



(12) 发明专利

(10) 授权公告号 CN 112541621 B

(45) 授权公告日 2021.08.31

(21) 申请号 202011392382.9

G06F 16/29 (2019.01)

(22) 申请日 2020.12.02

G06K 9/62 (2006.01)

(65) 同一申请的已公布的文献号

(56) 对比文件

申请公布号 CN 112541621 A

CN 104239556 A, 2014.12.24

(43) 申请公布日 2021.03.23

审查员 余汉鸣

(73) 专利权人 香港理工大学深圳研究院

地址 518057 广东省深圳市南山区粤海街  
道高新技术产业园南区粤兴一道18号  
香港理工大学产学研大楼205室

(72) 发明人 史文中 沈泉麒

(74) 专利代理机构 深圳市君胜知识产权代理事

务所(普通合伙) 44268

代理人 陈专

(51) Int. Cl.

G06Q 10/04 (2012.01)

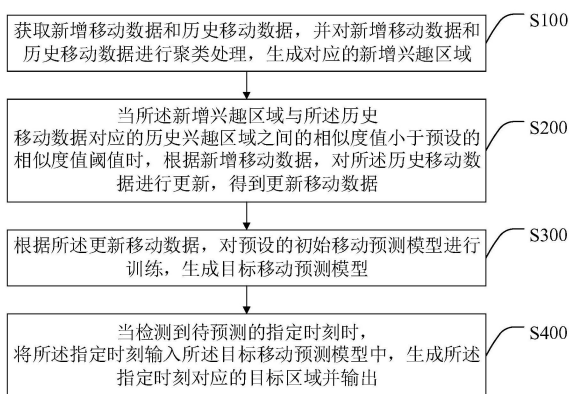
权利要求书3页 说明书15页 附图3页

(54) 发明名称

一种移动预测方法、智能终端及存储介质

(57) 摘要

本发明公开了一种移动预测方法、智能终端及存储介质,所述方法包括:获取新增移动数据和历史移动数据,并对所述新增移动数据和所述历史移动数据进行聚类处理,生成对应的新增兴趣区域;当所述新增兴趣区域与预设的历史兴趣区域之间的相似度值符合预设的相似度规则时,根据所述新增移动数据,对所述历史移动数据进行更新,得到更新移动数据;根据所述更新移动数据,对预设的初始移动预测模型进行训练,生成目标移动预测模型;当检测到待预测的指定时刻时,将所述指定时刻输入所述目标移动预测模型中,生成所述指定时刻对应的目标区域并输出。本发明在进行移动预测时,结合新旧数据,对兴趣区域进行实时更新,以提高后续预测结果的准确性。



1. 一种移动预测方法,其特征在于,所述移动预测方法包括:

获取新增移动数据和历史移动数据,并对所述新增移动数据和所述历史移动数据进行聚类处理,生成对应的新增兴趣区域;

当所述新增兴趣区域与预设的历史兴趣区域之间的相似度值符合预设的相似度规则时,根据所述新增移动数据,对所述历史移动数据进行更新,得到更新移动数据,其中,所述历史兴趣区域为根据所述历史移动数据聚类得到的兴趣区域;

根据所述更新移动数据,对预设的初始移动预测模型进行训练,生成目标移动预测模型;其中,所述根据所述更新移动数据,对预设的初始移动预测模型进行训练,生成目标移动预测模型,具体包括:

对所述更新移动数据进行去重复处理,得到有效移动数据;

对所述有效移动数据进行聚类处理,生成对所述更新移动数据对应的更新兴趣区域;

将各个所述有效移动数据对应的更新兴趣区域作为对应的当前访问区域,并根据各个所述有效移动数据对应的更新兴趣区域以及所述有效移动数据之间的移动关系,提取各个所述有效移动数据对应的空间特征和时间特征,得到各个有效移动数据对应的移动事件;

根据所述移动事件,对预设的初始预测模型进行训练,生成目标移动预测模型;

当检测到待预测的指定时刻时,将所述指定时刻输入所述目标移动预测模型中,生成所述指定时刻对应的目标区域并输出。

2. 根据权利要求1所述的移动预测方法,其特征在于,所述更新移动数据包括采集时刻和采集坐标;所述对所述更新移动数据进行去重复处理,得到有效移动数据,具体包括:

根据所述采集时刻的先后顺序,对所述更新移动数据进行排序,生成移动数据序列;

依次计算所述移动数据序列中第N更新移动数据和第M更新移动数据之间的采集时刻的差值,得到移动时间,其中,N和M为小于等于所述更新移动数据的数量的正整数,且N小于M;

当所述移动时间小于等于预设的移动时间阈值时,根据所述第N更新移动数据和所述第M更新移动数据之间的采集坐标,计算对应的移动距离;

当所述移动距离小于等于预设的移动距离阈值时,将所述第N更新移动数据作为有效移动数据。

3. 根据权利要求1所述的移动预测方法,其特征在于,所述空间特征包括先前访问区域;所述将各个所述有效移动数据对应的更新兴趣区域作为对应的当前访问区域,并根据各个所述有效移动数据对应的更新兴趣区域以及所述有效移动数据之间的移动关系,提取各个所述有效移动数据对应的空间特征和时间特征,得到各个有效移动数据对应的移动事件,具体包括:

根据所述有效移动数据的采集时刻的先后顺序,对各个所述有效移动数据进行排序,生成有效移动数据序列;

将所述有效移动数据序列中第i有效移动数据对应的更新兴趣区域作为所述有效移动数据序列中第i+1有效移动数据对应的先前访问区域,将所述有效移动数据序列中第i+1有效移动数据对应的更新兴趣区域作为对应的当前访问区域,并根据所述先前访问区域和所述当前访问区域,生成所述第i有效移动数据对应的空间特征,以及

根据预设的时刻划分规则和各个所述有效移动数据的采集时刻,生成各个所述有效移

动数据对应的时间特征,其中, $i$ 为小于所述有效移动数据的数量的正整数。

4. 根据权利要求3所述的移动预测方法,其特征在于,所述时刻划分规则包括周划分规则和日划分规则。

5. 根据权利要求1所述的移动预测方法,其特征在于,所述初始预测模型包括特征权重对象,所述特征权重对象包括所述空间特征和所述时间特征;所述根据所述移动事件,对预设的初始预测模型进行训练,生成目标移动预测模型,具体包括:

针对每一个所述特征权重对象,根据所述移动事件,计算该特征权重对象对应各个所述更新兴趣区域的特征训练概率值;

将数值最大的特征训练概率值对应的更新兴趣区域作为特征训练预测区域,并针对每一个所述移动事件,根据该移动事件中的当前访问区域、所述特征训练预测区域和预设的权重调整公式,对该特征权重对象对应的权重值进行调整,生成目标移动预测模型。

6. 根据权利要求5所述的移动预测方法,其特征在于,所述初始预测模型包括多个候选预测模型,所述根据所述移动事件,对预设的初始预测模型进行训练,生成目标移动预测模型,具体包括:

针对每一个所述移动事件,基于所述候选预测模型、该移动事件中的时间特征和先前访问区域,计算该移动事件对应各个所述更新兴趣区域的模型训练概率值;

将数值最大的模型训练概率值对应的更新兴趣区域为模型训练预测区域作为该移动事件对应的模型训练预测区域,并根据该移动事件中的当前访问区域、所述训练预测区域和预设的权重调整公式,对各个所述候选模型对应的权重值进行调整,得到各个候选模型对应的目标权重值;

将数值最大的目标权重值对应的候选模型作为所述移动预测模型。

7. 根据权利要求6所述的移动预测方法,其特征在于,所述权重调整公式

$$\begin{cases} W_{\text{object}} = \frac{W_{\text{object}}}{c^{1+(p-0.5)}} \text{ 当 } r_{\text{max}} = r_{\text{ture}} \\ W_{\text{object}} = W_{\text{object}} \times c \text{ 当 } r_{\text{max}} \neq r_{\text{ture}} \end{cases}, \text{ 其中, object 为权重对象, 所述权重对象包括所}$$

述候选预测模型和所述特征权重对象,  $w_{\text{object}}$  为各个所述特征权重对象对应的权重值,  $c$  为系数,  $p$  为  $r_{\text{max}}$  对应的训练概率值, 所述训练概率值包括特征训练概率值和模型训练概率值,  $r_{\text{max}}$  为训练预测区域, 所述训练预测区域包括特征预测区域和模型预测区域,  $r_{\text{ture}}$  为各个所述移动事件中的当前访问区域。

8. 根据权利要求6所述的移动预测方法,其特征在于,所述候选模型包括带有自适应窗口的朴素贝叶斯模型和带有衰减因子的朴素贝叶斯模型;基于所述带有衰减因子的朴素贝叶斯模型,预测测试事件发生的测试概率值为

$$P(r_{\text{st}} = r_i | x_1 = v_{x_1}, \dots, x_n = v_{x_n}, t_1 = v_{t_1}, \dots, t_m = v_{t_m}) \propto \prod_j^n \frac{N_{v_{x_j}, r_i}}{N_{r_i}} \prod_k^m \frac{N_{v_{t_k}, r_i}}{N_{r_i}} \frac{N_{r_i}}{N_{r_{\text{st}}}}$$

其中,所述测试事件为所述空间特征为  $v_{x_1}$ 、 $v_{x_2}$ 、 $\dots$ 、 $v_{x_n}$ , 所述当前访问区域为  $r_i$ , 且所述时间特征为  $v_{t_1}$ 、 $v_{t_2}$ 、 $\dots$ 、 $v_{t_m}$  的移动事件,  $n$  为所述空间特征的数量正整数,  $m$  为所述时间特征的数量正整数;

$r_{st}$ 为所述更新兴趣区域,  $r_i$ 为所述测试事件中的当前访问区域,  $x_j$ 为所述空间特征,  $t_k$ 为所述时间特征;

$N_{v_{x_j}, r_i}$ 为所述空间特征为 $v_{x_j}$ 且所述当前访问区域为 $r_i$ 的移动事件对应的计数之和,  $N_{v_{t_k}, r_i}$ 为所述时间特征为 $t_k$ 且所述当前访问区域为 $r_i$ 的移动事件对应的计数之和,  $N_{r_i}$ 为所述当前访问区域为 $r_i$ 的移动事件对应的计数之和,  $N_{r_{st}}$ 为所有的移动事件对应的计数之和;

当所述测试事件对应的测试时刻为 $t$ 时, 各个所述移动事件对应的计数公式为  $W_{i_o, t} = \lambda^{t-t_{i_o}}$ ,  $i_o$ 为所述移动事件,  $\lambda$ 为预设的衰减因子。

9. 根据权利要求5所述的移动预测方法, 其特征在于, 所述当所述新增兴趣区域与所述历史移动数据对应的历史兴趣区域之间的相似度值符合预设的相似度规则时, 根据所述新增移动数据, 对所述历史移动数据进行更新, 得到更新移动数据, 具体包括:

当所述新增移动数据的数量大于等于预设的新增数据阈值时, 将所述新增移动数据和所述历史移动数据进行聚类处理, 生成多个新增兴趣区域;

根据各个所述新增兴趣区域和各个所述历史兴趣区域, 对所述新增移动数据和所述历史移动数据分别进行区域标注, 得到与各个所述新增兴趣区域对应的新增区域数据集和与各个所述历史兴趣区域对应的历史区域数据集;

针对每一个新增区域数据集, 计算该新增区域数据集与各个所述历史区域数据集之间的区域相似度值;

将大于等于预设的区域相似度阈值的区域相似度值对应的新增区域数据集和历史兴趣区域集标注为同一区域标签;

根据各个所述新增区域数据集和各个所述历史区域数据集对应的区域标签, 计算所述新增区域数据集和所述历史兴趣区域集之间的整体相似度值;

当所述整体相似度值小于预设的整体相似度值时, 根据所述新增移动数据, 对所述历史移动数据进行更新, 得到更新移动数据。

10. 根据权利要求9所述的移动预测方法, 其特征在于, 所述当所述新增移动数据的数量大于等于预设的新增数据阈值时, 将所述新增移动数据和所述历史移动数据进行聚类处理, 生成多个新增兴趣区域之前, 还包括:

根据所述新增移动数据的采集坐标, 判断是否存在与所述新增移动数据对应的历史兴趣区域;

若是, 则根据所述新增移动数据, 对所述初始移动预测模型的权重值进行调整。

11. 一种智能终端, 其特征在于, 所述智能终端包括: 存储器、处理器及存储在所述存储器上并可在所述处理器上运行的移动预测程序, 所述移动预测程序被所述处理器执行时实现如权利要求1-10任一项所述的移动预测方法的步骤。

12. 一种存储介质, 其特征在于, 所述存储介质存储有移动预测程序, 所述移动预测程序被处理器执行时实现如权利要求1-10任一项所述的移动预测方法的步骤。

## 一种移动预测方法、智能终端及存储介质

### 技术领域

[0001] 本发明涉及地理信息技术领域,尤其涉及一种移动预测方法、智能终端及存储介质。

### 背景技术

[0002] 个人移动的研究是智慧城市、社会科学等领域的重要问题之一,具有广泛的应用,例如根据个人的移动实时提供天气预报。目前基于个人移动的预测方法有基于马尔科夫链的、基于序列模型的、基于分类模型的、及基于其他模型如神经网络的四大类。其中基于分类模型的方法具有综合利用时空特征且在稀疏、密集数据上都较为适用的优点。但是目前的基于分类模型的预测方式较为固定,缺少对个人行为变化的考虑,例如以下两点:

[0003] (1) 活动区域的变化:随着时间的推移,个人可能由于搬家、转校等原因,活动区域范围发生了变化,包括探索新的活动区域,以及离开旧的活动区域,现有方法无法实时更新活动区域,同时检测其变化;

[0004] (2) 区域内移动规律的变化:由于时间(如季节)、交通(道路维修)、个人偏好等原因,个人对区域的访问时间、访问顺序等可能会产生变化,现有方法无法检测移动规律变化。

### 发明内容

[0005] 本发明的主要目的在于提供一种移动预测方法、智能终端及存储介质,旨在解决现有技术中个人移动模式变化带来的移动预测不精确的问题,以提高移动预测的可靠度和精确性。

[0006] 为实现上述目的,本发明提供一种移动预测方法,所述移动预测方法包括如下步骤:

[0007] 获取新增移动数据和历史移动数据,并对所述新增移动数据和所述历史移动数据进行聚类处理,生成对应的新增兴趣区域;

[0008] 当所述新增兴趣区域与预设的历史兴趣区域之间的相似度值符合预设的相似度规则时,根据所述新增移动数据,对所述历史移动数据进行更新,得到更新移动数据,其中,所述历史兴趣区域为根据所述历史移动数据聚类得到的兴趣区域;

[0009] 根据所述更新移动数据,对预设的初始移动预测模型进行训练,生成目标移动预测模型;

[0010] 当检测到待预测的指定时刻时,将所述指定时刻输入所述目标移动预测模型中,生成所述指定时刻对应的目标区域并输出。

[0011] 可选地,所述移动预测方法,其中,所述根据所述更新移动数据,对预设的初始移动预测模型进行训练,生成目标移动预测模型,具体包括:

[0012] 对所述更新移动数据进行去重复处理,得到有效移动数据;

[0013] 对所述有效移动数据进行聚类处理,生成对所述更新移动数据对应的更新兴趣区

域；

[0014] 将各个所述有效移动数据对应的更新兴趣区域作为对应的当前访问区域，并根据各个所述有效移动数据对应的更新兴趣区域以及所述有效移动数据之间的移动关系，提取各个所述有效移动数据对应的空间特征和时间特征，得到各个有效移动数据对应的移动事件；

[0015] 根据所述移动事件，对预设的初始预测模型进行训练，生成目标移动预测模型。

[0016] 可选地，所述移动预测方法，其中，所述更新移动数据包括采集时刻和采集坐标；所述对所述更新移动数据进行去重复处理，得到有效移动数据，具体包括：

[0017] 根据所述采集时刻的先后顺序，对所述更新移动数据进行排序，生成移动数据序列；

[0018] 依次计算所述移动数据序列中第N更新移动数据和第M更新移动数据之间的采集时刻的差值，得到移动时间，其中，N和M为小于等于所述更新移动数据的数量的正整数，且N小于M；

[0019] 当所述移动时间小于等于预设的移动时间阈值时，根据所述第N更新移动数据和所述第M更新移动数据之间的采集坐标，计算对应的移动距离；

[0020] 当所述移动距离小于等于预设的移动距离阈值时，将所述第N更新移动数据作为有效移动数据。

[0021] 可选地，所述移动预测方法，其中，所述空间特征包括先前访问区域；所述将各个所述有效移动数据对应的更新兴趣区域作为对应的当前访问区域，并根据各个所述有效移动数据对应的更新兴趣区域以及所述有效移动数据之间的移动关系，提取各个所述有效移动数据对应的空间特征和时间特征，得到各个有效移动数据对应的移动事件，具体包括：

[0022] 根据所述有效移动数据的采集时刻的先后顺序，对各个所述有效移动数据进行排序，生成有效移动数据序列；

[0023] 将所述有效移动数据序列中第i有效移动数据对应的更新兴趣区域作为所述有效移动数据序列中第i+1有效移动数据对应的先前访问区域，将所述有效移动数据序列中第i+1有效移动数据对应的更新兴趣区域作为对应的当前访问区域，并根据所述先前访问区域和所述当前访问区域，生成所述第i有效移动数据对应的空间特征，以及

[0024] 根据预设的时刻划分规则和各个所述有效移动数据的采集时刻，生成各个所述有效移动数据对应的时间特征，其中，i为小于所述有效移动数据的数量的正整数。

[0025] 可选地，所述移动预测方法，其中，所述时刻划分规则包括周划分规则和日划分规则。

[0026] 可选地，所述移动预测方法，其中，所述初始预测模型包括特征权重对象，所述特征权重对象包括所述空间特征和所述时间特征；所述根据所述移动事件，对预设的初始预测模型进行训练，生成目标移动预测模型，具体包括：

[0027] 针对每一个所述特征权重对象，根据所述移动事件，计算该特征权重对象对应各个所述更新兴趣区域的特征训练概率值；

[0028] 将数值最大的训练概率值对应的更新兴趣区域作为特征训练预测区域，并针对每一个所述移动事件，根据该移动事件中的当前访问区域、所述特征训练预测区域和预设的权重调整公式，对该特征权重对象对应的权重值进行调整，生成目标移动预测模型。

[0029] 可选地,所述移动预测方法,其中,所述初始预测模型包括多个候选预测模型,所述根据所述移动事件,对预设的初始预测模型进行训练,生成目标移动预测模型,具体包括:

[0030] 针对每一个所述移动事件,基于所述候选预测模型、该移动事件中的时间特征和先前访问区域,计算该移动事件对应各个所述更新兴趣区域的模型训练概率值;

[0031] 将数值最大的模型训练概率值对应的更新兴趣区域为模型训练预测区域作为该移动事件对应的模型训练预测区域,并根据该移动事件中的当前访问区域、所述训练预测区域和预设的权重调整公式,对各个所述候选模型对应的权重值进行调整,得到各个候选模型对应的目标权重值;

[0032] 将数值最大的目标权重值对应的候选模型作为所述移动预测模型。

[0033] 可选地,所述移动预测方法,其中,所述权重调整公式

$$\begin{cases} W_{\text{object}} = \frac{W_{\text{object}}}{c^{1+(p-0.5)}} \text{ 当 } r_{\text{max}} = r_{\text{ture}} \\ W_{\text{object}} = W_{\text{object}} \times c \text{ 当 } r_{\text{max}} \neq r_{\text{ture}} \end{cases}, \text{ 其中, object 为权重对象, 所述权重对象包括所}$$

述候选预测模型和所述特征权重对象,  $w_{\text{object}}$  为各个所述特征权重对象对应的权重值,  $c$  为系数,  $p$  为  $r_{\text{max}}$  对应的训练概率值, 所述训练概率值包括特征训练概率值和模型训练概率值,  $r_{\text{max}}$  为训练预测区域, 所述训练预测区域包括特征预测区域和模型预测区域,  $r_{\text{ture}}$  为各个所述移动事件中的当前访问区域。

[0034] 可选地,所述移动预测方法,其中,所述候选模型包括带有自适应窗口的朴素贝叶斯模型和带有衰减因子的朴素贝叶斯模型;基于所述带有衰减因子的朴素贝叶斯模型,预测测试事件发生的测试概率值为

$$P(r_{\text{st}} = r_i | x_1 = v_{x_1}, \dots, x_n = v_{x_n}, t_1 = v_{t_1}, \dots, t_m = v_{t_m}) \propto$$

$$\prod_j^n \frac{N_{v_{x_j}, r_i}}{N_{r_i}} \prod_k^m \frac{N_{v_{t_k}, r_i}}{N_{r_i}} \frac{N_{r_i}}{N_{r_{\text{st}}}};$$

[0035] 其中,所述测试事件为所述空间特征为  $v_{x_1}$ 、 $v_{x_2}$ 、 $\dots$ 、 $v_{x_n}$ , 所述当前访问区域为  $r_i$ , 且所述时间特征为  $v_{t_1}$ 、 $v_{t_2}$ 、 $\dots$ 、 $v_{t_m}$  的移动事件,  $n$  为所述空间特征的数量正整数,  $m$  为所述时间特征的数量正整数;

[0036]  $r_{\text{st}}$  为所述更新兴趣区域,  $r_i$  为所述测试事件中的当前访问区域,  $x_j$  为所述空间特征,  $t_k$  为所述时间特征;

[0037]  $N_{v_{x_j}, r_i}$  为所述空间特征为  $v_{x_j}$  且所述当前访问区域为  $r_i$  的移动事件对应的计数之和,  $N_{v_{t_k}, r_i}$  为所述时间特征为  $t_m$  且所述当前访问区域为  $r_i$  的移动事件对应的计数之和,  $N_{r_i}$  为所述当前访问区域为  $r_i$  的移动事件对应的计数之和,  $N_{r_{\text{st}}}$  为所有的移动事件对应的计数之和;

[0038] 当所述测试事件对应的测试时刻为  $t$  时, 各个所述移动事件对应的计数公式为  $W_{i_0, t} = \lambda^{t-t_{i_0}}$ ,  $i_0$  为所述移动事件,  $\lambda$  为预设的衰减因子。

[0039] 可选地,所述移动预测方法,其中,所述当所述新增兴趣区域与所述历史移动数据对应的历史兴趣区域之间的相似度值符合预设的相似度规则时,根据所述新增移动数据,对所述历史移动数据进行更新,得到更新移动数据,具体包括:

[0040] 当所述新增移动数据的数量大于等于预设的新增数据阈值时,将所述新增移动数据和所述历史移动数据进行聚类处理,生成多个新增兴趣区域;

[0041] 根据各个所述新增兴趣区域和各个所述历史兴趣区域,对所述新增移动数据和所述历史移动数据分别进行区域标注,得到与各个所述新增兴趣区域对应的新增区域数据集和与各个所述历史兴趣区域对应的历史区域数据集;

[0042] 针对每一个新增区域数据集,计算该新增区域数据集与各个所述历史区域数据集之间的区域相似度值;

[0043] 将大于等于预设的区域相似度阈值的区域相似度值对应的新增区域数据集和历史兴趣区域集标注为同一区域标签;

[0044] 根据各个所述新增区域数据集和各个所述历史区域数据集对应的区域标签,计算所述新增区域数据集和所述历史兴趣区域集之间的整体相似度值;

[0045] 当所述整体相似度值小于预设的整体相似度值时,根据所述新增移动数据,对所述历史移动数据进行更新,得到更新移动数据。

[0046] 可选地,所述移动预测方法,其中,所述当所述新增移动数据的数量大于等于预设的新增数据阈值时,将所述新增移动数据和所述历史移动数据进行聚类处理,生成多个新增兴趣区域之前,还包括:

[0047] 根据所述新增移动数据的采集坐标,判断是否存在与所述新增移动数据对应的历史兴趣区域;

[0048] 若是,则根据所述新增移动数据,对所述初始移动预测模型的权重值进行调整。

[0049] 此外,为实现上述目的,本发明还提供一种智能终端,其中,所述智能终端包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的移动预测程序,所述移动预测程序被所述处理器执行时实现如上所述的移动预测方法的步骤。

[0050] 此外,为实现上述目的,本发明还提供一种存储介质,其中,所述存储介质存储有移动预测程序,所述移动预测程序被处理器执行时实现如上所述的移动预测方法的步骤。

[0051] 与现有的移动预测方法相比,本发明考虑移动模式的变化包括:

[0052] (1) 为了解决活动区域的变化问题,从历史数据中提取兴趣区域后,每次获取新数据,实时更新兴趣区域,并提出了一种相似性度量方法来检测兴趣区域是否发生变化;

[0053] (2) 为了克服活动区域内移动规律的变化问题,提出里一种检测模式变化并削弱过时数据影响的预测模型,该模型基于一种新提出的分类模型,并将其与一种已有的流数据分类方法结合,增强了预测模型的稳定性,其中特征与模型的权重通过一种全新的方法进行实时更新。

[0054] 通过以上两点,本发明实施例能够取得较优的预测对象移动预测结果。

## 附图说明

[0055] 图1是本发明移动预测方法提供的较佳实施例的流程图;

[0056] 图2是本发明移动预测方法提供的较佳实施例中模型训练过程的示意图;



[0057] 图3是本发明移动预测方法提供的较佳实施例中历史移动数据进行更新过程的示意图；

[0058] 图4是本发明移动预测方法提供的较佳实施例中移动模式发生改变的示意图；

[0059] 图5为本发明智能终端的较佳实施例的运行环境示意图。

### 具体实施方式

[0060] 为使本发明的目的、技术方案及优点更加清楚、明确，以下参照附图并举实施例对本发明进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

[0061] 本发明较佳实施例所述的移动预测方法，如图1所示，所述移动预测方法包括以下步骤：

[0062] 步骤S100，获取新增移动数据和历史移动数据，并对所述新增移动数据和所述历史移动数据进行预处理，生成对应的新增兴趣区域。

[0063] 具体地，移动数据是指预先保存通过智能终端采集的与预测对象定位相关的数据，例如社交媒体数据、手机信令数据、GPS定位数据等。当预测对象，也就是智能终端的用户，在通过智能终端在社交网站发送信息时，同时会通过全球导航卫星系统(Global Navigation Satellite System, GNSS)、信令、蓝牙、wifi等方式获取当前的坐标，从而显示预测对象当前的位置。手机信令具有全覆盖性，只要智能终端开启，基站会获取对应的信号，从而产生手机信令。针对手机信令数据，可预先将不同的基站进行编号，由于基站的基站位置坐标是固定的，因此将各个基站位置坐标与基站编号对应后，获取手机信令时，根据其对应的基站编号即可确定对应的基站位置坐标，并将该基站位置坐标作为智能终端对应的位置坐标。每一条移动数据都至少由两部分组成，一部分是采集时刻，一部分是位置坐标，采集时刻是指该移动数据生成的时刻，位置坐标是指该采集时刻智能终端所在位置。除这两部分外，还可对不同的智能终端进行预测对象编号，以标识不同的预测对象，从而做到同时采集多人的历史移动数据。将同一预测对象编号对应的历史移动数据作为一个历史移动数据集，以进行后续的移动预测。为描述简洁，本实施例以单一预测对象进行描述。

[0064] 本实施例中，移动数据分为历史移动数据和新增移动数据，历史移动数据是指在前一次用于确定兴趣区域的移动数据，新增移动数据就是指在确定历史兴趣区域后采集到的与智能终端定位相关的数据。其中，通过对历史移动数据进行聚类得到的兴趣区域为历史兴趣区域。当获取新增移动数据和历史移动数据时，根据新增移动数据和历史移动数据中的定位信息，对新增移动数据和历史移动数据进行聚类，从而生成多个与新增移动数据和历史移动数据对应的新增兴趣区域。值得注意的是，这里的聚类对象新增移动数据和历史移动数据这一整体，并非分别对新增移动数据和历史移动数据聚类。

[0065] 其中，聚类处理可采用聚类算法实现或无监督学习的分类模型实现。以聚类算法为例，聚类算法是根据数据之间的相似性将数据分为不同的类别。将新增移动数据和历史移动数据作为处理移动数据，预测对象的移动一般有一定的规律，例如工作时间主要集中在办公区域。因此可通过聚类算法，将处理移动数据分为多组，将预测对象的移动描述为从一个组到另一个组的过程。由于每一处理移动数据在空间上表现为一个点，因此最后得到的处理移动数据集可代表一个区域，预测对象在区域中出现多次，因此将其作为兴趣区域，

同时聚类方法可采用基于密度的聚类算法,对处理移动数据继续聚类。常用的基于密度的聚类算法包括具有噪声的基于密度的聚类方法(Density-Based Spatial Clustering of Applications with Noise,DBSCAN)、密度最大值聚类算法(Maximum Density Clustering Application,MDCA)等。但DBSCAN需要确定两个参数,难以自适应选择参数,而采用固定参数会无法适应多个预测对象。因此本实施例优选采用基于密度的噪声应用空间聚类(Hierarchical Density-Based Spatial Clustering of Applications with Noise,HDNSCAN),HDNSCAN仅需要一个主要参数,且已在提取兴趣区域中被证明有效,也有研究报道根据聚类数量要求和聚类程度实现参数自适应,因此本实施例采用该算法将处理移动数据分为多个处理移动数据集。针对每一个处理移动数据集,根据该处理移动数据集中的处理移动数据的采集坐标,生成该处理移动数据集对应的数据区域,该数据区域可以将对应的处理移动数据集中的所有移动数据包括在内,生成数据区域可采用凸包算法等方式。最后将所述数据区域作为该处理移动数据集中各个处理移动数据对应的更新兴趣区域。

[0066] 步骤S200,当所述新增兴趣区域与预设的历史兴趣区域之间的相似度值符合预设的相似度规则时,根据所述新增移动数据,对所述历史移动数据进行更新,得到更新移动数据,其中,所述历史兴趣区域为根据所述历史移动数据聚类得到的兴趣区域。

[0067] 具体地,在获取历史移动数据后,先对所述历史移动数据进行聚类处理,得到对应的历史兴趣区域。聚类处理的方式可采用上述可实现聚类效果的处理方式,在此不再赘述。所述相似度规则为用于衡量所述新增兴趣区域与所述历史兴趣区域之间是否相似的规则。例如预设一个相似度阈值用于评价新增兴趣区域和之前历史兴趣区域之间的相似度值。若所述新增兴趣区域与所述历史移动数据对应的历史兴趣区域之间的相似度值小于预设的相似度值阈值,则说明两者并不相似,因此新增移动数据对预测对象的移动行为产生了较大干扰。故而将新增移动数据增加至历史移动数据中,以对所述历史移动数据进行更新,得到更新移动数据。其中,所述区域和区域之间的相似度值可根据两个区域所对应的数据的重合率计算,也可采用两个区域所对应的面积重合率等方式。

[0068] 进一步地,为提高计算所述新增兴趣区域与所述历史移动数据之间相似度的准确性,步骤S200包括:

[0069] 步骤A10,当所述新增移动数据的数量大于等于预设的新增数据阈值时,将所述新增移动数据和所述历史移动数据进行聚类处理,生成多个新增兴趣区域;

[0070] 具体地,在获取新增移动数据和历史移动数据时,计算新增移动数据对应的历史兴趣区域。获得的新增移动数据中应包含预测对象编号、时间、位置,基于新增移动数据的位置,推测其所对应的历史兴趣区域,结果可能为不属于任何历史兴趣区域,判定为噪声。

[0071] 当新增移动数据达到预设的新增数据阈值 $\eta$ 时,将所述新增移动数据和所述新增移动数据进行聚类处理,生成多个新增兴趣区域。

[0072] 进一步地,当所述新增移动数据的数量小于预设的新增数据阈值时,也就是新增移动数据还未积累到一定程度时,也可根据新增移动数据对初始移动预测模型进行调整。具体过程为:

[0073] 根据所述新增移动数据的采集坐标,判断是否存在与所述新增移动数据对应的兴趣区域;

[0074] 若是,则根据所述新增移动数据,对所述初始移动预测模型的权重值进行调整。

[0075] 其中,对初始预测模型的权重值的调整在后文对初始预测模型的训练中有详细描述,在此不再赘述。

[0076] 步骤A20,根据各个所述新增兴趣区域和各个所述历史兴趣区域,对所述新增移动数据和所述历史移动数据分别进行区域标注,得到与各个所述新增兴趣区域对应的新增区域数据集和与各个所述历史兴趣区域对应的历史区域数据集;

[0077] 具体地,新增兴趣区域和历史兴趣区域两组兴趣区域,历史兴趣区域集合为 $RS_p$ ,大小为 $k_p$ ,其中的历史兴趣区域称为 $r_i$  ( $0 \leq i \leq k_p$ );新增兴趣区域集合 $RS_c$ ,大小为 $k_c$ ,其中的新增兴趣区域称为 $r_j$  ( $0 \leq j \leq k_c$ )。由于所述新增移动数据和所述历史移动数据所对应的兴趣区域可能不同,因此,将所述新增移动数据和所述历史移动数据进行不同的区域标注,分别得到新增区域数据集和历史区域数据集。

[0078] 例如,新增移动数据A对应的历史兴趣区域为区域1,而对应的新增移动区域为区域2。而进行区域标注后,所有新增区域数据集的数据量与历史区域数据集的数据量是相等的。其中,在历史兴趣区域或新增兴趣区域中,针对作为噪声的新增移动数据,可将其单独分为一个特殊的兴趣区域。为便于理解,本实施例中,以新增兴趣区域A、新增兴趣区域数据B、历史兴趣区域数据C,以及三者分别对应的新增区域数据集A、新增区域数据集B和历史区域数据集C为例进行描述。

[0079] 步骤A30,针对每一个新增区域数据集,计算该新增区域数据集与各个所述历史区域数据集之间的区域相似度值;

[0080] 具体地,计算新增区域数据集A和新增区域数据集B与历史区域数据集C之间的区域相似度值。区域相似度值可采用面积的重合率,数据集之间的相似性进行计算。本实施例优选采用的计算方式为根据新增区域数据集与历史区域数据集中移动数据的重合率来计算。

[0081] 计算公式可为 $S_r = \frac{N_{r_i \cap r_j}}{N_{r_i \cup r_j}}$ ,其中 $N_{r_i \cap r_j}$ 指新增区域数据集和历史区域数据集之

间,根据各个移动数据的采集坐标确定的地理位置相同的移动数据的交集, $N_{r_i \cup r_j}$ 指新增区域数据集和历史区域数据集之间,根据各个移动数据的采集坐标确定的地理位置相同的移动数据的并集; $S_r$ 为区域相似度值。

[0082] 步骤A40,将大于等于预设的区域相似度阈值的区域相似度值对应的新增区域数据集和历史兴趣区域集标注为同一区域标签;

[0083] 具体地,如果区域相似度达到预设的区域相似度阈值,那么将两个兴趣区域认为相同,例如新增区域数据集A和历史区域数据集C之间的区域相似度值达到预设的区域相似度阈值,则将新增区域数据集A和历史区域数据集C都标注为区域1;新增区域数据集B和历史区域数据集C之间的区域相似度值未达到预设的区域相似度阈值,则将新增区域数据集B和历史区域数据集C标注为不同的区域标签,如将新增区域数据集B标注为区域2。这一过程也可被称为标签匹配。

[0084] 步骤A50,根据各个所述新增区域数据集和各个所述历史区域数据集对应的区域标签,计算所述新增区域数据集和所述历史兴趣区域集之间的整体相似度值;

[0085] 具体地,将各个所述新增区域数据集与各个所述历史区域数据集进行区域标签标

注后,计算所述新增区域数据集和所述历史兴趣区域集之间的整体相似度值。可通过计算新增区域数据集与历史区域数据集对应同一区域标签的数量得到,本实施例中,计算公式可为 $S_{RS} = \frac{N_{s1}}{N_1}$ ,其中 $N_{s1}$ 指新增区域数据集和历史区域数据集对应的区域标签相同的数量, $N_1$ 指新增区域数据集或历史区域数据集的数量,若相似性达到阈值,则认为兴趣区域没有变化,若没有达到阈值,认为兴趣区域发生变化。

[0086] 步骤A60,当所述整体相似度值小于预设的整体相似度值时,根据新增移动数据,对所述历史移动数据进行更新。

[0087] 具体地,若整体相似度值大于等于预设的整体相似度值,则认为兴趣区域没有变化,若小于整体相似度值,认为兴趣区域发生变化,因此,根据新增移动数据,对所述历史移动数据进行更新,即将新增移动数据和历史移动数据整合,得到更新移动数据。

[0088] 步骤S300,根据所述更新移动数据,对预设的初始移动预测模型进行训练,生成目标移动预测模型。

[0089] 具体地,根据得到的更新移动数据,再对预设的初始移动预测模型进行训练,生成目标移动预测模型,其中,初始移动预测模型可以为背景中介绍的基于个人移动预测的预测模型。

[0090] 进一步地,为减少更新移动数据中累赘无效数据的影响,在对初始预测模型的训练过程中还会进行去重复处理,训练过程为:

[0091] 步骤B10,对所述更新移动数据进行去重复处理,得到有效移动数据;

[0092] 具体地,为减少更新移动数据中累赘无效数据的影响,在对初始预测模型的训练过程中还会进行去重复处理,将更新移动数据中重复的数据删除。例如用户在时刻a位于更新兴趣区域A发送了消息,一分钟后,当前时刻为时刻b时,用户仍然位于更新兴趣区域A并发送了消息,这种数据一方面数据量过大影响计算,一方面并未对预测结果带来更好的效果。可根据距离相近的采集时刻的更新移动数据之间距离的变化来将累赘的数据删除,以达到提高数据计算速度。

[0093] 进一步地,本实施例中采用的更新移动数据去重复处理的过程为:

[0094] 根据采集时刻的先后顺序,对所述更新移动数据进行排序,生成移动数据序列;

[0095] 依次计算移动数据序列中第N更新移动数据和第M更新移动数据之间的采集时刻的差值,得到移动时间,其中,N和M为小于等于所述更新移动数据的数量的正整数,且N小于M;

[0096] 当所述移动时间小于等于预设的移动时间阈值时,根据所述第N更新移动数据和所述第M更新移动数据之间的采集坐标,计算对应的移动距离;

[0097] 当所述移动距离小于等于预设的移动距离阈值时,将所述第N更新移动数据作为有效移动数据。

[0098] 具体地,先将所有更新移动数据的采集时刻转换为秒数,再从小到大排序,生成移动数据序列。由于个人可能在一定时间内,在同一个位置发送多条信息,但其移动状态没有发生改变,为此依据时间和位置进行去重。首先依据时间,依次计算移动数据序列中第N更新移动数据和第M更新移动数据之间的采集时刻的差值,得到移动时间,当多条更新移动数据的时间间隔小于阈值时,进一步检查其移动距离是否小于预设的移动距离阈值。可根据

所述第N更新移动数据和所述第M更新移动数据之间的采集坐标,计算得到两者的移动距离。由于本实施例中采集到的采集坐标形式可能不相同,而且大地坐标系并非平面坐标系,类似于椭球形,因此基于大地坐标系计算距离较为不便,因此为计算移动距离,预设一个标准的平面坐标系,然后将采集坐标投影至该平面坐标系,再根据投影后的坐标值计算两者之间的距离。若移动距离小于移动距离阈值,则认为该多条更新移动数据重复,则将多条数据中时间最早的保留,也就是第N更新移动数据,其他作为重复数据删除。

[0099] 步骤B20,对所述有效移动数据进行聚类处理,生成对所述更新移动数据对应的更新兴趣区域。

[0100] 具体地,此步骤与上述聚类处理过程类似,在此不再赘述。

[0101] 步骤B30,根据各个所述有效移动数据对应的更新兴趣区域以及所述有效移动数据之间的移动关系,将各个所述有效移动数据对应的更新兴趣区域作为对应的当前访问区域,并提取各个所述有效移动数据对应的空间特征和时间特征,得到各个有效移动数据对应的移动事件。

[0102] 具体地,空间特征为表明有效移动数据在空间上移动变化的特征,例如先前访问区域、前往访问区域与当前访问区域之间的距离;时间特征为有效移动数据在时间上移动变化的特征,例如在先前访问区域和当前访问区域之间的时差、发生移动的时刻所在哪一周、在一天中的白天或晚上等。通过空间特征和时间特征的结合,可得到有效移动数据对应的移动事件,移动事件可被定义为在某一时刻预测对象访问在某一更新兴趣区域。

[0103] 进一步地,预测对象的移动一般可根据之前的行为进行预测,例如某一个预测对象如果经过了书店,则有一定概率可能会去图书馆。因此在本实施例中,移动事件在空间上的变化可被定义为 $\langle r_a, r_b \rangle$ ,其中, $r_a$ 为预测对象移动前访问的历史兴趣区域,即先前访问区域(Pervious visit), $r_b$ 为预测对象移动后访问的历史兴趣区域,即当前访问区域。因此,步骤A30包括:

[0104] 根据所述有效移动数据的采集时刻的先后顺序,对各个所述有效移动数据进行排序,生成有效移动数据序列;

[0105] 将所述有效移动数据序列中第i有效移动数据对应的更新兴趣区域作为所述有效移动数据序列中第i+1有效移动数据对应的先前访问区域,将所述有效移动数据序列中第i+1有效移动数据对应的更新兴趣区域作为对应的当前访问区域,并根据所述先前访问区域和所述当前访问区域,生成所述第i有效移动数据对应的空间特征,以及

[0106] 根据预设的时刻划分规则和各个所述有效移动数据的采集时刻,生成各个所述有效移动数据对应的时间特征,其中,i为小于所述有效移动数据的数量的正整数。

[0107] 其中,时刻划分规则包括周划分规则和日划分规则。其中,日划分规则为一天中的时间区间的划分规则,指预测对象访问某个更新兴趣区域的时间在一天中的区间段。本实施例采用一种已有的时间段划分方法,具体为:凌晨(0:00-5:59),清晨(6:00-9:59),早上(10:00-13:59),下午(14:00-17:59),傍晚(18:00-20:59),深夜(21:00-23:59)。周划分规则为一周中的某一天的划分规则,指预测对象访问某个更新兴趣区域的时间在一周中为哪一天,该特征的使用是因为预测对象在周末和工作日呈现不同的移动模式,并且即使在工作日也有一定程度的区别。此外,使用者还可设置得到移动事件发生的月份的月划分规则,移动事件发生于白天亦或晚上的日夜划分规则等,因此时间特征的具体形式,以及包含的

类型可多种多样。

[0108] 步骤B40,根据所述移动事件,对预设的初始预测模型进行训练,生成目标移动预测模型。

[0109] 具体地,根据所述移动事件,对预设的初始预测模型进行训练,例如将移动事件的时间特征作为输入值,初始预测模型计算该时间特征对应的历史兴趣区域,然后将移动时间的空间特征作为真实值,将预测的历史兴趣区域与真实值进行比较,并根据两者之间的差距对初始预测模型进行参数调整,直至所有的移动事件都已训练,生成目标移动预测模型。

[0110] 进一步地,本实施例中初始预测模型包括特征权重对象,所述特征权重对象包括所述空间特征和所述时间特征;因此,具体的训练过程可为:

[0111] 步骤C10,针对每一个所述特征权重对象,根据所述移动事件,计算该特征权重对象对应各个所述更新兴趣区域的特征训练概率值。

[0112] 具体地,以初始预测模型的基础模型为朴素贝叶斯分类器为例进行描述,则初始预测模型可表示为 $P(r_{st}|pv, tm, dw) \propto P(pv|r_{st})^{w_{pv}}P(tm|r_{st})^{w_{tm}}P(dw|r_{st})^{w_{dw}}P(r_{st})^{w_{st}}$ ,

[0113] 其中 $r_{st}$ 为所述当前访问区域, $pv$ 为所述先前访问区域, $tm$ 为根据日划分规则得到的时间特征,所述 $dw$ 为根据周划分规则得到的时间特征。 $w_{pv}, w_{tm}, w_{dw}, w_r$ 分别表示四个特征权重对象对应的权重值。权重值越大,特征权重对象在模型中越重要。如果一个特征权重对象使得最后输出的结果更倾向于预测结果的真值 $r_{true}$ ,那么该特征权重对象比较重要。因此,针对每一个所述特征权重对象,根据所述移动事件,计算该特征权重对象对应各个所述更新兴趣区域的特征训练概率值。

[0114] 由上公式可知,共有四个需要训练的权重值 $w_{pv}, w_{tm}, w_{dw}, w_r$ ,分别对应 $P(pv|r_{st}), P(tm|r_{st}), P(dw|r_{st}), P(r_{st})$ 。以 $P(pv|r_{st})$ 为例,假设用于训练的移动事件为预测对象从区域A转移到区域B,那么 $pv$ 的值为区域A, $r_{st}$ 是需要预测的更新兴趣区域,该移动事件所对应的真值 $r_{true}$ 为B,因此对 $r_{st}$ 取不同的值, $P(pv|r_{st})$ 会有对应的概率值,将该概率值作为特征训练概率值。

[0115] 步骤C20,将数值最大的特征训练概率值对应的更新兴趣区域作为特征训练预测区域,并针对每一个所述移动事件,根据该移动事件中的当前访问区域、所述特征训练预测区域和预设的权重调整公式,对该特征权重对象对应的权重值进行调整,生成目标移动预测模型。

[0116] 具体地,选择令特征训练概率值 $P(pv|r_{st})$ 最大的 $r_{st}$ 作为训练预测区域 $r_{max}$ 。然后依据 $r_{max}$ 和该移动时间的当前访问区域 $r_{true}$ 调整权重 $w_{pv}$ 。其余三个权重调整方式与其相同。

[0117] 例如输入的移动事件为周一,先前访问区域为区域A,当前访区域为区域B,根据该移动事件可知 $pv$ 为区域A, $r_{true}$ 为区域B,在进行训练的时候,需要预测输入的移动事件中当前访区域是哪一个更新兴趣区域。例如,若待调整的权重值为 $w_{pv}$ ,那么计算当前访区域取不同值的时候,先访区域为区域A的概率,比如当前访区域为区域B时,特征训练概率值 $P(\text{区域A}|\text{区域B})$ 为 $2/3$ ,当前访区域为区域C时,特征训练概率值 $P(\text{区域A}|\text{区域C})$ 为 $1/3$ ,选择使特征训练概率值中的最大值所对应的当前访区域作为 $r_{max}$ ,也就是 $r_{max} = \text{区域B}$ ,因此 $r_{true} = r_{max}$ ;如果特征训练概率值 $P(\text{区域A}|\text{区域B})$ 为 $1/3$ ,特征训练概率值 $P(\text{区域A}|\text{区域C})$ 为 $2/3$ ,那么选择区域C为 $r_{max}$ ,这时 $r_{true}$ 不等于 $r_{max}$ 。如果 $r_{max}$ 和 $r_{true}$ 相同,那么该特征权重对象被认

为是对公式有积极作用,并增加该特征权重对象的权重,也就是增加周一对应的权重值;若不相同,则对应地减少权重值。调整权重值可根据预设的权重调整公式实现。

[0118] 本实施例提供一权重调整公式:

$$[0119] \quad \begin{cases} W_{\text{object}} = \frac{W_{\text{object}}}{c^{1+(p-0.5)}} \text{ 当 } r_{\text{max}} = r_{\text{ture}} \\ W_{\text{object}} = W_{\text{object}} \times c \text{ 当 } r_{\text{max}} \neq r_{\text{ture}} \end{cases};$$

[0120] 其中,object为权重对象,所述权重对象包括所述候选预测模型和所述特征权重对象, $w_{\text{object}}$ 为各个所述特征权重对象对应的权重值,c为系数,p为 $r_{\text{max}}$ 对应的训练概率值,所述训练概率值包括特征训练概率值和模型训练概率值, $r_{\text{max}}$ 为训练预测区域,所述训练预测区域包括特征预测区域和模型预测区域, $r_{\text{ture}}$ 为各个所述移动事件中的当前访问区域。当p比0.5大的时候,结果更加确定,若其等于真值,那么权重扩大更多。而当p小于0.5的时候,即使其等于真实值,但因为不够可靠,权重增加少。如果结果是错误,不等于真实值,则无论p是否大于0.5,都被认为是不可靠的,权重直接减小。

[0121] 进一步地,本实施例中初始预测模型包括多个候选预测模型;因此,具体的训练过程可为:

[0122] 针对每一个所述移动事件,基于所述候选预测模型、该移动事件中的时间特征和先前访问区域,计算该移动事件对应各个所述更新兴趣区域的模型训练概率值;

[0123] 将数值最大的模型训练概率值对应的更新兴趣区域为模型训练预测区域作为该移动事件对应的模型训练预测区域,并根据该移动事件中的当前访问区域、所述训练预测区域和预设的权重调整公式,对各个所述候选模型对应的权重值进行调整,得到各个候选模型对应的目标权重值;

[0124] 将数值最大的目标权重值对应的候选模型作为所述移动预测模型。

[0125] 具体地,为获得更好的预测效果,初始预测模型中可包括多个候选预测模型,而对每一个候选模型都设置一个权重值。然后采用移动事件,对各个候选预测模型进行训练,并根据预测结果的准确性,对候选预测模型对应的权重值进行调整,得到目标权重值。在进行使用时,将数值最大的目标权重值对应的候选模型作为所述移动预测模型。此过程与上述对特征权重对象对应的权重进行调整的过程类似,在此不再赘述。该种更新方式使得在近期具有较高预测精度的候选模型的目标权重值更大。目标权重值的上界设为1,下界设为0.1而不是0,使得权重可以在有限次数内到达下界,并快速增长。

[0126] 进一步地,本实施例中,候选预测模型包括带有自适应窗口的朴素贝叶斯模型(A Naïve Bayes model with adaptive windowing,NB-ADWIN)和带有衰减因子的朴素贝叶斯模型(ANaïve Bayes model with decay factors,NB-DF)。

[0127] 其中,NB-ADWIN对应的基本模型公式为 $P(r_{\text{st}}|x_{1\sim n}, t_{1\sim n}) \propto$

$$\prod_{j=1}^n \frac{P(x_j \wedge r_{\text{st}})}{P(r_{\text{st}})} \prod_{k=1}^m \frac{P(t_k \wedge r_{\text{st}})}{P(r_{\text{st}})} P(r_{\text{st}})$$

。为了计算公式中的各个概率值,用来对每个概率计算所需的数据建立多个窗口,每一个窗口用于检测对应的数据分布是否一致,若不一致,则剔除对应的采集时刻在前的数据,也就是旧的数据,并保留分布一致的数据,然后使用这些数据进行概率计算。由于随着时间变化,预测对象移动模式会发生改变,例如原先

在周一频繁访问兴趣区域A,后改为周二频繁访问兴趣区域A,原先最频繁去的是兴趣区域A,但因为搬家等原因,频繁访问的为兴趣区域B。因此基于新增移动数据和基于历史移动数据所引导的预测结果存在差距,为减少这一误差,通过数据的均一性,将历史移动数据中与新增移动数据并不均一的数据剔除,从而减少了历史移动数据的干扰,提高预测的准确性。

[0128] 此外,NB-ADWIN是基于窗口的技术,其具有一定的缺点,当移动模式变化比较慢的时候,窗口可能无法适应。移动模式变化较慢是指移动的规律变化较慢,其可以是移动距离、位置的访问顺序、访问时间等移动特征随着时间的变化,渐渐的发生改变了,比如从每天移动1km到1.1km再到1.2、1.3、1.4、1.5km等,数值是缓慢变化的,这种情况包括与其类似的包括但不限于其他特征的变化缓慢,可以称为移动模式变化比较慢。此外由于用户移动的不确定性,也可能导致错误的检测,因此一个稳定且能够削弱过时数据影响的方法是有必要的。基于此,本实施例提出一种带有衰减因子的朴素贝叶斯模型(A Naïve Bayes model with decay factors,NB-DF)。针对每一个输入的测试事件,计算对应的测试概率值为:

[0129] 该模型的基本公式为  $P(r_{st} = r_i | x_1 = v_{x_1}, \dots, x_n = v_{x_n}, t_1 = v_{t_1}, \dots, t_m = v_{t_m}) \propto \prod_j^n \frac{N_{v_{x_j}, r_i}}{N_{r_i}} \prod_k^m \frac{N_{v_{t_k}, r_i}}{N_{r_i}} \frac{N_{r_i}}{N_{r_{st}}}$ ;

[0130] 其中,所述测试事件为所述空间特征为 $v_{x_1}$ 、 $v_{x_2}$ 、 $\dots$ 、 $v_{x_n}$ ,所述当前访问区域为 $r_i$ ,且所述时间特征为 $v_{t_1}$ 、 $v_{t_2}$ 、 $\dots$ 、 $v_{t_m}$ 的移动事件,n为所述空间特征的数量正整数,m为所述时间特征的数量正整数;

[0131]  $r_{st}$ 为所述更新兴趣区域, $r_i$ 为所述测试事件中的当前访问区域, $x_j$ 为所述空间特征, $t_k$ 为所述时间特征;

[0132]  $N_{v_{x_j}, r_i}$ 为所述空间特征为 $v_{x_j}$ 且所述当前访问区域为 $r_i$ 的移动事件对应的计数之和, $N_{v_{t_k}, r_i}$ 为所述时间特征为 $t_m$ 且所述当前访问区域为 $r_i$ 的移动事件对应的计数之和, $N_{r_i}$ 为所述当前访问区域为 $r_i$ 的移动事件对应的计数之和, $N_{r_{st}}$ 为所有的移动事件对应的计数之和。

[0133] 一般每一个移动事件发生都计数为1,但是NB-DF考虑了测试事件发生的时刻与各个移动事件发生的时刻之间的时间长短对预测结果的影响。本实施例中,各种类型的移动事件之和可表示为: $N_{i_o, t} = \sum_{i_o \in S_{i_o}} W_{i_o, t}$ ,其中, $i_o$ 为所述移动事件, $S_{i_o}$ 为所有所述移动事件 $i_o$ 的集合,t为该测试时间对应的发生时刻, $W_{i_o, t}$ 表示在时刻t发生移动事件 $i_o$ 对应的计数, $W_{i_o, t} = \lambda^{t-t_{i_o}}$ , $\lambda$ 为预设的衰减因子, $\lambda \in (0, 1)$ 。随着移动事件发生的时刻,也就是 $t_{i_o}$ ,与测试事件发生的时刻,也就是t,之间的时间的长短,即 $(t-t_{i_o})$ ,移动事件对测试事件对应的预测结果影响越小,因此在通过计算事件发生次数来预测结果时,时间越久的数据应当减少其影响力。

[0134] 进一步地,当所述时间特征包括多种类型的特征时,例如上述根据周划分规则划分的周特征和日划分规则的日特征,则根据测试时间,计算得到的测试概率值可表示为P



$(r_{st} = r_i | pv = v_{pv}, td = v_{td}, dw = v_{dw})$ , 其中,  $pv$  为先前访问区域,  $td$  为日特征,  $dw$  为周特征,  $v_{td}$  为测试事件的日特征,  $v_{dw}$  为测试事件的周特征。因此, 上述模型的基本公式可对应地进行调整, 例如上述NB-DF模型, 其基本公式应调整为

$$P(r_{st} = r_i | pv = v_{pv}, td = v_{td}, dw = v_{dw}) \propto \frac{N_{v_{pv}, r_i}}{N_{r_i}} \frac{N_{v_{td}, r_i}}{N_{r_i}} \frac{N_{v_{dw}, r_i}}{N_{r_i}} \frac{N_{r_i}}{N_{rst}},$$

其中  $N_{v_{td}, r_i}$  为在  $v_{td}$  访问  $r_i$  的移动事件的数量,  $N_{v_{dw}, r_i}$  为在  $v_{dw}$  访问  $r_i$  的移动事件的数量。

[0135] 步骤S400, 当检测到待预测的指定时刻时, 将所述指定时刻输入所述目标移动预测模型中, 生成所述指定时刻对应的目标区域并输出。

[0136] 具体地, 目标移动预测模型用于预测预测对象在某一特定时刻, 最可能移动到的历史兴趣区域。因此, 当检测到待预测的指定时刻时, 将所述指定时刻输入所述目标移动预测模型中, 目标移动预测模型根据指定时刻, 计算对应各个历史兴趣区域的概率值, 并将概率值最大的历史兴趣区域作为目标区域并输出。

[0137] 进一步地, 执行该移动预测方法的主体可以是安装于软件上的插件也可以独立包装为单独的软件。此外, 该移动预测方法可与多种功能进行结合, 例如天气预报, 针对某一目标用户, 提前根据目标用户的上下班时间, 预测对应的目标区域, 然后结合天气预报, 将目标区域的在上下班时刻的天气推精准推送给目标用户, 以方便目标用户提前准备好衣物和雨具。除天气预报外, 交通实时报导, 实时预警等都可与移动预测结合的程序以及移动预测方法都可进行结合, 从而实现精准推送。

[0138] 为了说明本发明实施例提供的预测对象移动预测方法的预测效果, 图4示出了某一个用户的位置分布, 将该用户的位置坐标以某一时刻为分界线分为两组, 在该时刻以前得到的移动数据归为第一组, 在该时刻之后得到的移动数据归为第二组, 其中, 第一组移动数据的位置坐标用圆形标注, 第二组移动数据的位置坐标用方形标注, 椭圆框和方框框选的区域A、B、C和D为四个人工标注的活动区域, 从图中可看出第一组移动数据和第二组移动数据差别较大, 在活动区域A、B、C中只有第一组移动数据的圆形标注, 而在活动区域D中只有后一组移动数据的方形标注, 因此用户在该时刻前后活动的兴趣偏好变化较大, 如果不更新兴趣区域, 那么在活动区域D范围内可能存在的兴趣区域无法被提取, 只有当存在更新策略的时候, 新的兴趣区域才能被提取, 并且更符合实际情况。

[0139] 为了进一步说明本发明实施例提供的预测对象移动预测方法相对于其他主流方法的优势, 下表示出了本方法NB-AWDF与主流方法多特征加权贝叶斯模型(A Multi-feature Weighted Bayesian Model, MWBM)、稀疏移动马尔科夫链模型(A Sparse Mobility Markov Chain Model, SMMC)的精度对比, 精度为预测正确次数与预测总次数的比值,  $\eta$  为所述新增数据阈值。从表中可以看出本方法精度在社交网站Instagram数据最高达70.70%, 在社交网站Twitter数据最高达59.95%, 相对于MWBM在两个数据上最高分别提升6.45%和7.34%, 平均提升5.66%和6.39%, 相对于SMMC在两个数据上最高分别提升8.73%和7.91%, 平均提升7.91%和7.54%。因此, 本发明实施例提供的预测对象移动预测方法相对其他主流方法有优势。

	$\eta$	NB-AWD F	MWBM	SMMC	相 对 MWBM 提升	相 对 SMMC 提 升
[0140] Instagra m 数据	10	68.99%	62.67%	61.38%	6.32%	7.61%
	30	69.55%	63.09%	61.89%	6.45%	7.66%
	50	69.79%	64.49%	62.14%	5.30%	7.65%
	100	70.70%	66.13%	61.96%	4.57%	8.73%
平均提升					5.66%	7.91%
[0141] Twitter 数据	10	59.95%	52.62%	52.79%	7.34%	7.17%
	30	59.43%	52.46%	52.20%	6.96%	7.22%
	50	59.56%	53.59%	51.65%	5.97%	7.91%
	100	58.43%	53.14%	50.58%	5.29%	7.85%
平均提升					6.39%	7.54%

[0141] 进一步地,如图5所示,基于上述移动预测方法,本发明还相应提供了一种智能终端,所述智能终端包括处理器10、存储器20及显示器30。图5仅示出了智能终端的部分组件,但是应理解的是,并不要求实施所有示出的组件,可以替代的实施更多或者更少的组件。

[0142] 所述存储器20在一些实施例中可以是所述智能终端的内部存储单元,例如智能终端的硬盘或内存。所述存储器20在另一些实施例中也可以是所述智能终端的外部存储设备,例如所述智能终端上配备的插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)等。进一步地,所述存储器20还可以既包括所述智能终端的内部存储单元也包括外部存储设备。所述存储器20用于存储安装于所述智能终端的应用软件及各类数据,例如所述安装智能终端的程序代码等。所述存储器20还可以用于暂时地存储已经输出或者将要输出的数据。在一实施例中,存储器20上存储有移动预测程序40,该移动预测程序40可被处理器10所执行,从而实现本申请中移动预测方法。

[0143] 所述处理器10在一些实施例中可以是一中央处理器(Central Processing Unit, CPU),微处理器或其他数据处理芯片,用于运行所述存储器20中存储的程序代码或处理数据,例如执行所述移动预测方法等。

[0144] 所述显示器30在一些实施例中可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。所述显示器30用

于显示在所述智能终端的信息以及用于显示可视化的用户界面。所述智能终端的部件10-30通过系统总线相互通信。

[0145] 在一实施例中,当处理器10执行所述存储器20中移动预测程序40时实现以下步骤:

[0146] 获取新增移动数据和历史移动数据,并对所述新增移动数据和所述历史移动数据进行聚类处理,生成对应的新增兴趣区域;

[0147] 当所述新增兴趣区域与所述历史移动数据对应的历史兴趣区域之间的相似度值符合预设的相似度条件时,根据所述新增移动数据,对所述历史移动数据进行更新,得到更新移动数据;

[0148] 根据所述更新移动数据,对预设的初始移动预测模型进行训练,生成目标移动预测模型;

[0149] 当检测到待预测的指定时刻时,将所述指定时刻输入所述目标移动预测模型中,生成所述指定时刻对应的目标区域并输出。

[0150] 本发明还提供一种存储介质,其中,所述存储介质存储有移动预测程序,所述移动预测程序被处理器执行时实现如上所述的移动预测方法的步骤。

[0151] 当然,本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关硬件(如处理器,控制器等)来完成,所述的程序可存储于一计算机可读的存储介质中,所述程序在执行时可包括如上述各方法实施例的流程。其中所述的存储介质可为存储器、磁碟、光盘等。

[0152] 应当理解的是,本发明的应用不限于上述的举例,对本领域普通技术人员来说,可以根据上述说明加以改进或变换,所有这些改进和变换都应属于本发明所附权利要求的保护范围。

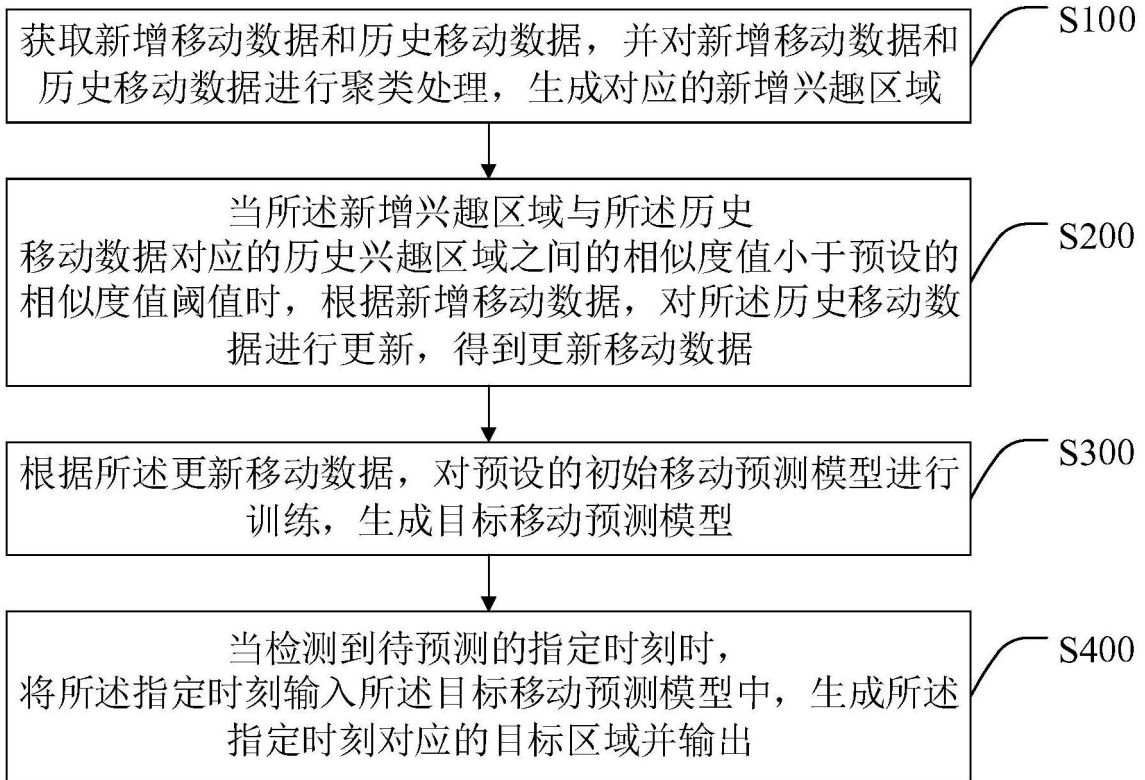


图1

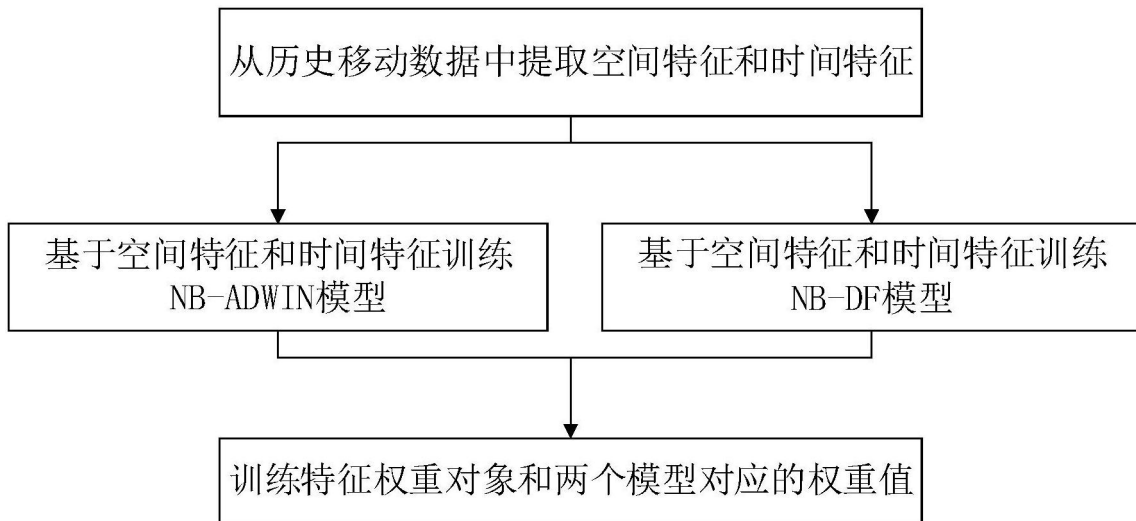


图2

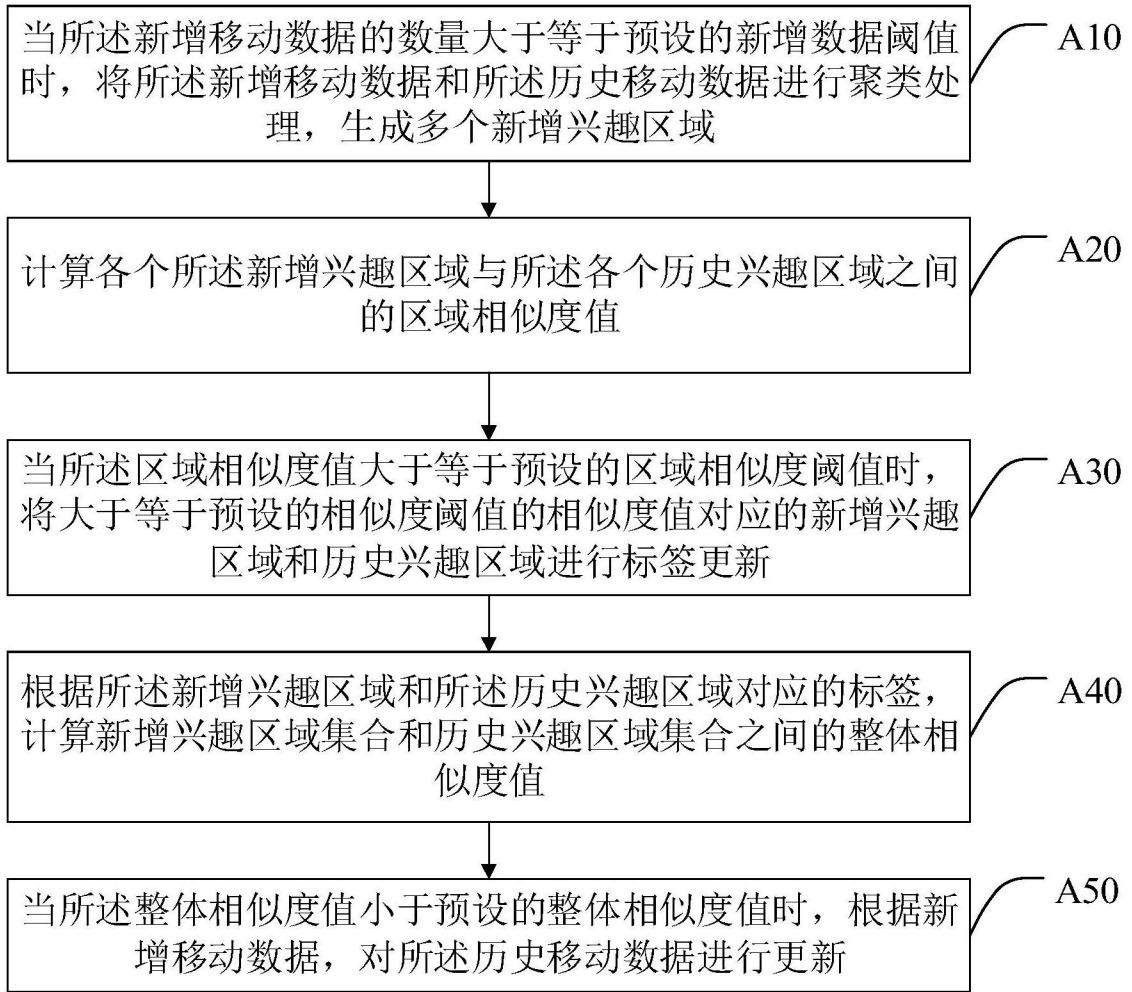


图3

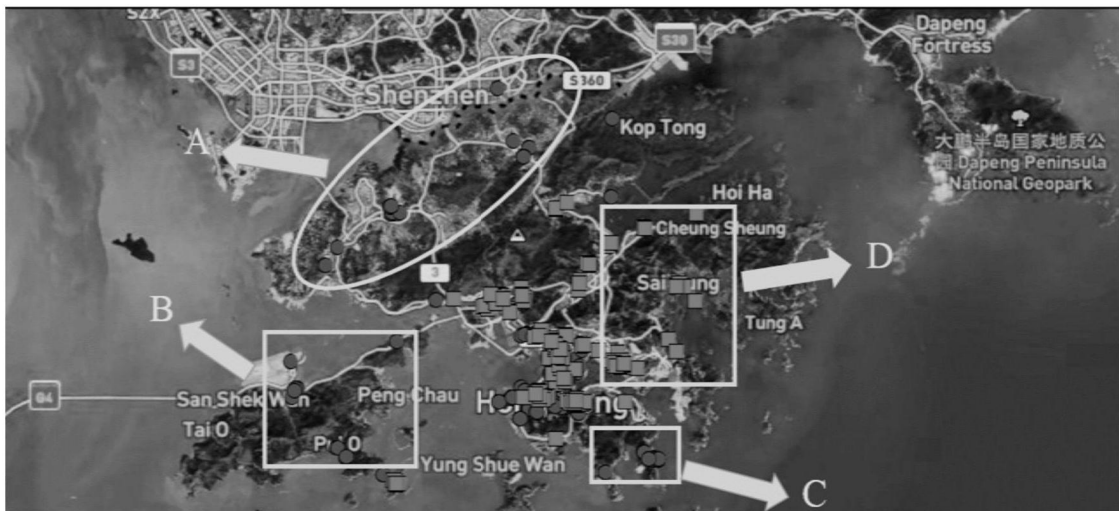


图4

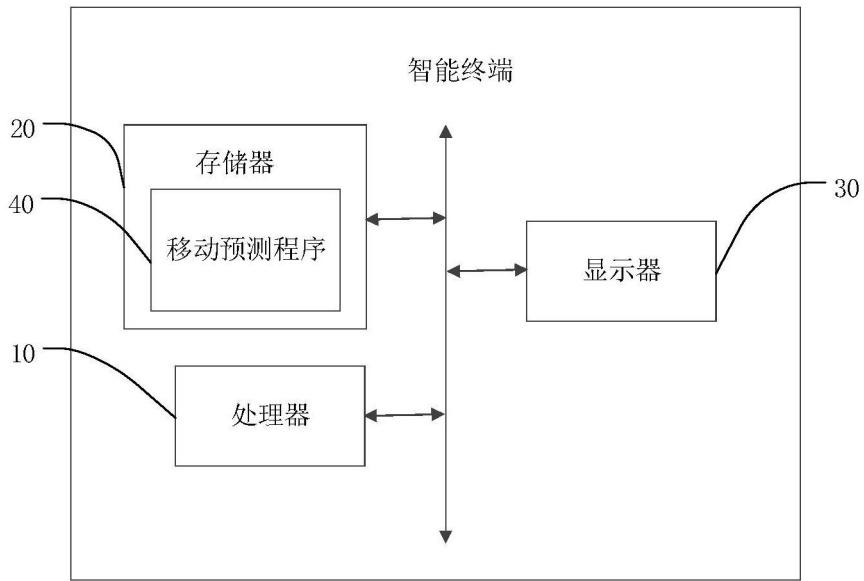


图5