

# 基于用户轨迹挖掘的智能位置服务

郑宇 谢幸  
微软亚洲研究院

关键词：基于位置的服务 GPS轨迹 轨迹数据挖掘

## 用户轨迹数据的产生

随着各种定位技术（如全球定位系统（Global Positioning System, GPS）和无线蜂窝网）的发展和普及，用户可以方便地获取个人位置信息和使用各种基于位置的服务，并将自己的移动过程以轨迹的形式记录下来。用户记录这些轨迹的目的有很多，例如：

1. 为了保证车辆的安全和有效调度以及对交通流量的分析，很多出租车、邮政车和货运车辆都配备了GPS设备。这些设备以一定的频率向某些特定的管控中心定时发送自己的坐标，将这些点按时间顺序连接起来就可得到车辆的运行轨迹。

2. 爱好自行车、远足和登山等户外运动的人通常利用GPS设备记录下自己的行程<sup>[1-2]</sup>。一方面，这些记录能保障他们沿原路返回，避免迷路；另一方面，运动结束后，爱好者也可从这些轨迹中提炼出丰富的信息，如本次运动的总里程数、海拔高度变化、速度和加速度变化以及爬坡难度等。这些信息有助于提高他们的运动水平，也可成为他们日后征服某些挑战的见证，同时还可为其他户外运动爱好者提供路线的选择。

3. 通过将轨迹与旅行过程中拍摄的照片自动关联并在电子地图上显示，用户不仅可以很直观地再现自己的行程，回忆起所见所闻，也可方便地与朋友分享旅行经历<sup>[3-6]</sup>。与枯燥而抽象的文字描述相比，这些关联了照片的线路可以在短时间内向他人呈现更直观、更生动的信息（如关键的转向、上车及换乘地点、景点的游览顺序和停留时间、风景的

特色和美丽程度等）。当他人希望去跟随和重复这次旅行时，这些轨迹又可成为最有效的导游，节约用户规划行程的时间，并在一些关键点上向给用户给出及时、有效的提示。

## 轨迹和基于位置的服务

轨迹本身的价值及其产生的便捷性催生了一批基于轨迹的位置服务，如前面提到的路线分享<sup>[1-6]</sup>以及车辆调度和安保等。然而单纯的可视化和轨迹交换并未充分发掘出轨迹中隐藏的知识。轨迹实际上记录了用户在真实世界的活动，而这些活动将在一定程度上体现个人的意图、喜好和行为模式。例如某个用户的轨迹经常出现在运动场馆，表明该用户可能会喜欢体育活动；经常穿越湖光山色的路线表征用户对户外活动的喜好。此外，对轨迹还可以有更细粒度的分析，例如可以根据用户经常光顾的餐馆类别（如川菜馆、湘菜馆）来判别出用户的口味。因此，如何挖掘轨迹中蕴含的知识就变得尤其重要。

这些知识可以从个人数据中挖掘出的用户行为、意图、经验和生活模式<sup>[7-12]</sup>；也可以是集合多人数据来发现热点地区和经典线路<sup>[13]</sup>，甚至理解人和人之间的相关性<sup>[14-15]</sup>，以及人在地域之间的活动模式<sup>[16-17]</sup>。如果了解了个人的生活规律，基于位置的服务就可以在适当的时候给予用户有效的推荐。例如，若服务提供商分析出用户通常在周末的下午要去超市买东西，就可以在合适的时间智能地将超市的折扣信息发送到该用户的移动终端；如果掌

握了一个城市的热点地区和经典旅行线路，服务提供商就可根据游客当前的位置来推荐周边的景点，并在游客访问完某个景点后智能地推荐下一个景点和具体的游玩线路。在“云”计算被认为是大势所趋的今天，只有将所有的个人轨迹汇集和整合到“云”端，对这些轨迹进行深度的分析和挖掘，完成“数据→信息→知识→智能”的过程，并最终用这些知识和智能来更好地为“端”服务，才能为用户提供更深层次、更人性化、更有效的位置服务<sup>[18]</sup>。

## 理解个人历史轨迹

### 地图匹配

在通常情况下，民用GPS使用CA码（GPS卫星发出的一种伪随机码，用于粗测距和捕获GPS卫星）传输信号，其所能达到的定位精度一般在10~20米之间。在某些道路稠密的地段，这个误差将使用户的当前位置映射到错误的道路上，从而给用户的定位和导航带来很大的麻烦。因此，国际上开展了一系列有关地图匹配的研究工作，主要研究如何将带有位置偏差的GPS轨迹映射到正确的道路上等问题。由于其功能的重要性和基础性，地图匹配也是很多基于轨迹的位置服务中必不可少的模块。

对于高采样率的轨迹，即GPS设备以高频率（如每2秒1个点）输出定位坐标，目前已有成熟的技术和产品用于解决该问题。人们日常生活中所用的各种GPS导航软件都采用了相应的地图匹配技术。但随着GPS移动终端数目的不断增加，实时获取GPS设备位置信息的通信和存储代价不断加大，导致对能源的消耗也不断增加。于是，在很多应用里，人们并不希望GPS设备以很高的频率输出定位坐标。例如，北京拥有约7万辆出租车，为了管理和调度这些车辆，每辆车都配备了GPS装置。如果每辆车每秒钟都向交通管理部门发送自己的坐标，一天的产生的数据量就会达到1TB。这将带来沉重的存储和通信负担，并造成巨大的能源浪费。同时，对于车辆的调度和管理，这样高的采样频率也

没有必要。因此，在实际情况中，一辆车基本上每分钟或每几分钟才向交通管理部门发送1个坐标。此类应用产生了大量的低采样率GPS轨迹，也为以前的地图匹配算法带来了新的挑战。

文献[8]针对低采样率的GPS轨迹，提出了一种基于全局信息的匹配方法。其主要思想包括（结合图1给出的事例具体说明，其中红点为记录的GPS点，而蓝色虚线为实际走过的道路）：

**前后位置信息** GPS采样点 $c$ 被映射到哪条道路上不仅取决于点 $c$ 到周边道路的距离，也取决于它的相邻GPS采样点 $a$ 、 $b$ 、 $d$ 和 $e$ 的映射情况。如果可以确定 $a$ 、 $b$ 、 $d$ 和 $e$ 属于水平路段，则点 $c$ 也很可能属于水平路段，即使点 $c$ 距离垂直路段更近。

**相互影响** 当处于映射点 $c$ 时，前后临近的点 $a$ 、 $b$ 、 $d$ 和 $e$ 将作为参考对象来辅助点 $c$ 的映射。而反过来，在映射点 $d$ 时，点 $c$ 将作为 $d$ 的参考点。因此，一个点既会受周边点的影响，也会影响周边的点。

**影响随距离衰减** 当处于映射点 $c$ 时，相对于点 $f$ 而言， $b$ 和 $d$ 距离 $c$ 更近，所以它们对点 $c$ 的影响应大于点 $f$ ，即点 $f$ 与 $c$ 将被映射到哪条道路上没有太大的关系。

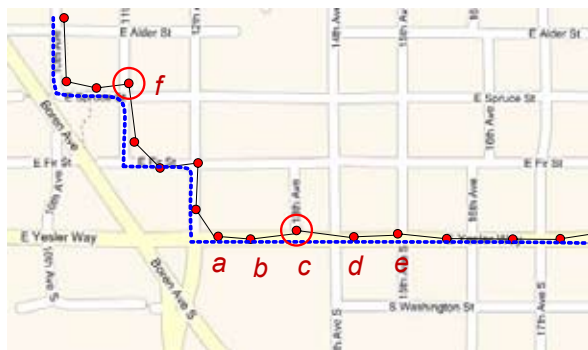


图1 GPS轨迹的地图匹配样例

### 交通方式判别

用户在记录自己的行程时可能采用不同的交通方式，如步行或开车；在同一次行程中也会变换自己的交通方式，如先坐车后步行。如果能从用户的GPS轨迹中自动学到线路产生时用户选用的交通模式，那对线路的分类和推荐将有很大的帮助。比

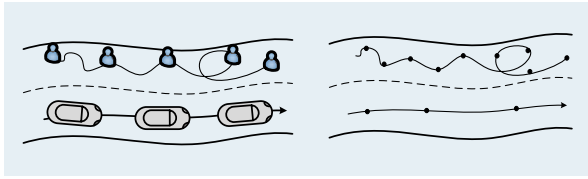


图2 驾车和步行的方向改变情况

如，帮助寻找自驾游用户保留开车的线路，而过滤掉那些乘坐公交、骑车和步行线路。

基于GPS轨迹的交通工具判别的主要挑战来自于两个方面。其一，交通工具的速度通常会受到交通状况的影响。比如，堵车的时候，汽车的运行速度可能跟正常路段的骑车速度相差无几。因此，单单基于速度的判别方法的识别精度小于50%；其二，用户在一次行程中通常会变换交通方式，如从开车到步行。这使得同一条线路可能会包含多种交通方式。如不能进行有效的分割，判断的结果也会包含不可避免的错误。

这项研究工作的主要创新点来源于以下三个方面<sup>[10-12]</sup>：

1. 提出了一种有效的路线分割方法。其主要的思想是利用步行路段来分割轨迹。在典型的场景下，人们在变换一种交通方式之前必须采用步行，即使步行的长度很短。例如，从公交车下车到乘坐另一辆出租车的过程中，用户必须通过走路来过渡。

2. 发现了一些受交通状况影响不大的特征，如方向改变率等，并结合监督学习的方法来训练一个分类模型。通常情况下，人在驾车时不可能像步行一样随意而灵活地改变自己的前进方向；而人在有意识地走直线时，产生的线路仍免不了要出现蜿蜒（如图2所示）。

3. 提出了一种后处理方法。该方法可从大量的线路中挖掘出一个隐含的地图，并分析不同路段上各种交通工具的使用概率和交通工具之间的转移概率，从而利用自然常识、地理限制和地图信息来修正错误的判别。

## 频繁模式挖掘<sup>[9]</sup>

用户历史轨迹中出现的频繁模式反映了个人的生活习惯和行为规律。如果服务提供商能够从轨迹中有效地理解这些知识，将可以为用户提供更深入、更个性化的位置服务。而要从轨迹中挖掘这些频繁模式，首先要面临的困难就是如何对个人的历史轨迹建模。

一条GPS轨迹通常由一系列带有时间戳的坐标点组成（如图3所示）。每个坐标点包含了经度、纬度和海拔高度等基本信息。一个人在一段时间内的活动可记录为一条连续的轨迹。在这条轨迹中，我们可以通过算法检测出一些用户停留过的地方。这个停留点并不是指速度为零的点，而是由一组实际的GPS点构成，如图3中 $p_3$ 、 $p_4$ 、 $p_5$ 和 $p_6$ 构成了一个停留点 $S$ 。它表示用户在某个区域内滞留的时间超过了一定的时间范围。与其他GPS点相比，这些停留点含有更重要的语义信息，如用户去过的餐馆和电影院等。基于这些停留点，一个用户的历史轨迹就可以表达为一个停留点序列，

如 $S_1 \xrightarrow{\Delta t_1} S_2 \xrightarrow{\Delta t_2} \dots \xrightarrow{\Delta t_{n-1}} S_n$ 。这个序列抓住了用户行为的重点，同时也大大减少了数据处理量。

虽然用户多次访问同一地点，但所产生的停留点可能不会完全一致（坐标会有偏差），因此直接对停留点进行比较并不可行。因此，我们需要对从轨迹中提取出来的停留点进行聚类使得相近的停留点就会被分配到同一个聚类中；再用各个停留点所归属的聚类来替换这个停留点，将停留点序列进一步转化为聚类的序列。这样，用户在不同时间段的

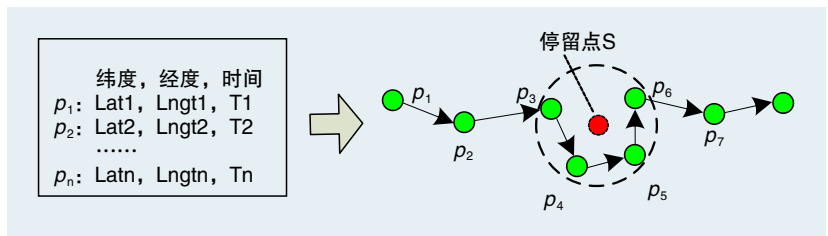


图3 一条GPS轨迹样例

历史轨迹就可以比较了。

有了用户历史轨迹的模型，我们可以用多种算法（如FP-growth、Closet+等）来挖掘这个数据中的频繁项集，比如用户A经常在周末早上去中关村或经常在周五晚上去超市等。而且这些频繁模式可以相互组合和连接，从而发现一些表征了用户生活、行为规律的顺序模式（Sequential Pattern）。例如，通常用户A在周末早上会去中关村看电影，然后下午去西单买东西。当然，这些学习到的模式会受到隐私保护，并只为用户个人所用。

## 基于多人轨迹数据的大众旅游推荐<sup>[13~16]</sup>

当外地的游客来到陌生的城市，他通常想知道这个城市中最有意思的地方以及最佳的游玩线路。而导游手册的旅行线路包含的景点和餐馆无论春夏秋冬都是千篇一律，缺乏实效性和对一个城市本地化的了解。如果汇集大量用户的轨迹数据，可从中自动发现某个时间段里一个城市里的热点地区以及地区之间的关联随着时间的推移，这些结果也会发生变化；从中还能发现本地居民倾向的地点和线路。这些知识为用户的旅行线路规划提供了有利的帮助。

### 大众轨迹数据建模

在挖掘有意思的地点和经典旅行线路时，首要的任务就是对不同用户的轨迹数据建模。其建模过程为（如图4所示）：先从每个用户的每条线路中提取出停留点（在图中表示为绿色小点），并把它们放在一个集合中；然后，利用一种基于密度的聚类算法，对这个停留点集合进行层次化聚类，在不

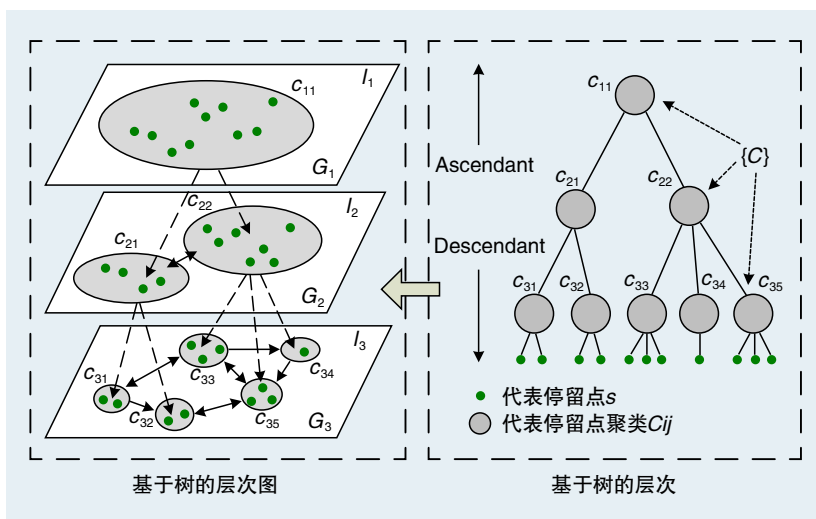


图4 基于层次图模型的多用户轨迹聚合

同的地理尺度上，将相近的停留点划分到同一个聚类（图中灰色节点，如 $C_{21}$ 等）。这样就得到一个如图4右半部分所示的一个层次树。树中的节点代表不同的停留点聚类，不同层次表示不同的地理空间尺度，层次越深，粒度越细，代表的地理空间也越小。随后，将不同用户的轨迹映射到这棵树的各个层次，就可以将不同的聚类连接起来，从而得到不同的图模型（如图4左半部分所示）。

### 发掘兴趣点和旅行专家

基于上述的层次图模型，我们可以通过以下方式来计算每个地区（聚类）的兴趣度以及每个用户的经验值（可理解为人对一个地域的了解程度和所具备地理知识）；然后，将其对地点按照兴趣度排名，对人按经验值排名；最后就可把那些具有较高排名的地区作为有趣的景点、把经验值较高的人作为本地的专家推荐给用户。

**景点的兴趣度** 一个地方的有趣程度不仅取决于访问过这个地方的人数，也取决于这些人的经验值。例如，火车站作为一个城市的门户，很可能是被游客访问次数最多的地方。如果只看访问人数，那么火车站恐怕要成为各个城市最有趣、最值得推荐的景点了。这显然不是一个游客希望得到的旅行推荐。但如果我们分析这些访问者的经验值就可以



发现大多数游客并不了解这个城市，而具备丰富本地旅行经验的专家也不会把火车站作为景点去游览。因此，只要考虑到用户的经验值，就可以避免把火车站这种热门地点作为有趣的景点推荐出来。

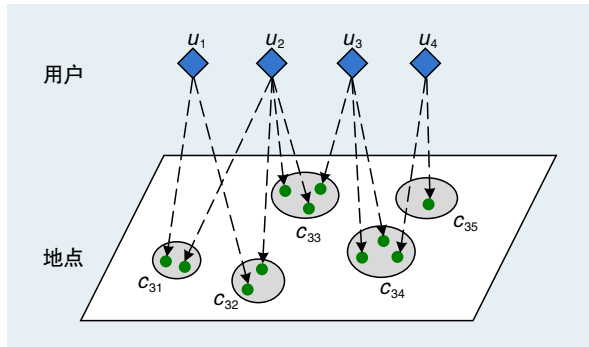


图5 人的经验值和地点的兴趣的相互依赖和关联关系

**互依赖和相关性** 人的经验值和地点的兴趣度相关联，例如，一个有经验的人可能去很多有趣的地方，而一个有趣的地方应该更能吸引很多有经验的人访问（如图5所示）。例如，一个北京本地人比外来的游客更容易找到一些高品质的餐馆和商场；而一个真正好的特色餐馆可能并不为很多游客所知，却能引来一些对北京食文化很有研究的美食专家。也就是说，一个人的经验值可表示成他去过地方的兴趣度的和；反过来，一个地方的兴趣度可表示为所有访问过该地点的人的经验值的和（类似网页排名算法HITS<sup>[19]</sup>的思想）。因此，为地点的兴趣度和人的经验值随机赋以非全零的初值，便可通过一个简单迭代的算法计算出最终的收敛结果。

**区域性** 一个人的经验值跟地域有关。例如，生活在美国西雅图的人可能非常了解西雅图，但对北京一无所知。同理，一个北京的旅行专家，可能对上海并不了解。对于一些大城市，人们对城市中不同区域的熟悉程度也不一样。因此，以上描述的经验值和兴趣度计算方法要以不同的区域为条件，即一个人将具备多个与区域相关的经验值。在不同的区域中，选用相应的经验值按人进行专家排名。

### 检测经典旅行线路

基于上面得到的地点兴趣度和人的经验值，我

们认为一个线路（地点间的访问顺序）的经典程度由以下三个因素决定（与传统意义上的“经典”有些不同，这里的“经典”不是指经过的人数最多、历史最悠久，而主要是指更有效、更合理，并受到有经验的专家认可和推荐）：

1. 选择这个顺序的人的经验值之和，即选择这个顺序的人越有经验，这个访问顺序可能越合理、越经典；
2. 这个顺序中包含的地点的兴趣度，如果一条线路中包含的景点越有趣，它可能越值得被推荐；
3. 人们选择一个顺序的概率。

如图6所示，线路A→C的经典值

$$S_{AC} = 5 \times \left( \frac{5}{7} \times I_A + \frac{5}{8} \times I_C \right) + \sum_{u_k \in U_{AC}} e^k。$$

这里， $I_A$ 和 $I_C$ 分别表示地点A和C的兴趣度， $e^k$ 为用户 $u_k$ 的经验值。5/7表示有7个人从A点离开，其中5个选择了A→C这个线路。而5/8则表示，有8人到达C，其中有5人来自于A→C线路。因此，A点的兴趣度只有5/7可以传播到这条表上来，而C点的兴趣度也只能有5/8分配到A→C。

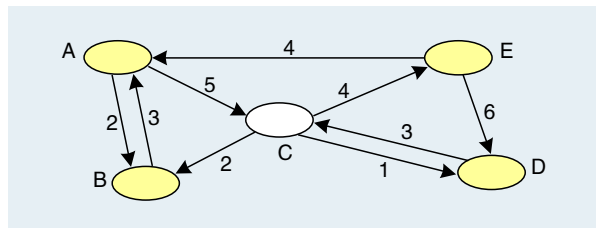


图6 计算一个顺序的经典程度

## 个性化朋友和地点推荐

### 基于历史轨迹的人的相似度计算<sup>[14]</sup>

正如前面所述，轨迹隐含了人的行为和喜好。因此，人们在地理空间移动的相似性，也在一定程度上反映了不同人之间的品味和爱好的相似性。我们首先用不同的层次图来为每个用户的历史轨迹建模，然后成对地比较图与图之间的相似性（如图7所示）。与之前提到过的大众数据建模方法一致，我们仍然利用层次化聚类的思想将所有用户的停留

点转化为一个公共的层次树（图7中间的部分），树中的各个节点（停留点聚类）表示不同尺度和粒度的地点。此后，将每个用户的线路分别导入这个公共的框架，便可得到用户各自的层次图（图7的左右两个部分分别表示用户1和2的层次图）。

在通过匹配两个层次图来计算用户相似性时，我们考虑以下两点因素：

**层次** 指两个人的相似性，可表示为两个层次图中各个对应层次上的图的相似性的加权和。这里的权重是由层次的深度来决定。由于较深的层次具有较细的空间粒度和尺度，两个用户越在较深层次上的图相似，则越表明他们的活动轨迹相似。因此，深层次的匹配结果应被赋予较大的权重。例如，两个人都在中国就不如两个人都在北京市相似。如果能发现两个人在代表学校和景点这种更细粒度的层次上有重叠，则说明这两个人相似度高。

**相似序列的长度** 指同一图层上两幅图的相似性，可表示为这两个图共享序列的相似性的和。而序列的相似性又取决于序列的长度。因此，用户共享的序列越多，序列的长度越长，则这两个图的相似性越大。例如A、B和C三个用户，A和B共同走过了一个长度为2的序列 $C_{32} \rightarrow C_{31}$ ，而A和C两个共同走过一个长度为3的序列 $C_{32} \rightarrow C_{31} \rightarrow C_{33}$ 。显然，与B相比，用户C与A更相似。

### 个性化朋友和地点推荐<sup>[15]</sup>

虽然利用大规模轨迹数据实现的大众化旅行推荐方法可找出一些公认的热门景点和经典旅行线路，但实际上不同的用户有不同的喜好，各种景点的排名在每个人的心目中也不一样。例如，喜欢自然风光的用户可能对故宫这样的历史古迹并不是特别感兴趣；喜欢美食的游客可能会更关注小吃聚集

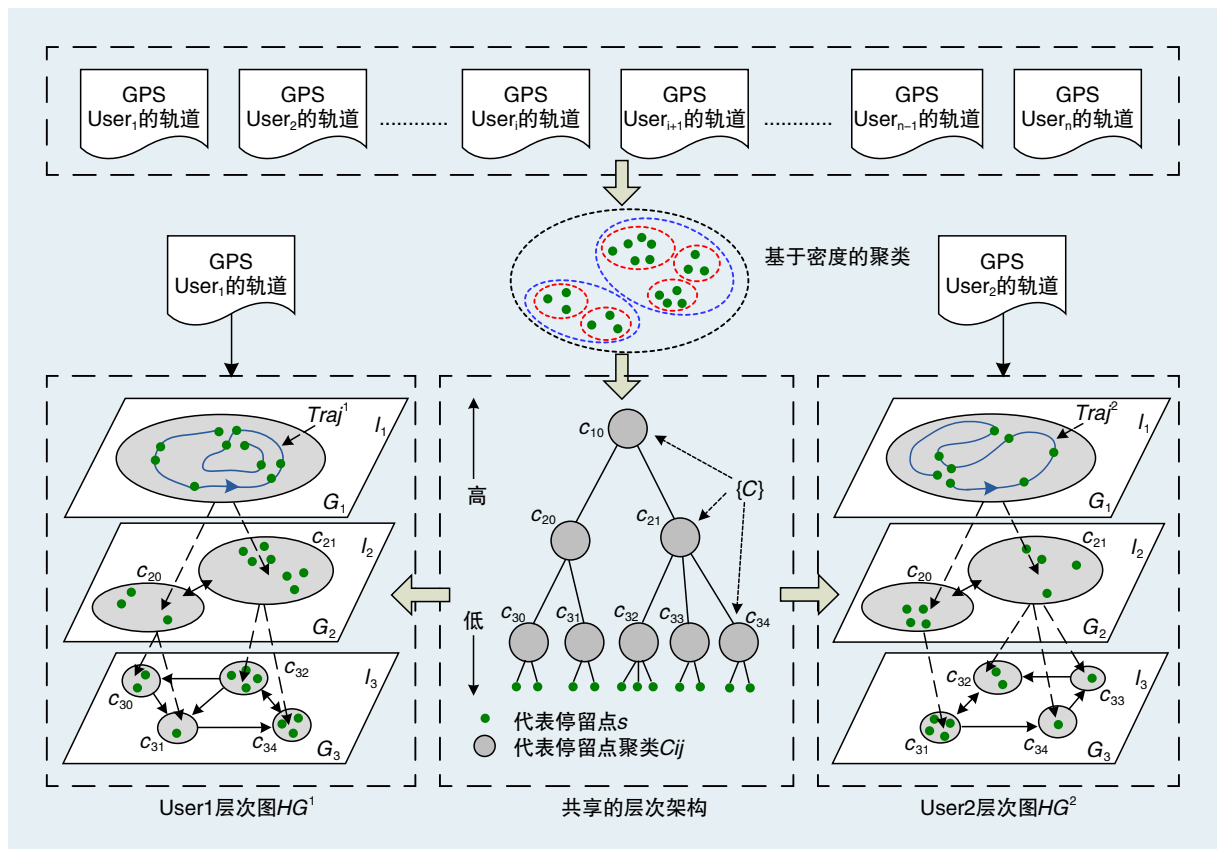


图7 利用层次图来比较用户的相似性

的街道。因此，针对个人的喜好来做个性化的推荐才是更人性化、更有效的位置服务。

我们设计的基于轨迹的个性化朋友和地点推荐包含以下三步：

1. 利用用户的历史轨迹计算出用户之间的相似性（参见上一节描述的方法），为某个用户找出最相似的 $n$ 个人作为潜在的朋友，完成个性化朋友推荐。也许他们在现实生活中多次擦肩而过，却从来没有认识的机会。因此，当在论坛中发起一些活动的时候（如自驾游和登山等），利用该方法就能帮助用户更精准地找到那些兴趣相投的人。

2. 从这些潜在朋友的历史轨迹中查找出一些用户自己没有去过的地点，并利用协同过滤的方法来估计该用户对这些地点的兴趣度。如图8所示，如果把用户和他们去过的地点用一个矩阵来表示，矩阵中的每个值表示用户曾去过这个地方的次数。那么我们就可以像亚马逊（Amazon）根据用户的买书记录来推荐图书那样，使用协同过滤来计算用户对未曾去过的地方的兴趣度。这里有个很重要的思想，即相似的人通常会做出类似的决定，所以越相似的人的经历越具有参考价值。

3. 按估算的兴趣度对用户未曾去过的地点排名，并把排名较高的 $m$ 个地点推荐给用户。作为个性化的地点推荐，这个推荐是根据用户过去的经历分析出来的。

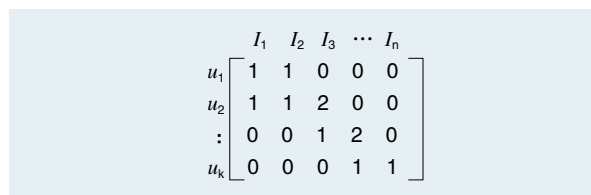


图8 用户和访问地点之间关系的矩阵表达方式

## 结语

现有的位置服务通常直接使用用户提供的位置数据，缺乏对这些数据的分析和挖掘，忽略了这些位置数据中蕴含的信息和知识。我们以在过去一段时间里完成的一些工作为例，介绍了如何利用数据

挖掘和机器学习等技术来从个人及大众用户轨迹数据中获取知识，并用这些知识来定制更加智能的基于轨迹的位置服务。

在不久的将来，尤其是“云”计算到来的时代，只有对用户的轨迹数据进行更有效的整合、更深入的分析 and 挖掘，才能获取更丰富的知识和智能，从而为用户提供更人性化、更有效的位置服务。■



郑宇

微软亚洲研究院互联网搜索与挖掘组研究员。主要研究方向为基于地理位置的搜索和服务、时空数据挖掘、地理信息系统和普适计算等。

yuzheng@microsoft.com



谢幸

CCF高级会员。微软亚洲研究院互联网搜索与挖掘组主管研究员。主要研究方向为空间数据挖掘、基于位置的服务、移动和普适计算等。

Xing.Xie@microsoft.com

## 参考文献

- [1] Mountain Bike. <http://www.mtb-routes.co.uk/northyorkmoors/default.aspx>
- [2] SportsDo. <http://sportsdo.net/Activity/ActivityBlog.aspx>
- [3] Yu Zheng, Longhao Wang, Xing Xie, Wei-Ying Ma. GeoLife-Managing and understanding your past life over maps, In Proceedings of International conference on Mobile Data Management (MDM 2008), Beijing China
- [4] Yu Zheng, Xing Xie, Wei-Ying Ma. Searching Your Life on Web Maps, SIGIR workshop on mobile information retrieval, 2008, Singapore
- [5] Yu Zheng, Yukun Chen, Xing Xie, Wei-Ying Ma. GoLife2.0: A Location-Based Social Networking Service. In proceedings of International Conference on Mobile Data Management 2009 (MDM 2009)
- [6] Yukun Chen, Kai Jiang, Yu Zheng. Trajectory Simplification Method for Location-Based Social Networking Services. In Proceedings of ACM GIS workshop on Location-based social networking services. 2009

更多参考文献请访问：[www.ccf.org.cn](http://www.ccf.org.cn)的“中国计算机学会通讯”栏目

- [7] 谢幸, 郑宇. 基于地理信息的用户行为理解, 中国计算机学会通讯, 10. 2008
- [8] Chengyang Zhang, Yu Zheng, Xing Xie. Map-Matching for Low-Sampling-Rate GPS Trajectories. In Proceedings of ACM SIGSPATIAL Conference on Geographical Information Systems (ACM GIS 2009)
- [9] Yang Ye, Yu Zheng, Yukun Chen, Xing Xie. Mining Individual Life Pattern Based on Location History. In proceedings of the International Conference on Mobile Data Management 2009 (MDM 2009)
- [10] Yu Zheng, Like Liu, Longhao Wang, Xing Xie. Learning Transportation Modes from Raw GPS Data for Geographic Application on the Web, In Proceedings of International conference on World Wild Web (WWW 2008), Beijing, China
- [11] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie. Understanding Mobility Based on GPS Data. In Proceedings of ACM conference on Ubiquitous Computing (UbiComp 2008), Seoul, Korea
- [12] Yu Zheng, Yukun Chen, Quannan Li, Xing Xie, Wei-Ying Ma. Understanding transportation modes based on GPS data for Web applications. To appear in ACM Transaction on the Web. 2009
- [13] Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of International conference on World Wild Web (WWW 2009), Madrid Spain
- [14] Quannan Li, Yu Zheng, Yukun Chen, Xing Xie. Mining user similarity based on location history. In Proceedings of ACM SIGSPATIAL conference on Geographical Information Systems (ACM GIS 2008), Irvine, CA, USA
- [15] Yu Zheng, Lizhu Zhang, Xing Xie. Recommending friends and locations based on individual location history. To appear in ACM Transaction on the Web, 2009
- [16] Yu Zheng, Lizhu Zhang, Xing Xie. Mining Correlation between Locations Using Human Location History. In Proceedings of ACM SIGSPATIAL Conference on Geographical Information Systems (ACM GIS 2009)
- [17] Vincent Wenchen Zheng, Yu Zheng and Qiang Yang. Joint Learning User's Activities and Profiles from GPS Data. In appear in Proceedings of ACM GIS workshop on Location-based social networking services. 2009
- [18] 郑宇, 谢幸, 马维英. 基于“云”计算的位置服务, 数字移动时代
- [19] Kleinberg, Jon (1999). Authoritative sources in a hyperlinked environment. Journal of the ACM 46 (5): 604 ~ 632