



## (12) 发明专利申请

(10) 申请公布号 CN 105447504 A

(43) 申请公布日 2016. 03. 30

(21) 申请号 201510751897. 6

(22) 申请日 2015. 11. 06

(71) 申请人 中国科学院计算技术研究所

地址 100190 北京市海淀区中关村科学院南路 6 号

(72) 发明人 陈益强 忽丽莎 谷洋 王晋东  
王双全

(74) 专利代理机构 北京泛华伟业知识产权代理有限公司 11280

代理人 王勇 李科

(51) Int. Cl.

G06K 9/62(2006. 01)

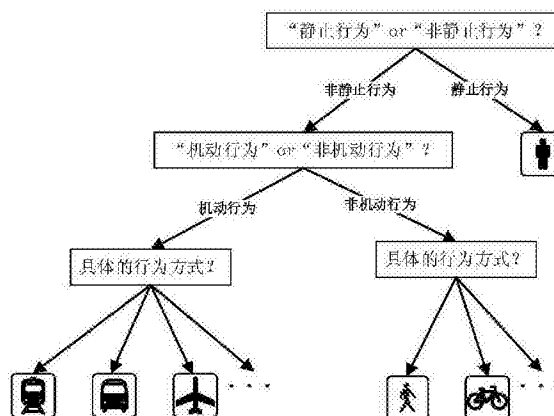
权利要求书2页 说明书6页 附图2页

### (54) 发明名称

一种交通模式行为识别方法及相应的识别模型构建方法

### (57) 摘要

本发明提供一种交通模式行为识别模型构建方法,包括:1) 对具体交通模式行为进行多层次的归类,建立相应的分类树;2) 对于分类树中每一个父节点的分类问题,基于随机森林模型,根据识别精度选择对应该分类问题的测试样本的最优特征集;3) 对于分类树中每一个父节点的分类问题,基于步骤2) 所得出的该父节点所对应的最优特征集,选择一个最优分类模型作为该父节点所对应的子分类模型。本发明还提供了相应的交通模式行为识别方法。本发明能够更加准确地区分具体的交通模式行为,且计算复杂度较小。



1. 一种交通模式行为识别模型构建方法,其特征在于,包括下列步骤:

1) 对具体交通模式行为进行多层次的归类,建立相应的分类树,其中所述分类树的每一个节点都代表一个类别,所述分类树的父节点下的子节点代表对该父节点对应类别进一步分类后所得到的子类,所述分类树的叶节点代表具体交通模式行为;

2) 对于分类树中每一个父节点的分类问题,基于随机森林模型,根据识别精度选择对应该分类问题的测试样本的最优特征集;其中所述父节点的分类问题是如何将样本从该父节点对应类别划分至其子节点对应类别的分类问题;

3) 对于分类树中每一个父节点的分类问题,基于步骤2)所得出的该父节点所对应的最优特征集,选择一个最优分类模型作为该父节点所对应的子分类模型。

2. 根据权利要求1所述的识别模型构建方法,其特征在于,所述步骤2)中,所述识别精度是:基于随机森林模型准确识别出类别的测试样本个数除以测试样本总个数;所述测试样本是:对已知的具体交通模式行为所采集的多种传感器数据的特征向量;所述特征向量包括所述传感器数据的多个特征,所述最优特征集中的特征在所述特征向量的范围内选择。

3. 根据权利要求2所述的识别模型构建方法,其特征在于,所述步骤2)中,所述传感器数据包括:加速度计、陀螺仪和磁力计数据。

4. 根据权利要求3所述的识别模型构建方法,其特征在于,所述步骤2)中,所述传感器数据还包括:基于卫星定位技术获得的速度数据。

5. 根据权利要求1所述的识别模型构建方法,其特征在于,所述步骤1)中,所述分类树的第一层子类包括“静止”和“非静止行为”;所述“非静止行为”的子类包括:“机动行为”和“非机动行为”。

6. 根据权利要求5所述的识别模型构建方法,其特征在于,所述步骤1)中,所述“非机动行为”的子类包括:“走路”、“跑步”和“骑车”;所述“机动行为”的子类包括:“乘公交”、“乘地铁”和“乘火车”。

7. 根据权利要求1所述的识别模型构建方法,其特征在于,所述步骤2)中,选择最优特征集的过程包括:

21) 根据基于特征集FeatureSet1的训练样本集训练随机森林模型RandomForest1,并计算模型的识别精度Accuracy1;所述训练样本集是所述测试样本所组成的集合;

22) 统计特征集FeatureSet1中各个特征在模型RandomForest1中的特征使用频率,将特征按照频率从大到小排序;其中特征使用频率是:随机森林中使用该特征的决策树个数除以随机森林中包含的决策树总个数;

23) 从特征集FeatureSet1中删掉使用频率取值较小的特征,剩余的特征组成特征集FeatureSet2;

24) 根据基于特征集FeatureSet2的训练样本集训练随机森林模型RandomForest2,计算该随机森林模型RandomForest2的识别精度Accuracy2;

25) 对比Accuracy2与Accuracy1,若Accuracy2小于Accuracy1,则进入步骤26);反之,用特征集FeatureSet2替换FeatureSet1,模型RandomForest2替换模型RandomForest1,精度Accuracy2替换精度Accuracy1,重新执行步骤22);

26) 将特征集FeatureSet1作为最优特征集。

8.根据权利要求2~7中任意一项所述的识别模型构建方法,其特征在于,所述步骤2)中,测试样本中初始的特征向量包括:均值、标准方差、过零率、百分位数、关联系数、功率谱密度、频域熵和谱峰位置。

9.一种交通模式行为识别方法,其特征在于,包括下列步骤:

a)获取基于多种传感器数据的待识别样本;  
b)读取多层次异构的交通模式行为识别模型的分类树,将根节点作为当前节点;  
c)根据当前节点子分类问题所对应的最优特征子集,生成基于最优特征子集的测试样本;

d)根据当前子分类问题所对应的最优分类模型,对基于最优特征子集的测试样本进行分类,得到下一层的类别;

e)如果所识别出的下一层类别为最底层类别,输出识别结果,转步骤f);否则,将所识别出的下一层类别的节点作为当前节点,待识别样本进入当前节点,转步骤c);

其中所述交通模式行为识别模型是根据权利要求1~7中任意一项所述的识别模型构建方法所生成的识别模型。

## 一种交通模式行为识别方法及相应的识别模型构建方法

### 技术领域

[0001] 本发明涉及普适计算、移动互联网及城市规划等技术领域,具体地说,本发明涉及一种交通模式行为识别方法及相应的识别模型构建方法。

### 背景技术

[0002] 利用传感数据识别用户的日常行为是普适计算领域的重要研究问题。交通模式行为(走路、骑单车、坐公交等)作为众多日常行为的一个子类,蕴含大量与用户的运动轨迹、生活规律等相关的信息。精准而有效地捕捉用户日常的交通模式行为,进而分析用户行为习惯和日常生活轨迹,对智能移动服务、健康监护、城市规划等多个领域都具有重要的作用。例如:利用交通模式行为识别技术,可以挖掘用户的日常轨迹信息和具体的出行方式;可以针对开车或坐车等进行合理的劝导,倡导绿色出行和健康出行。另外,大量用户交通模式行为数据的积累有利于城市规划、道路设计、公交线路优化等,具有广泛而重要的价值。随着移动终端的日益普及和终端内嵌的传感器件的发展,基于传感器非干扰地识别用户日常的交通模式行为已成为普适计算领域的研究热点。

[0003] 近年来,基于传感器的非干扰交通模式识别主要是:从传感器所采集数据中提取一系列特征作为特征集,通过对特征集进行分析处理建立和训练识别模型,利用该模型识别用户的多种交通模式行为,例如中国专利CN 201210185649等。这些方案都是试图寻找由某些固定特征组成的最优特征集以及寻找某个特定的最佳数学模型作为识别模型,然后再用这个最优特征集和识别模型来识别出具体的交通模式。然而,交通模式行为的类别众多,不同交通模式行为之间的相似度高,同种交通模式行为内部的差异性大,这些因素导致目前的交通识别方案的识别精度均不能令人满意,越来越难以满足实际应用的要求。

[0004] 因此,当前迫切需要一种能够对用户交通模式的准确区分的基于传感器数据的交通模式识别解决方案。

### 发明内容

[0005] 因此,本发明的任务是提供一种能够克服现有技术的上述缺陷的交通模式解决方案。

[0006] 根据本发明的一个方面,提供了一种交通模式行为识别模型构建方法,其特征在于,包括下列步骤:

[0007] 1)对具体交通模式行为进行多层次的归类,建立相应的分类树,其中所述分类树的每一个节点都代表一个类别,所述分类树的父节点下的子节点代表对该父节点对应类别进一步分类后所得到的子类,所述分类树的叶节点代表具体交通模式行为;

[0008] 2)对于分类树中每一个父节点的分类问题,基于随机森林模型,根据识别精度选择对应该分类问题的测试样本的最优特征集;其中所述父节点的分类问题是如何将样本从该父节点对应类别划分至其子节点对应类别的分类问题;

[0009] 3)对于分类树中每一个父节点的分类问题,基于步骤2)所得出的该父节点所对应

的最优特征集,选择一个最优分类模型作为该父节点所对应的子分类模型。

[0010] 其中,所述步骤2)中,所述识别精度是:基于随机森林模型准确识别出类别的测试样本个数除以测试样本总个数;所述测试样本是:对已知的具体交通模式行为所采集的多种传感器数据的特征向量;所述特征向量包括所述传感器数据的多个特征,所述最优特征集中的特征在所述特征向量的范围内选择。

[0011] 其中,所述步骤2)中,所述传感器数据包括:加速度计、陀螺仪和磁力计数据。

[0012] 其中,所述步骤2)中,所述传感器数据还包括:基于卫星定位技术获得的速度数据。

[0013] 其中,所述步骤1)中,所述分类树的第一层子类包括“静止”和“非静止行为”;所述“非静止行为”的子类包括:“机动行为”和“非机动行为”。

[0014] 其中,所述步骤1)中,所述“非机动行为”的子类包括:“走路”、“跑步”和“骑车”;所述“机动行为”的子类包括:“乘公交”、“乘地铁”和“乘火车”。

[0015] 其中,所述步骤2)中,选择最优特征集的过程包括:

[0016] 21)根据基于特征集`FeatureSet1`的训练样本集训练随机森林模型`RandomForest1`,并计算模型的识别精度`Accuracy1`;所述训练样本集是所述测试样本所组成的集合;

[0017] 22)统计特征集`FeatureSet1`中各个特征在模型`RandomForest1`中的特征使用频率,将特征按照频率从大到小排序;其中特征使用频率是:随机森林中使用该特征的决策树个数除以随机森林中包含的决策树总个数;

[0018] 23)从特征集`FeatureSet1`中删掉使用频率取值较小的特征,剩余的特征组成特征集`FeatureSet2`;

[0019] 24)根据基于特征集`FeatureSet2`的训练样本集训练随机森林模型`RandomForest2`,计算该随机森林模型`RandomForest2`的识别精度`Accuracy2`;

[0020] 25)对比`Accuracy2`与`Accuracy1`,若`Accuracy2`小于`Accuracy1`,则进入步骤26);反之,用特征集`FeatureSet2`替换`FeatureSet1`,模型`RandomForest2`替换模型`RandomForest1`,精度`Accuracy2`替换精度`Accuracy1`,重新执行步骤22);

[0021] 26)将特征集`FeatureSet1`作为最优特征集。

[0022] 其中,所述步骤2)中,测试样本中初始的特征向量包括:均值、标准方差、过零率、百分位数、关联系数、功率谱密度、频域熵和谱峰位置。

[0023] 根据本发明的另一方面,提供了一种交通模式行为识别方法,其特征不在于,包括下列步骤:

[0024] a)获取基于多种传感器数据的待识别样本;

[0025] b)读取多层次异构的交通模式行为识别模型的分类树,将根节点作为当前节点;

[0026] c)根据当前节点子分类问题所对应的最优特征子集,生成基于最优特征子集的测试样本;

[0027] d)根据当前子分类问题所对应的最优分类模型,对基于最优特征子集的测试样本进行分类,得到下一层的类别;

[0028] e)如果所识别出的下一层类别为最底层类别,输出识别结果,转步骤f);否则,将所识别出的下一层类别的节点作为当前节点,待识别样本进入当前节点,转步骤c);

[0029] 其中所述交通模式行为识别模型是根据前文所述的识别模型构建方法所生成的识别模型。

[0030] 与现有技术相比,本发明具有下列技术效果:

[0031] 1、本发明能够更加准确地区分具体的交通模式行为,例如:走路、跑步、骑车、乘公交车、开车、乘地铁、乘火车、乘轻轨等。

[0032] 2、本发明的计算复杂度相对较小。

## 附图说明

[0033] 以下,结合附图来详细说明本发明的实施例,其中:

[0034] 图1示出了本发明一个实施例中建立和训练该识别模型的过程的流程图;

[0035] 图2示出了本发明一个实施例中的多层次交通模式行为识别的构架图;

[0036] 图3示出了本发明一个实施例中每个特定分类的特征选择过程的流程图。

## 具体实施方式

[0037] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图,对本发明提出的基于多源传感器的层次异构交通模式行为识别模型及相应的交通模式行为识别方法进一步详细说明。应当理解,此处所描述的具体实施方式仅仅用以解释本发明,并不用于限定本发明。

[0038] 根据本发明的一个实施例,提供了一种基于多源传感器的层次异构交通模式行为识别模型,图1示出了建立和训练该识别模型的过程,它主要包括三个步骤:层次划分、特征选择、模型选择。下面分别介绍这三个步骤。

[0039] 一、层次划分

[0040] 交通模式行为包含多种,如走路、骑车、乘公交车、开车、乘地铁、乘火车、乘轻轨等。基于传感器数据对这些交通模式行为进行识别,其主要难度在于:

[0041] 1)交通模式行为的类别个数多。日常生活中,用户的交通模式行为包括:“走路”、“跑步”、“骑车”、“乘公交车”、“开车”、“乘地铁”、“乘火车”、“乘轻轨”等。同时识别多种交通模式行为的难度较大。

[0042] 2)不同交通模式行为之间的相似度高。例如:在路况比较拥堵的情况下,利用常规的运动传感器采集“开车”与“乘公交车”两种行为的数据就比较难区分。

[0043] 3)同种交通模式行为内部的差异性大。用户在日常生活中的交通模式行为受路况、天气等多方面因素的影响,具有很大的差异性;另外,由于个体的差异性,不同用户的同种交通模式行为也会有所不同。

[0044] 为了提高识别的精度,降低计算复杂度,本实施例中采用了一种分层次的异构识别模型。分层次是指将上述具体的交通模式行为进行多层次的归类,把对具体的交通模式行为的识别问题转化为多层识别问题。例如,本实施例中,按照运动方式的不同将交通模式行为识别自上而下划分为三个层次:第一层识别当前的交通模式行为是“静止行为”或“非静止行为”;针对第一层识别出的“非静止行为”,第二层将其划分为“机动行为”或“非机动行为”;针对第二层识别出的“机动行为”和“非机动行为”,第三层分别将“机动行为”和“非机动行为”进一步识别为具体的交通模式行为,即“走路”、“跑步”、“骑车”、“乘公交车”、“开

车”、“乘地铁”、“乘火车”、“乘轻轨”等。图2示出了本实施例的多层次交通模式行为识别的构架图。如图2所示,“机动行为”包括“乘公交车”、“开车”、“乘地铁”、“乘火车”、“乘轻轨”等,“非机动行为”包括“走路”、“跑步”、“骑车”等。层次划分实际上就是建立一个分类树,该分类树包括第一层节点(例如:“静止行为”节点和“非静止行为”节点),第二层节点(例如:“机动行为”或“非机动行为”节点),以及第三层节点(即叶节点,例如:“走路”、“跑步”、“骑车”、“乘公交车”、“开车”、“乘地铁”、“乘火车”、“乘轻轨”等节点)。上述每一个节点都代表一个类别,父节点下的子节点代表对父节点类别进一步分类后所得到的子类。这样,交通模式行为识别问题就转化为了多个分类子问题。采用多个子分类模型分别解决分类树中每一个父节点处的分类问题,即可识别出具体的交通模式行为。

## [0045] 二、特征选择

[0046] 本实施例中,为了解决分类树中各个父节点处的分类问题,首选需要采集各类交通模式行为已知的多源传感数据并对这些数据进行预处理和特征提取,生成用于交通模式行为识别的训练样本集,具体包括下列步骤:

[0047] 步骤101:采集不同交通模式行为(如静止、走路、跑步、骑车、乘公交、乘地铁、乘火车等)的多种不同来源的传感器(如加速度计、陀螺仪、磁力计等)数据,利用长度固定的窗口对每种传感器数据分别进行截取;

[0048] 步骤102:针对每个窗口的传感器数据,利用预处理算法对数据进行处理。预处理算法包括但不限于空缺值填补、数据滤波等。空缺值填补用于补充丢失的数据,使之成为连续的时间序列数据;数据滤波用于过滤掉孤立点数据或高频噪声数据等;

[0049] 步骤103:从预处理之后的窗口数据提取特征,典型的特征包括但不限于均值、标准方差、过零率、百分位数、关联系数、功率谱密度、频域熵、谱峰位置等;

[0050] 步骤104:将具有相同时间标签的窗口提取特征,按固定次序(如加速度特征、陀螺仪特征、磁力计特征等)组成向量,并加上对应交通模式行为的类别标号,生成训练样本集。

[0051] 在生成训练样本集后,针对分类树中每一个父节点处的分类问题(例如:已知样本属于“非静止行为”,如何将其进一步划分至“机动行为”或者“非机动行为”的分类问题;或者已经样本属于“非机动行为”,如何将其进一步划分至“走路”、“跑步”、“骑车”等具体的交通模式行为的分类问题),基于训练样本集选出适合于这一分类问题的特征集,即提取出适合于该父节点分类问题的特征组合。

[0052] 本实施例中,基于随机森林模型选择最优特征集。假设包含前文步骤103所有特征的特征集为FeatureSet1,那么每个特定分类的特征选择过程如图3所示,包括下列子步骤:

[0053] 步骤201:根据特征集FeatureSet1,生成训练样本集;所述训练样本集是所述测试样本所组成的集合;

[0054] 步骤202:根据FeatureSet1训练随机森林模型RandomForest1,并计算模型的识别精度Accuracy1;本实施例中,模型识别精度=模型准确识别出类别的测试样本个数/测试样本总个数;

[0055] 步骤203:统计特征集FeatureSet1中各个特征在模型RandomForest1中的使用频率,将特征按照频率从大到小排序;其中特征使用频率是:随机森林中使用该特征的决策树个数除以随机森林中包含的决策树总个数;

[0056] 步骤204:从特征集FeatureSet1中删掉使用频率取值较小的特征,剩余的特征组

成特征集FeatureSet2;本实施例中,每次可删掉预设比例(例如特征总个数的20%)的使用频率取值较小的特征;

[0057] 步骤205:根据基于特征集FeatureSet2的训练样本集训练随机森林模型RandomForest2,计算该随机森林模型RandomForest2的识别精度Accuracy2;

[0058] 步骤206:对比Accuracy2与Accuracy1,若Accuracy2小于Accuracy1,则进入步骤208;反之,进入步骤207;

[0059] 步骤207:用特征集FeatureSet2替换FeatureSet1,模型RandomForest2替换模型RandomForest1,精度Accuracy2替换精度Accuracy1,跳转到步骤203;

[0060] 步骤208:将特征集FeatureSet1作为最优特征集。

[0061] 三、模型选择

[0062] 在确定每一个父节点处的分类问题所对应的特征集后,基于该特征集选择一个最优分类模型作为该父节点所对应的子分类模型。这样,每个子分类模型用于将待识别的样本从一个类别划分至其下的一个子类别。例如:从“非静止行为”划分至“机动行为”或者“非机动行为”;或者从“非机动行为”划分至“走路”、“跑步”、“骑车”等具体的交通模式行为。多个这样的子分类模型就组成了分层次的异构识别模型。

[0063] 本步骤中,在确定各个子分类模型所对应的特征集后,进一步为各个子分类问题选择最优的分类模型,其方法是:对于每一层的每个子分类模型,利用包含最优特征集的样本,训练多个候选分类模型,候选分类模型包括但不限于支持向量机、朴素贝叶斯、决策树、K近邻、神经网络等分类模型。计算各个分类模型的召回率,将召回率取值最高的分类模型作为该特定分类的最优分类模型。需说明的是,本实施例中虽然采用了召回率来选择最优分类模型,但本发明并不限于此,在替换的实施例中,也可以采用其它的指标函数来选择最优分类模型。

[0064] 使用上述方法,能够针对不同的子分类模型选择最适合的特征组合,以及最适合的分类模型,这样能够更好地适应交通模式行为识别的特点,达到在提高识别精度的同时降低计算复杂度的技术效果。

[0065] 为了更加便于理解,以下结合一个具体示例进行说明,应当理解,该示例仅是示意性地,不应解释为对本发明保护范围的限制。该示例如下:

[0066] 采集加速度计、陀螺仪、磁力计和GPS四种传感器数据,基于每种传感器读数分别提取一系列特征,包括:均值、标准差、最小值、最大值、谱能量、过均值点个数、谱熵、功率谱密度等。另外,本示例中,GPS数据是指基于GPS定位技术所获得的速度测量值。

[0067] 在第一层区分“静止行为”与“非静止行为”时,特征集包含加速度的均值、标准差、最小值、最大值等特征,识别模型使用决策树;

[0068] 在第二层区分“机动行为”与“非机动行为”时,特征集包含加速度与陀螺仪的谱能量、过均值点个数、谱熵、功率谱密度等特征,识别模型使用朴素贝叶斯;

[0069] 在第三层区分“乘公交、乘地铁、乘火车”等具体的“机动行为”时,特征集包含加速度、磁力计的所有特征,识别模型使用随机森林;

[0070] 在第三层区分“走路、跑步、骑车”等具体的“非机动行为”时,特征集包含加速度、陀螺仪的所有特征,识别模型使用支持向量机。

[0071] 进一步地,根据本发明的另一实施例,还提供了一种基于上述多层次异构交通模



式行为识别模型的交通模式行为识别方法,包括下列步骤:

[0072] 步骤401:利用手机内嵌的加速度计、陀螺仪、磁力计等采集用户的交通模式行为数据;

[0073] 步骤402:对采集的传感器数据进行预处理;

[0074] 步骤403:针对预处理之后的每种传感器数据提取特征,组成待识别样本;

[0075] 步骤404:读取多层次异构交通模式行为识别模型的分类树,待识别样本进入分类树的第零层(本实施例中将分类树的根节点作为第零层),执行步骤步骤405;

[0076] 步骤405:根据当前节点子分类问题所对应的最优特征子集,生成包含最优特征子集的测试样本;

[0077] 步骤406:根据当前子分类问题所对应的最优分类模型,对包含最优特征子集的测试样本进行分类,得到识别结果,即得到下一层的类别;

[0078] 步骤407:如果所识别出的下一层类别即为最底层类别(即叶节点的类别),输出识别结果,转步骤408;否则,将所识别出的下一层类别的节点作为当前节点,待识别样本进入当前节点,转步骤5);

[0079] 步骤408:如果需要进行继续检测,转步骤401,否则结束。

[0080] 在实际测试中,针对六种典型的交通模式行为(静止、走路、骑车、开车、乘公交车、乘火车、乘地铁),发明人采集了四种传感器(加速度计、电子罗盘、磁力计、GPS)数据。初步的实验结果显示,基于多源传感器的层次异构交通模式行为识别方法比传统的基于单个识别模型的方法在识别性能上有明显的提升(精度在90%以上)。

[0081] 最后应说明的是,以上实施例仅用以描述本发明的技术方案而不是对本技术方法进行限制,本发明在应用上可以延伸为其它的修改、变化、应用和实施例,并且因此认为所有这样的修改、变化、应用、实施例都在本发明的精神和教导范围内。

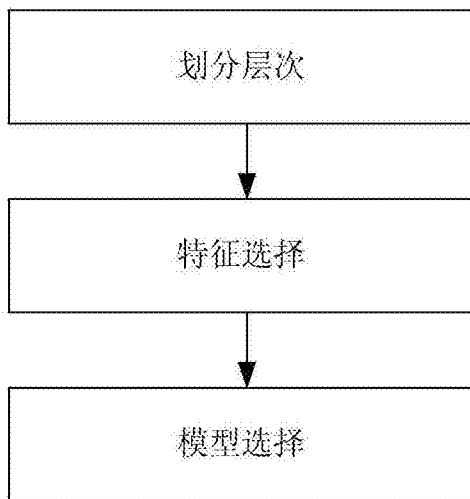


图1

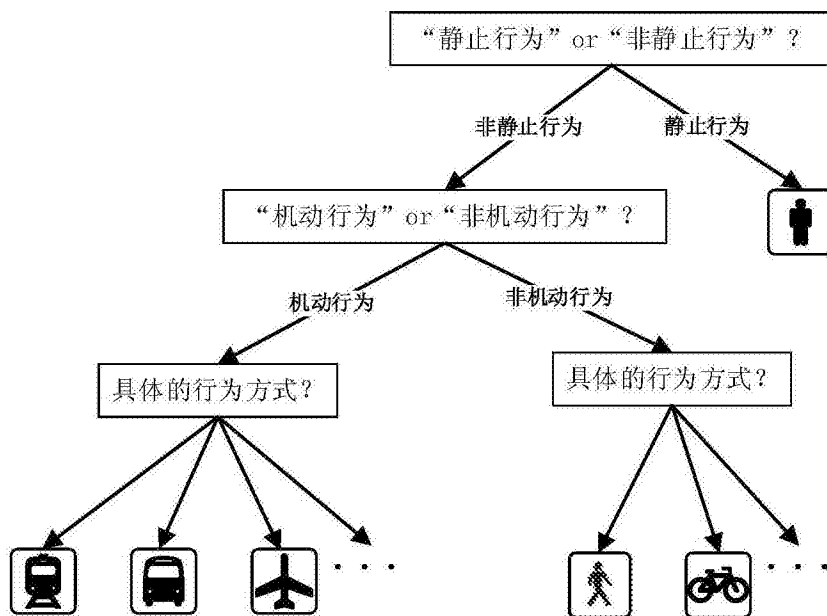


图2

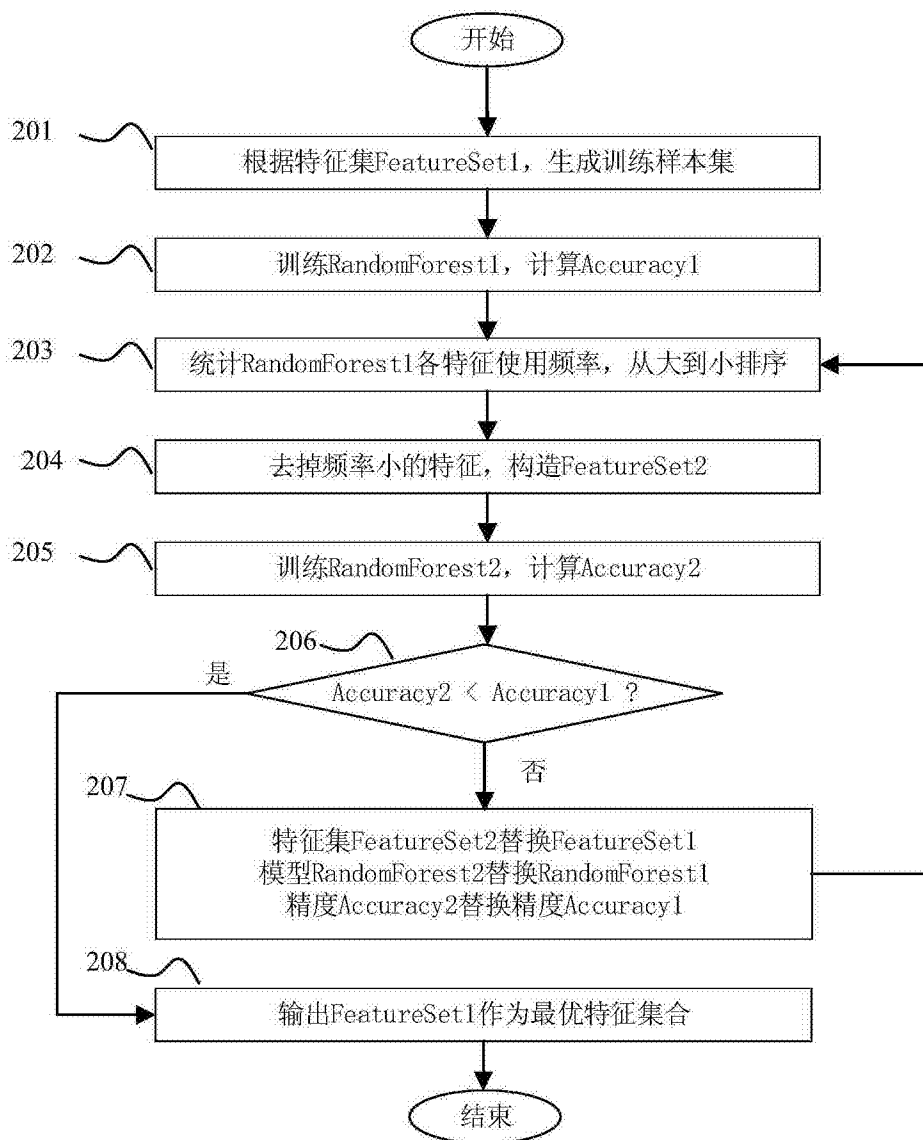


图3