



(12)发明专利申请

(10)申请公布号 CN 107194011 A

(43)申请公布日 2017.09.22

(21)申请号 201710488445.2

(22)申请日 2017.06.23

(71)申请人 重庆邮电大学

地址 400065 重庆市南岸区南山街道崇文路2号

(72)发明人 尚凤军 刘海昇

(74)专利代理机构 北京国坤专利代理事务所 (普通合伙) 11491

代理人 姜彦

(51) Int. Cl.

G06F 17/30(2006.01)

G06N 5/04(2006.01)

G06Q 50/00(2012.01)

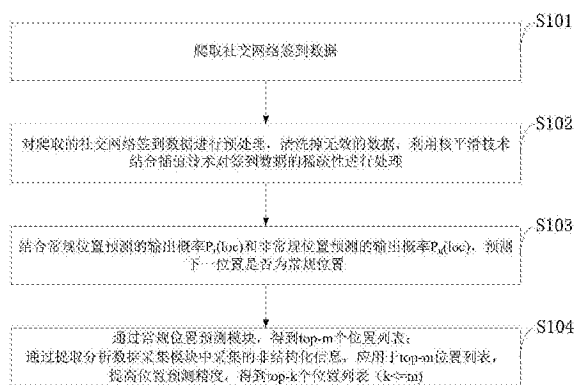
权利要求书2页 说明书5页 附图2页

(54)发明名称

一种基于社交网络的位置预测系统及方法

(57)摘要

本发明属于社交网络位置预测技术领域,公开了一种基于社交网络的位置预测系统及方法,包括:爬取社交网络签到数据;对爬取的社交网络签到数据进行预处理,清洗掉无效的数据,利用核平滑插值技术对签到数据的稀疏性进行处理;结合常规位置预测的输出概率和非常规位置预测的输出概率,预测下一位置是否为常规位置;通过常规位置预测得到top-m个位置列表;通过提取分析数据中采集的非结构化信息,应用于top-m位置列表,得到top-k个位置列表。通过知识图谱和位置推荐系统对非常规位置预测。本发明解决了位置预测中冷启动问题、非常规位置预测问题;提高了预测精度。



1. 一种基于社交网络的位置预测方法,其特征在于,所述基于社交网络的位置预测方法包括以下步骤:

步骤一,爬取社交网络签到数据;

步骤二,对爬取的社交网络签到数据进行预处理,过滤掉签到次数小于平均签到次数的数据,清洗掉无效的数据,利用核平滑插值技术对签到数据的稀疏性进行处理;在 $f(x)$ 中,若使用邻域样本的均值进行插值,则使 $f(x)$ 不平滑,所以使用一个核函数对估计值平滑;具体使用核加权平均,公式为:

$$f(x_0) = \frac{\sum_{i=1}^N K(x_0, x_i) y_i}{\sum_{i=1}^N K(x_0, x_i)} ;$$

其中 $K(\cdot)$ 采用高斯核函数,可见,离 x_0 越近的影响力越大,对应输出的权越大,符合签到数据的实际模拟;

步骤三,结合常规位置预测的输出概率 $P_r(1oc)$ 和非常规位置预测的输出概率 $P_u(1oc)$,预测下一位置是否为常规位置;

步骤四,通过常规位置预测模块,得到 $top-m$ 个位置列表;通过提取分析数据采集模块中采集的非结构化信息,应用于 $top-m$ 位置列表,提高位置预测精度,得到 $top-k$ 个位置列表, $k \leq m$ 。

2. 如权利要求1所述的基于社交网络的位置预测方法,其特征在于,所述预测下一位置是否为常规位置公式为:

$$P(1oc) = \lambda P_r(1oc) + (1-\lambda) P_u(1oc) ;$$

其中 $P_r(1oc)$ 为常规位置预测概率, $P_u(1oc)$ 为非常规位置预测概率, λ 为调节参数, $\lambda \in \{0,1\}$ 。

3. 如权利要求1所述的基于社交网络的位置预测方法,其特征在于,所述常规位置预测采用MHMM算法,HMM结合时间特征和空间特征对位置进行预测;选用混合HMM算法对下一位置进行预测;

$L_{t+1} = \arg \max P(C_{t+1} | S_t, \bar{x})$,其中 C_{t+1} 为 $t+1$ 时刻的位置类别, S_t 为 t 时刻的观察序列状态, \bar{x} 为时间空间向量。

4. 如权利要求1所述的基于社交网络的位置预测方法,其特征在于,所述非常规位置预测结合构建知识图谱,挖掘社交关系,采用融合社交关系的马尔科夫模型结合位置推荐系统对非常规位置进行预测;首先利用签到数据集构建知识图谱,在知识图谱上进行推理,其次基于历史签到数据并融合相似用户训练一个马尔科夫模型对下一位置进行预测;最后将马尔科夫模型和位置推荐系统结合在一起,提高位置预测精度。

5. 如权利要求4所述的基于社交网络的位置预测方法,其特征在于,所述非常规位置预测结合构建知识图谱,挖掘社交关系,采用融合社交关系的马尔科夫模型结合位置推荐系统对非常规位置进行预测;具体包括:

用签到数据集作为数据来源,构建社交知识图谱,在知识图谱上进行推理;推理方法包括三类:Embedding-based技术,Pathranking algorithms和Probabilistic graphical

models概率模型;社交关系的推理采用Embedding-based技术;Embedding-based技术以隐式因子模型为基本思想的方法,为基于低维向量的表示方法,将知识图谱中的实体和关系在低维的向量空间里进行表达,然后进行推理;

首先将实体和关系进行向量表示;其次,定义打分函数来衡量关系成立的可能性;再者,参数估计,根据打分函数推理相似用户;相似用户推理综合打分函数结构化信息和非结构化分析模块提取的兴趣相似度,表示如下:

$$\text{sim}(u, v) = \alpha s(u, v) + (1-\alpha) w(u, v);$$

其中, α 是调节参数,取值为 $[0, 1]$,反应结构化信息和非结构化信息相似度所占比重, $s(u, v)$ 代表结构化信息相似度, $w(u, v)$ 代表非结构信息相似度;非结构化信息相似度是非结构化信息分析模块提取兴趣关键字,然后根据余弦相似度计算用户之间的相似度;计算如下:

$$w(u, v) = \frac{u_i \cdot u_j}{|u_i| |u_j|};$$

其中 u_i, u_j 表示用户 i 和用户 j 兴趣关键字向量表示;余弦值取值范围为 $[0, 1]$,0表示完全不同,1表示完全相同;

在历史签到数据的基础上,融合相似用户的影响,训练马尔科夫模型对位置进行预测,即 $L_m = \max P(A_m | H, \text{sim}(u, v))$,其中 A_m 代表马尔科夫算法, H 代表历史签到数据, $\text{sim}(u, v)$ 代表用户相似度;

最后,综合融合相似用户的马尔科夫模型和位置推荐系统,提高位置预测精度;其公式如下:

$$L_{\text{top-n}} = \beta L_m + (1-\beta) S_r;$$

其中 L_m 代表融合相似用户的马尔科夫模型, S_r 代表位置推荐系统, β 为调节权重,取0.6;综合融合相似用户的马尔科夫模型和位置推荐系统,得到top-n位置列表。

6.一种如权利要求1所述基于社交网络的位置预测方法的基于社交网络的位置预测系统,其特征在于,所述基于社交网络的位置预测系统包括:

数据采集模块,应用爬虫系统,爬取社交网络签到数据;

数据预处理模块,对爬取的社交网络签到数据进行预处理,清洗掉无效的数据,利用核平滑插值技术对签到数据的稀疏性进行处理;

判断模块,结合常规位置预测的输出概率 $Pr(loc)$ 和非常规位置预测的输出概率 $Pu(loc)$,预测下一位置是否为常规位置;

常规位置预测模块,用于对预测位置进行分类,先预测位置的类别,预测位置;

非结构化数据分析模块,通过常规位置预测模块,得到top-m个位置列表;通过提取分析数据采集模块中采集的非结构化信息,应用于top-m位置列表,提高位置预测精度,得到top-k个位置列表;

非常规位置预测模块,结合构建知识图谱,挖掘相似用户,采用融合相似用户的马尔科夫模型结合位置推荐系统对非常规位置进行预测。

7.一种应用权利要求1~5任意一项所述基于社交网络的位置预测方法的社交网络终端。

一种基于社交网络的位置预测系统及方法

技术领域

[0001] 本发明属于社交网络位置预测技术领域,尤其涉及一种基于社交网络的位置预测系统及方法。

背景技术

[0002] 随着互联网的快速发展和可定位设备的大量普及,基于地理位置服务的网络应用越来越普及,如定向广告(targeted advertisement)、跟踪人口流动、预防疾病蔓延、网络安全、性能优化等,地址位置作为一种质量极高的信息资源被广泛应用。同时伴随着在线社交网络的发展,位置服务和在线社交网络逐渐趋于融合,即产生了LBSN。基于位置的LBSN是位置和社交的结合体,它支持用户随时随地在社交平台记录并分享自己的地理信息,它是以通信网络为媒介,以智能终端为主要载体的新型平台。在LBSN中,大量用户通过签到向朋友分享位置信息或地理标签。位置社交网络让基于位置的社交成为一种新的社交模式,使得线上社交和线下社交得到有机的结合,极大的改变了人们的生活方式。社交网络催生了许多基于位置的服务,为了提供更好的服务,预测用户最有可能的下一个位置是非常重要的。如通过预测用户下一个位置,商家可以更加有效的投放定向广告。现有预测方法有基于GPS轨迹历史数据的位置预测,有基于社交网络签到数据的位置预测。社交网络签到数据和GPS轨迹历史数据有着明显的区别。社交网络签到历史数据较稀疏,位置预测范围较大。相比于社交网络签到数据,连续记录的GPS数据之间间隔5-10米。但是GPS数据仅仅包括经度、纬度和时间戳信息,没有包括语义信息,无法根据社交关系进行位置预测。现有的基于社交网络的位置预测主要有运动轨迹的预测和下一地点的预测。运动轨迹的预测相对复杂,开销较大,对周期性轨迹预测表现良好,但是对周期性不明显的轨迹预测精度较差。现有基于社交网络下一位置预测假设下一位置用户曾经访问过,下一位置仅仅从个人历史位置中选择,容易造成“冷启动”,导致常规位置预测良好,非常规位置预测精度较低。

[0003] 综上所述,现有技术存在的问题是:现有基于GPS历史数据位置预测不包含语义信息,无法根据社交关系进行位置预测。现有的基于社交网络轨迹相似性位置预测存在运动轨迹预测相对复杂,开销较大,对周期性不明显的轨迹预测精度较差,容易造成“冷启动”。

发明内容

[0004] 针对现有技术存在的问题,本发明提供了一种基于社交网络的位置预测系统及方法。

[0005] 本发明是这样实现的,一种基于社交网络的位置预测方法,所述基于社交网络的位置预测方法包括以下步骤:

[0006] 步骤一,爬取社交网络签到数据;

[0007] 步骤二,对爬取的社交网络签到数据进行预处理,过滤掉签到次数小于平均签到次数的数据,清洗掉无效的数据,利用核平滑插值技术对签到数据的稀疏性进行处理;在 $f(x)$ 中,若使用邻域样本的均值进行插值,则使 $f(x)$ 不平滑,所以使用一个核函数对估计值

平滑;具体使用核加权平均,公式为:

$$[0008] \quad f(x_0) = \frac{\sum_{i=1}^M K(x_0, x_i) y_i}{\sum_{i=1}^M K(x_0, x_i)},$$

其中 $K(\cdot)$ 采用高斯核函数,可见,离 x_0 越近的影响力越大,对

应输出的权越大,符合签到数据的实际模拟;

[0009] 步骤三,结合常规位置预测的输出概率 $P_r(1oc)$ 和非常规位置预测的输出概率 $P_u(1oc)$,预测下一位置是否为常规位置;

[0010] 步骤四,通过常规位置预测模块,得到top-m个位置列表;通过提取分析数据采集模块中采集的非结构化信息,应用于top-m位置列表,提高位置预测精度,得到top-k个位置列表, $k \leq m$ 。

[0011] 进一步,所述预测下一位置是否为常规位置公式为:

$$[0012] \quad P(1oc) = \lambda P_r(1oc) + (1-\lambda) P_u(1oc)。$$

[0013] 其中 $P_r(1oc)$ 为常规位置预测概率, $P_u(1oc)$ 为非常规位置预测概率, λ 为调节参数, $\lambda \in \{0, 1\}$ 。

[0014] 进一步,所述常规位置预测采用MHMM算法,HMM结合时间特征和空间特征对位置进行预测。不考虑时间和空间的影响,给定相同的观测序列,HMM总是得到相同的预测结果;考虑到社交用户的签到行为受到时间和空间的影响,选用混合HMM算法对下一位置进行预测。

[0015] $L_{t+1} = \arg \max P(C_{t+1} | S_t, \bar{x})$,其中 C_{t+1} 为t+1时刻的位置类别, S_t 为t时刻的观察序列状态, \bar{x} 为时间空间向量。

[0016] 进一步,所述非常规位置预测结合构建知识图谱,挖掘社交关系,采用融合社交关系的马尔科夫模型结合位置推荐系统对非常规位置进行预测。利用签到数据集构建知识图谱,在知识图谱上进行推理,挖掘相似用户,基于历史签到数据并融合相似用户训练一个马尔科夫模型对下一位置进行预测。最后将马尔科夫模型和位置推荐系统结合在一起,提高位置预测精度。

[0017] 进一步,利用签到数据集作为数据来源,构建社交知识图谱,在知识图谱上进行推理。推理方法有三类:Embedding-based技术,Pathranking algorithms,和Probabilistic graphical models概率模型。社交关系的推理采用Embedding-based技术。Embedding-based技术是以隐式因子模型为基本思想的方法,它是基于低维向量的表示方法,将知识图谱中的实体和关系在低维的向量空间里进行表达,然后进行推理。首先将实体和关系进行向量表示;其次,定义打分函数来衡量关系成立的可能性。再者,参数估计,根据打分函数推理相似用户。相似用户推理综合打分函数结构化信息和非结构化分析模块提取的兴趣相似度,表示如下:

$$[0018] \quad \text{sim}(u, v) = \alpha s(u, v) + (1-\alpha) w(u, v);$$

[0019] 其中, α 是调节参数,取值为 $[0, 1]$,反应结构化信息和非结构化信息相似度所占比重, $s(u, v)$ 代表结构化信息相似度, $w(u, v)$ 代表非结构信息相似度。非结构化信息相似度是非结构信息分析模块提取兴趣关键字,然后根据余弦相似度计算用户之间的相似度。计算如下:

$$[0020] \quad w(u, v) = \frac{u_i \cdot u_j}{|u_i| |u_j|};$$

[0021] 其中 u_i, u_j 表示用户 i 和用户 j 兴趣关键字向量表示。余弦值取值范围为 $[0, 1]$, 0表示完全不同, 1表示完全相同。

[0022] 在历史签到数据的基础上, 融合相似用户的影响, 训练马尔科夫模型对位置进行预测, 即 $L_m = \max P(A_m | H, \text{sim}(u, v))$, 其中 A_m 代表马尔科夫算法, H 代表历史签到数据, $\text{sim}(u, v)$ 代表用户相似度。

[0023] 最后, 综合融合相似用户的马尔科夫模型和位置推荐系统, 提高位置预测精度。其公式如下:

$$[0024] \quad L_{\text{top-n}} = \beta L_m + (1 - \beta) S_r;$$

[0025] 其中 L_m 代表融合相似用户的马尔科夫模型, S_r 代表位置推荐系统, β 为调节权重, 取0.6。综合融合相似用户的马尔科夫模型和位置推荐系统, 得到top-n位置列表。

[0026] 本发明的另一目的在于提供一种所述基于社交网络的位置预测方法的基于社交网络的位置预测系统包括:

[0027] 数据采集模块, 应用爬虫系统, 爬取社交网络签到数据;

[0028] 数据预处理模块, 对爬取的社交网络签到数据进行预处理, 清洗掉无效的数据, 利用核平滑插值技术对签到数据的稀疏性进行处理;

[0029] 判断模块, 结合常规位置预测的输出概率 $\text{Pr}(\text{loc})$ 和非常规位置预测的输出概率 $\text{Pu}(\text{loc})$, 预测下一位置是否为常规位置;

[0030] 常规位置预测模块, 用于对预测位置进行分类, 先预测位置的类别, 预测位置;

[0031] 非结构化数据分析模块, 通过常规位置预测模块, 得到top-m个位置列表; 通过提取分析数据采集模块中采集的非结构化信息, 应用于top-m位置列表, 提高位置预测精度, 得到top-k个位置列表;

[0032] 非常规位置预测模块, 结合构建知识图谱, 挖掘相似用户, 采用融合相似用户的马尔科夫模型结合位置推荐系统对非常规位置进行预测。

[0033] 本发明的另一目的在于提供一种应用所述基于社交网络的位置预测方法的社交网络终端。

[0034] 本发明的优点及积极效果为: 基于社交网络签到数据, 结合模糊聚类、知识图谱和位置推荐对下一地点进行预测, 既适用于常规模式位置预测, 又适用于非常规模式位置预测。本发明结合时间和位置因素对位置预测的影响, 充分挖掘社交网络签到信息中的语义信息, 解决了常规位置预测问题, 同时通过对非常规位置预测解决了位置预测中的“冷启动”问题。本发明运用知识图谱, 挖掘隐含相似用户; 运用概率论知识预测下一位置为常规位置或非常规位置。对于非常规位置, 结合位置推荐系统, 提高非常规位置预测精度。

附图说明

[0035] 图1是本发明实施例提供的基于社交网络的位置预测方法流程图。

[0036] 图2是本发明实施例提供的基于社交网络的位置预测方法实现流程图。

[0037] 图3是本发明实施例提供的基于社交网络的位置预测系统结构示意图。

具体实施方式

[0038] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合实施例,对本发明进行进一步详细说明。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。

[0039] 下面结合附图对本发明的应用原理作详细的描述。

[0040] 如图1所示,本发明实施例提供的基于社交网络的位置预测方法包括以下步骤:

[0041] S101:爬取社交网络签到数据;

[0042] S102:对爬取的社交网络签到数据进行预处理,清洗掉无效的数据,利用核平滑插值技术对签到数据的稀疏性进行处理;

[0043] S103:结合常规位置预测的输出概率 $P_r(\text{loc})$ 和非常规位置预测的输出概率 $P_u(\text{loc})$,预测下一位置是否为常规位置;

[0044] S104:通过常规位置预测模块,得到top-m个位置列表;通过提取分析数据采集模块中采集的非结构化信息,应用于top-m位置列表,提高位置预测精度,得到top-k个位置列表($k \leq m$)。

[0045] 常规位置预测采用MHMM算法,HMM结合时间特征和空间特征对位置进行预测。不考虑时间和空间的影响,给定相同的观测序列,HMM总是得到相同的预测结果;考虑到社交用户的签到行为受到时间和空间的影响,选用混合HMM算法对下一位置进行预测。

[0046] $L_{t+1} = \arg \max P(C_{t+1} | S_t, \bar{x})$,其中 C_{t+1} 为t+1时刻的位置类别, S_t 为t时刻的观察序列状态, \bar{x} 为时间空间向量。

[0047] 所述非常规位置预测结合构建知识图谱,挖掘社交关系,采用融合社交关系的马尔科夫模型结合位置推荐系统对非常规位置进行预测。利用签到数据集构建知识图谱,在知识图谱上进行推理,挖掘相似用户,基于历史签到数据并融合相似用户训练一个马尔科夫模型对下一位置进行预测。最后将马尔科夫模型和位置推荐系统结合在一起,提高位置预测精度。

[0048] 利用签到数据集作为数据来源,构建社交知识图谱,在知识图谱上进行推理。推理方法有三类:Embedding-based技术,Path ranking algorithms,和Probabilistic graphical models概率模型。社交关系的推理采用Embedding-based技术。Embedding-based技术是以隐式因子模型为基本思想的方法,它是基于低维向量的表示方法,将知识图谱中的实体和关系在低维的向量空间里进行表达,然后进行推理。首先将实体和关系进行向量表示;其次,定义打分函数来衡量关系成立的可能性。再者,参数估计,根据打分函数推理相似用户。相似用户推理综合打分函数结构化信息和非结构化分析模块提取的兴趣相似度,表示如下:

[0049] $\text{sim}(u, v) = \alpha s(u, v) + (1 - \alpha) w(u, v)$;

[0050] 其中, α 是调节参数,取值为[0, 1],反应结构化信息和非结构化信息相似度所占比重, $s(u, v)$ 代表结构化信息相似度, $w(u, v)$ 代表非结构信息相似度。非结构化信息相似度是非结构信息分析模块提取兴趣关键字,然后根据余弦相似度计算用户之间的相似度。计算如下:

$$[0051] \quad w(u, v) = \frac{u_i \cdot u_j}{|u_i| |u_j|};$$

[0052] 其中 u_i, u_j 表示用户 i 和用户 j 兴趣关键字向量表示。余弦值取值范围为 $[0, 1]$, 0表示完全不同, 1表示完全相同。

[0053] 在历史签到数据的基础上, 融合相似用户的影响, 训练马尔科夫模型对位置进行预测, 即 $L_m = \max P(A_m | H, \text{sim}(u, v))$, 其中 A_m 代表马尔科夫算法, H 代表历史签到数据, $\text{sim}(u, v)$ 代表用户相似度。

[0054] 最后, 综合融合相似用户的马尔科夫模型和位置推荐系统, 提高位置预测精度。其公式如下:

$$[0055] \quad L_{\text{top-n}} = \beta L_m + (1-\beta) S_r;$$

[0056] 其中 L_m 代表融合相似用户的马尔科夫模型, S_r 代表位置推荐系统, β 为调节权重, 取0.6。综合融合相似用户的马尔科夫模型和位置推荐系统, 得到top-n位置列表。

[0057] 图2是本发明实施例提供的基于社交网络的位置预测方法实现流程图。

[0058] 如图3所示, 本发明实施例提供的基于社交网络的位置预测系统包括:

[0059] 数据采集模块, 应用爬虫系统, 爬取社交网络签到数据。

[0060] 数据预处理模块, 对爬取的社交网络签到数据进行预处理, 清洗掉无效的数据, 然后利用核平滑插值技术对签到数据的稀疏性进行处理。

[0061] 判断模块, 结合常规位置预测的输出概率 $P_r(\text{loc})$ 和非常规位置预测的输出概率 $P_u(\text{loc})$, 预测下一位置是否为常规位置:

$$[0062] \quad P(\text{loc}) = \lambda P_r(\text{loc}) + (1-\lambda) P_u(\text{loc}), \lambda \in \{0, 1\}。$$

[0063] 常规位置预测模块, 常规位置即频繁模式, 周期性模式; 如8点半上班, 中午12点在公司附近午餐, 下午6点半下班回家, 回家后在家看电视不外出。常规位置预测精度受到时间因素、地理因素和历史数据的影响。采用GHMM算法, HMM结合时间特征和空间特征对位置进行预测。为了克服预测范围大的困难, 首先对预测位置进行分类, 先预测位置的类别, 然后进一步预测位置。

[0064] 非结构化数据分析模块, 通过常规位置预测模块, 得到top-m个位置列表。非结构化预测模块通过提取分析数据采集模块中采集的非结构化信息, 应用于top-m位置列表, 提高位置预测精度, 得到top-k个位置列表($k \leq m$)。

[0065] 非常规位置预测模块, 位置预测不仅涉及常规位置, 由于人们的新奇特性, 随众模式, 在移动模式上会表现出对非常规位置的探索。如周六去看电影, 周天去购物等。非常规位置预测模块结合非结构化分析模块, 同时构建知识图谱, 挖掘相似用户, 采用融合相似用户的马尔科夫模型结合位置推荐系统对非常规位置进行预测。

[0066] 以上所述仅为本发明的较佳实施例而已, 并不用以限制本发明, 凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等, 均应包含在本发明的保护范围之内。

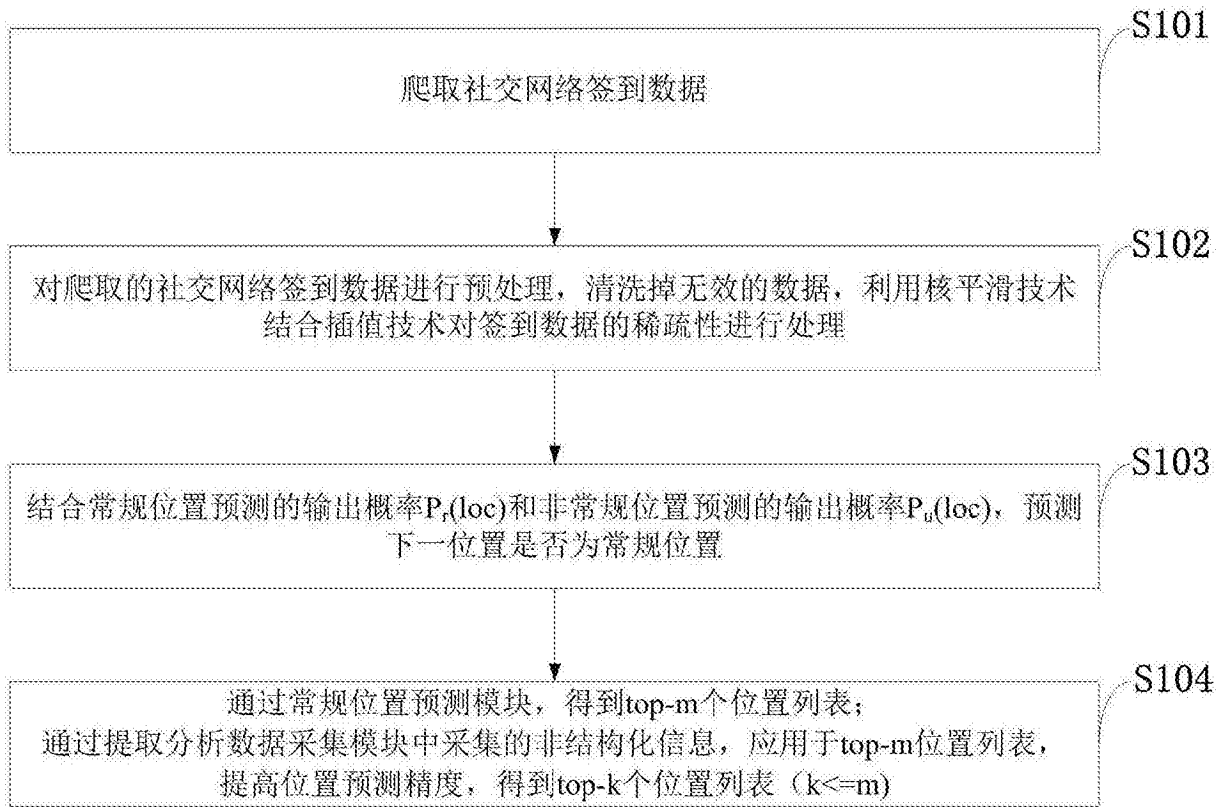


图1

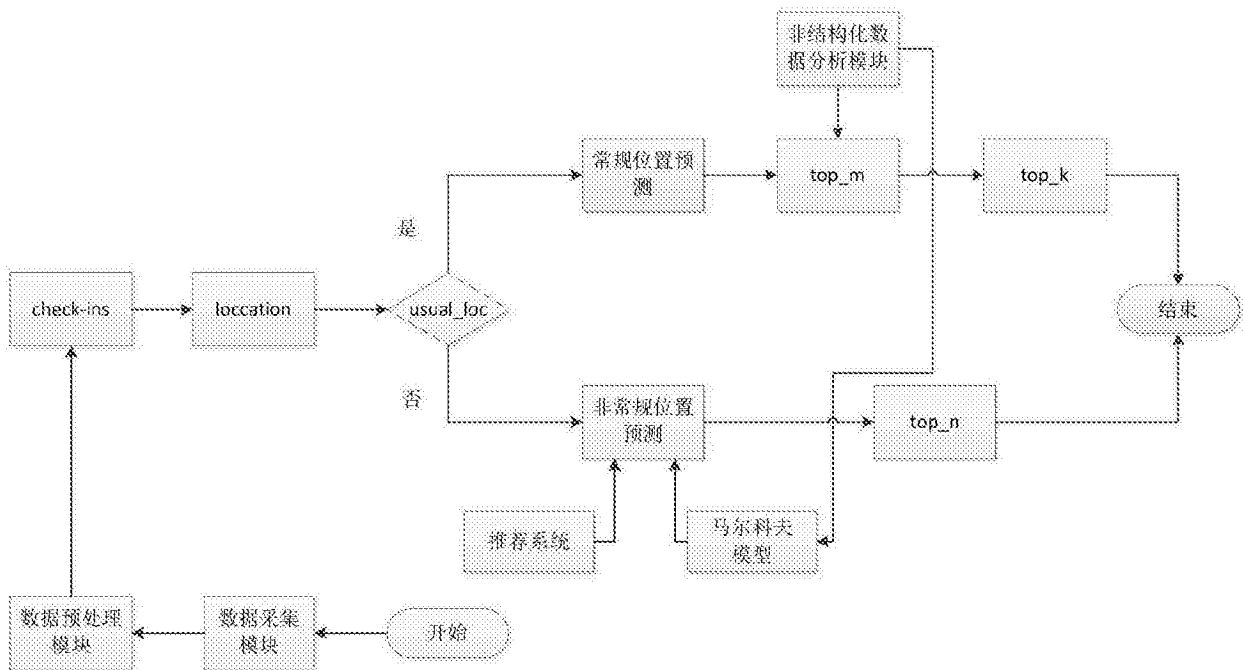


图2

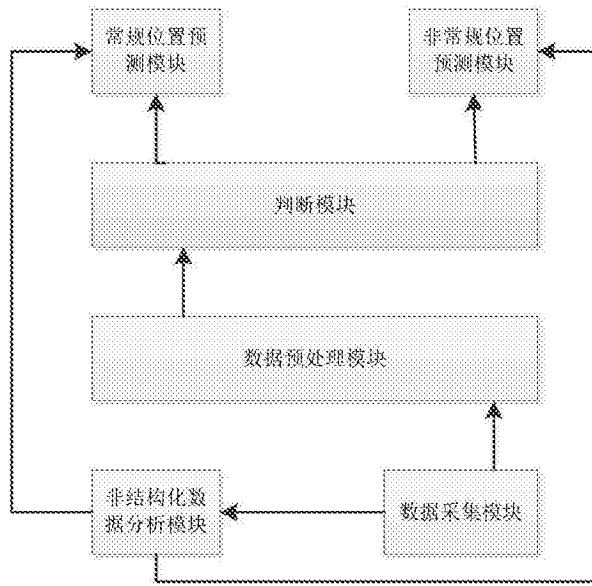


图3