

社交媒体中的时空轨迹模式挖掘

彭文志

新竹交通大学

关键词：时空数据挖掘

目前已有许多便携式设备（比如手机、导航系统和GPS记录仪等）能感知位置信息。使用这些装有GPS的设备，人们可以在互联网上记录和分享他们的路线^[1,2]。因为这些技术的进步，轨迹数据的挖掘、管理及搜索吸引了相当多的研究学者。这些分享的路线也显示了使用者在不同情境下的移动经历。我们根据这些数据，针对旅游^[3,4]和导航^[5~7]等应用进行了相关的研究。

基于旅游行程行为模式的轨迹挖掘

轨迹搜索已经成为许多应用程序及网站的基本功能，比如，EveryTrail（一个分享旅程的网站）。在这种网站中，用户可以分享他们的旅程，也可以与其他旅行者联络；可以指定一个特定的城市（比如台北市）或区域范围，获取所有经过这个区域的轨迹。轨迹搜索还可采用其他形式，例如系统以轨迹样本作为查询，获取相似的轨迹；或者根据用户给定的一些感兴趣的区域（regions of interest, ROIs），找出与这些区域贴近的历史轨迹。虽然这些搜索形式是一个重要的进展，但却局限于假设用户在计划轨迹前已拥有ROIs的先期知识。否则使用者可能需要花费很多时间去尝试不同的ROIs。我们将剔除这些假设，提出一个轨迹搜索的框架，具有新颖的轨迹挖掘、排名以及搜索功能，能够支持更为高效的轨迹搜索。

近年来，虽然出现了许多行程推荐的服务，但它们往往都只是提供一个ROIs的访问序列而没有

提供详细的路线说明。此外，用户在游玩各个城市及景点时，也可能有不同的行程喜好。例如，有些人在安排行程时，希望能够游玩的景点越多越好，而花在每个景点游玩的时间较少（广度旅游）；有些人却偏好深入游玩较少的景点（深度旅游）。因此，我们的目标是开发一个能够从现有的旅游轨迹中找出经过热门ROIs的轨迹搜索服务。查询的形式如下：

使用者给定一个特定的区域范围及对深度旅游或广度旅游的偏好，系统将会从轨迹数据库中计算并返回前K条有趣的轨迹，且每条轨迹都会经过一些该区域范围内热门的ROIs。

为了获取符合要求的深度与广度旅游的轨迹，需要迎接下列挑战：

- 推断找出ROIs旅程轨迹吸引人的程度；
- 设计出深度旅行与广度旅行轨迹的评价方式；
- 开发一个能高效支持在线搜索的算法。

为此，我们开发了一个基于用户移动模式的旅游路径推荐系统PATS。PATS系统由旅程行为挖掘和轨迹搜索两部分组成。PATS返回的路径是许多其他用户深度或广度旅游的宝贵经验。

首先找出ROIs，即有一定数量轨迹经过的区域。例如图1中，假设密度门限值为2，则ROIs的集合为{R1,R2,R3,R4}。然而，即使给定了一组ROIs，用户对如何计划行程仍然毫无头绪。一个行程不仅需要一组ROIs，也需要对如何游玩这些ROIs进行规划，例如某个ROIs的行程规划可能是R2→R1→R4，但是在图1中并不存在任何一条完全包含这些ROIs的轨迹。所以，我们将从轨迹数据库

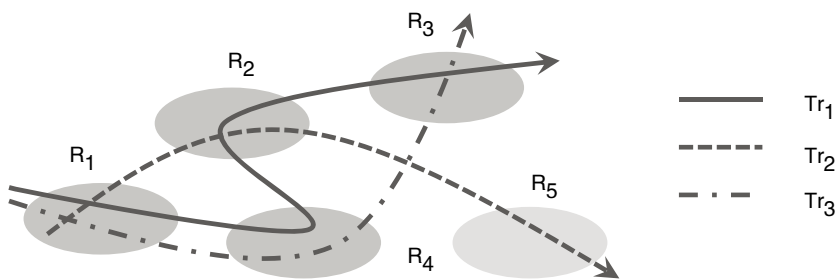


图1 旅游行为模式样例

中挖掘出隐藏在各轨迹中相关的ROIs，将其合并，并推荐有趣的行程。综合分析ROIs在旅程顺序中的关系，旅程模式不仅仅代表着旅游景点访问顺序的关系，也表示着区域间的移动概率。而且，用户还可为ROI吸引人的程度评分。ROIs的可达性越高，代表用户经过该ROI的概率越高。我们建立了一个用户的“移动图”，图中顶点代表一个ROI，边代表ROIs间可能的移动方式。基于用户的移动图，我们提出了一个使用马尔科夫过程作为吸引力评分（attractive score, AS）的算法。

图1中，因为经过了4个ROIs，轨迹1（即Tr1）

可能为广度旅行的轨迹，轨迹2（即Tr2）与轨迹3（即Tr3）则可能为深度旅行的轨迹。但对于是深度旅行还是广度旅行的判断不能光靠经过的ROIs数来确定，还必须考虑该ROI的吸引力评分。我们结合ROIs数和吸引力评分设计了两种对轨迹的评分方式：深度旅行评分（depth-trip score function, DT）和广度旅行评分（breadth-trip score function, BT）。对在线查询来说，响应时间可以说是非常重要的。因此，我们提供了一个有约束范围的轨迹搜索算法（bounded trajectory search, BTS），以

便得到在两种评分方式中分数前K高的轨迹。BST算法的主要思想是逐步减少搜索空间，而不计算用户指定范围内的所有轨迹。通过真实轨迹数据验证，结果表明，PATS不仅高效，而且能有效地找到包含热门ROIs的前K条轨迹。

如前所述，给定一个区域范围和使用者对深度、广度旅游的喜好，系统返回K条轨迹。这些轨迹都包含或经过特定的区域，而且是真实的，因此这样的行程信息是可信的。在PATS中，对轨迹有趣程度的评分是基于ROIs的数量与它们的吸引力。如何对轨迹评分是PATS系统面临的一大挑战。我们开发了一个离线旅程行为挖掘组件和一个在线轨迹搜索组件。具体介绍如下：

旅程行为挖掘组件 该组件的主要任务是找出一组ROIs并计算它们吸引力的分数。我们采用了文献[8,9]中基于密度的方法来确定ROIs。有了ROIs之后，考虑行程顺序与ROIs的关系，通过用户移动

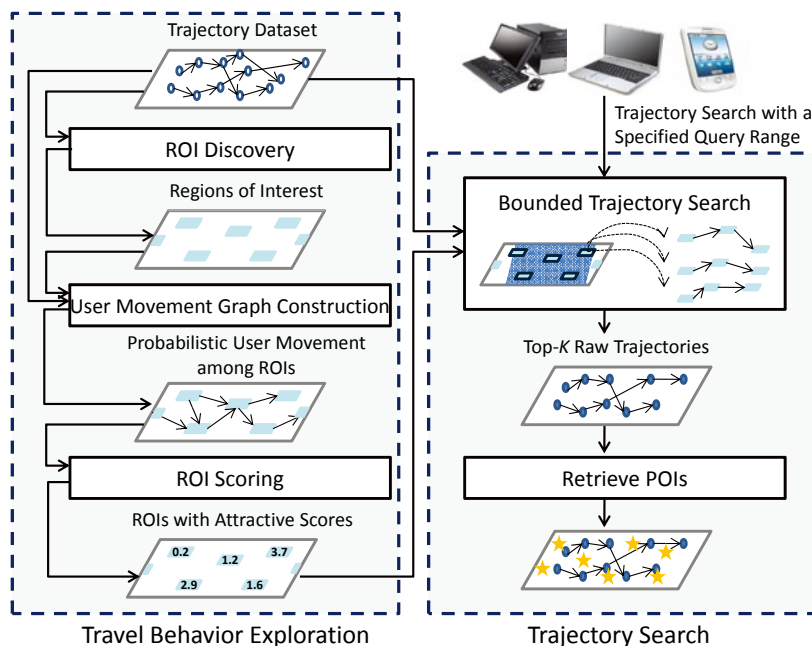
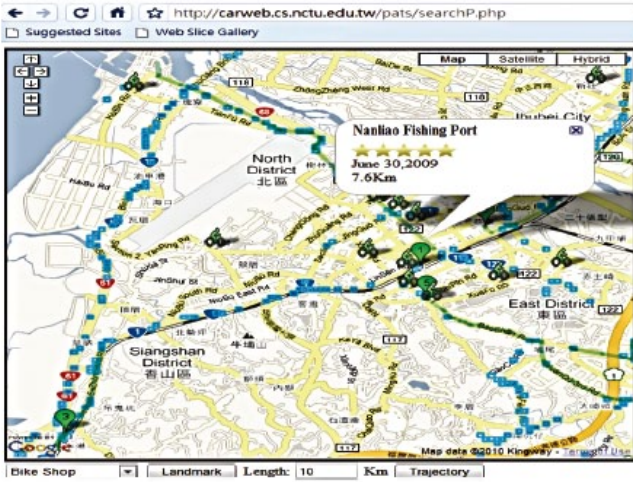


图2 PATS架构概貌



(a) 网页界面



(b) Android智能手机界面

图3 PATS操作界面

图获取行程顺序的关系。在用户移动图中，顶点代表一个ROI，边代表ROIs间可能的移动方式。基于用户的移动图，我们提出了一种是用马尔可夫过程对ROI做吸引力评分的算法。

轨迹搜索组件 利用该在线组件，用户可以通过网页界面，如图3 (a)，或安卓 (Android) 应用程序，如图3 (b)，来获取前K条有趣的轨迹。用户可以调整区域范围来查询，如图3 (a) 中黑框内的区域范围和图3 (b) 中显示的整个地图可作为PATs的搜索范围。给定查询的范围与用户深度和广度旅游的喜好后，地图上将显示出前K条轨迹与附近的景点。图3 (b) 的右图为返回的

前K条轨迹，选择其中一条轨迹后，路线就会在地图上显示出来，如图3 (b) 左图所示。

此外，图4是我们在台湾的日月潭旅游区根据不同需求查询的结果。图4 (a) 为日月潭旅游区的景点；图4 (b) 展示了根据深度旅游喜好查询的结果；图4 (c) 展示了根据广度旅行喜好查询的结果。

个性化导航系统

目前，基于位置或路线导航的软件系统越来越多，这些系统通常可以帮助用户记录所走过的路



(a) 日月潭景点



(b) DT评分最高的轨迹



(c) BT评分最高的轨迹

图4 在不同模式下推荐的游览日月潭的轨迹

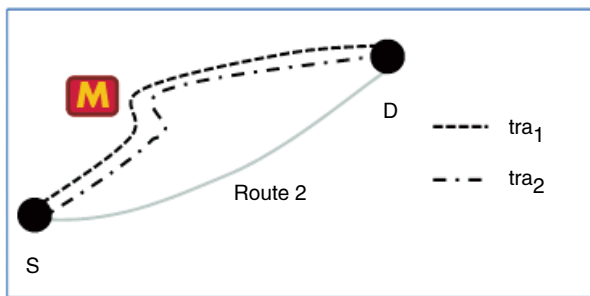


图5 个性化路径样例

线并规划导航路线。然而，随着GPS导航系统的普及，如何改善传统的导航系统，使之变得聪明和个性化成为亟待解决的问题。

何谓个性化

人们在日常生活中的活动范围看似很大，但实际观察后就会发现，每个人经常去的地方总是只有那几处。以上班族为例，他常去的地方不外乎是住所、工作单位、餐厅或商场，并且根据个人习惯偏好只会去某几个地方。如图5，用户从起点（S）前往终点（D）并不总会以最短路径行进，可能因为习惯顺路购买食物而经常走路径1（即Route1），所以即使路径2（Route2）距离比路径1短，用户也不一定会考虑。因此，导航系统的个性化是指从用户过去的移动轨迹挖掘出用户习惯行经的区域或道路，将其作为导航系统推荐

的路线。

一般的导航系统只是找出起始点到目的地的最短路径或最快路径。我们的系统与之不同，而是先根据用户平时的路径习惯，找出他所熟悉的区域和道路，进而为他推荐一条从起点到终点最为熟悉的路线。其中的关键点有：

1 利用用户的路径数据，找出他所熟悉的区域和路段。

通过利用用户的GPS历史数据，挖掘出用户的驾驶行为以及熟悉的路段，然后用这些特性来规划导航路线，以便让用户驾驶起来更为方便和安全。

2 利用用户熟悉的区域做规划，找出从起点到终点用户最熟悉的路段。

通过设计算法，利用所找出的熟悉区域来计算和找出一条较快且用户熟悉的路线。这要求系统必须大幅提升算法的性能，以使其响应时间变短。

关于系统设计，首先要寻找用户熟悉的路段。将GPS历史数据映射到地图上，并以经过的频率作为熟悉程度。如图6，根据不同的 θ 值，提取路段。此外，计算最佳的导航路径不能只依赖熟悉程度，还应该考虑起止点的距离。我们实现了一个模拟导航系统的界面，如图7所示。系统可以为使用者提供个性化路径（如图8），用户能够在界面上选取起点和终点，也可以调整参数控制熟悉程度和距离的影响程度。



(a) $\theta = 50$



(b) $\theta = 100$



(c) $\theta = 150$

图6 熟悉路段的提取

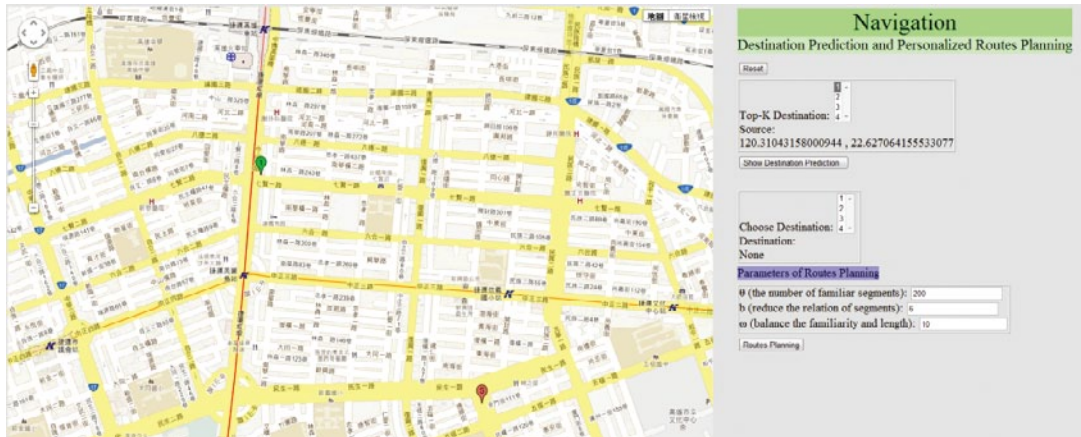


图7 操作界面 (选取起点和终点)

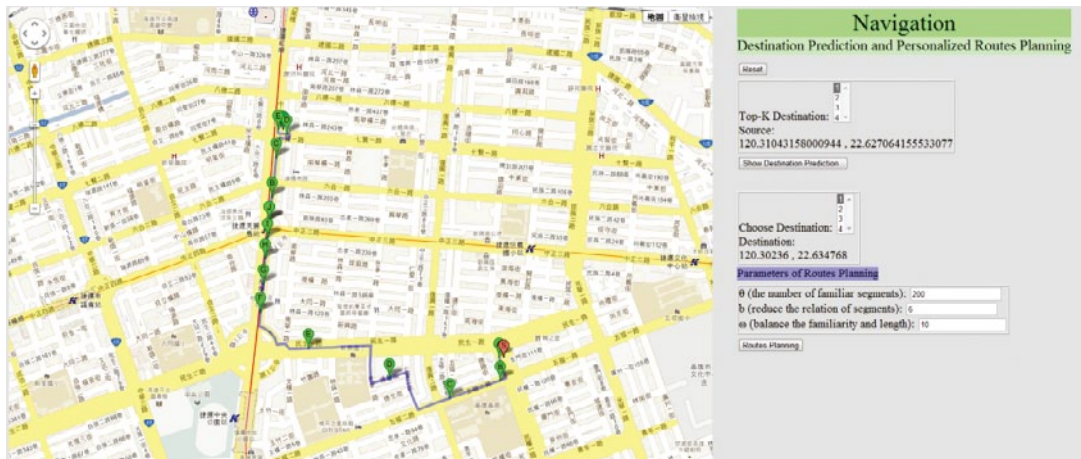


图8 显示个性化路线

结语

基于时空数据挖掘技术,通过分析用户过去的驾驶或旅游经验,我们的系统为人们出行提供了一种便利的方式。未来,我们将提供更加个性化的旅游或导航服务,并结合云计算技术,让用户可以随时随地轻松地获取并使用我们所开发的应用服务。■



彭文志

台湾新竹交通大学副教授。主要研究方向为移动数据管理、数据挖掘、数据库等。

wcpeng@cs.nctu.edu.tw

参考文献

- [1] C.-H. Lo, W.-C. Peng, C.-W. Chen, T.-Y. Lin and C.-S. Lin, CarWeb: A Traffic Data Collection Platform, Proceedings of the 9th International Conference on Mobile Data Management, Beijing, China, April 27-30, 2008
- [2] Y.-M. Chang, L.-Y. Wei, C.-S. Lin, C.-H. Jung, W.-C. Peng and I.-H. Chen, Exploring GPS Data for Traffic Status Estimation, Proceedings of the 10th International Conference on Mobile Data Management, Taipei, Taiwan, May 18-20, 2009
- [3] L.-Y. Wei, W.-C. Peng, B.-C. Chen, and T.-W. Lin, PATS: A Framework of Pattern-Aware Trajectory Search, Proceedings of the 1st Workshop on Uncertain Mobile Data Management and Mining (In conjunction with MDM), Kansas, Missouri, USA, May 23, 2010

- [4] L.-Y. Wei, W.-C. Peng, and W.-C. Lee, Exploring Pattern-Aware Travel Routes for Trajectory Search, *ACM Transactions on Intelligent Systems and Technology*
- [5] K.-P. Chang, L.-Y. Wei, M.-Y. Yeh and W.-C. Peng, Discovering Personalized Routes from Trajectories, *Proceedings of the third International Workshop on Location-based Social Networks (in conjunction with ACM SIGSPATIAL GIS 2011)*, Chicago, Illinois, USA, Nov. 1-4, 2011
- [6] T.-W. Lin, B. Zheng, L.-Y. Wei, and W.-C. Peng, Exploring Dynamic Query Shapes for Nearby Traffic Monitoring Queries in Road Networks, *Proceedings of ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, Illinois, USA, Nov. 1-4, 2011
- [7] L.-Y. Wei, W.-C. Peng, C.-S. Lin and C.-H. Jung, Exploring Spatio-Temporal Features for Traffic Estimation on Road Networks, *Proceedings of the 11th International Symposium on Spatial and Temporal Databases (SSTD 2009)*, Aalborg, Denmark, July 8-10, 2009
- [8] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, Trajectory pattern mining, *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Jose, California, USA, August 12-15, 2007
- [9] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, Mining periodic behaviors for moving objects, *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, USA, July 25-28, 2010