(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106600053 A
(43)申请公布日 2017.04.26

(21)申请号 201611141121.3

(22)申请日 2016.12.12

(71)申请人 西安交通大学
地址 710049 陕西省西安市碑林区咸宁西路28号

(72)发明人 王平辉 孙飞扬 王迪 管晓宏 陶敬 张岩 曹鹏飞 贾鹏 胡小雨 曹宇 兰林

(74)专利代理机构 西安智大知识产权代理事务所 61215
代理人 段俊涛

(51)Int.Cl.
G06Q 10/04(2012.01)
G06F 17/30(2006.01)
G06K 9/62(2006.01)

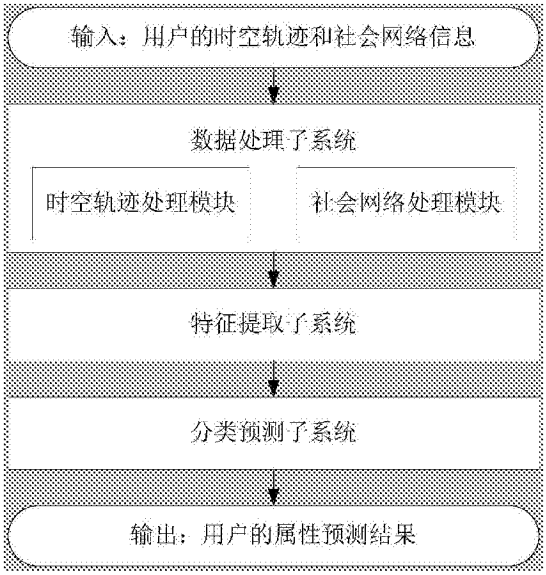
权利要求书2页 说明书6页 附图2页

(54)发明名称

一种基于时空轨迹和社会网络的用户属性预测系统

(57)摘要

本发明提供一种基于时空轨迹和社会网络的用户属性预测系统，通过对用户的行为模式进行分析，预测用户的真实身份属性。包括数据处理，特征提取和分类预测这三个子系统；对时空轨迹数据和社会网络数据进行分析；提出原创的非负张量分解(NTF)算法自动提取出用户的隐含特征；利用用户的隐含特征使用多种分类器对用户属性进行预测。本发明可用于用户属性真实性检测；也可用于根据预测出的属性进行精准推广。



CN 106600053 A

1.一种基于时空轨迹和社会网络的用户属性预测系统，其特征在于，包括：

数据处理子系统，包括时空轨迹处理模块和社会网络处理模块，时空轨迹处理模块将所有用户的时空轨迹处理成容易进行后续操作的三阶张量形式，社会网络处理模块将所有用户之间的社交关系处理成容易进行后续操作的邻接矩阵形式。

特征提取子系统，降低用户时空轨迹的维度，从用户的时空轨迹数据中提取出有价值的特征，使提取出的特征适用于现有的分类算法；

分类预测子系统，利用用户的隐含特征训练多种分类器，使用已知属性的用户隐含特征训练多种分类器对用户，用目标用户的隐含特征进行预测。

2.根据权利要求1所述基于时空轨迹和社会网络的用户属性预测系统，其特征在于，所述时空轨迹处理模块中，所需的原始的时空轨迹记录包括用户标识，地理位置标示和时间标识，时空轨迹处理模块建立一个元素全为零的三阶张量，其中行数＝用户标识数、列数＝地理位置标识数、管数＝时间段标识数，即三阶张量的每一行代表一个用户，每一列代表一个地点，每一管代表一个时间段。

3.根据权利要求1所述基于时空轨迹和社会网络的用户属性预测系统，其特征在于，所述社会网络处理模块中，所需的数据为用户的社会网络信息，用户的社会网络信息表现为用户间存在的某种关系，对这些信息进行提取，建立一个反映用户间社交关系的邻接矩阵。

4.根据权利要求1所述基于时空轨迹和社会网络的用户属性预测系统，其特征在于，所述特征提取子系统应用非负张量分解(NTF)算法来提取有价值的特征，所述非负张量分解(NTF)算法对时空轨迹张量进行分解，用社会网络信息进行约束，得到三个二阶矩阵，分别代表了每个用户、每个地理位置和每个时间段的隐含特征。

5.根据权利要求4所述基于时空轨迹和社会网络的用户属性预测系统，其特征在于，所述非负张量分解(NTF)算法包括：输入一个张量$\mathcal{X}\in\mathbb{R}^{m\times n\times h}$，其中m是用户数量，n是地理位置数量，h是时间段数；输入社会网络矩阵$A\in\mathbb{R}^{m\times m}$，若用户$u_i$和$u_j$有社交关系，则A(i,j)＝1，否则A(i,j)＝0，NTF算法将解决如下的优化问题：

$$\min_{U,V,T\geq 0} O=\underbrace{\left\|\mathcal{X}-\{U,V,T\}\right\|^2}_{O_{TRA}}+\alpha\underbrace{\mathrm{Tr}(U^{\mathrm{T}}LU)}_{O_U}+\gamma\underbrace{(\|U\|^2+\|V\|^2)}_{正则项}$$

上述优化问题也即目标函数，其中O_{TRA}代表对时空轨迹信息的分解，O_U代表利用社会网络进行约束，$\{U,V,T\}=\sum_{j=1}^{r}u_j\circ v_j\circ t_j$；U,V,T分别是用户、地理位置和时间段的隐含特征表示，$U\in\mathbb{R}^{m\times r},V\in\mathbb{R}^{n\times r},T\in\mathbb{R}^{h\times r}$是待学习的非负矩阵，r是隐含特征的维度；$\circ$表示向量外积；$u_{:j},v_{:j},t_{:j}$分别是矩阵U,V,T的第j列；L＝D-A，$D\in\mathbb{R}^{m\times m}$，$D(i,i)=\sum_{j=1}^{m}A(i,j)$；α，γ为调节参数。

6.根据权利要求5所述基于时空轨迹和社会网络的用户属性预测系统，其特征在于，所述目标函数的乘法更新规则(multiplicative updating rule)为：

$$U(i,j)\leftarrow U(i,j)\sqrt{\frac{\left[\mathcal{X}_{(1)}(T\odot V)+\alpha AU\right](i,j)}{\left[U(T^{\mathrm{T}}T*V^{\mathrm{T}}V)+\alpha DU+\gamma U\right](i,j)}}$$

$$V(i,j) \leftarrow V(i,j) \sqrt{\frac{\left[X_{(2)}(T \odot U)\right](i,j)}{\left[V(T^{\mathrm{T}}T * U^{\mathrm{T}}U) + \gamma V\right](i,j)}}$$

$$T(i,j) \leftarrow T(i,j) \sqrt{\frac{\left[X_{(3)}(V \odot U)\right](i,j)}{\left[T(V^{\mathrm{T}}V * U^{\mathrm{T}}U) + \gamma T\right](i,j)}}$$

其中$X_{(1)}$,$X_{(2)}$,$X_{(3)}$分别是张量X的模-1,模-2,模-3展开;⊙表示Khatri-Rao乘积;*表示Hadamard乘积,U,V,T矩阵的初始值随机生成,但必须保证非负,最终迭代得到用户的隐含特征矩阵U。

7.根据权利要求1所述基于时空轨迹和社会网络的用户属性预测系统,其特征在于,所述分类预测子系统使用多种分类器对用户属性进行预测,最后综合判断用户属性。

一种基于时空轨迹和社会网络的用户属性预测系统

技术领域

[0001] 本发明属于数据挖掘技术领域，特别涉及一种基于时空轨迹和社会网络的用户属性预测系统。

背景技术

[0002] 随着互联网技术的应用与发展，互联网用户越来越多。互联网具有虚拟性，用户在互联网上的资料并不一定是其真实属性，为了提高互联网的安全性，需要确保用户身份的真实性。

[0003] 移动通信技术的发展和智能移动设备(如智能手机、平板电脑)的快速普及，使移动设备与用户之间的联系越来越密切，而许多移动设备和APP能够记录用户的行动。于是利用用户行为对用户属性做出推断这一问题吸引了很多研究人员的关注。

[0004] 这里主要关注移动设备记录下的用户地理位置信息。例如，许多用户喜欢在微信、微博等社交平台上发布自己的消息；使用移动设备上的购物或团购APP；使用地图和导航功能；为了能够随时使用这些功能，大多数用户会长时间开启GPS、WIFI或4G通讯。开发上述APP的第三方供应商以及网络运营商能够获取到用户的使用记录，再通过一些方法分析出这些记录产生的时间和地点。例如，如果一个用户用手机发布了一条微博，APP可以通过4G基站信息和手机内置的GPS功能获取当前的地理位置；网络运营商可以通过多个基站对用户的地理位置进行定位。将一个用户的每一条时间地点记录组成一个序列，就得到了一个用户的时空轨迹。时空轨迹反映了用户的行动模式。

[0005] 目前已经有一些方法通过分析用户的时空轨迹来推断用户属性，但是这些方法都是基于地理位置的语义信息来做的。例如，一个微博用户在几个不同地点发布了微博消息，为了判断这名用户的属性(如性别、职业)，传统方法需要知道微博发布地点的信息(如商场、公司、饭店或游乐园)。显然，地点的语义信息并不是总能明确获取的，例如一栋高层建筑的不同楼层可能有不同的功能。这对传统方法的效果有很大的影响。此外，由于每个用户在同一个地点出现的目的都是不同的，只凭时空轨迹来推断用户属性必然存在瓶颈，需要加入新的特征来突破。

[0006] 社会网络是由用户的好友关系建立起来的网络图，图中每个节点代表一个用户，每条边代表一对好友关系。有研究统计发现，社会网络中的好友具有“同质性”，即一对好友具有一项或几项相同属性的概率很高。但若仅使用社会网络来推断用户属性，首先需要知道社会网络中大多数节点的属性，但由于隐私问题，这在实际应用中是难以获取的。

发明内容

[0007] 为了克服上述现有技术的缺点，本发明的目的在于提供一种基于时空轨迹和社会网络的用户属性预测系统，与传统方法相比，本发明的一项优势在于输入的时空轨迹不需要具有详细语义信息的地理位置数据，因此适用于多种不同类型的数据集；本发明的另一项优势在于输入的社会网络不需要具有用户的身份信息，因此适用于不同的社会网络。

[0008] 为了实现上述目的，本发明采用的技术方案是：

[0009] 基于时空轨迹和社会网络的用户属性预测系统，包括：

[0010] 数据处理子系统，包括时空轨迹处理模块和社会网络处理模块。

[0011] 时空轨迹处理模块用于将所有用户的时空轨迹处理成容易进行后续操作的三阶张量形式。

[0012] 具体地，时空轨迹处理模块中，所需的原始的时空轨迹记录包括用户标识，地理位置标示和时间标识，时空轨迹处理模块建立一个元素全为零的三阶张量，其中行数＝用户标识数、列数＝地理位置标识数、管数＝时间段标识数，即三阶张量的每一行代表一个用户，每一列代表一个地点，每一管代表一个时间段。

[0013] 所述社会网络处理模块用于将所有用户之间的社交关系处理成容易进行后续操作的邻接矩阵形式。所需的数据为用户的社会网络信息用户间必定存在某种关系(如好友，关注，点赞等)，对这些信息进行提取，建立一个反映用户间社交关系的邻接矩阵。

[0014] 具体地，邻接矩阵的行数和列数都等于用户数，用户u_i和用户u_j的关系反映在矩阵的第i行j列中。

[0015] 特征提取子系统，降低用户时空轨迹的维度，从用户的时空轨迹数据中提取出有价值的特征，使提取出的特征适用于现有的分类算法。

[0016] 具体地，本发明提出了一种非负张量分解(NTF)算法来提取有价值的特征，对时空轨迹张量进行分解，用社会网络信息进行约束，得到三个二阶矩阵，分别代表了每个用户、每个地理位置和每个时间段的隐含特征。其中本发明最关心用户隐含特征矩阵，它能反映每个用户的特征，用于分类器的训练和预测，同时特征的维度可以根据需要自行设定，满足高效、准确的要求。

[0017] 分类预测子系统，利用用户的隐含特征训练多种分类器，使用已知属性的用户隐含特征训练多种分类器对用户，用目标用户的隐含特征进行预测。

[0018] 具体地，本发明目前使用了SVM，Logistic回归和线性回归三种分类器，这三种分类器的优点是实现简单，运行效率高，分类准确率高。

[0019] 与现有技术相比，本发明的有益效果是：

[0020] 1、突破了现有的基于时空轨迹的用户属性预测技术依赖地理位置信息的限制。

[0021] 本发明所需的时空轨迹信息不需要任何地理位置特征，可以用简单的标识(如地点1等)替代，这就大大提高了本发明的适用性，同时由于加入了社会网络的信息，预测精度相对现有技术有明显提高。

[0022] 2、结合社会网络信息，预测能力得到明显提高。

[0023] 本发明将社会网络数据和时空轨迹数据结合在一起，相较于独立使用社会网络的预测技术，不需要事先知道网络中重点用户的属性，不存在隐私问题和数据缺乏的问题。

[0024] 3、能够处理大数据的分类预测问题。

[0025] 在时空轨迹数据量极大时，由于特征可能高于训练样本数量，现有技术往往会遇到过拟合问题，严重影响预测能力。本发明提出了一种非负张量分解算法，对时空轨迹进行了降维处理，能够自行设定特征数量，彻底克服了这个问题。

附图说明

[0026] 图1为本发明系统整体结构图。

[0027] 图2为本发明时空轨迹处理模块流程图。

[0028] 图3为本发明社会网络处理模块流程图。

[0029] 图4为本发明特征提取子系统流程图。

[0030] 图5为本发明分类预测子系统流程图。

具体实施方式

[0031] 为使本发明实施例的目的、技术方案和优点更加清楚，下面结合附图和实施例详细说明本发明的实施方式。

[0032] 如图1所示，本系统由三个子系统组成，分别是数据处理子系统，特征提取子系统和分类预测子系统。系统的输入数据包括两部分：用户的时空轨迹信息和社会网络信息。其中用户的时空轨迹信息应该包括三部分：用户标识，地理位置标识和时间标识。即每一行表示：某个用户在某时间去过某地点。社会网络信息同样包括三部分：用户1标识，用户2标识，社交关系。即每一行表示：用户1和用户2具有社会关系。

[0033] 值得说明的是，输入数据中有一部分用户的属性是未知的(称为目标用户)，需要预测这些用户的属性，为此需要一些属性已知的用户的时空轨迹数据。与现有技术不同是，本发明并不是必需获取与目标用户具有密切社会关系的用户的信息，而只需要与目标用户去过相同地点的用户信息；同时也不需获取地理位置的语义信息，这大大降低了信息获取难度。

[0034] 首先，将上述数据输入数据处理子系统，该子系统包括时空轨迹处理模块，用于将所有用户的时空轨迹处理成容易进行后续操作的三阶张量形式；该子系统还包括社会网络处理模块，用于将所有用户之间的社交关系处理成容易进行后续操作的邻接矩阵形式。

[0035] 接着，处理后的时空轨迹张量和社会网络矩阵被送入特征提取子系统，使用一种原创的非负张量分解(NTF)算法得到每一个用户的隐含特征。该非负张量分解算法，对时空轨迹张量进行分解，用社会网络信息进行约束，得到三个二阶矩阵，分别代表了每个用户、每个地理位置和每个时间段的隐含特征。其中本发明最关心用户隐含特征矩阵，它能反映每个用户的特征，用于分类器的训练和预测，同时特征的维度可以根据需要自行设定，满足高效、准确的要求。

[0036] 最后，将用户的隐含特征矩阵送入分类预测子系统，使用已知属性的用户隐含特征训练多种分类器对用户，用目标用户的隐含特征进行预测。本发明目前使用了SVM，Logistic回归和线性回归三种分类器，这三种分类器的优点是实现简单，运行效率高，分类准确率高。

[0037] 本发明中各个子系统的详细介绍如下：

[0038] 1、数据处理子系统

[0039] 主要实现输入数据的预处理，包括将所有用户的时空轨迹处理成容易进行后续操作的三阶张量形式，以及将所有用户之间的社交关系处理成容易进行后续操作的邻接矩阵形式。

[0040] 数据处理子系统包括时空轨迹处理模块和社会网络处理模块：

[0041] 其中时空轨迹处理模块用于将所有用户的时空轨迹处理成三阶张量形式。原始的

时空轨迹记录包括用户标识，地理位置标示和时间标识；三阶张量的每一行代表一个用户，每一列代表一个地点，每一管代表一个时间段。首先获取用户的时空轨迹，包括用户标识、地理位置标识和时间段标识。建立一个元素全为零的三阶张量，其中行数＝用户标识数、列数＝地理位置标识数、管数＝时间段标识数。将每个用户的时空轨迹填入张量，每个元素代表用户某时间段在某地点的出现次数。这样就得到一个反映用户时空轨迹的张量，能够区分不同时间段、不同地理位置对属性预测的不同影响。

[0042] 值得注意的是，在通常的数据量下，该张量的规模也是相当大的。假设有一万个用户，一千个不同的地理位置和一百个时间段，则张量的规模已经达到了10000*1000*100＝10^9，每一个用户的特征数量为1000*100＝10^5个，特征数量过多会导致分类器过拟合，影响预测效果，因此必须对张量进行处理。

[0043] 数据处理子系统还包括社会网络处理模块，用于将所有用户之间的社交关系处理成容易进行后续操作的邻接矩阵形式。首先获取用户的社会网络信息，用户间必定存在某种关系(如好友，关注，点赞等)，对这些信息进行提取，建立一个元素全为零的二阶矩阵，其中行数＝列数＝用户数。将用户间的社交关系填入矩阵，每个元素代表所在行、列的用户是否有社交关系。例如，用户u_i和用户u_j的关系反映在矩阵的第i行j列中。这样就得到一个反映用户间社交关系的邻接矩阵。

[0044] 或者，可以采用如下格式：用户1标识，用户2标识，社交关系，即每一行表示：用户1和用户2具有社会关系。

[0045] 2、特征提取子系统

[0046] 主要功能是降低用户时空轨迹的维度，从用户的时空轨迹数据中提取出有价值的特征，使提取出的特征适用于现有的分类算法。

[0047] 具体地，本发明应用了一种原创的非负张量分解(NTF)算法，输入数据是从上述数据处理子系统获得的时空轨迹张量和社会网络矩阵，通过非负张量分解算法得到每一个用户的隐含特征，其中使用社会网络信息进行约束，得到三个二阶矩阵，分别代表了每个用户、每个地理位置和每个时间段的隐含特征。其中本发明最关心用户隐含特征矩阵，它能反映每个用户的特征，用于分类器的训练和预测，同时特征的维度可以根据需要自行设定，满足高效、准确的要求。

[0048] 上述非负张量分解(NTF)算法详细描述如下：

[0049] 输入一个张量$\mathcal{X}\in\mathbb{R}^{m\times n\times h}$，其中m是用户数量，n是地理位置数量，h是时间段数。NTF算法是将张量X分解成三个低维矩阵的形式，实际求解时，将此问题转化为如下的优化问题：

[0050] $$\min_{U,V,T\geq 0} O_{\mathrm{TRA}}=\left\|\mathcal{X}-\hat{\mathcal{X}}\right\|^2 \qquad (1)$$

[0051] 其中$\hat{\mathcal{X}}=\{U,V,T\}\equiv\sum_{j=1}^{r}u_j\circ v_j\circ t_j$；$U\in\mathbb{R}^{m\times r},V\in\mathbb{R}^{n\times r},T\in\mathbb{R}^{h\times r}$是待学习的非负矩阵，r是隐含特征的维度；{•}表示秩-1张量和；o表示向量外积；$u:_j,v:_j,t:_j$分别是矩阵U,V,T的第j列。U,V,T分别是用户、地理位置和时间段的隐含特征表示。

[0052] NTF算法具有可扩展性，可以根据需求加入先验知识，以提高隐含特征的准确性。本发明中加入了社会网络信息作为先验知识。

[0053] 输入社会网络矩阵$A\in\mathbb{R}^{m\times m}$，若用户$u_i$和$u_j$有社交关系，则A(i,j)＝1，否则A(i,j)＝0。由于具有社交关系的用户更可能具有相同的属性，对应于上述优化问题，将最小化以下损失函数：

[0054] $$O_U=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|u_i-u_j\right\|^2A(i,j) \tag{2}$$

[0055] 如果具有社交关系的用户u_i和u_j有不同的隐含特征，上述损失函数将给予惩罚，大小为$||u_{:i}-u_{:j}||^2$。

[0056] 建立一个对角矩阵$D\in\mathbb{R}^{m\times m}$，使得$D(i,i)=\sum_{j=1}^{m}A(i,j)$，取L＝D-A，根据一系列推导，可以将上述(2)式重写成如下形式：

[0057] $$\begin{aligned}O_U&=\sum_{i=1}^{m}\sum_{j=1}^{m}(u_iA(i,j)u_i^{\mathrm{T}}-u_iA(i,j)u_j^{\mathrm{T}})\\&=\mathrm{Tr}(U^{\mathrm{T}}LU)\end{aligned}$$

[0058] 结合上述两式，就可以得到NTF算法的目标函数：

[0059] $$\min_{U,V,T\geq0}O=\underbrace{\left\|X-\{U,V,T\}\right\|^2}_{O_{\mathrm{fitting}}}+\alpha\underbrace{\mathrm{Tr}(U^{\mathrm{T}}LU)}_{O_U}+\gamma\underbrace{(\left\|U\right\|^2+\left\|V\right\|^2)}_{\text{正则项}} \tag{3}$$

[0060] 其中第一项是时空轨迹信息；第二项是社会网络信息；第三项是正则项，用来防止过拟合；α，γ参数可以调节。

[0061] 求解该优化问题用到了拉格朗日乘子法(Lagrange Multiplier)和KKT条件(Karush-Kuhn-Tucker)。最终得到如下的乘法更新规则(multiplicative updating rule)：

[0062] $$U(i,j)\leftarrow U(i,j)\sqrt{\frac{\left[X_{(1)}(T\odot V)+\alpha AU\right](i,j)}{\left[U(T^{\mathrm{T}}T*V^{\mathrm{T}}V)+\alpha DU+\gamma U\right](i,j)}} \tag{4}$$

[0063] $$V(i,j)\leftarrow V(i,j)\sqrt{\frac{\left[X_{(2)}(T\odot U)\right](i,j)}{\left[V(T^{\mathrm{T}}T*U^{\mathrm{T}}U)+\gamma V\right](i,j)}} \tag{5}$$

[0064] $$T(i,j)\leftarrow T(i,j)\sqrt{\frac{\left[X_{(3)}(V\odot U)\right](i,j)}{\left[T(V^{\mathrm{T}}V*U^{\mathrm{T}}U)+\gamma T\right](i,j)}} \tag{6}$$

[0065] 其中$X_{(1)}$，$X_{(2)}$，$X_{(3)}$分别是张量X的模-1，模-2，模-3展开；⊙表示Khatri-Rao乘积；*表示Hadamard乘积。U,V,T矩阵的初始值随机生成，但必须保证非负，这样的话由于使用了乘法更新规则，可以保证最终得到的U,V,T矩阵非负。经过数轮迭代之后，NTF算法能够收敛。NTF算法整体的时间复杂度为O(mnhr)，在普通计算机上即可实现，并且NTF算法很容易扩展到分布式系统上并行处理。总体来说，NTF算法是一种高效，准确的非负张量分解算法。

[0066] 综上，特征提取子系统的流程如下：

[0067] 输入用户的时空轨迹张量和社会网络信息，使用(4)，(5)，(6)三式迭代更新U,V,T

矩阵，最终得到用户的隐含特征矩阵U。

[0068] 3、分类预测子系统

[0069] 主要功能是利用用户的隐含特征训练多种分类器，使用已知属性的用户隐含特征训练多种分类器对用户，用目标用户的隐含特征进行预测。

[0070] 本发明目前使用了SVM，Logistic回归和线性回归三种分类器。现有的scikit-learn工具提供了大量分类器算法，可以利用其中的算法来实现分类预测子系统中的部分功能。scikit-learn是一个基于Python的科学计算库，提供了数种分类算法可供选择，分类预测子系统选择了SVM分类器(sklearn.svm)，Logistic回归(sklearn.linear_model.LogisticRegression)，和线性回归(sklearn.linear_model.LinearRegression)。

[0071] 从特征提取子系统获得的用户隐含特征矩阵U，其中一部分用户的属性是已知的，使用这部分用户的隐含特征作为训练集，训练分类器，再使用该分类器对其余用户的属性进行预测。由于分类器可能有错判，因此分类预测子系统使用了三种分类器同时对用户进行预测，如果多数分类器预测出了同一结果，则取该预测结果作为最终判断。

[0072] 综上，本发明提供的一种基于时空轨迹和社会网络的用户属性预测系统，通过对用户的行为模式进行分析，预测用户的真实身份属性。本发明可用于用户属性真实性检测；也可用于根据预测出的属性进行精准推广。

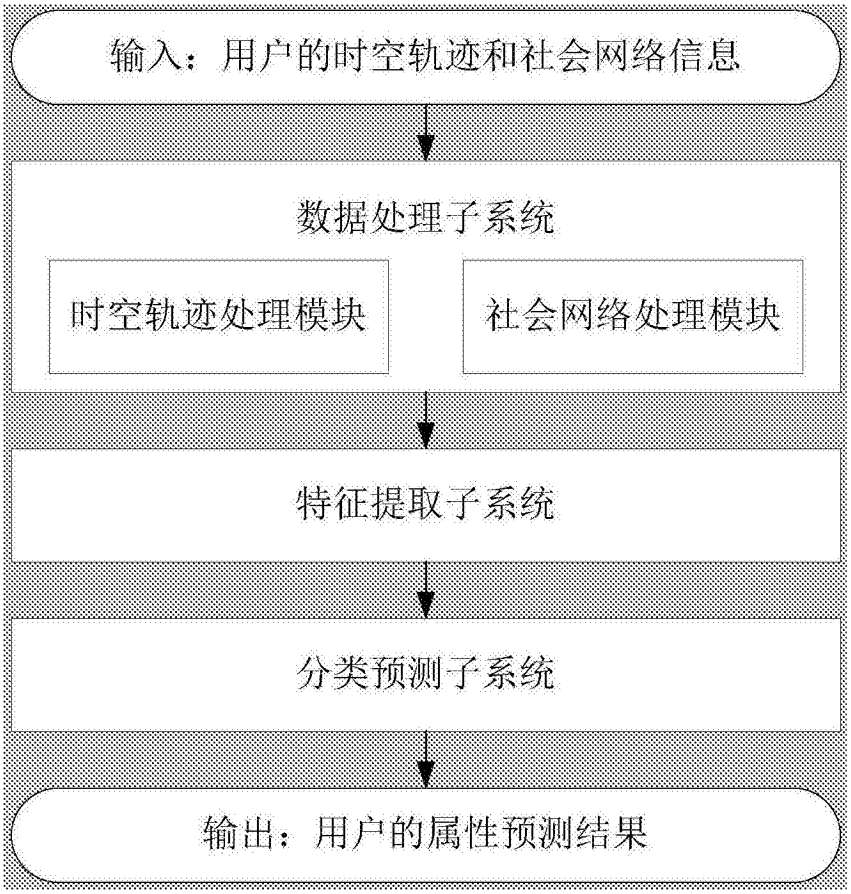

输入：用户的时空轨迹和社会网络信息
数据处理子系统
时空轨迹处理模块
社会网络处理模块
特征提取子系统
分类预测子系统
输出：用户的属性预测结果

图1

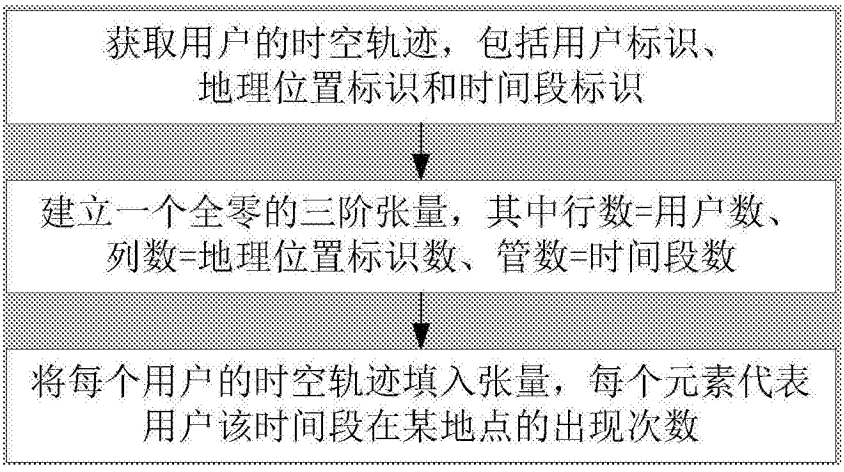

获取用户的时空轨迹，包括用户标识、
地理位置标识和时间段标识
建立一个全零的三阶张量，其中行数=用户数、
列数=地理位置标识数、管数=时间段数
将每个用户的时空轨迹填入张量，每个元素代表
用户该时间段在某地点的出现次数

图2

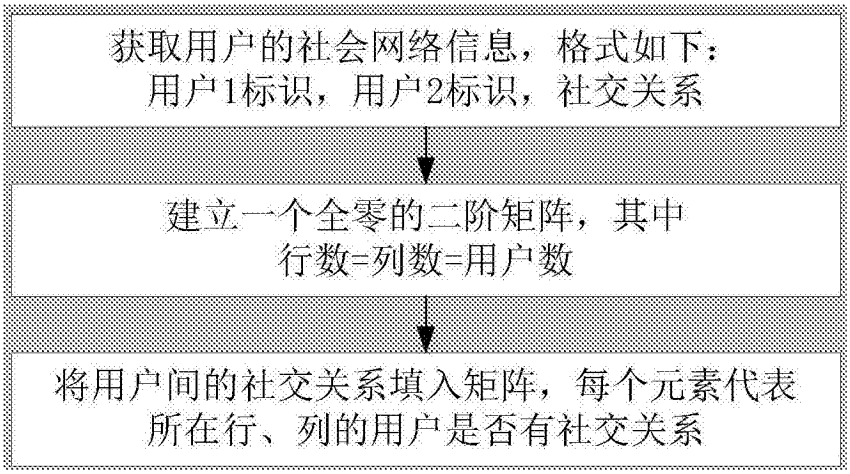

获取用户的社会网络信息，格式如下：
用户1标识，用户2标识，社交关系
建立一个全零的二阶矩阵，其中
行数=列数=用户数
将用户间的社交关系填入矩阵，每个元素代表
所在行、列的用户是否有社交关系

图3

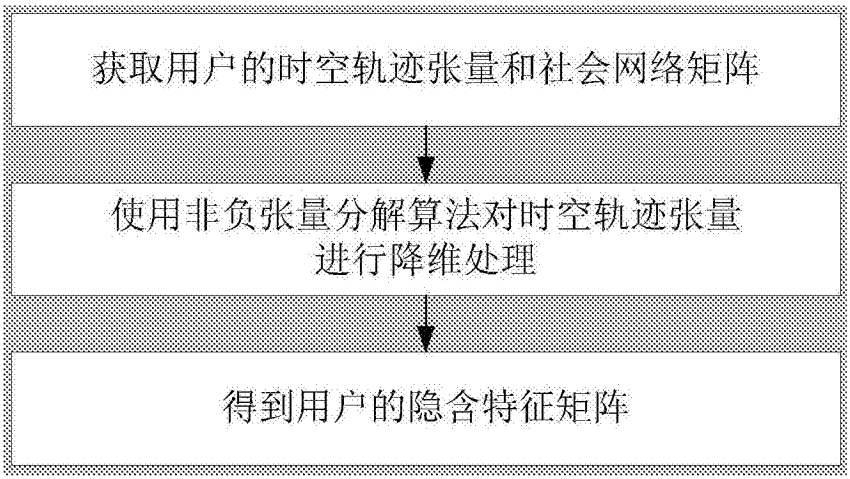

获取用户的时空轨迹张量和社会网络矩阵
使用非负张量分解算法对时空轨迹张量
进行降维处理
得到用户的隐含特征矩阵

图4

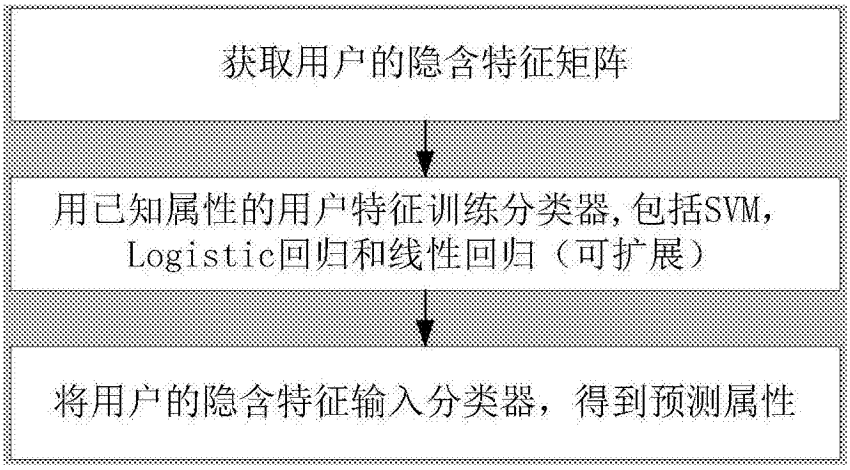

获取用户的隐含特征矩阵
用已知属性的用户特征训练分类器,包括SVM,
Logistic回归和线性回归（可扩展）
将用户的隐含特征输入分类器，得到预测属性

图5