(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 112488384 B
(45)授权公告日 2021.08.31

(21)申请号 202011358914.7
(22)申请日 2020.11.27
(65)同一申请的已公布的文献号
申请公布号 CN 112488384 A
(43)申请公布日 2021.03.12
(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 刘哲维
(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 朱阳波 王永文
(51)Int.Cl.
G06Q 10/04(2012.01)
G06K 9/62(2006.01)
G06N 20/00(2019.01)
(56)对比文件
CN 105740401 A,2016.07.06
CN 110570044 A,2019.12.13
审查员 袁一民

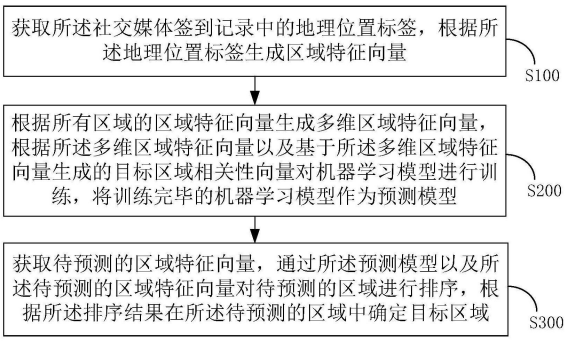
权利要求书3页 说明书9页 附图3页

(54)发明名称

基于社交媒体签到预测目标区域的方法、终端及存储介质

(57)摘要

本发明公开了基于社交媒体签到预测目标区域的方法、终端及存储介质，通过获取所述社交媒体签到记录中的地理位置标签，根据所述地理位置标签生成区域特征向量；根据所有区域的区域特征向量生成多维区域特征向量，根据所述多维区域特征向量以及基于所述多维区域特征向量生成的目标区域相关性向量对机器学习模型进行训练，将训练完毕的机器学习模型作为预测模型；获取待预测的区域特征向量，通过所述预测模型以及所述待预测的区域特征向量对待预测的区域进行排序，根据所述排序结果在所述待预测的区域中确定目标区域。本发明将确定用户常驻区域的任务抽象为一个排序问题，并使用机器学习模型对用户访问过的各区域进行排序，最终成功预测用户的常驻区域。



CN 112488384 B

1.基于社交媒体签到预测目标区域的方法,其特征在于,所述方法包括:

获取所述社交媒体签到记录中的地理位置标签,根据所述地理位置标签生成区域特征向量;

根据所有区域的区域特征向量生成多维区域特征向量,根据所述多维区域特征向量以及基于所述多维区域特征向量生成的目标区域相关性向量对机器学习模型进行训练,将训练完毕的机器学习模型作为预测模型;

获取待预测的区域特征向量,通过所述预测模型以及所述待预测的区域特征向量对待预测的区域进行排序,根据所述排序结果在所述待预测的区域中确定目标区域;

所述根据所有区域的区域特征向量生成多维区域特征向量,根据所述多维区域特征向量以及基于所述多维区域特征向量生成的目标区域相关性向量对机器学习模型进行训练,将训练完毕的机器学习模型作为预测模型包括:

获取所有区域的区域特征向量并进行整合,将整合得到的向量作为多维区域特征向量;

获取基于所述多维区域特征向量生成的目标区域相关性向量;

将所述多维区域特征向量作为机器学习模型的输入数据,将所述基于所述多维区域特征向量生成的目标区域相关性向量作为所述机器学习模型的输出数据,对所述机器学习模型进行训练;

将训练完毕的机器学习模型作为预测模型;

所述获取基于所述多维区域特征向量生成的目标区域相关性向量包括:

根据所述多维区域特征向量中的每一个区域与目标区域的相关性,对每一个区域进行评分,得到所述每一个区域的目标区域相关性分值;

对所述每一个区域的目标区域相关性分值进行整合,将整合得到的向量作为目标区域相关性向量。

2.根据权利要求1所述的基于社交媒体签到预测目标区域的方法,其特征在于,所述获取所述社交媒体签到记录中的地理位置标签,根据所述地理位置标签生成区域特征向量包括:

获取所述社交媒体签到记录中的地理位置标签,通过所述地理位置标签对所述社交媒体签到记录进行分类;

根据分类结果生成区域签到频率数据;

根据分类结果生成区域活跃天数频率数据;

根据分类结果生成区域在预设时间段的活跃频率数据;

根据所述区域签到频率数据、所述区域活跃天数频率数据以及所述区域在预设时间段内的活跃频率数据,生成各区域的区域特征向量。

3.根据权利要求2所述的基于社交媒体签到预测目标区域的方法,其特征在于,所述根据分类结果生成区域签到频率数据包括:

根据所述分类结果计算用户在各个区域发布的社交媒体签到的数量;

根据所述地理位置标签获取用户发布的社交媒体签到的总数量;

将所述用户在各个区域发布的社交媒体签到的数量与所述用户发布的社交媒体签到的总数量的比值作为各个区域的区域签到频率数据。

4.根据权利要求2所述的基于社交媒体签到预测目标区域的方法，其特征在于，所述根据分类结果生成区域活跃天数频率数据包括：

根据所述分类结果计算用户在各个区域的活跃天数；所述活跃天数为用户至少发布过一条社交媒体签到记录的天数；

对计算出的用户在各个区域的活跃天数进行加法运算，得到用户的活跃总天数；

将所述用户在各个区域的活跃天数与所述用户的活跃总天数的比值作为各个区域的区域活跃天数频率数据。

5.根据权利要求2所述的基于社交媒体签到预测目标区域的方法，其特征在于，所述区域在预设时间段的活跃频率数据包括：夜晚活跃频率数据、夏季活跃频率数据以及周末活跃频率数据，所述根据分类结果生成区域在预设时间段的活跃频率数据包括：

所述区域在预设时间段的活跃频率数据包括：夜晚活跃频率数据、夏季活跃频率数据以及周末活跃频率数据；所述根据分类结果生成区域在预设时间段的活跃频率数据包括：

根据所述分类结果计算所述区域的夜晚活跃天数；所述夜晚活跃天数为用户在夜间预设时间段内位于所述区域至少发布过一条的天数；

对计算出的用户在各个区域的夜晚活跃天数进行加法运算，得到夜晚活跃总天数；

将所述区域的夜晚活跃天数与所述夜晚活跃总天数的比值作为所述区域的夜晚活跃频率数据；

根据所述分类结果计算用户在预设月份之间位于所述区域发布的社交媒体签到数量；

根据所述地理位置标签获取用户在所述预设月份之间在各个地区发布的社交媒体签到总数量；

将所述用户在预设月份之间位于所述区域发布的社交媒体签到数量与所述用户在所述预设月份之间在各个地区发布的社交媒体签到总数量的比值作为所述夏季活跃频率数量；

根据所述分类结果计算用户在周末位于所述区域发布的社交媒体签到数量；

根据所述地理位置标签获取用户在周末位于各个地区发布的社交媒体签到总数量；

将所述用户在周末位于所述区域发布的社交媒体签到数量与所述用户在周末位于各个地区发布的社交媒体签到总数量的比值作为所述周末活跃频率数据。

6.根据权利要求1所述的基于社交媒体签到预测目标区域的方法，其特征在于，所述获取待预测的区域特征向量，通过所述预测模型以及所述待预测的区域特征向量对待预测的区域进行排序，根据所述排序结果在所述待预测的区域中确定目标区域包括：

获取待预测的区域特征向量，将所述待预测的区域特征向量输入所述预测模型中；

获取所述预测模型输出的基于所述待预测的区域特征向量生成的评分；

基于所述评分对待预测的区域进行排序，根据排序结果在所述待预测的区域中确定目标区域。

7.一种移动终端，其特征在于，包括：处理器、与处理器通信连接的存储介质，所述存储介质适于存储多条指令；所述处理器适于调用所述存储介质中的指令，一致性实现上述权利要求1-6任一项所述的基于社交媒体签到预测目标区域的方法的步骤。

8.一种计算机可读存储介质，其上存储有多条指令，其特征在于，所述指令适用于由处理器加载并执行，以实现上述权利要求1-6任一项所述的基于社交媒体签到预测目标区域

的方法的步骤。

基于社交媒体签到预测目标区域的方法、终端及存储介质

技术领域

[0001] 本发明涉及地理信息领域，尤其涉及的是一种基于社交媒体签到预测目标区域的方法、终端及存储介质。

背景技术

[0002] 使用带有地理位置标签的社交媒体签到数据，推测用户的常驻区域是地理信息科学、人类移动模式研究等领域的重要研究手段。从技术手段上看，常用的用户常驻地推测方法，多使用简单统计方法。该类方法多基于人的经验直觉，缺乏严谨的证明及理论基础，探测结果精度较低。

[0003] 因此，现有技术还有待改进和发展。

发明内容

[0004] 本发明要解决的技术问题在于，针对现有技术的上述缺陷，提供一种基于社交媒体签到预测目标区域的方法、终端及存储介质，旨在解决现有技术中根据社交媒体签到数据难以准确预测用户的目标区域的问题。

[0005] 本发明解决问题所采用的技术方案如下：

[0006] 第一方面，本发明实施例提供一种基于社交媒体签到预测目标区域的方法，其中，所述方法包括：

[0007] 获取所述社交媒体签到记录中的地理位置标签，根据所述地理位置标签生成区域特征向量；

[0008] 根据所有区域的区域特征向量生成多维区域特征向量，根据所述多维区域特征向量以及基于所述多维区域特征向量生成的目标区域相关性向量对机器学习模型进行训练，将训练完毕的机器学习模型作为预测模型；

[0009] 获取待预测的区域特征向量，通过所述预测模型以及所述待预测的区域特征向量对待预测的区域进行排序，根据所述排序结果在所述待预测的区域中确定目标区域。

[0010] 在一种实施方式中，所述获取所述社交媒体签到记录中的地理位置标签，根据所述地理位置标签生成区域特征向量包括：

[0011] 获取所述社交媒体签到记录中的地理位置标签，通过所述地理位置标签对所述社交媒体签到记录进行分类；

[0012] 根据分类结果生成区域签到频率数据；

[0013] 根据分类结果生成区域活跃天数频率数据；

[0014] 根据分类结果生成区域在预设时间段的活跃频率数据；

[0015] 根据所述区域签到频率数据、所述区域活跃天数频率数据以及所述区域在预设时间段内的活跃频率数据，生成各区域的区域特征向量。

[0016] 在一种实施方式中，所述根据分类结果生成区域签到频率数据包括：

[0017] 根据所述分类结果计算用户在各个区域发布的社交媒体签到的数量；

[0018] 根据所述地理位置标签获取用户发布的社交媒体签到的总数量；

[0019] 将所述用户在各个区域发布的社交媒体签到的数量与所述用户发布的社交媒体签到的总数量的比值作为各个区域的区域签到频率数据。

[0020] 在一种实施方式中，所述根据分类结果生成区域活跃天数频率数据包括：

[0021] 根据所述分类结果计算用户在各个区域的活跃天数；所述活跃天数为用户至少发布过一条社交媒体签到记录的天数；

[0022] 对计算出的用户在各个区域的活跃天数进行加法运算，得到用户的活跃总天数；

[0023] 将所述用户在各个区域的活跃天数与所述用户的活跃总天数的比值作为各个区域的区域活跃天数频率数据。

[0024] 在一种实施方式中，所述区域在预设时间段的活跃频率数据包括：夜晚活跃频率数据、夏季活跃频率数据以及周末活跃频率数据，所述根据分类结果生成区域在预设时间段的活跃频率数据包括：

[0025] 所述区域在预设时间段的活跃频率数据包括：夜晚活跃频率数据、夏季活跃频率数据以及周末活跃频率数据；所述根据分类结果生成区域在预设时间段的活跃频率数据包括：

[0026] 根据所述分类结果计算所述区域的夜晚活跃天数；所述夜晚活跃天数为用户在夜间预设时间段内位于所述区域至少发布过一条的天数；

[0027] 对计算出的用户在各个区域的夜晚活跃天数进行加法运算，得到夜晚活跃总天数；

[0028] 将所述区域的夜晚活跃天数与所述夜晚活跃总天数的比值作为所述区域的夜晚活跃频率数据；

[0029] 根据所述分类结果计算用户在预设月份之间位于所述区域发布的社交媒体签到数量；

[0030] 根据所述地理位置标签获取用户在所述预设月份之间在各个地区发布的社交媒体签到总数量；

[0031] 将所述用户在预设月份之间位于所述区域发布的社交媒体签到数量与所述用户在所述预设月份之间在各个地区发布的社交媒体签到总数量的比值作为所述夏季活跃频率数量；

[0032] 根据所述分类结果计算用户在周末位于所述区域发布的社交媒体签到数量；

[0033] 根据所述地理位置标签获取用户在周末位于各个地区发布的社交媒体签到总数量；

[0034] 将所述用户在周末位于所述区域发布的社交媒体签到数量与所述用户在周末位于各个地区发布的社交媒体签到总数量的比值作为所述周末活跃频率数据。

[0035] 在一种实施方式中，所述根据所有区域的区域特征向量生成多维区域特征向量，根据所述多维区域特征向量以及基于所述多维区域特征向量生成的目标区域相关性向量对机器学习模型进行训练，将训练完毕的机器学习模型作为预测模型包括：

[0036] 获取所有区域的区域特征向量并进行整合，将整合得到的向量作为多维区域特征向量；

[0037] 获取基于所述多维区域特征向量生成的目标区域相关性向量；

[0038] 将所述多维区域特征向量作为机器学习模型的输入数据，将所述基于所述多维区域特征向量生成的目标区域相关性向量作为所述机器学习模型的输出数据，对所述机器学习模型进行训练；

[0039] 将训练完毕的机器学习模型作为预测模型。

[0040] 在一种实施方式中，所述获取基于所述多维区域特征向量生成的目标区域相关性向量包括：

[0041] 根据所述多维区域特征向量中的每一个区域与目标区域的相关性，对每一个区域进行评分，得到所述每一个区域的目标区域相关性分值；

[0042] 对所述每一个区域的目标区域相关性分值进行整合，将整合得到的向量作为目标区域相关性向量。

[0043] 在一种实施方式中，所述获取待预测的区域特征向量，通过所述预测模型以及所述待预测的区域特征向量对待预测的区域进行排序，根据所述排序结果在所述待预测的区域中确定目标区域包括：

[0044] 获取待预测的区域特征向量，将所述待预测的区域特征向量输入所述预测模型中；

[0045] 获取所述预测模型输出的基于所述待预测的区域特征向量生成的评分；

[0046] 基于所述评分对待预测的区域进行排序，根据排序结果在所述待预测的区域中确定目标区域。

[0047] 第二方面，本发明实施例还提供一种移动终端，其中，包括：处理器、与处理器通信连接的存储介质，所述存储介质适于存储多条指令；所述处理器适于调用所述存储介质中的指令，一致性实现上述任一项所述的基于社交媒体签到预测目标区域的方法的步骤。

[0048] 第二方面，本发明实施例还提供一种计算机可读存储介质，其中，所述指令适用于由处理器加载并执行，以实现上述任一项所述的基于社交媒体签到预测目标区域的方法的步骤。

[0049] 本发明的有益效果：本发明实施例通过获取所述社交媒体签到记录中的地理位置标签，根据所述地理位置标签生成区域特征向量；根据所有区域的区域特征向量生成多维区域特征向量，根据所述多维区域特征向量以及基于所述多维区域特征向量生成的目标区域相关性向量对机器学习模型进行训练，将训练完毕的机器学习模型作为预测模型；获取待预测区域的特征向量，通过所述预测模型对所述待预测区域进行排序，根据所述排序结果在所述待预测区域中确定目标区域。本发明将确定用户常驻区域的任务抽象为一个排序问题，并使用机器学习模型对用户访问过的各区域进行排序，最终成功预测用户的常驻区域。

附图说明

[0050] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明中记载的一些实施例，对于本领域普通技术人员来讲，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的附图。

[0051] 图1是本发明实施例提供的基于社交媒体签到预测目标区域的方法的流程示意

图。

[0052] 图2是本发明实施例提供的生成区域特征向量的流程示意图。

[0053] 图3是本发明实施例提供的获取预测模型的流程示意图。

[0054] 图4是本发明实施例提供的确定目标区域的流程示意图。

[0055] 图5是本发明实施例提供的终端的原理框图。

[0056] 图6是本发明实施例提供的用于评价本发明技术效果的实验结果图。

具体实施方式

[0057] 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。

[0058] 需要说明，若本发明实施例中有涉及方向性指示(诸如上、下、左、右、前、后……)，则该方向性指示仅用于解释在某一特定姿态(如附图所示)下各部件之间的相对位置关系、运动情况等，如果该特定姿态发生改变时，则该方向性指示也相应地随之改变。

[0059] 近年来，随着移动定位设备的普及和基于位置服务的兴起，传统社交网络与定位技术融合产生了一种新型的在线社交媒体——基于地理位置的社交网络，支持用户随时随地分享自己的位置信息。该类应用的用户的典型行为是进行社交媒体签到或者针对签到地点进行评论等等。个人的签到数据可以表示个人的历史移动轨迹，大量用户的签到数据可以揭示人类的移动模式和生活规律。由于签到数据是带有地理信息的社交网络数据，因此它既可以反映用户的社交网络行为，又可以反映用户的移动行为。同时由于其获取方式简单、成本低，近年来越来越多的学者采用签到数据进行研究。

[0060] 其中一项研究就包括使用带有地理位置标签的社交媒体签到数据，推测用户的常驻区域。该方法是地理信息科学、人类移动模式研究等领域的重要研究手段。从技术手段上看，常用的用户常驻地推测方法，多使用简单统计方法。该类方法多基于人的经验直觉，缺乏严谨的证明及理论基础，探测结果精度较低。

[0061] 基于现有技术的上述缺陷，本发明提供了一种确定目标区域的方法，以实现准确确定用户的常驻区域。该方法通过将确定用户常驻区域的任务抽象为一个排序问题，并使用机器学习模型对用户访问过的各区域进行排序，最终实现成功预测用户的常驻区域。

[0062] 如图1所示，所述方法包括如下步骤：

[0063] 步骤S100、获取用户的社交媒体签到记录，根据所述社交媒体签到记录生成区域特征向量。

[0064] 用户的社交媒体签到记录结合了用户当前地理位置这一要素。在用户的社交媒体签到记录里面，用户可以对地点进行签到，公开他们的地理位置，在签到地留下评论信息。当社交媒体与位置结合在一起以后，便能对用户活动的时间、地点进行分析，更好地了解用户在各个区域的活动规律。因此需要首先获取用户的社交媒体签到记录，根据所述社交媒体签到记录生成各个区域的区域特征向量。

[0065] 如图2所示，在一种实现方式中，所述步骤S100具体包括如下步骤：

[0066] 步骤S110、获取所述社交媒体签到记录中的地理位置标签，通过所述地理位置标签对所述社交媒体签到记录进行分类；

[0067] 步骤S120、根据分类结果生成区域签到频率数据；

[0068] 步骤S130、根据分类结果生成区域活跃天数频率数据；

[0069] 步骤S140、根据分类结果生成区域在预设时间段的活跃频率数据；

[0070] 步骤S150、根据所述区域签到频率数据、所述区域活跃天数频率数据以及所述区域在预设时间段内的活跃频率数据，生成各区域的区域特征向量。

[0071] 获取到用户的社交媒体签到记录以后，为了实现基于所述社交媒体签到记录分析用户在各个区域的活动情况，需要首先根据所述社交媒体签到记录中的地理位置标签对所述社交媒体签到记录以区域为单位进行分类，然后根据分类结果生成区域签到频率数据、区域活跃天数频率数据以及区域在预设时间段内的活跃频率数据，最后基于这几类频率数据构成一个区域的特征向量。在一种实现方式中，可以根据所述分类结果计算用户在各个区域发布的社交媒体签到的数量，然后根据所述用户的社交媒体签到记录获取用户发布的社交媒体签到的总数量，最后将所述用户在各个区域发布的社交媒体签到的数量与所述用户发布的社交媒体签到的总数量的比值作为各个区域的区域签到频率数据。

[0072] 此外，还可以根据所述分类结果计算用户在各个区域的活跃天数，然后对计算出的用户在各个区域的活跃天数进行加法运算，得到用户的活跃总天数；所述活跃天数为用户至少发布过一条社交媒体签到记录的天数，最后将所述用户在各个区域的活跃天数与所述用户的活跃总天数的比值作为各个区域的区域活跃天数频率数据。

[0073] 并且，还可以根据所述分类结果计算用户在预设时间段位于各个区域发布的社交媒体签到的数量，然后对计算出的用户在预设时间段位于各个区域发布的社交媒体签到的数量进行加法运算，得到用户在预设时间段内发布的社交媒体签到的总数量，最后将所述用户在预设时间段位于各个区域发布的社交媒体签到的数量与所述用户在预设时间段内发布的社交媒体签到的总数量的比值作为各个区域的区域在所述预设时间段内的活跃频率数据。

[0074] 在一种实现方式中，所述区域在预设时间段的活跃频率数据包括：夜晚活跃频率数据、夏季活跃频率数据以及周末活跃频率数据。为了获取这三类频率数据，本实施例需要根据所述分类结果计算所述区域的夜晚活跃天数，所述夜晚活跃天数为用户在夜间预设时间段内位于所述区域至少发布过一条的天数。然后对计算出的用户在各个区域的夜晚活跃天数进行加法运算，得到夜晚活跃总天数。最后将所述区域的夜晚活跃天数与所述夜晚活跃总天数的比值作为所述区域的夜晚活跃频率数据。

[0075] 此外，还需要根据所述分类结果计算用户在预设月份之间位于所述区域发布的社交媒体签到数量。然后，根据所述地理位置标签获取用户在所述预设月份之间在各个区域发布的社交媒体签到总数量。之后，将所述用户在预设月份之间位于所述区域发布的社交媒体签到数量与所述用户在所述预设月份之间在各个区域发布的社交媒体签到总数量的比值作为所述夏季活跃频率数量。

[0076] 另外，还需要根据所述分类结果计算用户在周末位于所述区域发布的社交媒体签到数量。之后，根据所述地理位置标签获取用户在周末位于各个区域发布的社交媒体签到总数量。最后将所述用户在周末位于所述区域发布的社交媒体签到数量与所述用户在周末位于各个区域发布的社交媒体签到总数量的比值作为所述周末活跃频率数据。

[0077] 在一种实现方式中，鉴于相关研究表明，用户夜间在社交媒体签到相较于白天签

到，更能反应用户的常驻区域；同理，用户夏季5月至9月在社交媒体签到，相较于冬季更能反应用户的常驻区域；用户周末在社交媒体上签到，相较于工作日更能反应用户的常驻区域。因此本实施例基于上述相关研究，可以将夜间预设时间段设置为晚上19点至早上7点，预设月份之间设置为5月至9月，周末则是传统意义上的周六至周日。

[0078] 举例说明，对于某用户u_i，其社交媒体记录中的地理位置标签显示其在某区域r_j出现过，则该区域的特征向量$ur^j_i=(rt_p,rt_{ad},rt_{np},rt_{an},rt_s,rt_w)$。其中，$rt_p$为用户$u_i$在区域$r_j$发布的社交媒体签到数量占其所有社交媒体签到数量的比例；rt_{ad}为用户u_i在区域r_j的活跃天数占其所有活跃天数的比例，其中活跃天数定义为用户在该天至少发布过一条社交媒体签到。则rt_{an}为用户u_i在区域r_j的活跃夜晚数占其所有活跃夜晚数的比例，其中活跃夜晚定义为用户在该天的晚上19点至早上七点之间至少发布了一条社交媒体签到；rt_s为用户u_i在区域r_j的夏季5月至9月的社交媒体签到数量占其所有夏季社交媒体签到数量的比例；rt_w为用户u_i在区域r_j的周末的社交媒体数量占其所有周末的社交媒体签到数量的比例。

[0079] 为了实现对机器学习模型的训练，如图1所示，所述方法还包括如下步骤：

[0080] 步骤S200、根据所有区域的区域特征向量生成多维区域特征向量，根据所述多维区域特征向量以及基于所述多维区域特征向量生成的目标区域相关性向量对机器学习模型进行训练，将训练完毕的机器学习模型作为预测模型。

[0081] 具体地，本实施例采用的是监督学习类的机器学习模型，即将机器学习模型的训练过程变成一种学习任务，通过建立输入变量和输出变量之间的数学关系，使得机器学习模型学习如何从输入变量中预测输出变量。因此需要首先获取作为输入数据的多维区域特征向量以及作为输出数据的目标区域相关性向量，然后根据这两个向量对机器学习模型进行训练。训练完毕的机器学习模型就可以作为预测模型，例如用于预测用户的常驻区域。

[0082] 如图3所示，在一种实现方式中，所述步骤S200具体包括如下步骤：

[0083] 步骤S210、获取所有区域的区域特征向量并进行整合，将整合得到的向量作为多维区域特征向量；

[0084] 步骤S220、获取基于所述多维区域特征向量生成的目标区域相关性向量；

[0085] 步骤S230、将所述多维区域特征向量作为机器学习模型的输入数据，将所述基于所述多维区域特征向量生成的目标区域相关性向量作为所述机器学习模型的输出数据，对所述机器学习模型进行训练；

[0086] 步骤S240、将训练完毕的机器学习模型作为预测模型。

[0087] 具体地，本实施例通过获取所有区域的区域特征向量，将其进行整合，得到多维区域特征向量。然后根据所述多维区域特征向量生成的目标区域相关性向量。为了生成目标区域相关性向量，在一种实现方式中，本实施例通过根据所述多维区域特征向量中的每一个区域与目标区域的相关性，对每一个区域进行评分，得到所述每一个区域的目标区域相关性分值，然后对所述每一个区域的目标区域相关性分值进行整合，将整合得到的向量作为目标区域相关性向量。该步骤每一个区域的目标区域相关性分值可以正确指示每一个区域与目标区域的相关性远近。获取到所述多维区域特征向量以及所述目标区域相关性向量以后，将所述多维区域特征向量作为机器学习模型的输入数据，将所述目标区域相关性向量作为所述机器学习模型的输出数据，对所述机器学习模型进行训练，最后将训练完毕的机器学习模型作为预测模型。

[0088] 举例说明，当需要预测的目标区域为用户的常驻区域时，本实施例需要事先收集用户的常驻区域。具体地，可以通过网络爬虫，爬取用户的社交媒体的个人主页信息，或者在微博、推特、照片墙等社交媒体中，查询用户在个人信息中填入自己的现在所在城市，根据这些信息确定用户的常驻区域。对于该用户访问过的所有区域，若该区域为用户的常驻区域，则其相关性打分值为1；否则该区域相关性打分值为0。具体方法为，对于用户访问过的所有区域(r_1,r_2,…,r_m)，若该区域为用户的常驻区域，则其相关性打分值为1；否则该区域相关性打分值为0，构成m维向量(0,0,…,1,…,0)，直到每一个区域的区域相关性分值都计算完成。

[0089] 在一种实现方式中，鉴于多重决策树LambdaMART模型对于建立“排序学习”框架的搜索排序算法十分有效，因此本实施例中采用多重决策树LambdaMART模型作为预测模型。LambdaMART是一种Listwise类型的LTR算法，它基于LambdaRank算法和MART(Multiple Additive Regression Tree)算法，将搜索引擎结果排序问题转化为回归决策树问题。MART实际就是梯度提升决策树(GBDT,Gradient Boosting Decision Tree)算法。GBDT的核心思想是在不断的迭代中，新一轮迭代产生的回归决策树模型拟合损失函数的梯度，最终将所有的回归决策树叠加得到最终的模型。现有技术中，LambdaMART是一个非常成熟的模型，其整个训练过程已经十分地流程化。在模型训练时，只需要构造该模型的输入数据和输出数据作为训练数据即可实现对该模型的训练过程。

[0090] 获取到用于训练机器学习模型的输入数据以及输出数据以后，为了预测出用户的目标区域，如图1所示，所述方法还包括如下步骤：

[0091] 步骤S300、获取待预测的区域特征向量，通过所述预测模型以及所述待预测的区域特征向量对待预测的区域进行排序，根据所述排序结果在所述待预测的区域中确定目标区域。

[0092] 预测模型已经训练完毕，因此其能够根据输出数据自动预测出正确的输出数据。为了预测用户的目标区域，首先需要根据前述步骤生成所有待预测的区域的区域特征向量，然后将这些待预测的区域特征向量输入所述预测模型中，通过所述预测模型给这些待预测的区域进行排序，再根据所述排序结果确定所述待预测的区域中每一个区域与目标区域的相关性的大小，进而在所述待预测的区域中确定要预测的用户的目标区域。

[0093] 如图4所示，在一种实现方式中，所述步骤S300具体包括如下步骤：

[0094] 步骤S310、获取待预测的区域特征向量，将所述待预测的区域特征向量输入所述预测模型中；

[0095] 步骤S320、获取所述预测模型输出的基于所述待预测的区域特征向量生成的评分；

[0096] 步骤S330、基于所述评分对待预测的区域进行排序，根据排序结果在所述待预测的区域中确定目标区域。

[0097] 具体地，将这些待预测的区域的区域特征向量输入所述预测模型中，通过所述预测模型给这些待预测的区域进行目标区域相关性分值的评分，然后基于评分结果对这些待预测的区域进行排序，再根据所述排序结果在所述待预测的区域中确定预测的目标区域。在一种实现方式中，可以根据目标区域相关性分值的大小，由大到小对所述待预测区域中的每一个区域进行排序，当所要预测的目标区域为用户的常驻区域时，则可以将位于第一

顺序位上的区域作为该用户的常驻区域。

[0098] 为了说明本发明实施例提供的基于社交媒体签到预测目标区域的方法的效果，本发明实施例采用真实数据进行实验。图6是本发明使用真实的照片墙社交媒体签到数据得到的实验结果。本实施例使用Accuracy,F-measure,Balanced Accuracy等指标对本发明方法及其他对比方法进行量化评价。量化结果显示，本方法能得到较大的Accuracy,F-measure,Balanced Accuracy，证明了本方法较其他方法的优越性，当所要预测的目标区域为用户的常驻区域时，相比其他预测方法，本发明可以更准确地预测社交媒体用户的常驻区域。

[0099] 基于上述实施例，本发明还提供了一种智能终端，其原理框图可以如图5所示。该智能终端包括通过系统总线连接的处理器、存储器、网络接口、显示屏。其中，该智能终端的处理器用于提供计算和控制能力。该智能终端的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该智能终端的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现基于社交媒体签到预测目标区域的方法。该智能终端的显示屏可以是液晶显示屏或者电子墨水显示屏。

[0100] 本领域技术人员可以理解，图5中示出的原理框图，仅仅是与本发明方案相关的部分结构的框图，并不构成对本发明方案所应用于其上的智能终端的限定，具体的智能终端可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0101] 在一种实现方式中，所述智能终端的存储器中存储有一个或者一个以上的程序，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行基于社交媒体签到预测目标区域的方法的指令。

[0102] 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的硬件来完成，所述的计算机程序可存储于一非易失性计算机可读取存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本发明所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0103] 综上所述，本发明公开了一种基于社交媒体签到预测目标区域的方法，其特征在于，所述方法包括：获取所述社交媒体签到记录中的地理位置标签，根据所述地理位置标签生成区域特征向量；根据所有区域的区域特征向量生成多维区域特征向量，根据所述多维区域特征向量以及基于所述多维区域特征向量生成的目标区域相关性向量对机器学习模型进行训练，将训练完毕的机器学习模型作为预测模型；获取待预测区域的特征向量，通过所述预测模型对所述待预测区域进行排序，根据所述排序结果在所述待预测区域中确定目标区域。本发明通过将确定用户常驻区域的任务抽象为一个排序问题，并使用机器学习模型对用户访问过的各区域进行排序，最终成功预测用户的常驻区域。

[0104] 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。

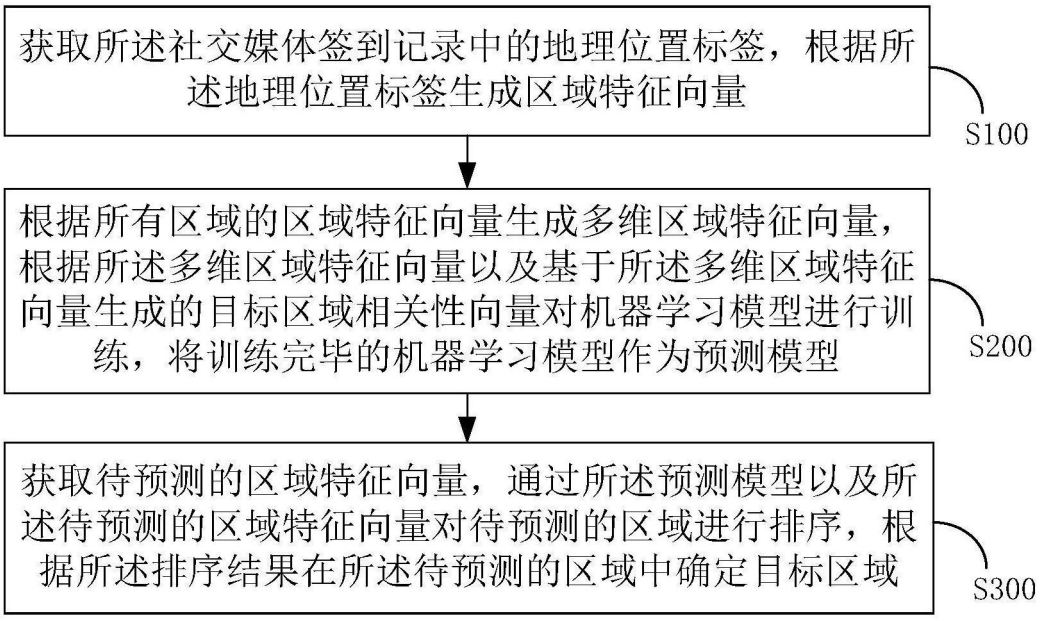

获取所述社交媒体签到记录中的地理位置标签，根据所述地理位置标签生成区域特征向量
S100
根据所有区域的区域特征向量生成多维区域特征向量，根据所述多维区域特征向量以及基于所述多维区域特征向量生成的目标区域相关性向量对机器学习模型进行训练，将训练完毕的机器学习模型作为预测模型
S200
获取待预测的区域特征向量，通过所述预测模型以及所述待预测的区域特征向量对待预测的区域进行排序，根据所述排序结果在所述待预测的区域中确定目标区域
S300

图1

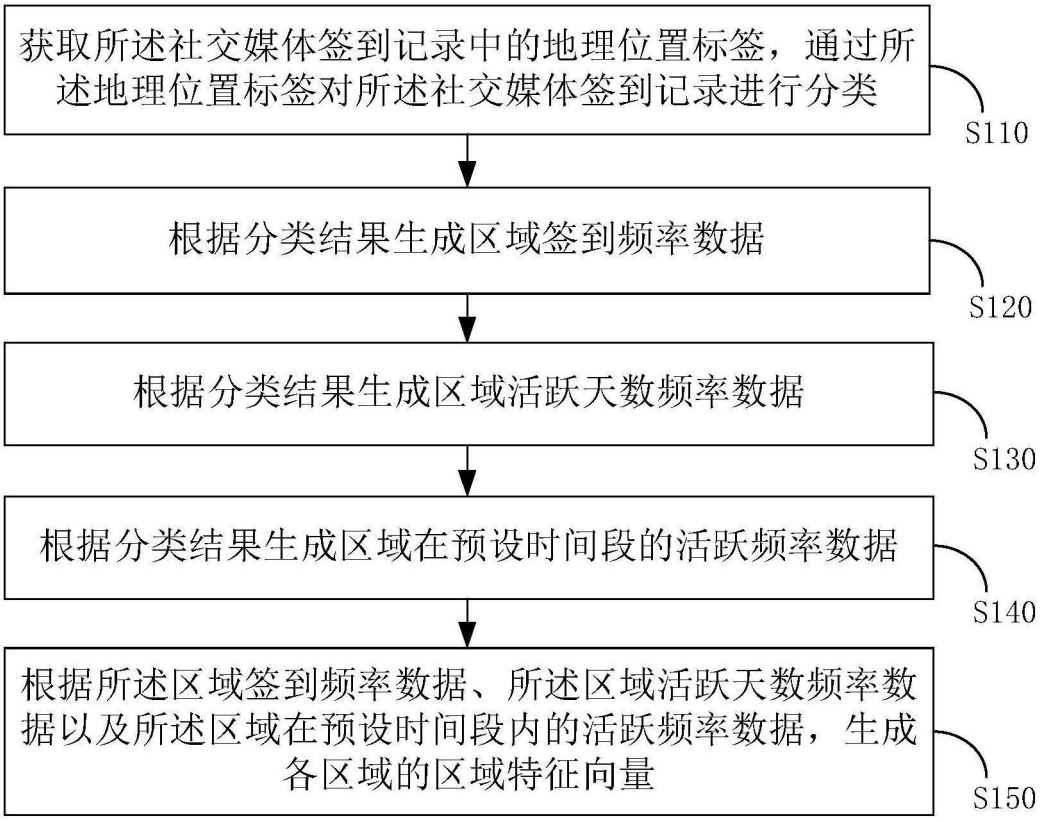

获取所述社交媒体签到记录中的地理位置标签，通过所述地理位置标签对所述社交媒体签到记录进行分类
S110
根据分类结果生成区域签到频率数据
S120
根据分类结果生成区域活跃天数频率数据
S130
根据分类结果生成区域在预设时间段的活跃频率数据
S140
根据所述区域签到频率数据、所述区域活跃天数频率数据以及所述区域在预设时间段内的活跃频率数据，生成各区域的区域特征向量
S150

图2

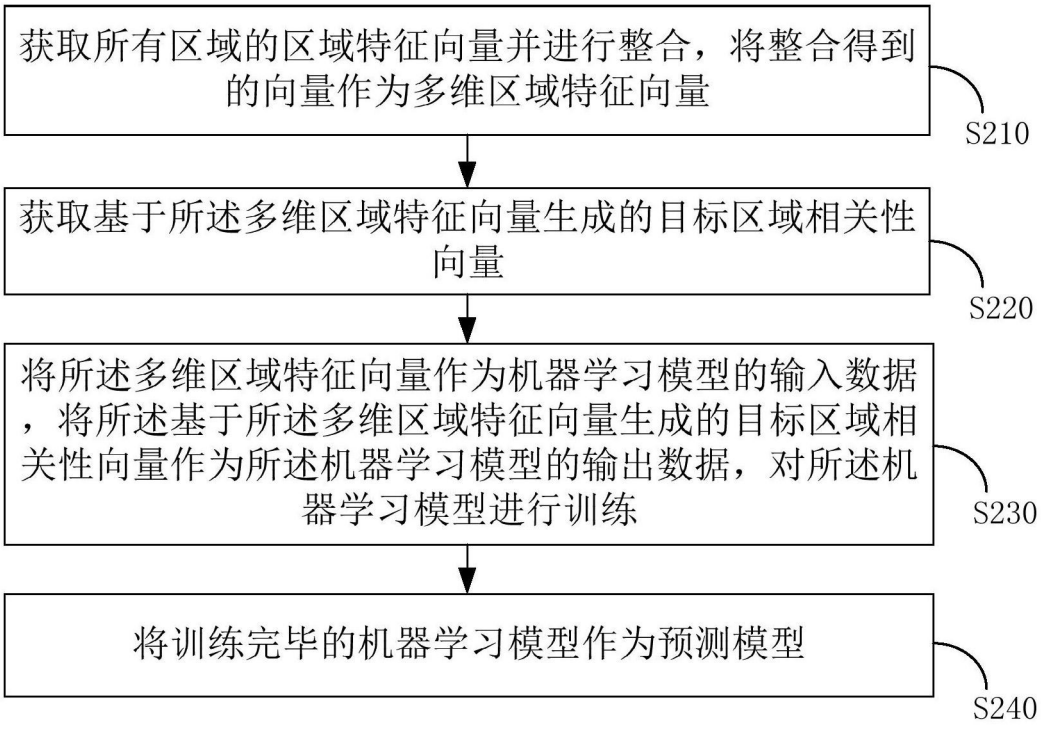

获取所有区域的区域特征向量并进行整合，将整合得到的向量作为多维区域特征向量
S210
获取基于所述多维区域特征向量生成的目标区域相关性向量
S220
将所述多维区域特征向量作为机器学习模型的输入数据，将所述基于所述多维区域特征向量生成的目标区域相关性向量作为所述机器学习模型的输出数据，对所述机器学习模型进行训练
S230
将训练完毕的机器学习模型作为预测模型
S240

图3

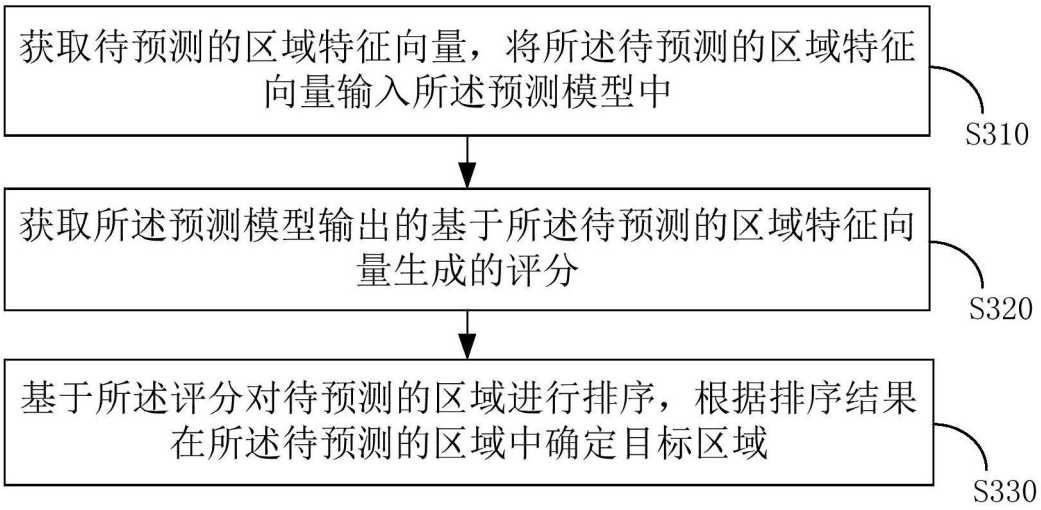

获取待预测的区域特征向量，将所述待预测的区域特征向量输入所述预测模型中
S310
获取所述预测模型输出的基于所述待预测的区域特征向量生成的评分
S320
基于所述评分对待预测的区域进行排序，根据排序结果在所述待预测的区域中确定目标区域
S330

图4

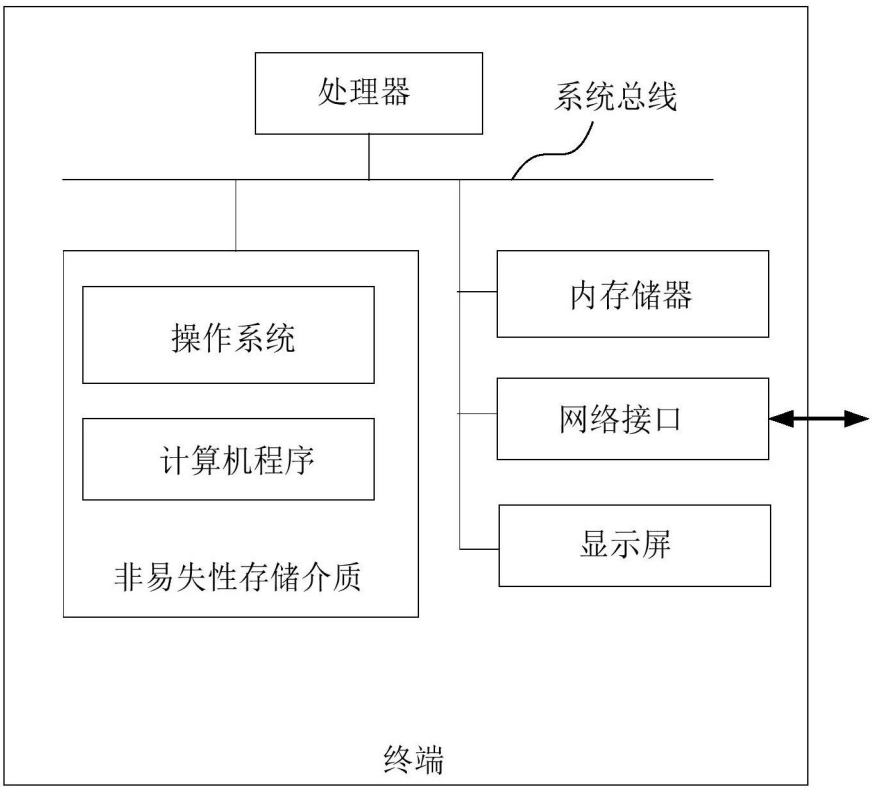


图5

	Accuracy (Home Country Detection)		F-measure (Recognize Tourist)		Accuracy (Recognize Tourist)		Balanced Accuracy (Recognize Tourist)	
	number	rank	number	rank	number	rank	number	rank
LambdaMART-HD	84.9%	1	74.3%	1	88.2%	1	80.4%	1
MP	82.5%	2	68.4%	4	86.7%	3	76.2%	4
MP-19-7	76.3%	4	67.0%	5	85.3%	5	75.8%	5
MAN	78.4%	3	69.6%	3	86.5%	4	77.3%	3
MFV	75.4%	5	73.5%	2	87.6%	2	80.3%	2

图6