维加速度信号处理成与放置方式无关的时间序列方法如下,其中三维加速度信号是指原始的三维加速度信号或实时的三维加速度信号:

步骤 1-1,将原始三维加速度信号进行标准化处理,在各维度上均变成均值为 0,标准差为 1 的标准序列,得到标准三维序列;

步骤 1-2,计算标准三维序列的协方差矩阵;

步骤 1-3,计算协方差矩阵的特征值和对应特征值的特征向量,取最大特征值对应的特征向量;

步骤 1-4,将步骤 1-1 中得到的标准三维序列乘以步骤 1-3 中得到的特征向量得到与放置方式无关的时间序列。

3. 如权利要求 1 所述基于 shapelet 特征的加速度传感器放置无关化运动识别方法,其特征在于,在步骤 2 中,利用 SAX 方法来进行时间子序列的离散化表示。

4. 如权利要求 1 所述基于 shapelet 特征的加速度传感器放置无关化运动识别方法,其特征在于,在步骤 3 中,从离散化序列的集合中得到候选 shapelet 的方法如下:

步骤 3-1,对于离散化序列的集合中每个离散化序列,随机遮罩其中的若干列;

步骤 3-2,使用计数表来统计当前随机遮罩后的离散化序列在各个类别中出现的次数;

步骤 3-3,重复步骤 3-1 至步骤 3-2 若干次,将各次所得计数表进行累加,得到遮罩前离散化序列与其对应类别的最终计数表:

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{mn} \end{bmatrix},$$

在计数表中,有 m 个遮罩前离散化序列,每个遮罩前离散化序列具有 n 个类别;
计算计数表 G 的补表:

$$\hat{G} = \begin{bmatrix} \hat{g}_{11} & \hat{g}_{12} & \cdots & \hat{g}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{g}_{m1} & \hat{g}_{m2} & \cdots & \hat{g}_{mn} \end{bmatrix},$$

其中对于补表中任意一个元素 $\hat{g}_{ij} = \max_{i=1}^{tom}(g_{ij}) - g_{ij}$;

对于 m 个遮罩前离散化序列,其中第 i 个遮罩前离散化序列对应的权重为:

$$w_i = \sum_{k=1}^n \text{abs}(g_{ik} - \hat{g}_{ik});$$

根据权重对遮罩前离散化序列进行排序,取权重较大的若干遮罩前离散化序列所对应的时间子序列作为候选 shapelet。

5. 如权利要求 1 所述基于 shapelet 特征的加速度传感器放置无关化运动识别方法,其特征在于,在步骤 4 中,从候选 shapelet 中计算得到 shapelet 的方法如下:

步骤 4-1,对于一个候选 shapelet,计算候选 shapelet 与数据集中所有时间序列的距离,得到时间距离集合;

步骤 4-2,从时间距离集合中查找最佳分割阈值,使得该候选 shapelet 的信息增益最大,将最佳分割阈值对应的信息增益作为该候选 shapelet 的信息增益;

步骤 4-3,重复步骤 4-1 至 4-2,直至得到对所有候选 shapelet 及其最佳分割阈值,并比较各个候选 shapelet 的信息增益,将信息增益最大的候选 shapelet 作为 shapelet。

6. 如权利要求 5 所述基于 shapelet 特征的加速度传感器放置无关化运动识别方法,其特征在于,在步骤 5 中,根据 shapelet 将当前数据集划分为两个数据集的方法为,根据 shapelet 所对应的最佳分割阈值划将数据集划分为两个数据集,其中,第一数据集中任意时间序列与 shapelet 之间的距离均小于最佳分割阈值,且第二数据集中任意时间序列与 shapelet 之间距离均大于或等于最佳分割阈值。

7. 如权利要求 1 所述基于 shapelet 特征的加速度传感器放置无关化运动识别方法,其特征在于,在步骤 6 中,构建的分类器为二叉树,对满足停止条件的数据集,将该数据集的类别作为该数据集的 shapelet 所在节点的叶节点。

8. 如权利要求 1 所述基于 shapelet 特征的加速度传感器放置无关化运动识别方法,其特征在于,在步骤 5 中,停止条件为以下任意一者:

a) 当前数据集中的数据数目小于第一预设值;

b) 当前数据集中最多的同类别时间子序列数目占当前数据集数目的比例达到第二预设值;

c) 重复次数超过第三预设值。

9. 如权利要求 7 所述基于 shapelet 特征的加速度传感器放置无关化运动识别方法,其特征在于,在步骤 7 中,将时间序列与分类器中的 shapelet 进行距离比较,识别出实时的三维加速度信号所属的类别方法为:从二叉树的根节点开始,对于当前节点,计算时间序列和当前节点 shapelet 的距离,并根据距离到达对应的子节点,直至到达叶节点:距离小于当前节点 shapelet 所对应的最佳分割阈值,到达当前节点的左边子节点;否则,到达当前节点的右边子节点。

基于 shapelet 特征的加速度传感器放置无关化运动识别方法

技术领域

[0001] 本发明涉及活动识别领域,具体涉及一种基于 shapelet 特征的加速度传感器放置无关化运动识别方法。

背景技术

[0002] 近年来,加速度传感器作为常用惯性传感器被广泛应用在各项智能手持设备上,如智能手机、智能腕带、智能手表等,基于传感器的人机交互也随之越来越受到重视,在人机交互中使用者的运动状态常作为重要的输入信息,因此利用加速度传感器进行运动识别成为当前热门的研究课题。同时,由于智能手持设备总是被用户随身携带,且计算能力越来越强,利用智能手持设备中内置加速度传感器对人体运动进行识别变得可行,并在交互游戏、健康监护等领域有广泛的应用。

[0003] 基于加速度传感器的运动识别方法主要由特征抽取、模型建立、活动识别三部分构成。现有方法主要抽取均值、方差、最大值、最小值等时域特征和频域熵等频域特征,但是由于常见智能手持设备如智能手机,有多种不同的放置方式,如手机的屏幕背向或面向人体,手机正立或者反转放置等,且智能手机可以放置在胸口、裤子口袋等多个不同的放置位置,这些放置方式和放置位置的差异使得提取出的特征在分布上差异很大,因此,现有利用加速度传感器进行人体运动识别的方法要求加速度传感器必须以特定的方式固定在人体特定的位置,当不按照指定方式佩戴传感设备时会极大影响识别的准确率。然而,传感设备在使用过程中很容易发生平移或旋转,偏离预定的放置方式和放置位置,导致识别性能下降。该现象在使用智能手持设备的情况下尤其严重,因为用户对智能手持设备的放置方式和位置的自由度很高,且智能手持设备在用户运动过程中一般无法固定。

[0004] 现有基于加速度传感器的放置方式和位置无关运动识别方法引入模体概念来减少传感器放置位置对加速度信号的影响,但是模体仅关注时间序列中频繁出现的序列,这样的序列可能在所有运动类别的时间序列中均频繁出现,不利于运动识别。在时序处理领域,shapelet 是指最能代表某一类别的一个时间序列,这样的序列具有频繁出现在某一类别,基本不出现在其他类别的特点,对运动识别很有帮助。加速度信号作为一种时间序列信号,引入 shapelet 可以更有效地找到加速度信号中最能代表各项运动的信号序列,解决放置不同对运动识别造成的影响,提高运动识别的准确率。发明内容

[0005] 本发明要解决的问题是如何在传感设备的放置方式和位置无法固定的情况下识别人体运动。为了解决上述问题,本发明提出了一种基于加速度传感器,面向智能手持设备用户,与传感设备放置方式和位置无关的运动识别方法。该方法利用加速度信号本身变化的特点,利用降维和 shapelet 特征,提取出最能代表各项运动的加速度信号序列,这样的序列具有不随传感设备放置方式和位置变化而改变的特点,从而解决非固定加速度传感器位置下的运动识别问题。

[0006] 一种基于 shapelet 特征的加速度传感器放置无关化运动识别方法,包括:

[0007] 获取若干原始的三维加速度信号,将每个原始三维加速度信号处理成与放置方式无关的时间序列,得到数据集,其中原始的三维加速度信号带有类别标注;

[0008] 步骤 2,从数据集中获取设定长度的时间子序列并进行离散化表示,得到离散化序列的集合;

[0009] 步骤 3,利用随机映射从离散化序列的集合中得到候选 shapelet,并更改设定长度,返回步骤 2,直至数据集中所有长度的时间子序列均已被获取;

[0010] 步骤 4,从候选 shapelet 中计算得到 shapelet;

[0011] 步骤 5,判断步骤 2 中所处理的数据集是否符合预设的停止条件;

[0012] 符合,对符合停止条件的数据集停止操作;

[0013] 不符合,根据步骤 4 所得 shapelet 将不符合停止条件的数据集划分为两个数据集,针对每个划分后得到的数据集重复步骤 2 至 4;

[0014] 步骤 6,利用所得的全部 shapelet 构建分类器;

[0015] 步骤 7,获取实时的三维加速度信号,将实时的三维加速度信号处理为与放置方式无关的时间序列,将时间序列与分类器中的 shapelet 进行距离比较,识别出实时的三维加速度信号所属的类别。

[0016] 步骤 1 至步骤 6 为数据训练阶段,步骤 7 为运动识别阶段。此处步骤 1 所得时间序列为一维信号,在步骤 2 中,数据集包括若干时间序列,每个时间序列包括不同长度的时间子序列,每次获取并且进行离散化的时间子序列均为数据集中设定长度的时间子序列,因此在步骤 3 中更改设定长度并返回步骤 2,直至数据集中所有长度的时间子序列均已被获取,保证在筛选时涵盖数据集中所有长度的时间子序列。

[0017] 将三维加速度信号处理成与放置方式无关的时间序列方法如下,其中三维加速度信号是指原始的三维加速度信号或实时的三维加速度信号:

[0018] 步骤 1-1,将原始三维加速度信号进行标准化处理,在各维度上均变成均值为 0,标准差为 1 的标准序列,得到标准三维序列;

[0019] 步骤 1-2,计算标准三维序列的协方差矩阵;

[0020] 步骤 1-3,计算协方差矩阵的特征值和对应特征值的特征向量,取最大特征值对应的特征向量;

[0021] 步骤 1-4,将步骤 1-1 中得到的标准三维序列乘以步骤 1-3 中得到的特征向量得到与放置方式无关的时间序列。

[0022] 其中标准序列的协方差矩阵计算公式如下:

$$[0023] \quad C = \begin{pmatrix} cov(X, X) & cov(X, Y) & cov(X, Z) \\ cov(Y, X) & cov(Y, Y) & cov(Y, Z) \\ cov(Z, X) & cov(Z, Y) & cov(Z, Z) \end{pmatrix}$$

[0024] 用 $cov(X_i, X_j)$ 表示协方差矩阵中的任意元素,其为序列 X_i 和序列 X_j 的协方差,其中 X_i 和 X_j 取值均为 X、Y 或 Z,计算公式为:

$$[0025] \quad cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)]$$

[0026] μ_i, μ_j 分别为序列 X_i, X_j 的均值。

[0027] 在步骤 2 中,利用 SAX 方法来进行时间子序列的离散化表示。

[0028] 通过该方法得到的离散化序列为字符串形式。

[0029] 在步骤 3 中,从离散化序列的集合中得到候选 shapelet 的方法如下:

[0030] 步骤 3-1,对于离散化序列的集合中每个离散化序列,随机遮罩其中的若干列;

[0031] 步骤 3-2,使用计数表来统计当前随机遮罩后的离散化序列在各个类别中出现的次数;

[0032] 步骤 3-3,重复步骤 3-1 至步骤 3-2 若干次,将各次所得计数表进行累加,得到遮罩前离散化序列与其对应类别的最终计数表:

[0033]

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{mn} \end{bmatrix},$$

[0034] 在计数表中,有 m 个遮罩前离散化序列,每个遮罩前离散化序列具有 n 个类别;

[0035] 计算计数表 G 的补表:

[0036]

$$\hat{G} = \begin{bmatrix} \hat{g}_{11} & \hat{g}_{12} & \cdots & \hat{g}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{g}_{m1} & \hat{g}_{m2} & \cdots & \hat{g}_{mn} \end{bmatrix},$$

[0037] 其中对于补表中任意一个元素 $\hat{g}_{ij} = \max_{i=1 \text{ to } m} (g_{ij}) - g_{ij}$;

[0038] 对于 m 个遮罩前离散化序列,其中第 i 个遮罩前离散化序列对应的权重为:

[0039] $w_i = \sum_{k=1}^n \text{abs}(g_{ik} - \hat{g}_{ik})$;

[0040] 根据权重对遮罩前离散化序列进行排序,取权重较大的若干遮罩前离散化序列所对应的时间子序列作为候选 shapelet。

[0041] $\max_{i=1 \text{ to } m} (g_{ij})$ 表示第 j 列中取值最大的 g_{ij} 。

[0042] 由于原始的三维加速度信号带有类别标注,因此其中的每个时间序列均具有类别标注(即标注为某个类别),因此每个时间序列中抽取的时间子序列以及对应的离散化序列也具有类别。离散化序列具有若干列,每列均用一个字符表示,随机遮罩若干列表示取出未遮罩的列所对应的维度进行类别统计,对于任意一个离散化序列,通过遮罩后所得字符串在各个类别中出现的次数,来统计该遮罩后所得字符串对应的离散化序列出现次数,用于改变离散化序列的权重,从而选出权重最大的离散化序列所对应的时间子序列作为候选 shapelet。离散化序列所对应的时间子序列是指离散化序列的时间子序列,因此候选 shapelet 也是未经离散化的时间子序列。

[0043] 步骤 3-3 中权重较大的离散化序列的数目为预设值。

[0044] 在步骤 4 中,从候选 shapelet 中计算得到 shapelet 的方法如下:

[0045] 步骤 4-1,对于一个候选 shapelet,计算候选 shapelet 与数据集中所有时间序列的距离,得到时间距离集合;

[0046] 步骤 4-2,从时间距离集合中查找最佳分割阈值,使得该候选 shapelet 的信息增益最大,将最佳分割阈值对应的信息增益作为该候选 shapelet 的信息增益;

[0047] 步骤 4-3,重复步骤 4-1 至 4-2,直至得到对所有候选 shapelet 及其最佳分割阈值,并比较各个候选 shapelet 的信息增益,将信息增益最大的候选 shapelet 作为

shapelet。

[0048] 对于一个候选 shapelet,最佳分割阈值为使当前的候选 shapelet 的信息增益最大的距离。

[0049] 在步骤 5 中,根据 shapelet 将当前数据集划分为两个数据集的方法为,根据 shapelet 所对应的最佳分割阈值划将数据集划分为两个数据集,其中,第一数据集中任意时间序列与 shapelet 之间的距离均小于最佳分割阈值,且第二数据集中任意时间序列与 shapelet 之间距离均大于或等于最佳分割阈值。

[0050] 此处第一、第二不具有时间或逻辑上的先后顺序,仅用于区分两个数据集。

[0051] 在步骤 6 中,构建的分类器为二叉树,对满足停止条件的数据集,将该数据集的类别作为该数据集的 shapelet 所在节点的叶节点。

[0052] 数据集的类别为当前数据集中同类别的时间子序列数目最多的类别。

[0053] 在步骤 5 中,停止条件为以下任意一者:

[0054] a) 当前数据集中的数据数目小于第一预设值;

[0055] b) 当前数据集中最多的同类别时间序列数目占当前数据集数目的比例达到第二预设值;

[0056] c) 重复次数超过第三预设值。

[0057] 第一预设值可为如最初的数据集中数据数目的 10%。

[0058] 第二预设值可为 90%。

[0059] 当构建的分类器为二叉树时,c) 中重复次数即表示二叉树的深度,第三预设值可为 $\log_2(\text{类别数目})+2$ 。

[0060] 在步骤 7 中,将时间序列与分类器中的 shapelet 进行距离比较,识别出实时的三维加速度信号所属的类别方法为:从二叉树的根节点开始,对于当前节点,计算时间序列和当前节点 shapelet 的距离,并根据距离到达对应的子节点,直至到达叶节点:距离小于当前节点 shapelet 所对应的最佳分割阈值,到达当前节点的左边子节点;否则,到达当前节点的右边子节点。

[0061] 在构建二叉树时,根据最佳分割阈值划分的两个数据集满足与最佳分割阈值的大小关系,其中一者中所有时间序列与 shapelet 的距离均小于最佳分割阈值,另一者中所有时间序列与 shapelet 的距离均大于或等于最佳分割阈值。

[0062] 本发明是一种新的基于加速度传感器运动识别方法,和现有的运动识别方法相比优点在于:

[0063] (1) 在加速度传感器放置无关活动识别中引入 shapelet 特征,提取出运动信号中不随放置方式和放置位置变化的特征;

[0064] (2) 装备加速度传感器的设备可以按照任意的放置方式放置在任意的的位置;

[0065] (3) 在识别过程中,装备加速度传感器的设备发生平移或者偏转基本不影响运动识别的准确率。

附图说明

[0066] 图 1 为数据训练阶段的流程图;

[0067] 图 2 为本发明一个实施例放置位置无关化的方法流程图;

- [0068] 图 3 为本发明当前实施例时间子序列离散化的方法流程图；
- [0069] 图 4 为本发明当前实施例从离散化序列中获取候选 shapelet 的方法流程图；
- [0070] 图 5 为本发明当前实施例随机映射过程示意图；
- [0071] 图 6 为本发明当前实施例进行一次随机映射后的计数表示例图；
- [0072] 图 7 为本发明当前实施例权重计算示意图；
- [0073] 图 8 为本发明当前实施例得到 shapelet 的步骤流程图；
- [0074] 图 9 为本发明当前实施例运动识别部分的流程图；
- [0075] 图 10 为本发明当前实施例将得到的序列 \bar{T} 根据高斯分布划分成 N 个等概率的区间方法示意图。

具体实施方式

[0076] 现结合具体实施例和附图对本发明方法进行详细解释。本发明提出一种基于 shapelet 特征的加速度传感器放置无关化运动识别方法,该方法包括模型训练和运动识别两个部分。

[0077] 模型训练部分主要功能是处理原始加速度信号,训练识别模型,该部分处理的对象是带正确标注的原始加速度信号序列的集合,主要过程包括放置方式无关化、数据离散化、筛选候选 shapelet、计算 shapelet、构建二叉树五个步骤。流程如图 1 所示,具体步骤如下:

[0078] 步骤 1,获取若干原始的三维加速度信号,将每个原始三维加速度信号处理成与放置方式无关的时间序列,得到数据集,其中原始的三维加速度信号带有类别标注。

[0079] 这一步骤主要进行放置方式无关化:对加速度传感器所获取的原始三维加速度信号进行处理,得到与设备放置方式无关的时序信号,且该时序信号为时间序列,流程如图 2 所示,主要过程如下:

[0080] 步骤 1-1,将原始三维加速度信号进行标准化处理,在各维度上均变成均值为 0,标准差为 1 的标准三维序列 (X, Y, Z) ;

[0081] 步骤 1-2,计算协方差矩阵,即:

$$[0082] \quad C = \begin{pmatrix} \text{cov}(X, X) & \text{cov}(X, Y) & \text{cov}(X, Z) \\ \text{cov}(Y, X) & \text{cov}(Y, Y) & \text{cov}(Y, Z) \\ \text{cov}(Z, X) & \text{cov}(Z, Y) & \text{cov}(Z, Z) \end{pmatrix} \quad (1)$$

[0083] 用序列 X_i 和序列 X_j 表示协方差矩阵中的序列 X, Y 和 Z ,则 $\text{cov}(X_i, X_j)$ 表示和序列 X_j 的协方差,计算公式为:

$$[0084] \quad \text{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \quad (2)$$

[0085] μ_i, μ_j 分别为序列 X_i 和序列 X_j 所对应的均值。

[0086] 步骤 1-3,计算协方差矩阵 C 的特征值和特征向量,取最大特征值对应的特征向量 w^T ;

[0087] 步骤 1-4,将步骤 1-1 中得到的标准三维序列 (X, Y, Z) 乘以步骤 1-3 中得到的特征向量 w^T 得到与放置方式无关的时间序列 $T = (t_1, t_2, \dots, t_n)$,其中:

$$[0088] \quad t_i = (x_i, y_i, z_i) \times w^T \quad (i=1, \dots, n) \quad (3)$$

[0089] 其中时间序列 T 由 n 个时间数据组成, 第 i 个时间数据 t_i 的三维坐标为 (x_i, y_i, z_i) 。

[0090] 步骤 2, 从数据集中获取所有指定长度的时间子序列并进行离散化表示, 得到离散化序列的集合。

[0091] 步骤 2 采用 SAX 方法进行, 设定长度 l , 对于数据集中的每一个时间序列, 将这些时间序列中所有长度为 l 的时间子序列进行离散化表示, 流程如图 3 所示, 主要过程如下:

[0092] 步骤 2-1, 将时间子序列进行标准化处理, 变成均值为 0, 标准差为 1 的标准时间序列 T ;

[0093] 步骤 2-2, 将长度为 l 的标准时间序列 T 表示为长度为 w 的序列 $\bar{T} = (\bar{t}_1, \dots, \bar{t}_w)$, 即得到分段序列, $w < l$, 其中 \bar{T} 中的第 i 个元素的计算公式为:

$$[0094] \quad \bar{t}_i = \frac{w}{l} \sum_{j=\frac{l}{w}(i-1)+1}^{\frac{l}{w}i} T_j \quad (i = 1, \dots, w) \quad (4)$$

[0095] 步骤 2-3, 如图 10 所示, 将得到的序列 \bar{T} 根据高斯分布划分成 N 个等概率的区间, 并将同一个概率区间的序列值用同一字符表示, 得到相应的字符串, 每个时间子序列所对应的字符串即为该时间子序列的离散化表示, 所得所有的离散化序列形成离散化序列的集合。其中横坐标表示时间, 纵坐标为概率分布空间。沿横坐标方向画直线, 划分高斯曲线, 所得高斯曲线与直线以及纵坐标之间的 N 个面积相等的区域, 所对应的纵坐标区间即为 N 个等概率的区间, 本发明当前实施例 N 取值为 3。

[0096] 步骤 3, 利用随机映射从离散化序列的集合中得到候选 shapelet, 并更改设定长度, 返回步骤 2, 直至数据集中所有长度的时间子序列均已被获取。对步骤 2 中所得离散化序列集合, 采用随机映射的方式, 筛选出最可能成为 shapelet 的离散化序列。流程如图 4 所示, 主要过程如下:

[0097] 步骤 3-1, 对于离散化序列的集合中每个离散化序列, 随机遮罩其中的若干列。如图 5 所示, 第一个运动“跑步”中有三个离散化序列, 第二个运动“跑步”中包含两个离散化序列, 第三个运动“走路”有两个离散化序列, 离散化序列构成了离散化序列集合, 在其中一次遮罩中, 每个离散化序列的第二列和第三列被遮罩, 第一个离散化序列从“acbac”变成较短的序列“aac”, 从而完成随机映射。

[0098] 步骤 3-2, 使用计数表来统计当前随机遮罩后的离散化序列在各个类别中出现的次数。

[0099] 如图 6 所示, 编号 1 中的离散化序列“acbac”在经过随机映射后变成较短的离散化序列“aac”, 这个离散化序列在类别跑步中出现了 1 次, 在类别走路中出现了 0 次, 更新计数表中对应的项, 编号 4 中的离散化序列“babac”和编号 6 的离散化序列“bcaac”在随机映射后都变成了短的离散化序列“bac”, 这个序列在类别跑步和类别走路中都只出现了一次, 因此计数表中编号 4 的计数均置为 1, 依此类推。

[0100] 步骤 3-3, 重复步骤 3-1 至步骤 3-2 若干次, 将各次所得计数表进行累加, 得到遮罩前离散化序列与其对应类别的最终计数表:

[0101] $G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{mn} \end{bmatrix}$, 即共有 m 个离散化序列, 每个离散化序列具有 n

个类别, 对于任意元素 g_{ij} 表示: 对于第 j 个类别, 将第 i 个遮罩前离散化序列在每次遮罩后所得的离散化序列在该类别中出现的次数进行累加所得的最终次数。

[0102] 其中重复次数为预设值, 本发明当前实施例中设为 10。

[0103] 计算得到最终计数表的补表:

[0104] $\widehat{G} = \begin{bmatrix} \widehat{g}_{11} & \widehat{g}_{12} & \cdots & \widehat{g}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{g}_{m1} & \widehat{g}_{m2} & \cdots & \widehat{g}_{mn} \end{bmatrix}$, 其中:

[0105] $\widehat{g}_{ij} = \max_{i=1}^m (g_{ij}) - g_{ij}$ (5)

[0106] 从而计算权重, 第 i 个离散化序列所对应的权重为:

[0107] $w_i = \sum_{k=1}^n \text{abs}(g_{ik} - \widehat{g}_{ik})$ (6)

[0108] 根据权重对离散化序列进行排序, 取权重较大的若干离散化序列。如图 7 所示, 假设计数表如图 7 中最左边的部分, 在跑步列中最大值为 4, 因此该计数表所对应补表的跑步列是通过最大值 4 减去计数表的跑步列对应元素计算所得, 如图 7 中间的部分所示, 权重的计算过程如图 7 右边的部分所示, 从图 7 中可以看出, 编号为 7 的时间子序列的权重最高, 这样的时间子序列最有可能为 shapelet, 因此可将该时间子序列选择为候选 shapelet。

[0109] 更改步骤 2 中的时间子序列长度 1, 重复步骤 2 及步骤 3, 直到覆盖全部的时间子序列, 得到离散化表示集合。

[0110] 步骤 4, 从候选 shapelet 计算得到 shapelet。

[0111] 处理步骤 3 中得到离散化序列构成的集合, 得到这些离散化序列所对应的时间子序列, 构成一系列可能成为 shapelet 的时间子序列(即候选 shapelet)的集合。从候选 shapelet 的集合中计算出 shapelet, 流程如图 8 所示, 过程如下:

[0112] 步骤 4-1, 对于一个候选 shapelet, 计算候选 shapelet 与数据集中所有时间子序列的距离, 得到时间距离集合。

[0113] 步骤 4 中, 数据集可能为最初的数据集 D , 也可能为划分后的数据集 D_R 或 D_L , 此处以最初的数据集 D 为例进行计算, 对于划分后的每个数据集, 计算 shapelet 的方式相同。计算候选 shapelet 与数据集 D 中所有时间序列的距离, 对于长度为 m 的时间序列 $T = (t_1, \dots, t_m)$ 和长度为 m' 的时间序列 $T' = (t'_1, \dots, t'_{m'})$, 设 $m \leq m'$, 计算时间序列 T 和 T' 之间的距离:

[0114] $\text{dist}(T, T') = \text{dist}(T', T) = \begin{cases} \sqrt{\sum_{i=1}^m (t_i - t'_i)^2}, & m = m' \\ \min_{C' \in D'} (\text{dist}(C', T)), & m < m' \end{cases}$ (7)

[0115] 其中, D' 是 T' 中所有长度为 m 的时间序列的集合, C' 是 D' 中长度为 m 的时间子序列, $\text{dist}(C', T)$ 的计算公式为 $\sqrt{\sum_{i=1}^m (t_i - t'_i)^2}$, 则 $\min_{C' \in D'} (\text{dist}(C', T))$

表示在 $\sqrt{\sum_{i=1}^m (t_i - t'_i)^2}$ 中取距离最小的值。

[0116] 步骤 4-2, 从计算出的距离集合中查找出最佳分割阈值。

[0117] 对于数据集 D 和作为候选 shapelet 的时间子序列 s , 从得到的距离集合中选择一个距离 d_{th} 将整个数据集 D 分为两个数据集 D_L 和 D_R , 使得属于数据集 D_L 中的任意时间序列 T_L 均满足 $\text{dist}(T_L, s) < d_{th}$, 属于数据集 D_R 中的任意时间序列 T_R 均满足 $\text{dist}(T_R, s) \geq d_{th}$ 。数据集 D_L 和 D_R 中的时间序列数目分别为 n_L 和 n_R , 则这次分割的信息增益为:

$$[0118] \quad \text{Gain}(s, d_{th}) = E(D) - \frac{n_L}{n} E(D_L) - \frac{n_R}{n} E(D_R) \quad (8)$$

[0119] 其中, $E(D)$ 为训练数据集 D 的熵。对于由 c 个不同类别的共 n 个时间子序列数据构成的数据集 D , 其中 $n \geq c$, 第 i 个类别的数目是 n_i , 则第 i 个类别出现的概率为 $p_i = \frac{n_i}{n}$,

定义数据集 D 的熵为:

$$[0120] \quad E(D) = -\sum_{i=1}^c p_i \log(p_i) \quad (9)$$

[0121] 将距离集合中每个距离作为分隔距离阈值, 计算对应的信息增益, 从中找出使得信息增益最大的距离, 称之为最佳分隔阈值 d_{osp} , 即对于距离集合中的任意距离 d_{th}^* , 均有:

$$[0122] \quad \text{Gain}(s, d_{osp(D,s)}) \geq \text{Gain}(s, d_{th}^*) \quad (10)$$

[0123] 步骤 4-3, 重复步骤 4-1 至 4-2, 直至得到对所有候选 shapelet 及其最佳分割阈值, 并比较各个候选 shapelet 的信息增益, 将信息增益最大的候选 shapelet 作为 shapelet。

[0124] 对于第一个作为候选 shapelet 的时间子序列 s , 将当前时间子序列及其对应的最佳分割阈值 $d_{osp}(D, s)$ 作为 shapelet, 否则将当前时间子序列 s 及其对应的最佳分割阈值 $d_{osp(D,s)}$ 分割数据集 D 得到的信息增益 $\text{Gain}(s, d_{osp(D,s)})$ 和与之前的 shapelet 分割数据集 D 得到的信息增益 $\text{Gain}(\text{shapelet})$ 进行比较, 若 $\text{Gain}(s, d_{osp(D,s)}) > \text{Gain}(\text{shapelet})$, 则更新 shapelet 为当前时间子序列 s 及其对应的最佳分割阈值 $d_{osp(D,s)}$ 。

[0125] 通过不断更新 shapelet, 得到信息增益最大的候选 shapelet 作为 shapelet。

[0126] 步骤 5, 判断步骤 2 中所处理的数据集是否符合预设的停止条件:

[0127] 符合, 对符合停止条件的数据集停止操作;

[0128] 不符合, 根据步骤 4 所得 shapelet 将不符合停止条件的数据集划分为两个数据集, 针对每个划分后得到的数据集重复步骤 2 至 4。

[0129] 根据步骤 4 所得 shapelet 和其对应的最佳分割阈值 d_{th} 将不符合停止条件的数据集(此处以最初的数据集 D 分为例)划分为两个数据集 D_L 和 D_R , 使得属于数据集 D_L 中的任意时间序列 T_L 均满足 $\text{dist}(T_L, s) < d_{th}$, 属于数据集 D_R 中的任意时间序列 T_R 均满足 $\text{dist}(T_R, s) \geq d_{th}$ 。将划分后的每个数据集进行停止条件的判断, 停止条件为以下任意一者:

[0130] a) 当前数据集中的数据数目小于第一预设值, 本发明当前实施例中为最初的数据集中数据数目的 10%;

[0131] b) 当前数据集中最多的同类别时间子序列数目占当前数据集数目的比例达到第二预设值,本发明当前实施例中为 90%;

[0132] c) 重复次数超过第三预设值,本发明当前实施例中为 $\log_2(\text{类别数目})+2$ 。

[0133] 符合,对符合停止条件的数据集停止操作;

[0134] 不符合,根据步骤 4 所得 shapelet 将不符合停止条件的数据集划分为两个数据集,针对每个划分后得到的数据集重复步骤 2 至 4。

[0135] 对满足停止条件的数据集,将该数据集的类别作为该数据集的 shapelet 所在节点的叶节点。

[0136] 步骤 6,利用所得的全部 shapelet 构建分类器。

[0137] 这里构建好的二叉树即为分类器,二叉树上所有的非叶节点均为 shapelet,这些 shapelet 集合构成了 shapelet 特征。

[0138] 运动识别部分主要功能是处理原始加速度信号,利用数据训练部分训练出的分类器对信号进行识别,该部分的处理对象是单条三维加速度信号序列。该部分包括放置方式无关化、根据模型识别运动两个过程,如图 9 所示,其中放置方式无关化和数据训练部分一致。第二阶段的具体内容如下:

[0139] 根据模型识别运动。获取实时的三维加速度信号,将实时的三维加速度信号处理为与放置方式无关的时间序列,将时间序列与分类器中的 shapelet 进行距离比较,识别出实时的三维加速度信号所属的类别方法为:从二叉树的根节点开始,对于当前节点,计算时间序列和当前节点 shapelet 的距离,并根据距离到达对应的子节点,直至到达叶节点:距离小于当前节点 shapelet 所对应的最佳分割阈值,到达当前节点的左边子节点;否则,到达当前节点的右边子节点。

[0140] 本发明方法可运行在装有加速度传感器的设备如智能手机、智能平板等设备上,在任意的放置方式和位置下均可提供准确的运动识别结果。

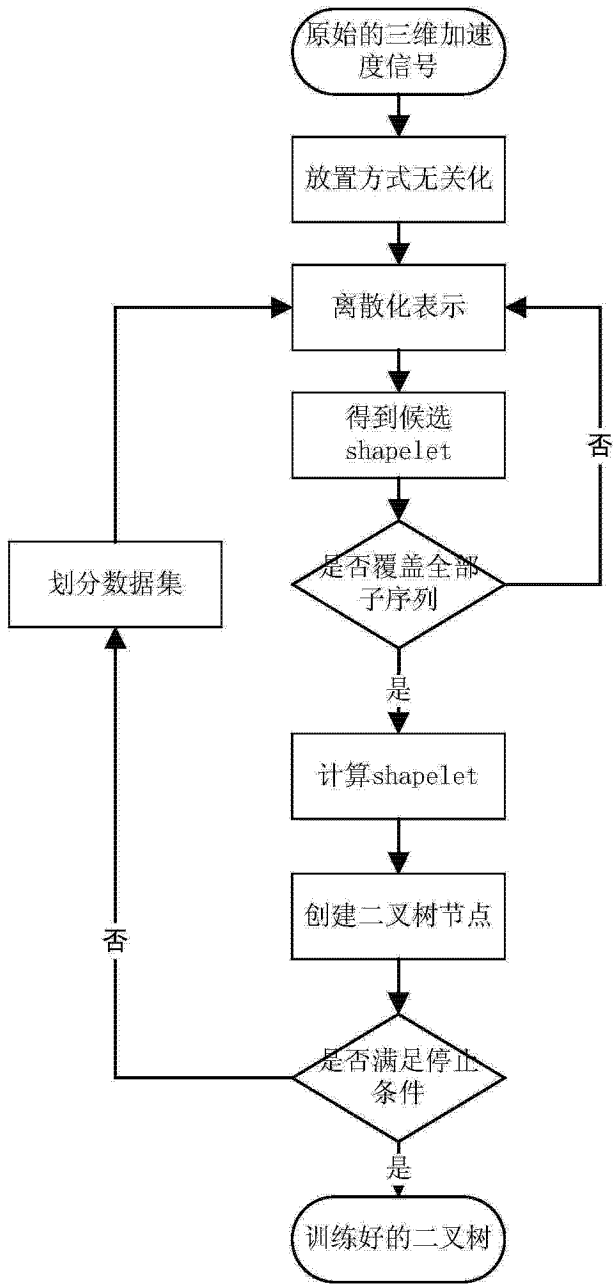


图 1

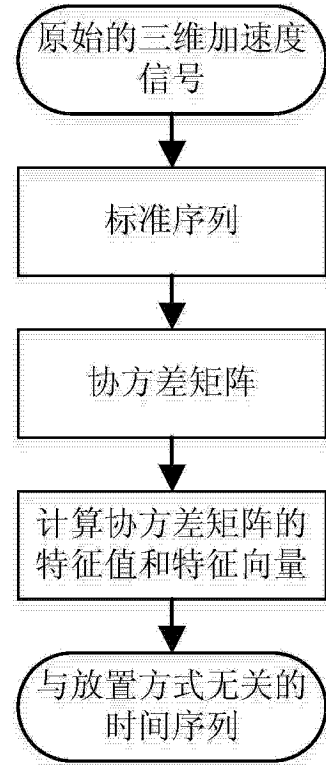


图 2

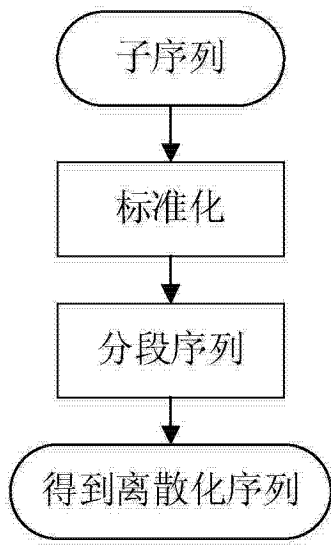


图 3

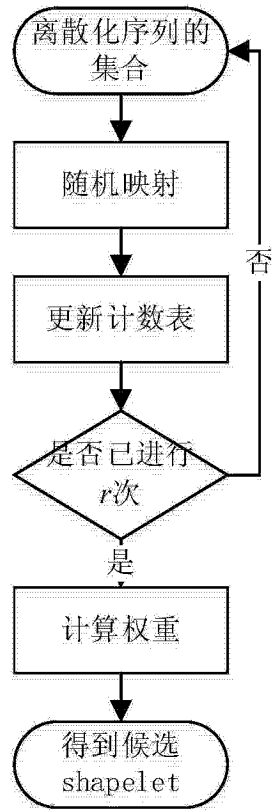


图 4

| 运动 | 随机映射过程 | | | | |
|----|--------|---|---|---|---|
| 跑步 | a | c | b | a | c |
| | b | c | c | c | a |
| | b | a | b | c | c |
| 跑步 | b | c | b | a | c |
| | b | c | a | c | c |
| 走路 | b | c | a | a | c |
| | b | b | a | c | b |

图 5

| 运动 编号 | | 随机映射 | | | | | 计数表 | | | | | |
|-------|---|------|---|---|---|---|-----|----|----|---|---|---|
| 跑步 | 1 | a | c | b | a | c | 编号 | 跑步 | 走路 | | | |
| | 2 | b | c | c | c | a | | | | 1 | 1 | 0 |
| | 3 | b | a | b | c | c | | | | 2 | 1 | 0 |
| 跑步 | 4 | b | c | b | a | c | | | | 3 | 2 | 0 |
| | 5 | b | c | a | c | c | | | | 4 | 1 | 1 |
| | 6 | b | c | a | a | c | | | | 5 | 2 | 0 |
| 走路 | 7 | b | b | a | c | b | | | | 6 | 1 | 1 |
| | | | | | | | 7 | 0 | 1 | | | |

图 6

| 计数表 | | | 计数表补表 | | | 权重 | |
|-----|----|----|-------|----|----|----|---------------------|
| 编号 | 跑步 | 走路 | 编号 | 跑步 | 走路 | 编号 | 权重 |
| 1 | 2 | 0 | 1 | 2 | 2 | 1 | $ 2-2 + 0-2 = 2$ |
| 2 | 2 | 0 | 2 | 2 | 2 | 2 | $ 2-2 + 0-2 = 2$ |
| 3 | 3 | 0 | 3 | 1 | 2 | 3 | $ 3-1 + 0-2 = 4$ |
| 4 | 3 | 2 | 4 | 1 | 0 | 4 | $ 3-1 + 2-0 = 4$ |
| 5 | 4 | 1 | 5 | 0 | 1 | 5 | $ 4-0 + 1-1 = 4$ |
| 6 | 3 | 2 | 6 | 1 | 0 | 6 | $ 3-1 + 2-1 = 4$ |
| 7 | 0 | 2 | 7 | 4 | 0 | 7 | $ 0-4 + 2-0 = 6$ |

图 7

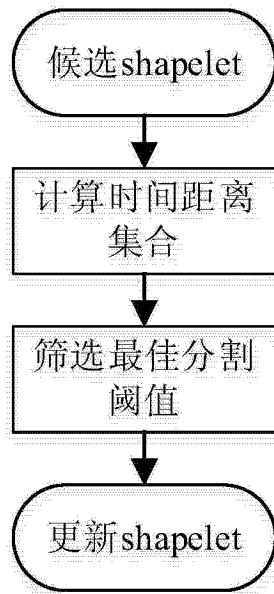


图 8

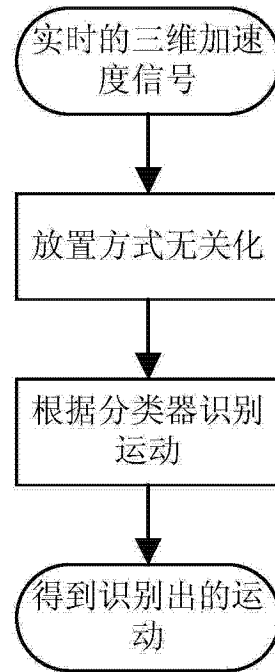


图 9

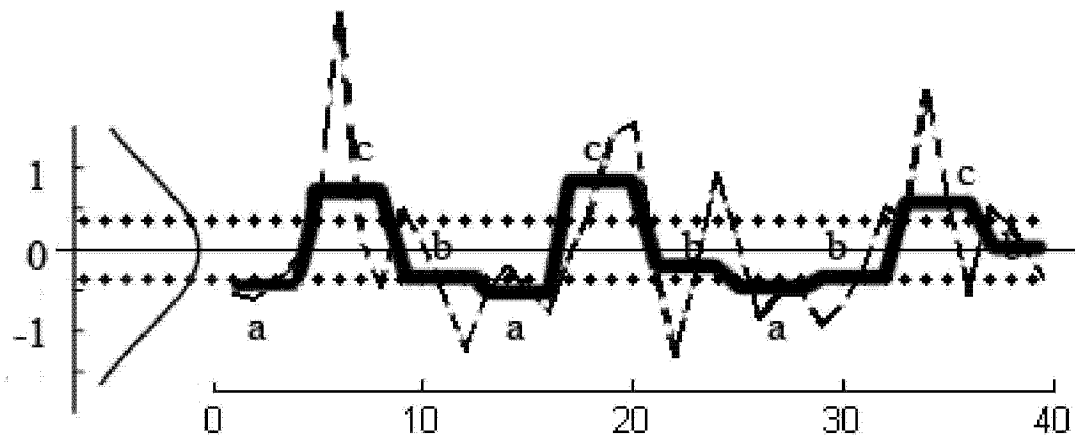


图 10