电 子 科 技 大 学

硕士文献阅读综述

班 学 号：201621060221

姓　　名：高崇铭

指导教师：邵俊明 教授

学科专业：计算机科学与技术

所在学院：计算机科学与工程学院

指导教师阅读后签字：

摘 要

近年来，随着移动设备飞快普及与硬件存储、计算能力的飞快提升，每天都有海量的轨迹数和带地点标签的签到数据以惊人的速度产生。高效地对这些数据进行表征和挖掘，将在现有的经济、环境、交通等领域中产生巨大的效益。因此，结合社交网络进行 POI 推荐（Point-of-Interest recommendation in Location-Based Social Networks）问题成了一个重要的问题。解决好这个推荐问题，将使得人们的生活得到极大的提升，也能为不同规模的商业经济带来巨大的效益。在本文中，我将首先对国际上主流的在 LBSNs 中的轨迹表征方法进行总结。之后再对 POI 推荐的算法做一个简短的综述，将现有的模型分为四类：（1）基于签到数据的纯 POI 推荐。（2）考虑地理因素的 POI 推荐。（3）考虑社交网络的 POI 推荐。（4）考虑时间影响的 POI 推荐。我将分别列举每类推荐算法中经典的例子，详述其工作原理，并指出这些模型共有的问题和缺陷：（a）由于涉及到时空的原始轨迹数据通常有着长度不固定、采样率不固定等天生的缺陷，虽然各种研究方法分别基于自己的假设提出了很多独立的轨迹特征提取以及预处理手段，但在富含多源语义的各种应用场景下，这些方法在表征轨迹的时候都只考虑了轨迹数据的单反面或少数方面信息，没有将多源信息有机地结合起来。（b）作为影响 POI 推荐正确率程度最大的地理因素，几乎所有方法都将 POI 的位置信息视为一个平面上的点，忽略了地点是有层次结构的，一个 POI 隶属于一个区，一个区，一个城市，不同层次的地理结构给 POI 带来的影响应该分别考虑。（c）由于作为辅助的社交网络数据、地理数据以及时间数据本身也存在稀疏、有噪声等缺陷，使得用它们来增强推荐系统的有效性和准确性都将受到限制。最后，我将对四类模型做出总结性归纳，针对以上提出的问题，指出在 LSBNs 中进行 POI 推荐的未来方向，并基于此提出我在本次毕业设计工作中解决问题的思路。

关键词：轨迹表征，轨迹挖掘，POI 推荐，LBSNs

1. 国内外研究现状

二十一世纪以来，随着带有 GPS、GSM 及 RFID 等移动设备的普及，大量具有时间与空间属性的数据被收集与存储下来，研究如何从这些海量数据中提取出固定模式，催生了大量的轨迹挖掘任务。而后，由于社交网络的兴起与多领域数据的整合，轨迹与基于地点的签到数据 (check-in data) 被关联了大量的语义信息，如社交网络，如时间关联，人们将这样语义数据称为 LBSNs（Location-Based Social Networks），而轨迹数据在 LBSNs 中的语义分析与挖掘成为了关注的热点。这其中，POI（Point-of-interest）推荐成为了一个非常热门且重要的任务。在现实中，进行 POI 推荐之前，轨迹需要良好的表征。然而，目前的 POI 推荐算法多将注意力集中到了推荐模块，而忽视了数据表征模块。事实上，数据表征的结果直接影响到推荐的有效性和准确性。

另外，由于社交网络与地点推荐在某些软件下的高度结合，为用户推荐感兴趣的地点也是最近的热点问题。与传统商品、音乐、电影推荐不一样，在地点推荐中，人们的选择不再单纯地由地点的吸引程度来决定，还将由社交网络中用户与朋友的关系和朋友的偏好来决定，此外，用户的日常活动范围也将限制用户的选择。在这样的情境下，如何结合用户的社交网络、地点的地理信息，来给用户推荐准确地进行推荐刚兴趣的地点，是地点推荐系统中最重要的研究问题。

于是，本文我将先对轨迹数据在 LBSNs 的表征做出简单总结，之后再对 LBSNs 中的 POI 推荐算法做出总结。

2. LBSNs 中轨迹语义表征方法总结

由不同设备采集到的原始轨迹数据只是一个带着时间戳的序列，其中包含的信息需要进一步筛选和提取。不同的研究工作根据后续的挖掘任务，提出了不同的轨迹特征提取方法以及语义表达模型，此处将常见的轨迹表达模型分类为三种:

(1) 基于关键点提取的轨迹表征

将轨迹数据用有限且紧凑的关键点来表达是一种最常见的方式。而关键点有两种定义方式: (a)全局关键点。这种类型的关键点又被成为 Points-of-Interest (POIs)，通常来说，这些点是提前选定的，比如地图上的学校、餐馆、加油站等标志性点或区域，也可以简单地将地图均匀地划分为小方格得到[1,2,3,4,5,6,7,8]。然而也有很多研究工作致力于从原始轨迹中提取出这样的全局关键点(区域)，比如一些工作[9,10]采用基于划分的 K-means 的思想来对轨迹集中的 GPS 点进行聚类得到全局关键点，同样地，另外一些工作则使用基于密度方法来聚类，例如 DBSCAN，OPTICS，KDE，以及重力模型[11,12,13,14,15,16]。(b)基于单条轨迹的关键点。这种方式通常是基于后续工作的需求，对轨迹的某种特征进行提取。例如文献[17,18]提取每一条轨迹的 stay points 来代表该轨迹，stay points 即轨迹进行停留了一段时间的点，能结合停留时间与地点场景，刻画轨迹产生者的行为，如购物、吃饭、睡觉等。而文献[19]则提取出重要的驻点以及拐点作为一条轨迹的关键点。

(2) 基于关键线段提取的轨迹表征

类似于关键点表征，关键线段表征也是一种直观的轨迹表示方法。轨迹的关键线段同样可分为两种: (a) 全局关键线段。在城市中，车辆的交通轨迹应当被道路所约束，因此将轨迹表示为一系列道路段的拼接是一种直观而合理的方式，这也催生了一些路网匹配的方法[20,21,22]，同样，也有从轨迹集合中提取出线段的工作，例如文献[23]用分段聚类的方法，用最小描述长度(MDL)作为评价因子将轨迹中的代表性线段聚类找出。(b) 基于单条轨迹的关键线段。这一类的方法也较为直观，例如文献[24,15]将每条轨迹划分为行走段和非行走段，而文献[25,26,27]则用压缩的思想，将原轨迹段集合表示为最少的特征段，并在这个过程中保持最少的误差。

(3) 轨迹高层结构提取与表达

根据后续的挖掘任务，很多工作直接从轨迹数据中提取出高层的数据结构。例如为了轨迹检索，轨迹被投影到了树的结构上[28,29,30]，文献[31]则提出了用户与位置的二分图与树形层次图结构来表征原始轨迹中用户与地点的关系。文献[5,4,32]则用用户与地点和时间的三阶张量来对轨迹进行信息抽取。

现有的轨迹表征方法通常针对某种特定轨迹挖掘任务而设计的，其只考虑了某些特定领域的信息，没有将各种语义信息综合进表征中。且在预处理过程中，每条轨迹都是分开对待的，随着轨迹数据的飞速增长，相似的轨迹与地点将被冗余地处理多次，这种分开处理的策略将占用越来越多的存储空间和处理时间，因此，我们需要提出一种基于全局的轨迹表征。为了使这种全局表征方式更加完备，轨迹的相似性度量以及各种语义都应以某种方式融入到这种表征中来。

3. LBSNs 中的 POI 推荐系统

在 LBSNs 中，为用户推荐地点是一个非常实际的问题，这不仅能给用户提供合适的休闲产所和社交平台，还能同时结合 LBSNs 中富饶的信息（社交关系、签到或者轨迹历史信息等）来推测用户的偏好，故能完善用户画像，这能给维护信息的大型社交网络以及实体商业的商家带来大量利益。虽然传统的推荐系统已经被研究得很深入，并且成熟地投放进例如亚马逊、Netflix、淘宝等大型网站中，但是 POI 推荐有着独特的性质，使得这些传统推荐方法不能直接应用到 POI 推荐中。在这里，我们首先总结 POI 推荐独有的性质。再总结目前四种主流的 POI 推荐系统。

3.1 POI 推荐的特性

(1) 地理影响

如同 1970 年的 Tobler 第一地理定律指出:“任何东西都有关联的东西，但距离近的东西关联的更多”[33]。对于 LBSNs，Tobler 第一定律表明用户会更为偏好距离近的 POI 而非距离远的 POI，且喜欢已经喜欢的 POI 附近的 POI 的概率更大。事实上，地理影响对用户的偏 好影响是最大的。

(2) 隐式反馈与稀疏性

在传统的推荐系统中，用户通常将自己对商品(书、电影、音乐等)的偏好表达为一个评分矩阵，其中每一个元素都是一个固定范围内的数值(如 [1,5])，越大则表示某用户对该商品的喜好程度越高。而在 POI 推荐系统中，没有这样的 评分矩阵数据，用户对于地点的交互数据只是一个访问频率，这

个数值是离散的整数。这也就造成一个问题，大多数用户频繁访问的地点都是极少的，这些地方可能访问量高达上千次。而对于其他地方，用户可能仅仅访问过一到两次，这并不能体现用户对该地点的喜爱程度。举一个例子，在 Netflix 电影推荐数据集中，数据的空缺值占 99%，但在 Gowalla 签到数据中，有值的地方仅占 $2.08 \times 10{-}4$。这种访问频率的极大反差表现了 POI 推荐中的一大挑战。

(3) 社交属性

通过 LSBNs 中的数据可以观察得到假设:用户会从其朋友处得到并接受喜好推荐的建议，传统推荐系统将社交关系与评分洗好结合起来以增强推荐系统的性能。一些工作[34, 35]也表明了社交关系数据的融入确实能增强推荐系统的性能。然而在 POI 推荐系统中，有工作[36]表明 96%的用户分享了不到 10%的公共地点，这表明大量的用户并没有分享 POI 给自己的朋友，因此，社交网络在 POI 推荐中的影响并不如传统推荐那样明显。

3.2 LBSNs 中的 POI 推荐系统

(1) 基于纯地点的 POI 推荐系统

传统的推荐系统的思想很朴实，给用户推荐商品，并不用局限于社交网络下的地点推荐系统(LBSNs)。然而，由于用户在社交网络中的签到数据的地理信息可以挖掘出用户某些偏好，2011 年左右，文献[43,44]开始将传统推荐系统算法应用于推测 POI 中。此时的推荐仍然是考虑最经典的协同过滤思想，将 POIs 视为商品，那基于用户的推荐系统[45]和基于商品的推荐系统[46,47]分别考虑相似用户和相似商品进行推荐。之后，随着基于矩阵分解的推荐方法开始流行[48,49,50]，文献[44]提出了加约束项的矩阵分解的 POI 推荐方法，文献[51]基于概率矩阵分解(PMF)与概率因子模型(PFM)的 POI 推荐系统。这些方法仅仅基于 LBSNs 中用户与地点的交互数据，没有考虑其他方面的语义。因此，后续工作开始加入其他方面的语义信息来增强推荐性能。

(2) 考虑地理影响的 POI 推荐系统

在地点推荐系统中，自然要考虑地理信息对用户决策的影响。1970 年的托比第一地理定律指出:“任何东西都有关联的东西，但距离近的东西关联的更多”[33]。这给了地点推荐系统两个指示:(a)用户访问地点应该遵循就近原则。(b)用户对自己喜欢的地点周边的地点更加刚兴趣。同时，大量的文献[46, 47, 37, 48, 49]也在研究中发现了空间位置聚集的现象，并把这些规律应用于地点推荐系统中，对已有模型进行了改进。其中文献 [48, 37]将用户访问两个地点之间的概率建模为两个地点之间距离的幂律分布。而[45]则将这一概率建模为多维混合高斯分布，且用实验表明其性能好于纯地点推荐系统中的 PMF 与 PFM 方法。而[49]则引入了核函数来建模这一概率，且用实验表明其性能好于用幂律分布来进行建模的 POI 推荐系统。而文献[47]进一步提出了基于贝叶斯概率的非负矩阵分解算法(BNMF)来对地理影响进行刻画，且用实验表明齐心更好于 PMF 与 RMF 算法。

(3) 考虑社交网络的 POI 推荐系统

事实上，基于用户社交关系的推荐系统在 LBSN 概念前久被广泛使用了，分为基于存储的[50, 51, 52]与基于模型[34, 35]的两种方法。之后，文献[45, 36]将这些方法推广到了地点推荐系统中，并取得了良好的效果。具体地，文献[36]提出了基于朋友的 POI 协同过滤推荐算法(FCF)，其考虑了朋友的偏好，而非 LSBNs 中其他非朋友用户的爱好。注意 FCF 更加注重推荐的准确性而非可行性，这意味

着 FCF 可能不能产生很多推荐，但能保证推荐的质量。而文献[45]提出了加入社交约束的概率矩阵分解方法(PMFSR)，做出了朋友之间的隐向量应该尽量相似的假设，并将这化为一个约束带入到普通的矩阵分解方法中。

(4) 考虑时间影响的 POI 推荐系统

在传统的推荐系统中存在考虑时间影响的工作，比如有基于矩阵分解[53]的方法和随机游走[54]的方法。在这些传统方法中，时间影响是以衰减因子的形式加入到推荐系统中的。相反地，在 POI 推荐中，时间影响则是以推荐不同时间阶段的 POI 作为出发点。其中文献[48]考虑了用户倾向于在不同时间阶段访问不同类型的 POI，于是提出了时间敏感的 POI 推荐系统，并在实验中表明了加入时间的 POI 推荐系统系统性能好于不加的。而文献[55]则进一步加入不同时间阶段的用户偏好约束:用户在一天中的不同阶段考虑的 POI 类型是不同的，而在相邻时间阶段的偏好则是相似的，于是他们对用户在不同时间阶段的偏好做了约束，并取得了不错的效果。

现有方法大多数是利用社交网络、地理信息来对地点进行推荐，然而在推荐的过程中却没有考虑到其他信息的数据是否全面，是否与用户地点交互数据一致的问题。于是一个问题浮现:能否考虑一个地点推荐系统，不仅将社交网络以及地理信息用于推荐，反之还能用推荐信息来补齐社交网络以及地理信息(社团挖掘问题、区域划分问题)，这样能使得信息最大效益的利用，而不同的任务也能互相促进。另一方面，无论是社交网络，还是用户地点交互数据都是非常稀疏的，对于 check-in 数据，其交互元素大多是用户访问特定地点的次数，而大量的空缺值中，如何将用户不感兴趣的负样本从潜在的未访问的元素中区分开来，是一个困扰推荐系统的问题。

4. 存在挑战与待解决关键问题

(1) 现有轨迹表征方法未考虑 LBSNs 中的多源语义

现有的基于 LBSNs 进行轨迹挖掘或者 POI 推荐的工作都将轨迹表征问题进行简化，其仅仅考虑单方面的语义信息，这使得轨迹的表征单调，并且没有良好的解决稀疏、长度不固定和噪声多的问题。这使得不同轨迹挖掘或 POI 推荐的工作的数据集不能共享，也使得其挖掘或推荐的性能依赖于表征的好坏。如何融入 LBSNs 中丰富的语义信息是一个急需解决的问题。

(2) 未充分挖掘 POI 在地理分布中的层级结构

几乎所有的 POI 推荐工作都将 POI 视为地图中的点集，而在事实上，POI 的分布是具有层次性的。比如一个 POI 属于一条特定街道，而这条街道属于某个区，进而属于某个城市。在推荐中，城市、区和街道对 POI 也是有影响的，这种影响应该用一个层次结构刻画并融入到 POI 的推荐工作中。

(3) LBSNs 中的社交网络数据稀疏，提升 POI 推荐效果不明显

为了解决 POI 推荐中用户与地点交互的稀疏问题，很多工作引入 LBSNs 中的社交网络。但是，社交网络的信息本身也是稀疏和缺失的，这就造成了上文提到的社交网络不能很好的促进 POI 推荐这一现象。要使得社交网络能够促进 POI 推荐算法的正确性，需要首先考虑解决社交网络数据本身的稀疏和缺失性问题。

基于以上三个问题，我将借鉴目前已有算法，分别提出相应的解决方法，并将在后续的毕设中实现。我将用综述中提到的现有算法作为对比算法，在多个数据集上综合测试，并在测试中不断改进自己的算法。

参考文献

[1] F. Gianno i, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pa ern mining, " in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2007, pp. 330--339.

[2] L.-Y. Wei, Y. Zheng, and W.-C. Peng, "Constructing popular routes from uncertain trajectories, " in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 195--203.

[3] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city, " in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014, pp. 1027--1036.

[4] Y.Wang, Y.Zheng, and Y.Xue, "Travel time estimation of a path using sparse trajectories, " in Proceedings of the 20th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, 2014, pp. 25--34.

[5] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories, " IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 3, pp. 712--725, 2015.

[6] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Huang, and Z. Xu, "Destination prediction by sub-trajectory synthesis and privacy protection against such prediction, " in 2013 IEEE 29th International Conference on Data Engineering (ICDE), 2013, pp. 254--265.

[7] B.Zheng, N.J.Yuan, K.Zheng, X.Xie, S.Sadiq, and X.Zhou, "Approximate keyword search in semantic trajectory database, "2015 IEEE 31st International Conference on Data Engineering (ICDE), 2015, pp. 975--986.

[8] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks, " in Proceedings of the 17th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, 2011, pp. 1082--1090.

[9] D. Ashbrookand T. Starner, "Learning significant locations and predicting user movement with gps, " in Proceedings. Sixth International Symposium on Wearable Computers, 2002.(ISWC 2002), 2002, pp. 101--108.

[10] D. Ashbrook and T. Starner, "Using gps to learn significant locations and predict movement across multiple users, " Personal and Ubiquitous computing, vol. 7, no. 5, pp. 275--286, 2003.

[11] Z.Li, B.Ding, J.Han, R.Kays, and P.Nye, "Mining periodic behaviors for moving objects, " in Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010, pp. 1099--1108.

[12] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma, "Recommending friends and locations based on individual location history, " ACM Transactions on the Web (TWEB), vol. 5, no. 1, p. 5, 2011.

[13] H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen, "Discovery of convoys in trajectory databases, " Proceedings of the VLDB Endowment, vol. 1, no. 1, pp. 1068-1080, 2008.
[14] Z. Chen, H. T. Shen, and X. Zhou, "Discovering popular routes from trajectories, " in Proceedings of the 27th International Conference on Data Engineering, 2011, pp. 900--911.
[15] Y.Zheng, Q.Li, Y.Chen, X.Xie, and W.-Y.Ma, "Understanding mobility based on gps data, " in Proceedings of the 10th international conference on Ubiquitous computing. ACM, 2008, pp. 312--321.
[16]Y.Wang, N.J.Yuan, D.Lian, L.Xu, X.Xie, E.Chen, and Y.Rui, "R"egularity and conformity: Location prediction using heterogeneous mobility data, " in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1275--1284.
[17] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, "Mining user similarity based on location history, " in Proceedings of the 16th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, 2008, p. 34.
[18] Y. Zheng and X. Xie, "Learning travel recommendations from user-generated gps traces, " ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 1, p. 2, 2011.
[19] N. Adrienko and G. Adrienko, "Spatial generalization and aggregation of massive movement data, " IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 2, pp. 205--219, 2011.
[20] J. S. Greenfeld, "Matching gps observations to locations on a digital map, " in 81th annual meeting of the transportation research board, vol. 1, no. 3, 2002, pp. 164--173.
[21] W. Chen, M. Yu, Z. Li, and Y. Chen, "Integrated vehicle navigation system for urban applications, " 2003.
[22] P.NewsonandJ.Krumm, "Hiddenmarkovmapmatchingthroughnoiseandsparseness, " in Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM, 2009, pp. 336--343.
[23] J.-G. Lee, J. Han, and K.-Y. Whang, "Trajectory clustering: a partition-and-group framework, " in Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, 2007, pp. 593--604.
[24] Y. Zheng, L. Liu, L. Wang, and X. Xie, "Learning transportation mode from raw gps data for geographic applications on the web, " in Proceedings of the 17th International Conference on World Wide Web, 2008, pp. 247--256.
[25] W.-C. Lee and J. Krumm, "Trajectory preprocessing, " in Computing with Spatial Trajectories, 2011, pp. 3--33.
[26] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, " Cartographica: The International Journal for Geographic Information and Geovisualization, vol. 10, no. 2, pp. 112--122, 1973.
[27] R. Bellman, "On the approximation of curves by line segments using dynamic programming, " Communications of the ACM, vol. 4, no. 6, p. 284, 1961.
[28] A. Gu man, R-trees: a dynamic index structure for spatial searching. ACM, 1984, vol. 14, no. 2.
[29] L.Wang, Y.Zheng, X.Xie, and W.-Y.Ma, "A flexible spatio-temporal indexing scheme for large-scale gps track retrieval, " in 9th International Conference on Mobile Data Management, 2008, pp. 1--8.

[30] D. Pfoser, C. S. Jensen, Y. Theodoridis et al., "Novel approaches to the indexing of moving object trajectories." in VLDB, 2000, pp. 395--406.
[31] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories, " in Proceedings of the 18th International Conference on World Wide Web, 2009, pp. 791--800.
[32] Y. Liu, C. Liu, B. Liu, M. Qu, and H. Xiong, "Unified point-of-interest recommendation with temporal interval assessment, " in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1015-1024.
[33] W. R. Tobler, "A computer movie simulating urban growth in the detroit region, " Economic geography, vol. 46, no. sup1, pp. 234--240, 1970.
[34] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks, " in Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010, pp. 135--142.
[35] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: social recommendation using probabilistic matrix factorization, " in Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008, pp. 931--940.
[36] M. Ye, P. Yin, and W.-C. Lee, "Location recommendation for location-based social networks, " in Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. ACM, 2010, pp. 458--461.
[37] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation, " in Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, 2011, pp. 325--334.
[38] B.BerjaniandT.Strufe, "Arecommendationsystemforspotsinlocation-basedonline social networks, " in Proceedings of the 4th Workshop on Social Network Systems. ACM, 2011, p. 4.
[39] J.S.Breese, D.Heckerman, and C.Kadie, "Empirical analysis of predictive algorithms for collaborative filtering, " in Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1998, pp. 43--52.
[40] B.Sarwar, G.Karypis, J.Konstan, and J.Riedl, "Item-based collaborative filtering recommendation algorithms, " in Proceedings of the 10th international conference on World Wide Web. ACM, 2001, pp. 285--295.
[41] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering, " IEEE Internet computing, vol. 7, no. 1, pp. 76--80, 2003.
[42] A.MnihandR.R.Salakhutdinov, "Probabilisticmatrixfactorization, "inAdvancesin neural information processing systems, 2008, pp. 1257--1264.
[43] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization, " in Advances in neural information processing systems, 2001, pp. 556--562.
[44] Y.Koren, R.Bell, and C.Volinsky, "Matrix factorization techniques for recommender systems, " Computer, vol. 42, no. 8, 2009.
[45] C.Cheng, H.Yang, I.King, and M.R.Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks." in Aaai, vol. 12, 2012, pp. 17--23.

[46] H. Gao, J. Tang, and H. Liu, "gscorr: modeling geo-social correlations for new checkins on location-based social networks, " in Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012, pp. 1582-1586.
[47] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-ofinterest recommendation, " in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013, pp. 1043--1051.
[48] Q.Yuan, G.Cong, Z.Ma, A.Sun, and N.M.Thalmann, "Time-aware point-of-interest recommendation, " in Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2013, pp. 363--372.
[49] J.-D.Zhang, C.-Y.Chow, and Y.Li, "i"georec: A personalized and efficient geographical location recommendation framework, " IEEE Transactions on Services Computing, vol. 8, no. 5, pp. 701--714, 2015.
[50] J. Golbeck, "Generating predictive movie recommendations from trust in social networks, " Trust Management, pp. 93--104, 2006.
[51] P. Massa and P. Avesani, "Trust-aware recommender systems, " in Proceedings of the 2007 ACM conference on Recommender systems. ACM, 2007, pp. 17--24.
[52] M. Jamali and M. Ester, "Trustwalker: a random walk model for combining trustbased and item-based recommendation, " in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009, pp. 397--406.
[53] Y. Koren, "Collaborative filtering with temporal dynamics, " Communications of the ACM, vol. 53, no. 4, pp. 89--97, 2010.
[54] L. Xiang, Q. Yuan, S. Zhao, L. Chen, X. Zhang, Q. Yang, and J. Sun, "Temporal recommendation on graphs via long-and short-term preference fusion, " in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010, pp. 723--732.
[55] H.Gao, J.Tang, X.Hu, and H.Liu, "Exploring temporal effects for location recommendation on location-based social networks, " in Proceedings of the 7th ACM conference on Recommender systems. ACM, 2013, pp. 93--100.
[56] J.Shao, X.He, C.Böhm, Q.Yang, and C.Plant, "Synchronization-inspired partitioning and hierarchical clustering, " IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 4, pp. 893--905, 2013.
[57] J. Shao, Z. Han, Q. Yang, and T. Zhou, "Community detection based on distance dynamics, " in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1075--1084.
[58] J. Shao, Q. Yang, H.-V. Dang, B. Schmidt, and S. Kramer, "Scalable clustering by iterative partitioning and point a ractor representation, " ACM Transactions on Knowledge Discovery from Data, vol. 11, no. 1, p. 5, 2016.
[59] J. Shao, F. Huang, Q. Yang, and G. Luo, "Robust prototype-based learning on data streams, " IEEE Transactions on Knowledge and Data Engineering, 2017.
[60] J. Shao, C. Gao, W. Zeng, J. Song, and Q. Yang, "Synchronization-inspired coclustering and its application to gene expression data, " in 2017 IEEE 11th International Conference on Data Mining (ICDM). IEEE.

[61] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model, " Journal of machine learning research, vol. 3, no. Feb, pp. 1137--1155, 2003.

[62] R.Collobert, J.Weston, L.Bo ou, M.Karlen, K.Kavukcuoglu, and P.Kuksa, "Natural language processing (almost) from scratch, " Journal of Machine Learning Research, vol. 12, no. Aug, pp. 2493--2537, 2011.

[63] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality, " in Advances in Neural Information Processing Systems 26, 2013, pp. 3111--3119.

[64] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation, " in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532--1543.

Abstract

With the rapid growing of mobile devices as well as the storage capability and computation capability, the trajectory data and location-based check in data are mounting in a dazzling speed. To leverage these data can benefit fields ranged from economy, ecology to transportation and so on. There are a myriad of spatial-temporal data representation and knowledge mining tasks emerged both in academia and industry. Point-of-Interest (POI) recommendation in Location-Based Social Networks (LBSNs) is a prominent problem that call for attentions, since well tackled this problem will bring significant improvement to human daily life and help LBSNs server providers to increase revenues. Differing from traditional recommender systems, POI recommender systems have some unique characteristics. In this paper, I present a review of existing POI recommendation algorithms and discuss some research directions for POIs recommendation. According to the type of additional information integrated with check-in data by POI recommendation algorithms, I classify POI recommendation algorithms into four categories: pure check-in data based POI recommendation approaches, geographical influence enhanced POI recommendation approaches, social influence enhanced POI recommendation approaches and temporal influence enhanced POI recommendation approaches.

(1) Pure check-in data based POI recommendation. Traditional recommender systems make recommendations by exploiting explicit ratings for items, which are not available in LBSNs. However, the frequencies of check-in recorded by LBSNs implicitly reflect users' preferences for POI. Hence, in order to produce POI recommendations, several studies adopted traditional recommendation algorithms to infer users' personalized tastes for POI by mining the check-in patterns of users. And those methods can be generally divided as Item-based and User-based POI recommendation. The former assumes that similar users have similar tastes for locations and makes POI recommendations based on the opinions of most similar neighbors. On the other hand, item-based POI recommendation approach assumes that users are interested in similar POIs.

(2) Geographical influence enhanced POI recommendation. In LBSNs, there are physical interactions between users and POIs, which is a unique property distinguishing POI recommendation from traditional item recommendation. Moreover, the Tobler's First Law of Geography reported that "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). The Tobler's First Law of Geography is also represented as geo- graphical clustering phenomenon in users' check-in activities. Two intuitions contribute this phenomenon: 1) users prefer to visit nearby POIs rather than distant ones; 2) users may be interested in POIs surrounded a POI that users prefer. Several studies argue that geographical clustering phenomenon in users' check-in activities, known as geographical influence, can be utilized to improve the POI recommender systems.

(3) Social influence enhanced POI recommendation. Social influence enhanced recommendation approaches have been extensively explored in traditional recommender systems, include memory-based methods and model-based methods. Inspired by the assumption that friends of LBSNs share more common interests than non-friends, several POI recommendation approaches improve the quality of recommendation by taking social influence into consideration. Although social influence shows an important impact on the performance of traditional recommender system, the experimental results of above mentioned social influence enhanced POI recommendation approaches show that social influence weights litter than geographical influence and check-in activities.

(4) Temporal influence enhanced POI Recommendation. There exists studies that consider temporal influence in traditional recommender systems, such as matrix factorization based approach, random walk based approach. However, in traditional recommendation systems, temporal influence is used to as a factor that decays the weights of ratings. On the contrary, POI recommendation systems generally use temporal influence to make POI recommendation for a specific temporal state.

The common issues and disadvantages of these models are listed, which included but not limited to following points: (a) The trajectory data and other spatial-temporal data are naturally faulty duo to the non-fixed length, volatile sampling rate and other factors. Although the existing methods have proposed many feature extraction and pre-processing models to tackle those drawbacks, almost all of them only considered one or few aspects of rich semantic information of spatial-temporal data, which in turn leads to a bunch of non-thoughtful trajectory representation frameworks. (b) Geographical information takes the leader part in enhancing the performance of POI recommendation. However, almost all methods consider POIs as sole points on a flat space, which is not sensible since there is a hierarchical structure existing in geographical space, i.e. each POI belongs to a street, a region and a city. When recommending POIs to user, the hierarchical structure should be taken into account. (c) Besides, the data from social network, geographical statistics and temporal side may suffer from sparsity and noisy, which will limit the effectiveness and accuracy when the data are considered in POI recommendation.

After summarized these issues, some promising directions are listed, First, in LBSNs, the frequencies of check-in for POIs vary dramatically and users' check- in frequency intuitively reflects the degree of users' preferences for POIs, and Rank-based collaborative filtering approaches may be applicable to POI recommendation since rank-based collaborative filtering approaches infer users' preferences from pairwise comparisons rather than numerical ratings. Second, since users often involve several social networks, information derived from other social networks would be beneficial for POI recommendation in LBSNs. In this case, POI recommendation based on transfer learning is a potential research direction. Third, LBSNs provide rich additional information for enhancing the performance of POI recommender systems, i.e. check-ins, geographical information, social relationships and temporal information. A unified recommendation framework is desirable for POI recommender systems to boost the performance of POI recommender systems by joint all kinds of additional information. Finally, As the rapidly growing amount of users and POIs available in LBSNs, POI recommender systems suffer seriously from scalability problem. Hence, parallelized computing methods, e.g. MapReduce and Spark, are worthy of exploiting to speed up the computation process of POI recommendation. Following those directions, I will propose several models to conduct POI recommendation and complete my graduate project.

The rest of this paper is organized as follows. I formalize the problem of POI recommendation in LBSNs. First I summarize the trajectory representation methods as three classes. Then four types of POI recommendation framework: pure check-in data based POI recommendation approaches, geographical influence enhanced POI recommendation approaches, social influence enhanced POI recommendation approaches and temporal influence enhanced POI recommendation approaches are surveyed one by one, I discuss the merit and drawbacks of each category and give the idea of my work.